Cascaded Refinement Network for Point Cloud Completion - Supplementary Material

Xiaogang Wang Marcelo H Ang Jr Gim Hee Lee
National University of Singapore, Singapore
xiaogangw@u.nus.edu {mpeangh,gimhee.lee}@nus.edu.sg

A. Evaluation Metrics

We adopt the Fréchet Point Cloud Distance (FPD)[1] to evaluate the point cloud quality in addition to the Chamfer Distance. Similar to FID, FPD can be considered as the probability measurement between the real and fake results, where it calculates the Gaussian measurement on the point features. Specially, we extracted the global features (1808-dimension) obtained from a point classification module with an accuracy of 98%. We show quantitative comparison on FPD evaluation in Table 1. Here we only calculate FPD on 2048 points since it is not practical to calculate the covariance matrix for a larger point size. The lowest FPD value in Table 1 demonstrates that our generated outputs obtain more realistic probability measurements than other approaches, which verifies the effectiveness of the joint optimization on the adversarial loss and reconstruction loss.

Dataset	Methods		
	PCN [4]	TopNet [3]	Ours
Ours	7.91	9.43	**5.21**
TopNet	7.58	9.47	**5.84**

Table 1. FPD comparison among different methods. The lower the better.

Cat.	Pla.	Cab.	Car	Cha.	Lam.	Sof.	Tab.	Ves.
mIoU (%)	35.09	1.67	10.54	2.90	3.86	4.32	2.62	27.02

Table 2. Mean IoU (mIoU) for each category.

B. Explanation of Mean Shape

Mean shapes are used as a good initialization for the iterative refinement on the point features. We quantify the distinctions between the mean shapes and the testing data for each category by calculating mean IoU [1]. Table 2 shows large distinctions for all categories since majority of values are very small. Although airplane has the highest IoU, the intersection only occupies around 1/3 of the whole object. We set the resolution of voxels to be 48 and 0.5 as the threshold when calculating the mean IoU.

C. Network Architecture Details

We illustrate the details of coarse reconstruction, dense reconstruction and discriminator in this section.

The coarse reconstruction stage consists of three fully-connected layers with the size of 1024, 1024 and 1536 units and a reshaping operation to obtain 512×3 points. For the lifting module in the dense reconstruction stage, a shared MLPs block with $\{128, 64\}$ neurons is used to transfer f_s to the feature of size 64. $\mathcal{C}_C(\cdot)$ and $\mathcal{C}_E(\cdot)$ are two MLPs blocks with $\{64, 128\}$ and $\{64\}$ neurons, respectively. Finally we take another MLPs block which has $\{512, 512, 3\}$ neurons to obtain the complete points. ReLU activation is used for all the convolution layers except for the last one.

Our discriminator consists of patch selection, hierarchical feature integration and value regression.

Patch selection. Similar with PointNet++ [2], we first sample 256 seed points by FPS from the complete point sets. We then adopt the ball query method to find points (an upper limit number of K is set) that are within a radius to the seed points. The three radii are $\{0.1, 0.2, 0.4\}$ and the corresponding Ks are $\{16, 32, 128\}$.

Hierarchical feature integration. The inputs are 256 local regions of points with the data size of $256 \times K \times 3$. Three shared MLPs blocks with $\{16, 16, 32\}$, $\{32, 32, 64\}$ and $\{32, 48, 64\}$ neurons followed by a max-pooling operation are used to extract locally global point features. The output is the concatenation of three features with the size of 256×160.

Value regression. The confidence scores for each local patch are regressed by a MLP with one single neuron. The output data size is 256×1.

[1] https://github.com/seowok/TreeGAN

D. More Qualitative Results

We show more qualitative results for different resolutions on our created datasets in Figures 1 - 2. Figure 3 shows more qualitative results for the dataset of PCN, in which the resolution of output is 16,384. Figures 4 - 5 show the qualitative results for the occluded dataset. Various experiments demonstrate that our method obtain more details and the output points are more evenly distributed.

References

[1] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 605–613, 2017.

[2] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pages 5099–5108, 2017.

