Unsupervised 3D Shape Completion through GAN Inversion – Supplementary Material –

Junzhe Zhang[1,3] Xinyi Chen[2,3] Zhongang Cai[3,4] Liang Pan[1]
Haiyu Zhao[3,4] Shuai Yi[3,4] Chai Kiat Yeo[2] Bo Dai[1] Chen Change Loy[1]
[1]S-Lab, Nanyang Technological University [2]Nanyang Technological University
[3]SenseTime Research [4]Shanghai AI Laboratory
{junzhe001,xchen032}@e.ntu.edu.sg,{liang.pan,asckyeo,bo.dai,ccloy}@ntu.edu.sg
{caizhongang,zhaohaiyu,yishuai}@sensetime.com

A. Additional Details of Evaluation Metrics

Minimum Matching Distance (MMD) [1] measures the fidelity of a set of generated shapes $\mathbb{S}_c$ with respect to the set of ground truth shapes $\mathbb{S}_{gt}$. Specifically, we match every shape $\mathbf{x}_{gt}$ in $\mathbb{S}_{gt}$ to the generated shape $\mathbf{x}_c$ in $\mathbb{S}_c$ with minimum distance, and report the average macthed distance.

Earth Mover's Distance (EMD) [10] is the solution, *i.e.*, the minimum cost, to transport one shape's point cloud to another. It is defined in Eq. (1), where ϕ is a bijection. The bijection is highly indicative of uniformity of the generated shapes. We use the implementation provided by MSN [8].

$$d_{EMD}(\mathbf{x}_A, \mathbf{x}_B) = \min_{\phi:\mathbf{x}_A \to \mathbf{x}_B} \frac{1}{|\mathbf{x}_A|} \sum_{p \in \mathbf{x}_A} ||p - \phi(p)||_2 \quad (1)$$

Unidirectional Chamfer Distance (UCD) measures the squared L2 distance from the partial input shape $\mathbf{x}_p$ to the complete output $\mathbf{x}_c$, as shown in Eq. (2).

$$d_{UCD}(\mathbf{x}_p, \mathbf{x}_c) = \frac{1}{|\mathbf{x}_p|} \sum_{p \in \mathbf{x}_p} \min_{q \in \mathbf{x}_c} ||p - q||_2^2 \quad (2)$$

Unidirectional Hausdorff Distance (UHD) [4, 12], similarly, measures the single-sided Hausdorff distance:

$$d_{UHD}(\mathbf{x}_p, \mathbf{x}_c) = \max_{p \in \mathbf{x}_p} \min_{q \in \mathbf{x}_c} ||p - q||_2^2 \quad (3)$$

F1 score, following pcl2pcl [4], is defined as the harmonic average of the **accuracy** and the **completeness**, where accuracy measures the fraction of points in the $\mathbf{x}_c$ that are matched in the corresponding ground truth shape $\mathbf{x}_{gt}$, and completeness measures the fraction of points in the $\mathbf{x}_{gt}$ that are matched in $\mathbf{x}_c$. Here, the status of matching is determined by a threshold $\epsilon = 0.03$ for L2 distance.

B. Additional Details of Datasets

In Sec. 4.3 of the main paper, both the virtual scans and the ball-holed partial shapes follow the same train-test split of ShapeNet [3], whereas the PartNet [9] follows the train-test split provided by MPC [12]. As some shapes in the PartNet test set are present in the ShapeNet train set, we remove them for evaluating the PartNet target domain. The ball-holed partial shapes in this section are generated as described in PF-Net [7], *i.e.*, removing exactly 512 points from the complete shape. But in Fig. 1 and Fig. 9 of the main paper, higher incompleteness levels are tested to assess the robustness of ShapeInversion.

The real-world partial scans are provided by pcl2pcl [4]. Both ScanNet [5] and KITTI [6] are split into train set and test set, and the mapping GAN of the pcl2pcl framework, which maps the latent space of partial shapes to that of the complete ones, is retrained on the real scan train set. Note that MatterPort3D [2] does not have its own train set, so pcl2pcl is evaluated on MatterPort3D using the model retrained on ScanNet, given that both these two datasets are captured by depth cameras. In contrast, ShapeInversion does not need the real scan train set, and is directly evaluated on the test set.

