International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 4 ǁ April 2023 ǁ PP. 426-432

Research on 3D point cloud classification method based on deep learning

Xiaolin Wang[1], Gang Chen[1], Haoran Zhou[1]
*[*1] University of Shanghai for Science and Technology, Shanghai, China*
Corresponding Author: Xiaolin Wang

Abstract
As an important 3D geometric data, point cloud is widely used and has great research prospects. As a key link in 3D data processing, point cloud classification has a wide range of applications in the fields of 3D reconstruction, 3D object detection and cultural relic protection. This paper reviews the research progress of 3D point cloud classification methods based on deep learning techniques. Classify the point cloud classification methods according to the form of input data through literature analysis, and summarize the basic ideas, advantages and disadvantages of various methods. Finally, the challenges and prospects of point cloud classification are prospected.
Keywords: *point cloud classification deep learning, deep learning, convolutional neural network.*

Date of Submission: 06-04-2023 Date of acceptance: 18-04-2023

I. INTRODUCTION

In recent years, with the rapid development of various 3D scanning technologies, the acquisition of point cloud data of objects has become faster, more accurate and more convenient, which has promoted the continuous expansion of its application fields and scenarios. Among them, 3D point cloud classification is an object recognition, The basic work of various processing tasks such as segmentation and matching. At present, 3D point cloud classification methods have been widely used in fields such as augmented reality [1], reverse engineering, autonomous driving [2] and cultural relics protection [3]. Therefore, point cloud classification has become a research hotspot in the field of 3D computer vision. However, there are still huge challenges in point cloud classification due to the unordered, unstructured, high-dimensionality of point cloud data, as well as factors such as noise in the measurement process, complex scenes, and occlusions.

The point cloud classification task is generally divided into two steps, extracting local and global representative feature information from a large amount of point cloud data, and then using the learned feature information to classify each point into a predefined semantic category. The early traditional point cloud classification methods mainly use hand-designed 3D feature descriptors and machine learning-based classifiers for classification and recognition. Common classifiers include Gaussian mixture model [4,5], support vector machine [6,7], random forest [8,9], conditional random field [10] and Markov random field [11,12]. Due to the weak representation ability of the feature information extracted by manual design, the correlation between local point clouds and contexts has not been taken into account. Traditional 3D point cloud classification methods generally have problems of insufficient robustness and accuracy bottlenecks. With the formal definition of deep learning [13] in the journal Nature in 2015, researchers have proposed many methods for point cloud classification related tasks using deep learning technology.

At present, there are some review articles summarizing the point cloud classification methods based on deep learning. This paper adds the latest related methods proposed on the basis of predecessors, and summarizes the working principles and key technologies of classic and cutting-edge point cloud classification methods. Finally, through the elaboration of the research status of related methods, the future development direction of point cloud classification technology based on deep learning is discussed.

II. CLASSIFICATION METHOD BASED ON DEEP LEARNING

The 3D point cloud classification method based on deep learning technology can be divided into three types according to the representation of the input data: voxel-based, multi-view based and point cloud-based classification methods directly. And select a classic representative network structure for display.

2.1 POINT CLOUD DATA SET AND POINT CLOUD CLASSIFICATION EVALUATION INDEX

2.1.1 Related classification datasets

ModelNet [14] was proposed in the Princeton Laboratory in 2015. All the models in the dataset are regular and clean, and the structure of the dataset is relatively good. Therefore, ModelNet is widely used in tasks such as point cloud model retrieval and classification. The ModelNet dataset contains two types, ModelNet40 and ModelNet10. The ModelNet40 dataset contains 3D CAD models of 40 common grid forms, consisting of 9843 training models and 2468 test models. The ModelNet10 dataset contains 10 categories of 3D CAD models, including 3991 training models and 908 test models.

The ScanObjectNN [15] dataset is an indoor scene object created using CAD data. It is a more powerful real-world object dataset, which contains about 15,000 actual scanned objects in 15 categories. Although it has fewer categories than the ModelNet40 dataset, ScanObjectNN is more practically challenging than artificially synthesized datasets due to complex real scenes, noise occlusion, and partial absence of 3D objects.

