Supplementary Material of Symmetric Shape-Preserving Autoencoder for Unsupervised Real Scene Point Cloud Completion

In the supplementary material, we first show more details about network, metric, and multi-category loss. Next, we conduct more experiments including the robustness of noises and the corrupted inputs, and the error bar of the training experiments. Finally, more completion on real objects and more qualitative comparisons are given to show the advantages of our USSPA.

1. More Details

1.1. Network Details

Inspired by PCN encoder [8], we design the structures of the encoders E_R, E_A, and the point discriminator $\mathcal{D}_p$ in our USSPA as shown in Figure 1. The parameter of the weight-shared MLPs[1] of our encoders are [3, 128, 256] and [512, 512]. And the parameters of the weight-shared MLPs of our point discriminator are [3, 128, 256] and [512, 256]. Besides, to get the probability, we also employ an MLP in our point discriminator after extracting the global feature. The parameter of this MLP is [256, 64, 64, 1].

The decoders D_R, D_A, D_{sl} and the feature discriminator $\mathcal{D}_f$ are MLPs with parameters [512, 512, 512, 1024, 1024, 1536], [512, 512, 512, 1024, 1024, 1536], [512, 128, 128, 2] and [512, 64, 64, 1] separately. The parameters of the weight-shared MLPs of the upsampling refinement module are [3, 256, 256, 256], [512, 256, 256, 256], [512, 256, 256, 256], respectively. The details of our upsampling operation in our upsampling refinement module are shown in Figure 1 as we mentioned in our paper. For a point p with feature f, the upsampling operation generates k points p_0, p_1, ..., p_{k-1}. The parameter of the MLP is [515, 512, 512, 12].

For multi-category data, the channel numbers of the output layer in our discriminators are set to $11 = 10 + 1$ where 10 is the category number and 1 represents the fake label mentioned in our paper.

In practice, we alternately optimize the generator and the discriminator of our USSPA.

[1]For a MLP, $[c_1, c_2, ..., c_n]$ means that there are n layers in MLP. The input layer has c_1 channels, the output layers have c_n channels, and $c_2, ..., c_{n-1}$ denote the channel numbers of hidden layers.

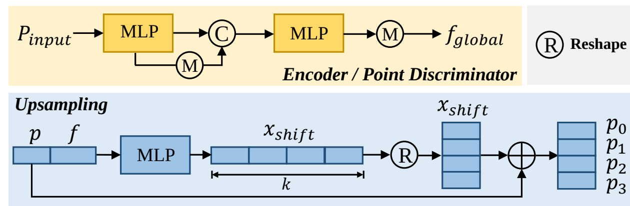

Figure 1. Detailed structures of our encoders and point discriminator (yellow) and the upsampling operation (blue).

1.2. Metric Details

We employ pre-point L1 Chamfer Distance cd_{l1} and F-score $F^{0.1\%}_{score}$, $F^{1\%}_{score}$ to measure the difference between the predicted point cloud and the referential ground truth. Following the definition in [1], the Chamfer Distance with L1-norm between two point clouds P_1 and P_2 is calculated as:

$$\begin{aligned} cd^{l1}_{P_1 \leftrightarrow P_2} =& \frac{1}{2}\Big(\frac{1}{|P_1|}\sum_{\boldsymbol{x}\in P_1}\min_{\boldsymbol{y}\in P_2}||\boldsymbol{x}-\boldsymbol{y}|| \\ &+ \frac{1}{|P_2|}\sum_{\boldsymbol{x}\in P_2}\min_{\boldsymbol{y}\in P_1}||\boldsymbol{x}-\boldsymbol{y}||\Big). \end{aligned} \tag{1}$$

