3D scenes semantic segmentation using deep learning based Survey

Amani Y. Noori¹, Shaimaa H. Shaker² and Raghad abdulaali azeez³
¹AL-Mustansiriyah university, Iraq
²University of technology-Iraq
³University of Baghdad /collage of education ibn-ruhed
amani.yousif@uomustansiriyah.edu.iq
Shaimaa.h.shaker@uotechnology.edu.iq
raghad.azeez@ircoedu.uobaghdad.edu.iq

Abstract
Semantic segmentation realization and understanding is a stringent task not just for computer vision but also in the researches of the sciences of earth, semantic segmentation decompose compound architectures in one elements, the most mutual object in a civil outside or inside senses must classified then reinforced with information meaning of all object, it’s a method for labeling and clustering point cloud automatically. Three dimensions natural scenes classification need a point cloud dataset to representation data format as input, many challenge appeared with working of 3d data like: little number, resolution and accurate of three Dimensional dataset. Deep learning now is the power and popular tool for data and image processing in computer vision, used for many applications like “image recognition”, “object detection”, “semantic segmentation”, In this research paper, provide survey a background for many techniques designed to 3 Dimensions point cloud semantic segmentation in different domains on many several available free datasets and also making a comparison between these methods.

Keywords: point cloud, semantic segmentation, deep learning, and dataset.

1. Introduction

Three dimensions computer graphics, or 3 dimensional computer graphics represent geometric data in 3D and stored in the computer in form of (X,Y,Z) and have intensity and color information for the purposes of performing calculations, many method to represent 3d data like point cloud unstructured, voxel, multi view every method has it advantage or dis advantage, when the work with normal scenes the better to work with point cloud dataset, no many large number of three dimensions point cloud dataset there is indoor and outdoor data set the famous Benchmark dataset used for this purpose. [1]

Semantic network is branch of computer vision called “point cloud classification”, “point labeling” and “semantic segmentation / classification/ labeling,” means specify each point cloud to a semantic labeling. The variance between “instance segmentation” and “semantic segmentation”, If we take an image to do classification for every pixel into” meaningful classes
of objects” for example we can separated pixels of whole image of van color with white color this semantic segmentation, instance segmentation is not separate for each pixel but separated for each thing inside image[2].

Deep learning uses more than two hidden layers to get height dimensions features by train input data, the type of deep learning fully convolution neural network is became very common used in remote senses, CNN components of many layers convolution layer doing convolution operations by filter the output result is feature map , pooling is operation of down sampling feature map produced by convolution layer , fully connected perform on a flattened input all neurons connected one to each other. FC found at end of CNN architectures used for produce class scores for each feature [3].

3D convolution networks are divided into two types: “discrete convolution networks” and “continuous convolution networks”, the first method used convolution Kernels to continuous area, the neighboring points in area are linked with spatial allocation with related to point in center, Multi-Layer Perception (MLP) for learning weights. The discrete CNN used convolution Kernels for orderly grids, the neighboring point’s weight is located to offsets of the point in center [4].

2-Related works

1- PointNet

PointNet is first method introduced which provide a model that consume three dimension point cloud used to classify object, part of segmentation and semantic of segmentation the architecture of PointNet takes point clouds directly as input for deep learning there is no use convolution neural network just using Multi Layer Perceptions (MLP). [2]

The basic architecture of PointNet model is using of two transformation network “T-network”, used for transform features, “S-network” used to provide normalization and max pool used for reduce vector of global features, the problem with this model cannot extract enough local feature information, First introduce PointNet then PointNet++ ,[ 5]

2- PointNet++

PointNet++, a powerful architecture of neural network to processing point cloud directly. The general idea of PointNet++ is simple, the method of this model is throw distance Equation segment set of points into local overlapping regions using for features extractor Convolution neural network from that neighborhood parts, This method increase number of features especially local features between points use K- nearest neighbor search , this method still have a problem that cannot provide enough features since geometric relation of one point and its contiguous points is neglected [ 6].
3- **PointCNN**

This model is a generalization, improved of CNN for treating non-order unstructured point cloud. This model has a best result on shape datasets, the idea of this model to provide X-Convolution operator for each point and feature that permutes and gives weight before they are manipulated by convolution [7].

4- **PC-CNN**

This method accepts input raw point clouds directly. PC-CNN performs better convolutional neural network models. CNN groups using KNN to select continuous points for convolution. Operation, PC-CNN improved PointNet that provides local feature extraction and enhances the segmentation by adding one max-pool function and increasing the number of layers to be 10 and one concat layer that reduces the high cost computation of PointNet++. [8]

5- **Point neighborhoods**

This method introduces deep learning that contains two components. The feature space neighborhood NF is found by computing to find all points within the (Knn) K nearest neighbors algorithm in the space of learned features and the world space neighborhood NW is computed by grouping each point with a clustering algorithm using K-means algorithm into world space. Its main components are NF- and NW-modules working on incorporating neighborhood information from the world space and feature space. This method inserts the pairwise distance loss L pair and the centroid loss L cent in the model of three dimensions semantic segmentation. This model works with outdoor and indoor datasets [9].

