약간 감독되는 포인트 클라우드 분석에서 일반 로컬 트랜스포머 네트워크

Anh-Thuan Tran[1], 이태호[1], Hoanh-Su Le[3], 최필주[1], 이석환[2], 권기룡[1]
[1]부경대학교 인공지능융합학과, [2]동아대학교 컴퓨터공학과
[3]경제법학과 정보시스템학부 베트남국립대학교

thuantran@pukyong.ac.kr, krkwon@pknu.ac.kr

General Local Transformer Network in Weakly-supervised Point Cloud Analysis

Anh-Thuan Tran[1], Tae Ho Lee[1], Hoanh-Su Le[3], Philjoo Choi[1], Suk-Hwan Lee[2], Ki-Ryong Kwon[1]
[1]Department of Artifical Intelligence Convergence, Pukyong National University
[2]Department of Computer Engineering, Dong-A University
[3]Faculty of Information Systems, University of Economics and Law, Vietnam National University

Abstract

Due to vast points and irregular structure, labeling full points in large-scale point clouds is highly tedious and time-consuming. To resolve this issue, we propose a novel point-based transformer network in weakly-supervised semantic segmentation, which only needs 0.1% point annotations. Our network introduces general local features, representing global factors from different neighborhoods based on their order positions. Then, we share query point weights to local features through point attention to reinforce impacts, which are essential in determining sparse point labels. Geometric encoding is introduced to balance query point impact and remind point position during training. As a result, one point in specific local areas can obtain global features from corresponding ones in other neighborhoods and reinforce from its query points. Experimental results on benchmark large-scale point clouds demonstrate our proposed network's state-of-the-art performance.

1. Introduction

Three-dimensional large-scale point cloud segmentation has received much attention due to its challenging and practical applications in autonomous driving or robotics. Specifically, PointNet++ [1], one of the prominent studies, introduces farthest point sampling and extract features through neighborhoods. On the other sides, RandLA-Net [2] concentrate on geometric positions to address irregular structures. Recently, Point Transformer [3] is one of the first studies to propose a self-attention network on raw points with spatial feature supplements as position encoding.

Nevertheless, these point clouds are restricted due to their vast points, which require costly point-wise annotations. To avoid this exhausting annotation, weakly supervised methods in point clouds have been growing. Several works generated pseudo labels by semi-supervised learning. Π Model [4] achieved this goal by self-ensembling in different regularization and augmentation. Then, MT [5] was proposed to resolve unstable and complexity in the temporal ensembling model. Other works performed point cloud segmentation through limited point annotations. Zhang et al. [6] utilized self-supervised pretext tasks to transfer prior knowledge from the large unlabeled point cloud. Xu and Lee [7] proposed a weakly point cloud segmentation network using a small fraction of points by learning gradient approximation and color smoothness constraints. PSD [8] improved performance through the perturbed branch and enforced predictive consistency. SQN [9] explored semantic similarity between neighboring points without self-supervised pretraining or hand-crafted constraints. Although these methods perform weakly supervised on large-scale point clouds, they are built on fully supervised architecture and are consequently unstable due to massive sparse points.

To resolve this challenge, we propose General Local Transformer Network strengthen point impacts with coherent neighborhoods in weakly supervised point cloud semantic segmentation. The attention module contains two primary components called general local features and point attention. General local features represent global features from different neighborhoods based on their order positions. Point attention shares weights to its local features to reinforce impacts. Based on the idea that distribution of points in local area is relatively homogeneous [1], our method plays a vital role in weakly-supervised to define neighborhood labels. Geometric encoding consolidates position coordinates and supplemental features to give an overview of point position and balance point impacts.

2. Proposed Method

Although neighboring points have vital features, their order positions based on distances significantly impact point representation. We explore the principal factors of each position in different neighborhoods to get general local features. However, each point in 3D point cloud can attend to several neighborhoods, so it has variant impacts on multiple points. To address this problem, we use encoded point weight to force neighboring points to focus on local areas with corresponding order positions. Points that share the same order position are combined.

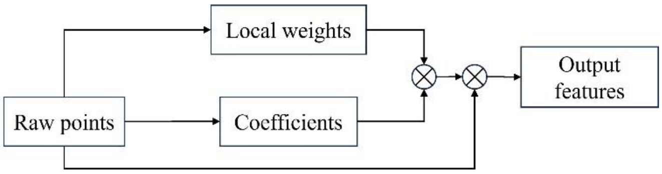

Figure 1. Attention mechanism architecture.

Query point weights are shared with their neighborhoods by generating point coefficients using input features. Therefore, local neighborhoods are reinforced by global factors from other areas and query point features. To decode general local features, we have matrix multiplication between general local features and input features to get local coefficients. Then, importance of neighboring points in specific local neighborhoods is determined. We integrate general local features and point attention with learnable parameters to regenerate local neighborhoods.

3. Experiment

We compare our proposed model to state-of-the-art methods in S3DIS Area-5, summarized in Table I. Best results in weakly labeled settings are represented through bold parts and underlined areas show fully labeled settings. We perform experiments using 100 epochs, batch size 2 and AdamW optimizer with learning rate 1e-5 and weight decay 1e-4. The network achieves outstanding performance with 65.1% mIoU and surpasses several fully-supervised studies.

Table 1. Semantic segmentation results on S3DIS Area-5.

Setting	Method	mIoU
100%	PointNet++[1]	50.0
	RandLA-Net [2]	63.0
	Point Transformer [3]	<u>70.4</u>
10%	Π Model [4]	46.3
	MT [5]	47.9
	Xu and Lee [7]	48.0
1%	Zhang et.al [6]	61.8
	PSD [8]	63.5
0.1%	RandLA-Net [2]	52.9
	SQN [9]	61.4
	Ours	**65.1**

4. Conclusion

We propose a local attention network to perform point cloud segmentation under weakly-supervised settings. Local attention considers relationships between query points and their neighborhoods to obtain more coherent features. As a result, our network has the ability to handle these point cloud problems with a tiny fraction of labeled points.

5. Acknowledgement

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-2020-0-01797) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and the Pukyong National University Industry-university Cooperation Research Fund in 2023(202312400001).

References

[1] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”, in *Advances in Neural Information Processing Systems*, 2017.

[2] Q. Hu, Bo Yang, L. Xie, S. Rosa, and Y. Guo, “RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds”, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11105–11114, 2020.

[3] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, “Point transformer”, in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 16259–16268, 2021.

[4] S. Laine and T. Aila, “Temporal ensembling for semi-supervised learning”, in *ICLR*, 2017.

[5] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results”, in *Advances in neural information processing systems*, vol. 30, 2017.

[6] Y. Zhang, Z. Li, Y. Xie, Y. Qu, C. Li, and T. Mei, “Weakly supervised semantic segmentation for large-scale point cloud,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, pp. 3421–3429, 2021.

[7] X. Xu and G. H. Lee, “Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13706–13715, 2020.

[8] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15520–15528, 2021.

[9] Q. Hu, B. Yang, G. Fang, Y. Guo, A. Leonardis, N. Trigoni, and A. Markham, “Sqn: Weakly-supervised semantic segmentation of large-scale 3d point clouds,” in *Computer Vision–ECCV 2022: 17th European Conference*, Tel Aviv, Israel, October 23–27, Proceedings, Part XXVII, pp. 600–619, 2022.