C. Additional Details on Implementation

Pre-training. We follow Shu *et al.* [11] to train the tree-GAN baseline. The PatchVariance module has $n = 100$ patches, each with $k = 30$ points. This would ensure the entire generated shape (with 2048 points) is sampled via FPS. This setting works for r-GAN [1] as well, as shown in Fig. 1. tree-GAN is trained on eight Nvidia V100 GPUs for 2000 epochs. For the four categories with more than 3000 shapes, *i.e.*, plane, car, chair, and table, the batch size is 512; for the other four categories with fewer shapes, *i.e.*, cabinet, lamp, sofa, and boat, the batch size is 128 to train for enough iterations.

Inversion. The k in k-Mask is 5. We manually split the inversion stage into four sub-stages, each with different learning rates for z and θ: $\alpha_z = [1\times10^{-2}, 1\times10^{-4}, 1\times10^{-5}, 1\times$

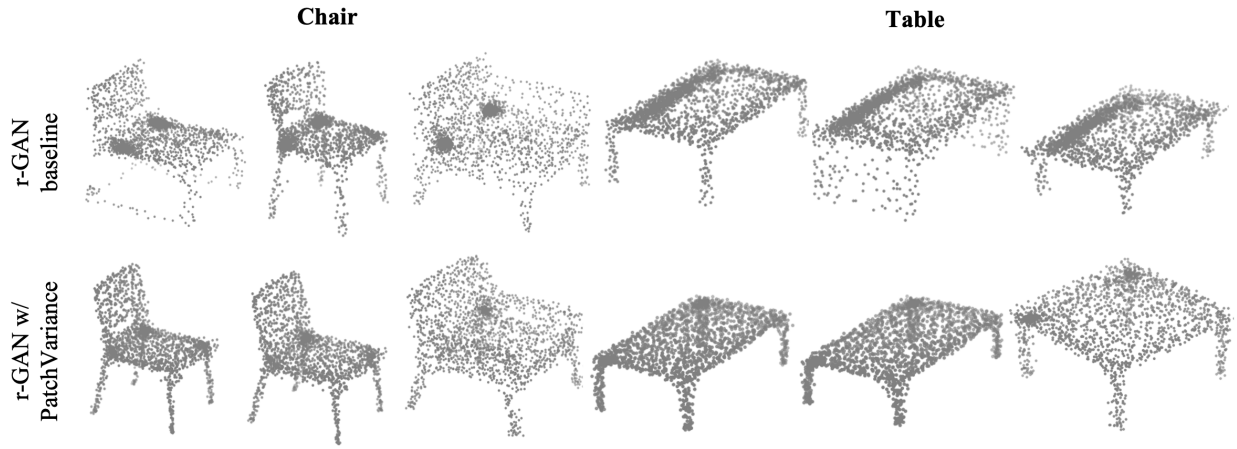

Figure 1. Validation of the significance of PatchVariance on r-GAN [1]. Non-uniformity issue is a common problem faced by point cloud GANs, our PatchVariance loss directly works on the generated point cloud, and is invariant to the architecture of the generator, hence it effectively enhances the uniformity of generated point clouds for r-GAN as well. Darker regions have a higher density of points

