3D Point Cloud Classification Method Based on Multiple
Attention Mechanism and Dynamic Graph Convolution

3D Point Cloud Classification Method Based on Multiple Attention Mechanism and Dynamic Graph Convolution

Yu Zhang, Zilong Wang

School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai, China; e-mail: zhangyu95900@163.com

Yongjian Zhu

School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai, China; e-mail: zhuyongjian_hn@126.com

Corresponding author: zhuyongjian_hn@126.com

In order to solve the problem of uneven density and the low classification accuracy of 3D point cloud, a 3D point cloud classification method fuses multi-attention machine is proposed. It is principally based on the traditional point cloud dynamic graph convolution classification network, into multiple attention mechanisms, including self-attention, spatial attention and channel attention mechanisms. The self-attention mechanism can reduce the dependence on irrelevant points while aligning point clouds, and input the processed point cloud into the classification network. Then the missing geometric information in the classification network is compensated by the integration of spatial and channel attention mechanisms. The experimental results on the public data set ModelNet40 indicate that compared with the DGCNN classification network, the improved network model improves the classification accuracy of the data set by 0.5 % and the average accuracy by 0.9 %. Meantime, the classification accuracy outstrips other contrast classification algorithms.

KEYWORDS: 3D point cloud classification, dynamic graph convolution, self-attention, spatial attention, channel attention.

1. Introduction

With the gradual rise of laser radar, point cloud processing has gradually become a hot focus. Deep learning of point cloud is widely used in computer vision, artificial intelligence, autonomous driving and other fields. However, compared with image data, point cloud data is characterized by sparsity, disorder and uneven spatial distribution, makes it difficult to process point cloud data directly. Owing to the special way of processing point cloud by neural network model, the processing method based on point cloud is still in the rising stage. With increasing the development of deep learning recently, it has provided strong support for the research of point cloud, and put forward some innovative algorithm models to deal with point cloud.

For the shape classification problem of 3D point cloud, methods can be divided into traditional methods [4] and the rise of Transformer-based methods [26, 10, 5, 30]. Traditional methods mainly involve: point-based method, volume-based method, and multi-view-based method. In point-based method for point classification, it can be divided into point-by-point MLP method [11], graph-based method [29, 15, 7, 28], convolution-based method [1] and hierarchical data structure-based method [6]. Qi et al. proposed the PointNet model [11] that can use point cloud directly as input to realize the replacement invariance of symmetric functions, and use MLP layer to learn point cloud features and the maximum pooling layer to extract global features. However, PointNet cannot capture local structural details between points, making it difficult to obtain good results with less extracted feature information effect. Based on the original model, Qi proposed PointNet++ model [12]. By superimposing the sampling layer, the grouping layer and learning layer of PointNet, PointNet++ can learn features from local geometric structure and abstract local features layer by layer. Using the distance of the metric space, local features can be learned by using the growth of the context scale. The direct processing of the original points increases the computational complexity of the model. At the same time, due to the sparsity of the original points, the performance and accuracy of the model are limited. Li et al. proposed the PointCNN model [8] that can integrate the CNN convolution module into the point cloud processing. The point cloud is converted by X-Conv, and then the transformed features are convoluted. However, the selection of feature points is based on random

down-sampling and farthest point sampling method, the resulting point cannot represent the core point. Te et al. proposed the RGCNN model [18], which regards the characteristics of point as graphic signals, uses the feature matrix and adjacency matrix of irregular point cloud as inputs, and updates the graphic Laplacian matrix in each layer of RGCNN to adaptively capture the structure of the dynamic graph. Converting point cloud features into dynamic graph features can reduce the problem of insufficient feature extraction caused by local missing of point cloud data to a certain extent, and the dynamic graph can ensure that point cloud features are not lost and the accuracy of the model is guaranteed. Therefore, this paper abandons the method of directly extracting point cloud features in the original intention of model design, and transforms point cloud features into dynamic graph features for extraction and operation. Simonovsky et al. proposed the ECC model [15], which can solve graphs of different sizes and connectivity by extending the convolution operation from grid to arbitrary graphs while avoiding the frequency domain. Wang et al. proposed the DGCNN model [22], taking EdgeConv as the core layer, constructing a graph in the feature space, and dynamically updating it after each layer of the network. However, due to the limitation of edge convolution, local geometric information is lost in classification. Volume-based methods usually convert point cloud voxels into 3D meshes, and then apply 3D convolutional neural network (CNN) to spatial representation for shape classification. Maturana et al. introduced a volume occupancy network called VoxNet model [9] is used to achieve robust 3D object recognition. Wu et al. proposed ShapeNet model [25] to learn the distribution of point clouds from different 3D shapes. However, volume-based methods are difficult to expand to dense 3D data, and with the increase of data, the calculation amount and memory usage of the model will also increase. Li et al. proposed the SO-Net model [6], which can systematically adjust the receptive field overlap to perform hierarchical feature extraction by performing point-to-node KNN search on SOM.

