Journal of Engineering, Science and Mathematics

Volume 02, Number 02, Pages 36-55, Year 2023
https://jesm.in/archives/

Research Paper

Machine Learning Based Deep Cloud Model to Enhance Robustness and Noise Interference

Mohan Raparthi[1], Abhishek Agarwal [2]

[1]Software Engineer, alphabet Life Science, Dallas Texas, 75063, US

[2]Vermont State University, USA

Correspondence should be addressed to Mohan Raparthi; raparthimohan.scholar@gmail.com

Handling Editor: Abolfazl Mehbodniya

To improve the ability of the 3-D point cloud deep network classification model, this paper presents a competitive attention fusion module that may be applied to various classification networks. The intrinsic similarity between representation and intermediate features is used to redistribute the weights of intermediate feature channels. Implementation of the present module in the standard networks such as Point net++ and Point ASNL, were included to conduct the experiments. The suggested block is autonomous and transferrable, according to the results, and it concentrates on the core and backbone elements that are better suited for 3D point cloud form categorization. The presented module improves the model's resistance to point cloud disturbance noise, outlier noise, and random noise when compared to the state-of-the-art network without affecting classification accuracy. When the random noise numbers are 0, 10, 50, 100, and 200, the accuracies reach 93.2%, 92.9%, 85.7%, 78.2%, and 63.5%, respectively.

Keywords: *Machine Learning, Deep Learning, Cloud computing, Classification, Cloud disturbance, Anti-interference*

1. Introduction

In computer vision applications, 3D point cloud data largely makes up for the lack of spatial structure information of 2D images. The learning methods based on deep neural networks adopted by many researchers can be divided into following

types based on different 3D data expression methods; Methods for handcrafted feature preprocessing [1], multi-view [2], voxel [3] and raw point cloud data [4-5]. To improve learning outcomes, multiview learning attempts to combine facts from several points of view. These views are frequently available from a variety of sources or feature subsets. Voxels are isolated 3D units that were before volumes, locations, and properties, analogous to pixels in a picture. They can be used to physically describe distinct points in an ontologically explicit and information-rich way. For each attribute, a separate support vector machine is trained and then the collection of customized learners is integrated using the sum rule. The set of classifiers is primarily based on local binary pattern variations. The original 3D data can show the three-dimensional representation of the object, and the 3D point cloud is used as input to avoid unnecessary volume division caused by inputting regularized data such as multi-view and voxels in the convolutional network. Affected by the acquisition equipment and coordinate system, the arrangement order of 3D point-cloud data is very different. Data points are a by-product of 3D scanning operations and have a wide range of uses. The task of semantic segmentation is increasingly challenging, the dataset is more complicated, and 3D point-cloud deep network research is a significant area of research. Through the distribution of the multilevel feature extrusion excitation submodule and the feature intrinsic correlation self-attention submodule, 3D points cloud classification networks adaptively weight the intermediate feature channel weights. The performance in the semantic segmentation task is not as excellent as in the classification task when the CAF model is employed for the intermediate weight redistribution of the feature channel weight.

For the classification and segmentation of disordered point cloud data, reference [4] proposed PointNet network, directly deal with sparse unstructured point clouds. Reference [5] based on PointNet, down-sampled and grouped point clouds, and proposed PointNet++. PCPNet [6] is a novel multiscale variant based on the PointNet system. A patch-based learning method, Reference [7] improved PointNet by using local structure more effectively and performed recursive feature aggregation on the nearest-neighbor graph to obtain local high-dimensional features. SO-Net [8] used a self-organizing map algorithm (SOM) to obtain key feature points and obtain model descriptors that contain spatial information through the PointNet module. To obtain model descriptors with spatial information through the Point Net module, SO-Net employed the self-organizing map (SOM) approach to obtain important feature points. With the introduction of the transformation approach, point CNN was able to realize the potential regularization of disordered point clouds. Dense Point generalized the convolution operator and extended the regular grid CNN to irregular point configurations. The data from several local areas is gathered by the attention technique used by Point2 Sequence. A CNN encoded independently using a circular convolution layer in accordance with the various distances of the features and local centre points. Reference [9] used the two-dimensional target detection algorithm that reduces the detection range of the three-dimensional model to the

viewing cone and uses PointNet to obtain high-dimensional features. Reference [10] proposed PointCNN and introduced the X-Conv transformation method to realize the potential regularization of disordered point clouds. DensePoint [11] extended the CNN of the regular grid to irregular point configurations by generalizing the convolution operator. Point2Sequence [12] uses an attention mechanism to aggregate information from different local regions. ACNN [13] according to the different distance information of features and local center points, encoded separately through circular convolutions. PointWeb [14] associates each set of pair of points in a local region to obtain more expressive local features. Some other works have introduced the use of graph convolutional networks to learn local graphs [15-16] or geometric elements [17], extracting local features of point clouds. The attention mechanism [18] calculates the degree of correlation between features. The commonly used attention mechanism in 2D image processing is the SE module [19]. In the PVNet 3D point cloud network [20], an integrated attention fusion module fuses intermediate features with global features. PointASNL [21] uses a general self-attention mechanism to update intragroup features in an adaptive sampling module. The greatest contribution of the CAF module to the classification network is to increase the model's robustness and noise interference resistance. The experiment trains and tests two cutting-edge networks using the Model Net40 dataset, and the findings demonstrate that the competitive attention fusion module is helpful. By comparing the model with the state-of-the-art network, the reliability of the prototype can be substantially increased while preserving and increasing the accuracy of point-cloud classification. The model also has strong anti-interference ability to different numbers of point-cloud disturbances, outliers, and random noise. End-to-end learning reduces training complexity while improving anti-noise performances as compared to traditional filtering.

In the study of deep 3D point cloud classification networks, optimized feature extraction ability, improving the resistance to point cloud disturbance, outliers, random noise, and other interference factors are prominent research areas. This paper presents a competitive attention fusion (CAF) module, which is a transferable intermediate feature channel optimization structure, introduces residual connections and channel competitiveness, and reassigns feature channel weights through learning with two types of attention as the core value. The CAF module contains 2 submodules: 1) a multilevel feature extrusion excitation submodule, which focuses on the extraction and fusion of global features at different levels; 2) a feature intrinsic correlation self-attention submodule, which measures the intrinsic similarity of intermediate features. The presented CAF module can be embedded in different point cloud classification networks, has transferability and scalability, improves the expression ability of point cloud global features, and strengthens the robustness of the model to point cloud noise.

2. Competitive Attention Fusion Module

The point cloud feature extraction network uses 2 or more intermediate feature abstraction layers. The intermediate

features are often a collection of global and local features, which greatly affects the accuracy of the classification results. The CAF module learns from the 2-layer intermediate output features. Fusion weights, which can represent the importance and expressiveness of the intermediate feature channels of the current layer, and redistribute the channel features through the weights to obtain optimized new intermediate features. In short, the CAF module uses the central idea of the attention mechanism, aggregates salient features, stimulates channel features that are more important and have a greater impact on the results, suppresses invalid or inefficient channel features, reduces noise interference, and improves the model robustness.

