

Learning to Deblur Polarized Images

Chu Zhou^{1†} · Minggui Teng^{2,3} · Xinyu Zhou⁴ · Chao Xu⁴ · Imari Sato¹ · Boxin Shi^{2,3}

Received: date / Accepted: date

Abstract A polarization camera can capture four linear polarized images with different polarizer angles in a single shot, which is useful in polarization-based vision applications since the degree of linear polarization (DoLP) and the angle of linear polarization (AoLP) can be directly computed from the captured polarized images. However, since the on-chip micro-polarizers block part of the light so that the sensor often requires a longer exposure time, the captured polarized images are prone to motion blur caused by camera shakes, leading to noticeable degradation in the computed DoLP and AoLP. Deblurring methods for conventional images often show degraded performance when handling the polarized images since they only focus on deblurring without considering the polarization constraints. In this paper, we propose a polarized image deblurring pipeline to solve the problem in a polarization-aware manner by adopting a divide-and-conquer strategy to explicitly decompose the problem into two less ill-posed sub-problems, and design a two-stage neural network to handle the two sub-problems respectively. Experimental results show that our method achieves state-of-the-art performance on both synthetic and real-world images, and can improve the performance of polarization-based vision applications such as image dehazing and reflection removal.

Keywords Image deblurring · Polarized images · Polarization camera · Deep learning

✉ Boxin Shi
E-mail: shiboxin@pku.edu.cn

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

1 National Institute of Informatics, Japan

2 National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, China

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

4 National Key Laboratory of General AI, School of Intelligence Science and Technology, Peking University, China

1 Introduction

Polarization is an important property of light in addition to its amplitude and phase. By virtue of the physical information encoded in the polarization-relevant parameters (*e.g.*, the degree of linear polarization (DoLP) and the angle of linear polarization (AoLP) of the incoming light to the sensor), it has demonstrated unique advantages in various computer vision tasks, such as glass segmentation (Mei et al, 2022), color constancy (Ono et al, 2022), and shape reconstruction (Deschaintre et al, 2021). Besides, it also plays a crucial role in applications for autonomous driving, such as road detection (Li et al, 2020a), reflection removal (Lyu et al, 2019, 2022), dehazing (Zhou et al, 2021), and HDR reconstruction (Zhou et al, 2023a), conquering the bottleneck of conventional vision algorithms in this field.

With the development of the division of focal plane (DoFP) technology, modern polarization cameras (*e.g.*, Lucid Vision Phoenix polarization camera¹) can capture four spatially-aligned and temporally-synchronized linear polarized images with different polarizer angles in a snapshot, bringing convenience to the acquisition of the DoLP and AoLP. However, the captured polarized images are prone to motion blur caused by camera shakes, since the on-chip micro-polarizers block part of the light so that the sensor often requires a longer exposure time. Therefore, it is of practical interest to deblur polarized images for acquiring the DoLP and AoLP accurately.

Recent advances in image deblurring (Nah et al, 2017; Kupyn et al, 2019; Cho et al, 2021; Ji et al, 2022) have shown their effectiveness in a large variety of scenes. They adopt deep neural networks to remove the motion blur in a single unpolarized image by extracting image features and priors from a large amount of training data. However, deblurring

¹ <https://thinklucid.com/product/phoenix-5-0-mp-polarized-model/>

multiple polarized images (captured by a polarization camera in a snapshot) could be a much more challenging problem than deblurring a single unpolarized image due to several fundamental issues:

- Polarization constraints (the numerical relationship between multiple polarized images constrained by the physical meanings of polarization) have to be preserved across multiple images in addition to removing the motion blur, making the problem highly ill-posed. Since the deblurring methods designed for conventional images can only process the polarized images in a polarization-unaware manner (*i.e.*, treat multiple polarized images as ordinary frames and ignore the physical meanings of polarization), they have no chance to take the preservation of polarization constraints into consideration.
- The DoFP technology adopted by modern polarization cameras induces a higher level of sensor noise (compared to the same sensor without micro-polarizers) (Kumar et al, 2007; Qiu et al, 2019; Abubakar et al, 2018), leading to degradation of low-level features (*e.g.*, structural and boundary information). Since current deblurring methods usually aim to deal with the images captured by conventional sensors, they cannot take advantage of the information and priors provided by the polarization-related parameters to mitigate the difficulties during the feature extraction and fusion procedure.
- Since the on-chip micro-polarizers attenuate scene radiance and the extent of attenuation is scene-dependent (Wu et al, 2020; Zhou et al, 2023a), the exposures turn out to be spatially-variant and arbitrary, making high-level features (*e.g.*, semantic and contextual information) less distinctive. Since recent deblurring methods usually adopt a single-stage pipeline that depends strongly on the quality of high-level features to hallucinate the sharp contents to restore the sharp image from its blurry counterpart in an end-to-end manner, they tend to suffer from ringing artifacts and fake textures.

In this paper, we propose a polarized image deblurring pipeline to solve the problem in a polarization-aware manner, by adopting a divide-and-conquer strategy to explicitly decompose the problem into two less ill-posed sub-problems: deblurring the corresponding unpolarized image for restoring the sharp image content, without considering the polarization constraints; utilizing the restored sharp image content to guide the deblurring procedure of the polarized images, considering the recovery of the polarization constraints. Concretely, we have two key observations on the characteristics of polarized images:

- (1) *Blur resistance property of Stokes parameters*: two (out of three) Stokes parameters (which are polarization-related parameters and will be introduced later) have similar ap-

pearance to the gradient distribution of the corresponding unpolarized image and are less affected by motion blur.

- (2) *Semantic similarity between polarized and unpolarized images*: the polarized images usually have similar semantic and contextual information to the corresponding unpolarized image.

Based on these inspiring characteristics of polarized images, we specially build a two-stage neural network to handle the two sub-problems respectively: The first stage is an *unpolarized image estimator* that leverages the less-contaminated structural and boundary information encoded in the Stokes parameters for repairing degraded low-level features. The second stage is a *polarized image reconstructor* that transfers the semantic and contextual information from the estimated unpolarized image to the polarized images with robustly extracted high-level features. To summarize, this paper makes contributions by demonstrating three customized model designs to deal with the three issues mentioned above:

- a polarized image deblurring pipeline to solve the problem in a polarization-aware manner, by decomposing it into two less-ill-posed sub-problems and handling them via:
- leveraging the less-contaminated structural and boundary information encoded in the Stokes parameters for repairing degraded low-level features; and
- transferring semantic and contextual information from the estimated unpolarized image to the polarized images with robustly extracted high-level features.

Experimental results show that our method achieves state-of-the-art performance on both synthetic and real-world images, and can improve the performance of polarization-based vision applications such as image dehazing and reflection removal.

2 Related work

2.1 Deblurring unpolarized images

Generally, methods for deblurring unpolarized images can be divided into two categories: image deblurring and video deblurring methods. We will have an overview on them respectively in the following.

Image deblurring methods. Previous works turned the problem into a maximum *a posteriori* (MAP) estimation problem and adopted numerical optimization to enforce sharpness (Fergus et al, 2006; Xu et al, 2013; Pan et al, 2016; Yang and Ji, 2019; Chen et al, 2020, 2021a; Zhang et al, 2022c). Recently, learning-based approaches have been introduced to deal with this problem (Zhang et al, 2022b). By extracting features from a large amount of training data to reduce the ill-posedness, they often show better performance and higher

efficiency than numerical optimization-based methods. Some works proposed to restore the sharp image from its blurry counterpart in an end-to-end manner using neural networks with different kinds of architectures, including convolutional neural network (CNN) (Nah et al, 2017; Zhang et al, 2018a; Tao et al, 2018; Gao et al, 2019; Zhang et al, 2019; Suin et al, 2020; Cho et al, 2021; Chi et al, 2021; Mou et al, 2022; Ji et al, 2022; Chen et al, 2022; Liu et al, 2024), generative adversarial network (GAN) (Kupyn et al, 2018, 2019; Zhang et al, 2020b; Wen et al, 2021b; Li et al, 2022a), multi-layer perceptron (MLP) (Tu et al, 2022), diffusion probabilistic model (DPM) (Whang et al, 2022; Ren et al, 2023), vision transformer (ViT) (Zamir et al, 2022; Wang et al, 2022b; Tsai et al, 2022; Chu et al, 2022; Kong et al, 2023), *etc.*, demonstrating visually impressive results. Some works proposed to deblur in a two-stage manner, by adopting neural networks to estimate blur-related physical parameters (*e.g.*, blur kernels and motion flows) first and then utilizing them to guide the deblurring process (Kaufman and Fattal, 2020; Ren et al, 2020; Dong et al, 2021; Tran et al, 2021; Chen et al, 2021b; Chakrabarti, 2016; Sun et al, 2015; Chen et al, 2018; Gong et al, 2017; Yuan et al, 2020), showing good generalization ability. Some methods adopted unrolling algorithms to bridge the gap between classical MAP optimization and pure deep learning techniques, which are also useful in deblurring or other image enhancement tasks (Kruse et al, 2017; Eboli et al, 2020; Zhang et al, 2020a). However, they are not suitable for handling multiple polarized images since they can only deblur in a polarization-unaware manner by processing one polarized image at a time, without considering the polarization constraints.

Video deblurring methods. Video deblurring methods can process multiple frames instead of a single frame at a time. Early works adopted numerical optimization to perform blur kernel estimation and deconvolution for restoring sharp frames (Bar et al, 2007; Wulff and Black, 2014; Ren et al, 2017). With the development of deep-learning, neural networks have also been adopted to handle this problem. These methods could be divided into three categories: window-based methods, recurrent methods, and memory-based methods. Window-based methods take multiple frames as input and try to reconstruct the middle frame, by referencing neighboring frames within a small fixed-size window (Su et al, 2017; Pan et al, 2020; Suin and Rajagopalan, 2021; Zhang et al, 2022a). Recurrent methods deblur frames in sequence, by collectively aggregating frame features till up-to-date to provide restoration cues (Wieschollek et al, 2017a; Hyun Kim et al, 2017; Nah et al, 2019; Zhou et al, 2019; Zhong et al, 2020; Li et al, 2021; Zhu et al, 2022; Jiang et al, 2022; Wang et al, 2022a). Memory-based methods allow the feature propagation from the entire sequence when restoring the given frame, by using memory banks to provide appropriate spatio-temporal information for each region and

exploiting the features of all frames (Ji and Yao, 2022). Additionally, there are methods designed for deblurring burst images (Aittala and Durand, 2018; Wieschollek et al, 2017b), which can also process multiple frames. However, they are still not suitable for handling multiple polarized images since the polarized images do not contain temporal information as the video frames or burst images.

2.2 Polarization and image enhancement

Here, we will introduce the intersection of polarization and image enhancement in two aspects: enhancing polarized images and polarization-based image enhancement.

Enhancing polarized images. There are many methods aiming to enhance the polarized images (*e.g.*, joint chromatic and polarimetric demosaicing (Tyo et al, 2009; Qiu et al, 2019; Wen et al, 2019, 2021a; Pistellato et al, 2022; Liu et al, 2022; Yi et al, 2024; Guo et al, 2024), denoising (Zhang et al, 2017; Abubakar et al, 2018; Tibbs et al, 2018; Li et al, 2020b; Liang et al, 2022; Li et al, 2023), and low-light enhancement (Hu et al, 2020; Li et al, 2022b; Xu et al, 2022; Zhou et al, 2023b)) to acquire the DoLP and AoLP accurately for improving the performance of downstream applications of polarization-based vision. Some methods perform enhancement in a direct manner, by processing multiple polarized images simultaneously in the intensity domain (*e.g.*, IPLNet (Hu et al, 2020)). However, their performance may not that good, since enhancing in the intensity domain cannot make effective use of the information provided by polarization. On the contrary, some methods adopt a Stokes-domain enhancement pipeline (*e.g.*, ColorPolarNet (Xu et al, 2022) and PLIE (Zhou et al, 2023b)) to enhance the polarized images in an indirect manner. These methods usually show higher performance and robustness, since the polarization characteristics of the incoming light to the sensor are fully encoded in the Stokes parameters, making the enhancement process more polarization-aware. Although the above methods are specially designed for processing multiple polarized images, the enhancement tasks targeted by them have completely different properties from the deblurring problem.

Polarization-based image enhancement. Different from enhancing polarized images, polarization-based image enhancement aims to utilize the polarization information to handle the image enhancement problems. Since the reflection and scattering phenomena would alter the polarization property of light, polarization information can be used as priors to eliminate the reflection artifacts (Lyu et al, 2019; Lei et al, 2020; Lyu et al, 2022) or remove the haze (Zhou et al, 2021; Ma et al, 2024a,b) inside the images. Since the polarization cameras can not only offer spatially-variant exposures but also contain abundant structural and contextual information of the scene (encoded in the DoLP and AoLP), we can also use polarization information to achieve HDR

reconstruction (Wu et al, 2020; Ting et al, 2021; Zhou et al, 2023a)). Although the above methods may also utilize the polarization-related parameters to provide guidance, their underlying strategies and objectives diverge from those of methods designed for enhancing polarized images, as they are designed to solve different types of problems.

3 Polarized image deblurring pipeline

Denoting the total intensity of the incoming light to the sensor as $\mathbf{I}$ (*i.e.*, the unpolarized image captured without polarizers), 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):

$$\mathbf{I}_\alpha = \frac{1}{2} \mathbf{I} \cdot (1 - \mathbf{p} \cdot \cos(2(\alpha - \theta))), \quad (1)$$

where α is the polarizer angle, $\mathbf{p} \in [0, 1]$ and $\theta \in [0, \pi]$ denote the DoLP and AoLP of the incoming light to the sensor respectively. Note that in this paper, we assume the camera response function to be linear since polarization cameras usually output images with a linear camera response function, and focus on linear polarization (*i.e.*, do not consider circular polarization) since polarization cameras only equip linear polarizers.

By virtue of four-directional, on-chip micro-polarizers provided by the DoFP technology, a modern polarization camera can capture four spatially-aligned and temporally-synchronized linear polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ with different polarizer angles $\alpha_{1,2,3,4} = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ in a snapshot, which brings convenience to the acquisition of $\mathbf{p}$ and θ . This is because $\mathbf{p}$ can be calculated using

$$\mathbf{p} = \frac{\sqrt{\mathbf{S}_1^2 + \mathbf{S}_2^2}}{\mathbf{S}_0}, \quad (2)$$

and θ can be calculated using

$$\theta = \frac{1}{2} \arctan\left(\frac{\mathbf{S}_2}{\mathbf{S}_1}\right), \quad (3)$$

where $\mathbf{S}_{0,1,2}$ are called the Stokes parameters (Können, 1985) of the incoming light to the sensor, which can be directly obtained from the captured polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ according to their physical meanings²:

$$\begin{cases} \mathbf{S}_0 = \mathbf{I} = \frac{1}{2}(\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_3} + \mathbf{I}_{\alpha_4}) \\ \mathbf{S}_1 = \mathbf{I}_{\alpha_3} - \mathbf{I}_{\alpha_1} \\ \mathbf{S}_2 = \mathbf{I}_{\alpha_4} - \mathbf{I}_{\alpha_2} \end{cases}. \quad (4)$$

² $\mathbf{S}_0$ describes the total intensity of the light, and $\mathbf{S}_1$ ($\mathbf{S}_2$) describes the difference between the intensity of the vertical (135°) and horizontal (45°) polarized light (Können, 1985).

