IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 April 2023, accepted 14 June 2023, date of publication 20 June 2023, date of current version 23 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3287940

SURVEY

Exploring Semantic Information Extraction From Different Data Forms in 3D Point Cloud Semantic Segmentation

ANSI ZHANG[1], SONG LI[1], JIE WU[1], SHAOBO LI[1], AND BAO ZHANG[2]
[1]State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China
[2]School of Mechanical Engineering, Guizhou University, Guiyang 550025, China

Corresponding author: Song Li (songli970103@gmail.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1713300, in part by the Scientific Research Project for Introducing Talents from Guizhou University under Grant (2021)74, in part by the Guizhou Provincial Department of Education Youth Science and Technology Talent Growth Project under Grant [2022]142, in part by the Guizhou Province Higher Education Integrated Research Platform Project under Grant [2020]005, and in part by the Guizhou Provincial Colleges and Universities Talent Training Base Project under Grant [2020]009.

ABSTRACT As a critical step in 3D scene understanding, semantic segmentation of point clouds has broad application scenarios, including intelligent driving, augmented reality, smart factories, etc. Point cloud data is complex and irregular, and traditional machine learning methods are difficult to achieve ideal segmentation results. Deep learning techniques have yielded remarkable outcomes for researchers, leading to a surge in interest in investigating the semantic segmentation of point clouds. This article begins by examining the difficulties involved in segmenting point clouds by analyzing the inherent structural characteristics of point clouds. Then, commonly used datasets for point cloud semantic segmentation and evaluation metrics for assessing segmentation performance were introduced. Subsequently, an exploration was carried out on extracting semantic information from different data forms in point cloud semantic segmentation. Based on these findings, the experimental results of these methods on publicly available datasets are compared quantitatively. Lastly, several outlooks are presented regarding the future development of semantic segmentation techniques for 3D point clouds. The point cloud semantic segmentation techniques summarized in this paper are mainly from the state-of-the-art methods presented at top international conferences. The goal is to provide a comprehensive overview of this field's state of the art and can be used as a reference for researchers and beginners.

INDEX TERMS 3D point cloud, deep learning, public datasets, semantic segmentation.

I. INTRODUCTION

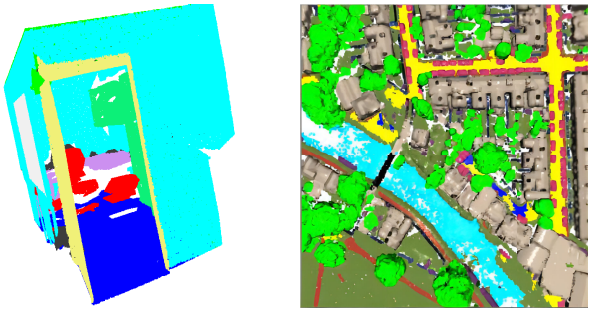
The task of point cloud semantic segmentation entails assigning a pre-defined semantic category to each point in the point cloud data, such as pedestrians, cars, buildings, etc. (as demonstrated in Fig.1, which exhibits the semantic segmentation outcomes of point clouds in various scenarios). This technology is the basis for fields such as autonomous driving [1], [2], indoor navigation [3], [4], and built environment analysis [5], [6], which can help computers better understand the environment and make more accurate decisions, and thus has a wide range of application scenarios.

Driven by the demand for practical applications, research on point clouds is gradually becoming popular due to the increasingly convenient acquisition and processing of point cloud data with the widespread use of sensors such as LiDAR and stereo cameras. The importance of point cloud semantic segmentation is that it can provide a high-precision understanding and analysis of the 3D environment. For example, point cloud semantic segmentation in autonomous driving can help vehicles identify roads, pedestrians, and obstacles better and thus make more accurate decisions.

The associate editor coordinating the review of this manuscript and approving it for publication was Eunil Park.

TABLE 1. Structural characteristics of 3D point clouds and the corresponding solutions proposed by researchers for these challenges on the semantic segmentation task.

Structural characteristics	Solution Ideas	Specific methods
Permutation invariance	Obtaining global features through symmetric functions	PointNet [11], RSNet [12]
	Transforming point clouds into ordered representations	Rethage's method [13], SO-Net [14]
	Proposing new convolution operations	PointCNN [15], EdgeConv [16]
	Sampling operation with permutation invariance	Yang's method [17]
Rotation invariance	Learning the rotation invariance feature	PointNet [11], You's method [18], SPHNet [19]
	Mapping the points to a specific space	SRINet [20], SFCNN [21]
	Exploiting potential geometric relationships	RIConv [22], RS-CNN [23]
Density inconsistency	Using specific sampling methods	KPConv [24], GACNet [25], RandLA-Net [26]
	Completing incomplete point cloud surfaces	Yi's method [27]
	Modifying the convolution kernel	InterpConv [28], Lei's method [29]



(a) Point cloud indoor scene segmentation

(b) Point cloud outdoor scene segmentation

FIGURE 1. The segmentation results of the public datasets using the PointNet++ [7] semantic segmentation network are shown, where different colors indicate different semantic categories, and the same color indicates the same categories in the Fig.. (a): The segmentation results for a conference room in area 5 of the S3DIS [8] dataset. (b): The segmentation result for an urban area of the large outdoor dataset SensatUrban [9].

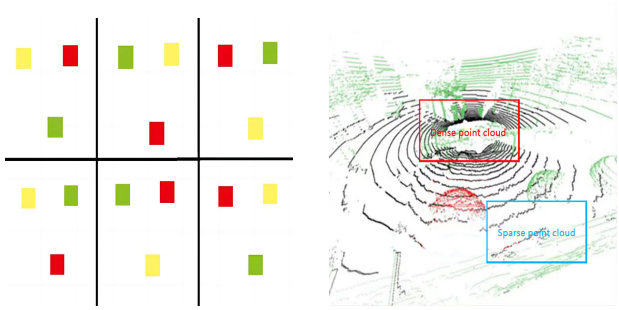
Because of the unique attributes of point cloud data, conventional 2D image processing techniques [10] are not directly transferable to 3D data. The primary obstacles that impede the semantic labeling of point clouds are their structure with permutation invariance, rotation invariance, and density inconsistency. The structural characteristics of point clouds are briefly described below, and some corresponding solutions are summarized, and Table 1 shows the correspondence between structural characteristics and specific solutions.

Permutation invariance refers to swapping the position of any point in the point cloud without affecting the object expressed by the point cloud, as shown in Fig.2(a), which requires algorithms that can learn consistent features from many permutations. Some existing network models address permutation invariance in the point clouds by four main classes of methods. The first class uses symmetric functions to obtain global or local features from point clouds. Network models through this approach include PointNet [11] and RSNet [12]. PointNet [11] uses symmetric functions to obtain global information of all points in a point cloud and then splices this global feature behind each point for semantic segmentation. RSNet [12] performs a maximum pooling operation on the points within a slice to generate a global feature representation in each slice, and all slice information constitutes a feature vector ordered sequence. The second type is the ordered representation achieved by transforming the unordered point cloud. In order to be able to deal with unordered 3D point cloud data directly, Rethage et al. [13] transformed the point cloud internally into an ordered structure by 3D convolution before semantic segmentation processing. SO-Net [14] employed a Self-Organizing Map [30] to Simulate the spatial arrangement of the points. This method compresses the point cloud into an isolated feature vector, ensuring the feature vectors' permutation invariance in theory. The third category is to propose new convolution operators. The χ-conv operator is proposed in PointCNN [15], which can consider the shape of the point clouds and arrange them into a potential canonical order. Similarly, in DensePoint [31], the new operator defined extends the regular grid CNN to irregular point cloud operations while satisfying local connectivity and weight sharing. In addition, Wang et al. [16] proposed the convolution operator EdgeConv based on GCNs. The EdgeConv is designed to be invariant to the order of its neighbors and thus has permutation invariance. The fourth category is the proposed sampling operation with permutation invariance. Yang et al. [17] proposed Gumbel Subset Sampling to solve the permutation invariance problem from the sampling perspective.

Rotation invariance means that the coordinates of almost all points change after rotating the point cloud but still represent the same object. Random directional perturbations of point clouds can make deep learning methods less effective and thus can limit the generalization in practical applications. Generally, three categories of approaches can be employed to address the issue of rotation invariance. The first category eliminates the effect of point cloud rotation

by learning the rotation invariance feature. PointNet [11] rotates the point cloud to a suitable position by adding a T-Net module and then semantically segments it. You et al. [18] propose a Pointwise Rotation-Invariant Network that addresses the rotation invariance of point clouds in spherical space from a deep learning perspective. In SPHNet [19], the idea of learning invariance from data is abandoned, and different spherical harmonics kernels are proposed. The second category addresses rotation invariance by mapping the points in the point cloud to a specific space. SRINet [20] obtains a rotation-invariant representation of the 3D point cloud by mapping the 3D coordinates to a 4D projection feature space. Also, based on projection, SFCNN [21] maps the original points to discrete spheres, which helps the model resist rotations and perturbations while maximizing the preservation of the input 3D shape details. The third category of approaches to address rotational invariance is to use the potential geometric relationships in point clouds. Zhang et al. [22] proposed the RIConv operator to acquire rotationally invariant low-level geometric features. Aided by relational learning, Relational Shape Convolution [23], developed by Liu et al., can incorporate the geometric interdependence of points. The resulting RS-CNN network built on this operator can withstand rotational disturbances due to the sturdy geometric topological relationships between the learned points.

Density inconsistency is manifested as the point cloud in the target scene is dense at some locations and sparse at others, which is more common in autonomous driving scenes, as shown in Fig.2(b). Possible reasons for this situation include the relative position between the object surface and the point cloud sampling device, the color of the object surface, and other factors. The researchers addressed the density inconsistency problem in three aspects. The first type of approach utilizes the idea of sampling, which is the most frequently used method. KPConv [24] combines the radius neighborhood with the conventional subsampling strategy to ensure the efficiency and robustness of point cloud data with different densities. Similarly, in GACNet [25], a directed graph is constructed for a given point cloud data and randomly sampled within a radius ρ to form a neighborhood, ensuring that the neighborhood is independent of the point density. In RandLA-Net [26], random sampling is used for point selection to reduce the computational complexity of high-density and large-scale point cloud scenarios and attenuate the efficiency impact due to inconsistent point cloud densities. The second type of method to addressing density inconsistency in point clouds involves filling in missing surface data through a process known as completing incomplete point cloud surfaces. For the problem of domain gaps in 3D point clouds acquired by different LiDAR sensors, Yi et al. [27] recovered the underlying 3D surface from sparse and incomplete LiDAR point clouds by using a sparse voxel completion network (SVCN). Furthermore, the domain adaptation problem was converted to a 3D surface completion task. The third type of method deals with density inconsistency by modifying the convolution kernel. Sparse invariant Interpolated Convolution (InterpConv) [28] operation was proposed by Mao et al. to normalize the points in the neighborhood of each kernel weight vector, ensuring the density invariance of InterpConv. The utilization of the fuzzy mechanism in the 3D point cloud spherical convolution kernel was proposed by Lei et al. [29]. They designed a fuzzy kernel that eliminates the traditional discrete spherical kernel weight assignment problem and exhibits natural robustness to missing data and point density.



(a) Permutation invariance (b) Density inconsistency

FIGURE 2. Structural characteristics of point clouds, where (a) shows the permutation invariance of the point cloud. For the three points in the figure, six different permutations exist, but the shapes they express are the same. (b) demonstrates the density inconsistency of the LIDAR point cloud for autonomous driving, where the points in the red box in the figure are dense, while the points in the blue box are sparse, which is caused by the different distance between the sensor and the target location.

The semantic segmentation methods proposed from the structural characteristics of point clouds have solved some of the problems of poor segmentation caused by structural characteristics. In order to enhance the comprehension of semantic information contained within point clouds, numerous exceptional methodologies have been proposed by researchers from diverse perspectives. This paper will organize and analyze these approaches in the upcoming sections. The succeeding parts of this paper will be presented in the following sequence: Section II introduces the public datasets and the evaluation metric for point cloud segmentation. Section III explores semantic information extraction from different data forms in point cloud semantic segmentation. Section IV is devoted to the quantitative analysis of the experimental results obtained from the methods discussed in the preceding sections. Furthermore, Section V presents several perspectives on the shortcomings of the current point cloud semantic segmentation and provides an outlook on future development.

II. DATASETS AND INDICATORS

This section introduces the commonly used 3D point cloud semantic segmentation datasets and gives the point cloud segmentation evaluation indicators. These datasets include indoor scenes, urban streets, and autonomous driving scenes, providing data support for developing point cloud segmentation models. For datasets with different scenarios, diverse

TABLE 2. Basic information on common public datasets for point cloud semantic segmentation. Data quantity in the table is measured in millions. "-" indicates that data is unavailable.

