S^3-NeRF: Neural Reflectance Field from Shading and Shadow under a Single Viewpoint

Wenqi Yang
The University of Hong Kong
wqyang@cs.hku.hk

Guanying Chen*
FNii and SSE, CUHK-Shenzhen
chenguanying@cuhk.edu.cn

Chaofeng Chen
Nanyang Technological University
chaofenghust@gmail.com

Zhenfang Chen
MIT-IBM Watson AI Lab
chenzhenfang2013@gmail.com

Kwan-Yee K. Wong
The University of Hong Kong
kykwong@cs.hku.hk

Abstract

In this paper, we address the "dual problem" of multi-view scene reconstruction in which we utilize single-view images captured under different point lights to learn a neural scene representation. Different from existing single-view methods which can only recover a 2.5D scene representation (i.e., a normal / depth map for the visible surface), our method learns a neural reflectance field to represent the 3D geometry and BRDFs of a scene. Instead of relying on multi-view photo-consistency, our method exploits two information-rich monocular cues, namely shading and shadow, to infer scene geometry. Experiments on multiple challenging datasets show that our method is capable of recovering 3D geometry, including both visible and invisible parts, of a scene from single-view images. Thanks to the neural reflectance field representation, our method is robust to depth discontinuities. It supports applications like novel-view synthesis and relighting. Our code and model can be found at https://ywq.github.io/s3nerf.

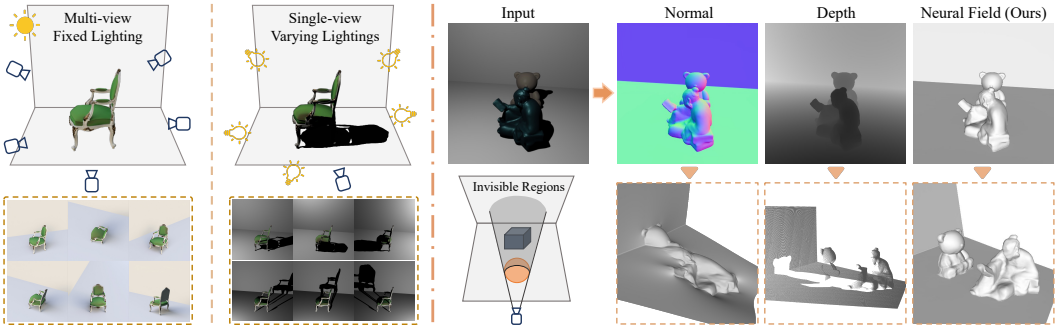
1 Introduction

3D reconstruction from images is a central and important problem in computer vision. Multi-view stereo methods, which capture a target scene from multiple viewpoints under a fixed lighting condition [12, 24, 45, 46], are the most widely adopted approach for scene reconstruction. These methods, however, often assume surfaces with Lambertian reflectance and have difficulties in recovering high-frequency surface details.

An alternative approach to scene reconstruction is to utilize images captured from a fixed viewpoint but under different point light sources (see Fig. 1 (a)). This setup is adopted by photometric stereo (PS) methods [15, 47, 56] where shading information is utilized to reconstruct surface details of non-Lambertian objects. Shadow is another cue that has been exploited for shape recovery by shape-from-shadow methods [11, 61, 67]. However, existing single-view methods typically adopt a single normal or depth map to represent the visible surface, making them incapable of describing back-facing and occluded surfaces (see Fig. 1 (b)). Besides, methods relying on surface normal representation struggle

*Corresponding author

36th Conference on Neural Information Processing Systems (NeurIPS 2022).



(a) Different capturing setups (b) Comparison of different scene representations

Figure 1: (a) Difference between multi-view fixed lighting and single-view varying lighting setups. (b) Comparison of the normal map, depth map, and neural field in representing a 3D scene. Obtaining accurate depth from normal integration is non-trivial [5], and depth map cannot describe the invisible regions. The adopted neural field is capable of modeling the complete scene geometry.

to deal with depth discontinuities [25]. It is desirable to obtain a more complete scene reconstruction (including both visible and invisible parts) from single-view images. In this paper, we realize this by exploiting both shading and shadow cues to recover both visible and invisible parts of a scene.

Recently, neural scene representations have achieved significant progress in multi-view reconstruction and novel-view synthesis [35, 48, 63]. These methods model a continuous 3D space (i.e., the scene) with a multi-layer perception (MLP) which maps 3D points to scene properties (*e.g.*, density and color in NeRF [35]). Despite its great success in multi-view scene modeling, neural scene representation has been less explored in single-view scene modeling.

In this paper, building on top of the recent advances in neural scene representation, we propose to optimize a neural field using images captured from a single viewpoint under different point lights. Our method is fundamentally different from existing works [35, 48, 63] in that, instead of relying on multi-view photo-consistency, we exploit monocular shading and shadow cues to optimize our neural field for scene reconstruction (see Sec. A in supplementary for intuitive explanations on shadow cues).

A straightforward idea would be to condition the color MLP of NeRF [35] also on the point light directions. However, we find such a naïve solution fails to recover scene geometry and appearance. To make better use of the photometric stereo images, we explicitly model the surface geometry and BRDFs with a reflectance field and adopt a physics-based rendering to obtain the 3D point color [2, 3]. The 2D pixel color of a sampled ray can then be computed using volume rendering. Differentiable shadow computation is considered explicitly by tracing a ray from a 3D point to the point light position to check the light visibility [71]. As evaluating the light visibility of all points sampled along a ray is computationally expensive [49], we accelerate the computation by only evaluating the light visibility at the expected surface point, making online shadow computation possible during optimization.

To summarize, our contributions are:

- We address a novel problem of 3D neural reflectance field optimization from single-view images captured under different point lights. Different from existing neural scene representation methods that rely on multi-view photo-consistency, our method exploits monocular shading and shadow cues for neural field optimization.
- Our method jointly recovers the geometry and BRDFs of a scene, and adopts an efficient online shadow computation to fully exploit the information-rich shading and shadow cues.
- Experiments on multiple challenging datasets show that our method can faithfully reconstruct a complete scene geometry from single-view images. Our method is robust to depth discontinuities. It supports applications like novel-view synthesis and relighting.

