Research on street tree extraction method based on improved RandLA-Net

Liangjun Gao *

College of Civil Engineering, Hefei University of Technology, Hefei, China, 230009

* Corresponding Author Email: 17856000972@163.com

Abstract. In view of the challenges faced by street tree extraction in the current urban road scene, such as the problem of accurate segmentation caused by insufficient automation and the overlapping of ground object point clouds, this study is committed to improving the automatic recognition and extraction ability of street trees by improving the RandLA-Net network. In the data preprocessing stage, the ground points are first removed by calculating the point cloud slope value, and the outliers are removed by statistical filtering algorithm. Subsequently, two key improvements were made to the RandLA-Net network: first, the cross-entropy loss function in the original network was replaced by the focus loss function to reduce the impact of the imbalance in sample size on network segmentation, and to enhance the sensitivity of the model to samples that are difficult to classify; Second, the Efficient Channel Attention (ECA) module is introduced to strengthen the model's ability to learn and express street tree features. In order to verify the effectiveness of the proposed improved method, this study uses the road dataset of Yongchang Road in Beijing to carry out comprehensive training and testing. Experimental results show that the improved model has significant improvement in segmentation accuracy and Intersection over Union (IoU), which fully proves the effectiveness of the proposed improved strategy.

Keywords: Street Tree Extraction, RandLA-Net, Loss Function, ECA, Attention Mechanisms.

1. Introduction

Urban road greening is an essential component of the urban ecosystem and plays an irreplaceable role in improving the urban environment and enhancing the quality of life for residents. Among the various elements of road greening, street trees are a primary component, and information about their health status, distribution density, and species diversity is crucial for urban planning and management. However, traditional methods of obtaining information about street trees largely depend on manual surveys, which are not only time-consuming and labor-intensive but also difficult to ensure in terms of data accuracy and timeliness. In recent years, three-dimensional laser point cloud data has been widely applied in the fields of urban 3D modeling and object recognition [1]. Nevertheless, in urban road scenarios, due to the overlapping of ground object point clouds, the diverse shapes of street trees, and their dense distribution, achieving automatic recognition and precise extraction of street trees remains a challenging issue. To date, research on the extraction of street trees from point cloud data has made certain progress, with the main methods being as follows: (1) Methods based on clustering algorithms: For instance, the authors improved the traditional density-based spatial clustering of applications with noise (DBSCAN) algorithm to achieve clustering analysis of vehicle-mounted laser point cloud data [2]. By setting reasonable density thresholds and neighborhood radii, the algorithm can effectively separate street tree point clouds from the surrounding environmental point clouds. The authors employed the Euclidean clustering algorithm to perform cluster analysis on preprocessed point cloud data [3]. By setting a reasonable distance threshold, the point cloud data was divided into different cluster groups. (2) Methods based on deep learning: For instance, the study involved the use of convolutional neural networks and several classic object detection algorithms within deep learning to process panoramic image data collected by a vehicle-mounted mobile measurement system, achieving automatic recognition and extraction of street trees [4]. The PointCNN deep learning model was utilized to process vehicle-mounted laser point cloud data, enabling accurate extraction and segmentation of street trees from the point cloud data [5]. (3) Methods based on sample modeling: A model and a prior sample set were constructed based on the expressive features of the target pole-like objects [6]. Under the constraints of the model and semantic rules, candidate target pole-like point clouds were obtained, and the extraction of the target pole-like point clouds was finally achieved using the least squares algorithm. (4) Methods based on machine learning: A support vector machine (SVM) algorithm based on the radial basis kernel function was used to fuse and classify the extracted features, training an SVM model to accurately identify street tree targets [7]. Based on the morphological characteristics of pole-like ground objects, a longitudinal grid template was used to preliminarily extract pole-like objects, followed by training an SVM classification model with pre-established pole-like object samples to classify the extracted pole-like objects [8]. Despite certain achievements, current research still faces limitations and challenges. For example, issues such as the construction of point cloud data spatial indexing and the differentiation between street trees and man-made pole targets still need to be further addressed. At the same time, most existing extraction methods suffer from low levels of automation and insufficient segmentation accuracy, which fail to meet the demands of practical applications.

