BoostMIS: Boosting Medical Image Semi-supervised Learning with Adaptive Pseudo Labeling and Informative Active Annotation

Wenqiao Zhang [1]   Lei Zhu [1]   James Hallinan [2, *]   Shengyu Zhang [3]   Andrew Makmur [2]

Qingpeng Cai [1] Beng Chin Ooi [1]

[1] National University of Singapore, Singapore, [2] National University Hospital, Singapore, [3] Zhejiang University, China

`wenqiao@nus.edu.sg`, `e0203764@u.nus.edu`, {`james_hallinan,andrew_makmur`}`@nuhs.edu.sg`,
`sy_zhang@zju.edu.cn`, {`qingpeng,ooibc`}`@comp.nus.edu.sg`

Abstract

In this paper, we propose a novel semi-supervised learning (SSL) framework named BoostMIS that combines adaptive pseudo labeling and informative active annotation to unleash the potential of medical image SSL models: (1) BoostMIS can adaptively leverage the cluster assumption and consistency regularization of the unlabeled data according to the current learning status. This strategy can adaptively generate one-hot "hard" labels converted from task model predictions for better task model training. (2) For the unselected unlabeled images with low confidence, we introduce an Active learning (AL) algorithm to find the informative samples as the annotation candidates by exploiting virtual adversarial perturbation and model's density-aware entropy. These informative candidates are subsequently fed into the next training cycle for better SSL label propagation. Notably, the adaptive pseudo-labeling and informative active annotation form a learning closed-loop that are mutually collaborative to boost medical image SSL. To verify the effectiveness of the proposed method, we collected a metastatic epidural spinal cord compression (MESCC) dataset that aims to optimize MESCC diagnosis and classification for improved specialist referral and treatment. We conducted an extensive experimental study of BoostMIS on MESCC dataset. The experimental results verify our framework's effectiveness with a significant improvement over various state-of-the-art methods. Our work will be available at github [1].

1. Introduction

In recent years, the development of deep learning brings prosperity to the field of computer vision [26, 27, 29, 29, 30, 32, 54, 57–59] and applied in industrial application [22, 28,
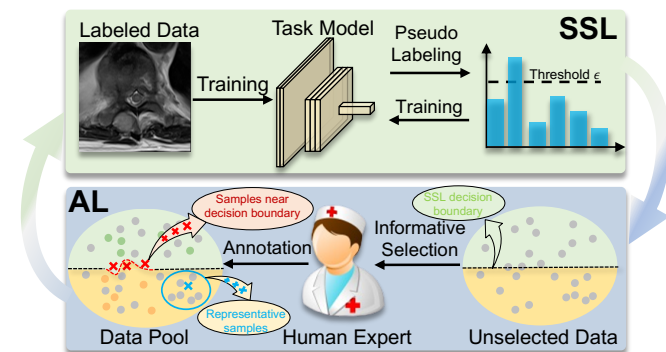

Figure 1. **An example of how does AL facilitates medical image SSL.** Orange and green dots represent two different classes in the data pool. Grey dots represent the unlabeled data. Red and blue marks indicate the annotation candidates, which can smooth the decision boundary and propagate representative label information to unlabeled data, respectively.

31, 46, 52, 56]. For instance, a significant trend in healthcare is to leverage the sizable well-annotated dataset using deep learning technology [7, 19, 40, 49, 55, 62] to achieve automated medical image analysis. However, annotating such a medical dataset is an expert-oriented, expensive, and time-consuming task. The situation is further exacerbated by large numbers of new raw medical images without labels. These are generated on a daily basis in clinical practice and there is an increasing demand for the exploitation of such data. Therefore, semi-supervised learning (SSL) that goes beyond traditional supervised learning by exploiting a limited amount of labeled data together with a huge amount of raw unlabeled data is gaining traction. Among the existing SSL methods, pseudo-labeling [25] is a specific variant where model predictions are converted to pseudo-labels, which is often used along with confidence-based thresholding that retains unlabeled examples only when the classifier is sufficiently confident. Some of them [4, 24, 42, 50], have

---

[1] `https://github.com/wannature/BoostMIS`
[*] Corresponding Author.

achieved great success on image classification datasets such as CIFAR-10/100 [23], demonstrating the potential value of utilizing unlabeled data.

Unfortunately, these natural image-based pseudo-labeling SSL methods may not satisfactorily resolve the medical imaging problems, which can be summarized as two key aspects: (1) **Poor Data Utilization**. Medical images (e.g., CT, MRI) are highly similar at the pixel level, making them hard to classify, even by human beings. That is, the pseudo-labeling may only produce a few pseudo-labels with high confidence above a fixed threshold for unlabeled medical images, especially at the early stage of the training process. It suffers from the poor utilization of data problem that is a considerable amount of unlabeled data are ignored. (2) **Missing Informative Sample**. The ignored unlabeled data with prediction confidence below the pre-defined threshold may have informative data (e.g., samples near the boundary of clusters, representative samples in the unlabeled distribution space) that can further improve the model's performance. Based on the aforementioned insights, a meaningful optimization goal of Medical Image SSL is to explore an effective learning approach to tap onto unlabeled medical data deeply.

Inspired by another alternative to leverage the unlabeled data, that is active learning (AL) [38], which aims to select the most informative samples to maximize the model performance with minimal labeling cost. AL seems to be a tempting approach to deal with the aforementioned problems in pseudo-labeling SSL methods. As illustrated in Figure 1, SSL already results in the embodiment of knowledge from selected unlabeled data with high confidence by pseudo-labeling. The judicious AL selection can reflect the value of additionally informative samples in unselected data on top of such embodied knowledge. Those informative cases can assist the SSL model to propagate extra valuable knowledge to unlabeled data, thereby improving the unlabeled data utilization for better SSL. In fact, AL and SSL are naturally related with respect to their common goal, i.e., utilizing unlabeled samples. From the perspective of machine learning, utilizing the correlation and similarity among related learning methods can be regarded as a form of inductive transfer. It can introduce the *inductive bias* [3] to make the combined learning method prefer the correct hypothesis, thereby improving the performance.

