Mirror-NeRF: Learning Neural Radiance Fields for Mirrors with Whitted-Style Ray Tracing

Junyi Zeng[*]
Zhejiang University
Hangzhou, China
zengjunyi@zju.edu.cn

Chong Bao[*]
Zhejiang University
Hangzhou, China
chongbao@zju.edu.cn

Rui Chen
Zhejiang University
Hangzhou, China
22221111@zju.edu.cn

Zilong Dong
Alibaba Group
Hangzhou, China
list.dzl@alibaba-inc.com

Guofeng Zhang
Zhejiang University
Hangzhou, China
zhangguofeng@zju.edu.cn

Hujun Bao
Zhejiang University
Hangzhou, China
bao@cad.zju.edu.cn

Zhaopeng Cui[†]
Zhejiang University
Hangzhou, China
zhpcui@zju.edu.cn

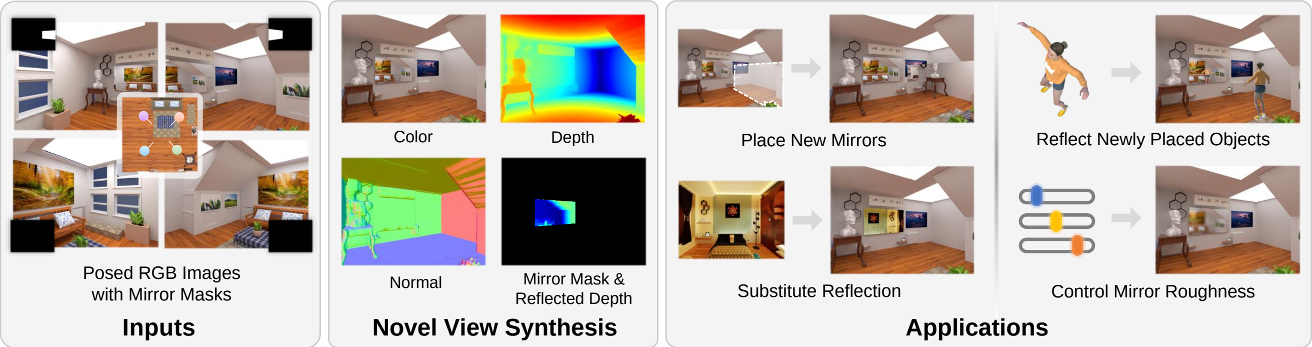


Figure 1: We present Mirror-NeRF, a novel neural rendering framework that incorporates Whitted Ray Tracing to achieve photo-realistic novel view synthesis in the scene with the mirror and supports various scene manipulation applications. Given the posed images with mirror reflection masks, we can learn the correct geometry and reflection of the mirror.

ABSTRACT

Recently, Neural Radiance Fields (NeRF) has exhibited significant success in novel view synthesis, surface reconstruction, *etc.* However, since no physical reflection is considered in its rendering pipeline, NeRF mistakes the reflection in the mirror as a separate virtual scene, leading to the inaccurate reconstruction of the mirror and multi-view inconsistent reflections in the mirror. In this paper, we present a novel neural rendering framework, named Mirror-NeRF, which is able to learn accurate geometry and reflection of the mirror and support various scene manipulation applications with mirrors, such as adding new objects or mirrors into the scene and synthesizing the reflections of these new objects in mirrors, controlling mirror roughness, *etc.* To achieve this goal, we propose a unified radiance field by introducing the reflection probability and tracing rays following the light transport model of Whitted Ray Tracing, and also develop several techniques to facilitate the learning process. Experiments and comparisons on both synthetic and real datasets demonstrate the superiority of our method. The code and supplementary material are available on the project webpage: https://zju3dv.github.io/Mirror-NeRF/.

CCS CONCEPTS

• **Computing methodologies → Computer vision**; **Rendering**.

KEYWORDS

neural rendering; ray tracing; scene reconstruction; scene editing

ACM Reference Format:
Junyi Zeng, Chong Bao, Rui Chen, Zilong Dong, Guofeng Zhang, Hujun Bao, and Zhaopeng Cui. 2023. Mirror-NeRF: Learning Neural Radiance Fields for Mirrors with Whitted-Style Ray Tracing. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29-November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3581783.3611857

1 INTRODUCTION

3D scene reconstruction and rendering is a long-standing problem in the fields of computer vision and graphics with broad applications in VR and AR. Although significant progress has been made over decades, it is still very challenging to reconstruct and re-render the scenes with mirrors, which exist ubiquitously in the real world.

[*] Junyi Zeng and Chong Bao contributed equally to this research. The authors from Zhejiang University are also affiliated with the State Key Lab of CAD&CG.
[†] Corresponding author: Zhaopeng Cui.

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada

ACM ISBN 979-8-4007-0108-5/23/10…$15.00
https://doi.org/10.1145/3581783.3611857

The "appearance" of the mirror is not multi-view consistent and changes considerably with the observer's perspective due to the physical reflection phenomenon where the light will be entirely reflected along the symmetric direction at the mirror.

Recently, Neural Radiance Fields (NeRF) [16] has exhibited significant success in novel view synthesis and surface reconstruction due to its capability of modeling view-dependent appearance changes. However, since the physical reflection is not considered in its rendering pipeline, NeRF mistakes the reflection in the mirror as a separate virtual scene, leading to the inaccurate reconstruction of the geometry of the mirror, as illustrated in Fig. 2. The rendered "appearance" of the mirror also suffers from multi-view inconsistency. Several techniques [22, 26, 50] decompose the object material and illuminations to model the reflection effect at the surface, but they all assume the surfaces with certain diffuse reflection to recover object surface first and then model the specular component. Thus they struggle to handle the mirrors with pure specular reflection due to the incorrect surface estimation of mirrors. NeRFReN [9] models reflection by separating the reflected and transmitted parts of a scene as two radiance fields and improves the rendering quality for the scenes with mirrors, while it still fails to model the physical specular reflection process. Thus, it cannot render the reflection that is not observed in the training views as shown in Fig.2, and cannot synthesize new reflections of the objects or mirrors that are newly placed in the scene.

In this paper, we propose a novel neural rendering framework, named Mirror-NeRF, to accomplish high-fidelity novel view synthesis in the scene with mirrors and support multiple scene manipulation applications. For clarity, we term the ray as the inverse of light. The rays emitted from the camera are termed as camera rays and rays reflected at the surface are termed as reflected rays. Exhaustively conducting ray tracing in a room-scale environment is prohibitively expensive. With the goal of achieving physically-accurate rendering of reflections in the mirror, we draw inspiration from Whitted Ray Tracing [37] where the ray is reflected at the mirror-like surface and terminates at a diffuse surface. Specifically speaking, we first define the probability that the ray is reflected when hitting a spatial point as the reflection probability. The reflection probability is parameterized as a continuous function in the spatial space by a Multi-Layer Perceptron (MLP). Then we trace the ray emitted from the camera. The physical reflection will take place when the ray hits the surface with a high reflection probability. We accumulate the density and radiance of the ray by the volume rendering technique and synthesize the image by blending the color of camera rays and reflected rays based on the reflection probability. Instead of taking the specular reflection as separate neural fields, our neural fields are unified, which is more reasonable to synthesize new physically sound reflection from novel viewpoints. As shown in Fig. 1, our representation further supports various types of scene manipulations, *e.g.*, adding new objects or mirrors into the scene and synthesizing the reflections of these new objects in mirrors, controlling the roughness of mirrors and reflection substitution.