2 Related Work

Photometric stereo (PS) PS methods can recover pixel-wise surface normals from images captured under different light directions [15, 56]. Traditional PS methods treat specular observations as

outliers [36, 57, 58] or fit sophisticated reflectance models [10, 18, 54] to handle non-Lambertian surfaces. Recent methods resort to deep learning technique to solve this problem. Supervised learning methods learn a mapping from image observations to surface normals using synthetic dataset with ground-truth normals [7, 9, 17, 26, 32, 43, 73]. Self-supervised methods optimize the network parameters using an image reconstruction loss [22, 51]. The above methods assume directional lightings. For near-field PS problem, methods based on PDE [40] and deep learning [29, 31, 34, 44] have been proposed. More recently, Li *et al.* [25] proposes a coordinate-based MLP to represent the normal map of the visible surface assuming directional lights. In contrast, our method represents a scene with a continuous volume and recovers the full 3D scene geometry under a near-field setup.

Shape from shadow Shadow has been exploited to estimate shape information [11]. Yu and Chang optimized a height map from shadow cues using a graph-based representation [67]. Shadowcuts [6] explicitly considers shadow in Lambertian photometric stereo. Yamashita *et al.* [61] introduced a 1D shadow graph to accelerate the shadow computation. Recently, DeepShadow [21] models the depth map of a scene by an MLP and optimizes the model with a shadow reconstruction loss. These methods can only recover a height map of the visible surface. Besides, they require the detected shadow regions as input, but shadow detection is itself a non-trivial problem.

Neural scene representation Neural scene representations have been successfully applied in novel-view synthesis and multi-view reconstruction [38, 48, 52, 59, 63]. The popular neural radiance field (NeRF) [35] represents a continuous space with an MLP, which regresses the volume density and RGB color of a 3D point from the point coordinates and view direction. Attracted by the photo-realistic rendering produced by NeRF, many follow-up works are introduced to improve the reconstructed surface quality [39, 55, 64], rendering speed [14, 30, 41], optimization speed [37, 50, 65], and robustness [1, 33, 68].

The above methods consider each 3D point as an emitter, making them not able to model the surface materials and lighting separately. Inverse rendering methods have been proposed to jointly recover shape, materials, and lightings in a casual capture setup [3, 4, 69, 71, 72]. NeRV [49] explicitly models shadow and indirect illumination assuming a known environment map. NRF [2] and IRON [70] adopt a co-located camera-light setup to simplify the image formation model. PS-NeRF [62] utilizes multi-view and multi-light images to induce regularizations for more accurate surface reconstruction.

There are some attempts to reconstruct a radiance field from a single-view image in a data-driven manner (e.g., conditioning the MLP input with image features [13, 42, 66]), or utilizing depth image as shape prior [60]. However, due to the strong ambiguity, these methods struggle to achieve high-quality reconstruction. Compared with the above approaches, our method extends neural scene representation to reconstruct accurate shape and materials from single-view photometric stereo images.

3 Method

Given N images captured from a single viewpoint under different near point lights, our method targets at recovering the geometry and materials for the scene (see Fig. 2). Following existing near-field photometric stereo methods [34, 44], we assume a calibrated perspective camera and known point light positions. Instead of representing the visible surface with a normal / depth map like others [25, 34, 44], we adopt a 3D neural field representation [3, 35, 39] to describe the 3D scene.

3.1 Neural Reflectance Field Representation

Our method is built on top of the recent neural radiance field (NeRF) [35]. Following UNISURF [39], we adopt an occupancy field instead of a density field to better represent the surface geometry. UNISURF uses an MLP to map a 3D point $\boldsymbol{x} \in \mathbb{R}^3$ and a view direction $\boldsymbol{d} \in \mathbb{R}^3$ to occupancy $o(\boldsymbol{x}) \in \mathbb{R}$ and color $c(\boldsymbol{x}, \boldsymbol{d}) \in \mathbb{R}^3$. An image can be generated through volume rendering in which the color of each pixel (or ray $\boldsymbol{r}$) is calculated by

$$\boldsymbol{C}(\boldsymbol{r}) = \sum_{i=1}^{N_V} o(\boldsymbol{x}_i) \prod_{j<i} \left(1 - o\left(\boldsymbol{x}_j\right)\right) c(\boldsymbol{x}_i, \boldsymbol{d}), \tag{1}$$

where $\boldsymbol{x}_i$ denotes a 3D point sampled along the ray $\boldsymbol{r} = \boldsymbol{o} + t\boldsymbol{d}$, with $\boldsymbol{o} \in \mathbb{R}^3$ being the camera center and $\boldsymbol{d} \in \mathbb{R}^3$ the ray direction specified by the pixel, and $N_V \in \mathbb{R}$ is the number of samples per ray.

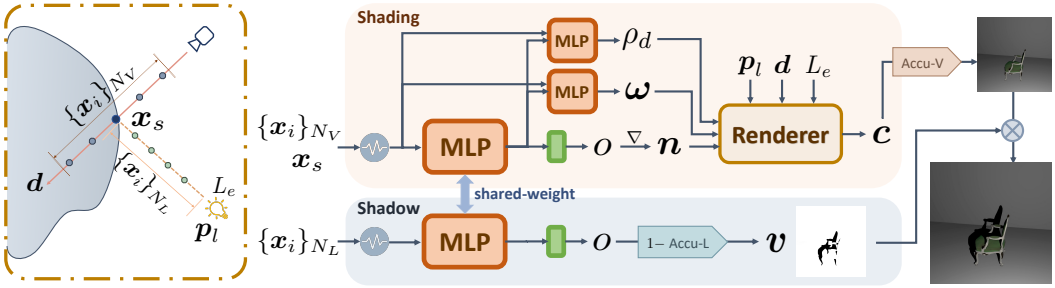


Figure 2: Overview of the method. For each camera ray, we first apply root-finding to locate the surface intersection point $\boldsymbol{x}_s$. N_V points on the camera ray are sampled within a relatively large interval around the surface to generate accumulated shading values. N_L points are sampled on the surface-to-light segment to calculate the light visibility, which is multiplied to the accumulated shading to output the final RGB value.

Given multi-view images, a radiance field can be optimized to reproduce the input images. However, applying NeRF-based methods to single-view images is non-trivial. A straightforward idea would be to condition the color MLP of NeRF also on light directions, but our experiments show that such a naïve solution fails to produce reasonable reconstruction due to the lack of constraints on scene geometry.

