MLNet: Multi-level Classification Network

Doudou Xue 1,a*, Yinglei Cheng 1,b, Pei Wen1,c, Liyin Zhang2,d
1 Information and Navigation College, Air Force Engineering University, Xi’an, Shaanxi
2 Materials Science and Engineering School, Nanyang Technological University, Singapore
a 1447551957@qq.com, b ylcheng718@163.com, c 2412288250@qq.com, d zzzly0116@126.com

Abstract. The point cloud classification network only uses point cloud structure features or edge features to construct the feature vector, so that the classification accuracy is low. To solve this problem, this paper proposes a multi-level classification network. First, the original point cloud is divided into small samples using the preprocessing algorithm to obtain batch input and improve training efficiency; then construct structural features and edge features for point cloud feature extraction, and obtain local fine-grained descriptions through multi-level expressions within and between points; then design a convolutional neural network for feature learning. With the deepening of the network level, the degree of feature abstraction level is higher and higher, and the degree of distinction increases, so as to effectively improve the accuracy. Using the Vaihingen dataset to test the MLMS-Net network, the accuracy rate reached 85.4%.

1. Introduction

With the development of science and technology, point clouds are gradually being used in fields such as autonomous driving, remote sensing mapping and robotics. In the field of remote sensing mapping, point clouds are widely used in 3D building reconstruction[1], road model extraction[2], and vegetation coverage estimation[3]. The key step is to effectively classify and segment the point cloud model. However, due to the unordered and irregular distribution, it is still a huge challenge to achieve high-precision classification and segmentation of point clouds.

The traditional point cloud classification algorithm needs to express the feature vector of each point, but the classification accuracy is low when the point cloud is classified by manually constructing the feature vector[4-6]. In recent years, convolutional neural networks have achieved great success in computer vision, and their algorithm accuracy far exceeds machine learning and traditional algorithms. The convolutional neural network extracts shallow neighborhood features by sharing the convolution kernel, and learns abstract deep features in the training network, so that the classification accuracy is higher. Therefore, literature [7] proposed that the use of convolutional neural networks to extract deep features of point clouds can improve the classification accuracy of point clouds. However, the main object of convolutional neural networks is regular rasterized images, while point clouds are irregular and unordered data. Therefore, it is necessary to consider how to effectively apply convolutional neural networks to point cloud processing.

The first proposed method is mainly to regularize the point cloud, which mainly includes: a method based on multi-view images[8-10] and a method based on voxels[11-13]. Su proposed MVCNN (Multi-View Convolution Neural Network)[8] in 2015. The main idea is to project the point cloud from multiple
perspectives, pool the multi-angle projection features into global features, and finally complete the
classification through the convolutional neural network. However, due to the two-dimensional image
ignores the three-dimensional structural characteristics of the point cloud, too much information is lost,
and the occluded part cannot be effectively processed, resulting in low classification accuracy. Zhou
proposed VoxelNet[11] in 2018. Its main idea is to divide the point cloud into three-dimensional grids
uniformly, extract the features in each grid, and use the three-dimensional convolutional neural network
to learn structural features to complete the points Cloud classification. This method preprocesses the
point cloud into a voxel structure. If the resolution is too low, the accuracy will be lost, and the memory
usage and time overhead will increase exponentially with the increase of the resolution.

Since the regularized preprocessing of the point cloud will affect the classification accuracy, Qi
proposed PointNet[14], which combines the original point cloud with the neural network for the first time,
and successfully extracts the global features of the 3D model to complete the classification task. However,
the network only considers global information, and does not consider the local information of the point
cloud. The key of the convolutional neural network is to capture the features of the neighborhood through
convolution to achieve accurate classification. In 2018, Qi proposed PointNet2[15]. The local
neighborhood was obtained by sampling the farthest point, and the hierarchical neural network was used
to extract local features. Experiments proved that adding local neighborhoods can effectively improve
the accuracy of network classification. The above model only considers the structural features of the point
cloud. If regarded the point cloud as a graph model, the point cloud also has another feature: edge features.
Simonovsky et al. regarded points as the vertices of the graph, obtained neighborhood information
through edge conditional convolutional networks[17], and used the VoxelGrid algorithm to achieve graph
optimization. Wang proposed a dynamic graph network, using edge feature space to build a probability
graph model[18], dynamically updating the feature map at each layer of the network, and completing the
point cloud classification task. However, this method only uses the edge features of the point cloud, and
the classification effect is poor.