The general idea of the CAF module is illustrated in Figure 1. In the figure, B is the batch data size (batch size), which is the number of point clouds after the current sample, N_i is downsampled in this layer, and C_i is the number of feature channels of the current sample in this layer. The module aggregates many layers channel features, and design two special submodules for parallel independent training, the purpose is to obtain the global attention weights that focus on different levels and the attention weights that pay attention to the internal correlation of intermediate features, and fuse them to obtain new weight coefficients. The weights of the intermediate channels are redistributed to realize the optimization of intermediate features. There are two specially designed modules, namely' multi-layer feature squeeze excitation' (MFSE) sub-module and feature intrinsic correlation self-attention' (feature inner connection self-attention, FICSA) sub-module.

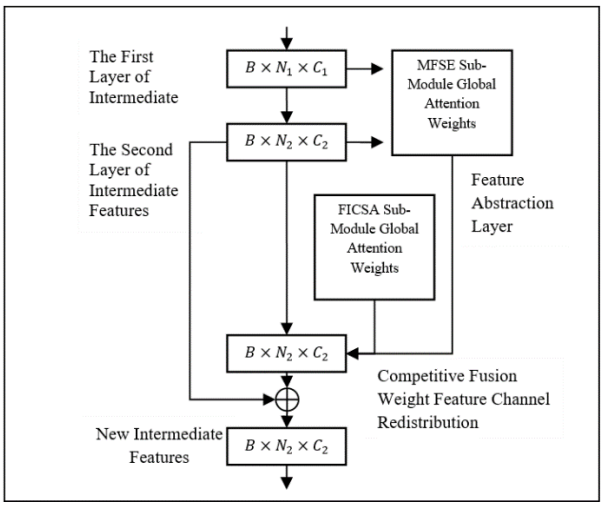

Figure 1: Schematic Diagram of the Competitive Attention Fusion Module

2.1. MFSE Sub-module

In the two-dimensional image classification task, the squeeze excitation net (SE-Net) [19] performs well, and the SE module adaptively adjusts the features of the channels according to the correlation between channels. The module can be added to a variety of two-dimensional classification networks as an independent structure to improve the network classification accuracy. Competitive squeeze excitation (CMPE-SE) [22] achieves a better mapping structure, based on the SE module by merging the competitive relationship between the residual mapping and the identity mapping, reimaging of the internal features of the two-dimensional image is realized. According to the specific task requirements, the intermediate channel of the fusion feature can be merged in parallel or convolutional fusion, as shown in Figure 2. In the figure, h, w, C_i are the height, width, and number of channels of the two-dimensional feature map, respectively?

The MFSE submodule in the CAF module is inspired by the 2D CMPESE module. The representation of disordered 3D point-cloud features is necessarily distinct from the regular 2D image feature map. The MFSE sub-module does not simply use the 2D CMPE-SE module. Based on the characteristics of the point cloud data structure and the three-dimensional classification network, the pooling layer, feature dimension, fully connected layer, feature fusion method, channel weighting, etc. are designed and adjusted independently accordingly. In order to prevent the convolution network's convolutional network from accidentally causing superfluous volume division, the three-dimensional representation of the object and the three-dimensional point cloud are used as input. The arranging order of 3D point-cloud data varies greatly depending on the acquisition apparatus and coordinate system. The PointNet network deals directly with sparse unstructured point clouds for the classification and segmentation of disordered point cloud data. An innovative multiscale variation of the PointNet system is PCPNet. PointNet, a patch-based learning technique, performed recursive feature aggregation on the nearest neighbor graph and made better use of spatial correlation.

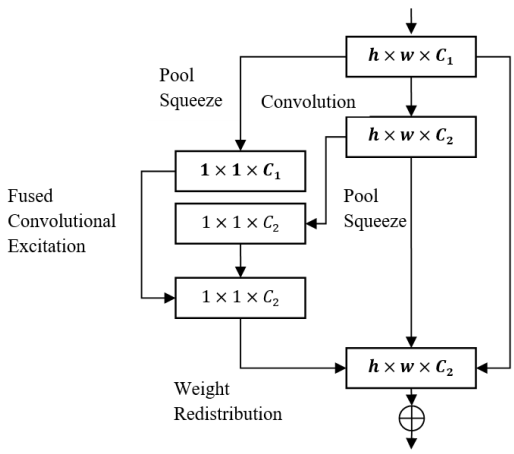

Figure 2: Structure Diagram of 2D Extrusion Excitation Sub-Modules

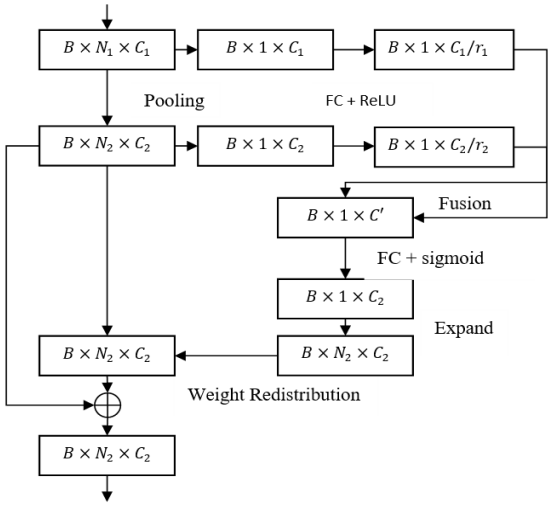

Figure 3: Structure Diagram of 3D Extrusion Excitation Sub-Modules

The network structure and the specific implementation process of the MFSE submodule are shown in Figure 3.

1) Parallel input: The two parallel input features are the intermediate features of the previous layer $f1 \in R^{N1\times C1}$ and the intermediate features of the current layer after the feature extraction layer is $f2 \in R^{N2\times C2}$.

2) Feature up-scaling: Pooling aggregates salient features on different channels, and two sets of input high-dimensional features are obtained from fully-connected layer and the activation function, respectively. The formula for feature up-scaling is:

$$\varphi_i(f_i) = \varphi_i(P_i(f_i))/r_i; i = 1,2 \ldots\ldots\ldots\ldots\ldots..(1)$$

In the formula: is the global feature aggregation, such as the maximum pooling function Max pooling; is the fully connected layer (FC) and the ReLU activation function, considering the training difficulty caused by the large number of parameters, set the channel scaling ratio to adjust the intermediate

channel number, $C_i \rightarrow C_i/r_i$ the number of channels of the output feature.

3) Feature fusion: The formula for the fusion of two groups of high-dimensional features is the following.

$$F(f1, f2) = \varphi_1(f1) \oplus \varphi_2(f2) \ldots\ldots\ldots (2)$$

Among them, the channels are connected proportionally by expanding the number of columns $F \in R^{1\times C'}$, and $C' = C_1/r_1 + C_2/r_2$ the global high-dimensional features of different levels of features are learned, as shown in Figure 3. After further calculation, the global attention weight of the feature channel of the current layer is obtained as:

$$W_{se}(f_1, f_2) = \emptyset\big(F(f_1, f_2)\big) \ldots\ldots\ldots\ldots (3)$$

Where: $\emptyset$ is the fully connected layer with the normalization function $Sigmoid$, $W_{se}(f_1, f_2) \in R^{1\times C_2}$ is the final global attention weight.

4) Weight expansion: The global attention weight obtained by the MFSE module represents the score corresponding to each channel of the intermediate feature of the current layer. The feature channel with a high score is more influential for the next layer training, and the feature channel with a low score represents the feature of the channel. It has a slight effect on the network classification results. To adapt to the point-cloud data structure, the dimension of the global attention weight is expanded, so that the dimensionN_2 is the count of sampling points of the sample in the current layer.

