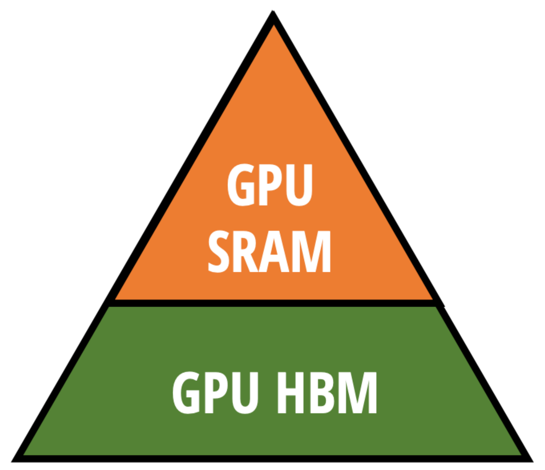
PointMamba: A Simple State Space Model for Point Cloud Analysis

Dingkang Liang[1*], Xin Zhou[1*], Wei Xu[1], Xingkui Zhu[1], Zhikang Zou[2], Xiaoqing Ye[2], Xiao Tan[2], Xiang Bai[1†]

[1]Huazhong University of Science and Technology, [2]Baidu Inc.

[*]Equal contribution. [†]Corresponding author.

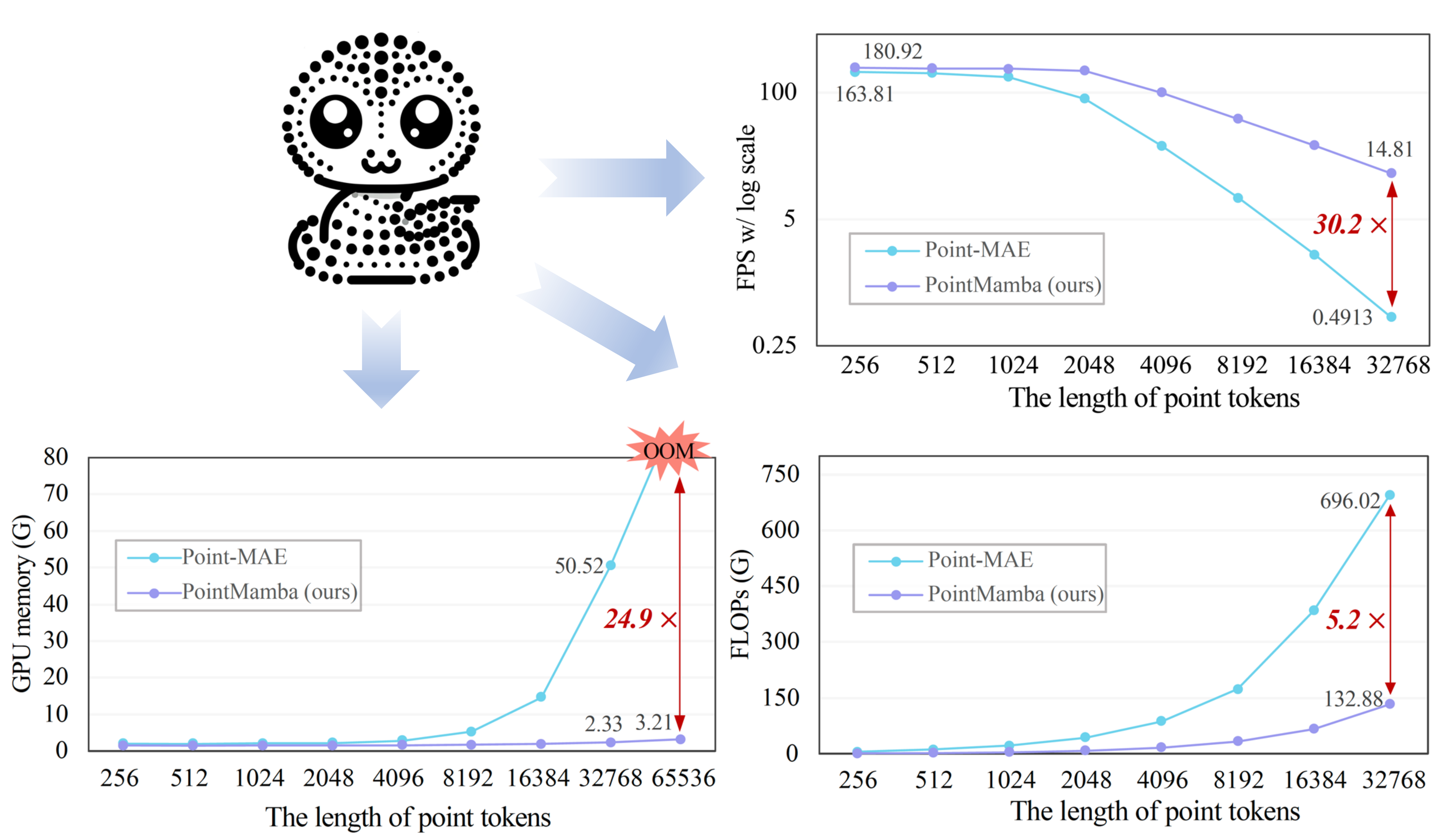
Problem & Motivation



