

Learning Relighting and Intrinsic Decomposition in Neural Radiance Fields

Supplementary Material

In this supplementary material, we present the following:

1. More detailed procedures for generating pseudo labels.
2. Specifications of experimental settings.
3. Additional qualitative results.

1. Pseudo Label Generation

Here, we elaborate on the post-processing steps (Fig. 1) in the main paper Sec. 3.2. It starts with generating pseudo shading based on Lambertian reflection principles. Under the assumption that the light intensity and color remain constant, shading can be approximated by the dot product between the normal ray and the light ray. The light ray encompasses both direct/indirect illumination and necessitates light visibility to account for occlusion effects.

1.1. Step A: obtain normal and light visibility

The normals are derived from the SDF network. The geometry network also provides depth information which is used to estimate the intersection points in conjunction with sphere tracing[1]. Light visibility, which indicates whether a point is directly illuminated, is obtained by sphere tracing based on the light position and intersection points.

1.2. Step B: generate pseudo shading

The generation of pseudo-shading follows the formula,

$$S' = ((\vec{N} \cdot \vec{L}) \cdot V)^\gamma \quad (1)$$

where the optimal shading S' is the multiplication of the light visibility V and the dot product of the normal \vec{N} and the light ray \vec{L} . $(\cdot)^\gamma$ represents for gamma correction. This correction is crucial because the human eye’s perception of brightness is not linear. Most images we see have undergone gamma correction to accommodate this perceptual effect. Therefore, calculating shading also necessitates gamma correction, yielding to our defined pseudo shading.

1.3. Step C: generate pseudo reflectance

This step entails inferring the most probable pseudo reflectance from the pseudo shading, principally based on the equation $R = I/S$. The approach has two main points that should be noted here.

First, the current pseudo shading only considers direct light. As seen in previous papers [2, 7, 10], solving for indirect light is a complex and computationally expensive process. Our novel approach leverages the trained model to generate multiple versions of images under different lighting conditions, each accompanied by respective pseudo

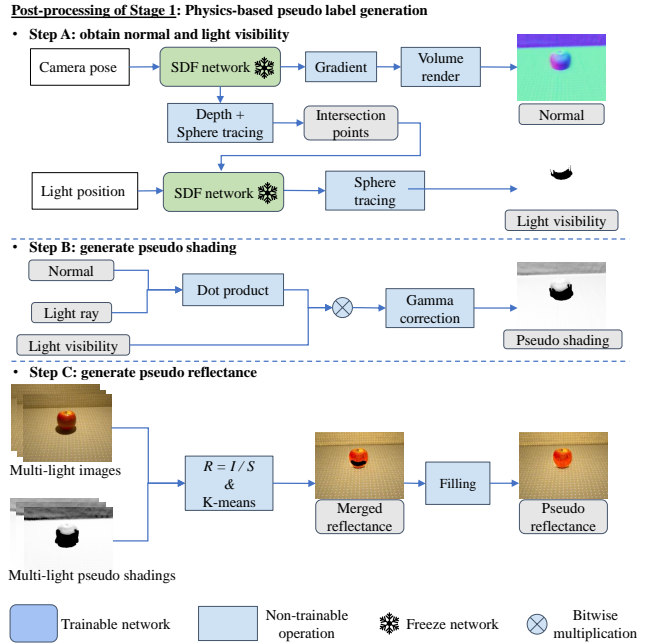


Figure 1. Post-processing and generating pseudo labels.

shadings. As direct light strengthens on a pixel, the influence of indirect light diminishes, making the reflectance derived from higher pseudo shading values more reliable. We compare the outcomes under multiple lighting conditions and synthesize the most credible reflectance for each pixel based on the intensity of pseudo shading.

Second, the residual term includes specular and other effects that are not considered in $R = I/S$. Specular highlights, which have high pseudo shading values, do not reflect the object color but rather the light source color (e.g., white reflections). By analyzing different lighting conditions, where highlights typically vanish except under specific angles, we can deduce the object color by selecting the most common reflectance outcomes.

Our implementation employs the K-means algorithm, incorporating the weights of pseudo shading. This approach allows us to achieve a merged reflectance under varied lighting conditions, as shown in the intermediate result at the bottom in Fig. 1. However, some regions within the merged reflectance may appear vacant due to the absence of direct illumination in all lighting conditions. So, we address these areas with a filling strategy. This strategy specifically considers the distance between void and non-void pixels, their normals, and their colors in the RGB image, thereby achieving the final pseudo reflectance.

