

Li, Q. and Zhao, Q. (2023) Weakly-Supervised Semantic Segmentation of Airborne LiDAR Point Clouds in Hong Kong Urban Areas. In: 2023 Joint Urban Remote Sensing Event (JURSE), Heraklion, Greece, 17-19 May 2023, ISBN 9781665493734 (doi: [10.1109/JURSE57346.2023.10144215](https://doi.org/10.1109/JURSE57346.2023.10144215))



Copyright © 2023 IEEE. Reproduced under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

For the purpose of open access, the author(s) has applied a Creative Commons Attribution license to any Accepted Manuscript version arising.

<https://eprints.gla.ac.uk/300530/>

Deposited on: 16 June 2023

# Weakly-Supervised Semantic Segmentation of Airborne LiDAR Point Clouds in Hong Kong Urban Areas

Qiaosi Li

Urban Big Data Centre  
School of Social and Political Sciences  
University of Glasgow  
Glasgow, UK  
Qiaosi.Li@glasgow.ac.uk

Qunshan Zhao

Urban Big Data Centre  
School of Social and Political Sciences  
University of Glasgow  
Glasgow, UK  
Qunshan.Zhao@glasgow.ac.uk

**Abstract**— Semantic segmentation of airborne LiDAR point clouds of urban areas is an essential process prior to applying LiDAR data to further applications such as 3D city modeling. Large-scale point cloud semantic segmentation is challenging in practical applications due to the massive data size and time-consuming point-wise annotation. This paper applied weakly-supervised Semantic Query Network and sparse points annotation pipeline to practical airborne LiDAR datasets for urban scene semantic segmentation in Hong Kong. The experiment result obtained the overall accuracy over 84% and the mean intersect over union over 75%. The capacity of intensity and return attributes of LiDAR data to classify the vegetation and construction was explored and discussed. This work demonstrates an efficient workflow of large-scale airborne LiDAR point cloud semantic segmentation in practice.

**Keywords**—Airborne LiDAR, Point cloud classification, Urban buildings and trees, Deep learning

## I. INTRODUCTION

Aerial LiDAR (light detection and ranging) point clouds that provide geographical and spatial information are superior data sources to construct the 3D city models. Essential information, for instance, footprints and height of buildings, vegetation coverage, and canopy height could be efficiently extracted from LiDAR data in city-wide scale. To achieve this goal, LiDAR point cloud classification (also known as semantic segmentation) is a required process to classify points into multiple homogeneous regions that have the same properties i.e., buildings, trees, roads, etc. in an urban scene [1]. In recent years, various deep learning networks have been proposed for point cloud classification/semantic segmentation tasks and showed powerful capacity in benchmark point cloud datasets [2]–[4]. However, there is a lack of understanding of the efficiency and the performance of these algorithms on large scale real datasets. Furthermore, algorithms applicable to fully annotated benchmark datasets are difficult to reproduce with large-scale real datasets due to the demanded time and effort to label billions of points before training the model. In this study, we investigated a weakly-supervised deep learning algorithm Semantic Query Network (SQN) [5] with real airborne LiDAR dataset to understand the performance of semantic segmentation of buildings and vegetation in the urban area with sparse point labels. We also explored the use of LiDAR attributes such as echo number and return intensity in point cloud classification.

## II. RELATED WORK

Point cloud classification or semantic segmentation is challenging due to the implicit structure, high redundancy, and uneven sampling density of the point cloud data [1]. Early research mostly relied on geometric hand-crafted features that can represent the local feature of each point e.g., normal, curvature and roughness. Some machine learning algorithms such as random forest and support vector machine are then applied to conduct the per-point classification based on hand-crafted features [6], [7]. Other studies attempted to improve classification by considering the neighborhood contextual features by incorporating conditional random field [8]–[10]. Nevertheless, the performance of these methods can largely depend on extraction and selection of suitable hand-crafted features that are required prior knowledge. The versatility and the performance of machine learning-based approaches are limited in large-scale datasets.

