Supplementary material for "Unsupervised Learning of Fine Structure Generation for 3D Point Clouds by 2D Projection Matching"

Chao Chen*
School of Software, BNRist, Tsinghua University, China
chenchao19@mails.tsinghua.edu.cn

Zhizhong Han*
Wayne State University
h312h@wayne.edu

Yu-Shen Liu
School of Software, BNRist, Tsinghua University, China
liuyushen@tsinghua.edu.cn

Matthias Zwicker
University of Maryland, College Park
zwicker@cs.umd.edu

1. Network Architecture

The network we leverage for point cloud reconstruction from single images includes a 2D encoder and a 3D point cloud decoder. The 2D encoder is a CNN with 7 layers. The first layer has a 5×5 kernel with 16 channels and a stride of 2. Each of the remaining layers has 3 kernels and comes in pairs, where one layer in the pair has a stride of 2 and the other has a stride of 1. The number of channels grows by a factor of 2 after each strided layer. These convolutional layers are followed by two fully connected layers with a dimension of 1024. The 3D point cloud decoder has one fully connected layer with a dimension of 1024 and then predicts a point cloud. The point cloud of N points is predicted as a vector with a dimension of $3N$ (point coordinates).

2. Single Image Reconstruction

We present more single image reconstruction in this section. We randomly select 100 shapes in the test set in each of Airplane, Car, and Chair classes in Fig. 1, Fig. 2, and Fig. 3, respectively, where each shape is represented by 16000 points. We can see that our method can reconstruct shapes with arbitrary topologies in high accuracy.

3. Novel Shape Generation

We present more generated novel shapes in this section. We randomly sample 100 noise in the 32 dimensional Gaussian space, which is further used to generate 100 novel shapes using the trained point decoder. We demonstrate the generated shapes with 16000 points in Fig. 4, which shows that our method can generate plausible and reasonable shapes.

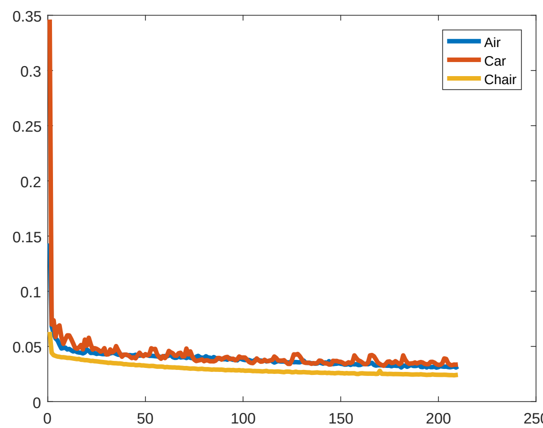

Figure 5. Training loss under Airplane, Car and Chair.

4. Loss visualization

We visualize the loss curves employed during the training with 16000 points under Airplane, Car, and Chair dataset in Fig. 5. We can see that our method trains a network to stably achieve to a minimum, which produces our good results.

5. Source Code

Our code, data and models are available at https://github.com/chenchao15/2D_projection_matching

*indicates the equal contribution. This work was supported by National Key R&D Program of China (2020YFF0304100), the National Natural Science Foundation of China (62072268), and in part by Tsinghua-Kuaishou Institute of Future Media Data, and NSF (award 1813583). The corresponding author is Yu-Shen Liu.

