Supplementary Materials for Point Cloud Upsampling via Cascaded Refinement Network

Hang Du*, Xuejun Yan*, Jingjing Wang, Di Xie, and Shiliang Pu(✉)

Hikvision Research Institute, Hangzhou, China
{duhang, yanxuejun, wangjingjing9, xiedi, pushiliang.hri}@hikvision.com

1 Network Details

In the following, we elaborate the detailed network architecture.

Feature extraction. In Fig.3 of the main text, we provide the overall structure of feature extraction module. Here, we give more details about the parameter settings. At the start of the feature extraction module, we feed the initial points to a set of MLPs, in which the numbers of output channel are 64, and 128, respectively. Then, the global feature is produced through a max pooling operation. After that, the duplicated global features are concatenated with the initial features and reduced to 128 channels via two MLP layers. We further adopt a point transformer layer [2] to refine the local shape context, and the channel number C' is set to 64 at the stage of feature transformations. After the local self-attention operation, the channel number of output point features is 128.

Feature expansion. As presented in the main text, we employ two branches for feature expansion. For transposed convolution-based branch, we set the output channel numbers of MLPs as 32, and then a one-dimensional deconvolution layer is utilized to produce expanded features with 128 channel numbers and r times point numbers. For duplicate-based branch, the output features also have r times point numbers with 128 output channel numbers. Then, we concatenate the two-branch features and obtain the expanded features using two MLP layers, in which the output channel numbers are 256 and 128, respectively.

Coordinate reconstruction. For regressing the per-point offset $\Delta\mathcal{P}$, the expanded features are gradually reduced to 64 and 3 channels through two MLP layers. Then, the per-point offset $\Delta\mathcal{P}$ is added on the r times duplicated input point clouds.

2 More Experimental Results

In this section, we provide more experimental results, including effect of training supervision, ablation study on refinement stage and visualization results on real-scanned data.

* Equal contribution.

Table 1: Effect of training Supervision. The values of CD, HD, and P2F are multiplied by 10^3. A smaller value denotes a better performance.

Model	Medium (1,024) input			Dense (2,048) input		
	CD	HD	P2F	CD	HD	P2F
Last stage	0.832	11.443	3.108	0.507	7.904	1.972
All stages	**0.808**	**10.750**	**3.061**	**0.471**	**7.123**	**1.925**

2.1 Effect of Training Supervision

In this section, we conduct an experiment to verify the effectiveness of the supervision on three stages.

From the results in Table 1, we can see that adding the supervision on all stages achieves a better result than only constraining on the last stage. We consider the reason behind is that the supervision on each stage enables to make its output more reliable and then provides a better initial shape for the next stage. Therefore, we calculate CD loss for three generation stages and optimize them simultaneously.

2.2 Ablation Study on Refinement Stage

Fig. 1 provides some visualized results on removing the refinement stage in the inference phase. From the results, we can find there are some outliers produced by the second upsampling stage (the second column). Then, the refiner enables to adjust them to a better position, and thus obtains a result with higher fidelity.

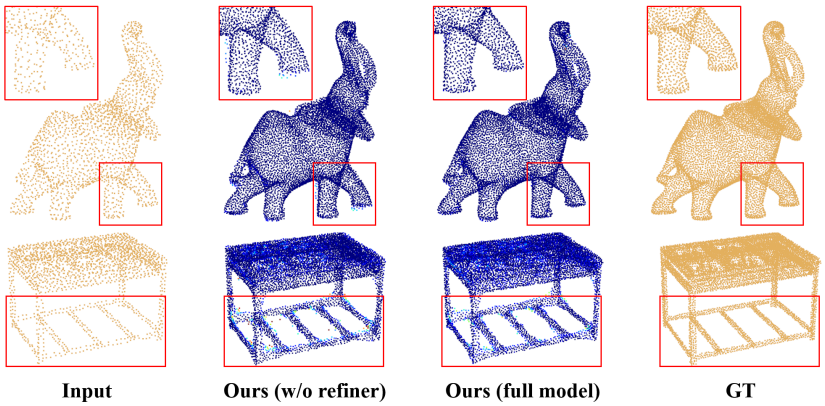

Fig. 1: Qualitative comparisons on refinement stage.

2.3 Visualization Results

Here, we give more visualization results on ScanObjectNN [1] dataset. As shown in Fig. 2, our method generate more uniform with detailed structures on various objects compared with other competitors. The 3D surface reconstruction are largely influenced by the quality of the upsampled point clouds. The proposed method is able to preserve the details in the sharp areas and smoothness in the smooth regions. The visualization results demonstrate that our upsampled point clouds are more uniform and close to the target surface.