To utilize shading information in photometric stereo images, we explicitly model the BRDFs of the scene and recover 3D point color with physics-based rendering [2, 3]. Observing that shadow provides strong cues for inferring the geometry of both visible and invisible surfaces in a scene, we compute shadow in an online manner by tracing a ray from a surface point to the light position to determine its light visibility. Give a point light located at $\boldsymbol{p}_l \in \mathbb{R}^3$ with emitted intensity $L_e \in \mathbb{R}$, Eq. (1) can be rewritten as

$$\boldsymbol{C}(\boldsymbol{r}) = \sum_{i=1}^{N_V} o(\boldsymbol{x}_i) \prod_{j<i} (1 - o(\boldsymbol{x}_j))\, f_v(\boldsymbol{p}_l; \boldsymbol{x}_i) f_c(\boldsymbol{d}, \boldsymbol{p}_l, L_e; \boldsymbol{x}_i), \tag{2}$$

where the 3D point color $c(\boldsymbol{x}, \boldsymbol{d})$ is replaced by the product of light visibility $f_v(\boldsymbol{p}_l; \boldsymbol{x})$ and physics-based rendered color $f_c(\boldsymbol{d}, \boldsymbol{p}_l, L_e; \boldsymbol{x})$, the details of which are given in the following subsections.

3.2 Physics-based Color Rendering

We consider non-Lambertian surfaces with spatially-varying BRDFs. The rendering equation for a surface point $\boldsymbol{x}$ viewed from a direction $\boldsymbol{d}$ under a near point light $(\boldsymbol{p}_l, L_e)$ can be written as

$$f_c(\boldsymbol{d}, \boldsymbol{p}_l, L_e; \boldsymbol{x}) = \underbrace{L_{int}(\boldsymbol{p}_l, L_e; \boldsymbol{x})}_{\text{Light Intensity}} \underbrace{f_m(\boldsymbol{d}, \boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x}); \boldsymbol{x})}_{\text{BRDF Value}} \underbrace{\max\left(\boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}), 0\right)}_{\text{Shading}}, \tag{3}$$

where $L_{int}(\boldsymbol{p}_l, L_e; \boldsymbol{x})$ denotes the incident light (taking light falloff into account), $\boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x})$ the incident light direction, and $f_m(\boldsymbol{d}, \boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x}); \boldsymbol{x})$ the BRDF value at $\boldsymbol{x}$. The normal at $\boldsymbol{x}$ can be derived from the gradient of the occupancy field as $\boldsymbol{n}(\boldsymbol{x}) = \nabla o(\boldsymbol{x}) / \|\nabla o(\boldsymbol{x})\|_2$ [39].

Lighting model Following previous works [34, 44], we adopt the inverse-square law for point light attenuation where light intensity L_{int} is proportional to the multiplicative inverse of the square of the distance s (*i.e.*, $L_{int} \propto 1/s^2$). The incident light direction $\boldsymbol{w}_i$ and light intensity L_{int} at a point $\boldsymbol{x}$ are given by

$$\boldsymbol{w}_i(\boldsymbol{p}_l; \boldsymbol{x}) = \frac{\boldsymbol{p}_l - \boldsymbol{x}}{\|\boldsymbol{p}_l - \boldsymbol{x}\|_2}, \qquad L_{int}(\boldsymbol{p}_l, L_e; \boldsymbol{x}) = \frac{L_e}{\|\boldsymbol{p}_l - \boldsymbol{x}\|_2^2}. \tag{4}$$

BRDF model Similar to [3, 69, 71], we adopt a BRDF model represented by a combination of diffuse color ρ_d and specular reflectance ρ_s, which is given by

$$f_m(\boldsymbol{w}_i, \boldsymbol{w}_o; \boldsymbol{x}) = \rho_d + \rho_s(\boldsymbol{w}_i, \boldsymbol{w}_o; \boldsymbol{x}). \tag{5}$$

Following [16, 25], we model the isotropic specular reflectance by a weighted combination of Sphere Gaussian (SG) bases, which demonstrates better results in modeling specular effects than the parametric Microfacet model [20]. The specular component ρ_s is hence written as $\rho_s = \boldsymbol{\omega}^T D(\boldsymbol{h}, \boldsymbol{n})$, where $D(\boldsymbol{h}, \boldsymbol{n}) = G(\boldsymbol{h}, \boldsymbol{n}; \lambda) = \left[e^{\lambda_1(\boldsymbol{h}^T\boldsymbol{n}-1)}, \cdots, e^{\lambda_k(\boldsymbol{h}^T\boldsymbol{n}-1)}\right]^T$ denotes the SG bases, with $\lambda_* \in \mathbb{R}_+$ controls the specular sharpness. The diffuse component ρ_d and SG weights $\boldsymbol{\omega}$ are estimated by two MLPs.

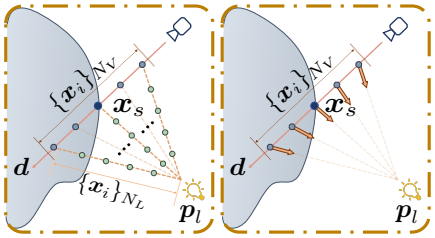
3.3 Online Shadow Computation

A 3D point $\boldsymbol{x}$ is shadowed if there is any occluders in its line of sight for the light position $\boldsymbol{p}_l$. It follows that light visibility $f_v(\boldsymbol{p}_l, \boldsymbol{x}) \in [0, 1]$ for a 3D point $\boldsymbol{x}$ can be computed by accumulating occupancies along this line (see Fig. 2), *i.e.*,

$$f_v(\boldsymbol{p}_l; \boldsymbol{x}) = 1 - \sum_{i=1}^{N_L} o(\boldsymbol{x}_i) \prod_{j<i} (1 - o(\boldsymbol{x}_j)), \tag{6}$$

where N_L is the number of points sampled along the line.

However, calculating light visibilities for all N_V points sampled along the ray for a pixel is computationally expensive (*i.e.*, $O(N_V N_L)$ MLP queries for each pixel / ray (see Fig. 3 (a)). To speed up shadow computation, previous methods either adopt an MLP to directly regress light visibility of a point [49] to reduce the queries for each ray to $O(N_V)$ (see Fig. 3 (b)), or pre-extracts the surface points (assuming a fixed scene geometry) [71] to reduce the number of MLP queries to $O(N_L)$. Instead, we first locate the expected surface points $\boldsymbol{x}_s$ along the ray by root-finding [39] and calculate its light visibility in an online manner. Eq. (2) can be reformulated for efficient color rendering as



(a) $O(N_V N_L)$ (b) $O(N_V)$

Figure 3: Alternative shadow modeling.

$$\boldsymbol{C}(\boldsymbol{r}) = f_v(\boldsymbol{p}_l; \boldsymbol{x}_s) \sum_{i=1}^{N_V} o(\boldsymbol{x}_i) \prod_{j<i} (1 - o(\boldsymbol{x}_j))\, f_c(\boldsymbol{x}_i, \boldsymbol{d}, \boldsymbol{p}_l, L_e). \tag{7}$$

3.4 Optimization

Different from shape-from-shadow methods [21, 53], our method does not require direct supervision for shadow rendering. We rely on image reconstruction loss for optimization.