The MFSE submodule pays attention to the global influence of intermediate features at different levels, and calculates the contribution of global features to the channel weights by competitively fusing the significant high-dimensional features of the two-layer features. This module is beneficial to improve the stability of the model classification results.

2.2. FICSA Submodule

To improve the expressive ability of the intermediate features and enhance the influence of the internal association of the intermediate features on the classification results, further research is done using the attention mechanism to learn the intermediate features.

The CAF module's FICSA submodule introduces the self-attention mechanism and establishes the intrinsic interest weights for evaluating the effects of intermediate characteristics on categorization outputs. The FICSA submodule differs in that it uses attention concurrently with the network backbone, independently learning the internal relationship of the intermediate features, and taking part in the redistribution of channel weights rather than in the feature abstraction layer or adaptable sampling. The 2D CMPE-SE module is not the only one that the MFSE sub-module uses. The pooling layer, feature dimension, fully connected layer, feature fusion method, channel weighting, etc. are created and changed separately according to the peculiarities of the point-cloud data structure and the three-dimensional categorization network. The network structure and the specific implementation process of the FICSA submodule are shown in Figure 4.

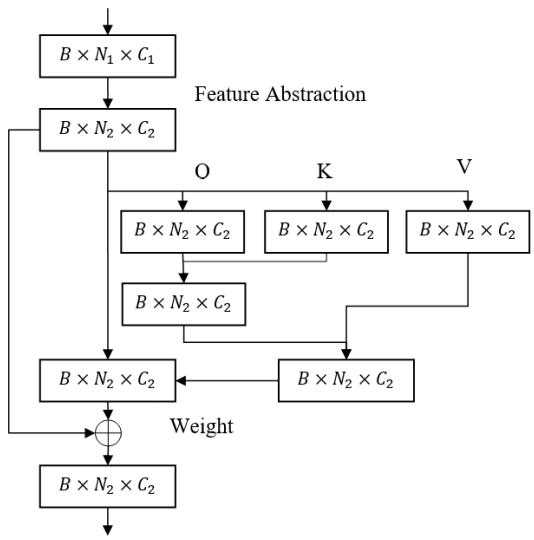

Figure 4: Structure Diagram of Feature Intrinsic Correlation Self-Attention Sub-Module

1) Parallel convolution: Input the intermediate feature $f_2 \in R^{N_2 \times C_2}$ of the current layer, and after 1×1 point convolution, the features of all channels at each point are linearly mapped to 3 parallel high-dimensional features, the formula is:

$$\begin{cases} V(f_2) = w_v * f_2, \\ Q(f_2) = w_q * f_2, \ldots\ldots\ldots\ldots (4) \\ K(f_2) = w_k * f_2 \end{cases}$$

In the formula: V, Q, K are 3 independent feature mapping functions respectively; w_i are different linear conversion coefficients, meaning that 1×1 point convolution is performed in the feature channel dimension, C_2 the number of convolution kernels is , and 3 the corresponding high-level features, the dimensions are $N_2 \times C_2$.

2) Similarity calculation: Obtaining the difference between Q and K through dot product operation, the formula is as follows.

$$A(f_2) = \gamma\left[Q(f_2)K^T(f_2) = \sqrt{C}\right] \ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

In the formula: A is the internal high-dimensional relationship of the intermediate features; γ is the aggregation function, such as the Softmax normalization function; $\sqrt{C}$ is the optional channel scaling factor, $W_{sa}(f_2) \in R^{N_2 \times C_2}$ the setting purpose is to reduce the number of training parameters. Obtain the global representation of the intrinsic correlation of features between points Attention weight, the formula is

$$W_{sa}(f_2) = \gamma\big(A(f_2)V(f_2)\big) \ldots\ldots\ldots\ldots\ldots (6)$$

Where: V is the dimension of the feature channel used to adjust A. As shown in Figure 4, after the operation of formula (5), $N_2 \times N_2$ the output feature dimension is, $N_2 \times C_2$ and the output feature dimension is adjusted to the FICSA sub-module, paying attention to the intrinsic similarity of intermediate features, and through this module, the ability of the network to extract global features can be improved, and the contribution of global features to channel weights can be obtained from another perspective. The robustness of the model against noise can be enhanced with the help of this module. The strengths of intermediate feature channels are redistributed on the basis of the intrinsic similarity between representation and intermediate features. Conducted experiments while implementing the presented module in well-known networks like Point net and Point ASNL. The results show that the proposed block is autonomous, portable, and focuses on the core and backbone components that are more appropriate for 3D point cloud form categorization. When compared to the most advanced network, the proposed module enhances the model's

resistance to point cloud disturbance noise, outlier noise, and random noise without reducing classification performance.

2.3. CAF Module

The CAF module is composed of the MFSE submodule and the FICSA submodule. The two submodules are independently learned in parallel, the competitive feature fusion, the residual learning is introduced, and the feature channel weights are re-divided:

$$W_{CAF} = \alpha W_{se}(f1, f2) + \beta W_{sa}(f2) \ldots\ldots\ldots\ldots (7)$$

In the formula: $\alpha and \beta$ are the global attention weights andthe scale coefficients of W_{se}, W_{sa} , respectively. Values are adjusted through experiments according to the characteristics of different data set samples. Usually, it can be set to 1:1. The MFSE sub-module is used to maintain the stability of the classification accuracy. Decreasing the FICSA sub-module enhances the robustness of the model, if the point cloud noise is more disturbed, it can be adjusted$\alpha < \beta$, and otherwise it can be adjusted as$\alpha > \beta$.

Through matrix addition, $\alpha and \beta$ the two global attention weights are fused according to different scale coefficients to obtain the final weight distribution coefficient$W_{CAF} \in R^{N_2 \times C_2}$, This coefficient contains global learning and intrinsic feature similarity measure for multilevel features, and two sets of weights compete to provide an optimization scheme for feature channels. New intermediate features are obtained through weight redistribution and residual connection is$f_2^* \in R^{N_2 \times C_2}$.

$$f_2^* = f_2 + W_{CAF} f_2 \ldots\ldots\ldots\ldots (8)$$

3. Network structure of point-cloud classification based on CAF module

The 3-D point cloud classification network generally obtains the high-dimensional features of the sample through two or more layers of feature extraction layers, obtains the global features through the pooling layer, and obtains the classification score of the point cloud through the fully connected layer learning. The first half of the high-dimensional feature extraction is the core of the classification network, which determines the information analysis ability, classification accuracy, robustness, and other evaluation indicators of the model. Based on various 3D data expression techniques, the learning techniques based on deep neural networks used by various researchers can be categorized into the following categories: voxel, multiview, and raw point cloud data pre-processing techniques using handmade features. The original 3D data can display the item in three dimensions, and the 3D point cloud is used as input to prevent unwanted volume division brought on by inserting regularized data into the convolutional network, such as multi-view and voxels. The arranging order of 3D point-cloud data varies greatly depending on the collecting apparatus and coordinate system.

The noise interference in the actual point cloud includes disturbances and outliers, which are often manifested as the position offset of some point sets of the sample, and there is background noise. When the model is tested, the noise point set is also regarded as part of the sample, which affects the classification result of the sample. The role of the CAF module in the

network is to make the model pay more attention to the core features that determine the type of sample by adjusting the weights of the intermediate feature channels. The two submodules learn from two different perspectives: a) the multi-level global features, b) the internal correlation of the intermediate features. Obtaining weights that are more helpful to focus on the core channel, improve the network's ability to learn global features, strengthen the model's anti-interference ability, and help solve difficult problems in point-cloud deep networks.