When motion blur occurs, denoting the blurry counterparts of $\mathbf{I}$ and $\mathbf{I}_{\alpha_i}$ as $\mathbf{B}$ and $\mathbf{B}_{\alpha_i}$ respectively ($i = 1, 2, 3, 4$), their relationships can be written as

$$\mathbf{B} = \mathbf{I} \otimes \mathbf{K} + \mathbf{N}, \quad (5)$$

and

$$\mathbf{B}_{\alpha_i} = \mathbf{I}_{\alpha_i} \otimes \mathbf{K} + \mathbf{N}_i \quad (i = 1, 2, 3, 4), \quad (6)$$

where $\otimes$ stands for the convolution operation, $\mathbf{K}$ denotes a spatially-variant blur kernel determined by the motion field, $\mathbf{N}$ and $\mathbf{N}_i$ represent noise terms depending on the input signal and camera configuration (Hu et al, 2020; Xu et al, 2022; Zhou et al, 2023b). Note that both $\mathbf{B}$ and $\mathbf{B}_{\alpha_i}$ ($i = 1, 2, 3, 4$) can be considered to share the same blur kernel $\mathbf{K}$, since the pixel shifts among the polarized images caused by mosaicing is small enough (usually 1 pixel for a 4×4 mosaicing of polarization and color filters inside a polarization camera), which will not cause significant changes to the motion field and is often ignored by other works designed for enhancing the polarized images (Hu et al, 2020; Li et al, 2022b; Xu et al, 2022; Zhou et al, 2023b).

We aim to restore $\mathbf{p}$ and θ from their degraded counterparts $\mathbf{p}^*$ and θ^* by deblurring $\mathbf{B}_{\alpha_{1,2,3,4}}$. Unlike conventional image deblurring problems which only need to focus on removing the blurry artifacts, our problem becomes more ill-posed since we need to consider how to preserve the polarization constraints across the polarized images in addition to deblurring. This is because according to Equation (2), Equation (3), and Equation (4), the relationship among $\mathbf{p}$, θ , and $\mathbf{I}_{\alpha_{1,2,3,4}}$ involves non-linear operations, which makes the accuracy of $\mathbf{p}$ and θ highly sensitive to the preservation of the polarization constraints (Hu et al, 2020; Zhou et al, 2023b). Therefore, we adopt a divide-and-conquer strategy to explicitly decompose the problem into two less ill-posed sub-problems, and propose a two-stage polarized image deblurring pipeline to handle them respectively in a polarization-aware manner:

- The first stage aims to deblur the corresponding unpolarized image for restoring the sharp image content, without considering the polarization constraints.
- The second stage aims to utilize the restored sharp image content to guide the deblurring procedure of the polarized images, considering the recovery of the polarization constraints.

This pipeline is implemented by building a two-stage neural network, as shown in Fig. 1, and we will detail its model designs in the following section.

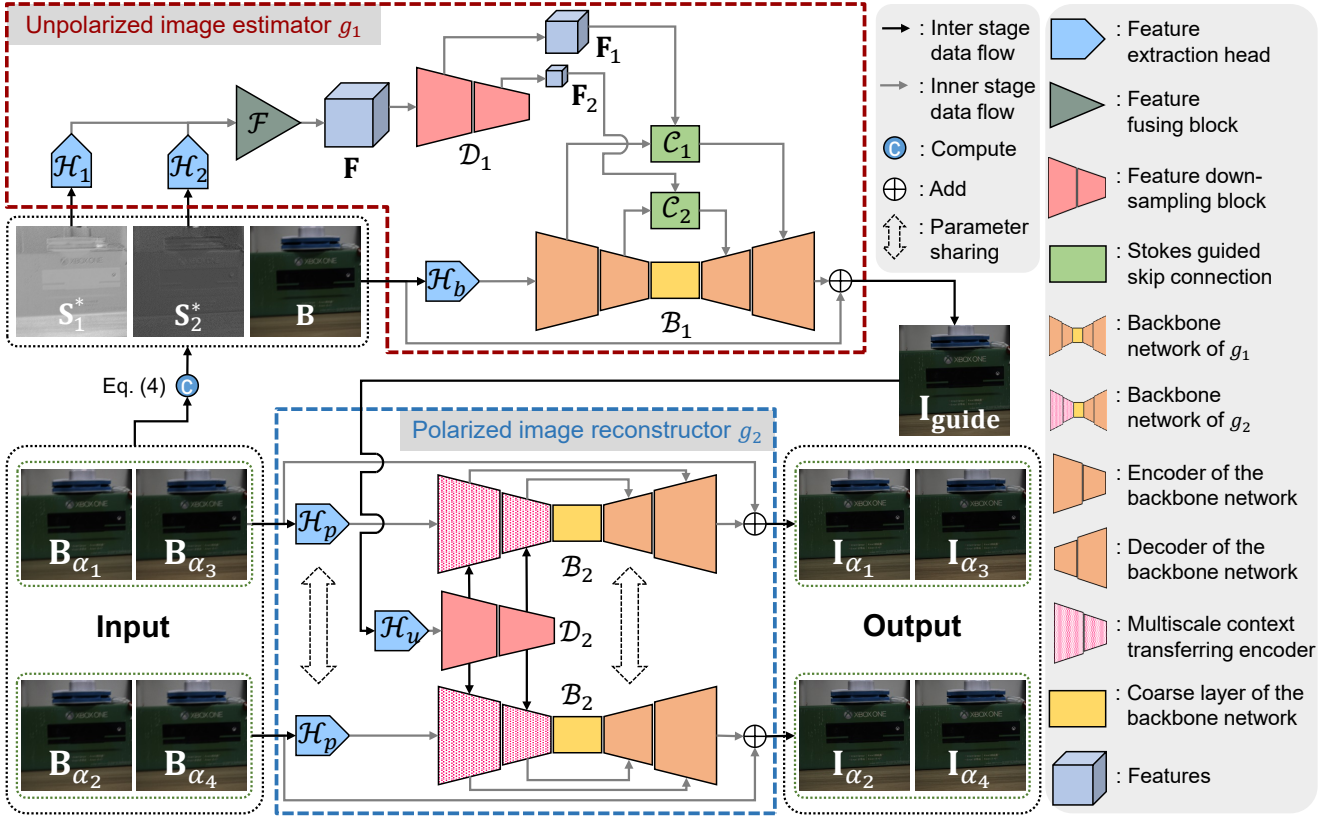


Fig. 1 Architecture of the neural network that implements the proposed polarized image deblurring pipeline. It consists of two stages: an unpolarized image estimator that uses $S_{1,2}^*$ (the blurry counterparts of the Stokes parameters $S_{1,2}$) as priors to estimate the sharp unpolarized image I_{guide} , and a polarized image reconstructor that reconstructs $I_{\alpha_{1,2,3,4}}$ under the guidance of I_{guide} . All images are normalized for visualization.

4 Network model design

4.1 Unpolarized image estimator

Overall architecture. We design the first stage as an unpolarized image estimator g_1 , aiming to estimate I from B . Apparently, the problem handled by g_1 is similar to deblurring a single conventional image, however, since B is computed from $B_{\alpha_{1,2,3,4}}$ which are captured by a polarization camera, the low-level features (*e.g.*, structural and boundary information) would suffer from severe degradation, making the problem more challenging. This is because the DoFP technology adopted by modern polarization cameras induces higher level sensor noise (compared to the same sensor without micro-polarizers) (Kumar et al, 2007; Qiu et al, 2019; Abubakar et al, 2018) so that the noise term N in Equation (5) and Equation (6) becomes non-negligible. Prominently, we notice that there are two Stokes parameters ($S_{1,2}$) having similar “appearance” to the gradient distribution G of the corresponding unpolarized image due to their “differential” nature (*i.e.*, both of them describe the difference between two polarized images according to Equation (4)), as shown in Fig. 2 (left); besides, $S_{1,2}$ are less affected by motion blur since they often have relatively stronger response near the

edges, which means that their blurry counterparts $S_{1,2}^*$ could retain less-contaminated structural and boundary information. To verify it, we design a simulation experiment on our training set (including 8000 images, see Section 5.2 for details) to quantitatively measure the gap between the sharp/blurry variable pairs, by computing the average value of LPIPS score (learned perceptual image patch similarity, which is suitable for measuring the similarity distance in the feature space) (Zhang et al, 2018b). Results are shown in Fig. 2 (middle), we can see that the degradation degree of $S_{1,2}$ is nearly half of that of G . Therefore, we expect that $S_{1,2}^*$ can compensate the degraded low-level features, and decide to use them as priors. As shown in the red dashed frame of Fig. 1, this stage can be written as the following equation:

$$I_{\text{guide}} = g_1(B, S_{1,2}^*), \quad (7)$$

where I_{guide} denotes the estimated sharp unpolarized image by g_1 , which will serve as an important guiding signal in the next stage.

Layer details. We first adopt two feature extraction heads (consisting of several convolution layers) $\mathcal{H}_1$ and $\mathcal{H}_2$ to extract features from S_1^* and S_2^* respectively, and adopt a feature fusing block (consisting of a convolution layer and a dense

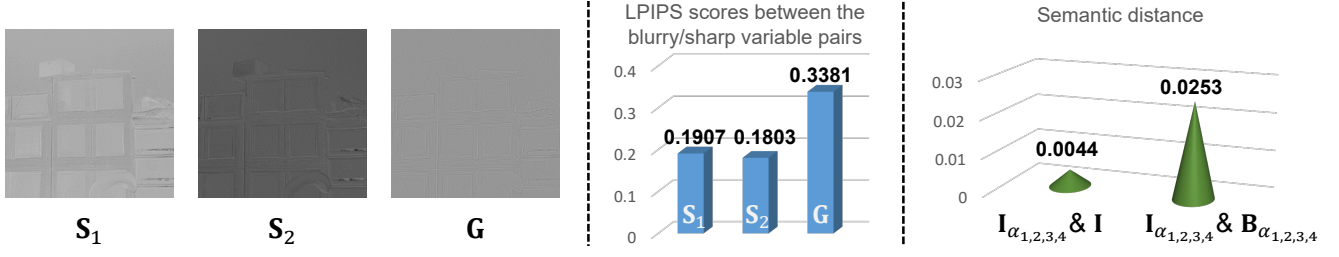


Fig. 2 Left: Two of the Stokes parameters ($S_{1,2}$) have similar “appearance” to the gradient distribution G of the corresponding unpolarized image (please zoom-in for better details). Middle: LPIPS (Zhang et al, 2018b) scores between the blurry/sharp variable pairs, including $S_{1,2}$ and G . Right: Semantic distance between $I_{\alpha_{1,2,3,4}}$ and I , $I_{\alpha_{1,2,3,4}}$ and $B_{\alpha_{1,2,3,4}}$.

block (Huang et al, 2017)) $\mathcal{F}$ to obtain the fused Stokes features F . Then, we adopt another feature extraction head $\mathcal{H}_b$ to extract features from B , and feed them along with F into a backbone network $\mathcal{B}_1$ (a modified autoencoder architecture (Hinton and Salakhutdinov, 2006) similar to the one used in (Zhou et al, 2023b)) to learn the residual between B and I . For fully utilizing the less-contaminated structural and boundary information retained in $S_{1,2}^*$, we need to allow F to directly affect B across the entire feature domain. Therefore, instead of directly feeding it into the encoder of $\mathcal{B}_1$, we propose to use Stokes guided skip connections (SGSC) to substitute the skip connections inside $\mathcal{B}_1$ for processing it. SGSC accepts two kinds of inputs: the features from the backbone network and the features from the Stokes parameters. It uses residual bottleneck blocks (He et al, 2016) to filter each feature channel, and uses a squeeze-and-excitation block (Hu et al, 2018) to recalibrate the features by learning normalized weights for recovering fine-grained high-order image structure. Specifically, we adopt a feature downsampling block (consisting of several convolution layers and residual blocks (He et al, 2016)) $\mathcal{D}_1$ to downsample F to get multiscale features of the Stokes parameters (F_1 and F_2), and adopt two SGSCs ($\mathcal{C}_1$ and $\mathcal{C}_2$) to process them respectively.

4.2 Polarized image reconstructor

Overall architecture. We design the second stage as a polarized image reconstructor g_2 , aiming to reconstruct $I_{\alpha_{1,2,3,4}}$ from $B_{\alpha_{1,2,3,4}}$ under the guidance of I_{guide} . As shown in the blue dashed frame of Fig. 1, this stage can be written as the following equation:

$$I_{\alpha_{1,2,3,4}} = g_2(B_{\alpha_{1,2,3,4}}, I_{\text{guide}}). \quad (8)$$

However, from Equation (1) we can see that the polarized image I_α can be regarded as a re-exposed version of the unpolarized image I with spatially-variant and arbitrary exposures (i.e., $I_\alpha = I \cdot M$ where $M = \frac{1}{2}(1 - \mathbf{p} \cdot \cos(2(\alpha - \theta))) \in [0, 1]$ is a scene-dependent exposure mask) (Wu et al, 2020; Zhou et al, 2023a), which poses a major challenge to efficiently extract high-level features (e.g., semantic and contextual information) from $B_{\alpha_{1,2,3,4}}$ for deblurring. Particularly, we

have observed that the polarized images usually have similar semantic and contextual information to the corresponding unpolarized image. To verify it, we design another simulation experiment on our training set to quantitatively measure the gap between I and $I_{\alpha_{1,2,3,4}}$, by using the contrastive language-image pre-training (CLIP) model (Radford et al, 2021) to obtain their encoded semantic feature maps and compute the average value of MSE between them (which we defined as the semantic distance) (Chefer et al, 2022). Results are shown in Fig. 2 (right), we can see that the semantic distance between $I_{\alpha_{1,2,3,4}}$ and I could be very small (even nearly one order of magnitude smaller than that between $I_{\alpha_{1,2,3,4}}$ and $B_{\alpha_{1,2,3,4}}$). Therefore, we choose to transfer the semantic and contextual information encoded in the estimated unpolarized image I_{guide} to the target polarized images $I_{\alpha_{1,2,3,4}}$.

