in conjunction with the Medical Image Computing and Computer-Assisted Interventions Conference

2017 International MICCAI BraTS Challenge

September 14, 2017

Quebec City, Quebec, Canada

(pre-conference proceedings)

Scope

BraTS has always been focused on the evaluation of state-of-the-art methods for the segmentation of brain tumors in magnetic resonance imaging (MRI) scans. **BraTS 2017** utilizes multi-institutional pre-operative MRI scans and **focuses on the segmentation of** intrinsically heterogeneous (in appearance, shape, and histology) **brain tumors**, namely gliomas. Furthermore, in order to pinpoint the clinical relevance of this segmentation task, BraTS'17 also focuses **on the prediction of patient overall survival**, via integrative analyses of radiomic features and machine learning algorithms.

Clinical Relevance

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. This intrinsic heterogeneity of gliomas is also portrayed in their imaging phenotype (appearance and shape), as their sub-regions are described by varying intensity profiles disseminated across multimodal MRI scans, reflecting varying tumor biological properties. Due to this highly heterogeneous appearance and shape, segmentation of brain tumors in multimodal MRI scans is one of the most challenging tasks in medical image analysis.

There is a growing body of literature on computational algorithms addressing this important task. Unfortunately, open data sets for designing and testing these algorithms are not currently available, and private data sets differ so widely that it is hard to compare the different segmentation strategies that have been reported so far. Critical factors leading to these differences include, but not limited to, i) the imaging modalities employed, ii) the type of the tumor (GBM or LGG, primary or secondary tumors, solid or infiltratively growing), and iii) the state of disease (images may not only be acquired prior to treatment, but also post-operatively and therefore show radiotherapy effects and surgically-imposed cavities). Towards this end, BraTS is making available a large dataset with accompanying delineations of the relevant tumor sub-regions.

Tasks

Task 1: Segmentation of gliomas in pre-operative scans: The participants are called to address this task by using the provided clinically-acquired training data to develop their method and produce segmentation labels of the different glioma sub-regions. **The sub-regions considered for evaluation are: 1) the "enhancing tumor" (ET), 2) the "tumor core" (TC), and 3) the "whole tumor" (WT)** [see figure below]. The ET is described by areas that show hyper-intensity in T1Gd when compared to T1, but also when compared to "healthy" white matter in T1Gd. The TC describes the bulk of the tumor, which is what is typically resected. The TC entails the ET, as well as the necrotic (fluid-filled) and the non-enhancing (solid) parts of the tumor. The appearance of the necrotic (NCR) and the non-enhancing (NET) tumor core is typically hypo-intense in T1-Gd when compared to T1. The WT describes the complete extent of the disease, as it entails the TC and the peritumoral edema (ED), which is typically depicted by hyper-intense signal in FLAIR. The labels in the provided data are: 1 for NCR & NET, 2 for ED, 4 for ET, and 0 for everything else.

Task 2: Prediction of patient overall survival (OS) from pre-operative scans: Once the participants produce their segmentation labels in the pre-operative scans, they will be called to use these labels to extract imaging features from the given MRI data, in an attempt to predict patient OS. The participants do not need to be limited to volumetric parameters, but can also consider intensity, morphologic, histogram-based, and textural features, as well as spatial information, and glioma diffusion properties extracted from glioma growth models. Three groups of survival are considered, i.e. long-survivors (e.g., >15 months), short-survivors (e.g., <10 months), and mid-survivors (e.g. between 10 and 15 months).

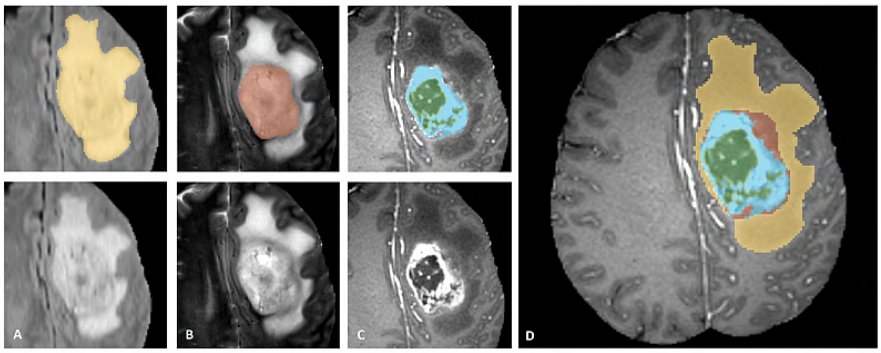


Fig.1: Glioma sub-regions. Shown are image patches with the tumor sub-regions that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (Fig.A), the tumor core (red) visible in T2 (Fig.B), the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C). The segmentations are combined to generate the final labels of the tumor sub-regions (Fig.D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue). (Figure taken from the BraTS IEEE TMI paper.)

Data

The datasets used in this year's challenge have been updated, since BraTS'16, with more routine clinically-acquired 3T multimodal MRI scans and all the ground truth labels have been manually-revised by expert board-certified neuroradiologists.
Ample multi-institutional routine clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG), with pathologically confirmed diagnosis and available OS, will be provided as the training, validation and testing data for this year's BraTS challenge. These multimodal scans describe a) native (**T1**) and b) post-contrast T1-weighted (**T1Gd**), c) T2-weighted (**T2**), and d) T2 Fluid Attenuated Inversion Recovery (**FLAIR**) volumes, and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions, mentioned as data contributors below. All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing tumor (NCR/NET — label 1), as described in the BraTS reference paper, published in IEEE Transactions for Medical Imaging (Fig.1). The provided data are distributed after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution (1 mm^3) and skull-stripped. The data provided during BraTS'17 differs significantly from the data provided during the previous BraTS challenges. Specifically, this year, expert neuroradiologists have radiologically assessed the complete original TCIA glioma collections and categorized each scan as pre- or post-operative. Subsequently, all the pre-operative TCIA scans were annotated by experts for the various sub-regions and included in this year's BraTS datasets. Participants are not allowed to use additional private data (from their own institutions) for data augmentation, since our intentions are to provide a fair comparison among the participating methods. The only case that this will be considered as a valid contribution is if they also report results using only the BraTS'17 data and discuss any potential difference in the results.
Validation data was released on June 30, allowing participants to obtain preliminary results in unseen data and also report it in their submitted papers, in addition to their cross-validated results on the training data. The ground truth of the validation data was not provided to the participants. Finally, all participants were presented with the same test data, for a limited controlled time-window (48h), before the participants are required to upload their final results in CBICA's IPP (ipp.cbica.upenn.edu).

Organizing Committee

- Spyridon (Spyros) Bakas, Ph.D., — [*Lead Organizer*]
  Center for Biomedical Image Computing and Analytics (CBICA), SBIA, UPenn, PA, USA
- Bjoern Menze, Ph.D.,
  Technical University of Munich (TUM), Germany
- Christos Davatzikos, Ph.D.,
  Center for Biomedical Image Computing and Analytics (CBICA), SBIA, UPenn, PA, USA
- Mauricio Reyes, Ph.D.,
  University of Bern, Switzerland
- Keyvan Farahani, Ph.D.,
  Cancer Imaging Program, National Cancer Institute (NCI), National Institutes of Health, USA

Data Contributors

- John B. Freymann & Justin S. Kirby - on behalf of The Cancer Imaging Archive (TCIA),
  Cancer Imaging Program, NCI, National Institutes of Health (NIH), USA
- Christos Davatzikos, Ph.D. & Spyridon (Spyros) Bakas, Ph.D.,
  Center for Biomedical Image Computing and Analytics (CBICA), SBIA, UPenn, PA, USA
- Hassan Fathallah-Shaykh, M.D., Ph.D.,
  University of Alabama at Birmingham, AL, USA
- Roland Wiest, M.D.,
  University of Bern, Switzerland
- Andras Jakab, M.D., Ph.D.,
  University of Debrecen, Hungary
- Marc-Andre Weber, M.D.,
  Heidelberg University, Germany

Brain Tumor Segmentation from Multi Modal MR images using Fully Convolutional Neural Network

Varghese Alex, Mohammed Safwan, and Ganapathy Krishnamurthi

Indian Institute of Technology Madras, Chennai, India,
gankrish@iitm.ac.in

Abstract. This paper explains the use of a fully convolutional neural network (FCNN) for segmentation of gliomas from Magnetic Resonance Images (MRI). A fully automatic voxel based classification was achieved by training a 23 layer deep FCNN on 2-D slices extracted from patient volumes. The network was trained on slices extracted from 130 patients and validated on 50 patients. The false positives in segmentation map generated by the FCNN were removed by connected component analysis. On the BraTS 2017 validation set, the proposed algorithm achieved a mean whole tumor, tumor core & active dice score of 0.83, 0.69 & 0.69 respectively.

Keywords: Deep Learning, Gliomas, MRI, FCN

1 Introduction

In this paper, we propose a 23 layer deep FCNN for the task of segmentation of gliomas from MR scans. In the field of medical image analysis, one of the oft used architecture is the the U-net architecture [1]. The network used in this work has a similar architecture as that of U-net. The network was trained on 2-D axial slices (240*240) extracted from FLAIR, T2, T1, and T1 post contrast sequences. The FCNN architecture enables to classify all the voxels in a slice using a single forward pass thereby reducing the time required for prediction considerably as opposed to a patch based technique.

Convolutional neural network and its variants being deterministic approaches tend to mis-classify voxels as lesion in regions like brain stem, cerebellum where occurrence of gliomas is anatomically impossible. We utilize connected component analysis to discard components below a certain threshold for false positive reduction. For the prognosis challenge, the overall survival rate was categorized into three groups namely short survivors (prognosis, $p < 3$ months), mid survivors (3 months $< p < 6$ months) & long survivors ($p > 6$ months). Features such as age, size of the lesion, relative size of various components of the lesion, extent of tumor, etc. were extracted from the label map generated by FCNN & fed as input to an SVM to predict the patient as short, mid or long survivor.

2 Materials and Methods

The proposed technique comprises of following stages:

1. Pre-processing of data
2. Training stage
3. Testing stage
4. Post-processing
5. Feature extraction for survival rate prediction.
6. Training & testing of SVM classifier for survival rate prediction.

The flowchart of the proposed the technique is given in Fig. (1).

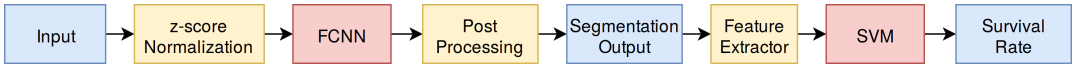


Fig. 1: Flow chart of the proposed network

2.1 Data

The network was trained and validated on the BraTS 2017 training data [2],[3], [4],[5]. The training data comprises of 210 HGG volumes and 75 LGG volumes collected from multiple centers. Each patient comprises of FLAIR, T2, T1, T1 post contrast and the associated ground truth labeled by experts. Each sequence was skull stripped and was re-sampled to 1mm*1mm*1mm (isotropic resolution).

For the overall survival challenge, age & prognosis of the patient post treatment were supplied by the organizers. The training set for the challenge comprised of 163 High Grade Glioma patients of which 84 patients had survival rate between 180 and 540 days (mid survivors), while 41 patients had prognosis less than 180 days (short survivors) and 38 patients had prognosis greater than 540 days (long survivors).

2.2 Fully Convolutional Neural Network

A typical FCNN comprises of convolution operations, max pooling layers and transpose convolution layers. The absence of fully connected layers in FCNNs leads to reduction of number of parameters in the network & enables feeding of inputs of arbitrary sizes. The max pooling layer helps in reduction of dimension of the feature maps in the deeper layers and also aid in capturing translation invariant features from the input images.

The dimensionality of the feature maps are brought back to size of the input by either using up-sampling modules such as bilinear interpolation of feature maps or transposed convolution. The use of transposed convolution in the networks makes the scaling procedure of feature maps a parameter to be learned

during the training process. Concatenation of feature maps or skip connections at various depths of the network leads to improvement of the network's performance.

FCNNs have an inherent advantage of classifying all pixels in the image by using single forward pass of the image and thus makes FCNNs an ideal choice for semantic segmentation related task. Similar to traditional CNNs, the parameters of the network are learned by minimizing the cross entropy.

3 Preprocessing of Data

3.1 Z-score Normalization

Multi center data and magnetic field inhomogeneities contribute to the intensity inhomogeneities in MR image. The volumes were normalized to have zero mean and unit variance.

4 Training of Network

4.1 Slice Extraction

Majority of the volumes in the dataset were acquired along the axial plane and hence had highest resolution this plane. Due to this reason, the networks were trained on axial slice extracted from all four sequences. The data imbalance in the dataset was addressed by training the network on slices that comprise of at least one voxel of lesion. The network was trained and validated on 7000 & 3000 axial slices respectively.

4.2 Network Architecture

The architecture of the network is given in Fig. (2 (a)). The numbers within each **Conv** block comprises of 2 sets of convolution by 3x3 kernels, batch normalization and a non linearity (ReLU), (Fig. (2 (b))). The number of learnable filters in each layer is depicted by the suffix in the **Conv** and **UpConv** block. The concatenation of feature maps is presented in the architecture as blue arrows.

The stride, kernel size & padding of the transposed convolution are chosen so as to produce feature maps of similar height and width as that of the feature maps of the adjoining **Conv** block. This enables concatenation of feature maps without the need of cropping feature maps from the **Conv** block. The network makes use of convolution with 1x1 filters in the hindmost convolution block and results in generating the segmentation map.

Training The network was trained with slices extracted from 120 HGG patients. The weights and biases in each layer was initialized using the Xavier initialization [6]. The network was trained for 30 epochs and the weights and biases were

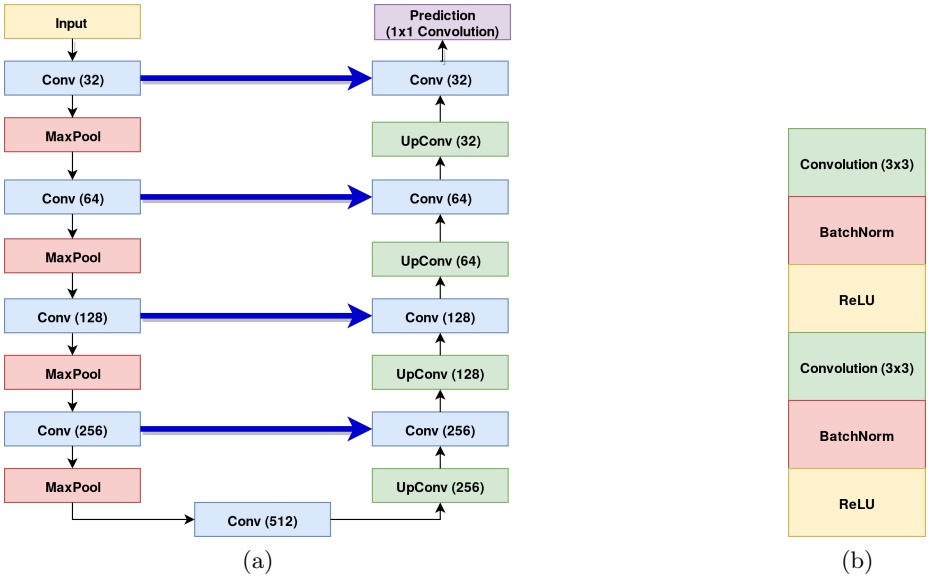


Fig. 2: Architecture of the proposed network. (a) Proposed FCNN. (b) Composition of the **Conv** block.

learned by minimizing the cross entropy loss function with ADAM [7] as the optimizer.

The class imbalance in the data was addressed by data augmentation and using a weighted cross entropy as the loss function. The data augmentation scheme comprises of horizontal flipping/ mirroring of the data & the corresponding ground truth. This augmentation scheme preserves the overall structure of brain. The weight assigned to normal:necrotic:edema:enhancing was in the order of 1:5:2:3.

4.3 Post processing

The predictions made by the network were post processed using connected component analysis. The components below a threshold (T=2000) were discarded.

4.4 Survival prediction

The overall survival of subjects were binned into 3 categories/classes namely short survivors, mid survivors and long survivors. Features such as age of patient, ratio of number of voxels of edema to number of voxels of lesion, ratio of number of voxels of necrosis to number of voxels of lesion, ratio of number of voxels of enhancing tumor to number of voxels of lesion, etc. were fed as input to a Support Vector Machine with a linear kernel. The list of normalized feature to train the SVM is given in Table (1).

Table 1: Features used to train SVM for survival rate. # -Number of voxels.

Sl.No	Feature	Normalized Feature
1	Age	Age/100
2	Extent of Lesion	# Slices with lesion/ Total number of slices
3	Presence of Edema	0 or 1
4	Amount of edema	# Edema voxels/ # Lesion voxels
5	Amount of necrosis	# Necrotic voxels/ # Lesion voxels
6	Amount of enhancement	# Enhancing tumor voxels/ # Lesion voxels
7	Amount of lesion	# Lesion Voxels / # Voxels in the brain

5 Results

The performance of network on the entire BraTS on the local test set (n=40) is given in Table (2). Fig.(3) shows the performance of the network on 2 different patients from the local test data.

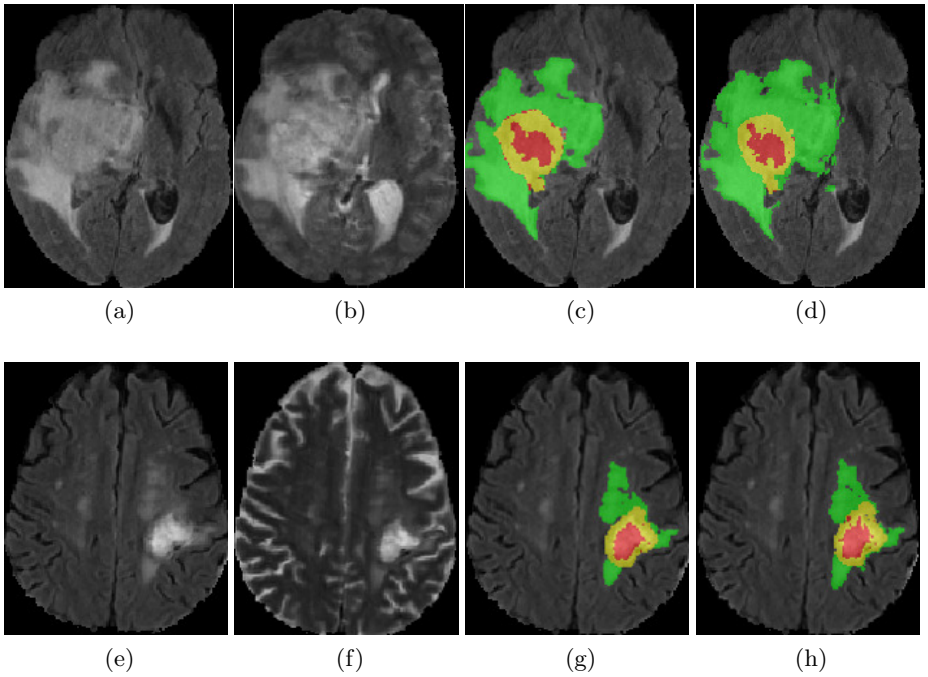
Table 2: Performance of the network on local test data

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.84	0.84	0.77
Std. Dev.	0.19	0.20	0.19
Median	0.90	0.90	0.83

The post processing technique improves the performance of the network. On the local test data, the improvement in performance was in the order of 2% for whole tumor dice score, 2 % for tumor core and 1% active tumor. Fig.(4) shows an example were the proposed post processing technique aids in eliminating false positives.

For the overall survival rate, using limited features, the SVM classifier was able to predict the overall survival rate of a patient with 60% accuracy. Larger number of mid survivors in the training data when compared to the other classes attributes to the classifier's better performance in predicting the mid survivors when compared to short or long survivors.

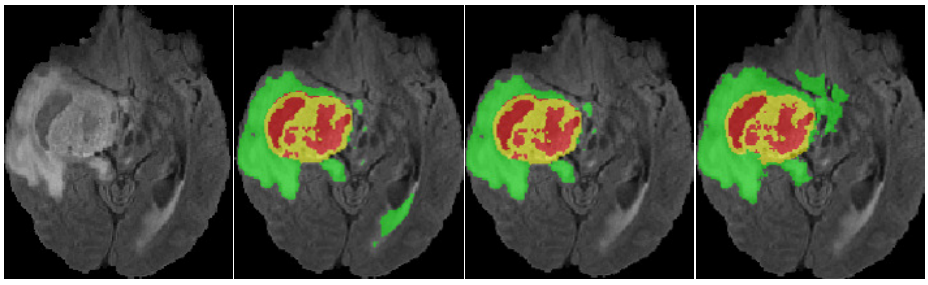
The performance of the network on the BraTS 2017 validation set is given in Table (3). It was observed that the network maintains similar whole tumor scores on the local test data and on the validation data. However, a dip in performance was observed in the tumor core & active tumor compartments.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3: Performance of the network on local test data. (a) FLAIR. (b) T2. (c) Prediction. (d) Ground Truth. (e) FLAIR. (f) T2. (g) Prediction. (h) Ground Truth. In figures c,d,g & h, green- Edema, yellow- Enhancing Tumor, red- Necrotic Core.



(a) (b) (c) (d)

Fig. 4: Reduction of False positive using connected components. (a) FLAIR. (b) Raw Prediction. (c) Post Processed image. (d) Ground Truth. In figures b, c, d, green- Edema, yellow- Enhancing Tumor, red- Necrotic Core.

Table 3: Performance of the network on BraTS 2017 validation data

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.83	0.69	0.69
Std. Dev.	0.16	0.30	0.32
Median	0.90	0.83	0.84

6 Conclusion

In this paper, we propose a fully automatic technique for segmetation of gliomas using FCNNs.

- An FCNN was trained to predict all voxels in a slice using a single forward pass.
- A single network was trained to segment both HGG and LGG volumes.
- Features from the segmentation map generated by the network was used to predict the overall survival rate of the patient. SVM with linear kernel was trained to classify the patient as short, mid or long survivor.

At present the authors plan to work on the following:

1. Improve performance of network in tumor core and active tumor regions.
2. Extract more features and & convert the prognosis task from a classification task to a regression task.

The networks were developed using Torch framework and was trained on Titan X. The entire pipeline (pre-processing, testing and post processing) takes approximately 30 seconds per patient.

References

1. Ronneberger O, Fischer P, Brox T. "U-net: Convolutional networks for biomedical image segmentation". In International Conference on Medical Image Computing and Computer-Assisted Intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.
2. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015
3. Bakas S et al. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]

4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
5. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
6. X. Glorot et al., "Understanding the difficulty of training deep feedforward neural networks, in International conference on artificial intelligence and statistics, pp. 249256, 2010
7. Kingma D, Ba J. "Adam: A method for stochastic optimization". arXiv preprint arXiv:1412.6980. 2014 Dec 22
8. Y. Bengio et al. "Greedy layer-wise training of deep networks." Advances in neural information processing systems 19 (2007): 153, 2007
9. G. Hinton et al.,"Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude". 2012

3D U-Nets For Brain Tumor Segmentation in MICCAI 2017 BraTS Challenge

Amorim, P. H. A. (pedro.amorim@einstein.br)[1*]; Chagas, V. S. (vinicius.chagas@inovacaoeinstein.com.br)[1*]; Escudero, G. G. (guilherme.escudero@einstein.br)[1*]; Oliveira, D. D. C. (diego.colombo@einstein.br)[1*]; Pereira, S. M. (silvio.pereira@einstein.br)[1*]; Santos, H. M. (heitor.santos@einstein.br)[1*] and Scussel, A. A. (artur.scussel@einstein.br)[1*]

[1] Hospital Israelita Albert Einstein, São Paulo, Brazil; *Indicates equal contribution

Abstract. In this paper we present our solution for brain tumor segmentation in MICCAI 2017 BraTS data set. Three different convolutional neural networks with the same 3D U-Net architecture were trained for each of the tumor segmentation targets (whole tumor, tumor core and enhancing tumor) with 3D patches as inputs. Preprocessing were done in each case separately, equalizing histogram on whole tumor and normalizing voxels on all modalities. Our solution yielded Dice coefficients of 0.9111, 0.9118 and 0.8272 on 30% of the training set (test split) and 0.8844, 0.7674 and 0.7261 on the leaderboard validation set (respectively for each segmentation target).

Keywords: MICCAI, BraTS, brain tumor segmentation, 3D U-Net, 3D patches

1. Introduction

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. Due to their highly heterogeneous appearance and shape, segmentation of brain tumors in multimodal MRI scans is one of the most challenging tasks in medical image analysis. [1]

Towards this end, BraTS is making available a large dataset with accompanying delineations of the relevant tumor sub-regions [1, 2, 3, 4].

In this work, we describe the methods used for the segmentation of each tumor region, as well as the results and future work.

2. Methods

To accomplish the segmentation task, we created 3 convolutional neural networks, each one for a target: whole tumor, tumor core, and enhancing tumor. All deep learning networks were implemented using Keras library in Python, with Theano as backend.

The dataset used to train these networks was the 2017 BraTS competition. It consists of 210 high grade glioma (HGG) and 75 low grade gliomas (LGG), all of them annotated by experts for all gliomas sub-regions[1, 2, 3, 4]. Only the HGG images were used to train our models. To analyze our results a split was made by patient, such that 70% of the data (147 patients) was the training set and 30% the validation set (63 patients).

A standard U-Net [5] was used to perform segmentation, but instead of 2D images, 3D images were used as input of the network.

Appropriate changes were made so that the network would be able to receive these kind of images. T1 and T2 weighted images had their histograms equalized in order to increase intensity contrast. On the other hand, post-contrast T1 and FLAIR images went through standard scaling and normalization before being used as inputs.

We briefly describe below the techniques used for the segmentation of each tumor region.

2.1. Whole tumor segmentation

Whole tumor regions correspond to the union of all tumor labels (1: necrotic and non-enhancing tumor; 2: peritumoral edema; and 4: GD-enhancing tumor), so their masks were built from that.

Segmentation was performed by a 3D U-Net convolutional neural network with 64 X 64 X 64 patches of T2, FLAIR, T1 and post-contrast T1 modalities as inputs (thus, the network has 4 64 X 64 X 64 channels as inputs).

Before being used as inputs, all images had their histograms equalized for brain voxels only, in order to increase contrast of brain tissues for all modalities. Only high grade glioma (HGG) cases were used as inputs.

2.2. Tumor core segmentation

Tumor core regions correspond to the union of necrotic and non-enhancing tumor (label 1) and GD-enhancing tumor (label 4) regions, so their masks were built from that.

Segmentation was performed by a 3D U-Net convolutional neural network with 64 X 64 X 64 patches of post-contrast T1 modality as inputs (thus, the network has 1 64 X 64 X 64 channel as input).

Only high grade gliomas (HGG) were used as training inputs, so the network was not able to find tumor core masks for low grade gliomas (LGG), yielding few labeled pixels as output. In those cases, the tumor core mask is replaced by the whole tumor mask in our pipeline to make up for that lack of a bigger segmented region.

2.3. Enhancing tumor segmentation

Gadolinium-enhancing tumor regions correspond to the ones labeled as number 4 on the dataset.

Segmentation was performed by a 3D U-Net convolutional neural network with 64 X 64 X 64 patches T1 and post-contrast T1 modalities as inputs (thus, the network has 2 64 X 64 X 64 channels as inputs).

Only high grade gliomas (HGG) were used as training inputs, so the network did not perform well for low grade gliomas (LGG), yielding few labeled pixels as output even when there was no enhancing tumor region. In those cases, the enhancing tumor mask is replaced by a blank mask (i.e. all-zeros) in our pipeline to make up for the network's inability to recognize tumors with no enhancing regions.

2.4. Overall survival estimation

The task was divided in two parts. The first is a multi class classification of short (> 10 months), medium (between 10 and 15 months) and long (> 15 months) overall survival. The second task is a regression one to predict the overall survival in days. To estimate overall survival in days for each patient, a XGBoost regressor [6] was used. Similarly, a XGBoost classifier [6] was chosen for the classification task on overall survival duration groups.

The model's input features include patient's age as well as morphological (volume) and statistical features (intensity mean and standard deviation, maximum and minimum intensities, entropy) of the tumor regions volumes in all modalities (T1, post-contrast T1, T2 and FLAIR). The ground truth mask were used to extract such features. Some dimensionality reduction algorithms were also used to increase the features of the dataset. The algorithms included PCA, ICA, Truncated SVD, Gaussian Random Projections and Sparse Random Projections. Each one of these returned 15 components, summing up to 75 more features, in a total of 120.

The training was made using 5 fold cross-validation, repeated 10 times, in order to avoid bias. The results of the validation set in each case were used to calculate the scores, that included accuracy and macro average F1 score on the classification problem. For regression the coefficient of determination (R2) and the root mean squared error (RMSE) were calculated.

3. Results

Results for the segmentation task on the test split of the training set as well as on the validation leaderboard are displayed on Table 1. Predictions were made passing patches to the net with stride 16. The regions where more than one patch coincided had their predictions averaged to get the final prediction mask and evaluate the Dice metric.

Results for the overall survival classification task are displayed on Table 2, and the scores for the overall survival regression task can be found on Table 3.

In the overall survival classification task, class 0 corresponds to overall survival up to 10 months; class 1 encompasses overall survival from 10 to 15 months; and class 2 implies overall survival above 15 months.

Table 1. Mean scores for each tumor segmentation target.

Segmentation target	Validation on 30% of the training set - test split (63 cases)		Validation leaderboard scores (46 cases)	
	Dice coefficient	Hausdorff distance	Dice coefficient	Hausdorff distance
Whole tumor	0.9111	19.8746	0.8845	13.6590
Tumor core	0.9118	11.7898	0.7674	22.3667
Enhancing tumor	0.8272	11.000	0.7261	13.8817

Table 2. Scores for overall survival duration group classification on cross-validation

Accuracy (%)	F1 Score (%)
48.5	47.1

Table 3. Scores for overall survival estimation in days on cross-validation.

R2 Score	RMSE
-196	234.818

4. Discussion

The idea of implementing neural networks using the different types of images together (T1, post-contrast T1, T2 and FLAIR) resulted in a promising solution for the task of segmentation of the brain tumor. We implemented different numbers of channels for each type of tumor (whole tumor, tumor core and enhanced tumor) based on the variations of the signal intensity of voxels. All whole tumor segmentation was designed with two networks using the four channels and two distinct methods of preprocessing. One of the methods used a simple standard scaler with normalization and the other one utilised a histogram equalization only in the brain section. The first network demonstrated less restriction to give high probabilities of the whole tumor in regions with high-intensity voxels. However, the network that used the four channels and the histogram equalization acquired a high degree of conservatism, because of these two results we used the mean probability between the two networks and acquired the values as illustrated in Table 1.

The tumor core network used only the post-contrast T1 channel with the standard scaling and normalization preprocessing. We believe that this channel has enough

contrast for the network to distinguish between the brain tumor edema and the brain tumor core. However, this contrast is more evident in HGG cases when compared to LGG cases. In the latter, our network resigned to guess the series tumor core. This problem is intended to be solved using a secondary network trained focused only in the LGG cases. The enhanced tumor segmentation used a network with T1 and post-contrast T1 channels, the idea was to use the contrast between this two channels to better select the voxels of enhanced tumor. Although the segmentation has a considerable dice in the training dataset (approximately 0.82), we confronted the same differences of voxels intensity between the HGG and LGG data which made difficulty the enhanced tumor segmentation in the LGG cases.

The proposed method in this short paper showed a good solution for the segmentation of the three types of tumor. However, the effectiveness is more evident in the HGG types when compared with the LGG.

The features extracted to perform overall-survival were not enough to get a good regression score. In future work will be analysed what features correlate to a shorter or longer overall-survival, and extract that features to do regression.

5. Conclusion

The decision of using deep learning, specifically the U-Net, was due the great results it had in other tasks of segmentation in biomedical images [5]. We observed that even in medical images that architecture has a promising performance and we focused more in the images pre and post-processing.

At the beginning of the project we've tried different approaches regarding the input for the segmentation tasks. We observed that the 64 X 64 X 64 3D patches had the best results.

Results shown in this paper are promising and we look forward to improving our scores on the segmentation task by training all networks with LGG cases (and not only HGG cases) in order to increase generalization. We are working to improve our performance on the overall survival tasks as well.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)“, IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)

2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”, Nature Scientific Data, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection”, The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection”, The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Ronneberger, O.; Fischer, P.; Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241. Springer, Cham. (2015)
6. Chen, T.; Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794. ACM (2016).

Multi-dimensional Gated Recurrent Units for Brain Tumor Segmentation

Simon Andermatt, Simon Pezold, and Philippe Cattin

Department of Biomedical Engineering, University of Basel, Switzerland

Abstract. Glioma segmentation is a difficult task due to strongly varying intensity and shape of the pathology. We propose to tackle it with multi-dimensional gated recurrent units, a recently developed method which incorporates recurrent neural networks in image segmentation. By competing in the BraTS 2017 segmentation challenge, we hope to be able to determine the strengths and weaknesses inherent to multi-dimensional gated recurrent units in the case of pathology segmentation.

1 Introduction

Ranking currently on third place in MrBrainS 2013 [6], a brain segmentation challenge which was held in conjunction with MICCAI 2013, Multi-dimensional Gated Recurrent Units [1] (MD-GRU) have shown promising results on medical images. However, we believe modelling anatomy to be a far easier problem than modelling pathological structures. Shapes, extent and intensity of pathologies can vary immensely between patients. We hence want to explore the capabilities of MD-GRU in pathology segmentation by evaluating it on BraTS 2017 [7, 2–4] images, using as little pre- and postprocessing on the available data as possible. In the following, we describe our method in detail.

2 Methods

2.1 Network

We model our network close to its original publication in [1], which contains a comprehensive overview of the method. In brief, the MD-GRU layer is the multi-dimensional analogy to one-dimensional bi-directional gated recurrent units [5], where the signal is processed in both directions and the sum of both states is returned as output. For each data dimension, i.e. three dimensions in the case of BraTS 2017, the data are processed along the selected dimension in the forward and backward direction. To make sure that the whole available spatial context is taken into consideration, the standard GRU formulas are adapted to include the local neighborhood when taking the weighed sum of input and state. This can be elegantly formulated by convolutions, which replace the previous matrix multiplications. We use the same network setup as described in [1], with three MD-GRU layers, followed by a voxelwise fully connected layer. The first two

of these layers are activated with a hyberbolic tangent, while the last one is directly fed into the softmax function, returning the probabilities for each of the five classes – we ignored the fact that class 3 is not represented in the data.

2.2 Preprocessing

Each scan is provided to the network as the original image as well as a high-pass filtered version. The high-pass filtered version is created by subtracting a 3d Gauss-filtered version from the data, where we use a Gauss kernel with $\sigma = 5$ voxels. The intensities of both the original scans and their high-pass filtered counterparts are then normalized to zero mean and a standard deviation of one. No further preprocessing is applied.

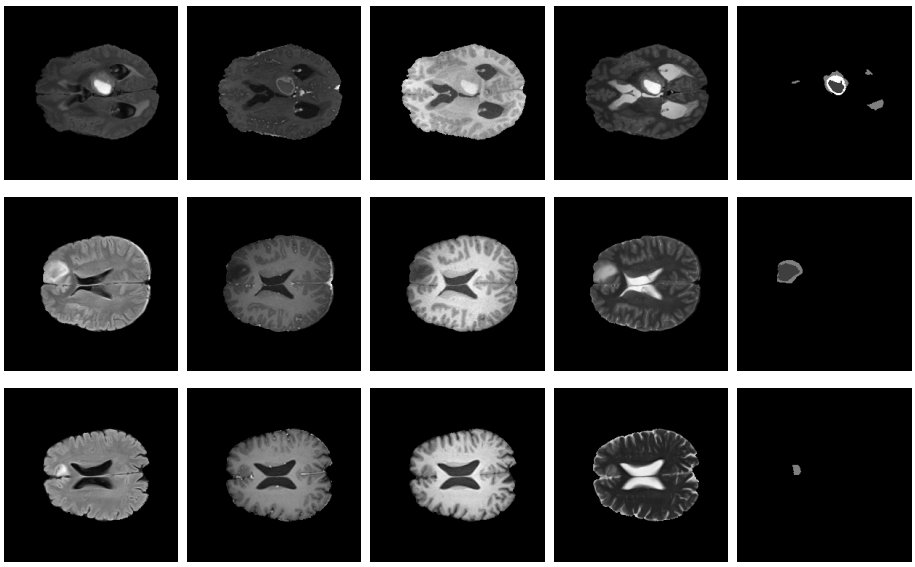


Fig. 1. Sample results. *Rows (top to bottom):* Slice 70 from local validation data Brats17_CBICA_ATX_1 and slices 87 and 88 from public validation data Brats17_TCIA_612_1 and Brats17_TCIA_613_1, respectively. *Columns (left to right):* FLAIR, T1CE, T1 and T2 sequences and segmentation results.

2.3 Optimization

For each iteration, a patch of $80 \times 80 \times 80$ voxels is extracted from a randomly selected scan at a random location. The patch is deformed using a randomly generated deformation field. On a low-resolution grid with a spacing of 75 voxels along each dimension, centered at the middle of the patch, random deformation vectors are drawn from $\mathcal{N}(0, 7)$. This grid is then interpolated to a full-resolution

grid using cubic interpolation, producing a smooth deformation field. We train the network for 40 000 iterations using Adadelta. For 50 000 more iterations, we train the network with additional data augmentation consisting of random rotations of drawn from $[-10°, 10°]$ and a scaling factor along each dimension drawn from $[0.8, 1.2]$. We apply DropConnect [8] on all state and input weights in the form of multiplicative Gaussian noise from $\mathcal{N}(1, 0.5)$. The network has been implemented, trained and evaluated in TensorFlow 1.1. We randomly chose three volumes to be used as a separate validation set (called *local* in the following) and excluded them from training.

2.4 Evaluation

For each subvolume, we evenly divide its voxel grid into blocks of $100 \times 100 \times 100$, maintaining a constant overlap of 20 voxels along all dimensions. This produces $3 \cdot 3 \cdot 2$ subvolumes, which are evaluated by the network. The individual subvolumes are then stitched using linear blending at the overlaps, to make up for bad predictions due to the missing information at the borders of the subvolume. The final class for each voxel is then determined by the largest probability. No postprocessing is applied on the network's classification results.

3 Results

Although the testing data has not yet been made available, the validation leaderboard still gives interesting information about the performance of the different teams' algorithms. In Table 1, we show the average performance for each label and score for the three validation patients. Table 2 lists the final validation scores. All in all, we were able to achieve results that are comparable to the top performing methods on the validation leaderboard. At the time of writing, using the average of the calculated validation leaderboard ranks in Table 2 to create an overall ranking, we managed to achieve rank 7 out of 45.

Table 1. Local validation set. Average Dice, sensitivity, specificity and 95th percentile Hausdorff for enhancing tumor (ET), whole tumor (WT) and tumor core (TC).

	Dice	Sensitivity	Specificity	Hausdorff95
ET	0.81589	0.81338	0.99948	3.14208
WT	0.90617	0.94959	0.99527	8.71794
TC	0.93871	0.93148	0.99924	1.66667

Table 2. Public validation set. *Rows:* Mean, standard deviation, median and 25th and 75th percentile are given for the evaluated metrics. The Leaderboard rank at the bottom is calculated from all complete entries in the validation leaderboard, where the values of the snapshot at Mon Jul 24 00:00:04 EDT 2017 were used for their calculation. *Columns:* Dice, sensitivity, specificity and 95th percentile Hausdorff for enhancing tumor (ET), whole tumor (WT) and tumor core (TC).

	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.7112	0.8932	0.7349	0.7355	0.8925	0.6843	0.9984	0.9948	0.9988	4.1870	4.6126	8.1886
StdDev	0.3044	0.0862	0.2991	0.2882	0.1279	0.3116	0.0036	0.0040	0.0022	6.1122	5.7321	13.8129
Median	0.8481	0.9130	0.8666	0.8349	0.9389	0.8334	0.9992	0.9954	0.9994	2.0000	3.0811	4.4721
25quantile	0.7465	0.8909	0.7085	0.7643	0.8870	0.5581	0.9985	0.9918	0.9988	1.4142	2.2361	2.0443
75quantile	0.8865	0.9452	0.9326	0.9120	0.9563	0.9122	0.9997	0.9982	0.9998	3.6056	4.2128	9.8615
LB rank	10	7	16	21	16	32	13	9	5	2	4	4

4 Discussion

Inspecting our local validation dataset and the results from the public validation set, we were able to identify two main issues with our method in the context of brain tumor segmentation. First, other types of pathology that are not labelled in this competition, as for instance the lesions shown in the first row of Figure 1, can lead to varying amounts of misclassification. Second, pathology classes that are only represented in a small amount of data or not at all, especially in the case of the *enhancing tumor* class, can lead to a large impact on any of the outcome measures. If the intensity distribution of a scan additionally differs from the distributions seen during training, classes with low voxel count can be missed entirely or classes missing in a patient might be segmented at the wrong location as outliers. Examples are rows two and three of Figure 1. In row two, the small enhancing tumor is not recognized for instance, leading to a Dice of 0. On the other hand, if a class is not present at all, it is very easy to achieve a Dice score of 1 (row three).

The difference between mean and median in Table 2 indicates, that across all classes and measures few outliers caused a great impact on the scores. Except for the normalization to zero mean and one standard deviation to both high-pass filtered and original data, we did not apply preprocessing to the data. We could implement specific counter measures, which would alleviate the influence of these outliers. For instance, we could implement histogram matching during training and testing, to ensure similiar intensity distributions, which should help with the misclassifications of the either small or non-existent classes. Furthermore, providing additional high-pass filtered versions of the input data could help reducing the influence of noise. However, we wanted to keep our method as close to the original implementation as possible and hence only applied the already published preprocessing. Using the original implementation allows for a direct comparison to other problem settings.

References

1. Andermatt, S., Pezold, S., Cattin, P.: Multi-dimensional Gated Recurrent Units for the Segmentation of Biomedical 3D-Data. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. pp. 142–151. Springer (2016)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features". Nature Scientific Data (2017), [in press]
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017), DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017), DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv:1406.1078 [cs, stat] (Jun 2014)
6. Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., de Bresser, J., Alansary, A., de Bruijne, M., Carass, A., El-Baz, A., Jog, A., Katyal, R., Khan, A.R., van der Lijn, F., Mahmood, Q., Mukherjee, R., van Opbroek, A., Paneri, S., Pereira, S., Persson, M., Rajchl, M., Sarikaya, D., Smedby, Ö., Silva, C.A., Vrooman, H.A., Vyas, S., Wang, C., Zhao, L., Biessels, G.J., Viergever, M.A.: MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. Computational Intelligence and Neuroscience 2015, e813696 (Dec 2015)
7. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Ç. Demiralp, Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993–2024 (Oct 2015), DOI: 0.1109/TMI.2014.2377694
8. Wan, L., Zeiler, M., Zhang, S., Cun, Y.L., Fergus, R.: Regularization of Neural Networks Using Dropconnect. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13). pp. 1058–1066 (2013)

Sequential 3D U-Nets for Brain Tumor Segmentation

Andrew Beers[1], Ken Chang[1], James Brown[1], Emmett Sartor[2], CP Mammen[3], Elizabeth Gerstner[4], Bruce Rosen[1], Jayashree Kalpathy-Cramer[1]

[1] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[2] Massachusetts General Hospital, Boston, MA, USA
[3] NVIDIA, India
[4] Athinoula A. Martinos Center for Biomedical Imaging, Department of Neuro-Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

abeers@mgh.harvard.edu
kenchang@mit.edu
qtim.lab@gmail.com

Abstract. The 2017 BraTS Competition challenges participants to develop a fully automatic or semi-automatic multi-modality tumor segmentation tool for enhancing, non-enhancing, and edema in glioblastoma patients.Our entry to the competition is a fully automatic pipeline that involves chaining together several unique 3D U-Net, a type of 3D patch-based convolutional neural network. Our pipeline takes advantage of the prior knowledge that enhancing and non-enhancing tumor are likely to be found within regions of edema and within proximity to each other by feeding the prediction outputs of earlier networks into later networks. We achieve greater context for our patch-based sampling method by predicting downsampled labels and then upsampling them using a separate 3D U-Net. We use a fine-tuning network and a candidate evaluation network to account for tissue border discrepancies and catastrophic segmentation failure. Preliminary results for an unoptimized version of this pipeline on validation data with unknown ground truth segmentations had mean dice coefficients of 0.78, 0.67, and 0.68 for whole tumor, enhancing, and non-enhancing tissue respectively.

Keywords: 3D CNN · U-Net · Sequential · Deep Learning · Upsampling · Patches

1 Introduction

Compared to traditional segmentation methods, deep learning does not rely on the generation of handcrafted features to distinguish tumor from normal brain anatomy. Instead, raw image intensities are taken as input subjected to many layers of convolutions, to calculate an output signal. The many degrees of freedom and inclusion of non-linearities allow the algorithm to learn complex patterns with a high level of abstraction[1]. Up until this point many of the deep learning algorithms that have been applied to brain tumor segmentation have been 2D Convolutional Neural Networks (CNNs), which do not take advantage of the full breadth of volumetric information. Recently, there has been an increase in popularity of 3D CNNs[2,3], which have been shown to be effective for this task, albeit at the expense of additional computational complexity. For example, the 3D U-Net architecture was successfully applied for the segmentation of the Xenopus kidney, a complex and highly variable structure[4].

The 2017 BraTS Challenge challenges participants to develop a fully automatic or semi-automatic multi-modality tumor segmentation tool for enhancing, non-enhancing, and edema in glioblastoma patients[5,6,7,8]. The 2017 BraTs Challenge patient cohort includes glioblastoma and low-grade glioma pre-operative patients. Participants are provided with coregistered and skull-stripped T2, pre-contrast T1, post-contrast-T1, and FLAIR images, and are then asked to generate segmentations that can then be compared against ground-truth segmentations of edema, enhancing tumor, and non-enhancing tumor. Ground-truth segmentations are manually drawn by one to four raters and then approved by expert neuro-radiologists. Our proposed segmentation method for BraTS 2017 involves the training of several 3D U-Nets concatenated end-to-end combined with a regime of hand-crafted pre- and post-processing operations on input data. The result is a fully automatic segmentation pipeline requiring no additional data aside from four coregistered input modalities and a practical computation time for batch processing.

2. Methods

Our segmentation method is a pipeline of different neural networks, pre-processing, and post-processing steps. We start by we normalizing whole brain intensities to zero mean and unit variance within each patient and within each MR modality. For the purposes of whole tumor segmentation, we also down-sampled a copy of each modality to isotropic 2mm resolution. All of our 3D CNNs took as input a patch of size 16x16x16 voxels, regardless of image resolution. For the whole tumor CNN, patches were sampled in the following ratio: 1% background, 29% normal brain, 70% tumor. For the enhancing tumor, non-enhancing, and fine-tuning CNNs, patches were sampled from only within the tumor region.

We use a slightly modified version of the U-Net laid out in Çiçek et al[4]. Our U-Net has a unique and separate downsampling branch for each input modality. Downsampling branches are concatenated for the purposes of the residual connections, such that there is only one multi-channel upsampling branch. The intention of this structure was to preserve anatomical information from each input modality, rather than let any one modality predominate. Each downsampling branch has proportionately fewer filters than the original U-Net, such that when concatenated they equal the same amount of filters for the original upsampling branch.

The start of our pipeline feeds pre-processed isotropic 2mm patches into a U-Net trained to predict the binary whole tumor labelmap provided by the BraTS organizers. Pseudo-probability maps outputted by this network are thresholded at an arbitrary value, and then fed into an upsampling U-Net. The upsampling U-Net is trained on labelmaps downsampled to isotropic 2mm resolution, and then upsampled back up to 1mm resolution using nearest-neighbor interpolation. These modified labelmaps serve as input data, and the given task is to predict the original, non-resampled labels as ground-truth. Using a trained version of this network, outputs of the whole tumor U-Net are upsampled into isotropic 1mm space.

The next four U-Nets in the pipeline are meant to create initial estimates for non-enhancing and enhancing segmentations by iteratively learning from the outputs of each of the previous networks. The first two U-Nets are trained on the original four modalities and the output of upsam-

pled whole tumor U-Net, with the goal of predicting enhancing and non-enhancing labels respectively. Probability maps output from these labelmaps and the output of the whole Tumor U-Net are then fed into the third and fourth U-Net. These networks are trained again to again predict the provided binary non-enhancing and enhancing labelmaps, but this time with the added advantage of being given an initial estimate of all of the other segmentations and probability maps produced by the preceding U-Nets.

The final U-Net takes in all three output segmentations (whole tumor, non-enhancing, and enhancing), and feeds them into what we refer to as a "fine-tuning" network. The fine-tuning network attempts to smooth incorrect borders from the previous U-Nets, and reconcile differences between the three previous three output segmentations.

Finally, we applied binary dilation followed by erosion to fill small gaps, and removed islands of whole tumor smaller than 1000 voxels. Connected components (islands) that remain are then identified. Representative axial slices from each island are then fed into a 2D ResNet, with the goal of identifying whether the given candidate is a true segmentation or an erroneous segmentation. ResNet is a popular 2D CNN architecture that has found success in similar medical image segmentation tasks[9]. Islands with a high pseudo-probability of being erroneous are removed. The remaining islands are submitted as entry to the competition for consideration as a fully-automatic segmentation method.

For consideration as a semi-automatic segmentation method, the output segmentations of these pipelines will be corrected for obviously erroneous borders, and incorrect islands that had escaped the proposed post-processing pipeline. A neuro-oncologist will perform the segmentation correction.

All U-Nets were trained on NVIDIA GPUs (P100s and K80s) for at least 30 epochs, or whenever validation accuracy stopped increasing within a certain predefined tolerance. Networks were designed in Keras with a TensorFlow backend.

3 Preliminary Results

We have submitted preliminary results to the competition leaderboard for Training and Validation phases. The segmentation results are based on neural networks that have yet to be optimized with respect their input parameters, and do not include the full post-processing pipeline described above. They are also fully automatic segmentations, with no manual corrections from neuro-oncologists or other readers.

For the training phase, our whole tumor, enhancing, and nonenhancing segmentation outputs achieved a mean dice coefficient across all cases of 0.79, 0.68, and 0.76, respectively, with a standard deviation of 0.16, 0.29, and 0.18. For the validation phase, our whole tumor, enhancing, and nonenhancing segmentation outputs achieved a mean dice coefficient across all cases of 0.78, 0.67, and 0.68 respectively, with a standard deviation 0.17, 0.32, and 0.28. We submitted segmentations for all available cases in both training and validation.

4 Conclusion

With the advent of ever-more-powerful GPUs, a segmentation pipeline composed of multiple, unique neural networks has become feasible. Each model in our pipeline adds value to the resulting segmentation, resulting in competitively accurate segmentations that, with little manual editing, can become of immediate use to the neuro-oncologist or radiologist.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444. doi:10.1038/nature14539.
2. Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, Ben Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Medical Image Analysis, Volume 36, 2017, Pages 61-78, ISSN 1361-8415, http://dx.doi.org/10.1016/j.media.2016.10.004.
3. Casamitjana A., Puch S., Aduriz A., Vilaplana V. (2016) 3D Convolutional Neural Networks for Brain Tumor Segmentation: A Comparison of Multi-resolution Architectures. In: Crimi A., Menze B., Maier O., Reyes M., Winzeck S., Handels H. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2016. Lecture Notes in Computer Science, vol 10154. Springer, Cham
4. Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox T., Ronneberger O. (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin S., Joskowicz L., Sabuncu M., Unal G., Wells W. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer, Cham
5. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
6. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017.
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017.
9. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conf Comput Vis Pattern Recognit [Internet]. IEEE; 2016 [cited 2017 Apr 12]. page 770–8.

Tumor segmentation from multi-parametric MRI using random forest with superpixel and tensor based feature extraction

H. N. Bharath[1,2], S. Colleman[3], D. M. Sima[1,2], S. Van Huffel[1,2]

[1] Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven, Belgium.
[2] Imec, Leuven, Belgium.
[3] Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium.

Abstract. Identification and localization of brain tumor tissues plays an important role in diagnosis and treatment planning of gliomas. A fully automated superpixel wise tumor tissue segmentation algorithm using random forest is proposed in this paper. Features for random forest classifier are extracted by constructing a tensor from multi-parametric MRI data and applying multi-linear singular value decomposition. The method is trained and tested on high grade glioma (HGG) patients from BRATS 2017 training database. It achieves a performance of 83%, 76% and 78% Dice scores for whole tumor, enhancing tumor and tumor core, respectively.

Keywords: Superpixel, Multilinear singular value decomposition, Random forest.

1 INTRODUCTION

Accurate characterisation and localization of tissue types play a key role in brain tumor diagnosis and treatment planning. Neuro-imaging methods in particular magnetic resonance imaging (MRI) provide anatomical and pathophysiological information about brain tumors and aid in diagnosis, treatment planning and follow-up of patients. Manual segmentation of tumor tissue is a tedious and time consuming job, it also suffers from inter and intra-rater variability. An automated brain tumor segmentation algorithm will help to overcome those problems. However, automation of brain tumor tissue segmentation is a difficult problem and often fails when applied on MRI images from different centres/scanners.

Performing superpixel-level image segmentation offers certain advantages over pixel-level segmentation like spatial smoothness, capturing image redundancy and reducing computational complexity [1, 2]. It has also been used in the context of brain tumor segmentation [2]. Recently, tensor decomposition has been used to extract features from high-dimensional data to use in classification algorithms [3]. Multi-parametric MRI consisting of T2, T1, T1+contrast and FLAIR imaging after co-registration and resampling to the same resolution, can be naturally represented as 3-D tensor. In this paper we develop a fully automatic tumor tissue segmentation algorithm using random forest

classifier, where both superpixel-level image segmentation and tensor decompositions methods are combined to extract features for the classifier.

2 METHOD

2.1 Preprocessing

First, each individual 3D image is scaled to the range [0-1]. Next, intensities are normalized by applying histogram equalization. Reference histogram is generated by selecting 10 random images from the training set and extracting a histogram from the combined image. Histogram equalization is applied separately to different modalities.

2.2 Feature extraction

The MR images are divided into smaller patches which are better aligned with intensity edges called superpixels [4]. The tissue assignment is done on superpixel-level instead of individual pixel, which helps to reduce computational cost and improve spatial smoothness [2]. For each of the superpixels a 4D tensor is constructed, where the first two modes are 5×5 image patches with main voxel at the centre, third mode is the modality and the difference image (e.g.: abs(T1-T2)) of the modalities and the fourth mode is the voxels within the superpixel. Features are extracted by applying multilinear singular value decomposition (MLSVD) [5] on the 4D Tensor. The fourth mode factor matrix is not used in the feature set. Another 3D tensor is constructed for each superpixel, where the frontal slices are the covariance matrix of pixel-level features. Pixel-level features consist of mean, median, standard deviation and entropy over a 5×5 window. Again, features are extracted by applying rank-2 MLSVD. Additional features like mean, entropy and standard deviation within each superpixel are also included in the feature set. These features are calculated for all four modalities and six difference image between the modalities.

2.3 Training and Tissue Segmentation

Tumor tissue segmentation was performed using a two-stage classifier. In the first stage a binary classification was performed on the superpixels to segment tumor and non-tumor regions. In the second-stage a multi-class classification was performed on the superpixels which are inside the estimated tumor region to segment active tumor, edema, necrosis and healthy tissue. The two stage operation is demonstrated in Fig. 1. For both stages a random forest classifier with 100 trees was used. Random forest classifiers are trained using a iterative method, Initially, 40 patients are randomly selected from the high grade glioma (HGG) dataset for training and the trained model is tested on the remaining dataset. Next, the patients which resulted in low Dice scores are included in training set and a new model is trained. After the first stage image filling operation is performed on the estimated whole tumor segmentation before going to the second stage. In the first stage, superpixels are obtained from FLAIR imaging modality because the entire tumor is brighter in this modality. Whereas, in the second stage, T1+contrast imaging modality is used to obtain superpixels.

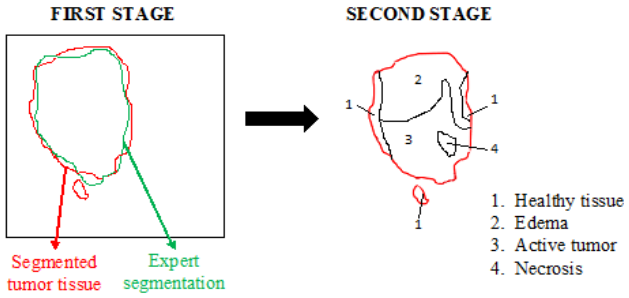


Fig. 1: Demonstration of whole tumor segmentation in first stage and sub-tissue segmentation in second stage.

3 Results

The BRATS 2017 high grade glioma database [6–9] containing 210 patients is split into training set (70%) and test set (30%). The trained model is tested on 63 HGG patients, boxplot of Dice score is shown in Fig. 2. Average Dice score and sensitivity obtained from the trained model over 63 HGG patients are shown in Table 1.

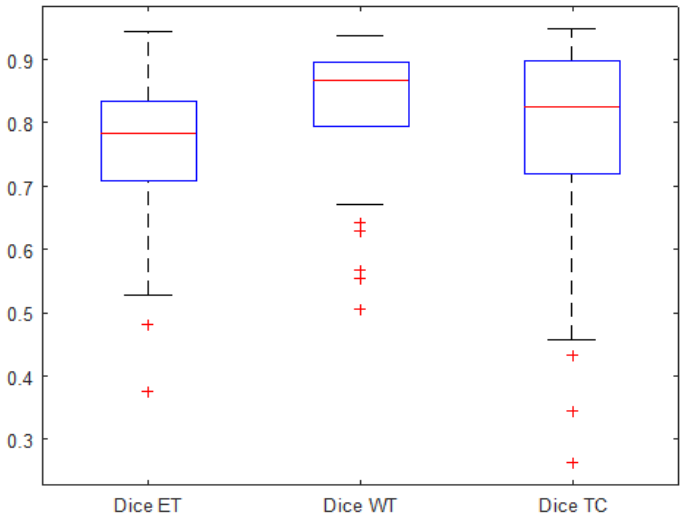


Fig. 2: Boxplots of Dice score for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) on 63 BRATS 2017 training dataset

The trained model is also tested on BRATS 2017 validation dataset [6–9], the results are shown in Fig. 3 and Table 2. The performance is worse when compared to HGG case

Table 1: Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and sensitivity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over 63 HGG patients.

	Dice ET	Dice WT	Dice TC	Sensitivity ET	Sensitivity WT	Sensitivity TC
Mean	0.761	0.833	0.783	0.855	0.815	0.777
Std	0.106	0.096	0.147	0.126	0.090	0.191
Median	0.783	0.867	0.824	0.886	0.837	0.826
25 quantile	0.708	0.795	0.723	0.820	0.769	0.721
75 quantile	0.833	0.895	0.898	0.941	0.884	0.908

with average Dice scores of 76%, 69% and 59% for whole tumor, enhancing tumor and tumor core, respectively.

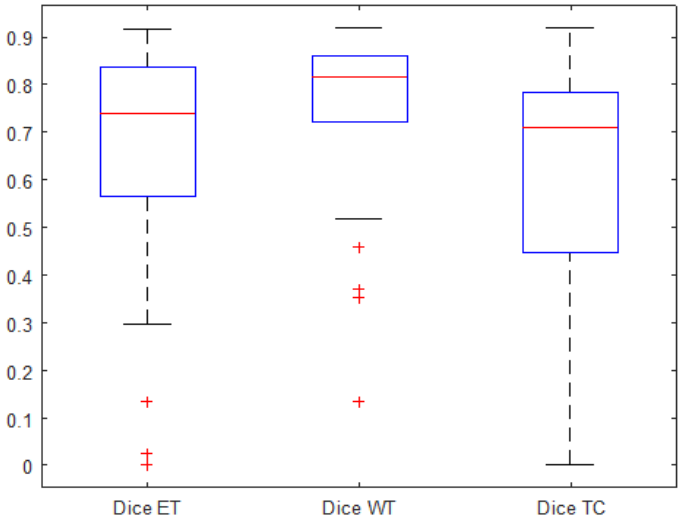


Fig. 3: Boxplots of Dice score for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) on 46 BRATS 2017 validation dataset.

4 Conclusion

In this paper, we developed a fully automated algorithm for brain tumor segmentation from multi-parametric MRI data. Superpixel and tensor based feature extraction algorithm is proposed to use with two stage random forest classifier for segmenting tumor tissue in HGG patients. The performance of algorithm is comparable to the state of art methods when applied only to HGG patients. However, its performance deteriorates when tested on BRATS 2017 validation database, which contains both low grade

Table 2: Mean, standard deviation, median 25 quantile and 75 quantile of Dice score and sensitivity for enhancing tumor (ET), whole tumor (WT) and tumor core (TC) over 46 validation dataset.

	Dice ET	Dice WT	Dice TC	Sensitivity ET	Sensitivity WT	Sensitivity TC
Mean	0.688	0.755	0.586	0.759	0.748	0.575
Std	0.259	0.166	0.285	0.242	0.150	0.299
Median	0.767	0.815	0.710	0.829	0.777	0.631
25 quantile	0.590	0.723	0.449	0.694	0.691	0.445
75 quantile	0.852	0.860	0.778	0.896	0.860	0.797

glioma (LGG) and HGG patients. To improve the performance on BRATS 2017 validation database LGG patients needs to be included in the training phase.

ACKNOWLEDGMENT

This research was supported by: EU: European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013): EU MC ITN TRANSACT 2012, #316679 and ERC Advanced Grant, #339804 BIOTENSORS. This paper reflects only the authors views and the Union is not liable for any use that may be made of the contained information.

References

1. Jia S, Zhang C. "Fast and robust image segmentation using an superpixel based fcm algorithm". Image Processing (ICIP), 2014 IEEE International Conference on. IEEE (2014).
2. Wu W, Chen AY, Zhao L, Corso JJ. "Brain tumor detection and segmentation in a CRF (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features". International journal of computer assisted radiology and surgery, 9(2), 241-253 (2014).
3. Fargeas A, Albera L, Kachenoura A, Dréan G, Ospina JD, Coloigner J, Lafond C, Delobel JB, De Crevoisier R, Acosta O. "On feature extraction and classification in prostate cancer radiotherapy using tensor decompositions". Medical engineering & physics, 37(1), 126-131 (2015).
4. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. "SLIC superpixels compared to state-of-the-art superpixel methods". IEEE transactions on pattern analysis and machine intelligence, 34(11), 2274-2282 (2012).
5. De Lathauwer L, De Moor B, Vandewalle J. "A multilinear singular value decomposition". SIAM journal on Matrix Analysis and Applications, 21(4), 1253-1278 (2000).
6. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM,

Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)”, IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).

7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. “Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features”, Nature Scientific Data, (2017) [In Press].
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection”, The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. “Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection”, The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

3D U-Net for Multimodal Brain Tumor Segmentation

Shilei Cao[1], Buyue Qian[1] Changchang Yin[1], Xiaoyu Li[1], and Shiyu Chang[2]

[1] Xi'an Jiaotong University, Xi'an, Shaanxi, China,
{shileicao, lentery, wemakefocus}@stu.xjtu.edu.cn, qianbuyue@xjtu.edu.cn
[2] IBM T. J. Watson Research, Yorktown Height, NY 10598, USA,
shiyu.chang@ibm.com

Abstract. We have witnessed the big success of 2D U-Net in biomedical image segmentation, the effect of whose 3D extension is under explored, especially for multimodal brain tumor segmentation. In this paper, we examine the effectiveness of 3D convolutional network on this task using a U-Net structure. We take data augmentation, data sampling strategy and class-specific weighted loss into consideration, which we think is important for a successful segmentation model. Experiment results show the effectiveness of our proposed methods.

Keywords: 3D U-Net, convolutional network, data augmentation, class balancing

1 Introduction

Since fully convolutional networks (FCN) on semantic segmentation [1] was proposed, FCN have attracted more and more attentions on segmentation tasks. As one of the members of this architecture, U-Net [2] has been successfully applied to many biomedical image segmentation task. Recently, its 3D extension has been steadily gaining more and more research interests on volumetric images, such as 3D U-NET [3] and [4]. Although increasing attention has been paid, only limited works focus on their effect on multimodal brain tumor segmentation task. In this paper, we will give an experimental study of the effectiveness of 3D U-Net on multimodal brain tumor segmentation. This paper is intended for Multimodal Brain Tumor Segmentation Challenge 2017 (Brats2017)[6–9] [3].

Since the cardinality of available dataset is low, we absorbed some useful augmentation methods into our proposed framework, such as random crop, rotation, flip and etc. As we all know, the sample number of each class is seriously imbalanced, which push us to absorb a weighted loss to solve this problem. Experiment results show both techniques can boost the performance of the segmentation tasks.

[3] https://www.med.upenn.edu/sbia/brats2017.html

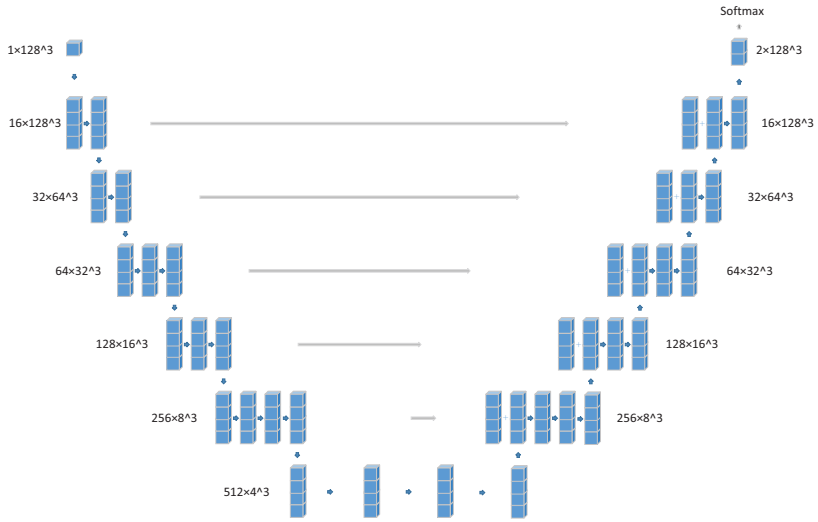


Fig. 1. The 3D u-net architecture.

2 The Method

2.1 Data augmentation

Data augmentation is important in the applications of computer vision, which is a pillar of a successful model especially in the case where only limited data samples are available. The augmentation method we use contains random crop, flip, zoom and etc.

2.2 Network Architecture

The network architecture can be seen in Fig. 1. We use patch-wise input whose central voxel point is averagely sampled across the 4 classes to eliminate the impact of the imbalance among different classes. Just like the classic U-Net [2], the network is composed of convolutions with downsampling and upsampling making the network look like the shape of 'U'. Batch normalization is absorbed to accelerate network training before nonlinear units (We use Relu[10] in this paper). The output is fed into a softmax layer and then cross entropy loss is added for training the network. To further eliminate the negative impact of the class imbalance, a weighted loss is proposed as follows:

$$\sum_{i=1}^{n} \mathbf{L} = \frac{1}{\|C_i\|} \mathbf{L}_i$$

Where n is the number of classes, $\mathbf{L}$ is the total loss, $\mathbf{L}_i$ is the class ith loss. $\|C_i\|$ is the cardinality of the class i.

3 Experiment Results

A preliminary result is shown in Table. 1.

Table 1. Evaluation results at BRATS2017 online judge system.

	ET	WT	TC
Sensitivity	0.9646	0.9995	0.9993
Specificity	0.5083	0.5014	0.5013
Dice	0.0240	0.1189	0.0575
Hausdorff95	98.97	86.89	99.54

References

1. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CoRR*, vol. abs/1411.4038, 2014. [Online]. Available: http://arxiv.org/abs/1411.4038
2. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597
3. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: Learning dense volumetric segmentation from sparse annotation," *CoRR*, vol. abs/1606.06650, 2016. [Online]. Available: http://arxiv.org/abs/1606.06650
4. F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, 2016. [Online]. Available: http://arxiv.org/abs/1606.04797
5. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167
6. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. R. Gerstner, M. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. T. Ryan, D. Sarikaya, L. H. Schwartz, H. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Székely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. B. Ünal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. V. Leemput, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015. [Online]. Available: https://doi.org/10.1109/TMI.2014.2377694
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI

collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
10. R. K. Srivastava, J. Masci, F. Gomez, and J. Schmidhuber, "Understanding locally competitive networks," *arXiv preprint arXiv:1410.1165*, 2014.

Volumetric Multimodality Neural Network For Brain Tumor Segmentation

Laura Silvana Castillo ⋆, Laura Alexandra Daza⋆, Luis Carlos Rivera⋆ and Pablo Arbeláez

Universidad de los Andes, Bogotá Colombia

Abstract. Brain lesion segmentation is one of the hardest tasks to be solved in computer vision with an emphasis on the medical field. We present a convolutional neural network that produces a semantic segmentation of brain tumors, capable of processing volumetric data along with information from multiple MRI modalities at the same time. This results in the ability to learn from small training datasets and highly imbalanced data. Our method is based on DeepMedic, a method in the state of the art in brain lesion segmentation. We develop a new architecture with more convolutional layers, organized in three parallel pathways with different input resolution, and additional fully connected layers. We tested our method over the 2017 BraTS Challenge dataset, reaching an average dice coefficient of 87% over training dataset and 86% over validation dataset for the whole tumor segmentation task.

Keywords: Semantic segmentation, Brain tumor, Deep learning, MRI

1 Introduction

Brain tumors are abnormal formations of mass that have different shapes, sizes and internal structures. As this formations grow, they apply pressure to the surrounding tissues, causing varied problems such as unexplained nausea or vomiting, seizures or personality changes [1]. The use of Magnetic Resonance Imaging (MRI) to visualize the brain of a patient allows doctors to look for any life-threatening abnormality. However, finding those structures in a medical image is a complicated task, highly prone to error [2]. In spite of the fact that the treatment selection is based directly on the diagnosis, nowadays that process is made manually, which causes it to be observer-dependent and increases the uncertainty of the patient's outcome.

Automatic brain lesion segmentation has been a topic of interest for more than a decade. Early approaches to this problem were based on the detection of abnormalities using healthy-brain atlases and probability models [3]. Then, results were upgraded with the use of deformable registration fields coupled with Markov Random Fields (MRF) [4]. Subsequent approaches using machine

⋆ Authors with equal contributions

learning techniques, such as Random Forest [5, 6], improved the results even more, reaching an average dice coefficient of 73% in the 2012 BraTS Challenge.

In recent years, Convolutional Neural Networks (CNN) have shown outstanding results in detection, classification and segmentation tasks, being able to match and sometimes outperform humans. Some of this success is due to the rapid improvement of machines computational power and to CNNs ability of abstracting features in different representations of an image [7]. Fully convolutional networks (FCN) have proven to be an effective way to do pixel-by-pixel classification, obtaining a mean IoU of 67% in the PASCAL-VOC dataset in 2012. This method offers the advantage of combining coarse and shallow semantic information from images with an arbitrary input size [8]. In 2015 a method called U-Net, based on FCN and specialized in the task of segmenting medical images, was developed. U-Net's architecture has a contracting path to extract context and an expanding path to locate the object within the whole image [9]. Recently, an expansion of this method to process tridimensional images was presented. V-Net demonstrated a remarkable behavior in the MICCAI 2012 PROMISE Challenge dataset for prostate segmentation in MRIs, obtaining an average dice of 82% [10]. Another important method to segment volumetric medical images is DeepMedic, a neural network implemented in *Theano* that takes as inputs multimodal 3D patches extracted from MRIs. It analyzes the information using two pathways, one for every resolution used, and combines their results with fully connected layers to generate a semantic segmentation by category [11].

In this project, we aim at providing an efficient, accurate way of automatically estimating the volume of a brain tumor. To do that, we use the 2017 BraTS Challenge dataset, which has 210 High Grade Glioma (HGG) cases and 75 Low Grade Glioma (LGG) cases [2, 12, 13, 14]. Each image has different modalities and annotations made by several specialists. We developed a neural network, based on V-Net and DeepMedic, with three contracting pathways and residual connections that receive patches centered on the same voxel, but with different spatial resolution. During the testing stage, the average dice coefficient was calculated to measure the performance of the methods. BraTS challenge have different segmentation evaluation tasks, the average dice coefficient, sensitivity, specificity and Hausdorff distance.

2 Methodology

2.1 Multimodality Volumetric Neural Network

As mentioned before, we used as starting points the methods V-Net and DeepMedic, given the good previous performance that these had in tasks of segmentation in MRIs. Figure 1 shows an overview of our approach, our neural network has three identical parallel pathways, each one with six convolutional layers and two residual connections, to extract features on specific resolution levels. All the paths receive patches centered at the same voxel, but extracted from different versions of the image (original and downsampled by factors of three and five). The patches have input sizes of 27^3, 17^3 and 15^3 for the normal, medium and low

resolution pathways, respectively. After those last two pathways, an upsample layer is used to make the three outputs of the same size. Finally, the results are concatenated and introduced in the fully connected layers to be combined and then classified. The classification layer is a convolution with kernel size of 1^3 and the final output is made using a softmax classifier.

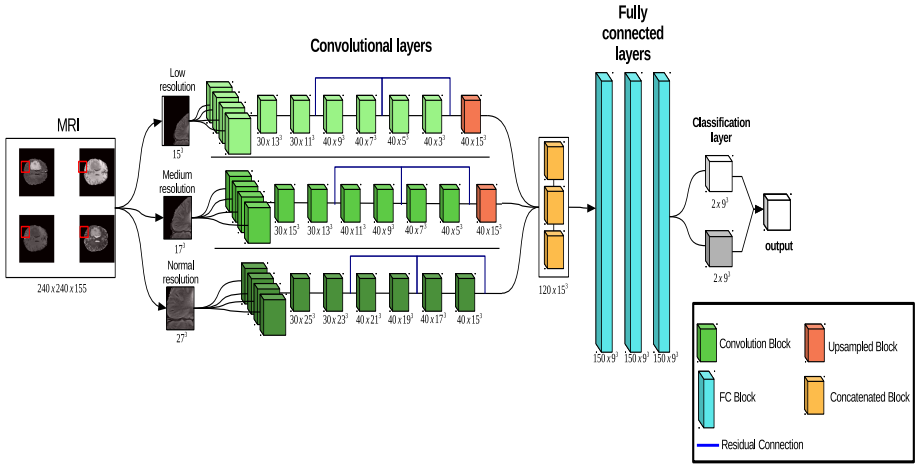


Fig. 1. Proposed Architecture. The kernels of the convolutions in the three pathways are 3^3 and no padding was made in those operations. The input of the 3 paths are centered in the same voxel, but the medium resolution and low patches are obtained from downsampled versions of the image by factors of 3 and 5, respectively

2.2 Data

The BraTS challenge (2017) training dataset includes 210 different MRI files from high grade glioma (HGG) cases and 75 MRIs from low grade glioma (LGG), and the BraTS challenge (2017) validation dataset includes 46 different MRI files. The images have four modalities (T1, T1 contrast-enhanced, T2 and FLAIR). The groundtruth annotations in this dataset contain five different categories, representing the internal structures of the tumor [2, 12, 13, 14]:

0. Everything Else.
1. Necrosis and Non-Enhancing tumor.
2. Edema.
4. Enhancing tumor.

Training The architecture was trained using the 285 cases from the training dataset (HGG and LGG). Our method input are patches of size 27^3 that are

extracted randomly, making sure that 50% of them are centered at a voxel labeled as tumor. The data is normalized individually by setting the mean to 0 and the variance to 1. Data augmentation is made to avoid overfitting of the model due to the small size of the training dataset, and it is performed on the fly to prevent memory issues. The process is made by reflecting randomly chosen volumes along the sagittal axis.

Validation To test the model, the 46 volumes from the validation dataset were evaluated with the network. In the testing stage, the patches are extracted at uniform intervals in the validation volumes.

2.3 Evaluation Metrics

Dice Coefficient The Dice-Coefficient (Equation 1) is calculated as performance metric. This measure states the similarity between clinical Ground Truth annotations and the output segmentation of the model. Afterwards, we calculate the average of those results to obtain the overall dice coefficient of the models.

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

Hausdorff Distance The Hausdorff Distance (Equation 2) is mathematically defined as the maximum distance of a set to the nearest point in the other set [15], in other words how close are the segmentation and the expected output.

$$H(A, B) = max\{min\{d(A, B)\}\} \tag{2}$$

Sensitivity and Specificity Are statistical measures used to evaluate the behavior of the predictions and the proportions of True Positives (TP), False Negatives (FN), False Positives (FP) and True Negatives (TN). The Sensitivity (Equation 3), also known as True Positive Rate, gives the proportion of true positives predicted correctly. The specificity (Equation 4), also known as True Negative Rate measures, how well the true negatives are predicted.

$$Sensitivity = TPR = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = TNR = \frac{TN}{TN + FP} \tag{4}$$

3 Experimental Results

As mentioned before, in our method (figure 1) we propose a three-pathway architecture, with different resolutions, in order to get information of the location of the tumor and, at the same time, acquire local data that helps to differentiate

the parts of the lesions, avoiding false positives. It was trained in the BraTS 2017 training dataset with a learning rate of $5 * 10^{-4}$ for 35 epochs. In Table 1 we present the results on the training and validation datasets:

Table 1. Dice coefficient, sensitivity, specificity and Hausdorff distance of our neural network for Enhanced Tumor, Whole Tumor and Core Tumor; evaluated over the training and validation datasets from BraTS 2017

	Dice			**Sensitivity**			**Specificity**			**Hausdorff95**		
	Enh.	**Whole**	**Core**	**Enh.**	**Whole**	**Core**	**Enh.**	**Whole**	**Core**	**Enh.**	**Whole**	**Core**
Train	0.74	0.87	0.78	0.78	0.87	0.74	0.99	0.99	0.99	7.52	21.32	9.18
Val	0.69	0.86	0.69	0.74	0.87	0.64	0.99	0.99	0.99	10.1	25.0	17.5

Overall, our approach, as shown on Table 1 reached a superior result in the whole tumor segmentation task with an average dice coefficient of 87% over training dataset and 86% over validation dataset and a sensitivity of 87% in both. Additionally, our neural network reached obtains a specificity of 99% in all the evaluated tasks. The usage of different level resolution and more fully connected layers helped improve the assemble of all feature maps in the trained data. In figure 2 we present some examples of the predictions against the ground truth where we can see the capability of our method to predict the exact area where the patient's tumor occurs with minimal noisy activations in other areas.

4 Conclusions

Based on V-Net and DeepMedic, we proposed a volumetric multimodality neural network. Our method receives as input 3D patches extracted from the dataset volumes and its architecture consist of three identical parallel pathways, each one with six convolutional layers and two residual connections (to extract features on three specific resolution levels), afterwards there are upsample layers (to make the three outputs of the same size) and finally, the results are concatenated and introduced in the fully connected layers to be combined and then classified. In this paper, we have presented preliminary results in the 2017 BraTS Challenge dataset (Training and Validation) reaching an average dice coefficient of 87% over training dataset and 86% over validation dataset for the whole tumor segmentation task (Table 1).

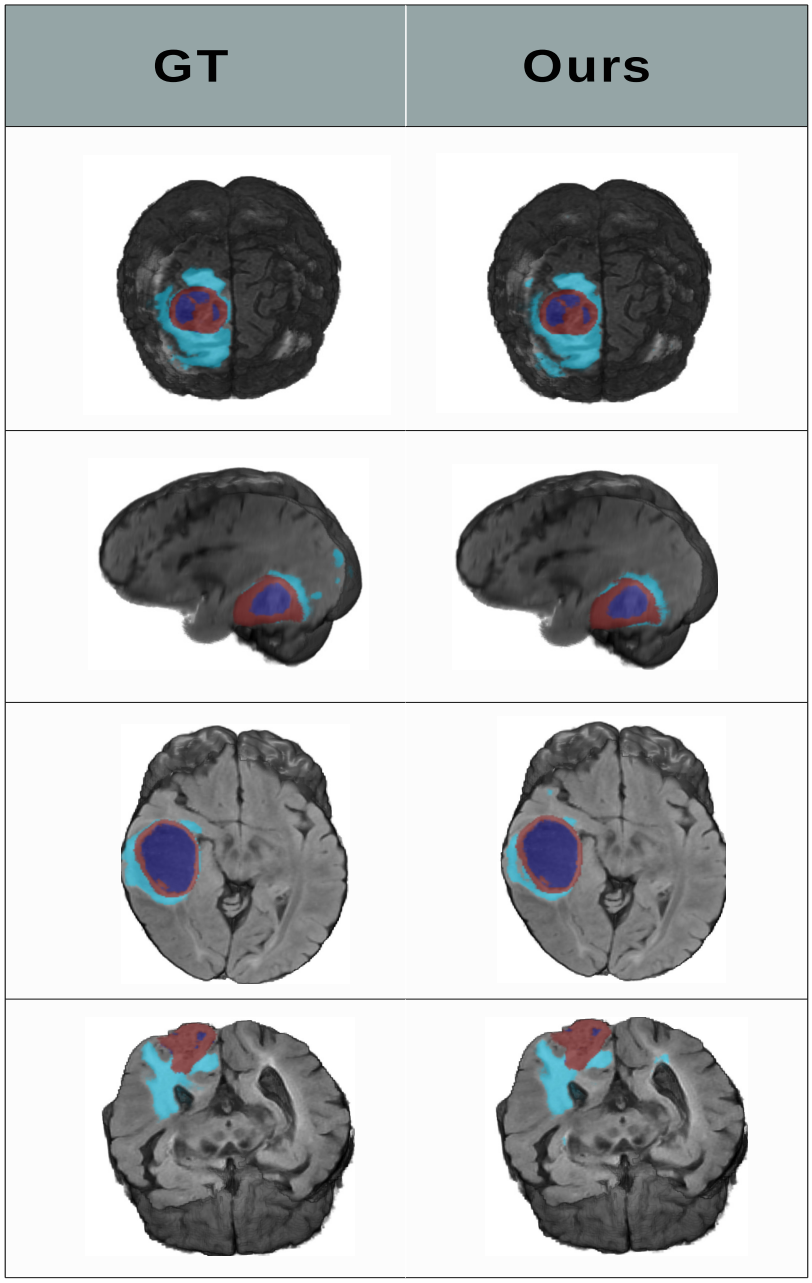


Fig. 2. Comparison between some results obtained by our neural network against the ground truth

References

[1] Mayo Clinic: Brain tumor - Symptoms and causes. http://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/dxc-20117134 Online; accessed July 19, 2017.

[2] Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S.J., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". IEEE Transactions on Medical Imaging **34**(10) (2015) 1993,2024

[3] Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Elsevier **8**(3) (2004) 275–283

[4] Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain mr images. Springer (2012) 651–658

[5] Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M.: Segmentation of brain tumor images based on integrated hierarchical classification and regularization. Proceedings MICCAI-BRATS (2012)

[6] Menze, B.H., Geremia, E., Ayache, N., Szekely, G.: Segmenting glioma in multimodal images using a generative-discriminative model for brain lesion segmentation. Proceedings MICCAI-BRATS (2012)

[7] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. ECCV (2014) 818–833

[8] Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. CoRR **abs/1605.06211** (2016)

[9] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. CoRR **abs/1505.04597** (2015)

[10] Milletari, F., Navab, N., Ahmadi, S.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. CoRR **abs/1606.04797** (2016)

[11] Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic on brain tumor segmentation. Athens, Greece Proc. BRATS-MICCAI (2016)

[12] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Nature Scientific Data ([In Press]) (2017)

[13] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: "segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection". The Cancer Imaging Archive (2017) doi:10.7937/K9/TCIA.2017.KLXWJJ1Q.

[14] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: "segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection". The Cancer Imaging Archive (2017) doi:10.7937/K9/TCIA.2017.GJQ7R0EF.

[15] Rote, G.: Computing the minimum hausdorff distance between two point sets on a line under translation. Information Processing Letters **38**(3) (1991) 123–127

Masked V-Net: an approach to brain tumor segmentation

Marcel Catà, Adrià Casamitjana, Irina Sánchez, Marc Combalia and Verónica Vilaplana

Signal Theory and Communications Department, Universitat Politècnica de Catalunya. BarcelonaTech, Spain
{adria.casamitjana,veronica.vilaplana}@upc.edu

Abstract. This paper introduces Masked V-Net architecture, a variant of the recently introduced V-Net[13] that reformulates the residual connections and uses a ROI mask to constrain the network to train only on relevant voxels. This architecture allows dense training on problems with highly skewed class distributions by performing data sampling on the output instead of in the input. We use Masked V-Net in the context of brain tumor segmentation and report results on the BraTS2017 Training and Validation sets.

1 Introduction

Glioma is the most common type of primary brain tumor arising from glial cells. Gliomas may have different degrees of aggressiveness, variable prognosis and several heterogeneous histological sub-regions that are described by varying intensity profiles across different Magnetic Resonance Imaging (MRI) modalities, which reflect diverse tumor biological properties [1]. Accurate segmentation and measurement of the different tumor sub-regions is critical for monitoring progression, surgery or radiotherapy planning and follow-up studies. However, the distinction between tumor and normal tissue is difficult as tumor borders are often fuzzy and there is a high variability in shape, location and extent across patients. Despite recent advances in automated algorithms for brain tumor segmentation in multimodal MRI scans, the problem is still a challenging task in medical imaging analysis.

Many different computational methods have been proposed to solve the task. Here we will only review some of the most recent approaches based on deep learning, which are the top-performing methods in BraTS challenge since 2014. Representative works based on other machine learning models include [2–6] and methods reviewed in [1].

* This work has been partially supported by the projects BIGGRAPH-TEC2013-43935-R and MALEGRA TEC2016-75976-R financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund (ERDF). Adrià Casamitjana is supported by the Spanish "Ministerio de Educación, Cultura y Deporte" FPU Research Fellowship.

As opposed to classical discriminative models based on pre-defined features, deep learning models learn a hierarchy of increasingly complex task specific features directly from data, which results in more robust features.

Some methods do not completely exploit the available volumetric information and use two-dimensional Convolutional Neural Networks (CNN), processing 2D slices independently or using three orthogonal 2D patches to incorporate contextual information [7, 8]. The model in [8] consists of two pathways, a local pathway that concentrates on pixel neighborhood information, and a global pathway, which captures global context of the slice. This two-path structure is adopted in a fully 3D approach named DeepMedic [9], consisting of two parallel 3D CNN pathways producing soft segmentation maps, followed by a fully connected 3D CRF that imposes generalization constraints and obtains the final labels. The network is extended in [10] by adding residual connections between the outputs of every two layers. They empirically show that the residual connections give modest but consistent improvement in sensitivity over all tumor classes. In [14] we compare the performances of three 3D CNN architectures inspired in two well known 2D models used for image segmentation [11, 12] and a variant of [9] showing the importance of the multi-resolution connections to obtain fine details in the segmentation of tumor sub-regions.

In the context of BraTS Challenge 2017, in this paper we propose a system for segmenting gliomas in multi-institutional, multi-modal pre-operative MRI scans. It involves two steps: a first CNN that performs tumor localization, concentrating on the simpler tumor vs non-tumor problem, and a second CNN which is dedicated to segmenting the different tumor sub-regions. Both networks use the same architecture, named Masked V-Net. Figure 1 shows the complete pipeline.

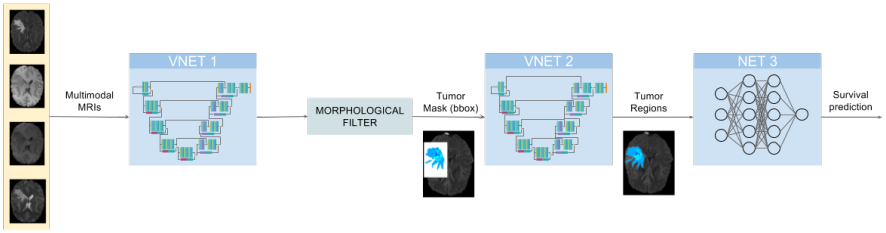


Fig. 1: The pipeline used for brain tumor segmentation

2 Method

2.1 Masked V-Net

In this work, we introduce a variant of the V-Net architecture [13], called Masked V-Net. Inspired by results in [18], we reformulate the residual connections such

that the propagation of the input signal through the network is minimally modified. To account for different dimensions in the shortcut connections, max-pooling operations are used for spatial correspondence, and 1x1x1 convolutions are used to match the number of channels. The final architecture is shown in Figure 2.

In order to constrain the network to train only on relevant, informative voxels, we use a ROI mask at the output. This ROI mask can be used, for example, to discard non-informative, background voxels present in MRI modalities, or to focus on specific brain regions. The ROI mask force all voxels outside the mask to belong to the background class with 100% confidence after the softmax activation.

2.2 Data sampling and training

One of the main problems in brain lesions is that they affect a small portion of the brain, making naive training strategies biased towards the trivial decision of null detection. Brain tumors normally correspond to 5% to 10% of the overall brain tissue, being each tumor region an even smaller region. Besides, a great portion of the image is non informative background masked out by skull-stripping algorithms. Hence, sampling strategies should be adopted to overcome this limitation. The approach adopted in this paper is shown in Figure 1 and uses a concatenation of two Masked V-Nets trained separately.

The first network outputs a raw segmentation of the whole tumor region. It uses the ROI mask of the brain in order to consider only brain tissue for training. To overcome the biased decision towards background class and avoid the use of weights to give importance to whole tumor class, the strategy adopted is based on using a modified dice coefficient as loss function that can be written as:

$$D = \frac{\sum_{i=1}^{N} p_i \cdot l_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} l_i} \tag{1}$$

where N is the total number of voxels, p_i is the softmax output of the i-th voxel, and l_i is the i-th voxel label ($l_i = 0, 1$).

The second network is trained separately using as ROI mask a prism-like mask build from the ground-truth labels, used to simulate non-perfect whole tumor predictions from the first network. Here, we use a combination of cross entropy (X_E) and the dice coefficient for each tumor sub-region (whole tumor (D_{WT}), enhancing tumor (D_{ET}) and tumor core (D_{TC})) for the loss function. We empirically choose the values for the weights in both parts of the function:

$$L = X_E + 0.5 * (D_{WT} + D_{ET} + D_{TC}) \tag{2}$$

During inference time, we concatenate the two networks and place in between a morphological filter to get rid of small spurious detections made by the first network.

