SymmNeRF: Learning to Explore Symmetry Prior for Single-View View Synthesis Supplementary Material

Xingyi Li[1], Chaoyi Hong[1], Yiran Wang[1], Zhiguo Cao[1], Ke Xian[2*], and Guosheng Lin[2]

[1] Key Laboratory of Image Processing and Intelligent Control, Ministry of Education
School of AIA, Huazhong University of Science and Technology, China
{xingyi_li,cyhong,wangyiran,zgcao}@hust.edu.cn
[2] S-lab, Nanyang Technological University
{ke.xian,gslin}@ntu.edu.sg

In this document, we provide additional visualization results, the network architecture of SymmNeRF as well as limitations and failure cases.

1 Additional Visualization Results

In this section, we provide additional qualitative results of the main paper.

1.1 Qualitative Comparisons on the ShapeNet-SRN Dataset

As shown in Fig. 1 and Fig. 2, compared with SRN [6] and PixelNeRF [8], our method can synthesize more photo-realistic and reasonable novel views with fine-grained details close to ground truths.

1.2 Novel View Synthesis on the ShapeNet-SRN Dataset

We further provide more visualization of novel view synthesis results in Fig. 3 and Fig. 4. As can be seen, SymmNeRF can always synthesize photo-realistic and reasonable novel renderings from totally different viewpoints. With the help of the symmetry priors and the hypernetwork, SymmNeRF accurately recovers the geometry information and texture details despite the occlusions in the reference view.

1.3 Generalization Results on Real-World Datasets

Here we provide additional generalization results on the real-world Pix3D [7] and Stanford Cars [3] datasets in Fig. 5 and Fig. 6. Compared with PixelNeRF [8], SymmNeRF can effectively infer the geometry and appearance of real-world chairs and cars.

* Corresponding author

1.4 Qualitative comparisons on the ShapeNet-NMR dataset under the category-agnostic single-view reconstruction setting

We provide additional qualitative comparisons on the ShapeNet-NMR [1,2] dataset under the category-agnostic single-view reconstruction setting. We show in Fig. 7 and Fig. 8 that SymmNeRF outperforms other state-of-the-art methods [4,5,6,8]. This also implies that symmetry priors benefit the reconstruction of almost all symmetric objects, and that our method can also deal with objects that are not perfectly symmetric. This is because a few asymmetric objects are also included in the training dataset. Our model can perceive and recognize asymmetry thanks to the global latent code and hypernetwork. SymmNeRF therefore adaptively chooses to utilize local features to reconstruct asymmetric objects.

1.5 Ablation Study

We also show more qualitative evaluation of different configurations of our method on the ShapeNet-SRN [1,6] dataset in Fig. 10. The baseline model (a) tends to render smoothly. Simply using pixel-aligned image features (b) still fails to fully understand 3D structure. In contrast, our full model (c) reproduces photo-realistic details from most viewpoints. The rendering quality of (d) deteriorates as the hypernetwork is not adopted. We have to emphasize that, *only including both the symmetry priors and the hypernetwork can accurately recovers the geometry information and texture details despite the occlusions.*

2 Network Architecture of SymmNeRF

Here we visualize the network architecture of SymmNeRF in Fig. 9.

3 Limitations and Failure Cases

Although symmetry can benefit single-view view synthesis, our method still suffers from some limitations. One is that when the object is not perfectly symmetric, symmetry priors may sometimes lead to erroneous rendering, see Fig. 11. This depends on how informative the reference view is. Another one is that the model trained on single-object scenes may not handle multiple-object scenes, because multiple objects, as a whole, are usually asymmetric.