Datasets	Years	Scenarios	Marking method	semantic categories	Data quantity	Acquisition sensors
S3DIS [8]	2016	Indoor scenes	Pointwise	13	215	Matterport scanner
ScanNet [32]	2017	Indoor scenes	Voxel	20	-	Structure sensor
Semantic3D [33]	2017	City streets	Pointwise	8	4009	Terrestrial Laser Scanner
SemanticKITTI [34]	2019	Automatic driving	Pointwise	19	4549	Velodyne HDL-64E

evaluation metrics are essential to more accurately assess the excellence of semantic segmentation models. Efficient and diverse datasets and targeted evaluation metrics can provide a solid foundation for theoretical research and facilitate the emergence of new methods.

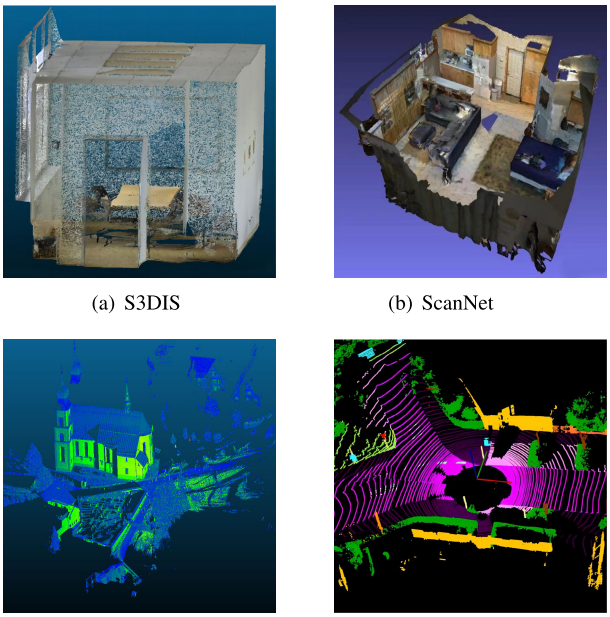
A. DATASETS

Dedicated to the development of point cloud segmentation techniques, several research institutions provide open and reliable datasets. This subsection presents several datasets most commonly used by researchers, including S3DIS [8], ScanNet [32], Semantic3D [33], and SemanticKITTI [34]. Table 2 lists the basic information of these datasets.

S3DIS: [8] The S3DIS dataset significantly contributes to computer vision, specifically for indoor scene analysis. Developed by a dedicated research group at Stanford University, it comprises five vast indoor areas spanning three distinct buildings, encompassing 6020 square meters and featuring over 215 million data points. The dataset offers diverse scenes, and Fig.3(a) represents a conference room scenario for Area-5 of this dataset. The dataset utilizes a Matterport scanner to scan 272 rooms and automatically generate point clouds for the site. The semantic tag is divided into 12 semantic categories, i.e., structural elements (ceiling, floor, walls, beams, windows, doors), ordinary furniture (tables, chairs, sofas, bookcases, wood paneling), and clutter, for a total of 13 categories.

ScanNet: [32] The ScanNet dataset is valuable for computer vision and 3D modeling research. It was developed jointly by Princeton University, Stanford University, and the Technical University of Munich. It is a dataset of indoor scenes in a natural environment, and Fig.3(b) shows an indoor scene from this dataset. This dataset is widely used for semantic voxel annotation, 3D object classification, and CAD model retrieval. It has 20 object class labels for the semantic segmentation task and 1 class for free space; each object class label corresponds to a furniture class.

Semantic3D: [33] The Semantic3D dataset is a large outdoor LiDAR dataset developed by researchers at ETH Zurich, Switzerland. The dataset comprises 30 ground-based LiDAR scans, including about 4 billion manually labeled points. Several scenes, including farms, town halls, sports fields, castles, and market squares, are covered, and some of the scene visualization results are shown in Fig.3(c). Semantic3D contains eight semantic categories, such as vegetation, buildings, cars,



(a) S3DIS (b) ScanNet

(c) Semantic3D (d) SemanticKITTI

FIGURE 3. Visualization of point clouds semantic segmentation dataset. (a) a conference room in the S3DIS dataset [8], (b) a living room in the ScanNet dataset [32], (c) a street scene in the Semantic3D dataset [35], and (d) an autonomous driving scene in the SemanticKITTI [34].

etc., and is one of the largest cloud semantic segmentation datasets available for outdoor attractions.

SemanticKITTI: [34]The SemanticKITTI dataset is a large outdoor scene dataset based on automotive LiDAR constructed by researchers at the University of Bonn, Germany. It shows traffic in the city center of Karlsruhe, Germany, residential areas, freeway scenes, and rural roads. The dataset was annotated with 19 semantic categories, and the annotated scene is shown in Fig.3(d). The data collection comprises 22 sequences, with the first 11 sequences (0 to 10) designated for use in training and the remaining 11 sequences (11 to 21) reserved for testing.

B. INDICATORS

In order to fairly reflect the superiority of the model, the effectiveness of point cloud segmentation needs to be evaluated from different aspects using some well-known evaluation metrics. Commonly adopted performance measures in prior studies encompass Overall Accuracy (OA),

TABLE 3. Comparison of the advantages and disadvantages of different types of point cloud semantic segmentation methods.

Method category	Advantages	Disadvantages
Methods based on dimensionality reduction	2D image segmentation technology is relatively mature	Loss of spatial information due to dimensionality reduction
Methods based on voxelisation	3D CNN can be applied directly to extract features	High computing power and memory requirements
Methods based on primitive points	Simple data pre-processing and complete structural information	Susceptible to noise and missing data
Methods based on multiple data formats	Obtain more comprehensive information for segmentation	Modal gaps are a difficult issue to resolve

Mean Intersection over Union (mIoU), memory consumption, and computation duration. Equations(1), (2), and (3) show the mathematical expressions of OA, IoU, and mIoU, respectively.

$$\mathrm{OA} = \frac{1}{\mathrm{N}} \sum_{i=0}^{k} TP_i \tag{1}$$

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{mIoU} = \frac{1}{\mathrm{k}+1} \sum_{i=0}^{k} IoU_i \tag{3}$$

TP, FN, and FP in (2) represent true positives, false negatives, and false positives, respectively. Expressly, for a given category α, T and F represent correct and incorrect classification, P and N represent a classification into α and non-α categories, respectively, i.e., In the context of point cloud segmentation, TP denotes the number of points that are accurately classified as category α, while FP represents the number of points that belong to category non-α but are incorrectly predicted as α. On the other hand, FN represents the number of non-α points that are correctly predicted as such. N is the total number of points in the point cloud, and k denotes the number of semantic categories. Although Overall Accuracy (OA) can provide an overall assessment of the classification performance, it could be biased towards categories with more points in the scene. This is because the impact of a wrong prediction for a class with fewer points is very small for OA evaluation criteria, while a correct prediction for a class with fewer points is very important. To mitigate this issue, the Mean Intersection over Union (mIoU) is more suitable as it captures the prediction accuracy of each semantic category, regardless of its point count.

III. SEGMENTATION METHOD

Due to advancements in deep learning technology, point cloud segmentation algorithms have improved accuracy in recent years compared to traditional machine learning methods [36], [37], [38]. This section categorizes current point cloud segmentation methods based on deep learning, considering the data processing format. The categorization includes four groups: methods based on dimensionality reduction, methods based on voxelization, methods based on primitive points, and methods based on multiple data formats. The advantages and disadvantages of these four types of methods are summarized in Table 3. Fig.4 depicts a detailed classification of these methods; Fig.5 shows a general network structure for extracting semantic information using four different forms of point cloud representation. The following subsections classify and summarize representative network models proposed in recent years.

TABLE 4. Overall description of the dimensionality reduction-based methods presented in this subsection.

Types	Technologies	Methods
Projection	Spherical projection	SqueezeSeg [39], SqueezeSegV2 [40], SqueezeSegV3 [41], Milioto's method [42], FPS-Net [43]
	BEV projection	VolMap [44], SalsaNet [45], PolarNet [46]
	Hybrid projection	MPF [47]
Multi-view	Virtual camera views	Lawin's method [48], SnapNet [49], SnapNet-R [50]
	Virtual tangent planes	Tatarchenko's method [51]

A. METHODS BASED ON DIMENSIONALITY REDUCTION

The point cloud segmentation methods based on dimensionality reduction benefit from the maturity of 2D image segmentation techniques. These methods reduce the 3D point cloud data into 2D image data that can be processed directly using mature 2D image segmentation methods and then remap the segmentation results into 3D space for point cloud segmentation. These methods can be further classified into two categories: one is dimensionality reduction by projection, and the other is multi-view dimensionality reduction. Table 4 lists the overall description of the dimensionality reduction-based methods presented in this subsection.

1) PROJECTION-BASED METHOD

The most frequently employed projection for point cloud segmentation are spherical projection, bird's eye view projection, and hybrid projection. Spherical projection maps the point cloud onto a sphere surrounding it; the bird's eye view projection method involves projecting the point cloud onto a two-dimensional plane from a top-down perspective; hybrid projection techniques combine multiple projection techniques to generate a more comprehensive representation of the 3D point cloud.

Spherical projection is a dimensionality reduction method by mathematically processing spherical coordinates to obtain

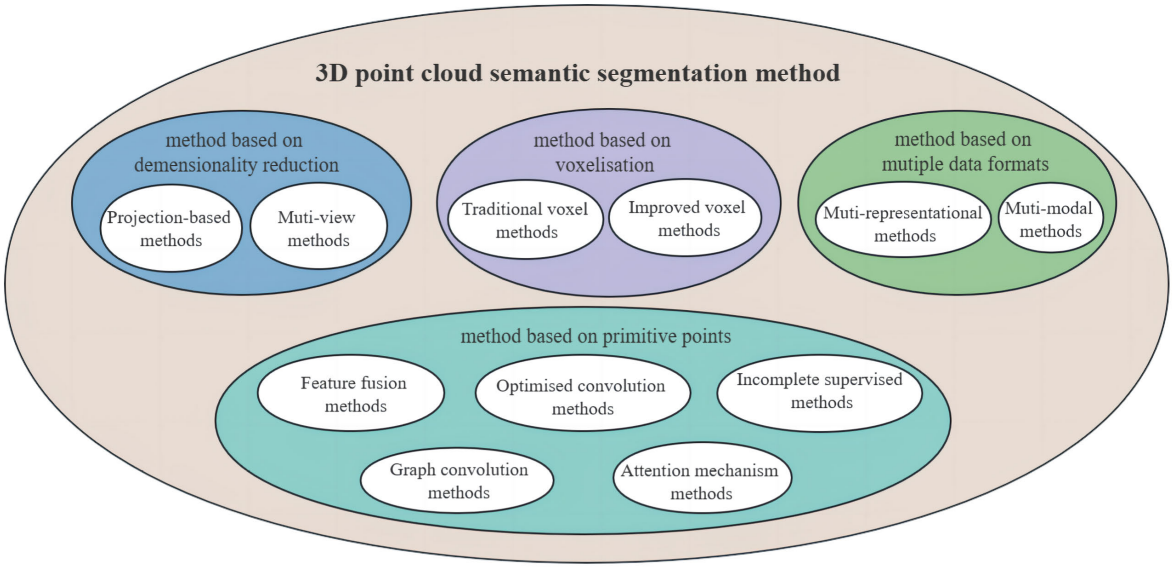


FIGURE 4. An overall overview of all methods. The methods are divided into four categories according to the data processing format, where each category further refines these methods according to the segmentation technique.

