ReLeaPS : Reinforcement Learning-based Illumination Planning for Generalized Photometric Stereo

Jun Hoong Chan[1,2,†] Bohan Yu[1,2,†] Heng Guo[3,4] Jieji Ren[5] Zongqing Lu[1,2] Boxin Shi[1,2,‡]

[1] National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University
[2] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[3] School of Artificial Intelligence, Beijing University of Posts and Telecommunications [4] Osaka University
[5] School of Mechanical Engineering, Shanghai Jiao Tong University

junhoong95@stu.pku.edu.cn, {ybh1998, zongqing.lu, shiboxin}@pku.edu.cn,
heng.guo@ist.osaka-u.ac.jp, jiejiren@sjtu.edu.cn

Abstract

Illumination planning in photometric stereo aims to find a balance between surface normal estimation accuracy and image capturing efficiency by selecting optimal light configurations. It depends on factors such as the unknown shape and general reflectance of the target object, global illumination, and the choice of photometric stereo backbones, which are too complex to be handled by existing methods based on handcrafted illumination planning rules. This paper proposes a learning-based illumination planning method that jointly considers these factors via integrating a neural network and a generalized image formation model. As it is impractical to supervise illumination planning due to the enormous search space for ground truth light configurations, we formulate illumination planning using reinforcement learning, which explores the light space in a photometric stereo-aware and reward-driven manner. Experiments on synthetic and real-world datasets demonstrate that photometric stereo under the 20*-light configurations from our method is comparable to, or even surpasses that of using lights from all available directions.*

1. Introduction

Photometric stereo estimates the surface normal of an object from images taken in a fixed camera viewpoint under different light directions. From the early work of Woodham [24], photometric stereo can be solved with at least three calibrated lights under the ideal Lambertian reflectance assumption. However, shape recovery in the real-world scene is more complex. For brevity, we denote *generalized photometric stereo* as recovering surface normal from image observations with consideration of general non-Lambertian reflectances and global illumination effects such as shadows and inter-reflections. To achieve generalized photometric stereo, more lights with uniform distribution are generally preferred in existing methods [17], which can be time-consuming and impractical. To overcome this limitation, it is desirable to develop appropriate illumination planning strategies that can achieve similar or better accuracy with a limited number of light directions.

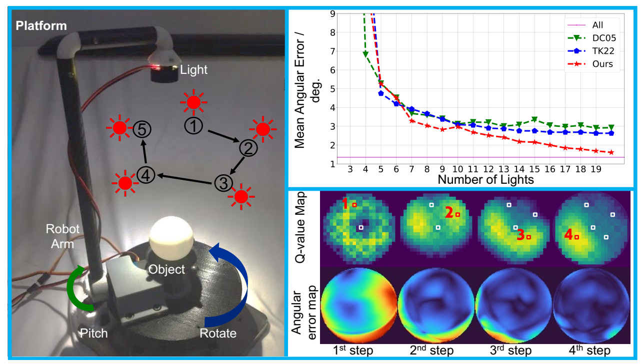

Figure 1. **(Left)** Our online illumination planning validation setup with the light direction being controlled by a robot arm. **(Top right)** Normal estimation error under light distributions from different illumination planning methods. Photometric stereo using 20 lights from our illumination planning outperforms existing methods (DC05 [3] and TK22 [21]), and can even obtain comparable results that use 100 light candidates (All). **(Bottom right)** Illumination planning results for a sphere object based on our ReLeaPS, where the light directions are updated iteratively from the Q-value map, driven by the reward of surface normal estimation errors.

Despite the importance of illumination planning, few methods have been proposed. An optimal offline light source placement for Lambertian photometric stereo [3] has

Project page: `https://jhchan0805.github.io/ReLeaPS`
† Equal contributions. ‡ Corresponding author.

been achieved by minimizing the normal uncertainty. On top of this, the illumination planning method for shadow-robust photometric stereo [21] is further designed by optimizing the light source direction adaptively based on previously captured images of an object. Both methods assume Lambertian reflectance, limiting their effectiveness to capture complex properties of real-world material and illumination on non-Lambertian surfaces.

In the context of photometric stereo, achieving optimal illumination planning involves addressing two primary concerns: 1) determining an appropriate learning strategy for illumination planning, and 2) integrating the image formation model of generalized photometric stereo into the pipeline. Illumination planning is a complex task that heavily depends on the object's shape, reflectance properties, and global illumination effects. Existing methods [3, 21] rely on heuristic rules, making it hard to handle online illumination planning under real-world reflectance, global illuminations, and adaptively fitting to generalized photometric stereo backbones. How to integrate a generalized image formation model with a deep neural network for illumination planning should also be considered.

Learning-based methods have demonstrated notable achievements in handling non-Lambertian reflectance properties in generalized photometric stereo in recent years [2, 4]. These methods have significantly boosted the accuracy of surface normal estimation. The success of learning-based techniques motivates us to integrate illumination planning into photometric stereo using a learning-based approach. However, the illumination planning problem is characterized by an enormous search space, which makes it difficult to obtain the ground truth light distribution. Consequently, applying supervised learning techniques to learn illumination planning becomes prohibitively difficult.

In this paper, we propose ReLeaPS, a **Re**inforcement **Lea**rning-based illumination planning for generalized **P**hotometric **S**tereo as shown in Fig. 1. Reinforcement learning (RL) is a learning method that trains an agent through trial-and-error, which enables the agent to efficiently explore the enormous search space in a reward-driven manner. To formulate the illumination planning with generalized photometric stereo into the RL framework, we employ a dueling deep Q-network (DQN) [22] to learn the optimal action based on input images and light directions. We propose a brute-force exploration strategy for the agent to explore the light direction and iteratively optimize its actions via Q-learning based on accumulated observations to maximize the reward. We further transform the sparse reward into a dense reward to facilitate the learning process in RL. In summary, ReLeaPS hopes to make *Re*markable *LeaPS* towards more time-efficient and accuracy-aware photometric stereo in practical application, via the following contribution:

- proposing the first RL approach for online illumination planning in a reward-driven manner;
- designing a dueling DQN specially tailored to generalized photometric stereo;
- enhancing the performance of different photometric stereo backbones with a smaller number of inputs by appropriate illumination planning; and
- evaluating RL-based illumination planning by building a real data validation setup.

