

Relighting Neural Radiance Fields with Shadow and Highlight Hints

Chong Zeng*

State Key Lab of CAD and CG,
Zhejiang University
Hangzhou, China
chongzeng2000@gmail.com

Guojun Chen

Microsoft Research Asia
Beijing, China
guoch@microsoft.com

Yue Dong

Microsoft Research Asia
Beijing, China
yuedong@microsoft.com

Pieter Peers

College of William & Mary
Williamsburg, USA
ppeers@siggraph.org

Hongzhi Wu

State Key Lab of CAD and CG,
Zhejiang University
Hangzhou, China
hwu@acm.org

Xin Tong

Microsoft Research Asia
Beijing, China
xtong@microsoft.com

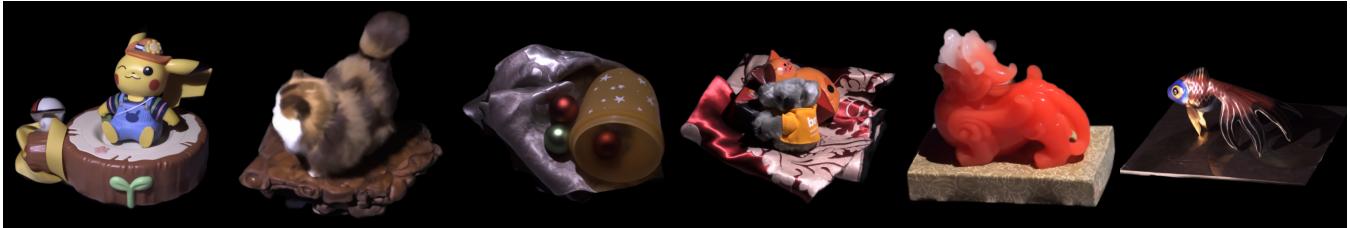


Figure 1: Free viewpoint relighting of neural radiance fields trained on 500–1,000 unstructured photographs per scene captured with a handheld setup.

ABSTRACT

This paper presents a novel neural implicit radiance representation for free viewpoint relighting from a small set of unstructured photographs of an object lit by a moving point light source different from the view position. We express the shape as a signed distance function modeled by a multi layer perceptron. In contrast to prior relightable implicit neural representations, we do not disentangle the different light transport components, but model both the local and global light transport at each point by a second multi layer perceptron that, in addition to density features, the current position, the normal (from the signed distance function), view direction, and light position, also takes shadow and highlight hints to aid the network in modeling the corresponding high frequency light transport effects. These hints are provided as a suggestion, and we leave it up to the network to decide how to incorporate these in the final relit result. We demonstrate and validate our neural implicit representation on synthetic and real scenes exhibiting a wide variety of shapes, material properties, and global illumination light transport.

CCS CONCEPTS

- Computing methodologies → Image-based rendering; Reflectance modeling.

KEYWORDS

Relighting, Free-viewpoint, Neural Implicit Modeling

ACM Reference Format:

Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. 2023. Relighting Neural Radiance Fields with Shadow and Highlight Hints. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings (SIGGRAPH '23 Conference Proceedings)*, August 6–10, 2023, Los Angeles, CA, USA. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3588432.3591482>

1 INTRODUCTION

The appearance of real-world objects is the result of complex light transport interactions between the lighting and the object's intricate geometry and associated material properties. Digitally reproducing the appearance of real-world objects and scenes has been a long-standing goal in computer graphics and computer vision. Inverse rendering methods attempt to undo the complex light transport to determine a sparse set of model parameters that, together with the chosen models, replicates the appearance when rendered. However, teasing apart the different entangled components is ill-posed and often leads to ambiguities. Furthermore, inaccuracies in one model can adversely affect the accuracy at which other components can be disentangled, thus requiring strong regularization and assumptions.

In this paper we present a novel, NeRF-inspired [Mildenhall et al. 2020], neural implicit radiance representation for free viewpoint relighting of general objects and scenes. Instead of using analytical reflectance models and inverse rendering of the neural implicit representations, we follow a data-driven approach and refrain from decomposing the appearance in different light transport components. Therefore, unlike the majority of prior work in relighting neural implicit representations [Boss et al. 2021a, 2022; Kuang et al. 2022;

*Work done during internship at Microsoft Research Asia.

Srinivasan et al. 2021; Zheng et al. 2021], we relax and enrich the lighting information embedded in handheld captured photographs of the object by illuminating each view from a random point light position. This provides us with a broader unstructured sampling of the space of appearance changes of an object, while retaining the convenience of handheld acquisition. Furthermore, to improve the reproduction quality of difficult to learn components, we provide shadow and highlight *hints* to the neural radiance representation. Critically, we do not impose how these hints are combined with the estimated radiance (*e.g.*, shadow mapping by multiplying with the light visibility), but instead leave it up to the neural representation to decide how to incorporate these hints in the final result.

Our hint-driven implicit neural representation is easy to implement, and it requires an order of magnitude less photographs than prior relighting methods that have similar capabilities, and an equal number of photographs compared to state-of-the-art methods that offer less flexibility in the shape and/or materials that can be modeled. Compared to fixed lighting implicit representations such as NeRF [Mildenhall et al. 2020], we only require a factor of five times more photographs and twice the render cost while gaining relightability. We demonstrate the effectiveness and validate the robustness of our representation on a variety of challenging synthetic and real objects (*e.g.*, Figure 1) containing a wide range of materials (*e.g.*, subsurface scattering, rough specular materials, etc.) variations in shape complexity (*e.g.*, thin features, ill-defined furry shapes, etc.) and global light transport effects (*e.g.*, interreflections, complex shadowing, etc.).

2 RELATED WORK

We focus the discussion of related work on seminal and recent work in image-based relighting, inverse rendering, and relighting neural implicit representations. For an in-depth overview we refer to recent surveys in neural rendering [Tewari et al. 2022], (re)lighting [Einabadi et al. 2021], and appearance modeling [Dong 2019].

Image-based Relighting. The staggering advances in machine learning in the last decade have also had a profound effect on image-based relighting [Debevec et al. 2000], enabling new capabilities and improving quality [Bemana et al. 2020; Ren et al. 2015; Xu et al. 2018]. Deep learning has subsequently been applied to more specialized relighting tasks for portraits [Bi et al. 2021; Meka et al. 2019; Pandey et al. 2021; Sun et al. 2019, 2020], full bodies [Guo et al. 2019; Kanamori and Endo 2018; Meka et al. 2020; Yeh et al. 2022; Zhang et al. 2021a], and outdoor scenes [Griffiths et al. 2022; Meshry et al. 2019; Philip et al. 2019]. It is unclear how to extend these methods to handle scenes that contain objects with ill-defined shapes (*e.g.*, fur) and translucent and specular materials.

Our method can also be seen as a free-viewpoint relighting method that leverages highlight and shadow hints to help model these challenging effects. Philip et al. [2019] follow a deep shading approach [Nalbach et al. 2017] for relighting, mostly diffuse, outdoor scenes under a simplified sun+cloud lighting model. Relit images are created in a two stage process, where an input and output shadow map computed from a proxy geometry is refined, and subsequently used, together with additional render buffers, as input to a relighting network. Zhang et al. [2021a] introduce

a semi-parametric model with residual learning that leverages a diffuse parametric model (*i.e.*, radiance hint) on a rough geometry, and a learned representation that models non-diffuse and global light transport embedded in texture space. To accurately model the non-diffuse effects, Zhang et al. require a large number (~ 8,000) of structured photographs captured with a light stage. Deferred Neural Relighting [Gao et al. 2020] is closest to our method in terms of capabilities; it can perform free-viewpoint relighting on objects with ill-defined shape with full global illumination effects and complex light-matter interactions (including subsurface scattering and fur). Similar to Zhang et al. [2021a], Gao et al. embed learned features in the texture space of a rough geometry that are projected to the target view and multiplied with *radiance cues*. These radiance cues are visualizations of the rough geometry with different BRDFs (*i.e.*, diffuse and glossy BRDFs with 4 different roughnesses) under the target lighting with global illumination. The resulting images are then used as guidance hints for a neural renderer trained per scene from a large number (~ 10,000) of unstructured photographs of the target scene for random point light-viewpoint combinations to reproduce the reference appearance. Philip et al. [2021] also use radiance hints (limited to diffuse and mirror radiance) to guide a neural renderer. However, unlike Zhang et al. and Gao et al., they pretrain a neural renderer that does not require per-scene fine-tuning, and that takes radiance cues for both the input and output conditions. Philip et al. require about the same number as input images as our method, albeit lit by a single fixed natural lighting conditions and limited to scenes with hard surfaces and BRDF-like materials. All four methods rely on multi-view stereo which can fail for complex scenes. In contrast our method employs a robust neural implicit representation. Furthermore, all four methods rely on an image-space neural renderer to produce the final relit image. In contrast, our method provides the hints during volume rendering of the neural implicit representation, and thus it is independent of view-dependent image contexts. Our method can relight scenes with the same complexity as Gao et al. [2020] while only using a similar number of input photographs as Philip et al. [2021] without sacrificing robustness.