Volume rendering loss The first loss is the L1 reconstruction loss between the volume rendered image $\boldsymbol{C}_v$ (*i.e.*, the computed $\boldsymbol{C}(\boldsymbol{r})$ in Eq. (7)) and the input image:

$$\mathcal{L}_v = \sum \|\boldsymbol{C}_v - I\|_1. \tag{8}$$

Surface rendering loss UNISURF [39] proposes to combine the volume rendering and surface rendering by gradually shortening the sampling range in a ray to refine the surface region. However, we empirically found that the model will start to degrade when the sampling interval is decreased as there is no multi-view information to constrain the non-sampled regions. We therefore propose to adopt a joint volume and surface rendering strategy. We additionally compute the surface rendering color $\boldsymbol{C}_s(\boldsymbol{r})$ using the expected surface point $\boldsymbol{x}_s$ and calculate the L1 loss, *i.e.*,

$$\mathcal{L}_s = \sum \|\boldsymbol{C}_s - I\|_1, \tag{9}$$

$$\boldsymbol{C}_s(\boldsymbol{r}) = f_v(\boldsymbol{p}_l; \boldsymbol{x}_s) f_c(\boldsymbol{d}, \boldsymbol{p}_l, L_e; \boldsymbol{x}_s). \tag{10}$$

Normal smoothness loss Similar to [39], we also include a regularization loss to promote smoothness in surface normal (ϵ is a small random perturbation):

$$\mathcal{L}_n = \sum \|\boldsymbol{n}(\boldsymbol{x}_s) - \boldsymbol{n}(\boldsymbol{x}_s + \epsilon)\|_2^2. \tag{11}$$

Table 1: Comparison with neural field methods on relighting and normal estimation results.

Method	BUDDHA		READING		BUNNY		CHAIR		LEGO		HOTDOG	
	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓
NeRF* [35]	38.57	70.12	39.50	72.60	37.41	68.35	35.25	88.46	**35.56**	91.09	**39.80**	72.07
UNISURF* [39]	41.51	54.86	40.54	60.59	38.48	54.27	34.98	47.80	34.55	45.81	38.64	51.00
Ours	**43.42**	**2.52**	**43.13**	**2.11**	**40.43**	**1.81**	**36.33**	**1.93**	35.54	**6.57**	38.01	**2.59**

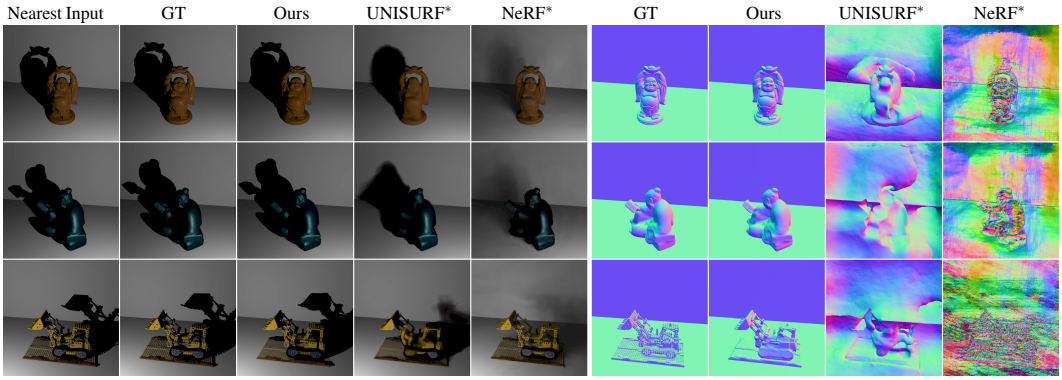


Figure 4: Comparison with neural field methods on relighting (left) and normal estimation (right).

Overall loss The overall loss function used for optimization is as follow with α set to 0.005:

$$\mathcal{L} = \mathcal{L}_v + \mathcal{L}_s + \alpha\mathcal{L}_n. \tag{12}$$

4 Experiments

4.1 Implementation Details

Similar to UNISURF [39], we use an 8-layer MLP (256 channels with softplus activation) to predict the occupancy o and output a 256-dimensional feature vector. Two additional 4-layer MLPs then take the feature vector and point coordinates as input to predict the albedo ρ_d and weights $\boldsymbol{\omega}$ of SG bases. We sample $N_V = 256$ points along the camera ray and $N_L = 256$ points along the surface-to-light line segment. We use Adam optimizer [23] with an initial learning rate of 0.0002 which decays at 200 and 400 epochs. We train each scene for 800 epochs on one Nvidia RTX 3090 card, which takes about 16 hours to converge.

Evaluation Metrics We adopt mean angular error (MAE) in degree for surface normal evaluation and L1 error in cm for depth assessment. PSNR is used to measure the quality of images rendered under novel view or novel lighting.

4.2 Datasets

Our method targets at recovering the complete scene by exploiting both shading and shadow. However, existing photometric stereo datasets are mostly interested in the object region and intentionally remove the influence of the background (*e.g.*, cover the background with black cloth to avoid inter-reflections [47]), which makes the shadow and shading information invisible in the background regions. Therefore, such datasets [34, 47] are not suitable to evaluate the full potential of our method.

Instead, we evaluate our method on multiple synthetic datasets with complicated scene geometry and materials. Specifically, we used 10 3D objects for data rendering, where 5 objects from DiLiGent-MV Dataset [27] (namely, *BEAR*, *BUDDHA*, *COW*, *POT2*, and *READING*), 2 objects from the internet (namely, *BUNNY* and *ARMADILLO*), and 3 objects from NeRF's blender dataset [35] (namely, *LEGO*, *CHAIR*, and *HOTDOG*). We rendered *LEGO*, *CHAIR*, and *HOTDOG* with Blender's Cycles pathtracer, and the other 7 objects with Mitsuba [19]. As our method does not explicitly model inter-reflections, we set the max bounces to 0 during rendering. During rendering, we created a scene by adding a horizontal and a vertical plane to model the desk and wall, and objects are placed on the

Table 2: Comparison with single-view normal / depth estimation methods (only object regions).

