Supplementary Material for "Efficient LiDAR Point Cloud Oversegmentation Network"

Le Hui[1], Linghua Tang[2], Yuchao Dai[1], Jin Xie[2*] and Jian Yang[2*]
[1]Northwestern Polytechnical University, [2]Nanjing University of Science and Technology
huile@nwpu.edu.cn, {tanglinghua, csjxie, csjyang}@njust.edu.cn, daiyuchao@gmail.com

A. Overview

This supplementary material provides more details about our method. In Sec. B, we compare our method with other oversegmentation methods. In Sec. C, we provide the details of our method, including visualization results, limitations, and impacts. We also discuss future work. We will release the code after the paper is published.

B. Oversegmentation Methods Comparison

B.1. Flexibility and Scalability

In order to show the advantages of our method, we compare the proposed SuperLiDAR with advanced point cloud oversegmentation methods, including SPG [9], SSP [7], and SPNet [6]. The detailed comparison is shown in Tab. 1. Specifically, we analyze three important factors, including data preprocessing, end-to-end network, and superpoint size control. SPG is an optimization based method that segments superpoints through a greedy graph-cut algorithm [8]. The greedy graph-cut algorithm uses different thresholds of energy to segment superpoints, so it cannot effectively control the minimum or maximum size of the superpoint. SSP uses deep features to replace handcrafted features used in SPG to generate superpoints. SSP is a two-stage method, and cannot precisely control superpoint size. Notably, SSP requires preparing data (constructing the graph and computing handcrafted features) as the network input in advance. SPNet follows the same data processing procedure as used in SSP. SPNet is a clustering based method that directly clusters superpoints from the input point cloud. The number of superpoints is determined by the initial seed points. The more numbers of seed points, the smaller the superpoint size is. Thus, SPNet cannot directly control the superpoint size. In this paper, the proposed SuperLiDAR uses the breadth-first search (BFS) to generate superpoints based on the learned low-level point embeddings. During BFS, our method can control the superpoint size by setting the minimum/maximum number of points in each superpoint. Therefore, our method is flexible to generate multi-scale superpoints. In addition, our method does not require data processing and directly takes raw point clouds as input to generate superpoints. Thanks to the simple yet effective designs, our method can be easily integrated into the current network to form an end-to-end framework. Compared with other oversegmentation methods, the proposed SuperLiDAR has good flexibility and scalability.

Method	No Data Preprocessing	End-to-End	Control Size (Min & Max)
SPG [9]	✗	✗	✗
SSP [7]	✗	✗	✗
SPNet [6]	✗	✓	✗
SuperLiDAR (ours)	✓	✓	✓

Table 1. Comparison of different oversegmentation methods.

B.2. Potential Applications

The proposed LiDAR point cloud oversegmentation method SuperLiDAR is flexible and scalable and will allow the application of a series of downstream tasks, such as 3D segmentation, 3D detection, and 3D tracking. Furthermore, the speed of our SuperLiDAR is $100\times$ faster than other oversegmentation methods.

In Fig. 1, we show the generated superpoints by our SuperLiDAR on the validation set of SemanticKITTI. In the figure, we cycle four categories, including pedestrians, cyclists, traffic signs, and cars. It can be observed that whether small objects (pedestrians, cyclists, and traffic signs) or large objects (cars), our method can generate good superpoints from the LiDAR point clouds. Essentially, the superpoint is a representation of the point cloud. Superpoint can adaptively capture the local geometric structures of point clouds. Compared with hard local neighborhoods,

*Corresponding authors.
Le Hui and Yuchao Dai are with Shaanxi Key Laboratory of Information Acquisition and Processing, China. Linghua Tang, Jin Xie and Jian Yang are with Key Lab of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, and Jiangsu Key Lab of Image and Video Understanding for Social Security, China.