The CAF module can be migrated as an independent optimization module and applied to a similar 3-D point cloud classification network framework, which is embedded in the network. The initial features of point cloud samples are usually normal vectors $F_0 \in R^{N_0 \times 3}$,after passing through the feature extraction layer, $F_1 \in R^{N_1 \times C_1}$ output intermediate features. Cancel the output connection of the original network feature extraction layer. After redistributing the weights of the feature channel through the CAF module, a new intermediate feature is obtained $F_1^* \in R^{N_1 \times C_1}$,and as the input of the next feature extraction layer and the second CAF module. After 2 layers of feature extraction after redistributing with the channel weights, the global features are obtained by the pooling layer and the classification score is calculated by the fully connected layer. When applied in different networks, the implementation method of the feature extraction layer is theoretically analyzed for the specific classification network, corresponding to the task target. Adjust the weight distribution ratio of the two sub-modules in the CAF module to attain a network structure with better performance.

4. Experiment and result analysis

The classification experiments, robustness analysis, and comparison are carried out on the 3-D point cloud dataset ModelNet40[3], including 9843 training samples and 2468 testing samples, all samples are divided into 40 categories. In addition, the experimental analysis the necessity of the module, and perform semantic segmentation experiments. All experiments are based on tensor flow, applying 1 GTX 2080Ti GPU.

Experiments verify the effectiveness and transferability of the CAF module on the state-of-the-art networks, namely Pointnet++ and Point ASNL, respectively. The results show that adding the CAF module can enhance the network to point cloud without reducing the average accuracy of the classification results. The anti-interference ability of disturbance, outliers and random noise, by adjusting the number of input points of training samples, the robustness of the model can be further improved while keeping the classification accuracy stable.

After many experiments, the training parameters are set as follows: the channel scaling ratio in formula (1) is set to 4, and the weight fusion ratio in formula (7) is set to 1. Unless otherwise specified, when the state-of-the-art network is Pointnet++, the count of input points is 1024, the batch size is 16, and the result is the average of 12 tests; when the state-of-the-art network is PointASNL, the count of input points is 1024, the batch size is 24, and the result is the average of 5 tests.

4.1. Shape Classification

The CAF module is added to Pointnet++, the normal vector is added in both training and testing, and the point cloud is randomly rotated to simulate the real scene during testing. Since the optimal training model is not provided in the Reference [5] (the number of input points is 5000, the optimal accuracy 91.9%) detailed training parameters, this study reproduces the optimal classification accuracy of the Reference [5] is 90.7%, the average classification accuracy after adding the CAF module is 91.0%, the improvement of the classification accuracy proves that the CAF is a CAF The validity and feasibility of the module. When the CAF module is added to PointASNL, when only the coordinate points are input, the classification accuracy is 92.9% (92.88%), which is not lower than 92.9% in the Reference [21] (the optimal classification accuracy in the actual test is 92.85%); when adding normal vectors to training and testing, the classification accuracy is 3.2% (93.19%), not lower than 93.2% in the Reference [21] (the optimal classification accuracy of the actual test is 93.15%), as shown in Table 1.

TABLE 1: AVERAGE CLASSIFICATION ACCURACY IN THE MODELNET40 DATASET

Method	**Enter**	$N_{in}/10^3$	$A_{cc}/\%$
PointNet	Pt	1	89.22
SO-Net	Pt, Nml	2	90.92
PointNet++	Pt, Nml	5	91.92
PointCNN	Pt	1	92.22
Point2Sequence	Pt	1	92.62
A-CNN	Pt, Nml	1	92.62
PointASNL	Pt	1	92.87
PointASNL	Pt, Nml	1	93.17
This Work	Pt	1	92.9
This Work	Pt, Nml	1	93.21

In the table, N_{in} is the count of input points, A_{cc} is the classification accuracy, Pt is the input 3-D point cloud coordinate data, and Nml is the input 3-D point cloud normal vector. The experimental results reveal the independence and transferability of the CAF module, and it is helpful to maintain the classification accuracy.

4.2. Robustness analysis

The biggest contribution of the CAF module to the classification network is to improve the model's resistance to noise interference and enhance the robustness of the model. Many classification models only consider the performance on the complete point cloud dataset, and do not consider the random background that may exist in actual situations. Therefore, some models with excellent classification performance do not necessarily have strong anti-interference ability. Several classification models do not account for the random background that may be present in real-world scenarios and merely evaluate performance on the entire point cloud dataset. As a result, not all models with high anti-interference capabilities also have excellent classification performance. The real point cloud gathering is generally

disorganized and noisy due to outliers and disturbances, which are restricted by the collection scene and sensor accuracy. Disturbance and noise are frequently present in point-cloud noise.

Outliers Point cloud noise: Disturbance and outliers in the actual point cloud data collection, limited by the collection scene and sensor accuracy, the collected point cloud is often disordered and noisy. The point cloud noise usually includes disturbance and noise. Disturbances represent data points that fluctuate within a certain range above and below the sampling plane, and outliers represent outliers that appear randomly at any location in space. In the experiment, Gaussian noise was appended to the point cloud to simulate the disturbance, and the standard normal distribution was used; random noise was appended to the point cloud to simulate outliers and the noise points. In Model Net 40's typical aeroplane model, for instance, there are a total of 10,000 data points, and Gaussian noise from a regular normal distribution is added. To test a single aircraft model using the most recent model, PointASNL, 51.60 msvoxel filtering, statistical filtering, and comprehensive filtering (radius filtering), respectively. The conventional normality assumption and Gaussian noise were added to the point cloud to represent disturbance; random noise was added to the point cloud to mimic outliers and noisy points. The experimental CAF module was used to examine outliers and disturbances using PointASNL as the leading network. Taking PointASNL as the state-of-the-art network, the experimental CAF module was used to analyze disturbances and outliers. Using the trained model to test, the results are depicted in Figure 5 in which n is the number of noise points. After the CAF module is added, the anti-interference performance of the model to the point cloud disturbance and outliers is significantly improved. Figure 6 shows Anti-Jamming Performance w.r.t. different Random noise types.

Model robustness: To further test the effectiveness of the CAF module in model robustness, referring to the methods for testing model robustness in Point ASNL [21] and KCNet [7], a certain number of original point sets are replaced by random noise, simulating the situation of data loss and noise interference at the same instant of time, the random noise number n is 0, 1, 10, 50, 100, respectively.

Figure 7 shows the classification accuracy of the Pointnet++ network with the CAF module and the original network on the random noise test set. The results show that as the amount of noise increases, the classification accuracy of the network with the CAF module decreases more slowly, and the robustness of the model is significantly improved.

The CAF module is added to the PointASNL network for robustness experiments. To test the impact of input points on the anti-interference performance of the CAF module, the input point cloud numbers are set to 1024 (1k), 2048 (2k), 3000 (3k), the experimental results are shown in Figure 8. It is observed that for the model trained with 1024 points input, under the same conditions, after adding the CAF module, the anti-interference ability of the network is improved under different amounts of random noise, if the model is considered to be applied to the missing points. In the live point cloud with a lot of random noise, appropriately increasing the

number of input points can obtain better anti-interference ability while keeping the classification performance stable. With the increase of the number of interference points, CAF uses 2048 and 3000 input points during training, and the CAF module exhibits stronger classification performance and robustness compared to 1024 input points. Figure 9 shows the impact of the CAF module on the robustness of the model.

