



Review

Deep learning-based 3D point cloud classification: A systematic survey and outlook[☆]

Huang Zhang^a, Changshuo Wang^{b,c,d,e,1}, Shengwei Tian^{a,*}, Baoli Lu^{b,d,f}, Liping Zhang^{b,d,e}, Xin Ning^{b,e}, Xiao Bai^g

^a School of Software, Xinjiang University, Xinjiang 830000, China

^b Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China

^c Center of Materials Science and Optoelectronics Engineering & School of Microelectronics, University of Chinese Academy of Sciences, Beijing 100049, China

^d Beijing Key Laboratory of Semiconductor Neural Network Intelligent Sensing and Computing Technology, Beijing 100083, China

^e Cognitive Computing Technology Joint Laboratory, Wave Group, Beijing 102208, China

^f School of Microelectronics, University of Chinese Academy of Sciences, Beijing 100049, China

^g School of Computer Science and Engineering, State Key Laboratory of Software Development Environment, Jiangxi Research Institute, Beijing 100191, China

ARTICLE INFO

Keywords:

Deep learning

Point cloud

3D data

Classification

ABSTRACT

In recent years, point cloud representation has become one of the research hotspots in the field of computer vision, and has been widely used in many fields, such as autonomous driving, virtual reality, robotics, etc. Although deep learning techniques have achieved great success in processing regular structured 2D grid image data, there are still great challenges in processing irregular, unstructured point cloud data. Point cloud classification is the basis of point cloud analysis, and many deep learning-based methods have been widely used in this task. Therefore, the purpose of this paper is to provide researchers in this field with the latest research progress and future trends. First, we introduce point cloud acquisition, characteristics, and challenges. Second, we review 3D data representations, storage formats, and commonly used datasets for point cloud classification. We then summarize deep learning-based methods for point cloud classification and complement recent research work. Next, we compare and analyze the performance of the main methods. Finally, we discuss some challenges and future directions for point cloud classification.

1. Introduction

In recent computer vision field, the processing technology of two-dimensional images is close to maturity, and many researchers have shifted their research focus to three-dimensional scenes that are more in line with the real world [1,2]. In a 3D scene, point cloud plays an important role in representing the 3D scene because of its rich expression information. Therefore, point cloud has become a common form of data expression in the study of 3D vision. As technology advances, the acquisition of point cloud data is becoming increasingly intelligent and convenient, and there are many acquisition methods, such as: LIDAR laser detection, point cloud acquisition through 3D model calculation, point cloud acquisition through 3D reconstruction through 2D images [3], etc.

As the most basic point cloud analysis task, point cloud classification has been widely used in many fields such as security detection [4,5], target object detection [6,7], medicine [8,9], and three-dimensional

reconstruction [10,11]. The purpose of point cloud classification is to equip each point in the point cloud with a marker to identify the overall or part properties of the point cloud. Since the component attributes of point clouds belong to the category of point cloud segmentation, in this paper, we mainly focus on the overall attributes of point clouds, namely point cloud classification.

As shown in Fig. 1, 3D data comes in a variety of representations. Currently, it is possible to convert point clouds into mesh, voxel, or multi-view data to learn 3D object representation through indirect methods, but these methods are prone to problems such as loss of 3D geometric information of objects or excessive memory consumption. Before PointNet, due to the disorder and irregularity of point cloud, deep learning technology cannot directly process point cloud. Early point cloud processing used hand-designed rules for feature extraction, and then used machine learning-based classifiers (such as Support Vector Machines (SVM) [12], AdaBoost [13], Random Forest (RF) [14],

[☆] This paper was recommended for publication by Dr Shibiao Xu.

* Corresponding author.

E-mail address: 357348035@qq.com (S. Tian).

¹ The author contributed equally to this work and should be considered co-first author.

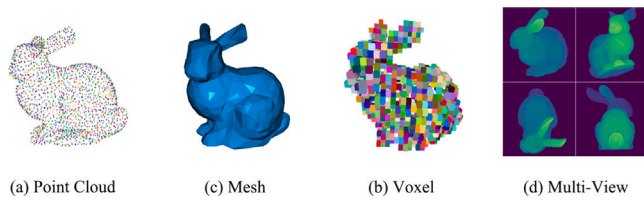


Fig. 1. 3D data representation.

etc.) to predict the class label of the point cloud, but these Class methods have poor adaptive ability and are prone to noise [15]. Some researches solved the noise problem by synthesizing context information, such as Conditional Random Field (CRF) [16], Markov Random Field (MRF) [17], etc., which improved the classification performance to a certain extent. However, the feature expression ability of hand-designed rule extraction is limited, especially in complex scenes, the accuracy and generalization ability of the model cannot meet the requirements of human beings, and this method relies heavily on researchers with professional knowledge and experience. With the rapid development of computer computing and data processing capabilities, the application of deep learning technology in point cloud analysis has also been promoted. The paper published by Charles et al. [18] of Stanford University in 2017 proposes a deep learning network, PointNet, that directly processes point clouds. This paper is a landmark, and methods of directly processing point clouds gradually dominate and become the future research trend. However, the sparsity and disorder of point cloud data make it difficult to obtain local or global features of the point cloud. The method of directly processing the original point cloud cannot be limited to the classification mechanism of convolution, graph, or attention mechanism in the past. Therefore, we explored the local feature and global feature mechanism based on the original point cloud, and made a more specific analysis of the point-based method, focusing on the latest research methods in recent years, which covers the hot research transformer in this field. (Table 1 shows a complete description of abbreviations).

In the past few years, point cloud processing in computer vision has been achieved through traditional manual feature extraction and classification. Point cloud processing based on deep learning has been a popular research direction in recent years. [19–21] summarized the semantic/part segmentation, object detection, and point cloud completion of point clouds based on deep learning. However, there is currently little systematic introduction to the latest developments in point cloud classification tasks. Our work differs from previous related work in several aspects: the form of point cloud data, the perspective of point cloud classification, and the differentiation of research results. This article extends the previous work and summarizes the latest progress in point cloud classification research in recent years into four categories (see Section 4 for details). For the distinction between local and non-local point cloud processing, we divide it into local feature aggregation (including 4 types) and non-local feature aggregation (including 3 types).

It is worth emphasizing that, although the number of 3D data processing research has grown exponentially in recent years, few surveys can be consistent with the current research on 3D point cloud classification, and clearly define the membership relationship of point cloud classification methods. It lacks a systematic summary of the whole process of point cloud classification, such as 3D form representation and storage methods. Previous research focused on manual extraction of features, voxels and multi view methods. After long-term review and research, we have attempted to propose solutions to these issues. This article investigates research publications as of July 2022, including a series of methods currently popular in this field. Based on the characteristics and implementation process of each point cloud classification method, this article elucidates the key points of different

classification strategies. The overall structure of the article is shown in Fig. 2.

Specifically, the main contributions of our work are as follows:

- We first give a detailed introduction to the 3D data and make a deeper interpretation of the point cloud for the reader's understanding, and then give the datasets used for point cloud classification and their acquisition methods.
- We summarize recently published research on point cloud classification reviews, building on which to complement state-of-the-art research methods. These methods are classified into four categories according to their characteristics, including multi-view-based, voxel-based, point-cloud-based methods, and polymorphic fusion-based methods. And then the point-cloud-based methods are subdivided.
- We discuss the advantages and limitations of subcategories of methods based on their classification. This classification is more suitable for researchers to explore these methods on actual needs.
- We give the evaluation metrics and performance comparisons of the methods to better demonstrate the performance of various methods on the dataset, and then analyze some current challenges and future trends in this field.

2. Research methodology

This literature review's survey and research methods are from journals, conferences, and seminar papers of several well-known literature databases [22], including Web of Science, IEEE Xplore, Science Direct, Google Scholar, etc. Searching for keywords related to contributions can initially generate relevant contributions of researchers, including "Point cloud", "Deep learning", "Classification" and "3D". Top conference publications such as International Conference on Computer Vision (ICCV), Computer Vision and Pattern Recognition (CVPR), and European Conference on Computer Vision (ECCV) and high-quality journals such as IEEE Transactions and Pattern Recognition are the primary search objects. Individual searches include point cloud classification datasets, point cloud data formats, and application contributions. Combined with published reviews and periodical conferences, the classification of methods is formulated to emphasize our contributions.

This research is not a negation of the research methods summarized by the predecessors, but rather our understanding of this research [23–25], such as the representation of 3D data, the storage format of point clouds, and the methods based on local features (point-by-point method, convolution, graph, attention mechanism), global feature aggregation (Transformer, global module, RNN and LSTM) and polymorphism fusion not mentioned in the previous research.

Due to the fact that point clouds have multiple datasets and varying degrees of complexity, there are limitations to the investigation process of experimental results using different methodologies. Therefore, we searched for results from comparative experiments in multiple papers to supplement the performance results of each method in more datasets.

3. 3D data

3.1. 3D data representation

There are various representations of 3D data, such as point cloud, mesh, and voxel. Here we introduce them.

Point cloud: A point cloud is essentially a large collection of tiny points drawn in 3D space, as shown in Fig. 1(a), which consists of a large collection of points captured using a 3D laser scanner. These points can express the spatial distribution and surface characteristics of the target. Each point in the point cloud contains rich information, such as: three-dimensional coordinates (x, y, z), color information (r, g, b) and surface normal vector, etc.

Mesh: 3D data can also be represented by a mesh grid, which can be viewed as a collection of points that build local relationships between points. Triangular mesh, also known as triangular patch (as shown in

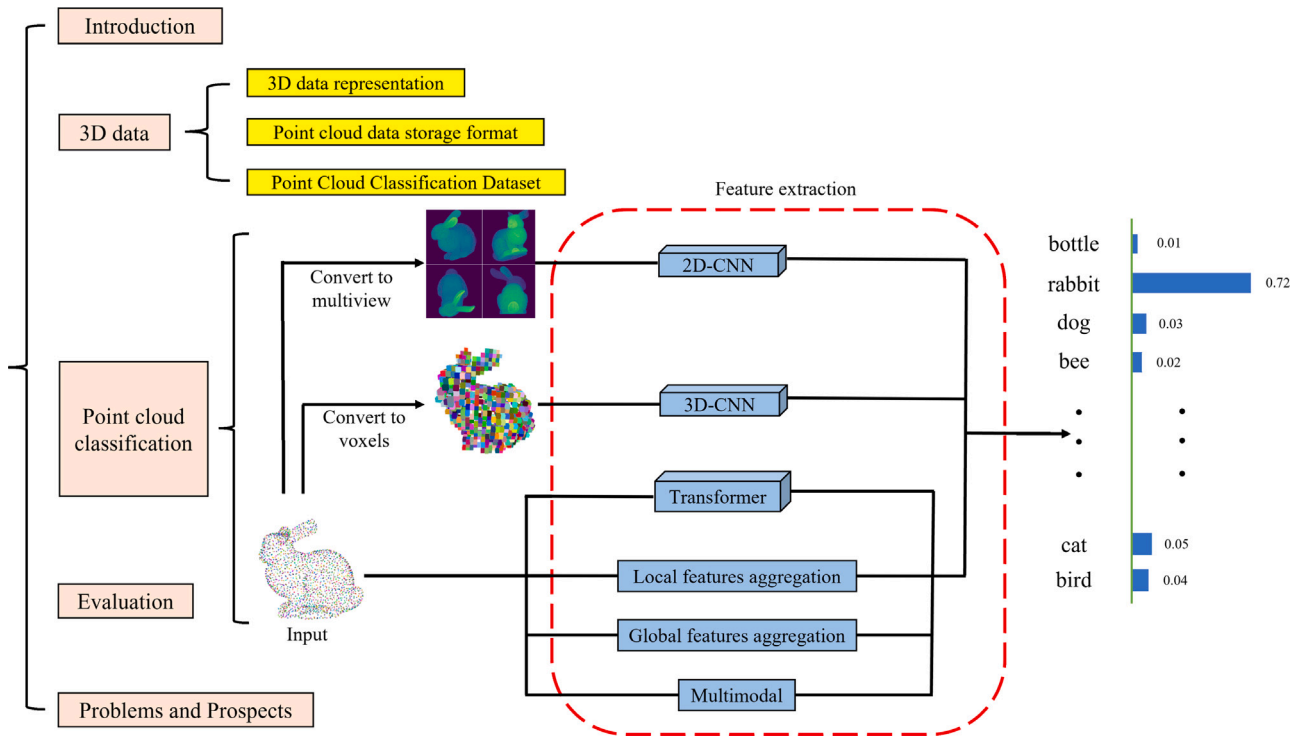


Fig. 2. Overall structure of the article. First, the point cloud is used as input, and the point cloud can be transformed by voxel or multi-view. Secondly, features are extracted from the original point cloud, the transformed voxels or multi-views. The final output is the probability value of each class.