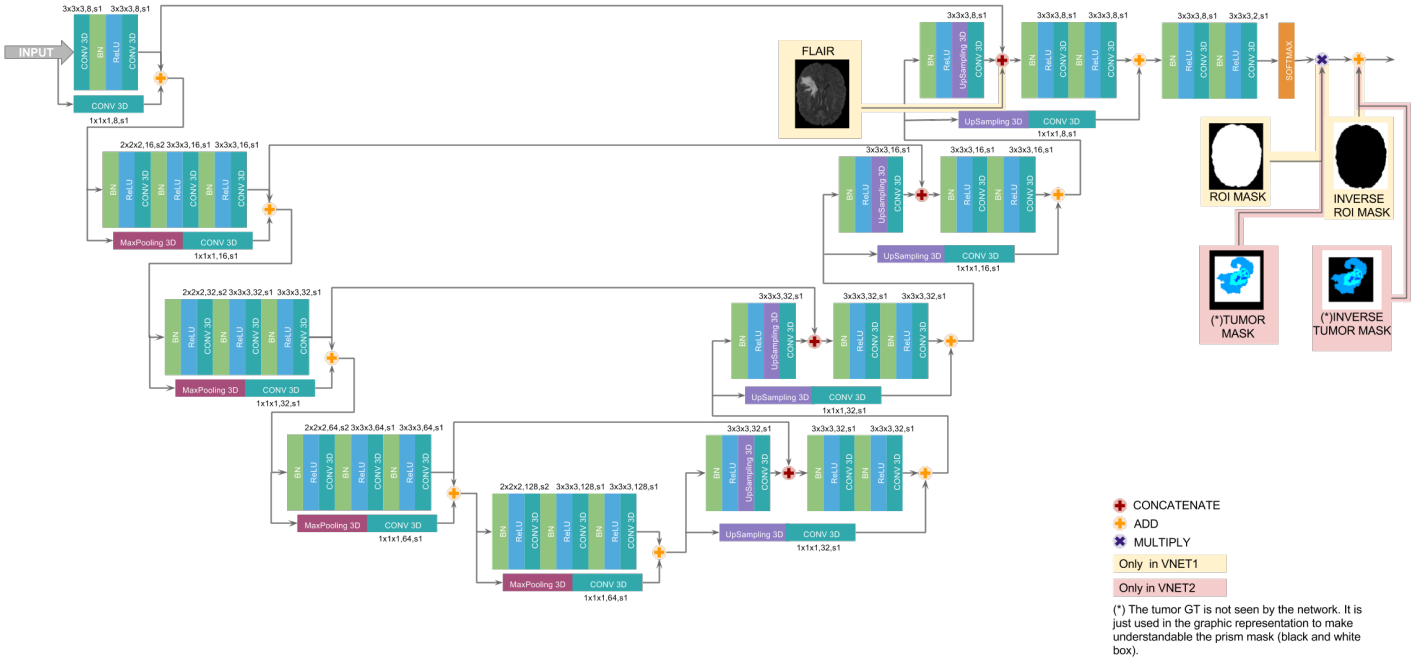


Fig. 2: Masked VNet

3 Results and discussion

3.1 Data

We use BraTS2017 training data [15–17], consisting of 210 pre-operative MRI scans of subjects with glioblastoma (GBM/HGG) and 75 scans of subjects with lower grade glioma (LGG), corresponding to the following modalities: native T1, post-contrast T1-weighted, T2-weighted and FLAIR, acquired from multiple institutions. Ground truth annotations comprise GD-enhancing tumor (ET, label 4), peritumoral edema (ED, label 2), necrotic and non-enhancing tumor (CNR/NET, label 1) as described in [1]. The data is distributed co-registered to the same anatomical template, interpolated to the same resolution ($1mm^3$) and skull-stripped. For the prediction of patient overall survival, the overall survival data (OS) defined in days is also included for the subjects in the training set. The validation set consists of 46 scans with no distinction between GBM/HGG and LGG, with OS data.

Each scan is individually normalized in mean and standard deviation. For training, we use data augmentation by adding scan reflections with respect to the sagittal plane.

3.2 Performance on BraTS2017 training and validation sets

Evaluation of the results is performed merging the predicted labels into three classes: enhancing tumor ET (label 1), whole tumor WT (labels 1, 2, 4), and tumor core TC (labels 1, 4), using Dice score, Hausdorff distance, Sensitivity and Specificity.

Preliminary results for the BraTS 2017 Training dataset have been obtained by hold-out using 70% of the data for training and the remaining 30% for validation purposes. In addition to that, the performance on the BraTS 2017 Validation set, reported on the challenge's leaderboard [1], is also presented in Table 1 and Table 2.

	Dice			Hausdorff		
	ET	WT	TC	ET	WT	TC
Validation (Train)	0.671	0.869	0.685	7.145	6.410	9.584
Validation set	0.714	0.877	0.637	5.434	8.343	11.173

Table 1: Results for BraTS 2017 data. Dice and Hausdorff metrics are reported.

Results presented in Table 1 show high performance on the Dice metric for the whole tumor (WT) region, but low values for enhancing tumor (ET) and tumor core (TC) regions, compared to state-of-the-art. Using the BraTS Validation set, we are able to compare to other participants in the challenge. In the case of

[1] https://www.cbica.upenn.edu/BraTS17/lboardValidation.html

	Sensitivity			Specificity		
	ET	WT	TC	ET	WT	TC
Validation (Train)	0.735	0.851	0.664	0.998	0.994	0.997
Validation set	0.723	0.879	0.619	0.998	0.994	0.998

Table 2: Results for BraTS 2017 data. Sensitivity and specificity are reported.

Dice-WT, our method is very close to the results obtained by the top performing methods while, again, our method achieves rather low Dice-ET and Dice-TC metrics. Hausdorff distances are higher than the best performing algorithms, being specially inflated the whole-tumor region, probably indicating some outlier predictions that increase the metric.

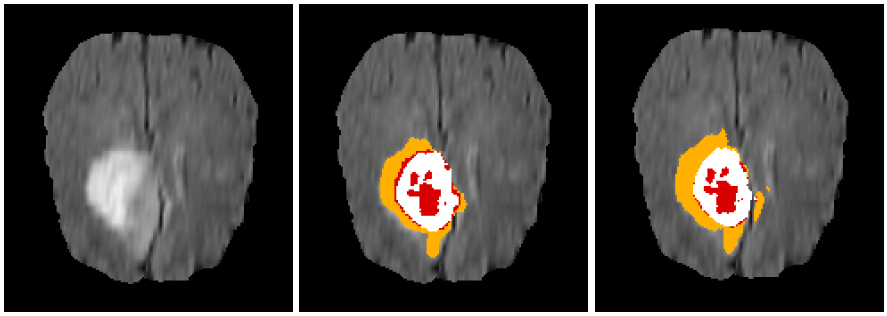
Even though specificity is not very informative for imbalanced classes, results from Table 2 show that we are able to properly represent background, probably due to the use of masks in the predictions. More interestingly, sensitivity shows that ET and TC regions might be underrepresented in our predicted segmentations. This results guide us to future improvements trying to overcome that behavior.

3.3 Visual analysis

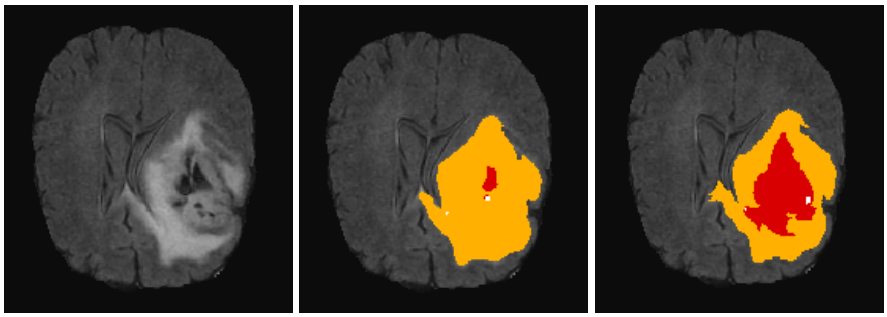
Figure 3 shows two subjects among the quantitatively better (first row) and poorer (second row) results. In both cases, it can be visually appreciated that our method correctly segments the whole tumor region. For the subject shown in Figure 3(a), the system is able to properly capture all tumor regions, meaning that the first network is able to correctly localize the tumor and the second network is able to capture differences between tumor regions. On the other hand, in Figure 3(b), we show a case where even though the tumor is correctly localized by the first network, the second isn't able to properly detect different tumor subregions. We see that edema (ED - label 2) is overrepresented in our segmentation to the detriment of smaller classes: GD-enhancing tumor (ET - label 4) and the necrotic and non-enhancing tumor (NCR/NET - label 1). This effect can also be inferred from lower values in ET and TC dice coefficients.

4 Conclusions

In this paper we introduce the Masked V-Net architecture that uses masks to focus training on relevant parts of the brain. We use it to solve the class imbalance problematic of the brain tumor segmentation task. We use a two-step process that (i) localizes brain tumor area and (ii) distinguishes between different tumor regions, ignoring all other background voxels. This scheme allows us to perfom dense-training on MR images. We finally show results on BraTS 2017 Training and Validation sets, showing that while the results obtained for the WT segmentation are competitive with other participants' algorithms, we aren't able to properly capture the less common regions (TC or ET).



(a)



(b)

Fig. 3: Segmentation results of two subjects: a) TCIA 479 b) TCIA 109. From left to right we show the FLAIR sequence, followed by Prediction and GT tumor segmentation. We distinguish intra-tumoral regions by color-code: enhancing tumor (white), peritumoral edema (orange) and necrotic and non-enhancing tumor (red).

References

1. Menze, B.H., Jakab, A., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). Medical Imaging, IEEE Trans. on 34(10), 1993–2024 (2015).
2. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. MICCAI. Springer, 2011.
3. Zikic, D., Glocker, B., Konukoglu, E. , et al. "Decision forests for tissue-speciffc segmentation of high-grade gliomas in multi-channel MR." MICCAI, Springer, 2012.
4. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain mr images. MICCAI, Springer, 2012.
5. Maier, O., Wilms, M. and Handels, H. "Image Features for Brain Lesion Segmentation Using Random Forests." BrainLes Workshop, MICCAI, Springer, 2015.
6. Tustison, N., Shrinidhi, K., Wintermark, M., et al. Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. Neuroinformatics, 2015
7. Pereira, S., et al. "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI." BrainLes Workshop, MICCAI, Springer, 2015.

8. Havaei, M., Davy, A., et al., "Brain tumor segmentation with deep neural networks." Medical Image Analysis, 2016.
9. Kamnitsas, K., Ledig, C., Newcombe, V. et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation", Medical Image Analysis, Volume 36, 2017.
10. Kamnitsas, K., Ferrante, E., Parisot, S., et al, "DeepMedic for Brain Tumor Segmentation", BrainLes Workshop, MICCAI, Springer, 2016.
11. Long, J., Shelhamer, E., Darrel, T.: Fully Convolutional Networks for Semantic Segmentation. CVPR, Boston, USA, 2015
12. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." MICCAI. Springer, 2015.
13. Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." 3D Vision (3DV), 2016 Fourth International Conference on. IEEE, 2016.
14. Casamitjana, A., Puch, S., Aduriz, A., Vilaplana, V., "3D Convolutional Neural Networks for Brain Tumor Segmentation: a comparison of multi-resolution architectures", BrainLes Workshop, MICCAI, 2016.
15. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 2017.
16. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
17. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
18. He, K., Zhang, X., et al. "Identity mappings in deep residual networks." European Conference on Computer Vision. Springer, 2016.

Brain Tumor Segmentation with Label Distribution Learning and Multi-Level Feature Representation

Shengcong Chen, Changxing Ding, and Chenhong Zhou

South China University of Technology, Guangdong, China

Abstract. Convolutional neural networks have been widely adopted for brain tumor segmentation. Patches-wise networks predict the label of central pixel, which is insufficient to represent the whole patch. Therefore, label distribution is introduced into the network to improve training. In addition, to distinguish subtle difference, multi-level feature maps are utilized to predict labels of central pixels. We participate in BRATS 2017 and Dice of complete tumor, tumor core and enhancing tumor in validation dataset is around 0.87, 0.74 and 0.65.

1 Introduction

To our knowledge, CNN-based methods for BRATS challenge are either global-based or patch-wise. Networks yield segmentation results of whole images directly, like fully-convolutional neural networks(FCN) [7] and SegNet [8], belong to global-based networks, while patch-wise networks predict labels of central pixels.

In this article, we study patch-wise convolutional neural networks. Patch-wise segmentation methods [5][6] predict labels of central pixels or voxels. But labels of central pixels cannot represent all the content within patches. We introduce label distribution and multi-feature networks into brain tumor segmentation.

2 Method

2.1 Overview

Figure 1 shows the proposed 13-layer patch-wise neural network, which includes two branch. The first branch predicts label distribution of the whole patch, which enables the network to extract global features. The other branch utilizes features cropped from multi layers, and predicts the label of central pixel in the patch. Features from higher layers include more abstract information, while those from lower layers include more detailed information.

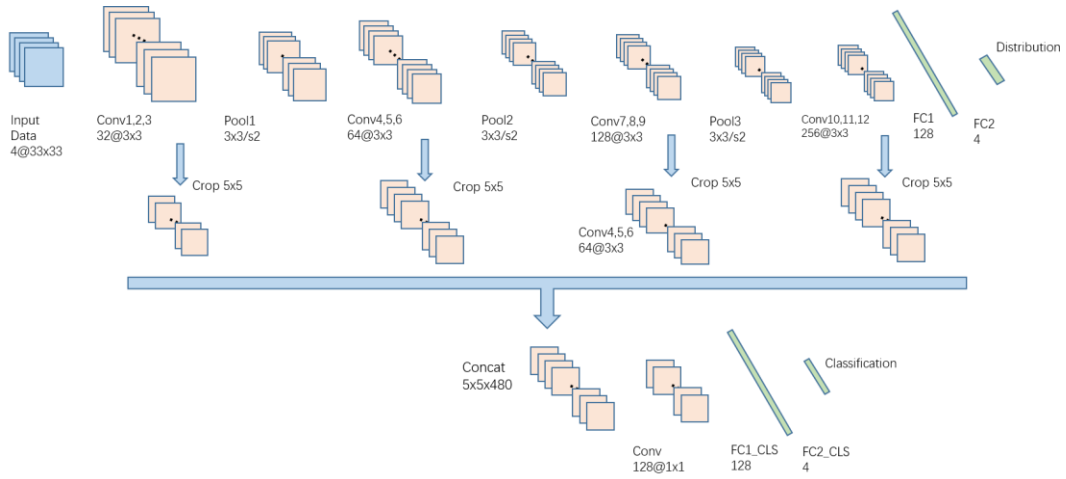


Fig. 1. proposed architecture

2.2 Label Distribution

Traditional patch-wise networks are designed for predicting labels of central pixels, which is not continuous in boundary, i.e., labels of central pixels cannot describe all the information included within patches. Therefore, we construct label distribution to describe content within patches. It is constructed through weighted sum of the label map in the whole patch, which can describe the patch better. To focus on information in central region, we employ Gaussian distribution to construct the weights of voxels, therefor the pixel closer to center has better weight, and is more important for the patch.
To optimize the network, we take symmetrical KL divergence as the loss function.

$$symKL(\hat{p}_i, d_c) = \sum_c (d_c - \hat{p}_c)\, log(\frac{d_c}{\hat{p}_c}) \quad (1)$$

2.3 Multi-level feature architecture

Generally, convolutional neural networks take features yielded from the last layer to make prediction, which is more abstract and represent global information within the patch. There are subtle differences within different tumor categories, which requires more detailed information. And features yielded from lower layers involve more detail information e.g. edges, corners.
Therefore, low-level information is also employed to make prediction. Considering that detail features far from the central pixels may be helpless for predicting task, we crop central regions from low-level feature maps and cascading them with feature maps yielded in the last convolutional layer.

3 Experiments

3.1 Preprocessing and Training

In BRATS 2017, there are four kinds of labels, e.g. GD-enhancing tumor, peritumoral edema, necrotic and non-enhancing tumor, and there are four different modalities for MRI data, e.g. native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). [1] There are 274 training objects provided in BRATS 2017 and we normalize pixels in brain region of each object using average and standard deviation compute from the region of each object for

pre-processing.

Similar to [6], we train the proposed network with class-balance training dataset and then finetune with randomly sampled dataset. At the first training stage, label distribution is required for optimizing the network, which help the network to extract features. And at the second stage, it is omitted because we optimize the last FC layer only, which transform the distribution of output to its original distribution.

3.2 Result

Table 1 shows the results of our proposed model evaluated in BRATS 2017 validation dataset. Complete region(WT) includes all labels 1,2,4. Core region includes labels 1,4. And enhancing tumor region includes only labels 4. There are three kinds of evaluation criteria, Dice, Sensitivity and Specificity.

$$DSC = \frac{2TP}{2TP+FP+FN} \tag{2}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{3}$$

$$Specificity = \frac{TP}{TP+FP} \tag{4}$$

Table 1. Evaluation result of our proposed method in BRATS 2017 validation dataset

	Dice	Sensitivity	Specificity
WT	0.86771	0.88816	0.99208
TC	0.73782	0.75746	0.99615
ET	0.64958	0.76636	0.99716

4 Conclusion

In this paper, label distribution is introduced to help patch-wise neural networks extract features, and multi-level features are utilized to distinguish subtle differences. For the future work, we may extend our proposed method to 3D networks. Our Dice of complete tumor, tumor core and enhancing tumor in BRATS 2017 validation dataset is around 0.87, 0.74 and 0.65.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Pereira, Sérgio, et al. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images[J]. IEEE Transactions on Medical Imaging 35.5 (2016): 1240-1251.
6. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with deep neural networks[J]. Medical image analysis, 2017, 35: 18-31.
7. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. arXiv preprint arXiv:1411.4038 (2014)
8. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolution decoder architecture for image segmentation. arxiv:1511.00561 (2015)

A Separate 3D-SegNet Architecture for Brain Tumor Segmentation

Shidu Dong [1]

[1] Computer Science and Engineer College, Chongqing University of Technology, China, 40054
shidu_dong@uri.edu

Abstract. We propose separate 3D-SegNet by applying combination 2D-SegNet with 1D-SegNet for brain tumor segmentation challenge. First, 2D-SegNet which has 4 encoding layers and 4 decoding layers is adopted to gain the 2D-features in a slice. Second, the features in the same position of each slice are integrated into 1D features. Third, to use context along the z-axis, the 1D-features are fed into 1D-SegNet which has 4 encoding and 4 decoding layers and then are classified feature-wise. Experimental results show that the separate 3D-SegNet can obtain higher

Keywords: SegNet, Deep Learning, Semantic Segmentation.

1 Introduction

Segmentation and the subsequent quantitative assessment of brain tumor provide valuable information for the analysis of neuropathologies and are important for planning of treatment strategies, monitoring of disease progression and prediction of patient outcome [1]. In the last decade, many automatic tumor segmentation methods have been proposed. many of them use hand-designed features method whose features such as local histograms, Gabor and region shape difference are extracted firstly, and then given to the classifier such as SVM and Random decision forest whose training procedure does not affect the nature of those features [2]. Many methods also utilize conditional random field (CRF) as post processing to improve automatic segmentation results [3].

In recent years, Deep convolution neural networks which have been shown to excel learning a hierarchy task-adapted complex feature from in-domain data has been seen huge success in image Classification[4,5,6], object detection and image semantic segmentation[7,8,9].Some of the brain tumor segmentation based on Deep convolution neural networks are also proposed. To gain more global contextual features in a slice, two paths and cascade CNN architectures are proposed [2], however, the context features among slices have not been employed. To incorporate 3D contextual, a dual pathway, 11-layers deep, three-dimensional Convolutional Neural Network are presented [1]. Due to being based on patch-wise, only limited space context features are explored, and many redundant convolution calculations are contained. FCN with deconvolution layers and Segnet with a symmetric encoder-decoder architecture train an end-to-end and pixel-to-pixel convolutional neural network for pixel-wise prediction with the whole image as inputting, to avoid to use patch [7].

Due to the huge volume of data of 3D brain scan images and limited GPU memory, it is impossible to extend directly FCN and SegNet to 3D to explore the 3D context for brain tumor segmentation challenge. We propose separate 3D-SegNet for brain tumor segmentation. The 2D SegNet is utilized to capture the 2d context features and 1D SegNet is adopted to get the context features among the slices along the z-axis. Without local patch-wise training and testing, our method can obtain more 3D context information with broader receive field and can avoid redundant convolutional calculate.

2 Architecture

Brain MRI images for Brats 2017 Challenge has 155 slices, 240×240 resolution, and 4 channels, i.e. FLAIR, T1, T1-contrast, and T2. As being illustrated in Fig. 1, our separate 3D-SegNet contains 2D-SegNet and 1D-SegNet parts. First, MRI images are normalized firstly, and then each slice of them are fed into 2D-SegNet to gain $240 \times 240 \times 64$ features. Second, the features in the same position of 155 slices are integrated into 1D features which contain 64 channels and then they are inputted to the 1D-SegNet to be classified into labels.

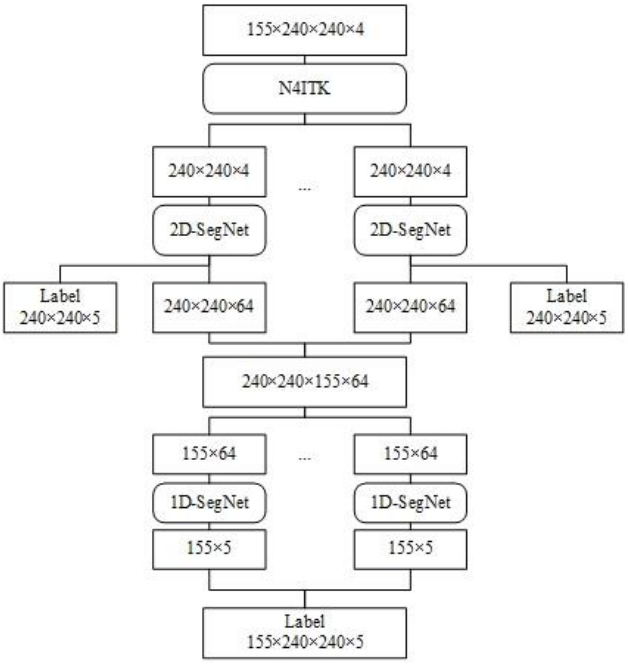


Fig. 1 An illustration of the separate 3D-SegNet architecture

Unlike [9], in this paper, the 2D-SegNet only adopts 4 encoding and decoding layers as being depicted in Fig.2, due to 240×240 resolution of brain MRI image. Each Encoder has 2 or 3 2-dimension convolution layers, whose kernel size are 3×3, and a

maxpooling layer with 2×2 window and stride 2, whose outputs contain features sub-sampled by a factor 2 and maxpooling indices (MaxInd for short) implying the location of the maximum feature value in each pooling window[2] (see Fig.2 B). Sub-sampling results in a large input image context (spatial window) for each pixel in the feature maps and maxpooling indices are to capture the edge information [2]. Each Decoder has an up-sample layer and 2 or 3 convolution layers (see Fig. 2 C). Upsample takes the outputs of the previous layer and the maxpooling indices of the corresponding encoding layer as input and its output is sparse feature map(s) up-sampled by a factor 2 (see Fig. 2 D [9]). 240×240×64 features of each slice outputted by the last Decoder are fed to the next stage, i.e. 1D-SegNet, and the softmax layer to be classified into labels which can be used in training.

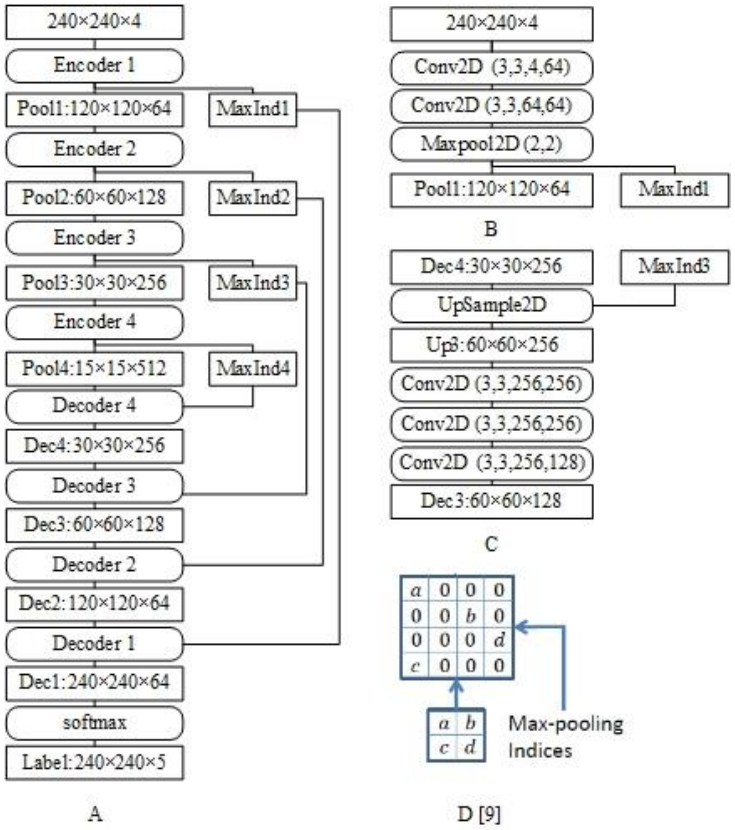


Fig.2 A. An illustration of the 2D-SegNet, B. Encoder 1, C Decoder 3

In order to capture the contexts among slices, we proposed 1D-SegNet to process the output of 2D-SegNet. As being illustrated in Fig. 3, the 1D-SegNet contains 3 encoder and decoder layers. Each Encoder has 2-3 1D convolution layers and a 1D max-pooling which can figure out the features sub-sampled by a factor 2 and maxpooling indices which denote the location of the maximum feature value in each pooling window by the following equation (see Fig.3 B),

$$p_i^l = max\left(f_{2i}^l, f_{min(2i+1,L_l)}^l\right)$$
$$MaxInd_l^i = \begin{cases} 2i & p_i^l = f_{2i}^l \\ 2i+1 & p_i^l \neq f_{2i}^l \end{cases} \quad i \in \left[0, \left\lceil \frac{L_l}{2} \right\rceil\right), \tag{1}$$

where p denotes the upsamped feature, and i, l denotes ith position and lth encoder layer. Each Decoder has an upsample layer, and 2-3 1D convolution layers (see Fig.3 C). With the following equation, upsample layer can restore the feature of position stored in MaxInd.

$$f_{MaxInd_i^l}^l = p_i^l, \quad i \in [0, \left\lceil \frac{L_l}{2} \right\rceil) \tag{2}$$

Training procedures are divided into 2 stages. First, like [9], 2D-SegNet is trained end-to-end with 2D images and their corresponding labels. And then, features of 155 slice of a patient image in training set extracted by the trained 2D-SegNet are reorganized into 240×240 1D features. Second, 1D-SegNet is trained with the reorganized feature and their corresponding label.

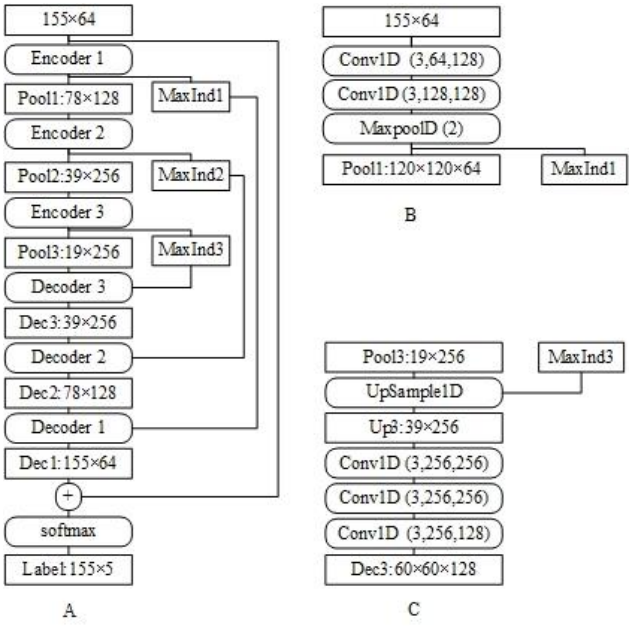


Fig.3 A. An illustration of the 1D-SegNet, B. Encoder 1, C Decoder 3

Brain tumor segmentation image exhibits highly-skewed class distribution. In training set, the background and health region (label 0) have 1637 times more pixels than the necrotic and non-enhancing tumor (label 1). To cope with this problem, we use the weighted cross-entropy loss [9] as the objective function for training the network, which assigns higher misclassification costs to the minority class than to the majority [12] as depicted in Tabel 1.

Tabel 1. The misclassification costs

Label	weight
0	0.00247
1	3.94
2	0.302
4	1.0

3 Experiments and Results

We use the real patient data from 2017 brain tumor segmentation challenge to test the performance of 2D-SetNet and 3D-SegNet [10, 11, 12]. Since only labels for the training set for Brats 2017 Challenge are available, we pick up the 10 percent of images as the testing set. As a result, only the 90 percent of training-set are utilized to train our network.

Table 2 shows the average dice distance of 2D-SegNet and separate 3D-SegNet using the HGG testing set images picked up from the training set. The second column is the result of 2D-SegNet with 4 Encoders and 4 Decoders as illustrated in Fig.2. The last column is the result of 3D-SegNet. From the table, we can see that 3D-SegNet obtain higher accurate than 2D-SegNet.

Table 3 shows the Performance of 2D-SegNet and Separate 3D-Segnet on validation set on the official web. From the table, we can see 3D-Segnet is better almost on all the evaluation metrics.

Table 2. Mean Dice Distance of 2D-SegNet and Separate 3D-Segnet

Label	2D-SegNet	3D-Segnet
1	0.47	0.54
2	0.54	0.69
4	0.69	0.71

Table 3. Performance of 2D-SegNet and Separate 3D-Segnet on validation set

	2D-SegNet	3D-Segnet
Dice_ET	0.58276	0.60036
Dice_WT	0.65568	0.79767
Dice_TC	0.63289	0.69021
Sensitivity_ET	0.72381	0.71095
Sensitivity_WT	0.95877	0.91027
Sensitivity_TC	0.79379	0.81491
Specificity_ET	0.99616	0.99695
Specificity_WT	0.94817	0.97916
Specificity_TC	0.98731	0.99049
Hausdorff95_ET	41.35682	21.0903
Hausdorff95_WT	56.62996	23.64621
Hausdorff95_TC	59.78678	26.39409

4 Conclusion

We presented the separate 3D-SegNet which apply combination the 2D-SegNet and 1D-SegNet to capture 3D-space context. The experimental results shows that the 3D-SegNet can obtain higher accurate than 2D-SegNet.

References

1. Kamnitsas, Konstantinos, et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation." Medical image analysis 36, 61-78 (2017).
2. Havaei, Mohammad, et al. "Brain tumor segmentation with deep neural networks." Medical image analysis 35, 18-31 (2017).
3. Krähenbühl, Philipp, and Vladlen Koltun. "Efficient inference in fully connected crfs with gaussian edge potentials." Advances in neural information processing systems (2011).
4. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems (2012).
5. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
6. Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9, (2015).
7. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015).
8. Noh, H., Hong, S., & Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision, 1520-1528 (2015).
9. Badrinarayanan, V., Kendall, A. and Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for scene segmentation. IEEE transactions on pattern analysis and machine intelligence (2017).
10. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).
11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press].
12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of

the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q

Using NiftyNet to ensemble convolutional neural nets for the BRATS challenge.

Zach Eaton-Rosen[1], Wenqi Li[1], Guotai Wang[1], Tom Vercauteren[1], Sotirios Bisdas[2,3], Sebastien Ourselin[1], M. Jorge Cardoso[1]

[1]Translational Imaging Group, Centre for Medical Image Computing, University College London, UK
[2] Department of Brain Repair and Rehabilitation, Institute of Neurology, University College London, UK
[3] Department of Neuroradiology, The National Hospital for Neurology and Neurosurgery, UCL Hospitals NHS Trust, London, UK

Abstract. Convolutional neural nets have become popular for a variety of applications in medical imaging, including image segmentation, image regression, super-resolution and volume-level classification.
Training neural nets is computationally expensive and labour-intensive. Although some nets are available, pre-trained, with open-source code, they still employ a variety of deep-learning platforms. This makes it burdensome to compare nets, and difficult to use them together (for example, to ensemble the results).
NiftyNet is an open-source medical imaging software package that implements state-of-the-art nets in TensorFlow. In this submission, we train a variety of neural nets for segmentation of the BRATS data. We then ensemble these independently-trained nets to improve the overall performance. This highlights the versatility of the NiftyNet platform.

1 Introduction

The BRATS challenge [1][2][3][4] consists of two parts. One is the long-running segmentation challenge for high- and low-grade gliomas. The aim of the challenge is to match expert segmentations as closely as possible.

The second part of the challenge is to predict survival time from MR images and the subject's age.

2 Methods

In this paper, we train several different convolutional neural nets independently on the provided data. These nets are available on the NiftyNet platform. The nets we used are detailed in Table 1.

The training data is preprocessed using histogram normalisation and whitened. The Adam optimiser [5] is used for all training cases, because of its robustness to the chosen learning rate parameter.

To ensemble the methods, we propose three different approaches:

Table 1. For training, data was augmented with a left-right flip with $p = 0.5$. We adopted a selective sampling strategy: 90% of the image patches examined have at least two labels (of the possible 4) by specification. The learning rate lr was determined through trial and error.

Net Name	**Notes**	**Parameters**	**Reference**
DeepMedic		$lr = 0.01$, optimiser=Adam	[6]
Dense VNet		$lr = 0.0001$, optimiser=Adam	[7]
High-Res net	Fit three variants	$lr = 0.01$, optimiser=Adam	[8]

– M1 — Majority voting: The modal value of the label prediction is used across all nets.
– M2 —Shallow ensembling: The predictions from each model are put through a neural network with one hidden layer to allow different weightings to the different nets.
– M3 —Deep ensembling: the input images are fed into a CNN, along with each net's predicted output, and trained to best mimic the ground truth.

M1 is the case of M2 with equal weights, with M2 being a special case of M3 in turn.

3 Results

Table 2. A summary of the individual net results and the results of majority voting.

Method	**Dice ET**	**Dice WT**	**Dice TC**	**Sensitivity ET**	**Sensitivity WT**	**Sensitivity TC**
Deepmedic	0.6822	0.83658	0.71912	0.75992	**0.88855**	0.78894
Dense Vnet	0.65226	0.836	0.70668	0.69923	0.85884	0.73753
HR (small)	0.6755	0.80537	0.67083	0.72231	0.76896	0.68916
HR (default)	0.67521	0.81748	0.6585	0.68578	0.77202	**0.69939**
HR (large)	**0.71195**	**0.87083**	0.66603	**0.77353**	0.8749	0.6324
Majority Vote	0.70288	0.83321	**0.73451**	0.72845	0.77727	0.69404
	Specificity ET	**Specificity WT**	**Specificity TC**	**Haussdorf 95 ET**	**Haussdorf 95 WT**	**Haussdorf 95 TC**
Deepmedic	0.99796	0.9905	0.99414	17.35533	27.49044	31.34963
Dense Vnet	0.99806	0.9908	0.99511	8.4005	24.90084	29.17855
HR (small)	0.99835	0.996	0.99693	6.89773	34.20315	57.31446
HR (default)	**0.99881**	0.99575	0.99486	6.047	8.25249	12.01527
HR (large)	0.9976	0.99459	0.99859	7.6967	10.61978	15.42364
Majority Vote	0.9986	**0.99773**	**0.9989**	**5.17102**	**7.35433**	**8.93822**

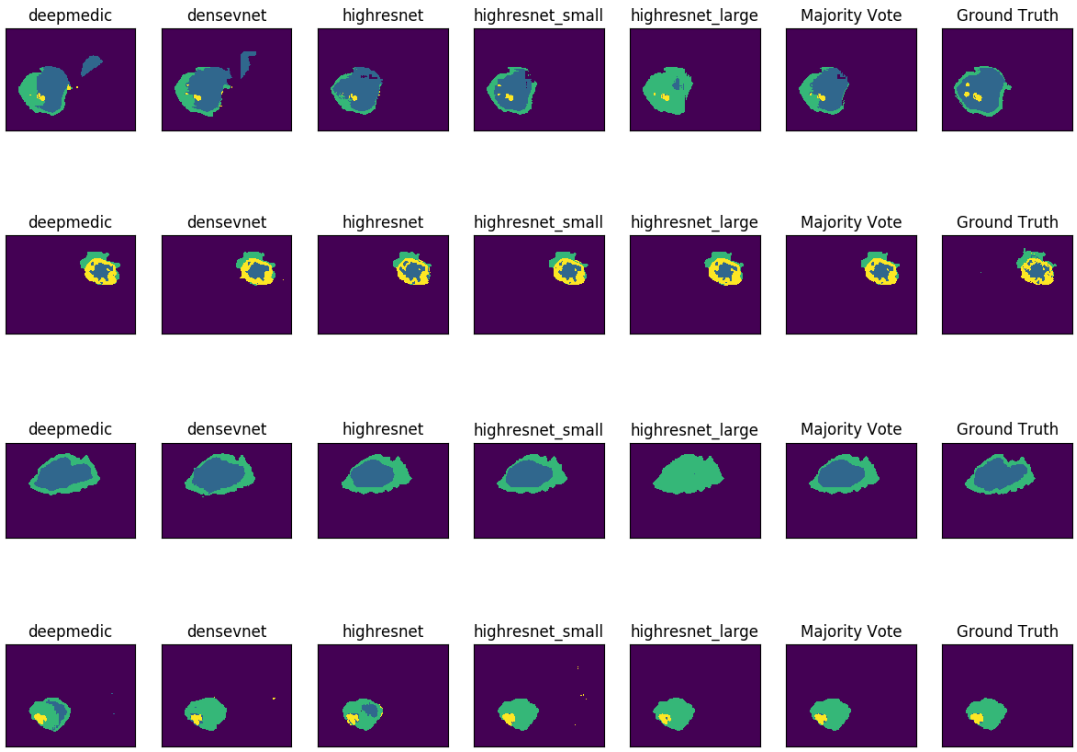


Fig. 1. Majority voting is effective at reducing spurious segmentation labels (four arbitrarily-chosen subjects).

In Figure 1 we can see some of the estimated tumour segmentations from the various models, and the result of 'majority voting'. In Table 2 we have the results on the validation data, assessed via the CBICA portal.

This is only the reults for 'M1': Precise architectural details of M2/M3 will be added after the challenge data is released, as they are not finalised.

4 Prediction Challenge

In this part of the BRATS Challenge, the object is to predict the post-scan survival time. The data provided is the MRI scans and the patient age, with survival time available for the test data. In this task, we try to use imaging features to predict survival time.

To generate imaging features, we used the images after histogram normalisation, as above. For each segmentation label, we first computed its volume (3 features). We also calculated the mean, standard deviation, and kurtosis of each modality for each tumour region (3 parameters $\times$ 3 labels $\times$ 4 modalities $= 36$ features). The location of the tumour will also be informative as to its effects. For the 'whole tumour' label, we computed the fraction of the tumour in each of 27 regions. Including subject age yields $(3 + 36 + 27 + 1) = 67$ features, with 163 labeled subjects.

This problem is especially prone to overfitting, because of the small number of available subjects and the lack of pertinent clinical features (for example, we do not have patient sex, blood type, etc.). We chose to use a Bayesian Ridge Regression to estimate the survival age.

Our best result was an R^2 score of 0.17 on train data, using 3-fold cross-validation (Figure 2).

5 Discussion

In this work, we have shown that ensembling models improves their performance on the Hausdorff metric. We have remaining work to do on how best to combine the predictions.

Majority voting does badly on the 'whole tumour' label, so probably is penalising discrepancies between estimations too much.

The code used to produce this paper will be released fully upon completion of the challenge.

Acknowledgements: We gratefully acknowledge the support of NVIDIA Corporation for the donation of a Titan X Pascal GPU used in this work.

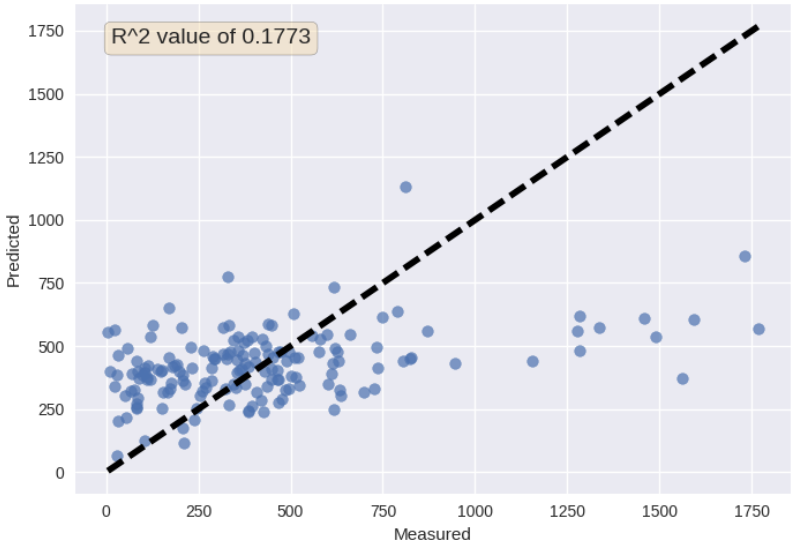


Fig. 2. This figure shows the 3-fold cross-validation result for predicted vs measured survival. The dashed line is unity.

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10) (2015) 1993–2024
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels for the pre-operative scans of the tcga-gbm collection (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels for the pre-operative scans of the tcga-lgg collection (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Nature Scientific Data (in press)
5. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
6. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Medical image analysis **36** (2017) 61–78
7. Gibson, E., Giganti, F., Hu, Y., Bonmati, E., Bandula, S., Gurusamy, K., Davidson, B., Pereira, S.P., Clarkson, M.J., Barratti, D.C.: Automatic multi-organ segmentation on abdominal ct with dense v-networks (in preparation)
8. Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T.: On the compactness, efficiency, and representation of 3d convolutional networks: Brain parcellation as a pretext task. In: International Conference on Information Processing in Medical Imaging, Springer (2017) 348–360

Patch-Based 3D U-Net for Brain Tumor Segmentation

Xue Feng[1[0000-0002-2181-9889]] and Craig Meyer[1, 2]

[1] Biomedical Engineering, University of Virginia, Charlottesville VA 22903, USA
[2] Radiology & Medical Imaging, University of Virginia, Charlottesville VA 22903, USA
xf4j@virginia.edu

Abstract. Accurate segmentation of different sub-regions of gliomas including peritumoral edema, necrotic core, enhancing and non-enhancing tumor core using multimodal MRI scans has important clinical relevance in diagnosis and treatment. However, due to the highly heterogeneous appearance and shape, segmentation of brain tumors is very challenging. Recent development using deep learning models has proved its effectiveness in the past few brain segmentation challenges as well as other semantic and medical image segmentation problems. Most models in brain tumor segmentation use a 2D/3D patch to predict the class label for the center pixel. Variant patch sizes and scales are used to improve the model performance. U-Net is a popular network structure for end-to-end full-image segmentation. However, the heterogeneous appearance, shape and locations of brain tumors and large image size make the application of U-Net on the whole image ineffective and impractical. We propose to use a 3D U-Net structure on extracted patches to predict the class labels for all pixels in the patch. In this paper we will describe the patch extraction, network structure, training and deploy methods and pre and post processing algorithms. Preliminary results showed effectiveness of this model.

Keywords: Brain Tumor Segmentation, 3D U-Net, Deep Learning.

1 Introduction

Gliomas are the most common primary brain malignancies, with different degrees of aggressiveness, variable prognosis and various heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. The Multimodal Brain Tumor Segmentation Challenge (BraTS) 2017 utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous brain tumors [1-2]. The dataset used in this challenge includes multiple-institutional clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and low grade glioma (LGG) containing a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) Fluid Attenuated Inversion Recovery (FLAIR) volumes [3-4]. 285 training volumes with annotated GD-enhancing tumor, pertumoral edema and necrotic and non-enhancing tumor and the overall survival data defined in days are provided. 46 volumes are used for validation with the

following two goals: 1) provide pixel-by-pixel label maps for the three sub-regions and background; 2) estimate the survival days.

Convolutional neural network (CNN) based models have proved their effectiveness and superiority over traditional medical image segmentation algorithms and are quickly becoming the mainstream in BraTS challenges. Due to the highly heterogeneous appearance and shape of brain tumors, small patches are usually extracted to predict the class for the center pixel. To improve model performance, multi-scale patches with different receptive field sizes are often used in the model [5]. U-Net is a very popular convolutional network structure that consists of a contracting path to capture context and a symmetric expanding path that enables precise localization with 3D extension [6-7]. In this paper we propose to apply the 3D U-Net structure on extracted patches to perform a full-patch segmentation. During test, a sliding window approach is used to predict class labels with adjustable overlap to improve accuracy. Since the CNN model only performs pixel level prediction, post processing methods are often used to refine the final label maps. As our proposed method is prone to false positives, a post processing method to compensate for that is also developed.

2 Methods

The steps in our proposed method for brain tumor segmentation include pre-processing of the images, patch extraction, training using a 3D U-Net structure, usage of the model for full volume prediction, post processing to refine label maps, as described below.

2.1 Image Pre-processing

To compensate for the MR inhomogeneity, the bias correction algorithm based on N4ITK is first applied to the T1 and T1Gd images [8]. To reduce the effect of the absolute pixel intensities to the model, an intensity normalization step is applied to each volume of all subjects by subtracting the mean and dividing by the standard deviation so that each MR volume will have a zero mean and unit variance. In practice, as the original uncropped volume is used in which the brain only takes the central region, the mean and standard deviation are estimated using the central region (0.25 - 0.75 on each dimension) of the volume.

2.2 Patch Extraction

For simplicity, we will use foreground to denote all tumor pixels and background to denote the rest. To increase the predicting power of the model, the extracted patches should cover all possible cases including all background, mostly background, mostly foreground and equally distributed. However, as the foreground labels contain much more variability and are difficult to segment, the patches should favor those with more

foreground pixels to learn a wide range of tumor characteristics and those with equally distributed foreground and background pixels to improve the boundary accuracy.

A total of 400 patches with size 64 * 64 * 64 are extracted per volume. To avoid clustering on a small region for patch extraction, a total of 3200 possible patch center positions are calculated by applying a small random perturbation to the equally distanced 3D grids. Different perturbations are applied to different subjects. The relative probability p that a patch center position is selected is given as:

$$p = (1 - abs(r - 0.5) * r \quad [1]$$

in which r is the ratio of foreground pixels in the patch. It can be easily proven that $r = 0.75$ corresponds to the maximum chance of being selected. To be able to select all background patches, p is adjusted to the minimum non-zero value when $r = 0$. The probabilities of all 3200 positions are normalized to have a sum of 1 before drawing 400 from them.

2.3 Network Structure and Training

A 3D U-Net based structure containing 3 encoding and 3 decoding layers is used in our model. Intensity normalized images with all MR protocols are concatenated as the 4-channel input for the model. The corresponding segmentation map of the input patch is given as the ground truth labels. The number of features in each encoding layers are 48, 96, 192, respectively. A VGG like network with two consecutive 3D convolutional layers with kernel size 3 followed by the rectified linear and batch norm layers is used in the encoding path. A major difference with the original U-Net structure is the handling of edges. Zero padding is used in our model to maintain the sizes of the decoding paths without cropping. The output of the model is 64 * 64 *64 * 4 matrix containing the probability of each input voxel belonging to each category (3 foreground classes and the background).

4 epochs are used on all extracted patches, which takes about 60 hours to train on a Titan X GPU. No data augmentation such as rotation and mirroring is used due to the concern of longer training time, although we hypothesize that it can further improve the model performance.

2.4 Volume Prediction Using the Model

To make a prediction for the full volume, a sliding window approach is used with stride size 16. Therefore, each voxel will be predicted by 4 models and the mean output is used as the final probability. For the given image size, it takes 5 minutes to generate the output for the entire volume on the same GPU.

2.5 Post Processing

Since more foreground pixels are used in the training steps than the background pixels, this model tends to have more false positives, causing low specificity and high Hausdorff distance. To remove false positives, we hypothesize that the tumors are most

likely to occur in one or multiple connected regions with a relatively large weighted area and thus the isolated regions with small weighted areas are false positives. The following steps are used in the removal process: 1) predict the label map by taking the maximum probability and binarize it to background (0) and foreground (1) pixels; 2) apply binary dilation to the label map with iterations 10; 3) perform the connection analysis to get all connected foreground regions; 4) for each foreground region, calculate the weighted area as the sum of the pixel probabilities belonging to foreground of all pixels in this region; 5) apply a thresholding so that the regions with weighted area smaller than ½ of the maximum weighted area are re-classified as background.

3 Results

All 285 training dataset are used in the model training process. With multiple submissions to the CBICA's Image Processing Portal, we are able to compare the performances of our models and show the logic of part of our methods, although many other choices are unverified. The results are based on all 46 validation dataset. The final mean dice indexes of the enhanced tumor (ET), whole tumor (WT) and tumor core (TC) are 0.751, 0.896 and 0.799.

The intensity normalization plays a major role in the success of the model. Using the original images with the lowest 1% and highest 99% percent pixels removed, the mean dice indexes using the same model with prediction stride size of 32 are 0.663, 0.841 and 0.748. A scrutiny of the dataset showed that the mean intensities of those with low scores often differ a lot with those in the training dataset, meaning that the variation in the mean intensities pose a challenge for the model. As in MR images, the contrast is much more important than the absolute pixel intensities, we want to normalize them without changing the contrast. Besides the intensity normalization, we also reduced the prediction stride size to 16 at the cost of increased prediction time. The mean dice indexes are 0.729, 0.887 and 0.796 without post processing. Although the dice indexes are pretty good, the Hausdorff distances are very large at 4.794, 24.124 and 10.120, meaning there are many false positives, which are verified with visual checking of the results. Therefore, we performed the connected region analysis on all training dataset and developed a post processing method to reduce false positives based on the hypothesis that all tumors are connected into relatively large regions with multiple dilation steps. This step reduced the Hausdorff distances to 4.755, 12.525 and 8.685, respectively and improved the dice index to the final result.

4 Discussion and Conclusions

In this paper we developed a patch-based 3D U-Net for brain tumor segmentation with intensity normalization in pre-processing and false positive reduction in post-processing. The preliminary results showed promises of this model as we are among the top performing teams in most evaluation metrics. However, there are some interesting findings and problems that need further investigation.

It is noted that the median metrics are significantly higher than the mean metrics. For example, the median dice indexes are 0.865, 0.925 and 0.880. It makes sense in that the theoretical maximum dice index is 1 and minimum dice index is 0. However, we noted that in several cases, the dice indexes are as low as 0 for ET and TC and 0.6 for WT. It is mostly due to the low sensitivity meaning that the model is not able to recognize the tumor regions. The possible reason for these failed regions is that their characteristics deviate a lot from the training dataset.

In the 3D U-Net model, we found that the batch norm layer was helpful in improving the model stability and performance. However, different with the canonical application of the batch norm layer, in which the batch statistics is used in training and the global statistics is used in deployment, it performed much better with batch statistics in deployment than global statistics. Since the batch size is 1, a per-channel normalization is actually performed by subtracting its own mean. One possible explanation could be that by doing such normalization, the model focuses on the differences of neighboring pixels in one channel and ignores the absolute values, which may help the segmentation process. However, a further investigation is needed to figure out the exact reason.

Compared with the patch-based model that only predicts the center pixel, when predicting the segmentation label maps for the full patch, different pixels are very likely to have different effective receptive field sizes due to the zero padding in the edge. We argue that a pixel should still be able to be predicted even based on partial receptive field, which, for the very edge pixel, corresponds to only half of the receptive field. Furthermore, the U-Net structure learns an optimized receptive field with multiple encoding and decoding paths and the connections in between and thus is superior than the multi-scale model. Furthermore, the significant overlap in prediction sliding windows can improve the prediction accuracy with more averages.

For post processing methods, we initially applied the conditional random forest (CRF) method but it often provided poor results unless the parameters in the model are very carefully chosen. It could also fail completely by predicting all background pixels with certain parameters. It is not surprising because the differences in intensities are often subtle and highly depend on the MR protocol. Therefore, we gave up this method and aimed to just reduce false positives in the post-processing step. However, one drawback of our method is that it may suffer from overfitting since the parameter selection is based on the feedback from the validation dataset.

In conclusion, we developed a patch-based 3D U-Net model for brain tumor segmentation by adapting the patch-based network structure which only predicts the center pixel to the 3D U-Net structure which predicts the full patch. Novel pre-processing and post-processing algorithms are also developed to improve the model performance. The code is available at https://github.com/xf4j/brats17.

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA,

Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)

2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. "Efficient Multi-Scale 3D CNN with fully connected CRF for Accurate Brain Lesion Segmentation", Medical Image Analysis 36, 61-78 (2017)
6. Ronneberger O, Fischer P, Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv:1505.04597 (2015)
7. Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation", arXiv:1606.06650 (2016)
8. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. "N4ITK: Improved N3 Bias Correction", IEEE Trans Med Imaging. 29(6), 1310-1320 (2010)

Deep Learning Neural Networks for Segmentation of brain Tumor in Multi Modal MRI

Naomi Fridman

Afeka Academic College of Engineering, Tel Aviv, Israel
Naomi.Fridman@s.afeka.ac.il

Abstract. In this project, I investigate the potential Convolutional Neural network(CNN) to learn and mark a brain tumor voxel on an MRI modality scan. Gliomas which is type of brain tumors, can be deadly. Accurate and fast tumor segmentation can improve medical treatment by reducing the lag between diagnosis and treatment, speed up surgery planning and follow up. Deep Convolutional Neural Network (CNN) has been proven lately to be efficient tool in object recognition, and they fit well medical imaging. The approach to the segmentation problem is to classify each pixel in the 2D plane along the z axis, by its nearby pixels. This classification is done with 4 layers convolutional layers and a hidden fully connected layer. I report preliminary, results obtained using BraTS 2017 Training dataset [8-11]. The conclusion of this work is that CNN's have potential to perform well on the segmentation task. More complex architecture, with combinations of few models, is needed to improve the results even further.

Keywords: Segmentation, BraTS challenge, Classification, Convolutional neural networks, Deep learning

1 Introduction

In this project, I investigate the potential CNN's to mark the Glioma tumor on an MRI modality scan. Gliomas are a type of brain tumor that can be deadly, and accurate segmentation of the tumor and its sub regions, is needed for treatment planning and surgery. The large amount of data of one MRI scan, meaning 4 models of 155 images of 240x240 pixels. To analyze and annotate this amount is time consuming task for human expert. Automatic tumor segmentation can improve medical treatment by providing an efficient and annotation of tumor and its sub-region location. The large amount of data of one MRI scan, meaning 4 models of 155 2D images of 240x240 pixels. To analyze and annotate this amount is time consuming task for human expert. Deep learning neural network, has proven lately as efficient tool for many tasks in image processing and analyzing. CNNs especially has become the state of the art methodology for object recognition. The nature of most medical images, that have high correlation among near voxel, make them good candidate for convolution sliding window.

2 Methods and Materials

2.1 MRI Dataset

The Training dataset of BraTS 2017 [7,8,9,10] comprises multimodal MRI scans of glioblastoma (GBM/HGG, 135) and lower grade glioma (LGG, 108). For each patient, there are four MRI scans as described: T1-, contrast enhanced T1- (T1c), T2- and T2-weighted FLAIR. The multimodal MRI scans, were acquired with different clinical protocols and various scanners from multiple (n=19) institutions, all images were pre-processes, meaning: co-registered to the same anatomical template, interpolated to the same resolution and skull-stripped.
All Images have been segmented and approved manually by experts. Annotations comprise the GD-enhancing tumor (ET - label 4), the peritumoral edema (ED - label 2), necrotic (NCR/NET - label 1) and non-enhancing tumor (label 3).

2.2 MRI Pre-processing

MRI data often contain artifacts produced by inhomogeneity in the magnetic field or small movements made by the patient during scan time, this often creates a bias across the resulting scans, which can affect the segmentation results particularly made by computer-based models. To correct that, I employed SimpleITK [5] N4 bias field correction filter [6] on all T1 and T1C images in the dataset. Since MRI intensities are expressed in arbitrary units and may vary between different machines, additional image pre-processing was made to standardize the voxel intensities, so each sequence was transformed to have zero mean and unit standard deviation.

2.3 Convolutional neural networks

Convolutional neural networks (CNN) proven to do well in image classification and segmentation tasks [3], especially when features are fully viable and not hidden one behind the other. Medical MRI images feature a high similarity and correlation in the intensities among neighboring voxels, so a local approach of patch classification, was chosen. It sims natural to select a 3D area around the classified voxel, but 3D convolutional networks need huge resources, so the approach of solving it slice by slice, is practical.

2.4 Model Architecture

I used sliding window convolution network with small kernel. The model has four-Convolutional layers, using. Pooling and dropout to improve accuracy and prevent over-fitting.

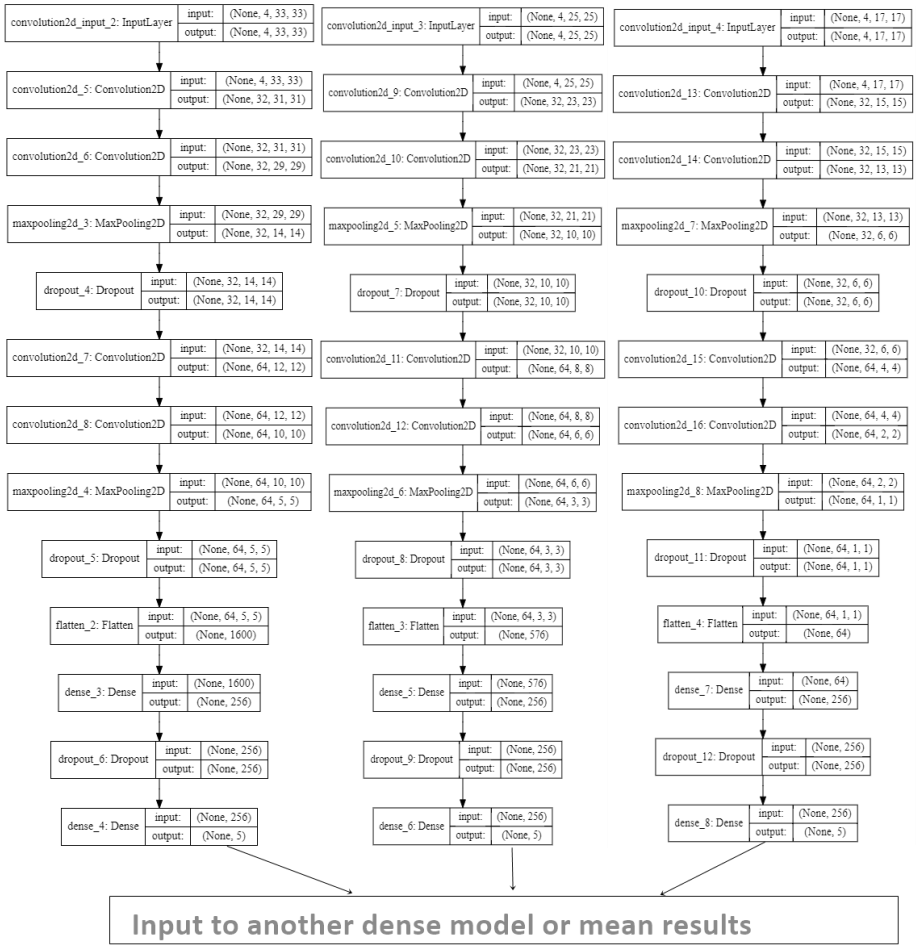


Fig. 1. Model architecture scheme. 3 Convolutional network models with different patch sizes: 33, 25, 17. Result is calculated as simple mean of all 3 outputs, or output of simple dense model on the 3 results.

2.5 Model training

Train data was split to train and test group. Each model was trained on 50,000 patches, selected randomly, evenly distributed according to voxel labels. Every few epochs, a new train group was created. Three models of different patch sizes (33,25 and 17), were trained, to result probabilities for each class. Mean of those probabilities, was the result. Model was created with Keras [11] and Theano [4].

3 Results

Preliminary results on train group, where promising Presented an example of a result of a model with Patch Size - 33 Train data 50,000 patches, 10,000 per class. First results, Fig3. Fig. 4. were calculated on Test data, which is relevant due to small and sparse data used to train the model. The second result shown, Fig 4. And Fig. 5. was produced on test data kept a side from the train set.

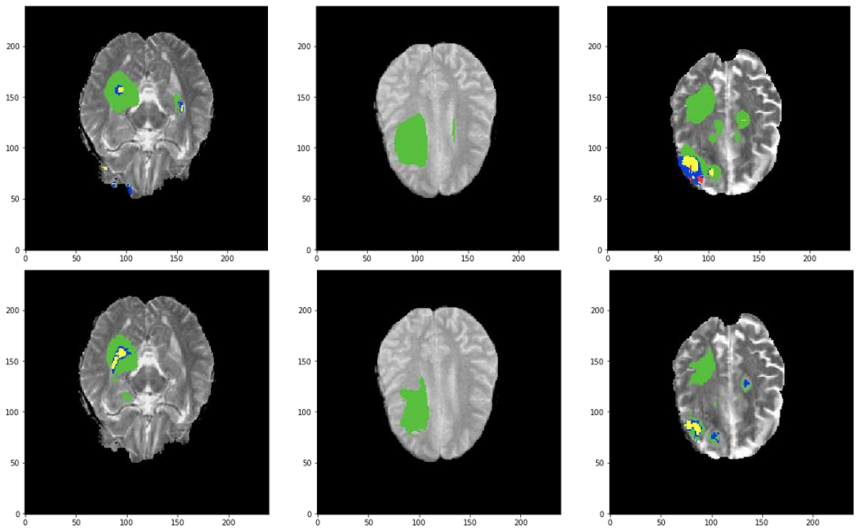
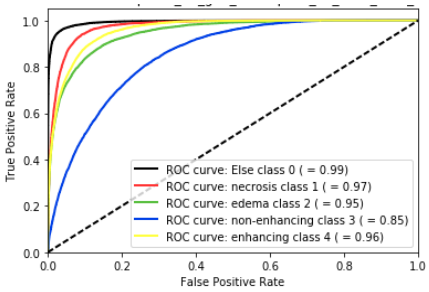


Fig. 2. Example of prediction on 3 randomly selected planes from train group.



CONFUSION MATRIX:

```
[[9455  32  436   59   18]
 [  27 8748  198  745  282]
 [ 755  185 7717 1096  247]
 [ 242 1705 1847 4298 1908]
 [  69  336  415 1228 7952]]
```

Fig. 3. ROC curve on 50,000 randomly selected. voxels, 10,000 per each class.

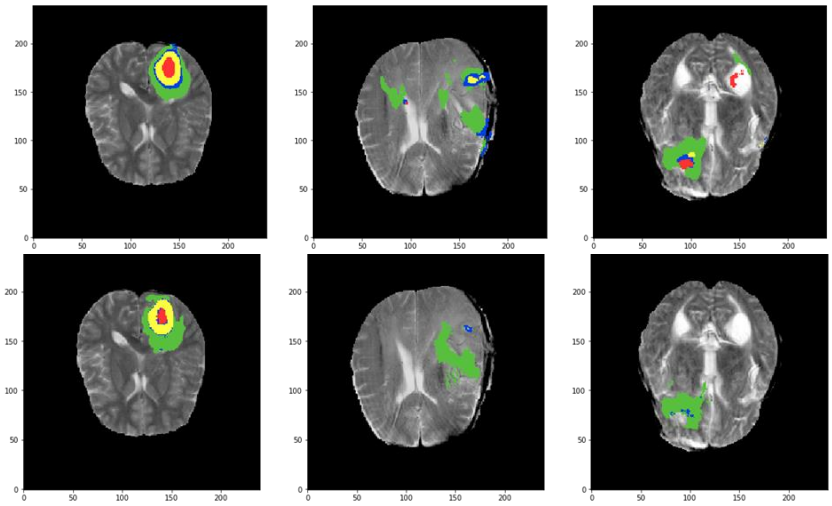


Fig. 4. Example of prediction on 3 randomly selected planes from test group.

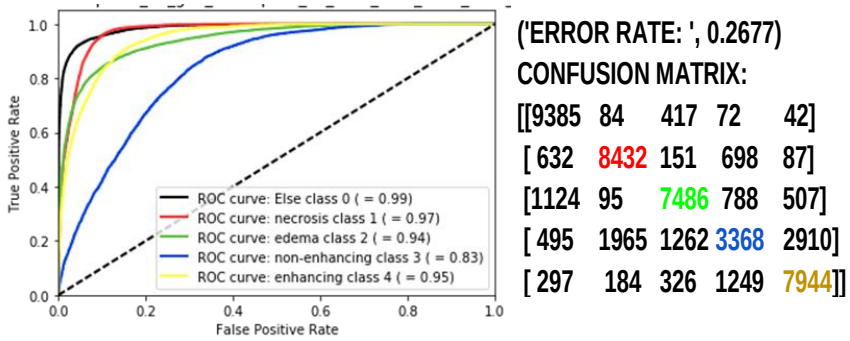


Fig. 5. ROC curve on 50,000 randomly selected voxels from test group, that was not part of the training data.

4 Discussion

CNN models yield good result in image classification and segmentation tasks, and while the model presented, yields promising results, in the medical context, there is no room for errors or false positives. To improve the model, we need to explore three main paths: one is in pre-processing the data and resampling techniques, the other is optional architecture.

An upsampling techniques should be applied and tested, it has the potential of improving the non-sufficient segmentation results on non-enhancing tumor (lable – 3). An architecture based on the one built by Havaei et al [2] which uses a two-pathway model to capture local and global features, can improve results as well.

5 Conclusion

Combining different technique, improves the result. In my model, results where improved when I mean results of different patch size models.

Sizes of tumors varies a lot. Large tumors are segmented better than small tumors. Segmentation results on small tumors, should be checked with different resampling methods.

References

1. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L.Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In ICLR, 2015.
2. Havaei, M. et. al, Brain Tumor Segmentation with Deep Neural Networks. arXiv preprint arXiv:1505.03540, 2015.Author, F., Author, S., Author, T.: Book title. 2nd edn. Publisher, Location (1999).
3. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234--241, 2015, available at arXiv:1505.04597 [cs.CV]
4. Theano Python library, https://github.com/Theano/Theano
5. N4BiasFieldCorrectionImageFilter Class Reference, https://itk.org/SimpleITKDoxygen/html/classitk_1_1simple_1_1N4BiasFieldCorrectionImageFilter.html
6. J.G. Sled, A.P. Zijdenbos and A.C. Evans. "A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in Data" IEEE Transactions on Medical Imaging, Vol 17, No 1. Feb 1998.
7. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of

the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
11. Keras: The Python Deep Learning library, https://keras.io/

Multimodal Brain Tumor Segmentation using 3D convolutional networks

Rodríguez Colmeiro Ramiro Germán[1,2,3] and Verrastro Claudio Abel[1,3]

[1] Universidad Tecnológica Nacional, Buenos Aires, Argentina
[2] Université de technologie de Troyes , Troyes, France
[3] Comisión Nacional de Energía Atómica, Buenos Aires, Argentina
Rodriguez.Colmeiro@Gmail.com

Abstract. Volume segmentation is one of the most time consuming and therefore error prone tasks in the field of medicine. The construction of a good segmentation requires cross-validation from highly trained professionals. In this context the creation of a method capable of automatically segment volumes is one of the most compelling challenges in the field of medical imaging. In order to address this problem we propose the use of deep convolutional networks. Using a 2 step procedure we first segment whole the tumor from a low resolution volume and then feed a second step which makes the fine tissue segmentation. The advantages of using DCN is that no interaction with the user is required, all parameters are self-learned and its accuracy can improve by feeding new examples to the trained network. The training dice-loss value surpass 0.8 and 0.7 for the coarse and fine segmentation networks respectively.

Keywords: Deep convolution network, medical image segmentation, volumetric semantic segmentation

1 Introduction

The manual segmentation of medical images is an exhausting and error prone procedure which is known to have a low inter-professional agreement [1]. The development of a fully automated and reliable segmentation procedure is one of the most challenging tasks in the field of medical imaging.
In the last years the deep convolutional networks (DCN) have been applied to image recognition tasks pushing the state of the art[2]. More recently DCN have been applied to semantic segmentation in 2-D images and 3-D volumes, achieving once more state of the art performances[3][4][5][6]. From the networks mentioned before the U-Net[7] is one of the most successful topologies and it was tested on several medical image segmentation tasks. As an extension to the the U-net topology the 3-D U-Net topology was derived[8]. This 3-D adaptation of the original topology replaces all 2-D convolutional layers with 3-D convolutional ones. Other types of nets, like the V-Net[9], where inspired in the U-Net and used in medical imaging segmentation problems.

It is straight forward to apply this 3-D segmentation topology in several medical imaging cases such as PET, TC, MRI and other voxelized medical image data. The BRATS challenge offers an unique opportunity to test this technique against other state of the art automatic segmentation methods [10][11][12].
In the following sections we describe the implementation of the 3-D U-Net for this particular task, starting with the data preprocessing, the objective segmented volume construction and full segmentation procedure description. Then we present the preliminary results from the first validation test followed by the discussion of this results. Finally the preliminary conclusions are drawn.

2 Methods

This section describes the operations of the automated segmentation process, starting with the description of the full segmentation process, followed by the input data and objective data treatment and construction of the coarse and fine segmentation networks and finally the training procedure.

2.1 Segmentation process

The segmentation was performed using a 2 step process. Both steps consists of a 3-D U-net which accomplish a different task. The whole process is described in figure 1, the process start with the input preprocessing which prepares the input volumes for the *Coarse Segmentation Network (CSN)*, its task is to detect the whole tumor within the brain volume, then by making use of the CSNs output the input volume is masked and cropped, obtaining a volume which consist of only tumorous tissue, this masked input is finally feed to the *Fine Segmentation Network (FSN)* wich is responsible of generating the desired output labels; *enhancing tumor* (ET), *tumor core* (TC) and *whole tumor* (WT) .

2.2 Data preprocessing

The input data consist of a set of 3D images acquired from 4 different MRI modalities; *flair, t1, t1ec and t2*. They where treated as different channels of the same volume. Each volume of 250x250x155 voxels was processed as a single volume with 4 channels, just as the RGB channels of a photograph. Voxels values were also normalized between 0 and 1 and coded with a 32 bit floating point number.
The objective data, which consists of a single volume with labels from 1 to 4, was decomposed into several objective volumes. For the CSN all labels where fussed in a single volume with value 1, which represented the whole tumor, in the case of the FSN all labels with different values where converted to different objective volumes, creating 4 binary objective volumes. In each case two new objective classes where added, one which consisted of non-tumorous brain tissue and an other which was *background*, their function was to help the convergence of the CNN.

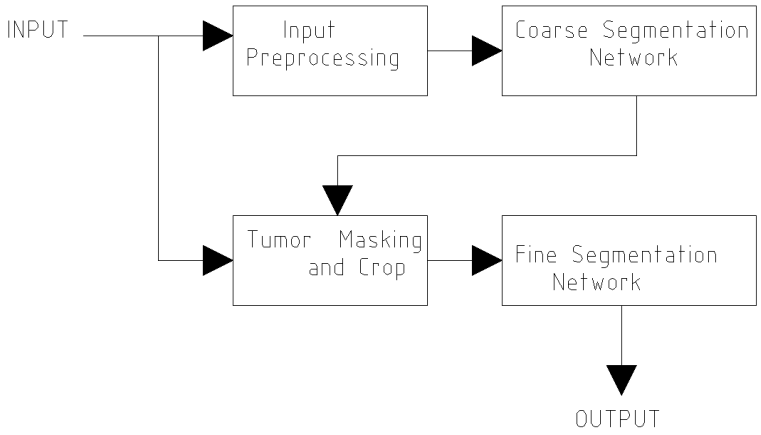


Fig. 1. Block diagram of the full segmentation process.

2.3 Deep Convolutional Network

The topology selected for the segmentation task was a 3-D version of the U-Net. This topology can be explained following figure 2. The topology consists of two parts.

The first one, follows the shape of a typical image recognition network, which performs convolutions over the input image then applies a non lineal function (a ReLU neuron in this case) and finally performs a pooling operation.

In this implementation the common *max-pooling* operation was replaced by a 2-voxel stride convolution operation as suggested in [9].

This first part decomposes the input volume in lower resolution volumes, dividing by two its dimensions in each step but duplicating the number of channels therefore maintaining the information.

The second part starts at the lower level (referred as base level) and is where the semantic segmentation starts. In the base level where instead of a 2-voxel stride convolution the network applies a 2-voxel stride up-convolution, which duplicates the image size and infers the values of the new voxels.

The following levels of this part not only receive information of the lower levels but also high-detailed information which is broadcasted from the levels of first part of the network which operate at their own precision (volume size) making possible detail rich segmentation of the input volume (see figure 2).

The broadcasted detail rich volume is concatenated at channel level with the volume coming from the lower level, i.e. if the volume at the current level has a size of 64x64x16 voxels and 16 channels, then the concatenated volume will have a size of 64x64x16 voxels and 32 channels.

The final segmentation is done by a convolutional layer followed by a *softmax* operation among the objective classes.

Both networks where constructed with the same number of levels, 4 levels in the downward path, 4 levels in the upward path and a base level.

Each level consists of a convolution part, for feature extraction and a down or up sampling procedure. The feature extraction part consists of two layers with a 3-D convolution followed by a ReLU neuron.

The network input has four channels (oner per input volume type) which is decomposed into 8 channels in the fist level, and ending in 64 channels at base level. The convolutional filters had a size of 3x3x3 voxels and the final segmentation filters had a size of 2x2x2 voxels.
In the following sub-sections some particularities of each net are explained.

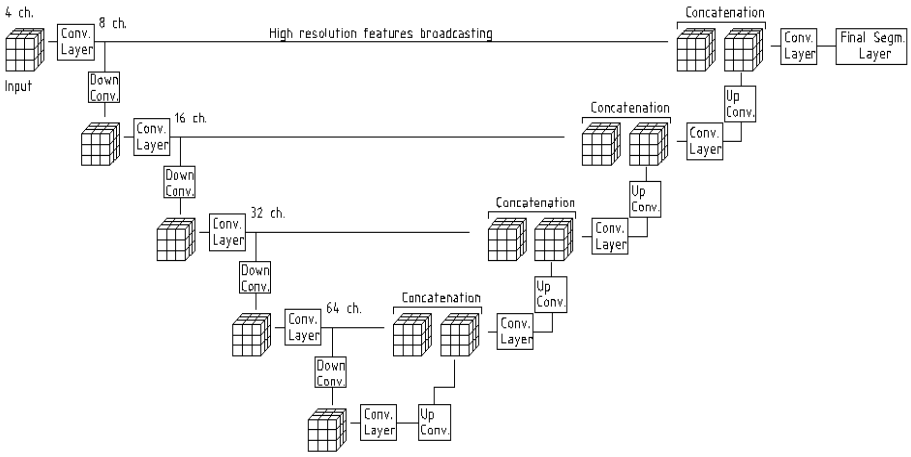


Fig. 2. 3-D U-Net topology.

Coarse Segmentation Network This network takes as input a volume of size 128x128x64 voxels with 4 channels , this volume is generated by down-sampling the input. This down-sampling was necessary due to the limited GPU memory (the calculations where done using a NVIDIA Kepler K20 GPU using the TensorFlow framework[13]). The output of the network is also a volume of size 128x128x64 voxels.

Fine Segmentation Network This network receives a 4 channel 96x96x96 voxel volume. When the network is in production its input is built using the output of the CSN and the original input. The output of the CSN is up-sampled to the original volume size and then used as a binary mask for the original input. Once the input volume is masked it is cropped to a 96x96x96 voxel volume. When the network is being trained, instead of the CSN output, the ground truth is used to mask the volume.

2.4 Training

As objective function the dice loss running over all outputs was chosen. The objective function can be seen in equation 1, where N_c is the number of objective classes, N_v the number of voxels in the volume, g_i are the voxels of the ground truth and p_i the values of the softmaxed output of the network.
The training of both networks was done separately using the ADAM optimizer [14] and with learning rate decay following $\alpha = \frac{\alpha_{ini}}{\sqrt{n_{step}}}$, where n_{step} is the step number. Batch normalization[15] was applied at each convolutional layer.

$$D = \sum_{c=0}^{N_c} \frac{2 \sum_i^{N_v} p_i g_i}{\sum_i^{N_v} p_i^2 + \sum_i^{N_v} g_i^2} \tag{1}$$

2.5 Data Enhancement

Since the training of DNN requires tenths of thousands of samples to converge, new input samples where generated from the original data. This new samples where created by applying two transformations: rotation and conform deformation.
Both transformation where performed by operating on an uniform mesh which extends over the whole input volume. The first operation applied to it is a random rotation in an interval of $-\frac{\pi}{4}$ to $\frac{\pi}{4}$ in the direction of a random versor. Then the deformation procedure is applied to the rotated mesh. The deformation consists of a sinusoidal distortion in every axis, this sinusoidal distortion has a random period between 10 and 25 pixels and a peak between 2 and 4 pixels.
After the fully transformed mesh is created, it is used to map the coordinates of all voxels in the input volumes and segmentation labels to their new position in the transformed sample. The transformed sample mapping is created using a trilinear interpolation.

3 Results

This section summarises the results of the validation set. The CSN was trained with 58000 examples achieving a dice loss over $D_c = 0.82$ while the FSN was trained with 67000 examples and achieved a dice loss over $D_f = 0.7$.
The table 1 shows the results of the segmentation performed over the BRATS17 validation set. The global score is shown as *mean* and the detail by institution is depicted in the first three rows.

4 Discussion

The validation results of this first approach showed a good correlation between the training accuracy and the validation. The segmentation procedure was able to perform as expected in the all of the dataset sources even when no samples

Table 1. Validation Results

	Dice ET	Dice WT	Dice TC	Hausdorff95 ET	Hausdorff95 WT	Hausdorff95 TC
CBICA	0.514	0.797	0.521	10.5	10.3	13.4
TCIA	0.412	0.827	0.611	9.47	6.56	11.0
UAB	0.545	0.869	0.508	11.9	6.13	12.4
Mean	0.469	0.822	0.566	13.8	9.56	14.7

of the UAB dataset where present in the training batch.
The lower performance in the *tumor core* and *enhancing tumor* was expected since the FOV of the FSN was smaller than the larger tumor, this can be easily solved with more GPU memory. Also the fine segmentation was more sensible to changes in the normalization of the input data.
The normalization of the imput samples showed to be critical during the validation phase. All samples from the new source where normalized to the best performing sample of the validation dataset. This normalization was performed by matching histograms.
The fact that the network was able to perform segmentation on unknown samples showed that the data enhancement procedure was effective and successfully expanded the original dataset more than 100 times. Due to the long training times of both networks (more than 5 days) only the described topology was tested.

5 Conclusions

The proposed automatic segmentation procedure based on a 3-D U-net topology showed to be effective when applied to the BRATS17 dataset. The networks where able to perform segmentation without any other information than the given examples and no specific knowledge of the problem was used to create the topology.
Even when the data enhancement techniques where simple and not problem specific they where capable of produce new valid data. Generating a problem specific data enhancement technique which focuses on the transformation of the tumor areas might speed up the process and help generalization.
The data normalization probed to be one of the critical points in the network validation. The normalization used during the training phase (only data scaling) was not accurate when samples from other sources where tested, like the ones from UAB. To fix this a more complex normalization procedure was used in the validation dataset which matches histogram of every volume to those of a correctly segmented sample. By making use of this new normalization the segmentation procedure was able to improve its score by a 50% without having to retrain any of the networks. Nevertheless a new training procedure with histogram-matched samples is being trained and is expected to achieve better results.

Also further refinement of the topology can be done, more problem-specific configuration and fine tuning of level and layer sizes might lead to networks that converge faster and better.
Given that the networks topology applied is more or less a default configuration, the results are not state of the art but encouraging to further study this segmentation technique which showed great potential and is applicable within the field of medical imaging as well as other computer vision problems.

References

1. Brent Foster, Ulas Bagci, Awais Mansoor, Ziyue Xu, and Daniel J Mollura. A review on segmentation of positron emission tomography images. *Computers in biology and medicine*, 50:76–96, 2014.
2. Jrgen Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.
3. Yun Liu, Krishna Gadepalli, Mohammad Norouzi, George E Dahl, Timo Kohlberger, Aleksey Boyko, Subhashini Venugopalan, Aleksei Timofeev, Philip Q Nelson, Greg S Corrado, et al. Detecting cancer metastases on gigapixel pathology images. *arXiv preprint arXiv:1703.02442*, 2017.
4. Marios Anthimopoulos, Stergios Christodoulidis, Lukas Ebner, Andreas Christe, and Stavroula Mougiakakou. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE transactions on medical imaging*, 35(5):1207–1216, 2016.
5. Xiaohong W Gao and Rui Hui. A deep learning based approach to classification of ct brain images. 2016.
6. Holger R Roth, Christopher T Lee, Hoo-Chang Shin, Ari Seff, Lauren Kim, Jianhua Yao, Le Lu, and Ronald M Summers. Anatomy-specific classification of medical images using deep convolutional nets. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 101–104. IEEE, 2015.
7. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
8. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432. Springer, 2016.
9. Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016.
10. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby JS., Freymann JB., Farahani K., and Davatzikos C. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 2017.
11. Bakas S., Sotiras A., Bilello M., Rozycki M., Kirby JS., Freymann JB., Farahani K., and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *Nature Scientific Data*, 2017.

12. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby JS., Freymann JB., Farahani K., and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *Nature Scientific Data*, 2017.
13. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
14. Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
15. Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

An Automatic Computerized Method for Segmentation Labels of Various Glioma Sub-regions on Supporting Brain Tumor Detection in Radiologic Imaging

Peifang Guo

Montreal, QC, Canada
Corresponding author. *E-mail address*: peif.guo@gmail.com
Received July 22, 2017

Abstract. In this study, to support the detection of brain tumors clinically, an automated method is proposed for the segmentation labels of Glioma sub-regions, which includes the steps of the calculation of the average values of image sub-region colors, the labeled image sub-regions mapped into color spaces and segmentation of labeled sub-regions. In the validation, the proposed method is implemented in the samples of image scans from the real-time data of the BraTS'17.

Keywords: Labels of Glioma sub-regions, Glioma sub-region segmentation, brain tumors, image mapping into colors, clinical decision support, radiologic imaging.

1 Introduction

An automatic method for the segmentation labels of the various Glioma sub-regions with a biomarker model for predicting disease patterns of brain tumors in radiologic imaging is important in the brain tumor diagnosis clinically (Bakas, et al., 2017A; Guo, 2017A; Bakas, et al., 2017B; Bakas, et al., 2017C; Menze, et al., 2015; Bauer, et al., 2013; Jack et al., 2010). This study seeks to address the challenge for the segmentation labels of the Glioma sub-regions and the Glioma sub-region segmentation in the samples of radiologic imaging on supporting the detection of brain tumors in clinical.

2 Method

In this study, the proposed method mainly includes the steps of the calculation of the average values of the sub-region colors, the output of labeled image sub-regions, the labeled image sub-regions mapped into color spaces, and segmentation of labeled sub-regions with colors. In the proposed approach, based on the average values of the sub-region colors, the minimum distance classifier is employed to measure an image

pixel between that image pixel and each image sub-region. Then the shortest distance would indicate that the pixel most closely matches that color of the sub-region in the classification.

3 Results

In the validation, the proposed method is implemented in the samples of image scans on the real-time data of the 2017 brain tumor segmentation challenge (BraTS'17). In the challenge of the BraTS'17, the image sub-regions need to be evaluated are the "enhancing tumor" (ET), the "tumor core" (TC) and the "whole tumor" (WT).

Fig. 1 shows the preliminary results from an example of experiments in the samples of image scans on the real-time data of the BraTS'17. The top row of Fig. 1 are the image patches, from the left to the right: the WT visible in FLAIR (A), the TC visible in T2 (B), the ET structures visible in T1c, surrounding the cystic/necrotic components of the core (C). The second row of Fig. 1 is the output of labeled image sub-regions, while the third row of Fig. 1 shows the labeled sub-regions when mapped into the color space. The output of segmentation labels of the different Glioma sub-regions are illustrated in the last row of Fig. 1, which shows three different sub-regions involving, from the left to the right, the WT, TC and ET.

After going through the recursive computation, Fig. 2 further shows the resulting segmented Glioma sub-region labels for the two different sub-regions, the TC (upper row) and ET (lower row), where Fig. 2(a) is the output of segmentation labels of the sub-regions (see the last row of Fig. 1); Fig. 2(b) is the output of labeled image sub-regions of the TC and ET, and Fig. 2(d) present the results of the sub-region segmentation involving the TC and ET.

On the data of the BraTS'17, the ground truth labels was annotated by experts for the various glioma sub-regions. Figs 3-4 illustrate an example of the experiments for the segmentation of the ground truth labels on the data of the BraTS'17. Figs 3-4 (a) are the ground truth labels which have been manually-revised by expert board-certified neuroradiologists. Figs. 3-4 (b) display the output of labeled images of the ground truth labels (with the edema (yellow), the non-enhancing solid core (red), the necrotic/cystic core (green) and enhancing core (blue)). Fig. 3 (c) is the outline of the ground truth labels. Fig. 3(d)-(e) and Fig. 4(c)-(g) show individually the segmented ground truth labels of the edema (yellow), the non-enhancing solid core (red), the necrotic/cystic core (green) and enhancing core(blue)).

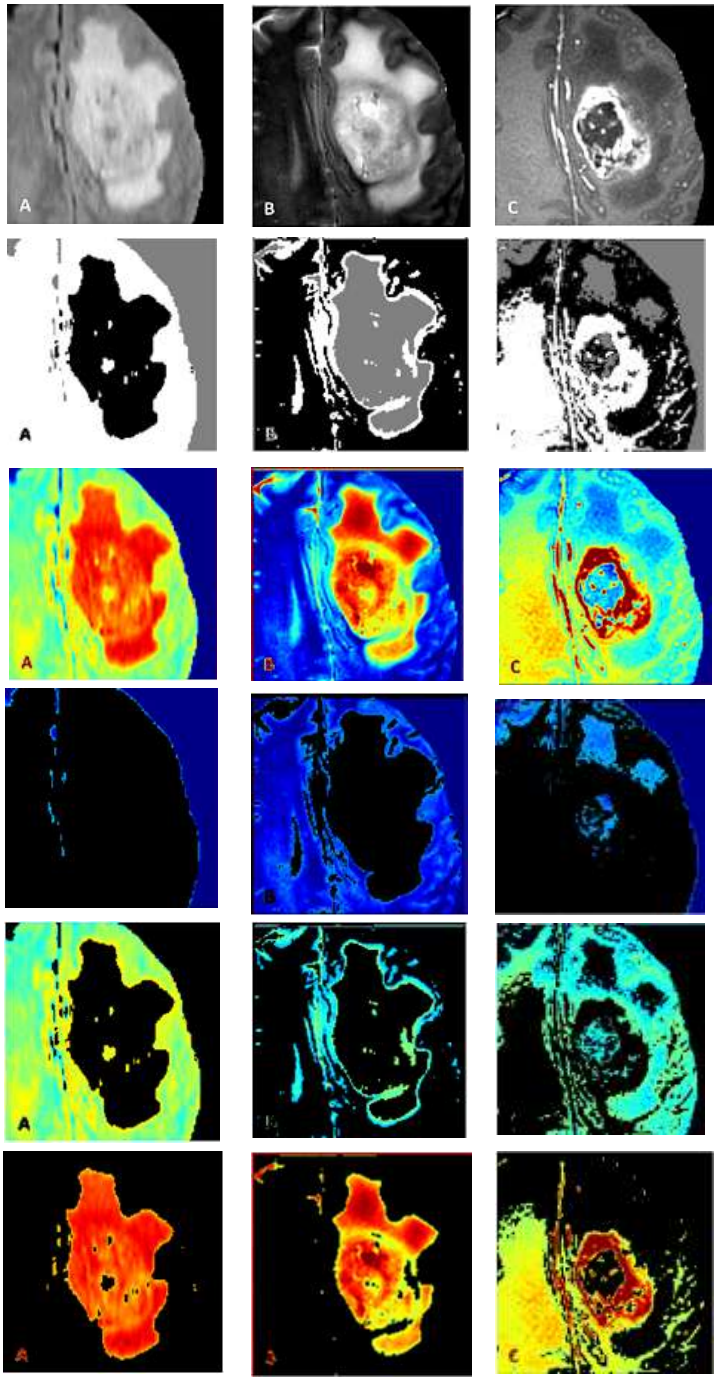


Fig. 1. The implementation results from an example of experiments on the real-time data of the BraTS'17.

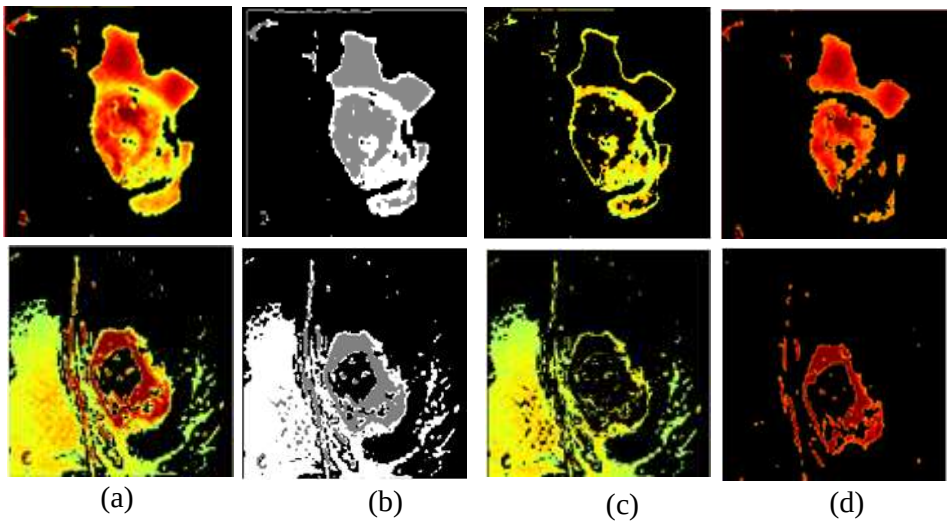


Fig. 2. Results of segmented sub-region labels, the TC (upper row) and ET (lower row) after the recursive computation, from an example of experiments on the real-time data of the BraTS'17.

