

Relit-NeuLF: Efficient Relighting and Novel View Synthesis via Neural 4D Light Field

Zhong Li*
OPPO US Research Center, InnoPeak
Technology, Inc
Palo Alto, CA, USA

Liangchen Song
State University of New York at
Buffalo
Buffalo, NY, USA

Zhang Chen
OPPO US Research Center, InnoPeak
Technology, Inc
Palo Alto, CA, USA

Xiangyu Du
OPPO US Research Center, InnoPeak
Technology, Inc
Palo Alto, CA, USA

Lele Chen
OPPO US Research Center, InnoPeak
Technology, Inc
Palo Alto, CA, USA

Junsong Yuan
State University of New York at
Buffalo
Buffalo, NY, USA

Yi Xu
OPPO US Research Center, InnoPeak
Technology, Inc
Palo Alto, CA, USA

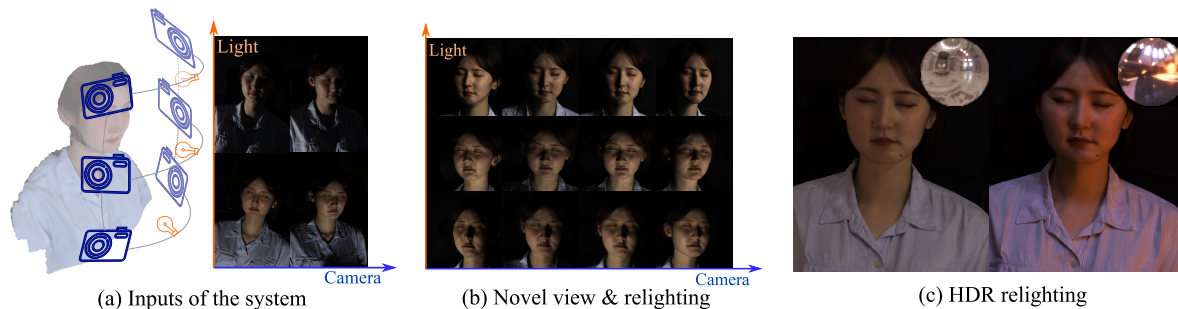


Figure 1: The proposed method takes (a) sparsely sampled multi-view images under different directional lights as input. After modeling the scene with the proposed Relit-NeuLF, our system can perform simultaneous novel view synthesis and relight with both (b) novel directional lights and (c) HDR light probes.

ABSTRACT

In this paper, we address the problem of simultaneous relighting and novel view synthesis of a complex scene from multi-view images with a limited number of light sources. We propose an analysis-synthesis approach called Relit-NeuLF. Following the recent neural 4D light field network (NeuLF) [22], Relit-NeuLF first leverages a two-plane light field representation to parameterize each ray in a 4D coordinate system, enabling efficient learning and inference. Then, we recover the spatially-varying bidirectional reflectance distribution function (SVBRDF) of a 3D scene in a self-supervised manner. A DecomposeNet learns to map each ray to its SVBRDF components:

*Corresponding author. Email: zhong.li@oppo.com

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

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

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0108-5/23/10...\$15.00

<https://doi.org/10.1145/3581783.3612160>

albedo, normal, and roughness. Based on the decomposed BRDF components and conditioning light directions, a RenderNet learns to synthesize the color of the ray. To self-supervise the SVBRDF decomposition, we encourage the predicted ray color to be close to the physically-based rendering result using the microfacet model. Comprehensive experiments demonstrate that the proposed method is efficient and effective on both synthetic data and real-world human face data, and outperforms the state-of-the-art results.

CCS CONCEPTS

• **Computing methodologies** → **Rendering**; **Computer vision problems**; **Virtual reality**.

KEYWORDS

Light Fields, 3D Deep Learning, View Synthesis, Image-based Rendering

ACM Reference Format:

Zhong Li, Liangchen Song, Zhang Chen, Xiangyu Du, Lele Chen, Junsong Yuan, and Yi Xu. 2023. Relit-NeuLF: Efficient Relighting and Novel View Synthesis via Neural 4D Light Field. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*, October 29–November 3,

2023, Ottawa, ON, Canada. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3581783.3612160>

1 INTRODUCTION

Novel view synthesis of real-world objects under different lighting conditions has been an important research topic for computer vision and computer graphics. A system that can perform simultaneous relighting and free-viewpoint rendering, thereby integrating a real-world object into a digital environment, has numerous commercial applications, especially when combined with real-time light estimation techniques [14, 24].

Traditional methods [8, 18] solve the free-viewpoint rendering problem based on the geometric reconstruction pipeline, which assumes the objects are mostly Lambertian. Since these methods do not fully model the complex surface reflection properties, the rendered results may be unsatisfactory. Taking human faces as an example, these methods do not model the spatially varying reflectance characteristics, such as diffuse and specular reflections.

To overcome these limitations, Light Stage [11] are used to capture the reflectance field of objects. With synchronized LED lights and cameras, a “One-Light-at-A-Time” (OLAT) image set can be obtained. Then, an input environment map can be decomposed into a linear combination of these OLAT images to relight the object in a photorealistic manner. This approach has a profound impact on the visual effects industry and inspires numerous follow-up works. To accurately capture the reflectance field, many light sources are required. However, due to physical constraints, only about one hundred light sources are usually installed on a light stage. Therefore, the system cannot produce sharp shadows and high-frequency details due to under-sampling. To solve this problem, Sun *et al.* [45] propose a learning-based method to perform high-precision resampling of the under-sampled OLAT image set. However, simultaneous novel view synthesis during relighting remains unsolved.

For image-based novel view synthesis, great progress has been made with the neural radiance field [29] (NeRF), based on a volumetric representation. One of the limitations of NeRF is that the color and density of sample points along each ray must be integrated. Therefore, rendering with NeRF is time consuming. A few works [12, 17, 39, 55] have attempted to accelerate NeRF by either using much more memory or a geometric *prior*. While there has been substantial progress in novel view synthesis, relatively less effort has been spent in free-viewpoint rendering with relighting.

In this paper, given sparsely sampled images illuminated by sparse LED light sources (as shown in Fig. 1), we present a two-stage analysis-synthesis MLP neural network for free-viewpoint rendering and relighting. Our model takes the camera ray direction and light direction as the input. We first analytically recover reliable normal, albedo, and roughness in a self-supervised manner (analysis stage), and then simulate the process of physically-based rendering to output the ray color (synthesis stage).

