Toronto-3D: A Large-scale Mobile LiDAR Dataset for Semantic Segmentation of Urban Roadways

Weikai Tan[1], Nannan Qin[1,2], Lingfei Ma[1], Ying Li[1], Jing Du[3], Guorong Cai[3], Ke Yang[4], Jonathan Li[1,4*]

[1]Department of Geography and Environmental Management,
University of Waterloo, Waterloo, ON Canada N2L 3G1

[2]Key Laboratory of Planetary Sciences, Purple Mountain Observatory,
Chinese Academy of Sciences, Nanjing, JS 210033, China

[3]College of Computer Engineering, Jimei University, Xiamen, FJ 361021, China

[4]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON Canada N2L 3G1

{weikai.tan, nannan.qin, l53ma, y2424li, ke.yang, junli*}@uwaterloo.ca,
jingdu@jmu.edu.cn, guorongcai.jmu@gmail.com

Abstract

Semantic segmentation of large-scale outdoor point clouds is essential for urban scene understanding in various applications, especially autonomous driving and urban high-definition (HD) mapping. With rapid developments of mobile laser scanning (MLS) systems, massive point clouds are available for scene understanding, but publicly accessible large-scale labeled datasets, which are essential for developing learning-based methods, are still limited. This paper introduces Toronto-3D, a large-scale urban outdoor point cloud dataset acquired by a MLS system in Toronto, Canada for semantic segmentation. This dataset covers approximately 1 km of point clouds and consists of about 78.3 million points with 8 labeled object classes. Baseline experiments for semantic segmentation were conducted and the results confirmed the capability of this dataset to train deep learning models effectively. Toronto-3D is released [1] to encourage new research, and the labels will be improved and updated with feedback from the research community.

1. Introduction

Accurate and efficient scene perception of urban environments are crucial for various applications, including HD mapping, autonomous driving, 3D model reconstruction, and smart city [3]. In the past decade, the largest portion of research in urban mapping is using 2D satellite and airborne imagery [5, 15], and autonomous driving researches also relied heavily on 2D images captured by digital cameras [24]. Compared with 2D images that are short of georeferenced 3D information, 3D point clouds collected by Light Detection and Ranging (LiDAR) sensors have become desirable for urban studies [16, 23]. However, point clouds are unstructured, unordered and are usually in a large volume [20]. Deep learning algorithms have shown advantages tackling these challenges in point cloud processing in various tasks, including semantic segmentation [2, 22], object detection [4, 35], classification [14, 21], and localization [6, 31].

Mobile platforms that integrate MLS sensors, location sensors (e.g. Global Navigation Satellite Systems (GNSS)), and 2D cameras (e.g. panoramic and digital cameras) are gaining popularity in urban mapping and autonomous driving due to the flexibility of data collection [11, 34], but training effective deep learning models is not feasible without high-quality labels of the point clouds [1]. The development of deep learning has always been driven by high-quality datasets and benchmarks [29]. They allow researchers to focus on improving performance of algorithms without the hassle of collecting, cleaning and labeling large amount of data. They also ensure the performance of the algorithms are comparable with each other.

In this paper, Toronto-3D, a new large-scale urban outdoor point cloud dataset acquired by a MLS system is presented. This dataset covers 1km of streets and consists of about 78.3 million points. A sample of the proposed dataset is shown in Fig. 1. The main contributions of this paper are to:

- present a large-scale point-wise labeled urban outdoor point cloud dataset for semantic segmentation,
- investigate an integrated network for point cloud semantic segmentation,
- provide an extensive comparison on the performance of state-of-the-art deep learning semantic segmentation methods on the proposed dataset.
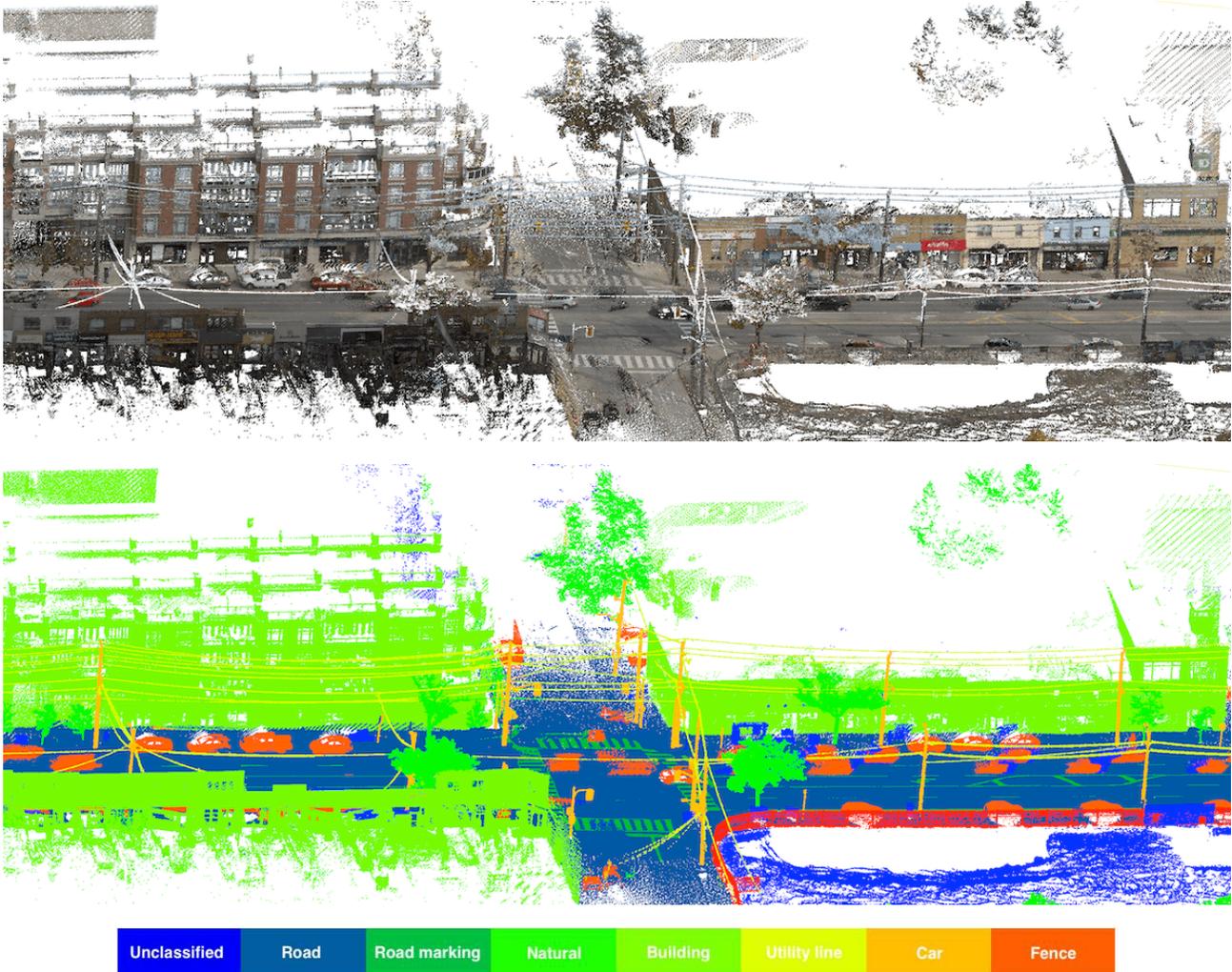
---

[1]https://github.com/WeikaiTan/Toronto-3D

1

Figure 1. Part of our dataset. Top: dataset with natural color (RGB). Bottom: class labels

2. Available point cloud datasets for 3D Semantic Segmentation

With the advancement of LiDAR and RGB-D sensors, and the development of autonomous driving and 3D vision, point cloud data has become more and more accessible. However, such datasets usually have a very large volume of data and contain lots of noise, making it difficult and time-consuming to produce high-quality manual labels. Popular accessible outdoor point cloud datasets for semantic segmentation are as follows:

Oakland 3-D [18] is one of the earliest outdoor point cloud datasets acquired by a MLS system mounted with a side-looking SICK LMS sensor. The sensor is a mono-fiber Li-DAR, and the point density is relatively low. This dataset contains about 1.6 million points and was labeled into 44 classes. However, only 5 classes: vegetation, wire, pole, ground, and facade, were evaluated in literature. This dataset is relatively small so that it is more suitable for developing and testing lightweight networks.

iQmulus [30] dataset comes from the IQmulus & TerraMobilita Contest acquired by a system called Stereopolis II [19] in Paris. A monofiber Riegl LMS-Q120i LiDAR was used to collect the point clouds. The full dataset has over 300 million points labeled into 22 classes, but only a small part of the dataset of 12 million points in a 200 m range with 8 valid classes was publicly available for the contest dataset. This dataset suffers unsatisfactory quality of classification due to occlusion from the monofiber LiDAR sensor and the annotation process [25].

Semantic3D [8] is collected by terrestrial laser scanners, and it has much higher point density and accuracy compared with the other datasets. 8 class labels were included in this dataset. However, only very limited viewpoints are feasible for static laser scanners, and similar datasets are not easily acquired in practice.

Table 1. Recent urban outdoor point cloud datasets for semantic segmentation

Dataset	Year	Primary fields	Length	# points	# classes labeled	# classes evaluated	LiDAR Sensor
Oakland [18]	2009	x, y, z, label	1510 m	1.6 M	44	5	SICK LMS
iQmulus [30]	2015	x, y, z, intensity, GPS time, scan origin, # echoes, object ID, label	200 m	12 M	22	8	Riegl LMS-Q120i
Semantic3D [8]	2017	x, y, z, R, G, B, intensity, label	-	4 B	8	8	Terrestrial Laser Scanner
Paris-Lille-3D [25]	2018	x, y, z, intensity, label	1940 m	143.1 M	50	9	Velodyne HDL-32E
SemanticKITTI [1]	2019	x, y, z, intensity, label	39.2 km	4.5 B	28	25	Velodyne HDL-64E
Toronto-3D (Ours)	2020	x, y, z, R, G, B, intensity, GPS time, scan angle rank, label	1000 m	78.3 M	8	8	Teledyne Optech Maverick

Paris-Lille-3D [25] is one of the most popular outdoor point cloud datasets in recent years. The dataset was collected with a MLS system using Velodyne HDL-32E LiDAR, with a point density and measurement accuracy closer to point cloud data acquired by autonomous driving vehicles. The dataset covers close to 2km with over 140 million points, and very detailed labels of 50 classes were provided. For benchmarks, the dataset uses 9 classes for the purpose of semantic segmentation.