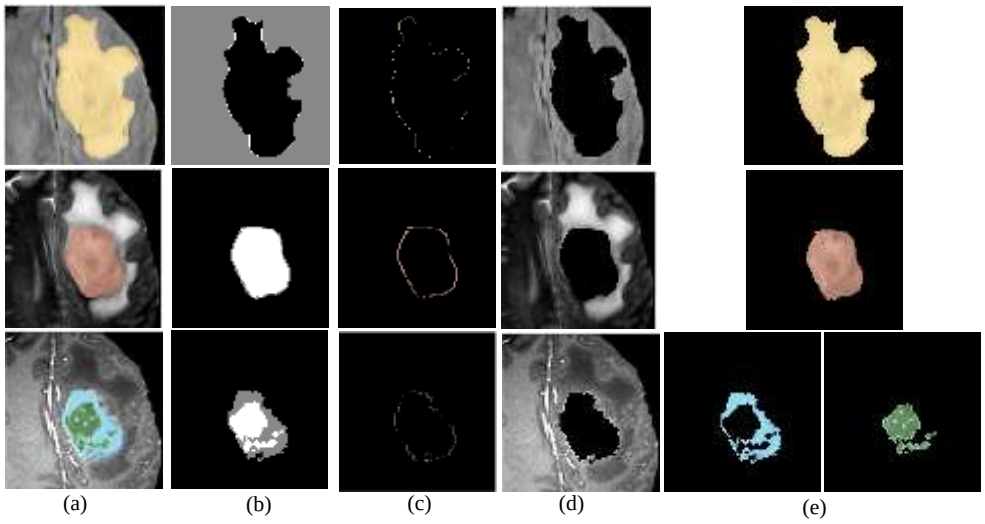
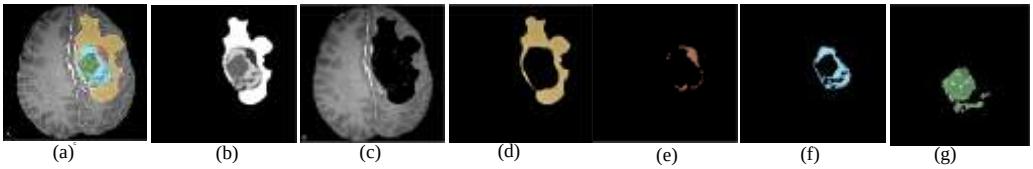


Fig. 3. An example of the experiments for the segmentation of the ground truth labels on the BraTS'17.



(a) (b) (c) (d) (e) (f) (g)

Fig. 4. An example of the experiments for the segmentation of the final labels of the tumor structures (the ground truth labels) on the BraTS'17.

4 Conclusion and Future Work

The future work will be dedicated on the analysis of the Glioma sub-regions qualitatively, using the techniques of iterative contour refinement and morphological reconstruction from the previous work (Guo, 2017B), for example, to remove the non-TC and non-ET (see Fig. 3(d)) with the ground truth labels (as the image masks). In addition, the histogram-based techniques from the previous work (Guo et al., 2014) would be useful to compute image intensities of the Glioma sub-regions quantitatively, as compared with the ground truth provided from the BraTS'17.

References

Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", *Nature Scientific Data*, (2017A) [In Press]

Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017B. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q

Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017C. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M. "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, pp. R97–R129, 2013.

Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. The nultimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* 34, pp. 1993–2024, 2015.

Guo, P. A tissue-based biomarker model for predicting disease patterns," *J. Knowledge-Based Sys.*, vol. 276, pp. 160-169, 2017A.

Guo, P. Brain tissue classification method for clinical decision-support systems. *J. Eng. Appli. Artific. Intell.*, vol. 64, pp. 232-241, 2017B.

Guo, P. and Bhattacharya, P. An evolutionary framework for detecting protein conformation defects," *J. Information Sciences*, vol. 276, pp. 332-342, 2014

Jack Jr., C. R., Knopman, D. S., Jagust, W.J ., Shaw, L. M., Aisen, P. S., Weiner, M.W., Petersen, R. C., Trojanowski, J. Q. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol*. vol. 9, pp. 119–128, 2010.

Automated Brain Tumor Segmentation Using A 3D Deep Detection-Classification Model

Yan Hu [1] and Yong Xia [1, 2*]

[1] Shaanxi Key Lab of Speech & Image Information Processing (SAIIP), School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, PR China
[2] Centre for Multidisciplinary Convergence Computing (CMCC), School of Computer Science and Technology, Northwestern Polytechnical University, Xi'an 710072, PR China

Abstract. Brain tumor segmentation plays a pivotal role in clinical practice and research settings. In this paper, we propose a 3D deep detection-classification model for automated segmentation of brain tumor and intra-tumor structures, including necrosis, edema, non-enhancing and enhancing tumor. This algorithm has two planks: a cascaded two-level U-Net for detecting the outline of tumors and a patch-based deep convolutional network for classifying tumor voxels to one of those four structures. We have evaluated the proposed algorithm on the training dataset provided by the Brain Tumor Segmentation 2017 (BraTS 2017) Challenge[1-4] and achieved a Dice similarity index of 0.72, 0.77 and 0.79 for the segmentation of active tumor volume, core tumor volume and entire tumor volume, respectively. Our pilot results show that the proposed algorithm has competitive performance in automated tumor segmentation.

Keywords: Brain tumor segmentation, deep learning, fully convolutional network (FCN), U-Nets, magnetic resonance imaging (MRI)

1 Introduction

Gliomas are the most common primary brain malignancies. At present, over 130 different types of 'high grade' and 'low grade' brain tumors are known, and it is reported about 560 teenagers under age 19 are diagnosed as brain tumor patients every year only in UK. Despite considerable advances in gliomas research, diagnosis, treatment plan and follow-up evaluation of brain tumors, in which accurate delineation of tumor volumes is an essential step, remain major challenges in related clinical practices. Magnetic resonance imaging (MRI) can provide high spatial resolution of anatomical details and unique contrast between soft tissues, and hence is suitable for this task. Currently, brain tumors are usually delineated manually by medical professionals, which requires a high degree of skill and concentration, and is time-consuming, expensive, and prone to operator bias. Computer-aided brain tumor segmentation using MRI would overcome these issues and provide medical professionals an unprecedented tool for efficient and reliable diagnosis, treatment and prognosis of brain tumors. Therefore, there has been considerable research directed to automated brain tumor segmentation in MRI images.

A number of brain tumor segmentation methods have been proposed in the literature. These methods can be roughly grouped into two categories: generative models utilizing detailed prior information about the appearance and spatial distribution of the different types of tissue regions and discriminative models directly learning the characteristic of different tissue regions from manually annotated training images. Recently, deep learning techniques, such as the convolutional neural network (CNN) and fully convolutional network (FCN), have been adopted to solve this medical image segmentation problem. Zhao et al.[5] jointly used a FCN and the conditional random fields (CRF). Konstantinos et al.[6] designed a dual pathway 3D CNN called DeepMedic, which contains 11 learnable layers. Pereira et al.[7] used two different CNNs for the segmentation of high grade gliomas (HGG) and low grade gliomas (LGG), respectively.

In this paper, we propose a 3D deep detection-classification model for automated segmentation of brain tumor and intra-tumor structures, including necrosis, edema, non-enhancing and enhancing tumor. At the detection stage, we designed a cascaded U-Nets to identify the entire tumor volume; and at the classification stage, we used a patch-based CNN to assign each tumor voxel to an intra-tumor structures. We have evaluated the proposed algorithm on the training dataset provided by the Brain Tumor Segmentation 2017 (BraTS 2017) Challenge and achieved a Dice similarity index of 0.72, 0.77 and 0.79 for the segmentation of active tumor volume, core tumor volume and entire tumor volume, respectively.

2 Dataset

This study was performed on the BraTS 2017 training dataset, which includes multimode brain MRI scans of 285 subjects. For each subject, there are four MRI sequences, including the T1-weighted (T1), T1 with gadolinium enhancing contrast (T1c), T2-weighted (T2) and FLAIR. All studies have been segmented manually, by one to four raters, and their annotations were approved by experienced neuro-radiologists. The segmentation ground truth identifies four types of intra-tumoral structures: necrosis, edema, non-enhancing and enhancing tumor.

3 Method

The brain tumor segmentation algorithm consists of two main procedures, i.e. tumor detection and tumor voxel classification. Tumor detection aims to locate the entire tumor volume and extract global spatial features, and tumor voxel classification targets at accurately delineating the tumor into four intra-tumor structures.

3.1 Tumor Detection

As shown in Fig. 1, tumor detection consists of three major steps. First, since there are four 3D MRI sequences for each subject, we take them as four independent input.

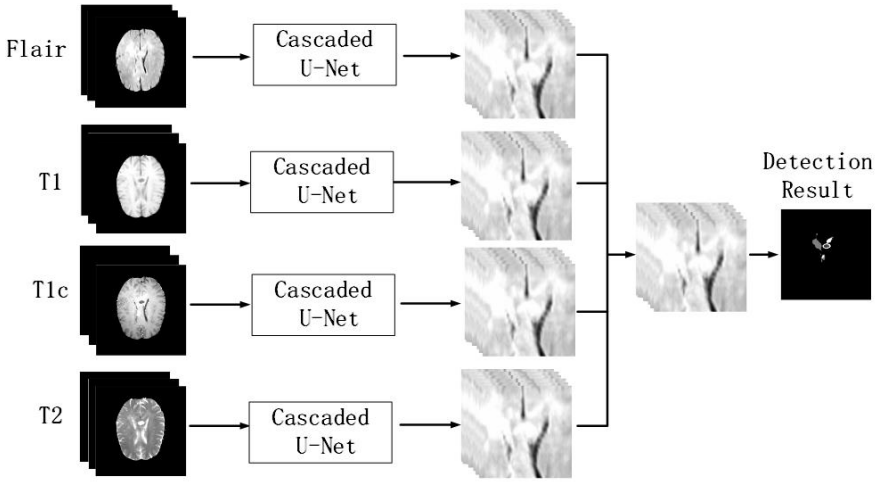


Fig 1. Architecture of the Tumor Detection Module

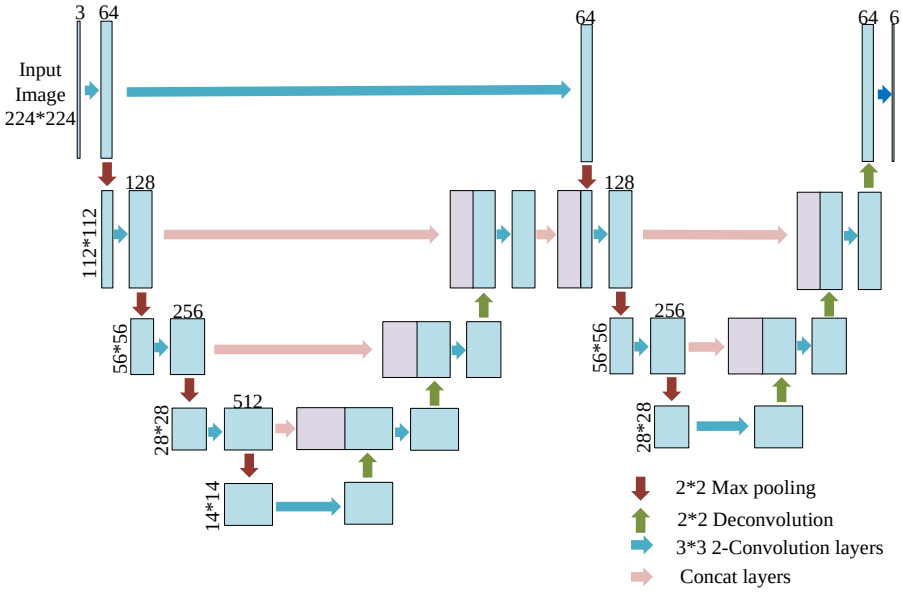


Fig 2. Architecture of the Cascaded U-Net

When processing the i_{th} slice, we consider the $i-1_{th}$, i_{th} and $i+1_{th}$ slices together to make fully use of the 3D information. Hence, we regard three adjacent slices of one modality as the three color channels of an image. Second, each modality of the MRI data is used to train a cascade two-level U-Nets[8] for feature extraction in an "end-to-end" way, whose architecture is displayed in Fig. 2. Third, four groups of feature maps are concatenated together as the input of a CNN with two convolutional layers for tumor volume detection.

Finally, the output of this model is a label map, which indicates the volume of brain tumor, and a group of feature maps with rich spatial information.

3.2 Classification

Based on the detected tumor boundary, we further classify each brain voxel to one of four intra-tumor structures. For each tumor voxel, we let a 45×45 window center on it. The image patch inside this window is combined with the feature maps obtained in the detection step to form an input to a pre-trained VGG-16 network[9]. The output of this CNN gives the class label of the corresponding tumor voxel. The architecture of this voxel classification module is shown in Fig. 3.

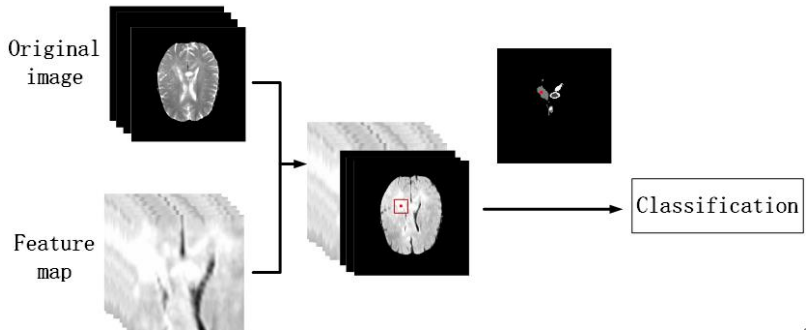


Fig.3 Architecture of Tumor Voxel Classification Module

4 Experiments and Results

We evaluated the proposed brain tumor segmentation algorithm on the BRATS 2017 training database with five-fold cross validation. Following the request of the challenge[10], four intra-tumor structures have been grouped into three mutually inclusive tumor regions: the active (enhancing tumor), core (necrosis + non-enhancing tumor + enhancing tumor) and the whole tumor (all four structures combined) for evaluation. The performance of segmenting each type of tumor regions was evaluated quantitatively by using the sensitivity and Dice similarity coefficient (DSC), which is defined as the ratio between the intersection and union of the obtained volume S_1 and ground truth volume S_2

$$DC(S_1, S_2) = \left(2|S_1 \cap S_2|\right) / \left(|S_1| + |S_2|\right) \quad (1)$$

The value of DSC ranges from 0 to 1, with a higher value representing a more accurate segmentation result. The segmentation performance measured in terms of sensitivity and DSC was given in Table 1. It reveals that the proposed algorithm is able to produce relatively accurate brain tumor segmentation.

Table 1. Result of the proposed algorithm on Brats 2017 Training Database

DSC			Sensitivity		
Whole	Core	Active	Whole	Core	Active
0.79	0.77	0.72	0.79	0.76	0.77

5 Discussion on Using Multi-Mode MRI Scans

To demonstrate the contribution of each imaging modality to solving this problem, we performed empirical evaluation and concluded that the proposed algorithm perform best when T1, T1c, T2 and Flair are used in detection and T1, T1c and T2 are used in classification.

6 Conclusion

In this paper, we propose a 3D deep detection-classification model for automated segmentation of brain tumor and intra-tumor structures using four modalities of MRI scans. The evaluation results on the BRATS 2017 training database indicate that the proposed algorithm is able to produce relatively accurate brain tumor segmentation.

References

1. BraTS 2017 Image Repository. http://www.med.upenn.edu/sbia/brats2017/data.html.
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
5. Zhao, XM., Wu, YH., Song, GD., Li, ZY., Fan, Y., Zhang, YZ.: "Brain tumor segmentation using a fully convolutional neural network with conditional random fields," in proc of BRATS-MICCAI (2016).
6. Konstantinos, K., Enzo, F., Sarah, P., Cristian, L., Aditya, Nori., Antonio, Criminisi., Daniel, Rueckert., Glocker, B.: "DeepMedic on Brain Tumor Segmentation," in MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS), 2016, pp. 18-22.

7. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240-1251(2016).
8. Ronneberger, O., Fischer, P., Brox,T.: "U-Net: Convolutional Networks for Biomedical Image Segmentation," in medical image computing and computer assisted intervention (2015).
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556 (2014).
10. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).

Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge

Fabian Isensee[1], Philipp Kickingereder[2], Wolfgang Wick[3], Martin Bendszus[2], and Klaus H. Maier-Hein[1]

[1] Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2] Department of Neuroradiology, University of Heidelberg Medical Center, Heidelberg, Germany
[3] Neurology Clinic, University of Heidelberg Medical Center, Heidelberg, Germany

Abstract. Quantitative analysis of brain tumors is critical for clinical decision making. While manual segmentation is tedious, time consuming and subjective, this task is at the same time very challenging to solve for automatic segmentation methods. In this paper we present our most recent effort on developing a robust segmentation algorithm in the form of a convolutional neural network. Our network architecture was inspired by the U-Net and has been carefully modified to maximize brain tumor segmentation performance. We use a dice loss function to cope with class imbalances and use extensive data augmentation to successfully prevent overfitting. Our method beats the current state of the art on BraTS 2015 and shows promising results on the BraTS 2017 validation set (dice scores of 0.896, 0.797 and 0.732 for whole tumor, tumor core and enhancing tumor, respectively). We furthermore take part in the survival prediction subchallenge by training an ensemble of a random forest regressor and a multilayer perceptron ensemble on shape features describing the tumor subregions. Our ensemble achieves 335.08 root mean squared error (232.76 mean absolute error) in a five fold cross-validation over the 163 training cases.

Keywords: CNN, Brain Tumor, Glioblastoma, Deep Learning

1 Introduction

Quantitative assessment of brain tumors provides valuable information and therefore constitutes an essential part of diagnostic procedures. Automatic segmentation is attractive in this context, as it allows for faster, more objective and potentially more accurate description of relevant tumor parameters, such as the volume of its subregions. Due to the irregular nature of tumors, however, the development of algorithms capable of automatic segmentation remains challenging.

The brain tumor segmentation challenge (BraTS) [1] aims at encouraging the development of state of the art methods for tumor segmentation by providing a large dataset of annotated low grade gliomas (LGG) and high grade glioblastomas (HGG). Unlike the previous years, the BraTS 2017 training dataset, which consists of 210 HGG and 75 LGG cases, was annotated manually by one to four raters and all segmentations were approved by expert raters [2–4]. For each patient a T1 weighted, a post-contrast T1-weighted, a T2-weighted and a FLAIR MRI was provided. The MRI originate from 19 institutions and were acquired with different protocols, magnetic field strengths and MRI scanners. Each tumor was segmented into edema, necrosis and non-enhancing tumor and active/enhancing tumor. The segmentation performance of participating algorithms is measured based on the DICE coefficient, sensitivity, specificity and Hausdorff distance. Additionally to the segmentation challenge, BraTS 2017 also required participants to develop an algorithm for survival prediction. For this purpose the survival (in days) of 163 training cases was provided as well.

Inspired by the recent success of convolutional neural networks, an increasing number of deep learning based automatic segmentation algorithms have been proposed. Havaei et al. [5] use a multi-scale architecture by combining features from pathways with different filter sizes. They furthermore improve their results by cascading their models. Kamnitsas et al. [6] proposed a fully connected multi-scale CNN that was among the first to employ 3D convolutions. It comprises a high resolution and a low resolution pathway that are recombined to form the final segmentation output. For their submission to the brain tumor segmentation challenge in 2016 [7], they enhanced their architecture through the addition of residual connections for improved segmentation performance. They addressed the class imbalance problem through a sophisticated training data sampling strategy. Kayalibay et al. [8] developed very successful adaptation of the popular U-Net architecture [9] and achieved state of the art results for the BraTS 2015 dataset. Notably, they employed a Jaccard loss function that intrinsically handles class imbalances. They make use of the large receptive field of their architecture to process entire patients at once, at the cost of being able to train with only one patient per batch. Here we propose our contribution to the BraTS 2017 challenge that is also based on the popular U-Net architecture [9]. Our network possesses twice as many filters than [8] while being trained with a slightly smaller input patch size and a larger batch size. We furthermore employ a multiclass adaptation of the dice loss [10] and make extensive use of data augmentation.

Image based tumor phenotyping and derived clinically relevant parameters such as predicted survival is typically done by means of radiomics. Intensity, shape and texture features are thereby computed from segmentation masks of the tumor subregions and subsequently used to train a machine learning algorithm. These features may also be complemented by other measures handcrafted to the problem at hand, such as the distance of the tumor to the ventricles [11]. Although our main focus was put on the segmentation part of the challenge, we developed a simple radiomics based approach combined with a random forest regressor and a multilayer perceptron ensemble for survival prediction.

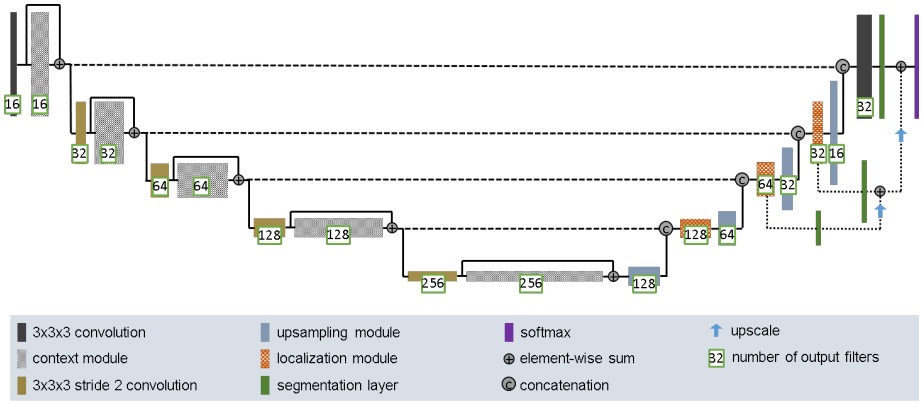


Fig. 1. Network architecture. Our architecture is inspired by the UNet [9]. The context pathway (left) aggregates high level information that is subsequently localized precisely in the localization pathway (right). Inspired by [8] we inject gradient signals deep into the network through deep supervision.

2 Methods

2.1 Segmentation

Data preprocessing With MRI intensity values being non standardized, normalization is critical to allow for data from different institutes, scanners and acquired with varying protocols to be processed by one single algorithm. This is particularly true for neural networks where imaging modalities are typically treated as color channels. Here we need to ensure that the value ranges match not only between patients but between the modalities as well in order to avoid initial biases of the network. We found the following simple workflow to work surprisingly well. First, we normalize each modality of each patient independently by subtracting the mean and dividing by the standard deviation of the brain region. We then clip the resulting images at $[-5, 5]$ to remove outliers and subsequently rescale to $[0, 1]$, with the non-brain region being set to 0.

Network architecture Our network is inspired by the U-Net architecture [9]. We designed the network to process large 3D input blocks of 128x128x128 voxels. In contrast to many previous approaches who manually combined different input resolutions or pathways with varying filter sizes, the U-Net based approach allows the network to intrinsically recombine different scales throughout the entire network. Just like the U-Net, our architecture comprises a context aggregation pathway that encodes increasingly abstract representations of the input as we progress deeper into the network, followed by a localization pathway that recombines these representations with shallower features to precisely localize the structures of interest. We refer to the vertical depth (the depth in the U shape)

as level, with higher levels being lower spatial resolution, but higher dimensional feature representations. The activations in the context pathway are computed by *context modules*. Likewise, we call the processing blocks in the localization pathway *localization modules*. Each context module is in fact a pre-activation residual block [12] with two 3x3x3 convolutional layers and a dropout layer ($p_{\text{drop}} = 0.3$) in between. Context modules are connected by stride 2 3x3x3 convolutions. We increase the feature map resolution in the localization pathway by means of upscaling (size 2, stride 2) followed by a 3x3x3 convolution that halves the number of feature maps (*upsampling module*). Following the upsampling, feature maps from the localization pathway are concatenated with feature maps from the context pathway and subsequently passed to a localization module. A localization module consists of a 3x3x3 convolution followed by a 1x1x1 convolution and halves the number of feature maps. Inspired by [8] we employ deep supervision in the localization pathway by integrating segmentation layers at different levels of the network and combining them via elementwise summation to form the final network output. Throughout the network we use leaky ReLu nonlinearities for all feature map computing convolutions. We furthermore replace the traditional batch with instance normalization [13] since we found that the stochasticity induced by small batch sizes destabilizes batch normalization.

Training Procedure Our network architecture is trained with randomly sampled patches of size 128x128x128 voxels and batch size 2. We refer to an epoch as an iteration over 100 batches and train for a total of 300 epochs. Training is done using the ADAM optimizer with an initial learning rate $\text{lr}_{\text{init}} = 5 \cdot 10^{-4}$, the following learning rate schedule: $\text{lr}_{\text{init}} \cdot 0.985^{\text{epoch}}$ and a l2 weight decay of 10^{-5}.

One challenge in medical image segmentation is the class imbalance in the data that hampers the training when using the conventional categorical crossentropy loss. In the BraTS 2017 training data for example, there is 166 times as much background (label 0) as there is enhancing tumor (label 4). We approach this issue by formulating a multiclass Dice loss function that is differentiable and can be easily integrated into deep learning frameworks:

$$\mathcal{L}_{\text{dc}} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_i^k v_i^k}{\sum_i u_i^k + \sum_i v_i^k} \tag{1}$$

where u is the softmax output of the network and v is a one hot encoding of the ground truth segmentation map. Both u and v have shape i by c with i being the number of pixels in the training patch and $k \in K$ being the classes.

When training large neural networks from limited training data, special care has to be taken to prevent overfitting. We address this problem by utilizing a large variety of data augmentation techniques. Whenever possible, we initialize these techniques using aggressive parameters that we subsequently attenuate over the course of the training. The following augmentation techniques were applied on the fly during training: random rotations, random scaling, random elastic deformations, gamma correction augmentation and mirroring.

The fully convolutional nature of our network allows to process arbitrarily sized inputs. At test time we therefore segment an entire patient at once, alleviating problems that may arise when computing the segmentation in tiles with a network that has padded convolutions. We furthermore use test time data augmentation by mirroring the images and averaging the softmax outputs.

2.2 Survival Prediction

The task of survival prediction underpins the clinical relevance of the BraTS challenge, but at the same time is very challenging, particularly due to the absence of treatment information. For this subchallenge, only the image information and the age of the patients was provided. Our approach to survival prediction is based on radiomics. We characterize the tumors using image based features that are computed on the segmentation masks. We compute shape features (13 features), first order statistics (19 features) and gray level co-occurence matrix features (28 features) with the pyradiomics package [14]. The tumor regions for which we computed the features were the edema (ede), enhancing tumor (enh), necrosis (nec), tumor core (core) and whole tumor (whole). We computed only shape features for edema and the whole tumor, shape and first order features for core and the entire feature set for necrosis and enhancing. With the image features being computed for all modalities, we extracted a total of 517 features.

These features are then used for training a regression ensemble for survival prediction. Random forests are well established in the radiomics community for performing well, especially when many features but only few training data are available. These properties make random forest regressors the prime choice for the scenario at hand (518 features, 163 training cases). We train a random forest regressor (RFR) with 1000 trees and the mean squared error as split criterion. Additionally, we designed an ensemble of multilayer perceptrons (MLP) to complement the output of the regression forest. The ensemble consists of 15 MLPs, each with 3 hidden layers, 64 units per layer and trained with a mean squared error loss function. We use batch normalization, dropout ($p_{\text{drop}} = 0.5$) and add gaussian noise ($\mu = 0, \sigma = 0.1$) in each hidden layer. The outputs of the RFR and the MLP ensemble are averaged to obtain our final prediction.

3 Results

Segmentation We trained and evaluated our network on the BraTS 2017 and 2015 training datasets via five fold cross-validation. No external data was used and the network was trained from scratch. Furthermore, we used the five networks obtained by the corresponding cross-validation as an ensemble to predict the respective validation(BraTS 2017) and test (BraTS 2015) set. Both the training set and validation/test set results were evaluated using the online evaluation platforms to ensure comparability with other participants.

Table 1 compares the performance of our algorithm to other state of the art methods on the BraTS 2015 test set. Our method compares favorably to other

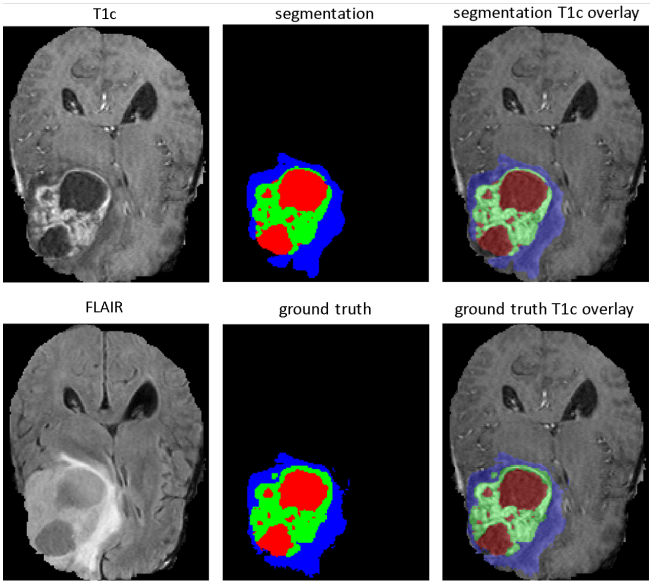


Fig. 2. Qualitative segmentation result. Our approach is capable of segmenting large as well as fine grained regions accurately.

	Dice			Sensitivity			PPV		
	whole	core	enh.	whole	core	enh.	whole	core	enh.
Kamnitsas et al. [6]	**0.85**	0.67	0.63	0.88	0.60	0.67	**0.85**	**0.86**	**0.63**
Kayalibay et al. [8]	**0.85**	0.72	0.61	**0.91**	**0.73**	0.67	0.82	0.77	0.61
ours	**0.85**	**0.74**	**0.64**	**0.91**	**0.73**	**0.72**	0.83	0.80	**0.63**

Table 1. BraTS 2015 test set results.

state of the art neural networks and is currently ranked first in the BraTS 2015 test set online leaderboard. In Table 2 we show an overview over the segmentation performance of our model on the BraTS 2017 dataset. A qualitative segmentation result (Brats17_TCIA_469_1) is shown in Figure 2. Notably, we achieve dice scores of 0.896, 0.797 and 0.732 for whole, core and enhancing, respectively, on the BraTS 2017 validation set. This result places us among the best performing methods according to the online validation leaderboard.

Survival Prediction We extensively evaluated the components of our regression ensemble as well as different feature sets with the aim of minimizing the mean squared error by running 5-fold cross-validations on the 163 provided training cases. A summary of our findings for both the ground truth and our segmentations is shown in Table 3. We observed that the random forest regressor performs very well across all feature sets while the MLP ensemble is much less stable. The overall best results were obtained by averaging the MLP ensem-

Dataset	Dice			Sensitivity			Specificity			Hausdorff Dist.		
	whole	core	enh.	whole	core	enh.	whole	core	enh.	whole	core	enh.
BraTS 2017 Train	0.895	0.828	0.707	0.890	0.831	0.800	0.995	0.997	0.998	6.04	6.95	6.24
BraTS 2017 Val	0.896	0.797	0.732	0.896	0.781	0.790	0.996	0.999	0.998	6.97	9.48	4.55

Table 2. Results for the BraTS 2017 dataset. Train: 5 fold cross-validation on the training data (285 cases). Val: Result on the validation dataset (46 cases).

Features	Ground Truth Segmentation			Our Segmentation		
	RFR	MLP ens	combined	RFR	MLP ens	combined
shape, age (66)	334.89	**352.00**	**339.61**	353.12	**343.19**	**335.08**
glcm, age (225)	348.14	462.16	381.25	**350.78**	388.99	357.41
first order, age (229)	358.69	388.44	362.20	354.66	381.42	355.89
shape, glcm, age (290)	**344.86**	431.96	367.14	346.40	378.73	349.13
shape, first order, age (294)	352.64	372.59	350.62	351.56	360.24	342.46
glcm, first order, age (453)	353.18	443.64	378.83	354.30	383.82	356.25
all (518)	350.40	385.66	354.86	352.95	372.04	348.55

Table 3. Survival prediction experiments. We trained a random forest regressor (RFR) and a MLP ensemble (MLP ens). Averaging RFR and MLP ensemble yields the *combined* result. The best root mean squared error is achieved when using RFR and MLP ensemble together with only shape features and the patients age.

ble output with the one from the random forest regressor (column *combined*) and using only shape features and the age of a patient. Interestingly, while the random forest performance is almost identical between ground truth and our segmentations, the MLP ensemble performs better on our segmentations for all feature sets, which is also reflected by the *combined* results. The best root mean squared error we achieved was 335.08 (mean absolute error 232.76).

4 Discussion

In this paper we presented contribution to the BraTS 2017 challenge. For the segmentation part of the challenge we developed a deep convolutional neural network architecture which was trained using extensive data augmentation and a dice loss formulation. We achieve state of the art results on BraTS 2015 and presented promising scores on the BraTS 2017 validation set. Training time was of about five days per network. Due to time restrictions we were limited in the number of architectural variants and data augmentation methods we could explore, yet we expect to find even better performing constellations for our final test set submission in the near future. Careful architecture optimizations already allowed us to train with large 128x128x128 patches and a batch size of 2 with 16 filters in the highest level, which is significantly more than in [8]. Training with larger batch sizes and more convolutional filters in a multi-GPU setup should yield further improvements, especially provided that we did not observe significant overfitting in our experiments. While most of our effort was concentrated on the segmentation part of the challenge, we also proposed an ensemble of a

random forest regressor and a multilayer perceptron ensemble for the survival prediction subchallenge. By using only shape based features, we achieved a root mean squared error of 335.08 and a mean absolute error of 232.76 in a five fold cross-validation on the training data and using our segmentations.

References

1. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE TMI*, vol. 34, no. 10, pp. 1993–2024, 2015.
2. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Nature Scientific Data*, 2017 (In Press).
3. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection," *TCIA*, 2017.
4. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection," *TCIA*, 2017.
5. M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *MIA*, vol. 35, pp. 18–31, 2017.
6. K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *MIA*, vol. 36, pp. 61–78, 2017.
7. K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, "DeepMedic for brain tumor segmentation," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer, 2016, pp. 138–149.
8. B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," *arXiv preprint arXiv:1701.03056*, 2017.
9. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.
10. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *International Conference on 3D Vision*. IEEE, 2016, pp. 565–571.
11. L. Macyszyn, H. Akbari, J. M. Pisapia, X. Da, M. Attiah, V. Pigrish, Y. Bi, S. Pal, R. V. Davuluri, L. Roccograndi *et al.*, "Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques," *Neuro-oncology*, vol. 18, no. 3, pp. 417–425, 2015.
12. K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *ECCV*. Springer, 2016, pp. 630–645.
13. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint arXiv:1607.08022*, 2016.
14. J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer Research (Accepted)*, 2017.

Fully Convolutional Network with Hypercolumn Features for Brain Tumor Segmentation

Mobarakol Islam[1] and Hongliang Ren[2]

[1] Department of NUS Graduate School for Integrative Sciences and Engineering (NGS), National University of Singapore, Singapore
mobarakol@u.nus.edu
[2] Department of Biomedical Engineering, National University of Singapore, Singapore
ren@u.nus.sg

Abstract. Brain tumor segmentation using multi-modal MRI data sets is important for diagnosis, surgery and follow up evaluation. In this paper, a fully convolutional network (FCN) with hypercolumns features (e.g. PixelNet) utilizes for automatic brain tumor segmentation containing low and high-grade glioblastomas. Though pixel level convolutional predictors like fully-convolutional networks (FCN), are computationally efficient, such approaches are not statistically efficient during learning precisely because spatial redundancy limits the information learned from neighboring pixels. PixelNet extracts features from multiple layers that correspond to the same pixel and samples a modest number of pixels across a small number of images for each SGD (Stochastic gradient descent) batch update. PixelNet has achieved whole tumor dice accuracy 90% and 87% for training and validation respectively.

Keywords: Brain tumor segmentation, Gliomas, BRATS, Deep Learning, Convolutional Neural Network, Pixel level segmentation, Hypercolumn.

1 Introduction

It is very important to segment the gliomas and its intra-tumoral structures as well as estimate relative volume to monitor the progression, assessment, treatment planning and follow-up studies. Generally, the segmentation of gliomas observes in various regions such as active tumorous tissue, necrotic tissue, and the peritumoral edematous which defined through intensity changes relative to the surrounding normal tissue. However, gliomas or glioblastomas are usually spread out, poorly contrasted and intensity information being disseminated across various modalities that make them difficult to segment [1]. The tumor intensity also differs across the patients like HGG patients the tumor consists of enhancing, non-enhancing and necrotic parts, while in the LGG patients it is not necessarily to include an enhancing part [2]. Due to inconsistency and diversity of MRI acquisition parameters [3] and hardware variations, there are large difference in appearance, shape and intensity ranges among the same sequences and acquisition scanners [5], which make the segmentation more challenging. Thus, physicians conventionally use rough evaluation or manual segmentation; however, manual

segmentation is time-consuming and laborious task that is inclined to misinterpretation and observer bias [4].

In recent years, Deep Learning (DL) have drawn increasing attention medical applications such as in object detection [6, 7], semantic segmentation [8] and classification [9]. DL models like convolutional neural networks (CNN) are capable of learning high level and task adaptive hierarchical features from training data and take part as an effective approach. Havae et al. [10] build a CNN based two-pathway cascade network which performs a two-phase training using both local and global contextual features and tackle difficulties related to the imbalance of tumor labels in data. Another similar approach DeepMedic [11] uses two convolutional parallel pathways and 3D CNN architecture with 11-layers for brain lesion segmentation. Later, modified version of DeepMedic with residual connection utilize for brain tumor segmentation [12]. On the other hand, Pandian et al. [13] and Casamitjana et al. [14] use 3D volumetric CNN to train sub- volume of multi-modal MRIs and show that 3D CNN performs well for segmentation as MRI acquires 3D information. The benefit of these architectures is that they performed well with a comparatively smaller dataset. However, they are computationally expensive as it needs 3D kernels and a large number of trainable parameters. Alex et al. [15] uses 5 layers deep Stacked Denoising Auto-Encoder (SDAE) and Randhawa et al [16] uses 8 layers CNN and Pereira et al. [17] uses deeper CNN architecture with small kernel for segmenting gliomas from MRI.

In spatially-invariant label prediction problem like semantic segmentation, every separate label per pixel predicts using a convolutional architecture. As a result, gradient based learning like Stochastic gradient descent (SGD) treats training data as sampled independently and form an identical distribution [18]. Hyvärinen et al. [19] demonstrate that pixel in a given image is highly correlated and neighbouring pixels are not independent. To capture the high-level global context and minimize the loss of the contextual information in higher convolutional layers, there have built many predictors based on multiscale feature extraction from multiple layers of a CNN [20]. Hariharan et al. [21] extracted features of the same pixels from multiple layers and accumulate in a feature vector called “Hypercolumns”. To extract feature, FCNs [22] efficiently implemented linear prediction in a coarse to fine manner. To reduce memory footprint DeepLab [23] incorporate filter dilation and linear-weighted fusion in fully connected layers. ParseNet [24] averages the pooling feature by normalization and concatenation to add spatial context for a layer response. PixelNet [25] adopt both Hariharan et al. [21] and ParseNet [24] to build hypercolumn and concatenate spatial context in the layer where the tradeoff between statistical and computational efficiency for convolutional learning. PixelNet shows state of art performance for in BRATS 2017 [27,28,29,30] training and validation dataset.

2 Methodology

PixelNet extract multi-scale convolution and feature and concatenate them as hypercolumn to ensure all local and global contextual information in the learning phase (Fig. 1). A hyper descriptor can be written as:

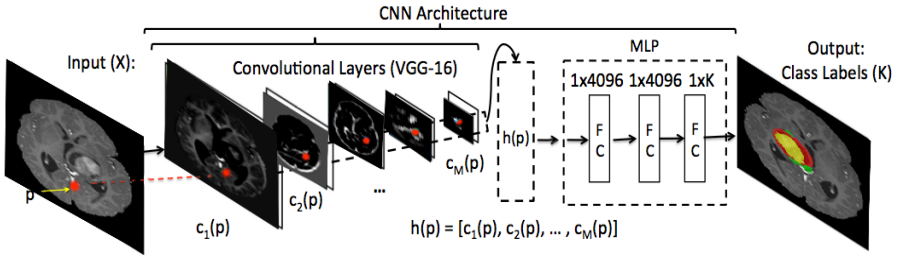


Fig. 1. Three modalitities (Flair, T1, T1c) input in a CNN and extract hypercolumn descriptor for a sample pixel from multiple convolutional layers. The hypercolumn descriptor is then fed to a multi-layer perception for non-linear optimization.

$$h_p = [c_1(p), c_2(p), ..., c_M(p)] \quad (1)$$

where h_p denote the multiscale hypercolumn features for the pixel p, and $c_i(p)$ denote the feature vector from layer i. PixelNet considers pixel wise prediction as operating over hypercolumn features. For example, the final prediction for pixel p,

$$f_{\theta,p}(X) = g(h_p(X)) \quad (2)$$

where θ represent both hypercolumn features h and pixel wise predictor g. θ updates by using SGD training. We use a series of fully connected layers followed by ReLU activation function similar to VGG-16 [26] to implement non-linear predictor. We adopt sparse pixel prediction at training time for efficient mini-batch generation. In sparse prediction, hypercolumn features h_p choose from dense convolutional responses at all layers by computing the 4 discrete locations in the feature map c_i (for i^{th} layer) closest to sampled pixel $p \in P$ and finally apply bilinear interpolation to get i^{th} layer response in hypercolumn.

3 Experiment

3.1 Dataset

BRATS 2017 (Brain Tumor Image Segmentation Benchmark) [27, 28, 29, 30] training database consists in total 285 cases of patients. It is a multi-modal MRI scans of 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) and 4 different modalities including T1 (spin- lattice relaxation), T1c (T1-contrasted), T2 (spin-spin relaxation) and FLAIR (fluid attenuation inversion recovery). Each scan is a continuous 3D volume of 155 2D slices of size 240x240. The volume of the various modalities is already skull-stripped, aligned with T1c and interpolated to 1 mm voxel resolution. The provided ground truth with manual segmentation includes three labels: GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing

tumor (NCR/NET — label 1). The predicted labels are evaluated by merging three regions: whole tumor (WT: all four labels), tumor core (TC: 1,2) and enhancing tumor (ET: 4).

3.2 Training and Evaluation for Segmentation Task

We use depth slicing images on axial orientation on the mixed HGG and LGG data (285 MRI cases in total). However, this dataset is highly imbalance where ground-truth contains 98% pixel are healthy tissue (label 0) and remaining are also unequal ratios of four ROI labels such as edema, enhancing, necrotic and non-enhancing. We deal this issue by ignoring all the blank slices in groundtruth (both background and healthy issue) and train PixelNet with corresponding 3 modalities such as flair, T1C and T2. So we ignore T1 scan to see the performance of the PixelNet. Though BRATS 2017 has in total 44175 (285x155) slices, we utilize only 18924 (43% data) slices corresponding to ground-truth with non-zero class (contains at least one class 1 or 2 or 4) in our first observation. We use Caffe deep learning platform to perform all of our experiments.

Table 1. Dice and Sensitivity for BRATS 2017 training dataset

Level	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.711	0.909	0.866	0.771	0.897	0.831
StdDev	0.293	0.070	0.115	0.231	0.091	0.139
Median	0.830	0.929	0.902	0.849	0.924	0.872
25quantile	0.706	0.896	0.837	0.727	0.872	0.786
75quantile	0.882	0.946	0.930	0.916	0.951	0.916

Table 2. Specificity and Hausdorff95 for BRATS 2017 training dataset

Level	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.998	0.995	0.998	6.946	7.275	6.103
StdDev	0.002	0.005	0.002	15.362	13.494	11.546
Median	0.999	0.996	0.999	2.000	3.000	3.162
25quantile	0.998	0.994	0.998	1.414	2.236	2.236
75quantile	1.000	0.998	0.999	3.606	4.899	4.583

After PixelNet prediction, we evaluate all the cases for training set (285 cases) and validation (46 cases) using online evaluation portal for BRATS 2017 challenge. Table 1 and 2 represent the training set evaluation results where whole tumor average Dice accuracy is 90% and Hausdorff distance is 7.3 which is quite promising. Table 3 and 4 shows the evaluation results of validation set where average 87% dice accuracy and 9.8 Hausdorff distance. Though enhance tumor and tumor core region have lower accuracy than whole tumor, the individual accuracy can be considered as state of art performance.

Fig. 2 shows some visualized examples of PixelNet prediction with comparing ground-truth.

Table 3. Dice and Sensitivity for BRATS 2017 validation dataset

Level	Dice			Sensitivity		
	ET	WT	TC	ET	WT	TC
Mean	0.689	0.876	0.761	0.720	0.861	0.710
StdDev	0.304	0.086	0.221	0.287	0.136	0.253
Median	0.829	0.902	0.849	0.840	0.906	0.795
25quantile	0.625	0.877	0.708	0.657	0.849	0.562
75quantile	0.881	0.929	0.912	0.883	0.949	0.910

Table 4. Specificity and Hausdorff95 for BRATS 2017 validation dataset

Level	Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Mean	0.998	0.995	0.998	12.938	9.820	12.361
StdDev	0.002	0.005	0.003	26.453	13.516	20.826
Median	0.999	0.996	0.999	2.449	4.581	5.050
25quantile	0.998	0.993	0.998	1.799	2.828	3.041
75quantile	1.000	0.998	1.000	8.569	8.093	11.176

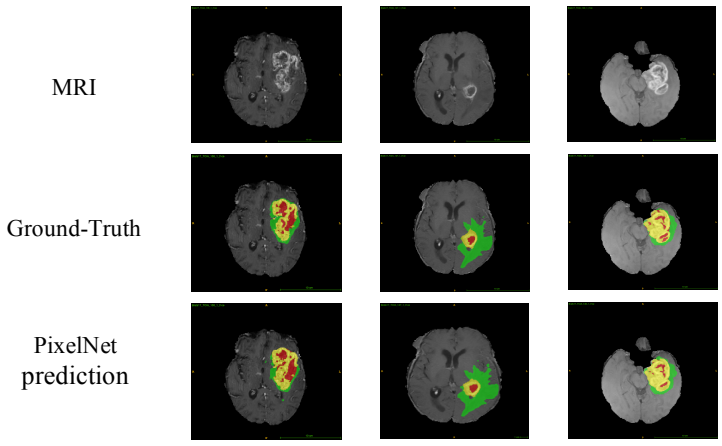


Fig. 2. PixelNet prediction for BRATS 2017 Training Dataset.

Table 5. Dice, Sensitivity and Hausdorff95 for BRATS 2017 validation dataset with reduced training data.

	Dice			Sensitivity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC
PixelNet (43% data)	0.689	0.876	0.761	0.720	0.861	0.710	12.938	9.820	12.361
PixelNet (29% data)	0.677	0.861	0.775	0.722	0.818	0.776	14.675	11.808	23.726

To see the small dataset effect to PixelNet we train our model with only 12930 (29% data) slices corresponding to ground-truth consisting all three classes (1, 2, 4) and compare the performance with trained model of all non-zero slices or at least one class slices (43% data: 18924 slices). Table 5 shows the Dice and Hausdorff95 evaluation for validation dataset by varying amount of training data. PixelNet achieves almost same performance for less dataset.

4 Conclusion

We presented an automatic brain tumor segmentation method based on pixel level semantic segmentation. We choose PixelNet which extracts multi layers convolutional feature and form hypercolumn. Hypercolumn contains useful contextual information and use sparse pixel prediction to generate efficient mini-batch which produces promising results for Brain tumor segmentation. Though the preliminary results of the ET and TC are not good as WT however we are still working on this model to achieve better accuracy in all the regions of the tumor.

Acknowledgments: I would like to thank Aayush Bansal, the author of PixelNet [25], for the assistance to implement PixelNet for this project.

References

1. Nyul, L.G., Udupa, J.K., Zhang, X.: New variants of a method of MRI scale standardization. IEEE Transactions on Medical Imaging 19(2),143-50 (2000).
2. Ellwaa, A., et al: Brain Tumor Segmantation using Random Forest trained on iterative selected patients. In proc. of BRATS-MICCAI, (2016).
3. Omuro, A., DeAngelis, L. M.: Glioblastoma and Other Malignant Gliomas: A Clinical Review. Jama, vol. 310, pp. 1842-1850, (2013).
4. Bauer, S., et al: A survey of MRI-based medical image analysis for brain tumor studies. Physics in Medicine and Biology, vol. 58, pp. R97, (2013).
5. Inda, Maria-del-Mar, Bonavia, R., Seoane, J.: Glioblastoma multiforme: A look inside its heterogeneous nature. Cancers 6.1, pp. 226-239 , (2014).
6. Baumgartner, C. F., Kamnitsas, K., Matthew, J., Smith, S., Kainz, B., Rueckert, D.: Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: Med Image Comput Comput Assist Interv. Vol. 9901 of Lect Notes Comput Sci. pp. 203–211, (2016).

7. Azizi, S., Imani, F., Ghavidel, S., Tahmasebi, A., Kwak, J. T., Xu, S., Turkbey, B., Choyke, P., Pinto, P., Wood, B., Mousavi, P., Abolmaesumi, P.: Detection of prostate cancer using temporal sequences of ultrasound data: a large clinical feasibility study. Int J Comput Assist Radiol Surg 11 (6), 947–956, (2016).
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597v1, (2015).
9. Milletari, F., Navab, N., Ahmadi, S. A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. arXiv:1606.04797v1, (2016).
10. Havaei, M., Davy, A., Warde-Farley, D.: Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis, vol. 35, pp. 18-31, (2017).
11. Kamnitsas, K., et al: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, vol. 36, pp. 61-78, (2017).
12. Kamnitsas, K., et al: DeepMedic on Brain Tumor Segmentation. In proc. of BRATS-MICCAI, (2016).
13. Pandian1, B., Boyle1, J., Orringer, D. A.: Multimodal Tumor Segmentation with 3D Volumetric Convolutional Neural Networks. In proc. of BRATS-MICCAI, (2016).
14. Casamitjana, A., et al: 3D Convolutional Networks for Brain Tumor Segmentation. In proc. of BRATS-MICCAI, (2016).
15. Alex, V., Krishnamurthi, G.: Brain Tumor Segmentation from Multi Modal MR images using Stacked Denoising Autoencoders. In proc. of BRATS-MICCAI, (2016).
16. Randhawa, R., Modi, A., Jain, P., Warier, P.: Improving segment boundary classification for Brain Tumor Segmentation and longitudinal disease progression. In proc. of BRATS-MICCAI, (2016).
17. Pereira, S., et al: Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Transactions on Medical Imaging, vol. 35, pp. 1240-1251, (2016).
18. Bottou, L.: Large-scale machine learning with stochastic gradient descent," in Proceedings of COMPSTAT' 2010, pp. 177–186, Springer, (2010).
19. Hyvärinen, A., Hurri, J., Hoyer, P. O.: Natural Image Statistics: A Probabilistic Approach to Early Computational Vision. (1. Aufl. ed.) 200939. DOI: 10.1007/978-1-84882-491-1, (2009).
20. Denton, E. L., Chintala, S., Fergus, R., et al: Deep generative image models using a laplacian pyramid of adversarial networks. In NIPS, (2015).
21. Hariharan, B., Arbelaez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In CVPR, (2015).
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional models for semantic segmentation. In CVPR, (2015).
23. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. In ICLR, (2015).
24. Liu, W., Rabinovich, A., Berg, A. C.: Parsenet: Looking wider to see better. arXiv preprint arXiv:1506.04579, (2015).
25. Bansal, A., et al: PixelNet: Representation of the pixels, by the pixels, and for the pixels. arXiv:1702.06506v1, (2017).
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, (2014).
27. Menze, B., et al: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging 34 (2015).
28. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby J. S., Freymann, J. B., Farahani, K., Davatzikos, C.: Advancing the Cancer Genome Atlas glioma MRI collections

with expert segmentation labels and radiomic features. Nature Scientific Data, (2017) [In Press].
29. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
30. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.

Brain Tumor Segmentation using a 3D FCN with Multi-Scale Loss

Andrew Jesson and Tal Arbel

McGill University., Montreal, Qc, Canada
ajesson@cim.mcgill.ca
arbel@cim.mcgill.ca
http://www.cim.mcgill.ca

Abstract. In this work, we present a multi-task, 3D Fully Connected Network (FCN) architecture for brain tumor segmentation. Our method includes a multi-scale loss function on predictions given at each resolution of the FCN. Using this approach, the higher resolution features can be combined with the initial segmentation at a lower resolution so that the FCN models context in both the image and label domains. The model is trained using a multi-scale loss function and a curriculum on sample weights is employed to address class imbalance, showing competitive results for brain tumor segmentation. Prediction results are not yet available.

1 Introduction

In this paper, we present a 3D fully connected network with multi-scale loss for the segmentation of brain tumours. Our framework was submitted to the 2017 MICCAI Brain Tumor Segmentation (BraTS) Challenge [1–3, 9]. The 2017 BraTS Challenge is comprised of two tasks: segmentation of high and low grade glioma in multi-channel MRI, and the prediction of patient survival time.

Brain tumor segmentation is a challenging task due, primarily, to three sources of variability across patient images: (1) Variability across size, shape, and texture of gliomas and surrounding edema, (2) variability in normal brain anatomy, and (3) variability in intensity range and contrast in qualitative MR imaging modalities. Additionally, the proportion of positive tumor classes to normal brain anatomy is very low, resulting in extreme class imbalance. Successful methods tackling this problem must then model local and global context in the raw imaging data, account for local and global interactions between classes, and address the difficulties arising from class imbalance.

Popular approaches to brain tumour segmentation include probabilistic graphical models [9, 11], classical machine learning [9], and deep learning [5, 6]. Probabilistic graphical models address the challenges by modelling the statistical distributions of image intensities and textures over the tumor classes, while potentially also incorporating atlas derived spatial prior probability maps for normal brain structures [11]. Incorporation of pixel and structure level Markov Random Field (MRF) modelling has also been used to capture local and global

interactions between classes [11]. Machine learning techniques capture local and global context using features derived from hand crafted filters, which are then used as input to Random Forrest or Support Vector Machine models for pixel wise classification [9].

Deep learning approaches to brain tumor segmentation have been explored, through the use of Fully Convolutional Networks (FCNs) [8] or UNets [10]. Fully Convolutional Networks have shown to be effective solutions for semantic segmentation in both natural [8] and medical images [5, 6, 10]. This is due to their ability to learn and combine meaningful multi-scale features for pixel classification. Furthermore, these models do not generally require extensive data preprocessing to give state of the art results [6]. In this manner, FCNs address the problem of modelling context in the imaging domain as well as, indirectly, the problems arising from image intensity and contrast variability in MRI. One limitation of FCNs is that they do not explicitly model context in the label domain. This limitation has been addressed by cascading an FCN with a graphical model such as a CRF [6]. This approach has been shown to improve the results of FCNs applied to the problem of stroke lesion segmentation [6], but generally requires a two stage model. Incorporating both an FCN and a CRF into a single, end-to-end model is still an open research area.

In this work, we present a multi-task 3D FCN architecture for brain tumor segmentation. Our method includes a multi-scale loss function on predictions given at each resolution of the FCN. In this way, the higher resolution features can be combined with the initial segmentation at a lower resolution so that the FCN models context in both the image and label domains. The model is trained using a multi-scale loss function and a curriculum on sample weights is employed to address class imbalance, showing competitive results for brain tumor segmentation.

The BRaTs challenge also involves a task for the prediction of patient survival. Prediction of clinical outcome in medical imaging present similar challenges in this context, compounded by the relatively small amount of training data available. Prediction tasks are usually carried out using anatomical imaging biomarkers (such as volume, shape, and texture of various structures), and clinical data as features for statistical, or machine learning models [1, 4]. We are in the process of incorporating patient survival prediction into the FCN network. Results for patient survival prediction have not yet been completed at the time of this submission.

2 Method

We now describe the FCN framework for brain tumour segmentation. The choice of method is informed by its capacity to address the four major challenges in brain tumor segmentation, namely: modelling context in the image domain, modelling context in the label domain, addressing the qualitative nature of MRI, and addressing the class imbalance persistent in medical imaging tasks.

Fully Convolutional Networks are are trained "online," which allows for the employment of a learning curriculum to address the class imbalance problem. Here we use pixel-wise sample weighting so that each training example will contribute to the loss based on the frequency of occurrence for its class. This forces the model to learn features for under represented classes that otherwise would have little influence on the objective function. The class weights are then decayed at each training iteration until all samples are equally weighted.

2.1 Model Architecture

The model architecture is shown in Figure 1. White boxes indicate model inputs, here a multi-channel MR image. Red boxes indicate model outputs. In this case confidence maps for each class to be segmented. Blue boxes indicate feature maps. Feature maps are extracted using learned 3D convolution operations. Green boxes indicate feature vectors. The feature vectors are produced by a global average pooling operation on each feature map. This results in feature vectors to be used for prediction. Each box contains the number of features and size of feature map if applicable. Arrows indicate operations. The color code for each operation is shown in the figure.

2.2 Loss Function

Here we describe the objective functions that are optimized during training. Let a given training data set have N pixels, truth segmentation distributions $\mathbf{p} = \{p_1 ... p_M\}$ for K classes, and a survival time $\mathtt{t}$. Further, let the model produce segmentation maps $\mathbf{q} = \{q_1 ... q_M\}$ at M resolutions, and a survival time prediction $\hat{t}$. Finally to each segmentation class corresponds a weight w_k and the regression target has a corresponding weight $w_{\mathtt{t}}$ The objective function for segmentation when using categorical cross entropy is then:

$$\mathcal{L}_S = -\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{l=1}^{K} w_k \times p_{i,j,k} \times \ln\left(q_{i,j,k}\right). \tag{1}$$

The objective function for survival time regression or classification is given by:

$$\mathcal{L}_T = w_t \times \mathcal{L}_t^{\star}\left(\mathtt{t}, \hat{t}\right), \tag{2}$$

where, $\mathcal{L}_t^{\star}$ is an appropriate objective function, for example, squared error for regression. The model can be jointly optimized for segmentation and outcome prediction by taking a weighted sum of these two losses:

$$\mathcal{L} = \lambda_S \mathcal{L}_S + \lambda_T \mathcal{L}_T. \tag{3}$$

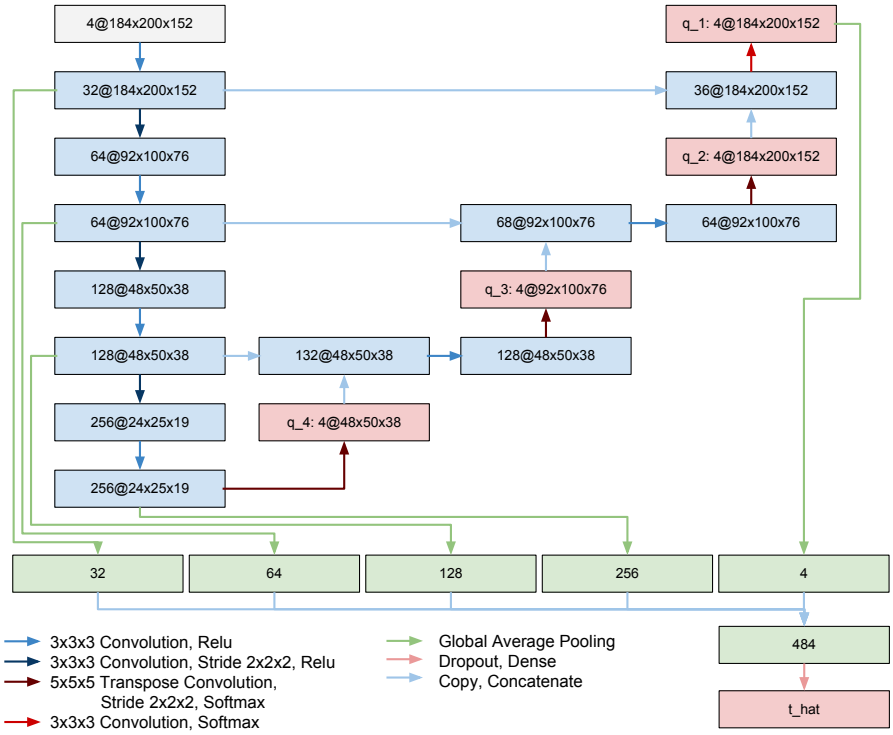


Fig. 1. Architecture of model.

2.3 Curriculum Sample Weighting

The sample weights, $\mathbf{w} = \{w_1 ... w_K\}$ and w_{t} are decayed over each epoch n according to the following curriculum:

$$w_\star(n) = \frac{1}{f_\star} \times r^n + 1, \tag{4}$$

where $f_\star$ is the frequency of occurrence of a given target over the training set and r is a rate parameter on $(0, 1)$. Notice that the weights converge to 1 as the number of epochs grows large ensuring that all samples receive an equal weight at the later training stages.

3 Experiments and Results

3.1 Data

BraTS 2017 Training Set. The BraTS 2017 training data set is comprised of 210 high-grade and 75 low-grade glioma patient data sets. Each data set

contains a T1, T1 post contrast (T1c), T2, and FLAIR MR image, along with an expert tumor segmentation. Each brain tumor is segmented into 3 classes: edema, necrotic/non-enhancing core, and enhancing tumor core. Survival time in days and age in years is provided for 164 of the high-grade data sets [1–3, 9].

BraTS 2017 Validation Set. The BraTS 2017 validation data set is comprised of 46 patient data sets. Each data set contains a T1, T1 post contrast (T1c), T2, and FLAIR MR image. No expert tumor segmentation masks are provided and the grade of each glioma is not specified [1–3, 9].

3.2 Preprocessing

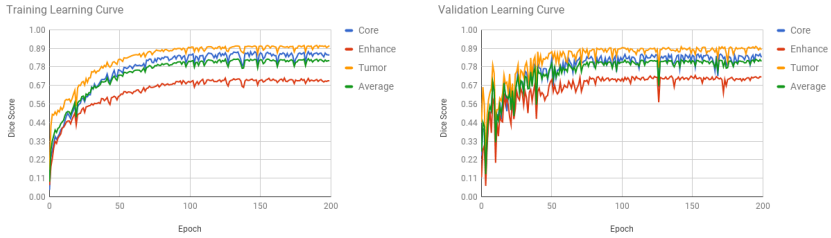
The challenge MRI data have been skull stripped and co-registered by the organizers [1–3, 9]. We have done minimal additional preprocessing. Namely, we standardize the intensities of each image using the mean and standard deviation over the masked region of a given MR image, and the images are cropped to 184x200x152.

3.3 5-Fold Cross Validation

We employ 5-fold cross validation on the training set for hyper-parameter optimization. The BraTS 2017 training set is randomly split into five folds with 57 patient data sets each. For each test fold a model is trained on three and validated on one of the remaining folds.

Parameters We optimize the loss function in equation 3 using Adam [7] with a learning rate of 0.0002 and batch size of 1. The decay rate r in equation 4 is set to 0.95. The initial weights in equation 4 are set to $[1, 210, 90, 280]$ for background, tumor core, edema, and enhancing tumor, respectively. We regularize the model using data augmentation, where at each training iteration a random affine transformation is applied to the image, segmentation mask, and sample weights. Random rotation and shear angles in degrees are drawn independently from a unit normal distribution. Random scales for each spatial dimension are drawn independently from $\mathcal{N}(1, (\frac{1}{15})^2)$. Images are also randomly flipped left to right.

Learning Curves Figure 2 shows an example of the evolution of the dice scores for tumor core, enhancing tumor, and whole tumor for one of the 5-fold cross validation experiments. We see that the curriculum results in stable training and that the model does not over fit.



(a) (b)

Fig. 2. Tracking dice scores during training. In these two plots we show the training curves for one of the five cross-validation experiments. Figure 2a shows the dice scores for training. Figure 2b shows the dice scores for validation.

3.4 Quantitative Results

Tables 1 and 2 show the results for the challenge metrics on the training and validation data sets, respectively. A visual comparison of these results between the training and validation sets is shown in Figure 3. We can see consistent performance on both the training and validation data, which indicates that this model generalizes well to unseen examples.

	Dice			Sensitivity			Specificity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.682	0.886	0.789	0.750	0.878	0.782	0.997	0.994	0.996	6.58	7.11	8.11
StdDev	0.285	0.077	0.213	0.248	0.114	0.222	0.003	0.006	0.007	10.3	10.5	10.4
Median	0.789	0.909	0.877	0.847	0.916	0.868	0.998	0.996	0.998	2.83	3.61	4.12
25quantile	0.638	0.865	0.734	0.678	0.845	0.710	0.997	0.993	0.997	2.00	2.24	2.24
75quantile	0.863	0.937	0.923	0.915	0.953	0.924	0.999	0.998	0.999	5.66	6.71	9.38

Table 1. Challenge Metric Statistics: 5-Fold Cross-Validation on BraTS 2017 Training Set. Results are specified for enhancing tumor (ET), whole tumor (WT), and tumor core (TC)

3.5 Qualitative Results

Figure 4 shows example segmentation masks on predicted high-grade test cases from the 5-fold cross validation experiment. Figure 5 shows example segmentation masks on predicted low-grade test cases from the 5-fold cross validation experiment. In both cases we see segmentation results consistent with the provided ground truth images. Furthermore, we see that the model has learned to

	Dice			Sensitivity			Specificity			Hausdorf-95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.713	0.899	0.751	0.732	0.904	0.720	0.998	0.995	0.998	6.98	4.16	8.65
StdDev	0.291	0.070	0.240	0.288	0.102	0.259	0.003	0.004	0.003	12.1	3.37	9.35
Median	0.844	0.908	0.820	0.834	0.939	0.839	0.999	0.996	0.999	3.00	3.08	5.65
25quantile	0.650	0.891	0.685	0.676	0.902	0.558	0.998	0.992	0.998	2.00	2.24	2.24
75quantile	0.891	0.947	0.935	0.905	0.962	0.924	0.999	0.998	0.999	4.24	5.26	11.7

Table 2. Challenge Metric Statistics: BraTS 2017 Validation Set. Results are specified for enhancing tumor (ET), whole tumor (WT), and tumor core (TC)

model context well by noticing that normal enhancements in the T1c images are not predicted as enhancing tumor. This holds even when those enhancements are adjacent to the predicted tumor.

Figure 6 shows example segmentation masks on predicted cases from the BraTS 2017 validation data. No ground truth masks are provided with this data. It appears as though each prediction is reasonable.

4 Conclusion

In this work, we showed how a multi-task, 3D FCN architecture can be successfully developed for the context of brain tumor segmentation. Our multi-scale network combines higher resolution features with the lower level segmentation results, permitting the FCN to model context in both the image and label domains. This architecture includes a multi-scale loss function on predictions given at each resolution of the segmentation FCN. The model is trained using a curriculum on sample weights to address class imbalance, showing competitive results for brain tumor segmentation. Prediction results are not yet available.

References

1. Bakas S, A.H., et al.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q (2017)
2. Bakas S, A.H., et al.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF (2017)
3. Bakas S, A.H., et al.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017 [In Press])
4. Gillies, R.J., Kinahan, P.E., et al.: Radiomics: Images are more than pictures, they are data. Radiology 278(2), 563–577 (2016), `https://doi.org/10.1148/radiol.2015151169`, pMID: 26579733

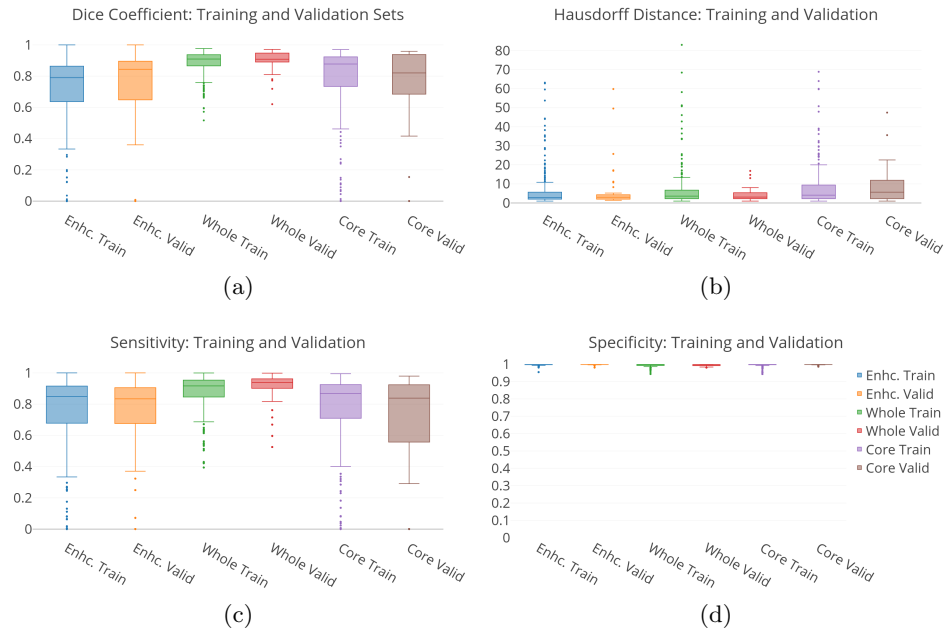


Fig. 3. Box plots of challenge metrics for the BraTS 2017 Training and BraTS 2017 Validation data sets. Results are shown in pairs for each data set and are given for enhancing tumor (enhc.), whole tumor, and tumor core.

5. Havaei, M., Davy, A., et al.: Brain tumor segmentation with deep neural networks. Medical image analysis 35, 18–31 (2017)
6. Kamnitsas, K., Ledig, C., et al.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. Medical image analysis 36, 61–78 (2017)
7. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
8. Long, J., Shelhamer, E., et al.: Fully convolutional networks for semantic segmentation. In: Proc, CVPR. pp. 3431–3440 (2015)
9. Menze, B.H., Jakab, A., et al.: The multimodal brain tumor image segmentation benchmark (brats). Tr. MI 34(10), 1993–2024 (2015)
10. Ronneberger, O., Fischer, P., et al.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
11. Subbanna, N.K., Precup, D., et al.: Hierarchical probabilistic gabor and mrf segmentation of brain tumours in mri volumes. In: Proc. MICCAI. pp. 751–758. Springer (2013)

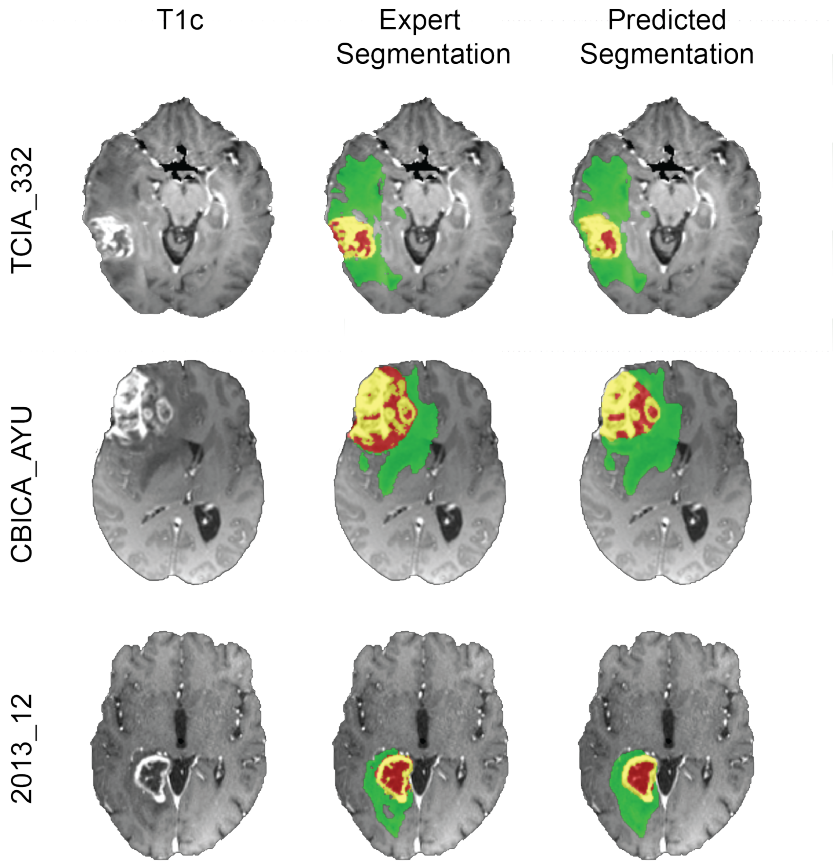


Fig. 4. Examples of high-grade glioma segmentation results for BraTS 2017 Training Data [1–3, 9] 5-fold cross-validation experiment. Segmentation images are overlaid on preprocessed T1c image. Each row is an axial slice taken from a different patient. The green label is edema, the red label is non-enhancing or necrotic tumor core, and the yellow label is enhancing tumor core.

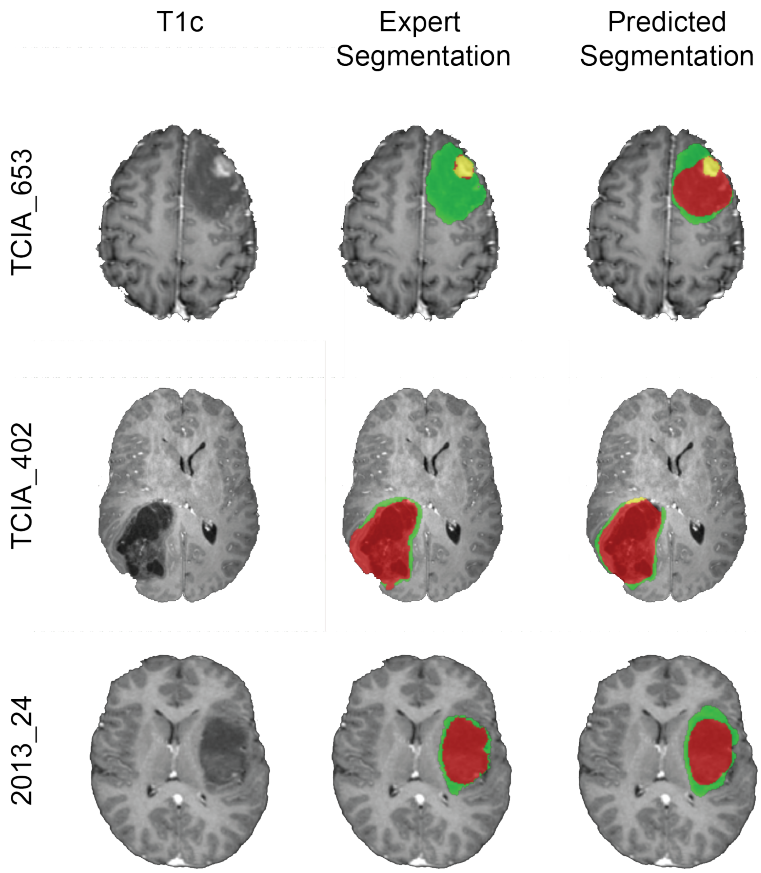


Fig. 5. Examples of low-grade glioma segmentation results for BraTS 2017 Training Data [1–3, 9] 5-fold cross-validation experiment. Segmentation images are overlaid on preprocessed T1c images. Each row is an axial slice taken from different patients. The green label is edema, the red label is non-enhancing or necrotic tumor core, and the yellow label is enhancing tumor core.

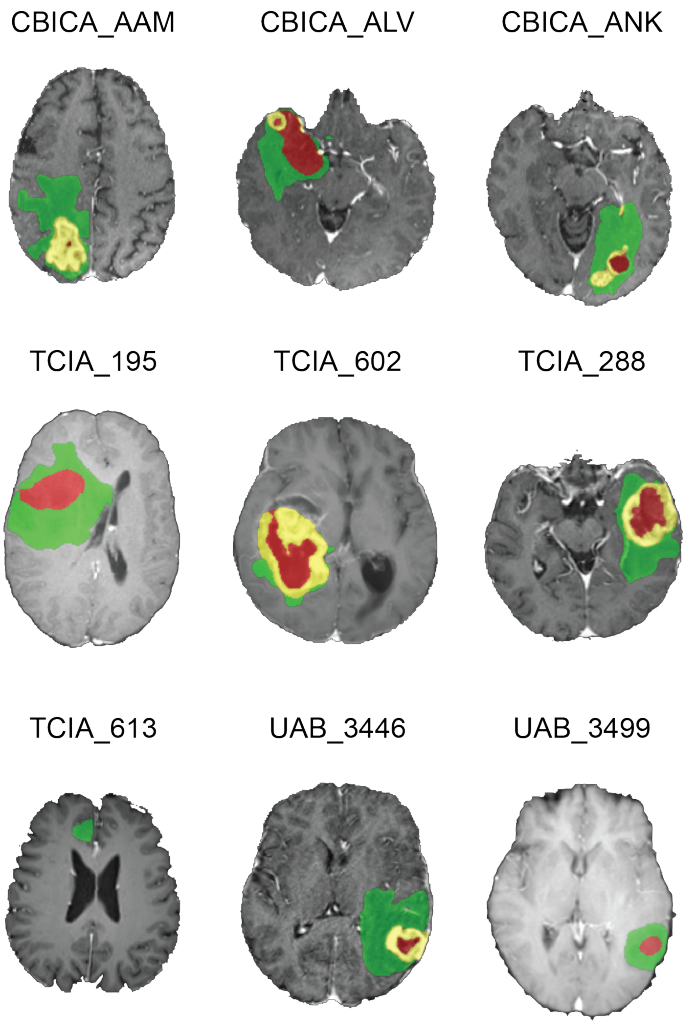


Fig. 6. Examples of segmentation results from BraTS 2017 Validation Data [1–3, 9]. Segmentation images are overlaid on unprocessed T1c images. Each image is an axial slice taken from different patients. The green label is edema, the red label is non-enhancing or necrotic tumor core, and the yellow label is enhancing tumor core.

Uncertainty-assisted Brain Tumor Segmentation and Survival Prediction

Alain Jungo[1], Richard McKinley[2], Raphael Meier[1], Urspeter Knecht[2], Luis Vera[3], Julián Pérez-Beteta[3], David Molina-García[3], Víctor M. Pérez-García[3], Roland Wiest[2], and Mauricio Reyes[1]

[1] Institute for Surgical Technology and Biomechanics, University of Bern
[2] Support Center for Advanced Neuroimaging, Institute for Diagnostic and Interventional Neuroradiology, University Hospital Inselspital and University of Bern
[3] Mathematical Oncology Laboratory Universidad de Castilla-La Mancha
alain.jungo@istb.unibe.ch

Abstract. Uncertainty measures of medical image analysis technologies, such as deep learning, are expected to facilitate their clinical acceptance and synergies with human expertise. Therefore, we propose a full-resolution residual convolutional neural network (FRRN) for brain tumor segmentation that quantifies the uncertainty by the principle of Bayesian Dropout. We further employ uncertainty information to perform uncertainty-assisted correction of segmentation results. The proposed Bayesian Dropout FRRN architecture achieves similar segmentation performance compared to a standard FRRN and a superior result compared to a clinically-validated state-of-the-art approach. A qualitative evaluation further suggests that uncertainty-assisted corrections can improve segmentation results.

Keywords: Deep Learning, Brain Tumor Segmentation, Uncertainty estimation

1 Introduction

Over the past years, large improvements could be observed in brain tumor segmentation. This is partly due to the adoption of the fast-evolving deep learning approaches from the field of computer vision. An even more important reason for the recent advances is the availability of public datasets and online benchmarks [10]. This progress has later guided research to focus on optimizing model architectures for achieving high segmentation performance. However, as the accuracy of these systems still requires expert monitoring of results, clinical applications such as radiological and high-throughput data analysis would benefit greatly from additional uncertainty information along with a good segmentation performance. Information on the segmentation uncertainty could be used to e.g. guide an operator in making manual corrections to the automatic segmentation

results. In this work, we thus focus on the largely unexplored aspect of quantifying model uncertainty in the context of brain tumor segmentation. Existing work of uncertainty in brain tumor segmentation include a perturbation-based approach for conditional random fields [1, 8] and a level set-based method defined via a Gaussian Process [7]. The limitations of these techniques are their lack of transferability to neural networks and their restriction to quantify predictive uncertainty of a given model only.

The aim of this work is to explore uncertainty estimation in deep learning-based methods for brain tumor segmentation. Therefore, as a baseline we adopt the methodology of the state-of-the-art full-resolution residual network (FRRN) [11] architecture. Then, we incorporate the idea of Bayesian Dropouts based on Gal et al. [5] to obtain model uncertainty. In a first experiment, we compare the Bayesian FRRN version (B-FRRN) performance to the standard FRRN and to the existing and clinically-validated brain tumor segmentation approach BraTumIA [9]. The impact of uncertainty-assisted corrections on the final segmentation result is studied in a second experiment.

2 Methods

In this section, the adopted CNN architecture and the incorporated Bayesian Dropout are presented.

2.1 FRRN Architecture

The adopted full-resolution residual network (FRRN) [11] is based on two streams; a residual stream and a pooling stream. The first one is responsible for maintaining a residual path between the network input and output. This improves the gradient flow and thus the training. Moreover, the residual stream allows the network to carry information at full image resolution required for precise segmentation of the image details [11]. The second stream reduces the resolution by pooling operations before returning to original resolution by upsampling. Due to the reduced resolution, the filters on the pooling stream can capture contextual information. An important aspect of the architecture are the connections between pooling and residual streams. This enables the network to simultaneously combine both global and local image information [11].

We propose a full-resolution residual network architecture with four max-pooling/upsampling steps (Figure 1) which takes axial slices of all four sequences (T1-weighted, T1-weighted post-contrast, T2-weighted, FLAIR) as input. A detailed view of the residual units (RU) and full-resolution residual units (FRRU) is shown in Figure 2.

2.2 Bayesian Uncertainty Estimation

As presented in [5], the Dropout regularization can be interpreted as an approximation for Bayesian inference over the weights of the network. A fully Bayesian

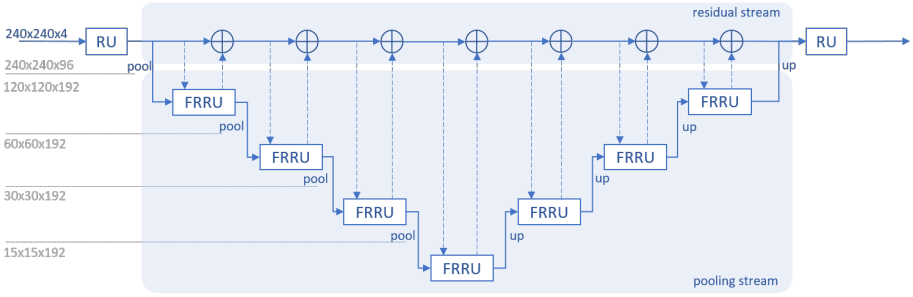
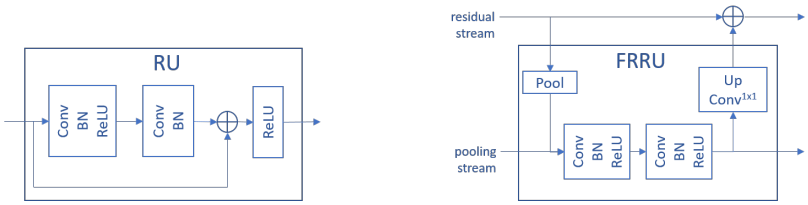


Fig. 1: Full-resolution residual network with four pooling steps. Dashed lines represent the exchange connections between the residual and pooling streams.



(a) (b)

Fig. 2: Detailed view of the units of the architecture in Figure 1. (a) The residual unit (RU) including its residual connection. (b) The full-resolution residual unit (FRRU) where *pool* and *up* adapt to the pooling and residual stream, respectively. The 1×1 convolution aligns the number of feature channels among the streams. Unless specified differently, the convolution kernels are of size 3×3.

network would require applying Dropout after each convolution layer. Following Kendall et al. [6], we omit the strongly regularizing fully Bayesian network, in favor of more computationally efficient training times. As depicted in Figure 3, the Dropout layers of the Bayesian Dropout FRRN (B-FRRN) are placed after each pooling and before each upsampling operation.

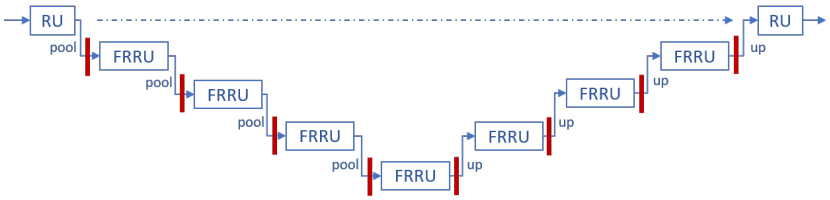


Fig. 3: Positions of the Dropout layers in the B-FRRN network. Residual path is omitted for the sake of clarity.

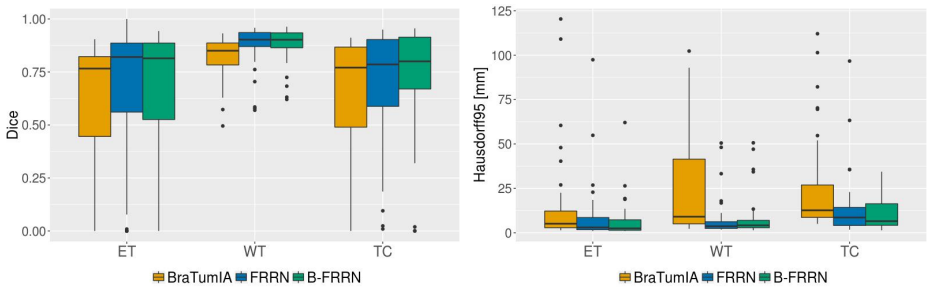
The Dropouts are applied during training and test time. At test time, the Dropouts produce randomly sampled networks, which can be viewed as Monte Carlo samples over the posterior distribution of the models. K network samples are used to produce one prediction with uncertainty estimation. The classification of one voxel is determined by the average of class probabilities over K predictions. As described in [6], the class uncertainty can be computed from the variance in the softmax probabilities of the K predictions. In order to correct a prediction with the help of the uncertainty, we compute the largest and most uncertain regions. These regions can efficiently guide the clinician to the locations to be revised.

3 Experiments & Results

In this section, we first focus on the performance of B-FRRN in comparison to BraTumIA [9] and a FRRN [11]. In a second experiment, we perform a qualitative evaluation of the uncertainty-assisted corrections.

3.1 Comparison between B-FRRN, FRRN & BraTumIA

According to Kendall et al. [6] a minimum of approximately $K = 6$ Dropout Monte Carlo samples are required to improve segmentation performance (on the CamVid dataset) compared to an architecture where the Dropout weights are averaged during testing. For the B-FRRN model we use a rather large $K = 20$. The reason is that, compared to [6], we are not only interested in the segmentation performance but also in the uncertainty comprised in the K predictions.



(a) (b)

Fig. 4: Boxplots for the Dice coefficient (a) and Hausdorff (95^{th} percentile) distance (b).

The comparison of the approaches was performed on the BraTS17 validation dataset [10, 2–4]. While the FRRN and B-FRRN models are trained on 100 randomly selected subjects of the BraTS17 training dataset [10, 2–4], we used a

Table 1: Quantitative results of the comparison between BraTumIA, FRRN and Bayesian FRRN. Bold numbers highlight the best result for a given metric and tumor region (ET=enhancing tumor, WT=whole tumor, TC=tumor core).

	Model	ET	WT	TC
Dice	BraTumIA	0.606 (±0.316)	0.827 (±0.094)	0.635 (±0.299)
	FRRN	**0.678** (±0.304)	0.876 (±0.095)	0.701 (±0.262)
	B-FRRN	0.674 (±0.316)	**0.884** (±0.080)	**0.726** (±0.251)
Sensitivity	BraTumIA	0.705 (±0.293)	**0.882** (±0.102)	0.640 (±0.319)
	FRRN	0.732 (±0.272)	0.848 (±0.138)	**0.712** (±0.283)
	B-FRRN	**0.759** (±0.271)	0.879 (±0.117)	0.705 (±0.266)
Specificity	BraTumIA	0.998 (±0.003)	0.988 (±0.008)	**0.997** (±0.003)
	FRRN	0.998 (±0.003)	**0.996** (±0.003)	0.996 (±0.005)
	B-FRRN	0.998 (±0.003)	0.994 (±0.005)	0.997 (±0.005)
Hausdorff95$_{[mm]}$	BraTumIA	14.92 (±25.94)	24.27 (±27.49)	24.85 (±26.01)
	FRRN	9.00 (±17.09)	**7.24** (±10.62)	13.34 (±16.86)
	B-FRRN	**6.63** (±10.80)	7.93 (±11.20)	**10.91** (±9.22)

BraTumIA version that was trained on a independent dataset consisting of 54 MRI examinations (described in more detail in [9]). Table 1 lists a summary of the achieved results for the three methods. Additionally, the distribution of the obtained Dice coefficients and Hausdorff (95^{th} percentile) distances are presented in Figure 4.

3.2 Uncertainty-assisted Correction

The uncertainty obtained from the B-FRRN model was used to perform the corrections with a brushing tool [13]. To guide the uncertainty-assisted corrections on the segmentation, the largest and most uncertain regions were visualized in form of a colored overlay, see Figure 5 (left).

Two qualitative results of segmentation errors that could be detected due to the uncertainty-guidance, are shown in Figure 5. Table 2 presents the Dice coefficients before and after uncertainty-assisted correction for the same two subjects.

