



BushNet: Effective semantic segmentation of bush in large-scale point clouds

Hejun Wei^a, Enyong Xu^{b,c,*}, Jinlai Zhang^a, Yanmei Meng^{a,*}, Jin Wei^a, Zhen Dong^a, Zhengqiang Li^a

^a College of Mechanical Engineering, Guangxi University, Nanning 530004, Guangxi, China

^b Dongfeng Liuzhou Motor Co., Ltd, Liuzhou 545005, Guangxi, China

^c College of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China



ARTICLE INFO

Keywords:

Deep learning
3D computer vision
Point cloud

ABSTRACT

Effective and robust semantic segmentation of bush is the fundamental problem of agroforestry environment understanding. However, the point cloud data of most large-scale agroforestry scenes is extremely large, and it is difficult to perform semantic segmentation on them. In order to realize the effective semantic segmentation of bush point cloud in large-scale agroforestry environment, this paper proposes BushNet, a novel point cloud segmentation network consists of three key components. Firstly, we propose the minimum probability random sampling module which can quickly and randomly sample a huge point cloud while avoiding the problem of random sampling easily causing re-sampling, reducing the consumption of computing resources and improving the convergence speed. Secondly, we propose the local multi-dimensional feature fusion module which makes the network more sensitive to bush point cloud features, thereby showing better bush segmentation performance. Thirdly, we propose the multi-channel attention module to achieve more accurate attention distribution and improved training efficiency. Experiments demonstrate that our approach significantly improves segmentation performance on multiple large-scale agroforestry point cloud data sets.

1. Introduction

Large-scale point cloud maps have become an indispensable part of robot navigation, drone navigation, autonomous driving and other fields. There have been more and more studies (Černáva et al., 2019; Liang et al., 2014; Pierzchała et al., 2018; Chen et al., 2019) using simultaneous localization and mapping (Durrant-Whyte and Bailey, 2006) (SLAM) technology to build agroforestry environmental maps to help the agriculture and forestry carry out related work. However, most point cloud maps can only allow robots to perceive the spatial location of objects, but the category of objects is unknown, which limits the performance of bush maintenance robots in automated maintenance. Realizing the semantic segmentation of the agroforestry environmental map is the fundamental problem of agroforestry environment understanding, especially the identification and segmentation of the plants that need to be maintained, and can help the robot to more accurately perceive the environmental information, and then achieve the optimal path planning and automatic maintenance. With the rapid development

of deep learning (LeCun et al., 2015), image semantic segmentation has also developed rapidly as a basic task. There are many successful algorithms (Long et al., 2015; He et al., 2017a; Ronneberger et al., 2015; Chen et al., 2017), however, point clouds have the characteristics of irregularity, disorder, unstructured, and large amount of data, the image semantic segmentation algorithms cannot be used directly on point clouds (Zhang et al., 2021a).

Pointnet (Qi et al., 2017a) has taken a big step forward in the process of point cloud semantic segmentation. It uses a globally shared multi-layer perceptron (MLP) to obtain effective information and achieve a better point cloud semantic segmentation performance. It cannot perceive the local feature information of the point cloud well, resulting in the loss of effective information, and the accuracy of segmentation is not high. Subsequently, some researchers have proposed a variety of local feature information perception modules optimized on the basis of Pointnet (Wang et al., 2018a; Defferrard et al., 2016; Shen et al., 2018; Klokov and Lempitsky, 2017), all of which have achieved good results. However, with the expansion of point cloud map applications, the scale

* Corresponding authors.

E-mail addresses: 13557729070@163.com (E. Xu), gxu_mengyun@163.com (Y. Meng).

of point cloud maps has risen to millions of points or even tens of thousands of square meters. Most point cloud segmentation algorithms can only be performed on small-scale point clouds due to excessive computing resources. The RandLAnet (Hu et al., 2020) solves this problem well. The point cloud is reduced by fast random downsampling. After the scale is reduced, the point cloud local feature learning is performed, and the receptive field is expanded through the superimposed local feature encoding module, which avoids the loss of information caused by random downsampling, and realizes the rapid segmentation of large-scale point clouds.

In the context of automated maintenance in agroforestry environment by robots through path planning, rapid semantic segmentation of bushes in large-scale point clouds is a key technology. Although the existing point cloud semantic segmentation algorithms can do the semantic segmentation of various types of objects, the segmentation performance of bush is not satisfactory. To the best of our knowledge, only a few algorithms (Wang et al., 2018a) can perform the segmentation of bushes well in large-scale point clouds but the efficiency of segmentation is relatively low. This paper proposes a novel point cloud segmentation network, named BushNet. Our main contribution lies in the design of multiple components that can improve the efficiency and accuracy of the bush point cloud segmentation in large-scale. Firstly, we propose the minimum probability random sampling module. Given the fact that the random sampling of point cloud used by current point cloud segmentation networks have the problem of re-sampling and lose important point information, a minimum probability random sampling method is proposed to optimize random sampling, which can reduce the consumption of computing resources and improving the convergence speed. Secondly, we propose the local multi-dimensional feature fusion module. Bushes have various and irregular shapes, and the normal vector characteristics of their point clouds are extremely obvious, which is different from other rules constituting objects. Adding simple points feature (SPF) can make the network more sensitive to bush point cloud features, thereby showing a better bush segmentation effect. Thirdly, we propose the multi-channel attention module, which can take into account the interactive information between multiple channels when inputting multi-channel features, which improves the rationality of feature weight distribution. Combining all the above components, the BushNet can better segment the bushes point cloud in a large-scale environment, and take into account the computation cost and efficiency.

2. Related work

In recent years, many scholars have also proposed a variety of point cloud segmentation networks to be applied to agricultural and forestry environments, which are mainly divided into the following categories: (1) Segmentation network based on multi-view and voxel; (2) Segmentation network based on graph optimization; (3) Point cloud based segmentation network.

2.1. Multi-view and voxel segmentation networks

Inspired by the convolutional neural network for image processing, in 2015, the Multi-view Convolutional Neural Network (MVCNN) (Su et al., 2015) was proposed by Su et al. This method uses multiple cameras arranged in different perspectives to capture 2D images to form multiple perspectives of the object. The 2D view is aggregated into a 3D descriptor through the set convolutional neural network, and then feature learning is performed to achieve classification. Subsequently, Feng et al. proposed the Group View Convolutional Neural Network (GVCNN) (Feng et al., 2018), which extracts view-level descriptors through Full Convolutional Networks (FCN), and then learns the connections between the views in groups and performs grouping to generate group-level descriptors and shape feature descriptors, and finally input the fully connected layer (FC) to complete the classification task.

The proposal of point cloud voxelization solves the unstructured and

disordered problems of point cloud. In this way, the voxelized point cloud can use 3D convolutional neural network CNN to extract semantic information. Also in 2015, Maturana et al. took the lead in mentioning VoxNet (Maturana and Scherer, 2015). Its core idea is to divide the entire point cloud into multiple 3D grids using the occupancy grid algorithm, and then normalize each grid unit, and finally enter the volume Build layers and maximum pooling layers. Gargoum et al. (2018) proposed a voxel-based method to identify the light poles of rural roads. First, the point cloud data was voxelized and the noise points were removed, and the three-dimensional clustering of the voxel data was performed through the connected region tags. The density-based method is used to combine the connected regions of the same object, and the light poles are extracted by geometric clustering. Zou et al. (2017) proposed a voxel-based deep learning method to identify tree species in a three-dimensional map. They extracted individual trees through point cloud density and used voxel rasterization to obtain features. Guan et al. (2015) used a voxel-based upward growth algorithm to remove the ground point cloud, and then segmented a single tree species by European clustering and a voxel-based normalization algorithm, and finally used waveform representation and depth Boltzmann machine Classify urban tree species.

