Supplementary Material for: Point Cloud Completion by Skip-attention Network with Hierarchical Folding

Xin Wen, Tianyang Li, Zhizhong Han, Yu-Shen Liu
{x-wen16,lity16}@mails.tsinghua.edu.cn h312h@umd.edu liuyushen@tsinghua.edu.cn

This is the supplementary material for *Point Cloud Completion by Skip-attention Network with Hierarchical Folding*[1] in CVPR 2020.

1. Implementation Details

Encoder. The encoder of SA-Net adopts three PointNet++ layers. Within each layer, the central point set is first sampled using farthest point sampling algorithm, and for each central point in the set, we use the k nearest neighbor (kNN) algorithm to find its 64 neighbor points. Then, followed by MLPs and max-pooling layer, the features of the central point with its neighbor points are grouped into a single feature vector as the local region feature. The settings of MLPs and each resolution level are detailed in Table 1.

Table 1. Detailed settings of encoder.

Level	FPS points	kNN points	MLPs channels
1	512	64	[64, 64, 128]
2	256	64	[128, 128, 256]
3	-	-	[256, 256, 512]

Decoder. The decoder initially duplicates the input 512-dimensional global representation by 16 times. The following three levels of decoder lift the number of point features to 64, 256 and 2,048, respectively. In folding block, we use three layers of MLPs in up-module to map the M_i^D-dimensional point features that concatenated with 2D grids to 3-dimensional codewords. The channels for MLPs are set to M_i^D, $\frac{M_i^D}{2}$ and 3, respectively. Each layer of MLPs is followed by batch normalization and relu activation except the last layer with 3 channels. We use one layer of MLPs in down-module, which has the same channel as the input. Between each two levels, we use an additional MLPs layer to change the dimension of the point features. The number of point features along with their dimensions are detailed in Table 2.

Table 2. Detailed settings of decoder.

Level	Feature Num.	Feature Dim.
3	64	256
2	256	128
1	2048	3

Training. Given predicted point clouds $\hat{X} = \{\hat{\boldsymbol{x}}_i | i = 1, 2, 3, ..., N\}$ and the ground truth point cloud $X = \{\boldsymbol{x}_i | i = 1, 2, 3, ..., N\}$, the Chamfer distance is defined as:

$$\mathcal{L}_{\mathrm{CD}}(\mathrm{X}, \hat{\mathrm{X}}) = \sum_{\boldsymbol{x} \in \mathrm{X}} \min_{\hat{\boldsymbol{x}} \in \hat{\mathrm{X}}} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\| + \sum_{\hat{\boldsymbol{x}} \in \hat{\mathrm{X}}} \min_{\boldsymbol{x} \in \mathrm{X}} \|\hat{\boldsymbol{x}} - \boldsymbol{x}\|. \quad (1)$$

The Earth Mover distance is defined as:

$$\mathcal{L}_{\mathrm{EMD}}(\hat{\mathrm{X}}, \mathrm{X}) = \min_{\phi: \hat{\mathrm{X}} \to \mathrm{X}} \frac{1}{\hat{\mathrm{X}}} \sum_{\hat{\boldsymbol{x}} \in \hat{\mathrm{X}}} \|\hat{\boldsymbol{x}} - \phi(\hat{\boldsymbol{x}})\|, \quad (2)$$

where ϕ is a bijection that minimizes the average distance between corresponding points in $\hat{X}$ and X. Since finding the optimal ϕ is too computational expensive, we follow the simplified algorithm in PCN [2] to estimate the approximation of EMD during training. The total loss for training is the weighted sum of the CD and EMD, defined as:

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{EMD}} + \lambda \mathcal{L}_{\mathrm{CD}}, \quad (3)$$

where λ is the weight parameter fixed to 10 for the experiments in our paper.

2. More Experiments

Ablation studies on folding block. To verify the effectiveness of each part of folding block in SA-Net, we design 2 variations: (a) The *No-UDU* is a variation that only keeps one up-module and does not use the up-down-up framework. (b) The *No-self* is a variation that remove the self attention in the folding block. The comparison of this two variations with the *Full* model is shown in Table 3.

Table 3. Effect of each part in folding block.

