

A Reflective Object Reconstruction Method Based on Neural Radiance Field

Lifen Li^{1, 2, 3}, Zhengqin Yu^{1, 2, 3} and Ronghua Zhang^{1, 2, 3, *}

¹ Department of Computer, North China Electric Power University, Baoding 071066, China

² Hebei Key Laboratory of Knowledge Computing for Energy & Power, Baoding 071066, China

³ Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education, Baoding 071066, China

*Corresponding author Email: zronghua88@gmail.com

Abstract: We propose a neural rendering-based 3D reconstruction method for reconstructing the geometry and BRDF of reflective objects from multi-view images captured in unknown environments. Multi-view reconstruction of reflective objects is extremely challenging because specular reflections are view-dependent and therefore violate multi-view consistency, which is the cornerstone of most multi-view reconstruction methods. Recent neural rendering techniques can model the interaction between ambient light and object surfaces to adapt to view-dependent reflections, making it possible to reconstruct reflective objects from multi-view images. However, accurately modeling ambient light in neural rendering is difficult. We propose a two-step approach to solve this problem. First, by introducing view-dependent photometric losses, our method accurately reconstructs the geometry of reflective objects. Then, with the object geometry fixed, we use more accurate sampling to recover the BRDF of the ambient light and object. Experiments show that our method is able to accurately reconstruct the geometry and BRDF of reflective objects from only RGB images without knowing the ambient light and object masks.

Keywords: Neural Radiance Field, Neural Implicit Surface, Multi-view Reconstruction, Computer Vision.

1. Introduction

In the sphere of computer graphics and vision, multi-view 3D reconstruction serves as an integral undertaking. Since its initial proposition in 2003 by Hartley and Zisserman^[1], the art has seen considerable growth. However, there persists a critical and pressing issue, namely the reconstruction of reflective objects commonly found in real-world scenarios. These objects, bearing glossy exteriors, display varying coloration from different viewing angles due to reflected light. This property poses a significant challenge to a majority of multi-view reconstruction strategies that depend upon view consistency for stereo matching. Consequently, this has become a crucial hindrance to the quality of reconstruction inherent in today's technological solutions.

Recently, a rising trend in the domain of multi-view reconstruction—neural rendering-based techniques—has indicated immense feasibility in managing complex objects. Within these neural rendering-based reconstruction methodologies, the foundational surface geometry is commonly depicted as an implicit function, such as a signed distance function (SDF) encoded via a multilayer perceptron (MLP). By emulating the view-aligned color information and lessening the discrepancy between the rendered image and the input image, these tactics fine-tune the neural implicit function, subsequently accomplishing the geometry's reconstruction.

Nonetheless, despite the efficacy neural rendering-based methods demonstrate when tackling generic reconstruction tasks, they grapple with difficulties when confronted with reflective objects. This challenge originates from the fact that color functions in these techniques primarily target the relationship between color, viewing angle, and surface geometry, while overlooking the inherent optical dynamics associated with reflection. Thus, when attempting to fit specular color alterations on the surface across varying

directions of view, these methods might unintentionally generate flawed geometry. This predicament persists even with the increase in position encoding frequency or with employing deeper and wider MLP networks.

Consequently, there is a demand for specialized processing and schemas for reconstructing surfaces of reflective objects, ensuring that the influences of reflective phenomena on color changes are accurately captured to satisfy the requirements of practical applications. Prospective avenues for research could revolve around fabricating more sophisticated color functions for explicitly modeling reflection processes, and investigating alternate strategies for augmenting the performance of neural reconstruction techniques when applied to reflective objects.

2. Related Work

2.1. Traditional Multi-view Reconstruction.

Traditional multi-view reconstruction methods mainly rely on discrete representation techniques, such as depth maps, point clouds and voxel grids, which are widely used in the field of 3D reconstruction. Among them, depth map has become the most widely used representation method because of its high flexibility and suitability for parallel computing. Depth-based methods usually start by analyzing the correlation between multiple input images and build a three-dimensional model by calculating depth maps. After obtaining the depth map, further fusion is performed to generate a point cloud, and grid algorithms such as Poisson surface reconstruction^[2], spherical projection^[3] or Delaunay triangulation^[4] are used to convert the point cloud into a grid form.

