IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9554610

SEMANTIC SEGMENTATION OF UAV LIDAR POINT CLOUDS OF A STACK INTERCHANGE WITH DEEP NEURAL NETWORKS

Weikai Tan[1], Student Member, IEEE, Dedong Zhang[2], Lingfei Ma[3], Lanying Wang[1], Nannan Qin[4], Yiping Chen[5], Senior Member, IEEE, Jonathan Li[1,2], Senior Member, IEEE*

[1]Department of Geography and Environmental Management, University of Waterloo
[2]Department of Systems Design Engineering, University of Waterloo
[3]School of Statistics and Mathematics, Central University of Finance and Economics
[4] Key Laboratory of Planetary Sciences, Purple Mountain Observatory, Chinese Academy of Sciences
[5]School of Informatics, Xiamen University
[*]Corresponding author: junli@uwaterloo.ca

ABSTRACT

Stack interchanges are essential components of transportation systems. Mobile laser scanning (MLS) systems have been widely used in road infrastructure mapping, but accurate mapping of complicated multi-layer stack interchanges are still challenging. This study examined the point clouds collected by a new Unmanned Aerial Vehicle (UAV) Light Detection and Ranging (LiDAR) system to perform the semantic segmentation task of a stack interchange. An end-to-end supervised 3D deep learning framework was proposed to classify the point clouds. The proposed method has proven to capture 3D features in complicated interchange scenarios with stacked convolution and the result achieved over 93% classification accuracy. In addition, the new low-cost semi-solid-state LiDAR sensor Livox Mid-40 featuring a incommensurable rosette scanning pattern has demonstrated its potential in high-definition urban mapping.

Index Terms— LiDAR, UAV, mobile laser scanning, road infrastructure, deep learning, semantic segmentation

1. INTRODUCTION

Urbanization and population growth have brought demands and pressure for urban transportation infrastructures. Multi-layer interchanges have transformed road intersections from 2D to 3D spaces to reduce traffic flow interference to improve traffic efficiency and driving safety [1]. Accurate 3D mapping of road infrastructure provides spatial information for various applications, including navigation, traffic management, autonomous driving and urban landscaping [2]. Multi-layer interchanges are complicated road objects with different designs in various terrain environments, making them difficult to be mapped and modelled.

LiDAR sensors are favorable in 3D urban mapping due to the capability of capturing 3D information directly, Airborne laser scanning (ALS) and vehicle-mounted MLS systems are commonly used in interchange bridge mapping and reconstruction. Interchange bridge extraction from ALS point clouds were usually performed by topography removal [3] or segment extraction with threshold-based algorithms [1]. MLS systems have been widely used in 3D road inventory mapping, but a large portion of existing methods require road surface extraction as the first step [4]. Moreover, most existing studies on stack interchange mapping only focused on the road surfaces, while the underneath structures including piers and beams were disregarded since little data could be collected from ALS or MLS sensors. UAV is a thriving platform in urban 3D mapping due to the flexibility of data collection. UAV LiDAR systems can collect flyover bridge structures underneath the bridges. Therefore, previous methods focusing on road surfaces may not apply in UAV point clouds.

Deep learning methods have outperformed conventional threshold-based methods and methods using hand-crafted features in recent years in capturing features from the massive amounts of unordered and unstructured point clouds [5]. Among the various methods, the point-based network KPFCNN [6] has shown superior performance in a number of outdoor point cloud datasets in the semantic segmentation task. However, mapping large road infrastructures like stacked interchanges requires the algorithms to capture features at large scales. Multi-scale grouping (MSG) was proposed in [7] to deal with uneven distribution of point density, but it helps increase the effective receptive field (ERF) as well. However, MSG is generally very time and space-consuming in computation to be implemented in the networks [8]. Stacking point convolution layers can increase the ERF [9], and it has shown promising results in semantic segmentation tasks in outdoor scenarios [10]. In this study, an improved segmentation network with stacked kernel point convolutions (KPConv) [6] is explored in a multi-layer stack interchange scene collected by a UAV LiDAR system.

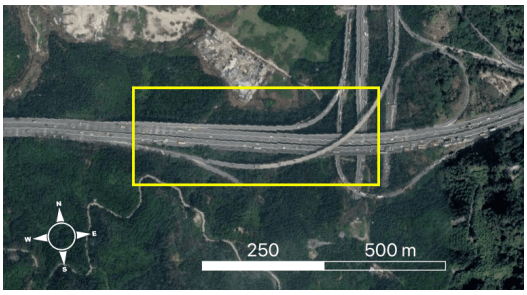

Fig. 1. Fulong Flyover and point cloud coverage

Fig. 2. UAV LiDAR system with Livox Mid-40[1]

