GRNet: Gridding Residual Network for Dense Point Cloud Completion Supplementary Material

Haozhe Xie[1,2,3][0000−0001−9596−5179], Hongxun Yao[1,2][0000−0003−3298−2574], Shangchen Zhou[4][0000−0001−8201−8877], Jiageng Mao[5][0000−0003−2571−8767], Shengping Zhang[2,6][0000−0001−5200−3420], and Wenxiu Sun[3][0000−0001−5026−8820]

[1]State Key Laboratory of Robotics and System, Harbin Institute of Technology
[2]Faculty of Computing, Harbin Institute of Technology
[3]SenseTime Research [4]Nanyang Technological University
[5]The Chinese University of Hong Kong [6]Peng Cheng Laboratory
https://haozhexie.com/project/grnet

Overview

In this supplementary material, we provide additional information to complement the manuscript. First, we present details of *Gridding*, *Gridding Reverse*, and *Cubic Feature Sampling* (Section 1). Second, we provide additional quantitative results on ShapeNet, Completion3D, and KITTI (Sections 2, 3, and 4). Third, we present additional ablation studies (Section 5). At last, we present more qualitative results compared to other methods (Section 6).

1 More Explanations on Gridding, Gridding Reverse, and Cubic Feature Sampling

1.1 Gridding

According to the manuscript, given a vertex v_i and its neighboring points $p \in \mathcal{N}(v_i)$. The proposed *Gridding* layer computes the corresponding value w_i of this vertex v_i as

$$w_i = \sum_{p \in \mathcal{N}(v_i)} \frac{w(v_i, p)}{|\mathcal{N}(v_i)|} \tag{1}$$

where $|\mathcal{N}(v_i)|$ is the number of neighboring points of v_i and $w(v_i, p)$ is defined as

$$w(v_i, p) = (1 - |x_i^v - x|)(1 - |y_i^v - y|)(1 - |z_i^v - z|) \tag{2}$$

Based on Equations 1 and 2, the partial derivative with respect to x can be calculated as follows

$$\frac{\partial w_i}{\partial x} = \begin{cases} -\frac{1}{|\mathcal{N}(v_i)|} \sum_{p \in \mathcal{N}(v_i)} (1 - |y_i^v - y|)(1 - |z_i^v - z|), & x > x_i^v \\ \frac{1}{|\mathcal{N}(v_i)|} \sum_{p \in \mathcal{N}(v_i)} (1 - |y_i^v - y|)(1 - |z_i^v - z|), & x \leq x_i^v \end{cases} \tag{3}$$

where x and x_i^v are the x-coordinates of the point p and vertex v_i, respectively. Similarly, the partial derivative with respect to y and z can be calculated as follows

$$\frac{\partial w_i}{\partial y} = \begin{cases} -\frac{1}{|\mathcal{N}(v_i)|}\sum_{p\in\mathcal{N}(v_i)}(1-|x_i^v-x|)(1-|z_i^v-z|), & y > y_i^v \\ \frac{1}{|\mathcal{N}(v_i)|}\sum_{p\in\mathcal{N}(v_i)}(1-|x_i^v-x|)(1-|z_i^v-z|), & y \leq y_i^v \end{cases} \tag{4}$$

$$\frac{\partial w_i}{\partial z} = \begin{cases} -\frac{1}{|\mathcal{N}(v_i)|}\sum_{p\in\mathcal{N}(v_i)}(1-|x_i^v-x|)(1-|y_i^v-y|), & z > z_i^v \\ \frac{1}{|\mathcal{N}(v_i)|}\sum_{p\in\mathcal{N}(v_i)}(1-|x_i^v-x|)(1-|y_i^v-y|), & z \leq z_i^v \end{cases} \tag{5}$$

where y and y_i^v are the y-coordinates of the point p and vertex v_i, respectively. z and z_i^v are the z-coordinates of the point p and vertex v_i, respectively.