However, learning both geometry- and reflection-accurate mirror with the proposed new representation is not trivial. First, the reflection at a surface point is related to the surface normal. The analytical surface normal from the gradient of volume density has significant noise since the density cannot concentrate precisely on the surface. Thus, we exploit an MLP to parameterize a smooth distribution of surface normal. Second, the reconstruction of mirror surface is ambiguous and challenging, since the "appearance" of mirror is from other objects and not consistent from different viewpoints. Based on the fact that mirrors in real world usually have planar surfaces, we leverage both plane consistency and forward-facing normal constraints in a joint optimization manner to guarantee the smoothness of the mirror geometry and reduce the ambiguity of the reflection. Moreover, a progressive training strategy is proposed to stabilize the geometry optimization of the mirror.

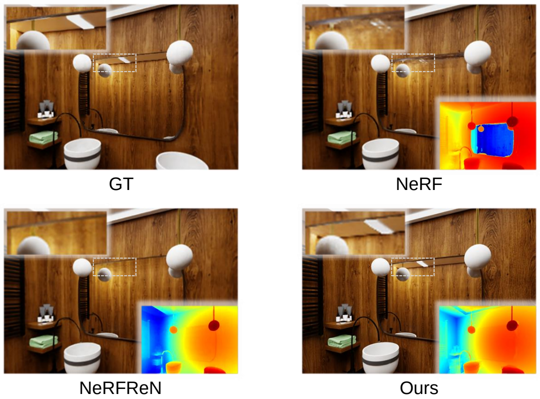


Figure 2: Comparison of the novel views synthesized by different methods. NeRF [16] mistakes the reflection in the mirror as a separate virtual scene, leading to inaccurate depth of the mirror. NeRFReN [9] uses two radiance fields to learn the color inside and outside the mirror separately. They synthesize the reflection in the mirror by interpolating the memorized reflection and cannot infer the reflection unobserved in the training views, *e.g.*, the missing ceiling. Instead, we successfully synthesize new reflections in the mirror with the accurate depth of the mirror due to our ray tracing pipeline.

Our contributions can be summarized as follows. **1)** We propose a novel neural rendering framework, named Mirror-NeRF, that resolves the challenge of novel view synthesis in the scene with mirrors. Different from NeRF [16] and NeRFReN [9] that tend to learn a separate virtual world in the mirror, Mirror-NeRF can correctly render the reflection in the mirror in a unified radiance field by introducing the reflection probability and tracing the rays following the light transport model of Whitted Ray Tracing [37]. The physically-inspired rendering pipeline facilitates high-fidelity novel view synthesis with accurate geometry and reflection of the mirror. **2)** To learn both accurate geometry and reflection of the mirror, we leverage several techniques, including a surface normal parametrization to acquire smooth distribution of surface normal, the plane consistency and forward-facing normal constraints with joint optimization to ensure the accurate geometry of the mirror, and a progressive training strategy to maintain the stability of training. **3)** The proposed Mirror-NeRF enables a series of new scene manipulation applications with mirrors as shown in Fig. 1, such as object placement, mirror roughness control, reflection substitution, *etc.* Extensive experiments on real and synthetic datasets demonstrate that Mirror-NeRF can achieve photo-realistic novel view synthesis. A large number of scene manipulation cases show the physical correctness and flexibility of the proposed method.

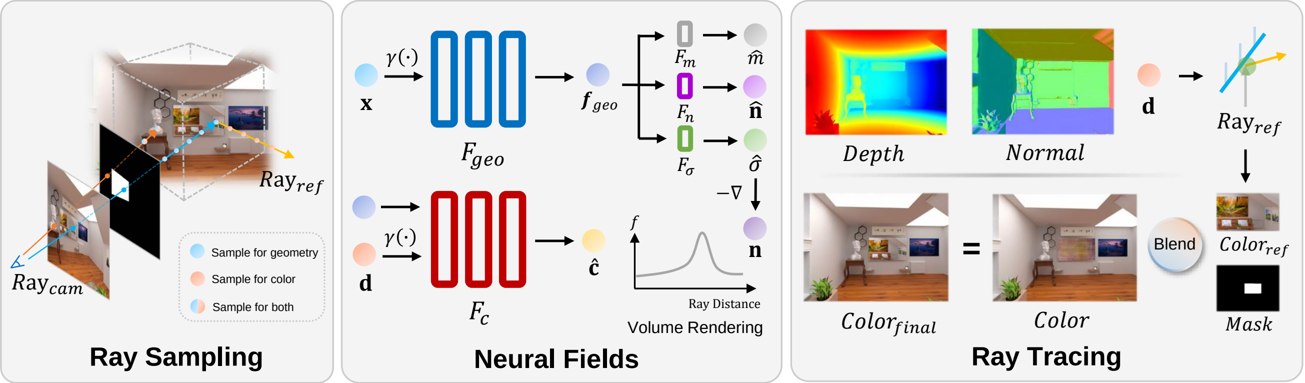


Figure 3: Framework. We trace the rays physically in the scene and learn a unified radiance field of the scene with the mirror. The neural field takes as input spatial location x, view direction d, and outputs the volume density $\hat{\sigma}$, radiance $\hat{\mathbf{c}}$, surface normal $\hat{\mathbf{n}}$ and reflection probability $\hat{m}$. The final color is blended by the color of the camera ray and the reflected ray based on the reflection probability.

2 RELATED WORK

2.1 Neural Rendering

The goal of neural rendering is to synthesize photorealistic images or videos by computing the light transport in a 3D scene. Lots of works [15, 21, 43] have been proposed to push the envelope of rendering quality in this field. One of the most notable approaches is NeRF [16], which models the radiance field of a scene using the MLP. By training on a set of posed images, NeRF learns to infer the radiance and density of each sampled point and accumulates them along the ray with volume rendering techniques to render the color. This enables NeRF to generate photorealistic images of the scene from a novel viewpoint. Several extensions and improvements have been proposed to apply NeRF to more challenging problems, such as scene reconstruction [1, 8, 13, 29, 30, 32, 36, 38, 39, 44, 48], generalization [24, 33], novel view extrapolation [35, 45], scene manipulation [2, 28, 40–42], SLAM [23, 54], segmentation [20, 53], human body [18, 31] and so on. Furthermore, some NeRF-variants provide various applications, such as supersampling [29] and controllable depth-of-field rendering [39]. However, these NeRF-variants struggle to model mirror reflection since they assume that all lights in the scene are reflected at Lambertain surfaces.

