BFANet: Revisiting 3D Semantic Segmentation with Boundary Feature Analysis

Supplementary Material

7. Differences among Four Proposed Metrics

'Merge error' refers to our observation that some small objects are directly merged into *one* larger object. 'Displacement' refers to the phenomenon where the boundaries of an object shift relative to *one or more* adjacent objects due to unclear semantic features at these boundaries. For 'Displacement', most of mask points must be correctly predicted to calculate boundary blurriness. That's why we introduced the θ parameter. The essential difference between 'false response' and 'region classification error' is that the former involves a complete GT mask with many small erroneous regions being predicted, whereas the latter involves the complete GT mask being classified as another object of similar shape.

7.1. Real-Time Boundary Pseudo-Label Calculation

We provide the pseudo-code of our PBPLC as below:

Algorithm 1 Proposed Parallel Boundary Pseudo-Label Calculation (PBPLC)

Input: Point cloud coordinates $\mathcal{C} \in \mathbb{R}^{N\times 3}$
Point cloud semantic labels $\mathcal{S} \in \mathbb{R}^{N\times 1}$
Parameter: Radius threshold r
Number of points in the Point cloud: N
Euclidean distance between two points: ED$(\cdot)$
Output: Binary boundary pseudo-label $\hat{E} \in \{0,1\}^N$
Procedure:

Initialize $\mathcal{B}$ as an all-zero queue
Put the $\mathcal{C}_i$, $\mathcal{S}_i$ on i_{th} CUDA parallel threads.
Operation on N threads simultaneously:
for $j < N$ **do**
  if $\mathcal{S}_i == \mathcal{S}_j$ or $r <$ ED$(\mathcal{C}_i, \mathcal{C}_j)$ **then**
    continue
  else
    $\hat{E}_i = 1$
  end if
end for
return $\hat{E}$

8. Additional Comparison to SOTAs

In the manuscript, we provide results of our method on the general datasets ScanNet200 and ScanNetv2. Here, we report the results on another common dataset, S3DIS [1], as well. The S3DIS dataset comprises 271 scenes across 6 areas, with 13 semantic classes labeled in these scenes. We report the results specifically for area 5 for semantic segmentation evaluation, while the other areas are used for training. All results are from the publicly available data of the method, and "-" indicates that the method does not provide the corresponding data.

To the best of our knowledge, our proposed method is the first octree-based method validated on the S3DIS dataset for 3D semantic segmentation. As reported in Tab. 8, Our method achieves state-of-the-art performance on this dataset.

Methods	Venue	Input	mIoU↑
SparseUNet [12]	CVPR'18	voxel	-
MinkUNet [6]	CVPR'19	voxel	65.4
LargeKernel3D [4]	CVPR'23	voxel	-
OA-CNN [39]	CVPR'24	voxel	71.1
PTv1 [69]	ICCV'21	point	70.4
Stratified-Tr [24]	CVPR'22	point	72.0
FastPoint-Tr [37]	CVPR'22	point	70.4
PTv2 [56]	NeurIPS'22	point	71.6
PTv3 [57]	CVPR'24	point	73.4
SPG [15]	ECCV'24	point	73.3
O-CNN [49]	TOG'17	octree	-
OctFormer [48]	TOG'23	octree	-
Ours	-	octree	**73.7**

Table 8. Evaluation on S3DIS Area 5 with Traditional Metrics.

8.1. Additional Comparison with Proposed Metrics

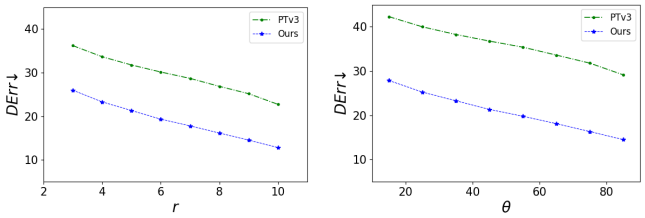


Figure 6. Additional Comparison to PTv3 with the Proposed Metric

In Section 5.2.2, We provide comparison results of Err with respect to changes in boundary distance r. Since DErr$_\theta$ is also related to the mask threshold θ, we provide comparison results of DErr$_\theta$ with respect to changes in θ in this section. As depicted in Fig. 6, our method clearly outperforms state-of-the-art approaches, demonstrating its superior ability to mitigate displacement error.