Motivated by the recent work NeuLF [22], we use two-plane representation to parameterize a ray in space into a 4D coordinate. The analysis stage involves a *DecomposeNet*, whose inputs are rays and the outputs are the decomposed surface normal, albedo, and roughness. The synthesis stage is a *RenderNet*. With the predicted SVBRDF components from the *DecomposeNet* and a light direction,

RenderNet is trained to generate the color for the input ray. The entire network is trained end-to-end and is supervised with both photometric loss and an SVBRDF rendering loss (using a microfacet model, *e.g.*, [20]). The SVBRDF parameters are self-supervised. We found that the SVBRDF decomposition of a ray makes it easier for the network to learn the implicit rendering process.

Compared to NeRF’s volumetric representation [29], our method utilizes a 4D light field representation [22]. By directly mapping a ray and a light direction to a color value, we achieve efficient inference and low memory consumption. While this approach limits rendering to front views, it suffices for most free-viewpoint relighting applications, such as remote 3D video calls with changing backgrounds. Our contributions are:

- We propose a two-stage analysis-synthesis MLP network for efficient relighting and novel view synthesis, which maps ray and light direction to a color value directly.
- Our proposed *DecomposeNet* can recover the SVBRDF (surface normal, albedo, and roughness) of a 3D scene in a self-supervised manner using a set of camera viewpoints and a limited number of light sources as inputs.
- We created a multi-identity multi-view OLAT dataset of real human faces and a synthetic multi-view multi-lighting dataset. We plan to release our dataset to facilitate future research on relighting and novel view synthesis.

2 RELATED WORKS

Our method simultaneously synthesizes novel views and upsamples the sparse light sources. These two problems subsume many computer vision and graphics-related tasks and have been key topics for decades. Below we briefly review the related work on light angular up-sampling and novel view synthesis, as well as recent efforts on these using neural rendering.

Light angular upsampling: Traditional methods upsample discrete light source input using geometric modeling, a technique distinct from light stage data upsampling [11]. Such methods undertake geometric and photometric stereo reconstruction of objects, revealing albedo, normal, and specular normal [26, 50, 51], and subsequently employ predefined BRDFs for different lighting conditions. Tunwattanapong *et al.* employ parametric models [47]. Deep learning adaptations, distinct from traditional light stage data, have been applied for single-image human relighting [19, 31, 43]. Shu *et al.* presented a human face illumination transfer, with some limitations in compatibility [35]. Challenges persist with predefined BRDFs, especially when interpreting complex facial features. Methods from both Masselus *et al.* [27] and Rainer *et al.* [32, 33] seek to address these challenges, but results may lack detail. Sengupta *et al.* highlight challenges in expressing objects with complex materials, pointing out issues with predefined analytic BRDFs [34]. Xu *et al.* introduced a neural network for image-based relighting, although shadow accuracy remains problematic [52]. Our approach accentuates shadow accuracy and includes novel view synthesis. Lastly, Sun *et al.* [45] provided a method for relighting faces, albeit without supporting novel view synthesis.

Novel view synthesis: For generating novel views from a set of images, traditionally, multi-view stereo algorithms [13, 61] can

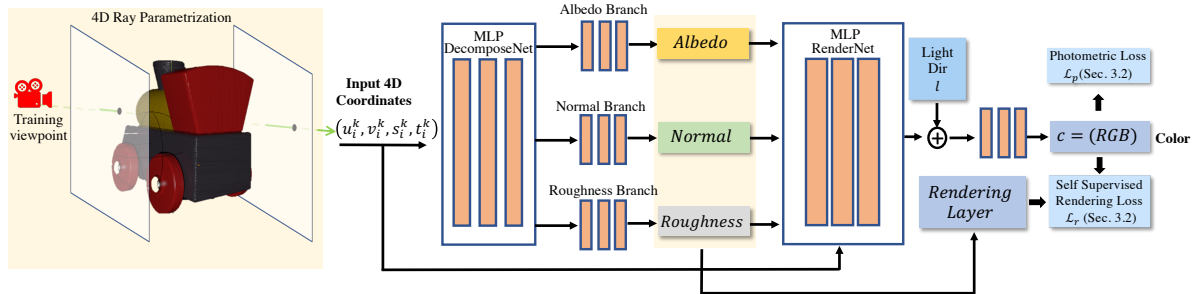


Figure 2: An overview of our proposed Relit-NeuLF. The input is the 4D coordinate of a ray and a light direction. The output is the RGB radiance of the ray under the light direction. Our DecomposeNet first takes the 4D coordinate as input and outputs SVBRDF parameters. Next, the SVBRDF parameters together with the light direction are fed into an implicit rendering network (RenderNet) to synthesize the target color. The network is trained end-to-end with photometric loss and self-supervised rendering loss.

be used to reconstruct a textured mesh. To achieve higher photo realism, this approach requires more vertices and high-resolution texture maps. It does not model view-dependent effects either. Image-based rendering, such as light field [21] and lummigraph [15], can render view-dependent effects by reducing the 5D radiance field in space to a 4D light field, but the quality of the novel view synthesis is affected by the sampling density of the input [7]. Subsequent works use geometry proxy for better view interpolation [6, 49], but the results also depend on the accuracy of the proxy and are sensitive to occlusions. Recently, deep learning has been applied to generate novel views from sparse images. Neural networks are typically used to generate an implicit intermediate representation, and a rendering algorithm is developed based on the representation [28, 37, 38, 46, 64]. A recent seminal work NeRF [29, 48] generate unprecedented results. It uses an MLP network to regress 5D coordinates (location in space and view direction) to the view-dependent radiance and density volume. Nevertheless, it requires aggregating the samples along each ray during rendering, which is time-consuming. Although there are recent solutions to accelerate NeRF [12, 17, 55], they typically need to consume more memory. On the other hand, neural light field approaches [1, 22, 36, 40, 42, 44] use a neural network to directly establish the mapping between ray direction and color; therefore, no complicated ray marching is needed. These methods improve the rendering speed while maintaining low memory consumption. But none of these methods model environment lighting and thus cannot relight the scene under unseen illumination.