2.2 Neural Rendering With Reflection

Plenty of works [3, 5, 6, 10, 12, 17, 49, 51, 52] have been working on making NeRF understand physical reflection. PhySG [46] simplifies light transport by modeling the environment illumination and material properties as mixtures of spherical Gaussians and integrating the incoming light over the hemisphere of the surface. InvRender [50] extends PhySG to model the indirect light by using another mixture of spherical Gaussians to cache the light that bounces off from other surfaces. These approaches assume that surfaces are diffuse with a simple BRDF and environment lighting is far away from the scene. For a room with the mirror, they cannot handle the complex reflection and material diversity in the scene. As for NeRF, it will treat the reflection in mirrors as real geometry, which reconstructs the inaccurate depth of the mirror. RefNeRF [26] decomposes the light as diffuse and specular components and learns the reflection using a radiance field conditioned by the reflected view direction. NeRFReN [9] employs two radiance fields to learn the color inside and outside the mirror and depth constraints to recover the depth of the mirror. However, these methods generate mirror reflection from new viewpoints by interpolating the previously learned reflections, and are limited in accurately inferring reflections that were not observed during training and synthesizing reflections for newly added objects or mirrors in the scene. By introducing the physical ray tracing into the neural rendering pipeline, our method can correctly render the reflection in the mirror and support multiple scene manipulation applications.

3 MIRROR-NERF

We introduce Mirror-NeRF, a physically inspired neural rendering framework that supports photo-realistic novel view synthesis of scenes with mirrors and reconstructs the accurate geometry and reflection of mirrors. As illustrated in Fig. 3, we leverage unified neural fields to learn the volume density, normal, reflection probability and radiance inside and outside the mirror (Sec. 3.1). With the intention of generating physically-accurate reflections in the mirror, we employ the light transport model in Whitted Ray Tracing [37] and trace the volume rendered ray in the scene (Sec. 3.2). Besides, some regularization constraints for the mirror surface (Sec. 3.3) and a progressive training strategy (Sec. 3.4) are proposed to improve the reconstruction quality of the mirror and stabilize the training.

3.1 Unified Neural Fields

We design several neural fields to learn the properties of the scene, which are unified for parts inside and outside the mirror (Fig. 3).

3.1.1 Geometry and Color. Following the implicit representation in NeRF [16], we use a geometry MLP $\mathcal{F}_{geo}$ to encode the geometry feature f_{geo} at an arbitrary spatial location $\mathbf{x}$. The volume density field is presented by a volume density MLP $\mathcal{F}_\sigma$ which takes f_{geo} as input, and the radiance field is presented by a radiance MLP $\mathcal{F}_c$ which takes f_{geo} and view direction $\mathbf{d}$ as input:

$$\begin{aligned} f_{geo} &= \mathcal{F}_{geo}(\gamma_x(\mathbf{x})), \\ \sigma &= \mathcal{F}_\sigma(f_{geo}), \\ \mathbf{c} &= \mathcal{F}_c(f_{geo}, \gamma_d(\mathbf{d})), \end{aligned} \tag{1}$$

where $\gamma_x(\cdot)$ and $\gamma_d(\cdot)$ are respectively the positional encoding function of spatial position and view direction. σ and $\mathbf{c}$ are volume density and radiance respectively. To render an image from a specific viewpoint, we follow the volume rendering techniques in NeRF. The volume-rendered color $\hat{C}$ of a ray $\boldsymbol{r}$ is calculated by accumulating the volume densities σ_i and radiances $\mathbf{c}_i$ of sampled points x_i along the ray:

$$\begin{aligned}\hat{C}(\boldsymbol{r}) &= \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i,\\ T_i &= \exp\left(-\sum_{j=1}^{i-1}\sigma_j\delta_j\right),\\ \alpha_i &= 1-\exp\left(-\sigma_i\delta_i\right),\end{aligned} \tag{2}$$

where N is the number of sampled points on the ray $\boldsymbol{r}$, and δ_i is the sampling distance between adjacent points along the ray.

3.1.2 Smooth Surface Normal. Prior works [4, 22] have analyzed the acquisition of surface normal in NeRF that the negative gradient of volume density *w.r.t.* $\mathbf{x}$ can give a differentiable approximation of the true normal:

$$\mathbf{n} = -\frac{\nabla\sigma(\mathbf{x})}{||\nabla\sigma(\mathbf{x})||}. \tag{3}$$

However, such parametrization tends to produce an unsmooth surface normal distribution since the volume density cannot concentrate precisely on the surface. The noise in the surface normal will severely hamper tracing the correct direction of the reflected rays at the mirror. To obtain a smooth distribution of surface normal, we utilize an MLP $\mathcal{F}_n$ that takes f_{geo} as input and predicts the smoothed surface normal $\hat{\mathbf{n}}$:

$$\hat{\mathbf{n}} = \mathcal{F}_n(f_{geo}). \tag{4}$$

We supervise the optimization of $\mathcal{F}_n$ by the analytical surface normal $\mathbf{n}$:

$$\mathcal{L}_n = ||\hat{\mathbf{n}} - \mathbf{n}||_2^2. \tag{5}$$

To compute the surface normal at the intersection point of a ray $\boldsymbol{r}$ and the surface, we follow the Eq. (2) by:

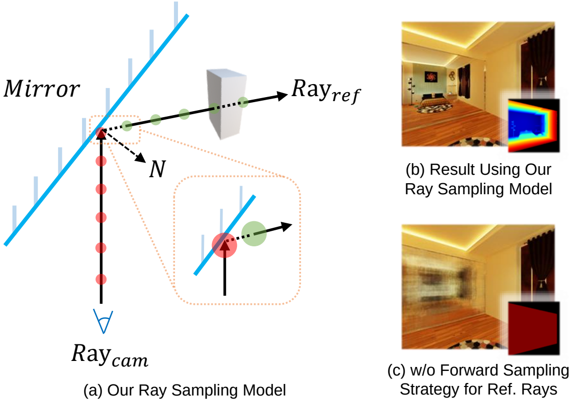
$$\hat{\mathbf{N}}(\boldsymbol{r}) = \sum_{i=1}^{N} T_i\alpha_i\hat{\mathbf{n}}_i. \tag{6}$$

3.1.3 Reflection Probability. To model the reflection and perform the Whitted-style ray tracing described in Sec. 3.2, we also utilize an MLP $\mathcal{F}_m$ to predict the probability m that rays will be reflected at a spatial point:

$$m = \mathcal{F}_m(f_{geo}), \tag{7}$$

where m ranges in $[0, 1]$. To determine the reflection probability $\hat{M}$ of a ray $\boldsymbol{r}$ hitting the solid surface, we perform the volume rendering like Eq. (2):

$$\hat{M}(\boldsymbol{r}) = \sum_{i=1}^{N} T_i\alpha_i m_i. \tag{8}$$



(a) Our Ray Sampling Model

(b) Result Using Our Ray Sampling Model

(c) w/o Forward Sampling Strategy for Ref. Rays

Figure 4: Our strategy for sampling points on rays is shown in (a). We sample points on both the camera ray and the reflected ray. For the reflected ray, we forward a distance from the origin to start sampling points to avoid the reflected ray terminating unexpectedly near the origin due to the "foggy" geometry. The effectiveness of this design is demonstrated by the comparison of (b) and (c) where mirror reflection is corrupted without the forward sampling strategy. The bottom right images in (b) and (c) show the reflected depth of the mirror.