Taking the classic depth-based method COLMAP^[5] as an example, it first extracts features in the image and matches them between different views to estimate depth information. Subsequently, a dense point cloud is generated by fusing these predicted depth maps. Finally, an additional meshing step is

applied to obtain the surface representation from the point cloud. However, such methods with complex processing pipelines often have some problems. Since errors may be introduced at each processing stage, these errors gradually accumulate and may eventually lead to incomplete 3D models. Especially for non-Lambertian surfaces (that is, objects whose surface reflection characteristics do not satisfy Lambertian reflection law), traditional methods are often difficult to effectively handle the color changes related to the view, further affecting the accuracy and detail performance of the reconstruction.

Therefore, current research is actively exploring new methods to overcome these limitations. Among them, neural radiation fields, as an emerging technology, are receiving widespread attention. This method aims to directly learn the mapping from images to 3D space through deep learning models, thereby achieving more accurate and detailed 3D model reconstruction.

2.2. Reflective object reconstruction.

The presence of reflective surfaces is extremely common in real-life scenes and has a profound impact on the immersive quality of virtual experiences. Such reflective objects usually have smooth surfaces that can reflect incident light of different intensities. Therefore, the reflection effect will cause the color of the object to appear inconsistent at different viewing angles. In addition, reflection phenomena may also cause significant interference in multi-view consistency, posing a challenge to existing 3D reconstruction methods. Although neural representation methods such as NeRF[6] and NeuS[7] have achieved unprecedented reconstruction quality, they still face difficulties in accurately capturing and reconstructing the appearance of reflective surfaces, especially complex specular reflections such as those on glass or mirrors.

This limitation mainly stems from the complexity of modeling the scene behind reflections, such as the handling of surface translucency (such as fog effects). Furthermore, these methods usually consider all spatial points along the camera ray path when interpreting the color of points on reflective surfaces, but ignore the changes in the color of reflective surfaces when viewed from different directions.

To address the challenge of reflective surface reconstruction, Verbin et al.[8] proposed Ref-NeRF, which replaces the original radiation dependence by enhancing NeRF and introducing a representation of the reflection direction, thereby generating a more accurate radiation field.

At the same time, other studies have proposed volume density regularization methods, aiming to improve the reconstruction quality of NeRF normal vectors to render specular reflections more realistically. Similarly, NeFRen[9] extends NeRF by segmenting the scene into transmission and reflection components, modeling these components with separate neural radiation fields, and synthesizing new perspectives through weighted combinations. Both qualitative and quantitative results demonstrate the superiority of this method in PSNR and its ability to accurately reproduce specular reflections.

Additionally, there are studies exploring view-dependent reflective appearance rendering by decomposing the scene into multiple physical components such as shape, reflectivity, and lighting. However, Ge et al.[10] pointed out that this decomposition may lead to inaccuracies, which in turn affects the reconstruction performance. Therefore, they introduced Ref-NeuS based on NeuS and proposed a strategy to reduce the importance of reflective surfaces. Ref-NeuS leverages multi-view context to estimate explicit reflection scores, helps identify reflective surface locations, and designs photometric losses to dynamically reduce blurring. Compared with Ref-NeRF, Ref-NeuS achieves higher PSNR, significantly improving reconstruction accuracy and accuracy of reflective surface normals.

In recent research, a novel neural reconstruction method named NeRO[11] is proposed, which can accurately reconstruct the geometry of reflective objects without knowing the ambient lighting conditions and object masks. The core idea of this method involves incorporating the explicit integration of the rendering equation into a neural reconstruction framework, significantly enhancing the representational power of the neural implicit function, enabling it to capture high-frequency specular color changes.

3. Method

The method for reflective surfaces proposed in this chapter can be divided into two stages. The first stage is to reconstruct the geometric surface of the reconstructed object. The purpose is to reconstruct a geometric surface as accurate as possible. The work in this stage is mainly based on the third stage. It is carried out based on the improvement of the work in Chapter 3; the second stage is to perform special processing on the reflective surface, repair the reconstruction results of the first stage, and reconstruct the material information of the reflective surface. The overall process of the overall reconstruction method is shown in Figure 3-1 below:

