

# Supplementary Materials

## Multi-fineness Boundaries and the Shifted Ensemble-aware Encoding for Point Cloud Semantic Segmentation

Anonymous Authors

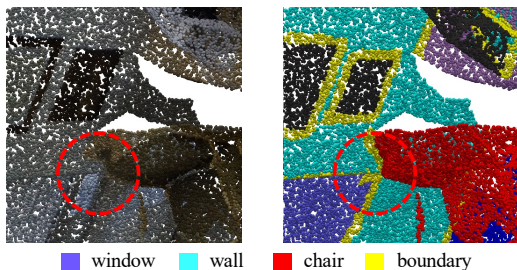
### A APPENDIX

In the appendix, we provide supplementary materials to complement the main manuscript, including more detailed descriptions of training strategies, experiments, visualizations, and theoretical analyses.

### B TRAINING STRATEGIES

For fairness, we follow the training strategies of our baseline, PointNext [11]. For semantic segmentation on S3DIS [1], we train for 100 epochs, and the training set is repeated 30 times. The number of input points is set to 24k, where the criterion is CrossEntropy. The initial learning rate is 0.01 with a weight decay of 0.0004. For ScanNet [2], the number of input points is set to 64k, and the criterion is similar to S3DIS. The number of training epochs is set to 100, and the training set is repeated six times. The initial learning rate is set to 0.01, and a multi-step learning rate decay strategy with a drop rate of 0.1 is used at 60 to 90 epochs. For ShapeNetPart [16], the number of input points is set to 2,048. MBSE is trained for 400 epochs and takes FocalLoss as a criterion. Similar to ScanNet, the strategy of multi-step learning rate decay is used with a 0.1 decay rate but at 210 to 270 epochs.

### C ENCLAVE PHENOMENON



**Figure 1: The illustration of the enclave phenomenon. The yellow points indicate boundary points queried by KNN.**

Here, we provide a supplementary explanation of the enclave phenomenon mentioned in subsection 3.1.1 of the manuscript and analyze the reasons for the poor performance of query boundaries with KNN. In sparse regions, KNN queries may lead to wrong boundary queries, resulting in enclave phenomena, where distant points may be grouped into the same boundary neighbor due to the absence of spatial distance constraints. As depicted by the red dashed circle in Figure 1, some wall points, despite being closer to the window and chair, respectively, are divided into the same boundary neighbor. Their features are imposed similarity constraints in the neighbor. Although they have the same segmentation category,

these features should not exhibit similar semantics due to the different surrounding environments. It is the main reason for the decreased accuracy of the multi-fineness boundary feature constraints with KNN queries in MBC, as shown in Table 4 in the main manuscript.

