

Article

ReDDLE-Net: Reflectance Decomposition for Directional Light Estimation

Jiangxin Yang^{1,2}, Binjie Ding^{1,2}, Zewei He^{1,2,3}, Gang Pan⁴, Yanpeng Cao^{1,2}, Yanlong Cao^{1,2,*} and Qian Zheng^{4,*}

¹ State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China

² Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, School of Mechanical Engineering, Zhejiang University, Hangzhou 310027, China

³ School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China

⁴ College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China

* Correspondence: sdcaoyl@zju.edu.cn (Y.C.); qianzheng@zju.edu.cn (Q.Z.)

Abstract: The surfaces of real objects can visually appear to be glossy, matte, or anywhere in between, but essentially, they display varying degrees of diffuse and specular reflectance. Diffuse and specular reflectance provides different clues for light estimation. However, few methods simultaneously consider the contributions of diffuse and specular reflectance for light estimation. To this end, we propose ReDDLE-Net, which performs **R**eflectance **D**ecomposition for **D**irectional **L**ight **E**stimation. The primary idea is to take advantage of diffuse and specular clues and adaptively balance the contributions of estimated diffuse and specular components for light estimation. Our method achieves a superior performance advantage over state-of-the-art directional light estimation methods on the DiLiGenT benchmark. Meanwhile, the proposed ReDDLE-Net can be combined with existing calibrated photometric stereo methods to handle uncalibrated photometric stereo tasks and achieve state-of-the-art performance.

Keywords: uncalibrated photometric stereo; diffuse reflectance; specular reflectance; convolutional neural network



Citation: Yang, J.; Ding, B.; He, Z.; Pan, G.; Cao, Y.; Cao, Y.; Zheng, Q. ReDDLE-Net: Reflectance Decomposition for Directional Light Estimation. *Photonics* **2022**, *9*, 656. <https://doi.org/10.3390/photonics9090656>

Received: 11 August 2022

Accepted: 9 September 2022

Published: 15 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Photometric stereo aims at estimating a surface through a series of images, which are illuminated by different directional lights, with a fixed camera [1] (near-field photometric stereo [2–4] is not considered in this paper). In practice, the application of calibrated photometric stereo methods [5,6] is restricted by the tedious light source calibration procedure. In contrast, uncalibrated photometric stereo (UPS) technology can estimate the surface normal from the observed images without prior knowledge of light directions and intensities. Hence, it has recently attracted increasing attention from both industrial and academic research communities. The calibration of light directions is more laborious than that of the light intensity. There are numerous semi-calibrated photometric stereo methods [3,7,8] that can accurately estimate the light intensity with captured images and light directions. Thus, this paper aims at the estimation of the light direction.

Without a priori knowledge of light information, there is a 3×3 linear ambiguity [9] between the estimated normal and the normal ground truth. If the surface is integrable, the ambiguity can be degraded to a three-parameter generalized bas-relief (GBR) ambiguity [10,11]. Most of the existing traditional methods [12–14] for UPS are based on the Lambertian assumption and focus on resolving the GBR ambiguity. Methods such as those of [15], and [16] can handle surfaces with general bidirectional reflectances (BRDFs), which only work under assumptions of uniformly distributed light sources.

Recently, deep-learning-based methods have been introduced to the UPS task. Chen et al. [5] tried to directly learn the surface normals from observed images and

proposed a one-stage model called UPS-FCN. Later, Chen et al. [17] presented a two-stage model called self-calibrating deep photometric stereo networks (SDPS-Net), which consisted of a light calibration network (LCNet) and a normal estimation network (NENet). Then, they extended the SDPS-Net by introducing shading as the supervision in the light calibration task and presented a guided calibration network (GCNet), which performed light calibration for spatially varying albedo objects [18]. Kaya et al. [19] proposed an uncalibrated photometric stereo method without using the ground-truth surface normal. They first trained a light estimation network, which was supervised by the ground truth, and then estimated the surface normal through an inverse rendering network in an unsupervised way. We found that the state-of-the-art UPS methods are generally based on two-stage networks. They first estimated the light information from captured images. Acquired light information was concatenated with image data and fed into a normal estimation network. Following this strategy, there is a convenient but efficient way to provide a framework for UPS tasks by combining the directional light estimation module with existing calibrated photometric stereo methods. A number of calibrated photometric stereo methods have been proposed to improve the performance of normal estimation for non-Lambertian surfaces [5,20–31], and they have achieved great performance on both real datasets [32–34] and synthetic datasets [5,25]. As the light estimation error will accumulate in a normal estimation network, improving the accuracy of light direction estimation is expected to improve the performance of the overall UPS framework.

However, there are few methods focusing on the improvement of the directional light estimation. The reflective component of real-world objects varies with their surface characteristics. Diffuse and specular components have been experimentally proven to be effective clues for learning-based light direction estimation methods [18], and these two kinds of components show themselves differently. Specifically, diffuse reflectance exhibits slow variations in its intensity and obeys Lambert's cosine law, i.e., the illumination of a surface is directly proportional to the cosine of the angle between the normal and the direction of incident light. Specular reflectance is more concentrated and sharp, and it is often described as a mirror-like reflectance of light from a surface. The normal vector of an illuminating point in specular reflectance is on the angle bisector of the incident light direction and viewing direction. Diffuse and specular reflectance provides different and valid clues. However, existing learning-based methods implicitly process them together in the network and focus on a single type of reflectance information during the feature extraction phase, as shown in Figure 1.

To address this problem, we propose a reflectance decomposition (ReD) mechanism to take full advantage of the reflectance information and adjust the attention of diffuse and specular reflectance information for light direction estimation, as shown in Figure 1. We incorporate the ReD mechanism into a two-branch deep convolutional neural network, which extracts diffuse and specular components through two sub-networks, and we present a **Re**flEctance **D**ecomposition **D**irectional **L**ight **E**stimation Network (ReDDLE-Net). ReDDLE-Net can be seamlessly integrated into the existing photometric stereo methods for UPS tasks. In summary, the contributions of this paper are as follows. (1) We propose the ReDDLE-Net, which explicitly considers the reflectance property of an object for directional light estimation. (2) We present a solution for building the training data with labels of reflectance properties based on the retroreflective response. (3) We demonstrate that the proposed ReDDLE-Net can adaptively adjust the attention for diffuse and specular components according to the inputs and that it outperforms state-of-the-art methods regarding light estimation and the follow-up tasks of uncalibrated photometric stereo.

