Towards a Unified Complementary Fusion Framework for Robust Polarimetric Imaging

Chu Zhou*, Yixing Liu, Minggui Teng, Chao Xu, Boxin Shi*, *Senior Member, IEEE* and Imari Sato, *Member, IEEE*

***Abstract*—Polarization, as an intrinsic property of light alongside amplitude and phase, has demonstrated great potential in a variety of downstream applications by providing valuable physical cues encoded in the degree of polarization (DoP) and the angle of polarization (AoP). Polarimetric imaging aims to acquire these polarimetric parameters by capturing polarized snapshots. However, compared to conventional imaging, it faces greater difficulties due to the presence of polarizers, which attenuate light intensity in a spatially variant manner. Such attenuation complicates exposure control: a short exposure leads to low signal-to-noise ratio and color distortion, whereas a relatively long exposure increases the risk of motion blur and saturation. To address these challenges, this work proposes PolFusion+, a unified framework that robustly produces clean and sharp polarized snapshots by complementarily fusing a degraded pair of short-exposed noisy and long-exposed blurry inputs. Building upon a polarization-aware three-phase fusion scheme, PolFusion+ introduces two key advancements. First, to handle saturation in the blurry snapshot, the irradiance restoration phase extracts and rectifies color information from both inputs, effectively mitigating saturation-induced degradation. Second, to ensure physically faithful polarization reconstruction, the framework explicitly models the individual characteristics and interdependencies of the DoP and AoP, enabling their joint restoration. These improvements are supported by a degradation-oriented neural network tailored to the fusion scheme. Experimental results demonstrate that PolFusion+ achieves state-of-the-art performance, effectively benefiting downstream applications.**

***Index Terms*—Polarimetric imaging, polarization-based vision, deep learning**

I. Introduction

POLARIZATION is an intrinsic property of light, alongside amplitude and phase. It encodes valuable physical cues through polarimetric parameters such as the degree of polarization (DoP) and the angle of polarization (AoP). By fully exploiting these physical cues, polarization-based methods often outperform conventional image-based approaches, demonstrating great potential in a variety of downstream applications such as reflection removal [26], [35], [57], image dehazing [44], [69], and shape estimation [1], [3]. Since the effectiveness of these applications is closely tied to the reliability of the physical cues, it is of practical significance to obtain high-quality polarimetric parameters.

As conventional imaging approaches cannot directly capture polarimetric parameters, the concept of *polarimetric imaging* has been introduced to acquire them in an indirect manner by capturing polarized images. For instance, retrieving the DoP and AoP requires a minimum of three polarized images with different polarizer angles. While a rotating polarizer offers a way to achieve this, the necessity of multiple shots makes the capturing process quite cumbersome and the captured images vulnerable to pixel misalignment. In contrast, polarization cameras (*e.g.*, Lucid Vision Phoenix[1]), empowered by the division of focal plane (DoFP) technique, utilize the four-directional on-chip micro-polarizers to record all necessary polarization information in a single shot. The resulting set of four polarized images, each corresponding to a different polarizer angle ($0°$, $45°$, $90°$, and $135°$ respectively), is referred to as a *polarized snapshot*. The use of polarization cameras greatly simplifies the capturing process and eliminates pixel misalignment, bringing polarimetric imaging closer to the convenience of conventional imaging.

However, despite their convenience, polarization cameras still pose greater challenges than conventional cameras. This is primarily because polarizers attenuate light intensity in a spatially variant manner [66], making it difficult to determine an appropriate exposure time. As shown in Figure 1 (a), a short exposure time leads to a low signal-to-noise ratio (SNR) and color distortion, which negatively impacts the accuracy of the DoP and AoP. In contrast, a relatively long exposure time may improve the SNR but introduces a higher risk of motion blur and saturation, resulting in overly smoothed textures and numerical errors in the DoP and AoP, as shown in Figure 1 (b). To address these challenges, a number of post-processing approaches have been proposed to perform low-light enhancement [28], [56], [70] and deblurring [71] on the degraded polarized snapshots. These methods are designed to be polarization-aware, meaning that they can explicitly take the preservation of polarization properties into consideration, and therefore often achieve better performance than their counterparts developed for conventional images [5], [25], [48], [65]. However, due to the inherently ill-posed nature of the underlying tasks, the quality of the restored results remains limited.

Notably, we observe that different types of degraded polarized snapshots contain complementary information. Specifically, short-exposed snapshots tend to be sharp but noisy, while long-exposed ones are cleaner but often suffer from blur. This

* Corresponding authors: zhou_chu@hotmail.com and shiboxin@pku.edu.cn

Chu Zhou and Imari Sato are with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan.

Yixing Liu, Minggui Teng, and Boxin Shi are with the State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100080, China, and also with the National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing 100080, China.

Chao Xu is with the National Key Laboratory of General Artificial Intelligence, School of Intelligence Science and Technology, Peking University, Beijing 100080, China.

[1]https://thinklucid.com/product/phoenix-5-0-mp-polarized-model/

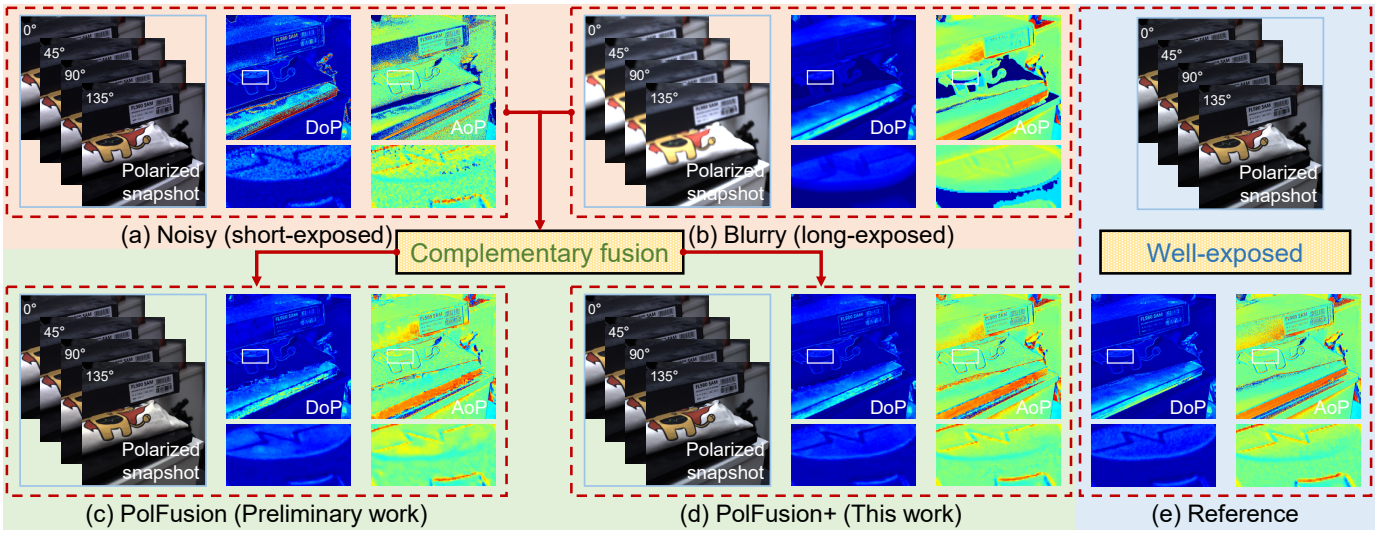


Fig. 1. In polarimetric imaging, the spatially variant light attenuation caused by polarizers makes it difficult to determine an appropriate exposure time. (a) A short exposure results in low SNR and color distortion, which degrades the accuracy of the DoP and AoP. (b) A longer exposure improves SNR but introduces motion blur and saturation, leading to over-smoothed textures and numerical errors in the DoP and AoP. (c) Our preliminary work PolFusion [67] introduced the first polarization-aware complementary fusion framework, achieving improved quality and property preservation, but still struggles under challenging conditions. (d) In this work, we propose PolFusion+, a unified complementary fusion framework that extends PolFusion [67] for robust polarimetric imaging. (e) A well-exposed polarized snapshot captured with a tripod, serving as the reference. Throughout this paper, for better visualization, the brightness of the noisy (short-exposed) polarized snapshots is manually aligned to the reference by normalization, and the DoP and AoP are visualized using color maps after normalizing and averaging the RGB channels.

