Appendix: Self-Ordering Point Clouds

Pengwan Yang, Cees G. M. Snoek, Yuki M. Asano
University of Amsterdam
p.yang3@uva.nl

1. Further details

Backbone structure. The backbone is vanilla PointNet without any transformation layers to preserve the positional information. It consists of five multi-layer perceptions (MLP). See Figure A for more details.

Class split. To build the zero-shot transfer learning (Table 8 in paper), we remove the overlaping classes between the training data and test data. We report the class splits in Table A.

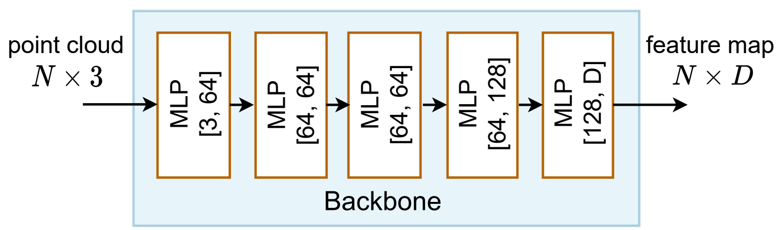

Figure A: **Structure of the backbone.** The backbone consists of five MLPs. N denotes the number of points and D denotes the feature dimension size.

training data		test data	
ModelNet	airplane, bathtub, bed, bookshelf, sofa, bottle, car, chair, cone,curtain, desk, dresser, keyboard, door, glass box, xbox, lamp, laptop, mantel, monitor, person, night stand, plant, radio, tent, stairs, toilet, tv stand, wardrobe, range hood	ModelNet	bench, bowl, cup, plower pot, guitar, piano, sink, vase, stool, table
ModelNet	airplane, bathtub, bed, bookshelf, sofa, bottle, car, chair, cone,curtain, desk, dresser, laptop, door, glass box, keyboard, xbox, lamp, mantel, monitor, person, night stand, plant, tv stand, range hood, radio, stairs, tent, toilet, wardrobe, table, bench, bowl, plower pot, guitar, cup, piano, sink, vase, stool	ShapeNet Core55	rifle, watercraft, loudspeaker, cabinet, display, telephone, bus, faucet, clock, flowerpot, jar, cap bookshelf, knife, train, trash bin, motorbike, bag, pistol, file cabinet, stove, mug, washer, printer, helmet, microwaves, skateboard, tower, camera, basket, can, pillow, mailbox, dishwasher, rocket, birdhouse, earphone, microphone, remote, bicycle
ModelNet	airplane, bathtub, bed, bookshelf, sofa, car, chair, curtain, desk, dresser, laptop, door, glass box, keyboard, xbox, lamp, mantel, monitor, person, night stand, plant, tv stand, range hood, radio, stairs, tent, toilet, wardrobe, table, bench, bowl, plower pot, guitar, cup, piano, sink, vase, stool	3D MNIST	all classes

Table A: **Class split** for transfer learning. We remove the overlapping classes between training data and test data.

	part segmentation (mIoU)				object detection (mAP@0.25)			
eval. #points	2048	256	128	64	4096	512	256	128
FPS	83.7	62.5	43.8	26.0	66.7	50.6	44.2	30.5
Lang *et al*.	83.7	76.3	68.8	60.1	66.7	58.8	53.0	49.2
This paper	83.7	79.8	75.4	69.1	66.7	61.5	59.1	56.7

Table B: **Evaluation for part segmentation on ShaperNet Core55 and object detection on S3DIS.** For evaluation, PointNet is used as the part segmentation network and FCAF3D [1] is used as the object detection network. The metric for part segmentation is mIoU(%) and the metric for object detection is mAP@0.25 (%).

	number of points					
	16	32	64	128	256	1024
PointNet	52.8	73.5	80.2	82.7	85.7	86.2
PointNet++	59.1	79.7	85.2	88.4	90.3	90.7
DGCNN	55.0	71.9	81.4	84.6	88.3	92.0

Table C: **Different base network in evaluation task**. We evaluate on classification for ModelNet40 with different base network. Our self-ordering is effective regardless of the used base network in evaluation, highlighting its general nature.

2. Further results

Evaluation segmentation and object detection. We further assess the effectiveness of the learned ordering on two additional complex tasks: part segmentation and object detection (Table B). The results demonstrate that self-ordering is capable of tackling complex 3D scene understanding tasks.

Different base network in evaluation task. We compare three base networks on ModelNet40 for classification, namely PointNet, PointNet++, and DGCNN [2].The results in Table C show that we obtain effective orderings across all three base networks. For all networks 80% to 90% of the classification scores are maintained using only 3% of the points per cloud.

