EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

Analysis of Current Advancement in 3D Point Cloud Semantic Segmentation

Koneru Pranav Sai[1], Sagar Dhanraj Pande[2,*]

[1,2]School of Computer Science, VIT-AP University, Amaravati, AP, India,

Abstract

INTRODUCTION: The division of a 3D point cloud into various meaningful regions or objects is known as point cloud segmentation.
OBJECTIVES: The paper discusses the challenges faced in 3D point cloud segmentation, such as the high dimensionality of point cloud data, noise, and varying point densities.
METHODS: The paper compares several commonly used datasets in the field, including the ModelNet, ScanNet, S3DIS, and Semantic 3D datasets, ApploloCar3D, and provides an analysis of the strengths and weaknesses of each dataset. Also provides an overview of the papers that uses Traditional clustering techniques, deep learning-based methods, and hybrid approaches in point cloud semantic segmentation. The report also discusses the benefits and drawbacks of each approach.
CONCLUSION: This study sheds light on the state of the art in semantic segmentation of 3D point clouds.

Keywords: Point cloud, Semantic segmentation, Datasets, Deep learning

Received on 14 September 2023, accepted on 20 November 2023, published on 28 November 2023

doi: 10.4108/eetiot.4495

1. Introduction

The accessibility and cost-effectiveness of 3D sensors, such as LiDARs and RGB-D cameras, have revolutionized numerous disciplines by providing rich 3D data that was previously unavailable or difficult to acquire. The utilization of 3D data offers a significant benefit in that it furnishes insights into the geometry and morphology of objects, which cannot be obtained through conventional 2D imagery. Autonomous vehicles, robots, remote sensing, and even medical care could all benefit greatly from the availability of 3D information [1].

*Corresponding author. Email: sagarpande30@gmail.com

The point cloud format is a common choice for the representation of three-dimensional data. Unlike other formats, such as meshes or volumetric grids, point clouds do not require any discretization or surface approximation, and therefore they can preserve the original geometric information of the object or scene being represented. A three-dimensional point cloud is the representation of the coordinates of an object or surface, and it is composed of a group of data points that span all three dimensions. These points are often captured using 3D scanners or LIDAR (Light Detection and Ranging) technology, which can collect millions of points in a single scan. A point cloud is a collection of points, each of which has its own distinct x, y, and z coordinates that characterize its location in 3D space. The resolution of the scanner used to get the data can change how dense the point cloud is, or how many points there are per unit of space. [2]. The aim of point cloud segmentation is to split the points in a cloud into distinct segments based on their shared characteristics, such as the presence or absence of geometric structures like planes, spheres, or cylinders. Color, texture, shape, and density are all examples of segmentation criteria. The enhanced object recognition, data

visualization, and precise depth perception made 3D point segmentation a valuable instrument for a variety of applications.

Semantic, instance, and part segmentation are broad classifications for 3D point cloud segmentation techniques. The process of assigning a label, often describing the kind of item or surface that the point belongs to each individual point in the cloud is what is known as semantic segmentation. This is helpful for activities such as analyzing the scene or detecting and identifying objects in the scene. Instance segmentation, on the other hand, involves identifying and segmenting individual objects within the point cloud. This can be useful for tasks such as object tracking or robot navigation. Part segmentation involves segmenting objects into smaller sub-parts or components. This can be helpful for operations like moving or putting things together. Different deep-learning strategies for segmenting point clouds were provided in the reviewed publications [3-7]. In this research, we present a high-level summary of point-cloud-based 3D semantic segmentation. Section II outlines the key issues facing the field that serve as the driving forces behind the different approaches and describes the commonly available 3D point cloud datasets

2. Challenges and Datasets

2.1. Challenges

In 3D data, we can precisely identify an object's shape, size, and other characteristics. However, it is not a simple task to extract features from 3D point clouds. Cloud information is typically unorganized, scarce, and noisy. There is also no clear pattern in the data, suggesting that the surface's form is more likely to be artificial rather than naturally occurring.
3D point cloud segmentation faces the following challenges:
(i) Disorganisation, including both high- and low-density areas.
(ii) Unstructured/no grid: Point clouds, unlike other types of 3D data like voxel grids, do not have a predefined grid layout.
(iii) Unordered: The points in a point cloud are not arranged in any pattern.
(iv) Noise and Outliers: Point clouds can contain noise and outliers due to sensor inaccuracies, occlusions, or reflections.
(v) Large data sets: The processing time, storage space, and memory requirements of point clouds can be prohibitive if the clouds include many points.
(vi) Limited sensor resolution: The resolution of 3D sensors used to capture point clouds can be limited, resulting in loss of detail and accuracy.