observation motivates an intuitive strategy for high-quality polarimetric imaging by *fusing a noisy and blurry pair of polarized snapshots in a complementary manner*. Such a strategy can serve as a form of boosting, where multiple suboptimal inputs are combined to produce clean and sharp results. Moreover, it offers a way to make use of existing degraded polarized snapshots rather than discarding them. However, most existing fusion methods [4], [38], [63], [64] are designed for conventional images and are therefore not suitable for polarized snapshots, as they do not account for the preservation of polarization properties during the fusion process. This often leads to unreliable acquisition of the DoP and AoP. To deal with this issue, our preliminary work [67], referred to as *"PolFusion"*, was the first to propose a polarization-aware complementary fusion framework, which enables quality-improved and property-preserved polarimetric imaging. Nevertheless, its performance is limited by the assumption that the blurry snapshot is not impacted by overexposure, which weakens its robustness under real-world saturation. More importantly, it treats the DoP and AoP as independent quantities, whereas recent findings show that the polarimetric parameters are intrinsically coupled (*i.e.*, changes in one parameter inevitably influence the others) [73]. Figure 1 (c) presents an example where PolFusion [67] struggles to produce satisfying results.

In this work, to address the limitations of our preliminary work PolFusion [67], we extend it into a unified complementary fusion framework for robust polarimetric imaging, referred to as *PolFusion+*. While PolFusion+ builds upon and redesigns the overall architecture of PolFusion [67]—namely the polarization-aware three-phase fusion scheme (irradiance restoration, polarization reconstruction, and artifact suppression) and the degradation-oriented network tailored to this scheme—it introduces several key improvements to the execution of each fusion phase and the corresponding network modules. First, in the irradiance restoration phase, to handle the issue of saturation in the blurry snapshot, we propose to extract color information not only from the blurry snapshot but also from the noisy one, and accordingly adapt the network to rectify degraded color features caused by saturation. Second, in the polarization reconstruction phase, to ensure the accuracy of the DoP and AoP, we go beyond treating them as independent physical quantities; instead, we explicitly model their individual characteristics and interrelationships to enable joint restoration, which is further supported by the network design that facilitates such interaction. As illustrated in Figure 1 (d), PolFusion+ achieves more visually compelling and physically accurate results compared to PolFusion [67], showing its effectiveness under more challenging real-world conditions. To summarize, this work makes contributions by demonstrating:

- We extend our preliminary work PolFusion [67] into a unified polarimetric imaging framework, PolFusion+, which features a redesigned polarization-aware three-phase fusion scheme and a degradation-oriented network tailored to this scheme. Key improvements include:
- Improved irradiance restoration by extracting color information not only from the blurry snapshot but also from the noisy one, accompanied by network adaptations that effectively mitigate saturation-induced degradation.
- Enhanced polarization reconstruction by jointly modeling the individual characteristics and interdependencies of the DoP and AoP, supported by dedicated network modules that ensure physically faithful reconstruction.

Experimental results demonstrate that our PolFusion+ not only achieves state-of-the-art performance on both synthetic and

real data, but also effectively benefits downstream applications.

II. Related Work

A. Polarimetric Imaging

Most existing polarimetric imaging methods focus on polarized image demosaicing and super-resolution. Polarized image demosaicing aims to address the mosaic artifacts introduced by the color-polarization filter array (CPFA) used in DoFP polarization cameras. Based on numerical optimization, early methods [31], [33], [37], [42] incorporated handcrafted priors to suppress mosaic artifacts. To achieve higher performance, learning-based approaches have been explored, utilizing neural networks [10], [32], [39], [53], [61] and dictionary learning techniques [34], [52], [62]. Polarized image super-resolution aims to perform resolution enhancement while preserving the polarization properties. Hu *et al*. [16] designed a neural network for super-resolving monochrome polarized images. Yu *et al*. [59] extended this idea to color polarized images. Recently, Zhou *et al*. [74] proposed a complementary polarized image demosaicing and super-resolution framework. Despite their effectiveness, these methods are not applicable to tasks such as low-light enhancement or deblurring.

In addition, several polarimetric imaging methods have been specifically developed for low-light enhancement and deblurring. For low-light scenarios, IPLNet [15] and ColorPolarNet [56] employed residual dense blocks as the backbone to process multiple polarized noisy images simultaneously. Li *et al*. [28] introduced a realistic noise modeling approach for low-light polarization data and built a powerful vision Transformer-based network to effectively suppress noise. PLIE [70] proposed a Stokes-domain enhancement strategy and a dual-branch network to mitigate artifacts in the DoP and AoP. For deblurring, PolDeblur [71] designed a divide-and-conquer strategy and a two-stage neural network. However, due to the inherent ill-posedness of the underlying tasks, these methods often struggle to recover fine details.

There are also benchmark datasets tailored for polarimetric imaging. For example, Hwang *et al*. [19] introduced PolarBurstSR, a dataset intended for developing and evaluating burst-based super-resolution models; Jeon *et al*. [20] presented both trichromatic and hyperspectral Stokes datasets covering diverse indoor and outdoor scenes under different illumination conditions. Although these datasets have substantially supported research on polarimetric imaging, they do not align with the objectives and settings of our task.

B. Fusing Noisy and Blurry Pairs

Compared to low-light enhancement [5], [65] and deblurring [25], [48] methods that only take a single degenerated image as the input, fusing a pair of noisy and blurry images generally achieves higher performance and better generalization ability by leveraging complementary information from both types of data. Early approaches were primarily based on numerical optimization. Yuan *et al*. [60] combined residual and gain-controlled deconvolution to suppress ringing artifacts. Choi *et al*. [6] designed a novel camera system capable of capturing two blurry images and one noisy image in a single shot, and proposed a motion-based fusion algorithm. Son and Park [47] introduced a patch-based method for estimating the point spread function (PSF) by extracting the structure information from the noisy image, along with channel-dependent deblurring method to reconstruct the blur-free image. Later, Son *et al*. [46] proposed an iterative scheme that alternates between PSF estimation and deconvolution, guided by the noisy observation. Gu *et al*. [9] adopted a Gaussian mixture model to estimate the underlying intensity distribution before fusion. However, these optimization-based methods are very time-consuming, which cannot meet the requirement of real-time fusing.

Recently, deep learning has also been applied to this problem. They usually run much faster than optimization-based methods. LSD2 [38] and LSFNet [4] introduced convolutional neural networks to fuse image pairs in an end-to-end manner. SelfIR [63] proposed a self-supervised learning strategy to restore clean and sharp content. D2HNet [64] further improved image quality through a two-phase design. However, since these methods are developed for conventional images, they tend to perform poorly on polarized snapshots due to their inability to preserve the polarization properties during the fusion process.

C. Polarization-based Downstream Applications

Polarization-based algorithms have overcome the performance limitations of conventional image-based approaches in a wide range of downstream applications. Among the most prominent are biomedical imaging and clinical diagnosis [11], which rely on measuring the vectorial properties of light or the transformations imparted by biological tissues. Due to the distinctive polarimetric responses exhibited by transparent objects, polarization has been extensively applied in robotics for transparent object segmentation [22], [36]. In the field of autonomous driving, it has been employed for tasks such as road detection [27] and scene understanding [29]. Additionally, polarization is widely used for stress analysis in birefringent materials [50], as stress-induced changes in birefringence directly affect the polarization state of transmitted light. Furthermore, since material defects often alter polarization properties, polarization has also been adopted for defect detection and analysis in industrial inspection scenarios [30], [58].

Another important application area of polarization is computer vision. Since the polarization properties of light are influenced by surface normals, textures, and material characteristics of scene objects, polarization has been used to solve a variety of 3D vision problems such as shape estimation [1], [3], inverse rendering [7], and depth sensing [21], [49]. In addition, because reflection and scattering alter the polarization state of light, polarization information has been applied to suppress reflection artifacts [26], [35], [57] and remove haze from images [44], [69]. Due to the fact that the DoP values of achromatic pixels are always achromatic regardless of the illumination, polarization can facilitate color constancy [41]. Furthermore, the spatially variant light attenuation introduced by polarizers can be utilized for high dynamic range imaging [55], [66]. Since direct sunlight and ambient skylight possess different polarization properties, polarization has also been explored for shadow removal [72].