2.1.2 Evaluation indicators

In order to more intuitively and fully compare the effectiveness of the performance of various methods, it is necessary to have a unified evaluation standard. For point cloud classification tasks, this paper selects the two most important classification evaluation indicators of overall accuracy OA and average accuracy AA to analyze and compare different point cloud classification methods.

2.2 VOXEL-BASED METHODS

The voxel-based point cloud classification method converts irregular point cloud data into regular raster data to complete the classification task. The VoxNet network proposed by Maturana et al. [16] was the first to apply the idea of point cloud voxelization to point cloud classification tasks. This method uses the normalized value of the grid unit as the network convolution layer The input, and then use the 3D neural network to extract features, and use the maximum pooling operation for non-overlapping voxel blocks. Wu et al. at Princeton University proposed 3D ShapeNets [14] networks using probabilistic joint distributions, which represent 3D point cloud geometric features as binary probability distributions on a 3D voxel grid.Both VoxNet and 3DShapeNet networks suffer from large memory consumption and low accuracy.In order to solve the sparsity of point clouds and reduce computational overhead, Riegler et al. [17] proposed OctNet using a flexible octree structure instead of a voxel grid. The OctNet network uses a series of octree structures to divide the 3D space hierarchically. Each leaf node of the octree stores the corresponding pooling features. This method greatly reduces the amount of computation and memory consumption. Inspired by OctNet, Wang et al. [18] of Tianjin University proposed OCNN, which uses 3D convolutional neural network to extract the sparse and discrete normal vector features contained in the octree. This model saves computational overhead and computational memory to a certain extent.because the local geometric structure information is not fully utilized, the classification accuracy is not high.

Wang et al. [19] proposed a multi-scale convolutional network MSNet based on the multi-scale voxelization technology, which divides the point cloud space into voxels of multiple scales to extract local features of different spatial resolutions, and at the same time obtained according to the constructed CRF model Point cloud space context information, and finally predict the category probability of point cloud according to the fusion of local features and global context information. The MSNet model retains more feature information, which improves the classification accuracy. Voxel representation point clouds are inherently sparse. Therefore, using CNN to extract spatially sparse voxel data will lead to long training time and often inefficient. Le et al. [20] mixed point and grid data into a network model PointGrid for 3D shape understanding, sampling a constant number of points in each voxel grid unit, enabling the network framework to use 3D convolutions to Extract local geometric detail features. Through experimental comparison, the network model shows superior classification performance.

The voxel-based approach mainly transforms point cloud data into voxel grid data, which can clearly encode the unstructured point cloud spatial information. The above-mentioned methods effectively solve the difficulties encountered in the voxelization process, but the classification accuracy is lost due to the blurring of point cloud boundaries, the loss of feature information caused by the voxelization process, and the complexity of storage and computation overhead.

2.3 MULTI-VIEW BASED METHODS

The most representative multi-view-based method is the MVCNN network proposed by su et al [21], the essence of which is to render 3D objects into multiple views and then use the well-established convolutional network framework to identify and classify 3D objects. This is specifically shown in flowchart 1. The method does not fully utilize the feature relationships among multiple views, which to some extent limits the discrimination of the final aggregated feature descriptors, thus making the classification accuracy not so ideal.