4 Survival Prediction

The survival prediction for the Brats 2017 database was carried out in two phases:

1. **Feature selection:** filtering with extensive cross-validation processes on the training set based on several information measurements (e.g. Gini impurity, variance reduction with respect to target attribute)

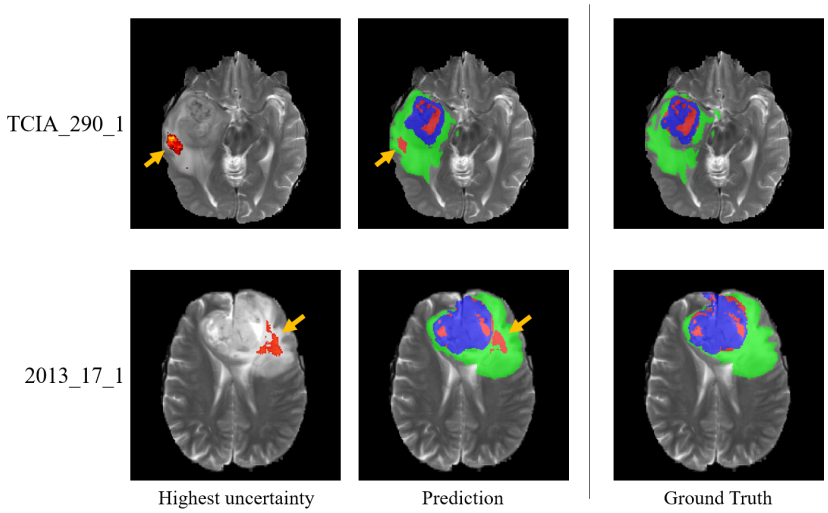


Fig. 5: Uncertainty-assisted correction for two subjects. (left) Uncertain regions that are used to locate erroneously segmented tissues (middle, right). The arrows point at the areas of interest.

Table 2: Comparison between the original B-FRRN segmentation and the uncertainty-assisted correction of two subjects with rather large uncertain areas (ET=enhancing tumor, WT=whole tumor, TC=tumor core).

Subject	Description	ET	WT	TC
TCIA_290_1	B-FRRN	0.899	0.905	0.846
	Corrected	0.899	0.905	**0.925**
2013_17_1	B-FRRN	0.939	0.929	0.916
	Corrected	0.939	0.929	**0.937**

2. **Construction of prediction models:** standard machine learning prediction models were trained, together with extensive cross-validation, using with subsets of the most informative features. The models used were:
   - SVM with RBF Kernel.
   - Neural Network with linear activation function.
   - Sparse Grids.

5 Discussion & Conclusion

The evaluation results (Table 1) reveal that the proposed B-FRRN architecture achieves results comparable to the standard FRRN and an overall improved segmentation performance compared to BraTumIA. Particularly the Dice coefficient and Hausdorff distance (95^{th} percentile) results are superior for the the B-FRRN

approach. Further, Figure 4 shows that the variation in the aforementioned metrics is unfailingly higher for BraTumIA compared to the FRRN and B-FRRN method. The similarity between the results of FRRN and B-FRRN demonstrate that the Bayesian Dropout approach is equivalent or slightly more performant then the standard approach.

The quantitative results of the second experiment show that a uncertainty-assisted correction can enhance the segmentation results. We could primarily observe large uncertain areas where edema is erroneously classified as non-enhancing tumor (shown in Figure 5). Correcting these errors led to an improved segmentation of the tumor core (Table 2). In the case where many small uncertain areas are detected, the correction is more time-consuming and the measurable gain marginal. In conclusion, the results showed good segmentation performance of the proposed network and first evidence suggests that uncertainty-assisted corrections can further improve the segmentation results. For survival prediction, the following observations were made

- Age was identified as the parameter containing more information regarding survival.
- Several image-based parameters were also significant features for survival prediction.
- Subsets of attributes containing age, surface and distance measurements provided the outstanding prediction results for the mentioned models.

Acknowledgments. This work was supported by the Swiss National Foundation by grant number 169607.

References

1. Alberts, E., Rempfler, M., Alber, G., Huber, T., Kirschke, J., Zimmer, C., Menze, B.H.: Uncertainty quantification in brain tumor segmentation using CRFs and random perturbation models. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). pp. 428–431. IEEE (apr 2016), http://ieeexplore.ieee.org/document/7493299/
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017), [In Press]
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (jan 2017), https://wiki.cancerimagingarchive.net/display/DOI/Segmentation+Labels+and+Radiomic+Features+for+the+Pre-operative+Scans+of+the+TCGA-LGG+collection
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (jan 2017), https://wiki.cancerimagingarchive.

net/display/DOI/Segmentation+Labels+and+Radiomic+Features+for+the+Pre-operative+Scans+of+the+TCGA-GBM+collection;jsessionid=C2BE9FB8F9D5532DCA9E5CD294787DBC
5. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning https://arxiv.org/pdf/1506.02142.pdf
6. Kendall, A., Badrinarayanan, V., Cipolla, R.: Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding (nov 2015), http://arxiv.org/abs/1511.02680
7. Lê, M., Unkelbach, J., Ayache, N., Delingette, H.: GPSSI: Gaussian process for sampling segmentations of images. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 9351, pp. 38–46. Springer (2015), http://link.springer.com/10.1007/978-3-319-24574-4{_}5
8. Meier, R., Knecht, U., Jungo, A., Wiest, R., Reyes, M.: Perturb-and-MPM: Quantifying Segmentation Uncertainty in Dense Multi-Label CRFs (mar 2017), http://arxiv.org/abs/1703.00312
9. Meier, R., Knecht, U., Loosli, T., Bauer, S., Slotboom, J., Wiest, R., Reyes, M.: Clinical Evaluation of a Fully-automatic Segmentation Method for Longitudinal Brain Tumor Volumetry. Scientific Reports 6, 23376 (mar 2016), http://dx.doi.org/10.1038/srep23376http://10.0.4.14/srep23376http://www.nature.com/articles/srep23376{\#}supplementary-information
10. Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, L., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Antonio Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S.J., Riklin-Raviv, T., Reza, S., Ryan, M., Schwartz, L., Shin, H.C., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Hye Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging p. 33 (2014), https://hal.inria.fr/hal-00935640
11. Pohlen, T., Hermans, A., Mathias, M., Leibe, B.: Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes (nov 2016), http://arxiv.org/abs/1611.08323
12. Porz, N., Bauer, S., Pica, A., Schucht, P., Beck, J., Verma, R.K., Slotboom, J., Reyes, M., Wiest, R.: Multi-Modal Glioblastoma Segmentation: Man versus Machine. PLoS ONE 9(5), e96873 (may 2014), http://dx.plos.org/10.1371/journal.pone.0096873
13. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. NeuroImage 31(3), 1116–1128 (jul 2006), http://www.ncbi.nlm.nih.gov/pubmed/16545965http://linkinghub.elsevier.com/retrieve/pii/S1053811906000632

Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation

K. Kamnitsas, W. Bai*, E. Ferrante*, S. McDonagh*, M. Sinclair*
N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, D. Rueckert, B. Glocker

Biomedical Image Analysis Group, Imperial College London, UK
* Equal contribution, in alphabetical order

Abstract. Deep learning approaches such as convolutional neural nets have consistently demonstrated to outperform previous methods on challenging tasks such as dense, semantic segmentation. However, different models perform differently, with behaviour largely influenced by architectural choices and training settings. This paper explores ensembles of multiple models and architectures (EMMA) for robust performance through aggregation of predictions from a wide range of methods, and thus reducing the influence of the meta-parameters of individual models and the risk of overfitting the configuration to a particular database. EMMA can be seen as an unbiased, generic deep learning model which is shown to yield excellent performance on the BRATS 2017 challenge.

1 Introduction

Brain tumours are among the most fatal types of cancer [1]. Among the tumours that originally develop in the brain, gliomas are the most frequent [2]. They arise from glioma cells and, depending on their aggressiveness, they can be broadly categorized into high and low grade gliomas [3,4]. High grade gliomas (HGG) develop rapidly and aggressively, forming abnormal vessels and often a necrotic core, accompanied by surrounding oedema and swelling [5,2]. They are highly malignant, commonly leading to patient death in less than two years even after treatment [4]. Low-grade gliomas can be benign or malignant, grow slower, but they may recur and evolve to high grade gliomas, thus their treatment is also warranted. For treatment of brain tumours, patient undergo radiation, chemotherapy and surgery [1].

Firstly for diagnosis and monitoring the tumour's progression, then for treatment planning, and further for assessing the effect of treatment strategy, various neuro-imaging protocols are employed. Magnetic Resonance imaging (MRI) is widely employed in both clinical routine and research studies, as it facilitates tumour analysis by allowing estimation of its extent, its location and investigation of its subcomponents [2]. This however requires accurate delineation of the tumour in the images, which proves challenging due to its complex structure and appearance, the 3D nature of the MR images and the multiple MR sequences that need to be consulted in parallel for informed judgement. These factors not

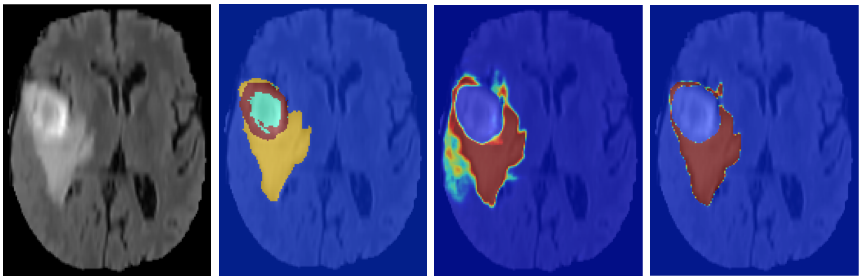
only make manual delineation time-consuming, but also subject to significant inter- and intra-rater variability [6].

Automatic segmentation systems aim to provide an efficient, objective and scalable solution. Representative early works are the atlas-based outlier detection method of [7] and the joint segmentation-registration framework, often guided by a tumour growth model [8,9,10]. The past few years saw rapid developments of machine learning methods, with Random Forests being among the most successful [11,12,13]. Individual decision trees exhibits low classification bias but high variance. By averaging over the predictions of multiple trees, de-correlated by training them with different subsets of the input features, random forests "average away" the variance without increasing the bias, which results to their good generalization. More recently, deep learning based methods, mostly represented by convolutional neural networks (CNN), have shown very promising results for segmentation of brain tumour [14,15,16]. Their power is commonly attributed to the automatic learning of data-driven, task-specific feature detectors during the optimization.

A large variety of CNN architectures has been proposed in the recent literature. At the same time, these models typically have a vast number of meta parameters that need configuration and which greatly influence performance. When one tries to study the behaviour of CNNs on a given task with respect to a certain factor, such a pre- or post-processing technique, the results can be biased by the choice of the architecture and the rest configuration. Additionally, a configuration highly optimized on a given database may have over-fitted it and not generalize well to other data or tasks. In this work we push towards constructing a more *generic deep learning model*. Inspired by the paradigm of random forests, we bring together a multitude of different CNN architectures, configured and trained in different ways in order to achieve high variance between their results. By bagging together all models, we construct an *Ensemble of Multiple Models and Architectures* (EMMA), with the aim of *averaging away* model- and configuration-specific behaviours, which then allows studying the behaviour of this unbiased *generic deep learning model*. This is in contrast to the commonly used ensembles of the same architecture, trained with small variations such as initial seeds, which results in highly correlated outputs, largely defined by the main architectural choices. In this preliminary work, we evaluate EMMA on the validation data of Brain Tumour Segmentation (BRATS) challenge 2017 [28,29,30,31], where we achieve top performance among 45+ competing teams. This indicates the quality of the model and paves the way for its use in further analysis.

2 Ensembles of Multiple Models and Architectures

A large number of CNN architectures have been proposed in the literature and shown promising for the segmentation task. Regarding the architectures, models commonly differ in depth, number of filters and the way they process multi-scale context among others. Such architectural choices influence not only the overall



(a) (b) (c) (d)

Fig. 1: (a) FLAIR and (b) manual annotation of subject 2013_10_1 in training database of BRATS 2017, depicting oedema (yellow), enhancing (red), non enchancing/necrotic core (light blue). (c) Confidence map for the oedema class of a network trained to minimize cross-entropy or (d) maximize Intersection over Union (IoU) loss. The choice of loss influences the model's behaviour. IoU and Dice (not depicted) losses tend to binarize the decision boundaries. The models were not trained on the depicted subject.

performance of the model, but also its behaviour. For instance, models with large receptive fields tend to be influenced less by small details in the image, and can present improved localisation capabilities. However they may be less precise in following image details than models emphasizing only local information.

Another key aspect that heavily influences a model's performance as analysed in [16] is the class imbalance in a task, appropriate handling of which can make the difference between low or state of the art performance. This issue is commonly tackled by employing class-weighted sampling during training or using a loss function that weights samples appropriately. However weighted sampling or class-weighted loss influences the bias and sensitivity of the model to each class, defining its behaviour. Other loss functions change the model's behaviour in different manners. For example, it was observed (Fig. 1) that networks trained with cross-entropy, an information-theoretic loss, tend to present softer class-confidence maps, in comparison to networks that optimize Intersection over Union (IoU), Dice or similar losses [17]. This can lead to highly confident but wrong predictions.

Finally, the choice of different hyper-parameters for the optimization and regularization can heavily affect the performance of a model. It is often observed by practitioners that the choice of optimizer and its configuration, for instance its learning rate, may make the difference between bad and good segmentation. The sensitivity to all these meta-parameters is a great practical problem, magnified by the fact that the same meta-parameter setting is not guaranteed to be well behaved among different network architectures, or even on different tasks and data. As a result, it is often difficult to draw generic and confident conclusions without spending a considerable amount of time in optimizing the experimental settings.

We build a model that is less biased by such architectural choices and more robust to suboptimal configuration by ensembling different state-of-the-art architectures, configured and trained under widely different settings. We introduce vast difference between models, with the aim of bagging singletons that are not strongly correlated. For example, sensitive models may synergize well with specific models, correcting or supporting each others predictions, with such an ensemble behaving better than a fusion of sensitive-only models. This way EMMA, the resulting *Ensemble of Multiple Models and Architectures* is a collection of models with low bias but high variance. By averaging over the outputs of all models, the variance is reduced, random unbiased errors from individual models are smoothed out, and the final output is more representative of a *generic CNN model*.

In the following we describe the individual models used to construct EMMA.

2.1 DeepMedic

Model description: The first network architecture we employ is deepMedic, originally presented in [18,16]. It is a fully 3D, multi-scale convolutional network, designed with a focus on efficient processing of 3D images. To achieve this, it utilizes parallel convolutional pathways that take as input context of the image at different scales. Pathways operating at lower scales avoid convolving large volumes at full resolution and thus remain computationally cheap, while they process information from large spatial extent. Although originally developed for segmentation of brain lesions, deepMedic was found promising on diverse tasks, such as segmentation of the placenta from motion corrupted MR ([19]), which makes it a good candidate for an ensemble that aims to be generic and robust. We train two different deepMedic models to construct EMMA. The first is the residual version of the model, as previously employed in BRATS 2016 [20] and depicted in Fig. 2. The second model is a wider variant, where the number of feature maps at each layer were doubled.

Training details: The models were trained by extracting multi-scale segments of the image with a 50% probability centred on healthy tissue and 50% probability centred on tumour. As previously analysed in [16], this strategy allows the class-balance to be implicitly regulated by the size of the sampled lesion and the size of input, which was found to behave well in various tasks. The networks are trained with cross-entropy, with all meta-parameters adopted from the original configuration.

2.2 FCN

Model description: We employ 3D fully convolutional networks (FCN) [21] as components of the ensemble. In its encoding part, FCN performs a number of convolutions to extract image features and down-samples the feature maps after every few convolutions to increase the receptive field and learn features at a more global scale. However, in order for the segmentation to be at high resolution,

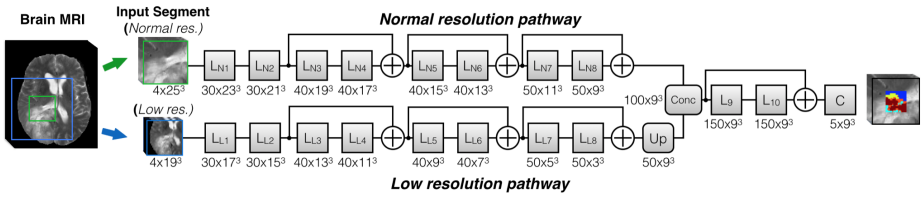


Fig. 2: We used two DeepMedics [16] in our experiments. The smaller of the two is depicted, where the number of feature maps and their dimension at every layer are depicted in the format ($Number \times Size$). The second model used in the ensemble is wider, with double the number of feature maps at every layer. All kernels and feature maps are 3D, even though not depicted for simplicity.

down-sampling has to be reversed, which is done in the decoder of the network. FCN performs this by creating multiple segmentation maps from features of different scales. Segmentation maps from lower scales are up-sampled and then fused with the segmentations from higher resolutions. In our implementations down-sampling is performed via convolution with stride 2 and up-sampling via bilinear interpolation.

We experiment with two FCN architectures. FCN-VGG, depicted in Fig. 3, uses an adaptation of the VGG network [22] as its encoder. The segmentation maps from each scale are up-sampled straight to high resolution, are concatenated and finally fused into the final classification. The second one, FCN-ResNet, was constructed by doubling the number of feature maps at every convolutional layer and adding residual connections [23]. Additionally, the decoder of this architecture follows the original FCN [21], with segmentations from lower scales getting gradually upsampled, first to the resolution of the previous scale, get fused via element-wise addition with the segmentations from that scale, and then are up-sampled further. In both architectures, all kernels in the down-sampling part of the network are of size 3×3×3 and followed by batch normalization and ReLUs. Zero-padding is used at every layer to equalize input and output dimensions.

Training details: We randomly draw 64×64×64 patches for training, with an equal probability from each label. We train 3 FCNs for EMMA. First, an FCN-VGG is trained via maximizing the mean Dice coefficient across label classes as the loss. Two similar FCN-ResNets are trained, one optimized by maximizing the Dice loss and one by maximizing the IoU loss. Optimization is performed using Adam [24] with a learning rate of 0.001.

2.3 U-Net

Model description: We employ a modified 3D version of the original U-Net [25] architecture as another component of the ensemble. Our U-Net model consists of a contractive path formed by max-pooling and convolutional layers, followed

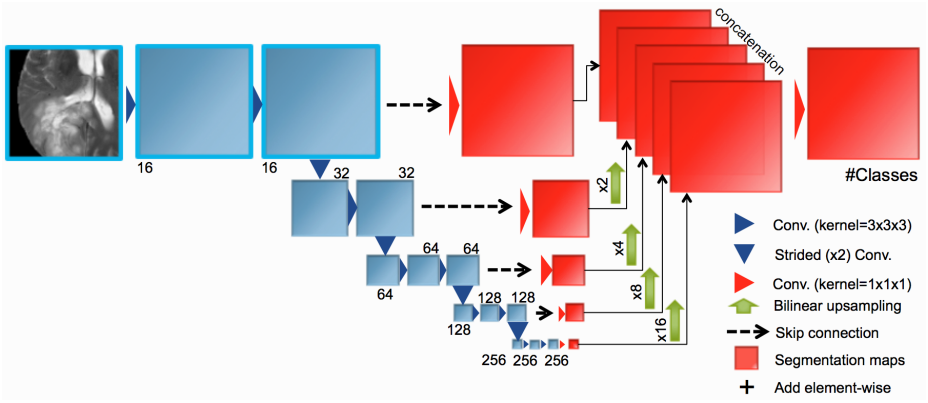


Fig. 3: Our implementation of FCN-VGG, one of the two FCN architectures used in EMMA. Depicted are the number of feature maps at every layer. All kernels and feature maps are 3D, even though not depicted for simplicity.

by an expansive phase where the original resolution is achieved through up-convolutions (up-sampling of the feature maps with repetition, followed by convolutional layers) (Fig. 4). Similarly to [25], we employ skip-connections to incorporate information from the corresponding feature maps (from the contracting path) into the expansive phase. However, instead of concatenating these features as proposed by [25], we follow the strategy suggested in [26] replacing concatenations with summations to reduce model complexity. Batch normalization was used after every hidden convolutional layer. Dropout, at a rate of 50% was used after the last convolution layer of the contracting path, and after the second and fourth convolution of the expansive path. Appropriate padding is used at every layer to equalize input and output dimensions.

Training Details: The U-Net was trained with patches of size 64×64×64. The patches were sampled with an equal probability being centred around a voxel from each label, with background as an extra class. Only voxels within the brain mask were considered. We used stochastic gradient descent for optimization with AdaDelta [27] as strategy to adapt the learning rate. The initial learning rate was set to 0.1 while the AdaDelta decay rate was set to 0.95. We used mini-batches of size 8, categorical cross-entropy for the loss function and weight decay (L^2 weighting factor = 0.00001) for regularization. We considered two data augmentation strategies: (1) patches where randomly flipped along the left-right axis in the axial plane and (2) Gaussian noise (μ=0, σ=0.01) was added to the patches while training.

2.4 Ensembling

The two DeepMedics, the three FCN models and the Unet are trained completely separately. At testing time, each model segments individually an unseen image

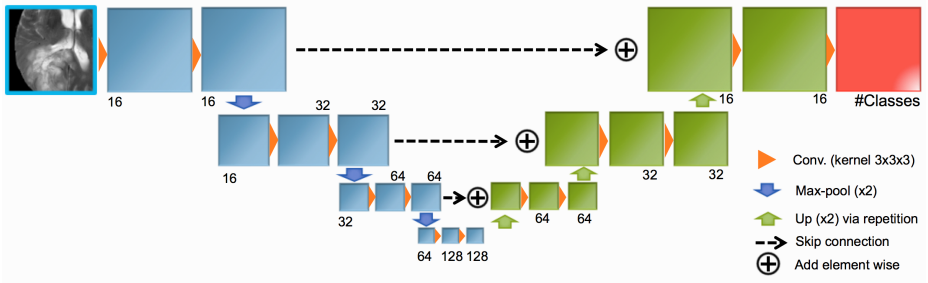


Fig. 4: Schematic of the adapted Unet used in our experiments. Depicted are the number of feature maps at every layer. All kernels and feature maps are 3D, even though not depicted for simplicity.

and outputs its normalized class-confidence maps. The models are then ensembled into EMMA. The ensemble's confidence maps for each class are created by calculating for each voxel the average confidence of the individual models for the voxel to belong to this class. The final segmentation made by the EMMA is performed by assigning to each voxel the class with the highest confidence.

2.5 Implementation details

The original implementation of DeepMedic was used for the corresponding two models, along with the default meta-parameters, publicly available on `https://biomedia.doc.ic.ac.uk/software/deepmedic/`. The FCN-ResNet models were implemented using DLTK, a deep learning library with a focus on medical imaging applications that allowed quick implementation and experimentation (`https://github.com/DLTK/DLTK`). Finally, our adaptation of the Unet will be released on `https://gitlab.com/eferrante`.

3 Evaluation

3.1 Material

Our system was evaluated on the data from the 2017 Brain Tumour Segmentation Challenge (BRATS) [28,29,30,31]. The training set consists of 210 cases with high grade glioma (HGG) and 75 cases with low grade glioma (LGG), for which corresponding manual segmentations are provided. The segmentations include the following tumour tissue labels: 1) necrotic core and non enhancing tumour, 2) oedema, 4) enhancing core. Label 3 is not used. The validation set consists of 46 cases of both HGG and LGG but the grade is not revealed. Reference segmentations for the validation set are hidden and evaluation is carried out via an online system. For evaluation, the 3 predicted labels are merged into different sets of whole tumour (all labels), the core (labels 1,4) and the enhancing tumour (label 4). For each subject, four MRI sequences are available, FLAIR, T1, T1

contrast enhanced (T1ce) and T2. The datasets are pre-processed by the orgazers and provided as skull-stripped, registered to a common space and resampled to isotropic $1mm^3$ resolution. Dimensions of each volume are $240 \times 240 \times 155$.

3.2 Pre-processing

We apply only minimal pre-processing. Each MR sequence and separately for each subject, is individually normalized by subtracting from each voxel the mean intensity of brain tissue in that image. We then divide each voxel with the standard deviation of the intensities within the brain area of the image.

3.3 Results

The six individual models described in Sec. 2 were trained on the whole training database individually. They were then applied on the blinded validation database and their predicted class-confidence maps were fused into EMMA predictions. EMMA's segmentations were submitted for online evaluation. Despite that the individual models were not highly optimized, EMMA achieved highest scores among multiple competing entries. We account this on the wide variation between the ensembled models, which present significantly different types of mistakes and behaviours, but when fused, they correct each other to form a low variance, unbiased system (Fig. 5). Table 1 shows the achieved performance, along with the next two top performing entries at the time of manuscript submission.

Table 1: Average performance of EMMA on the blinded validation data of BRATS 2017 as computed on the online evaluation platform and comparison to the two highest performing teams out of the 45 entries visible at the time of manuscript submission. Numbers in bold indicate best performance.

	DSC			Sensitivity			Specificity			Hausdorff_95		
	Enh.	Whole	Core	Enh.	Whole	Core	Enh	Whole	Core	Enh.	Whole	Core
biomedia1 (EMMA)	**75.7**	**90.2**	82.0	**79.0**	90.9	78.3	99.8	99.5	**99.9**	**4.22**	4.56	**6.11**
UCL-TIG	75.2	89.7	**82.5**	77.1	**91.2**	**83.9**	99.8	99.4	99.7	4.78	**3.97**	7.60
MIC_DKFZ	73.2	89.6	79.7	**79.0**	89.6	78.1	99.8	**99.6**	**99.9**	4.55	6.97	9.48

4 Conclusion

This paper has introduced EMMA, an ensemble of widely varying CNN architectures, configured and trained under very different settings. By bagging such heterogeneous collection of networks we aim at constructing a deep-learning model that is less influenced by individual architectural choices as well as robust

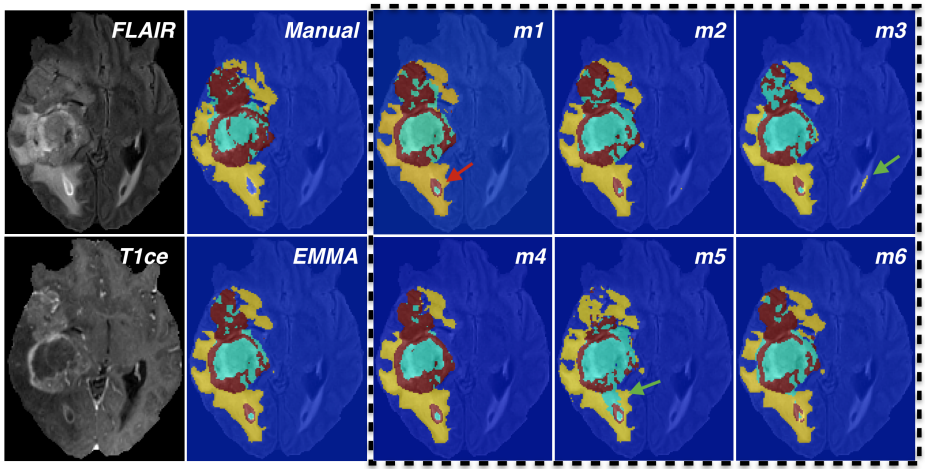


Fig. 5: FLAIR, T1ce and manual annotation for subject CBICA_ABO_1 of the training database. Also depicted are predictions of individual models, obtained in preliminary experiments with two-fold validation on the training database, and their ensembling into EMMA. Green arrows point inconsistent mistakes that are corrected by the ensembling, while red arrow shows a consistent mistake.

to suboptimal configuration, in order to enable more representative and unbiased analysis. Even though the individual networks were not highly optimized for the tumour segmentation task, EMMA achieved top performance in the validation stage of the BRATS 2017 challenge, indicating its robustness, which will be further evaluated in the testing stage of the challenge. This also indicates that diverse ensembles such as EMMA may also offer strong performance on different problems without extensive meta-parameter search, which we aim to explore in the future, and which would enable more straightforward and confident analysis.

Such generic ensemble models could be useful to explore effect of various factors, such as different data pre-processing techniques or sufficient number of training data, less biased by individual network choices. A particularly important practical issue to investigate is whether EMMA is more robust to the domain shift between the training and testing data, which is strongly affecting large scale multi-center studies and commonly reduces performance of individual networks when no domain adaptation methods are employed [32]. Finally, one could investigate the uncertainty of the ensemble, represented by the entropy in the predictions of the individual models, which could serve as a better confidence indication than the normalized scores of the individual models [33], the behaviour of which is highly influenced by training settings (Fig. 1). Considering uncertainty measures that are not model-specific could reveal what type of patients or tumours are most challenging for deep-learning based systems to learn.

5 Acknowledgements

This work is supported by the EPSRC (EP/N023668/1, EP/N024494/1, EP/P001009/1) and partially funded under the 7th Framework Programme by the European Commission (CENTER-TBI: https://www.center-tbi.eu/). KK is supported by the Presidents PhD Scholarship of Imperial College London. NP is supported by Microsoft Research through its PhD Scholarship Programme and the EPSRC Centre for Doctoral Training in High Performance Embedded and Distributed Systems (HiPEDS, Grant Reference EP/L016796/1). We gratefully acknowledge the support of NVIDIA Corporation with the donation of GPUs for our research.

References

1. DeAngelis, L.M.: Brain tumors. New England Journal of Medicine **344**(2) (2001) 114–123
2. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of mri-based medical image analysis for brain tumor studies. Physics in medicine and biology **58**(13) (2013) R97
3. Kleihues, P., Burger, P.C., Scheithauer, B.W.: The new who classification of brain tumours. Brain pathology **3**(3) (1993) 255–268
4. Louis, D.N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., Ellison, D.W.: The 2016 world health organization classification of tumors of the central nervous system: a summary. Acta neuropathologica **131**(6) (2016) 803–820
5. Giese, A., Bjerkvig, R., Berens, M., Westphal, M.: Cost of migration: invasion of malignant gliomas and implications for treatment. Journal of clinical oncology **21**(8) (2003) 1624–1636
6. Mazzara, G.P., Velthuizen, R.P., Pearlman, J.L., Greenberg, H.M., Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated mri segmentation. International Journal of Radiation Oncology Biology Physics **59**(1) (2004) 300–312
7. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical image analysis **8**(3) (2004) 275–283
8. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2011) 532–540
9. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Joint tumor segmentation and dense deformable registration of brain mr images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2012) 651–658
10. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: Glistrboost: combining multimodal mri segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer (2015) 144–155

11. Criminisi, A., Shotton, J., Konukoglu, E., et al.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision **7**(2–3) (2012) 81–227
12. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2012) 369–376
13. Le Folgoc, L., Nori, A.V., Ancha, S., Criminisi, A.: Lifted auto-context forests for brain tumour segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer (2016) 171–183
14. Urban, G., Bendszus, M., Hamprecht, F., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. in proc of BRATS-MICCAI (2014)
15. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging **35**(5) (2016) 1240–1251
16. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis **36** (2017) 61–78
17. Nowozin, S.: Optimal decisions from probabilistic models: the intersection-over-union case. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2014) 548–555
18. Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., Glocker, B.: Multi-scane 3d convolutional neural networks for lesion segmentation in brain mri. in proc of ISLES-MICCAI (2015)
19. Alansary, A., Kamnitsas, K., Davidson, A., Khlebnikov, R., Rajchl, M., Malamateniou, C., Rutherford, M., Hajnal, J.V., Glocker, B., Rueckert, D., et al.: Fast fully automatic segmentation of the human placenta from motion corrupted mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2016) 589–597
20. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: Deepmedic for brain tumor segmentation. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer (2016) 138–149
21. Long, J., et al.: Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2015) 3431–3440
22. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR. (2015) 1–14
23. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 770–778
24. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations. (2015)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2015) 234–241

26. Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., Wolz, R., Valdes-Hernandez, M., Dickie, D., Wardlaw, J., et al.: White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks. arXiv preprint arXiv:1706.00935 (2017)
27. Zeiler, M.D.: Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701 (2012)
28. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10) (2015) 1993–2024
29. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017) [In Press]
30. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. The Cancer Imaging Archive (2017) DOI:10.7937/K9/TCIA.2017.KLXWJJ1Q.
31. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. The Cancer Imaging Archive (2017) DOI:10.7937/K9/TCIA.2017.GJQ7R0EF.
32. Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al.: Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: International Conference on Information Processing in Medical Imaging, Springer (2017) 597–609
33. Gal, Y.: Uncertainty in deep learning. PhD thesis, PhD thesis, University of Cambridge (2016)

Radiomics-based Convolutional Neural Network (RadCNN) for Brain Tumor Segmentation on Multi-parametric MRI

Ayush Karnawat*, Prateek Prasanna*, Anant Madabushi, and Pallavi Tiwari

Case Western Reserve University, Cleveland OH 44106, USA,
axk840, pxp238, axm788, pxt130@case.edu *

Abstract. Accurate segmentation of gliomas on routine magnetic resonance image (MRI) scans plays an important role in disease diagnosis, prognosis, and patient treatment planning. In this paper, we present a fully automated approach, radiomics-based convolutional neural network (RadCNN), for segmenting both high and low grade gliomas using multi-modal MRI volumes (T1c, T2w, FLAIR). RadCNN incorporates radiomic texture features (i.e. Haralick, Gabor, Laws) within DeepMedic (a deep 3D CNN framework; a top performing method in the BraTS 2016 challenge) to further augment the performance of brain tumor sub-compartment segmentation. We first identify textural radiomic representations that best separate the different compartments (enhancing tumor, whole tumor, and tumor core) and then feed these representations as inputs to the CNN classifier for prediction of different sub-comartments. We hypothesize that textural radiomic representations of lesion sub-compartments will enhance the seperation of sub-compartment boundaries, and hence providing these features as inputs to the deep CNN, over and above raw intensity values alone, will improve the sub-compartment segmentation. Using a training set of N=241 patients, validation set of N=44, and test set of N=46 patients, RadCNN method achieved Dice Similarity Coefficient scores (DSC) of 0.71, 0.89, 0.73 for enhancing tumor, whole tumor, and tumor core, respectively. Compared to the DeepMedic model, RadCNN showed improvement in DSC scores for both the enhancing and whole tumors, and demonstrated comparable results in segmenting the tumor core. Similarly, smaller Hausdroff distance measures were obtained with RadCNN as compared to the DeepMedic model, across all the sub-compartments. [1]

Keywords: Gliomas, Segmentation, Radiomics, Feature selection, CNN

1 Introduction

Gliomas, one of the most common types of primary brain tumors, exhibit phenotypically heterogeneous sub-regions comprising of the enhancing and non-enhancing lesion, necrotic core, and the surrounding edema, each of which contains relevant diagnostic and prognostic information [1]. As such, accurate estimation of the information (i.e. volume/position) contained within these regions is critical for diagnosis, treatment planning, and long-term survival assessment. Towards that end, accurate delineation of the tumor and its sub-regions is required, which, due to their variable shape and size, poses a significant challenge. In particular, manual segmentation of the region boundaries is both time-consuming and prone to misinterpretation and human error, often resulting in high inter-rater variability [2].

Accurate, automatic segmentation frameworks aim to solve this problem, while providing a more efficient and scalable solution for clinical applicability. In recent years, convolutional neural networks (CNN) have drawn increasing attention for problems involving classification and semantic segmentation, especially in the field of image recognition. In fact, leading methods from previous years of the brain tumor segmentation (BRATS) challenge [3,4,5] have consistently employed CNN-based architectures. However, ability of hand-crafted radiomic features in conjunction with deep CNN networks have been largely unexplored. Radiomic textural features allow for capture of higher

* Research was supported by 1U24CA199374-01, R01CA202752-01A1, R01CA208236-01A1, R21CA179327-01, R21CA195152-01, R01DK098503-02, 1C06-RR12463-01, PC120857, LC130463, the DOD Prostate Cancer Idea Development Award, W81XWH-16-1-0329, the Case Comprehensive Cancer Center Pilot Grant, VelaSano Grant from the Cleveland Clinic, I-Corps program, Ohio Third Frontier Program, and the Wallace H. Coulter Foundation Program in the Department of Biomedical Engineering at Case Western Reserve University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[1] *Equal contribution

order quantitative measurements (e.g. co-occurrence matrix homogeneity, neighboring gray-level dependence matrix, multi-scale Gaussian derivatives), for modeling macro and micro-scale morphologic attributes within and around the lesion area from across different MRI protocols. Texture analysis has previously shown promise in distinguishing different grades of brain tumors [6] as well as identifying brain tumors from treatment confounders [7]. In fact, texture features have previously been used in conjunction with Random Forests [8] to segment brain tumor lesions. In this work, we hypothesize that textural radiomic representations of lesion sub-compartments will enhance the seperation of sub-compartment boundaries, and hence providing these features as inputs to the deep CNN, over and above raw intensity values alone, will improve the sub-compartment segmentation.(Figure 1).

In this paper, we present radiomics-based CNN (RadCNN), a multi-scale convolutional neural network architecture that incorporates optimized radiomic texture features as an input to a deep CNN model, for improved estimation of tumor sub-compartments: enhancing tumor, non-enhancing tumor, necrosis, and edema. In the following sections, we illustrate our approach and describe, in detail, the features of our architecture followed by its application to the segmentation of gliomas in brain tumors.

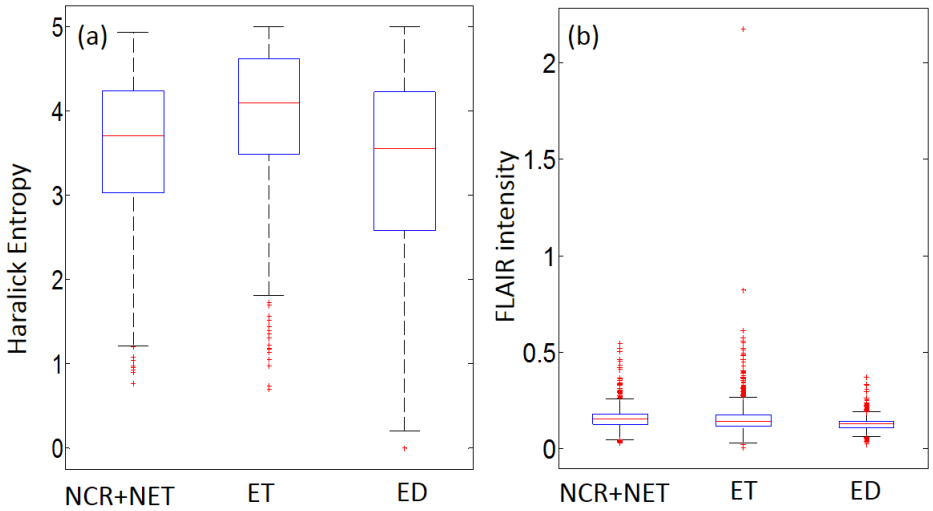


Fig. 1. Box plots showing distribution of (a) Haralick Entropy features and (b) FLAIR intensities in randomly chosen voxels across cases with poor DSC using an intensity-only CNN. The separation in radiomic texture representation across different sub-compartments serves as the motivation for using radiomic features as input channels to the CNN model.

2 Methodology

Figure 2 below shows the workflow of our presented RadCNN model. Briefly, testural radiomic features are first extracted from the multi-parametric MRI scans. This is followed by selection of the most relevant features for differentiating the various sub-compartments. Finally, the best features, in addition to the raw intensity channels, are provided as input to a multi-resolution CNN for multi-class classification.

2.1 Feature Extraction

For each 3D MRI volume, we choose to primarily extract textural radiomic features including Gabor, Haralick, and Laws due to their ability to capture micro, and macro-level morphologic attributes relating to intensity, edges, and gradient-specific differences across different compartments on routine MRI scans (T1c, T2w, FLAIR) (Figure 1). A Gabor filter can be defined as the modulation of a complex sinusoidal by a Gaussian function and is controlled by scale λ and orientation θ parameters. Gabor features, which are modeled according to human visual perception, are

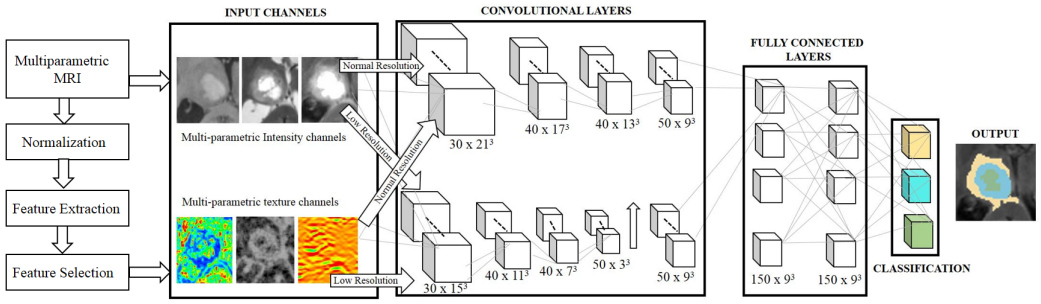


Fig. 2. RadCNN pipeline: Workflow of our proposed framework, which comprises two stages. First, we extract and select the top features. We then use the selected texture map volumes, along with the normalized multi-parametric MRI (T1c, T2w, FLAIR) scans as channel inputs to a 3D CNN for classification.

extracted as a response to convolution of an image with distinct Gabor filters obtained by varying each of the associated parameters across the filter bank. Haralick features capture gray-level co-occurrence patterns, where a matrix of co-occurring gray-level pairs in the image is constructed, from which second-order statistical texture features can be derived. Second order intensity statistics such as angular second moment, contrast, and difference entropy are used to characterize the MRI images. Laws features use 5×5 separable masks that are symmetric or anti-symmetric to extract level (L), edge (E), spot (S), wave (W), and ripple (R) patterns on an image. The convolution of these masks with every image yields distinct Laws features. In particular, we computed 40 Gabor filter responses with varying $\lambda = 2, 4, 8, 16, 32$ and $\theta = 0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°$ values, 13 Haralick features, and 25 Laws features. In total, we extracted 78 radiomic features for each sequence (T1w, T2, FLAIR), resulting in a total of 234 texture features per study.

2.2 Feature Selection

In order to create a well-defined discriminative model, our goal was to select only the features that would improve the segmentation of the sub-compartments that were getting over, or under-segmented using intensity-based CNN (i.e. DeepMedic model). Hence, our feature selection experiments were driven by the output of the DeepMedic model during the training stage. We observed the following consistent trends while training the cases with the DeepMedic model, (a) oversegmentation of edema regions, and (b) undersegmentation of necrosis+non-enhancing regions. We hence designed two separate feature selection experiments to address the aforementioned problems, where our primary goal was to identify the best texture features to (a) distinguish edema from background (non brain tissue) voxels, and (b) distinguish necrosis+non-enhancing regions from the rest. We used a minimum redundancy maximum relevance (mRMR) algorithm in conjunction with a Random Forest classifier, and evaluated the importance of each of the 234 feature vectors in a 3-fold cross-validation setting on a random subset of texture voxels from the training dataset.

In each of the 100 iterations, we assigned weights to the selected features based on their rank of occurrence. The cumulative weights over 100 runs are then used to select the best features. The most unimportant features were then eliminated, and the top 2 features for each classification experiment were retained. The top features included Haralick entropy, energy, inverse difference moment, and correlation co-occurrence statistics.

2.3 3D Convolutional Neural Network (CNN)

The CNN portion of our pipeline is based on the DeepMedic framework by Kamnitsask et al. [3], which has been shown to provide the best-performing automated brain sub-compartment segmentations in BRATS'16 challenge benchmark datasets. The network consists of an 11-layer deep multi-scale 3D CNN consisting of two parallel convolutional pathways that processes the input at both a normal resolution and one at a lower scale to achieve a large receptive field for classification and segmentation [3].

The CNN architecture, as seen in Figure 2, comprises 11 layers: eight consecutive convolutional-pooling layers followed by two fully connected layers and one classification layer along two pathways. Each convolutional-pooling layer uses the same fixed $3 \times 3 \times 3$ convolutional kernel and $2 \times 2 \times 2$

	DSC	Sensitivity	Specificity	Hausdroff
DeepMedic	(0.703, 0.880, 0.732)	(0.747, 0.891, 0.712)	(0.998, 0.994, 0.997)	(5.689, 11.993, 11.776)
RadCNN	(0.710, 0.890, 0.732)	(0.749, 0.891, 0.697)	(0.998, 0.995, 0.998)	(5.244, 6.533, 10.065)

Table 1. Performance of RadCNN model (trained on intensities + radiomics features as inputs) compared with DeepMedic (trained on intensities alone). Each 3-tuple represents the respective average scores for (enhancing tumor (ET), whole tumor (WT), and tumor core (TC)) across the test cohort (N =46).

pooling kernel with 30, 30, 40, 40, 40, 40, 50, and 50 neurons, respectively. Both fully connected layers have 150 neurons which are connected to the four final neurons to determine each voxel's region subtype. We utilized the Adam optimizer which has been shown to work well in practice and compares favorably to other adaptive learning-method algorithms [10]. Most importantly, our network includes additional input channels in the form of selected textural radiomic maps extracted from the original multi-parametric MRI scans.

To measure the value of preselecting certain low-level handcrafted features as input to a deep 3D CNN framework, we quantitatively compared the performance of our model against a standard DeepMedic model trained on multi-parametric intensities by modifying the number of input channels into the network. The validation metric used is the Dice similarity coefficient (DSC), a spatial overlap index.

3 Experiments and Results

3.1 Data and preprocessing

The training dataset provided by BraTS '17 challenge [1,11,12,13], consists of 210 multi-modal MRI (T1, T1c, T2w, FLAIR) scans of patients with high- (HGGs) and 75 patients with low-grade gliomas (LGGs). The images were skull-stripped, co-registered to a common space, and re-sampled to a 1 mm^3 voxel resolution, with the final dimensions of each volume being $240 \times 240 \times 155$ voxels. Each volume was normalized by subtracting the mean and dividing by the standard deviation of the intensities. These were also affine-aligned to the same space to ensure consistency between the data. The ground truth provided for each voxel consisted of four sub-regions: non-tumor, necrosis + non-enhancing tumor (NCR + NET), edema (ED), and enhancing tumor (ET).

Similarly, the images within the validation dataset, which included about 46 total cases were also pre-processed using the same pipeline. The ground-truth was not included for evaluation. Instead, the evaluation was performed on the online CBICA portal provided by the University of Pennsylvania [14] with the official results being obtained by combining the predicted segmentation into three labels: Whole Tumor, WT (NCR + NET + ED + ET), Tumor Core, TC (NCR + NET + ET), and just the ET.

3.2 Preliminary Results

To evaluate the impact of using textural radiomic maps in conjunction with intensity-only scans as inputs to a CNN model (RadCNN versus DeepMedic), we first trained, validated, and tested the DeepMedic CNN on a subset of the HGG and LGG cases from the BraTS '17 training and validation dataset, using the T1c, T2w, FLAIR protocols. We then evaluated RadCNN on the same data (i.e using the same cases and parameters for training, validation, and testing). Table 1 shows the results gathered from the online evaluation platform [14] for both models. We also compared the DSC values of the RadCNN model with those obtained from the other models in the BRATS 2016 challenge using the corresponding 2016 dataset. It may be observed from Table 2 that inclusion of radiomic features resulted in an increase in the performance across most of the top-performing measures.

Compared to the intensity-based DeepMedic model, the proposed pipeline showed an improvement to both the enhancing and whole tumor DSC and comparable results in segmenting the tumor core. Although the specificity of the TC does decrease slightly when texture features are incorporated, the Hausdorff distance, a measure of how far two topographical objects are from each other, between the prediction and the ground truth decreases quite substantially (about 50%

	Whole Tumor	Tumor Core	Enhancing
Chang	0.87	0.81	0.72
Dera	0.91	0.91	0.84
Krishnamurthi	0.84	0.71	0.81
Randhawa	0.87	0.75	0.81
Song	0.86	0.70	0.73
Vilaplana	0.89	0.76	0.37
Zeng	0.85	0.82	0.80
Zhao	0.87	0.82	0.76
DeepMedic	0.90	0.75	0.72
RadCNN	0.90	0.82	0.80

Table 2. Comparison of top performing methods from BraTS 2016 challenge with RadCNN using the BraTS '16 Training Dataset (highlighted in bold).

for the whole tumor, and 14% for the tumor core). This suggests that the addition of pre-selected textural radiomic maps results in an overall improvement in the predicted segmentations. Figure 3 shows an example of the improvement in sub-compartment segmentations using RadCNN, when compared to using just DeepMedic.

To compare the effects of selecting texture features on each individual label, we also calculated the percentage of under- and over-segmentation of each label on a per-patient basis. With the addition of texture features, the under-segmentation percentages of the each label tend to decrease overall. On average, for the cross-validation cases, this error rate decreases by 1% (20% vs. 19%) for the ED label, and shows comparable results with respect to the NCR + NET, and ET labels.

4 Discussion and Conclusion

In this paper, we presented RadCNN, a radiomics-based convolutional neural network approach to improve brain sub-comartment segmention, by providing optimized radiomic representations of the multi-parametric MRI scans as inputs to a 3D-CNN classifier. Our results suggest that radiomic features, owing to their ability to provide complementary phenotypic information [15], in conjunction with 3D CNN can augment lesion segmentation performance. We are working on further optimization of input channels that will incorporate additional low-level hand-crafted features to help improve segmentation predictions. These newer features will be evaluated and compared to our current results to create a more optimized feature set. Furthermore, we will apply post-processing techniques that combine segmentations from different models to create more robust segmentation predictions. As a part of the challenge, we will subsequently use an optimized set of radiomic measurements from the tumor sub-compartments to build a prognostic model, which will be evaluated on the test set.

References

1. Menze, BH, Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., others: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging 34(10), 1993-2024 (2015)
2. Mazzara, G. P, Velthuizen, R. P, Pearlman, J. L, Greenberg, H. M, Wagner, H.: Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. International Journal of Radiation Oncology* Biology* Physics 59(1), 300–312 (2004)
3. Konstantinos, K., Ledig, C., Newcombe, V. FJ, Simpson, J. P, Kane, A. D, Menon, D. K, Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, 61–78 (2017)
4. Lun, TK, Hsu, W.: Brain tumor segmentation using deep convolutional neural network. Proceedings of BRATS-MICCAI (2016)
5. Meier, R., Knecht, U., Wiest, R., and Reyes, M.: CRF-based brain tumor segmentation: alleviating the shrinking bias. International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries 100–107 (2016)

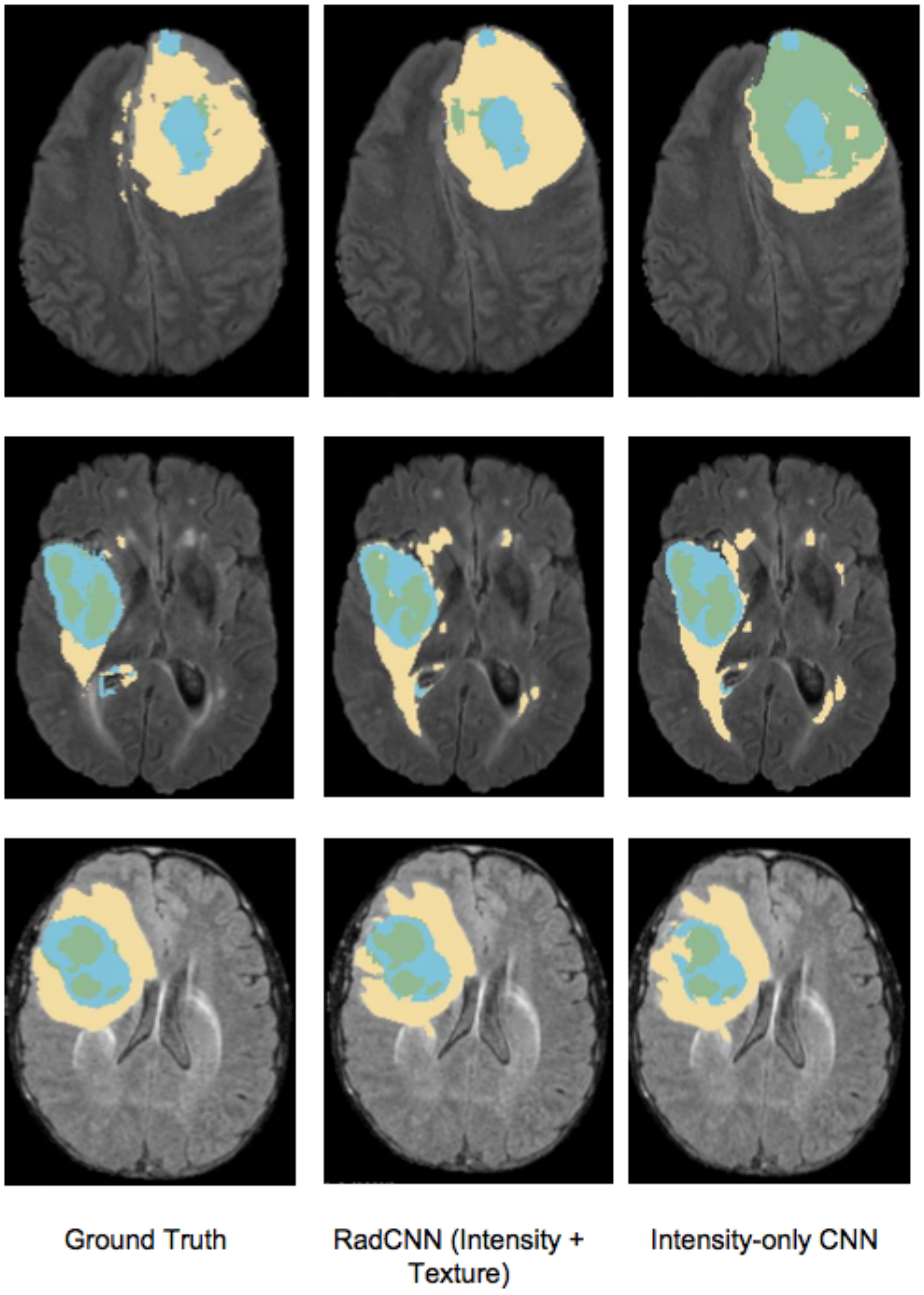


Fig. 3. Examples of segmentation results. From left to right: Ground-truth, RadCNN segmentation prediction, and an intensity-only DeepMedic CNN segmentation prediction. Blue, green, and yellow labels denote the necrosis + non-enhancing tumor (NCR + NET), enhancing tumor (ET), and edema (ED), respectively.

6. Ryu, Y.J., Choi, S.H., Park, S.J., Yun, T.J., Kim, J., Sohn, C.: Glioma: application of whole-tumor texture analysis of diffusion-weighted imaging for the evaluation of tumor heterogeneity. Public Library of Science (2014)
7. Prasanna, P., Tiwari, P., Madabhushi, A.: Co-occurrence of Local Anisotropic Gradient Orientations (CoLlAGe): a new radiomics descriptor. Scientific reports (2016)
8. Goetz, M., Weber, C., Bloecher, J., Stieltjes, B., Meinzer, H., Maier-Hein, K.: Extremely randomized trees based brain tumor segmentation. Proceeding of BRATS challenge-MICCAI (2014)
9. Rathi, VP, Palani, S.: Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis. arXiv preprint arXiv:1208.2128 (2012)
10. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, JS, Freymann, JB, Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017)
12. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, JS, Freymann, JB, Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
13. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, JS, Freymann, JB, Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)
14. Center for Biomedical Image Computing and Analytics, University of Pennsylvania: Image Processing Portal - https://ipp.cbica.upenn.edu/. A web accessible platform for imaging analytics (2015)
15. Prasanna, P., Patel, J., Partovi, S., Madabhushi, A., Tiwari, P.: Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. European radiology (2016)

Brain Tumor Segmentation using Deep U-Net

Geena Kim

Regis University, College of Computer and Information Sciences, Denver, CO, USA

Abstract. In this study, various modifications are made to a U-Net to segment brain tumor substructures. A simple mini U-Net model trained on a subset of BRATS 17 training data yields Dice scores of 0.8 (ET), 0.84 (WT), and 0.81 (TC) for HGG and 0.64 (ET), 0.83 (WT) and 0.68 (TC) for HGG+LGG BRATS 17 validation set. Modifications such as double convolution layers and inception modules are added to take advantage of deeper layers. Without any data augmentation, those modifications improved the Dice score to 0.68/0.86/0.70 (ET/WT/TC) and 0.66/0.87/0.72 (ET/WT/TC) on the BRATS 17 validation dataset.

Keywords: Brain Tumor Segmentation, Fully Convolutional Neural Networks, Deep Convolutional Neural Networks

1 Introduction

Convolutional neural network (CNN) has become the most popular approach in the field of computer vision since the AlexNet won by a large margin in the Large Scale Visual Recognition Challenge (LSVRC), a computer vision challenge with a task of classifying objects in natural images [6] in 2012. Various CNN models such as VGGNet, GoogLeNet (also called InceptionNet) and ResNet showed that CNNs with a deep-layer architecture can learn more complicated features from images [7–9]. Since then, the CNN approach has improved the accuracy beyond the human-level [10].

A deep CNN model has typically millions of parameters, thus requires a large amount of training images. The success of CNNs in recognizing natural images was largely due to the large public data repositories such as the ImageNet, a massive dataset with 14 million images which annotation was crowd-sourced via Amazon Mechanical Turk [5].

Analyzing medical images is more challenging than natural images by nature and it is also costly to obtain annotated data. Recent efforts organizing automated medical image analysis challenges and creating annotated datasets such as the Brain Tumor Segmentation (BRATS) challenge and its image collections [1–4] have been immensely helpful for the improvement of automated medical image analysis algorithms. Thanks to the considerable size and quality of the annotated data in BRATS, a number of deep CNN architectures have been implemented and showed top performances [16–20]. With performances comparable to the top-performing non-neural network algorithms [11–15], the

CNN-based algorithms have been steadily increasing their population in the challenge leaderboard, yet there are still rooms to improve.

Different from image classification tasks, the final output is 2D (or higher) in image segmentation tasks. Therefore, in segmentation, the classification of one pixel is highly correlated with that of adjacent pixels. One popular method to classify a pixel based on the neighboring pixels is the patch-based approach, which a model classifies the center pixel based on the information from surrounding pixels (patch) [23]. Many CNN architectures previously implemented in BRATS use the patch-based approach which has an added benefit that it can help mitigating severe class imbalance by selecting more patches from the tumor region. Recently, the fully convolutional neural network (FCN) approach has shown that it is more efficient than the patch-based approach [24], and has been implemented also in BRATS [18, 20]. Certain FCN-based models such as U-Net and V-Net tested on other medical images have symmetric contracting and expanding paths and merge the feature maps from contracting path to the feature maps in expanding path to preserve local features [25, 26].

In this paper, with an assumption that the modules from deep CNN models such as VGGNet, InceptionNet and ResNet developed for recognizing natural images are also useful in medical images, we add double convolutions and inception modules to the U-Net architecture ro segment brain tumor substructures.

2 Method

2.1 Preprocessing

The BRATS 17 Training data is normalized per patient scan using the median pixel value from the histogram of brain pixel values. Then the training data is split into training, validation, and test subsets based on the tumor location and size calculated from the segmentation labels such that cases in each location-size combined class are well distributed among training, validation and test sets. The 2D image slices are then shuffled within each set. Though it may be useful, N4 biasfield correction was omitted in the preprocessing as it is time consuming. Also, no data augmentation was used during training data preparation.

2.2 Architecture

Model A. Mini U-Net. Fig. 1 shows the base model (mini U-Net). It has total 11 2d-convolution layers. Except the last convolution layer for output which has 1×1 convolution, all other convolution layer has 3×3 filters. After every convolution, a batch normalization and ReLU activation is applied. Maxpooling is used to contract and upsampling is used to expand spatially. An upsampling layer uses simple resizing by nearest neighbor interpolation. As shown in the diagram, a feature map from the last layer of each downsampling step is concatenated depth-wise with the upsampled layer.

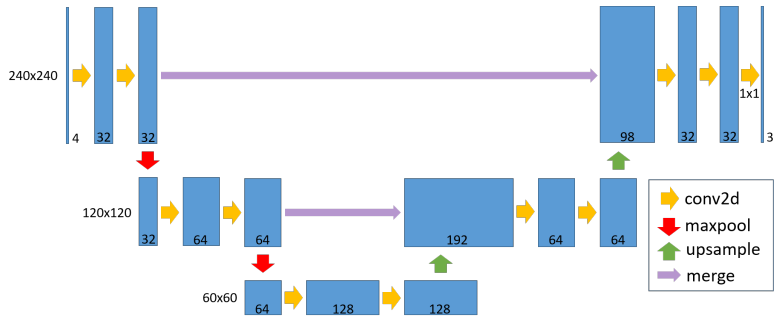


Fig. 1: Model A: mini U-Net architecture.

Model B. Mini U-Net with double convolution layers. This model has the same structure with the model A except that there are two 3×3 convolution layers before the activation to increase the receptive field (field of view). Having two successive 3×3 convolution layers is equivalent to having one 5×5 convolution but is computationally less expensive by factor of 25/9.

Model C. Mini U-Net with inception modules. Another variant of U-Net tested in this work uses modified inception modules. Typically, an inception module consists of 1×1, 1×1-3×3, 1×1-5×5 and 3×3 maxpool-1×1 paths running in parallel then merging depth-wise. In the InceptionNet, 1×1 convolutions before multiple convolution paths make it computationally less expensive, thus help constructing a deep layer architecture with less number of parameters. Inception-like modules which replace 5×5 convolutions with two 3×3 convolutions as mentioned above are added to the mini U-Net. Fig. 2 shows the overall architecture of the model C. A module A has 1×1, 1×1-3×3, and 1×1-3×3-3×3 convolution paths with stride = 1 and padding. A module B has a 3×3 max pooling layer with stride 2 (denoted as /2 by convention), 1×1-3×3/2, and 1×1-3×3-3×3/2 convolution layers. A module C is a reverse of a module B and consists of an upsampling layer, 1×1-3×3/2, and 1×1-3×3-3×3/2 deconvolution (transpose convolution) layers.

2.3 Training

The models are trained using Keras and TensorFlow python libraries and a NVIDIA TitanX GPU. Models A, B, and C have roughly 0.5 M, 1 M and 0.5 M parameters and the training takes 0.5-2 days on 20k 2D images. All models use an Adam optimizer with a learning rate between 0.001 to 0.01. Batch normalization is applied before every ReLU activation to reduce overfitting and use a higher learning rate.

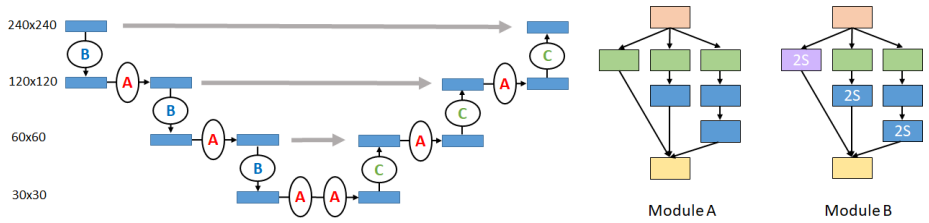


Fig. 2: Model C: mini U-Net architecture with inception-like modules. In the diagrams of Module A and B- orange squares represent the input layers to the modules, green squares represent 1×1 convolution layers, blue squares represent 3×3 convolution layers, and yellow squares represent depth-wise concatenation.

Table 1: Model performance: Mean 3D Dice score. In the Val17 column, the † symbol means that the Dice scores are obtained using a subset (HGG) of Val17 dataset.

Experiment	Model	Trained on	Dice score (ET/WT/TC)				
			Test data				
			TR15HGG	TR17HGG	TR17LGG	TR17ALL	Val17
1	A	TR15HGG	83/89/82	77/80/78	26/77/40	63/80/68	
2	A	TR17HGG		79/84/83	20/66/32	64/81/72	80/83.5/81†
3	A	TR17ALL		79/83/81		63/83/75	64/83/68
4	A	TR17HGG		79/84/85	22/72/32	64/82/71	
5	B	TR17HGG		80/85/86	32/76/31	65/83/71	
6	A	TR17LGG		62/72/50	24/82/66	52/75/54	
7	B	TR17LGG		67/76/52	20/82/63	54/78/56	
8	C	TR17LGG		66/77/49	24/82/66	55/79/54	
9	B	TR17ALL		79/85/82	22/83/65	64/85/77	68/86/70
10	C	TR17ALL		80/87/82	25/83/60	65/86/77	66/87/72

3 Result and Discussion

A summary of model performances is shown in Table 1. The performance of the model A (mini U-Net with a single convolution per layer) trained on the BRATS 15 training dataset versus on the BRATS 17 training dataset is shown in the experiments 1 and 2. The performance of the model A on the BRATS 17 validation dataset is similar to that of the BRATS 17 training dataset. The model A gives Dice scores of 0.84 (whole tumor), 0.80 (enhancing tumor), and 0.81 (tumor core) in HGG cases. Since the LGG data contains rare and confusing cases where there is zero to little enhancing tumor region while whole tumor region is large, a performance a model on the LGG data is generally poorer than one on the HGG data. On the HGG + LGG mixed dataset, the model A gave the Dice scores of 0.64, 0.83, and 0.68 for enhancing tumor, whole tumor and tumor core.

Experiments 1-3 and 4-10 use different normalization methods for preprocessing the images. The method used in experiments 4-10 uses the median pixel value in the brain pixels histogram instead of the 1-99 percentile pixel values to

normalize. The previous method (1-99) used in experiments 1-3 was considered mainly to mitigate an effect of outlier pixel values. However, it has a drawback of pushing the brightest 1% pixels to the same larger pixel values regardless of its original brightness, which makes more false positives for enhancing tumor. The second method (normalized by median) performs slightly better on the LGG data but shows little difference on the HGG+LGG data. Experiments 6-8 show performances of models trained on LGG data. Deeper layers seem to increase the overall performance, but it is not clear whether it improves segmentation of difficult cases in LGG data. Experiments 9-10 with the models B and C show improved results for the HGG+LGG data by implementing deeper layer architectures. The model C did not seem performing better than the model B within above experiments. However, further model optimization is needed to better test the effect of inception modules.

Since the number of model parameters are bigger in deeper layers, it is easier to overfit. Also, the class imbalance is severe when segmenting tumor substructures. The percentage of 2D slices with at least one pixel from tumor substructure is less than 50%, and even worse, the percentage of 2D slices with pixels from enhanced tumor is much smaller. As shown above, the model performances on the LGG data is not as good as on the HGG or mixed cases. In the LGG data, there are certain cases with a large whole tumor region without any enhancing tumor pixels. Also, there are rare cases both in HGG and LGG such as tumors occurring top and bottom parts of the brain. The difficulty from a large number of parameters in the model and the high class imbalance can be alleviated by aggressive data augmentation and resampling, which we aim to do as the next step.

Acknowledgement

We thank NVIDIA for their kind donation of a TitanX GPU.

References

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging. 34, 1993-2024 (2015).

2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017) [In Press].
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017).
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017).
5. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision. 115, 211-252 (2015).
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097-1105 (2012).
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. (2014).
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1-9 (2015).
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770-778 (2016).
10. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1026-1034 (2015).
11. Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., Davatzikos, C.: GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 1st International Workshop, Brainles 2015. pp. 144-155. Springer International Publishing (2016).
12. Maier, O., Wilms, M., Handels, H.: Image Features for Brain Lesion Segmentation Using Random Forests. In: Crimi, A., Menze, B., Maier, O., Reyes, M., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 1st International Workshop, Brainles 2015. pp. 119-130. Springer International Publishing (2016).
13. Meier, R., Karamitsou, V., Habegger, S., Wiest, R., Reyes, M.: Parameter Learning for CRF-Based Tissue Segmentation of Brain Tumors. In: Crimi, A., Menze, B., Maier, O., Reyes, M., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 1st International Workshop, Brainles 2015. pp. 156-167. Springer International Publishing (2016).
14. Song, B., Chou, C.-R., Chen, X., Huang, A., Liu, M.-C.: Anatomy-Guided Brain Tumor Segmentation and Classification. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 2nd International Workshop, BrainLes 2016. pp. 162-170. Springer International Publishing (2016).

15. Zeng, K., Bakas, S., Sotiras, A., Akbari, H., Rozycki, M., Rathore, S., Pati, S., Davatzikos, C.: Segmentation of Gliomas in Pre-operative and Post-operative Multimodal Magnetic Resonance Imaging Volumes Based on a Hybrid Generative-Discriminative Framework. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 2nd International Workshop, BrainLes 2016. pp. 184-194. Springer International Publishing (2016).
16. Havaei, M., Dutil, F., Pal, C., Larochelle, H., Jodoin, P.-M.: A Convolutional Neural Network Approach to Brain Tumor Segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 1st International Workshop, Brainles 2015. pp. 195-208. Springer International Publishing (2016).
17. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Deep Convolutional Neural Networks for the Segmentation of Gliomas in Multi-sequence MRI. In: Crimi, A., Menze, B., Maier, O., Reyes, M., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 1st International Workshop, Brainles 2015. pp. 131-143. Springer International Publishing (2016).
18. Chang, P.D.: Fully Convolutional Deep Residual Neural Networks for Brain Tumor Segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 2nd International Workshop, BrainLes 2016. pp. 108-118. Springer International Publishing (2016).
19. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., Rueckert, D., Glocker, B.: DeepMedic for Brain Tumor Segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 2nd International Workshop, BrainLes 2016. pp. 138-149. Springer International Publishing (2016).
20. Zhao, X., Wu, Y., Song, G., Li, Z., Fan, Y., Zhang, Y.: Brain Tumor Segmentation Using a Fully Convolutional Neural Network with Conditional Random Fields. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S., and Handels, H. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 2nd International Workshop, BrainLes 2016. pp. 75-87. Springer International Publishing (2016).
21. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Medical Image Analysis. (2016).
22. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis. 36, 61–78 (2017).
23. Ciresan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Deep neural networks segment neuronal membranes in electron microscopy images. In: Advances in neural information processing systems. pp. 2843-2851 (2012).
24. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431-3440 (2015).
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234-241. Springer (2015).
26. Milletari, F., Navab, N., Ahmadi, S.-A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. arXiv:1606.04797 [cs]. (2016).

Enhancing DeepMedic Architecture by Bottleneck Blocks and U-Net-like skip connections

Egor Krivov e.a.krivov@gmail.com[1] and
Mikhail Belyaev m.belyaev@skoltech.ru[2, 1]

[1]Institute for Information Transmission Problems
[2]Skolkovo Institute of Science and Technology

Abstract

DeepMedic architecture is the state-of-the-art solution for 3d medical image segmentation. In this paper, we try to improve it by using residual connections with linear projection, bottleneck blocks and direct connections from U-Net.
Keywords: CNN, Segmentation, MRI, BraTS

1 Introduction

Over the last few years, deep learning (DL) has become the state-of-the-art solution for image analysis. DL was originally successfully applied to two-dimensional images in such competition as Imagenet [1] and COCO [2]. Usually, top-performing solutions for these competitions turn into templates for DL applications in similar areas.

We try to use a similar approach here and apply bottleneck blocks [3], residual connections with linear projection and U-Net connections to improve one of the best solutions for BraTS 2016, DeepMedic architecture [4].

2 Method

2.1 Original DeepMedic

DeepMedic architecture (Figure 1) has proven to be one of the best solutions for medical image segmentation. It is based on volumetric convolutional layers 3-dimensional which allows us to use information for all three directions. The network also has context pathways, which extract context information about image region. Finally, it is a fully convolutional network, allowing fast and effective evaluation on large images.

2.2 Direct connection from U-Net

U-Net [5] uses direct connections from the detailed input image to the final layers of segmentation, improving object boundary detection. We try to apply the same scheme to DeepMedic, by directly connecting detailed input to the block that concatenates detailed and contextual information.

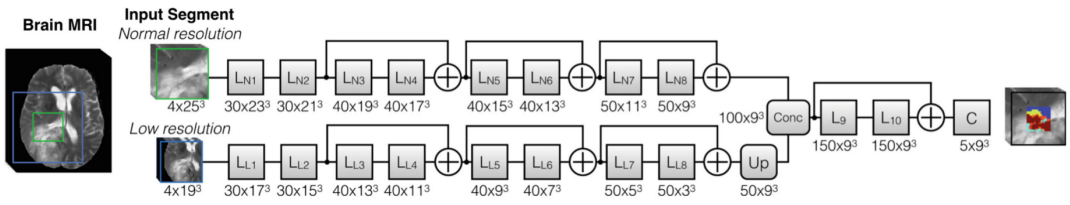


Figure 1: Original DeepMedic architecture from [4]

Architecture	WT	TC	ET
original DeepMedic	0.842 (0.163)	0.765 (0.244)	0.658 (0.321)
with direct connection	0.848 (0.149)	0.736 (0.264)	0.650 (0.322)
with linear projection connection	0.848 (0.146)	0.763 (0.253)	0.663 (0.324)
with bottleneck block	0.844 (0.158)	0.759 (0.250)	0.643 (0.320)

Table 1: Dice scores and std across patients for whole tumor (WT) tumor core (TC) and enhancing tumor (ET)

2.3 Residual connections with linear projections

To our knowledge, the latest version of DeepMedic uses residual connections that simply add features for available channels, meaning that if the number of input channels for a residual block is less than the number of output channels, some channels will not get any residual connection. This problem can be solved by using linear projections, which are implemented as $1 \times 1 \times 1$ convolution, linking input and output of the residual block [3]. We change all simple residual connections in DeepMedic with residual connections via linear projections.

2.4 Bottleneck block

The bottleneck block from [3] is used as a residual connection which compresses the input before performing convolution, therefore, decreasing the number of required parameters, which was beneficial for the winning solution of 2015 Imagenet competition. We replace all residual blocks with bottleneck blocks.

Since bottleneck blocks have only one convolution with kernel size more than zero, we use kernel size of 5, instead of 3 (as in original DeepMedic) to get the same output shape. Number of output channels for the bottleneck block is the same as in original DeepMedic, while the number of inner channels is four times smaller, following [3] approach.

3 Experiments and results

To evaluate suggested improvements, we have implemented DeepMedic and suggested improvements with tensorflow [6]. As a dataset we used BraTS 2017 [7] [8] [9] [10]. We have randomly split the dataset (285 scans) into three parts: training (78%), validation (6 scans) and testing (20%). We minimized logarithmic loss function on the training set using stochastic gradient descent with Nesterov momentum. We started with learning rate equal to 0.1 and deceased it twice each time validation loss plateaus. Patches are equally sampled from healthy and cancerous tissues. Batch size was equal to 64, batches per epoch to 100. Number of epochs was equal to 80, because after that point both training and validation loss plateaus completely. The whole training process takes about 4command hours on Nvidia GTX980Ti per architecture.

We were predicting three masks: whole tumor (WT), tumor core (TC) and enhancing tumor (ET). To transform probability predictions to discrete predictions on test dataset, we picked three thresholds by maximizing dice scores on the validation set.

Results are presented in table 1. It is worth mentioning, that our results are lower than usual for DeepMedic.

4 Conclusion

In this work we have tried to improve DeepMedic with simple modifications. Although ideas were reasonable, it didn't really improve or change dice scores. It is signaling that this architecture is quite robust to changes.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[2] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, Springer, 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[4] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, "Deepmedic for brain tumor segmentation," in *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 138–149, Springer, 2016.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.

[6] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[7] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[8] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," 2017.

[9] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels for the pre-operative scans of the tcga-gbm collection," 2017.

[10] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels for the pre-operative scans of the tcga-lgg collection," 2017.

Brain Tumor Segmentation Using an Adversarial Network

Zeju Li[1], Yuanyuan Wang[1,2*], Jinhua Yu[1,2*]

[1] Department of Electronic Engineering, Fudan University, Shanghai, China
[2] Key laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Shanghai, China

Abstract. Recently, the convolutional neural network (CNN) has been successfully applied to the task of brain tumor segmentation. However, the effectiveness of a CNN-based method is limited by the small receptive field, and the segmentation results don't perform well in the spatial contiguity. Therefore, many attempts have been made to strengthen the spatial contiguity of the network output. In this paper, we proposed an adversarial training approach to train the CNN network. A discriminator network is trained along with a generator network which produces the synthetic segmentation results. The discriminator network is encouraged to discriminate the synthetic labels from the ground truth labels. Adversarial adjustments provided by the discriminator network are fed back to the generator network to help reduce the differences between the synthetic labels and the ground truth labels and reinforce the spatial contiguity with high-order loss terms. The presented method is evaluated on the Brats2017 training dataset. The experiment results demonstrate that the presented method could enhance the spatial contiguity of the segmentation results and improve the segmentation accuracy.

Keywords: Brain Tumor Segmentation, Adversarial Network, Deep Learning

1 Introduction

Automatic segmentation of brain tumors in magnetic resonance imaging (MRI) images is of great clinical value. Multimodal brain tumor segmentation challenge (Brats) provides a great platform for the evaluation of different segmentation methods [1][2][3][4]. Among the existing segmentation methods, the convolutional neural network (CNN) provides very outstanding results and attracts increasing attentions [5]. However, one defect of a CNN-based segmentation method is that the receptive field is always limited by the size of convolutional kernels. This problem could become more apparent in brain tumor segmentation because the appearance of brain tumors is unpredictable and MRI images are inhomogeneous in the intensity. Therefore, the segmentation results of CNN-based methods always have rough boundaries and perform poorly on details of tumor sub-regions. Instead of adding the context information or conditional markov random field (CRF), we want to strength the spatial contiguity by using an auxiliary high-order loss term, which can make the net-

work more perceptive towards spatial-connected tumor regions. This idea is inspired by the generative adversarial network (GAN) [6]. The GAN, through the use of adversarial loss, has successfully been applied to the generation of real-life images in an unsupervised way [7].

In this paper, we present a novel CNN-based brain tumor segmentation method by using an adversarial network. The presented method was evaluated on the Brats2017 training dataset.

2 Method

The proposed method consists of two CNN networks, named the generator network and the discriminator network. These two networks are tightly connected. The detailed structure is described in Fig. 1.

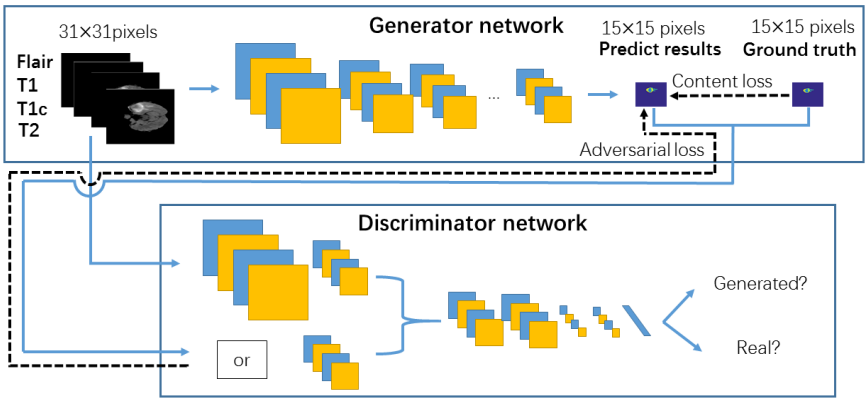


Fig. 1. Schematic diagram of the proposed framework. Forward propagations are depicted using blue lines and backward propagations are depicted using black dotted lines.

2.1 Generator Network

Generator network is designed to produce segmentation results. The lost function of the generator network consists of two parts, called the content loss and the adversarial loss. Specifically, the loss function is calculated as:

$$L^G = L_{softmax}(G(I), label_{GT}) + L_{Ad}^G \quad (1)$$

The first term $L_{softmax}$, namely the content loss, is the pixel-wise derivatives. I represent the patches of four modalities MRI images. $label_{GT}$ represent the given ground truth labels which are corresponding to I. The second term L_{Ad}^G, namely the adversarial loss, is produced simultaneously by the adversarial network and is discussed in the next section.

2.2 Discriminator network

Adversarial Training. The discriminator network is a core of the presented method. The synthetic labels and the ground truth labels, with the form of the one-hot coding, are input to the discriminator network. The original MRI images are also input to the network as references. Leaky relu layers are utilized but no pooling layers are included in the discriminator network, as suggested by a previous GAN study [7]. Images are down sampled to the same size of the labels using convolutional layers with stride 2. The discriminator network is trained to distinguish two kinds of labels obtained from different places. The loss function is calculated as:

$$L^D = \frac{1}{2}\left[L_{bce}(I, D(label_{GT}), 0) + L_{bce}\left(I, D\left(G(I)\right), 1\right)\right] \quad (2)$$

As usual, the binary cross entropy (bce) loss is calculated to differ two types of inputs.

Adversarial Loss. The discriminator network provides the adversarial loss by minimizing the probability that the adversarial predicts $D(label_{GT})$ are similar to the one of synthetic labels. The adversarial loss is calculated as:

$$L_{Ad}^G = L_{bce}(I, D\left(G(I)\right), 0) \quad (3)$$

The adversarial loss contains high order derivatives and are fed back to the generator network in (1). The effectiveness of pretrain of the model is evaluated in this study. During the pretrain procedure, the generator network is firstly trained for a while separately, and the discriminator is trained with settled generator network until the network is stable.

2.3 Post-processing

The largest three-dimensional (3D) connection region of the segmentation results is firstly chosen as the tumor candidate. 3D bounding box slightly larger than the tumor candidate is built, and then recognition results inside the bounding box are finally identified as the tumor region.

3 Experimental Results

It is interesting for us to observe the effect of the addition of the discriminator network. Thus, two experiments were carried out with exactly the same condition except for the existence of the adversarial network. Qualitative results could be found in Fig. 2. To better illustrate the effectiveness of the adversarial loss, the segmentation results were extracted directly from the generator network without any post-processing. With the help of adversarial loss, the amount of over segmentation is reduced. False positive segmentation results are also decreased. Moreover, the spatial contiguity of segmentation of all tumor sub-regions is enhanced. The segmentation results become

smoother in the sub-regions and preserve good boundaries. It is hard to achieve by the post-processing of CRF with certain parameters.

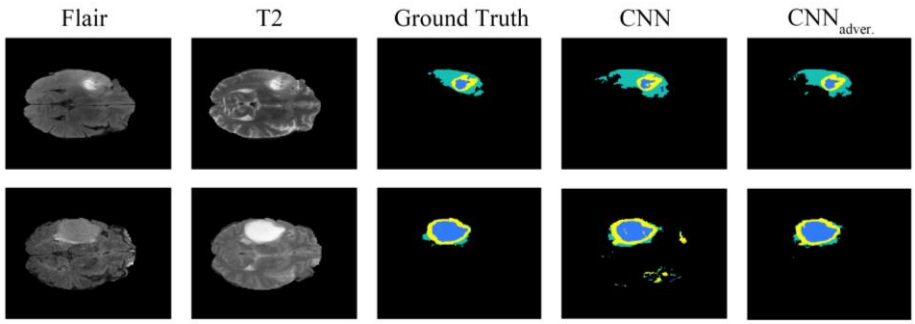


Fig. 2. Visual segmentation results of two examples using the direct outputs of CNN network.

To quantitative exhibit the effectiveness of the adversarial loss, the segmentation results of all 285 training data are summarized in Table 1. The effectiveness of pre-train of the model is also evaluated.

Table 1. Quantitative segmentation results of 285 training data using the direct outputs of a CNN network. $CNN_{adver.}$ corresponds to the network trained with the adversarial loss. Results with post-processing is marked followed by *.

Method	Whole			Core			Enhancing		
	DSC	PPV	Sen.	DSC	PPV	Sen.	DSC	PPV	Sen.
CNN	86.1	82.2	90.0	86.3	85.5	87.2	77.1	74.2	80.1
$CNN_{adver.}$	87.9	86.8	89.0	86.8	86.6	87.0	77.5	77.1	77.9
$CNN_{adver.}$*	89.5	90.6	88.4	88.0	89.0	87.0	78.4	79.1	77.8
$CNN_{adver.}$-pretrain*	**89.7**	89.8	88.5	**88.4**	**89.7**	**87.2**	**79.1**	78.1	**80.0**

4 Discussion

The discriminator network is trained by discriminating the synthetic labels from the ground truth labels with the use of plenty of training patches. An adversarial network can discover the mutual characteristics of mistakes by learning from the training data, and deliver the information to the generator network in a high-order form. This kind of high-order loss has a stronger adaptability, contains more global information, and can't be provided by per-pixel loss. Thus, this kind of loss is more meaningful.

The pretrain of the model seems to be useful. Understandably, the better the discriminator is trained, the more effective the adversarial loss is. The experiment results illustrate this point. Segmentation results of the pretrained model show more accurate recognition ability, especially in details.

5 Conclusion

In this study, a novel CNN-based tumor segmentation method with adversarial network is presented. A high-order adversarial loss provided simultaneously by the discriminator network is added to encourage the generator network to produce more precise results. It should be mentioned that the presented method is also efficient and energy-saving since the complexity is not added to the network at test time. In the future, we will extent the presented method into 3D to better make use of the 3D information of MRI images. Further, we will exploit the best way to take advantage of the adversarial loss.

Acknowledgments. This work was supported by the National Basic Research Program of China (2015CB755500), the National Natural Science Foundation of China (11474071).

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Pereira S, Pinto A, Alves V, Silva CA. "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images", IEEE Transactions on Medical Imaging 35(5), 1240-1251 (2016)
6. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. "Generative Adversarial Networks", NIPS, pp. 2672-2680 (2014)
7. Radford A, Metz L, Chintala S. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", arXiv: 1511.06434 (2015)

MvNet: Multi-view Deep Learning Framework for Multimodal Brain Tumor Segmentation

Yuexiang Li and Linlin Shen*

Computer Vision Institute, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
{yuexiang.li, llshen}@szu.edu.cn

Abstract. Brain tumor segmentation plays an important role in the disease diagnosis. In this paper, we proposed a multi-view deep learning framework to address the challenge of multimodal brain tumor segmentation. The proposed deep learning framework uses three U-net based sub-networks to segment multimodal brain images from different view-point, i.e. slices along x, y, z axis. The three sub-networks produce independent segmentation results and vote for the final outcome. The proposed multi-view deep learning framework was evaluated on BraTS 17 validation set and achieved a competing result, i.e. Dice scores of 0.84, 0.66, 0.54 for whole tumor, enhancing tumor and tumor core, respectively.

Keywords: Deep Learning, Multi-view, Tumor Segmentation

1 Introduction

Brain tumor is a severe disease threating the health of human-being. An accurate automatic tumor segmentation framework can significantly improve the efficiency of disease diagnosis and help design appropriate treatment strategy. In recent years, we witnessed the development of deep learning algorithm and were impressed by its powerful performance. Increasing numbers of studies tried to employ deep learning algorithm to process medical images. In previous challenges, i.e. BraTS 15-16 [1-3], various 2D and 3D deep learning networks have been proposed for the segmentation of multimodal brain tumor. For example, Lun et al. evaluated three types of 2-D convolutional networks, i.e. Patch-Wise, FCN [4] and SegNet [5] for BraTS 15 dataset. Kamnitsas et al. extended a 3-D CNN architecture, i.e. DeepMedic [6], with residual connections for brain lesion segmentation. However, most of the proposed 2D networks only use the slices along z-axis, which do not fully explore the spatial information compared to the 3D approaches. As the 3D convolutional network is not computational-efficient, in this paper, we proposed a 2.5D deep learning framework, with three sub-networks for different view-points, to take advantages of both 2D and 3D approaches. Henceforth, our proposed 2.5D framework is named as Multi-view net (MvNet).

2 Multi-view Deep Learning Network

2.1 Network Architecture

The proposed MvNet framework consists of three sub-networks and processes multi-modal brain images along different axis, as illustrated in **Fig. 1**. FLAIR voxel data is taken for example, though four brain image models, i.e. T1, T1Gd, T2 and FLAIR, are available. The slices from different modal were concatenated as input for each U-net [7].

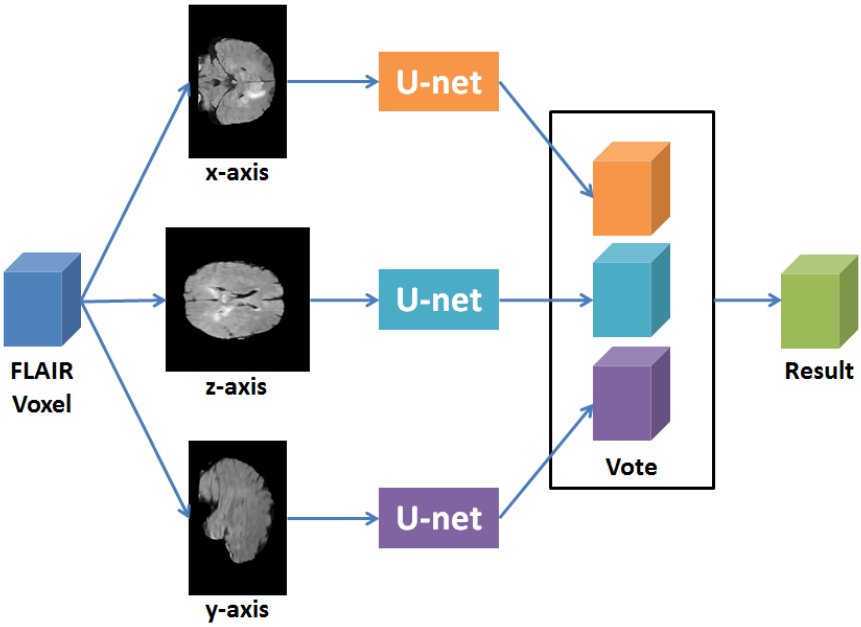
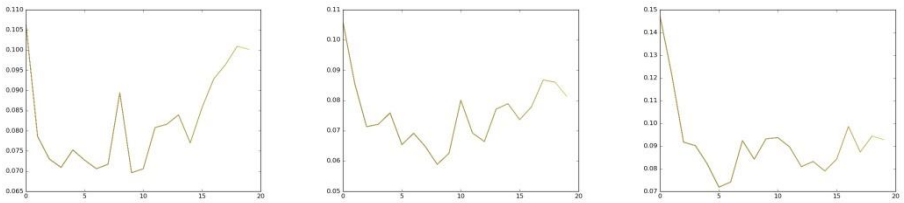


Fig. 1. Architecture of MvNet framework

2.2 Implementation

The proposed MvNet is established using PyTorch toolbox. The network is trained on two K80 with a mini-batch size of 30. Adam is used as the optimizer [8]. The start learning-rate is set to 0.0002. The BraTS 17 training dataset [9, 10] is separated to training and validation sets according to the ratio of 80:20. The three sub-networks are separately trained on training set and validated on validation set. **Fig. 2** presents the loss curves for U-net along different axis. From **Fig. 2**, the three sub-networks are found to produce the best performances on validation set after about 10 epochs of training.

