Supplementary Materials for PatchRD: Detail-Preserving Shape Completion by Learning Patch Retrieval and Deformation

Bo Sun[1], Vladimir G. Kim[2], Noam Aigerman[2], Qixing Huang[1], and Siddhartha Chaudhuri[2,3]

[1] UT Austin
[2] Adobe Research
[3] IIT Bombay

1 More Results

We show more qualitative results of shape completion results on random-crop dataset (Figure 1 and Figure 2), ScanNet[1] objects (Figure 3), shapes with large missing areas (Figure 4) and novel categories (Figure 5).

2 More Results on PCN Benchmark

We show the quantitative results and more qualitative results on the PCN dataset [6] in Table 1 and Figure 6 respectively. Our method is quantitatively a little worse than the best-performing SnowFlakeNet because our method might fail when there's no reference details in the input shape, and CD-L1 is more sensitive to structure than details. Importantly, visual results in Figure 6 indicate that our method produces more clean and plausible shapes, especially in the missing areas.

3 More Training Details

For the **coarse completion**, the input shape is the partial detailed shape and the ground truth is the coarse version ($4\times$ downsampled) of the detailed full shape. The loss function is the cross-entropy loss between the GT and the output. We use Kaiming Uniform method for weight initialization and the Adam optimizer to train 100 epochs for each shape category on a single Titan X. Training takes $\sim$3 hrs for the 3D CNN, $\sim$12 hrs for retrieval, and $\sim$2 hrs for deformation and blending. Inference for a shape with 128^3 voxels takes $\sim$20s on a single 12GB Titan X.

4 Failure Cases Analysis

Our pipeline has 3 stages: (1) coarse completion, (2) patch retrieval, and (3) patch deformation and blending. If one stage fails, the result might be different from

the GT shape. However, our method can still produce plausible output, i.e. semantically correct and smoothly connected shapes. If stage 1 fails, the overall structure will be different from the GT shape. If stage 2 fails, the local details will be inaccurate. If stage 3 fails, the connection between patches will not be smooth, causing irregular or noisy shapes. Some examples of failure cases from each step are shown in Fig. **??**.

5 Network Architectures

We show the detailed network architectures for coarse completion, retrieval metric learning, and deformation and blending weight prediction in Figure 8, Figure 9, and Figure 10 respectively.

	TopNet[3]	GRNet[5]	SnowFlakeNet[4]	PatchRD(Ours)
CD-L_1	13.43	9.37	7.78	8.79

Table 1: Quantitative results on chair class of the PCN Dataset[6]. All methods are trained on chair class only. We report the L_1 chamfer distance $\times 10^{-}3$.

References

1. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE (2017)
2. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A.: Convolutional occupancy networks. In: European Conference on Computer Vision (ECCV) (2020)
3. Tchapmi, L.P., Kosaraju, V., Rezatofighi, S.H., Reid, I., Savarese, S.: Topnet: Structural point cloud decoder. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
4. Xiang, P., Wen, X., Liu, Y.S., Cao, Y.P., Wan, P., Zheng, W., Han, Z.: Snowflakenet: Point cloud completion by snowflake point deconvolution with skip-transformer. In: ICCV (2021)
5. Xie, H., Yao, H., Zhou, S., Mao, J., Zhang, S., Sun, W.: Grnet: Gridding residual network for dense point cloud completion. In: ECCV (2020)
6. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: Pcn: Point completion network. In: 3D Vision (3DV), 2018 International Conference on 3D Vision (2018)