Deep learning-based methods such as convolution neural network (CNN) produced high accuracy in 2D image classification/segmentation tasks. Early studies transferred 3D point clouds to 2D rasters, such as the digital terrain model and digital surface model to feed into the network [11], [12]. This method still requires hand-crafted features and can cause information loss in data transformation between 3D point cloud and 2D raster. In recent years, deep learning networks were developed to approach the semantic segmentation task for the raw point clouds. Influential studies include pioneering work PointNet [13] and PointNet++ [13], [14], KPConv [3], and network designed for large-scale datasets Superpoint Graphs [15] and RandLA-Net [2]. The state-of-art networks proposed for airborne LiDAR point clouds including GraNet [16] and LGENet [17]. These methods are directly applied to raw point clouds and show higher capacity than traditional methods requiring hand-crafted features. It is noted that deep learning algorithms commonly demand a large number of training samples to train the model. It is extremely time-consuming to label billions of points in a large-scale LiDAR dataset in practical application. The emergence of the weak-supervision deep learning method provides a promising solution from point cloud annotation to semantic segmentation. Some studies introduced super-point-based active learning [18] and self-supervised pre-train methods to fine-tune the network to approach weak supervision [19], [20]. For airborne LiDAR point clouds, pseudo-labeling strategy was applied to create additional supervisory sources, which was demonstrated to produce comparable accuracy with full

supervision network [21], [22]. The recently proposed SQN method does not require pre-training, post-processing, or active labeling processing, which has great potential in large-scale real world LiDAR datasets classification [5].

### III. METHODS

#### A. Semantic Query Network

Semantic Query Network (SQN) was proposed to give 0.1% of labeled points to train the model for large-scale point cloud semantic segmentation tasks. SQN consists of two main components: the point local feature extractor for learning visual patterns and the point feature query network for collecting relevant semantic features to train the model. The entire raw point clouds were firstly encoded into a set of hierarchical latent representations through RandLA-Net encoder[2] that includes four layers of local feature aggregation followed by a random sampling operation. Arbitrary 3D point position was taken as input to query latent representations within a local neighborhood. These queried representations are compressed into a compact representation for the queried points. After obtaining the unique and representative feature vector of the queried points, it was fed into a series of multilayer perceptrons (MLPs) to infer the final semantic label. Overall, given a sparse number of annotated points, neighboring point features are queried in parallel with training, which allows useful training signals to be back-propagated to a wider spatial context (Fig. 1).

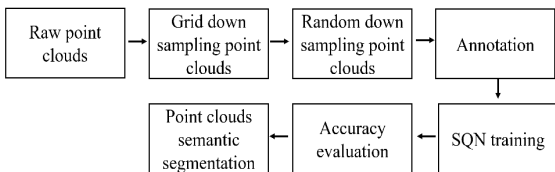


Fig. 1. The workflow of applying weakly-supervised semantic query network (SQN) in LiDAR point clouds semantic segmentation.

A user-friendly annotation pipeline was developed based on the open-source software CloudCompare (<https://www.danielgm.net/cc/>). A grid down sampling was applied to the raw point cloud and then a random down sampling to 0.1% of the total points was conducted in CloudCompare. Since the remaining points are sparse, the original point cloud was used as a reference to carry on point-wise annotation. It was reported that using the designed pipeline took 18 hours to annotate 0.1% of the large urban-scale SensatUrban dataset instead of 600 hours for full points labelling. More details about the algorithm and annotation strategy can be found in paper of SQN [5].

The versatility of SQN has been validated with various point clouds benchmark datasets such as aerial photogrammetry point clouds SensatUrban, aerial LiDAR point clouds DELAS, terrestrial LiDAR point clouds Toronto3D, and Semantic3D. Other than 3D point position XYZ, RGB color attributes of points served as input features if applicable in the dataset. The results showed that the performance of SQN trained with 0.1% labeled points was comparable with full points supervised network (e.g., RandLA-Net) with overall accuracy (OA) of 91%-97% and mean intersect over union (MIoU) of 70.9%-77.7% [5].