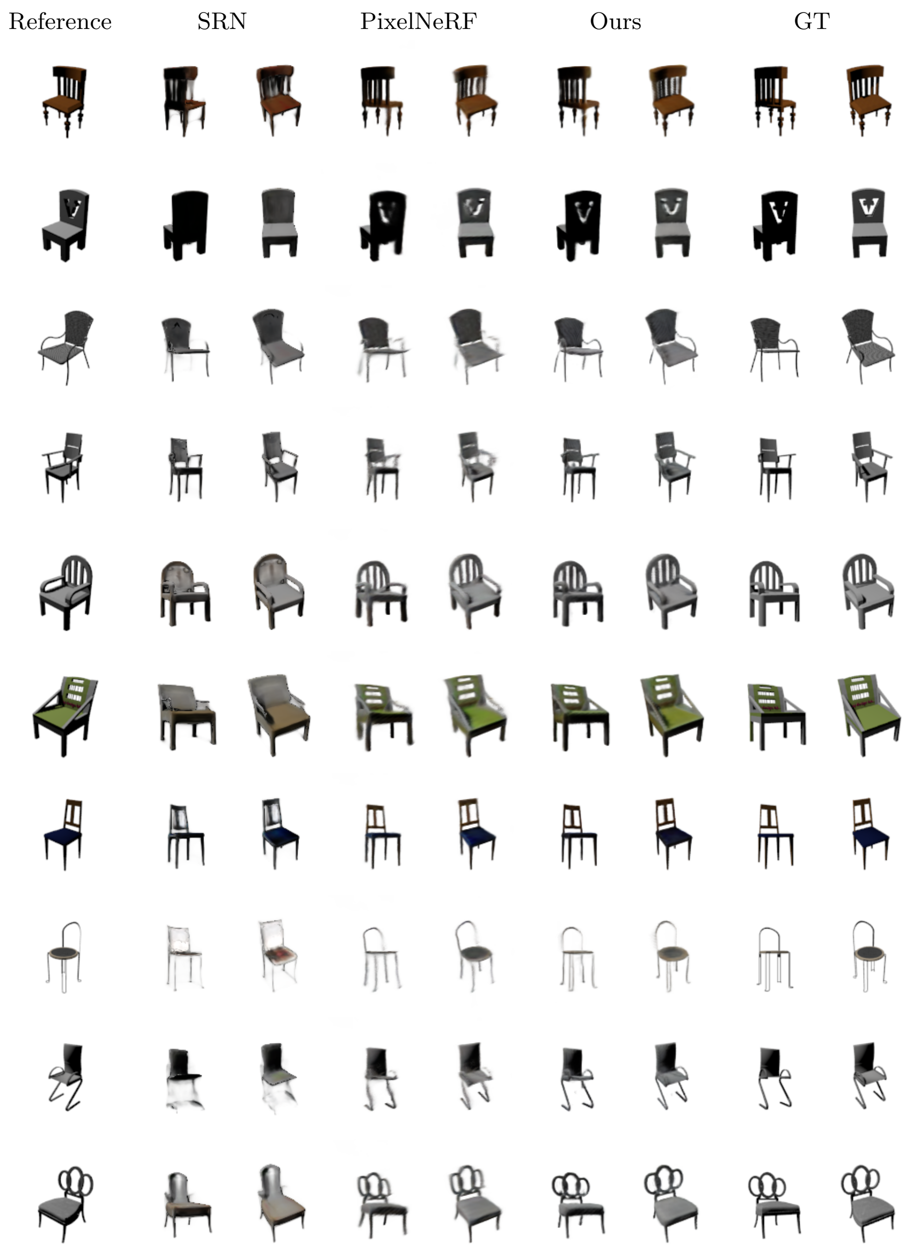

Fig. 1: Additional qualitative comparisons on the "Chairs" category of the ShapeNet-SRN [1,6] dataset. Compared with SRN [6] and PixelNeRF [8], SymmNeRF yields more photo-realistic and reasonable novel views with fine-grained details close to ground truths.

Fig. 2: Additional qualitative comparisons on the "Cars" category of the ShapeNet-SRN [1,6] dataset. Compared with SRN [6] and PixelNeRF [8], SymmNeRF yields more photo-realistic and reasonable novel views with fine-grained details close to ground truths.