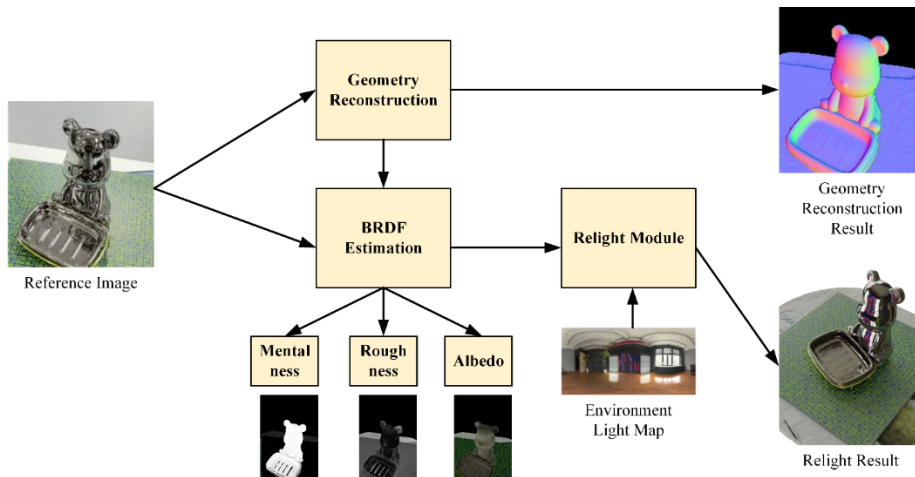


Figure 3-1. The framework of our method

3.1. Geometry Reconstruction

In the first stage of reconstruction, a geometric surface reconstruction is required. However, due to the various problems mentioned above on reflective surfaces, it is impossible to accurately reconstruct the geometric surface information of reflective objects. Here, the design in Ref-NeuS is adopted, using a color loss L_{color} that adaptively reduces the weight assigned to reflective pixels to overcome reconstruction failures caused by color inconsistencies observed on glossy surfaces. Next, we will explain how to derive L_{color} .

First, the reflection score needs to be defined in order to identify the pixels corresponding to the reflective surface. The same method used by NeRF-W^[11] is to make the rendering color of the light obey the Gaussian distribution:

$$\hat{C}(r) \sim (\bar{C}(r), \bar{\beta}^2(r)) \quad (3-1)$$

Among them, $\bar{C}(r)$ and $\bar{\beta}^2(r)$ are the mean and variance respectively. The variance here is determined based on the color in multiple images corresponding to the same point. For the reflective surface, the same point The colors observed at different viewing angles are likely to be different, so the variance will be larger than the points on the normal surface. In order to obtain the multi-view pixel color $\{C_i(r)\}_{i=1}^N$, project the surface point x to all multi-view images $\{I_i\}_{i=1}^N$. And use bilinear interpolation to obtain the corresponding pixel color:

$$M = K \cdot (R \cdot x + T) \quad (3-2)$$

$$C = f(I, M) \quad (3-3)$$

Among them, $f()$ represents the bilinear interpolation function, K represents the internal calibration matrix, R represents the rotation matrix, T represents the translation matrix, \cdot is matrix multiplication, and C is the target pixel color.

Considering that only some local areas of the image contain reflections, the localization of reflections is regarded as an anomaly detection problem. It is expected to identify the reflective surface as an abnormal area and assign a high reflection score. Therefore, an anomaly detector with Mahalanobis distance as the reflection score is used to estimate $\bar{\beta}^2(r)$:

$$\bar{\beta}^2(r) = \gamma \frac{1}{N-1} \sum_{j=1}^{N-1} \sqrt{(C_i(r) - C_j(r))^T \Sigma^{-1} (C_i(r) - C_j(r))} \quad (3-4)$$

where γ is the proportional numerator that controls the scale of the reflection fraction, and Σ^{-1} is the empirical covariance matrix.

