NeIF: Representing General Reflectance as Neural Intrinsics Fields for Uncalibrated Photometric Stereo

Anonymous Author(s)
Affiliation
Address
email

Abstract

Uncalibrated photometric stereo (UPS) is challenging due to the inherent shape-light. Existing solutions alleviate this ambiguity by either explicitly associating reflectance to light estimation or resolving light conditions in a supervised manner. This paper establishes an implicit relation between light clues and light estimation and solves UPS in an unsupervised manner. The key idea is to represent the reflectance as four neural intrinsics fields, *i.e*, specular, shadow, light, and position, based on which the neural light field is implicitly associated to light clues of specular reflectance and cast shadow. The unsupervised training of neural intrinsics fields fully exploits all observed pixel values, which facilitates solving UPS. Our method demonstrates a superior performance advantage over state-of-the-art UPS methods on public datasets and promising results under the challenging setting of sparse UPS. We will release our codes and pre-trained models once the paper was accepted.

1 Introduction

Photometric stereo (PS) [47] aims at recovering the surface normal from several light-varying images captured at a fixed viewpoint. As compared with other stereo approaches (*e.g.*, multi-view stereo [38], active sensor-based solutions [55]), photometric stereo is excellent at recovering fine-detailed surfaces and has been widely used for Hollywood movies [9], industrial quality inspection [46], and biometrics [52]. Calibrating accurate lighting directions is crucial to the performance of photometric stereo methods [51]. However, lighting calibration is often tedious, dramatically limiting the applicability in the real-world. To this end, researchers develop uncalibrated photometric stereo (UPS) methods that estimate surface normal with unknown lights.

Uncalibrated photometric stereo suffers from General Bas-Relief (GBR) ambiguity [6] for an integrable surface. Early solutions address the ambiguity by explicitly associating reflectance to light, *i.e*, adopting analytic reflectance models (*e.g.*, Lambertian reflectance [4], [34], parametric specular reflection [18], specular spikes [53], inter-reflection [10]) or imposing priors from reflectance properties [3, 22, 21, 41]. Thus, these methods could be less effective for unknown reflectance due to the strong reliance reflectance assumption. Besides, these methods ignore the clue of cast shadow and even fail for shadow regions due to the shadow-free surface assumption. Further, most of them require the light intensity to be identical for robust estimation. Recently, deep learning-based approaches address the ambiguity by estimating light direction and intensity before recovering surface normal [12, 13, 25, 37]. They train a light estimation network using a large-scale amount of rendered data in a supervised manner. However, the training data bias [30] can hardly be eliminated and could produce unexpected estimation for real-world data. Because rendered training data inevitably contain the domain gap from real ones and could scarcely cover all surfaces with different geometry and reflectance in the real-world. Besides, such a two-step solution could bring accumulating errors for surface normal

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

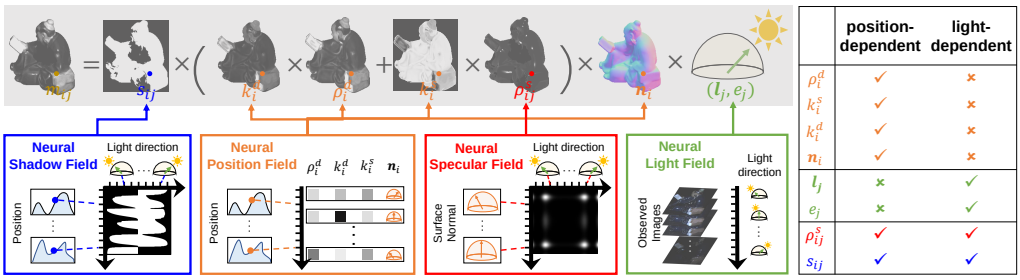

Figure 1: Illustration of our neural intrinsics fields. Left-top: the rendering equation. Left-bottom: our four neural intrinsic fields, *i.e*, from left to right: shadow, position, specular, and light fields, respectively. Each sub-figure in the left-bottom illustrates the mutual information across dimensions of position-light, position, normal-light, observed images, respectively. Left figure shows how the neural intrinsics fields are imposed to render a pixel. The definition of notations could be found from Eq. (2) and Eq. (4) Right: a summary of our intrinsics w.r.t. the dependence on position or light.

estimation. Further, all these methods assume the light intensity to distribute in a pre-defined range (*i.e*, $[0.2, 2]$), limiting their applicability.

To this end, we propose **NeIF**, representing general reflectance as **Ne**ural **I**ntrinsics **F**ields for uncalibrated photometric stereo. Our method differs from previous methods in three aspects: 1) it fully considers clues of specular reflectance and cast shadow from each observed pixel for light estimation so that it is expected to produce accurate estimation for both light conditions and surface normal; 2) it does not make explicit assumptions about the reflectance or light so that it could work with general surface reflectance and flexible light settings; 3) it infers light and surface normal in an unsupervised manner so that it could be free from training data bias and achieve stable performance for data from difference sources.

Our key idea is to represent the general reflectance as four neural intrinsics fields (*i.e*, specular reflectance, cast shadow, light, and position, see Fig. 1), implemented by four multi-layer perceptrons (MLPs). The connection of these four fields is based on the implicit relation (or dependence) of these intrinsics so that no explicit assumption is imposed, *e.g.*, we take the estimated light as the input to recover specular reflectance and cast shadow instead of explicitly exploiting them for light estimation. These intrinsics fields are optimized to reconstruct each pixel value from observed images, which fully exploit mutual information across different dimensions, as shown in Fig. 1. The reconstruction error backpropagates to the neural light field through neural specular and shadow fields, so that clues of specular and shadow can be implicitly and fully considered for light estimation. Our contributions are summarized as

- We represent general reflectance as four intrinsic neural fields to implicitly associate per-pixel reflectance to light, which fully considers clues of specular reflectance and cast shadow for light estimation.
- We propose the NeIF, an uncalibrated photometric stereo method trained in an unsupervised manner, which works with general surface reflectance, flexible light settings, and training data bias.
- We show that our method achieves a superior performance over uncalibrated and unsupervised methods. We also demonstrate its excellent generalization capacity to data from different sources and promising performance with the challenging setup of sparse uncalibrated photometric stereo.

2 Related Work

In this section, we mainly review the latest works in neural reflectance representation, and related works unsupervised calibrated methods and uncalibrated photometric stereo methods. Readers may refer to [39] and [12] for a more comprehensive summary.