At present, most point cloud classification networks only consider structural features or edge features
unilaterally, and do not fully extract point cloud features. In response to the above problems, this paper
proposes a point cloud-oriented multi-level classification network MLNet. The network takes irregular
point clouds as input, extracts features from two levels of structural features and edge features, and uses
neural networks to obtain point cloud deep structure features and edge features, finally through the post-
processing module to complete the point cloud classification tasks.

2. Multi-level Classification Network

This section introduces the multi-level network MLNet. The network includes two parts: preprocessing
algorithm, feature extraction algorithm. point cloud preprocessing algorithm is mainly to segment large-
scale point clouds, structure feature extraction is to extract point structure features, edge feature extraction
is to learn point cloud edge features, and post-processing is to fusion structure features and edge features.

2.1. Preprocessing Algorithm

Since the experimental data in this paper are large-scale laser LiDAR point cloud data, the point cloud
needs to be segmented. The specific steps are shown in Table I.

| Table I: Preprocessing Algorithm |
|----------------------------------|
| **Input:** Original Point Cloud P, Segmentation threshold \( l_{\text{min}} \) |
| **output:** Point Cloud Patchs \( P_{\text{patch}} \) |

1. Create an initial segmented cube based on the point cloud boundary \( B \), side length is \( \text{Side}_{B} \);
2. Compare \( l_{\text{min}} \) and \( \text{Side}_{B} \). If \( l_{\text{min}} \) is more bigger, Break; If \( \text{Side}_{B} \) is more bigger, run next step;
3. Divide \( P \) into eight pieces \( P_{\text{temp}} = \{ P_{p1}, P_{p2}, \ldots, P_{pn} \} \);
4. For \( P_i \) in \( P_{\text{temp}} \), Perform the first three steps.
2.2. Feature Extraction Algorithm

Get the local neighborhood: First, make a ball with point \( p_i \) as the center and \( r \) as the radius. If the number of points in the ball is greater than \( k \), randomly select \( k \) points. If it is less than \( k \), randomly copy the points until the number of neighbor points reaches \( k \).

Construct Edge Feature: Combine \( L_1 \) norm, \( L_2 \) norm, \( L_\infty \) norm to construct the edge feature \( e_{ij} = (L_1, L_2, L_\infty) \). \( x, y, z \) is the Coordinate information.

\[
L_1 = |x_i - x_j| + |y_i - y_j| + |z_i - z_j| \tag{1}
\]

\[
L_2 = \left( |x_i - x_j|^2 + |y_i - y_j|^2 + |z_i - z_j|^2 \right)^{1/2} \tag{2}
\]

\[
L_\infty = \max \left( |x_i - x_j|, |y_i - y_j|, |z_i - z_j| \right) \tag{3}
\]

Construct structure features: Combine the coordinates of point \( i \) and \( j \) to get the structure features \( s_{ij} = (P_i, P_j) \);

Construct Multi-level Feature: After obtain the Edge Feature and structure features, Cascading those features to obtain multi-level features \( f_{ij} = (e_{ij}, s_{ij}) \);

Feature learning: First, Pass the multi-level features through the fully connected layer to get the Low-level features. When the neural network is trained, the features will pass through the convolutional layer, the batchnorm layer, the activation layer and the pooling layer in turn; the convolutional layer uses the feature update function to obtain a deep representation of the structural features, and the activation layer nonlinearizes the output and The pooling layer uses a symmetric function to ensure that the disorder of the point cloud will not affect the classification result. Common activation functions are Sigmod and ReLU. The Sigmod will cause the gradient become smaller during the training process, so the activation function used in this article is the ReLU function, which unidirectionally suppresses the negative output and can solve the problem of gradient disappearance. The specific network structure is shown in the Fig 1 below.