Fig. 1(b)), is one of the commonly used mesh grids to describe 3D objects. A collection of points and edges of a slice is called a mesh.

Voxel: In 3D object representation, voxels are also an important form of 3D data representation, as shown in Fig. 1(c), voxels are good at representing non-uniformly filled regularly sampled spaces, therefore, voxels can effectively represent point cloud data with a lot of empty or evenly filled spaces. By voxelizing the point cloud data, it is beneficial to increase the data computing efficiency and reduce the access to random memory, but the voxelization of the point cloud data will inevitably bring a certain degree of information loss.

Multi-view: A multi-view image (as shown in Fig. 1(d)) is also a representation of point cloud data, which is derived from a single-view image, and is an image that renders a 3D object into multiple viewpoints at a specific angle. The challenges are mainly the choice of perspective and perspective fusion.

3.2. Application of point cloud

In the processing process of the point cloud, the same kind of point cloud data not only has similar reflection intensity, color, and other information but also often has similar macroscopic characteristics and local geometric characteristics. Therefore, these features can be used to form feature vectors and classify and extract laser point clouds.

a. Autonomous driving: The automatic point cloud classification method based on machine learning has good applicability in-vehicle laser point cloud processing and has high research value for improving the automatic and intelligent processing of point clouds. The point cloud feature vector is constructed by the contextual semantic environment, local geometric features, and spatial distribution of the vehicle-mounted laser point cloud, which improves the automation and intelligence level of point cloud classification.

b. Medical field: Point clouds can intuitively simulate the 3D structure of biological organs, tissues, etc. Algorithms such as classification based on medical point cloud data can assist doctors in more accurate diagnosis and treatment and have important application value in clinical medicine, medical device-aided design, and 3D printing.

c. 3D reconstruction: As a high-density spatial 3D information data, point cloud plays an important role in 3D model reconstruction tasks. Similar to images, point clouds and meshes have a unique distribution rather than randomness. The three-dimensional reconstruction of the point cloud uses the algorithm to extract discrete spatial information and establish the connection relationship between points, so as to obtain the network model. Point cloud data can effectively balance user needs and economic costs in reconstruction tasks.

3.3. Point cloud data storage format

There are hundreds of 3D file formats available for point clouds, and different scanners produce raw data in many formats. The biggest difference between point cloud data files is the use of ASCII and binary. Binary systems store data directly in binary code. Common point cloud binary formats include FLS, PCD, LAS, etc. Several other common file types can support both ASCII and binary formats. These include PLY, FBX. The E57 stores data in both binary and ASCII formats, combining many of the advantages of both in one file type. Below we introduce some commonly used point cloud data storage formats:

Obj: The point cloud file in obj format is developed by Wavefront Technologies. It is a text file. It is a simple data format that only represents the geometry, normal, color, and texture information of the 3D data. This format is usually represented in ASCII, but there are also proprietary obj binary versions.

Las: The las format is mainly used to store LIDAR point cloud data, which is essentially a binary format file. A LAS file consists of three parts: the header file area (including total number of points, data range, dimension information of each point), variable-length record area (including coordinate system, extra dimensions, etc.), and point set record area (including point coordinate information, R, G, B information, classification information, intensity information, etc.), the las format takes into account the characteristics of LIDAR data, the structure is reasonable, and it is easy to expand.

Ply: The full name of ply is Polygon File Format, which is inspired by obj and is specially used to store 3D data. Ply uses a nominally flat

Table 1
List of Abbreviations.

Abbreviation	Full name
SVM	Support Vector Machines
RF	Random Forest
CRF	Conditional Random Field
MRF	Markov Random Field
MVCNN	Multi- View Convolutional Neural Network
GIFT	GPU acceleration and Inverted File
GVCNN	Group-View Convolutional Neural Networks
RCPCNN	Recurrent Clustering and Pooling CNN
MHBN	Multi-view Harmonized Bilinear Network
MVTN	Multi-View Transition Network
FPNN	Field Probing Neural Networks
MSNet	Multi-Scale Network
VV-Net	Voxel VAE Net
AFA	Adaptive Feature Adjustment
SO-Net	Self-Organizing Network
SOM	self-organizing map
MLP	Multi-Layer Perceptron
GDANet	the geometric disentanglement attention network
PointSCNet	Point Cloud Structure and Correlation Net
RepSurf	Representative Surfaces
PCNN	Point Convolutional Neural Network
RS-CNN	Relational Shape Convolutional Networks
DNNSP	Deep Neural Network with Spatial Pooling
A-CNN	Annular Convolutional Neural Network
SCN	ShapeContextNet
MCCNN	Monte Carlo Convolution Neural Network
InterpCNNs	Interpolating Convolutional Neural Networks
SPHNet	Spherical Harmonics Network
SC-CNN	Spatial Coverage Convolutional Neural Network
SCOP	Spatial Coverage Operator
GNN	Graph Neural Network
ECC	Edge Conditional Convolutional
SpecGCN	Spectral Graph Convolution Network
CAGQ	Coverage-Aware Network Query
DGCNN	Dynamic Graph CNN
LDGCNN	Linked Dynamic Graph Convolutional Neural Network
PAT	Point Attention Transformer
GSA	Group-Shuffle Attention
MHA	Multi-Head Attention
GSS	Gumbel Subset Sampling
FPA	Feature Pyramid Attention
GAU	Global Attention Upsampling
LSA	Local Spatial Awareness
GAC	Graph Attention Convolution
Point TnT	Point Transformer-in-Transformer
VT	Visual Transformer
DETR	Detection Transformer
PCT	Point Cloud Transformer
3DMedPT	3D Medical Point Transformer
BERT	Bidirectional Encoder Representation from Transformers
VSA	Voxel-based Set Attention
VoxSeT	Voxel Set Transformer
AS	Adaptive Sampling module
L-NL	local non-local module
FPS	Farthest Point Sampling
RNN	Recurrent Neural Network
LSTM	Long Short-term Memory
DQN	Deep Q Network
RSNet	Recurrent Slice Networks
PVT	Point-Voxel Transformer
SWA	Sparse Window Attention Module
RA	Relative Attention module

list of polygons to represent objects. It can store information including color, transparency, surface normal vector, texture coordinates and data confidence, and can set different properties for the front and back sides of the polygon. There are two versions of this file, an ASCII version, and a binary version.

E57: E57 is a vendor neutral file format for point cloud storage. It can also be used to store image and metadata information generated by laser scanners and other 3D imaging systems and is a strict format using fixed-size fields and records. It saves data using ASCII and binary

codes and provides most of the accessibility of ASCII and the speed of binary, and it can store 3D point cloud data, attributes, images.

PCD: PCD is the official designated format of Point Cloud Library. It consists of two parts: header file and point cloud data. It is used to describe the overall information of point cloud. It has two data storage types, ASCII, and binary, but the header file of PCD file must use ASCII. Encoding, a nice benefit of PCD is that it adapts well to PCL, resulting in the highest performance compared to PCL applications.

3.4. 3D point cloud public datasets

Today, there are many point cloud datasets provided by industries and universities. The performance of different methods on these datasets reflects the reliability and accuracy of the methods. These datasets consist of virtual or real scenes, which can provide ground-truth labels for training the network. In this section, we will introduce some commonly used point cloud classification datasets, and the division of each dataset is shown in Table 1.

ModelNet40 [26]: The dataset was developed by the Vision and Robotics Laboratory at Princeton University. The ModelNet40 dataset consists of synthetic CAD objects. As the most widely used benchmark for point cloud analysis, ModelNet40 is popular for its diverse categories, clear shapes, and well-structured datasets. The dataset consists of objects of 40 categories (e.g. airplane, car, plant, lamp), of which 9843 are used for training and 2468 are used for testing. The corresponding points are uniformly sampled from the mesh surface and then further preprocessed by moving to the origin and scaling to a unit sphere. Download link: <https://modelnet.cs.princeton.edu/>

ModelNet-C [27]: The ModelNet-C set contains 185,000 different point clouds and was created based on the ModelNet40 validation set. This dataset is mainly used to benchmark damage robustness for 3D point cloud recognition, with 15 damage types and 5 severity levels for each damage type, such as noise, density, etc. Helps to understand the robustness of the model. Download link: <https://sites.google.com/umich.edu/modelnet40c>

ModelNet10 [26]: ModelNet10: ModelNet10 is a subset of ModelNet40, the dataset contains only 10 classes and it is divided into 3991 training and 908 testing shapes. Download link: <https://modelnet.cs.princeton.edu/>

Sydney Urban Objects [28]: The dataset, collected in Sydney CBD, contains a variety of common urban road objects, including 631 scanned objects in the categories of vehicles, pedestrians, signs, and trees. Download link: <https://www.acfr.usyd.edu.au/papers/SydneyUrbanObjectsDataset.shtml>

ShapeNet [29]: ShapeNet is a large repository of 3D CAD models developed by researchers at Stanford University, Princeton University, and the Toyota Institute of Technology in Chicago, USA. The repository contains over 300 million models, of which 220,000 models are classified into 3,135 classes arranged using WordNet hypernym-hyponymy relations. ShapeNetCore is a subset of ShapeNet that includes nearly 51,300 unique 3D models. It provides 55 common object categories and annotations. ShapeNetSem is also a subset of ShapeNet, which contains 12,000 models. It is smaller in scale but has a wider coverage, including 270 categories. Download link: <https://shapenet.org/>

ScanNet [30]: ScanNet is an instance-level indoor RGB-D dataset containing 2D and 3D data. It is a collection of labeled voxels, not points or objects. As of now, ScanNet v2, the latest version of ScanNet, has collected 1513 annotated scans with about 90% surface coverage. In the semantic segmentation task, this dataset is labeled with 20 classes of annotated 3D voxelized objects. Download link: <http://www.scan-net.org/>

ScanObjectNN [31]: ScanObjectNN is a real-world dataset consisting of 2902 3D objects divided into 15 categories, which is a challenging point cloud classification dataset due to background, missing parts, and deformations in the dataset. Download link: <https://hkust-vgd.github.io/scanobjectnn/>

Table 2
Point cloud classification datasets.