Simultaneous free-viewpoint rendering and relighting:

To render novel views with relighting effects simultaneously, an effective method is to sample sufficient reflection fields from multiple viewpoints and under multi-directional point lights and then upsample across viewpoints and illumination. A pioneer work of Debevec *et al.* [11] built a light stage to collect One-Light-at-A-Time (OLAT) data, and then linearly combine reflectance fields to relight the subject. Although the light stage can collect reflectance fields from hundreds of light directions, the sampling rate is still limited. It can only be applied to low-resolution environment maps for re-lighting, so high-frequency lighting effects are not supported.

Moreover, the light stage can only perform relighting in a few fixed viewing angles.

A recent work models texture maps to achieve relighting effects with known subject geometry [59]. Zhang *et al.* and Boss *et al.* propose methods to decompose the shape and reflectance of the subject to achieve new perspective rendering and relighting without known lighting conditions [2–5, 30, 54, 57, 58, 60, 62]. However, both methods can only leverage a relatively low-resolution environment lighting map, hence the results have no high-frequency details. NeRV [41] and PS-NeRF [53] improve on NeRF and decomposes shape and reflectance in the case of known multi-illumination. Our method is based on the light field representation [22] and proposes a two-step network for simultaneous novel view synthesis and relighting. With internet image collections, Zhang *et al.* [56] reconstruct both the geometry and material appearance of the object, however, it can only model the genus-zero topology. Our method does not require a geometry proxy for high-frequency relighting and achieves fast free-viewpoint rendering with less memory consumption.

3 PROPOSED METHOD

In this paper, we present a simple yet efficient framework that can generate novel views under novel directional, given sparse multi-view One-Light-at-A-Time (OLAT) images as input. To achieve this, we extend a neural 4D light field representation NeuLF [22] (Section 3.1) with relighting capability by using SVBRDF decomposition. Given the set of (ray, color) pairs from known views, to guide the network with physically based rendering, we first design a DecomposeNet that disentangles each ray’s intrinsic SVBRDF components (albedo, normal, and roughness). We then use a RenderNet to use albedo, normal, roughness, and ray direction as input with light direction as conditioning to regress ray color (Section 3.2). Fig. 2 shows our pipeline.

3.1 NeuLF Overview

We review the NeuLF representation in this section. As the radiance along a ray is constant if viewing from outside the convex hull of the scene, a 3D scene can be represented as a 4D light field using

two plane parameterization [21]. Each ray r from a camera with known camera poses will intersect with both the uv and the st plane and can be uniquely represented using the coordinate of the intersections $r = \{u, v, s, t\}$.

NeuLF [22] formulates a mapping function f_c with a multi-layer perceptron (MLP). Given a set of N calibrated input images $\{I_1, I_2, \dots, I_n\}$, it extracts the 4D ray parameterization of each pixel and its corresponding color: $(u_i^k, v_i^k, s_i^k, t_i^k) \rightarrow c_i^k, (k = 1 \dots N, i = 1 \dots N_k)$, where c_i^k is the color of the i -th pixel in the k -th image. N_k is the total number of pixels in the k -th image. The goal of the network f_c is to learn the mapping from the 4D coordinate of a ray to the corresponding color value. The MLP network parameters Θ can be learned by minimizing the following photometric loss \mathcal{L}_p :

$$\mathcal{L}_p = \sum_{k=1}^M \sum_{i=1}^{N_k} \left\| f_c(u_i^k, v_i^k, s_i^k, t_i^k | \Theta) - c_i^k \right\|_2. \quad (1)$$

To render a novel view, NeuLF simply queries the network for each ray in the novel view.

3.2 Relit-NeuLF

The NeuLF representation does not consider the changes of incident light direction. In other words, NeuLF is only able to represent a scene under a fixed baked-in illumination. Given a set of OLAT input images $\{I_i^l, i = 1, \dots, N, l = 1, \dots, L\}$, we have N different viewpoints and each viewpoint is illuminated by L different directional lights. In total, we have $N \times L$ images. To perform lighting interpolation and novel view synthesis simultaneously, one straightforward approach is to extend NeuLF to take lighting direction as input. Instead of only mapping 4D coordinate to color value as in Eq. 1, we can train a mapping $f_c^l((u_k, v_k, s_k, t_k), l) \rightarrow c_k^l$ for each input view i . With this, we can simulate the light transport (LT) of a given 4D coordinate ray, *i.e.*, how color value changes with varying lighting direction.

However, this vanilla method simply regresses a light-conditioned ray to a color value. Without implicitly modeling the scene representation, synthesis results obtained by such a method lacks high-frequency detail and are unable to recover complex reflection and refraction effects (see Sec. 4.4 for details). We hence propose a two-stage network designed with principles inspired by physically-based rendering techniques. In the first stage, we create a DecomposeNet that analyzes a ray r 's SVBRDF components, including normal, albedo, and roughness. Then, these SVBRDF components are fed into a second network named RenderNet to learn the inverse rendering function. The SVBRDF components can be recovered without ground truth by self-supervision. We will describe each network in detail.

DecomposeNet: Given the input $N \times L$ images (Fig. 4 shows sample images), we build a MLP network that takes 4D coordinate $r = \{u, v, s, t\}$ for each pixel as input, and outputs SVBRDF parameters: surface normal (N), diffuse albedo (A) and specular roughness (R). Considering that the SVBRDF components are closely correlated, the DecomposeNet first extracts a shared feature and then uses three decoders for three different components. Compared with using a completely separate branch for each SVBRDF component, this shared structure improves efficiency and reduces the risk of

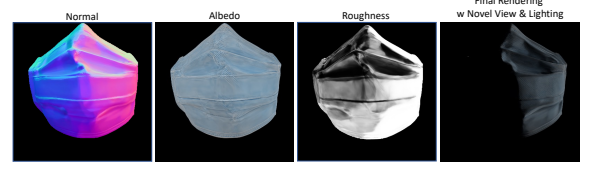


Figure 3: Our SVBRDF decomposition results. From left to right: surface normal map, surface albedo, surface roughness, and rendering results under novel viewpoint and light direction.

over-fitting. It is also worth nothing that our method can recover SVBRDF components in a self-supervised fashion without knowing the ground truth. We will illustrate our self-supervised loss later in the following paragraph. The network DecomposeNet(\cdot) is represented as:

$$N, A, R = \text{DecomposeNet}(r | \Theta_d), \quad (2)$$

where $r \in \{I_i^l, i = 1, \dots, N, l = 1, \dots, L\}$.