Layer details. Similar to g_1 , we adopt another backbone network $\mathcal{B}_2$ to learn the residual between $B_{\alpha_{1,2,3,4}}$ and $I_{\alpha_{1,2,3,4}}$. Since we want semantic information to flow from I_{guide} to $B_{\alpha_{1,2,3,4}}$ for guidance, instead of directly feeding the concatenated features of $B_{\alpha_{1,2,3,4}}$ and I_{guide} into $\mathcal{B}_2$, we propose to use multiscale context transferring encoder (MCTE) to substitute the encoder inside $\mathcal{B}_2$ for processing them. MCTE adopts several convolution layers to perform context transferring in each scale, and then uses residual blocks (He et al, 2016) to refine the features for enriching details. Specifically, we adopt another feature downsampling block $\mathcal{D}_2$ to downsample the features of I_{guide} extracted by a feature extraction head $\mathcal{H}_u$, and then feed them into MCTE along with the features of $B_{\alpha_{1,2,3,4}}$ extracted by another feature extraction head $\mathcal{H}_p$. Besides, plugging the values of $\alpha_{1,2,3,4}$ into Equation (1), we can see that $I = I_{\alpha_1} + I_{\alpha_3} = I_{\alpha_2} + I_{\alpha_4}$. So, we propose a polarization-based grouping (PG) strategy, by dividing $B_{\alpha_{1,2,3,4}}$ into two groups ($B_{\alpha_{1,3}}$ and $B_{\alpha_{2,4}}$) and adopting two sets of network modules with shared parameters (including $\mathcal{H}_p$ and $\mathcal{B}_2$) to deal with them respectively. In this way, the problem could be further decomposed into two guided image decomposition problems: separating two image layers $I_{\alpha_{1,3}}$ (or $I_{\alpha_{2,4}}$) from a “mixture image” I_{guide} under the guidance of their blurry counterparts $B_{\alpha_{1,3}}$ (or $B_{\alpha_{2,4}}$), which could be more robust than jointly processing $B_{\alpha_{1,2,3,4}}$ in a direct manner.

5 Implementation details

5.1 Loss function

The total loss function L can be written as

$$L(\mathbf{I}_{\text{guide}}, \mathbf{I}_{\alpha_{1,2,3,4}}, \mathbf{S}_{0,1,2}) = \lambda_1 L_c(\mathbf{I}_{\text{guide}}) + \lambda_2 \sum_{i=1}^4 L_c(\mathbf{I}_{\alpha_i}) + \lambda_3 L_s(\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2), \quad (9)$$

where $\lambda_{1,2,3}$ are set to be 0.5, 1, and 1 respectively, L_c (content loss) and L_s (Stokes loss) denote two kinds of basic loss functions used during the training process respectively. We will detail each of them in the following.

Content loss. The content loss L_c aims to bridge the gap between the restored image $\mathbf{I}_V$ (which can be any one of $\mathbf{I}_{\text{guide}}$ and $\mathbf{I}_{\alpha_{1,2,3,4}}$) and its corresponding ground truth for sharper image content, which is defined as

$$L_c(\mathbf{I}_V) = \lambda_{c1} L_1(\mathbf{I}_V) + \lambda_{c2} L_2(\mathbf{I}_V) + \lambda_{c3} L_{\text{perc}}(\mathbf{I}_V) + \lambda_{c4} L_{\text{edge}}(\mathbf{I}_V), \quad (10)$$

where L_1 , L_2 , L_{perc} , and L_{edge} denote the ℓ_1 , ℓ_2 , perceptual, and edge loss respectively, and $\lambda_{c1,2,3,4}$ are 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 extracted by $VGG_{3,3}$ convolution layer of VGG-19 network (Simonyan and Zisserman, 2014) pretrained on ImageNet (Russakovsky et al, 2015), and the edge loss L_{edge} is defined as the ℓ_2 loss computed using the edges extracted by Laplace kernels.

Stokes loss. The Stokes loss L_s aims to bridge the gap between the restored Stokes parameters $\mathbf{S}_{0,1,2}$ (computed from the restored polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ using Equation (4)) and their corresponding ground truth for more accurate polarization constraints, which is defined as

$$L_s(\mathbf{S}_0, \mathbf{S}_1, \mathbf{S}_2) = \lambda_{s1} L_2(\mathbf{S}_0) + \lambda_{s2} L_2(\mathbf{S}_1) + \lambda_{s3} L_2(\mathbf{S}_2) + \lambda_{s4} L_2\left(\frac{\mathbf{S}_2}{\mathbf{S}_1}\right), \quad (11)$$

where $\lambda_{s1,2,3,4}$ are set to be 20, 500, 500, and 500 respectively. Note that the fourth term of L_s is adopted to optimize the ratio between $\mathbf{S}_2$ and $\mathbf{S}_1$ since the calculation of the AoLP depends on it (see Equation (3)).

5.2 Dataset preparation

There is no public dataset for such a polarized image deblurring task, and it is difficult to obtain pairwise blurry and sharp polarized images at a large scale. Besides, existing deblurring benchmark datasets (*e.g.*, GOPRO (Nah et al, 2017), HIDE (Shen et al, 2019), and RealBlur (Rim et al, 2020)) do not contain any polarization information, which cannot be used to generate polarized images. Therefore, we propose to generate a synthetic dataset for training our network. The generating process can be described as the following steps:

- (1) Collect polarized images from two public polarized low-light image enhancement datasets (LLCP (Xu et al, 2022) and PLIE (Zhou et al, 2023b)) to serve as the ground-truth sharp images.
- (2) Randomly generate camera shake trajectories using the algorithm proposed by Boracchi and Foi (2012).
- (3) Adopt the spatially-variant blur model proposed by Whyte et al (2012) to make the trajectories pixel-wise.
- (4) Move the corresponding ground-truth sharp images along the pixel-wise trajectories to get multiple sharp latent frames.
- (5) Averaging those sharp latent frames to synthesize the blurry images.
- (6) Add noise to the blurry images (as done in other works (Hu et al, 2020; Xu et al, 2022; Zhou et al, 2023b)) to better simulate the real situation.

The LLCP (Xu et al, 2022) and PLIE (Zhou et al, 2023b) datasets provide 430 different scenes in total, and we randomly split them into two parts that contain 400 and 30 scenes for making the training and test sets respectively. For each scene in the training (test) set, we randomly generate 20 (10) different camera shake trajectories and further perform data augmentation (*e.g.*, flipping and rotating) so that the training (test) set contains 8000 (300) different images finally. Note that data augmentation is only applied when generating the training set. The images are resized and randomly cropped to 256×256 (512×512) pixels in the training (test) set.

5.3 Training strategy

We implement our method using PyTorch with 4 NVIDIA 1080Ti GPUs and apply a two-stage training strategy. First, to ensure a stable initialization of the training process, we train the unpolarized image estimator and the polarized image reconstructor independently for 600 and 300 epochs respectively with a learning rate of 0.002. Then, we finetune the entire network in an end-to-end manner for another 100 epochs with a learning rate of 0.001.

For optimization, we use Adam optimizer (Kingma and Ba, 2014) with $\beta_1 = 0.5$, $\beta_2 = 0.999$. The network parameters are initialized with Xavier initialization (Glorot and Bengio, 2010). We add an instance normalization (Ulyanov et al, 2016) layer and a ReLU activation function after each convolution layer.

6 Experiments

6.1 Evaluation on synthetic data

Since there is no existing approach specially designed for deblurring polarized images, we choose to compare our method

Table 1 Quantitative comparisons on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFilter (Liu et al, 2024). **Bold** font indicates the best performance.

	PSNR-p	SSIM-p	PSNR- θ	SSIM- θ	PSNR-I	SSIM-I
Ours	29.12	0.774	17.77	0.330	29.67	0.868
IPLNet (Hu et al, 2020)	19.98	0.702	17.01	0.267	26.08	0.822
ColorPolarNet (Xu et al, 2022)	28.41	0.740	16.99	0.279	25.09	0.812
PLIE (Zhou et al, 2023b)	26.32	0.744	17.24	0.269	28.40	0.853
MIMO (Cho et al, 2021)	26.61	0.730	16.51	0.280	25.92	0.830
XYDeblur (Ji et al, 2022)	22.74	0.698	15.83	0.272	26.68	0.850
STDAN (Zhang et al, 2022a)	28.28	0.756	17.06	0.309	27.49	0.862
MISCFilter (Liu et al, 2024)	28.75	0.763	17.16	0.311	28.45	0.865

Table 2 Quantitative comparisons on real data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFilter (Liu et al, 2024). **Bold** font indicates the best performance.

	CLIP-IQA-p	CLIP-IQA- θ	CLIP-IQA-I
Ours	0.506	0.677	0.439
IPLNet (Hu et al, 2020)	0.503	0.628	0.428
ColorPolarNet (Xu et al, 2022)	0.479	0.592	0.366
PLIE (Zhou et al, 2023b)	0.484	0.589	0.274
MIMO (Cho et al, 2021)	0.500	0.667	0.343
XYDeblur (Ji et al, 2022)	0.504	0.624	0.359
STDAN (Zhang et al, 2022a)	0.479	0.593	0.366
MISCFilter (Liu et al, 2024)	0.433	0.582	0.361

to two groups of methods that solve the closest problems. The first group is originally designed for handling the polarized image enhancement tasks, including IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), and PLIE (Zhou et al, 2023b). The second group is originally designed for handling the unpolarized image/video deblurring tasks, including MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFilter (Liu et al, 2024). All methods in the first group can process the polarized images simultaneously in a direct manner. In the second group, STDAN (Zhang et al, 2022a) can process the polarized images simultaneously in an indirect manner since it is designed for deblurring continuous video frames, while MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), and MISCFilter (Liu et al, 2024) can only process the polarized images in a frame-by-frame manner since it is designed for deblurring a single image. Note that since STDAN (Zhang et al, 2022a) reconstructs the middle frame by referencing 5 neighboring frames, when compared to it we set its input as a video whose frames are stacked by the blurry polarized images in the order of “ $[\mathbf{B}_{\alpha_3}, \mathbf{B}_{\alpha_4}, \mathbf{B}_{\alpha_1}, \mathbf{B}_{\alpha_2}, \mathbf{B}_{\alpha_3}, \mathbf{B}_{\alpha_4}, \mathbf{B}_{\alpha_1}, \mathbf{B}_{\alpha_2}]$ ” to ensure that the output frames are in the order of “ $[\mathbf{B}_{\alpha_1}, \mathbf{B}_{\alpha_2}, \mathbf{B}_{\alpha_3}, \mathbf{B}_{\alpha_4}]$ ”.

All the compared methods are trained on our dataset (either finetuned or retrained for adaptation to our dataset) under the supervision of the polarized images $\mathbf{I}_{\alpha_1,2,3,4}$ (the very same outputs of our method). In practice, we choose to finetune (*i.e.*, train from pre-trained weights) the methods designed for deblurring unpolarized images (Cho et al, 2021;

Ji et al, 2022; Liu et al, 2024; Zhang et al, 2022a) and retrain (*i.e.*, train from scratch) the methods designed for enhancing polarized images (Hu et al, 2020; Xu et al, 2022; Zhou et al, 2023b). Such a training strategy for the compared methods could be reasonable and often adopted by other methods (Hu et al, 2020; Xu et al, 2022; Zhou et al, 2023b), since finetuning is commonly used when adapting a model to a specific task or domain that is closely related to the original task the model was trained on (*e.g.*, deblurring unpolarized images *vs.* deblurring polarized images), while retraining is typically used when the task is different from the original tasks that pre-trained models cannot provide a good starting point (*e.g.*, enhancing low-light polarized images *vs.* deblurring polarized images). Besides, for a fair comparison, we apply our loss function to the methods designed for enhancing polarized images (Hu et al, 2020; Xu et al, 2022; Zhou et al, 2023b), while preserving the original loss functions for the methods designed for deblurring unpolarized images (Cho et al, 2021; Ji et al, 2022; Liu et al, 2024; Zhang et al, 2022a) during training. This is because our loss function includes a Stokes loss term (Equation (11)), which requires multiple polarized images to compute the polarization-related parameters, while multiple polarized images cannot be obtained within a single forward pass for the methods designed for deblurring unpolarized images. Additionally, forcibly modifying the forward/backward propagation logic of these methods to accommodate our loss function may degrade their performance rather than improve it.

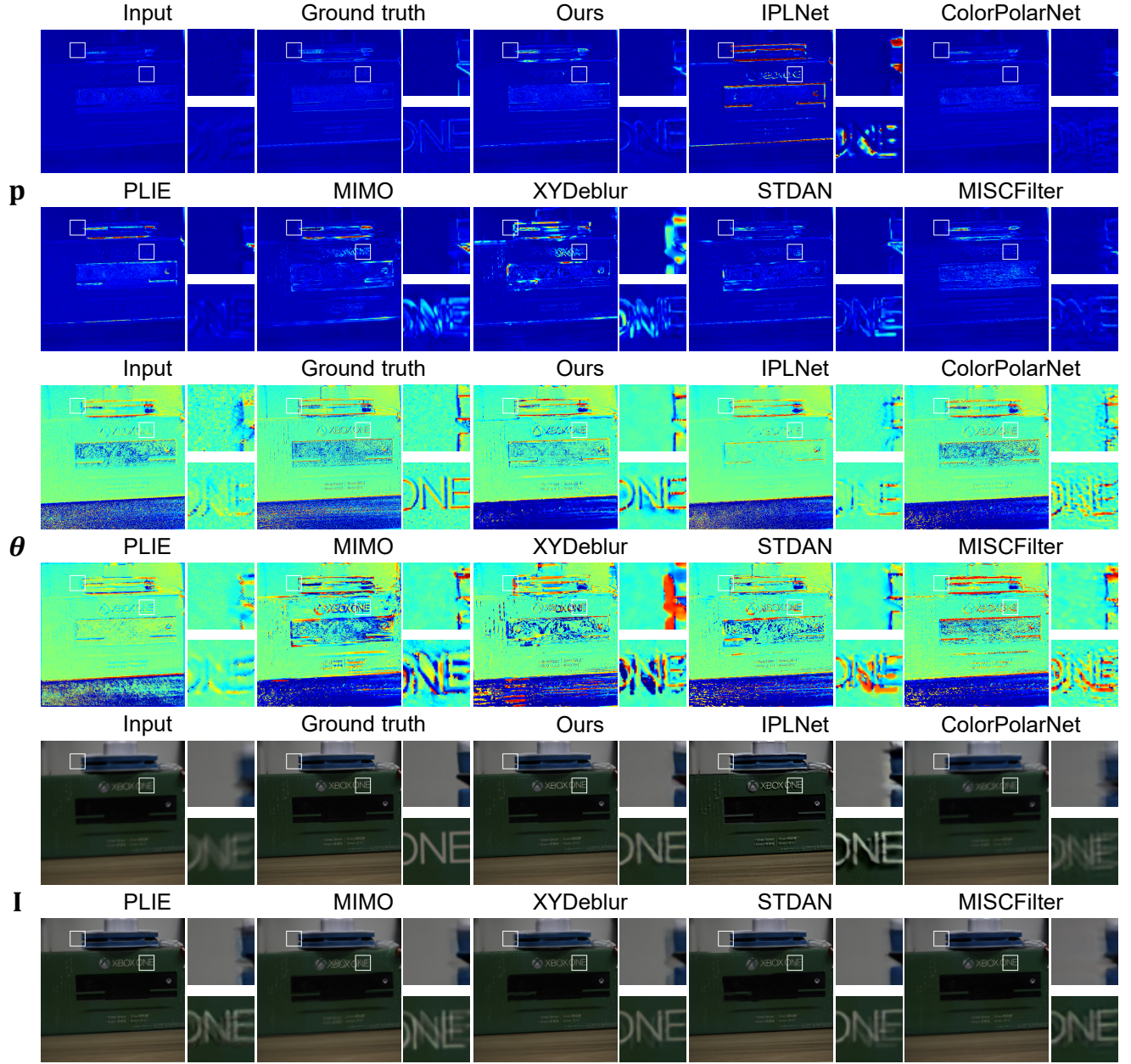


Fig. 3 Qualitative comparisons on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFILTER (Liu et al, 2024). The DoLP $\mathbf{p}$ and AoLP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023b)). Please zoom-in for better details.

