Polarization Guided Mask-Free Shadow Removal

Chu Zhou[1 †], **Chao Xu**[2], **Boxin Shi**[3,4 #]

[1]National Institute of Informatics, Japan
[2]National Key Laboratory of General Artificial Intelligence, School of IST, Peking University, China
[3]State Key Laboratory for Multimedia Information Processing, School of CS, Peking University, China
[4]National Engineering Research Center of Visual Technology, School of CS, Peking University, China

Abstract

Shadow is a phenomenon that degenerates image quality and decreases the performance of downstream vision algorithms. Despite the fact that current image shadow removal methods have achieved promising progress, many of them require an externally obtained shadow mask as a necessary part of the input data, which not only introduces additional workload but also leads to degenerated performance near the shadow boundary due to the inaccuracy of the mask. Some of them do not require the shadow mask, however, they need to simultaneously consider the restoration of the brightness and color information along with the preservation of the texture and structure information inside the shadow region without external clues, which poses highly ill-posedness and makes the results prone to artifacts. In this paper, we propose **Pol-ShaRe**, the first **Pol**arization-guided image **Sha**dow **Re**moval solution, to remove shadow in a mask-free manner with fewer artifacts. Specifically, it consists of a two-stage pipeline to relieve the ill-posedness and a neural network tailored to the pipeline to suppress the artifacts. Experimental results show that our Pol-ShaRe achieves state-of-the-art performance on both synthetic and real-world images.

Introduction

Shadow is a frequently occurring phenomenon present in natural images when the light is partially or completely blocked in a certain region. The existence of shadow degenerates the image quality with color and brightness degradation, which would decrease the performance of downstream vision applications and lead to poor photography experience of users. Unlike the common image enhancement tasks that solve global degradation problems (*e.g.*, dehazing and deraining in the whole image plane), shadow removal aims to solve a spatially non-uniform partial degradation problem (*i.e.*, recovering the original pixel values inside the shadow region), posing unique challenges (Guo et al. 2023a,b).

Early works (Tian and Tang 2011; Guo, Dai, and Hoiem 2012; Yang, Tan, and Ahuja 2012; Khan et al. 2015) use priors from image statistics to solve this problem. However, their applicability is limited since they are based on time-consuming numerical optimization. Recently, deep-learning has been introduced to handle it with higher efficiency, by adopting deep neural networks with different architectures (*e.g.*, CNN (Li et al. 2023; Niu et al. 2022; Wan et al. 2022), GAN (Liu et al. 2023a,b; Zhang et al. 2020), vision Transformer (Guo et al. 2023a), and diffusion model (Guo et al. 2023b)) to extract image features from a large amount of training data. Despite that these learning-based methods could produce plausible results, they still encounter two key issues: (1) *Dependency on the shadow mask*: As shown in Fig. 1 (a), many of them require a shadow mask as a necessary part of the input data (Li et al. 2023; Niu et al. 2022; Wan et al. 2022; Liu et al. 2023a; Guo et al. 2023a,b), while obtaining the mask usually relies on external shadow detection or manual annotation approaches, introducing additional workload; besides, as mentioned in (Guo et al. 2023b), since the externally obtained mask is always not that accurate, their performance would degenerate near the shadow boundary. (2) *Artifacts inside the shadow region*: there are also some methods do not require the shadow mask (Liu et al. 2023b; Zhang et al. 2020), however, the problem they face is highly ill-posed since it requires to simultaneously consider the restoration of the brightness and color information along with the preservation of the texture and structure information inside the shadow region without external clues, making the results prone to artifacts (*e.g.*, false color and less-distinctive textures, as shown in Fig. 1 (b)).

In this paper, we analyze the shadow image formation model, and propose **Pol-ShaRe**, the first **Pol**arization-guided image **Sha**dow **Re**moval solution, as shown in Fig. 1 (c). By exploiting the priors in both the intensity and polarization domains to indicate the per-pixel shadow confidence, our solution can get rid of the dependency on the shadow mask. It consists of a processing pipeline and a neural network tailored to the pipeline. To relieve the ill-posedness of the problem, we design the pipeline to be two-stage to explicitly decouple the restoration of the brightness and color information from the preservation of the texture and structure information, by reformulating the shadow removal problem into two consecutive guided information reconstruction problems. Specifically, under the guidance of the priors, the first stage aims to reconstruct the total intensity modulated by the degree of polarization of the incoming light to the

† Most of this work was done as a PhD student at Peking University.

# Corresponding author: shiboxin@pku.edu.cn

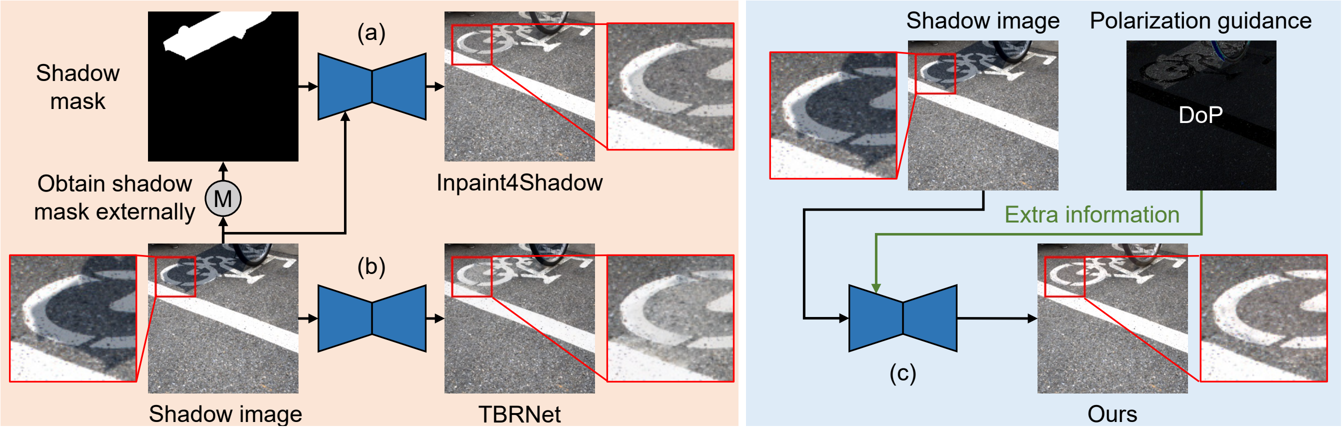

Figure 1: (a) Many shadow removal methods require an externally obtained shadow mask (*e.g.*, Inpaint4Shadow (Li et al. 2023)), which not only introduces additional workload but also leads to degenerated performance near the shadow boundary due to the frequently occurring inaccuracy of the mask. (b) There are also some methods that do not require the shadow mask (*e.g.*, TBRNet (Liu et al. 2023b)), however, they tend to suffer from artifacts inside the shadow region, such as false color and less-distinctive textures. (c) Our Pol-ShaRe can remove shadow in a mask-free manner with fewer artifacts thanks to the extra information provided by the guidance of polarization from the DoP (degree of polarization).

sensor, which not only encodes distinctive texture and structure information but also has better resistance to shadow; the second stage aims to reconstruct the shadow-free image using the recovered modulated total intensity as an extra guidance, focusing on the brightness and color information. To suppress the artifacts inside the shadow region, we design our network to be composed of dual domain prior fusion (DDPF) and texture guided demodulation (TGD) modules by making full use of the physical properties lying in the degradation procedure. To summarize, this paper makes contributions by showing: (1) the first polarization-guided image shadow removal solution exploiting the dual domain priors from the shadow image formation model without the dependency on the shadow mask, consisting of: (2) a two-stage pipeline to relieve the ill-posedness of the problem, by explicitly decoupling the restoration of the brightness and color information from the preservation of the texture and structure information; and (3) a neural network tailored to the pipeline to suppress the artifacts inside the shadow region, by integrating physics-oriented modules fully taking into account the degradation procedure.

Experimental results show our Pol-ShaRe achieves state-of-the-art performance on both synthetic and real images.

Related work

Image shadow removal. Early works attempted to solve this problem by adopting numerical optimization (Tian and Tang 2011; Guo, Dai, and Hoiem 2012; Yang, Tan, and Ahuja 2012; Khan et al. 2015) based on the priors from natural image statistics. With the development of deep neural networks, learning-based methods have also been adopted to handle this problem, showing higher efficiency. Generally, these learning-based methods could be divided into three categories: supervised, unsupervised (or weakly-supervised), and semi-supervised methods. Supervised methods usually achieve better performance in recovering the original pixel values, by extracting image features from a large amount of paired data. They adopt different strategies, such as directly reconstructing the corresponding shadow-free image (Guo et al. 2023b; Niu et al. 2022; Wan et al. 2022; Zhu et al. 2022a; Chen et al. 2021; Cun, Pun, and Shi 2020; Hu et al. 2019a; Liu et al. 2023c; Jin et al. 2024), explicitly learning the residual (Guo et al. 2023a; Liu et al. 2023a,b; Yücel et al. 2023) or multiplicator (Zhu et al. 2022b; He et al. 2021; Qu et al. 2017) (or both of them (Zhang et al. 2020)) between the shadow and shadow-free images, predicting the shadow parameters (Le and Samaras 2019), estimating and fusing multiple images with different exposures (Fu et al. 2021) or gammas (Sen et al. 2023), treating the shadow removal task as an inpainting task (Li et al. 2023), *etc*. Unsupervised methods, which do not rely on any paired data for training (Guo et al. 2023c; Liu et al. 2021b; Le and Samaras 2021; Liu et al. 2021a; Hu et al. 2019b), usually have better generalization ability and convenience. Semi-supervised methods (Ding et al. 2019) adopt a compromise approach. They only use a small amount of paired data during training, aiming to balance the performance and generalization ability. However, since these methods are image-based ones that cannot make use of the information provided by other modalities (*e.g.*, the polarization-relevant parameters), they are more likely to encounter performance bottlenecks.