Although voxelization solves the unstructured nature of 3D point clouds, other problems arise. For example, point clouds are only data on the surface points of the object. Except for the surface of the object, there are data holes in the rest of the 3D space. A simple conversion will cause unnecessary calculation costs. In addition, each voxel of the voxelized point cloud has a very regular shape. The size of the voxel greatly affects the resolution, and may also cause the loss of information. If the voxel is too small, it can be making the resolution high will cause a lot of calculation cost and memory usage, and too large voxels will cause the resolution to be too low, which will reduce the accuracy of point cloud classification. The network proposed in this paper directly operates on points, which can avoid information loss and memory resource occupation caused by voxel segmentation, and can obtain rich local and global features through the spatial position information of points, which is more helpful for bushes point cloud feature learning and recognition.

2.2. Graph segmentation networks

The excellent performance of graph convolutional neural networks in image processing has led some studies to try to apply it to 3D point clouds. The Mix-cut algorithm proposed by Golovinskiy and Funkhouser (2009) searches for the K nearest neighbors of key points and constructs the graph, and then uses Graph neural network for training, this algorithm realizes outdoor city target detection. Simonovsky and Komodakis (2017) proposed a method that uses edge labels to perform edge conditional convolution (ECC) on the neighborhood of a local graph, and uses an asymmetric edge function to calculate the spatial feature relationship between local points. In order to solve the problem that the greedy maximum pooling of most algorithm networks may cause the network to ignore some important information, Wang et al. replaced the maximum pooling with a recursive clustering strategy and proposed SpecGCN (Wang et al., 2018a). Point to construct a neighborhood map and perform spectrogram convolution, then perform continuous recursive clustering, and finally output the classification or segmentation results. Underwood et al. (2016) used Markov random field probabilistic ground model for segmentation on the three point cloud map of the almond orchard. The system can effectively draw the flower and fruit distribution map.

Although the related methods of graph optimization can obtain the local feature information of the point cloud better, the calculation is relatively expensive and it is difficult to apply to the semantic segmentation of large-scale point clouds. In this paper, we use minimum probability random sampling(MRS) to greatly increase the speed of point cloud sampling. At the same time, simple geometric calculations are used to directly extract the spatial characteristics of the key point K

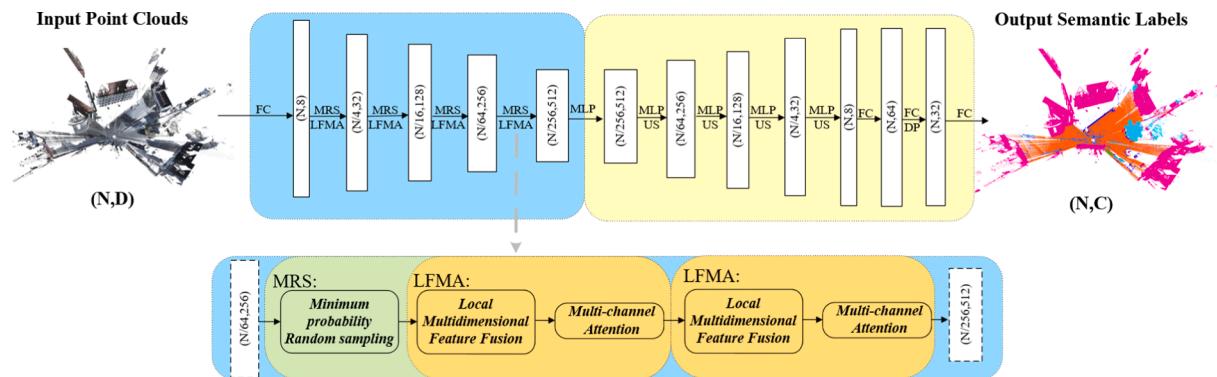


Fig. 1. BushNet's overall network structure diagram. (N, D) represents the number of input point cloud points and point cloud feature dimensions, (N, C) represents the number of input point cloud points and the number of output point cloud labels, FC is the fully connected layer, MRS is Minimum probability random sampling, and LFMA is a combination of two modules: Local Multidimensional Feature Fusion module and Multi-channel Attention module. US is up-sampling, and MLP is a multi-layer perceptron.

neighborhood, which greatly reduces the computation cost and reduces the memory usage.

2.3. Point cloud segmentation networks

PointNet (Qi et al., 2017a) is the first point cloud segmentation algorithm that directly operates on points. It uses a multi-layer perceptron (MLP) to approximate the function h , which represents the local features extracted by each point. PointNet connects the aggregated global features and the local features of each point (Concat). Subsequently, MLP extracts new features from the features of the merged points. Based on them, the semantic label corresponding to each point can be predicted. The original PointNet does not use any local structure information in adjacent points. In further research, Qi et al. (2017b) proposed PointNet++ that uses a hierarchical neural network to capture local geometric features to improve the original PointNet model. Due to the fact that K-nearest neighbor search in PointNet++ may cause K-nearest neighbors to fall into one direction, Jiang et al. (2018) designed PointSIFT to capture local features from eight directions. In the entire architecture, the PointSIFT module implements multi-scale representation by stacking several directional encoding (OE) units.

The attention layer of the convolutional neural network is very important. Some effective two-dimensional attention mechanisms have been proposed. For example, Xue et al. (2019) proposed a dual attention network (DANet) to adaptively correlate local features with their global

dependencies. Huang et al. (2019) proposed a crisscross network (CCNet) to obtain full image context information in a very effective and efficient manner. Ding et al. (2019) proposed an asymmetric convolution block (ACB), which is an architecture-neutral structure as a CNN building block that uses one-dimensional asymmetric convolution to strengthen the square convolution kernel. On the point cloud semantic segmentation, Peng et al. (2021) introduced a multi-attention semantic segmentation model MASS, which uses two novel attention and pillar attention to form a multi-attention (MA) mechanism to better aggregate features and improve the performance of dense top view semantics. In addition to the field of autonomous driving, point cloud semantics can also do some valuable things. For example, Liu et al. (2021) proposed HIDA, a lightweight assistance system based on 3D point cloud instance segmentation and solid-state LiDAR sensors, which can help the visually impaired people to detect and avoid the indoor obstacle.

Although point-based methods perform well on small-scale point clouds, they are still affected by computational cost. The above methods are still not suitable for large-scale point clouds. This paper proposes an effective and fast semantic segmentation network of bush point cloud in large-scale environment has found a good balance between accuracy and computational resource cost.

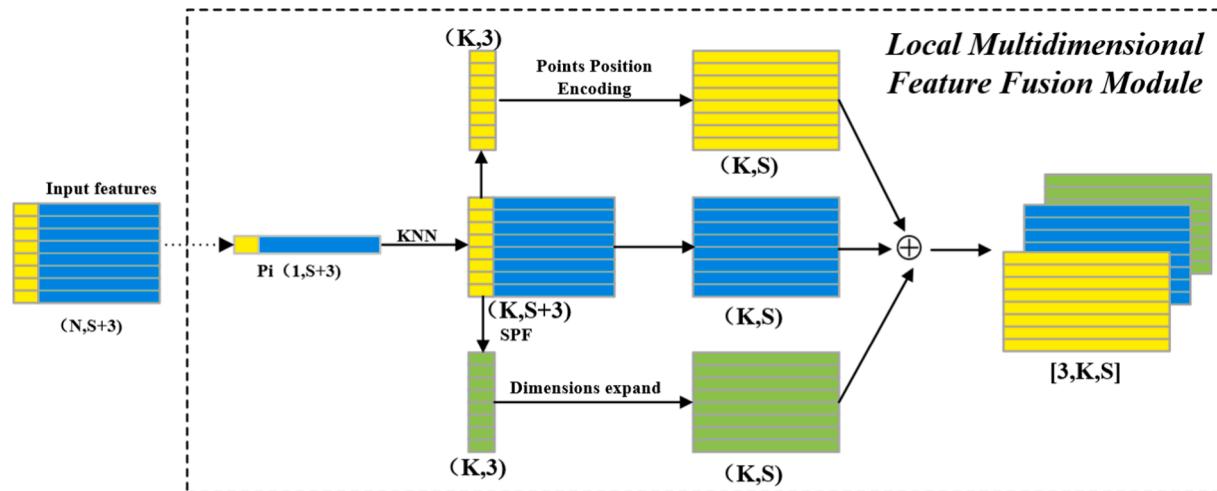


Fig. 2. Schematic diagram of the local multidimensional feature fusion module, P_i is a sampling point, N is the number of point sets, $(1, S+3)$ is the feature information dimension of the point, K is the number of neighbor points searched by the KNN algorithm, SPF is the simple points feature.