III. Background and Problem Formulation

In this section, we introduce the background in Section III-A, and detail the problem formulation in Section III-B.

A. Background

When a linear polarizer with angle α is placed in front of a camera, the captured polarized image $\mathbf{I}_\alpha$ can be modeled by Malus' law [13] as:

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

where $\mathbf{I}$ denotes the total intensity of the light, which can be regarded as the unpolarized image (*i.e.*, captured without the polarizer), $\mathbf{p} \in [0, 1]$ and $\boldsymbol{\theta} \in [0, \pi]$ represent the DoP and AoP of the incoming light to the sensor, respectively. Here, we assume a linear camera response function, which is commonly adopted by polarization cameras. In addition, we focus solely on linear polarization, as polarization cameras are typically equipped with linear polarizers and do not account for circular polarization.

Rewriting Equation (1) in polynomial form, $\mathbf{I}_\alpha$ becomes a linear combination of three parameters $\mathbf{S}_{0,1,2}$:

$$\mathbf{I}_\alpha = \frac{1}{2}\mathbf{S}_0 - \frac{1}{2}\cos(2\alpha) \cdot \mathbf{S}_1 - \frac{1}{2}\sin(2\alpha) \cdot \mathbf{S}_2, \tag{2}$$

where

$$\begin{cases} \mathbf{S}_0 = \mathbf{I} \\ \mathbf{S}_1 = \mathbf{I} \cdot \mathbf{p} \cdot \cos(2\boldsymbol{\theta}) \\ \mathbf{S}_2 = \mathbf{I} \cdot \mathbf{p} \cdot \sin(2\boldsymbol{\theta}) \end{cases} \tag{3}$$

are the Stokes parameters [24] of the incoming light to the sensor. Once $\mathbf{S}_{0,1,2}$ become available, the DoP $\mathbf{p}$ and AoP $\boldsymbol{\theta}$ can be computed as:

$$\mathbf{p} = \frac{\sqrt{\mathbf{S}_1^2 + \mathbf{S}_2^2}}{\mathbf{S}_0} \quad \text{and} \quad \boldsymbol{\theta} = \frac{1}{2}\arctan(\frac{\mathbf{S}_2}{\mathbf{S}_1}). \tag{4}$$

Polarization-based downstream applications often rely on accurate acquisition of the DoP $\mathbf{p}$ and AoP $\boldsymbol{\theta}$ to provide physical cues, necessitating the solution of Equation (2). As this equation represents a linear system with three unknowns, $\mathbf{S}_{0,1,2}$, solving it requires at least three polarized images captured at different polarizer angles. While one could use a conventional camera and rotate the polarizer to capture multiple shots, polarization cameras offer a more convenient solution. This is because polarization cameras are able to capture four polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ with different polarizer angles $\alpha_{1,2,3,4} = 0°, 45°, 90°, 135°$ in a single shot. These four polarized images comprise a polarized snapshot $\mathcal{I}$, which can be written as

$$\mathcal{I} = \mathbf{I}_{\alpha_{1,2,3,4}}. \tag{5}$$

By plugging the values of $\alpha_{1,2,3,4}$ into Equation (1), Equation (2), and Equation (3), we can deduce that the Stokes parameters $\mathbf{S}_{0,1,2}$ can be directly calculated from $\mathbf{I}_{\alpha_{1,2,3,4}}$:

$$\begin{cases} \mathbf{S}_0 = \frac{1}{2}(\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_3} + \mathbf{I}_{\alpha_4}) \\ \quad\;\; = \mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_3} = \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_4} \\ \mathbf{S}_1 = \mathbf{I}_{\alpha_3} - \mathbf{I}_{\alpha_1} \\ \mathbf{S}_2 = \mathbf{I}_{\alpha_4} - \mathbf{I}_{\alpha_2} \end{cases}. \tag{6}$$

This makes the acquisition of the DoP $\mathbf{p}$ and AoP $\boldsymbol{\theta}$ significantly more efficient and practical.

B. Problem Formulation

As discussed in Section III-A, the analysis implicitly assumes that the polarized snapshot $\mathcal{I}$ is well-exposed. However, due to the spatially variant light attenuation introduced by polarizers, determining an appropriate exposure time in real-world polarimetric imaging remains a challenging task. As shown in Figure 1, a short exposure results in low SNR and color distortion, while longer exposure improves SNR but introduces motion blur and saturation. Here, following Equation (5), we denote the short-exposed noisy and long-exposed blurry polarized snapshots as

$$\mathcal{L} = \mathbf{L}_{\alpha_{1,2,3,4}} \quad \text{and} \quad \mathcal{B} = \mathbf{B}_{\alpha_{1,2,3,4}}, \tag{7}$$

respectively. Substituting $\mathbf{I}_{\alpha_{1,2,3,4}}$ in Equation (6) with $\mathbf{L}_{\alpha_{1,2,3,4}}$ and $\mathbf{B}_{\alpha_{1,2,3,4}}$, we can obtain the noisy and blurry Stokes parameters $\mathbf{S}^{\mathrm{L}}_{0,1,2}$ and $\mathbf{S}^{\mathrm{B}}_{0,1,2}$, respectively. Similarly, the corresponding degenerated DoP and AoP (denoted as $\mathbf{p}^{\mathrm{L}}$ and $\boldsymbol{\theta}^{\mathrm{L}}$, $\mathbf{p}^{\mathrm{B}}$ and $\boldsymbol{\theta}^{\mathrm{B}}$) can be computed by substituting $\mathbf{S}_{0,1,2}$ in Equation (4) with $\mathbf{S}^{\mathrm{L}}_{0,1,2}$ and $\mathbf{S}^{\mathrm{B}}_{0,1,2}$, respectively.

Our objective is to develop a framework for reconstructing a well-exposed polarized snapshot $\mathcal{I}$ that is clean, sharp, and free from color distortion or saturation, thereby enabling high-quality polarimetric imaging. Given that the short-exposed input $\mathcal{L}$ tends to be sharp but noisy, while the long-exposed input $\mathcal{B}$ is cleaner yet typically suffers from blur, a more promising strategy is to fuse them in a complementary manner, rather than relying solely on low-light enhancement of $\mathcal{L}$ or deblurring of $\mathcal{B}$. Accordingly, the reconstruction process can be formulated as a maximum a posteriori (MAP) estimation of the output $\mathcal{I}$ conditioned on the inputs $\mathcal{L}$ and $\mathcal{B}$, with a fusion function f parameterized by Ψ:

$$\underset{\Psi}{\operatorname{argmax}}\, f(\mathcal{I}|\mathcal{L}, \mathcal{B}, \Psi). \tag{8}$$

IV. Polarization-Aware Three-Phase Fusion Scheme

To solve the MAP problem defined in Equation (8), it is essential to design an appropriate fusion scheme for the fusion function f. Although numerous fusion methods have been developed for conventional images [4], [38], [63], [64], they are not directly applicable to polarized snapshots, as they primarily focus on mitigating blur and noise while neglecting the preservation of polarization properties. To address this issue, we propose a polarization-aware fusion scheme, as illustrated in Figure 2, which comprises three key phases: irradiance restoration, polarization reconstruction, and artifact suppression. These phases are introduced in detail in the following.