Method	BUDDHA		READING		BUNNY		CHAIR		LEGO		HOTDOG	
	MAE↓	Depth L1↓	MAE↓	Depth L1↓	MAE↓	Depth L1↓	MAE↓	Depth L1↓	MAE↓	Depth L1↓	MAE↓	Depth L1↓
ZL18 [28]	37.52	19.84	37.29	25.97	31.41	17.68	39.53	41.19	46.83	34.56	39.75	18.02
QY18 [40]	**12.27**	3.81	40.84	26.13	14.23	4.10	29.69	15.95	33.09	17.87	16.82	8.98
HS20 [44]	18.40	6.47	27.11	18.94	16.93	10.96	29.56	13.99	33.54	13.27	27.25	13.22
Ours	14.25	**1.50**	**7.02**	**2.09**	**9.42**	**1.63**	**17.44**	**4.74**	**31.13**	**7.31**	**14.67**	**1.68**

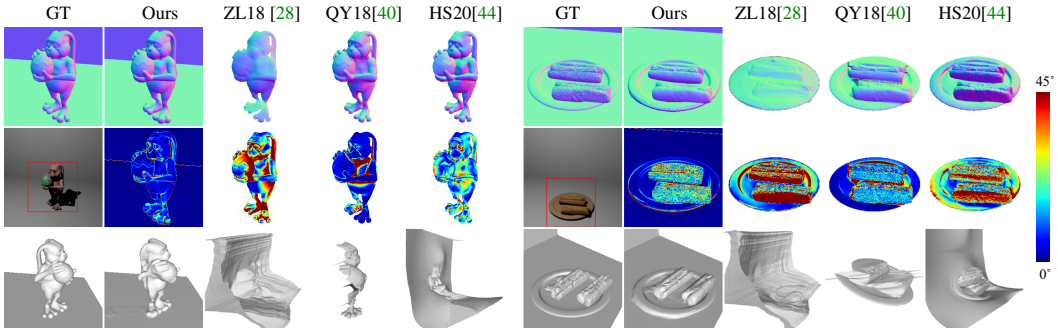


Figure 5: Comparison with single-view normal / depth estimation baselines. Row 1 and Row 2 show the normal and error maps. Row 3 shows the side-view of the reconstructed surfaces.

horizontal plane. Each scene was rendered under 128 uniformly sampled near point lights, and the rendered images are in linear space with a resolution of 512×512.

4.3 Comparisons with Existing Methods

To justify the effectiveness of our method, we compare it with three types of methods, namely, neural field methods, photometric stereo methods, and single-image shape estimation methods.

Neural radiance field methods We first verify the design of our method by comparing it with two simple baselines (*i.e.*, adapting NeRF [35] and UNISURF [39] for this problem by conditioning the color MLP on light direction). Table 1 and Fig. 4 show the normal estimation and relighting results in the training view. Although the baseline methods can achieve reasonable rendering results in terms of PSNR, they fail to predict accurate cast-shadow and cannot reconstruct the geometry of the scene (with a large average MAE of 77.12/52.39). In contrast, our method is able to accurately reconstruct the shape with an average MAE of 2.92, and achieves the best rendering results (average PSNR of 39.48). This result indicates that simply conditioning the color MLP on light direction does not provide sufficient constraint to regularize the scene geometry.

Single-view shape estimation methods We then compare with three state-of-the-art single-view normal / depth estimation methods, including two near-field PS methods (QY18 [40] and HS20 [44]) and one single-image shape estimation method (ZL18 [28]). QY18 [40] and HS20 [44] consider exactly the same setup as our method (multiple images captured under near point lights), so the input are the same as our method. ZL18 [28] assumes an image captured under co-located flash light as input, so we choose the image illuminated by a point light that is closest to the camera as its input. As these methods are designed to estimate the shape in the object region and have difficulty in dealing with the background, we only report the normal and depth estimation results on the object region for the training view in Table 2. Since ZL18 [28] and HS20 [44] require depth alignment before evaluation, we align the estimated depth with the ground truth for all the methods for fair comparison. We can see that our method achieves the best results for both normal and depth estimation. Moreover, as shown in Fig. 5, our method can faithfully reconstruct both visible and invisible parts of the scene, which is not possible by methods that rely on the normal or depth representation.

4.4 Method Analysis

We next conduct ablation study for different components of our method, and evaluate our method on different setups to further analyze its behavior.

Table 3: Quantitative results for the ablation study.

Method	Train View						Novel Views					
	CHAIR		BUNNY		BUDDHA		CHAIR		BUNNY		BUDDHA	
	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑	MAE↓	PSNR↑
w/o shading	32.49	33.71	40.19	38.72	35.69	41.43	–	–	–	–	–	–
w/o shadow	3.52	30.81	2.38	33.45	3.45	34.30	12.34	22.31	11.98	24.43	16.65	23.27
w/o $\mathcal{L}_s$	2.55	35.85	2.81	39.73	3.82	43.04	**5.18**	**28.58**	6.31	29.11	8.52	28.61
Ours	**1.93**	**36.33**	**1.81**	**40.43**	**2.52**	**43.42**	5.53	26.82	**6.15**	**29.55**	**6.93**	**31.53**

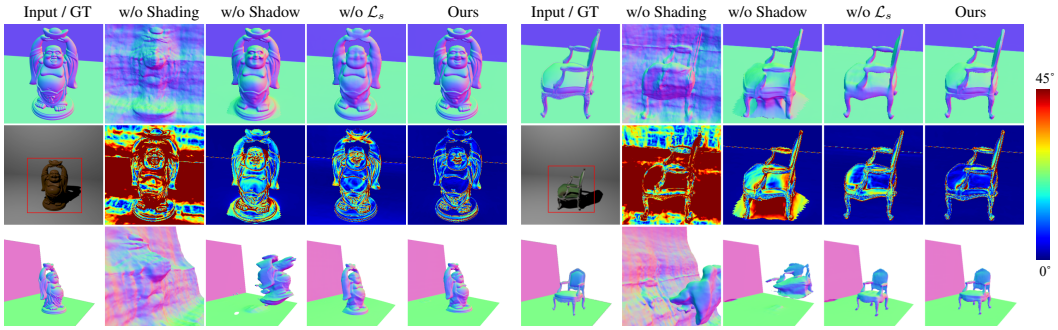


Figure 6: Visual results for the ablation study. Row 1 is the normal of train view, and row 2 shows its error map compared with ground truth. Row 3 shows the normal of a novel view.