- The complexity of attention mechanisms is quadratic, bringing significant computational cost, which is not friendly to low-resource devices.
- Mamba has emerged as a promising architectures for sequence modeling thanks to its strong representation ability and linear-time complexity.

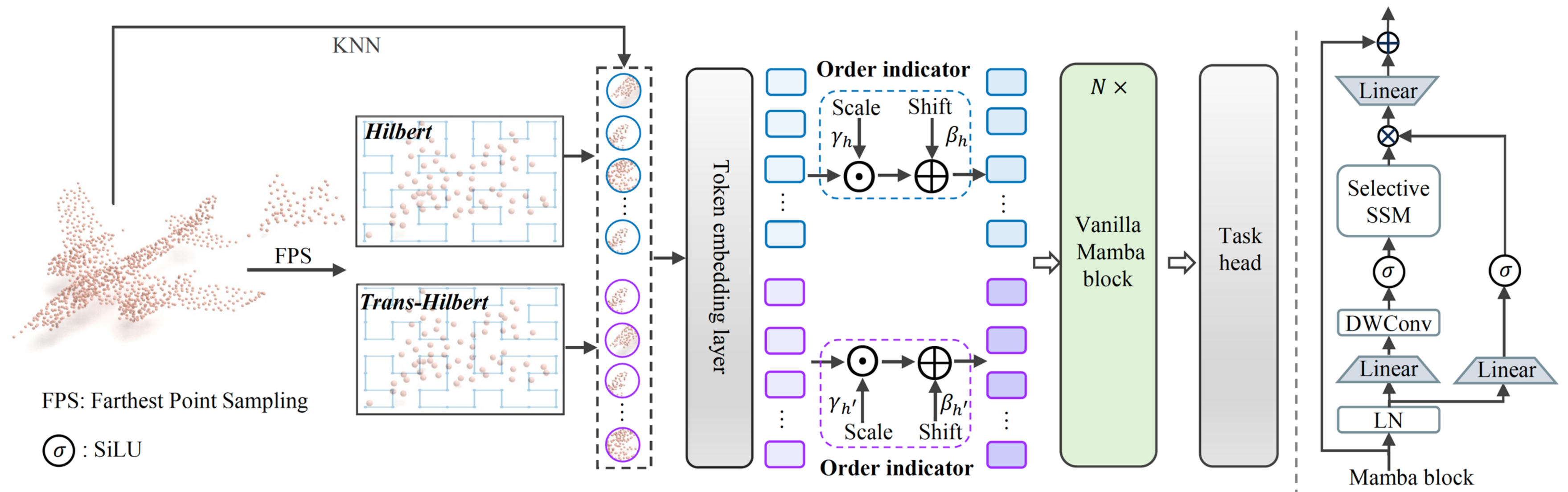
Q: *How to design a simple, elegant method that operates with linear complexity, thereby retaining the benefits of global modeling for point cloud analysis?*

