

RESEARCH

Open Access



MP-DGCNN for the semantic segmentation of Chinese ancient building point clouds

Caochenyu Zhou^{1,2}, Youqiang Dong^{1,2*}, Miaole Hou^{1,2}, Yuhang Ji^{1,2} and Caihuan Wen³

Abstract

Point cloud semantic segmentation is a key step in the scan-to-HBIM process. In order to reduce the information in the process of DGCNN, this paper proposes a Mix Pooling Dynamic Graph Convolutional Neural Network (MP-DGCNN) for the segmentation of ancient architecture point clouds. The proposed MP-DGCNN differs from DGCNN mainly in two aspects: (1) to more comprehensively characterize the local topological structure of points, the edge features are redefined, and distance and neighboring points are added to the original edge features; (2) based on a Multilayer Perceptron (MLP), an internal feature adjustment mechanism is established, and a learnable mix pooling operator is designed by fusing adaptive pooling, max pooling, average pooling, and aggregation pooling, to learn local graph features from the point cloud topology. To verify the proposed algorithm, experiments are conducted on the Qutan Temple point cloud dataset, and the results show that compared with PointNet, PointNet++, DGCNN, GACNet and LDGCNN, the MP-DGCNN segmentation network achieves the highest OA, mIOU and mAcc, reaching 90.19%, 65.34% and 79.41%, respectively.

Keywords Ancient architectural buildings, Deep learning, Semantic segmentation, DGCNN, Mixed pooling

Introduction

The ancient architectures were the important components of the Chinese cultural heritage [1, 2]. Suffering from weathering, fires and rotting, a mass of Chinese historical buildings with wooden structural frame has disappeared. The digital archiving was the important measurement for the protection of historical buildings [3, 4] and 3D point cloud which can provide the extract spatial geometry of built heritage with a complex shape has been widely used in the documentation of Chinese ancient architectures [5]. However, the captured original

point cloud lacked the structured information, such as semantics and hierarchy between parts, which disturbed the usage of point cloud in other application fields [6–8].

Point cloud semantic segmentation divides the original point cloud into subdatasets with semantic meaning. Based on the results of semantic segmentation, the annotated point cloud can be utilized to reconstruct parametric geometries manageable in H-BIM platforms [9–12]. In practical projects, manual methods involving segmentation through visual recognition and manual labeling of semantic information on the point cloud have been widely adopted by operators. Recently, numerous studies on deep learning (DL) techniques, offering novel and more effective solutions for point cloud semantic segmentation, have been conducted with the aim of replacing manual operations [13–16].

However, existing methods still face challenges in segmenting complex structures, such as ancient Chinese architecture. 3D modeling of such intricate models is also a prominent topic in the fields of graphics and digital modeling. This article introduces a Mix Pooling Dynamic

*Correspondence:

Youqiang Dong
dongyouqiang@bucea.edu.cn

¹ School of Geomatics and Urban Spatial Informatics, Engineering and Architecture, Beijing University of Civil, Beijing 102616, China

² Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring, Beijing 100044, China

³ Space Information Technology Application Research Center of Hebei Geological and Mineral Exploration and Development Bureau, Geological Geomatics Institute of Hebei, Shijiazhuang 050031, Hebei, China

Graph Convolutional Neural Network (MP-DGCNN) designed for the segmentation of point clouds representing ancient architecture.

Related works

Nowadays, various of point cloud semantic segmentation methods including machine learning (ML) techniques, deep learning (DL) techniques and hybrid methods have been proposed and achieved good performance [17].

The ML techniques involved multiple stages including neighborhood selection, feature extraction, feature selection and semantic segmentation [18]. Aiming at each stage, the researchers proposed several of strategies so that the ML can obtain better performance on the point cloud semantic segmentation of historical buildings. Grilli et al. [19] analyzed the efficacy of the geometric covariance features and the impact of the different features calculated on spherical neighborhoods at various radius sizes. Simone Teruggi et al. [20] extended a machine learning (ML) classification method with a multi-level and multi-resolution (MLMR) approach to segment the Pomposa Abbey (Italy) and the Milan Cathedral (Italy). Dong et al. [21] fused the geometric covariance features and the features from construction regulation to classify the roof point cloud of ZHONG HE Temple into 9 categories. The proper features promote the higher accuracy of point cloud segmentation [22]. The different types of historical architectures have different appearances, it is necessary to design proper segmentation features. This limits the application of machine learning.

The hybrid approach first utilizes an over-segmentation or point cloud segmentation algorithm as the initial pre-segmentation. Subsequently, prior knowledge or supervised methods are employed to label segments rather than individual points. In [23], the point clouds were initially clustered into segments with the assistance of a multi-resolution super voxel algorithm. Following this, Vosselman et al. [24] employed the Hough Transform (HT) to generate planar patches within their PCSS algorithm framework as the preliminary segmentation step. Similarly, Landrieu and Simonovsky [25] employed a super point structure in the pre-segmentation phase and introduced a contextual PCSS network that combines super point graphs with Point-Net and contextual segmentation.

In the past decades, the successful application of deep learning based on the image has promoted the scholars to extend research into 3D point cloud data, and some representative methods and theories have been developed [26–29]. Deep neural networks have made significant progress in research on indoor scenes, urban streets,

and remote sensing, and have also provided new ideas for component recognition in ancient architectural scenes.

DL approaches on the semantic segmentation of point cloud can be classified into 3 categories: projection-based [30, 31], voxel-based [32–34], and point-based. Now, the point-based networks have established as mainstream method for point cloud semantic segmentation. Point-Net and its later improvement PointNet++ [35, 36] were considered as pioneer works. Compared with the regular supervise machine learning, the DL methods do not need to design the features.

Inspired by PointNet, Wang et al. [37] proposed DGCNN. DGCNN make use of the KNN to establish a local graph structure and designs the EdgeConv operator to process the edge features of the graph structure for learning the topological relationships between points. Due that the local graph constructed by DGCNN is dynamically updated, the convolutional receptive field can cover the entire diameter of the point cloud and allowed the information to diffuse within the diameter of the point cloud. Zhang et al. [38] proposed LDGCNN (Linked Dynamic Graph CNN) based on DenseNet [39]. This method can mitigate the problem of gradient vanishing by densely connecting the features at different levels and extracting the richer semantic information from multi-level local features. To learn the important features, Sun et al. [40] designed an edge feature weighting function for DGCNN to weaken the interference of distant points and strengthen the features of nearby points considering the feature distribution of local points. Wang et al. [41] introduced the residual network idea [42] to increase the depth of the network, making the network training more stable and improving the feature extraction capability.

DGCNN achieved excellent performance on the public datasets such as ModelNet40 [43], ShapeNetPart [44], and S3DIS [45]. The scholars in the field of architectural heritage introduced this method into the point cloud semantic segmentation process of the historical building. Pierdicca et al. [46] make use of DGCNN to label the elements of churches, porches, and monasteries in the ArCH dataset. Given the resemblance in color among components sharing the same structural characteristics in historical buildings, this method added color features in HSV space and normal vectors to the input layer of DGCNN and extended the edge convolution operation layer by constructing the local graph in the feature space. To improve the accuracy of segmentation results, Matrone et al. [47] added the geometric features such as verticality and planarity to the input layer of DGCNN on the basis of the performance evaluation of the various machine models and deep learning models on the ArCH dataset. Massimo et al. [47]

described a strategy using DGCNN for point cloud semantic segmentation of the historical building. This method firstly expanded the ArCH data set by rotating, cropping, translating, coordinate perturbation, and scaling; subsequently, the transfer learning technology was applied to pre-train parameters of the DGCNN model on the extended ArCH dataset; last, a small portion of the point clouds from a new scene was selected to adjust parameters for semantic segmentation of the remaining point clouds.

DGCNN make use of the local graph structures to describe spatial points and learns the topological relationship between points through edge convolution. Although DGCNN can achieve superior performance on some semantic segmentation tasks, it was still a challenge for the semantic segmentation of Chinese historical buildings. One hand, the components of Chinese historical buildings are overlapped and the scales of the components varied greatly. On the other hand, the randomness of input points exacerbates the complexity of point clouds. The deep neural networks should capture more discriminative geometric information from local structures of point clouds so that its generalization ability and stability need to be further improved when facing ancient architectural scenes. To overcome this limitation, this article focuses on the edge features and pooling functions of DGCNN, proposes an improved strategy for autonomously learning

pooling rules, and puts forward a more robust mixed-pooling DGCNN point cloud semantic segmentation network.

