Rethinking Few-shot 3D Point Cloud Semantic Segmentation
Supplementary Material

Zhaochong An[1,2], Guolei Sun[1*], Yun Liu[3*], Fayao Liu[3], Zongwei Wu[4], Dan Wang[2], Luc Van Gool[1], Serge Belongie[2]
[1] Computer Vision Laboratory, ETH Zurich
[2] Pioneer Centre for Artificial Intelligence, University of Copenhagen
[3] Institute for Infocomm Research, A*STAR
[4] Computer Vision Lab, CAIDAS & IFI, University of Wurzburg

1. More Details about Foreground Leakage

As discussed in Sec. 3.2, the current few-shot 3D point cloud semantic segmentation (FS-PCS) setting [3, 5–10] employs a non-uniform sampling mechanism with a bias toward foreground classes. This biased sampling algorithm samples more points from foreground objects than from the background, resulting in a noticeable point density disparity between foreground and background.

More precisely, the biased sampling algorithm can be outlined in Alg. 1[1]. In line 1, it firstly obtains the input foreground point set $\mathbf{P}_{\mathrm{FG}}$ that includes all the input points belonging to the foreground class C with respect to the current few-shot task. Then, from lines 2 to 6, it calculates the quantity N_{FG} that will be used for sampling foreground points in the output. N_{FG} maintains a proportional relationship to the presence of foreground points in the input data when $n \geq m$. Next, in line 7, it selects N_{FG} points exclusively from the input foreground point set $\mathbf{P}_{\mathrm{FG}}$. However, in line 8, the remaining $m - N_{\mathrm{FG}}$ points are sampled from the entire input points $\mathbf{X} = \{\mathbf{P}_1, .., \mathbf{P}_{\mathrm{n}}\}$, which still includes the foreground points in $\mathbf{P}_{\mathrm{FG}}$. Consequently, this double-sampling of foreground points in these two steps leads to foreground objects having a denser distribution of points in the final output than their background counterparts.

We also present additional visualizations in Fig. 1. Both the theoretical analysis and visualizations clearly demonstrate that this biased sampling leaks foreground class information to models through density disparity. Consequently, the models no longer need to excel at learning essential knowledge adaptation patterns for few-shot tasks; instead, they can simply segment the target by detecting denser regions. This foreground leakage undermines the validity of existing benchmarks of previous models.

Algorithm 1: The biased sampling algorithm

Data: input point cloud $\mathbf{X}$ with n points $\{\mathbf{P}_1, .., \mathbf{P}_{\mathrm{n}}\}$, sampling number m, foreground class C with respect to current few-shot task
Result: sampled points $\{\mathbf{P}_{\mathrm{i}_1}, .., \mathbf{P}_{\mathrm{i}_\mathrm{m}}\}$ from $\mathbf{X}$

1 $\mathbf{P}_{\mathrm{FG}} \leftarrow \{\mathbf{P}_{\mathrm{i}} \mid label\_of(\mathbf{P}_{\mathrm{i}}) = C\}$;
2 **if** $n < m$ **then**
3 $\quad N_{\mathrm{FG}} \leftarrow |\mathbf{P}_{\mathrm{FG}}|$;
4 **else**
5 $\quad N_{\mathrm{FG}} \leftarrow m\frac{|\mathbf{P}_{\mathrm{FG}}|}{n}$;
6 **end**
7 $\mathbf{Res}_1 \leftarrow$ sample N_{FG} points from $\mathbf{P}_{\mathrm{FG}}$;
8 $\mathbf{Res}_2 \leftarrow$ sample $m - N_{\mathrm{FG}}$ points from $\mathbf{X}$;
9 $\{\mathbf{P}_{\mathrm{i}_1}, .., \mathbf{P}_{\mathrm{i}_\mathrm{m}}\} \leftarrow \mathbf{Res}_1 \cup \mathbf{Res}_2$;

2. More Implementation Details

We employ the first three blocks from the Stratified Transformer [4] as our backbone. Our backbone architecture aligns with the one used for the S3DIS dataset [1] in [4], indicating that we maintain consistency in backbone architectures for both S3DIS and ScanNet [2]. Unlike [4], we do not employ different Stratified Transformer architectures for these two datasets. The momentum coefficient μ within the BPC module is set to 0.995. For both datasets, our input features include both the XYZ coordinates and RGB colors. The training and testing are using 4 RTX 3090 GPUs.