3. BushNet

3.1. Overview

The overall structure of the BushNet proposed in this paper is as shown in Fig. 1, which is the classic Encoder-Decoder structure. The point cloud input network first passes through a Fully Connected layer to increase the feature dimension to $(N, 8)$, where the N denotes the number of input point cloud points. Then after 4 layers of Minimum Probability Random Sampling (MRS) and stacked twice LFMA (Local Multidimensional Feature Fusion and Multi-channel Attention) module, the dimension of the deep feature information of the network reaches $(N/256, 512)$. And then through the shared MLP, continuous up-sampling to restore the original feature information, by the fully connected layer and Dropout layer (Srivastava et al., 2014) output the semantically segmented point clouds.

3.2. Minimum probability random sampling

In the literature (Liang et al., 2014), it has been proved that random sampling is the fastest method when facing large-scale point cloud downsampling, which can compare with the traditional methods such as Farthest Point Sampling (FPS) (Qi et al., 2017b) and Inverse Density Importance Sampling (IDIS) (Groh et al., 2018). The algorithm improves the sampling speed by more than a hundred times. Random sampling selects K points uniformly from the original N points, its computational complexity is $O(K)$, which has nothing to do with the total number of input points, that is, it is constant time, so it has inherent scalability. Compared with FPS and IDIS, random sampling has the highest computational efficiency regardless of the size of the input point cloud. However, random sampling also has certain drawbacks, that is, because of randomness, it is easy to lose important point information and repeated sampling. The problem of missing information is solved by the local multidimensional feature fusion module proposed in Section 3.3, and for the problem of repeated sampling, a minimum probability random sampling method is proposed to optimize random sampling (see Fig. 2).

First, it randomly assigns a probability to each point of the input original point cloud to obtain the point with the smallest probability as the center point, and search the n points nearest the center point through the pre-built KDtree (Foley and Sugerman, 2005). If the number of points in the scene is insufficient, find all points in the scene, and update the probability of the selected point at the same time. The closer the point is to the center, the greater the probability, which reduces the probability of being selected, so as to reduce the occurrence of repeated selection of points. The processing of randomly assigned probability and selected point probability change can be expressed as follows:

The probability ζ_i of randomly assigning point i is as follows:

$$\zeta_i = R(0, 1) \quad (1)$$

where $R(0, 1)$ is the function to generate random numbers from 0 to 1. Then we need to get the index containing the smallest probability point, as follows:

$$\text{cloudindex} = \text{index}(\min\{\zeta_1, \dots, \zeta_n\}) \quad (2)$$

Then update the point probability of the area near the candidate point to avoid repeated sampling points. The probability update can be expressed by the following formula:

$$\zeta'_i = \zeta_i + \sqrt{(1 - d_i/d_m)} \times W \quad (3)$$

where d_i is the Euclidean distance between the sampled point and the center point, d_m is the maximum Euclidean distance from the center point among the n nearest neighbors, and W is the weight of the sample point class in the whole data set.

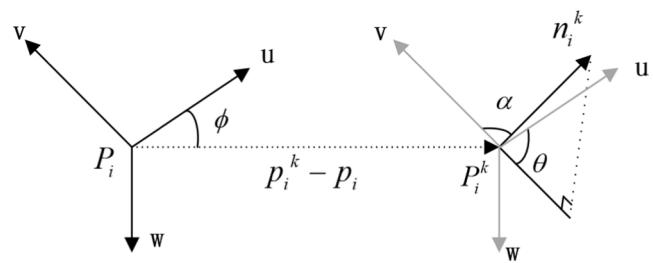


Fig. 3. The coordinate system of key points p_i and its neighboring points p_i^k , The α, Φ, θ are the three characteristic elements of the two-point normal vector.

3.3. Local multidimensional feature fusion module

The local multidimensional feature fusion module aims to aggregate the multidimensional local feature information in the K neighborhood of key points, so that the network can capture enough feature information for training to achieve better results. The multi-dimensionality we proposed mainly refers to spatial location features, RGB and simple points feature (SPF).

The previous point cloud segmentation network generally uses the relative position of the point cloud (spatial location features) for point cloud feature learning. Although the relative position information brings a better point cloud segmentation effect, the redundant and single relative position information still cannot reflect the spatial geometric characteristics of various types of objects. The structure of bushes is diverse and irregular, and the normal vector characteristics of the point cloud are extremely obvious, which is different from other regular objects. We therefore use three angular deviations of point cloud normal vector to capture the surface variation features in the neighborhood of key points, which we called the simple points feature (SPF). The specific calculation process is as follows:

Step1: Firstly, we use KDtree (Foley and Sugerman, 2005) to quickly determine the neighboring points of the key points, and convert the problem of solving the normal vector into a problem of fitting the points in the range to a plane by the least square method. Then, in our fitting plane optimization function, it is necessary to find a plane for all neighboring points, so that all neighboring points are on this plane, or the distance between all points and the plane is the smallest. Finally, similar to the PCA principle (Pearson, 1903), solving the shortest distance from all point clouds to the fitted local plane can be simplified as solving the smallest eigenvalue of the covariance matrix, selecting the eigenvector corresponding to the smallest eigenvalue, and performing unitization, then the vector is the point cloud normal vector.

Step2: In order to calculate the deviation (α, Φ, θ) between the normal vector of key point p_i and the normal vector of neighborhood points $p_i^k \in \{P_i^1, P_i^2, \dots, P_i^K\}$, a fixed coordinate system $v-u-w$ is first defined on the key point p_i , as shown in Fig. 3.

In Fig. 3, u is the normal vector of the key point. Based on it, a space coordinate system is established. The conversion relationship between the coordinate axes is obtained by the following formulas:

$$u = n_i \quad (4)$$

$$v = u \times \frac{(p_i^k - p_i)}{\|p_i^k - p_i\|} \quad (5)$$

$$w = u \times v \quad (6)$$

where $\|p_i^k - p_i\|$ is the Euclidean distance between the two points.

Step3: We establish the translation of the new coordinate system on the neighbor point p_i^k as the origin. After the coordinate system is established, the SPF of the key point p_i will be calculated. We firstly get the index of the key point p_i and its K neighbor point p_i^k . Then calculate the α, Φ, θ between each key point p_i and its neighboring points p_i^k as

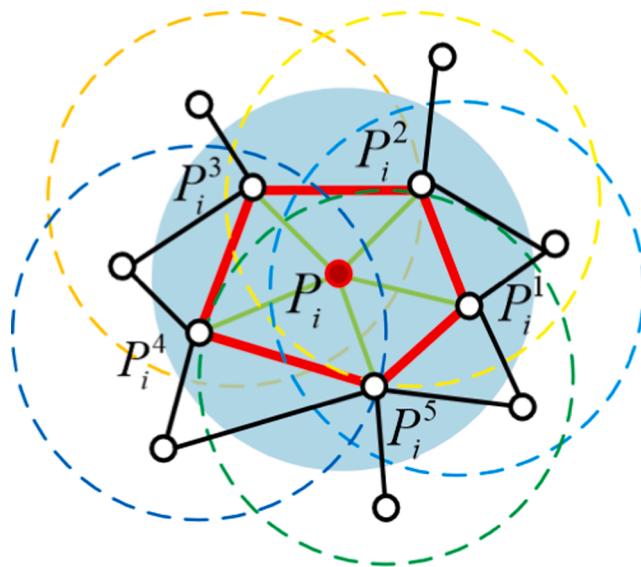


Fig. 4. The coverage area of SPF includes key points p_i and their neighbors p_i^k .

follows:

$$a_k = v \cdot n_i^k \quad (7)$$

$$\phi_k = u \cdot \frac{(p_i^k - p_i)}{d} \quad (8)$$

$$\theta_k = \arctan(w \cdot n_i^k, u \cdot n_i^k) \quad (9)$$

where $d = \|p_i^k - p_i\|$, n_i^k is the normal vector of p_i^k . let $SPF_k = (a_k, \phi_k, \theta_k)$, then aggregate all SPF into $(K, 3)$ -dimensional SPF feature, that is, the final SPF value of the key point. The SPF coverage area is shown in Fig. 4.