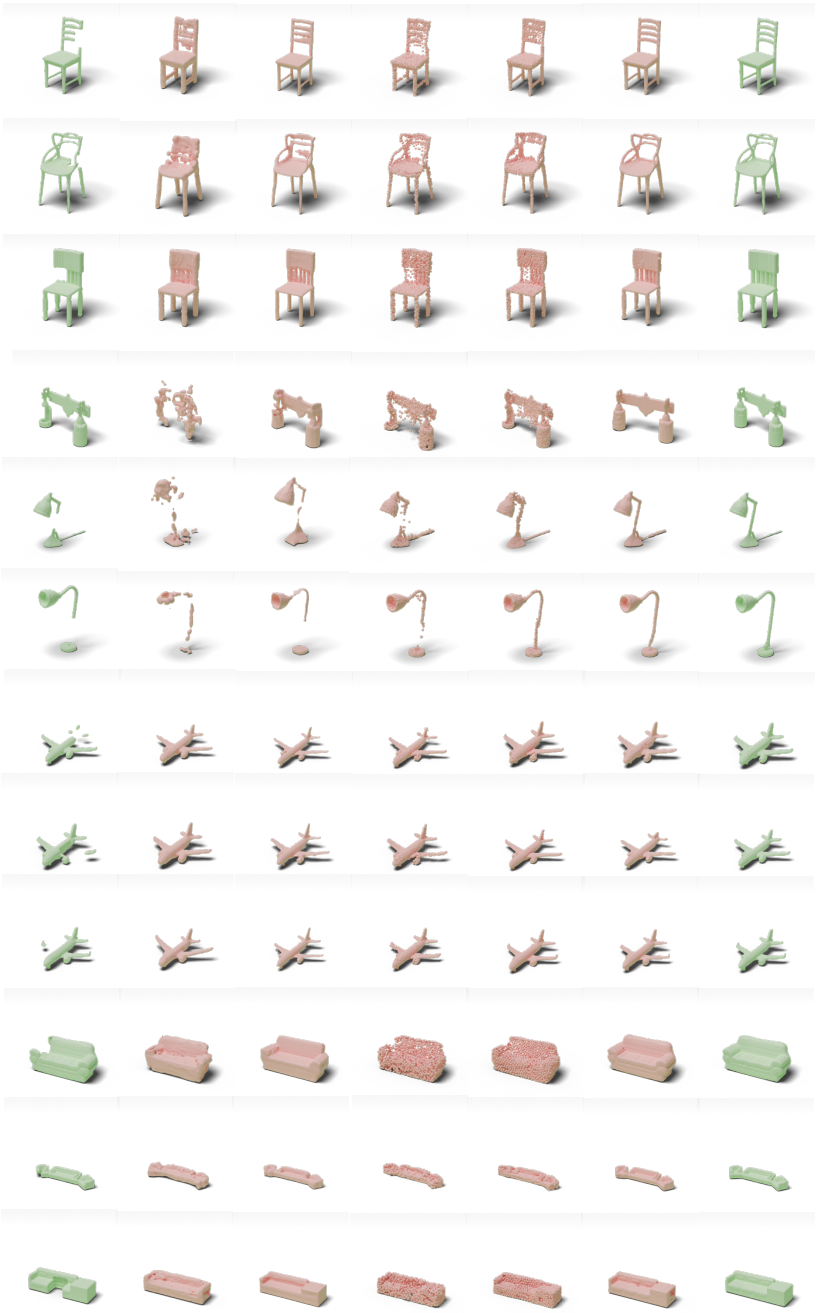

Fig. 1: More qualitative shape completion results on the Random-Crop Dataset.