Qualitative results. We show more qualitative examples in Figure B.

References

[1] Danila Rukhovich, Anna Vorontsova, and Anton Konushin. Fcaf3d: Fully convolutional anchor-free 3d object detection. In *ECCV*, 2022.

[2] 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 Transactions on Graphics*, 2019.

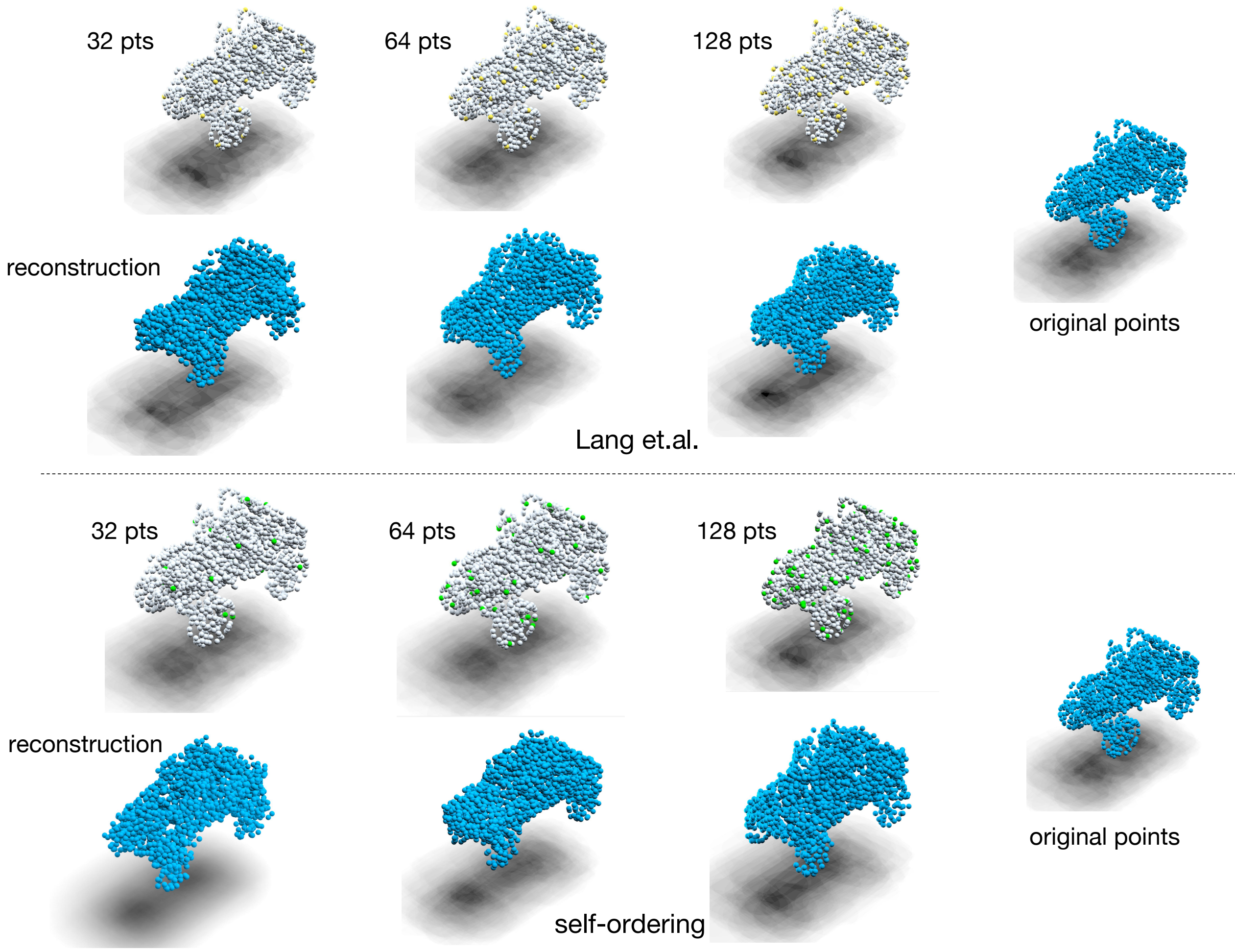

Figure B: **Qualitative results on reconstruction.** We show qualitative examples of reconstruction recovered from only 32 points to 128 points. At upper part is the supervised baseline and at bottom part is our self-ordering. Our reconstructions are closer to the original shape.