LERPS: LIGHTING ESTIMATION AND RELIGHTING FOR PHOTOMETRIC STEREO

Ashish Tiwari *Shanmuganathan Raman*

CVIG Lab, Indian Institute of Technology Gandhinagar
{ashish.tiwari, shanmuga}@iitgn.ac.in

ABSTRACT

Photometric stereo is a method to obtain surface normals of an object using its images captured under varying illumination directions. The existing deep learning-based methods require multiple images of an object captured using complex image acquisition systems. In this work, we propose a deep learning framework to perform three tasks jointly: *(i) lighting estimation*, *(ii) image relighting*, and *(iii) surface normal estimation*, all from a single input image of an object with non-Lambertian surface and general reflectance. The network explicitly segregates global geometric features and local lighting-specific features of the object from a single image. The local features resemble attached shadows, shadings, and specular highlights, providing valuable lighting estimation and relighting cues. The global features capture the lighting-independent geometric attributes that effectively guide the surface normal estimation. The joint training transfers valuable insights to achieve significant improvements across all three tasks. We show that the proposed single-image-based relighting framework outperforms several existing photometric stereo methods which require multiple images of a static object.

***Index Terms*—** Photometric Stereo, Image-based Relighting.

1. INTRODUCTION

Overview. Photometric stereo estimates surface orientations of a static object imaged under different lightings with a constant viewing direction [1, 2]. Lighting (light direction and intensity) estimation and scene relighting have been studied for a long time in the computer vision and graphics community for better 3D understanding. Photometric stereo has applications in visual effects, virtual and augmented reality, and e-commerce. While there has been significant progress in the methods for relighting the natural outdoor scenes [3, 4], indoor scenes [5], and portraits [6, 7, 8], none of them have addressed the problem in a photometric stereo setup. In general, surface normal estimation becomes challenging due to global illumination effects like shadows and inter-reflections. Inspired by their success, several works have adopted deep learning approaches for photometric stereo [9, 10, 11, 12, 13]. However, the existing learning-based methods rely on a large number of images. While taking multiple images of objects under different lightings helps in understanding their surface geometry and the reflectance properties, they require bulky image acquisition systems such as light stage [14]. Such acquisition setups sparsely sample the light space due to the limited light sources (on a spherical dome). Therefore, these methods fail to produce photorealistic images outside the limited light source locations.

Contributions. (i) In this work, we propose a multi-purpose deep neural network called *Lighting Estimation and Relighting for Photometric Stereo (LERPS)* that estimates lighting, relights the image under target novel lightings, and estimates per-pixel surface normals, all from a single image. (ii) Through necessary feature visualization, we demonstrate that the network learns to segregate the global (lighting-independent) geometric features and the local lighting-specific features using multiple images with different lightings provided one at a time. During inference, it requires just a single image to perform all three tasks. The proposed joint training transfers valuable insights to achieve significant improvements across all three tasks. While the local features provide valuable lighting estimation and relighting cues [15], the global features are shown to provide the lighting-independent geometric attributes for the network to guide the surface normal estimation effectively. (iii) The image relighting results show that these images can be used for data augmentation for photometric stereo-based applications.

2. RELATED WORK

The literature for photometric stereo is very diverse. In this section, we discuss only the deep learning-based approaches and refer the reader to [2] for an exhaustive survey on classical photometric stereo techniques. Deep neural network for non-Lambertian photometric stereo was first introduced in *Deep Photometric Stereo Network (DPSN)* [13]. It adopts dropout in a fully connected network to capture the effects of cast shadows and regress the surface normals. However, the number-and order-fixed manner of estimating surface normals limits its practicality. For better handling of the non-Lambertian objects and taking full advantage of the neighborhood information, subsequent methods deployed the convolutional neural networks (CNN) [9, 11, 10, 16, 17]. The works [11] and [18] use channel max-pooling over a fully-convolutional network (FCN) to estimate surface normals from an arbitrary number of input images. CNN-PS [9] introduced an observation map to tackle the restriction of fixed inputs. Wang *et al.* [19] proposes to decouple the surface normal from the reflectance function using an inverse reflectance model with collocated lightings. Yao *et al.* [20] proposed a graph-based approach that unifies per-pixel and all-pixel approaches to incorporate both inter-image and intra-image information. UPS-FCN [11], SDPS-Net [10] and GCNet [15] address the photometric stereo problem in an uncalibrated setting. SDPS-Net uses a two-stage approach: Light Calibration Network (LCNet) for lighting estimation and Normal Estimation Network (NENet) for normal estimation using the lightings estimated by LCNet. GCNet explicitly leverages surface normals and shading information for improved lighting estimation. We take a step further by relighting images under arbitrary lightings along with estimating the associated lightings and further estimating surface normals. Some methods combine image reconstruction with surface normal estimation [21] for non-Lambertian surfaces. However, none of these methods have addressed surface normal estimation with image-based relighting.