In this paper, MLP represents the feature update function of the convolutional layer, the activation function isReLU. After obtaining the low-level local features by fully connected layer, then send it to the Conv(1,3,1), Conv(1,1,64) layer and the batchnorm layer to obtain a shallow feature map; then send the feature map to the Conv(1,1,128), Conv(1,1,1024) layer to obtain 1024 dimension global feature. With the deepening of the network, the obtained features become more and more abstract, and the degree of differentiation of features gradually increases.

3. Experiment and Analysis

This article uses the Vaihingen dataset provided by the ISPRS competition. The data set has 753,876 training points and 411,722 test points, with an average density of 4/m². the data is adaptively adjusted in advance for training, the roof and the facade are merged into buildings, and the power line points are removed. Therefore, the total number of categories is seven, which are: buildings, low vegetation, trees, ground, car, fence, and shrub.

When performing point cloud segmentation, \( l_{\text{min}} \) is 5m. The initial learning rate of the MLNet network is 0.01, and the learning rate is half after 2000 training; the momentum stochastic gradient descent method is used for training, where the momentum value is 0.9; the batch block size is 32, and the dropping rate is 0.1; use six in the double cross-validation scheme, the network input is the coordinate information of the point cloud. The machine configuration is as follows: GPU is GTX 2060, CPU is i7-8750H, and the experimental environment is python 3.6, tensorflow.
3.1. Experimental Results

Fig 2 is the classification result diagram of the test data, where I represents the original point cloud diagram, II represents the accuracy diagram, the green is the correct point, and the red is the wrong point. 1-7 represent building, fences, low vegetation, ground, shrub, trees, and car, respectively. The number of car points is too few, so they are not displayed here.

It can be seen from the figure that the network performs well in extracting the details of the edge of the building. The accuracy of the ground, low vegetation and trees is high, while the results of fences and bushes are poor.

Table II shows the user classification accuracy of each category when the Vaihingen dataset is used for testing. 1-7 respectively represent the same meaning as above. The overall classification accuracy of
the network can reach 85.8%, and the Kappa coefficient is 0.92. It can be seen from the table that the classification accuracy of the ground can reach 93%, the roof’s classification accuracy is 94%, the classification accuracy of low vegetation can reach 80%, the classification accuracy of trees is 82%, and the classification result is good; The main reasons are the following three aspects: (1) These four types of training data are bigger than other data, so there are more neurons and training parameters in the network to learn the corresponding features; (2) During training, these four types of ground Objects can be better classified by combining structural features and relational features; (3) Because there are more data points than other points, the impact of incorrect points on the classification accuracy is much smaller than other categories. In contrast, the classification accuracy of shrubs, fences and cars is low. 11% of shrubs are classified as fences, and 13% of shrubs are classified as trees.

Table 2

|   | 1   | 2    | 3    | 4    | 5    | 6    | 7    |
|---|-----|------|------|------|------|------|------|
| 1 | 93.2| 1.61 | 2.22 | 0.24 | 5.92 | 6.67 | 3.25 |
| 2 | 0.5 | 72.01| 1.76 | 0.15 | 5.82 | 1.07 | 4.08 |
| 3 | 1.44| 8.94 | 80.05| 9.54 | 13.01| 1.94 | 5.47 |
| 4 | 0.03| 0.37 | 7.48 | 91.66| 0.24 | 0.05 | 2.04 |
| 5 | 1.9 | 12.78| 6.08 | 0.19 | 55.17| 8.7  | 11.15|
| 6 | 2.08| 3.92 | 2.17 | 0.03 | 20.99| 81.82| 4.65 |
| 7 | 0.12| 1.14 | 0.26 | 0.13 | 1.07 | 0.05 | 68.68|

Fig 3 shows the classification results of fences and shrubs. Among them, I represents the correct classification result, and II represents the network classification result of this article. It can be seen from the figure that the fence point and the bush are relatively close in position. When using lidar scanning, the generated point cloud is easy to mix, and the elevation difference between the two is small, so the classification accuracy is low. In summary, the multi-scale and multi-level point cloud classification network proposed in this paper can achieve accurate classification of LiDAR point cloud data.