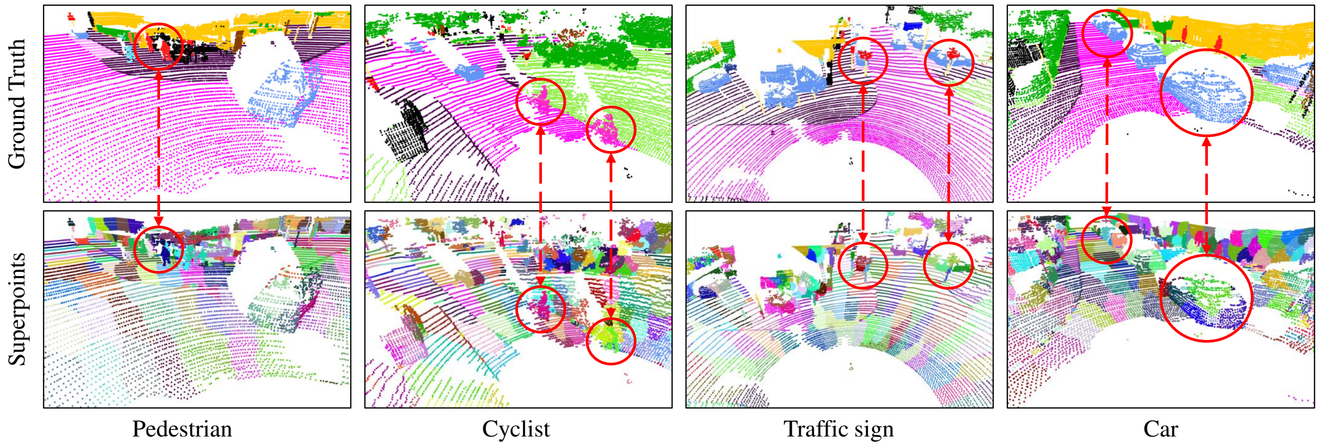


Figure 1. Visualization of superpoints generated by our SuperLiDAR on the validation set of SemanticKITTI. Note that the superpoints are randomly colored (zoom in for a better view).

such as voxel based, ball query based, and k-NN based neighborhoods, adaptive neighborhoods ($i.e.$, superpoint) are built along the geometric structures of the point cloud surface. By applying superpoints, we can obtain more discriminative local features, thereby improving the performance of downstream tasks.

In this paper, we have demonstrated that using superpoints can effectively improve the performance of semantic segmentation. In our SuperLiDAR, we use semantic labels to formulate the local discriminative loss for learning low-level point embeddings of point clouds. Note that the semantic labels are used to learn low-level point embedding rather than high-level semantic information. Thus, semantic labels can be replaced by different supervision signals that can distinguish different objects. In 3D detection, we can formulate a new detection framework based on superpoints. Specifically, we first extract superpoint-level features of point clouds. Then, we use a foreground prediction branch to filter out potential objects by masking background superpoints. Finally, we can build a superpoint clustering branch ($e.g.$, cluster superpoints of car in Fig. 1) to obtain proposals for classification and bounding box prediction. Similarly, in 3D multi-object tracking, we can formulate a new framework based on the tracking-by-detection scheme. Specifically, we first detect objects in each frame, and then track objects based on the similarity between superpoints in adjacent frames. Notably, the proposed SuperLiDAR will not heavily burden the downstream tasks due to the simple yet effective design.

C. Our SuperLiDAR

C.1. More Results

Quantitative Results In the main paper, we show the results of semantic segmentation on the online test sets of SemanticKITTI [2] and nuScenes [3]. In Tab. 2, we also provide the results of semantic segmentation on the validation sets of SemanticKITTI and nuScenes. It can be observed that our method can significantly improve the baseline results.

Method	SemanticKITTI (mIoU)	nuScenes (mIoU)
Baseline	65.6	77.8
SuperLiDAR (ours)	**69.3**	**79.5**

Table 2. Semantic segmentation results on the validation sets of SemanticKITTI and nuScenes.

In addition, we mainly focus on autonomous driving applications, so we conduct experiments on outdoor SemanticKITTI and nuScenes. Our method can be applied to indoor S3DIS [1] and ScanNet [4]. As shown in Tab. 3, our method obtains better results compared to the previous state-of-the-art method SPNet [6].

Method	OOA	BR	BP	F1
S3DIS Area 5				
SPNet [6]	96.5	82.1	13.1	22.6
SuperLiDAR (ours)	**97.3**	**83.3**	**14.0**	**23.9**
ScanNet				
SPNet [6]	96.4	79.1	13.5	23.1
SuperLiDAR (ours)	**97.4**	**80.4**	**14.1**	**24.0**

Table 3. Comparison results of generated superpoints on the S3DIS and ScanNet.

Ablation Studies Here we provide more ablation studies of oversegmentation and semantic segmentation on the validation set of SemanticKITTI [2].