9. Implementation Detail.

9.1. Test Time Augmentation (TTA)

In response to the challenges of model robustness and generalization, most SOTA methods [36, 48, 57] incorporate test-time augmentation approaches during inference to enhance semantic segmentation. In our work, we exploit three existing test-time augmentation approaches: rotation, superpoint pooling, and multiple checkpoint ensemble. For rotation, we input the initial view with a yaw angle offset of 120 degrees for each pass and perform score maximization across the three results. In addition, we directly apply the superpoint pooling method borrowed from [36, 59, 70] to perform mean pooling operations on points within the same superpoint. Finally, we apply the multiple checkpoint ensemble method from our baseline model, OctFormer [48], to further refine the semantic segmentation results.

Due to rotation, superpoint pooling has become a consensus and is widely adopted by most state-of-the-art methods [25, 36, 48, 57, 59, 70]. Therefore, in our manuscript, we follow the approach of OctFormer [48] and focus on ablation studies of checkpoint ensemble. Since the S3DIS dataset does not include superpoints, we omit this TTA operation for its validation.

9.2. Training Data Augmentations

Our data augmentation strategy during training closely aligns with that of PTv3. Detailed comparisons are provided in Tab. 9 and Tab. 10. For clearer visual reference, differences are highlighted in green. Specifically, our method entirely dispenses with the need for a voxel network, hence we remove the grid sampling operation. Furthermore, given that our method is octree-based, we reduce the number of sphere-crop points to enhance training speed and reduce memory consumption.

Augmentations	Parameters
random dropout	dropout ratio: 0.2, p: 0.2
random rotate	axis: z, angle: [-1, 1], p: 0.5
	axis: x, angle: [-1 / 64, 1 / 64], p: 0.5
	axis: y, angle: [-1 / 64, 1 / 64], p: 0.5
random scale	scale: [0.9, 1.1]
random flip	p: 0.5
random jitter	sigma: 0.005, clip: 0.02
elastic distort	params: [[0.2, 0.4], [0.8, 1.6]]
auto contrast	p: 0.2
color jitter	std: 0.05; p: 0.95
sphere crop	ratio: 1.0, max points: 102400
normalize color	p: 1.0
scene mixup	num: 2

Table 9. Training Data Augmentations in Proposed Method.

Augmentations	Parameters
random dropout	dropout ratio: 0.2, p: 0.2
random rotate	axis: z, angle: [-1, 1], p: 0.5
	axis: x, angle: [-1 / 64, 1 / 64], p: 0.5
	axis: y, angle: [-1 / 64, 1 / 64], p: 0.5
random scale	scale: [0.9, 1.1]
random flip	p: 0.5
random jitter	sigma: 0.005, clip: 0.02
elastic distort	params: [[0.2, 0.4], [0.8, 1.6]]
auto contrast	p: 0.2
color jitter	std: 0.05; p: 0.95
grid sampling	grid size:0.02
sphere crop	ratio: 1.0, max points: 128000
normalize color	p: 1.0
scene mixup	num: 2

Table 10. Training Data Augmentations in PTv3 [57].

10. Additional Visualizations

In this section, we provide the comparative segmentation results in Fig. 7, the snapshot of the ScanNet200 online leaderboard in Fig. 9, and the visual representation of octree construction in Fig. 8.

Comparative Segmentation Results. We provide additional visual comparisons, as illustrated in Fig. 7. The comparative results from the first and second rows demonstrate that our method effectively mitigates Displacement errors. Furthermore, the comparative results in the first row demonstrate that our method effectively overcomes False Response errors.

Octree Construction. As shown in Fig. 8, we construct an octree with a depth of 9 from the input point cloud and visualize the octree structure at each depth along with the corresponding point cloud. Adjacent green cubes across different depths illustrate a parent-child relationship, with the cube at the lower depth acting as the parent node to the corresponding cube at the higher depth. In addition, the octree structure captures both global information (at lower depths) and local details (at higher depths), which is different from the voxelization methods.

Snapshot of the ScanNet200 Online Leaderboard. Considering that the ScanNet200 online leaderboard may update in real-time, we have included a screenshot from October 22, 2024. As illustrated in Fig. 9, our BFANet ranks 2nd on the ScanNet200 official benchmark challenge, presenting the highest mIoU so far if excluding the 1st place winner that, however, involves large-scale training with auxiliary data.