The Transformer-based method [18, 23, 14, 31] is an algorithm that has emerged in the past two years. Transformer is based on self-attention (SA), which was initially used for natural language processing, and then gradually applied to computer vision due to

its strong feature representation ability. Vaswani et al. proposed the Transformer model [19], which uses the attention mechanism to construct the global dependency between input and output. With the wide application of attention mechanisms, new attention mechanisms have gradually emerged. Woo et al. proposed the CBAM model [24] to represent the attention mechanism module of the convolution module, replacing a single attention mechanism module with a combination of spatial attention mechanism module (SAM) and channel attention mechanism module (CAM). The spatial and channel attention mechanisms are fused to make the network pay more attention to useful features and enhance the expression ability of features according to the attention scores between different points. Therefore, when designing the model, this paper adds the CBAM module to the classification task to enhance the performance of point cloud features in the classification task. Guo et al. applied Transformer to the region of point cloud processing firstly and achieved remarkable effects [3]. By taking advantage of the inherent order invariance of Transformer, the definition of point cloud order is avoided, and feature learning is conducted through the attention mechanism. Wang et al. improved Transformer's capture of point cloud semantics and distinguishing features, and proposed a new point cloud Transformer Info PCT model [20]. This model combines self-attention mechanism and mutual information maximization, which is helpful for deep learning of point clouds. Although Transformer has powerful features, there are some drawbacks. Transformer performs poorly in extracting low-level features, which can easily lead to wrong predictions of small targets and loss of local information. Moreover, completely based on the self-attention mechanism, there is a certain loss of location information between points.

In view of the single function of spatial transform in DGCNN and the inability to filter irrelevant points and the lack of local information points in edge convolution, a method to improve the accuracy of point cloud classification is proposed in this paper. Based on the existing model DGCNN, the self-attention mechanism is used to replace the original model module spatial transform. While dealing with the invariance of point cloud displacement, it can more effectively extract point cloud features, remove irrelevant

points, and input them to the classification backbone network. However, the defect of local information loss in the classification network is still presence. Therefore, on its classification backbone network, spatial attention and channel attention mechanisms are added to further improve the classification accuracy by fusing spatial and channel attention mechanisms to compensate for missing geometric information and enhance local feature information.

The main contributions of this paper are as follows:
- First, the self-attention mechanism is introduced to replace the spatial transform module in DGCNN. The self-attention mechanism has inherent substitution invariance when processing a series of points, which can well replace the spatial transform module. Meanwhile, it can decrease the feature extraction of irrelevant points and reduce the calculation of the model.
- Second, based on self-attention mechanism, spatial attention and channel attention mechanisms are introduced. A channel attention and a spatial attention module are added after the edge convolution, and the two attention modules are fused to construct a local neighborhood graph of the points by the relative positional relationship between the points. The spatial attention mechanism looks for the most important part of the network to process. The channel attention mechanism can strengthen or suppress different channels for other different tasks by modeling the importance of each feature channel.
- Third, the network with self-attention mechanism, channel and spatial attention mechanism can effectively use the spatial information of point cloud data, improve the classification accuracy of point cloud, and perform the classification task on the public datasets ModelNet40, and achieve good results.

2. Related Work

Zhai et al. proposed a multi-scale dynamic GCN model [27] for point cloud classification, combining the farthest point sampling algorithm and K-NN graph method to sparse the scale of the entire point cloud set scale and locate the neighborhood for each central data point. At the same time, the edge convolu-

tion operation is used to extract the composite features between the adjacent connection points and the center point, and the global maximum pool layer is used to extract the global semantic information for classification. This also shows that the dynamic graph method has a unique role in point cloud classification tasks. At the same time, Wang et al. proposed the DGCNN model [22] which takes EdgeConv as the core layer, constructs a graph in the feature space, and updates dynamically after each layer of the network. The backbone network mainly inputs point cloud data into the spatial transform module for coordinate rotation transformation. Then the processed point cloud data is input into four identical edge convolutions. Finally, the point cloud features are extracted by once full connection maximum pooling operation and twice MLP operation to achieve the goal of classification. The DGCNN classification backbone network utilizes a new neural network module edge convolution and runs the edge convolution module throughout the classification network module. Compared with traditional 2D images, EdgeConv uses k-nearest neighbor (KNN) graphs to construct local features and perform convolution operations. Edge-Conv dynamically updates the features of the points in the point cloud, taking into account both the features of the points and the proximity of k points to the center point in the current feature space. These adjacent points can form a graphical structure, which can be considered as local features. But the model also has some shortcomings. The function of Spatial transform is only to transform the point cloud data without any other operations, which has little effect on the whole model. Although the edge convolution in the classification network can be used to extract features in a dynamic graph way, there will be a lack of local geometric information, which will have a certain impact on the classification of point clouds. In summary, the dynamic graph convolution can be dynamically updated during feature map extraction, making point cloud feature extraction more accurate than global extraction, but it also has inevitable lack of local geometric information.