Dataset	Year	Samples	Classes	Training	Test	Type	Form
ModelNet40 [26]	2015	12311	40	9843	2468	Synthetic	Mesh
ModelNet40-C [27]	2015	185000	15	–	–	Synthetic	Point Cloud
ModelNet10 [26]	2015	4899	10	3991	605	Synthetic	Mesh
Sydney Urban Objects [28]	2013	588	14	–	–	Real Word	Point Cloud
ShapeNet [29]	2015	51190	55	–	–	Synthetic	Mesh
ScanNet [30]	2017	12283	17	9677	2606	Real Word	RGB-D
ScanObjectNN [31]	2019	2902	15	2321	581	Real Word	Point Cloud

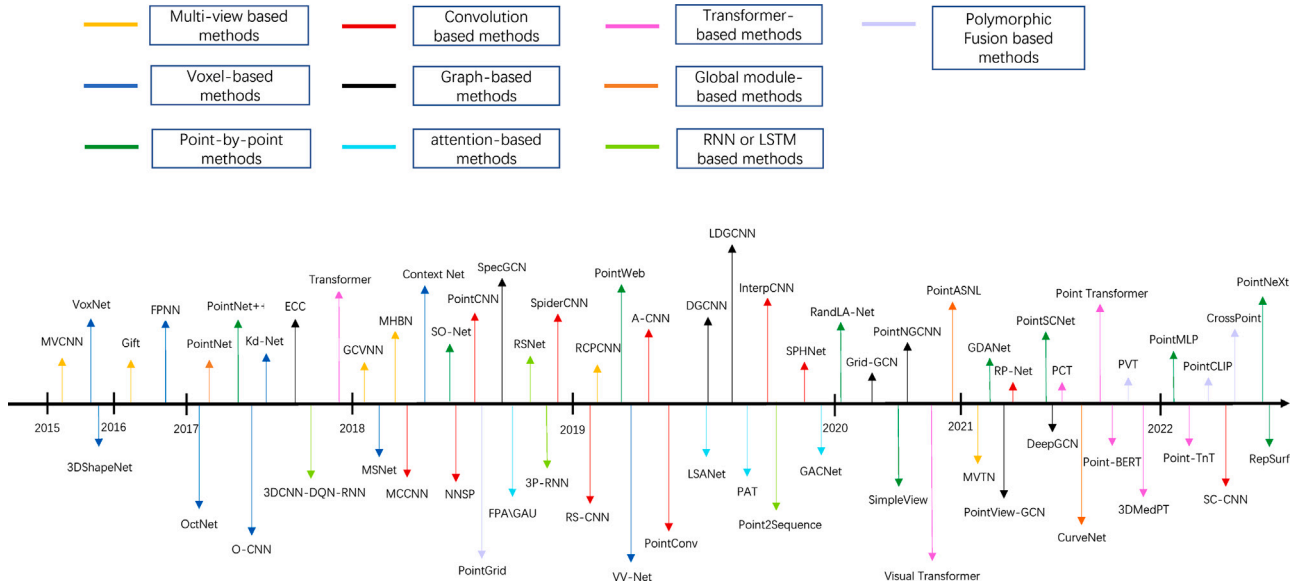


Fig. 3. Timeline of the development of classification methods. Different color lines represent different kinds of point cloud classification.

4. Point cloud classification method based on deep learning

The point cloud classification model based on deep learning has been widely used in point cloud analysis due to its advantages of strong generalization ability and high classification accuracy. This section provides a detailed division of deep learning-based point cloud classification methods and supplements some recent research works. Fig. 3 shows the publication timeline of each classification method.

4.1. Multi-view-based methods

Multi-view learning is a machine learning framework in which data is represented by multiple distinct feature groups, each of which is called a specific view. The multi-view-based method is a deep learning based on 2D images. This method is divided into three steps: First, project the 3D image into multiple views. Second, extract the view function. Third, fuse these functions to accurately classify 3D shapes.

In 2015, Su et al. [32] first proposed Multi-View convolutional neural network (MVCNN). Since the collection of 2D views can provide a lot of information for 3D shape recognition, this method recognizes 3D shapes from the collection of rendering views on 2D images. Representing 3D Shapes" is a longstanding problem. This method first demonstrates a standard CNN architecture trained to recognize shapes independently of rendered views and shows that 3D shapes can be recognized even from a single view. When multiple views of an object are provided, the recognition rate is further improved by using the CNN architecture to combine the information of multiple views into a compact shape descriptor. The network architecture is shown in Fig. 4. This method requires a large amount of computation in the projection retrieval process, and when converting multiple view features into global features through max-pooling, other non-maximum

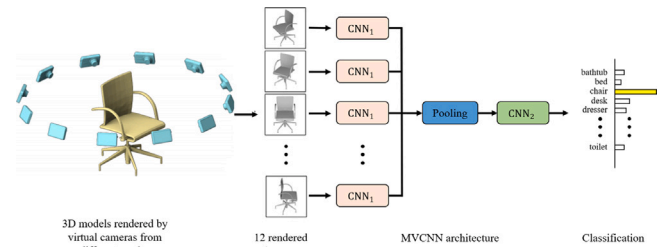


Fig. 4. Firstly, the multi-view of the object is obtained through 12 cameras with different angles to represent the 3D shape of the object. Secondly, the features of different views are extracted through CNN. Finally, the compact shape descriptor is obtained through pooling and CNN to achieve classification.

element information is ignored, so it will inevitably cause information missing.

Therefore, in view of a large amount of computation and efficiency of MVCNN, Bai et al. [33] proposed a real-time 3D shape search engine, namely GIFT (GPU acceleration and Inverted File), this method uses GPU acceleration in the stages of projection and view feature extraction, which greatly shortens the time spent on retrieval tasks, and has high efficiency and ability to handle large-scale data.

Based on MVCNN, the Group-View convolutional neural networks (GVCNN) architecture proposed by Feng et al. [34] introduces the "view-group-shape" architecture to extract descriptors, which can effectively utilize the feature relationship between views. In order to overcome the problem of information loss due to max pooling, Wang et al. [35] introduced the view clustering and pooling layer based on the dominant set to improve the MVCNN method and proposed recurrent clustering and pooling CNN (RCPCNN). It entails building

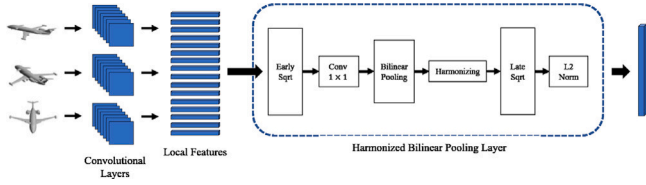


Fig. 5. The architecture of Multi-view Harmonized Bilinear Network (MHBN [36]).

a view similarity graph, clustering nodes (views) in this graph based on dominance sets, and pooling information from within each cluster. The recurrent clustering and pooling layer are to aggregate multi-view features in a way that provides more discriminative power for 3D object recognition. The RCPCNN is plugged into ready-made pre-trained CNN which improves the performance of multi-view 3D object recognition.

The Multi-view harmonized bilinear network (MHBN) proposed by Yu et al. [36] aggregates local convolutional features by bilinear pooling to obtain efficient 3D object representations. Fig. 5 shows the network architecture of this method. To achieve an end-to-end trainable framework, this method coordinates the merging of bilinear pooling into one layer in the network, from the perspective of patch-to-patch similarity measurement, to address the problem of view-based methods pooling view features that differ.

To avoid the lack of dynamism of current multi-view methods, Hamdi et al. [37] propose the Multi-View Transition Network (MVTN), which proposes a differentiable module that predicts the best view-point for a task-specific multi-view network, regressing the best for 3D shape recognition Viewpoint. MVTN can be trained end-to-end with any multi-view network, showing significant performance gains in 3D shape classification and retrieval tasks without additional training supervision.

Summary: Compared with the traditional manual extraction feature classification, the multi-view-based method has a better effect in point cloud classification, but it is still difficult to make full use of information. The application of large-scale scenes, and the inherent geometric relationship of 3D data is the challenge we need to face.

4.2. Voxel-based methods

This type of approach converts a 3D point cloud model into voxel form that approximates the shape of an object, each voxel block contains a set of associated points, and then uses 3D CNNs to classify the voxels.

Maturana et al. [38] proposed a convolutional neural network architecture called VoxNet to represent 3D information with a volumetric occupancy grid. VoxNet is the earliest voxel-based 3DCNN model. As shown in Fig. 6, this method normalizes each grid, then constructs a feature map through convolution and max pooling a single voxel block. This architecture uses 2.5D to represent the features of the local description scan, and adopts a full volume representation, which improves the ability to express environmental information and enables powerful 3D object recognition.

Wu et al. [26] proposed 3D ShapeNets to recognize 3D objects. The model represents a 3D shape as a probability distribution of binary variables on a grid of 3D voxels, each of which can be represented by a binary tensor, and predicts the next best view in the presence of uncertain initial recognition. Finally, 3D ShapeNets can incorporate new views to identify objects in common with all views.

Since 3D convolutions are computationally expensive, the spatial resolution increases model complexity when convolving 3D voxels. In response to this phenomenon, Li et al. [39] proposed a Field Probing neural networks (FPNN), which represented 3D data as a 3D field, then sampled the input field through a set of detection filters, and finally used the field detection filter to characterize extract. The field detection

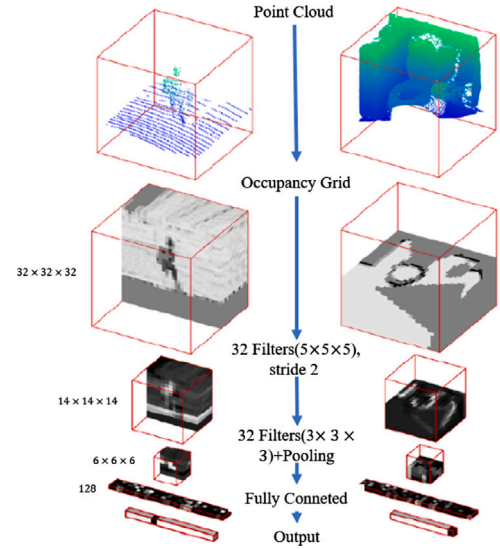


Fig. 6. First, the input layer is the grid of $32 \times 32 \times 32$ voxels. Second, the double-layer convolution takes the spatial information of the point cloud and feature map as the four-dimensional input. Then, the features are extracted through a pooling process. Finally, the output category is obtained through full connection.

layer can be used with other inference layers, but this approach makes the semantic segmentation results lower resolution.

To reduce memory consumption and improve computational efficiency, some scholars use octree structure instead of fixed-resolution voxel structure. Riegler et al. [40] proposed OctNet, which exploits the sparsity of 3D data to hierarchically partition the space with a set of unbalanced octrees, where each leaf node stores a pooled feature representation. This method enables deeper networks without affecting their resolution. Wang et al. [41] proposed an octree-based convolutional neural network called O-CNN, which uses octrees to represent 3D data information and discretize its surface. The 3D CNN operation is only performed on the octant occupied by the 3D shape surface, which improves the computational efficiency and power.