Additionally, we compute weight maps W_R and W_S for

both pseudo reflectance and pseudo shading based on the edges of pseudo shading and visibility. Areas with higher pseudo shading values, or those further from visibility edges (where visibility calculations may be prone to errors), exhibit greater credibility in their pseudo labels; conversely, areas closer to visibility edges or with lower pseudo shading values are deemed less reliable.

2. Experimental Settings

Datasets. To validate our approach, we conduct experiments on both synthetic and real-world datasets.

For the synthetic dataset, models are obtained from NeRF [4], with lighting configurations borrowed from Zeng et al. [8]. To facilitate quantitative analysis, GT for reflectance, shading, and residuals are rendered in Blender. Each scene comprises 500 images for training, 100 for validation, and 100 for testing, including intrinsic components for each image. Importantly, adhering to the configurations in [8], the settings for lighting and camera poses are managed independently.

The real dataset we use is the ReNe dataset [5], where lighting and camera poses are grid-sampled. This dataset features 2000 images across scenes, captured from 50 different viewpoints under 40 lighting conditions. Following their dataset split, we use 1628 images (44 camera poses \times 37 light positions) for training.

Additionally, given that the settings of lights and cameras are dependent on the former one and grid-sampled in the latter, our proposed method is designed to accommodate both configurations.

Metrics. To evaluate the comparison between predicted images and ground truth (GT), we employ the following metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [6], and Learned Perceptual Image Patch Similarity (LPIPS) [9].

Implementation details. Our model’s hyperparameters include a batch size of 2048 and each stage was trained for 500k iterations. We implemented the model in PyTorch and used the AdamW [3] optimizer with a learning rate of $1e^{-3}$ for optimization. The experiments can be conducted on a single Nvidia RTX 3090 or A40 GPU. The weights of losses, w_{eik} , w_{curv} , $w_{\text{intrinsic}}$, w_{reg} are set to 0.1, $5e^{-4}$, 1.0, and 1.0, respectively.

3. Additional qualitative results

We present additional results in this section. Fig. 2 displays additional examples comparing our method with others. Fig. 3 - Fig. 8 shows more qualitative results of our method on the ReNe dataset. Furthermore, Fig. 9 - Fig. 11 demonstrate additional qualitative results of our method on real scenes from [8].

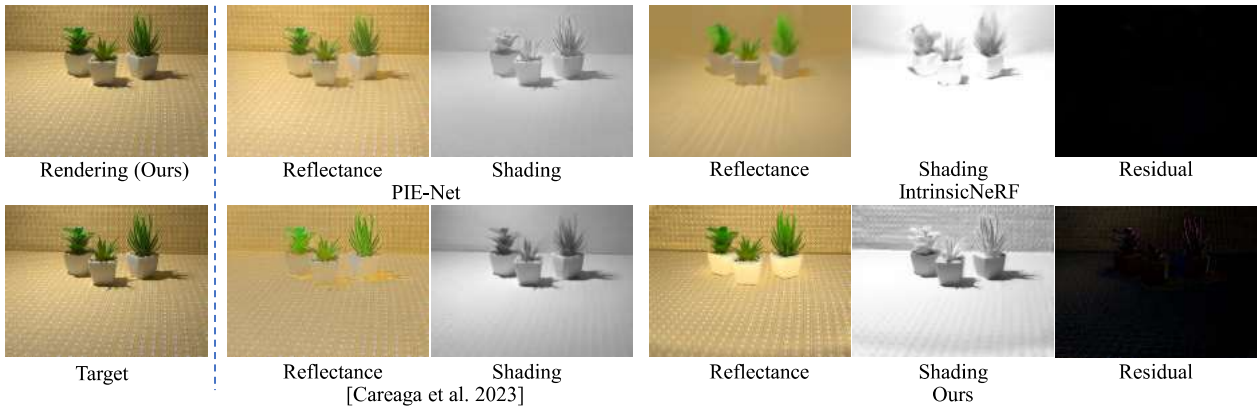


Figure 2. Compare with other methods on the ReNe dataset.