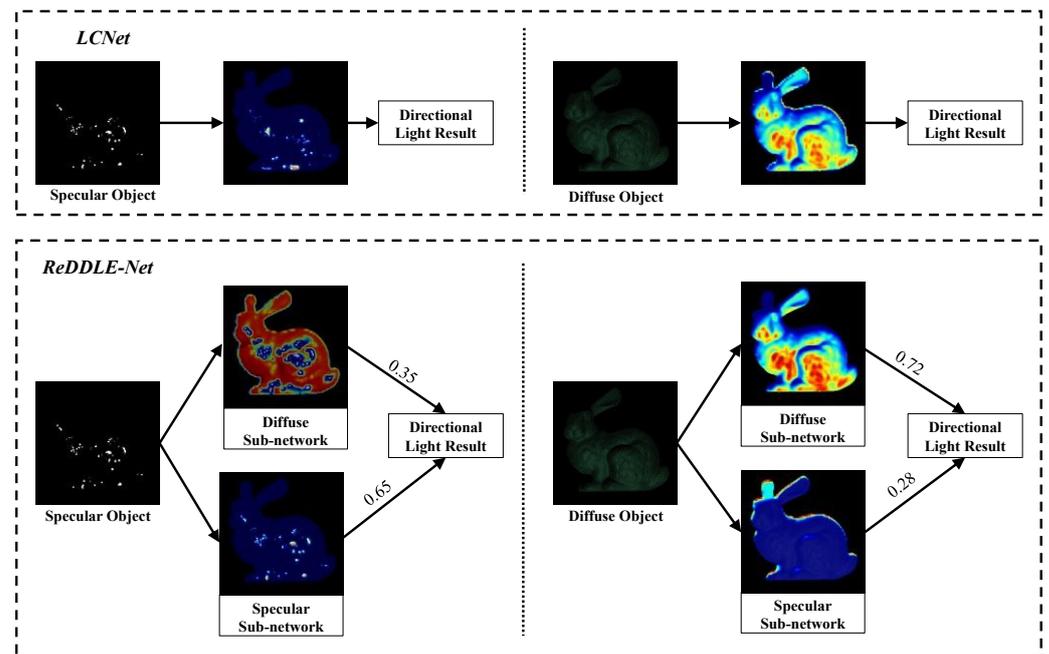


Figure 1. A comparison of the workflows of LCNet and ReDDLE-Net. LCNet uses a single-branch structure to extract features, while our ReDDLE-Net uses two-branch feature extractors (diffuse sub-network and specular sub-network). The number indicates the reflectance weight calculated by the reflectance decomposition (ReD) mechanism of ReDDLE-Net, which represents the contributions of the diffuse sub-network and specular sub-network. Top left: For a specular object, the LCNet focuses only on the highlighted part. Bottom left: ReDDLE-Net extracts diffuse and specular information using two separate sub-networks. The ReD mechanism gives a larger reward to the estimated light from the specular branch than that from the diffuse branch. Top right: For a diffuse object, the LCNet focuses on the diffuse part. Bottom right: When the object’s surface is diffuse, the ReD mechanism pays more attention to the diffuse component and improves the contribution of the diffuse branch.

2. Method

Following the operation of SDPS-Net [17], the azimuth and elevation of the light direction space are divided into 36 bins. Thus, light direction estimation can be converted into a classification task. The azimuth and elevation of a light direction are estimated separately. Figure 2 shows the proposed ReDDLE-Net framework.

2.1. Data Preparation

The dichromatic reflectance model describes how surface reflectance is composed of diffuse and specular reflectance. Diffuse reflectance (e.g., shading) exhibits small variations in its intensity. On the contrary, specular reflectance (e.g., highlighting) is more concentrated and intense. Chen et al. [18] found that both highlights and shadings are useful clues in deep-learned-based light estimation. Intending to provide a priori knowledge of reflectance to ReDDLE-Net during the training phase, we classified the 100 MERL materials [35] into three categories (i.e., diffuse, specular, and mixed samples).

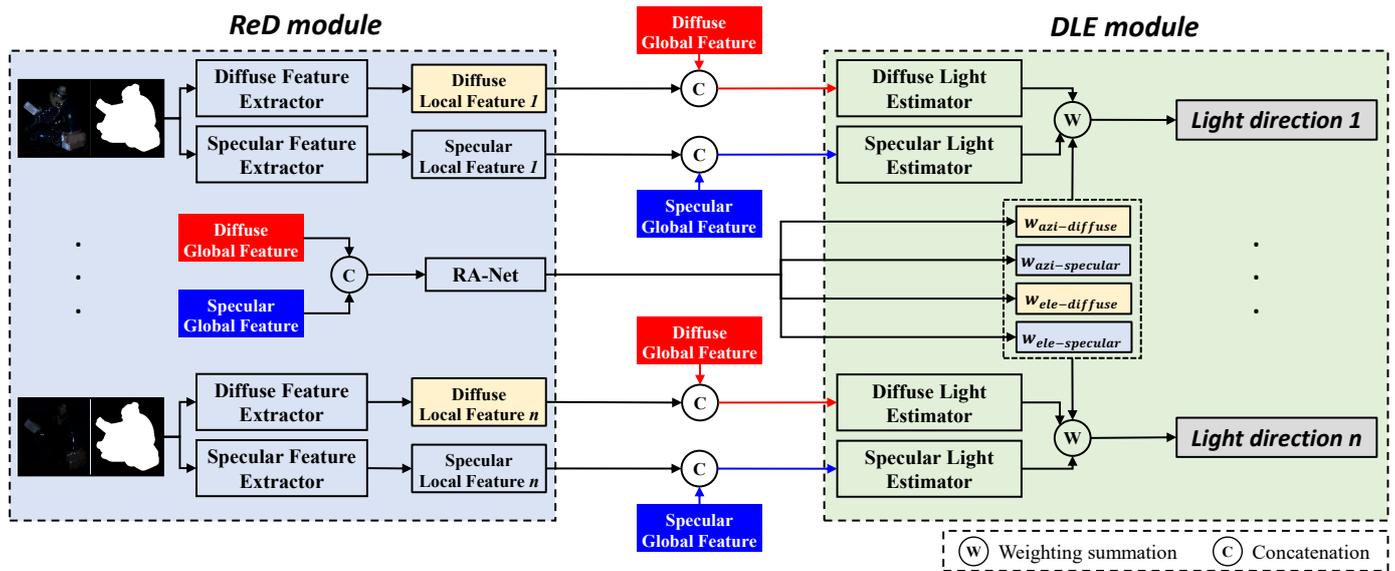


Figure 2. The proposed ReDDLE-Net framework. Our ReDDLE-Net consists of two integrated processing modules, namely, the reflectance decomposition module (ReD module) and the directional light estimation module (DLE module). The ReD module is used to obtain the local diffuse and specular features from pairs of images and masks. Then, max-pooling is utilized to aggregate local features and acquire global features for the diffuse and specular components, respectively (the details of feature aggregation can be found in Figure 3). Taking the concatenation of local features and global features as the input, the reflectance-aware network (RA-Net) produces the reflectance weights, which indicate the proportions of diffuse and specular components. The DLE module fuses the outputs of the ReD module, including the local features, global features, and reflectance weight, to generate the final light direction estimation results.

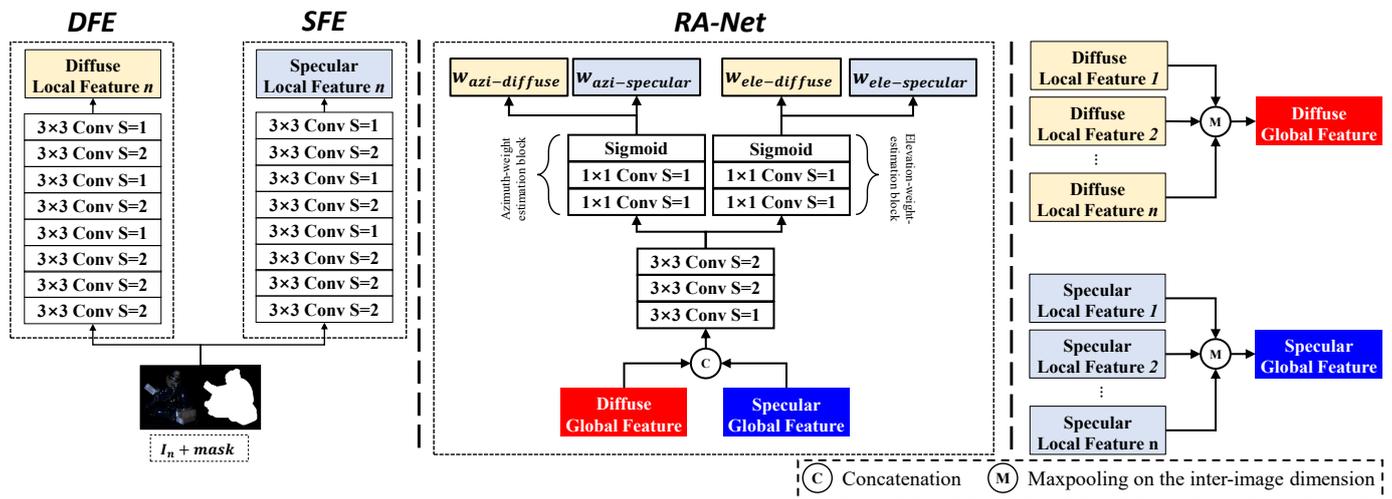


Figure 3. The proposed reflectance decomposition (ReD) module. **Left:** the network structure of the diffuse feature extractor (DFE) and specular feature extractor (SFC). **Middle:** The network structure of the reflectance-aware network (RA-Net). **Right:** Schematic diagram of the acquisition of global features.