Hao et al. from Tsinghua University applied the self-attention mechanism to the field of point cloud classification and achieved remarkable results [3]. The proposed model has good semantic feature learning ability and achieves the most advanced performance on several tasks, especially the shape classification of point clouds and parts segmentation. The core of the model is the self-attention mechanism. The self-attention mechanism can effectively capture the global features of point cloud data, and generate accurate attention features based on the input features of the global context, so that accurate feature classification can be performed. On this basis, Gao et al. proposed a model to improve the deep learning of point cloud transformer [2]. Aiming at the problem of single attention calculation and local feature embedding loss in point cloud Transformer, this paper proposes to split multiple heads in the calculation of self-attention weight matrix and calculate the attention score independently to calculate self-attention from multiple aspects. Then, in the neighborhood coding part, multi spatial scale neighborhood feature embedding is added to fuse multiple spatial features. Finally, the spatial position coding embedding is added to the attention feature to enhance the spatial position feature. However, the extracted features are only static features. Compared with the dynamic graph method, there is still a problem that the features are not accurate enough. Wang et al. proposed a graphic attention convolution method [21] for point cloud semantic segmentation. By introducing the graphic attention mechanism, the point cloud is transformed into a graph structure to record neighborhood relationship, and the down-sampling of point cloud features is realized to extract local point cloud features and up-sample point cloud features. However, although the influence of irrelevant points on the extraction process is removed when using the attention mechanism method to extract features, there is still the problem of local information loss, which also limits the further improvement of the attention effect.

3. Our Approach

This paper improves the DGCNN classification network model. In the process of data processing, due to the original spatial transform module has not significantly improved the classification effect obviously, this paper removes the original module and improves it with the self-attention mechanism module. Self-attention mechanism can reduce influence of irrelevant points while performing point cloud coordinate transformation. In the classification network, the

edge convolution only considers the Euclidean distance between the point coordinates and the neighborhood points, thus ignoring the direction between the adjacent points, resulting in the loss of some local geometric information. In this paper, the defects of edge convolution are improved and enhanced. For the defect of local geometric information loss, the channel and spatial attention modules are introduced after each layer of edge convolution, and the two modules are fused to make up for the missing geometric information and enhance the extraction and processing of point cloud features.

The backbone network mainly inputs the point cloud data into the SA module in the form of n × 3, performs coordinate transformation and calculates the attention score, and removes the influence of unrelated points without changing the other characteristics of the point cloud. Then the processed data is input into the classification network through edge convolution processing with 64 channels. The classification network consists of three identical edge convolutions with 64, 64, 128 channels and three identical spatial and channel attention fusion modules. In the end, the point cloud features are extracted by one fully con-

nected maximum pooling and two MLP to achieve the purpose of classification. The edge convolution does not directly generate the feature of the points from the embedding, but generates the edge features describing the relationship between the points and their adjacent points through KNN, which can capture the local geometric structure while maintaining the position invariance.

3.1. Self-Attention Replace Spatial Transform

This paper replaces the spatial transform module. The self-attention mechanism has good substitution invariance, and can well replace the spatial transform module. At the same time, when processing point cloud data, it can reduce the dependence on irrelevant points, so as to better extract the internal features between cloud data.

The sub-attention mechanism Q, K, and V are the query matrix, key matrix, and value matrix generated by the linear transformation of the input features, respectively. First, the way to modify Transformer for point clouds is to treat the entire point cloud as a sentence and each point as a word. Second, when considering a simple point embedding, the interac-

Figure 1

Structure of this model
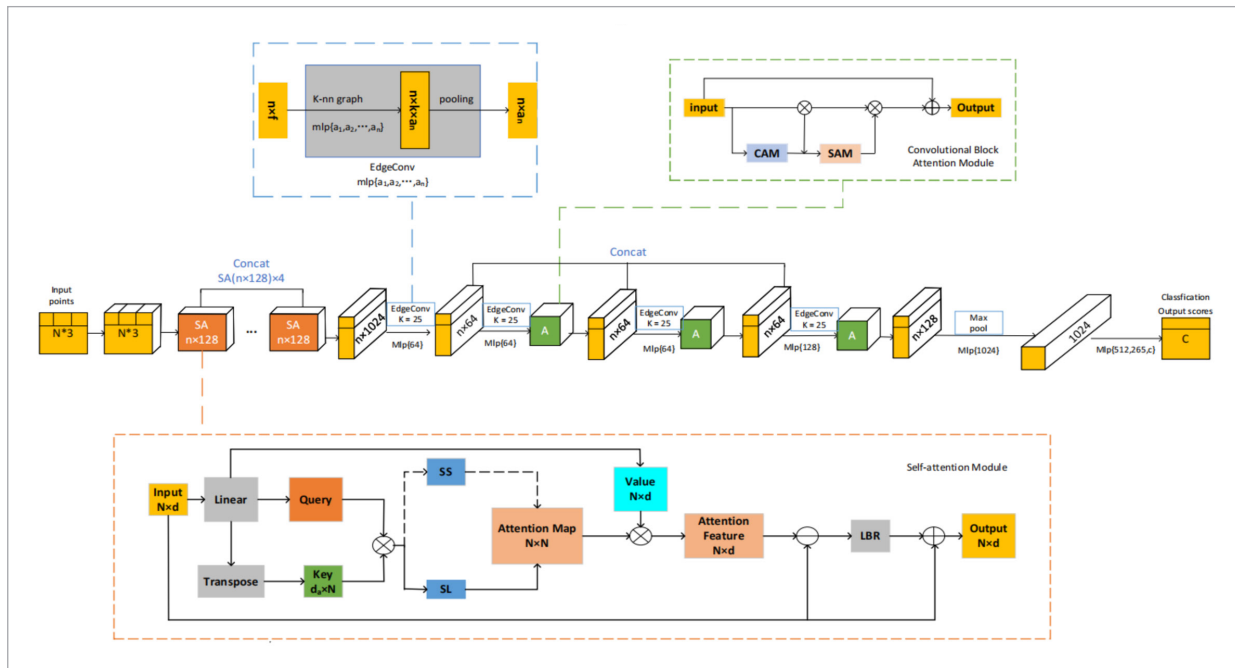
tion between points is ignored. Like word embedding, the purpose of point embedding is to make the point closer to the embedded space in the case of more similar semantics. Specifically, a point cloud is embedded into a d-dimensional feature Fin, and only the 3D coordinates of the point are used as its input feature description. Among them, d = 3 can filter out some distant points and reduce the influence of irrelevant points on feature extraction. Finally, the query, key, value features are fused and nonlinearity is introduced through the activation layer. Specifically, we use the embedded point cloud as query, and value, and the transposed point cloud as key to calculate the attention score and extract the features. The calculated features are fused with the input data to obtain new point cloud features.