2.1 Neural Reflectance Representation

Neural Radiance Fields (NeRFs) [31] focuses on the 3D geometries without explicitly modeling the interaction between light and objects via photometric image formation model. Many subsequent works explore its application in various computer vision problems, such as relighting [14], photometric stereo [26], and multi-view photometric stereo [24]. These works either require known lighting conditions [**bi2022neural**, 43, 58, 26] or adopt simple reflectance model [56]. There are also some works [7, 57] considering general reflectance, but they require multi-view information and does not explicitly consider cast shadow effects. Kaya *et al.* [24] implemented NeRF in multi-view photometric stereo, but their work also leverages multi-view information. Different from these methods, our method considers general reflectance, cast-shadow effects, and unknown light conditions for uncalibrated photometric stereo by taking images captured at a single viewpoint.

2.2 Unsupervised Calibrated Photometric Stereo

Classical methods solved the calibrated photometric stereo problem without knowing the ground truth surface normal. Therefore, we classify them as unsupervised methods. The least square based algorithm [47] provides the simplest solution, which assumes the object to be Lambertian. The algorithm's performance is robust in a wide range of datasets, but its strong assumption on the reflectance model remains much space for improvement. The following works either regard the non-Lambertian reflectance components as the outliers [5, 16, 33, 48, 23, 49] or apply analytic reflectance models including Torrance-Sparrow[18], the Ward model[16], a mixture of multiple Ward models [19], [1] *etc.* to consider the non-Lambertian effects. However, the performance of those methods can only deal with limited types of materials. There are also more advanced methods that utilize the general refletance features such as reciprocity, isotropy [2], and monotonicity [21]. Those methods give a reliable estimation for objects with a broad range of materials. Inspired by them, our approach considers the general reflectance properties.

With the progress of deep learning, many learning-based frameworks have been proposed for calibrated photometric stereo. Taniai *et al.* [44] propose the first unsupervised learning-based photometric stereo method through a rendering equation. However, their reflectance model does not separately consider shadow, specular highlights, and diffuse components.

Different from these methods that need the known lights, we address the uncalibrated photometric stereo. We also train our method in an unsupervised manner. Because it is expected to be free from training data bias and thus obtain a stable estimation across data from different sources.

2.3 Uncalibrated Photometric Stereo

Previous works study the phenomenon where Lambertian objects under arbitrary directions can generate identical images when similar transformations are applied to both the object's normal and the light's direction. Belhumeur *et al.* [6] reduce the dimension of the transformation to a three parameters GBR transformation by considering integrability constraints. Based on that, extra clues on reflectance, such as half-vector symmetry [28], albedo clustering [42], specular spikes [53], or assumptions on light source distribution, such as ring light [60], symmetry light [32], or uniform-distributed light sources [54, 4, 42, 50, 34, 28], are used to resolve the GBR ambiguity. However, all these methods require the light intensity to be identical, which is inapplicable in the real-world datasets such as DiLiGenT [39], Apple & Gourd [2], and Light Stage Data Gallery [9]. Cho *et al.* [15] put up a semi-calibrated method to deal with non-uniform light intensity, but they assumed the light directions to be known. Quéau *et al.* [36] address the photometric stereo problem under inaccurate lighting calibration, while the accuracy could significantly drop when non-Lambertian components become dominant.

Recently, many deep learning methods have been proposed for uncalibrated photometric stereo. Chen *et al.* [12] propose a supervised uncalibrated framework, SDPS-Net, which can simultaneously estimate the light conditions (intensity and direction) and the surface normal. They suggest treating light estimation as a classification problem and separate the normal and light prediction to reduce the complexity. Their following work, GC-Net[13], improved the performance of SDPS-Net by adding shading as an extra channel to the input of the light estimation network. However, as a common problem for all supervised methods, an over-fitting problem may occur due to the training data bias [30]. In contrast, unsupervised methods do not have such a concern. Another obvious benefit is

that there is no need to synthesize training sets for unsupervised network training. To utilize the advantage of unsupervised methods, Kaya *et al.* [25] propose a compromised method that trains the light estimation network in a supervised manner (similar to [12]), but estimate the surface normal in an unsupervised way (similar to [44]. However, they still suffer from the training data bias during light estimation. Compared to previous works, our method neither makes a strict assumption on reflectance nor needs special light source distribution. Instead, we solve the uncalibrated photometric stereo problem in an unsupervised manner for general reflectance surfaces with unknown light source information.

3 Method

3.1 Problem Formulation

Given a set of observations $\boldsymbol{I} \triangleq (I_0, I_1, ..., I_f)$ of a static surface, illuminated by f unknown directional illuminations distributing on the upper-hemisphere, uncalibrated photometric stereo aims at recovering light directions $\boldsymbol{L} \triangleq (\boldsymbol{l}_0, \boldsymbol{l}_1, ..., \boldsymbol{l}_f)$, light intensities $\boldsymbol{E} \triangleq (e_0, e_1, ..., e_f)$, and surface normal $\boldsymbol{N} \triangleq \{\boldsymbol{n}_i | i \in \mathbb{P}\}$, $\mathbb{P}$ is the set of all positions on the surface. The solution is achieved by solving the optimization problem,

$$\underset{\boldsymbol{L},\boldsymbol{E},\boldsymbol{N}}{\arg\min} \sum_{i=1}^{\#\mathbb{P}} \sum_{j=1}^{f} \mathrm{D}(\bar{m}_{ij}, m_{ij}), \tag{1}$$

where $\bar{m}_{ij} \in I_j$ is the observed pixel intensity at position i, $\#\mathbb{P}$ is the number of elements in $\mathbb{P}$[1], m_{ij} is the corresponding rendered pixel intensity, $\mathrm{D}(\cdot,\cdot)$ is a metric describing their difference. Under an orthographic camera with linear radiometric response, m_{ij} is formulated as (simplified in a per-pixel form),

$$m_{ij} = e_j \rho(\boldsymbol{n}_i, \boldsymbol{l}_j, \boldsymbol{v}) \max(\boldsymbol{n}_i^\top \boldsymbol{l}_j, 0) = e_j \rho_{ij} \max(\boldsymbol{n}_i^\top \boldsymbol{l}_j, 0), \tag{2}$$

where $\boldsymbol{v} = [0, 0, 1]$ is the view direction pointing toward the viewer, ρ_{ij} describes the general reflectance, $\max(\boldsymbol{n}_i^\top \boldsymbol{l}_j, 0)$ represents the attach shadow.