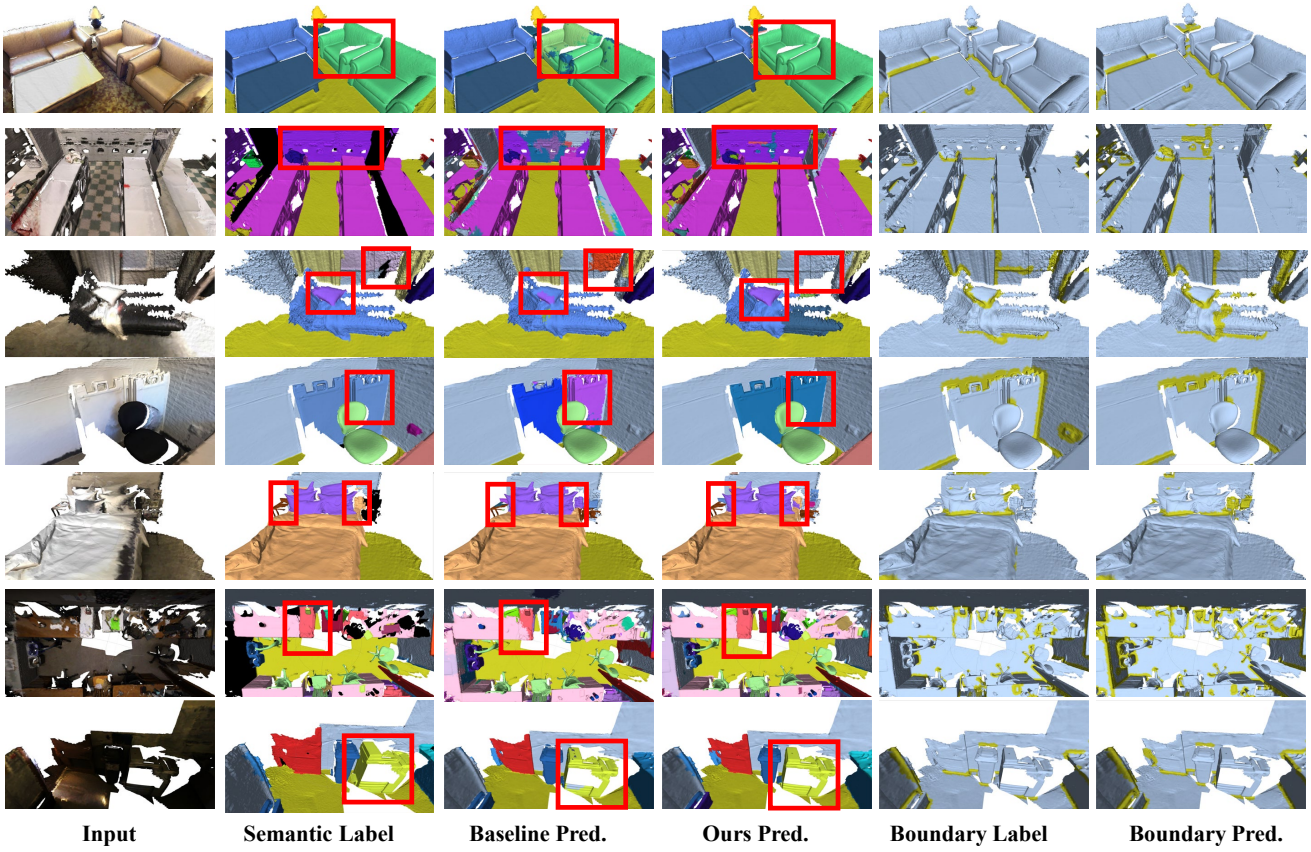


Figure 7. Qualitative Comparison. Pred. stands for Prediction. The red rectangular boxes indicate areas of particular interest.