[1]This work is supported by SERB IMPRINT-2 grant.

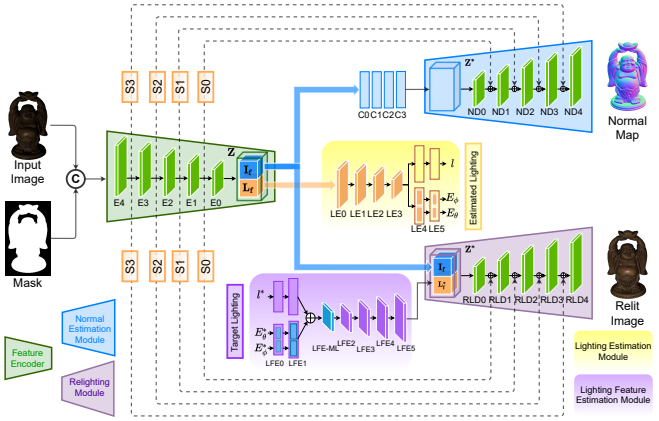

Fig. 1. Detailed network architecture of LERPS designed for lighting estimation, image relighting, and normal estimation from a single image. The design of the network is inspired by the hourglass network [22]

3. METHOD

3.1. Image Formation Model

Consider an anisotropic non-Lambertian surface f characterised by the Bidirectional Reflectance Distribution Function (BRDF) ρ imaged with an orthographic camera under directional lighting. For each surface point (x, y) with normal $\boldsymbol{n} \in \mathbb{R}^3$ illuminated by the j^{th} incoming lighting with direction $\boldsymbol{\ell}_j \in \mathbb{R}^3$ and intensity $e_j \in \mathbb{R}$, the image formation of the surface viewed from the direction $\boldsymbol{v} \in \mathbb{R}^3$ is given by Equation 1.

$$\boldsymbol{I}_j(x, y) = e_j \rho(\boldsymbol{n}, \boldsymbol{\ell}_j, \boldsymbol{v}) \psi_{f,s}(x, y) \left[\boldsymbol{n}^T \boldsymbol{\ell}_j\right] + \epsilon_j \tag{1}$$

Here, $\psi_{f,s}(x, y)$ specifies the attached and cast shadows. It is equal to 0, if (x, y) is shadowed and equal to 1, otherwise. ϵ_j represents the global illumination effects and noise effect.

3.2. Network Architecture

We propose a multi-purpose deep learning framework called *Lighting Estimation and Relighting for Photometric Stereo (LERPS)* that jointly performs lighting estimation, image relighting, and surface normal estimation. The detailed architecture is provided in Fig. 1. The design of the lighting estimation and relighting module follows that of [8]. The normal estimation module is similar to that of relighting module. However, it is trained with different weights.

(a) Feature Encoder: It pools feature $\mathbf{Z}$ from the image of an object (concatenated with a binary mask). The feature $\mathbf{Z}$ is further divided into two features: the global object-level feature $\mathbf{I_f}$ and the local lighting-specific feature $\mathbf{L_f}$. While the existing methods [10, 11] adopt a feature fusion strategy to combine local and global features from multiple images, LERPS achieves the same (in the form of $\mathbf{Z}$) from a single image itself.