Unknown light brings two ambiguities when solving Eq. (1), *i.e*, shape-light ambiguity, which is denoted as an invertable matrix $\boldsymbol{G} \in \mathbb{R}^{3\times3}$, and reflectance-light ambiguity, which is denoted as a non-zero scalar $c_j \in \mathbb{R}$,

$$m_{ij} = e_j (c_j c_j^{-1}) \rho_{ij} \max(\boldsymbol{n}_i^\top (\boldsymbol{G}\boldsymbol{G}^{-1}) \boldsymbol{l}_j, 0), 0). \tag{3}$$

3.2 Neural Intrinsics Fields

General reflectance decomposition. To exploit clues of specular reflectance and cast shadow for light estimation, we decompose the general reflectance as the cast shadow term s_{ij} multiplying the bidirectional reflectance term,

$$\rho_{ij} = s_{ij}(k_i^d \rho_i^d + k_i^s \rho_{ij}^s), \tag{4}$$

where subscript 'i' and 'j' indicate pixel position- (or normal-) and light-dependent factors, respectively, the cast shadow term s_{ij} is either 0 or 1, ρ_i^d, ρ_{ij}^s represent the diffuse and specular reflectance, and k_i^d and k_i^s are coefficients that balance out the effects of specular and diffuse reflectance[2].

Neural fields of specular reflectance and cast shadow. Previous methods identify specular features from *specific* pixels and associate an *explicit* relation to light for light estimation. However, this scheme could fail for the surface where the specular features are invisible or the explicit relation is violated. Besides, leaving out the clue of cast shadow could obstruct producing competitive performance.

In contrast, we leverage *both* clues of specular and shadow for light estimation, which is achieved by building the neural specular field and neural shadow field and associating the fields to light conditions. These neural fields make the utmost of *all* observed pixels, which fully exploits the mutual information across different normal-light (specular) and depth-light (shadow). Two MLPs, namely *SpecularNet* and *ShadowNet*, implement these neural fields, respectively. Instead of estimating light conditions

[1] Without loss of generality, we put '#' before a set symbol to represent its number of elements in this paper.

[2] We think these coefficients of a point will not change under different lights.

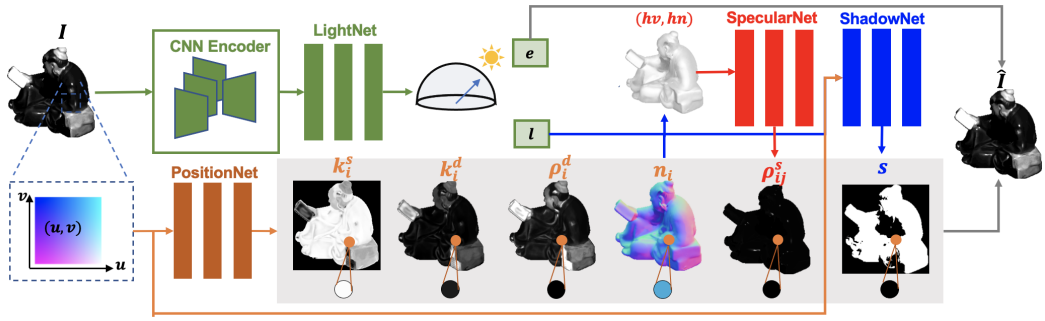

Figure 2: The framework of the proposed NeIF. LightNet takes the observed image $\boldsymbol{I}_j$ as the input and outputs light intensity e_j and direction $\boldsymbol{l}_j$. SpecularNet takes $\cos < \boldsymbol{h}_j, \boldsymbol{n}_i >$ and $\cos < \boldsymbol{h}_j, \boldsymbol{v} >$ as the input and outputs specular reflectance ρ_{ij}^s. PositionNet takes the input of positional code and outputs diffuse reflectance ρ_i^d, surface normal $\boldsymbol{n}_i$, and coefficients k_i^d, k_i^s. ShadowNet takes inputs of positional code and light direction $\boldsymbol{l}_j$ and output shadow indicator s_{ij}. All the intrinsics are used to render the observed pixel value m_{ij}.

from SpecularNet or ShadowNet, we take the light direction as the input of these networks to achieve their *implicit* association to light conditions. Since the specular reflectance and cast shadow are normal- and pixel position-dependent, we also feed the estimated surface normal and the positional code to them, respectively. SpecularNet and ShadowNet output ρ_{ij}^s and s_{ij}, respectively, as shown in Fig. 2.

Neural light field. There is mutual information across different observed images, *i.e*, observed images with similar appearances are expected to be illuminated by similar lights. To fully consider and exploit this mutual information, we build the neural light field by concatenating a CNN encoder to an MLP, namely *LightNet*. The encoder extracts a light code from each observed image[3], and the LightNet outputs the corresponding light conditions (*i.e*, $e_j, \boldsymbol{l}_j$) from the light code, as shown in Fig. 2.

Neural position field. There is mutual information across different positions on a surface, *i.e*, the consistency of shape and diffuse reflectance in the spatial domain. To fully consider and exploit this mutual information, we establish a neural position field, namely *PositionNet*, implemented by an MLP. The neural position field output position-dependent, light-independent factors[4], *i.e*, $\boldsymbol{n}_i, \rho_i^d, k_i^d, k_i^s$. The PositionNet takes the positional code as the input, as shown in Fig. 2.

3.3 Optimizing Neural Intrinsics Fields

We adopt the reconstruction loss function with the ℓ_1 metric to optimze our NeIF,

$$\mathcal{L}_{\text{rec}} = \sum_{i=1}^{\#\mathbb{P}} \sum_{j=1}^{f} |\bar{m}_{ij} - e_j s_{ij}(k_i^d \rho_i^d + k_i^s \rho_{ij}^s) \max(\boldsymbol{n}_i^\top \boldsymbol{l}_j, 0)|. \tag{5}$$

Silhouette constraint. $\mathcal{L}_{\text{rec}}$ cannot resolve the shape-light ambiguity in an unsupervised manner due to the inherently severe ill-posedness. Therefore, we introduce the silhouette constraint (similar to those in [**8758025**, 13]) to stabilize the training of PositionNet. To be specific, we use polynomial fitting with a moving window block to traverse and pre-compute the contour's normal of the given objects, represented as $\hat{\boldsymbol{N}}^{\text{si}} \triangleq \{\hat{\boldsymbol{n}}_k^{\text{si}} \in \mathbb{R}^{2\times 1}, k \in \mathbb{S}\}$, $\mathbb{S}$ is the point set of the contour. We consider $\hat{\boldsymbol{N}}^{\text{si}}$ could guide the prediction of the azimuth of boundary normal (at the same positions) and

[3]We use light code instead of positional code because the light code contains discriminative features of light intensity. Please refer to our supplementary material for the correlation analysis on pixel and light intensity.