The calculation process of the attention feature is as follows:

First, the input feature is transformed by a linear transformation into a query vector Query, a key vector Key and a value vector Value, that is, each input feature is multiplied by three coefficients to obtain three vectors

$$Q = W^q \times F_{in} \tag{1}$$

$$K = W^k \times F_{in} \tag{2}$$

$$K = W^k \times F_{in} \tag{3}$$

Second, the correlation between Query and Key vectors obtained by linear transformation is used to calculate the weight of the self-attention mechanism, and the point multiplication between the two vectors is used to obtain the self-attention weight.

$$A = K^T \times Q. \tag{4}$$

Then use the softmax function to process the weight of the self-attention mechanism to generate the weight matrix of the self-attention mechanism.

$$A = soft\max(K^T \times Q). \tag{5}$$

Finally, the weight matrix A of the self-attention mechanism is multiplied by the Value vector to obtain the final output feature.

$$F_{out} = A \times V + F_{in} \tag{6}$$

3.2. Convergence Channel and Spatial Attention Mechanism

The whole classification network is improved. For the defect of local geometric information loss in edge convolution, channel and spatial attention module are introduced to make up for the defect of geometric information loss. The channel attention mechanism compresses the spatial dimension to the greatest extent while maintaining the channel dimension unchanged, and uses the channel relationship between features to generate the channel attention map. Differ from channel attention, the spatial attention mechanism compresses the channel dimension to the maximum extent, and uses the spatial relationship between features to generate the spatial attention map, which is also a supplement to the channel attention module.

The channel attention formula is:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + $$
$$+ MLP(MaxPool(F)))$$
$$= \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{avg}^c))). \tag{7}$$
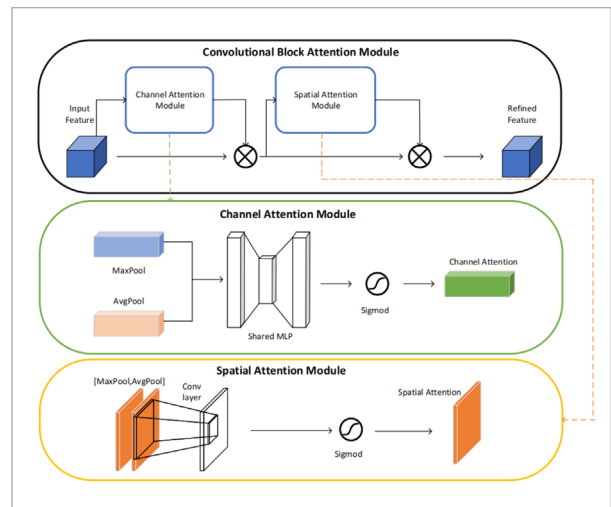
Among them, σ Is sigmoid function, and W_0 and W_1 are attention figures

The spatial attention formula is:

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)]))$$
$$= \sigma(f^{7\times7}([F_{avg}^s; F_{max}^s])). \tag{8}$$

Figure 2
Structure of channel attention and spatial attention module

Among them, σ It is a sigmoid function, and is 7×7 size convolution kernel.

By combining the channel attention module with the spatial attention module, the channel attention map output by the channel attention is input into the spatial attention module. The final output point cloud data can better have local geometric features and make up for the lack of a single edge volume.

