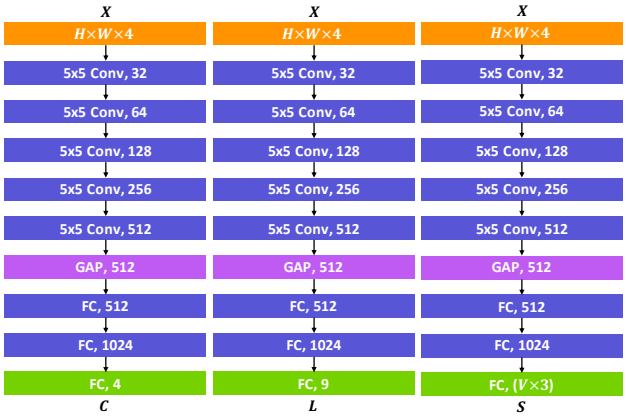
Supplementary of Self-Supervised 3D Mesh Reconstruction from Single Images

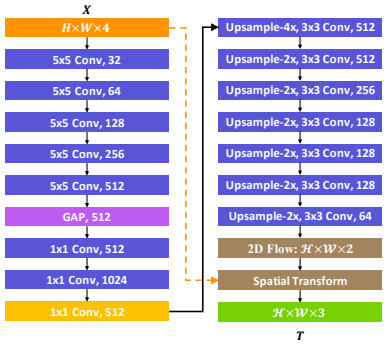
Tao Hu[1] Liwei Wang[1] Xiaogang Xu[1] Shu Liu[2] Jiaya Jia[1,2]
[1] The Chinese University of Hong Kong [2] SmartMore
{taohu,lwwang,xgxu,leojia}@cse.cuhk.edu.hk, sliu@smartmore.com

1. Network Structure

The input data $X = [I, M] \in \mathbb{R}^{H\times W\times 4}$ is concatenated by the original RGB single image $I \in \mathbb{R}^{H\times W\times 3}$ and its silhouette mask $M \in \mathbb{R}^{H\times W\times 1}$. Our 3D mesh reconstruction model consists of 4 sub-encoders, *i.e.* Camera Encoder E_c, Light Encoder E_l, Shape Encoder E_s, and Texture Encoder E_t. These sub-encoders separately encode corresponding attribute feature to avoid mutual effect, as illustrated in Fig. 1. The landmark feature extractor E_f is the conventional U-Net network whose output resolution is equal to the input resolution $H \times W$ and output channel is 128. The landmark classification network is a MLP network with three layers whose output dimension is V.



(a) Camera Encoder. (b) Light Encoder. (c) Shape Encoder.



(d) Texture Encoder.

Figure 1: Reconstructed Model.

2. Limitations and Failure Cases

Although our SMR can effectively reconstruct 3D mesh object on multiple category-specific datasets with only 2D silhouette annotations, there are still two main limitations to be overcome in the near future: *Limitation 1: Silhouette Annotations.* It still requires silhouette annotations. For the sake of simplification, we have not considered the influence of background information. For some category-specific objects, such as horse, we annotate their silhouette annotations by detectron2. Thus the incorrect annotations will influence the reconstructed results. The failure case is illustrated in Fig. 2. *Limitation 2: Non-Rigid Objects.* It is not fit for non-grid objects, like the humans or flowers. Since it is difficult to determine the canonical viewpoints and the topological limitations of mesh representation. To avoid these limitations, we will further predict the silhouette masks by self-supervised learning and build a fully unsupervised 3D reconstruction model for more general objects.

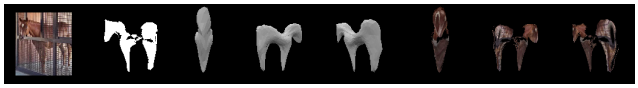


Figure 2: Failure case: The incorrect silhouette of the horse will affect the accuracy of reconstructed object.

3. More Reconstruction Results

3D Reconstruction On ShapeNet On the Shapenet dataset, since our SMR aims to model category-specific object, we perform experiments on these 13 categories one-by-one. We introduce the ground truth camera parameters so as to evaluate the reconstructed accuracy and compare with other supervised methods to demonstrate SMR's effectiveness. The visualization results of our results on ShapeNet are shown in Fig. 3.