SemanticKITTI [1] is one of the most recent and largest publicly available datasets serving the purpose of semantic segmentation. The dataset was further annotated on the widely used KITTI dataset [7]. This dataset contains about 4.5 billion points covering close to 40 km, and it is labeled by each sequential scan with 25 classes for the evaluation of semantic segmentation. This dataset is more focused on algorithms towards autonomous driving.

Development and validation of deep learning algorithms demand more datasets with various object labels. Toronto-3D is introduced in this paper to provide an additional high-quality point cloud dataset for 3D semantic segmentation with new labels. Table 1 shows a comparison of comprehensive indicators with the above-mentioned datasets.

3. New dataset: Toronto-3D

3.1. Data acquisition

The point clouds in this dataset were acquired with a vehicle-mounted MLS system: Teledyne Optech Maverick[2]. The system consists of a 32-line LiDAR sensor, a Ladybug 5 panoramic camera, a GNSS system, and a Simultaneous Localization and Mapping (SLAM) system. The

LiDAR sensor can capture point clouds at up to 700,000 points per second at a vertical field of view covering from -10° to +30°, with an accuracy of better than 3 cm. The collected point clouds were further processed with LMS Pro[3] software. Natural color (RGB) was assigned to each point with reference to the imaging camera.

3.2. Description of the dataset

This dataset was collected on Avenue Road in Toronto, Canada, covering approximately 1 km of road segment with approximately 78.3 million points. This dataset is divided into four sections, and each section covers a range of about 250 m. An overview of the approximate boundary of each section is illustrated in Fig. 2. This dataset is collected using a 32-line LiDAR sensor, and the point clouds have high density of about 1000 points/m^2 on the ground on average. The dataset covers the full range of the MLS sensor of approximately 100 m away from the road centerline without trimming. Limited post-processing was done to resemble real-world point cloud collection scenarios without trimming far points and resampling.

Each of the four sections of the dataset was saved separately in *.ply* files. The point clouds were classified and point-wise labels were assigned manually using CloudCompare[4] software. Each point cloud file has the following 10 attributes:

- *x, y, z*: Position of each point recorded in meters, in NAD83 / UTM Zone 17N
- *R, G, B*: Natural color reflectance of red, green, blue of each point, recorded in integer [0, 255]
- *Intensity*: LiDAR intensity of each point, normalized to integer [0, 255]

---

[2]http://www.teledyneoptech.com/en/products/
mobile-survey/maverick/

[3]https://www.teledyneoptech.com/en/products/
software/lms-pro/
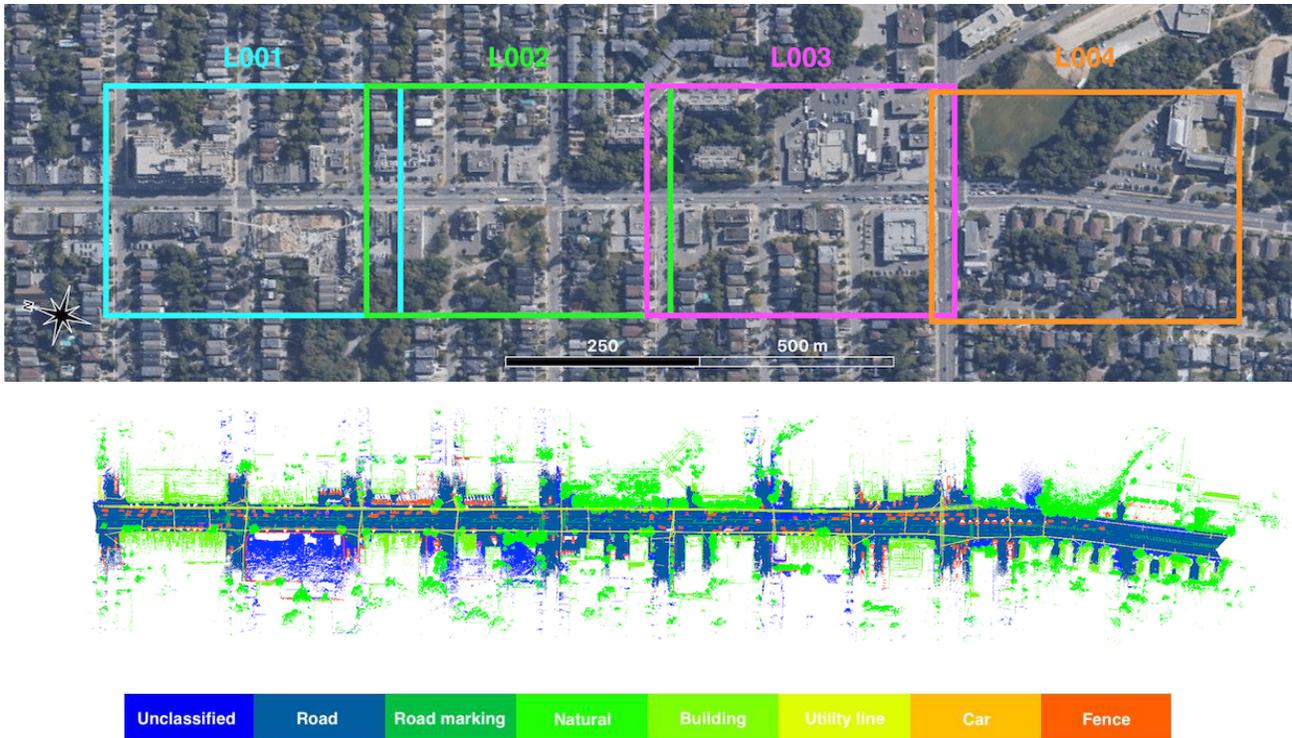
[4]https://www.cloudcompare.org