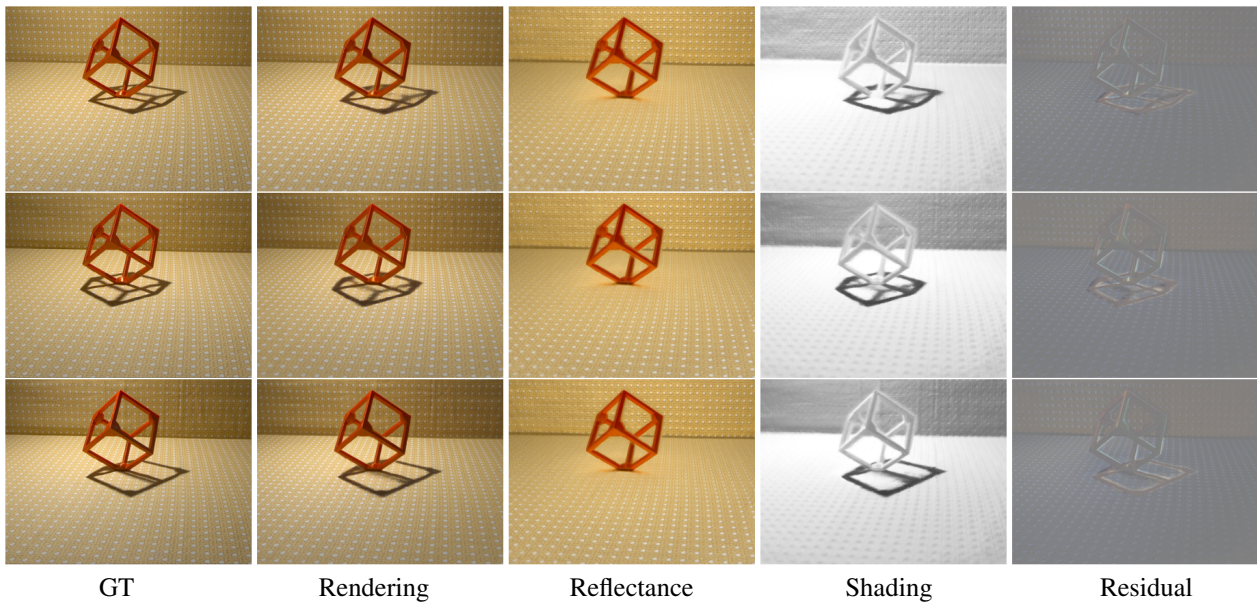


Figure 3. Qualitative results on the ReNe dataset (Cube).

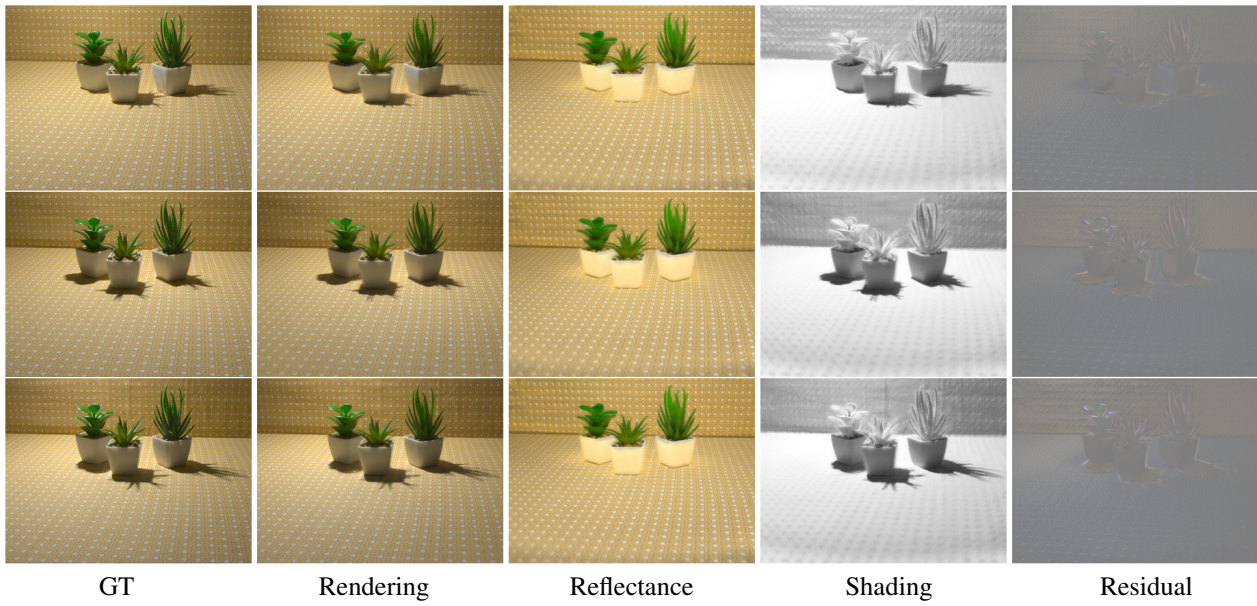


Figure 4. Qualitative results on the ReNe dataset (Garden).