Polarization-based vision. Polarization is an important property of light in addition to its amplitude and phase, and it has been widely introduced into the field of computer vision. Recent polarization-based vision algorithms can be divided into two categories: the first category aims to solve the high-level vision problems, such as transparent object segmentation (Kalra et al. 2020; Mei et al. 2022), road scene analysis and understanding (Li et al. 2020; Liang et al. 2022),

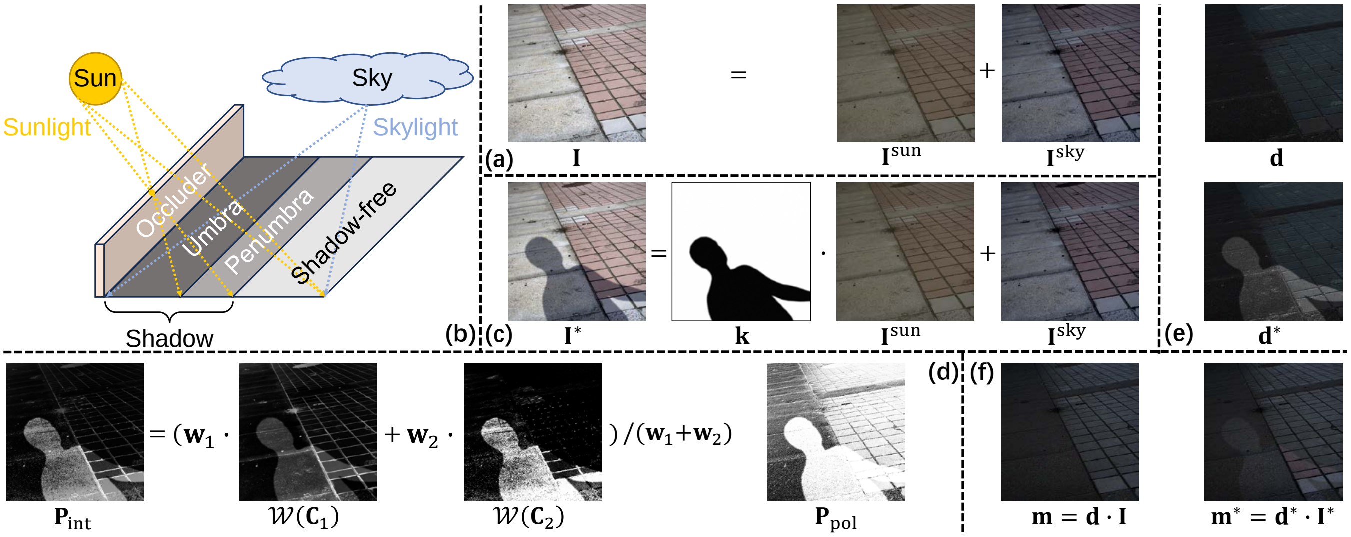
Sun
Sky
Sunlight
Skylight
Occluder
Umbra
Penumbra
Shadow-free
Shadow
(a)
$\mathbf{I}$
=
$\mathbf{I}^{\text{sun}}$
+
$\mathbf{I}^{\text{sky}}$
$\mathbf{d}$
(b)
(c)
$\mathbf{I}^*$
=
$\mathbf{k}$
·
$\mathbf{I}^{\text{sun}}$
+
$\mathbf{I}^{\text{sky}}$
(e)
$\mathbf{d}^*$
$\mathbf{P}_{\text{int}}$
$= (\mathbf{w}_1 \cdot$
$\mathcal{W}(\mathbf{C}_1)$
$+ \mathbf{w}_2 \cdot$
$\mathcal{W}(\mathbf{C}_2)$
$)/(\mathbf{w}_1 + \mathbf{w}_2)$
$\mathbf{P}_{\text{pol}}$
(d)
(f)
$\mathbf{m} = \mathbf{d} \cdot \mathbf{I}$
$\mathbf{m}^* = \mathbf{d}^* \cdot \mathbf{I}^*$

Figure 2: (a) When taking photos in outdoor scenes lit by daylight under sunny weather, the total intensity of the captured image $\mathbf{I}$ mainly includes two components: the sunlight $\mathbf{I}^{\text{sun}}$ and skylight $\mathbf{I}^{\text{sky}}$. (b) Shadow occurs when the sunlight is blocked by the occluder in the background completely (causing umbra) or partially (causing penumbra). (c) When shadow occurs, $\mathbf{I}$ degenerates into $\mathbf{I}^*$, which depends on the soft blocking matte $\mathbf{k}$. (d) $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$: The priors in the intensity and polarization domains respectively (both $\mathbf{w}_{1,2}$ are set to be $\mathbf{1}$ here for visualization). (e) $\mathbf{d}$ and $\mathbf{d}^*$: the DoP (degree of polarization) of the incoming light to the sensor without and with shadow respectively. (f) $\mathbf{m}$ and $\mathbf{m}^*$: the total intensity modulated by the DoP without and with shadow respectively, which often have closer appearance to each other compared with $\mathbf{I}$ and $\mathbf{I}^*$.

car detection (Blin et al. 2019), *etc*.; the second category aims to solve the low-level vision problems, such as shape estimation (Ba et al. 2020; Lyu et al. 2023), inverse rendering (Dave, Zhao, and Veeraraghavan 2022), depth sensing (Kadambi et al. 2017; Tian et al. 2023), image enhancement (*e.g*., reflection removal (Lei et al. 2020; Lyu et al. 2022), image dehazing (Zhou et al. 2021), color constancy (Ono et al. 2022), and HDR reconstruction (Zhou et al. 2023a)), *etc*. By modeling the formation of image in a polarization perspective and fully using the unique information encoded in the polarization-relevant parameters, these algorithms often achieve higher performance compared with the image-based ones. Our Pol-ShaRe for the first time explores the polarization guidance for solving the image shadow removal problem, which belongs to the second category.

Method

Dual domain priors

Intensity domain. Considering the outdoor scenes lit by daylight under sunny weather (denoted as $\mathbf{I}$), there are mainly two light sources: direct sunlight $\mathbf{I}^{\text{sun}}$ and ambient skylight $\mathbf{I}^{\text{sky}}$ (Tian and Tang 2011), as shown in Fig. 2 (a). Assume for a moment some occluders in the background block the sunlight completely or partially in a certain region, which cast shadow in the image plane (Tian and Tang 2011), as shown in Fig. 2 (b), the captured image can be described as $\mathbf{I}^*$[1]:

$$\mathbf{I}^* = \mathbf{k} \cdot \mathbf{I}^{\text{sun}} + \mathbf{I}^{\text{sky}}, \tag{1}$$

where $\mathbf{k} \in [0, 1]$ is a soft blocking matte, and $\cdot$ denotes element-wise multiplication, as shown in Fig. 2 (c). For a certain pixel $\mathbf{x}$, if $\mathbf{k}(\mathbf{x}) = 1$, $\mathbf{x}$ is in the shadow-free region; otherwise, $\mathbf{x}$ is in the shadow region. The shadow region can be further divided into the umbra region ($\mathbf{k}(\mathbf{x}) = 0$), which is the inner area and often occupies a large part, and the penumbra region ($0 < \mathbf{k}(\mathbf{x}) < 1$), which is near the boundaries and often occupies a small part. According to Eq. (1), we can see in the shadow region the brightness is attenuated and $\mathbf{I}^{\text{sky}}$ accounts for a dominant proportion (*i.e*., $\mathbf{I}^*$ would have similar properties to $\mathbf{I}^{\text{sky}}$). Since the skylight illumination often becomes bluish caused by Rayleigh scattering in such a condition (Holstein 1999), the blue channel is generally brighter than the other channels in the shadow region (Huang et al. 2011; Inoue and Yamasaki 2020; Ono et al. 2022). In short, $\mathbf{I}^*$ usually satisfies both of the following *two common senses* in the shadow region: (I) The pixel values are relatively smaller. (II) The pixel values in the blue channel are relatively larger than the other ones. Here, inspired by Ono et al. (2022), we adopt a weighting function $\mathcal{W}(\cdot)$ to turn the above common senses into an intensity domain numerical prior $\mathbf{P}_{\text{int}}$:

$$\mathbf{P}_{\text{int}} = (\mathbf{w}_1 \cdot \mathcal{W}(\mathbf{C}_1) + \mathbf{w}_2 \cdot \mathcal{W}(\mathbf{C}_2))/(\mathbf{w}_1 + \mathbf{w}_2), \tag{2}$$

where $\mathbf{C}_1 = (1 - \mathbf{I}^*_{\text{R}}) \cdot (1 - \mathbf{I}^*_{\text{G}}) \cdot (1 - \mathbf{I}^*_{\text{B}})$ and $\mathbf{C}_2 = (2\mathbf{I}^*_{\text{B}} - \mathbf{I}^*_{\text{R}} - \mathbf{I}^*_{\text{G}})/(2\mathbf{I}^*_{\text{AVG}})$ are two conditions satisfied by common senses (I) and (II) respectively[2], and $\mathbf{w}_{1,2} \in [0, 1]$

[1] We use the superscript $*$ to mark the degenerated variable (caused by shadow) throughout this paper.