Overall, we propose a novel framework BoostMIS that aims to boost medical image SSL, which comprises the following parts: (1) *Medical Image Task Model*. The task model (medical image classification in this paper) is first trained through weakly-augmented medical images with supervised labels. (2) *Consistency-based Adaptive Label Propagator*. This module propagates the label information for unlabeled data using both pseudo-labeling and consistency regularization. Since the learning ability and performance of the models are different in each training stage, we define a dynamically adaptive threshold based on the current learning status to produce pseudo-labels for better unlabeled data utilization. Then the consistency regularization forces the model to produce the same prediction of weakly-augmented and strongly-augmented data as the regularization criterion for better generalization capacity. (3) *Adversarial Unstability Selector*. To boost SSL by AL, we introduce the virtual adversarial perturbation to choose the unstable samples that lying on the clusters' boundaries as annotation candidates. Specifically, the SSL model would be weaker or even inconsistent on the clusters' boundaries, the adversarial unstability selector could identify samples near the boundaries by measuring the inconsistency between samples and the corresponding virtual adversarial examples. (4) *Balanced Uncertainty Selector*. To further identify the informative cases in the unlabeled pool, we use the density-aware entropy of the SSL model to evenly select the samples with high uncertainty in each predicted class as the complementary set to balance the subsequent training . The union of adversarial unstability samples and balanced uncertainty samples will be the final annotation candidates to expand the labeled pool. In summary, our proposed four modules work in a loop and make progress together to boost the medical image SSL.

To summarize, the major contributions of our paper are as follows:

- To the best of our knowledge, we are the first to incorporate AL into SSL to unleash the potential of unlabeled data for better medical image analysis.

- We propose adaptive pseudo-labeling and informative active annotation that reasonably leverage the unlabeled medical images and form a closed-loop structure to boost the medical image SSL.

- We collected a metastatic epidural spinal cord compression (MESCC) dataset for method development and extensive evaluation, which aims to optimize MESCC diagnosis and classification for improved specialist referral and treatment.

- The consistent superiority of the proposed BoostMIS is demonstrated on the MESCC dataset that outperforms existing SSL methods by a large margin.

2. Related Work

Semi-supervised Learning. Semi-supervised learning (SSL) is concerned with learning from both unlabeled and labeled samples, where typical methods ranging from entropy minimization [16,21,47], pseudo-labeling (also called self-training) [1, 11, 25, 51] and consistency regularization [2, 10, 45, 50] to improve model's performance. Re-

cent work FixMatch [42] that proposes to combine pseudo-labeling and consistency regularization simultaneously has shown promising performance in semi-supervised tasks. FixMatch [42] achieves SoTA results by combining these techniques with weak and strong data augmentations and uses cross-entropy loss as the regularization criterion. However, FixMatch suffers from the poor utilization of data problem that is it ignores a considerable amount of unlabeled data with confidence below the labeling threshold. Therefore, an important goal of Medical Image SSL is to utilize the unlabeled data appropriately.

Semi-supervised Learning in Medical Image Analysis. Due to the difficulty of data annotation, SSL is widely used in medical imaging processing, e.g., medical image detection [48, 60], classification [18, 34, 39] and segmentation [36, 61]. [48] proposes an SSL method for 3D medical image detection that uses a generic generalization of the focal loss to regularize the SSL training. [61] improves the performance of disease grading and lesion segmentation by SSL. [36] uses an attention-based SSL learning method to boost the performance of medical image segmentation. However, these previous works are limited to SSL, which may fail to mine the potential of unlabeled data deeply.

Semi-supervised Active Learning. Recently, few works are proposed that combine SSL and AL. [12, 41] generally adopt VAE-GAN structure in the SSL manner to learn latent representations of labeled and unlabeled samples through a min-max game, and then conduct AL-based on learned semantic distribution. The latest works [15, 17, 43] make the combination based on prediction inconsistency given a set of data augmentations. However, all of these methods are designed for natural image classification tasks, while this paper focuses on the complicated medical image SSL task.

3. Method

This section describes the proposed medical image semi-supervised (SSL) framework BoostMIS. We will describe each module and introduce the training strategy.

3.1. Problem Formulation

Before presenting our method, we first introduce some basic notions and terminologies. We consider a labeled pool of training samples $\mathcal{X}$ with the corresponding labels $\mathcal{Y}$, and an unlabeled pool of training samples $\mathcal{U}$. The goal is to use all paired samples $(\mathcal{X}, \mathcal{Y})$ and unlabeled samples $\mathcal{U}$ to train the proposed medical image SSL framework BoostMIS. As shown in Figure 2, we use the following training pipeline $M(.; \theta)$ to show how the BoostMIS works:

$$\underbrace{M((\mathcal{X}, \mathcal{Y}, \mathcal{U}); (\theta_{\mathcal{S}}, \theta_{\mathcal{A}}))}_{\text{BoostMIS}} = \underbrace{T((\mathcal{Y}, \mathcal{P}^s) | (\mathcal{X}, \mathcal{U}^s)); \theta_{\mathcal{S}})}_{\text{SSL Training}}$$
$$\leftrightarrow \underbrace{S((\mathcal{C} | \mathcal{U}^u); (\theta_{\mathcal{S}}, \theta_{\mathcal{A}}))}_{\text{AL Annotation}} \quad (1)$$

$T(\cdot; \theta)$ is the task model (medical image classification in this paper) with SSL and $S(\cdot; \theta)$ is the annotation candidate selection through active learning (AL). **In SSL aspect**, $T(\cdot; \theta)$ is trained by labeled samples $\mathcal{X}$ with weak augmentation in *Medical Image Task Model* and propagates the label information to unlabeled samples $\mathcal{U}$ in *Consistency-based Adaptive Label Propagator*. The unlabeled samples with their posterior probability above the adaptive threshold are selected as training examples $\mathcal{U}^s$ with pseudo labels $\mathcal{P}^s$, which combine $\mathcal{X}$ to train $T(\cdot; \theta)$ with augmentation-aware consistency regularization. **In AL aspect**, once the SSL training process is finished, we feed unselected samples $\mathcal{U}^u$ with perturbation to $T(\cdot; \theta)$ to generate the adversarial samples in *Adversarial Unstability Selector*. After that, we select out top-K samples with the biggest KL-divergence between unselected samples $\mathcal{U}^u$ and their adversarial samples. To further identify the informative cases in the unlabeled pool, we use the entropy of $T(\cdot; \theta)$ to evenly select the top-K samples with high uncertainty in each predicted class as the complementary set to balance the subsequent training. The final candidates $\mathcal{C}$ composed of parallelly selected N_A samples ($\leq 2K$ due to the intersection of candidates) are provided to human experts for annotation. As a consequence, the sizes of the labeled pool and unlabeled pool will be updated to perform better SSL. The loop will be repeated until the task model's performance meets requirements or the annotation budget has run out.

