# ReNeRF: Relightable Neural Radiance Fields with Nearfield Lighting

Yingyan Xu[1,2] Gaspard Zoss[2] Prashanth Chandran[2]
Markus Gross[1,2] Derek Bradley[2] Paulo Gotardo[2*]
[1]ETH Zürich [2]DisneyResearch|Studios
{yingyan.xu,grossm}@inf.ethz.ch, gotardop@gmail.com
{gaspard.zoss,prashanth.chandran,derek.bradley}@disneyresearch.com

## Abstract

*Recent work on radiance fields and volumetric inverse rendering (*e.g.*, NeRFs) has provided excellent results in building data-driven models of real scenes for novel view synthesis with high photorealism. While full control over viewpoint is achieved, scene lighting is typically "baked" into the model and cannot be changed; other methods only capture limited variation in lighting or make restrictive assumptions about the captured scene. These limitations prevent the application on arbitrary materials and novel 3D environments with complex, distinct lighting. In this paper, we target the application scenario of capturing high-fidelity assets for neural relighting in controlled studio conditions, but without requiring a dense light stage. Instead, we leverage a small number of area lights commonly used in photogrammetry. We propose* ReNeRF*, a relightable radiance field model based on the intuitive and powerful approach of image-based relighting, which implicitly captures global light transport (for arbitrary objects) without complex, error-prone simulations. Thus, our new method is simple and provides full control over viewpoint and lighting, without simplistic assumptions about how light interacts with the scene. In addition, ReNeRF does not rely on the usual assumption of distant lighting – during training, we explicitly account for the distance between 3D points in the volume and point samples on the light sources. Thus, at test time, we achieve better generalization to novel, continuous lighting directions, including nearfield lighting effects.*

## 1. Introduction

Neural 3D scene representations such as Neural Radiance Fields (NeRFs) and volumetric neural rendering [28, 29] have recently demonstrated the ability to model the real 3D world with impressive ease and photorealism, making this task as easy as capturing a short video using a smart phone. NeRFs can synthesize novel images from controlled and arbitrary viewpoints, thus enabling virtual fly-throughs within captured scenes [24, 46, 47]. In addition, these techniques are highly valuable in many applications (*e.g.*, feature films, games, VR, AR, telepresence), whose virtual environments can be populated with photoreal 3D assets simply by photographing the real world. In this paper, we target the application scenario of capturing such high-fidelity rendering assets within controlled studio conditions, leveraging existing photogrammetry hardware.

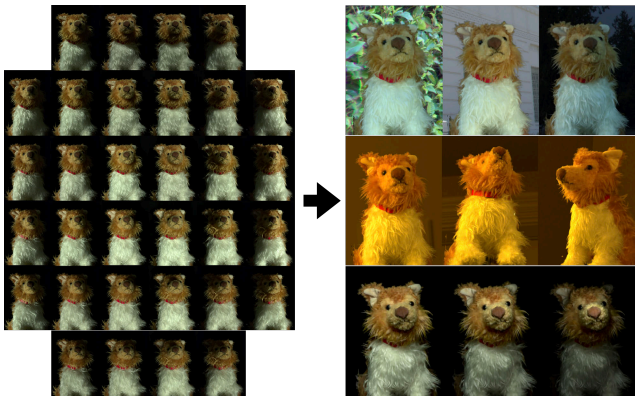

Figure 1. We propose ReNeRF, a relightable neural radiance field trained on multi-view imagery captured under simple OLAT area lights (left). Once trained, our method allows re-rendering under novel viewpoints and illumination conditions, including both environment maps and nearfield light sources (right).

NeRFs represent 3D objects and scenes using neural networks (multilayer perceptrons, or MLPs) to model a continuous look-up function that encodes the density and color at any 3D point within the observed volume of space. As a key limitation, the scene lighting is typically fixed ("baked") into the model's colors and cannot be altered. As such, a model captured within a particular lighting environment (*e.g.*, in studio) cannot be rendered photorealistically within another environment that has its own, distinct lighting conditions (*e.g.*, in a feature film). Therefore, we require the ability to photorealistically relight these neural field mod-

*Now at Google

els in order to extend their usefulness to a larger variety of applications. Although there have already been attempts at making NeRFs relightable (see Section 2), the techniques proposed so far provide limited control over lighting, photorealism, generalization to different scene materials and illumination, or ease of capture and training. Most previous methods further break down in the presence of nearfield illumination changes, which pose additional challenges over the traditional assumption of distant illumination.

To address these limitations, we present a technique for building high-quality, relightable neural radiance fields (ReNeRFs) from images captured in studio conditions, that can achieve full control over camera viewpoint while relighting for any illumination, including nearfield lighting. Our method is inspired by the simplicity and effectiveness of image-based relighting (IBRL) [35, 13], a data-driven approach to model complex global illumination in scenes with arbitrary materials, without explicit light transport simulations. IBRL exploits light superposition to photorealistically perform relighting by linearly combining a *discrete* number of basis images, captured under one-light-at-a-time (OLAT) conditions. While IBRL is essentially 2D and assumes distant lighting and fixed viewpoints, our ReNeRF method is fully 3D and supports novel views while explicitly accounting for the distance between 3D points in the scene and on the light sources, thus capturing nearfield lighting effects. In addition, ReNeRF learns to nonlinearly interpolate a *continuous* OLAT basis at variable distances, effectively performing OLAT super-resolution and generalizing to novel lighting conditions. We show how this model can be learned using only a few area light sources that are often employed for in-studio photogrammetry, therefore not requiring complex and dense light stage setups.

## 2. Related Work

The literature on NeRF models has been growing incredibly fast, with many works devoted to improving rendering quality [2, 48, 27], geometry representation [36, 50], deformation [49, 41, 15, 38, 39], scene size [24, 46, 47], and rendering speed [12, 14, 16, 30, 54]. As most of that work is orthogonal to our contributions, this section focuses on relightable NeRFs, 2D IBRL and its extensions of to 3D.

Related work aimed at making NeRFs relightable has focused on decomposing scene shape, reflectance, and illumination. Given the highly ill-posed inverse rendering problem, all these methods require at least one of the following: (1) precomputed surface geometry (difficult with thin structures, unsuitable for non-solids) [22, 59, 53], (2) predefined BRDF or phase function [44, 6, 5, 56, 60, 32, 55], (3) simplified lighting model (distant lighting, smooth lighting) [20, 7, 43], or (4) simplified light transport model (no shadowing, no translucency, direct-only or simplistic indirect illumination) [44, 58, 56, 52, 32, 7, 9, 19, 8]. Additionally, most methods represent materials and light transport with a static (colored) albedo and neutral shading; these simplistic models cannot account for complex effects such as changes in surface color due to light passing through thin volumes (*e.g.*, human ears, nostrils). Combined, the above assumptions limit relighting quality on actual real-world scenes with complex geometry, materials, and light transport (*e.g.*, human heads with intricate reflectance and geometry). Our work is free of such assumptions and performs photorealistic relighting in free viewpoints without prescribed models for shape, material, or light transport.

