A Comparative Study of Semantic Segmentation Using Deep Neural Networks in a GNSS-Denied Underground Parking Lot

Kristie Hu, Jing Du, Xinghan Gong, Lingfei Ma and Jonathan Li

May 12, 2023

A Comparative Study of Semantic Segmentation Using Deep Neural Networks in a GNSS-denied Underground Parking Lot

[redacted]

Abstract: Deep neural networks (DNNs) in intelligent point cloud processing have achieved remarkable progress in recent years. Most existing methods and models were adopted on either outdoor or indoor scenes while very few previous studies were conducted in GNSS-denied environments. In this paper, we carried out a comparative study in semantic segmentation outputs using different DNNs in an underground parking lot dataset. Manually labeled indoor point cloud data were trained and tested using 7 different DNNs (e.g., PointNet, KPConv, FPConv, BAAF-Net, etc.). Our experiments demonstrated how well different DNNs perform in GNSS-denied environments with performance assessments in mIoU, Mean Accuracy (mAcc), Overall Accuracy (OA), as well as visualization outputs. The main contribution of this comparative study is to compare state-of-the-art DNN algorithms' performance in semantic segmentation directly on the raw indoor mobile laser scanning (iMLS) data from a GNSS-denied underground parking lot and evaluate the effectiveness and potentials of different DNNs in underground 3D taskings. Draw upon that, which current algorithms are optimal and how future work in GNSS-denied environments can be inspired and implemented would be discussed.

Keywords: Deep learning, semantic segmentation, indoor point cloud processing, GNSS-denied environment, underground parking lot

1. Introduction

Nowadays, with widespread and profound technological advancement in point cloud processing, the applications regarding navigation, autonomous parking, digital twins development as well as geodata management are getting mature [1]. Driven by rapid urbanization and city advancement, underground parking lots play as important roles in city daily commutes and urban space planning as they become great alternatives for relieving space and traffic burdens on the ground level. However, the existing iMLS datasets are not rich enough to stimulate large-scale Public Participatory GIS (PPGIS) collaboration like outdoor point cloud datasets. To support the development of autonomous driving and parking, there is an increasing demand for standardizing 3D point cloud management as well as indoor facility planning, especially in a GNSS-denied environment like underground parking lots. However, our current knowledge and understanding are insufficient to evaluate how well technologies can be used to support accurate planning and 3D geodata tasking in underground environments.

Due to the inherent nature of point clouds such as irregularity and lack of orders, automating the taskings and fitting optimized DNNs are essential for the development of point cloud processing. Since traditional methods encounter various barriers in point cloud processing (e.g., rule-based and threshold-based models), deep learning-based approaches, which have great potential in 3D taskings, become the mainstream in the field of remote sensing.

In this study, we examine the feasibility and accuracy level of semantic segmentation using seven popular existing deep neural networks, including PointNet [2], PointNet++ [3], SPG [4], KPConv [5], FPConv [6], BAAF-Net [7], and Stratified Transformer [8].

Table 1. DNNs' Semantic Segmentation Results on S3DIS, ModelNet40, and ScanNet (n: with Normal)

DNNs	OA (%)		
	S3DIS	ModelNet40	ScanNet
PointNet [2]	78.6	89.2	78.6
PointNet++ [3]	81.9	90.7/ 91.9^{n}	81.0
SPG [4]	85.5	73.0	85.5
KPConv [5]	-	79.1	68.0
FPConv [6]	88.3	68.9	-
BAAF-Net [7]	88.9	83.1	88.9
Stratified Transformer [8]	91.5	78.1	-

According to Table 1, the first six models are the traditional DNNs that were published before and around 2021. The Stratified Transformer model [8] in the bottom is the only transformer-based DNN that has been used in our comparison study, which is a newly emerging model proposed around mid-2022. Even though there are many existing research studies related to semantic segmentation, many barriers and challenges emerging from indoor point cloud processing are still unsolved. Moreover, few previous studies attempted to directly detect and segment 3D objects in GNSS-denied indoor environments. Hence, inspiration is needed to be stimulated and elaborated to generate better solutions in this emerging field for digital twins and autonomous driving. Overall, the key contributions of this paper are fourfold:

1. We performed semantic segmentation directly on point clouds collected from GNSS-denied environments using seven different state-of-the-art DNNs.
2. We conducted a comparative study based on the results of these DNN models. By comparing the mean Intersection over Union (mIoUs) and Overall Accuracy (OA) from the segmentation results, the accuracy of each class has been classified and discussed.
3. Labeling and label re-assembly will be incorporated into our work before conducting semantic segmentation on this GNSS-denied underground parking lot dataset.
4. Our experiments are expected to inspire future work in exploring the feasibility of optimizing models' segmentation performance within GNSS-denied environments.