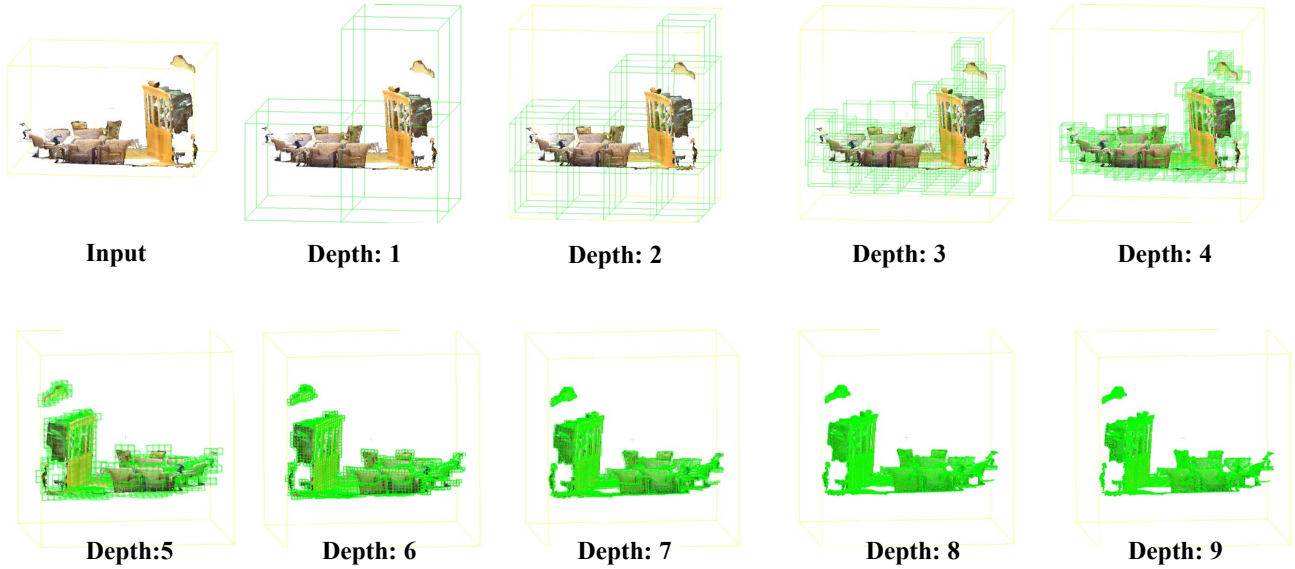
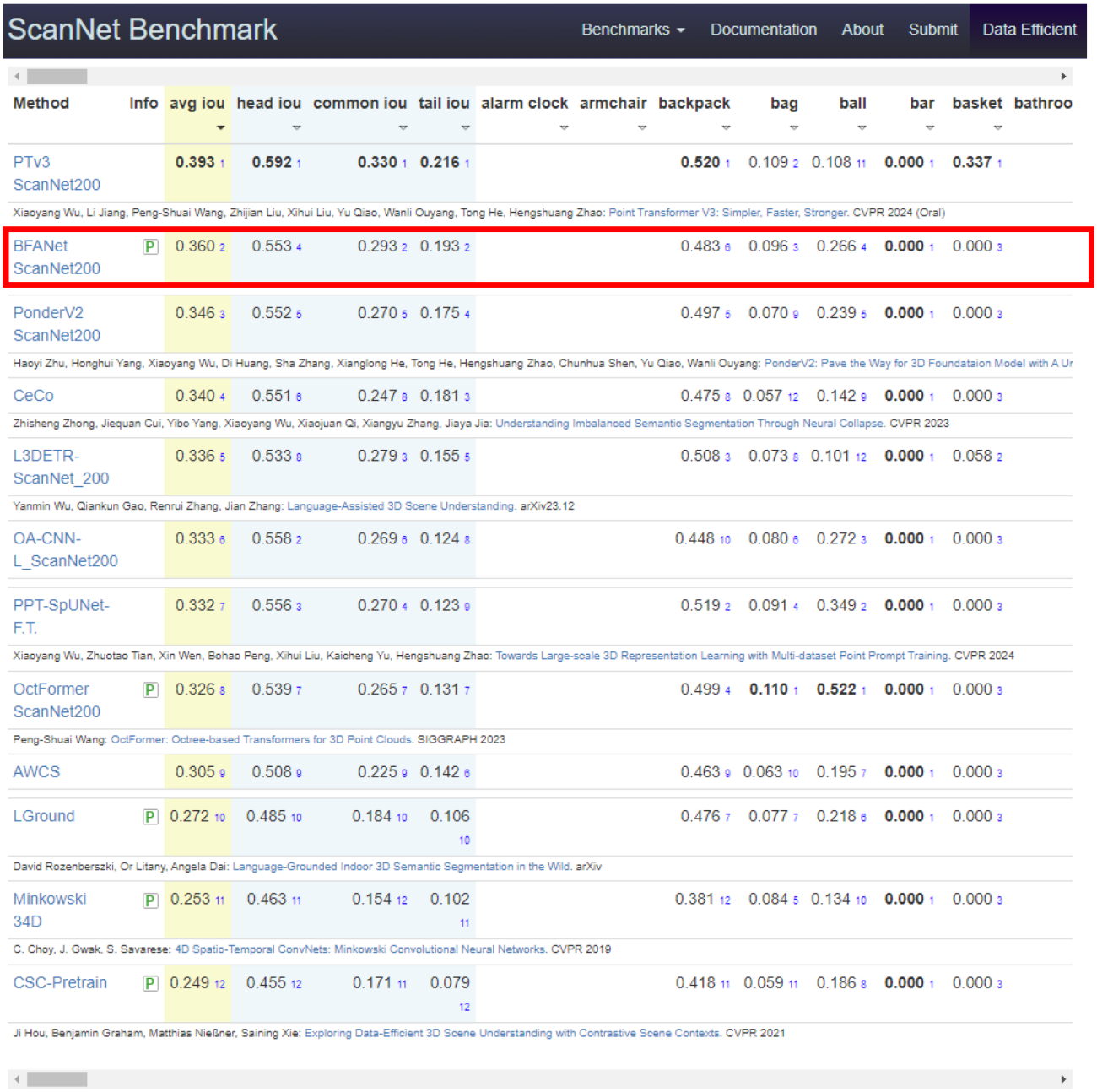