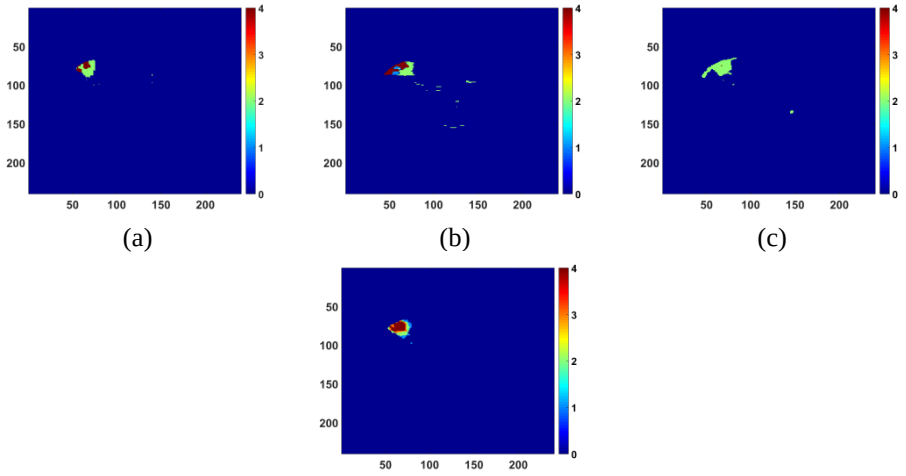


(a) (b) (c)

Fig. 2. Loss curves on validation set. (a) is for x-axis U-net. (b) is for y-axis U-net and (c) is for z-axis U-net.

2.3 Results

Fig. 3 presents the segmentation results from different U-net branches. The labels in the mask represent different parts of brain tumor, i.e. label 4 is for GD-enhancing tumor, label 2 is for peritumoral edema and label 1 is for necrotic and non-enhancing tumor. It can be observed from **Fig. 3** that the final result (**Fig. 3** (d)) removes the segmentation errors by fusing results from different U-nets (**Fig. 3** (a-c)). We compared the Dice coefficient of z-axis only framework and our MvNet on the BraTS 17 validation set. In **Table 1**, ET represents enhancing tumor. WT represents whole tumor and TC represents tumor core. Although a drop of Dice_TC, i.e. 0.08, is found by adding x and y axis U-nets, the proposed MvNet significantly improves the segmentation performances for ET and WT, i.e. 0.17 and 0.07.



(a) (b) (c)

(d)

Fig. 3. Segmentation results (view from z-axis). (a) Result from x-axis U-net. (b) Result from y-axis U-net. (c) Result from z-axis U-net. (d) Final result after voting.

Table 1. Dice coefficient on validation set

	Dice_ET	Dice_WT	Dice_TC
z-axis U-net	0.49	0.77	0.62
MvNet	0.66	0.84	0.54

3 Survival Prediction

Survival prediction is another task introduced in BraTS 17. We developed a convolutional network to predict the survival periods of patients. The convolutional network (CNN) includes the four brain image modals and the tumor segmentation result from MvNet as input and finally predicts the survival days. In the train-validation process, the CNN achieves an error of ±40 days on our validation set.

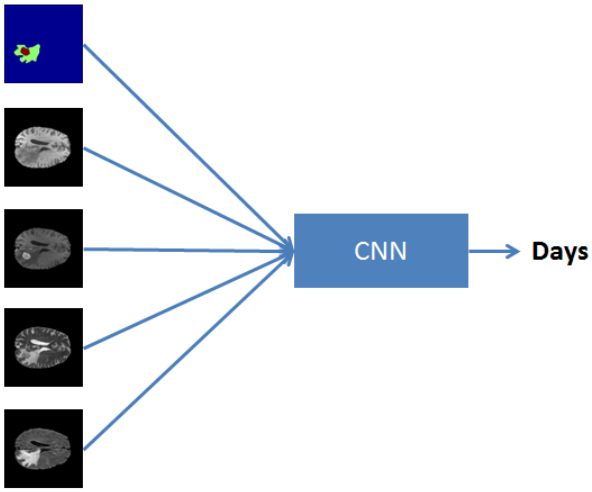


Fig. 4. Flowchart for survival prediction

4 Conclusion

In this paper, we proposed a novel deep learning framework, so-called MvNet, for multi-modal brain tumor segmentation. The proposed MvNet employs three sub-networks to process the brain images along different axis. The results on BraTS 17 validation set show the competing performance of MvNet, i.e. Dice scores of 0.84, 0.66 and 0.54 were achieved for whole tumor, enhancing tumor and tumor core, respectively.

Reference

1. Menze, B.H., Jakab, A., Bauer, S., Kalpathycramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34, 1993-2024 (2015)
2. Previous BraTS Challenges: http://www.braintumorsegmentation.org/
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, C, D.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data [In press] (2017)
4. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431-3440 (2015)
5. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence PP, 1-1 (2017)
6. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, 61 (2016)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer Assisted Interventions, vol. 9351, pp. 234-241 (2015)
8. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv e-print arXiv:1412.6980 (2014)
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, C, D.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q, (2017)
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, C, D.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF, (2017)

Appendix

Table 2. Detailed information of U-net used in MvNet. The pipeline consists of Input layer (**I**), Convolutional layer (**C**), Batch Normalization layer (**BN**), LeakyReLU (**L**), Max-pooling layer (**M**), Convolution Transpose layer (**CT**), Concatenation layer (**CON**) (the number in this layer tells it concatenates the output from which layer) and loss layer, i.e. LogSoftmax.

Layer	Input Size	Type	Kernel size & amount
1	224x224x4	I-C-BN-L-C-BN-L-M	3x3, 32
2	112x112	C-BN-L-C-BN-L-M	3x3, 64
3	56x56	C-BN-L-C-BN-L-M	3x3, 128
4	28x28	C-BN-L-C-BN-L-M	3x3, 256
5	14x14	C-BN-L-C-BN-L	3x3, 512
6	14x14	CT-CON(4)	4x4, 256
7	28x28	C-BN-L-C-BN-L	3x3, 256
8	28x28	CT-CON(3)	4x4, 128
9	56x56	C-BN-L-C-BN-L	3x3, 128
10	56x56	CT-CON(2)	4x4, 64
11	114x114	C-BN-L-C-BN-L	3x3, 64
12	114x114	CT-CON(1)	4x4, 32
13	224x224	C-BN-L-C-BN-L	3x3, 32
14	224x224	C-LogSoftmax	3x3, 4

Brain Tumor Segmentation via 3D Fully Dilated Convolutional Networks

Xiaogang Li*, Xiang Zhang†, and Zhigang Luo†

*College of Science, National University of Defense Technology, Changsha 410073, China
†College of Computer, National University of Defense Technology, Changsha 410073, China

Abstract. In this paper we propose a novel two-pathways 3D fully convolutional networks (3DFCN) architectures for brain tumor segmentation task. By using dilated convolutions, we can effectively increase the size of receptive field without additional layers, which enable the networks efficiently capture large-scale semantic contextual. The two-pathways networks with different dilated factors aggregate multi-scale contextual and enhance the space consistence of the segmentation results, which dramatically decreases the false positive in our experiments.

Keywords: FCN, Dilated Convolutions, Multi-Scale

1 Introduction

Gliomas are the most common brain tumors that occur in adults [6,10]. Due to the gliomas have significantly varies size, shape, and location, it is difficult to extract the exact region of brain tumor from the MRI volumes, which is important for clinical diagnosis and treatment. In this case, high accuracy automatic segmentation methods play an important role in practice. As the benchmark of this area, the BRATS challenges provide a uniform criteria for evaluating brain tumor segmentation algorithms [6]. There are two tasks in BRATS 2017. In this short paper we only focus on the first one: segmentation of gliomas in pre-operative scans [8,9].

The deep learning methods based on [5] have reached the start of art performance in image semantic segmentation task. Recently this excellent idea has been introduced in medic image segmentation [1–3,7]. Considering the characteristic of MRI, some 3D convolutional networks methods were proposed in order to utilize the 3D contextual information in brain MRI volumes [1,3].

In addition, several strategies were considered to improve the space consistence of segmentation results, such as using fully connected conditional random field (CRF) [4] as post-processing step [3], using multi-pathways networks architectures [1–3] and so on.

2 Method

2.1 3D Dilated Convolutions

The definition of 2D dilated convolutions was introduced by **Fisher Yu** for multi-scale context aggregation in image segmentation task [11]. As below, we simply general this definition into 3D format.

Let $F : \mathbb{Z}^3 \to \mathbb{R}$ be a discrete function. Let $\Omega_r = [-r, r]^3 \cap \mathbb{Z}^3$ and let $k : \Omega_r \to \mathbb{R}$ be a discrete filter of size $(2r + 1)^3$. The 3D dilated convolution operator $*_l$ can be defined as

$$(F *_l k)(\mathbf{p}) = \sum_{\mathbf{s}+l\mathbf{t}=\mathbf{p}} F(\mathbf{s})k(\mathbf{t}). \tag{1}$$

The operator $*_l$ is usually referred as $l-$dilated convolution operator [11]. Obviously, $1-$dilated convolution is the ordinary convolution. Consider the sequence dilated convolutional operations:

$$F_{i+1} = F_i *_{l_i} k_i \quad \text{for } i = 0, 1, \ldots, n-1. \tag{2}$$

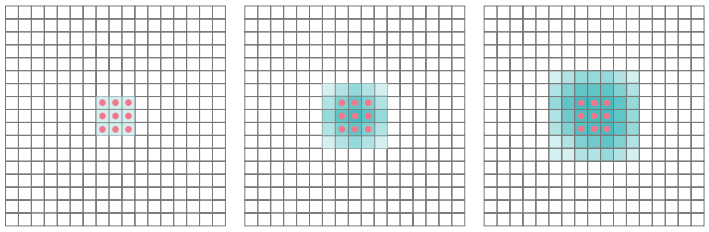
The receptive field of one element in F_i can be defined as the set of elements in F_0 that impact the value of this element. Denote the receptive field as I. Then the size of receptive field is the number of element in I, *i.e.* $|I|$. We illustrate the middle slice of these two kind of convolutions in figure 1.

2.2 3D Fully Dilated Convolutional Network Architecture

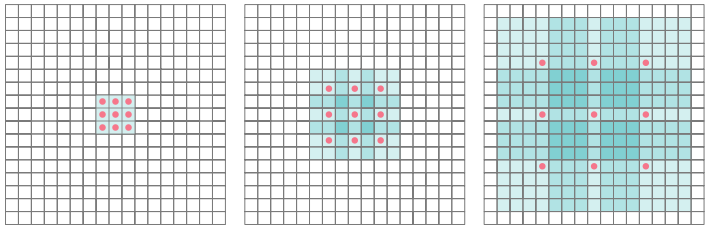
We design a two-pathways architectures with 3D dilated convolutional layers, as displaying in figure 2. The first pathway network uses multi-modals MRI as input and adopts a larger dilated factors (1, 2, 2, 2, 1, 1), which concentrate on large-scale contextual information. The second pathway network uses both the segmentation results of the first pathway and the multi-modals MRI as input, which employs a smaller dilated factors (1, 1, 2, 1, 1, 1) and focus on the small-scale contextual information. The second pathway plays a role in improving the space consistence of the segmentation results.

3 Experiments and Results

As this is a short paper, we only focus on the brain tumor segmentation task of this year. We use 190 HGG cases and 70 LGG cases to train our networks and use 20 HGG cases and 5 LGG cases for validation. Our primary results about *Dice Score*, *Spec*, *Sens* equal 0.42, 0.99, 0.66 respectively.



(a) The receptive field of three ordinary convolution layers which equal three 1-dilated convolution layers.



(b) The receptive field of three dilated convolution layers with dilated factor equals 1, 2, 4 respectively.

Fig. 1: Comparing those two kind of convolutions, it is obvious that the dilated convolution can expand the size of receptive field without loss of resolution [11]. By changing the value of dilated factor, we can control the size of receptive field, which provides a facile way to exploit different scale contextual information without changing the architectures of the networks.

4 Discussion and Conclusion

We use this two-pathways networks to integrate multi-scale information. With help of dilated convolutions we can achieve large receptive field with less layers and change receptive field size of network by adopt different dilated convolution factors. Because we only exam small networks, our primary results are not very outstanding. In the future work, we will increase the depth of our networks and exploit some other post-processing methods such as fully connected conditional random field [4].

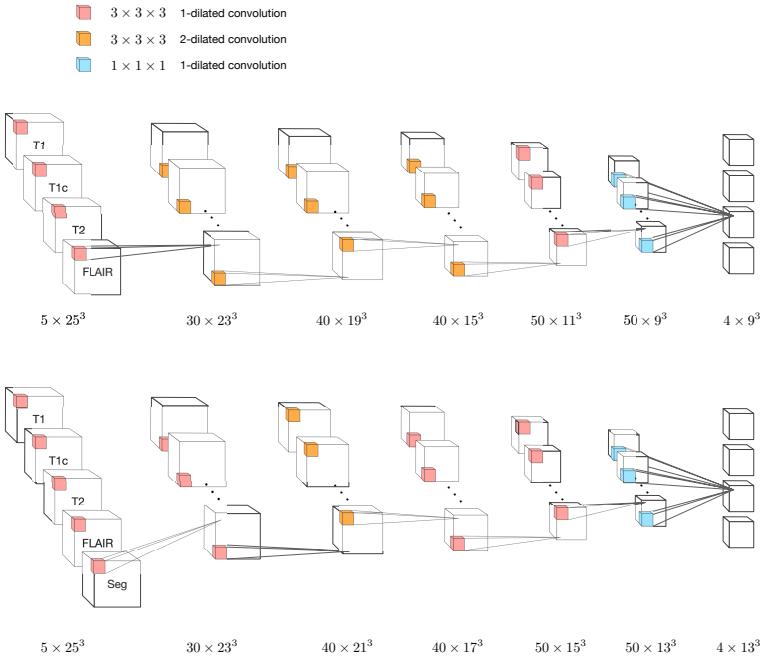


Fig. 2: The 3D fully dilated convolutional network architecture. Every training data volumes have been normalized before training. The smaller convolution kernels ($3 \times 3 \times 3$) can effectively decrease the number of parameters for those networks. And due to the convolution kernels are dilated, we do not lose receptive field. In the training section, we first train the first pathway network with multi-modals MRI training data. Then we use the trained first network to segment the training data and get the prediction volumes. Next, utilize the multi-modals MRI training data and its corresponding prediction volumes, generated by the first pathway network, to train the second pathway network.

References

1. Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study. *NeuroImage*, 2017.
2. Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.
3. Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.
4. Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.
5. Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.
6. Menze, Bjoern H, Jakab, Andras, Bauer, Stefan, Kalpathy-Cramer, Jayashree, Farahani, Keyvan, Kirby, Justin, Burren, Yuliya, Porz, Nicole, Slotboom, Johannes, Wiest, Roland, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2015.
7. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive*, 2017.
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive*, 2017.
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, and Davatzikos C. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 2017.
11. Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

A Location Sensitive Brain Tumor Segmentation Method

Luyan Liu[1], Dong Nie[2], Qian Wang[1] and Dinggang Shen[2]

[1] Med-X Research Institute, School of Biomedical Engineering, Shanghai Jiao Tong University, China
[2] Department of Radiology BRIC, University of North Carolina at Chapel Hill, USA
dgshen@med.unc.edu

Abstract. Automatic brain tumor segmentation from magnetic resonance imaging (MRI) is very important as it is significantly beneficial for early diagnosis, radiotherapy planning, outcome prediction and follow-up assessment of patients in (high or low grade) brain tumors. As we all known, brain tumor is an aggressive which can grow and spread in any areas of brain. Tumor also includes different subtypes and different kinds of tumor correspond to different treatments due to their different characteristics. Thus, pre-delineation of tumor's subcomponents is in urgent demand to the better diagnosis of life-threatening brain tumor. In this paper, we proposed a novel location sensitive brain tumor segmentation framework which can not only detect brain tumors in high accuracy quickly, but also improve the accuracy on the state-of-the-art brain tumor segmentation techniques. Our proposed method consists of two stages: 1) tumor detection and 2) tumor segmentation. In the first stage, we used a location sensitive region based fully convolutional network to generate a 3D location sensitive map. The location sensitive map can be used to guide the tumor segmentation in the second stage which could be any deep learning based methods. The experimental results show that our brain tumor segmentation method can improve the accuracy of state-of-the-art brain tumor segmentation method. Notably, our proposed brain tumor segmentation method is a general framework which can be easily applied to any other existing brain tumor segmentation methods.

Keywords: Location Sensitive Map, Convolutional Neuronal Network, Brain Tumor Segmentation.

1 Introduction

In worldwide, there are almost 238,000 patients are diagnosed as brain tumors every year [1]. Among them, gliomas are the most common brain tumors among adults, which accounts for 70% of adult primary brain tumors [2]. Based on World Health Organization (WHO) grade [3], gliomas can be divided into two groups: high-grade glioma (HGG) and low-grade glioma (LGG). High-grade gliomas are more aggressive (growing faster) and usually have worse outcome with two years or less survival time

[4]. The low-grade gliomas are growing slower which come with a longer life expectancy. Though there are considerable progresses in glioma related research, the diagnosis and treatment are still dissatisfactory. In the current clinical routine and studies, Magnetic Resonance Imaging (MRI) becomes a primary diagnostic evaluation tools and has played an important role in the radiological of assessment of gliomas. MRI has been widely used to evaluate the progression of brain tumor and help decide treatment plan, since it provides critical and detailed structural information of brain. For example, the hyper-intense tissue appearance shown in contrast-enhances T1-weighted presents enhancing tumor, while the hyper-intensity on T2/FLAIR-weighted MRI presents edema tumor. The different proportion of different tumor's sub-components would lead to different kinds of brain tumors and result in different choices of treatment strategy and outcome.

Manual brain tumor segmentation is very difficult, time consuming, and prone to errors. It is highly required to develop accurate and automatic brain tumor segmentation methods. However, it is quite challenging to develop the segmentation techniques because tumors are usually diffused, low contrast and have a lot of extend tentacle-like structures. The automatic segmentation technique will also highly suffer from intensity changes inside and around tumor, variation of tumor shape, size, position and tumor type, all of which are widely variable due to brain tumors can appear in any anywhere in brain with any shape and size. Moreover, the intensity values from the same tumor can vary drastically due to using different types of MRI scanners (e.g. 1.5, 3 or 7 tesla) or different imaging acquisition protocol (field of view value, voxel resolution, gradient strength, b0 value, etc.) from different hospitals. Additionally, it is difficult to distinguish the tumor tissues from the healthy tissues as the boundaries of the tumors are usually fuzzy. All of the above factors greatly increase the difficulties of developing accurate brain tumor segmentation algorithms. Multi-modality MRI (e.g., T1w, T1 contrasted, T2w, FLAIR sequences) are employed to settle the above problems, since almost one unique signature of tumor's sub-component can be given based on imaging information of different modalities. With multi-modality MRI sequences, we propose a novel automatic location sensitive brain tumor segmentation technique in this study.

Our proposed method includes two stages: tumor detection and tumor segmentation. In the first stage, we create a location map for each subject using a region-based position sensitive object detection method to detect tumors in FLAIR sequence. In the second stage, the location map is used to guide the segmentation of deep learning-based brain tumor segmentation techniques. It can be considered as one input of any existing traditional or state-of-the-art deep learning-based brain tumor segmentation method. In this framework, we use the "deepMedic" [5] brain tumor segmentation technique, which achieves top ranking performance in both BRATS and ISLES 2015 Challenges, in the second stage. There are three contributions in this work: (1) we propose a novel automatic and supervised location sensitive method for brain tumor segmentation; (2) our proposed framework can perform both brain tumor detection and segmentation with robust and accurate performance; (3) our proposed model can lead to the faster running time and there is no need of post-processing (e.g., morphological operations, CRF or MRF).

2 Method

In this study, we proposed a novel unsupervised location sensitive brain tumor segmentation method, which conduct brain tumor segmentation with the guide of results from brain tumor detection. As we all know that brain tumor is diffusive and do not have a clear boundary. Moreover, it varies a lot in shape and location. All these make brain tumor segmentation a more challenging task. In order to solve this challenging problem, we proposed a novel location sensitive brain tumor segmentation method which is not only sensitive to tumor location, but also sensitive to tumor shape. The framework of our proposed method is shown in Figure 1. Our proposed method includes two stages. At first, we create a location sensitive map for each subject via a region based object detection method in the first stage; then, the generated location sensitive maps are used to guide the segmentation in the second stage in two ways. The first way is considering the tumor location map as one input channel of the second stage in order to provide location information to guide the segmentation. The second way is taking the location map as a weighted location mask to refine the segmentation results of the second stage by removing tissues outside the map.

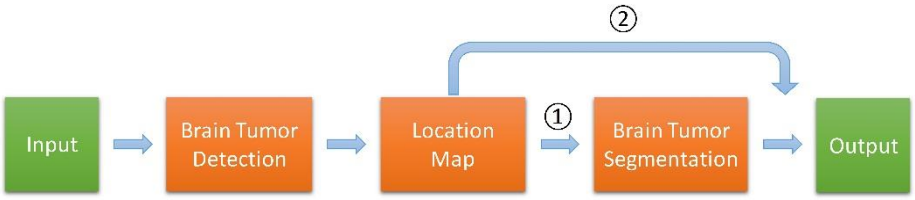


Fig. 1. Frame work of proposed location sensitive brain tumor segmentation method. (① represents the pipeline of the first way, ② represents the pipeline of the second way.)

2.1 Location Sensitive Map

Since the brain tumor is diffusive and can grow in any region of brain, it does not have a clear and inerratic boundary, which makes it much difficult to delineate the abnormal tissues from normal brain tissues. To solve these problems, we propose a novel location sensitive brain tumor segmentation technique which utilizes a novel location sensitive map as guidance for segmentation. Specifically, region-based fully convolutional network (RFCN) [6] is employed to create the ingenious location sensitive map.

In order to generate the location sensitive map, we need to generate the annotation of tumor location for the training dataset since we only have ground truth for segmentation rather than annotation for tumor location. The annotation of tumor location can be generated within the following steps: firstly, the connected components of labels for training dataset are computed; secondly, we get a distance graph by computing the Euclidean distance of each pair of connected components; finally, a breadth-first

search (BFS) algorithm is used to determine the nearby tumor components for each patient and then a rectangle for each tumor component will be drawn as the corresponding annotation. With the annotation of tumor location, we rearrange the 3D FLAIR T2 image into 2D slices with respect to 3 dimensions (i.e., coronal, sagittal, axial), and utilize RFCN to detect tumor locations (coordinates of rectangles which mark the detected tumors) on those 2D slices in each dimension. The 3D location map can be obtained by averaging the score/probability of each rectangle, in three dimensions (i.e., coronal, sagittal, axial), which represents the score/probability of real brain tumor appear at this location. Fig. 2 shows the generated location sensitive maps, and the white areas in the location maps mean that tumors are most likely located at. The white part of location map has high score/probability to be tumor, while the dark part of location map has low score/probability to be tumor. From the Fig. 2, we can see that the generated location sensitive map is able to accurately predict the location of brain tumor and can even capture the outline of tumor.

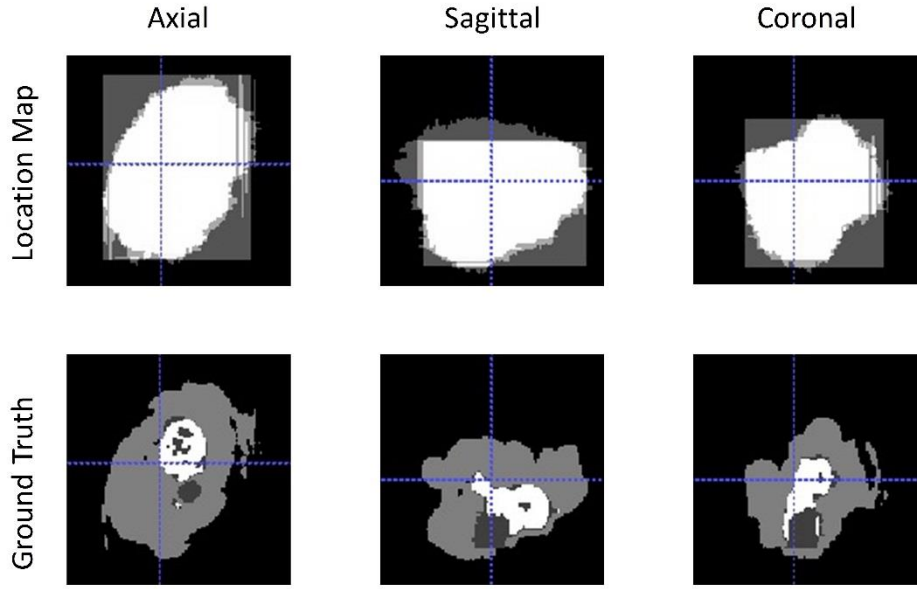


Fig. 2. The generated location map. (The first row is one case of the generated location map in three dimensions (i.e., coronal, sagittal and axial), the second row is ground truth.)

2.2 Location Sensitive Brain Tumor Segmentation

From the Fig. 2, we know that the location map can accurately reflect the position of brain tumor. Also, it is worth to note that the edge of the white part (the part with high score/probability) in the location map is also very similar with the shape of the tumor. With such vital information provided in location sensitive map, we can greatly improve the segmentation accuracy of the existing state-of-the-art brain tumor segmentation methods by using the guidance from the location and shape sensitive map. The

whole pipeline of our proposed location sensitive brain tumor segmentation method is illustrated in Fig. 3. Our proposed method is very simple but efficient, and easy to implement, moreover, it is able to achieve great improvement compared with the state-of-the-art methods.

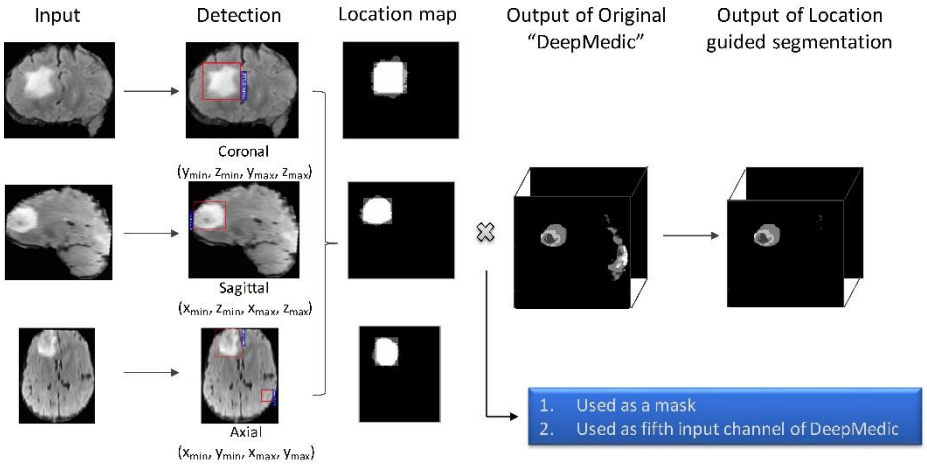


Fig. 3. The framework of our proposed location sensitive brain tumor segmentation method

The key point of our method is that we propose to segment the brain tumor with the guidance of location map which is sensitive to not only tumor location but also tumor outline. Specifically, we propose two approaches to utilize the location map to guide the tumor segmentation:

1. The location map is considered as a weighted location mask and can be applied to the segmentation results from any existing brain tumor segmentation methods. According to subsection 3.1, we know that location map can accurately reflect the tumor location and shape. Since we average the scores/probabilities in three dimensions to get the 3D location map, there may be some outliers which are detected in only on dimension. Therefore, we set a threshold value ($1/3 \approx 0.3333$) on the location map to remove the low score part and make the value of high score part equal to 1. Then the threshold location map is applied onto the segmentation results of any existing methods, which will lead to a better result in which only the segmentation results that on the binarized location map are kept.
2. The location map can be treated as one of the input channel of any existing state-of-the-art brain tumor segmentation methods. Since the location map contains important information about tumor location, it will somehow influence and guide the deep learning model to focus on the part that tumor would most likely located in and further improve the segmentation accuracy of original methods. Fig. 3 shows the effect of brain tumor segmentation method with the guide of location sensitive map. We don't need any other post-processing techniques with this approach.

3 Experimental Results

Dataset and Preprocessing. Our models are training on the data from the 2017 MICCAI BRATS Challenge training data [7], which consists of 275 cases with 210 high grade gliomas (HGG) cases and 75 low grade glioma (LGG) cases. The corresponding reference segmentation are provided for all cases in training data which include three tumor tissue classes this year: necrotic core (labeled 1), edema (labeled 2), and enhancing core (labeled 4). Unlike the data provides in the previous years, the reference segmentation results are all manually segmented by experts and there is no non-enhancing core (labeled 3) labeled this year. The validation data are released at the end of June, and there is no ground truth and tumor grade provided. All results are reported from the online evaluation platform — CBICA's Image Processing Portal. In the evaluation stage, three labels are merged into different sets which include whole tumor (all three classes), the core (classes 1 and 4), and the enhancing tumor (class 4). For each subject, there are four modalities provides which are FLAIR, T1, T1-contrast and T2. The pre-processing of dataset is done by the organizers which includes skull-strip, register to a common space and resample to isotropic 1 〖mm〗 ^3 resolution. The dimensions of each volume are 240 × 240 × 155. We only normalized the data to have a zero mean and unit variance. In order to reduce the problems caused by data imbalance between HGG cases and LGG cases, we augmented the LGG data by reflecting the data with respect to the sagittal axis, and then all data, including the additional augmented LGG data, were augmented through the same way to resolve the small sample problem.

Experimental Setting. In order to develop and evaluate our proposed methods before the releasing of unlabeled validation and test data, we separate the labeled BRATS 2017 training dataset (210/HGG, 150/LGG after the first augmentation only applied on LGG data) into three part: training, validation and test subsets. Firstly, 30 percent of training data are randomly selected and divided into test subset with 63/HGG and 45/LGG, then 30 percent of the remaining data are randomly selected and divided into validation subset with 44/HGG and 31/LGG, and finally the remaining data are split into training subset with 103/HGG and 74/LGG. After the release of validation and test data, we will train our proposed method on the while training set in order to utilize all available imaging information to get better estimation of segmentation results. Our experiments are implemented on an NVIDIA GTX Titan X GPU with 12G memory using cuDNN v5.0.

Experimental Results. In this study, we used the state-of-the-art brain tumor segmentation method "deepMedic" as our baseline, as well as the original segmentation method in our proposed framework. The data we used is the BRATS 2017 which includes 285 multi-model/channel images of patients with 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG). The images were registered to a common space and resampled to isotropic 〖1mm〗 ^3 resolution with dimensions of

240×240×155 by organizers. Each modality of images was normalized by subtracting the mean and dividing by the standard deviation of the intensities within the brain. Training data were augmented via reflection with respect to the sagittal axis, and LGG training data were augmented twice in order to alleviate problems caused by data imbalance. Table 1 shows the dice scores of using difference brain tumor segmentation methods. From Table 1, we can see that our proposed method outperforms the state-of-the-art method "deepMedic".

Table 1. Dice scores of our proposed method compared with the state-of-the-art ("DeepMedic") on the online evaluation platform

Evaluation	DeepMedic	Proposed 1	Proposed 2
Dice_ET	73.04	**75.37**	73.18
Dice_WT	87.94	**89.39**	87.52
Dice_TC	74.17	**74.34**	74.03
Sen_ET	76.10	**76.59**	75.20
Sen_WT	**89.30**	88.90	85.81
Sen_TC	**73.14**	73.13	70.90
Spe_ET	99.82	99.82	99.82
Spe_WT	99.40	99.52	**99.56**
Spe_TC	99.71	99.71	**99.82**
Hau95_ET	6.909	5.510	**4.435**
Hau95_WT	20.71	7.872	**4.464**
Hau95_TC	14.18	11.07	**10.94**

4 Discussion and Conclusions

In this study, we proposed a novel location sensitive brain tumor segmentation method which is robust, efficient and accuracy. Since the brain tumor is diffusive, without clear boundary and can grow anywhere in brain, we firstly generate a location sensitive map for each case which can accurately locate the position of tumors and give an accurate range of tumor regions. The location sensitive maps are then guide the deep learning-based brain tumor segmentation techniques to get accurate delineation of brain tumors. The experimental results show that our proposed method can high improve the dice ratio of segmentation results with the state-of-the-art brain tumor segmentation method as baseline.

References

1. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008.* International journal of cancer, 2010. **127**(12): p. 2893-2917.
2. Ricard, D., et al., *Primary brain tumours in adults.* The Lancet, 2012. **379**(9830): p. 1984-1996.
3. Louis, D.N., et al., *The 2007 WHO classification of tumours of the central nervous system.* Acta neuropathologica, 2007. **114**(2): p. 97-109.
4. Ohgaki, H. and P. Kleihues, *Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas.* Journal of Neuropathology & Experimental Neurology, 2005. **64**(6): p. 479-489.
5. Kamnitsas, K., et al., *Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.* Medical Image Analysis, 2017. **36**: p. 61-78.
6. Li, Y., K. He, and J. Sun. *R-fcn: Object detection via region-based fully convolutional networks*. in *Advances in Neural Information Processing Systems*. 2016.
7. Menze, B.H., et al., *The multimodal brain tumor image segmentation benchmark (BRATS).* IEEE transactions on medical imaging, 2015. **34**(10): p. 1993-2024.
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
9. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
10. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

Dilated Convolutions for Brain Tumor Segmentation in MRI Scans

Marc Moreno Lopez[1] and Jonathan Ventura[2]

[1] Department of Computer Science
University of Colorado Colorado Springs
mmorenol@uccs.edu
[2] Department of Computer Science
University of Colorado Colorado Springs
jventura@uccs.edu

Abstract. In this work, we present a novel method to segment brain tumors using a dilated Fully Convolutional Network [7]. An accurate brain tumor segmentation is key for a patient to get the right treatment and for the doctor who must perform surgery. Due to the genetic differences that exist in different patients, even between the same kind of tumor, an accurate segmentation is crucial.To evaluate our algorithm we use the evaluation tool from the Brain Tumor Segmentation challenge, BraTS from 2017.

Keywords: Deep learning · Brain tumor segmentation · Dilated convolutions · Fully convolutional network

1 Introduction

Cancer is one of the leading causes of death in the world. In the US, cancer is the 2nd leading cause exceeded only by heart disease [1]. To put it in perspective, one out of every four deaths in the US is caused by cancer. Due to this high-death ratio, scientist all over the world have tried to find a cure for cancer. In this work, my intention is to find a faster and more efficient way to detect cancer in time [8, 11, 12, 13].

According to the National Cancer Institute [2], checking for cancer (or for conditions that may become cancer) in people who have no symptoms is called screening. Screening can help doctors find and treat several types of cancer early. Early detection is important because when abnormal tissue or cancer is found early, it may be easier to treat. By the time symptoms appear, cancer may have begun to spread and is harder to treat. Several screening tests have been shown to detect cancer early and to reduce the chance of dying from that cancer. But it is important to keep in mind that screening tests can have potential harms as well as benefits. Some screening tests may cause bleeding or other health problems.

Screening tests can have false-positive results – the test indicates that cancer may be present even though it is not. False-positive test results can cause

anxiety and are usually followed by additional tests and procedures that also have potential harms. Screening tests can also have false-negative resultsthe test indicates that cancer is not present even though it is. False-negative test results may provide false reassurance, leading to delays in diagnosis and possibly causing an individual to put off seeking medical care even if symptoms develop[14]. Finally, overdiagnosis is also possible. This happens when a screening test correctly shows that a person has cancer, but the cancer is slow growing and would not have harmed that person in his or her lifetime. Treatment of such cancers is called overtreatment.

In this work, we want to combine a screening method, such as Magnetic Resonance Imaging, MRI, with the latest technology on machine learning, deep learning. Deep Learning is a new area of Machine Learning research, which has been introduced with the objective of moving Machine Learning closer to one of its original goals: Artificial Intelligence. There have been some approaches using this method [9] [6], but they havent presented an approach that can beat human performance.

A correct segmentation is key for many reasons. The most important one is so that the patient can get the best possible treatment. An accurate tumor quantification is needed so that the patient gets the amount of treatment that he needs. A rightful segmentation is crucial too in life-threatening cases. These are cases where the tumor is next to or on top of one of the cerebellum, or to similar sensitive parts. Therefore, we need to do a correct segmentation, especially in the boundaries between tumor and edema. This last part is central when planning a brain tumor extraction. Doctors need to know what they are facing before performing any surgery.

One of the main factors for a correct image segmentation is having enough images to train the network [5]. When working with medical images this can become an issue.

We are going to work towards a correct segmentation using a Deep Convolutional Neural Network with dilated filters instead of pooling filters. Moreover, instead of training it with the whole image, we will use a patch-based training approach.

2 Methods

We apply a fully convolutional network approach in order to produce a per-pixel segmentation output. Our network is applied to each slice in a scan separately.

2.1 Dilation

An issue with traditional convolutional neural network architectures that use max pooling is that they downsample the image and thus produce a segmentation with resolution smaller than the input size.

In [15], Yu et al. develop a new convolutional network module that is specifically designed for dense prediction. They present a model that uses dilated

Layer	Type	Configuration	Dilation
1	Convolutional	3x3x1x32	1
2	Batch normalization		
3	ReLU		
4	Convolutional	3x3x32x32	1
5	Batch normalization		
6	ReLU		
7	Convolutional	3x3x32x32	2
8	Batch normalization		
9	ReLU		
10	Convolutional	3x3x32x32	4
11	Batch normalization		
12	ReLU		
13	Convolutional	3x3x32x32	8
14	Batch normalization		
15	ReLU		
16	Convolutional	3x3x32x32	16
17	Batch normalization		
18	ReLU		
19	Convolutional	3x3x32x32	1
20	Batch normalization		
21	ReLU		
22	Convolutional	1x1x32x4	1

Table 1. Configuration of the CNN

convolutions. This model is designed to systematically aggregate multi-scale contextual information without losing resolution. All their work is based on the fact that dilated convolutions support exponential expansion of the receptive field without losing resolution or coverage.

Let $F : \mathbb{Z}^2$ be a discrete function. Let $\Omega_r = [-r, r]^2\mathbb{Z}^2$ and let $k : \Omega_r \to \mathbb{R}$ be a discrete filter of size $(2r+1)^2$. The discrete convolution opreator can be defined as

$$(F * k)(p) = \sum_{s+t=p} F(s)k(t) \tag{1}$$

We now generalize this operator. Let l be a dilation factor and let $*_l$ be defined as

$$(F *_l k)(p) = \sum_{s+lt=p} F(s)k(t) \tag{2}$$

We will refer to $*_l$ as a dilated convolution or and l-dilated convolution.

2.2 Patch-based training

The dataset exhibits severe class imbalance, i.e. the tumor pixels are vastly outnumbered by the non-tumor pixels. This poses a problem when training the network, because the non-tumor pixels influence the total loss function much more strongly than the tumor pixels.

To address this issue, we adopt a patch-based training approach. During training, we randomly sample patches from the images to form a batch. Each patch is exactly the size of receptive field of the network.

We sample patches using a uniform distribution over the classes. In other words, we ensure that each batch has the same number of examples of each class. This effectively remedies the class imbalance that would be caused by simply randomly sampling the patches. Moreover, we don't sample patches from pixels with zero intensity.

Because the network is fully convolutional, we can use the same network trained on patches to test on images. At test time, we use whole images as input to the network to produce the full-resolution output.

3 Results

3.1 Implementation details

To develop the experiment, we used Keras for Python 2.7. All training was done in one machine with Ubuntu and Nvidia Titan X card with 12 GB of memory.

To train the network, we used a batch size of 120 patches, 1000 batches per epoch and 1000 epochs. We used an Adagrad optimizer [4] with a learning rate of 0.01. One training epoch takes 146 seconds.

3.2 Evaluation on validation dataset

To evaluate our algorithm, we used the BRATS evaluation tool. To measure the performance of the algorithms, we have to evaluate recall (or sensitivity), specificity and the dice score (also called F1) [3]. Recall (also known as sensitivity) is the fraction of relevant instances that are retrieved, therefore it measures the proportion of positives that are correctly identified as such. Specificity measures the proportion of negatives that are correctly identified as such. The dice score or F1, is a is a statistic used for comparing the similarity of two samples. It is calculated using the precision and recall parameters.

DICE	Sensitivity	Specificity	Hausdorff95
0.52788	0.68864	0.99495	32.01849

Table 2. Mean results for Enhanced Tumor

DICE	Sensitivity	Specificity	Hausdorff95
0.63685	0.71717	0.90964	36.46434

Table 3. Mean results for Tumor Core

DICE	Sensitivity	Specificity	Hausdorff95
0.73717	0.86329	0.97142	43.38166

Table 4. Mean results for Whole Tumor

In Figures 1, 2 and 3 we have included some examples of segmented MRI. Figure 1 and 2 are good examples of how the segmentation should look like. In FIgure 3 there are some issues with the edema detection that we will discuss later.

4 Discussion

Based on the results that we have obtained, we can say that the balancing of data is key to avoid getting too many pixels labels as background or normal tissue. In Figure 1 and Figure 2, we can see that the segmentation is correct, but we still need to address some issues. We have a balanced algorithm, but we still have room for improvement until we reach human performance.

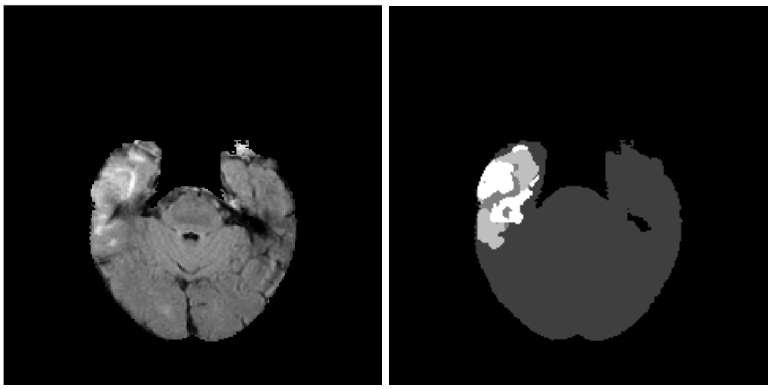


Fig. 1. MRI with its corresponding segmentation

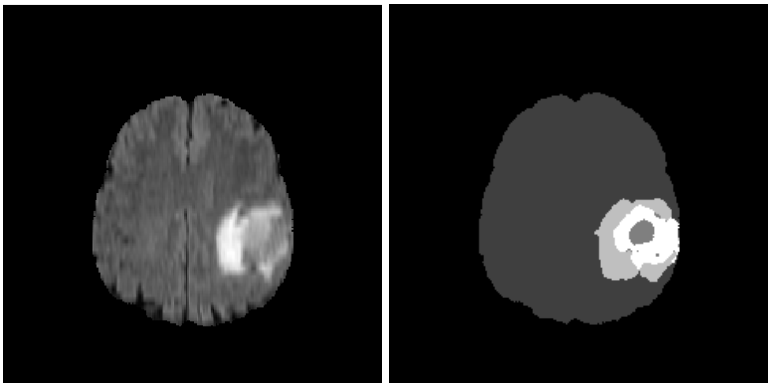


Fig. 2. MRI with its corresponding segmentation

One of our main areas of improvement is edema detection. There are some issues with the detection of edema, since too many pixels are detected as edema and this lowers the precision of the algorithm. This is caused by the similarity of some of the edema pixels with adjacent brain pixels. In Figure 3 we can see a good example of this problem.

As for the tumor pixels, we can see a problem with boundaries or edges. The edges between edema and tumor core are not segmented well enough. Many tumor pixels are labeled as edema. And the same thing happens with enhanced tumor and tumor pixels.

As for enhanced tumor, the algorithm behaves similarly for most of the cases, and it has a regular behavior in most of the images. However, there are some issues with enhanced tumor detection, where the dice score is low. We think that this is due to the low number of enhanced tumor pixels in some of the MRI. In these cases, the MRI contains few Enhanced Tumor pixels (10-20) and our algorithm misses all or most of them.

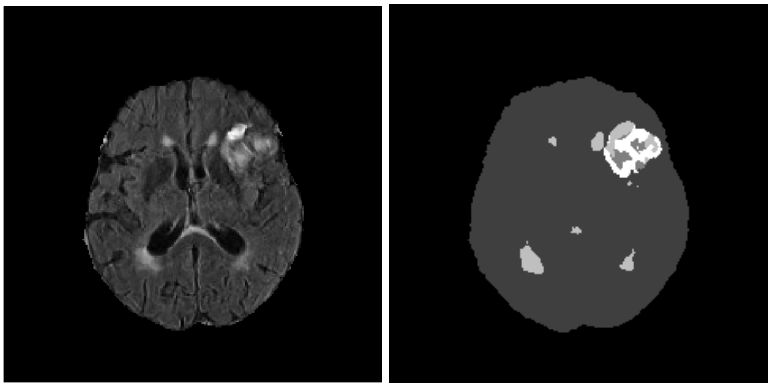


Fig. 3. MRI with its corresponding segmentation

5 Conclusions

Results are encouraging, especially for whole tumor and lead us to believe that, even though we still have work to do until we reach our objective, we can keep working in this direction. The lack of medical data is one of the main problems that we have encountered in this project. To publish medical data, the three parts involved, doctor, institution and patient, must agree to publish it. Therefore, if you arent working with a medical institution, its difficult to obtain lots of medical data or even trustworthy data. As a future work we would like to implement a data augmentation model using elastic image deformation, to overcome this scarcity of data.

Finding the right network configuration to work with such a small amount of data was an enormous challenge. We went through many different networks and different configurations to try to find the most suitable network for our needs. The dilated convolution is effective for brain tumor segmentation to introduce context without losing output spatial resolution and it is an interesting direction to explore in the future. One of the main ideas we want to apply is combining well-known structures like U-Net [10] with the dilation model. Another interesting approach due to its effectiveness for segmentation, is the one proposed by Yu et al. in [16], where they use Dilated Residual Networks for image segmentation.

To increase the accuracy when detecting Enhanced Tumor, we could address this by changing the balance of classes during patch sampling and/or adding class weights.

Brain tumor segmentation isn't an easy task. Due to the genetics of cancer, it remains being a task for which the doctors help is needed. However, they dont have to do all by themselves, since with tools like the one that we have designed in this work, we can help them and we can contribute in the fight against cancer.

References

[1] http://www.who.int/mediacentre/factsheets/fs297/en/.

[2] https://www.cancer.gov/.

[3] Jesse Davis and Mark Goadrich. "The Relationship Between Precision-Recall and ROC Curves". In: *Proceedings of the 23rd International Conference on Machine Learning*. ICML '06. Pittsburgh, Pennsylvania, USA: ACM, 2006, pp. 233–240. ISBN: 1-59593-383-2. DOI: 10.1145/1143844.1143874. URL: http://doi.acm.org/10.1145/1143844.1143874.

[4] John Duchi. "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". In: *Journal of Machine Learning Research* 12 (2011), pp. 2121–2159.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[6] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. "Brain tumor segmentation with deep neural networks". In: *Medical Image Analysis* 35 (2017), pp. 18–31. arXiv: arXiv:1505.03540.

[7] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully Convolutional Networks for Semantic Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2015). arXiv: arXiv:1605.06211.

[8] Bjoern Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elisabeth Gerstner, Marc-Andre Weber, Tal Arbel, Brian Avants, Nicholas Ayache, Patricia Buendia, Louis Collins, Nicolas Cordier, Jason Corso, Antonio Criminisi, Tilak Das, Herve Delingette, Cagatay Demiralp, Christopher Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan Iftekharuddin, Raj Jena, Nigel John, Ender Konukoglu, Danial Lashkari, Jose Antonio Mariz, Raphael Meier, Sergio Pereira, Doina Precup, S. J. Price, Tammy Riklin-Raviv, Syed Reza, Michael Ryan, Lawrence Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos Silva, Nuno Sousa, Nagesh Subbanna, Gabor Szekely, Thomas Taylor, Owen Thomas, Nicholas Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)". In: *IEEE Transactions on Medical Imaging* 34.10 (2015), pp. 1993–2024.

[9] Sergio Pereira, Adriano Pinto, Victor Alves, and Carlos A Silva. "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images". In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1240–1251.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (2015),

pp. 234–241. ISSN: 16113349. DOI: 10.1007/978-3-319-24574-4_28. arXiv: 1505.04597.

[11] Bakas S, Akbari H, Sotiras A, Rozycki M Bilello M, Kirby J, Freymann J, Farahani K, and Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features". In: *Nature Scientific Data* (2017). [In press].

[12] Bakas S, Akbari H, Sotiras A, Rozycki M Bilello M, Kirby J, Freymann J, Farahani K, and Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection". In: *The Cancer Imaging Archive* (2017). DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.

[13] Bakas S, Akbari H, Sotiras A, Rozycki M Bilello M, Kirby J, Freymann J, Farahani K, and Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection". In: *The Cancer Imaging Archive* (2017). DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.

[14] R. Weinberg. *The Biology of Cancer, Second Edition*. Taylor & Francis Group, 2013. ISBN: 9781317963462. URL: https://books.google.com/books?id=MzMmAgAAQBAJ.

[15] Fisher Yu and Vladen Koltun. "Multi-Scale Context Aggregation by Dilated Convolutions". In: *International Conference on Learning Representations (ICLR)* (2016). arXiv: arXiv:1511.07122v3.

[16] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. "Dilated Residual Networks". In: *Computer Vision and Pattern Recognition (CVPR)* (2017). arXiv: arXiv:1705.09914.

SIBIA-GlS: Scalable Biophysics-Based Image Analysis for Glioma Segmentation

Andreas Mang[2], Sameer Tharakan[1], Amir Gholami[1], Naveen Himthani[1], Shashank Subramanian[1], James Levitt[1], Muneeza Azmat[1], Klaudius Scheufele[3], Miriam Mehl[3], Christos Davatzikos[4] Bill Barth[4], and George Biros[1]

[1] Institute for Computational Engineering and Science, U Texas, Austin
[2] Department of Mathematics, University of Houston
[3] Department of Computer Science, University of Stuttgart
[4] Department of Radiology, University of Pennsylvania
[5] Texas Advanced Computing Center, U Texas, Austin

1 Introduction

SIBIA-GlS is baaed on Gooya et al [8] and on SIBIA (Scalable Integrated Biophysics-based Image Analysis) [6], a set of algorithms and software for biophysics-based image analysis. The SIBIA-GlS pipeline comprises two main steps: the ***ML-step*** in which we use supervised machine learning to create probability maps for the target classes ("whole tumor", "edema", "tumor core", and "enhancing tumor"); and the ***SIBIA-step*** in which we combine these probabilities with a biophysical model of tumor growth coupled with large-scale diffeomorphic registration to implicitly impose spatial correlations Besides [8], our work borrows from [10] and the algorithms and workflows in [22] that summarizes the BRATS 2012 and 2013 competitions. In this work, we used the training, validation, and testing datasets of the BRATS 2017 competition [1–3]. The BRATS 17 competition has several metrics to assess the quality of a segmentation. In this paper we just focus on the *"whole tumor" Dice score*. We achieve 0.87 median Dice score (0.84 mean) and 0.84 median Dice score (0.83 mean) for the training and validation images respectively. The histograms are shown below.

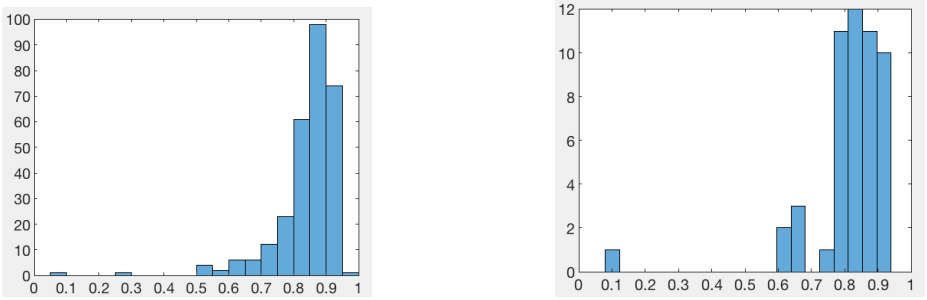


Fig. 1 *Dice histograms for "whole-tumor" label. Left: training (285 images); Right: validation (46 images)*

2 Methods

In this section, we discuss the methodology and the overall formulation for SIBIA-GlS. First, a few words regarding preprocessing. We normalize the intensities by centering their mean after removing the bottom and top 1% outliers (for each modality separately). Then, we affinely register all images to a normal (segmented) probabilistic atlas image using an L^2 similarity measure and the T1 patient image. All the classification steps take in the atlas space.

Notation: With boldface, we denote vector fields; with normal face fonts we denote scalar fields. We define the following probability fields: π_G (gray matter), π_W (white matter), π_F (cerebrospinal fluid and ventricles), π_{ED} (edema), π_{TC} (tumor core), π_{EN} (enhancing tumor), and the π_{WT} (whole tumor) probability. We also define $\boldsymbol{\pi}^A$ to be the vector probability for a reference brain (that may or may not have a tumor, depending on the context), so that $\boldsymbol{\pi}^A = \{\pi_G, \pi_W, \pi_F, \pi_{WT}\}$. We define $\boldsymbol{\pi}^S$ to be the patient vector probability map.

We use SIBIA-GlS for the "whole-tumor" label and then binary classification for the other labels. SIBIA-GlS consists of four main components. First, the ***inverse tumor growth model*** is used to biophysically constrain the ML classifier. Second, the ***registration problem*** used for atlas-based segmentation (with or whithout tumor). Third, ***supervised machine learning*** framework that provides the initial "whole-tumor" probability. Fourth, the ***overall coupling*** that combines the three first components to produce a final "whole-tumor" probability. We briefly describe these components below.

Tumor model: Given $\boldsymbol{\pi}^A(0) := \boldsymbol{\pi}^A(x, 0)$, the probability map of a healthy brain (i.e., the *"atlas"*), the *forward* tumor operator $\boldsymbol{\mathcal{T}}$ is given by

$$\boldsymbol{\pi}^A(1) := \boldsymbol{\pi}^A(\boldsymbol{x}, 1) = \boldsymbol{\mathcal{T}}(\boldsymbol{g}, \boldsymbol{\pi}^A(0)). \tag{1}$$

Here $\boldsymbol{g}$ are ***tumor growth model parameters*** that control the tumor growth. $\boldsymbol{\pi}^A(0)$ comprises π_W, π_G, and π_F, where $\boldsymbol{\pi}^A(1)$ comprises π_W, π_G, π_F, and π_{WT}. To simplify the notation, we suppress the dependence on the normal atlas $\boldsymbol{\pi}^A(0)$; we simply write $\boldsymbol{\pi}^A = \boldsymbol{\mathcal{T}}(\boldsymbol{g})$, where $\boldsymbol{g}$ is the vector of parameters that control the tumor growth dynamics. In this work we used a simple reaction-diffusion model and g is the initial condition for the tumor parameterized by 125 Gaussians.

In the ***inverse tumor problem***, given $\boldsymbol{\pi}^A_*$ (data with tumor) and $\boldsymbol{\pi}^A(0)$ (a normal brain), we solve an optimization problem for $\boldsymbol{g}$: $\min_{\boldsymbol{g}}(\boldsymbol{\pi}^A(1) - \boldsymbol{\pi}^A_*)^2$, were $\boldsymbol{\pi}^A(1)$ is given by (1). We have omitted (due to space limitations) an additional regularization term [5–7, 11, 16] that controls the reconstruction of g.

Registration: We use a velocity-based formulation for diffeomorphic registration [4, 9, 12–15]. Given a vector field $\boldsymbol{\pi}^S(0)$ and a velocity field $\boldsymbol{v}$, the *forward* image registration problem computes a deformation of $\boldsymbol{\pi}^S(0) := \boldsymbol{\pi}^S(x, 0)$, let's call it $\boldsymbol{\pi}^S(1) := \boldsymbol{\pi}^S(x, 0)$. We abstract this operation using $\boldsymbol{\mathcal{R}}$. That is,

$$\boldsymbol{\pi}^S(1) = \boldsymbol{\mathcal{R}}(\boldsymbol{v}, \boldsymbol{\pi}^S(0)). \tag{2}$$

In the ***inverse registration problem***, we're given two vector fields $\boldsymbol{\pi}^S(0)$ and $\boldsymbol{\pi}^S_*$ and we seek to compute $\boldsymbol{v}$ such that the difference between $\boldsymbol{\pi}^S(1)$ and

$\boldsymbol{\pi}_*^S$ is a small as possible. Formally, $\min_{\boldsymbol{v}}(\boldsymbol{\pi}^S(1) - \boldsymbol{\pi}_*^S)^2$ such that $\boldsymbol{\pi}^S(1)$ is given by (2). We have omitted the necessary regularization for the velocity [13].

Supervised classification: We first identify "whole-tumor" voxels using binary classification. Then, we classify the "whole-tumor" voxels to "tumor core" and "edema". Lastly, we classify the "tumor-core" voxels to "enhancing" and "non-enhancing". All these are binary classifications. As featured, we use 288 2D Gabor features per voxel, and we use 50,000,000 training voxels. For the machine learning step we use nearest neighbor classification (using an in-house code [20,21]) for the "whole tumor" label. For the binary classifications to distinguish edema, and enhancing tumor we used 25,000,000 points and used LightGBM, an open-source, random-forest classifier. Both classifiers return probability maps. The whole-tumor probability maps are passed to the next step to introduce spatial correlation. Then, the updated π_{WT}, along with the π_{TC} and π_{EN} from LightGBM are combined and threshold to produce the final labels.

Coupled formulation (SIBIA): The inputs to our problem are $\boldsymbol{\pi}^S(0)$ (initial patient probability map) and $\boldsymbol{\pi}^A(0)$ (normal atlas without tumor). The outputs are $\boldsymbol{g}$ and $\boldsymbol{v}$, and $\boldsymbol{\pi}^S(1)$ and $\boldsymbol{\pi}^A(1)$, which contains π_{WT}-the main output of the SIBIA-GlS part in the pipeline. Formally, the optimization problem (omitting, for notational simplicity, regularization terms for g and u) is given by

$$\begin{aligned} \min_{\boldsymbol{v},\,\boldsymbol{g}} \ (\boldsymbol{\pi}^A(1) - \boldsymbol{\pi}^S(1))^2 \qquad \text{such that} \quad & \boldsymbol{\pi}^A(1) = \boldsymbol{\mathcal{T}}(\boldsymbol{g}, \boldsymbol{\pi}^A(0)), \\ & \boldsymbol{\pi}^S(1) = \boldsymbol{\mathcal{R}}(\boldsymbol{v}, \boldsymbol{\pi}^S(0)). \end{aligned} \tag{3}$$

Here we assume we have probability maps $\boldsymbol{\pi}^A(0)$ and $\boldsymbol{\pi}^S(0)$ for both atlas and patient images and then we try to match them. That means that we need to segment gray matter, white matter, and CSF and ventricles, in addition to tumor probabilities. We need this information to be able to calibrate the tumor growth model through the determination of $\boldsymbol{g}$. Tumor doesn't grow in ventricles and material properties differ in white and grey matter. We obtained these tissue-type probability maps using probabilistic atlas segmentation averaging registration with 10 normal brains and two large-deformation diffeomorphic registration algorithms, DEMONS [17], and our own CLAIRE [14].

3 Results

3.1 SIBIA results

First, we report results for the SIBIA component, assuming a correct and known π_{WT} and shows how we can grow a tumor in atlas space and then map it to a given tumor in the patient space, to test our joint inversion / segmentation approach. By adjusting the regularization parameters, we can adjust the regularization provided by SIBIA. We report representative results in Figure 2. SIBIA can effectively match the given segmentation. All SIBIA runs are in reduced 128^3 resolution. The total run time for SIBIA is under 1 minute using a 10-node configuration (2-socket Xeon E5-2690 v3 (Haswell) with 12 cores/socket).

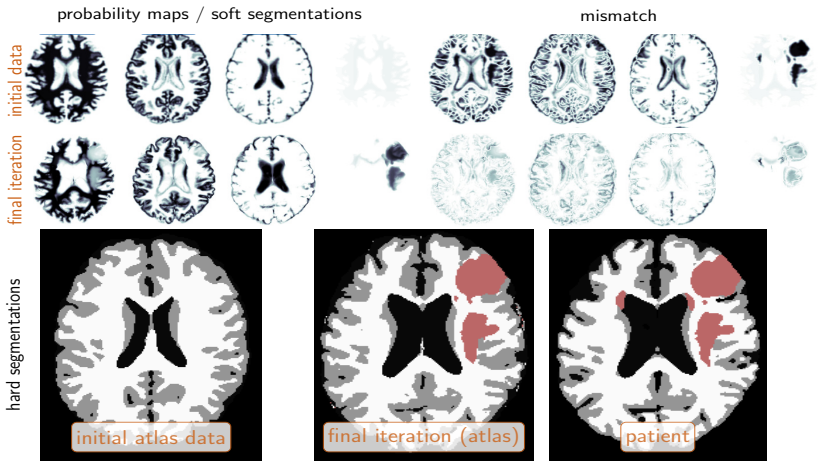


Fig. 2 *The top row shows the initial configuration (4 images to the left: probability maps* $\boldsymbol{\pi}^A(0)$ *(in particular, from left to right,* π_G*,* π_W*,* π_F *and* π_{WT}*) at iteration zero; 4 images to the right: mismatch (pointwise residual) between* $\boldsymbol{\pi}^A(1)$ *and* $\boldsymbol{\pi}^S(1)$ *at iteration zero. The second row shows the same configuration at the final iteration of our coupled tumor inversion and registration scheme. The three images on the bottom show the corresponding hard segmentation (left: initial label maps given for the atlas image; middle: deformed configuration of the atlas image (registered to the patient data) with simulated tumor; right: synthetic patient image. The obtained atlas based segmentation (middle image) and the ground truth segmentation for the patient are very similar.*

3.2 BRATS Validation results

In figure 3, we present segmentation for seven different BRATS17 cases (from the training set) for which we get different Dice scores. We can recover multifocal tumors and quite complex shapes, but we do have cases with quite bad dice scores both because of false negatives and positives. Overall, there are several technical reasons that create problems in our algorithm and we will discuss them in a longer version of this paper.

4 Discussion

We presented preliminary results for our joint formulation for combining atlas-based and machine learning-based segmentation. Below we list some observations on our efforts.

- The processing time per patient is about one hour using 10 dual-socket x86 nodes. The most expensive parts is the nearest neighbor classifier and the 20 diffeomorphic registrations for the gray matter, white matter, and CSF. Both of them are in full resolution ($204^2 \times 155$). SIBIA is much faster because we use 128^3 resolution.
- The nearest-neighbor classifier requires 8 nodes and it uses distributed memory parallelism based on MPI so that it can handle the problem of finding nearest neighbors. Notice for each brain we have to find the neighbors of 1.5M voxels in a dataset of 50M voxels. SIBIA and CLAIRE use MPI. The

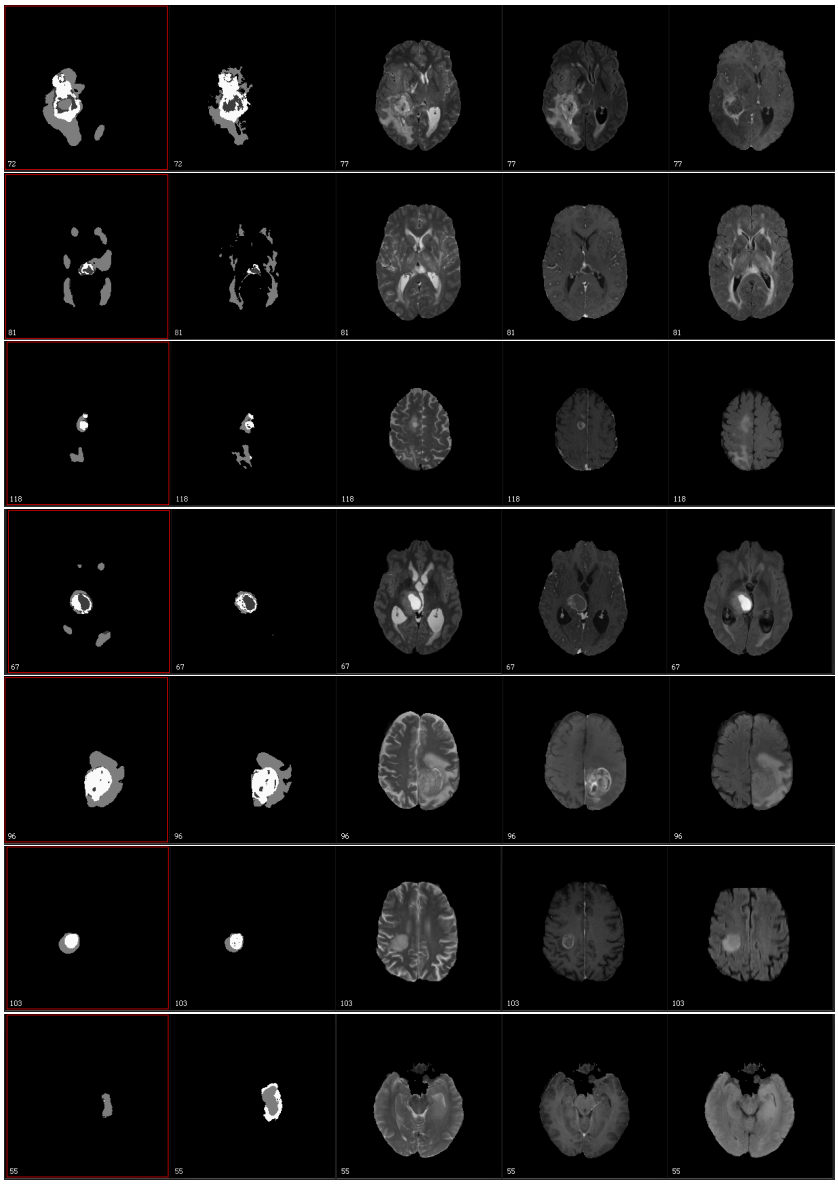


Fig. 3 *Images from the training set:* ***Top-to-bottom:*** *different BRATS brains: CBICA_ABO (0.87), CBICA_AWH (0.81), CBICA_ATB (0.81), CBICA_ATX (0.53), TCIA_242 (0.93), 2013_26 (0.53), TCIA_177 (0.23). In parenthesis, we report the "whole-tumor" Dice score.* ***Left-to-right****: In the first two columns we show our SIBIA-GlS segmentation (outlined with a red box) followed by the ground truth segmentation (provided by the BRATS17 organizers). In the last three columns, we show the T2, T1ce and FLAIR MRI images for each case. In the segmentation images (the first two columns), white is enhancing tumor, light gray is edema, and dark gray is non-enhancing tumor.*

nearest neighbors were run on the Stampede 2 system, CLAIRE and SIBIA on Lonestar 5 system, and combining the segmentations on the Maverick system, all at TACC.

- The primary classification is key for SIBIA. We need an acceptable initial guess for the segmentation based on machine learning for good results. SIBIA cannot correct really bad initial segmentations.
- Our approach is fully automatic. We only need to specify algorithm parameters (for the inverse solvers, iterative solvers, etc) . The results are not overly sensitive to these parameters. The only important parameter is the thresholding parameter for π_{WT}, which is used to produce a hard segmentation. Although 0.5 seems obvious, using 0.3 produces much better results (meaning that a voxel is labeled as "whole-tumor" if $\pi_{WT} > 0.3$).
- Our approach integrates biophysical simulations with machine learning, optimization, and image analysis. It can provide bio-physical parameters (growth rate, mass effect, and others) that might be critical for clinical studies to assess the current and/or future state of an individual patient.
- The solvers for the individual building blocks are based on state-of-the-art technology in scientific computing [6, 12, 14, 18–20].
- Unlike [8], we do not iterate to update the SIBIA segmentation. That's a limitation of our scheme and will address in future work.
- High-grade and low-grade gliomas have different characteristics but we have not taken this into account. We simple merged all the training data.

5 Conclusion

We have presented preliminary results for SIBIA-GlS—a new framework for biophysics-based image analysis for glioma segmentation. We demonstrated that our approach yields promising results. However, several issues remain open. We obtain excellent results for SIBIA-GlS if the initial proposal for the segmentation does not contain significant noise. Improving on this initial segmentation will be key for our future work. In addition to that we will extend our biophysical model to, e.g., include edema and mass effect. We expect that this will significantly improve our current results.

Acknowledgments

This material is based upon work supported by NSF grant CCF-1337393; by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Numbers DE-SC0010518 and DE-SC0009286; and by NIH grant 10042242. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the DOE, NIH, and NSF. Computing time on the Texas Advanced Computing Centers Stampede system was provided by an allocation from TACC and the NSF. We're particularly grateful to the stuff of TACC for providing the resources that enabled us to participate in this competition.

References

1. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017), in press
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels for the pre-operative scans of the TCGA-GBM collection (2017), http://doi.org/10.7937/k9/tcia.2017.klxwjj1q
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation labels for the pre-operative scans of the TCGA-LGG collection (2017), http://doi.org/10.7937/k9/tcia.2017.gjq7r0ef
4. Chen, K., Lorenz, D.A.: Image sequence interpolation using optimal control. Journal of Mathematical Imaging and Vision 41, 222–238 (2011)
5. Gholami, A., Mang, A., Biros, G.: An inverse problem formulation for parameter estimation of a reaction-diffusion model of low grade gliomas. Journal of Mathematical Biology 72(1), 409–433 (2016)
6. Gholami, A., Mang, A., Scheufele, K., Davatzikos, C., Mehl, M., Biros, G.: A framework for scalable biophysics-based image analysis. In: Proc ACM/IEEE Conference on Supercomputing (2017), (accepted)
7. Gholami, A., Scheufele, K., Davatzikos, C., Mang, A., Mehl, M., Biros, G.: A framework for scalable biophysics-based image analysis. In: Proceedings of SC17. The SCxy Conference series, ACM/IEEE, Denver, Colorado (November 2017)
8. Gooya, A., Pohl, K.M., Bilello, M., Cirillo, L., Biros, G., Melhem, E.R., Davatzikos, C.: GLISTR: Glioma image segmentation and registration. Medical Imaging, IEEE Transactions on 31(10), 1941–1954 (2013)
9. Hart, G.L., Zach, C., Niethammer, M.: An optimal control approach for deformable registration. In: Proc IEEE Conference on Computer Vision and Pattern Recognition. pp. 9–16 (2009)
10. Hogea, C., Davatzikos, C., Biros, G.: Brain-tumor interaction biophysical models for medical image registration. SIAM Journal on Scientific Computing 30(6), 3050–3072 (2008)
11. Hogea, C., Davatzikos, C., Biros, G.: An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. Journal of Mathematical Biology 56(6), 793–825 (2008)
12. Mang, A., Biros, G.: An inexact Newton–Krylov algorithm for constrained diffeomorphic image registration. SIAM Journal on Imaging Sciences 8(2), 1030–1069 (2015)
13. Mang, A., Biros, G.: Constrained H^1-regularization schemes for diffeomorphic image registration. SIAM Journal on Imaging Sciences 9(3), 1154–1194 (2016)
14. Mang, A., Gholami, A., Biros, G.: Distributed-memory large-deformation diffeomorphic 3D image registration. In: Proc ACM/IEEE Conference on Supercomputing. No. 72 (2016)
15. Mang, A., Ruthotto, L.: A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity- preserving diffeomorphic image registration. SIAM Journal on Scientific Computing (2017), in press
16. Mang, A., Toma, A., Schuetz, T.A., Becker, S., Eckey, T., Mohr, C., Petersen, D., Buzug, T.M.: Biophysical modeling of brain tumor progression: from unconditionally stable explicit time integration to an inverse problem with parabolic PDE constraints for model calibration. Medical Physics 39(7), 4444–4459 (2012)

17. Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N.: ilogdemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. International Journal of Computer Vision 92(1), 92–111 (2011)
18. March, W.B., Xiao, B., Biros, G.: Askit: Approximate skeletonization kernel-independent treecode in high dimensions. SIAM Journal on Scientific Computing 37(2), A1089–A1110 (2015)
19. March, W.B., Xiao, B., Yu, C., Biros, G.: An algebraic parallel treecode in arbitrary dimensions. In: Proceedings of IPDPS 2015. 29th IEEE International Parallel and Distributed Computing Symposium, Hyderabad, India (May 2015), http://dx.doi.org/10.1109/IPDPS.2015.86
20. March, W.B., Xiao, B., Yu, C.D., Biros, G.: Askit: An efficient, parallel library for high-dimensional kernel summations. SIAM Journal on Scientific Computing 38(5), S720–S749 (2016), http://dx.doi.org/10.1137/15M1026468
21. March, W.B., Yu, C., Xiao, B., Biros, G.: LIBASKIT home page (2015), http://padas.ices.utexas.edu/libaskit
22. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The multimodal brain tumor image segmentation benchmark (BRATS). Medical Imaging, IEEE Transactions on 34(10), 1993–2024 (2015)

Pooling-free fully convolutional networks with dense skip connections for semantic segmentation, with application to brain tumor segmentation

Richard McKinley[1], Alain Jungo[2], Roland Wiest[1], Mauricio Reyes[2]

[1] Support Centre for Advanged Neuroimaging Inselspital, University Hospital, University of Bern, Switzerland
[2] Institute For Surgical Technology and Biomechanics, University of Bern, Switzerland,

Abstract. Segmentation of medical images requires multi-scale information, combining local boundary detection with global context. State-of-the-art convolutional neural network (CNN) architectures for semantic segmentation are often composed of a downsampling path which computes features at multiple scales, followed by an upsampling path, required to recover those features at the same scale as the input image. Skip connections allow features discovered in the downward path to be integrated in the upward path. The downsampling mechanism is typically a pooling operation. However, pooling was introduced in CNNs to enable translation invariance, which is not desirable in segmentation tasks. For this reason, we propose an archgitecture, based on the recently proposed Densenet, for semantic segmentation, in which pooling has been replaced with dilated convolutions. We present results on the validation dataset of the Multimodal Brain Tumor Segmentation Challenge 2017.

Introduction

We present a network architecture for semantic segmentation, heavily inspired by the recent Densenet architecture for image classification [1], in which pooling layers are replaced by heavy use of dilated convolutions [2]. Densenet employs dense blocks, in which the ouput of each layer is concatenated with its input before passing to the next layer. A typical Densenet architecture consists of a number of dense blocks separated by transition layers: the transition layers contain a pooling operation, which allows some degree of translation invariance and downsamples the feature maps. A Densenet architecture adapted for semantic segmentation was presented in [3], which adopted the now standard approach of U-net [4]: a downsampling path, followed by an upsampling path, with skip connections passing feature maps of the sample spatial dimension from the downsampling path to the upsampling path.

In this paper, we describe an alternative architecture adapting Densenet for semantic segmentation: in this architecture, which we call DeepSCAN, there are no transition layers and no pooling operations. Instead, dilated convolutions are used to increase the receptive field of the classifier. The absence of transition layers means that the whole network can be seen as a single dense block, enabling gradients to pass easily to the deepest layers.

We describe the general architecture of DeepSCAN, plus the particular features of the network as applied to brain tumor segmentation, and report preliminary results on the validation portion of the BRATS 2017 dataset.

Dense blocks

The fundamental unit of a densenet architecture is the densely connected block, or dense block. In such a block, the output of each layer (where a layer here means some combination of convolutional filters, nonlinearities and perhaps batch normalization) is concatenated to its input before passing to the next layer.

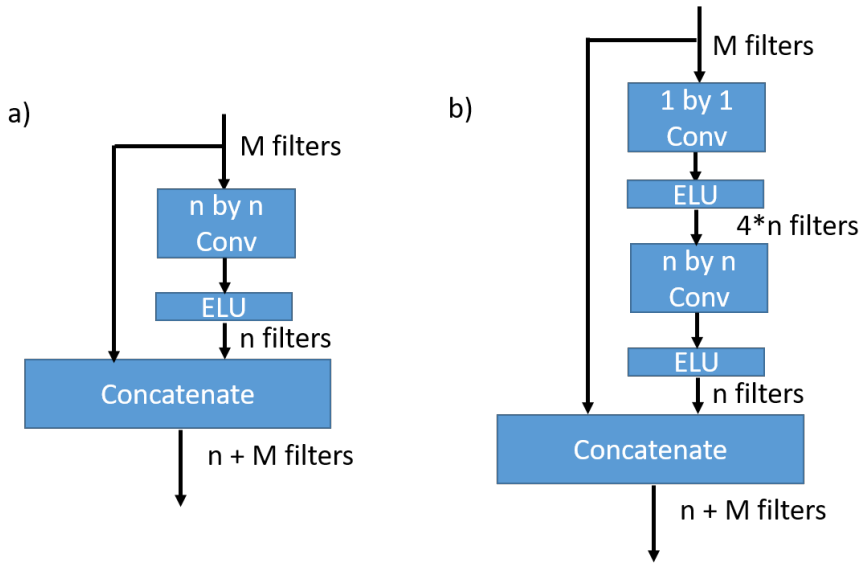


Fig. 1. Dense units, as used in the DeepSCAN architecture a) a dense unit without bottleneck, and b) a dense unit with bottleneck

The layers in our dense blocks have the shape shown in Figure 1. Depending on its position in the network, the convolution might have kernel size 3 by 3 or 5 by 5, and might or might not be dilated. At deeper levels of the network (where the feature depth is rather high) a "bottleneck" is used, meaning that

before the 2D convolution a convolution with 1 by 1 kernels is performed to reduce the number of parameters. As a nonlinearity, we use Exponential Linear Units (ELU) [5] rather than the combination of rectified linear unit and Batch Normalization [6] used in the original Densenet paper. There are two reasons for this: the first is that densenets are very momery intensive: removing batch norm layers reduces the overall memory footprint of network. Secondly eliminating batch normalization makes training less sensitive to high levels of variance in batches.

Dilated convolutions

We use dilated convolutions to aggregate features at multiple scales. Dilated convolutions, sometimes called atrous convolutions, can be best visualised as convolutional layers "with holes": a 3 by 3 convolutional layer with dilation 2 is a 5 by 5 convolution, in which only the centre and corner values of the filter are nonzero, as illustrated in Figure 3. Dilated convolutions are a simple way to increase the receptive field of a classifier without losing spatial information.

Application to Brain Tumor Segmentation

We trained a DeepSCAN classifier on 100 cases (50 LGG, 50 HGG) from the BRATS 2017 training dataset [7]–[10]. The network used is pictured in Figure 2. The network was built using Keras [11] and Tensorflow [12], and trained using stochastic gradient descent with momentum for 100 epochs. Rather than using a softmax layer to classify the three labels (edema, enhancing, other tumor) we employ a multi-task approach to hierarchically segment the tumor into the three overlapping targets: whole tumor, tumor core and enhancing: thus the output of the network is three sigmoid units, one for each target. The network segments the volume slice-by slice: the input data is five consecutive slices from all four modalities, Ground truth for such a set of slices is the lesion mask of the central slice.

Slices from all three directions (sagittal, axial, coronal) were fed to the classifier for training, and in testing the results of these three directions were ensembled by averaging. When applied to the BRATS 2017 validation dataset, the mean Dice scores for Whole Tumor, Tumor core and enhancing tumor were 0.87, 0.68 and 0.71 respectively. After some mild postprocessing (removing small connected components), the Dice coefficient for whole tumor increased to 0.88, and for tumor core to 0.70.

References

[1] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the iEEE conference on computer vision and pattern recognition*, 2017.

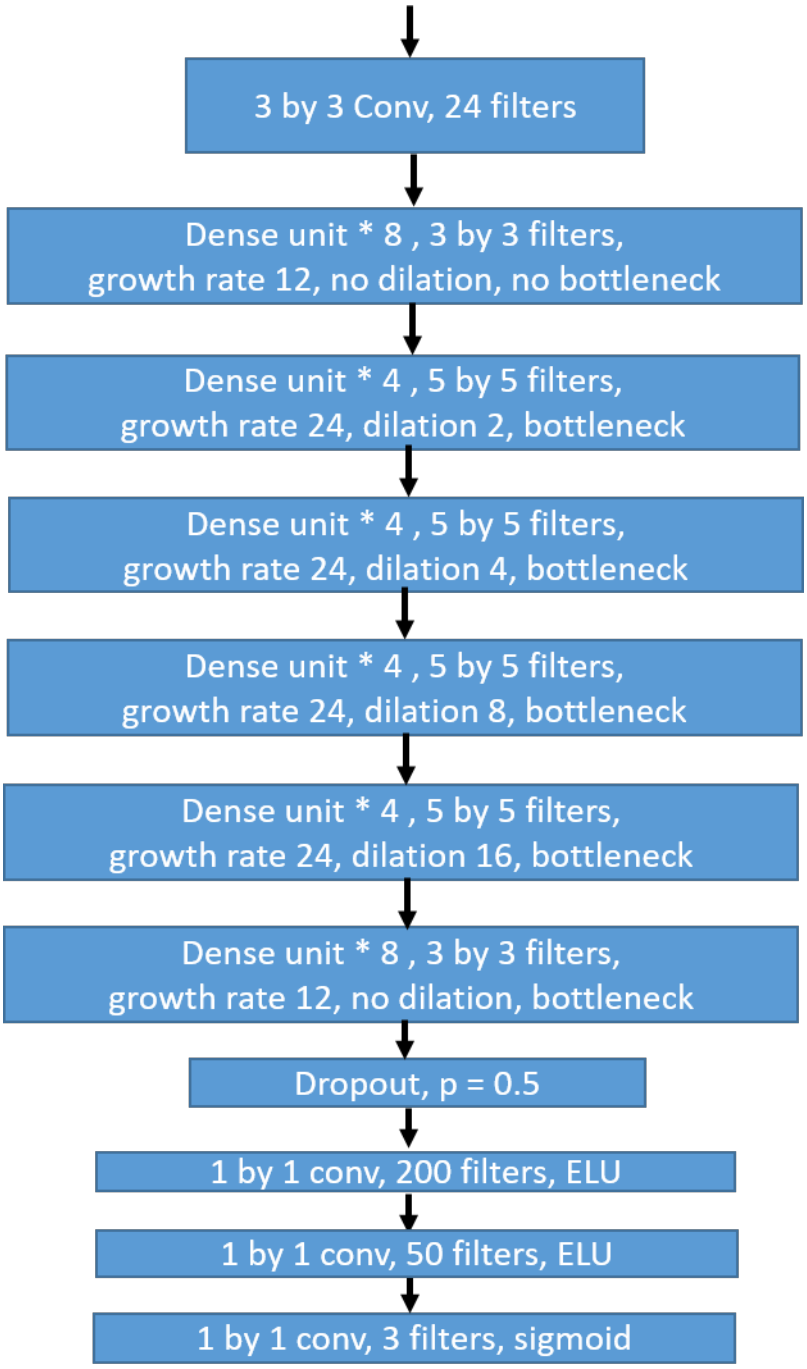


Fig. 2. The DeepSCAN architecture, as applied to brain tumor segmentation

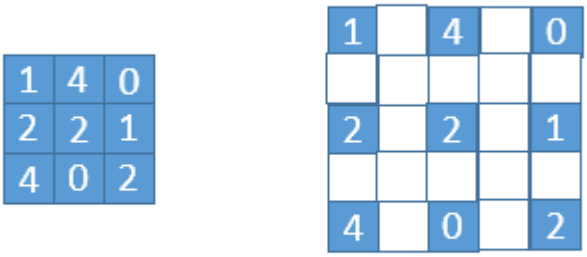


Fig. 3. Left, a 3 by 3 kernel. Right, a 3 by 3 kernel with dilation 2, visualised as a 5 by 5 kernel

[2] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proceedings of international conference on learning representations (iCLR 2017)*, 2017.

[3] S. Jégou, M. Drozdzal, D. Vázquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional denseNets for semantic segmentation," in *CoRR*, 2016, vol. abs/1611.09326.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, 2015.

[5] D. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (eLUs)," *CoRR*, vol. abs/1511.07289, 2015.

[6] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *arXiv:1502.03167*, pp. 1–11, 2015.

[7] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. V. Leemput, "The multimodal brain tumor image

segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct 2015.

[8] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mRI collections with expert segmentation labels and radiomic features," in *In press, nature scientific data*, 2017.

[9] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tCGA-gBM collection," in *The cancer imaging archive*, 2017.

[10] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tCGA-lGG collection," in *The cancer imaging archive*, 2017.

[11] F. Chollet and others, "Keras." https://github.com/fchollet/keras; GitHub, 2015.

[12] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems." 2015.

Fully Automated Glioma Brain Tumors Segmentation and Patient Overall Survival Prediction with SVMs Learning Algorithms: BraTS'2017 Challenge

Alexander F. I. Osman[0000-0002-1286-475X]

American University of Beirut Medical Center, Beirut, 1107 2020 Riad El-Solh, Lebanon
alexanderfadul@yahoo.com

Abstract. The aim of this work was to develop a model for accurate auto-segmentation of the glioma brain tumors in multimodal MRIs and prediction of patient overall survival based on SVMs algorithms. BraTS'2017 datasets were used in this study. We developed a model based on an SVMs algorithms for auto-segmentation of gliomas. Image intensity features were extracted for this purpose as well as pre- and post-processing on the MRIs were employed. The model was trained using the provided training datasets. Then the trained model was used to produce segmentation labels on the validation datasets. Also, the auto-segmented labels were called in combination the patient age parameter for OS classifications prediction. The OS prediction model was trained using the BraTS'17 OS data. A confusion matrix was plotted to evaluate the OS predictor performance in the three classification categories, i.e. long, short, and mid-survivors. The evaluated segmentation results of the edema sub-region tumor on Flair MRIs is not reported. However, for OS classification prediction the algorithm's sensitivity and specificity metrics between the automatically predicted and clinically obtained OS were 100.0% for long-survivors (>15 months); short-survivors (<10 months); and mid-survivors (10 to 15 months), respectively. The OS prediction algorithm's accuracy was 100.0% for classification. The segmentation algorithm took approximately 1.0 to 10 minutes to segment edema in an MR scan of 155 slices and predict the patient OS on Intel (R) Core (TM) 8.00 GB RAM, CPU @ 2.50 GHz processor with 64-bit Operating System. We developed a model for glioma brain tumors segmentation in multimodal MRIs and patient OS predictions based on the SVM learning algorithm. In its present form, the model is fully automated, fairly accurate, easy to implement and efficient. Further improving in the model could provide a robust, and cost-effective supplement to traditional segmentation methods.

Keywords: Glioma Brain Tumors, MRI, Image Segmentation, SVMs Learning Algorithm, GBM Overall Survival.

1 Introduction

Brain tumors, glioma types, are the most common primary brain malignancies. Glioma tumors tend to have different degrees of aggressiveness, variable prognosis and various

heterogeneous histological sub-regions, i.e. peritumoral edema, necrotic core, enhancing and non-enhancing tumor core [1]. Glioma brain tumors are broadly divided clinically into high-grade and low-grade. High-grade tumors are invasive tumors that aggressively grow in a relatively short period of time, leading to the patient's death. Glioblastoma multiform (GBM) is the most common and aggressive high grade (grade IV) glioma tumors with a median survival rate of two years or less and require immediate treatment [2-3]. In contrast, low-grade tumors are slow-growing, with a life expectancy of several years.

Accurate segmentation (or contouring) of the glioma tumors (with its sub-regions) is an essential step in the patient radiation therapy chain for a patient who is a candidate for external beam radiation therapy. Multimodalities of magnetic resonance images (MRIs) are used for glioma tumors segmentation fused on a computed tomography (CT) image acquired during the patient treatment simulation which is standard for patient treatment planning. Moreover, glioma tumors segmentation is crucial during the radiation therapy course for adaptive tumor targeting and after the treatment course for follow up purposes to evaluate the progression of the disease and the success of the treatment strategy.

Mostly, the glioma tumors contouring/segmentation are done manually by the radiation oncologist. This task needs a considerable time and effort as well as sometimes the oncologist may seek for consultation with a neuroradiologist in some complicated cases which adds extra time and effort. In addition, for a given tumor segmentation task there are significant variations among the oncologists themselves in contouring/segmenting the tumors. Therefore, introducing the image processing routines and state-of-the-art machine learning (ML) algorithm techniques that can computationally auto-segment the glioma brain tumors in multimodal MRI scans could have the potential for improved diagnosis, treatment planning, and follow-up of individual patients. Besides, it could serve as a supporting decision tool or replace the existence traditional method. Finally, computational auto-segmentation of gliomas for radiotherapy planning would be useful for standardizing and significantly expediting clinic workflow.

However, developing automated brain tumor segmentation techniques is technically challenging in comparison with healthy/normal tissues/organs. One of the reasons is lesion areas are only defined through intensity changes that are relative to surrounding normal tissue, and even manual segmentations by expert raters show significant variations when intensity gradients between adjacent structures are smooth or obscured artifacts [4]. Also, glioma tumors and its sub-structures vary considerably through patients in its appearance and shape, and localization, size, etc. In this work, we developed a state-of-the-art machine learning model based on support vector machines (SVMs) for; a) automated segmentation of glioma brain tumors in pre-operative multimodal MRI scans and b) prediction of GBM patient overall survival (OS) via integrative analyses of some radiomic features.

2 Materials and Methods

2.1 Dataset

Brain tumors MRI scans datasets and patient overall survival data used in this work were provided by BraTS'17 Challenge [4-7]. The number of patients included was n=210 with HGG/GBM and n=75 with LGG tumors as training datasets, and n=46 patients combing HGG/GBM and LGG as validation datasets. The OS data were included with correspondences to the pseudo-identifiers of the GBM/HGG imaging data with n=163 as training data and n=33 as validation data. The multimodal MRI training datasets of HGG/GBM and LGG were provided with their ground truth segmentation labels of various glioma sub-regions. The manually segmented volume structures were performed by experts following the same annotation protocol, and their annotations were revised and approved by board-certified neuroradiologists. Annotation labels included were the GD-enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor (NCR/NET). The pre-operative multimodal MRI dataset for each patient included four scans; a) T1-weighted (highlights fat locations), post-contrast/gadolinium T1c-weighted (taken after the injection of the contrast agent gadolinium), c) T2-weighted (highlights water locations), and d) T2-Flair (Fluid Attenuated Inversion Recovery, an MR imaging technique that produces images similar to T2-weighted images, but with free water suppressed) volumes. The scans were acquired during the clinical routine with different clinical protocols and various scanners from multi-institutions. The provided multimodal MRI dataset for each patient was co-registered with aligning the volumes of the four MRI modalities to the same anatomical template. Also, the data are interpolated to the same resolution of $1\times1\times1$ mm^3 and skull-stripped to correlate all four images for each patient.