Following the use of the octree structure in the 3D data representation, the Kd-tree structure is also used in the point cloud classification model. The Kd-network proposed by Klovov et al. [42] adopts, compared with voxel and mesh, the ability of Kd-tree to index and structure 3D data has been improved, so Kd-network has a smaller memory footprint and more efficient computation during training and testing. The 3D Context Net proposed by Zeng et al. [43] utilizes the method of local and global context cues imposed by Kd-tree for semantic segmentation.

Octree structure and Kd-tree structure reduce memory consumption to a certain extent and improve computational efficiency, but due to the influence of voxel boundary value, these two structures cannot make full use of local data features and the accuracy is affected. Wang et al. [44] proposed a multi-scale convolutional network (MSNet). This method first divides the space into voxels of different scales, then applies MSNet on multiple scales simultaneously to learn local features, and finally uses a CRF. The prediction results of MSNet are globally optimized to achieve a more accurate point cloud classification task.

The VV-Net Voxel VAE Net (VV-Net) proposed by Meng et al. [45] uses a kernel-based interpolating variational autoencoder (VAE) to encode localities in voxels, each voxel is further subdivided into sub-voxels, which are within voxels interpolate sparse point samples to efficiently handle noisy point cloud datasets.

Summary: Compared with the multi-view method, the voxel-based method pays attention to the relationship between 3D data, and groups the point clouds with internal connections into a set of points, thereby establishing voxels. Although the voxel-based model solves the point

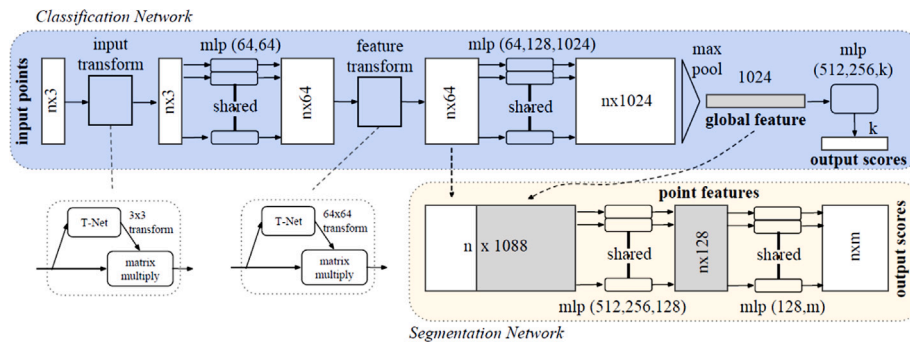


Fig. 7. PointNet [18] network architecture. The blue area indicates that the network realizes point cloud classification, and the yellow part indicates that the network realizes point cloud segmentation.

cloud disorder and unstructured problem, the sparseness and incomplete information of point cloud data lead to the low efficiency of the classification task, so the information in the point cloud cannot be fully utilized.

4.3. Point cloud-based method

Many current research methods focus more on directly processing point clouds using deep learning techniques. Feature aggregation operator is the core of processing point cloud, which realizes the information transfer of discrete points. Feature aggregation operators are mainly divided into two categories: local feature aggregation and global feature aggregation. In this section, from the perspective of feature aggregation, the two categories of methods are divided and introduced in more detail.

In 2017, the PointNet proposed by Qi et al. [18] (shown in Fig. 7) is a pioneering study of point cloud-based methods, which is a method of global feature aggregation. This method directly takes the point cloud as input, transforms it through the T-Net module, then learns each point by sharing the full connection, and finally aggregates the features of the point into global features through the maximum pooling function. Although PointNet is a pioneer based on deep learning, it still has defects. For example, PointNet only captures the feature information of a single point and global points but does not consider the relational representation of adjacent points, which makes PointNet unable to effectively perform fine-grained classification.

4.3.1. Local feature aggregation

a. Point-by-point method. Qi et al. [46] successively proposed PointNet++ based on PointNet. It needs to solve two problems: dividing the generated point set and aggregating local features through a local feature learner. It processes point clouds in a hierarchical manner, and each layer consists of a sampling layer, a grouping layer, and a PointNet layer, as shown in Fig. 8. Among them, the sampling layer obtains the centroid of the local neighborhood, the grouping layer constructs a subset of the local neighborhood, and the PointNet layer obtains the relationship between the points in the local area. But PointNet++ still ignores the prior relationship between points.

Based on PointNet++, Qian et al. [47] improved the training and training strategy to improve the performance of PointNet++ and introduced a separable MLP and an inverted residual bottleneck design in the PointNet++ framework, named its framework PointNeXt (as shown in Fig. 9).

Zhao et al. [48] proposed a method of extracting features from the local neighborhood of point clouds — PointWeb, which connects the points in the local neighborhood with each other, and proposes a new module, Adaptive Feature Adjustment (AFA), which is used to find information transfer between points. This method makes full use of the local features of points and realizes adaptive adjustment.

Hu et al. [49] proposed an efficient and lightweight neural architecture - RandLA-Net, which uses random point sampling, and increases the receptive field of each point through an efficient local feature aggregation module, so as to better capturing complex local structures reduces memory footprint and computational cost, but this approach may discard some key features of sparse points.

The Self-Organizing Network (SO-Net) proposed by Li et al. [50] constructs a self-organizing map (SOM), extracts the features of each point and SOM node hierarchically, and uses a single feature vector to represent the point cloud, by appending a 3-layer perception to the encoded global feature vector Multi-Layer Perceptron (MLP) is used to classify point clouds. SO-Net has good parallelism and simple structure, but it has limitations in processing large-scale point cloud data.

In order to fully capture the most critical geometric information, Xu et al. [51] proposed the geometric disentanglement attention network — GDANet, and introduced the Geometry-Disentangle module to decompose the original point cloud into two parts: contour and plane, so as to capture and refine 3D semantics to supplementing local information. The method has good robustness.

Goyal et al. [52] proposed a projection-based method, SimpleView, which showed that training and evaluation factors independent of the network architecture have a large impact on point cloud classification performance. The PointSCNet proposed by Chen et al. [53] is used to capture the geometric information and local area information of the point cloud. It consists of three modules: a space-filling curve-guided sampling module, an information fusion module, and a channel spatial attention module. In PointSCNet the raw point cloud is fed to the sampling and grouping block, which is sampled using a Z-order curve to obtain the high correlation between points and local regions. After extracting the sampled point cloud features, a feature fusion module is designed to learn the structure and related information. Finally, the keypoint features are enhanced by the channel space module.

Ma et al. [54] noticed that detailed local geometric information may not be the key to analyzing point clouds, so they introduced a pure residual network called PointMLP, which does not have a complex local geometry extractor, but is equipped with a lightweight geometric affine module, there is a significant improvement in inference speed. Huang et al. [55] calculated the rate of change of recognition confidence when a shape-invariant perturbation with explicit constraints was added to each point in the point cloud, and according to this method, they proposed a point cloud sensitivity map, which was then used to propose a shape-invariant adversarial point cloud.

Ran et al. [56] proposed using representative surfaces (RepSurf) to represent point clouds, which has two variants: Triangular RepSurf and Umbrella RepSurf. This representation can be used as a module in the point cloud classification and segmentation framework to improve the performance of the point cloud framework.

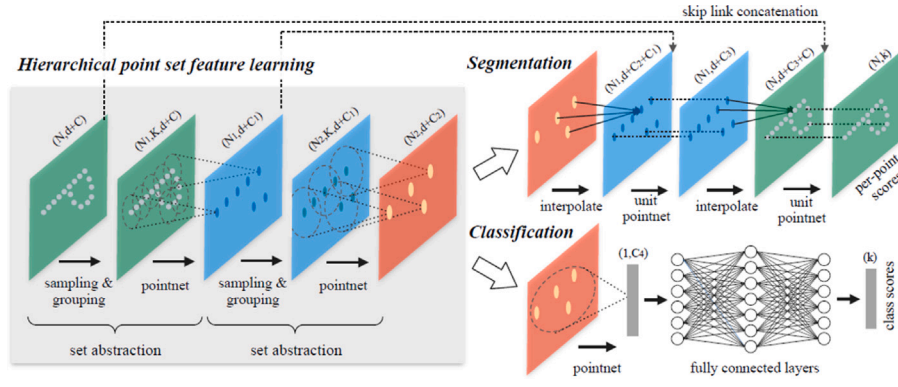


Fig. 8. PointNet++[46] point cloud segmentation and classification architecture diagram. The feature map obtained after twice Sampling&Grouping is used to extract global features through a PointNet, and then the classification results are obtained through the fully connected network.

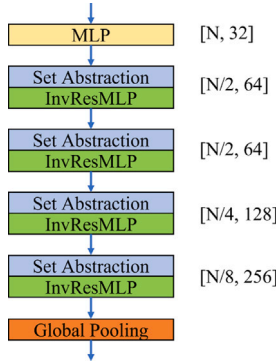


Fig. 9. PointNet classification architecture. It has the same Set Abstraction module as PointNet++. The InvResMLP module represents the inverted residual MLP used to scale the model architecture.

b. Convolution-based methods. Convolutional Neural Network (CNN) plays an important role in deep learning and is the most basic deep learning model. Its excellent performance in the field of 2D image processing has led researchers to apply it to 3D point clouds and design point convolution for point cloud classification.

Atzmon et al. [57] proposed the point convolutional neural network (PCNN), which applied convolutional Neural Network (CNN) to point cloud. First, the function on the point cloud is extended to a continuous volume function in space; then a continuous volume convolution is applied to the function; the final result is a constrained point cloud.

Liu et al. [58] proposed Relational Shape Convolutional Networks (RS-CNN) to extend regular CNNs to irregular point clouds for analysis of point clouds. Yousefhusien et al. [59] proposed an one-dimensional fully convolutional network. Wang et al. [60] proposed a deep neural network with spatial pooling (DNNSP) to classify large-scale point clouds. This method can learn the features from the entire region to the center point of the point cluster, to achieve a robust representation of point features. Komarichev et al. [61] propose a point cloud-based annular convolutional neural network (A-CNN) model. Ran et al. [62] proposed RPNNet based on a group relation aggregation module, which is robust to rigid transformations and noise. The ShapeContextNet (SCN) proposed by Xie et al. [63] is represented by using the shape context as a building block, so that it can capture and propagate object part information. SCN is an end-to-end model.

Since directly convolving kernels with point-related features would result in the discarding of shape information and variance in point ordering, Li et al. [64] proposed PointCNN to address this problem, which confirmed the development of local structures for point cloud classification networks importance.

Due to the sparsity, irregularity, and disorder of the point cloud, it is difficult to directly perform the convolution operation on it. Wu et al. [65] proposed to apply the dynamic filter to the convolution operation called PointConv, which is simple and reduces the computer storage pressure. Monte Carlo Convolution Neural Network (MCCNN) [66] represents the convolution kernel itself as a multilayer perceptron and describes the convolution as a Monte Carlo integration. SpiderCNN [67] inherits the multi-scale hierarchical structure of CNN and consists of SpiderConv units, which extend the convolution operation from regular grids to irregular point sets that can be embedded in n -dimensional space by parameterizing a series of convolution filters to effectively extract geometric features from point clouds. Mao et al. [68] designed interpolating convolutional neural networks (InterpCNNs) based on combining InterConv(interpolating convolution operation).