Results using ReLeaPS indicate that the accuracy of the recovered normal benefitted from RL-based illumination planning surpasses existing methods [3, 21] (Fig. 1 right).

2. Related Work

This section mainly discusses the influence of the number of illuminations for photometric stereo, optimal illumination planning approaches, and reinforcement learning. We recommend relevant survey papers for comprehensive reviews and benchmarks about non-learning [18] and learning-based [14, 7] photometric stereo. We also mention RL for vision problems very briefly.

Photometric stereo. As discussed in the survey on photometric stereo [18], despite that classical photometric stereo [24] requires a minimum of 3 non-coplanar light directions to solve the surface normal, more images under varying lights are preferred to handle general reflectances and complicated effects in photometric stereo. For example, typically $10 \sim 20$ images are required for photometric stereo with shadow analysis [1]. For outlier-based photometric stereo methods [25, 16, 15, 12, 26, 20, 11], about $50 \sim 100$ images are used to reject shadows and specular highlights. Most photometric stereo methods handling general BRDFs (*e.g.*, PS-FCN [2], CNN-PS [4], PX-Net [9]) show better performance by taking $50 \sim 100$ images as input. To reduce the number of lights, existing methods such as SPLINE-Net [28] and LMPS [8] consider designing specific network structures to achieve a photometric stereo under a sparse light distribution. Instead of developing new photometric stereo methods, we reduce the number of lights while maintaining the normal estimation accuracy by better planning the light distribution during data capture.

Optimal illumination planning. We summarize illumination planning approaches in photometric stereo in Table 1 by considering various factors such as the object shape, general non-Lambertian reflectance, online/offline manner, and global illumination effects (*e.g.*, cast shadows, interreflections). Drbohlav and Chantler (DC05) [3] proposed the first illumination planning work in Lambertian photometric stereo considering the camera noise, showing that the distribution of orthogonal triplets of lights is optimal. This illumination setting is offline as the optimized lights

Table 1. Summary of illumination planning methods.

Method	Reflectance model	Offline/ online	Shape aware	Global illumination	
				Shadows	Inter-reflection
DC05 [3]	Lambertian	Offline	✗	✗	✗
TK22 [21]	Lambertian	Online	✓	✓	✗
IK23 [5]	Non-Lambertian	Offline	✗	✓	✓
Ours	Non-Lambertian	Online	✓	✓	✓

are independent of the shape and reflectance of the scene. Tanikawa *et al*. (TK22) [21] presented an online illumination planning related to the shape and shadows in the scene. The light direction is iteratively chosen based on the criteria that the shadow effects are minimized in the newly captured images. To handle non-Lambertian reflectance, Iwaguchi and Kawasaki (IK23) [5] proposed a near-light photometric stereo method and found optimal light patterns for diffuse and Phong materials from the training data. However, the optimal illumination is offline in the test phase and only adapted to the corresponding near-light photometric stereo setup [5]. We propose an online illumination planning method based on reinforcement learning, producing optimal light directions iteratively considering the shape and non-Lambertian reflectance, and can be adapted to different photometric stereo backbones.

RL for vision problems. Starting from the [19], RL is first proposed as a mathematical problem in which an agent learns through interaction with an environment. The agent receives rewards or punishments for each action it takes, allowing it to learn which actions lead to better outcomes. Compared to supervised learning, RL is particularly useful in scenarios where it is not feasible to obtain labeled data. RL has now been used in low-level vision tasks such as low-light image enhancement [27], distorted color enhancement [13], and denoise [13]. However, RL has not been explored in the task of photometric stereo and illumination planning. We introduce a novel approach for light distribution selection in generalized photometric stereo by integrating a reward function with RL principles.

3. Problem Formulation

This section begins with an introduction to the image formation model and a definition of the photometric stereo task. Subsequently, we describe the problem of illumination planning in generalized photometric stereo and present how it can be formulated using RL.

3.1. Image Formation Model

Given an orthographic camera with a linear radiometric response and T calibrated directional lights, we capture T image observations $\mathcal{I} = \{\boldsymbol{I}_1, \boldsymbol{I}_2, \cdots, \boldsymbol{I}_T\} \in \mathbb{R}^{H\times W\times T}$ by turning on/off the lights one after another. The image intensity profile $\mathcal{I}(\boldsymbol{p}) \in \mathbb{R}^T$ at pixel position $\boldsymbol{p}$ under varying lights can be formulated as

$$\mathcal{I}(\boldsymbol{p}) = \boldsymbol{s} \odot \boldsymbol{\rho} \odot \max(\boldsymbol{L}\boldsymbol{n}, \boldsymbol{0}), \tag{1}$$

where $\boldsymbol{s}$ and $\boldsymbol{\rho}$ are T-dimensional vectors representing the global illumination effects (*e.g*., cast shadows, inter-reflections) and non-Lambertian reflectance under T light directions $\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_T] \in \mathbb{R}^{T\times 3}$, $\boldsymbol{n} \subset \mathcal{S}^2 \in \mathbb{R}^3$ represents the surface normal at the pixel position $\boldsymbol{p}$.