The main contributions of the presented MP-DGCNN are three folds:

- 1) The modified edge features of local graph are proposed by adding the position, direction, and distance information. The modified edge features enable MP-DGCNN to learn richer structural information and indirectly enhances the ability to extract local features.
- 2) A hybrid pooling operator integrating the maximum pooling, average pooling, aggregation pooling, and adaptive pooling that can learn pooling rules independently from point clouds is designed, which can efficiently autonomously learn pooling rules to extract local feature vectors of points.

The structure of this paper is organized as follows: the second chapter introduces the proposed MP-DGCNN which was the applied algorithm and net in this paper; the third chapter evaluated the performance on the test data; in the last chapter, the author summarized and forecast the methodology.

MP-DGCNN

The MP-DGCNN network proposed in this paper for semantic segmentation of ancient buildings is shown in Fig. 1. The core component of MP-DGCNN is the

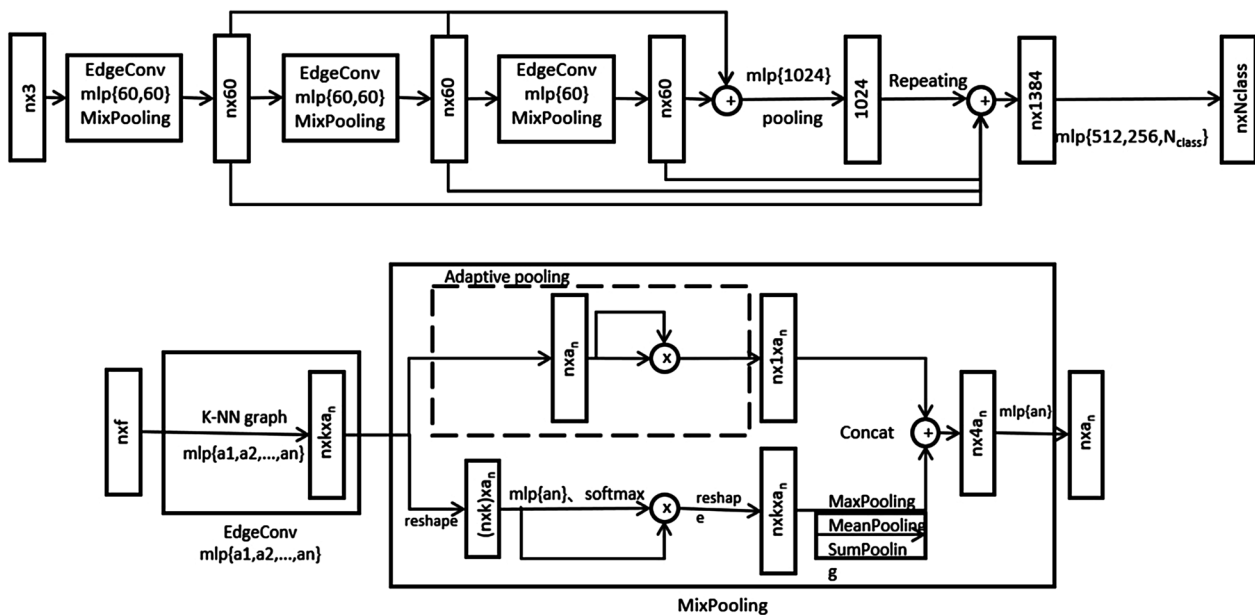


Fig. 1 MP-DGCNN

improved EdgeConv layer, which captures fine-grained local geometric structures by stacking three EdgeConv layers. The output features of the three EdgeConv layers are concatenated and pooled to form a one-dimensional global feature descriptor. Then, the descriptor is concatenated with the local features of each point in the output features of the three EdgeConv layers, so that each point's features include both local detail information and global feature descriptors. Finally, MLP is used to fuse depth semantic information, and dropout technology is used to alleviate overfitting. The segmentation score of each point was calculated to complete the segmentation task.

The improved edge convolution operator is the core concept of MP-DGCNN, which mainly includes defining edge features, extracting graph features with MLP, and merging mixed pooling.

Edge feature definition

The MP-DGCNN network redefines edge features by introducing squared distance and position coordinates of neighbor node, enhancing the characterization of local graph structures and the richness of neighborhood information, thereby achieving a more comprehensive representation of the topology of points. The defined edge features are:

$$L_i = \{l_{i_0}, l_{i_1}, \dots, l_{i_j}\} \tag{1}$$

In Eq. (1), L_i represents the edge feature vector set of the local graph and l_{i_j} represents the edge features between the center point p_{i_0} and one of its neighboring points $p_{i_j}, j = \{0, 1, 2, \dots, k - 1\}$.

$$l_{i_j} = (p_{i_0}, p_{i_j}, e_{i_j}, e_{i_j}^T e_{i_j}) \tag{2}$$

l_{i_j} is a feature vector composed of four parts, p_{i_0} expresses the coordinate information of the center point, p_{i_j} describes the position information of neighboring nodes, e_{i_j} describes the direction of the graph structure, and $e_{i_j}^T e_{i_j}$ explicitly describes the distance. In particular, In particular, when $j = 0$, e_{i_j} was zero vector. If the point p_i is a row vector with n_{p_i} elements, the vector l_{i_j} has $3n_{p_i} + 1$ elements.

Graph feature extraction based on MLP

The neural network extracted graph features from the local topological structure of points relying on the feature extraction function f_e . The input is edge feature L_i . Output local graph feature vector f_i .

$$f_i = f_e(G(V_i, E_i)) = MixPooling\{h(l_{i_0}), h(l_{i_1}), \dots, h(l_{i_{k-1}})\} \tag{3}$$

$$l_{i_j} = h(p_{i_0}, p_{i_j}, e_{i_j}, e_{i_j}^T e_{i_j}) \tag{4}$$

In Eq. (3), $MixPooling\{\}$ is the mixed pooling function and $h()$ is the function which extracted the feature vector $h(l_{i_j})$ of the hidden layer from edge features using parameter shared MLP.

$$h_{c'}(p_{i_0}, p_{i_j}, e_{i_j}, e_{i_j}^T e_{i_j}) = \sum_{c=1}^C (w_{c'c} p_{i_0c} + w_{c'(c+C)} p_{i_jc} + w_{c'(c+2C)} e_{i_jc}) + w_{c'} e_{i_j}^T e_{i_j} + b_{c'} \tag{5}$$

$$h(p_{i_0}, p_{i_j}, e_{i_j}, e_{i_j}^T e_{i_j}) = (h_1(p_{i_0}, p_{i_j}, e_{i_j}, e_{i_j}^T e_{i_j}), \dots, h_C(p_{i_0}, p_{i_j}, e_{i_j}, e_{i_j}^T e_{i_j})) \tag{6}$$

In Eq. (6), p_{i_0c} , p_{i_jc} and e_{ij} presented the element values of the feature vectors of the i_{th} center point, j_{th} neighboring point, and j_{th} edge on the c_{th} channel, respectively; C is the number of channels of the center point p_{i_0} , that is $C = n_{p_i}$; C' is the number of neurons in the MLP layer, and c' identifies the neuron ordinal; $w_{c'c}$, $w_{c'(c+C)}$, $w_{c'(c+2C)}$, $w_{c'}$ and $b_{c'}$ are trainable parameters for MLP.

