

# NeRF as a Non-Distant Environment Emitter in Physics-based Inverse Rendering

Jingwang Ling

lingjw20@mails.tsinghua.edu.cn  
School of Software and BNRIst,  
Tsinghua University  
China

Ruihan Yu

auroraryan0301@gmail.com  
Department of Physics,  
Tsinghua University  
China

Feng Xu\*

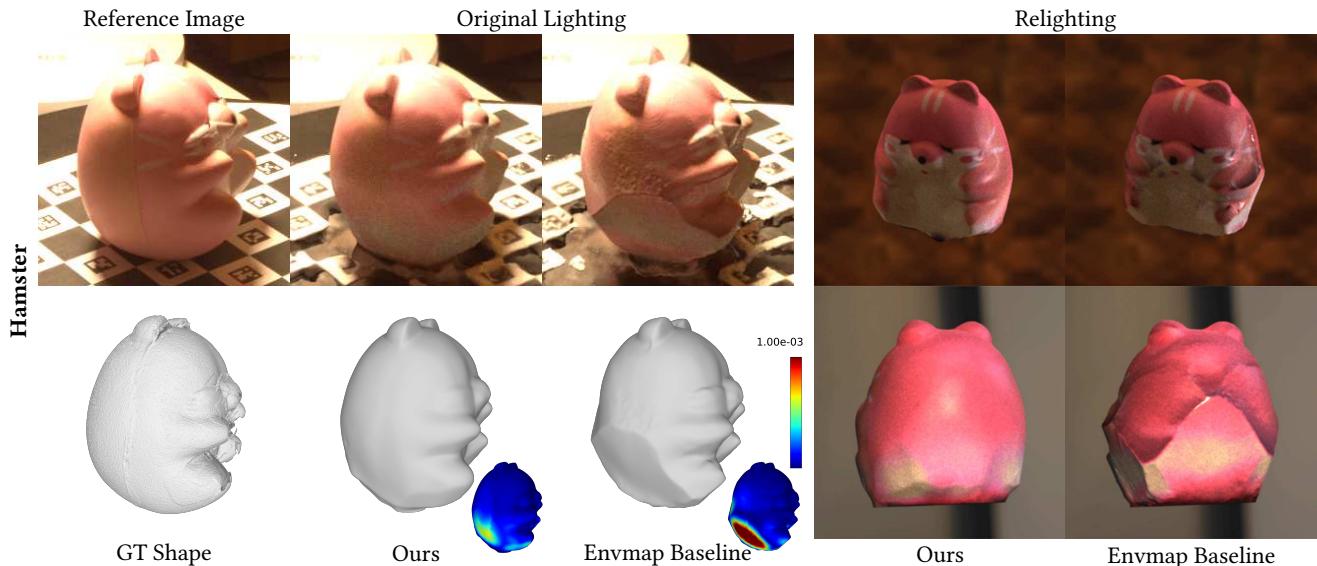
xufeng2003@gmail.com  
School of Software and BNRIst,  
Tsinghua University  
China

Chun Du

duchun@utibet.edu.cn  
Tibet University  
China

Shuang Zhao

shz@ics.uci.edu  
University of California, Irvine  
USA



**Figure 1:** We perform physics-based inverse rendering on real captured data featuring non-distant lighting. In this scenario, the commonly used environment map inaccurately models the lighting, leading to artifacts visible in relighting and shape reconstruction results. In contrast, our proposed NeRF emitter accurately models non-distant lighting, achieving high-quality inverse rendering.

## ABSTRACT

Physics-based inverse rendering enables joint optimization of shape, material, and lighting based on captured 2D images. To ensure accurate reconstruction, using a light model that closely resembles the captured environment is essential. Although the widely adopted distant environmental lighting model is adequate in many

\*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

SIGGRAPH Conference Papers '24, July 27–August 01, 2024, Denver, CO, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 979-8-4007-0525-0/24/07  
<https://doi.org/10.1145/3641519.3657404>

cases, we demonstrate that its inability to capture spatially varying illumination can lead to inaccurate reconstructions in many real-world inverse rendering scenarios. To address this limitation, we incorporate NeRF as a non-distant environment emitter into the inverse rendering pipeline. Additionally, we introduce an emitter importance sampling technique for NeRF to reduce the rendering variance. Through comparisons on both real and synthetic datasets, our results demonstrate that our NeRF-based emitter offers a more precise representation of scene lighting, thereby improving the accuracy of inverse rendering.

## CCS CONCEPTS

- Computing methodologies → Rendering; Reconstruction.

## KEYWORDS

Inverse Rendering, Neural Radiance Fields, Lighting Estimation

**ACM Reference Format:**

Jingwang Ling, Ruihan Yu, Feng Xu, Chun Du, and Shuang Zhao. 2024. NeRF as a Non-Distant Environment Emitter in Physics-based Inverse Rendering. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers '24 (SIGGRAPH Conference Papers '24), July 27–August 01, 2024, Denver, CO, USA*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3641519.3657404>

## 1 INTRODUCTION

Reconstructing the shape and material of an object as well as the lighting condition from 2D images has been a long-standing challenge in computer graphics and vision, offering numerous applications in 3D reconstruction and scene digitization. Thanks to recent advancements, physics-based inverse rendering has gained popularity for its ability to accurately simulate light transport within complex environments and to produce high-quality reconstructions. Also known as *analysis by synthesis*, inverse rendering is typically formulated as optimization problems that seek for shape, material, and lighting parameters to minimize the difference (measured with some predetermined *loss*) between captured input images and renderings generated with these parameters. To solve these optimizations efficiently using gradient-based methods (such as stochastic gradient descent), it is desired to incorporate *differentiable rendering* in inverse-rendering pipelines.

To accurately simulate light transport for inverse rendering, having an accurate scene model is crucial. However, in commonly used object-centric capture scenarios, the limitation in the number of captured images and the relatively localized camera positions pose challenges in reconstructing the entire scene. As a result, inverse rendering approaches tend to model the object of interest in terms of shape and material, while approximating the rest of the scene as a distant “environment”. The environment map [Debevec and Malik 1997] has thus become a popular method for approximating lighting around objects. However, we demonstrate that in situations where the light source is not infinitely distant, an environment map proves to be an inadequate approximation, leading to inaccurate inverse-rendering results.

The crucial issue lies in the distant lighting assumption of the environment map, where the lighting distribution is considered spatially invariant. However, when the light source is not distant, it exhibits strong parallax effects, as the light arrives from various directions at different positions on the object’s surface (see Figure 2 right). Consequently, a spatially invariant lighting model becomes an inadequate approximation for the actual lighting conditions. To effectively model non-distant lighting, a representation capable of synthesizing spatially varying incoming radiance distribution is necessary.

We have observed that a neural radiance field (NeRF) [Mildenhall et al. 2020] is well-suited for representing real, spatially varying lighting. While an environment map essentially models a 2D radiance field at an infinite distance, NeRF models a radiance field that resides within 3D non-distant volumetric densities. While NeRF is originally used for novel-view synthesis, we demonstrate its extension to light modeling, as moving the shading point is analogous to moving the viewpoint. With an HDR NeRF representing the unbounded scene surrounding the object, we can synthesize 3D-consistent incoming radiance at any shading point on the object.

Therefore, to achieve more accurate inverse rendering when the lighting is not distant, we propose a technique that utilizes a NeRF to model the environmental lighting. We model the scene in a hybrid manner, with the object of interest represented by a surface and the surrounding by a NeRF. To render these hybrid scenes, we generalize the standard surface-only rendering equation to use NeRF-based illumination. Additionally, we introduce an importance sampling scheme for NeRFs. Then, we build an inverse rendering pipeline that effectively reconstructs shape, material, and NeRF lighting from captured images.

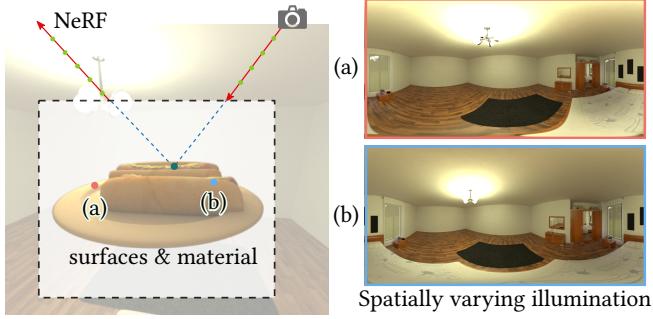
We capture both real and synthetic datasets featuring non-distant lighting to compare our method with a baseline pipeline that employs environment map emitters. Our results demonstrate that an environment map inaccurately models non-distant lighting, resulting in artifacts in inverse-rendering results. In contrast, our proposed NeRF emitter accurately models non-distant lighting and achieves high-quality reconstructions.

