Variational Relational Point Completion Network
– Supplementary Material –

A. Overview

In this supplementary material, we provide in-depth method analysis (Sec. B), inference details (Sec. C), detailed dataset comparisons (Sec. G), comprehensive ablation studies (Sec. E), resource usages (Sec. F), and the user study on real scans (Sec. D). Qualitative results with different settings are shown in the corresponding sections.

B. Analysis

B.1. Variational Modeling

Inspired by [13], our VMNet consists of two parallel paths: 1) a reconstruction path, and 2) a completion path. Both two paths follow similar variational auto-encoder structures, which generates complete point clouds using embedded global features and predicted distributions. During training, the encoded distributions (posterior) for incomplete point clouds (completion path) are guided by the encoded distributions (prior) for complete point clouds (reconstruction path). In this way, we mitigate the domain gap between the posterior and the prior distributions by regularizing the posterior distribution to approach the prior distribution. Consequently, the learned smooth complete shape priors are incorporated into our shape completion process. During inference, we only use the completion path. We randomly generate a sample from the learned posterior distribution $p_\psi(\mathbf{z_g}|\mathbf{X})$ for shape completion. Theoretically, diverse plausible coarse completions can be generated by using different samples from $p_\psi(\mathbf{z_g}|\mathbf{X})$. However, we observe similar predicted coarse completions for different samples (see Fig. 1). In other words, different samples do not influence our completion results. According to [13], employing a generative adversarial learning scheme can highly increase the shape completion diversity, which we leave as a future research direction.

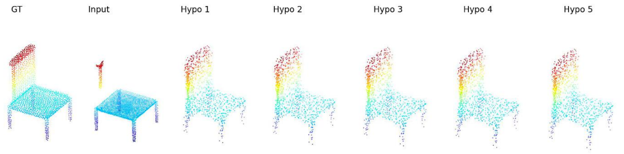

Figure 1: **Coarse Completion Results by Different Samples.** We can observe that the generated different hypotheses for coarse shape skeletons are similar with each other.

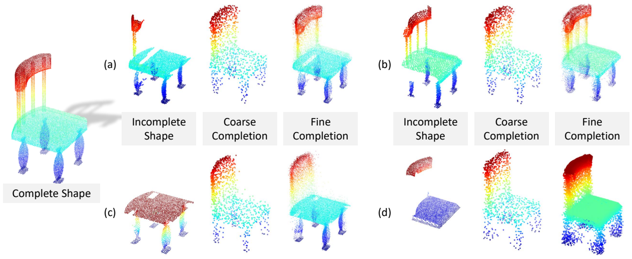

Figure 2: **Observing different parts of the chair, the VRCNet generates different complete chairs based on the partial observations and predicted shape skeletons.** In (a) and (b), the VRCNet predicts complete chair shapes by learning the shape geometrical symmetry from the partial observations. Both (c) and (d) show the incomplete shapes with large missing ratios, and the VRCNet predicts the fine complete shapes based on the coarse complete shapes.

B.2. Local Point Relation Learning

Relation operations [2, 12] (also known as self-attention operations) adaptively learn a meaningful compositional structure that is used to predict adaptive weights by exploiting relations among local elements. Comparing to conventional convolution operations that use fixed weights, relation operations adapt aggregation weights based on the composability of local elements. Motivated by the success of using relation operations in natural language process and image applications, we expand and use relation operations to learn point relations in neighboring points for point cloud completion. Previous methods [11, 4, 9] preserve observed local details from the incomplete point clouds in their completion results by learning local point features. However, they cannot generate fine-grained complete shapes for those missing parts. Consequently, their completion results often have high-quality observed shape parts and low-quality missing shape parts. In contrast, with the help of self-attention operations, our RENet can adap-