We use the SPF feature to be added in parallel to the local multi-dimensional feature fusion module, and the RGB and spatial position features of the point are fused into an effective point cloud local feature, so that the information contained in the local feature is richer, and the network can learn the bush features more effectively and sensitive to the semantic recognition of bushes, and at the same time, the receptive field of local features is improved to a certain extent, which makes up for the shortcomings that random sampling is easy to lose key information.

3.4. Multi-channel attention module

We argue that the attention mechanism that only considers the weight of each group of features can easily ignore the relationship between features. The algorithm in this paper has multi-channel local

features. Each category (including bushes) has certain rules for its own multi-channel features. If the relationship of multi-channel interaction is not considered, this may affect the distribution of feature weights and even the final classification. Inspired by ECA-net (Wang et al., 2020b), we propose a point cloud feature multi-channel attention module. After channel-level global average pooling without reducing the dimensionality, we use a fast one-dimensional convolution of size 3 to capture the cross-channel information. The basic network architecture of the multi-channel attention module is as follows (see Fig. 5):

The processing steps of the multi-channel attention module are as follows:

- (1) The input aggregation feature first passes through a global average pooling layer (GAP), which can be expressed by the following formula:

$$g(x) = \frac{1}{WH} \sum_{i=1,j=1}^{W,H} x_{ij} \quad (10)$$

where W is the width of the feature dimension, and H is the height of the feature dimension.

- (2) Let $y = g(x)$, then through a fast one-dimensional n -kernel convolution can calculate the channel weight, as follows:

$$w = \delta(C1D_n(y)) \quad (11)$$

where the $C1D_n$ denotes the one-dimensional convolution with n -kernel, and δ denotes the channel weight calculation function.

- (3) Finally, through the summation module and the shared multi-layer perceptron (MLP) output point neighborhood's local feature information.

After the local multi-dimensional feature fusion is completed, the K -nearest neighbor feature of the key point contains the three parts of RGB, spatial position features, and SPF . This feature is spliced to form a new three-channel aggregation feature and then input to the multi-channel attention pooling layer. After learning the weights of each channel feature, the network generates spatial features with attention weights through a weighting module, and then aggregates the local features of the key points through a shared MLP and saves them locally.

4. Experiment

Our experiment is run on a computer with AMD Ryzen 7 4800H CPU, 16 GB RAM and a Nvidia RTX 2060 GPU with 6 GB memory. To train the BushNet, we use the Adam optimizer (Kingma and Ba, 2015). And the initial learning rate is 0.01 and reduced by 5% after each epoch. The number of closest points K is set to 16.

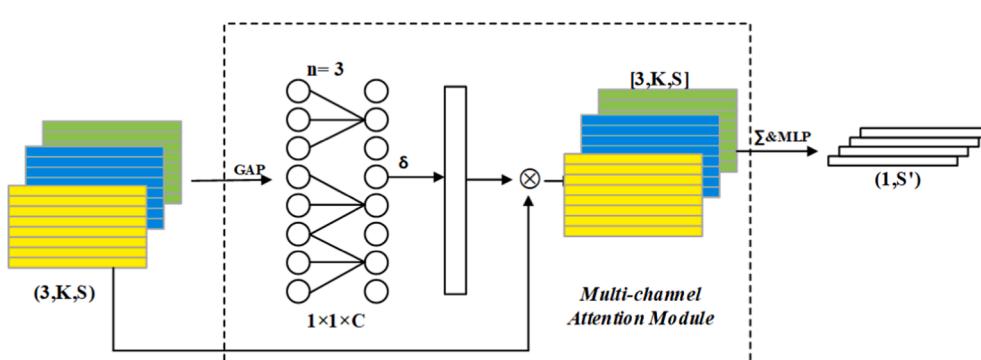


Fig. 5. Schematic diagram of the multi-channel attention module, $(3, K, S)$ is the feature dimension after local multi-dimensional feature fusion, and C is the feature channel dimension, GAP is the global average pooling, and n is the kernel size of convolution, σ represents the channel weight, $(1, S')$ is the feature dimension after attention weight allocation.

Table 1

Experiments on the Semantic 3D dataset.

	MIOU(%)	OA(%)	Man made.	Natural.	High veg.	Bush	Buildings	Hard scape	Cars	Scanning art.
ShellNet (Zhang et al., 2019)	68.6	92.2	96.0	89.0	83.6	40.0	93.8	34.2	69.5	42.8
GACNet (Wang et al., 2019)	70.6	90.8	86.8	79.4	87.8	60.1	94.0	36.7	77.1	43.2
SPG (Landrieu and Simonovsky, 2018)	71.4	93.4	96.7	91.9	87.0	43.5	83.1	30.8	75.4	63.1
KPConv (Thomas et al., 2019)	73.1	92.3	88.1	80.6	81.6	45.1	94.2	38.0	80.0	77.0
RandLA-Net (Hu et al., 2020)	74.0	93.7	95.0	89.2	82.1	50.1	94.5	39.7	78.1	63.3
Ours	76.3	94.5	94.7	89.5	82.3	62.1	95.2	41.0	78.9	66.7

4.1. Evaluation metrics

We use IOU(Intersection over Union), MIOU(Mean Intersection over Union) and OA(Overall Accuracy) to compare carious algorithms. IOU represents the ratio of the intersection of the prediction result *Pre* and the true value *Ture* of a certain category by the model to their union, which can be calculated by follows:

$$IOU = \frac{Pre \cap Ture}{Pre \cup Ture} \times 100\% \quad (12)$$

MIOU represents the ratio of the intersection and union between the predicted result of each category and the true value of the model. Assuming there are m categories, the result of summing and re-averaging can be obtained by following formula.

$$MIOU = \frac{1}{m} \sum_{i=1}^m IOU_i^m \quad (13)$$

We suppose that there are m categories of objects to be classified. When classifying the i_{th} category, a weight $\{W_1^i, W_2^i, \dots, W_l^i, \dots, W_m^i\}$ will be assigned to all categories, and the highest weight is used as the classification result, and the weight obtained for this category is W_1^i . OA can be obtained by following equation.

$$OA = \frac{\sum_{i=1}^m W_i^i}{\sum_{i=1}^m (W_1^i + W_2^i + \dots + W_k^i)} \times 100\% \quad (14)$$

The above three indicators are recognized as important indicators for evaluating the performance of semantic segmentation, which can well reflect the accuracy and robustness of algorithms, and are widely used by the computer vision community (Qi et al., 2017a,b; Zhang et al., 2021a; Thomas et al., 2019; Zhang et al., 2020, 2021b).