The retroreflective response $f(\theta)$ is defined as the value of the bidirectional reflection distribution function (BRDF) when the view direction v coincides with the direction of incident light l [36]. θ is the angle between the normal and the half-vector h , which is defined as $h = \frac{l+v}{|l+v|}$. The peak near $\theta = 0$ is called the specular peak $f(0)$, which can be

seen as an indicator of surface roughness. The roughness of the surface is a crucial factor that affects the proportion of reflective components on the surface [36]. As the surface of the object becomes smoother, the characteristics of the specular component will gradually become apparent, and vice versa. According to [36], the materials were always divided into two groups (smooth and rough) based on the specular peak (*smooth* : $f(0) > 0.5$, *rough* : $f(0) < 0.5$). As shown in Figure 4, the specular peak values of different materials have a broad range.

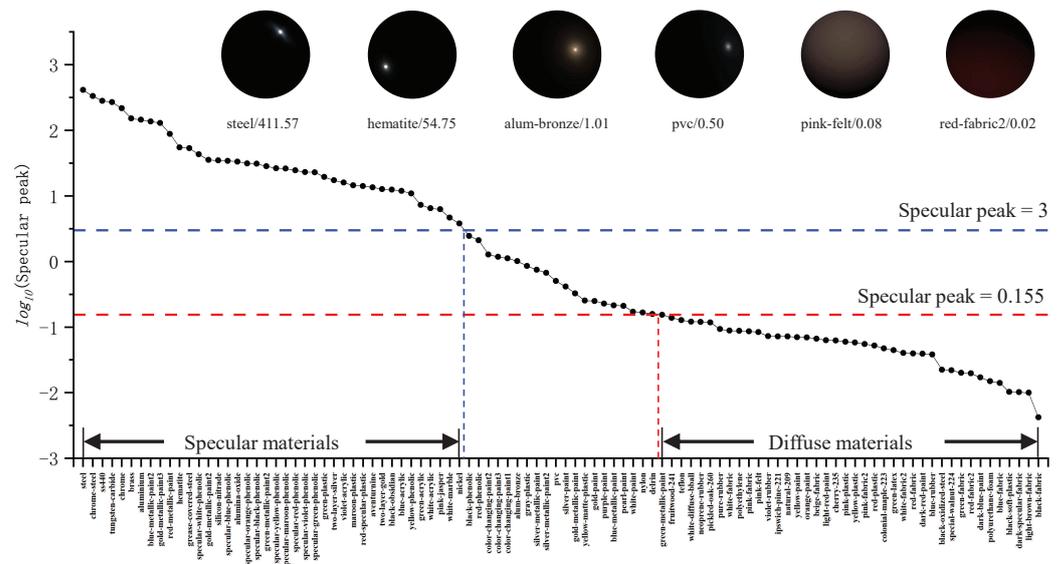


Figure 4. The specular peaks [36] of 100 MERL materials [35]. The vertical coordinate represents the value of the specular peak. We scaled the values through $\log_{10}(\text{specular peak})$ for a better display. The horizontal coordinate represents the name of each material in the dataset of 100 MERL materials. The spheres rendered by different materials are exhibited at the top of the pictures. Corresponding names and specular peaks are shown at the bottom of the object images.

The curve is flat at the ends and steep in the middle, and the materials are mainly clustered in two intervals. To draw a useful distinction of materials and ensure the same number of samples in the diffuse and specular categories, we used a dual-threshold classification method, which was defined as **specular samples**: $f(0) > 3$, **diffuse samples**: $f(0) < 0.155$. Thus, we divided the 100 MERL materials into 40 specular samples, 40 diffuse samples, and 20 mixed samples. The objects with diffuse and specular samples were employed as the training set for more effective supervision, and the mixed samples were used as the validation set.

2.2. Reflectance Decomposition Module (ReD Module)

As shown in Figure 3, the ReD module consisted of a diffuse feature extractor (DFE), specular feature extractor (SFC), and reflectance-aware network (RA-Net).

DFE and SFC. Since an object mask can provide strong information for the occluding contours of an object and effectively improve the performance of directional light estimation [17,18], we concatenated the mask and captured image as the input and used a two-branch structure as a feature extractor, as shown in Figure 3. Eight convolutional layers with a kernel size of 3×3 were employed to extract local features from the input, and each layer was followed by leaky ReLU activation. The DFE and SFE extracted local features from the pairs of images and masks, respectively, avoiding confusion between the two types of reflectance information in the feature extraction process.

RA-Net. Since the global feature contains implicit surface geometry and reflectance information of an object [17], we concatenated two types of global features as the input of the RA-Net to efficiently calculate the contributions of the two types of reflectance information, as shown in Figure 3. We first used one convolutional layer with $stride = 1$ to fuse both the diffuse and specular global features and three convolutional layers with $stride = 2$ to reduce the dimensions of the feature maps. All convolutional layers with a kernel size of 3×3 were followed by leaky ReLU activation. The processed features were fed into the azimuth-weight-estimation block and elevation-weight-estimation block. Each block contained a 1×1 convolutional layer, a leaky ReLU activation, and one 1×1 convolutional layer. A sigmoid was the final layer of the RA-Net for obtaining the output w_r and $w_s = 1 - w_r$. To confirm to the light direction estimation results, the reflectance weights were estimated separately for the azimuth and elevation.

2.3. Directional Light Estimation Module (DLE Module)

The DLE module consisted of a diffuse directional light estimator (DDLE) and specular directional light estimator (SDLE). The DLE module was also designed based on a two-stream structure and took the concatenation of local features and global features as input, as illustrated in Figure 5. Local features were extracted from the input of each image, and they only contained information relevant to each directional light. The global features aggregated all of the local features and implicitly contained the reflectance information (i.e., free from light) of the surface. We used one convolutional layer with a kernel size of 3×3 to fuse the local and global features and three convolutional layers with a kernel size of 3×3 to reduce the spatial dimension. Finally, azimuth classifiers and elevation classifiers, which consisted of two convolutional layers with a kernel size of 1×1 , were used to acquire the diffuse classification scores $Pred_{d-azi/ele}$ and specular classification scores $Pred_{s-azi/ele}$. The output of ReDDLE-Net could be expressed as the weighted sum of $Pred_{d-azi/ele}$ and $Pred_{s-azi/ele}$ as

$$Pred_{azi/ele} = Pred_{d-azi/ele} \times w_{diffuse-azi/ele} + Pred_{s-azi/ele} \times w_{specular-azi/ele}$$

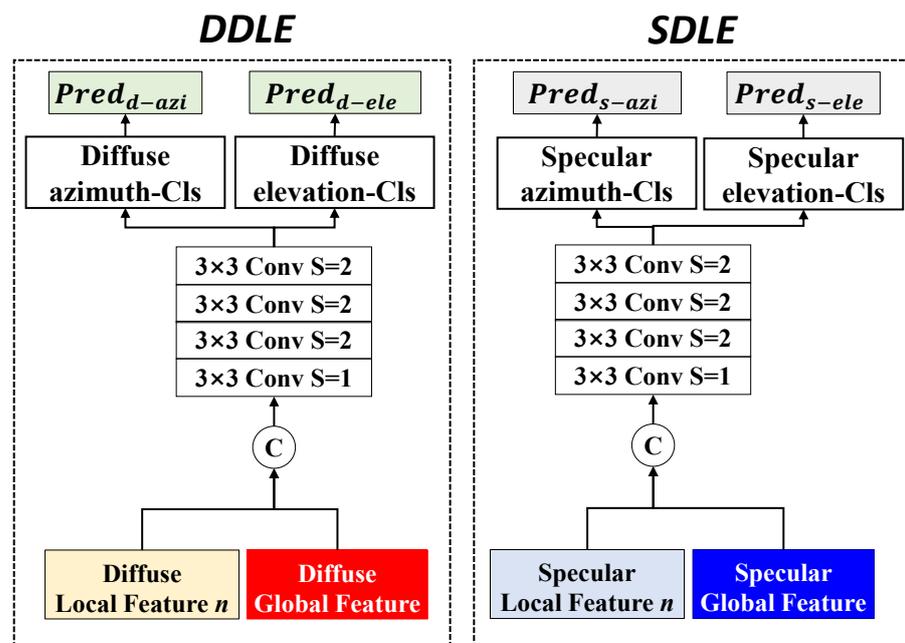


Figure 5. The proposed directional light estimation (DLE) module. Left: Diffuse directional light estimator (DDLE). Right: Specular directional light estimator (SDLE).