RenderNet: Prior works leverage predefined SVBRDF components to analytically render the scene [23]. However, using predefined SVBRDF (*e.g.*, microfacet renderer [20]) is not a universal rendering solution. It cannot model complex objects and materials at the same level of fidelity as real photos, such as human faces' subsurface scattering and specularly distribution. Failing to model them will compromise the photorealism of the rendering results. We propose to train a network to implicitly model the real-world rendering process, using the microfacet renderer and ground truth for weak supervision. The RenderNet takes the estimated SVBRDF components (albedo (A), normal (N), and roughness (R)) as inputs. To render a ray under specific light direction l , we train a fully connected network:

$$c_{pred} = \text{RenderNet}(N, A, R, r, l | \Theta_r), \quad (3)$$

where $r \in \{I_i^l, i = 1, \dots, N, l = 1, \dots, L\}$.

Loss function: We train both **DecomposeNet** and **RenderNet** in an end-to-end manner. For each ray in camera view i and light direction l , we can compute the predicted color c_{pred} as follows:

$$c_{pred} = \text{RenderNet}(\text{DecomposeNet}(r | \Theta_d), r, l | \Theta_r). \quad (4)$$

During training, we leverage two data loss terms \mathcal{L}_p and \mathcal{L}_m , and one normalization term \mathcal{L}_n . We use \mathcal{L}_p to minimize the multi-view photometric loss. To self-supervise the SVBRDF decomposition, we use \mathcal{L}_m to enforce the network output to be close to the physically-based microfacet rendering result. We choose the microfacet model [20] instead of the Phong model because the microfacet model can produce more realistic rendering results. To regularize the normal vector to be physically plausible, we use \mathcal{L}_n to encourage the normal N to have a norm of 1.

$$\mathcal{L}_p = \|c_{pred} - c\|_2, \quad (5)$$

$$\mathcal{L}_m = \|c_{pred} - M(N, A, R, v, l)\|_2, \quad (6)$$

$$\mathcal{L}_n = \left\| 1 - N^T N \right\|_2, \quad (7)$$

where M is the microfacet BRDF rendering model [20], A is the diffuse albedo, and v, l denote the view direction and light direction respectively. Let $h = \frac{v+l}{\|v+l\|}$ be the half vector, then we have

$$M(N, A, R, v, l) = \frac{A}{\pi} + \frac{D(h, N, R)F(v, h)G(l, v, N, R)}{4(N \cdot l)(N \cdot v)}, \quad (8)$$

where $D(h, N, R) = \frac{a^2}{\pi((N \cdot h)^2(a^2-1)+1)^2}$ is normal distribution function (NDF), $F(v, h) = F_0 + (1 - F_0)2^{-5.55473(v \cdot h) - 6.98316(v \cdot h)}$ is Fresnel term [25], and $G(l, v, N, R) = \frac{N \cdot v}{(N \cdot v)(1-k)+k} \cdot \frac{N \cdot l}{(N \cdot l)(1-k)+k}$ is geometry term. The extra variables in the equations are $F_0 = 0.05$, $a = R^2$, and $k = \frac{(R+1)^2}{8}$.

Fig. 3 shows an example of our decomposition and rendering results. It can be seen that **DecomposeNet** successfully decomposes the BRDF components with the self-supervision of the microfacet model, and can generate rendering results under unseen light direction.

3.3 Implementing Details

We train the networks using a combination of the photometric loss, microfacet rendering loss, and normal regularization loss:

$$\mathcal{L} = \lambda_m \mathcal{L}_m + \lambda_p \mathcal{L}_p + \lambda_n \mathcal{L}_n. \quad (9)$$

The weight coefficients for the three losses are $\lambda_m = 0.1$, $\lambda_p = 1$ and $\lambda_n = 0.01$. We conduct experiments on both synthetic data and real captured data. For real data, to estimate camera poses, we adopt [63] and use the planar pattern to calibrate the intrinsic and extrinsic parameters of each camera. To calibrate the directions of light sources, we capture images of a chrome ball under all the lighting conditions and use the brightest spot to identify the direction of each light.

We normalize the input 4D coordinate (u, v, s, t) to the range $[-1, 1]$, and then go through the 8 layers of fully connected ReLU in DecomposeNet to obtain the 256-dimensional shared feature vector. Then the feature vector is split into three SVBRDF parameter branches: N (Normal), A (Albedo), R (Roughness), each of which contains 1 fully-connected layer. The Normal and Albedo branches have 3-dimensional outputs and use ReLU activation, while the Roughness branch has a 1-dimensional output and uses sigmoid activation. Then we concatenate N, A, R with ray coordinate $\{u, v, s, t\}$ as the input to the RenderNet, which uses 8 layers of fully connected and ReLU to encode its input into a 128-dimensional feature vector. The feature vector is concatenated with 3-dimensional light direction l , then goes through two fully-connected layers with sigmoid activation, and finally outputs 3-channel RGB radiance. In terms of training details, we set the batch size of each iteration to 8192 and use Adam optimizer [9]. The initial learning rate is set to 3×10^{-4} and decays 0.995 per epoch. In the real data experiment, we use an input of 13 camera views with 500×750 resolution and 40 LED light sources. The training takes 5 hours to complete on 1 NVIDIA RTX 3090 graphics card, while the inference takes 150ms per frame under the same resolution.

4 EXPERIMENT

To demonstrate the effectiveness of the proposed framework for high-quality relighting and novel view synthesis, we conduct experiments on both rendered synthetic data and captured real-world datasets. We first discuss the acquisition of real data and synthetic data in Section 4.1. Then we show both quantitative and qualitative results of novel view synthesis and relighting in Section 4.2. We report rendering speed and quality with comparison to an extended NeRF-based method in Section 4.3. We also show our SVBRDF decomposition results and perform an ablation study to illustrate its importance in Section 4.4. We refer readers to our supplemental video, which showcases our results on dynamic relighting and view synthesis.