4.2. Experiments on the semantic 3D dataset

Semantic 3D (Hackel et al., 2017) is a large outdoor scene point cloud data set, covering a series of rich scenes such as urban streets, churches, squares, villages, castles, etc., providing a total of 8 semantic labels. Including artificial terrain (pavement), natural pavement (grass),

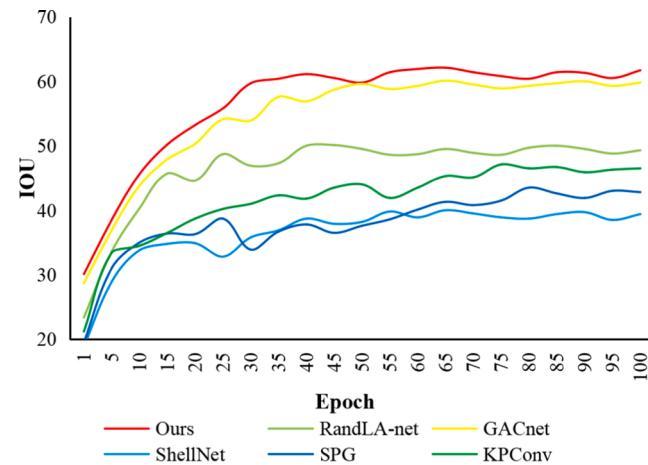


Fig. 7. The change of IOU with epoch for bush point cloud segmentation of different algorithms.

high vegetation (trees and large shrubs), bushes, buildings, artificial landscapes (fountains, etc.), scanning artifacts, cars and unmarked points. There are a total of 30 sub-scene data sets, 15 training sets and 15 test sets, with more than 1 billion points in the point cloud. This data set is rich and comprehensive. This section mainly tests the BushNet proposed in this paper on this data set, and compare it with the mainstream point cloud segmentation methods, such as ShellNet(Zhang et al., 2019), GACNet(Wang et al., 2019), SPG(Landrieu and Simonovsky, 2018), KPConv(Thomas et al., 2019), and RandLA-Net(Hu et al., 2020) in MIOU, OA, various object's IOU. The results are shown in Table 1.

It can be seen from Table 1 that the MIOU and OA of our BushNet is better than all the above algorithms. Compared with the RandLA-net algorithm, the MIOU is improved by 2.3%, and the OA is improved by 0.8%. It is worth noting that our accuracy is satisfactory on the bush's IOU. Although there is a slight decline in the accuracy of some classifications, the improvement has made a certain contribution to the accurate bush point cloud segmentation.

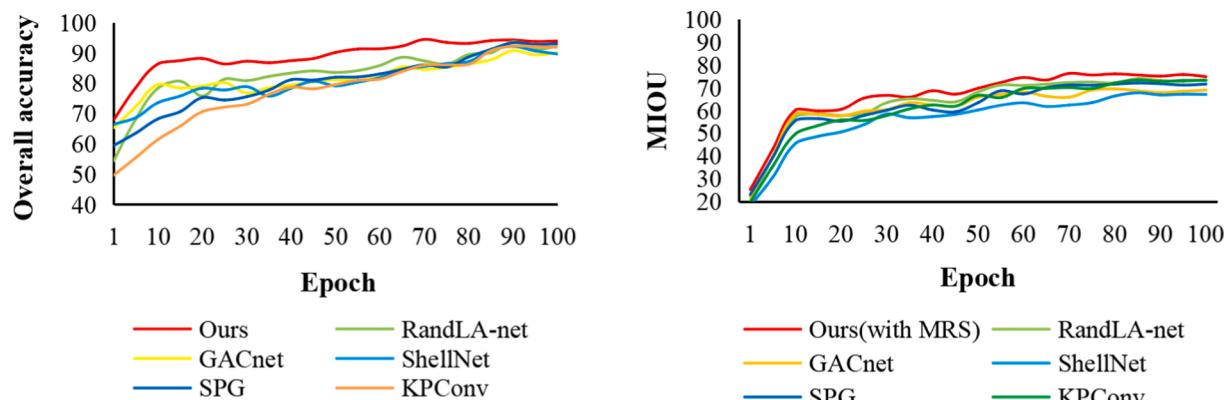


Fig. 6. Changes of MIOU and OA of different algorithms with epoch.

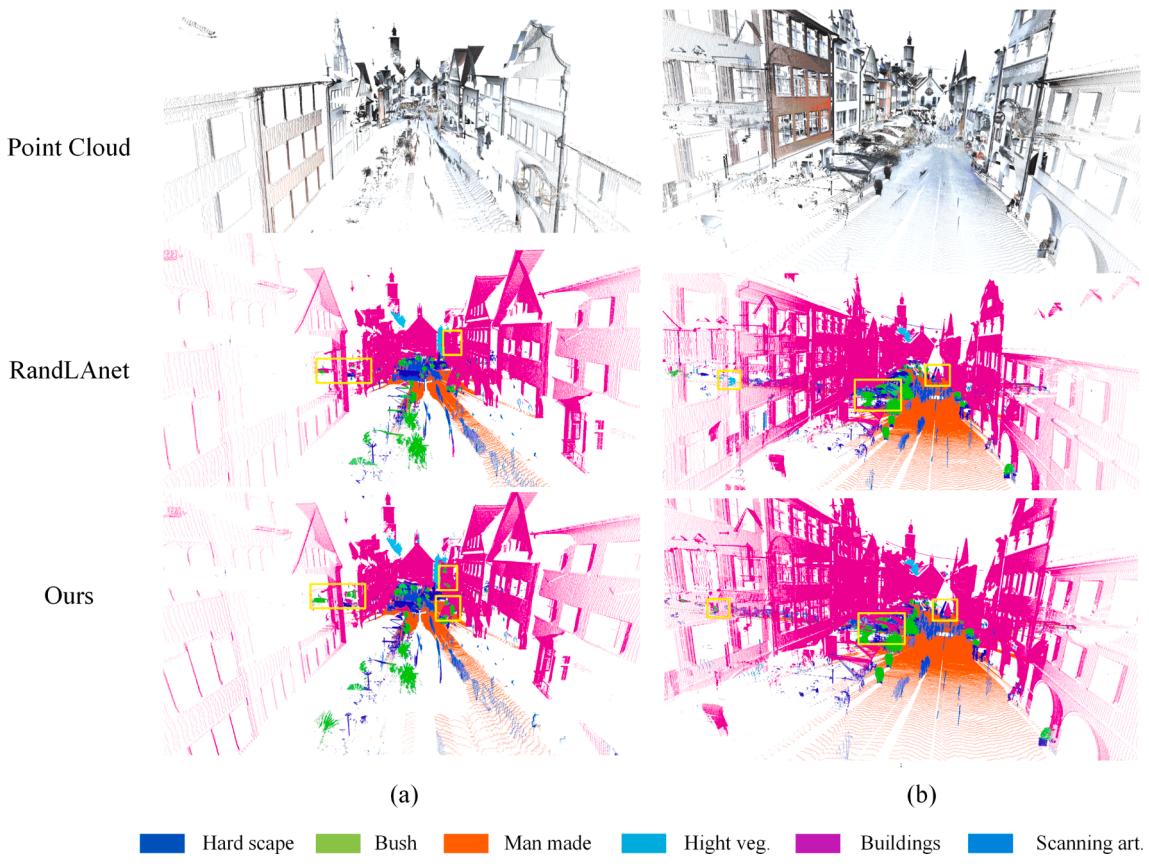


Fig. 8. Comparison of BushNet and RandLA-net semantic segmentation performance.

At the same time, we present the OA and MIOU convergence curves of each algorithm during the training process as shown in Fig. 6 and 7. From Fig. 6, we can see that our BushNet can quickly converge to the best state in these two dimensions and the growth curve does not fluctuate much. It can reach the convergence state around the 65th epoch of training (OA is 94.5%, MIOU is 74.5%). The other algorithms converge slowly and are prone to large fluctuations. In Fig. 7, the bush point cloud segmentation IOU of our BushNet is better than all other algorithms, and can reach the optimal convergence at the 55th epoch, which is a greater improvement than other algorithms.

In Fig. 8, in each point cloud scene, we marked the yellow box in the part with obvious effect contrast. From the (a), we can see that RandLA-net has missing segmentation in the yellow marked box. The main scenes in this range are bushes and buildings. RandLA-net has omitted the bushes in this space, and all of them are missing and marked as a building, and our algorithm can clearly distinguish between these two categories and separate them perfectly. RandLA-net has over-segmentation in the yellow mark box in (b), which incorrectly enlarges the segmentation range of bushes, which affects the classification of other categories of objects. In contrast, the segmentation boundary of

our algorithm is reasonable and clear, and it does not affect others categories.