4. Experiment

4.1. Dataset

In order to validate the effect of model classification, this paper performs a classification model of 3D point cloud object classification on the public datasets ModelNet40. The ModelNet40 datasets contains 12311 grid CAD models in 40 different categories. Of these, 9843 models were used for training and 2468 models were used for testing. Only the (x, y, z) coordinates of the sample point are used and the original grid is discarded. During training, the data is increased by randomly scaling the object and disturbing the location objects and points. The form of input point cloud is n × 3, and then the input from the self-attention module. In the self-attention mechanism module, it consists of four self-attention mechanism layers, each channel is 128, and finally fully connected to output. Then the form of n × 1024 points are input into the classification network. The classification network is composed of three channels of 64, 64 and 128 edge convolutions. After each edge convolution, three identical channels and spatial attention modules are embedded to make up for the deficiency of feature extraction. Finally, the point cloud is output to 1024 points by MLP, maximum pooling layer and full connection operations and the final classification is achieved by MLP with 512,265, c channels.

Details of experimental equipment: In order to verify the effect of the proposed 3D point cloud classification in the fusion of multiple attention mechanisms and DGCNN, the simulation experiment uses ModelNet40 dataset for training and testing in Ubuntu 20.04, Python 3.7, Python 1.12.1, CUDA11.4 environments. The hardware parameters of the computer and configuration environment used are processor Inter (R) Xeon (R) W-1270P CPU @3.80G Hz, RAM is 64G. NVIDIA GeForce RTX 3090 video card chip has a memory capacity of 24G.

4.2. Experiment Setup

In this paper, 1024 points are used for training and testing. The model achieves the best accuracy on the test set at the K=25, after 250 rounds of training. The evaluation indexes used in this paper are overall accuracy (OA) and average classification accuracy (MA) to evaluate the classification effect of point cloud.

4.3. Model Comparison

Table 1 lists the accuracy of the proposed method on the ModelNet40 and ModelNet10 data set with previous mainstream methods and classification models in late years. Compared with the previous PointNet model, the improved method in this paper improves the classification accuracy by 4.33%. Compared with DGCNN model, the accuracy of the proposed method is improved by 0.5%, and the average accuracy is improved by 0.9%. Compared with various classic and existing methods, this paper also has a lead of at least 0.4%. Through experimental analysis, in point cloud

Table 1

Accuracy comparison between this model and other models on ModelNet10

Algorithm	Input	ModelNet10 MA (%)	ModelNet10 OA (%)
DGCNN	1024	93.6	93.7
DGCNN	2048	*	*
PointNet	1024	91.5	91.8
PointGCN Global pooling	1024	91.4	91.8
PointGCN Mutil-resolution pooling	1024	91.6	91.9
SpiderCNN	1024	*	*
Multi-Scale Dynamic Graph Convolution	1024	*	*
RGCNN	1024	*	*
PointConv	1024	*	*
SO-Net	2048	*	*
Ours	1024	93.9	94.1
Ours	2048	*	*

Table 2

Accuracy comparison between this model and other models on ModelNet40

| Algorithm	Input	ModelNet40 MA (%)	ModelNet40 OA (%)
DGCNN	1024	89.06	92.2
DGCNN	2048	89.2	92.3
PointNet	1024	85.35	88.37
PointGCN Global pooling	1024	86.19	89.27
PointGCN Mutil-resolution pooling	1024	86.05	89.51
SpiderCNN	1024	*	92.04
Multi-Scale Dynamic Graph Convolution	1024	88.1	92.2
RGCNN	1024	*	90.5
PointConv	1024	*	92.5
SO-Net	2048	87.3	90.9
Ours	1024	89.9	92.7
Ours	2048	89.5	92.6

processing stage, the self-attention mechanism is used to replace the spatial transform. When the basic point cloud processing is completed, the point cloud data is further filtered to remove irrelevant points to reduce the impact on classification tasks. At the same time, on this basis, the spatial and channel attention mechanisms are fused to complete and enhance the features extracted by edge convolution, so as to minimize the lack of classification accuracy caused by the problem of missing local geometric information. This shows that the improvement points proposed for DGCNN have improved effect. In addition, Figures 3-5 show the training curves of the model in different data sets and different points.

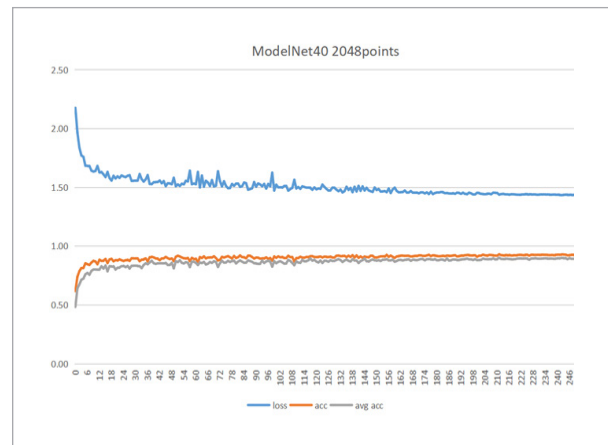## 4.4. Relationship between Model Classification Accuracy and K Value