[4]Although specular reflectance and cast shadow are pixel position-dependent, they are also light-dependent. Considering them requires the input of light condition which increases the complexity of the neural field.

introduce the silhouette loss function,

$$\mathcal{L}_{\text{si}} = \sum_{k=1}^{\#\mathbb{S}} |\text{Nor}(\text{C}(\boldsymbol{n}_k)) - \hat{\boldsymbol{n}}_k^{\text{si}}|, \tag{6}$$

where $\boldsymbol{n}_k \in \mathbb{R}^{1\times 3}$ represents the estimated surface normal from PositionNet, $\text{C}(\cdot)$ cuts off the 3rd dimension of $\boldsymbol{n}_k$ (*i.e*, $\text{C}(\boldsymbol{n}_k) \in \mathbb{R}^{1\times 2}$), and $\text{Nor}(\cdot)$ is the vector normalization operation.

Warm-up loss functions. To avoid local minimum and achieve faster convergence, we warm up the NeIF in early-stage during training with three additional loss functions. We use the azimuth of lighting direction estimated by YS97 [54] to guide the training of LightNet,

$$\mathcal{L}_{\text{az}} = \frac{1}{f}\sum_{j=1}^{f} |\text{Nor}(\text{C}(\boldsymbol{l}_j)) - \text{Nor}(\text{C}(\boldsymbol{l}_j^{\text{az}}))|_2, \tag{7}$$

where $\boldsymbol{l}_j^{\text{az}}$ are estimated light directions by [54]. We adopt the gradient penalty [20] $\mathcal{L}_{\text{gp}}$ to stabilize the training of SpecularNet,

$$\mathcal{L}_{\text{gp}} = \text{ReLU}(-\nabla_{\boldsymbol{n}_i^\top \boldsymbol{h}_j} \rho_{ij}^s). \tag{8}$$

The intuition is from Blinn-Phong model, where the specular reflectance is monotonically increasing w.r.t the $\boldsymbol{n}_i^\top \boldsymbol{h}_j$, *i.e*, $\nabla_{\boldsymbol{n}_i^\top \boldsymbol{h}_j} \rho_{ij}^s > 0$. We also supervise the training of ShadowNet using pseudo shadow maps $\hat{\boldsymbol{S}}_j, j = 1, 2, ..., f$,

$$\mathcal{L}_{\text{shadow}} = \frac{1}{\#\mathbb{P}} \sum_{j=1} f|\boldsymbol{S}_j - \hat{\boldsymbol{S}}_j|_2. \tag{9}$$

The pseudo shadow maps are obtained by binarizing the observed images, *i.e*, considering an observed pixel to be cast shadow if its intensity value is smaller than $0.2\times$ the mean intensity values of this image. After early-stage training, we discard these loss functions for a broad range of reflectance.

NeIF training. We train NeIF with the warm-up loss function in first 10 epochs,

$$\mathcal{L}_{\text{warmup}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{si}}\mathcal{L}_{\text{si}} + \lambda_{\text{az}}\mathcal{L}_{\text{az}} + \lambda_{\text{gp}}\mathcal{L}_{\text{gp}} + \lambda_{\text{shadow}}\mathcal{L}_{\text{shadow}}, \tag{10}$$

where $\lambda_{\text{si}} = 5, \lambda_{\text{az}} = 0.1, \lambda_{\text{gp}} =, \lambda_{\text{shadow}} = 10$. We then train NeIF untill 500 epochs or converging with the loss function,

$$\mathcal{L}_{\text{NeIF}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{si}}\mathcal{L}_{\text{si}} + \lambda_{\text{shadow}}\mathcal{L}_{\text{recShadow}}, \tag{11}$$

where $\mathcal{L}_{\text{recShadow}}$ is the another shadow map supervison loss to train the ShadowNet. The loss function is the same to $\mathcal{L}_{\text{shadow}}$. The only difference is the calculation of $\hat{\boldsymbol{S}}_j$. For $\mathcal{L}_{\text{recShadow}}$, we calculate $\hat{\boldsymbol{S}}_j$ by rendering a depth map with the estimated $\boldsymbol{l}_j$. The depth map is reconstructed from the estimated surface normal map $\boldsymbol{N}$ by method [8][5]. $\mathcal{L}_{\text{recShadow}}$ is used to align outputs of ShadowNet to those of PositionNet.

Implementation details. We generate the positional code from the coordinate (in the image plane) by the same method in [31]. With the assumption of isotropic reflectance, we simplify the input of SpecularNet from $\{\boldsymbol{n}_i^\top \boldsymbol{l}_j, \boldsymbol{n}_i^\top \boldsymbol{v}, \boldsymbol{l}_j^\top \boldsymbol{v}\}$ to $\{\boldsymbol{n}_i^\top \boldsymbol{h}_j, \boldsymbol{h}_j^\top \boldsymbol{v}\}$[6] [28] for easier model training. Similar to the Cook-Torrance reflectance model [17], we assume $k_i^d + k_i^s = 1$ to reduce the numbers of unknowns. The CNN decoder with declining channels processes the down-sampled images with a dimension of 256×256. The LightNet takes flatten features to predict $\boldsymbol{l}_j, e_j$ in two different branches. For the ShadowNet, we also generate the positional code for $\boldsymbol{l}_j$ and concatenate it to the feature in 9-th layer. The output of ShadowNet is either 0 or 1, which is realized by a step function with a similar implementation in [35][7].

4 Experiments

Training details. Our main framework is implemented in PyTorch, while the pre-calculations (method [54] and silhouette fitting) are implemented in MATLAB. We use Adam as the optimizer

[5] Please find the details in the supplementary material.