3. More Qualitative Results

We present additional qualitative results in Fig. 2, comparing our method (5th column) with the previous best-performing method, QGE (6th column). Besides, Fig. 3 showcases more visual comparisons between our models

*Corresponding authors: Guolei Sun and Yun Liu
[1]The corresponding source code can be found at the link.

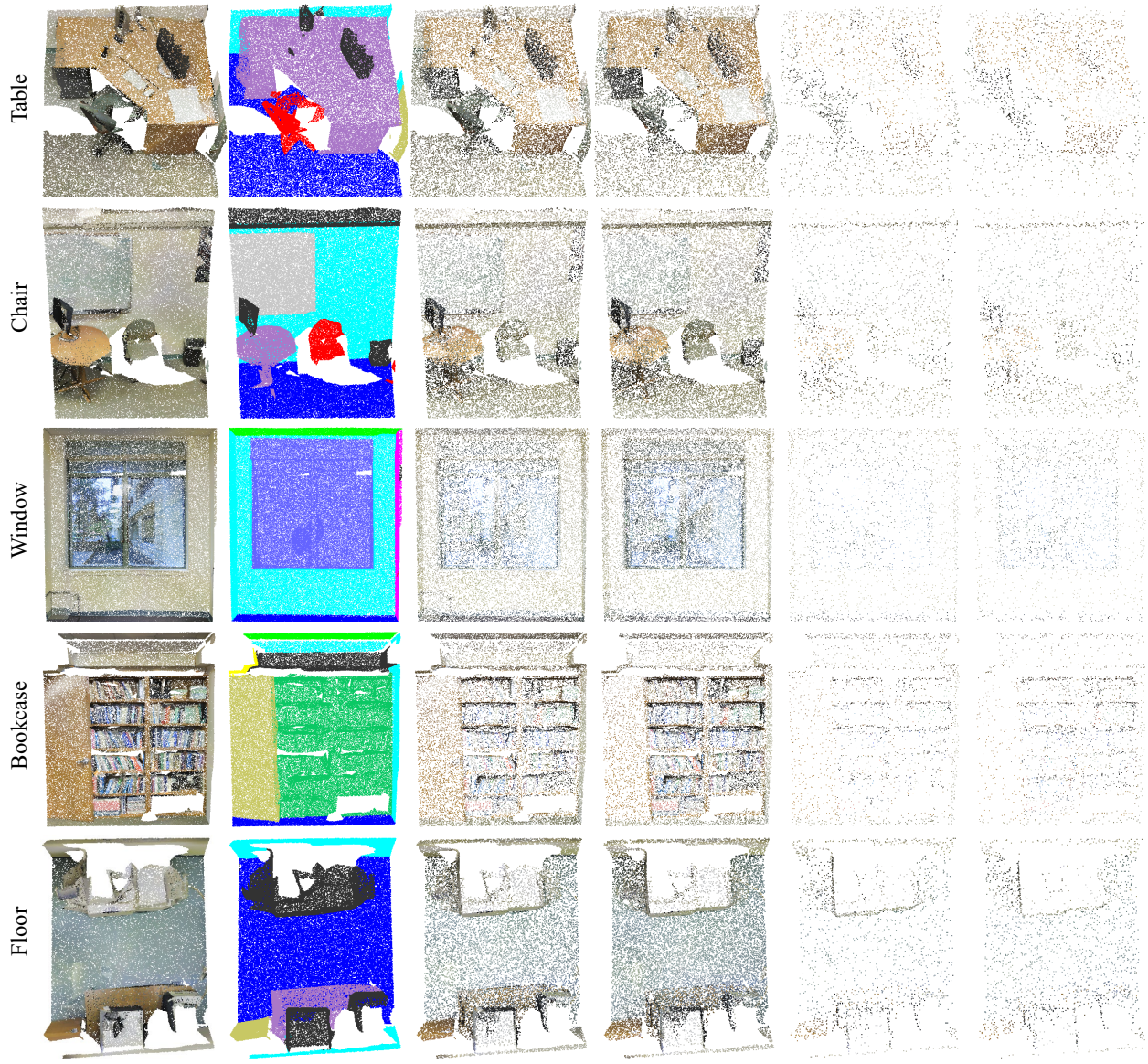

Figure 1. **Visualization of various scenes from the S3DIS dataset [1], with the target class for the 1-way few-shot task labeled at the leftmost of each scene.** Each scene includes six types of point clouds, arranged *from left to right:* (1) The original point cloud; (2) Ground truth of all categories; (3) Our corrected input with 20,480 points in a uniform distribution; (4) Input with 20,480 points in a biased distribution; (5) Input with 2,048 points in a uniform distribution; (6) Input with 2,048 points in a biased distribution, as adopted by previous works.