A. Irradiance Restoration

Since $\mathbf{S}_0$ physically represents the total intensity of the light, it inherently contains polarization-unrelated irradiance information that can serve as a valuable prior for guiding subsequent polarization-aware processing. Moreover, as noted

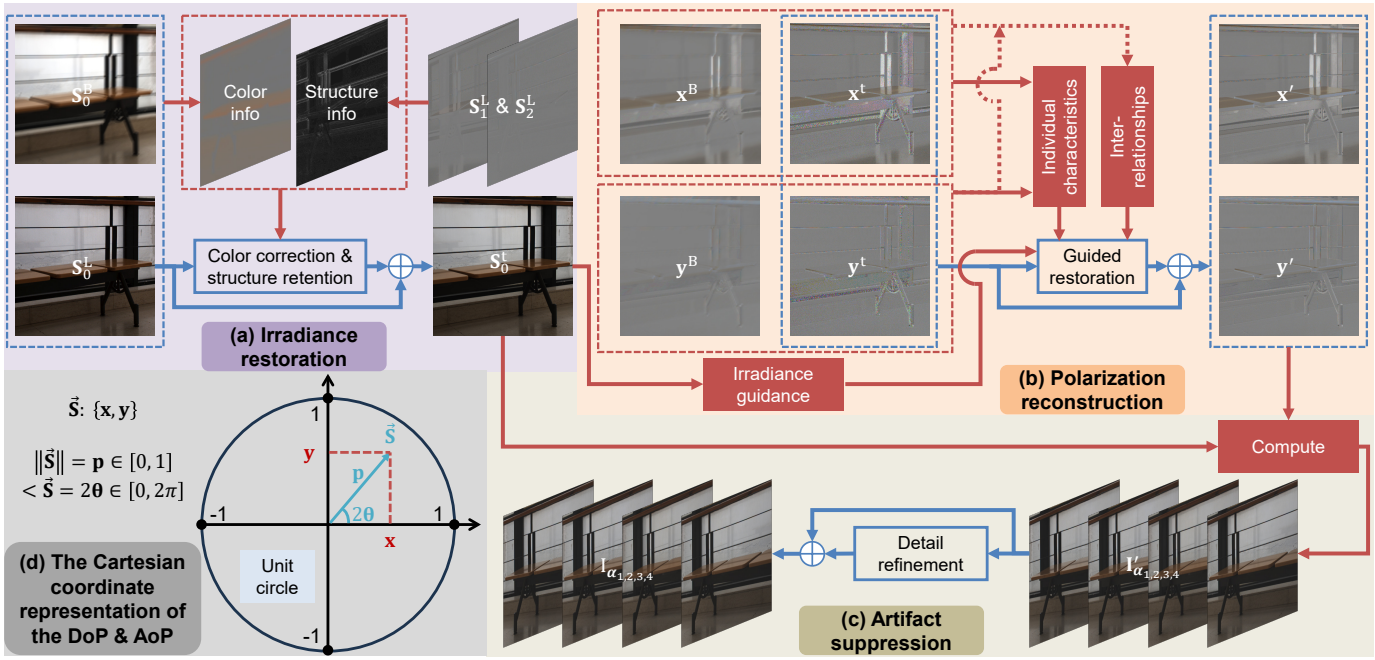


Fig. 2. Our polarization-aware fusion scheme consists of three phases: (a) irradiance restoration, aiming to restore the polarization-unrelated high-level irradiance information of the scene by recovering the total intensity of the light; (b) polarization reconstruction, aiming to establish the physical correlation among the polarized images by reconstructing the DoP and AoP in a Cartesian coordinate representation; (c) artifact suppression, aiming to suppress the artifacts lying in the details by performing refinement in the image domain. In addition, (d) a toy example is provided to illustrate the Cartesian coordinate representation of the DoP and AoP.

in [71], polarized and unpolarized images captured from the same scene exhibit strong semantic similarity. This implies that $\mathbf{S}_0$, which is equivalent to the unpolarized image $\mathbf{I}$, may share similar high-level features with the corresponding polarized snapshot $\mathcal{I}$. Leveraging these characteristics, we formulate the first phase as irradiance restoration, which aims to restore the polarization-unrelated high-level irradiance information of the scene by recovering the total intensity of the light.

As illustrated in Figure 2 (a), the short-exposed polarized snapshot $\mathcal{L}$ tends to preserve sharper contours compared to the long-exposed one $\mathcal{B}$. Therefore, we choose to learn the residual between $\mathbf{S}_0^{\mathrm{L}}$ and $\mathbf{S}_0^{\mathrm{t}}$ instead of the residual between $\mathbf{S}_0^{\mathrm{B}}$ and $\mathbf{S}_0^{\mathrm{t}}$. Here, $\mathbf{S}_0^{\mathrm{t}}$ denotes the recovered coarse value of $\mathbf{S}_0$. However, $\mathbf{S}_0^{\mathrm{L}}$ often suffers from color distortion and significant noise, which hampers effective feature extraction and leads to incorrect global tone and indistinct local structures. Fortunately, although $\mathbf{S}_0^{\mathrm{B}}$ is affected by motion blur, it retains reliable color information thanks to the relatively high SNR of $\mathcal{B}$. Additionally, $\mathbf{S}_{1,2}^{\mathrm{L}}$ provide valuable structural information, as both represent the differences between pairs of polarized images (see Equation (6)) and tend to highlight regions with high gradients. Thus, we propose to incorporate both the color information from $\mathbf{S}_0^{\mathrm{B}}$ and the structure information from $\mathbf{S}_{1,2}^{\mathrm{L}}$ to perform color correction and structure retention on $\mathbf{S}_0^{\mathrm{L}}$, respectively. In summary, this phase (denoted as f_1) can be described as:

$$\mathbf{S}_0^{\mathrm{t}} = f_1(\mathbf{S}_0^{\mathrm{L}}, \mathbf{S}_0^{\mathrm{B}}, \mathbf{S}_{1,2}^{\mathrm{L}}) + \mathbf{S}_0^{\mathrm{L}}. \tag{9}$$

Compared to our preliminary work PolFusion [67], PolFusion+ explicitly accounts for saturation in $\mathcal{B}$ and introduces a more robust strategy that extracts color information not only from $\mathcal{B}$ but also from $\mathcal{L}$. This is motivated by the observation that pixels saturated in $\mathcal{B}$ generally correspond to bright regions of the scene. These regions remain unsaturated in $\mathcal{L}$ while maintaining sufficient intensity to ensure a good signal-to-noise ratio, thereby reducing the impact of color distortion.

B. Polarization Reconstruction

Once the polarization-unrelated high-level irradiance information is restored, the next step is to establish the physical correlation among the polarized images. As indicated by Equation (4) and Equation (6), this correlation is governed by the DoP and AoP, given the irradiance of the scene. Accordingly, we define the second phase as polarization reconstruction, which aims to reconstruct the high-quality DoP and AoP that faithfully preserves the polarization properties. To achieve this, previous polarimetric imaging methods typically perform indirect reconstruction by restoring the degraded values in either the image domain [15], [28], [56], [71] or the Stokes domain [70]. This strategy is mainly adopted because the degradation patterns of the DoP and AoP are generally more complex than those of the polarized images or Stokes parameters, due to the inherent non-linearity in their formulation (see Equation (4)). However, such indirect approaches often increase the ill-posedness of the problem and suffer from error accumulation, which in turn reduces the overall effectiveness of the optimization.

Notably, we observe that by interpreting the DoP $\mathbf{p}$ as the radial distance and twice the AoP $2\boldsymbol{\theta}$ as the angular component in a polar coordinate system, they together form a vector $\vec{\mathbf{S}}$ that lies within the unit circle, as illustrated in Figure 2 (d). By