Figure 8. Visualization of Octree Construction



Method	Info	avg iou	head iou	common iou	tail iou	alarm clock	armchair	backpack	bag	ball	bar	basket	bathroo
PTv3 ScanNet200		**0.393** 1	**0.592** 1	**0.330** 1	**0.216** 1			**0.520** 1	0.109 2	0.108 11	**0.000** 1	**0.337** 1	
Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao: Point Transformer V3: Simpler, Faster, Stronger. CVPR 2024 (Oral)													
BFANet ScanNet200	P	0.360 2	0.553 4	0.293 2	0.193 2			0.483 6	0.096 3	0.266 4	**0.000** 1	0.000 3	
PonderV2 ScanNet200		0.346 3	0.552 5	0.270 5	0.175 4			0.497 5	0.070 9	0.239 5	**0.000** 1	0.000 3	
Haoyi Zhu, Honghui Yang, Xiaoyang Wu, Di Huang, Sha Zhang, Xianglong He, Tong He, Hengshuang Zhao, Chunhua Shen, Yu Qiao, Wanli Ouyang: PonderV2: Pave the Way for 3D Foundataion Model with A Ur													
CeCo		0.340 4	0.551 6	0.247 8	0.181 3			0.475 8	0.057 12	0.142 9	**0.000** 1	0.000 3	
Zhisheng Zhong, Jiequan Cui, Yibo Yang, Xiaoyang Wu, Xiaojuan Qi, Xiangyu Zhang, Jiaya Jia: Understanding Imbalanced Semantic Segmentation Through Neural Collapse. CVPR 2023													
L3DETR-ScanNet_200		0.336 5	0.533 8	0.279 3	0.155 5			0.508 3	0.073 8	0.101 12	**0.000** 1	0.058 2	
Yanmin Wu, Qiankun Gao, Renrui Zhang, Jian Zhang: Language-Assisted 3D Scene Understanding. arXiv23.12													
OA-CNN-L_ScanNet200		0.333 6	0.558 2	0.269 6	0.124 8			0.448 10	0.080 6	0.272 3	**0.000** 1	0.000 3	
PPT-SpUNet-F.T.		0.332 7	0.556 3	0.270 4	0.123 9			0.519 2	0.091 4	0.349 2	**0.000** 1	0.000 3	
Xiaoyang Wu, Zhuotao Tian, Xin Wen, Bohao Peng, Xihui Liu, Kaicheng Yu, Hengshuang Zhao: Towards Large-scale 3D Representation Learning with Multi-dataset Point Prompt Training. CVPR 2024													
OctFormer ScanNet200	P	0.326 8	0.539 7	0.265 7	0.131 7			0.499 4	**0.110** 1	**0.522** 1	**0.000** 1	0.000 3	
Peng-Shuai Wang: OctFormer: Octree-based Transformers for 3D Point Clouds. SIGGRAPH 2023													
AWCS		0.305 9	0.508 9	0.225 9	0.142 6			0.463 9	0.063 10	0.195 7	**0.000** 1	0.000 3	
LGround	P	0.272 10	0.485 10	0.184 10	0.106 10			0.476 7	0.077 7	0.218 6	**0.000** 1	0.000 3	
David Rozenberszki, Or Litany, Angela Dai: Language-Grounded Indoor 3D Semantic Segmentation in the Wild. arXiv													
Minkowski 34D	P	0.253 11	0.463 11	0.154 12	0.102 11			0.381 12	0.084 5	0.134 10	**0.000** 1	0.000 3	
C. Choy, J. Gwak, S. Savarese: 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. CVPR 2019													
CSC-Pretrain	P	0.249 12	0.455 12	0.171 11	0.079 12			0.418 11	0.059 11	0.186 8	**0.000** 1	0.000 3	
Ji Hou, Benjamin Graham, Matthias Nießner, Saining Xie: Exploring Data-Efficient 3D Scene Understanding with Contrastive Scene Contexts. CVPR 2021													

Figure 9. ScanNet200 Benchmark Challenge. Recorded on October 22, 2024.

