Hindawi
International Journal of Antennas and Propagation
Volume 2022, Article ID 5417440, 11 pages
https://doi.org/10.1155/2022/5417440

Research Article

DANC-Net: Dual-Attention and Negative Constraint Network for Point Cloud Classification

Hang Sun,[1,2] **Yuanyue Zhang**,[1] **Jinmei Shi**,[3] **Shuifa Sun**,[1] **Guanqun Sheng**,[1] **and Yirong Wu**[1]

[1]*College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China*
[2]*Hubei Engineering Technology Research Center for Farmland Environment Monitoring, China Three Gorges University, Yichang 443002, Hubei, China*
[3]*College of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, Hainan 571158, China*

Correspondence should be addressed to Shuifa Sun; watersun@ctgu.edu.cn

Received 17 January 2022; Revised 30 March 2022; Accepted 25 April 2022; Published 28 May 2022

Academic Editor: Hervé Aubert

Convolutional neural networks, as a branch of deep neural networks, have been widely used in multidimensional signal processing, especially in point cloud signal processing. Nevertheless, in point cloud signal processing, most point cloud classification networks currently do not consider local feature correlation. In addition, they only adopt ground-truth as positive information to guide the training of networks while ignoring negative information. Therefore, this paper proposes a network model to classify point cloud signals based on feature correlation and negative constraint, DANC-Net (dual-attention and negative constraint on point cloud classification). In the DANC-Net, the dual-attention mechanism is utilized to strengthen the interaction between local features of point cloud signal from both channel and space, thereby improving the expression ability of extracted features. Moreover, during the training of the DANC-Net, the negative constraint loss function ensures that the features in the same categories are close and those in the different categories are far away from each other in the representation space, so as to improve the feature extraction capability of the network. Experiments demonstrate that the DANC-Net achieves better classification performance than the existing point cloud classification algorithms on synthetic datasets ModelNet10 and ModelNet40 and real-scene dataset ScanObjectNN. The code is released at https://github.com/sunhang1986/DANC-Net.

1. Introduction

Signal processing is usually understood as the processing of electronic signals [1–5]. Point cloud processing can be described as the processing of point cloud, a kind of multidimensional signal. However, the classification task of point cloud is still facing enormous challenges due to its unordered and sparse characteristics.

3D objects can be represented in two ways according to the spatial distribution of the 3D point cloud. (1) regular structure representation, which is represented by multi-view and voxel representation, and (2) irregular and unstructured representation, which is represented by point cloud and grid representation. Point cloud processing methods based on regular structured representations include 3D volumetric convolutional neural networks (CNNs) [6–8] and the multi-view CNN [9, 10]. These methods transform irregular/unstructured point clouds to regular/structured images (or volume grids), and use two-dimensional (2D) CNNs to extract local features and global features of the point cloud. Although these methods solve the unordered distribution issues of point clouds, they bring a lot of challenges in calculation and issues in memory consumption. Octree-based method [11] alleviates these problems to a certain extent and can apply 3D CNN to higher resolution grid. Le and Duan [12] and Hua et al. [13] studied different 3D convolution operators based on grid cells, which can better learn local features. On the contrary, methods based on irregular unstructured representation do not need to transform the representation of point cloud. They can learn point cloud features using special CNNs designed for raw point cloud data [14–16]. Because of low memory consumption and simple structure of this type of representation, point cloud classification methods based on irregular unstructured representation have attracted more and more attention from researchers.

In the study of point cloud classification based on irregular and unstructured representations, Qi et al. [14] designed a PointNet network capable of point-by-point coding in order to use deep learning to process point cloud data. However, the details are lost because the whole PointNet network does not divide the point cloud regions and extract the region features. PointNet++ [15], which is based on PointNet, adopts a hierarchical structure that allows repeated capture of local information. Therefore, the overall accuracy (OA) of PointNet++ in ModelNet40 dataset is greatly improved compared with the OA of PointNet, which effectively demonstrates the importance of local information. However, because the processes of extracting local features are mutually independent, information is not exchanged between subclouds, resulting in a loss of structural information. Since then, in order to simplify the training process and save computing resources, a large number of researchers have proposed methods based on CNNs, such as PointCNN [16], tangent convolutions [17], and point cloud classification networks [18], which strengthen the geometric structure acquisition of point cloud data. However, these methods do not consider the effects of local structure relationship that are essential in 3D object recognition.

In summary, how to efficiently learn in-depth local features and their relationship from point cloud has become a pressing problem. In addition, most of the existing point cloud classification networks only use positive information to guide the training of network, lacking of effective use of negative information, which limits the network capability to extract more distinguishing features for point cloud classification.

In order to efficiently learn the correlation between local features of point cloud signals and utilize the negative information which is crucial to the classification results, we propose an effective point cloud classification network. Our point cloud classification network, based on a dual-attention mechanism and contrastive learning constraints, is named DANC-Net. The main components of the network are the channel attention and self-attention (CASA) module and the negative constraint loss function (NC-loss). The CASA module is used before the global features are aggregated. Channel attention and self-attention are used to capture the relationship between local features. In NC-loss, the output point cloud features with local feature relationships are divided into the output feature, positive sample features, and negative sample features. The output feature is constrained by negative information, in order to be approach-positive sample features and stay away from negative sample features. Positive information and negative information are used effectively at the same time, which improves the classification ability of our DANC-Net.

To sum up, our contributions are three-folds as follows:

(1) We propose a dual-attention module, CASA. It can strengthen the extraction of local feature correlation from channel and spatial, thereby helping the network to further develop the geometric structure between points.

(2) We propose a negative constraint loss function, NC-loss. Besides the positive information constraints, the effective constraint of negative information has also been strengthened; thus, the ability of the network to extract more distinctive features is improved.

(3) We propose a dual-attention negative constraint network, DANC-Net, which achieves superior performance compared with the recently proposed point cloud classification methods on open datasets ModelNet10 [8] and ModelNet40 [8] and the real-scene dataset ScanObjectNN [19].

