Point-based Method를 사용한 포인트 클라우드 연구 동향

한정우*, 김종국*
*고려대학교 전기전자공학부
hanjungwoo1@korea.ac.kr, jongkook@korea.ac.kr

A Survey on Point Cloud Research Paradigm Using Point – based Method

Jung-Woo Han*, Jong-Kook Kim*
*School of electrical engineering, Korea University

ABSTRACT

In recent years, the use of LiDAR sensors is increasing as autonomous driving, robot control, and drones are considered more. Contrary to ordinary cameras, LiDAR sensors make it possible to handle challenging problems by calculating the distance between objects. This crucial characteristic makes more active research on deep learning models dealing with point clouds which are data of LiDAR. In this paper, among the schemes of using the point cloud, the Point-based approach is mainly discussed. Furthermore, future streams and insights can be considered by looking at solving methods and the limitations.

1. INTRODUCTION

3D data are considered more to solve many practical problems in the real world. Autonomous driving, robot controls, and drone are the most frequent application examples. LiDAR (Light Detection And Ranging) sensor is a preeminent tool for collecting 3d data. It calculates return time after shooting the laser pulse and represents by using point clouds that consist of a large amount points.

The point cloud is unstructured and unordered data. It represents objects using total N number points, and concurrently its coordinates (X Y Z) and point features are treated. To handle this featured data, symmetric function and equivalent calculation are essential to all points. Therefore, it is a more arduous approach than the typical deep learning method using 2D image data.

In general, which pre-processing is used for the data is the criterion for the classification. There are a Volumetric-base Method using a voxelized point cloud (substituting each point into space and solidifying it), and a Multi view-based method using several images obtained by projecting the point into a specific plane. However, in this paper, we focus on a Point-based method that does not take pre-processing. Because it takes a larger amount of data and calculations than2D images, it cannot include preprocessing process for practical use.

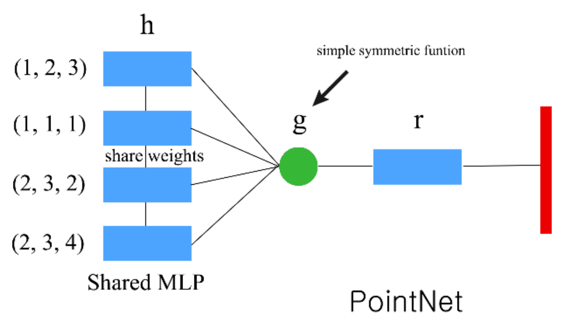

Figure 1. PointNet. Shared MLP

PointNet [1] is the first paper that introduced the Point-based method. They devise a shared MLP layer(Figure 1.) that can be used without pre-processing by adopting the symmetric Convolution Neural Network (CNN) function. This Layer makes the research paradigm for Point-based Method in recent years.

As the follow-up study, PointNet++ [2] adopts the Sampling method to minimize linearly increasing computation and to interpret segmented structures. Moreover, a Grouping method is taken to extract geometric information between points and sampled peripheral points. The overall model structure is composed using the above two methods, and the Point-based method takes the PointNet++ model structure as the basis. Each paper has its strategy to select which sampled points. And grouping means a feature extraction from the sampled points.

In the last of the paper, the research trends of Sampling and Grouping that are the basis of the point-based method will be discussed and their limitations, future works further proposed.

2. BACKGROUND

2.1 Classification

Classification is generally performed by learning the features of each point and then gradually learning the entire shape features. The last layer of Fully Connected (FC) Layer concludes the classification.

2.2 Segmentation

Segmentation requires an understanding of the overall geometric structure and the subdivided structure of each point . The methodologies of segmentation are Semantic Segmentation (frame unit, Scene level), Instance Segmentation (individual unit, Object level), and Part Segmentation (partial unit, Part level). In general, the basis of models is Semantic Segmentation and the goal is to separate points into several subsets according to their meaning.