The development of a segmentation algorithm is a challenging task considering these issues [6]. Section III discusses various methods of segmentation addressing these challenges.

2.1. Datasets

In 3D point semantic segmentation, data sets play a crucial role in developing and training effective algorithms for accurately classifying and labeling points in a cloud. This section, we will discuss a set of datasets that sees regular use in 3D point cloud segmentation:

ModelNet [8]

The ModelNet40 dataset has become a standard benchmark for evaluating 3D object recognition and classification algorithms. It has been used to evaluate the efficiency of various algorithms and to evaluate the impact of different factors, such as point cloud resolution and object occlusion, on recognition accuracy. CAD-generated meshes totalling 12,311 in 40 different classes, including plants, tables, chairs, desks, lamps, vehicles, and airplanes, are included in the Mod- elNet40 benchmark dataset. There are 9,843 models in the training set and 2,468 models in the test set of the ModelNet40 dataset.

PC-Urban (Urban Point Cloud) [9]

The PC-Urban dataset was obtained using a 64-channel Ouster LiDAR sensor. The dataset consists of approximately 4 billion 300 million points collected over 66 thousand sensor frames. The labeled data is separated into raw and registered point cloud frames, with a variable number of consecutive registered frames associated with each raw frame. It provides 25 class identifiers for the 23-million-point dataset with 5,000 instances. End users can simply extend labeling, which is performed by PC- Annotate, using the same application.

Semantic3D [10]

This benchmark dataset contains over 3 billion points and 8 class labels. It contains 15 training and test scenes. It is a valuable resource for building and testing machine learning models for 3D object recognition, scene interpretation, and similar applications. The diversity of scenes included in the dataset, ranging from churches to soccer fields, should help ensure that models trained on this dataset are robust and generalizable to a wide variety of real-world environments.

Joint 2D-3D-semantic data for understanding indoor scenes [11]

The dataset is gathered in six sizable interior spaces that are part of three separate, primarily office and educational buildings. When viewed from the same coordinate system, there is a one-to-one relationship between the various modalities at each point. The collection is an assemblage of 1,413 equirectangular RGB images alongside a grand total of 70,496 conventional RGB photographs. Additionally, it encompasses the depths, outer layer averages, semantic annotations, global XYZ OpenEXR form, and camera metadata for each individual image.

Semantic KITTI [12]

A sizable dataset featuring outstanding point-wise annotation detail and a total of 28 classes, which can be put to a variety

of diverse uses. This dataset stands out from others because of the meticulous scan-wise tagging of the sequences. Annotation work has been performed on all 22 sequences that make up the odometry benchmark that is part of the KITTI Vision Benchmark. This dataset consists of close to 43,000 scans in total. In addition, the data is labeled for the rotating laser sensor's whole horizontal field of view, which indicates that annotations are provided for each point that is acquired by the sensor as it scans its surroundings. This can be seen by the fact that the field of view spans 360 degrees horizontally.

Habitat-Matterport 3D Semantics Dataset [13]

There are 3,100 rooms and 142,646 object instance annotations spread throughout two hundred and sixteen 3D spaces. The quantity, caliber, and variety of object annotations are all significantly higher than in previous datasets. The texture data usage to identify pixel-accurate object boundaries distinguishes HM3DSEM from other datasets in a significant way. Using various techniques, we show how well the HM3DSEM dataset performs for the Object Goal Navigation task. HM3DSEM-trained policies outperform those that were trained using earlier datasets.

ApolloCar3D [14]

ApolloCar3DT data set contains 5,000 high-quality images and more than 60,000 car instances, this dataset provides a rich source of data for training and testing object detection and recognition models. Models trained on this dataset should be very accurate and reliable because they are based on precise 3D CAD representations of individual vehicles, complete with measured model sizes and semantically labeled critical points. In addition, the dataset's substantial size renders it a prime asset for the development and assessment of deep learning models. These models frequently demand extensive data to attain remarkable levels of precision. ApolloCar3DT represents a remarkable leap in the advancement and assessment of computer vision models specifically tailored for driving scenarios, owing to its impressive dimensions and exceptional quality.

3. Methods for Semantic Segmentation of Point Clouds

Semantic segmentation is a computer vision technique that splits an image into many segments or areas, each one representing a different object or class. In contrast to object detection, which only identifies the location of an object in an image, semantic segmentation labels every pixel in the image, providing a more detailed understanding of the image content [15][16]. The board classification of methods of semantic segmentation is shown in Figure 1.