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*, pages 1534–1543, 2016. 5, 1

[2] Kaixin Cai, Pengzhen Ren, Yi Zhu, Hang Xu, Jianzhuang Liu, Changlin Li, Guangrun Wang, and Xiaodan Liang. Mixreorg: Cross-modal mixed patch reorganization is a good mask learner for open-world semantic segmentation. In *ICCV*, pages 1196–1205, 2023. 1

[3] Linwei Chen, Lin Gu, and Ying Fu. When semantic segmentation meets frequency aliasing. *ICLR*, 2024. 2, 3

[4] Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, and Jiaya Jia. Largekernel3d: Scaling up kernels in 3d sparse cnns. In *CVPR*, pages 13488–13498, 2023. 6, 1

[5] Bowen Cheng, Ross Girshick, Piotr Dollár, Alexander C Berg, and Alexander Kirillov. Boundary iou: Improving object-centric image segmentation evaluation. In *CVPR*,

pages 15334–15342, 2021. 3

[6] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *CVPR*, pages 3075–3084, 2019. 1, 2, 6, 7

[7] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, pages 5828–5839, 2017. 5

[8] Angela Dai, Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Christian Theobalt. Bundlefusion: Real-time globally gonsistent 3d reconstruction using on-the-fly surface reintegration. *ACM TOG*, 36(4):1, 2017. 1

[9] Xueqing Deng, Peng Wang, Xiaochen Lian, and Shawn Newsam. Nightlab: A dual-level architecture with hardness detection for segmentation at night. In *CVPR*, pages 16938–16948, 2022. 2

[10] Lunhao Duan, Shanshan Zhao, Nan Xue, Mingming Gong, Gui-Song Xia, and Dacheng Tao. Condaformer: Disassembled transformer with local structure enhancement for 3d point cloud understanding. In *NeurIPS*, pages 23886–23901, 2023. 2

[11] Jingyu Gong, Jiachen Xu, Xin Tan, Jie Zhou, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Boundary-aware geometric encoding for semantic segmentation of point clouds. In *AAAI*, pages 1424–1432, 2021. 3

[12] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *CVPR*, pages 9224–9232, 2018. 6, 1

[13] Zhangxuan Gu, Li Niu, Haohua Zhao, and Liqing Zhang. Hard pixel mining for depth privileged semantic segmentation. *IEEE TMM*, 23:3738–3751, 2020. 2

[14] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE TPAMI*, 43(12):4338–4364, 2020. 1, 2

[15] Jiawei Han, Kaiqi Liu, Wei Li, and Guangzhi Chen. Subspace prototype guidance for mitigating class imbalance in point cloud semantic segmentation. In *ECCV*, 2024. 6, 1

[16] Lei Han, Tian Zheng, Yinheng Zhu, Lan Xu, and Lu Fang. Live semantic 3d perception for immersive augmented reality. *IEEE TVCG*, 26(5):2012–2022, 2020. 1

[17] Wenkai Han, Chenglu Wen, Cheng Wang, Xin Li, and Qing Li. Point2node: Correlation learning of dynamic-node for point cloud feature modeling. In *AAAI*, pages 10925–10932, 2020. 5

[18] Yining Hong, Zishuo Zheng, Peihao Chen, Yian Wang, Junyan Li, and Chuang Gan. Multiply: A multisensory object-centric embodied large language model in 3d world. In *CVPR*, pages 26406–26416, 2024. 1

[19] Ji Hou, Benjamin Graham, Matthias Nießner, and Saining Xie. Exploring data-efficient 3d scene understanding with contrastive scene contexts. In *CVPR*, pages 15587–15597, 2021. 6

[20] Wenbo Hu, Hengshuang Zhao, Li Jiang, Jiaya Jia, and Tien-Tsin Wong. Bidirectional projection network for cross dimension scene understanding. In *CVPR*, pages 14373–14382, 2021. 2

[21] Zeyu Hu, Mingmin Zhen, Xuyang Bai, Hongbo Fu, and Chiew-lan Tai. Jsenet: Joint semantic segmentation and edge detection network for 3d point clouds. In *ECCV*, pages 222–239. Springer, 2020. 3

[22] Shihua Huang, Zhichao Lu, Ran Cheng, and Cheng He. Fapn: Feature-aligned pyramid network for dense image prediction. In *ICCV*, pages 864–873, 2021. 3

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 5

[24] 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 *CVPR*, pages 8500–8509, 2022. 6, 1

[25] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *CVPR*, pages 4558–4567, 2018. 4, 5, 2

[26] Hong Joo Lee, Jung Uk Kim, Sangmin Lee, Hak Gu Kim, and Yong Man Ro. Structure boundary preserving segmentation for medical image with ambiguous boundary. In *CVPR*, pages 4817–4826, 2020. 3

[27] Xiangtai Li, Ansheng You, Zhen Zhu, Houlong Zhao, Maoke Yang, Kuiyuan Yang, Shaohua Tan, and Yunhai Tong. Semantic flow for fast and accurate scene parsing. In *ECCV*, pages 775–793, 2020. 3