2. Related Work

In recent years, deep learning continues to make breakthroughs in computer vision [20–23]. The early point cloud classification methods based on deep learning transform point cloud to regular volume grids and then extract features from the point cloud by using 3D CNNs [6, 8]. However, 3D CNN takes up more computing resources than 2D CNN. To make computation affordable, the volume grids are usually in low resolution, resulting in the loss of geometric information of 3D mesh shape, especially when dealing with large-scale point cloud. Therefore, the 3D point cloud is mapped to the 2D space, and then, the 2D image CNNs are used to classify [7, 10]. With well-engineered image CNNs, these methods have achieved the expected performance. Nevertheless, the selection of projection angle and projection plane has a significant impact on the classification accuracy, so the generalization ability of these models is poor.

PointNet [14], a kind of end-to-end network, is the first method to deal with point cloud directly based on deep learning. The method takes N points as input and uses a 3×3 affine transformation matrix (T-Net) to realize input alignment and feature alignment. The aligned point cloud learns global feature vectors through multiple three-layer perceptrons (MLPs) and max pooling, and finally realizes end-to-end point cloud classification. However, vital local information is ignored in the PointNet. PointNet++ [15] proposed by Qi et al. is a point cloud classification network based on PointNet. It refers to the feature extraction method of PointNet to process each group of point clouds independently. Then, the global features are aggregated using max pooling. The hierarchical structure of PointNet++ exploits local information to a certain extent. In PointNet++, multi-scale algorithm is used to group point clouds. In the process of grouping, it is inevitable that there will be repeated grouping points, which will result in local information redundancy and reduce the classification ability of the network. For the purpose of reducing the redundancy of local information, the authors of A-CNN [24] proposed the constraint-based k-nearest neighbor (k-NN) algorithm and annularly convolution on the basis of hierarchical structure. As shown in Figure 1, the input point cloud is sampled and the constraint-based k-NN algorithm is used to construct groups in each layer of the network. Then, the features within each group are extracted by combining annular convolution with max pooling. Compared with multi-scale

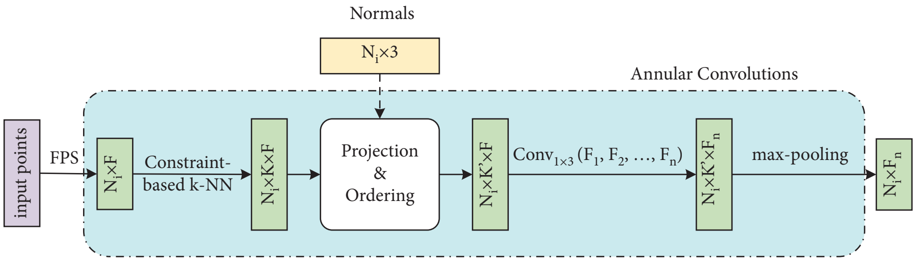

Figure 1: A-CNN abstract layer structure.

grouping, rings of annularly convolution do not contain duplicate points, which allows the network to learn more discriminant features. Therefore, A-CNN achieved a higher classification performance than PointNet++ on the ModelNet dataset. The progress of the above end-to-end point cloud classification methods is undeniable, However, their approach of extracting local features independently leads to inadequate identification of correlations between points or local neighborhoods.

Recently, attention mechanism [25] has achieved remarkable achievements in natural language processing, image recognition [26], and other fields. In point cloud classification, Bhattacharyya et al. [27] proposed an altitude attention model, which can achieve superior classification performance of airborne laser scanning (ALS) by considering the altitude information of points. Lee et al. [28] proposed a simple and efficient network based on self-attention, called set transformer, which can process set data, such as a point cloud. Shajaha et al. [29] proposed a multi-view CNN with self-attention. Multiple views of a roof point cloud were taken as the input, an adaptive weight learning algorithm was used to assign weights corresponding to each view, and the category of the roof was the output. However, the generalization ability of the model [29] is poor and is limited to special field. On the contrary, the DANC-Net we proposed can be applied to any point cloud classification tasks.

Currently, most of the point cloud classification networks only use ground-truth as positive information to guide the training of the network while negative information is ignored, which leads to the limitation of network discrimination capabilities. Therefore, in order to further explore the correlation of local features of point cloud and the constraints of negative information on features, this paper proposes the DANC-Net based on dual-attention mechanism and negative information constraints.

3. Method

This section details the proposed DANC-Net in this paper. First, the architecture of DANC-Net point cloud classification method is introduced in Section 3.1. Then, Section 3.2 performs detailed analysis of dual-attention CASA module for capturing correlations between local features. Next, Section 3.3 presents the loss function NC-loss under the negative information constraint. Finally, Section 3.4 summarizes the total loss function of the DANC-Net.

3.1. DANC-Net Architecture. For a clear understanding of our DANC-Net, we show the network architecture and the output feature map size of each layer in the network in Figure 2. Our DANC-Net consists of five layers.

(1) Input layer: for a given 3D shape point cloud, the coordinates and normals of N points are used as input.

Feature map size: each input consists of a 3D coordinates (x, y, z) and a normal, i.e., two $N\times3$-dimensional tensors.

(2) A-CNN layer: local features are extracted from point cloud. This layer performs two feature extractions, and each feature extraction includes two operations, namely, the farthest point sampling (FPS) algorithm [30] and the A-CNN abstraction layer.

Feature map size: after two feature extractions, the previous-level output point cloud is divided into N_1 and N_2 local regions, and the number of channels for each local region feature is 128 and 256, respectively.

(3) CASA layer: a new feature map $\boldsymbol{f_r}$ with geometric relationship and positional relationship among local features of point cloud is obtained, and then, the global feature vector $\mathbf{f}_g$ is aggregated through the PointNet [14] layer.

Feature map size: the point cloud global feature vector $\mathbf{f}_g$ with correlation between local feature output by this layer has 1024 channels.

(4) Negative information constraint layer: the $\boldsymbol{f_r}$ is used to construct the NC-loss under the constraint of negative information, so as to restrain the mutual interference between similar categories (such as nightstands and dressing tables with similar spatial structure).

Feature map size: the feature map $\boldsymbol{f_r}$ that is fed into the loss function has 256 channels.

(5) Output layer: MLPs are used to obtain the probability score of the point cloud belonging to the c category.

Feature map size: the final output vector size of our DANC-Net is $1\times c$.