[3] Lyne P Tchapmi, Vineet Kosaraju, S. Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[4] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *2018 International Conference on 3D Vision*, pages 728–737. IEEE, 2018.

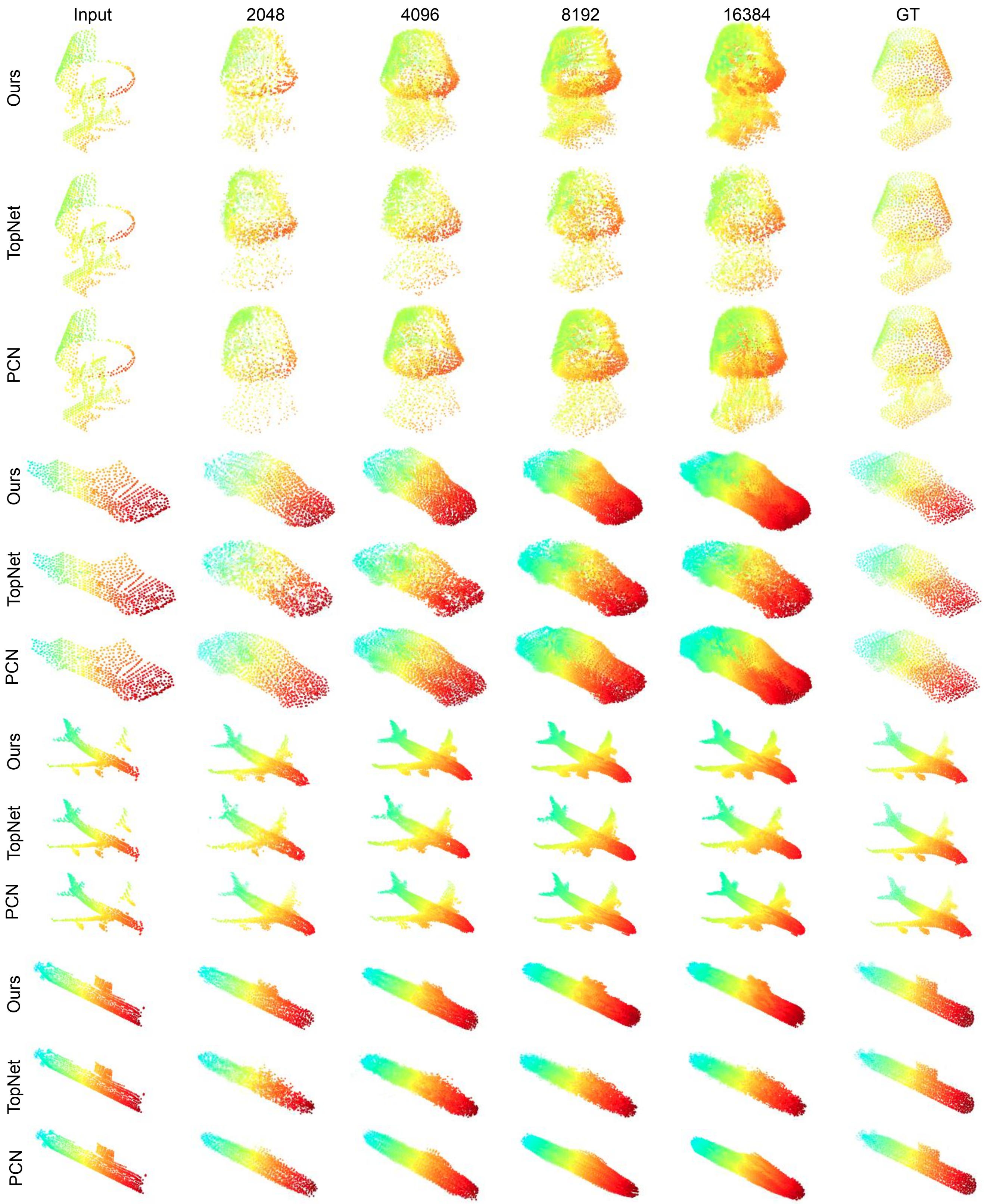

Figure 1. Qualitative comparison results with different resolutions on our created dataset (1/2).