3.1. Weak And Semi-Supervised Semantic Segmentation

The challenge of effectively labeling substantial volumes of data poses a significant obstacle in the realm of Semantic segmentation. One approach to address this challenge is to use weakly supervised or semi-supervised learning techniques (SSL). To train a model that can execute pixel-level segmentation, these methods make use of supplementary sources of information, such as image-level or region-level labels. Semantic segmentation techniques that rely only on semi-supervision or weak supervision are discussed at length in [16–24].

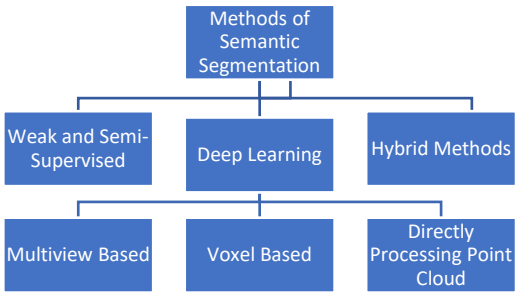

Figure. 1. Methods for semantic segmentation

Yassine Ouali et al. [16] have made a significant contribution to the field of semantic segmentation by introducing a novel semi-supervised approach through the utilization of cross-consistency training. A well-known method of semi-supervised learning is called consistency training. The approach described here leverages unlabeled data and is based on the principle of clustering, which posits that the decision boundary should be situated in regions of sparse data density. The authors posit that the low-density regions exhibit a greater prominence in the hidden representations as compared to the inputs for semantic segmentation. To solve this problem, the authors offered cross-consistency training, in which the predictions stay the same even when the encoder's outputs are changed in different ways. This makes the encoder's representations better, which makes the division more accurate.

Dong-Hyun Lee [17] proposed a method for semi-supervised semantic segmentation called Entropy Minimization with Pseudo-Labels (EMPL), which is a combination of supervised and unsupervised learning. The method employed in this approach involves the simultaneous utilization of both labeled and unlabeled data. It incorporates entropy regularization to the unlabeled data within the context of semi-supervised learning. A common assumption made in this scenario is the presence of a low-density separation between classes. The entropy regularization encourages the model to develop this separation between classes with low density. The method uses Pseudo- Labels, which are the predicted labels with the highest probability as if they were true labels for the unlabeled data. This approach is equal to

entropy regularization, where the model is encouraged to produce low-entropy outputs for the unlabeled data.

With bounding box annotations, which are simpler to get than pixel-level segmentation masks, J. Dai et al. [18] presented the "BoxSup" approach for semantic segmentation. The procedure is divided into two stages: autonomously generating region recommendations and training convolutional networks. To restore segmentation masks and gradually enhance the networks, these processes are repeated. The proposed method has several advantages over fully supervised methods. First, obtaining bounding box annotations is significantly simpler than pixel-level segmentation masks, which can be expensive and time-consuming. Second, BoxSup offers the opportunity to utilize an extensive array of bounding box annotations, thereby enhancing the overall performance of the model. Thirdly, the technique is adaptable and readily applicable to other datasets and tasks.

In their notable work, Nasim Souly et al. [19] introduced a compelling solution to tackle the scarcity of pixel-level labeled data in the field of semantic segmentation. Their approach, rooted in the principles of Generative Adversarial Networks (GANs), offers a promising semi-supervised framework. The proposed methodology incorporates an extensive collection of unlabeled or partially labeled data, alongside artificially generated images produced by GANs. A multi-class classifier serving as the discriminator in the GAN framework is supported by the system's generator network, which also supplies additional training instances. The classifier la- bels each sample with one of K potential labels or flags it as a bogus sample (extra class). The system drives genuine samples to be closer together in the feature space by introducing a lot of fictitious visual data, which enhances multiclass pixel classification.

In their scholarly work, Hung et al. [20] present a meticulous approach aimed at augmenting the precision of semantic segmentation. This approach ingeniously leverages the combined power of labeled and unlabeled data, thereby offering a comprehensive strategy for achieving superior results in this domain. The main advance is the creation of a pixel-level (rather than an image-level) fully convolutional discriminator. This discriminator is trained to distinguish between the ground-truth segmentation distribution and the anticipated probability maps generated by the segmentation model, while also considering the resolution of the maps. The technique recommended involves the integration of the adversarial loss and the conventional cross-entropy loss within the segmentation model. This integration enhances the accuracy of the segmentation process, particularly in regions of the image where the model is prone to producing erroneous outcomes.