Model-based Inverse Rendering. An alternative to data-driven relighting is inverse rendering (a.k.a. analysis-by-synthesis) where a set of trial model parameters are optimized based on the difference between the rendered model parameters and reference photographs. Inverse rendering at its core is a complex non-linear optimization problem. Recent advances in differentiable rendering [Li et al. 2018; Loper and Black 2014; Nimier-David et al. 2019; Xing et al. 2022] have enabled more robust inverse rendering for more complex scenes and capture conditions. BID-R++ [Chen et al. 2021] combines differentiable ray tracing and rasterization to model spatially varying reflectance parameters and spherical Gaussian lighting for a known triangle mesh. Munkberg et al. [2022] alternate between optimizing an implicit shape representation (*i.e.*, a signed distance field), and reflectance and lighting defined on a triangle mesh. Hasselgren et al. [2022] extend the work of Munkberg et al. [2022] with a differentiable Monte Carlo renderer to handle area light sources, and embed a denoiser to mitigate the adverse effects of Monte Carlo noise on the gradient computation to drive the non-linear optimizer. Similarly, Fujun et al. [2021] also employ a differentiable

Monte Carlo renderer for estimating shape and spatially-varying reflectance from a small set of colocated view/light photographs. All of these methods focus on direct lighting only and can produce suboptimal results for objects or scenes with strong interreflections. A notable exception is the method of Cai *et al.* [2022] that combines explicit and implicit geometries and demonstrates inverse rendering under known lighting on a wide range of opaque objects while taking indirect lighting in account. All of the above methods eventually express the shape as a triangle mesh, limiting their applicability to objects with well defined surfaces. Furthermore, the accuracy of these methods is inherently limited by the representational power of the underlying BRDF and lighting models.

Neural Implicit Representations. A major challenge in inverse rendering with triangle meshes is to efficiently deal with changes in topology during optimization. An alternative to triangle mesh representations is to use a volumetric representation where each voxel contains an opacity/density estimate and a description of the reflectance properties. While agnostic to topology changes, voxel grids are memory intensive and, even with grid warping [Bi *et al.* 2020], fine-scale geometrical details are difficult to model.

To avoid the inherent memory overhead of voxel grids, NeRF [Mildenhall *et al.* 2020] models the continuous volumetric density and spatially varying color with two multi layer perceptrons (MLPs) parameterized by position (and also view direction for color). The MLPs in NeRF are trained per scene such that the accumulated density and color ray marched along a view ray matches the observed radiance in reference photographs. NeRF has been shown to be exceptionally effective in modeling the outgoing radiance field of a wide range of object types, including those with ill-defined shapes and complex materials. One of the main limitations of NeRF is that the illumination present at capture-time is baked into the model. Several methods have been introduced to support post-capture relighting under a restricted lighting model [Li *et al.* 2022; Martin-Brualla *et al.* 2021], or by altering the color MLP to produce the parameters to drive an analytical model of the appearance of objects [Boss *et al.* 2021a, 2022, 2021b; Kuang *et al.* 2022; Srinivasan *et al.* 2021; Yao *et al.* 2022; Zhang *et al.* 2021c], participating media [Zheng *et al.* 2021], or even whole outdoor scenes [Rudnev *et al.* 2022].

Due to the high computational cost of ray marching secondary rays, naïvely computing shadows and indirect lighting is impractical. Zhang *et al.* [2021c], Li *et al.* [2022], and Yang *et al.* [2022] avoid tracing shadow rays by learning an additional MLP to model the ratio of light occlusion. However, all three methods ignore indirect lighting. Zheng *et al.* [2021] model the indirect lighting inside a participating media using an MLP that returns the coefficients of a 5-band expansion. NeILF [Yao *et al.* 2022] embeds the indirect lighting and shadows in a (learned) 5D incident light field for a scene with *known* geometry. NeRV [Srinivasan *et al.* 2021] modifies the color MLP to output BRDF parameters and a visibility field that models the distance to the nearest ‘hard surface’ and lighting visibility. The visibility field allows them to bypass the expensive ray marching step for shadow computation and *one-bounce* indirect illumination. A disadvantage of these solutions is that they do not guarantee that the estimated density field and the occlusions are coupled. In contrast, our method directly ties occlusions to the estimated implicit

geometry reproducing more faithful shadows. Furthermore, these methods rely on BRDFs to model the surface reflectance, precluding scenes with complex light-matter interactions.

NeLF [Sun *et al.* 2021] aims to relight human faces, and thus accurately reproducing subsurface scattering is critical. Therefore, Sun *et al.* characterize the radiance and global light transport by an MLP. We also leverage an MLP to model local and global light transport. A key difference is that our method parameterizes this MLP in terms of view and light directions, whereas NeLF directly outputs a full light transport vector and compute a relit color via an inner-product with the lighting. While better suited for relighting with natural lighting, NeLF is designed for relighting human faces which only exhibit limited variations in shape and reflectance.

Similar in spirit to our method, Lyu *et al.* [2022] model light transport using an MLP, named a Neural Radiance Transfer Field (NRTF). However, unlike us, Lyu *et al.* train the MLP on synthetic training data generated from a rough BRDF approximation obtained through physically based inverse rendering on a triangle mesh extracted from a neural signed distance field [Wang *et al.* 2021] computed from unstructured observations of the scene under static natural lighting. To correct the errors due the rough BRDF approximation, a final refinement step of the MLP is performed using the captured photographs. Similar to Lyu *et al.* we also use an MLP to model light transport, including indirect lighting. However, unlike Lyu *et al.* we do not rely solely on an MLP to model high frequency light transport effects such as light occlusions and specular highlights. Instead we provide shadow and highlight hints to the radiance network and let the training process discover how to best leverage these hints. Furthermore, we rely on a neural representation for shape jointly optimized with the radiance, allowing us to capture scenes with ill-defined geometry. In contrast, Lyu *et al.* optimize shape (converted to a triangle mesh) and radiance separately, making their method sensitive to shape errors and restricted to objects with a well-defined shape.

An alternative to using an implicit neural density field, is to model the shape via a signed distance field (SDF). Similar to the majority of NeRF-based methods, PhySG [Zhang *et al.* 2021b] and IRON [Zhang *et al.* 2022a] also rely on an MLP to represent volumetric BRDF parameters. However, due to the high computational cost, these methods do not take shadowing or indirect lighting in account. Zhang *et al.* [2022b] model indirect lighting separately, and train an additional incident light field MLP using the incident lighting computed at each point via ray casting the SDF geometry. While our method also builds on a neural implicit representation [Wang *et al.* 2021], our method does not rely on an underlying parametric BRDF model, but instead models the full light transport via an MLP. Furthermore, we do not rely on an MLP decoupled from the estimated geometry to estimate shadowing, but instead accumulate light occlusion along a single shadow ray per view ray, ensuring consistency between the shadows and the estimated geometry.