An approach that does not require explicit parameterization and simulation is to condition the model on a learned latent code that captures the changes in appearance due to changes in the lighting conditions [24, 46, 47]; some methods combine a coarse scale NeRF-like model with convolutional 2D upsampling on the rendered image plane [11, 34, 18]. This conditioning strategy is often used to tolerate small variability in lighting but does not provide an explicit and intuitive mechanism to control lighting.

IBRL [35, 13] is a technique for hyper-realistically modifying the lighting of a real scene after it has been photographed (or for relighting rendered virtual scenes) from fixed viewpoints. Since light is additive, relighting is achieved via a simple linear combination of a set of *basis images*, which are captured under time-multiplexed, directional (OLAT) lighting from a small number of camera viewpoints [25]. IBRL guarantees photorealistic results with full camera resolution but it is essentially a 2D technique. As another drawback, IBRL requires capturing images under the illumination of a dense light stage with complex and expensive hardware. Attempts at extending IBRL from 2D to 3D have so far predominantly used mesh-based representations, leading to the natural application of IBRL in 2D texture space [17, 26, 57, 4]. In comparison to NeRFs, these mesh-based methods are limited to modeling relatively smooth surfaces and cannot properly represent non-solids and high-frequency geometry in thin structures such as fur and human hair. Here, we extend the idea underlying IBRL to enable the relighting of NeRFs, thus allowing for full control over both (continuous and arbitrary) lighting and viewpoint. Other 2D methods use U-Nets to predict albedo and normal maps and relight portrait images [37, 33]; these methods cannot synthesize novel views and present inconsistent results across real views.

Two works are closer in spirit to our approach. Neural light-transport fields (NeLFs) [45] train an MLP to output, for each 3D point, a *discrete* OLAT basis (a 128-vector) that is multiplied with a small $8 \times 16$ environment map. Neural radiance transfer fields (NRTFs) [22] train an MLP to model a *continuous* light transport function within the field. These two methods share common limitations: (1) they assume distant lighting and do not model nearfield lighting;

(2) to avoid requiring a dense light stage to capture training data, synthetic OLAT images are rendered and used for training. Thus, to simulate light stage data, these methods go back to inaccurate 3D meshes and require simplifying assumptions about materials and light transport in the scene. They still require explicit and complex simulations for training, and the real light transport is not captured in the real training data, going against the key ideas of IBRL. As we will show, ReNeRF captures more comprehensive volumetric light transport including nearfield lighting using a simple capture setup, allowing to relight complex objects like human faces in arbitrary environments from novel views.

## 3. Relightable Neural Radiance Fields

ReNeRF is inspired by IBRL, where the idea is to capture a number of images of an object lit from many different OLAT directions and blend the resulting 2D images linearly to create synthetic renditions of the object under novel lighting. We extend this concept to NeRFs by capturing multi-view images of an object lit by a few small area-OLATs and construct a NeRF model that disentangles point-OLATs and allows to relight the object from novel viewpoints and novel light positions, including nearfield lights.

Next, we will describe our capture setup and data preparation methods (Section 3.1), followed by our light transport model for nearfield lighting within a NeRF framework (Section 3.2). We then outline the ReNeRF architecture for rendering with point (Section 3.3) and more general lighting (Section 3.4), followed by the training details (Section 3.5).

### 3.1. Data Capture and Preprocessing

As we target the application scenario of capturing high-quality rendering assets under controlled in-studio conditions, we seek to leverage existing photogrammetry hardware that is considerably simpler than dense light stages. Our capture setup consists of 10 video cameras and 32 area light sources (LED bars), spread around the frontal hemisphere of a small capture volume of about $0.3\text{m}^3$, as illustrated in Fig. 2 (left). On average, the LED bars and cameras are placed at about 1 meter from the captured objects. All cameras and light sources operate in sync. The LED bars are triggered as OLATs, and we add a "full ON" and "full OFF" condition (for dark frame subtraction) for a total of 34 lighting conditions that are captured in under 1.5 seconds at 24 fps. An example set of OLAT images obtained from a frontal camera is illustrated in Fig. 2 (right). Our 32 light sources are in stark contrast with the 331 sources used in recent work [25, 17, 26, 57] and 460 sources in Bi *et al*. [4]. In our experiments in Section 4, we also show results from training on only 16 area-OLATs. We do not require polarization filters for lights and cameras.

We calibrate camera positions (extrinsics and intrinsics) using standard stereo vision tools and color-correct images

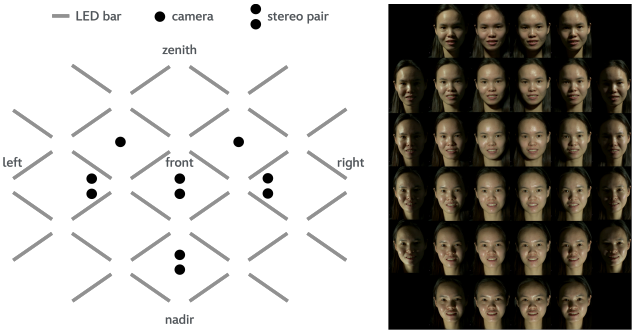


Figure 2. Capture setup for in-studio photogrammetry: layout of 10 video cameras and 32 area lights (LED bars) placed around the frontal hemisphere of the captured volume (left); and 32 area-OLAT images captured under the lighting of one LED bar (right).

using a color chart. As another standard practice, we capture multi-view images of a mirror sphere showing all light sources. Given the known sizes of the mirror sphere and rectangular LED bars, we triangulate their 3D positions following the method of Xu *et al*. [51]. Although not needed by our new method, we construct LatLong environment maps of all 34 lighting conditions for comparison with baseline methods in Section 4. This complete calibration is done only once and then re-used for different captured objects. Finally, precise calibration of (white) LED-bar intensities is not needed for our new method because they are optimized together with our neural network weights during training.

Our cameras include four narrow-baseline stereo pairs, Fig. 2 (left), used for traditional depth estimation via binocular stereo [3]. The resulting depth provides weak supervision for our 3D density fields (Section 3.5). Finally, when capturing non-static, live objects such as human heads, we follow the usual procedure and apply optical flow [10] alignment against the full-ON image in each view [25], to remove small involuntary motion from the set of OLATs.