## 2 RELATED WORK

*Neural Radiance Fields (NeRF)* [Mildenhall et al. 2020] employ neural networks to model the radiance field in 3D scenes, enabling novel-view synthesis. Mip-NeRF 360 [Barron et al. 2022] proposes a scene contraction method that extends NeRF to unbounded scenes, enabling NeRF to model the surround environment of an object of interest. Instant-NGP [Müller et al. 2022] accelerates NeRF by proposing a hybrid grid-network field representation. NeRFStudio [Tancik et al. 2023] and Zip-NeRF [Barron et al. 2023] combine the works by Barron et al. [2022]; Müller et al. [2022] to accelerate the modeling of an unbounded environment using NeRF. RawNeRF [Mildenhall et al. 2022] and VR-NeRF [Xu et al. 2023a] explore extending NeRF to high dynamic range, thereby improving the novel-view synthesis quality in dark-light or virtual reality scenarios. Considering that light sources in physics-based rendering are often HDR, integrating HDR technique into NeRF holds the potential to represent an HDR environment emitter.

Based on the powerful modeling capabilities of neural networks, several research efforts have delved into leveraging neural networks to represent complex luminaires. The works by Zhu et al. [2021]; Condor and Jarabo [2022] employ neural networks to model the light field or radiance field of virtual luminaries, thereby accelerating physics-based rendering by reducing variance. DMRF [Qiao et al. 2023] explores the insertion of a virtual mesh into the trained NeRF and simulates the interactions between them. While these works focus on using pretrained networks to model light in forward rendering, our work considers employing NeRF in inverse rendering as an emitter to reconstruct scene parameters.

*Differentiable Rendering* involves differentiating the rendering equation to obtain gradients of scene parameters, which is the foundation of gradient-based optimization in the subsequent inverse rendering. A key challenge in differentiable rendering lies in the differentiation of visibility discontinuities, which is essential to enable shape optimization. The works by Li et al. [2018]; Zhang et al. [2019, 2020] explicitly sample edges at the silhouettes of shapes to estimate the boundary term that arises when differentiating the integral over visibility discontinuities. Another approach, based on warped-area sampling [Bangaru et al. 2020], sidesteps visibility discontinuities



**Figure 2: The region within the bounding box is modeled by surfaces and material, while the region outside is handled by NeRF to account for environmental lighting. The NeRF-synthesized illumination viewed from two shading points (red (a) and blue (b) dots) is visualized on the right.**

by converting boundary integrals into area integrals. This approach can be extended from meshes to signed distance functions (SDF) to enable differentiable SDF rendering [Vicini et al. 2022; Bangaru et al. 2022], and has also recently been extended to path space to differentiate path integrals [Xu et al. 2023b]. Deriving differentiable volume rendering has been previously explored by Zhang et al. [2019], but our model distinguishes itself by incorporating emissive volume and excluding scattering.

*Inverse Rendering* aims to utilize gradient descent to minimize rendering loss between rendered and captured images to reconstruct the scene shape, material, and lighting parameters. Accurate lighting models are crucial to achieve high-quality reconstruction via inverse rendering. However, current inverse rendering methods often rely on simplistic light models, which either restrict the capture setup or lead to inaccurate approximations of real-world scene lighting. Some works require specific capture configurations in controlled environments, such as a co-located camera-light setup [Bi et al. 2020b;a; Zhang et al. 2022a] or a point light source [Yang et al. 2022a,b; Ling et al. 2023]. For inverse rendering in ordinary environments, an environment map [Debevec and Malik 1997] commonly serves as a lighting approximation. Numerous works utilize inverse rendering for reconstructing individual objects [Zhang et al. 2021; Boss et al. 2021a,b, 2022], outdoor buildings [Rudnev et al. 2021], and humans [Chen and Liu 2022]. There is also a trend in inverse rendering to be more physically accurate by considering shadows [Hasselgren et al. 2022; Srinivasan et al. 2021], interreflections [Zhang et al. 2022b; Wu et al. 2023; Jin et al. 2023], and differentiable path tracing [Sun et al. 2023]. All of the above methods employ an environment map as the scene lighting model. However, an environment map assumes that the scene lighting is infinitely distant, which is a rare case in real-world capture setups. We demonstrate that using an environment map can result in degraded reconstruction quality when this assumption is violated. Therefore, we propose a NeRF emitter module designed to be a universal component applicable across inverse rendering systems, avoiding the distant lighting assumption.

### 3 PRELIMINARIES

We revisit differentiable surface rendering [Nimier-David et al. 2020] and NeRF [Mildenhall et al. 2020], comparing their similarities and distinctions. Image pixels  $I_1, I_2, \dots, I_k, \dots, I_N$  measure the integral of the dot product between the sensor importance function  $W$  and the incident radiance  $L_i$  at a position  $\mathbf{p}$ :

$$I_k = \int_{\mathcal{A}} \int_{\mathcal{S}^2} W_k(\mathbf{p}, \omega) L_i(\mathbf{p}, \omega) d\omega^\perp d\mathbf{p}, \quad (1)$$

where  $\omega$  denotes the light direction and  $\omega^\perp$  is the projected solid angle. Surface and volume rendering exhibit differences in assumptions about the scene, affecting  $L_i(\mathbf{p}, \omega)$ . Surface rendering assumes a vacuum between surfaces, where incoming radiance equals outgoing radiance at the first intersection point of the ray, thus  $L_i$  is given by the surface version  $L_i^s$ :

$$L_i^s(\mathbf{p}, \omega) = L_o(\mathbf{r}(\mathbf{p}, \omega, t_0), -\omega), \quad (2)$$

where  $t_0$  is the distance to the closest intersection point  $\mathbf{r}(\mathbf{p}, \omega, t_0)$  along direction  $\omega$ . Outgoing radiance  $L_o$  involves an integral over the hemisphere, expressed by the rendering equation:

$$L_o(\mathbf{p}, \omega) = L_e(\mathbf{p}, \omega) + \int_{\mathcal{S}^2} L_i(\mathbf{p}, \omega') f_s(\mathbf{p}, \omega, \omega') d\omega'^\perp, \quad (3)$$

where  $L_e$  represents the surface emitted radiance and  $f_s$  denotes the BSDF. The  $L_i$  on the right-hand side involves recursive computation and can be implemented using path tracing.

On the other hand, NeRF makes different assumptions about the scene. It assumes that the scene is filled with emissive volume, without surfaces or volumetric scattering. In this case,  $L_i$  is determined by the volumetric version  $L_i^v$ , calculated using the (non-scattering) volume rendering equation:

$$L_i^v(\mathbf{p}, \omega) = \int_0^\infty T(\mathbf{p}, \omega, t) \sigma(\mathbf{r}(\mathbf{p}, \omega, t)) \mathbf{c}(\mathbf{r}(\mathbf{p}, \omega, t), -\omega) dt, \quad (4)$$

where  $\sigma$  represents the volumetric density,  $\mathbf{c}$  is the emission, and  $T$  denotes the transmittance, which models the absorption effect of occluding densities:

$$T(\mathbf{p}, \omega, t) = \exp \left( - \int_0^t \sigma(\mathbf{r}(\mathbf{p}, \omega, s)) ds \right). \quad (5)$$

In differentiable surface rendering, we need to differentiate Equation 3. Assuming we eliminate the discontinuities using techniques like reparameterization [Bangaru et al. 2020; Vicini et al. 2022], we obtain

$$\begin{aligned} \partial_{\mathbf{x}} L_o(\mathbf{p}, \omega) = \partial_{\mathbf{x}} L_e(\mathbf{p}, \omega) + \int_{\mathcal{S}^2} & [\partial_{\mathbf{x}} L_i(\mathbf{p}, \omega') f_s(\mathbf{p}, \omega, \omega') \\ & + L_i(\mathbf{p}, \omega') \partial_{\mathbf{x}} f_s(\mathbf{p}, \omega, \omega')] d\omega'^\perp, \end{aligned} \quad (6)$$

where  $\partial_{\mathbf{x}} L_i(\mathbf{p}, \omega')$  represents the derivative provided by the light source when it is directly visible from the direction  $\omega'$ .

The differentiation of NeRF is straightforward. Equation 4 can be computed via ray marching and is trivially differentiable. Gradients can be computed using automatic differentiation such as PyTorch [Paszke et al. 2019].