To evaluate the results quantitatively, we choose to adopt two frequently-used metrics including PSNR and SSIM. Results are shown in Table 1. Our model consistently outperforms the compared methods on all metrics. Visual quality comparisons are shown in Fig. 3³. Note that we not only evaluate the accuracy of the restored DoLP $\mathbf{p}$ and AoLP θ , but also evaluate the quality of the restored unpolarized image $\mathbf{I}$. As for $\mathbf{p}$ and θ , our method achieves better performance than other methods, because we explicitly consider preserv-

ing the polarization constraints. Taking the second box of $\mathbf{p}$ as an example, the compared methods either suffer from severe artifacts (*e.g.*, IPLNet (Hu et al, 2020), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), and STDAN (Zhang et al, 2022a)) or tend to generate over-smooth edges (*e.g.*, ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), and MISCFILTER (Liu et al, 2024)). This is because the methods designed for deblurring can only process the polarized images in a polarization-unaware manner, while the methods designed for enhancing polarized images do not take the

³ Additional results can be found in the supplementary material.

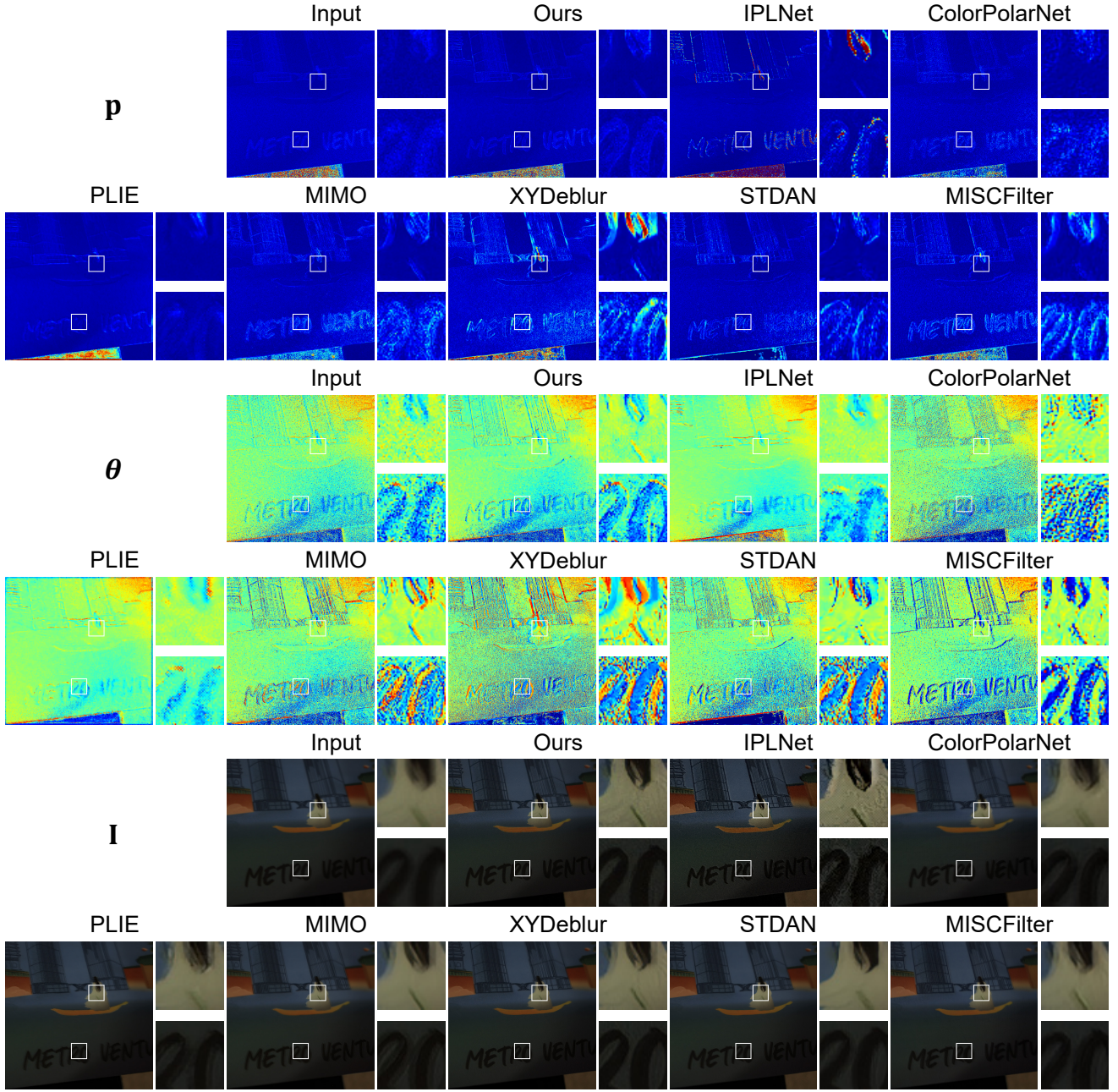


Fig. 4 Qualitative comparisons on real data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFilter (Liu et al, 2024). The DoLP $\mathbf{p}$ and AoLP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023b)). Please zoom-in for better details.

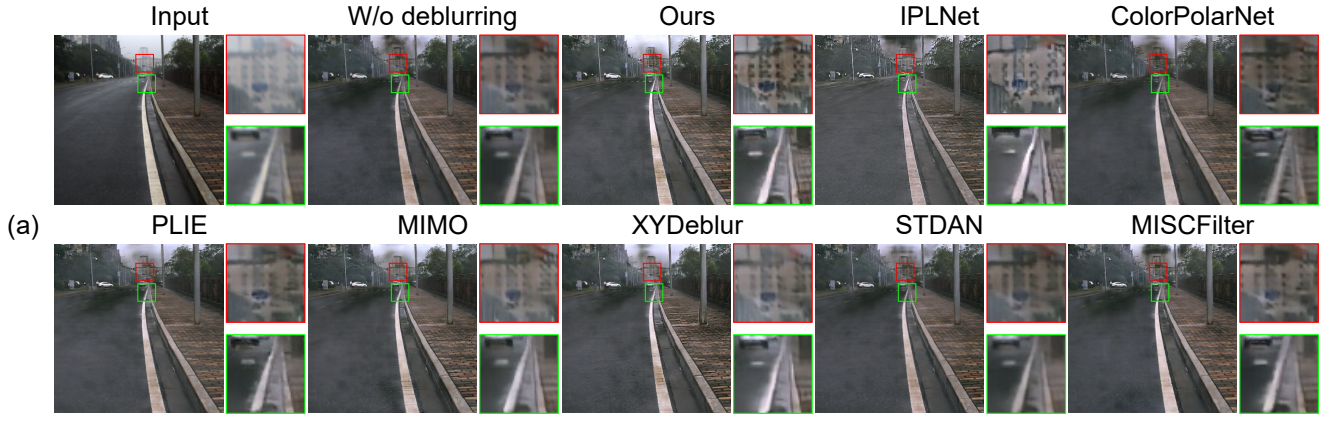
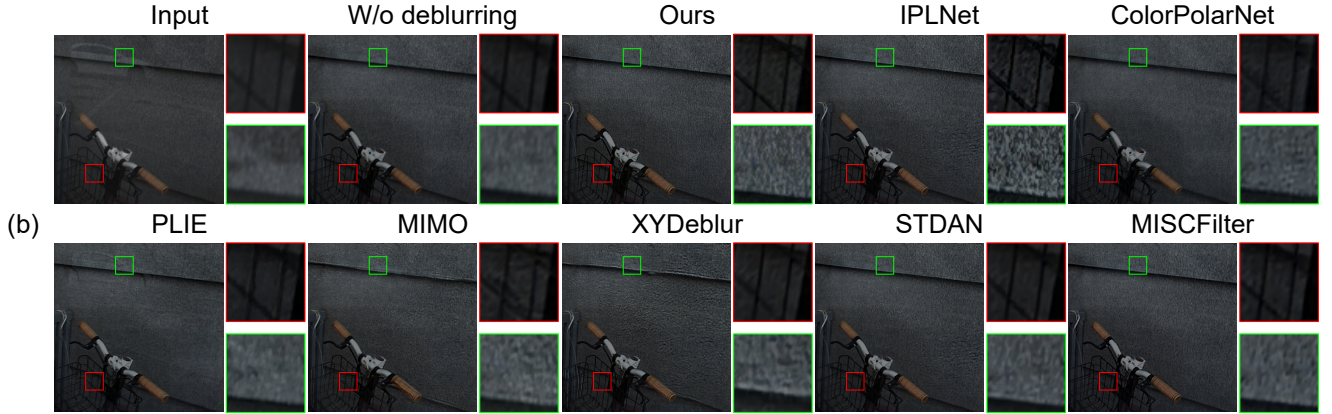
unique properties of motion blur into consideration. As for **I**, our results resemble the ground truth more closely with less distortion. This is because our method can properly deal with the degraded low-level features and less distinctive high-level features.

6.2 Evaluation on real data

To demonstrate the strong generalization capability of our method on complex and diverse real-world data, we capture a real-world dataset using a Lucid Vision Phoenix polarization camera. The dataset comprises 40 different scenes, which exhibits a range of conditions (*e.g.*, different environments, varying blur patterns, *etc.*). Note that the dataset does not provide the ground truth for reference.

Table 3 Quantitative evaluation results of ablation study on synthetic data. **Bold** font indicates the best performance.

	PSNR- $\mathbf{p}$	SSIM- $\mathbf{p}$	PSNR- θ	SSIM- θ	PSNR- $\mathbf{I}$	SSIM- $\mathbf{I}$
W/o $\mathbf{I}_{\text{guide}}$	27.12	0.730	17.14	0.241	28.36	0.856
W/o $\mathbf{S}_{1,2}^*$	28.41	0.772	17.67	0.329	29.16	0.863
$\mathbf{G}$ instead of $\mathbf{S}_{1,2}^*$	28.43	0.768	17.70	0.329	29.19	0.864
W/o SGSC	29.11	0.773	17.71	0.327	28.99	0.862
W/o MCTE	28.63	0.760	17.17	0.267	29.38	0.866
W/o PG	29.10	0.773	17.52	0.293	29.50	0.866
W/o L_s	29.03	0.769	17.48	0.314	29.62	0.867
Our complete model	29.12	0.774	17.77	0.330	29.67	0.868

Application: image dehazing**Application: reflection removal****Fig. 5** Results of polarization-based vision applications (including (a) image dehazing (Zhou et al, 2021) and (b) reflection removal (Lyu et al, 2019, 2022)) without deblurring and deblurred by our method and the compared methods (IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFILTER (Liu et al, 2024)). Please zoom-in for better details.

To evaluate the results quantitatively, we choose to adopt a non-reference metric, CLIP-IQA (Wang et al, 2023) (contrastive language-image pretraining for image quality assessment). The results, presented in Table 2, indicate that our model outperforms the compared methods, showcasing superior performance across the dataset. Visual quality comparisons are shown in Fig. 4⁴. From the results we can see

that our method suffer less from artifacts. Taking the first box of $\mathbf{p}$ as an example, the compared methods tend to produce inaccurate results (shown as abnormal colors). Taking the first box of $\mathbf{I}$ as another example, we can see that IPLNet (Hu et al, 2020) brings about ringing artifacts with higher noise level, while other compared methods fail to restore sharp image content.

⁴ Additional results can be found in the supplementary material.

Table 4 Computational complexity analysis on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023b), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022a), and MISCFILTER (Liu et al, 2024). Here, the size of the input images is 512×512 pixels (3 color channels), and the inference time is the time taken to reconstruct all 300 scenes.

	Number of parameters	FLOPs	Memory	Inference time
Ours	3.14 M	360.20 G	2.39 G	4.36 s
IPLNet (Hu et al, 2020)	1.80 M	944.97 G	1.89 G	3.57 s
ColorPolarNet (Xu et al, 2022)	1.03 M	943.94 G	1.51 G	1.90 s
PLIE (Zhou et al, 2023b)	1.50 M	453.94 G	4.28 G	4.84 s
MIMO (Cho et al, 2021)	16.11 M	1207.58 G	2.46 G	13.56 s
XYDeblur (Ji et al, 2022)	4.92 M	550.40 G	2.18 G	12.19 s
STDAN (Zhang et al, 2022a)	13.84 M	3766.04 G	5.94 G	129.73 s
MISCFILTER (Liu et al, 2024)	15.99 M	993.29 G	2.82 G	60.92 s

6.3 Ablation study

To verify the validity of each design choice, we conduct a series of ablation studies, which can be expressed as

- (1) W/o $\mathbf{I}_{\text{guide}}$: deblurring $\mathbf{B}_{\alpha_{1,2,3,4}}$ directly without estimating $\mathbf{I}_{\text{guide}}$ first (to demonstrate the contribution of our two-stage deblurring pipeline).
- (2) W/o $\mathbf{S}_{1,2}^*$: removing $\mathbf{S}_{1,2}^*$ from g_1 (to verify the necessity of using $\mathbf{S}_{1,2}^*$ as priors).
- (3) $\mathbf{G}$ instead of $\mathbf{S}_{1,2}^*$: substituting the Stokes parameters $\mathbf{S}_{1,2}^*$ with the gradient distribution $\mathbf{G}$ (to show the advantage of using $\mathbf{S}_{1,2}^*$ over using $\mathbf{G}$).
- (4) W/o SGSC: directly feeding the fused Stokes features $\mathbf{F}$ into the encoder of $\mathcal{B}_1$ (to show the importance of SGSC used in g_1).
- (5) W/o MCTE: directly feeding the concatenated features of $\mathbf{B}_{\alpha_{1,2,3,4}}$ and $\mathbf{I}_{\text{guide}}$ into $\mathcal{B}_2$ (to show the importance of MCTE used in g_2).
- (6) W/o PG: using a single backbone network to process $\mathbf{B}_{\alpha_{1,2,3,4}}$ jointly (to validate the effectiveness of adopting the PG strategy in g_2).
- (7) W/o L_s : removing the Stokes loss (to show the benefit of using such kind of loss in our setting).

Quantitative comparisons are shown in Table 3. These results demonstrate that our complete model achieves the optimal performance with these specific design choices.