3.2 Whitted-Style Ray Tracing

NeRF [16] does not take into account the physical reflection in the rendering pipeline. When applied to the scene with the mirror, NeRF cannot reconstruct the geometry of the mirror and treats the reflection in the mirror as a separate virtual scene. To handle the reflection at the mirror, we draw inspiration from Whitted Ray Tracing [37] where the ray is reflected at the mirror-like surface and terminates at the diffuse surface. As shown in Fig. 4, when a ray is reflected, we first compute the location $\hat{\mathbf{X}}$ of the intersection point of the ray $\boldsymbol{r}$ and the surface by:

$$\begin{aligned}\hat{\mathbf{X}}(\boldsymbol{r}) &= \mathbf{o}(\boldsymbol{r}) + \hat{D}(\boldsymbol{r})\mathbf{d}(\boldsymbol{r}),\\ \hat{D}(\boldsymbol{r}) &= \sum_{i=1}^{N} T_i\alpha_i t_i,\end{aligned} \tag{9}$$

where $\hat{D}$, $\mathbf{o}$ and $\mathbf{d}$ are the expected termination depth, origin and direction of the ray $\boldsymbol{r}$ respectively. T_i and α_i are the same as Eq. (2).

To trace the reflected ray $\boldsymbol{r}_{ref}$ of a ray $\boldsymbol{r}$, we set $\hat{\mathbf{X}}(\boldsymbol{r})$ as its origin, and compute its direction by:

$$\mathbf{d}(\boldsymbol{r}_{ref}) = \mathbf{d}(\boldsymbol{r}) - 2\left(\hat{\mathbf{N}}(\boldsymbol{r})\cdot\mathbf{d}(\boldsymbol{r})\right)\hat{\mathbf{N}}(\boldsymbol{r}). \tag{10}$$

Here all direction vectors are normalized.

Then, we use the volume rendering technique to compute the color of the ray $\boldsymbol{r}$ and its reflected ray $\boldsymbol{r}_{ref}$. The radiances of the sampled points on $\boldsymbol{r}$ and $\boldsymbol{r}_{ref}$ are attained by querying the same neural radiance field. Since the density-based representation always induces a "foggy" geometry, the reflected ray may terminate unexpectedly near the origin as illustrated in Fig. 4(c). To solve the problem, we start sampling points on the reflected ray at a distance from the origin as shown in Fig. 4(a).

We blend the color of the ray $\boldsymbol{r}$ and its reflected ray $\boldsymbol{r}_{ref}$ according to the volume-rendered reflection probability of the ray $\hat{M}(\boldsymbol{r})$ as:

$$\hat{C}^P(\boldsymbol{r}) = \hat{C}(\boldsymbol{r})\left(1-\hat{M}(\boldsymbol{r})\right) + \hat{C}^P(\boldsymbol{r}_{ref})\hat{M}(\boldsymbol{r}). \tag{11}$$

Note that $\hat{C}^P$ is defined recursively, and the recursion terminates when $\hat{M}$ is zero or the specified maximum recursion depth is reached.

For each pixel, we generate a ray from the camera and trace it in the scene. The set of these camera rays is denoted as R_{cam}. The pixel color is rendered by Eq.(11) with $\boldsymbol{r} \in R_{cam}$. We supervise the rendered pixel color by the ground truth pixel color C^I with a photometric loss:

$$\mathcal{L}_c = \sum_{\boldsymbol{r} \in R_{cam}} ||\hat{C^P}(\boldsymbol{r}) - C^I(\boldsymbol{r})||_2^2. \tag{12}$$

To guide the optimization of reflection probability $\hat{M}$, we calculate the binary cross entropy loss between the rendered reflection probability $\hat{M}$ and the mirror reflection mask M:

$$\mathcal{L}_m = \sum_{r \in R_{cam}} -\left(M(\boldsymbol{r}) \log \hat{M}(\boldsymbol{r}) + (1 - M(\boldsymbol{r})) \log \left(1 - \hat{M}(\boldsymbol{r})\right)\right), \tag{13}$$

where M is obtained by using the off-the-shelf segmentation tools like [11] on the ground-truth images.

3.3 Regularization

We design a novel rendering pipeline based on Whitted Ray Tracing for the mirror, while a naïve training without regularization always leads to unstable convergence at the mirror where the "appearance" of the mirror is blurred. We find that the bumpy surface of the mirror will greatly affect the quality of reflection due to underconstrained density at the mirror. Thus, we introduce several regularization terms into our optimization process.

3.3.1 Plane Consistency Constraint. As far as we observe, mirrors typically have planar surfaces in the real world. To make full use of this property, we apply the plane consistency constraint proposed by [7] to the surface of the mirror. Specifically, we randomly sample four points A, B, C, D on the surface of the mirror and enforce the normal vector of the plane ABC to be perpendicular to the vector $\vec{AD}$:

$$\mathcal{L}_{pc} = \frac{1}{N_p} \sum_{i=1}^{N_p} |\vec{A_iB_i} \times \vec{A_iC_i} \cdot \vec{A_iD_i}|, \tag{14}$$

where N_p denotes the number of the 4-point sets randomly selected from the planes.

3.3.2 Forward-facing Normal Constraint. With regard to the reflection equation Eq. (10), we find that it still holds when the surface normal rotates 180 degrees and points to the inside of the surface. This ambiguity will incur the incorrect depth of the mirror. To tackle this issue, we follow [26] to enforce that the analytical surface normal $\hat{\mathbf{n}}$ of sampled points makes an obtuse angle with the direction $\mathbf{d}$ of the camera ray $\boldsymbol{r}$, *i.e.*, the surface normal should be forward-facing to the camera.

$$\mathcal{L}_{n_{reg}} = \max(0, \hat{\mathbf{n}} \cdot \mathbf{d}(\boldsymbol{r}))^2. \tag{15}$$

3.3.3 Joint Optimization. In practice, we jointly optimize all networks with the aforementioned losses. In other words, each loss will eventually have an impact on the volume density field and radiance field:

$$\begin{aligned} \mathcal{L} = \lambda_c \mathcal{L}_c + \lambda_m \mathcal{L}_m + \lambda_{pc} \mathcal{L}_{pc} \\ + \lambda_n \mathcal{L}_n + \lambda_{n_{reg}} \mathcal{L}_{n_{reg}}, \end{aligned} \tag{16}$$

where λ is the coefficient of each loss term. Joint optimization will bring three main advantages. First, the surface normal loss $\mathcal{L}_n$ not only influences the $\mathcal{F}_n$ but also encourages $\mathcal{F}_{geo}$ to produce a smooth feature distribution, which makes the volume density uniformly concentrate on the surface to strengthen the flatness of the surface. Second, the reflection probability loss $\mathcal{L}_m$ will promote the volume density field to reach a peak at the mirror, thereby producing an unbiased depth for the mirror. Both of the losses regulate the $\mathcal{F}_{geo}$ through f_{geo}. Third, in spite of the employment of plane and normal constraints, any tiny error of the surface normal will be amplified during the reflection. Through joint optimization, these errors will be iteratively refined since the photometric loss $\mathcal{L}_c$ will implicitly adjust the surface normal $\hat{\mathbf{N}}$ to the desired direction through the differentiable reflection equation.