3.2. Medical Image Task Model

For task model development, we train a convolutional prototypical network [53] with Resnet50 [20] as its backbone. Given labeled data $\mathcal{X} = \{\mathbf{x}_i |_{i=1}^{N_l}\}$ with weak augmentation (e.g., using only flip-and-shift data augmentation) and their corresponding labels $\mathcal{Y} = \{\mathbf{y}_i |_{i=1}^{N_l}\}$, N_l is the number labeled samples. The task model is trained via the mapping function $F : \mathcal{X} \rightarrow \mathcal{Y}$, with the cross-entropy loss ℓ_s minimized through supervised label information.

$$\ell_s(\theta_{\mathcal{S}}) = \frac{1}{N_l} \sum_{i=1}^{N_l} D_{ce}(\mathbf{y}_i, P_m(\mathbf{p}_i | A_w(\mathbf{x}_i))) \quad (2)$$

where $P_m(\cdot | \cdot)$ represents the posterior probability distribution of the task model. $D_{ce}(\cdot, \cdot)$ denote the cross-entropy between two probability distributions. $\mathbf{p}_i$ is the predicted label. $A_w(.)$ is a weak augmentation function.

3.3. Consistency-based Adaptive Label Propagator

Inspired by FixMatch [42], SSL label propagator in BoostMIS is a combination of two approaches: pseudo-labeling and consistency regularization. The pseudo-label of an unlabeled weakly-augmented image is computed based on the task model's prediction when its confidence is above the defined threshold, and the pseudo-label is

Figure 2. **Overview of BoostMIS.** It consists of four modules: (a) *Medical Image Task Model* trains the basic task model with weakly-augmented labeled data by supervised learning. (b) *Consistency-based Adaptive Label Propagator* propagates the label information to unlabeled samples by pseudo-labeling with dynamically adjusted threshold and augmentation based consistency regularization. (c) *Adversarial Unstability Selector* estimates KL-Divergence of unselected data and their corresponding virtual adversarial examples with perturbation to select the annotation candidates. (d) *Balanced Uncertainty Selector* evenly selects the samples with high uncertainty as the complementary annotation set to balance the subsequent SSL training.

enforced against the task model's output for a strongly-augmented image. However, the learning ability and performance of the models are different in each training step. Thus, we introduce adaptive threshold (**AS**) ϵ_t that can be dynamically adjusted at different learning statuses to utilize unlabeled data better. Formally speaking, in each iterative training step t, the AS ϵ_t for the class with fewer samples is defined as bellow:

$$\epsilon_t = \begin{cases} \alpha \cdot Min\{1, \frac{Count_{\epsilon_t}}{Count_{\epsilon_{t-1}}}\} + \frac{\beta \cdot N_A}{2K}, & \text{if } t < T_{max} \\ \alpha + \beta, & \text{otherwise} \end{cases} \quad (3)$$

where coefficients α and β are pre-defined thresholds. T_{max} is a defined value of the iterative learning step. N_A is the number of active learning selected annotation candidates. $Count_{\epsilon_t}$ is a counting function to estimate the learning status at time step t. Given $\mathcal{U} = \{\mathbf{u}_i|_{i=1}^{N_u}\}$ (N_u is the number of unlabeled data), $Count_{\epsilon_t}$ computed as follows:

$$Count_{\epsilon_t} = \sum_{i=1}^{N_u} \mathbb{1}(P_m(\mathbf{p}_i|A_w(\mathbf{u}_i)) > \alpha + \beta) \quad (4)$$

$Count_{\epsilon_t}$ records the number of unlabeled data $\mathcal{U}$ whose prediction scores are above the high confidence threshold $\alpha + \beta$. We argue that the learning effect of the SSL model can be reflected by the number of samples whose predictions are above the threshold. The AS ϵ_t can be dynamically adjusted based on the current learning status and amount of AL informative selection to encourage the better utilization of unlabeled data (prediction of the class with fewer samples) until the iteration step exceeds T_{max}.

Thus, we feed the unlabeled data set $\mathcal{U}$ with weak augmentation into the task model. The predictions of selected samples $\mathcal{U}^s = \{\mathbf{u}_i^s|_{i=1}^{N_u^s}\}$ are converted to one-hot pseudo-labels $\mathcal{P}^s = \{\mathbf{p}_i^s|_{i=1}^{N_u^s}\}$ when their predicted confidences are above the adaptive threshold ϵ_t, where N_u^s denote the number of SSL selected unlabeled samples. Then we introduce

the augmentation-aware consistency regularization, which computes the model's prediction for a strong augmentation of the same image. The model is trained to enforce its prediction on the strongly-augmented image to match the pseudo-label via a cross-entropy loss.

$$\ell_u(\theta_\mathcal{S}) = \frac{\mu}{N_u^s} \sum_{i=1}^{N_u^s} D_{ce}(P_m(\mathbf{p}_i^s | A_w(\mathbf{u}_i^s)), P_m(\mathbf{p}_i^s | A_s(\mathbf{u}_i^s))) \quad (5)$$

where μ is a fixed scalar hyperparameter denoting the relative weight of the unlabeled loss, $A_s(.)$ represents the strong augmentation function RandAugment [9].

3.4. Adversarial Unstability Selector

Above SSL model label propagator can deliver the label information from labeled data to pseudo-labeled samples. However, there are many informative cases among the unselected samples. These samples fail to acquire enough label information, and the task model has a vague understanding of the local data distribution of these samples. In this sense, selecting informative samples in unselected data for further annotation is more valuable, which can help the SSL model smooth the decision boundary and propagate the representative label information to unlabeled data. We divide the unselected samples $\mathcal{U}^u = \{\mathbf{u}_i^u\}|_{i=1}^{N_u^u}$ (N_u^u is the number of unselected samples) into two categories: *unstable* and *uncertain*. In this section, we aim to find the unstable samples by introducing the adversarial unstability selector (**AUS**).