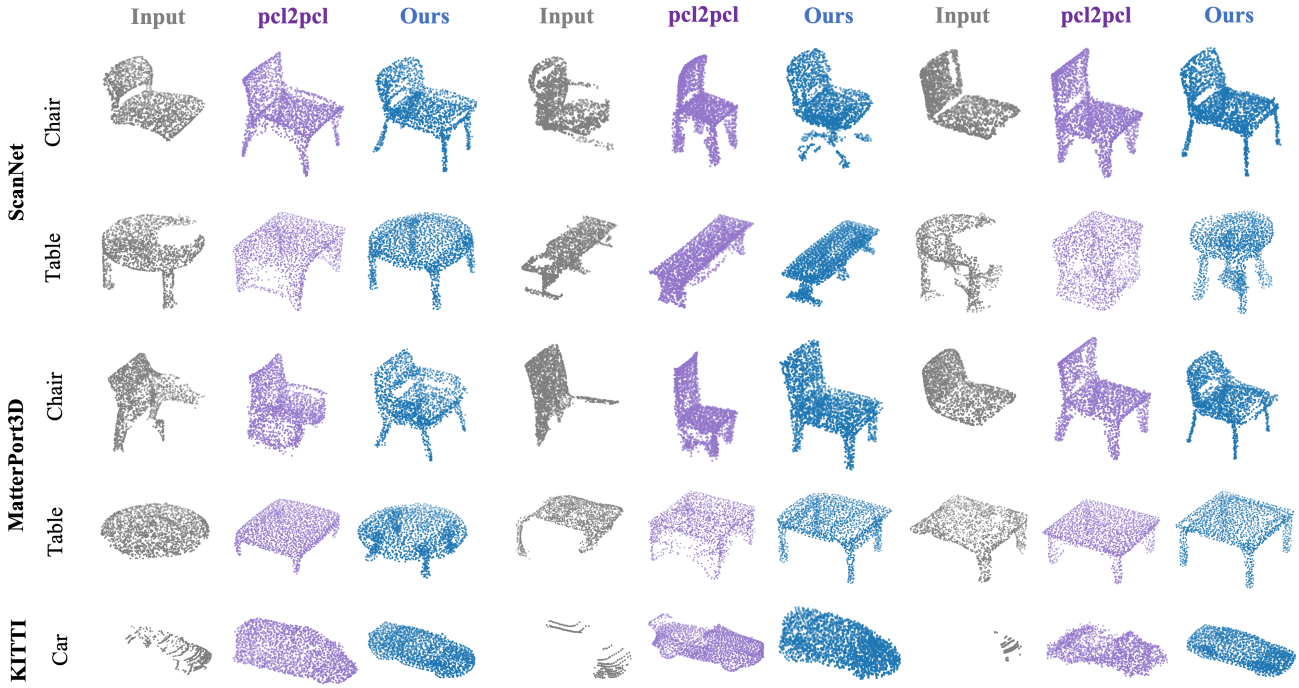

Figure 2. More qualitative results for shape completion on real-world partial scans. Note that pcl2pcl is retrained with real-world partial shapes (together with synthetic complete shapes in a unpaired manner [4]). In contrast, ShapeInversion does not use any real scans, yet, reconstructs high-fidelity shapes that are more faithful to the partial input

10^{-6}], and $\alpha_\theta = [2 \times 10^{-7}, 1 \times 10^{-6}, 1 \times 10^{-6}, 2 \times 10^{-7}]$. For bulk structures, *i.e.*, car, couch, cabinet, and plane, each sub-stage consists of 30 iterations; for thin structures, *i.e.*, chair, lamp, table, and boat, each sub-stage consists of 200 iterations.

D. Additional Qualitative Results

We show more qualitative results on shape completion for virtual scans (Fig. 3), cross-data validation (Fig. 4), real scans (Fig. 2), shape jittering (Fig. 5), shape morpging (Fig. 6), as well as on the effectiveness of PatchVariance on r-GAN (Fig. 1).