3.2. Dual-Attention (CASA) Unit. Most point cloud classification networks only enhance the expression ability of the network from the perspective of enhancing local feature

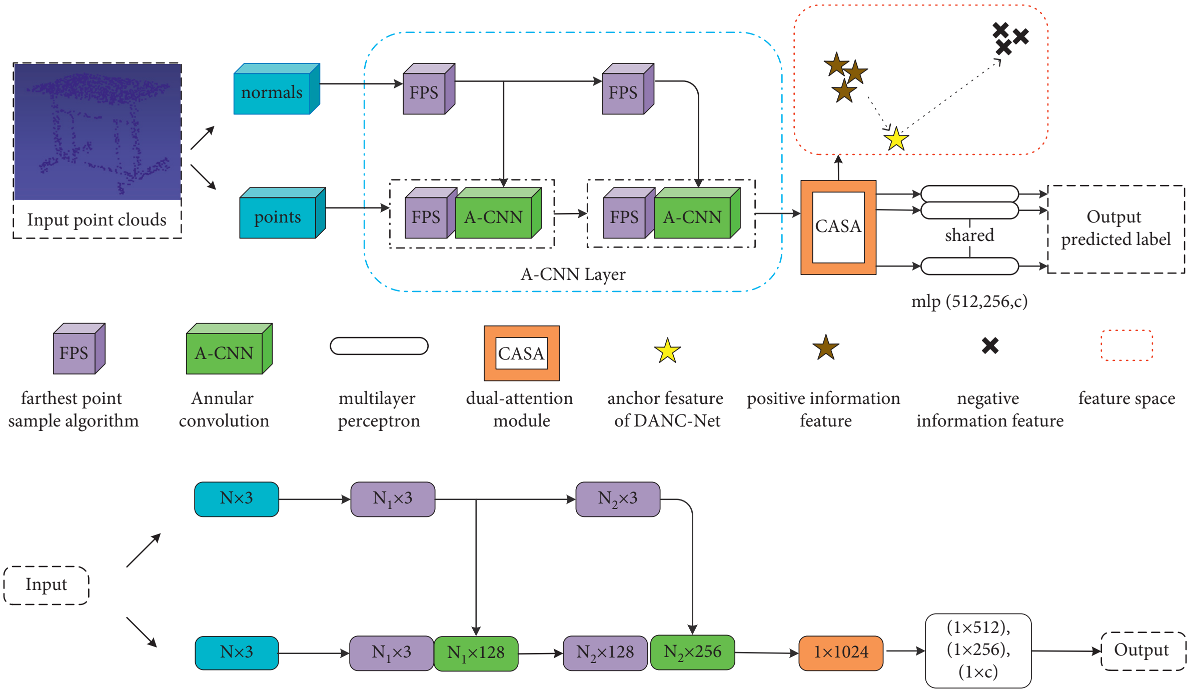

Figure 2: The architecture of DANC-Net (top) and the feature map size of each node of the model (bottom). The blue dashed box on the left of the network structure is the local feature extraction layer (A-CNN layer).

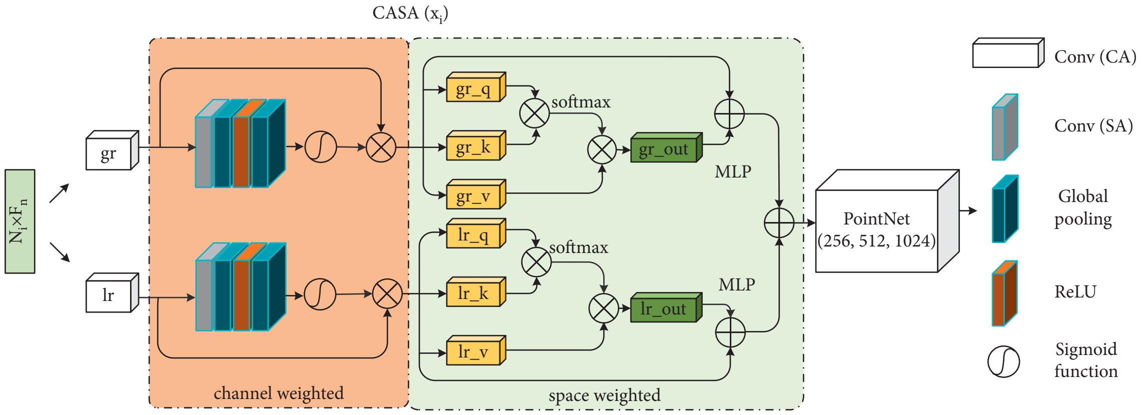

Figure 3: CASA module. ***gr*** and ***lr***, respectively, represent the geometric and positional features of the sub-cloud. $\boldsymbol{x_i}$ represents the ith input feature of our CASA module.

extraction, while ignoring the exchange of information between local features. CASA unit can adaptively learn feature weights and capture the correlation between local features. As shown in Figure 3, the CASA module consists of two branches, each consisting of channel attention [31] and self-attention [32]. Among them, the upper branch is used to extract the geometric feature relationship and the lower branch is used to extract the positional feature relationship. The CASA module considers that the features of different channels contain completely a different weighting information and point distribution is not uniform in different spatial positions. CASA treats different features and points unequally, providing extra flexibility in processing different types of information and expanding CNN's expressive ability.

For a set of points containing n points $P = \{P_i, \quad i = 1, \ldots, n\}$, A-CNN is used to extract the geometric feature vector $gr_i \in \mathbb{R}^d$ and location feature vector $lr_i \in \mathbb{R}^3$ of the local sub-cloud $\mathbf{G_i}$; the two feature vectors are then input into the CASA module. The output is a high-level global feature $\mathbf{f}_g$ that incorporates context information. The key to this process is how to generate different weights for each point feature. The detailed implementation process of the CASA module is analyzed as follows.