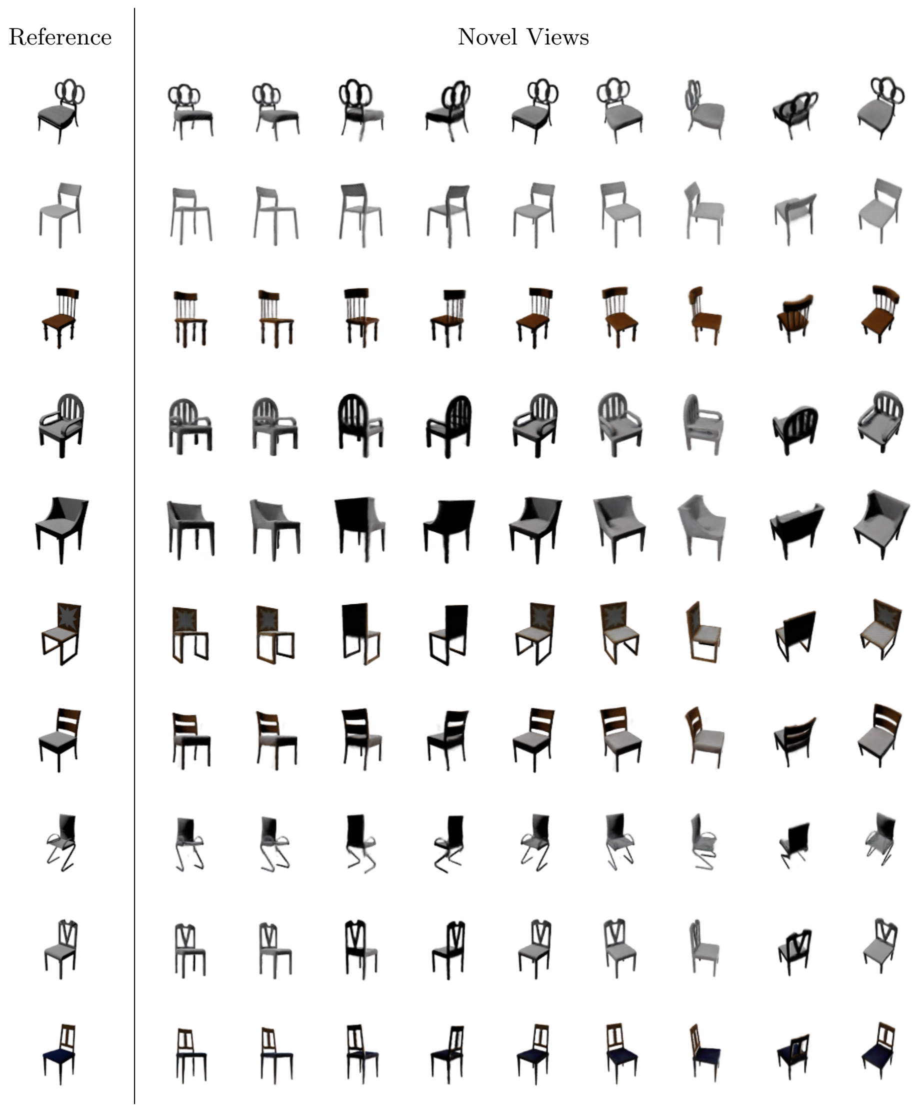

Fig. 3: Additional novel view synthesis results on the "Chairs" category of the ShapeNet-SRN [1,6] dataset. As can be seen, ours can always synthesize photo-realistic and reasonable novel renderings from totally different viewpoints.