2.4. Loss Function

We adopted cross-entropy loss function for the light azimuth and elevation estimation, which were $\mathcal{L}_{dir-azi}$ and $\mathcal{L}_{dir-ele}$, respectively. Given N images, the light azimuth loss or light elevation loss can be expressed as:

$$\mathcal{L}_{dir-azi/ele} = \frac{1}{N} \sum_n \left(- \sum_{i=1}^{32} \{y_i^n = 1\} \log(p_i^n) \right) \quad (1)$$

where $\{\cdot\}$ is a binary indicator (0 or 1) function, p_i^n is the predicted probability for bin i of the n^{th} image, and $y_i^n = 1$ is the ground label (0 or 1). The reflectance decomposition loss item is defined as \mathcal{L}_{ReD}

$$\mathcal{L}_{ReD} = -\hat{w}_d \cdot \log(w_d) - \hat{w}_s \cdot \log(w_s) \quad (2)$$

where w_d and $w_s = 1 - w_d$ are the estimated reflectance weights for diffuse and specular samples, and \hat{w}_d and \hat{w}_s are the ground truths of the reflectance weights. We set $\hat{w}_d = 1$ for the diffuse samples and $\hat{w}_s = 1$ for the specular samples. To perform directional light estimation based on reflectance decomposition, we combined the loss terms (\mathcal{L}_{ReD} and \mathcal{L}_{dir}), and our final network loss function became

$$\begin{aligned} \mathcal{L} = & \mathcal{L}_{dir-azi} + \mathcal{L}_{dir-ele} + \lambda_{azi} \cdot \mathcal{L}_{ReD-azi} \\ & + \lambda_{ele} \cdot \mathcal{L}_{ReD-ele} \end{aligned} \quad (3)$$

where λ_{azi} and λ_{ele} balanced $\mathcal{L}_{ReD-azi}$ and $\mathcal{L}_{ReD-ele}$. We divided the dataset into pseudo-diffuse and pseudo-specular groups according to the specular peak indicator. We set $\lambda_{azi} = \lambda_{ele} = 0.05$ during training.

3. Results

All experiments were tested on a GeForce RTX 2080Ti. The Adam optimizer was used to optimize our networks during the training phase with the default parameters [37]. We used the *Blobby* and *Sculpture* datasets provided by [5] as our training dataset. The training process took around 27 h with 20 epochs. We applied the same data pre-processing technique as that suggested for LCNNet [17], except for the addition of extra noise disturbances. The proposed method had 8.82 million parameters and spent 7.28 s on processing the DiLiGenT benchmark with an input of 96 images.

3.1. Evaluation of Light Direction Estimation

In this section, we systematically evaluate the performance of our ReDDLE-Net and compare it with a state-of-the-art light direction estimation method on synthetic datasets and real datasets. The mean angular error (MAE; in degrees) was used to measure the accuracy of the predicted light direction.

3.1.1. Performance of Light Direction Estimation on Synthetic Datasets

The 100 MERL synthetic datasets was obtained by rendering *Bunny* and *Dragon* shapes with 100 BRDFs from the MERL dataset. We compared our method with LCNNet [17], which has a better ability to estimate the light direction on a surface with isotropic materials than traditional methods. As illustrated in Figure 6, the proposed method performed better on most materials and reduced the average light direction error by 0.17° and 0.77° on the *Bunny* and *Dragon* objects, respectively.

In order to increase the interpretability of the proposed method and validate the behavior of our method, which could adjust the attention according to the input, we used class activation maps (CAMs) to visualize the predicted class scores on the diffuse and specular sub-networks, respectively. In order to study the behavior of our method, which could adjust the attention according to the input, we used Eigen-CAM [38] to visualize the predicted class scores on the diffuse and specular sub-networks, respectively. A CAM [39] visualizes the predicted scores by projecting them on the input and highlights the discriminative features produced by the network. Eigen-CAM [38] improves CAMs by using the principal components of the features from the network, which are robust to adversarial noise. Figure 7 shows the class activations that were computed using Eigen-CAM for LCNNet and our method. We found that LCNNet used less information than our method and focused on a single type of reflectance information during feature extraction. Most objects have both diffuse and specular reflections on their surfaces, and both types of reflections provide useful clues for light direction estimation. Whereas LCNNet focused on a single type of information during feature extraction, our method improved performance in light direction estimation by extracting both diffuse and specular information through its diffuse and specular sub-networks. Smooth objects present both diffuse and specular reflectance, such as with the *blue metallic paint* of the *Bunny* and *green metallic paint* of the *Dragon*. The diffuse sub-network focused more on information in the diffuse area and bypassed the specular area. The specular sub-network, on the contrary, focused more on the information in the specular areas. The surfaces of rough materials exhibit Lambertian properties, i.e, diffuse reflective properties without sharply highlighted areas, such as with the *blue rubber* of the *Bunny* and *blue fabric* of the *Dragon*. For rough materials, the diffuse sub-network focused more on areas with local diffuse maxima. Meanwhile, the specular sub-network reduced diffuse materials' contributions to the results. Based on the experimental results, the ReDDLE-Net worked in line with our design: The two sub-networks focused on different information, and the network could adjust the contributions of the sub-networks to the final result according to the input of the surface characteristics.

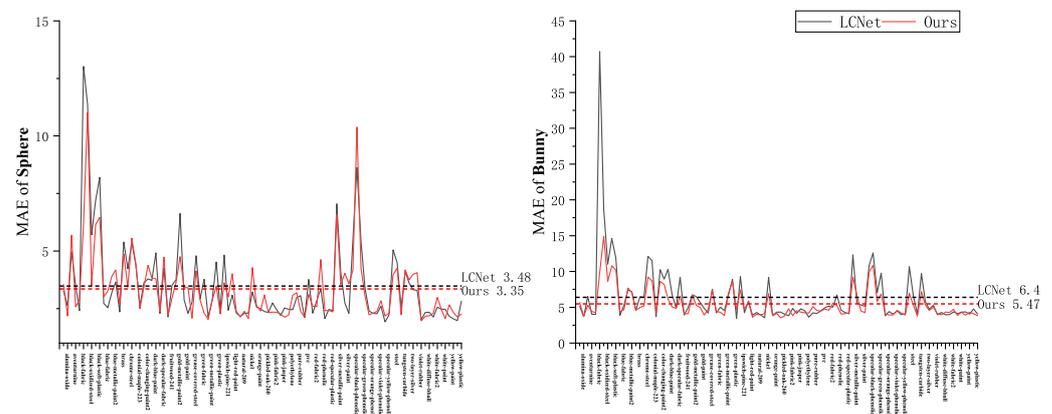


Figure 6. Quantitative comparison of LCNNet and our ReDDLE-Net for the *Bunny* and *Dragon* objects made of 100 MERL materials. The horizontal axis represents the names of the materials and the vertical axis represents the MAE. The horizontal coordinate represents the names of the materials in the 100 MERL dataset. The dashed line is the average error reference line for each method.