### 3.2. Image Formation Model

To derive our new model for nearfield light transport, we begin with the familiar rendering equation [40], which defines a generic representation of the radiance $L_o(\mathbf{x}, \omega_o)$ leaving the 3D point $\mathbf{x}$ in the outgoing direction $\omega_o$,

$$L_o(\mathbf{x}, \omega_o) = L_e(\mathbf{x}, \omega_o) + \int_\Omega f(\mathbf{x}, \omega_i, \omega_o) |\cos\theta_i| L_i(\mathbf{x}, \omega_i) \mathrm{d}\omega_i. \quad (1)$$

This equation accounts for radiance that is emitted, reflected, and transmitted within the scene. $L_e(\cdot)$ and $L_i(\cdot)$ are the emitted and incident radiances, while $f(\cdot)$ is the bidirectional scattering distribution function (BSDF); $\theta_i$ and $\mathrm{d}\omega_i$ are the incidence and solid angles, respectively.

To capture global light transport without a prescribed BSDF $f(\cdot)$, nor complex recursive simulation, we encap-

sulate this model into a data-driven light transport function $F_t(\cdot)$ for a given light source. We model a small, nearfield area light $A$ as a set of $N_p$ uniformly-spaced *near point lights*, each with a position $\mathbf{p}$ and an intensity $L_e(\mathbf{p})$,

$$L_o(\mathbf{x}, \omega_o, A) = \frac{1}{N_p} \sum_{\mathbf{p} \in A} F_t\left(\mathbf{x}, \frac{\mathbf{p}-\mathbf{x}}{d_{\mathbf{px}}}, \omega_o\right) \frac{L_e(\mathbf{p})}{d_{\mathbf{px}}^2}. \quad (2)$$

This model explicitly accounts for nearfield lighting by deriving the incident lighting direction $\omega_i = \frac{\mathbf{p}-\mathbf{x}}{d_{\mathbf{px}}}$ from the 3D positions $\mathbf{p}$ and $\mathbf{x}$ and the Euclidean distance between them, $d_{\mathbf{px}}$, which attenuates the intensity of light. In contrast, related work [45, 22] considers only distant point sources and $\omega_i$ is therefore spatially constant. Borrowing from IBRL, $\omega_i$ does not represent the actual path along which light arrives at $\mathbf{x}$; it is simply *a label used to index the OLAT basis*. The continuous OLAT basis learned by ReNeRF will implicitly capture the complex light transport in the scene.

Volumetric rendering is accomplished by incorporating Eq. 2 into the rendering procedure of a standard NeRF, as illustrated in Fig. 3(a). The RGB color $\mathcal{C}$ of each image pixel $\mathbf{u}$ under area light $A$ is rendered by integrating 3D points $\mathbf{x}_i$ sampled along the pixel's ray $\omega_o(\mathbf{u})$ that is cast through the volume,

$$\mathcal{C}(\omega_o, A) = \sum_i T_i \left(1 - e^{-\sigma_i \delta_i}\right) L_o(\mathbf{x}_i, \omega_o, A), \quad (3)$$

where $\delta_i = \|\mathbf{x}_{i+1} - \mathbf{x}_i\|_2$ and $T_i$ is the accumulated transmittance defined by the densities $\sigma_i = \sigma(\mathbf{x}_i)$ [28]. Full control over viewpoint and lighting is thus achieved with the rendering model in Eq. 3. Given the principle of light superposition, this new model generalizes to arbitrary lighting (including distant lighting in an environment map) by summing the contributions of multiple area lights at different positions and sizes. Rendering can be implemented efficiently for scenes with many lights; most MLP layers are evaluated only once per sample $\mathbf{x}_i$ (Section 3.4).

By training a NeRF MLP to encode the new radiance model $L_o(\cdot)$ in Eq. 2, we enable our ReNeRF method to learn global light transport $F_t(\cdot)$ from training images captured under area-OLAT lighting. In this process, ReNeRF learns to disentangle individual point-OLAT sources $\mathbf{p}$ and achieves good generalization over other nearlight positions (distances and directions) not captured in the training data.

### 3.3. ReNeRF Architecture

As in a standard NeRF, ReNeRF represents the captured volume implicitly within an MLP, whose architecture is outlined in Fig. 3(b). This neural network includes the usual input and outputs of other NeRFs, with a novel input: the point light sources within the area light $A$. ReNeRF represents the light transport function $F_t(\cdot)$ in Eq. 2 using three main elements: a *NeRF MLP*, an *OLAT MLP*, and a *spherical codebook* of learned OLAT codes, as described next.

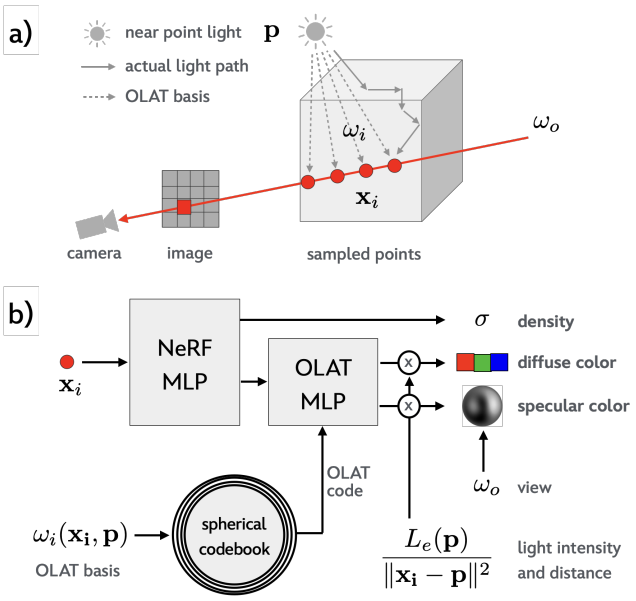


Figure 3. Our image formation model and ReNeRF architecture: (a) volumetric rendering with a nearfield OLAT basis (indexed by the light direction $\omega_i$) that continuously varies across the 3D points $\mathbf{x}_i$ within the volume. Our ReNeRF architecture (b) includes a new OLAT MLP that is conditioned on NeRF features and a learned OLAT code associated with $\omega_i$. The output comprises density, diffuse color, and spherical harmonics coefficients for specular color.