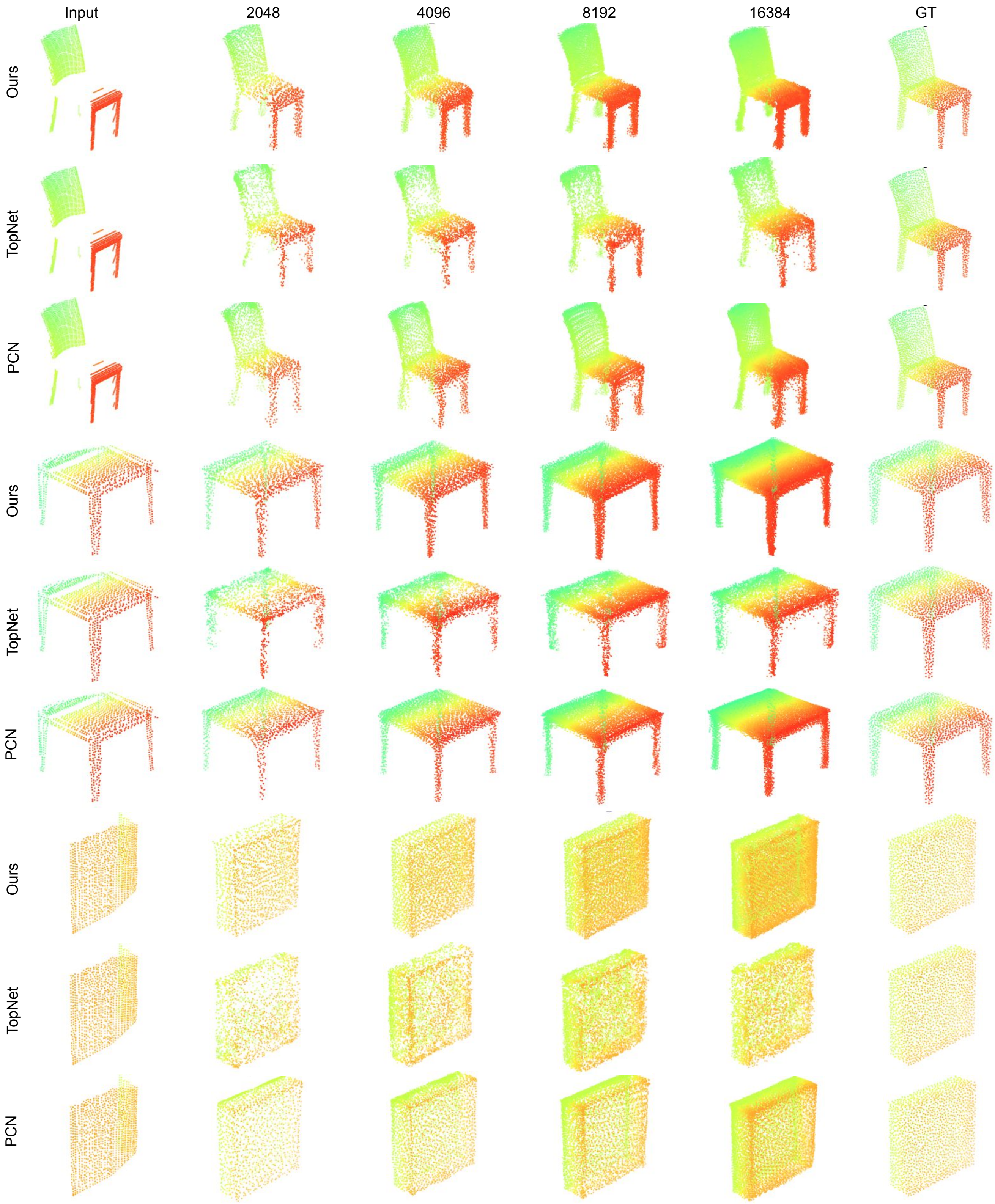

Figure 2. Qualitative comparison results with different resolutions on our created dataset (2/2).