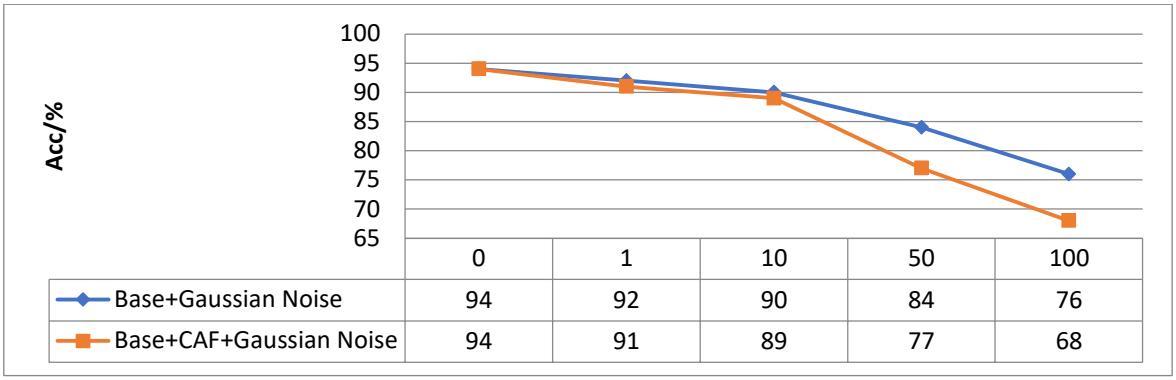

Figure 5: Anti-jamming performance w.r.t. different Gaussian noise types

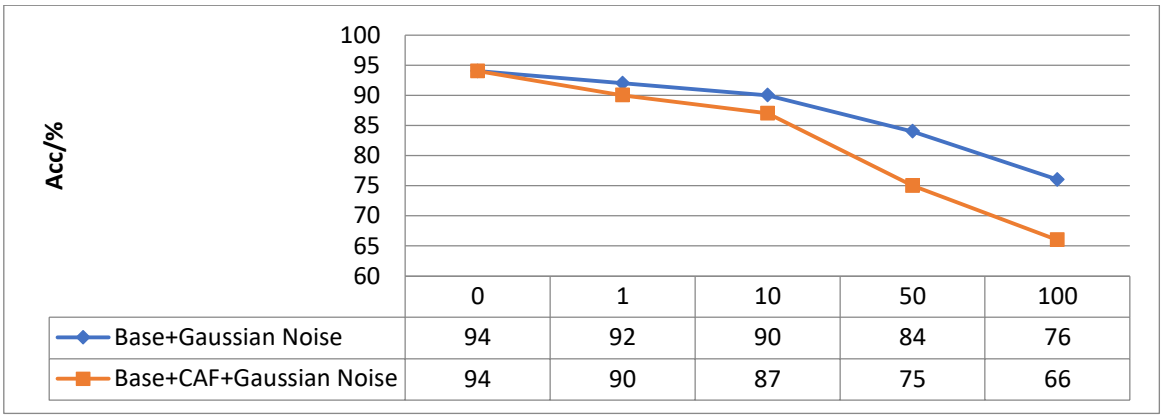

Figure 6: Anti-Jamming Performance w.r.t. different Random noise types

To further verify the contribution of the CAF module to the robustness of the model, experiments on the extreme anti-interference ability and classification performance are carried out on the PointASNL network with the CAF module, as shown in Fig. 5(c). Comparing Figs. 5(a), 5(b) and 5(c), if the classification accuracy is 50% as the acceptable limit of the noisy point cloud, when the number of training points is 1024, Pointnet++ can only contain (50±5) interference points, and after adding the CAF module, it is allowed to contain (90±5). There are 240±5 points of interference points in PointASNL, 280±5 points are allowed after adding the CAF module, and the number of interference points can be increased by increasing the count of training points to 300±5 points.