(b) Lighting Estimation Module: The local lighting-specific feature $\mathbf{L_f}$ is fed to the *lighting estimation module* to determine the lighting information: direction $\mathbf{E}_\phi$ (azimuth), $\mathbf{E}_\theta$ (elevation), and intensity $\boldsymbol{\ell}$. We consider a discretized light space[1] and model the lighting estimation as a classification problem, similar to that of [10].

[1]The azimuth angle $\phi \in [0°, 180°]$ and the elevation angle $\theta \in [-90°, 90°]$ describing the upper-hemisphere are divided into $K_d = 36$ bins. The intensities (in the range [0.2, 2]) are divided into $K_e = 20$ bins.

(c) Relighting Module: The specified target lightings $\mathbf{E}^*_\phi$, $\mathbf{E}^*_\theta$, and $\boldsymbol{\ell}^*$ are mapped to the new lighting feature map $\mathbf{L}^*_\mathbf{f}$ using *lighting feature estimation module*. The global object-level feature $\mathbf{I_f}$ and the new lighting feature $\mathbf{L}^*_\mathbf{f}$ are then concatenated to form $\mathbf{Z}^*$ which is fed to the *relighting module* to obtain the relit image $\mathbf{I}^*$.

(d) Normal Estimation Module: It takes global object-level feature $\mathbf{I_f}$ passed through a set of four convolutional blocks as input and is designed to obtain the surface normal map. The skip connections provide information flow between the *feature encoder* and *relighting* and *normal estimation module* at the respective scales.

3.3. Loss Functions

The network jointly optimizes the lighting estimation loss $\mathcal{L}_{\text{light}}$, the image relighting loss $\mathcal{L}_{\text{relight}}$, and the normal loss $\mathcal{L}_{\text{normal}}$, as described in Equation 2.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{light}} + \mathcal{L}_{\text{relight}} + \mathcal{L}_{\text{normal}} \tag{2}$$

Lighting Estimation Loss: Following [10], we use multi-class cross entropy loss [23] for lighting estimation, as described in Equation 3.

$$\mathcal{L}_{\text{light}} = \lambda_{l_a} \mathcal{L}_{l_a} + \lambda_{l_e} \mathcal{L}_{l_e} + \lambda_e \mathcal{L}_e \tag{3}$$

Here, $\mathcal{L}_{l_a}$, $\mathcal{L}_{l_e}$, and $\mathcal{L}_e$ are the loss terms (cross-entropy loss) for the azimuth direction ($\mathbf{E}^*_\phi$), elevation direction ($\mathbf{E}^*_\theta$), and the light intensity ($\boldsymbol{\ell}$), respectively. The weights $\lambda_{l_a} = 1.0$, $\lambda_{l_e} = 1.0$, and $\lambda_e = 2.0$ are used for model training.

Image Relighting Loss: The total image relighting loss ($\mathcal{L}_{\text{relight}}$) described in Equation 4.

$$\mathcal{L}_{\text{relight}} = \lambda_{L_1} \mathcal{L}_{L_1} + \lambda_{\text{GAN}} \mathcal{L}_{\text{GAN}} + \lambda_{\text{FM}} \mathcal{L}_{\text{FM}} \tag{4}$$

Here, $\lambda_{L_1} = 1.0$, $\lambda_{\text{GAN}} = 1.0$, and $\lambda_{\text{FM}} = 0.5$ are the weights for L_1-loss ($\mathcal{L}_{L_1}$), GAN loss ($\mathcal{L}_{\text{GAN}}$), and feature matching loss ($\mathcal{L}_{\text{FM}}$) terms, respectively. The loss $\mathcal{L}_{L_1}$ in turn is a conjunction of L_1-losses over images (target image I_t and predicted image I^*_t) and gradient images (∇I_t and ∇I^*_t), as given in Equation 5.

$$\mathcal{L}_{L_1} = \lambda_{img} \frac{1}{N_I} (\parallel I_t - I^*_t \parallel_1) + \lambda_{grad} \frac{1}{N_I} (\parallel \nabla I_t - \nabla I^*_t \parallel_1) \tag{5}$$

Here, $\lambda_{img} = 1.0$ and $\lambda_{grad} = 1.0$ are the weights for L_1 image loss and image gradient loss, respectively. The gradient loss preserves the edges and avoids blurring. We use *Least Square Generative Adversarial Network (LS-GAN)* [24] for GAN loss ($\mathcal{L}_{\text{GAN}}$) as described in Equation 6.