Contribution



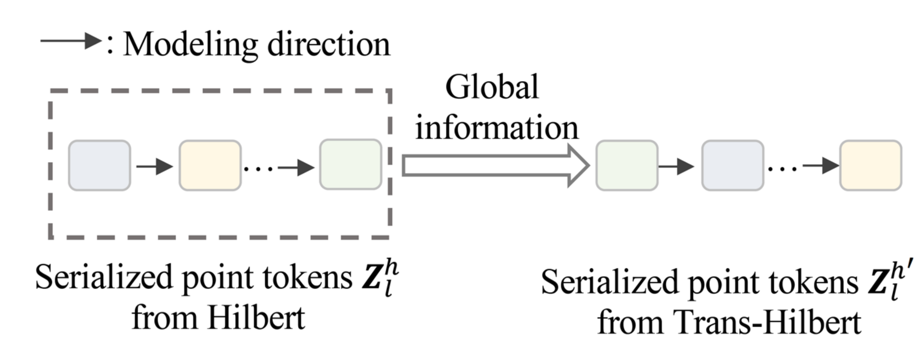
- We introduce the first state space model for point cloud analysis, named PointMamba , which features global modeling with linear complexity. Despite the absence of elaborate or complex structural designs, PointMamba demonstrates its potential as an optional model for 3D vision applications.
- Our PointMamba exhibits impressive capabilities, including structural simplicity (e.g., vanilla Mamba), low computational cost, and knowledge transferability (e.g., support for self-supervised learning).

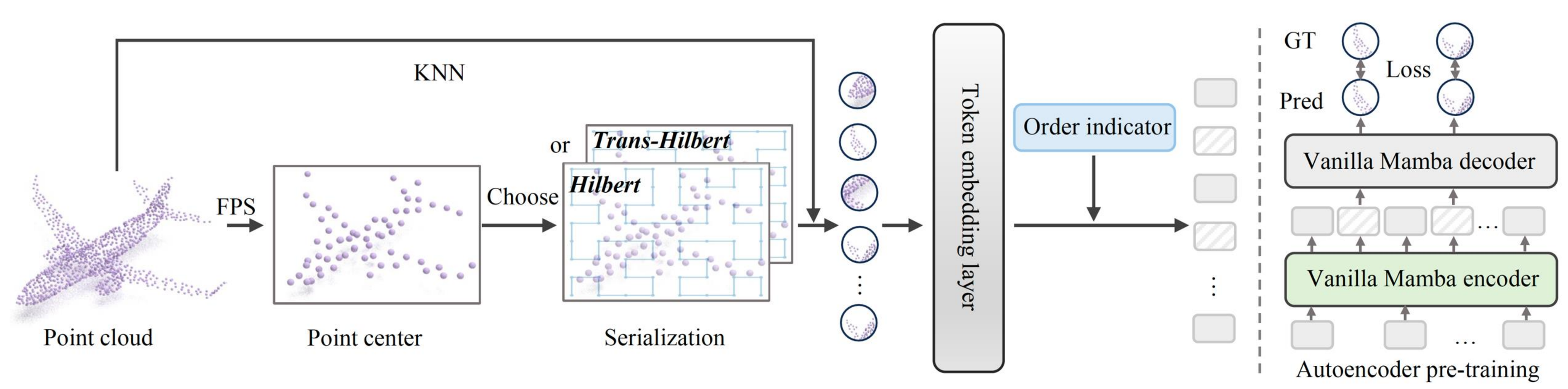
Methodology



Overview: The goal of this paper is to provide a simple yet solid Mamba baseline for point cloud analysis tasks and explore the potential of plain and non-hierarchical Mamba. Thus, in the spirit of *Occam's razor*, we make the structure as simple as possible without any complex or elaborate design.