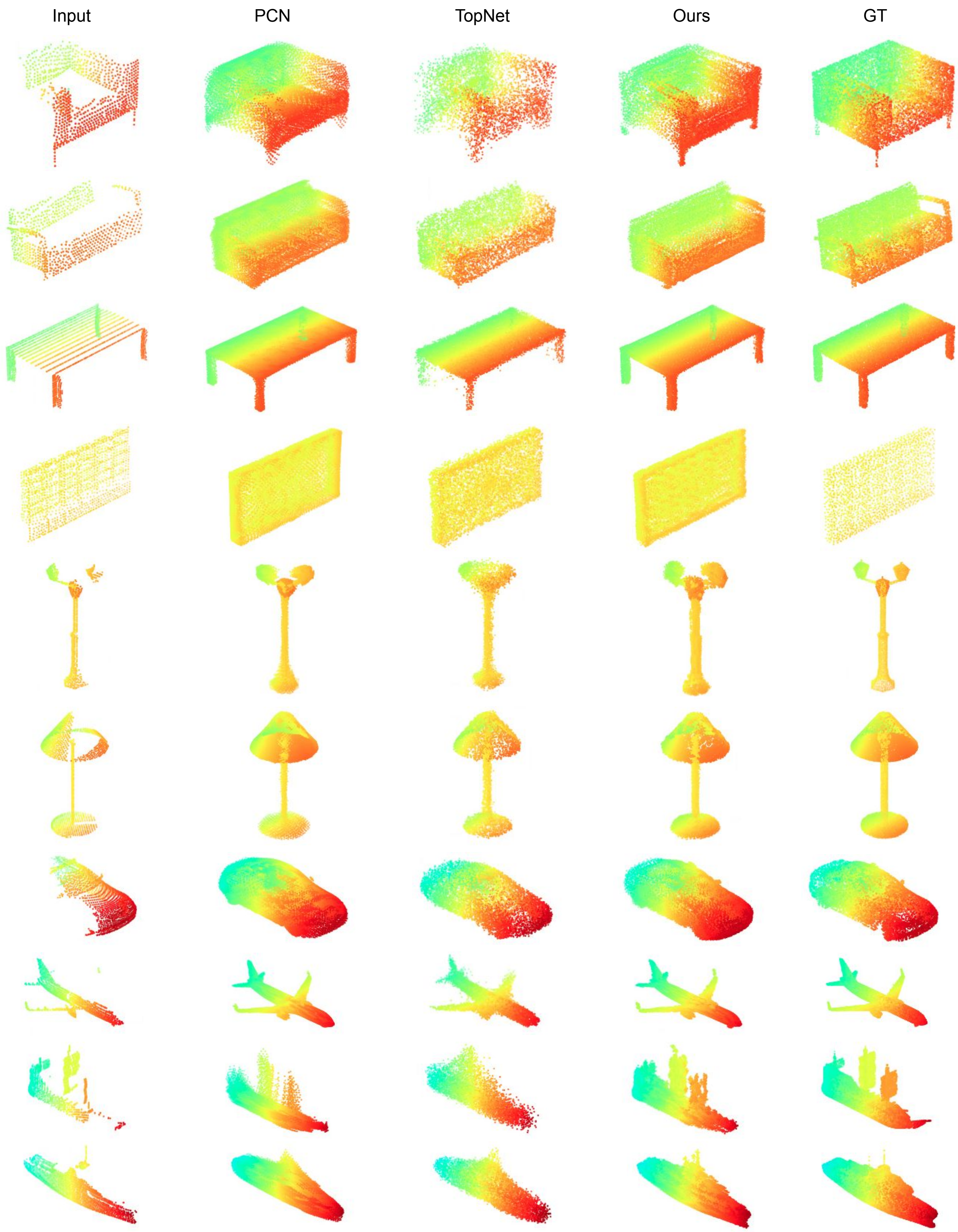

Figure 3. Qualitative comparison results on the dataset of PCN. The resolutions of output and ground truth are 16,384.