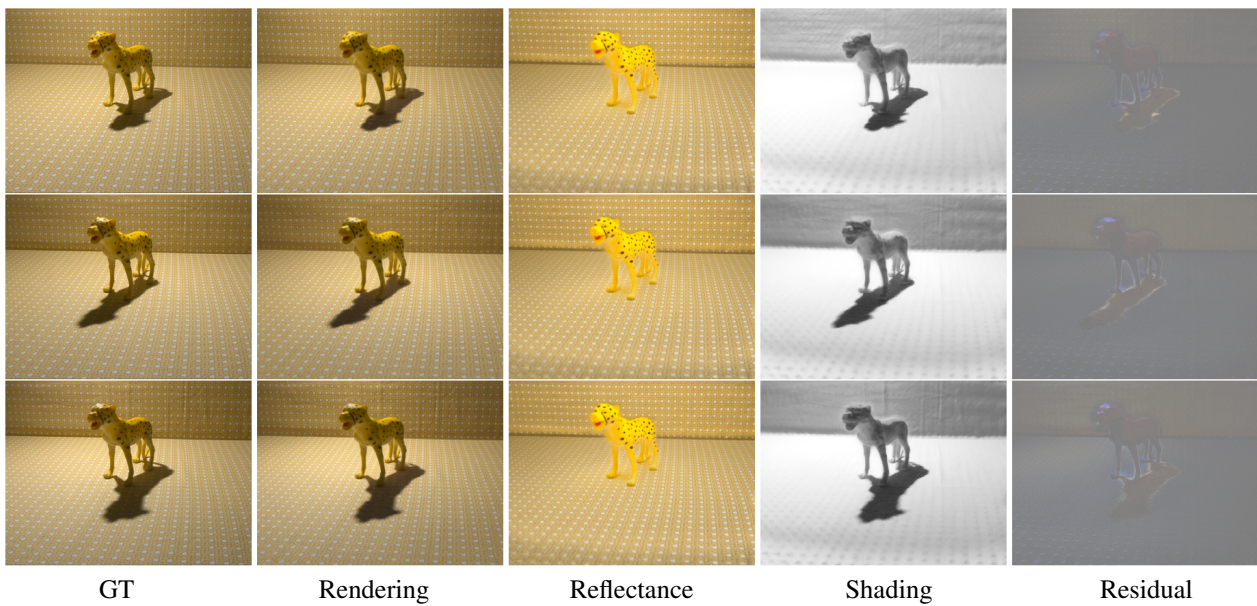


Figure 5. Qualitative results on the ReNe dataset (Cheetah).