First, the input geometric information and location information are weighted by channel attention. In this process, we use global average pooling to transform the global information of the channel to the channel descriptors; that is, the channel dimension is kept unchanged, but the other dimensions of the feature map are reduced to 1. In order to obtain the weights $\boldsymbol{CA}$ of the channels, channel descriptors sequentially pass through convolution layer, ReLU activation function, convolution layer, and sigmoid function.

$$CA = S\left(Conv\left(L\left(Conv\left(g_c\left(F_c\right)\right)\right)\right)\right), \tag{1}$$

where $\boldsymbol{F_c}$ is the feature graph of the input, g_c is the global pooling function, and S and L represent the sigmoid function and ReLU function, respectively. Finally, the input feature map $\boldsymbol{F_c}$ and channel weight $\boldsymbol{CA}$ are multiplied and the weighted feature map is obtained as shown in

$$F_c^* = CA \otimes F_c. \tag{2}$$

Second, the feature map is spatially weighted. Three 1 × 1 traditional convolutions of the input feature map are performed, and the three feature matrixes $\boldsymbol{q}$, $\boldsymbol{k}$, and $\mathbf{v}$ are obtained. The correlation matrix $\boldsymbol{M}$ is obtained by multiplying matrixes $\boldsymbol{q}$ and $\boldsymbol{k}$. The softmax normalization operation is performed on the correlation matrix $\boldsymbol{M}$ to obtain the attention weight in the range [0, 1]. The weight coefficient is applied to the feature matrix $\mathbf{v}$, and the residual connection is made with the input feature map F_c^*, so that each local feature is weighted by all local features. The weighted feature map obtained is shown in

$$F^* = \operatorname{softmax}\left(q \otimes k^T\right) \otimes v + F_c^*. \tag{3}$$

Finally, the feature graph $\boldsymbol{F}^*$, which has been weighted by channel and space, is fused by the MLP, and local features $\boldsymbol{f_r}$ containing context correlation are obtained by matrix addition. All local regional features are then aggregated by PointNet to obtain the global feature $\mathbf{f}_g$. To demonstrate the correctness of the CASA module, we verify its point cloud classification effect in the ablation study (Section 4.3).

3.3. Negative Constraint Loss Function (NC-Loss) Unit.

Inspired by [33, 34], in order to further improve the discrimination ability of the point cloud classification network, a loss function (NC-loss) with negative information constraints is proposed in this paper. The point cloud with the same label as the output feature is called positive information, while the point cloud with different labels from the output feature is called negative information. As shown in Figure 2, the red dotted box represents the feature space of the constructed NC-loss. In the feature space, NC-loss can not only close the features of positive information and the output features, but also push the features of negative information and the output features farther.

In the proposed NC-loss, an output feature is selected from the local features $\boldsymbol{f_r}$ containing context correlation. In this paper, the features of all input point cloud samples are traversed to ensure that the feature of each point cloud sample has the opportunity to be selected as output features. In the feature space, the distances between point clouds of the same class are minimized and the distances between point clouds of different classes are maximized. Therefore, the regularization loss function of contrastive learning is defined in

$$L_{nc} = \sum_{i=1}^{B}\left[D\left(F_i, F_P\right) - D\left(F_i, F_N\right)\right], \tag{4}$$

where D (x, y) represents the L_1 distance between x and y. The number of input point cloud samples is B, and $i \in I \equiv \{1, \ldots, B\}$ represents the ith point cloud sample selected as the output feature. $P(i)$ is the set of positive point cloud samples, which contains all the point clouds in the B samples with the same label as the output feature. $N(i)$ is the set of negative point cloud sample, which contains all the point clouds in the B samples with the same label as the output feature. $F = \mathrm{MLP}\left(f_r\right) \in \mathbb{R}^{D_M}$ means that the local features $\boldsymbol{f_r}$ containing context correlation become the feature matrixes $\boldsymbol{F}$ after MLP mapping; i.e., $\boldsymbol{F_i}$, $\boldsymbol{F_P}$, and $\boldsymbol{F_N}$ represent the output feature matrixes, the positive sample feature matrixes, and negative information sample feature matrixes, respectively, where D_M is a constant 256.

3.4. DANC-Net Totally Loss Function.

For the task of point cloud classification, we use the cross-entropy loss function to measure the distance between the predicted values of 3D point cloud samples and the ground-truth. The calculation method of cross-entropy loss is as follows:

$$L(P) = -\sum_{k=1}^{c} \hat{y}_k \ln\left(y_k\right), \tag{5}$$

where $\hat{y}_k \in \{0, 1\}$ indicates the kth value in the label vector. $y_k \in [0, 1]$ indicates the probability that the prediction sample P belongs to the kth class.

Therefore, the finally loss function L of our DANC-Net consists of a classification loss function and a negative constraint loss function (NC-loss). The final loss function L can be expressed as follows:

$$L = L(p) + \lambda L_{nc} = L(p) + \lambda \sum_{i=1}^{B} D\left(F_i, F_P\right) - D\left(F_i, F_N\right). \tag{6}$$

In formula (6), λ is the penalty parameter used to balance the classification loss and NC-loss. The ablation experiment results show that the classification accuracy of the DANC-Net based on the dual-attention CASA module can be further improved by using NC-loss.

4. Experiments and Result Analysis

We used three benchmark datasets, ModelNet10, ModelNet40, and ScanObjectNN, to compare our DANC-Net with the state-of-the-art point cloud classification algorithms [6, 8, 14–16, 24, 35–40]. The synthetic datasets ModelNet10 and ModelNet40 are subsets of ModelNet (a large 3D CAD model dataset). Each point cloud sampled from the grid contains 10,000 points and normal vectors, and the coordinates are normalized to unit spheres. ScanObjectNN is a real-world dataset of point cloud objects, constructed from indoor scene scanning. More details about three benchmark datasets can be found in Table 1.

Table 1: Distribution of training and test sets.

Datasets	Class number	Training models	Testing models	Total models
ModelNet10	10	3991	908	4899
ModelNet40	40	9843	2468	12331
ScanObjectNN	15	80% of the total models were randomly selected	20% of the total models were randomly selected	2902

In robustness test, 80% and 20% of the total models were randomly selected as the training set and test set. In the experiment using ModelNet10 and ModelNet40, 1024 points with normals were sampled, and the normals were used in the ordering algorithm in A-CNN.

The hardware environment of the experiments included an RTX 2080 Ti graphics card, 12 GB video memory, Ubuntu 18.04 operating system, and CUDA 10.1 + cuDNN 7.6.5 + TensorFlow 1.3.0 + Python 3.6. In Tables 2–4, the highest, second highest, and third highest classification accuracies are indicated by bold, underline, and italic text, respectively.