The NeRF MLP takes as input a single 3D point $\mathbf{x}_i$, and outputs density $\sigma_i$ and a 256-d vector $\mathbf{f}_i$ of neural features,

$$[\sigma_i, \mathbf{f}_i] = \mathrm{NeRF}(\mathbf{x}_i). \quad (4)$$

In contrast to a standard NeRF, we replace the subsequent radiance MLP by our novel OLAT MLP, which is conditioned on both $\mathbf{f}_i$ and another 256-d vector with a learned point-OLAT code, $\mathbf{o}_{\omega_i}$, and produces diffuse color $\mathbf{c}_{id}$ and specular coefficients $\mathbf{c}_{is}$ as

$$[\mathbf{c}_{id}, \mathbf{c}_{is}] = \mathrm{OLAT}(\mathbf{f}_i, \mathbf{o}_{\omega_i}). \quad (5)$$

The view-dependent specular coefficients $\mathbf{c}_{is} \in \mathbb{R}^{\hat{K}}$ are modeled within a low-frequency spherical harmonics (SH) subspace of order 3 ($\hat{K} = 9$). This constraint enforces smoothness across viewpoints $\omega_o$ [54, 49] and can be relaxed when training with denser sets of camera views. The outgoing radiance at $\mathbf{x}_i$, under point light $\mathbf{p}$, is then

$$L_o(\mathbf{x}_i, \omega_o, \mathbf{p}) = \left(\mathbf{c}_{id} + \mathbf{c}_{is}^T \mathrm{SH}(\omega_o)\right) \frac{L_e(\mathbf{p})}{\|\mathbf{x}_i - \mathbf{p}\|^2}. \quad (6)$$

Finally, the *spherical codebook* provides a natural embedding of the full set of learned OLAT codes $\mathbf{o}_{\omega_i}$ on a 3D sphere. This representation brings two key benefits: (1) it provides an intuitive way to index the learned OLAT

codes using the direction $\omega_i$; and (2) it defines the natural space (neighborhood) for interpolating novel codes not seen during training, thus helping achieve generalization to new lighting at test time. To implement a continuous codebook that encodes our continuous OLAT basis and allows for smooth interpolation, a straightforward solution is to adopt an order-$K$ spherical harmonics parameterization,

$$\mathbf{o}_{\omega_i} = \mathbf{C}\,\mathrm{SH}(\omega_i) \quad \text{with} \quad \mathbf{C} \in \mathbb{R}^{256\times K^2}. \tag{7}$$

This representation effectively couples the optimization of all OLAT codes and the problem becomes solving for a vector-valued function defined on the 3D sphere and compactly represented by the learned coefficient matrix $\mathbf{C}$ above. We set $K = 5$ for a continuous, smooth OLAT basis.

To implement this spherical codebook, combined with our OLAT MLP, note that matrix $\mathbf{C}$ is itself a linear operator (fully-connected layer) that can be trivially embedded into the first layer of the OLAT MLP. Therefore, in practice, we simply concatenate $\mathbf{f}_i$ and $\mathrm{SH}(\omega_i)$ and feed this longer vector to the OLAT MLP, which computes the OLAT code $\mathbf{o}_{\omega_i}$ only implicitly. While this can be seen as position-encoding $\omega_i$ with the SH basis, we will leverage the interpretation in Eq. 7 in our training loss as a way to regularize the codebook $\mathbf{C}$ (a sub-block in the first MLP layer). For a more detailed description of the layers in our network architecture, please refer to the supplementary material.

### 3.4. Using Area Lights & Environment Maps

The ReNeRF architecture in Fig. 3(b) considers only a single point light $\mathbf{p}$ at a time. To render with an area light or environment map we can approximate the complex lighting by multiple point sources that are either near or very distant, respectively. We then return to the idea of IBRL and create multiple renders (one per point source) and combine them in image-space to create the final render. In this process, the NeRF MLP needs to be evaluated only once for each 3D point $\mathbf{x}_i$. The output neural features $\mathbf{f}_i$ are then re-used to evaluate the OLAT MLP multiple times, once per point source $\mathbf{p}$. The contributions of the light sources are aggregated simply by summing the output diffuse and specular colors. At test time, when rendering with a LatLong environment map (distant lighting), we feed a direction $\omega_i$ that is taken from each pixel position in the LatLong and use the pixel intensity to directly replace the term $\frac{L_e(\mathbf{p})}{\|\mathbf{x}_i-\mathbf{p}\|^2}$. Here, $L_e(\mathbf{p})$ can be any RGB color. However, at training time, our calibrated area sources are modeled as white lights.

### 3.5. Training

During training, ReNeRF's network weights (including the codebook) and the area OLAT intensities $L_e(\mathbf{p})$ (uniform over each LED bar) are optimized together as to minimize the sum of the different losses defined below. Unless stated otherwise, we train each ReNeRF model using 32 area-OLAT images captured under 10 camera viewpoints, for a total of 320 images for each captured object.

Our rendering loss encourages the model to reproduce the color $\mathcal{C}^*_n(\mathbf{u})$ of each pixel $\mathbf{u}$ in the $n$-th training image,

$$\mathcal{L}_{RGB} = \lambda_{RGB} \sum_{n,\mathbf{u}} \big\|\gamma(\mathcal{C}^*_n(\mathbf{u})) - \gamma(\mathcal{C}_n(\mathbf{u}))\big\|_2^2. \tag{8}$$

Here, $\mathcal{C}_n(\mathbf{u})$ is the rendered pixel obtained from Eq. 3 and Eq. 6, and $\lambda_{RGB}$ is a predefined, fixed weight for the loss. This loss is taken after applying standard sRGB gamma $\gamma(\cdot)$ correction to the pixel values, to better fit shadowed areas (undertones), while de-emphasizing highlights slightly.

To constrain the density field, we compute background (matting) masks $\mathcal{M}_n$ and use them to penalize density along rays that go through empty space. The depth maps $\mathcal{D}^*_n(\mathbf{u})$, precomputed by our pairs of stereo cameras (Section 3.1), provide weak supervision for the NeRF's expected depth $\mathcal{D}_n(\mathbf{u})$ at each pixel,

$$\mathcal{L}_\sigma = \sum_{\substack{\mathbf{u}\in\mathcal{M}_n \\ n,i}} \lambda_m|\sigma_i| + \sum_{\substack{\mathbf{u}\notin\mathcal{M}_n \\ n}} \lambda_{\mathcal{D}}\|\mathcal{D}^*_n(\mathbf{u}) - \mathcal{D}_n(\mathbf{u})\|_2^2. \tag{9}$$

When the precomputed depth is inaccurate (*e.g.*, for furry or hairy objects), we enable depth supervision only in the first 10K iterations, as a means to speed up convergence.