### D 3D OBJECT CLASSIFICATION

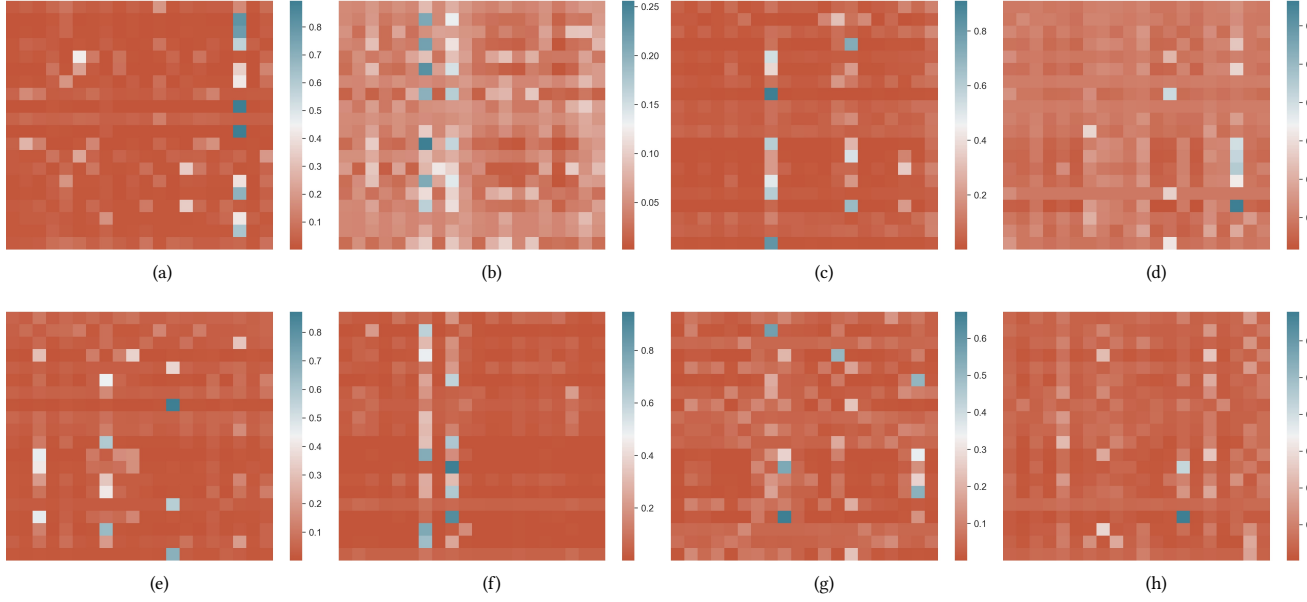
We further demonstrate the effectiveness of our method on ModelNet40 [13]. ModelNet40 is a commonly used point cloud dataset for 3D object classification. It contains 40 categories, 9,843 models in the training set, and 2,468 models in the validation set. Because there is no boundary for object classification, we apply SEP on PointNext [11]. Following the baseline, MBSE is trained with an initial learning rate of 0.001 and weight decay of 0.05 for 600 epochs. The number of input points is set to 1,024. The criterion used for classification is SmoothCrossEntropy. The sampled sub-point clouds are equally divided into (2, 2, 2, 2) ensembles, respectively. The results are shown in Table 1. The model can better perceive objects' spatial and contextual structure in large receptive fields through long-range correlation capture. SEP improves OA and mAcc by 0.3% and 0.6% over the baseline, respectively, and achieves state-of-the-art performance.

**Table 1: The comparison results on ModelNet40 for 3D object classification.**

Method	OA	mAcc
PointNet [9]	89.2	86.2
PointCNN [7]	92.2	88.1
DGCNN [12]	92.9	90.2
DeepGCN [6]	93.6	90.9
ASSANet [10]	92.9	-
SimpleView [4]	93.0	90.5
Point Cloud Transformer [5]	93.2	-
GDANet [15]	93.8	-
CurveNet [14]	93.8	-
PointMLP [8]	94.1	-
PointVector [3]	93.7	91.5
PointNext	94.0	91.1
<b>+SEP</b>	<b>94.3</b>	<b>91.7</b>

### E LONG-RANGE CORRELATIONS

The SEP significantly improves the performance of baselines with minimal computational cost. We visualize the long-range correlations within SEP to explore it further, as shown in Figure 2. SEP



**Figure 2: The visualization of long-range correlations in our SEP. We take a randomly selected scene from S3DIS Area 5 as the input of the trained MBSE model. Due to typographic space limitations, the figure displays eight long-range correlation maps of its sub-point cloud downsampled three times. Without intermediate processing that may cause feature blur or loss, such as feature aggregation, the long-range correlations captured by SEP contain direct point-to-point relationship information, which is beneficial for point cloud semantic segmentation.**

can independently capture direct point-to-point long-range correlations in the point cloud from various aspects, which is difficult to achieve with traditional local neighborhood feature aggregation. The values in the figure’s  $i$ -th row and  $j$ -th column represent the degree of long-range correlation between the  $i$ -th and  $j$ -th points in the ensemble. In the complex space of point clouds, long-range correlations depend not only on the segmentation categories and the positions of points but also on their surrounding environments. The semantic associations are bidirectional and inconsistent, so the visualizations of the long-range correlation do not exhibit diagonal symmetry. The larger the value, the stronger the one-way correlation between the two points. The direct information interactions between distant points assist the model in better comprehending the overall structure and semantics of the point cloud, leading to more accurate feature representations, which is one of the main reasons for the effectiveness of SEP.