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{\text{I}_\text{t}}(D(I_t))^2 + \mathbb{E}_{\text{I}_\text{s}}(1 - D(G(I_s, L_t))^2) \tag{6}$$

Here, G and D represent a generator (*feature encoder*) and a ResNet [25] based discriminator network, respectively. In order to ensure that the images of the same object under different lightings have same global geometrical features, we also enforce a feature matching loss $\mathcal{L}_{\text{FM}}$, as described in Equation 7.

$$\mathcal{L}_{\text{FM}} = \frac{1}{N_F} (\mathbf{I_{f_1}} - \mathbf{I_{f_2}})^2 \tag{7}$$

Here, $\mathbf{I_{f_1}}$ and $\mathbf{I_{f_2}}$ are geometric features of two input images $\mathbf{I_1}$ and $\mathbf{I_2}$ of the same object under different lightings, and N_F is the number of elements in $\mathbf{I_f}$. In the proposed design, $\mathcal{L}_{\text{FM}}$ allows the network to segregate the lighting-independent geometrical object feature $\mathbf{I_f}$ and the lighting feature $\mathbf{L_f}$ (Section 4.1).

Normal Loss: The per-pixel surface normals are obtained by minimizing the normal loss, as described in Equation 8.

$$\mathcal{L}_{\text{normal}} = \frac{1}{N}\sum_{i=1}^{N}(1 - \boldsymbol{n}_i^T \widetilde{\boldsymbol{n}}_i) \tag{8}$$

Here, N is the number of pixels, and $\boldsymbol{n}_i$ and $\widetilde{\boldsymbol{n}}_i$ are the ground truth and estimated normals at the i^{th} pixel, respectively.

3.4. Training Details

We adopt the widely used Blobby and Sculpture dataset [11] containing complex surface normal distributions and diverse real-world materials for training using Adam optimizer [26]. These surfaces are obtained from the MERL dataset [27] containing 100 BRDFs. We train the network for 30 epochs by randomly sampling 64 different pairs of images (of the same object) under different lightings per sample in each epoch.

4. EXPERIMENTAL ANALYSIS

In this section, we perform qualitative and quantitative analysis of the proposed network on both real and synthetic datasets.

Test Dataset. The test dataset includes the DiLiGenT benchmark [28] consisting of objects with non-Lambertian surfaces. Further, it also includes the *Sphere* and the *Bunny* shape rendered using the Mitsuba raytracer [29] with 100 BRDFs from the MERL dataset [27] under 100 different light directions.

Evaluation Metrics. We adopt the popularly used *Mean Angular Error (MAE)* (in degree) for measuring the accuracy of the predicted light directions and surface normals such that $\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left(\cos^{-1}(\boldsymbol{z}_i^T \widetilde{\boldsymbol{z}}_i)\right)$. Here, $\boldsymbol{z}_i$ and $\widetilde{\boldsymbol{z}}_i$ are the ground truth and predicted vectors, respectively, for the i^{th} image.

The scale-invariant *relative error* is used for measuring the accuracy of light intensities such that $E_{\text{err}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|(s\times\widetilde{e}_i - e_i|}{e_i}\right)$ Here, N is the number of images, $\widetilde{e}_i$ and e_i are the predicted and ground truth light intensities, respectively, for the i^{th} image. The scale factor s is obtained by minimizing $\sum_{i=1}^{n}\left((s\times\widetilde{e}_i - e_i\right)^2$ using least-squares. For assessing the quality of relit images, we use the average *Relative Error (REL)* and the *Structural Similarity Index (SSIM)* [30].