3 METHOD

Our goal is to extend neural implicit representations such as NeRF [Mildenhall *et al.* 2020] to model variations in lighting. NeRF has proven to be exceptionally efficient for viewpoint interpolation. In contrast to ray tracing with solid surfaces, NeRF relies on ray

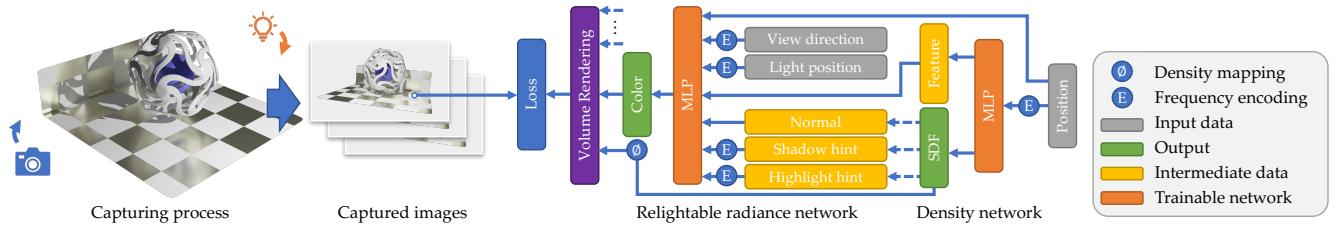


Figure 2: Overview: our neural implicit radiance representation is trained on unstructured photographs of the scene captured from different viewpoints and lit from different point light positions. The neural implicit radiance representation consists of two multi layer perceptron (MLP) networks for modeling the density field and for modeling the light transport. The MLP for modeling the density takes as input the position, and outputs the signed distance function of the shape and a feature vector that together with the current position, the normal extracted from the SDF, the view direction, the light source position, and the light transport hints, are passed into the radiance MLP that then computes the view and lighting dependent radiance.

marching through the volume, requiring at least an order of magnitude more computations. Not only does this ray marching cost affect rendering, it also leads to a prohibitively large training cost when secondary rays (e.g., shadows and indirect lighting) are considered. Instead of building our method on NeRF, we opt for using NeuS [Wang et al. 2021], a neural implicit signed distance field representation, as the basis for our method. Although NeuS does not speed up ray marching, it provides an unbiased depth estimate which we will leverage in subsection 3.2 for reducing the number of shadow rays.

Following prior work, our neural implicit radiance representation relies on two multi layer perceptrons (MLPs) for modeling the density field (following NeuS) and for modeling the (direct and indirect) radiance based on the current position, the normal derived from the density field, the view direction, the point light position, and the features provided by the density network. In addition, we also provide light transport *hints* to the relightable radiance MLP to improve the reproduction quality of difficult to model effects such as shadows and highlights. Figure 2 summarizes our architecture.

To train our neural implicit relightable radiance representation, we require observations of the target scene seen from different viewpoints and lit from different point light positions. It is essential that these observations include occlusions and interreflections. Colocated lighting (e.g., as in [Luan et al. 2021; Nam et al. 2018]) does not exhibit visible shadows and is therefore not suited. Instead we follow the acquisition process of Deferred Neural Lighting [Gao et al. 2020] and capture the scene from different viewpoints with a handheld camera while lighting the scene with a flash light of a second camera from a different direction.

We opt for parameterizing the radiance function with respect to a point light as the basis for relighting as this better reflects the physical capture process. A common approximation in prior relighting work that relies on active illumination (e.g., Light Stage) is to ignore the divergence of incident lighting due to the finite light source distance, and parameterize the reflectance field in terms lighting directions only. Similarly, we can also *approximate* distant lighting with point lighting defined by projecting the light direction onto a large sphere with a radius equal to the capture distance.

3.1 Representation

Density Network. Our neural implicit geometry representation follows NeuS [Wang et al. 2021] which uses an MLP to encode a Signed Distance Function (SDF) $f(\mathbf{p})$ from which the density function is derived using a probability density function $\phi_s(f(\mathbf{p}))$. This probability density function is designed to ensure that for opaque objects the zero-level set of the SDF corresponds to the surface. The width of the probability distribution models the uncertainty of the surface location. We follow exactly the same architecture for the density MLP as in NeuS: 8 hidden layers with 256 nodes using a Softplus activation and a skip connection between the input and the 4th layer. The input (*i.e.*, current position along a ray) is augmented using a frequency encoding with 6 bands. In addition, we also concatenate the original input signal to the encoding. The resulting output from the density network is the SDF at \mathbf{p} as well as a latent vector that encodes position dependent features.

Relightable Radiance Network. Analogous to the color MLP in NeRF and NeuS that at each volumetric position evaluates the view-dependent color, we introduce a *relightable radiance* MLP that at each volumetric position evaluates the view and lighting dependent (direct and indirect) light transport. We follow a similar architecture as NeRF/NeuS' color MLP and extend it by taking the position dependent feature vector produced by the density MLP, the normal derived from the SDF, the current position, the view direction, and the point light position as input. Given this input, the radiance MLP outputs the resulting radiance which includes all light transport effects such as occlusions and interreflections. We assume a white light source color; colored lighting can be achieved by scaling the radiance with the light source color (*i.e.*, linearity of light transport).

Given the output from the density network f as well as the output from the radiance network s , the color C along a view ray starting at the camera position \mathbf{o} in a direction \mathbf{v} is given by:

$$C(\mathbf{o}, \mathbf{v}) = \int_0^\infty w(t)s(\mathbf{p}, \mathbf{n}, \mathbf{v}, \mathbf{l}, f, \Theta) dt, \quad (1)$$

where the sample position along the view ray is $\mathbf{p} = \mathbf{o} + t\mathbf{v}$ at depth t , \mathbf{n} is the normal computed as the normalized SDF gradient:

$$\mathbf{n} = \nabla f(\mathbf{p}) / \| \nabla f(\mathbf{p}) \|, \quad (2)$$

\mathbf{v} is the view direction, \mathbf{l} is the point light position, f the corresponding feature vector from the density MLP, and Θ is a set of additional hints provided to the radiance network (described in subsection 3.2). Analogous to NeuS, the view direction, light position, and hints are all frequency encoded with 4 bands. Finally, $w(t)$ is the unbiased density weight [Wang et al. 2021] computed by:

$$w(t) = T(t)\rho(t), \quad (3)$$

$$T(t) = \exp\left(-\int_0^t \rho(u) du\right), \quad (4)$$

$$\rho(t) = \max\left(\frac{\frac{d\Phi_s}{dt}(f(t))}{\Phi_s(f(t))}, 0\right), \quad (5)$$

with T the transmittance over opacity ρ , Φ_s the CDF of the PDF ϕ_s used to compute the density from the SDF f . To speed up the computation of the color, the integral in Equation 1 is computed by importance sampling the density field along the view ray.

In the spirit of image-based relighting, we opt to have the relightable radiance MLP network include global light transport effects such as interreflections and occlusions. While MLPs are in theory universal approximators, some light transport components are easier to learn (e.g., diffuse reflections) than others. Especially high frequency light transport components such as shadows and specular highlights pose a problem. At the same time, shadows and specular highlights are highly correlated with the geometry of the scene and thus the density field. To leverage this embedded knowledge, we provide the relightable radiance MLP with additional *shadow* and *highlight hints*.

3.2 Light Transport Hints

Shadow Hints. While the relightable radiance network is able to roughly model the effects of light source occlusion, the resulting shadows typically lack sharpness and detail. Yet, light source occlusion can be relatively easily evaluated by collecting the density along a shadow ray towards the light source. While this process is relatively cheap for a single shadow ray, performing a secondary ray march for each primary ray’s sampled position increases the computation cost by an order of magnitude, quickly becoming too expensive for practical training. However, we observe that for most primary rays, the ray samples are closely packed together around the zero level-set in the SDF due to the importance sampling of the density along the view ray. Hence, we propose to approximate light source visibility by shooting a single shadow ray at the zero level-set, and use the same light source visibility for each sample along the view ray. To determine the depth of the zero level-set, we compute the density weighted depth along the view ray:

$$D(\mathbf{o}, \mathbf{v}) = \int_0^\infty w(\mathbf{p})t dt. \quad (6)$$

While for an opaque surface a single shadow ray is sufficient, for non-opaque or ill-defined surfaces a single shadow ray offers a poor estimate of the light occlusion. Furthermore, using the shadow information as a hard mask, ignores the effects of indirect lighting. We therefore provide the shadow information as an additional input to the radiance network, allowing the network learn whether to include or ignore the shadowing information as well as blend any indirect lighting in the shadow regions.

Highlight Hints. Similar to shadows, specular highlights are sparsely distributed high frequency light transport effects. Inspired by Gao et al. [2020], we provide specular highlight hints to the radiance network by evaluating 4 microfacet BRDFs with a GGX distribution [Walter et al. 2007] with roughness parameters {0.02, 0.05, 0.13, 0.34}. Unlike Gao et al., we compute the highlight hints using local shading which only depends on the surface normal computed from the SDF (Equation 2), and pass it to the radiance MLP as an additional input. Similar to shadow hints, we compute one highlight hint per view ray and reused it for all samples along the view ray.