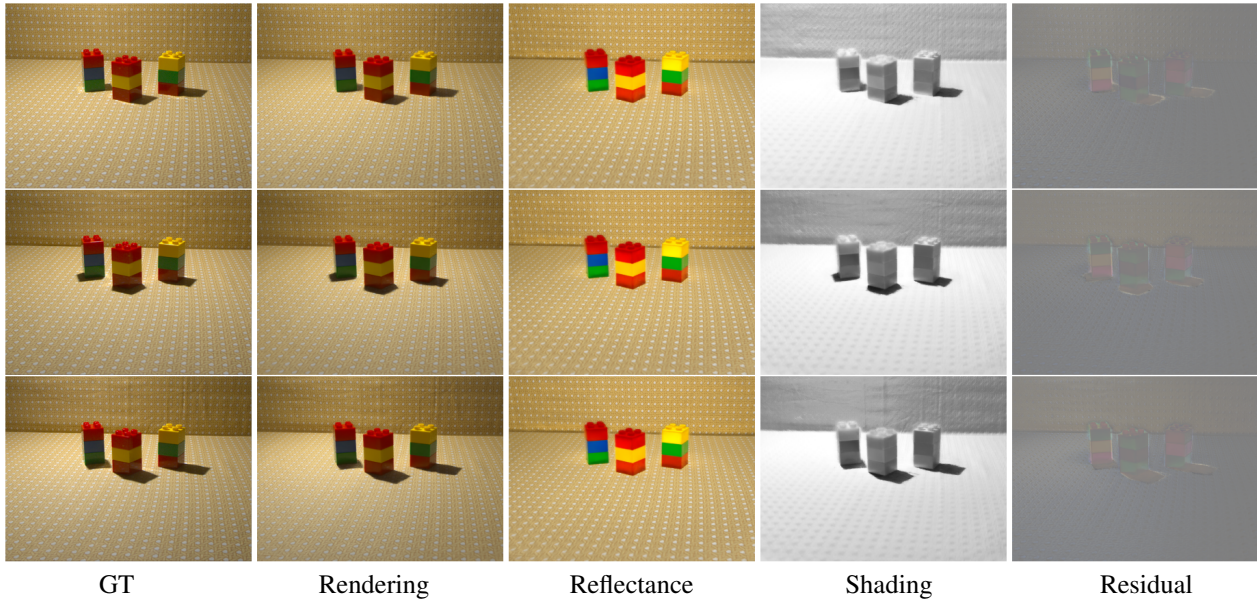


Figure 6. Qualitative results on the ReNe dataset (Lego).

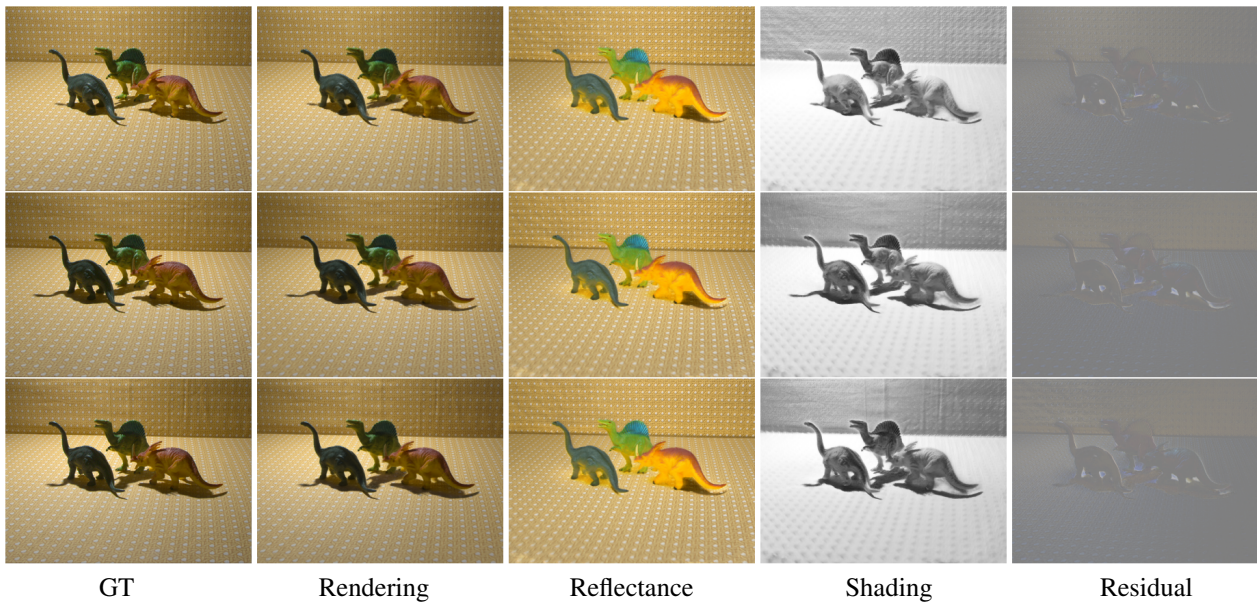


Figure 7. Qualitative results on the ReNe dataset (Dinosaurs).