4.1. Effect of Loss Terms and Visualizing Learned Features

Table 4.1 validates that LERPS trained with the proposed combination of loss terms (ID3) obtains the best performance in lighting estimation and image relighting (high SSIM and low REL) on the DiLiGenT benchmark.

ID	Loss function	MAE ↓ (direction)	Relative error ↓ (intensity)	SSIM ↑	REL ↓
ID1	$\mathcal{L}_{\text{L}_1}$	6.142	0.0957	0.879	0.212
ID2	$\mathcal{L}_{\text{L}_1}$+$\mathcal{L}_{\text{GAN}}$	4.067	0.0681	0.941	0.176
ID3	$\mathcal{L}_{\text{L}_1}$+$\mathcal{L}_{\text{FM}}$+$\mathcal{L}_{\text{GAN}}$	**3.341**	**0.0553**	**0.953**	**0.154**

Table 1. Effect of loss terms on the DiLiGenT benchmark.

Fig. 2 shows some representative features selected from the global object-level feature $\mathbf{I_f}$ and the local lighting-specific feature $\mathbf{L_f}$ (each of 256-channel dimension) learned by the network. The features corresponding to $\mathbf{L_f}$ are highly correlated with the scene attributes such as attached shadows, shadings, and specular highlights. These attributes provide strong cues for lighting estimation. However, they are not directly available as input in practice. Therefore, $\mathbf{L_f}$ becomes crucial in learning these attributes and guiding the accurate lighting estimation. The features $\mathbf{I_f}$ capture the global geometrical structure of the object (independent of lighting) owing to the constraint imposed through $\mathcal{L}_{\text{FM}}$. We believe this is because $\mathcal{L}_{\text{FM}}$ forces the relit images of the same object to have similar global features and preserve the geometrical information in the relit images.

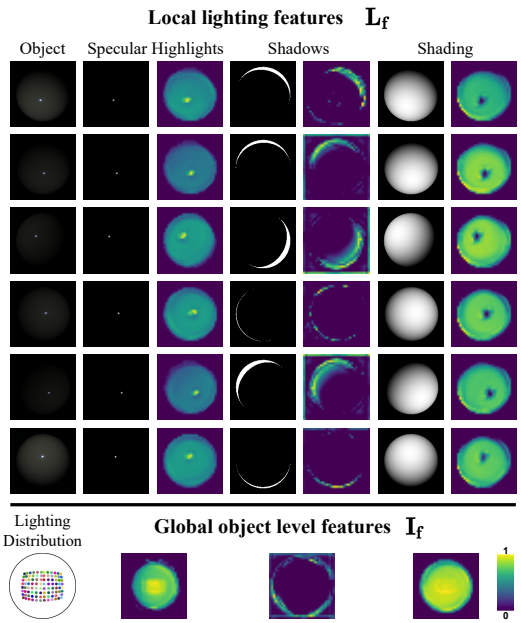

Fig. 2. Visualising the features learned by LERPS on the object BALL with a non-Lambertian surface. All the features are normalized to [0, 1] and color coded.

4.2. Results on lighting estimation and relighting

Table 2 and 3 quantify the network performance on test datasets for lighting estimation (in comparison with LCNet and GCNet) and relighting, respectively. LERPS clearly outperforms LCNet on the test datasets for lighting estimation. However, it's performance is comparable to that of GCNet on the DiLiGenT benchmark. Since there hasn't been any work on relighting over photometric data, we report the SSIM and REL obtained using LERPS in Table 3. Note that the values are averaged over 96 random image pairs (and associated lightings) per object for relighting. On average, the relit images have over 95% structural similarity with the desired target images on DiLiGenT benchmark. Fig. 3 shows the qualitative results for samples drawn from the test datasets. Our method is able to handle cast and attached shadows (CAT with regions marked in green color) and complex specularities (red-phenolic, BALL, and HARVEST) to a reasonable extent.