3.3 Loss & Training

We jointly train the density and radiance network using an image reconstruction loss \mathcal{L}_c and an SDF regularization loss \mathcal{L}_e . The image reconstruction loss is defined as the L_1 distance between the observation $\bar{C}(\mathbf{o}, \mathbf{v})$ and the corresponding estimated color $C(\mathbf{o}, \mathbf{v})$ computed using Equation 1: $\mathcal{L}_c = ||\bar{C} - C||_1$, for a random sampling of pixels (and thus view rays) in the captured training images (subsection 3.4). Furthermore, we follow NeuS, and regularize the density MLP with the Eikonal loss [Gropp et al. 2020] to ensure a valid SDF: $\mathcal{L}_e = (||\nabla f(\mathbf{p})||_2 - 1)^2$. For computational efficiency, we do not back-propagate gradients from the shadow and highlight hints.

3.4 Data Acquisition

Training the implicit representation requires observations of the scene viewed from random viewpoints and lit from a different random light position such that shadows and interreflections are included. We follow the procedure from Gao et al. [2020]: a handheld camera is used to capture photographs of the scene from random viewpoints while a second camera captures the scene with its collocated flash light enabled. The images from the second camera are only used to calibrate the light source position. To aid camera calibration, the scene is placed on a checkerboard pattern.

All examples in this paper are captured with a Sony A7II as the primary camera, and an iPhone 13 Pro as the secondary camera. The acquisition process takes approximately 10 minutes; the main bottleneck in acquisition is moving the cameras around the scene. In practice we capture a video sequence from each camera and randomly select 500–1,000 frames as our training data. The video is captured using S-log encoding to minimize overexposure.

For the synthetic scenes, we simulate the acquisition process by randomly sampling view and light positions on the upper hemisphere around the scene with a random distance between 2 to 2.5 times the size of the scene. The synthetic scenes are rendered with global light transport using Blender Cycles.

3.5 Viewpoint Optimization

Imperfections in camera calibration can cause inaccurate reconstructions of thin geometrical features as well as lead to blurred results. To mitigate the impact of camera calibration errors, we jointly optimize the viewpoints and the neural representation.

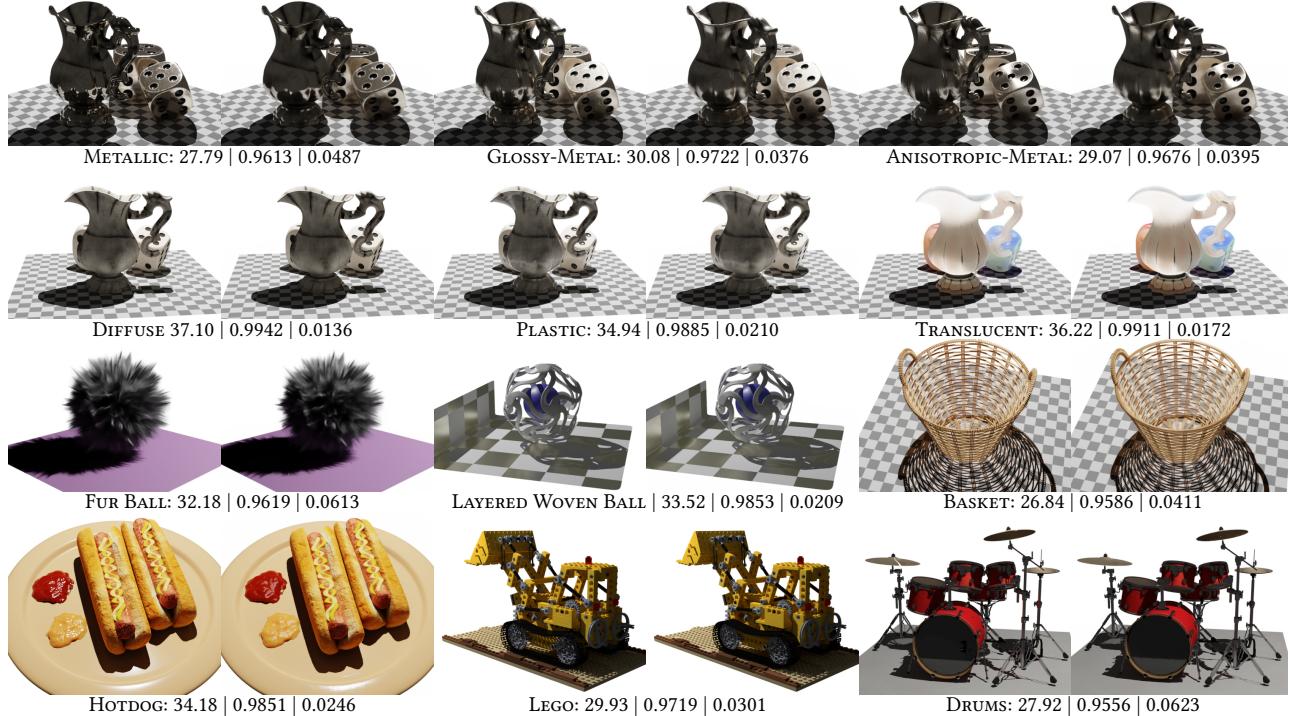


Figure 3: Qualitative comparison between synthetic scenes relit (right) for a novel viewpoint and lighting direction (not part of the training data) and a rendered reference image (left). For each example we list average PSNR, SSIM, and LPIPS computed over a uniform sampling of view and light positions.

Given an initial view orientation R_0 and view position t_0 , we formulate the refined camera orientation R and position t as:

$$R = \Delta R \cdot R_0, \quad (7)$$

$$t = \Delta t + \Delta R \cdot t_0, \quad (8)$$

where $\Delta R \in \text{SO}(3)$ and $\Delta t \in \mathbb{R}^3$ are learnable correction transformations. During training, we back-propagate the reconstruction loss, in addition to the relightable radiance network, to the correction transformations. We assume that the error on the initial camera calibration is small, and thus we limit the viewpoint changes by using a $0.06\times$ smaller learning rate for the correction transformations.

4 RESULTS

We implemented our neural implicit radiance representation in PyTorch [Paszke et al. 2019]. We train each model for 1,000k iterations using the Adam optimizer [Kingma and Ba 2015] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ with 512 samples per iteration randomly drawn from the training images. We follow the same warmup and cosine decay learning rate schedule as in NeuS [Wang et al. 2021]. Training a single neural implicit radiance representation takes approximate 20 hours on four Nvidia V100 GPUs.

We extensively validate the relighting capabilities of our neural implicit radiance representation on 17 synthetic and 7 captured scenes (including 4 from [Gao et al. 2020]), covering a wide range of different shapes, materials, and lighting effects.

Synthetic Scenes. Figure 3 shows relit results of different synthetic scenes. For each example, we list PSNR, SSIM, and LPIPS [Zhang et al. 2018] error statistics computed over 100 test images different from the 500 training images. Our main test scene contains a vase and two dice; the scene features a highly concave object (vase) and complex interreflections between the dice. We include several versions of the main test scene with different material properties: DIFFUSE, METALLIC, GLOSSY-METAL, ROUGH-METAL, ANISOTROPIC-METAL, PLASTIC, GLOSSY-PLASTIC, ROUGH-PLASTIC and TRANSLUCENT; note, some versions are only included in the supplemental material. We also include two versions with modified geometry: SHORT-FUR and LONG-FUR to validate the performance of our method on shapes with ill-defined geometry. In addition, we also include a FUR-BALL scene which exhibits even longer fur. To validate the performance of the shadow hints, we also include scenes with complex shadows: a BASKET scene containing thin geometric features and a LAYERED WOVEN BALL which combines complex visibility and strong interreflections. In addition to these specially engineered scenes to systematically probe the capabilities of our method, we also validate our neural implicit radiance representation on commonly used synthetic scenes in neural implicit modeling: HOTDOG, LEGO and DRUMS [Mildenhall et al. 2020]. Based on the error statistics, we see that the error correlates with the geometric complexity of the scene (vase and dice, HOTDOG, and LAYERED WOVEN BALL perform better than the Fur scenes as well as scenes with small details such as the LEGO and the DRUMS scene),



Figure 4: Qualitative comparison between captured scenes relit (right) for a novel viewpoint and lighting direction (not part of the training data) and a reference photograph (left). For each example we list average PSNR, SSIM, and LPIPS computed over randomly sampled view and light positions.