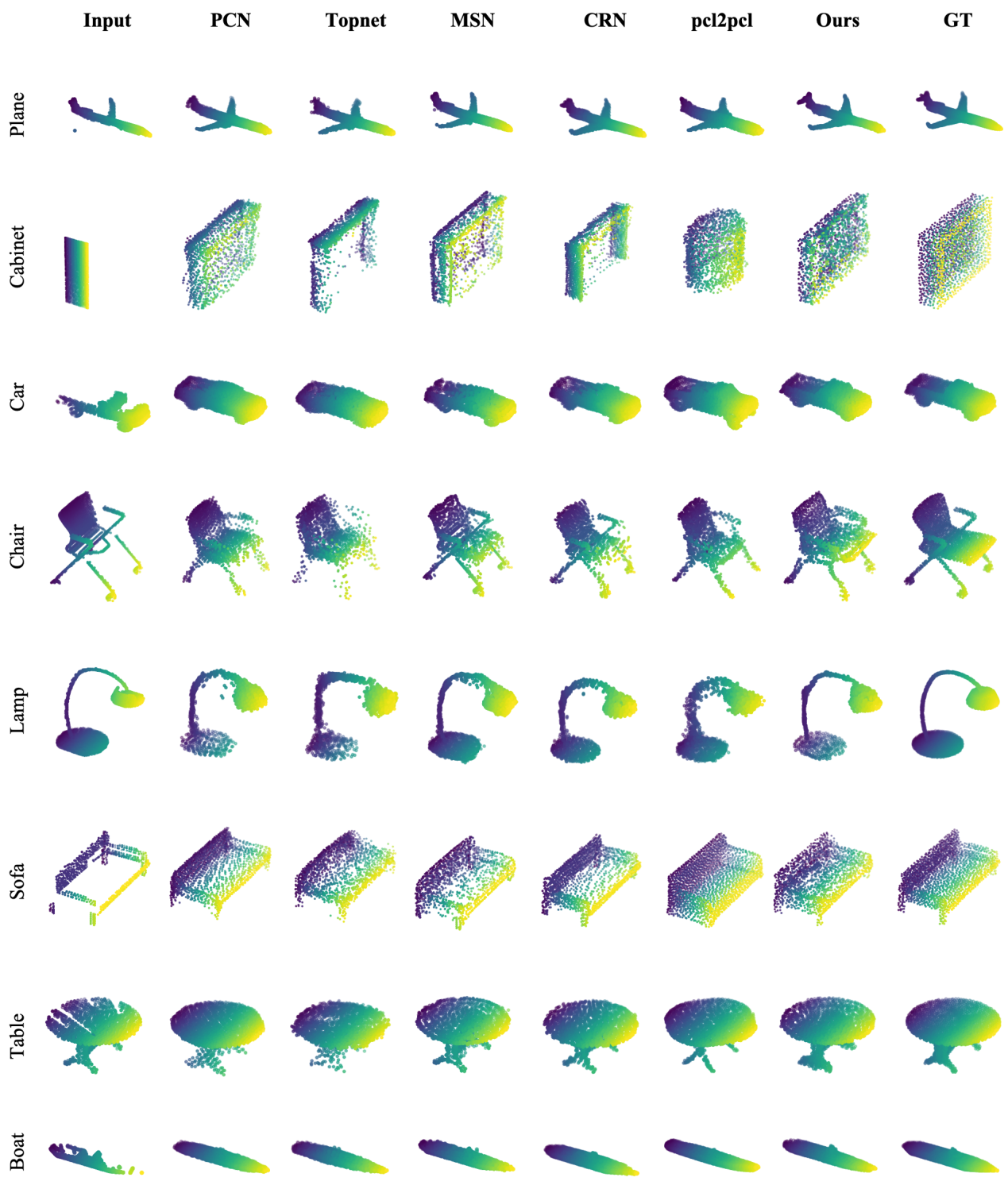

Figure 3. Qualitative comparison on the ShapeNet benchmark. The resolution of partial inputs, complete shapes, and ground truths are all 2048 points