3.2. Experimental Comparison
Compared with other point cloud classification networks provided by ISPRS, the classification results are shown in Table III. Among them, the IIS_1 network uses machine learning algorithms to segment point clouds through superpixels\cite{12}, and uses spectral and geometric features to divide the point clouds into 6 categories with a classification accuracy of 69.9%; HM_1 uses artificially designed geometric features and random forests to perform point clouds. Classification\cite{4}; UM uses genetic algorithms to complete point cloud classification through point cloud texture features and geometric features, with a classification accuracy of 80.86%; LUH network uses point geometric features and context classification to reduce point cloud classification errors\cite{1}; BIJ_W uses minimum spanning tree Describe the spatial characteristics of the point cloud, and the classification accuracy has reached 81.6%\cite{16}; WhuY3
introduces deep learning, which converts the point cloud into a two-dimensional image by calculating the geometric features and full waveform features of the neighborhood\cite{7}; NANJ2 uses a multi-scale neural network to extract three scales Geometric features, the network structure is relatively simple, the training accuracy is 85.2%, and the calculation time is 54.2s\cite{21}. It can be obtained from Table 3 that compared with HM\_1, UM (only using point cloud structure features) and BIJ\_W (only using point cloud relationship features), the overall classification accuracy of the network in this paper is greatly improved. The classification accuracy on the ground reaches 93.6%, which is better than other networks; the classification accuracy of the roof is 94%, and the average classification accuracy of low vegetation and trees is 82.4%. The classification accuracy of vehicles is lower than other networks, because the number of vehicle points is too few and the feature discrimination is low. Even though the classification accuracy of bushes and fences is lower than that of roofs and grounds, the classification accuracy is still higher than that of other networks. The main reason is that the network in this paper uses edge features and structural features for classification, which has a high degree of feature discrimination. In terms of time cost, the core of HM\_1 and UM is traditional machine learning, with small time cost and an average calculation time of 3.5s. BIJ\_W, WhuY3, NANJ2 and the algorithm in this paper all introduce neural networks to extract deep features. The network calculation time in this paper is 68.6s. Compared with the BIJ\_W network, it is reduced by 16.7s, mainly because the original point cloud data is used as input in this paper, which is not converted into grids and images. The MLNet network can effectively improve the classification accuracy of point clouds while ensuring the time cost.

| Algorithm | ground | roof | low_veg | tree | car | overall |
|-----------|--------|------|---------|------|-----|---------|
| IIIS\_1   | 85     | 82.8 | 65.8    | 48.2 | 64  | 69.9    |
| HM\_1     | 91.5   | 91.6 | 73.8    | 80.2 | 58.4| 80.5    |
| UM        | 89.1   | 92   | 79      | 77.9 | 47.7| 80.8    |
| LUH       | 91.1   | 94.2 | 77.5    | 83.1 | 73.1| 81.6    |
| BIJ\_W    | 90.5   | 92.2 | 78.5    | 78.4 | 56.4| 81.5    |
| WhuY3     | 90.1   | 93.4 | 81.4    | 78   | 63.4| 82.3    |
| NANJ2     | 91.2   | 93.6 | 88.8    | 82.6 | 66.7| 85.2    |
| MLNet     | 91.6   | 93.2 | 80.5    | 81.8 | 68.6| 85.4    |

4. Conclusion

At present, the network cannot make full use of the point cloud features. This paper proposes a multi-level network MLNet, designs a feature extraction module to extract point cloud structure features and edge features. Experiments show that the classification accuracy reaches 85.4%. However, this article does not consider the relationship between the local area of the point cloud. Future experiments can start from the perspective of extracting the edge relationship feature of the local point cloud to improve the classification accuracy of the point cloud.