Esteves et al. [69] use multi-valued spherical functions to model 3D data, and propose a spherical convolutional network that implements them by implementing precise convolutions on spheres in the spherical harmonic domain, thus solving the 3D rotation variance problem in convolutional neural networks. SPHNet [70] is based on Point CNN to achieve rotation invariance by using spherical harmonics in different layers of the network.

Since the local features of point clouds are difficult to aggregate and transfer effectively, Wang et al. [71] proposed a spatial coverage convolutional neural network (SC-CNN), the core of which is the spatial coverage convolution (SC-Conv). Anisotropic spatial geometry is constructed in the point cloud to implement depthwise separable convolution, replacing depthwise convolution with the spatial coverage operator (SCOP).

c. Graph-based methods. The graph neural network (GNN) was first proposed by Scarselli et al. [72]. Bruna et al. [73] are the first to apply convolution to low-dimensional graph structures to effectively represent depth. Kipf et al. [74] further proposed that a graph convolutional network (GCN) works well in semi-supervised classification tasks, and in fact, GCN is an optimization of CNN.

Simonovsky [75] proposed an ECC(edge conditional convolutional) network that can be applied to any graph structure in combination with the application of edge labels. This method uses the points as the vertices of the graph and the distance between the points as the weight, and performs a weighted average convolution operation, using the maximum sampling the information of aggregated vertices. It can be used for large-scale point cloud segmentation, but the amount of computation is large.

SpecGCN [76] replaces the standard max-pooling step with a recursive clustering and pooling strategy. Grid-GCN [77] uses a coverage-aware network query (CAGQ), which improves spatial coverage and reduces theoretical time complexity by exploiting the efficiency of grid space.

Mohammadi et al. [78] proposed PointView-GCN with a multi-level graph convolutional network (GCN) to hierarchically aggregate shape features of single-view point clouds, which facilitates encoding of object geometric cues and multi-view relationships, resulting in more specific global features.

Wang et al. [79] proposed a dynamic graph CNN (DGCNN) for point cloud learning and proposed an edge convolutional (EdgeConv) network module, which can better capture the local geometric features of point clouds and maintain arrangement invariance, which demonstrates the importance of local geometric features for 3D recognition tasks. Zhang et al. [80] further optimized DGCNN and proposed a Linked Dynamic Graph Convolutional Neural Network (LDGCNN), which removed the transformation network in DGCNN to simplify the network model and optimized the network by connecting hierarchical features of different dynamic graphs, which can better solve the gradient vanishing problem.

The PointNGCNN proposed by Lu et al. [81] describes the relationship between points in the neighborhood in a neighborhood graph and uses neighborhood graph filters to extract neighborhood feature information and spatial distribution information in feature space and Cartesian space.

d. Attention-based methods. The basic idea of the attention mechanism is to apply human perception to the machine, but human perception selectively focuses on part of the scene instead of processing the entire scene at a time, so researchers focus on the attention mechanism to carry out research and apply to the field of point cloud classification.

Yang et al. [82] developed a point attention transformer (PAT) based on point cloud reasoning. It is proposed to use efficient GSA (Group-Shuffle Attention) instead of expensive MHA (Multi-Head Attention) for modeling the relationship between points, and propose a method called Gumbel Subset Sampling (GSS) to select a subset of representative points, which reduces the computational cost.

Li et al. [83] proposed Feature Pyramid Attention Module (FPA) and Global Attention Upsampling Module (GAU) by combining attention mechanism and spatial pyramid. Chen et al. [84] designed a Local Spatial Awareness (LSA) layer and proposed an LSANet network architecture based on the LSA layer. LSA can learn the spatial relationship layer in the local area to generate spatially distributed weights, so that spatially independent operations can be performed. The spatial information extraction function of this method is powerful.

Wang et al. [85] proposed GACNet based on graph attention convolution (GAC). The GAPointNet proposed by Chen et al. [86] combines the self-attention mechanism with graph convolution, learns local information representation by embedding a graph attention mechanism in stacked MLP layers, and uses a parallel mechanism to aggregate the attention features of different GAPLayer layers, where GAPLayer layers and attention layers can be embedded into existing trained models to better extract local contextual features from unordered point clouds.

4.3.2. Global feature aggregation

a. Transformer-based methods. Since the transformer was first proposed in 2017 [87], it has achieved world-renowned results in the field of computer vision. Many researchers also use this structure in point cloud processing.

Engel et al. [88] propose a deep neural network Point Transformer that operates directly on unordered and unstructured point sets, and propose a learning score-based focus module, ScorNet, as part of the Point Transformer. The point cloud is used as the input of the Point Transformer, and local and global features are extracted from it, and then SortNet is used to rank the local features, and finally the local global attention is used to associate the local global features, as shown in Fig. 10.

Berg et al. [89] found that the self-attention operator grows rapidly and inefficiently with the growth of the input point set, and the attention mechanism is difficult to find the relationship between each

point in the global, so they propose a two-stage method — Point TnT, this method enables a single point and a point set to pay attention to each other effectively.

The Visual Transformer (VT) proposed by Wu et al. [90], which applies Transformer to label-based images from feature maps, can learn and associate high-level concepts of sparse distributions more efficiently. Carion et al. [91] propose a method to treat object detection as a direct ensemble prediction problem called Detection Transformer (DETR), which is an end-to-end detection transformer that takes CNN features as input and uses a Transformer encoder-decoder to generate boundaries.

Guo et al. [92] proposed a Transformer-based point cloud learning framework — Point Cloud Transformer (PCT), and proposed offset attention with implicit Laplacian operator and normalization refinement, the framework has permutation invariance and is more suitable for point cloud learning

3DMedPT is a Transformer network proposed by Yu et al. [9] for 3D medical point cloud analysis.

Inspired by BERT (Bidirectional Encoder Representation from Transformers), Yu et al. [93] proposed a new method for learning Transformer called Point-BERT. This method first divides the point cloud into several local blocks and generates discrete point labels of local information through a point cloud marker, and then by randomly masking some input point clouds and feeding them into the backbone Transformer, this method can generalize the concept of BERT to point clouds.

Pang et al. [94] proposed Point-MAE, which is a masked autoencoder method for point cloud self-supervised learning to solve the problems caused by point cloud location information leakage and uneven information density. question.

He et al. [95] introduced a voxel-based set attention module (VSA) to establish the Voxel Set Transformer (VoxSeT) architecture. VoxSeT can manage point clusters through the VSA module and process them in parallel with linear complexity. This method combines the high performance of Transformer with the high efficiency of voxel-based model, it has good performance in point cloud modeling.

b. Global module-based methods. Wang et al. [96] propose a global module that computes the response at a location as a weighted sum of all location features, providing a solution to aggregating global features, but the global point-to-point mapping may still be insufficient to extract the underlying patterns implied by the point cloud shape.

Yan et al. [97] propose an end-to-end network PointASNL that combines an adaptive sampling module (AS) and a local non-local module (L-NL), which can effectively deal with noisy point clouds. The AS module updates the features of the points by reasoning, then normalizes the weight parameters and re-weights the initial sampling points, which can effectively alleviate the bias effect. The L-NL module consists of local and non-local units of points, reducing the sensitivity of the learning process to noise.

Li et al. [98] adopted some CNN methods to support a deep GCN architecture, called DeepGCN, which consists of three blocks: GCN Backbone block for input point cloud feature transformation, fusion block for generating and fusing global features, MLP block prediction block used to predict labels. To solve the problem of gradient disappearance during GCN training, it is possible to train deeper GCN networks.

Xiang et al. [99] proposed a method based on aggregating hypothetical curves in point clouds, CurveNet, and effectively implemented the aggregation strategy, including a curve grouping operator and a curve aggregation operator. The network consists of a bunch of building blocks, FPS represents the farthest point sampling method.

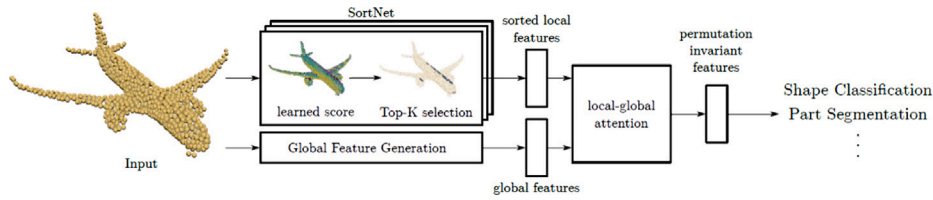


Fig. 10. Architecture of Point Transformer [88]. Local and global features are extracted from the input point cloud. SortNet sorts the local features to focus on the features of interest and captures their geometric relationship and shape information.

c. RNN or LSTM-based methods. Recurrent Neural Network (RNN) can usually effectively utilize context information to process sequence data. Long Short-term Memory (LSTM) is a special RNN, which can effectively solve the problems of gradient disappearance and gradient explosion in the training process of longer sequence data.

Engelmann et al. [100] extended the input-level context information and output-level context information based on PointNet, which enables PointNet to be applied in large-scale scenarios. It can make PointNet applied in large-scale scenarios.

Liu et al. [101] proposed a 3DCNN-DQN-RNN method, which fuses 3D Convolutional Neural Network (CNN), Deep Q Network (DQN) and residual RNN. First, the feature representation of points is obtained by 3DCNN. Second, DQN can detect and localize objects, and can automatically perceive the scene and adjust the feedback of 3DCNN. Finally, the RNN is used to identify the connections and differences of multi-scale features. Where LSTM units are used to prevent vanishing gradients and make the network have long-term memory, the method improves the accuracy of processing large-scale point clouds.

The RSNet network proposed by Huang et al. [102] takes the original point cloud as input, and then performs feature extraction, and then passes through the slice pooling layers in the three directions of x, y, and z. Each layer uses a bidirectional RNN to extract local features, and then uses slice parsing. The layer assigns the features of the point cloud sequence to each point, and finally outputs the predicted semantic label of each point.

Ye et al. [103] proposed an end-to-end semantic segmentation method, 3P-RNN, by combining CNN and RNN, which consists of a point pyramid module and a bidirectional hierarchical RNN module. In the work of distinguishing the same semantics, this method has certain limitations.

Point2Sequence proposed by Liu et al. [104] uses RNN to capture fine-grained contextual information to learn 3D shape features, it introduces an attention mechanism to enhance feature extraction.

4.4. Polymorphic fusion-based methods

The strategy of combining PointGrid with grids proposed by Le et al. [105] is to mix points and grids for representation. PointGrid is composed of several convolution blocks, which represent the features of different layers through maximum pooling. Each convolution layer includes a convolution kernel, and the over-fitting phenomenon is controlled by the pooling layer, and then completed by full connection. For inference, PointGrid has two fully connected layers, and finally performs regression with a softmax classifier, which can better identify fine-grained models and represent local shapes.

Zhang et al. [106] proposed a novel point cloud learning method, PVT (Point-Voxel Transformer), combining sparse window attention module (SWA) and relative attention module (RA), which combines voxel-based and point-based model idea, this method excels in the accuracy of point cloud classification.

PointCLIP proposed by Zhang et al. [107] learns point clouds based on pre-trained CLIP. Encoding the point cloud by projecting it into a multi-view depth map without rendering, enables zero-shot recognition by transferring the 2D pretrained knowledge to the 3D domain. And

an inter-view adapter is designed to better extract global features. The network architecture is shown in Fig. 11.