As mentioned in [2], the Chamfer Distance indicator is sometimes misleading because of its sensitivity to outliers. We then further take advantage of F-score [2] to deliver more comprehensive evaluations. F-score is defined as:

$$F^d_{score} = \frac{2P(d)R(d)}{P(d)+R(d)}, \tag{2}$$

where P is precision:

$$P(d) = \frac{1}{|P_1|}\sum_{\boldsymbol{x}\in P_1}[\min_{\boldsymbol{y}\in P_2}||\boldsymbol{x}-\boldsymbol{y}|| < d], \tag{3}$$

and R is recall:

$$R(d) = \frac{1}{|P_2|}\sum_{\boldsymbol{x}\in P_2}[\min_{\boldsymbol{y}\in P_1}||\boldsymbol{x}-\boldsymbol{y}|| < d]. \tag{4}$$

Both P and R take a parameter d to control the strictness. Smaller d means a smaller neighborhood, which brings both lower precision and recall. As suggested, we set d at 0.01 and 0.001 separately.

1.3. Multi-Category Loss Details

Let $\mathcal{D}_p(P_{R_0})_i$ represents the probability that the prediction P_{R_0} belongs to the i-th category, and g denotes the ground truth label of the artificial input point cloud $\mathcal{P}_A$. As we mentioned in the paper, the label 0 denotes fake data while the label $l > 0$ denotes real data belonging to category l. Thus, for multi-category data with K categories, we change the loss $\mathcal{L}_{F\to R}$, $\mathcal{L}_R$ and $\mathcal{L}_F$ into:

$$\begin{aligned}\mathcal{L}_{F\to R} = -\big(&\log[1-\mathcal{D}_p(P_{R_0})_0] + \log[1-\mathcal{D}_p(P_{R_2})_0] \\ &+ \log[1-\mathcal{D}_f(f_R)_0]\big),\end{aligned} \tag{5}$$

$$\begin{aligned}\mathcal{L}_F = -\big(&\log \mathcal{D}_p(P_{R_0})_0 + \log \mathcal{D}_p(P_{R_2})_0 \\ &+ \log \mathcal{D}_f(f_R)_0\big),\end{aligned} \tag{6}$$

and

$$\mathcal{L}_R = -\sum_{i=0}^{K}\big(g_i \log \mathcal{D}_p(\mathcal{P}_A)_i + g_i \log \mathcal{D}_f(f_A)_i\big), \tag{7}$$

respectively.

2. More Experiments

2.1. Robustness of Noises

We conduct an experiment by adding random shifts to the points of input point clouds to evaluate the robustness of the noises of our method. For an input point cloud $\mathcal{P}_R \in \mathbb{M}_{n_0\times 3}$, we randomly generate a shift matrix $x_{shift} \in [-s, s]_{n_0\times 3}$ through uniform sampling, where s indicates the noise scale. Then, we add the shift matrix to the input point cloud by $\mathcal{P}_R \leftarrow \mathcal{P}_R + x_{shift}$ to randomly shift the points. The Chamfer Distance cd^{l1} and F-Score $F^{0.01\%}_{score}$ of our prediction on chair category with different noise scales are shown in Figure 2. The results show the robustness of the noises of our method when the noise scale s is lower than 0.1.

2.2. Robustness of Corrupted Inputs

To evaluate the robustness of corrupted inputs and verify the necessity of our symmetry learning module, we remove the points of input point cloud along the X-axis as shown in figure 3. The completion results are nearly the same with different keeping ratios of input point clouds. These results show the robustness of corrupted inputs of our method which benefits from our carefully designed symmetry learning module. Our symmetry learning module can generate the symmetrical point cloud of input which fully leverages the existing information.

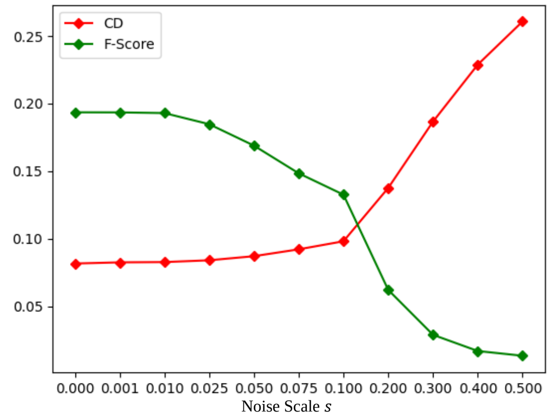

Figure 2. The Chamfer Distance cd^{l1} and F-Score $F^{0.01\%}_{score}$ of our prediction on chair category with different noise scales

Figure 3. The completion results (bottom) of the inputs (top) with different keeping ratios (100%, 75%, 50%, 25%).