Due to the introduction of multiple attention mechanisms based on DGCNN to calculate the attention of point clouds, more points are needed to partici-
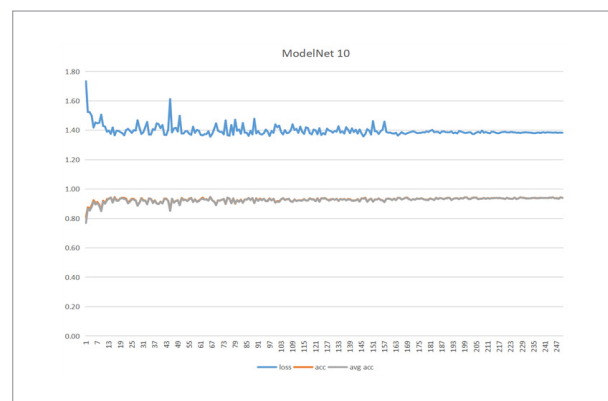
Figure 3

Training curve of 1024 points on ModelNet40
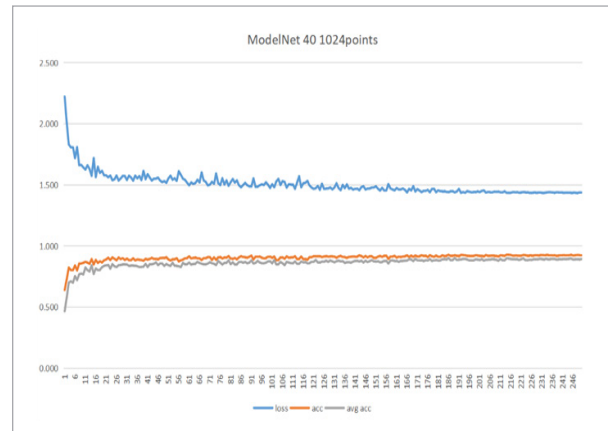
Figure 4

Training curve of 2048 points on ModelNet40

Figure 5

Training curve on ModelNet10

pate in the process of calculating attention scores. The classification task of DGCNN is the best when K = 20. However, due to the small number of nearest neighbor points selected, the original dynamic graph receiving domain is no longer suitable for the existing attention score calculation. Therefore, it is necessary to re-determine the K value. Therefore, this paper designs an experiment on K value. Based on the original K = 20, the K value is explored. Through the enumeration method, from 10 to 50, each interval is between 5 and 10, and the best K value is obtained through experiments.

Table 2 lists the relationship between different K values and classification accuracy. Figure 6 shows the relationship between K values and classification accuracy. This paper finds that the model has the best classification effect when K=25, so the K value is select as 25. Through the analysis of experimental results, the relationship between K value and accuracy is roughly normal distribution. When the K value is from 10 to 25, the accuracy continues to rise until the highest point. As the K value increases, the accuracy rate decreases.

Table 3

Relationship between model classification accuracy and K value

| KNN point numbers	MA(%)	OA(%)
10	89.2	92.3
15	89.3	92.1
20	89.3	92.4
25	89.9	92.7
30	89.0	92.1
35	89.1	92.2
40	89.6	92.3
50	88.8	92.1

4.5. Ablation Experiment

The effectiveness of the modules proposed is verified by ablation experiment. The ablation experiment is based on the point cloud shape classification task.

1 spatial attention mechanism: Remove the spatial attention mechanism module from the model separately, and the other modules remain unchanged. The experiment shows that the accuracy of the

Figure 6

Relationship between different K values and classification accuracy
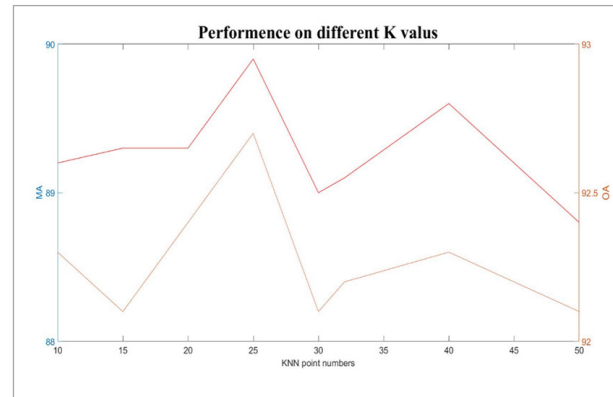
model decreases by 0.89%, which indicates that the spatial attention module can process the data in the point cloud space well and improve the accuracy of extracted feature.

2 Channel attention mechanism: Remove the channel attention mechanism from the model independently, and the other modules remain unchanged. The experiment shows that the accuracy of the model decreases by 0.4%, which shows that the channel attention mechanism can well enhance the feature processing of the related channels and further improve the accuracy of the model.

3 (3) Space and channel attention fusion module: Remove the attention fusion module from the model, while the other modules remain unchanged. The experiment results show that the accuracy of the model is decreased by 0.7%, which indicates that the attention fusion module can make up for the shortage of point cloud features extracted by convolution.

4 Self-attention mechanism: Remove the self-attention mechanism from the model, and the other modules remain unchanged. The experiment shows that the precision of the model decreases by 0.8%, which indicates that the self-attention mechanism can well replace spatial transform module to process point cloud.