Figure 2. Overview of the dataset. Top: Approximate boundary of each section in our dataset (Satellite imagery from Google Maps). Bottom: Overview of labels (each class in different colors).

- *GPS time*: GPS time of when each point was collected, recorded in float format
- *Scan Angle Rank*: Scan angle of each point in degree, recorded in integer [-13, 31]
- *Label*: Object class label of each point, recorded in integer [0, 8]

A sample of the Toronto-3D dataset is shown Fig. 1. Similar to previous datasets, the object class labels were defined as follows:

- *Road (label 1)*: Paved road surfaces, including sidewalks, curbs, parking lots
- *Road marking (label 2)*: Pavement markings including driving lines, arrows, pedestrian crossings
- *Natural (label 3)*: Trees, shrubs, not including grass and bare soil
- *Building (label 4)*: Any parts of low and multi-story buildings, store fronts
- *Utility line (label 5)*: Power lines, telecommunication lines over the streets
- *Pole (label 6)*: Utility poles, traffic signs, lamp posts
- *Car (label 7)*: Moving cars and parked cars on road sides and parking lots
- *Fence (label 8)*: Vertical barriers, including wooden fences, walls of construction sites
- *unclassified (label 0)*

A summary of number of points and distribution of labels in each section is shown in Table 2.

3.3. Challenges of Toronto-3D

The Toronto-3D dataset is comparable to Paris-Lille-3D in several aspects. They are both urban outdoor large-scale scenes collected by a vehicle-mounted MLS system with a 32-line LiDAR. Toronto-3D covers approximately half the distance of Paris-Lille-3D and includes half the number of points. They are both labeled with a similar number of classes for the purpose of semantic segmentation. Different from Paris-Lille-3D, the Toronto-3D dataset has the following characteristics that bring more challenges to effective point cloud semantic segmentation algorithms.

Full coverage of LiDAR measurement range. The MLS system that acquired this dataset has a valid measurement distance of approximately 100 m. Different from Paris-Lille-3D where only points within approximately 20 m away from the road centerline are available, Toronto-3D keeps all collected points within about 100 m without trimming. The full coverage of measurement range of Toronto-3D resembles point cloud data collection in real-world scenarios, and it brings challenges of variations of point density at different distances, inclusion of more noise, and inclusion of more objects further away from the sensor.

Table 2. Number of labeled points for each class (thousand)

Section	Road	Road marking	Natural	Building	Utility line	Pole	Car	Fence	unclassified	Total
L001	11,178	433	1,408	6,037	210	263	1,564	83	391	21,567
L002	6,353	301	1,942	866	84	155	199	24	360	10,284
L003	20,587	786	1,908	11,672	332	408	1,969	300	1,760	39,722
L004	3,738	281	1,310	525	37	71	200	4	582	6,748
Total	41,856	1,801	6,568	19,100	663	897	3,932	411	3,093	78,321

Variation of point density. Unlike the relatively small variation of point density in Paris-Lille-3D, the Toronto-3D dataset has a larger variation of point density of objects caused majorly by two reasons: inclusion of all points within full LiDAR measurement range, and repeated scans during point cloud collection. The variations of point density are illustrated in Fig. 3. As illustrated in the scene, the cars (colored in orange) on the streets have much higher point density compared to the parked cars at the upper-middle in the image. The cars with lower density are approximately 30-40 m away from the road centerline, which means such scenarios would not be included in Paris-Lille-3D. In addition, at the center area in the scene, point density is significantly higher (over 10 times higher) compared to other parts of the scene, and this is caused by repeated scans when the vehicle stopped at the intersection during data collection. The repeated scans resulted in variations of point density on the same building at the same distance to the sensor at different locations, and no resampling process was performed. The large variation of point density would be challenging to test the robustness of algorithms to capture features effectively.