## 4 NERF-BASED NON-DISTANT EMITTERS

### 4.1 Hybrid Rendering of Surfaces and NeRF

We propose a NeRF emitter to model the environment lighting surrounding the object. We adopt a hybrid scene representation containing surfaces and NeRF. As depicted in Figure 2 left, we partition the scene into two regions based on the bounding box of the object: the internal region is represented by surface and material, while the external region is represented by the NeRF-based environmental lighting. Rendering the hybrid scene involves a rendering equation that accounts for both surfaces and NeRF. According to our scene assumption, the incoming radiance  $L_i$  consists of  $L_i^s$  from surfaces and  $L_i^v$  from the NeRF:

$$L_i(\mathbf{p}, \omega) = L_i^s(\mathbf{p}, \omega) + L_i^v(\mathbf{p}, \omega). \quad (7)$$

Considering the surface contribution  $L_i^s$ , it follows the surface light transport, with additional consideration of the occluding NeRF:

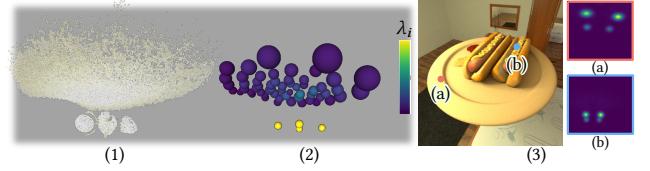
$$L_i^s(\mathbf{p}, \omega) = T(\mathbf{p}, \omega, t_0)L_o(\mathbf{r}(\mathbf{p}, \omega, t_0), -\omega), \quad (8)$$

where  $T$  represents the NeRF transmittance defined in Equation 5. The NeRF contribution  $L_i^v$  resembles Equation 4, but with zero densities within the bounding box. Furthermore, we integrate from 0 to  $t_0$  instead of  $\infty$ , where  $t_0$  denotes the distance to the first surface intersection.  $t_0$  goes to  $\infty$  when there is no surface intersection. The  $L_i$  defined in Equation 7 is used recursively in the computation of Equation 3 in path tracing. When the light path eventually exits the bounding box and goes to infinity, NeRF is queried to calculate the incoming radiance from environment lighting, akin to the behavior of a conventional light source.

Physics-based rendering necessitates its components, such as NeRF in our case, to adhere to physical correctness. While not often used in a physics-based context, we recognize that NeRF, fundamentally representing a radiance field, is well suited for modeling the physical quantity of “radiance” within a scene. NeRF simplifies light transport outside the bounding box by considering interreflections as emissions, yet the light paths within the bounding box are preserved. Although this simplification cannot capture the illumination changes resulting from scene edits after reconstruction, it effectively models interreflections during the capture phase, hence suitable for inverse rendering.

Our formulation allows for arbitrary bounding box delineation of the regions. We allow users to place a bounding box to denote the object of their interest. Light sources within the bounding box are accounted for with the self-emission  $L_e$  in Equation 3. Any part of the scene not enclosed within the bounding box is treated as part of the environmental light and modeled using NeRF. In the experiments, to mitigate inverse rendering ambiguity, we introduce a prior by constraining emission within the bounding box to zero. Imposing this prior however requires all emissive objects, such as area lights, to be located outside the bounding box. This prior is optional and not a fundamental limitation.

The scene properties also simplify NeRF computations to at most two rays in a light path. Assuming consistently zero densities within the bounding box, and given its convex nature, any rays originating and terminating within it remain inside, with zero densities along them. This leaves only the necessity to compute NeRF for the two rays entering or exiting the bounding box.



**Figure 3: We generate importance sampling distributions by (1) creating a point cloud of the bright parts of the NeRF, (2) clustering it into Gaussian mixtures, and (3) projecting these Gaussians to vMFs at shading points (red (a) and blue (b) dots).**

### 4.2 Emitter Importance Sampling for NeRF

To integrate the NeRF emitter into physically-based rendering, Monte Carlo sampling is required. This entails the ability to sample directions and compute the corresponding probability densities. Importance sampling for NeRF presents unique challenges due to its nature as a volumetric light source governed by a density field with without fixed topology. Hence, a specific importance sampling strategy is imperative for NeRF.

Our basic idea is to approximate regions of significant radiance contribution in NeRF using simple geometric primitives. Specifically, we employ a Gaussian mixture model to fit a point cloud extracted from the bright regions of NeRF, acting as a proxy model during importance sampling. First, we randomly sample rays originating from positions inside the bounding box. We use Equation 4 to compute the radiance along this ray and record the depth value that maximally contributes to the rendered radiance

$$t_{\max} = \arg \max_t T(\mathbf{p}, \omega, t)\sigma(\mathbf{r}(\mathbf{p}, \omega, t))\mathbf{c}(\mathbf{r}(\mathbf{p}, \omega, t), -\omega). \quad (9)$$

Then we add a sample point to the point cloud at the position  $\mathbf{r}(\mathbf{p}, \omega, t_{\max})$  with the weight  $Y(L_i^v(\mathbf{p}, \omega))$ , where  $Y$  is the operator that converts RGB radiance to monochrome luminance. We aim to approximate the radiance distribution of the point cloud via fitting an isotropic Gaussian mixture model of  $M = 64$  components. To efficiently minimize rendering variance using a limited number of components, inspired by MIS compensation [Karlik et al. 2019], we subtract the mean weight  $\bar{Y}(L_i^v(\mathbf{p}, \omega))$  from the weights of all the points. Only points with positive weights are retained. As a result, the remaining points (Figure 3 (a)) can focus on accurately capturing the shape of very bright regions, while remaining unbiased via the multiple-importance sampling combined with BSDF sampling. Based on the compensated weights, we cluster the point cloud into  $M$  isotropic Gaussians (Figure 3 (b)), each characterized by a mean  $\mu_i$ , covariance  $\kappa_i^{-1}\mathbf{I}$  and weight  $\lambda_i$ , with  $i \in [1, M]$ . When performing emitter importance sampling for a shading point  $\mathbf{p}$ , we first project the isotropic Gaussians into von Mises–Fisher distributions ( $\hat{\mu}_i, \hat{\kappa}_i, \hat{\lambda}_i$ ):

$$\hat{\mu}_i = \frac{\mu_i - \mathbf{p}}{\|\mu_i - \mathbf{p}\|_2}, \quad \hat{\kappa}_i = \frac{\kappa_i}{\|\mu_i - \mathbf{p}\|_2}, \quad \hat{\lambda}_i = \lambda_i, \quad (10)$$

and use the projected vMF lobes (Figure 3 (c)) to perform sampling according to the work by Jakob [2012]. In inverse rendering, as NeRF parameters update during joint optimization, the brightness distribution may change. To ensure the sampling distribution

matches the updated NeRF, we periodically repeat the above process to establish an updated Gaussian mixture model.

### 4.3 Differentiable Hybrid Rendering

Differentiable rendering of a hybrid scene, as defined in Section 4.1, requires specific considerations. First, we must address  $\partial_x L_i(\mathbf{p}, \omega')$  in Equation 6, which calculates the gradient of the light with respect to an arbitrary scene parameter  $x$ . As Equation 6 relies on detached sampling [Zeltner et al. 2021],  $\omega'$  does not receive gradients and  $\partial_{\omega'} L_i(\mathbf{p}, \omega')$  is therefore zero. However, it remains crucial to compute  $\partial_p L_i(\mathbf{p}, \omega')$ , essential for differentiating shape parameters. Note that  $\mathbf{p}$  also serves as the ray starting position input to NeRF. Therefore, we need to compute gradients of the NeRF radiance with respect to the ray starting position via automatic differentiation and propagate it to Equation 6.

Second, when we differentiate the volume rendering equation, we obtain a boundary term due to the dependence of the upper integration limit,  $t_0$ , on the scene's shape parameters:

$$\begin{aligned} \partial_x L_i'(\mathbf{p}, \omega) = & T(\mathbf{p}, \omega, t_0) \sigma(\mathbf{r}(\mathbf{p}, \omega, t_0)) \mathbf{c}(\mathbf{r}(\mathbf{p}, \omega, t_0), -\omega) \partial_x t_0 \\ & + \int_0^{t_0} \partial_x(T(\mathbf{p}, \omega, t)) \sigma(\mathbf{r}(\mathbf{p}, \omega, t)) \mathbf{c}(\mathbf{r}(\mathbf{p}, \omega, t), -\omega) dt. \end{aligned} \quad (11)$$

However, densities  $\sigma$  approach zero continuously at the bounding box's boundaries and are zero within it. Since  $\mathbf{r}(\mathbf{p}, \omega, t_0)$  always stays within the bounding box, this boundary term always evaluates to zero.