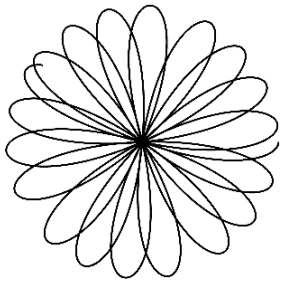

Fig. 3. Scanning pattern of Livox LiDAR

2. DATA

The data used in this study was a part of Fulong Flyover interchange located in Shenzhen, China. A satellite image of Fulong Flyover with the LiDAR point clouds' approximate coverage is illustrated in Fig. 1. The point clouds span approximately 400 m of road segments and cover several flyover bridges. The UAV flew along the flyover bridge at about 15-20 m distance several times to collect the point clouds. Livox High-Precision Mapping[1] software was used to stitch the scans to produce the point cloud map.

The MLS system features a Livox Mid-40 LiDAR[2] mounted on an APX-15 UAV[3], as shown in Fig. 2. The Livox Mid-40 is a robotic prism-based LiDAR different from conventional multi-line LiDAR commonly used in autonomous vehicles such as Velodyne HDL-32E. The new type of low-cost LiDAR sensor features an incommensurable scanning pattern with peak angular density at the center of field-of-view (FOV) [11], as illustrated in Fig. 3. Livox Mid-40 has a circular FOV of 38.4° with an angular precision over 0.1°, and it has a range of up to 260 m with a range precision of 2 cm. In addition, it is capable of capturing 100,000 points/second. Compared with conventional revolving mechanical LiDAR sensors, the Livox LiDAR has a smaller FOV but the incommensuarble scanning pattern could increase the point density over time, filling up to 20% FOV at 0.1 second and 90% at 1 second. Therefore, the Livox LiDAR could acquire points with high density at a fraction of the cost of high-end LiDARs in some applications where full 360° FOV are not required.

The point cloud consists of over 65 million points with xyz coordinates and intensity. All the points were manually labelled into 6 classes: natural, bridge, road, car, pole and guardrail. The road surface density at near range could be estimated as 500-1000 points/m^2, matching the performance of common vehicle-mounted 32-line LiDAR sensors.

This data poses several challenges to point cloud processing algorithms. First, the flyover bridges are very close to the mountain, and vegetation appears over, beside and under the bridges. Conventional threshold-based algorithms would be difficult to separate the bridge structure. Second, the road surface points are rare and incomplete because of the slant angle and heavy traffic at data collection. Algorithms rely on detecting road surfaces may not perform as expected. Finally, the bridge components are very large objects so that capturing features at a large scale would be challenging.

[1]https://github.com/Livox-SDK/livox_high_precision_mapping

[2]https://www.livoxtech.com/mid-40-and-mid-100

[3]https://www.applanix.com/products/dg-uavs.htm

3. PROPOSED METHOD

The task of semantic segmentation is to assign class labels to each point of the point clouds. The proposed method is an end-to-end semantic segmentation network taking raw point clouds directly to produce point-wise classification labels.

The KPConv [6] operation g at point coordinate x applies different weights W_k to each region with regard to a linear correlation between x'_i, the relative position of a point in space x_i within radius r, and the kernel points $\tilde{x}_k$:

$$g(x) = \sum_{k<K} max(0, 1 - \frac{||x'_i - \tilde{x}_k||}{d}) W_k \qquad (1)$$

where K is the total number of kernel points, and d refers to the influence distance. In this study, same settings as [6] were used: $K = 15$, $d = 1.5r$.

As illustrated in Fig. 4, the proposed network uses 5-layer U-Net styled architecture built upon KPFCNN [6]. The architecture mainly contains the following blocks: stacked convolution blocks, pooling blocks and upsampling blocks. Stacked convolution blocks are implemented with three sets of batch normalization-KPConv-Leaky ReLU operations.

The more challenging eastern 1/3 section with multiple layers of bridges was selected as the testing set. The rest 2/3 of the dataset with fewer layers of bridges and less curved roads was used as the training set. The point clouds were resampled into 10 cm grids prior to training, and only point coordinates were used in this study. Data augmentation methods includes random shuffling, scaling, and random rotation around z axis. The network was trained on a NVIDIA RTX 2080 Ti with batch size set to 6.