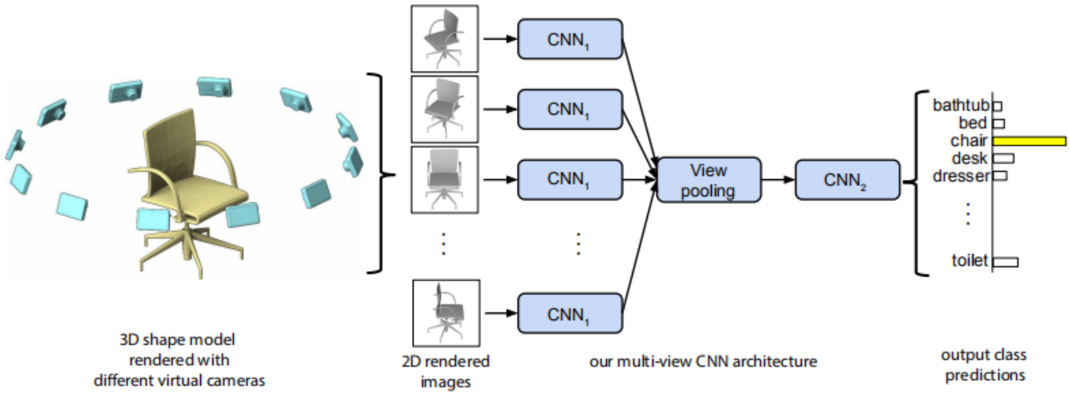

Figure 1: Network framework diagram of MVCNN

Bai et al [22] improved on the basis of MVCNN and proposed the Multi-loop View Convolutional Neural Network (MLVCNN), which extracts multiple views from multiple loop directions and multiple scales for 3D data, and extracts feature descriptors from the views as the basis for classification, and this method significantly improves the performance of 3D object classification. In order to extract more distinguishing features, Feng et al [23] proposed the GVCNN model through the idea of visual graph-group-shape, firstly, FCN is used to obtain view-level descriptors, then the grouping module groups and generates group-level descriptors according to the correlation and distinction between views, and the group-level descriptors are weighted and fused to generate shape descriptors for the point cloud classification task. The schematic diagram of GVCNN is shown in Figure 2. The proposed GVCNN method will make more use of view information to find more discriminative information, and the performance of 3D point cloud classification tasks has been significantly improved. In order to make full use of the correlation between views, Ma et al. [24] combined a convolutional neural network (CNN) with a long-term short-term memory (LSTM) according to the sequence between views. CNN is used to extract multiple view components. The view sequence features, LSTM and sequence voting layers aggregate these features into shape descriptors. This method effectively utilizes the advantages of CNN and LSTM. In the same year, the RotationNet [25] network used the pose label of the view as a potential variable on the basis of MVCNN, and used the unaligned data set to perform unsupervised self-alignment optimization on the pose of each view during the training process. In addition, literature methods [26-28] also have class contributions in the multi-view based research direction.

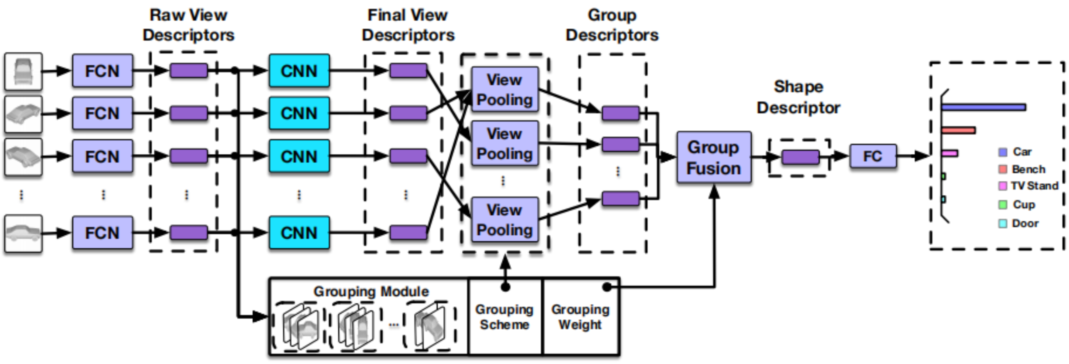

Figure 2: Network framework diagram of GVCNN

The method based on multi-view has achieved good results in point cloud classification tasks, but this type of algorithm needs to pre-set the selection of view angles and numbers, and different selection methods will have a great impact on the classification results. In addition, in the process of rendering point cloud data into multiple view representations, some geometric feature information will be lost, and computational overhead such as pre-training will also be increased.