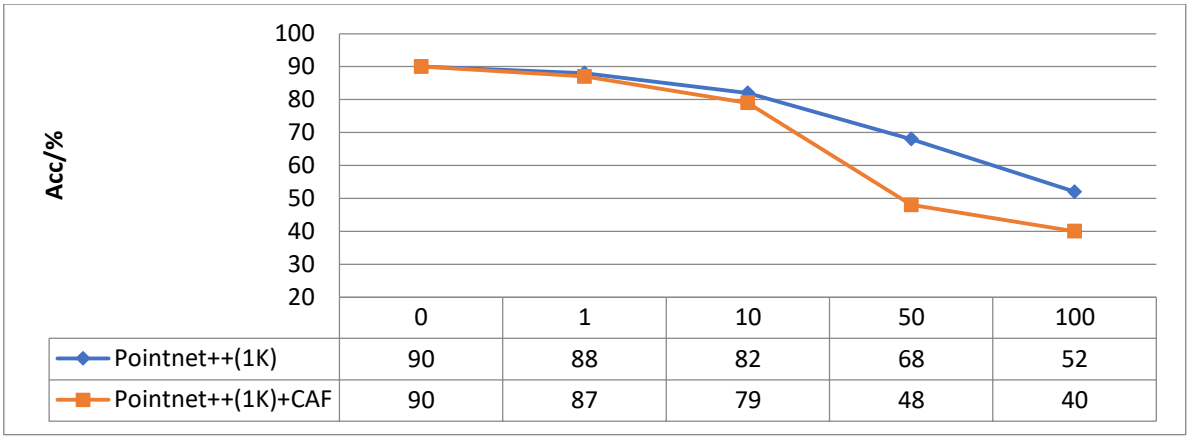

Figure 7: Pointnet++

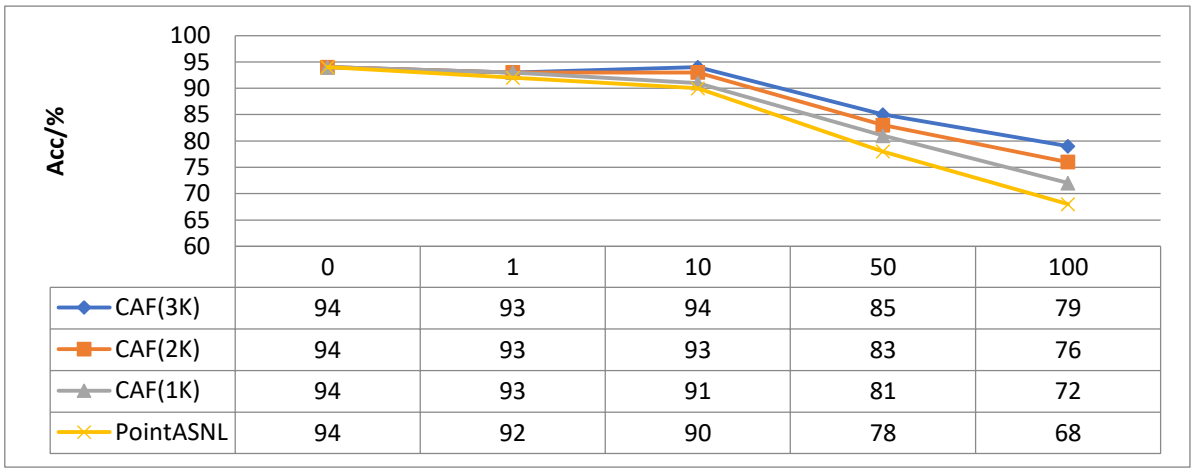

Figure 8: PointASNL

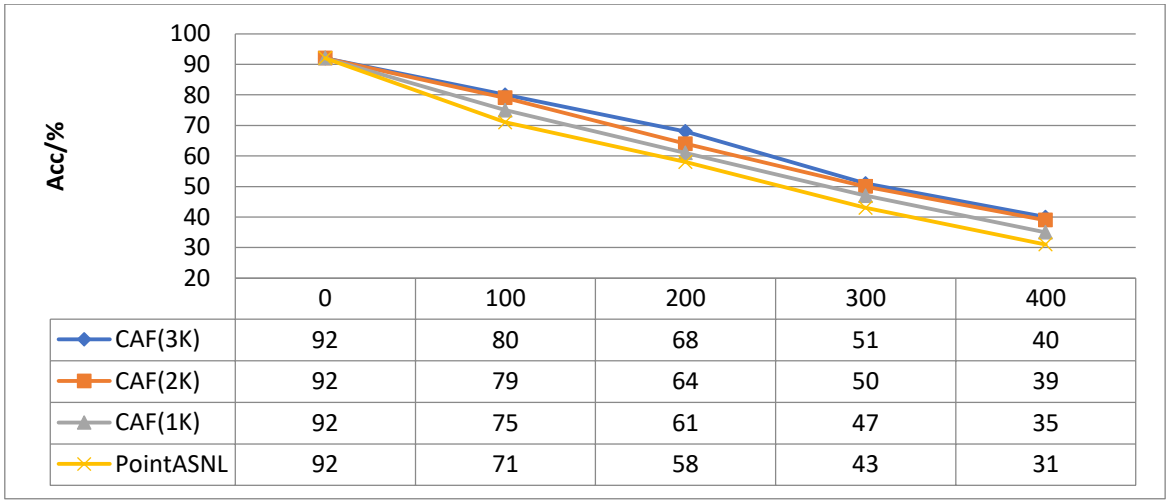

Figure 9: Impact of the CAF Module on Model Robustness

Comparative experiments: The traditional point cloud filtering methods are as follows. Straight-through filtering. (1) Set the settings based on the unique characteristics of the point cloud and filter out points outside the range determined by the coordinate axis. (2) The density of the points in the radius is filtered and the density is used to determine whether to keep or remove the point. (3) Bilateral

filtering, combining spatial structure and distance between points, applicable only to ordered point clouds. (4) Voxel filtering uses the AABB bounding box to voxelize the point cloud and reduce noise according to the voxel grid. (5) Statistical filtration, determine the distance distribution between each point and its k closest neighbors. This method is frequently used to remove outliers and other problematic points caused by measurement errors.

Use the AABB bounding box to voxelize the point cloud and reduce noise according to the voxel grid. This method is frequently used to remove outliers and other problematic points caused by misspecification. PointASNL is used as the state-of-the-art network, adding voxel filtering, statistical filtering, comprehensive filtering method, and the CAF module, respectively, to carry out comparative experiments comparing the performance of the CAF module and the conventional filtration technique in anti-noise achievement. Radius filtration is used as the initial step in the comprehensive filtering process, and statistical filtering is used to eliminate disturbance points. The unordered point cloud is used as input. In the traditional pre-processing process, statistical filtering is often used to filter such point clouds. To compare the performance of the CAF module and the traditional filtering method in anti-noise performance, PointASNL is used as the state-of-the-art network, adding voxel filtering, statistical filtering, comprehensive filtering method, and the CAF module, respectively, to carry out comparative experiments. The comprehensive filtering method is designed to first remove outliers by radius filtering and then filter out disturbance points by statistical filtering.

1) Anti-noise performance: As shown in Figure 10, voxel filtering filters out a few key points in the dense point cloud while filtering out noise, which influences the accuracy of the classification. The effect of improving the robustness of the model is small. Compared to statistical filtering, comprehensive filtering further improves the robustness of the model. Compared to traditional filtering, the CAF module has a more obvious effect in improving the model robustness. The anti-noise performance of the model is similar to the model with only the CAF module added, and the CAF module plays a leading role. From the perspective of classification accuracy, in the case of the same number of noise points, the CAF module > comprehensive filtering > statistical filtering >voxel filtering, so the introduction of the CAF module effectively improves the anti-noise performance of the network. From the point of view of the number of noise points, while maintaining the same classification accuracy, the CAF can withstand the interference of more noise points, so the robustness of the network is stronger. Table 2: shows the data for the CAF Module w.r.t. Figure 10.

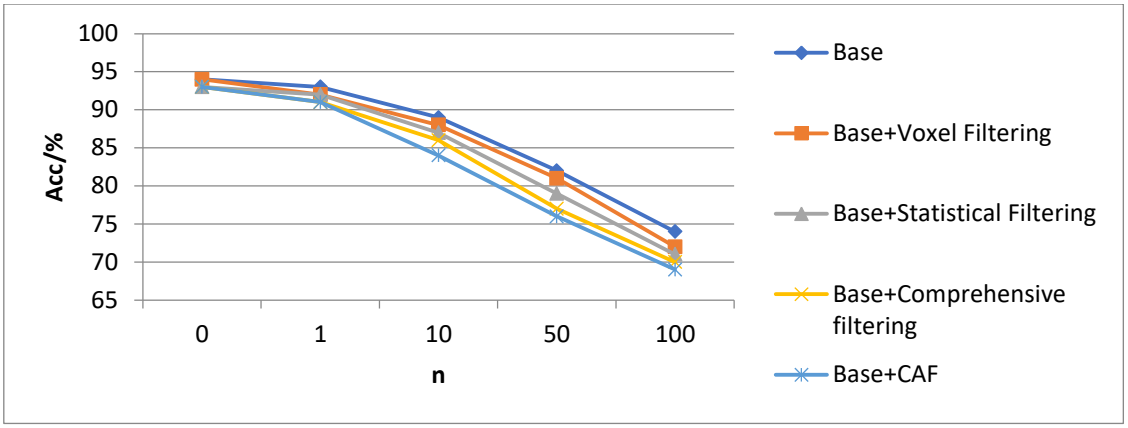

Figure 10: Comparison of CAF module and traditional filtering w.r.t. anti-interference performance

Table 2: Data for CAF Module w.r.t. Figure 10

Base	94	93	89	82	74
Base+Voxel Filtering	94	92	88	81	72
Base+Statistical Filtering	93	92	87	79	71
Base+Comprehensive filtering	93	91	86	77	70
Base+CAF	93	91	84	76	69

2) Time complexity: Traditional filtering involves preprocessing all original data separately, which requires a large amount of calculations and is separated from the network. It is an independent processing part. When processing any new input data, independent filtering processing is required. The CAF module is embedded in the network to achieve end-to-end training and testing, and achieve better anti-noise performance than traditional filtering with very little computation and lower time complexity. Taking the standard aircraft model in ModelNet40 as an example, in this model there are a total of 10000 data points, and Gaussian noise of standard normal distribution is introduced. When using the state-of-the-art model is PointASNL, to test a single aircraft model, 51.60 msvoxel filtering, statistical filtering, and comprehensive filtering (radius filtering) are used respectively. Statistical filtering to preprocess the point cloud data with added noise and input them into the network for feature extraction and classification. The test time increments for a single noisy aircraft model are54.30,60.71, and 84.24 ms. Finally, the noise data of the aircraft model without preprocessing is input into the state-of-the-art network with CAF module for testing, and the test time increment is 15.50 ms. Compared to the four, the network running time with the CAF module is much faster than that of traditional filtering. In actual use, the CAF module not only takes into account the classification accuracy, but also ensures a certain real-time performance.