2. RELATED WORK

Semantic segmentation of point clouds is one of the most essential 3D processing tasks for users to better understand the patterns and distinguish each class/feature's global-local relationships within a certain scene. Present methods of deep learning-based semantic segmentation usually achieve a global shape embedding based on some point-wise pre-embedding operations along with an aggregation method. Previous works regarding deep learning-based semantic segmentation of point clouds have been reviewed and presented below.

Point-based DNNs. The point-based method directly works on the irregular points, adopting the point features and position information as the inputs, thus keeping the extraction results more intact and cutting down the loss of information compared to projection-based methods. PointNet [2] was designed for effective learning and processing the point-wise dispersed information and global features, using shared MLPs and symmetrical pooling functions respectively. As a point-wise MLP method varied from PointNet, PointNet++ [3] captured local geometric patterns based on neighboring feature pooling. In particular, PointNet++ established a hierarchical structure to group the points and aggregate features progressively. Point convolution methods tend to propose effective convolution operations for the unordered and irregular point clouds. Operations of KPConv [5] convolution weights are located in Euclidean space by kernel points while BAAF-Net [7] accesses the local information of large-scale point clouds via a bilateral structure. Moreover, different choices of the number of kernel points have contributed more flexibility to KPConv than normal convolutions with fixed grids. Besides, Lin et al. [6] proposed a novel surface-style convolution operator namely FPConv to learn local flattening while omitting the intermediate representation transformation-like approaches based on 3D grids or graphs. FPConv [6] has made significant improvements compared to previous surface-style convolution-based methods. Graph-based semantic segmentation methods are also developed to extract the underlying shapes and geometric structures of 3D point clouds. As an attributed directed graph, SPG [4] captured contextual structures of large-scale point clouds, implemented by a graph-based convolutional network.

Transformer-Based DNNs. Transformer and attention-based algorithms have inspired the development of 2D image recognition in recent years [9] and throw light on revolutions of 3D point cloud processing. Since point clouds are essential sets embedded irregularly in a continuous space, Point Transformer [10] attempted to build a transformer layer based on a vector self-attention to maximize local feature extraction. It used the subtraction relation to generate the attention weights and enhance position coding, but it also suffered from non-stationary upon multiple perturbations and information redundancy because of the elaborate point-wise operations. Moreover, massive linear transformation layers may also lead to high computational and memory costs [10]. Point Cloud Transformer [11] adopted PointNet [2] architecture by replacing the shared MLP layers with standard transformer blocks based on the offset-attention mechanism. By enhancing input embedding with the farthest and neighbor point search, PCT [11] makes impressive progress in global feature aggregation in semantic segmentation. Although the transformer-based architectures [8] [10] [11] have achieved state-of-the-art performance in point cloud segmentation, the attempts at transformer-based point cloud processing remain limited, especially for GNSS-denied underground environments.