In response to the aforementioned issues, this study aims to enhance the automatic recognition and extraction capabilities of street trees by improving the deep learning model RandLA-Net. RandLA-Net, as an advanced point cloud processing network, has demonstrated excellent performance in 3D point cloud segmentation tasks. However, for the specific task of street tree extraction, the original network still has some deficiencies. Therefore, this study will make targeted improvements to RandLA-Net, introducing a focal loss function and an ECA (Efficient Channel Attention) attention module. These enhancements will increase the model's sensitivity to challenging samples and strengthen its ability to learn and express features of street trees, thereby achieving more efficient and accurate extraction of street trees. Finally, this paper validates the effectiveness of the improved method using vehicle-mounted laser point cloud data from Yongchang Road in Beijing for training and testing.

2. Research on street tree extraction method

2.1. Data preprocessing

2.1.1. Ground point filtering

Ground points constitute a significant portion of road scene point clouds, and they do not provide useful information for subsequent semantic segmentation tasks of street trees and other ground objects. Instead, they increase the complexity of data processing. Therefore, this paper initially employs a slope-based method for the removal of ground points.

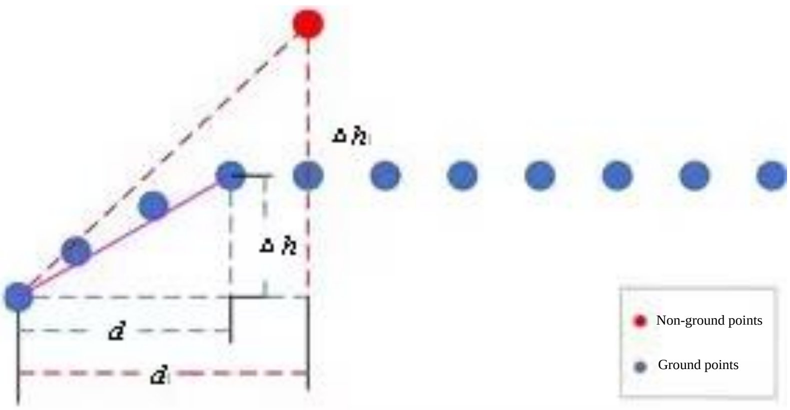


Figure 1. Schematic diagram of slope calculation

As shown in Figure 1, the fundamental principle of the gradient method is predicated on the premise that the variation in gradients of terrestrial points is relatively minor, which is leveraged for the purpose of filtration. The algorithmic procedure is as follows:

(1) Define the grid dimensions and the gradient threshold, and ascertain the number of grids.

(2) Cycle through the grids sequentially, find the lowest point of each grid, and then calculate the slope values for the other points in the grid according to the formula, assuming the lowest point is $\min_p$.

$$i = \frac{\Delta h}{d} = \frac{z - \min_p(z)}{\sqrt{(x - \min_p(x))^2 + (y - \min_p(y))^2}} \tag{1}$$

In this formula, i represents the gradient value of other points within the grid, Δh represents the elevation difference between the point of interest and the lowest point within the grid, and d represents the distance between the point of interest and the lowest point. The coordinates of the point of interest are denoted by x, y, z, while $\min_p(x)$, $\min_p(y)$, $\min_p(z)$ represent the coordinates of the lowest point within the grid.

(3) For each point within every grid, compare the gradient value with the gradient threshold to distinguish between ground points and non-ground points.

(4) Upon completion of the iteration, output the results.

2.1.2. Noise handling

After the filtering of ground points, the point cloud data may still contain outliers or noise points, which can affect the accuracy of subsequent semantic segmentation. Therefore, this paper employs a statistical filtering method for noise reduction. The basic idea of the statistical filtering method is to utilize statistical principles to perform statistical analysis on the neighborhood of each point and remove points that significantly differ from their neighboring points. The specific steps are as follows:

(1) Search for the k-nearest neighbors for each point.

(2) Calculate the average distance from each point to its k-nearest neighbors.

(3) Compute the global mean and standard deviation of distances based on the average distances of all points.

