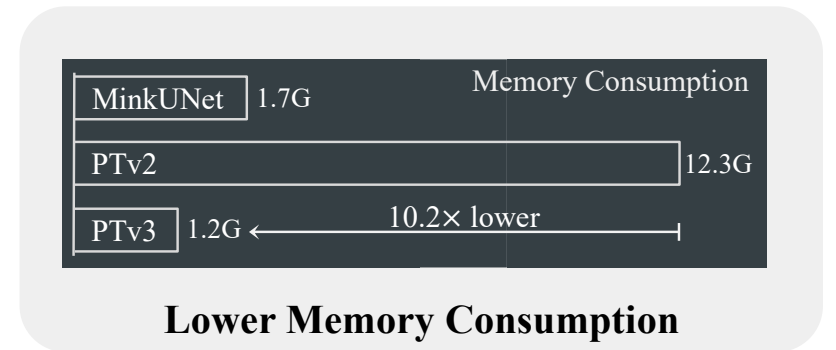
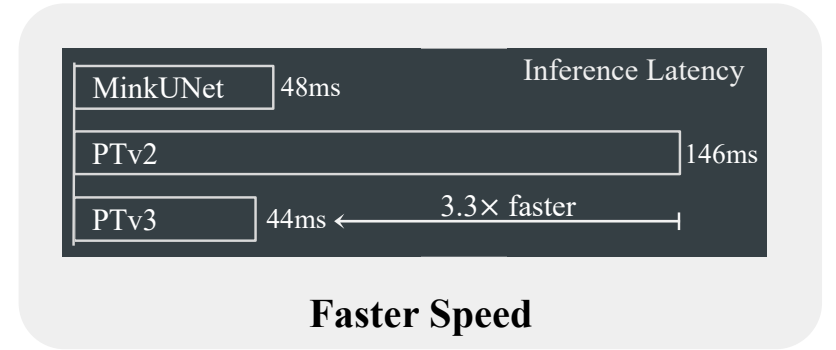
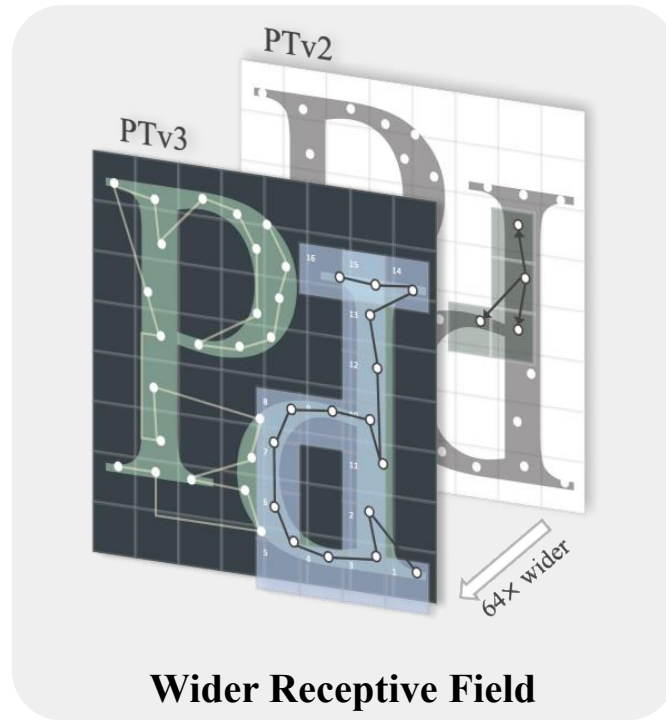