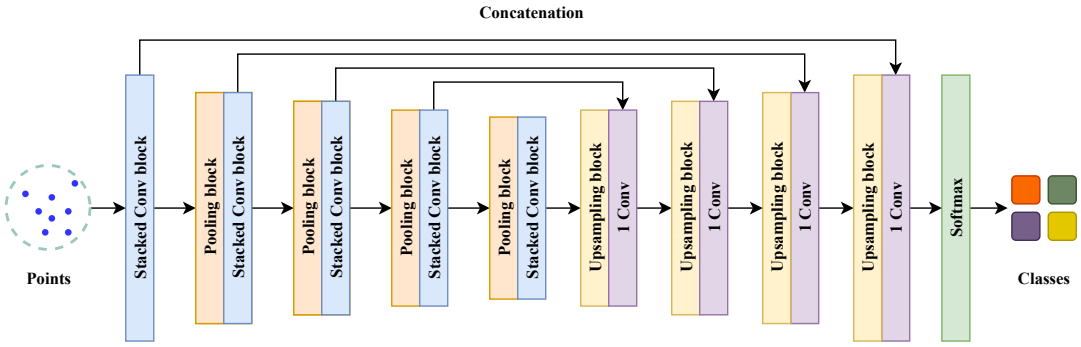

Fig. 4. Framework of network architecture for semantic segmentation

Intersection over union (IoU) of each class, overall accuracy (OA) and mean IoU ($mIoU$) are used to evaluate semantic segmentation results.

$$IoU = \frac{TP}{TP + FP + FN} \qquad OA = \frac{\sum TP}{N} \tag{2}$$

where TP, FP and FN represent the numbers of predicted points of true positives, false positives and false negatives respectively, and N stands for the total number of points. IoU of each class measures the performance on each class. $mIoU$ is the mean of IoUs of all classes evaluated. OA and $mIoU$ evaluate the overall quality of semantic segmentation.

4. RESULTS AND DISCUSSIONS

Table 1 shows the results of the proposed method in comparison to some recent 3D semantic segmentation algorithms. Compared with KPFCNN [6], our proposed method utilizing stacked convolutions has shown improvements in both OA and $mIoU$. The most significant improvement was on pole identification with 14% increase in IoU. The performance on car, bridge and road classification also improved by a noticeable margin. The convolution blocks with triple KPConv operations achieved the highest performance, which agrees with the findings of [9]. With an OA of over 93% and a $mIoU$ over 88%, the semantic segmentation result of the proposed method could provide confident guidance on high-definition mapping and 3D reconstruction of urban road infrastructures.

The testing set scene is much more complicated than the training set, with multi-layered bridges with many occlusions. The significant errors could be attributed to the misclassification of multiple classes to natural. The testing scene is in the middle of the stack interchange far from the UAV's trajectory, the point density at the testing scene is relatively sparse. The occlusions by the upper-level bridges make the point density of lower-level bridges even sparser, and some errors could be observed in the left half of the scene. Moreover, the flyover intersection is very close to the mountains and the vegetation, resulting in further difficulties in semantic segmentation. The confusion between guardrail and bridge could be the next most significant confusion due to only one side of the bridges was visible to the UAV LiDAR, which resulted in some confusions in structure. In terms of the errors on poles, errors could be found on the two street lamps' lamp parts on the right, while the pole parts are correctly classified. Finally, some flat surfaces underneath the bridge at the right corner were classified as road due to the flatness.

Even though most of the observed errors in this study could be attributed to the structures underneath the flyover bridges, these structures are often not visible in ALS point clouds acquired from above in previous studies [1]. The UAV LiDAR system provides a different view of road infrastructures compared to ALS and vehicle-based MLS systems so that structures underneath the bridges could be mapped and modelled with an additional perspective. Multi-angle and multi-directional scans from the UAV would potentially increase the performance of semantic segmentation if more data could be collected.

This study intends to serve as a conceptual and preliminary experiment on stack interchange mapping using the new type of LiDAR sensor mounted on a UAV system. There are some limitations could be addressed in future research in the following directions. First, the dataset is relatively small so that the algorithms are prone to overfit, and transfer learning could be applied to take advantage of previously trained weights on larger datasets. But this study demonstrated that accurate results of the complicated scenarios could be achieved even with little training data using the proposed algorithm. Second, there are a few calibration issues can be improved to reduce some distortions and artifacts possibly due to vibrations. LiDAR odometry and mapping (LOAM) algorithms [12] could be incorporated in addition to the GNSS records at post processing to increase the quality of registration. Next, only point coordinates were used, and the LiDAR reflectance could contribute to better semantic segmentation results. Last but not least, the semantic segmentation can serve as the initial step of scene recognition, and 3D reconstruction methods could be applied to create accurate 3D models and fill data gaps caused by occlusions.