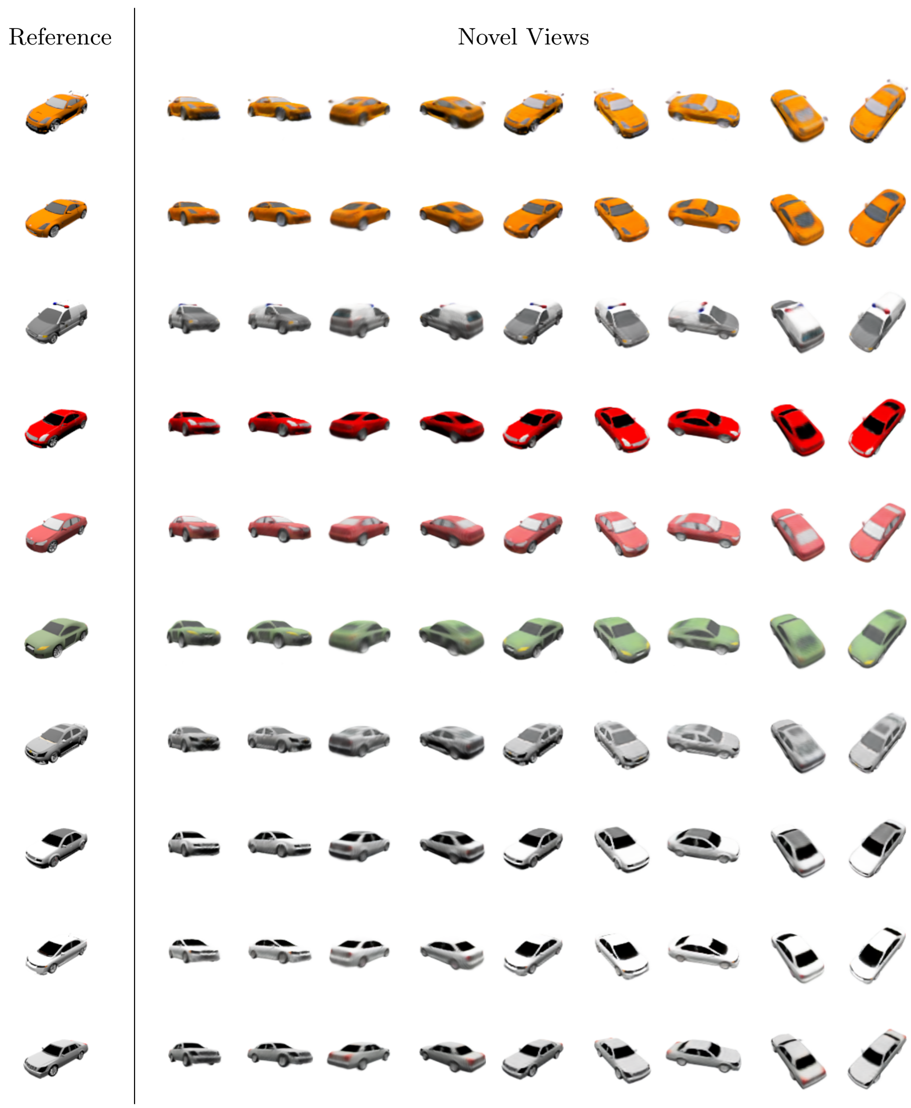

Fig. 4: Additional novel view synthesis results on the "Cars" category of the ShapeNet-SRN [1,6] dataset. As can be seen, ours can always synthesize photo-realistic and reasonable novel renderings from totally different viewpoints.

Fig. 5: Additional qualitative comparisons with PixelNeRF [8] on the real-world Pix3D [7] dataset. Compared with PixelNeRF, SymmNeRF yields better generalization.