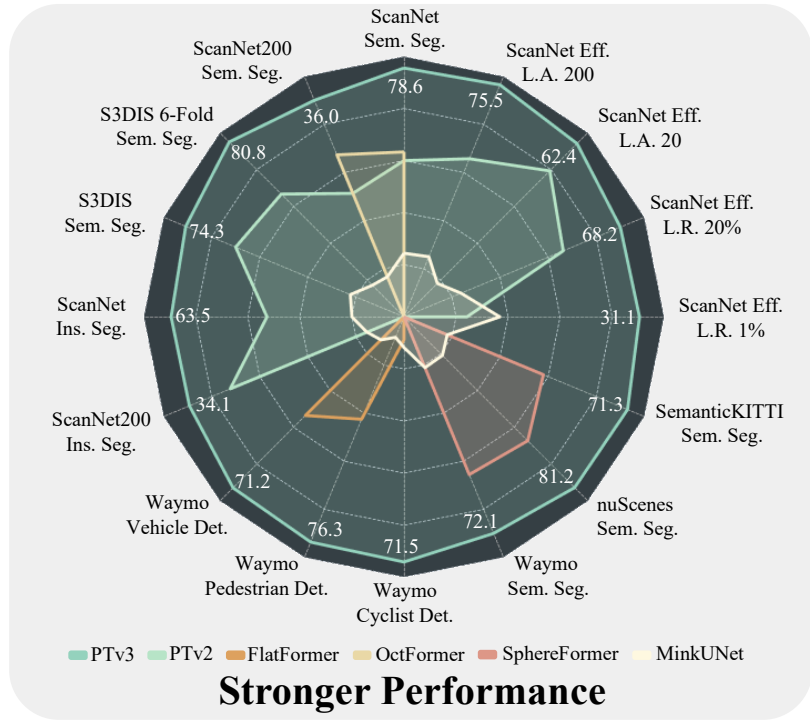
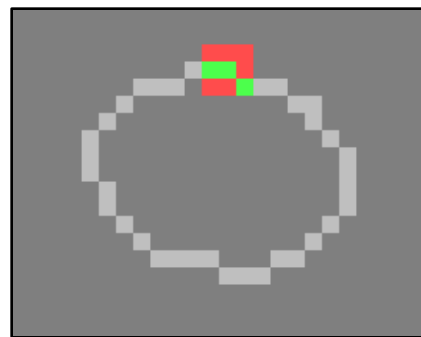
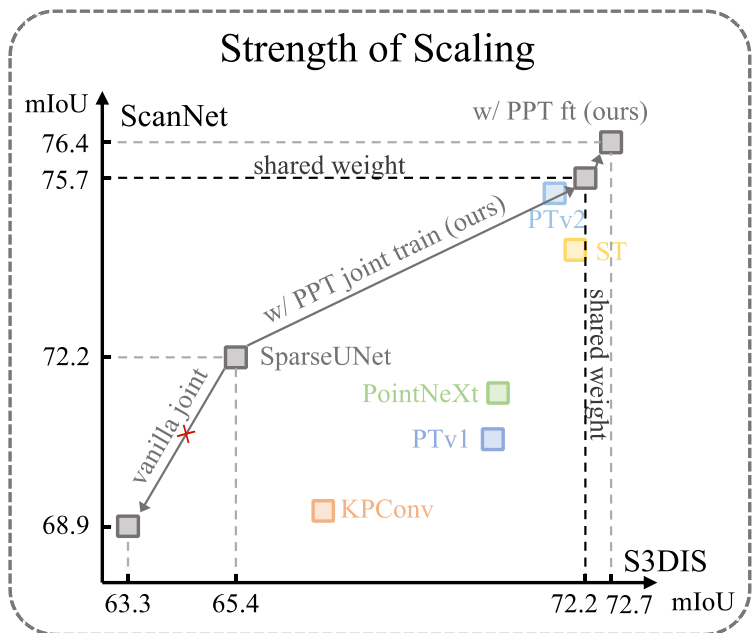