4.3. Submodule Necessity

To verify the necessity of the design of two independent submodules of the CAF module, the MFSE submodule and the FICSA submodule are separately applied to the state-of-the-art network PointASNL, and the trained model is used for comparative experiments. The

experimental results are depicted in Figure 11 and Table 3. Although the network applying the MFSE sub-module maintains the classification accuracy without noise, but the ability to resist interference decreases rapidly with the increase of the number of noise points. It can be observed from the table that the network applying the FICSA submodule has a greater number of noise interference. The resistance ability is stronger than that of the state-of-the-art network, but the classification accuracy is reduced when there is no noise or a small amount of noise. The network applying two submodules at the same time, namely the CAF module, can boost the model's resistance to noise interference while maintaining the average classification accuracy and ability.

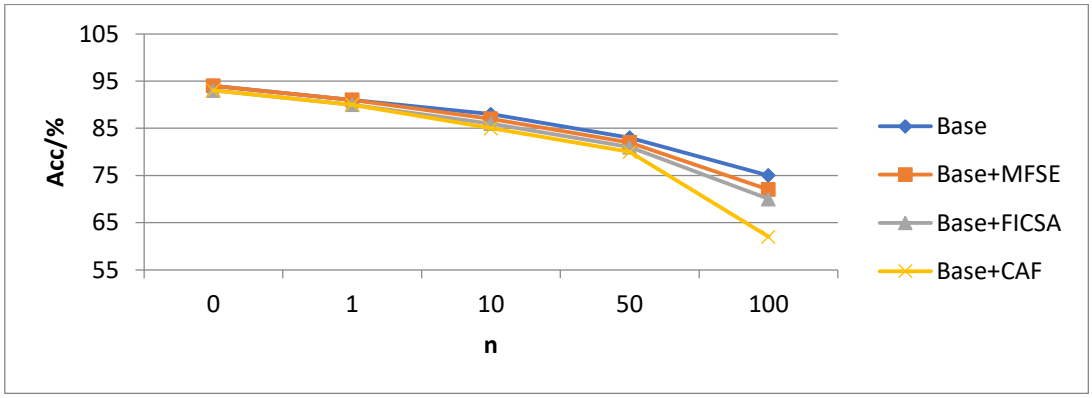

Figure 11: Influence of independent submodules on model robustness

Table 3: Influence on model robustness w.r.t. of Independent Sub-Modules

Base	94	91	88	83	75
Base+MFSE	94	91	87	82	72
Base+FICSA	93	90	86	81	70
Base+CAF	93	90	85	80	62

The MFSE submodule is designed to maintain the stability of the classification accuracy, and the FICSA submodule is designed to boost the model robustness. The two have their own advantages and disadvantages when applied independently. The CAF module retains the excellent performance of the two sub-modules to avoid their respective defects adverse effects.

4.4. Semantic Segmentation

The CAF module is applied to the 3-D point cloud segmentation network, and PointASNL is used as the state-of-the-art network to embed the CAF module in the middle of the feature extraction layer to conduct semantic segmentation experiments. Semantic part segmentation is performed on the ShapeNet Part data set [23], which includes 16 categories, 50 parts, and a total of 16881 samples. 2048 points are randomly sampled as input and the batch size is 16. The experimental results are shown in Table 4. The results show that the average intersection and union ratio mIoU in this experiment is close to the state-of-the-art network level, and the segmentation performance (intersection and union ratio IoU) in some categories is better than the state-of-the-art

model (areo, earphone, guitar, lamp, mug, pistol).

In indoor scenes Semantic scene segmentation is performed on dataset S3DIS [24], including 271 rooms obtained from 6 regions of 3 buildings; each point has a semantic label to classify it into one of 13 classes of objects. Experiments are performed on 6-fold cross-validation is utilized to compare the average intersection ratio in the region. The experimental results are shown in Table 4. The results show that the scene segmentation performance in this experiment is slightly lower than that of the state-of-the-art model, and the intersection ratio of some categories is better than the state-of-the-art model (beam, table and bookcase), where OA is the overall classification accuracy and mAcc is the average classification accuracy. Furthermore, when the CAF model is used for the intermediate redistribution of the weight of the feature channel, the performance in the semantic segmentation task is not as good as in the classification task. The semantic segmentation task is more difficult and the dataset is more complex, for semantic segmentation the research of 3D point cloud deep network is also an important research direction.

Table 4: Semantic Segmentation Performance with 6-Fold Cross-Validation in S3DIS

Method	OA	mAcc	mIoU	IoU			
				Ceiling	Floor	Table	Bookcase
PointNet	79.13	66.73	47.98	88.70	89.41	42.34	9.68
A-CNN	88.00		63.40	93.14	97.17	67.03	57.36
PointCNN	88.80	76.20	65.92	95.56	98.08	72.17	61.69
PointWeb	88.00	76.81	67.23	94.25	94.95	71.97	63.20
PointASNL	89.51	79.63	69.25	96.06	98.68	71.87	60.88
This Work	88.91	79.33	68.85	95.86	98.08	72.68	61.29

5. Conclusions

This study proposes a competitive attention fusion module, which can be migrated and embedded in different 3D point cloud classification networks, and adaptively weights the intermediate feature channel weights through the multilevel feature extrusion excitation sub-module and the feature intrinsic correlation self-attention sub-module distribution, improve the global feature extraction and expression ability of the network, and enhance the robustness of the model to noise interference. The experiment uses two state-of-the-art networks to train and test on the ModelNet40 dataset, and the results show that the addition of the competitive attention fusion module helps Maintaining and improving the accuracy of point-cloud classification, compared with the state-of-the-art network, the robustness of the model can be significantly enhanced, and it has strong anti-interference ability to different numbers of point cloud disturbances, outliers, and random noise. Compared with traditional filtering, the training complexity is reduced through end-to-end learning, and the anti-noise performance is better. The performance of the proposed module in semantic segmentation is inferior to that of classification tasks, and how to improve its impact on semantic segmentation performance needs further study.

References

[1] Authors, "A Brief Survey on 3D Semantic Segmentation of Lidar Point Cloud with Deep Learning," 2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2021, pp. 405-408, doi: 10.1109/NILES53778.2021.9600493.

[2] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu and M. Bennamoun, "Deep Learning for 3D Point Clouds: A Survey," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 12, pp. 4338-4364, 1 Dec. 2021, doi: 10.1109/TPAMI.2020.3005434.

[3] L. Zhang and Z. Zhu, "Unsupervised Feature Learning for Point Cloud Understanding by Contrasting and Clustering Using Graph Convolutional Neural Networks," 2019 International Conference on 3D Vision (3DV), 2019, pp. 395-404, doi: 10.1109/3DV.2019.00051.

[4] Z. Zhang et al., "Hierarchical Aggregated Deep Features for ALS Point Cloud Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 2, pp. 1686-1699, Feb. 2021, doi: 10.1109/TGRS.2020.2997960.