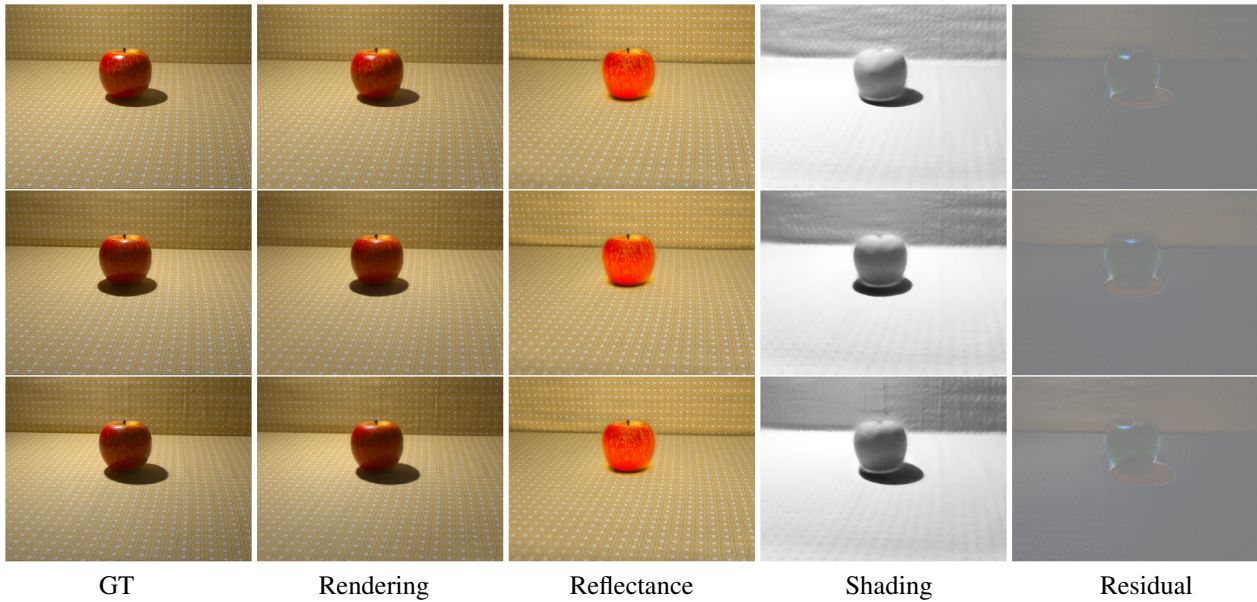


Figure 8. Qualitative results on the ReNe dataset (Apple).