3. DATASET

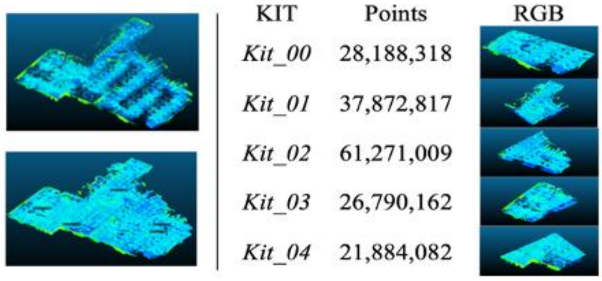

Fig. 1. Overview of the Underground Parking Lot Dataset

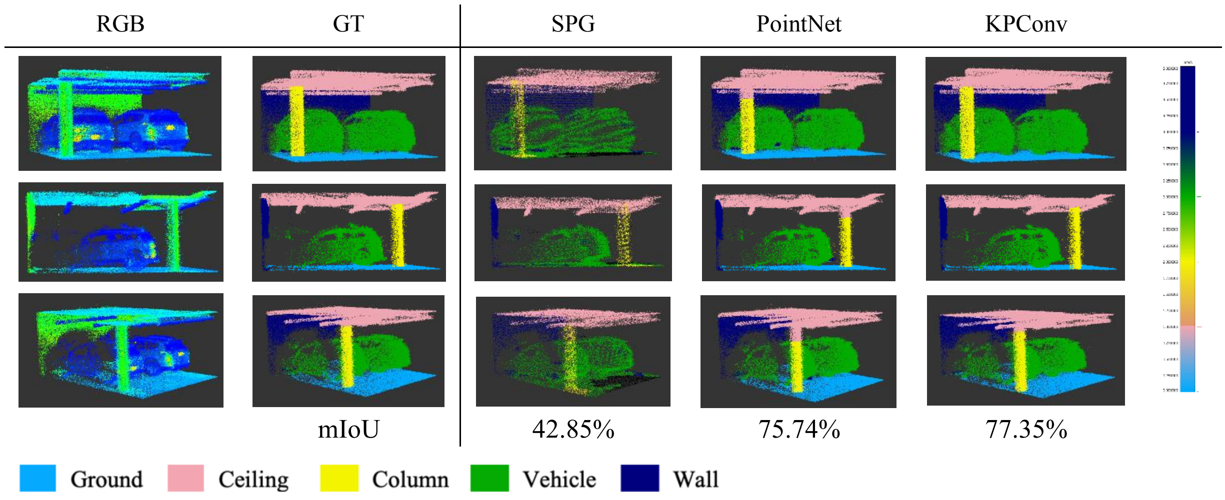

Fig. 2. Visualization of Sub-sample of Semantic Segmentation of the Parking Lot Dataset

The dataset used in the study is a GNSS-denied point cloud set collected from an underground parking lot using a Backpack Laser Scanning (BLS) system [12] by the GIM lab, University of Waterloo. For better visualization purposes, the entire dataset, which contains approximately 154,122,306 points, has been sliced into 5 separate kits shown in Fig. 1. There are 6 manually categorized classes included in this underground parking lot dataset:

- Ground (label 1): the ground surfaces with speed rubber bumps and manholes.
- Ceiling (label 2): ceilings and beams on the top of parking lots, as well as pipes.
- Column (label 3): all pillar and pole-like objects.
- Vehicle (label 4): includes sedans, SUVs, and trucks.
- Walls (label 5): walls at the boundary and dividing walls in the middle of the parking lot.
- Unclassified (label 6): unrecognizable objects.

Label Re-assembly. Since labels in the raw data are 1-indexed, the existing labels will be pre-processed to 0-indexed. Besides, there are some legibility and formatting issues in the labels listed in Table 2. All the class labels highlighted in bold indicate the need for label alteration.

Table 2. GNSS-denied underground data and original manually label fields (* refers to empty label)

Kit 00, 01,02,04	Kit 03	Fixed
Ground - 0	**Ceiling - 1**	Ground - 0
Ceiling - 1	**Ground - 2**	Ceiling -1
Column - 3	**Column - 3**	Column - 2
Vehicle - 4	**Vehicle - 4**	Vehicle - 3
Wall - 5	**Wall - 5**	Wall - 4
Unclassified - 6	[ ]*	Unclassified - 5

4. EXPERIMENTS

Accuracy Assessment. The metrics that we evaluated in the comparative study are overall accuracy (OA), mean accuracy (mAcc), mean intersection over union (mIoU), and IoU for each class respectively. All experiments were carried out on one desktop, the info for the machine is listed in Table 3.

Table 3. Machine Configuration

	Machine A
Processor	Intel (R) Xeon(R) Silver 4210
GPU	NVIDIA GeForce GTX 1080 Ti
RAM	6 GB

To better compare and analyze the semantic segmentation results, visualization outputs have been generated along with these accuracy factors. In Fig. 2, subsamples of the outputs for different DNNs have been demonstrated. The first column refers to the raw data presented in the RGB format while the second column includes the GT, which is the ground truth for training and testing the accuracy of each DNN's performance. Subsamples of semantic segmentation outputs for different models were presented in the following columns. From the visualization outputs in Fig. 2, the result for the SPG model is not ideal as the classification for labels "vehicle", "ground", and "column" are not clear. These have also been reflected in the accuracy measure metric in Fig. 2, as mIoU for the SPG is only about 42.82%. Besides, mIoUs for the PointNet, 75.74%, is relatively higher and the visualization is also more closely related to the representation in the GT. The mIoU for the KPConv is the highest among the three models shown in Fig. 2, which is about 77.35%.