(4) Set a threshold (typically a multiple of the global mean and standard deviation), and identify points with average distances greater than this threshold as outliers and remove them.

2.2. Street tree extraction based on improved RandLA-Net

RandLA-Net [9] is an efficient and lightweight 3D point cloud semantic segmentation network. The network initially introduces random point sampling techniques, thereby avoiding complex sampling technologies and cumbersome preprocessing steps. To compensate for the potential information loss due to random sampling, RandLA-Net ingeniously designs a Local Feature Aggregator (LFA) module. This module not only effectively captures the geometric details of the point cloud but also gradually expands the receptive field of each point, thereby enhancing the network's effectiveness. This feature enables RandLA-Net to directly process large-scale 3D point cloud data and demonstrates broad applicability in practical applications. However, for the specific task of street tree extraction, the original network still has some shortcomings. Therefore, this paper will make targeted improvements to RandLA-Net, introducing a focal loss function and an Efficient Channel Attention (ECA) module to enhance the model's sensitivity to difficult-to-classify samples and strengthen the model's learning and expressive ability for street tree features, thereby achieving more efficient and accurate street tree extraction.

2.2.1. Loss equalization based on focus loss function

The RandLA-Net network model employs the cross-entropy loss function. In the official datasets used by RandLA-Net, such as the S3DIS dataset, the distribution of point clouds to be segmented across various categories is relatively balanced, allowing the cross-entropy loss function to perform effectively. However, in road point cloud data collected by vehicular LiDAR, data such as ground, buildings, and other miscellaneous objects constitute a larger proportion of the point cloud, while the number of street tree point clouds to be extracted is significantly less than other data. This leads to the issue that the continued use of the cross-entropy loss function results in insufficient training for the underrepresented street tree class, failing to meet the requirements for the segmentation of street tree point clouds.

When the target of interest constitutes a small proportion of the data to be detected, this presents an imbalanced classification problem. The focal loss function can effectively address such issues. By diminishing the impact of easy samples on the total loss, the focal loss function places greater emphasis on hard-to-classify samples, thereby mitigating the impact of class imbalance on network segmentation.

Cross-entropy loss function is given by:

$$L=\begin{cases} -\ln y', & y=1 \\ -\ln(1-y'), & y=0 \end{cases} \tag{2}$$

Focal loss function is given by:

$$L_{\mathrm{fl}}=\begin{cases} -\alpha(1-y')^{\gamma}\ln y', & y=1 \\ -(1-\alpha)y'^{\gamma}\ln(1-y'), & y=0 \end{cases} \tag{3}$$

In this formula, y' represents the predicted class probabilities given by the model, and y represents the true sample labels. Compared to the cross-entropy loss function, the focal loss function introduces a weight of $(1-y')^{\gamma}$. The larger the probability of the correct category, the smaller $(1-y')^{\gamma}$ becomes, indicating that the loss weight for correctly classified samples is smaller, and vice versa, the loss weight for incorrectly classified samples is larger. Additionally, the focal loss function includes a parameter α, which reduces the weight of positive or negative samples to address the issue of sample imbalance. This paper adopts a focal loss function with parameters $\alpha=0.3$ and $\gamma=1.5$.

2.2.2. ECA Attention Mechanism

ECA [10] is a lightweight and efficient attention mechanism designed to overcome the high computational complexity of traditional attention mechanisms, which is detrimental to real-time applications. ECA replaces the fully connected layer following global average pooling with a one-dimensional convolution, thereby achieving cross-channel information interaction. This not only retains the advantages of channel attention mechanisms but also significantly reduces computational overhead and the number of parameters. This characteristic makes ECA particularly suitable for embedding in deep neural networks that require efficient processing of large-scale point cloud data. In the one-dimensional convolution process, the size of the one-dimensional convolution kernel is not fixed but is determined by a function related to the number of channels. This adaptive function considers the number of channels, allowing the size of the convolution kernel to adjust automatically according to the number of channels. The specific adaptive function is:

$$k=\psi(C)=\left|\frac{\mathrm{lb}(C)}{\gamma}+\frac{b}{\gamma}\right| \tag{4}$$

In this formula, k represents the size of the one-dimensional convolution kernel, C represents the number of channels, γ and b are preset parameters, usually $\gamma=2$ and $b=1$.

