

— Supplementary Material —

MambaTron: Efficient Cross-Modal Point Cloud Enhancement using Aggregate Selective State Space Modeling

Sai Tarun Inaganti
Robotics Institute, University of Minnesota
inaga015@umn.edu

Gennady Petrenko
Homothereum
gennady@homothereum.org

Table 1. Effects of APR

APR	affine	CD (avg)	
		known	novel
✗	✗	1.354	2.737
✓	✗	1.317	2.611
✓	✓	1.199	2.333

Table 2. ModelNet40 dataset

Method	OA (%)
PCM [6]	93.4
PointMamba [2]	93.6
Point-MAE [3]	93.8
Joint-MAE [1]	94.0
Ours	94.3

1. Effects of the APR scheme

The Adjacency Preserving Reordering (APR) scheme is space-filling, similar to the Hilbert [2] and XYZ [6] ordering techniques. Additionally, the scheme theoretically provides an infinite ways to order the points, only limited by the number of points, thanks to a learnable affine transformation layer. We look at the effects of each component through a further ablation study on the ShapeNet-ViPC dataset in Table 1.

2. Performance statistics on additional datasets

The task of cross-modal point cloud completion is a recent idea introduced [7] along with the ShapeNet-ViPC dataset, with no equivalent dataset that is just as comprehensive, i.e. contains class-mapped pointclouds with corresponding reference view images. For unimodal analysis, in addition to the real-world 3D object classification on the ScanObjectNN dataset, we perform experiments that we describe below with the same encoder configuration as that of the classification task on ScanObjectNN. The training data is subject to random rotation and slight scaling operations at each epoch. For comparison purposes, we take the results reported in the other research works.

Classification on ModelNet40. This [4] is a synthetic CAD dataset with 40 classes that contains a total of 12,311 noiseless pre-aligned models. The classification head is based on that of PointMamba [2]. We report the Overall Accuracy (OA) in percentage in Table 2.

Table 3. ShapeNetPart dataset

Method	mIoU (%)	
	class	instance
Point-MAE [3]	84.2	86.1
Joint-MAE [1]	85.4	86.3
PointMamba [2]	84.4	86.2
PCM [6]	87.0	85.3
Ours	87.4	86.9

Segmentation on ShapeNetPart. We perform the task of part segmentation on the ShapeNetPart [5] dataset which contains 16,881 shapes from 16 categories, with 50 segmentation parts in total. We report the mean Intersection over Union (mIoU) percentages at both the class level and the instance level in Table 3. Our model demonstrates superior performance by a good margin, especially on the instance level.

References

- [1] Ziyu Guo, Renrui Zhang, Longtian Qiu, Xianzhi Li, and Pheng-Ann Heng. Joint-mae: 2d-3d joint masked autoencoders for 3d point cloud pre-training. In Edith Elkind, editor, *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 791–799. International Joint Conferences on Artificial Intelligence Organization, 8 2023. Main Track. 1

- [2] Ding kang Liang, Xin Zhou, Wei Xu, Xingkui Zhu, Zhikang Zou, Xiaoqing Ye, Xiao Tan, and Xiang Bai. Pointmamba: A simple state space model for point cloud analysis. In *Advances in Neural Information Processing Systems*, 2024. [1](#)
- [3] Yatian Pang, Wenxiao Wang, Francis E. H. Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, page 604–621, Berlin, Heidelberg, 2022. Springer-Verlag. [1](#)
- [4] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. [1](#)
- [5] Li Yi, Vladimir G. Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. A scalable active framework for region annotation in 3d shape collections. 35(6), Dec. 2016. [1](#)
- [6] Tao Zhang, Haobo Yuan, Lu Qi, Jiangning Zhang, Qianyu Zhou, Shunping Ji, Shuicheng Yan, and Xiangtai Li. Point cloud mamba: Point cloud learning via state space model, 2024. [1](#)
- [7] Xuancheng Zhang, Yutong Feng, Siqu Li, Changqing Zou, Hai Wan, Xibin Zhao, Yandong Guo, and Yue Gao. View-guided point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15890–15899, 2021. [1](#)