CrossPoint [108] achieves 2D-to-3D correspondence by maximizing the point cloud and the corresponding rendered 2D image in invariant space and keeping the point cloud unchanged across transformations.

Summary: Compared with the multi-view-based method and the voxel-based method, the point cloud-based method directly processes the original points and can make full use of the point cloud information. Therefore, the point cloud-based method is also a future research direction, and the transformer-based method will be more widely used in the future.

5. Evaluation

The evaluation index is used to evaluate the performance of the point cloud classification method. The accuracy, space complexity, execution time, etc. are the evaluation indexes of the method, and the accuracy is the key index for evaluating various methods. Generally, accuracy (Acc), recall (Rec), and intersection-over-union (IoU) are used to evaluate the accuracy of the method. We list the results of the survey in Table 3. For the methods that did not cover all datasets in the experimental part, we investigate their comparative experiments from the authoritative literature [64,79,82,97] to show more perfect results.

- Accuracy refers to the ratio of the number of correctly predicted samples to the total number of predicted samples.
- Recall refers to the ratio of samples predicted to be positive classes to the total number of true positive classes.
- Intersection-over-Union refers to the ratio of the intersection and union of the predicted value and the true value.

The accuracy, recall, and intersection-over-union can be calculated by the following formulas: TP is the sample that is predicted to be a positive class and is actually a positive class, TN is a sample that is predicted to be a positive class and is actually a negative class, and FP is the sample. The samples that are predicted to be negative classes are actually negative classes, and FN is the samples that are predicted to be negative classes and are actually positive classes. Suppose there are N classes:

The accuracy of the i th class:

$$Acc_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (1)$$

Recall:

$$Rec = \frac{TP}{TP + FN} \quad (2)$$

The intersection ratio of the i th class:

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (3)$$

The current accuracy of point cloud classification is measured by indicators: overall accuracy (OA), average accuracy (MA), and average intersection-over-union ratio (mIoU), which are calculated as follows:

$$OA = \frac{\sum_{i=1}^N TP_i}{\sum_{i=1}^N (TP_i + FP_i)} \quad (4)$$

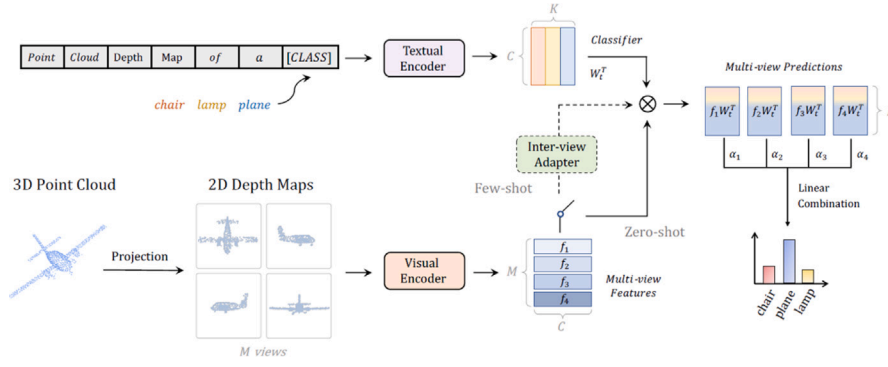


Fig. 11. The Pipeline of PointCLIP [107]. It projects the point cloud onto the multi-view depth map to fill the multimodal gap and extracts the global features through the interview adapter.

$$MA = \frac{1}{N} \sum_{i=1}^N Acc_i \quad (5)$$

$$mIoU = \frac{1}{N} \sum_{i=1}^N IoU \quad (6)$$

This section summarizes the performance of the mentioned methods on different datasets in Table 3.

6. Discussion

From the results in Table 3, it can be seen that the OA of most methods can reach over 90% on ModelNet40, while some cutting-edge methods such as RepSurf, PointMLP, and RP-Net can achieve 94%. The OA of PiontView-GCN is 95.4%, which meets the results of SOTA. PointNeXt, PointMLP, PointCNN, and others also perform well in large-scale and complex scenarios ScanObjectNN, but further breakthroughs are needed. Although significant achievements have been made in research based on deep learning, existing methods still have some limitations. We need to analyze its limitations, accurately grasp the future development of the three-dimensional field, and fill the current research gaps. This section will discuss the shortcomings of current point cloud classification methods based on deep learning, summarize research conclusions, and look forward to future research directions, as follows:

- From the experimental results of various methods (Table 3), it can be seen that some classification methods have outstanding performance in the ModelNet dataset, but their results are not outstanding in the more complex ScanObjectNN. This is a limitation caused by the difficulty in extracting and utilizing the features of complex scene point clouds. The classification network loses the original point cloud information during the process of feature sampling, aggregation, and extraction. [51] distinguishes the impact of point clouds in different regions on overall classification results by dividing point cloud features into sharpening and smoothing. Therefore, reasonable utilization of point clouds' local or non-local effective features is the key to solving this problem. It is also a future research trend based on original point cloud methods.
- From a data perspective, with the development of 3D scanning technology, the acquisition of large-scale point clouds is no longer a challenge. However, the quality of point cloud data significantly impacts the robustness of classification networks. The challenge lies in processing and correcting the original point cloud data. In addition, 3D data obtained through optical radar scanning often have varying degrees of noise, but current algorithms have little research on point cloud noise. In the future, we should explore the impact of noise on point cloud classification.
- At the application level, most methods currently focus on the analysis of relatively simple indoor scenes. However, in practical applications, it is often outdoor scenes with more information and complex

structures, resulting in limitations of classification algorithms in applications such as autonomous driving and forest monitoring. The key to breaking through this limitation is to improve the efficiency of classification algorithms and reduce the number of computational parameters.

- The method based on the origin cloud has certain advantages in algorithm performance, but the network model is more complex. Due to the integrity of the input raw point cloud data, simple point-based methods are still the future research trend.

7. Finding

Direct processing of point clouds has become a hot topic in recent years and a trend in point cloud processing, as it can maximize the utilization of multi-dimensional information of points to avoid information loss.

With the increasing demand for point cloud classification methods in various fields, researchers continue to propose new methods to improve accuracy and efficiency, which promote the development of deep learning and 3D applications. The point-based network model is the most frequent research content in the point cloud task, but the local features of the point cloud usually guarantee its neighborhood information by querying the index, which will inevitably cause the loss of some key information and the decline of computational efficiency. Effectively addressing this problem is extremely challenging.

Some of the current point cloud classification methods focus on improving accuracy, while others focus on improving efficiency. We need to solve the problem of “how to achieve high accuracy while being efficient?” In the network model we will propose in the future, we should pay attention to the optimization of the network architecture, which can reduce the computational complexity and memory usage while facing the complex and irregular point cloud.

8. Conclusion

This paper provides a comprehensive survey and discussion of deep learning-based point cloud classification methods in recent years.

First, we introduced the point cloud and its application in the introduction and discussed the characteristics and processing difficulties of the point cloud. Secondly, in the second section, the 3D data is introduced, and the commonly used 3D data representations, point cloud data storage formats and point cloud classification datasets are summarized. Building on the previous sections, we comprehensively review deep learning-based point cloud classification methods, classifying these methods into four broad categories: multi-view-based methods, voxel-based methods, point cloud-based methods, and polymorphic fusion-based methods. And then we compare the performance of existing methods. Finally, this paper points out the problems of the current method and prospects the future research direction.

Table 3
Accuracy comparison of point cloud classification methods on different datasets.

Methods	Year	Accuracy/% on different datasets											
		ModelNet 40			ModelNet 10			ScanObjectNN			ShapeNet		
		OA	MA	mIoU	OA	MA	mIoU	OA	MA	mIoU	OA	MA	mIoU
Multi-view based methods	MVCNN [32]	2016	90.10	78.90	–	–	–	–	–	–	–	–	–
	GIFT [33]	2016	83.10	81.94	–	92.35	91.12	–	–	–	–	–	–
	GCVNN [34]	2018	93.10	79.70	–	–	–	–	–	–	–	–	–
	MHBN [36]	2018	94.91	–	–	92.93	–	–	–	–	–	–	–
	RCPCNN [35]	2019	93.80	–	–	–	–	–	–	–	–	–	–
	MVTN [37]	2021	92.00	93.80	–	–	–	82.80	–	–	–	–	–
Voxel-based methods	VoxNet [38]	2015	85.90	83.00	–	92.00	–	–	–	–	–	–	–
	3D shapeNet [26]	2015	84.70	77.30	–	83.50	–	–	–	–	–	–	–
	FPNN [39]	2016	87.50	–	–	–	–	–	–	–	–	–	–
	OctNet [40]	2017	86.50	–	90.90	–	–	–	–	–	–	–	–
	O-CNN [41]	2017	90.60	–	85.90	–	–	–	–	–	–	–	–
	Kd-Net [42]	2017	91.80	88.50	–	94.00	93.50	77.20	–	–	–	–	82.30
	3D Context Net [43]	2018	–	–	–	–	–	–	–	–	–	–	–
	MSNet [44]	2018	–	–	–	–	–	–	–	–	–	–	–
	VV-Net [45]	2019	–	–	–	–	–	–	–	–	–	–	87.40
Point cloud-based method	PointNet++[46]	2017	91.90	–	–	–	–	77.90	75.40	–	–	–	85.10
	PointNetXt [47]	2022	–	–	–	–	–	87.70	85.80	–	–	–	87.20
	PointWeb [48]	2019	92.30	89.40	–	–	–	–	–	–	–	–	–
	RandLA-Net [49]	2020	–	–	–	–	–	–	–	–	–	–	–
	SO-Net [50]	2018	90.80	87.30	–	94.10	93.90	–	–	–	–	–	84.60
	GDANet [51]	2021	93.80	–	–	–	–	88.50	–	–	–	–	86.50
	SimpleView [52]	2020	93.00	–	–	–	–	80.50	–	–	–	–	–
	PintSCNet [53]	2021	93.70	–	–	–	–	–	–	–	–	–	–
	PointMLP [54]	2022	94.50	91.40	–	–	–	85.40	83.90	–	–	–	–
	RepSurf [56]	2022	94.70	91.70	–	–	–	84.60	81.90	–	–	–	–
	PCNN [57]	2018	92.30	–	–	94.90	–	–	–	–	–	–	–
	RS-CNN [58]	2019	93.60	–	–	–	–	–	–	–	–	–	86.20
	DNNSP [60]	2018	–	–	–	–	–	–	–	–	–	–	–
	A-CNN [61]	2019	92.60	90.30	–	95.50	95.30	–	–	–	–	–	86.10
	RP-Net [62]	2021	94.10	–	–	–	–	–	–	–	–	–	–
	SCN [63]	2018	90.00	–	–	–	–	–	–	–	–	–	84.60
	PointCNN [64]	2018	92.20	88.10	–	–	–	87.90	–	–	–	–	–
	PointCove [65]	2019	92.50	–	–	–	–	–	–	–	–	–	–
	MCCNN [66]	2018	90.90	–	–	–	–	–	–	–	–	–	–
	SpiderCNN [67]	2018	92.40	–	–	–	–	73.70	69.80	–	–	–	85.30
	InterpCNN [68]	2019	93.00	–	–	–	–	–	–	–	–	–	86.30
	SPHNet [70]	2019	87.70	–	–	–	–	–	–	–	–	–	–
	SC-CNN [71]	2022	93.8	–	–	96.4	–	–	–	–	–	–	86.40
	ECC [75]	2017	–	83.20	–	90.00	–	–	–	–	–	–	–
	SpecGCN [76]	2018	91.50	–	–	–	–	–	–	–	–	–	84.60
	Grid-GCN [77]	2020	93.10	91.30	–	97.50	97.40	–	–	–	–	–	–
	PointView-GCN [78]	2021	95.40	–	–	–	–	85.50	–	–	–	–	–
	DGCNN [79]	2019	92.20	90.20	–	–	–	86.20	–	–	–	–	–
	LDGCNN [80]	2019	92.90	90.30	–	–	–	–	–	–	–	–	–
	PointNGCNN [81]	2020	92.80	–	–	–	–	–	–	–	–	–	85.60
	PAT [82]	2019	91.70	–	–	–	–	–	–	–	–	–	–
	FPA/GAU [83]	2018	–	–	–	–	–	–	–	–	–	–	–
	LSANet [84]	2019	92.30	89.20	–	–	–	–	–	–	–	–	83.20
	GACNet [85]	2019	–	–	–	–	–	–	–	–	–	–	–
	GAPointNet [86]	2021	92.40	89.70	–	–	–	–	–	–	–	–	–
	Point Transformer [88]	2021	92.80	–	–	–	–	–	–	–	85.90	–	–
	Visual Transformer [90]	2020	–	–	–	–	–	–	–	–	–	–	–
	PCT [92]	2021	93.20	–	–	–	–	–	–	–	–	–	–
	3DMedPT [9]	2021	93.40	–	–	–	–	–	–	–	–	–	–
	Point-BERT [93]	2021	93.80	–	–	–	–	83.07	–	–	–	–	–
	Point-TnT [89]	2022	92.60	–	–	–	–	84.60	82.60	–	–	–	–
	Point-MAE [94]	2022	93.80	–	–	–	–	85.18	–	–	–	–	86.10
	PointNet [18]	2017	89.20	86.20	–	–	–	68.20	63.40	–	–	–	83.70
	PointASNL [97]	2020	93.20	–	–	95.90	–	–	–	–	–	–	–
	CurveNet [99]	2021	94.20	–	–	96.30	–	–	–	–	–	–	86.80
	DeepGCNs [98]	2021	93.20	90.30	–	–	–	–	–	–	–	–	–
	3DCNN-DQN-RNN [101]	2017	–	–	–	–	–	–	–	–	–	–	–
	RSNet [102]	2018	–	–	–	–	–	–	–	–	–	–	–
	3P-RNN [103]	2018	–	–	–	–	–	–	–	–	–	–	–
	Point2Sequence [104]	2019	92.60	90.40	–	95.30	95.10	–	–	–	–	–	–
Polymorphic Fusion based methods	PointGrid [105]	2018	92.00	88.90	–	–	–	–	–	–	86.10	80.50	–
	PVT [106]	2021	94.00	–	–	–	–	–	–	–	–	–	86.60
	PointCLIP [107]	2022	94.05	–	–	–	–	–	–	–	–	–	–
	CrossPoint [108]	2022	94.90	–	–	–	–	79.00	–	–	–	–	85.50