## 5 INVERSE RENDERING USING NERF EMITTER

Building upon the derivations above, we've implemented a pipeline that integrates NeRF as an emitter into physics-based inverse rendering. While previous inverse rendering methods typically require object foreground masks to exclude background influences, our pipeline can simultaneously obtain scene information from both foreground and background pixels and jointly optimize shape, material, and NeRF emitter based on the gradients of the rendering loss. Physics-based inverse rendering often requires proper initialization. Therefore, we designed a multi-stage optimization process that utilizes NeRF to aid the initialization of inverse rendering and facilitate the optimization process. Additionally, due to the widespread use of environment maps, existing datasets tend to avoid using nearby light sources, leading to a lack of thorough testing for scenes with close light sources. We established a capture system to thoroughly test such scenarios.

### 5.1 Multi-Stage Optimization

Compared to surface-based inverse rendering, NeRF demonstrates robustness in parameter initialization. Therefore, we first conduct ordinary NeRF training to initialize the scene, with the interior of the bounding box represented using either NeRF densities or NeuS [Wang et al. 2021; Yariv et al. 2023; Ge et al. 2023]. We then extract the geometry within the bounding box using truncated signed distance function (TSDF) fusion [Curless and Levoy 1996] to initialize the object shape, and clear the densities inside the bounding box. The object shape will be continually refined by inverse rendering. At this point, NeRF has a reasonable initial lighting approximation,

and the object's geometry has been initialized through NeRF. Finally, we perform joint optimization via inverse rendering until convergence.

### 5.2 Non-Distant Emitter Capture System

To thoroughly test inverse rendering on scenes with non-distant emitters, we establish a capture setup to capture a dataset featuring such scenarios. We additionally provide multiple lighting conditions by rotating the object to mitigate ambiguity under single-light conditions. Each scene includes a nearby light source placed around the object. We employ a turntable to rotate the object, capturing multi-view images with a camera for each rotation pose. While the camera's viewpoint remains focused on the object, lighting cues can be observed from background pixels, providing accurate lighting information. To accurately capture the lighting, we utilize a DSLR camera to capture HDR images.

## 6 IMPLEMENTATION DETAILS

### 6.1 HDR NeRF Parameterization and Training

Our NeRF model builds upon the nerfacto proposed by Tancik et al. [2023]. To faithfully represent the emitter, it's essential to train NeRF to output HDR radiance values. Our inverse rendering pipeline takes HDR images as input. Following the approach in RawNeRF [Mildenhall et al. 2022], we use an exponential activation function  $\exp(x - 5)$  as the output activation function for NeRF radiance. We utilize the relative L1 Loss as the rendering loss function:

$$\mathcal{L}_{\text{render}} = \frac{1}{N} \sum_{k=1}^N \left\| \frac{I_k - \widehat{I}_k}{\text{sg}(I_k) + \epsilon} \right\|_1, \quad (12)$$

where  $I_k$  represents the rendered pixel,  $\text{sg}(\cdot)$  indicates a stop-gradient operator,  $\widehat{I}_k$  is the pixel on the captured image, and  $\epsilon = 10^{-3}$  is a small number introduced to downweight excessively dark pixels. We use gradient scaling [Philip and Deschaintre 2023] to prevent NeRF from generating floaters. We adopt the  $L_\infty$  contraction proposed by Tancik et al. [2023] to handle unbounded environmental lighting.

### 6.2 Integrating NeRF into Differentiable Rendering

Our inverse rendering pipeline is implemented by combining NeRF-Studio [Tancik et al. 2023] and Mitsuba 3 [Jakob et al. 2022b]. While NeRFStudio utilizes neural networks to represent a NeRF, Mitsuba 3 is a megakernel-based differentiable renderer that cannot incorporate neural networks within a megakernel. Therefore, integrating these two systems necessitates special techniques.

We encapsulate NeRF and its importance sampling approach into Mitsuba 3 as an emitter module. Importance sampling is executed within Mitsuba 3 using Dr.Jit [Jakob et al. 2022a]. During light source evaluation, the ray starting positions and directions are transferred to PyTorch for NeRF computation. The computed radiance or derivatives are then passed back to Mitsuba 3. Since only the emitter involves neural networks, we employ one-sample MIS [Veach 1997] to allow emitter queries for a batch of paths to take place simultaneously. We split the rendering operation into two megakernels, with the neural network operations occurring

between them. Specifically, the first kernel generates NeRF query rays, and the second kernel collects NeRF evaluations for pixel reconstruction.

Due to NeRF's large memory footprint, it supports only a limited ray batch size. However, typical differentiable rendering systems operate on an image, exceeding NeRF's maximum batch size capacity. Therefore, we divide the rays into batches to be sequentially processed by NeRF. To save memory, we do not store computation graphs, but instead recompute them during gradient computation. During forward rendering, we conduct detached gradient NeRF evaluation, and record random seeds to enable the replay [Vicini et al. 2021] of random operations. During back-propagation, we recover the computation graph for each batch of rays for automatic differentiation.

As the NeRF evaluation constitutes the runtime bottleneck, we have developed a multi-GPU training system to enhance its parallelism. The primary GPU hosts a Mitsuba 3 scene and performs surface rendering. For light evaluation and gradient computation, this GPU is tasked with distributing NeRF queries to the other GPUs, as well as collating computation results.

### 6.3 Shape and Material Optimization

Our differentiable surface rendering builds upon the differentiable SDF proposed by Vicini et al. [2022]. This implementation effectively handles visibility discontinuities, facilitating shape optimization. We represent the shape and material using voxel grids with a resolution of  $256^3$ , and conduct inverse rendering for 320 iterations. In each iteration, we randomly sample 6 images from the dataset for training. Following a coarse-to-fine optimization approach, we start optimization using  $128^2$  resolution images and  $64^3$  voxel grids, progressively scaling up images to  $512^2$  and voxel grids to  $256^3$  at iterations 128 and 256. We use a Laplacian loss [Vicini et al. 2022] and a curvature loss [Rosu and Behnke 2023] to maintain the smoothness of voxel grids and redistance the SDF after each iteration. Both shape and material voxel grids, along with the NeRF network, are optimized using the Adam optimizer [Kingma and Ba 2015], with a learning rate of 3e-3 for shape, 2e-2 for material and 1e-2 for NeRF. We employ 512 primal and 128 adjoint samples per pixel.

### 6.4 Data Capture Details

Our dataset features two distinctive characteristics: high dynamic range and multiple rotation poses. To obtain HDR captures, we use a Canon 5D Mark III camera with bracketed exposures. Each HDR image is synthesized from seven differently exposed photographs using HDRUtils [Hanji et al. 2020; Hanji and Mantiuk 2023]. This synthesis process ensures that HDR images accurately record the physical radiance. Employing a turntable, we capture four rotational poses for each object, rotating it by 90 degrees each time. Multi-view images were captured for each rotation, and their camera intrinsic and extrinsic parameters were determined using Metashape. To align the four rotational poses, we placed two ChArUco boards during the capture: one on the turntable and the other beside it. These boards aided in estimating the transformation of the bounding box with respect to the scene. Inspired by recent works that jointly optimize NeRF and camera transformations [Lin et al. 2021], we

incorporate joint optimization of rotation transformations during the NeRF pretraining stage to refine the estimated transformations. In NeRF, we assigned a unique appearance embedding to each rotation pose to facilitate modeling of environment lighting variations as the object rotates.

## 7 EXPERIMENTS

The datasets used are first introduced, followed by the comparison with the environment map emitter in the task of inverse rendering and the evaluation of our emitter importance sampling.

### 7.1 Datasets

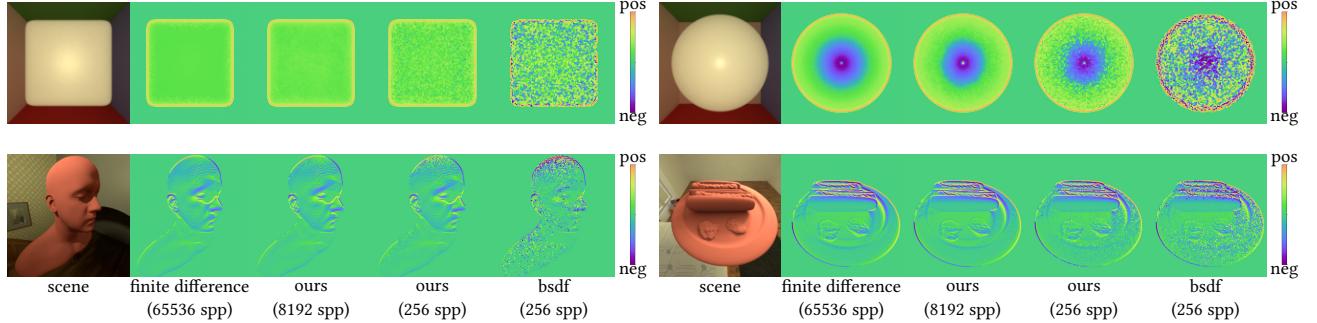
Following the aforementioned capture setup, we acquire four real-world data samples in an indoor environment. We place a lamp beside the object to simulate non-distant lighting conditions. Each sample consists of four rotation poses, with each rotation pose comprising 50-80 HDR multi-view images. We use a Konica Minolta Vivid 9i scanner to obtain the ground truth geometry. Following the real-world data capture setup, we synthesize four synthetic data samples using Mitsuba 3. For each set of synthetic data, we place an object in an indoor scene and render it from four different rotation poses. For each rotation pose, we synthesize 50 multi-view images.