2.3 Dataset

Several datasets have been collected to evaluate deep learning algorithms for various applications of 3D point cloud. To evaluate classification, ModelNet40[3], ShapeNet40[4] that are synthesized data and ScanNet using real-world data exist. The real-world data may include corruption due to the background noise or overlapping situations. Therefore, synthesized data ModelNet40, ShapeNet40 are mainly used for the evaluation. Data are collected through various tools such as Mobile Laser Scanners (MLS)[Semantic KITTI][5], Terrestrial Laser Scanners (TLS) [Semantic3D][6], RGBD camera and 3D Scanner [S3DIS][7]. S3DIS, Semantic3D, Semantic KITTI have mainly used evaluation methods, indicating more difficulty in order.

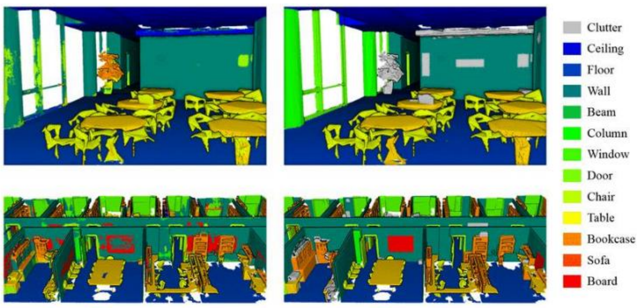

Figure 2. S3DIS DataSet

2.4 Model Evaluation Method

Many evaluation metrics are used for the experiment on various point cloud data. Overall Accuraccy (OA) and mean Class accuracy (mAcc) are the examples. OA indicates the average of overall experiments, and mAcc refers to mean accuracy for each class experiment. For the evaluation, OA, mean Intersection over Union (mIoU, mAcc) is the leading methods. IoU is to find the result of each pixel or point that makes an intersection with the correct answer. mIoU means the average of all experiments.

3. PARADIGM – POINT-BASED METHOD

In the case of point-based methods that have been processed primarily, research is being conducted based on point net++. By using shared MLP from PointNet to the base layer, additional concepts are given to sampling and grouping to create a new model.

3.1 Sampling Method

Farthest Point Sampling(FPS) - PointNet++ adopt FPS method. This method selects the farthest point from each other when selecting S sampling from N, and is generally evenly distributed without randomly selecting to represent N. An operation is required for all points and a time complexity for the amount of operation requires $O(N^2)$.

CP-Net(Critical Point Net) – CP-Net[8] selects high-value features with an index among the result values from each layer and uses it as a sampling criterion. Since it is used based on the result values that occur in the process of learning the relationship between points, it is an attempt to select more critical points. The time complexity of the computation amount requires $O(N\log N)$. Since it is a learnable sampling method, it can be seen as the middle of bias and variation. Since it is a learnable sampling method, it can be seen as the middle of bias and variation.

Random Sampling – RandLA-Net[9] uses Random Sampling to minimize the amount of computation required in this process. The time complexity for the required computation amount is $O(1)$. Compared to other methods, it shows the best performance in terms of computational volume. However, it is difficult to say that it has a valid representative method because it does not have a variance in the sampling method.

The sampling method should minimize the number of points and select the best. However, there has been a contradiction in the method of performing sampling to reduce the amount of computation. In an attempt to the best selection, sampling has become common to account for most of the entire model.

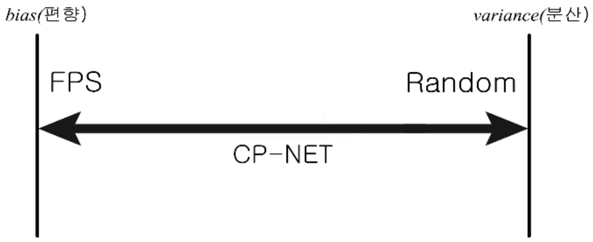

Figure 3. Sampling technique tradeoff between bias and variance.