Given image observations $\mathcal{I}$ and calibrated lights $\boldsymbol{L}$, photometric stereo (denoted as a function $\mathrm{PS}(\cdot)$) aims at recovering a surface normal map $\boldsymbol{N} \in \mathbb{R}^{H\times W\times 3}$ containing the normal vector $\boldsymbol{n}$ estimated at each pixel position, *i.e*.

$$\boldsymbol{N} = \mathrm{PS}(\mathcal{I}, \boldsymbol{L}). \tag{2}$$

3.2. Illumination Planning Driven by RL

As shown in Eq. (2), the accuracy of the estimated surface normal is related to the choice of photometric stereo backbones $\mathrm{PS}(\cdot)$ and light distribution $\boldsymbol{L}$. We define the illumination planning problem for generalized photometric stereo as follows,

Definition 1 *Suppose that the surfaces are expected to be illuminated by a total of T lights with the first light direction fixed to $\boldsymbol{l}_1$, the problem is to find the next $T-1$ optimal light directions such that under these total T light directions $\boldsymbol{L}_T^*$ and the corresponding image observations the estimated surface normal from a given photometric stereo backbone* $\mathrm{PS}(\cdot)$ *has the lowest angular error w.r.t. the ground truth $\tilde{\boldsymbol{N}}$,* i.e.*,*

$$\boldsymbol{L}_T^* = \operatorname*{argmin}_{\boldsymbol{L}_T} \frac{1}{HW} \sum_{i,j} \left(1 - \mathrm{PS}(\mathcal{I}_T, \boldsymbol{L}_T)_{ij} \cdot \tilde{\boldsymbol{N}}_{ij}\right). \tag{3}$$

The ground truth of light distribution is difficult to obtain as it is jointly influenced by complex non-Lambertian reflectance, global illuminations, shape variations, and photometric stereo backbones. Even for a fixed scene with ground truth normal and a specific photometric stereo backbone, finding the combination of $T-1$ optimal light directions in a brute-force manner still has an enormous searching space as T increases. Therefore, it is unrealistic to formulate illumination planning as a supervised learning task.

Instead, we solve the illumination planning via an iterative trial-and-error manner in a total of $T-1$ steps, which can be formulated as an RL problem. As shown in Fig. 2, beginning from an initial light direction and its corresponding image observation, we stack the image sequence and the light directions at each step, $t \in \{1, \ldots, T\}$, as state $\mathcal{S}_t$ in RL. The illumination planning algorithm corresponding to the agent in RL and denoted as a function $\mathrm{Agent}(\cdot)$ generates the next light directions denoted as $\mathcal{A}_t$, which corresponds to the action in RL. This illumination planning and

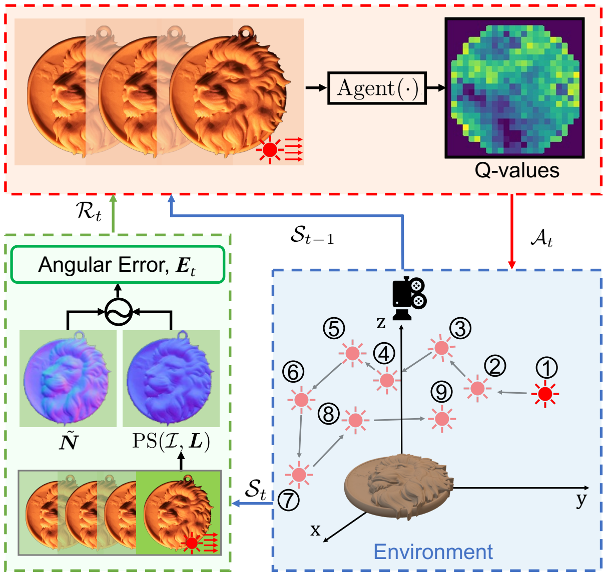

Figure 2. The illumination planning pipeline of ReLeaPS. **(Red block)** The agent observes the state $\mathcal{S}_{t-1}$ and selects action $\mathcal{A}_t$ from Q-values. **(Blue block)** The environment captures a new image based on action $\mathcal{A}_t$ to form a new state $\mathcal{S}_t$. **(Green block)** The angular error $\boldsymbol{E}_t$ between the predicted $\mathrm{PS}(\mathcal{I}, \boldsymbol{L})$ and ground truth normals $\tilde{\boldsymbol{N}}$ is calculated to form the reward $\mathcal{R}_t$. Then the network is updated based on reward $\mathcal{R}_t$. The process is a repetitive cycle with the above elements until the episode is terminated.

state update process can be formulated as:

$$\begin{aligned} \mathcal{A}_{t+1} &= \boldsymbol{l}_{t+1} = \mathrm{Agent}(\mathcal{S}_t), \\ \mathcal{S}_{t+1} &= \mathcal{S}_t \cup \{\boldsymbol{I}_{t+1}, \boldsymbol{l}_{t+1}\}\,. \end{aligned} \tag{4}$$

By repeating the above equations iteratively for $T-1$ times, the light distributions $\boldsymbol{L}_T$ and the corresponding image set $\mathcal{I}_T$ are generated as one illumination planning trial, which is then fed to a $\mathrm{PS}(\cdot)$ for estimating the surface normal. We define the reward in RL with the angular error between the estimated and the ground truth surface normal, and use it as feedback to update the agent for better illumination planning in the next trial.

Given the definition of illumination planning and its formulation based on RL, we show the details of the reward and agent network in RL in the next section, which is specifically designed to fit the context of photometric stereo.

4. Proposed Method

In this section, we introduce our proposed ReLeaPS, where the reward function, agent network, and exploration strategy are specifically designed based on photometric stereo prior knowledge for effective illumination planning.

4.1. Sparse-to-dense Reward Design

As shown in Fig. 2, we define the angular error map $\boldsymbol{E}_t \in \mathbb{R}^{H\times W}$ between the ground truth and estimated surface normal from images and lights at the t-th state,

$$\boldsymbol{E}_t = \left(1 - \mathrm{PS}(\mathcal{I}_t, \boldsymbol{L}_t)_{ij} \cdot \tilde{\boldsymbol{N}}_{ij}\right). \tag{5}$$

Following Definition 1, the optimization target is to minimize the mean of the angular error map at the T-th state, which can be directly converted to a single reward map at time T:

$$\mathcal{R}_t = \begin{cases} \mathbf{0}, & 2 \le t < T, \\ -\boldsymbol{E}_t, & t = T. \end{cases} \tag{6}$$

However, this direct reward strategy is only defined for the last state, leading to the problem of sparse reward settings. RL under sparse rewards is challenging as the agent can only receive limited feedback and underestimate the value of actions taken from other states. The sparse reward can lead to slow learning, exploration problems, lack of feedback, and difficulty in generalization. This highlights the importance of designing effective reward functions and exploring alternative techniques to encourage faster learning. First, we transform the angular error in the last state to the delta angular error between neighboring states, *i.e.*,

$$\mathcal{R}_t = \boldsymbol{E}_{t-1} - \boldsymbol{E}_t, \quad 2 \le t \le T. \tag{7}$$

As the accumulation of sparse and dense $\mathcal{R}_t$ are the same in all $T-1$ steps, we keep the optimization target but avoid the propagation problem in sparse rewards, which are shown to be effective for accelerating the RL training.