## F QUALITATIVE RESULTS

In order to show the effectiveness and the superiority of our method more intuitively, we visualize the multi-fineness boundary queries and semantic segmentation results, respectively.

### F.1 Multi-fineness Boundary Query

We conduct boundary queries at three different fineness degrees. Figure 3 illustrates that boundary queries at different fineness degrees yield distinct sets of boundary points. Querying with a higher fineness degree can yield more precise boundaries but may result in the omission of boundary points. Conversely, querying at a

lower fineness degree can produce more comprehensive boundaries but may misclassify some non-boundary points as boundary points. Multi-fineness boundary queries strike a good balance between boundary accuracy and its integrity, which can provide higher-quality boundary information support for boundary feature constraints, thereby achieving better boundary segmentation performance for point clouds.

### F.2 Semantic Segmentation Results

As depicted in Figure 4, MBSE has a more precise segmentation of objects with high boundary proportions, such as photo frames (belonging to the clutter), beams, and columns, due to its excellent multi-fineness boundary feature constraints and the capability to capture long-range correlations. MBSE exhibits outstanding semantic segmentation performance on both boundaries and non-boundaries.

## REFERENCES

- [1] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 2016. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1534–1543.
- [2] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5828–5839.
- [3] Xin Deng, WenYu Zhang, Qing Ding, and XinMing Zhang. 2023. Pointvector: a vector representation in point cloud analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9455–9465.
- [4] Ankit Goyal, Hei Law, Bowei Liu, Alejandro Newell, and Jia Deng. 2021. Re-visiting point cloud shape classification with a simple and effective baseline. In