6.4 Applications

To show the benefits of deblurring polarized images, we choose two different polarization-based vision applications including image dehazing (Zhou et al, 2021) and reflection removal (Lyu et al, 2019, 2022) as the validation methods respectively, and show that deblurring can improve their performance. Comparisons of the haze-removed and reflection-removed unpolarized images are shown in Fig. 5 (a) and (b) respectively⁵, which demonstrate that our method brings superior performance gain than the compared methods. Taking

the red box in Fig. 5 (a) as an example, we can see that the haze cannot be removed robustly without deblurring, and the results deblurred by the compared methods are still hazy, while the result deblurred by our method has higher clarity. Taking the green box in Fig. 5 (b) as another example, we can see that the result deblurred by our method has sharp textures with less reflection contamination, while IPLNet (Hu et al, 2020) alters the contrast of the input image excessively and generates fake textures that originally did not exist, and other compared methods fail to produce sharp textures.

6.5 Computational complexity analysis

To evaluate the computational complexity, we compare the number of parameters, FLOPs, memory, and inference time of different methods on our synthetic test dataset using a single NVIDIA 4090 GPU, as shown in Table 4. Here, the size of the input images is 512×512 pixels (3 color channels), and the inference time is the time taken to reconstruct all 300 scenes. Besides, since the methods designed for handling the unpolarized image deblurring tasks (MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), and MISCFILTER (Liu et al, 2024)) cannot simultaneously process four polarized images, the total time for processing four polarized images is taken as the inference time. From the results we can see that the computational complexity of our method is not very high, and it can realize real-time deblurring of polarized images.

7 Conclusion

We present a learning-based solution to deblur multiple polarized images for acquiring the DoLP and AoLP accurately. Specifically, we propose a polarized image deblurring pipeline to solve the problem in a polarization-aware manner by decomposing it into two less-ill-posed sub-problems, and build a two-stage neural network tailored to the pipeline to handle the two sub-problems respectively. Experimental results not only show that our method achieves state-of-the-art performance on both synthetic and real-world images, but also

⁵ Additional results can be found in the supplementary material.

demonstrate that using our method to deblur polarized images can improve the performance of polarization-based vision applications such as reflection removal and image dehazing.

Limitations. Since our method is specially designed for deblurring multiple polarized images captured by a polarization camera in a snapshot, it cannot be used to deblur videos since it does not explicitly consider the temporal information encoded in consecutive frames. Besides, since the formation of motion blur could be different from other types of degeneration (such as haze, rain, and noise), our method, which is based on the observations about motion blur, may not generalize to other low-level tasks.

Discussions. In the future, we aim to extend our work to address the polarized video deblurring problem by developing new modules that can better explore and utilize temporal connections. Additionally, we plan to generalize our approach to handle other types of image degradation by integrating comprehensive observations about polarized images, ultimately working towards a unified framework for enhancing polarized images.

Acknowledgement

This work was supported in part by National Natural Science Foundation of China under Grant No. 62136001, 62088102, 62276007, Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012), and JST-Mirai Program Grant Number JPMJM123G1.

References

- Abubakar A, Zhao X, Li S, Takruri M, Bastaki E, Bermak A (2018) A block-matching and 3-D filtering algorithm for Gaussian noise in DoFP polarization images. *IEEE Sensors Journal* 18(18):7429–7435
- Aittala M, Durand F (2018) Burst image deblurring using permutation invariant convolutional neural networks. In: *Proc. of European Conference on Computer Vision*, pp 731–747
- Bar L, Berkels B, Rumpf M, Sapiro G (2007) A variational framework for simultaneous motion estimation and restoration of motion-blurred video. In: *Proc. of International Conference on Computer Vision*, pp 1–8
- Boracchi G, Foi A (2012) Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing* 21(8):3502–3517
- Chakrabarti A (2016) A neural approach to blind motion deblurring. In: *Proc. of European Conference on Computer Vision*, pp 221–235
- Chefer H, Benaim S, Paiss R, Wolf L (2022) Image-based CLIP-guided essence transfer. In: *Proc. of European Conference on Computer Vision*, pp 695–711
- Chen H, Gu J, Gallo O, Liu MY, Veeraraghavan A, Kautz J (2018) Reblur2Deblur: Deblurring videos via self-supervised learning. In: *Proc. of International Conference on Computational Photography*, pp 1–9
- Chen L, Fang F, Lei S, Li F, Zhang G (2020) Enhanced sparse model for blind deblurring. In: *Proc. of European Conference on Computer Vision*, pp 631–646
- Chen L, Zhang J, Lin S, Fang F, Ren JS (2021a) Blind deblurring for saturated images. In: *Proc. of Computer Vision and Pattern Recognition*, pp 6308–6316
- Chen L, Zhang J, Pan J, Lin S, Fang F, Ren JS (2021b) Learning a non-blind deblurring network for night blurry images. In: *Proc. of Computer Vision and Pattern Recognition*, pp 10542–10550
- Chen L, Chu X, Zhang X, Sun J (2022) Simple baselines for image restoration. In: *Proc. of European Conference on Computer Vision*, pp 17–33
- Chi Z, Wang Y, Yu Y, Tang J (2021) Test-time fast adaptation for dynamic scene deblurring via meta-auxiliary learning. In: *Proc. of Computer Vision and Pattern Recognition*, pp 9137–9146
- Cho SJ, Ji SW, Hong JP, Jung SW, Ko SJ (2021) Rethinking coarse-to-fine approach in single image deblurring. In: *Proc. of International Conference on Computer Vision*, pp 4641–4650
- Chu X, Chen L, Chen C, Lu X (2022) Improving image restoration by revisiting global information aggregation. In: *Proc. of European Conference on Computer Vision*, pp 53–71
- Deschaintre V, Lin Y, Ghosh A (2021) Deep polarization imaging for 3D shape and SVBRDF acquisition. In: *Proc. of Computer Vision and Pattern Recognition*
- Dong J, Roth S, Schiele B (2021) Learning spatially-variant MAP models for non-blind image deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 4886–4895
- Eboli T, Sun J, Ponce J (2020) End-to-end interpretable learning of non-blind image deblurring. In: *Proc. of European Conference on Computer Vision*, Springer, pp 314–331
- Fergus R, Singh B, Hertzmann A, Roweis ST, Freeman WT (2006) Removing camera shake from a single photograph. In: *Proc. of ACM SIGGRAPH*, pp 787–794
- Gao H, Tao X, Shen X, Jia J (2019) Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: *Proc. of Computer Vision and Pattern Recognition*, pp 3848–3856
- Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: *Proc. of International Conference on Artificial Intelligence and Statistics*, pp 249–256
- Gong D, Yang J, Liu L, Zhang Y, Reid I, Shen C, Van Den Hengel A, Shi Q (2017) From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In: *Proc. of Computer Vision and Pattern Recognition*, pp 2319–2328
- Guo Y, Dai X, Wang S, Jin G, Zhang X (2024) Attention-based progressive discrimination generative adversarial networks for polarimetric image demosaicing. *IEEE Transactions on Computational Imaging*
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proc. of Computer Vision and Pattern Recognition*
- Hecht E, et al (2002) *Optics*, vol 5. Addison Wesley San Francisco
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
- Hu H, Lin Y, Li X, Qi P, Liu T (2020) IPLNet: A neural network for intensity-polarization imaging in low light. *Optics Letters* 45(22):6162–6165
- Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: *Proc. of Computer Vision and Pattern Recognition*, pp 7132–7141
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *Proc. of Computer Vision and Pattern Recognition*
- Hyun Kim T, Mu Lee K, Scholkopf B, Hirsch M (2017) Online video deblurring via dynamic temporal blending network. In: *Proc. of International Conference on Computer Vision*, pp 4038–4047
- Ji B, Yao A (2022) Multi-scale memory-based video deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 1919–1928
- Ji SW, Lee J, Kim SW, Hong JP, Baek SJ, Jung SW, Ko SJ (2022) XYDeblur: Divide and conquer for single image deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 17421–17430
- Jiang B, Xie Z, Xia Z, Li S, Liu S (2022) ERDN: Equivalent receptive field deformable network for video deblurring. In: *Proc. of European*

- Conference on Computer Vision, pp 663–678
- Kaufman A, Fattal R (2020) Deblurring using analysis-synthesis networks pair. In: Proc. of Computer Vision and Pattern Recognition, pp 5811–5820
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980
- Kong L, Dong J, Ge J, Li M, Pan J (2023) Efficient frequency domain-based transformers for high-quality image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 5886–5895
- Können G (1985) Polarized light in nature. CUP Archive
- Kruse J, Rother C, Schmidt U (2017) Learning to push the limits of efficient FFT-based image deconvolution. In: Proc. of International Conference on Computer Vision, pp 4586–4594
- Kumar R, Tyo JS, Ratliff BM (2007) Motion-based nonuniformity correction in DoFP polarimeters. In: Polarization Science and Remote Sensing III, vol 6682, pp 368–376
- Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) DeblurGAN: Blind motion deblurring using conditional adversarial networks. In: Proc. of Computer Vision and Pattern Recognition, pp 8183–8192
- Kupyn O, Martyniuk T, Wu J, Wang Z (2019) DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In: Proc. of International Conference on Computer Vision, pp 8878–8887
- Lei C, Huang X, Zhang M, Yan Q, Sun W, Chen Q (2020) Polarized reflection removal with perfect alignment in the wild. In: Proc. of Computer Vision and Pattern Recognition, pp 1750–1758
- Li D, Xu C, Zhang K, Yu X, Zhong Y, Ren W, Suominen H, Li H (2021) ARVo: Learning all-range volumetric correspondence for video deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 7721–7731
- Li D, Zhang Y, Cheung KC, Wang X, Qin H, Li H (2022a) Learning degradation representations for image deblurring. In: Proc. of European Conference on Computer Vision, pp 736–753
- Li N, Zhao Y, Pan Q, Kong SG, Chan JCW (2020a) Full-time monocular road detection using zero-distribution prior of angle of polarization. In: Proc. of European Conference on Computer Vision, pp 457–473
- Li X, Li H, Lin Y, Guo J, Yang J, Yue H, Li K, Li C, Cheng Z, Hu H, et al (2020b) Learning-based denoising for polarimetric images. Optics Express 28(11):16309–16321
- Li Z, Jiang H, Zheng Y (2022b) Polarized color image denoising using Pocoformer. arXiv preprint arXiv:2207.00215
- Li Z, Jiang H, Cao M, Zheng Y (2023) Polarized color image denoising. In: Proc. of Computer Vision and Pattern Recognition, pp 9873–9882
- Liang JA, Guo Yf, Liu B (2022) BM3D-based denoising method for color polarization filter array. Optics Express 30(12):22107–22122
- Liu C, Wang X, Xu X, Tian R, Li S, Qian X, Yang MH (2024) Motion-adaptive separable collaborative filters for blind motion deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 25595–25605
- Liu X, Li X, Chen SC (2022) Enhanced polarization demosaicking network via a precise angle of polarization loss calculation method. Optics Letters 47(5):1065–1069
- Lyu Y, Cui Z, Li S, Pollefeys M, Shi B (2019) Reflection separation using a pair of unpolarized and polarized images. In: Proc. of Advances in Neural Information Processing Systems
- Lyu Y, Cui Z, Li S, Pollefeys M, Shi B (2022) Physics-guided reflection separation from a pair of unpolarized and polarized images. IEEE Transactions on Pattern Analysis and Machine Intelligence
- Ma T, Zhang L, Sun B, Fan C (2024a) Multi-scale decomposition dehazing with polarimetric vision. In: International Conference on Multimedia Modeling, pp 112–126
- Ma T, Zhou J, Zhang L, Fan C, Sun B, Xue R (2024b) Image dehazing with polarization boundary constraints of transmission. IEEE Sensors Journal
- Mei H, Dong B, Dong W, Yang J, Baek SH, Heide F, Peers P, Wei X, Yang X (2022) Glass segmentation using intensity and spectral polarization cues. In: Proc. of Computer Vision and Pattern Recognition, pp 12622–12631
- Mou C, Wang Q, Zhang J (2022) Deep generalized unfolding networks for image restoration. In: Proc. of Computer Vision and Pattern Recognition, pp 17399–17410
- Nah S, Hyun Kim T, Mu Lee K (2017) Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 3883–3891
- Nah S, Son S, Lee KM (2019) Recurrent neural networks with intra-frame iterations for video deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 8102–8111
- Ono T, Kondo Y, Sun L, Kurita T, Moriuchi Y (2022) Degree-of-linear-polarization-based color constancy. In: Proc. of Computer Vision and Pattern Recognition, pp 19740–19749
- Pan J, Hu Z, Su Z, Yang MH (2016) L_0 -regularized intensity and gradient prior for deblurring text images and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(2):342–355
- Pan J, Bai H, Tang J (2020) Cascaded deep video deblurring using temporal sharpness prior. In: Proc. of Computer Vision and Pattern Recognition, pp 3043–3051
- Pistellato M, Bergamasco F, Fatima T, Torsello A (2022) Deep demosaicing for polarimetric filter array cameras. IEEE Transactions on Image Processing 31:2017–2026
- Qiu S, Fu Q, Wang C, Heidrich W (2019) Polarization demosaicking for monochrome and color polarization focal plane arrays. In: Vision, Modeling and Visualization
- Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, et al (2021) Learning transferable visual models from natural language supervision. In: Proc. of International Conference on Machine Learning, pp 8748–8763
- Ren D, Zhang K, Wang Q, Hu Q, Zuo W (2020) Neural blind deconvolution using deep priors. In: Proc. of Computer Vision and Pattern Recognition, pp 3341–3350
- Ren M, Delbracio M, Talebi H, Gerig G, Milanfar P (2023) Multi-scale structure guided diffusion for image deblurring. In: Proc. of International Conference on Computer Vision, pp 10721–10733
- Ren W, Pan J, Cao X, Yang MH (2017) Video deblurring via semantic segmentation and pixel-wise non-linear kernel. In: Proc. of International Conference on Computer Vision, pp 1077–1085
- Rim J, Lee H, Won J, Cho S (2020) Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Proc. of European Conference on Computer Vision
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. International Journal of Computer Vision 115(3):211–252
- Shen Z, Wang W, Lu X, Shen J, Ling H, Xu T, Shao L (2019) Human-aware motion deblurring. In: Proc. of International Conference on Computer Vision, pp 5572–5581
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
- Su S, Delbracio M, Wang J, Sapiro G, Heidrich W, Wang O (2017) Deep video deblurring for hand-held cameras. In: Proc. of Computer Vision and Pattern Recognition, pp 1279–1288
- Suin M, Rajagopalan A (2021) Gated spatio-temporal attention-guided video deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 7802–7811
- Suin M, Purohit K, Rajagopalan A (2020) Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 3606–3615
- Sun J, Cao W, Xu Z, Ponce J (2015) Learning a convolutional neural network for non-uniform motion blur removal. In: Proc. of Computer Vision and Pattern Recognition, pp 769–777
- Tao X, Gao H, Shen X, Wang J, Jia J (2018) Scale-recurrent network for deep image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 8174–8182