[6] $\boldsymbol{h}_j$ is the bisector of $\boldsymbol{l}_j$ and $\boldsymbol{v}$, $\boldsymbol{h}_j = \frac{\boldsymbol{l}_j + \boldsymbol{v}}{\|\boldsymbol{l}_j + \boldsymbol{v}\|}$.

[7] Please refer to our supplementary material for more details

Table 1: Quantitative comparison in terms of mean angular error for surface normal on DILIGENT benchmark. This table summarizes comparison methods. 'Semi' indicates certain method leverage partial information of light. '✓' (or '✗') represents that certain methods (do not) adopt supervised learning for normal $\boldsymbol{n}$ or light $\boldsymbol{l}$ estimation. 'Identical' means certain methods require the light intensity of different illuminating to be identical.

Method	PS/ UPS	$\boldsymbol{n}$ Supervision	$\boldsymbol{l}$ Supervision	e Supervision	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	AVG
LS[47]	PS	✗	N.A.	N.A.	4.10	8.39	14.92	8.41	25.6	18.5	30.62	8.89	14.65	19.8	16.64
TM18[44]	PS	✗	N.A.	N.A.	1.47	5.79	10.36	5.44	6.32	11.47	22.59	**6.09**	7.76	**11.03**	8.83
CH18(16)[11]	PS	✗	N.A.	N.A.	3.31	7.47	8.76	7.64	8.78	9.81	17.48	8.14	8.22	14.09	9.37
YS97[54]	UPS	✗	✗	Identical	39.12	41.30	43.02	39.1	47.18	42.25	79.56	42.94	41.88	41.06	45.74
AM07[4]	UPS	✗	✗	Identical	7.27	16.81	32.81	31.45	54.72	46.54	61.70	18.37	49.16	53.65	37.25
SM10[42]	UPS	✗	✗	Identical	8.90	11.98	15.54	19.84	22.73	48.79	73.76	16.68	50.68	26.93	29.58
WT13[50]	UPS	✗	✗	Identical	4.39	6.42	13.19	36.55	19.75	20.57	55.51	9.39	14.52	58.96	23.93
LM13[29]	UPS	✗	✗	Identical	22.43	15.44	25.76	25.01	22.53	29.16	34.45	32.82	20.57	48.16	27.63
PF14[34]	UPS	✗	✗	Identical	4.77	9.07	14.92	9.54	19.53	29.93	29.21	9.51	15.90	24.16	16.65
LC17[28]	UPS	✗	✗	Identical	9.30	10.90	19.00	12.60	15.00	18.30	28.00	12.40	15.70	22.30	16.35
CH18[11]	UPS	✓	✗	✓	3.96	7.19	13.06	12.16	11.84	18.07	27.22	11.13	11.11	20.46	13.62
CH19[12]	UPS	✓	✓	✓	2.77	6.89	8.97	8.06	8.48	11.91	17.43	8.14	7.50	14.90	9.51
CW20[13]	UPS	✓	✓	✓	2.50	5.60	**8.60**	7.90	7.80	9.60	**16.20**	7.20	7.10	14.90	8.71
CM20[15]	Semi	✗	Known	✗	2.78	8.07	13.38	8.05	26.9	18.18	33.35	9.47	19.58	14.19	15.4
KK21[25]	UPS	✗	✓	✓	3.78	5.96	13.14	7.91	10.85	11.94	25.49	8.75	10.17	18.22	11.62
SK22[37]	UPS	✓	✓	✓	3.46	5.48	10.00	8.94	**6.04**	9.78	17.97	7.76	7.1	15.02	9.15
Ours(16)	UPS	✗	✗	✗	2.34	4.52	10.33	7.87	7.68	11.23	25.07	8.00	6.35	20.42	10.38
Ours(96)	UPS	✗	✗	✗	**1.15**	**4.41**	8.78	**5.08**	6.14	**9.49**	17.68	7.94	**6.12**	11.82	**7.86**

with a learning rate $\alpha = 5 \times 10^{-4}$ to train our framework in 500 epochs for each scene separately, and the warm-up stage takes up ten epochs. The last 100 epochs use a lower learning rate $\alpha = 5 \times 10^{-5}$ for fine-tunning. The batch size is 32 for lighting and 256 for spatially random sampling. After every epoch, the depth map is reconstructed according to the predicted normal map by method [8]. Each scene takes from 2 hours to 6 hours on one RTX 2080Ti 12GB GPU, depending on the resolution of the objects.

Evaluation metrics. We adopt the same metric in [12], the scale-invariant relative error, to measure the accuracy of recovered light intensity as

$$E_{\text{int}} = \frac{1}{f} \sum_{j=1}^{f} \left(\frac{|\eta e_j - \tilde{e}_j|}{\tilde{e}_j} \right) \tag{12}$$

where, η is calculated by solving $\arg\min_{\eta} \sum_{j=1}^{f} (\eta e_j - \tilde{e}_j)^2$ by least squares minimization. The metric to measure the accuracy of the predicted light directions and surface normal is the widely used mean angle error (MAE) in degree.

4.1 Evaluation on Public Datasets

Since the proposed NeIF is an unsupervised uncalibrated photometric stereo method, we compare its performance with state-of-the-art uncalibrated and unsupervised photometric stereo methods. Three real-world datasets, including the benchmark dataset DILIGENT [39], APPLE & GOURD dataset [2], and the LIGHT STAGE DATA GALLERY [9], are used for evaluation.

Quantitative comparison for normal map. Table 1 lists relevant works for a comprehensive surface normal estimation comparison. As summarized in Table 1, our method is the only method that addresses UPS without the supervision of $\boldsymbol{N}$, $\boldsymbol{L}$, or $\boldsymbol{E}$. However, our method achieves the best performance and maintains a considerable advantage over other competitors[12, 13] (*i.e*, 7.86 for NeIF v.s. 8.71 for [13]). Objects in DILIGENT differ from shapes and reflectance. The numbers of our method are either best or competitive (less than 1° as compared with the best performing UPS method), which shows its good generalization capacity to general reflectance and various shapes. Because our NeIF fully considers mutual information by building up intrinsics fields and such implicit modeling facilitates modeling general reflectance.