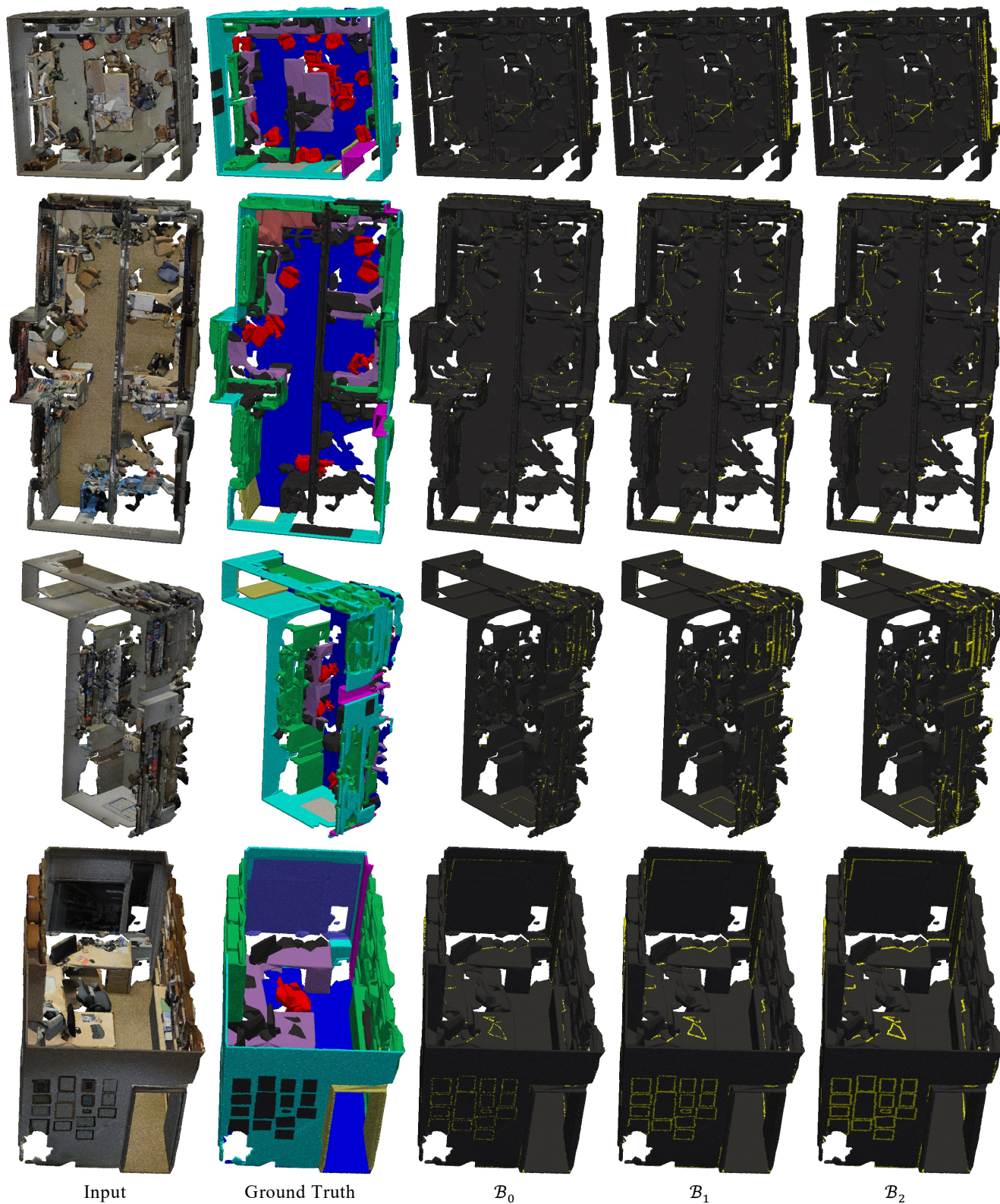


Figure 3: The visualization of boundary queries at three different fineness degrees (i.e.  $M = 3$ ).