[2] All pixel values are normalized to $[0, 1]$, and the subscript RGB and AVG denote the color channel index and the channel average respectively throughout this paper.

denote the weights which are learned parameters to refine the quality of $\mathbf{P}_{\text{int}}$. A visual example can be found on the left side of Fig. 2 (d), where both $\mathbf{w}_{1,2}$ are set to be $\mathbf{1}$ here for visualization. We can see that $\mathbf{P}_{\text{int}}$ has similar boundaries to the soft blocking matte $\mathbf{k}$, and it could be used to indicate the per-pixel shadow confidence for further shadow removal. *Details of $\mathcal{W}(\cdot)$ can be found in the supplementary material.*

Polarization domain. When placing a linear polarizer in front of the camera, the captured polarized image $\mathbf{I}_\alpha$ can be calculated using Malus' law (Hecht et al. 2002):

$$\mathbf{I}_\alpha = \mathbf{I} \cdot (1 - \mathbf{d} \cdot \cos(2\alpha - 2\boldsymbol{\theta}))/2, \tag{3}$$

where $\alpha \in [0, \pi]$ is the polarizer angle (the orientation of the polarizer), $\mathbf{d} \in [0, 1]$ and $\boldsymbol{\theta} \in [0, \pi]$ are the DoP (degree of polarization) and AoP (angle of polarization) of the incoming light to the sensor respectively. Since the skylight illumination tends to be more significantly polarized in outdoor scenes lit by daylight under sunny weather (Sekera 1957; Zhou et al. 2021), the overall DoP value in the shadow region is often relatively larger (Lin et al. 2006), as shown in Fig. 2 (e). However, considering the fact that the DoP value is often imbalanced across the three color channels due to its strong correlation to the wavelength (Pust and Shaw 2012), we could know that even in the shadow-free region there could still be a great number of pixels with relatively larger DoP values in one of the color channels. To avoid this issue, we choose to compute the dark channel of the overall DoP $\min(\mathbf{d}^*_{\text{R}}, \mathbf{d}^*_{\text{G}}, \mathbf{d}^*_{\text{B}})$ as the condition instead of using the overall DoP $\mathbf{d}^*$ itself. Similar to Eq. (2), we can obtain the following prior in the polarization domain $\mathbf{P}_{\text{pol}}$:

$$\mathbf{P}_{\text{pol}} = \mathcal{W}(\min(\mathbf{d}^*_{\text{R}}, \mathbf{d}^*_{\text{G}}, \mathbf{d}^*_{\text{B}})). \tag{4}$$

A visual example is shown on the right side of Fig. 2 (d).

As the dual domain priors ($\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$) become available, our solution can not only make use of the modality information of polarization, but also get rid of the dependency on the shadow mask, reducing the workload and increasing the robustness near the shadow boundary.

Two-stage shadow removal pipeline

We aim to reconstruct the shadow-free image $\mathbf{I}$ from the corresponding degenerated image $\mathbf{I}^*$ under the guidance of the DoP $\mathbf{d}^*$. To acquire $\mathbf{d}^*$, in addition to the unpolarized image $\mathbf{I}^*$, it requires at least two extra polarized images $\mathbf{I}^*_{\alpha_{1,2}}$ with different polarizer angles $\alpha_{1,2}$. Note that $\mathbf{I}^*_{\alpha_{1,2}}$ can be easily captured by placing a linear polarizer in front of the camera with two different orientations. Here we first explain how to acquire the DoP from these images. Rewriting Eq. (3), $\mathbf{I}_\alpha$ can be expressed as a linear combination of $\mathbf{S}_{0,1,2}$:

$$\begin{aligned} &\mathbf{I}_\alpha = (\mathbf{S}_0 - \cos(2\alpha)\mathbf{S}_1 - \sin(2\alpha)\mathbf{S}_2)/2, \text{ where} \\ &\mathbf{S}_0 = \mathbf{I}, \mathbf{S}_1 = \mathbf{I} \cdot \mathbf{d} \cdot \cos(2\boldsymbol{\theta}), \text{and } \mathbf{S}_2 = \mathbf{I} \cdot \mathbf{d} \cdot \sin(2\boldsymbol{\theta}) \end{aligned} \tag{5}$$

are called the Stokes parameters (Können 1985). From Eq. (5) we can see that $\mathbf{S}_{0,1,2}$ can be directly solved if we have an unpolarized image and at least two polarized images with different polarizer angles (or if we have at least three polarized images with different polarizer angles). Then, the DoP $\mathbf{d}$ could be acquired from $\mathbf{S}_{0,1,2}$ using:

$$\mathbf{d} = \sqrt{\mathbf{S}_1^2 + \mathbf{S}_2^2}/\mathbf{S}_0 = \sqrt{\mathbf{S}_1^2 + \mathbf{S}_2^2}/\mathbf{I}. \tag{6}$$

According to Eq. (2) and Eq. (4), $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$ can be directly computed from $\mathbf{I}^*$ and $\mathbf{d}^*$. However, despite that we have the priors $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$ for guidance, reconstructing $\mathbf{I}$ directly from its degenerated counterpart $\mathbf{I}^*$ is still challenging due to the large appearance difference between them in the shadow region (see Fig. 2 (a) and (c)). This is because in the shadow region, the pixel values of $\mathbf{I}^*$ are usually much smaller compared with $\mathbf{I}$ (*i.e.*, $\mathbf{k} \cdot \mathbf{I}^{\text{sun}} \ll \mathbf{I}^{\text{sun}}$ in most pixels, aligning with the fact that the umbra region often occupies a large part). Prominently, we notice the total intensity modulated by the DoP, denoted as $\mathbf{m}$, which can be written as

$$\mathbf{m} = \mathbf{d} \cdot \mathbf{I} = \sqrt{\mathbf{S}_1^2 + \mathbf{S}_2^2}, \tag{7}$$

have some attractive properties: It not only contains distinctive texture and structure information (*e.g.*, the object contours and edges are salient) due to the differential nature of $\mathbf{S}_{1,2}$ (Zhou et al. 2023b), but also has better resistance to shadow (*i.e.*, $\mathbf{m}$ and $\mathbf{m}^*$ often have closer appearance to each other compared with $\mathbf{I}$ and $\mathbf{I}^*$) since in the shadow region $\mathbf{d}^*$ is often larger while $\mathbf{I}^*$ is often smaller, resulting in smaller numerical gap. To verify it, we perform simulation on our synthetic dataset to quantitatively compute their mean absolute errors (MAE) respectively. We find that the MAE between $\mathbf{m}$ and $\mathbf{m}^*$ is around 15 times smaller than $\mathbf{I}$ and $\mathbf{I}^*$. A visual example can be found in Fig. 2 (f). Therefore, we design our pipeline to be two-stage where $\mathbf{m}$ serves as the intermediate bridge. Specifically, our pipeline reformulates the shadow removal problem into *two consecutive guided information reconstruction problems*: (I) restoring $\mathbf{m}$ from $\mathbf{m}^*$ under the guidance of $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$, focusing on the texture and structure information; (II) recovering $\mathbf{I}$ from $\mathbf{I}^*$ under the guidance of $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$ along with the restored $\mathbf{m}$, concentrating on the brightness and color information.

In this way, the restoration of the brightness and color information is explicitly decoupled from the preservation of the texture and structure information, which relieves the ill-posedness of the problem.

Network module designs

Overall architecture. We design a network tailored to our pipeline, as shown in Fig. 3. Given three degenerated polarized images $\mathbf{I}^*_{\alpha_{1,2,3}}$, instead of directly feeding them into the network, we pre-compute the degenerated total intensity $\mathbf{I}^*$ and its modulated version $\mathbf{m}^*$, along with the priors in both the intensity and polarization domains ($\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$ respectively). In the first stage (f_1), we first adopt a feature extraction (FE) block and a dual domain prior fusion (DDPF) module to extract features and obtain cross-domain priors from $\mathbf{m}^*$ along with $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$ respectively, then concatenate the features and process them using a backbone network (BN) to learn the residual between $\mathbf{m}^*$ and $\mathbf{m}$. This stage can be written as

$$\begin{aligned} \mathbf{m} &= f_1(\mathbf{m}^*, \mathbf{P}_{\text{int}}, \mathbf{P}_{\text{pol}}) \\ &= \text{BN}(\text{CAT}(\text{FE}(\mathbf{m}^*), \text{DDPF}(\mathbf{P}_{\text{int}}, \mathbf{P}_{\text{pol}}))) + \mathbf{m}^*, \end{aligned} \tag{8}$$

where CAT is channel concatenation. In the second stage (f_2), we first adopt another FE block and another DDPF module to process $\mathbf{I}^*$ along with $\mathbf{P}_{\text{int}}$ and $\mathbf{P}_{\text{pol}}$ respectively,

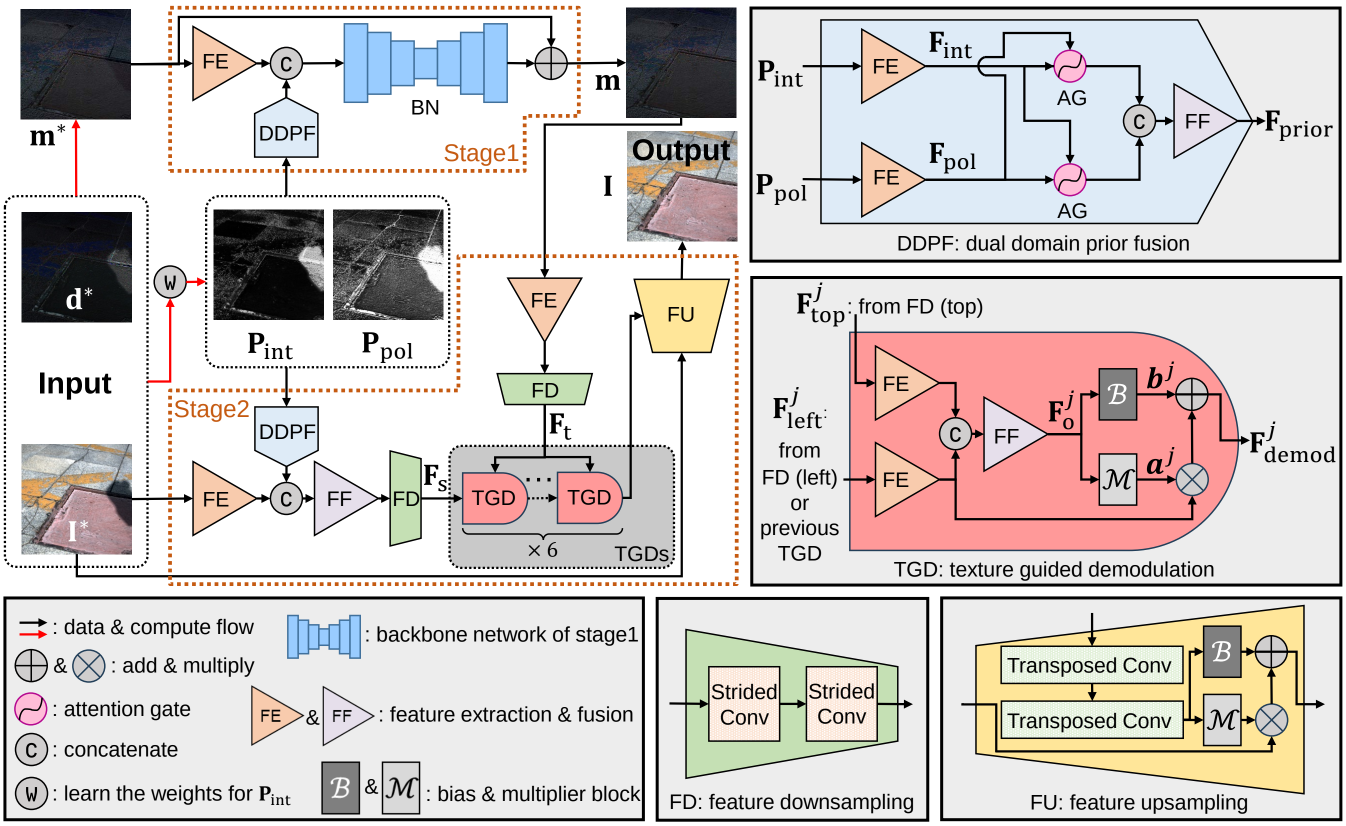

Figure 3: Architecture of our network tailored to our two-stage shadow removal pipeline.