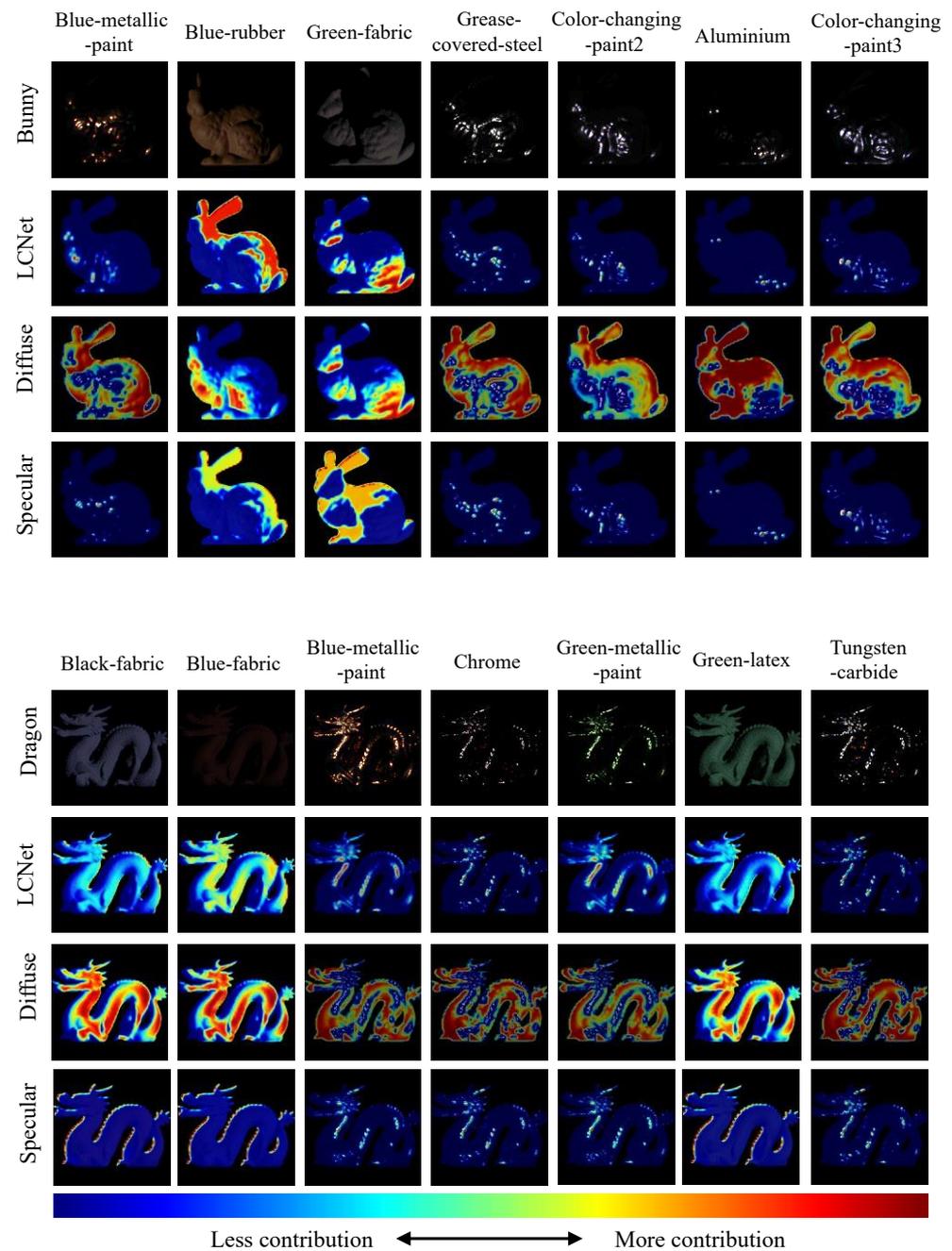


Figure 7. Schematic representation of the feature activation for LCNet and our ReDDLE-Net. The features in the red regions indicate more contributions, while blue regions indicate fewer contributions. The different columns represent the results of objects with different materials. Rows 1 and 5: Input images of eight *Bunny* and *Dragon* objects. The texts on the left indicate the names of the objects. Rows 2 and 6: Feature maps of LCNet [17]. Rows 3 and 7: Feature maps of our diffuse sub-network. The texts on the left indicate the name of our sub-network. Rows 4 and 8: Feature maps of our specular sub-network. The texts on the left indicate the name of our sub-network.

3.1.2. Performance of Directional Light Estimation on the DiLiGenT Benchmark Dataset

We compared our proposed ReDDLE-Net with an optimization-based method (PF14 [40]) and a learning-based method (LCNet [17]). Quantitative evaluation results on the DiLiGenT benchmark dataset are reported in Table 1. The proposed ReDDLE-Net achieved the lowest average light direction error on eight real-world objects from the DiLiGenT dataset. Our method performed worse than the others on *POT1*, *POT2*, and *CAT*. These three objects

were mainly diffuse and had multiple colors, which made it more difficult to estimate the light source. For diffuse objects, our ReD mechanism increased the contribution of the diffuse sub-network to the results and reduced the contribution of the specular sub-network. However, the highlights were not disturbed by color variations and were a valid clue for estimating the light directions with multicolored objects [18]. This was the reason for why our method performed poorly on diffuse multicolored objects (*POT1*, *POT2*, and *CAT*) and well on specular multicolored objects (*HARVEST*, *READING*). Figure 8 shows the light error distributions for objects in the DiLiGenT benchmark resulting from the two methods. We found that our method was much more effective in calculating light directions that were far from the viewing direction. This was because increasing the use of information by adjusting the attention to diffuse and specular reflectance could maintain the performance of the method. We also show the visual quality results for the Light Stage Data Gallery dataset [34] in Figure 9.