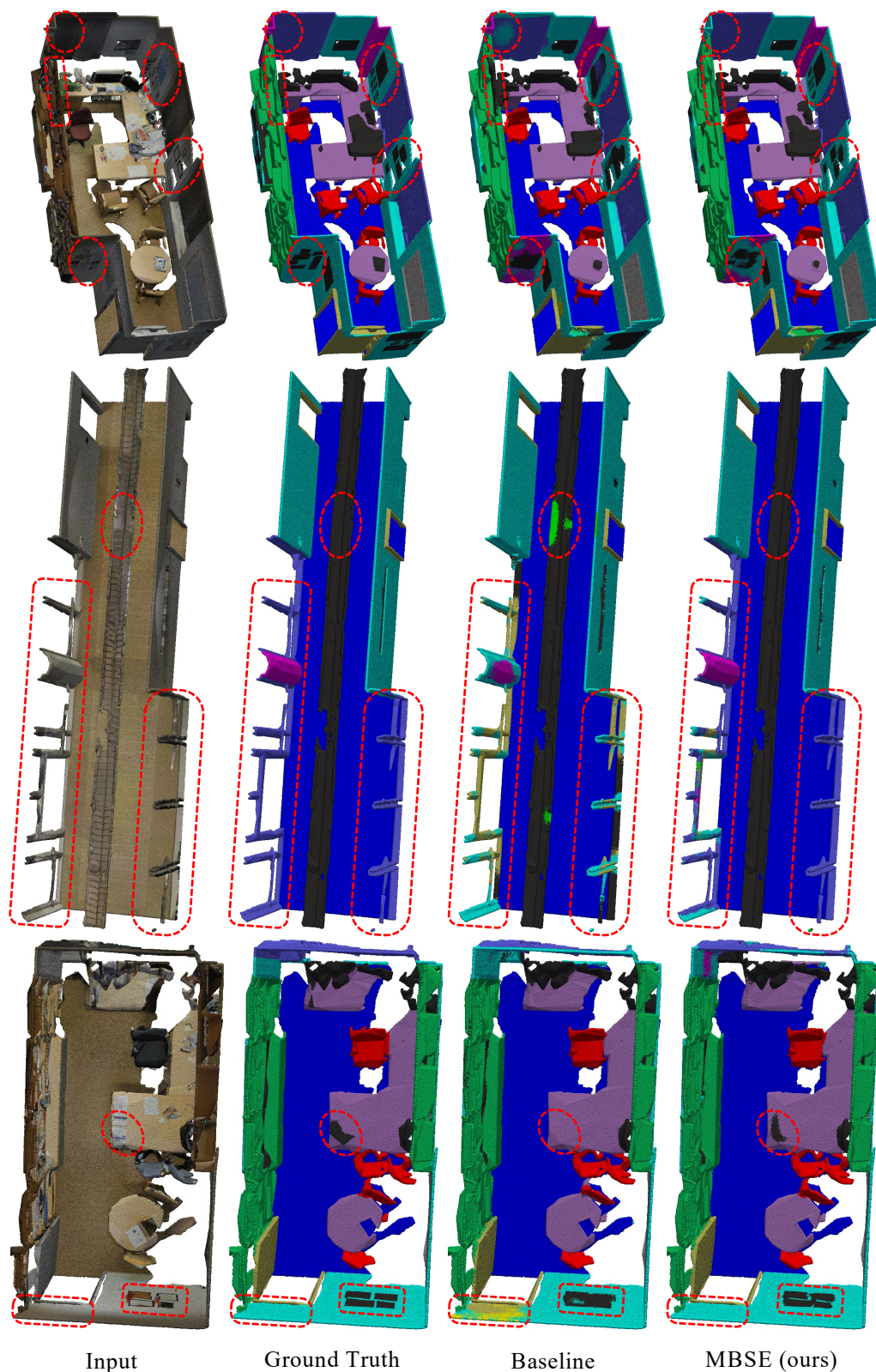


Figure 4: Comparison of semantic segmentation results between MBSE and baseline. It is best to zoom in for details.



- International Conference on Machine Learning. PMLR, 3809–3820.
- [5] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R Martin, and Shi-Min Hu. 2021. Pct: Point cloud transformer. *Computational Visual Media* 7 (2021), 187–199.
- [6] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. 2019. Deep-gcns: Can gcns go as deep as cnns?. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9267–9276.
- [7] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. 2018. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems* 31 (2018).
- [8] Xu Ma, Can Qin, Haoxuan You, Haoxi Ran, and Yun Fu. 2022. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. *arXiv preprint arXiv:2202.07123* (2022).
- [9] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660.
- [10] Guocheng Qian, Hasan Hammoud, Guohao Li, Ali Thabet, and Bernard Ghanem. 2021. Assanet: An anisotropic separable set abstraction for efficient point cloud representation learning. *Advances in Neural Information Processing Systems* 34 (2021), 28119–28130.
- [11] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. 2022. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in Neural Information Processing Systems* 35 (2022), 23192–23204.
- [12] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. 2019. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)* 38, 5 (2019), 1–12.
- [13] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1912–1920.
- [14] Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. 2021. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 915–924.
- [15] Mutian Xu, Junhao Zhang, Zhipeng Zhou, Mingye Xu, Xiaojuan Qi, and Yu Qiao. 2021. Learning geometry-disentangled representation for complementary understanding of 3d object point cloud. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 3056–3064.
- [16] Li Yi, Vladimir G Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. 2016. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (ToG)* 35, 6 (2016), 1–12.

523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580