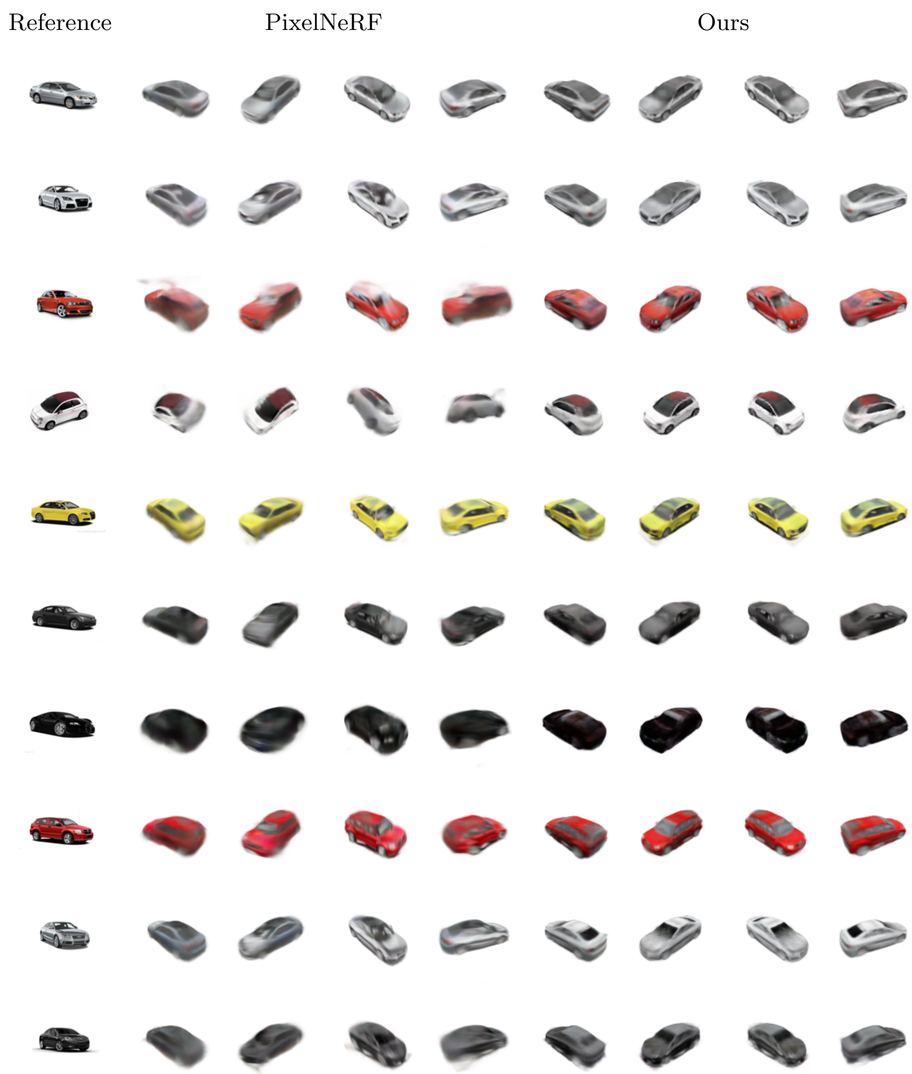

Fig. 6: Additional qualitative comparisons with PixelNeRF [8] on the real-world Stanford Cars [3] dataset. Compared with PixelNeRF, SymmNeRF yields better generalization.

Fig. 7: Addtional qualitative comparisons on the ShapeNet-NMR [1,2] dataset under the category-agnostic single-view reconstruction setting.

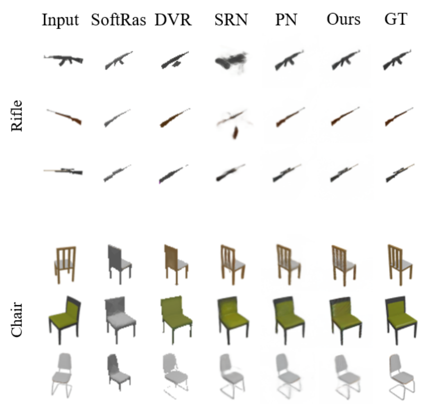

Fig. 8: Addtional qualitative comparisons on the ShapeNet-NMR [1,2] dataset under the category-agnostic single-view reconstruction setting.

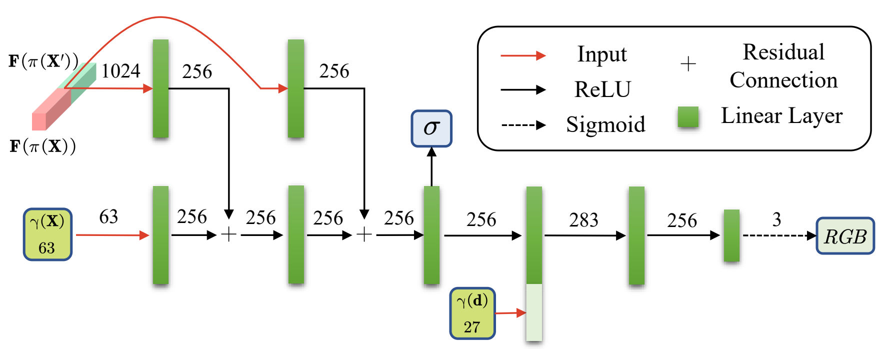

Fig. 9: SymmNeRF architecture. $\mathbf{X}' \in \mathbb{R}^3$ is the corresponding symmetric 3D point of $\mathbf{X}$, π denotes the process of projecting the 3D point onto the image plane using known intrinsics, $\mathbf{F}$ is the feature volume extracted by the image encoder network f, and $\gamma_{\mathbf{X}}(\cdot)$ and $\gamma_{\mathbf{d}}(\cdot)$ are positional encoding functions for spatial locations and viewing directions, respectively.