The major innovation of [21]'s approach is the use of changing dilation rates to transfer prejudiced information to non-discriminative portions of the image, which aids in the appearance of these regions in object localization maps. The suggested method will enhance both weak and semi-supervised segmentation using semantics by providing dense and trustworthy object localization maps.

Lee et. al., in [22] proposed a novel approach based on FickleNet addresses a critical limitation of weakly supervised semantic image segmentation by generating more ac- curate localization maps that capture both discriminative and non-discriminative parts of objects. The approach is simple yet effective, and its ability to generate ensemble effects from a single network is particularly noteworthy.

Geoff French et al. [23] came up with a new method based on the CutMix regularizer, which is a data augmentation tool that replaces rectangular patches of one image with patches from another image during training. This promotes better generalization and the learning of stable, invariant properties by the model. By adapting CutMix to semantic segmentation, the authors were able to apply consistency regularization to improve segmentation accuracy.

A promising method for poorly supervised semantic segmentation utilizing bounding box annotations was presented by C. Song et al. in [24]. The proposed method consists of two parts: an adaptive loss based on the filling rate (FR-Loss) and a class masking model based on boxes (BCM). Using bounding box annotations as a guide, the BCM is used to delete unnecessary portions of each class.

The accuracy of the segmentation is improved since attention is focused only on the relevant regions. The bounding box annotations' pixel-level segment recommendations are used to calculate each class's mean filling rates, which form the FR- Loss. The filling ratios act as a crucial prior cue that directs the machine to disregard incorrectly labeled pixels in the proposals. This enhances the accuracy of the segmentation and lessens the detrimental effects of proposals with inaccurate labels.

3.2. Semantic Segmentation Using Deep Learning

Deep learning-based semantics is a cutting-edge technique employed in the realm of computer vision. Its primary purpose is to partition an image into numerous segments, with the added ability to assign each individual pixel within the image to a distinct class or category. This process is performed using deep learning algorithms, particularly convolutional neural networks (CNNs). These deep neural networks can learn high-dimensional representations or features from raw data, such as images, audio, or text, without the need for domain-specific knowledge or feature engineering. In contrast, it is worth noting that traditional machine learning methods frequently depend on meticulously crafted features that have been carefully developed based on domain-driven expertise and intuition. The utilisation of deep learning techniques the categorization of point-based, voxel-based, and multi-view-based cloud semantic segmentation algorithms can be attributed to the structural characteristics of the input data in neural networks. The categories in question encompass multi-view, voxel, and point-based approaches, each with their own unique characteristics and applications.

Multi-view-based semantic segmentation refers to the task of segmenting objects or regions of interest in 3D scenes using multiple views of the same scene. This task involves combining information from multiple views to obtain more accurate segmentation. In terms of overall performance improvements, multi-view max-pooling of feature maps excels at both single-view segmentation and the combination of several views. Both scenarios have the potential to benefit from these changes. Very few studies are available on multi-view-based PCSS. A deep neural network strategy was put forth by L. Ma et al. [25] for the prediction of semantic segmentation from RGB-D sequences. The main advancement is the network's self-supervised training to forecast multi-view consistent semantics. The network is constructed using a novel approach to deep learning called single-view learning, which integrates RGB and depth information for semantic object-class segmentation. This method is improved by minimizing loss on several scales via multi-scale optimization, which further boosts the performance of the network.
The multi-view representation learning approach proposed in [26], aims to simultaneously extract useful features and learn a shared representation in a joint feature space. This strategy offers an effective mechanism for capturing underlying correlations. To enhance the training data as well as learning capacity, the model is trained via groups of video frames.

One significant disadvantage of Multiview-based approaches is that they can result in geometric information loss because 2D Multiview images are simply approximations of 3D situations. This limitation can be especially problematic for complex tasks such as point cloud semantic segmentation (PCSS), where the performance can be limited and unsatisfactory due to the loss of geometric structure. Multiple perspectives are necessary to cover all the spaces containing points, which is another drawback of Multiview-based approaches. It can be difficult to choose enough appropriate perspectives for Multiview projection in vast and complicated scenes, which might result in a lack of comprehensive point cloud coverage and the omission of crucial details.