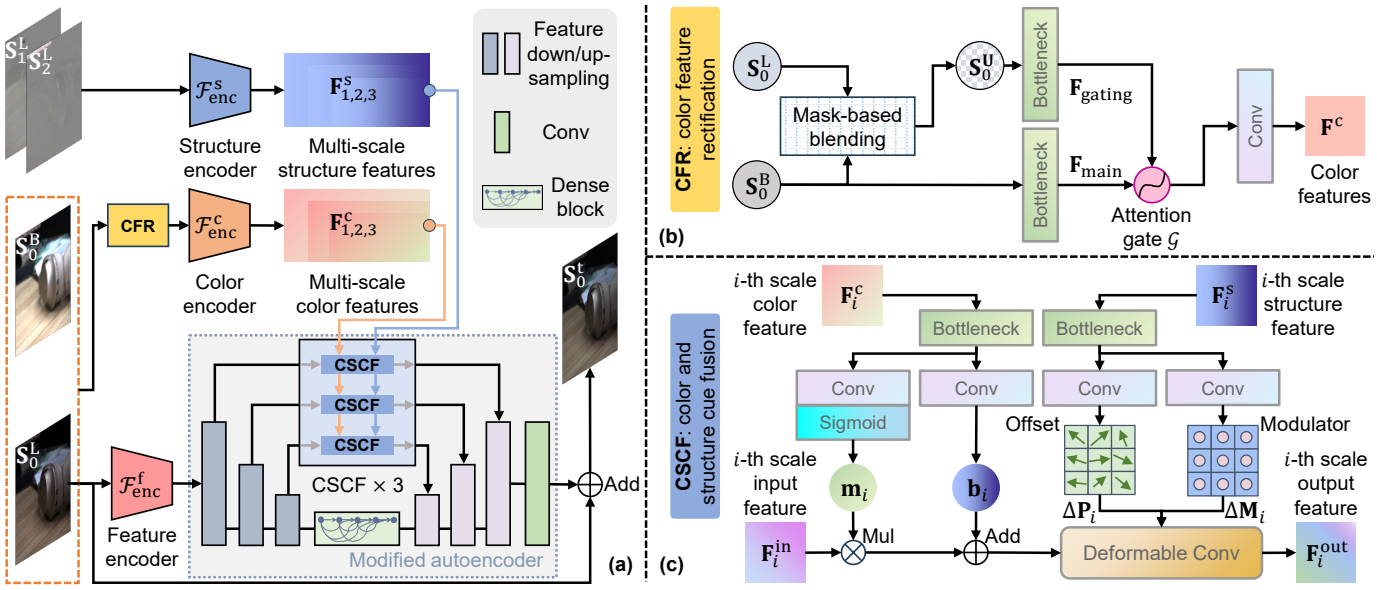


Fig. 3. Illustration of (a) the irradiance restoration network architecture, including the design details of (b) the CFR (color feature recticfication) and (c) CSCF (color and structure cue fusion) modules.

converting the polar coordinates into Cartesian coordinates, we obtain two components $\mathbf{x}$ and $\mathbf{y}$, representing the projections of $\vec{\mathbf{S}}$ along the horizontal and vertical axes, respectively. We refer to the pair $\{\mathbf{x}, \mathbf{y}\}$ as the Cartesian coordinate representation of the DoP and AoP, which can be computed as:

$$\mathbf{x} = \mathbf{p} \cdot \cos(2\boldsymbol{\theta}) \quad \text{and} \quad \mathbf{y} = \mathbf{p} \cdot \sin(2\boldsymbol{\theta}). \tag{10}$$

Combining Equation (3) and Equation (10), we can derive the following equation:

$$\mathbf{x} = \frac{\mathbf{S}_1}{\mathbf{S}_0} \quad \text{and} \quad \mathbf{y} = \frac{\mathbf{S}_2}{\mathbf{S}_0}, \tag{11}$$

which reveals that $\{\mathbf{x}, \mathbf{y}\}$ can be interpreted as the normalized $\mathbf{S}_{1,2}$, with the normalization performed by the total intensity of the light $\mathbf{S}_0$. This indicates that reconstructing $\{\mathbf{x}, \mathbf{y}\}$ not only mitigates error accumulation by maintaining a direct optimization of the DoP and AoP, but also exhibits a comparable level of ill-posedness to prior indirect reconstruction approaches in polarimetric imaging [15], [28], [56], [71], as $\{\mathbf{x}, \mathbf{y}\}$ keep a linear relationship with the polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ (see Equation (6)).

As shown in Figure 2 (b), we propose to utilize $\mathbf{S}_0^{\mathrm{t}}$, obtained from the irradiance restoration phase, to perform an initial estimation first:

$$\mathbf{x}^{\mathrm{t}} = \frac{\mathbf{S}_1^{\mathrm{L}}}{\mathbf{S}_0^{\mathrm{t}}} \quad \text{and} \quad \mathbf{y}^{\mathrm{t}} = \frac{\mathbf{S}_2^{\mathrm{L}}}{\mathbf{S}_0^{\mathrm{t}}}, \tag{12}$$

where $\{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}$ denote the estimated coarse values of $\{\mathbf{x}, \mathbf{y}\}$. Subsequently, we propose to refine $\{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}$ by learning their residuals with respect to the enhanced versions $\{\mathbf{x}', \mathbf{y}'\}$, leveraging the irradiance information provided by $\mathbf{S}_0^{\mathrm{t}}$ along with the additional guidance from $\{\mathbf{x}^{\mathrm{B}}, \mathbf{y}^{\mathrm{B}}\}$, which are defined as:

$$\mathbf{x}^{\mathrm{B}} = \frac{\mathbf{S}_1^{\mathrm{B}}}{\mathbf{S}_0^{\mathrm{B}}} \quad \text{and} \quad \mathbf{y}^{\mathrm{B}} = \frac{\mathbf{S}_2^{\mathrm{B}}}{\mathbf{S}_0^{\mathrm{B}}}. \tag{13}$$

In summary, this phase (denoted as f_2) can be described as:

$$\{\mathbf{x}', \mathbf{y}'\} = f_2(\{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}, \mathbf{S}_0^{\mathrm{t}}, \{\mathbf{x}^{\mathrm{B}}, \mathbf{y}^{\mathrm{B}}\}) + \{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}. \tag{14}$$

In our previous work PolFusion [67], $\mathbf{x}^{\mathrm{t}}$ and $\mathbf{y}^{\mathrm{t}}$ were treated as two independent physical quantities. In contrast, PolFusion+ performs guided restoration in a joint manner, allowing both their individual characteristics and interrelationships to be fully exploited. This joint modeling significantly improves the accuracy of the reconstructed DoP and AoP.

C. Artifact Suppression

With $\mathbf{S}_0^{\mathrm{t}}$ and $\{\mathbf{x}', \mathbf{y}'\}$ available, the coarse values of the polarized images (denoted as $\mathbf{I}'_{\alpha_{1,2,3,4}}$) can be computed using Equation (10) and Equation (1). However, directly adopting $\mathbf{I}'_{\alpha_{1,2,3,4}}$ as the final output is inadvisable, as their quality remains suboptimal. This is primarily attributed to the phase-wise decoupling: $\mathbf{S}_0^{\mathrm{t}}$ and $\{\mathbf{x}', \mathbf{y}'\}$ are estimated independently in two separate phases, potentially resulting in inconsistencies between the irradiance and polarization information. Such inconsistencies often give rise to noticeable artifacts, manifesting as unnatural intensity variations across the image plane, which are particularly prominent in textureless or homogeneous background regions.

To tackle this issue, as shown in Figure 2 (c), we explicitly introduce a third phase, termed artifact suppression, which is dedicated to performing detail refinement on $\mathbf{I}'_{\alpha_{1,2,3,4}}$ in the image domain to eliminate the artifacts caused by irradiance-polarization inconsistencies. Instead of seeking guidance from other physical quantities, we propose to directly learn the residual between $\mathbf{I}'_{\alpha_{1,2,3,4}}$ and $\mathbf{I}_{\alpha_{1,2,3,4}}$ in a blind manner. This design leverages the rich contextual information already present in $\mathbf{I}'_{\alpha_{1,2,3,4}}$, while avoiding the increased computational cost associated with incorporating auxiliary inputs that offer limited