New challenging classes. There are two class labels not commonly seen in other popular datasets listed in Table 1, i.e., road marking and utility line. Road markings include various pavement markings on the road surface, including pedestrian crossings and lane arrows, with various sizes and shapes, and they are difficult to distinguish from road surfaces. Though wires, defined similar to utility lines this dataset, were included in Oakland 3-D dataset [18], sample size was limited due to the small number of points in the dataset. The utility lines are thin linear objects that are challenging to identify, especially in areas where they overlap with poles, trees and are close to buildings. In addition, the fence class that covers various wall-like vertical structures is also challenging to identify.

4. Methods

4.1. Recent studies

Semantic segmentation of point clouds is to make predictions on each point to assign a semantic label. With the recent development of 3D deep learning, semantic segmen-
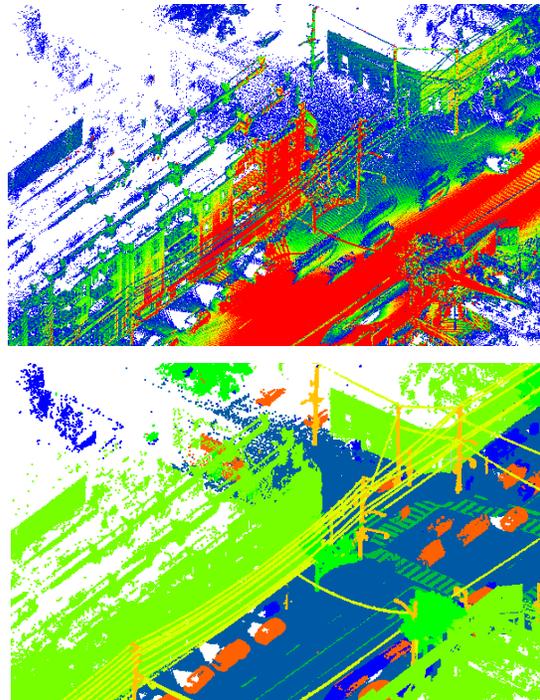

Figure 3. Variations of point density. Top: Point density high to low illustrated in color red to blue. Bottom: Labels of point cloud (each class in different colors).

tation tasks can be achieved by end-to-end deep neural networks. Existing 3D deep learning models on point clouds can be roughly generalized into three categories: view-based models, voxel-based models and point-based models. The view-based models such as MVCNN [27] project 3D point clouds into multiple views as 2D images, but they do not fully use the rich 3D information. Voxel-based models such as VoxNet [17] and 3D-CNN [10] structure unordered point clouds into voxel grids, so that known structures and methods of 2D images can be extended to 3D space. However, the nature of point clouds that they are sparse and have varying densities make voxelization inefficient.

Point-based methods directly process unordered and unstructured point clouds to capture 3D spatial features. Starting from PointNet [20] which learns point-wise spatial feature with multi-layer perceptron (MLP) layers, point-based

methods have been greatly developed, followed by Point-Net++ [22], PointCNN [12]. Graph models were also applied to extract spatial features in point-based models, and such methods include ECC [26] and DGCNN [32].

4.2. Baseline approaches for semantic segmentation

Six state-of-the-art point-based deep learning models for semantic segmentation were tested on the proposed dataset as baseline approaches:

PointNet++ [22] applies PointNet [20], which is the pioneer method using MLPs to process point cloud directly, to local neighborhoods of each point to capture local features, and a hierarchical approach is taken to capture both local and global features.

DGCNN [32] constructs graphs to extract local geometric features from local neighborhoods, and applies EdgeConv as a convolution-like operation. EdgeConv is isotropic about input features with convolutional operations on graph nodes and their edges.

KPFCNN [28] introduces a convolutional operator called KPConv to capture local features with weights defined by a set of kernel points. KPConv is robust to varying point densities and is computationally efficient. KPFCNN is currently ranked first in Paris-Lille-3D benchmark.

MS-PCNN [16] is an end-to-end point cloud segmentation network combining point convolutions with edge information. It applies revised PointConv [33] operations with edge features from revised EdgeConv [32] operations.

TGNet [13] introduces a novel graph convolution function called TGConv defined as products of point features from local neighborhoods. The features are extracted with Gaussian weighted Taylor kernel functions. It is an end-to-end semantic segmentation network with hierarchical TGConv followed by a conditional random field (CRF) layer.

MS-TGNet is proposed in this study with a revised structure of TGNet. Considering the full range of approximately 100 m from the road centerline was preserved in this dataset, there is a large difference in point density. Multi-scale grouping (MSG) proposed in PointNet++ [22] was designed to capture features more effectively in point clouds with large variations in point density. A MSG layer was implemented in the second layer of the original TGNet architecture to capture local geometric features at three different radii (0.2 m, 0.4 m and 0.8 m) through testing.

4.3. Evaluation metrics

For the evaluation of semantic segmentation results, intersection over union (IoU) of each class, overall accuracy (OA) and mean IoU ($mIoU$) are used.

$$IoU_n = \frac{TP_n}{TP_n + FP_n + FN_n} \quad (1)$$

$$OA = \frac{\sum TP_n}{Total\ number\ of\ points} \quad (2)$$

$$mIoU = \frac{\sum IoU_n}{N} \quad (3)$$

where N is the total number of labels, n is the nth label in N, TP, FP and FN represent number of points of true positives, false positives and false negatives of the predictions respectively. OA and $mIoU$ evaluate the overall quality of semantic segmentation, and IoU of each class measures the performance on each class.