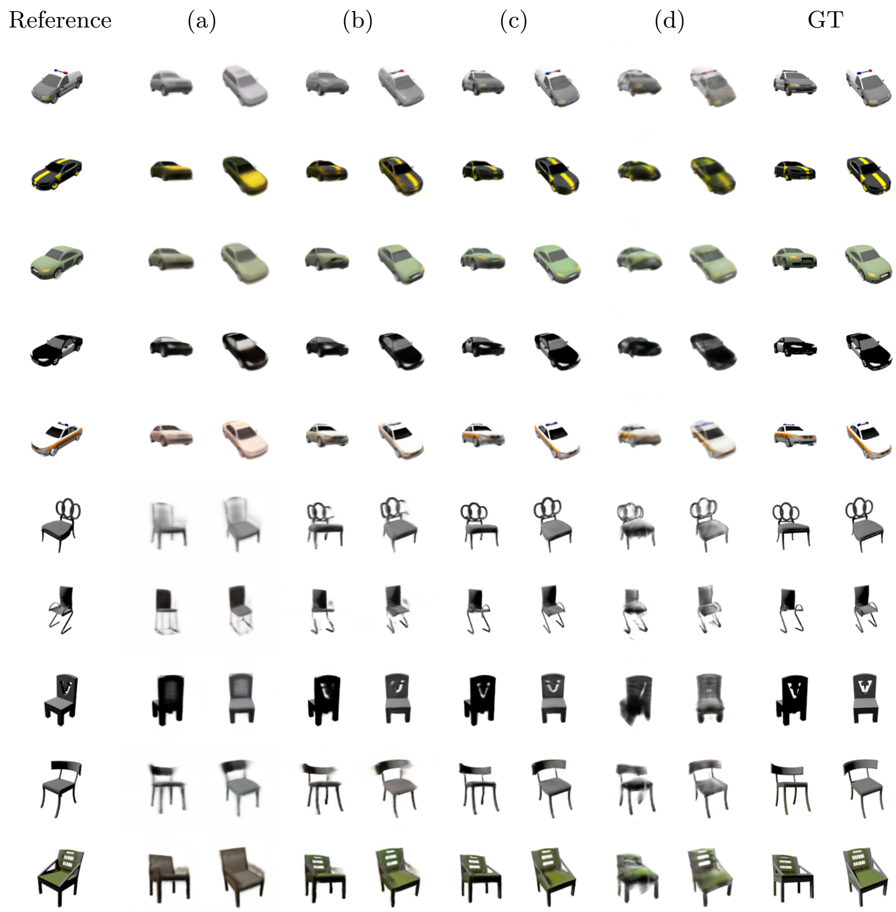

Fig. 10: Additional qualitative evaluation of different configurations of our method on the ShapeNet-SRN [1,6] dataset. (a): a minimalist version of our method only including the image encoder network and the hypernetwork, without taking pixel-aligned features and symmetric features as input to the neural radiance field. (b): adding pixel-aligned image features, compared to (a). (c): adding pixel-aligned and symmetric image features, compared to (a). (d): removing the hypernetwork, compared to (c).

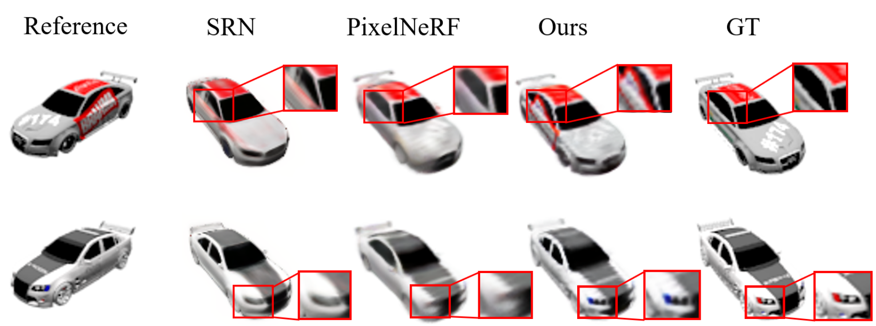

Fig. 11: Failure cases. When the object is not perfectly symmetric and the reference view is not informative enough, symmetry priors may sometimes lead to erroneous rendering.

References

1. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012 (2015) 2, 3, 4, 5, 6, 9, 10, 11
2. Kato, H., Ushiku, Y., Harada, T.: Neural 3d mesh renderer. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018) 2, 9, 10
3. Krause, J., Stark, M., Deng, J., Fei-Fei, L.: 3d object representations for fine-grained categorization. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 554–561 (2013) 1, 8
4. Liu, S., Li, T., Chen, W., Li, H.: Soft rasterizer: A differentiable renderer for image-based 3d reasoning. The IEEE International Conference on Computer Vision (ICCV) (Oct 2019) 2
5. Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A.: Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3504–3515 (2020) 2
6. Sitzmann, V., Zollhoefer, M., Wetzstein, G.: Scene representation networks: Continuous 3d-structure-aware neural scene representations. Advances in Neural Information Processing Systems **32**, 1121–1132 (2019) 1, 2, 3, 4, 5, 6, 11
7. Sun, X., Wu, J., Zhang, X., Zhang, Z., Zhang, C., Xue, T., Tenenbaum, J.B., Freeman, W.T.: Pix3d: Dataset and methods for single-image 3d shape modeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2974–2983 (2018) 1, 7
8. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4578–4587 (2021) 1, 2, 3, 4, 7, 8