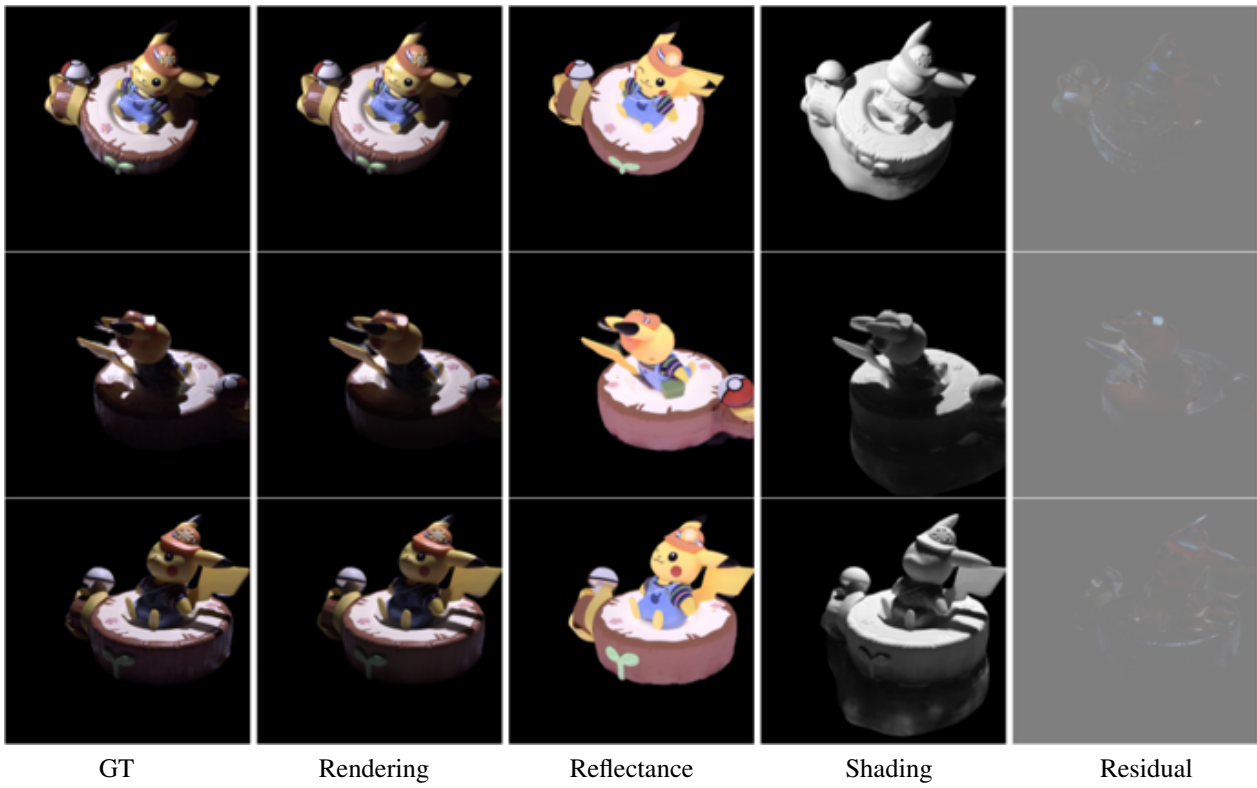


Figure 9. Qualitative results on the real scene (Pikachu).

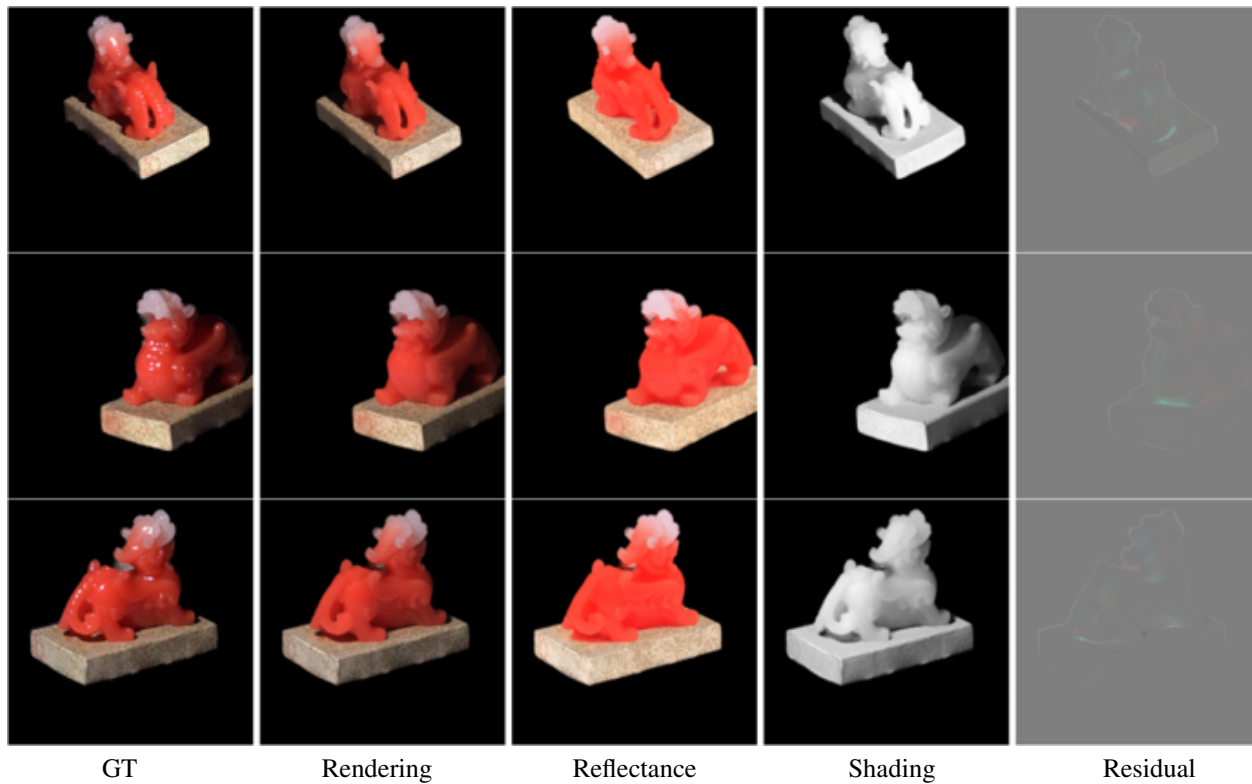


Figure 10. Qualitative results on the real scene (Pixiu).