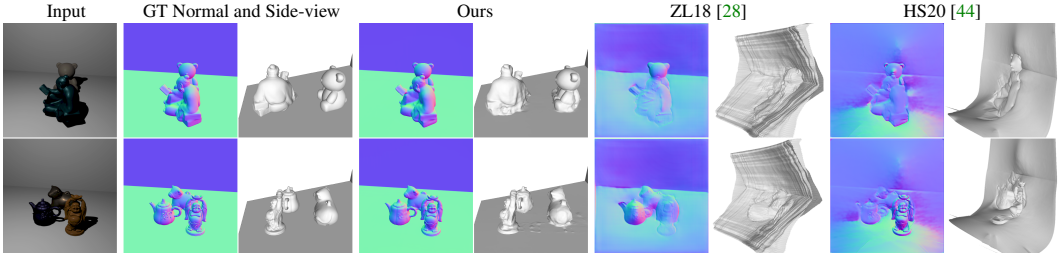


Figure 7: Results on scenes with multiple occluding objects.

Joint shading and shadow modeling Our method exploits both shading and shadow information for scene reconstruction. To analyze the effect of both components in reconstructing the scene geometry, we trained two variant models where (a) *"w/o shading"* replaces the BRDF module with a 4-layer MLP to directly predict RGB values with additional light location input; and (b) *"w/o shadow"* removes the shadow module and only output the shading. Results are summarized in Table 3 and Fig. 6. We evaluate the results for both trained view and novel views, and report MAE of normal maps and PSNR of rendered images. As the *"w/o shading"* model fails to estimate proper depth and the recovered surfaces totally deviate from the ground truth, we omit its results for novel views. Without shadow information, the model may still predict proper surface normal for the trained view. However, the model fails to predict the depth and shape of the object since there is no constraint on invisible regions. By exploiting both shading and shadow cues, our method can well reconstruct the full scene.

Joint volume and surface rendering We also analyze the effectiveness of the surface rendering loss $\mathcal{L}_s$ by comparing our full model with the one without $\mathcal{L}_s$. Results in Table 3 and Fig. 6 show that surface rendering loss can effectively refine the surface normals of the object surface.

Effect of occlusion/discontinuity and unseen region We further demonstrate the potential of our method in reconstructing the complete geometry of the scene, especially when there are occlusions or discontinuous surfaces. Figure 7 shows the reconstructions for two scenes with multiple objects occluding each other. It is very difficult to identify the shape of the invisible regions just from the single-view images. However, by effectively leveraging the shadow information, our method successfully predicts the shape (*i.e.*, the occupancy field) of the invisible regions, which is not feasible for existing works.

We also investigate the performance of our method on surface with challenging invisible shapes. Note that the shape of invisible regions are mainly constrained by the shadow (which indicates the

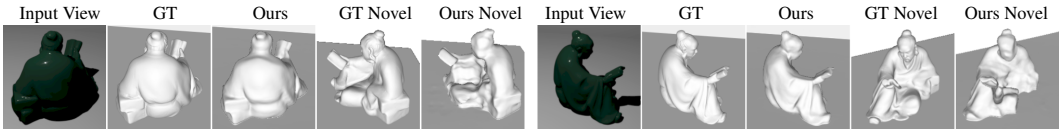


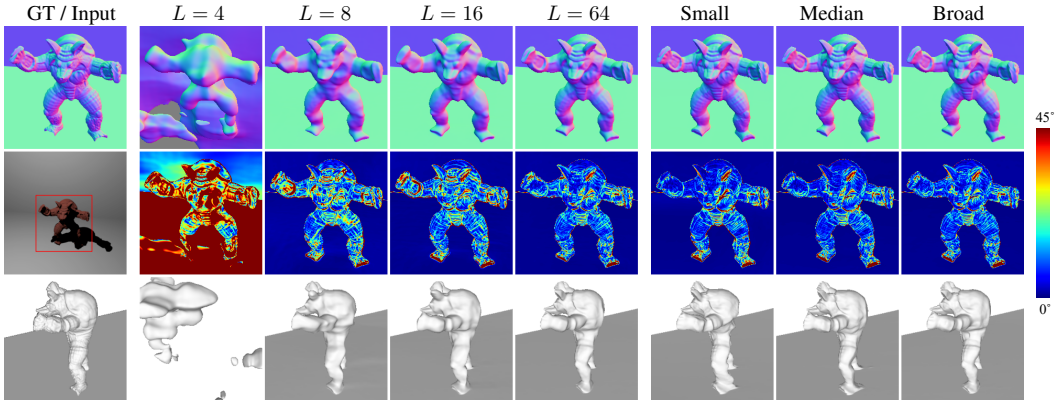
Figure 8: Results on two scenes with challenging invisible shapes.

Table 4: Analysis on Light numbers.

	CHAIR			ARMADILLO		
Light#	MAE↓	Depth↓	PSNR↑	MAE↓	Depth↓	PSNR↑
4	30.44	181.62	19.39	46.87	93.66	15.87
8	2.40	10.99	35.60	2.59	5.98	37.17
16	2.19	**8.11**	36.20	2.47	**5.89**	37.50
32	2.06	8.77	36.23	2.14	6.27	39.20
64	**1.91**	8.64	**36.52**	2.08	7.18	39.78
128	1.93	9.04	36.33	**1.96**	6.65	**40.13**

Table 5: Analysis on Light Range.

	CHAIR			ARMADILLO		
Range	MAE↓	Depth↓	PSNR↑	MAE↓	Depth↓	PSNR↑
small	3.96	18.79	30.15	2.38	6.86	35.26
median	2.08	**8.75**	35.96	**1.84**	**4.59**	38.92
broad	**1.93**	9.04	**36.33**	1.96	6.65	**40.13**



(a) Different Number of Lights (b) Different Light Distributions

Figure 9: Analysis on different number of lights and light distributions.

occupancy along the light path). In Fig. 8, we show the reconstruction of *READING* object which is posed to make the concave surface invisible. From the results of novel views, we can see that our method can properly recover the invisible irregular surface, though some invisible regions are not fully consistent with the ground truth shape. This result demonstrates that shadow provides strong cue for shape recovery especially for unseen regions.