The AUS estimates the stability of the model's prediction on samples by calculating the inconsistency between prediction on samples and corresponding virtual adversarial samples. Specifically, given an unselected sample $\mathbf{u}_i^u$ ($i \leq N_u^u$), its representation $\mathbf{r}_i^u$ is extracted from the last layer of task model and prediction is $\mathbf{p}_i^u$. We feed $\mathbf{r}_i^u$ and $\mathbf{p}_i^u$ simultaneously into the generator to get virtual adversarial perturbation $\mathbf{r}_i^p$. After that, the perturbed representation $\bar{\mathbf{r}}_i^u = \mathbf{r}_i^u + \mathbf{r}_i^p$ is fed into the task model to get its perturbed prediction $\bar{\mathbf{p}}_i^u$. The perturbation is formulated as:

$$\mathbf{r}_i^p = \underset{\Delta r, ||\Delta r|| \leq \tau}{\arg\max} Div(P_m(\mathbf{p}_i^u | \mathbf{r}_i^u), P_m(\bar{\mathbf{p}}_i^u | \mathbf{r}_i^u + \Delta r)) \quad (6)$$

where τ a hyper-parameter that indicates the maximum perturbation step, Δr is a randomly sampled vector, $Div(\cdot, \cdot)$ is a non-negative function measuring the divergence between two distributions, and we use KL divergence in practice.

The evaluation of $\mathbf{r}_i^p$ cannot be performed using Eq.(6) because the gradient of $Div(\cdot, \cdot)$ with respect to r is always 0 at $r = 0$. To solve this problem, [35] proposed to approximate $\mathbf{r}_i^p$ using the second order Taylor expansion and solve the $\mathbf{r}_i^p$ via the power iteration method. Specifically, we can approximate $\mathbf{r}_i^p$ by repeatedly applying the following update n_t times (n_t is a hyper-parameter):

$$\mathbf{r}_i^p \leftarrow \tau \overline{\nabla_{\Delta r} Div(P_m(\mathbf{p}_i^u | \mathbf{r}_i^u), P_m(\bar{\mathbf{p}}_i^u | \mathbf{r}_i^u + \Delta r))} \quad (7)$$

the computation of $\nabla_{\Delta r} Div$ can be performed with one iteration of backpropagation for the neural network. Once the $\mathbf{r}_i^p$ is solved, we can estimate the variance $Var(\cdot; \cdot)$ of unlabeled sample $\mathbf{r}_i^u$ by measuring the KL-divergence:

$$Var(\mathbf{r}_i^u; \theta_A) = Div(P_m(\mathbf{p}_i^u | \mathbf{r}_i^u), P_m(\bar{\mathbf{p}}_i^u | \mathbf{r}^u + \mathbf{r}_i^p)) \quad (8)$$

Finally, AUS selects top-K samples with the largest variance from unstable samples as an initial recall set of the AL annotation candidates. These unstable samples often lie on the clusters' boundary that can smooth the decision boundary for the SSL model to output more correct predictions.

3.5. Balanced Uncertainty Selector

Besides unstable samples, there are abundant uncertain samples among unselected data that still hold low prediction confidence are also informative to the task model. To further find the uncertain samples to boost the SSL model, we introduce the balanced uncertainty selector (BUS), which evenly chooses the unselected samples with high uncertainty in each predicted class. These selected samples can significantly reduce model uncertainty and balance the subsequent SSL. In practice, we use the task model's entropy to estimate the uncertainty of samples, and then evenly select top-K samples with the largest entropy for human annotation. The entropy $Ent'(\cdot; \cdot)$ for the unlabeled sample $\mathbf{u}_i^u$ can be calculated with the following formulation:

$$Ent'(\mathbf{u}_i^u; \theta_S) = \sum_{c \in C} P_m(\mathbf{p}_i^c | A_w(\mathbf{u}_i^u)) log P_m(\mathbf{p}_i^c | A_w(\mathbf{u}_i^u)) \quad (9)$$

where C is the set of possible classes in the dataset. $\mathbf{p}_i^c$ indicates the predicted class of $\mathbf{u}_i^u$ belonging to the c-th class.

The above entropy-based formula only estimates information certainty of each sample and fails to take the distribution relations among samples into account. As a result, the metric may run into the risk of selecting some outliers or unrepresentative samples in the distribution space. To alleviate this issue, we re-weight the uncertainty metric with a representativeness factor and explicitly consider the data distribution. We denote this density-aware uncertainty as:

$$Ent(\mathbf{u}_i^u; \theta_S) = Ent'(\mathbf{u}_i^u; \theta_S)(\frac{1}{M} \sum_{j=1}^{M} Sim(\mathbf{u}_i^u, \mathbf{u}_j^u)) \quad (10)$$

where $Sim(\cdot, \cdot)$ estimates the cosine similarity of $\mathbf{r}^u$ and its nearest M samples in the distribution space. Each class samples $\lfloor \frac{K}{N_c} \rfloor$ images as uncertain annotation candidates.

Finally, the union of unstable and uncertain samples composes the final annotation candidates to expand the labeled pool, which brings informative cases to improve the subsequent SSL.

4. Experiments

We evaluate the effectiveness of our proposed BoostMIS in two datasets that are the metastatic epidural spinal cord

compression (MESCC) dataset, followed by a discussion of BoostMIS's property with controlled studies.

4.1. Dataset and Setting

Dataset. We collected a metastatic epidural spinal cord compression (MESCC) dataset that aims to aid MESCC diagnosis and classification. It consists of 7,295 medical images collected from adult patients ($\geq$ 18 years) with a random selection of studies across different MRI scanners (GE and Siemens 1.5 and 3.0T platforms). When drawing each bounding box, the annotating radiologist employed the Bilsky classification [5] that consists of patients with six types of grades that are b0, b1a, b1b, b1c, b2, and b3. In general, we also can classify these grades into low-grade (i.e., b0, b1a, and b1b) and high-grade (i.e., b1c, b2, and b3) Bilsky MESCC. Specifically, patients with low-grade are amenable to radiotherapy, while the patients with high-grade are more likely to require surgical decompression. The dataset is randomly split into 70% (5207) / 15% (1011) / 15% (1077) for the training/validation/test sets, respectively. More details of the MESCC dataset can be seen in the appendix.

Implementation Details. We employ the Wide ResNet-50 [20] as the backbone model to conduct the medical image classification task. To train the AL-based SSL model with a balanced initialization, we set up the initial labeled pool (10% data) by uniformly sampling in each class. The initial labeled pool is a subset of the training set (30% data) for pure SSL models and other data are randomly sampled. We train BoostMIS with a standard stochastic gradient descent (SGD) [6] optimizer in all experiments. Besides, we use an identical set of hyperparameters (μ=1, Mo=0.9, α=0.9, β=0.05, T_{max}=50,000, τ=1, B=64, AC=30, IP=10%, SS=30%)[2] across all datasets.

Metrics. To evaluate the medical image classification task, we employ the following metrics: Accuracy (ACC), Macro Precision (MP), Macro F1 score (MF1) [8], Macro recall (MRC) [8] and Error Rate (ER).