4.3. Experiments on the RELLIS 3D dataset

The data of RELLIS 3D (Jiang et al., 2020) was collected at Texas A&M University in the United States. This is a multi-modal data set collected in a field environment. It contains 13,556 LiDAR scans and 6,235 image annotations. Among them, the point cloud labels are 14 categories. We chose this data set because the data set has a large number of bushes in the point clouds. The motivation of the BushNet in this paper is to improve the accuracy of bush segmentation in a large environment. We believe that this data set can verify the accuracy of the BushNet in this paper. We still choose the recently proposed ShellNet, KPconv, GACNet, SPG, and RandLA-net algorithms to compare with BushNet. They perform well on the Semantic3D public data set, especially the bush segmentation accuracy of GACNet reaches 60%. We take MIOU, OA, and IOU of each category as the main evaluation indicators. The experimental results are shown in Table 2.

From Table 2, it can be seen that the BushNet proposed in this paper

Table 2

Comparison of point cloud semantic segmentation algorithms for 11 types of objects on the RELLIS3D dataset. Evaluation indicators include MIOU, OA, and category level segmentation accuracy. Note that the original dataset has 14 types of annotations, but because of the number of 3 types of annotations are too little, so we do not include them.

	MIOU(%)	OA(%)	Grass	Tree	Vehicle	Log	Person	Bush	Concrete	Barrier	Puddle	Mud	Rubble
ShellNet (Zhang et al., 2019)	38.9	56.5	41.3	80.2	32.6	3.8	83.8	57.0	21.2	72.0	8.1	7.5	20.4
GACNet (Wang et al., 2019)	41.9	58.7	45.3	84.9	38.9	7.8	85.3	61.0	23.4	75.3	10.3	8.3	21.4
SPG (Landrieu and Simonovsky, 2018)	40.8	59.9	46.5	83.6	36.7	5.1	83.4	48.3	28.4	74.6	11.6	6.9	23.5
KPConv (Thomas et al., 2019)	27.4	23.8	56.3	49.3	2.1	1.0	81.3	57.3	33.2	3.5	0.8	5.3	11.2
RandLA-Net (Hu et al., 2020)	42.2	61.8	44.1	81.8	40.8	5.3	88.1	58.4	23.5	72.1	6.5	11.1	32.3
Ours	44.2	63	55.6	80.7	38.5	4.6	86.4	69.8	24.7	74.2	5.8	12.6	33.1

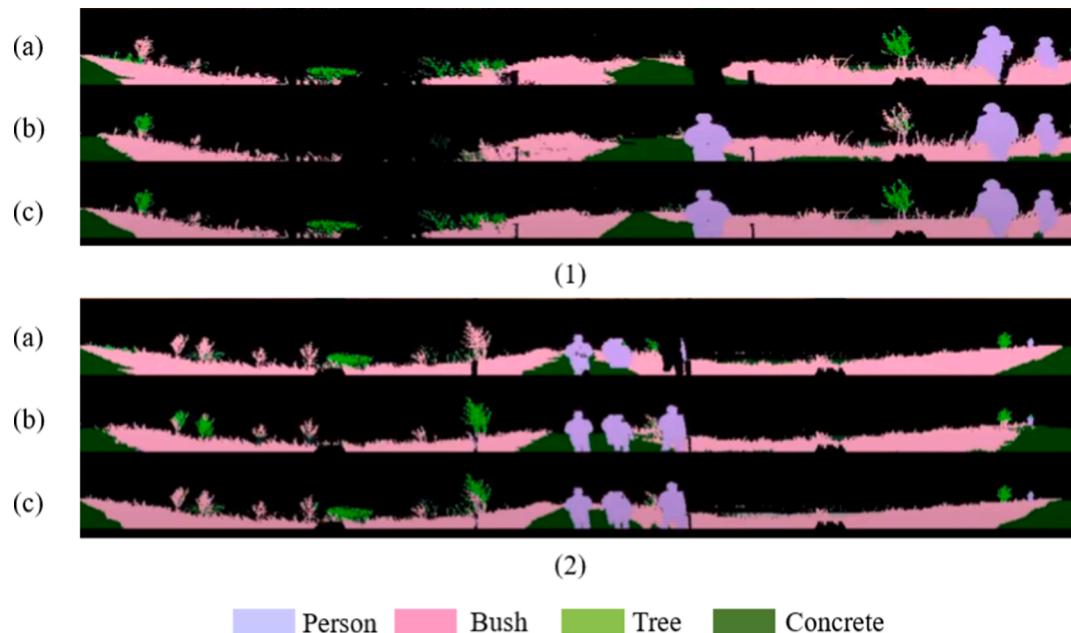


Fig. 9. Comparison of segmentation performance on the RELLIS3D dataset. (a) is the original point cloud annotation, (b) is the segmentation results of RandLA-net, (c) is the segmentation results of the BushNet. We can see that the segmentation of the algorithm in the Bush class is more complete and basically consistent with the ground truth.

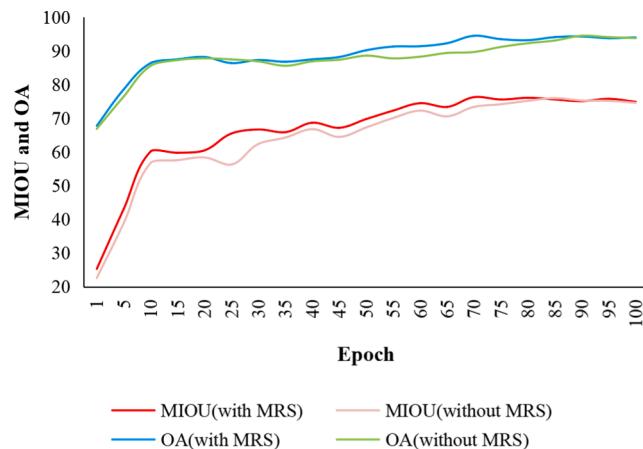


Fig. 10. Effects of Minimum probability random sampling.

performs best in the dimensions of MIOU, OA, Bush, Mud and Rubble, which are improved by 1.5%, 1.2%, 11.4%, 1.5% and 0.9% respectively compared with the RandLA-net. The segmentation accuracy of bush we care about has been improved very well, which also shows that the BushNet can show good semantic segmentation performance on bushes no matter in urban, rural or field environments. Comparison of segmentation effects on the RELLIS3D dataset are shown in Fig. 9.

4.4. Ablation study

All experiments in this section are trained on 3 scene point cloud data including stgallencathedral_station1-3 of Semantic3D, and tested in stgallencathedral_station6.

4.4.1. Effects of minimum probability random sampling (MRS)

In order to verify that our proposed minimum probability random sampling improves the convergence speed of BushNet training, we record the convergence of the minimum probability random sampling (MRS) and the BushNet without MRS, as shown in Fig. 10, the Bush-Net

Table 3
Effects of Local multidimensional feature fusion module.

Method	OA(%)	MIOU(%)	Bush(%)
RandLA-Net(SPO + RGB)	93.7	74.0	50.1
Ours(RGB + SPF)	76.3	58.7	53.9
Ours(SPO + SPF)	83.2	66.9	56.7
Ours(RGB + SPO + SPF)	94.1	75.2	58.3

with MRS can reach the optimal convergence state after training to about 70th epoch on both MIOU and OA, but the optimal convergence of BushNet without MRS is delayed to 80th epoch. It can be seen that MRS is of great help to the convergence speed of the algorithm.