1.2 Gridding Reverse

Point Coordinates Normalization. *Gridding Reverse* generates point $p_i^c = (x_i^c, y_i^c, z_i^c)$ for the i-th grid cell by a weighted combination of eight vertices $\{v_\theta|\theta \in \Theta^i\}$ and the corresponding values $\{w'_\theta|\theta \in \Theta^i\}$ in this cell, which is calculated as

$$p_i^c = \frac{\sum_{\theta\in\Theta^i} w'_\theta v_\theta}{\sum_{\theta\in\Theta^i} w'_\theta} \tag{6}$$

where $\sum_{\theta\in\Theta^i} w'_\theta \neq 0$ and $\Theta^i = \{\theta_j^i\}_{j=1}^8$ represents the index set of vertices of this 3D grid cell. Let $(x_\theta^v, y_\theta^v, z_\theta^v)$ be the coordinate of the vertex v_θ, where $x_\theta^v, y_\theta^v, z_\theta^v \in \{-\frac{N}{2}, -\frac{N}{2}+1, \dots, -\frac{N}{2}-1\}$ and N is the resolution of the 3D grid. The x-, y-, and z- coordinates of p_i^c is calculated as

$$x_i^c = \frac{\sum_{\theta\in\Theta^i} w'_\theta x_\theta^v}{\sum_{\theta\in\Theta^i} w'_\theta} \tag{7}$$

$$y_i^c = \frac{\sum_{\theta\in\Theta^i} w'_\theta y_\theta^v}{\sum_{\theta\in\Theta^i} w'_\theta} \tag{8}$$

$$z_i^c = \frac{\sum_{\theta\in\Theta^i} w'_\theta z_\theta^v}{\sum_{\theta\in\Theta^i} w'_\theta} \tag{9}$$

Since the coordinate $(x_i^{gt}, y_i^{gt}, z_i^{gt})$ of the point in the ground truth point cloud satisfies $-1 < x_i^{gt}, y_i^{gt}, z_i^{gt} < 1$. The coordinates of the point p_i^c are normalized to $(-1, 1)$ by dividing $-\frac{N}{2}$.

Backward of Gridding Reverse. The partial derivative with respect to w'_θ can be calculated as

$$\begin{aligned} \frac{\partial x_i^c}{\partial w'_\theta} &= \frac{x_\theta^v}{\sum_{\theta \in \Theta^i} w'_\theta} - \frac{\sum_{\theta \in \Theta^i} w'_\theta x_\theta^v}{\left(\sum_{\theta \in \Theta^i} w'_\theta\right)^2} \\ &= \frac{x_\theta^v}{\sum_{\theta \in \Theta^i} w'_\theta} - \frac{1}{\sum_{\theta \in \Theta^i} w'_\theta} \cdot x_i^c \\ &= \frac{x_\theta^v - x_i^c}{\sum_{\theta \in \Theta^i} w'_\theta} \end{aligned} \tag{10}$$

Similarly,

$$\frac{\partial y_i^c}{\partial w'_\theta} = \frac{y_\theta^v - y_i^c}{\sum_{\theta \in \Theta^i} w'_\theta} \tag{11}$$

$$\frac{\partial z_i^c}{\partial w'_\theta} = \frac{z_\theta^v - z_i^c}{\sum_{\theta \in \Theta^i} w'_\theta} \tag{12}$$

1.3 Cubic Feature Sampling

Point Coordinates Normalization. *Cubic Feature Sampling* aggregates features $F^c = \{f_i^c\}_{i=1}^m$ of the coarse point cloud $P^c = \{p_i^c\}_{i=1}^m$ from the 3D feature map $\mathcal{F} = \{f_i^v\}_{i=1}^{t^3}$, where $f_i^c, f_i^v \in \mathbb{R}^c$, c is the number of channels of $\mathcal{F}$, m is the number of points in the coarse point cloud, and t is the resolution of $\mathcal{F}$. According to the manuscript, the features f_i^c for $p_i^c = (x_i^c, y_i^c, z_i^c)$ are calculated as

$$f_i^c = [f_{\theta_1^i}^v, f_{\theta_2^i}^v, \dots, f_{\theta_8^i}^v] \tag{13}$$