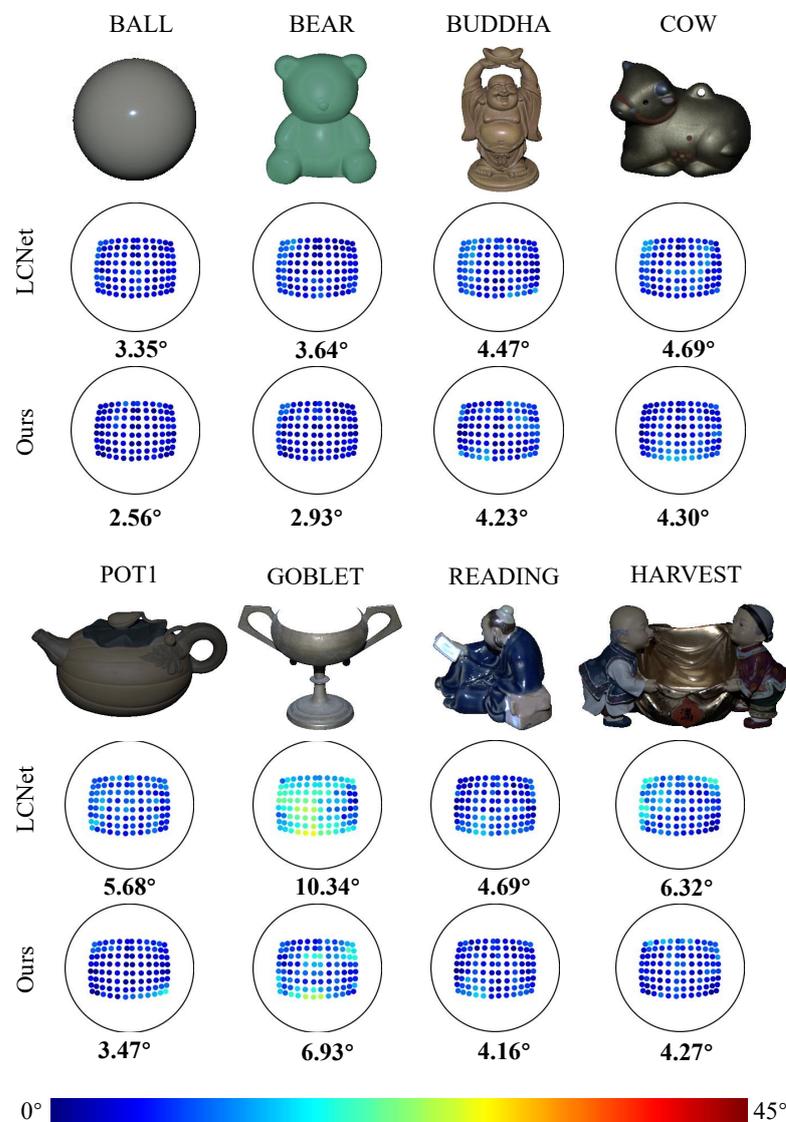


Figure 8. Quantitative light direction evaluation results on the DiLiGenT benchmark dataset. Rows 1 and 4: Captured images of eight samples in the DiLiGenT benchmark dataset. Rows 2–3 and 5–6: Light direction error distributions of corresponding samples for LCNet and ReDDLE-Net, respectively. Following [17], we mapped the light source $[x, y, z]$ to a point $[x, y]$. The color of the points corresponds to the magnitude of the error.

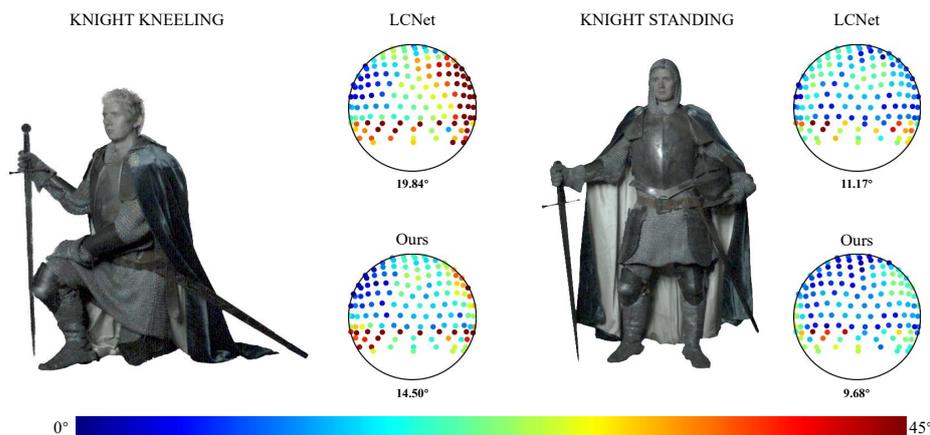


Figure 9. Quantitative light direction evaluation results on the Light Stage Data Gallery dataset. Left: Captured image and light direction error distribution on *KNIGHT KNEELING* for LCNet and ReDDLE-Net. Right: Captured image and light direction error distribution on *KNIGHT STANDING* for LCNet and ReDDLE-Net. We mapped the light source $[x, y, z]$ to a point $[x, y]$. The color of points corresponds to the magnitude of the error.

Table 1. The average light direction error on the DiLiGenT dataset. The numbers represent the MAE (°).

	BALL	CAT	POT1	BEAR	POT2	BUDDHA	GOBLET	READING	COW	HARVEST	Avg.
PF14 [40]	4.90	5.31	2.43	5.24	13.52	9.76	33.22	21.77	16.34	24.99	13.75
LCNet [17]	3.35	4.15	5.68	3.64	2.77	4.47	10.34	4.69	4.69	6.32	5.01
Ours	2.56	6.54	3.47	2.93	4.15	4.23	6.93	4.16	4.30	4.27	4.35

3.1.3. Performance of Reflectance Decomposition (Red) Mechanism

To verify the generality of the ReD mechanism, we applied the proposed method to cross-domain datasets (including synthetic and real datasets) and analyzed the relationship between the output weights and the reflective properties of the object surface.

Since the training datasets were obtained by rendering based on the 100 MERL BRDF dataset [35], we first applied the ReD mechanism on the CyclesPS dataset [25], which was created with the Disney BSDF model [36], for cross-domain evaluation. As shown in Figure 10, there was a significant difference in the reflective properties between the diffuse and metallic categories in CyclesPS. The proportion of specular reflectance was much greater in the metallic samples than in the diffuse samples. In contrast, the diffuse reflectance was dominant in the diffuse samples. We thus used the results from both categories for comparison, as shown in the right part of Figure 10. We found that the RA-Net made an effective distinction between the metallic and diffuse categories. The diffuse weights of diffuse objects were generally larger than those of metallic objects for the same sample. To qualitatively analyze the performance of the RA-Net on real objects, we show the output weights of the proposed method on the DiLiGenT dataset in Table 2. The results of the weight estimation were consistent with our perception of the objects’ surfaces. Objects (*BEAR*, *POT1*) that included a dominant Lambertian component had a large diffuse weight. Objects with metallic paint (*COW*) or strong specular spikes (*READING*) had a large specular weight. Since *BALL* and *POT2* had sparse specular spikes and a large area with a diffuse component, the ReD mechanism improved the contribution of diffuse feature extractor.

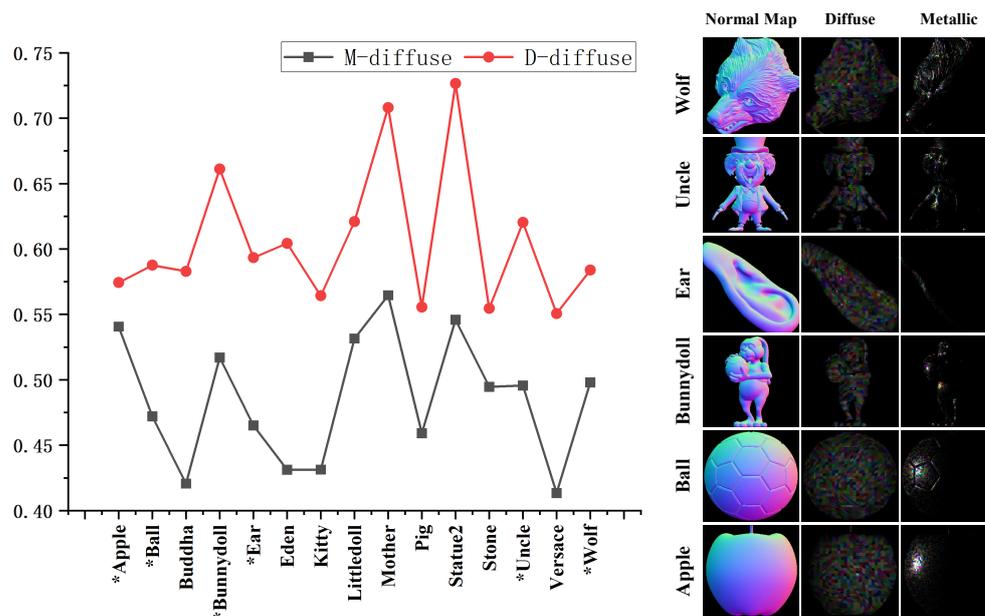


Figure 10. Left: Diffuse weight for *diffuse* and *metallic* samples in the CyclesPS dataset. The vertical coordinate represents the value of the reflectance weight. The horizontal coordinate represents the names of the objects in the CyclesPS dataset. The objects marked with * are visualized in the diagram on the right. Right: Six objects from the CyclesPS dataset. Each example consists of a normal map, an image of a *diffuse* sample, and an image of a *specular* sample.

Table 2. The average weights ($\frac{w_{azi} + w_{ele}}{2}$) of 10 objects in the DiLiGenT dataset.