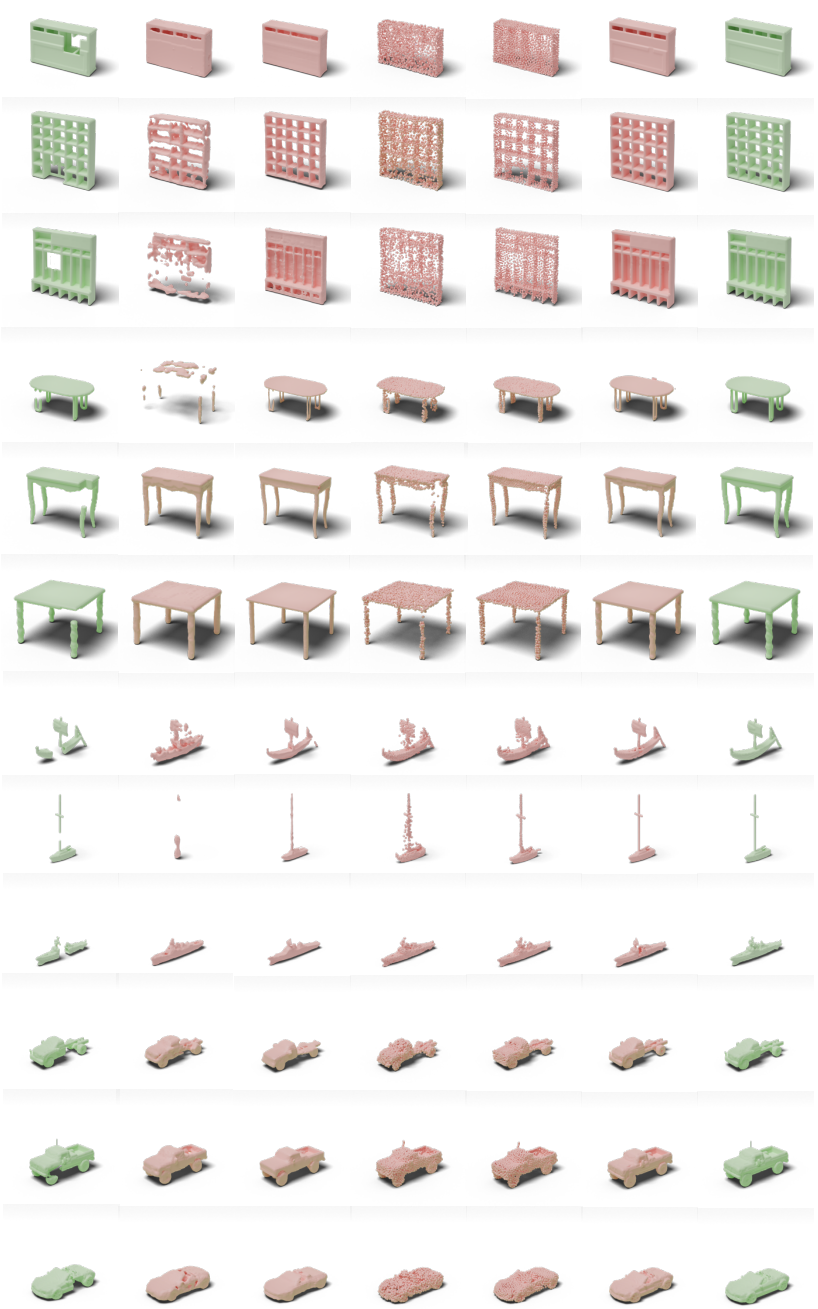

Fig. 2: More qualitative shape completion results on the Random-Crop Dataset.

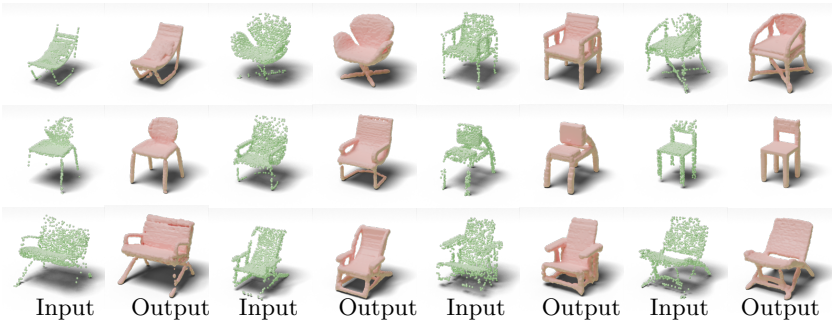

Fig. 3: More shape completion results on real scans for ScanNet[1] objects.

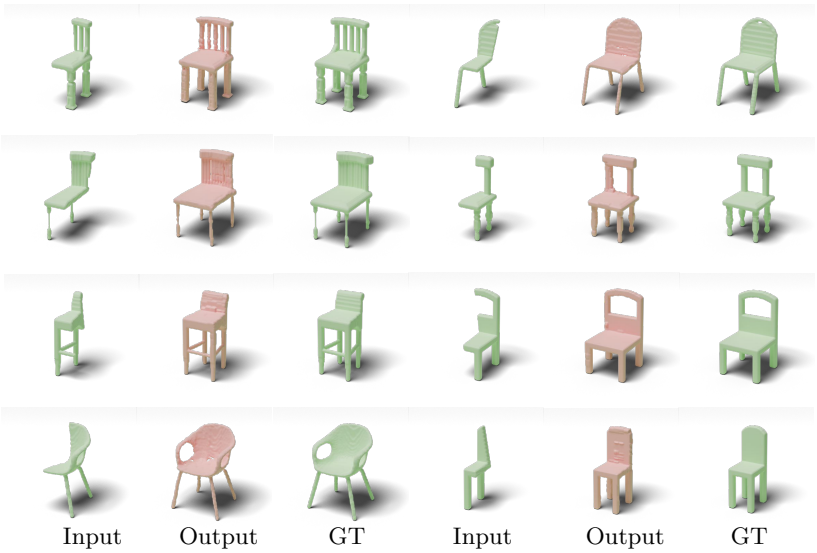

Fig. 4: More shape completion results on shapes with large missing areas.

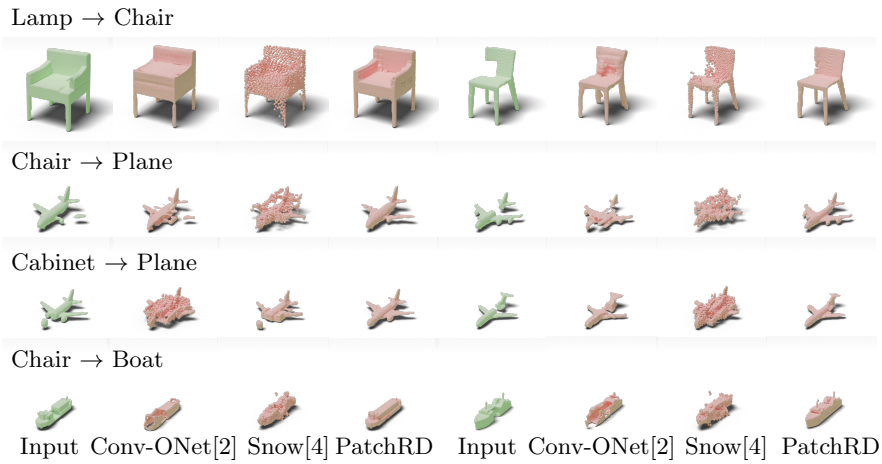

Fig. 5: More testing results on novel categories. For each row, we note the training categories and testing categories on the left top corners. Lamp→Chair means training on lamp and testing on chair shapes.