2.2 Machine Learning Model

We used Support Vector Machines (SVMs) to develop a fully auto-segmentation model with no human interaction and OS prediction for patients with GBM. SVMs [8-9] are a set of related supervised machine learning (ML) methods used for classification. The SVM algorithm creates a hyperplane that separates the data into two classes with the maximum-margin. Given training examples labeled (supervised learning) either "1" or "0", a maximum-margin hyperplane is identified which splits the "1" from the "0" training examples, such that the distance between the hyperplane and the closest examples (the margin) is maximized. Hence, this makes SVM algorithm a superior choice for classification tasks. Some binary classifications do not have a simple hyperplane as a useful separating criterion. For those classifications, there is a mathematical approach that retains nearly all the simplicity of an SVM separating hyperplane to create non-linear classifiers by applying kernels [10]. These class of kernel functions includes polynomials, radial basis function (Gaussian), and multilayer perceptron or sigmoid (neural network). The mathematical approach using kernels relies on the computational method of hyperplanes. Therefore, nonlinear kernels can use identical calculations and solution algorithms, and obtain classifiers that are nonlinear. The resulting classifiers are hypersurfaces in some space *S*.

Segmentation Model. A fully automated model based on SVMs was developed to produce segmentation labels of the different glioma tumor sub-regions. For a given patient, the segmentation task was performed with five steps. First, the 3-D MRI data and the ground truth data are uploaded to the model. Second, the uploaded images are pre-processed and image features are extracted. A 3-D median filter is applied to remove the noises in the uploaded volumetric MRI data. Also, the image data are normalized to the global maximum intensity in the image and a thresholding technique (reweighting) with an adequate constant value is used for selecting features in the image. Moreover, the image data and the segmentation labels data are transformed to vectors format. Third, a supervised SVM classifier using the radial basis kernel is used to train the model with the extracted features in combination with its manual segmentation labels. A radial basis function kernel, which have given best result among the kernels, is used for data classification in the model training. After the classification process is completed the model is checked whether it converged or not. Fourth, after the model is successfully converged post-processing is applied to the training SVM segmented data. Holes filling and something filters are employed to fill the cavities in the segmented tumor sub-region and smooth the edges. Also, Gaussian filters are applied to enhance the segmentation and remove the noise in the segmented image. Finally, after the model is being well trained with enough training data, the trained SVM model is used to predict the segmentation labels for the new unseen different glioma sub-regions on the validation data. A prediction function [12] is used to return a vector of predicted class labels for the new MRI data based on the trained SVM classification model. The new MRI image is uploaded, pre-processed and post-processed in a similar way in the previous steps with no ground truth data is used. The whole process is fully automated with no human interaction required.

OS prediction model. An SVM algorithm was also used for this task. The extracted features (i.e. tumor size and location) combined with patient age data were used to develop the prediction model. OS data were clustered into three groups: long-survivors (>15 months), mid-survivors (5 to 15 months), and short-survivors (<5 months). A linear SVM classifier was trained using the training data with ground truth. A fraction of data was used for cross-validating the model.

2.3 Model Evaluation

The predicted segmentation labels with our model on a number of evaluation datasets were generated. The model performance on the evaluation dataset is not reported due to some technical problem on uploading the predicted labels data to the evaluation system. The predictive power of the patient overall survival model for classification was evaluated with plotting a confusion matrix. Performance metrics of model sensitivity, specificity, and accuracy in the three classification categories i.e. long-survivors, short-survivors, and mid-survivors were reported.

3 Results

The predicted glioma brain tumor, edema structure, segmentation results with an SVM learning model on the validation dataset are shown in Fig. 1. The quantitative measures of the model for evaluation dataset segmentation are not reported in the results.

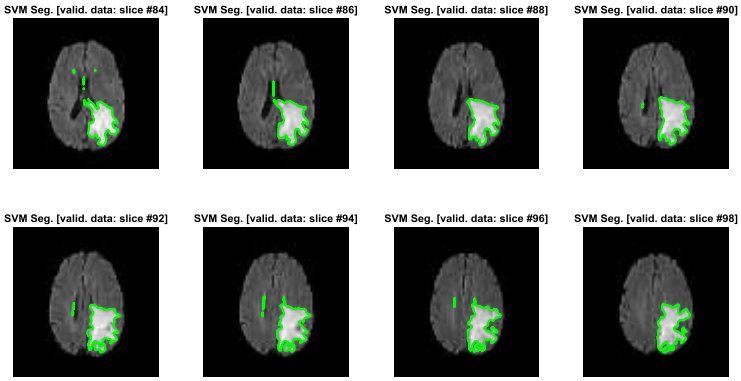


Fig. 1. Predicted segmentation labels (green), edema tumor, on a T2-Flair MRI validation data (slice# 84:2:98) with our automated model.

The segmentation algorithm took approximately 1.0 to 10 minutes to segment a glioma tumor, edema, in an MR scan of 155 slices and predict the patient OS in an Intel(R) Core(TM) CPU @ 2.50 GHz processor with 64-bit Operating System on our local computer infrastructure.

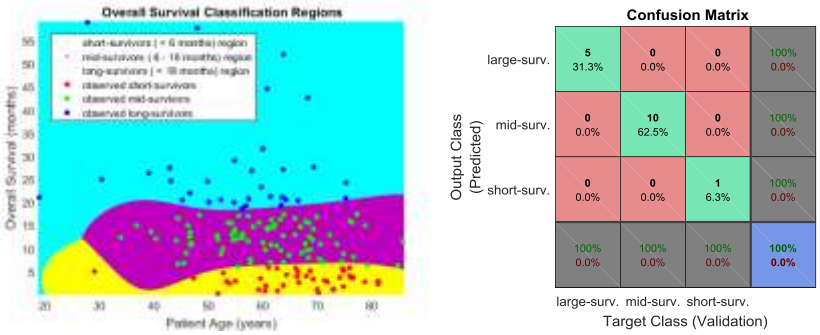


Fig. 2. The plot of the OS classification predicted regions (*Left*) and the confusion matrix (*Right*). On the confusion matrix, the rows correspond to the predicted class (Output Class), and the columns show the true class (Target Class). The diagonal cells show for how many (and what percentage) of the examples the trained algorithm correctly estimates the classes of observations. That is, it shows what percentage of the true and predicted classes match. The off diagonal cells show where the classifier has made mistakes. The column on the far right of the plot shows the accuracy for each predicted class, while the row at the bottom of the plot shows the accuracy for each true class. The cell in the bottom right of the plot shows the overall accuracy.

The patient overall survival predicted results from the segmented data and patient's age parameter were reported in Fig. 2 for GBM.

4 Discussion

Two tasks of glioma brain tumors segmentation and patient overall survival predictions using prior patient data were successfully implemented in this work with machine learning approach.

Brain tumors, gliomas, segmentation was predicted with a developed fully automated model based on SVMs. The reason behind choosing SVM is that it is a superior algorithm for classification tasks. It is based on a maximum-margin distance between the hyperplane and the closest examples (group data). Twelve edema structure segmentation labels were predicted with the model on T2-weighted Flair MR Scans. The quantitative evaluation measures were not reported due to a technical problem in uploading the data to the system. Qualitatively, the segmentation results in Fig.1 are fairly acceptable at this stage. However, the predicted segmentation labels could further be improved. For example, standardizing image intensities of imaging modalities/scanners and patient images with statistical-parametric-mapping could significantly reduce the effect of the patients/modalities variations and improve the predicted segmentations.

Prediction of patient's overall survival (OS) using the segmented tumor in combination with patient age and tumor radiomic features could serve as a survival-predictor and provide an informative clue about the patient treatment outcome. The evaluated results and metric scores were not received yet. The number (and what percentage) of the OS that our algorithm was correctly predicted the classification (Fig. 2) from the validation data was 5 (31.3%) as long-survivors, 10 (62.5%), mid-survivors, and 1 (6.2%) as short-survivors. Thus, the reported sensitivity, defined as correctly classified positive samples divided by true positive samples, for three classification categories was 100.0% where the true and predicted classes are matching. The number of misclassified predictions were 0 (0.0%) for the three classification categories. Hence, the specificity, defined as correctly classified negative samples divided by true negative samples, for three classification categories was 100.0% where our algorithm has made no classification mistakes. The overall accuracy achieved by our algorithm reported for OS classification predictions was 100.0%.

5 Conclusions

We developed a fully automated model for accurate auto-segmentation of the glioma brain tumors in multimodal MRIs and prediction of patient overall survival based on SVMs learning algorithms. The segmentation model was trained the BraTS'2017 training datasets. Then the trained model was used to produce segmentation labels on the validation datasets. In addition, the auto-segmented labels were called in combination the patient age parameter for OS classifications prediction. The OS prediction model was trained using the BraTS'17 OS data. The evaluated segmentation results of the edema sub-region tumor on Flair MRIs is not reported. However, for OS classification prediction the algorithm's sensitivity and specificity metrics between the automatically predicted and clinically obtained OS were 100.0% for long-survivors (>15 months); short-survivors (<10 months); and mid-survivors (10 to 15 months), respectively. The OS prediction algorithm's accuracy was 100.0% for classification. The segmentation algorithm took approximately 1.0 to 10 minutes to segment edema in an MR scan of 155 slices and predict the patient OS on Intel (R) Core (TM) 8.00 GB RAM, CPU @ 2.50 GHz processor with 64-bit Operating System. In its present form, the model is fully automated, fairly accurate, easy to implement and efficient. Further improving in the model could provide a robust, and cost-effective supplement to traditional segmentation methods

References

1. Holland, E.: Progenitor cells and glioma formation. Current Opinion in Neurology 14(6), 683–688 (2001).
2. Ohgaki, H., and Kleihues, P.: Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. Journal of Neuropathology & Experimental Neurology 64(6), 479–489 (2005).
3. Louis, D., Ohgaki, H., Wiestler, O., and Cavanee, W.: WHO classification of tumours of the central nervous system. 4th edn. WHO/IARC, Lyon, France (2007).
4. Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, D., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S., Raviv, T., Reza, S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).
5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, (2017). [In Press]
6. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative

Scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
7. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
8. Christianini, N., and Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, UK, (2000).
9. Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, second edition. Springer, New York, USA, (2008).
10. Fan, R-E., Chen, P-H., and Lin, C-J.: Working set selection using second order information for training support vector machines. Journal of Machine Learning Research 6, 1889–1918 (2005).
11. Platt, J.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in Large Margin Classifiers. MIT Press, 61–74 (1999).

Residual Encoder and Convolutional Decoder Neural Network for Glioma Segmentation

Kamlesh Pawar*[1,2], Zhaolin Chen[3,4], N. Jon Shah[1,3,4], and Gary Egan[1,2]

[1] Monash Biomedical Imaging, Monash University, Australia
[2] School of Psychological Sciences, Monash University, Australia
[3] Electrical and Computer System Engineering, Monash University, Australia
[4] Institute of Medicine, Research Centre Juelich, Germany

Abstract. A deep learning approach to glioma segmentation is presented. An encoder and decoder pair deep learning network is designed which takes T1, T2, T1-CE (contrast enhanced) and T2-Flair (fluid attenuation inversion recovery) images as input and outputs the segmented labels. The encoder is a 49 layer deep residual learning architecture that encodes the $240 \times 240 \times 4$ input images into $8 \times 8 \times 2048$ feature maps. The decoder network takes these feature maps and extract the segmented labels. The decoder network is fully convolutional network consisting of convolutional and upsampling layers. Additionally, the input images are downsampled using bilinear interpolation and are inserted into the decoder network through concatenation. This concatenation step provides spatial information of the tumor to the decoder, which was lost due to pooling/downlsampling during encoding. The network is trained on the BRATS-17 training dataset and validated on the validation dataset. The dice score, sensitivity and specificity of the segmented whole tumor, core tumor and enhancing tumor is computed on validation dataset.

Keywords: deep learning, image segmentation, computer vision, CNN

Introduction

Gliomas are the tumors of the central nervous system which arises from glial cells. The gliomas are classified into two types depending on the aggressiveness of the tumor. The more aggressive are called high grade gliomas (HGG) and less aggressive are called low grade (LGG), both types of tumors are malignant and need treatment. The accurate segmentation of gliomas is important in grading, treatment and monitoring of the tumor progression. Mutiple MR image contrasts are used to evaluate the type and extent of tumor. The different contrasts T1, T2, T1-CE and T2-Flair are analysed by a radiologists and tumor regions are manually segmented. Segmenting brain tumor is a comprehensive task, and large intra-rater variability is often reported, e.g. 20% [1]. Thus it is imperative to

* correspondence to: kamlesh.pawar@monash.edu

have a reliable automatic segmentation algorithm that standardizes the process of segmentation, resulting in more precise planning, treatment and monitoring.

Deep learning methods based on convolutional neural networks (CNN) [2, 3] have demonstrated highly accurate results in image classification and segmentation [4]. However, selection of the number of CNN layers is a complex task. On one hand, increasing the number of layers improves complexity of the network and leads to more accurate results. On the other hand, designing more deep CNN may result in degradation of performance due to exploding/vanishing gradients. This problem is partially solved by the batch normalization layers [5] that minimizes the chances of exploding/vanishing gradients. Another limitation of designing deep neural network is that training them becomes difficult after a certain depth and the network ceases to converge. However recently introduced residual networks [6], consisting of short cut connections can be trained to the larger depths. In this paper we present an encoder-decoder based CNN architecture to solve the tumor segmentation problem. The network consist of two parts, first is a 49 layer deep residual encoder network followed by a 10 layers of fully convolutional decoder network. The performance for the whole tumor, core tumor and enhancing tumor were 0.824, 0.627 and 0.575 respectively.

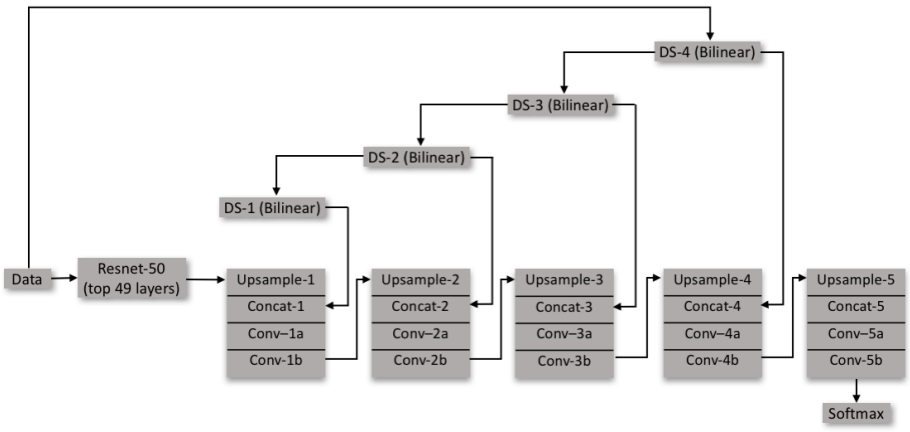


Fig. 1. Residual encoder and convolutional decoder network; the encoder is a 49 layer deep residual network and the decoder is a 10 layer deep fully convolutional network with bilinear upsampling layers. The input data is also downsampled using bilinear interpolation and is inserted back into the decoder through concatenation.

Methods

The method presented here is based on residual learning convolutional neural network [6]. The network is a 2D CNN, which performs segmentation on individual slices. The network is designed as an encoder-decoder pair, the four input

Layer	Input dimension	Output dimension
Upsample-1	$8 \times 8 \times 2048$	$15 \times 15 \times 2048$
Concat-1	$15 \times 15 \times 2048$	$15 \times 15 \times 2052$
Conv-1a	$15 \times 15 \times 2052$	$15 \times 15 \times 1024$
Conv-1b	$15 \times 15 \times 1024$	$15 \times 15 \times 1024$
Upsample-2	$15 \times 15 \times 1024$	$30 \times 30 \times 1024$
Concat-2	$30 \times 30 \times 1024$	$30 \times 30 \times 1028$
Conv-2a	$30 \times 30 \times 1028$	$30 \times 30 \times 512$
Conv-2b	$30 \times 30 \times 512$	$30 \times 30 \times 512$
Upsample-3	$30 \times 30 \times 512$	$60 \times 60 \times 512$
Concat-3	$60 \times 60 \times 512$	$60 \times 60 \times 516$
Conv-3a	$60 \times 60 \times 516$	$60 \times 60 \times 256$
Conv-3b	$60 \times 60 \times 256$	$60 \times 60 \times 256$
Upsample-4	$60 \times 60 \times 256$	$120 \times 120 \times 256$
Concat-4	$120 \times 120 \times 256$	$120 \times 120 \times 260$
Conv-4a	$120 \times 120 \times 260$	$120 \times 120 \times 128$
Conv-4b	$120 \times 120 \times 128$	$120 \times 120 \times 128$
Upsample-5	$120 \times 120 \times 128$	$240 \times 240 \times 128$
Concat-5	$240 \times 240 \times 128$	$240 \times 240 \times 132$
Conv-5a	$240 \times 240 \times 132$	$240 \times 240 \times 64$
Conv-5b	$240 \times 240 \times 64$	$240 \times 240 \times 64$
SotmaxwithLoss	$240 \times 240 \times 64$	$240 \times 240 \times 4$

Table 1. Decoder network input and output dimensions for each layer; all the kernal sizes in the decoder were 3×3

images of size 240×240 are given as input to the encoder that encoded them into $8 \times 8 \times 2048$ data. This encoded data are provided to a fully convolutional decoder network that predict the labels for the glioma segmentation. The Resent-50 [6] which was the winner of ILSRVC 2015 image classification challenge was used as an encoder network followed by fully convolutional layers of decoder network.

Fig.1, shows the network architecture; it consist of the first 49 layers of Resnet-50 as the encoder. The decoder network consist of upsampling layers that enlarges the dimension of images by a factor of 2. The weights of upsampling layer were fixed to bilinear upsampling and were not learned. During the encoding process the spatial information is lost due to pooling/downsampling, therefore the spatial information is reintroduced into the decoding network by concating the original images scaled by bilinear interpolation after each upsampling layer. After each convolutional layer the batch normalization and scaling was performed followed by an ReLU non linear activation function. The layer parameters for encoder network, which is derived from first 49 layers of Resnet-50 are presented in [6]. The decoder networks' input and output dimensions are shown in the Table.1, where each layer consist of convolutional kernal of size 3×3. The output of the decoder network layer is class probabilities for labels.

The training dataset [7–10] used to train the network was provided by organiser which consisted of 3D brain images of 284 patient with four different contrast: T1, T1-CE, T2 and T2-Flair. The images were normalized using histogram matching, the reference histogram used was obtained by averaging histograms of all the training dataset. The data is divided into 254 training and 30 test dataset. The caffe [11] framework was used to train the network using stochastic gradient descent.

Results

The trained network was tested on the validation data and the results of segmentation were uploaded on the computing portal provided by the organisers. The dice score, sensitivity, specificity and Hausdorff distance were computed on the segmented labels. The Table.2, shows the result of segmentation on 46 different validation dataset. The segmentation network was also tested on test dataset derived from the training dataset, this was the dataset on which the network was not trained. The results of the segmentation on one slice from 3 different patient are shown in Fig.2

Metric	Whole Tumor	Core Tumor	Enhancing Tumor
Dice Score (mean)	0.824	0.627	0.575
Dice Score (median)	0.865	0.728	0.724
Sensitivity (mean)	0.831	0.669	0.595
Sensitivity (median)	0.885	0.746	0.690
Specificity (mean)	0.993	0.994	0.999
Specificity (median)	0.994	0.997	0.999

Table 2. Results for the segmentation on validation dataset of 46 patients

Discussion

The median of dice score, sensitivity and specificity are all greater than the mean, which indicates that the proposed method performed well for most of the dataset but did not performed well for a few, that took the mean to the lower side. The boundaries of the labels segmented by the proposed algorithm are smooth compared to the ground truth (Fig.2(g)), this may be due to the fact that some of the ground truth were created using automated algorithms rather than human rater.

Conclusion

In this paper, we developed a 59-layer deep encoder-decoder CNN network that takes a 2D slices of 3D MRI mutimodal images as input and outputs the seg-

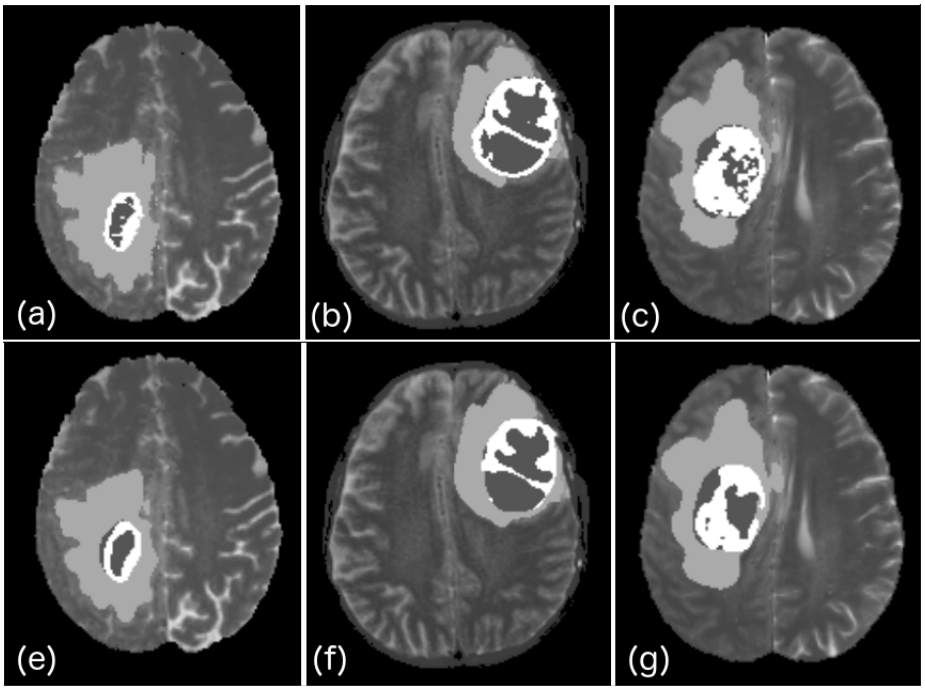


Fig. 2. Segmentation results on test dataset for 3 different patients; **(a, b, c)** are the ground truth; **(d, e, f)** are segmentation results from the proposed residual encoder - convolutional decoder network

mented labels. The average and median dice score for whole tumor were 0.824 and 0.865 respectively.

References

1. G. P. Mazzara, R. P. Velthuizen, J. L. Pearlman, H. M. Greenberg, and H. Wagner, "Brain tumor target volume determination for radiation treatment planning through automated mri segmentation," *International Journal of Radiation Oncology* Biology* Physics*, vol. 59, no. 1, pp. 300–312, 2004.
2. Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396–404.
3. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
4. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

5. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
7. B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.
8. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Nature Scientific Data*, 2017.
9. S. Bakas, H. Sotiras, M. Bilello, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection," *The Cancer Imaging Archive*, 2017, DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
10. S. Bakas, H. Sotiras, M. Bilello, J. Kirby, J. Freymann., K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection," *The Cancer Imaging Archive*, 2017, DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
11. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," *arXiv preprint arXiv:1408.5093*, 2014.

Multimodal Brain Tumor Segmentation using Ensemble of Forest Method

Ashish Phophalia[1] and Pradipta Maji[2]

ashish_p@iiitvadodara.ac.in, pmaji@isical.ac.in
[1]Indian Institute of Information Technology, Vadodara, Gandhinagar Campus, Gujarat, India-382028
[2]Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India - 700108

Abstract. In this paper, we propose a cascaded ensemble method based on Random Forest, named as **Ensemble-of-Forest, (EoF)**. Instead of classifying huge amount of data with a single forest, we proposed two stage ensemble method for Multimodal Brain Tumor Segmentation problem. Identification of Tumor region and its sub-regions poses challenge in terms of variations in intensity, location etc. from patient to patient. We identify the initial region of interest (ROI) by linear combination of FLAIR and T2 modality. For each training scan/ROI, we define a Random Forest as first stage of ensemble method. For a test ROI, collect a set of similarly seen ROI and hence forest based on mutual information criteria and collect majority voting to classify voxels in it. We have reported results on BRATS 2017 dataset in this paper.

Keywords: Brain Tumor Segmentation, Ensemble method, Random Forest

1 Introduction

Gliomas is one of the well known form of brain tumor from glial cells. It is categorized as either Low Grade Gliomas (LGG) or High Grade Gliomas (HGG). HGG is found to be more aggressive in growth leading to death. The tumor region is also divided into sub-regions inside named as necrotic, enhancing and non-enhancing part. The automated methods must delineate tumor region and segment them from healthy brain. However, presence of tumor is highly varies in terms of intensity, texture, appearance etc. from patient to patient. Hence, one needs to take many parameters in to account for classification and segmentation task [1], [2].

The Multimodal **Bra**in **T**umor Image **S**egmentation (BraTS) challenge have seen various methods in past few years [3]. The Deep Learning/CNN, SVM, Random Forest and CRF based methods are most commonly used approaches [4], [5], [6], [7], [8], [9].

Random Forest (RF) [10], being a supervised learning method has been explored in the brain tumor segmentation context led by work of Zikic et. al. [11]. The authors have proposed Gaussian Mixture model based prior probability estimation for various tissues which are used to train Decision Forest in the later

stage. Goetz et. al. [12] have proposed their method based on Extremely Randomized Tree method [13] to leverage the selection of features in splitting in trees as in classical random forest. Meier et. al. [14] proposed Random forest method followed by Conditional Random Field (CRF) as spatial regularization. Malmi et. al. [15] proposed two stage RF architecture where first stage classifies Tumor and Non-Tumor tissues. Second stage involve classification of tumorous tissues to various sub-parts. The method follows multiple postprocessing steps using morphological operations and Markov Random Field (MRF) as spatial regularization. Folgoc et. al. [16] proposed multilayer RF architecture where each layer sequentially refines predefined ROI in to sub parts of tumor region/classes. Ellwaa et. al. [17] have proposed a new way to select sample from few patients only instead of selecting samples from whole database.

Inspired by previous methods with RF, we propose a novel method based on two stage architecture. The crux of method is to find the relevance of the tumor/patient seen in the past. In essence, we tries to capture the resemblance of the input patient/tumor from given database as a past examples. The proposed architecture is shown in Figure 1. We model the characteristics each example via RF. Given an input test pattern, selected similar RF cast their votes and again ensemble in majority fashion as second stage predication. Since, brain tumor segmentation is a imbalanced class problem, the method give much scope to select samples from small classes. Also, the method can be extended easily with more data with respect to time for better predication. The section below describe all the step in detail followed by result and discussion section.

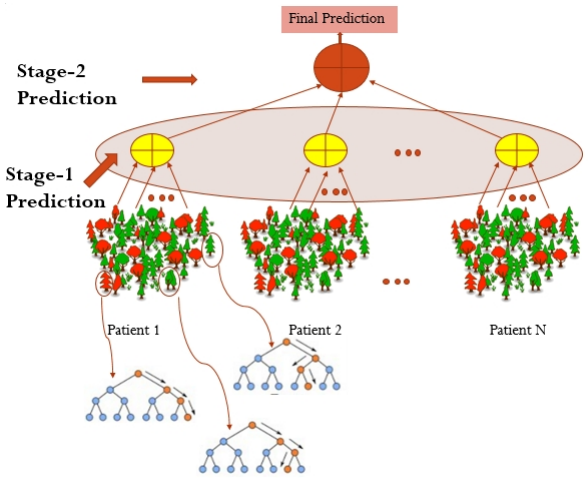


Fig. 1. The proposed architecture as Ensemble of Forest (EoF) method

2 Method

2.1 Dataset

We have used MICCAI BRATS 2017 dataset in this work [3], [18], [19], [20]. The dataset contains total 285 cases, 210 HGG and 75 LGG patients. Each patient has 4 modalities, namely, FLAIR, T1, T1-Contrast enhanced, T2 and ground truth label volume of size $240 \times 240 \times 155$. The database comprises of contribution from various organizations, this make challenging to cop-up with variety in data. The sub-regions considered for evaluation are: 1) the "enhancing tumor" (ET), 2) the "tumor core" (TC), and 3) the "whole tumor" (WT). The labels in the provided data are: 1 for necrotic (NCR) and the non-enhancing (NET), 2 for edema (ED), 4 for ET, and 0 for everything else. The TC entails the ET, as well as the necrotic (fluid-filled) and the non-enhancing (solid) parts of the tumor. The WT describes the complete extent of the disease, as it entails the TC and the peritumoral edema (ED). All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol. The provided data are distributed after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution ($1mm^3$) and skull-stripped.

2.2 Pre-Processing

The N4ITK bias correction method [21] was applied to all the images in dataset using 3DSlicer software. Afterward, all the images were rescaled in the range $[0 - 255]$.

2.3 Initial Segmentation

We have considered linear combination of FLAIR and T2 imaging modalities as $I_{fused} = \alpha * I_{FLAIR} + (1 - \alpha) * I_{T2}$ (where α is adjusted experimentally) [9]. This helps to maintain the hyper-intensity of the tumor region and, at the same time, to suppress the intensity of other irrelevant areas [22]. An intensity at T^{th} percentile is defined as threshold in I_{fused}, τ, and volume intensity above τ is converted to 1, otherwise 0, thus converted to binary mask. Afterward, M largest connected components in 3D is extracted from the binary mask as initial region of interest (ROI). The selection of multiple components is done in order to capture presence of abmornal tissues at different locations. Here, we have considered three largest connected components. The image I_{fused} is also considered for feature generation along with those four modalities in the next step.

2.4 Feature Extraction

After extracting initial ROI for each patient, we have derived 280 features at each voxel locations. The features used are as follows:

1. Appearance - Intensity in each modality ($1 \times 5 = 5$ features)
2. Texture - a) Image Gradient in three directions in each modality, b) Min, Max, Mean, Median, Mode intensities after Gaussian smoothing at scale, $\in 0.5, 1.0, 1.5$, and window siez $3 \times 3 \times 3$ and $5 \times 5 \times 5$ ($3 \times 5 = 15 + 5 \times 3 \times 2 \times 5 = 150$ features)
3. Statistical - Max, Min, Mean, Median, Mode, Standard deviation, kurtosis, skewness and three central order moments in each modality with window size $3 \times 3 \times 3$ and $5 \times 5 \times 5$ ($11 \times 5 \times 2 = 110$ features)

2.5 Forest Generation

Random Forest is ensemble method utilizing numerous of trained decision trees for decision making [10]. It pass test input to each tree and collect class information and aggregate them to make final decision. The most commonly used aggregation method is Majority Voting. The construction of RF takes limited subset of features to make an split. In this work, we propose a Random Forest to each patient data. Since data is quite large and requires more trees in the training. Hence, we proposed two stage hierarchical ensemble method. For each training scan, a Random Forest is constructed considering features at only initial ROI locations.

2.6 Ensemble of Forests

The idea is to find a resemblance with the patients in the database. Here, we tries to model given test example, with already seen past examples. Hence, a mutual information based criteria is used to find out M similar cases from the database. So, for each test case, class information is collected from each those *similar* forests and aggregated again in majority voting fashion as final class label with equal weights. Note that, mutual information with itself is considered to be zero and hence it is not considered in similar M cases for training database.

3 Results

We have used only BRATS 2017 training dataset released by Challenge Organizers. The dataset contains 285 patient volumes where 210 are HGG cases and rest are LGG cases. The parameter values estimated experimentally as: $M = 11$, $\alpha = 0.7$ and $T = 98$ percentile. The Figure 3 shows the boxplot for the training database and mean of whole tumor, tumor core and enhance part are 0.64, 0.49 and 0.47 respectively.

4 Discussion

In this paper, a novel two stage based Random forest is proposed. A forest of small trained forests is build to classify test pattern aggregation voting from all

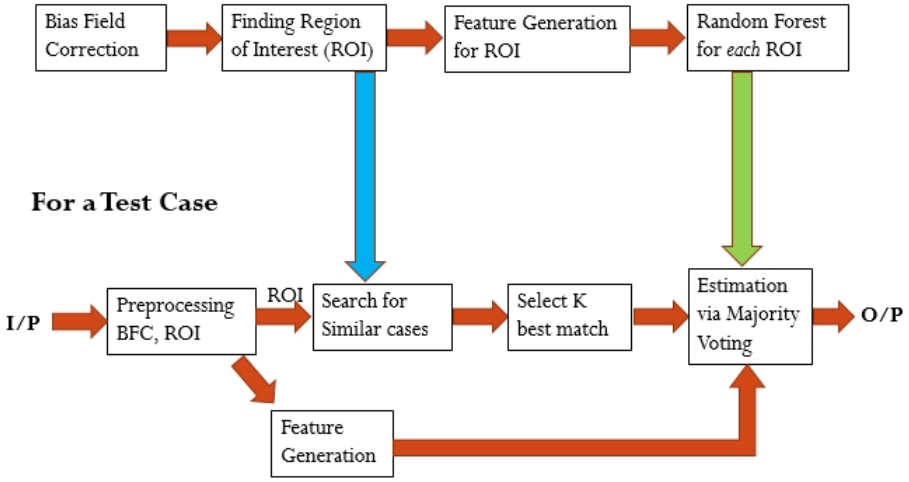


Fig. 2. The flowchart for proposed method

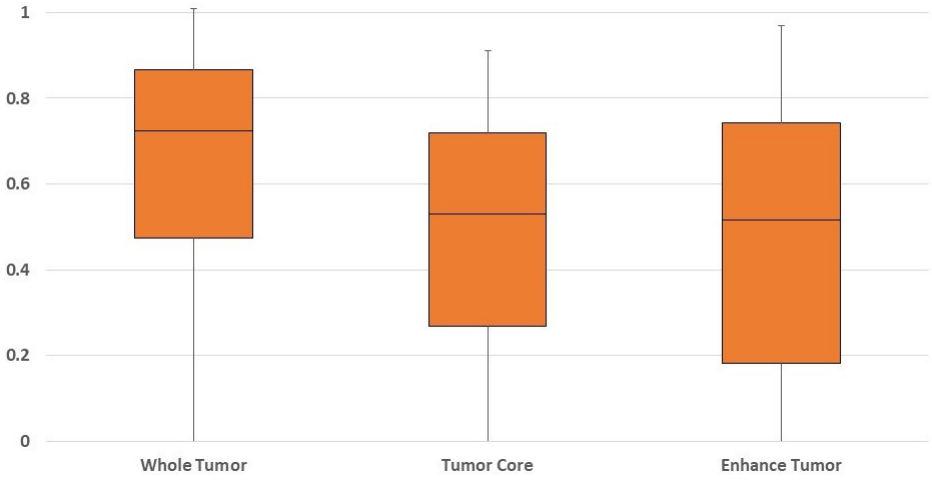
such small forest. For each patient, a small forest is trained based on the features extracted from initial ROI. The results presented here are satisfactory. However, a details analysis is sought for construction of those small forests and trees inside in them.

Acknowledgement

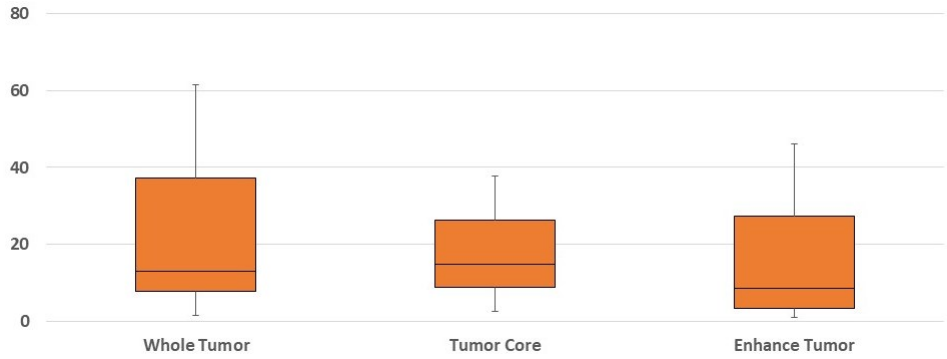
This work was supported in part by the Dept. of Electronics and Information Technology, Govt. of India (PhD-MLA/4(90)/2015-16).

References

1. Ivana Despotović, Bart Goossens, and Wilfried Philips. Mri segmentation of the human brain: challenges, methods, and applications. *Computational and mathematical methods in medicine*, 2015, 2015.
2. Nelly Gordillo, Eduard Montseny, and Pilar Sobrevilla. State of the art survey on mri brain tumor segmentation. *Magnetic resonance imaging*, 31(8):1426–1438, 2013.
3. Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2015.
4. Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: State of the art and future directions. *Journal of Digital Imaging*, pages 1–11, 2017.

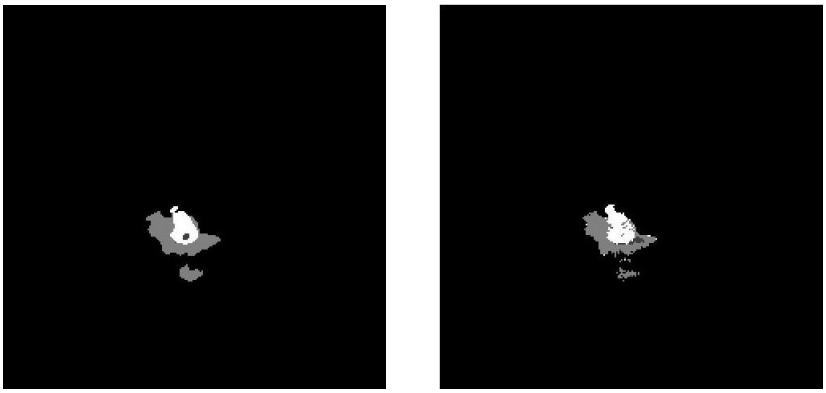


(a) Dice Coefficient



(b) Hausdorff Distance

Fig. 3. The performance of proposed method on BRATS 2017 training database. Parameters used are mentioned in the text.

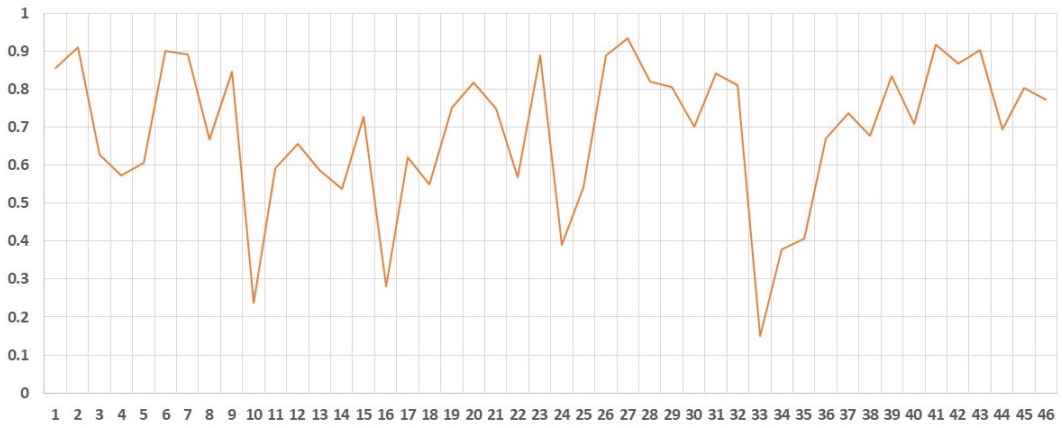


(a) Ground Truth

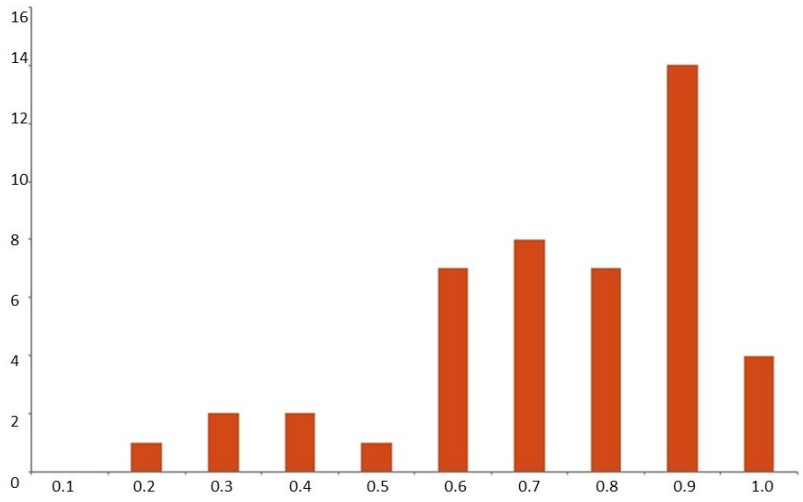
(b) Segmentation from proposed method

Fig. 4. The ground truth image (on left) and result of proposed method (on right) showing for case $TCIA-231$ slice 75. The dice score for whole tumor is 0.9239.

5. Ali Işın, Cem Direkoğlu, and Melike Şah. Review of mri-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*, 102:317–324, 2016.
6. Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *arXiv:1505.03540v3 [cs.CV]*, 2016.
7. Adriano Pinto Victor Alves Srgio Pereira. Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Transactions on Medical Imaging*, 2015.
8. László Lefkovits, Szidónia Lefkovits, and László Szilágyi. Brain tumor segmentation with optimized random forest. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 88–99. Springer, 2016.
9. Bi Song, Chen-Rui Chou, Xiaojing Chen, Albert Huang, and Ming-Chang Liu. Anatomy-guided brain tumor segmentation and classification. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 162–170. Springer, 2016.
10. Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
11. Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, C Demiralp, Jamie Shotton, Owen M Thomas, Tilak Das, Raj Jena, and Stephen J Price. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 369–376. Springer, 2012.
12. Michael Goetz, Christian Weber, Josiah Bloecher, Bram Stieltjes, Hans-Peter Meinzer, and Klaus Maier-Hein. Extremely randomized trees based brain tumor segmentation. *Proceeding of BRATS challenge-MICCAI*, pages 006–011, 2014.



(a)



(b)

Fig. 5. The performance graph of Dice coefficient for whole tumor on BRATS 2017 validation dataset

13. Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.
14. Raphael Meier, Stefan Bauer, Johannes Slotboom, Roland Wiest, and Mauricio Reyes. Appearance-and context-sensitive features for brain tumor segmentation. *Proceedings of MICCAI BRATS Challenge*, pages 020–026, 2014.
15. Eric Malmi, Shameem Parambath, Jean-Marc Peyrat, Julien Abinahed, and Sanjay Chawla. Cabs: A cascaded brain tumor segmentation approach. *Proceedings of MICCAI BRATS Challenge*, pages 042–047, 2015.
16. Loic Le Folgoc, Aditya V Nori, Siddharth Ancha, and Antonio Criminisi. Lifted auto-context forests for brain tumour segmentation. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 171–183. Springer, 2016.
17. Abdelrahman Ellwaa, Ahmed Hussein, Essam AlNaggar, Mahmoud Zidan, Michael Zaki, Mohamed A Ismail, and Nagia M Ghanem. Brain tumor segmantation using random forest trained on iteratively selected patients. In *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 129–137. Springer, 2016.
18. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, and Davatzikos C. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, page In Press, 2017.
19. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive*, 2017.
20. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive*, 2017.
21. Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320, 2010.
22. Rupsa Saha, Ashish Phophalia, and Suman K. Mitra. Brain tumor segmentation from multimodal mr images using rough sets. In *Workshop on Computer Vision Application, ICVGIP*, page In Press. Springer, 2016.

Brain Tumor Segmentation in MRI Scans using Deeply-Supervised Neural Networks

Reza Pourreza, Ying Zhuge, Holly Ning, Robert Miller

Radiation Oncology Branch, National Cancer Institute, National Institute of Health, Bethesda, MD 20814, USA
reza.pourrezashahri@nih.gov

Abstract. This paper presents a deeply-supervised neural network method based on Holistically-Nested Edge Detection (HED) for automatic segmentation of brain tumor from multi-modal magnetic resonance images (MRI). The HED method, which is originally developed for image edge detection using deep convolutional neural networks, is extended for multiple-object segmentation. The performance of the proposed method is measured in terms of Dice score. Experiments on 2017 Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) challenge datasets demonstrate that the method performs well. The assessments revealed the Dice scores of 0.86, 0.60, and 0.69 for whole tumor (WT), tumor core (TC), and enhancing tumor (ET) classes, respectively.

Keywords: HED, Brain Tumor, MRI.

1 Introduction

Gliomas, a type of brain tumor originating from glial cells, are the most frequent primary brain tumors in adults [1]. Improved quantification of the various aspects of a glioma requires accurate segmentation of the tumor in MRI scans. Since the manual segmentation is time-consuming and subject to human error and irreproducibility, automatic segmentation has received a lot of attention recently. However, automated segmentation of brain tumor is difficult because tumors appear in different shapes, sizes, locations, and intensities. The presented methods in previous rounds of BRATS challenge [2-5] mostly fall under two categories. The first category includes patch-wise methods where the input images are divided into many patches. The patches are then fed to the classifiers individually and the classifier determines the class of the center voxel of the patch. Examples include Support Vector Machine (SVM), Random Forest, and Convolutional Neural Network (CNN) based methods [6-8]. Unlike the patch-wise methods, the second category of segmentation methods feed the whole slice of the MRI to the classifier as once. Almost all the methods in the second category are based on Fully Convolutional Networks (FCN) [9]. While the methods in the first category are very time-consuming, the ones in the second category are very efficient.

An automatic brain tumor segmentation method is presented in this paper which falls under the second category. The presented method is a deeply supervised neural network [10] based on a modified HED [11]. The rest of the paper is organized as follows: the used dataset is presented in section 2, the proposed method is descried in section 3, the achieved results are provided in section 4, and the paper is concluded in section 5.

2 Material

The MRI scans used in this study include data from the BRATS 2017 challenge that include 210 high grade glioma (HGG) and 75 lower grade glioma (LGG) cases. The image datasets share the following four MRI contrasts: T1, post-contrast T1-weighted (T1c), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR). All the images are co-registered and resampled to the same voxel resolution: 155 slices of 240×240 pixels. The scans are annotated by domain experts and four labels are used: edema, non-enhancing (solid) core, necrotic (or fluid-filled) core, and enhancing core. In this study, T1 modality is disregarded and the remaining three modalities are used. It is also worth to mention that the WT class includes all the annotated labels, the TC class includes all the labels but edema, and the ET class includes only enhancing core label.

3 Proposed Method

The proposed method consists of three steps: preprocessing, classification, and post-processing. These steps are explained in the next subsections.

3.1 Preprocessing

Unlike computed tomography (CT) where the measurements are done in absolute units, MRI scans are expressed in arbitrary units that differ between study visits and subjects. In this work, N4BiasCorrect filter followed by a simple piece-wise linear histogram matching algorithm is used to normalize the intensity of the MRI scans. The same procedure is done for all three used modalities. First, a histogram of the scan is calculated. The peak of the histogram corresponds to the white matter of the brain. Also, the 1 and 99 percentiles of the cumulative histogram are calculated. Then, the values between the 1 percentile and the peak index are mapped to n_1 and n_2 interval, and the values between the peak index and 99 percentiles are mapped to n_2 to n_3 interval. In this study, n_1, n_2, and n_3 are chosen as 0, 127, and 255, respectively. It is worth to mention that the histogram is calculated using signal-only voxels and the blank voxels are excluded. By using this simple preprocessing algorithm, all modalities of all MRI scans are normalized to n_1 to n_3 interval.

3.2 Classification

The classification is done via a HED-based neural network. HED was originally proposed for edge detection but it is extensible to other applications such as segmentation. The core of a HED is a FCN. However, HED has multiple side outputs located at each convolutional layer together with a fusion of all the side outputs. All the outputs are compared to the ground truth during the training and this deep supervision makes the network to be trained more effectively compared to FCNs. More details about HED is available in [11].

Although HED is introduced for binary classification (edge/non-edge), the architecture is changed in this work for multiple-class classification tasks. The MRI scans are fed to the network slice-by-slice. In other words, the three modalities of a single preprocessed slice are fed to the network and the network generates three binary classification maps for WT, TC, and ET classes where each pixel in the binary maps corresponds to a voxel in the input images.

The number of background voxels is much higher than the number of tumor voxels in the dataset. Two simple solutions are sought in this work to reduce the class skew. First, those slices with no signals are excluded from training. Second, a window of 200×200 pixels is cropped from the center of the slice where the original slice size is 240×240. The removed pixels correspond to the blank area around the skull. This simple solution reduces the number of background pixels at least by 30 percent and consequently lowers the skew.

3.3 Postprocessing

Once the whole scan is passed through the neural network slice-by-slice, the generated binary maps for each class are put together to generate the 3d tumor maps for the entire scan. The 3d binary maps include false alarms. Postprocessings for the three classes are conducted individually to reduce the false alarms and improve the accuracy.

For the WT class, postprocessing includes maintaining the largest connected object and getting rid of the other objects. But before applying this, a morphological binary opening operator is applied to the 3d binary map to separate those false alarms that are connected using weak connections and subsequently reduced their sizes.

Postprocessing for TC class includes multiplying its 3d binary map by the postprocessed WT 3d binary map. By doing this, all the false alarms outside the whole tumor segment are removed. Like the CT class, postprocessing for ET class includes multiplying its 3d binary map by both the postprocessed WT and CT 3d binary maps.

4 Results and Discussions

The neural network was implemented in Caffe and the rest of the pipeline was implemented in Python. All the available training data from BRATS 2017 were used in training the neural network. The training was carried out via a Stochastic Gradient Descent strategy with a batch size of 30 on a NVIDIA Titan Xp graphic card.

The performance was measured using the BRATS online evaluation system for the validation dataset. The achieved Dice scores are reported in Table 1.

Table 1. Dice scores for the validation dataset

Dataset	WT	TC	ET
Validation	0.86	0.60	0.69

A sample segmentation outcome together with the ground truth is shown in Fig. 1.

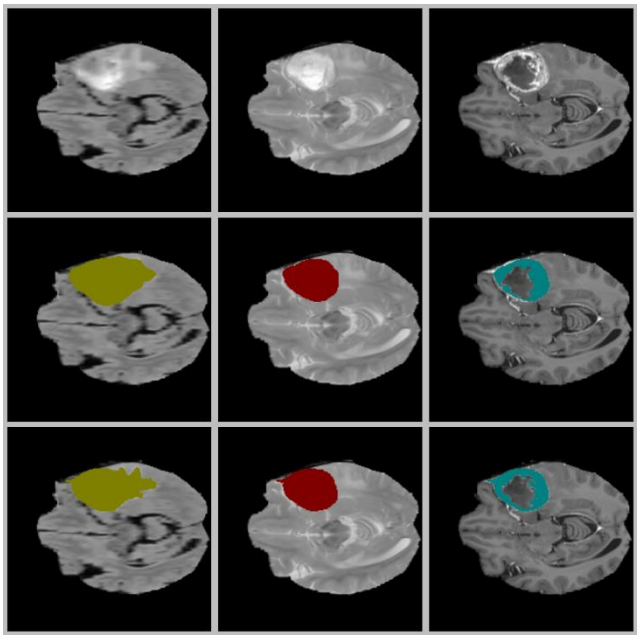


Fig. 1. Sample segmentation results. Top row from left to right: FLAIR, T2, and T1c. Middle row: segmented WT, TC, and ET. Bottom row: ground truth WT, TC, and ET.

5 Conclusions

An automatic method for brain tumor segmentation method in MRI scans was presented in this paper. The classification core of the presented method is a deeply supervised neural network based on HED. Low computational complexity, yet very effective preprocessing and postprocessing steps were added to the pipeline that highly improved the accuracy of the segmentation. The achieved results are very promising and demonstrate the effectiveness of the presented method in terms of the segmentation accuracy. Moreover, the method has very low computational burden as the neural network as well as the preprocessing and postprocessing steps are very computationally efficient.

6 Acknowledgements

This research was supported by the Intramural Research Program of the National Cancer Institute, NIH.

References

1. Holland, E.C.: Progenitor cells and glioma formation. Current opinion in neurology 14(6), 683-688 (2001).
2. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Nature Scientific Data, (2017) [In Press]
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection, The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
5. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection, The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
6. Arikan, M., Fröhler, B., Möller, T.: Semi-automatic brain tumor segmentation using support vector machines and interactive seed selection. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 13-24 (2016).
7. Lefkovits, L., Lefkovits, S., Szilagyi, L.: Brain tumor segmentation with optimized random forest. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 13-24 (2016).
8. Lun, T.K., Hsu, W.: Brain tumor segmentation using deep convolutional neural network. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 26-29 (2016).
9. Zhao, X., Wu, Y., Song, G., Li, Z., Fan, Y., Zhang, Y.: Brain tumor segmentation using a fully convolutional neural network with conditional random fields. In: Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge, pp. 77-80 (2016).
10. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: Proceedings of AISTATS, pp. 562-570 (2015).
11. Xie, S., Tu, Z.: Holistically-nested edge detection. International Journal of Computer Vision, 1-16 (2017).

Fully Automatic Brain Tumour Segmentation using Random Forests and Patient survival prediction using XGBoost

Karthik Revanuru-karthikrvnr@gmail.com
Nameetha Shah-nameeta.shah@gmail.com

July 31, 2017

Abstract

Accurate and automated brain tumour segmentation from MRI images is important for performing quantitative analysis of MRI data. There is great potential to utilize quantitative imaging data as prognostic and predictive biomarkers for glioma patients. In this article we present two approaches, one for brain tumour segmentation task and the other for survival prediction. For segmentation task we propose a fully automatic method using random forests algorithm. For prediction task we compare XGBoost, random forests and logistic regression algorithms. We utilize T1, T2, FLAIR and T1 post contrast image series along with clinical parameter of age for predicting survival.

Keywords Brain Tumour segmentation, Survival Prediction, Random Forests, XGBoost, Gradient Magnitude, ResNet .

1 Introduction

In this paper we propose a novel approach for segmentation task using Random Forest. Features for random forest[6] include pixel location, pixel intensity, average of pixel intensities over a 5*5*5 region, gradient magnitude over 5*5*5 region and texture over a 20*20*20 region. These features are extracted from various MRI sequences like T1,T2, FLAIR and T1 post contrast images. For prediction task we propose an approach using XGBoost[7]. Features for XGBoost are extracted from MR images using ResNet[8]. We detail our methods in section 2 , experiments and results in section 3 and in section 4 we present conclusion and future ideas for work.

2 Methods

2.1 Segmentation

In segmentation task we have used random forest as our classifier and we have classified each pixel into one of the four classes (1 for NCR NET, 2 for ED, 4 for ET, and 0 for everything else). We have chosen spatial location of pixel, intensity of each pixel along average of intensities around the pixel over a 20*20*20 and 5*5*5 window, gradient magnitude over a 5*5*5 window and texture. Idea behind using this is to get both global context and local context at pixel level for classification.

Before giving the entire data to the classifier we have down sampled the data to remove skewness as well as to speed up the training step. We randomly pick pixels for each label so that each class has the same number of data points.

The parameter settings used within the Random Forest algorithm include:

- n estimators = 10 : The number of trees per forest.
- max features = sqrt(n features) : The number of features to consider when looking for the best split.
- min samples split = 2 : The minimum number of samples required to split an internal node.
- min samples leaf = 1 : The minimum number of samples required to be at a leaf node.
- n features = The number of features when fit is performed.

2.2 Survival Prediction

In survival prediction we have used ResNet(with soft-max layer removed) to extract features from each of T1,T2, FLAIR and T1 post contrast images. ResNet is a pre-trained network for image recognition which won the 1st place on the ILSVRC 2015 classification task. These extracted features along with age are then passed on to XGBoost for training and prediction.

XGBoost stands for extreme gradient boosting. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance. The features which make this algorithm dominating over others are:

- Sparse Aware implementation with automatic handling of missing data values.
- Block Structure to support the parallelization of tree construction.
- Continued Training so that you can further boost an already fitted model on new data.

Along with XGBoost we have also used Random Forest and Logistic Regression and compared their performance.

3 Experiments and Results

All the experiments were carried out using Scikit learn[9] and opencv[5] on python3.5. Code that has been developed for our experiments can be accessed here[1] and data-set[4][1][2][3] that has been used is provided by the BraTS 2017 organizers.

3.1 Segmentation

For segmentation task currently we have trained on ten scans from training data and we have not included texture as a feature. We would be adding texture as a feature and use all data for training by the final deadline. We have evaluated our method on data which was used in training and completely unseen data. The results are tabulated below.

.	0	1	2	4
0	72996	3	25	2
1	4	72820	40	162
2	15	141	72839	31
4	3	410	32	72581

The table above is a confusion matrix which has the count of pixels for each label and the data is seen in training. The accuracy is more than 99 percent for all classes

.	0	1	2	4
0	16380337	86308	1029549	217641
1	440	2137	9544	19555
2	734	3186	61373	13711
4	2357	979	6831	21318

The table above is a confusion matrix which has the count of pixels for each label and the data is completely unseen. Accuracy in test scenario is 92.2 percent for all classes.

[1] `https://github.com/karthikrvnr/Brain_Tumor_Segmentation`

3.2 Survival Prediction

For survival prediction we have divided the given data into 70 and 30 for training and testing respectively. The results for both classification and regression are tabulated below.

3.2.1 Regression

Regressor	Mean Avg Error in days
Logistic Regression	253
XGBoost	225
Random Forest	203

3.2.2 Classification

We have used XGBoost and Random Forest for classification.

.	Short	Medium	Long
Short	32	28	16
Medium	8	12	12
Long	20	24	28

The table above is a confusion matrix using XGBoost Classifier.

.	Short	Medium	Long
Short	40	20	16
Medium	8	12	12
Long	20	28	24

The table above is a confusion matrix using Random Forest Classifier.

4 Conclusion and future work

In this paper we have presented our approach for brain tumour segmentation and survival prediction using Random Forests. Continuing this work for final submission we want to include texture as one of the features and improve accuracy by using all the data for training. For survival task we would like to use Weighted Majority algorithm instead of giving age along with ResNet features to the classifier and see how it affects the accuracy.

References

[1] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. In *Nature Scientific Data*, 2017.

[2] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. In *The Cancer Imaging Archive*, 2017.

[3] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin Kirby, John Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. In *The Cancer Imaging Archive*, 2017.

[4] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, and Van Leemput K. The multimodal brain tumor image segmentation benchmark (brats). In *IEEE Transactions on Medical Imaging 34(10)*, pages 1993–2024, 2015.

[5] G. Bradski. Open cv. *Dr. Dobb's Journal of Software Tools*, 2000.

[6] Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.

[7] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[9] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830, November 2011.

Conditional Adversarial Network for Semantic Segmentation of Brain Tumor

Mina Rezaei, Konstantin Harmuth, Willi Gierke, Thomas Kellermeier, Martin Fischer, Haojin Yang, Christoph Meinel

Hasso Plattner Institute for Digital Engineering,
Prof.-Dr.-Helmert-Strae 2-3, 14482 Potsdam, Germany
{mina.rezaei,haojin.yang,christoph.meinel}@hpi.de{konstantin.harmuth,
willi.gierke,thomas.kellermeier,martin.fischer}@student.hpi.uni-potsdam.
de

Abstract. Automated medical image analysis has a significant value in diagnosis and treatment of lesions. Brain tumors segmentation has a special importance and difficulty due to the difference in appearances and shapes of the different tumor regions in magnetic resonance images. Additionally the data sets are heterogeneous and usually limited in size in comparison with the computer vision problems. The recently proposed adversarial training has shown promising results in generative image modeling. In this paper we propose a novel end-to-end trainable architecture for brain tumor semantic segmentation through conditional adversarial training. We exploit conditional Generative Adversarial Network (cGAN) and train a semantic segmentation Convolution Neural Network (CNN) along with an adversarial network that discriminates segmentation maps coming from the ground truth or from the segmentation network for BraTS 2017 segmentation task[15, 4, 2, 3]. We also propose an end-to-end trainable CNN for survival day prediction based on deep learning techniques for BraTS 2017 prediction task [15, 4, 2, 3]. The experimental results demonstrate the superior ability of the proposed approach for both tasks. The proposed model achieves on validation data a DICE score, Sensitivity and Specificity respectively 0.68, 0.99 and 0.98 for the whole tumor, regarding online judgment system.

Keywords: Conditional Generative Adversarial Network, Brain Tumor Semantic Segmentation, Survival day prediction

1 Introduction

Medical imaging plays an important role in disease diagnosis and treatment planning as well as clinical monitoring. The diversity of magnetic resonance imaging (MRI) acquisition regarding its settings (e.g. echo time, repetition time, etc.) and geometry (2D vs. 3D) also the difference in hardware (e.g. field strength, gradient performance, etc.) can yield variation in the appearance of the tumors that makes the automated segmentation challenging [8]. An accurate brain lesion segmentation algorithm based on multi-modal MR images might be able to improve

the prediction accuracy and efficiency for a better treatment planning and monitoring the disease progress. As mentioned by Menze et al. [15], in last few decades the number of clinical study for automatic brain lesion detection has grown significantly. In the last three years, Generative Adversarial Network(GAN) [6] become a very popular approach in various computer vision studies for example for classification [18, 13], object detection [11, 24], video prediction [14, 5, 23], image segmentation[9] and even mass segmentation for mammogram analysis [25]. In this work we address two tasks by BraTS-2017 [15, 4, 2, 3] challenges by two different approaches. Semantic segmentation is the task of classifying parts of images together that belong to the same object class. Inspired by the power of cGAN networks [25, 9], we propose an end-to-end trained adversarial deep structural network to perform brain High and Low Grade Glioma (HGG/LGG) tumor segmentation. We also illustrate how this model could be used to learn a multi-modal images, and provide preliminary results of an application for semantic segmentation. To this end we consider patient-wise "U-Net" [19] as a generator and "Markovian GAN" [10] as an discriminator. For the second task of BraTS-2017 [15, 4, 2, 3], we designed an end-to-end trainable CNNs on clinical data which enables to predict the survival day. The architecture use parallel CNN which one way is responsible to learn patient-wise MR images and another learned representation of clinical data. A detailed evaluation of the parameters variations and network architecture is provided. The contribution of this work can be summarized as following:

- We proposed a robust solution for brain tumors segmentation through conditional GAN. We achieved promising results on two type of brain tumor segmentation (The overall Dice for whole-tumor region is 0.68, Specificity 0.99 and Sensitivity 0.98).
- We proposed an automatic and trainable deep learning architecture for survival day prediction based on clinical data and MR images.

The rest of the paper is organized as follows: Chapter 2 describes the proposed approaches for semantic segmentation and survival day prediction, Chapter 3 presents the detailed experimental results. Chapter 4 concludes the paper and gives an outlook on future work.

2 Methodology

In this chapter we will describe first our proposed approach to the brain tumor sub-region segmentation based on deep learning and then our approach to the survival day prediction. The core techniques applied in our approach are depicted as well. In the GAN theory [6], the Discriminator Network (D) tries to decide if a certain input is sourced from the reference distribution, or has been generated by the Generator Network (G). The training procedure in G uses the pixel labels of certain multi-modal images and D tries to distinguish this certain boundary regions (we have three sub region tumor) comes from reference distribution or generative network. In order to incorporate more classes to this output

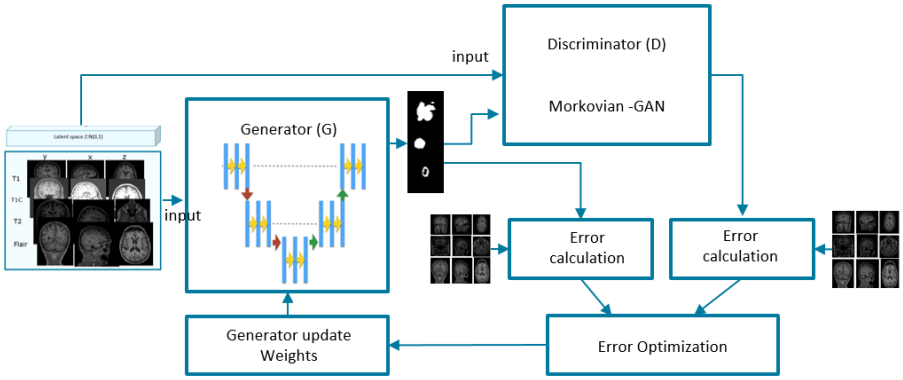


Fig. 1. The proposed architecture for semantic segmentation of brain tumor

while keeping with the GAN spirit of distinguishing distribution class instead of one example class, we could add additional input sources. As suggested by Goodfellow [6], one can consider the cGAN models with multi-class labels as:

1. GAN model with class-conditional models: which make the input label rather than the output. We ask GAN to generate specic classes. [16]
2. GAN model with N different output classes: that network trained by N different "real" and no "fake" classes. [21]
3. GAN models with N+1 different output classes: which the network train by N different "real" and an additional "fake" class. This type works very well for semi-supervised learning when it combined with feature matching GANs e.g. [20]

Therefor our proposed method lies in the second category as we consider for each multi-modal image three segmentation classes. Figure 1 describes the proposed approach to the brain tumor segmentation. In continue we describe the detail of techniques of pixel label classes for prediction in section 2.1 and for survival day prediction in section 2.2.

2.1 Brain Tumor Semantic Segmentation

We adapt the generator and discriminator architectures from [17, 9]. We applied Virtual-BatchNorm-Convolution [7] on generator network to make the "U-Net" [19] patient-wise. We choose "U-Net" architecture as generator because most of the deep learning approaches are patch-wise learning models, which ignore the contextual information within the whole image region. Like winner of BraTS-2016 [1], we come over this problem by leveraging global-based CNN methods (e.g. Seg-Net, Encoder-Decoder and FCN) and incorporating multi-modal of MRI data. We use Virtual-BatchNorm [7] in the generator network and Reference-BatchNorm [7]in the discriminator network to reduce over-fitting.

The discriminative network is based on "Markovian GAN" [17]. Then two models trainable simultaneously through back propagation, corresponds to a minimax two-player game. An "U-Net" generative model G; Captures the data distribution, pixel segmentation and train to minimize the probability of D making a mistake. A "Markovian GAN" discriminative model D: to estimate the probability that a sample came from the training data rather than G.

2.2 Survival Day Prediction

Figure 2 describes our solution for survival day prediction. We proposed a two path way architecture which one has several CNN and it is responsible for multi-modal image representation and another learned the clinical data features. The extracted features from each path way, concatenated in next step to shared the learned features. Then they passed to two fully connected layers to learn the survival day. We use Virtual-BatchNorm [7] on the CNNs network which learned image representation. To prevent over-fitting, we generated augmented images through horizontal and vertical flipping and re-scaling. We applied Mean squared error as Loss function. We mapped the clinical data (Ages and survival days) into float[0,1].

3 Experiments

In order to evaluate the performance of the proposed cGANs method, we test the method on two types of brain tumor data provided by BraTs 2017 challenge [15, 4, 2, 3]. We applied a bias field correction on the MR images to correct the intensity non-uniformity in MR images by using N4ITK [22]. In next step of pre-processing we applied histogram matching normalization [12]. We train both the generator and the discriminator to make them stronger together and avoid making one network significantly stronger than the other by taking turn. We consider multi-madal images from same patient in each batch during training and use all the released data by BraTS 2017 challenge[15, 4, 2, 3] in training time which is 75 patients with Low Grade Glioma(LGG) and 210 patients with High Grade Glioma(HGG). We used all prepared image-modal from three axes of x,y,z (3x4x155x285) that the input and output are 4-3 channel images(4:image-modal; 3:three sub-region of each tumor type). We get better result when don't shuffle input data in generator network. In generator network Sign function helps for noise reduction. The generator for all layers use ReLU activation function except output layer which use Tanh. Qualitative results are shown in Figures 3. On this size data sets (530100 2D images with the size of 250x250) training took around 72 hours on parallel Pascal Titan X GPUs. Table 11 shows the results of the proposed models evaluated at BraTS 2017 online judge system. The evaluation system uses three tasks. The online system provides the results as follows: The tumor structures are grouped in three different tumor regions. This is mainly due to practical clinical applications. As described by BraTS 2017 [15, 4, 2, 3], tumor regions are defined as:

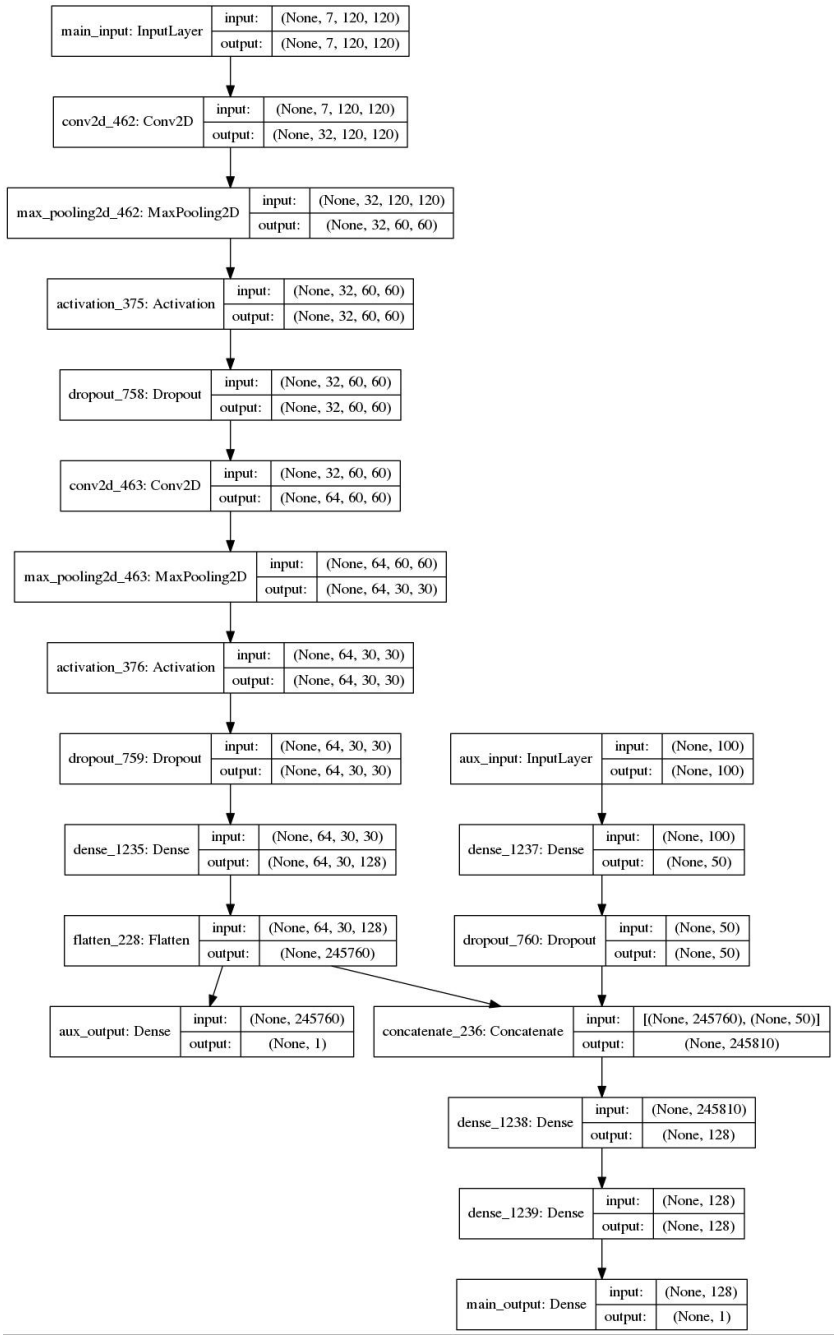


Fig. 2. The propose architecture for survival day prediction

1. WT: Whole tumor region represents the area with all labels 1,2,3,4 which 0 for normal tissue, 1 for edema, 2 for non-enhancing core, 3 for necrotic core, 4 shows enhancing core.
2. CT: Core tumor region represent only tumor core region, it measures label 1,3,4.
3. ET: Enhancing tumor region (label 4)

There are four kinds of evaluation criteria for segmentation task like Dice score, Hausdorff distance, Sensitivity and Specificity has provided by BraTS 2017 challenge organizer as an online judgment system.

Table 1. Preliminary results till now from BraTS-2017 online judge system on Validation data(unseen data)

	Whole Tumor	Core of Tumor	Enhanced Tumor
Dice	0.70	0.55	0.40
Sensitivity	0.68	0.52	0.99
Specificity	0.99	0.99	0.99

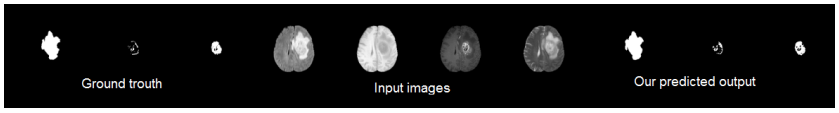


Fig. 3. The output segmentation result on training data

Table 1 shows the preliminary results but our work is still on the progress. Table2 shows the survival day prediction results.

Table 2. Preliminary results on survival day prediction. We used 70% of the data (115 patients) for training, 10% (16 patients) for validation and 20% (32 patients) for testing. The first path way of CNN has seven input channel which four from multi-modal images and three from segmented regions. We translated ages from interval [0, 100] into float [0,1] and also for survival day did from [0-1750] days into float of [0,1].

Data	Accuracy
Validation	73.1%
Test	64.08%

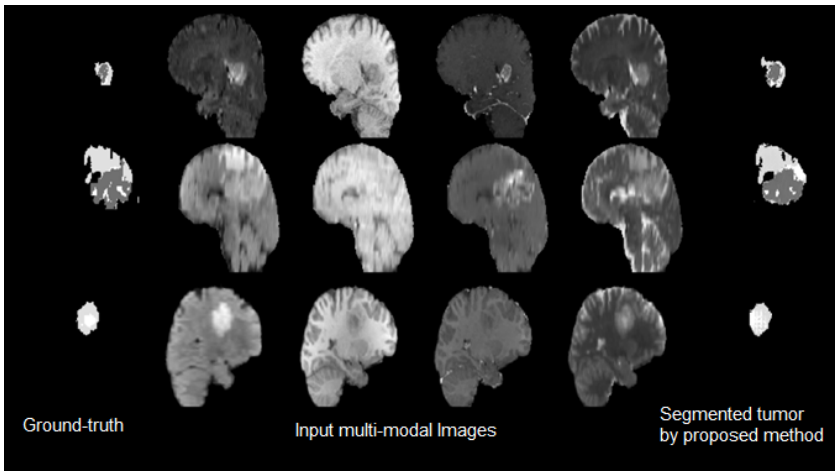


Fig. 4. The output segmentation result on training data

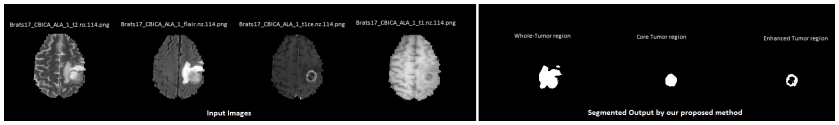


Fig. 5. The preliminary segmentation result on validation data

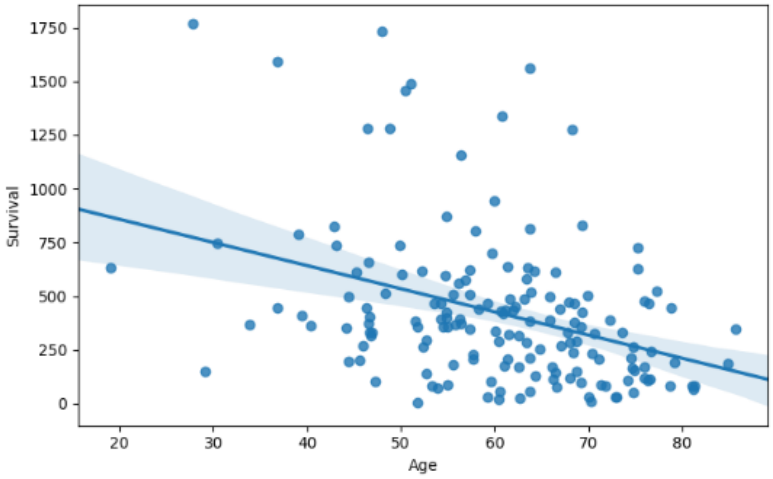


Fig. 6. clinical data distribution from training set

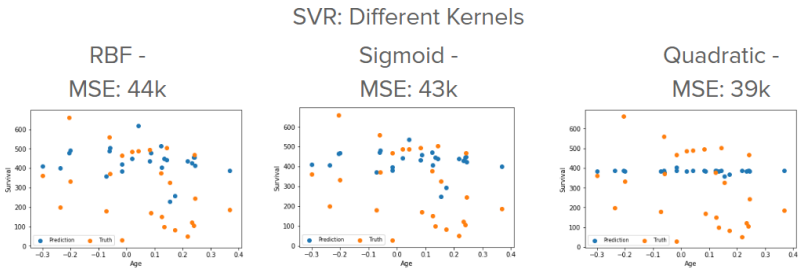


Fig. 7. different regression techniques (e. g. Support Vector Regression, Polynomial Regression,) for survival day prediction.

4 Conclusion

In this paper, we propose and evaluated approaches for two important clinical tasks: brain tumor segmentation and prediction of survival day after tumor diagnosis. The proposed approach for tumor segmentation is end-to-end trainable based on the newly proposed conditional generative adversarial network. Furthermore, adversarial training is used to handle the global-based CNN in generator to reduce over-fitting and increase robustness. We proposed an automated trainable parallel convolution neural network to predict the survival day as the second task in the challenge. These networks learn a loss adapted to the task and data at hand, which makes it applicable in unseen data. For the future work, we look for further improvement on generative network by incorporating recurrent neural network(RNN) inside of our Encoder-Decoder.

References

1. https://www.cbica.upenn.edu/sbia/Spyridon.Bakas/MICCAI_BraTS/MICCAI_BraTS_2016_proceedings.pdf
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017)
5. Finn, C., Goodfellow, I.J., Levine, S.: Unsupervised learning for physical interaction through video prediction. CoRR abs/1605.07157 (2016), http://arxiv.org/abs/1605.07157

6. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. ArXiv e-prints (2014)
7. Goodfellow, I.J.: NIPS 2016 tutorial: Generative adversarial networks. CoRR abs/1701.00160 (2017)
8. Inda, Maria-del-Mar, R.B., Seoane, J.: Glioblastoma multiforme: A look inside its heterogeneous nature. In: Cancer Archive 226-239 (2014)
9. Isola, P., Zhu, J., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. CoRR abs/1611.07004 (2016), http://arxiv.org/abs/1611.07004
10. Li, C., Wand, M.: Precomputed real-time texture synthesis with markovian generative adversarial networks. CoRR abs/1604.04382 (2016), http://arxiv.org/abs/1604.04382
11. Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., Yan, S.: Perceptual Generative Adversarial Networks for Small Object Detection. ArXiv e-prints (Jun 2017)
12. Lszl G. Nyl, J.K.U., Zhang, X.: New variants of a method of mri scale standardization. IEEE transactions on medical imaging 29(6) (2000)
13. Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I.J.: Adversarial autoencoders. CoRR abs/1511.05644 (2015), http://arxiv.org/abs/1511.05644
14. Mathieu, M., Couprie, C., LeCun, Y.: Deep multi-scale video prediction beyond mean square error. CoRR abs/1511.05440 (2015), http://arxiv.org/abs/1511.05440
15. Menze, B., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M., Arbel, T., Avants, B., Ayache, N., Buendia, P., Collins, D., Cordier, N., Corso, J., Criminisi, A., Das, T., Delingette, H., Demiralp, ., Durst, C., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, D., Hamamci, A., Iftekharuddin, K., Jena, R., John, N., Konukoglu, E., Lashkari, D., Mariz, J., Meier, R., Pereira, S., Precup, D., Price, S., Raviv, T., Reza, S., Ryan, S., Sarikaya, D., Schwartz, L., Shin, H., Shotton, J., Silva, C., Sousa, N., Subbanna, N., Szekely, G., Taylor, T., Thomas, O., Tustison, N., Unal, G., Vasseur, F., Wintermark, M., Ye, D., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging 34(10), 1993–2024 (2015)
16. Mirza, M., Osindero, S.: Conditional generative adversarial nets. CoRR abs/1411.1784 (2014), http://arxiv.org/abs/1411.1784
17. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR abs/1511.06434 (2015), http://arxiv.org/abs/1511.06434
18. Reed, S.E., Akata, Z., Mohan, S., Tenka, S., Schiele, B., Lee, H.: Learning what and where to draw. In: Lee, D.D., Sugiyama, M., Luxburg, U.V., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems 29, pp. 217–225. Curran Associates, Inc. (2016), http://papers.nips.cc/paper/6111-learning-what-and-where-to-draw.pdf
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer International Publishing (2015)
20. Salimans, T., Goodfellow, I.J., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. CoRR abs/1606.03498 (2016), http://arxiv.org/abs/1606.03498

21. Springenberg, J.T.: Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks. ArXiv e-prints (Nov 2015)
22. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging 29(6), 1310–1320 (2010)
23. Vondrick, C., Pirsiavash, H., Torralba, A.: Generating videos with scene dynamics. CoRR abs/1609.02612 (2016), http://arxiv.org/abs/1609.02612
24. Wang, X., Shrivastava, A., Gupta, A.: A-fast-rcnn: Hard positive generation via adversary for object detection. arXiv preprint arXiv:1704.03414 (2017)
25. Zhu, W., Xie, X.: Adversarial deep structural networks for mammographic mass segmentation. CoRR abs/1612.05970 (2016), http://arxiv.org/abs/1612.05970

Brain tumor segmentation using CNN based method

Sara Sedlar

sara.sedlar@gmail.com

Abstract. In this paper an automatic brain tumor segmentation approach based on Convolutional Neural Networks (CNN) is proposed. The model architecture is composed of two small CNNs, one extracting local and the other one features from larger region [5]-[6]. In addition to the information extracted from the differences in the appearance between healthy and tumorous tissues, the method exploits the facts that very often tumor introduces high asymmetry to the brain and that the most difficult task is the classification within tumor (between different tumor stadiums) and within tumor's proximity (between healthy and tumorous tissue). The method training and hyper-parameter tuning were performed on the BRATS 2017 training dataset and the final results are reported on the BRATS 2017 validation dataset. The Dice scores for the validation dataset are 0.64182, 0.80571 and 0.65275 for enhancing tumor, whole tumor and tumor core, respectively.

Keywords: brain tumor, tumor segmentation, CNN segmentation

1 Introduction

Tumor segmentation is an important step in the evaluation of tumor's grade, monitoring and modeling of its progress, what is necessary for successful therapy and surgery planning. Manual tumor segmentation is time consuming and it is prone to the intra and inter rater variability. In recent years, the challenge of the automatic tumor segmentation has attracted a significant attention of scientist and engineers. Numerous discriminative and generative approaches were proposed and the methods based on neural networks are becoming the most dominant [5]-[7]. The article is structured as follow. In the Section 1 a short overview of the database used for training and validation is provided. In the Section 2 the proposed method is described. In the Section 3 experiment details are given. The results are presented in the Section 4 and at the end conclusions and future work are provided in the Section 5.

2 Database

BraTS 2017 training database is composed of 210 high grade glioma (HGG) and 75 low grade glioma (LGG) tumor cases. For each case four magnet resonance

imaging (MRI) modalities (T1, T2, T1CE, flair) are provided and ground truth for four classes, namely healthy tissue, necrotic and non-enhancing tumor, peritumoral edema and enhancing tumor [1]-[2]. BraTS 2017 validation database contains 46 cases without information about tumor's grade and ground truth segmentation. In order to validate algorithm's performance as in real world scenario, training database is split into two subset. The first subset contains two thirds of training data (140 HGG, 50 LGG) and it is used for segmentation model training. The second subset contains one third of the training data (70 HGG, 25 LGG) and it is used to determine when to stop algorithm's training, to guide the training process by changing hyper-parameters and to tune post-processing thresholds. The validation set is not seen during any part of the training process.

3 Methods

3.1 Pre-processing

Each volume is normalized using mean and standard deviation of its non-zero voxels.

3.2 Segmentation model

The proposition of this brain tumor segmentation method is motivated by the success of the methods that are using parallel CNNs to extract features from the regions of different sizes [5]-[6]. The model architecture is composed of two small CNNs. One CNN is trained to extract features from patches of larger size and to exploit information about brain's asymmetry and patch position (large region CNN). The other CNN is trained to extract local features from small patches (small region CNN) and it could be seen as means of segmentation refinement. The small and large region CNN features are merged by a fully connected layer and are further used for voxel classification.

Another important information is related to the fact that the patches capturing tumorous regions are much more difficult to discriminate than those capturing exclusively healthy and mostly tumorous tissues. This is addressed by representing segmentation process as a combination of two classification problems. The first problem is two class problem of discrimination between healthy and tumorous tissues. Training patches for this part are sampled from the entire brain region. Another, more difficult, problem is classification of the voxels into four classes: healthy tissue, necrotic and non-enhancing tumor, peritumoral edema and enhancing tumor. Training patches for the four class segmentation are randomly selected from the tumor region and its close neighborhood.

Large region CNN architecture and input data. The input to the large region CNN is composed of 11 patches. For a given voxel for each modality a 33 x 33 patch and its corresponding pair from the position mirrored with respect to the sagittal plane are extracted. This enables the model to exploit asymmetry

introduced by tumor. In addition to those 8 volume patches, three patches that contain normalized distances from the brain's center along all three axis are extracted as well.

The structure of CNN used for large region feature extraction.

1. Convolutional layer 1
   (a) convolution 5 x 5 x 11 x 32
   (b) ReLU neuron activation
2. Convolutional layer 2
   (a) convolution kernels 5 x 5 x 32 x 64
   (b) ReLU neuron activation
   (c) max pooling with stride 2 x 2
3. Convolutional layer 3
   (a) convolution 5 x 5 x 64 x 128
   (b) ReLU neuron activation
4. Fully connected layer 1
   (a) fully connected 8192 x 512
   (b) ReLU neuron activation
5. Fully connected layer 2
   (a) fully connected 512 x 32
   (b) ReLU neuron activation

Small region CNN architecture and input data. The input to the small region CNN is composed of 4 patches. For a given voxel for each modality a 13 x 13 sized patch is extracted.

The structure of CNN used for small region feature extraction.

1. Convolutional layer 1
   (a) convolution 5 x 5 x 4 x 32
   (b) ReLU neuron activation
2. Convolutional layer 2
   (a) convolution 5 x 5 x 32 x 64
   (b) ReLU neuron activation
3. Fully connected layer 1
   (a) fully connected 1600 x 256
   (b) ReLU neuron activation
4. Fully connected layer 2
   (a) fully connected 256 x 32
   (b) ReLU neuron activation

Small and large region feature merging. The structure of the fully connected network used for merging small and large region features.

1. Fully connected layer 1
   (a) fully connected 64 x 32
   (b) ReLU neuron activation
2. Fully connected layer 2
   (a) fully connected 32 x 4
   (b) Softmax

Loss function is a linear combination of log-losses of two and four class classification problems.

In order to prevent over-fitting to the segmentation model training data the disturb labels approach is used [8]. In this approach one part of the labels is set to a label selected in a random manner from set of existing labels.

3.3 Post-processing

In the post-processing stage all the voxels classified as tumorous with low prediction score are set as healthy. The remaining regions classified as tumorous are labeled using connected-component method. If the size of a region is below certain threshold or the mean prediction score of the region is not high enough, it is re-classified as a healthy region.

4 Implementation details

The algorithm implementation was done in Python programming language. For data loading and saving NiBabel library was used, volume pre and post-processing were performed by the standard libraries numpy and scipy and the segmentation model was created and trained with tensorflow. The hardware employed was composed of nVidia's GeForce GTX 980 Ti (6 GB) GPU and Intel Core i7-6700K CPU @ 4.00 GHz (32 GB). The training lasted approximately 20h and the testing lasts around 3.5 mins per volume.

In order to facilitate training process some meta data were computed before training such as normalization parameters, brain masks and tumor distance maps. Brain mask corresponds to the volume's region where all modalities are non-zero. Tumor distance map contains distances to the tumor for each voxel that is present in the brain mask and it is used to select data for two and four class classification problems. It is computed using 3d dilation morphology operator.

5 Results

In the Table 1 Dice scores obtained on the validation dataset, not seen by the training algorithm, are provided.

Table 1. Results of tumor segmentation on BraTS 2017 validation dataset.

	Enhancing Tumor	Whole Tumor	Tumor Core
Dice score	0.64182	0.80571	0.65275

6 Discussion and conclusions

In this paper a CNN based approach for brain tumor segmentation is presented. It was shown that a small segmentation model can achieve considerably promising results. The future work will include increase of model's size and training on larger amount of data (entire training dataset and additional data produced by augmentation).

References

1. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Kamnitsas Konstantinos, Ferrante Enzo, Parisot Sarah, Ledig Christian, Nori Aditya, Criminisi Antonio, Rueckert Daniel, Glocker, Ben. (2016). DeepMedic for Brain Tumor Segmentation. 138-149. 10.1007/978-3-319-55524-9_14.
6. Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle, "Brain tumor segmentation with Deep Neural Networks", Medical Image Analysis, Volume 35, 2017, Pages 18-31, ISSN 1361-8415
7. Liya Zhao and Kebin Jia, Multiscale CNNs for Brain Tumor Segmentation and Diagnosis, Computational and Mathematical Methods in Medicine, vol. 2016, Article ID 8356294, 7 pages, 2016. doi:10.1155/2016/8356294
8. Lingxi Xie and Jingdong Wang and Zhen Wei and Meng Wang and Qi Tian, "DisturbLabel: Regularizing CNN on the Loss Layer"

Brain Tumor Segmentation using Dense Fully Convolutional Neural Network

Mazhar Shaikh, Gagan Acharya, Ganesh Anand, Abhijit Amrutkar, Varghese Alex, and Ganapathy Krishnamurthi

Medical Imaging and Reconstruction Lab, Department of Engineering Design, Indian Institute of Technology Madras, Chennai, India
gankrish@iitm.ac.in

Abstract. Accurate tumor segmentation is crucial for treatment and survival prediction of cancer patients. Manual segmentation of brain tumor is often time consuming and the performance of the segmentation varies based on the operators experience. This leads to the requisition of fully automatic method for brain tumor segmentation. In this paper, we propose a fully automatic method for the segmentation of brain tumor from multi modal MR images, which is evolved by integrating a densely connected fully convolutional neural network (FCNN), followed by post-processing using a Dense Conditional Random Field (DCRF). The proposed FCNN consists of blocks of densely connected layers, transition down layers in down-sampling path and transition up layers in up-sampling path. The method was tested on dataset provided by Multi modal Brain Tumor Segmentation Challenge (BraTS) 2017. The training data is composed of 210 high-grade brain tumor and 75 low-grade brain tumor cases. On the BraTS 2017 validation data, the proposed network achieves a mean whole tumor, tumor core & active tumor dice score of 0.87, 0.68 & 0.65. respectively.