with BPC (w/ BPC, 5th column) and without BPC (w/o BPC, 6th column).

We have the following observations from the visual comparisons: (1) Our method yields visually better results than the previous best-performing method, highlighting the superiority of our proposed correlation optimization paradigm in enhancing the generalization ability for few-shot tasks. (2) The lightweight BPC module, equipped with non-parametric base prototypes, effectively mitigates the base susceptibility issue inherent in models. This ensures accurate segmentation of novel classes, further validating the efficacy of our approach.

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 *Proceedings of the IEEE conference on computer vision and pattern recog-*

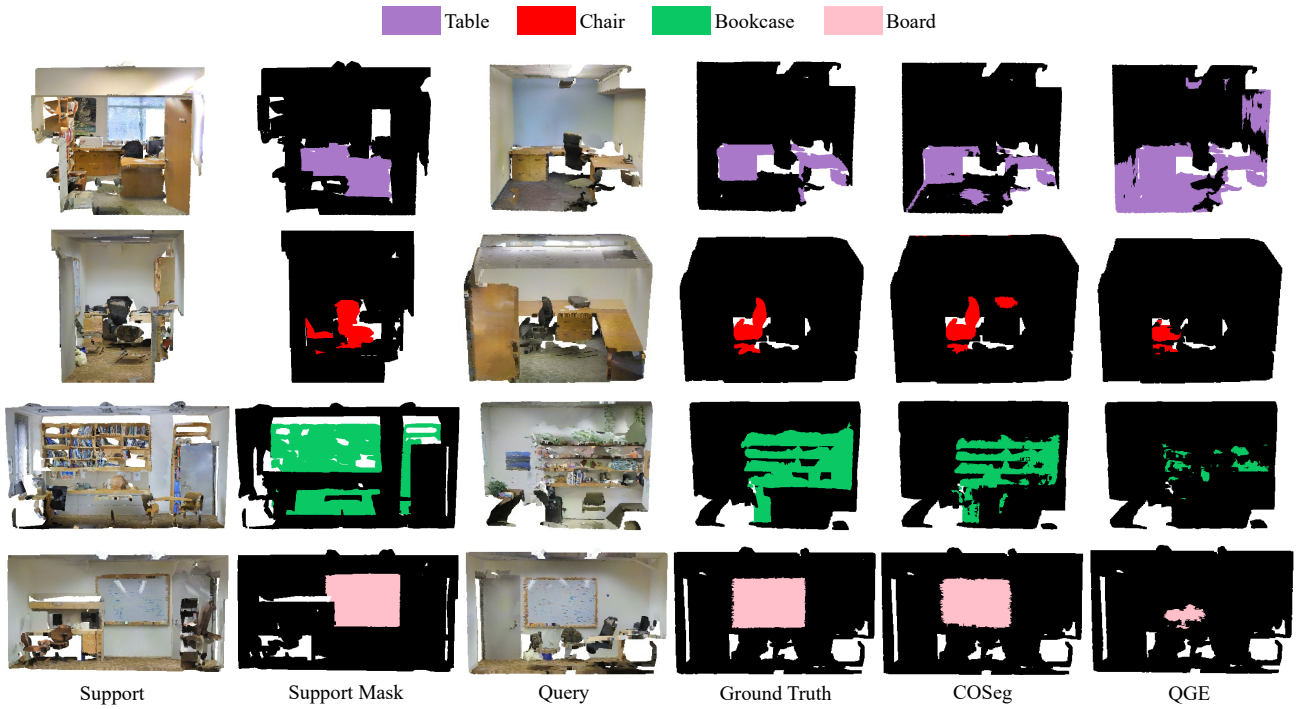

Figure 2. Qualitative comparisons between our proposed model COSeg and QGE [6]. Each row, from top to bottom, represents the 1-way 1-shot task with the target category as table (purple), chair (red), bookcase (green) and board (pink), respectively.