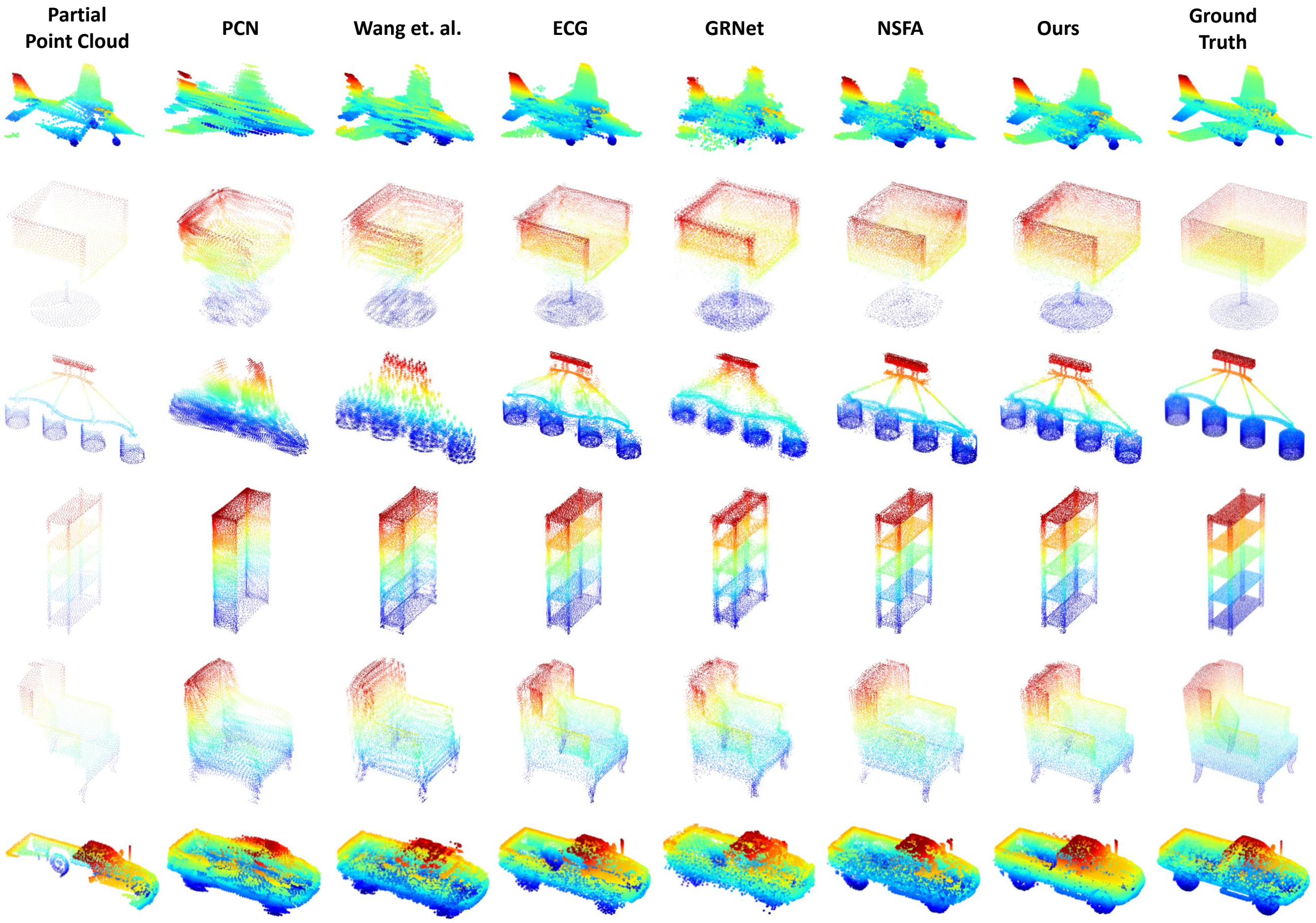

Figure 3: **Qualitative Results on the MVP Dataset by Different Methods.** Note that we use different point sizes for different shapes to achieve better visualizations.

tively recover fine-grained complete shapes by implicitly predicting shape structural relations, such as geometrical symmetries, regular arrangements and surface smoothness. More qualitative results by different methods are shown in Fig. 3, which show that the RENet can effectively learn structural relations for shape completion. For example, the lamp completion results (the 3rd row of Fig. 3) show that the VRCNet can recover those cylinder bulbs by the learned geometrical symmetry. In particular, given different observed incomplete shapes, the VRCNet can generate different complete 3D shapes with the help of both structural relations from the partial observations and the generated coarse overall shape skeletons (shown in Fig. 2). Moreover, the VRCNet can generate pluralistic complete shapes for real-scanned incomplete point clouds (see Fig. 4), which validates its strong robustness and generaliability.

