

# SGG-Nets: Generic Rotation-Invariant Plugin Networks for Point Cloud Analysis

## Supplementary Materials

Jian Zhu, *Member, IEEE*, Jianrong Yan, Jiebin Huang, Yongwei Nie, *Member, IEEE*, Bin Sheng, *Member, IEEE*, and Tong-Yee Lee, *Senior Member, IEEE*

**Abstract**—This document includes a description of the network architecture of SGG-PaRot, along with several supplementary experiments for the main paper.

### I. NETWORK ARCHITECTURE OF SGG-PAROT

To further demonstrate the versatility of SGG-Net as a plug-in module, we select the recent PaRot [9] method as an additional baseline and integrate SGG-Net into it. The network structure of PaRot [9] is shown in the gray area of Figure 11, with detailed explanations available in the original paper. Similar to DGCNN [7], PaRot [9] does not have any subsampling operations, and the neighborhood graph only needs to be constructed once. Therefore, a single M-SGGConvMdl module is sufficient to construct the plugin to produce the final SGG-PaRot, as illustrated in Figure 11. Specifically, considering the network structure of PaRot [9], we equip an SGGConv layer for each LBR layer in PaRot and concatenate the outputs of both to feed into the next LBR layer. Ultimately, our M-SGGConvMdl consists of three sequentially connected SGGConv layers. To achieve a balance between model performance and computational efficiency, we set the values of parameters  $S$  and  $K$  in SGGConvMdl (see Table I) to 64 and 30, respectively. This straightforward integration process enables the incorporation of rotation-invariant features from SGG-Net into PaRot [9] with minimal modifications.

### II. SUPPLEMENTARY EXPERIMENTS

#### A. Classification on FG3D

**Data.** To further validate the effectiveness of our proposed SGG-Net, we conduct comprehensive classification experiments on the recent fine-grained 3D shape dataset (FG3D) [5]. This dataset comprises three distinct categories: airplane, car, and chair. Each category contains several subcategories with very subtle discrepancies between them, requiring the network to have a higher level of capability to capture details of shape. The airplane category consists of 4,173 models, categorized into 13 subcategories, with 3,441 models allocated for training and 732 for testing. The chair category comprises 13,054 models, divided into 33 subcategories, with 11,124 models allocated for training and 1,930 for testing. Similarly, the car category contains 8,325 models, categorized into 20 subcategories, with 7,010 models allocated for training and 1,315 for

testing. All models are rotated with  $z/SO3$  case to evaluate the performance of our SGG-Nets.

**Implementation.** The network architecture of each model follows the same structure as trained with ModelNet40 [8], but the output dimensions of the classifier are adjusted according to the selected categories from FG3D [5], specifically 13 for airplane, 33 for chair, and 20 for car. Since the classification results on FG3D are not provided in the original publications of the baseline networks, we adopt the experimental configuration established for ModelNet40 [8]. All models are trained for 300 epochs with a batch size of 32.

**Results.** The experimental results are shown in Table XII and overall accuracy (OA) is used as the evaluation metric. Three categories from FG3D [5] exhibit significant differences in difficulty, with most methods performing best in the airplane category and worst in the car category. Accuracies across all the categories of two regular networks (i.e., PointNet++ [6] and DGCNN [7]) are significantly improved by integrating SGG-Net. Specifically, SGG-Net demonstrates the most substantial impact when integrated with PointNet++ [6], achieving a 23.22% accuracy improvement in the car category. Additionally, when combined with DGCNN [7] it yields a substantial improvement of 11.92% in classification accuracy for the chair category. Even for specialized networks, i.e., RICov++ [10] and PaRot [9], which already perform excellently, accuracies can still be improved to some extent. Overall, SGG-PaRot outperforms all methods across three categories, achieving the highest accuracy. These results further demonstrate the effectiveness of our proposed SGG-Net.

TABLE XII: Classification results on FG3D with  $z/SO3$  rotation.

	Method	airplane	chair	car
Baseline	PointNet++ [6]	79.98	48.04	39.43
	DGCNN [7]	74.31	41.03	44.41
	RICov++ [10]	96.01	79.64	76.03
	PaRot [9]	97.26	81.15	77.87
Ours	SGG-PointNet++	93.25 (↑ 13.27)	59.98 (↑ 11.94)	62.65 (↑ 23.22)
	SGG-DGCNN	89.89 (↑ 15.58)	52.95 (↑ 11.92)	61.44 (↑ 17.03)
	SGG-RICov++	97.37 (↑ 1.36)	81.37 (↑ 1.73)	76.86 (↑ 0.83)
	SGG-PaRot	<b>97.45</b> (↑ 0.19)	<b>81.53</b> (↑ 0.38)	<b>78.12</b> (↑ 0.25)