#### B. Experimental Dataset and Settings

The experiment area is in Shatin, Hong Kong, China with an area of 7.2 km<sup>2</sup> (Fig. 2). There are multiple types of

construction in the experiment area including high-rise residential buildings, low-rise village houses, and large public buildings. Green spaces are mainly composed of wood areas in open spaces (e.g., in parks and hills) and planted trees in residential gardens and nearby roads. The terrain of the experiment area is rough in the hill areas but flat in the land reclamation area along the river. The LiDAR data was collected by Optech Galaxy Prime (Optech, In., Toronto, Canada) scanner in January 2020 (<https://www.geomap.cedd.gov.hk/GEOOpenData/eng/Default.aspx>). The scanner emitted the near-infrared (1064 nm) and recorded up to five laser pulses. The average point density of the dataset is 44 points/m<sup>2</sup>.

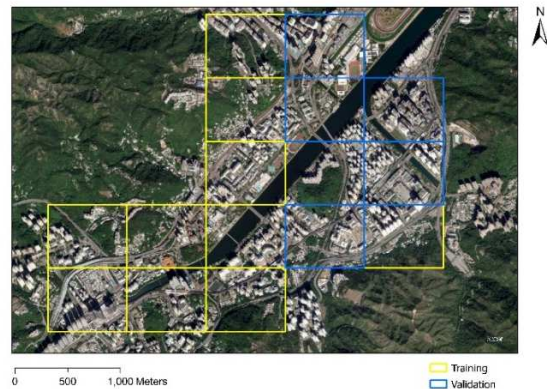


Fig. 2. The experiment area in Shatin, Hong. 11 tiles in yellow boundary were used for training while 5 tiles in blue boundary were used for validation. The basemap is high-resolution satellite images from Digital Globe (Esri, USA).

Other than X, Y, and Z coordinates, the LiDAR data contain laser pulse returns information and intensity values, but without RGB attributes. A previous study demonstrated that RGB attributes did not give improvement to semantic segmentation since the roof of buildings can have the same color as roads [23]. We observed intensity values of buildings are significantly higher than vegetation and the return information can reflect the internal structure under the surface, which was widely used in vegetation studies[24]. However, the intensity and the return attributes were rarely used for urban space point cloud classification in previous research. In this work, three LiDAR attributes namely intensity, return number, and the number of returns were served as input features besides the XYZ position of points. The intensity values range from 0 to ~60000 which were normalized to 0 to 1 to improve the model stability and performance. We followed the instruction of the annotation pipeline to label down sampling points [5]. As the main objective of the project is to extract 3D information of buildings and trees from the LiDAR data, therefore, points were only labeled into four semantic categories: ground, buildings, trees (including arbor and shrub but excluding lawn), and others in this study. The training points were used to conduct the One-way ANOVA analysis and the post Tukey's Test for each LiDAR attribute (i.e., intensity, return number, and number of returns) among 4 classes to understand if the attribute can help the classification. Moreover, to evaluate the capacity of LiDAR attributes, multiple SQN models were trained and evaluated by feeding with XYZ coordinates alone or coordinates with extra attributes.

The hyperparameter of the SQN and the training parameter (e.g., training number, epoch, learning rate, etc) were set to

follow the SQN code (<https://github.com/QingyongHu/SQN>). All experiments were conducted on a PC with Intel Xeon Gold 6234 CPU, an NVIDIA Quadro RTX 8000 GPU, and the RAM of 128 GB.

The Overall accuracy (OA) (1) and mean intersect over union (MIoU) were used to evaluate the performance of point cloud semantic segmentation. IoU is computed from the ratio of the overlap area to the combined area of prediction and ground truth to evaluate the correctness of segmentation (2). It is noted that the full point cloud (i.e., raw point cloud) of validation data were used to evaluate the segmentation performance.

$$\text{Overall accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Intersect over union} = \frac{TP}{TP + FP + FN} \quad (2)$$

Where TP is true positive; FN is false negative; FP is false positive; TN is true negative.