Voxel-based semantic segmentation can be used to segment and classify individual points included inside a point cloud. This is achieved by assigning a semantic designation to each cloud point. It involves separating the point cloud to voxels, which are small volumetric elements, and extracting features from each voxel to classify the points contained within. These features can include the raw coordinates, as well as other attributes such as shape, orientation, and connectivity. To further characterize point clusters and facilitate supervised or un- supervised classification, F. Poux et al. [27] suggested a voxel-based feature engineering approach for point cloud data processing. To provide compatibility between frame-works, the technique offers various feature generalization degrees. The first feature set (SF1) is shape-based and uses only the point cloud's basic X, Y, and Z properties.
The second feature set (SF2) produces a 3D structural connectivity feature set by de- riving relationships and topologies between voxel elements. For planar-dominant classes, the method outperforms both innovative and cutting-edge deep learning methods on the entire S3DIS dataset, with an F1-score of > 85% and a low barrier to integration.

Point-Voxel CNN (PVCNN), a new 3D deep learning model that makes use of sparsity and has a smaller memory footprint, was proposed by Zhijian Liu et al. [28]. It encodes 3D input data as point clouds. In addition, the model employs voxel-based convolution to create contiguous memory access patterns, which enhances performance in comparison to existing point-based models. Numerous tests on various tasks show that PVCNN beats voxel-based baselines while using 10 times less memory, on average, and 7 times more speed than the most advanced point-based models. Overall, PVCNN is a promising method for 3D deep learning since it provides excellent accuracy while also offering memory and compute savings.

Methods for Processing Point Cloud Data Directly: Processing point cloud data directly entails conducting tasks such as the process of segmentation, recognition of objects, and 3D reconstruction without first converting the data into another format such as a mesh or voxel grid. The following are some of the many techniques available for direct processing of point cloud data.

PointNet: Point clouds are simpler and more unified structures than meshes, which can have varying topologies and connectivity. This makes point clouds easier to work with and learn from in many cases. Meshes can have many combinatorial irregularities and complexities that can make it difficult to extract useful features or to perform operations like convolutions. Point clouds, on the other hand, are just collections of 3D points, and can be processed using simple geometric operations and shared MLPs in point-based networks like PointNet [29]. The performance of PointNet on large outdoor point clouds of metropolitan landscapes is assessed by A. Nurunnabi et al. in their study [30], and they conclude that PointNet has the potential for semantic segmentation. PointNet does not account for the neighborhood structure provided by the measurement space created by its neighbors. The study also examines how input vectors influence the efficacy of Point-Net and how sensitive it is to meta-parameters such as the size of the batch, block partitioning, and block point count.

PointNet++: PointNet++ is an extension of PointNet that fixes the problem of not considering local structure, which is caused by the measurement space made by its nearby neighbors. By utilizing the distance measure inherent to the underlying space, this method effectively partitions the set of points into local regions that overlap with one another. Convolutional neural networks (CNNs) employ a similar approach by examining small local regions to identify distinctive features. Higher-level features are generated by processing these local features after they have been aggregated into larger units. Up until the entire point set's features are obtained, this process is repeated. Partitioning the point set and decoupling point collections or local feature

learners utilising a local feature learner are two challenges that must be overcome in the design of PointNet++. To share the local feature learner weights just as the convolutional setting, the point set must be partitioned such that common structures exist across partitions [31]. The experimental findings presented in reference [32] showcase the remarkable performance of various point-based deep learning techniques. Among these techniques, it is worth noting that PointNet++ stands out as the frontrunner, exhibiting the highest level of segmentation accuracy. The benefit of PointNet++ is that it offers flexibility in the sizes of the point cloud data's hierarchical organization. All architectures, except for PointNet, performed better after pre-training with artificial 3D models.

Point Convolution: Point clouds are irregular and unordered, making convolution difficult, unlike regular dense grids in images. Convolution kernels consist of the weight and density functions, which are treated as non-linear functions based on the regional coordinates of the 3D points, they represent. The kernel density estimation method is used to teach density functions, while perceptron networks are used for weight functions. Wu et al. [33] developed a technique called PointConv that allows for convolution on point sets in 3D space that is both translation- and permutation-invariant. Since point clouds are typically unstructured and irregular, this method offers a way to execute convolutional operations on them. The PointConv technique provides an answer for restoring the original resolution of a point cloud that has been subsampled. It achieves this by functioning as a deconvolution operator, which allows for the up sampling of the feature representation. The biggest achievement of this work is the development of a novel method for efficiently computing weight functions. This allows for the network to be scaled up, which in turn leads to a large gain in performance. When applied to challenging semantic segmentation benchmarks on 3D point clouds, such as ModelNet40, ShapeNet, and ScanNet, the suggested technique achieves results that are at the cutting edge of the field.