4.4.2. Effects of local multidimensional feature fusion module

In order to verify the effectiveness of our local multi-dimensional feature fusion module, we conducted a total of four sets of ablation experiments on the local multi-dimensional feature fusion module, including a combination of RGB, spatial position features (SPO) and SPF three feature information: (1) RandLA-Net Local features include RGB and SPO, (2) BushNet's local features include RGB, SPO and SPF, (3) Bushnet's local features include RGB and SPF, (4) Bushnet's local features include SPO and SPF. the experimental results are shown in Table 3. It is proved that the newly added SPF feature is the main influencing factor of OA, MIOU and bush segmentation accuracy. It can be seen that after incorporating the new SPF features, the OA increased by 0.4%, MIOU increased by 1%, and the IOU of bushes increased by 8.2%. It proved that the new local multi-dimensional feature fusion was

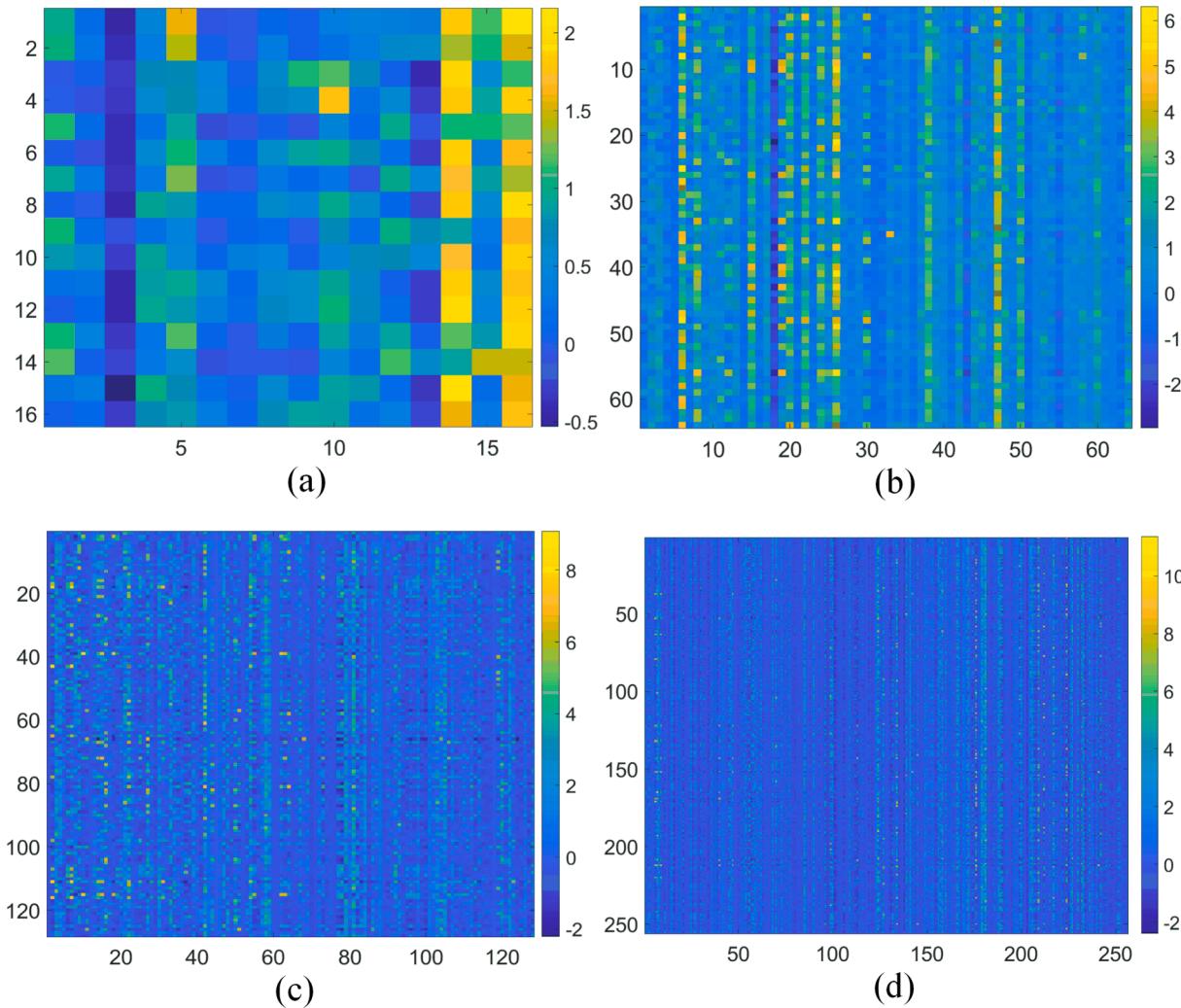
Table 4
Effects of Multi-channel attention module.

Attention method	OA(%)	MIOU(%)	Bush(%)
Att(Hu et al., 2020)	88.5	75.1	58.3
MAM(Ours)	94.5	76.3	62.1
ECA-Net(Wang et al., 2020a)	92.7	74.9	59.8
SE(Hu et al., 2019)	92.3	74.5	58.7
Non-local attention(Wang et al., 2018b)	89.6	73.4	56.7
Axial attention(Ho et al., 2019)	90.1	74.3	57.5

Table 5

The computational cost of each algorithm.

	Parameters(millions)	Inference Time(ms)	MIOU(%)	Bush(%)	Maximum points(millions)	Memory(G)
PointNet(Qi et al., 2017a)	0.23	270	18.5	36.3	0.49	3.28
PointNet++(Qi et al., 2017b)	0.31	428	21.6	39.7	0.98	3.75
PointCNN(Li et al., 2018)	3.5	1323	25.8	41.3	0.04	–
ShellNet(Zhang et al., 2019)	0.15	249	36.2	50.5	0.49	2.78
GACNet(Wang et al., 2019)	0.18	10875	39.1	55.6	0.01	–
Kpconv(Thomas et al., 2019)	4.3	313	25.6	42.7	0.52	3.53
SPG(Landrieu and Simonovsky, 2018)	0.12	10550	38.2	51.8	0.01	–
RandLA-net(Hu et al., 2020)	0.25	66	40.7	53.4	1.03	2.77
Ours	0.36	81	42.3	60.1	0.96	2.86

**Fig. 11.** The attention weight map under 4 different feature dimensions. The dimensions of (a) to (d) are 16*16, 64*64, 128*128, 256*256, where yellow represents a higher attention weight.

more effective and achieved satisfactory bush segmentation accuracy.

4.4.3. Effects of multi-channel attention module

We conducted a comparative experiment of BushNet using different attention method to Table 4. As shown in Table 4, the MAM proposed in this paper is more effective than other attention mechanisms. It increases BushNet's OA by 1.8%-6%, MIOU by 1.2%-2.9%, and Bush's IOU by 2.3%-5.4%. This result shows that MAM can learn effective features in multi-channel features and assign corresponding attention scores, so that the entire network can learn effective features.

4.5. Discussion

We further evaluate the efficiency of BushNet. The parameters and inference time(per frame) can reflect this problem well, which are the commonly used in the computer vision community (He et al., 2016; Huang et al., 2017; He et al., 2017b; Ren et al., 2015). So in this section we will evaluate the parameters, inference time(per frame) of each algorithm. We implement this experiment on the 0002 sequence of RELLIS 3D. Since it will take a lot of time and cost to evaluate the entire data set, the 0002 sequence has the most abundant point cloud information, which contains a large number of bush point clouds, therefore we chose

this sequence as a compromise. In addition, we also measured the maximum number of input points for each algorithm in the inference to prove that BushNet can be used in a large-scale environment.

From the experimental results in Table 5, we can know that KPconv has the largest amount of parameters and the smallest is SPG. However, the SPG inference takes a lot of time due to the related operations of the graph structure. The RandLA-net and the BushNet in this paper have reached a good balance of inference time and performance. RandLA-net has the fastest inference time, which can reach 66 ms. BushNet's inference time can also reach 81 ms, which is several times faster than the inference speed of other networks. Our BushNet has superior accuracy and inference speed while the memory requirement only slightly increases, because BushNet adds more feature information, which increases the memory requirements by 3% compared to RandLA-net, but it is still within an acceptable range.

Moreover, the attention maps or feature maps can give us a more intuitive understanding of how the network learns data features. However, the BushNet only learns features from 16 local points at a time. Such a small range cannot be visualized as a specific category, so we extract the attention weight of feature encoding layer to visually analyzed. As shown in Fig. 11, we visualize the attention weight map under 4 different feature dimensions. The attention weight tends to give a higher weight to the 14–16 channel features in the first encoding layer, and then there are more channels have gained some attention. This shows that attention pooling tends to choose prominent or key point features at the beginning. After the point cloud is continuously down-sampled and the feature dimension is continuously deepened, the attention layer tends to retain more features of these points.