Acknowledgment

This work was supported by Youth Program of National Natural Science Foundation of China 41601436. Thanks Cuizhu Huo for helping with baseline experiments setup.

References

[1] NIEMEYER J, ROTTENSTEINER F, SOERGEL U, et al. Hierarchical higher order crf for the classification of airborne LiDAR point clouds in urban areas[J]. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016, XL1-B(3):655-662.

[2] YANG Bisheng, FANG Lina, LI Qingquan, et al. Automated extraction of road markings from mobile lidar point clouds[J]. Photogrammetric Engineering & Remote Sensing, 2012, 78(4): 331-338.
[3] DALPONTE M, BRUZZONE L, GIANELLE D. A system for the estimation of single-Tree stem diameter and volume using multireturn lidar data[J]. IEEE Transactions on Geoscience & Remote Sensing, 2011, 49(7): 2479-2490.

[4] STEINSIEK M, POLEWSKI P, YAO Wei, et al. Semantische analyse von als- und mls-daten in urbanen gebieten mittels conditional random fields[C]// 37. Wissens chaft lich-Technische Jahrestagung der DGPF. 2017: 65-87.

[5] LI Xiaotian, JIANG Gang, ZHANG Yu. Classification of airborne LiDAR point cloud based on improved analytic hierarchy process[J]. Journal of Gansu Sciences, 2019, 31(1): 86-91.

[6] Horvat D, Zalik B, Mongus D. Context-dependent detection of non-linearly distributed points for vegetation classification in airborne LiDAR[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 116(6), 1–14.

[7] YANG Zhishuang, JIANG Wanshou, XU Bo, et al. A convolutional neural network-based 3d semantic labeling method for als point clouds[J]. Remote Sensing, 2017, 9(9):936-948.

[8] SU hang, MAJI S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3d shape recognition[C]// 2015 IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2015: 945-953.

[9] Riegler G, Ulusoys AO, Bischof H, et al. Octnetf-usion: learning depth fusion from data[C]// International Conference on 3D Vision, Qingdao, China: 3DV, 2017: 57-66.

[10] HE Elong, WANG Hongping, CHEN Qi, et al. An improved contextual classification method of point cloud[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(3): 362-370.

[11] ZHOU YIN, TUEL O, Voxelnet: end-to-end learning for point cloud based 3d object detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 4490-4499.

[12] ANANDAKUMAR M, RAMIYA, Rama R, et al. A super voxel-based spectro-spatial approach for 3d urban point cloud labelling[J]. International Journal of Remote Sensing, 2016, 37(17), 4172-4200.

[13] SHAO Lei, DONG Guangjun, YU Ying, et al. A point cloud classification method based on multi-scale voxel and higher order random fields. Journal of Computer-Aided Design & Computer Graphics[J], 2019, 31(3): 385-392.

[14] Charles RQ, SU HAO, Kaichun M, et al. Pointnet: deep learning on point sets for 3d classification and segmentation[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 77-85.

[15] Charles RQ, YI LI, SU HAO, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space[C]// Advances in Neural Information Processing Systems 30, NY: Curran Associates, Inc. 2017: 5099-5108.

[16] WANG Zhen, ZHANG Lliqiang, ZHANG Liang, et al. A deep neural network with spatial pooling (DNNSP) for 3-d point cloud classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8):4594-4604.

[17] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2017: 29-38.

[18] WANG Yue, Sun Yongbin, Liu Ziwei, et al. Dynamic graph cnn for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 112-124.

[19] Fauvel M, Chanussot J, Benediktsson J A, et al. A spatial-spectral kernel-based approach for the classification of remote-sensing images[J]. Pattern Recognition, 2012, 45(1): 381-392.

[20] Yousefhussien M, Kelbe D, Lentilucci EJ, et al. A multi-scale fully convolutional network for semantic labeling of 3d point clouds[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 143(5): 191-204.

[21] ZHAO Ruibing, PANG Mingyong, Wang Jidong. Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network[J]. International Journal of Geographical Information Science, 2018, 32(5):960-979.