Point Transformer V3: Simpler, Faster, Stronger

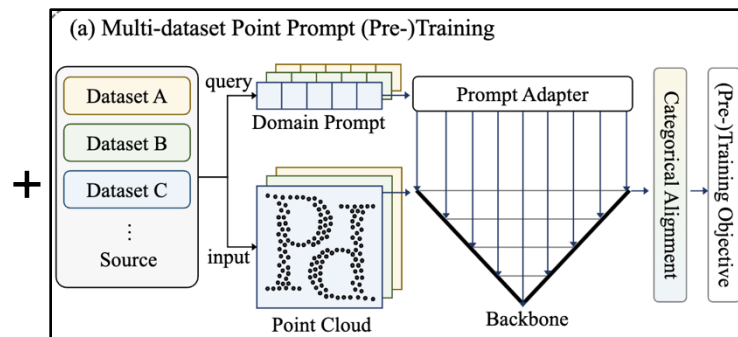
- ⇒ State-of-the-art performance on over **20** downstream tasks that span both **indoor** and **outdoor** scenarios.
- ⇒ Expanding the receptive field from **16** to **1024** points while remaining efficient.
- ⇒ **3x** increase in processing speed and **10x** improvement in memory efficiency compared with PTV2



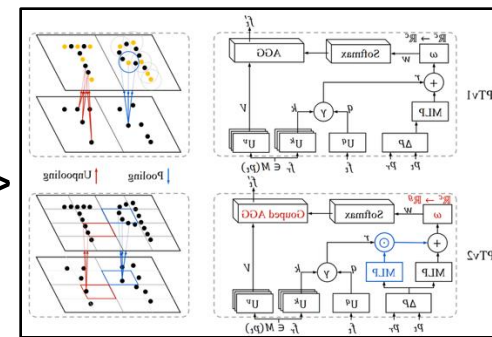
Point Transformer V3: Scaling Principle



SparseUNet (2019)



PPT (Large-scale Training)

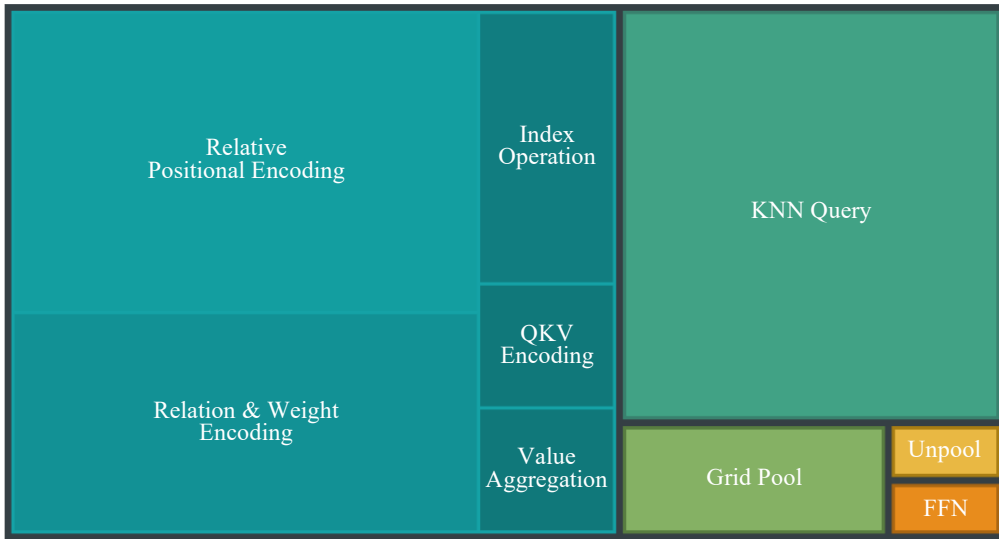


PTv2 (2022)

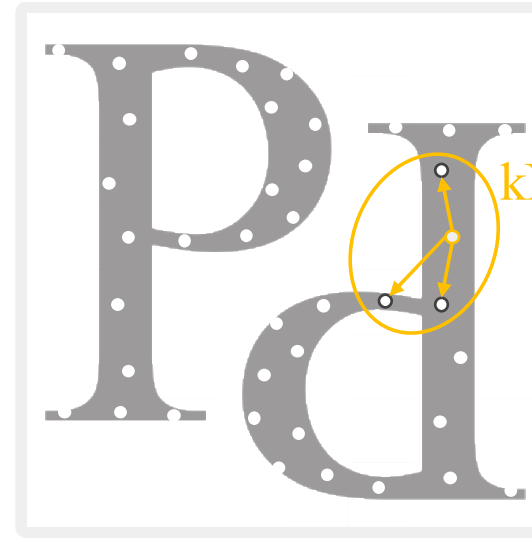
- ⇒ Enhanced with large-scale pre-training, **SparseUNet** surpasses **Point Transformers** in accuracy while remaining efficient;
- ⇒ **Point Transformers** fails to scale up due to limitations in efficiency;
- ⇒ We hypothesize that model performance is **more significantly influenced by scale** than by complex design details;
- ⇒ We should **prioritize simplicity and efficiency over the accuracy** of certain mechanisms;
- ⇒ **Efficiency enable scalability** and further enable stronger accuracy.