### 7.2 Comparisons with Environment Map

We compare the proposed NeRF emitter with the commonly employed environment map emitter, prevalent in recent inverse rendering frameworks [Vicini et al. 2022; Boss et al. 2022; Jin et al. 2023; Sun et al. 2023]. Since the NeRF emitter is designed to be a universal component applicable across inverse rendering systems, in principle it could be integrated into these frameworks to achieve similar comparison results. For fair assessment, we choose the differentiable SDF [Vicini et al. 2022] framework for its physics-based principles, and build both our method and the baseline upon it, thereby highlighting the differences stemming from the emitter.

We adapt the environment map baseline method to suit our data. Firstly, we need an environment map input for the baseline. For synthetic data, we render environment maps using the ground truth scene in Mitsuba 3. For real data, we train a NeRF with captured HDR images and render environment maps from the resulting NeRF. In both cases, we place a virtual spherical camera at the center of the object and render environment maps. When rendering the environment map, we retain the object of interest in the scene, and set the spherical camera's near plane at the object's bounding box. This approach captures global illumination effects caused by the object while preventing occlusion of the spherical camera. We render an environment map for each rotation pose to capture the variations in global illumination caused by object rotations.

Secondly, the rendered background does not match the input images when using environment maps due to parallax effects. Therefore, for the synthetic dataset, we generate ground truth object masks and set the background pixels in the input images to black. During inverse rendering, the baseline renders only the object and hides the environment map, thus generating masked rendered images that match the background-processed input images. For the real dataset, we lend the background modeling ability of our NeRF



**Figure 4: Comparing the gradient images rendered by our emitter importance sampling and pure BSDF sampling.**

**Table 1: Quantitative comparison with the environment map baseline.**

Method	Novel-view Synthesis			Relighting			Shape CD $\downarrow$
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	
Ours	30.70	0.97	0.019	27.99	0.96	0.040	8.35e-6
Envmap	22.41	0.95	0.054	21.32	0.92	0.088	1.20e-4

emitter to the baseline. Specifically, we use NeRF to render a background image behind the bounding box and a semi-transparent image to model occlusions (if any) in front of the bounding box. These images are composited with the rendered object image to obtain the final image in the baseline.

Thirdly, the baseline employs a similar coarse-to-fine optimization procedure to ours. We use the same NeRF-extracted geometry as the shape initialization for the baseline. After these adaptations, the key difference between our method and the baseline remains in different emitter models.

Figure 5 presents the qualitative comparison results of the rendering and shape reconstruction between our method and the environment map baseline on the synthetic datasets. We visualize rendered images under the original lighting, reconstructed shape, and relighting results in two novel lighting conditions. For the Head and Hotdog scene, we optimize the diffuse reflectance parameter using a diffuse BSDF following the approach by Vicini et al. [2022]. In other scenes, we optimize the base color and roughness parameters using a principled BSDF [Burley 2012]. In the **Head** scene, shadows cast by the head onto the shoulder and chest are evident. While the baseline appears to match these cast shadows when rendering in the original lighting, the relighting results reveal local shape bumps. These stem from inaccuracies during inverse rendering, where shadows rendered by the environment map fail to align with input images at shadow borders. In the **Hotdog** scene, inaccurate shadows rendered by the environment map compromise the rendering results under original lighting. These artifacts persist during relighting and shape visualization. Given our method’s use of a more accurate light modeling approach using NeRF, rendered shadows better align with input images under original lighting, and our relighting and shape results do not exhibit such issues. In the **Teapot** scene, glossy highlights are apparent. However, due to environment map inaccuracies, baseline-rendered highlights do not match reference images under original lighting, resulting in inaccurate material parameters and geometry bumps, visible in

relighting results. In contrast, our method more closely replicates glossy highlights in both original lighting and relighting results. The **Boar** scene is captured in a Cornell box with a red wall on one side and a green wall on the other. The baseline appears to show illumination effects from the red wall on the boar’s hind leg in the original lighting. However, relighting results reveal baked illumination effects (green on the boar’s face, red on the boar’s hind leg). Our method accurately renders illumination effects under original lighting without baking reflections in relighting results.

In Table 1, we present the average quantitative results of our method and the environment map baseline on the synthetic dataset. We use ground truth object masks to set background pixels to black when evaluating novel-view synthesis and relighting. This approach provides a more precise representation of the object foreground’s quality, excluding the impact of the inaccurate background in the baseline. The PSNR metrics are exclusively computed within the object foreground mask. Notably, our method consistently outperforms the baseline across novel-view synthesis, relighting, and shape reconstruction tasks.

Figure 6 and 1 show the comparison results on real-world data. In the **Cabbage**, **Hamster** and **Dog** scenes, we observe attached shadows in the original lighting. It is evident in the relighting results that many attached shadows are baked into the baseline, causing dark artifacts. Our method, in contrast, produces more reasonable relighting results, indicating a more accurate material reconstruction. The **RealChair** scene exhibits strong self-shadows. We observe inaccurately cast shadows rendered by the baseline in the original lighting. These mismatched shadow borders are also present in the reconstructed materials during relighting. Our method accurately renders cast shadows in the original lighting and yields cleaner relighting results. We visualize the shape errors as heatmaps in the insets. From these visualized shape and error maps, we observe that our reconstructed shape is more accurate than the baseline. A part of the checkerboard is enclosed within the bounding box, thus modeled by surface rendering, while the rest is modeled by NeRF. During shape visualization, we manually crop the checkerboard part to visualize the object only.

### 7.3 Ablation Study

We compare our emitter importance sampling for NeRF with pure BSDF sampling in Figure 4.

In the top row, we show two examples containing, respectively, a rough cube and a rough sphere with a NeRF emitter trained

using multi-view renderings of the Cornell box. The camera is positioned in a top-down perspective view. The derivative images are computed with respect to the vertical displacement of the cube and the radius of the sphere.

The bottom row of Figure 4 show two examples under more complex non-distant indoor illuminations, both exhibiting self-shadowing effects. We compute derivative images with respect to vertical displacements of the objects.

In all four examples, converged derivative estimates (using 8192 spp) generated with our approach closely match finite-difference references (using 65536 spp). Using lower sample counts (256 spp), our importance sampling outperforms pure BSDF sampling significantly.

## 8 DISCUSSION AND CONCLUSION

*Discussion.* The main drawback of employing our proposed NeRF emitter is the increased computational cost in evaluating NeRF. With our multi-GPU implementation, our inverse rendering optimization requires approximately 4.5 hours on 8 RTX 4090 GPUs, whereas the baseline using an environment map completes in about one hour on a single RTX 4090. However, accelerating NeRF remains an active area of research, and our method stands to benefit from advancements in this direction. Recent works such as Adaptive Shells [Wang et al. 2023a] can potentially accelerate our approach by improving NeRF evaluation speed. Additionally, research progress aimed at reducing the samples per pixel in inverse rendering, exemplified by the works of Nicolet et al. [2023]; Wang et al. [2023b]; Chang et al. [2023], may further accelerate our approach by reducing the number of NeRF evaluation rays.

Our emitter sampling method for NeRF does not account for the directional dependency of the radiance function  $c$ . While this assumption performs well with diffuse emitters, it can result in increased rendering variance with highly directional-dependent emitters. It also overlooks occlusion introduced by the surface to NeRF, which should decrease the sampling probabilities of occluded emitters. However, as an initial step in the emitter importance sampling for NeRF emitters, our proposed method achieves more accurate light modeling and effective light sampling in inverse rendering. Future work could explore this direction further to reduce the rendering variance when using NeRF as an emitter, possibly employing more advanced guiding techniques [Müller et al. 2017].