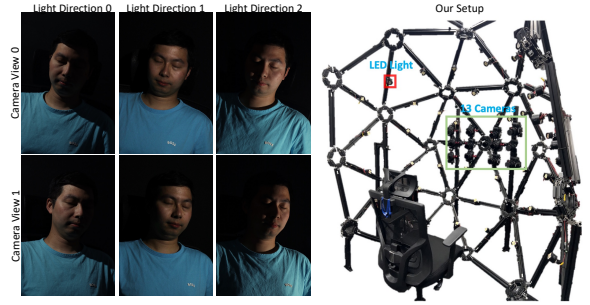


Figure 4: Our multi-camera & multi-light capture system, with sample captured images. The system contains 13 synchronized DSLR cameras and 40 LED lights. We use an Arduino microcontroller to control the cameras and lights.

4.1 Real and Synthetic Data Setup

Real data acquisition Obtaining multi-view images under different lighting directions is not an easy task. To the best of our knowledge, there is no such dataset publicly available. Similar to [16], we built a hemisphere multi-view OLAT (one light at a time) image acquisition system. As shown in Fig. 4, our system consists of 13 synchronized Canon 850D DSLR cameras and 40 LED light sources. We have shown 4 subjects in our experiments. For each subject, we acquire 13×40 OLAT images. These OLAT images sparsely collected the reflectance data on the front hemisphere of the subject.

During data collection, we require all subjects to remain approximately still. It takes 5 seconds to complete the OLAT acquisition of 13 viewing angles and 40 LED light sources. From the collected data of each subject, we randomly selected 1 out of 13 views and 3 out of 40 light conditions for testing, and use the rest for model training.

Synthetic data rendering Additionally, we used Blender’s physically based path tracer renderer and rendered 3 textured objects: synthetic face, wood train, and face mask. We set up 5×5 camera views on the front hemisphere, set 105 directional light sources around the full sphere, and render at a resolution of 800×800 pixels. Each camera differs by 10 degrees and each light source differs by 25 degrees on the sphere. We randomly select 2 out of 25 views and 6 out of 105 lights for testing, and the rest for model training.

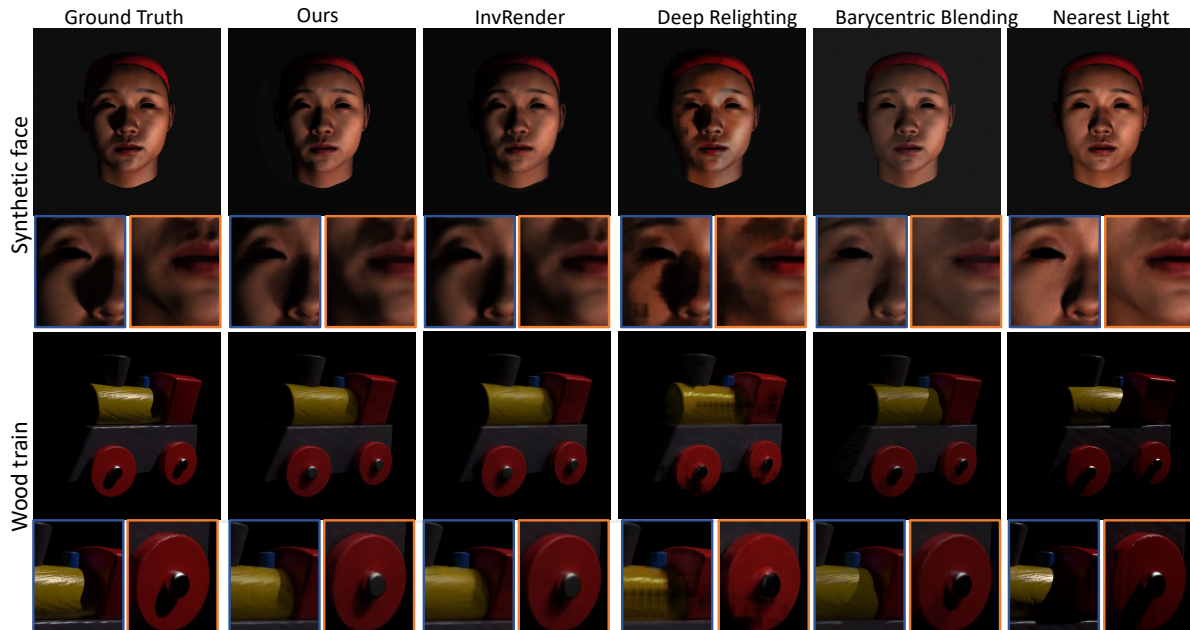


Figure 5: Qualitative results of relighting with an unseen light direction under fixed viewpoints. From left to right: Ground truth, our method, InvRender [62], Deep Relighting [52], barycentric blending, and nearest light image.

Table 1: Metrics on relighting (averaged across test lights) for each scene in the Synthetic/Real dataset.

	PSNR \uparrow					SSIM \uparrow					LPIPS \downarrow				
	Nearest	Barycentric	Deep Relighting	InvRender	Ours	Nearest	Barycentric	Deep Relighting	InvRender	Ours	Nearest	Barycentric	Deep Relighting	InvRender	Ours
Toy train	32.79	33.60	33.43	33.82	34.40	0.914	0.918	0.879	0.919	0.930	0.078	0.082	0.084	0.086	0.075
Face Cover	28.03	28.16	27.34	27.96	28.79	0.793	0.801	0.763	0.804	0.810	0.150	0.163	0.203	0.160	0.130
Face	32.65	32.48	31.42	32.88	33.14	0.820	0.808	0.754	0.827	0.833	0.071	0.062	0.068	0.081	0.075
Real	27.49	27.07	27.23	27.86	28.33	0.732	0.745	0.708	0.747	0.771	0.143	0.162	0.155	0.141	0.132

4.2 Relighting and View Synthesis Results

Relighting Our Relit-NeuLF model can generate rendering results under novel viewpoints and novel lighting directions. As shown in Fig. 5, we show qualitative relighting results for different synthetic data. We compared our method with two state-of-the-art solutions: InvRender [62], which generates novel view synthesis and relighting under unknown illumination, and Deep relighting [52], which presents an image-based relighting method using sparse predefined directional lights. In addition, we implement two baseline solutions for comparison: barycentric blending and nearest light in the training set. Compared to the hold-out GT validation image, our results surpass the other methods in terms of shadow shape and clarity (e.g., the nose shadows in the zoom-in figures of the first row). Although the barycentric blending and nearest light methods can maintain a high-frequency appearance, their shadows have largely incorrect coverage. This demonstrates that these methods, although widely adopted for relighting, have high requirements on the denseness of input lighting directions and are not suitable when the input is sparse. Compared with Deep Relighting, our method produces more accurate shadow shapes, more natural specular highlights, and much fewer appearance artifacts, leading to more realistic results. While InvRender produces relatively good

results, the outputs are slightly blurry and exhibit less accurate shadows, potentially due to the method’s inability to generalize to single point light illuminations.