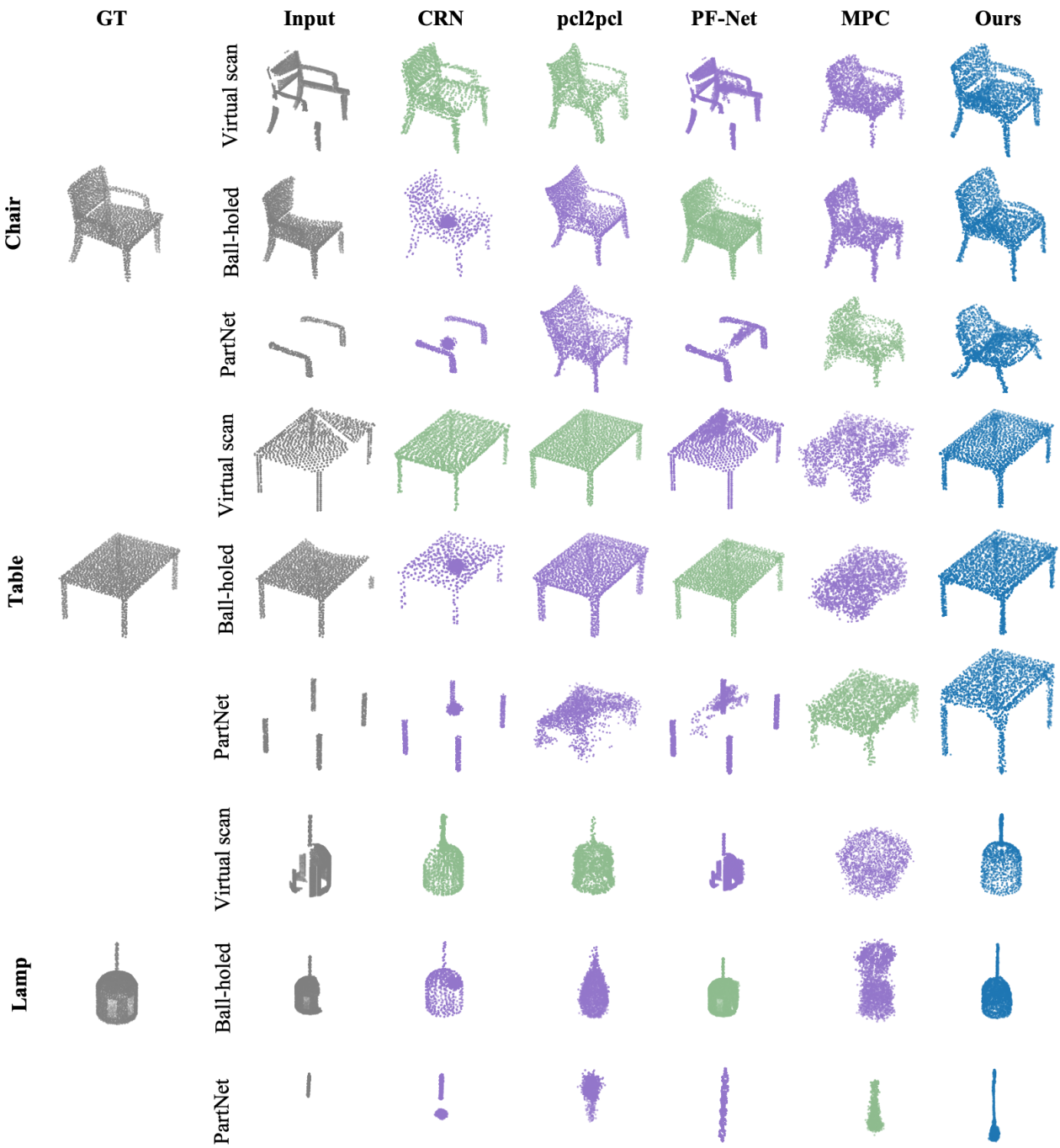

Figure 4. Additional visualization of cross-domain validation. Different partial forms of the same object are tested. In-domain results are in dark green whereas out-of-domain ones are in purple. Supervised methods CRN and PF-Net show significant performance drop with domain change; unsupervised methods pcl2pcl and MPC show relatively better results for out-of-domain inputs. ShapeInversion constantly provides plausible and accurate outputs for all partial forms. Note that CRN leverages the partial input during the refinement stage; PF-Net only predicts the missing regions and combines the partial input as the final output