*Conclusion.* Our work showcased the efficacy of integrating NeRF as a non-distant environment emitter within the physics-based inverse rendering pipeline. By addressing the shortcomings of the widely adopted distant environmental lighting model, we developed a novel approach to capture spatially varying illumination for accurate 3D reconstruction under real-world configurations. Furthermore, our introduction of an emitter importance sampling technique offered significant variance reduction for rendering with NeRF-based lighting. We demonstrated the effectiveness of our technique using both real and synthetic inverse-rendering experiments.

## ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2023YFC3305600, 2018YFA0704000), the NSFC (No.62021002),

and the Key Research and Development Project of Tibet Autonomous Region (XZ202101ZY0019G). This work was also supported by THUIBCS, Tsinghua University, and BLBCI, Beijing Municipal Education Commission. Feng Xu is the corresponding author.

## REFERENCES

- Sai Praveen Bangaru, Michaël Gharbi, Fujun Luan, Tzu-Mao Li, Kalyan Sunkavalli, Milos Hasan, Sai Bi, Zexiang Xu, Gilbert Bernstein, and Frédéric Durand. 2022. Differentiable Rendering of Neural SDFs through Reparameterization. In *SIGGRAPH Asia 2022 Conference Papers, SA 2022, Daegu, Republic of Korea, December 6-9, 2022*, Soon Ki Jung, Jehee Lee, and Adam W. Bargteil (Eds.). ACM, 5:1–5:9. <https://doi.org/10.1145/3550469.3555397>
- Sai Praveen Bangaru, Tzu-Mao Li, and Frédéric Durand. 2020. Unbiased warped-area sampling for differentiable rendering. *ACM Trans. Graph.* 39, 6 (2020), 245:1–245:18. <https://doi.org/10.1145/3414685.3417833>
- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022*. IEEE, 5460–5469. <https://doi.org/10.1109/CVPR52688.2022.00539>
- Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2023. Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1–6, 2023*. IEEE, 19640–19648. <https://doi.org/10.1109/ICCV51070.2023.01804>
- Sai Bi, Zexiang Xu, Pratul P. Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Milos Hasan, Yannick Hold-Geoffroy, David J. Kriegman, and Ravi Ramamoorthi. 2020a. Neural Reflectance Fields for Appearance Acquisition. *CoRR* abs/2008.03824 (2020). arXiv:2008.03824 <https://arxiv.org/abs/2008.03824>
- Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Milos Hasan, Yannick Hold-Geoffroy, David J. Kriegman, and Ravi Ramamoorthi. 2020b. Deep Reflectance Volumes: Relightable Reconstructions from Multi-view Photometric Images. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III (Lecture Notes in Computer Science, Vol. 12348)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer, 294–311. [https://doi.org/10.1007/978-3-030-58580-8\\_18](https://doi.org/10.1007/978-3-030-58580-8_18)
- Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P. A. Lensch. 2021a. NeRD: Neural Reflectance Decomposition from Image Collections. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10–17, 2021*. IEEE, 12664–12674. <https://doi.org/10.1109/ICCV48922.2021.01245>
- Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan T. Barron, Hendrik P. A. Lensch, and Varun Jampani. 2022. SAMURAI: Shape And Material from Unconstrained Real-world Arbitrary Image collections. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (Eds.). [http://papers.nips.cc/paper\\_files/paper/2022/hash/a8f2713b5c6bdcd3d264f1aa9b9c6f03-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/a8f2713b5c6bdcd3d264f1aa9b9c6f03-Abstract-Conference.html)
- Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan T. Barron, and Hendrik P. A. Lensch. 2021b. Neural-PIL: Neural Pre-Integrated Lighting for Reflectance Decomposition. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6–14, 2021, virtual*, Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (Eds.). <https://proceedings.neurips.cc/paper/2021/hash/58ae749f25ed636f486bc85feb3f0ab-Abstract.html>
- Brent Burley. 2012. Physically Based Shading at Disney. In *SIGGRAPH 2012 Course Notes*. <https://www.disneyanimation.com/publications/physically-based-shading-at-disney/>
- Wesley Chang, Venkataram Sivaram, Derek Nowrouzezahrai, Toshiya Hachisuka, Ravi Ramamoorthi, and Tzu-Mao Li. 2023. Parameter-space ReSTIR for Differentiable and Inverse Rendering. In *ACM SIGGRAPH 2023 Conference Proceedings, SIGGRAPH 2023, Los Angeles, CA, USA, August 6–10, 2023*, Erik Brunvand, Alla Sheffer, and Michael Wimmer (Eds.). ACM, 18:1–18:10. <https://doi.org/10.1145/3588432.3591512>
- Zhaoxi Chen and Ziwei Liu. 2022. Relighting4D: Neural Relightable Human from Videos. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIV (Lecture Notes in Computer Science, Vol. 13674)*, Shai Avidan, Gabriel J. Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer, 606–623. [https://doi.org/10.1007/978-3-031-19781-9\\_35](https://doi.org/10.1007/978-3-031-19781-9_35)
- Jorge Condor and Adrián Jarabo. 2022. A Learned Radiance-Field Representation for Complex Luminaires. In *33rd Eurographics Symposium on Rendering, EGSR 2022 - Symposium Track, Prague, Czech Republic, 4–6 July 2022*, Abhijeet Ghosh and Li-Yi Wei (Eds.). Eurographics Association, 49–58. <https://doi.org/10.2312/SR.20221155>
- Brian Curless and Marc Levoy. 1996. A Volumetric Method for Building Complex Models from Range Images. In *Proceedings of the 23rd Annual Conference on Computer*