2.4 POINT CLOUD BASED METHODS

The above voxel-based and multi-view point cloud classification methods have high computational complexity and cannot make full use of the attributes of the original point cloud data. For this reason, researchers have tried and proposed a variety of classification methods directly based on point clouds. Among them, PointNet [29] is a milestone for directly processing point clouds, which uses MLP to learn the features of each point independently, and aggregates the features of the learned points into global features through the maximum pooling layer (as shown in Figure 3). However, PointNet ignores the local structure caused by the metric. Based on this problem, Qi et al. proposed the PointNet++[30] network, which achieves fine local geometric structure feature extraction through the introduction of a layered neural network, and classifies small domain features into large Units extract higher-level feature information, but this method lacks point-to-point correlations in different regions. Subsequent research focuses on better integrating the contextual information of the local area of the point cloud from the three perspectives of convolutional neural network, graph convolutional neural network and attention mechanism.

The most representative method based on convolutional neural network is PointCNN proposed by Li et al. [31]. The perceptron uses the χ transformation after extracting the point cloud features. The experimental results show that the X-Conv structure in PointCNN performs better than the maximum pooling operation in the PointNet network. Subsequently, Liu et al. [32] proposed the RS-CNN model by learning 3D objects through geometric relationships. This model can use RS-Conv to map the low-dimensional feature information of the point cloud into high-dimensional information, and can extract more space. Contextual feature information enables high-precision point cloud classification tasks.