For quantitative evaluation, we use three metrics: PSNR (Peak Signal-to-Noise Ratio, higher is better), SSIM (Structural Similarity Index Measure, higher is better), and LPIPS (Learned Perceptual Image Patch Similarity, lower is better). The first three rows of Table. 1 shows the metrics on the synthetic data. Our method consistently outperforms the other three with noticeable margins. The last row compares the metrics on real captured human faces, where our method also has the highest performance.

HDRI relighting As shown in Fig. 7, we demonstrate the ability of our method to synthesize visually pleasing relighting under arbitrary HDRI environment maps. Because our method can recover the reflectance under novel lighting directions with a high angular resolution, we can relight the object by treating an HDRI environment map as a collection of OLAT lighting conditions. Unlike ordinary light stage which can only sample hundreds of OLAT images, we use our trained model to generate 3096 OLAT images, each corresponding to a different lighting direction. Then these OLAT images are linearly superimposed to produce the final relighting result, where the linear combination weights are computed based on the

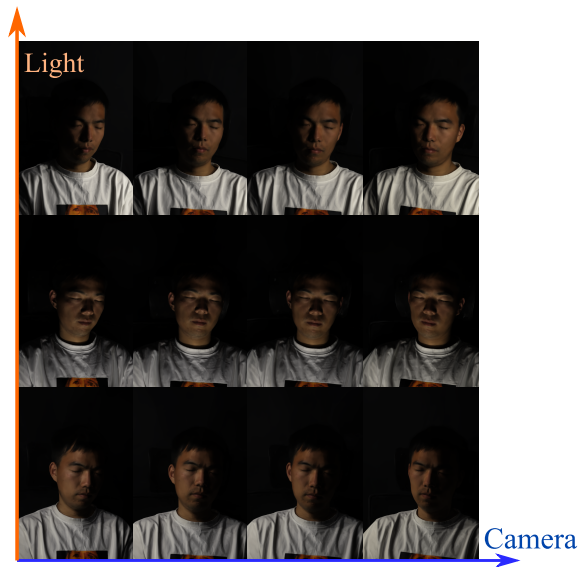


Figure 6: Qualitative results on simultaneous novel view synthesis and relighting. We show results on real data. The x-axis is different novel viewpoints while the y-axis is different lighting directions



Figure 7: Our HDRI relighting results. Each subject is relit by two different environment maps. Note that the subject’s appearance correctly reflects the dominant light source in the environment map.

HDRI environment map. We have shown four HDRI relighting results across different human identities using the light probe images from [10]. Note that for each row in Fig. 7, we rotate the light probe images by different angles. Our relighting results are physically natural, and the specular highlights on the faces correctly reflect the color and position of the highlights in the light probe images. Meanwhile, the cast shadows are also visually plausible. Please see our supplemental video for more dynamic relighting results.

Simultaneous view synthesis and relighting As shown in Fig. 6, we qualitatively show our rendering results of two real

human subjects under simultaneous novel view synthesis and relighting. In the figure, the x-axis is different novel viewpoints while the y-axis is different lighting directions. The lighting effects (e.g., shadows and diffuse reflection) are consistent across different viewpoints. At the same time, high-frequency details (e.g., specular highlights on the foreheads) are successfully recovered and naturally move in accordance with viewpoint and lighting direction. We reason that this is because our RenderNet implicitly learns the light transport properties of the scene with the help of self-supervised SVBRDF components from our DecomposeNet.

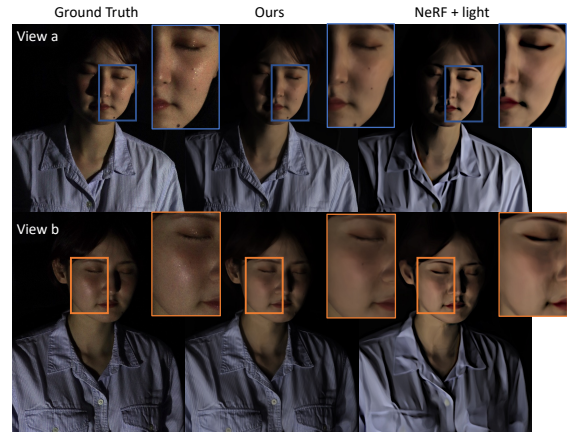


Figure 8: Qualitative comparison on simultaneous relighting and view synthesis. For each row, we showed a synthesis view with unseen lighting from ground truth, ours, and NeRF+light. Our results exhibit a more natural looking and are less blurry than NeRF+light.

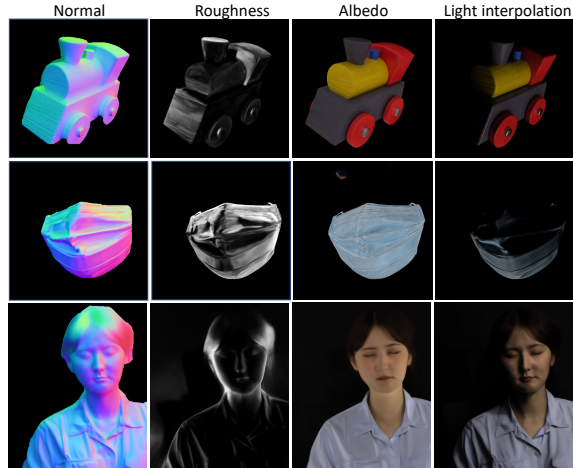
For recent NeRF-based methods such as NeRFactor [60], InvRender [62], NeRD [2] and Neural-PIL [4], although they target for simultaneous novel view synthesis and relighting same as our approach, their inputs have fixed illumination which is different from our method. Therefore, for a fair comparison, we extend the volumetric rendering method NeRF [29] with relighting capability, denoted as “NeRF+light”. To implement NeRF+light, we condition the NeRF’s output radiance on the lighting direction. As shown in Fig. 8, we compare our method with NeRF+light on two novel views with novel light directions. It can be seen that our method produces more natural-looking and less blurry synthesis results than NeRF+light (e.g., the wrinkles on the cloth, and the facial details in the close-up insets). This is because our SVBRDF decomposition module enables the model to better understand the scene and maintain geometry/texture/material details under viewpoint/lighting variation.