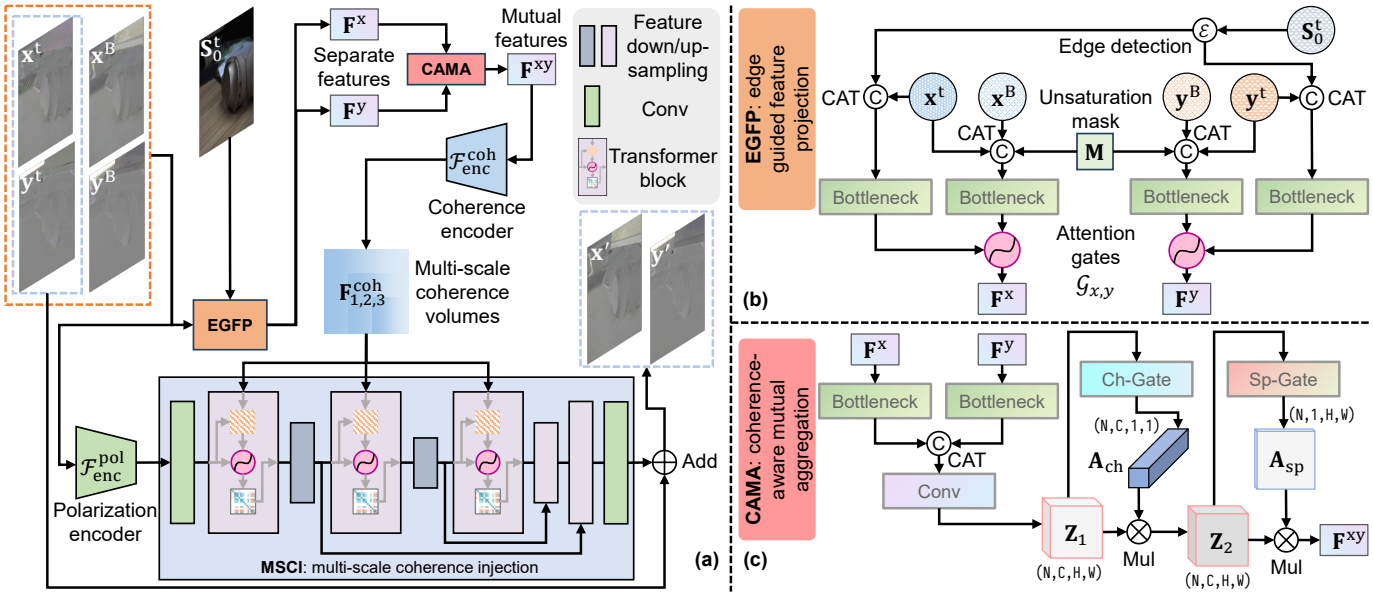


Fig. 4. Illustration of (a) the polarization reconstruction network architecture, including the design details of (b) the EGFP (edge guided feature projection) and (c) CAMA (coherence-aware mutual aggregation) modules.

performance gain. In summary, this phase (denoted as f_3) can be described as:

$$\mathbf{I}_{\alpha_{1,2,3,4}} = f_3(\mathbf{I}'_{\alpha_{1,2,3,4}}) + \mathbf{I}'_{\alpha_{1,2,3,4}}. \tag{15}$$

V. Degradation-Oriented Neural Network

To implement the polarization-aware three-phase fusion scheme described in Section IV, we adopt an efficient learning-based approach instead of relying on time-consuming numerical optimization. Specifically, we design a neural network composed of three dedicated sub-networks, each corresponding to one of the three fusion phases. These sub-networks are carefully designed to account for the underlying degradation processes, thereby ensuring the overall fusion performance. The architectures of these sub-networks are detailed in the following.

A. Irradiance Restoration Network

Overall architecture. This sub-network aims to implement the irradiance restoration phase (f_1). According to Equation (9), it is designed to take $\mathbf{S}_0^{\mathrm{L}}$, $\mathbf{S}_0^{\mathrm{B}}$, and $\mathbf{S}_{1,2}^{\mathrm{L}}$ as input and output $\mathbf{S}_0^{\mathrm{t}}$. Specifically, as shown in Figure 3 (a), we first employ a structure encoder $\mathcal{F}_{\mathrm{enc}}^{\mathrm{s}}$ to extract multi-scale structure features $\mathbf{F}_{1,2,3}^{\mathrm{s}}$ from $\mathbf{S}_{1,2}^{\mathrm{L}}$. In parallel, a CFR (color feature recticfication) module followed by a color encoder $\mathcal{F}_{\mathrm{enc}}^{\mathrm{c}}$ is used to extract the multi-scale color features $\mathbf{F}_{1,2,3}^{\mathrm{c}}$ from both $\mathbf{S}_0^{\mathrm{B}}$ and $\mathbf{S}_0^{\mathrm{L}}$. Meanwhile, a feature encoder $\mathcal{F}_{\mathrm{enc}}^{\mathrm{f}}$ processes $\mathbf{S}_0^{\mathrm{L}}$ to produce base features, which are then fed into a modified autoencoder [14] enhanced with a dense block [18] at the coarsest layer to learn the residual between $\mathbf{S}_0^{\mathrm{L}}$ and $\mathbf{S}_0^{\mathrm{t}}$. To make full use of the guidance provided by $\mathbf{F}_{1,2,3}^{\mathrm{c}}$ and $\mathbf{F}_{1,2,3}^{\mathrm{s}}$, we replace the standard skip connections in the autoencoder with three CSCF (color and structure cue fusion) modules at corresponding scales, rather than injecting these features directly into the downsampling part. All encoders (including $\mathcal{F}_{\mathrm{enc}}^{\mathrm{s}}$, $\mathcal{F}_{\mathrm{enc}}^{\mathrm{c}}$, and $\mathcal{F}_{\mathrm{enc}}^{\mathrm{f}}$) are designed to be convolutional layers.

CFR: color feature recticfication. The CFR module is designed to rectify degraded color features caused by saturation. As shown in Figure 3 (b), we first adopt a mask-based blending strategy to produce an updated version of $\mathbf{S}_0^{\mathrm{B}}$, denoted as $\mathbf{S}_0^{\mathrm{U}}$:

$$\mathbf{S}_0^{\mathrm{U}} = \mathbf{S}_0^{\mathrm{B}} \cdot \mathbf{M} + \mathbf{S}_0^{\mathrm{L}} \cdot (1 - \mathbf{M}), \tag{16}$$

where $\mathbf{M}$ is a binary unsaturation mask defined as:

$$\mathbf{M}(k) = \begin{cases} 1 & \text{if } \mathbf{S}_0^{\mathrm{B}}(k) \text{ is unsaturated} \\ 0 & \text{otherwise} \end{cases}, \tag{17}$$

with k denoting the pixel index. This operation ensures that saturated regions in $\mathbf{S}_0^{\mathrm{B}}$ are replaced with their counterparts from $\mathbf{S}_0^{\mathrm{L}}$, which typically retain more reliable color information. After the mask-based blending, two bottleneck blocks [12], denoted as B_B and B_U, are used to extract features from $\mathbf{S}_0^{\mathrm{B}}$ and $\mathbf{S}_0^{\mathrm{U}}$, respectively. An attention gate [40] $\mathcal{G}$ is then applied to these features, followed by a convolutional layer C_G to produce initial color features $\mathbf{F}^{\mathrm{c}}$. This process can be formulated as:

$$\mathbf{F}^{\mathrm{c}} = C_G(\mathcal{G}(\mathbf{F}_{\mathrm{main}}, \mathbf{F}_{\mathrm{gating}})), \tag{18}$$

where $\mathbf{F}_{\mathrm{main}}$ and $\mathbf{F}_{\mathrm{gating}}$ are the main and gating signals for the attention mechanism, obtained by:

$$\begin{cases} \mathbf{F}_{\mathrm{main}} = B_B(\mathbf{S}_0^{\mathrm{B}}) \\ \mathbf{F}_{\mathrm{gating}} = B_U(\mathbf{S}_0^{\mathrm{U}}) \end{cases}. \tag{19}$$

CSCF: color and structure cue fusion. The CSCF module is designed to mitigate issues related to incorrect global tone and indistinct local structures in the feature space. Without loss of generality, we illustrate the workflow of CSCF at the i-th scale ($i = 1, 2, 3$) in Figure 3 (c), where it takes $\mathbf{F}_i^{\mathrm{in}}$ as input and produces the output $\mathbf{F}_i^{\mathrm{out}}$, guided by the corresponding

color features $\mathbf{F}_i^{\mathrm{c}}$ and structure features $\mathbf{F}_i^{\mathrm{s}}$. First, we derive a multiplier $\mathbf{m}_i$ and a bias $\mathbf{b}_i$ from $\mathbf{F}_i^{\mathrm{c}}$:

$$\begin{cases} \mathbf{m}_i = Sigmoid(C_m(B_c(\mathbf{F}_i^{\mathrm{c}}))) \\ \mathbf{b}_i = C_b(B_c(\mathbf{F}_i^{\mathrm{c}})) \end{cases}, \tag{20}$$

where B_c denotes a bottleneck block [12] for feature projection, and C_m, C_b are two convolutional layers. These parameters are used to apply an affine transformation to $\mathbf{F}_i^{\mathrm{in}}$, effectively adjusting the color to correct global tone:

$$\mathbf{F}_i^{\mathrm{t}} = \mathbf{m}_i \cdot \mathbf{F}_i^{\mathrm{in}} + \mathbf{b}_i, \tag{21}$$

where $\mathbf{F}_i^{\mathrm{t}}$ is the transformed feature. To further enhance local structure clarity, we apply a deformable convolution layer [75] D to adaptively align gradient details:

$$\mathbf{F}_i^{\mathrm{out}} = D(\mathbf{F}_i^{\mathrm{t}}, \Delta\mathbf{P}_i, \Delta\mathbf{M}_i), \tag{22}$$

where $\Delta\mathbf{P}_i$ and $\Delta\mathbf{M}_i$ are the learned spatial offsets of sampling points and modulation scalars, respectively, obtained from the structure features:

$$\begin{cases} \Delta\mathbf{P}_i = C_P(B_s(\mathbf{F}_i^{\mathrm{s}})) \\ \Delta\mathbf{M}_i = C_M(B_s(\mathbf{F}_i^{\mathrm{s}})) \end{cases}, \tag{23}$$

where B_s is another bottleneck block and C_P, C_M are additional convolutional layers.