Although we have made a systematic investigation on point cloud classification, there are still some limitations that need to be addressed by researchers in future research. The point cloud classification method is applied to many real scenes, such as indoors, roads, railways, and gardens. In these scenes, we cannot specifically judge the advantages of many point cloud classification methods. Therefore, it is necessary for researchers to select a classification algorithm suitable for the required scenario according to the actual situation, which is also the problem of dataset shortage mentioned in the discussion. In addition, researchers need to further explore and supplement the polymorphic fusion classification algorithm.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgment

This work is supported by Xinjiang University School-Enterprise Joint Project.

References

- [1] C. Ma, W. Meng, B. Wu, S. Xu, X. Zhang, Efficient joint gradient based attack against sor defense for 3d point cloud classification, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 1819–1827.
- [2] C. Ma, W. Meng, B. Wu, S. Xu, X. Zhang, Towards effective adversarial attack against 3d point cloud classification, in: 2021 IEEE International Conference on Multimedia and Expo, ICME, IEEE, 2021, pp. 1–6.
- [3] C. Wang, C. Wang, W. Li, H. Wang, A brief survey on RGB-D semantic segmentation using deep learning, *Displays* 70 (2021) 102080.
- [4] M. Sturari, M. Paolanti, E. Frontoni, A. Mancini, P. Zingaretti, Robotic platform for deep change detection for rail safety and security, in: 2017 European Conference on Mobile Robots, ECMR, IEEE, 2017, pp. 1–6.
- [5] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, C. Pan, Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network, *IEEE Trans. Geosci. Remote Sens.* 55 (6) (2017) 3322–3337.
- [6] D. Fernandes, A. Silva, R. Névoa, C. Simões, D. Gonzalez, M. Guevara, P. Novais, J. Monteiro, P. Melo-Pinto, Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy, *Inf. Fusion* 68 (2021) 161–191.
- [7] Y. Zhou, O. Tuzel, Voxnet: End-to-end learning for point cloud based 3d object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4490–4499.
- [8] Y. Wang, S. Wu, Y. Duan, Y. Huang, A point cloud-based deep learning strategy for protein-ligand binding affinity prediction, *Brief. Bioinform.* 23 (1) (2022) bbab474.
- [9] J. Yu, C. Zhang, H. Wang, D. Zhang, Y. Song, T. Xiang, D. Liu, W. Cai, 3D medical point transformer: Introducing convolution to attention networks for medical point cloud analysis, 2021, arXiv preprint arXiv:2112.04863.
- [10] T. Xu, D. An, Y. Jia, Y. Yue, A review: Point cloud-based 3d human joints estimation, *Sensors* 21 (5) (2021) 1684.
- [11] B. Yang, R. Huang, J. Li, M. Tian, W. Dai, R. Zhong, Automated reconstruction of building lods from airborne lidar point clouds using an improved morphological scale space, *Remote Sens.* 9 (1) (2016) 14.
- [12] M.A. Hearst, S.T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, *IEEE Intell. Syst. their Appl.* 13 (4) (1998) 18–28.
- [13] R. Kumar, R. Verma, Classification algorithms for data mining: A survey, *Int. J. Innov. Eng. Technol.* 1 (2) (2012) 7–14.
- [14] Z. Gan, L. Zhong, Y. Li, H. Guan, A random forest based method for urban object classification using lidar data and aerial imagery, in: 2015 23rd International Conference on Geoinformatics, IEEE, 2015, pp. 1–4.
- [15] L. Zhou, Y. Liu, Z. Pengcheng, B. Xiao, Y. Yazhou, J. Zhou, G. Lin, T. Harada, E.R. Hancock, Information bottleneck and selective noise supervision for zero-shot learning, *Mach. Learn.* (2022).
- [16] N. Plath, M. Toussaint, S. Nakajima, Multi-class image segmentation using conditional random fields and global classification, in: Proceedings of the 26th Annual International Conference on Machine Learning, 2009, pp. 817–824.
- [17] D. Munoz, J.A. Bagnell, N. Vandapel, M. Hebert, Contextual classification with functional max-margin markov networks, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 975–982.
- [18] C.R. Qi, H. Su, K. Mo, L.J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660.
- [19] A. Nguyen, B. Le, 3D point cloud segmentation: A survey, in: 2013 6th IEEE Conference on Robotics, Automation and Mechatronics, RAM, IEEE, 2013, pp. 225–230.
- [20] E. Arnold, O.Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, A. Mouzakitis, A survey on 3d object detection methods for autonomous driving applications, *IEEE Trans. Intell. Transp. Syst.* 20 (10) (2019) 3782–3795.
- [21] B. Fei, W. Yang, W.-M. Chen, Z. Li, Y. Li, T. Ma, X. Hu, L. Ma, Comprehensive review of deep learning-based 3d point cloud completion processing and analysis, *IEEE Trans. Intell. Transp. Syst.* (2022).
- [22] D.P. Kumar, T. Amgoth, C.S.R. Annavarapu, Machine learning algorithms for wireless sensor networks: A survey, *Inf. Fusion* 49 (2019) 1–25.
- [23] E.O. Mkpogio, O.E. Okeke-Uzodike, E.I. Emmanuel, Quality attributes for an LMS cognitive model for user experience design and evaluation of learning management systems, in: 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), Atlantis Press, 2021, pp. 234–242.
- [24] E.O. Mkpogio, A. Hussain, M.O. Agbudu, Security issues in the use of mobile educational apps: A review, *Int. J. Interact. Mobile Technol.* 15 (6) (2021).
- [25] S. Dustdar, V.C. Pujol, P.K. Donta, On distributed computing continuum systems, *IEEE Trans. Knowl. Data Eng.* (2022).
- [26] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3D shapenets: A deep representation for volumetric shapes, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1912–1920.
- [27] J. Sun, Q. Zhang, B. Kailkhura, Z. Yu, C. Xiao, Z.M. Mao, Benchmarking robustness of 3D point cloud recognition against common corruptions, 2022, arXiv preprint arXiv:2201.12296.
- [28] M. De Deuge, A. Quadros, C. Hung, B. Douillard, Unsupervised feature learning for classification of outdoor 3d scans, in: Australasian Conference on Robotics and Automation, 2, University of New South Wales Kensington, Australia, 2013, p. 1.
- [29] A.X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, et al., Shapenet: An information-rich 3d model repository, 2015, arXiv preprint arXiv:1512.03012.
- [30] A. Dai, A.X. Chang, M. Savva, M. Halber, T. Funkhouser, M. Nieß ner, ScanNet: Richly-annotated 3D reconstructions of indoor scenes, in: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.
- [31] M.A. Uy, Q.-H. Pham, B.-S. Hua, D.T. Nguyen, S.-K. Yeung, Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data, in: International Conference on Computer Vision, ICCV, 2019.
- [32] H. Su, S. Maji, E. Kalogerakis, E. Learned-Miller, Multi-view convolutional neural networks for 3D shape recognition, in: IEEE International Conference on Computer Vision, 2015.
- [33] B. Song, B. Xiang, Z. Zhou, Z. Zhang, L.J. Latecki, GIFT: A real-time and scalable 3D shape search engine, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016.
- [34] Y. Feng, Z. Zhang, X. Zhao, R. Ji, G. Yue, GVCNN: Group-view convolutional neural networks for 3D shape recognition, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
- [35] C. Wang, M. Pelillo, K. Siddiqi, Dominant set clustering and pooling for multi-view 3d object recognition, 2019, arXiv preprint arXiv:1906.01592.
- [36] T. Yu, J. Meng, J. Yuan, Multi-view harmonized bilinear network for 3d object recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 186–194.
- [37] A. Hamdi, S. Giancola, B. Ghanem, Mvtn: Multi-view transformation network for 3d shape recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1–11.
- [38] D. Maturana, S. Scherer, Voxnet: A 3d convolutional neural network for real-time object recognition, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2015, pp. 922–928.
- [39] Y. Li, S. Pirk, H. Su, C.R. Qi, L.J. Guibas, Fpnn: Field probing neural networks for 3d data, *Adv. Neural Inf. Process. Syst.* 29 (2016).
- [40] G. Riegler, A. Osman Ulusoy, A. Geiger, Octnet: Learning deep 3d representations at high resolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3577–3586.
- [41] P.-S. Wang, Y. Liu, Y.-X. Guo, C.-Y. Sun, X. Tong, O-cnn: Octree-based convolutional neural networks for 3d shape analysis, *ACM Trans. Graph.* 36 (4) (2017) 1–11.
- [42] R. Klokov, 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, 2017, pp. 863–872.
- [43] W. Zeng, T. Gevers, 3Dcontextnet: Kd tree guided hierarchical learning of point clouds using local and global contextual cues, in: Proceedings of the European Conference on Computer Vision (ECCV) Workshops, 2018.
- [44] L. Wang, Y. Huang, J. Shan, L. He, MSNet: Multi-scale convolutional network for point cloud classification, *Remote Sens.* 10 (4) (2018) 612.
- [45] H.-Y. Meng, L. Gao, Y.-K. Lai, D. Manocha, Vv-net: Voxel vae net with group convolutions for point cloud segmentation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 8500–8508.
- [46] C.R. Qi, L. Yi, H. Su, L.J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [47] G. Qian, Y. Li, H. Peng, J. Mai, H.A.A.K. Hammoud, M. Elhoseiny, B. Ghanem, PointNetx: Revisiting PointNet++ with improved training and scaling strategies, 2022, arXiv preprint arXiv:2206.04670.
- [48] H. Zhao, L. Jiang, C.-W. Fu, J. Jia, Pointweb: Enhancing local neighborhood features for point cloud processing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 5565–5573.
- [49] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, A. Markham, Randla-net: Efficient semantic segmentation of large-scale point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11108–11117.
- [50] J. Li, B.M. Chen, G.H. Lee, So-net: Self-organizing network for point cloud analysis, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9397–9406.
- [51] M. Xu, J. Zhang, Z. Zhou, M. Xu, X. Qi, Y. Qiao, Learning geometry-disentangled representation for complementary understanding of 3d object point cloud, in: Proceedings of the AAAI Conference on Artificial Intelligence, 35, (4) 2021, pp. 3056–3064.
- [52] A. Goyal, H. Law, B. Liu, A. Newell, J. Deng, Revisiting point cloud classification with a simple and effective baseline, 2020.