6- **Dynamic Graph CNN (DGCNN)**

DGCNN is related to approaches, PointNet and graph CNNs, this procedure named EdgeConv for extracted edge features and with stall of keeping invariance permutation and in each convolution layer K-nearest neighbor used for grouping [10].

7- **ShellNet**

This method provides efficiency convolution operator directly consuming point cloud based on statistical and shell uses statistics from concentric spherical shells to define features and resolve the ambiguity of order of point, this method uses traditional convolution to perform on such features. The efficacy of model ShellNet appears on classification for object, segmentation of object, and “semantic scene segmentation” beside making the network very high speed in training stage [11] .

8- **A-CNN**

This method operates on direct 3D point cloud input to new convolution model that improve the capture of local geometric in 3D shape format where using “Farthest Point Sampling” (FPS) perform extraction centroids random distribution on object surface then using KNN for extraction from each local region neighbors of center. Finally doing convolutions operation on set of in order points max function feature extractor throw
neighbors to generate feature vectors unique with each local region. This method operates on ScanNet, ShapeNet and S3DIS datasets [12].

8- RandLA-Net
This method is lightweight and effective memory and computationally semantic segmentation neural network for big 3D point clouds data set that accept direct point cloud no selection point method use random sampling point, propose a new and efficient local feature grouping module to capture local complex structures over small set of point using (KNN) K- nearest neighbors algorithm. This method uses benchmarks data set Semantic3D, S3DIS and SemanticKITTI [13].

9- SAC (Structure Aware Convolution)
SAC is a method to generalize (improved) deep learning neural network regular grids to unstructured three Dimensions point clouds. The idea of SAC is to matching the neighborhoods of point clouds with a numbers of three dimensions kernels, each kernel like consider as a “geometric template” formed by a series of learning 3D points. So, the input point cloud structure is activated by the matching kernels. To evaluate the efficiency of the proposing method, embedded SAC in one of networks of deep learning for 2D point cloud like PointNet and PointNet++ as a lightweight module, used in on both segmentation tasks and classification [14]. all method above will appear in figure (1) bellow:

| Model                      | Year | Dataset       | Method                                                                 | mIOU | Accuracy |
|----------------------------|------|---------------|------------------------------------------------------------------------|------|----------|
| 1- PointNet                | 2017 | S3DIS         | No convolution MPL, Use two T-net network and max pooling layer as a symmetric function | 47.71 | 78.5     |
|                            |      | ScanNet       |                                                                        | 33.9 | 73.9     |
|                            |      |               |                                                                        |      |          |
| 2- pointNet++              | 2017 | ScaNet        | Sample method(FPS) and (MSG) and (MRG) grouping                        | 55.7 | 84.5     |
|                            |      |               |                                                                        |      |          |
| 3- PointCNN                | 2018 | ScaNet        | X-Convolution operator for weighted feature and Knn for grouping       | 45.8 | 85.1     |
|                            |      | S3DIS         |                                                                        | 65.39| 88.1     |
|                            |      |               |                                                                        |      |          |
| 4- PC-CNN                  | 2018 | S3DIS         | Convolution with KNN and KD structure                                 | 68.63%| 88.01   |
|                            |      |               |                                                                        |      |          |
| 5- DGCNN                   | 2019 | S3DIS         | EdgeConv to extract edge features and used for grouping Knn           | 56.1 | 84.1     |
| 6- Point neighborhoods[4]  | 2017 | ScanNet       | KNN for feature And Nw used K-means for world space                   | 45.15| Overall point accuracy (oAcc) 75.53 |
3-Data Sets

Datasets can be classified into two divisions: indoor datasets created with Kinect and outdoor captured by 3D scanners like LIDAR. These free and public datasets bellow have their pros and cons [15]

1- ScanNet
   Indoor and large dataset improved by Stanford university, used for semantic segmentation provide twenty categories, include 1513 scenes from 707 various indoor environments contains 2.5M RGB-D image [15].