- Tibbs AB, Daly IM, Roberts NW, Bull DR (2018) Denoising imaging polarimetry by adapted bm3d method. *Journal of the Optical Society of America* 35(4):690–701
- Ting J, Wu X, Hu K, Zhang H (2021) Deep snapshot HDR reconstruction based on the polarization camera. In: *Proc. of International Conference on Image Processing*
- Tran P, Tran AT, Phung Q, Hoai M (2021) Explore image deblurring via encoded blur kernel space. In: *Proc. of Computer Vision and Pattern Recognition*, pp 11956–11965
- Tsai FJ, Peng YT, Lin YY, Tsai CC, Lin CW (2022) Stripformer: Strip transformer for fast image deblurring. In: *Proc. of European Conference on Computer Vision*, pp 146–162
- Tu Z, Talebi H, Zhang H, Yang F, Milanfar P, Bovik A, Li Y (2022) MAXIM: Multi-axis MLP for image processing. In: *Proc. of Computer Vision and Pattern Recognition*, pp 5769–5780
- Tyo JS, LaCasse CF, Ratliff BM (2009) Total elimination of sampling errors in polarization imagery obtained with integrated microgrid polarimeters. *Optics Letters* 34(20):3187–3189
- Ulyanov D, Vedaldi A, Lempitsky V (2016) Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:160708022*
- Wang J, Chan KC, Loy CC (2023) Exploring clip for assessing the look and feel of images. In: *Proc. of the AAAI Conference on Artificial Intelligence*, pp 2555–2563
- Wang Y, Lu Y, Gao Y, Wang L, Zhong Z, Zheng Y, Yamashita A (2022a) Efficient video deblurring guided by motion magnitude. In: *Proc. of European Conference on Computer Vision*, pp 413–429
- Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H (2022b) Uformer: A general U-shaped transformer for image restoration. In: *Proc. of Computer Vision and Pattern Recognition*, pp 17683–17693
- Wen S, Zheng Y, Lu F, Zhao Q (2019) Convolutional demosaicing network for joint chromatic and polarimetric imagery. *Optics Letters* 44(22):5646–5649
- Wen S, Zheng Y, Lu F (2021a) A sparse representation based joint demosaicing method for single-chip polarized color sensor. *IEEE Transactions on Image Processing* 30:4171–4182
- Wen Y, Chen J, Sheng B, Chen Z, Li P, Tan P, Lee TY (2021b) Structure-aware motion deblurring using multi-adversarial optimized CycleGAN. *IEEE Transactions on Image Processing* 30:6142–6155
- Whang J, Delbracio M, Talebi H, Saharia C, Dimakis AG, Milanfar P (2022) Deblurring via stochastic refinement. In: *Proc. of Computer Vision and Pattern Recognition*, pp 16293–16303
- Whyte O, Sivic J, Zisserman A, Ponce J (2012) Non-uniform deblurring for shaken images. *International Journal of Computer Vision* 98(2):168–186
- Wieschollek P, Hirsch M, Scholkopf B, Lensch H (2017a) Learning blind motion deblurring. In: *Proc. of International Conference on Computer Vision*, pp 231–240
- Wieschollek P, Scholkopf B, Lensch HP, Hirsch M (2017b) End-to-end learning for image burst deblurring. In: *Proc. of Asian Conference on Computer Vision*, pp 35–51
- Wu X, Zhang H, Hu X, Shakeri M, Fan C, Ting J (2020) HDR reconstruction based on the polarization camera. *IEEE Robotics and Automation Letters* 5(4):5113–5119
- Wulff J, Black MJ (2014) Modeling blurred video with layers. In: *Proc. of European Conference on Computer Vision*, pp 236–252
- Xu L, Zheng S, Jia J (2013) Unnatural L_0 sparse representation for natural image deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 1107–1114
- Xu X, Wan M, Ge J, Chen H, Zhu X, Zhang X, Chen Q, Gu G (2022) ColorPolarNet: Residual dense network-based chromatic intensity-polarization imaging in low-light environment. *IEEE Transactions on Instrumentation and Measurement* 71:1–10
- Yang L, Ji H (2019) A variational EM framework with adaptive edge selection for blind motion deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 10167–10176
- Yi Y, Zhang P, Chen Z, Zhang H, Luo Z, Zhang G, Li W, Zhao Y (2024) A demosaicing method based on an inter-channel correlation model for DoFP polarimeter. *Optics and Lasers in Engineering* 181:108388
- Yuan Y, Su W, Ma D (2020) Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training. In: *Proc. of Computer Vision and Pattern Recognition*, pp 3555–3564
- Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH (2022) Restormer: Efficient transformer for high-resolution image restoration. In: *Proc. of Computer Vision and Pattern Recognition*, pp 5728–5739
- Zhang H, Dai Y, Li H, Koniusz P (2019) Deep stacked hierarchical multi-patch network for image deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 5978–5986
- Zhang H, Xie H, Yao H (2022a) Spatio-temporal deformable attention network for video deblurring. In: *Proc. of European Conference on Computer Vision*, pp 581–596
- Zhang J, Luo H, Liang R, Zhou W, Hui B, Chang Z (2017) PCA-based denoising method for division of focal plane polarimeters. *Optics Express* 25(3):2391–2400
- Zhang J, Pan J, Ren J, Song Y, Bao L, Lau RW, Yang MH (2018a) Dynamic scene deblurring using spatially variant recurrent neural networks. In: *Proc. of Computer Vision and Pattern Recognition*, pp 2521–2529
- Zhang K, Gool LV, Timofte R (2020a) Deep unfolding network for image super-resolution. In: *Proc. of Computer Vision and Pattern Recognition*, pp 3217–3226
- Zhang K, Luo W, Zhong Y, Ma L, Stenger B, Liu W, Li H (2020b) Deblurring by realistic blurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 2737–2746
- Zhang K, Ren W, Luo W, Lai WS, Stenger B, Yang MH, Li H (2022b) Deep image deblurring: A survey. *International Journal of Computer Vision* 130(9):2103–2130
- Zhang M, Fang Y, Ni G, Zeng T (2022c) Pixel screening based intermediate correction for blind deblurring. In: *Proc. of Computer Vision and Pattern Recognition*, pp 5892–5900
- Zhang R, Isola P, Efros AA, Shechtman E, Wang O (2018b) The unreasonable effectiveness of deep features as a perceptual metric. In: *Proc. of Computer Vision and Pattern Recognition*
- Zhong Z, Gao Y, Zheng Y, Zheng B (2020) Efficient spatio-temporal recurrent neural network for video deblurring. In: *Proc. of European Conference on Computer Vision*, pp 191–207
- Zhou C, Teng M, Han Y, Xu C, Shi B (2021) Learning to dehaze with polarization. In: *Proc. of Advances in Neural Information Processing Systems*
- Zhou C, Han Y, Teng M, Han J, Li S, Xu C, Shi B (2023a) Polarization guided HDR reconstruction via pixel-wise depolarization. *IEEE Transactions on Image Processing* 32:1774–1787
- Zhou C, Teng M, Lyu Y, Li S, Xu C, Shi B (2023b) Polarization-aware low-light image enhancement. In: *Proc. of the AAAI Conference on Artificial Intelligence*, pp 3742–3750
- Zhou S, Zhang J, Pan J, Xie H, Zuo W, Ren J (2019) Spatio-temporal filter adaptive network for video deblurring. In: *Proc. of International Conference on Computer Vision*, pp 2482–2491
- Zhu C, Dong H, Pan J, Liang B, Huang Y, Fu L, Wang F (2022) Deep recurrent neural network with multi-scale bi-directional propagation for video deblurring. In: *Proc. of the AAAI Conference on Artificial Intelligence*, pp 3598–3607

Supplementary Material: Learning to Deblur Polarized Images

Chu Zhou^{1†} · Minggui Teng^{2,3} · Xinyu Zhou⁴ · Chao Xu⁴ · Imari Sato¹ ·
Boxin Shi^{2,3}

Received: date / Accepted: date

8 Details of the proposed new blocks

In this section, we provide the details of the proposed new blocks (including Stokes guided skip connection (SGSC) and multiscale context transferring encoder (MCTE)), as shown in Fig. 6.

9 Additional results on synthetic data

In this section, we provide additional visual quality comparisons on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024), as shown in Fig. 7, Fig. 8, and Fig. 9, corresponding to Footnote 3 of the paper.

10 Additional results on real data

In this section, we provide additional visual quality comparisons on real data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024),

as shown in Fig. 10, Fig. 11, and Fig. 12, corresponding to Footnote 4 of the paper.

11 Additional results of polarization-based vision applications

In this section, we provide additional results of polarization-based vision applications (including image dehazing (Zhou et al, 2021) and reflection removal (Lyu et al, 2019, 2022)) without deblurring and deblurred by our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024), as shown in Fig. 13 and Fig. 14 respectively, corresponding to Footnote 5 of the paper.

References

- Cho SJ, Ji SW, Hong JP, Jung SW, Ko SJ (2021) Rethinking coarse-to-fine approach in single image deblurring. In: Proc. of International Conference on Computer Vision, pp 4641–4650
- Hu H, Lin Y, Li X, Qi P, Liu T (2020) IPLNet: A neural network for intensity-polarization imaging in low light. Optics Letters 45(22):6162–6165
- Ji SW, Lee J, Kim SW, Hong JP, Baek SJ, Jung SW, Ko SJ (2022) XYDeblur: Divide and conquer for single image deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 17421–17430
- Liu C, Wang X, Xu X, Tian R, Li S, Qian X, Yang MH (2024) Motion-adaptive separable collaborative filters for blind motion deblurring. In: Proc. of Computer Vision and Pattern Recognition, pp 25595–25605
- Lyu Y, Cui Z, Li S, Pollefeys M, Shi B (2019) Reflection separation using a pair of unpolarized and polarized images. In: Proc. of Advances in Neural Information Processing Systems
- Lyu Y, Cui Z, Li S, Pollefeys M, Shi B (2022) Physics-guided reflection separation from a pair of unpolarized and polarized images. IEEE Transactions on Pattern Analysis and Machine Intelligence

✉ Boxin Shi

E-mail: shiboxin@pku.edu.cn

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

1 National Institute of Informatics, Japan

2 National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, China

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

4 National Key Laboratory of General AI, School of Intelligence Science and Technology, Peking University, China

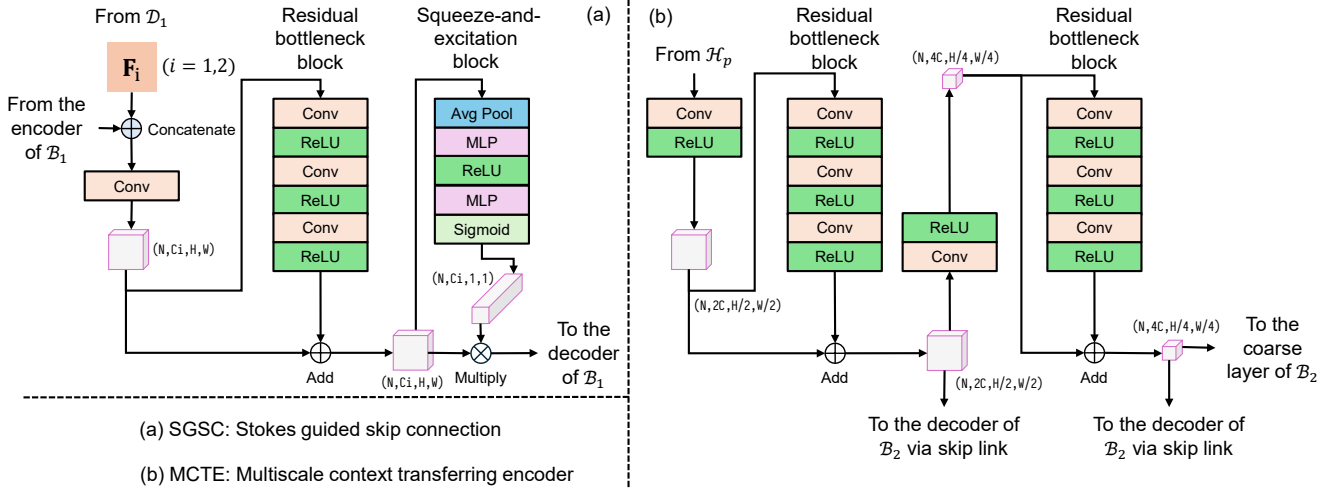


Fig. 6 The details of the proposed Stokes guided skip connection (SGSC) and multiscale context transferring encoder (MCTE) blocks.

- Xu X, Wan M, Ge J, Chen H, Zhu X, Zhang X, Chen Q, Gu G (2022) ColorPolarNet: Residual dense network-based chromatic intensity-polarization imaging in low-light environment. *IEEE Transactions on Instrumentation and Measurement* 71:1–10
- Zhang H, Xie H, Yao H (2022) Spatio-temporal deformable attention network for video deblurring. In: *Proc. of European Conference on Computer Vision*, pp 581–596
- Zhou C, Teng M, Han Y, Xu C, 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, Shi B (2023) Polarization-aware low-light image enhancement. In: *Proc. of the AAAI Conference on Artificial Intelligence*, pp 3742–3750

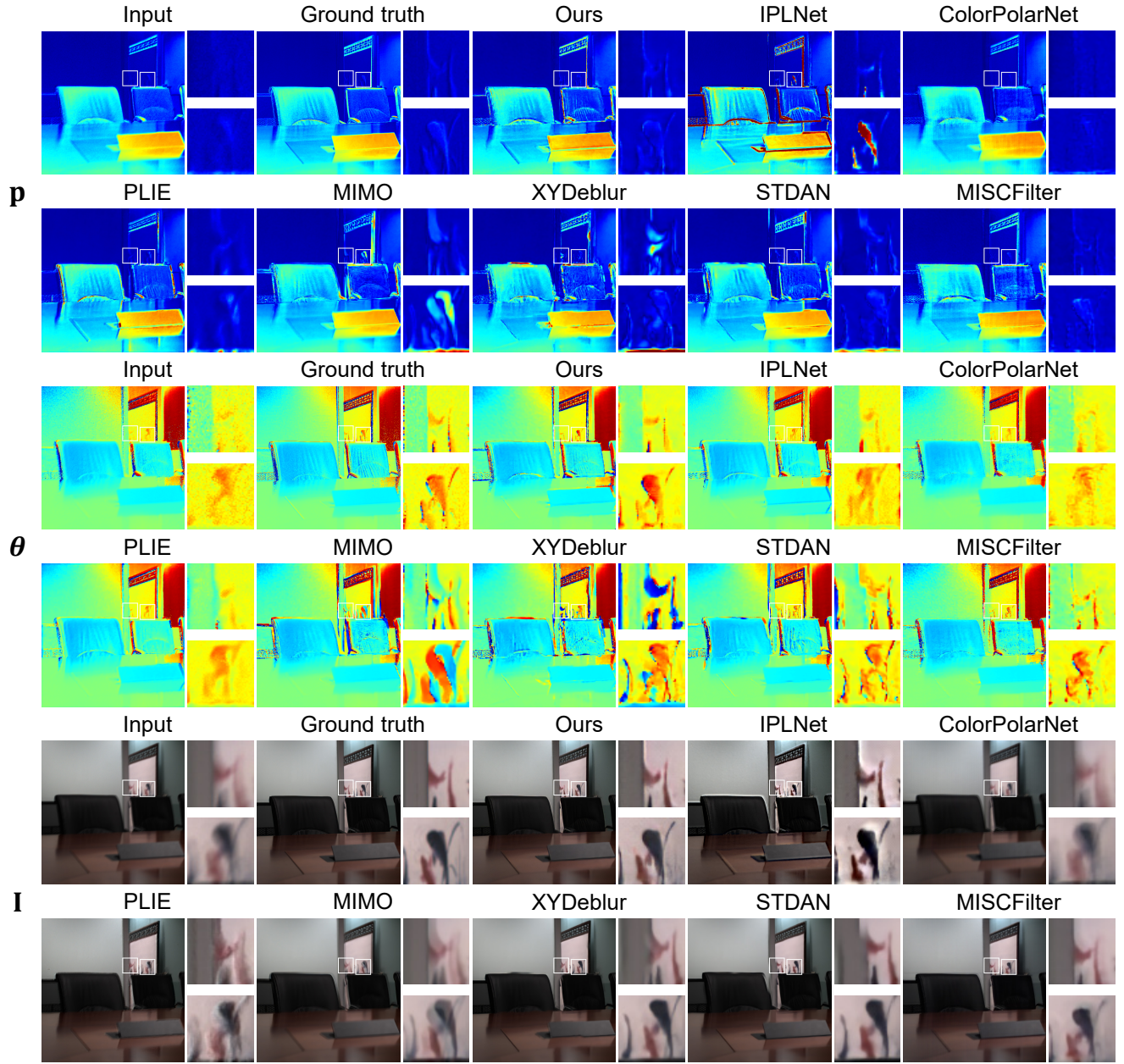


Fig. 7 Additional qualitative comparisons on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFILTER (Liu et al, 2024) (part 1). The DoP p and AoP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023)). Please zoom-in for better details.