where $\{f_{\theta_j^i}^v\}_{j=1}^8$ denotes the features of eight vertices of the i-th 3D gird cell where p_i^c lies in. Specifically, the coordinates of the eight vertices $\{(x_{\theta_j^i}^v, y_{\theta_j^i}^v, z_{\theta_j^i}^v)\}_{j=1}^8$ satisfy $x_{\theta_j^i}^v \in \{\lfloor \frac{t}{2} x_i^c \rfloor, \lceil \frac{t}{2} x_i^c \rceil\}$, $y_{\theta_j^i}^v \in \{\lfloor \frac{t}{2} y_i^c \rfloor, \lceil \frac{t}{2} y_i^c \rceil\}$, and $z_{\theta_j^i}^v \in \{\lfloor \frac{t}{2} z_i^c \rfloor, \lceil \frac{t}{2} z_i^c \rceil\}$, respectively.

Backward of Cubic Feature Sampling. During backward propagation, the partial derivative with respect to $f_{\theta_j^i}^v$ can be presented as

$$\frac{\partial f_{i,j}^c}{\partial f_{\theta_j^i}^v} = 1 \tag{14}$$

where $j \in \{1, 2, \dots, 8\}$ and $f_{i,j}^c$ denotes the j-th element in f_i^c.

Since $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ is not differentiable, the partial derivatives with respect to x_i^c, y_i^c, and z_i^c are 0 [1], which can be formulated as follows:

$$\frac{\partial f_{i,j}^c}{\partial x_i^c} = 0 \tag{15}$$

$$\frac{\partial f_{i,j}^c}{\partial y_i^c} = 0 \tag{16}$$

Table 1. Results of point cloud completion on ShapeNet compared using the Chamfer Distance (CD) with L1 norm computed on 16,384 points and multiplied by 10^3. The best results are highlighted in bold.

Methods	Airplane	Cabinet	Car	Chair	Lamp	Sofa	Table	Watercraft	Overall
AtlasNet [2]	6.366	11.943	10.105	12.063	12.369	12.990	10.331	10.607	10.847
PCN [7]	**5.502**	10.625	**8.696**	10.998	11.339	11.676	8.590	9.665	9.636
FoldingNet [6]	9.491	15.796	12.611	15.545	16.413	15.969	13.649	14.987	14.308
TopNet [4]	7.614	13.311	10.898	13.823	14.439	14.779	11.224	11.124	12.151
MSN [3]	5.596	11.963	10.776	10.620	10.712	11.895	8.704	9.485	9.969
GRNet	6.450	**10.373**	9.447	**9.408**	**7.955**	**10.512**	**8.444**	**8.039**	**8.828**

$$\frac{\partial f_{i,j}^c}{\partial z_i^c} = 0 \tag{17}$$

2 Additional Quantitative Results on ShapeNet

According to the manuscript, the Chamfer Distance is with L2 norm. However, PCN [7] adopts the Chamfer Distance with L1 norm as an evaluation metric, which can be formulated as follows

$$\text{CD} = \frac{1}{2}\left(\frac{1}{n_{\mathcal{T}}}\sum_{t\in\mathcal{T}}\min_{r\in\mathcal{R}}||t-r|| + \frac{1}{n_{\mathcal{R}}}\sum_{r\in\mathcal{R}}\min_{t\in\mathcal{T}}||t-r||\right) \tag{18}$$

where $\mathcal{T} = \{(x_i, y_i, z_i)\}_{i=1}^{n_{\mathcal{T}}}$ is the ground truth and $\mathcal{R} = \{(x_i, y_i, z_i)\}_{i=1}^{n_{\mathcal{R}}}$ is the reconstructed point set being evaluated. $n_{\mathcal{T}}$ and $n_{\mathcal{R}}$ are the numbers of points of $\mathcal{T}$ and $\mathcal{R}$, respectively.

Table 1 shows the results of point cloud completion using the Chamfer Distance calculated with Equation 18. The values of PCN are **exactly** the same as Table 4 in the original paper [1].