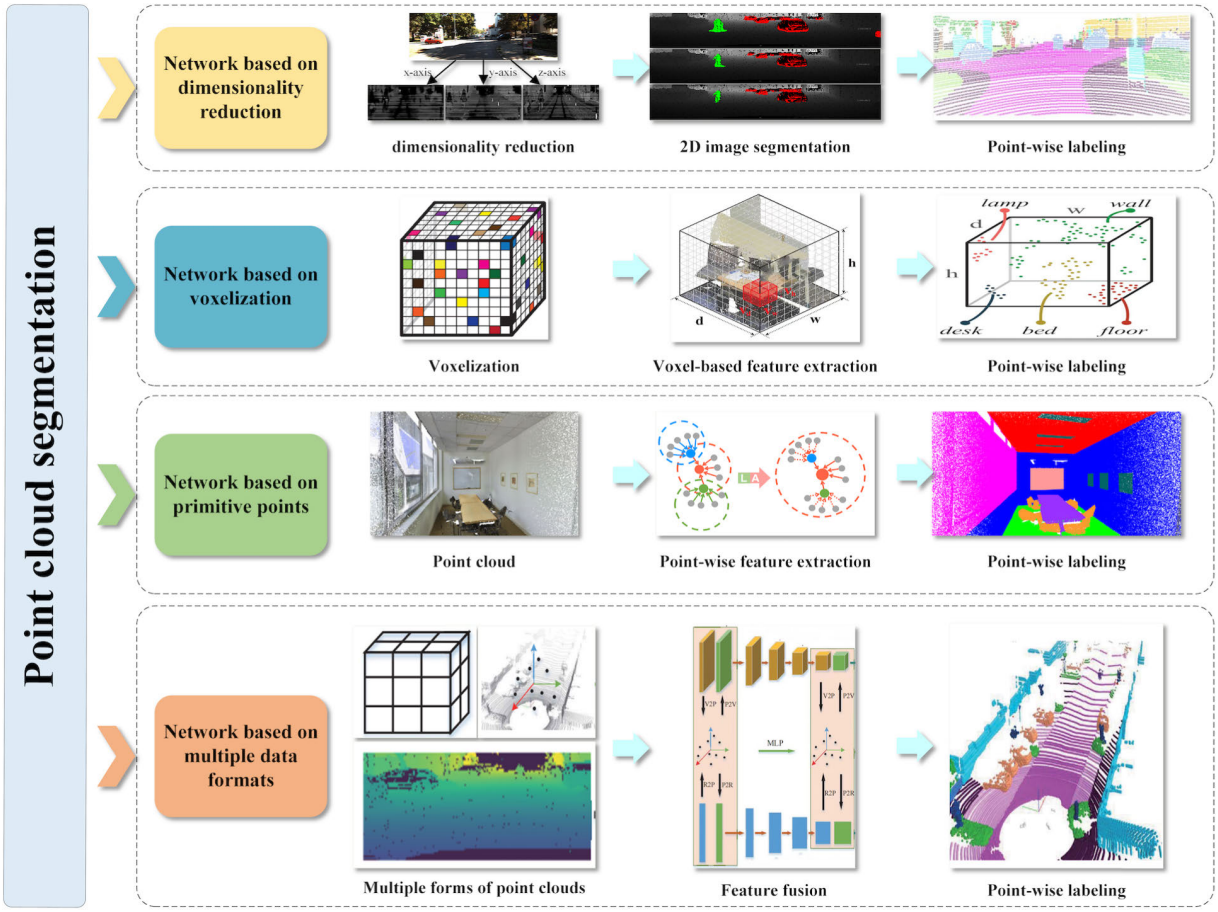
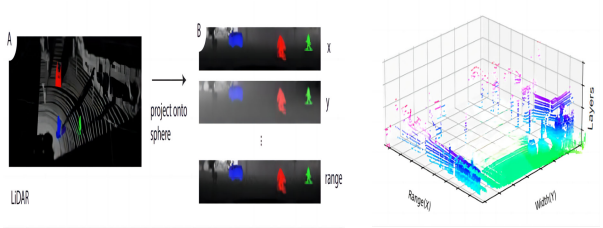


FIGURE 5. A comprehensive outline for performing semantic segmentation using four different types of point cloud representations, highlighting their structural characteristics.

pixel coordinates. Wu et al. [39] introduced SqueezeSeg, a novel end-to-end network that employs SqueezeNet [52] as the base architecture. The network projects the LiDAR point cloud onto a sphere (Fig.6), producing a dense grid-based representation that serves as the input to the CNN. The first few layers in the SqueezeSeg [39] network significantly impact

segmentation accuracy due to dropout noise, so the new Context Aggregation Module (CAM) is proposed in SqueezeSegV2 [40] to reduce the sensitivity to dropout noise. A mask channel is added to the projection image to further improve the model's accuracy. Due to the significant variation in feature distribution across different image locations in LiDAR images and the limitation of standard 2D convolution in capturing only local features in specific regions of the image, SqueezeSegV3 [41] introduces a novel technique called SAC. SAC utilizes different filters to extract projected image features for different parts of the image, allowing the network to capture the varied features present in the LiDAR images effectively. Based on SqueezeSeg [39] and SqueezeSegV2 [40], Milioto et al. [42] used an efficient GPU-based k-nearest neighbor search post-processing step to alleviate discretization errors and fuzzy inference output to address the bleeding phenomenon when transferring semantic labels from 2D projection images to 3D point clouds. Xiao et al. [43] conducted a study and discovered that modality gaps exist in the different message images produced by spherical projection. Consequently, directly overlaying these channels as regular images often leads to sub-optimal segmentation results. An end-to-end network based on spherical projection, FPS-Net, was designed to solve this problem. Unlike the above methods, FPS-Net uses an entirely new structure where each channel image is first learned individually. Then the learned features are fused and applied to the LIDAR point cloud segmentation.

Spherical projections are prone to quantization errors, such as different points in a point cloud being projected onto the same 2D grid, even if they are far apart. Such errors can reduce the accuracy of subsequent processing. In order to tackle this issue, Beltrán et al. [53] proposed the utilization of the Bird's Eye View (BEV) projection, which can somewhat alleviate the factor mentioned above. The BEV projection provides a top-down view of the point cloud while preserving scale and distance information. It is widely used in LIDAR detection [54], [55], [56] and recently in point cloud segmentation. Drawing on the method of VoxelNet [57] and the idea of bird's-eye projection, the LiDAR point cloud is represented as a voxel grid map in VolMap [44]. Then the points on the X- and Y-axis component planes are discretized into a projection grid with a specific resolution, the Z-axis is represented as a channel layer of the projected image to avoid losing height features, and the visualization results are shown in Fig.6(a). Aksoy et al. [45] proposed a projection-agnostic model, SalsaNet, which can input both spherical and bird's eye view projections for semantic segmentation. For the first time, comparative experiments demonstrated the effectiveness of a bird's eye view and spherical projections in segmenting different objects. Because of the more compact spherical projection, small objects are better when using spherical projection for semantic segmentation. In contrast, large objects perform better by the bird's-eye view projection as an input. To improve the effectiveness of bird's-eye view projection for segmenting fine-grained semantic classes, Zhang et al. [46] proposed a new 2D representation of point



(a) Spherical projection (b) BEV projection

FIGURE 6. The 3D structure of the point cloud is transformed into a 2D format by different projection methods. Where (a) is the spherical projection method used by SqueezeSeg [39], and (b) is the bird's eye projection method used by VolMap [44].

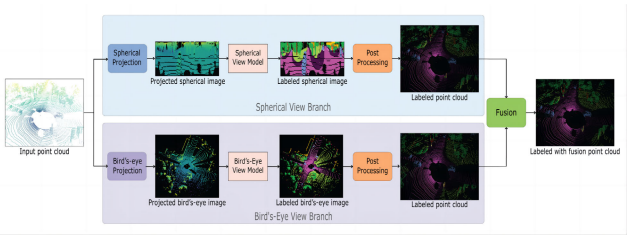


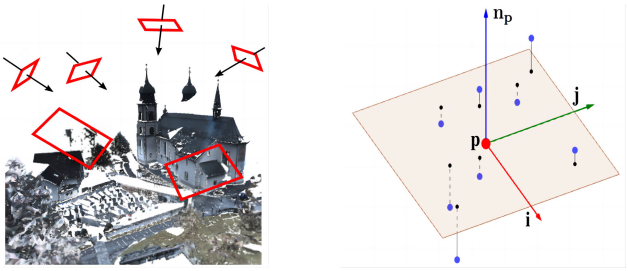
FIGURE 7. The pipeline of MPF [47]. The upper branch of the MPF is responsible for performing semantic segmentation on the original point cloud using spherical projection, while the lower branch employs bird's eye view projection for the same purpose. The final output of the MPF is the semantic segmentation of the original input point cloud, achieved by integrating the results obtained from both branches.

cloud data, the polar coordinate bird's-eye view representation, which distributes points more evenly and reduces the burden on the predictor.

While spherical-based and bird's-eye view-based projection techniques demonstrate effectiveness, the information projected onto the 2D plane by the spherical projection and the bird's-eye view projection differs. Based on this assumption, a Multi-Projection Fusion (MPF) framework was proposed by Alnaggar et al. [47] to compensate for the intrinsic information loss in single-projection methods by fusing multiple projections (Fig.7). The framework uses two independent and efficient 2D full convolutional models to segment the spherical projection and the bird's eye view projection, respectively, with MobileNetV2 [48] as the lightweight skeleton for the spherical projection model network and a lightweight modification of U-Net [49] as the skeleton for the bird's eye view model. Finally, a soft voting mechanism is used to fuse the segmentation results of the spherical and bird's-eye projections. The experimental findings indicate that mixed projection techniques can yield superior segmentation results compared to single projection techniques.

2) MULTI-VIEW METHOD

The multi-view method obtains a series of 2D views containing different side information from different directions of the point cloud by simulating different perspectives of human observation of the object. Semantic segmentation is



(a) Generate snapshots from different perspectives

(b) Tangential projection

FIGURE 8. Multi-view reduced-dimensional point cloud. Where (a) is SnapNet [50] acquiring a 2D snapshot of a 3D target from different views by multiple virtual cameras, and (b) is tangent convolutions [51] downscaling the surrounding points with the views of different points in the point cloud.

performed at the pixel level based on these views using a mature 2D segmentation framework. Finally, the labels obtained from the segmentation on the 2D views are back-projected onto the 3D point cloud to achieve the effect of the point clouds' semantic segmentation.

To avoid the limitations of 3D CNN, Lawin et al. [48] projected point clouds into multiple virtual camera views for the first time based on the idea of multi-view dimensionality reduction. In order to provide more spatial and textural information to the point cloud semantic segmentation, information such as color, depth, and surface normals are also added to the projection view. Using the same projection strategy, Boulch et al. [49] proposed SnapNet, which generates a composite image of RGB views and geometric features by selecting many suitable snapshots of the point cloud (the selection of snapshot views is shown in Fig.8), then using a fully convolutional network to label each group of 2D snapshots pixel by pixel and finally back-projecting the semantic segmentation results into the original point cloud. Based on SnapNet [49], Guerry et al. [50] proposed an improved multi-view convolutional neural network, SnapNet-R. In contrast to the above approach, Tatarchenko et al. [51] form a series of virtual tangent planes using each point in the point cloud as a tangent point and project the local points onto these virtual tangent planes to produce a set of tangent images, each of which is treated as a regular 2D mesh supporting planar convolution(As shown in Fig.8(b)), and then use the proposed tangent convolution to segment the point cloud.

The point clouds' segmentation method based on dimensionality reduction has the advantages of fast speed and low memory consumption, so it is often used in some embedded applications that require high real-time performance, such as autonomous driving. At the same time, the point cloud after dimensionality reduction also brings the following problems: (1) several different points in the space may be projected onto the same grid in the 2D plane, and (2) the spatial geometric structure of the point cloud will be damaged. Due to these problems, the accuracy of point cloud segmentation is reduced.

TABLE 5. Overview of the semantic segmentation method for point clouds based on voxelisation.

Types	Technologies	Methods
Traditional voxel	Dense voxel division	Huang's method [58], SEGCloud [59], VV-Net [60]
Improved voxel	Efficient data structure representations	OctNet [61], Kd-tree [62]
	Sparse convolution operations	SSC [63], Choy's method [64]
	New voxelization techniques	SplatNet [65], Cylinder3D [66], LessNet [67]

B. METHODS BASED ON VOXELISATION

Given the excellent results of CNNs in image semantic segmentation and the similarity between voxels and pixels regarding data organization, the point clouds were converted into voxels by researchers, who subsequently introduced some 3D-based neural network models to perform semantic segmentation on the point clouds, which is known as voxel-based methods. In the voxel-based methodology, the initial step involves breaking down the entire point cloud into some voxels, followed by utilizing a 3D convolutional neural network (3D CNN) to extract features and employ the employment of voxels as the fundamental unit for predictive classification. The voxel-based techniques can be categorized into traditional and improved voxel methods. The voxel-based point cloud semantic segmentation approaches introduced in this subsection are enumerated in Table 5.

1) TRADITIONAL VOXEL METHODS

The traditional dense voxel-based methods aim to divide the 3D space containing point clouds into ordered voxels and then perform convolutional operations using standard 3D CNN to extract point cloud features. When segmenting voxels, the points within the same voxel are divided into the same semantic labels. VoxNet [68] dominates this class of methods, but it is used for object recognition. Later researchers proposed a series of traditional dense voxel-based semantic segmentation methods for 3D point clouds.