From *Towards Large-scale 3D Representation Learning with Multi-dataset Point Prompt Training*

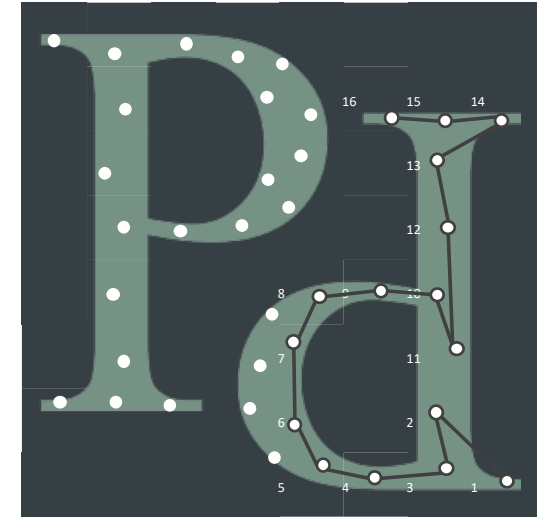
Point Transformer V3: breaking the curse of permutation invariance



PTv2 Forward Latency



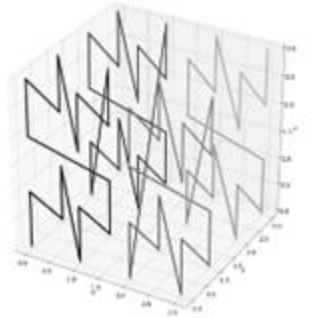
PTv2



PTv3

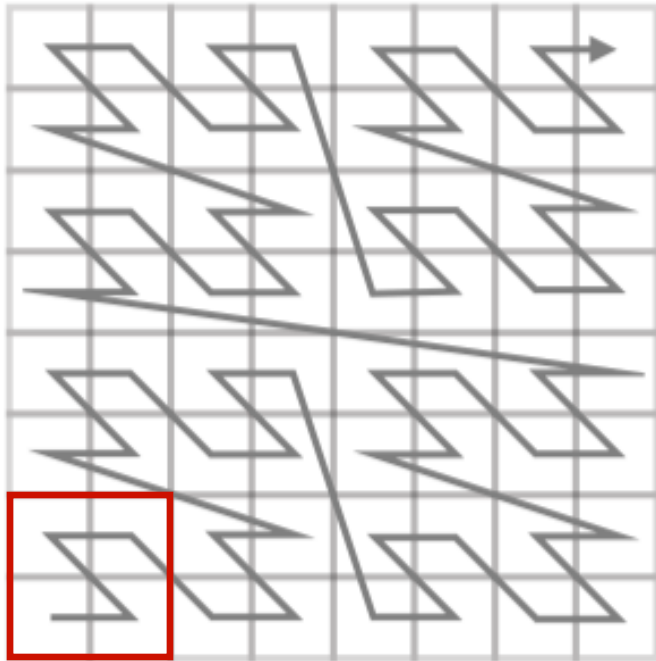
- ⇒ Classical point cloud transformers build upon point-based backbones, which treat point clouds as **unstructured data** and rely on neighboring query algorithms like **kNN**;
- ⇒ Yet kNN is extremely inefficient due to difficulty in parallelization. (28% latency)
- ⇒ **Do we really need the accurate neighbors** queried by kNN? **No**, attention is adaptive to kernel shape, all we need to do is relatively precise and **enlarge the kernel shape**;
- ⇒ Inspired by OctFormer and FlatFormer, we move away from the traditional paradigm, which treats point clouds as unordered sets. we choose to **break the curse by serializing point cloud into a structured format**.

Point Transformer V3: making unstructured sparse data structured!