Visual quality comparison for normal map. Fig. 3 illustrates the visual quality comparison in terms of recovered surface normal maps and corresponding error maps on DILIGENT. The comparison is conducted with two state-of-the-art UPS methods[12, 13] and a state-of-the-art unsupervised PS method [44]. Our method is less sensitive to spatially-varying albedo due to the per-pixel manner (red boxes in the left subfigure of Fig. 3). However, although the positional code considers the

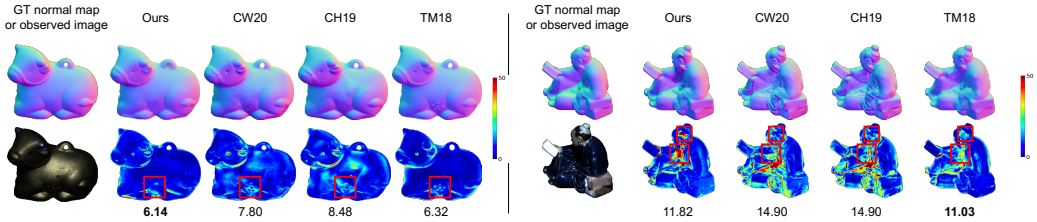

Figure 3: Visual quality comparison in terms of normal map and error map.

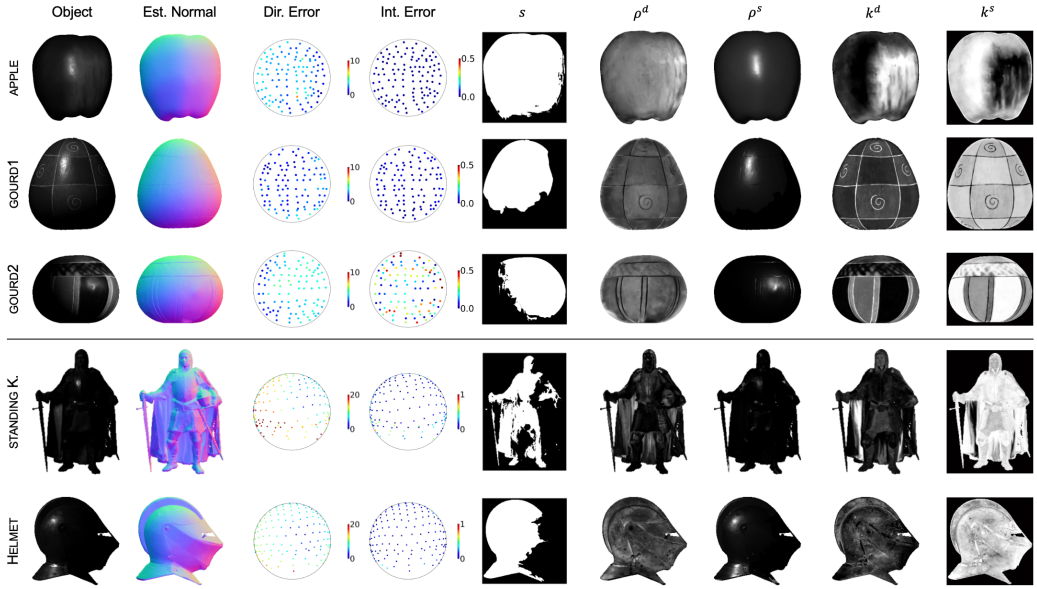

Figure 4: From left to right: the reference image, estimated surface normal map from our method, the lighting direction error, light intensity error, the estimated shadow map, diffuse reflectance map, specular reflectance map, diffuse scaling coefficient map, and specular coefficient map. Top three objects are from APPLE & GOURD [2], and bottom two objects are from LIGHT STAGE DATA GALLERY [9].

global shape effect, this per-pixel manner is less effective in modeling complex shape information as compared with the all-pixel one. It could fail for regions with cast shadow (or overexposure) under most light directions (red boxes in the right subfigure of Fig. 3). We also show the visual quality results for APPLE & GOURD dataset [2] and the LIGHT STAGE DATA GALLERY [9] in Fig. 4[8]. Our method could produce reliable estimation for most regions, thanks to the full exploitation of mutual information across different dimensions.

Quantitative comparison for light directions and intensities. As can be observed in Table 2 and Table 3, our method achieves a superior performance advantage over unsupervised methods[34, 4, 42, 50, 29, 28] while maintaining competitive performance as compared with supervised methods. Although the supervised methods pre-define the intensity range, they produce less accurate estimation as compared with our approach due to the possible overfitting of supervised learning. In contrast, our LightNet builds the neural light field, which could fully consider the mutual information among different observed images, resulting in the accurate estimation of the scale-invariant light intensity.

4.2 Evaluation on Sparse Uncalibrated Photometric Stereo

As compared with other per-pixel PS methods (*e.g.*, [59, 27]), the proposed NeIF exploits much more constraints to estimate the normal at a point, *i.e*, $\#\mathbb{P} \times f$ v.s. f. Because it learns fields instead of regressing intensity profile [40] to surface normal. Therefore, we investigate the effectiveness of our

[8]More visual quality illustration and comparison could be found in our supplementary material.

Table 2: Quantitative comparison in terms of light direction and intensity errors on DILIGENT benchmark [39].

Model	BALL		BEAR		BUDDHA		CAT		COW		GOBLET		HARVEST		POT1		POT2		READING		AVG	
	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.
YS97[54]	12.41	0.334	14.06	0.260	11.68	0.300	13.75	0.318	15.79	0.251	15.24	0.316	59.41	0.586	12.99	0.322	12.58	0.283	13.08	0.266	18.10	0.320
PF14[34]	4.90	0.036	5.24	0.098	9.76	0.053	5.31	0.059	16.34	0.074	33.22	0.223	24.99	0.156	2.43	**0.017**	13.52	0.044	21.77	0.122	13.75	0.088
CH19[12]	3.27	0.039	3.47	0.061	4.34	0.048	4.08	0.095	4.52	0.073	10.36	0.067	6.32	0.082	5.44	0.058	2.87	0.048	**4.50**	0.105	4.92	0.068
CW20[13]	1.75	**0.027**	**2.44**	0.101	2.86	**0.032**	4.58	0.075	**3.15**	**0.031**	**2.98**	0.042	**5.74**	0.065	**1.41**	0.039	**2.81**	0.059	5.47	0.048	**3.32**	0.052
Ours(16)	2.83	0.012	4.14	0.014	5.77	0.033	8.45	0.038	8.91	0.047	10.09	0.062	6.44	**0.041**	6.16	0.062	3.50	0.026	15.85	0.136	7.21	0.050
Ours(96)	**1.69**	0.030	3.96	**0.010**	**1.73**	**0.032**	**2.92**	**0.021**	4.98	0.050	6.82	**0.040**	7.06	**0.032**	3.33	0.134	3.71	**0.028**	7.45	**0.042**	4.37	**0.042**

Table 3: Quantitative comparison in terms of light direction and intensity errors on APPLE & GOURD and LIGHT STAGE DATA GALLERY.