In order to effectively integrate the ECA module into the RandLA-Net architecture, this paper improves the key feature extraction layer of the network. Specifically, an ECA attention module is inserted after each local aggregation module to dynamically adjust the feature response of each channel. The purpose of this design is to weight the locally aggregated feature maps through the ECA module, enhance those feature channels that are important for street tree recognition, and suppress irrelevant or noisy information, so as to improve the sensitivity and segmentation accuracy of the model to street tree categories.

3. Experiments and analysis of results

3.1. Overview of the study area and data introduction

This paper takes Yongchang Road in Beijing as the study area and uses a mobile vehicle-mounted laser scanner to obtain laser point cloud data of the scenes on both sides of the road. The LiDAR data of the study area covers a diverse range of ground features, including but not limited to the ground, various buildings, traffic facilities, natural vegetation, and dynamic targets such as moving vehicles, as shown in Figure 2. From the overall distribution of the point cloud data, the ground point cloud is continuous and extensive, covering the entire study area with slight undulations in terrain. There are multiple buildings distributed on both sides of the road, which, when analyzed from the XOY plane view of the point cloud data, generally present a strip shape with a certain directional arrangement. In addition, the point cloud data also includes green areas, where a variety of vegetation, such as tall trees and low shrubs, are densely distributed. These vegetations exhibit diverse point cloud morphologies with significant differences in elevation. Regarding traffic facilities, street lamps and utility poles appear in the point cloud data with regular cylindrical shapes, presenting a uniform and slender characteristic. Meanwhile, the point clouds of public facilities such as billboards and bus stops are also clearly visible. Although their elevation is relatively low, their sporadic distribution does not cause confusion with other ground features. Near the road, randomly distributed vehicle point clouds can be observed. These dynamic targets have non-fixed point cloud locations, usually lower than fixed installations such as street lamps, utility poles, and street trees.

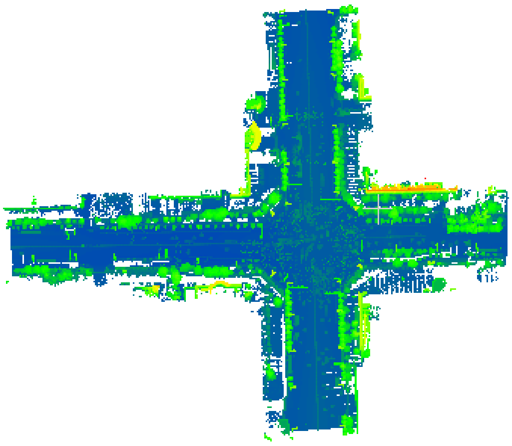
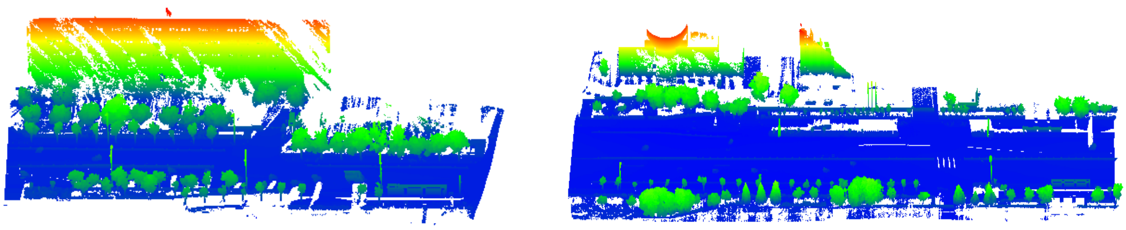


Figure 2. Study area points cloud data

3.2. Experiments and analysis of results

In this paper, the experimental environment is as follows: the CPU is i5-10200, the GPU is GTX1650, the memory is 16 GiB, and the deep learning framework is Pytorch. In order to verify the effectiveness of the proposed method, two sections of road data in the study area are selected for testing and analysis, among the selected two sections of point cloud data, the point cloud of street tree in experimental data 1 has more data and more dense distribution than that of other non-ground ground objects, and the point cloud of street tree in experimental data 2 has less data and more discrete distribution than that of other non-ground ground objects, as shown in Figure 3.

