Point Cloud Pre-training with Natural 3D Structures –Supplementary Material–

Ryosuke Yamada[1*] Hirokatsu Kataoka[1*] Naoya Chiba[2] Yukiyasu Domae[1] Tetsuya Ogata[1,2]
[1]National Institute of Advanced Industrial Science and Technology (AIST) [2]Waseda University

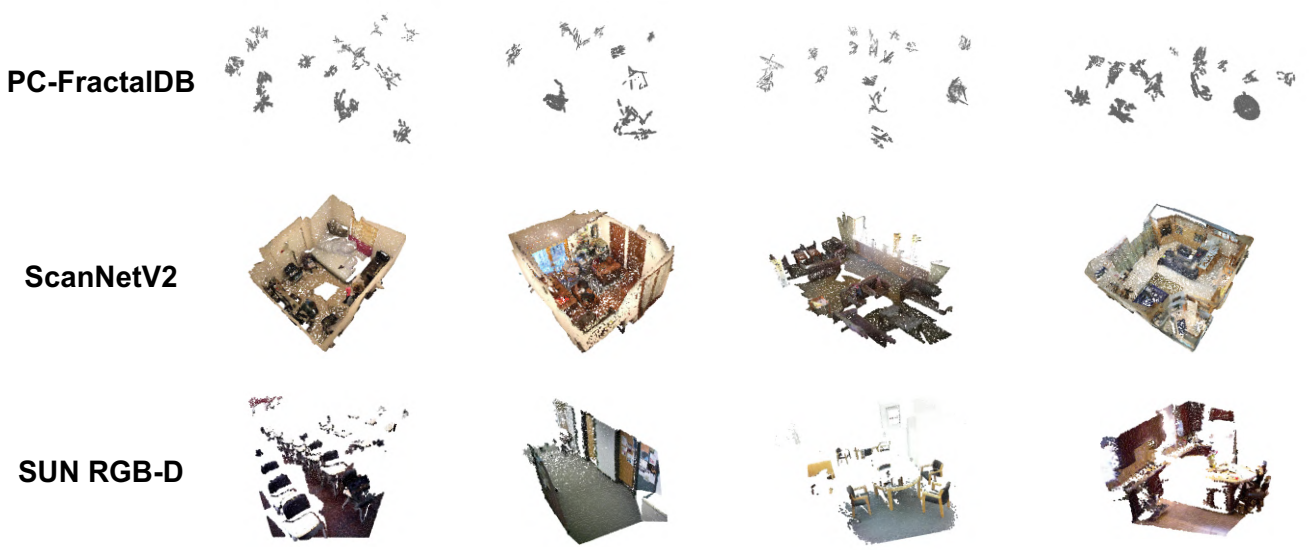

Figure 1. **Examples of 3D fractal scenes in the PC-FractalDB.** The first row is the visualization of automatically generated 3D fractal scenes. In the present study, we do not give RGB color to a 3D fractal model. Therefore a 3D fractal model has only geometry coordinates. The second row is the visualization of 3D scenes in ScanNetV2, and the third row is the visualization of 3D scenes in SUN RGB-D.

Appendix

This supplemental document describes the details of our PC-FractalDB pre-trained model construction in Section A. We show more exploratory study on the variance threshold and the fraction of noise mixed in FractalNoiseMix in Section B. Detailed per-category results on 3D object detection benchmarks are shown in Section C. In addition, Qualitative results of 3D object detection in ScanNetV2 are presented in Section D.

A. Visualization of the PC-FractalDB

In this section, we visualize the PC-FractalDB used for pre-training (see Fig. 1). We compare the samples with 3D scenes randomly sampled from ScanNetV2 [1] and SUN RGB-D [3]. A 3D fractal scene is generated from 3D fractal models only, so it has no background. It is possible to generate a background, but pre-training tends not to succeed. We consider that this because the number of categories has increased compared to the previous 3D datasets. Therefore, Pre-training 3D fractal scenes with a background, we think challenging task in the future.

B. More exploratory study

Effects of variance threshold (see Table 1). This exploratory experiment attempts to clarify the optimal parameters for the variance threshold for defining fractal categories under the PC-FractalDB (1,000 categories, 500 instances, 10,000 scenes) condition. Table 1 showed that 0.15 is better than 0.10 in variance threshold σ. The higher the variance threshold, the more the 3D fractal model tends to have more shape patterns with fractal features, which means that the inter-category variance is significant, and the effect of pre-training is enhanced. Additionally, we found that

Table 1. The variance threshold comparisons for 3D fractal model creation. We also investigated 0.10, 0.15, and 0.20 in the w/ variance adjustment. However, the 0.20 score requires a large amount of time to find a fractal category.

Variance threshold	ScanNetV2 mAP@0.25	SUN RGB-D mAP@0.25
w/ variance (0.10)	59.2	56.8
w/ variance (0.15)	**61.9**	**59.0**
w/ variance (0.20)	N/A	N/A

Table 2. Comparisons in the fractal noise rate of FractalNoiseMix.

Noise rate	ScanNetV2 mAP@0.25	SUN RGB-D mAP@0.25
Fractal noise: 10%	60.3	57.0
Fractal noise: 20%	**61.9**	**59.0**
Fractal noise: 30%	60.3	58.8
Fractal noise: 40%	60.7	57.8

the variance threshold of 0.20 requires a more significant amount of time to definite the fractal category. Therefore, we set the variance threshold σ as 0.15 in the present paper.