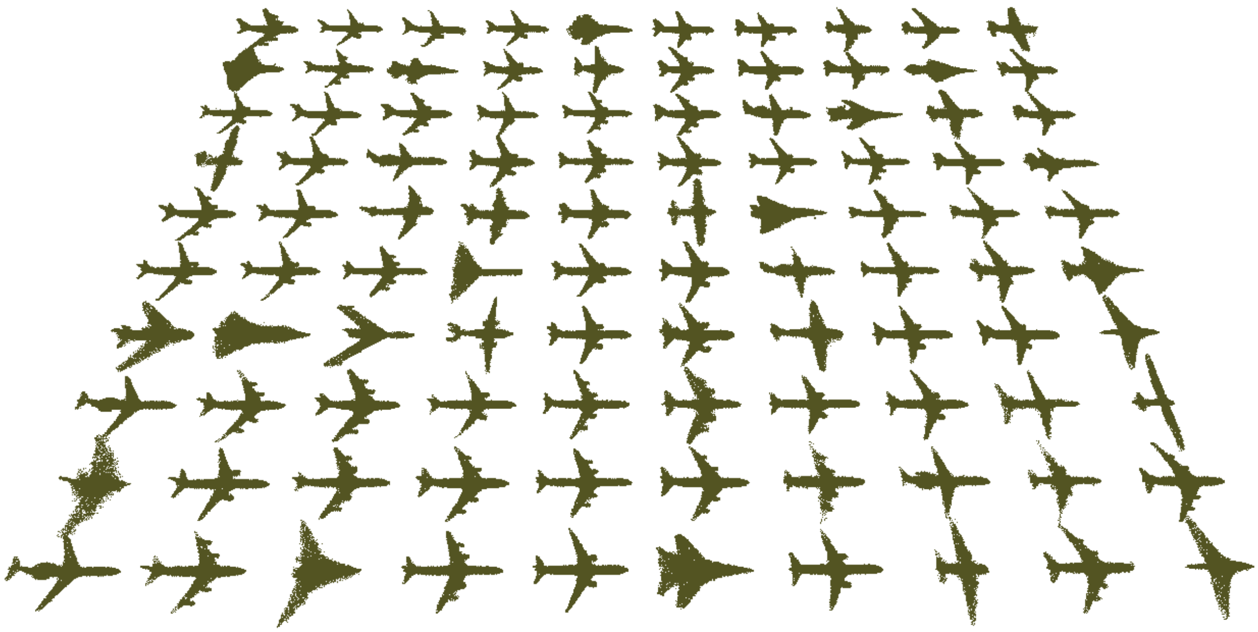

Figure 1. More single image reconstructions under Airplane class.

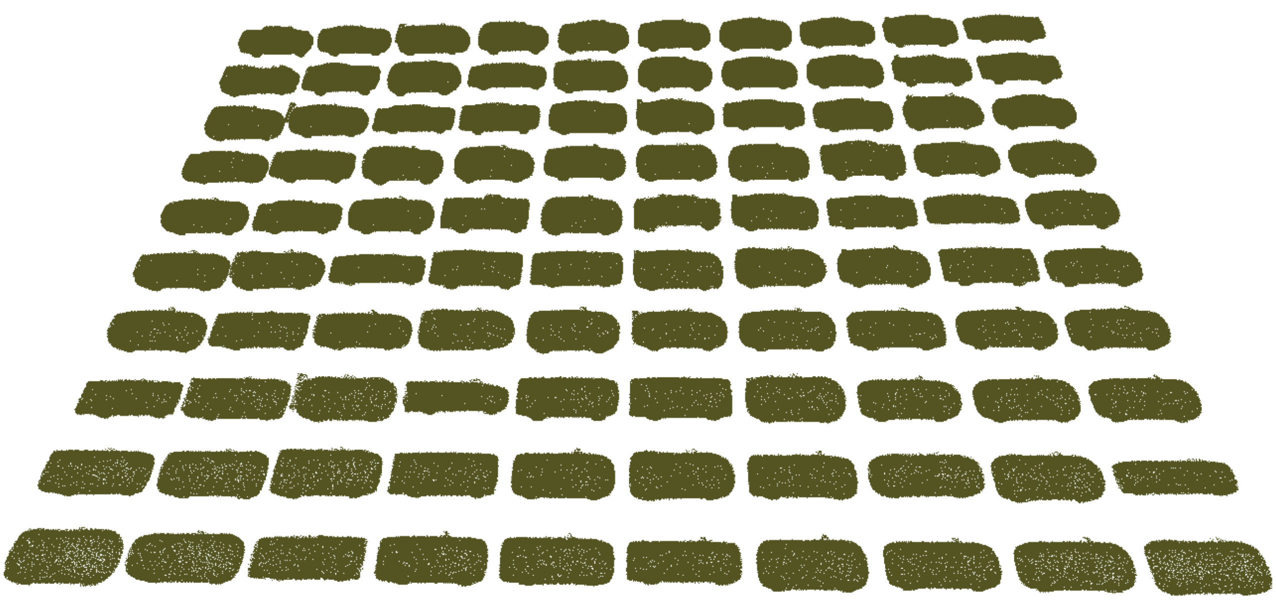

Figure 2. More single image reconstructions under Car class.