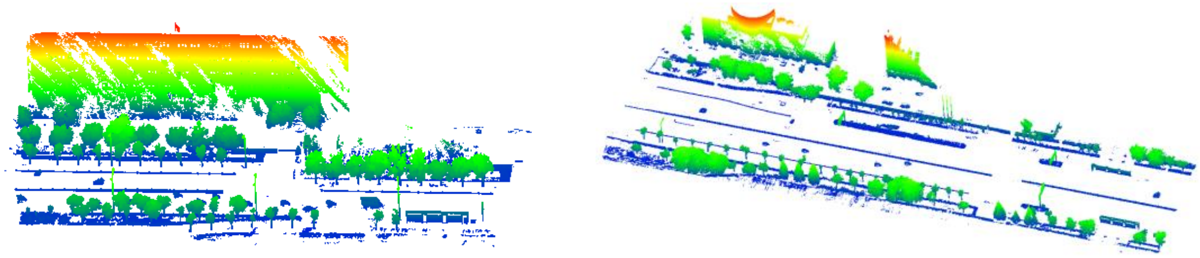


(a) Experimental data 1 (b) Experimental data 2

Figure 3. Experimental data

3.2.1. Data preprocessing

In this paper, the slope filtering algorithm is used to segment the ground point cloud to reduce the amount of data processed by the subsequent network. For some noise and outliers, a statistical filtering algorithm is used to filter them out. The parameters of the slope filtering algorithm are set to 2.0 meters for grid size and 0.1 for slope threshold, and the parameters for the statistical filtering algorithm are set to $k = 50$ and the threshold value is set to 1.5 times standard deviation. The preprocessing results are shown in Figure 4.



(a) Preprocessing results of experimental data 1 (b) Preprocessing results of experimental data 2

Figure 4. Experimental data preprocessing results

3.2.2. Substitution of the loss function

In this paper, Intersection over Union (IoU) and Accuracy were used as evaluation criteria. Intersection union ratio, which is the ratio of the intersection of the true value and the predicted value to the union of the true value and the predicted value. Accuracy, which is the percentage of the total number of samples that are correctly predicted. The cross-entropy loss function and the focus loss function (improved model 1) were used in the two groups of experiments, respectively, and the other experimental conditions, such as experimental data, training strategies and parameters, were the same, and the experimental results are listed in Table 1.

Table 1. Comparison of the segmentation results of improved model 1 and RandLA-Net model

Data	Model	IoU/%	mIoU/%	Accuracy /%
Experimental data 1	RandLA-Net	82.4	79.6	85.2
	Improved Model 1	89.3	90.8	91.5
Experimental data 2	RandLA-Net	76.5	78.8	80.6
	Improved Model 1	87.7	90.5	90.8

As can be seen from Table 1, after replacing the cross-entropy loss function in the original RandLA-Net network with the focus loss function, the accuracy of IoU and segmentation of the two sets of experimental data is greatly improved, and the experimental data 1 IoU is increased by 6.9% and the accuracy is increased by 6.3%. Experimental data2 IoU increased by 11.2% and accuracy by 10.2%.

3.2.3. ECA Attention Module

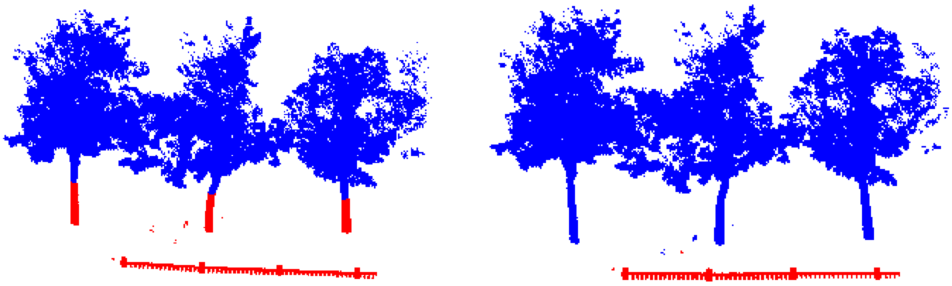
In the experiment, the original RandLA-Net network feature extraction structure and the improved feature extraction inserted into the ECA attention module were used respectively structure (improved model 2), and the other experimental conditions, such as experimental data, training strategy and parameters, are the same, and the experimental results are listed in Table 2.