Model	APPLE		GOURD1		GOURD2		AVG		STANDING KNIGHT		HELMET		AVG	
	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.	dir.	int.
PF14[34]	6.68	0.109	21.23	0.096	25.87	0.329	17.92	0.178	33.81	1.311	25.40	0.576	46.26	2.198
CH19[12]	9.31	0.106	4.07	0.048	7.11	**0.186**	6.83	0.113	11.60	0.286	6.57	0.212	13.61	0.221
CW20[13]	10.91	0.094	4.29	0.042	7.13	0.199	7.44	0.112	**5.31**	0.198	**5.33**	0.096	9.20	0.163
Ours	**2.65**	**0.011**	**1.76**	**0.029**	**3.21**	0.230	**2.54**	**0.090**	13.38	**0.189**	8.12	**0.082**	10.75	**0.135**

method under the challenging setting of sparse UPS. We randomly select 16 images illuminated by different lights from DILIGENT and test our method using these images. We repeat this process 30 times, similar to [59]. We compare the performance with that from the classical PS method PF14 [34], as deep learning-based approaches (*e.g.*, [12, 13]) perform poorly without re-training. Due to the fewer reconstruction terms to train our NeIF, we slightly increase the dimension of the positional encoding module from 4 to 6 by increasing the frequency [45] to stabilize the training, *i.e*, strengthening the role of positional code to reduce the variance of estimated intrinsics. Besides, we drop $\mathcal{L}_{\text{az}}$ during early stage warm-up because YS97 [54] fails for objects with centralized specular components (*e.g.*, BALL).

We report the mean results over 30 trails in Table 1. As can be observed, our method achieves promising performance on light and normal estimation for sparse UPS.[9]

5 Conclusion

This paper proposes NeIF for UPS. By representing the general reflectance as four neural intrinsics fields, it implicitly associates the light clues of specualr reflectance and cast shadow to light estimation, which faciliate solving UPS with general reflectance. The proposed NeIF could fully exploit mutual informace from all observed pixel values so that it could produce stable estimation even for sparse UPS. The unsupervised training manner of NeIF is beneficial to the generalization capacity to data from different sources.[10]

Limitations. Although our method produces promising results for light condition and surface normal estimation, it has several limitations. First, the estimated intrinsics such as shadow, specular and diffuse reflectance, and their balancing coefficients are less accurate, as shown in Fig. 1. Second, we need to balance the role of positional code and estimated intrinsics when applying our method for sparse UPS. However, how to find an optimal balancing strategy is unknown. Third, as shown in Fig. 3, our method is less effective on global effects, especially for regions with cast shadow (or overexposure) under most light directions due to the per-pixel manner. Besides, the inference of our method is inefficient due to the unsupervised training manner and the per-pixel estimation, which is inapplicable for real-time applications.

References

[1] Jens Ackermann et al. "Photometric stereo for outdoor webcams". In: *2012 IEEE conference on computer vision and pattern recognition*. IEEE. 2012, pp. 262–269.

[9] More results and comparison could be found in our supplementary material.

[10] More ablation studies could be found in our supplementary material.

[2] Neil Alldrin, Todd Zickler, and David Kriegman. "Photometric stereo with non-parametric and spatially-varying reflectance". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2008, pp. 1–8.

[3] Neil G Alldrin and David J Kriegman. "Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach". In: *2007 IEEE 11th International Conference on Computer Vision*. IEEE. 2007, pp. 1–8.

[4] Neil G Alldrin, Satya P Mallick, and David J Kriegman. "Resolving the generalized bas-relief ambiguity by entropy minimization". In: *2007 IEEE conference on computer vision and pattern recognition*. IEEE. 2007, pp. 1–7.

[5] Svetlana Barsky and Maria Petrou. "The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.10 (2003), pp. 1239–1252.

[6] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. "The bas-relief ambiguity". In: *International journal of computer vision* 35.1 (1999), pp. 33–44.

[7] Mark Boss et al. "NeRD: Neural Reflectance Decomposition From Image Collections". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Oct. 2021, pp. 12684–12694.

[8] Xu Cao et al. "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 (CVPR)*. June 2021, pp. 2382–2391.

[9] Charles-Félix Chabert et al. "Relighting human locomotion with flowed reflectance fields". In: *ACM SIGGRAPH 2006 Sketches*. 2006, 76–es.

[10] Manmohan Krishna Chandraker, Fredrik Kahl, and David J Kriegman. "Reflections on the generalized bas-relief ambiguity". In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. IEEE. 2005, pp. 788–795.

[11] 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)*. Sept. 2018.

[12] Guanying Chen et al. "Self-calibrating deep photometric stereo networks". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 8739–8747.

[13] Guanying Chen et al. "What is learned in deep uncalibrated photometric stereo?" In: *European Conference on Computer Vision*. Springer. 2020, pp. 745–762.

[14] Xingyu Chen et al. "Hallucinated Neural Radiance Fields in the Wild". In: *arXiv preprint arXiv:2111.15246* (2021).

[15] Donghyeon Cho et al. "Semi-calibrated photometric stereo". In: *IEEE transactions on pattern analysis and machine intelligence* 42.1 (2018), pp. 232–245.

[16] Hin-Shun Chung and Jiaya Jia. "Efficient photometric stereo on glossy surfaces with wide specular lobes". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2008, pp. 1–8.

[17] Robert L Cook and Kenneth E. Torrance. "A reflectance model for computer graphics". In: *ACM Transactions on Graphics (ToG)* 1.1 (1982), pp. 7–24.

[18] Athinodoros S Georghiades. "Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo". In: *Computer Vision, IEEE International Conference on*. Vol. 3. IEEE Computer Society. 2003, pp. 816–816.

[19] Dan B Goldman et al. "Shape and spatially-varying brdfs from photometric stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.6 (2009), pp. 1060–1071.

[20] Ishaan Gulrajani et al. "Improved training of wasserstein gans". In: *Advances in neural information processing systems* 30 (2017).