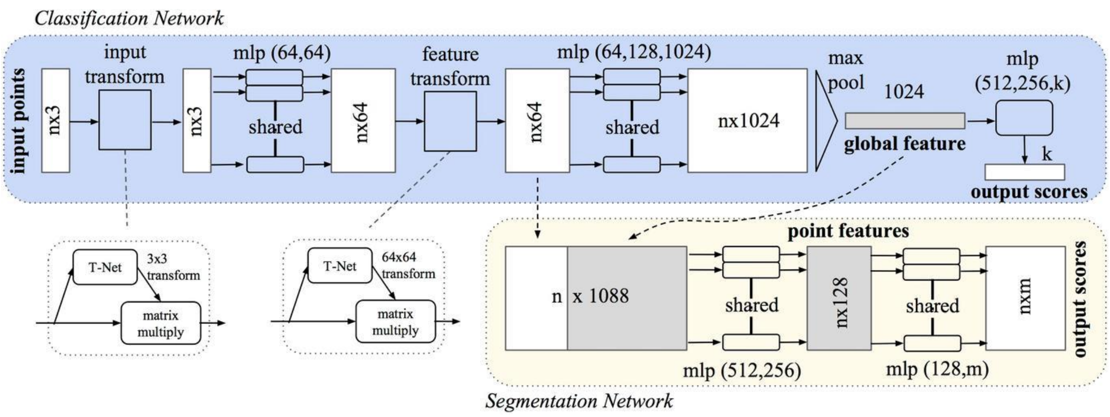

Figure 3: Network framework diagram of PointNet

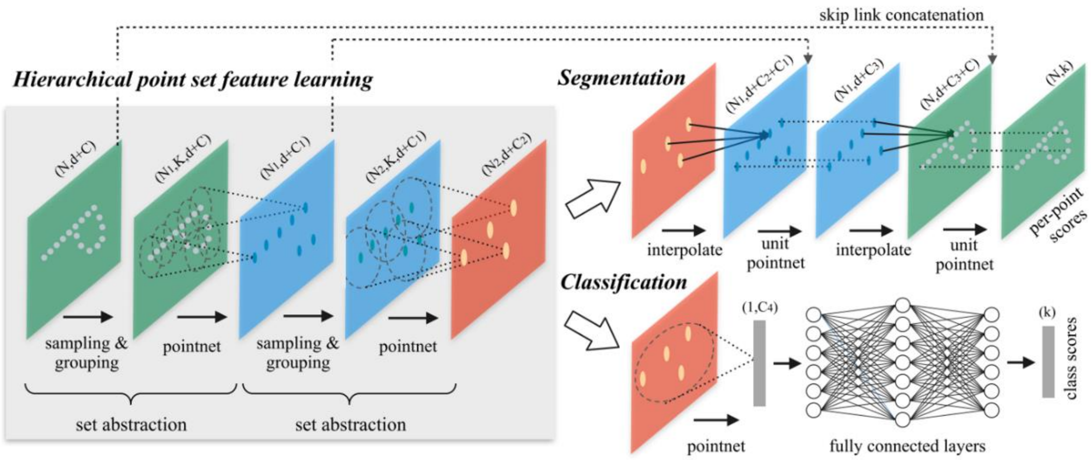

Figure 4: Network framework diagram of PointNet++

PointConv [33] builds a deep convolutional network and uses mlp to learn the weight function, mainly considering the convolution kernel as a nonlinear function of the local coordinates of the three-dimensional point cloud composed of weight and density functions. This network effectively improves the accuracy of point cloud classification while reducing memory usage. ConvPoint [34] can generalize CNN. The network uses continuous convolution instead of discrete convolution to process point cloud data. ConvPoint is not only flexible but also reaches an advanced level in classification tasks. PAConv [35] is a position-adaptive convolution, which mainly uses a dynamically assembled basic weight matrix to construct a learning convolution kernel, in which the parameters of the weight matrix are learned through ScoreNet self-adaptation. Embedding PAConv into the PointNet network can get better point cloud classification results.

In recent years, deep learning has been widely used in the field of graph neural network, among which the concept of graph neural network was first proposed by S et al. [36]. On this basis, Bruna et al. [37] proposed a convolutional neural network based on graph modeling for non-Euclidean domains. Kipf et al. [38] formally proposed graph convolutional neural networks, which achieved promising results in semi-supervised classification tasks. In order to improve the effectiveness of point cloud descriptors, Dynamic Graph CNN (DGCNN) [39] uses the constructed EdgeConv layer to capture local geometric structure features, which can dynamically update and generate edge features related to embedded points and neighboring points. However, DGCNN mainly focuses on the coordinate distance between points and ignores the vector direction between adjacent points, resulting in the loss of some local geometric feature information. Zhang et al. [40] proposed Linked Dynamic Graph Convolutional Neural Network (LDGCNN) inspired by DGCNN. It can aggregate the dynamic graph features of different layers, learn more edge vector feature information, and avoid the problem of gradient disappearance to a certain extent. Grid-GCN [41] can quickly scale point cloud learning, reduce theoretical time complexity and improve spatial coverage. Zhou et al. [42] proposed adaptive graph convolution AdaptConv. The AdaptConv network can adaptively adjust the dynamically learned point cloud features, which greatly improves the flexibility of graph convolution and can effectively capture different semantic parts. point cloud relationship.

The working principle of the attention mechanism is to make the system focus on the main information and ignore the secondary information. The transform composed of self-attention mechanism has achieved good results in 2D image classification, natural language processing and other fields. Therefore, many scholars have introduced transform into multiple tasks of point cloud processing.The most representative one is the high-precision PCT [43] network proposed by Tsinghua scholars. The specific framework diagram is shown in Figure 4. This network maps point cloud data to a higher-dimensional feature space, and then uses offset- The attention and normalization mechanisms are used to learn the geometric features of the local point cloud and obtain the semantic similarity at different scales, and finally aggregate the local and global feature information to realize the classification task of the point cloud. In addition, researchers have proposed a more powerful architecture based on transform. For example JSENet[44], Point Transformer[45], etc. In these networks, Point Transformer designs Point Transformer layers for point clouds, and builds networks for tasks such as point cloud classification by stacking them.