C. Inference Details

Our VRCNet consists of two consecutive sub-networks, PMNet and RENet. PMNet generates overall shape skeletons (i.e. coarse completions) using probabilistic modeling, and RENet enhances structural relations at multiple scales to generate our fine completions. For inference, PMNet only uses its completion path to predict coarse completions based on the incomplete point clouds. A coarse complete point cloud that consists of 1024 points can be regarded as 3D adaptive points to facilitate learning local point relations. The coarse completion is combined with the incomplete point cloud (2048 points) as the input (3072 points) to the RENet. After exploiting multi-scale point features, RENet uses the Edge-aware Feature Expansion (EFE) module [4] to upsample and expand the point feature so as to generate complete point clouds with different resolutions. For example, we generate complete shapes with 16384 points by 1) upsampling to 18432 points (3072 × 5) and thus 2) downsampling to 16384 points using farthest

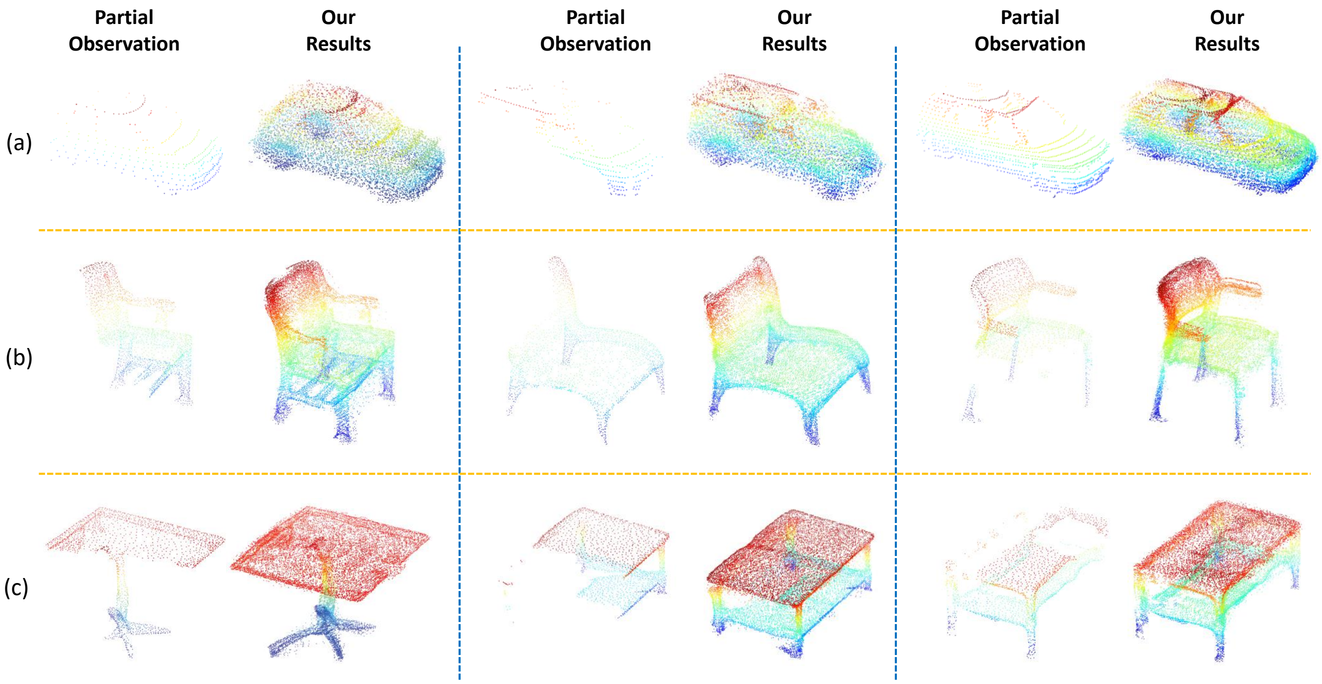

Figure 4: **Our completion results for real-scanned incomplete point clouds.** (a) shows our completion results for incomplete cars from the KITTI dataset [1]. (b) and (c) show our results for scanned incomplete chairs and tables, respectively.

Table 1: A user study of completion quality on real scans. The values are average scores given by volunteers (3 points for best result, 1 point for the worst result). VRCNet is the most preferred method overall

Category	PCN [10]	NSFA [11]	VRCNet
Car (KITTI)	**2.87**	1.07	2.07
Chair (ScanNet)	1.60	1.73	**2.67**
Table (ScanNet)	1.27	2.20	**2.60**
Overall	1.91	1.67	**2.45**

Table 2: Ablation studies (2,048 points) for the proposed network modules, including Point Self-Attention Kernel, Dual-Path Architecture and Point Selective Kernel Module.