Mix pooling

For a local graph structure G_i , each vector in the graph features F_i of G_i is located in the feature space with the a_n dimension as is seen in Eq. (7).

$$F_i = (f_0, f_1, \dots, f_{k-1}) \subseteq R^{a_n} \tag{7}$$

The MLP with a_n neurons in one layer (without bias constant and activation function) is used to adjust the graph features.

$$h(F_i) = F_i W_{adj} = \begin{bmatrix} f_{01} & \dots & f_{0a_n} \\ \vdots & \ddots & \vdots \\ f_{k-11} & \dots & f_{k-1a_n} \end{bmatrix} \begin{bmatrix} w_{11} & \dots & w_{1a_n} \\ \vdots & \ddots & \vdots \\ w_{a_n1} & \dots & w_{a_na_n} \end{bmatrix} \tag{8}$$

In Eq. (8), F_i represents the local graph feature, W_{adj} represents the weight matrix of MLP, f_j represents the j_{th} element of the i_{th} feature vector of the local graph feature, and w_{ij} represents the learnable weight parameter. If the number of neurons in a single-layer MLP is a_n . W_{adj} is an $a_n \times a_n$ matrix which is shared among different local graph G_i . According to the contribution degree of each element of each feature vector, make use of the activation function $Softmax()$ to stimulate the adjusted graph features into the attention weight coefficient W_{att} . Finally, multiply the adjusted graph feature F_i and the attention coefficient W_{att} to generate Graph Features F'_i . The optimized Graph Features was described as F'_i .

$$W_{att} = Softmax(h(F_i)) = Softmax(F_i W_{adj}) \tag{9}$$

$$F'_i = F_i * W_{att} = F_i * Softmax(F_i W_{adj}) \tag{10}$$

In Eq. (9), F_i is the graph feature after optimization, $*$ is Hadamard product, and $Softmax()$ is the activation function. Neural networks can dynamically generate attention coefficients W_{att} based on different F_i , and dynamically optimize local graph features to obtain F'_i .

Mix-pooling is composed of four pooling methods: max pooling, mean pooling, sum pooling, and adaptive pooling. It is a local structure information fusion strategy based on a dynamic feature adjustment mechanism. The strategy mainly consists of three parts:

1. The construction of max pooling, mean pooling and sum pooling

The Max pooling vector, mean pooling vector, and Sum pooling vector are further extracted along each axis direction of all feature vectors.

$$\begin{cases} f'_{max} = max(F'_i) = max\{f'_0, f'_1, \dots, f'_{k-1}\} \subseteq R^{a_n} \\ f'_{mean} = mean(F'_i) = mean\{f'_0, f'_1, \dots, f'_{k-1}\} \subseteq R^{a_n} \\ f'_{sum} = sum(F'_i) = sum\{f'_0, f'_1, \dots, f'_{k-1}\} \subseteq R^{a_n} \end{cases} \tag{11}$$

In Eq. (11), f'_{max} , f'_{mean} and f'_{sum} represented Max Pooling vector, Mean Pooling vector and Sum Pooling vector, respectively. f'_{max} describes the main features of the graph structure, f'_{mean} reflects the internal properties of the graph structure, and f'_{sum} aggregates pooling vectors to solve the problem of important feature loss caused by random sampling. It can effectively handle several topological structures of the same semantic point.

2. The construction of adaptive pooling vector

Although the above three different pooling methods can extract graph features with different characteristics, the Max Pooling vector, Mean Pooling vector and Sum Pooling vector are designed based on the fixed rules. In order to abstract feature vectors representing local features of point, adaptive pooling vector is proposed. Make use of the 2D convolution kernel with size $1 \times k$ to perform convolution operator on the local graph to reconstruct the feature vectors of points and adaptively extracted pooling vectors directly from the graph features. Finally, applied the feature adjustment mechanisms to this vector to obtain the adaptive pooling vectors $f'_{adapt} \subseteq R^{a_n}$ of points.

3. The construction of adaptive pooling vector

Concatenated the pooling vectors of f'_{max} , f'_{mean} , f'_{sum} and f'_{adapt} into f'_{concat} along the feature dimension as is shown in Eq. (12). Then, make use of a single-layer MLP with a_n neurons to map f'_{concat} back to the R^{a_n} space, and calculate the local feature vector $f'_{Edgeconv}$ of the points extracted by the edge convolution operator as is shown in Eq. (13).

$$f'_{concat} = (f'_{max}, f'_{mean}, f'_{sum}, f'_{adapt}) \subseteq R^{4a_n} \tag{12}$$

$$f'_{Edgeconv} = h(f'_{concat}) \subseteq R^{a_n} \tag{13}$$

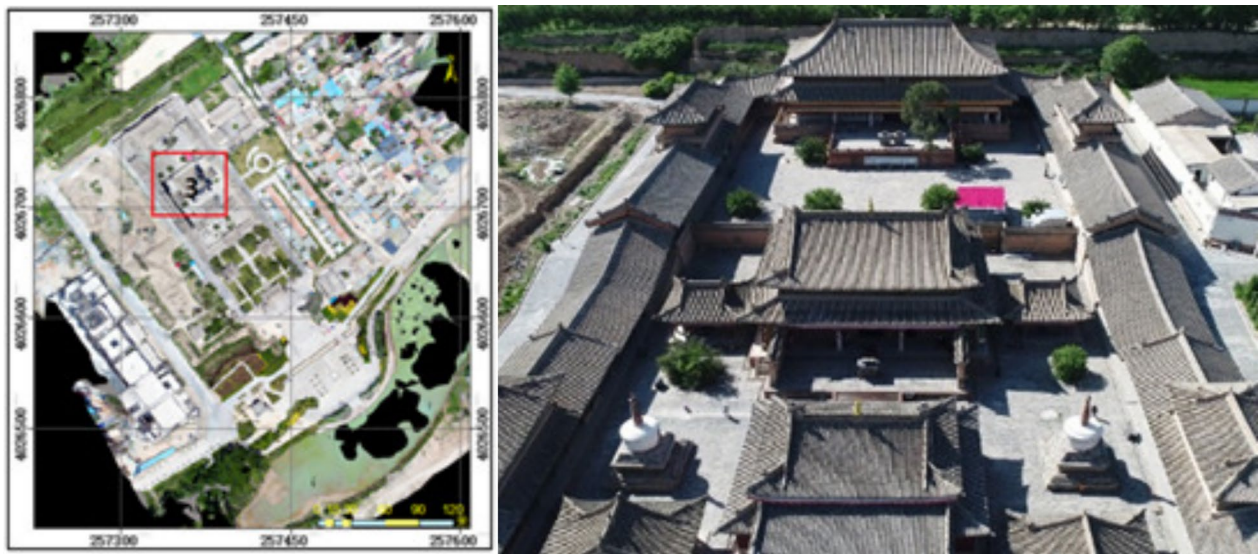


Fig. 2 The location of Qutan Monastery and the overview of Qutan Monastery

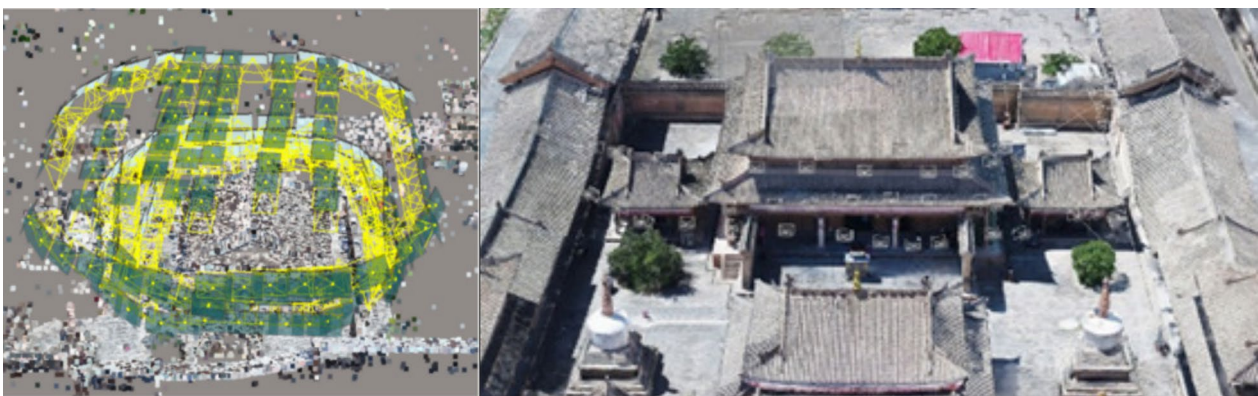


Fig. 3 A fully structured Color point cloud of BaoGuang building

Experiments and results

Experimental data

Qutan Monastery is located at Mabugoukou, 21 km south of Ledu County, Haidong City, Qinghai Province as is shown in Fig. 2. It was firstly built in 1392 with a construction area of 27,000 hectare, carrying a history of more than 600 years till now. It keeps the most intact Ming-style building groups in northwestern China.