4.4. Parameters and configurations

L002 was selected as the testing set among the four sections due to its smaller size and balanced number of points of each label, while the other three sections were used for training and validation. For fair comparison, only coordinates (x, y, z) of point clouds were used.

The parameter settings of PointNet++ and DGCNN were directly used from the networks for indoor scenes, which may limit the performance of these two algorithms to some extent. The parameter settings of KPFCNN, MS-PCNN and TGNet were used from the networks tested on Paris-Lille-3D dataset in literature. The network structures and parameter settings of these algorithms may not be directly comparable, and parameter tuning does not guarantee the fairness of comparison. In this study, the results are for baseline illustration purpose only, and better results could be potentially achieved with further tuning. The models were trained and tested on a NVIDIA RTX 2080Ti with 11G of RAM, and batch sizes were adjusted accordingly.

5. Results and discussions

5.1. Performance of baseline approaches

The results for semantic segmentation baseline approaches using Toronto-3D are shown in Table 3.

PointNet++ [22] achieved highest IoU in road class. However, the PointNet++ model with MSG modules did not perform as well as the base PointNet++ architecture with the published parameter settings of indoor scenes. DGCNN [32] performed the worst in terms of both OA and $mIoU$ in our dataset. Since DGCNN uses KNN for construction of graphs to capture local features, it may not perform well in this dataset with varying point density.

KPFCNN [28] is on the top spot of Paris-Lille-3D benchmark at the moment, and it achieved the highest OA and second highest $mIoU$ among the tested baseline algorithms. KPFCNN achieved the highest IoU in building, utility line, pole and fence segmentation. MS-PCNN [16] and TGNet [13] both achieved $mIoU$ of over 58% following the performance of KPFCNN.

The proposed MS-TGNet achieved comparable results with KPFCNN, achieving highest $mIoU$ of 60.96% and second highest OA of 91.69% among the baseline approaches. It has highest IoUs in road marking and natural classes.

Table 3. Semantic segmentation results of different methods (%)

Methods	OA	$mIoU$	Road	Rd mrk.	Natural	Building	Util. line	Pole	Car	Fence
PointNet++ [22]	91.21	56.55	**91.44**	7.59	89.80	74.00	68.60	59.53	53.97	7.54
PointNet++ (MSG) [22]	90.58	53.12	90.67	0.00	86.68	75.78	56.20	60.89	44.51	10.19
DGCNN [32]	89.00	49.60	90.63	0.44	81.25	63.95	47.05	56.86	49.26	7.32
KPFCNN [28]	**91.71**	60.30	90.20	0.00	86.79	**86.83**	**81.08**	**73.06**	42.85	**21.57**
MS-PCNN [16]	91.53	58.01	91.22	3.50	90.48	77.30	62.30	68.54	53.63	17.12
TGNet [13]	91.64	58.34	91.39	10.62	91.02	76.93	68.27	66.25	**54.10**	8.16
MS-TGNet (Ours)	91.69	**60.96**	90.89	**18.78**	**92.18**	80.62	69.36	71.22	51.05	13.59

A visual comparison of semantic segmentation results of KPFCNN and MS-TGNet is shown in Fig. 4.

5.2. Areas for improvements

From the results of the baseline approaches, all algorithms performed poorly in road marking and fence with IoUs lower than 22%, and the accuracy of utility line, pole and car can be improved. Four examples of errors in the results from the best performing two algorithms, KPFCNN and MS-TGNet, are further illustrated in Table 4.

In *Box a* where some concrete blocks are placed as a barrier, KPFCNN classified them as buildings while MS-TGNet classified them as cars. The concrete blocks may have similar structures to cars and buildings but they are smaller in size. In the same scene, part of trunks of the trees were misclassified as poles in KPFCNN but they are correctly grouped into natural class in MS-TGNet possibly due to the edge features. It explained the higher performance of KPFCNN on pole and higher performance of MS-TGNet on natural class to some extent.

Box b shows a truck partially misclassified by both algorithms. They both correctly classified the lower part of the truck, but the upper part with box shape was classified as building by KPFCNN and fence by MS-TGNet. The truck is a moving object and only one side was completely scanned, adding the vertical structure of the truck, so that these conditions possibly resulted in the confusion.

Box c shows a scenario with road markings. KPFCNN was not able to identify road markings with a IoU close to 0. MS-TGNet captured part of the road markings but at a low IoU of 18.78%, and missed the pedestrian crossings. The road markings on pavements are difficult to distinguish with point coordinates only, and additional color and intensity information would make road markings easier to identify.

Box d illustrates a scene with utility lines at the boundary of the test point cloud. KPFCNN distinguished the utility lines correctly, while MS-TGNet classified the horizontal line as a mixture of several classes including road and road markings. The horizontal utility line in *Box d* is at the boundary of the point cloud without road surface underneath to provide contextual information.