The classification method that directly processes the point cloud is better than the classification method that preprocesses the point cloud structure. Because the process of converting point clouds into voxels and multi-views will lead to the loss of feature information, resulting in a decline in point cloud classification performance. In addition, the convolutional neural network and graph convolution methods in this type of method require a large calculation and storage overhead when implementing classification tasks, and then need to be optimized in terms of calculation and memory usage.

III. CONCLUSION

This paper presents a detailed overview of classification methods for 3D point clouds based on deep learning techniques. First, some commonly used data sets and evaluation indicators related to point cloud classification are introduced; then, according to different data representations of point clouds, point cloud classification methods are divided into voxel-based, multi-view and direct point cloud methods. Each method has advantages and disadvantages. Voxel-based and multi-view methods have strong practicability in some specific scenes, but the process of converting unstructured point cloud data into multiple views and voxelization will lose a lot of effective depth information. Although the method of directly processing the original point cloud requires a large amount of calculation in the convolution process, it can greatly improve the accuracy of point cloud classification. Although deep learning technology has achieved excellent results by automatically extracting feature information. However, both the voxel-based multi-view and the original point cloud-based methods have further optimization problems. For example: how the network model can achieve high-precision and high-efficiency classification at the same time; how to fully and flexibly utilize the correlation between points, etc. The research on the application of deep learning technology in point cloud processing is constantly being promoted, and more innovative and effective methods are expected on this basis.

REFERENCES

[1]. Adikari, S. B., Ganegoda, N. C., Meegama, R. G., and Wanniarachchi, I. L. (2020). Applicability of a single depth sensor in real-time 3d clothes simulation: augmented reality virtual dressing room using kinect sensor. Adv. Hum. Comput. Interact. 2020, 1314598. doi: 10.1155/2020/1314598

[2]. Yu, D., Ji, S., Liu, J., and Wei, S. (2021). Automatic 3D building reconstruction from multi-view aerial images with deep learning. ISPRS J. Photogram. Remote Sens. 171, 155–170. doi: 10.1016/j.isprsjprs.2020.11.011

[3]. Pan Y, Dong Y, Wang D, et al. Three-dimensional reconstruction of structural surface model of heritage bridges using UAV-based photogrammetric point clouds[J]. Remote Sensing, 2019, 11(10): 1204.

[4]. Lalonde J F, Unnikrishnan R, Vandapel N, et al. Scale selection for classification of point-sampled 3D surfaces[C]//Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05). IEEE, 2005: 285-292.

[5]. Lalonde J F, Vandapel N, Huber D F, et al. Natural terrain classification using three‐dimensional ladar data for ground robot mobility[J]. Journal of field robotics, 2006, 23(10): 839-861.

[6]. Gao Z, Liu X. Support vector machine and object-oriented classification for urban impervious surface extraction from satellite imagery[C]//2014 The Third International Conference on Agro-Geoinformatics. IEEE, 2014: 1-5.

[7]. Colgan M S, Baldeck C A, Féret J B, et al. Mapping savanna tree species at ecosystem scales using support vector machine classification and BRDF correction on airborne hyperspectral and LiDAR data[J]. Remote Sensing, 2012, 4(11): 3462-3480.

[8]. Gan Z, Zhong L, Li Y, et al. A random forest based method for urban object classification using LiDAR data and aerial imagery[C]//2015 23rd International Conference on Geoinformatics. IEEE, 2015: 1-4.

[9]. Wang C, Shu Q, Wang X, et al. A random forest classifier based on pixel comparison features for urban LiDAR data[J]. ISPRS journal of photogrammetry and remote sensing, 2019, 148: 75-86.

[10]. Niemeyer J, Rottensteiner F, Soergel U. Contextual classification of lidar data and building object detection in urban areas[J]. ISPRS journal of photogrammetry and remote sensing, 2014, 87: 152-165.

[11]. Munoz D, Bagnell J A, Vandapel N, et al. Contextual classification with functional max-margin markov networks[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 975-982.

[12]. Shapovalov R, Velizhev A, Barinova O. Non-associative markov networks for point cloud classification[C]//Proceeding of the ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis, Paris, France. 2010: 1-3.