To obtain the 3D point cloud of Qutan Monastery, Using FARO X330 and DJI drones to collect point clouds and images of Qutan Temple. The 1620 UAV images are collected by DJI Phantom4 which was composed of a FC6310R camera with a $13.2 \times 8.8 \text{ mm}^2$ sensor size and a $2.41 \mu\text{m}$ pixel size. The flight path surrounded the building as is shown in Fig. 3. The distance from the exposure points to this building varied from 20 to 85

m. Considering the 8.8 mm focal length and the photographic distance, the ground sampling distance (GSD) for all cameras ranged from 0.5 cm to 2 cm. Relying on the commercial software package Bentley, this DIM point cloud was generated. The density of the generated DIM point cloud was $59,737 \text{ points/m}^2$. Finally, cut out the scene of the BaoGuang Hall for the convenience of subsequent annotation work. The integrated complete point cloud is shown in Fig. 3.

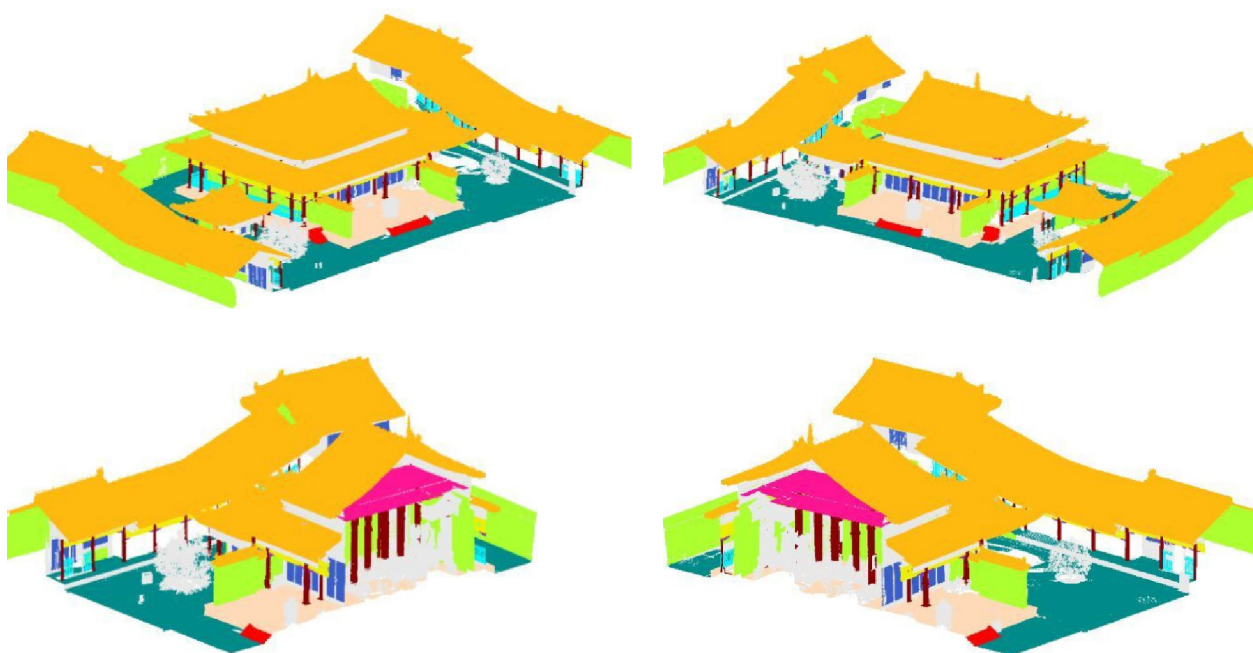
The collected internal and external data of the BaoGuang Hall (Zhongdian) and its surrounding scenes.

Evaluation criteria

Neural networks find it challenging to effectively learn all structural features when training models with only a small portion of point clouds and inferring the rest.



(a) A fully structured Color point cloud of Baoguang building



(b) Point cloud manual annotation results

Fig. 4 Experimental data display

Considering the scarcity of benchmark datasets for semantic segmentation of Chinese ancient architecture, this paper adopts a two-fold cross-validation method as the evaluation standard for experiments.

The twofold Cross Validation method calculates the average overall accuracy (OA), mean intersection over union (mIOU) and mean accuracy (mAcc) to evaluate the segmentation performance of different network models on the ancient building dataset. Initially, the left part of the point cloud is used as training data, and the right part is used as testing data to compute the OA, mIOU and mAcc of the network model on the right part

of the point cloud. Subsequently, the network model is retrained using the right part of the point cloud, and the OA, mIOU and mAcc on the left part of the point cloud are calculated. Finally, the average of the two is taken as the final result.

Training process

Training data

The Qutan Temple dataset comprises 48,788,720 labeled points, and this article divides the ancient building point cloud into 12 categories: roof, ceiling, column, bracket, sparrow, doors and windows, wall, plinth, railing, steps,

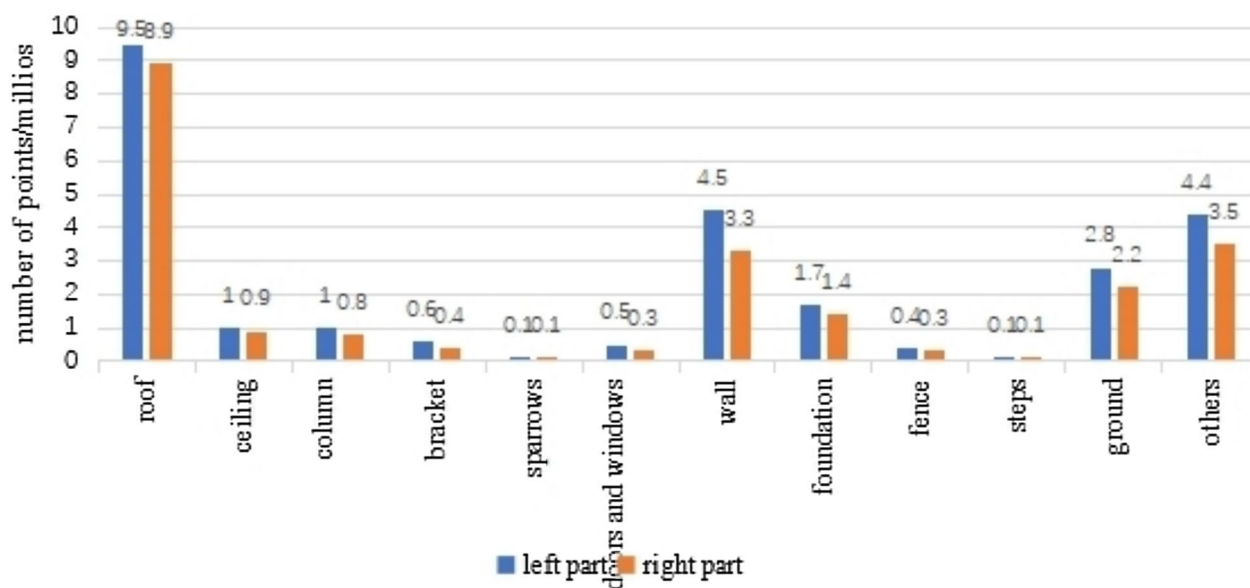


Fig. 5 The quantity distribution statistics of points cloud dataset

Table 1 Segmentation results produced by different DNN model

Index	PointNet	PointNet++	DGCNN	LDGCNN	GACNet	MP-DGCNN (ours)
OA	0.7497	0.7699	0.8628	0.8035	0.8564	0.9019
mIOU	0.3828	0.3537	0.5696	0.5099	0.5311	0.6534
mAcc	0.4811	0.4724	0.6671	0.6283	0.6399	0.7941

ground, and others. The semantic label of the Baoguang Hall point cloud was manually implemented, and the labeled results are shown in Fig. 4. After labeling, the scene was cut along the central axis into left and right point clouds for subsequent twofold cross-validation experiments. The left half of the point cloud was used as the training set in the first fold, with the right half serving as the testing set. For the second fold, these roles were reversed, with the right half used for training and the left half for testing. This division guarantees that both subsets maintain a consistent distribution of architectural features.

The number of points within each component is shown in Fig. 5. The roof category contained the highest number of points, approximately 9 million points, while the categories such as Sparrow and Step only have 10,000 points. To balance space distribution of the data, the training data is resampled at 0.01m intervals to ensure uniform distribution of points. After sampling, the left part of the point cloud contains 26,531,655 points, while the right part contains 22,257,065 points.

Experiment environment and parameter setting

This experiment was conducted on Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz; NVIDIA GeForce RTX 2080Ti with 11GB of VRAM; Pycharm IDE, Python3.7, and Tensorflow2.3 environment.