4.3. Results on Normal Estimation

Table 4 shows the average MAE on the DiLiGenT benchmark dataset in comparison with recent deep learning based methods (with 96 input images) and some traditional non-deep learning

Dataset	LCNet [10]		GCNet [15]		LERPS (ours)	
	MAE ↓	Rel. Error ↓	MAE ↓	Rel. Error ↓	MAE ↓	Rel. Error ↓
Sphere	3.47	0.082	2.61	0.074	**2.48**	**0.071**
Bunny	5.38	0.089	4.01	0.085	**3.97**	**0.082**
DiLiGenT	4.92	0.068	**3.32**	**0.052**	3.34	0.055

Table 2. Average MAE (direction) and Relative error (intensity) values on test datasets.

Dataset	SSIM ↑	REL ↓
Sphere	0.961	0.146
Bunny	0.957	0.159
DiLiGenT	0.953	0.154

Table 3. Quantitative results on image relighting.

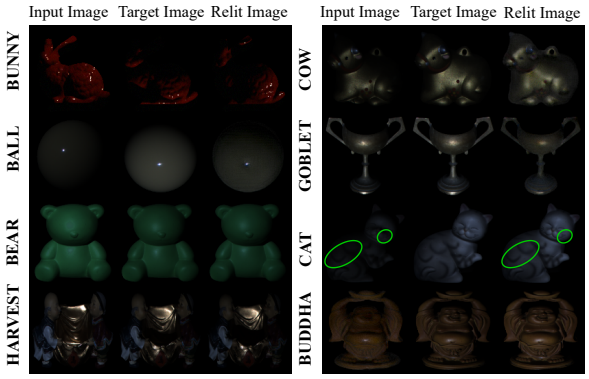

Fig. 3. Qualitative results on image relighting.

based methods. Our method ranks best with an average MAE of 7.89° and a standard deviation of ± 0.24 units. We report the standard deviation since for each object in the DiLiGenT benchmark dataset, there are 96 images under different lightings and each one can be used as an input to the network. In Fig. 4, we show the qualitative comparison of the existing methods on the DiLiGenT benchmark dataset. Unlike our single-image-based method, these methods take a total of 96 images as input. The error maps produced by LERPS show a relatively lower MAE (see the marked regions), especially for complex objects like HARVEST and READING, and even over lesser complex surfaces like COW and POT. Although our single image-based method is comparable to multi-image-based methods, we do not deny the fact that multiple images can provide more information. Our sole attempt is to show how an appropriate network design can help learn a significant amount of information from a single image itself.

DL-based methods	UPS-FCN [11]	UPS-FCN*	DPSN [11]	SDPS-Net [10]	SDPS-Net*	GCNet+ PS-FCN [15]	GCNet+ PS-FCN	CNN-PS [9]	DR-PSN [21]	LERPS
Average MAE	16.02	16.46	9.41	9.51	9.49	8.75	9.05	8.52	7.90	**7.89** ±0.24
Traditional Methods	AM07 [31]		SM10 [32]	WT13 [33]		LM13 [34]		PF14 [35]	LC18 [36]	
Average MAE	37.25		29.59	23.93		27.63		16.66	16.30	

Table 4. Quantitative comparison of different methods over the DiLiGenT Benchmark Dataset. * indicates the testing performed with relit images generated by our method.

5. CONCLUSION

We introduce a deep framework, called LERPS, for joint single-image-based lighting estimation, relighting, and normal estimation. With necessary feature visualisations, we observe that the network captures and disentangles the global lighting-independent and the local lighting-specific features of the object. We compare the performance of the proposed method with several calibrated and uncalibrated photometric stereo methods through extensive qualitative and quantitative evaluation on both synthetic and real datasets.

Fig. 4. Qualitative results on normal estimation over the DiLiGenT benchmark. Regions are marked in red color and MAE is reported below each normal map for better comparison.

6. REFERENCES

[1] Robert J Woodham, "Photometric method for determining surface orientation from multiple images," *Optical engineering*, vol. 19, no. 1, pp. 191139, 1980.

[2] Jens Ackermann, Michael Goesele, et al., "A survey of photometric stereo techniques," *Foundations and Trends® in Computer Graphics and Vision*, vol. 9, no. 3-4, pp. 149–254, 2015.

[3] Xin Jin, Yannan Li, Ri Xu, Xiaokun Zhang, and Xiaodong Li, "A method of single reference image based scene relighting," *MethodsX*, vol. 5, pp. 933–938, 2018.