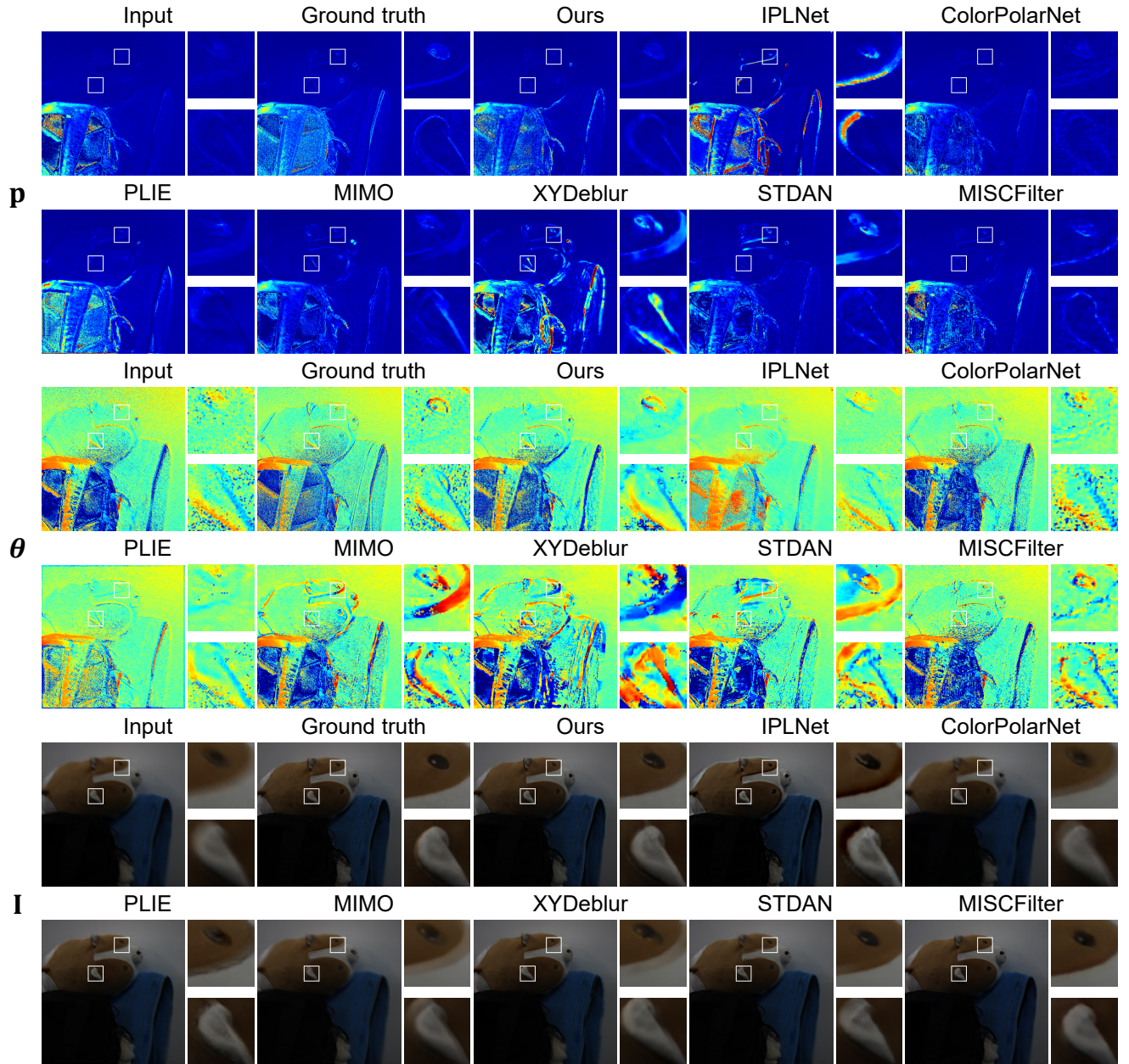


Fig. 8 Additional qualitative comparisons on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024) (part 2). The DoP $\mathbf{p}$ and AoP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023)). Please zoom-in for better details.

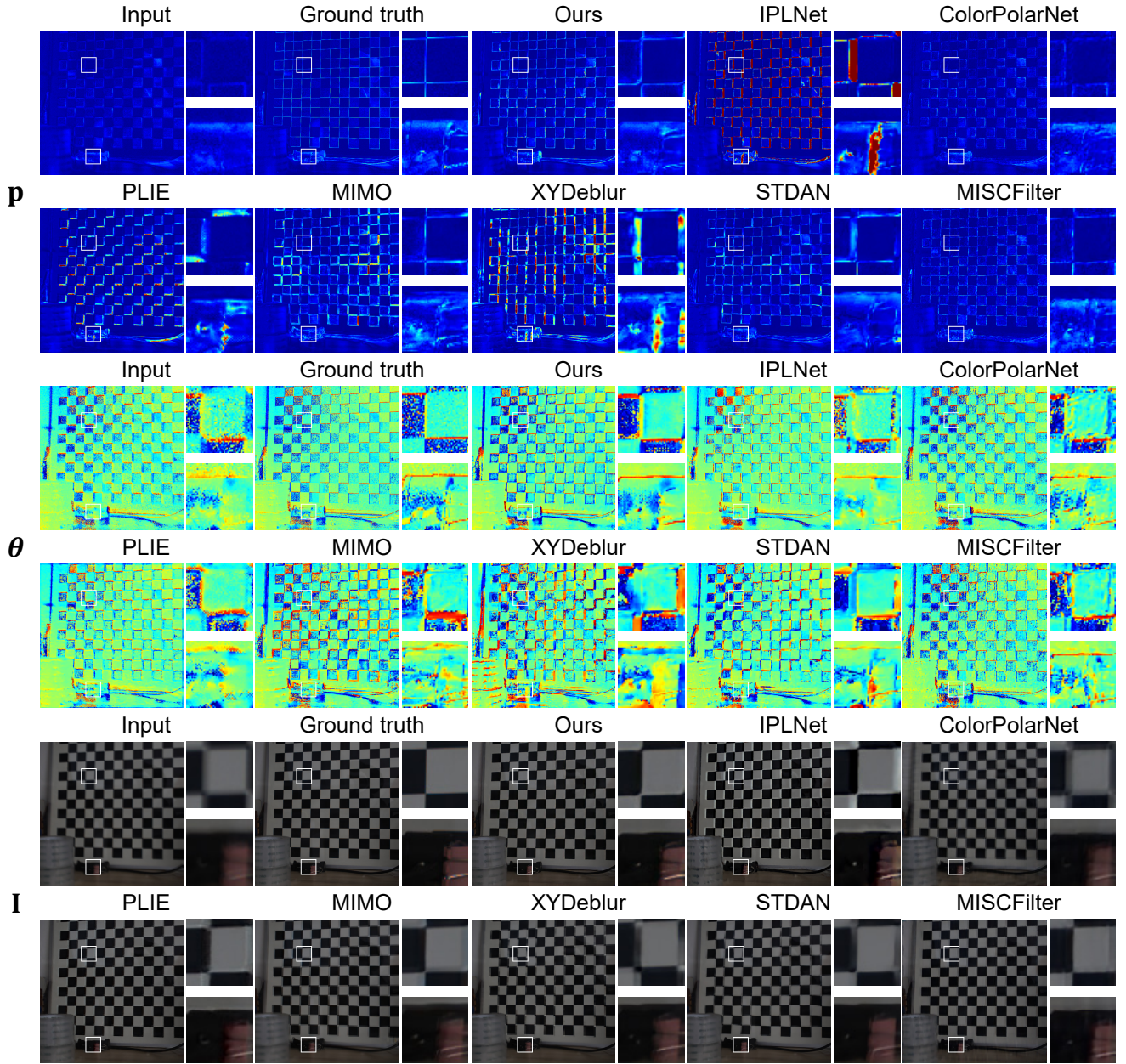


Fig. 9 Additional qualitative comparisons on synthetic data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFILTER (Liu et al, 2024) (part 3). The DoP $\mathbf{p}$ and AoP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023)). Please zoom-in for better details.

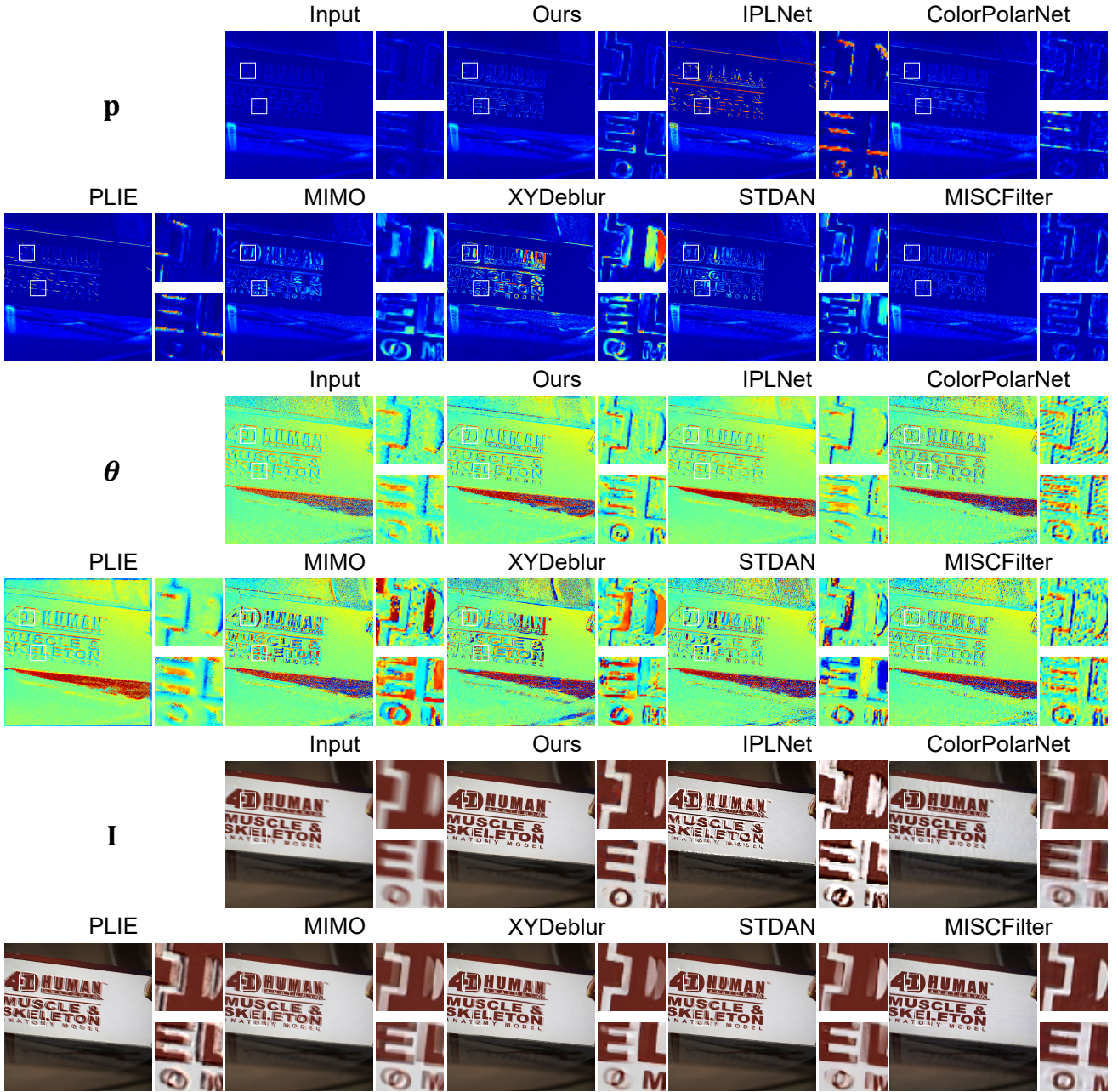


Fig. 10 Additional qualitative comparisons on real data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024) (part 1). The DoP p and AoP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023)). Please zoom-in for better details.