At the same time, in terms of the maximum number of input points in the inference, BushNet can handle a point cloud scale close to 10^6 , proving that it can perform semantic segmentation in a large-scale point cloud environment. In summary, the BushNet improves the MIOU and IOU of bushes, and the increase in the amount of parameters and inference time is within an acceptable range, achieving a good balance between accuracy and computational cost.

5. Conclusion

This paper proposes BushNet, an algorithm for the semantic segmentation of bush point clouds in large-scale environments. We integrate SPF features into the local feature aggregation module, which can better capture the local feature information of bush point clouds, and introduce the multi-channel attention module that considers the interaction between multiple features. The algorithm can significantly improve the segmentation accuracy while ensuring good computation cost and memory usage. The segmentation accuracy of bushes has been improved by 12%, and multiple ablation experiments and data set verification have been carried out. The results prove the superiority of our algorithm. Regrettably, there are very few open source point cloud datasets with bush annotations. In the future, we will work to build a rich bush annotation point cloud dataset, further optimize the semantic segmentation performance, and improve the efficiency of BushNet. We hope that the algorithm in this paper can promote the automation of bush conservation to a certain extent, and also provide some ideas for other related researchers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Cerňava, J., Mokroš, M., Tuček, J., Antal, M., Slatkovská, Z., 2019. Processing chain for estimation of tree diameter from gnss-imu-based mobile laser scanning data. *Remote Sens.* **11**, 615.
- Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587.
- Chen, X., Milioto, A., Palazzolo, E., Gigliére, P., Behley, J., Stachniss, C., 2019. SuMa++: Efficient LiDAR-based Semantic SLAM. In: Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS).
- Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inform. Process. Syst.* **29**, 3844–3852.
- Ding, X., Guo, Y., Ding, G., Han, J., 2019. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1911–1920.
- Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localization and mapping: part i. *IEEE Robot. Autom. Mag.* **13**, 99–110.
- Feng, Y., Zhang, Z., Zhao, X., Ji, R., Gao, Y., 2018. Gvcnn: Group-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 264–272.
- Foley, T., Sugerman, J., 2005. Kd-tree acceleration structures for a gpu raytracer. In: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware, pp. 15–22.
- Gargouri, S.A., Koch, J.C., El-Basyouny, K., 2018. A voxel-based method for automated detection and mapping of light poles on rural highways using lidar data. *Transp. Res.* **2672**, 274–283.
- Golovinskiy, A., Funkhouser, T., 2009. Min-cut based segmentation of point clouds. In: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops. IEEE, pp. 39–46.
- Groh, F., Wieschollek, P., Lensch, H., 2018. Flex-convolution (million-scale point-cloud learning beyond grid-worlds). arXiv preprint arXiv:1803.07289.
- Guan, H., Yu, Y., Ji, Z., Li, J., Zhang, Q., 2015. Deep learning-based tree classification using mobile lidar data. *Remote Sens. Lett.* **6**, 864–873.
- Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. Semantic3d.net: A new large-scale point cloud classification benchmark. arXiv preprint arXiv:1704.03847.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017a. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017b. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp. 2961–2969.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
- Ho, J., Kalchbrenner, N., Weissenborn, D., Salimans, T., 2019. Axial attention in multidimensional transformers. arXiv:1912.12180.
- Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2019. Squeeze-and-excitation networks. arXiv:1709.01507.
- Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. Randla-net: Efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11108–11117.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
- Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W., 2019. Ccnet: Criss-cross attention for semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 603–612.
- Jiang, M., Wu, Y., Zhao, T., Zhao, Z., Lu, C., 2018. Pointsift: A sift-like network module for 3d point cloud semantic segmentation. arXiv preprint arXiv:1807.00652.
- Jiang, P., Osteen, P., Wigness, M., Saripalli, S., 2020. Rellis-3d dataset: Data, benchmarks and analysis. arXiv preprint arXiv:2011.12954.
- Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: ICLR (Poster).
- Klokov, R., Lempitsky, V., 2017. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 863–872.
- Landrieu, L., Simonovsky, M., 2018. Large-scale point cloud semantic segmentation with superpoint graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4558–4567.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* **521**, 436–444.
- Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018. Pointcnn: Convolution on \mathcal{X} -transformed points. arXiv:1801.07791.
- Liang, X., Hyppä, J., Kukko, A., Kaartinen, H., Jaakkola, A., Yu, X., 2014. The use of a mobile laser scanning system for mapping large forest plots. *IEEE Geosci. Remote Sens. Lett.* **11**, 1504–1508.
- Liu, H., Liu, R., Yang, K., Zhang, J., Peng, K., Stiefelhagen, R., 2021. Hida: Towards holistic indoor understanding for the visually impaired via semantic instance segmentation with a wearable solid-state lidar sensor. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1780–1790.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440.
- Maturana, D., Scherer, S., 2015. Voxnet: A 3d convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 922–928.

- Pearson, K., 1903. I. mathematical contributions to the theory of evolution.—xi. on the influence of natural selection on the variability and correlation of organs. *Philos. Trans. Roy. Soc. Lond. Ser. A Contain. Pap. Math. Phys. Character* 200, 1–66.
- Peng, K., Fei, J., Yang, K., Roitberg, A., Zhang, J., Bieder, F., Heidenreich, P., Stiller, C., Stiefelhagen, R., 2021. Mass: Multi-attentional semantic segmentation of lidar data for dense top-view understanding. arXiv preprint arXiv:2107.00346.
- Pierzchala, M., Giguère, P., Astrup, R., 2018. Mapping forests using an unmanned ground vehicle with 3d lidar and graph-slam. *Comput. Electron. Agric.* 145, 217–225.
- Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 652–660.
- Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: *Adv. Neural Inform. Process. Syst.*, p. 30.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Information Process. Syst.* 28, 91–99.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.
- Shen, Y., Feng, C., Yang, Y., Tian, D., 2018. Mining point cloud local structures by kernel correlation and graph pooling. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4548–4557.
- Simonovsky, M., Komodakis, N., 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3693–3702.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E., 2015. Multi-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE international conference on computer vision, pp. 945–953.
- Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6411–6420.
- Underwood, J.P., Hung, C., Whelan, B., Sukkarieh, S., 2016. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. *Comput. Electron. Agric.* 130, 83–96.
- Wang, C., Samari, B., Siddiqi, K., 2018a. Local spectral graph convolution for point set feature learning. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 52–66.
- Wang, L., Huang, Y., Hou, Y., Zhang, S., Shan, J., 2019. Graph attention convolution for point cloud semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10296–10305.
- Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020a. Eca-net: Efficient channel attention for deep convolutional neural networks. arXiv:1910.03151.
- Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020b. Eca-net: efficient channel attention for deep convolutional neural networks, 2020 IEEE. In: *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.
- Wang, X., Girshick, R., Gupta, A., He, K., 2018b. Non-local neural networks. arXiv: 1711.07971.
- Xue, H., Liu, C., Wan, F., Jiao, J., Ji, X., Ye, Q., 2019. Danet: Divergent activation for weakly supervised object localization. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6589–6598.
- Zhang, J., Chen, L., Ouyang, B., Liu, B., Zhu, J., Chen, Y., Meng, Y., Wu, D., 2021a. Pointcutmix: Regularization strategy for point cloud classification. arXiv preprint arXiv:2101.01461.
- Zhang, J., Meng, Y., Wu, J., Qin, J., Yao, T., Yu, S., et al., 2020. Monitoring sugar crystallization with deep neural networks. *J. Food Eng.* 280, 109965.
- Zhang, K., Liu, M., Zhang, J., Dong, Z., 2021b. Pa-mvsnet: Sparse-to-dense multi-view stereo with pyramid attention. *IEEE Access* 9, 27908–27915.
- Zhang, Z., Hua, B.S., Yeung, S.K., 2019. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1607–1616.
- Zou, X., Cheng, M., Wang, C., Xia, Y., Li, J., 2017. Tree classification in complex forest point clouds based on deep learning. *IEEE Geosci. Remote Sens. Lett.* 14, 2360–2364.