Table 2. Comparison of the segmentation results of improved model 2 and RandLA-Net model

Data	Model	IoU/%	mIoU/%	Accuracy /%
Experimental data 1	RandLA-Net	82.4	79.6	85.2
	Improved Model 2	93.7	90.3	94.8
Experimental data 2	RandLA-Net	76.5	78.8	80.6
	Improved Model 2	90.9	92.5	92.8

As can be seen from Table 2, after inserting the ECA attention module into the original RandLA-Net network, the experimental data1 IoU is increased by 11.3% and the accuracy is increased by 9.6%. Experimental data2 IoU increased by 14.4% and accuracy by 12.2%.

Fig. 5 shows the local area diagram of the model segmentation results after adding the ECA attention module, and it can be seen from Fig. 5 that before adding the ECA attention module, the original model is not complete enough to learn the features of the street tree part, resulting in the misclassification of the tree part into other ground object classes, and after adding the ECA attention module, the model can learn the relevant features of the tree part well, so as to effectively solve the problem of tree part classification error.



(a) No ECA attention module (b) ECA attention module

Figure 5. Comparison of segmentation results

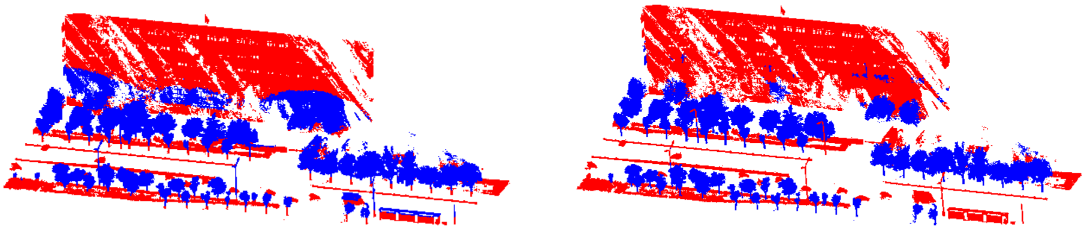
3.2.4. Focus loss function + ECA attention module

The original RandLA-Net network model is used for experiments, and then the model after replacing the loss function and adding the ECA module, that is, the improved model 3, is used to carry out the experiment. The other experimental conditions, such as experimental data, training strategies and parameters, are the same, and the experimental results are listed in Table 3.

Table 3. Comparison of the segmentation results of improved model 3 and RandLA-Net model

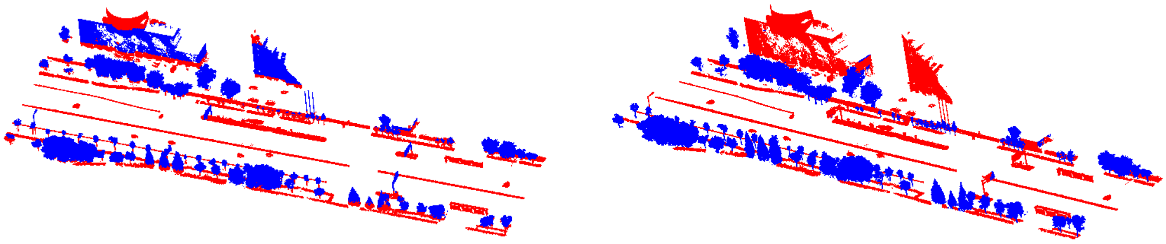
Data	Model	IoU/%	mIoU/%	Accuracy /%
Experimental data 1	RandLA-Net	82.4	79.6	85.2
	Improved Model 3	94.6	92.9	95.9
Experimental data 2	RandLA-Net	76.5	78.8	80.6
	Improved Model 3	92.8	93.7	94.2

It can be seen from Fig. 6 and Fig. 7 that when the original RandLA-Net model is segmented, a large number of building facades and street lamps and other clutter classes are mis divided into street trees, and the tree parts of street trees are mis divided into other clutter classes, and the improved model avoids this kind of problem well when segmentation, and greatly improves the segmentation accuracy. As can be seen from Table 3, the segmentation accuracy and IoU of the improved model in this paper are significantly improved compared with the original RandLA-Net model, in which the IoU value of experimental data 1 is increased by 12.2% and the accuracy is increased by 10.7%, and the IoU value of experimental data 2 is increased by 16.3% and the accuracy is increased by 13.6%, which further proves the effectiveness of the improvement strategy proposed in this paper.