[13]. LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[14]. Wu Z, Song S, Khosla A, et al. 3d shapenets: A deep representation for volumetric shapes[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1912-1920.

[15]. Uy M A, Pham Q H, Hua B S, et al. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 1588-1597.

[16]. Maturana D, Scherer S. Voxnet: A 3d convolutional neural network for real-time object recognition[C]//2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2015: 922-928.

[17]. Riegler G, Osman Ulusoy A, Geiger A. Octnet: Learning deep 3d representations at high resolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3577-3586.

[18]. Yu T, Meng J, Yuan J. Multi-view harmonized bilinear network for 3d object recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 186-194.

[19]. Wang L, Huang Y, Shan J, et al. MSNet: Multi-scale convolutional network for point cloud classification[J]. Remote Sensing, 2018, 10(4): 612.

[20]. Le T, Duan Y. Pointgrid: A deep network for 3d shape understanding[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 9204-9214.

[21]. Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3d shape recognition[C]//Proceedings of the IEEE international conference on computer vision. 2015: 945-953.

[22]. Jiang J, Bao D, Chen Z, et al. MLVCNN: Multi-loop-view convolutional neural network for 3D shape retrieval[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 8513-8520.

[23]. Feng Y, Zhang Z, Zhao X, et al. Gvcnn: Group-view convolutional neural networks for 3d shape recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 264-272.

[24]. Ma C, Guo Y, Yang J, et al. Learning multi-view representation with LSTM for 3-D shape recognition and retrieval[J]. IEEE Transactions on Multimedia, 2018, 21(5): 1169-1182.

[25]. Kanezaki A, Matsushita Y, Nishida Y. Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5010-5019.

[26]. Yu T, Meng J, Yuan J. Multi-view harmonized bilinear network for 3d object recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 186-194.

[27]. Yang Z, Wang L. Learning relationships for multi-view 3D object recognition[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 7505-7514.

[28]. Wei X, Yu R, Sun J. View-gcn: View-based graph convolutional network for 3d shape analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1850-1859.

[29]. Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652-660.

[30]. Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.

[31]. Li Y, Bu R, Sun M, et al. Pointcnn: Convolution on x-transformed points[J]. Advances in neural information processing systems, 2018, 31.

[32]. Liu Y, Fan B, Xiang S, et al. Relation-shape convolutional neural network for point cloud analysis[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 8895-8904.

[33]. Wu W, Qi Z, Fuxin L. Pointconv: Deep convolutional networks on 3d point clouds[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. 2019: 9621-9630.

[34]. Boulch A. ConvPoint: Continuous convolutions for point cloud processing[J]. Computers & Graphics, 2020, 88: 24-34.

[35]. Xu M, Ding R, Zhao H, et al. Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 3173-3182.

[36]. Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15. Springer International Publishing, 2018: 593-607.

[37]. Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[J]. arXiv preprint arXiv:1312.6203, 2013.

[38]. Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.

[39]. Wang Y, Sun Y, Liu Z, et al. Dynamic graph cnn for learning on point clouds[J]. Acm Transactions On Graphics (tog), 2019, 38(5): 1-12.

[40]. Zhang K, Hao M, Wang J, et al. Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features[J]. arXiv preprint arXiv:1904.10014, 2019.

[41]. Xu Q, Sun X, Wu C Y, et al. Grid-gcn for fast and scalable point cloud learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 5661-5670.
[42]. Zhou H, Feng Y, Fang M, et al. Adaptive graph convolution for point cloud analysis[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 4965-4974.
[43]. Guo M H, Cai J X, Liu Z N, et al. Pct: Point cloud transformer[J]. Computational Visual Media, 2021, 7: 187-199.
[44]. Hu Z, Zhen M, Bai X, et al. Jsenet: Joint semantic segmentation and edge detection network for 3d point clouds[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16. Springer International Publishing, 2020: 222-239.
[45]. Zhao H, Jiang L, Jia J, et al. Point transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 16259-16268.