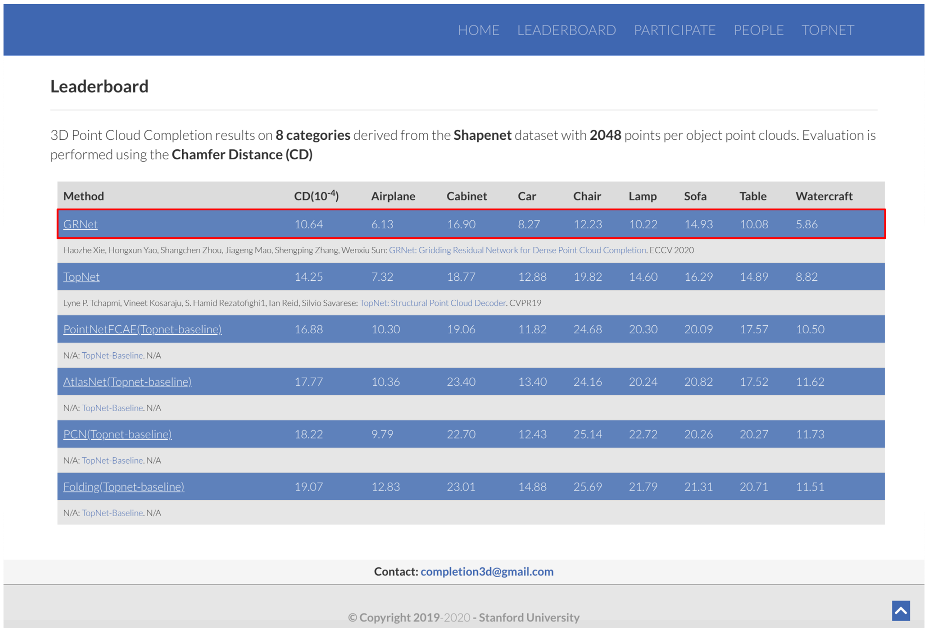
3 Quantitative Results on Completion3D

Figure 1 is the screenshot of the leaderboard results on the Completion3D benchmark, which is available online at https://completion3d.stanford.edu/results.

4 Additional Quantitative Results on KITTI

PCN [7] uses the Fidelity Distance (FD) and Minimal Matching Distance (MMD) as evaluation metrics for KITTI. FD is the average distance from each point in the input to its nearest neighbor in the output, which can be defined as follows

[1] https://arxiv.org/pdf/1808.00671



HOME LEADERBOARD PARTICIPATE PEOPLE TOPNET

Leaderboard

3D Point Cloud Completion results on **8 categories** derived from the **Shapenet** dataset with **2048** points per object point clouds. Evaluation is performed using the **Chamfer Distance (CD)**

Method	CD(10^{-4})	Airplane	Cabinet	Car	Chair	Lamp	Sofa	Table	Watercraft
GRNet	10.64	6.13	16.90	8.27	12.23	10.22	14.93	10.08	5.86
Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, Wenxiu Sun: GRNet: Gridding Residual Network for Dense Point Cloud Completion. ECCV 2020									
TopNet	14.25	7.32	18.77	12.88	19.82	14.60	16.29	14.89	8.82
Lyne P. Tchapmi, Vineet Kosaraju, S. Hamid Rezatofighi1, Ian Reid, Silvio Savarese: TopNet: Structural Point Cloud Decoder. CVPR19									
PointNetFCAE(Topnet-baseline)	16.88	10.30	19.06	11.82	24.68	20.30	20.09	17.57	10.50
N/A: TopNet-Baseline. N/A									
AtlasNet(Topnet-baseline)	17.77	10.36	23.40	13.40	24.16	20.24	20.82	17.52	11.62
N/A: TopNet-Baseline. N/A									
PCN(Topnet-baseline)	18.22	9.79	22.70	12.43	25.14	22.72	20.26	20.27	11.73
N/A: TopNet-Baseline. N/A									
Folding(Topnet-baseline)	19.07	12.83	23.01	14.88	25.69	21.79	21.31	20.71	11.51
N/A: TopNet-Baseline. N/A									

Contact: completion3d@gmail.com

© Copyright 2019-2020 - Stanford University

Fig. 1. The screenshot of the Completion3D benchmark results. Available online at https://completion3d.stanford.edu/results
.