As observed by Li *et al.* [20], we found that L2-norm regularization prevents undesired attenuation of the rendered specular colors, encoded by the SH coefficients $\mathbf{c}_{is}$ in Eq. 5. To enforce smoothness and avoid overfitting, we also apply similar regularization to the SH coefficient matrix $\mathbf{C}$ of our OLAT codebook, within the first layer of the OLAT MLP,

$$\mathcal{L}_{SH} = \lambda_s \sum_{n,i,\mathbf{u}} \|\mathbf{c}_{is}\|_2^2 + \lambda_{\mathrm{w}}\|\mathbf{C}\odot\mathbf{W}\|_F^2. \tag{10}$$

Here, $\|\cdot\|_F$ is the Frobenius norm, $\odot$ is the element-wise product, and $\mathbf{W}$ is a matrix with weights where $\mathbf{W}_{i,j} = (k+1)$ when $\mathbf{C}_{i,j}$ corresponds to a SH coefficient of order $k$. These weights are used to penalize high-frequency terms more strongly.

To encourage smoothness throughout the ReNeRF network, we also apply a simplified Lipschitz regularizer [21]. Let $\mathbf{M}_l$ be the weight matrix of MLP layer $l$ and let $\|\mathbf{M}_l\|_\infty$ be the weight with the highest absolute value, we define

$$\mathcal{L}_{\mathrm{w}} = \lambda_{\mathrm{w}} \sum_l \|\mathbf{M}_l\|_\infty. \tag{11}$$

We train ReNeRF with weights $\lambda_{RGB} = 1$, $\lambda_{\mathcal{D}} = 0.002$, $\lambda_m = \lambda_s = 0.01$, $\lambda_{\mathrm{w}} = 0.001$. A high $\lambda_s = 10$ is set in the first 10K iterations to force radiance to be predominantly

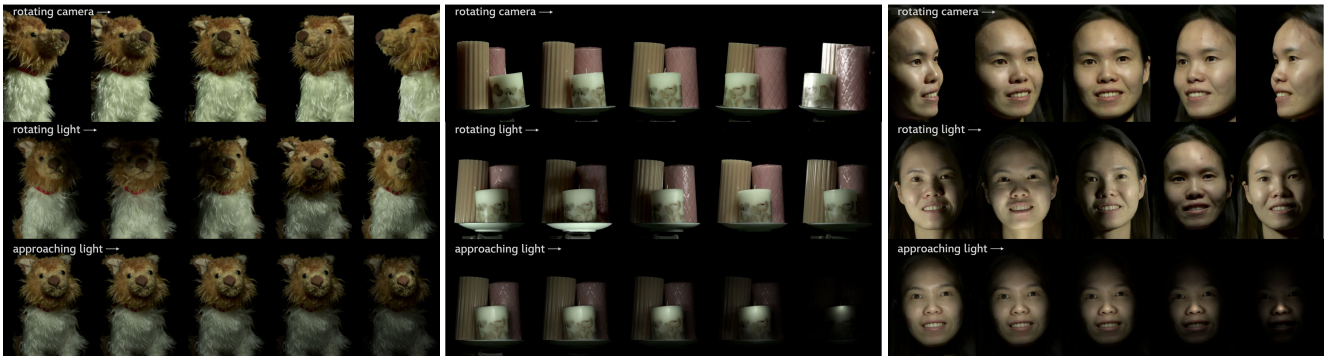


Figure 4. ReNeRF renderings for three captured real scenes: a furry dog, a plate with wax candles, and a human head. These scenes include a variety of materials with complex light transport. We show ReNeRF's ability to render the scene under novel views (fixed lighting), novel lighting directions (fixed view), and under an approaching near point light. For additional examples, please see the supplemental video.

explained by the diffuse component. We position-encode the input points $\mathbf{x}_i$ using 10 frequency bands. During the first 150K iterations, we train with $N_p = 1$ point source $\mathbf{p}$ per LED bar (at the center); we then add two other sources ($N_p = 3$) at the extremities of each bar, which is sufficient due to the smoothness of our codebook. As we use a plain NeRF MLP without recent extensions to accelerate rendering [30], training time can be quite long. For the results presented next, we limit the image resolution to $384 \times 512$ (unless stated otherwise). At this resolution, we train each ReNeRF model with an Adam optimizer for about 200K iterations with a batch of 8K rays, or approximately 28 hours on a single NVidia RTX3090 GPU.

## 4. Results

This section presents results of ReNeRF, trained on both real and synthetic datasets, and demonstrates the ability to generate renditions of different objects under novel views and lighting conditions with high photorealism. The new results are also analyzed in comparison to those obtained with alternative model designs and baseline methods in the literature. Please refer to the supplemental material for the complete set of renderings and comparisons.

**Novel views and relighting.** We now demonstrate how ReNeRF effectively extends the novel view synthesis capabilities of a NeRF with the simplicity and photorealism of IBRL. We trained a ReNeRF model for different real scenes with a variety of complex materials, as captured within our photogrammetry setup. The toy dog in Fig. 4 (left) has shiny eyes and long furs that introduce complex self-occlusion and self-shadowing, making it particularly challenging to capture and re-render (see mesh-based result in the supplemental material). The ceramic plate with three wax candles in Fig. 4 (middle) also shows complex shadows and light transport including subsurface scattering. In addition, the human head in Fig. 4 (right) is comprised of semi-translucent skin, shiny eyes and teeth, and thin structures

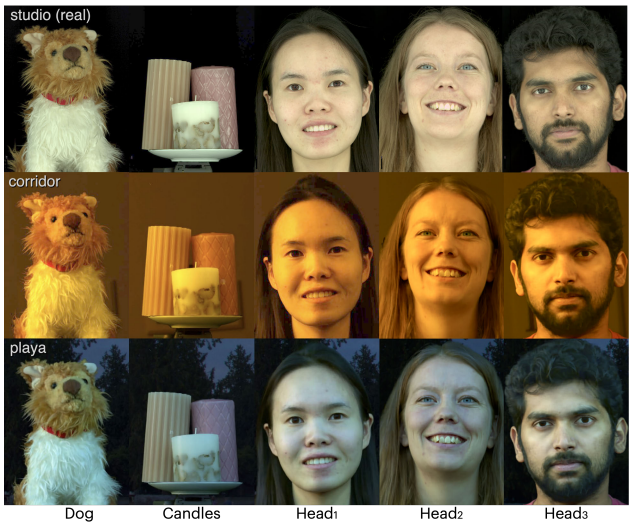


Figure 5. ReNeRF models rendered within two new, distant lighting environments encoded by LatLong maps (shown on the background). The real full-ON image (top) is also shown for reference.

on scalp hair that cannot be easily modeled by traditional mesh-based methods [42, 23]. As ReNeRF is purely data-driven and free of simplifying assumptions, it successfully learns to render these complex scenes under novel views and lighting with high photorealism. This includes not only rendering the object under a continuously varying lighting direction, but also under a point light that moves extremely close to the scene and intensifies the nearfield lighting effects. A shown in Fig. 4 and in the supplemental video, the motion of surface highlights and cast shadows realistically correspond to the 3D motion of the near point source.