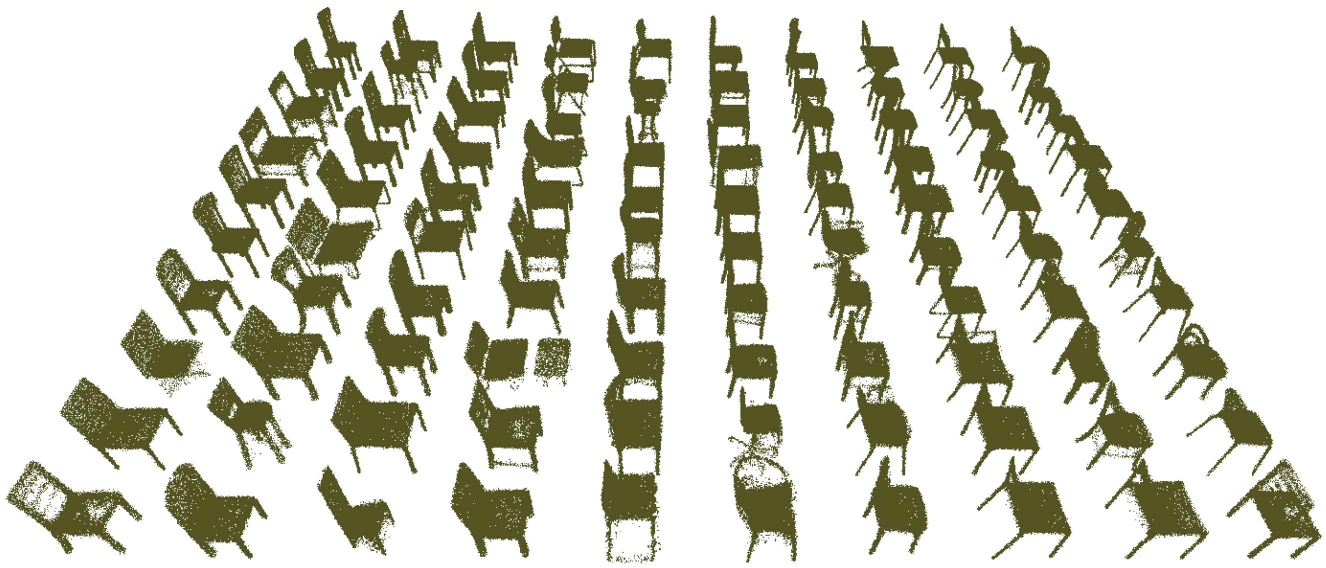

Figure 3. More single image reconstructions under Chair class.

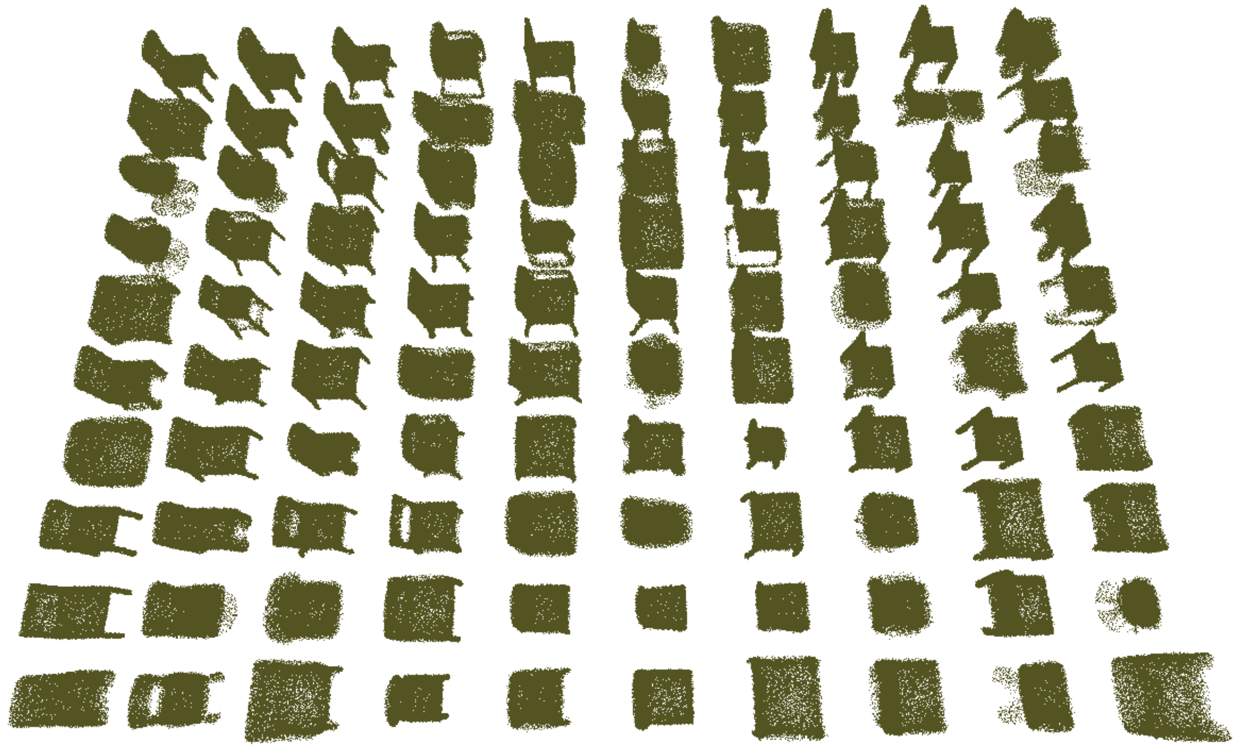

Figure 4. More novel chairs generated by our method.