and use a feature fusion (FF) block and a feature downsampling (FD) block to fuse and downsample the concatenated features into the semantic features $\mathbf{F}_\mathrm{s}$, then cascade 6 texture guided demodulation (TGD) modules to process $\mathbf{F}_\mathrm{s}$ under the guidance of the texture features $\mathbf{F}_\mathrm{t}$ extracted by an FE block and an FD block from the restored $\mathbf{m}$, and use a feature upsampling (FU) block to recover $\mathbf{I}$. This stage can be written as

$$\mathbf{I} = f_2(\mathbf{I}^*, \mathbf{P}_\mathrm{int}, \mathbf{P}_\mathrm{pol}, \mathbf{m}) = \mathrm{FU}(\mathrm{TGDs}(\mathbf{F}_\mathrm{s}, \mathbf{F}_\mathrm{t})), \tag{9}$$

where TGDs denotes the cascaded 6 TGD modules, and

$$\begin{cases} \mathbf{F}_\mathrm{s} = \mathrm{FD}(\mathrm{FF}(\mathrm{CAT}(\mathrm{FE}(\mathbf{I}^*), \mathrm{DDPF}(\mathbf{P}_\mathrm{int}, \mathbf{P}_\mathrm{pol})))) \\ \mathbf{F}_\mathrm{t} = \mathrm{FD}(\mathrm{FE}(\mathbf{m})) \end{cases} . \tag{10}$$

Information about layer and training details can be found in the supplementary material.

Dual domain prior fusion. Despite that both $\mathbf{P}_\mathrm{int}$ and $\mathbf{P}_\mathrm{pol}$ can independently indicate the shadow confidence, we should not simply concatenate them and extract their features jointly. This is because they belong to different domains so that the same value in $\mathbf{P}_\mathrm{int}$ and $\mathbf{P}_\mathrm{pol}$ would usually correspond to different degrees of shadow degeneration. Besides, the outliers (*i.e.*, the pixels with erroneous shadow confidence scores) would always exist in both $\mathbf{P}_\mathrm{int}$ and $\mathbf{P}_\mathrm{pol}$, making the features hard to be extracted in a joint manner. Fortunately, we have observed that $\mathbf{P}_\mathrm{int}$ and $\mathbf{P}_\mathrm{pol}$ have complementary characteristics: $\mathbf{P}_\mathrm{int}$ usually have smaller values while $\mathbf{P}_\mathrm{pol}$ usually have larger values, and the outliers in $\mathbf{P}_\mathrm{int}$ tend to be in the shadow region while the outliers in $\mathbf{P}_\mathrm{pol}$ tend to be in the shadow-free region (see Fig. 2 (d)). To this end, we propose a dual domain prior fusion (DDPF) module to fuse the priors in both the intensity and polarization domains. As shown in Fig. 3 (top right), we first use two FE blocks to extract their domain-specific features $\mathbf{F}_\mathrm{int}$ and $\mathbf{F}_\mathrm{pol}$ respectively, and then adopt a mutual attention mechanism to handle each other's outliers by fully exploiting their complementary characteristics to bridge the domain gap. This mechanism adopts two attention gates (AG) (Oktay et al. 2018) to reweight one of $\mathbf{F}_\mathrm{int}$ and $\mathbf{F}_\mathrm{pol}$ with the other one serving as the gating signal for emphasizing the domain correlations, and then use an FF block to further refine the concatenated features for obtaining cross-domain priors $\mathbf{F}_\mathrm{prior}$. The working flow of the DDPF module can be described as

$$\begin{aligned} \mathbf{F}_\mathrm{prior} &= \mathrm{DDPF}(\mathbf{P}_\mathrm{int}, \mathbf{P}_\mathrm{pol}) \\ &= \mathrm{FF}(\mathrm{CAT}(\mathrm{AT}(\mathbf{F}_\mathrm{int}, \mathbf{F}_\mathrm{pol}), \mathrm{AT}(\mathbf{F}_\mathrm{pol}, \mathbf{F}_\mathrm{int}))), \\ \text{where } \mathbf{F}_\mathrm{int} &= \mathrm{FE}(\mathbf{P}_\mathrm{int}),\ \ \mathbf{F}_\mathrm{pol} = \mathrm{FE}(\mathbf{P}_\mathrm{pol}), \end{aligned} \tag{11}$$

and $\mathrm{AT}(\mathbf{v}_\mathrm{input}, \mathbf{v}_\mathrm{gating})$ denotes the attention gate that aims to reweight $\mathbf{v}_\mathrm{input}$ with $\mathbf{v}_\mathrm{gating}$ serving as the gating signal. With the high-fidelity cross-domain priors $\mathbf{F}_\mathrm{prior}$ available, both two stages can benefit from the useful shadow clues encoded in them, improving the generalization ability to unseen shadow patterns.

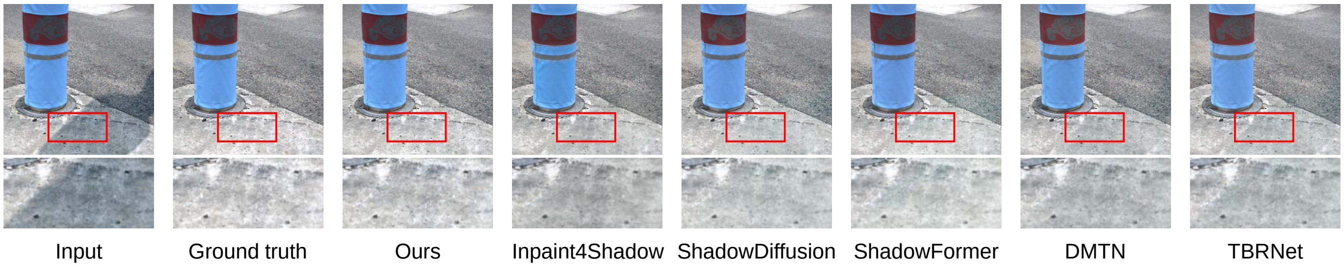

Figure 4: Examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on synthetic data. We zoom-in the red box regions below each image. *More examples can be found in the supplementary material.*

Texture guided demodulation. Since the second stage aims to reconstruct $\mathbf{I}$ with the help of $\mathbf{m}$, which is similar to dealing with a signal demodulation problem (see Eq. (7) for the relationship between $\mathbf{m}$ and $\mathbf{I}$), an effective approach should be proposed to ensure that $\mathbf{m}$ could provide texture guidance uniformly across the entire spatial domain. Besides, the modality misalignment should also be taken into consideration due to the sparsity difference between them. Therefore, we propose a texture guided demodulation (TGD) module to acquire high-quality guidance from $\mathbf{m}$ and overcome the modality misalignment, by explicitly performing demodulation-like operations in the latent space. And we propose to cascade 6 TGD modules to simulate the procedure of multiple rounds of iterative demodulation. Concretely, for the j-th TGD module, as shown in Fig. 3 (middle right), denoting the guiding signal as $\mathbf{F}^j_{\text{top}}$ (which comes from the top FD block, *i.e.*, it is always set to be the texture features $\mathbf{F}_{\text{t}}$) and the input signal as $\mathbf{F}^j_{\text{left}}$ (which comes from the left FD block if $j = 1$ else the $(j-1)$-th TGD module), we first use two FE blocks to extract their features respectively and adopt an FF block to fuse the concatenated features into a demodulation operator $\mathbf{F}^j_{\text{o}}$, then apply a multiplier block $\mathcal{M}$ along with a bias block $\mathcal{B}$ to $\mathbf{F}^j_{\text{o}}$ to estimate a multiplier $\mathbf{a}^j$ and a bias $\mathbf{b}^j$ respectively, and finally use $\mathbf{a}^j$ and $\mathbf{b}^j$ to demodulate $\mathbf{F}^j_{\text{left}}$. Denoting the output of the j-th TGD module as $\mathbf{F}^j_{\text{demod}}$, its working flow can be described as

$$\begin{aligned}\mathbf{F}^j_{\text{demod}} &= \text{TGD}(\mathbf{F}^j_{\text{left}}, \mathbf{F}^j_{\text{top}}) = \mathbf{a}^j \cdot \mathbf{F}^j_{\text{left}} + \mathbf{b}^j \\ \text{where } \mathbf{a}^j &= \mathcal{M}(\mathbf{F}^j_{\text{o}}),\ \mathbf{b}^j = \mathcal{B}(\mathbf{F}^j_{\text{o}}), \\ \text{and } \mathbf{F}^j_{\text{o}} &= \text{FF}(\text{CAT}(\text{FE}(\mathbf{F}^j_{\text{left}}), \text{FE}(\mathbf{F}^j_{\text{top}}))).\end{aligned} \tag{12}$$

Thanks to the demodulation-like operations, our network is not only adept at adjusting the brightness and color, but also capable of preserving the fine-grained texture details inside the shadow region with fewer artifacts.