In the LiDAR point grouping algorithm, the threshold γ of distance constraint in breadth-first search (BFS) is an important parameter. In Tab. 4, we show the oversegmentation results under different thresholds $\gamma \in$

Setting	OOA	BR	BP	F1
$\gamma = 0.5$	95.80	62.08	**22.15**	**32.65**
$\gamma = 1.0$	96.01	65.11	21.02	31.78
$\gamma = 1.5$ (default)	96.21	**65.52**	20.52	31.25
$\gamma = 2.0$	**96.30**	65.49	19.60	30.17
$\gamma = 2.5$	96.23	65.91	18.42	28.79
$K = 1$	94.81	57.72	17.98	27.41
$K = 3$	95.67	63.13	18.78	28.94
$K = 5$ (default)	96.21	65.52	**20.52**	**31.25**
$K = 7$	96.13	**65.74**	20.38	31.11
$K = 9$	**96.41**	65.27	20.09	30.72

Table 4. Oversegmentation results on the validation set of SemanticKITTI under different settings.

$\{0.5, 1.0, 1.5, 2.0, 2.5\}$. It can be observed that when γ is greater than 1.5m, we can obtain higher OOA, which measures the theoretical upper bound of semantic segmentation using superpoints. Considering the performance of other metrics BR, BP, and F1, we set $\gamma = 1.5$m in this paper.

In the superpoint refinement module, the hyperparameter K indicates the number of candidate superpoints that are used to learn point-superpoint affinity. In Tab. 4, we show the oversegmentation results at different values of K. Note that $K = 1$ means that we directly assign the point to the nearest superpoint in the coordinate space. It can be observed that as the value of K increases, the performance of OOA increases. In addition, BR, BP, and F1 achieve the best results when the value of K is about 5. It is worth noting that as the value of K increases, the computational cost gradually increases. To balance the computational cost and performance, we set $K = 5$ in this paper.

In LiDAR semantic segmentation, we propose a multi-scale superpoint aggregation module and integrate it with our LiDAR oversegmentation network to form an end-to-end LiDAR semantic segmentation framework. In Tab. 5, we show the ablation study of semantic segmentation with different numbers of superpoint scales. Note that different scales of superpoints are generated by adjusting the minimum size (N_{min}) and maximum size (N_{max}) in the LiDAR point grouping algorithm. It can be observed that our method achieves the best results when using three scales of superpoints ("SuperLiDAR (scale num. = 3, default)").

Ideal superpoints are located inside the instance, but not cross different instances. Since the semantic segmentation loss $\mathcal{L}_{ce}$ cannot distinguish different instances, it could generate cross-instance superpoints, resulting in lower performance. In contrast, the proposed local discriminative loss $\mathcal{L}_{sp}$ can identify different instances by capturing the discriminability of local geometric structures, thereby generating high-quality superpoints. As a result, we conduct ablation studies on the validation set of SemanticKITTI. The OOA/BR/BP/F1 are 95.0/63.6/18.1/28.1 ($\mathcal{L}_{ce}$) and **96.2/65.5/20.5/31.2** ($\mathcal{L}_{sp}$).

Setting	mIoU
SuperLiDAR (scale num. = 1)	67.8
SuperLiDAR (scale num. = 2)	68.5
SuperLiDAR (scale num. = 3, default)	**69.3**
SuperLiDAR (scale num. = 4)	69.0

Table 5. Semantic segmentation results on the validation set of SemanticKITTI under different settings.

Visualization Results Here we provide more visualization results of oversegmentation and semantic segmentation to show the effectiveness of the proposed SuperLiDAR.

In Fig. 2, we show the visualization results of superpoints before applying the superpoint refinement module. In the figure, the points with black color are usually the boundary points (in red circles) that are not assigned to the superpoints in LiDAR point grouping algorithm. After applying the superpoint refinement module, it can be observed that these black points can be accurately assigned to the corresponding superpoints.

In Figs. 3 and 4, we show the results (*i.e.*, superpoints) of LiDAR point cloud oversegmentation on the validation sets of SemanticKITTI [2] and nuScenes [3], respectively. In the figures, the superpoints are randomly colored. Please zoom in for a better view. It can be observed that compared with SPG [9], SSP [7], and SPNet [6], our SuperLiDAR can generate high-quality superpoints with clear boundaries.

In Figs. 5 and 6, we show the results of semantic segmentation on the validation sets of SemanticKITTI and nuScenes, respectively. From the error maps, it can be observed that using superpoints can effectively improve the performance of semantic segmentation.