We observe that the existing method [21] conducts the illumination planning by only considering the one-step expected angular error $\boldsymbol{E}_t$ if adding the light $\boldsymbol{l}_t$. Motivated by their illumination planning practice on the photometric stereo, we improve the reward design by adding more attention to the one-step angular error $\boldsymbol{E}_t$. This technique is called reward shaping, formulated as

$$\mathcal{R}_t = \boldsymbol{E}_{t-1} - \boldsymbol{E}_t - \alpha \boldsymbol{E}_t, \quad 2 \le t \le T, \tag{8}$$

where α is the weight of the reward shaping term. Although the target is different from the optimization target in Definition 1, we empirically discover that this new dense reward strategy is more robust and efficient in RL training.

4.2. Agent Network for Illumination Planning

Overview of dueling DQN. First, we briefly introduce the background knowledge about relevant RL networks used in our design. We adopt dueling DQN [22] as our agent network to predict optimal light directions. DQN defines Q-value as the measure of expected cumulative reward that an agent can obtain by taking a specific action in a given state. Instead of choosing the action with maximum one-step reward in Eq. (8), DQN selects the one with maximum Q-value to determine the action $\mathcal{A}_t$ leading to the maximum

Algorithm 1 Illumination planning in generalized photometric stereo with reinforcement learning

```
repeat
    Initialize l_1 at (0, 0, 1)^⊤ direction
    Select observation image I_1 from dataset
    Combine l_1 and I_1 to form initial state S_1
    for t = 2, ..., T do
        Agent choose the l_t (i.e., A_t) based on state S_{t-1}
        Select observation image I_t from dataset
        Combine l_t, I_t, and previous S_{t-1} to form state S_t
        Estimate normal PS(I_t, L_t) in state S_t
        Calculate angular error E_t between PS(I_t, L_t) & Ñ
        Compute reward R_t from angular error E_t
        Store S_{t-1}, A_t, R_t, S_t in replay buffer
        Sample batch from replay buffer
        Update neural network
    end for
until Maximum number of episodes has been reached
```

expected long-term reward at state $\mathcal{S}_t$. To improve the efficiency and stability of DQN, dueling DQN, as a variant of DQN, separates the estimation of the state value function $\mathrm{Val}(\mathcal{S}_t)$ and the advantage function $\mathrm{Adv}(\mathcal{S}_t, \mathcal{A}_{t+1})$. The state value records the expected cumulative reward that an agent can obtain from a given state $\mathcal{S}$, while the advantage represents the additional expected reward by taking a specific action $\mathcal{A}$ in that state $\mathcal{S}$. The Q-value function for dueling DQN is a combination of the state value and advantage:

$$\begin{aligned}\mathrm{Q}(\mathcal{S}_t, \mathcal{A}_{t+1}) &= \mathrm{Val}(\mathcal{S}_t) + \mathrm{Adv}(\mathcal{S}_t, \mathcal{A}_{t+1}) \\ &\quad - \frac{1}{|\mathcal{A}|} \sum_{\mathcal{A}_{t+1}} \mathrm{Adv}(\mathcal{S}_t, \mathcal{A}_{t+1}),\end{aligned} \tag{9}$$

where $|\mathcal{A}|$ is size of action space. The Q-Network is trained by the Bellman equation and the loss function $\mathcal{L}$ is:

$$\begin{aligned}\mathrm{Q}'(\mathcal{S}_t, \mathcal{A}_{t+1}) &= \mathcal{R}_{t+1} + \gamma \max_{\mathcal{A}_{t+2}} \mathrm{Q}_{\mathrm{target}}(\mathcal{S}_{t+1}, \mathcal{A}_{t+2}), \\ \mathcal{L} &= \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\mathrm{Q}(\mathcal{S}_t, \mathcal{A}_{t+1}) - \mathrm{Q}'(\mathcal{S}_t, \mathcal{A}_{t+1}) \right)^2,\end{aligned} \tag{10}$$

where γ is the discount factor in RL, $\mathrm{Q}_{\mathrm{target}}$ is the target network in RL, which is saved from Q every 1000 step. We use the gradient descent method to update Q parameters.

During the test stage, we only need to evaluate the advantage network while discarding the state value network. The selected light direction $\mathcal{A}^*_{t+1}$ is:

$$\mathcal{A}^*_{t+1} = \underset{\mathcal{A}_{t+1}}{\mathrm{argmax}} \frac{1}{HW} \sum_{i,j} \mathrm{Adv}(\mathcal{S}_t, \mathcal{A}_{t+1})_{ij}. \tag{11}$$

We use a replay buffer in our dueling DQN design to learn from prior experiences and enhance sampling efficiency. Specifically, a replay buffer stores past experiences of the agent in the form of transitions $(\mathcal{S}_{t-1}, \mathcal{A}_t, \mathcal{R}_t, \mathcal{S}_t)$. The agent samples a mini-batch of transitions from the replay buffer to update its Q-value function. We specially design an exploration strategy to facilitate exploration and assist in network training (details in supplementary material). The overall pipeline of our agent network for illumination planning is summarized in Algorithm. 1.

Architecture design of dueling DQN. We design the network structure of the state value network and the advantage network following the insights from learning-based photometric stereo methods. As discussed in [29, 7], the network architecture in learning-based photometric stereo are usually divided into the per-pixel branch and the all-pixel branch. The methods in the per-pixel branch such as CNN-PS [4] address the inter-image intensity variation at each pixel via an observation map, while the methods in the all-pixel branch such as PS-FCN [2] pay more attention to the intra-image intensity variation of each input image via image feature extraction. The dueling DQN architecture is designed by leveraging priors from the above two branches together.