3.4 Progressive Training Strategy

In the early stage of training, the neural field is unstable and easily falls into the local optimum. We conclude the degeneration situations as two cases: 1) The reflection in the mirror might be learned as a separate scene with inaccurate depth just like NeRF in the case the color converges faster than the geometry. 2) The color may be stuck in a local optimum and blurry if strong geometric regularization is enabled at the beginning. To make training stable, we progressively train the image area inside and outside the mirror and schedule the coefficients of losses at different stages of training. In the initial stage, we enable λ_c and disable the remaining coefficients to maintain the stability of the neural field and avoid the geometry of the mirror being ruined. Furthermore, we replace the $\mathcal{L}_c$ with masked photometric loss $\mathcal{L}_{cm}$:

$$\mathcal{L}_{cm} = \sum_{\boldsymbol{r} \in R_{cam} \bigcap \overline{R_M}} ||\hat{C^P}(\boldsymbol{r}) - C^I(\boldsymbol{r})||_2^2 + \sum_{\boldsymbol{r} \in R_{cam} \bigcap R_M} ||\hat{C^P}(\boldsymbol{r}) - K||_2^2, \tag{17}$$

where R_M is the set of rays hitting the mirror-like surface and $\overline{R_M}$ is the complementary set of R_M. K is a constant vector, which we use $(0, 0, 0)$ in our experiments. The use of K for the image region inside the mirror is to learn an initial rough shape of the mirror without learning its reflection, which will be discussed in Sec. 4.3.2. $\mathcal{L}_{cm}$ is used until the last stage. After a few epochs, we activate the λ_m, λ_{pc}, λ_n, $\lambda_{n_{reg}}$ to regularize the location and geometry of the mirror. After this stage, the accurate depth of the mirror is expected to have been learned by the neural fields. At last, we use $\mathcal{L}_c$ instead of $\mathcal{L}_{cm}$ to jointly optimize the reflection part and refine the geometry of the mirror.

4 EXPERIMENTS

4.1 Datasets

To the best of our knowledge, there is no room-level dataset containing mirrors publicly available for the task of novel view synthesis. Therefore, we create 5 synthetic datasets and capture 4 real datasets with mirrors. Each synthetic dataset is an indoor room downloaded from the BlendSwap [14], including living room, meeting room, washroom, bedroom, and office. Real datasets are captured in real indoor scenes using IPad Pro, including clothing store, lounge, market and discussion room. In each dataset, images are captured 360 degrees around the scene. We split the images as training and test sets to perform the quantitative and qualitative comparison. We use the off-the-shelf segmentation tool [11] to segment the mirror reflection mask in the image.

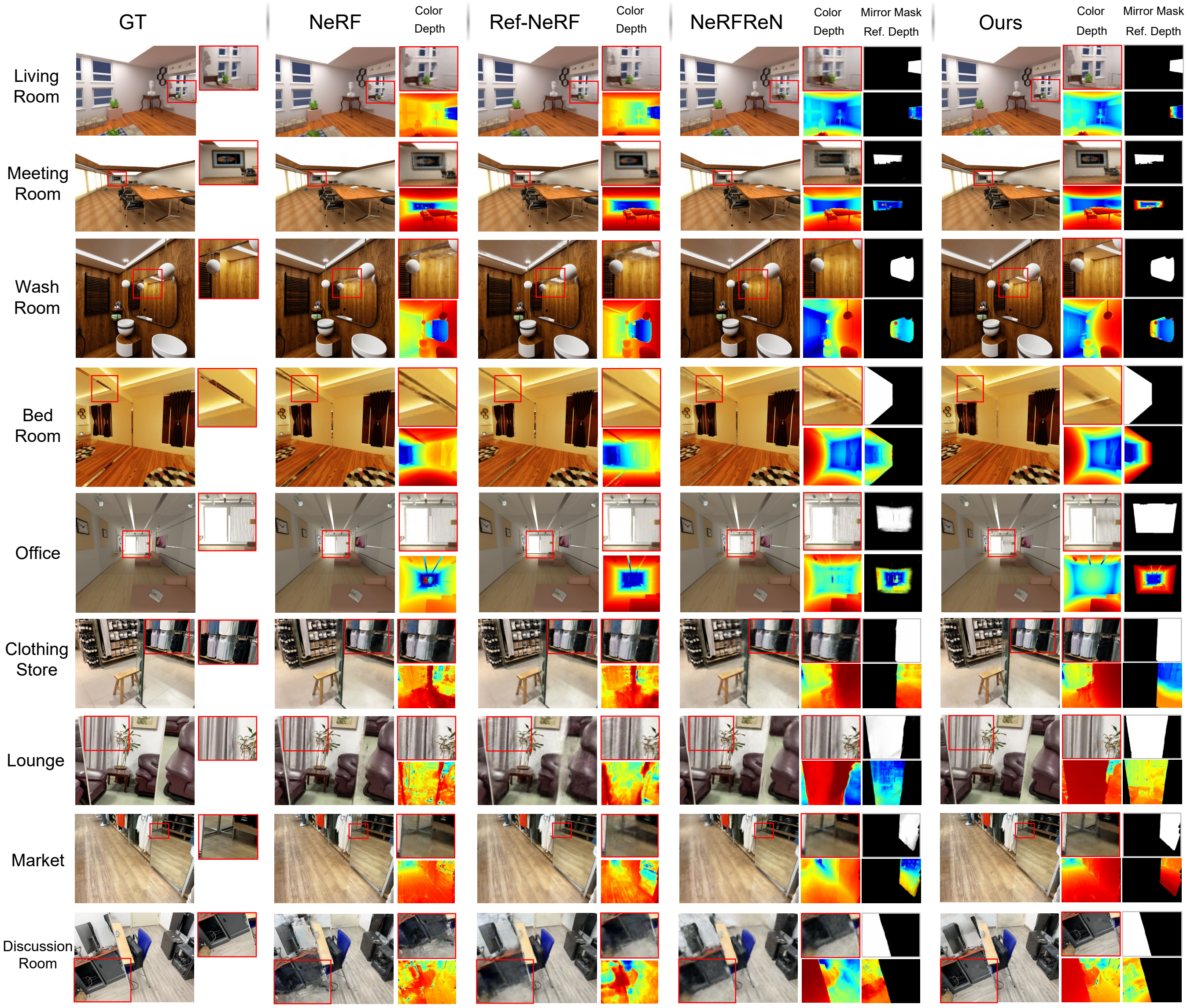


Figure 5: Qualitative comparison of novel view synthesis on synthetic and real scenes with mirrors.

4.2 Comparisons

We compare our method with NeRF [16] and the state-of-the-art neural rendering methods dealing with the reflection, *i.e.*, Ref-NeRF [26] and NeRFReN [9]. The same mirror masks are provided for our method and NeRFReN.