**Rendering with environment maps.** One of the main application scenarios that we target is in capturing high-fidelity 3D assets that can be rendered photorealistically in novel environments with their own distinct lighting. Fig. 5 and the supplemental material show pre-trained ReNeRF models rendered within different environments whose dis-


Figure 6. Diffuse and specular radiances are disentangled by ReNeRF without polarization filters, keeping the capture setup simple and facilitating inspection of the light transport model. These renders correspond to the first and last columns, third row of Fig. 4.

tant lighting is encoded in a LatLong map. These results suggest that different virtual environments could be populated with high-quality assets simply by capturing images in-studio, using a photogrammetry hardware setup that is more affordable and less complex than dense light stages.

**Diffuse-specular separation.** ReNeRF can successfully learn to represent diffuse radiance separately from specular radiance, without requiring polarization filters in front of lights or cameras [42, 23], keeping our setup simple. Fig. 6 shows the separate diffuse and specular rendering layers for the models in Fig. 4 (bottom row), under a near point light. While the toy dog always shows specular reflections due to the complex fur geometry, the smoother surfaces of the candles and face become more diffuse as the point light gets very close and specular reflection is directed towards the side views. This results shows that ReNeRF is generalizing well to represent nearfield lighting not seen at training time.

**Decomposing area-OLATs.** Instead of a dense array of distant point lights, our capture setup includes a small number of near area lights. Thus, the ReNeRF model not only needs to account for and interpolate near point lighting, but it also has to decompose the area-OLAT lighting in the training data into point-OLAT lighting. Fig. 7 and the supplementary video show that modeling the lighting of each LED-bar as the contribution of three near point lights is enough to learn a continuous basis of point-OLATs, due to our smooth spherical codebook representation.

**Lighting extrapolation.** Our 32 area lights are spread out as to cover most of (but not all) the frontal hemisphere around the captured scene. Here we evaluate how well a pre-trained ReNeRF model generalizes to novel lighting directions outside this incomplete hemisphere. As a reference for comparison, we also train a modified version of our method, ReNeRF$_{(D)}$, that models distant (directional) lighting, as in [22, 45]. We train both models on our real head dataset of Fig. 4 (right) and, at test time, render a near point light that is initially 20cm in front of the nose, Fig. 8 (left). This light source then gradually rotates about

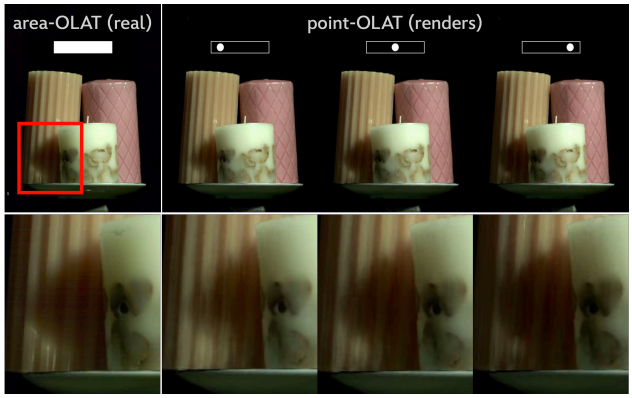


Figure 7. A training image captured under the light of a single LED bar (top left), and the 3 point-OLATs sampled during training (top right) for learning our continuous point-OLAT basis. The insets show the moving shadow under the learned point-OLATs.

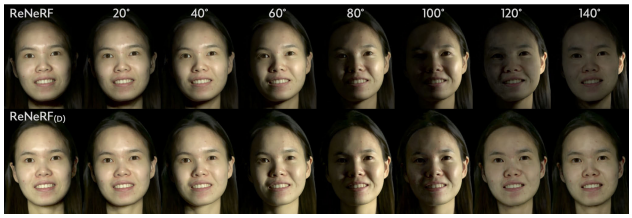


Figure 8. Lighting extrapolation with a near point light rotating to the left of the person, within a horizontal plane.

the face, within a horizontal plane, towards the left side of the person, Fig. 8 (right). As shown in the figure, extrapolation with the distant lighting model ReNeRF$_{(D)}$ breaks down at $80^o$, earlier than $120^o$ with our nearfield model, which sees a bigger variety of directions $\omega_i$ during training.

**Quantitative relighting evaluation.** We now evaluate ReNeRF trained on five real datasets with only half of the 32 area-OLAT sources, holding the images of the other 16 area-OLATs for validation. We compare ReNeRF against the distant lighting methods ReNeRF$_{(D)}$, NeRF-SHL [20], NeRFactor [58] and a simplified NeRF-W [24] with 64D and 128D lighting codes. For each method and dataset, we computed PSNR over the complete validation set with 16 OLATs in 10 views, as shown in Table 1. Since NeRF-W cannot take an environment map for controllable relighting, at test time we had to fit its lighting code to the GT validation images. This is done only for PSNR comparison. ReNeRF achieves comparable PSNR scores with NeRF-W$_{(128D)}$, while no additional test time optimization is needed, and outperforms the other baselines on all datasets. The supplemental file shows that ReNeRF is better than NeRF-W at inter- and extrapolating light codes due to its novel codebook model. Fig. 9 (top) shows relighting errors of ReNeRF and the baselines on example validation OLAT lighting. To compare relighting errors un-