4.3 Speed Analysis

Our pipeline is based on the light field based neural networks [22, 36]. Compared with the volumetric NeRF-based method, we directly regress ray color rather than aggregate radiance with time-consuming ray marching. Hence, as shown in Table. 2, our method is significantly faster than NeRF+light, while still having higher rendering PSNR.

Table 2: Comparison of rendering speed and quality with NeRF+light. We compute average PSNR across all real data.

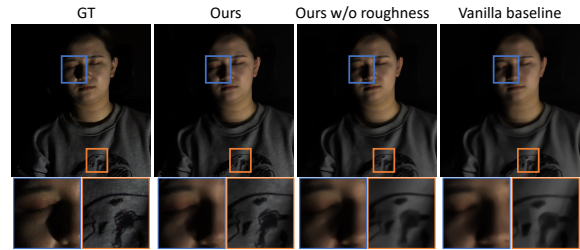
method	Input	Viewpoint range	second/frame	PSNR \uparrow
Ours	4D + light	front views	0.151	28.33
NeRF+light	5D + light	360°	22.143	26.12

**Figure 9: Qualitative results on SVBRDF decomposition. From left to right: surface normal map, surface roughness, surface albedo, and our rendering under a novel light direction.**

4.4 SVBRDF Decomposition Results

Fig. 9 shows the decomposition results on the synthetic toy train, face cover, and real human face scene. The results show that our DecomposeNet can generate reasonable view-independent normal, albedo, roughness field by using self-supervised microfacet rendering loss. We can see that in these three subjects, the estimated normal map is in line with the object geometry and contains shape details such as the wrinkles on the mask and cloth, and the bumpy surface on the wood train. The estimated roughness map plausibly represents the roughness of the object’s surface. And the estimated albedo restores the intrinsic color of the objects. Although the albedo still contains a small amount of baked-in lighting effects due to inter-reflection, it has little impact on our final rendering quality. Rather, with such reasonable decomposition results, our RenderNet can already learn a good balance between the microfacet model and out-of-model effects. The last column of Fig. 9 is the relighting result given a novel lighting direction. The results are realistic, where the shadows correctly depict the geometry and lighting.

Ablation study on SVBRDF decomposition To demonstrate the importance of the SVBRDF decomposition from DecomposeNet, we conduct the following ablation study. We compared our full method with the vanilla baseline method described in Sec 3.2 (our baseline) and a baseline method that only decomposes normal and albedo without considering roughness (our w/o roughness). This ablation is conducted on both real and synthetic datasets, where Table 3 shows the average metrics. Fig. 10 shows the novel view relighting results on an example data. In particular, compared with

**Figure 10: Ablation study on SVBRDF Decomposition. From left to right: ground truth, ours, ours w/o roughness, and our baseline. The bottom row shows close-up views for easier comparison.**

ground truth, we observe that ours w/o roughness and our baseline exhibit more over-smoothing effects than ours, such as the highlight details on the face and the pattern on the cloth. Our full solution outperforms the other two baselines both quantitatively and qualitatively. This is because, with SVBRDF parameters as a prior, our RenderNet implicitly learns a more physically-correct rendering process.

Table 3: Ablation study on SVBRDF Decomposition. The metrics are averaged over the test views and light directions in both real and synthetic datasets.

method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Ours baseline	30.88	0.921	0.067
Ours w/o roughness	31.15	0.917	0.042
Ours	31.84	0.932	0.038

5 CONCLUSIONS

In summary, we have presented Relit-NeuLF, a two-stage MLP network that simultaneously achieves highly-detailed relighting and novel view synthesis with weak supervision. Our method employs DecomposeNet to decompose SVBRDF components with self-supervision, based on sparse camera views and limited light directions. By utilizing the decomposed SVBRDF components, our implicit RenderNet generates high-fidelity novel view synthesis and relighting results, trained end-to-end with self-supervised rendering loss and photometric loss. Our approach achieves high rendering quality with fast rendering speed and low memory cost, thanks to the 4D light field representation. Although our method has limitations, such as requiring hard-to-acquire inputs and being unable to perform non-frontal view synthesis, we suggest future work to extend our method with a more flexible light field representation, explore light field network-based relighting with unconstrained input illumination, and investigate alternative spatial or frequency embedding approaches that generalize well in light field representation.

6 ACKNOWLEDGEMENT

We wish to convey our gratitude to our previous intern, Yuqi Ding, for his foundational efforts on the dome structure and the development of the multi-lighting system.