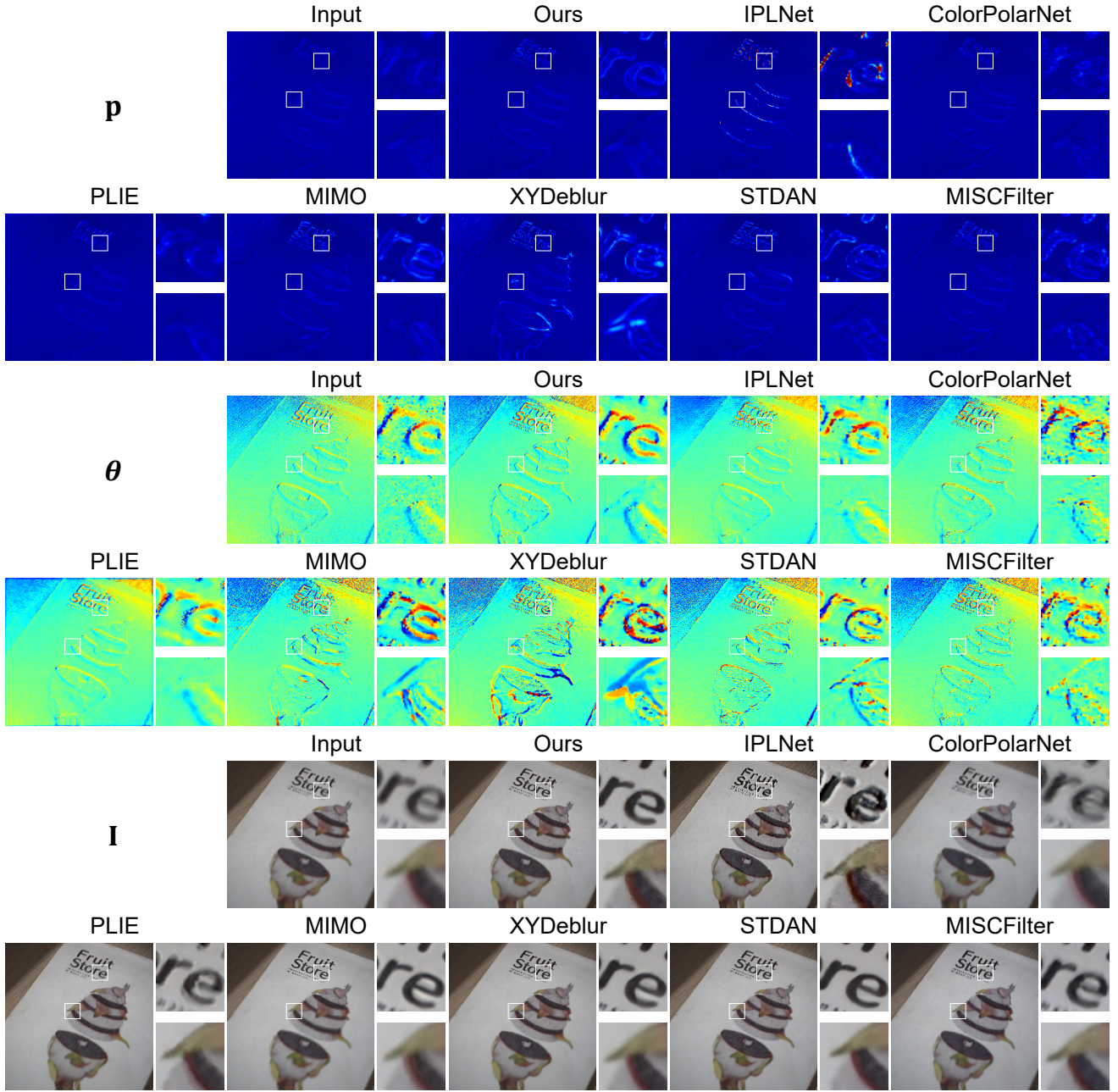


Fig. 11 Additional qualitative comparisons on real data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024) (part 2). The DoP p and AoP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023)). Please zoom-in for better details.

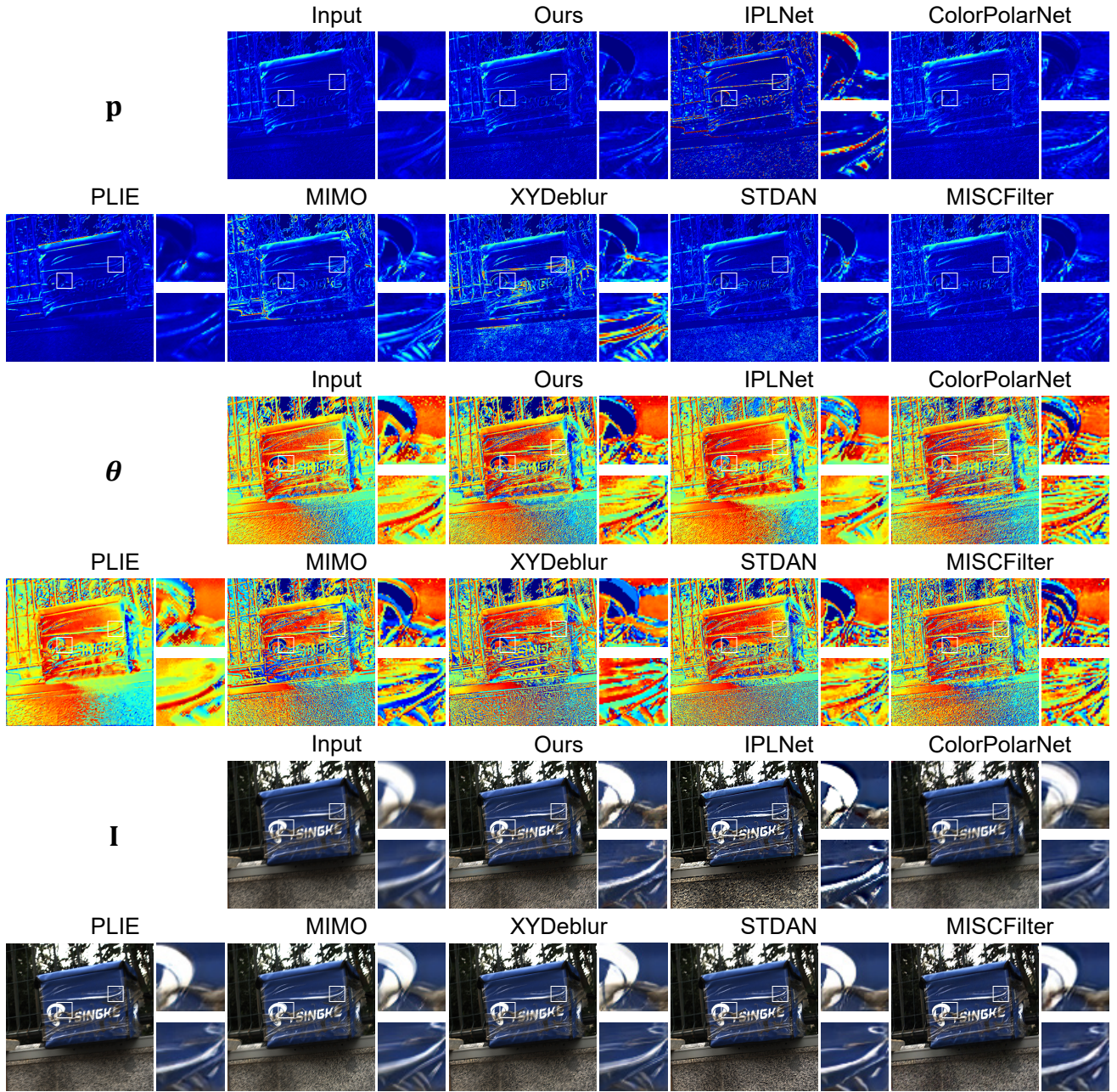


Fig. 12 Additional qualitative comparisons on real data among our method, IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFilter (Liu et al, 2024) (part 3). The DoP p and AoP θ are visualized using color maps after normalizing and averaging the RGB channels (as done in (Hu et al, 2020; Zhou et al, 2023)). Please zoom-in for better details.

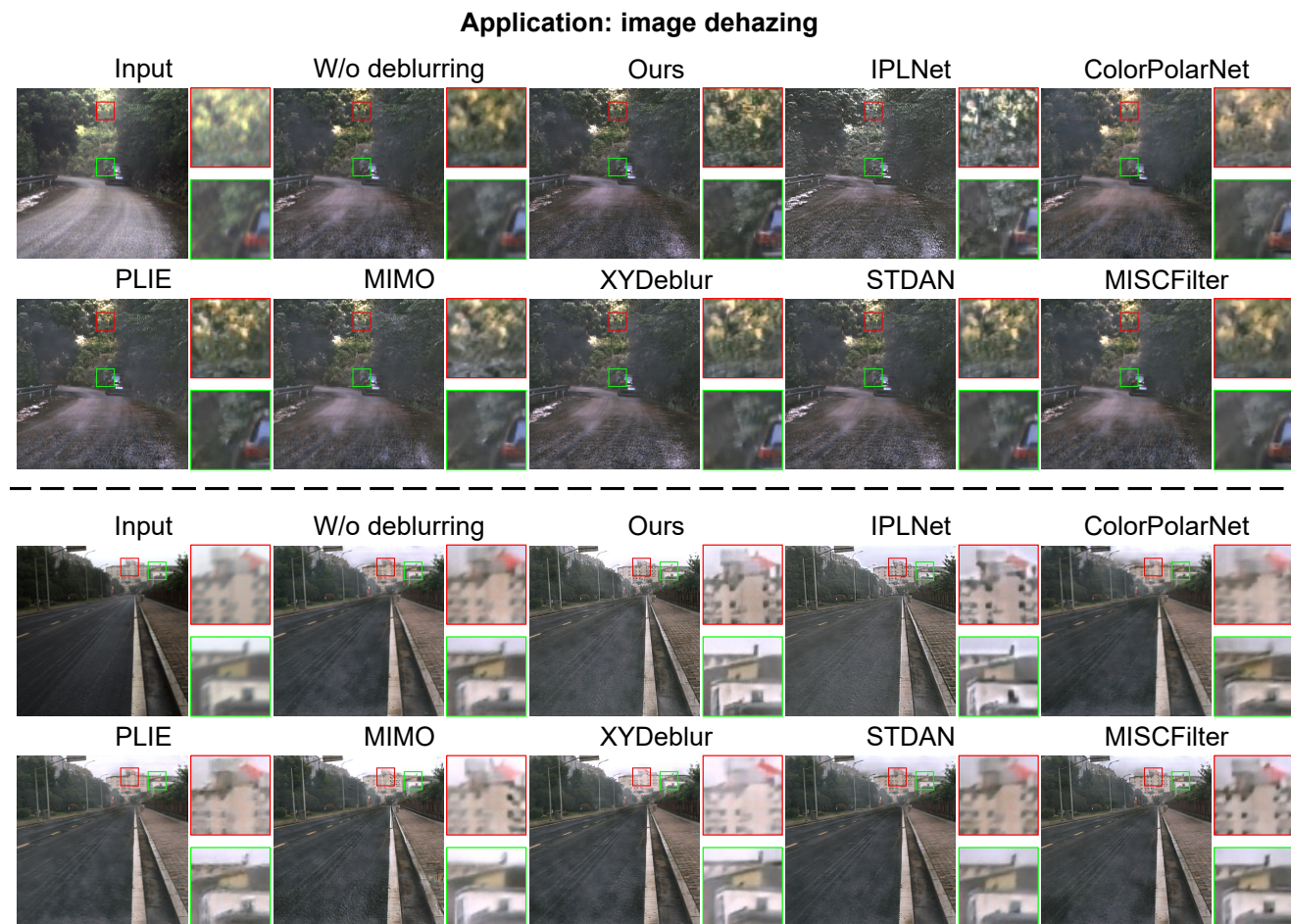


Fig. 13 Additional results of polarization-based vision applications (image dehazing (Zhou et al, 2021)) without deblurring and deblurred by our method and the compared methods (IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFILTER (Liu et al, 2024)). Please zoom-in for better details.

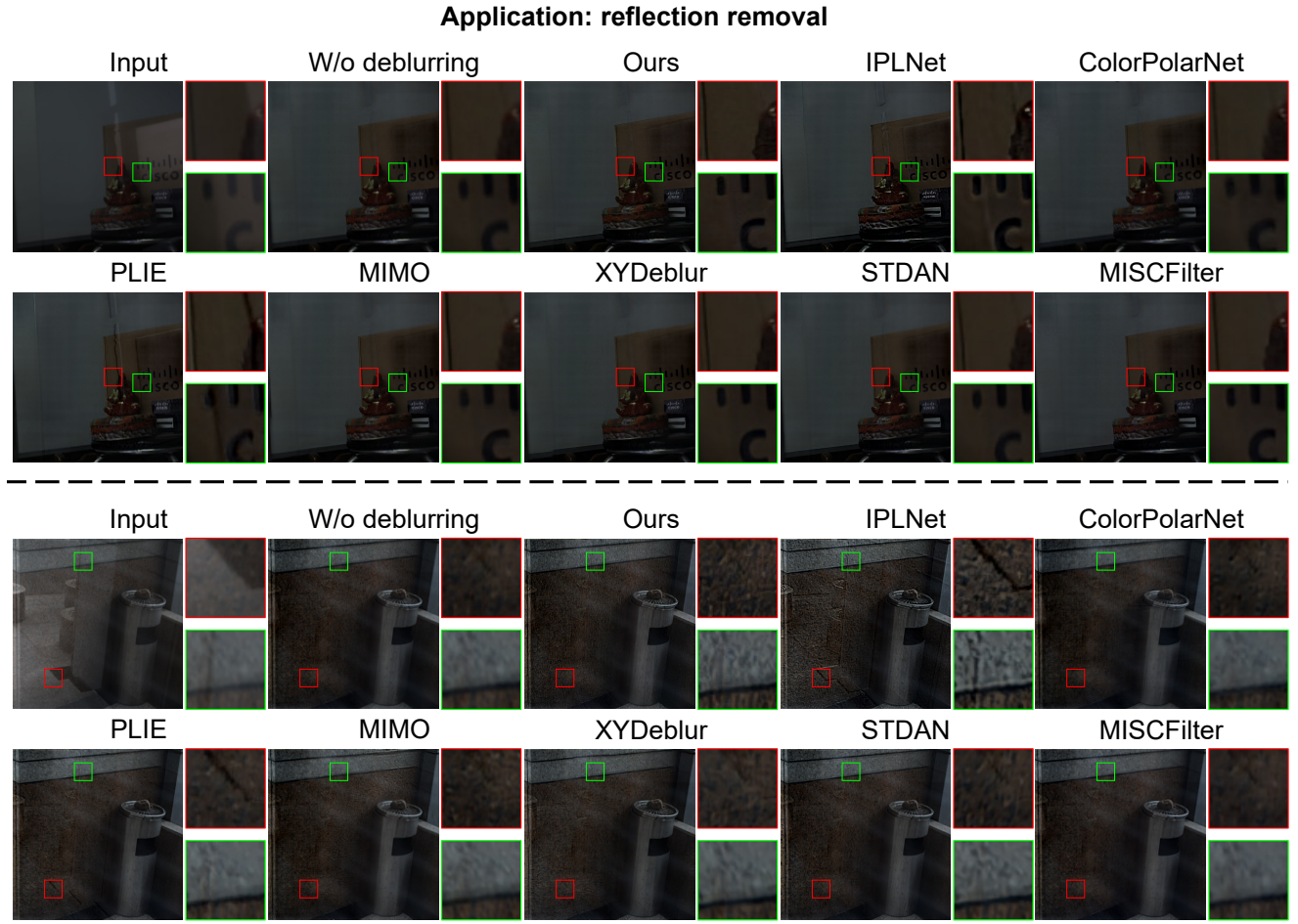


Fig. 14 Additional results of polarization-based vision applications (reflection removal (Lyu et al, 2019, 2022)) without deblurring and deblurred by our method and the compared methods (IPLNet (Hu et al, 2020), ColorPolarNet (Xu et al, 2022), PLIE (Zhou et al, 2023), MIMO (Cho et al, 2021), XYDeblur (Ji et al, 2022), STDAN (Zhang et al, 2022), and MISCFILTER (Liu et al, 2024)). Please zoom-in for better details.