C.2. Limitations and Impacts

Limitations The proposed SuperLiDAR is a supervised method that requires annotated labels (*e.g.*, semantic labels and instance labels) to train the network. Therefore, it can be integrated into downstream tasks (*e.g.*, semantic segmentation and instance segmentation) to form a multi-task framework in a supervised manner. However, for unsupervised, semi-supervised, and self-supervised downstream tasks, such as unsupervised point cloud correspondence, weakly supervised point cloud semantic segmentation and 3D detection, the proposed method is limited to these tasks.

Impacts The proposed method can be applied to self-driving cars and transportation. For self-driving cars, some objects (such as cars, trucks, and pedestrians) on the road may be incorrectly recognized by the proposed method. These issues require further research and consideration when building upon this work in self-driving cars.

Ethical Consideration This work is able to facilitate

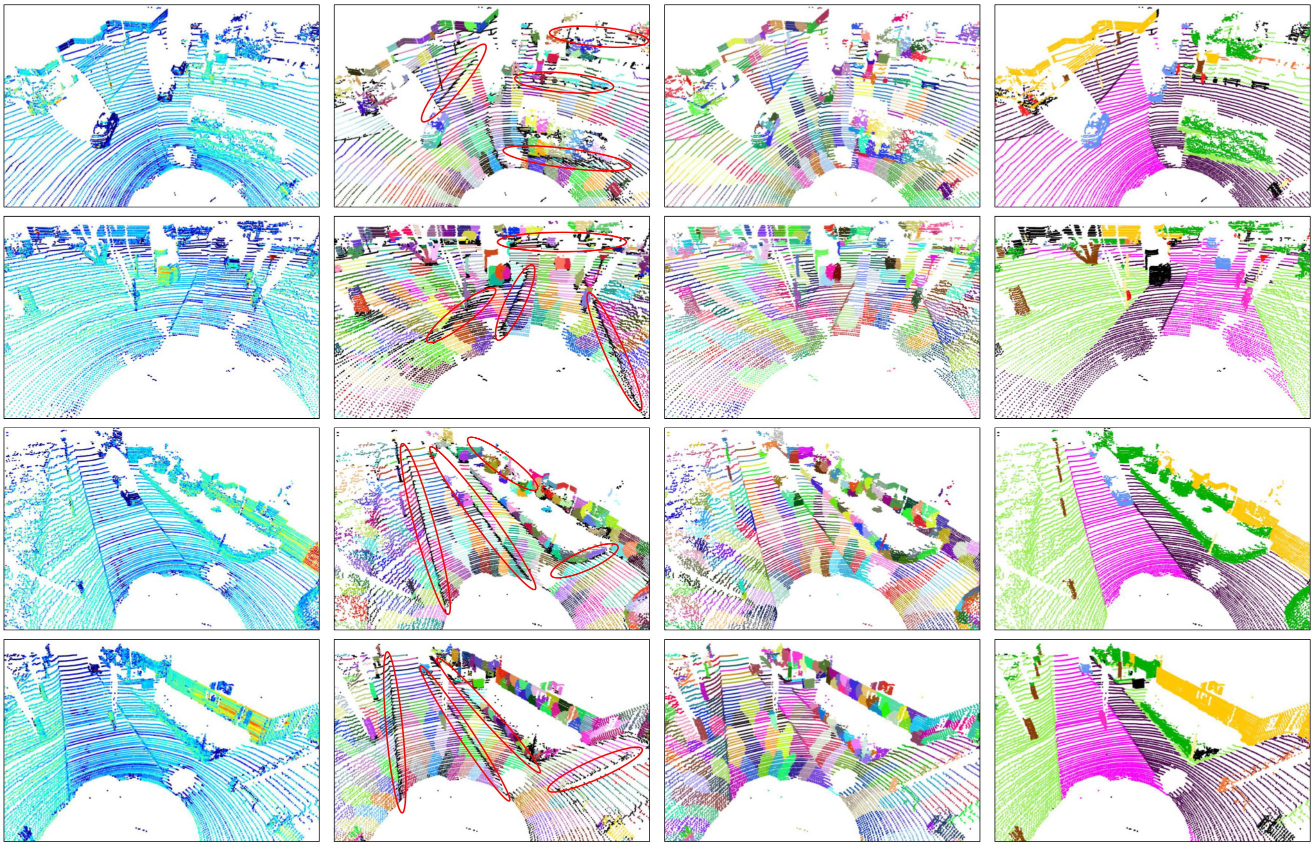
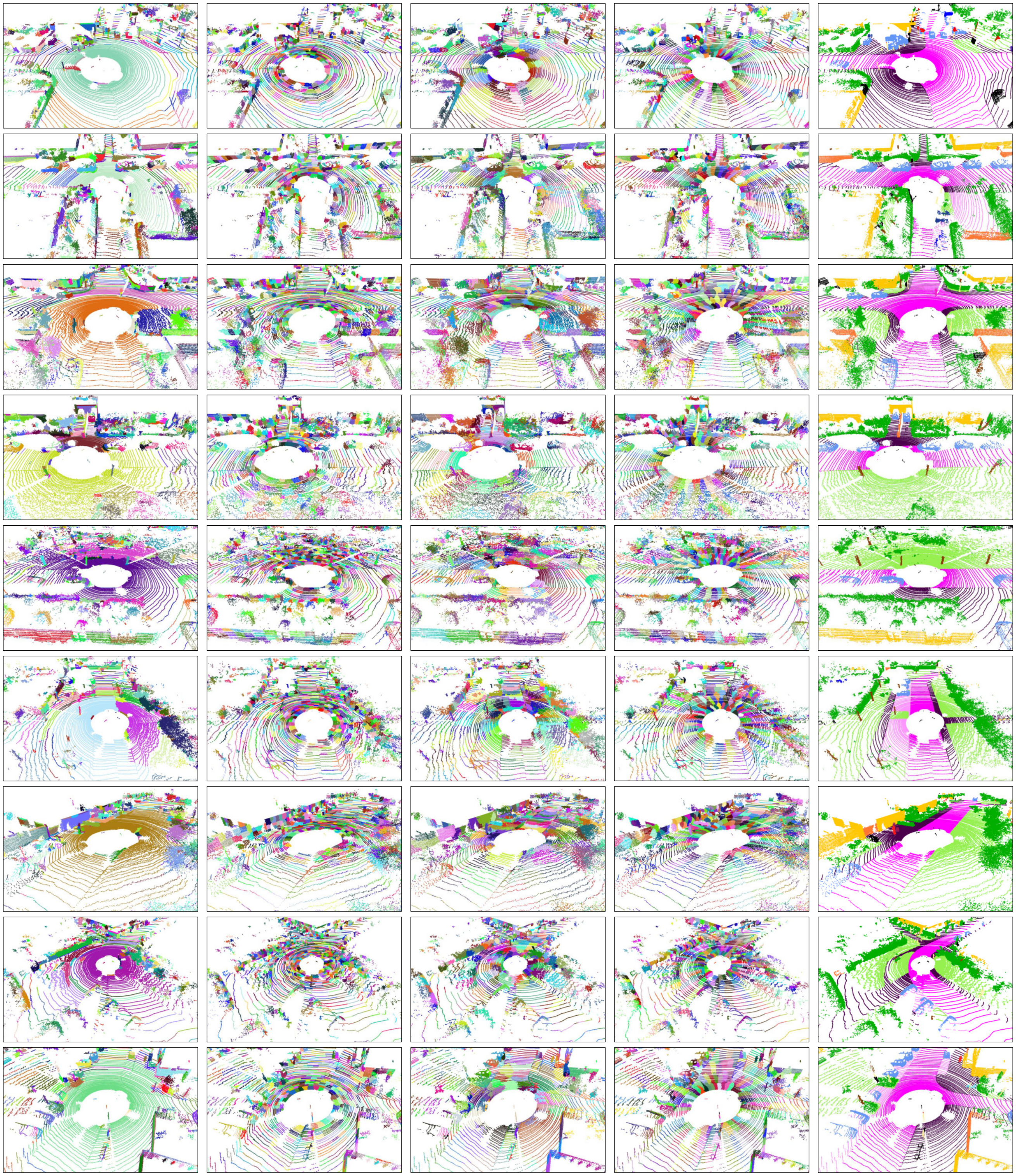