[28] Zhijie Lin, Zhaoshui He, Xu Wang, Bing Zhang, Chang Liu, Wenqing Su, Ji Tan, and Shengli Xie. Dbganet: dual-branch geometric attention network for accurate 3d tooth segmentation. *IEEE TCSVT*, 34(6):4285–4298, 2023. 3

[29] Sun-Ao Liu, Yiheng Zhang, Zhaofan Qiu, Hongtao Xie, Yongdong Zhang, and Ting Yao. Learning orthogonal prototypes for generalized few-shot semantic segmentation. In *CVPR*, pages 11319–11328, 2023. 1

[30] Zhijian Liu, Xinyu Yang, Haotian Tang, Shang Yang, and Song Han. Flatformer: Flattened window attention for efficient point cloud transformer. In *CVPR*, pages 1200–1211, 2023. 2

[31] Carlos Lopez-Molina, Bernard De Baets, and Humberto Bustince. Quantitative error measures for edge detection. *PR*, 46(4):1125–1139, 2013. 3

[32] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. In *ICLR*, 2016. 6

[33] Dmitrii Marin, Zijian He, Peter Vajda, Priyam Chatterjee, Sam Tsai, Fei Yang, and Yuri Boykov. Efficient segmentation: Learning downsampling near semantic boundaries. In *CVPR*, pages 2131–2141, 2019. 3

[34] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *IROS*, pages 922–928, 2015. 2

[35] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3DV*, pages 565–571, 2016. 5

[36] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann. Mix3d: Out-of-context data augmentation for 3d scenes. In *3DV*, pages 116–125, 2021. 4, 5, 6, 2

[37] Chunghyun Park, Yoonwoo Jeong, Minsu Cho, and Jaesik Park. Fast point transformer. In *CVPR*, pages 16949–16958, 2022. 6, 1

[38] Jinyoung Park, Sanghyeok Lee, Sihyeon Kim, Yunyang Xiong, and Hyunwoo J Kim. Self-positioning point-based transformer for point cloud understanding. In *CVPR*, pages 21814–21823, 2023. 2

[39] Bohao Peng, Xiaoyang Wu, Li Jiang, Yukang Chen, Hengshuang Zhao, Zhuotao Tian, and Jiaya Jia. Oa-cnns: Omni-adaptive sparse cnns for 3d semantic segmentation. In *CVPR*, pages 21305–21315, 2024. 6, 1

[40] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, pages 652–660, 2017. 2

[41] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, pages 5099–5108, 2017. 2

[42] Wonseok Roh, Hwanhee Jung, Giljoo Nam, Jinseop Yeom, Hyunje Park, Sang Ho Yoon, and Sangpil Kim. Edge-aware 3d instance segmentation network with intelligent semantic prior. In *CVPR*, pages 20644–20653, 2024. 2, 5, 8

[43] David Rozenberszki, Or Litany, and Angela Dai. Language-grounded indoor 3d semantic segmentation in the wild. In *ECCV*, pages 125–141, 2022. 5, 6

[44] Liyao Tang, Yibing Zhan, Zhe Chen, Baosheng Yu, and Dacheng Tao. Contrastive boundary learning for point cloud segmentation. In *CVPR*, pages 8489–8499, 2022. 3, 8

[45] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *ICCV*, pages 6411–6420, 2019. 2

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 4

[47] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE TPAMI*, 43(10): 3349–3364, 2020. 4

[48] Peng-Shuai Wang. Octformer: Octree-based transformers for 3d point clouds. *ACM TOG*, 42(4):1–11, 2023. 1, 2, 4, 6, 7

[49] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-cnn: Octree-based convolutional neural networks for 3d shape analysis. *ACM TOG*, 36(4):1–11, 2017. 2, 4, 6, 1

[50] Shaohu Wang, Fangbo Qin, Yuchuang Tong, Xiuqin Shang, and Zhengtao Zhang. Probabilistic boundary-guided point cloud primitive segmentation network. *IEEE TIM*, 2023. 3

[51] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM TOG*, 38(5): 1–12, 2019. 2

[52] Ziyi Wang, Yongming Rao, Xumin Yu, Jie Zhou, and Jiwen Lu. Semaffinet: Semantic-affine transformation for point cloud segmentation. In *CVPR*, pages 11819–11829, 2022. 2

[53] Dongyue Wu, Zilin Guo, Aoyan Li, Changqian Yu, Changxin Gao, and Nong Sang. Conditional boundary loss for semantic segmentation. *IEEE TIP*, 32:3717–3731, 2023. 3

[54] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *CVPR*, pages 9621–9630, 2019. 2

[55] Weijia Wu, Yuzhong Zhao, Mike Zheng Shou, Hong Zhou, and Chunhua Shen. Diffumask: Synthesizing images with pixel-level annotations for semantic segmentation using diffusion models. In *ICCV*, pages 1206–1217, 2023. 1