References

1. Uy, M.A., Pham, Q.H., Hua, B.S., Nguyen, D.T., Yeung, S.K.: Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In: ICCV. pp. 1588–1597 (2019)
2. Zhao, H., Jiang, L., Jia, J., Torr, P.H.S., Koltun, V.: Point transformer. In: ICCV (2021)

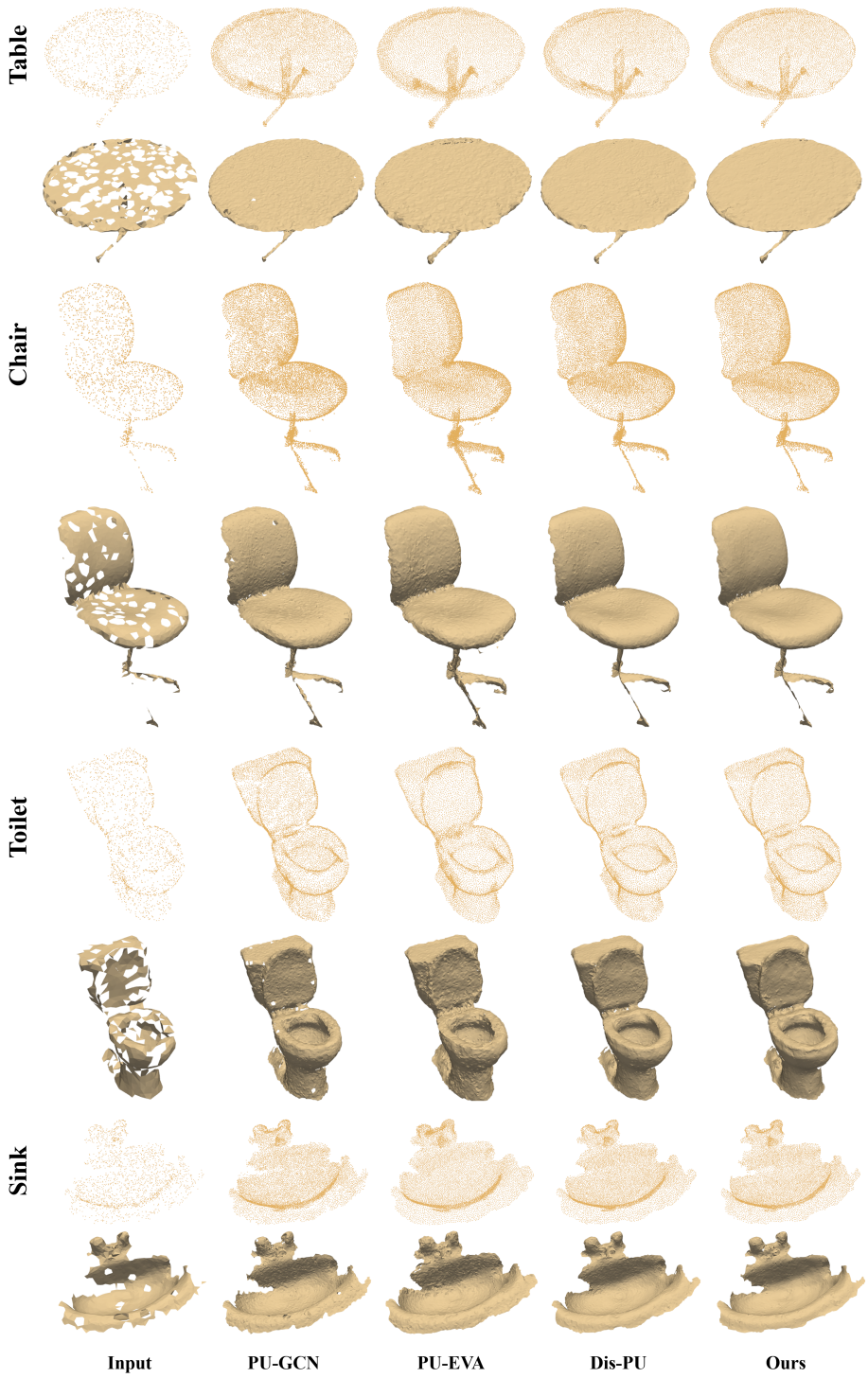

Fig. 2: Point cloud upsampling (×4) results on real-scanned sparse inputs. Compared with the other methods, our upsampled point clouds are more uniform and proximity-to-surface. One can zoom in for details.