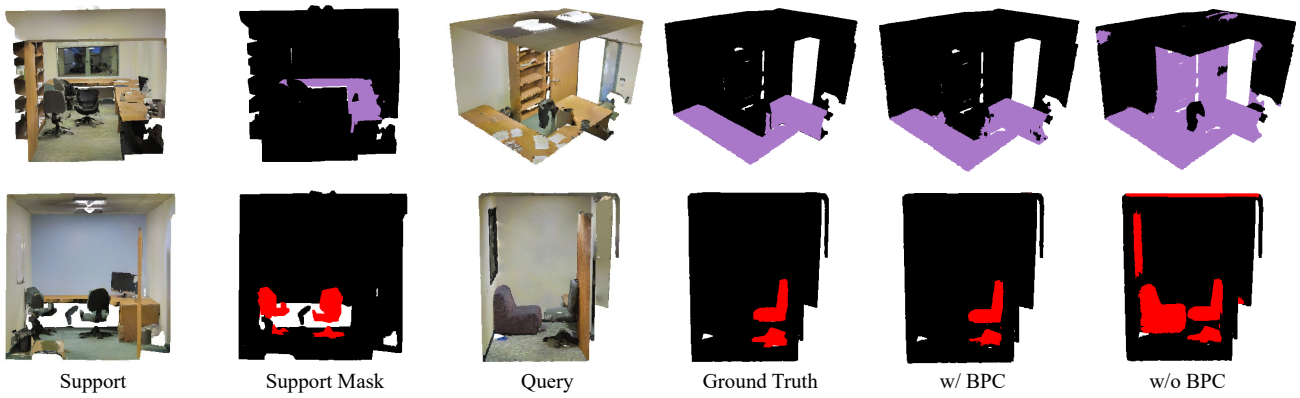

Figure 3. Qualitative comparisons between our models with BPC (w/ BPC) and without BPC (w/o BPC). Each row has the target class under the 1-way 1-shot task as table (purple) and chair (red), respectively, arranged from top to bottom.

nition, pages 1534–1543, 2016. 1, 2

[2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5828–5839, 2017. 1

[3] Shuting He, Xudong Jiang, Wei Jiang, and Henghui Ding. Prototype adaption and projection for few-and zero-shot 3d point cloud semantic segmentation. *IEEE Transactions on Image Processing*, 2023. 1

[4] Xin Lai, Jianhui Liu, Li Jiang, Liwei Wang, Hengshuang Zhao, Shu Liu, Xiaojuan Qi, and Jiaya Jia. Stratified transformer for 3d point cloud segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8500–8509, 2022. 1

[5] Yongqiang Mao, Zonghao Guo, LU Xiaonan, Zhiqiang Yuan, and Haowen Guo. Bidirectional feature globalization for few-shot semantic segmentation of 3d point cloud scenes. In *2022 International Conference on 3D Vision (3DV)*, pages 505–514. IEEE, 2022. 1

[6] Zhenhua Ning, Zhuotao Tian, Guangming Lu, and Wenjie Pei. Boosting few-shot 3d point cloud segmentation via query-guided enhancement. *arXiv preprint arXiv:2308.03177*, 2023. 3

[7] Jiahui Wang, Haiyue Zhu, Haoren Guo, Abdullah Al Mamun, Cheng Xiang, and Tong Heng Lee. Few-shot point cloud semantic segmentation via contrastive self-supervision and multi-resolution attention. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2811–2817. IEEE, 2023.

[8] Canyu Zhang, Zhenyao Wu, Xinyi Wu, Ziyu Zhao, and Song Wang. Few-shot 3d point cloud semantic segmentation via stratified class-specific attention based transformer network. In *AAAI*, 2023.

[9] Na Zhao, Tat-Seng Chua, and Gim Hee Lee. Few-shot 3d point cloud semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8873–8882, 2021.

[10] Guanyu Zhu, Yong Zhou, Rui Yao, and Hancheng Zhu. Cross-class bias rectification for point cloud few-shot segmentation. *IEEE Transactions on Multimedia*, 2023. 1