From t images of shape $H \times W$ in state $\mathcal{S}_t$, we apply the feature extraction module from PS-FCN [2] to obtain the 4D global feature map of shape $H \times W \times t \times C_1$, where C_1 is the dimension of the feature vector (Fig. 3 (a)). In the advantage network, we treat the global feature map as an image where each pixel contains t stacked feature vectors of size C_1 (Fig. 3 (b)). For every pixel, we project each of its feature vectors to a position on a per-pixel 3D observation feature [4] with the size of $L_y \times L_x \times C_1$ where L_y and L_x represent the number of discretized light directions vertically and horizontally, according to the light directions in $\mathcal{S}_t$. Each observation feature is passed to a series of shared-weight convolutional layers to get the advantage feature of shape $L_y \times L_x$. This feature from all pixels is combined to form the advantage map $\boldsymbol{A} \in \mathbb{R}^{H \times W \times L_y \times L_x}$.

In the state value network, the global feature map is passed through multiple convolutional layers to estimate the state value map $\boldsymbol{V} \in \mathbb{R}^{H \times W \times 1 \times 1}$ (Fig. 3 (c)). Given the output of the advantage map and state value map, we can calculate the Q-value map, $\boldsymbol{Q} \in \mathbb{R}^{H \times W \times L_y \times L_x}$ following Eq. (9), which is used to infer the optimal light direction for the next state.

4.3. Datasets and Implementation

In this paper, we employ rendered images from synthetic datasets (Blobby [6] and Sculpture [23]) for model training and evaluate the generalization ability using real datasets (DiLiGenT [18] and DiLiGenT10[2] [14]).

Synthetic dataset processing. Based on the synthetic datasets in [6, 23], we randomly split their geometries into training and test sets in a 4:1 ratio and apply random scale, rotation, and translation for data augmentation. This results

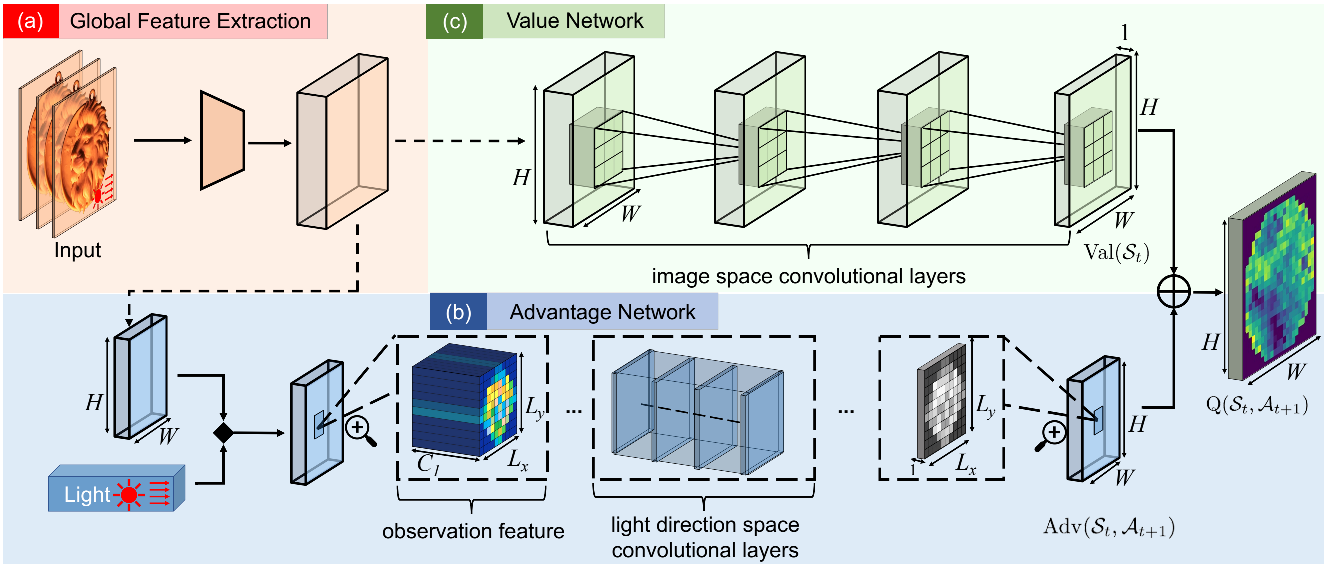

Figure 3. The network architecture of dueling DQN [22] for ReLeaPS, which contains **(a)** the global feature extraction, **(b)** the advantage network, and **(c)** the state value network. The global feature extraction is to extract the features pixel-wisely from input images into a feature map; the advantage network transforms the feature map into an observation feature and predicts the action based on light directions; the state value network extracts the non-Lambertian effects from the feature map for prediction. The Q-value function in Eq. (9) is expressed as the sum of the state value and advantage function.

in 500 training samples and 50 test samples for each dataset. We use a physically-based ray-tracing renderer to generate images under different light distributions, with a spatial resolution of 256×256.

Real dataset acquisition and benchmark. We evaluate the generalization ability of our proposed method using two real datasets, DiLiGenT [18] and DiLiGenT10^2 [14]. To obtain performance results for DiLiGenT10^2 [14], we submit the normal map to the benchmark website[1]. To further validate the efficacy of our method in real-world scenarios, we build a setup with a movable light source that enables illumination from any direction within the upper hemisphere, as shown in Fig. 1. We show one of the captured scenes LionHead in the main paper[2].

The real datasets in [18, 14] for evaluation are captured in advance and exhibit limited coverage of images under a fixed light distribution, leading to a restricted action space for our method during evaluation. To tackle this issue and allow our approach to accommodate varying light distributions, we map each light direction to its nearest neighbor within our predefined settings. We exclude the Q-values of light directions that are not observed and select the action from the remaining set.