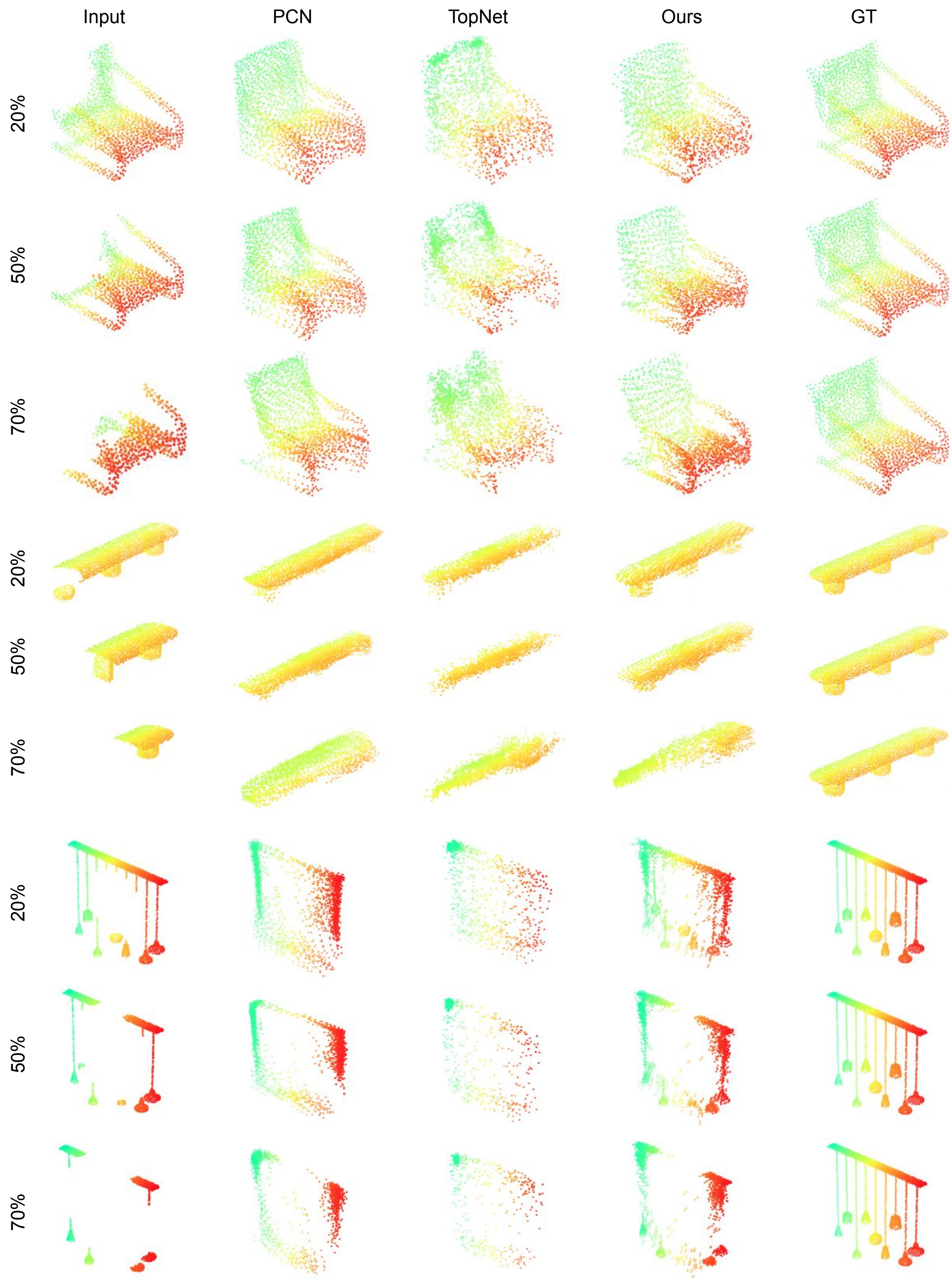

Figure 4. Qualitative comparison results for the occluded dataset. We show different results for three occlusion ratios: 20%, 50% and 70%. The resolutions of output and ground truth are 2048 (1/2).