Loss function. The total loss function of our network L consists of two terms: modulation loss L_{mod} and image loss L_{img}, which is defined as

$$L(\mathbf{m}, \mathbf{m}_{\text{gt}}, \mathbf{I}, \mathbf{I}_{\text{gt}}) = L_{\text{mod}}(\mathbf{m}, \mathbf{m}_{\text{gt}}) + \beta L_{\text{img}}(\mathbf{I}, \mathbf{I}_{\text{gt}}), \tag{13}$$

where β is empirically set to be 0.1, the subscript gt denotes the ground truth, and both L_{mod} and L_{img} are defined as the following basic loss function (here we use $\mathbf{v}$ and $\mathbf{v}_{\text{gt}}$ to denote the input and ground truth variables respectively):

$$\begin{aligned}L_{basic}(\mathbf{v}, \mathbf{v}_{\text{gt}}) &= \beta_1 L_1(\mathbf{v}, \mathbf{v}_{\text{gt}}) + \beta_2 L_2(\mathbf{v}, \mathbf{v}_{\text{gt}}) \\ &+ \beta_3 L_{\text{perc}}(\mathbf{v}, \mathbf{v}_{\text{gt}}) + \beta_4 L_{\text{grad}}(\mathbf{v}, \mathbf{v}_{\text{gt}}),\end{aligned} \tag{14}$$

where L_1, L_2, L_{perc}, and L_{grad} denote the ℓ_1, ℓ_2, perceptual, and gradient loss (ℓ_2 loss in the gradient domain) respectively, and $\beta_{1,2,3,4}$ are empirically set to be 10, 100, 0.1, and 10 respectively. The perceptual loss L_{perc} is defined as the ℓ_2 loss computed using the feature maps of $VGG_{3,3}$ convolution layer of VGG-19 network (Simonyan and Zisserman 2014) pretrained on ImageNet (Russakovsky et al. 2015).

Experiments

Evaluation on synthetic and real data

We compare our results to five latest learning-based shadow removal methods, including four methods that require an externally obtained shadow mask as a necessary part of the input data (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), and DMTN (Liu et al. 2023a)), along with a method that does not require the shadow mask (TBRNet (Liu et al. 2023b)). To show that polarization could provide better guidance than shadow masks, in our experiments on synthetic data, we offer the ground truth masks (which are crafted from the ground truth $\mathbf{k}$ using binarization) to the compared methods. In such a setting, the compared methods can reach their theoretical performance upper bounds which cannot be reached practically due to the imperfection of the externally obtained mask. All compared methods are retrained on our synthetic dataset for a fair comparison. *Information about our synthetic dataset can be found in the supplementary material.*

Visual quality comparisons on synthetic data are shown in Fig. 4. Our result resembles the ground truth more closely, while the compared methods fail to restore the brightness and color information; besides, the compared methods tend to yield blurry texture details in the shadow region, destroying the image structures. This is because our method can make full use of the polarization property of light to reduce the ill-posedness, while the compared methods cannot. We also adopt PSNR, SSIM, and RMSE (the root mean square error in the LAB color space) to evaluate the results on synthetic data quantitatively, following the previous works (Li

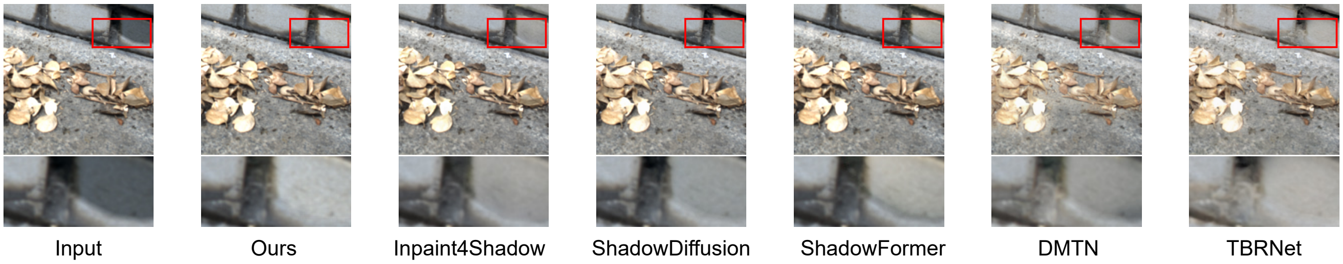

Figure 5: Examples of shadow removal results on real data. See the caption of Fig. 4 for explanation. *More examples can be found in the supplementary material.*

Table 1: The quantitative results of shadow removal using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on synthetic data.

	Shadow region			Shadow-free region			All pixels		
	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓
Input image	18.81	0.917	41.57	31.43	0.986	**0.973**	18.49	0.913	8.423
Inpaint4Shadow (Li et al. 2023)	31.02	0.978	11.27	34.56	0.989	2.319	30.49	0.957	4.105
ShadowDiffusion (Guo et al. 2023b)	29.93	0.967	11.54	33.18	0.986	3.082	29.21	0.951	4.326
ShadowFormer (Guo et al. 2023a)	32.92	0.985	10.63	35.21	0.990	2.016	31.16	0.973	3.237
DMTN (Liu et al. 2023a)	30.69	0.974	11.38	35.69	0.992	1.961	30.78	0.971	3.451
TBRNet (Liu et al. 2023b)	29.74	0.961	12.06	32.89	0.984	3.104	29.13	0.950	4.414
Ours	**34.68**	**0.990**	**9.648**	**36.44**	**0.993**	1.938	**32.24**	**0.981**	**2.858**

Table 2: Quantitative evaluation results of ablation study.

	PSNR↑	SSIM↑	RMSE↓
W/o polarization	27.83	0.934	5.104
W/o priors	28.75	0.952	5.061
Mask instead of $\mathbf{P}_{\text{pol}}$	30.06	0.952	4.293
Single stage	29.06	0.964	4.933
W/o DDPF	31.48	0.977	3.136
W/o TGD	30.92	0.970	3.812
Our complete model	**32.24**	**0.981**	**2.858**

et al. 2023; Guo et al. 2023b,a; Liu et al. 2023a,b). Results are shown in Tab. 1. Our method consistently outperforms the compared ones on all metrics. *Computational complexity analysis can be found in the supplementary material.*

To demonstrate the generalization ability, we capture several real images containing shadows. Qualitative results are shown in Fig. 5. We can see that our method can output clearer images with less false color. For example, the color of the wall tiles is correctly recovered by our method, while other methods tend to produce yellowish results. This is because our Po-ShaRe take advantage of the modality information of polarization, while the compared methods cannot.

Ablation Study

We conduct a series of ablation studies to verify the validity of each design choice. First, we show the significance of using polarization information by modifying the network to a single-image shadow removal network (W/o polarization). Besides, we show the effectiveness of the priors by not explicitly extracting them (W/o priors) and using the shadow mask to substitute $\mathbf{P}_{\text{pol}}$ (Mask instead of $\mathbf{P}_{\text{pol}}$). Then, we verify the necessity of our two-stage pipeline by comparing with a model that reconstruct $\mathbf{I}$ in a single stage without estimating $\mathbf{m}$ (Single stage). In addition, we show the importance of our DDPF and TGD modules by substituting them with convolution layers (W/o DDPF and W/o TGD). As shown in Tab. 2, our complete model achieves the first performance.

Conclusion

We propose Pol-ShaRe, the first polarization-guided image shadow removal solution. By exploiting the dual domain priors from the shadow image formation model, it can remove shadow in a mask-free manner. Specifically, it consists of a two-stage pipeline that decouples the restoration of the brightness and color information from the preservation of the texture and structure information, and a network tailored to the pipeline that integrates physics-oriented modules.

Limitations. Our Pol-ShaRe may face challenges when the prior in the polarization domain introduces incorrect guidance. For instance, in indoor scenes illuminated by artificial light sources or outdoor scenes under adverse weather conditions (*e.g.*, rain, haze, or snow), the captured images often do not conform to the assumptions we made about the priors. Furthermore, since obtaining multiple polarized images requires multiple shots, performance may degrade due to misalignment among the polarized images, limiting its applicability in dynamic scenes.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grant No. 62136001, 62088102, 62276007 and JST-Mirai Program Grant Number JPMJM123G1.

References

Ba, Y.; Gilbert, A.; Wang, F.; Yang, J.; Chen, R.; Wang, Y.; Yan, L.; Shi, B.; and Kadambi, A. 2020. Deep shape from polarization. In *Proc. of European Conference on Computer Vision*, 554–571.

Blin, R.; Ainouz, S.; Canu, S.; and Meriaudeau, F. 2019. Adapted learning for polarization-based car detection. In *Proc. of International Conference on Quality Control by Artificial Vision*, volume 11172, 312–318.

Chen, Z.; Long, C.; Zhang, L.; and Xiao, C. 2021. CANet: A context-aware network for shadow removal. In *Proc. of International Conference on Computer Vision*, 4743–4752.

Cun, X.; Pun, C.-M.; and Shi, C. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In *Proc. of the AAAI Conference on Artificial Intelligence*, 10680–10687.

Dave, A.; Zhao, Y.; and Veeraraghavan, A. 2022. PANDORA: Polarization-aided neural decomposition of radiance. In *Proc. of European Conference on Computer Vision*, 538–556.

Ding, B.; Long, C.; Zhang, L.; and Xiao, C. 2019. ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proc. of International Conference on Computer Vision*, 10213–10222.

Fu, L.; Zhou, C.; Guo, Q.; Juefei-Xu, F.; Yu, H.; Feng, W.; Liu, Y.; and Wang, S. 2021. Auto-exposure fusion for single-image shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 10571–10580.

Guo, L.; Huang, S.; Liu, D.; Cheng, H.; and Wen, B. 2023a. ShadowFormer: Global context helps shadow removal. In *Proc. of the AAAI Conference on Artificial Intelligence*, 710–718.