During the experimental training phase, the number of iterations was set as 150, the initial learning rate was set as 0.001, the learning decay rate was 0.5, and the Batch Size parameter was set to 4. The k-nearest neighbor parameter for MP-DGCNN was set to 20.

Experimental results and analysis

Experimental results

The point cloud semantic segmentation experiments were conducted on the Baoguang Temple dataset using PointNet, PointNet++, DGCNN, LDGCNN, GACNet and MP-DGCNN. Based on the twofold cross-validation, the segmentation accuracy of different network models was shown in Table 1.

Table 1 showed that the DL method based on the graph structure including DGCNN, LDGCNN, and MP-DGCNN achieved better segmentation accuracy

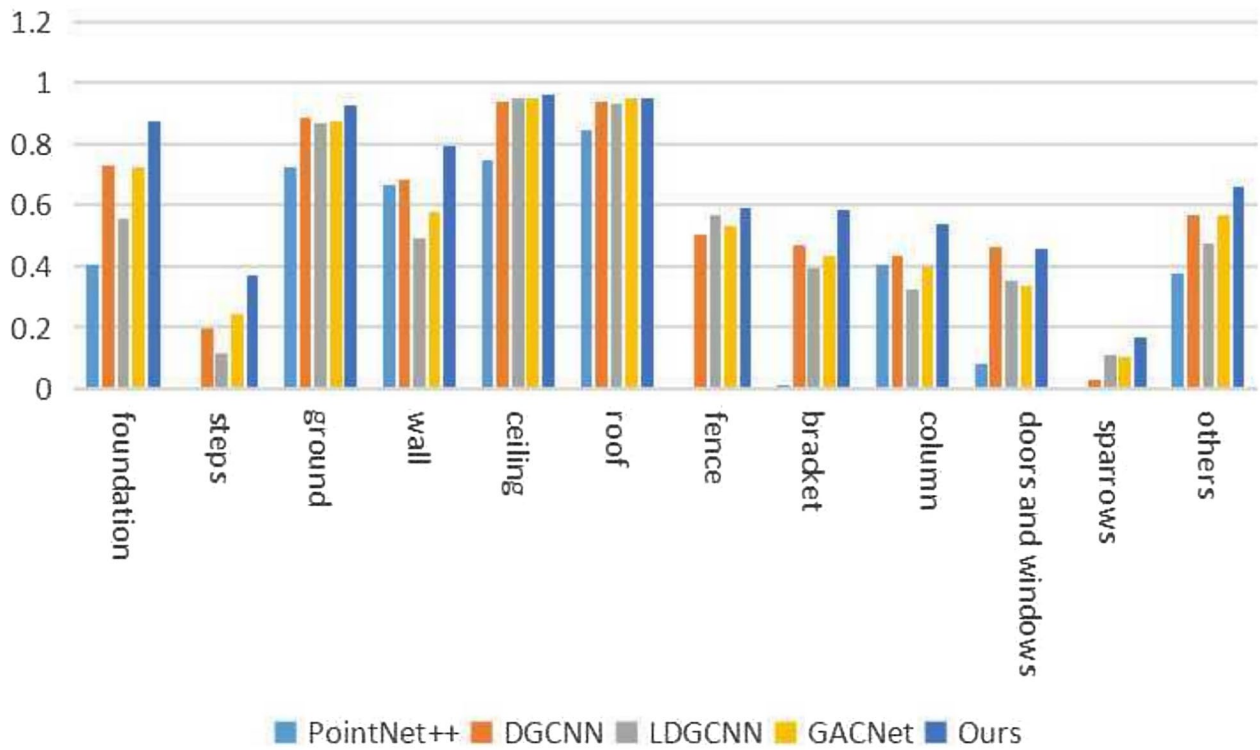


Fig. 6 OA with different DNN model

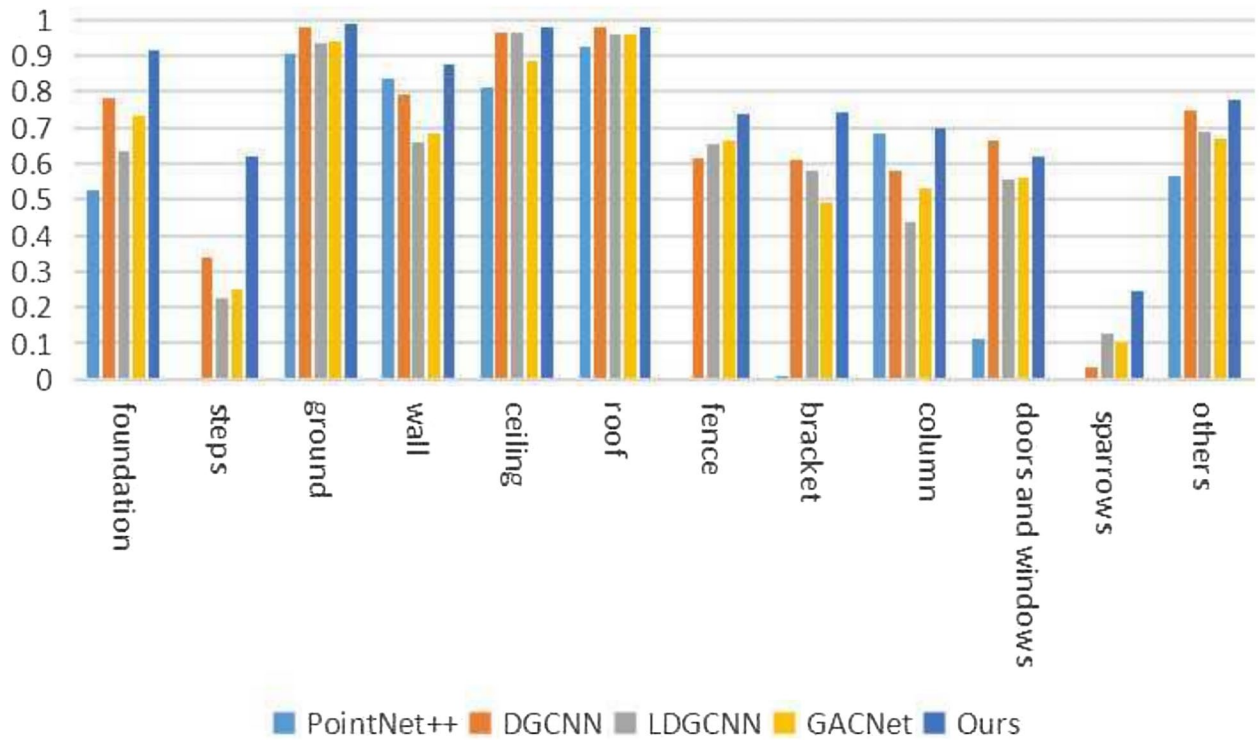


Fig. 7 IOU with different DNN model

than that of PointNet and PointNet++ due that point's local structures are learned through edge convolution which can capture the topological relationship of points and better handle the unordered nature of point clouds. Among the methods based on the graph structure, the MP-DGCNN designed in this paper achieved an overall accuracy (OA) of 90.19%, a mean intersection over union (mIOU) of 65.34% and a mean accuracy (mAcc) of 79.41% for the point cloud semantic segmentation on the Baoguangdian dataset. Compared with DGCNN,

the improved MP-DGCNN increased the overall OA by 3.91%, mIOU by 8.38% and mAcc by 16.58%.

Figures 6 and 7 showed the OA and IoU of different components by different networks. MP-DGCNN achieved good results and the segmentation precision of pedestals, floors, walls, ceilings, roofs, fences, brackets, and columns and reached about 70% or more. Except for windows and doors, the segmentation IoU of all components also increased.