Point Self-Attention	Dual-Path Architecture	Kernel Selection	CD	F1
			6.64	0.476
	✓		6.43	0.488
✓			6.35	0.484
	✓	✓	6.35	0.490
✓	✓		6.15	0.492
✓	✓	✓	5.96	0.499

point sampling.

D. User Study on Real Scans

We conduct a user study on the performances of various methods in Tab 1. Specifically, we gather a group of 15 volunteers to rank the quality of complete point cloud predicted by PCN, NSFA, and our VRCNet, on the real scans of three object categories: car, chair and table. For each object category, the volunteers are given three anonymous groups of results, produced by three methods. The volunteers are instructed to give the best, middle, and worst results 3, 2, and 1 point(s) respectively. We then compute the average scores of all volunteers for each method and class category. The evaluation is conducted in a double-blind manner (the methods are anonymous to both the instructor and the volunteers) and the order of the groups are shuffled for each category. Our VRCNet is the most favored method overall amongst the three. PCN obtains higher score for car completion because it generates smooth mean shapes for all cars, even though few observed shape details of those cars are preserved in their completion results. For the other two categories, chair and table, the VRCNet receives the highest scores due to its effectiveness on reconstructing complete shapes using predicted shape symmetries.

E. Ablation Studies

We provide the detailed ablation studies in Table 2, which reports the evaluation results with different combi-

Table 3: **Comparing MVP with existing datasets.** MVP has many appealing properties, such as 1) diversity of uniform views; 2) large-scale and high-quality; 3) rich categories. Note that both PCN and C3D only randomly render **One** incomplete point cloud for each CAD model to construct their testing sets. (C3D: Completion3D; Cat.: Categories; Distri.: Distribution; Reso.: Resolution; PC: Point Cloud; FPS: Farthest Point Sampling; PDS: Poisson Disk Sampling. Point cloud resolution is shown as multiples of 2048 points.)

	#Cat.	Training Set		Testing Set		Virtual Camera			Complete PC		Incomplete PC	
		#CAD	#Pair	#CAD	#Pair	Num.	Distri.	Reso.	Sampling	Reso.	Sampling	Reso.
PCN [10]	8	28974	∼200k	1200	1200	8	Random	160×120	Uniform	8×	Random	∼3000
C3D [5]	8	28974	28974	1184	1184	1	Random	160×120	Uniform	1×	Random	1×
MSN [3]	8	28974	∼1.4m	1200	1200	50	Random	160×120	Uniform	4×	Random	∼ 5000
Wang et. al. [6]	8	28974	28974	1200	1200	1	Random	160×120	Uniform	1×	Random	1×
SANet [7]	8	28974	∼200k	1200	1200	8	Random	160×120	Uniform	1×	Random	1×
NSFA [11]	8	28974	∼200k	1200	1200	7	Random	160×120	Uniform	8×	Random	1×
MVP	**16**	2400	62400	1600	**41600**	**26**	**Uniform**	**1600×1200**	**PDS**	**1/2/4/8×**	**FPS**	1×