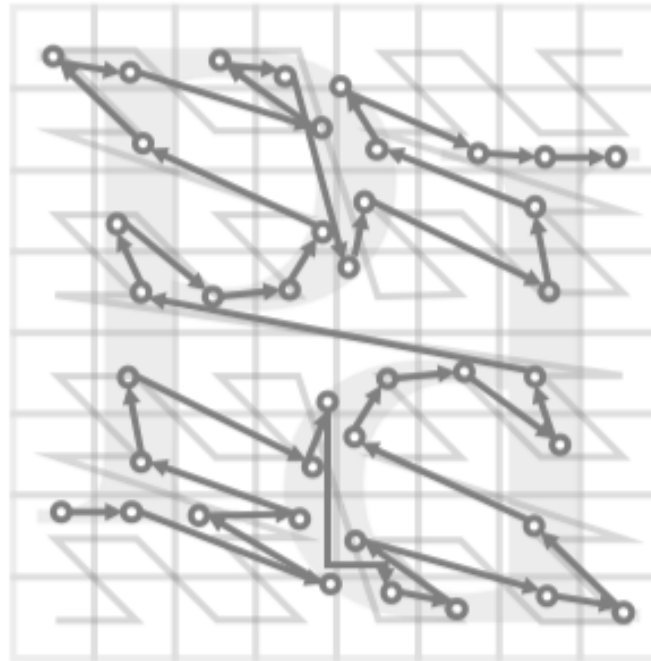


Do precise neighborhood by KNN really matter?

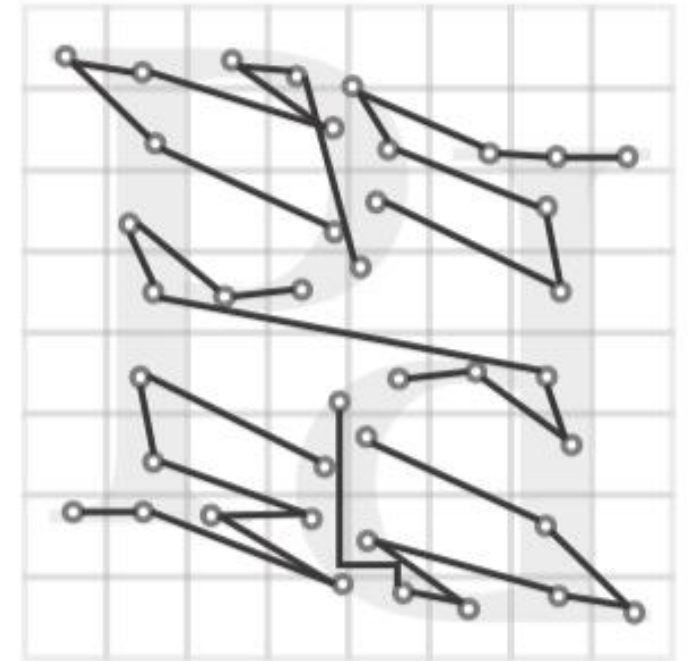
=> No! Attention is adaptive! We just need to make sure the kernel is large! 16 => 1024



Space-filling Curve



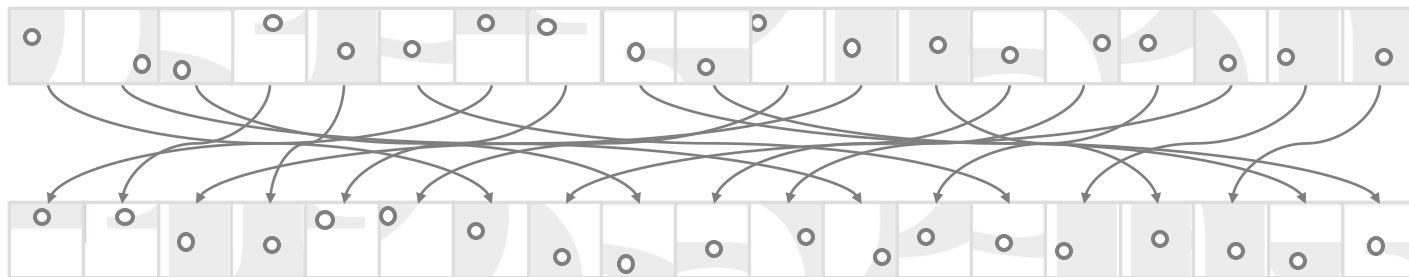
Ordered Point Cloud
structured 1D format