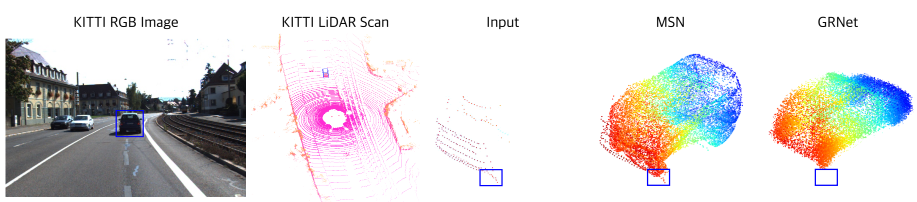


Fig. 2. The clutters in the KITTI LiDAR Scan, as shown in the blue bounding box. Compared to MSN, GRNet recovers the complete point cloud while removing the clutters in the input point cloud.

$$\mathrm{FD} = \frac{1}{n_{\mathcal{I}}} \sum_{i \in \mathcal{I}} \min_{r \in \mathcal{R}} ||i - r||_2^2 \tag{19}$$

where $\mathcal{I}$ denotes the input point cloud. MMD is the Chamfer Distance (CD) between the output and the car point cloud from ShapeNet that is the closest to the output point cloud in terms of CD. The Fidelity and MMD on KITTI of the compared methods are shown in Table 2.

However, both FD and MMD are not suitable metrics for KITTI. As shown in Figure 2, real-world LiDAR scans usually contain clutters which should be

Table 2. Results of point cloud completion on KITTI compared using Fidelity Distance (CD) and Minimal Matching Distance (MMD) computed on 16,384 points. Note that both FD and MMD are with L2 norm. The best results are highlighted in bold.

Methods	FD ($\times 10^3$)	MMD ($\times 10^3$)
AtlasNet [2]	1.759	2.108
PCN [7]	2.235	1.366
FoldingNet [6]	7.467	**0.537**
TopNet [4]	5.354	0.636
MSN [3]	**0.434**	2.259
GRNet	0.816	0.568

Table 3. The Chamfer Distance (CD) and F-Score@1% on ShapeNet with different numbers of points sampled from the coarse point cloud. The best results are highlighted in bold.

# Points	CD ($\times 10^{-4}$)	F-Score@1%
1024	2.775	0.697
2048	**2.723**	**0.708**
4096	2.832	0.681

removed in the recovered point cloud. MSN [3] incorporates the minimum density sampling (MDS) to preserve the structure of the input point cloud. Although MSN outperforms other methods in terms of FD, the clutters in the input point cloud are also preserved. MMD s measures how much the output resembles the cars in ShapeNet. However, cars from ShapeNet cannot cover all types of cars in the real-world.