Effect of light distributions To analyze the robustness of our method on different light distribution, we evaluate it on scenes illuminated by different number of lights or different ranges (see our supplementary material for visualization of light distributions), and the numerical results are summarized in Table 4 and Table 5 (depth errors are calculated in object regions). The model fails to reconstruct the scene with only 4 light inputs, but can reconstruct faithful shape of the scene with 8 light inputs. With more lights used, the surface of the object is further refined (see Fig. 9). We can also observe that our method can still work for small range of light distributions to recover invisible regions. Overall, our method is robust to different number and different range of lights.

Results on real scenes To further demonstrate the practicality of our method, we evaluate on three real scenes, which were captured using a fixed camera (with 28mm focal length) and a handheld cellphone flashlight (see Fig. 10). The object was put on the table and close to the wall. We turned off all the environmental light sources and only kept the flashlight on, which was randomly moved around to capture images illuminated under different light conditions. For each object we took around 70 images.

Our setup does not require manual calibration of lights. Instead, we applied the state-of-the-art self-calibrated photometric stereo network (SDPS-Net [8]) for light direction initialization, and roughly measured the camera-object distance as initialization of light-object distance. After initialization, the position and direction of lights are jointly optimized with the shape and BRDF during training. Please refer to our supplementary materials for more training details. Sample inputs and results are

Figure 10: The data capturing setup and three testing objects.

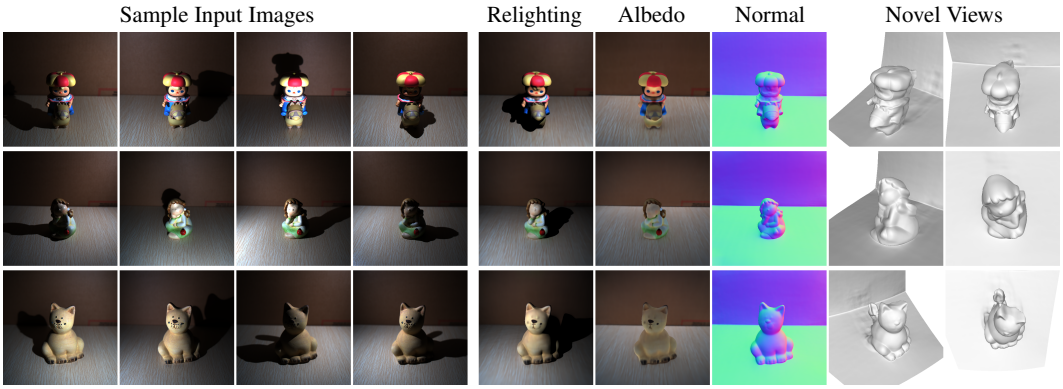


Figure 11: Results on the real captured data. From top to bottom: *BOTANIST*, *GIRL*, *CAT*.

shown in Fig. 11. Even with this casual capturing setup and uncalibrated lights, our method achieved satisfactory results in normal prediction and full 3D shape reconstruction.

5 Conclusion

In this paper, we have introduced a method to optimize a neural reflectance field for a non-Lambertian scene from single-view images captured under different near point lights. Our method jointly recovers the geometry (*i.e.*, occupancy field) and BRDFs of the scene by fully utilizing the shading and shadow cues. Interestingly, our results on scenes with complicated shapes and materials show that the complete scene geometry can be faithfully reconstructed just from single-view photometric images. Moreover, comprehensive method analysis demonstrates that our method is robust to scenes with different geometry, materials, light number, and light distributions. Additionally, our method supports applications like novel-view synthesis and relighting.

Limitation First, like existing near-field PS methods [44], our method requires known light positions, which requires additional efforts for lighting calibration. Second, as our method relies on shadow cue for invisible shape reconstruction, its performance may decrease if the scene background is highly complicated as the background geometry will affect the appearance of shadows. Third, although the shape of the invisible parts can be well reconstructed, the reflectance of those regions are not well constrained by shadow. Last, our method ignores inter-reflection effects in image formation. In the future, we will further extend our method to tackle these limitations.

Acknowledgements This work was partially supported by the National Key R&D Program of China (No.2018YFB1800800), the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone, NSFC-62202409, and the Research Grant Council of Hong Kong (SAR), China (project no. 17203119).

References

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 3

[2] 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 preprint arXiv:2008.03824*, 2020. 2, 3, 4

[3] 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 (ICCV)*, pages 12684–12694, 2021. 2, 3, 4

[4] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan Barron, and Hendrik Lensch. Neural-PIL: Neural pre-integrated lighting for reflectance decomposition. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021. 3

[5] Xu Cao, Boxin Shi, Fumio Okura, and Yasuyuki Matsushita. Normal integration via inverse plane fitting with minimum point-to-plane distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2382–2391, 2021. 2

[6] Manmohan Chandraker, Sameer Agarwal, and David Kriegman. Shadowcuts: Photometric stereo with shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 3

[7] Guanying Chen, Kai Han, and Kwan-Yee K. Wong. PS-FCN: A flexible learning framework for photometric stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–18, 2018. 3

[8] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Self-calibrating deep photometric stereo networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8739–8747, 2019. 9

[9] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K. Wong. Deep photometric stereo for non-Lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 44(1):129–142, 2020. 3

[10] Hin-Shun Chung and Jiaya Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 3

[11] Michael Daum and Gregory Dudek. On 3-d surface reconstruction using shape from shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 461–468, 1998. 1, 3

[12] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 32(8):1362–1376, 2009. 1

[13] Chen Gao, Yichang Shih, Wei-Sheng Lai, Chia-Kai Liang, and Jia-Bin Huang. Portrait neural radiance fields from a single image. *arXiv preprint arXiv:2012.05903*, 2020. 3

[14] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. FastNeRF: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14346–14355, 2021. 3

[15] Hideki Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 1994. 1, 2

[16] Zhuo Hui and Aswin C Sankaranarayanan. Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2017. 5

[17] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 3

[18] Satoshi Ikehata and Kiyoharu Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[19] Wenzel Jakob. Mitsuba renderer, 2010. 6

[20] Brian Karis and Epic Games. Real shading in unreal engine 4. *Proc. Physically Based Shading Theory Practice*, 4(3):1, 2013. 5