Effects of FractalNoiseMix (see Table 2). We need to augment a 3D fractal model in intra-category because there is only one 3D fractal model for each fractal category after defining a fractal category. Therefore, we augment a 3D fractal model in the present study by mixing another category as fractal noise. This experiment explores the most effective fractal noise ratio for FractalNoiseMix under the PC-FractalDB (1,000 categories, 500 instances, 10,000 scenes) condition. Table 2 showed that the fractal noise ratio of 20 % was the best parameter.

C. Per-category result

This section shows the detailed average precision (AP) of each category when the IoU threshold is set to 0.25. as a supplement to the PC-FractalDB pre-training model benchmark. 3D object detection results for ScanNetV2 and SUN RGB-D are shown in Table 3 and Table 4.

As shown in Table 3 and Table 4, we show that our proposed pre-training with the PC-FractalDB can improve the fine-tuning in the 3D detection task with the point cloud. In particular, when the backbone network is PointNet++×2, we confirm the improvement of the average accuracy from RandomRooms in 13 out of 18 categories in ScanNetV2. In addition, we confirm the improvement of the average accuracy from RandomRooms in 7 out of 11 categories in SUN RGB-D.

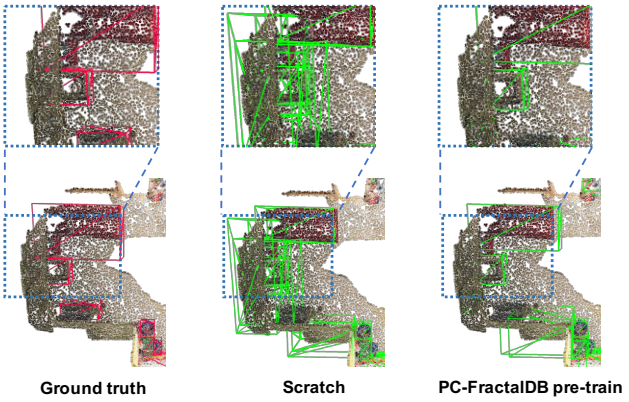

Figure 2. In the figures, we compare point cloud fractal database pre-trained VoteNet with ground truth and VoteNet training from scratch.

D. Qualitative result

In this section, we visualize the detection results of the baseline VoteNet trained from scratch and the pre-trained model using our method on ScanNetV2. As shown in Fig. 2, the PC-FractalDB pre-trained model results in fewer false positives, and the 3D bounding boxes estimated tend to be close to the ground truth than scratch.

References

[1] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5828–5839, 2017. 1

[2] Yongming Rao, Benlin Liu, Yi Wei, Jiwen Lu, Cho-Jui Hsieh, and Jie Zhou. Randomrooms: Unsupervised pre-training from synthetic shapes and randomized layouts for 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3283–3292, 2021. 3

[3] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 567–576, 2015. 1

Table 3. 3D object detection on ScanNetV2.

	cab	bed	chair	sofa	tabl	door	wind	bkshf	pic	cntr	desk	curt	fridg	showr	toil	sink	bath	ofurn	mAP@0.25
Scratch (PointNet++)	36.3	87.4	88.3	86.4	62.1	42.1	36.1	49.4	4.7	56.3	66.0	46.5	48.4	60.1	95.7	48.3	89.4	37.8	57.9
Scratch (SR-UNet)	34.0	76.5	89.1	83.9	57.2	46.7	35.1	45.4	8.0	57.4	65.9	51.1	44.1	56.7	96.3	51.5	81.5	45.8	57.0
RandomRooms (PointNet++) [2]	37.2	87.4	88.9	89.8	61.9	45.3	42.6	53.5	7.8	51.7	67.2	53.5	54.0	66.4	96.8	62.6	92.0	43.6	61.3
PC-FractalDB (PointNet++)	38.0	88.1	88.3	88.8	60.7	51.9	39.3	58.5	9.3	57.2	68.1	56.7	51.7	63.0	97.1	60.1	92.5	44.1	**61.9**
PC-FractalDB (PointNet++×2)	40.4	88.1	88.4	89.1	61.6	53.1	43.1	59.5	15.6	54.4	68.7	53.8	54.8	74.7	99.6	57.3	92.5	46.2	**63.4**
PC-FractalDB (SR-UNet)	29.1	82.7	89.8	83.7	58.5	48.6	38.6	54.2	7.7	61.1	69.4	50.8	46.2	62.5	95.0	57.3	84.9	48.6	59.4

Table 4. 3D object detection on SUN RGB-D.

	bathtub	bed	bookshelf	chair	desk	dresser	night stand	sofa	table	toilet	mAP@0.25
Scratch (PointNet++)	77.6	83.6	28.6	73.7	23.6	25.5	57.9	64.4	48.9	90.1	57.4
Scratch (SR-UNet)	69.2	81.7	29.7	74.7	22.7	24.2	57.5	63.1	48.8	89.3	56.1
RandomRooms (PointNet++) [2]	76.2	83.5	29.2	76.7	25.1	33.2	64.2	63.8	49.0	91.2	59.2
PC-FractalDB (PointNet++)	78.4	85.3	32.8	74.6	26.3	33.5	63.8	62.9	50.3	87.3	**59.4**
PC-FractalDB (PointNet++×2)	79.8	83.6	32.1	75.4	28.5	30.2	67.3	64.4	50.8	88.9	**60.2**
PC-FractalDB (SR-UNet)	71.4	82.3	30.7	75.0	24.7	26.4	58.4	64.3	48.3	89.5	57.1