Graph based convolution: Graph Convolution Neural Networks (GCN) are a fascinating class of neural networks that specialize in processing graphs, enabling the analysis of diverse data structures. The spatial interactions between points in a cloud can be represented by a graph, with each point serving as a node. GCNs operate on these graphs to learn representations that capture the structure and features of the point cloud. Point cloud semantic segmentation was given a new approach by Yuan [34], who proposed using GCNs to learn both local and global properties of the point cloud. The method also includes a novel approach for generating complete pseudo labels, which can be used to increase the amount of labeled information available for training the GCN.Wang et al. [35] presented EdgeConv, a new neural network module aimed to restore topology and improve point cloud representation ability. The EdgeConv module is well-suited for the CNN-based high-level point cloud segmentation and classification tasks. The EdgeConv module utilizes dynamically generated graphs from each network layer. It includes local neighbourhood knowledge and can be stacked to learn global shape fea- tures. The module can be simply inserted into pre-existing designs and is differentiable. The authors evaluated the performance of the EdgeConv module on several standard benchmarks, including ModelNet40, ShapeNetPart, and S3DIS. They demonstrate that their model is superior to other approaches when it comes to classifying and segmenting point clouds.
The HDGCN, a groundbreaking innovation proposed by Z. Liang et al. [36], offers a novel approach to categorizing point clouds according to their semantic significance. Capturing local patterns or relationships is one of the major difficulties in learning from point clouds. A potent method for obtaining neighborhood shape information is graph convolution. The authors suggest a depthwise graph convolution that, in comparison to earlier graph convolution techniques, uses less memory and is inspired by depthwise convolution. In contrast to pointwise convolution, the depthwise graph convolution method accumulates characteristics on a per-channel basis. The authors have introduced a novel block, referred to as Depthwise graph Convolution (DGConv), which is designed for local feature extraction. Combining depth-wise graph convolution with point-wise convolution yields this obstruct. The DGConv block exhibits invariance to diverse point ordering and facilitates the transfer of features to neighboring points, as well as the extraction of features from individual points. The HDGCN model is meticulously crafted by incorporating a sequence of DGConv modules. These modules are designed with a hierarchical structure, enabling them to effectively capture global as well as local characteristics of point cloud data. The model's performance was deemed satisfactory in both the S3DIS and Paris-Lille-3D datasets.

3.3. Hybrid Methods

Hybrid methods for point cloud semantic segmentation refer to the combination of multiple techniques or algorithms to achieve more precise and effective segmentation results on point cloud data. When a voxel contains points from many classes, voxel-based convolutions may generate confusing or incorrect predictions. Other techniques, such as PointNets and point-wise convolutions, have high memory and processing costs, however. When it comes to the segmentation of point clouds, this offers a hurdle.

These issues were addressed by the FusionNet design proposed by F. Zhang et al. [37], which supplements voxel-based representations with a mini-PointNet for feature learning and a fusion module for quick feature aggregation. Compared to standard methods, FusionNet appears to have several benefits. It can learn point-wise predictions that are more accurate while also using less memory and processing resources than voxel-based convolutional networks. Because of this, it excels at jobs involving large-scale point cloud segmentation. By using a structure called a super point graph

(SPG), which divides the scene into geometrically homogeneous elements, the method proposed by
The work of L. Landrieu et al. [38] can capture contextual relationships between object pieces in a representation that is both compact and rich. This allows for the semantic segmentation of massive point clouds.

A hybrid methodology for 3D semantic scene labeling is described by L. Jiang et al. [39], and it entails giving semantic labels to specific places inside a 3D scene. The method models the edges between points and their contextual neighbors to represent the semantic links between the two. Two branches make up the strategy: an edge branch that creates edge features that incorporate point features, and an encoder-decoder branch that predicts point labels. The graph starts out with a crude layer in the hierarchical graph architecture, and it gradually becomes richer as the point decoding process moves along. The point prediction algorithm is improved by the edge branch, that forecasts label of every edge in the final graph to show the semantic consistency of the two connected points. Additionally, edge characteristics are provided into the relevant point module at several layers to incorporate contextual data for message transmission improvement in local areas.

A quick overview of the papers reviewed along with methodologies and the data sets used is given in Table 1.