Guo, L.; Wang, C.; Yang, W.; Huang, S.; Wang, Y.; Pfister, H.; and Wen, B. 2023b. ShadowDiffusion: When degradation prior meets diffusion model for shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 14049–14058.

Guo, L.; Wang, C.; Yang, W.; Wang, Y.; and Wen, B. 2023c. Boundary-aware divide and conquer: A diffusion-based solution for unsupervised shadow removal. In *Proc. of International Conference on Computer Vision*, 13045–13054.

Guo, R.; Dai, Q.; and Hoiem, D. 2012. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12): 2956–2967.

He, S.; Peng, B.; Dong, J.; and Du, Y. 2021. Mask-ShadowNet: Toward shadow removal via masked adaptive instance normalization. *IEEE Signal Processing Letters*, 28: 957–961.

Hecht, E.; et al. 2002. *Optics*, volume 5. Addison Wesley San Francisco.

Holstein, B. R. 1999. Blue skies and effective interactions. *American Journal of Physics*, 67(5): 422–427.

Hu, X.; Fu, C.-W.; Zhu, L.; Qin, J.; and Heng, P.-A. 2019a. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11): 2795–2808.

Hu, X.; Jiang, Y.; Fu, C.-W.; and Heng, P.-A. 2019b. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In *Proc. of International Conference on Computer Vision*, 2472–2481.

Huang, X.; Hua, G.; Tumblin, J.; and Williams, L. 2011. What characterizes a shadow boundary under the sun and sky? In *Proc. of International Conference on Computer Vision*, 898–905.

Inoue, N.; and Yamasaki, T. 2020. Learning from synthetic shadows for shadow detection and removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(11): 4187–4197.

Jin, Y.; Ye, W.; Yang, W.; Yuan, Y.; and Tan, R. T. 2024. DeS3: Adaptive attention-driven self and soft shadow removal using ViT similarity. In *Proc. of the AAAI Conference on Artificial Intelligence*, 2634–2642.

Kadambi, A.; Taamazyan, V.; Shi, B.; and Raskar, R. 2017. Depth sensing using geometrically constrained polarization normals. *International Journal of Computer Vision*, 125: 34–51.

Kalra, A.; Taamazyan, V.; Rao, S. K.; Venkataraman, K.; Raskar, R.; and Kadambi, A. 2020. Deep polarization cues for transparent object segmentation. In *Proc. of Computer Vision and Pattern Recognition*, 8602–8611.

Khan, S. H.; Bennamoun, M.; Sohel, F.; and Togneri, R. 2015. Automatic shadow detection and removal from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3): 431–446.

Können, G. 1985. *Polarized light in nature*. CUP Archive.

Le, H.; and Samaras, D. 2019. Shadow removal via shadow image decomposition. In *Proc. of International Conference on Computer Vision*, 8578–8587.

Le, H.; and Samaras, D. 2021. Physics-based shadow image decomposition for shadow removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 9088–9101.

Lei, C.; Huang, X.; Zhang, M.; Yan, Q.; Sun, W.; and Chen, Q. 2020. Polarized reflection removal with perfect alignment in the wild. In *Proc. of Computer Vision and Pattern Recognition*, 1750–1758.

Li, N.; Zhao, Y.; Pan, Q.; Kong, S. G.; and Chan, J. C.-W. 2020. Full-time monocular road detection using zero-distribution prior of angle of polarization. In *Proc. of European Conference on Computer Vision*, 457–473.

Li, X.; Guo, Q.; Abdelfattah, R.; Lin, D.; Feng, W.; Tsang, I.; and Wang, S. 2023. Leveraging inpainting for single-image shadow removal. In *Proc. of International Conference on Computer Vision*, 13055–13064.

Liang, Y.; Wakaki, R.; Nobuhara, S.; and Nishino, K. 2022. Multimodal material segmentation. In *Proc. of Computer Vision and Pattern Recognition*, 19800–19808.

Lin, S.-S.; Yemelyanov, K. M.; Pugh, E. N.; and Engheta, N. 2006. Separation and contrast enhancement of overlapping cast shadow components using polarization. *Optics Express*, 14(16): 7099–7108.

Liu, J.; Wang, Q.; Fan, H.; Li, W.; Qu, L.; and Tang, Y. 2023a. A decoupled multi-task network for shadow removal. *IEEE Transactions on Multimedia*.

Liu, J.; Wang, Q.; Fan, H.; Tian, J.; and Tang, Y. 2023b. A shadow imaging bilinear model and three-branch residual network for shadow removal. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, Y.; Guo, Q.; Fu, L.; Ke, Z.; Xu, K.; Feng, W.; Tsang, I. W.; and Lau, R. W. 2023c. Structure-informed shadow removal networks. *IEEE Transactions on Image Processing*.

Liu, Z.; Yin, H.; Mi, Y.; Pu, M.; and Wang, S. 2021a. Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Transactions on Image Processing*, 30: 1853–1865.

Liu, Z.; Yin, H.; Wu, X.; Wu, Z.; Mi, Y.; and Wang, S. 2021b. From shadow generation to shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 4927–4936.

Lyu, Y.; Cui, Z.; Li, S.; Pollefeys, M.; and Shi, B. 2022. Physics-guided reflection separation from a pair of unpolarized and polarized images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2151–2165.

Lyu, Y.; Zhao, L.; Li, S.; and Shi, B. 2023. Shape from polarization with distant lighting estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 13991–14004.

Mei, H.; Dong, B.; Dong, W.; Yang, J.; Baek, S.-H.; Heide, F.; Peers, P.; Wei, X.; and Yang, X. 2022. Glass segmentation using intensity and spectral polarization cues. In *Proc. of Computer Vision and Pattern Recognition*, 12622–12631.

Niu, K.; Liu, Y.; Wu, E.; and Xing, G. 2022. A boundary-aware network for shadow removal. *IEEE Transactions on Multimedia*.

Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N. Y.; Kainz, B.; Glocker, B.; and Rueckert, D. 2018. Attention U-Net: Learning where to look for the pancreas.

Ono, T.; Kondo, Y.; Sun, L.; Kurita, T.; and Moriuchi, Y. 2022. Degree-of-linear-polarization-based color constancy. In *Proc. of Computer Vision and Pattern Recognition*, 19740–19749.

Pust, N. J.; and Shaw, J. A. 2012. Wavelength dependence of the degree of polarization in cloud-free skies: simulations of real environments. *Optics Express*, 20(14): 15559–15568.

Qu, L.; Tian, J.; He, S.; Tang, Y.; and Lau, R. W. 2017. DeshadowNet: A multi-context embedding deep network for shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 4067–4075.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; Berg, A. C.; and Fei-Fei, L. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3): 211–252.

Sekera, Z. 1957. Polarization of skylight. In *Geophysik II/Geophysics II*, 288–328. Springer.

Sen, M.; Chermala, S. P.; Nagori, N. N.; Peddigari, V.; Mathur, P.; Prasad, B.; and Jeong, M. 2023. SHARDS: Efficient shadow removal using dual stage network for high-resolution images. In *Proc. of Winter Conference on Applications of Computer Vision*, 1809–1817.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tian, C.; Pan, W.; Wang, Z.; Mao, M.; Zhang, G.; Bao, H.; Tan, P.; and Cui, Z. 2023. DPS-Net: Deep polarimetric stereo depth estimation. In *Proc. of International Conference on Computer Vision*, 3569–3579.

Tian, J.; and Tang, Y. 2011. Linearity of each channel pixel values from a surface in and out of shadows and its applications. In *Proc. of Computer Vision and Pattern Recognition*, 985–992.

Wan, J.; Yin, H.; Wu, Z.; Wu, X.; Liu, Y.; and Wang, S. 2022. Style-guided shadow removal. In *Proc. of European Conference on Computer Vision*, 361–378.

Yang, Q.; Tan, K.-H.; and Ahuja, N. 2012. Shadow removal using bilateral filtering. *IEEE Transactions on Image Processing*, 21(10): 4361–4368.

Yücel, M. K.; Dimaridou, V.; Manganelli, B.; Ozay, M.; Drosou, A.; and Saa-Garriga, A. 2023. LRA&LDRA: Rethinking residual predictions for efficient shadow detection and removal. In *Proc. of Winter Conference on Applications of Computer Vision*, 4925–4935.

Zhang, L.; Long, C.; Zhang, X.; and Xiao, C. 2020. RIS-GAN: Explore residual and illumination with generative adversarial networks for shadow removal. In *Proc. of the AAAI Conference on Artificial Intelligence*, 12829–12836.

Zhou, C.; Han, Y.; Teng, M.; Han, J.; Li, S.; Xu, C.; and Shi, B. 2023a. Polarization guided HDR reconstruction via pixel-wise depolarization. *IEEE Transactions on Image Processing*, 32: 1774–1787.

Zhou, C.; Teng, M.; Han, Y.; Xu, C.; and Shi, B. 2021. Learning to dehaze with polarization. In *Proc. of Advances in Neural Information Processing Systems*.

Zhou, C.; Teng, M.; Lyu, Y.; Li, S.; Xu, C.; and Shi, B. 2023b. Polarization-aware low-light image enhancement. In *Proc. of the AAAI Conference on Artificial Intelligence*.

Zhu, Y.; Huang, J.; Fu, X.; Zhao, F.; Sun, Q.; and Zha, Z.-J. 2022a. Bijective mapping network for shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 5627–5636.

Zhu, Y.; Xiao, Z.; Fang, Y.; Fu, X.; Xiong, Z.; and Zha, Z.-J. 2022b. Efficient model-driven network for shadow removal. In *Proc. of the AAAI Conference on Artificial Intelligence*, 3635–3643.