[21] Asaf Karnieli, Ohad Fried, and Yacov Hel-Or. Deepshadow: Neural shape from shadow. *arXiv preprint arXiv:2203.15065*, 2022. 3, 5

[22] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Uncalibrated neural inverse rendering for photometric stereo of general surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3804–3814, 2021. 3

[23] Diederik Kingma and Jimmy Ba. ADAM: A method for stochastic optimization. In *Proceedings of the The International Conference on Learning Representations (ICLR)*, 2015. 6

[24] Vladimir Kolmogorov and Ramin Zabih. Multi-camera scene reconstruction via graph cuts. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 82–96, 2002. 1

[25] Junxuan Li and Hongdong Li. Neural reflectance for shape recovery with shadow handling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2, 3, 5

[26] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita. Learning to minify photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[27] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: a robust solution and benchmark dataset for spatially varying isotropic materials. *IEEE Transactions on Image Processing (TIP)*, 2020. 6

[28] Zhengqin Li, Zexiang Xu, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics (TOG)*, 37(6):1–11, 2018. 7, 8

[29] Daniel Lichy, Soumyadip Sengupta, and David W Jacobs. Fast light-weight near-field photometric stereo. *arXiv preprint arXiv:2203.16515*, 2022. 3

[30] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 15651–15663, 2020. 3

[31] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. A cnn based approach for the near-field photometric stereo problem. *arXiv preprint arXiv:2009.05792*, 2020. 3

[32] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. Px-net: Simple and efficient pixel-wise training of photometric stereo networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12757–12766, 2021. 3

[33] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7210–7219, 2021. 3

[34] Roberto Mecca, Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. Luces: A dataset for near-field point light source photometric stereo. *arXiv preprint arXiv:2104.13135*, 2021. 3, 4, 6

[35] 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 *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 405–421, 2020. 2, 3, 6, 7

[36] Yasuhiro Mukaigawa, Yasunori Ishii, and Takeshi Shakunaga. Analysis of photometric factors based on photometric linearization. *JOSA A*, 2007. 3

[37] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989*, 2022. 3

[38] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3504–3515, 2020. 3

[39] Michael Oechsle, Songyou Peng, and Andreas Geiger. UNISURF: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5589–5599, 2021. 3, 4, 5, 6, 7

[40] Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, François Lauze, and Jean-Denis Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision*, 60(3):313–340, 2018. 3, 7

[41] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. KiloNeRF: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 14335–14345, 2021. 3

[42] Konstantinos Rematas, Ricardo Martin-Brualla, and Vittorio Ferrari. Sharf: Shape-conditioned radiance fields from a single view. *Proceedings of the ACM International Conference on Machine Learning (ICML)*, 2021. 3

[43] Hiroaki Santo, Masaki Samejima, Yusuke Sugano, Boxin Shi, and Yasuyuki Matsushita. Deep photometric stereo network. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017. 3

[44] Hiroaki Santo, Michael Waechter, and Yasuyuki Matsushita. Deep near-light photometric stereo for spatially varying reflectances. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 137–152, 2020. 3, 4, 7, 8, 10

[45] Johannes L Schönberger, Enliang Zheng, Jan-Michael Frahm, and Marc Pollefeys. Pixelwise view selection for unstructured multi-view stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 501–518, 2016. 1

[46] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 1

[47] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2019. 1, 6

[48] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019. 2, 3

[49] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7495–7504, 2021. 2, 3, 5

[50] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct Voxel Grid Optimization: Super-fast convergence for radiance fields reconstruction. *arXiv preprint arXiv:2111.11215*, 2021. 3

[51] Tatsunori Taniai and Takanori Maehara. Neural inverse rendering for general reflectance photometric stereo. In *Proceedings of the ACM International Conference on Machine Learning (ICML)*, 2018. 3

[52] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Yifan Wang, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. *arXiv preprint arXiv:2111.05849*, 2021. 3

[53] Kushagra Tiwary, Tzofi Klinghoffer, and Ramesh Raskar. Towards learning neural representations from shadows. *arXiv preprint arXiv:2203.15946*, 2022. 5

[54] Silvia Tozza, Roberto Mecca, M Duocastella, and A Del Bue. Direct differential photometric stereo shape recovery of diffuse and specular surfaces. *Journal of Mathematical Imaging and Vision*, 2016. 3

[55] 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. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021. 3

[56] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 1980. 1, 2

[57] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2010. 3

[58] Tai-Pang Wu and Chi-Keung Tang. Photometric stereo via expectation maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 2010. 3

[59] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. *arXiv preprint arXiv:2111.11426*, 2021. 3

[60] Dejia Xu, Yifan Jiang, Peihao Wang, Zhiwen Fan, Humphrey Shi, and Zhangyang Wang. Sinnerf: Training neural radiance fields on complex scenes from a single image. *arXiv preprint arXiv:2204.00928*, 2022. 3

[61] Yukihiro Yamashita, Fumihiko Sakaue, and Jun Sato. Recovering 3d shape and light source positions from non-planar shadows. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 1775–1778, 2010. 1, 3

[62] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K. Wong. Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In *European Conference on Computer Vision (ECCV)*, 2022. 3

[63] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Ronen Basri, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 2492–2502, 2020. 2, 3

[64] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34, 2021. 3

[65] Alex Yu, Sara Fridovich-Keil, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. *arXiv preprint arXiv:2112.05131*, 2021. 3

[66] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3

[67] Yizhou Yu and Johnny T Chang. Shadow graphs and 3d texture reconstruction. *International Journal of Computer Vision (IJCV)*, 62(1):35–60, 2005. 1, 3

[68] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. NeRF++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 3

[69] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. 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 (CVPR)*, pages 5453–5462, 2021. 3, 4

[70] Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. Iron: Inverse rendering by optimizing neural sdfs and materials from photometric images. *arXiv preprint arXiv:2204.02232*, 2022. 3

[71] 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 Transactions on Graphics (TOG)*, 40(6), 2021. 2, 3, 4, 5

[72] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. *arXiv preprint arXiv:2204.06837*, 2022. 3

[73] Qian Zheng, Yiming Jia, Boxin Shi, Xudong Jiang, Ling-Yu Duan, and Alex C. Kot. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 3

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] See Section 5 Limitation.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] We discuss it in supplementary materials.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See supplementary materials for training details.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See supplementary materials.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes] See supplementary materials.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]