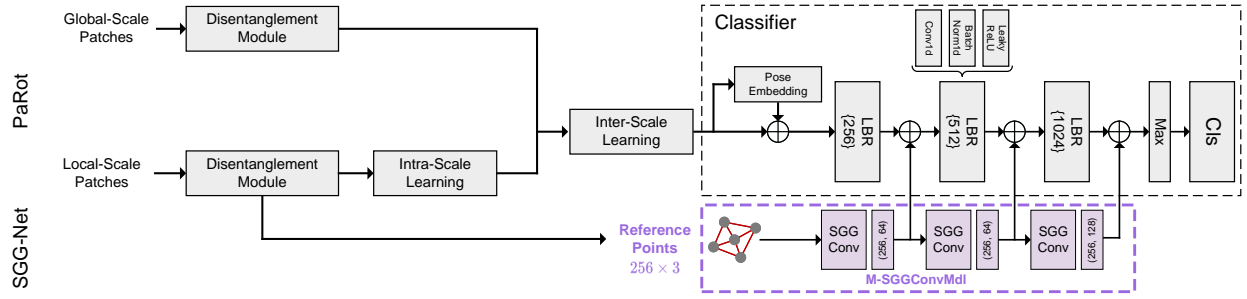


Fig. 11: Network architecture of SGG-PaRot. We construct our SGG-Net using a single M-SGGConvMdl module, which consists of three SGGConv layers, and seamlessly integrate it with the baseline PaRot [9] by concatenating each SGGConv layer to the corresponding LBR layer in PaRot [9].

### B. Semantic Segmentation on S3DIS

**Data.** We conduct semantic segmentation experiments on the Stanford Large-Scale 3D Indoor Spaces Dataset (S3DIS) [1]. The dataset includes 6 indoor areas, which contain a total of 272 rooms, and 13 semantic categories can be classified in each room, e.g. board, bookcase, chair, ceiling, beam, and clutter. Following previous work [6] [7], we sample 4096 points in each block as input, and each point is represented as a 9D vector including 3D coordinate values ( $x$ ,  $y$ ,  $z$ ), RGB, and normalized spatial coordinates.

TABLE XIII: Semantic segmentation results on S3DIS.

Method	$z/SO3$	$SO3/SO3$
PointNet++ [6]	10.1	68.2
DGCNN [7]	12.8	68.0
SGG-PointNet++ (ours)	25.3 ( $\uparrow$ 15.2)	69.4 ( $\uparrow$ 1.2)
SGG-DGCNN (ours)	20.0 ( $\uparrow$ 7.2)	70.0 ( $\uparrow$ 2.0)

**Results.** Following the work of DGCNN [7], all networks are trained on Area-1 to Area-5 and tested on Area-6. Due to the rarity of methods that perform segmentation testing on the S3DIS [1] dataset under the three challenging rotation-based training/testing settings, we only compare the performance of the baseline networks before and after integrating SGGConvs, and only select PointNet++ [6] and DGCNN [7] as the baseline.

Table XIII presents the semantic segmentation results under two rotation settings:  $z/SO3$  and  $SO3/SO3$ . After integrating our SGGConvs, SGG-PointNet++ showed a remarkable increase of 15.2% in mIoU for the  $z/SO3$  case, and an improvement of 1.2% for the  $SO3/SO3$  case. The mIoU of SGG-DGCNN has increased by 7.2% compared to DGCNN [7] in the  $z/SO3$  case, and by 2.0% in the  $SO3/SO3$  case. These results further demonstrate that our SGG-Net can enhance the rotation-invariance of the baseline networks. Besides, we visualize the semantic segmentation results of some selected rooms in Area-6 (see Figure 12), from which it can be seen that the improved networks handle some of the boundary regions more finely, resulting in better segmentation results.

### REFERENCES

- [1] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3d semantic parsing of large-scale indoor spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1534–1543.
- [2] J. Fei and Z. Deng, "Rotation invariance and equivariance in 3d deep learning: a survey," *Artificial Intelligence Review*, vol. 57, no. 7, p. 168, 2024.
- [3] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 12, pp. 4338–4364, 2020.
- [4] Z. Kuang, J. Yu, S. Zhu, Z. Li, and J. Fan, "Effective 3-d shape retrieval by integrating traditional descriptors and pointwise convolution," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3164–3177, 2019.
- [5] X. Liu, Z. Han, Y.-S. Liu, and M. Zwicker, "Fine-grained 3d shape classification with hierarchical part-view attentions," *IEEE Transactions on Image Processing*, 2021.
- [6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.
- [7] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.
- [8] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3d shapenets: A deep representation for volumetric shapes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1912–1920.
- [9] D. Zhang, J. Yu, C. Zhang, and W. Cai, "PaRot: Patch-wise rotation-invariant network via feature disentanglement and pose restoration," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, pp. 3418–3426, Jun. 2023.
- [10] Z. Zhang, B.-S. Hua, and S.-K. Yeung, "Rconv++: Effective rotation invariant convolutions for 3d point clouds deep learning," *International Journal of Computer Vision*, vol. 130, no. 5, pp. 1228–1243, 2022.