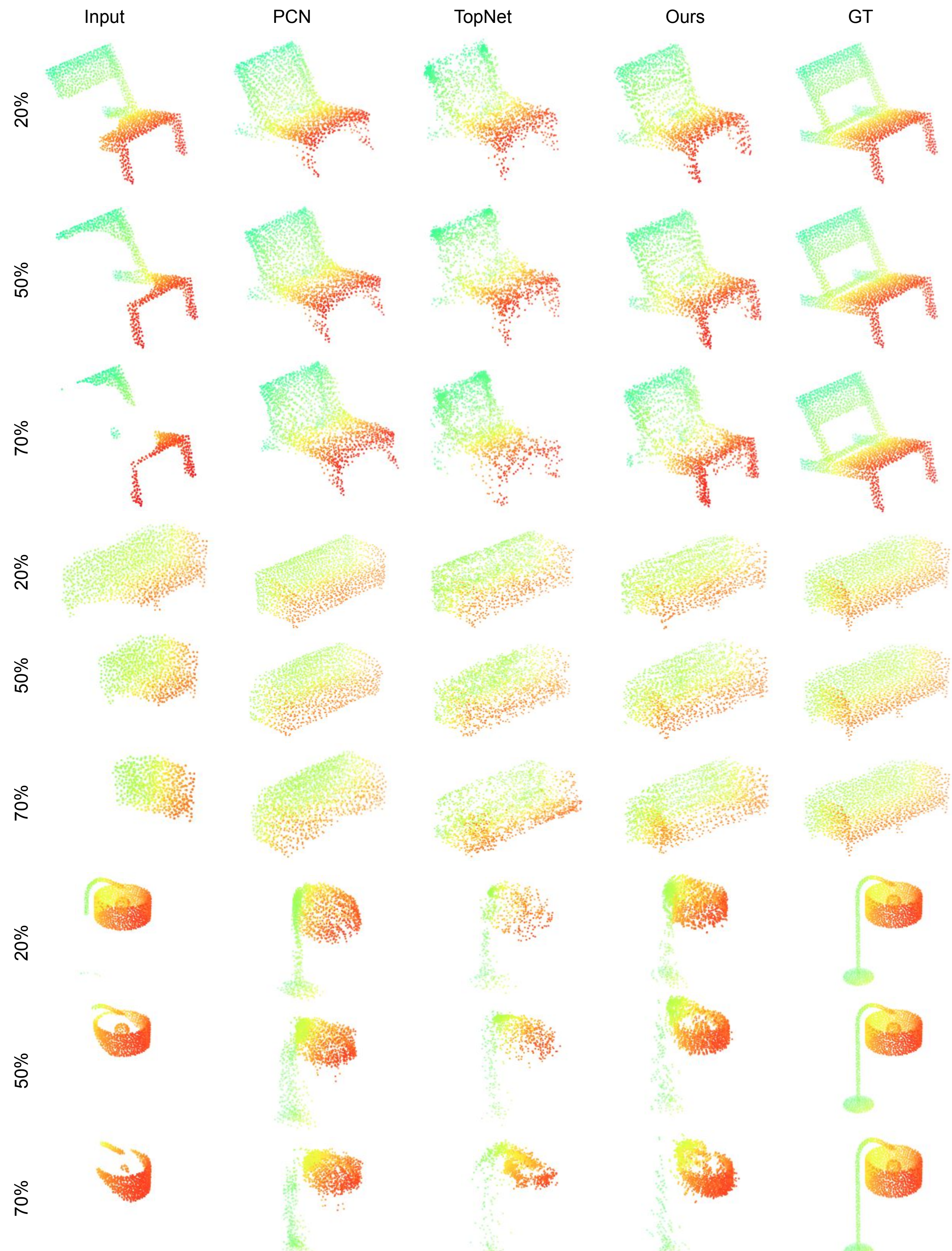

Figure 5. Qualitative comparison results for the occluded dataset. We show different results for three occlusion ratios: 20%, 50% and 70%. The resolutions of output and ground truth are 2048 (2/2).