3.1 Grouping Method

<Table 1> The result of S3DIS DataSet, 6-fold Cross-validation used

	OA(%)	mAcc(%)	mIoU(%)
RandLA-Net	88.0	**82.0**	70.0
PointASNL	88.8	79.0	68.7
PointTransformer	**90.2**	81.9	**73.5**

Local Feature Aggregation(LFA) – LFA is a module used after random sampling in RandLA-Net[9]. Surrounding points for each point are found using K-Nearest Neighbor (KNN). The features of it are collected to the central point and the attention mechanism makes features aggregated. The reason of the aggregation is to ensure features of randomly vanished points. LFA shows the best performance in feature extraction and the fast model using the Random sampling method.

Local-Non Local(L-NL) – L-NL is a module used by PointASNL[10], which simultaneously encodes feature on a nearby section and an overall section. Near features are used for the structure of the Transformer [11], and grouping is processed through KNN. Meanwhile, overall section feature is made using attention mechanism. Above two features are combined and used for the eventual feature.

Point Transformer Layer – Point Transformer[12] introduces this method and one example that adopt Transformer structure with an accurate comprehension of architecture. They substitute the data to appropriate for the Multi-head Attention [11] using Linear Layer and Shared MLP. It is an example of Feature extraction and enables high speed through parallelization.

However, in the case of RandLA-Net approach, they move out the KNN process which requires a huge amount of computation to the data loader. PointASNL adopts FPS but, attempts to remove specific data by adding a module called Adaptive Sampling. Nonetheless, more computation causes lower speed. Although, Point Transformer takes good Feature Extraction method, using many layers and FPS make hard to applicate in the real-world.

<Table 2> Each Module's Method and the Complexity

$N \sim 10^6$	Method	Complexity
RandLA-Net	KNN	$O(N^2)$
PointASNL	FPS	$O(N^2)$
PointTransformer	FPS	$O(N^2)$

4. RESEARCH FLOW

Models based on Point-based Method are constructed through Sampling and Grouping architectures. Sampling should be the way of selecting the best points but minimizing the computation cost. Moreover, the tradeoff in Bias and Variance is crucial. The Grouping methods should upgrade the performance of the Feature Extraction and perform parallelization to ensure no speed degradation.

Although, there are many approaches using Point Cloud data for the real-world application and problems, it is essential to low computation architecture for practical use.

Acknowledgement

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2018-0-01433) supervised by the IITP(Institute for Information & communications Technology Promotion)

Reference

[1] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[2] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017.

[3] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes, 2015. 5

[4] Li Yi, Vladimir G. Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. A scalable active framework for region annotation in 3d shape collections. SIGGRAPH Asia, 2016.

[5] 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. of the IEEE/CVF International Conf. on Computer Vision (ICCV), 2019.

[6] Timo Hackel, N. Savinov, L. Ladicky, Jan D. Wegner, K.Schindler, and M. Pollefeys. SEMANTIC3D.NET: A newlarge-scale point cloud classification benchmark. InISPRSAnnals of the Photogrammetry, Remote Sensing and SpatialInformation Sciences, volume IV-1-W1, pages 91–98, 2017.
[7] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1534–1543, 2016.
[8] Ehsan Nezhadarya, Ehsan Taghavi, Ryan Razani, BingbingLiu, and Jun Luo. Adaptive hierarchical down-sampling forpoint cloud classification. InProceedings of the IEEE/CVFConference on Computer Vision and Pattern Recognition(CVPR), June 2020.
[9] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, YulanGuo, Zhihua Wang, Niki Trigoni, and Andrew Markham.Randla-net: Efficient semantic segmentation of large-scalepoint clouds. InProceedings of the IEEE/CVF Conferenceon Computer Vision and Pattern Recognition (CVPR), June2020.
[10] u Yan, Chaoda Zheng, Zhen Li, Sheng Wang, andShuguang Cui. Pointasnl: Robust point clouds processingusing nonlocal neural networks with adaptive sampling. InProceedings of the IEEE/CVF Conference on Computer Vi-sion and Pattern Recognition (CVPR), June 2020.
[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszko-reit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and IlliaPolosukhin. Attention is all you need, 2017.
[12] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip H. S. Torr, andVladlen Koltun. Point transformer.CoRR, abs/2012.09164,2020.