3D Reconstruction In The Wild We present more reconstructed object in the wild to demonstrate the generalization of our SMR, as shown in Fig. 4, Fig. 5, and Fig. 6.

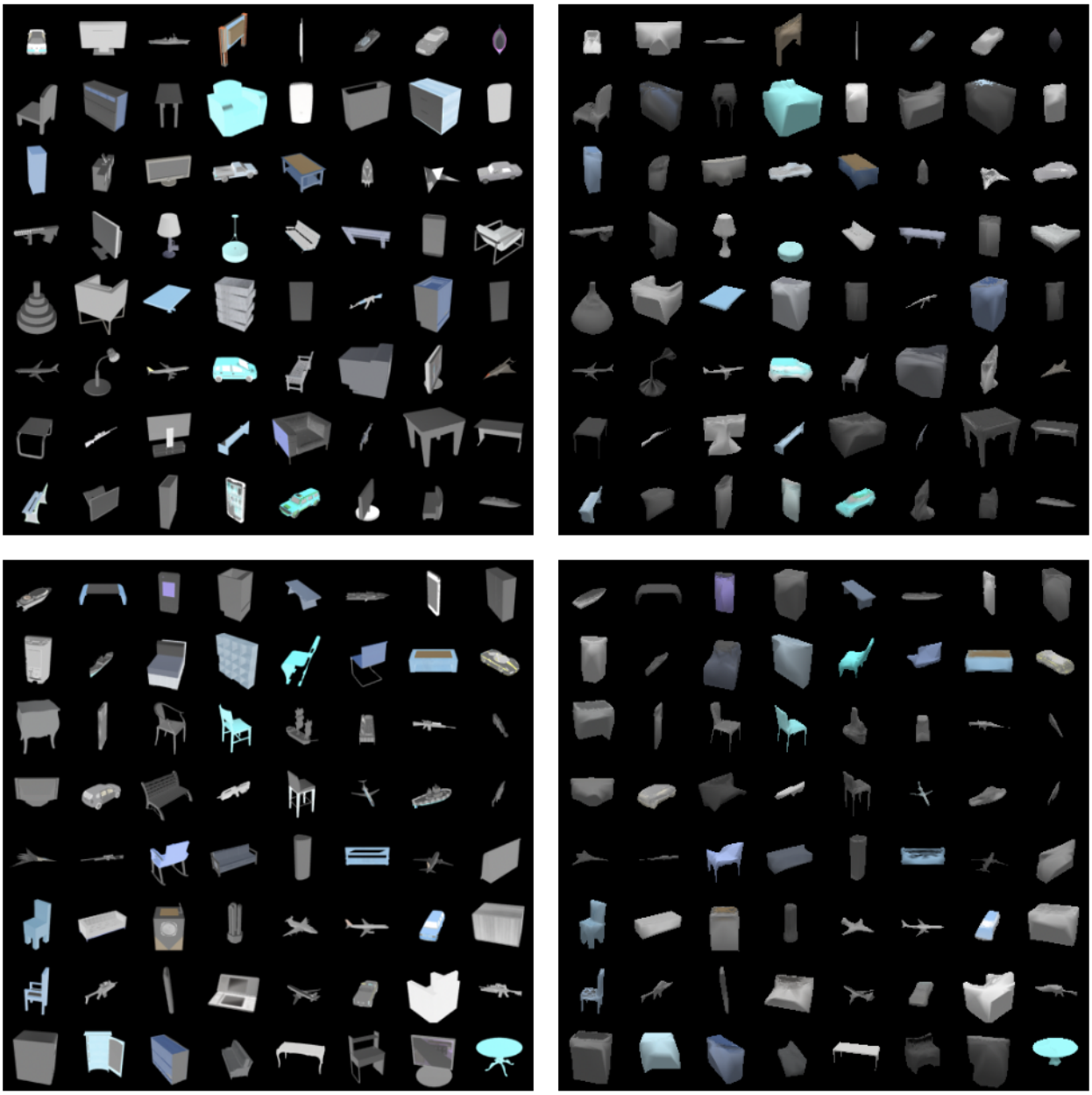


Figure 3: 3D Object Reconstruction on ShapeNet by our SMR.