Comparison of Methods. For quantifying the efficacy of the proposed framework, we use several baselines for performance comparison according to different aspects. On the SSL aspect, we consider three SSL baselines: **P-Labeling** [25], **MixMatch** [4], **FixMatch** [42]. On the AL aspect, we choose the following recent methods as baselines: **R-Labeling** [13], **DBAL** [14], **VAAL** [41] and **CSAL** [15]. More details of baselines are in the appendix.

Further, we also investigate effectiveness of each individual component in BoostMIS via evaluating the following variants: (1) **BoostMIS (-AS)** uses a fixed confidence threshold ($\alpha + \beta$) to propagate the label information to unlabeled data. (2) **BoostMIS (-BUS)** only considers the incon-

sistency in AL annotation to choose the unselected medical images. (3) **BoostMIS (-VAR)** selects the AL annotation candidates without Adversarial Unstability Selector. (4) **SSL+RS** is a simplified version of BoostMIS, which consists of the SSL model in BoostMIS and applies a random sampling strategy.

4.2. Experimental Results

The overall quantitative results of our framework and baselines on the test set of the MESCC dataset are listed in Table 1. From this table, we can observe that our model consistently achieves superiority of performance from 15% to 30% labeled data. In particular, under 30% labeled data, BoostMIS outperforms other baselines in terms of accuracy by a large margin of **2.88% $\sim$ 10.21%** (two-grading) and **2.70% $\sim$ 18.31%** (six-grading) for MESCC medical image classification task. It is worth noting that almost all AL-based SSL models can improve the model performance by employing an appropriate annotation strategy, except for the R-Labeling. This phenomenon is reasonable as the randomly selected annotation candidates are not informative enough for the SSL model. In addition, compared to the pure SSL model Fixmatch, BoostMIS brings a huge reduction in **10% annotation cost (500 samples)** to reach the similar performance (BoostMIS: 91.46%, Fixmatch: 91.09%). In other words, benefiting from the adversarial unstability selector (AUS) and balanced uncertainty selector (BUS), BoostMIS can provide more informative samples for the task model with less annotation cost and a significant performance improvement.

We also report the Macro Precision, the Macro F1 score and the Macro Recall of the compared models in Table 2. Our BoostMIS once again achieves the best score in terms of all the metrics. Notably, the results of AL-based SSL models fail to maintain consistent superiority compared with the global accuracy in Table 1. For instance, the Macro Recall scores of AL-based SSL baselines are inferior to Fix-Match. This is because those AL models tend to select annotation candidates with class b0/low-grade, which has the largest data proportion in MESCC dataset. The imbalanced annotation candidates without balanced supervised label information may lead to the misclassification of other class data. Meanwhile, the balanced uncertainty selector in BoostMIS uniformly chooses the samples with high uncertainty in each predicted class, which helps to maintain remarkable results consistently.

4.3. In-Depth Analysis

We further validate several vital issues of the proposed method by answering the four questions as follows.

Q1: What is the contribution of the individual components in BoostMIS to boost the medical image SSL? We conduct an ablation study to illustrate the effectiveness

---

[2]B and Mo refers to batch size and momentum in SGD optimizer. IP=10% and SS=30% are the sampling ratio of the initial AL labeled pool and SSL training set. AC indicates the max AL training cycle.

Table 1. **Performance comparison on the MESCC dataset.** Superscript † indicates that the model only employs the SSL algorithm. AL* indicates that the SSL model (FixMatch) uses the corresponding AL annotation strategy. We report the accuracy of all these models when the percentage of labeled pool reaches 15%/20%/25%/30%. A larger score indicates better performance, and the top two scores of accuracy are in bold. Acronym notations of each model can be found in Section 4.1.

Methods	AL*	MESCC two-grading				MESCC six-grading		
		15% Labels	20% Labels	25% Labels	30% Labels	20% Labels	25% Labels	30% Labels
P-Labeling† [25]		78.64± 5.20	81.34± 4.64	82.54± 5.85	85.61± 4.74	32.43± 5.39	37.88± 3.53	43.16± 3.81
MixMatch† [4]		85.42± 2.69	86.44± 2.88	88.77± 1.39	90.81± 1.58	43.18± 4.18	46.98± 1.67	52.83± 0.93
FixMatch† [42]		85.52± 2.04	87.37± 3.06	90.34± 1.49	91.09± 0.30	41.41± 4.27	48.65± 1.76	54.32± 1.02
R-Labeling [13]	✓	83.66± 5.29	84.12± 5.91	87.47± 3.34	88.67± 3.53	38.07± 6.69	42.34± 5.76	50.88± 4.83
DBAL [14]	✓	86.72± 4.09	88.67± 3.06	89.04± 1.67	91.74± 0.37	45.13± 3.90	49.58± 1.76	54.87± 0.65
VAAL [41]	✓	86.07± 3.90	**88.95**± 2.88	90.71± 1.49	92.11± 1.21	**46.89**± 2.69	50.60± 0.84	57.85± 0.56
CSAL [15]	✓	**88.39**± 2.51	88.49± 2.60	**91.27**± 1.49	**92.94**± 0.37	46.80± 4.64	**52.83**± 1.21	**58.77**± 0.62
BoostMIS	✓	**89.14**± 3.16	**91.46**± 1.49	**93.13**± 1.67	**95.82**± 0.28	**48.19** ± 1.67	**57.38**± 1.58	**61.47**± 0.56

Table 2. **Experimental results of Macro Precision (MP), Macro F1 score (MF1), Macro Recall (MRC) on MESCC dataset.**

Methods	MESCC two-grading			MESCC six-grading		
	MP	MF1	MRC	MP	MF1	MRC
P-Labeling† [25]	61.15	62.73	65.46	23.13	22.68	32.20
MixMatch† [4]	73.15	77.20	84.50	28.33	29.28	38.81
FixMatch† [42]	73.75	77.97	**85.60**	30.11	31.70	**40.24**
R-Labeling [13]	65.86	66.68	67.64	27.58	27.91	34.34
DBAL [14]	76.97	78.89	81.24	28.13	28.54	33.61
VAAL [41]	75.60	**78.91**	83.79	31.76	33.38	39.75
CSAL [15]	77.87	78.87	79.97	**32.55**	**34.07**	39.83
BoostMIS	**86.09**	**87.54**	**89.15**	**36.74**	**38.77**	**45.06**

Table 3. **Ablation study of different modules in BoostMIS over MESCC.** The results are reported when the AL selected samples reaches 20% data.