Table 1. A Taxonomy of the 3D point cloud semantic segmentation methods

	Methods	Data sets used
weakly supervised or semi-supervised learning techniques	Cross consistency training [16]	PASCAL VOC, Cityscapes, Camvid, SUN-RGB-D
	Entropy Minimization with Pseudo-Labels (EMPL) [17]	MNIST handwritten dig dataset
	BoxSup[18]	PASCAL VOC 2012 PASCAL-CONTEXT
	GANs[19]	PASCAL, Sift FLow CamVid
	adversarial loss + standard cross-entropy loss [20]	PASCAL VOC 2012 Cityscapes
	Dilated Convolution [21]	PASCAL VOC
	FickleNet [22]	PASCAL VOC 2012
	CutMix regularizer [23]	CamVid
	FR-Loss +BCM [24]	CamVid
Deep learning	Multi-view based [25]	NYUDv2
	Multi-view based [26]	UAV, DAVIS16
	Voxel-based [27]	S3DIS
	Voxel-based [28]	PVCNN
	Pointnet[29]	ShapeNet
	Pointnet[30]	ALS
	Pointnet++[31]	MNIST, ModelNet40 SHREC15, ScanNet
	Pointnet++[32]	ROSE-X
	Point Convolution[33]	ModelNet40
	Graph based Convolution [34-36]	ShapeNet, S3DIS ModeNet40, Paris- Lille 3D
Hybrid Models	FusionNet Architecture [37]	S3DIS, ScanNet Semantic KITTI
	Super Point Graphs [38]	Semantic3D, S3DIS
	Hierarchical point-edge interaction network [39]	S3DIS, ScanNet v2

4. Conclusion

Semantic segmentation of three-dimensional point clouds has been briefly discussed in this research. We presented a thorough taxonomy of the techniques and present a comparison of their performance. We also talk about the pros and cons of each method. Frequently used datasets for 3D point cloud semantic segmentation are also presented; these can be used as standards when comparing algorithms' results. In the future, hybrid datasets can be used for comparison of 3D Point cloud semantic segmentation.

5. References

1. **Conference Paper** Chen, X.. Ma, H.. Wan, J.. Li, B.. Xia, T.: Multi-view 3D object detection network for autonomous driving. CVPR, (2017).
2. **Book** Juana, V. H., Abhinav, V.: Chapter 12 - Semantic scene segmentation for robotics, Editor(s): Alexandros Iosifidis, Anastasios Tefas, Deep Learning for Robot Perception and Cognition. Academic Press, 279-311, (2022).
3. **Review Paper** Huang, X., Mei, G., Zhang, J., Abbas, R.: A comprehensive survey on point cloud registra- tion, arXiv:2103.02690, (2021).
4. **Review Paper** Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep learning for 3D point clouds: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 43(12), 4338–4364 (2021).
5. **Review Paper** Lu, H., Shi, H.: Deep learning for 3D point cloud understanding: A survey. arXiv:2009.08920, (2020).
6. **Review Paper** Nguyen, A., Le, B.: 3D point cloud segmentation: A survey. 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM), pp. 225-230, Philippines, (2013).
7. **Journal Article** Garcia, G. A., O-Escolano, S., Oprea, S., Martinez, V.V., Rodriguez, G. J.: A Review on Deep Learning Techniques Applied to Semantic Segmentation. https://arxiv.org/abs/1704.06857, (2017).
8. **Dataset** http://3dshapenets.cs.princeton.edu/
9. **Dataset** Ibrahim, M., Akhtar, N., Wise, M., Mian, M.: PC-Urban Outdoor dataset for 3D point Cloud semantic segmentation. IEEE Dataport, (2021).
10. **Dataset** https://www.semantic3d.net/
11. **Dataset** http://3Dsemantics.stanford.edu/.
12. **Dataset** www.semantic-kitti.org.
13. **Dataset** https://aihabitat.org/datasets/hm3dsemantics/
14. **Dataset** https://apolloscape.auto/car_instance.html
15. **Conference Paper** Gould, S., Fulton, R., Koller, D.: Decomposing a Scene into Geometric and Semantically Consistent Regions. Proceedings of the IEEE International Conference on Computer Vision pp. 1-8, (2009).
16. **Journal Article** Ouali, Y., Hudelot, C., Tami, M.: Semi-Supervised Semantic Segmentation with Cross-Consistency Training, arxiv.org/pdf/2003.09005, CVPR, (2020).
17. **Article** Lee, D-H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on Challenges in Representation Learning, ICML, vol. 3, pp. 2, (2013).