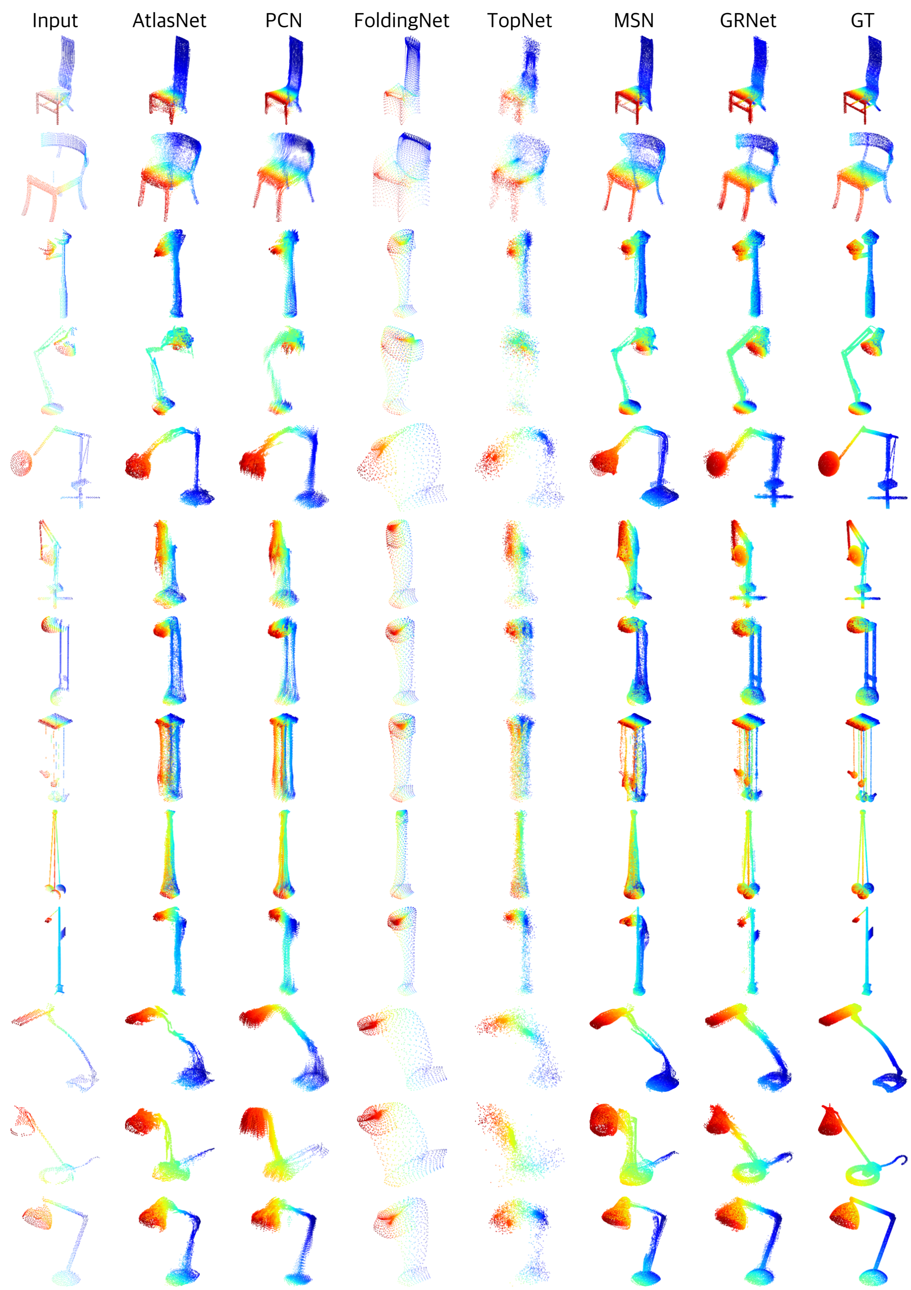
5 Additional Ablation Studies

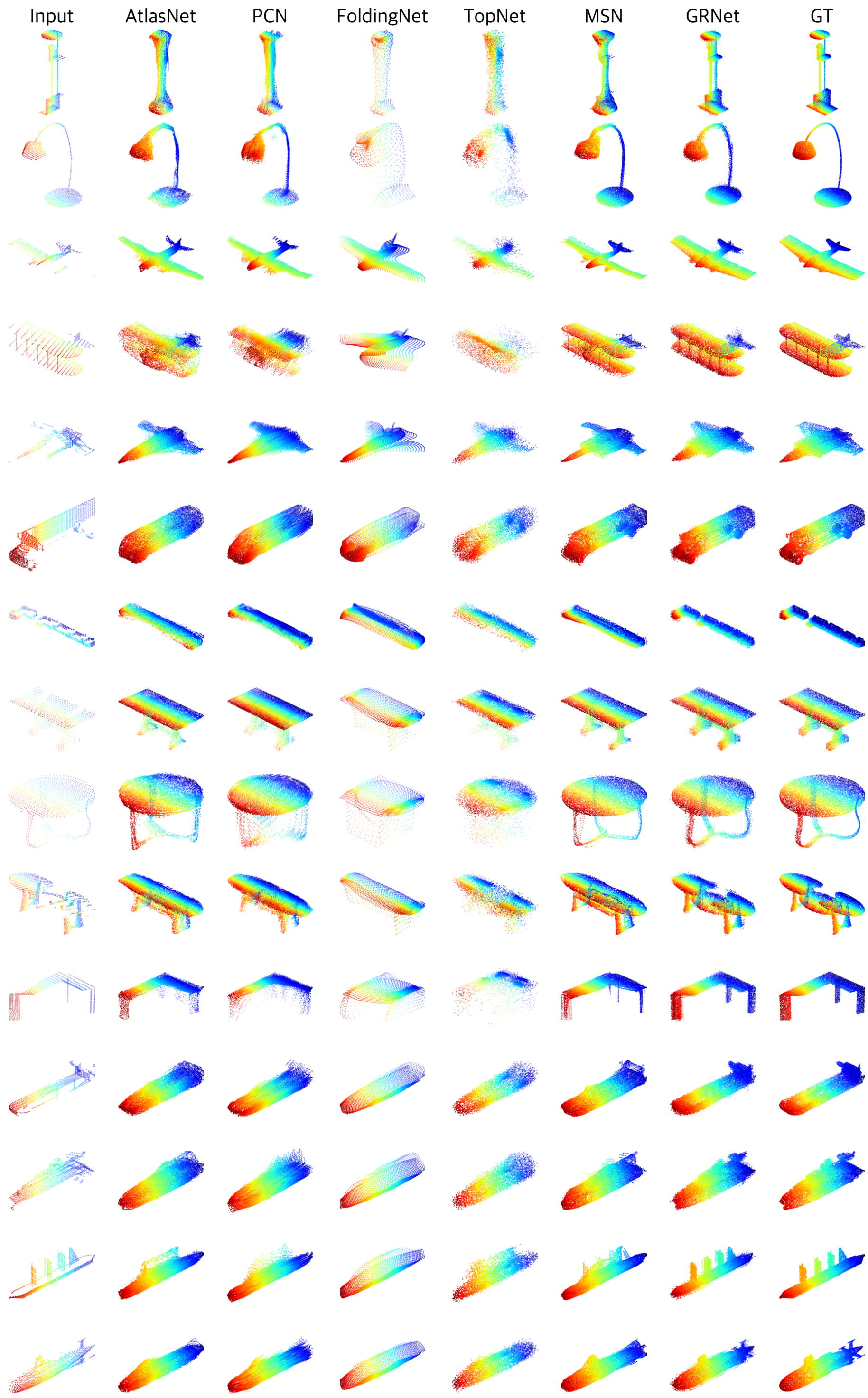
Number of Sampling Points. *Gridding Reverse* generates a coarse point cloud from a 3D grid. We randomly sample 2,048 points from the coarse point cloud to generate a point cloud containing a fixed number of points for the following MLP. Table 3 shows the Chamfer Distance (CD) and F-Score@1% with different numbers of points sampled.

Experimental results indicate that sampling 2,048 points from the coarse point clouds archives the best performance in terms of CD and F-Score. The coarse point cloud of an object usually contains about 3,000-4,000 points, over-sampling 4,096 points from the coarse point cloud leads to redundant information in the sampled point cloud. Sampling 1,024 points from the coarse point cloud may lose too much information for the subsequent processing.

6 Qualitative Comparisons

In this section, we provide more visual comparisons with the state-of-the-art methods [7,2,6,4,3] for point cloud completion on the ShapeNet dataset [5].




Input
AtlasNet
PCN
FoldingNet
TopNet
MSN
GRNet
GT

References

1. Esser, S.K., McKinstry, J.L., Bablani, D., Appuswamy, R., Modha, D.S.: Learned step size quantization. In: ICLR 2020 (2020) 3
2. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: A papier-mâché approach to learning 3D surface generation. In: CVPR 2018 (2018) 4, 6, 7
3. Liu, M., Sheng, L., Yang, S., Shao, J., Hu, S.M.: Morphing and sampling network for dense point cloud completion. In: AAAI 2020 (2020) 4, 6, 7
4. Tchapmi, L.P., Kosaraju, V., Rezatofighi, H., Reid, I.D., Savarese, S.: TopNet: Structural point cloud decoder. In: CVPR 2019 (2019) 4, 6, 7
5. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: A deep representation for volumetric shapes. In: CVPR 2015 (2015) 7
6. Yang, Y., Feng, C., Shen, Y., Tian, D.: FoldingNet: Point cloud auto-encoder via deep grid deformation. In: CVPR 2018 (2018) 4, 6, 7
7. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: PCN: point completion network. In: 3DV 2018 (2018) 4, 6, 7