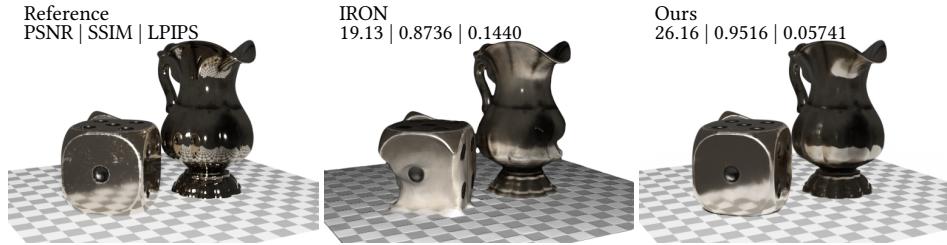


Figure 5: Comparison to inverse rendering results from IRON [Zhang et al. 2022a] (from 500 colocated training images) on the METALLIC scene. Our model is evaluated under colocated point lights. IRON is affected by the interreflections and fails to accurately reconstruct the geometry.

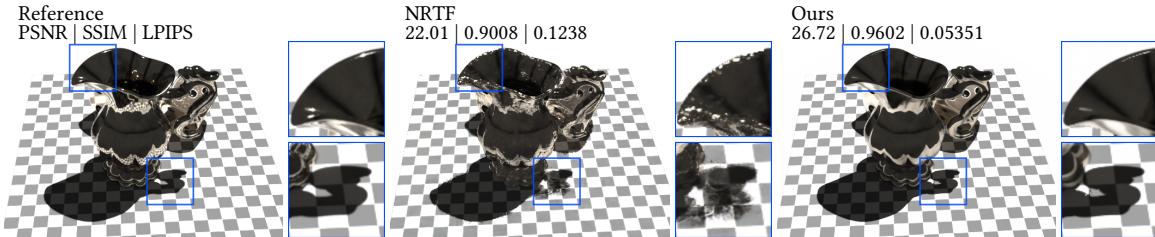


Figure 6: A comparison to Neural Radiance Transfer Fields (NRTF) trained on 500 OLAT reference images and reference geometry. To provide a fair comparison, we also train our network on the same directional OLAT images (without reference geometry) instead of point lighting. NRTF struggles to correctly reproduce shadow boundaries and specular interreflections (see zoom-ins).

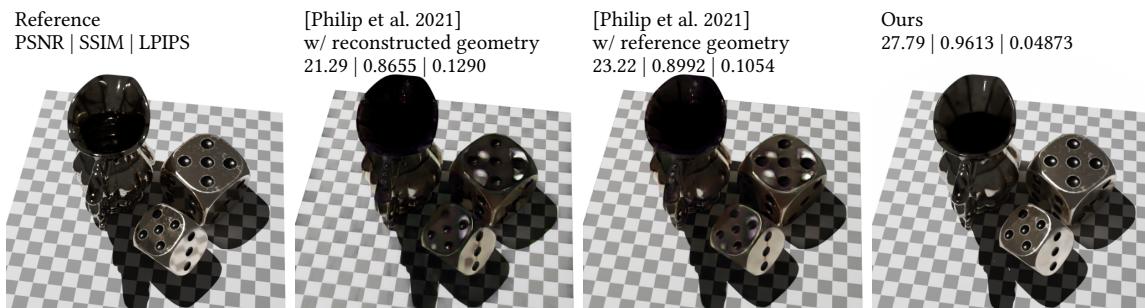


Figure 7: Comparison to the pretrained relighting network of Philip et al. [2021] on 500 input images of the METALLIC scene rendered with the target lighting. Even under these favorable conditions, their method struggles to reproduce the correct appearance for novel viewpoints.

and with the material properties (highly specular materials such as METALLIC and ANISOTROPIC-METAL incur a higher error). Visually, differences are most visible in specular reflections and for small geometrical details.

Captured Scenes. We demonstrate the capabilities of our neural implicit relighting representation by modeling 3 new scenes captured with handheld setups (Figure 4). The PIKACHU STATUE scene contains glossy highlights and significant self-occlusion. The CAT ON DECOR scene showcases the robustness of our method on real-world objects with ill-defined geometry. The CUP AND FABRIC scene exhibits translucent materials (cup), specular reflections of the balls, and anisotropic reflections on the fabric. We refer to the supplementary material for additional video sequences of these scenes visualized for rotating camera and light positions.

Comparisons. Figure 5 compares our method to IRON [Zhang et al. 2022b], an inverse rendering method that adopts a neural representation for geometry as a signed distance field. From these results, we can see that IRON fails to correctly reconstruct the shape and reflections in the presence of strong interreflections. In a second comparison (Figure 6), we compare our method to Neural Radiance Transfer Fields (NRTF) [Lyu et al. 2022]; we skip the fragile inverse rendering step and train NRTF with 500 reference OLAT images and the reference geometry. To provide a fair comparison, we also train and evaluate our network under the same directional OLAT images by conditioning the radiance network on light direction instead of point light position. From this test we observe that NRTF struggles to accurately reproduce shadow edges and specular interreflections, as well as that our method can also be successfully trained with directional lighting. Figure 7 compares our method to the pre-trained neural relighting network of Philip *et al.* [2021] on the challenging METALLIC test scene. Because multiview stereo [Schönberger and Frahm 2016] fails for this scene, we input geometry reconstructed from the NeuS SDF as well as ground truth geometry. Finally, we also render the input images under the reference target lighting; our network is trained without access to the target lighting. Even under these favorable conditions, the relighting method of Philip *et al.* struggles to reproduce the correct appearance. Finally, we compare our method to Deferred Neural Lighting [Gao et al. 2020] (using their data and trained model). Our method is able to achieve similar quality results from ~500 input images compared to ~10,000 input images for Deferred Neural Lighting. While visually very similar, the overall errors of Deferred Neural Lighting are slightly lower than with our method. This is mainly due to differences in how both methods handle camera calibrations errors. Deferred Neural Lighting tries to minimize the differences for each frame separately, and thus it can embed camera calibration errors in the images. However, this comes at the cost of temporal “shimmering” when calibration is not perfect. Our method on the other hand, optimizes the 3D representation, yielding better temporal stability (and thus requiring less photographs for view interpolation) at the cost of slightly blurring the images in the presence of camera calibration errors.

Table 1: Ablation results on synthetic scenes

Ablation Variant	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Full hints	32.02	0.9727	0.0401
w/o highlight hint	31.96	0.9724	0.0407
w/o shadow hint	27.67	0.9572	0.0610
w/o any hints	27.54	0.9568	0.0620
1 basis material	31.54	0.9707	0.0428
2 basis materials	31.54	0.9707	0.0429
4 basis materials	32.02	0.9727	0.0401
8 basis materials	31.98	0.9726	0.0401
50 training images	24.29	0.9335	0.0706
100 training images	27.96	0.9572	0.0520
250 training images	30.36	0.9666	0.0456
500 training images	32.02	0.9727	0.0401

Table 2: Ablation results of viewpoint optimization on real captured scenes

Ablation Variant	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
w/ viewpoint optimization	34.72	0.9762	0.0695
w/o viewpoint optimization	33.62	0.9719	0.0794

5 ABLATION STUDIES

We perform several ablation experiments (visual and quantitative) on the synthetic datasets to evaluate the impact of each of the components that comprise our neural implicit radiance representation.

Shadow and Highlight Hints. A key contribution is the inclusion of shadow and highlight hints in the relightable radiance MLP. Figure 9 shows the impact of training without the shadow hint, the highlight hint, or both. Without shadow hints the method fails to correctly reproduce sharp shadow boundaries on the ground plane. This lack of sharp shadows is also reflected in the quantitative errors summarized in Table 1. Including the highlight hints yield a better highlight reproduction, e.g., in the mouth of the vase.

Impact of the Number of Shadow Rays. We currently only use a single shadow ray to compute the shadow hint. However, we can also shoot multiple shadow rays (by importance sampling points along the view ray) and provide a more accurate hint to the radiance network. Figure 10 shows the results of a radiance network trained with 16 shadow rays. While providing a more accurate shadow hint, there is marginal benefit at a greatly increased computational cost, justifying our choice of a single shadow ray for computing the shadow hint.

