STRUCTURE-GUIDED SINGLE VIEW 3D RECONSTRUCTION

Haoliang Jiang Fang Shu Xiaobo Wang

1 INTRODUCTION

Humans can easily recognize the object from a single 2D image and imagine the 3D shape in their mind, even if parts of the object are hidden. It is because after human see the object, they will have an idea of what the object is and what the general shape and features it has based on existing information in the image. From this insight, we would like to improve the current state-of-art model of reconstructing 3D shape from a single image called GenRe [1]. Their approach uses the 2.5D image to generate the reconstructed model which does not make enough sense because depth can't show the back of the object. To handle this problem, we propose a structure-guided reconstruction method that takes advantage of both detailed recoveries using depth cue and structure recovery using prior constraints based on recent research of human perception.

Our approach offers some unique advantages. First, our estimated 3D structure encodes the symmetry, compactness and planarity constraints explicitly of given objects, which can help us understand the reconstruction in a more transparent way. As some papers argue previous black-box neural networks suffer from domain adaption [2] because of memorization and overfitting. Second, the combination of surface and depth information ensures the reconstruction contains full details of varied objects sharing similar structure as we want to synthesize 3D object.

2 METHODS

The GenRe method integrates these representations for generalizable, high-quality 3D shape reconstruction. Experiments demonstrate that GenRe achieves state-of-the-art performance on shape reconstruction for both seen and unseen classes. Building upon this base-line model, we adopt a new branch starting from the original image which predicts the structure of the object from the original image using Network D, shown in Fig.1. It will then be added to the input of Network B and Network C as spherical maps and voxel data. Specifically, network D is implemented from im2struct [3] . It can convert a 2D RGB image to a 3D simple structure which contains a simple mesh surface. We then convert it to spherical maps and voxel data which are shown in Figure 2 and Figure 3 respectively, matching the input data structure of network C and network D. We concatenate the newly generated structure with the UV map generated from the depth image to strengthen the structure-guided concept in this network. In this way, the information of the full 3D model can be embedded into the UV map from depth information, so that network can actually see the unseen parts.

3 RESULTS AND DISCUSSION

In general, our model outputs reconstruction results of good quality in the class of chair. Furthermore, in cases where occupancy happens in an image, our model tends to predict a complete chair instead of a chair with an incomplete structure. Also, because of the additional structure information, our model generates 3D models showing apparent structures when the object structure in an image is not clear. Figure 4 shows the comparison of three predictions between our model and the baseline model. 3D objects are displayed by vertexes so that we can see the entire structure of the reconstruction shape.

As shown in the case a in Figure 4, although our model predicts the position of legs correctly, it is still hard for it to predict the correct length and shape of the leg. We assume this is because the camera poses. Images in our dataset are shot by different poses. In contrast, for the Im2struct [3] dataset, only the canonical pose is utilized. The difference in rotation orientations of the predictions in two streams might end up with ambiguous geometry information in reconstruction. Moreover, we also test the model on the image from an unseen class. As a result, the reconstruction quality is limited by the fact that only the class of chair is available for structure prediction dataset.

4 CONCLUSIONS

We find our structure network successfully predicts the general shape of the object from the original 2D image, as hidden parts can be seen in the 3D mesh model. After using the structure information, the UV map will contain the back of the object, demonstrating the effectiveness of our method. We show our structure-guided network has achieved better performance compared with the baseline model and has more precise predicted 3D model especially when parts of the object are unseen in the 2D image. However, our model fails to recover structures for object categories unseen from the training set. For further work, we will show its efficiency in multiple classes instead of a single class of chair. Also, T-net or other techniques could be introduced to the network to solve the problem.