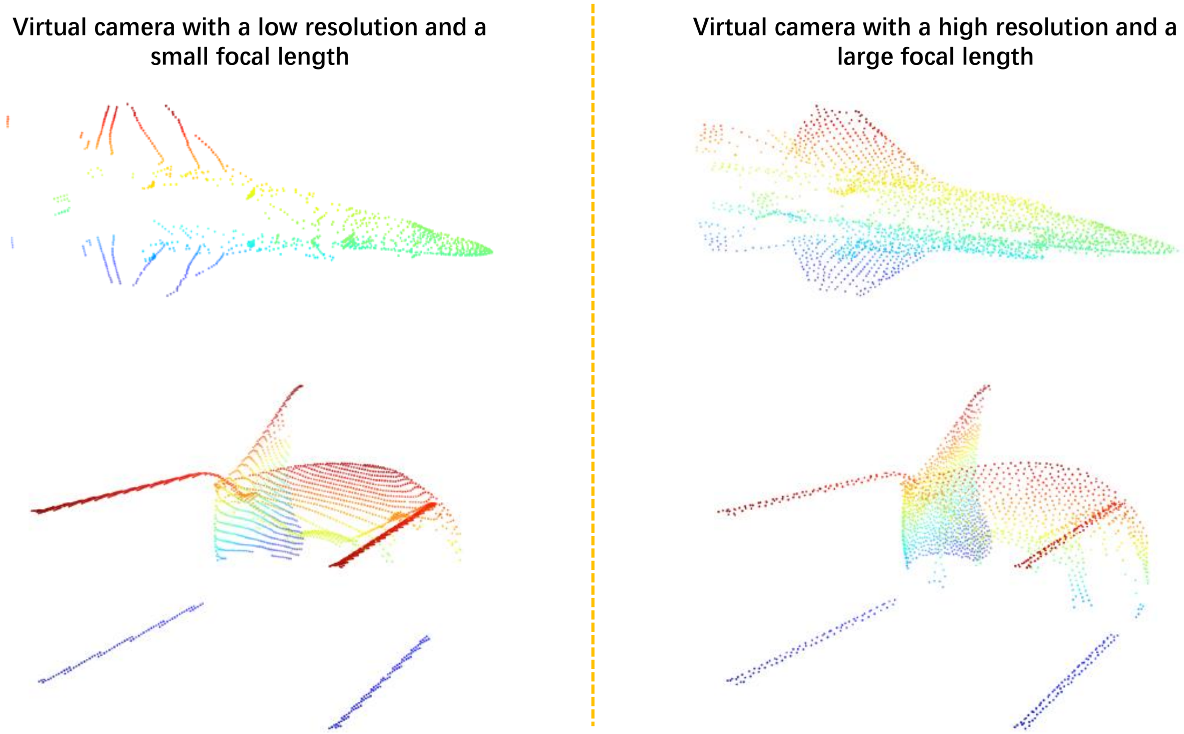
Virtual camera with a low resolution and a small focal length
Virtual camera with a high resolution and a large focal length

Figure 5: **Rendered Incomplete Point Clouds with Different Camera Resolutions.** We use a high camera resolution to capture more realistic shapes than using low resolutions.

nations of the proposed modules, Point Self-Attention Kernel (PSA), Dual-Path Architecture (DP) and Point Selective Kernel (PSK) Module. As reported in Table 2, it is obvious that using the proposed modules can improve the completion accuracy. During training and evaluation, the effectiveness of using PSA and DP are very straightforward. PSK may lead to fluctuating evaluation results during training, but it can highly improve the point cloud completion performance.

F. Resource Usage

We report the resource usages by PCN [10], NSFA [11] and our VRCNet in the Table 4. PCN and our VRCNet are implemented using pytorch, and we use the official implementation (by tensorflow) for NSFA. To achieve a fair comparison for the inference (Inf.) time, we use the same batch size 32 and test all methods by using an NVIDIA V100 GPU on the same workstation. Note that, NSFA has many non-trainable operations (such as ball query, grouping and sampling), and hence it takes the longest inference time al-