Model	No-UDU	No-self	Full
CD	2.31	2.33	2.18

Ablation studies on λ. λ is the factor to balance between EMD and CD loss, it is set to 10 during the whole experiments. In Table 4, we show the performance of SA-Net under different settings of λ.

Visualization Results. In Figure 1, we visualize more completion results on KITTI dataset. And in Figure 2 and Figure 3, we show more visualized completion results of SA-Net on ShapeNet dataset.

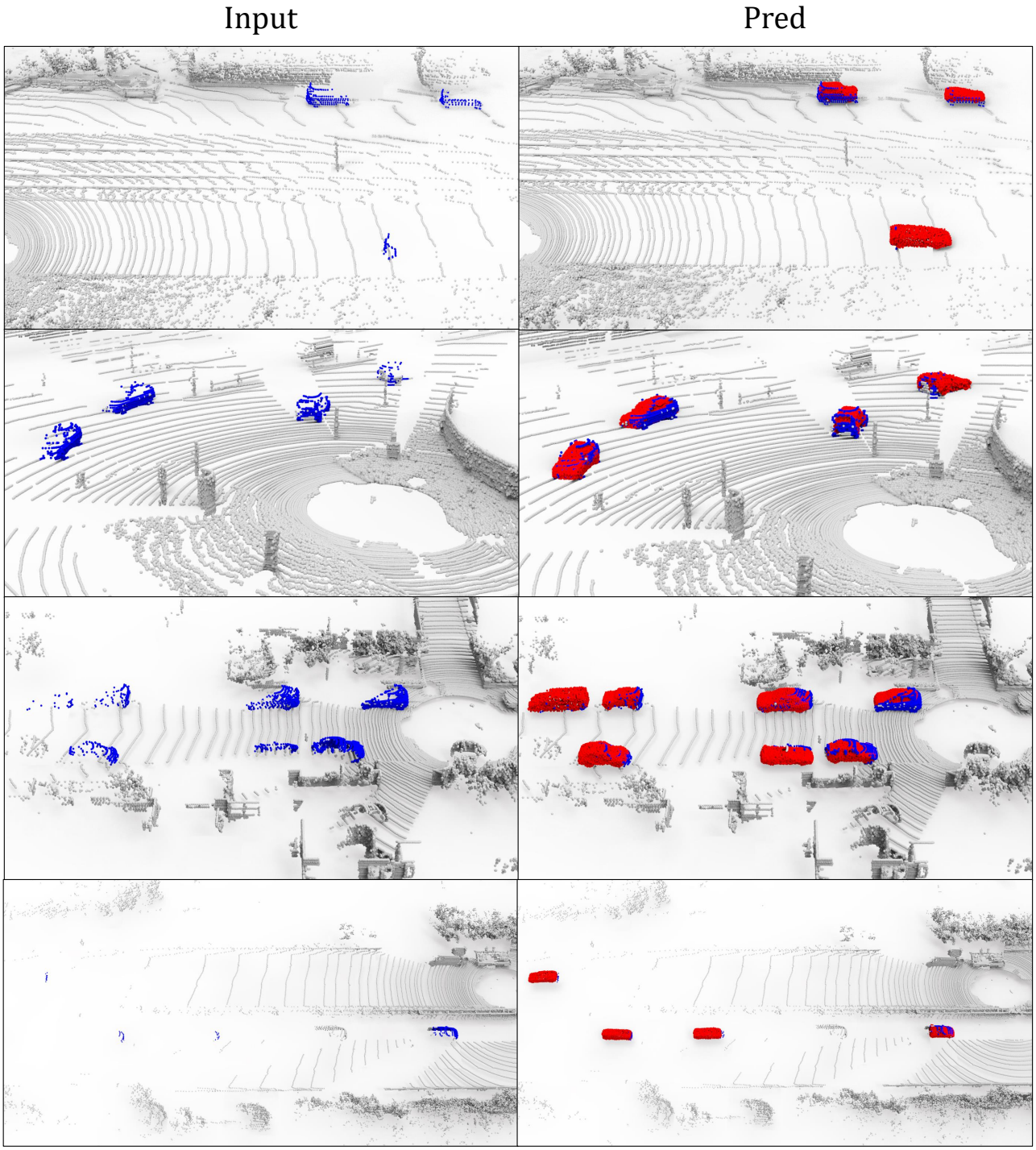

Figure 1. More completion results on KITTI dataset.