We perform the quantitative comparisons of novel view synthesis on the metrics PSNR, SSIM [34], and LPIPS [47]. As demonstrated in Tab. 1, on the regular test viewpoints, our method outperforms the SOTA methods handling the reflection (*i.e.*, Ref-NeRF and NeRFReN) on both synthetic and real datasets, and is comparable with NeRF. Note that NeRF does not reconstruct the physically sound geometry of the mirror and just interpolates the memorized reflection when performing novel view synthesis, while our method recovers the correct depth of the mirror and enables synthesizing reflections unobserved in training views and multiple applications due to the physical ray-tracing pipeline. Since the above test viewpoints are close to the distribution of training viewpoints, NeRF can generate visually reasonable reflection by interpolating the reflection of nearby views. To compare the correctness of modeling reflection, we capture a set of more challenging test images with more reflections unobserved in the training views. We quantitatively compare the reflection in the mirror, as shown in Tab. 2. Our method surpasses all the compared methods since we can faithfully synthesize the reflection by tracing the reflected ray in the scene. Please refer to the supplementary material for more details.

Qualitative comparisons on the synthetic and real datasets are shown in Fig. 5. NeRF models the scene as a volume of particles that block and emit light [25], and conditions the view-dependent reflection by view direction input. The assumption is suitable well for the Lambertian surface but fails in resolving the reflection in the mirror. The multi-view inconsistent reflection in the mirror will mislead NeRF to learn a separate virtual scene in the mirror, *e.g.*, the inaccurate depth results shown in Fig. 5, since NeRF does not

Methods	Synthetic Datasets			Real Datasets		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
NeRF	28.501	0.903	0.066	25.399	0.788	0.209
Ref-NeRF	28.703	0.905	0.079	24.544	0.730	0.294
NeRFReN	28.483	0.902	0.080	23.191	0.686	0.367
Ours	29.243	0.907	0.077	25.173	0.785	0.205

Table 1: Quantitative comparison of novel views at regular test viewpoints on synthetic and real scenes with mirrors. The best is marked in red and the second is marked in orange.

Methods	Synthetic Datasets			Real Datasets		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
NeRF	23.326	0.964	0.027	19.749	0.886	0.117
Ref-NeRF	22.828	0.964	0.028	20.188	0.897	0.122
NeRFReN	23.542	0.966	0.030	19.174	0.871	0.148
Ours	25.677	0.975	0.021	22.705	0.912	0.085

Table 2: Quantitative comparison of reflections inside the mirror from challenging novel viewpoints out of the training set distribution on synthetic and real scenes.

Settings	PSNR ↑	SSIM ↑	LPIPS ↓
w/o Surface Normal Param.	20.464	0.720	0.349
w/o $\mathcal{L}_{cm}$	28.331	0.878	0.103
w/o Plane Consistency	30.687	0.916	0.058
w/o Forward. Normal Reg.	31.108	0.923	0.052
w/o Joint Optimization	27.691	0.875	0.106
Full Model	32.422	0.933	0.047

Table 3: We quantitatively analyze our model design and training schemes on the synthetic bedroom.

consider the physical reflection in the rendering pipeline. Despite Ref-NeRF's attempt to reproduce reflections by reparameterizing the radiance field using the reflected ray direction and surface materials, it encounters the same limitation as NeRF in reconstructing the mirror's geometry. NeRFReN takes two neural radiance fields to model the scene inside and outside the mirror respectively and can produce the smooth depth of the mirror. However, the above methods synthesize the reflection by interpolating the memorized reflection. The common drawback of these methods is that they cannot synthesize the reflections unobserved in the training set from new viewpoints, *e.g.*, the missing statue in the mirror of the living room, the vanishing ceiling in the mirror of the washroom, and broken cabinet in the mirror of the discussion room in Fig.5. With our neural rendering framework based on physical ray tracing, we can synthesize the reflection of any objects in the scene from arbitrary viewpoints. Moreover, NeRF, Ref-NeRF, and NeRFReN struggle to produce the reflection of the objects whose reflection has high-frequency variations in color, *e.g.*, the distorted hanging picture in the mirror of the meeting room, the blurry curtain in the mirror of the office and the lounge, and the "fogged" clothes in the mirror of the clothing store in Fig.5. By contrast, our method renders detailed reflections of objects by tracing the reflected rays. Compared to NeRFReN, our method can also recover smoother depth of the mirror, *e.g.*, the depth of the mirror from NeRFReN is damaged by the reflection of distant light on the office while our method recovers the mirror depth accurately.

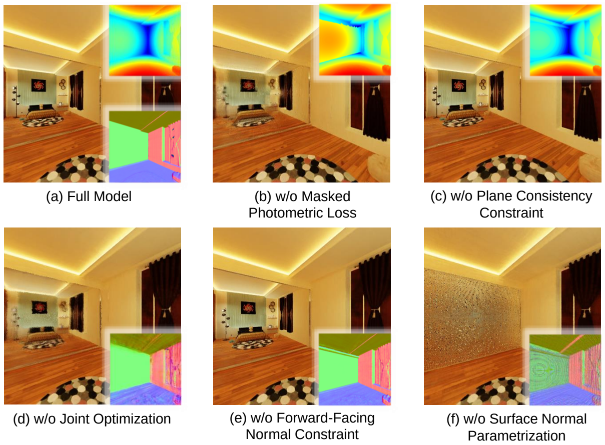

(a) Full Model
(b) w/o Masked Photometric Loss
(c) w/o Plane Consistency Constraint
(d) w/o Joint Optimization
(e) w/o Forward-Facing Normal Constraint
(f) w/o Surface Normal Parametrization

Figure 6: Ablation studies. We qualitatively analyze our model design and training schemes. The top right and bottom right images in each subfigure show the depth and normal map respectively.

4.3 Ablation Studies

We qualitatively and quantitatively analyze our model design and training schemes on the synthetic bedroom in this section, as shown in Fig. 6 and Tab. 3. For more ablation studies, please refer to the supplementary material.

4.3.1 *Smooth Surface Normal Parametrization.*

We first inspect the effectiveness of our surface normal parametrization (Sec.3.1) by using the analytical surface normal from Eq. (3) to calculate the direction of the reflected ray. As depicted in Fig. 6(f) and Tab. 3, the reflection in the mirror is collapsed due to the inevitable noise in the analytical surface normal of the mirror. Instead, our parametrization provides a smooth surface normal with less noise to guide the optimization of the reflection in the mirror.