4.1. Parameters. For experiments on three benchmark datasets, 1024 points from 3D meshes are sampled randomly as the input of the DANC-Net. The data augmentation method was the same as that of A-CNN. The loss included classification loss and comparison loss, as defined in (6), and the classification loss used the cross-entropy loss function. Using the Adam optimizer, the initial learning rate was set to 0.001 and attenuated at a decay rate of 0.7 per 200,000 steps. The classification model was trained for 250 epochs with a batch_size of 16. In the experiment, the penalty parameter was set to 1.0.

4.2. Comparison Experiments and Analysis. We demonstrate the effectiveness of our DANC-Net on ModelNet40 and ModelNet10 datasets though comparison experiments. As shown in Tables 2 and 3, using the datasets ModelNet10 and ModelNet40, our DANC-Net was compared with the state-of-the-art point cloud classification methods based on deep learning. The quantitative evaluation of the classification performance of models in experiments adopts the commonly used evaluation metrics for point cloud classification: mean per-class accuracy (mA) and overall accuracy (OA). mA and OA are defined by

$$mA = \sum_{i=1}^{c} \frac{\mathrm{num}\,(TP)_i}{\mathrm{num}_i} \times \frac{1}{c}. \tag{7}$$

$$OA = \frac{\sum_{i=1}^{c} \mathrm{num}\,(TP)_i}{T}. \tag{8}$$

where $\mathrm{num}(TP)_i$ represents the number of 3D meshed shapes correctly classified into category i; num_i represents the number of 3D meshed shapes that belong to category i; and T represents the total number of 3D meshed shapes to be predicted. In Tables 2 and 3, the top three mA and OA are highlighted by bold, underline, and italic text, respectively. Furthermore, the classification results of the methods used for comparison are obtained from corresponding papers. If the data and results are not given in the paper, we will download the codes of the models. The classification results are obtained by training and testing the models in the corresponding experimental environment. Moreover, "-" means that the dataset cannot be used by the methods or the codes are not provided in the paper.

4.2.1. Comparison Experiments on ModelNet10 Dataset. The DANC-Net achieves the best classification performance with 95.5 OA and 95.4 mA on ModelNet10 (Table 2).

(1) Comparison with the deep learning point cloud classification methods based on voxel grid.

In Table 2, we compare DANC-Net with 3DShapeNets and VoxNet methods based on voxel grid on ModelNet10. We observe that our DANC-Net achieves 12.0 OA and 3.5 OA more than 3DShapeNets and VoxNet. For 3DShapeNets and VoxNet, the classification performance is low because the invariance of point cloud cannot be maintained when point cloud is converted to 3D voxel grids. For our DANC-Net, conversion of point cloud to other forms is not required, and the invariance of point cloud is retained, achieving high classification accuracy.

(2) Comparison with the deep learning classification methods based on point cloud.

As shown in Table 2, we can observe that (1) compared with the classical PointNet and PointNet++, mA of our DANC-Net is increased by 1.2 and 0.7, and OA is increased by 1.1 and 0.6, respectively, on ModelNet10. (2) Compared with Kd-Net [35] and DGCNN [36], mA is increased by 1.9 and 0.6, and OA is improved by 1.5 and 0.6, respectively. (3) Compared with A-CNN, our DANC-Net has achieved an improvement of 1.0 mA and 0.9 OA on ModelNet10. Our DANC-Net achieves high performance because the CASA module can adjust both spatial weights of features and channel weights of features. Furthermore, our DANC-Net adds negative information constraint, so the classification accuracy is higher than that of other methods.

4.2.2. Contrast Experiments on ModelNet40 Dataset. The DANC-Net achieved the highest classification accuracy with 92.9 OA and 90.5 mA on ModelNet40 (Table 3).

(1) Comparison with the deep learning point cloud classification methods based on voxel grid.

Table 2: Classification performance on ModelNet10. (The top three accuracies are highlighted by bold, underline, and italic.)

Methods	Input	Points (k)	mA (%)	OA (%)
VoxNet (IROS 2015)	Points	1	—	92.0
3DShapeNets (CVPR 2016)	Points	1	—	83.5
PointNet (CVPR 2017)	Points	1	94.2	94.4
PointNet++ (CVPR 2017)	Points + normal	5	*94.7*	94.9
Kd-Net (ICCV 2017)	Points	32	93.5	94.0
DGCNN (TOG 2019)	Points	1	94.8	94.9
A-CNN (CVPR 2019)	Points + normal	1	94.4	*94.6*
Ours	Points + normal	1	**95.4**	**95.5**

Input and points represent the input data type and the number of sampling points, respectively.

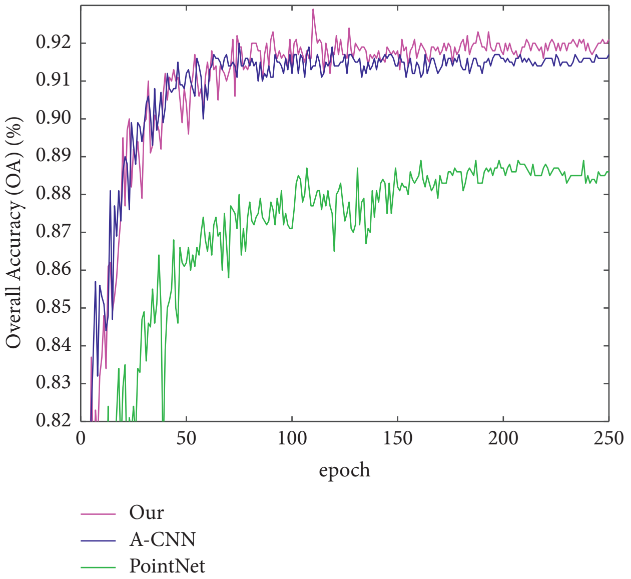

Figure 4: Variation of the overall accuracy (OA) in % on ModelNet40 with 250 epochs.