Table 4. Effect of λ.

Model	0.1	1.0	10	100
CD	2.30	2.26	2.18	2.24

References

[1] Xin Wen, Tianyang Li, Zhizhong Han, and Yu-Shen Liu. Point cloud completion by skip-attention network with hierarchical folding. In *Proceedings of International Conference on Computer Vision*, 2020. 1

[2] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. PCN: Point completion network. In *2018 International Conference on 3D Vision (3DV)*, pages 728–737. IEEE, 2018. 1

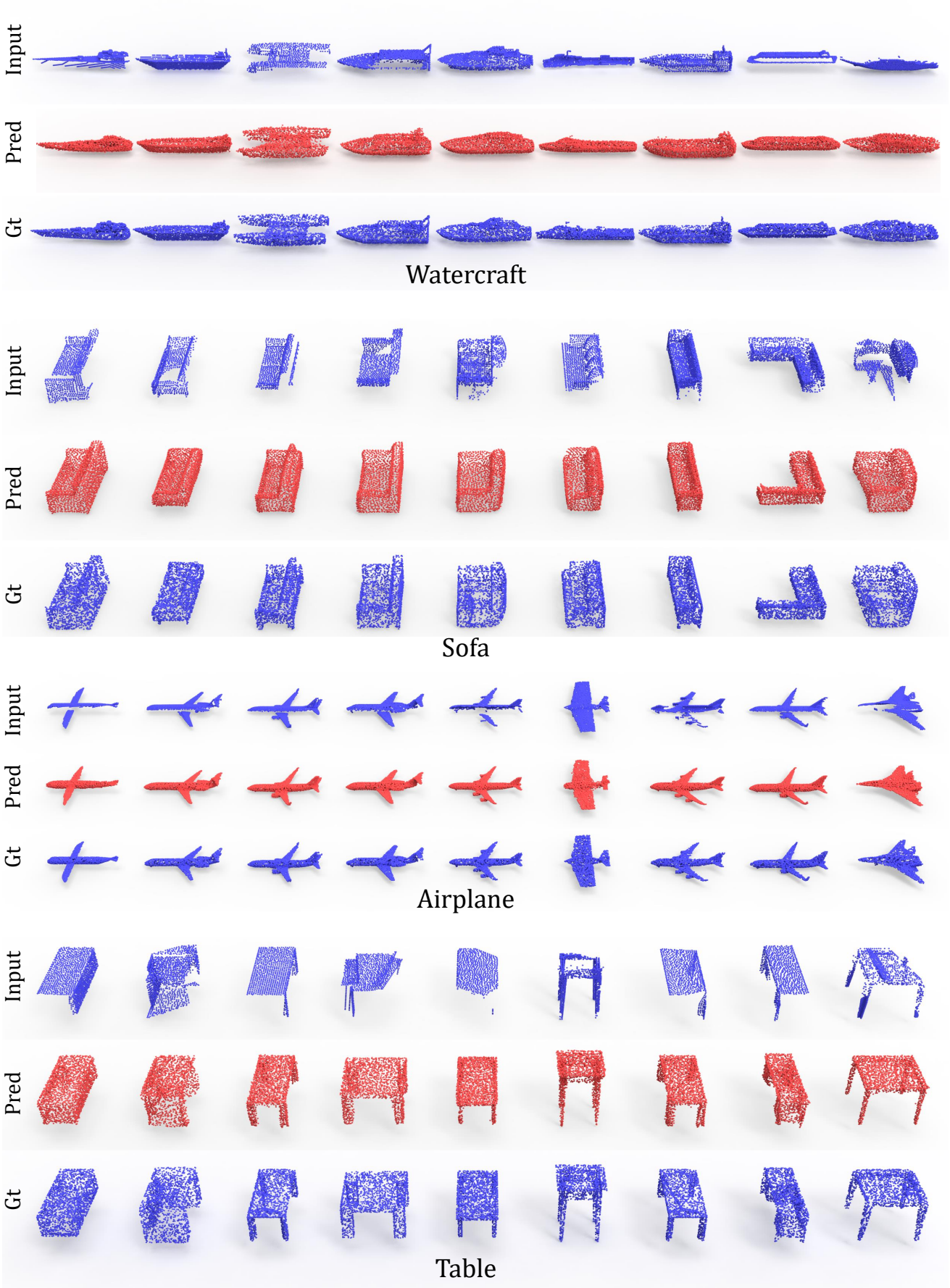

Figure 2. More completion results on ShapeNet dataset.