2.3. Error Bar of Training Experiments

The error bars of our training experiments on 10 categories are shown in Figure 4, which shows that the performance of our USSPA is stable and repeatable.

3. More Completion on Real Object

More completion on real objects by Unpaired [6], ShapeInv [9], and our USSPA are shown in Figure 5. The results of the TV on the first line show that our prediction is more uniform with the vertical right boundary. And the results of the table on the second line show that our prediction is cleaner with fewer outliers. More comparisons indicate the superiority and generalization ability of our USSPA on real objects.

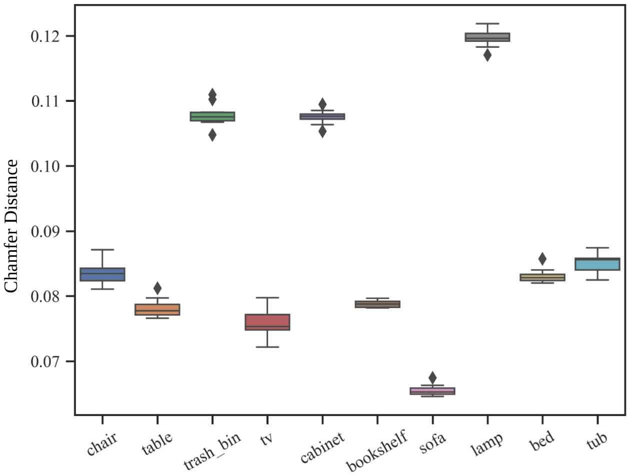

Figure 4. The box image of our method in terms of the Chamfer Distance cd^{l1}, where diamonds indicate the outliers.

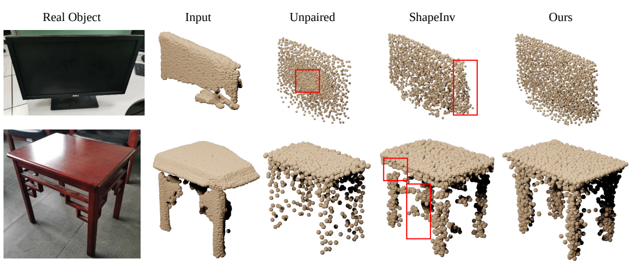

Figure 5. Completion on real objects (TV(top) and table(bottom)) by Unpaired [6], ShapeInv [9] and our USSPA.

4. More Qualitative Comparisons

More qualitative comparisons of our USSPA and other works [3–7, 9] are shown in Figures 6, 7 and 8.

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. 1

[2] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3405–3414, 2019. 1

[3] 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. 3

[4] Yida Wang, David Joseph Tan, Nassir Navab, and Federico Tombari. Learning local displacements for point cloud completion. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1558–1567, 2022. 3

[5] Xin Wen, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Yu-Shen Liu. Cycle4completion: Unpaired point cloud completion using cycle transformation with missing region coding. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13075–13084, 2021. 3

[6] Niloy J. Mitra Xuelin Chen, Baoquan Chen. Unpaired point cloud completion on real scans using adversarial training. In *2020 International Conference on Learning Representations 2020 (ICLR)*, 2020. 2, 3

[7] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12498–12507, 2021. 3

[8] 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

[9] Junzhe Zhang, Xinyi Chen, Zhongang Cai, Liang Pan, Haiyu Zhao, Shuai Yi, Chai Kiat Yeo, Bo Dai, and Chen Change Loy. Unsupervised 3d shape completion through gan inversion. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1768–1777, 2021. 2, 3

