

# PointTPA: Dynamic Network Parameter Adaptation for 3D Scene Understanding

## Supplementary Material

### S1. Additional Experiments

#### S1.1. Analysis on Different Rank

One of our core hyperparameters is the rank  $r$ , which controls the feature dimension of the projection and has a significant impact on both the number of parameters and feature extraction capability. As shown in Tab. 1a, we observe that the model achieves the best performance when  $r = 64$ .

#### S1.2. Analysis on Grouping

We fix the number of points per group, resulting in dynamic group counts across different scenes. As shown in Tab. 1b, using a fixed number of points yields performance slightly inferior to our method. This may be because fixed-point sampling struggles to adapt to point clouds with varying numbers of points. Therefore, we ultimately adopt the strategy of using a fixed number of groups for grouping.

#### S1.3. Results on ScanNet200

To further verify the robustness of our method, we perform experiments on the ScanNet200 [8] dataset. Built upon the ScanNet [3] benchmark, ScanNet200 expands the labeled categories, making the segmentation task substantially more challenging. As shown in Tab. 2, our PointTPA obtains 83.0% in allAcc, a 0.7% improvement over DAPT [13], and achieve 33.2% in mIoU, surpassing PointGST [7] by 1.0%.

Table 1. Ablation on ranks and grouping strategies. Here, N denotes the number of points per group.

(a) Ablation on different ranks.				(b) Ablation on grouping.		
$r$	mIoU	mAcc	allAcc	N	mIoU	mAcc
32	77.4	85.5	91.7	100	78.1	86.2
64	<b>78.4</b>	<b>86.3</b>	<b>92.2</b>	200	77.9	86.3
96	77.9	85.9	92.2	400	77.8	85.7
128	77.7	85.6	92.1	<b>Ours</b>	<b>78.4</b>	<b>86.3</b>

#### S2. Detailed Time Discussion

As shown in the table, DPP and SNG incur comparable training time per scene. During inference, they introduce negligible overhead compared to the setting without them, which is consistent with our expectations.

Method	Train <sub>(ms)</sub>	Infer <sub>(ms)</sub>
SNG	+9	+2
DPP	+10	+0
Both	+22	+2

#### S3. More Analysis

**Segmentation Comparison** Fig. 1 provides qualitative comparisons between FFT, 3D PEFT methods (e.g.,

Table 2. Validation results on ScanNet200 [8].

	Param. (M)	Reference	ScanNet200 Val [8]	
			allAcc	mIoU
<i>Full Fine-tuning</i>				
Sonata(w/o dec.) [9]	108.5	CVPR 25	83.8	33.9
<i>General PEFT methods</i>				
Sonata(linear probing)	0.2 (0.2%)	CVPR 25	81.1	28.4
+ Prefix Tuning [6]	1.2 (1.1%)	ACL 21	81.5(+0.4)	29.4(+1.0)
+ BitFit [11]	0.4 (0.4%)	ACL 22	82.3(+1.2)	31.1(+2.7)
+ LoRA [4]	1.2 (1.1%)	ICLR 22	82.5(+1.4)	31.7(+3.3)
+ VeRA [5]	0.3 (0.3%)	ICLR 24	81.8(+0.7)	29.8(+1.4)
+ RandLoRA [1]	0.9 (0.8%)	ICLR 25	82.0(+0.9)	30.3(+1.9)
<i>PEFT methods for point cloud</i>				
+ IDPT [12]	2.8 (2.6%)	ICCV 23	81.2(+0.1)	27.9(-0.5)
+ DAPT [13]	1.4 (1.3%)	CVPR 24	82.3(+1.2)	33.2(+4.8)
+ PointGST [7]	1.3 (1.2%)	TPAMI 25	82.1(+1.0)	32.2(+3.8)
+ PointTPA (ours)	1.4 (1.3%)	–	<b>83.0(+1.9)</b>	<b>33.2(+4.8)</b>

IDPT [12], DAPT [13]), and our PointTPA. As highlighted in Fig. 1(a), PointTPA uniquely distinguishes the tables from the sofa where other methods fail. Furthermore, in Fig. 1(b), the predictions of PointTPA for fine-grained structures like bench legs are significantly closer to the FFT than IDPT and DAPT. These results demonstrate PointTPA’s superior local feature extraction and its robustness in adapting to complex 3D scenes.

**Qualitative Analysis** In Fig. 2, we provide additional visualization results on ScanNet, ScanNet200, ScanNet++ and S3DIS with different viewpoints. In detail, we select one scenario with different viewpoints from the original scenario point cloud of each dataset. It is evident that PointTPA produces superior segmentation results.

#### S4. Test on ScanNet

As shown in Tab. 3, we evaluate our method on the ScanNet test set. PointTPA outperforms DAPT by 1.7% and achieves performance on par with the FFT reference, with a marginal gap of 0.1%. These results further demonstrate the strong generalizability of PointTPA.