- Graphics and Interactive Techniques, SIGGRAPH 1996, New Orleans, LA, USA, August 4-9, 1996*, John Fujii (Ed.). ACM, 303–312. <https://doi.org/10.1145/237170.237269>
- Paul E. Debevec and Jitendra Malik. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997, Los Angeles, CA, USA, August 3-8, 1997*, G. Scott Owen, Turner Whitted, and Barbara Mones-Hattal (Eds.). ACM, 369–378. <https://doi.org/10.1145/258734.258884>
- Wenhang Ge, Tao Hu, Haoyu Zhao, Shu Liu, and Ying-Cong Chen. 2023. Ref-Neus: Ambiguity-Reduced Neural Implicit Surface Learning for Multi-View Reconstruction with Reflection. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*. IEEE, 4228–4237. <https://doi.org/10.1109/ICCV51070.2023.00392>
- Param Hanji and Rafal K. Mantiuk. 2023. Robust Estimation of Exposure Ratios in Multi-Exposure Image Stacks. *IEEE Trans. Computational Imaging* 9 (2023), 721–731. <https://doi.org/10.1109/TCI.2023.3301338>
- Param Hanji, Fangcheng Zhong, and Rafal K. Mantiuk. 2020. Noise-Aware Merging of High Dynamic Range Image Stacks Without Camera Calibration. In *Computer Vision - ECCV 2020 Workshops - Glasgow, UK, August 23-28, 2020, Proceedings, Part III (Lecture Notes in Computer Science, Vol. 12537)*, Adrià Bartoli and Andrea Fusello (Eds.). Springer, 376–391. [https://doi.org/10.1007/978-3-030-67070-2\\_23](https://doi.org/10.1007/978-3-030-67070-2_23)
- Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. 2022. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, Sammi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (Eds.). [http://papers.nips.cc/paper\\_files/paper/2022/hash/8fc27984bf16ca03cad043244ec470d-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/8fc27984bf16ca03cad043244ec470d-Abstract-Conference.html)
- Wenzel Jakob. 2012. Numerically stable sampling of the von Mises-Fisher distribution on  $S^2$  (and other tricks). *Interactive Geometry Lab, ETH Zürich, Tech. Rep* (2012), 6.
- Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. 2022b. *Mitsuba 3 renderer*. <https://mitsuba-renderer.org>.
- Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022a. DRJIT: a just-in-time compiler for differentiable rendering. *ACM Trans. Graph.* 41, 4 (2022), 124:1–124:19. <https://doi.org/10.1145/3528223.3530099>
- Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. 2023. TensorIR: Tensorial Inverse Rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 165–174. <https://doi.org/10.1109/CVPR52729.2023.00024>
- Ondrej Karlik, Martin Sik, Petr Vévoda, Tomáš Skrivan, and Jaroslav Krivánek. 2019. MIS compensation: optimizing sampling techniques in multiple importance sampling. *ACM Trans. Graph.* 38, 6 (2019), 151:1–151:12. <https://doi.org/10.1145/3355089.3356565>
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- Tzu-Mao Li, Miika Aittala, Frédéric Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo ray tracing through edge sampling. *ACM Trans. Graph.* 37, 6 (2018), 222. <https://doi.org/10.1145/3272127.3275109>
- Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. 2021. BARF: Bundle-Adjusting Neural Radiance Fields. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*. IEEE, 5721–5731. <https://doi.org/10.1109/ICCV48922.2021.000569>
- Jingwang Ling, Zhibo Wang, and Feng Xu. 2023. ShadowNeuS: Neural SDF Reconstruction by Shadow Ray Supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 175–185. <https://doi.org/10.1109/CVPR52729.2023.00025>
- Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. 2022. NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*. IEEE, 16169–16178. <https://doi.org/10.1109/CVPR52688.2022.01571>
- 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 *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12346)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer, 405–421. [https://doi.org/10.1007/978-3-030-58452-8\\_24](https://doi.org/10.1007/978-3-030-58452-8_24)
- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* 41, 4 (2022), 102:1–102:15. <https://doi.org/10.1145/3528223.3530127>
- Thomas Müller, Markus H. Gross, and Jan Novák. 2017. Practical Path Guiding for Efficient Light-Transport Simulation. *Comput. Graph. Forum* 36, 4 (2017), 91–100. <https://doi.org/10.1111/cgf.13227>
- Baptiste Nicolet, Fabrice Rousselle, Jan Novák, Alexander Keller, Wenzel Jakob, and Thomas Müller. 2023. Recursive Control Variates for Inverse Rendering. *ACM Trans. Graph.* 42, 4 (2023), 62:1–62:13. <https://doi.org/10.1145/3592139>
- Merlin Nimier-David, Sébastien Speierer, Benoît Ruiz, and Wenzel Jakob. 2020. Radiative backpropagation: an adjoint method for lightning-fast differentiable rendering. *ACM Trans. Graph.* 39, 4 (2020), 146. <https://doi.org/10.1145/3386569.3392406>
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (Eds.), 8024–8035. <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f9f2fbaf9701272740-Abstract.html>
- Julien Philip and Valentin Deschaintre. 2023. Radiance Field Gradient Scaling for Unbiased Near-Camera Training. *CoRR* abs/2305.02756 (2023). <https://doi.org/10.48550/ARXIV.2305.02756>
- Yi-Ling Qiao, Alexander Gao, Yiran Xu, Yue Feng, Jia-Bin Huang, and Ming C. Lin. 2023. Dynamic Mesh-Aware Radiance Fields. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*. IEEE, 385–396. <https://doi.org/10.1109/ICCV51070.2023.00042>
- Radu Alexandru Rosu and Sven Behnke. 2023. PermutoSDF: Fast Multi-View Reconstruction with Implicit Surfaces Using Permutohedral Lattices. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 8466–8475. <https://doi.org/10.1109/CVPR52729.2023.00818>
- Viktor Rudnev, Mohamed Elgharib, William A. P. Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. 2021. Neural Radiance Fields for Outdoor Scene Relighting. *CoRR* abs/2112.05140 (2021). arXiv:2112.05140 <https://arxiv.org/abs/2112.05140>
- 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 *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*. Computer Vision Foundation / IEEE, 7495–7504. <https://doi.org/10.1109/CVPR46437.2021.00741>
- Cheng Sun, Guangyan Cai, Zhengqin Li, Kai Yan, Cheng Zhang, Carl Marshall, Jia-Bin Huang, Shuang Zhao, and Zhao Dong. 2023. Neural-PBIR Reconstruction of Shape, Material, and Illumination. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*. IEEE, 18000–18010. <https://doi.org/10.1109/ICCV51070.2023.01654>
- Matthew Tancik, Ethan Weber, Evonne Ng, Rui long Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David McAllister, Justin Kerr, and Angjoo Kanazawa. 2023. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In *ACM SIGGRAPH 2023 Conference Proceedings, SIGGRAPH 2023, Los Angeles, CA, USA, August 6-10, 2023*, Erik Brunvand, Alla Sheffer, and Michael Wimmer (Eds.). ACM, 72:1–72:12. <https://doi.org/10.1145/3588432.3591516>
- Eric Veach. 1997. *Robust Monte Carlo methods for light transport simulation*. Ph.D. Dissertation. Stanford University, USA. <https://searchworks.stanford.edu/view/3911108>
- Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2021. Path replay backpropagation: differentiating light paths using constant memory and linear time. *ACM Trans. Graph.* 40, 4 (2021), 108:1–108:14. <https://doi.org/10.1145/3450626.3459804>
- Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2022. Differentiable signed distance function rendering. *ACM Trans. Graph.* 41, 4 (2022), 125:1–125:18. <https://doi.org/10.1145/3528223.3530139>
- 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. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (Eds.), 27171–27183. <https://proceedings.neurips.cc/paper/2021/hash/e41e164f7485ec4a28741a2d0ea41c74-Abstract.html>
- Yu-Chen Wang, Chris Wyman, Lifen Wu, and Shuang Zhao. 2023b. Amortizing Samples in Physics-Based Inverse Rendering Using ReStIR. *ACM Trans. Graph.* 42, 6 (2023), 214:1–214:17. <https://doi.org/10.1145/3618331>
- Zian Wang, Tianchang Shen, Merlin Nimier-David, Nicholas Sharp, Jun Gao, Alexander Keller, Sanja Fidler, Thomas Müller, and Zan Gojcic. 2023a. Adaptive Shells for Efficient Neural Radiance Field Rendering. *ACM Trans. Graph.* 42, 6 (2023), 260:1–260:15. <https://doi.org/10.1145/3618390>
- Haoqian Wu, Zhipeng Hu, Lincheng Li, Yongqiang Zhang, Changjie Fan, and Xin Yu. 2023. NeFI: Inverse Rendering for Reflectance Decomposition with Near-Field Indirect Illumination. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 4295–4304. <https://doi.org/10.1109/CVPR52729.2023.00418>
- Linning Xu, Vasu Agrawal, William Laney, Tony Garcia, Aayush Bansal, Changil Kim, Samuel Rota Bulò, Lorenzo Porzi, Peter Kortschied, Aljaz Bozic, Dahu Lin, Michael Zollhöfer, and Christian Richardt. 2023a. VR-NeRF: High-Fidelity

- Virtualized Walkable Spaces. In *SIGGRAPH Asia 2023 Conference Papers, SA 2023, Sydney, NSW, Australia, December 12–15, 2023*, June Kim, Ming C. Lin, and Bernd Bickel (Eds.). ACM, 43:1–43:12. <https://doi.org/10.1145/3610548.3618139>
- Peiyu Xu, Sai Praveen Bangaru, Tzu-Mao Li, and Shuang Zhao. 2023b. Warped-Area Reparameterization of Differential Path Integrals. *ACM Trans. Graph.* 42, 6 (2023), 213:1–213:18. <https://doi.org/10.1145/3618330>
- Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. 2022a. PS-NeRF: Neural Inverse Rendering for Multi-view Photometric Stereo. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 13661)*, Shai Aviran, Gabriel J. Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer, 266–284. [https://doi.org/10.1007/978-3-031-19769-7\\_16](https://doi.org/10.1007/978-3-031-19769-7_16)
- Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. 2022b. S<sup>3</sup>-NeRF: Neural Reflectance Field from Shading and Shadow under a Single Viewpoint. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 – December 9, 2022*, Sammi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (Eds.). [http://papers.nips.cc/paper\\_files/paper/2022/hash/0a630402ee92620dc2de3b704181de9b-Abstract-Conference.html](http://papers.nips.cc/paper_files/paper/2022/hash/0a630402ee92620dc2de3b704181de9b-Abstract-Conference.html)
- Lior Yariv, Peter Hedman, Christian Reiser, Dor Verbin, Pratul P. Srinivasan, Richard Szeliski, Jonathan T. Barron, and Ben Mildenhall. 2023. BakedSDF: Meshing Neural SDFs for Real-Time View Synthesis. In *ACM SIGGRAPH 2023 Conference Proceedings, SIGGRAPH 2023, Los Angeles, CA, USA, August 6–10, 2023*, Erik Brunvand, Alla Sheffer, and Michael Wimmer (Eds.). ACM, 46:1–46:9. <https://doi.org/10.1145/3588432.3591536>
- Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob. 2021. Monte Carlo estimators for differential light transport. *ACM Trans. Graph.* 40, 4 (2021), 78:1–78:16. <https://doi.org/10.1145/3450626.3459807>
- Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020. Path-space differentiable rendering. *ACM Trans. Graph.* 39, 4 (2020), 143. <https://doi.org/10.1145/3386569.3392383>
- Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. 2019. A differential theory of radiative transfer. *ACM Trans. Graph.* 38, 6 (2019), 227:1–227:16. <https://doi.org/10.1145/3355089.3356522>
- Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. 2022a. IRON: Inverse Rendering by Optimizing Neural SDFs and Materials from Photometric Images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022*. IEEE, 5555–5564. <https://doi.org/10.1109/CVPR52688.2022.00548>
- 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 *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19–25, 2021*. Computer Vision Foundation / IEEE, 5453–5462. <https://doi.org/10.1109/CVPR46437.2021.00541>
- Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. 2022b. Modeling Indirect Illumination for Inverse Rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022*. IEEE, 18622–18631. <https://doi.org/10.1109/CVPR52688.2022.01809>
- Junqiu Zhu, Yaoyi Bai, Zilin Xu, Steve Bakó, Edgar Velázquez-Armendáriz, Lu Wang, Pradeep Sen, Milos Hasan, and Ling-Qi Yan. 2021. Neural complex luminaires: representation and rendering. *ACM Trans. Graph.* 40, 4 (2021), 57:1–57:12. <https://doi.org/10.1145/3450626.3459798>

