

Supplementary Material: Polarization Guided Mask-Free Shadow Removal

Chu Zhou^{1 †}, Chao Xu², Boxin Shi^{3,4 #}

¹National Institute of Informatics, Japan

²National Key Laboratory of General Artificial Intelligence, School of IST, Peking University, China

³State Key Laboratory for Multimedia Information Processing, School of CS, Peking University, China

⁴National Engineering Research Center of Visual Technology, School of CS, Peking University, China
zhou_chu@hotmail.com, xuchao@cis.pku.edu.cn, shiboxin@pku.edu.cn

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

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

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

[#] Corresponding author.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

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

$$\mathbf{L}^{\text{sun}} > \mathbf{L}^{\text{sky}} \quad (14)$$

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

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

where $\mathbf{R}(\lambda)$ and $\mathbf{Q}(\lambda)$ are the reflectance and camera sensitivity function respectively, \mathbf{I}^{sun} and \mathbf{I}^{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

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

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

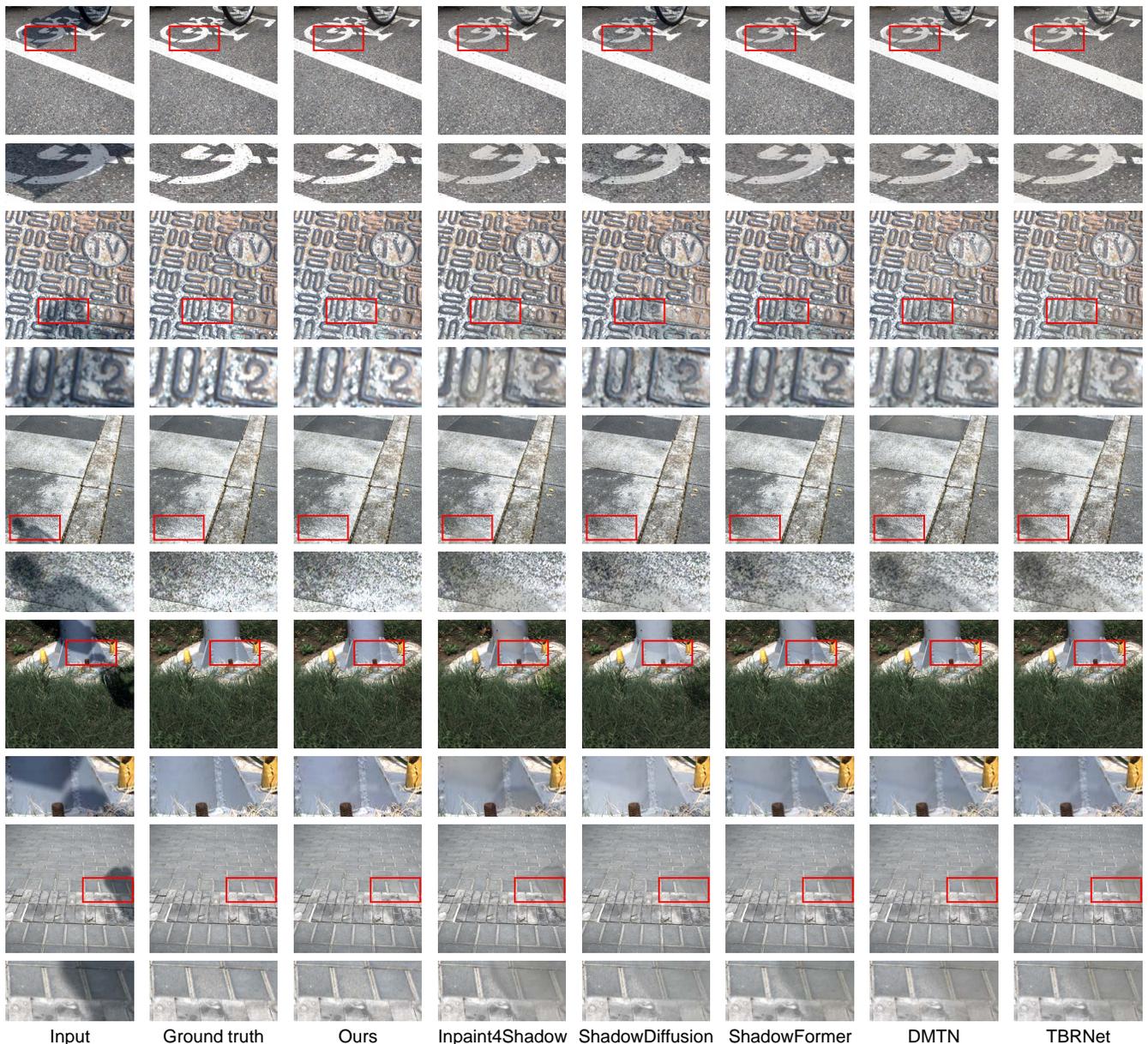


Figure 6: Additional examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on synthetic data. The close-up views of red box regions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

block consists of two strided convolution layers to down-sample 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 I^* to obtain the final output 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 $\beta_1 = 0.5$, $\beta_2 = 0.999$.