Methods	Threshold		AL Selection		ACC	
	Adaptive	Fixed	Uncertain	Unstable	Two	Six
SSL+RS	✓				91.45	53.76
- AS		✓	✓	✓	94.98	60.07
- BUS	✓			✓	**94.06**	**60.44**
- AUS	✓		✓		92.57	58.03
BoostMIS	✓		✓	✓	**95.82**	**61.47**

of each component in Table 3. Comparing BoostMIS and BoostMIS(-AT) (Row 2 vs Row 4), the adaptive threshold significantly contributes 0.84% and 1.40% improvement on accuracy. The results of Row 3 and Row 4 show the accuracy improvement of the AUS (unstable sampling) and BUS (uncertain sampling) in AL module, respectively. Meanwhile, SSL+RS (Row 1) which combines the SSL module in BoostMIS and AL random sampling obtains the worst performance. This further validates the superiority of our AL informative selection. Finally, the results indicate that the introduced two AL selection strategies can boost the medical image SSL in a mutually rewarding way.

Q2: How to determine the optimal start size of the initial labeled pool? We conduct an exploratory analysis to systematically infer a proper starting size of the initial labeled pool (IP) for BoostMIS. As shown in Figure 3, using uniform sampling to select different starting sizes, BoostMIS and its variants achieve different accuracies (optimal start size is 10% data). For example, the model starting with IP=5% labeled data clearly under-performs the model starting with IP = 10% labeled samples, when

both models reach 30% labeled samples. We observe that an SSL model with a small start size of labeled data may blindly predict some unlabeled samples into a certain class, due to the fact that the model lacks the data distribution information of these samples. This may affect the subsequent AL selection which is based on the prediction from the task model. On the contrary, a big start size may also reduce an AL-based SSL ability (e.g., 10% and 15% labeled samples). Although the task model may learn a relatively strong data distribution, the capacity of the AL budget is limited. There are not enough AL selected data to improve the robustness of the task model. In summary, naively choosing a start size is not advisable, because it may lead to under-utilization of AL-based selection.

Q3: Are these annotation candidates from AL selection valuably informative? To build insights on the superior performance of our AL selection method, we perform an in-depth analysis of accuracy improvement that benefits from the informative annotation candidates. Figure 4 summarizes the task model's performance using different labeled samples in supervised learning, which indicates the following: First, the performance is similar when using the same scale of randomly selected data and original data with
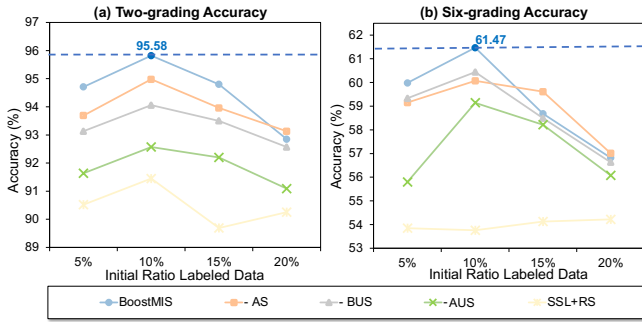
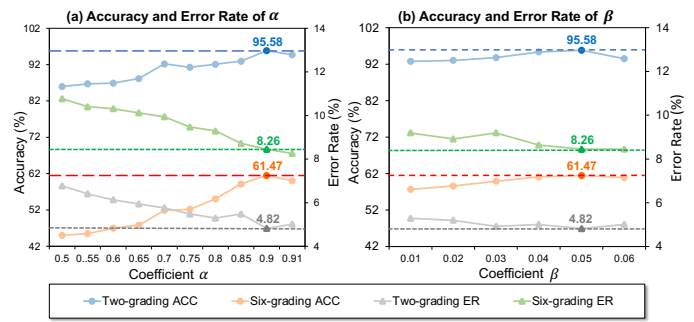Figure 3. **Performance comparison under different size of initial labeled data pool** IP.

Figure 5. **Ablation studies with respect to different coefficient** α **and** β **of adaptive threshold (AS).**
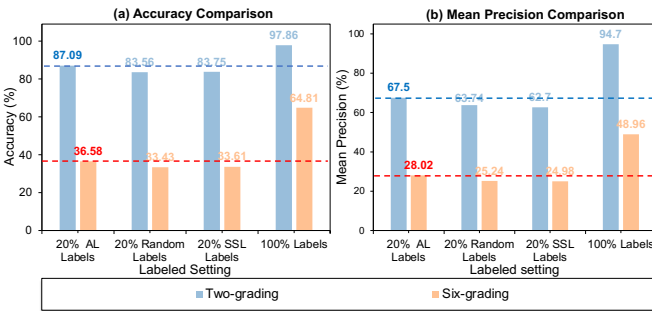

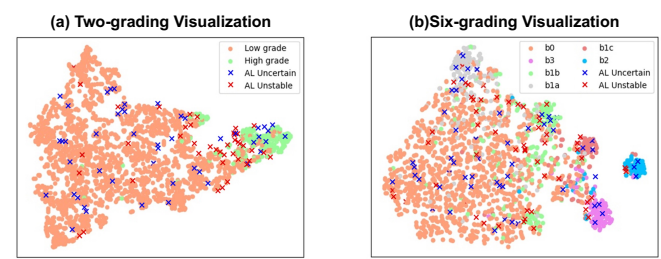Figure 4. **Performance comparison under supervised learning with different labeled data setting on the MESCC dataset.**

Figure 6. **The tSNE embeddings of the MESCC dataset and the AL selection behavior of BoostMIS.**

the supervised label. This proves that the original data are not biased and have not hindered the model's learning ability. Second, training a task model with AL selected samples can represent the global data distribution better. The accuracy results of **20% AL samples are competitive** even comparing the training model with 100% labeled data.

To further analyze where the improvement comes from, we visualize our method's sample selection behavior via tSNE embedding [44] in Figure 6. This visualization result suggests that the AUS tends to select unstable samples near the decision boundary and the BUS can select the representative uncertain samples. These unstable and uncertain samples can smooth the decision boundary for the medical image SSL model and propagate representative label information to unlabeled data, respectively.

Q4: Can the model trust the pseudo-labels based on adaptive threshold? We investigate the impact of hyper-parameters α and β of AS for SSL ability. The classification error rate and accuracy of different hyper-parameters on MESCC dataset are shown in Figure 5. We can observe that the lowest error rate of generated pseudo-labels are 4.83% and 8.26% for two-grading and six-grading. This figure also suggests that the optimal choice of α and β are around 0.9 and 0.05, respectively, either increasing or decreasing these values results in a performance decay.