#### IV. RESULTS AND DISCUSSION

##### A. LiDAR Point Cloud Annotation

Following the annotation pipeline, the number of points that need to be labeled was largely reduced, allowing great error tolerance to decide the annotation in boundary areas. Table I compares the total number of points before and after down sampling for training data.

TABLE I. A COMPARISON OF THE TOTAL NUMBER OF POINTS BEFORE AND AFTER DOWN SAMPLING ON TRAINING DATA (11 YELLOW TILES IN FIG. 2).

Dataset	Area	Grid size	Raw points	Grid sampled points	Annotation points
Shatin	4.95 km <sup>2</sup>	0.32	254M	90M	90212

M is millions.

##### B. Experimental Results

The one-way ANOVA results showed that all of LiDAR attributes were significantly different among 4 classes in the training dataset with p-value less than 0.05. Specific to Tukey test results, differences in intensity could be told between trees and buildings, trees and others, and buildings and ground, but it is difficult to tell the difference between buildings and others. There were significant differences in return number for any two classes. The number of returns of trees was significantly different from the other 3 classes. Therefore, there is potential to improve the point cloud classification by including LiDAR intensity and return information to train the SQN model.

It took approximately 72 hours to finish 100 epochs of training. Full point clouds (Fig. 2, 5 blue tiles) were used to evaluate segmentation. Although fewer points were used to train the model, the SQN achieved semantic segmentation with OA over 84 % and MIoU over 75%. Table II compares the accuracy of point cloud semantic segmentation by using coordinates (XYZ) with and without LiDAR attributes. The inclusion of LiDAR attributes 2 slightly improved the OA of 0.09% and MIoU of 0.22%, which is likely related to the slight improvement in classifying ground and buildings. Using raw intensity value even decreased the overall accuracy, while it obtained the best IoU for trees at 93.96%. However, we did observe that there is a significant variation in raw intensity

value due to the sensor internal error or inter-sensor difference, which may limit the intensity to work even after the normalization. A sensor-based calibration for intensity value may be needed in such case to make good use of intensity attributes.

TABLE II. OVERALL ACCURACY (OA, IN %) AND INTERSECT OVER UNION (ALONE, IN %) OF EACH CLASS ON VALIDATION DATA.

Point cloud feature	OA	MIoU	IoU			
			G	T	B	O
XYZ	87.00	75.54	78.93	93.43	82.70	47.09
XYZ + Attributes 1	86.40	74.51	78.15	93.96	81.95	44.99
XYZ + Attributes 2	87.13	75.76	79.22	93.50	83.28	47.05

Attributes 1: the original intensity, return number and, number of returns.  
Attributes 2: the normalized intensity, return number, number of returns.  
G: Ground; T: Trees; B: Buildings; O: Others.

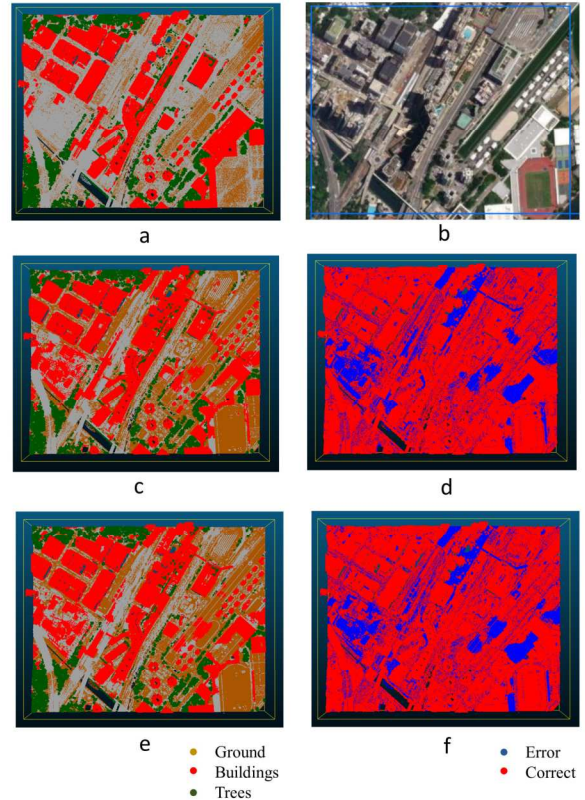


Fig. 3. An example tile of SQN prediction results. Map a, ground truth point clouds; b, high-resolution satellite images from Digital Globe (Esri, USA); c, predicted point clouds using XYZ coordinates; d, error map between a and c; e, predicted point clouds using XYZ coordinates and attributes 2 (see TABLE II); f, error map between a and e