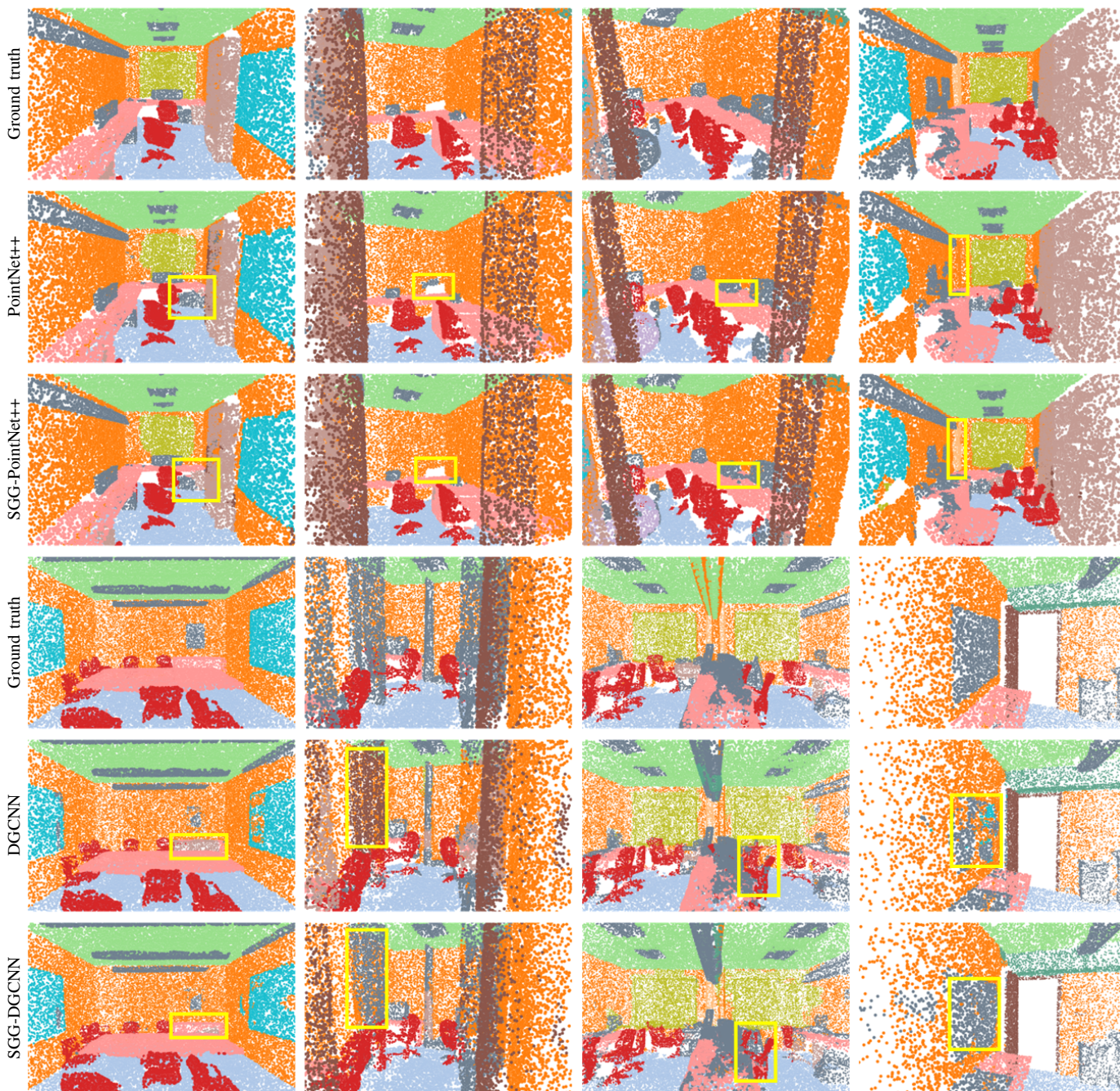


Fig. 12: Visualization of the semantic segmentation results of some selected rooms in Area-6 of the S3DIS dataset in the  $SO3/SO3$  case. From top to bottom: Ground truth, PointNet++, SGG-PointNet++, Ground truth, DGCNN, and SGG-DGCNN.