2- Semantic3D
   Is high quality outdoor and density dataset the numbers of points is more than 4 billion provide eight classes of semantic including cars, artifacts of scanning, low plant, high plant, natural soil and handmade soil.[16]

3- S3DIS(“Stanford Large-scale 3D Indoor Spaces Dataset”)
   Its Indoor data set, large- scale RGB-D dataset it contain above 215 million number of points was collected by Stanford university et al.it contains a grouped of three types of buildings ,271 rooms with area of covering equal 6000 m². [17]

4- WHU MLS
   RSD46-WHU is a public outdoor data set very large dataset used for classify senses in remote sensing applications. This dataset is aggregated from “Google Earth and Tianditu”. The resolution of utmost classes is 0.5m, and the others are about 2m. There are image from 500 to 3000 with each class. The number of image is 117000 images with forty six classes. [13]
4- Conclusion

In this article research, we reviewed different approaches for 3D semantic segmentation network depends on deep learning neural network through the interval of time (2017-2020). Many methods improvement throw years every time the author trying to design method with low memory consume and high accuracy and using indoor and outdoor dataset. Every technology has many disadvantage, advantage and appears many drawbacks result to exist little number of 3D dataset so the work with this field is challenge.

References
1- Rui Zhang, Yuan Wang, Guangyun Li, Zhen Han, Junpeng Li, and Chunying Li, “3D Convolutional Neural Network for Semantic Scene Segmentation based on Unstructured Point Clouds” international journal for formability Engineering, vol. 14, no. 7, July 2018, pp. 1503-1512, DOI: 10.23940/ijpe.18.07.p14.15031512, PC-CNN.
2- Yuxing Xie, Jiaojiao Tian, Member, IEEE and Xiao Xiang Zhu, Senior Member, IEEE, “A Review of Point Cloud Semantic Segmentation”, arXiv:1908.08854v1 [cs.CV] 23 Aug 2019.
3- Afshine Amidi and Shervine Amidi, “convolutional Neural Networks cheatsheet” stanford university, www.stanford.edu, 2018.
4- Yulan Guo, Wenyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun, “Deep learning For 3D Point Clouds: A Survey”, arXiv:1912.12033v1 [cs.CV] 27 Dec 2019.
5- Charles R. Qi, Hao Su, Kaichun Mo “PointNet: Deep Learning on Point Sets for 3D Classification and segmentation”, conference on Computer Vision and Pattern Recognition (CVPR), 2017.
6- Charles R. Qi Li Yi Hao Su Leonidas J. Guibas, “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
7- Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhua D Baoquan Cheni, “PointCNN: Convolution On X-Transformed Points”, arXiv:1801.07791v5 [cs.CV] 5 Nov 2018.
8- Rui Zhang, Y. Wang, Guangyun Li, Zhen Han, Junpeng Li and Chunying Li, “3D Convolutional Neural Network for Semantic Scene Segmentation based on Unstructured Point Clouds”, International Journal of Performability Engineering 14(7):1503-1512, DOI: 10.23940/ijpe.18.07.p14.15031512, July 2018.
9- Francis Engelmann, Theodora Kontogianni, Jonas Schult, and Bastain Leibe”Know What Your Neighbors Do: 3D Semantic Segmentation of Point Clouds”, arXiv:1810.01151 [cs.CV], DOI: 10.1007/978-3-030-11015-4_29, 2018.
10- YUE WANG and YONGBIN SUN, ZIWEI LIU, SARMA, MICHAEL M. BRONSTEIN, JUSTIN M. SOLOMON, “Dynamic Graph CNN for Learning on Point Cloud”, ACM Transactions on Graphics, Vol. 38, No. 5, Article 146. Publication date: October 2019.
11- Zhiyuan Zhang, Binh-Son Hua, Sai-Kit Yeung, “ShellNet: Efficient Point Cloud Convolutional Neural Networks using Concentric Shells Statistics”, arXiv:1908.06295v1 [cs.CV] 17 Aug 2019.
12- A. Komarichev, Z. Zhong, J. Hua, “A-cnn: Annularly convolutional neural networks on point clouds, in: the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
13- Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, Andrew Markham, “RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds”, arXiv:1911.11236v2 [cs.CV] 30 Mar 2020.

14- Lei Wang, Yuxuan Liu, Shenman Zhang, Jixing Yan, and Pengjie Tao, “Structure-Aware Convolution for 3D Point Cloud Classification and Segmentation” Remote Sens. 2020, 12, 634; doi: 10.3390/rs12040634.

15- Eman Ahmed, Alexander Saint, Abdel Rahman Shabayek Habayek, Kseniya Cherenkova, Srig Das, GLEB Gusev, Djamila Aouada, “A survey on Deep Learning Advances on Different 3D Data Representations”, arXiv:1808.01462v2 [cs.CV] 6 Apr 2019.

16- Charles Ruizhongtai Qi, “Deep Learning on Point Sets for 3D Classification and Segmentation”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI, USA. 21–26 July 2017; pp. 77–85.

17- Saifullahi Aminu Bello, Shangshu Yu, Cheng Wang, senior member, IEEE“ Review: deep learning on 3D point clouds “, arXiv:2001.06280v1 [cs.CV] 17 Jan 2020.