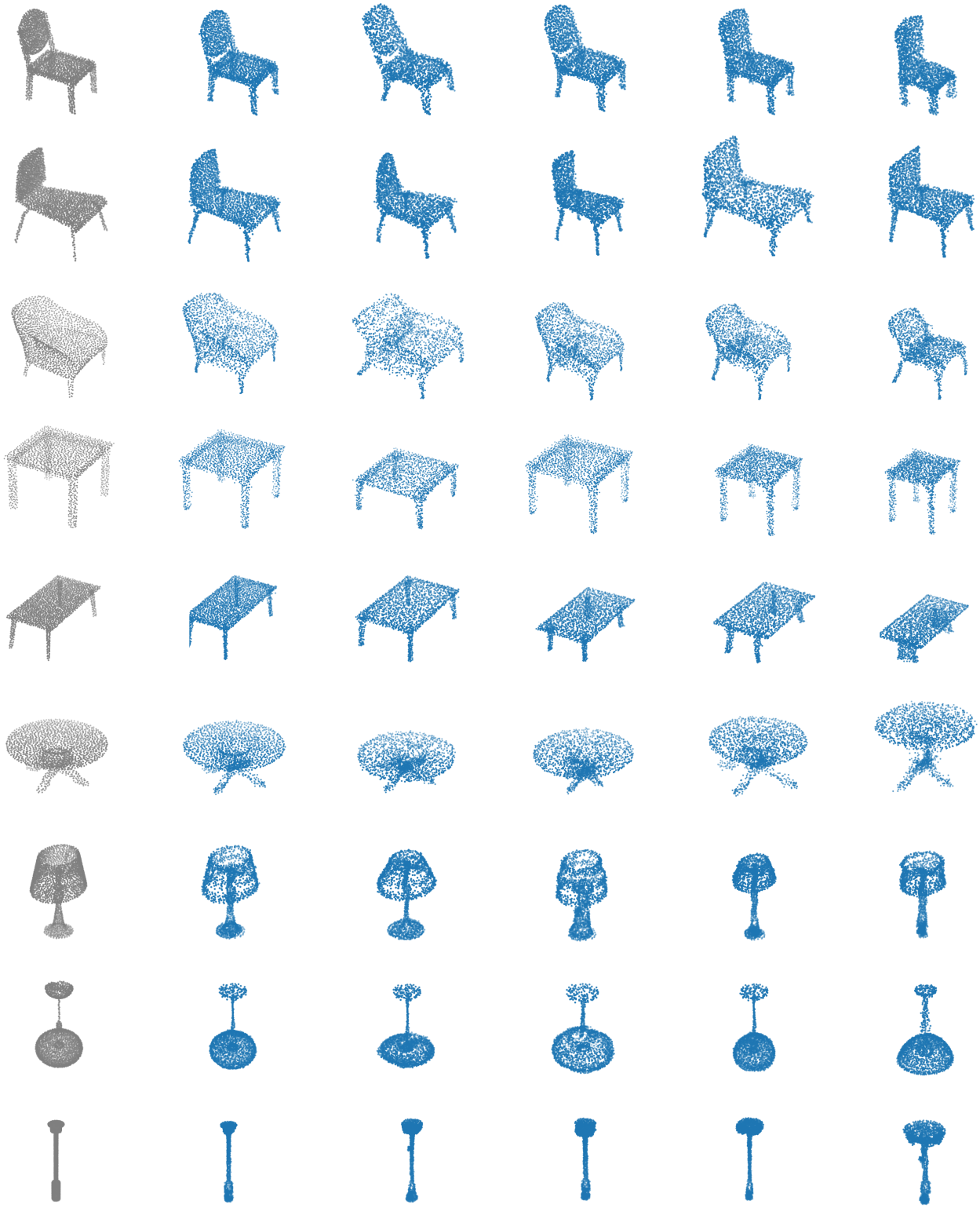

Figure 5. Additional qualitative results for shape random jittering by ShapeInversion. Random jittering changes one shape into other plausible shapes of different geometries by introducing perturbation in the latent space