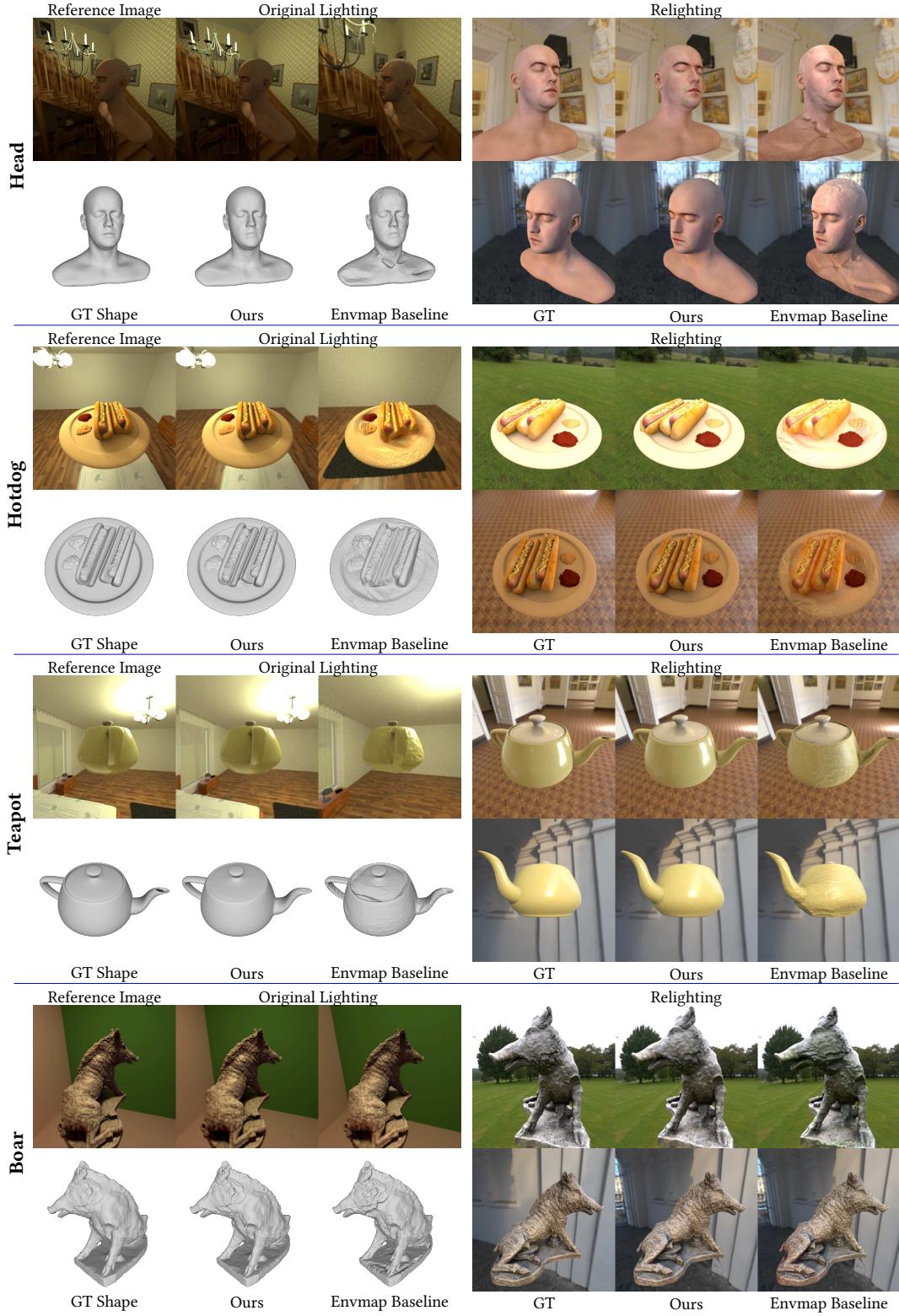
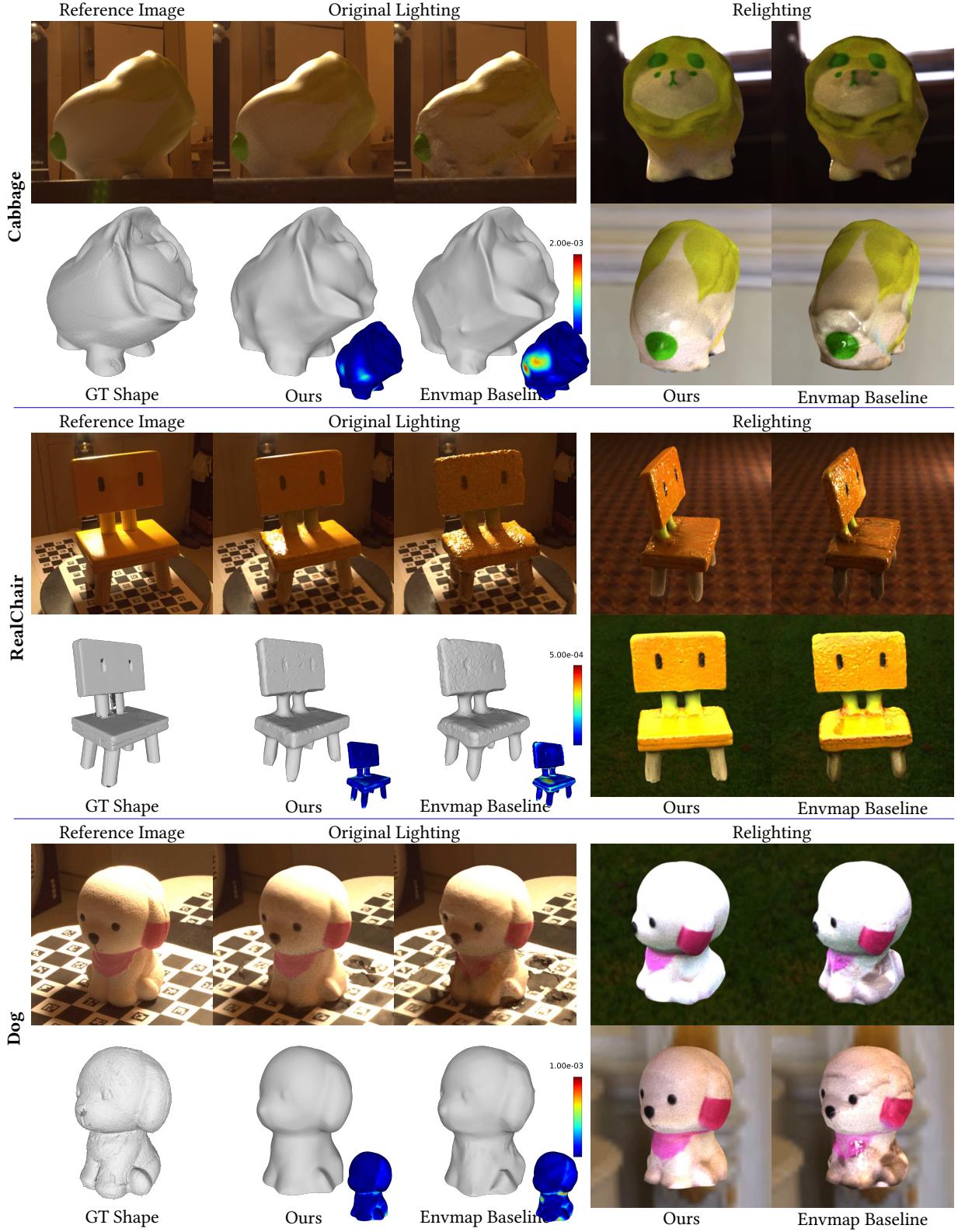


Figure 5: Comparison with the environment map baseline on synthetic datasets.

**Figure 6: Comparison with the environment map baseline on real datasets.**