Implementation details. Our framework is implemented in PyTorch. The total steps for illumination planning are 20, i.e., $T = 20$. The other parameters are set to $\gamma = 0.9$ and $\alpha = 0.5$. We train our model using AdamW [10] optimizer for $10,000$ episodes. The learning rate is 10^{-5} and weight decay is 10^{-8} respectively. The replay buffer size is set to $16,384$. It takes about 2 days to train on a single NVIDIA GeForce RTX 3090 graphics card.

5. Experiments

Our experimental evaluation consists of four subsections: 1) improving photometric stereo backbones, 2) ilumination planning method comparison, 3) quantitative analysis on real benchmarks, and 4) ablation studies.

5.1. Improving Photometric Stereo Backbones

Theoretically, our proposed method can be employed in conjunction with any photometric stereo backbone that utilizes a set of images and light directions as input. We evaluate ReLeaPS on three representative photometric stereo backbones: the least square (LS) [24] method representing conventional methods, CNN-PS [4] representing per-pixel learning methods, and PS-FCN [2] representing all-pixel learning methods. In accordance with PS-FCN [2], we employ a random selection of 20 light directions to represent the vanilla performance of the existing photometric stereo backbone method under limited illumination. Next, we compare the vanilla performance of photometric stereo backbones by evaluating them using ReLeaPS under 20 light directions to demonstrate the improved performance achieved by our method. The experimental results, presented in Table 2, demonstrate a reduction in mean angular error across all datasets and photometric stereo backbones

[1] `https://photometricstereo.github.io/diligent102.html`

[2] Please refer to the supplementary video showing our online illumination planning and capture process.

Table 2. Quantitative comparisons of different illumination planning approaches in terms of mean angular error on Blobby [6], Sculpture [23], DiLiGenT [18], and DiLiGenT10^2 [14] datasets using different photometric stereo backbones with 20 lights. 'Rnd.' stands for the random selection of light directions averaged over 10 evaluations and 'BF' stands for brute-force strategy.

Dataset	PS backbone	Rnd.	Illumination planning			Oracle	
			DC05 [3]	TK22 [21]	Ours	All	BF *
Blobby [6]	LS [24]	3.4	3.4	3.5	**2.9**	3.3	1.9
	CNN-PS [4]	7.2	12.3	7.2	**4.8**	2.3	N/A
	PS-FCN [2]	3.1	3.4	3.6	**2.6**	2.5	
Sculpture [23]	LS [24]	10.6	11.3	10.2	**9.7**	10.5	7.0
	CNN-PS [4]	12.3	20.9	10.2	**7.6**	5.2	N/A
	PS-FCN [2]	7.0	7.4	7.1	**6.5**	5.9	
DiLiGenT [18]	LS [24]	14.0	13.9	**13.8**	**13.8**	13.9	12.6
	CNN-PS [4]	10.2	**9.9**	12.1	10.1	7.4	N/A
	PS-FCN [2]	8.3	7.8	8.1	**7.7**	7.5	
DiLiGenT10^2 [14]	LS [24]	24.6	24.7	24.4	**23.5**	22.1	N/A
	CNN-PS [4]	18.3	21.3	18.3	**17.7**	16.4	
	PS-FCN [2]	15.9	15.9	16.6	**15.8**	15.9	

*Some of the results are not available (N/A) due to high computational cost (CNN-PS [4] and PS-FCN [2]) or lack of ground truth normal (DiLiGenT10^2 [14]).

when comparing the 'Rnd.' column (vanilla performance) with the 'Ours' column (enhanced performance). On average, ReLeaPS achieves an improvement in a mean angular error of 1.02 degrees. Remarkably, for rows 1, 4, 7, and 12, our performance using only 20 lights even outperforms that of using all (about 100) input images, demonstrating the effectiveness of our method in enhancing many different photometric stereo backbones.

5.2. Illumination Planning Method Comparison.

We conduct a comparative study of ReLeaPS with two existing illumination planning approaches, namely DC05 [3] and TK22 [21], under 20 lights. For DC05 [3], we select the observation configuration with a vertical light direction as it yields better performance. The results presented in Table 2 demonstrate that our proposed method outperforms other illumination planning methods in terms of a mean angular error, except for the CNN-PS [4] backbone on the DiLiGenT [18] benchmark. We attribute this to the narrow light distribution (centered in a small range) in DiLiGenT, which limits the ability to demonstrate the advantage of using RL for illumination planning. DiLiGenT [18] use light sources in a planar grid, which only covers a small portion of the full hemisphere, while our rendered synthetic datasets and DiLiGenT10^2 [14] set the light sources on the full hemisphere. This design difference may have an impact on the performance of illumination planning methods. Overall, our method improves the performance of mean angular error using 20 light directions by 2.46 compared to DC05 [3] and 1.03 compared to TK22 [21].

Additionally, we employ a brute-force (BF) strategy for illumination planning in the LS [24] method on specific rows (1, 4, and 7). This method exhaustively searched all possible light directions for one step and selected the one with the lowest mean angular error compared to the ground truth. The results demonstrate that this oracle method achieves an average improvement of 1.63 in mean angular error compared to our method, highlighting the potential for further improvements in illumination planning.

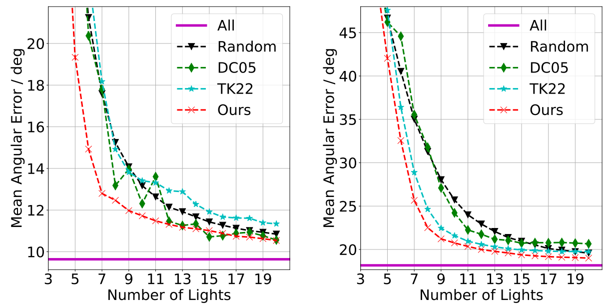

Figure 4. Quantitative evaluation of illumination planning methods w.r.t. to the different number of light directions on real-world benchmarks: DiLiGenT [18] **(left)** and DiLiGenT10^2 [14] **(right)**. The mean angular error is averaged among LS [24], CNN-PS [4], and PS-FCN [2] backbones.