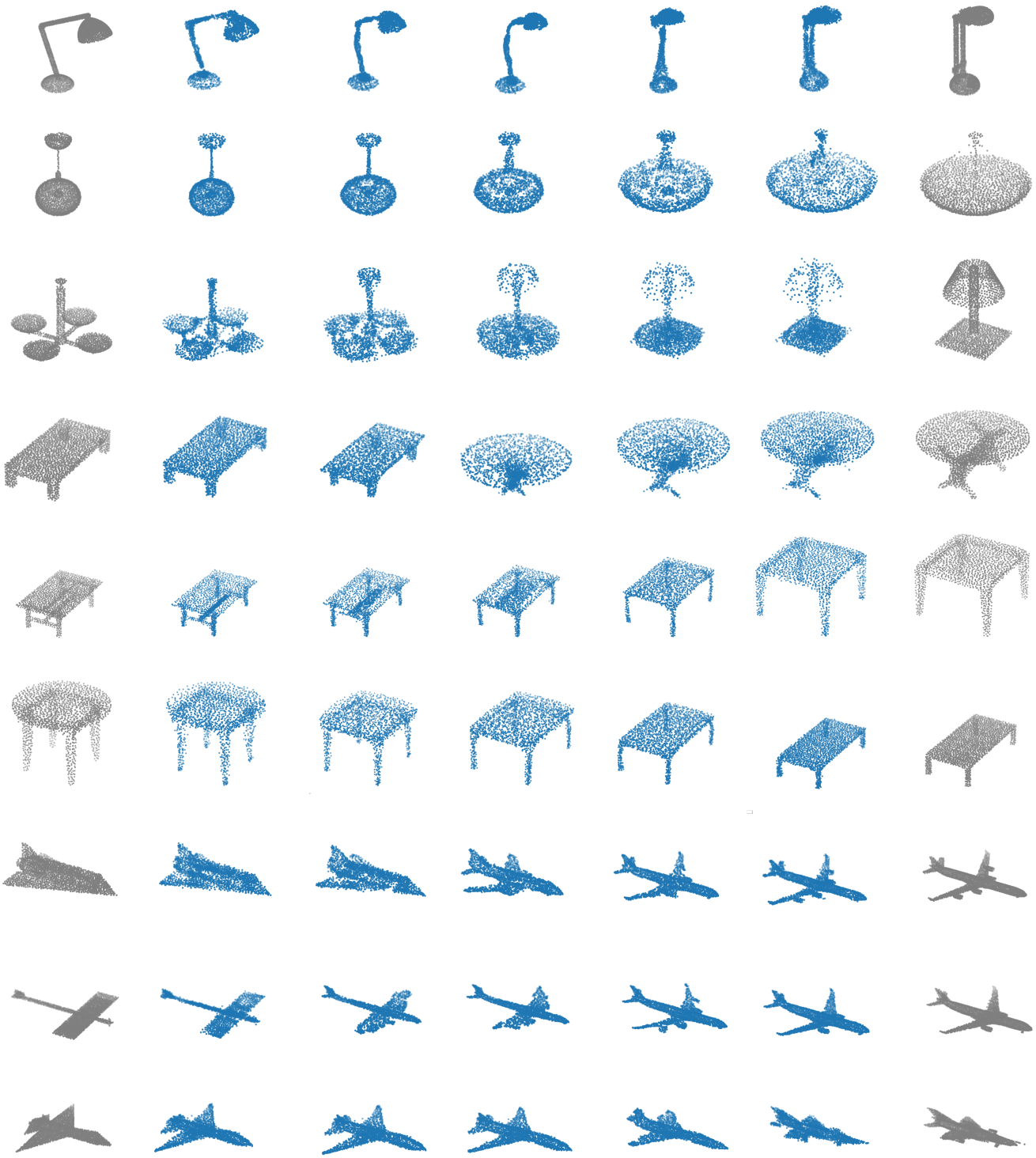

Figure 6. Additional qualitative results for shape interpolation (morphing) by ShapeInversion. A sound transition from one shape to another is achieved by interpolation between their corresponding latent vectors z and generator parameters θ

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3D point clouds. In *ICML*, 2018.

[2] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3D: Learning from RGB-D data in indoor environments. In *3DV*, 2017.

[3] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[4] Xuelin Chen, Baoquan Chen, and Niloy J Mitra. Unpaired point cloud completion on real scans using adversarial training. In *ICLR*, 2020.

[5] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In *CVPR*, 2017.

[6] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *CVPR*, 2012.

[7] Zitian Huang, Yikuan Yu, Jiawen Xu, Feng Ni, and Xinyi Le. PF-Net: Point fractal network for 3D point cloud completion. In *CVPR*, 2020.

[8] Minghua Liu, Lu Sheng, Sheng Yang, Jing Shao, and Shi-Min Hu. Morphing and sampling network for dense point cloud completion. In *AAAI*, 2019.

[9] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In *CVPR*, 2019.

[10] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vis.*, 40(2):99–121, 2000.

[11] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3D point cloud generative adversarial network based on tree structured graph convolutions. In *ICCV*, 2019.

[12] Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks. In *ECCV*, 2020.