Then $\bar{\beta}^2(r)$ is expanded to the original color loss L_{color} , so that it can perceive the reflection area, using a method similar to that in NeRF-W:

$$L_{color} = \sum_{r \in R} \frac{\|C(r) - \bar{C}(r)\|_2^2}{2\bar{\beta}^2(r)} + \frac{\log \bar{\beta}^2(r)}{2} \quad (3-5)$$

R is the number of rays in the batch. Since $\bar{\beta}^2(r)$ is estimated by Equation 4-4, for any certain point, $\bar{\beta}^2(r)$ is a constant and can be deleted from the loss function, and L1 loss can be used instead of L2 loss, so the loss function can be simplified to

$$L_{color} = \sum_{r \in R} \frac{|C(r) - \bar{C}(r)|}{\bar{\beta}^2(r)} \quad (3-6)$$

3.2. BRDF estimation

When calculating lighting effects, the material information of the surface is an important calculation parameter. After the first stage of geometric reconstruction, a relatively accurate geometric shape of the reflective object can be obtained. In the second stage, it is necessary to reconstruct the material information of the surface of the reflective object based on the accurate geometric shape obtained in the first stage, that is, metalness, albedo and roughness. At this stage, the geometric shape reconstructed in the first stage will be directly used to predict the material information of the reconstructed object surface.

In this module, U-Net is used as the main network of the material estimation network, and a differentiable renderer is used to render the material information output by the material estimation network and the first-stage geometric information into corresponding images. A differentiable renderer is a special type of renderer that uses a differentiable rendering function. It can not only input various scene parameter information (including material information and geometric information, etc.) to obtain a two-dimensional image, but also The derivatives of this function for different scene parameters can be calculated.

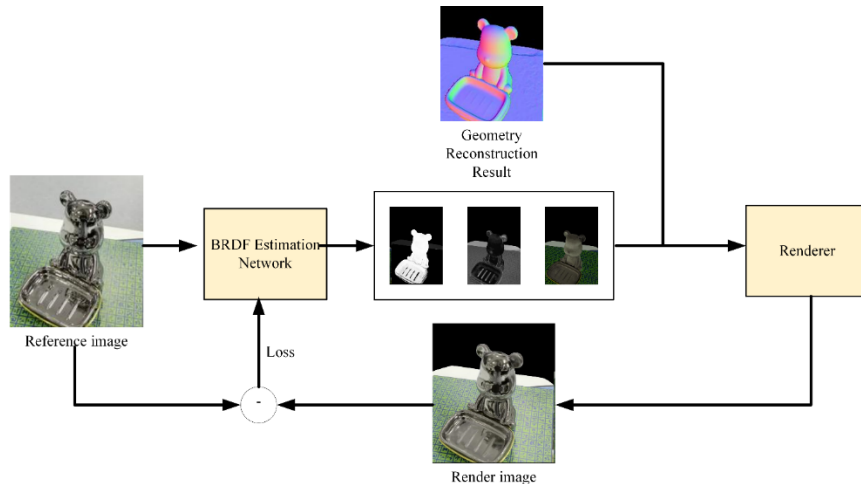


Figure 3-2. The framework of BRDF Estimation

After obtaining the corresponding rendered image, the difference between the rendered image and the original image is calculated, and the loss is passed to the material estimation network in the form of back propagation to optimize network parameters.

4. Experiments

4.1. Dataset

The dataset used is the Glossy-Real dataset used in NeRO. This is a real dataset that contains five objects (coral, lucky cat, bear, rabbit and vase). The surface of each reconstructed object has Smooth surface with strong specular reflection

effect. The multi-view image set of each reconstructed object consists of 100-130 images with a resolution of 1024×768, and COLMAP is used to reconstruct the camera pose of the captured image. In order to more accurately locate the corresponding camera posture of the captured image, all reconstruction objects are placed on a calibration plate with strong texture, and all images are captured with mobile phones. The real surface data of the reconstructed object is captured by the RGBD sensor EinScan Pro 2X. In this article, corals, rabbits and bears are selected as reconstruction objects. The schematic diagram of the selected data sets is shown in Figure 4-1.

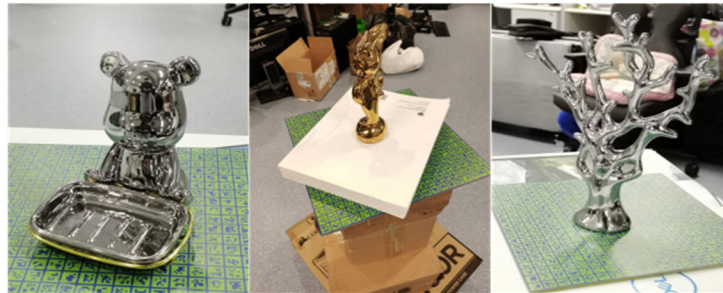


Figure 4-1. Glossy-Real Dataset

4.2. Geometry reconstruction result

As shown in Figure 4-2, the experimental results of the first stage of geometric surface reconstruction are shown. In the first column, the input reference images are shown, which provide the basic data for the subsequent reconstruction process. The second column shows the target image rendered by the model. This process aims to enable the model to effectively learn the three-dimensional shape of the reconstructed object contained in the multi-view image by