[4] Zhuo Hui, Ayan Chakrabarti, Kalyan Sunkavalli, and Aswin C Sankaranarayanan, "Learning to separate multiple illuminants in a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3780–3789.

[5] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi, "Deep image-based relighting from optimal sparse samples," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–13, 2018.

[6] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras, "Portrait lighting transfer using a mass transport approach," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1, 2017.

[7] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi, "Single image portrait relighting," *ACM Transactions on Graphics (Proceedings SIGGRAPH)*, 2019.

[8] Christoph Rhemann, Graham Fyffe, Jay Busch, Jonathan T Barron, Paul Debevec, Ravi Ramamoorthi, Tiancheng Sun, Xueming Yu, Yun-Ta Tsai, and Zexiang Xu, "Single image portrait relighting," 2019.

[9] Satoshi Ikehata, "Cnn-ps: Cnn-based photometric stereo for general non-convex surfaces," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–18.

[10] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong, "Self-calibrating deep photometric stereo networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8739–8747.

[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)*, 2018, pp. 3–18.

[12] Yakun Ju, Kin-Man Lam, Yang Chen, Lin Qi, and Junyu Dong, "Pay attention to devils: A photometric stereo network for better details," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020, pp. 694–700.

[13] 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*, 2017, pp. 501–509.

[14] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar, "Acquiring the reflectance field of a human face," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 145–156.

[15] Guanying Chen, Michael Waechter, Boxin Shi, Kwan-Yee K Wong, and Yasuyuki Matsushita, "What is learned in deep uncalibrated photometric stereo?," in *European Conference on Computer Vision*. Springer, 2020, pp. 745–762.

[16] 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*, 2019, pp. 7568–7576.

[17] 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*, 2019, pp. 8549–8558.

[18] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee Kenneth Wong, "Deep photometric stereo for non-lambertian surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[19] Xi Wang, Zhenxiong Jian, and Mingjun Ren, "Non-lambertian photometric stereo network based on inverse reflectance model with collocated light," *IEEE Transactions on Image Processing*, vol. 29, pp. 6032–6042, 2020.

[20] Zhuokun Yao, Kun Li, Ying Fu, Haofeng Hu, and Boxin Shi, "Gps-net: Graph-based photometric stereo network," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[21] Yakun Ju, Junyu Dong, and Sheng Chen, "Recovering surface normal and arbitrary images: A dual regression network for photometric stereo," *IEEE Transactions on Image Processing*, vol. 30, pp. 3676–3690, 2021.

[22] Alejandro Newell, Kaiyu Yang, and Jia Deng, "Stacked hourglass networks for human pose estimation," in *European conference on computer vision*. Springer, 2016, pp. 483–499.

[23] Claude Elwood Shannon, "A mathematical theory of communication," *ACM SIGMOBILE mobile computing and communications review*, vol. 5, no. 1, pp. 3–55, 2001.

[24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros, "Image-to-image translation with conditional adversarial networks," *CVPR*, 2017.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[26] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[27] Wojciech Matusik, *A data-driven reflectance model*, Ph.D. thesis, Massachusetts Institute of Technology, 2003.

[28] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan, "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.

[29] Wenzel Jakob, "Mitsuba renderer, 2010," 2010.

[30] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[31] 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.

[32] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan, "Self-calibrating photometric stereo," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 1118–1125.

[33] 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.

[34] Feng Lu, Yasuyuki Matsushita, Imari Sato, Takahiro Okabe, and Yoichi Sato, "Uncalibrated photometric stereo for unknown isotropic reflectances," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1490–1497.

[35] Thoma Papadhimitri and Paolo Favaro, "A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima," *International journal of computer vision*, vol. 107, no. 2, pp. 139–154, 2014.

[36] Feng Lu, Xiaowu Chen, Imari Sato, and Yoichi Sato, "Symps: Brdf symmetry guided photometric stereo for shape and light source estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 1, pp. 221–234, 2017.