As shown in Table 3, our DANC-Net achieves 13.2 mA and 7.5 mA more than 3DShapeNets and VoxNet, respectively, on ModelNet40 dataset, and OA is increased by 8.2 and 7.0. Our DANC-Net has achieved the above improvement because it is more effective in learning the features of point cloud than 3DShapeNets and VoxNet methods based on voxel.

(2) Comparison with the deep learning classification methods based on point cloud.

As shown in Table 3, we can observe that (1) mA of our DANC-Net is increased by 4.3 and 2.3, and OA is improved by 3.7 and 2.3, respectively, compared with the classical PointNet and PointNet ++. (2) Compared to Kd-Net based on KD tree, mA of our DANC-Net is increased by 2.0, and OA is improved by 1.1. (3) Compared to DGCNN, PointCNN, and A-CNN based on convolution, mA of our DANC-Net is increased by 1.1, 2.4, and 0.6, respectively, and OA is increased by 1.1, 0.7, and 0.7, respectively. (4) Compared with the recent point cloud classification network PointHop [37] and MRFGAT [38], mA of our DANC-Net is increased by 6.1 and 0.4, and OA is increased by 3.8 and 0.4. Our DANC-Net demonstrates high classification performance because it extracts local features of point clouds to obtain more information. In addition, in order to obtain higher classification accuracy, CASA is added to our DANC-Net, which can take advantage of local feature correlations.

Meanwhile, mA of our DANC-Net is increased by 1.1, and OA is increased by 0.6, respectively, compared with DGANet [39]. There are two reasons for achieving high performance: (1) our DANC-Net can dynamically weight local features by CASA module. (2) Loss function with negative constraint is used to eliminate the interference between point cloud categories with similar structures. In contrast, DGANet introduces offset attention into graph-based methods. The accuracy of constructing local graph impacts the results of feature extraction. As a consequence, the classification accuracy of DGANet is inferior to our DANC-Net.

Besides, we compare our method with SRN-PointNet++ [40], which can extract geometrical relationship between points. OA of our DANC-Net is 1.4 higher than that of SRN-PointNet++. We think that (1) our DANC-Net uses CASA to assign different weights to features, improving the flexibility of network. SRN-PointNet++ uses MLPs to obtain geometrical relationship between points, and the weights are the same for each local feature. Therefore, the classification accuracy of SRN-PointNet++ is not as good as that of our DANC-Net. (2) ModelNet40 contains much more categories of 3D point cloud shape than ModelNet10. Therefore, the point clouds in ModelNet40 have higher shape similarity and smaller distance between categories than those in ModelNet10. Under the circumstances, the loss function with negative constraint plays a prominent role, which further improves the classification performance of DANC-Net. To sum up, our proposed DANC-Net obtains the better performance in the final results than other deep learning methods on the task of point cloud classification.

The OA changes of our DANC-Net, PointNet, and A-CNN with epoch times are depicted in Figure 4. It can be found that (1) the OA of our DANC-Net is consistently higher than that of A-CNN when the OA reaches a stationary stage. (2) The OA of our DANC-Net is always higher than that of PointNet throughout the testing process.

4.3. Ablation Study. In order to verify the effectiveness of the dual-attention CASA module and the loss function with negative constraint (NC-loss) in our DANC-Net network,

Table 3: Classification performance on ModelNet40. (The top three accuracies are highlighted by bold, underline, and italic.)

Methods	Input	Points (k)	mA (%)	OA (%)
VoxNet (IROS 2015)	Points	1	83.0	85.9
3DShapeNets (CVPR 2016)	Points	1	77.3	84.7
PointNet (CVPR 2017)	Points	1	86.2	89.2
PointNet++ (CVPR 2017)	Points + normal	5	87.9	91.9
Kd-Net (ICCV 2017)	Points	32	88.5	91.8
PointCNN (NeurIPS 2018)	Points + normal	1	88.1	92.2
DGCNN (TOG 2019)	Points	1	90.2	*92.3*
A-CNN (CVPR 2019)	Points + normal	1	89.9	92.2
SRN-PointNet++ (CVPR 2019)	Points	1	—	91.5
PointHop (IEEE *T* MULTIMEDIA 2020)	Points	1	84.4	89.1
DGANet (remote sensing 2021)	Points	1	89.4	*92.3*
MRFGAT (INT *J* ANTENN PROPAG 2021)	Points	1	90.1	92.5
Ours	Points + normal	1	**90.5**	**92.9**

Input and points represent the input data type and the number of sampling points, respectively.

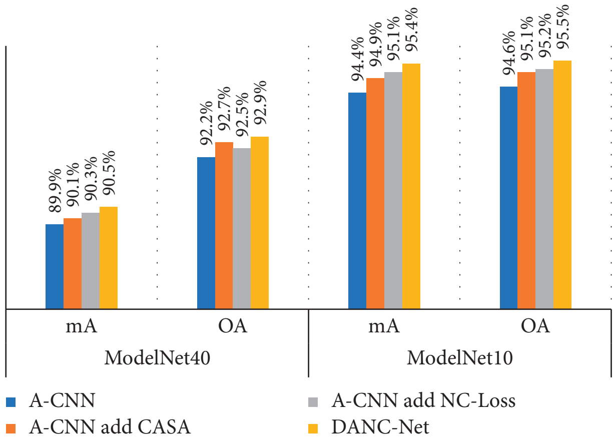

Figure 5: Classification effects of ablation experiments.

Table 4: Overall accuracy in % on ScanObjectNN. (The top three accuracies are highlighted by bold, underline, and italic.)

Methods	OBJ_BG	PB_T25	PB_T25_R	PB_T50_R	PB_T50_RS	Mean of OA
PointNet	79.0	74.5	73.2	69.3	67.8	72.7
PointNet++	83.5	85.4	82.8	80.9	78.7	82.2
DGCNN	85.3	**85.7**	83.8	*80.9*	*81.0*	83.3
A-CNN	85.1	83.2	83.5	81.8	81.4	*83.0*
Ours	**86.4**	84.3	**84.0**	**82.7**	**81.8**	**83.8**

this paper conducts ablation experiments on ModelNet10 and ModelNet40 datasets. Besides our DANC-Net, three additional models are designed in these experiments, including A-CNN, A-CNN + CASA module, and A-CNN + NC-loss.