Figure 2. Visualization of superpoints generated by our method on the validation set of SemanticKITTI. Note that the superpoints are randomly colored (zoom in for a better view). The points with black colors are usually the boundary points (in red circles) that are not assigned to the superpoints in the LiDAR point grouping algorithm.

the development of certain applications. For example, it can help domestic robots avoid potential obstacles in indoor environments. In assisted driving, it can help the driver recognize potential objects that may affect driving in advance. In addition, all datasets used in this paper are publicly available as academic research, and the evaluation metrics used in the experiments are also standard. For negative outcomes, it depends on a specific task and the criteria for assessing positive and negative.

C.3. Future Work

In future, we consider utilizing the masked auto-encoder (MAE) [5] to improve the proposed SuperLiDAR to a self-supervised method. In this way, it can be potentially applied to a wider range of tasks, such as supervised, self-supervised, and unsupervised tasks in 3D scene understanding.



SPG SSP SPNet SuperLiDAR (ours) Ground Truth

Figure 3. Visualization of superpoints generated by different methods on the validation set of SemanticKITTI. Note that the superpoints are randomly colored (zoom in for a better view).

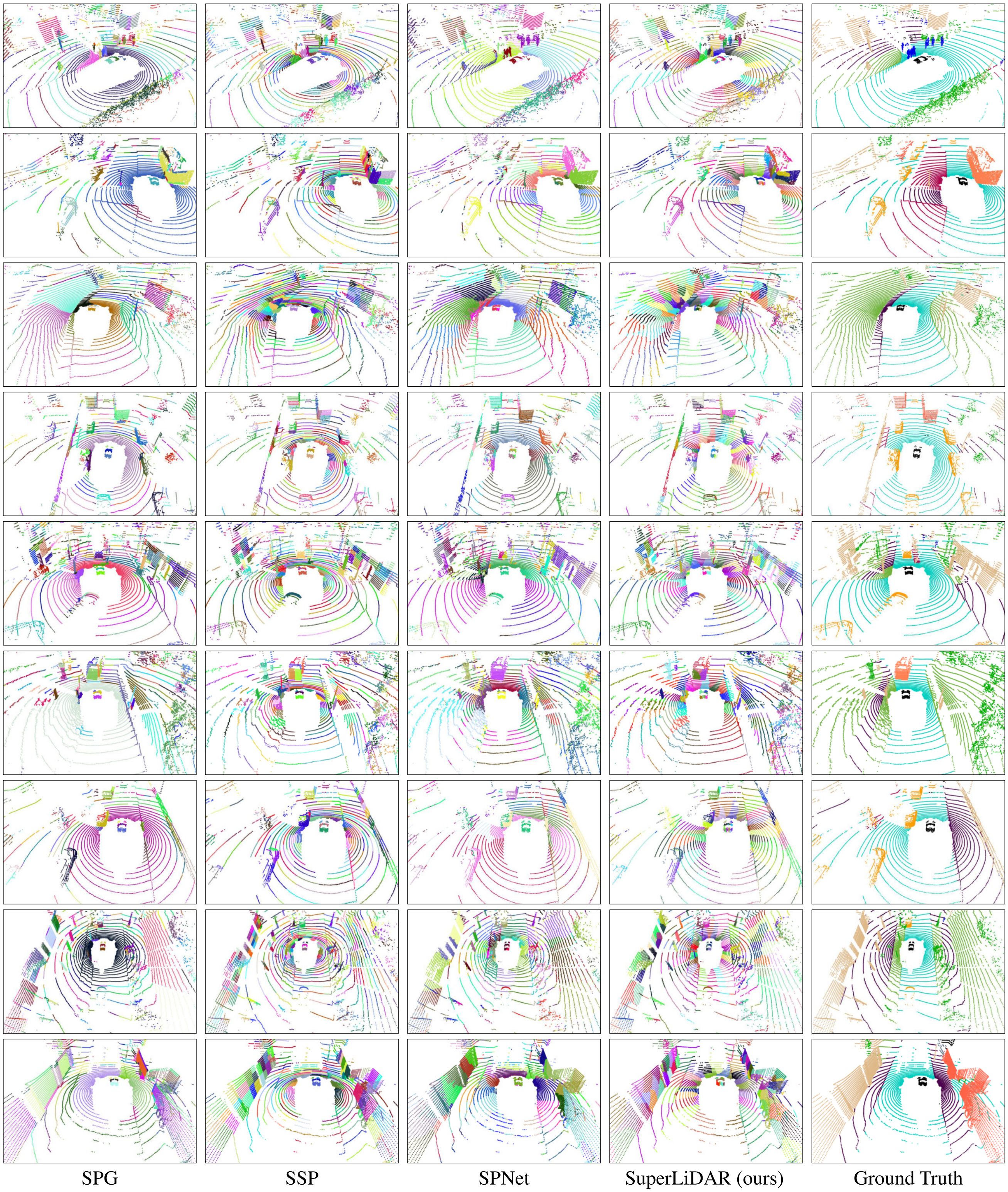


Figure 4. Visualization of superpoints generated by different methods on the validation set of nuScenes. Note that the superpoints are randomly colored (zoom in for a better view).