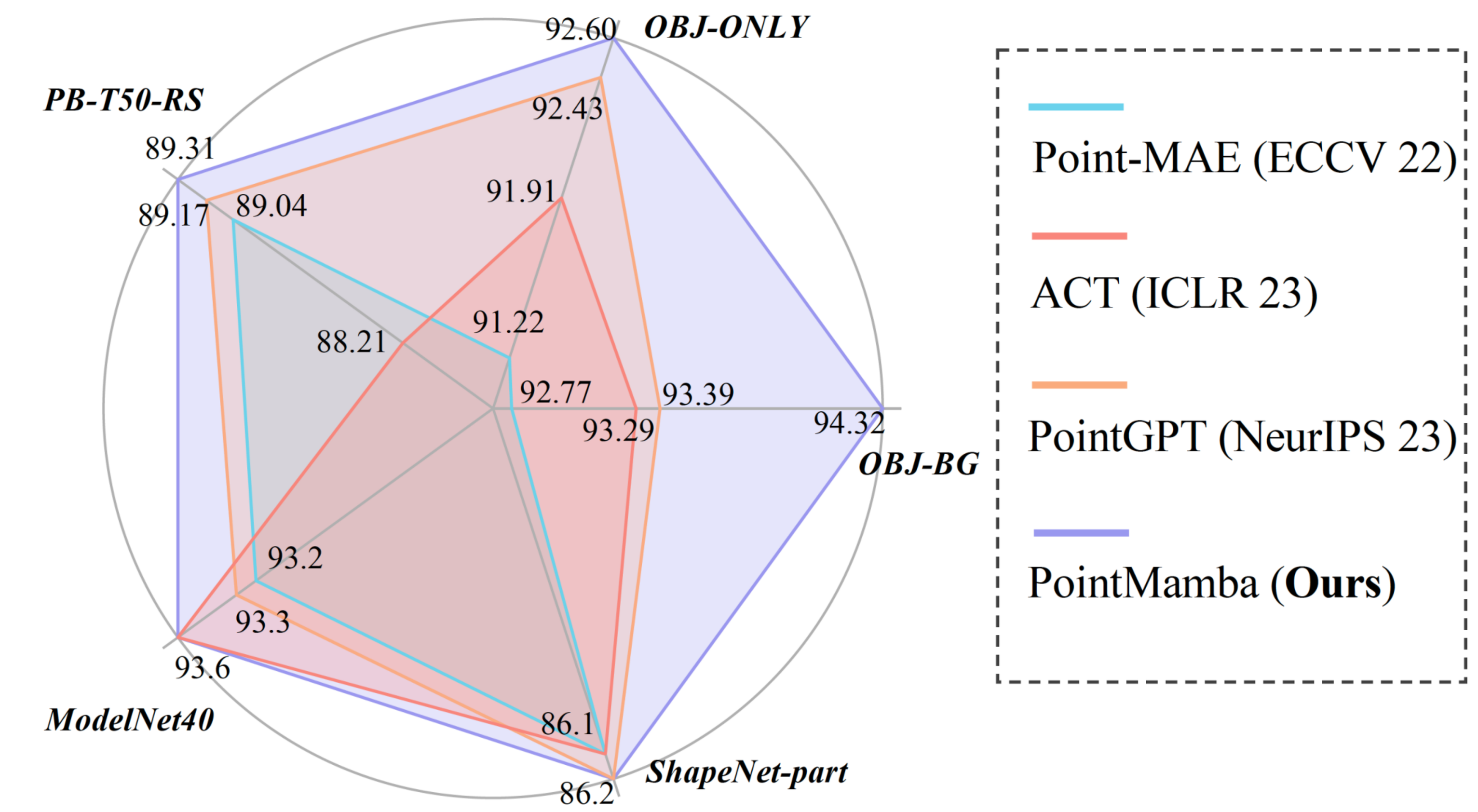
- **Point scanning strategy:** We propose to utilize two types of space-filling curves, including Hilbert and Trans-Hilbert, to generate the serialized key points.
- **Order indicator:** Maintaining the distinct characteristics of these different scanning strategies is important for preserving the integrity of the spatial information.
- **Mamba encoder:** For each Mamba block, layer normalization (LN), Selective SSM, depth-wise convolution (DW), and residual connections are employed.
- **Global modeling mechanism:** Ensure that each serialized point in the Trans-Hilbert sequence is informed by the entire history of the previously processed Hilbert sequence, enabling a more contextually rich and globally aware modeling process.