In Figure 5, we observe that (1) CASA module and NC-loss contribute different degrees of improvement in classification performance from A-CNN to A-CNN + CASA and from A-CNN to A-CNN + NC-loss. (2) When both CASA module and NC-loss are added to the A-CNN, the classification accuracy will reach the maximum.

4.4. Robustness Test. We employ ScanObjectNN dataset to test the robustness of our DANC-Net. The objects in ScanObjectNN, which are selected from SceneNN [41] and ScanNet [42] scenes, are screened by the bounding boxes. In Table 4, we summarize the OAs of our DANC-Net and the state-of-the-art methods on ScanObjectNN dataset. In Table 4, OBJ_BG, PB_T25, PB_T25_R, PB_T50_R, and PB_T50_RS are five subsets of ScanObjectNN. OBJ_BG is point cloud with background, and PB denotes point cloud with random disturbance. T25 or T50 denotes point cloud after translation of 25% or 50% is performed. *R* and S denote rotation and scaling, respectively.

As presented in Table 4, we find that (1) our DANC-Net achieves the highest OA of 86.4 in OBJ_BG (without perturbation), compared with other methods. (2) Our DANC-Net also outperforms other methods on disturbed PB_T25_R, PB_T50_R, and PB_T50_RS datasets. (3) Overall, our DANC-Net has the highest average OA of 83.8

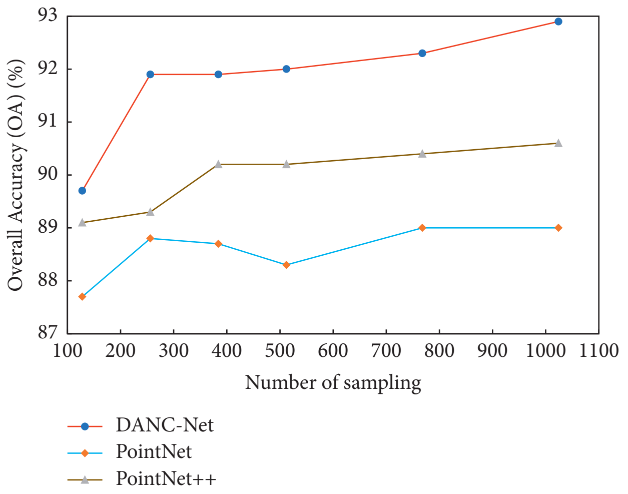

Figure 6: Robustness test of our DANC-Net to a sparse point cloud representation.

on ScanObjectNN. In summary, our DAC-Net performs better than other methods, which demonstrates its robustness on real-world datasets.

Figure 6 shows the robustness of our DANC-Net when ModelNet40 dataset is used for the test, in which 25%, 50%, 62.5%, 75%, and 87.5% of the input sampling points are randomly selected and discarded. The number of sampling points for training and testing is the same.

As shown in Figure 6, (1) our DANC-Net also achieves the highest classification accuracy, no matter whether the input point cloud is dense or sparse. (2) Compared with DANC-Net when 1024 sampling points are used, when 25% of sampling points are randomly dropped, the OA of the DANC-Net is only 0.6 lower. (3) Compared with PointNet when 1024 sampling points are used, when 87.5% of sampling points are randomly dropped, the OA of the DANC-Net is higher. It is shown that our DANC-Net is robust to point cloud sparsity.

5. Conclusions

At present, most point cloud classification models fail to explore the correlation between local regional features, and they use ground-truth as positive information to guide the network training, ignoring negative information. In view of this issue, we propose a new model of dual-attention and negative constraint network (DANC-Net). Our DANC-Net strengthens the interaction between local features of point cloud signals from both channel and space. At the same time, positive information and negative information are used effectively to improve the classification ability of our DANC-Net. Experimental results on synthetic datasets demonstrate that our DANC-Net successfully achieves high classification performance on point cloud classification tasks. Experimental results on real-world datasets confirm that our DANC-Net is robust. In the contrastive learning, the hard negative example is beneficial to enhance the ability of the network to distinguish between signal and noise. Therefore, in the future, we will explore new strategies to increase the number of the hard negative samples, so as to improve the classification ability of the DANC-Net.

Data Availability

The datasets used in this manuscript are available at https://github.com/artemkomarichev/a-cnn and https://hkust-vgd.github.io/scanobjectnn/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Nos. 61871258 and U1703261), Natural Science Foundation of Hubei Province of China (No. 2021CFB004), and Open Fund of the Key Laboratory of Oil and Gas Resources and Exploration Technology of the Ministry of Education (K2021-01).

References

[1] J. Cong, X. Wang, C. Yan, L. T. Yang, M. Dong, and K. Ota, “CRB weighted source localization method based on deep neural networks in multi-UAV network,” *IEEE Internet of Things Journal*, p. 1, 2022.

[2] X. Wang, L. T. Yang, D. Meng, M. Dong, K. Ota, and H. Wang, “Multi-UAV cooperative localization for marine targets based on weighted subspace fitting in SAGIN environment,” *IEEE Internet of Things Journal*, vol. 9, 2021.

[3] F. Wen, J. Shi, and Z. Zhang, “Generalized spatial smoothing in bistatic EMVS-MIMO radar,” *Signal Processing*, vol. 193, Article ID 108406, 2022.

[4] G. Zheng, Y. Song, and C. Chen, “Height measurement with meter wave polarimetric MIMO radar: signal model and MUSIC-like,” *Algorithms*, vol. 190, Article ID 108344, 2022.

[5] J. Shi, Z. Yang, and Y. Liu, “On parameter identifiability of diversity-smoothing-based MIMO radar,” *IEEE Transactions on Aerospace and Electronic Systems*, p. 1, 2021.

[6] D. Maturana and S. Scherer, “VoxNet: a 3D convolutional neural network for real-time object recognition,” in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, Hamburg, Germany, October 2015.

[7] C. R. Qi, H. Su, M. Niebner, A. Dai, M. Yan, and L. J. Guibas, “Volumetric and multiview cnns for object classification on 3D data,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5648–5656, Las Vegas, NV, USA, June 2016.

[8] Z. Wu, S. Song, A. Khosla et al., “3D ShapeNets: a deep representation for volumetric shapes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1912–1920, Boston, MA, USA, June 2015.