Fig. 6: More qualitative comparison on PCN Dataset[6].

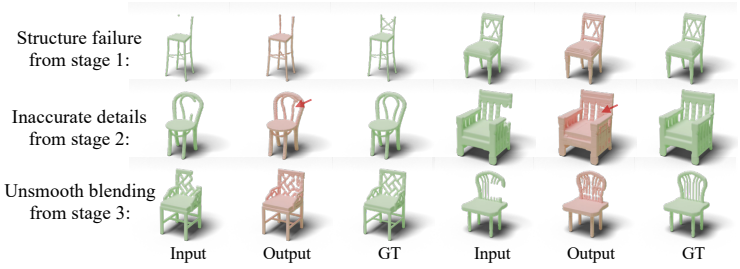

Fig. 7: Failure cases caused by different steps.

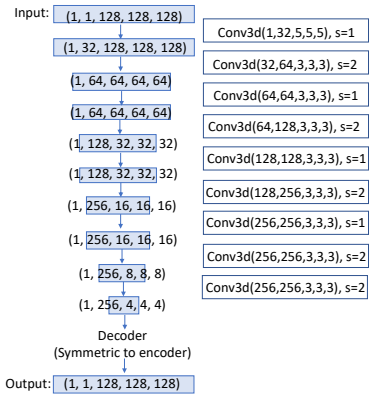

Fig. 8: Architecture of the coarse completion network. The input is a partial shape with size $(128, 128, 128)$ and the output is a coarse shape with the same size. We only show the encoder in detail here. The decoder is symmetric to the encoder. In the figure, blue boxes are tensors and white boxes are layers between two tensors. The array after $Conv3d$ means (input channel, output channel, kernel size, kernel size,kernel size). s means stride.

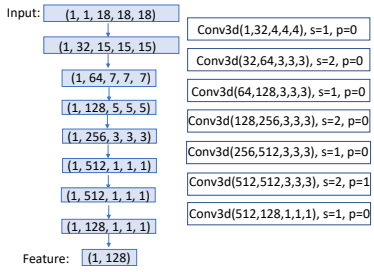

Fig. 9: Architecture of the feature encoder in the retrieval learning part. The input is a patch with size $(18, 18, 18)$. The output is a feature vector with size 128. In the figure, blue boxes are tensors and white boxes are layers between two tensors. The array after $Conv3d$ means (input channel, output channel, kernel size, kernel size,kernel size). s means stride. p means padding.

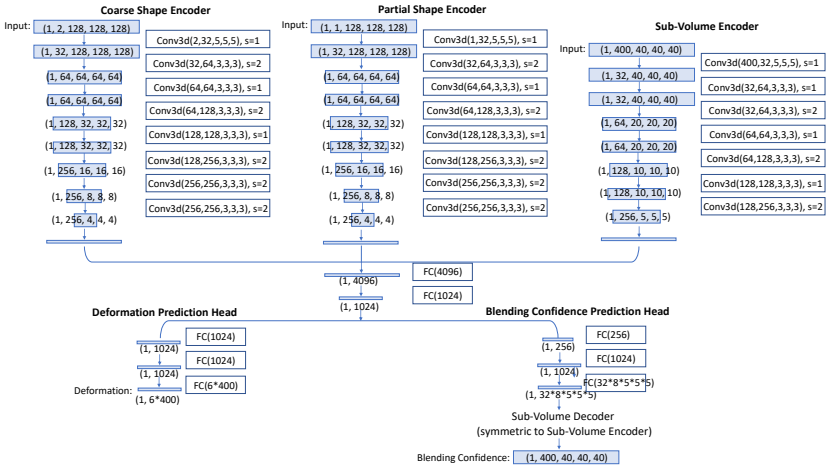

Fig. 10: Architecture of the patch deformation and blending weight prediction network. There are 3 branches to encode the coarse shape, partial shape and the sub-volume to one-dimensional feature vectors. Then two heads decode the concatenated feature vectors to deformation and blending weights. In the figure, blue boxes are tensors and white boxes are layers between two tensors. The array after $Conv3d$ means (input channel, output channel, kernel size, kernel size,kernel size). s means stride.