NeuS vs. NeRF Density MLP. While the relightable radiance MLP learns how much to trust the shadow hint (worst case it can completely ignore unreliable hints), the radiance MLP can in general not reintroduce high-frequency details if it is not included in the shadow hints. To obtain a good shadow hint, an accurate depth estimate of the mean depth along the view ray is needed. Wang *et al.* [2021] noted that NeRF produces a biased depth estimate, and they introduced NeuS to address this problem. Replacing NeuS by NeRF for the density network (Figure 10) leads to poor shadow reproduction due to the adverse impact of the biased depth estimates on the shadow hints.

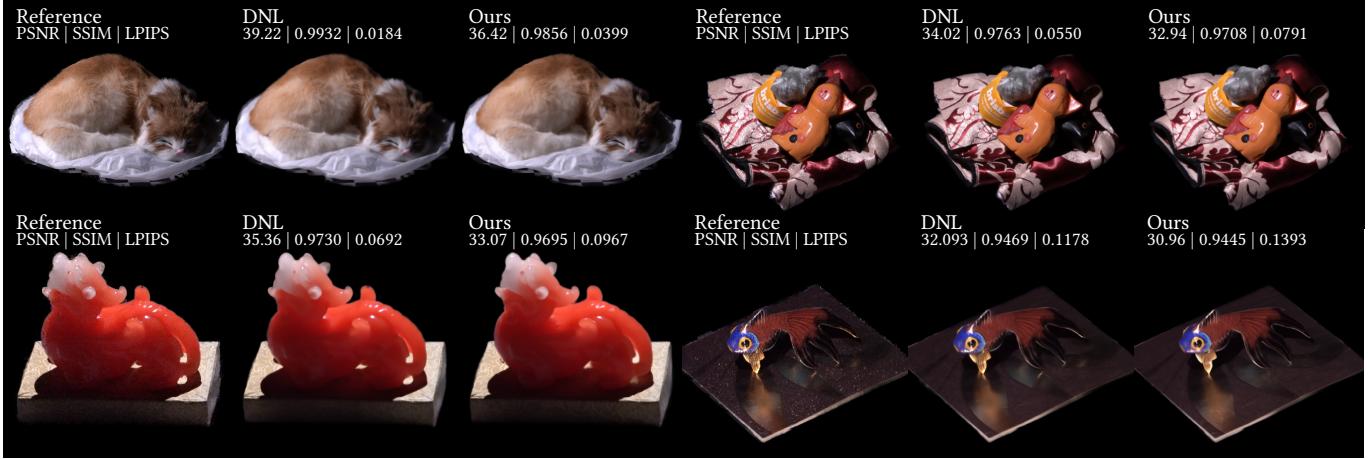


Figure 8: Comparison with Deferred Neural Lighting [Gao et al. 2020]. We train our neural implicit radiance representation using only 1/25th (~500) randomly selected frames for Gao *et al.*'s datasets, while achieving comparable results.

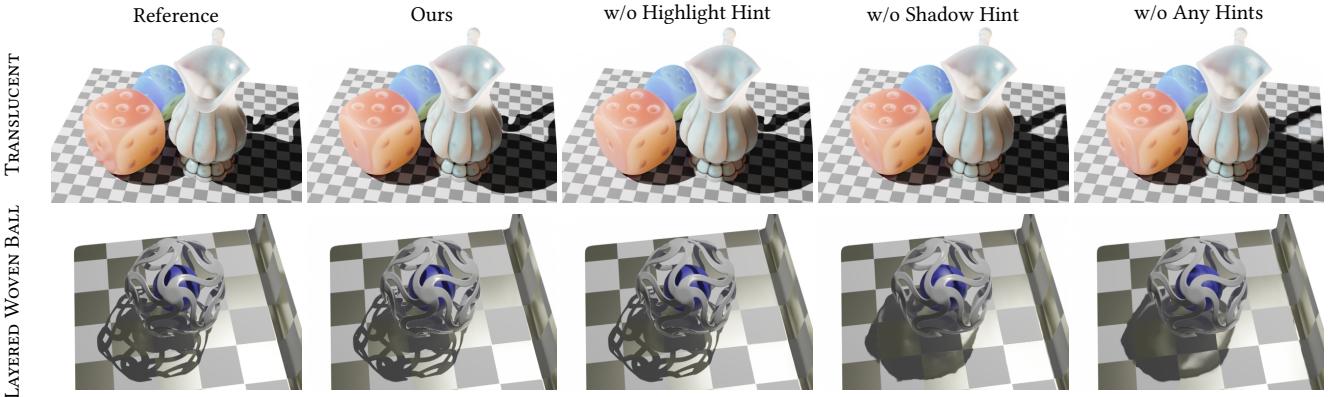


Figure 9: Impact of shadow and highlight hints; without the hints the network fails to accurately reproduce the desired effect.



Figure 10: Impact of the number of shadow rays and the underlying implicit shape representation demonstrated on the BASKET scene. Using 16 shadow rays only provides marginal improvements at the cost of significant computation overhead. Using NeRF as the basis for the neural implicit shape yields degraded shadow quality due to depth biases.

Impact of the number of Basis Materials for the Highlight Hints. Table 1 shows the results of using 1, 2, 4 and 8 basis materials for computing the highlight hints. Additional highlights hints improve the results up to a point; when too many hints are provided erroneous correlations can increase the overall error. 4 basis materials

strike a good balance between computational cost, network complexity, and quality.

Impact of Number of Training Images. Figure 11 and Table 1 demonstrate the effect of varying the number of input images from

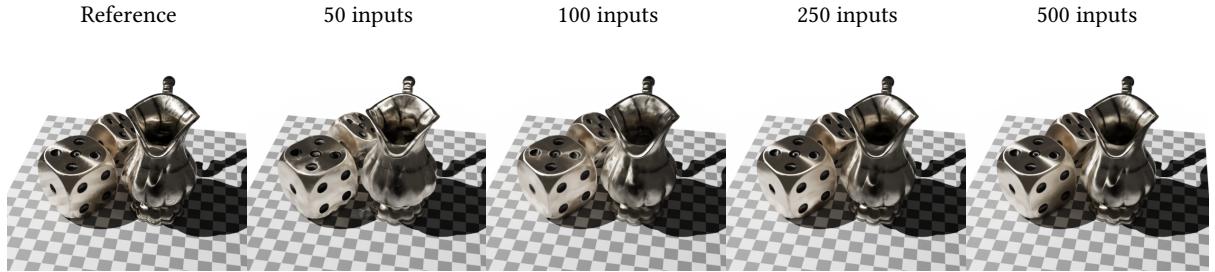


Figure 11: Impact of the number of captured training images. Increasing the number of training images improves the quality. The quality degrades significantly when the number of images is less than 250.

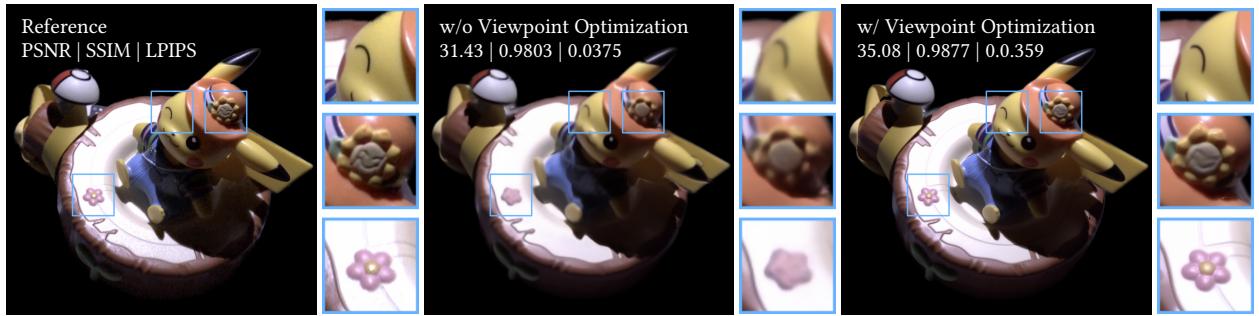


Figure 12: Effectiveness of Viewpoint Optimization. Using viewpoint optimization greatly enhances the image quality in terms of sharpness and detail.

50, 100, 250 to 500. As expected, more training images improve the results, and with increasing number of images, the increase in improvement diminishes. With 250 images we already achieve plausible relit results. Decreasing the number of training images further introduces noticeable appearance differences.