Inspired by the success of deep learning of 2D images and the 3D CNN proposed by Ji et al. [69] for human action recognition in video data, Huang et al. [58] first proposed a full 3D voxel-based convolutional neural network for solving the point cloud labeling problem. To improve the accuracy of voxel segmentation, the authors also propose that when sufficient computational resources are available, the centers can be shifted, the segmentation process can be re-run, and finally, a voting mechanism is used to decide which label to assign to each point. To address the coarse voxel prediction of Huang's method, Tchapmi et al. [59] introduced SEGCloud to effectively combines neural networks, trilinear interpolation [70], and FC-CRF to achieve point-level segmentation of the 3D point cloud. The proposed network architecture, as depicted in Fig.9, is thoughtfully designed to optimize semantic segmentation performance. Typically, each voxel

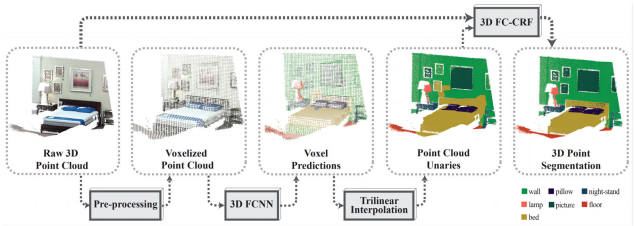
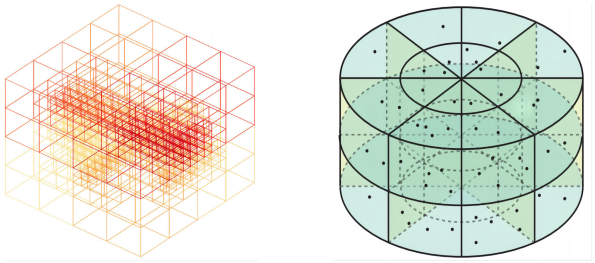


FIGURE 9. Segcloud's pipeline [59]. First, the point cloud is voxelized, and then coarsely semantically annotated at the voxel level by 3D FCNN. This coarse output is transferred back to the original 3D point representation by Trilinear Interpolation, and the obtained 3D point scores are used for 3D FC-CRF inference to produce the final fine-grained results.



(a) Octree representation (b) Cylinder partition

FIGURE 10. Improved point cloud voxel partitioning. (a) describes an octree representation [61] to divide voxels, which allows more and smaller voxels to be obtained where points are denser so that computing power can be focused on proper places. (b) describes the way of dividing voxels by cylinders [66], which ensures a more uniform distribution of points within each voxel in the case of varying densities between distant and near point clouds of the autonomous driving environment.

contains only Boolean occupancy states (i.e., occupied or unoccupied) and not other detailed point distributions, so only little details can be captured. Therefore, Meng et al. [60] proposed a new point cloud segmentation network, VV-Net, whose key idea is to efficiently encode the point distribution within each voxel.

2) IMPROVED VOXEL METHODS

Although representing point clouds as voxels to solve point cloud segmentation is conceptually simple, many potential challenges still need proper algorithmic optimization. Firstly, a traditional dense voxel representation would quickly exceed the memory limit of a computer, and secondly, it would consume too much computational time. Therefore, the main goal of later research is to address the extensive time overhead and memory cost, with specific research results including efficient data structure representations [61], [62], sparse convolution operations [63], [64], and new voxelization techniques [65], [66].

The high activation of the traditional dense voxel method occurs only near the object's boundary. At the same time, the 3D data is usually sparse, and the computational resources consumed by each voxel are equal, so the traditional dense voxel method is challenging to extend to high-resolution voxel scenes. In order to focus the calculation on valuable areas, Riegler et al. [61] proposed a new structural representation of point cloud data, the octree representation, shown in Fig.10(a). In octree representation, the deeper the tree is in the place with high point density, the smaller the voxels are divided in the place with high point density so that the computing power can be concentrated in the place with dense points. In addition to improving the structure of voxels, some researchers can also improve the convolution operation to reduce the computation. Graham et al. [63] introduced Submanifold Sparse Convolutional (SSC) to perform sparse point clouds. Unlike previous implementations of sparse convolutional networks [71], [72], this convolutional operation has the same number of active loci at each layer, so sparsity remains constant, thus avoiding the "submanifold" dilation phenomenon, and it is feasible to train deep networks using this operator. The LiDAR point cloud of outdoor scenes shows that the density in the immediate area is much larger than in the outlying area. Suppose the traditional equal volume voxel is used to divide the LiDAR point cloud of outdoor scenes will be an uneven distribution of more points in the near-area voxels and fewer points in the far-area voxels. To solve the above phenomenon, Zhou et al. [66] proposed a new voxelization method, i.e., 3D cylinder partition, as shown in Fig.10(b), which can allocate more voxels to the dense points in the near region. Based on this, Cylinder3D is developed, whose 3D convolution operation is inspired by the sparse convolution technique employed in SECOND [73], resulting in a reduction of the computational resources required for feature extraction. Based on cylinder voxel segmentation, Feng et al. [67] proposed LessNet to better encode voxel features by aggregating point features within voxels without querying neighboring points to improve the semantic information.

The existing voxel-based semantic segmentation models address the disordered and unstructured characteristics of point clouds. However, voxel-based methods have many drawbacks, such as the size of voxels cannot be readily determined, high arithmetic power and memory requirements, and blurred segmentation boundaries, which limit the further development of voxel-based methods.

C. METHODS BASED ON PRIMITIVE POINTS

Since dimensionality reduction-based methods are prone to spatial structure loss, and voxel-based methods are computationally resource intensive and require high memory capacity, in recent years, primitive point-based methods guided by PointNet [11] do not require excessive pre-processing of point cloud data have become a new research hotspot. Primitive point-based methods can be classified into the following five categories: feature fusion methods, graph convolution methods, optimized convolution methods, attention mechanism methods, and incomplete supervised methods. A summary of the ideas of these five categories of primitive point-based methods and some representative specific methods corresponding to each type of technology are shown in Table 6.

TABLE 6. The classification of methods based on primitive points, the approximate idea of each category, and the corresponding representative methods of recent years.

Method category	General idea	Representative methods
Feature fusion methods	Capture the semantic information of point clouds by aggregating local and global information or fusing features at different levels.	PointNet [11], PointNet++ [7], Cheng's method [74], SalsaNext [75], Qiu's method [35], LG-Net [76], Nie's method [77], Lu's method [78], Fan's method [79], LGENet [80]
Graph convolution methods	The point cloud is represented as a graph structure, and feature extraction is performed directly on the graph structure to capture the semantic information.	RGCNN [81], SPG [82], Liu's method [83], DeepGCNs [84], Lin's method [85], LDGCNN [86], MuGNet [87], Li's method [88]
Optimised convolution methods	Improve the traditional CNN by adapting it to the structural features of the point cloud and running it directly on the primitive point cloud.	piderConv [89], PCC [90], Komarichev's method [91], Kumawat's method [92], PointConv [93], Lei's method [94], diffConv [95]
Attention mechanism methods	Make the neural network ignore irrelevant information and focus on the critical information of the task to enhance learning ability.	GAPNet [96], PyramNet [97], Point2Sequence. [98], LAE-Conv [99], DAPnet [100], Zhao's method [101], S3Net [102], Tang's method [103]
Incomplete supervised methods	The semantic category of each point is obtained through information propagation by feature learning on weakly labeled data.	Liu's method [104], Zhang's method [105], Jiang's method [106], HybridCR [107], Yang's method [108], Zhang's method [109], Shi's method [110], GaIA [111], Unal's method [112], Wei's method [113], Ren's method [114], PNAL [115]

1) FEATURE FUSION METHODS

In order to better achieve fine-grained point cloud segmentation, the structural features between points are the critical consideration. At present, a multitude of network models has been devised to capture the intrinsic connections within point clouds. These models leverage techniques such as aggregating information from neighboring points and fusing regional features across different levels to enhance the accuracy of semantic segmentation for point clouds.

PointNet [11] is the first network model that uses primitive points for point cloud segmentation. However, they only partially consider the point clouds' local features and only splice a global feature obtained by maximum pooling after the feature vector of each point, resulting in some fine-grained wrong segmentation in the segmentation results. PointNet++ [7] addresses this problem using a layered neural network with many set abstraction levels (Fig.11). In PointNet++, the sampling&grouping layer selects a subset of points from the input data, which are then organized into a localized region for further processing. The local region is transformed into a feature vector using mini-PointNet within the Pointnet layer. Like the idea of CNN, the ''seen'' range of ''local area'' becomes broader after multi-level feature extraction, and the fusion from local features to global features is gradually completed. Using PointNet [11] and PointNet++ [7] as a reference, researchers later proposed many feature fusion methods by combining local and global features. Cheng et al. [74] introduced a novel cascaded non-local network model composed of three types of non-local blocks (Neighborhood-level, Superpoint-level, and Global-level), which collaboratively aggregate local features to enhance point cloud semantic segmentation performance. Drawing upon the advancements made in SalsaNet [45], Cortinhal et al. [75] proposed a novel module that adds a residual dilated convolution stack to the front end of the encoder in the network pipeline. This module effectively fuses receptive fields at multiple scales to capture a comprehensive range of contextual information. Similarly, Qiu et al. [35] introduced a bilateral block to perceive the structural features of nearby points, which enhances the local environment by using the geometric and semantic features of the surrounding points. During the downsampling of point clouds, it is common for the local structure to be compromised. The DFC module, developed by Zhao et al. [76], enables retaining important features as point sets decrease in size. This module is combined with the GCM module to learn long-term dependencies and compensate for the lack of general perceptual information in local features. The resulting LG-Net, consisting of these two modules, is particularly effective for point cloud detail segmentation. To investigate how to effectively integrate features at different scales and stages in a point cloud segmentation network, Nie et al. [77] devised a scale pyramid architecture that allows information to flow more freely and systematically. To address the aggregation of different category points, Lu et al. [78] suggest utilizing distinct aggregation methods for data within the same category and across different categories and present a customized module called Category Guided Aggregation. The fusion of multi-scale features in large-scale point clouds has been an active research area, with notable successes achieved by researchers. Fan et al. [79] introduced a learnable module SCF, which effectively extracts spatial contextual features from voluminous point clouds. SCF comprises three key constituent blocks: the LPR block and the DDAP block, which capture distinct local features, and the GCF block, which extracts global semantic features. Lin et al. [80] introduced LGENet, which employs 2D and 3D point convolution to extract features and learn local geometries for ALS point cloud segmentation, followed by a global encoder to utilize this contextual information.

2) GRAPH CONVOLUTION METHODS

Point clouds and graphs share similarities in being unstructured and sparse. The graph convolution method combines convolution operations with graph structure representation to

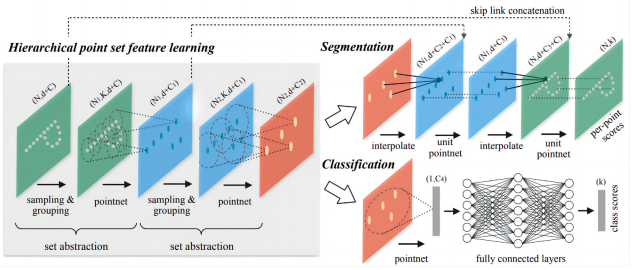


FIGURE 11. The pipeline of PointNet++ [7]. Firstly, the point cloud features are aggregated to the sampled partial points by multiple Set Abstraction levels, then these points with overall features are combined with the original points for interpolation operation, and finally, the semantic class of each point in the original point cloud is obtained.

enable convolution neural networks to operate on the graph structure and capture dependency relationships, leading to a more comprehensive understanding of the underlying relationships.