[9] H. Huang, E. Kalogerakis, S. Chaudhuri, D. Ceylan, V. G. Kim, and E. Yumer, “Learning local shape descriptors from Part Correspondences with multiview convolutional networks,” *ACM Transactions on Graphics*, vol. 37, no. 1, pp. 1–14, 2018.

[10] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, “Multiview convolutional neural networks for 3d shape recognition,” in *Proceedings of the IEEE International Conference on*

Computer Vision, pp. 945–953, Santiago, Chile, December 2015.

[11] G. Riegler, A. Osman Ulusoy, and A. Geiger, "Octnet: learning deep 3d representations at high resolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3577–3586, Honolulu, HI, USA, July 2017.

[12] T. Le and Y. Duan, "PointGrid: a deep network for 3d shape understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9204–9214, Salt Lake City, UT, USA, June 2018.

[13] B.-S. Hua, M.-K. Tran, and S.-K. Yeung, "Pointwise convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.

[14] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: deep learning on point sets for 3D classification and segmentation," in *Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 77–85, Honolulu, HI, USA, July 2017.

[15] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: deep hierarchical feature learning on point sets in a metric space," in *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*, pp. 5100–5109, Long Beach CA USA, December 2017.

[16] Y. Li, R. Bu, M. C. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: convolution on x-transformed points," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 820–830, Montréal, Canada, December 2018.

[17] M. Tatarchenko, J. Park, V. Koltun, and Q. -Y. Zhou, "Tangent convolutions for dense prediction in 3d," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3887–3896, Salt Lake City, UT, USA, June 2018.

[18] S. Qiu, S. Anwar, and N. Barnes, "Geometric back-projection network for point cloud classification," *IEEE Transactions on Multimedia*, vol. 24, pp. 1943–1955, 2017.

[19] M. A. Uy, Q.-H. Pham, B.-S. Hua, T. Nguyen, and S. -K. Yeung, "Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1588–1597, Seoul, Korea (South), November 2019.

[20] J. Liu, Z. Chen, B. Du, and D. Tao, "ASTS: a unified framework for arbitrary shape text spotting," *IEEE Transactions on Image Processing*, vol. 29, pp. 5924–5936, 2020.

[21] X. Huang, B. Du, and W. Liu, "Multichannel color image denoising via weighted schatten p-norm minimization," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence IJCAI*, Yokohama, July 2020.

[22] H. Sun, Y. Zhang, P. Chen et al., "Scale-free heterogeneous cycleGAN for defogging from a single image for autonomous driving in fog," *Neural Computing & Applications*, pp. 1–15, 2021.

[23] H. Sun, J. Li, J. Chang, B. Du, and Z. Su, "Efficient compressive sensing tracking via mixed classifier decision," *Science China Information Sciences*, vol. 59, no. 7, pp. 072102–072115, 2016.

[24] A. Komarichev, Z. Zhong, and J. A. Hua, "CNN: annularly convolutional neural networks on point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7421–7430, Long Beach, CA, USA, June 2019.

[25] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the Advances in neural information processing systems*, Long Beach, CA, USA, December 2017.

[26] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," in *Proceedings of the Advances in Neural Information Processing System*, Vancouver, BC, Canada, December 2019.

[27] P. Bhattacharyya, C. Huang, and K. Czarnecki, "Sa-det3d: self-attention based context-aware 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3022–3031, Montreal, BC, Canada, October 2021.

[28] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. hoi, and Y. W. Teh, "Set transformer: a framework for attention-based permutation-invariant neural networks," in *Proceedings of the International Conference on Machine Learning*, pp. 3744–3753, PMLR, Long Beach, CA, USA, June 2019.

[29] D. A. Shajahan, V. Nayel, and R. Muthuganapathy, "Roof classification from 3-D LiDAR point clouds using multiview CNN with self-attention," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1465–1469, 2020.

[30] C. Moenning and N. A. Dodgson, *Fast Marching Farthest point Sampling*, University of Cambridge, Computer Laboratory, Cambridge, 2003.

[31] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: feature fusion attention network for single image dehazing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 11908–11915, New York, NY, USA, February 2020.

[32] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7794–7803, Salt Lake City, UT, USA, June 2018.

[33] P. Khosla, P. Teterwak, C. Wang et al., "Supervised contrastive learning," in *Proceedings of the Advances in Neural Information Processing Systems*, vol. 33, pp. 18661–18673, Curran Associates, Inc, December 2020.

[34] I. Misra and L. van der Maaten, "Self-supervised learning of pretextinvariant representations," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR*, Seattle, WA, USA, June 2020.

[35] R. Klokov and V. Lempitsky, "Escape from cells: deep kd-networks for the recognition of 3d point cloud models," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 863–872, Venice, Italy, October 2017.

[36] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics*, vol. 38, no. 5, pp. 1–12, 2019.

[37] M. Zhang, H. You, P. Kadam, S. Liu, and C.-C. J. Kuo, "Pointhop: an explainable machine learning method for point cloud classification," *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1744–1755, 2020.

[38] X. A. Li, L. Y. Wang, and J. Lu, "Multiscale receptive fields graph attention network for point cloud classification," *Complexity*, vol. 2021, Article ID 8832081, 9 pages, 2021.

[39] J. Wan, Z. Xie, Y. Xu, Z. Zeng, D. Yuan, and Q. Qiu, "DGANet: a dilated graph attention-based network for local feature extraction on 3D point clouds," *Remote Sensing*, vol. 13, no. 17, p. 3484, 2021.

[40] Y. Duan, Y. Zheng, J. Lu, J. Zhou, and Q. Tian, "Structural relational reasoning of point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 949–958, Long Beach, CA, USA, June 2019.

[41] B. S. Hua, Q. H. Pham, D. T. Nguyen, M. -K. Tran, L. -F. Yu, and S. -K. Yeung, “Scenenn: a scene meshes dataset with annotations,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. No.01, pp. 8778–8785, Honolulu, HI, USA, February 2019.

[42] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, “Scannet: richly-annotated 3d reconstructions of indoor scenes,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5828–5839, Honolulu, HI, USA, July 2017.