Method	OA(%)	mIoU(%)	Natural(%)	Bridge(%)	Road(%)	Car(%)	Pole(%)	Guardrail(%)
PointNet++ [7]	84.86	70.81	75.98	77.82	83.79	79.70	59.40	48.19
MS-TGNet [8]	83.80	69.03	78.19	69.43	81.65	65.93	68.24	50.74
KPFCNN [6]	90.89	82.60	87.25	78.23	88.72	88.55	77.14	75.72
Ours - double Conv	91.59	86.54	86.70	82.71	**90.45**	**95.01**	89.56	74.83
Ours - triple Conv	**93.54**	**88.71**	**90.52**	**86.53**	88.79	93.09	**91.41**	**81.94**

Table 1. Evaluation of semantic segmentation

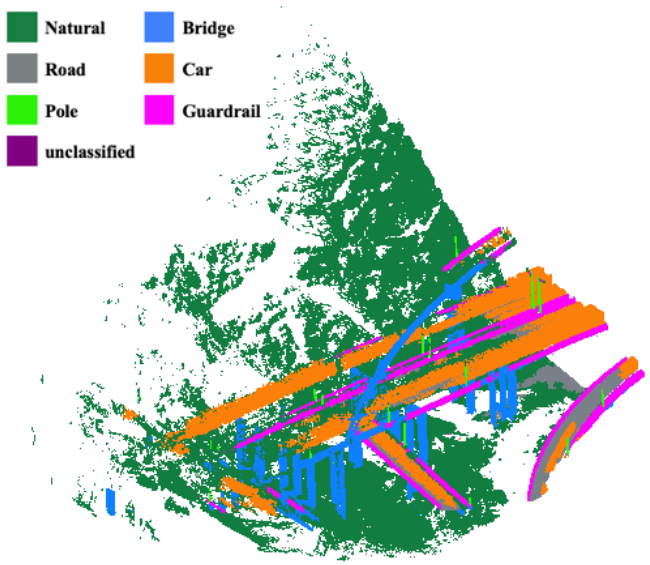

Fig. 5. Result of proposed method

5. CONCLUSION

This study tested a UAV LiDAR system on urban road infrastructure mapping with a case study on semantic segmentation of a multi-layer stack interchange. An end-to-end point cloud semantic segmentation network was adopted to classify components of the stack interchange, i.e. natural, bridge, road, car, pole and guardrail. The method achieved over 93% overall accuracy and over 88% $mIoU$, despite the challenges of lack of road surfaces and complicated structures. The stacked convolutional layers were effective in increasing the ERF and improving semantic segmentation performance. The results could be extended to various tasks, including urban high-definition 3D mapping and 3D model reconstruction. In addition, this study also showcases the potential capability of the UAV-mounted Livox Mid-40 MLS system on urban high-definition mapping of complicated scenarios.

6. ACKNOWLEDGMENTS

We would like to thank Livox Technology Ltd. for providing the point cloud data. This work was supported in part by the National Natural Science Foundation of China (42001400).

7. REFERENCES

[1] L. Cheng, Y. Wu, Y. Wang, L. Zhong, Y. Chen, and M. Li, "Three-Dimensional Reconstruction of Large Multilayer Interchange Bridge Using Airborne LiDAR Data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 2, pp. 691–708, 2015.

[2] R. Wang, J. Peethambaran, and D. Chen, "LiDAR Point Clouds to 3-D Urban Models: A Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 2, pp. 606–627, 2018.

[3] S. J. Oude Elberink and G. Vosselman, "3D Information Extraction from Laser Point Clouds Covering Complex Road Junctions," *The Photogram. Record*, vol. 24, no. 125, pp. 23–36, 2009.

[4] L. Ma, Y. Li, J. Li, C. Wang, R. Wang, and M. Chapman, "Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review," *Remote Sens.*, vol. 10, no. 10, pp. 1531, 2018.

[5] Y. Li, L. Ma, Z. Zhong, F. Liu, M. A. Chapman, D. Cao, and J. Li, "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review," *IEEE Trans. Neural Netw. Learning Syst.*, 2020, DOI: 10.1109/TNNLS.2020.3015992.

[6] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and Deformable Convolution for Point Clouds," in *Proc. IEEE ICCV*, Seoul, Korea (South), Oct. 2019, pp. 6410–6419, IEEE.

[7] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," in *Adv. NIPS*, 2017, pp. 5099–5108.

[8] W. Tan, N. Qin, L. Ma, Y. Li, J. Du, G. Cai, K. Yang, and J. Li, "Toronto-3D: A Large-Scale Mobile LiDAR Dataset for Semantic Segmentation of Urban Roadways," in *Proc. IEEE CVPRW*, June 2020, pp. 202–203.

[9] F. Engelmann, T. Kontogianni, and B. Leibe, "Dilated Point Convolutions: On the Receptive Field Size of Point Convolutions on 3D Point Clouds," *arXiv:1907.12046 [cs]*, 2020.

[10] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, "RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds," in *Proc. IEEE CVPR*, June 2020, pp. 11108–11117.

[11] Z. Liu, F. Zhang, and X. Hong, "Low-cost Retina-like Robotic Lidars Based on Incommensurable Scanning," *arXiv:2006.11034 [cs]*, 2020.

[12] J. Lin and F. Zhang, "Loam_livox: A fast, robust, high-precision LiDAR odometry and mapping package for LiDARs of small FoV," *arXiv:1909.06700 [cs, eess]*, 2019.