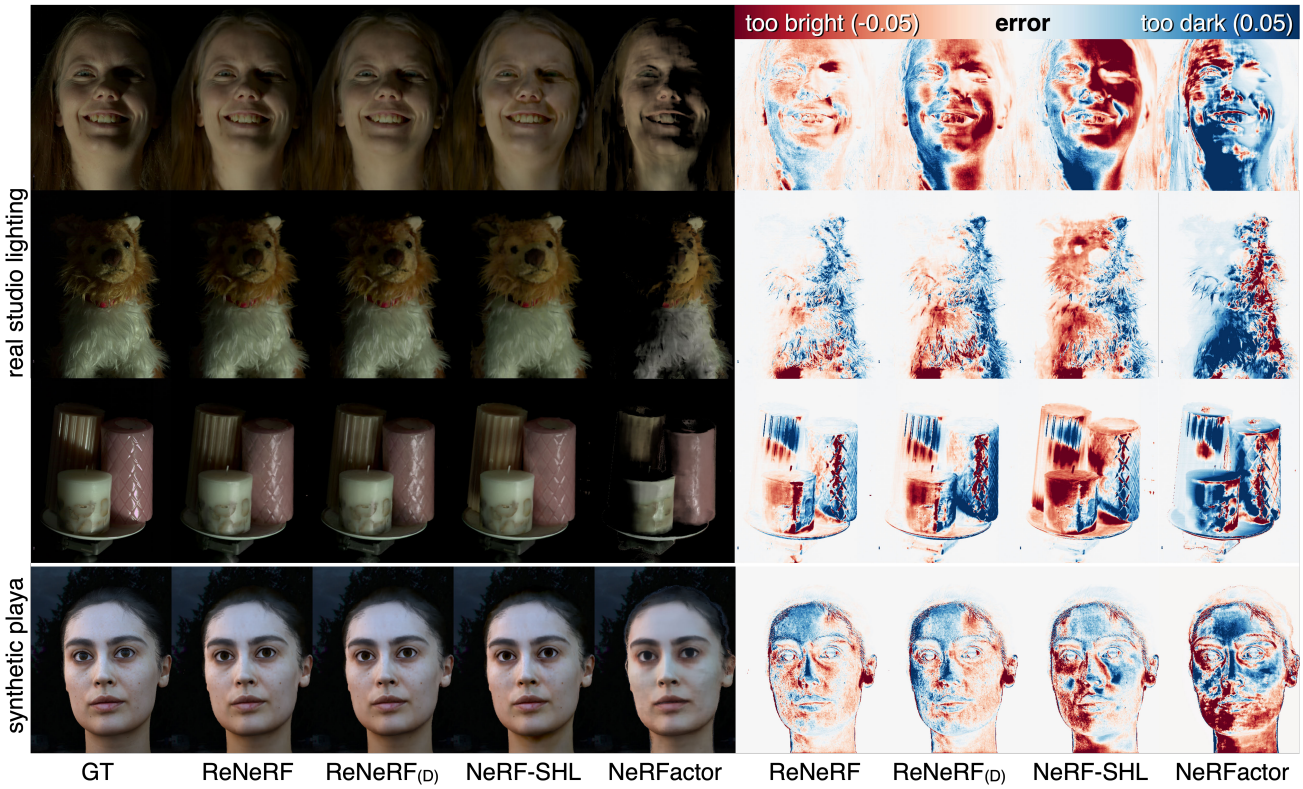


Figure 9. Relighting error on unseen validation images with novel illumination: (top) real datasets under an area light (OLAT), (bottom) synthetic dataset under envmap lighting. ReNeRF provides better relighting (smaller errors) in comparison to the three baseline methods.

der environment lighting (since no GT is captured), we additionally train ReNeRF and the baselines on a synthetic dataset [1] (32 point OLATs, 10 cameras) and results are shown in the last row of Fig. 9. NeRFactor models only direct illumination and heavily relies on surfaces extracted from a pretrained NeRF; it cannot produce convincing renderings. NeRF-SHL often cannot reproduce sharp shadows and tends to render with more uniform (smooth) lighting. Finally, another issue is that many related baselines (*e.g.*, NeRF-SHL, NeRFactor) are parameterized by albedo and its diffuse shading has a neutral color. Thus, they cannot capture changes in light chromaticity as it passes through thin object layers. To illustrate this issue, we trained a modified ReNeRF-albedo model with similar parameterization, and verified that it could not represent light going through the nostril of the person in Fig. 10. The ReNeRF model that we propose here does capture this effect. More details on the experimental setup, baseline implementations, and other comparisons are given in the supplemental material.

## 5. Conclusion

We present ReNeRFs, relightable neural radiance fields inspired by the simple and powerful image-based relighting method. A ReNeRF is trained on multiview images of an object captured under a small number of OLAT area lights, removing the need for dense light stage hardware. The new model is simple and does not require complex light transport simulations nor overly simplistic assumptions about how light interacts with the observed scene. Once trained, a ReNeRF model can render the neural field from novel viewpoints and under novel illumination, with high photorealism and including nearfield lighting. ReNeRF is thus a practi-

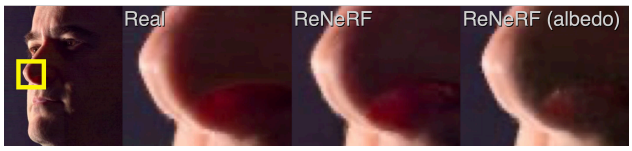


Figure 10. White light passing through a thin layer (nostril) picks up color before exiting. Models parameterized by albedo and neutral-color radiance (right) cannot model this effect.

Table 1. Relighting PSNR on validation OLATs.

| Method | Dog | Candles | $\text{Head}_1$ | $\text{Head}_2$ | $\text{Head}_3$ |
|---|---|---|---|---|---|
| NeRF-SHL | 30.31 | 27.97 | 32.46 | 31.74 | 37.27 |
| NeRFactor | 24.50 | 24.77 | 28.04 | 27.09 | 32.32 |
| NeRF-W$_{(64D)}$ | 30.20 | 29.12 | 32.78 | 33.03 | 38.14 |
| NeRF-W$_{(128D)}$ | 31.63 | 30.14 | 34.40 | 32.73 | 38.14 |
| ReNeRF$_{(D)}$ | 30.64 | 28.94 | 32.98 | 32.21 | 37.87 |
| ReNeRF | 31.17 | 29.49 | 34.04 | 33.04 | 38.14 |

cal approach for capturing high-quality rendering assets to populate virtual environments in different applications.

A limitation of our current implementation is the long training and evaluation time as we need to query the OLAT MLP once per point light (a 512x384 image takes 25s to render with a point light, and 11min with a 144-direction environment map). However, ReNeRF will naturally benefit from faster NeRF architectures [12, 31] in future work. Also, ReNeRF at times renders smoothed specular highlights due to (1) the SH model for specular radiance that handles sparse training views, and (2) the sRGB gamma correction applied in the render loss.