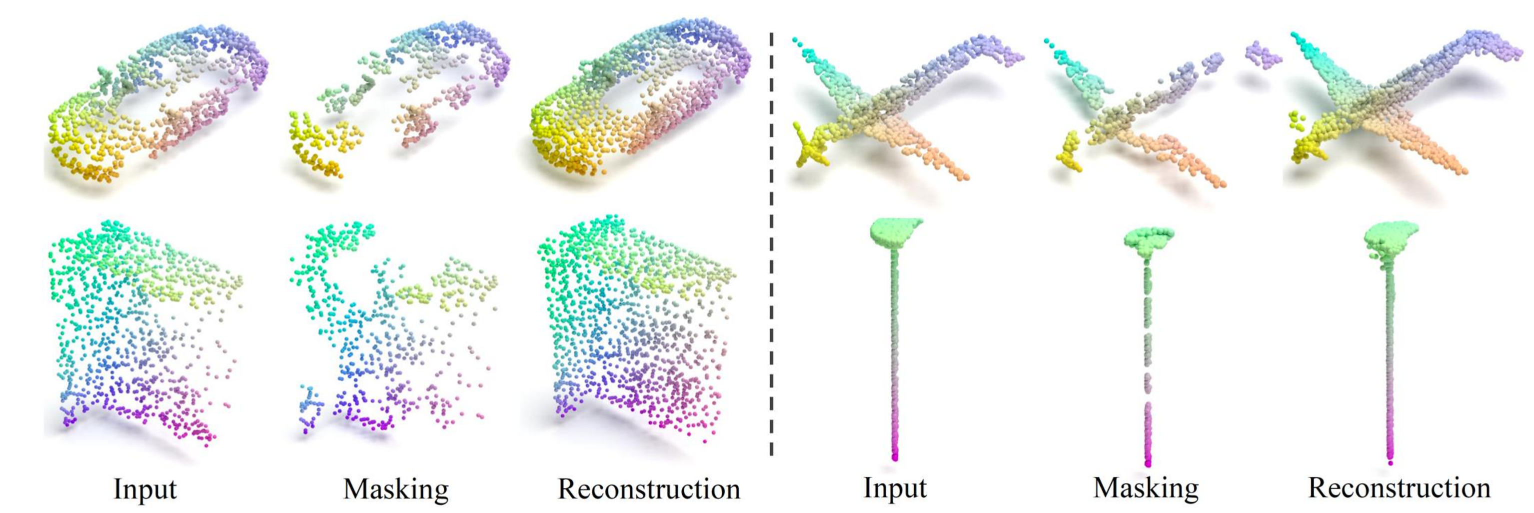


Serialization-based mask modeling: considering the unidirectional modeling of Mamba, we customize a simple yet effective serialization-based mask modeling paradigm. During the pre-training, we randomly choose one space-filling curve to generate the serialized point tokens for mask modeling, and different serialized point tokens have different order indicators.

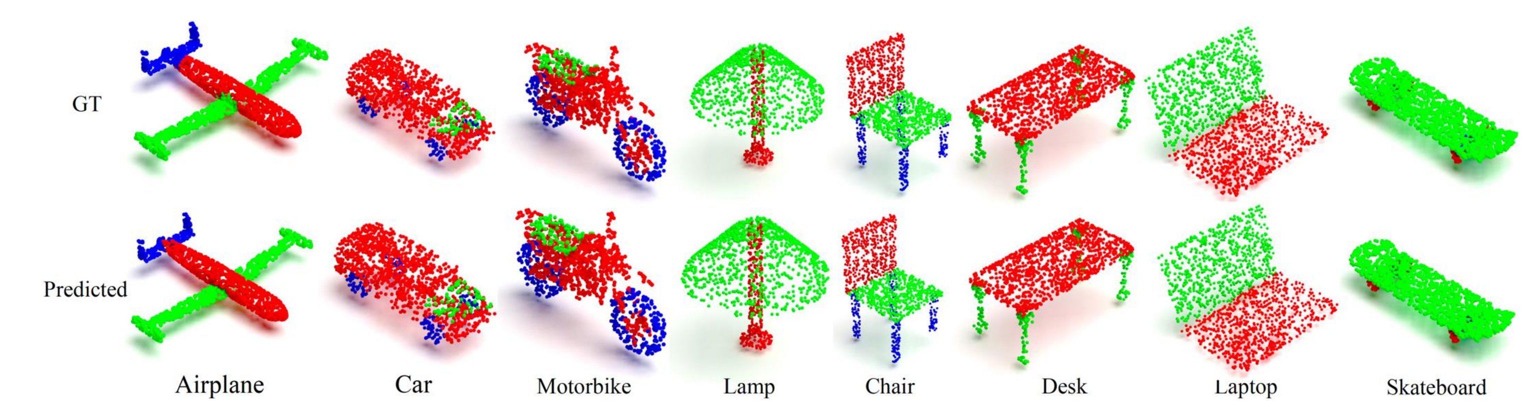
Experiments



Without bells and whistles, our PointMamba achieve better performance than the representative Transformer-based methods on the various point cloud analysis datasets.



The qualitative results of mask predictions of our PointMamba on ShapeNet validation set.



The qualitative results of part segmentation of our PointMamba on ShapeNetPart.

References

[1] PointGPT: Auto-regressively Generative Pre-training from Point Clouds. NeurIPS 23.

[2] Autoencoders as Cross-Modal Teachers. ICLR 23.

[3] Masked Autoencoders for Point Cloud Self-supervised Learning. ECCV 22.