Supplementary Material: Polarization Guided Mask-Free Shadow Removal

Chu Zhou[1 †], **Chao Xu**[2], **Boxin Shi**[3,4 #]

[1]National Institute of Informatics, Japan
[2]National Key Laboratory of General Artificial Intelligence, School of IST, Peking University, China
[3]State Key Laboratory for Multimedia Information Processing, School of CS, Peking University, China
[4]National Engineering Research Center of Visual Technology, School of CS, Peking University, China

About the problem scope and usability

Our Pol-ShaRe aims to solve the very same problem as current image shadow removal methods (Li et al. 2023; Guo et al. 2023a; Liu et al. 2023a,b) (*i.e.*, restoring image content only in shadow regions, which is a partial degradation problem) under the guidance of polarization. As far as we know, there is no existing polarization-based method can do the same thing. The most relevant works could be the following ones: Lin et al. (2006) proposed a polarization-based method to separate the overlapping cast shadows and enhance the contrast, however, it directly computes the degree of polarization of the incoming light to the sensor and treats it as the result of contrast enhancement, which cannot recover the original pixel values and can only handle the grayscale images; Reda, Shen, and Zhao (2019) proposed a polarization-based method to enhance the images where all pixels are in the shadow region with extremely low illumination, which solves a global degradation problem more like low-light image enhancement.

Considering that current image shadow removal methods (Li et al. 2023; Guo et al. 2023a; Liu et al. 2023a,b) primarily address outdoor scenes lit by daylight under sunny weather, due to the lighting conditions of existing datasets (Qu et al. 2017; Wang, Li, and Yang 2018; Le and Samaras 2019), our Pol-ShaRe is also designed for such scenes to ensure practical usability. Regarding image capturing, our Pol-ShaRe is as convenient as current shadow removal methods, as capturing polarized images merely requires placing a polarizer in front of the lens.

About the shadow image formation model

Considering the outdoor scenes lit by daylight under sunny weather, there are mainly two light sources: direct sunlight and ambient skylight (Tian and Tang 2011). Denoting their illumination spectral power distribution (SPD) as $\mathbf{L}(\lambda)$, $\mathbf{L}^{\text{sun}}(\lambda)$, and $\mathbf{L}^{\text{sky}}(\lambda)$ respectively (where λ is the wavelength), the relationship between them can be written as

$$\mathbf{L}(\lambda) = \mathbf{L}^{\text{sun}}(\lambda) + \mathbf{L}^{\text{sky}}(\lambda). \tag{13}$$

Here, the sunlight component is often stronger (Tian, Sun, and Tang 2009), *i.e.*,

$$\mathbf{L}^{\text{sun}} > \mathbf{L}^{\text{sky}} \tag{14}$$

holds for most cases. According to the photometric model proposed by Tian, Sun, and Tang (2009), when taking photos in such scenes, the total intensity of the captured image $\mathbf{I}$ can be described as

$$\begin{aligned}\mathbf{I} &= \int_\lambda \mathbf{L}(\lambda)\cdot\mathbf{R}(\lambda)\cdot\mathbf{Q}(\lambda)\mathrm{d}\lambda\\ &= \int_\lambda (\mathbf{L}^{\text{sun}}(\lambda)+\mathbf{L}^{\text{sky}}(\lambda))\cdot\mathbf{R}(\lambda)\cdot\mathbf{Q}(\lambda)\mathrm{d}\lambda\\ &= \int_\lambda \mathbf{L}^{\text{sun}}(\lambda)\cdot\mathbf{R}(\lambda)\cdot\mathbf{Q}(\lambda)\mathrm{d}\lambda + \int_\lambda \mathbf{L}^{\text{sky}}(\lambda)\cdot\mathbf{R}(\lambda)\cdot\mathbf{Q}(\lambda)\mathrm{d}\lambda\\ &= \mathbf{I}^{\text{sun}} + \mathbf{I}^{\text{sky}},\end{aligned} \tag{15}$$

where $\mathbf{R}(\lambda)$ and $\mathbf{Q}(\lambda)$ are the reflectance and camera sensitivity function respectively, $\mathbf{I}^{\text{sun}}$ and $\mathbf{I}^{\text{sky}}$ denote the intensity components of sunlight and skylight respectively.

About the weighting function used for extracting priors

The idea of designing a weighting function $\mathcal{W}(\mathbf{v})$ to filter the pixels with relatively larger values in $\mathbf{v}$ is inspired from Ono et al. (2022). Specifically, $\mathcal{W}(\mathbf{v})$ can be written as

$$\mathcal{W}(\mathbf{v}) = \frac{1}{(1+e^{a(\mathbf{v}-b)})}, \tag{16}$$

where the hyper-parameters a and b are set to -50 and 0.08 respectively, which are the same as the ones used by Ono et al. (2022). From Eq. (16) we can see for a certain pixel in $\mathbf{v}$, a larger value of $\mathcal{W}(\mathbf{v})$ indicates that the pixel has higher confidence to be larger. And the effectiveness of the selection of the hyper-parameters is verified by Ono et al. (2022).

Layer and training details

Layer details. Both the FE block, multiplier block, and bias block are designed to be bottleneck blocks (He et al. 2016). The FF block consists of a convolution layer and a squeeze-and-excitation block (Hu, Shen, and Sun 2018). The FD

† Most of this work was done as a PhD student at Peking University.

# Corresponding author: shiboxin@pku.edu.cn

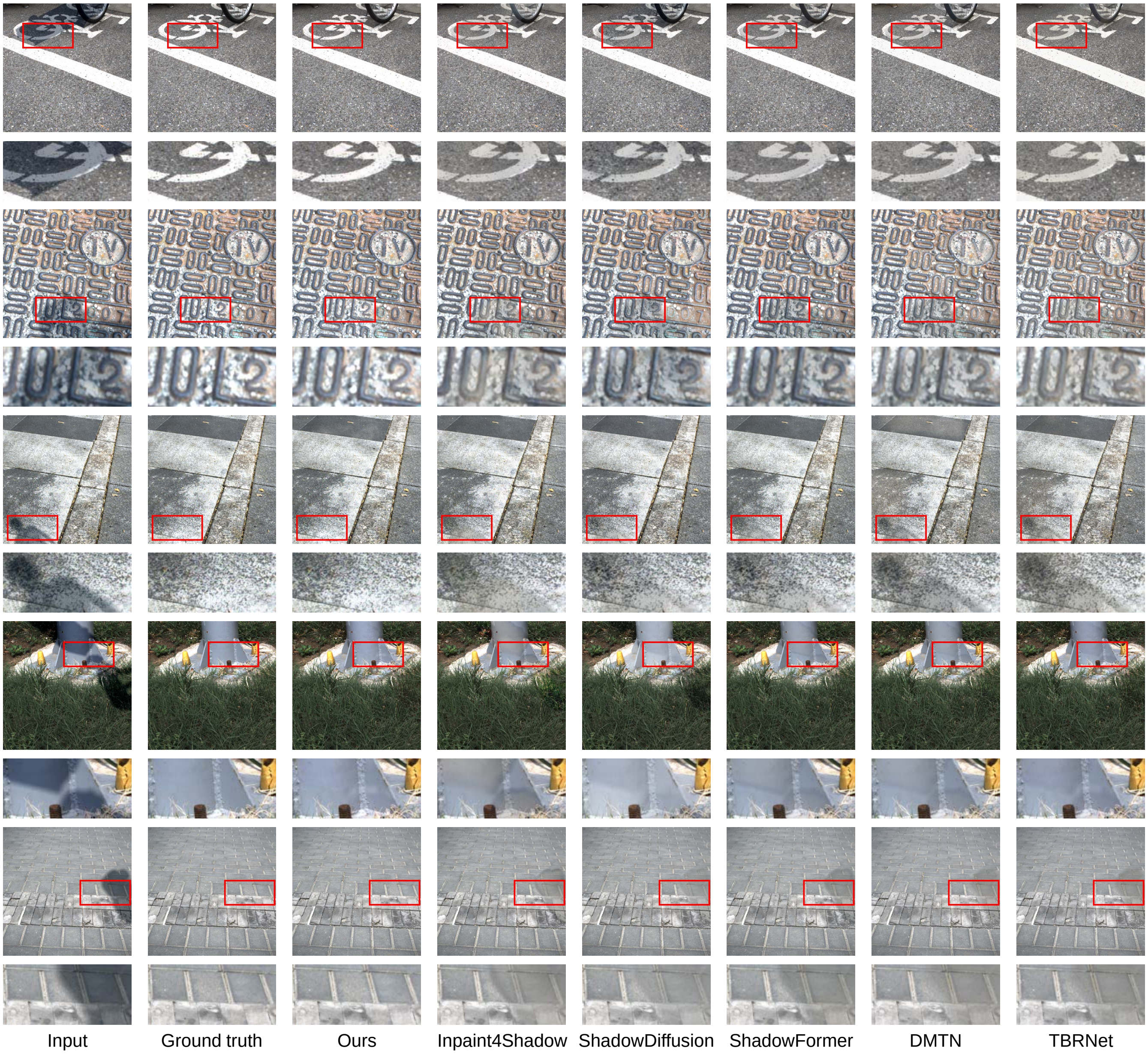

Figure 6: Additional examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on synthetic data. The close-up views of red box regions are displayed below each image.

block consists of two strided convolution layers to downsample the features. The FU block first adopts two transposed convolution layers to upsample the features outputted by the TGD module and estimates a multiplier and a bias from them using a multiplier block and a bias block respectively, and then performs demodulation-like operations on $\mathbf{I}^*$ to obtain the final output $\mathbf{I}$. As for the backbone network of the first stage, we choose the U-Net architecture (Ronneberger, Fischer, and Brox 2015) due to its excellent performance on dense prediction tasks. Instance normalization (Ulyanov, Vedaldi, and Lempitsky 2016) and `LeakyReLU` are added after each convolution layer.

Training details. We implement the network using PyTorch with 4 NVIDIA 1080Ti GPUs, and apply a two-phase training strategy: first, training two stages for 100 epochs respectively in an independent manner to ensure a stable initialization; then, finetuning the entire network in an end-to-end manner for another 100 epochs. The batch size is set to 4, and the learning rate is set to 0.01. For optimization, we use Adam optimizer (Kingma and Ba 2014) with $\beta_1 = 0.5$, $\beta_2 = 0.999$.