## References


[1] Animatable digital double of louise by eisko©. http://www.eisko.com. 8

[2] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. *CVPR*, 2022. 2

[3] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. High-quality single-shot capture of facial geometry. *ACM TOG*, 29(3), 2010. 3

[4] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM TOG*, 40(4), 2021. 2, 3

[5] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv*, 2020. 2

[6] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 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 16*, pages 294–311. Springer, 2020. 2

[7] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12684–12694, 2021. 2

[8] Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan T Barron, Hendrik Lensch, and Varun Jampani. Samurai: Shape and material from unconstrained real-world arbitrary image collections. *arXiv preprint arXiv:2205.15768*, 2022. 2

[9] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems*, 34:10691–10704, 2021. 2

[10] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, pages 25–36, 2004. 3

[11] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient geometry-aware 3D generative adversarial networks. In *CVPR*, 2022. 2

[12] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXII*, pages 333–350. Springer, 2022. 2, 9

[13] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proc. SIGGRAPH*, page 145–156. ACM, 2000. 2

[14] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 2

[15] Guy Gafni, Justus Thies, Michael Zollhöfer, and Matthias Nießner. Dynamic neural radiance fields for monocular 4d facial avatar reconstruction. In *Proc. CVPR*, pages 8649–8658, June 2021. 2

[16] Stephan J. Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. FastNeRF: High-fidelity neural rendering at 200fps. In *Proc. ICCV*, 2021. 2

[17] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, 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. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM TOG*, 38(6), 2019. 2, 3

[18] Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. HeadNeRF: A real-time NeRF-based parametric head model. In *CVPR*, 2022. 2

[19] Zhengfei Kuang, Kyle Olszewski, Menglei Chai, Zeng Huang, Panos Achlioptas, and Sergey Tulyakov. Neroic: neural rendering of objects from online image collections. *ACM Transactions on Graphics (TOG)*, 41(4):1–12, 2022. 2

[20] Gengyan Li, Abhimitra Meka, Franziska Mueller, Marcel C. Buehler, Otmar Hilliges, and Thabo Beeler. EyeNeRF: A hybrid representation for photorealistic synthesis, animation and relighting of human eyes. *ACM TOG (Proc. Siggraph)*, 41(4), 2022. 2, 5, 7

[21] Hsueh-Ti Derek Liu, Francis Williams, Alec Jacobson, Sanja Fidler, and Or Litany. Learning smooth neural functions via lipschitz regularization. In *ACM SIGGRAPH 2022 Conference Proceedings*. ACM, 2022. 5

[22] Linjie Lyu, Ayush Tewari, Thomas Leimkuehler, Marc Habermann, and Christian Theobalt. Neural radiance transfer fields for relightable novel-view synthesis with global illumination. In *ECCV*, 2022. 2, 4, 7

[23] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spher-

ical gradient illumination. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, EGSR'07, pages 183–194. Eurographics Association, 2007. 6, 7

[24] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR*, 2021. 1, 2, 7

[25] Abhimitra Meka, Christian Haene, Rohit Pandey, Michael Zollhoefer, 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. Deep reflectance fields - high-quality facial reflectance field inference from color gradient illumination. volume 38, 2019. 2, 3

[26] Abhimitra Meka, Rohit Pandey, Christian Haene, Sergio Orts-Escolano, Peter Barnum, Philip Davidson, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, Chloe Legendre, Wan-Chun Ma, Ryan Overbeck, Thabo Beeler, Paul Debevec, Shahram Izadi, Christian Theobalt, Christoph Rhemann, and Sean Fanello. Deep relightable textures - volumetric performance capture with neural rendering. *ACM TOG (Proc. Siggraph Asia)*, 39(6), 2020. 2, 3

[27] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. NeRF in the dark: High dynamic range view synthesis from noisy raw images. *arXiv*, 2021. 2

[28] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 4

[29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Comm. ACM*, 65(1):99–106, 2021. 1

[30] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM TOG*, 41(4), 2022. 2, 6

[31] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 9

[32] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8280–8290, 2022. 2

[33] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas Lehrmann. Learning physics-guided face relighting under directional light. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5124–5133, 2020. 2

[34] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *Proc. CVPR*, 2021. 2

[35] J.S. Nimeroff, E. Simoncelli, and J. Dorsey. Efficient rerendering of naturally illuminated environments. In *Eurographics Workshop on Rendering*, page 359–373, 1994. 2

[36] Michael Oechsle, Songyou Peng, and Andreas Geiger. Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proc. ICCV*, 2021. 2

[37] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Transactions on Graphics (TOG)*, 40(4):1–21, 2021. 2

[38] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. *Proc. ICCV*, 2021. 2

[39] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. HyperNeRF: A higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph.*, 40(6), dec 2021. 2

[40] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann Publishers Inc., 3rd edition, 2016. 3

[41] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *Proc. CVPR*, 2020. 2

[42] Jérémy Riviere, Paulo Gotardo, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. Single-shot high-quality facial geometry and skin appearance capture. *ACM TOG (Proc. Siggraph)*, 39(4), 2020. 6, 7

[43] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Neural radiance fields for outdoor scene relighting. *arXiv preprint arXiv:2112.05140*, 2021. 2

[44] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *CVPR*, 2021. 2

[45] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. Nelf: Neural light-transport field for portrait view synthesis and relighting. In *EGSR*, 2021. 2, 4, 7

[46] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul Srinivasan, Jonathan T. Barron, and Henrik Kretzschmar. Block-NeRF: Scalable large scene neural view synthesis. In *CVPR*, 2022. 1, 2

[47] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-NeRF: Scalable construction of large-scale NeRFs for virtual fly-throughs. In *CVPR*, 2022. 1, 2

[48] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022. 2

[49] Daoye Wang, Prashanth Chandran, Gaspard Zoss, Paulo Bradley, and Paulo Gotardo. MoRF: Morphable radiance

fields for multiview neural head modeling. In *Proc. SIGGRAPH*, 2022. 2, 4

[50] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Proc. NeurIPS*, 2021. 2

[51] Yingyan Xu, Jérémy Riviere, Gaspard Zoss, Prashanth Chandran, Derek Bradley, and Paulo Gotardo. Improved lighting models for facial appearance capture. In *Proc. Eurographics (short paper)*, 2022. 3

[52] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. 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*, pages 266–284. Springer, 2022. 2

[53] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. Neilf: Neural incident light field for physically-based material estimation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*, pages 700–716. Springer, 2022. 2

[54] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. PlenOctrees for real-time rendering of neural radiance fields. In *Proc. ICCV*, 2021. 2, 4

[55] Jason Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. Ners: neural reflectance surfaces for sparse-view 3d reconstruction in the wild. *Advances in Neural Information Processing Systems*, 34:29835–29847, 2021. 2

[56] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5565–5574, 2022. 2

[57] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, Jonathan T. Barron, Ravi Ramamoorthi, and William T. Freeman. Neural light transport for relighting and view synthesis. *ACM TOG*, 40(1), 2021. 2, 3

[58] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. NeRFactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM TOG (Proc. Siggraph Asia)*, 40(6), 2021. 2, 7

[59] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18643–18652, 2022. 2

[60] Quan Zheng, Gurprit Singh, and Hans-Peter Seidel. Neural relightable participating media rendering. *Advances in Neural Information Processing Systems*, 34:15203–15215, 2021. 2