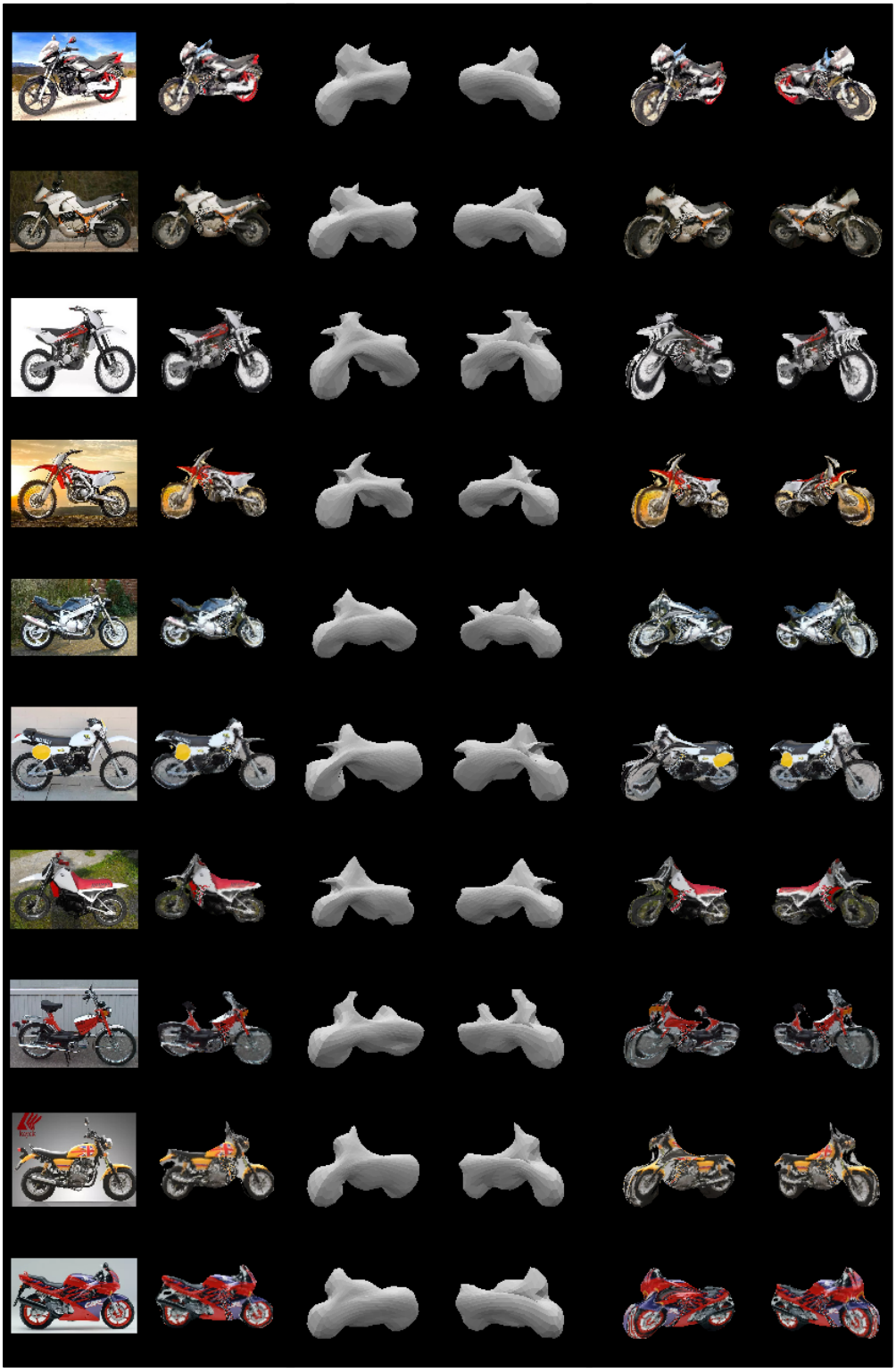


Figure 4: 3D Motorbike Reconstruction in the wild.

Figure 5: 3D Cow Reconstruction in the wild.

Figure 6: 3D Horse Reconstruction in the wild.