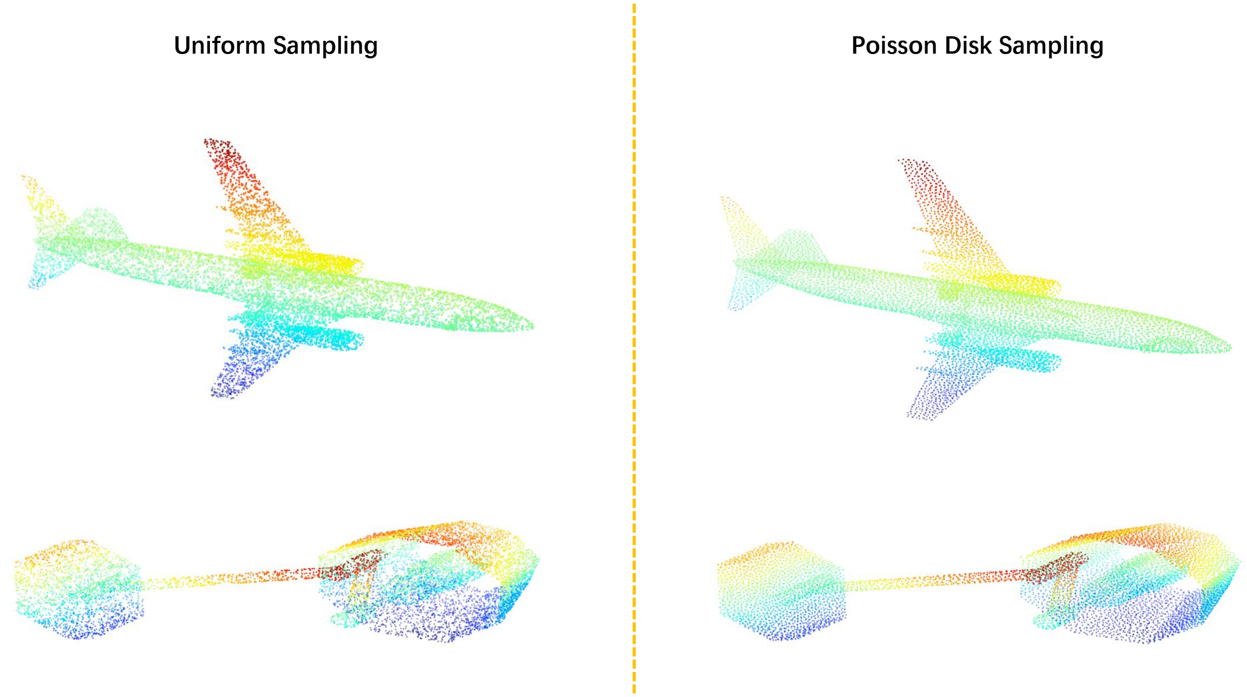

Figure 6: **Sampling Complete Point Clouds with Different Sampling Methods.** Unlike previous methods that use uniform sampling, we use Poisson Disk Sampling to generate complete point clouds, which can better describe the underlying 3D shape surfaces.

Table 4: Resource Usages.

Method	#Params. (M)	Model Size (Mib)	Inf. Time (ms)
PCN [29]	6.86	26	2.7
NSFA [30]	5.38	64	764.9
VRCNet	17.47	67	183.3

though it has the least parameters. Our VRCNet achieves significant improvements in completion qualities with an acceptable increment in the computational cost.

G. Dataset Comparisons

As stated in the main paper, previous methods usually use two datasets for incomplete point cloud: ShapeNet [8] by PCN [10] and Completion3D [5]. Because the incomplete point cloud dataset created by PCN is too massive, most following works (including the Completion3D) use a subset of ShapeNet [8] derived from PCN [10]. However, they do not have a unified and standardized dataset setting, which makes it difficult in directly comparing their performance. Furthermore, their generated shapes (incomplete and complete point clouds) all have low qualities, which makes their data unrealistic. In view of this, we create the Multiple-View Partial point cloud (MVP) dataset, which can be a high-quality and unified benchmark for partial point clouds. The detailed comparisons between different datasets are reported in Table 3. The proposed MVP dataset has more shape categories (16 v.s. 8), more testing data (e.g. 41600 v.s. 1200) and higher quality point clouds by using better data preparation methods (e.g. PDS, FPS and Uniformly distributed camera poses) than previous datasets. The qualitative comparisons for rendered incomplete shapes are visualized in Fig. 5. By using a high resolution and a large focal length, our rendered partial point clouds are more realistic than using low resolutions and small focal lengths. Moreover, the qualitative comparisons for sampled complete point clouds by different sampling methods are shown in Fig 6. The MVP dataset uses the Poisson Disk Sampling (PDS) method, which can yield smoother complete point clouds than using uniform sampling.

More qualitative comparisons can be found in our supplementary video.

References

[1] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 3

[2] Han Hu, Zheng Zhang, Zhenda Xie, and Stephen Lin. Local relation networks for image recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3464–3473, 2019. 1

[3] Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, and Shi-Min Hu. Morphing and sampling network for dense point cloud completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11596–11603, 2020. 4

[4] Liang Pan. Ecg: Edge-aware point cloud completion with graph convolution. *IEEE Robotics and Automation Letters*, 2020. 1, 2

[5] Lyne P Tchapmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 383–392, 2019. 4, 5

[6] Xiaogang Wang, Marcelo H Ang Jr, and Gim Hee Lee. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 790–799, 2020. 4

[7] Xin Wen, Tianyang Li, Zhizhong Han, and Yu-Shen Liu. Point cloud completion by skip-attention network with hierarchical folding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1939–1948, 2020. 4

[8] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. 5

[9] Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. Grnet: Gridding residual network for dense point cloud completion. *arXiv preprint arXiv:2006.03761*, 2020. 1

[10] 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. 3, 4, 5

[11] Wenxiao Zhang, Qingan Yan, and Chunxia Xiao. Detail preserved point cloud completion via separated feature aggregation. *arXiv preprint arXiv:2007.02374*, 2020. 1, 3, 4

[12] Hengshuang Zhao, Jiaya Jia, and Vladlen Koltun. Exploring self-attention for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10076–10085, 2020. 1

[13] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic image completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1438–1447, 2019. 1