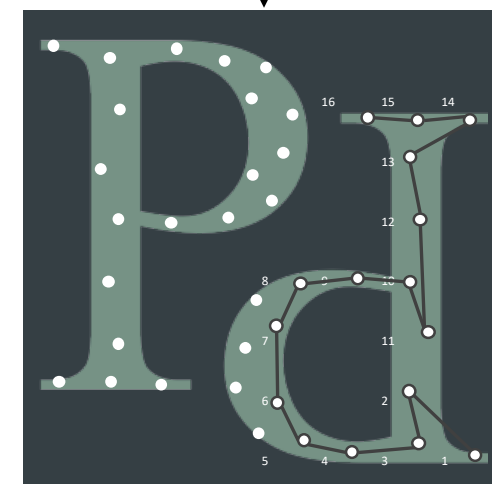
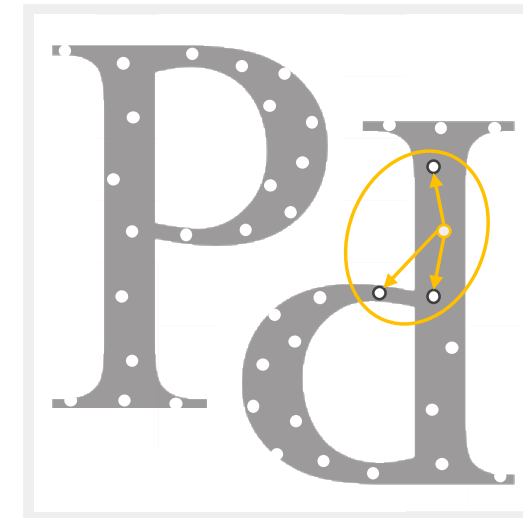
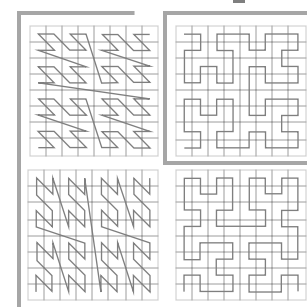
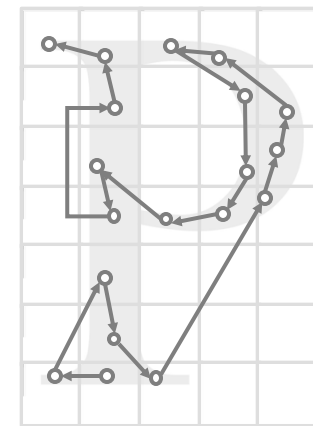
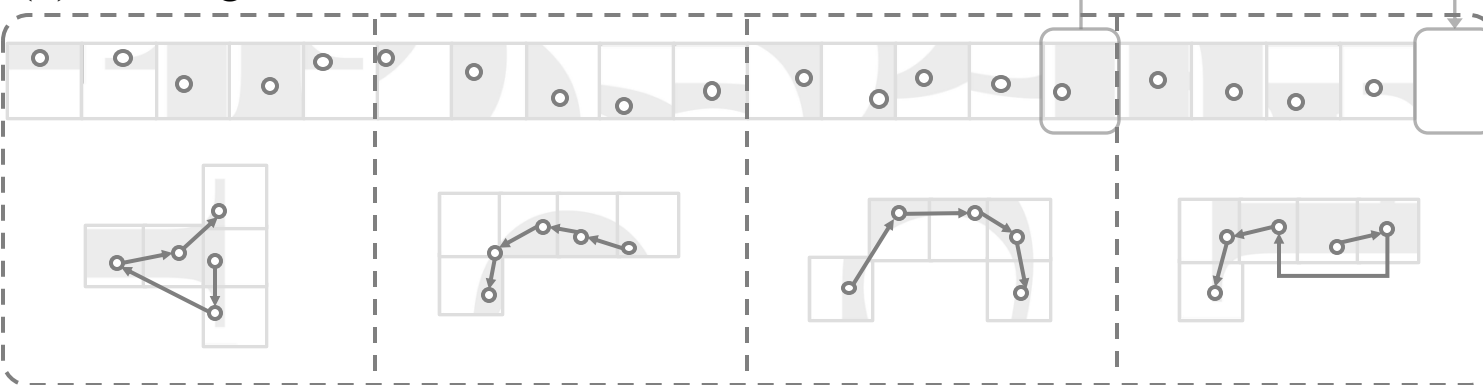
Patch Partition
preserving spatial
neighbor relationships