REFERENCES

- [1] Benjamin Attal, Jia-Bin Huang, Michael Zollhöfer, Johannes Kopf, and Changil Kim. 2022. Learning neural light fields with ray-space embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19819–19829.
- [2] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. 2021. Nerf: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12684–12694.
- [3] Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan Barron, Hendrik Lensch, and Varun Jampani. 2022. Samurai: Shape and material from unconstrained real-world arbitrary image collections. *Advances in Neural Information Processing Systems* 35 (2022), 26389–26403.
- [4] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. 2021. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. *Advances in Neural Information Processing Systems* 34 (2021), 10691–10704.
- [5] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. 2020. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3982–3991.
- [6] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen. 2001. Unstructured Lumigraph Rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*. Association for Computing Machinery, New York, NY, USA, 425–432.
- [7] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum. 2000. Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 307–318.
- [8] Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics (ToG)* 34, 4 (2015), 1–13.
- [9] Kingma Da. 2014. A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [10] Paul Debevec. 2008. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Acm siggraph 2008 classes*. 1–10.
- [11] 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.
- [12] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5501–5510.
- [13] Yasutaka Furukawa and Jean Ponce. 2009. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence* 32, 8 (2009), 1362–1376.
- [14] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. 2019. Fast spatially-varying indoor lighting estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6908–6917.
- [15] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen. 1996. The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 43–54.
- [16] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. 2019. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (ToG)* 38, 6 (2019), 1–19.
- [17] Peter Hedman, Pratul P Srinivasan, Ben Mildenhall, Jonathan T Barron, and Paul Debevec. 2021. Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5875–5884.
- [18] Takeo Kanade and PJ Narayanan. 2007. Virtualized reality: perspectives on 4D digitization of dynamic events. *IEEE Computer Graphics and Applications* 27, 3 (2007), 32–40.
- [19] Yoshihiro Kanamori and Yuki Endo. 2018. Relighting humans: occlusion-aware inverse rendering for full-body human images. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–11.
- [20] Brian Karis and Epic Games. 2013. Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice* 4, 3 (2013), 1.
- [21] Marc Levoy and Pat Hanrahan. 1996. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 31–42.
- [22] Zhong Li, Liangchen Song, Celong Liu, Junsong Yuan, and Yi Xu. 2022. NeuLF: Efficient Novel View Synthesis with Neural 4D Light Field. In *Eurographics Symposium on Rendering*.
- [23] Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. 2018. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–11.
- [24] Celong Liu, Lingyu Wang, Zhong Li, Shuxue Quan, and Yi Xu. 2022. Real-time lighting estimation for augmented reality via differentiable screen-space rendering. *IEEE Transactions on Visualization and Computer Graphics* 29, 4 (2022), 2132–2145.
- [25] Alexander I Lvovsky. 2013. Fresnel equations. *Encyclopedia of Optical Engineering* 27 (2013), 1–6.
- [26] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, Paul E Debevec, et al. 2007. Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. *Rendering Techniques 2007*, 9 (2007), 10.
- [27] Vincent Masselus, Pieter Peers, Philip Dutré, and Yves D Willemsy. 2004. Smooth reconstruction and compact representation of reflectance functions for image-based relighting. In *Proceedings of the fifteenth eurographics conference on rendering techniques*. 287–298.
- [28] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. 2019. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.
- [29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*. Springer, 405–421.
- [30] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting triangular 3d models, materials, and lighting from images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8280–8290.
- [31] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. 2021. Total relighting: learning to relight portraits for background replacement. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–21.
- [32] Gilles Rainer, Adrien Bousseau, Tobias Ritschel, and George Drettakis. 2022. Neural Precomputed Radiance Transfer. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 365–378.
- [33] Gilles Rainer, Wenzel Jakob, Abhijeet Ghosh, and Tim Weyrich. 2019. Neural bit compression and interpolation. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 235–244.
- [34] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D Castillo, and David W Jacobs. 2018. Sfsmnet: Learning shape, reflectance and illuminance of faces in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6296–6305.
- [35] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. 2017. Portrait lighting transfer using a mass transport approach. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1.
- [36] Vincent Sitzmann, Semon Rezhchikov, William T. Freeman, Joshua B. Tenenbaum, and Fredo Durand. 2021. Light Field Networks: Neural Scene Representations with Single-Evaluation Rendering. In *Proc. NeurIPS*.
- [37] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. 2019. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2437–2446.
- [38] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. 2019. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems* 32 (2019).
- [39] Liangchen Song, Anpei Chen, Zhong Li, Zhang Chen, Lele Chen, Junsong Yuan, Yi Xu, and Andreas Geiger. 2023. Nerfplayer: A streamable dynamic scene representation with decomposed neural radiance fields. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2732–2742.
- [40] Liangchen Song, Zhong Li, Xuan Gong, Lele Chen, Zhang Chen, Yi Xu, and Junsong Yuan. 2023. Harnessing Low-Frequency Neural Fields for Few-Shot View Synthesis. *arXiv preprint arXiv:2303.08370* (2023).
- [41] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. 2021. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7495–7504.
- [42] Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. 2022. Light Field Neural Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8269–8279.
- [43] 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 (2019), 79–1.
- [44] Tiancheng Sun, Kai-En Lin, Sai Bi, Zexiang Xu, and Ravi Ramamoorthi. 2021. Nelf: Neural light-transport field for portrait view synthesis and relighting. *EGSR* (2021).
- [45] 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 Transactions on Graphics (TOG)* 39, 6 (2020), 1–12.

- [46] Justus Thies, Michael Zollhöfer, and Matthias Nießner. 2019. Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–12.
- [47] Borom Tunwattapanong, Abhijeet Ghosh, and Paul Debevec. 2011. Practical image-based relighting and editing with spherical-harmonics and local lights. In *2011 Conference for Visual Media Production*. IEEE, 138–147.
- [48] 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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 5481–5490.
- [49] Daniel N Wood, Daniel I Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H Salesin, and Werner Stuetzle. 2000. Surface light fields for 3D photography. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 287–296.
- [50] Robert J Woodham. 1980. Photometric method for determining surface orientation from multiple images. *Optical engineering* 19, 1 (1980), 139–144.
- [51] Yi Xu and Daniel G Aliaga. 2010. High-resolution modeling of moving and deforming objects using sparse geometric and dense photometric measurements. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1237–1244.
- [52] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. 2018. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–13.
- [53] 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 *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part I*. Springer, 266–284.
- [54] Yao Yao, Jingyang Zhang, Jingbo Liu, Yihang Qu, Tian Fang, David McKinnon, Yanghai Tsin, and Long Quan. 2022. NeLF: Neural Incident Light Field for Physically-based Material Estimation. In *European Conference on Computer Vision (ECCV)*.
- [55] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. 2021. Plenotrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5752–5761.
- [56] Jason Zhang, Gengshan Yang, Shubham Tulsiani, and Deva Ramanan. 2021. NeRS: Neural reflectance surfaces for sparse-view 3d reconstruction in the wild. *Advances in Neural Information Processing Systems* 34 (2021), 29835–29847.
- [57] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. 2022. 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*. 5565–5574.
- [58] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 2021. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5453–5462.
- [59] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. 2021. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics (TOG)* 40, 1 (2021), 1–17.
- [60] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. 2021. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–18.
- [61] Yingliang Zhang, Zhong Li, Wei Yang, Peihong Yu, Halting Lin, and Jingyi Yu. 2017. The light field 3D scanner. In *2017 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–9.
- [62] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. 2022. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18643–18652.
- [63] Zhengyou Zhang. 2000. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* 22, 11 (2000), 1330–1334.
- [64] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. 2018. Stereo magnification. *ACM Transactions on Graphics* 37, 4 (2018), 1–12.