Supplementary Material

Peng-Shuai Wang
Microsoft Research Asia
penwan@microsoft.com

Yang Liu
Microsoft Research Asia
yangliu@microsoft.com

Xin Tong
Microsoft Research Asia
xtong@microsoft.com

1. Training details for the experiments

For the experiments in Section 4.1 and 4.2, we use the same set of training parameters. Specifically, we set the batch size to 8 and the weight decay to 0.0005, use stochastic gradient descent (SGD) with a momentum of 0.9. The initial learning rate is set as 0.1 and decreases by a factor of 10 after about 6 epochs. The training process stops after 25 epochs.

2. Discussions

Comparison to implicit function-based approaches Recently, implicit functions are used in deep learning as the 3D shape representation [1, 2, 3]. We did not conduct experiments to compare with these methods since they are *orthogonal* to our method: they focus on the shape representation whereas we focus on network structures. Technically, DeepSDF [3] was applied to shape completion by optimizing the latent code to match the partial data while completing the missing part, per shape, in a computationally expensive and memory-costly way. Our method can directly output the shape in one single forward pass. OccNet [2] use the autoencoder architecture directly without skip-connection, the partial input cannot be well preserved. IM-Net [1] has not been tested the completion task.

Ablation study on skip connections l_2 and l_3 in Figure 2 3DEPN is based on a dense U-Net and the decoder of SGC-Net is also a dense network. We regard them as comparable dense networks and do the comparisons with similar amount of parameters and network depth in Section 4. 3DEPN and SGCNet use full skip-connections, including l_2 and l_3. In the ablation studies, our network without l_2 and l_3 achieves better results.

3. Shape reconstruction from a meso-skeleton

To demonstrate the flexibility of the proposed method, we also conduct experiment on the task of reconstructing a complete 3D shape from its meso-skeleton.

Dataset We use the chair and plane datasets provided by [5], which include the synthesized meso-skeletons and the corresponding 3D shapes. The chair dataset contains 889 training and 100 testing pairs, and the airplane dataset contains 626 training and 100 testing pairs. The meso-skeletons are represented as point cloud containing 2048 points, with which we build the octree directly. For the 3D shapes, we use the virtual scanner to convert them into dense point cloud with oriented normals [4], then build the target octrees. The depth of octree is set as 6. The two datasets are trained *separately*, which is the same as P2P-NET.

Implementation details We use the same network as the one used in shape completion. To avoid overfitting, we rotate each skeleton and the corresponding ground-truth object along with the upright axis 12 times for data augmentation. The batch size is set as 24, and the network is trained using SGD with a momentum of 0.9 and a weight decay of 0.0005. The initial learning rate is set as 0.1 and decreases by a factor of 2 after 60 epochs. The training process stops after 120 epochs. We use the Chamfer distance defined in as the evaluation metric.

Comparison We do a comparison with P2P-Net [5]. Since there is no explicit point correspondence between the meso-skeleton and the target shape, P2P-Net relies on a loss function enforcing a shape-wise similarity between the predicted and the target point sets during the training stage to build the correspondence. We directly build the correspondence between the input meso-skeleton and output shape with the proposed skip connections. On the dataset plane and chair, the medians of Chamfer distances are 1.04 and 5.55 for our methods, 1.66 and 6.06 for P2P-Net.

Visual results The visual results are shown in Figure 1. Compared with P2P-Net, the point clouds produced by our method are regularly distributed. Since the point normal is also regressed with the loss function, the output point cloud can be directly used as the input of the Poisson Reconstruction method. The reconstructed meshes are shown in the fourth column of Figure 1. However, it is very hard and even impossible to reconstruct surface mesh from the point cloud of P2P-Net, since the point cloud of P2P-Net is scattered and the internal volume structure of the shape is not kept, which makes it extremely difficult to define the inside and outside for the shape.

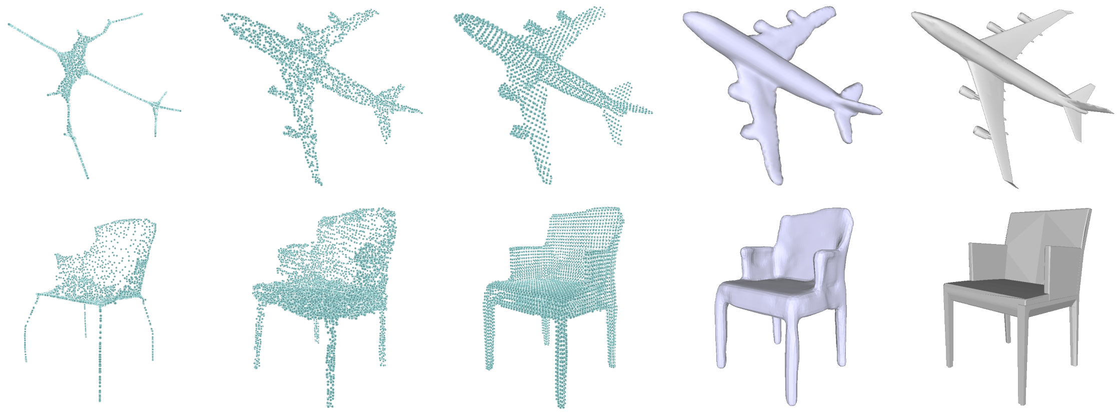

Figure 1: Visual results of shape reconstruction from meso-skeletons. Compared with P2P-Net, the point cloud of our method is regularly distributed. And since the normal is also regressed, the complete meshes can be reconstructed, which is hard or even impossible for the point clouds of P2P-Net.

References

[1] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[2] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[3] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[4] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. *ACM Trans. Graph. (SIGGRAPH)*, 36(4), 2017. 1

[5] Kangxue Yin, Hui Huang, Daniel Cohen-Or, and Hao Zhang. P2P-NET: Bidirectional Point Displacement Net for Shape Transform. *ACM Trans. Graph. (SIGGRAPH)*, 37(4), 2018. 1