reducing the difference between the rendered image and the original image. information. In the third column, the estimated depth map of the reconstructed object is shown, which intuitively reflects the depth change of the object surface. Finally, in the fourth column, the reconstructed geometric model based on the above process is shown. By comparing and analyzing these results, it can be clearly seen that the improved method can more completely reconstruct the geometric information of the reconstructed object, thus verifying the effectiveness of the method.

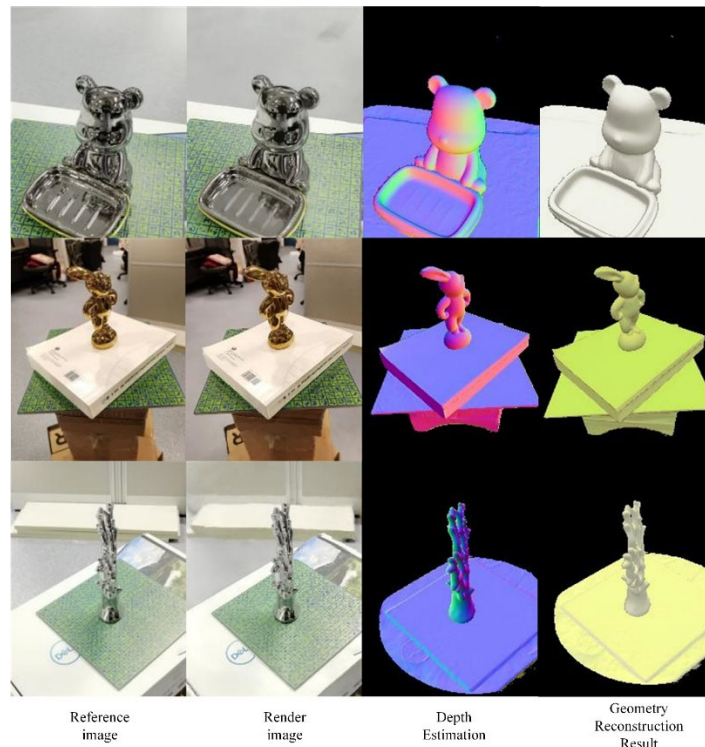


Figure 4-2. Geometry Reconstruction Result

4.3. BRDF estimation result

As shown in Figure 4-3, the detailed results of the second stage of material reconstruction are shown. The first and second columns in the figure present the reference image and the corresponding rendered image respectively. By comparing the two columns of images, the differences and similarities in the reconstruction process can be visually observed. During the reconstruction process, the obtained material parameters are refined into three key parameters:

reflectivity, metallicity, and roughness. From the third column to the fifth column, the reconstruction results of reflectivity, metallicity and roughness are shown in sequence. After analysis, it can be found that these prediction results are consistent with the actual material information in the reference image. In particular, the reconstruction effect of metallicity and roughness can more accurately reflect the material characteristics in the reference image. Therefore, it can be considered that this stage The material reconstruction results are reasonable and effective.