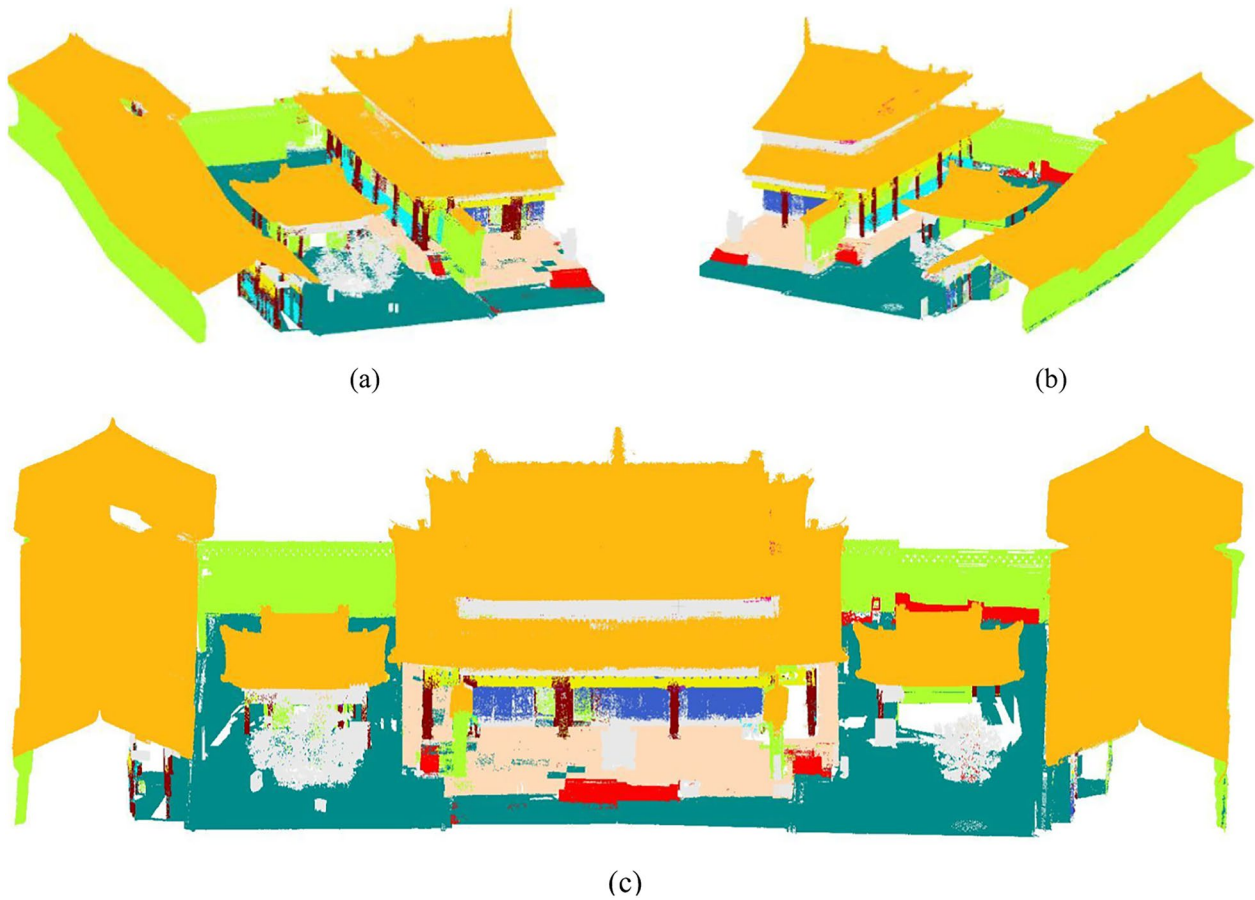


Fig. 8 Points cloud semantic segmentation for Baoguang building using MP-DGCNN

Table 2 Segmentation influence with different selection of pooling function

Index	The designed edge feature	Max-pooling	Mean-pooling	Sum-pooling	Mix-pooling	OA	mIOU
1	–	√	–	–	–	86.28	56.96
2	√	√	–	–	–	89.39	63.08
3	√	–	√	–	–	87.02	58.39
4	√	–	–	√	–	87.67	60.11
Ours	√	–	–	–	√	90.19	65.34

- a. MP-DGCNN achieved segmentation accuracies of 91.34%, 62.15%, and 98.94% for the foundation, treads, and ground (with surface features), respectively. The accuracies were 28.07%, 13.34%, and 1.24% higher than those achieved by DGCNN. This indicated that MP-DGCNN had advantages in capturing subtle differences between structures and can learn fine-grained features of point cloud structures compared with DGCNN.
- b. Due that edge features were strengthened by introducing distance and the coordinates of neighboring points and the mixed pooling strategy which can retain more useful features and alleviate information loss was applied in the process of MP-DGCNN, the segmentation accuracy of fence, bracket, and column also increased by 12.19%, 13.58%, and 11.79% compared with DGCNN.
- c. PointNet and PointNet++ was hardly separated sparrows from the large-scale point clouds scene and the accuracy of sparrows based on DGCNN only reached 3.16%. Compared with DGCNN, MP-DGCNN has a significant improvement in overall accuracy of sparrow segmentation, with an increase of 21.50%, but the overall accuracy is still not high.

Based on the above analysis, the experimental results showed that the proposed MP-DGCNN in this study is more effective and has stronger generalization ability compared to other networks. Figure 8 showed the semantic segmentation results of MP-DGCNN.

Experimental results analysis

(1) The Impact of edge features and Mix Pool function

To analyze the advantages of the designed edge features and mixed pooling function on the semantic segmentation results, several of comparative experiments were conducted. DGCNN was used in the first experiment; the method used in the second, third and fourth experiments applied the designed edge features and the pooling function was max-pooling, mean-pooling and sun-pooling; the MP-DGCNN was used in the fifth experiment.

In Table 2, the symbol " $\sqrt{\text{m}}$ " and "-" indicates the corresponding method is used. As is shown in Table 2, OA of the five experimental groups are 86.28%, 89.39%, 87.02%, 87.67%, and 90.19%, and the mIOUs are 56.96%, 63.08%, 58.39%, 60.11%, and 65.34%, respectively. Specifically,

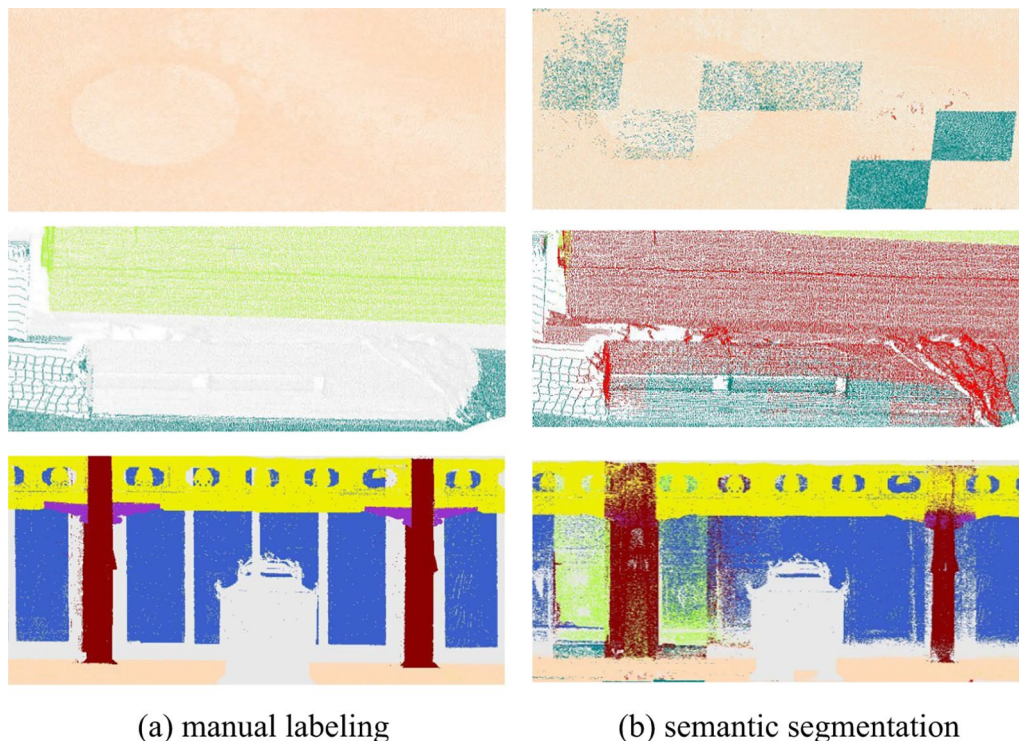


Fig. 9 Some over-segmentation using MP-DGCNN model

- (a) In the first and second experiments, the same pooling function (max pooling) was used. Both OA and mIOU of the second experiment were higher than that of the first experiment. The experimental results showed the designed edge feature which contains the distance information of graph nodes and the neighborhood information is more comprehensive and can better characterize the topological relationship between points.
- (b) According to the second, third and fourth experiments, the max-pooling method in the second experiment obtained the best performance. These experimental results are consistent with the conclusions in references.
- (c) Due that the mix pooling function was applied in MP-DGCNN, OA and mIOU of the fifth experiments increased by 0.8% and 2.26% compared with the second experiment in which the designed edge feature was also used. This further confirmed that the proposed mixed pooling module alleviates the problem of information loss caused by max-pooling to a certain extent. This is because the mixed pooling function optimizes the distribution of graph features through self-learning of pooling rules and retained more feature information.