A detailed comparison of the subsample's visualization in PointNet vs. KPConv has been demonstrated in Fig. 3, and segmentation differences have been highlighted in green boxes. According to the comparison, the PointNet

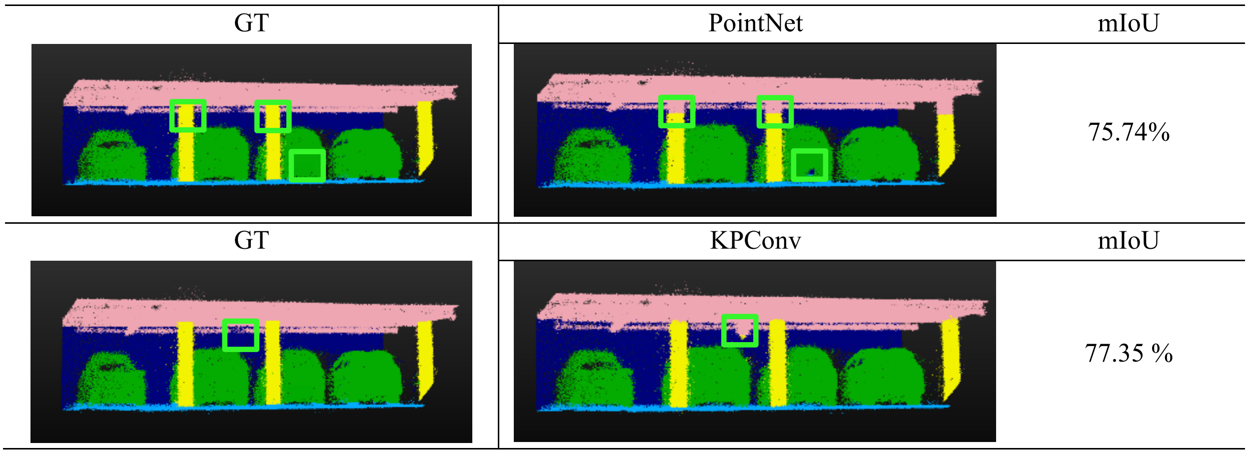

Fig. 3. Ground Truth vs. Visualization of PointNet & KPConv Subsample

Table 4. Quantitative Result of Segmentation Performance over 7 different DNNs in OA, mAcc, mIoU, and class IoUs

DNNs	*OA*	*mAcc*	*mIoU*	ground	ceiling	column	vehicle	wall	unclassified
PointNet [2]	*96.35*	*82.47*	*75.74*	97.8	95.8	72.2	94.3	82.7	11.7
PointNet++ [3]	*96.78*	*92.15*	*83.79*	97.5	95.0	75.5	96.4	90.3	48.0
SPG [4]	*86.97*	*50.90*	*42.85*	93.27	42.16	87.05	33.61	1.00	0.00
KPConv [5]	-	-	*77.35*	97.50	95.73	79.33	97.20	93.31	1.03
FPConv [6]	*96.62*	-	*76.38*	97.23	94.65	76.48	96.04	89.87	4.03
BAAF-Net [7]	-	-	*83.61*	97.43	94.30	74.57	96.79	91.39	47.16
Stratified Transformer [8]	*98.00*	*64.22*	*62.07*	97.02	98.01	82.22	95.02	0.00	0.15

subsample's errors are mainly in the "column", where the upper column has been segmented as part of the "ceiling". As for the KPConv model, it segmented part of the "wall" into the "ceiling" label. Combined with the quantitative result shown in Table 4, the performance accuracy for each DNN has been illustrated. In PointNet, the mIOU is 75.74%. The lowest class IoU is "Column" at 72.2% while the highest class IoU is "Ground" at 97.8%. Compared to PointNet, the IoU in PointNet++ has increased by 8%, to 83.79% in total. SPG has the lowest mIoU at 42.85%. Specifically, it was not effective in segmenting "ceiling" (42.16%) and "vehicle" (33.61%), which is the main factor that brought down the mIoU in the model. Both KPConv and FPConv models performed well in the segmentation of the underground parking lot as their mIoUs are 77.35% and 76.38% respectively with balanced class IoU for segmentation in each individual class. Besides, the BAAF-Net model has the highest mIoU (83.61%) while the Stratified Transformer (ST) has the highest OA (98.00%) among all the DNNs. The mIoU for ST is only 62.07% caused by the imbalanced classes.