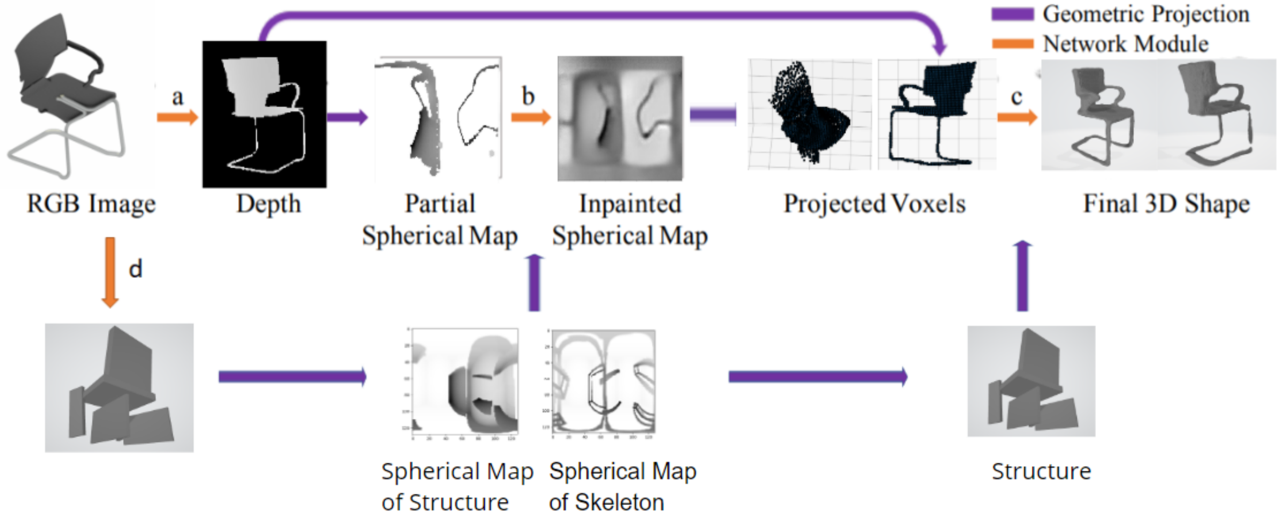

FIGURE 1. Proposed network architecture. We use two steams in our architecture for reconstruction. The upper stream uses the main idea proposed by Zhang et al. [1]. The bottom stream is introduced to capture structure information. Network D is built upon [3].

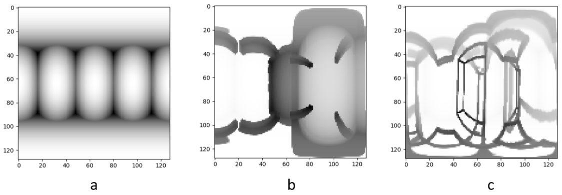

FIGURE 2. Spherical maps. a is an example of spherical map of a unit cube. b is an example of the spherical map of a solid structure of a chair. c is an example of the spherical map of a skeleton structure of a chair.

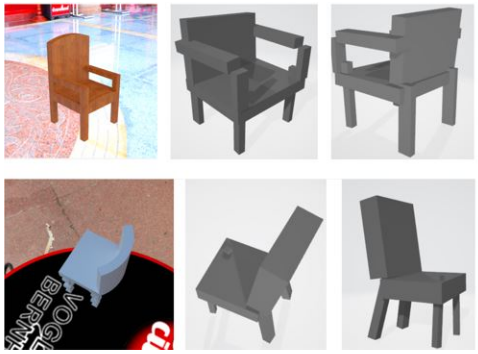

FIGURE 3. Output of network D: the left column is the original image, and the two following comluns are the resulting 3D mesh model. The hidden parts in the image can be generated.

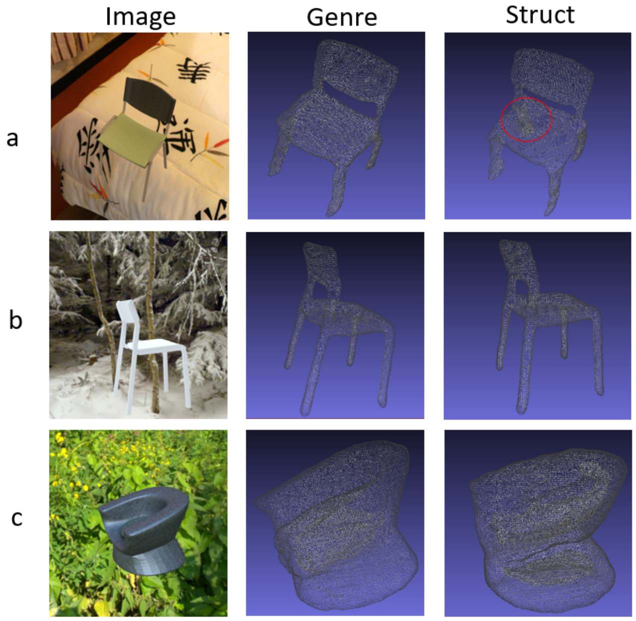

FIGURE 4. Comparison of 3D shapes generated by the baseline and our model. 3D objects are displayed using only vertexes for visualization.

REFERENCES

[1] Zhang, X. et al. Learning to Reconstruct Shapes from Unseen Classes. In *Advances in Neural Information Processing Systems (NIPS)* (2018).

[2] Wu, Jiajun and Wang, Yifan and Xue, Tianfan and Sun, Xingyuan and Freeman, Bill and Tenenbaum, Josh Marrnet: 3d shape reconstruction via 2.5 d sketches. *Advances in neural information processing systems* , page 540–550 (2017)

[3] Niu C, Li J, Xu K: Im2Struct: Recovering 3D shape structure from a single RGB image. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018).

[4] J. Li, K. Xu, S. Chaudhuri, E. Yumer, H. Zhang, and L. Guibas. Grass: Generative recursive autoencoders for shape structures. arXiv preprint arXiv:1705.02090, 2017.