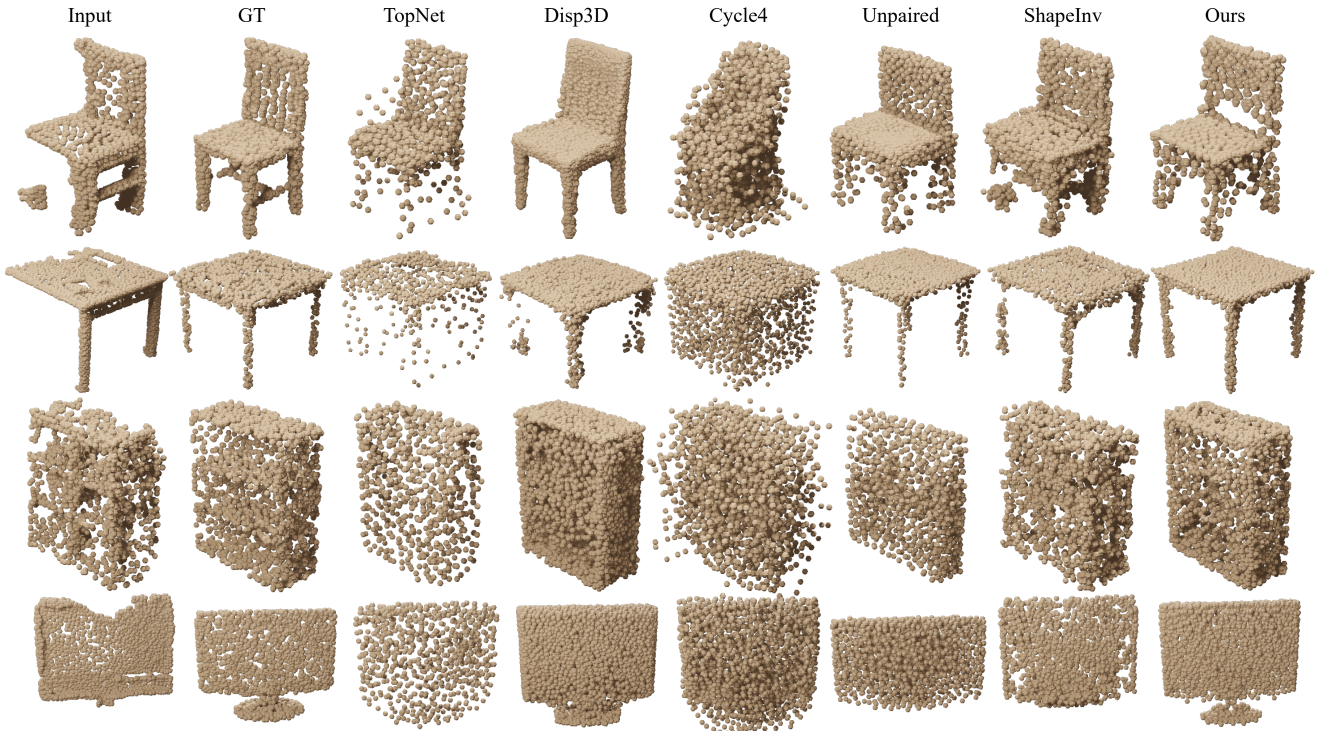

Figure 6. Visualization of completion results by our USSPA and other works. From top to bottom: chair, table, bookshelf, and TV.