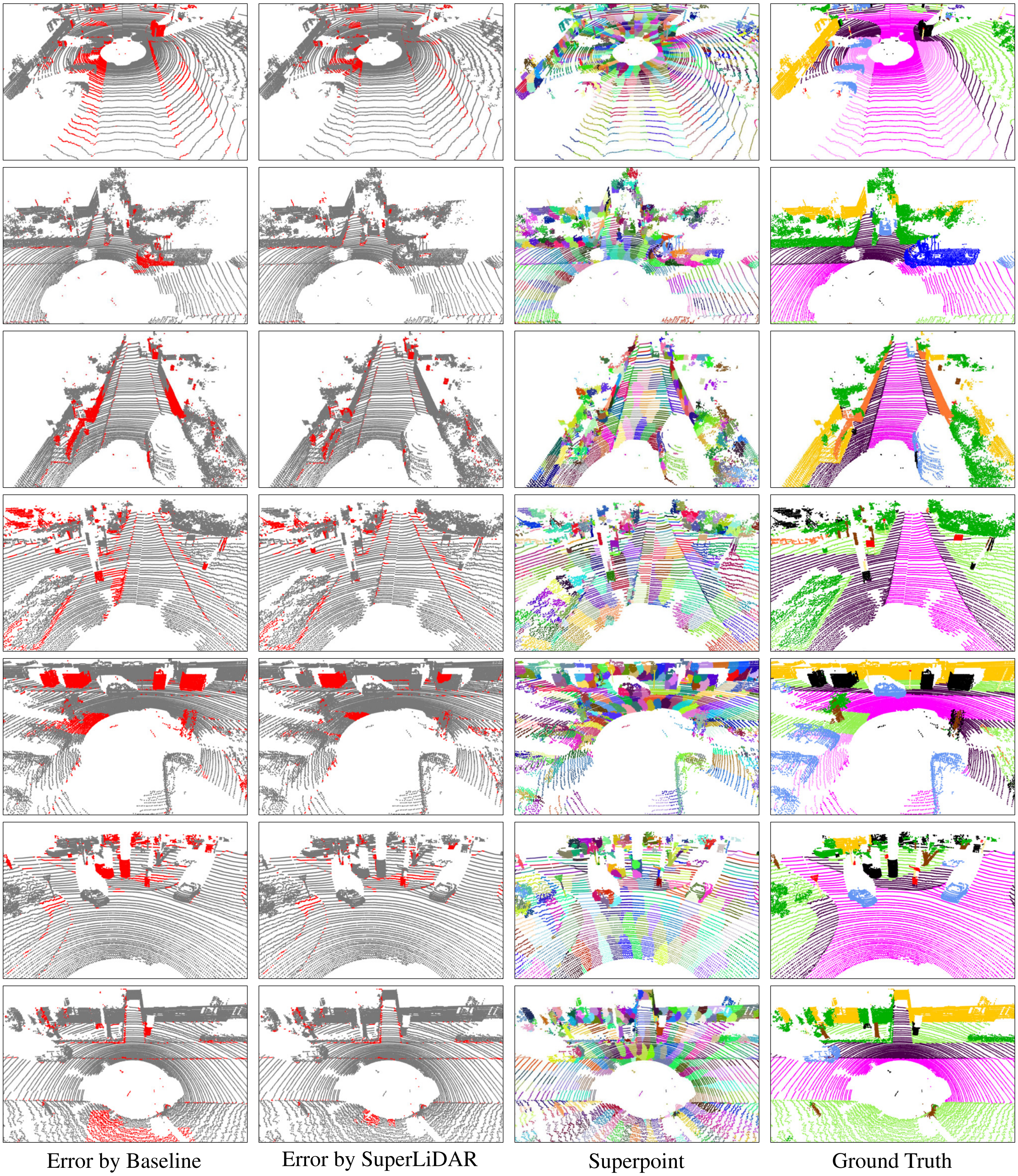


Figure 5. Visualization of semantic segmentation on the validation set of SemanticKITTI. Note that the superpoints are randomly colored (zoom in for a better view).