With the development of graph convolutional neural networks (GCNs), several researchers have used GCNs for the segmentation task of point clouds. Te et al. [81] first used GCNs for point cloud segmentation tasks and proposed the RGCNN. RGCNN takes the features of points as graph signals, with the point cloud feature matrix and adjacency matrix as inputs. Due to the complexity of the graphs constructed using point clouds, Landrieu et al. [82] introduced the concept of a Super Points Graph (SPG) for effectively addressing semantic segmentation challenges associated with processing massive point clouds. SPG considers each geometric shape after geometric partitioning as a super point to construct a super point graph. It can provide a detailed description of the interconnections between adjacent targets, effectively solving the problems of too-independent operation of each point and lack of contact between points. Liu et al. [83] propose a dynamic point aggregation module based on GCNs to overcome the limitations of previous methods that only sample and group points in Euclidean space resulting in limited adaptability to different scenarios. This approach allows for a more flexible and robust hierarchical point set learning model than those relying on fixed point aggregation strategies. To tackle the vanishing gradients issue that restricted traditional GCNs, Li et al. [84] introduced innovative concepts such as residual connections, dense connections, and dilated convolutions into GCNs. These techniques enable the creation of deeper graph neural network models better suited for complex data. Combining the idea of multi-scale feature fusion, Lin et al. [85] designed a novel graph max pooling operation based on GCNs and applied it to the 3D-GCN network they constructed to summarize features at different scales. Due to the shape and weights of each kernel in 3D-GCN being learnable during training, it is robust to the movement and scaling changes of 3D point clouds. Point cloud semantic segmentation models based on GCNs tend to have an enormous time complexity, and later, researchers have proposed methods dedicated to improving the training speed of the models. Based on the improvement of DGCNN [16], Zhang et al. [86] removed the transform network in DGCNN to reduce the size of the network and proposed a linked dynamic graph CNN (LDGCNN) to classify and segment the point clouds directly. Aimed at the disorder and non-uniformity of point clouds, Xie et al. [87] proposed MuGNet, a framework with graph convolution that effectively converts point clouds into graphical representations with reduced computational requirements. The computational effort is reduced by using the GCNs on preformed point cloud graphs, and the segmentation accuracy is maintained by using bidirectional networks that fuse different resolution feature embeddings. From a mathematical theory perspective, Li et al. [88] proposed an improved KNN search and MLP algorithm to optimize the computational process of GCNs, which reduces the time and space complexity of existing GCNs.

3) OPTIMISED CONVOLUTION METHODS

Thanks to its local connectivity and weight sharing, the standard convolutional neural network performs remarkably well in tasks such as image recognition. However, applying them to 3D point clouds that lack stable structure is still challenging. To address this situation, some researchers have developed convolutional operations for irregular point clouds to perform point cloud segmentation tasks by improving traditional convolutional processes.

Many researchers have modified traditional CNNs to adapt to unique structures and extract features from point clouds better. To address the challenge of extending the convolution operation from regular lattices to irregular point sets, Xu et al. [89] proposed a new convolution operation called SpiderConv. The SpiderCNN, composed of SpiderConv, incorporates a filter design that combines a simple step function and a Taylor polynomial, with the step function capturing local geodesic information and the Taylor polynomials ensuring expressiveness. Wang et al. [90] proposed a new operator for operating on non-grid structured data, Parametric Continuous Convolution (PCC), whose convolution kernel function is parameterized by a multilayer perceptron and spans the entire continuous domain. Komarichev et al. [91] introduce annular convolution to perform feature extraction directly on 3D point clouds. This new convolution operator can define an arbitrary kernel size on each local annular region and better specify the annular structure and orientation in the computation to accurately depict the geometric properties of the local neighborhood, helping to capture a better geometric representation of the 3D shape. Kumawat et al. [92] proposed the ReLPV block, as a replacement for the conventional 3D convolutional layer, involves extracting phase information from the 3D local neighborhood, resulting in a more efficient capture of phase information and an improved feature representation of the input data. Wu et al. [93] proposed PointConv, which treats the convolution kernel as a non-linear function incorporating local coordinates of each 3D point, accurately approximates filter weights and density functions,

and achieves both permutation and translation invariance, making it an efficient operation for 3D point cloud feature extraction. Unlike the above methods, Lei et al. [94] constructed a new data structure using the sparsity of irregular point clouds and proposed an octree-guided neural network architecture to segment 3D point clouds directly. The octree in this structure differs from the octree in the voxel representation [61], which improves the algorithm's efficiency by coarsening the data hierarchically and avoiding searching for domain points using the KNN algorithm. Lin et al. [95] considered the existence of regular inductive bias through local point feature learning and proposed a new graph convolution method called ''difference graph convolution'' (diffConv), which includes techniques such as density-dilated ball query, Laplace smoothing, and masked attention. By using these techniques, diffConv can achieve better feature learning performance without the constraints of the regular view.

4) ATTENTION MECHANISM METHODS

Recent deep learning tasks have widely adopted attention mechanisms, whose basic idea is to enable neural networks to ignore irrelevant information and focus on crucial details for the study. This technique has been shown to have significant effects in practice. For point cloud segmentation tasks, many researchers have also introduced attention mechanisms to focus on point clouds' fine-grained and critical features, thus improving the segmentation accuracy.

Exploiting the fine-grained semantic features of point clouds is essential for improving the segmentation accuracy. Chen et al. [96] embed a graphical attention mechanism in stacked MLP layers to better note the fine-grained features of point clouds. Zhiheng et al. [97] designed two new modules, Graph Embedding Module (GEM) use the covariance matrix to explore the relationships between points to enhance the local feature representation of the network, and Pyramid Attention Network (PAN) assigns robust semantic features to each point to preserve the delicate geometric features. Furthermore, he proposed PyramNet based on these two modules to better learn point clouds' spatial local geometric features. To highlight the importance of different scale regions in the local division of point clouds, Liu et al. [98] proposed an RNN-based sequence model Point2Sequence. This model divides each local area into multi-scale regions and introduces an attention mechanism to improve the critical regional scale features.

In addition to focusing on the fine-grained features of the point cloud, some attention-based methods also aim to fuse local features with those of the larger region. Feng et al. [99] introduced a point-wise spatial attention module to adaptively integrate local point features and long-range contextual information. Likewise, Chen et al. [100] introduced the DAPnet, which incorporates the point attention module (PAM) and the group attention module (GAM). The PAM utilizes the inter-region correlation of point clouds to assign varying weights, while the GAM enhances the inter-group correlation.

Taking inspiration from the achievement of the Transformer model [116] in both natural language and image processing, Zhao et al. [101] applied it to point cloud processing and proposed the Point Transformer. Point clouds are a collection of vectors in space, matching the Transformer's self-attention operator, which makes Point Transformer naturally superior to other convolutional models for point cloud processing. Cheng et al. [102] proposed S3Net, a novel approach that utilizes a Transformer encoder-decoder structure for point cloud semantic segmentation while employing SIntraAM and SInterAM to capture intra- and inter-feature information. Additionally, the method utilizes the Sparse Residual Tower to process the obtained detailed information and extract global features. Also, based on the encoder-decoder structure, Tang et al. [103] proposed a voxel-based encoder for local and global feature extraction, decoding features and segmenting point clouds by cross-attention and self-attention in a Transformerd-based decoder.

5) INCOMPLETE SUPERVISED METHODS

With the development of point cloud acquisition devices, the acquisition of large-scale point clouds has become more accessible; however, annotating this data at the point level is difficult. Therefore, in recent years, many researchers have started to use incompletely labeled or Coarse-grained scene class-level labeled point cloud segmentation datasets for incompletely supervised model training. They have achieved results comparable to supervised learning methods.

Performing semantic segmentation on sparsely labeled points is the most common incompletely supervised method, and the commonly used strategies include self-training and pre-training. Liu et al. [104] put forward a weakly supervised point cloud semantic segmentation model based on self-training, borrowing from self-training in 2D image understanding. The network performs label expansion by iteratively executing the graph transfer module and combines ''Relation Net'' to learn similar features among super-voxel of complex 3D structures. So only one point per object needs to be labeled for the input point cloud to achieve good segmentation results. To establish the topology of labeled and unlabeled points and to perform efficient information propagation, Zhang et al. [105] introduce perturbation branching and context-aware modules. By constraining the prediction consistency between the perturbed and original data and with the help of the context-aware module, the GCNs is driven to establish the fine-grained graph topology of the point cloud. Building on the success of contrastive loss in self-supervised learning, Jiang et al. [106] introduced guided point contrastive loss, a novel approach that aims to improve feature representation and model generalization in a semi-supervised setting. Similarly, Li et al. [107] propose a hybrid contrastive regularization (HybridCR) framework in a weakly supervised environment. Influenced by the idea of pre-training, Yang et al. [108] improve the shortcoming of failing to integrate spatial information well in PointContrast

[117] and improve efficient learning of 3D data by clustering pre-trained point features in limited annotations scenarios with fine-tuning and active labeling strategies. Based on similar ideas, Zhang et al. [109] used the color information of the point cloud as a self-supervised signal to learn prior knowledge, combined it with local perceptual regularization to learn contextual knowledge, and then initialized the weakly supervised network using pre-trained encoder parameters. Shi et al. [110] introduced a novel approach for training segmentation models on outdoor 3D point cloud sequences with extremely sparse annotations (i.e., only 0.001% of points are labeled). Their approach is based on a spatio-temporal framework that comprises two main modules: the first module is a temporal dimensional matching module, enabling the propagation of pseudo labels across different frames, while the second module is a spatial dimensional graph propagation module facilitating information propagation from the pseudo labels to each frame's point cloud. These modules allow effective training segmentation models on sparsely annotated 3D point cloud sequences. From the perspective of reducing the cognitive uncertainty of unlabeled points, Lee et al. [111] proposed the GaIA, the main idea of which is to reduce the category uncertainty of unlabeled points by the reliable information of labeled points. The specific approach is to calculate the relative entropy between the target point's entropy and its neighbors' entropy to represent the information's reliability by graphical information gain. Furthermore, combined with the proposed ArcPoint loss, embed those unlabeled points with high entropy values into reliable labeled points to reduce the entropy value and then increase the information reliability of unlabeled points. To facilitate the development of Scribble-Supervised methods for point cloud segmentation tasks, Unal et al. [112] published ScribbleKITTI, the first scribble annotation dataset for LiDAR semantic segmentation. Furthermore, it proposed a weakly supervised LiDAR semantic segmentation pipeline based on scribble annotation.

Learning point-level labels using subcloud-level or scene-level labels is one of the incompletely supervised methods for point clouds, which significantly reduces the labeling requirements of the dataset compared to supervised learning. Wei et al. [113] employed a sub-cloud annotation strategy to annotate 3D point cloud datasets and proposed a novel approach to training point cloud semantic segmentation by utilizing weak labels at the cloud level in the original 3D space. This method represents the first instance of using cloud-level weak labels to train point cloud semantic segmentation. Ren et al. [114] combine semantic segmentation and target detection using scene-level labeling, coupling their predictions through cross-task consistency loss to obtain significantly better results than a single-task baseline.

Learning semantic information about point clouds in tags that contain a few errors is also a form of incomplete supervised learning. To address the effect of mislabelling in the dataset on the performance of segmentation models, Ye et al. [115] proposed the Point Noise-Adaptive Learning (PNAL) framework to address this problem. PNAL is noise-rate blind to deal with the unique problem of noise-rate variation in point clouds. The framework generates the best possible labels by introducing point-wise confidence selection, cluster-wise label correction, and voting strategies.

TABLE 7. Overall description of the multiple data formats methods presented in this subsection.

Types	Methods	Ideas
Multi-representational	PVCNN [118]	Extraction of features with different granularity by two branches with different data resolutions.
	SPVNet [119]	Changing the voxel branch in PVCNN [118] to sparse convolution to extract features.
	DRINet [120]	Iteratively run two branches to aggregate and propagate point and voxel features.
	RPVNet [121]	Adaptively select the best feature representation for each point through interactive learning.
Multi-modal	Madawi's method [122]	Converts RGB images into a polar coordinate mapped representation of LiDAR.
	FuseSeg [123]	Represents LiDAR point clouds as range images.
	TSNet [124]	Projecting sparse point clouds into the camera coordinate system.
	2DPASS [125]	Semantic segmentation is done by supervision of pure 3D tags.
	MM-TTA [126]	Obtain cross-channel pseudotags by adaptively selecting pseudotags generated from different modal data.

Compared with the method based on dimensionality reduction and the method based on voxels, the method based on primitive points does not require data transformation and avoids the loss of information and computational complexity. However, semantic segmentation using point cloud data directly requires a higher quality of point cloud data, which is susceptible to noise and missing data. Multiple data representations or multimodal data can be utilized to compensate for the impact of data quality so that the information can complement each other.

D. METHODS BASED ON MULTIPLE DATA FORMATS

For the same segmentation scene, different forms of data representation often contain information that is not parallel. Using multiple data formats as input to a point cloud segmentation model allows meaningful information to be extracted from various sources, which significantly helps improve the segmentation accuracy of the point cloud. We further divide such methods into multi-representational and multi-modal methods. The specific method for multiple data formats described in this subsection is shown in Table 7.

1) MULTI-REPRESENTATIONAL METHODS

In recent years, some researchers have combined the representation of point clouds with different forms of images, voxels, and point sets and proposed hybrid data input

models that have contributed significantly to improving the segmentation accuracy of point clouds, which we call multi-representation methods.