	BALL	CAT	POT1	BEAR	POT2	BUDDHA	GOBLET	READING	COW	HARVEST
Diffuse weight	0.59	0.57	0.65	0.59	0.67	0.55	0.59	0.45	0.45	0.48
Specular weight	0.41	0.43	0.35	0.41	0.33	0.45	0.41	0.55	0.55	0.52

3.2. Evaluation of Normal Estimation

The light direction estimation network serves the uncalibrated photometric stereo problem. We combined ReDDLE-Net (ReDDLE-Net utilizes the raw images as inputs and acquires the light direction; then, we make use of the ground-truth light intensity in the datasets to calibrate the images; the images calibrated for light direction and intensity are employed as the inputs of the calibrated photometric stereo) with MT-PS-CNN [6] to deal with the uncalibrated photometric stereo task. MT-PS-CNN constructs inter- and intra-frame representations for accurate normal estimation of non-Lambertian objects, consisting of an *inter-frame feature extractor*, *intra-frame feature extractor*, and *normal map estimation*. For details, please refer to [6]. We compared the performance with that of traditional [16,17,40–44] and learning-based [17] (for a fair comparison, we corrected the light intensity of the input images before estimating the normal vector) uncalibrated photometric stereo methods on objects in the DiLiGenT benchmark, as reported in Table 3. The proposed model with MT-PS-CNN achieved the lowest average MAE (8.21°) on 10 real-world objects of DiLiGenT and performed the best on eight objects. In Figure 11, we visualize the surface normal estimation results from CH19 [17] and the proposed method. Due to the improved accuracy of light direction estimation, our method kept the normal map error to a low level, regardless of whether the surface was rough or smooth.

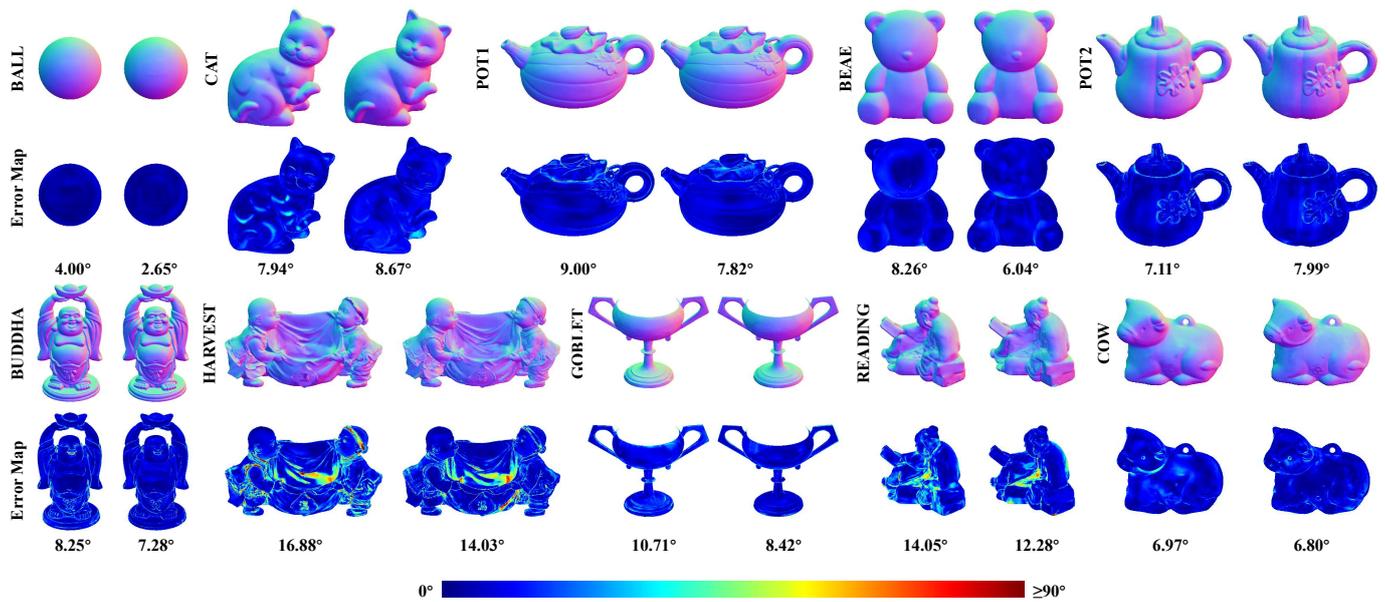


Figure 11. The surface normal estimation results of CH19 [17] and the proposed self-calibrating photometric stereo framework. Rows 1 and 3 are estimated normal maps. Rows 2 and 4 are error maps with the ground truth. The images on the left of each object are the results of CH19 [17], and those on the right are the results of the proposed method.

Table 3. The normal errors for 10 objects from the DiLiGenT dataset. The numbers represent the MAE (°).

	BALL	CAT	POT1	BEAR	POT2	BUDDHA	GORLET	READING	COW	HARVEST	AVG.
AM07 [41]	7.27	31.45	18.37	16.81	49.16	32.81	46.54	53.65	54.72	61.70	37.25
SM10 [42]	8.90	19.84	16.68	11.98	50.68	15.54	48.79	26.93	22.73	73.86	29.59
WT13 [43]	4.39	36.55	9.39	6.42	14.52	13.19	20.57	58.96	19.75	55.51	23.93
LM13 [16]	22.43	25.01	32.82	15.44	20.57	25.76	29.16	48.16	22.53	34.45	27.63
PF14 [40]	4.77	9.54	9.51	9.07	15.90	14.92	29.93	24.18	19.53	29.21	16.66
LC17 [44]	9.30	12.60	12.40	10.90	15.70	19.00	18.30	22.30	15.00	28.00	16.35
CH19 [17]	4.00	7.94	9.00	8.26	7.11	8.25	10.71	14.05	6.97	16.88	9.32
Ours	2.65	8.76	7.82	6.04	7.99	7.28	8.42	12.28	6.80	14.03	8.21

4. Discussion

As demonstrated by the results shown in Section 3, ReDDLE-Net performed well on both synthetic and real datasets. Existing learning-based methods [17–19] extract reflectance features implicitly, and it is hard for them to make full use of both types of reflectance clues, as we mentioned in Section 3.1.1. Our ReDDLE captured the diffuse and specular reflections separately through a two-branch structure and adjusted the contributions of different reflectances to the results through the RA-Net. We built the training data with labels of reflectance properties and use them to guide the diffuse and specular sub-networks to focus on the corresponding reflectance information.

Although our method achieved good performance on different datasets, there are still some limitations. Since we used the specular peak as a single indicator to classify the material datasets, the training datasets were actually divided into pseudo-diffuse and pseudo-specular materials. However, those data were not purely diffuse or specular. To address this problem, in future work, we will use more physical variables as indicators for classification or adapt an unsupervised approach in order to adjust the contributions of diffuse and specular information.

5. Conclusions

In this paper, a novel reflectance decomposition (ReD) mechanism was developed to adaptively adjust the attention of diffuse and specular reflectance with material changes. We built a two-branch CNN architecture called ReDDLE-Net and achieved competitive performance on synthetic and real datasets. We used the class activation mapping (CAM) method to verify that the diffuse and specular sub-networks focused on diffuse and specular reflectance, respectively. The experimental results showed that the ReD mechanism provided an effective stage for promoting light estimation. In addition, the proposed ReDDLE-Net could be seamlessly integrated into learning-based photometric stereo methods to handle the uncalibrated photometric stereo task.