Point Transformer V3: Serialized Attention

(a) Reordering

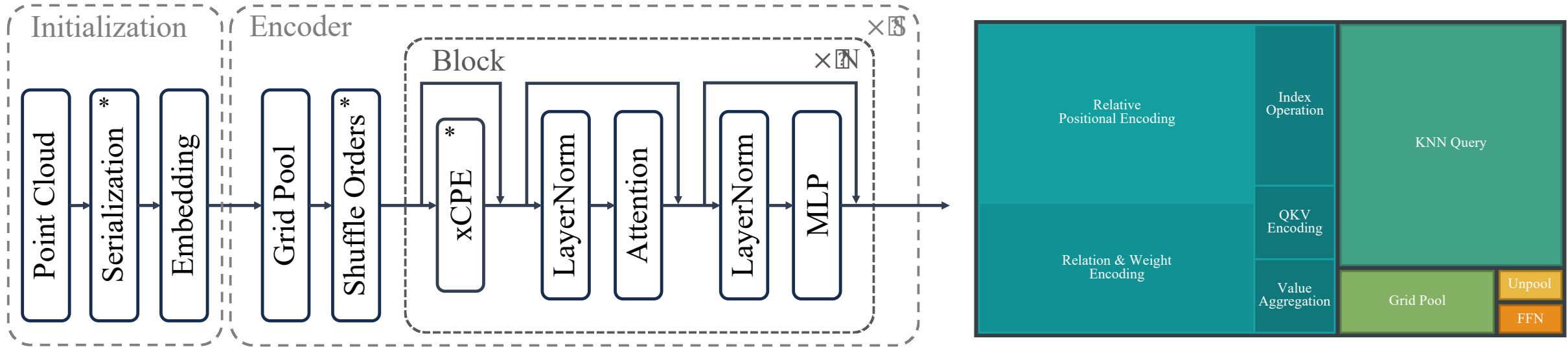


(b) Padding



- ⇒ **Computing the order** of given unstructured point cloud and space-filling curve pattern.
- ⇒ **Making** the point cloud **structured** with the computed order after padding.
- ⇒ Then the unstructured data is arranged as an 1D array just as **language tokens**.
- ⇒ We can directly **apply** well-optimized **attention operator** designed for structured data.

Point Transformer V3: xCPE



- ⇒ **Relative positional encoding** is also time-consuming for point cloud transformers.
- ⇒ Use **Sparse Convolution** as a replacement for relative positional encoding (xCPE).
- ⇒ If Sparse Convolution is not easy to deploy,
 - ⇒ Use **O-CNN PyTorch** by Peng Shuang Wang as a replacement (Full PyTorch Code)
 - ⇒ <https://github.com/octree-nn/ocnn-pytorch>

Point Transformer V3 - Performance

Results

Metric: Top-1 IoU

Code available in Pointcept =>



Methods	MIOU	AIR VENT	BACKPACK	BAG	BASKET	BED	BINDER	BLANKET	BLIND RAIL	BLINDS	BOARD	BOOK	BOOKSHELF	BOTTLE	BOWL	BOX	BUCKET	CABINET	CEILING	CEILING LAMP	CHAIR	CLOCK	CLOTH	CL	
PTv3 - PPT	0.464	0.034	0.591	0.427	0.007	0.812	0.000	0.745	0.629	0.876	0.000	0.171	0.494	0.382	0.118	0.507	0.327	0.366	0.908	0.921	0.711	0.732	0.000		
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																									
PTv3	0.458	0.057	0.613	0.423	0.023	0.710	0.000	0.707	0.626	0.871	0.002	0.239	0.524	0.421	0.204	0.452	0.140	0.383	0.915	0.915	0.749	0.706	0.000		
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																									
PT-Fusion-All	0.450	0.058	0.467	0.301	0.076	0.815	0.000	0.689	0.566	0.821	0.080	0.346	0.578	0.378	0.136	0.468	0.211	0.342	0.925	0.935	0.723	0.596	0.000		
PTv2	0.427	0.073	0.463	0.219	0.003	0.679	0.000	0.667	0.597	0.873	0.007	0.187	0.523	0.435	0.295	0.461	0.101	0.369	0.916	0.902	0.712	0.727	0.000		
Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, Hengshuang Zhao. Point Transformer V2: Grouped Vector Attention and Partition-based Pooling . NeurIPS 2022																									
PonderV2-SparseUNet-base	0.386	0.000	0.389	0.110	0.054	0.739	0.000	0.565	0.530	0.822	0.000	0.382	0.548	0.325	0.000	0.361	0.315	0.396	0.917	0.929	0.730	0.547	0.000		
Haoyi Zhu, Honghui Yang, Xiaoyang Wu, Di Huang, Sha Zhang, Xianglong He, Hengshuang Zhao, Chunhua Shen, Yu Qiao, Tong He, Wanli Ouyang. PonderV2: Pave the Way for 3D Foundation Model with A Universal Pre-training Paradigm . Arxiv, 2023																									
MinkowskiNet	0.292	0.000	0.323	0.116	0.039	0.719	0.000	0.418	0.031	0.726	0.005	0.217	0.481	0.178	0.000	0.212	0.084	0.286	0.879	0.880	0.627	0.211	0.000		
Christopher Choy, JunYoung Gwak, Silvio Savarese. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks . CVPR 2019																									