Fig. 3 compares the ground truth, the SQN predicted results, and the error maps of one validation tile. Consistent with the accuracy results, trees were well identified acquiring the best IoU of over 93%. This may benefit from the structural differences between the vegetation and the artificial construction i.e., the crown shape, canopy structure, and roughness of the canopy surface, etc. Most of the buildings and ground were correctly classified reaching IoU of 83% and 75% respectively. However, there was some confusion between the roof and ground due to their similar characteristic of flat and smooth surfaces and intensity values. The others class obtained the lowest accuracy of 47%. It is difficult for the network to learn some specific features for the others class as this class includes everything other than trees, buildings,

and ground that contains many inconsistent features. A subcategory of the others class (e.g., the wall, bridge, road, rail, car, etc.) may eliminate the above-mentioned problem. There is no significant difference found between using XYZ coordinates along and combining XYZ coordinates with intensity and return attributes to predict the label of point clouds in this experiment.

## V. CONCLUSIONS

This study shows that the weakly-supervised SQN method has great potential to conduct semantic segmentation with large-scale airborne LiDAR in urban areas. An efficient workflow to process LiDAR point cloud from annotation, model training, and prediction was demonstrated. Our experiment achieved good performance of semantic segmentation in the real urban scene of Shatin, Hong Kong with an OA over 84% and MIOU over 75%. The experiment results demonstrated that XYZ coordinates play the most important role in buildings and trees segmentation. The inclusion of LiDAR intensity and return information to XYZ coordinates has potential to improve semantic segmentation. However, the variation of intensity caused by inter-sensor differences may reduce the performance of intensity. Therefore, a calibration or normalization to intensity is suggested before feeding it to the network. This study paves the way for city-wide LiDAR point semantic segmentation, which will facilitate the LiDAR data in subsequent research such as 3D model construction and tree structure mapping.

## ACKNOWLEDGMENT

This work is supported by the ESRC to the Urban Big Data Centre [ES/L011921/1 and ES/ S007105/1] and is also funded by the Glasgow City Council. The authors would like to thank the CEED, HKSAR for providing data, and the developer (Dr. Qingyong Hu) of RandLA-Net and SQN for giving us suggestions for this research.