(a) Segmentation results of the RandLa-Net model (b) Segmentation results of the improved model

Figure 6. Comparison of the segmentation results of experimental data 1



(a) Segmentation results of the RandLa-Net model

(b) Segmentation results of the improved model

Figure 7. Comparison of the segmentation results of experimental data 2

4. Conclusion

In this paper, an improved model based on RandLA-Net network is used to realize the segmentation task of street tree point cloud in road scenes. Firstly, the experimental data were preprocessed to complete the filtering of ground point clouds and the removal of noise points, which effectively improved the processing speed of subsequent tasks. Then, the following improvements are made for the RandLA-Net model: (1) the cross-entropy loss function in the original network is replaced by the focus loss function to enhance the sensitivity of the model to the samples that are difficult to classify, and solve the problem of unbalanced category proportion; (2) The ECA attention module is introduced to strengthen the learning and expression ability of the model on street tree features, and improve the segmentation accuracy of the model. Finally, through experiments conducted on a self-built dataset, the improved model proposed in this paper achieved a segmentation accuracy and IoU of 95.9% and 94.6% on experimental dataset 1, respectively, which is an increase of 10.7% and 12.2% compared to the original RandLA-Net model. On experimental dataset 2, the segmentation accuracy and IoU achieved were 94.2% and 92.8%, respectively, representing an improvement of 13.6% and 16.3% over the original RandLA-Net model. The model essentially completed the extraction task of street tree point clouds, demonstrating the feasibility of the improved model in the task of extracting street tree point clouds in road scenarios.

References

[1] Deng Yutong, Li Feng, Zhou Siqi, et al. Characteristic object extraction and 3D reconstruction of urban road based on vehicle point cloud data [J]. Journal of Beijing University of Technology, 2024, 50 (4): 498 - 507.

[2] Zhang Yan. Street trees and attribute information extraction from vehicle-borne laser point cloud [D]. Qingdao: Shandong University of Science and Technology, 2019.

[3] LI H, MENG T, ZHANG X M, et al. An improved conditional Euclidean clustering point cloud segmentation method [C]//International Conference on Computing, Control and Industrial Engineering. Singapore: Springer Nature Singapore. [S. l.: s. n.], 2023: 655 - 662.

[4] Zhang Liang. Research on street tree recognition method based on deep learning [D]. Nanjing: Southeastern University, 2019.

[5] Sun Guoqiang, Pangyan. Point cloud segmentation method of vehicle-mounted street tree based on PointCNN [J]. Beijing Surveying and Mapping, 2023, 37 (1): 61 - 66.

[6] Li Yongqiang, Dong Yahan, Zhang Xitong, et al. Point cloud information extraction for streetlights with vehicle-borne LiDAR [J]. Journal of Surveying and Mapping, 2018, 47 (2): 247 - 259.

[7] Li Qiujie, Zheng Jiaqiang, Zhou Hongping, et al. Point cloud recognition of street tree target based on variable-scale grid index and machine learning [J]. Transactions of the CSAM, 2018, 49 (6): 32 - 37.

[8] Li Yongqiang, Li Pengpeng, Dong Yahan, et al. Automatic extraction and classification of pole-like objects from vehicle LiDAR point cloud [J]. Journal of Surveying and Mapping, 2020, 49 (6): 724 - 735.

[9] HU Q Y, YANG B, XIE L H, et al. Randla-net: Efficient semantic segmentation of large-scale point clouds [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.: s. n.], 2020: 1 - 16.

[10] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.: s. n.], 2020: 11531 - 11539.