(2) Limitations

Although the MP-DGCNN can obtain better performance, there are still some limitations as is shown in Fig. 9. Some points belong to ground are misclassified as pedestals. The shape of these misclassified points appeared as regular squares. This is because the shape of the fragmented ground which was derived by the method for dividing building blocks is similar to the shape of pedestal. The difference between the fragmented ground and pedestal is the absolute position. Similarly, the incense burner within other categories on the ground appeared as a stepped shape and was misclassified as pedestals.

Moreover, the components connecting to the windows and doors are also misclassified due that the position, shape, and color of the windows, doors and these components are similar. This resulted that segmentation boundaries were clear and are misclassified as walls.

It can be seen that the MP-DGCNN semantic segmentation network can learn point cloud features from data. However, when the points within the different categories similar in position, shape, and color, the misclassification of points still occurred. One hand, the proposed method can't capture all the features. On the other hand, besides on the attributes such as geometric shape and color, other knowledge (the connection) should be considered in the process of recognize the components of ancient architecture. Therefore, there is still a lot of room for research on semantic segmentation networks for ancient architecture.

Robustness test

To evaluate the robustness of the MP-DGCNN, we conducted experiments on the Qutan Temple dataset with varying point cloud densities. Specifically, we modified the density of the point clouds to 25%, 50%, and 75% of their original density to test how well MP-DGCNN maintains its performance under conditions of data sparsity. For each density level, the point clouds were randomly sub-sampled to the specified percentage, ensuring a representative but reduced set of data points. The model was then tested on these modified datasets to measure any impacts on the Overall Accuracy (OA) and mean Intersection over Union (mIOU). The results are shown in Table 3. The result shows a minimal decrease in performance metrics with a reduction in point cloud density, demonstrating the model's robustness to variations in input data quality. MP-DGCNN achieves high robustness and generalization across varying point cloud densities through advanced feature extraction and pooling strategies. These methods effectively capture and maintain crucial geometric and contextual information, enabling the model to deliver consistent and accurate segmentation results even under diverse data conditions.

Conclusion

The semantic segmentation of ancient building point clouds serves as the cornerstone for the 3D reconstruction of ancient structures. In this study, we introduced MP-DGCNN, a deep learning-based approach tailored for the semantic segmentation of ancient building point clouds. This method effectively labels the semantics of ancient architectural elements from point cloud data with robustness and high precision. By incorporating modified edge features and a hybrid pooling strategy

Table 3 Segmentation influence with different selection of pooling function

Point cloud density	OA	mIOU
25%	0.8745	0.6229
50%	0.8897	0.6355
75%	0.8966	0.6478

into the original DGCNN framework, MP-DGCNN demonstrates improved capability in capturing nuanced features. The performance of MP-DGCNN was rigorously evaluated using the Qutan Temple dataset, yielding an overall accuracy of 90.19%, a mean Intersection over Union (mIOU) of 65.34% and a mean accuracy (mAcc) of 79.41%. Furthermore, comparative analysis with other point cloud segmentation algorithms such as PointNet++, PointNet, DGCNN, LDGCNN and GAC-Net revealed that MP-DGCNN consistently outperforms these methods in terms of overall accuracy and mIOU.

While our preliminary research demonstrates the enhanced accuracy of ancient building semantic segmentation achieved by MP-DGCNN, future research endeavors will focus on several fronts.

Firstly, we intend to establish a benchmark for typical components of ancient architecture to build a comprehensive and robust data foundation. This is crucial for enabling the DNNs to accurately learn the intricate details of ancient architecture. It also standardizes testing conditions, allowing researchers to conduct controlled, comparable experiments. Additionally, we aim to integrate targeted attention mechanisms to address the issue of incorrect semantic segmentation often caused by similar shapes. This approach will focus on refining our feature extraction methods to more precisely and efficiently capture the complex features of ancient wooden buildings. This will expand the application scope of DNNs in cultural heritage scenarios. Finally, we aim to investigate the possibility of incorporating prior knowledge of ancient architecture into semantic segmentation tasks. By incorporating this specialized knowledge, the DNNs can form more accurate and contextually appropriate segmentations, enhancing the reliability and precision of the models in analyzing and interpreting architectural elements from historical periods.

Acknowledgements

This research was funded by National Key Research and Development Program of China, Grant Number 2022YFF0904300, National Natural Science Foundation of China, Grant Number 42171356, 42171444, and 42301516, Beijing Municipal Education Commission- Municipal Education Commission Joint Fund Project, grant number KZ202110016021, research Project of Beijing Municipal Education Commission—General Project of Science and Technology Plan, grant number KM202110016005 and Beijing University of Civil Engineering and Architecture Special funds for basic scientific research business expenses of municipal universities, grant number X20043.

Author contributions

Conceptualization, Y.D. and C.Z.; methodology, C.Z.; software, C.Z.; resources, Y.D.; writing—original draft preparation, C.Z.; writing—review and editing, Y.D., C.Z. and M.H.; supervision, Y.D.; project administration, C.Z., Y.D. and M.H.; funding acquisition, M.H. All authors have read and agreed to the published version of the manuscript.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Competing interests

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.

Received: 15 February 2024 Accepted: 21 May 2024

Published online: 22 July 2024

References

- Liu J, Wu ZK. Rule-based generation of ancient Chinese architecture from the song dynasty. *J Comput Cult Herit*. 2015;9(2):1–22. <https://doi.org/10.1145/2835495>.
- Hu Q, Wang S, Fu C, Ai M, Yu D, Wang W. Fine surveying and 3D modeling approach for wooden ancient architecture via multiple laser scanner integration. *Remote Sensing*. 2016;8(4):270. <https://doi.org/10.3390/rs8040270>.
- Biryukova MV, Nikonova AA. The role of digital technologies in the preservation of cultural heritage. *Muzeológia a kultúrne dedičstvo*. 2017;5:1.
- Adane A, Chekole A, Gedamu G. Cultural heritage digitization: challenges and opportunities. *Int J Comput Appl*. 2019;178(33):1–5. <https://doi.org/10.5120/ijca2019919180>.
- Shizhen J, Yi L, Yuqing X, Bo Z, Xiangbin M, Ke Q. Conservation and management of Chinese classical royal garden heritages based on 3D digitalization - a case study of Jianxin courtyard in Jingyi garden in fragrant hills. *J Cult Herit*. 2022;58:102–11. <https://doi.org/10.1016/j.culher.2022.09.020>.
- Lyn W, Alastair R, Adam F, James H. 3D digital documentation for disaster management in historic buildings: applications following fire damage at the Mackintosh building, The Glasgow School of Art. *J Cult Herit*. 2018;31:24–32. <https://doi.org/10.1016/j.culher.2017.11.012>.
- Yusheng X, Xiaohua T, Uwe S. Voxel-based representation of 3D point clouds: methods, applications, and its potential use in the construction industry. *Autom Constr*. 2021;126: 103675. <https://doi.org/10.1016/j.autcon.2021.103675>.
- Xiaoqiang T, Deke G, Yulan G, Xiaolei Z, Zhong L. CloudNavi: toward ubiquitous indoor navigation service with 3D point clouds. *ACM Trans Sensor Netw*. 2019;15(1):1–28. <https://doi.org/10.1145/3216722>.
- Croce V, Caroti G, De Luca L, Jacquot K, Piemonte A, Véron P. From the semantic point cloud to heritage-building information modeling: a semiautomatic approach exploiting machine learning. *Remote Sens*. 2021;13(3):461. <https://doi.org/10.3390/rs13030461>.
- Xiucheng Y, Pierre G, Mathieu K, Hélène M, Arnadi M, Tania L. Review of built heritage modelling: integration of HBIM and other information techniques. *J Cult Herit*. 2020;46:350–60. <https://doi.org/10.1016/j.culher.2020.05.008>.
- Li L, Tang L, Zhu H, Zhang H, Yang F, Qin W. Semantic 3D modeling based on CityGML for ancient Chinese-style architectural roofs of digital heritage. *ISPRS Int J Geo Inf*. 2017;6(5):132. <https://doi.org/10.3390/ijgi6050132>.
- Juan M, Javier L, Juan E, Nieto J, Silvana B. Semantic interpretation of architectural and archaeological geometries: point cloud segmentation for HBIM parameterization. *Autom Constr*. 2021;130: 103856. <https://doi.org/10.1016/j.autcon.2021.103856>.
- Mingtao F, Liang Z, Xuefei L, Syed ZG, Ajmal M. Point attention network for semantic segmentation of 3D point clouds. *Pattern Recogn*. 2020;107: 107446. <https://doi.org/10.1016/j.patcog.2020.107446>.
- Feng C, Fei W, Guangwei G, Yimu J, Jing X, Guoping J, Xiaoyuan J. JSPNet: learning joint semantic & instance segmentation of point clouds via feature self-similarity and cross-task probability. *Pattern Recogn*. 2022;122: 108250. <https://doi.org/10.1016/j.patcog.2021.108250>.
- Hejun W, Enyong X, Jinlai Z, Yanmei M, Jin W, Zhen D, Zhengqiang L. BushNet: effective semantic segmentation of bush in large-scale point clouds. *Comput Electr Agric*. 2022;193: 106653. <https://doi.org/10.1016/j.compag.2021.106653>.