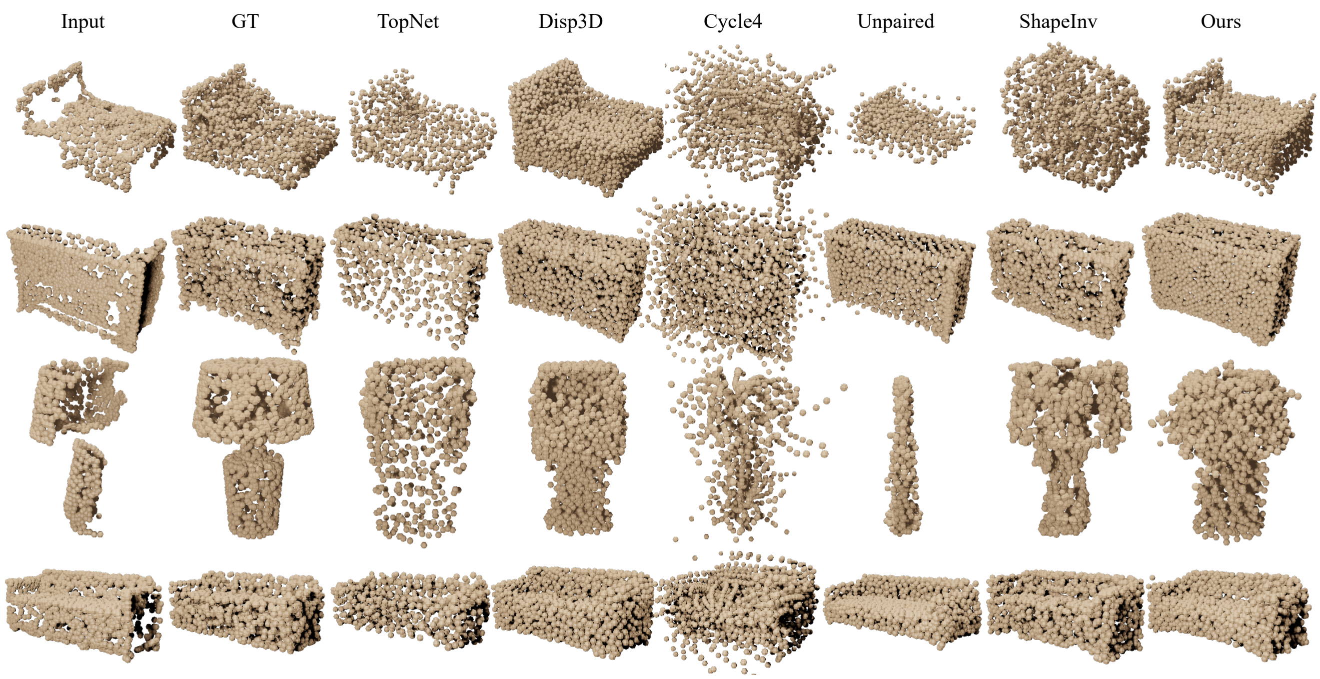

Figure 7. Visualization of completion results by our USSPA and other works. From top to bottom: bed, cabinet, lamp, and sofa.

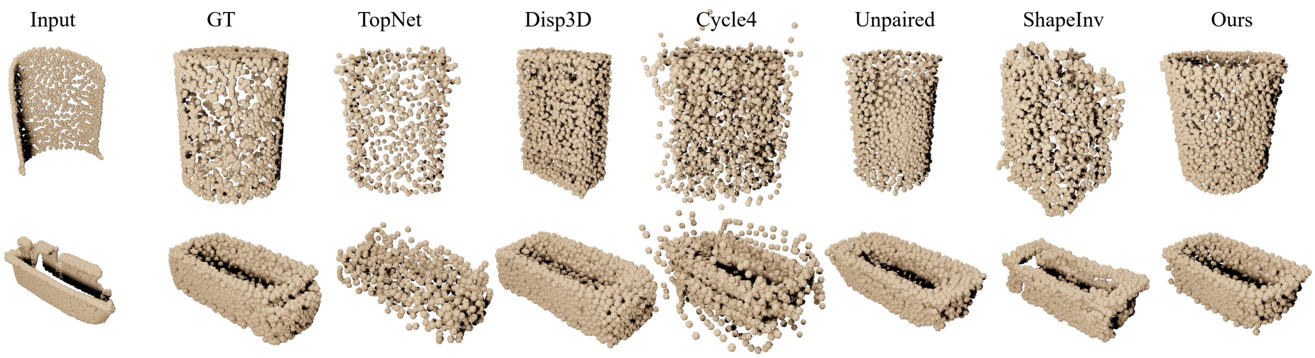

Figure 8. Visualization of completion results by our USSPA and other works. From top to bottom: trash bin, and tub.