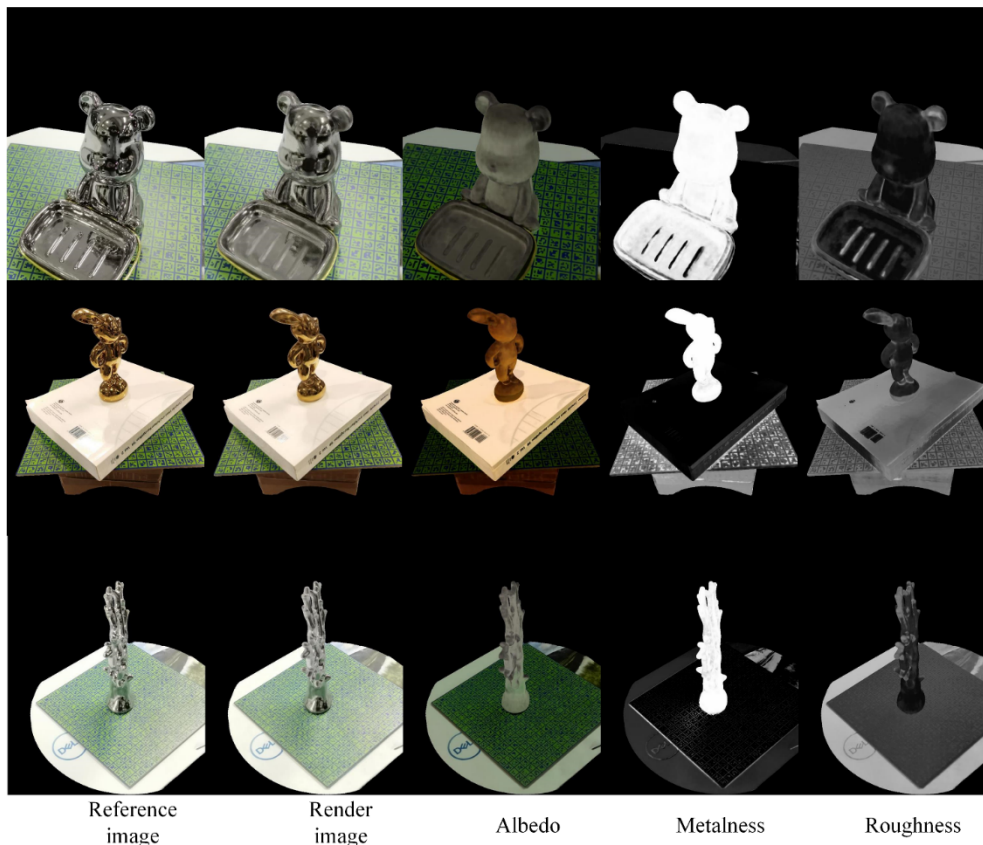


Figure 4-3. BRDF Estimation Result

5. Conclusion

For objects with reflective surfaces that are common in daily life, traditional multi-view 3D reconstruction methods are often difficult to achieve ideal results due to the complexity of their surface reflections. To this end, this paper proposes a reflective surface reconstruction method based on neural radiation fields. The method is divided into two stages and aims to accurately reconstruct objects with reflective surfaces. In the first stage, an occupancy loss is introduced to estimate the geometric shape to optimize the reconstruction accuracy of the geometric model. Through this improvement, a more accurate geometric model can be obtained. In the second stage, based on the reconstructed accurate geometric model, the material parameter estimation model is used to restore the material parameters of the reconstructed object, thereby achieving a complete reconstruction of the entire reconstructed object. Experimental results show that in the first stage, the improved method in this paper can reconstruct a more accurate geometric model than the original method. In the second stage, satisfactory surface material information of the reconstructed object was successfully reconstructed.

References

- [1] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge university press, 2003.
- [2] Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction[C]//Proceedings of the fourth Eurographics symposium on Geometry processing. 2006, 7(4).
- [3] Cardozo N, Allmendinger R W. Spherical projections with OSXStereonet[J]. Computers & Geosciences, 2013, 51: 193-205.
- [4] Chen L, Xu J. Optimal delaunay triangulations[J]. Journal of Computational Mathematics, 2004: 299-308.
- [5] Schonberger J L, Frahm J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4104-4113.
- [6] Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[J]. Communications of the ACM, 2021, 65(1): 99-106.

- [7] Wang P, Liu L, Liu Y, et al. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction[J]. arXiv preprint arXiv:2106.10689, 2021.
- [8] Verbin D, Hedman P, Mildenhall B, et al. Ref-nerf: Structured view-dependent appearance for neural radiance fields[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 5481-5490.
- [9] Guo Y C, Kang D, Bao L, et al. Nerfren: Neural radiance fields with reflections[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 18409-18418.
- [10] Ge W, Hu T, Zhao H, et al. Ref-neus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 4251-4260.
- [11] Liu Y, Wang P, Lin C, et al. Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images[J]. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-22.
- [12] Martin-Brualla R, Radwan N, Sajjadi M S M, et al. Nerf in the wild: Neural radiance fields for unconstrained photo collections[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7210-7219.