1st Place Solution for ScanNet++ Dataset Challenge in Semantic Segmentation

Point Transformer V3 - Performance



Results

Code available in Pointcept =>

Metric: Top-3 IoU

Methods	MIOU	AIR VENT	BACKPACK	BAG	BASKET	BED	BINDER	BLANKET	BLIND RAIL	BLINDS	BOARD	BOOK	BOOKSHELF	BOTTLE	BOWL	BOX	BUCKET	CABINET	CEILING	CEILING LAMP	CHAIR	CLOCK	CLOTH	CL	
PTv3 - PPT	0.710	0.306	0.863	0.730	0.162	0.985	0.000	0.981	0.945	0.987	0.126	0.686	0.684	0.560	0.348	0.706	0.770	0.638	0.991	0.969	0.882	0.785	0.018		
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																									
PTv3	0.697	0.204	0.815	0.711	0.132	0.957	0.002	0.969	0.957	0.985	0.060	0.819	0.788	0.691	0.371	0.693	0.705	0.578	0.993	0.979	0.889	0.782	0.000		
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																									
PTv2	0.665	0.204	0.810	0.521	0.054	0.979	0.001	0.895	0.942	0.970	0.139	0.687	0.713	0.651	0.461	0.665	0.475	0.660	0.992	0.974	0.885	0.785	0.003		
Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, Hengshuang Zhao. Point Transformer V2: Grouped Vector Attention and Partition-based Pooling . NeurIPS 2022																									
PT-Fusion-All	0.662	0.206	0.759	0.621	0.183	0.977	0.003	0.925	0.939	0.956	0.148	0.679	0.819	0.605	0.352	0.671	0.644	0.599	0.991	0.985	0.918	0.681	0.031		
MinkowskiNet	0.531	0.004	0.650	0.383	0.110	0.915	0.000	0.809	0.913	0.926	0.055	0.577	0.751	0.348	0.117	0.511	0.482	0.555	0.988	0.984	0.862	0.373	0.009		
Christopher Choy, JunYoung Gwak, Silvio Savarese. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks . CVPR 2019																									
KPCConv	0.460	0.000	0.684	0.326	0.034	0.912	0.000	0.845	0.898	0.883	0.125	0.497	0.757	0.021	0.000	0.512	0.471	0.542	0.986	0.984	0.906	0.000	0.036		
Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, Leonidas J. Guibas. KPCConv: Flexible and Deformable Convolution for Point Clouds . ICCV 2019																									
PointNet++	0.389	0.000	0.263	0.237	0.000	0.791	0.000	0.749	0.704	0.897	0.000	0.362	0.669	0.239	0.000	0.387	0.524	0.523	0.983	0.981	0.893	0.311	0.000		
Charles R. Qi, Li Yi, Hao Su, Leonidas J. Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space . NIPS 2017																									

Point Transformer V3 - Information

Oral Session Information =>



Oral

Point Transformer V3: Simpler Faster Stronger

Xiaoyang Wu · Li Jiang · Peng-Shuai Wang · Zhijian Liu · Xihui Liu · Yu Qiao · Wanli Ouyang · Tong He · Hengshuang Zhao


Summit Flex Hall C Oral #1

[[Abstract](#)] [Visit [Orals 2C 3D from multiview and sensors](#)]

Wed 19 Jun 1 p.m. – 1:18 p.m. PDT ([Bookmark](#))

[Poster](#) presentation: [Point Transformer V3: Simpler Faster Stronger](#)

Wed 19 Jun 5 p.m. PDT – 6:30 p.m. PDT ([Bookmark](#))

[ OpenReview]

[[Paper Metadata for Authors \(e.g. Slide Uploads...\)](#)]