5.3. Quantitative Analysis on Real Benchmarks

ReLeaPS is designed to minimize the mean angular error under 20 lights while also improving performance across a range of 3 to 20 lights. We compare our proposed method with other illumination planning methods under varying numbers of light directions on two real datasets, DiLiGenT [18] and DiLiGenT10^2 [14]. We show the average performance of mean angular error among all three photometric stereo backbones (LS [24], CNN-PS [4], and PS-FCN [2]) in Fig. 4. Our results demonstrate that ReLeaPS outperforms other illumination planning methods for the most number of light directions, except for $14 \sim 20$ lights on the DiLiGenT [18] dataset. We attribute this to the narrow light distribution in DiLiGenT [18], as described in Section 5.2. On the DiLiGenT10^2 [14] side, we achieve the same 20-light performance as TK22 [21] using only 14 lights. Overall, ReLeaPS achieves similar or better performance than other illumination planning methods, even with a smaller number of lights.

Furthermore, the effectiveness of ReLeaPS is being validated through qualitative comparisons in Fig. 5, where it is compared to other methods using real data. Specifically, on the left side of the figure, the presence of shadows in the red circle of READING from the DiLiGenT [18] dataset results in a significant angular error. Although DC05 [3] and TK22 [21] can recover these areas using 15 lights, ReLeaPS accomplishes comparable results with only 7 lights. On the right side of Fig. 5, the LIONHEAD object, captured using our setup, is challenging to reconstruct using DC05 [3]. However, ReLeaPS efficiently recovers the normal map of the object with only 11 lights. These results authenticate the

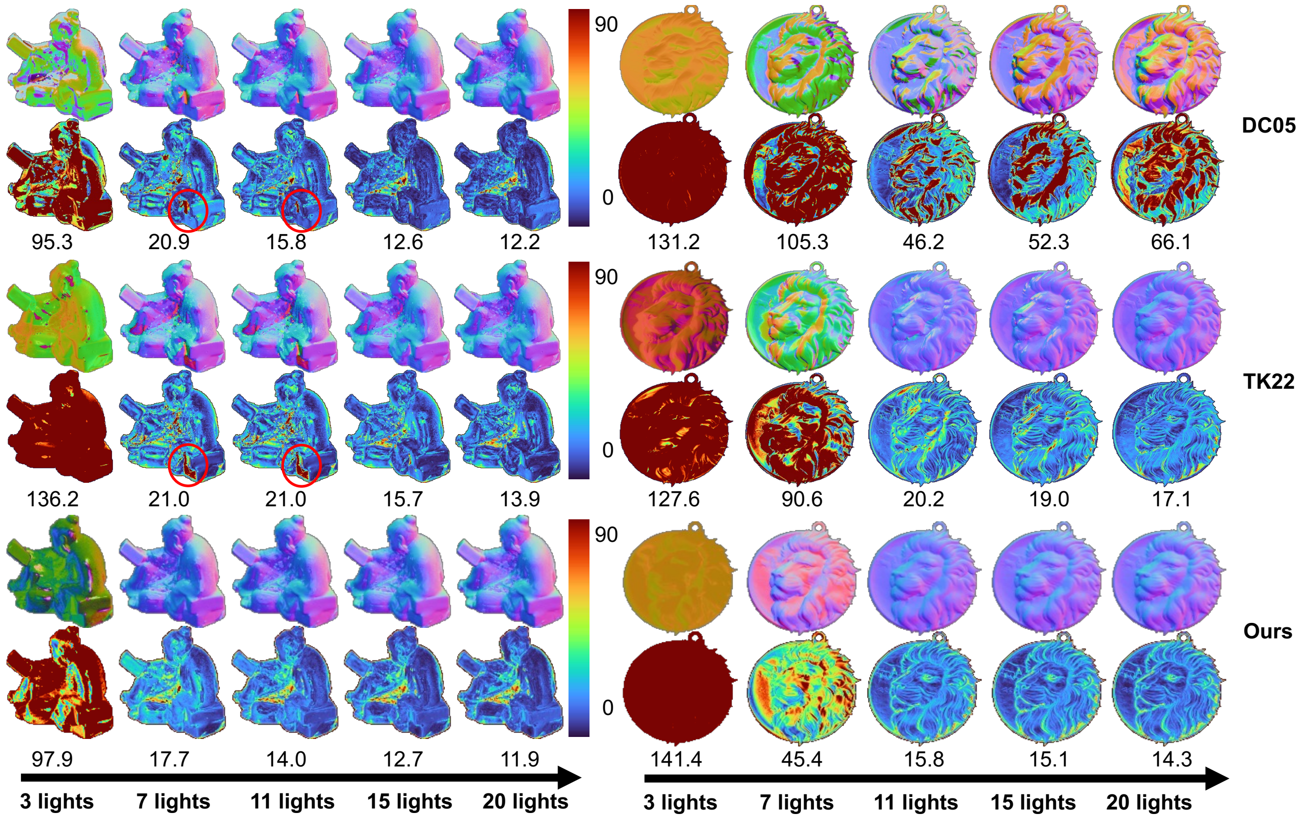

Figure 5. Qualitative comparison of recovered surface normals and error maps for **(left)** READING (from DiLiGenT [18]) and **(right)** LIONHEAD (captured using our setup) using different illumination planning methods (*i.e.*, DC05 [3], TK22 [21], and Ours) with increasing light directions (3, 7, 11, 15, and 20 lights) in CNN-PS [4] backbone. The red circle indicates a region with shadows that cannot be effectively recovered by CNN-PS [4], resulting in a large angular error.

Table 3. Ablation studies on ReLeaPS with 20 lights on Blobby [6], Sculpture [23], DiLiGenT [18], and DiLiGenT10^2 [14] datasets using the LS [24] backbone: including generalized image formation, reward design, and network design (simplified to 'ReLeaPS w/ Lambert', 'ReLeaPS w/o reward', and 'ReLeaPS w/ MLP' respectively).