Author Contributions: Conceptualization, J.Y. and B.D.; methodology, B.D. and Y.C. (Yanlong Cao); software, Z.H.; validation, G.P., Y.C. (Yanpeng Cao), and Y.C. (Yanlong Cao); writing—original draft preparation, Q.Z.; writing—review and editing, Q.Z.; visualization, Q.Z.; supervision, Y.C. (Yanlong Cao). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China: 2020YFB1711400, National Natural Science Foundation of China: 52075485, Zhejiang Provincial Key Research and Development Program: 2021C03112, The National Science Fund for Distinguished Young Scholars: 61925603 and Zhejiang Lab: 2018EB0ZX01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Woodham, R.J. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* **1980**, *19*, 191139. [[CrossRef](#)]
2. Mecca, R.; Wetzler, A.; Bruckstein, A.M.; Kimmel, R. Near field photometric stereo with point light sources. *Siam J. Imaging Sci.* **2014**, *7*, 2732–2770. [[CrossRef](#)]
3. Logothetis, F.; Mecca, R.; Cipolla, R. Semi-calibrated near field photometric stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 941–950.
4. Logothetis, F.; Budvytis, I.; Mecca, R.; Cipolla, R. A CNN based approach for the near-field photometric stereo problem. *arXiv* **2020**, arXiv:2009.05792.
5. Chen, G.; Han, K.; Wong, K.Y.K. PS-FCN: A flexible learning framework for photometric stereo. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–18.
6. Cao, Y.; Ding, B.; He, Z.; Yang, J.; Chen, J.; Cao, Y.; Li, X. Learning inter- and intraframe representations for non-Lambertian photometric stereo. *Opt. Lasers Eng.* **2022**, *150*, 106838. [[CrossRef](#)]
7. Cho, D.; Matsushita, Y.; Tai, Y.W.; Kweon, I.S. Semi-calibrated photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 232–245. [[CrossRef](#)] [[PubMed](#)]
8. Quéau, Y.; Wu, T.; Cremers, D. Semi-calibrated near-light photometric stereo. In *Lecture Notes in Computer Science: Proceedings of the International Conference on Scale Space and Variational Methods in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 656–668.
9. Hayakawa, H. Photometric stereo under a light source with arbitrary motion. *JOSA A* **1994**, *11*, 3079–3089. [[CrossRef](#)]
10. Belhumeur, P.N.; Kriegman, D.J.; Yuille, A.L. The bas-relief ambiguity. *Int. J. Comput. Vis.* **1999**, *35*, 33–44. [[CrossRef](#)]
11. Yuille, A.; Snow, D. Shape and albedo from multiple images using integrability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 158–164.
12. Yuille, A.L.; Snow, D.; Epstein, R.; Belhumeur, P.N. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability. *Int. J. Comput. Vis.* **1999**, *35*, 203–222. [[CrossRef](#)]
13. Epstein, R.; Yuille, A.L.; Belhumeur, P.N. Learning object representations from lighting variations. In *Lecture Notes in Computer Science: Proceedings of the International Workshop on Object Representation in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 179–199.
14. Kriegman, D.J.; Belhumeur, P.N. What shadows reveal about object structure. *JOSA A* **2001**, *18*, 1804–1813. [[CrossRef](#)] [[PubMed](#)]
15. Sato, I.; Okabe, T.; Yu, Q.; Sato, Y. Shape reconstruction based on similarity in radiance changes under varying illumination. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8.

16. Lu, F.; Matsushita, Y.; Sato, I.; Okabe, T.; Sato, Y. Uncalibrated photometric stereo for unknown isotropic reflectances. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1490–1497.
17. Chen, G.; Han, K.; Shi, B.; Matsushita, Y.; Wong, K.Y.K. Self-calibrating deep photometric stereo networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8739–8747.
18. Chen, G.; Waechter, M.; Shi, B.; Wong, K.Y.K.; Matsushita, Y. What is learned in deep uncalibrated photometric stereo? In *Lecture Notes in Computer Science: European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 745–762.
19. Kaya, B.; Kumar, S.; Oliveira, C.; Ferrari, V.; Van Gool, L. Uncalibrated neural inverse rendering for photometric stereo of general surfaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3804–3814.
20. Liu, H.; Yan, Y.; Song, K.; Yu, H. SPS-Net: Self-Attention photometric stereo network. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–13. [[CrossRef](#)]
21. Yao, Z.; Li, K.; Fu, Y.; Hu, H.; Shi, B. Gps-net: Graph-based photometric stereo network. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 10306–10316.
22. Ikehata, S. PS-Transformer: Learning sparse photometric stereo network using self-attention mechanism. *BMVC* **2021**, *2*, 11.
23. Logothetis, F.; Budvytis, I.; Mecca, R.; Cipolla, R. PX-NET: Simple and efficient pixel-wise training of photometric stereo networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 12757–12766.
24. Zheng, Q.; Jia, Y.; Shi, B.; Jiang, X.; Duan, L.Y.; Kot, A.C. SPLINE-Net: Sparse photometric stereo through lighting interpolation and normal estimation networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8549–8558.
25. Ikehata, S. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Computer Vision Foundation: Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–18.
26. Ju, Y.; Dong, J.; Chen, S. Recovering surface normal and arbitrary images: A dual regression network for photometric stereo. *IEEE Trans. Image Process.* **2021**, *30*, 3676–3690. [[CrossRef](#)] [[PubMed](#)]
27. Wang, X.; Jian, Z.; Ren, M. Non-lambertian photometric stereo network based on inverse reflectance model with collocated light. *IEEE Trans. Image Process.* **2020**, *29*, 6032–6042. [[CrossRef](#)]
28. Ju, Y.; Lam, K.M.; Chen, Y.; Qi, L.; Dong, J. Pay attention to devils: A photometric stereo network for better details. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, Yokohama, Japan, 7–15 January 2021; pp. 694–700.
29. Li, J.; Robles-Kelly, A.; You, S.; Matsushita, Y. Learning to minify photometric stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7568–7576.
30. Honzátko, D.; Türetken, E.; Fua, P.; Dunbar, L.A. Leveraging Spatial and Photometric Context for Calibrated Non-Lambertian Photometric Stereo. In Proceedings of the 2021 International Conference on 3D Vision (3DV), London, UK, 1–3 December 2021; pp. 394–402.
31. Santo, H.; Samejima, M.; Sugano, Y.; Shi, B.; Matsushita, Y. Deep photometric stereo networks for determining surface normal and reflectances. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 114–128. [[CrossRef](#)] [[PubMed](#)]
32. Shi, B.; Wu, Z.; Mo, Z.; Duan, D.; Yeung, S.K.; Tan, P. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3707–3716.
33. Alldrin, N.; Zickler, T.; Kriegman, D. Photometric stereo with non-parametric and spatially-varying reflectance. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
34. Chabert, C.F.; Einarsson, P.; Jones, A.; Lamond, B.; Ma, W.C.; Sylwan, S.; Hawkins, T.; Debevec, P. Relighting human locomotion with flowed reflectance fields. In *ACM SIGGRAPH 2006 Sketches*; Association for Computing Machinery: Boston, MA, USA, 2006; p. 76.
35. Matusik, W. A Data-Driven Reflectance Model. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2003.
36. Burley, B.; Studios, W.D.A. Physically-based shading at disney. In Proceedings of the ACM SIGGRAPH, 2012; Volume 2012, pp. 1–7. Available online: <https://www.disneyanimation.com/publications/physically-based-shading-at-disney/> (accessed on 10 August 2022).
37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
38. Muhammad, M.B.; Yeasin, M. Eigen-cam: Class activation map using principal components. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
39. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
40. Papadhimetri, T.; Favaro, P. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *Int. J. Comput. Vis.* **2014**, *107*, 139–154. [[CrossRef](#)]
41. Alldrin, N.G.; Mallick, S.P.; Kriegman, D.J. Resolving the generalized bas-relief ambiguity by entropy minimization. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–7.

42. Shi, B.; Matsushita, Y.; Wei, Y.; Xu, C.; Tan, P. Self-calibrating photometric stereo. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1118–1125.
43. Wu, Z.; Tan, P. Calibrating photometric stereo by holistic reflectance symmetry analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1498–1505.
44. Lu, F.; Chen, X.; Sato, I.; Sato, Y. Symps: Brdf symmetry guided photometric stereo for shape and light source estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 221–234. [[CrossRef](#)] [[PubMed](#)]