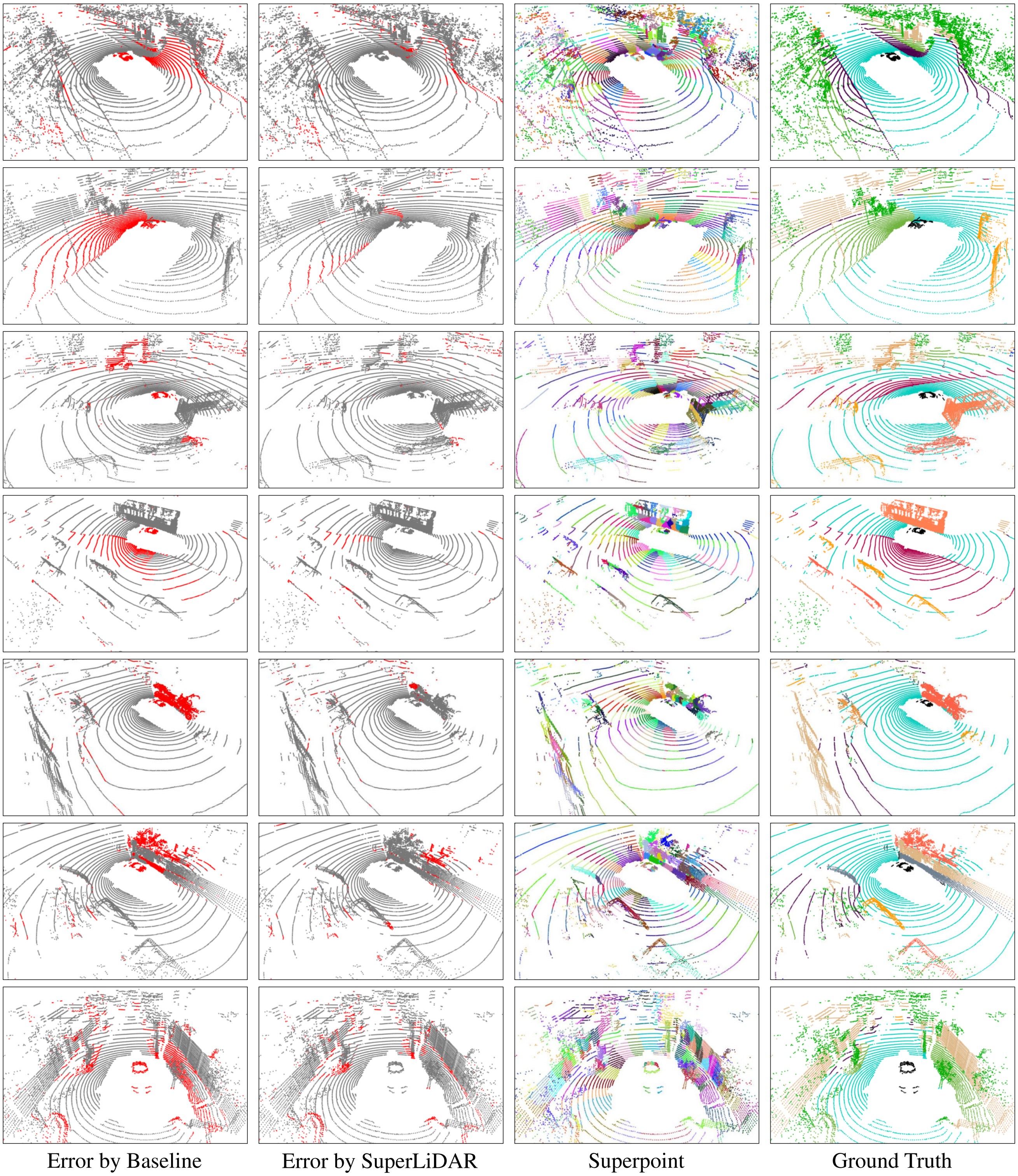


Figure 6. Visualization of semantic segmentation on the validation set of nuScenes. Note that the superpoints are randomly colored (zoom in for a better view).

References

[1] 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 *CVPR*, 2016. 2

[2] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. SemanticKITTI: A dataset for semantic scene understanding of lidar sequences. In *ICCV*, 2019. 2, 3

[3] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuScenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020. 2, 3

[4] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *CVPR*, 2017. 2

[5] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, 2022. 4

[6] Le Hui, Jia Yuan, Mingmei Cheng, Jin Xie, and Jian Yang. Superpoint network for point cloud oversegmentation. In *ICCV*, 2021. 1, 2, 3

[7] Loic Landrieu and Mohamed Boussaha. Point cloud oversegmentation with graph-structured deep metric learning. In *CVPR*, 2019. 1, 3

[8] Loic Landrieu and Guillaume Obozinski. Cut Pursuit: fast algorithms to learn piecewise constant functions. In *AISTATS*, 2016. 1

[9] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *CVPR*, 2018. 1, 3