[5] Parashar, A., Parashar, A., Ding, W., Shekhawat, R. S., & Rida, I. (2023). Deep learning pipelines for recognition of gait biometrics with covariates: A comprehensive review. Artificial Intelligence Review, 1-65.

[6] S. Xu, X. Zhou, W. Ye and Q. Ye, "Classification of 3-D Point Clouds by a New Augmentation Convolutional Neural Network," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 7003405, doi: 10.1109/LGRS.2022.3141073.

[7] S. Sahebdivani, H. Arefi and M. Maboudi, "Deep Learning based Classification of Color Point Cloud for 3D Reconstruction of Interior Elements of Buildings," 2020 International Conference on Machine Vision and Image Processing (MVIP), 2020, pp. 1-6, doi: 10.1109/MVIP49855.2020.9116894.

[8] Rida, I. (2018). Feature extraction for temporal signal recognition: An overview. arXiv preprint arXiv:1812.01780.

[9] Y. Wu, S. Zhang, H. Ogai, H. Inujima and S. Tateno, "Realtime Single-Shot Refinement Neural Network With Adaptive Receptive Field for 3D Object Detection From LiDAR Point Cloud," in IEEE Sensors Journal, vol. 21, no. 21, pp. 24505-24519, 1 Nov.1, 2021, doi: 10.1109/JSEN.2021.3114345.

[10] C. -C. Lin, C. -H. Kuo and H. -T. Chiang, "CNN-Based Classification for Point Cloud Object With Bearing Angle Image," in IEEE Sensors Journal, vol. 22, no. 1, pp. 1003-1011, 1 Jan.1, 2022, doi: 10.1109/JSEN.2021.3130268.

[11] Y. Wang, C. Yue and X. Tang, "A Geometry Feature Aggregation Method for Point Cloud Classification and Segmentation," in IEEE Access, vol. 9, pp. 140504-140511, 2021, doi: 10.1109/ACCESS.2021.3119622.

[12] Rida, I., Al Maadeed, S., Jiang, X., Lunke, F., & Bensrhair, A. (2018, April), An ensemble learning method based on random subspace sampling for palmprint identification, In 2018 IEEE International conference on acoustics, speech and signal processing (ICASSP) (pp. 2047-2051). IEEE.

[13] S. Ullah, U. Qayyum and A. J. Choudhry, "Time and Memory Efficient 3D Point Cloud Classification," 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 2019, pp. 521-525, doi: 10.1109/IBCAST.2019.8667269.

[14] T. Zhang, "Spherical-GMM: A Rotation and Scale Invariant Method for Point Cloud Classification," 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), 2021, pp. 156-161, doi: 10.1109/ICHCI54629.2021.00040.

[15] R. Bao, K. Palaniappan, Y. Zhao, G. Seetharaman and W. Zeng, "GLSNet: Global and Local Streams Network for 3D Point Cloud Classification," 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), 2019, pp. 1-9, doi: 10.1109/AIPR47015.2019.9174587.

[16] Li, W., Zhou, X., Yang, C., Fan, Y., Wang, Z., & Liu, Y. (2022). Multi-objective optimization algorithm based on characteristics fusion of dynamic social networks for community discovery. Information Fusion, 79, 110-123.

[17] X. Yao, J. Guo, J. Hu and Q. Cao, "Using Deep Learning in Semantic Classification for Point Cloud Data," in IEEE Access, vol. 7, pp. 37121-37130, 2019, doi: 10.1109/ACCESS.2019.2905546.

[18] Cheng, H., Yuan, S., Li, W., Yu, X., Liu, F., Liu, X., & Bezabih, T. T. (2024), De-accumulated error collaborative learning framework for predicting Alzheimer's disease progression, Biomedical Signal Processing and Control, 89, 105767.

[19] G. Sakr, A. Berberian and P. Habib, "Comparing Deep Learning Models for Road Asset Detection and Classification in LiDAR Point Cloud," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2019, pp. 138-145, doi: 10.1109/SITIS.2019.00033.

[20] A. Bayu, A. Wibisono, H. A. Wisesa, N. S. Intizhami, W. Jatmiko and A. Gamal, "Semantic Segmentation of Lidar Point Cloud in Rural Area," 2019 IEEE International Conference on Communication, Networks and Satellite (Comnetsat), 2019, pp. 73-78, doi: 10.1109/COMNETSAT.2019.8844074.

[21] Luvembe, A. M., Li, W., Li, S., Liu, F., & Wu, X. (2024). CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection. Information Processing & Management, 61(3), 103653.

[22] S. V. Sheshappanavar and C. Kambhamettu, "Dynamic Local Geometry Capture in 3D Point Cloud Classification," 2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR), 2021, pp. 158-164, doi: 10.1109/MIPR51284.2021.00031.

[23] O. Poursaeed, T. Jiang, H. Qiao, N. Xu and V. G. Kim, "Self-Supervised Learning of Point Clouds via Orientation Estimation," 2020 International Conference on 3D Vision (3DV), 2020, pp. 1018-1028, doi: 10.1109/3DV50981.2020.00112.

[24] Gupta, A., & Awasthi, L. K. (2011). Peers-for-peers (P4P): an efficient and

reliable fault-tolerance strategy for cycle-stealing P2P applications. International Journal of Communication Networks and Distributed Systems, 6(2), 202-228.

[25] Khan, R., Shabaz, M., Hussain, S., Ahmad, F., & Mishra, P. (2022). Early flood detection and rescue using bioinformatic devices, internet of things (IOT) and Android application. World Journal of Engineering, 19(2), 204-215.

[26] F. Azizmalayeri, S. M. M. Peyghambarzadeh, H. Khotanlou and A. Salarpour, "Kernel Correlation Based CNN for Point Cloud Classification Task," 2018 8th International Conference on Computer and Knowledge Engineering (ICCKE), 2018, pp. 200-204, doi: 10.1109/ICCKE.2018.8566273.

[27] M. Zhang, Y. Wang, P. Kadam, S. Liu and C. . -C. Jay Kuo, "Pointhop++: A Lightweight Learning Model on Point Sets for 3D Classification," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 3319-3323, doi: 10.1109/ICIP40778.2020.9190740.

[28] Khan, R., Dhingra, N., & Bhati, N. (2022). Role of artificial intelligence in agriculture: A comparative study. In Transforming Management with AI, Big-Data, and IoT (pp. 73-83). Cham: Springer International Publishing.

[29] Y. Li and G. Baciu, "Local Learning in Point Clouds based on Spectral Pooling," 2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), 2020, pp. 84-91, doi: 10.1109/ICCICC50026.2020.9450222.

[30] M. A. Uy, Q. -H. Pham, B. -S. Hua, T. Nguyen and S. -K. Yeung, "Revisiting Point Cloud Classification: A New State-of-the-art Dataset and Classification Model on Real-World Data," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1588-1597, doi: 10.1109/ICCV.2019.00167.

[31] F. Gomez-Donoso, F. Escalona, S. Orts-Escolano, A. Garcia-Garcia, J. Garcia-Rodriguez and M. Cazorla, "3DSliceLeNet: Recognizing 3D Objects Using a Slice-Representation," in IEEE Access, vol. 10, pp. 15378-15392, 2022, doi: 10.1109/ACCESS.2022.3148387.

[32] A. Paigwar, O. Erkent, C. Wolf and C. Laugier, "Attentional PointNet for 3D-Object Detection in Point Clouds," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 1297-1306, doi: 10.1109/CVPRW.2019.00169.