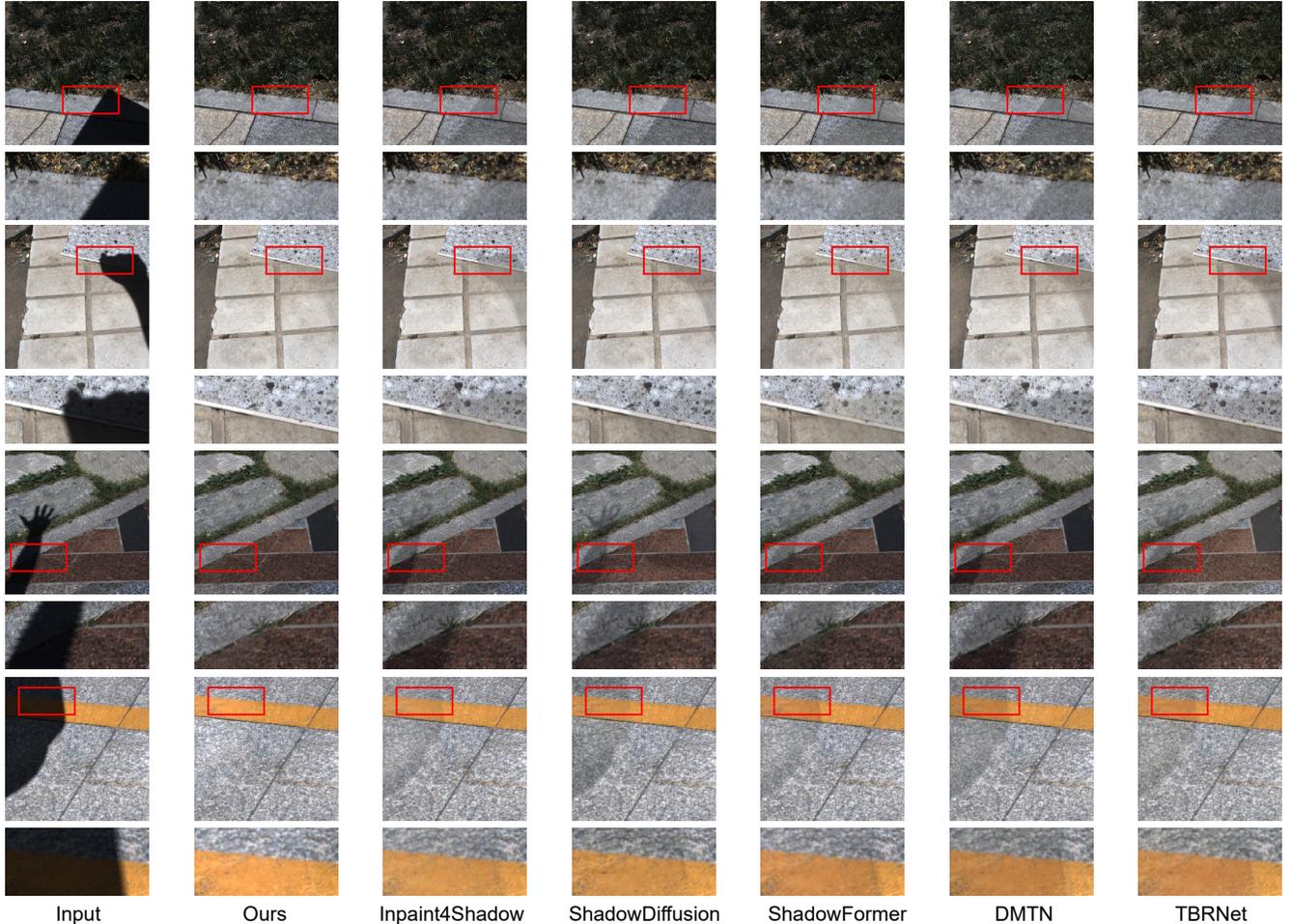


Figure 7: Additional examples of shadow removal results using our method and current ones (Inpaint4Shadow (Li et al. 2023), ShadowDiffusion (Guo et al. 2023b), ShadowFormer (Guo et al. 2023a), DMTN (Liu et al. 2023a), and TBRNet (Liu et al. 2023b)) on real data. The close-up views of red box regions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

| | 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 |

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)).

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 Pol-ShaRe 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 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. De-shadowNet: 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 III-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.