Table 3. Test on ScanNet [3]

Method	Reference	mIoU
Sonata (w/o dec.)	CVPR 25	75.9
Sonata (lin)	CVPR 25	70.0
+ DAPT	CVPR 24	74.1(+4.1)
+ <b>PointTPA (ours)</b>	–	<b>75.8(+5.8)</b>

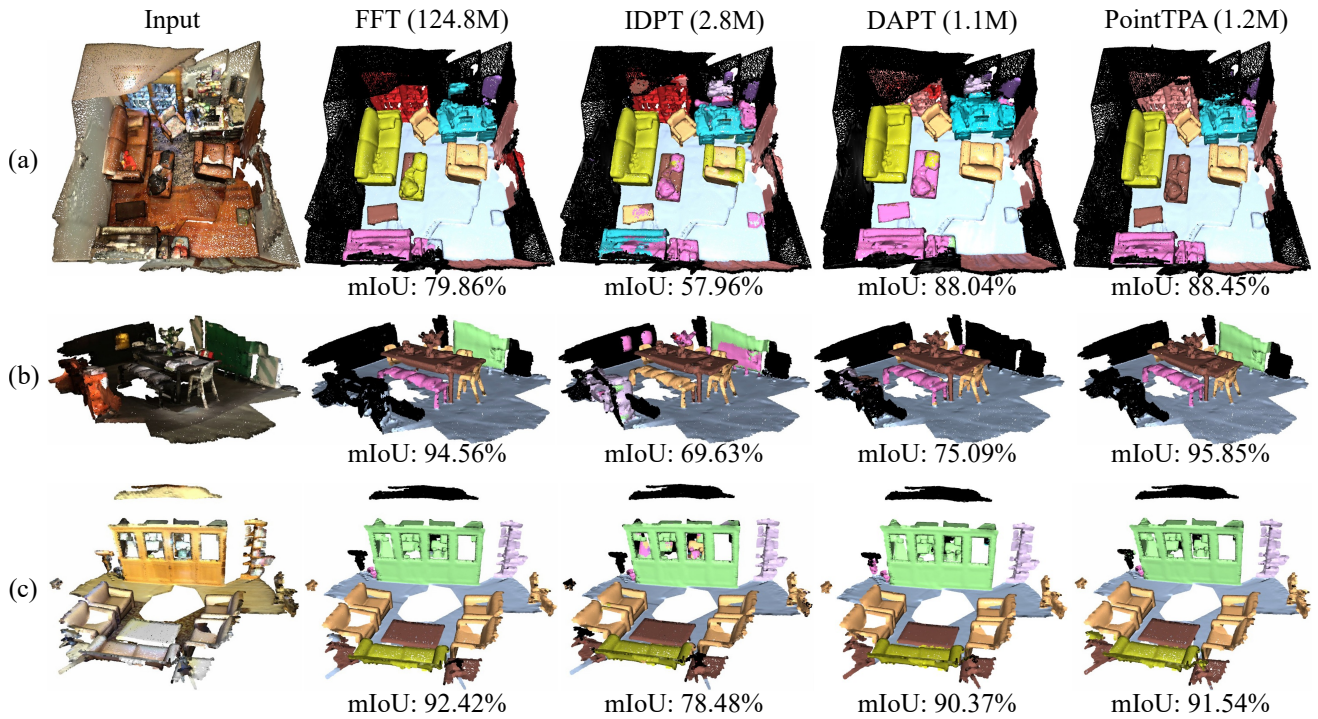


Figure 1. A comparison of FFT, IDPT [12], DAPT [13], and PointTPA on segmentation performance, evaluated on ScanNet [3].