B. Polarization Reconstruction Network

Overall architecture. This sub-network aims to implement the polarization reconstruction phase (f_2). According to Equation (14), it is designed to take $\{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}$, $\mathbf{S}_0^{\mathrm{t}}$, and $\{\mathbf{x}^{\mathrm{B}}, \mathbf{y}^{\mathrm{B}}\}$ as input and output $\{\mathbf{x}', \mathbf{y}'\}$. Specifically, as shown in Figure 4 (a), we first employ an EGFP (edge guided feature projection) module to project $\mathbf{x}^{\mathrm{t}}$ and $\mathbf{x}^{\mathrm{B}}$, $\mathbf{y}^{\mathrm{t}}$ and $\mathbf{y}^{\mathrm{B}}$ into the feature space separately, guided by $\mathbf{S}_0^{\mathrm{t}}$. This yields their respective projected features, $\mathbf{F}^{\mathrm{x}}$ and $\mathbf{F}^{\mathrm{y}}$. Subsequently, a CAMA (coherence-aware mutual aggregation) module is adopted to aggregate $\mathbf{F}^{\mathrm{x}}$ and $\mathbf{F}^{\mathrm{y}}$ into their mutual features $\mathbf{F}^{\mathrm{x,y}}$, which are then encoded into multi-scale coherence volumes $\mathbf{F}_{1,2,3}^{\mathrm{coh}}$ by a coherence encoder $\mathcal{F}_{\mathrm{enc}}^{\mathrm{coh}}$. Following this, a polarization encoder $\mathcal{F}_{\mathrm{enc}}^{\mathrm{pol}}$ extracts features from $\{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}$, which are then processed by a MSCI (multi-scale coherence injection) module to explicitly inject the coherence information from $\mathbf{F}_{1,2,3}^{\mathrm{coh}}$ into the residual learning between $\{\mathbf{x}^{\mathrm{t}}, \mathbf{y}^{\mathrm{t}}\}$ and $\{\mathbf{x}', \mathbf{y}'\}$. All encoders (including $\mathcal{F}_{\mathrm{enc}}^{\mathrm{coh}}$ and $\mathcal{F}_{\mathrm{enc}}^{\mathrm{pol}}$) are designed to be convolutional layers.

EGFP: edge guided feature projection. The EGFP module is designed to explore the individual characteristics of the horizontal and vertical axes of the Cartesian coordinate representation. As shown in Figure 4 (b), it explicitly utilizes the fine-grained edges extracted from $\mathbf{S}_0^{\mathrm{t}}$ via edge detection $\mathcal{E}$, together with the unsaturation mask $\mathbf{M}$ (defined in Equation (17)) as guidance. Since the acquisition processes for $\mathbf{F}^{\mathrm{x}}$ and $\mathbf{F}^{\mathrm{y}}$ are analogous, we focus on the derivation of $\mathbf{F}^{\mathrm{x}}$ without loss of generality. Specifically, we first use two bottleneck blocks [12], denoted as B_1 and B_2, to process the concatenated inputs: B_1 takes $\mathtt{CAT}(\mathbf{x}^{\mathrm{t}}, \mathbf{x}^{\mathrm{B}}, \mathbf{M})$, while B_2 takes $\mathtt{CAT}(\mathbf{x}^{\mathrm{t}}, \mathcal{E}(\mathbf{S}_0^{\mathrm{t}}))$, yielding the main and gating signals for the subsequent attention mechanism. An attention gate [40] $\mathcal{G}_x$ is then adopted to produce $\mathbf{F}^{\mathrm{x}}$, formulated as:

$$\mathbf{F}^{\mathrm{x}} = \mathcal{G}_x(B_1(\mathtt{CAT}(\mathbf{x}^{\mathrm{t}}, \mathbf{x}^{\mathrm{B}}, \mathbf{M})), B_2(\mathtt{CAT}(\mathbf{x}^{\mathrm{t}}, \mathcal{E}(\mathbf{S}_0^{\mathrm{t}})))). \tag{24}$$

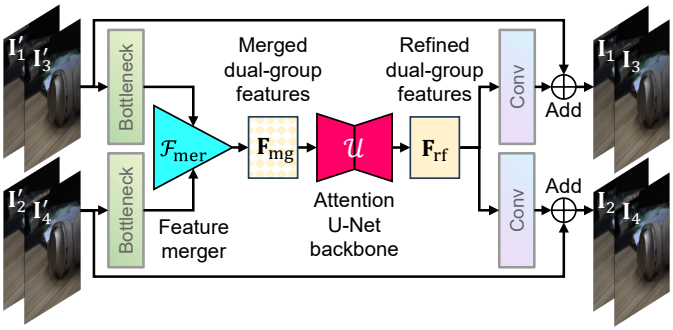


Fig. 5. Illustration of the artifact suppression network architecture.

CAMA: coherence-aware mutual aggregation. The CAMA module is designed to capture the interrelationships between $\mathbf{F}^{\mathrm{x}}$ and $\mathbf{F}^{\mathrm{y}}$. As shown in Figure 4 (c), the module begins by independently processing $\mathbf{F}^{\mathrm{x}}$ and $\mathbf{F}^{\mathrm{y}}$ through two bottleneck blocks [12], which enhance their respective feature representations. The enhanced features are then concatenated and fused via a convolutional layer to produce $\mathbf{Z}_1 \in \mathbb{R}^{N\times C\times H\times W}$, where N, C, H, and W denote the batch size, number of channels, height, and width, respectively. The resulting $\mathbf{Z}_1$ serves as an initial estimation of the mutual features $\mathbf{F}^{\mathrm{x,y}}$. Inspired by CBAM [54], which effectively models feature interrelationships, we propose to learn a channel attention vector $\mathbf{A}_{\mathrm{ch}} \in \mathbb{R}^{N\times C\times 1\times 1}$ and a spatial attention vector $\mathbf{A}_{\mathrm{sp}} \in \mathbb{R}^{N\times 1\times H\times W}$ using a channel gate $\mathcal{C}_h$ and spatial gate $\mathcal{S}_p$, respectively. These attention maps are utilized sequentially to recalibrate $\mathbf{Z}_1$ as follows:

$$\mathbf{F}^{\mathrm{x,y}} = \mathbf{A}_{\mathrm{sp}} \cdot \mathbf{Z}_2 = \mathbf{A}_{\mathrm{sp}} \cdot (\mathbf{A}_{\mathrm{ch}} \cdot \mathbf{Z}_1). \tag{25}$$

Specifically, $\mathbf{A}_{\mathrm{sp}}$ and $\mathbf{A}_{\mathrm{ch}}$ are obtained as:

$$\begin{cases} \mathbf{A}_{\mathrm{ch}} = \mathcal{C}_h(\mathbf{Z}_1) = Sigmoid(\mathtt{MLP}_1(\mathcal{P}_{\mathrm{sp}}(\mathbf{Z}_1))) \\ \mathbf{A}_{\mathrm{sp}} = \mathcal{S}_p(\mathbf{Z}_2) = Sigmoid(\mathtt{MLP}_2(\mathcal{P}_{\mathrm{ch}}(\mathbf{Z}_2))) \end{cases}, \tag{26}$$

where $\mathtt{MLP}_1$ and $\mathtt{MLP}_2$ denotes two different multi-layer perceptrons, $\mathcal{P}_{\mathrm{sp}}$ and $\mathcal{P}_{\mathrm{ch}}$ represent global average pooling operations along the spatial and channel dimensions, respectively.