Dataset	ReLeaPS w/ Lambert	ReLeaPS w/o reward	ReLeaPS w/ MLP	Ours
Blobby [6]	3.7	3.5	3.3	**2.9**
Sculpture [23]	12.9	11.9	10.6	**9.7**
DiLiGenT [18]	14.8	13.9	13.9	**13.8**
DiLiGenT10^2[14]	24.2	24.1	24.5	**23.5**

effectiveness of ReLeaPS in comparison to other illumination planning techniques.

5.4. Ablation Studies

We conduct ablation studies by removing the important components of ReLeaPS, *i.e.*, the inclusion of complex light transport effects in image formation (simplified to 'ReLeaPS w/ Lambert'), a deep dueling network for generalized photometric stereo (simplified to 'ReLeaPS w/ MLP'), and a sparse-to-dense reward function (simplified to 'ReLeaPS w/o reward'). More details about the simplified versions of ReLeaPS are shown in the supplementary material.

The results in Table 3 demonstrate that removing each part of ReLeaPS causes harm to the performance, proving the effectiveness of our specific designs in ReLeaPS.

6. Conclusion

We propose ReLeaPS, an RL-based online illumination planning method for generalized photometric stereo, which shows promising results under a limited number of illuminations. However, there are still some limitations:

- Compared with devices using fixed light sources, the moving parts in our validation setup make it less portable and slow in movement;
- ReLeaPS assumes the classic photometric stereo setup of distant light and orthographic camera. Near light and perspective camera models are not considered.

We hope our work shows promising potential for further improvements in illumination planning for photometric stereo and related applications.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant No. 62136001, 62088102. Heng Guo was supported by JSPS KAKENHI (Grant No. JP23H05491).

References

[1] Manmohan Chandraker, Sameer Agarwal, and David Kriegman. Shadowcuts: Photometric stereo with shadows. In *Proc. of Computer Vision and Pattern Recognition*, 2007.

[2] Guanying Chen, Kai Han, and Kwan-Yee K. Wong. PS-FCN: A flexible learning framework for photometric stereo. In *Proc. of European Conference on Computer Vision*, 2018.

[3] Ondrej Drbohlav and Mike Chantler. On optimal light configurations in photometric stereo. In *Proc. of International Conference on Computer Vision*, 2005.

[4] Satoshi Ikehata. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. In *Proc. of European Conference on Computer Vision*, 2018.

[5] Takafumi Iwaguchi and Hiroshi Kawasaki. Surface normal estimation from optimized and distributed light sources using dnn-based photometric stereo. In *Proc. of IEEE Winter Conference on Applications of Computer Vision*, 2023.

[6] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. In *Proc. of Computer Vision and Pattern Recognition*, 2011.

[7] Yakun Ju, Kin-Man Lam, Wuyuan Xie, Huiyu Zhou, Junyu Dong, and Boxin Shi. Deep learning methods for calibrated photometric stereo and beyond: A survey. *arXiv preprint arXiv:2212.08414*, 2022.

[8] Junxuan Li, Antonio Robles-Kelly, Shaodi You, and Yasuyuki Matsushita. Learning to minify photometric stereo. In *Proc. of Computer Vision and Pattern Recognition*, 2019.

[9] Fotios Logothetis, Ignas Budvytis, Roberto Mecca, and Roberto Cipolla. PX-Net: Simple and efficient pixel-wise training of photometric stereo networks. In *Proc. of International Conference on Computer Vision*, 2021.

[10] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *Proc. of International Conference on Learning Representations*, 2019.

[11] Daisuke Miyazaki, Kenji Hara, and Katsushi Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision*, 86:229–242, 2009.

[12] Yasuhiro Mukaigawa, Yasunori Ishii, and Takeshi Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 24:3326–34, 11 2007.

[13] Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. Distort-and-recover: Color enhancement using deep reinforcement learning. In *Proc. of Computer Vision and Pattern Recognition*, 2018.

[14] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. $DiLiGenT10^2$: A photometric stereo benchmark dataset with controlled shape and material variation. In *Proc. of Computer Vision and Pattern Recognition*, 2022.

[15] Yasuyuki Matsushita Satoshi Ikehata, David P. Wipf and Kiyoharu Aizawa. Photometric stereo using sparse bayesian regression for general diffuse surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(9):1078–1091, 2014.

[16] Yasuyuki Matsushita Satoshi Ikehata, David Wipf and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *Proc. of Computer Vision and Pattern Recognition*, 2012.

[17] Boxin Shi, Zhipeng Mo, Zhe Wu, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):271–284, 2019.

[18] 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 *Proc. of Computer Vision and Pattern Recognition*, 2016.

[19] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 1998.

[20] Kam-Lun Tang, Chi-Keung Tang, and Tien-Tsin Wong. Dense photometric stereo using tensorial belief propagation. In *Proc. of Computer Vision and Pattern Recognition*, 2005.

[21] Hirochika Tanikawa, Ryo Kawahara, and Takahiro Okabe. Online illumination planning for shadow-robust photometric stereo. In *International Workshop on Frontiers of Computer Vision*, 2022.

[22] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. Dueling network architectures for deep reinforcement learning. In *Proc. of International Conference on Machine Learning*, 2016.

[23] Olivia Wiles and Andrew Zisserman. SILNet: Single- and multi-view reconstruction by learning from silhouettes. *arXiv preprint arXiv:1711.07888*, 2017.

[24] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 01 1980.

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

[26] Chanki Yu, Yongduek Seo, and Sang Wook Lee. Photometric stereo from maximum feasible lambertian reflections. In *Proc. of European Conference on Computer Vision*, 2010.

[27] Rongkai Zhang, Lanqing Guo, Siyu Huang, and Bihan Wen. Rellie: Deep reinforcement learning for customized low-light image enhancement. In *Proc. of ACM International Conference on Multimedia*, 2021.

[28] 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 *Proc. of International Conference on Computer Vision*, 2019.

[29] Qian Zheng, Boxin Shi, and Gang Pan. Summary study of data-driven photometric stereo methods. *Virtual Reality & Intelligent Hardware*, 2(3):213–221, 2020.