16. Lee MS, Yang SW, Han SW. Gaia: Graphical information gain based attention network for weakly supervised point cloud semantic segmentation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2023:582–591. <https://doi.org/10.48550/arXiv.2210.01558>.
17. Xie Y, Tian J, Zhu XX. Linking points with labels in 3D: a review of point cloud semantic segmentation. *IEEE Geosci Remote Sens Mag*. 2020;8(4):38–59. <https://doi.org/10.1109/MGRS.2019.2937630>.
18. Martin W, Boris J, Stefan H, Clément M. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS J Photogramm Remote Sens*. 2015;105:286–304. <https://doi.org/10.1016/j.isprsjprs.2015.01.016>.
19. Grilli E, Remondino F. Classification of 3D digital heritage. *Remote Sens*. 2019;11(7):847. <https://doi.org/10.3390/rs11070847>.
20. Teruggi S, Grilli E, Russo M, Fassi F, Remondino F. A hierarchical machine learning approach for multi-level and multi-resolution 3D point cloud classification. *Remote Sens*. 2020;12(16):2598. <https://doi.org/10.3390/rs12162598>.
21. Dong Y, Li Y, Hou M. The point cloud semantic segmentation method for the Ming and Qing Dynasties' official-style architecture roof considering the construction regulations. *ISPRS Int J Geo Inf*. 2022;11(4):214. <https://doi.org/10.3390/ijgi11040214>.
22. Dong Y, Hou M, Xu B, Li Y, Ji Y. Ming and Qing dynasty official-style architecture roof types classification based on the 3D point cloud. *ISPRS Int J Geo Inf*. 2021;10(10):650. <https://doi.org/10.3390/ijgi10100650>.
23. Li H, Yongmei L, Chaoguang M. A novel 3D point cloud segmentation algorithm based on multi-resolution supervoxel and MGS. *Int J Remote Sens*. 2021;42(22):8492–525. <https://doi.org/10.1080/01431161.2021.1978583>.
24. George V, Maximilian C, Franz R. Contextual segment-based classification of airborne laser scanner data. *ISPRS J Photogramm Remote Sens*. 2017;128:354–71. <https://doi.org/10.1016/j.isprsjprs.2017.03.010>.
25. Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018;4558–4567. <https://doi.org/10.1109/cvpr.2018.00479>.
26. Ulku I, Akagündüz E. A survey on deep learning-based architectures for semantic segmentation on 2d images. *Appl Artif Intell*. 2022;36(1):2032924. <https://doi.org/10.1080/08839514.2022.2032924>.
27. Jiang B, An X, Xu S, Chen Z. Intelligent image semantic segmentation: a review through deep learning techniques for remote sensing image analysis. *J Indian Soc Remote Sens*. 2023;51(9):1865–78. <https://doi.org/10.1007/s12524-022-01496-w>.
28. Cui Y, Chen R, Chu W, Chen L, Tian D, Li Y, et al. Deep learning for image and point cloud fusion in autonomous driving: a review. *IEEE Trans Intell Transp Syst*. 2021;23(2):722–39. <https://doi.org/10.1109/TITS.2020.3023541>.
29. Geng X, Ji S, Lu M, Zhao L. Multi-scale attentive aggregation for LiDAR point cloud segmentation. *Remote Sens*. 2021;13(4):691. <https://doi.org/10.3390/rs13040691>.
30. Hu X, Yuan Y. Deep-learning-based classification for DTM extraction from ALS point cloud. *Remote Sens*. 2016;8(9):730. <https://doi.org/10.3390/rs8090730>.
31. Alexandre B, Joris G, Bertrand LS, Nicolas A. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. *Comput Graph*. 2018;71:189–98. <https://doi.org/10.1016/j.cag.2017.11.010>.
32. Qin N, Hu X, Wang P, Shan J, Li Y. Semantic labeling of ALS point cloud via learning voxel and pixel representations. *IEEE Geosci Remote Sens Lett*. 2019;17(5):859–63. <https://doi.org/10.1109/LGRS.2019.2931119>.
33. Zhou Y, Tuzel O. Voxelnet: End-to-end learning for point cloud based 3d object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018;4490–4499. <https://doi.org/10.1109/CVPR.2018.00472>.
34. Klokov R, Lempitsky V. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models. *Proceedings of the IEEE international conference on computer vision*. 2017;863–872. <https://doi.org/10.1109/ICCV.2017.99>.
35. Qi CR, Su H, Mo K, Guibas LJ. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017;652–660. <https://doi.org/10.1109/cvpr.2017.16>.
36. Qi CR, Yi L, Su H, Guibas, L. J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*. 2017;30. <https://doi.org/10.48550/arXiv.1706.02413>.
37. Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM. Dynamic graph cnn for learning on point clouds. *ACM Trans Graph*. 2019;38(5):1–12. <https://doi.org/10.1145/3326362>.
38. Zhang K, Hao M, Wang J, CW de Silva, C Fu. Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features. *arXiv preprint*. 2019; [arXiv:1904.10014](https://arxiv.org/abs/1904.10014). <https://doi.org/10.48550/arXiv.1904.10014>.
39. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017;4700–4708. <https://doi.org/10.1109/CVPR.2017.243>.
40. Yijun S, Hui H. A weighted point cloud classification network based on dynamic graph convolution. *Computer Engineering and Applications*. 2021:1–8.
41. Wang JG, He J, Pang DW. Point cloud classification and segmentation network based on dynamic graph convolutional network. *Laser Optoelectron Prog*. 2021;58(12):1215008.
42. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770–778. <https://doi.org/10.1109/IEEECONF.1997.85951>.
43. Wu Z, Song S, Khosla A, F Yu, L Zhang, X Tang, J Xiao. 3d shapenets: a deep representation for volumetric shapes. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015:1912–1920. <https://doi.org/10.1109/CVPR.2015.7298801>.
44. Chang AX, Funkhouser T, Guibas L, Hanrahan P, Huang Q, Li Z, et al. Shapenet: an information-rich 3d model repository. *arXiv preprint*. 2015; [arXiv:1512.03012](https://arxiv.org/abs/1512.03012). <https://doi.org/10.48550/arXiv.1512.03012>.
45. Armeni I, Sax S, Zamir A R, S Savarese. Joint 2d-3d-semantic data for indoor scene understanding. *arXiv preprint*. 2017; [arXiv:1702.01105](https://arxiv.org/abs/1702.01105). <https://doi.org/10.48550/arXiv.1702.01105>.
46. Pierdicca R, Paolanti M, Matrone F, Martini M, Morbidoni C, Malinverni ES, Frontoni E, Lingua AM. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sens*. 2020;12(6):1005. <https://doi.org/10.3390/rs12061005>.
47. Matrone F, Martini M. Transfer learning and performance enhancement techniques for deep semantic segmentation of built heritage point clouds. *Virtual Archaeol Rev*. 2021;12(25):73–84. <https://doi.org/10.4995/var.2021.15318>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.