Keywords: Fully convolutional neural networks, multi modal MRI segmentation, conditional random fields

1 Introduction

Segmentation of the gliomas from MR images is the preliminary step for treatment and surgical planning. Manual segmentation of gliomas are tedious and results in inter rater variability. In this paper, we propose a 103 layer deep fully convolution neural network (FCNN) for automatic segmentation of gliomas. The network was trained on 2-D axial slices of the brain. Densely connected Conditional Random Fields was incorporated into the framework as a post processing tool. Additionally, connected components analysis was used to remove false positives generated by the network.

2 Materials & Method

2.1 Data

The images used to train and validate this model were obtained from the BraTS 2017 challenge dataset [1], [2]. The training dataset consisted of multi modal MR images of 284 patients, with 210 patients from the high grade gliomas category (HGG) and 74 patients from the low grade gliomas (LGG) category. The following MRI modalities were provided for each patient: T2-weighted fluid attenuated inversion recovery (FLAIR), T1-weighted (T1), T1-weighted contrast-enhanced (T1ce), and T2-weighted (T2). The provided images were co-registered to the same anatomical template, interpolated to the same resolution (1 mm^3) and skull-stripped. The image dimension is 240×240×155, with 155 being the number of slices in the axial direction. Manually annotated ground truth segmentations were provided for three classes : GD-enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor (NCR/NET label 1). The network was trained on slices extracted from the axial plane.

2.2 Pre-processing

Multi modal scans can vary between patients depending on several factors including the instrument used, image acquisition axis, etc. In order to account for the patient-to-patient variation in the MR images, we adopted z-score normalization where we subtract the mean and divide by the standard deviation of the entire volume for each of the four channels of an individual's scan.

2.3 Densely connected FCNN model

Our segmentation technique is based on the One Hundred Layers Tiramisu model proposed for semantic segmentation originally by Simon Jgou et al [6]. Like most state-of-the-art models, the Tiramisu model involves a down-sampling path and up-sampling path, where a single slice of the brain is provided as input at the beginning of the model and class-wise probabilities for every pixel is output at the end of the up-sampling path.

The Tiramisu model ,shown in Fig. (1), consists of dense blocks(DB) used in DenseNet [5], which are made up of repeated Batch Normalization layers, ReLU, 3×3 convolutions and small skip connections. The dense blocks are paired with transition down layers(TD) in the down-sampling path and transition up layers(TU) in the up-sampling path. The transition down layer consists of 1×1 convolutions followed by 2×2 max-pool layer with stride 2. Transition up layer is composed of 3×3 transpose convolutions with stride 2. The various blocks used in the model are shown in Fig. (2).

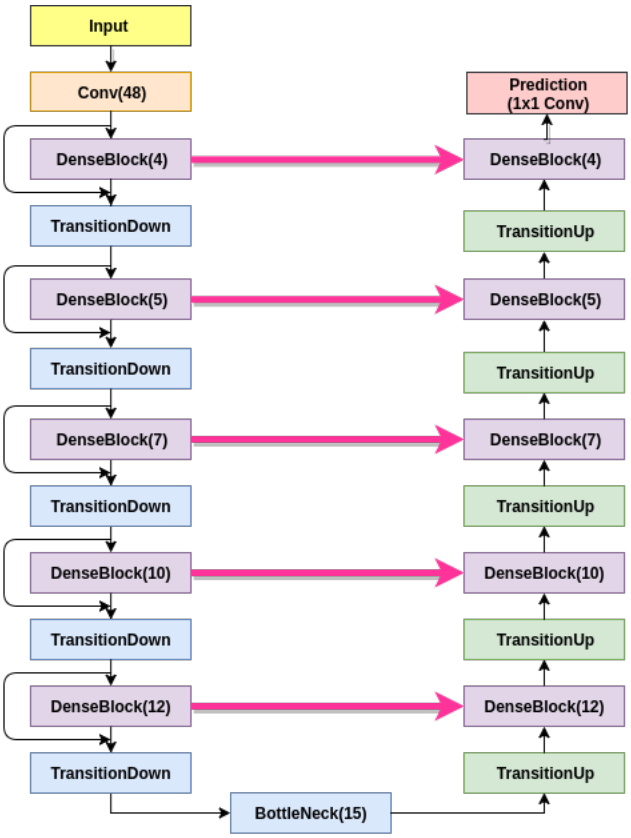
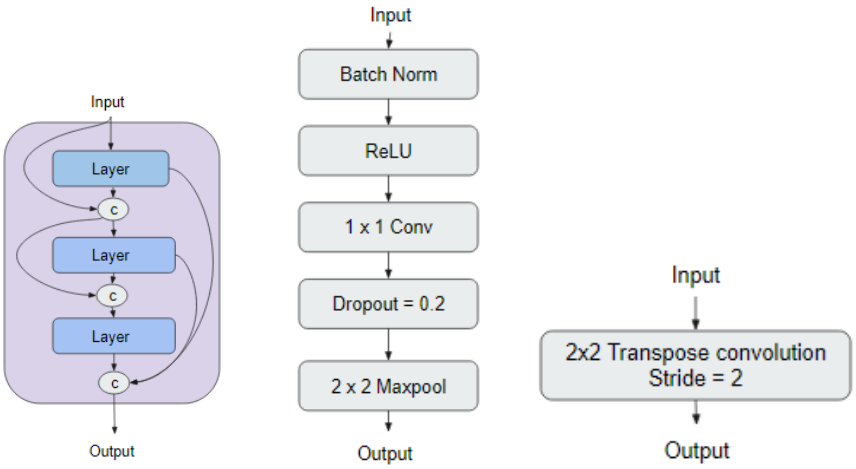


Fig. 1: Architecture of the proposed network



(a) (b) (c)

Fig. 2: Blocks used in the model. a) Dense Block. b) Transition Down. c) Transition Up

2.4 Post-processing using Dense-CRFs and Connected Components Analysis

To smoothen the segmentation predicted by the above model, we used fully connected conditional random fields with Gaussian edge potentials as proposed by Krähenbühl et al [7]. The unary potentials used by the CRF was computed using the predicted softmax probabilities. Further, the MR brain slice (all four modalities) along with the computed unary potentials was used for inferring the pixel labels.

The false positives in the prediction were further reduced by using connected component analysis, wherein only the largest component was retained.

3 Results

The performance of the proposed technique on the local HGG test data (n=21) is shown in Table (1) & Fig. (3). On the local test HGG data, the network achieved as mean whole tumor, tumor core and active tumor dice score of 0.84, 0.83, 0.80 respectively. The proposed post processing technique (CRF+ connected component analysis) yield a 1 % improvement in the whole tumor dice score, 1% in tumor core and 0.5 % in active tumor respectively.

On the local LGG test data (n=8), The performance of the proposed technique is shown in Table (2). Compared to HGG, the model underperforms on

Tumor core segmentation, while maintaining good performance on the whole tumor segmentation.

Table 1: Results of local test HGG data

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.84	0.83	0.80
Std Deviation	0.16	0.18	0.14
Median	0.89	0.87	0.84

Table 2: Results of local test LGG data

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.82	0.43	–
Std Deviation	0.11	0.29	–
Median	0.85	0.44	–

The performance of the proposed technique on the BraTS 2017 validation data (mixture of HGG and LGG) in given in Table (3). For whole tumor segmentation, the network maintains its performance on the validation data. However, a dip in performance was observed in the tumor core and active tumor regions on the validation set when compared to the local test data. The poor performance of the proposed technique on LGG tumor core segmentation negatively skews the performance statistics of our method on the validation data.

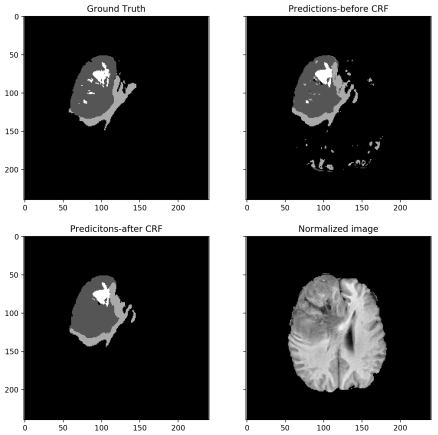
Table 3: Results of BraTS 2017 validation data, (n= 40)

	Whole Tumor	Tumor Core	Active Tumor
Mean	0.87	0.68	0.65
Std Deviation	0.11	0.34	0.32
Median	0.91	0.82	0.78

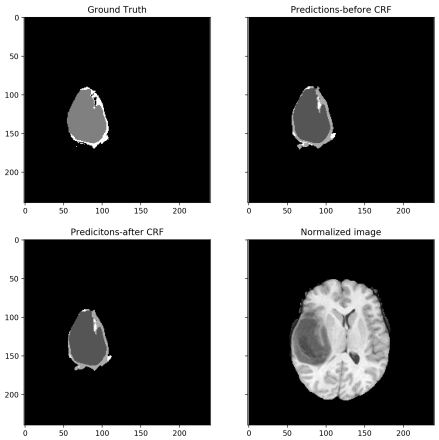
4 Conclusion

In this paper,we propose an automatic technique to segment gliomas from MR scans.

– A 103 layer deep network was implemented for segmentation of the gliomas from MR scans.



(a)



(b)

Fig. 3: Results of the proposed network on local test data. For each subfigure (Left to Right, Top to Bottom), Ground truth, Prediction(before post-processing), Prediction(after post-processing) , and the normalized FLAIR image slice, in that order.

– A single network was used for the segmentation task, irrespective of the grade of the glioma.
– The Dense CRF improved the performance of the network in all compartments.
– The proposed network completes the entire pipeline (preprocessing, prediction & post processing) under 30 seconds.

The network was developed using TensorFlow framework.

References

1. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
2. Menze BH et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Huang G, Liu Z, Weinberger KQ, van der Maaten L. "Densely connected convolutional networks. arXiv preprint arXiv:1608.06993. 2016 Aug 25.
6. Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. arXiv preprint arXiv:1611.09326. 2016 Nov 28.
7. Krhenbhl, Philipp, and Vladlen Koltun. "Efficient inference in fully connected crfs with gaussian edge potentials." Advances in neural information processing systems. 2011.

Glioblastoma and Survival Prediction

Zeina Shboul, Lasitha Vidyaratne, Mahbubul Alam, Syed M. S. Reza and Khan M. Iftekharuddin,

Vision Lab, Electrical & Computer Engineering, Old Dominion University
{zshbo001, lvidy001, malam001, sreza002, kiftekha}@odu.edu

Abstract. Glioblastoma is a stage IV highly invasive astrocytoma tumor. Its heterogeneous appearance in MRI poses critical challenge in diagnosis, prognosis and survival prediction. This work extracts a total of 1207 different types of texture and other features, tests their significance and prognostic values, and then utilizes the most significant features with Random Forest regression model to perform survival prediction. We use 163 cases from BraTS17 training dataset for evaluation of the proposed model. A 10-fold cross validation offers normalized root mean square error of 30% for the training dataset and the cross validated accuracy of 67%, respectively.

1 Introduction

Glioblastoma (GB) is categorized as a World Health Organization (WHO) stage IV brain cancer that originates in a star-shaped brain cells in the cerebrum called astrocytes [1] [2]. GB is the most invasive brain tumor and its highly diffusive infiltrative characteristics makes glioma a lethal disease [3] with a median survival of 14.6 months with radiotherapy and temozolomide, and 12.1 months with radiotherapy alone [4]. In addition, heterogeneity in GBM [5] poses further challenge not just for diagnosis, but also for prognosis and survival prediction using MR imaging.

In [6], the authors use the different subtype tumor volumes, extent of resection, location, size and other imaging features in order to evaluate the capability of these features in predicting survival. The authors in [7] use comprehensive visual features set known as VASARI (Visually AcceSAble Rembrandt Images) [8] in order to predict survival and correlate these features for genetic alterations and molecular subtypes. In [9], the authors quantify large number of radiomic image features including shape and texture in computed tomography images of lung and head-and-neck cancer patients.

This work discusses overall survival prediction using Random forest regression model based on different structural multiresolution texture features, volumetric, and histogram features. However, accurate representative tumor features requires accurate tumor segmentation. The recent developments in deep learning domain have opened up new avenues in various medical image processing research. Several recent studies [10] [11] apply Convolutional Neural Network (CNN) based deep learning techniques to

solve brain tumor segmentation problem successfully. Consequently, this work implements a state-of-the-art CNN architecture following [10] to enhance brain tumor segmentation task.

2 Dataset

In this study, we use MR images of 163 high grade GBM patients with overall survival (in days) data from BtaTS17 training data set [12] [13] [14] [15] (the median age, 61.167 years; range, 18.975-78.762 years). The available scans of the MRI are native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. The dataset are co-registered, re-sampled to 1 mm^3 and skull-stripped.

3 Methodology

3.1 Brain tumor segmentation

Accurate segmentation of tumor from the MRI is pre-requisite for survival prediction as most potent features are derived from the affected region. The complete pipeline for survival prediction is shown in Fig. 1. Note this paper primarily explains the proposed survival model.

Our previous works on multiclass MRI brain tumor segmentation using texture based features [16] [17] have yielded important results. The detail description of multiclass abnormal brain tumor segmentation is found in [16] [17]. Our texture features are extracted from raw (T1Gd, T2, and Flair) modalities tumor volumes. The texture representations are piecewise triangular prism surface area (PTPSA) [18], multifractional Brownian motion (mBm) [19], Generalized multifrational Brownian motion (GmBm) [20] [21], and five representations of Texton filters [22], respectively. This work further improves the segmentation performance by employing a two-stage process in which the outcomes from the deep learning based method are fused with that of a handcrafted feature based method that utilizes Random Forest (RF) for the classification task.

The input to the CNN are image patches where the 3rd dimension is comprised of the four MRI modalities: T1, T1Gd, T2, and FLAIR. The output of CNN is the classification of five tissues such as background, enhanced tumor, edema, necrosis and non-enhanced tumor, respectively. All the inputs to the CNN are pre-processed with N4-ITK bias correction, and intensity normalization for inter-volume consistency [10]. The training set is image patches randomly obtained from the BRATS 2017 training MRI volume set. The sufficiently trained CNN is subsequently used for the segmentation of testing data as shown in Fig. 1.

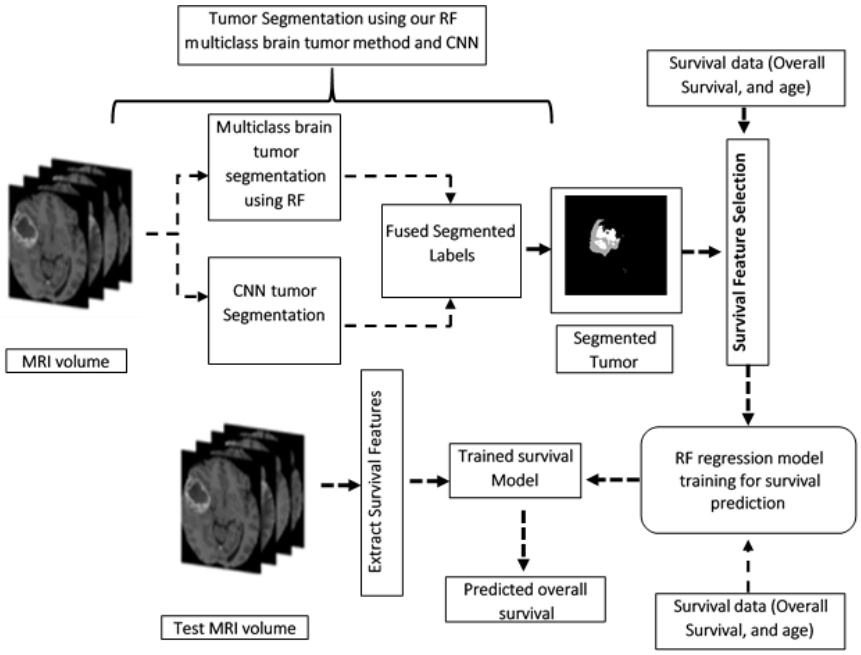


Fig. 1. Pipeline for tumor segmentation and survival prediction.

3.2 Survival Prediction

In this study we address the association between structural multiresolution texture features and overall survival. We extract 42 features from each raw MRI modality and texture representation of the whole tumor volume. These features are described by histogram, the co-occurrence matrix (measure the texture of image), the neighborhood gray tone difference matrix (measure a grayscale difference between pixels with certain grayscale and their neighboring pixels) and the run length matrix (capture the coarseness of a texture). In addition, we extract 5 volumetric and 6 histogram features from the tumor and the different tumor sub-regions (edema, enhancing tumor, and tumor core). Further, the tumor locations and the spread of the tumor in the brain are also considered. Finally, 9 area properties are extracted from the whole tumor from three viewpoints (view are set along x, y and z axis).

Feature selection is performed in three steps; first significant features were selected using a univariate cox regression model. Then, another univariate cox regression is applied on the quantified significant features. This ensure that these features are able to split the dataset into short vs. long survival. A total of two hundred and forty (240)

features out of one thousand and two hundred and seven (1207) features (~ 20 % of the total extracted features) are found to be significant in the previous two feature selection steps.

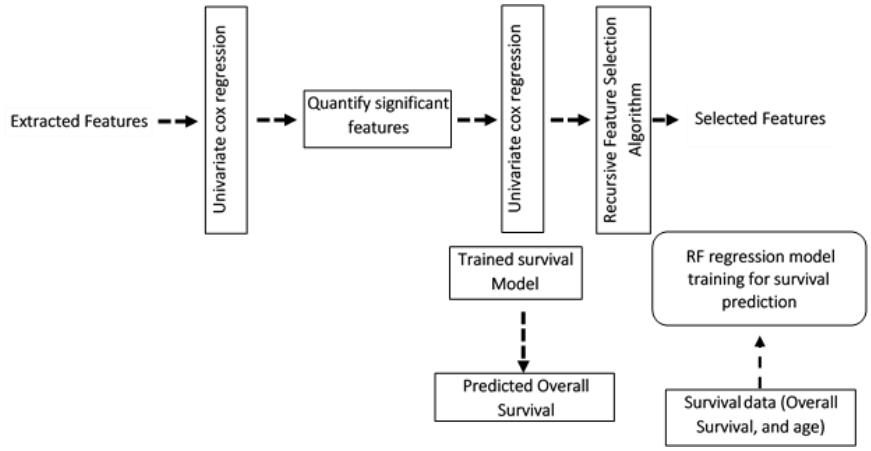


Fig. 2. Feature selection steps and survival model Pipeline

Finally, the two hundred and forty features are reduced to forty significant features using a recursive feature selection algorithm. The steps of feature selection and survival prediction is shown in Fig. 2. The selected features are then fed into Radom Forest regression model [23] for survival prediction.

4 Experimental Results

We perform tenfold cross-validation on survival prediction features extracted from the ground truth, which is available with the BraTS17 training dataset (163 patients), in order to evaluate the performance of the proposed RF survival regression model. We use the normalized root mean square error (NRMSE) of the overall survival predicted values as metric for evaluation. However, the survival prediction features are extracted from the fused segmented tumor using the BraTS17 validation dataset as described in Fig. 1.

The cross validated NRMSE of the training dataset is 30%. In addition, to evaluate the performance of the survival model based on classification, the overall survival is divided into three classes (long, medium and short) survivors corresponds to (>15 months, >10 months and <15 months, <10months), respectively. The cross validated accuracy is 67%. At the time of writing this paper, the evaluation of the validation dataset is not available.

5 Conclusion

In this work, we present a complete pipeline to perform survival prediction starting with brain tumor segmentation. The method uses BraTS17 validation dataset in a two-stage process for segmentation in which the outcomes from the deep learning based method are fused with that of our handcrafted feature based method. The segmented tumor volumes along with other volumetric, and histogram features are used in an RF regression model for survival prediction. We achieve cross validated NRMSE of 30% on the BraTS17 dataset and cross validated accuracy of 67%, respectively.

6 Acknowledgements

This work was funded by NIBIB/NIH grant# R01 EB020683.

References

[1] Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, and Kleihues P, "The 2007 WHO Classification of Tumours of the Central Nervous System," *Acta Neuropathologica,* vol. 114, no. 2, p. 97–109, 2007.

[2] Kliehues P, Burger PC, Collins VP, Newcombe EW, Ohgaki H, and Cavenee WK, "Glioblastoma," in *WHO Classification of Tumours: Pathology and Genetics of Tumours of the Nervous System*, Lyon, France, International Agency for Research on Cancer, 2000, pp. 29-39.

[3] Claes A, Idema AJ, and Wesseling P, "Diffuse Glioma Growth: a Guerilla War," *Acta Neuropathologica,* vol. 114, no. 5, p. 443–458, 2007.

[4] Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, and Mirimanoff RO, "Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma," *New England Journal of Medicine,* vol. 352, no. 10, pp. 987-996, 2005.

[5] Soeda A, Hara A, Kunisada T, Yoshimura S, Iwama T, and Park DM, "The Evidence of Glioblastoma Heterogeneity," *Scientific Reports* , vol. 5, p. 9630, 2015.

[6] Pope WB, Sayre J, Perlina A, Villablanca JP, Mischel PS, and Cloughesy TF, "MR Imaging Correlates of Survival in Patients with High-Grade Gliomas," *American Journal of Neuroradiology,* vol. 26, no. 10, pp. 2466-2474, 2005.

[7] Gutman DA, Cooper LA, Hwang SN, Holder CA, Gao J, Aurora TD, Dunn WD Jr, Scarpace L, Mikkelsen T, Jain R, Wintermark M, Jilwan M, Raghavan P, Huang E, Clifford RJ, Mongkolwat P, Kleper V, Freymann J, Kirby J, Zinn PO, Moreno CS, Jaffe C, Colen R, Rubin DL, Saltz J, Flanders A, and Brat DJ, "MR Imaging Predictors of Molecular

Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set," *Radiology,* vol. 267, no. 2, pp. 560-569, 2013.

[8] "VASARI Research Project - Cancer Imaging Archive Wiki," https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project.

[9] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, and Lambin P, "Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach," *Nature Communications,* vol. 5, 2014.

[10] Pereira S, Pinto A, Alves V, and Silva CA, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE transactions on medical imaging,* vol. 35, pp. 1240-1251, 2016.

[11] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jordoin PM Larochelle H, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis,* p. 2016.

[12] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, and Van Leemput K, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging,* vol. 34, no. 10, pp. 1993-2024, 2015.

[13] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, and Davatzikos C, "Advancing The Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features," *Nature Scientific Data,* 2017 [In Press].

[14] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, and Davatzikos C, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection," The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.

[15] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, and Davatzikos C, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection," The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.

[16] Ahmed S, Iftekharuddin KM, and Vossough A , "Efficacy of Texture, Shape, and Intensity Feature Fusion for Posterior-Fossa Tumor Segmentation in MRI," *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, pp. 206-213, 2011.

[17] Reza S, and Iftekharuddin KM, "Multi-fractal texture features for brain tumor and edema segmentation," in *SPIE, Medical Imaging 2014: Computer-Aided Diagnosis*, San Diego, California, 2014.

[18] Iftekharuddin KM, Jia W, and Marsh R, "Fractal analysis of tumor in brain images," *Machine Vision and Applications,* vol. 13, p. 352–362, 2003.

[19] Islam A, Reza SMS, and Iftekharuddin KM, "Multifractal Texture Estimation for Detection and Segmentation of Brain Tumors," *IEEE Transactions of Biomedical Engineering,* vol. 60, no. 11, pp. 3204-3215, 2013.

[20] Ayache A, and Vehel JL, "On the Identificaiton of the Pointwise Holder Exponent of the Generalized Multifractional Brownian Motion," *Stochastic Process Application,* vol. 111, pp. 119-156, 2004.

[21] Ayache A, and Vehel JL, "Generalized Multifractional Brownian Motion: Definition and Preliminary Results," *Theory and Applications in Engineering,* pp. 17-32, 1999.

[22] Leung T, and Jitendra M, "Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons," *International Journal of Computer Vision,* vol. 43, pp. 29-44, 2001.

[23] "caret: Classification and Regression Training," https://CRAN.R-project.org/package=caret.

Symmetry-driven Fully Convolutional Network for Brain Tumor Segmentation

Haocheng Shen, Ruixuan Wang, Jianguo Zhang, and Stephen McKenna

Computing, School of Science and Engineering, Univeristy of Dundee, UK

Abstract. We present a symmetry-driven fully convolutional network for brain tumor (and sub regions) segmentation in multimodal MR images. Our structure consists of a downsampling path and three upsampling paths, which extract multi-level contextual information by concatenating hierarchical feature representatio. Meanwhile, we introduce a symmetry-driven FCN by the proposal of using symmetry difference images. The model was evaluated on BRATS17 training and validation datasets.

Keywords: tumor segmentation, FCN, symmetry-driven

1 Introduction

Precise localization of brain tumors in 3D MR images is clinically crucial to make treatment plans, guide surgery and monitor the rehabilitation progress. Since manually segmenting is time-consuming, computer-aided automatic and reliable segmentation of brain tumor is necessary. Among brain tumors, gliomas appear most frequent [1], either at high grade (HG) or low grade (LG) according to the aggressive form of the disease. Due to the diversity and variation of tumor size, shape, location, and appearance of gliomas, multimodal MRs are often taken from patients to enhance the contrast of potential tumor and its structures. Normally the tumor region could be divided into four different subregions: edema, necrosis, non-enhancing and enhancing, where the combination of the last three structures is also called *tumor core*.

The automatic segmentation of gliomas is often formulated as a patch-level or voxel-level classification problem, where each (either 2D or 3D) patch in the 3D MR is classified as one type of the sub-structures. Deep convolutional neural networks (CNNs) automatically learning high-level discriminative feature representations, are not surprisingly achieving state-of-the-art results when applied to MRI brain tumor segmentation [5–7].

Different from traditional CNN models, fully convolutional networks (FCNs) were recently proposed by removing all the fully connected layers and have achieved promising results for medical image segmentation [11–13]. In FCNs, up-sampling layers are added on top of the down-sampling convolutional layers, in order to gain the same spatial size as that of the original input. Compared to CNNs, FCNs run only once on the whole input and would generate the classification result for each voxel (or pixel), which is more computationally efficient.

We present a symmetry-driven fully convolutional network for brain tumor (and sub regions) segmentation in multimodal MR images. Due to the property of FCN, our model enables end-to-end training and fast inference. We evaluate our model on BRATS17 dataset and report the initial results.

2 Methods

2.1 FCN structure

The architecture of the proposed method is illustrated in Fig 1. It contains two modules, i.e., one downsampling path with convolutional and maxpooling layers and three upsampling paths with upsampling and convolutional layers. The downsampling path aims at enlarging receptive fields to encode high level abstract and contextual information to detect tumors, while the upsampling paths reconstruct the fine details such as tumor boundaries. We designed the upsampling paths in a hierarchical manner to take full advantage of multiple scale feature maps from downsampling path.

The downsampling path is similar to VGG-16 network [9], but instead of using total 5 convolutional blocks (one convolutional blocks contains two or three convolutional layers with 3×3 kernels and 1 maxpooling layer with 2×2 strides), we only use the first 3 convolutional blocks. For upsampling paths, we simply upsample the feature maps from the last convolutional layer of each convolutional block (before maxpooling layer) to the original spatial size. Then another three convolutional layers are applied to encode multi-scale feature representations. The resulting feature maps from three upsampling paths are concatenated before the final classification layer. We formulate the training of whole network as a per-pixel classification problem with respect to the ground-truth segmentation masks and choose categorical cross entropy as the loss function. The 2D slices split from 3D MR volumes from axial view as the input of the proposed network.

2.2 Symmetry Difference Maps

It was noted that symmetry in axial view is an important cue for brain tumor segmentation as tumors usually break symmetric appearance of a health brain. We encode brain symmetry information to the CNN framework by adding extra symmetry maps. Our symmetry maps are computed as follows: 1) we first locate the symmetric axis in T1 modality axial slices through the approach presented in [8]; 2) given the symmetric axis, we found the corresponding matching pixel pairs and calculated their intensity differences. In order to reduce the effects of the errors of symmetric axis and image noises, each image was smoothed beforehand using a Gaussian filter with 5×5 kernel. The most matched pixel was searched in a 11×11 local window centered on the mirrored the location w.r.t the symmetry axis. The resulting intensity differences are then converted into range $[0, 1]$ by a sigmoid function. We calculated symmetry difference maps for each MR modality and combined them with the four original images as the inputs of our CNN framework as show in Fig. 1.

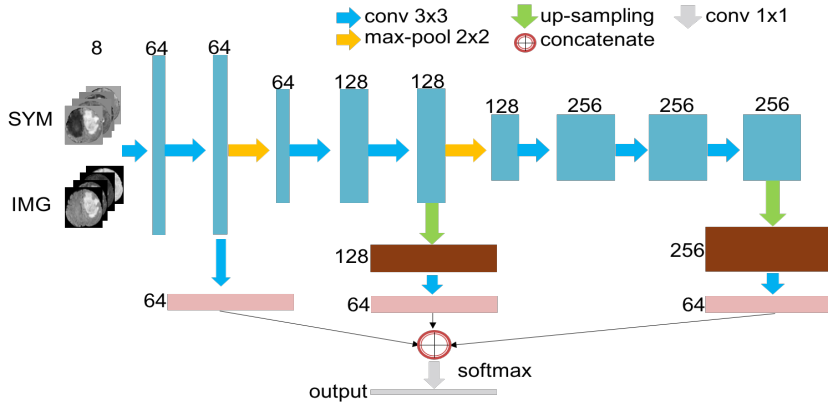


Fig. 1: FCN structure. Images and symmetry maps are concatenated as the input to the net [10]. Colored rectangles represent feature maps with numbers nearby being the number of feature maps. Best viewed in color.

3 Evaluation

Our model was evaluated on BRATS17 dataset [1–4] which contains 210 HGG and 75 LGG training data with known ground-truth segmentation maps and 46 validation data without revealing tumor grade information. Each subject's data includes 4 modalities (T1, T1-contrast or T1c, T2, and Flair) which were skull-stripped and co-registered. Quantitative evaluation is performed on three sub-tasks: 1) the *complete* tumor; 2) the tumor *core*; 3) the *enhancing* tumor region. For each sub-task, *Dice*, *Sensitivity*, *Specificity* and *Hausdorff Distance* are computed.

Our network model was implemented in Keras with Theano as backend. The network was trained using the Adam optimizer, with learning rate 0.001. The down-sampling path was initialized with VGG-16 weights [9] while up-sampling paths were initialized randomly using He's method [14].

We randomly split 210 HG and 75 LG in BRATS17 training set into two subsets at a ratio of 7:3, resulting in 201 training data (147 HG and 54 LG) and 84 test data (63 HG and 21 LG). The performance curves on 84 test data along epochs are shown in Fig 2. It is observed that all sub-tasks performances were saturated at about the 15th epoch, resulting in 0.87, 0.76, 0.69 in terms of *Dice* for *Complete*, *Core* and *Enhancing* tasks, respectively.

For the 46 validation data, we used the 15th epoch model trained by the 201 training data. The results are shown in Table 1. From both training and validation results, we found segmenting *Enhancing* is the most difficult task as LG cases usually contains small or even no regions of enhancing parts. Misclassifying the enhancing parts in LG cases may cause a significantly drop in terms of *Dice* performance.

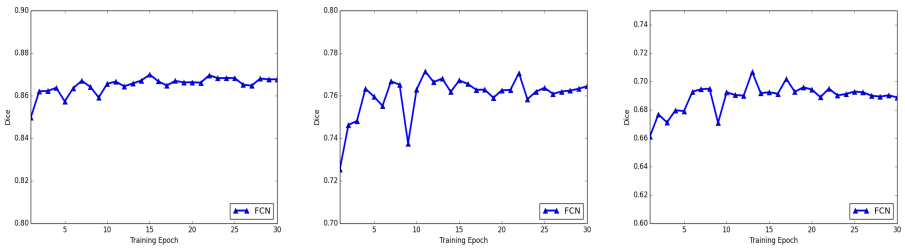


Fig. 2: Training-validation results on BRATS17. From left to right: *Complete*, *Core* and *Enhancing* tumor task. The vertical axis is Dice while horizontal axis is the number of epochs.

Table 1: Performance on the BRATS17 46 validation set

	Complete	Core	Enhancing
Dice	0.885 ± 0.082	0.752 ± 0.224	0.695 ± 0.293
Sensitivity	0.874 ± 0.115	0.750 ± 0.251	0.784 ± 0.257
Specificity	0.995 ± 0.004	0.997 ± 0.004	0.998 ± 0.004
Hausdorff	7.347 ± 17.632	11.569 ± 17.853	6.987 ± 11.090

4 Conclusion

We propose a symmetry-driven FCN and achieved reasonable segmentation results on BRATS17 dataset. We are still working on multiple potential extensions in order to improve our performance including data augmentations, multi-task FCN and classification on HG and LG cases.

References

1. Menze, B.H., Jakab, A., Bauer, S., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). Medical Imaging, 34(10), 1993–2024 (2015)
2. Bakas, S., Akbari, H., Sotiras, A., et al.: Advancing The Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features. Nature Scientific Data, (2017) [In Press]
3. Bakas, S., Akbari, H., Sotiras, A., et al.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
4. Bakas, S., Akbari, H., Sotiras, A., et al.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
5. Pereira, S., Pinto, A., Alves, V., et al.: Brain Tumor Segmentation Using Convolutional Neural Networks in MRI images. Medical Imaging, 35(5), 1240–1251 (2016)

6. Havaei, M., Davy, A., Warde-Farley, D., et al.: Brain Tumor Segmentation with Deep Neural Networks. Medical Image Analysis, 35, 18-31 (2017)
7. Kamnitsas, K., Ledig, C., Newcombe, V.F., et al.: Efficient Multi-scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. Medical Image Analysis, 36, 61-78 (2017)
8. Loy, G. and Eklundh, J.: Detecting Symmetry and Symmetric Constellations of Features. In: ECCV, pp. 508-521 (2016)
9. Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv preprint arXiv:1409.1556 (2014)
10. Shen, H., Zhang, J., Zheng, W.: Efficient Symmetry-driven Fully Convolutional Network for Multimodal Brain Tumor Segmentation. In: ICIP, to appear (2017)
11. Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional Networks for Biomedical Image Segmentation. In: MICCAI, pp. 234-241 (2015)
12. Chen, H., Qi, X.J., Cheng, J.Z. and Heng, P.A.: Deep Contextual Networks for Neuronal Structure Segmentation. In: AAAI (2016)
13. Chen, H., Qi, X., Yu, L. and Heng, P.A.: DCAN: Deep Contour-aware Networks for Accurate Gland Segmentation. In: CVPR, pp. 2487-2496 (2016)
14. He, K., Zhang, X., Ren, S. and Sun, J.: Delving Deep into Rectifiers: Surpassing Human-level Performance on Imagenet Classification. In: ICCV, pp. 1026-1034 (2015)

MRI Brain Tumor Segmentation using Random Forests and Fully Convolutional Networks

Mohammadreza Soltaninejad[1], Lei Zhang [1], Tryphon Lambrou[1], Guang Yang[2], Nigel Allinson[1], and Xujiong Ye [1]

[1]Laboratory of Vision Engineering, School of Computer Science, University of Lincoln, UK
[2]National Heart & Lung Institute, Imperial College London, UK
{msoltaninejad,lzhang,tlambrou,nallinson,xye}@lincoln.ac.uk,
g.yang@imperial.ac.uk

Abstract. In this paper, we propose a novel learning based method for automated segmentation of brain tumor in multimodal MRI images, which incorporates two sets of machine -learned and hand crafted features. Fully convolutional networks (FCN) forms the machine learned features and texton based features are considered as hand-crafted features. Random forest (RF) is used to classify the MRI image voxels into normal brain tissues and different parts of tumors, i.e. edema, necrosis and enhancing tumor. The method was evaluated on BRATS 2017 challenge dataset. The results show that the proposed method provides promising segmentations. The mean Dice overlap measure for automatic brain tumor segmentation against ground truth is 0.86, 0.78 and 0.66 for whole tumor, core and enhancing tumor, respectively.

Keywords: Fully Convolutional Networks, Random Forest, Deep Learning, Texton, MRI, Brain Tumor Segmentation.

1 Introduction

Accurate segmentation of brain tumor may aid the fast and objective measurement of tumor volume and also find patient-specific features that aid diagnosis and treatment planning [1]. Due to the recent advances in deep neural networks (DNN) in recognition of the patterns in the images, most of the recent tumor segmentations have focused on deep learning methods. Recently, fully convolutional networks (FCN) have been suggested for dense (i.e. per-pixel) classification with the advantage of end-to-end learning [2], without requiring those additional blocks in convolutional neural networks (CNN) based approaches. Despite the advantage of dense pixel classification, FCN-based methods still have limitations of considering the local dependencies in higher resolution (pixel) level. The loss of spatial information, which occurs in the pooling layers, results in coarse segmentation. This limitation will be addressed in our work by incorporating high resolution hand-crafted textural features which consider local dependencies of the pixel. Texton feature maps [3] provides significant information on multi-resolution image patterns in both spatial and frequency domains.

In this paper, a novel fully automatic learning based segmentation method is proposed, by applying hand-crafted and machine-learned features to random forest (RF) classifier. The machine-learned FCN based features detect the coarse region of the tumor while the hand-crafted texton descriptors consider the spatial features and local dependencies to improve the segmentation accuracy.

2 Methods

Our method is comprised of four major steps (pre-processing, FCN, Texton map generation, and RF classification) that are depicted in Fig. 1. In the pre-processing stage, the intensities were normalized for each protocol by subtracting the average of intensities of the image and dividing by their standard deviation. Then, the histogram of each image was normalized and matched to the one of the patient images which is selected as the reference.

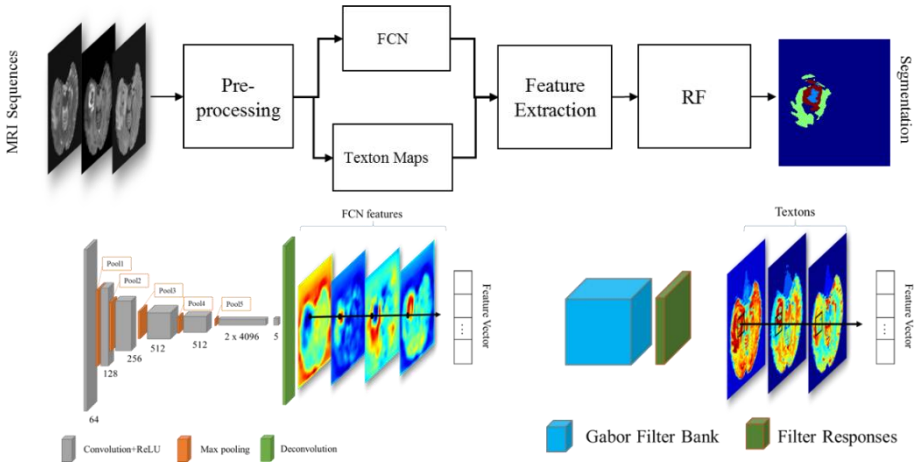


Fig. 1. Flowchart of the proposed method. The FCN architecture, machine learned feature extraction, and texton features.

2.1 FCN-based Features

A FCN-8s architecture in [2] was adopted for segmentation of brain tumor in multimodal MRI images, where the VGG16 [4] was employed as CNN classification net. The FCN-8s was constructed from FCN-16s skip net and FCN-32s coarse net, which was implemented by fusing predictions of shallower layer (Pool3) with 2 × upsampling of the sum of two predictions derived from pool4 and last layer. Then the stride 8 predictions were upsampled back to the image.

The feature vector is generated for each voxel based on the score map from the FCN. For each class label, a score map is generated, 4 maps are generated using the standard BRATS17 labelling system. The values of each map layer corresponding to each voxel are considered as machine-learned features of that voxel.

2.2 Texton features

The texton based features were applied in the proposed method as hand-crafted features to support the machine-learned features and improve the segmentation results. Textons are obtained by convolving the image with a Gabor filter bank. To cover all orientations six different filter directions were used: [0^o, 30^o, 45^o, 60^o, 90^o, 120^o]. Filter sizes were [0.3, 0.6, 0.9, 1.2, 1.5] and the wavelength of sinusoid coefficients of the Gabor filters were [0.8, 1.0, 1.2, 1.5].

The MR images were convolved with the Gabor filters, then the filter responses are merged together and clustered using *k-means* clustering. The number of clusters 16 was selected as the optimum value for the number of clusters in texton map. The texton map is created by assigning the cluster number to each voxel of the image. The texton feature for each voxel is the histogram of textons in a neighborhood window of 5×5 around that voxel.

The normalized intensity value of the voxels in each modality which is obtained from the pre-processing stage is also included in the feature vector. Therefore, in total 55 features were collected (4 FCN score maps, 3 protocol intensities and 48 texton histograms) for usage in the next step.

2.3 RF Classification

The potential tumor area detected by the FCN output was considered as the initial region of interest (ROI). This ROI was selected as a confidence margin of 10 voxels in 3D space around the detected initial tumor area which was calculated by morphological dilation. The feature vectors for voxels in this ROI were fed to the random forests. The main parameters in designing RF are tree depth and the number of trees. RF parameters were tuned by examining different tree depths and number of trees on training datasets and evaluating the classification accuracy using 5-fold cross validation. The number of 50 trees with depth 15 provided an optimum generalization and accuracy. Based on the classes assigned for each voxel in the validation dataset, the final segmentation mask was created by mapping back the voxel estimated class to the segmentation mask volume. Finally, the bright regions in the healthy part of the brain near to the skull were eliminated using a connected component analysis.

3 Results

The proposed method was performed on MATLAB 2016b on a PC with CPU Intel Core i7 and RAM 16 GB with the operating system windows 8.1. The FCN was implemented using MatCovNet toolbox [5]. GPU GeForce gtx980i was used to run FCN. The RF was implemented using open source code provided in [6] which is a specialized toolbox for RF classification based on MATLAB.

Both FCN and RF are trained on BRATS 2017 [7–10] training dataset which include 220 high grade glioma (HGG) and 75 low grade glioma (LGG) patient cases. The method was evaluated on BRTAS 2017 validation dataset which include 46 patient cases.

The evaluation measure which are provided by the CBICA's Image Processing Portal, i.e. Dice score, sensitivity, specificity, Hausdorff distance, were used to compare the segmentation results with the gold standard (blind testing). Figure 2 shows segmentation results of the proposed method on some cases of BRATS 2017 validation dataset. Table 1 provides the evaluation results obtained by applying the proposed method on BRATS 2017 validation dataset.

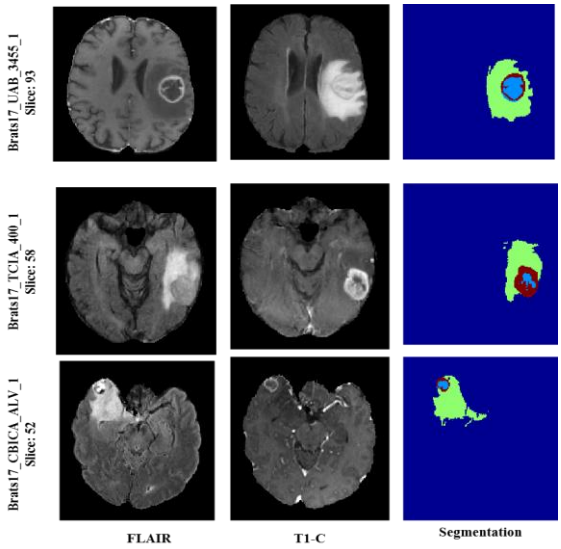


Fig. 2. segmentation masks for validation dates, using the proposed method. Light blue: necrosis and on-enhancing, green: edema, dark red: enhancing.

Table 1. Segmentation results for validation dataset which was provided by CBICA portal. ET: enhancing tumor, WT: whole tumor, TC: tumor core.

	Dice			Sensitivity			Specificity			Hausdorff (95%)		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Mean	0.66	0.86	0.78	0.57	0.83	0.72	1.00	1.00	1.00	3.76	7.61	8.70
STD	0.28	0.09	0.19	0.28	0.13	0.21	0.00	0.01	0.00	4.38	12.99	13.52

4 Conclusion

In this paper, a novel method was proposed in which the machine-learned features extracted using FCN were combined with hand-crafted texton features to encode global information and local dependencies into feature representation. The score map with pixel-wise predictions was used as a feature map which was learned from multi-modal BRATS2017 training dataset using the FCN. The machine-learned features, along with hand-crafted texton features were then applied to random forests to classify each MRI image voxel.

The results of the FCN based method showed that the application of the RF classifier to multimodal MRI images using machine-learned features based on FCN and hand-crafted features based on textons provides promising segmentations. The mean Dice overlap measure for automatic brain tumor segmentation against ground truth is 0.86, 078 and 0.66 for the whole tumor, core and enhancing tumor, respectively.

References

1. Gordillo, N., Montseny, E., Sobrevilla, P.: State of the art survey on MRI brain tumor segmentation. Magn Reson Imaging. 31, 1426–1438 (2013)
2. Long, J., Shelhamer, E., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
3. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour Detection and Hierarchical Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 33, 898–916 (2011)
4. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs]. (2014)
5. Vedaldi, A., Lenc, K.: MatConvNet: Convolutional Neural Networks for MATLAB. In: Proceedings of the 23rd ACM International Conference on Multimedia. pp. 689–692. ACM, New York, NY, USA (2015)
6. Taormina, R.: MATLAB_ExtraTrees - File Exchange - MATLAB Central, http://uk.mathworks.com/matlabcentral/fileexchange/47372-rtaormina-matlab-extratrees.
7. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Leemput, K.V.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging. 34, 1993–2024 (2015)
8. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Rozycki, M., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data, (2017) [In Press]
9. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017), DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
10. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, (2017), DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

An Expert Automated Preliminary Diagnostic System For Identifying Brain Tumors' Features and Types

Berkan Ural[1], Mustafa Yüksek[2], Furkan Muratdağı[3]

[1]PhD Candidate, Gazi University, Department of Electrical Electronics Engineering, Ankara/TURKEY, uralberkan@gmail.com/aliberkanural@gazi.edu.tr
[2]Associate Professor, Kafkas University, Department of Electrical Electronics Engineering, Kars/TURKEY, mustafa_yuksek2001@yahoo.com
[3]BSc Candidate, Gazi University, Department of Mechanical Engineering, Ankara/TURKEY, furkanmuratdagi@outlook.com

Abstract. This paper presents an expert automated system for detecting and preliminary diagnosing the human brain tumors using T1-weighted Magnetic Resonance Images with contrast. The system uses computer based procedures to detect tumor areas and to identify the type of tumor as benign or malignant with using multi-class classification. Generally, this system is consisted from three main distinct modules: detecting brain tumors with image processing and pattern recognition methods, feature extraction according to the important medical parameters, multi-class hybrid classification. According to the first part, an approximate reasoning is used to enhance the quality of MRI scans and segment them into special ROI (Region of Interest) areas. In the second part, the extraction of texture features of specialized ROIs has been achieved by using Gray Level Cooccurence Matrix (GLCM). Then, these features are compared according to the stored features in the Training Base. Finally, preliminary diagnosing and differentiating brain tumors and identifying the tumor types according to the hybrid classification are performed. However, brain tumors have different characteristics in different slice planes. In order to eliminate the inaccurate results, several consecutive planes are used for processing. Finally, the performance of this system is obtained as %95 with using 100 images from BraTS'17 dataset (50 images are used for training and 50 images are used for testing). All results show that the system works efficiently in tumor detection and classification of the samples.
Keywords: Brain Tumor Detection, Medical Image Processing, Magnetic Resonance Imaging (MRI), Multi-Class Classification

1 Introduction

Brain tumors among humans have gained importance for decades and generally differ according to the sex, age, race and nationality. Like other tumor types, some of malignant brain tumors such as Glioblastomas might develop suddenly from lower grades [1]. Therefore, preliminary diagnosing of the brain tumors in an appropriate time is very crucial for further treatments [2].

In recent years, the usage of the imaging tools in neuroscience department has been significantly increased. Especially, Magnetic Resonance Imaging (MRI) has been used

as a powerful brain imaging model that allows to take non-invasive measurements and 3D assessment of the tissue morphology [3]. MRI system mainly provides an unparalleled inside view of the human body. Also, the reputation of this imaging model has greatly increased and the knowledge of the normal and abnormal anatomy for important medical research are become a useful critical component in diagnosing and treatment areas.

Reliability and time-saving are the important parameters systems for fast detection and classification of brain tumors for the doctors. Common practices that are based on specialized technicians are become slow and possess a degree of subjectivity that is hard for quantifying [4].

The designed system is consisted from a successful hybrid algorithm for detecting and identifying the brain tumor from a given MRI image of patients. This system also finds sufficient usage under tumor detection in the medical area such as computer aided preliminary diagnosis. Diagnosing of brain tumors is a complex issue and brain cancer can be classified into 120 different types. According to this, benign brain tumors can be explained as life-threatening as malignant tumors, as they squeeze out normal brain tissue and disrupt some of the brain functions [5].

The purpose of this paper is to design, implement, interpret and evaluate the system which is given as the expert automated system to preliminary diagnose the brain tumors and to identify the tumor types, especially in T1-weighted images with contrast. However there are many uncertainities in images which are very difficult to handle, the offered system can be able to model such uncertainities, so this system have the sufficient potential to provide better performance. According to the first part of the system, an approximate reasoning is used to enhance the quality of MRI scans and segment them into special ROI (Region of Interest) areas. In the second part, the extraction of texture features of specialized ROIs has been achieved by using Gray Level Co-occurence Matrix (GLCM). Then, these features are compared according to the stored features in the Training Base. Finally, preliminary diagnosing and differentiating brain tumors and identifying the tumor types according to the hybrid classification are performed.

The rest of the paper is organized as follows: Section 2 presents the methods of the expert automated preliminary diagnosis system for identifying brain tumors' features and types. Image processing and pattern recognition stages, feature extraction and image classification techniques are discussed in Section 2. The results obtained from evaluation and conclusions are given in Section 3 and 4, respectively.

2 Methods

This work generally involves processing of raw T1-weighted MR images that are affected by brain cancer for humans for tumor detection, identification and classification of different types of brain tumors. In addition, images used for this study are generally MRI images and they are obtained from BraTS'17 image dataset [6, 7, 8 ,9]. The whole system is also designed with a Graphical User Interface (GUI) with using MATLAB 2016b version according to the Nielson's usability criteria [10]. Also, in the GUI development phase, Furkan Muratdağı helped Berkan URAL for increasing the success and the stability rates of the system.

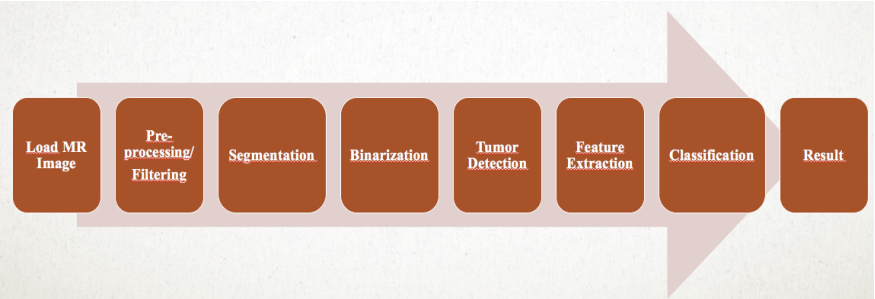


Fig. 1. Flowchart of the System

According to Fig. 1, all steps are given below, respectively.

Pathological Detection

All slices of MR images processed by the given system are classified as abnormal and also, this is known that these images contain a tumor or tumors based on the radiologist pathology report.

2.1 Load MR Image

For this step of the approach, images taken from the test subjects are generally in gray-level color format. Image size for each patient is approximately 255 um x 255 um and "uint8" is the class type of the dataset.

2.1 Pre-processing/Filtering

For this step, this is seen that most of the images have different noise types. Especially, Gaussian noise is the most common noise type among the images and this noise type is eliminated with using 5x5 Median filter with a standard deviation [11]. Then, this step consists the histogram equalization part. The main problem in the process for detecting the edge of the tumor area is that the tumor area is generally seen very dark on the image. To overcome this issue, histogram equalization is mainly performed. The aim of the fundamental enhancement is to increase the contrast between the brain and the tumor region. Finally, to enhance contrast between the normal brain and tumor area, a special sharpening filter is used on the MR images.

2.2 Segmentation

Segmentation is an essential issue for image classification. Currently, in many clinical works, segmentation is done manually or strongly supervised by a human expert. The knowledge level of operator affects the performance of the segmentation method. In addition, manual segmentation makes the original segmentation step more deteriorating. Hence, there is a great need for an expert automated brain tumor detection tool. In the segmentation phase, automated techniques are mainly used. These are; thresholding, edge detection, clustering and region extraction.

The aim of this step is to subdivide the images into their constituent object parts. The level of the subdivision depends on the rules and criteria in image processing and the segmentation has to stop when the edge of tumor is able to be detected and also isolating the tumor from the background is the other important problem. Otsu thresholding is used as segmentation part in order to obtain a thresholded image with gray

level 1 or 0. “1” represents the tumor area and “0” represents the background, respectively. Therefore, the thresholding index is obtained by using (1) [12];

$$T = \frac{\sum_{I=0}^{M-1} \sum_{J=0}^{N-1} e_{ij} * M_{ij}}{\sum_{I=0}^{M-1} \sum_{J=0}^{N-1} M_{ij}} \quad (1)$$

where M is the MRI image and Mij represents the image’s pixels.

2.4 Binarization

The main purpose for this step is to differentiate the contrast value between normal and tumor regions in the images. In this step, median filtered images are filtered through a standard 15 pixel Gaussian filter, so some important details are exposed.

2.5 Tumor Detection

In this phase, some morphological operations such as erosion, dilation and closing methods are used for filling the broken gaps at the edges and to have continuities at the boundaries [13]. After the dilation operation, a filling operator is applied to fill the close contours. After this step, the centroids of special Region Of Interest (ROI) areas are calculated to localize the regions. The final region after extraction is then logically operated for extraction of the main region in the MR image.

2.6 Feature Extraction

This stage involves obtaining important features of the extracted region in the MR images. The features can mainly be used as giving the property of the texture and the information are stored in the knowledge base for training of the system. Then, the extracted features are compared with the features of the other out-of-sample images for the classification step. For this occasion, Gray Level Co-occurence Matrix (GLCM) features are mainly used to separate and distinguish the result between the normal and the tumor area in the brain [14]. Generally, three GLCM matrices are used for this implementation according to the axial, sagittal and coronal slice diagonals (0, 90, 135 degrees).

2.7 Classification

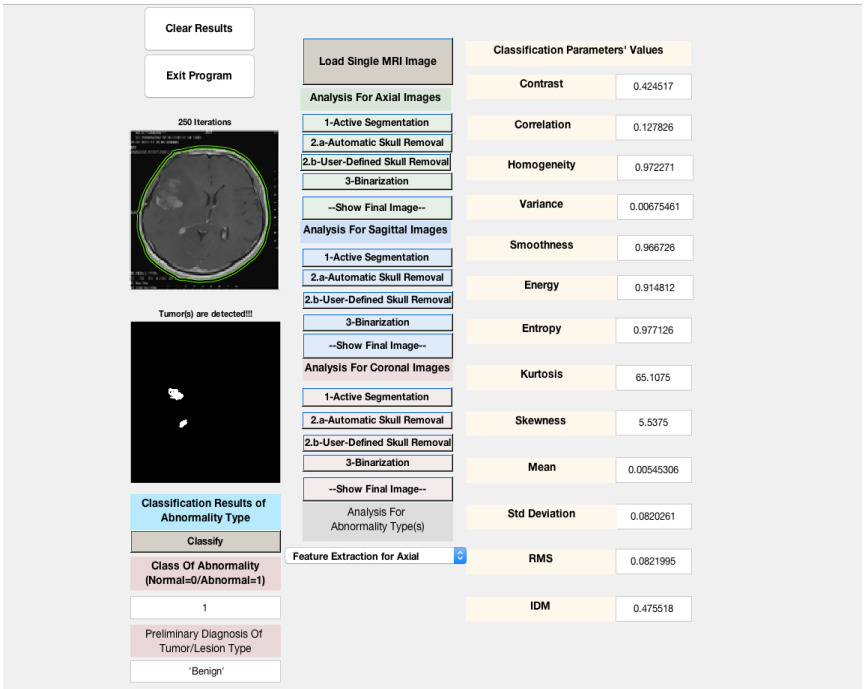
A hybrid classification method which is included the common k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) techniques is mainly used to detect candidate circumscribed tumor for this step and also to identify the final class of the processed data [15]. Obtained hybrid classification method is an adaptive system that learns to perform a function from input data. Adaptive means that the system parameters can be changed during the processes. In the training section, different MR images are used and the features and the tumor types are stored in the knowledge base. After the training phase, the parameters of the classification are fixed and the system performance is highly increased with using the deploy tool in MATLAB. This system can be similar in many ways with Artificial Neural Network during training phase. The networks of training phase consist of some mapping values and these values are generally between a set of input and output values. With using a learning algorithm or changing the weight values, the training phase can be controlled simultaneously.

3 Results and Discussion

The developed interface successfully classifies an input MR image into the type of the tumor such as benign or malignant. 50 sample images are used in the testing phase,

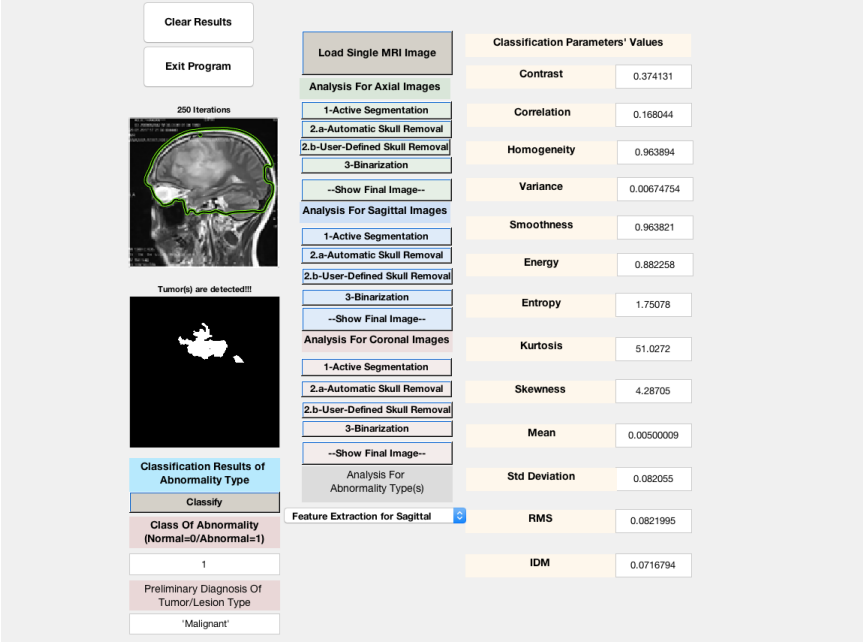
respectively. Then, the test images are compared with stored features from the training phase. According to the statistical analysis, %95 success rate is obtained from ROC curve analysis of the whole system.

Figure 2 shows the main objectives of the interface which are mentioned in the study. Also, in Fig. 2 and Fig. 3, a tumor region extraction process is specifically shown as step by step and also a classification example according to a sagittal image is given in Fig. 3 in detail.



(a)

6



(b)

Fig. 2. Results of Extraction and Classification For Axial (a) and Sagittal (b) Images

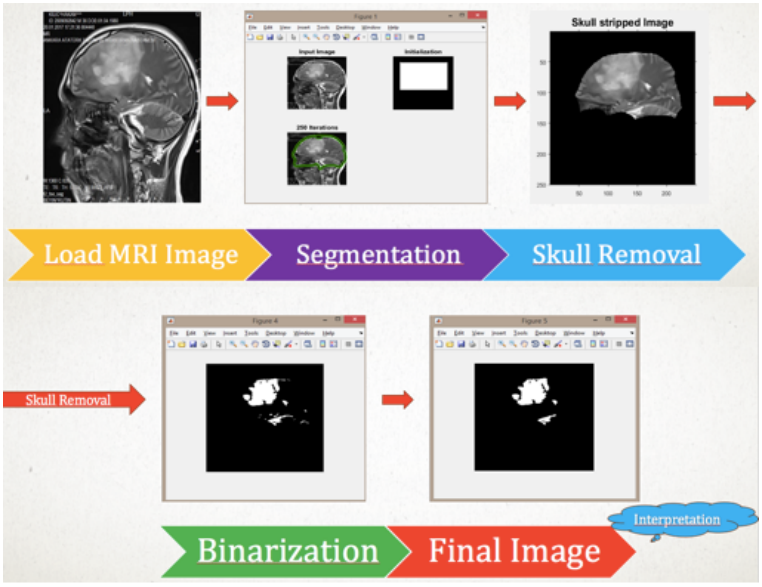


Fig. 3. Tumor Extraction Procedure for Sagittal Image

The whole results given above show that the system works efficiently with %95 success rate for tumor detection and identification stages. The truth of the obtained final

results is also supported by the brain tumor specialists from Ataturk Education and Research Hospital, from Ankara/TURKEY. The quality rate of the system can motivate us to extend this work to develop the system as a whole body tumor diagnosis scan system.

4 Conclusion

In this research, an expert automated preliminary diagnosis system for identifying brain tumors' features and types has been developed. The proposed system can help the doctors for diagnosing the human brain tumors, for further treatments. The main contributions in this paper are the several important image processing and pattern recognition methods, a specific segmentation and a special classification methods. In addition, the presented system has been encoded and the main modules of the system have been tested and validated in order to increase the diagnosing rate positively.

This research has some potential future developments. The preliminary diagnostic results can demonstrate the high classification accuracy for the image processing and classification structures. Moreover, the preliminary results also motivate us to extend this framework to detect and localize the tumors in the other organs. When this work is adapted to the different fields, this can be given a chance to early diagnose the different types of diseases. Also, if the methods of this study are improved, the tissue recognition can be done successfully. This can help doctors for diagnosing important diseases on time.

References

1. Fisher, J.L., Schwartzbaum, J.A., Wrensch, M., Wiemels, J.L.: Epidemiology of brain tumors. Neurologic Clinics 25, 867–890 (2007).

2. Lau, P.Y., Ozawa, S.: A Simple Method for Detecting Tumor in T2-Weighted MRI Brain Images: An Image-Based Analysis. Department of Information and Computer Science, Keio University, Yokohama-shi, 223-8522, Japan (2006).

3. Lefohn, A., Cates, J., Whitaker, R.: Interactive GPU-Based level sets for 3D Brain Tumor Segmentation (2003).

4. Cline, H.E., Lorensen, E., Kikinis, R., Jolesz.: Three-dimensional segmentation of MR images of the head using probability and connectivity. F. J Computer Assist Tomography, 14:1037-1045 (1990).

5. Dong, A., Wang, B.: Feature selection and analysis on mammogram classification, in Proceedings of the IEEE Pacific Rim Conference on Communications. Computers and Signal Processing (PACRIM '09), 731735, Victoria, BC, Canada (2009).

6. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A.,

Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).
7. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data [In Press] (2017).
8. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017).
9. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani K, Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017).
10. MacDonald, L.W.: Tutorial: Using Color Effectively in Computer Graphics. IEEE Computer Graphics and Applications, 19(4), 20-35 (1999).
11. Chung, L.A.W., Yan, H.: Current methods in the automatic tissue segmentation of 3D magnetic resonance brain images. Current Medical Imaging Reviews 2, 2(1), 91-103 (2006).
12. Clark, M.C., Hall, L.O.: Automatic tumor segmentation using Knowledge based techniques. IEEE transaction on medical imaging, 17(2) (1998).
13. Sonka, Hlavac, M., Boyle, V.R.: Image processing, Analysis, and Machine Vision, II Edition, Vikas Publishing House, New Delhi (2004).
14. Descombes, X., Kruggel, F., Wollny, G., Gertz, H.J.: An objectbased approach for detecting small brain lesions: Application to Virchow-robin spaces. IEEE Trans Med. Imaging, 23(2), 246-255 (2004).
15. Chaplot, S., Patnaik, L.M., Jagannathan, N.R.: Classification of magnetic resonance brain imagesusing wavelets as input to support vector machineand neural network. Biomedical Signal Processing and Control, 1(1), 86–92 (2006).

Automatic Brain Tumor segmentation Using 2.5D U-nets

Chunliang Wang and Örjan Smedby

School for Technology and Health (STH), KTH Royal institute of technology,
Hälsovägen 11C, SE-14152 Huddinge, Stockholm, Sweden
{chunliang.wang,orjan.smedby}@sth.kth.se
http://www.kth.se/sth

Abstract. In this report, we proposed an automatic brain tumor segmentation method using a deep learning framework based on the fully convolutional network. The proposed method carries out the segmentation using three 2D U-nets that are trained to segment brain tumors in the three orthogonal views, i.e., the axial, coronal and sagittial views. The final segmentation labels are created via a majority voting of these three U-nets. Common preprocessing steps including inhomogeneity correction and histogram alignment are applied on both training and testing images. The proposed method was tested on the BRATS 2017 brain tumor segmentation database. Two sets of U-nets were trained to segmentation glioblastoma and lower grade glioma separately. The Dice coefficients for the edema region, whole tumor and tumor core segmentation are 0.69, 0.85, 0.74 respectively.

Keywords: deep learning, fully convolutional network, brain tumor segmentation, U-net

1 Introduction

Brain tumor segmentation is one of the crucial steps for surgery planning and treatment evaluation. Despite the great amount of effort being put to address this challenging problem in the past two decades, segmentation of brain tumor remains to be one of the most challenging tasks in medical image analysis [1]. This is due to both the intrinsic nature of the tumor tissue being heterogeneous and the extrinsic problems with unsatisfactory image quality of clinical MRI scans. For example, the tumor mass of a glioma patient, the most common brain tumor, often consists of peritumoral edema, necrotic core, enhancing and non-enhancing tumor core. In addition to the complicated tumor tissue pattern, the MRI images can be further corrupted with a slowly varying bias field and/or motion artifacts, etc. In this study, we propose a hybrid method that consists of a pre-processing phase, where bias file correction and histogram alignment are performed and a segmentation phase using a deep learning framework. In our preliminary experiments, the proposed method delivered promising results on the public database of BRATS 2017 [2–4].

2 Methods

2.1 Data and Processing

In this study, we used the BRATS 2017 database, that consists of about 460 clinically-acquired pre-operative multimodal MRI scans of glioblastoma (GBM/HGG) and lower grade glioma (LGG). Each of the scans contains four MRI series, namely T1-weighted (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. The images were acquired from 19 different institutions. Manual segmentation of three types of tumor tissue, i.e. GD-enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor (NCR/NET) were provided for each case. In addition to the skull-striping and registration of four series, we added more pre-processing steps to fix the inhomogeneity of the MRI images and normalized the tissue intensity. For inhomogeneity correction, we used the N4ITK package proposed in [5]. Then the histogram of each volume is matched to a manually picked patient with HGG. This is done using Laszlo's approach implemented in ITK [6].

2.2 2.5D U-net

To perform the tumor segmentation, we adapted the U-net method proposed by Ronneberger et al. [7]. This is a variation of the fully convolutional network (FCN), the fully connected layers of classical CNNs are replaced by convolutional layers [8], which allows FCNs to be applied to images of any size and output label maps proportional to the input image and avoid the time-consuming sliding window operation. The original U-net was proposed to segment 2D images, in order to segment tumor in 3D, we trained three U-nets that will segment different tumor tissue types in 2D slices acquired in three orthogonal projects, i.e., in the axial, coronal and sagittal views. The final label map is generated by a majority voting of these three U-nets.

2.3 Implementation details

Our U-net implementation was based on the Keras framework with Theano backend (http://keras.io). The U-net architecture is identical to the one proposed in the original paper, with a little modification to allow the input image to have 4 channels. In addition to the histogram alignment, each channel is normalized group-wise to have 0 mean and standard deviation of 1. As HGG and LGG lesions are very different from each other, two sets of U-nets were trained to segment different types of tumor. In addition, a support vector machine (SVM) is trained to classify whether a case is HGG or LGG using the histogram of tumor core region from the HGG U-net sets, i.e., each case is segmented as HGG then classified with the SVM to see if the LGG U-net set need to be run.

For training all the U-nets, the categorical cross-entropy is used as the loss function for multi-structure segmentation. Stochastic gradient descent (SGD) is used as the optimizer in all training process.

Table 1. Segmentation accuracy evaluation on the validation data

Tumor type	Edema	Total tumor	Tumor core	Cases
HGG	0.820±0.105	0.879±0.087	0.819±0.155	36
LGG	0.229±0.411	0.749±0.282	0.442±0.391	10
total	0.695±0.322	0.851±0.156	0.742±0.272	46

3 Results

The proposed method was evaluated using the training and validation data of the BRATS 2017 database [2–4]. The HGG set was trained on 210 cases and the LGG set was trained on 75 cases. The validation set consists of 46 cases, among which 10 cases were classified by the SVM as LGG (the true labels are unknown).

Overall, the Dice coefficients for the edema region, whole tumor and tumor core segmentation are 0.69, 0.85, 0.74 respectively. However, the Dice scores are much better for the HGG cases than the LGG cases. Table 1 summarized the results on the validation data.

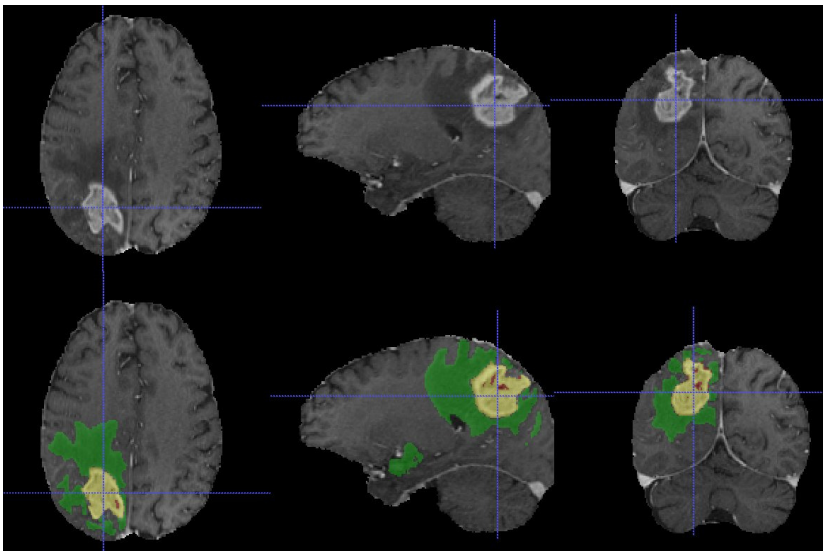


Fig. 1. An example of relatively well segmented cases. In this case, the Dice coefficients for the edema region, whole tumor and tumor core segmentation are 0.945, 0.877, 0.937 respectively

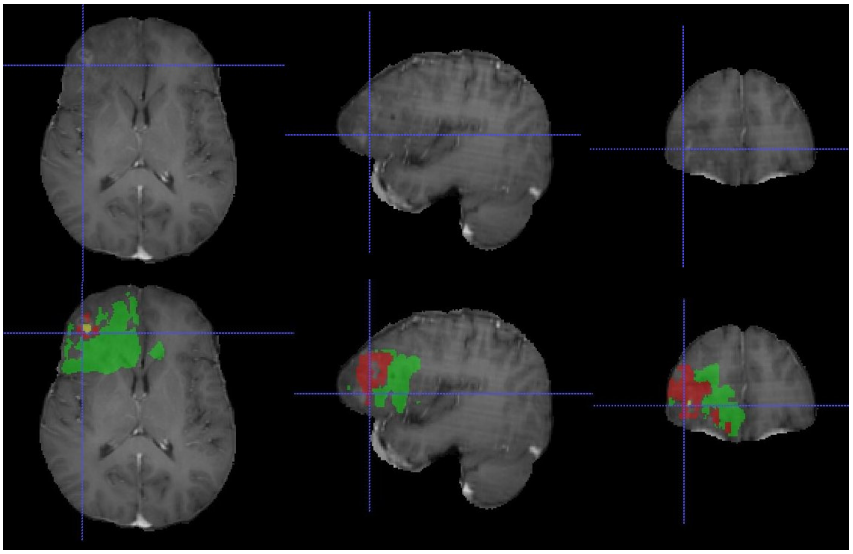


Fig. 2. An example of not well segmented cases. In this case, the Dice coefficients for the edema region, whole tumor and tumor core segmentation are 0.201,0.585, 0.193 respectively

4 Discussion and conclusion

In this study, we proposed a relative simple 2.5D U-net approach for tumor segmentation. While implementing a 3D U-net could potentially improve the segmentation accuracy, the 3D U-net approach faces some piratical challenges, such limited GPU memory that limits the depth of U-net and a limited number of training samples that can cause overfitting. The 2D U-nets are much easier to train and use larger number training samples (by converting 3D volumes to 2D slices). However the direct comparison with 3D U-net is not available as the leadboard is anonymous at the moment. From our experience, the LGG cases are more difficult to segment. This may be due to both the inconsistent appearance and the limited number of training data. Some future work include to add dropout layer to reduce the overfitting of the U-nets. Auto-context as reported in other studies may also help to further improve the segmentation accuracy.

In conclusion, we have proposed a hybrid image segmentation methods based on deep neural network. In our preliminary experiments, the proposed method delivered promising results on the public database of BRATS 2017

Acknowledgments. This research has been partially funded by the Swedish Hjärnfonden, grant no. FO2016-0175, and the Swedish Barncancerfonden, grant no. MT2016-0016.

References

1. Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2015.
2. S. Bakas, H Akbari, A. Sotiras, M. Bilello, M Rozycki, JS Kirby, JB Freymann, K Farahani, and C. Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data,*, page in press, 2017.
3. S. Bakas, H Akbari, A. Sotiras, M. Bilello, M Rozycki, JS Kirby, JB Freymann, K Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *Cancer Imaging Archive, DOI:10.7937/K9/TCIA.2017.KLXWJJ1Q*, 2017.
4. S. Bakas, H Akbari, A. Sotiras, M. Bilello, M Rozycki, JS Kirby, JB Freymann, K Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *Cancer Imaging Archive, DOI:10.7937/K9/TCIA.2017.GJQ7R0EF*, 2017.
5. Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320, 2010.
6. László G Nyúl, Jayaram K Udupa, and Xuan Zhang. New variants of a method of mri scale standardization. *IEEE transactions on medical imaging*, 19(2):143–150, 2000.
7. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
8. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3431–3440, 2015-06.

Automatic Brain Tumor Segmentation using Cascaded Anisotropic Convolutional Neural Networks

Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren

Translational Imaging Group, CMIC, University College London, UK
Wellcome/EPSRC Centre for Interventional and Surgical Sciences, UCL, London, UK
guotai.wang.14@ucl.ac.uk

Abstract. A cascade of fully convolutional neural networks is proposed to segment multi-modality MR images with brain tumor into background and three subregions: enhanced tumor core, whole tumor and tumor core. The cascade is designed to decompose the multi-class segmentation into a sequence of three binary segmentations according to the subregion hierarchy. Segmentation of the first (second) step is used as a binary mask for the second (third) step. Each network consists of multiple layers of anisotropic and dilated convolution filters that were obtained by training each network end-to-end. Residual connections and multi-scale predictions were employed in these networks to boost the segmentation performance. Experiments on BRATS 2017 online validation set predicted average Dice scores of 0.764, 0.897, 0.825 for enhanced tumor core, whole tumor and tumor core respectively.

Keywords: Brain tumor, convolutional neural network, segmentation

1 Introduction

Gliomas are the most common brain tumors that arise from glial cells. They can be categorized into two basic grades: low-grade gliomas (LGG) that tend to exhibit benign tendencies and indicate a better prognosis for the patient, and high-grade gliomas (HGG) that are malignant and more aggressive. With the development of medical imaging, brain tumors can be imaged by various Magnetic Resonance (MR) modalities, such as T1, T1-contrast, T2 and Fluid Attenuation Inversion Recovery (FLAIR). Different modalities can provide complementary information to analyze different sub-regions of gliomas, such as tumor cores and edema regions.

Automatic segmentation of brain tumors and substructures is promising to provide accurate and reproducible measurements of the tumors. It has great potential for better diagnosis, surgical planning and treatment assessment for brain tumors [15, 2]. However, this segmentation task is challenging because 1) the size, shape, and localization of brain tumors have considerable variations

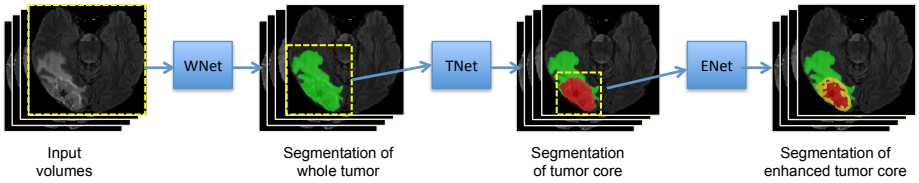


Fig. 1. The proposed triple cascaded framework for brain tumor segmentation. Three networks are proposed to hierarchically segment whole tumor (WNet), tumor core (TNet) and enhanced tumor core (ENet) sequentially.

among patient; 2) the boundaries between adjacent structures are often ambiguous due to the smooth intensity gradients, partial volume effects and bias field artifacts.

Discriminative methods based on deep neural networks have achieved state-of-the-art performance for multi-modality brain tumor segmentation tasks. Several key ideas to improve the performance of segmentation networks have been explored in the literature. These include efficient end-to-end training using a fully convolutional approach [1, 8], incorporating large visual contexts by employing a mixture of convolution and downsampling operations [12, 9], maintaining high resolution multi-scale features with dilated convolution and residual connection [17, 14, 5], and handling training data imbalance issue by designing new loss functions [7, 16] and sampling strategies [16].

Inspired by the previous work of cascaded neural networks for liver segmentation [6], we propose a cascade of CNNs for brain tumor subregion segmentation. We take advantage of dilated convolution, residual connection and multi-scale prediction to boost performance of the networks. In addition, we use anisotropic convolution to deal with 3D images as a trade-off between memory consumption and model complexity.

2 Methods

2.1 Triple Cascaded Framework

The proposed cascaded framework is shown in Fig. 1. We use three networks to hierarchically and sequentially segment substructures of brain tumors. The first network (WNet) segments the whole tumor from multi-modality 3D volumes of the same patient. The second network (TNet) segments the tumor core from the whole tumor region given by WNet, and the third network (ENet) segments the enhanced tumor core from the tumor core region given by TNet. Segmentation of the first (second) network is used as a binary mask for the second (third) network. These networks deal with binary segmentations and have different receptive fields. For WNet, the receptive field is the whole image region. The receptive field of TNet and ENet is the whole tumor region and tumor core region respectively. There are several benefits of using such a cascaded segmentation

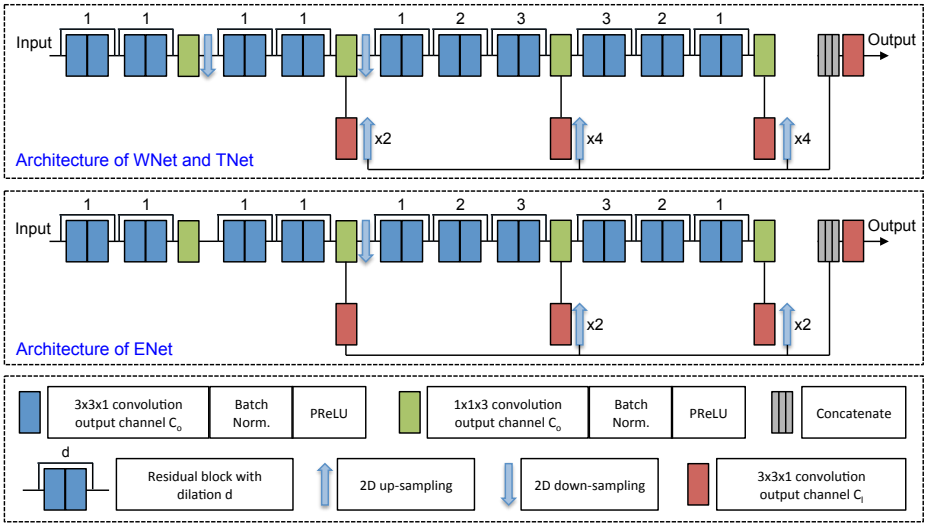


Fig. 2. Our anisotropic convolutional networks with dilated convolution, residual connection and multi-scale fusion. ENet uses only one downsampling layer considering its smaller input size.

framework. First, compared with training a single network for all substructures which requires complex network architectures, using three binary segmentation networks allows for a simpler network for each task. Therefore they are easier to train and can reduce over-fitting. Second, this helps reduce false positives since TNet only works on the region extracted by WNet and ENet only works on the region extracted by TNet. Third, this hierarchical pipeline follows the anatomic structure of tumors. It restricts the tumor core to be inside the whole tumor region and enhanced tumor core to be inside the tumor core region.

2.2 Anisotropic Convolutional Neural Networks

For 3D neural networks, the balance between memory consumption and feature representation ability should be considered. Many 2D networks take a whole 2D slice as input and can capture features in a large receptive field. However, taking a whole 3D volume as input consumes a lot of memory and therefore limits the resolution and number of features in the network, leading to a low representation ability. As a trade-off, we propose anisotropic networks that take a stack of slices as input with a large receptive field in 2D and a smaller receptive field along the direction orthogonal to the 2D slices. The architectures of our proposed MNet, TNet and ENet are shown in Fig. 2. All the networks are fully convolutional and use 10 residual connection blocks with anisotropic convolution, dilated convolution, and multi-scale fusion.

Anisotropic and Dilated Convolution. To deal with anisotropic receptive fields, we decompose a 3D kernel of size 3×3×3 into an intra-slice kernel with size 3×3×1 and an inter-slice kernel with size 1×1×3. Convolutional layers with either of these kernels have C_o output channels and each is followed by a batch normalization layer and an activation layer, as illustrated by blue and green blocks in Fig. 2. We use Parametric Rectified Linear Unit (PReLU) [10] in the activation layers. WNet and TNet use 20 intra-slice convolutional layers and four inter-slice convolutional layers with two 2D downsampling layers. ENets use the same set of convolutional layers as WNet but only one downsampling layer considering its smaller input size. We only employ up to two layers of downsampling in order to avoid large image resolution reduction and loss of segmentation details. After the downsampling layers, we use dilated convolution for intra-slice kernels to enlarge the receptive field within a slice. The dilation parameter is set to 1 to 3 as shown in Fig. 2.

Residual Connection. For effective training of deep CNNs, residual connections [11] were introduced to create identity mapping connections to bypass the parameterized layers in a network. Our MNet, TNet and ENet have 10 residual blocks. Each of the block contains two intra-slice convolutional layers, and the input of a residual block is directly added to the output, encouraging the block to learn residual functions with reference to the input. This can make information propagation smooth and speed the convergence of training [11, 14].

Multi-scale Prediction. In deep convolutional neural networks, sequential convolutional layers increase the receptive field and they capture features at different scales. Shallow layers learn to represent local and simple features while deep layers learn to represent global and abstract features. To combine both local and global features, we use three 1×3×3 convolutional layers at different scales of the networks to get intermediate predictions and upsample them to the resolution of the input. A concatenation of these predictions are fed into an additional 1×3×3 convolutional layer to obtain the final score map. These layers are illustrated by red blocks in Fig. 2. The outputs of these layers have C_l channels where C_l is the number of classes for segmentation in each network.

3 Experiments and Preliminary Results

Data and Implementation Details. We used the BRATS 2017[1] [15, 2, 4, 3] training and validation set for experiments. The training set contains images from 285 patients (210 HGG and 75 LGG). The BRATS 2017 validation set contains images from 46 patients with brain tumores of unknown grade. Each patient was scanned with four modalities: T1, T1c, T2 and FLAIR. We uploaded the segmentation results to the BRATS 2017 server which evaluated the segmentation and provided quantitative measurements in terms of Dice score,

[1] http://www.med.upenn.edu/sbia/brats2017.html

sensitivity, specificity and Hausdorff distance of enhanced tumor core, whole tumor, and tumor core respectively.

Our networks were implemented in Tensorflow[2] using NiftyNet[3]. We used Adaptive Moment Estimation (Adam) [13] for training, with initial learning rate 10^{-3}, weight decay 10^{-7}, batch size 5. Training was implemented on a an NVIDIA TITAN X GPU. We set C_o to 48 and C_l to 2 for MNet, TNet and ENet.

Table 1. Dice and Hausdorff measurements of our method (UCL-TIG) compared with top performance achieved by other teams. The results were provided by the BRATS 2017 validation leaderboard up to 18:00 July 23, 2017 (US eastern time). EN, WT, TC denote enhanced tumor core, whole tumor and tumor core respectively.

	Dice			Hausdorff		
	ET	WT	TC	ET	WT	TC
UCL-TIG*	0.7640	0.8970	0.8254	3.7133	3.9701	7.5424
MIC-DKFZ	0.7320	0.8964	0.7971	4.5470	6.9741	9.4767
CIAN	0.7112	0.8932	0.7349	4.1870	4.6126	8.1886
pvg	0.7129	0.8989	0.7514	6.9825	4.1616	8.6493
biomedia1	0.7570	0.9016	0.8202	4.2225	4.5576	6.1055

Table 2. Sensitivity and specificity measurements of our method (UCL-TIG) compared with top performance achieved by other teams. The results were provided by the BRATS 2017 validation leaderboard up to 18:00 July 23, 2017 (US eastern time). EN, WT, TC denote enhanced tumor core, whole tumor and tumor core respectively.

	Sensitivity			Specificity		
	ET	WT	TC	ET	WT	TC
UCL-TIG*	0.7748	0.9118	0.8412	0.9985	0.9942	0.9973
MIC-DKFZ	0.7900	0.8965	0.7807	0.9984	0.9956	0.9988
CIAN	0.7355	0.8925	0.6843	0.9983	0.9948	0.9988
pvg	0.7316	0.9037	0.7199	0.9982	0.9948	0.9982
biomedia1	0.7895	0.9088	0.7829	0.9982	0.9946	0.9986

Segmentation Results. Quantitative evaluation are shown on the BRATS 2017 leaderboard[4,5]. Table 1 presents Dice and Hausdorff measurements according to the leaderboard. It shows that our method achieves competitive results in terms of dice scores averaged over patients. Table 1 also shows our method

[2] https://www.tensorflow.org/
[3] http://niftynet.io/
[4] https://www.cbica.upenn.edu/BraTS17/lboardValidation.html
[5] Results retrieved on 18:00 23/July/2017 (US eastern time).

achieves low Hausdorff distances for different tumor subregions. Table 2 presents sensitivity and specificity measurements according to the leaderboard.

4 Conclusion

We developed a cascaded system to segment glioma subregions from multi-modality brain MR images. Results on BRATS 2017 online validation set predicted average Dice scores of 0.764, 0.897, 0.825 for enhanced tumor core, whole tumor and tumor core respectively.

Acknowledgements. This work was supported through an Innovative Engineering for Health award by the Wellcome Trust [WT101957], Engineering and Physical Sciences Research Council (EPSRC) [NS/A000027/1], the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact Initiative), UCL EPSRC CDT Scholarship Award [EP/L016478/1], a UCL Overseas Research Scholarship, a UCL Graduate Research Scholarship, and the Health Innovation Challenge Fund [HICF-T4-275, WT 97914], a parallel funding partnership between the Department of Health and Wellcome Trust.

References

1. B, M.H., Guizard, N., Chapados, N.: HeMIS : Hetero-Modal Image Segmentation. In: MICCAI. vol. 1, pp. 469–477 (2016)
2. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data (2017)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J., Freymann, J., Farahani, K., Davatzikos, C.: Segmentation Labels for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive (2017)
5. Chen, H., Dou, Q., Yu, L., Heng, P.A.: VoxResNet: Deep Voxelwise Residual Networks for Volumetric Brain Segmentation. NeuroImage pp. 1–9 (2016), http://arxiv.org/abs/1608.05895
6. Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbruster, M., Hofmann, F., Anastasi, M.D., Sommer, W.H., Ahmadi, S.a., Menze, B.H.: Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. In: MICCAI. vol. 1, pp. 415–423 (2016)
7. Fidon, L., Li, W., Garcia-peraza herrera, L.C.: Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation using Holistic Convolutional Networks. arxiv pp. 1–11 (2017), https://arxiv.org/abs/1707.00478

8. Fidon, L., Li, W., Garcia-peraza herrera, L.C., Ekanayake, J., Kitchen, N., Ourselin, S., Vercauteren, T.: Scalable multimodel convolutional networks for brain tumour segmentation. In: MICCAI (2017)
9. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain Tumor Segmentation with Deep Neural Networks. Medical Image Analysis 35, 18–31 (2016)
10. He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: ICCV (2015), http://arxiv.org/abs/1502.01852
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: CVPR (2016)
12. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. Medical Image Analysis 36, 61–78 (2017)
13. Kingma, D.P., Ba, J.L.: Adam: a Method for Stochastic Optimization. International Conference on Learning Representations 2015 pp. 1–15 (2015)
14. Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M.J., Vercauteren, T.: On the Compactness , Efficiency , and Representation of 3D Convolutional Networks : Brain Parcellation as a Pretext Task. In: IPMI (2017)
15. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, Ç., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). TMI 34(10), 1993–2024 (2015)
16. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. arXiv pp. 1–8 (2017), http://arxiv.org/abs/1707.03237
17. Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., Klusmann, M., David, A.L., Deprest, J., Vercauteren, T., Ourselin, S.: DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation. arXiv (2017), https://arxiv.org/abs/1707.00652

Overall Survival Time Prediction for High Grade Gliomas based on Sparse Representation Framework

Guoqing Wu[1], Yuanyuan Wang[1(✉)] and Jinhua Yu[1,2(✉)]

[1] Department of Electronic Engineering, Fudan University, Shanghai, China
[2] The key laboratory of medical imaging computing and computer assisted intervention of shanghai, China
yywang@fudan.edu.cn; jhyu@fudan.edu.cn

Abstract. In this paper, we propose a sparse representation-based radiomics framework to predict overall survival (OS) time of HGG. Firstly, we develop a patch-based sparse representation method to extract the high-throughput texture features in tumor region. Then, we propose to combine locality preserving projection and sparse representation to select discriminating features more effectively. Finally, we treat the OS time prediction as a multi-classification task and apply sparse representation to classification. Experimental results demonstrate that, with 10-fold cross-validation, the proposed method achieves the prediction accuracy of 90.97% and 93.55% by using T1 contrast-enhanced and T2 weighted magnetic resonance (MR) images, respectively.