## REFERENCES

- [1] A. Nguyen and B. Le, "3D point cloud segmentation: A survey," *IEEE Conference on Robotics, Automation and Mechatronics, RAM - Proceedings*, pp. 225–230, 2013, doi: 10.1109/RAM.2013.6758588.
- [2] Q. Hu et al., "RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11108–11117, 2020.
- [3] H. Thomas, C. R. Qi, J. E. Deschard, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-October, pp. 6410–6419, Oct. 2019, doi: 10.1109/ICCV.2019.00651.
- [4] T. Chaton and N. Chaulet, "Torch-Points3D: A Modular Multi-Task Framework for Reproducible Deep Learning on 3D Point Clouds", 2020 International Conference on 3D Vision (3DV) (pp. 1-10). IEEE.
- [5] Q. Hu et al., "SQN: Weakly-Supervised Semantic Segmentation of Large-Scale 3D Point Clouds", *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVII* (pp. 600-619). Apr. 2021, doi: 10.48550/arxiv.2104.04891.
- [6] H. B. Kim, G. Sohn, H. B. Kim -A, and G. Sohn -A, "Random Forests based multiple classifier system for power-line scene classification," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 38, pp. 253–258, 2011, doi: 10.5194/isprsarchives-XXXVIII-5-W12-253-2011.
- [7] J. Zhang, X. Lin, and X. Ning, "SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas," *Remote Sensing* 2013, Vol. 5, Pages 3749-3775, vol. 5, no. 8, pp. 3749–3775, Jul. 2013, doi: 10.3390/RS5083749.
- [8] M. Weinmann, B. Jutzi, and C. Mallet, "Semantic 3D scene interpretation: A framework combining optimal neighborhood size selection with relevant features," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3, pp. 181–188, Aug. 2014, doi: 10.5194/isprsannals-ii-3-181-2014.
- [9] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Contextual classification of lidar data and building object detection in urban areas," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 152–165, 2014, doi: 10.1016/j.isprsjprs.2013.11.001.
- [10] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 286–304, Jul. 2015, doi: 10.1016/j.isprsjprs.2015.01.016.
- [11] Z. Yang, W. Jiang, B. Xu, Q. Zhu, S. Jiang, and W. Huang, "A convolutional neural network-based 3D semantic labeling method for ALS point clouds," *Remote Sens (Basel)*, vol. 9, no. 9, Sep. 2017, doi: 10.3390/rs9090936.
- [12] R. Zhao, M. Pang, and J. Wang, "Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network," *International journal of geographical information science*, vol. 32, no. 5, pp. 960–979, May 2018, doi: 10.1080/13658816.2018.1431840.
- [13] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.
- [14] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [15] L. Landrieu and M. Simonovsky, "Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4558–4567, Nov. 2017, doi: 10.48550/arxiv.1711.09869.
- [16] R. Huang, Y. Xu, and U. Stilla, "GraNet: Global relation-aware attentional network for semantic segmentation of ALS point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 1–20, Jul. 2021, doi: 10.1016/j.isprsjprs.2021.04.017.
- [17] Y. Lin, G. Vosselman, Y. Cao, and M. Y. Yang, "Local and global encoder network for semantic segmentation of Airborne laser scanning point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 176, pp. 151–168, Jun. 2021, doi: 10.1016/j.isprsjprs.2021.04.016.
- [18] X. Shi, X. Xu, K. Chen, L. Cai, C. S. Foo, and K. Jia, "Label-Efficient Point Cloud Semantic Segmentation: An Active Learning Approach," *arXiv preprint arXiv:2101.06931*, Jan. 2021, doi: 10.48550/arxiv.2101.06931.
- [19] Z. Zhang, R. Girdhar, A. Joulin, and I. Misra, "Self-Supervised Pretraining of 3D Features on any Point-Cloud," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 10232–10243, Jan. 2021, doi: 10.48550/arxiv.2101.02691.
- [20] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep Convolutional Networks on 3D Point Clouds," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 9613–9622, Nov. 2018, doi: 10.48550/arxiv.1811.07246.
- [21] P. Wang and W. Yao, "A new weakly supervised approach for ALS point cloud semantic segmentation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 188, pp. 237–254, Jun. 2022, doi: 10.1016/j.isprsjprs.2022.04.016.
- [22] Y. Lin, G. Vosselman, and M. Y. Yang, "Weakly supervised semantic segmentation of airborne laser scanning point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 187, pp. 79–100, May 2022, doi: 10.1016/j.isprsjprs.2022.03.001.
- [23] S. B. Wicaksono, A. Wibisono, W. Jatmiko, A. Gamal, and H. A. Wisesa, "Semantic Segmentation on LiDAR Point Cloud in Urban Area using Deep Learning," 2019 International Workshop on Big Data and Information Security (IWBIIS), Oct. 2019, pp. 63–66. doi: 10.1109/IWBIIS.2019.8935882.
- [24] H. Qin, W. Zhou, Y. Yao, and W. Wang, "Individual tree segmentation and tree species classification in subtropical broadleaf forests using UAV-based LiDAR, hyperspectral, and ultrahigh-resolution RGB data," *Remote Sensing of Environment*, vol. 280, p. 113143, Oct. 2022, doi: 10.1016/j.rse.2022.113143.