5 Delete the entire attention module, leaving only the original classification network. The experiment shows that the accuracy of the model decreases by 1% without adding any attention mechanism module, which shows that the improvement on the original classification network is effective.

Table 4

Ablation experiment

Validation section	MA(%)	OA(%)
No self-attention	88.76	92.00
No spatial attention	89.00	91.82
No channel attention	89.31	92.40
No fusion module	89.20	92.00
None	88.30	91.80

4.6. Data Corruption Experiment

Based on the ModelNet40 dataset, the effect of dataset corruption [13] on ablation experiments was carried out. When the data set was deleted locally, the points in the data set were deleted according to the 10 % and 20 % criteria and the main ablation experiments were carried out. We performed ablation experiments on the main modules in the case of 10 % data loss, and compared the complete models in the case of 10 % and 20 % data loss. The experimental results show that the model has a gentle reduction in accuracy and good robustness in the face of corrupt data sets.

Table 5

Ablation experiment of corruption data set

Data set	Validation section	MA(%)	OA(%)
10 % dataset damaged	No self-attention	89.2	92.1
10 % dataset damaged	No fusion module	89.7	92.4
10 % dataset damaged	All	89.1	92.0
20 % dataset damaged	All	88.5	91.8
complete data set	All	89.9	92.7

4.7. Evaluating Indicator

We also list the evaluation indicators of the model. Although our model has a larger number of parameters and a larger model size than other models, it also indirectly reflects that our model also has a larger capacity, can accommodate more feature information, and has stronger fitting ability and expression ability. It also has better performance in the future deployment phase.

Table 6

Parameters and size of the model

Algorithm	Params(M)	Model Size(M)
DGCNN	3.3	80.5
PointNet	3.5	41.8
PointGCN	4.3	189
SpiderCNN	0.8	*
Multi-Scale Dynamic Graph Convolution	114.7	*
RGCNN	1.6	*
PointConv	2.7	0.3
SO-Net	2.5	*
Ours	7.27	87.7

5. Summary and Outlook

We propose a 3D point cloud classification that combines multiple attention mechanism and dynamic graph convolution, and introduce self-attention mechanism, spatial attention and channel attention mechanisms. The self-attention module replaces the spatial transform module of the original classification model, which reduces the processing of unrelated points while performing point cloud coordinate transformation and reduces the amount of calculation. Space and channel attention are fused to make up for the lack of local geometric information caused by edge convolution. The improved model is applied to ModelNet40 3D point cloud classification to effectively classify point cloud data. The experimental results indicate that the average accuracy rate is improved by 0.9% and the overall accuracy rate is effectively improved by 0.5% compared with DGCNN. Compared with other methods, the accuracy and average accuracy of the model are the highest, which fully proves the effectiveness of integrating multiple attention mechanism methods to classify point cloud. After repeated experiments, this paper gets some inspiration: when processing point cloud data, delete irrelevant points as much as possible to reduce the impact on the overall data and reduce the amount of data processing computation. In the process of classification, we try to make up for features, complete and enhance local information to achieve the purpose of improving the classification accuracy. This paper is

only about the classification of 3D point clouds, and the segmentation and semantic segmentation of specific scenes have not been studied. Later, we can try to apply this model to the automatic driving scene.

Declarations

There is no conflict of interest in this paper, and the data set used in the experiment is public and no confidential.

References

1. Arshad, S., Shahzad, M., Riaz, Q., Fraz, M. M. DPR-Net: Deep 3D Point Based Residual Network for Semantic Segmentation and Classification of 3D Point Clouds. IEEE Access, 2019, 7, 68892-68904. https://doi.org/10.1109/ACCESS.2019.2918862

2. Gao, J. et al. Deep Learning Model Based on Improved Point Cloud Transformer. Journal of North University of China (Nature Science Edition), 2021, 42(6), 9.

3. Guo, M.-H., Cai, J. X., Liu, Z. N., Mu, T. J., Martin, R. R., Hu, S. M. PCT: Point Cloud Transformer. Computational Visual Media, 2021, 7, 187-199. https://doi.org/10.1007/s41095-021-0229-5

4. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M. Deep Learning for 3D Point Clouds: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43, 4338-4364. https://doi.org/10.1109/TPAMI.2020.3005434

5. Li, C., Zhang, X., Yan, F., Zhuang, Y. Deep Place Recognition for 3D LiDAR Point Cloud with Transformer. China Automation Congress (CAC), 2021, 5640-5645. https://doi.org/10.1109/CAC53003.2021.9728443

6. Li, J., Chen, B. M., Lee, G. H. SO-Net: Self-Organizing Network for Point Cloud Analysis. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, 9397-9406. https://doi.org/10.1109/CVPR.2018.00979

7. Li, S., Zhou, X., Li, X., Zhang, X. Instance Segmentation of Point Cloud Based on Improved DGCNN for Robotic Grasping. 7th International Conference on Automation, Control and Robotics Engineering (CACRE), 2022, 328-332. https://doi.org/10.1109/CACRE54574.2022.9834115

8. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B. PointCNN: Convolution on X-Transformed Points. Advances in Neural Information Processing Systems, 2018, 31.