[56] Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, and Hengshuang Zhao. Point transformer v2: Grouped vector attention and partition-based pooling. In *NeurIPS*, pages 33330–33342, 2022. 1, 2, 6, 7

[57] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler, faster, stronger. In *CVPR*, pages 4840–4851, 2024. 1, 2, 4, 5, 6, 7

[58] Xiaoyang Wu, Zhuotao Tian, Xin Wen, Bohao Peng, Xihui Liu, Kaicheng Yu, and Hengshuang Zhao. Towards large-scale 3d representation learning with multi-dataset point prompt training. In *CVPR*, pages 19551–19562, 2024. 6

[59] Yizheng Wu, Min Shi, Shuaiyuan Du, Hao Lu, Zhiguo Cao, and Weicai Zhong. 3d instances as 1d kernels. In *ECCV*, pages 235–252, 2022. 2

[60] Yanmin Wu, Qiankun Gao, Renrui Zhang, and Jian Zhang. Language-assisted 3d scene understanding. *arXiv preprint arXiv:2312.11451*, 2023. 6

[61] Xiaoyang Xiao, Yuqian Zhao, Fan Zhang, Biao Luo, Lingli Yu, Baifan Chen, and Chunhua Yang. Baseg: Boundary aware semantic segmentation for autonomous driving. *NN*, 157:460–470, 2023. 3

[62] Xu Yan, Jiantao Gao, Chaoda Zheng, Chao Zheng, Ruimao Zhang, Shuguang Cui, and Zhen Li. 2dpass: 2d priors assisted semantic segmentation on lidar point clouds. In *ECCV*, pages 677–695, 2022. 2

[63] ChengKun Yang, MinHung Chen, YungYu Chuang, and YenYu Lin. 2d-3d interlaced transformer for point cloud segmentation with scene-level supervision. In *CVPR*, pages 977–987, 2023.

[64] Chaolong Yang, Yuyao Yan, Weiguang Zhao, Jianan Ye, Xi Yang, Amir Hussain, Bin Dong, and Kaizhu Huang. Towards deeper and better multi-view feature fusion for 3d semantic segmentation. In *ICONIP*, pages 3–15, 2023. 2

[65] Zetong Yang, Li Chen, Yanan Sun, and Hongyang Li. Visual point cloud forecasting enables scalable autonomous driving. In *CVPR*, pages 14673–14684, 2024. 1

[66] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *ICCV*, pages 12–22, 2023. 5

[67] Jianlong Yuan, Zelu Deng, Shu Wang, and Zhenbo Luo. Multi receptive field network for semantic segmentation. In *WACV*, pages 1894–1903, 2020. 3

[68] Jiaming Zhang, Ruiping Liu, Hao Shi, Kailun Yang, Simon Reiß, Kunyu Peng, Haodong Fu, Kaiwei Wang, and Rainer Stiefelhagen. Delivering arbitrary-modal semantic segmentation. In *CVPR*, pages 1136–1147, 2023. 1

[69] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *ICCV*, pages 16259–16268, 2021. 2, 6, 7, 1

[70] Weiguang Zhao, Yuyao Yan, Chaolong Yang, Jianan Ye, Xi Yang, and Kaizhu Huang. Divide and conquer: 3d point cloud instance segmentation with point-wise binarization. In *ICCV*, pages 562–571, 2023. 2, 5

[71] Weiguang Zhao, Guanyu Yang, Rui Zhang, Chenru Jiang, Chaolong Yang, Yuyao Yan, Amir Hussain, and Kaizhu Huang. Open-pose 3d zero-shot learning: Benchmark and challenges. *NN*, 181:106775, 2025. 2

[72] Zhisheng Zhong, Jiequan Cui, Yibo Yang, Xiaoyang Wu, Xiaojuan Qi, Xiangyu Zhang, and Jiaya Jia. Understanding imbalanced semantic segmentation through neural collapse. In *CVPR*, pages 19550–19560, 2023. 6

[73] Jinjing Zhu, Yunhao Luo, Xu Zheng, Hao Wang, and Lin Wang. A good student is cooperative and reliable: Cnn-transformer collaborative learning for semantic segmentation. In *ICCV*, pages 11720–11730, 2023. 1

[74] Liping Zhu, Cong Peng, Bingyao Wang, Chengyang Li, and Kaijie Zhu. Cbflnet: Cross-boundary feature learning for large-scale point cloud segmentation. *EAAI*, 126:106926, 2023. 3

[75] Qinfeng Zhu, Lei Fan, and Ningxin Weng. Advancements in point cloud data augmentation for deep learning: A survey. *PR*, page 110532, 2024. 6