Overall, except for the SPG model, all of the other DNNs performed well in a GNSS-denied environment. Among all the class accuracy in Table 4, the "column" is the class that has the lowest mIoU. Performance for this class can be optimized by exploring the feasibility of combining different DNNs. For instance, combine PointNet++ with ST to enhance the class IoU in "column" since ST has an excellent performance in segmenting "column" at around 82.22% in accuracy. Alternatively, since the OA in ST is the highest, future contributions can be elaborated on exploring the feasibility of transformer-based models in GNSS-denied environments by combining algorithms in different DNNs to optimize the accuracy in the "wall" class as it is the only imbalanced class (0%) in ST that result a low mIoU.

5. CONCLUSION

To sum up, we did the first comparative study in investigating the semantic segmentation performances of state-of-art DNNs within GNSS-denied environments. Based on the output, the mIoUs were mostly negatively affected by the accuracy in segmentation of "column" label. The BAAF-Net has the highest mIoU (83.61%) with high and average IoUs for each class while the ST model gets the highest OA (98.00%) with great output in all class expect for "wall".

Based on the quantitative result, future works can contribute to proposing practice measures in refining the 3D taskings in GNSS-denied environments with optimized DNNs. Besides, future contributions should focus on low-level taskings to generate better-quality underground datasets (e.g., through point cloud correction and completion). With an optimized and standardized model specifically designed for GNSS-denied scenes, better solutions for the development of digital twins and autonomous driving can be applied in future. After that, we shall move to the next stage for exploring, popularizing, and commercializing 3D tasks in the industry for application in GNSS-denied environments such as underground parking lots.

6. ACKNOWLEDGEMENTS

[redacted]

7. REFERENCES

[1] Y. Li, L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao and L. Jonathan, "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 32, no. 8, pp. 3412-3432, 2021.

[2] C. R. Qi, H. Su, K. Mo and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *CVPR,* pp. 652-660, 2017.

[3] C. R. Qi, L. Yi, H. Su and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *NeurIPS,* 2017.

[4] L. Landrieu and M. Simonovsky, "Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs," *CVPR,* pp. 4558 - 4567, 2018.

[5] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette and L. J. Guibas, "KPConv: Flexible and Deformable Convolution for Point Clouds," *ICCV,* pp. 6410-6419, 2019.

[6] Y. Lin, Z. Yan, H. Huang, D. Du, L. Liu, S. Cui and X. Han, "FPConv: Learning Local Flattening for Point Convolution," *CVPR,* pp. 4292-4301, 2020.

[7] S. Qiu, S. Anwar and N. Barnes, "Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion," *CVPR,* 2021.

[8] X. Lai, J. Liu, L. Jiang, L. Wang, H. Zhao, S. Liu, X. Qi and J. Jia, "Stratified Transformer for 3D Point Cloud Segmentation," *CVPR,* pp. 8490 - 8499, 2022.

[9] H. Hu, Z. Zhang, Z. Xie and S. Lin, "Local Relation Networks for Image Recognition," *ICCV,* pp. 3463 - 3472, 2019.

[10] H. Zhao, L. Jiang, J. Jia, P. H. Torr and V. Koltun, "Point Transformer," *ICCV,* pp. 16239-16248, 2021.

[11] M.-H. Guo, J.-X. Cai, Z.-N. Liu , T.-J. Mu and R. R. Martin , "PCT: Point cloud transformer," *CVM,* vol. 7, no. 2, pp. 187-199, 2021.

[12] Z. Gong, J. Li, Z. Luo, C. Wen, C. Wang and J. Zelek, "Mapping and Semantic Modeling of Underground Parking Lots Using a Backpack LiDAR System," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 2, pp. 734-746, 2021.