Previously, researchers performed semantic segmentation by projecting point clouds as pictures, converting them to voxels, or directly using the original point cloud. The voxel representation of the point cloud grows cubically with the increase of the input resolution in memory usage. The original point representation requires much time structuring unordered sparse data before feature extraction. Liu et al. [118] proposed Point-Voxel CNN (PVCNN) to combine both advantages, using low-resolution voxel-based branching to extract coarse-grained neighborhood information and high-resolution point-based branching to extract fine-grained point features as a complement to the voxel features. By utilizing a sparse and irregular point representation, this technique achieves efficient representation of 3D input data, while the dense and regular voxel representation is employed for convolution, which dispels the misconception that voxel-based convolution is inherently inefficient. Tang et al. [119] argue that PVCNN [118] can only provide limited voxel representation in large scenes while small objects occupy few voxels, so learning helpful information on voxel-based branches isn't easy. Combining the idea that sparse convolution can provide higher resolution than regular volume convolution [64] and the dual branching of PVCNN, Sparse Point-Voxel Convolution (SPVConv) was proposed to equip the sparse voxel-based branch with a high-resolution point-based branch that can effectively capture intricate details present in vast environments. Inspired by PVCNN [118] and SPVNet [119], Ye et al. [120] proposed DRINet, shown in Fig.12, which consists of two modules, SPVFE and SVPFE. The role of SVPFE is to generate point features from voxel features using the attention aggregation layer, and the role of SPVFE is to generate target-scale voxel features from point features using the multiscale pooling layer. The network inherits the advantages of the Point-Voxel two-branch representation. It runs both branches to iteratively aggregate and propagate point and voxel features to finally learn an advanced feature representation. In addition to fusing voxel and point features, Xu et al. [121] incorporated the range image representation of point clouds into the fusion framework and proposed the RPVNet. This framework transforms pixel and voxel features into point features. This methodology utilizes interactive learning to dynamically choose the optimal feature representation for each point, which is then fused onto points and transferred back to range images and voxels to enhance the features mutually beneficially. Compared with the above methods, RPVNet performs feature interaction at each feature extraction stage rather than through a final simple fusion, making the model more capable of learning.

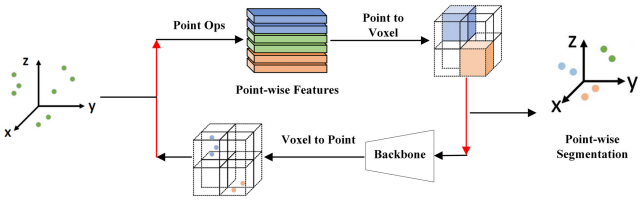


FIGURE 12. The pipeline of DRINet [120]. The upper branch of this model is Sparse Point-Voxel Feature Extraction (SPVFE), and the lower branch is Sparse Voxel-Point Feature Extraction (SVPFE). By iteratively running these two branches, the semantic features are transformed several times in the representation of points and voxels to refine the segmentation results gradually.

2) MULTI-MODAL METHODS

Multimodal learning, machine learning using information from multiple modalities, allows the aggregation of information from multiple data sources to make the representation learned by the model more complete. Over the past few years, there has been a growing interest in the application of multimodal learning techniques to point cloud segmentation tasks in the field of autonomous driving. This surge in interest can be attributed to the practical implications of these techniques.

Multiple types of sensors for autonomous cars can compensate for the shortcomings of a single sensor capturing environmentally relevant information, thus ensuring robust sensing in challenging environments. Most modern commercial autonomous vehicles are equipped with two complementary sensors, a camera, and a LIDAR, with the camera sensor providing color and the LIDAR providing depth information. To make fuller use of this information, Madawi et al. [122] convert RGB images into a polar-grid mapping representation to fuse the color information with the depth information from LIDAR. To better merge RGB images with LiDAR point clouds, Krispel et al. [123] proposed FuseSeg, which represents LiDAR point clouds as range images and fuses multi-modal data using only the standard 2D CNN. The feature warping module in FuseSeg warps the features extracted from the MobileNetV2 [127] branch to the SqueezeSeg [39] branch layer by layer and then splices the warped RGB features with the features from the range image for segmentation. Zhuang et al. [124] argued that the method of Madawi et al. [122] to project images' pixel coordinate into the LiDAR coordinate would result in the loss of appearance information and therefore proposed perspective projection. Then the proposed TSNet is used to learn features from RGB images and projected point cloud, and like FuseSeg, TSNet also fuses features layer by layer by two independent branches. When employing multimodal-based approaches for point cloud segmentation, it is often necessary to access both point clouds and images while ensuring a precise point-to-pixel mapping between the two modalities. The in-vehicle LIDAR has a 360-degree perception range, while the front camera has a narrower viewpoint perception. Yan et al. [125] proposed the 2DPASS method to perform semantic segmentation without strict multi-modal pairing constraints. The method by the MSFSKD block efficiently transfers complimentary 2D features into a 3D network, and the semantic segmentation is done by a pure 3D

TABLE 8. Semantic segmentation performance on S3DIS [8]. "*" indicates the result of cross-validation using 6-fold on that dataset, and "+" indicates the result of validation using "Area-5" on that dataset. "-" indicates that no experiments were performed on that data in the original paper. In the incomplete supervised method, the percentage inside "[]" indicates the percentage of annotation on the dataset.

Technology	Methods	Year	mIoU	OA
Multi-view	Tan-Conv [51]	2018	52.8$^+$	82.5$^+$
Traditional voxel	SEGCloud [59]	2017	48.92$^+$	-
Feature fusion	PointNL [74]	2020	68.4*	88.2*
	Qiu's method [35]	2021	72.2*	88.9*
	SCF-Net [79]	2021	71.6*	88.4*
	LG-Net [76]	2023	70.8$^+$	88.3$^+$
Graph convolution	SPG [82]	2018	58.4$^+$, 61.2*	86.38$^+$, 85.5*
	ResGCN-28 [84]	2019	60.0*	85.9*
	MuGNet [87]	2020	63.5$^+$, 69.8*	88.1$^+$, 88.5*
Attention mechanism	Feng's method [99]	2020	66.3*	88.95*
	Point Transformer [101]	2021	70.4$^+$, 73.5*	90.8$^+$, 90.2*
Multi-representation	DRINet [120]	2021	66.7$^+$	-
Incomplete supervise	Liu's method [104]	2021	[0.06%]55.3$^+$	-
	PSD [105]	2021	[0.03%]48.2$^+$, [1%]63.5$^+$, [100%]65.1$^+$, [1%]68.0*	-
	Zhang's method [109]	2021	[1%]61.8$^+$, [10%]64.0$^+$, [1%]65.9*, [10%]68.1*	-
	HybridCR [107]	2022	[1%]69.2*, [100%]70.7*	-
	GaIA [111]	2023	[1%]66.5$^+$, [0.02%]53.7$^+$, [1%]70.8*	-

TABLE 9. The table shows the experimental results of some methods on the Semantic3D [33] dataset, where "-" indicates that the data did not experiment in the original paper, and the percentage in "[]" in the incomplete supervised methods indicates the percentage of labeled data of the data.

Technology	Methods	Year	mIoU	OA
Multi-view	Lawin's method [48]	2017	58.5	88.9
	Tan-Conv [51]	2018	66.4	89.3
Traditional voxel	SEGCloud [59]	2017	61.3	-
Feature fusion	Qiu's method [35]	2021	75.4	94.9
	SCF-Net [79]	2021	77.6	94.7
Graph convolution	SPG [82]	2018	76.2	92.9
Inomplete supervise	HybridCR [107]	2022	[1%]76.8, [100%]77.4	-
	PSD [105]	2021	[1%]75.8	-
	Zhang's method [109]	2021	[1%]72.6, [10%]73.3	[1%]93.7, [10%]94.0

tagging supervised modal-specific decoder. There are uncontrollable changes in the natural environment, and models learned from previous datasets may not perform well in the new environment. In multi-modal 3D semantic segmentation, Shin et al. [126] proposed MM-TTA to allow models to learn new ''knowledge'' continuously. The model does not need to access source domain training data but quickly adapts to multi-modal test data. It is implemented by introducing two modules: Intra-PG, which updates the modal data models at different rates and generates reliable pseudo labels within each modality, and Inter-PR, which adaptively selects pseudo labels from both modalities. Together, these two modules generate the final cross-channel pseudo label to help test-time adaptation.

IV. EXPERIMENTAL RESULTS

In the previous section, we analyzed the idea of segmentation of point clouds qualitatively in several ways but did not consider the quantitative results of these methods. In this section, the effectiveness of these methods will be analyzed quantitatively based on the commonly used datasets presented in section II, and conclusions will be analyzed based on these experimental results.

S3DIS is a large-scale indoor scene dataset, and the performance of some of the methods presented in this paper on this dataset is shown in Table 8. For the two primary evaluation criteria, mIoU and OA, the feature fusion and attention mechanism-based approaches perform relatively well compared to the traditional voxelization approaches. This result occurs because the S3DIS dataset is an indoor scene dataset, and since there are many details and complex geometric structures, traditional voxel-based methods may have difficulty capturing these details and structures and therefore perform poorly. In contrast, methods based on attention mechanisms and feature fusion can better capture the details and structures in the point cloud because they can handle each point more flexibly, which leads to better extraction of local and global features, thus improving the accuracy of segmen-

TABLE 10. Comparison of quantitative results for some methods on the SemanticKITTI [34] dataset. The blue font in the incompletely supervised methods represents the percentage of annotated data.

Technology	Methods	Year	mIoU
Projection	SequeezeSeg [39]	2018	29.5
	SequeezeSegV2 [40]	2019	39.7
	SequeezeSegV3 [41]	2020	55.9
	RangNet++ [42]	2019	52.9
	FPS-Net [43]	2021	57.1
Improved voxel	Cylinder3D [66]	2020	61.8
	LessNet [67]	2022	58.8
Feature fusion	SalsaNext [75]	2020	59.5
	Qiu's method [35]	2021	59.9
	LG-Net [76]	2023	56.3
Attention mechanism	S3Net [102]	2021	66.8
	MPT-Net [103]	2022	68.2
Incomplete supervise	Jiang's method [106]	2021	[5%]41.8, [20%]58.8, [100%]67.7
	HybridCR [107]	2022	[1%]52.3, [100%]54.0
	Shi's method [110]	2022	[0.001%]44.8, [0.005%]52.3
multi-representation	SPVNAS [119]	2020	66.4
	DRINet [120]	2021	67.5
	RPVNet [121]	2021	70.3
multi-modal	PMF [124]	2021	63.9
	2DPASS [125]	2022	72.9

tation. In addition, the attention mechanism can help the model focus more on essential points and features and reduce the interference of noise and redundant information, thus improving the robustness and generalization ability of the model.

Semantic3D is a large-scale LiDAR point cloud dataset for outdoor scenes, and Table 9 shows the performance of some newly proposed methods on this dataset in recent years. Similar to the results on the indoor scene point cloud semantic segmentation dataset S3DIS, notable performance was achieved using feature fusion methods, as such methods can combine local features and global structure information in the point cloud for semantic segmentation. Notably, the graph convolution-based approach SPG achieved a mIoU score of 76.2% in 2018 alone, possibly because point clouds under the graph structure representation fit their structural features more closely, combined with the graph convolution using the neighborhood information of the point cloud for convolution operations to obtain higher-level features through multi-layer convolution. On the contrary, the performance is poor in multi-view-based methods because these methods cannot capture the stereo structure information of the point cloud well, leading to information loss and error accumulation, thus affecting segmentation accuracy. Although multi-view-based methods cannot beat other methods in terms of segmentation effect, they have the advantage of fast speed. Since incompletely supervised methods do not require many manual labels, these methods have achieved good development in recent years, as can be seen from Table 9, by labeling about 1% of the points to achieve the effect that many completely supervised methods are challenging to achieve.

SemanticKITTI dataset is a large outdoor scene dataset based on automotive LiDAR point clouds, which is mainly used to research algorithms related to the autonomous driving domain. Table 10 shows the quantitative results of some of the methods on this dataset. Due to the relative complexity of the outdoor road scene situation, the mIoU score of the SemanticKITTI dataset is relatively low compared to the experimental results of several datasets above. Due to the high real-time requirements of autonomous driving applications and the fast computational speed of projection-based methods, many of the methods on this dataset can be found to be projection-based. However, purely projection-based methods result in relatively low final mIoU scores due to the loss of spatial structure. Utilizing multi-representation methods or multi-modal data can reduce this impact. Encouraging results have recently been achieved with multi-data format methods, which will be a future direction as the computational power of embedded chips increases.

V. DISCUSSION AND PROSPECT

This paper summarizes four aspects of recent deep learning-based 3D point cloud segmentation methods and compares their performance on the corresponding datasets. Deep learning methods are more efficient in data feature extraction than traditional machine learning methods and achieve better segmentation accuracy. However, from the current research, there are still many unresolved issues, and how to solve these

unresolved problems is the focus of future research. Based on the review of point cloud segmentation methods in the above sections, we will then present a few of our views on the future development of point clouds.