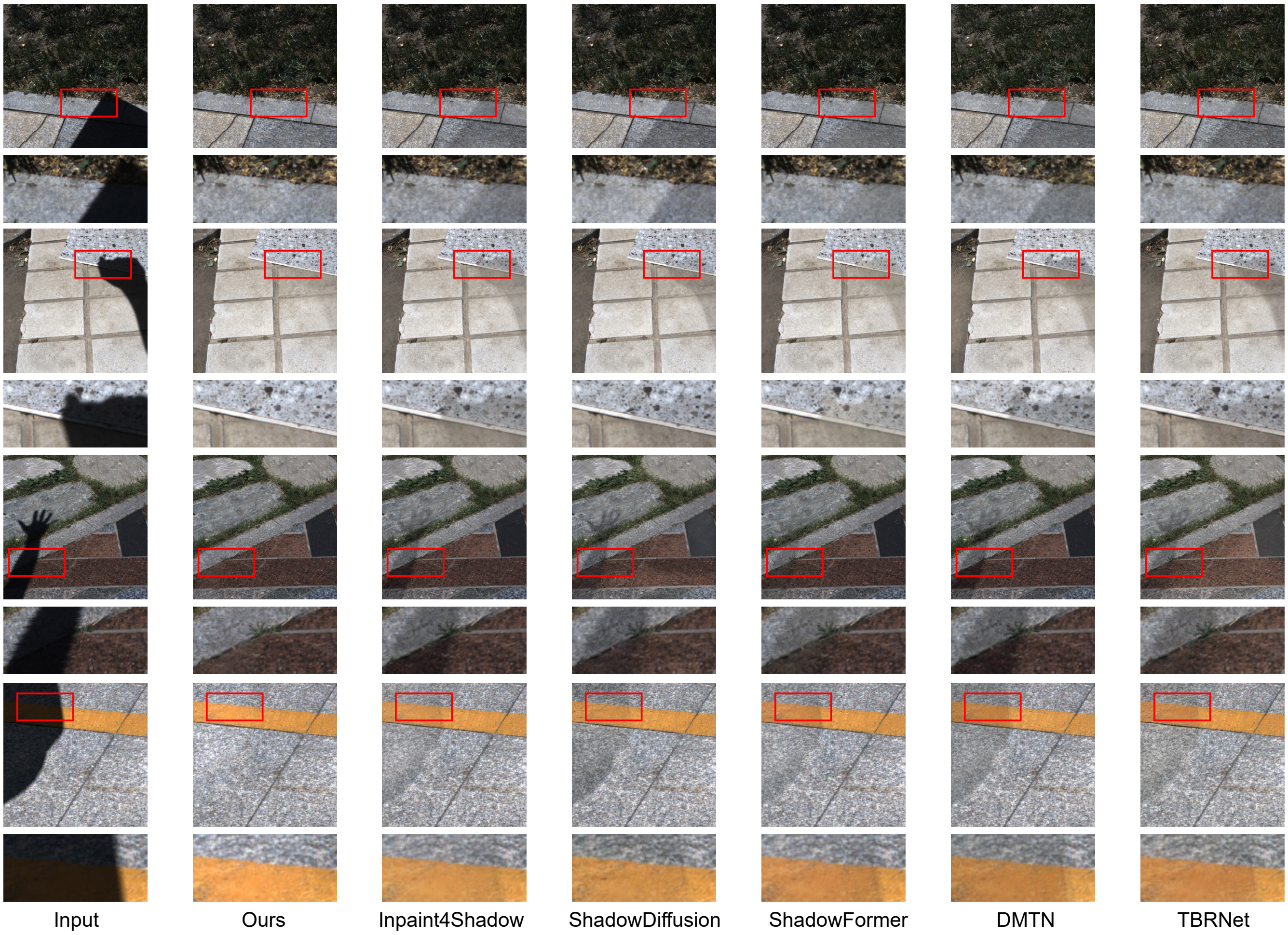

Figure 7: Additional examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on real data. The close-up views of red box regions are displayed below each image.

Table 3: Computational complexity analysis on synthetic data among our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)).

	Inpaint4Shadow (Li et al. 2023)	ShadowDiffusion (Guo et al. 2023b)	ShadowFormer (Guo et al. 2023a)	DMTN (Liu et al. 2023a)	TBRNet (Liu et al. 2023b)	Ours
Params (M)	23.9	55.5	11.3	45.6	69.9	10.4
MACs (G)	166.7	444.4	152.2	297.9	881.2	83.3

More information about the synthetic dataset

Considering the fact that there is no public dataset containing pairwise shadow and shadow-free images with polarized observations, and existing benchmark datasets (*e.g.*, SRD (Qu et al. 2017), ISTD (Wang, Li, and Yang 2018), and ISTD+ (Le and Samaras 2019)) do not contain any polarization information, we propose to generate a synthetic dataset for network training. Here, for obtaining a large number of polarized shadow-free images as the source data in a more convenient manner, we choose to use a Lucid Vision Phoenix polarization camera (RGB) instead of a linear polarizer to capture outdoor scenes lit by daylight under sunny weather, since the polarization camera can take four images with different polarizer angles ($0°$, $45°$, $90°$, and $135°$) at a single shot. Note that in practical applications, our PolShaRe does not require a polarization camera, and we only need to place a polarizer in front of the lens and rotating it for obtaining multiple polarized images.

After capturing, we can directly obtain $\mathbf{I}$, $\mathbf{d}$, and $\mathbf{m}$ using Eq. (5), Eq. (6), and Eq. (7) in the main paper as the ground truth for supervision. Then, we adopt the rendering-based simulation approach proposed by Inoue *et al.* (Inoue and Yamasaki 2020) to synthesize $\mathbf{I}^*$ as the input image from $\mathbf{I}$ with different shadow patterns by generating different $\mathbf{k}$, and generate reasonable polarization-related parameters according to the statistics of outdoor illumination (Sekera 1957; Kupinski et al. 2019) to obtain $\mathbf{d}^*$ as the input guidance. Besides, we add noise to better simulate the real situation. Specifically, we capture 100 different scenes in total, and we randomly split them into two parts that contain 90 and 10 scenes for making the training and test sets respectively. For each scene in the training (test) set, we randomly generate 90 (10) different shadow patterns so that the training (test) set contains 8100 (100) different images finally. The images are resized and cropped to 400×400.

More results on synthetic data

In this section, we provide additional examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on synthetic data, as shown in Fig. 6.

Computational complexity analysis

In this section, we evaluate the computational complexity of our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on our synthetic test dataset using a single NVIDIA 4090 GPU, as shown in Tab. 3.

More results on real data

In this section, we provide additional examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on real data, as shown in Fig. 7.

References

Guo, L.; Huang, S.; Liu, D.; Cheng, H.; and Wen, B. 2023a. ShadowFormer: Global context helps shadow removal. In *Proc. of the AAAI Conference on Artificial Intelligence*, 710–718.

Guo, L.; Wang, C.; Yang, W.; Huang, S.; Wang, Y.; Pfister, H.; and Wen, B. 2023b. ShadowDiffusion: When degradation prior meets diffusion model for shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 14049–14058.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proc. of Computer Vision and Pattern Recognition*.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proc. of Computer Vision and Pattern Recognition*, 7132–7141.

Inoue, N.; and Yamasaki, T. 2020. Learning from synthetic shadows for shadow detection and removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(11): 4187–4197.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kupinski, M. K.; Bradley, C. L.; Diner, D. J.; Xu, F.; and Chipman, R. A. 2019. Angle of linear polarization images of outdoor scenes. *Optical Engineering*, 58(8): 082419.

Le, H.; and Samaras, D. 2019. Shadow removal via shadow image decomposition. In *Proc. of International Conference on Computer Vision*, 8578–8587.

Li, X.; Guo, Q.; Abdelfattah, R.; Lin, D.; Feng, W.; Tsang, I.; and Wang, S. 2023. Leveraging inpainting for single-image shadow removal. In *Proc. of International Conference on Computer Vision*, 13055–13064.

Lin, S.-S.; Yemelyanov, K. M.; Pugh, E. N.; and Engheta, N. 2006. Separation and contrast enhancement of overlapping cast shadow components using polarization. *Optics Express*, 14(16): 7099–7108.

Liu, J.; Wang, Q.; Fan, H.; Li, W.; Qu, L.; and Tang, Y. 2023a. A decoupled multi-task network for shadow removal. *IEEE Transactions on Multimedia*.

Liu, J.; Wang, Q.; Fan, H.; Tian, J.; and Tang, Y. 2023b. A shadow imaging bilinear model and three-branch residual network for shadow removal. *IEEE Transactions on Neural Networks and Learning Systems*.

Ono, T.; Kondo, Y.; Sun, L.; Kurita, T.; and Moriuchi, Y. 2022. Degree-of-linear-polarization-based color constancy. In *Proc. of Computer Vision and Pattern Recognition*, 19740–19749.

Qu, L.; Tian, J.; He, S.; Tang, Y.; and Lau, R. W. 2017. DeshadowNet: A multi-context embedding deep network for shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 4067–4075.

Reda, M.; Shen, L.; and Zhao, Y. 2019. Image enhancement of shadow region based on polarization imaging. In *Proc. of Chinese Conference on Pattern Recognition and Computer Vision*, 736–748.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. In *Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention*, 234–241.

Sekera, Z. 1957. Polarization of skylight. In *Geophysik II/-Geophysics II*, 288–328. Springer.

Tian, J.; Sun, J.; and Tang, Y. 2009. Tricolor attenuation model for shadow detection. *IEEE Transactions on Image Processing*, 18(10): 2355–2363.

Tian, J.; and Tang, Y. 2011. Linearity of each channel pixel values from a surface in and out of shadows and its applications. In *Proc. of Computer Vision and Pattern Recognition*, 985–992.

Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2016. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*.

Wang, J.; Li, X.; and Yang, J. 2018. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proc. of Computer Vision and Pattern Recognition*, 1788–1797.