18. **Conference Paper** Dai, J., He, K., Sun, J.: BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1635-1643, Santiago, Chile, (2015).
19. **Conference Paper** Souly, N., Spampinato, C., Shah, M.: Semi supervised semantic segmentation using gener- ative adversarial network. In Proceedings of the IEEE International Conference on Com- puter Vision, pp. 5688–5696, (2017).
20. **Journal Article** Hung, W.C., Yi-Hsuan, T., Yan-Ting, L., Yen-Yu, L., Ming-Hsuan, Y.: Adversarial learn- ing for semi-supervised semantic segmentation. arXiv:1802.07934, (2018).
21. **Conference Paper** Wei, Y., Xiao, H., Shi, H., Jie, Z., Feng, J., S- Huang, T.: Revisiting dilated convolution: A simple approach for weakly- and semi-supervised semantic segmentation. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2018).
22. **Conference Paper** Lee, J., Kim, E., Lee, S., Lee, J., Yoon, S.: Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5267–5276, (2019).
23. **Journal Article** French, G., Aila, T., Laine, S., Mackiewicz, M., Finlayson, G.: Consistency regularization and cutmix for semi-supervised semantic segmentation. arXiv preprint arXiv:1906.01916, (2019).
24. **Conference Paper** Song,C., Huang, Y., Ouyang, W., Wang, L.: Box-driven class-wise region masking and fill- ing rate guided loss for weakly supervised semantic segmentation, 2019 IEEE/CVF Confer- ence on Computer Vision and Pattern Recognition, pp. 3131-31402019, (2019).
25. **Conference Paper** Lingni, M., Stückler, J., Kerl, C., Cremers, D.: Multi-view deep learning for consistent se- mantic mapping with RGB-D cameras. pp. 598-605, (2017).
26. **Conference Paper** Sellami, A., Tabbone, S.: Video semantic segmentation using deep multi-view representa- tion learning. 25th International Conference on Pattern Recognition (ICPR), pp. 1-7, Milan, Italy, (2021).
27. **Journal Article** Poux, F., Billen, R. Voxel-based 3D Point Cloud Semantic Segmentation: Unsupervised Geometric and Relationship Featuring vs Deep Learning Methods. ISPRS Int. J. Geo-Inf. 8(213), (2019).
28. **Conference Paper** Liu, Z., Tang, H., Y-L. Song, H.: Point-Voxel CNN for Efficient 3D Deep Learning, arxiv.org/abs/1907.03739v2, (2019).
29. **Conference Paper** Charles, R., Qi, Su, H., Mo, K., J.-Guibas, L.; PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 652-660, (2017).
30. **Article** Abdul, N., Teferle, N., Li, J., Lindenbergh, R., Parvaz, S.: Investigation of PointNet for Semantic Segmentation of Large-Scale Outdoor Point Clouds. ISPRS, (2021).
31. **Article** Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space, NeurIPS, (2017).
32. **Article** Turgut, K., Dutagaci, H., Galopin, G.: Segmentation of structural parts of rosebush plants with 3D point-based deep learning methods. Plant Methods. 18(20), (2022).
33. **Conference Paper** Wu, W., Qi, Z., Fuxin, L.: PointConv: Deep Convolutional Networks on 3D Point Clouds, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9613-9622, CA, USA, (2019).
34. **Article** Yuan, W., Point Cloud Semantic Segmentation using Graph Convolutional Network.
35. **Article** Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M. M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM TOG, (2019).
36. **Conference Paper** Liang, Z., Yang, M., Deng, L., Wang, C., Wang, B.: Hierarchical Depthwise Graph Convo- lutional Neural Network for 3D Semantic Segmentation of Point Clouds. 2019 International Conference on Robotics and Automation (ICRA), pp. 8152-8158, Canada, (2019).
37. **Journal Article** Zhang, F., Fang, J., Wah, B., Torr, P.: Deep FusionNet for Point Cloud Semantic Segmen- tation. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science, vol 12369. Springer, Cham. (2020).
38. **Conference Paper** Landrieu, L., Simonovsky, M.: Large-scale point cloud semantic segmentation with super- point graphs in Proceedings of the IEEE Conference on Computer Vision and Pattern Recog- nition, pp. 4558–4567, (2018).
39. **Conference Paper** Jiang, L., Zhao, H., Liu, S., Shen, X., Fu, C-H., Jia, J.: Hierarchical Point-Edge Interaction Network for Point Cloud Semantic Segmentation, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10432-10440, (2019).