- [53] X. Chen, Y. Wu, W. Xu, J. Li, H. Dong, Y. Chen, PointSCNet: Point cloud structure and correlation learning based on space-filling curve-guided sampling, *Symmetry* 14 (1) (2021) 8.
- [54] X. Ma, C. Qin, H. You, H. Ran, Y. Fu, Rethinking network design and local geometry in point cloud: A simple residual mlp framework, 2022, arXiv preprint arXiv:2202.07123.
- [55] Q. Huang, X. Dong, D. Chen, H. Zhou, W. Zhang, N. Yu, Shape-invariant 3D adversarial point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 15335–15344.
- [56] H. Ran, J. Liu, C. Wang, Surface representation for point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 18942–18952.
- [57] M. Atzmon, H. Maron, Y. Lipman, Point convolutional neural networks by extension operators, 2018, arXiv preprint arXiv:1803.10091.
- [58] Y. Liu, B. Fan, S. Xiang, C. Pan, Relation-shape convolutional neural network for point cloud analysis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8895–8904.
- [59] M. Yousefhussein, D.J. Kelbe, E.J. Ientilucci, C. Salvaggio, A multi-scale fully convolutional network for semantic labeling of 3D point clouds, *ISPRS J. Photogramm. Remote Sens.* 143 (2018) 191–204.
- [60] Z. Wang, L. Zhang, L. Zhang, R. Li, Y. Zheng, Z. Zhu, A deep neural network with spatial pooling (DNNSP) for 3-D point cloud classification, *IEEE Trans. Geosci. Remote Sens.* 56 (8) (2018) 4594–4604.
- [61] A. Komarichev, Z. Zhong, J. Hua, A-cnn: Annularly convolutional neural networks on point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7421–7430.
- [62] H. Ran, W. Zhuo, J. Liu, L. Lu, Learning inner-group relations on point clouds, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 15477–15487.
- [63] S. Xie, S. Liu, Z. Chen, Z. Tu, Attentional shapecontextnet for point cloud recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4606–4615.
- [64] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, B. Chen, Pointcnn: Convolution on x-transformed points, *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [65] W. Wu, Z. Qi, L. Fuxin, Pointconv: Deep convolutional networks on 3d point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9621–9630.
- [66] P. Hermosilla, T. Ritschel, P.-P. Vázquez, À. Vinacua, T. Ropinski, Monte carlo convolution for learning on non-uniformly sampled point clouds, *ACM Trans. Graph.* 37 (6) (2018) 1–12.
- [67] Y. Xu, T. Fan, M. Xu, L. Zeng, Y. Qiao, Spidercnn: Deep learning on point sets with parameterized convolutional filters, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 87–102.
- [68] J. Mao, X. Wang, H. Li, Interpolated convolutional networks for 3d point cloud understanding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1578–1587.
- [69] C. Esteves, C. Allen-Blanchette, A. Makadia, K. Daniilidis, Learning so (3) equivariant representations with spherical cnns, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 52–68.
- [70] A. Poulenard, M.-J. Rakotosaona, Y. Ponty, M. Ovsjanikov, Effective rotation-invariant point cnn with spherical harmonics kernels, in: 2019 International Conference on 3D Vision (3DV), IEEE, 2019, pp. 47–56.
- [71] C. Wang, X. Ning, L. Sun, L. Zhang, W. Li, X. Bai, Learning discriminative features by covering local geometric space for point cloud analysis, *IEEE Trans. Geosci. Remote Sens.* 60 (2022) 1–15.
- [72] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, *IEEE Trans. Neural Netw.* 20 (1) (2008) 61–80.
- [73] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, 2013, arXiv preprint arXiv:1312.6203.
- [74] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.
- [75] M. Simonovsky, N. Komodakis, Dynamic edge-conditioned filters in convolutional neural networks on graphs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3693–3702.
- [76] C. Wang, B. Samari, K. Siddiqi, Local spectral graph convolution for point set feature learning, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 52–66.
- [77] Q. Xu, X. Sun, C.-Y. Wu, P. Wang, U. Neumann, Grid-gcn for fast and scalable point cloud learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5661–5670.
- [78] S.S. Mohammadi, Y. Wang, A. Del Bue, PointView-GCN: 3D shape classification with multi-view point clouds, in: 2021 IEEE International Conference on Image Processing, ICIP, IEEE, 2021, pp. 3103–3107.
- [79] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein, J.M. Solomon, Dynamic graph cnn for learning on point clouds, *Acm Trans. Graph.* 38 (5) (2019) 1–12.
- [80] K. Zhang, M. Hao, J. Wang, C.W. de Silva, C. Fu, Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features, 2019, arXiv preprint arXiv:1904.10014.
- [81] Q. Lu, C. Chen, W. Xie, Y. Luo, PointNGCNN: Deep convolutional networks on 3D point clouds with neighborhood graph filters, *Comput. Graph.* 86 (2020) 42–51.
- [82] J. Yang, Q. Zhang, B. Ni, L. Li, J. Liu, M. Zhou, Q. Tian, Modeling point clouds with self-attention and gumbel subset sampling, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3323–3332.
- [83] H. Li, P. Xiong, J. An, L. Wang, Pyramid attention network for semantic segmentation, 2018, arXiv preprint arXiv:1805.10180.
- [84] L.-Z. Chen, X.-Y. Li, D.-P. Fan, K. Wang, S.-P. Lu, M.-M. Cheng, LSANet: Feature learning on point sets by local spatial aware layer, 2019, arXiv preprint arXiv:1905.05442.
- [85] L. Wang, Y. Huang, Y. Hou, S. Zhang, J. Shan, Graph attention convolution for point cloud semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 10296–10305.
- [86] C. Chen, L.Z. Fragonara, A. Tsourdos, GPointNet: Graph attention based point neural network for exploiting local feature of point cloud, *Neurocomputing* 438 (2021) 122–132.
- [87] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [88] N. Engel, V. Belagiannis, K. Dietmayer, Point transformer, *IEEE Access* 9 (2021) 134826–134840.
- [89] A. Berg, M. Oskarsson, M. O'Connor, Points to patches: Enabling the use of self-attention for 3D shape recognition, 2022, arXiv preprint arXiv:2204.03957.
- [90] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, P. Vajda, Visual transformers: Token-based image representation and processing for computer vision, 2020, arXiv preprint arXiv:2006.03677.
- [91] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: European Conference on Computer Vision, Springer, 2020, pp. 213–229.
- [92] M.-H. Guo, J.-X. Cai, Z.-N. Liu, T.-J. Mu, R.R. Martin, S.-M. Hu, Pct: Point cloud transformer, *Comput. Vis. Media* 7 (2) (2021) 187–199.
- [93] X. Yu, L. Tang, Y. Rao, T. Huang, J. Zhou, J. Lu, Point-bert: Pre-training 3d point cloud transformers with masked point modeling, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 19313–19322.
- [94] Y. Pang, W. Wang, F.E. Tay, W. Liu, Y. Tian, L. Yuan, Masked autoencoders for point cloud self-supervised learning, 2022, arXiv preprint arXiv:2203.06604.
- [95] C. He, R. Li, S. Li, L. Zhang, Voxel set transformer: A set-to-set approach to 3D object detection from point clouds, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8417–8427.
- [96] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.
- [97] X. Yan, C. Zheng, Z. Li, S. Wang, S. Cui, Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5589–5598.
- [98] G. Li, M. Müller, G. Qian, I.C.D. Perez, A. Abualshour, A.K. Thabet, B. Ghanem, Deepgcn: Making gcns go as deep as cnns, *IEEE Trans. Pattern Anal. Mach. Intell.* (2021).
- [99] T. Xiang, C. Zhang, Y. Song, J. Yu, W. Cai, Walk in the cloud: Learning curves for point clouds shape analysis, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 915–924.
- [100] F. Engelmann, T. Kontogianni, A. Hermans, B. Leibe, Exploring spatial context for 3D semantic segmentation of point clouds, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 716–724.
- [101] F. Liu, S. Li, L. Zhang, C. Zhou, R. Ye, Y. Wang, J. Lu, 3DCNN-DQN-rnn: A deep reinforcement learning framework for semantic parsing of large-scale 3D point clouds, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5678–5687.
- [102] Q. Huang, W. Wang, U. Neumann, Recurrent slice networks for 3d segmentation of point clouds, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2626–2635.
- [103] X. Ye, J. Li, H. Huang, L. Du, X. Zhang, 3D recurrent neural networks with context fusion for point cloud semantic segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 403–417.
- [104] X. Liu, Z. Han, Y.-S. Liu, M. Zwicker, Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network, in: Proceedings of the AAAI Conference on Artificial Intelligence, 33, (01) 2019, pp. 8778–8785.
- [105] T. Le, Y. Duan, Pointgrid: A deep network for 3d shape understanding, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9204–9214.
- [106] C. Zhang, H. Wan, X. Shen, Z. Wu, PVT: Point-voxel transformer for point cloud learning, 2021, arXiv preprint arXiv:2108.06076.
- [107] R. Zhang, Z. Guo, W. Zhang, K. Li, X. Miao, B. Cui, Y. Qiao, P. Gao, H. Li, Pointclip: Point cloud understanding by clip, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8552–8562.
- [108] M. Afham, I. Dissanayake, D. Dissanayake, A. Dharmasiri, K. Thilakarathna, R. Rodrigo, Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 9902–9912.