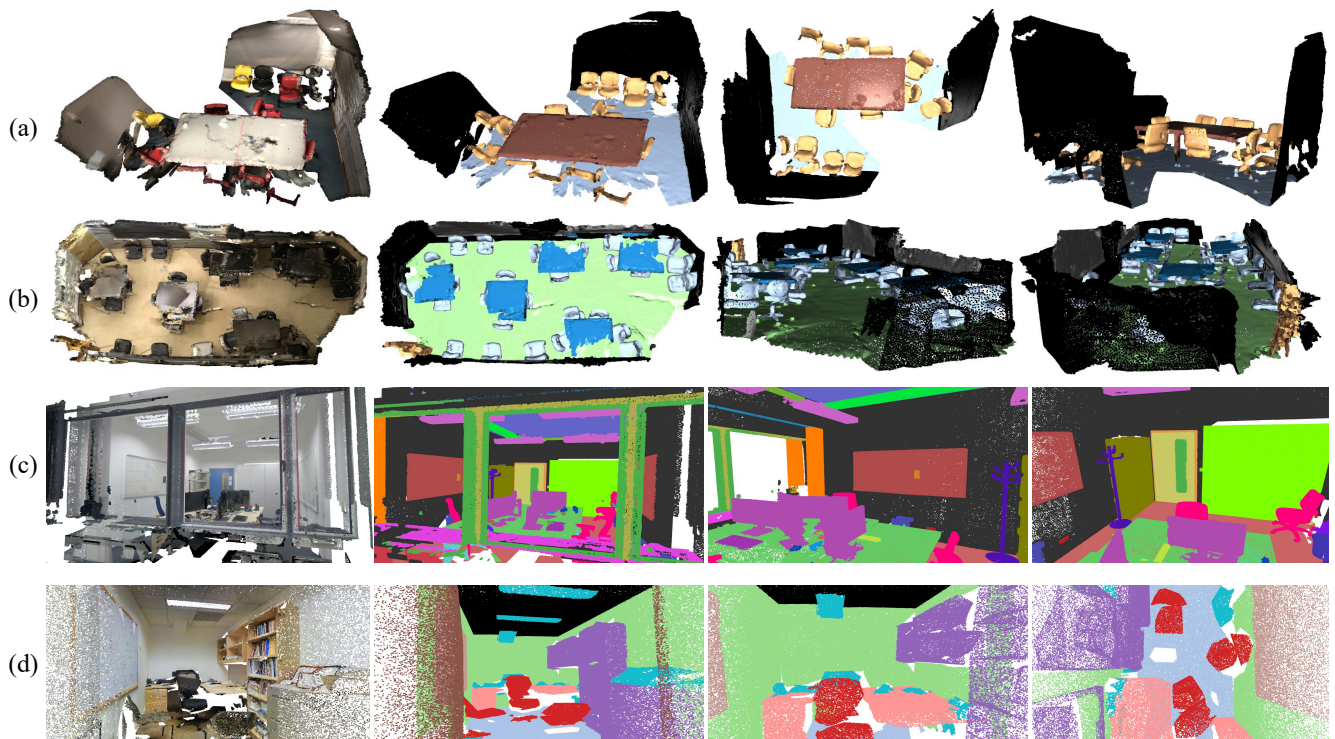


Figure 2. More visualizations of the semantic segmentation results of our PointTPA on four large-scale scene datasets. (a) ScanNet [3], (b) ScanNet200 [8], (c) ScanNet++ [10], (d) S3DIS [2] with 3 views.

## References

- [1] Paul Albert, Frederic Z Zhang, Cristian Rodriguez-Opazo, Hemanth Saratchandran, Anton van den Hengel, and Ehsan Abbasnejad. RandLora: full rank parameter-efficient fine-tuning of large models. 2024. 1
- [2] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 1534–1543, 2016. 2
- [3] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 5828–5839, 2017. 1, 2
- [4] Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. In *Proc. Int. Conf. Learn. Representations*, 2022. 1
- [5] Dawid Jan Kopiczko, Tijmen Blankevoort, and Yuki M Asano. Vera: Vector-based random matrix adaptation. In *Proc. Int. Conf. Learn. Representations*, 2024. 1
- [6] Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proc. Annual Meeting of the Association for Computational Linguistics*, pages 4582–4597, 2021. 1
- [7] Dingkan Liang, Tianrui Feng, Xin Zhou, Yumeng Zhang, Zhikang Zou, and Xiang Bai. Parameter-efficient fine-tuning in spectral domain for point cloud learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(12):10949–10966, 2025. 1
- [8] David Rozenberszki, Or Litany, and Angela Dai. Language-grounded indoor 3d semantic segmentation in the wild. In *Proc. Eur. Conf. Comput. Vis.*, pages 125–141, 2022. 1, 2
- [9] Xiaoyang Wu, Daniel DeTone, Duncan Frost, Tianwei Shen, Chris Xie, Nan Yang, Jakob Engel, Richard Newcombe, Hengshuang Zhao, and Julian Straub. Sonata: Self-supervised learning of reliable point representations. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 22193–22204, 2025. 1
- [10] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 12–22, 2023. 2
- [11] Elad Ben Zaken, Yoav Goldberg, and Shauli Ravfogel. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. In *Proc. Annual Meeting of the Association for Computational Linguistics*, pages 1–9, 2022. 1
- [12] Yaohua Zha, Jinpeng Wang, Tao Dai, Bin Chen, Zhi Wang, and Shu-Tao Xia. Instance-aware dynamic prompt tuning for pre-trained point cloud models. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 14161–14170, 2023. 1, 2
- [13] Xin Zhou, Dingkan Liang, Wei Xu, Xingkui Zhu, Yihan Xu, Zhikang Zou, and Xiang Bai. Dynamic adapter meets prompt tuning: Parameter-efficient transfer learning for point cloud analysis. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 14707–14717, 2024. 1, 2