Keywords: HGG, OS Time Prediction, MR imaging, Sparse Representation.

1 Introduction

Recently, radiomics has been successfully used in clinical diagnosis and prognosis. By converting medical images into mineable high-throughput features, radiomics provides a more comprehensive quantification of the entire tumor, and subsequently makes effective decision on these data. Yu et al. [1] exploit the extracted 671 radiomics features to noninvasively estimate isocitrate dehydrogenase 1 (IDH1) mutation for low grade gliomas. Multi-modality radiomics features were combined in [2] to predict the OS time for HGG. Among these applications of radiomics, extracting and selecting effective features for specific problem is very crucial, since good features lead directly to accurate classification.

Sparse representation (SR) has demonstrated great advantages in image restoration, feature selection and pattern recognition. In image processing, SR generally exploits adaptive learning dictionaries rather than the traditional analytically-designed dictionaries with fixed basis, such as wavelet, to represent images. Therefore it gave rise to the ability to extract or represent some small textures and details which usually play a decisive role in image classification. In addition, SR considers that natural signals can be represented linearly by a small number of atoms in dictionary. These atoms represent the essential features of target data and can be selected by the l_p -norm regulariz-

er of sparse model. Based on this principle, SR was successfully used for feature selection [3], canonical correlation analysis [4].

In this paper, we propose a SR-based radiomics framework to predict the OS time for HGG. Specifically, first, we develop a SR-based method to convert the statistical distribution of tumor texture into high-throughput texture features. Second, we propose a novel model combining locality preserving projection (LPP) and SR to select the most discriminative features. Particularly, a new structure preservation regularizer is introduced in feature selection model for considering the structures of within-class samples and between-class samples. Finally, we use sparse representation classification (SRC) to predict the OS time.

2 The Proposed Method

2.1 Sparse Representation-based Feature Extraction

Fig. 1 shows the flow chart of SR-based image texture feature extraction. First, we extract image patch sets $\mathbf{Y}_L \in \boldsymbol{R}^{n\times d}$, $\mathbf{Y}_M \in \boldsymbol{R}^{n\times d}$ and $\mathbf{Y}_S \in \boldsymbol{R}^{n\times d}$ from the segmented tumor images corresponding to long-survivors, mid-survivors and short-survivors, respectively. Where n and N denote the size and number of image patches, respectively. Then we learn three dictionaries $\mathbf{D}_L \in \boldsymbol{R}^{n\times k}$, $\mathbf{D}_M \in \boldsymbol{R}^{n\times k}$ and $\mathbf{D}_S \in \boldsymbol{R}^{n\times k}$ from $\mathbf{Y}_L \in \boldsymbol{R}^{n\times d}$, $\mathbf{Y}_M \in \boldsymbol{R}^{n\times d}$ and $\mathbf{Y}_S \in \boldsymbol{R}^{n\times d}$, respectively, by using the K-singular value decomposition algorithm [5].

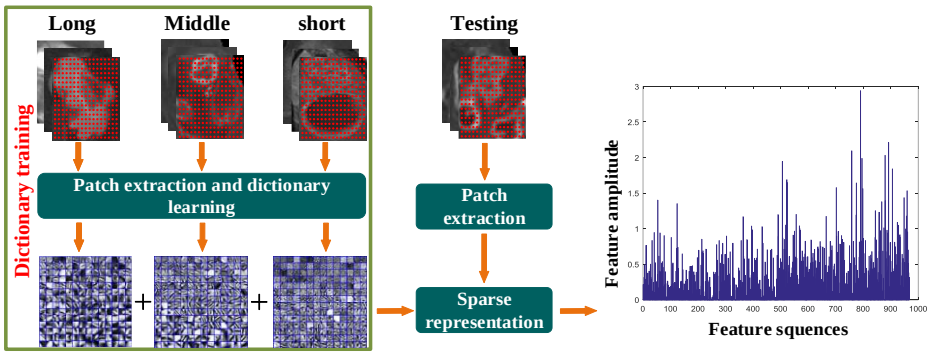


Fig. 1. Sparse representation-based feature extraction.

Finally, we extract the texture features by the following SR-based model:

$$\begin{cases} \hat{\boldsymbol{\Lambda}} = \sum_{i=1}^{m} \arg\min_{\boldsymbol{\alpha}_i} \left\| \boldsymbol{y}_i - \mathbf{D}\boldsymbol{\alpha}_i \right\|_2^2 + \phi \left\| \boldsymbol{\alpha}_i \right\|_0 \\ \boldsymbol{f} = \frac{1}{m} \sum_{i=1}^{m} \left| \boldsymbol{\alpha}_i \right| \end{cases}, \tag{1}$$

where $\mathbf{D} = [\mathbf{D}_L, \mathbf{D}_M, \mathbf{D}_S] \in \boldsymbol{R}^{n \times 3k}$, $\boldsymbol{y}_i$ denotes the i-th image patch in the tumor region of test image, m denotes the total number of image patches. $\hat{\boldsymbol{\Lambda}} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots \boldsymbol{\alpha}_m]$, $\boldsymbol{\alpha}_i \in \boldsymbol{R}^{3k}$ is the SR coefficients corresponding to $\boldsymbol{y}_i$. $\boldsymbol{f} \in \boldsymbol{R}^{3k}$ is the final obtained texture feature. ϕ is a regularization parameter. Orthogonal matching pursuit (OMP) algorithm could be used to solve the SR model in Eq. (1).

2.2 Sparse Representation-based Feature Selection

In this section, we propose a SR-based feature selection method to select the discriminative features. The proposed feature selection model is formulated as:

$$\min_{\mathbf{w}} \frac{1}{2}\left\|\hat{\mathbf{Y}} - \mathbf{w}^T\mathbf{F}\right\|_F^2 + \lambda_1[\beta tr(\mathbf{w}^T\mathbf{F}\mathbf{L}^w\mathbf{F}^T\mathbf{w}) - (1-\beta)tr(\mathbf{w}^T\mathbf{F}\mathbf{L}^b\mathbf{F}^T\mathbf{w})] + \lambda_2\left\|\mathbf{w}\right\|_{2,1}, \qquad (2)$$

where $\hat{\mathbf{Y}} \in \boldsymbol{R}^{c \times N}$ denotes a coding matrix of class labels, $\mathbf{F} = [\boldsymbol{f}_1 \cdots \boldsymbol{f}_i \cdots \boldsymbol{f}_N] \in \boldsymbol{R}^{d \times N}$ denotes high-dimensional feature data. d, N and c denote, respectively, the number of feature variables, subjects and classes. $\mathbf{w} \in \boldsymbol{R}^{d \times c}$ is a SR coefficient matrix. $\mathbf{L}^w \in \boldsymbol{R}^{N \times N}$ and $\mathbf{L}^b \in \boldsymbol{R}^{N \times N}$ denote the within-class and between-class graph laplacian matrix, respectively. $[\beta tr(\mathbf{w}^T\mathbf{F}\mathbf{L}^w\mathbf{F}^T\mathbf{w}) - (1-\beta)tr(\mathbf{w}^T\mathbf{F}\mathbf{L}^b\mathbf{F}^T\mathbf{w})]$ denotes a structure preservation term, which is designed to enable the within-class samples to be closer and the between-class samples to be far away in the new feature space. $0 \le \beta \le 1$ is a tuning parameter. Sparse regularization term $\left\|\mathbf{w}\right\|_{2,1}$ is designed to remove redundant features. λ_1 and λ_2 are the tuning parameters. Objective function (2) can be solved by the accelerated proximal gradient method [6].

Once we obtain the sparse matrix $\mathbf{w}$, we rank the l_2-norm value of each row of $\mathbf{w}$ in the descending order, then select features corresponding to the top-ranked rows. SRC proposed in [7] has achieved much success in pattern classification. Hence, in this paper, we apply SRC to predict the OS time.

3 Results and Discussion

We use the dataset provided by the MICCAI 'BraTS 2017' to validate the proposed method. This dataset was collected and organized by Bakas *et al* [8][9][10]. The segmentation benchmark proposed by [11] was used to segment the tumor images in the dataset. According to 'BraTS 2017', we cast the OS time prediction problem as a multi-classification task and divide the data into three categories based on survival, i.e. long-survivors (e.g., >15 months), mid-survivors (e.g. between 10 and 15 months) and short-survivors (e.g., <10 months). Our experimental dataset consists of 50 long-survivors, 42 mid-survivors and 63 short-survivors. For each case, we first extract 1452 texture features from T1 contrast-enhanced and T2-weighted MR images, respectively. Then we use the SR-based feature selection method to remove the redun-

dant features. Finally, the selected features are fed into SRC to predict OS time. We use leave-one-out cross validation (LOOCV) and 10-fold cross validation (Tenth) to validate the proposed model, respectively. The classification accuracy of the overall subjects and each class of subjects are calculated, respectively, to evaluate the prediction performance. The prediction results of the proposed framework is reported in Table 1.

As could be seen, the texture feature of T2 weighted MR image achieves the promising prediction result, with the highest accuracy of 96.13% in terms of LOOCV. For the two modalities, the gaps between the results of two validation methods is less than 4.5%. This demonstrates that the proposed method has high robustness in terms of the proportion of training samples to testing samples. In addition, the prediction performance of T2-weighted MR image is superior to that of T1 contrast-enhanced MR image as a whole.

Table 1. Prediction accuracy of different features and validation methods (%).

Method	ACC	ACC-long	ACC-middle	ACC-short
T1-Loocv	94.84	92.00	92.86	98.41
T1-Tenth	90.97	90.00	83.33	96.83
T2-Loocv	**96.13**	100.0	90.48	96.83
T2-Tenth	93.55	92.00	90.48	96.83

4 Conclusions

In this paper, we proposed a novel sparse representation framework to predict the (long, middle or short) OS time for HGG patients. Experimental results shown that our extracted features significantly improved the predictive accuracy of OS time. This further demonstrates that there is a close relationship between image texture and the OS time, even though these relationship is still poorly understood in clinical, the proposed method will help to understand it.

References

1. Yu, J. H., et al.: Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. Eur Radiol. pp. 1–14, (2016)
2. Liu, L., et al.: Outcome Prediction for Patient with High-Grade Gliomas from Brain Functional and Structural Networks. In: MICCAI, pp. 26–34, (2016)
3. Lin, D., et al.: Sparse models for correlative and integrative analysis of imaging and genetic data. J. Neurosci. Meth. 237, 69–78 (2014)
4. Lin, D., et al.: Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. Med. Image Anal. 18(6), 891–902 (2016)
5. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. 15(12), 3736–3745 (2006)

6. Zhu, X., et al.: Subspace regularized sparse multi-task learning for multi-class neuro-degenerative disease identification. IEEE Trans. Biomed Eng. 63(3), 607–618 (2016)
7. Wright, J., et al.: Robust face recognition via sparse representation," IEEE Trans. Pattern Anal. Mach. Intell. 31(2), 210–227 (2009)
8. Bakas, S., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data, (2017) [In Press]
9. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
10. Bakas, S., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
11. Menze, B. H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med Imaging. 34(10), 1993–2024 (2015)

Automatic segmentation of brain tumor from MR images using SegNet: selection of training data sets

Tsai-Ling Yang, Yu-Nian Ou, Teng-Yi Huang

Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan
tyhuang@mail.ntust.edu.tw

Abstract. In this study, we developed automatic segmentation of brain tumor using SegNet, a method based on a two-dimensional convolutional neural network and used the HGG data sets (n = 210) of BraTS 2017 for network training. We compared training schemes including or excluding slices without labeled tumor regions. The input images were FLAIR images. From the results, the dice similarity coefficients (dice=0.74, 2-fold cross-validation) obtained with training data sets excluding slices without labeled tumor region was significantly higher than those obtained with all slices ($P<0.05$, paired T- test). In the preliminary results, we were able to perform fully automatic segmentation of whole tumor region using SegNet. We aim to improve the performance of segmentation using more MR image contrasts (i.e., T1, T2, and T1Gd)

Keywords: Gliomas, Deep learning, SegNet, Image segmentation

1 Introduction

Brain gliomas are tumors associated with glial cells. They are the primary brain tumor among adults. Investigations improving diagnosis of gliomas are thus highly desirable. BraTS 2017 hosts the competition [1-4] for segmentations of brain gliomas with MR images. In this study, we attempt to use the recently advanced convolutional neural network architecture, SegNet [5] for this task. SegNet is a semantic segmentation architecture consisting of an encoder-decoder network followed by a pixel-wise classification layer. We compared different training schemes for SegNet.

2 Method

The training data sets provided by BraTS 2017 included multimodal MRI scans (HGG:210, LGG: 75). Four types of MR images (T1, T1 contrast enhanced, T2 and FLAIR) and images segmented manually by one to four raters (3 labels, 1: the necrotic and non-enhancing tumor, 2: the peritumoral edema, 4: GD-enhancing tumor) were provided. The images were all combined into a 3D volume (matrix size: 240x240x155) with the NIfTI file format. In this study, we used FLAIR images of 210 HGG data sets to evaluate the segmentation algorithm. The 3D volume of a subject was converted into 155 axial slices with a matrix size of 240x240. The data set was termed the group B. We further constructed a data set (the group A) excluding slices without labeled tumor

region. We used SegNet for pixel-wise semantic segmentation. The architecture of our SegNet implementation is shown in Fig. 1. The network consisted of encoder and decoder layers with 26 convolutional layers and an output layer connected a multi-class soft-max classifier. It was implemented with the Tensorflow framework. Two-fold cross-validation was used to evaluate the performance of the segmentation. The software procedure slice-by-slice segmented a whole brain FLAIR volumes and then merged the 2D slices into a 3D volume with the NIfTI file format. We subsequently evaluated the dice similarity coefficients between the true labels and the obtained whole-tumor segmentations. With groups A and B, we evaluated three schemes to train the SegNet. Two of the schemes were to train SegNet with A and B, respectively. The third scheme (A+B) is to use the SegNet trained with A as pretrain weights and then train the SegNet with B.

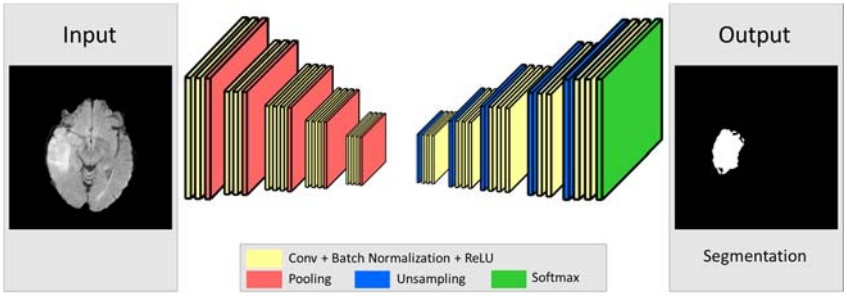


Fig. 1. SegNet architecture.

3 Results

Figure 2 displays an example of segmentation results. Figure 2(a) shows the true segmentation provided by BraTS 2017 and Figures 2(b-d) displays the segmentation results obtained with SegNet trained with different training schemes (i.e., A, A+B, B). The numbers underneath the figures are the dice similarity coefficients of the segmentation results. The red region indicated the segmented area of the whole tumor. In this demonstration, the scheme A produced a higher dice coefficient, which is consistent with visual observation. Table 1 lists the average dice coefficients of the segmentation results of 210 HGG subjects. The results suggest that dice coefficients produced by training schemes A and A+B are significantly higher than those generated by the scheme B ($P<0.05$, paired T-test). No prominent differences between the schemes A and A+B were identified.

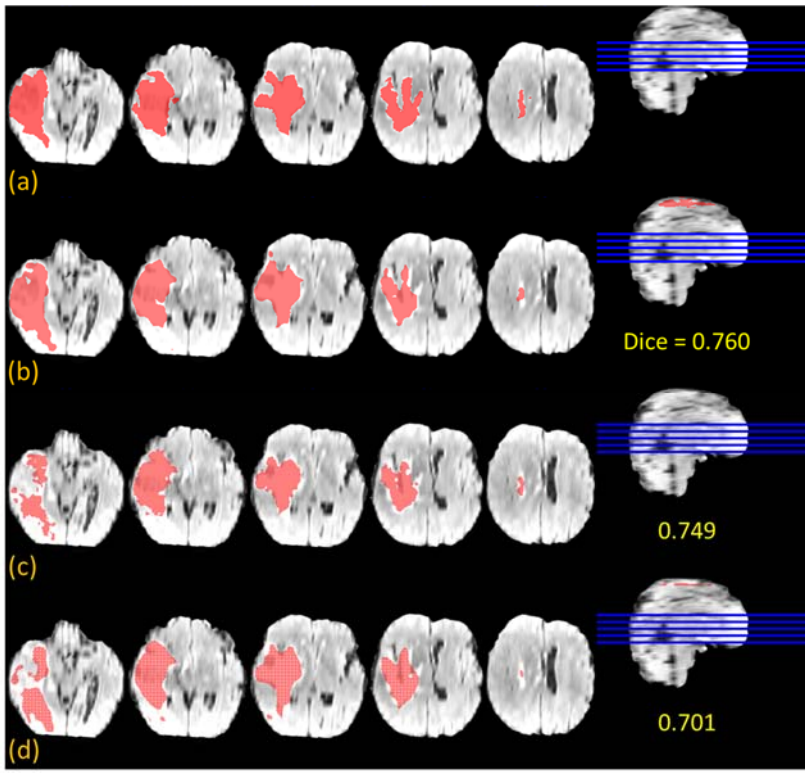


Fig. 2. The segmentation results of a subject: (a) reference (b)with the scheme A, (c) with the scheme A+B (d) with the scheme B

Table. 1. Dice coefficients of three schemes

	A	A+B	B
Average	0.740 ± 0.141	0.736 ± 0.16	0.729 ± 0.151
	A vs. A+B	A+B vs. B	A vs. B
T-Test	0.2787	0.0178	0.0013

Group A: excluding slices without labeled tumor region
Group B: all slices
Group A+B: pretraining with A, training with B

4 Discussions and Conclusions

In this study, we used SegNet for glioma segmentations with three training schemes. The dice metrics obtained using the schemes A and A+B were significantly higher than that obtained using the scheme B. No significant differences between the results obtained by A and A+B were identified. The results support using the scheme A as the

training method since training with A+B requires longer training period. During the BraTS 2017 competition, we continue to investigate methods to improve the segmentation results. For the results shown in this abstract, we only use the FLAIR images of the HGG data sets. The best segmentation results achieved the average dice coefficient of 0.74 for the whole HGG data set. We attempt to include more image contrasts (T1, T1ce, and T2) into SegNet to evaluate whether multi-modal images could improve the segmentation results.

References

1. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
2. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
3. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)
4. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press]
5. Vijay Badrinarayanan, Alex Kendall, Roberto Chipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Jan 2017.

3D segmentation of the brain tumor in MRI

Zeyu Zhang*, Mingen Zheng, and Lan Yi

Cisco Systems (China) Research and Development Co.,Ltd.
926 Yishan Rd. Caohejing Hi-Tech Park, Shanghai. P.R.C. 200233
{zeyzhang,zmingen,layi}@cisco.com

Abstract. This work mainly tries to improve the two-pathway model for segmenting the brain tumor from MRI data. We extend the model to directly process 3D data and predict tumor voxel. There are also some other optimizations like pre-classification of voxels, multiple resolution based convolution network to help improve the segmentation performance.

Keywords: Brain tumor segmentation, 3D deep neural network, Convolutional neural network, Multi-resolution deep network, Random Forest Classification

1 Introduction

There have been a lot of algorithms and model to do tumor segmentation in Magnetic Resonance Image (MRI) data. Details please refer to the survey [4] and we do not repeat here. Among these work, the two-pathway model [1] is a simple yet efficient model which have been examined for its good performance in previous BraTs Challenging with the real data. In our work, we did some optimizations to extend the original two-pathway models. Currently the work are still undergoing and the performance is still on tuning, so the report here is a little bit simple and corresponding method/model may even change in future.

2 Method

Our method of this brain tumor segmentation work is originated from the two-pathway model [1]. We believe that the learning directly on 3D would better utilize the spatial hints within the 3D MRI data thus improve the model performance.

[3] is a good work to directly model and learning 3D MRI data for brain tumor segmentation. I addition, it utilize the relationship between multi-scale image data to help improve the segmentation result. We first adopt its 3D modeling in modeling and experiments, and then we tried to scale the original data into different resolutions to understand how it can benefit the segmentation.

* This work is done when he was in internship in Cisco Systems (China) Research and Development Co.Ltd.

[2] did some interesting work to pre-classify the 2D images to help segmentation. Here we also do pre-classification on 3D data so that the model can early know the drafted category of 3D voxels.

3 Experiment Result

The experiments are done on BraTs Challenging 2017 training and validation data [5][6][7][8]. Currently the model is still under tuning and the performance looks not so good. Anyway, we would redesign some part of the model to overcome the slow convergence problem and the slow learning issues in the model. Besides, we hope the model tuning work later can help improve the accuracy.

Label	Dice_ET	Dice_WT	Dice_TC	Sensitivity_ET	Sensitivity_WT	Sensitivity_TC	Specificity_ET	Specificity_WT	Specificity_TC	Hausdorff95_ET	Hausdorff95_WT	Hausdorff95_TC
Brats17_CBICA_AAM_1	0	0.1409	0.0219	0	0.3701	0.1861	1	0.7494	0.8382		93.27915	142.3728
Brats17_TCIA_400_1	0	0.063	0.0164	0	0.283	0.1109	1	0.7747	0.8851		129.472	145.2205
Brats17_TCIA_617_1	0	0.1343	0.0725	0	0.2187	0.0913	1	0.8413	0.9262		102.645	112.39
Mean	0	0.1128	0.0369	0	0.2906	0.1294	1	0.7885	0.8832		108.4654	133.3277
StdDev	0	0.0432	0.031	0	0.076	0.0501	0	0.0475	0.044		18.78532	18.18848
Median	0	0.1343	0.0219	0	0.283	0.1109	1	0.7747	0.8851		102.645	142.3728
25quantile	0	0.0987	0.0191	0	0.2508	0.1011	1	0.762	0.8617		97.96209	127.3814
75quantile	0	0.1376	0.0472	0	0.3266	0.1485	1	0.808	0.9057		116.0585	143.7966

Fig. 1. Preliminary result of the model

4 Conclusion

Our method is more like a 3D version of two-pathway model. It is still not so good as we expected, as the network is still not so deep and we are afraid the segmentation capability may not be enough. Later we would first overcome the slow convergence problem and the tune the segmentation performance. After that, we would like to investigate the deeper network and if possible to utilize the advantage of ResNet structure to help the segmentation work.

References

1. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. arXiv preprint arXiv:1505.03540 (2015)
2. Eric Malmi, Shameem Parambath, Jean-Marc Peyrat, Julien Abinahed, Sanjay Chawla: CaBS: A Cascaded Brain Tumor Segmentation Approach (2015)

3. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. arXiv preprint arXiv:1603.05959 (2016)
4. Akkus Z, Galimzianova A, Hoogi A, et al. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions[J]. Journal of Digital Imaging, 2017: 1-11.
5. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp , Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).
6. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, (2017) [In Press].
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C. "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection", The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF

-

Automatic Brain Tumor Segmentation with 3D Deconvolution Network with Dilated Inception Block

Liang Zhao

[1] Siemens, USA
[2] State University of New York at Buffalo, USA

Abstract. Segmenting brain tumors from multi-modal imaging remains to be a challenging task despite the growing interest in the area. Brain tumors have a highly variable shape, appearance and spatial location. In this paper, we simplify 3D inception network [5] with dilated convolution [6] to capture multi-scale information. And we build a deconvolution network with dilated inception blocks to segment 3D images.

1 Method

1.1 Pre-processing

For each channel of each MRI case, we compute the standardized z-scores (zero mean and unit covariance) to put the data in the same scale.

1.2 Dilated Inception Residual Block

The original inception network make convolutions with kernel size 1, 3, 5, 7 in parallel way. Some following works replace a convolution with kernel size 5 with two convolutions with kernel size 3. To make the network simpler, we replace a convolution with kernel size $2 * n + 1,\ n = 1,\ 2,\ 3$ with a convolution with the kernel size 3 and the dilation n.

In this way, we can get the following two kinds of blocks (Fig. 1, Fig. 2).

1.3 Deconvolution Network

As the inception network should be able to get multi-scale information, we do not use up-sampling to merge features from different scales.

We use a deconvolution network to make the segmentation. (Fig. 3) We use two reduction blocks and one deconvolution layer to make the input and output be in the same scale.

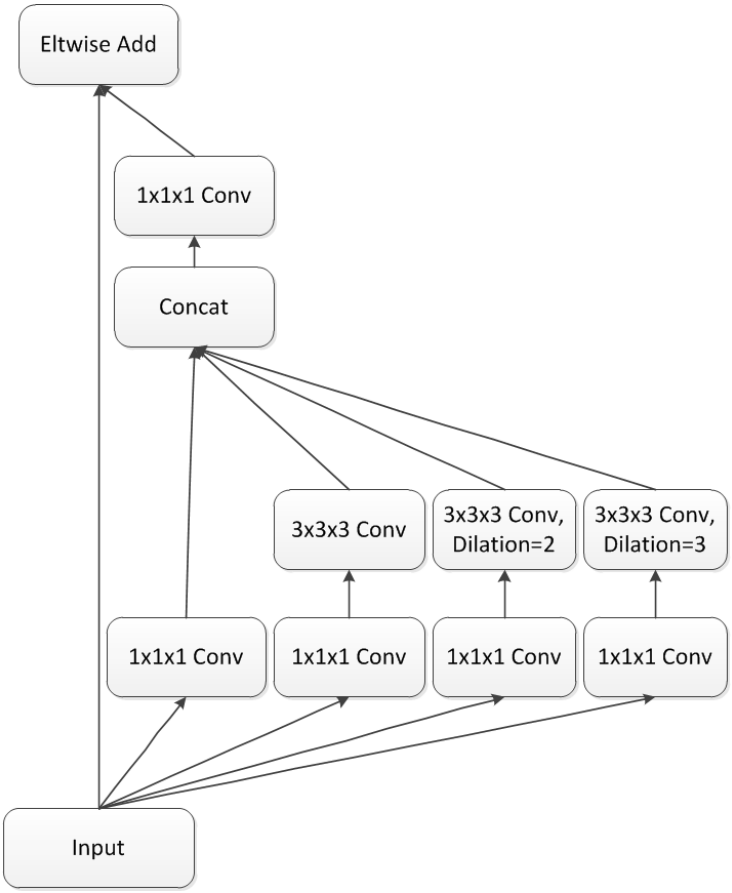


Fig. 1. Dilated inception resent block.

2 Experiments

We have evaluated our method on BRATS 2017 [4, 3, 1, 2] Validation data with DICE score.

The average DICE Score is 0.8709, 0.7824, 0.5777 for whole tumor, tumor core and enhancing tumor, receptively.

References

1. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *The Cancer Imaging Archive*, 2017.

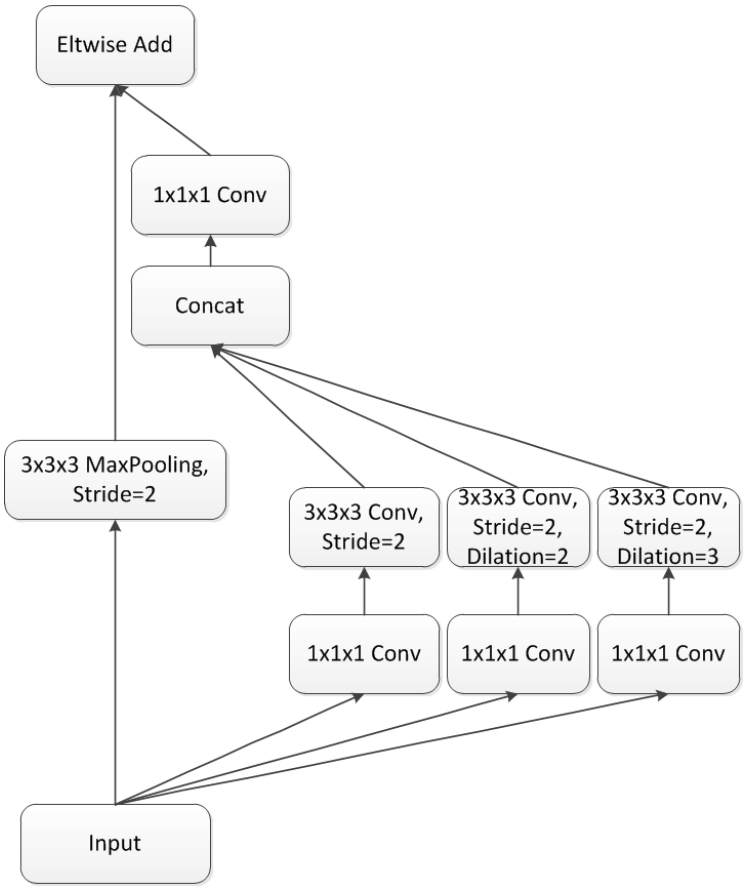


Fig. 2. Dilated reduction block.

2. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The Cancer Imaging Archive*, 2017.
3. S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 2017 [In Press].
4. B. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. Avants, N. Ayache, P. Buendia, L. Collins, N. Cordier, J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. Iftekharuddin, R. Jena, N. John, E. Konukoglu, D. Lashkari, J. Antonio Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. Riklin-Raviv, S. Reza, M. Ryan, L. Schwartz,

H.-C. Shin, J. Shotton, C. Silva, N. Sousa, N. Subbanna, G. Szekely, T. Taylor, O. Thomas, N. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. Hye Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 2015.
5. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *ICLR 2016 Workshop*, 2016.
6. F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *ICLR*, 2016.

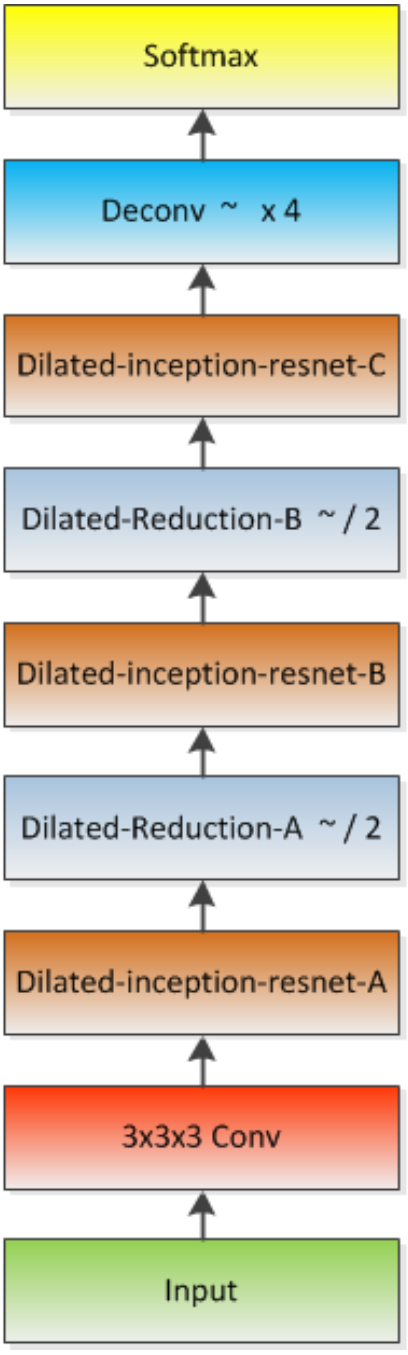


Fig. 3. Deconvolution network.

3D brain tumor segmentation through integrating multiple 2D FCNNs

Xiaomei Zhao[1,2] ✉, Yihong Wu[1] ✉, Guidong Song[3] , Zhenye Li[4] , Yazhuo Zhang[3,4,5,6] and Yong Fan[7]

[1] National Laboratory of Pattern Recognition,Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Beijing Neurosurgical Institute, Capital Medical University, Beijing, China
[4] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[5] Beijing Institute for Brain Disorders Brain Tumor Center
[6] China National Clinical Research Center for Neurological Diseases, Beijing, China
[7] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
zhaoxiaomei14@mails.ucas.ac.cn; yhwu@nlpr.ia.ac.cn

Abstract. The Magnetic Resonance Images (MRI) which can be used to segment brain tumors are of 3D. To make use of 3D information, 3 2D Fully Convolutional Neural Networks (FCNNs), each of which is trained to segment brain tumor images from axial, coronal, and sagittal views respectively, are integrated in this paper. The 3 FCNN models are integrated by fusing their segmentation results rather than by fusing into one deep network, which makes sure that each FCNN model is still allowed to be tested by 2D slices, guaranteeing the testing efficiency. A majority voting strategy is applied to do the fusing job. The proposed method can be easily extended to integrate more FCNN models which are trained to segment brain tumor images from more views, without retraining the FCNN models that we already have. In addition, Conditional Random Fields (CRFs) are applied to make sure the appearance and spatial consistency of our segmentation results. Experimental results show that, integrating the segmentation results of multiple 2D FCNNs obviously improve the segmentation accuracy.

Keywords: Brain Tumor Segmentation, Fully Convolutional Neural Networks, Conditional Random Fields, Multi-views

1 Introduction

Brain tumor segmentation results provide the volume, shape, and localization of brain tumors, which are crucial for brain tumor diagnosis and monitoring. Brain tumor segmentation technologies develop fast in recent years, especially those methods based on deep learning.

Besides in brain tumor segmentation area, deep learning has been successfully used in many other medical image segmentation areas. According to statistics, segmentation is the most common subject among the literatures that apply deep learning to medical images, and Convolutional Neural Networks (CNNs) are the most successful type of deep learning models for image analysis [1]. CNNs based methods have won many medical image segmentation challenges, such as Multimodal Brain Tumor Segmentation Challenge (BRATS) [2] and International Symposium on Biomedical Imaging (ISBI) cell tracking challenge [3].

Many kinds of medical images, such as the Magnetic Resonance Images (MRI) which can be used to segment brain tumors, are of 3D. To take full use of 3D information for medical image analysis, it is better to use 3D CNNs. However, 3D CNNs have large memory and training time requirements [4]. Therefore, many researchers have tried to integrate multiple 2D CNNs for segmenting 3D medical images, such as [4] and [5]. These methods integrated their multiple 2D CNNs into one deep network, and the 2D patches in multi-views centered at the same voxel should be sent into their deep networks at the same time. Under this situation, 3D images could only be segmented patch by patch, which is a very slow testing strategy, even if we change their CNNs into FCNNs. To improve the testing efficiency, we integrate multiple 2D FCNNs by integrating their segmentation results. Each FCNN model is trained by patches but tested by 2D slices, which improves the testing speed greatly. In this paper, we train 3 2D FCNN models using 2D patches of axial, coronal, and sagittal views respectively. During testing, we use these 3 networks to segment brain tumors slice by slice in 3 different views, yielding 3 segmentation results. Then we fuse these 3 segmentation results by voting. Experimental results show that this strategy is useful to improve segmentation accuracy. We also use CRF to make sure the appearance and spatial consistency of our segmentation results.

2 Method

The proposed segmentation method consists of 5 main steps: pre-processing, segmenting brain images slice by slice using 3 2D FCNN models, processing segmentation results using CRF, fusing segmentation results obtained in 3 different views, and post-processing. The fusing operation is performed before or after CRF, as shown in Fig. 1-(a) and Fig.1-(b) respectively.

2.1 Pre-processing

To make similar intensities in MRI scans of the same modality have similar tissue meanings, pre-processing steps are utilized. Our pre-processing steps include N4ITK [6] and intensity normalization [7].

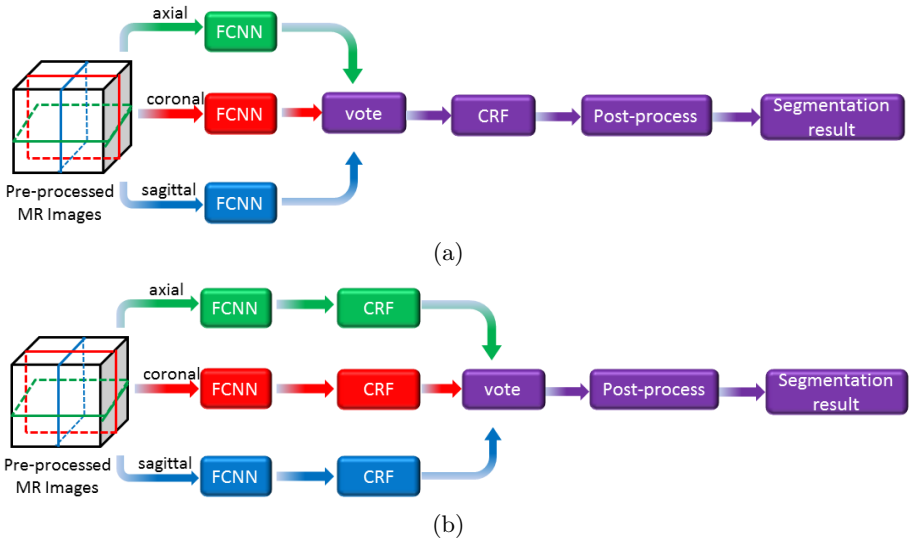


Fig. 1. Flowchart of our brain tumor segmentation method: (a) fusing the segmentation results obtained from different views before CRF; (b) fusing the segmentation results obtained from different views after CRF

2.2 Segmenting brain images by FCNNs

We use the same FCNN structure proposed in [7] as shown in Fig. 2, which has two different sizes of inputs. The large inputs pass through a series of convolutional and pooling layers and turn into feature maps with the same size of small inputs. These feature maps together with small inputs are used to predict their center pixel's label. Different from [7], we train 3 FCNN models in this paper, using 2D patches extracted from axial, coronal, and sagittal slices respectively. During testing, we use these 3 segmentation models to segment brain images slice by slice from 3 different views and obtain 3 segmentation results.

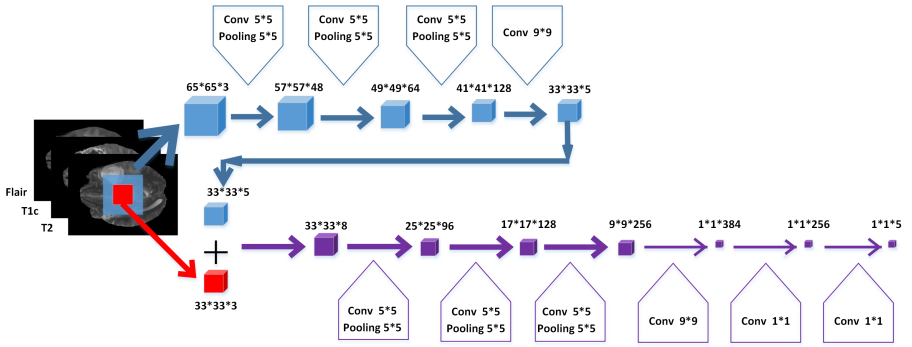


Fig. 2. The structure of our FCNN model

2.3 Processing segmentation results by 3D CRF

To make sure the appearance and spatial consistency of segmentation results, we use CRF to optimize our segmentation results. We have tried Conditional Random Fields as Recurrent Neural Networks (CRF-RNN) [8] and 3D CRF [9]. Experimental results showed that 3D CRF performed slightly better than CRF-RNN.

2.4 Fusing segmentation results obtained in 3 different views

As described in Section 2.2, 3 FCNN models are trained to segment brain images from 3 different views. During testing, their segmentation results are fused to make better use of the 3D information provided by the 3D MRI scans. In our experiments, the 3 segmentation results obtained in 3 different views are fused by majority voting.

We fuse the segmentation results as a post-processing step rather than fuse the 3 FCNN networks in one deep network, aiming to make sure that each FCNN model still has the ability to segment brain images slice by slice for efficiency. In this way, our method improves the efficiency of integrating multiple 2D CNNs while achieves better accuracy than using a single 2D CNN network.

2.5 Post-processing

We remove small isolated areas and correct some voxels label to post-process our segmentation results automatically by a simple thresholding method.

3 Experiment

3.1 Dataset

BRATS 2017 [10,11,12] separates its imaging data as 3 datasets, namely training dataset, validation dataset, and testing dataset. Each case in these 3 datasets contains multi-modal MRI scans, including T1, T1Gd, T2, and FLAIR. Similar to the previous BRATS challenges [13], all MRI scans of the same case have been skull-stripped, co-registered to the same anatomical template, and interpolated to the same resolution. Ground truth of each case in BRATS 2017 is produced by manual annotation. Up to now, the BRATS 2017 testing dataset is not available for participants. Therefore, we only use its training and validation dataset in our experiments.

BRATS 2017 training dataset contains 210 HGG (High Grade Glioma) cases and 75 LGG (Low Grade Glioma) cases. In this paper, we separate the 210 HGG into 2 subsets as 168 HGG and 42 HGG respectively. The subset of 168 HGG is used as the training dataset and the subset of 42 HGG is used as a testing subset. The ground truth of each training case is provided along with its MRI scans and we evaluate the segmentation results of the 42 HGG cases on our own personal computer.

BRATS 2017 validation dataset contains 46 cases with unknown grades. Since the ground truth of each validation case is not released to participants, its segmentation result can only be evaluated on its evaluation website[1].

3.2 Evaluation results

Our segmentation models are trained by the 168 HGG cases. The evaluation scores of our models on the other 42 HGG cases are shown in Tab. 1, where the method called Fusing(FCNNs)+3D CRF fuses the segmentation results before 3D CRF, as shown in Fig. 1-(a), and the method called Fusing(FCNNs+3D CRF) fuses the segmentation results after 3D CRF, as shown in Fig. 1-(b). The scores in Tab. 1 indicate that fusing the segmentation results obtained from different views obviously improve the segmentation accuracy. Tab. 1 also indicates that Fusing(FCNNs)+3D CRF+post-process has a better performance on PPV, while Fusing(FCNNs+3D CRF)+post-process has a better performance on Sensitivity. From the view of Dice, Fusing(FCNNs)+3D CRF+post-process performs slightly better on enhancing core, while Fusing(FCNNs+3D CRF)+post-process performs better on complete tumor.

Table 1. The average evaluation scores of 42 HGG cases

Methods	Dice			PPV			Sensitivity		
	Comp.	Core	Enh.	Comp.	Core	Enh.	Comp.	Core	Enh.
FCNNs(axial)	0.623	0.701	0.633	0.480	0.602	0.530	0.971	**0.916**	0.877
FCNNs(coronal)	0.666	0.738	0.675	0.578	0.664	0.594	0.963	0.896	0.865
FCNNs(sagittal)	0.662	0.703	0.653	0.526	0.607	0.574	0.957	0.912	0.846
Fusing(FCNNs)	0.696	0.800	0.730	0.562	0.753	0.663	**0.974**	0.901	0.877
Fusing(FCNNs)+3D CRF	0.859	0.867	0.824	0.931	**0.915**	0.804	0.816	0.850	0.880
Fusing(FCNNs)+3D CRF +post-process	0.864	**0.868**	**0.831**	**0.941**	0.909	**0.825**	0.818	0.854	0.873
FCNNs(axial)+3D CRF	0.857	0.843	0.787	0.873	0.840	0.731	0.857	0.875	**0.895**
FCNNs(coronal)+3D CRF	0.862	0.843	0.800	0.898	0.869	0.765	0.845	0.850	0.880
FCNNs(sagittal)+3D CRF	0.845	0.848	0.797	0.887	0.853	0.762	0.827	0.869	0.874
Fusing(FCNNs+3D CRF)	0.865	0.864	0.816	0.906	0.894	0.784	0.845	0.861	0.887
Fusing(FCNNs+3D CRF) +post-process	**0.873**	**0.868**	0.828	0.920	0.895	0.813	0.846	0.865	0.879
U-net(axial)	0.807	0.749	0.752	0.787	0.829	0.799	0.856	0.743	0.763
U-net(axial)+3D CRF	0.811	0.750	0.754	0.798	0.834	0.804	0.850	0.740	0.763

On the same 168 HGG cases, we train a U-net [3] to segment brain tumors and compare its performance with our method on the same 42 HGG cases. U-net won ISBI cell tracking challenge in 2015 and it is one of the most popular

[1] https://ipp.cbica.upenn.edu/

deep networks used to segment medical images in recently years. The evaluation scores in Tab. 1 show that the segmentation performance of U-net is much better than the performance of FCNNs. However, FCNNs+3D CRF works much better than U-net+3D CRF.

Table 2. The average Dice scores of BRATS 2017 validation dataset (46 cases)

Methods	Dice		
	Comp.	Core	Enh.
Fusing(FCNNs)+3D CRF +post-process	0.881	0.788	**0.754**
Fusing(FCNNs+3D CRF) +post-process	**0.888**	**0.792**	0.749

Apart from the 42 HGG cases, we also test our segmentation models on BRATS 2017 validation dataset. The Dice scores are shown in Tab. 2, which indicates that Fusing(FCNNs)+3D CRF+post-process performs slightly better on enhancing core, while Fusing(FCNNs+3D CRF)+post-process performs better on complete tumor and tumor core.

4 Conclusion

In this paper, we segment 3D brain images by integrating the segmentation results of multiple 2D FCNNs, which are trained to segment brain images from axial, coronal, and sagittal views respectively. Each of the 2D FCNN networks is tested slice by slice, guaranteeing the segmentation efficiency of our method. We also use CRF to optimize our segmentation results. Experimental results show that our fusing strategy improves segmentation accuracy. Moreover, the proposed method is not limited to fuse the 3 segmentation results obtained from 3 different views. It could be extended to fuse the more from more views.

Acknowledgements: This work was supported by the National High Technology Research and Development Program of China (2015AA020504) and the National Natural Science Foundation of China under Grant No. 61572499, 61421004.

References

1. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., and Snchez, C.I.: A survey on deep learning in medical image analysis. arXiv preprint arXiv:1702.05747 (2017)
2. Chang, P.D.: Fully Convolutional Deep Residual Neural Networks for Brain Tumor Segmentation. In International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 108-118 (2016)

3. Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234-241 (2015)
4. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., and Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In International conference on medical image computing and computer-assisted intervention, pp. 246-253 (2013)
5. Fritscher, K., Raudaschl, P., Zaffino, P., Spadea, M.F., Sharp, G.C., and Schubert, R.: Deep Neural Networks for Fast Segmentation of 3D Medical Images. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 158-165 (2016)
6. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C.: N4ITK: improved N3 bias correction. IEEE transactions on medical imaging 29, pp. 1310-1320 (2010)
7. Zhao, X., Wu, Y., Song, G., Li, Z., Fan, Y., and Zhang, Y.: Brain Tumor Segmentation Using a Fully Convolutional Neural Network with Conditional Random Fields. In International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 75-87 (2016)
8. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., and Torr, P.H: Conditional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1529-1537 (2015)
9. Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., and Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis 36, pp. 61-78 (2017)
10. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby JS., Freymann JB., Farahani K., Davatzikos C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data,(2017) [In Press]
11. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q
12. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, (2017). DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF
13. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 34 pp. 1993-2024 (2015)

Brain Tumor Segmentation with Cascaded Convolutional Neural Networks

Chenhong Zhou[1] , Changxing Ding[1] , Zhentai Lu[2] and Tianhao Zhang[3]

[1] South China University of Technology, Guangzhou, 510641, China
eezhouch@mail.scut.edu.cn
[2] Southern Medical University, Guangzhou, China
[3] General Electric Company

Abstract. In this paper, we propose a two-stage method based on U-Net and FusionNet for automatic brain tumor segmentation. In the first stage, we adopt a U-net architecture with a reweighted loss function to the 2D slices, in order to get a coarse segmentation mask with high sensitivity. Then, we dilate the segmentation mask generated from the first stage and divide mask into small cubic volumes. In the second stage, a 3D FusionNet is used to precisely segment these small cubic volumes and the predictions of cubic volumes are concatenated to get the final segmentation result. The proposed method has the following advantages: 1) it not only captures global information from whole slices but also grasps local details from small cubic volumes; 2) 2D and 3D information are both taken into consideration; 3) the network architectures in two stages are similar and both make dense pixel-wise predictions, leading to fast testing and training. Preliminary results on BRATS 2017 validation dataset demonstrate that proposed method can achieve good performance with DICE scores of 0.89, 0.80 and 0.75 for complete tumor, core tumor and enhancing tumor, respectively.

Keywords: Brain Tumor Segmentation, Cascaded Convolutional Neural Networks, Two Stage Method, U-Net, FusionNet, Multimodal MRI.

1 Introduction

With its outstanding performance in computer vision and pattern recognition, Convolutional Neural Networks (CNNs) have attracted enormous attention in medical image analysis fields. However, medical image data have two major characteristics different from natural image：1) medical data is usually three-dimensional (3D); 2) there exists a strong imbalance between the number of normal tissue and non-normal tissue. In order to deal with 3D data, most of CNN-based methods can be divided into 2D-CNNs and 3D-CNNs according to the dimensions of input. Although 3D-CNNs can make full use of the 3D information from MRI, 3D-CNNs usually use a relatively shallow network in order to alleviate increased computational cost and memory consumption, which leads to limited representation capability. However, 2D-CNNs can

utilize deeper network architecture to learn complicated representation to achieve good performance. Therefore, each method has its own advantages and disadvantages. On the other hand, these CNN-based methods can be roughly divided into single-prediction models and dense-prediction models. The former usually uses small patches to predict the label of its center pixel, while the later aims to dense label prediction, i.e., get the label of every voxel of the whole slices or large patches simultaneously. Clearly, the former only considers the local information and doesn't take spatial consistency into consideration; the later learns global information from whole slices or large patches, so it is much faster during testing and training.

Due to data imbalance, the learning process may converge to local minima of a sub-optimal loss function, thus predictions may strongly bias towards non-tumor tissue. To cope with this problem, single-prediction models may control sampling ratios of normal tissue and tumor, or use two phase training [1], where in the first phase, all labels of samples are equiprobable and in the second phase, labels of samples conform to the un-balanced nature of the data. On the other hand, other methods reweight loss function to alleviate this problem.

In this paper, we propose a two-stage segmentation method based on U-Net and FusionNet for automatic brain tumor segmentation. In the first stage, we employ 2D U-Net to the 2D slices to detect the locations of tumors roughly. Meanwhile we reweight the loss function to deal with data imbalance. Then we dilate coarse segmentation mask from the first stage and divide it into small cubic volumes. In the second stage, 3D FusionNet is used to precisely segment these cubic volumes and then we concatenate predictions of cubic volumes to get the ultimate and precise segmentation. The proposed method has the following advantages: 1) In the first stage, we use slices as input of 2D U-Net, so we can learn the global contextual information of whole slices to get coarse segmentation; in the second stage, we focus on small local cubic volumes to grasp the details to achieve precise predictions. 2) 2D and 3D information are both taken into consideration. 3) In both stages, the networks make dense pixel-wise predictions. Thus, it is fast during testing and training.

2 Data

We use the training set of BRATS 2017 challenge [2,3,7,8], which contains 210 high-grade glioma (HGG) and 75 low-grade glioma (LGG) patient scans. Each patient has four MRI sequences which are named T1, T1c, T2 and FLAIR. These are resampled to $1\times1\times1$ mm^3 and dimensions of each MRI sequence is $240\times240\times155$. The provided manual segmentations include four labels: 1 for necrotic (NCR) and the non-enhancing (NET) tumor, 2 for edema (ED), 4 for enhancing tumor (ET), and 0 for everything else, i.e. normal tissue and background (black padding). The official evaluation is calculated by merging the predicted labels into three sets: whole tumor (1,2,4), tumor core (1,4) and enhancing tumor (4).

3 Method

3.1 Pre-processing

To distinguish background from normal tissue, we relabel the background to label 3. Each sequence is individually normalized by subtracting the mean value and dividing by the standard deviation of all voxels in the sequence.

3.2 Two Stage Method

The First Stage. We apply a 2D U-Net to 2D slices to capture global information and get a coarse segmentation mask. The architecture is based on U-Net, consisting of one down-sampling path and one up-sampling path. Different from the original U-Net [4], we use zero padding for all the convolutional layers to keep the output dimension of both down-sampling and up-sampling path. And we employ batch normalization [5] ("BN") before each ReLU. The architecture is shown in Fig.1.

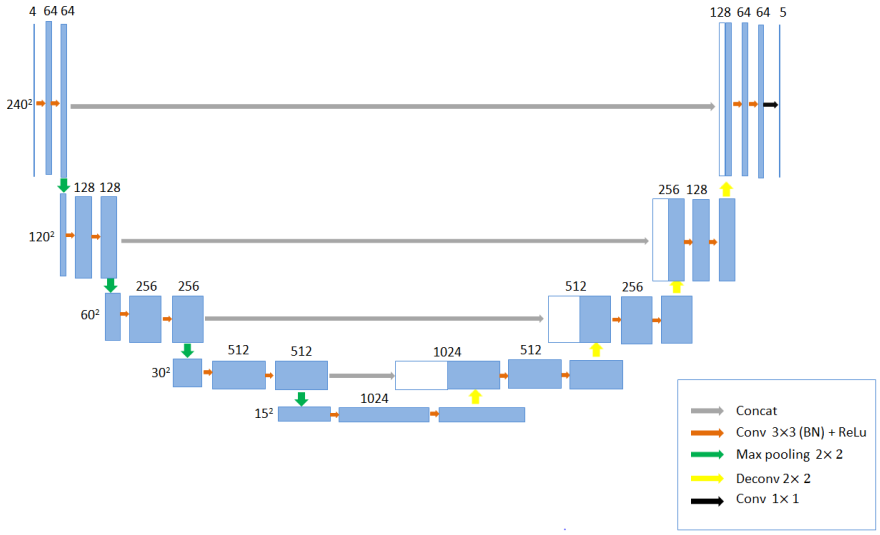


Fig. 1. The 2D U-Net architecture used in this work

Following [6], we adopt median frequency balancing method to handle the problem of label imbalance, where the weight assigned to a class in the loss function is the ratio of the median of class frequencies computed on the entire training set divided by the class frequency. The loss function that we used is categorical cross-entropy after softmax.

We use slicing images as inputs to train this network. The purpose of this stage is to obtain a segmentation mask which indicates if each voxel belongs to the tumor or

not. Then we dilate the segmentation mask by 10 voxels to reduce undetected tumor as much as possible.

The Second Stage. In this stage, we focus on local details and use a 3D FusionNet to predict the small cubic volumes to obtain precise and fine segmentation result. The 3D FusionNet is illustrated in Fig.2. Like the standard FusionNet [9], 3D FusionNet consists of an encoding path and a symmetric decoding path and four types of basic building blocks are used to construct the network. We use image patches of 32×32×16 voxels as input. Different from the standard FusionNet, all the convolutional layers are replaced with three-dimensional convolution kernels. In order to reduce the memory consumption, we decrease the depth of network and number of feature maps, therefore, there are four resolution steps, which means three max pooling layers are used and the number of feature maps are halved.

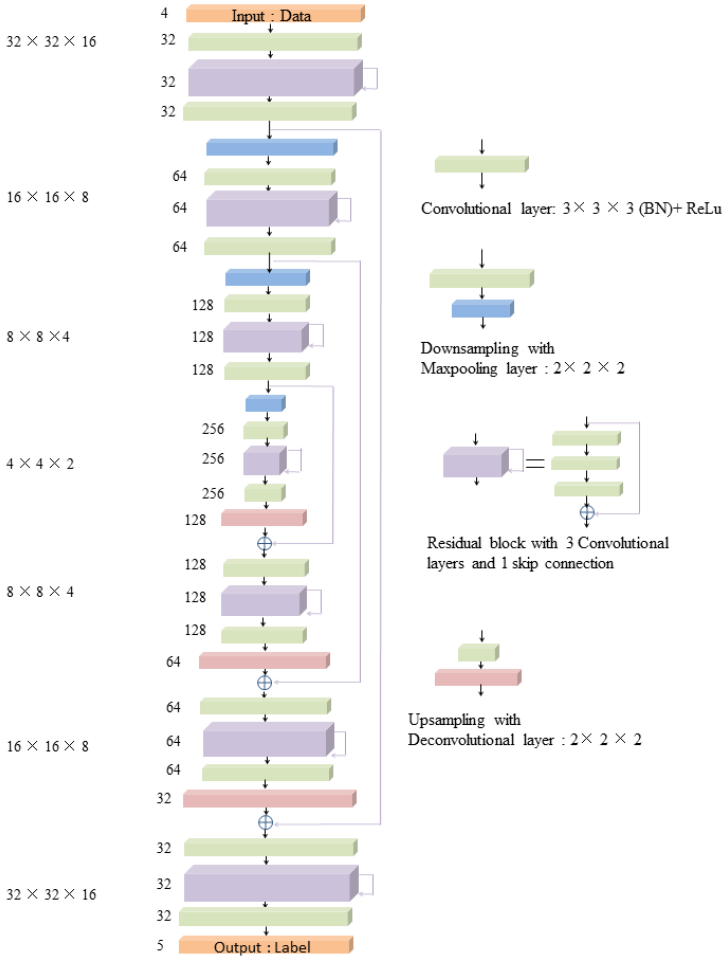


Fig. 2. The 3D FusionNet architecture used in this work

We sample 3D patches that are near to the tumor randomly as the training set for the 3D FusionNet. The dilated segmentation mask from the first stage is divided into small cubic volumes and the voxel labels are densely predicted by the 3D FusionNet. Then we concatenate the predictions of cubic volumes to get the final segmentation result.

3.3 Post-processing

We use a morphological filter to remove isolated small clusters. Some small clusters may be erroneously classified as tumor, thus we remove clusters in the segmentation obtained by the second stage if these clusters are smaller than a predefined threshold. We don't remove any clusters if the largest cluster is smaller than the threshold.

4 Results and Discussion

Preliminary results were obtained presented in Table 1. To evaluate the effect of our proposed method, we trained 2D U-Net without loss reweighting and compared it with the one-stage method (2D U-Net with loss reweighting) and two-stage method (2D U-Net with loss reweighting + 3D FusionNet).

Table 1. Performance on validation set.

	Dice		
	Complete	Core	Enhanced
2D U-net	0.87	0.75	0.69
One-stage	0.86	0.77	0.71
Two-stage	0.89	0.80	0.75

	Sensitivity		
	Complete	Core	Enhanced
2D U-net	0.85	0.70	0.75
One-stage	0.92	0.77	0.82
Two-stage	0.91	0.80	0.77

	Specificity		
	Complete	Core	Enhanced
2D U-net	0.99	0.99	0.99
One-stage	0.98	0.99	0.99
Two-stage	0.99	0.99	0.99

5 Conclusions

In this work, we propose a two-stage method based on U-Net and FusionNet for automatic brain tumor segmentation. The proposed method not only takes both 2D and 3D information into consideration, but also combines global contextual information of whole slices with local details extracted from small cubic volume. And both of stages make dense pixel-wise predictions, which accelerates predictions. Our empirical results indicate good performance is achieved. In the future, we would make more contributions to the network architecture and further promote the performance.

References

1. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y.,Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Med. Image Anal., 35, 18–31 (2017).
2. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp Ç, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).
3. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Nature Scientific Data, (2017) [In Press].
4. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. LNCS, vol. 9351, pp. 234-241. Springer (2015).
5. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167 (2015).
6. Eigen, David, and Rob Fergus.: Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In: Proceedings of the IEEE International Conference on Computer Vision (2015).
7. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection. The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.KLXWJJ1Q.
8. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby J, Freymann J, Farahani K, Davatzikos C.: Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-LGG collection. The Cancer Imaging Archive, 2017. DOI: 10.7937/K9/TCIA.2017.GJQ7R0EF.
9. Quan T M, Hilderbrand D G C, Jeong W K.: FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics. arXiv:1612.05360, 2016.

TP-CNN: A Two-Phase Convolution Neural Network based model to do automatic brain tumor segmentation by using BRATS 2017 data

Fan Zhou[b], Tengfei Li[a], Heng Li[a], Kaixian Yu[a], Yue Wang[b], and Hongtu Zhu[a,b]

Department of Biostatistics, University of Texas, MD Anderson Cancer Center[a]
Department of Biostatistics, University of North Carolina at Chapel Hill[b]

Abstract. This is the preliminary report for the BraTS 2017 challenge at MICCAI. The goal of this challenge is to utilize multi-institutional pre-operative MRI scans to segment out different tumor subregions, whose information will be then used to make predictions for patient overall survival. We build a two-phase patch-based convolution neural network (TP-CNN) to make classification for all the pixels in the brain regions, and refining the segmentation results by conditional random files (CRF) and certain post-processing procedures. With the segmentation results, we are able to extract different kinds of radiomic features and combine them with the age information to make subject-level prediction for the survival rate by using XGBoost method.

Keywords: Convolution Neural Network, XGBoost, Conditional Random Field, patch-based

1 Introduction and Data Summary

Brain tumor segmentation plays an important role in accurate diagnosis and efficient treatment of brain cancers, which has also drawn a lot of attention in the field of imaging analysis and computer vision. Certain kinds of brain tumors such as gliomas and glioblastomas are difficult to localize because they are often diffused, poorly contrasted, variously shaped and do not have fixed size or appearing location. Since manual segmentation may be time wasting and laborious, scientists take great effort in designing reliable automatic grain tumor segmentation methods and the BraTS 2017[3–6] challenge is one of such platforms such that all kinds of segmentation methods could be tested and evaluated.

The Brats 2017 challenge data set includes subjects from three different recourses, which are labeled with '2013', 'CBICA' and 'TCIA' respectively. Especially, 20 subjects from group '2013', 88 from group 'CBICA' and 102 from 'TCIA' are with high-grade gliomas (HGG), while the other 75 subjects from group '2013' and 'TCIA' are with low-grade gliomas (LGG). For each patient, four kinds of MRI scans are provided, including T2-weighted fluid attenuated inversion recovery (Flair), T1-weighted (T1), post-contrast T1-weighted (T1Gd), and T2-weighted (T2). All the four kinds of MRI scans are roughly registered

to a common template and resampled at 1mm resolution level, where the dimension of each MRI image is $240 \times 240 \times 144$. In the training data set, the annotation of three tumor regions, which are GD-enhancing tumor (ET label 4), the peritumoral edema (ED label 2), and the necrotic and non-enhancing tumor (NCR/NET label 1), are provided along with the MRI scans.

2 Method and Framework

2.1 Preprocessing

By checking the raw image data, we found that the MRI scans from different subjects may have varied intensity ranges and are suffering the bias which needs to be corrected. We use the the robust intensity normalization method proposed by [7] to do the preprocessing. The details of the procedures are described as the following steps:

1. Linearly transfer the original intensities of each individual MRI scans ($I_1(v)$) to the range between 0 and 255 ($I_2(v)$).
2. Calculate the intensity histogram with 256 bins. Then find the gray-value of highest histogram bin $x \in [0, 255]$ of $I_2(v)$.
3. Calculate the robust deviation which is formulated as $\sigma^2 = [\sum(v_i - x)^2]^{-1}, v_i \in I_2(v)$
4. Update the image value by $v_i^* = (v_i - x)/\sigma * c_1 + c_2$
5. Update the image value by $v_i^* = \min(\max(v_i^*, 0), 255)$

Figure 1 demonstrated the processing procedure of 2 subjects, which are randomly selected from group '2013' and 'TCIA' respectively.

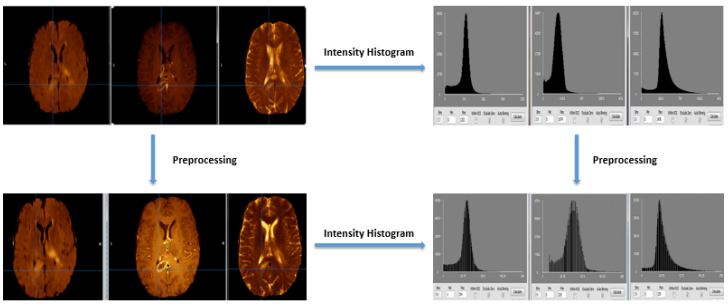


Fig. 1. Framework of TP-CNN procedure

2.2 Two-phase convolution neural network

After the preprocessing of the raw MRI images, we propose a patch-based two-phase convolution neural network (TP-CNN) to label all the pixels within the brain regions as non-tumor (0) or the three tumor classes (1, 2, 4). Specifically, the second-phase could be seen as a conditional optimization procedure, which is to refine the segmentation results within the tumor mask built by the first-phase CNN model. The input for both two CNN models will be the squared patches extracted from slices along the Z axis of the MRI images, where the corresponding label for each individual patch will the class that the center pixel of the patch belongs to. Since the four modalities T1, T2, T1Gd and Flair are co-registered, for pixels at the same location of T1Gd, T2 and Flair, we will generate three patches from each of them as the three channels of the input X for the CNN models, where T1 is dropped since the performance of our method with and without T1 do have show much difference and the model with fewer image channels have fewer parameters, which may dramatically reduce computational time in training and testing.

In the first phase of our CNN models, we extract patches in two different sizes, where the smaller one is $(18 + 3n) \times (18 + 3n)$ and the larger one is $(35 + 6n) \times (35 + 6n)$, where n denotes the kernel size, which determines the size of the two patches. The reason of this architecture choice is that we want the class prediction of a pixel to be decided by both the local features within a small region around the pixel and the larger 'context': the comparative location of the pixel inside the brain. We adopt the structure of the CNN model described in Zhao, et.al, 2017 in the first phase, where the larger input will be merged with the small input at a certain layer after passing through several convolution and pooling layers. The layer settings are shown in Figure 2. By experiment, we choose $n = 5$ to be optimal kernel size, with which we are able to mask out the whole tumor regions as the starting point for the second-phase procedure. In experiment, we pick out 100 subjects from all the five groups, and within each of the four classes randomly sampled 1000 pixels to build the two kinds of patches, so the total sample size will be around $2 \times 100 \times 4 \times 1000$. There are many more pixels belonging to the non-tumor class in brain regions, but we sampled equal number for the four classes to avoid the unbalance bias.

The motivation of adding the second-phase CNN model is that because of the large patch size we select in the first-phase CNN model, the whole tumor region could be masked out but a major issue is that many non-tumor regions will be misclassified as tumors. Specifically, the boundaries between different classes are difficult to classify, since pixels belonging to different groups along the boundaries may share similar features when the patch size is large. On the other hand, because of the memory limit, we are unable to sample enough patches for each individual subject to cover all the brain regions into the training data, which may also reduce the prediction power of the model. However, the sensitivities by the first-phase CNN model is high, meaning that most of the tumor regions could be masked out even with many misclassified non-tumor parts. Therefore, the main goal of phase 2 is to refine the segmentation results within the mask

built by phase 1. We adopt three convolution neural network with different input sizes and try to combine the prediction result by each of them. The first model still uses the same setting as the model in phase 1, but all the sampled points are within six-pixel distance from the tumor mask built by step 1. The second model has a single-size patch input, which is 13 by 13, and we combine classes 1 and 4 into a bigger class which helps us to better determine the regions of tumor core. Then within the tumor core specified by the second model, we use the third model with the patch size 7 by 7 to refine the segmentation performance of classes 1 and 4. The latter two models pay more attention to the local features.

To further improve the segmentation results, we extract the probability map, which are combined with the local features generated by ANTs to train a XGBoost model to re-do the pixel-level classification. Specifically, all the sample points used by XGBoost here will be selected from those pixels which have different prediction labels by different CNN models in phase 2. The whole framework of our method is illustrated in Figure 2. Figure 3 gives an example showing how the segmentation is improved after each refining step.

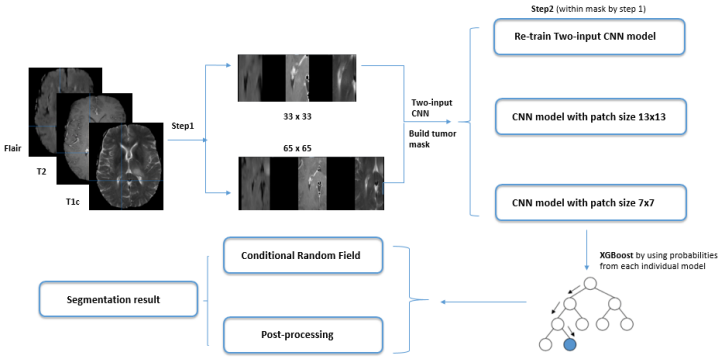


Fig. 2. Framework of TP-CNN procedure

2.3 Conditional Random Filed and Post-processing

The segmentation method shown in the previous subsection by using a two-phase CNN model is pixelwise, which does not consider the neighbor information around each pixel. Therefore, to obtain a spatially smooth segmentation results, we adopt conditional random fields (CRF) [8] as a post-processing step to refine the results from the TP-CNN procedures. Specifically, TP-CNN predicts the probability of assigning segmentation labels to each voxel, and CRF takes the prediction results and image information as input integrating with the pixel intensities and position information to globally smooth the shape of each tumor region.

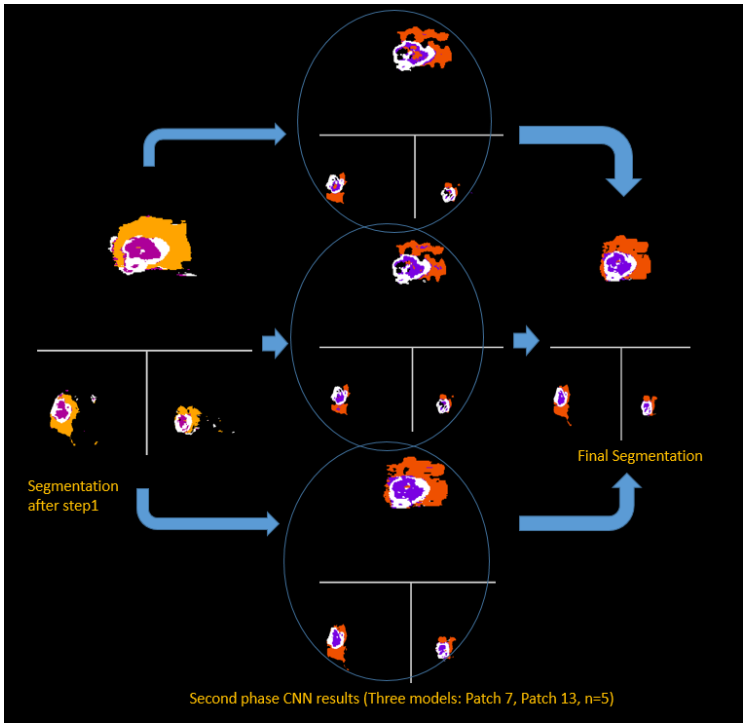


Fig. 3. Example segmentation through TP-CNN procedure

Besides CRF, we also further improve the segmentation results by applying following post-processing procedures:

1. The voxels within the tumor mask whose gray values of the corresponding flair image are less than 0.95 times the mean intensity value within that mask, and the gray value smaller than 150 in T1Gd images and 0.95 times mean intensities within the mask in the T2 images will be excluded from the tumor mask.
2. For those voxels within the enhancing tumor area whose gray value of T1Gd image are less than 110, the predicted label need to be changed to necrosis group.
3. Using FSL clustering method to segment the tumor mask into all connected components/clusters. If a cluster within which the mean gray value of both T2 and Flair images are larger than 150, voxels within that cluster will be removed from the tumor mask.
4. Voxels in clusters whose volume is less than 0.2 times the largest connected cluster volume will be reclassified as non-tumor; voxels in necrosis clusters whose volume is less than 0.01 times the largest connected neucrosis cluster volume will also be reclassified as non-tumor.

5. Fill in the holes within the tumor mask and assign voxels within the holes to necrosis area.

The performance by one-step CNN, two-phase CNN and final combined model regarding the dice-ratio of each tumor regions, corresponding sensitivities and 'Hausdorff distance' statistic are compared and shown in Table 1.

Method	Dice_ET	Dice_WT	Dice_TC	Sensitivity_ET	Sensitivity_WT	Sensitivity_TC	Specificity_ET	Specificity_WT	Specificity_TC	Hausdorff95_ET	Hausdorff95_WT	Hausdorff95_TC
Step1	0.47160	0.63001	0.52557	0.72245	0.92056	0.66802	0.98855	0.94825	0.98495	64.09213	64.88767	66.88475
Step2 n=5	0.59406	0.83233	0.66983	0.67431	0.91118	0.75906	0.99632	0.98399	0.99139	11.82644	9.39108	15.33965
Step2 Patch7	0.59829	0.83258	0.64347	0.66504	0.81070	0.69020	0.99753	0.99386	0.99419	16.17848	16.52437	22.24674
Step2 Patch13	0.63169	0.84269	0.65412	0.66929	0.81313	0.71633	0.99780	0.99530	0.99317	13.22915	12.42696	18.99704
Final (label fusion)	0.64779	0.85504	0.69934	0.74857	0.92260	0.78466	0.99671	0.98665	0.99250	10.74262	8.07866	13.98284
Final (Xgboost)	0.72217	0.85398	0.70818	0.75417	0.92754	0.76581	0.99717	0.98613	0.99408	6.6012	9.76032	13.61402

Table 1. Results comparison among step1, step2 and final combined two-phase results

3 Survival prediction by XGBoost

XGBoost [1] is short for Extreme Gradient Boosting, where the term Gradient Boosting is proposed in the paper Greedy Function Approximation: A Gradient Boosting Machine, by Friedman[2]. In brief, XGBoost is a tree-based boosting algorithm, which performs well in supervised learning problems and previous challenges. The advantage of using XGBoost here to predict the survival time for each patient is that we could throw out the unimportant features by based on AUC value and avoid over-fitting.

We extracted various imaging features of the tumor regions to predict the survival time. Nine shape features, including volume, surface, roundness and ratio between subregions, are calculated based on the segmentation results of each tumor subregion. And 10 intensity statistics features(such as maximum, minimum, median, kurtosis) 57 texture features(13 gray-level co-occurrence matrix features, 8 statistics of local binary patterns and 36 threshold adjacency statistics) and 72 wavelet transform features are separately extracted from all subregions and modalities.

Besides the global feature mentioned above, we also train a convolution neural network and extract the last dense layer as supplementary features. We take five 2-D slices in X-Y space with most pixels belonging to class 1 and 4 for each subject to generate the training data, which is augmented by flipping, rotating and pixel shifting. Then we train a CNN model using the VGG-16 setting and extract the dense layer before the softmax layer as features, which are combined with the global features and age for each subject as predictors (we are currently considering using 3D-CNN to extract the feature within the tumor regions roughly masked out based on our segmentation results because the way we use to select 2-D slides depend more on the accuracy of segmentation

predictions on validation data). We use Pearson chi-square test to analyze the marginal correlation between each of all m predictors with the survival time, and sort the m covariates by $p-$value from small to large, where the covariate with smaller $p-$value may show more significant association with the survival time, and we then use XGBoost to gradually add in covariates until AUC value will not increase for a certain number of steps. To test the performance of our model, we randomly split the training data into two parts, one of which is used to train the model and the other used to test the prediction accuracy. We do 10-fold cross-validation to test the performance of our method and use the classification evaluation by dividing the survival time into three classes using threshold 10 and 15 to compute the prediction accuracy. The highest prediction accuracy could reach 85% and the mean accuracy is around 70% (while the lowest accuracy could be 55%). The average MSE (mean square error) of the 10-times experiment is around 79680 (in days).

4 Conclusion and Discussion

Our two-phase convolution neural network together with post-processing and XGBoost is an automatic method to do brain tumor segmentation by using deep learning method. According to the performance of our method on the validation data set, our method could make an acceptable segmentation prediction for the non-tumor and three tumor regions in terms of dice-ratio and Hausdorff distance. The biggest advantage of our method is that we could achieve a very high sensitivity, which means that we do not have much tumor regions to be considered as non-tumor part. Another important characteristic of our method is that we do adopt some models with complicated structures such as Fully-CNN model, which may require a lot of computation power, but to combine the results from several simple models by using XGBoost and CRF, which is much cheaper in time consuming and machine cost. For the survival time prediction task, we use CNN model to extract the features from each of the four modalities and then use XGBoost to build the regression model, which could provide a relatively good prediction accuracy considering the small sample size we have.

We have to admit that there are at least two weaknesses of our method. First, the pixel-based prediction procedure may result in some singular points with a misclassified label within a regions having correct predictions, which may cause a huge reduce in the tumor core dice-ratio. Second, compared with FCNN or other slice-based method, our TP-CNN model could not acquire a smooth segmentation results, even though it can be improved by doing post-processing. Despite these potential disadvantages, the simple structure and the combining idea of our method will still be competitive when there are limited computing power and time, to build a reasonable results in segmenting tumor regions and predicting survival time of patients.

References

1. Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.
2. Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
3. Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2015.
4. Bakas S, Sotiras A Akbari H, Rozycki M Bilello M, Kirby J, Freymann J, Farahani K, and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *The Cancer Imaging Archive*, 2017.
5. Bakas S, Sotiras A Akbari H, Rozycki M Bilello M, Kirby J, Freymann J, Farahani K, and Davatzikos C. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection. *The Cancer Imaging Archive*, 2017.
6. Bakas S, Akbari H, Bilello M Sotiras A, Rozycki M, Freymann JB Kirby JS, Farahani K, and Davatzikos C. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Nature Scientific Data*, 2017.
7. Xiaomei Zhao, Yihong Wu, Song, Zhenye Guidong, Li, Yazhuo Zhang, and Yong Fan. A deep learning model integrating fcnns and crfs for brain tumor segmentation. *arXiv preprint arXiv:1702.04528*, 2017.
8. Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1529–1537, 2015.

A Multi-pathway 3D Dilated Convolutional Neural Network for Brain Tumor Segmentation

Jin Zhu[a], Duo Wang[a], Zhongzhao Teng[b], Pietro Liò[a]

[a] *Computer Laboratory, University of Cambridge, Cambridge, CB3 0FD, UK*
[b] *Department of Radiology, University of Cambridge, Cambridge, CB2 0QQ, UK*

Abstract

We developed a Multi-pathway 3D dilated Convolutional Neural Network (CNN) for BRATS Brain Tumour Segmentation challenge. Our CNN segment image volumes in a patch-wise manner. We extended a Vanilla Feed-forward 3D CNN in two useful ways. Firstly instead of stacking all four Magnetic Resonance (MR) sequences into images of 4 channels, we apply multi-pathway architectures with each pathway processing one MR sequence. Secondly we added a contextual information stream that process contextual voxels around a patch using dilated convolution. Our 3D CNN achieved a good performance with the all training and validation data from BRATS 2017(Dice_ET=0.66, Dice_WT=0.85, and Dice_TC=0.70).

Keywords: Brain Tumor Segmentation, Dilated Convolution, 3D CNN

1. Introduction

We developed a Multi-pathway 3D dilated Convolutional Neural Network (CNN) for BRATS Brain Tumour Segmentation challenge[1, 2, 3, 4]. Previous works [5, 6, 7] have applied standard feed-forward 3D CNNs on medical image segmentation and obtained promising results. In this work, We extended a vanilla Feed-forward 3D CNN with Multi-pathway architecture and dilated convolution in order to improve performance. Our CNN process MR volumes in a patch-wise manner, namely we divide a brain MR volume into a dense grid of smaller 3D patches, and train CNN to predict the segmentation maps in the central regions of patches. We are also inspired by

Email address: jz426@cam.ac.uk (Jin Zhu)

Kamnitsas et al[5] to include additional neural network stream that process contextual information surrounding the patch. However we applied dilated convolution which removes the need to downsample surrounding information, as performed in [5].

2. Method

Figure 1 shows an overview of our architecture. In our 3D CNN we have 2 input information processing streams, which process patch of size 27^3 and contextual information patch of size 39^3 respectively. The CNN is trained to segment the central 9^3 region of the patch. Inside each information processing stream, instead of stacking all four Magenetic Resonance (MR) sequences into images of 4 channels, we apply multi-pathway architectures with each pathway processing one MR sequence. Figure 2 illustrates the multiple pathways within each processing stream. Multi-pathway allows each stream to learn decoupled and independent features from each MR sequences.

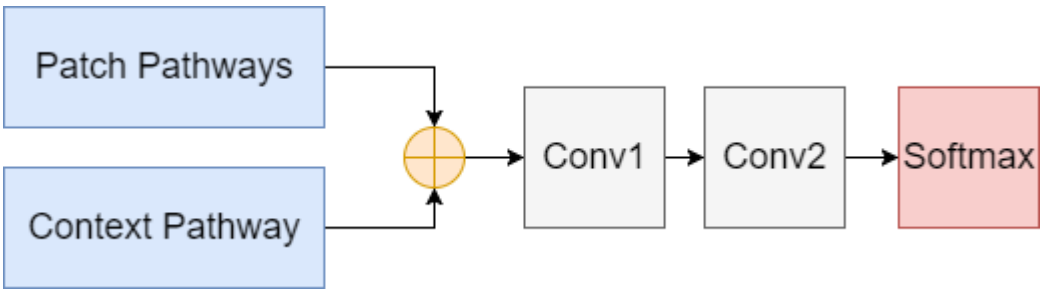


Figure 1: Overall 3D CNN architecture

Inside each pathway there are 9 3D convolutional layers. We also implemented residual connections as developed by He et al[8]. All convolutional layers have valid padding and $3\times3\times3$ convolutional kernels. For all pathways, the number of channels in each layer are $30-30-30-40-40-40-40-50-50$. For residual connections, the merging of two feature maps requires same feature map sizes except the channel dimension. Therefore we cropped the central part of the larger skip-passed feature maps to ensure the same feature map size. We also used concatenation instead of element-wise summation for feature map merging because concatenation is allows more flexibility of feature map merging.

For the contextual information, we use dilated convolution[9] to effectively increase the receptive field size without using any down-sampling or pooling

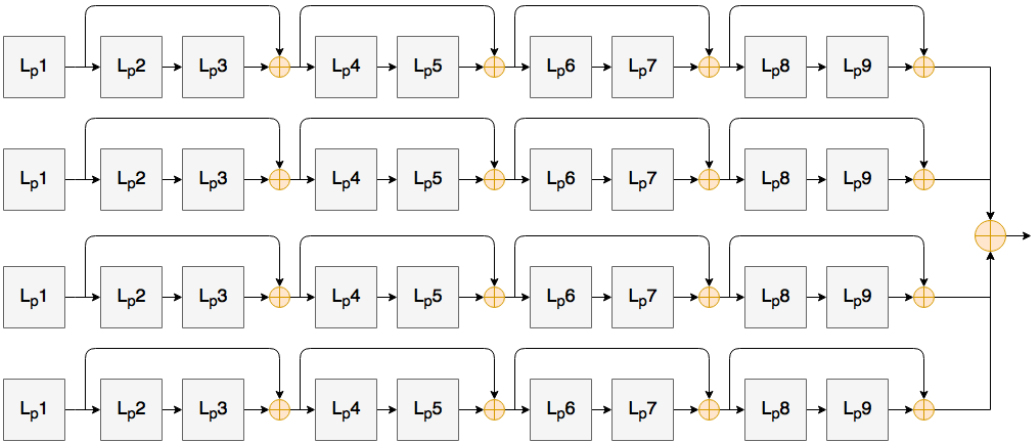


Figure 2: Pathways for each Magnetic Resonance (MR) sequence with Residual connections

layers. Figure 3 compares standard convolution with dilated convolution. Dilated convolution allows stride spacing between input positions, thereby increase the receptive field without increasing the number of kernel parameters. Therefore dilated convolution is well suited for modeling contextual information around each voxel. In our contextual pathways we used dilated convolution with a dilation factor of 2 except for the first convolutional layer.

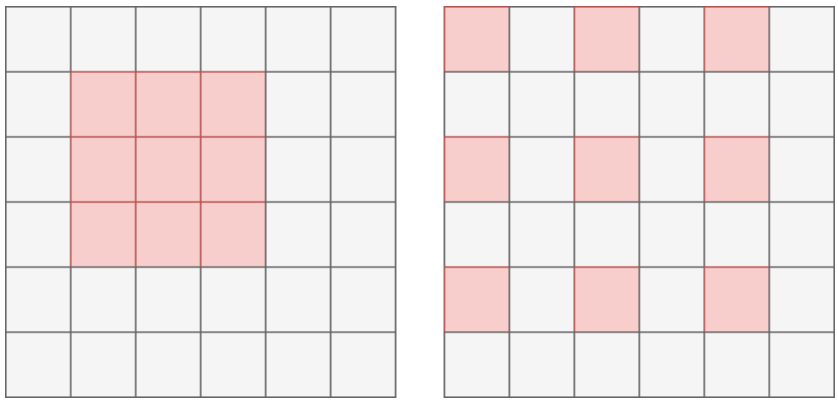


Figure 3: Comparison of standard convolution (left) and dilated convolution (right).

Our 3D CNN is trained end-to-end with Stochastic Gradient Descent Back-Propagation algorithm. We used RMSProp [10] optimizers with initial learning rate of 0.001 and batch size of 64.

We trained our 3D CNN on all 285 patients' images from BRATS 2017[1, 2, 3, 4] training data set. From images of each patient, the same number of patches were generated from foreground and background, in order to alleviate the class-imbalance in the original images. Specifically, 500 patches of the background and another 500 patches of foreground are randomly generated, namely 1000 patches of each patient in total. These amount of patches could lead to better estimation with reasonable memory and computation cost. In the training process, we loaded patches from 50 patients for each sub epoch, and shuffle all the patients at the start of each epoch, to avoid the over-fitting on a specific cluster of patients.

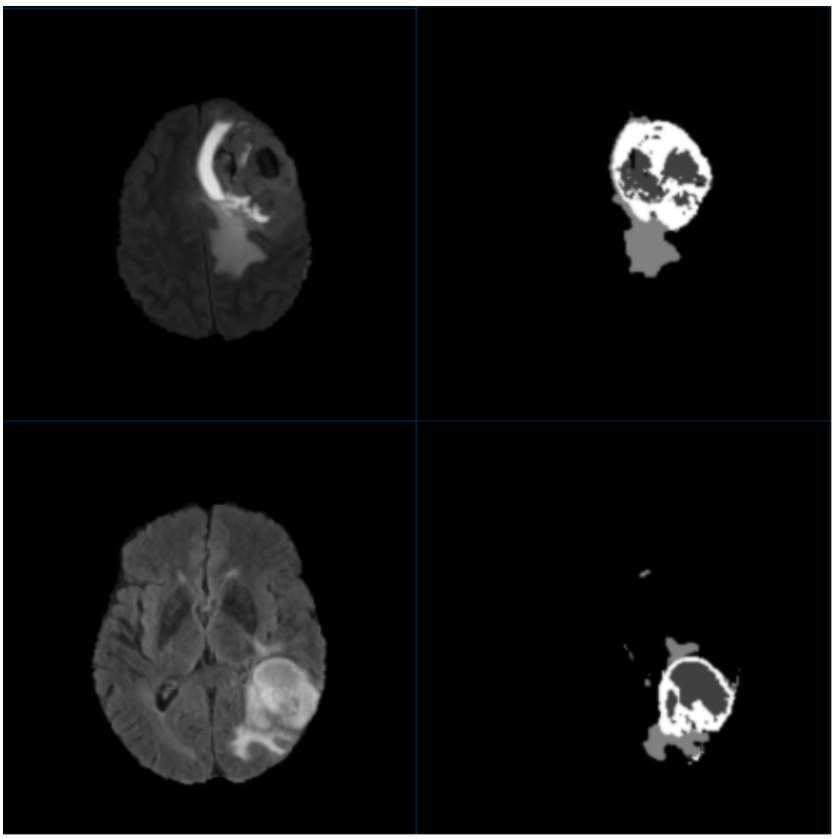


Figure 4: Flair (left) and segmentation (right).

3. Results

Our 3D CNN are validated with images of all 46 patients from BRATS 2017[1, 2, 3, 4] validation data set on CBICA Image Processing Portal[11]. Figure 4 shows two examples of our segmentations comparing with the original flair images, Table 1 shows the DICE values, and Table 2 shows the Hausdorff distances.

Table 1: Dice values

Dice_ET	**Dice_WT**	**Dice_TC**
0.6645	0.8516	0.6983

Table 2: Hausdorff distances

Hausdorff95_ET	**Hausdorff95_WT**	**Hausdorff95_TC**
15.29	17.47	19.32

4. Discussion and Conclusion

In this paper we proposed a multi-pathway 3D CNN architecture for medical image segmentation. The four separate pathways for all MRI modalities can flexibly learn the specific textual features of each sequence. The dilated convolutional pathway can use contextual information effectively.

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (brats), IEEE transactions on medical imaging 34 (2015) 1993–2024.

[2] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, C. Davatzikos, Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features, Nature Scientific Data (2017[In Press]).

[3] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, C. Davatzikos, Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection, The Cancer Imaging Archive (2017).

[4] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, C. Davatzikos, Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection, The Cancer Imaging Archive (2017).

[5] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, B. Glocker, Deepmedic for brain tumor segmentation, in: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer, pp. 138–149.

[6] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 3D Vision (3DV), 2016 Fourth International Conference on, IEEE, pp. 565–571.

[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 424–432.

[8] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

[9] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint arXiv:1511.07122 (2015).

[10] T. Tieleman, G. Hinton, Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude, COURSERA: Neural networks for machine learning 4 (2012) 26–31.

[11] C. for Biomedical Image Computing, U. o. P. Analytics, Image processing portal - https://ipp.cbica.upenn.edu/, A web accessible platform for imaging analytics (2015).