4.3.2 *Masked Photometric Loss $\mathcal{L}_{cm}$.*

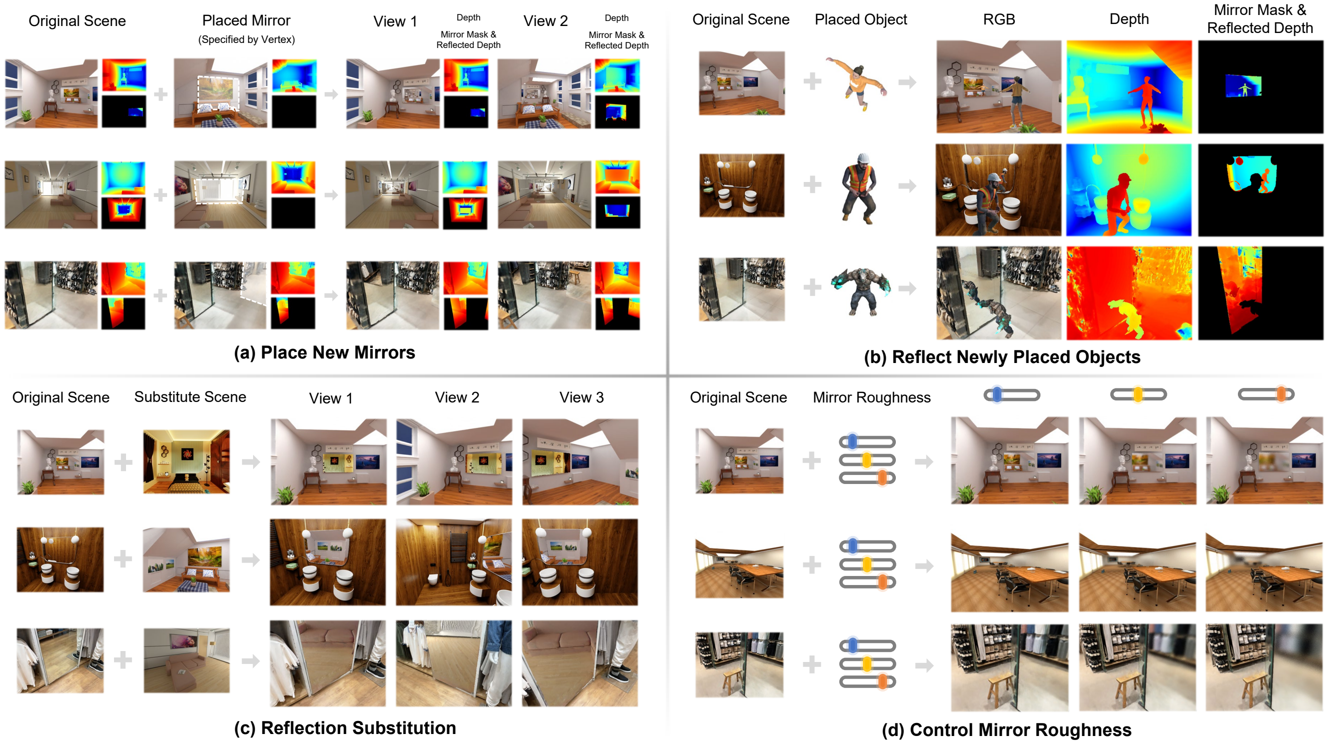
Without the usage of $\mathcal{L}_{cm}$ in the early stage (Sec. 3.4), the depth of the mirror is incorrectly recovered as depicted in Fig. 6(b). The reason for this is that color supervision inside the mirror may lead to the optimization of mirror geometry getting stuck in a local optimum during the initial stages while the mirror geometry has not yet converged.

4.3.3 *Regularization.*

We then analyze the efficacy of each regularization term (Sec.3.3) by turning it off during training. As demonstrated in Fig. 6(c) and Tab. 3, without plane consistency constraint, the discontinuities occur in the depth of the mirror which decreases the image quality. A similar effect happens for the forward-facing normal constraint as shown in Fig. 6 (e). This normal regularization can improve the image quality by correctly orienting the surface normal to the room. Without the joint optimization strategy, the reflection in the mirror is blurred due to the imprecise geometry of the mirror as shown in Fig. 6 (d). When all regularization terms are enabled, we successfully learn the precise reflection in the mirror with the highest image quality.

4.4 Applications

Due to the physical modeling of the mirror reflection, the proposed Mirror-NeRF supports various new scene manipulation applications with mirrors as shown in Fig. 7.



(a) Place New Mirrors

(b) Reflect Newly Placed Objects

(c) Reflection Substitution

(d) Control Mirror Roughness

Figure 7: Applications on synthetic and real scenes with mirrors.

4.4.1 Placing New Mirrors. By tracing the reflected rays at the mirror recursively, it is feasible for our method to integrate new mirrors into the original scene. As shown in Fig. 7(a), we enable the synthesis of novel views involving inter-reflection between the newly placed mirror and the original mirror, *e.g.*, the endless reflection of the room in the new and original mirrors in the first two rows, and the new reflection of the ground in the last row.

4.4.2 Reflecting Newly Placed Objects. We support the composition of multiple neural radiance fields and synthesize new reflections of the composite scenes in the mirror. Specifically, for each traced ray, we detect occlusion by comparing the volume-rendered depth from the radiance fields that have a collision with the ray. The ray will hit the surface with the minimum depth, and terminate or be reflected at the surface. Here we show the composite results of dynamic radiance field D-NeRF [19] with the scene modeled by our method in Fig. 7(b). The reflection of objects from D-NeRF is precisely synthesized in the mirror. This application might be of great use in VR and AR. Please refer to the supplementary video for the vivid dynamic composite results.

4.4.3 Reflection Substitution. In the film and gaming industries, artists may desire to create some magical visual effects, for example, substituting the reflections in the mirror with a different scene. Since we learn the precise geometry of the mirror, it can be easily implemented by transforming the reflected rays at the mirror into another scene and rendering the results of the reflected ray. As shown in Fig. 7(c), we can synthesize the photo-realistic view of the new scene in the mirror with multi-view consistency. Note that in consequence of tracing reflected rays in the new scene, the appearance in the mirror is flipped compared to the new scene.

4.4.4 Controlling the Roughness of Mirrors. According to the microfacet theory [27], the reason why a surface looks rough is that it consists of a multitude of microfacets facing various directions. We support modifying the roughness of the mirror by simulating the microfacet theory. Specifically, we trace the camera ray multiple times following Eq.10 with different random noises added on the surface normal and average the volume-rendered colors to get the final color of this ray. The roughness of the mirror is controlled by the magnitude of noise and the number of tracing times. With this design, we can generate reasonable reflections with different roughness as shown in Fig. 7(d).

5 CONCLUSION

We have proposed a novel neural rendering framework following Whitted Ray Tracing, which synthesizes photo-realistic novel views in the scene with the mirror and learns the accurate geometry and reflection of the mirror. Besides, we support various scene manipulation applications with mirrors. As a limitation, our method does not explicitly estimate the location of the light source in the room, which prevents us from relighting the room. The refraction is also not modeled in our framework since we focus on mirrors currently, and it is naturally compatible with our ray tracing pipeline and considered as future work.

ACKNOWLEDGMENTS

This work was partially supported by the NSFC (No. 62102356) and Ant Group.

REFERENCES

[1] Dejan Azinović, Ricardo Martin-Brualla, Dan B Goldman, Matthias Nießner, and Justus Thies. 2022. Neural rgb-d surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6290–6301.

[2] Chong Bao, Yinda Zhang, Bangbang Yang, Tianxing Fan, Zesong Yang, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. 2023. SINE: Semantic-driven Image-based NeRF Editing with Prior-guided Editing Field. *arXiv preprint arXiv:2303.13277* (2023).

[3] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824* (2020).

[4] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. 2021. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12684–12694.