The performance of algorithms with the backbone of PointNet++, including the proposed MS-TGNet, was probably limited by the number of points the structure is able to process with the limitation of RAM. KPFCNN uses a much larger number of points and outperformed MS-TGNet in most categories but showed some weakness in natural and road marking classification. Selection of KPConv kernels and network settings could be improved to capture road marking. New algorithms that can process a large number of points, such as RandLA-Net [9], would have advantages in outdoor scenes, and they will be tested on Toronto-3D.

6. Conclusions

This paper presents Toronto-3D, a new large-scale urban outdoor point cloud dataset collected by a MLS system. The dataset covers approximately 1km of road with over 78 million points in Toronto, Canada. All points were preserved in the range of data collection to resemble real-world application scenarios. This dataset was manually labeled into 8 categories, including road, road marking, natural, building, utility line, pole, car and fence. Five state-of-the-art end-to-end point cloud semantic segmentation algorithms and a proposed network named MS-TGNet were tested as baselines for this dataset.

The proposed MS-TGNet is able to produce comparative performance with state-of-the-art methods, achieving the highest $mIoU$ of 60.96% and a competitive OA of 91.69% in the new dataset. The Toronto-3D dataset provides new class labels including road markings, utility lines and fences, and all tested semantic segmentation methods need to improve on road markings and fences.

The intention of presenting this new point cloud dataset is to encourage developing creative deep learning models. The labels of this new dataset will be improved and updated with feedback from the research community.

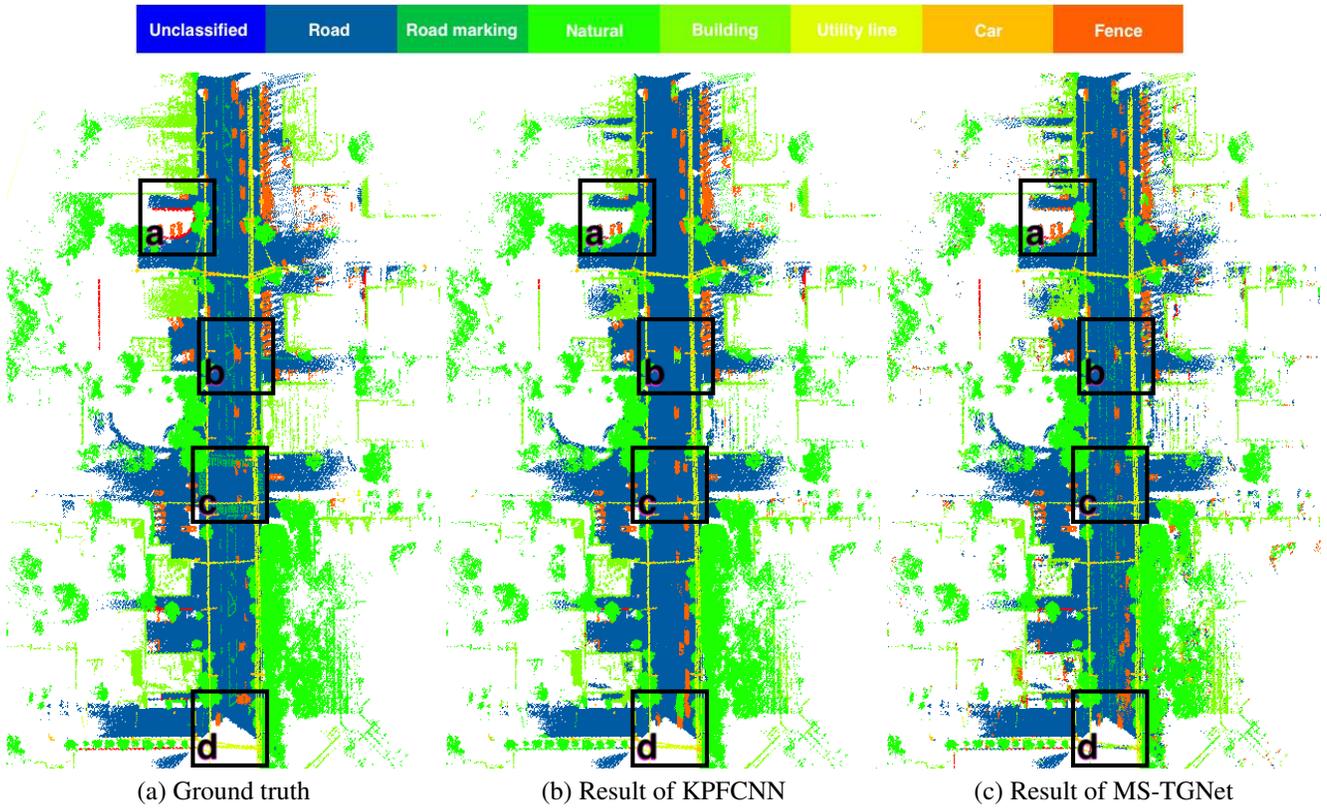(a) Ground truth   (b) Result of KPFCNN   (c) Result of MS-TGNet