Effectiveness of Viewpoint Optimization. Figure 12 and Table 2 demonstrate the effectiveness of viewpoint optimization on real captured scenes. While the improvement in quantitative errors is limited, visually we can see that viewpoint optimization significantly enhances reconstruction quality with increased sharpness and better preservation of finer details.

6 LIMITATIONS

While our neural implicit radiance representation greatly reduces the number of required input images for relighting scenes with complex shape and materials, it is not without limitations. Currently we provide shadow and highlight hints to help the relightable radiance MLP model high frequency light transport effects. However, other high frequency effects exist. In particular highly specular surfaces that reflect other parts of the scene pose a challenge to the radiance network. Naïve inclusion of ‘*reflection hints*’ and/or reparameterizations [Verbin et al. 2022] fail to help the network, mainly due to the reduced accuracy of the surface normals (needed to predict the reflected direction) for sharp specular materials. Resolving this limitation is a key challenge for future research in neural implicit modeling for image-based relighting.

7 CONCLUSION

In this paper we presented a novel neural implicit radiance representation for free viewpoint relighting from a small set of unstructured photographs. Our representation consists of two MLPs: one for modeling the SDF (analogous to NeuS) and a second MLP for modeling the local and indirect radiance at each point. Key to our method is the inclusion of shadow and highlight hints to aid the relightable radiance MLP to model high frequency light transport effects. Our method is able to produce relit results from just ~500 photographs of the scene; a saving of one to two order of magnitude compared to prior work with similar capabilities.

ACKNOWLEDGMENTS

Pieter Peers was supported in part by NSF grant IIS-1909028. Chong Zeng and Hongzhi Wu were partially supported by NSF China (62022072 & 62227806), Zhejiang Provincial Key R&D Program (2022C01057) and the XPLORER PRIZE.

REFERENCES

- Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. 2020. X-Fields: Implicit Neural View-, Light- and Time-Image Interpolation. *ACM Trans. Graph.* 39, 6 (2020).
- Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. 2021. Deep relightable appearance models for animatable faces. *ACM Trans. Graph.* 40, 4 (2021), 1–15.
- Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020. Deep Reflectance Volumes: Relightable Reconstructions from Multi-View Photometric Images. In *ECCV*. 294–311.
- 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 *ICCV*.

- 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 *NeurIPS*.
- Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik PA Lensch. 2021b. Neural-PIL: Neural Pre-Integrated Lighting for Reflectance Decomposition. In *NeurIPS*, Vol. 34. 10691–10704.
- Guangyan Cai, Kai Yan, Zhao Dong, Ioannis Gkioulekas, and Shuang Zhao. 2022. Physics-Based Inverse Rendering using Combined Implicit and Explicit Geometries. *Comp. Graph. Forum* 41, 4 (2022), 129–138.
- Wenzheng Chen, Joey Litallien, Jun Gao, Zian Wang, Clement Fuji Tsang, Sameh Khalis, Or Litany, and Sanja Fidler. 2021. DIB-R++: Learning to Predict Lighting and Material with a Hybrid Differentiable Renderer. In *NeurIPS*.
- Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the Reflectance Field of a Human Face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. 145–156.
- Yue Dong. 2019. Deep appearance modeling: A survey. *Visual Informatics* 3, 2 (2019), 59–68.
- Farshad Einabadi, Jean-Yves Guillemaut, and Adrian Hilton. 2021. Deep neural models for illumination estimation and relighting: A survey. In *Comp. Graph. Forum*, Vol. 40. 315–331.
- Duan Gao, Guojun Chen, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. 2020. Deferred Neural Lighting: Free-Viewpoint Relighting from Unstructured Photographs. *ACM Trans. Graph.* 39, 6, Article 258 (nov 2020).
- David Griffiths, Tobias Ritschel, and Julien Philip. 2022. OutCast: Outdoor Single-image Relighting with Cast Shadows. *Comp. Graph. Forum* 41, 2 (2022), 179–193.
- Amos Gropp, Lior Yaniv, Nir Haim, Matan Atzmon, and Yaron Lipman. 2020. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099* (2020).
- Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escalano, Rohit Pandey, Jason Dourgarian, Danhang Tang, Anastasia Tkach, Adarsh Kowdle, Emily Cooper, Mingsong Dou, Sean Fanello, Graham Fyffe, Christoph Rhemann, Jonathan Taylor, Paul Debevec, and Shahram Izadi. 2019. The Relightables: Volumetric Performance Capture of Humans with Realistic Relighting. *ACM Trans. Graph.* 38, 6, Article 217 (nov 2019).
- Jon Hasselgren, Nikolai Hofmann, and Jacob Munkberg. 2022. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. In *NeurIPS*.
- Yoshihiro Kanamori and Yuki Endo. 2018. Relighting Humans: Occlusion-Aware Inverse Rendering for Full-Body Human Images. *ACM Trans. Graph.* 37, 6, Article 270 (Dec. 2018).
- Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *ICLR*.
- Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. 2022. NeRoIC: Neural Rendering of Objects from Online Image Collections. *ACM Trans. Graph.* 41, 4, Article 56 (jul 2022).
- Quwei Li, Jie Guo, Yang Fei, Feichao Li, and Yanwen Guo. 2022. NeuLighting: Neural Lighting for Free Viewpoint Outdoor Scene Relighting with Unconstrained Photo Collections. In *SIGGRAPH Asia 2022 Conference Papers*. Article 13.
- Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo Ray Tracing through Edge Sampling. *ACM Trans. Graph.* 37, 6, Article 222 (dec 2018).
- Matthew M. Loper and Michael J. Black. 2014. OpenDR: An Approximate Differentiable Renderer. In *ECCV*. 154–169.
- Fujun Luan, Shuang Zhao, Kavita Bala, and Zhao Dong. 2021. Unified Shape and SVBRDF Recovery using Differentiable Monte Carlo Rendering. *Comp. Graph. Forum* 40, 4 (2021), 101–113.
- Linjie Lyu, Ayush Tewari, Thomas Leimkühler, Marc Habermann, and Christian Theobalt. 2022. Neural Radiance Transfer Fields for Relightable Novel-View Synthesis with Global Illumination. In *ECCV*, Vol. 13677. 153–169.
- Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*.
- Abhimitra Meka, Christian Häne, Rohit Pandey, Michael Zollhöfer, Sean Fanello, Graham Fyffe, Adarsh Kowdle, Xueming Yu, Jay Busch, Jason Dourgarian, Peter Denny, Sofien Bouaziz, Peter Lincoln, Matt Whalen, Geoff Harvey, Jonathan Taylor, Shahram Izadi, Andrea Tagliasacchi, Paul Debevec, Christian Theobalt, Julien Valentin, and Christoph Rhemann. 2019. Deep Reflectance Fields: High-Quality Facial Reflectance Field Inference from Color Gradient Illumination. *ACM Trans. Graph.* 38, 4, Article 77 (jul 2019).
- Abhimitra Meka, Rohit Pandey, Christian Häne, Sergio Orts-Escalano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, et al. 2020. Deep relightable textures: volumetric performance capture with neural rendering. *ACM Trans. Graph.* 39, 6 (2020), 1–21.
- Moustafa Mahmoud Meshry, Dan B Goldman, Sameh Khamis, Hugues Hoppe, Rohit Kumar Pandey, Noah Snavely, and Ricardo Martin Brualla. 2019. Neural Rerendering in the Wild. In *CVPR*.
- 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. *ECCV* (2020).
- J. Munkberg, W. Chen, J. Hasselgren, A. Evans, T. Shen, T. Muller, J. Gao, and S. Fidler. 2022. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *CVPR*. 8270–8280.
- O. Nalbach, E. Arabadzhyska, D. Mehta, H.-P. Seidel, and T. Ritschel. 2017. Deep Shading: Convolutional Neural Networks for Screen Space Shading. *Comp. Graph. Forum* 36, 4 (2017), 65–78.
- Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H. Kim. 2018. Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography. *ACM Trans. Graph.* 37, 6, Article 267 (Dec. 2018).
- Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. 2019. Mitsuba 2: A Retargetable Forward and Inverse Renderer. *ACM Trans. Graph.* 38, 6, Article 203 (nov 2019).
- Rohit Pandey, Sergio Orts Ecolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total relighting: learning to relight portraits for background replacement. *ACM Trans. Graph.* 40, 4 (2021), 1–21.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward 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 *NeurIPS*. 8024–8035.
- Julien Philip, Michaël Gharbi, Tinghui Zhou, Alexei A. Efros, and George Drettakis. 2019. Multi-view Relighting Using a Geometry-aware Network. *ACM Trans. Graph.* 38, 4, Article 78 (July 2019).
- Julien Philip, Sébastien Morgensthaler, Michaël Gharbi, and George Drettakis. 2021. Free-Viewpoint Indoor Neural Relighting from Multi-View Stereo. *ACM Trans. Graph.* 40, 5, Article 194 (sep 2021).
- Peiran Ren, Yue Dong, Stephen Lin, Xin Tong, and Baining Guo. 2015. Image Based Relighting Using Neural Networks. *ACM Trans. Graph.* 34, 4, Article 111 (jul 2015).
- Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. 2022. NeRF for Outdoor Scene Relighting. In *ECCV*.
- Johannes Lutz Schönberger and Jan-Michael Frahm. 2016. Structure-from-Motion Revisited. In *CVPR*.
- P. P. Srinivasan, B. Deng, X. Zhang, M. Tancik, B. Mildenhall, and J. T. Barron. 2021. NeRF: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *CVPR*.
- Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E. Debevec, and Ravi Ramamoorthi. 2019. Single image portrait relighting. *ACM Trans. Graph.* 38, 4, Article 79 (2019).
- Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. 2021. NeLF: Neural Light-transport Field for Portrait View Synthesis and Relighting. In *EGSR*. 155–166.
- Tiancheng Sun, Zexiang Xu, Xiuming Zhang, Sean Fanello, Christoph Rhemann, Paul Debevec, Yun-Ta Tsai, Jonathan T Barron, and Ravi Ramamoorthi. 2020. Light stage super-resolution: continuous high-frequency relighting. *ACM Trans. Graph.* 39, 6 (2020), 1–12.
- Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, W Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. 2022. Advances in neural rendering. In *Comp. Graph. Forum*, Vol. 41. 703–735.
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. 2022. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *CVPR*. 5481–5490.
- Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet models for refraction through rough surfaces. In *EGSR*. 195–206.
- 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. *NeurIPS* (2021).
- Jiankai Xing, Fujun Luan, Ling-Qi Yan, Xuejun Hu, Houde Qian, and Kun Xu. 2022. Differentiable Rendering Using RGBXY Derivatives and Optimal Transport. *ACM Trans. Graph.* 41, 6, Article 189 (nov 2022).
- Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. 2018. Deep Image-based Relighting from Optimal Sparse Samples. *ACM Trans. Graph.* 37, 4, Article 126 (July 2018).
- Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. 2022. PS-NeRF: Neural Inverse Rendering for Multi-view Photometric Stereo. In *ECCV*.
- Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. 2022. NeILF: Neural Incident Light Field for Material and Lighting Estimation. In *ECCV*.
- Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. 2022. Learning to Relight Portrait Images via a Virtual Light Stage and Synthetic-to-Real Adaptation. *ACM Trans. Graph.* 41, 6 (2022), 1–21.
- Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. 2022a. IRON: Inverse Rendering by Optimizing Neural SDFs and Materials from Photometric Images. In *CVPR*. 5555–5564.

- Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021b. PhySG: Inverse Rendering with Spherical Gaussians for Physics-based Material Editing and Relighting. In *CVPR*.
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*. 586–595.
- Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escalano, Philip Davidson, Christoph Rhemann, Paul Debevec, Jonathan T. Barron, Ravi Ramamoorthi, and William T. Freeman. 2021a. Neural Light Transport for Relighting and View Synthesis. *ACM Trans. Graph.* 40, 1, Article 9 (jan 2021).
- Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. 2021c. NeRFactor: Neural Factorization of Shape and Reflectance under an Unknown Illumination. *ACM Trans. Graph.* 40, 6, Article 237 (dec 2021).
- Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. 2022b. Modeling Indirect Illumination for Inverse Rendering. In *CVPR*.
- Quan Zheng, Gurprit Singh, and Hans-Peter Seidel. 2021. Neural Relightable Participating Media Rendering. In *NeurIPS*, Vol. 34. 15203–15215.

Relighting Neural Radiance Fields with Shadow and Highlight Hints

Chong Zeng*

State Key Lab of CAD and CG,
Zhejiang University
Hangzhou, China
chongzeng2000@gmail.com

Guojun Chen

Microsoft Research Asia
Beijing, China
guoch@microsoft.com

Yue Dong

Microsoft Research Asia
Beijing, China
yuedong@microsoft.com

Pieter Peers

College of William & Mary
Williamsburg, USA
ppeers@siggraph.org

Hongzhi Wu

State Key Lab of CAD and CG,
Zhejiang University
Hangzhou, China
hwu@acm.org

Xin Tong

Microsoft Research Asia
Beijing, China
xtong@microsoft.com

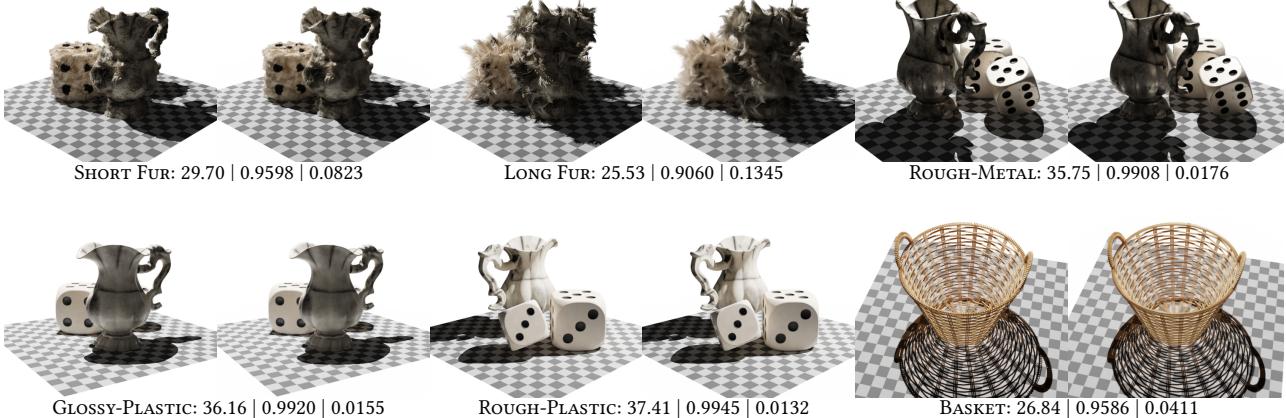


Figure 1: Qualitative comparison between additional synthetic scenes relit (right) for a novel viewpoint and novel lighting direction (not part of the training data) and a rendered reference image (left). For each example we list average PSNR, SSIM, and LPIPS computed over a uniform sampling of view and light positions.

ACM Reference Format:

Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. 2023. Relighting Neural Radiance Fields with Shadow and Highlight Hints. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings (SIGGRAPH '23 Conference Proceedings)*, August 6–10, 2023, Los Angeles, CA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3588432.3591482>

1 ADDITIONAL RESULTS

Figure 1 shows additional synthetic results to further test our method on scenes with different material properties. The BASKET scene is included in the ablation study figures, but not listed in Figure 3 (of the main paper); we include it here for completeness.

2 NETWORK ARCHITECTURE DETAILS

We follow exactly the same architecture as NeuS [Wang et al. 2021] for the density MLP: 8 hidden layers with 256 nodes using a Softplus

*Work done during internship at Microsoft Research Asia.

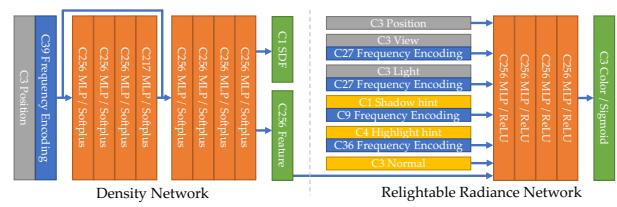


Figure 2: Detailed network architecture of the density and relightable radiance network. The number of output channels and activations are also marked.

activation and a skip connection between the input and the 4th layer. The input (i.e., current position along a ray) is augmented using a frequency encoding with 6 bands. The relightable radiance network has a similar network architecture as NeuS' color MLP: 4 hidden layers with 256 nodes using a ReLU activation. The final color is outputted after a Sigmoid activation, ensuring that the

output color is within the $(-1, 1)$ range. Figure 2 details network architecture of our method.

REFERENCES

- 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. *NeurIPS* (2021).