- **Diverse datasets in different fields.** Existing 3D point cloud datasets are mostly limited to object parts, indoor scenes, and urban street scenes. These limited datasets are insufficient for the diverse development of point cloud segmentation technology. Establishing more data-rich, effective, and comprehensive datasets is a prerequisite for developing point cloud segmentation technology. For example, the establishment of digital workshop point cloud datasets can promote the development of smart factories, the establishment of urban remote sensing point cloud datasets can encourage the construction of smart cities, and the establishment of farm point cloud datasets can promote the production of smart farms. In addition, the establishment of diverse annotated datasets is also a measure to facilitate the development of different point cloud segmentation methods.
- **Open-world point cloud segmentation.** Current closed-set point cloud segmentation methods are not robust enough for applications such as autonomous driving because the network's input can only be trained with specific class labels. New classes must be mislabelled, which can have disastrous consequences in complex and changing road scenarios. In order to facilitate more practical applications, a new research direction is to develop a class of Open-world point cloud segmentation methods that can be better adapted to the ''new environment.''
- **Enhancing the effect of boundary segmentation.** Clear segmentation boundaries between different categories are essential for a good algorithm in segmentation tasks. 3D point clouds are prone to confusion during feature fusion between different categories of points that are similar in appearance or spatially adjacent, resulting in poor boundary segmentation. However, because there are relatively few points at the boundary, its segmentation effect only contributes a little to the existing commonly used evaluation criteria such as mIoU or OA. The segmentation impact at the border is far more significant than elsewhere, so proposing new measures for assessing the effectiveness of boundary segmentation and new methods that work well for boundary segmentation is an urgent problem.
- **Domain adaptive learning.** The semantic segmentation models of point clouds are usually trained in a specific scene and used in this scene, for example, the models trained in indoor scenes are difficult to be applied to outdoor scenes, and the domain adaptive learning methods can be explored in the future to make the models applicable to point cloud data in different scenes and improve the generalization ability of the models.
- **More incomplete supervision methods.** Segmentation task has a finer granularity of labels than classification tasks and is very time-consuming and labor-intensive for collecting training datasets. Segmentation of point clouds is not only limited to public datasets. More specific applications require researchers to collect domain-specific point cloud data, and labeling hundreds of millions of points would take far more time than the algorithm itself. Incompletely supervised methods will be the focus of future research in large-scale and very large-scale point cloud segmentation because they require only a few points or weak types of object class labels to be labeled. They can be used for segmenting point clouds, and such methods can also help produce supervised method datasets.
- **Multi-source data fusion methods.** Equipping autonomous vehicles with different data acquisition devices like LiDAR and cameras can provide more comprehensive scene understanding information for autonomous driving. LiDAR point cloud segmentation using multi-source data is a hot topic of current research in autonomous driving. Compared to projection-based methods, the segmentation accuracy of multi-source data fusion methods is greatly improved, guaranteeing safe driving. However, higher computational power is required due to the need to process different data types. As the computing power of in-vehicle chips increases, more point cloud segmentation algorithms for autonomous driving will shift to multi-source data fusion methods.

REFERENCES

[1] J. Park, C. Kim, S. Kim, and K. Jo, ''PCSCNet: Fast 3D semantic segmentation of LiDAR point cloud for autonomous car using point convolution and sparse convolution network,'' *Expert Syst. Appl.*, vol. 212, Feb. 2023, Art. no. 118815.

[2] V. Vanian, G. Zamanakos, and I. Pratikakis, ''Improving performance of deep learning models for 3D point cloud semantic segmentation via attention mechanisms,'' *Comput. Graph.*, vol. 106, pp. 277–287, Aug. 2022.

[3] H. Thomas, B. Agro, M. Gridseth, J. Zhang, and T. D. Barfoot, ''Self-supervised learning of LiDAR segmentation for autonomous indoor navigation,'' in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 14047–14053.

[4] Q. Li, Y. Song, and X. Jin, ''EG-PointNet: Semantic segmentation for real point cloud scenes in challenging indoor environments,'' in *Proc. 16th ICME Int. Conf. Complex Med. Eng. (CME)*, Nov. 2022, pp. 91–94.

[5] Y. Perez-Perez, M. Golparvar-Fard, and K. El-Rayes, ''Segmentation of point clouds via joint semantic and geometric features for 3D modeling of the built environment,'' *Autom. Construct.*, vol. 125, May 2021, Art. no. 103584.

[6] Y. Perez-Perez, M. Golparvar-Fard, and K. El-Rayes, ''Artificial neural network for semantic segmentation of built environments for automated Scan2BIM,'' in *Proc. Comput. Civil Eng.*, Jun. 2019, pp. 97–104.

[7] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, ''PointNet++: Deep hierarchical feature learning on point sets in a metric space,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5105–5114.

[8] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, ''3D semantic parsing of large-scale indoor spaces,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1534–1543.

[9] Q. Hu, B. Yang, S. Khalid, W. Xiao, N. Trigoni, and A. Markham, ''SensatUrban: Learning semantics from urban-scale photogrammetric point clouds,'' *Int. J. Comput. Vis.*, vol. 130, no. 2, pp. 316–343, Feb. 2022.

[10] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[11] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 77–85.

[12] Q. Huang, W. Wang, and U. Neumann, "Recurrent slice networks for 3D segmentation of point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2626–2635.

[13] D. Rethage, J. Wald, J. Sturm, N. Navab, and F. Tombari, "Fully-convolutional point networks for large-scale point clouds," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 596–611.

[14] J. Li, B. M. Chen, and G. H. Lee, "SO-Net: Self-organizing network for point cloud analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9397–9406.

[15] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on X-transformed points," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 828–838.

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

[17] J. Yang, Q. Zhang, B. Ni, L. Li, J. Liu, M. Zhou, and Q. Tian, "Modeling point clouds with self-attention and Gumbel subset sampling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3318–3327.

[18] Y. You, Y. Lou, Q. Liu, Y.-W. Tai, L. Ma, C. Lu, and W. Wang, "Pointwise rotation-invariant network with adaptive sampling and 3D spherical Voxel convolution," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 12717–12724.

[19] A. Poulenard, M. Rakotosaona, Y. Ponty, and M. Ovsjanikov, "Effective rotation-invariant point CNN with spherical harmonics kernels," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 47–56.

[20] X. Sun, Z. Lian, and J. Xiao, "SRINet: Learning strictly rotation-invariant representations for point cloud classification and segmentation," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 980–988.

[21] Y. Rao, J. Lu, and J. Zhou, "Spherical fractal convolutional neural networks for point cloud recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 452–460.

[22] Z. Zhang, B. Hua, D. W. Rosen, and S. Yeung, "Rotation invariant convolutions for 3D point clouds deep learning," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 204–213.

[23] Y. Liu, B. Fan, S. Xiang, and C. Pan, "Relation-shape convolutional neural network for point cloud analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8887–8896.

[24] H. Thomas, C. R. Qi, J. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6410–6419.

[25] L. Wang, Y. Huang, Y. Hou, S. Zhang, and J. Shan, "Graph attention convolution for point cloud semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10288–10297.

[26] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11105–11114.

[27] L. Yi, B. Gong, and T. Funkhouser, "Complete & label: A domain adaptation approach to semantic segmentation of LiDAR point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15358–15368.

[28] J. Mao, X. Wang, and H. Li, "Interpolated convolutional networks for 3D point cloud understanding," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1578–1587.

[29] H. Lei, N. Akhtar, and A. Mian, "SegGCN: Efficient 3D point cloud segmentation with fuzzy spherical kernel," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11608–11617.

[30] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.

[31] Y. Liu, B. Fan, G. Meng, J. Lu, S. Xiang, and C. Pan, "DensePoint: Learning densely contextual representation for efficient point cloud processing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5238–5247.

[32] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "ScanNet: Richly-annotated 3D reconstructions of indoor scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2432–2443.

[33] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys, "Semantic3D.Net: A new large-scale point cloud classification benchmark," 2017, *arXiv:1704.03847*.

[34] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9296–9306.

[35] S. Qiu, S. Anwar, and N. Barnes, "Semantic segmentation for real point cloud scenes via bilateral augmentation and adaptive fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1757–1767.

[36] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Conditional random fields for LiDAR point cloud classification in complex urban areas," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vols. I–3, pp. 263–268, Jul. 2012.

[37] Y. Lu and C. Rasmussen, "Simplified Markov random fields for efficient semantic labeling of 3D point clouds," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2690–2697.

[38] J. Zhang, X. Lin, and X. Ning, "SVM-based classification of segmented airborne LiDAR point clouds in urban areas," *Remote Sens.*, vol. 5, no. 8, pp. 3749–3775, Jul. 2013.

[39] B. Wu, A. Wan, X. Yue, and K. Keutzer, "SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1887–1893.

[40] B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer, "SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 4376–4382.

[41] C. Xu, B. Wu, Z. Wang, W. Zhan, P. Vajda, K. Keutzer, and M. Tomizuka, "SqueezeSegV3: Spatially-adaptive convolution for efficient point-cloud segmentation," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, Aug. 2020, pp. 1–19.

[42] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet ++: Fast and accurate LiDAR semantic segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 4213–4220.

[43] A. Xiao, X. Yang, S. Lu, D. Guan, and J. Huang, "FPS-Net: A convolutional fusion network for large-scale LiDAR point cloud segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 176, pp. 237–249, Jun. 2021.

[44] H. Radi and W. Ali, "VolMap: A real-time model for semantic segmentation of a LiDAR surrounding view," 2019, *arXiv:1906.11873*.

[45] E. E. Aksoy, S. Baci, and S. Cavdar, "SalsaNet: Fast road and vehicle segmentation in LiDAR point clouds for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 926–932.

[46] Y. Zhang, Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh, "PolarNet: An improved grid representation for online LiDAR point clouds semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9598–9607.

[47] Y. A. Alnaggar, M. Afifi, K. Amer, and M. ElHelw, "Multi projection fusion for real-time semantic segmentation of 3D LiDAR point clouds," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1799–1808.

[48] F. J. Lawin, M. Danelljan, P. Tosteberg, G. Bhat, F. S. Khan, and M. Felsberg, "Deep projective 3D semantic segmentation," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Ystad, Sweden: Springer, Aug. 2017, pp. 95–107.

[49] A. Boulch, B. Le Saux, and N. Audebert, "Unstructured point cloud semantic labeling using deep segmentation networks," in *Proc. 3DOR@ Eurographics*, vol. 3, 2017, pp. 17–24.

[50] J. Guerry, A. Boulch, B. L. Saux, J. Moras, A. Plyer, and D. Filliat, "SnapNet-R: Consistent 3D multi-view semantic labeling for robotics," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 669–678.

[51] M. Tatarchenko, J. Park, V. Koltun, and Q. Zhou, "Tangent convolutions for dense prediction in 3D," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3887–3896.

[52] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*.

[53] J. Beltrán, C. Guindel, F. M. Moreno, D. Cruzado, F. García, and A. D. L. Escalera, "BirdNet: A 3D object detection framework from LiDAR information," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3517–3523.

[54] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12689–12697.

[55] Y. Wang, W. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8437–8445.

[56] B. Yang, W. Luo, and R. Urtasun, "PIXOR: Real-time 3D object detection from point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7652–7660.

[57] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4490–4499.

[58] J. Huang and S. You, "Point cloud labeling using 3D convolutional neural network," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 2670–2675.

[59] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "SEGCloud: Semantic segmentation of 3D point clouds," in *Proc. Int. Conf. 3D Vis. (3DV)*, Oct. 2017, pp. 537–547.

[60] H. Meng, L. Gao, Y. Lai, and D. Manocha, "VV-Net: Voxel VAE net with group convolutions for point cloud segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8499–8507.

[61] G. Riegler, A. O. Ulusoy, and A. Geiger, "OctNet: Learning deep 3D representations at high resolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6620–6629.

[62] R. Klokov and V. Lempitsky, "Escape from cells: Deep kd-networks for the recognition of 3D point cloud models," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 863–872.

[63] B. Graham, M. Engelcke, and L. V. D. Maaten, "3D semantic segmentation with submanifold sparse convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9224–9232.

[64] C. Choy, J. Gwak, and S. Savarese, "4D spatio-temporal ConvNets: Minkowski convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3070–3079.

[65] H. Su, V. Jampani, D. Sun, S. Maji, E. Kalogerakis, M. Yang, and J. Kautz, "SPLATNet: Sparse lattice networks for point cloud processing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2530–2539.