[21] Tomoaki Higo, Yasuyuki Matsushita, and Katsushi Ikeuchi. "Consensus photometric stereo". In: *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE. 2010, pp. 1157–1164.

[22] Michael Holroyd et al. "A photometric approach for estimating normals and tangents". In: *ACM Transactions on Graphics (TOG)* 27.5 (2008), pp. 1–9.

[23] Satoshi Ikehata et al. "Robust photometric stereo using sparse regression". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 318–325.

[24] Berk Kaya et al. “Neural Radiance Fields Approach to Deep Multi-View Photometric Stereo”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2022, pp. 1965–1977.

[25] Berk Kaya et al. “Uncalibrated neural inverse rendering for photometric stereo of general surfaces”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 3804–3814.

[26] Junxuan Li and Hongdong Li. “Neural Reflectance for Shape Recovery with Shadow Handling”. In: *arXiv preprint arXiv:2203.12909* (2022).

[27] Junxuan Li et al. “Learning to minify photometric stereo”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7568–7576.

[28] Feng Lu et al. “SymPS: BRDF symmetry guided photometric stereo for shape and light source estimation”. In: *IEEE transactions on pattern analysis and machine intelligence* 40.1 (2017), pp. 221–234.

[29] Feng Lu et al. “Uncalibrated Photometric Stereo for Unknown Isotropic Reflectances”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2013.

[30] Brian Mac Namee et al. “The problem of bias in training data in regression problems in medical decision support”. In: *Artificial intelligence in medicine* 24.1 (2002), pp. 51–70.

[31] Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In: *European conference on computer vision*. Springer. 2020, pp. 405–421.

[32] Kazuma Minami et al. “Symmetric-Light Photometric Stereo”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2022, pp. 2706–2714.

[33] Daisuke Miyazaki, Kenji Hara, and Katsushi Ikeuchi. “Median photometric stereo as applied to the segonko tumulus and museum objects”. In: *International Journal of Computer Vision* 86.2 (2010), pp. 229–242.

[34] Thoma Papadhimitri and Paolo Favaro. “A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima”. In: *International journal of computer vision* 107.2 (2014), pp. 139–154.

[35] Haotong Qin et al. “Binary neural networks: A survey”. In: *Pattern Recognition* 105 (2020), p. 107281.

[36] Yvain Quéau et al. “A non-convex variational approach to photometric stereo under inaccurate lighting”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 99–108.

[37] Francesco Sarno et al. “Neural Architecture Search for Efficient Uncalibrated Deep Photometric Stereo”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2022, pp. 361–371.

[38] Steven M Seitz et al. “A comparison and evaluation of multi-view stereo reconstruction algorithms”. In: *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06)*. Vol. 1. IEEE. 2006, pp. 519–528.

[39] Boxin Shi et al. “A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 3707–3716.

[40] Boxin Shi et al. “Bi-polynomial modeling of low-frequency reflectances”. In: *IEEE transactions on pattern analysis and machine intelligence* 36.6 (2013), pp. 1078–1091.

[41] Boxin Shi et al. “Elevation angle from reflectance monotonicity: Photometric stereo for general isotropic reflectances”. In: *European Conference on Computer Vision*. Springer. 2012, pp. 455–468.

[42] Boxin Shi et al. “Self-calibrating photometric stereo”. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. 2010, pp. 1118–1125.

[43] Pratul P. Srinivasan et al. “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)*. June 2021, pp. 7495–7504.

[44] Tatsunori Taniai and Takanori Maehara. “Neural inverse rendering for general reflectance photometric stereo”. In: *International Conference on Machine Learning*. PMLR. 2018, pp. 4857–4866.

[45] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[46] Eva Weigl et al. “Photometric stereo sensor for robot-assisted industrial quality inspection of coated composite material surfaces”. In: *Twelfth International Conference on Quality Control by Artificial Vision 2015*. Vol. 9534. SPIE. 2015, pp. 367–374.

[47] Robert J Woodham. “Photometric method for determining surface orientation from multiple images”. In: *Optical engineering* 19.1 (1980), pp. 139–144.

[48] Lun Wu et al. “Robust photometric stereo via low-rank matrix completion and recovery”. In: *Asian Conference on Computer Vision*. Springer. 2010, pp. 703–717.

[49] Tai-Pang Wu et al. “Dense photometric stereo: A markov random field approach”. In: *IEEE transactions on pattern analysis and machine intelligence* 28.11 (2006), pp. 1830–1846.

[50] Zhe Wu and Ping Tan. “Calibrating photometric stereo by holistic reflectance symmetry analysis”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 1498–1505.

[51] Limin Xie et al. “A practical means for calibrating an LED-based photometric stereo system”. In: *Optics and Lasers in Engineering* 64 (2015), pp. 42–50.

[52] Wuyuan Xie, Zhan Song, and Ronald C Chung. “Real-time three-dimensional fingerprint acquisition via a new photometric stereo means”. In: *Optical Engineering* 52.10 (2013), p. 103103.

[53] Sai-Kit Yeung et al. “Normal estimation of a transparent object using a video”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.4 (2014), pp. 890–897.

[54] Alan Yuille and Daniel Snow. “Shape and albedo from multiple images using integrability”. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE. 1997, pp. 158–164.

[55] Guangjun Zhang and Zhenzhong Wei. “A novel calibration approach to structured light 3D vision inspection”. In: *Optics & Laser Technology* 34.5 (2002), pp. 373–380.

[56] Jason Zhang et al. “NeRS: Neural reflectance surfaces for sparse-view 3d reconstruction in the wild”. In: *Advances in Neural Information Processing Systems* 34 (2021).

[57] Kai Zhang et al. “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)*. June 2021, pp. 5453–5462.

[58] Xiuming Zhang et al. “Nerfactor: Neural factorization of shape and reflectance under an unknown illumination”. In: *ACM Transactions on Graphics (TOG)* 40.6 (2021), pp. 1–18.

[59] Qian Zheng et al. “SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 8549–8558.

[60] Zhenglong Zhou and Ping Tan. “Ring-light photometric stereo”. In: *European Conference on Computer Vision*. Springer. 2010, pp. 265–279.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Line 9-13, 39-66
   (b) Did you describe the limitations of your work? [Yes] See Line 287-295
   (c) Did you discuss any potential negative societal impacts of your work? [No]
   (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? [Yes]
   (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] See Supplementary Material
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (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]
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? [No]
   (b) Did you mention the license of the assets? [No]
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No]
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]