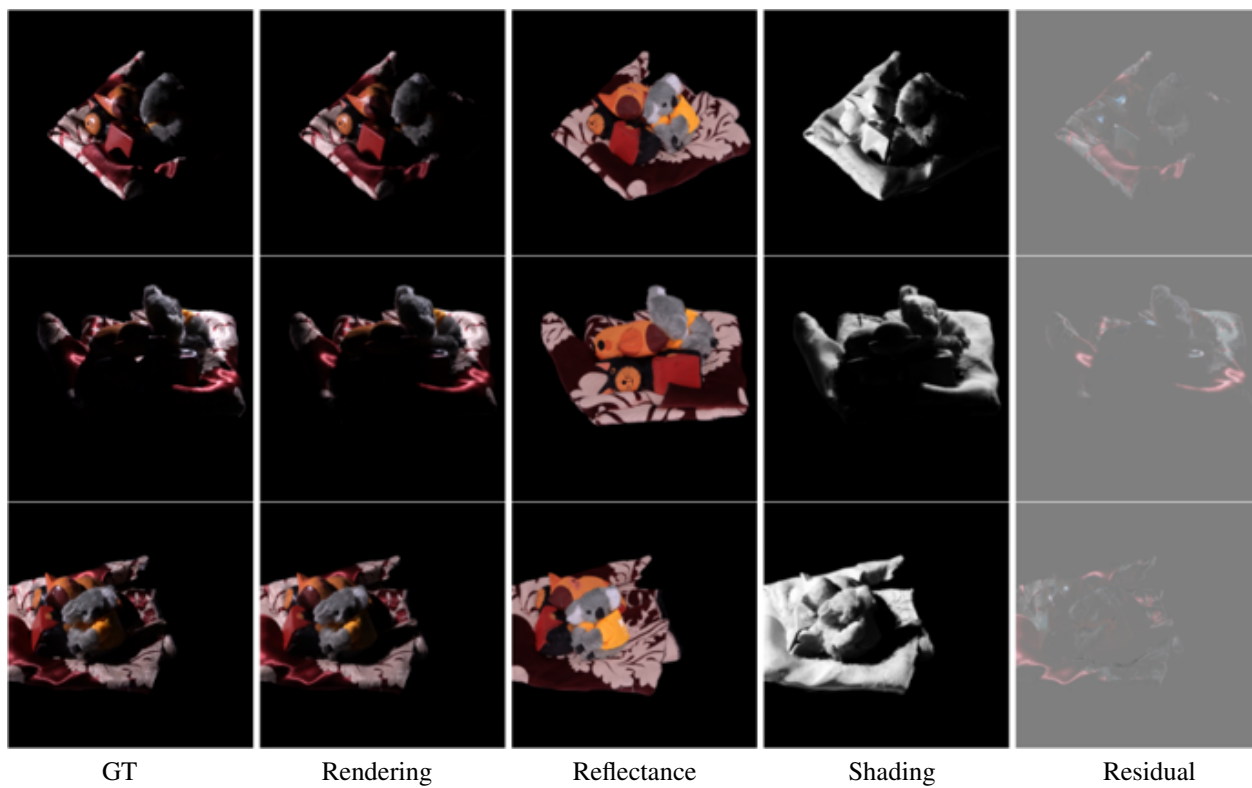


Figure 11. Qualitative results on the real scene (FurScene).

References

- [1] Ziyu Chen, Chenjing Ding, Jianfei Guo, Dongliang Wang, Yikang Li, Xuan Xiao, Wei Wu, and Li Song. L-tracing: Fast light visibility estimation on neural surfaces by sphere tracing. In Proceedings of the European Conference on Computer Vision (ECCV), 2022. 1
- [2] Haiyan Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensor: Tensorial inverse rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023. 1
- [3] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In International Conference on Learning Representations, 2018. 2
- [4] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1):99–106, 2021. 2
- [5] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 20762–20772, 2023. 2
- [6] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4):600–612, 2004. 2
- [7] Ziyi Yang, Yanzhen Chen, Xinyu Gao, Yazhen Yuan, Yu Wu, Xiaowei Zhou, and Xiaogang Jin. Sire-ir: Inverse rendering for brdf reconstruction with shadow and illumination removal in high-illumination scenes. arXiv preprint arXiv:2310.13030, 2023. 1
- [8] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In ACM SIGGRAPH 2023 Conference Proceedings, 2023. 2
- [9] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 586–595, 2018. 2
- [10] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In CVPR, 2022. 1