[66] H. Zhou, X. Zhu, X. Song, Y. Ma, Z. Wang, H. Li, and D. Lin, "Cylinder3D: An effective 3D framework for driving-scene LiDAR semantic segmentation," 2020, *arXiv:2008.01550*.

[67] G. Feng, W. Li, X. Zhao, X. Yang, X. Kong, T. Huang, and J. Cui, "LessNet: Lightweight and efficient semantic segmentation for large-scale point clouds," *IET Cyber-Syst. Robot.*, vol. 4, no. 2, pp. 107–115, Jun. 2022.

[68] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 922–928.

[69] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[70] E. Meijering, "A chronology of interpolation: From ancient astronomy to modern signal and image processing," *Proc. IEEE*, vol. 90, no. 3, pp. 319–342, Mar. 2002.

[71] B. Graham, "Sparse 3D convolutional neural networks," 2015, *arXiv:1505.02890*.

[72] M. Engelcke, D. Rao, D. Z. Wang, C. H. Tong, and I. Posner, "Vote3Deep: Fast object detection in 3D point clouds using efficient convolutional neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 1355–1361.

[73] Y. Yan, Y. Mao, and B. Li, "SECOND: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, Oct. 2018.

[74] M. Cheng, L. Hui, J. Xie, J. Yang, and H. Kong, "Cascaded non-local neural network for point cloud semantic segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 8447–8452.

[75] T. Cortinhal, G. Tzelepis, and E. E. Aksoy, "SalsaNext: Fast, uncertainty-aware semantic segmentation of LiDAR point clouds," in *Proc. Int. Symp. Visual Comput.*, San Diego, CA, USA: Springer, Oct. 2020, pp. 207–222.

[76] Y. Zhao, X. Ma, B. Hu, Q. Zhang, M. Ye, and G. Zhou, "A large-scale point cloud semantic segmentation network via local dual features and global correlations," *Comput. Graph.*, vol. 111, pp. 133–144, Apr. 2023.

[77] D. Nie, R. Lan, L. Wang, and X. Ren, "Pyramid architecture for multi-scale processing in point cloud segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17263–17273.

[78] T. Lu, L. Wang, and G. Wu, "CGA-Net: Category guided aggregation for point cloud semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11688–11697.

[79] S. Fan, Q. Dong, F. Zhu, Y. Lv, P. Ye, and F. Wang, "SCF-Net: Learning spatial contextual features for large-scale point cloud segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14499–14508.

[80] Y. Lin, G. Vosselman, Y. Cao, and M. Y. Yang, "Local and global encoder network for semantic segmentation of airborne laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 176, pp. 151–168, Jun. 2021.

[81] G. Te, W. Hu, A. Zheng, and Z. Guo, "RGCNN: Regularized graph CNN for point cloud segmentation," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 746–754.

[82] L. Landrieu and M. Simonovsky, "Large-scale point cloud semantic segmentation with superpoint graphs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4558–4567.

[83] J. Liu, B. Ni, C. Li, J. Yang, and Q. Tian, "Dynamic points agglomeration for hierarchical point sets learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7545–7554.

[84] G. Li, M. Müller, A. Thabet, and B. Ghanem, "DeepGCNs: Can GCNs go as deep as CNNs?" in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9266–9275.

[85] Z. Lin, S. Huang, and Y. F. Wang, "Convolution in the cloud: Learning deformable kernels in 3D graph convolution networks for point cloud analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1797–1806.

[86] K. Zhang, M. Hao, J. Wang, C. W. de Silva, and C. Fu, "Linked dynamic graph CNN: Learning on point cloud via linking hierarchical features," 2019, *arXiv:1904.10014*.

[87] L. Xie, T. Furuhata, and K. Shimada, "Multi-resolution graph neural network for large-scale pointcloud segmentation," 2020, *arXiv:2009.08924*.

[88] Y. Li, H. Chen, Z. Cui, R. Timofte, M. Pollefeys, G. Chirikjian, and L. Van Gool, "Towards efficient graph convolutional networks for point cloud handling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3732–3742.

[89] Y. Xu, T. Fan, M. Xu, L. Zeng, and Y. Qiao, "SpiderCNN: Deep learning on point sets with parameterized convolutional filters," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 87–102.

[90] S. Wang, S. Suo, W. Ma, A. Pokrovsky, and R. Urtasun, "Deep parametric continuous convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2589–2597.

[91] A. Komarichev, Z. Zhong, and J. Hua, "A-CNN: Annularly convolutional neural networks on point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7413–7422.

[92] S. Kumawat and S. Raman, "LP-3DCNN: Unveiling local phase in 3D convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4898–4907.

[93] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9613–9622.

[94] H. Lei, N. Akhtar, and A. Mian, "Octree guided CNN with spherical kernels for 3D point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9623–9632.

[95] M. Lin and A. Feragen, "DiffConv: Analyzing irregular point clouds with an irregular view," in *Proc. Eur. Conf. Comput. Vis.* Tel Aviv, Israel: Springer, Oct. 2022, pp. 380–397.

[96] C. Chen, L. Z. Fragonara, and A. Tsourdos, "GAPNet: Graph attention based point neural network for exploiting local feature of point cloud," 2019, *arXiv:1905.08705*.

[97] K. Zhiheng and L. Ning, "PyramNet: Point cloud pyramid attention network and graph embedding module for classification and segmentation," 2019, *arXiv:1906.03299*.

[98] X. Liu, Z. Han, Y.-S. Liu, and M. Zwicker, "Point2sequence: Learning the shape representation of 3D point clouds with an attention-based sequence to sequence network," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 8778–8785.
[99] M. Feng, L. Zhang, X. Lin, S. Z. Gilani, and A. Mian, "Point attention network for semantic segmentation of 3D point clouds," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107446.
[100] L. Chen, W. Chen, Z. Xu, H. Huang, S. Wang, Q. Zhu, and H. Li, "DAPNet: A double self-attention convolutional network for point cloud semantic labeling," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9680–9691, 2021.
[101] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16239–16248.
[102] R. Cheng, R. Razani, Y. Ren, and L. Bingbing, "S3Net: 3D LiDAR sparse semantic segmentation network," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 14040–14046.
[103] Z. J. Tang and T. Cham, "MPT-Net: Mask point transformer network for large scale point cloud semantic segmentation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 10611–10618.
[104] Z. Liu, X. Qi, and C. Fu, "One thing one click: A self-training approach for weakly supervised 3D semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1726–1736.
[105] Y. Zhang, Y. Qu, Y. Xie, Z. Li, S. Zheng, and C. Li, "Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 15500–15508.
[106] L. Jiang, S. Shi, Z. Tian, X. Lai, S. Liu, C.-W. Fu, and J. Jia, "Guided point contrastive learning for semi-supervised point cloud semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 6423–6432.
[107] M. Li, Y. Xie, Y. Shen, B. Ke, R. Qiao, B. Ren, S. Lin, and L. Ma, "HybridCR: Weakly-supervised 3D point cloud semantic segmentation via hybrid contrastive regularization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 14910–14919.
[108] J. Hou, B. Graham, M. Nießner, and S. Xie, "Exploring data-efficient 3D scene understanding with contrastive scene contexts," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15582–15592.
[109] Y. Zhang, Z. Li, Y. Xie, Y. Qu, C. Li, and T. Mei, "Weakly supervised semantic segmentation for large-scale point cloud," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 4, 2021, pp. 3421–3429.
[110] H. Shi, J. Wei, R. Li, F. Liu, and G. Lin, "Weakly supervised segmentation on outdoor 4D point clouds with temporal matching and spatial graph propagation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11830–11839.
[111] M. S. Lee, S. W. Yang, and S. W. Han, "GaIA: Graphical information gain based attention network for weakly supervised point cloud semantic segmentation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 582–591.
[112] O. Unal, D. Dai, and L. Van Gool, "Scribble-supervised LiDAR semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2687–2697.
[113] J. Wei, G. Lin, K. Yap, T. Hung, and L. Xie, "Multi-path region mining for weakly supervised 3D semantic segmentation on point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4383–4392.
[114] Z. Ren, I. Misra, A. G. Schwing, and R. Girdhar, "3D spatial recognition without spatially labeled 3D," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13199–13208.
[115] S. Ye, D. Chen, S. Han, and J. Liao, "Learning with noisy labels for robust point cloud segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 6443–6452.
[116] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.
[117] S. Xie, J. Gu, D. Guo, C. R. Qi, L. Guibas, and O. Litany, "PointContrast: Unsupervised pre-training for 3D point cloud understanding," in *Proc. Eur. Conf. Comput. Vis.*, Glasgow, U.K.: Springer, Aug. 2020, pp. 574–591.
[118] Z. Liu, H. Tang, Y. Lin, and S. Han, "Point-voxel CNN for efficient 3D deep learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 965–975.
[119] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, "Searching efficient 3D architectures with sparse point-voxel convolution," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, Aug. 2020, pp. 685–702.
[120] M. Ye, S. Xu, T. Cao, and Q. Chen, "DRINet: A dual-representation iterative learning network for point cloud segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 7427–7436.
[121] J. Xu, R. Zhang, J. Dou, Y. Zhu, J. Sun, and S. Pu, "RPVNet: A deep and efficient range-point-voxel fusion network for LiDAR point cloud segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16004–16013.
[122] K. El Madawi, H. Rashed, A. El Sallab, O. Nasr, H. Kamel, and S. Yogamani, "RGB and LiDAR fusion based 3D semantic segmentation for autonomous driving," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 7–12.
[123] G. Krispel, M. Opitz, G. Waltner, H. Possegger, and H. Bischof, "FuseSeg: LiDAR point cloud segmentation fusing multi-modal data," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1863–1872.
[124] Z. Zhuang, R. Li, K. Jia, Q. Wang, Y. Li, and M. Tan, "Perception-aware multi-sensor fusion for 3D LiDAR semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 16260–16270.
[125] X. Yan, J. Gao, C. Zheng, C. Zheng, R. Zhang, S. Cui, and Z. Li, "2DPASS: 2D priors assisted semantic segmentation on LiDAR point clouds," in *Proc. Eur. Conf. Comput. Vis.* Tel Aviv, Israel: Springer, Oct. 2022, pp. 677–695.
[126] I. Shin, Y. Tsai, B. Zhuang, S. Schulter, B. Liu, S. Garg, I. S. Kweon, and K. Yoon, "MM-TTA: Multi-modal test-time adaptation for 3D semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16907–16916.
[127] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

ANSI ZHANG received the B.S. degree in mechanical engineering from Shanghai University, China, in 2014, and the Ph.D. degree in mechanical manufacture and automation from the Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University, China, in 2019, with a focus on machine learning and fault diagnosis. He was a joint-training Ph.D. student with the University of South Carolina, Columbia, USA, from 2017 to 2018. He is currently an Associate Professor with the State Key Laboratory of Public Big Data, Guizhou University. His major research interests include computational intelligence, intelligent control systems, and fault diagnosis.

SONG LI received the B.S. degree in computer science and technology from the Jiangsu University of Science and Technology, China, in 2020. He is currently pursuing the master's degree with the State Key Laboratory of Public Big Data, Guizhou University, China. His research interests include deep learning, computer vision, and point cloud processing.

JIE WU received the bachelor's degree in management from China Jiliang University, in 2020. She is currently pursuing the master's degree in electronic information with Guizhou University. Her research interests include weakly supervised learning, deep learning, and point cloud data processing.

SHAOBO LI received the Ph.D. degree in computer software and theory from the Chinese Academy of Sciences, China, in 2003. From 2007 to 2015, he was the Vice Director of the Key Laboratory of Advanced Manufacturing Technology of Ministry of Education, Guizhou University (GZU), China, where he is currently the Director of the State Key Laboratory of Public Big Data. He is also a part-time Doctoral Tutor with the Chinese Academy of Sciences. He has published more than 200 papers in major journals and international conferences. His current research interests include the big data of manufacturing and intelligent manufacturing. His research has been supported by the National Science Foundation of China and the National High-Tech Research and Development Program (863 Program). He received honors and awards from New Century Excellent Talents in the University of Ministry of Education of China, Excellent Expert and Innovative Talent of Guizhou Province, the Group Leader of Manufacturing Informatization, and the Alliance Vice Chairperson of intelligent manufacturing industry in Guizhou Province.

BAO ZHANG received the bachelor's degree in mechanical engineering from Wenzhou University, in 2021. He is currently pursuing the master's degree in mechanical engineering with the School of Mechanical Engineering, Guizhou University. His research interests include indoor positioning, deep learning, point cloud data processing, and LIDAR.