[5] Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan Barron, Hendrik Lensch, and Varun Jampani. 2022. Samurai: Shape and material from unconstrained real-world arbitrary image collections. *Advances in Neural Information Processing Systems* 35 (2022), 26389–26403.

[6] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. 2021. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems* 34 (2021), 10691–10704.

[7] Zheng Chen, Chen Wang, Yuan-Chen Guo, and Song-Hai Zhang. 2022. Struct-NeRF: Neural Radiance Fields for Indoor Scenes with Structural Hints. *arXiv preprint arXiv:2209.05277* (2022).

[8] Haoyu Guo, Sida Peng, Haotong Lin, Qianqian Wang, Guofeng Zhang, Hujun Bao, and Xiaowei Zhou. 2022. Neural 3d scene reconstruction with the manhattan-world assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5511–5520.

[9] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. 2022. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18409–18418.

[10] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. 2023. TensoIR: Tensorial Inverse Rendering. *arXiv preprint arXiv:2304.12461* (2023).

[11] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]

[12] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. 2022. NeROIC: neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–12.

[13] Hai Li, Xingrui Yang, Hongjia Zhai, Yuqian Liu, Hujun Bao, and Guofeng Zhang. 2022. Vox-Surf: Voxel-based implicit surface representation. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[14] John Roper Matthew Muldoon. 2022. BlenderSwap. https://www.blenderswap.com/. Accessed: 2022-11-10.

[15] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. 2019. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.

[16] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proceedings of European Conference on Computer Vision*. 405–421.

[17] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8280–8290.

[18] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. 2021. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9054–9063.

[19] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. 2021. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10318–10327.

[20] Yawar Siddiqui, Lorenzo Porzi, Samuel Rota Buló, Norman Müller, Matthias Nießner, Angela Dai, and Peter Kontschieder. 2022. Panoptic Lifting for 3D Scene Understanding with Neural Fields. *arXiv preprint arXiv:2212.09802* (2022).

[21] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. 2020. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems* 33 (2020), 7462–7473.

[22] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. 2021. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7495–7504.

[23] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J Davison. 2021. iMAP: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 6229–6238.

[24] Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. 2022. Generalizable patch-based neural rendering. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*. Springer, 156–174.

[25] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. 2022. Advances in neural rendering. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 703–735.

[26] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. 2021. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. *arXiv preprint arXiv:2112.03907* (2021).

[27] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*. 195–206.

[28] Bing Wang, Lu Chen, and Bo Yang. 2022. DM-NeRF: 3D Scene Geometry Decomposition and Manipulation from 2D Images. *arXiv preprint arXiv:2208.07227* (2022).

[29] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. 2022. NeRF-SR: High Quality Neural Radiance Fields using Supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*. 6445–6454.

[30] Jingwen Wang, Tymoteusz Bleja, and Lourdes Agapito. 2022. Go-surf: Neural feature grid optimization for fast, high-fidelity rgb-d surface reconstruction. *arXiv preprint arXiv:2206.14735* (2022).

[31] Liao Wang, Ziyu Wang, Pei Lin, Yuheng Jiang, Xin Suo, Minye Wu, Lan Xu, and Jingyi Yu. 2021. ibutter: Neural interactive bullet time generator for human free-viewpoint rendering. In *Proceedings of the 29th ACM International Conference on Multimedia*. 4641–4650.

[32] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689* (2021).

[33] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul P Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. 2021. Ibrnet: Learning multi-view image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4690–4699.

[34] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[35] Frederik Warburg, Ethan Weber, Matthew Tancik, Aleksander Holynski, and Angjoo Kanazawa. 2023. Nerfbusters: Removing Ghostly Artifacts from Casually Captured NeRFs. *arXiv preprint arXiv:2304.10532* (2023).

[36] Yi Wei, Shaohui Liu, Yongming Rao, Wang Zhao, Jiwen Lu, and Jie Zhou. 2021. Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5610–5619.

[37] Turner Whitted. 2005. An improved illumination model for shaded display. In *ACM Siggraph 2005 Courses*. 4–es.

[38] Tong Wu, Jiaqi Wang, Xingang Pan, Xudong Xu, Christian Theobalt, Ziwei Liu, and Dahua Lin. 2022. Voxurf: Voxel-based Efficient and Accurate Neural Surface Reconstruction. *arXiv preprint arXiv:2208.12697* (2022).

[39] Zijin Wu, Xingyi Li, Juewen Peng, Hao Lu, Zhiguo Cao, and Weicai Zhong. 2022. Dof-nerf: Depth-of-field meets neural radiance fields. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1718–1729.

[40] Bangbang Yang, Chong Bao, Junyi Zeng, Hujun Bao, Yinda Zhang, Zhaopeng Cui, and Guofeng Zhang. 2022. Neumesh: Learning disentangled neural mesh-based implicit field for geometry and texture editing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVI*. Springer, 597–614.

[41] Bangbang Yang, Yinda Zhang, Yijin Li, Zhaopeng Cui, Sean Fanello, Hujun Bao, and Guofeng Zhang. 2022. Neural rendering in a room: amodal 3d understanding and free-viewpoint rendering for the closed scene composed of pre-captured objects. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–10.

[42] Bangbang Yang, Yinda Zhang, Yinghao Xu, Yijin Li, Han Zhou, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. 2021. Learning object-compositional neural radiance field for editable scene rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13779–13788.

[43] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. 2020. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems* 33 (2020), 2492–2502.

[44] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. 2022. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *arXiv preprint arXiv:2206.00665* (2022).

[45] Jian Zhang, Yuanqing Zhang, Huan Fu, Xiaowei Zhou, Bowen Cai, Jinchi Huang, Rongfei Jia, Binqiang Zhao, and Xing Tang. 2022. Ray priors through reprojection: Improving neural radiance fields for novel view extrapolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18376–18386.

[46] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5453–5462.

[47] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.

[48] Xiaoshuai Zhang, Sai Bi, Kalyan Sunkavalli, Hao Su, and Zexiang Xu. 2022. Nerfusion: Fusing radiance fields for large-scale scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5449–5458.

[49] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–18.

[50] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. 2022. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18643–18652.

[51] Youjia Zhang, Teng Xu, Junqing Yu, Yuteng Ye, Junle Wang, Yanqing Jing, Jingyi Yu, and Wei Yang. 2023. NeMF: Inverse Volume Rendering with Neural Microflake Field. *arXiv preprint arXiv:2304.00782* (2023).

[52] Boming Zhao, Bangbang Yang, Zhenyang Li, Zuoyue Li, Guofeng Zhang, Jiashu Zhao, Dawei Yin, Zhaopeng Cui, and Hujun Bao. 2022. Factorized and controllable neural re-rendering of outdoor scene for photo extrapolation. In *Proceedings of the 30th ACM International Conference on Multimedia*. 1455–1464.

[53] Shuaifeng Zhi, Tristan Laidlow, Stefan Leutenegger, and Andrew J Davison. 2021. In-place scene labelling and understanding with implicit scene representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15838–15847.

[54] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R Oswald, and Marc Pollefeys. 2022. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12786–12796.