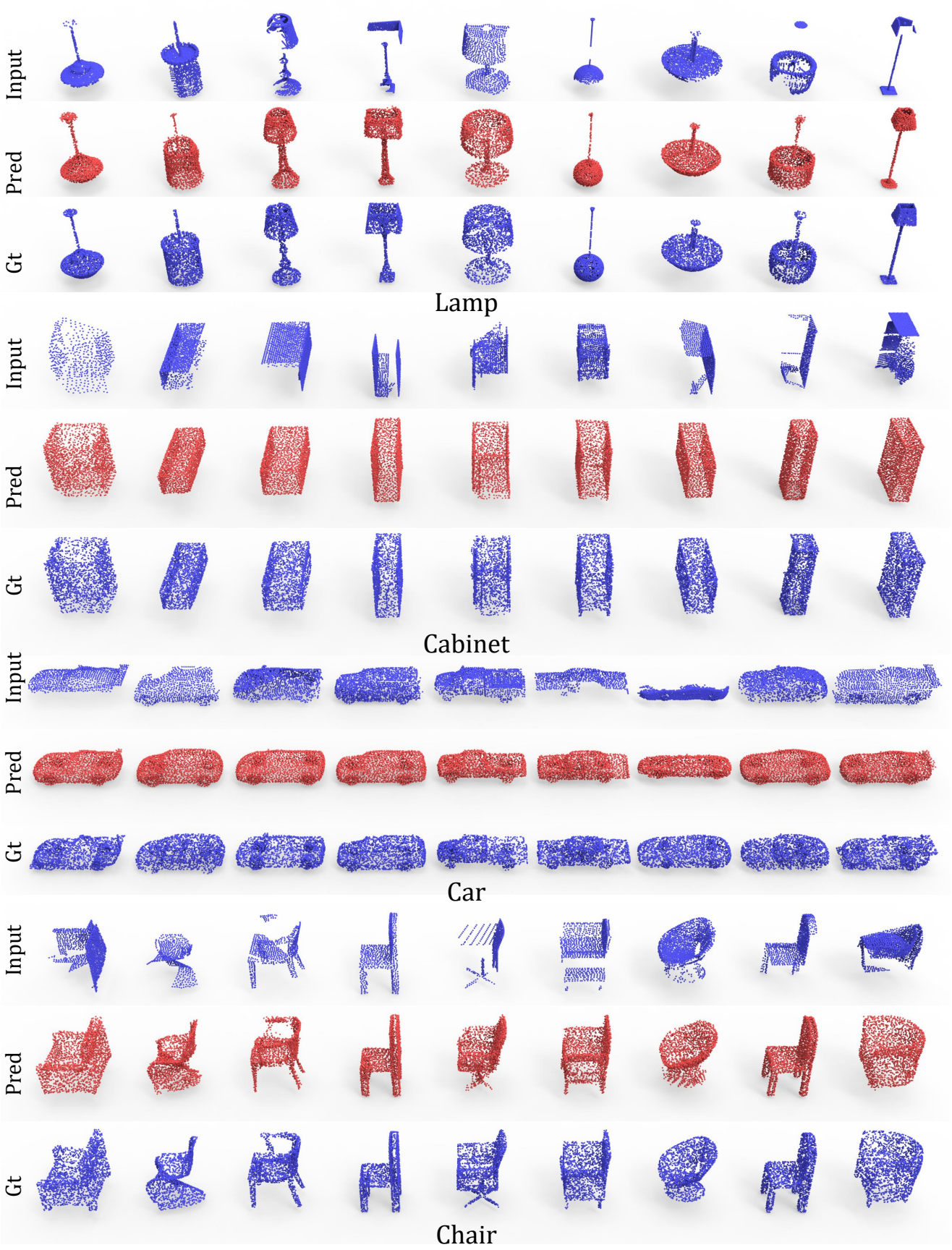

Figure 3. More completion results on ShapeNet dataset.