5. Conclusion

In this paper, we propose a medical image semi-supervised learning (SSL) framework BoostMIS that takes advantage of the active learning (AL) to unleash the potential of unlabeled data for better medical image analysis. In the algorithm, we propose adaptive pseudo-labeling and informative active annotation that leverage the unlabeled medical images and form a closed-loop structure to improve the performance of the medical image SSL model. The experimental results show that BoostMIS significantly outperforms SoTA SSL and AL approaches on medical image classification tasks. The proposed framework, which makes use of Apache SINGA [37] for distributed training, has been integrated into our MLCask [33] for handling healthcare images and analytics.

References

[1] Eric Arazo, Diego Ortego, Paul Albert, Noel E O'Connor, and Kevin McGuinness. Pseudo-labeling and confirmation bias in deep semi-supervised learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2020. 2

[2] Philip Bachman, Ouais Alsharif, and Doina Precup. Learning with pseudo-ensembles. *Advances in neural information processing systems*, 27:3365–3373, 2014. 2

[3] Jonathan Baxter. A model of inductive bias learning. *Journal of artificial intelligence research*, 12:149–198, 2000. 2

[4] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin Raffel. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249*, 2019. 1, 6, 7

[5] Mark H Bilsky, Ilya Laufer, Daryl R Fourney, Michael Groff, Meic H Schmidt, Peter Paul Varga, Frank D Vrionis, Yoshiya Yamada, Peter C Gerszten, and Timothy R Kuklo. Reliability analysis of the epidural spinal cord compression scale. *Journal of Neurosurgery: Spine*, 13(3):324–328, 2010. 6

[6] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010. 6

[7] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021. 1

[8] Nancy Chinchor and Beth M Sundheim. Muc-5 evaluation metrics. In *Fifth Message Understanding Conference (MUC-5): Proceedings of a Conference Held in Baltimore, Maryland, August 25-27, 1993*, 1993. 6

[9] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020. 5

[10] Zihang Dai, Zhilin Yang, Fan Yang, William W Cohen, and Ruslan Salakhutdinov. Good semi-supervised learning that requires a bad gan. *arXiv preprint arXiv:1705.09783*, 2017. 2

[11] Guodong Ding, Shanshan Zhang, Salman Khan, Zhenmin Tang, Jian Zhang, and Fatih Porikli. Feature affinity-based pseudo labeling for semi-supervised person re-identification. *IEEE Transactions on Multimedia*, 21(11):2891–2902, 2019. 2

[12] Sayna Ebrahimi, Mohamed Elhoseiny, Trevor Darrell, and Marcus Rohrbach. Uncertainty-guided continual learning with bayesian neural networks. *arXiv preprint arXiv:1906.02425*, 2019. 3

[13] Rosa L Figueroa, Qing Zeng-Treitler, Long H Ngo, Sergey Goryachev, and Eduardo P Wiechmann. Active learning for clinical text classification: is it better than random sampling? *Journal of the American Medical Informatics Association*, 19(5):809–816, 2012. 6, 7

[14] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2017. 6, 7

[15] Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan Ö Arık, Larry S Davis, and Tomas Pfister. Consistency-based semi-supervised active learning: Towards minimizing labeling cost. In *European Conference on Computer Vision*, pages 510–526. Springer, 2020. 3, 6, 7

[16] Yves Grandvalet, Yoshua Bengio, et al. Semi-supervised learning by entropy minimization. *CAP*, 367:281–296, 2005. 2

[17] Jiannan Guo, Haochen Shi, Yangyang Kang, Kun Kuang, Siliang Tang, Zhuoren Jiang, Changlong Sun, Fei Wu, and Yueting Zhuang. Semi-supervised active learning for semi-supervised models: exploit adversarial examples with graph-based virtual labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2896–2905, 2021. 3

[18] Prashnna Kumar Gyawali, Sandesh Ghimire, Pradeep Bajracharya, Zhiyuan Li, and Linwei Wang. Semi-supervised medical image classification with global latent mixing. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 604–613. Springer, 2020. 3

[19] James Thomas Patrick Decourcy Hallinan, Lei Zhu, Kaiyuan Yang, Andrew Makmur, Diyaa Abdul Rauf Algazwi, Yee Liang Thian, Samuel Lau, Yun Song Choo, Sterling Ellis Eide, Qai Ven Yap, et al. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine mri. *Radiology*, 300(1):130–138, 2021. 1

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3, 6

[21] Jui-Ting Huang and Mark Hasegawa-Johnson. Semi-supervised training of gaussian mixture models by conditional entropy minimization. In *Eleventh Annual Conference of the International Speech Communication Association*, 2010. 2

[22] Vijay Kakani, Van Huan Nguyen, Basivi Praveen Kumar, Hakil Kim, and Visweswara Rao Pasupuleti. A critical review on computer vision and artificial intelligence in food industry. *Journal of Agriculture and Food Research*, 2:100033, 2020. 1

[23] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 2

[24] Alex Kurakin, Colin Raffel, David Berthelot, Ekin Dogus Cubuk, Han Zhang, Kihyuk Sohn, and Nicholas Carlini. Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring. 2020. 1

[25] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896, 2013. 1, 2, 6, 7

[26] Juncheng Li, Siliang Tang, Fei Wu, and Yueting Zhuang. Walking with mind: Mental imagery enhanced embodied qa. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1211–1219, 2019. 1

[27] Juncheng Li, Siliang Tang, Linchao Zhu, Haochen Shi, Xuanwen Huang, Fei Wu, Yi Yang, and Yueting Zhuang. Adaptive hierarchical graph reasoning with semantic coherence for video-and-language inference. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1867–1877, 2021. 1

[28] Juncheng Li, Xin Wang, Siliang Tang, Haizhou Shi, Fei Wu, Yueting Zhuang, and William Yang Wang. Unsupervised reinforcement learning of transferable meta-skills for embodied navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12123–12132, 2020. 1

[29] Juncheng Li, Junlin Xie, Long Qian, Linchao Zhu, Siliang Tang, Fei Wu, Yi Yang, Yueting Zhuang, and Xin Eric Wang. Compositional temporal grounding with structured variational cross-graph correspondence learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 1

[30] Mengze Li, Ming Kong, Kun Kuang, Qiang Zhu, and Fei Wu. Multi-task attribute-fusion model for fine-grained image recognition. In *Optoelectronic Imaging and Multimedia Technology VII*, 2020. 1