Figure 4. Visual comparison of results of semantic segmentation

Table 4. Detailed views at semantic segmentation errors

References

[1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9297–9307, 2019. 1, 3

[2] Alexandre Boulch, Bertrand Le Saux, and Nicolas Audebert. Unstructured point cloud semantic labeling using deep segmentation networks. *3DOR*, 2:7, 2017. 1

[3] Mark Campbell, Magnus Egerstedt, Jonathan P How, and Richard M Murray. Autonomous driving in urban environments: approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1928):4649–4672, 2010. 1

[4] Xiaozhi Chen, Kaustav Kundu, Yukun Zhu, Andrew G Berneshawi, Huimin Ma, Sanja Fidler, and Raquel Urtasun. 3d object proposals for accurate object class detection. In *Advances in Neural Information Processing Systems*, pages 424–432, 2015. 1

[5] Sophie Crommelinck, Rohan Bennett, Markus Gerke, Francesco Nex, Michael Ying Yang, and George Vosselman. Review of automatic feature extraction from high-resolution optical sensor data for uav-based cadastral mapping. *Remote Sensing*, 8(8):689, 2016. 1

[6] Gil Elbaz, Tamar Avraham, and Anath Fischer. 3d point cloud registration for localization using a deep neural network auto-encoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4631–4640, 2017. 1

[7] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. 3

[8] Timo Hackel, N. Savinov, L. Ladicky, Jan D. Wegner, K. Schindler, and M. Pollefeys. SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume IV-1-W1, pages 91–98, 2017. 2, 3

[9] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. *arXiv preprint arXiv:1911.11236*, 2019. 7

[10] Jing Huang and Suya You. Point cloud labeling using 3d convolutional neural network. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2670–2675. IEEE, 2016. 5

[11] Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, et al. Towards fully autonomous driving: Systems and algorithms. In *2011 IEEE Intelligent Vehicles Symposium (IV)*, pages 163–168. IEEE, 2011. 1

[12] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in neural information processing systems*, pages 820–830, 2018. 6

[13] Ying Li, Lingfei Ma, Zilong Zhong, Dongpu Cao, and Jonathan Li. Tgnet: Geometric graph cnn on 3-d point cloud segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 2019. 6, 7

[14] Lorenzo Luciano and A Ben Hamza. Deep learning with geodesic moments for 3d shape classification. *Pattern Recognition Letters*, 105:182–190, 2018. 1

[15] Lei Ma, Manchun Li, Xiaoxue Ma, Liang Cheng, Peijun Du, and Yongxue Liu. A review of supervised object-based landcover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:277–293, 2017. 1

[16] Lingfei Ma, Ying Li, Jonathan Li, Weikai Tan, Yongtao Yu, and Michael A Chapman. Multi-scale point-wise convolutional neural networks for 3d object segmentation from lidar point clouds in large-scale environments. *IEEE Transactions on Intelligent Transportation Systems*, 2019. 1, 6, 7

[17] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928. IEEE, 2015. 5

[18] Daniel Munoz, J Andrew Bagnell, Nicolas Vandapel, and Martial Hebert. Contextual classification with functional max-margin markov networks. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 975–982. IEEE, 2009. 2, 3, 5

[19] Nicolas Paparoditis, Jean-Pierre Papelard, Bertrand Cannelle, Alexandre Devaux, Bahman Soheilian, Nicolas David, and Erwann Houzay. Stereopolis ii: A multi-purpose and multi-sensor 3d mobile mapping system for street visualisation and 3d metrology. *Revue française de photogrammétrie et de télédétection*, 200(1):69–79, 2012. 2

[20] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1, 5, 6

[21] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5648–5656, 2016. 1

[22] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017. 1, 6, 7

[23] Fabio Remondino. Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote sensing*, 3(6):1104–1138, 2011. 1

[24] German Ros, Sebastian Ramos, Manuel Granados, Amir Bakhtiary, David Vázquez, and Antonio M Lopez. Vision-based offline-online perception paradigm for autonomous driving. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 231–238. IEEE, 2015. 1

[25] Xavier Roynard, Jean-Emmanuel Deschaud, and François Goulette. Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6):545–557, 2018. 2, 3

[26] Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3693–3702, 2017. 6

[27] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015. 5

[28] Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J. Guibas. Kpconv: Flexible and deformable convolution for point clouds. *Proceedings of the IEEE International Conference on Computer Vision*, 2019. 6, 7

[29] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011. 1

[30] Bruno Vallet, Mathieu Brédif, Andrés Serna, Beatriz Marcotegui, and Nicolas Paparoditis. Terramobilita/iqmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49:126–133, 2015. 2, 3

[31] Peng Wang, Ruigang Yang, Binbin Cao, Wei Xu, and Yuanqing Lin. Dels-3d: Deep localization and segmentation with a 3d semantic map. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5860–5869, 2018. 1

[32] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 38(5):1–12, 2019. 6, 7

[33] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9621–9630, 2019. 6

[34] Bisheng Yang, Zhen Dong, Gang Zhao, and Wenxia Dai. Hierarchical extraction of urban objects from mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 99:45–57, 2015. 1

[35] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018. 1