9. Maturana, D., Scherer, S. A. VoxNet: A 3D Convolutional Neural Network for Real-time Object Recognition. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, 922-928. https://doi.org/10.1109/IROS.2015.7353481

10. Mazur, K., Lempitsky, V. S. Cloud Transformers: A Universal Approach to Point Cloud Processing Tasks. IEEE/CVF International Conference on Computer Vision (ICCV), 2021, 10695-10704. https://doi.org/10.1109/ICCV48922.2021.01054

11. Qi, C. R., Su, H., Mo, K., Guibas, L. J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 77-85.

12. Qi, C. R., 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.

13. Ren, J., Pan, L., Liu, Z. Benchmarking and Analyzing Point Cloud Classification Under Corruptions. In International Conference on Machine Learning, 2022, 18559-18575.

14. Shajahan, D. A., Varma, M., Muthuganapathy, R. Point Transformer for Shape Classification and Retrieval of Urban Roof Point Clouds. IEEE Geoscience and Remote Sensing Letters, 2022, 19, 1-5. https://doi.org/10.1109/LGRS.2021.3061422

15. Simonovsky, M., Komodakis, N. Dynamic Edge-Conditioned Filters in Convolutional Neural Networks on Graphs. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 29-38. https://doi.org/10.1109/CVPR.2017.11

16. Sun, L., Jin, Z., Wenju, W., Yusen, Z. Multi-Scale Transformer Lidar Point Cloud 3D Object Detection. Computer Engineering and Application, 2020, 58(8), 136-146.

17. Tao, Z., Zhu, Y., Wei, T., Lin, S. Multi-Head Attentional Point Cloud Classification and Segmentation Using Strictly Rotation-Invariant Representations. IEEE Access, 2021, 9, 71133-71144. https://doi.org/10.1109/ACCESS.2021.3079295

18. Te, G., Hu, W., Zheng, A., Guo, Z. RGCNN: Regularized Graph CNN for Point Cloud Segmentation. Proceedings

of the 26th ACM International Conference on Multimedia, 2018, 1-10. https://doi.org/10.1145/3240508.3240621

19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. Attention is All You Need. ArXiv abs/1706.03762, 2017, 1-15.

20. Wang, D., Yang, Z. InfoPCT: Mutual Information Maximization based Point Cloud Transformer. IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2022, 17-21. https://doi.org/10.1109/IPEC54454.2022.9777437

21. Wang, L., Huang, Y., Hou, Y., Zhang, S., Shan, J. Graph Attention Convolution for Point Cloud Semantic Segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, 10288-10297. https://doi.org/10.1109/CVPR.2019.01054

22. Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., Solomon, J. M. Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics (TOG), 2019, 38, 1-12. https://doi.org/10.1145/3326362

23. Wen, X., Xiang, P., Han, Z., Cao, Y. P., Wan, P., Zheng, W., Liu, Y. S. PMP-Net++: Point Cloud Completion by Transformer-Enhanced Multi-step Point Moving Paths. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 1-16. https://doi.org/10.1109/TPAMI.2022.3159003

24. Woo, S., Park, J., Lee, J. Y., Kweon, I. S. CBAM: Convolutional Block Attention Module. Proceedings of the European Conference on Computer Vision (ECCV), 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

25. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J. 3D ShapeNets: A Deep Representation for Volumetric Shapes. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 1912-1920.

26. Yin, J., Shen, J., Gao, X., Crandall, D., Yang, R. Graph Neural Network and Spatiotemporal Transformer Attention for 3D Video Object Detection from Point Clouds. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 1-1.

27. Zhai, Z., Zhang, X., Yao, L. Multi-Scale Dynamic Graph Convolution Network for Point Clouds Classification. IEEE Access, 2020, 8, 65591-65598. https://doi.org/10.1109/ACCESS.2020.2985279

28. Zhang, K., Hao, M., Wang, J., Chen, X., Leng, Y., de Silva, C. W., Fu, C Linked Dynamic Graph CNN: Learning through Point Cloud by Linking Hierarchical Features. 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), 2021, 7-12. https://doi.org/10.1109/M2VIP49856.2021.9665104

29. Zhang, Y., Rabbat, M. G. A Graph-CNN for 3D Point Cloud Classification. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, 6279-6283. https://doi.org/10.1109/ICASSP.2018.8462291

30. Zhen, N., Li, Y., Wang, Y., Xiao, Z., Zhao, N., Li, Y. 3D Point Cloud Target Detection Based on Attention Mechanism and PointRCNN. Electronic Measurement Technology, 2022, 045-009.

31. Zhou, Z., Zhao, C., Adolfsson, D., Su, S., Gao, Y., Duckett, T., Sun, L. NDT-Transformer: Large-Scale 3D Point Cloud Localisation using the Normal Distribution Transform Representation. IEEE International Conference on Robotics and Automation (ICRA), 2021, 5654-5660. https://doi.org/10.1109/ICRA48506.2021.9560932