[31] Mengze Li, Kun Kuang, Qiang Zhu, Xiaohong Chen, Qing Guo, and Fei Wu. Ib-m: A flexible framework to align an interpretable model and a black-box model. In *BIBM*, 2020. 1

[32] Mengze Li, Tianbao Wang, Haoyu Zhang, Shengyu Zhang, Zhou Zhao, Jiaxu Miao, Wenqiao Zhang, Wenming Tan, Jin Wang, Peng Wang, Shiliang Pu, and Fei Wu. End-to-end modeling via information tree for one-shot natural language spatial video grounding. 1

[33] Zhaojing Luo, Sai Ho Yeung, Meihui Zhang, Kaiping Zheng, Lei Zhu, Gang Chen, Feiyi Fan, Qian Lin, Kee Yuan Ngiam, and Beng Chin Ooi. Mlcask: Efficient management of component evolution in collaborative data analytics pipelines. *37th IEEE International Conference on Data Engineering (ICDE).*, 2021. 8

[34] Dwarikanath Mahapatra, Behzad Bozorgtabar, and Zongyuan Ge. Medical image classification using generalized zero shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3344–3353, 2021. 3

[35] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018. 5

[36] Dong Nie, Yaozong Gao, Li Wang, and Dinggang Shen. Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 370–378. Springer, 2018. 3

[37] Beng Chin Ooi, Kian-Lee Tan, Sheng Wang, Wei Wang, Qingchao Cai, Gang Chen, Jinyang Gao, Zhaojing Luo, Anthony KH Tung, Yuan Wang, et al. Singa: A distributed deep learning platform. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 685–688, 2015. 8

[38] Burr Settles. Active learning literature survey. 2009. 2

[39] Hong Shang, Zhongqian Sun, Wei Yang, Xinghui Fu, Han Zheng, Jia Chang, and Junzhou Huang. Leveraging other datasets for medical imaging classification: evaluation of transfer, multi-task and semi-supervised learning. In *International conference on medical image computing and computer-assisted intervention*, pages 431–439. Springer, 2019. 3

[40] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017. 1

[41] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5972–5981, 2019. 3, 6, 7

[42] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *arXiv preprint arXiv:2001.07685*, 2020. 1, 3, 6, 7

[43] Shuang Song, David Berthelot, and Afshin Rostamizadeh. Combining mixmatch and active learning for better accuracy with fewer labels. *arXiv preprint arXiv:1912.00594*, 2019. 3

[44] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 8

[45] Vikas Verma, Kenji Kawaguchi, Alex Lamb, Juho Kannala, Arno Solin, Yoshua Bengio, and David Lopez-Paz. Interpolation consistency training for semi-supervised learning. *Neural Networks*, 2021. 2

[46] Javier Villalba-Diez, Daniel Schmidt, Roman Gevers, Joaquín Ordieres-Meré, Martin Buchwitz, and Wanja Wellbrock. Deep learning for industrial computer vision quality control in the printing industry 4.0. *Sensors*, 19(18):3987, 2019. 1

[47] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 2

[48] Dong Wang, Yuan Zhang, Kexin Zhang, and Liwei Wang. Focalmix: Semi-supervised learning for 3d medical image detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3951–3960, 2020. 3

[49] Guotai Wang, Wenqi Li, Maria A Zuluaga, Rosalind Pratt, Premal A Patel, Michael Aertsen, Tom Doel, Anna L David, Jan Deprest, Sébastien Ourselin, et al. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE transactions on medical imaging*, 37(7):1562–1573, 2018. 1

[50] Qizhe Xie, Zihang Dai, Eduard Hovy, Minh-Thang Luong, and Quoc V Le. Unsupervised data augmentation for consistency training. *arXiv preprint arXiv:1904.12848*, 2019. 1, 2

[51] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF Conference on*

Computer Vision and Pattern Recognition, pages 10687–10698, 2020. 2

[52] Shuyuan Xu, Jun Wang, Wenchi Shou, Tuan Ngo, Abdul-Manan Sadick, and Xiangyu Wang. Computer vision techniques in construction: a critical review. *Archives of Computational Methods in Engineering*, 28(5):3383–3397, 2021. 1

[53] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3474–3482, 2018. 3

[54] Yi Yang, Yueting Zhuang, and Yunhe Pan. Multiple knowledge representation for big data artificial intelligence: framework, applications, and case studies. *Frontiers of Information Technology & Electronic Engineering*, 22(12):1551–1558, 2021. 1

[55] Jianpeng Zhang, Yutong Xie, Qi Wu, and Yong Xia. Medical image classification using synergic deep learning. *Medical image analysis*, 54:10–19, 2019. 1

[56] Wenqiao Zhang, Haochen Shi, Jiannan Guo, Shengyu Zhang, Qingpeng Cai, Juncheng Li, Sihui Luo, and Yueting Zhuang. Magic: Multimodal relational graph adversarial inference for diverse and unpaired text-based image captioning. *arXiv preprint arXiv:2112.06558*, 2021. 1

[57] Wenqiao Zhang, Siliang Tang, Yanpeng Cao, Shiliang Pu, Fei Wu, and Yueting Zhuang. Frame augmented alternating attention network for video question answering. *IEEE Transactions on Multimedia*, 22(4):1032–1041, 2019. 1

[58] Wenqiao Zhang, Siliang Tang, Jiajie Su, Jun Xiao, and Yueting Zhuang. Tell and guess: cooperative learning for natural image caption generation with hierarchical refined attention. *Multimedia Tools and Applications*, 80(11):16267–16282, 2021. 1

[59] Wenqiao Zhang, Xin Eric Wang, Siliang Tang, Haizhou Shi, Haochen Shi, Jun Xiao, Yueting Zhuang, and William Yang Wang. Relational graph learning for grounded video description generation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3807–3828, 2020. 1

[60] Hong-Yu Zhou, Chengdi Wang, Haofeng Li, Gang Wang, Shu Zhang, Weimin Li, and Yizhou Yu. Ssmd: Semi-supervised medical image detection with adaptive consistency and heterogeneous perturbation. *Medical Image Analysis*, page 102117, 2021. 3

[61] Yi Zhou, Xiaodong He, Lei Huang, Li Liu, Fan Zhu, Shanshan Cui, and Ling Shao. Collaborative learning of semi-supervised segmentation and classification for medical images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2079–2088, 2019. 3

[62] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018. 1