MSCI: multi-scale coherence injection. As shown in Figure 4 (a), the MSCI module functions as a backbone component that explicitly injects the multi-scale coherence volumes $\mathbf{F}_{1,2,3}^{\mathrm{coh}}$ into the restoration process of $\{\mathbf{x}', \mathbf{y}'\}$. To enlarge receptive field and include long-range association, we adopt Transformer blocks [8] with cross-attention layers as the core building units. To fully leverage the guidance provided by $\mathbf{F}_{1,2,3}^{\mathrm{coh}}$, we design the complete information flow following the paradigm of autoencoder [14], enabling hierarchical feature encoding and decoding for effective multi-scale processing.

C. Artifact Suppression Network

This sub-network aims to implement the artifact suppression phase (f_3). According to Equation (15), it is designed to take $\mathbf{I}'_{\alpha_{1,2,3,4}}$ as input and output $\mathbf{I}_{\alpha_{1,2,3,4}}$. Specifically, as

TABLE I
QUANTITATIVE COMPARISONS ON SYNTHETIC DATA. THE COMPARISONS INVOLVE OUR POLFUSION+, OUR PRELIMINARY WORK POLFUSION [67], TWO STATE-OF-THE-ART POLARIMETRIC IMAGING APPROACHES (PLIE [70] FOR LOW-LIGHT ENHANCEMENT AND POLDEBLUR [71] FOR DEBLURRING) AND THEIR CORRESPONDING MODIFIED VERSIONS (PLIE_M AND POLDEBLUR_M), FOUR STATE-OF-THE-ART FUSION METHODS DEVELOPED FOR CONVENTIONAL IMAGES (LSD2 [38], LSFNET [4], SELFIR [63], AND D2HNET [64]). THROUGHOUT THIS PAPER, WE USE THE **RED** (**BLUE**) TEXT TO HIGHLIGHT THE BEST (SECOND-BEST) RESULTS.

	PSNR (DoP $\mathbf{p}$)	SSIM (DoP $\mathbf{p}$)	PSNR (AoP $\boldsymbol{\theta}$)	SSIM (AoP $\boldsymbol{\theta}$)	PSNR (total intensity $\mathbf{S}_0$)	SSIM (total intensity $\mathbf{S}_0$)
PolFusion+	**29.59**	**0.825**	**17.12**	**0.407**	**40.47**	**0.986**
PolFusion [67]	**28.94**	0.793	**17.07**	**0.402**	39.82	**0.985**
PLIE [70]	27.92	**0.807**	16.99	0.397	39.35	0.983
PolDeblur [71]	25.28	0.671	15.28	0.266	23.45	0.795
PLIE_m	27.40	0.794	16.39	0.368	40.13	0.984
PolDeblur_m	24.49	0.780	16.89	0.401	**40.42**	0.979
LSD2 [38]	25.61	0.655	13.98	0.268	26.10	0.916
LSFNet [4]	26.43	0.722	16.16	0.319	28.71	0.897
SelfIR [63]	21.33	0.749	16.40	0.374	36.02	0.935
D2HNet [64]	23.32	0.663	15.20	0.255	24.35	0.791

shown in Figure 5, we begin by dividing $\mathbf{I}'_{\alpha_{1,2,3,4}}$ into two groups: $\mathbf{I}'_{\alpha_{1,3}}$ and $\mathbf{I}'_{\alpha_{2,4}}$. This division is based on Equation (6), that each group contains the complete irradiance information ($\mathbf{S}_0$) and contains one distinct component of the polarization properties ($\mathbf{S}_1$ and $\mathbf{S}_2$). We then apply two bottleneck blocks [12] to extract features from $\mathbf{I}'_{\alpha_{1,3}}$ and $\mathbf{I}'_{\alpha_{2,4}}$, respectively. These features are subsequently merged using a feature merger $\mathcal{F}_{\text{mer}}$, which consists of a residual block [12] followed by a squeeze-and-excitation block [17], yielding the merged dual-group features $\mathbf{F}_{\text{mg}}$. After that, we employ an attention U-Net [40] $\mathcal{U}$ to perform pixel-wise detail refinement on $\mathbf{F}_{\text{mg}}$, producing the refined dual-group features $\mathbf{F}_{\text{rf}}$, which can be formulated as:

$$\mathbf{F}_{\text{rf}} = \mathcal{U}(\mathbf{F}_{\text{mg}}) = \mathcal{U}(\mathcal{F}_{\text{mer}}(B_{g_1}(\mathbf{I}'_{\alpha_{1,3}}), B_{g_2}(\mathbf{I}'_{\alpha_{2,4}}))), \quad (27)$$

where B_{g_1} and B_{g_2} denote the bottleneck blocks corresponding to each group. Finally, two convolutional layers are applied to decode $\mathbf{F}_{\text{rf}}$ into the residuals corresponding to each group.

VI. IMPLEMENTATION DETAILS

A. Loss Function

To optimize our degradation-oriented neural network, we design three loss functions corresponding to the three sub-networks $f_{1,2,3}$: the irradiance loss L_{f_1}, the polarization loss L_{f_2}, and the refinement loss L_{f_3}. Each of these losses will be described in detail in the following.

Irradiance loss. The irradiance loss L_{f_1} is defined as:

$$L_{f_1} = \lambda_1 L_1(\mathbf{S}_0^{\text{t}}, \mathbf{S}_0^{\text{gt}}) + \lambda_2 L_{\text{perc}}(\mathbf{S}_0^{\text{t}}, \mathbf{S}_0^{\text{gt}}), \quad (28)$$

where $\lambda_{1,2}$ are set to 10.0 and 0.05, respectively. Here, L_1 denotes the ℓ_1 loss, and L_{perc} represents the perceptual loss. Throughout this paper, the subscript "gt" indicates the ground truth. The perceptual loss is defined as:

$$L_{\text{perc}}(\mathbf{S}_0^{\text{t}}, \mathbf{S}_0^{\text{gt}}) = L_2(\phi_h(\mathbf{S}_0^{\text{t}}), \phi_h(\mathbf{S}_0^{\text{gt}})), \quad (29)$$

where L_2 denotes the ℓ_2 loss, and ϕ_h is the feature map extracted from the h-th layer of a VGG-19 network [45] pretrained on ImageNet [43]. In our implementation, features from the $VGG3,3$ convolutional layer are used.

Polarization loss. The polarization loss L_{f_2} is defined as:

$$\begin{aligned} L_{f_2} = {} & \lambda_3 L_1(\{\mathbf{x}', \mathbf{y}'\}, \{\mathbf{x}^{\text{gt}}, \mathbf{y}^{\text{gt}}\}) + \lambda_4 L_{\text{tv}}(\{\mathbf{x}', \mathbf{y}'\}) \\ & + \lambda_5 R_1(\{\mathbf{x}', \mathbf{y}'\}, \{\mathbf{x}^{\text{gt}}, \mathbf{y}^{\text{gt}}\}), \end{aligned} \quad (30)$$

where $\lambda_{3,4,5}$ are set to 1.0, 0.5, and 1.0, respectively. Here, L_{tv} denotes the total variation loss, and R_1 stands for a polarization-based regularization term, which is defined as:

$$R_1(\{\mathbf{x}', \mathbf{y}'\}, \{\mathbf{x}^{\text{gt}}, \mathbf{y}^{\text{gt}}\}) = L_2(\mathbf{x}' \cdot \mathbf{y}^{\text{gt}}, \mathbf{y}' \cdot \mathbf{x}^{\text{gt}}). \quad (31)$$

Refinement loss. The refinement loss L_{f_3} is defined as:

$$\begin{aligned} L_{f_3} = {} & \lambda_6 L_1(\mathbf{I}_{\alpha_{1,2,3,4}}, \mathbf{I}^{\text{gt}}_{\alpha_{1,2,3,4}}) + \lambda_7 R_2(\mathbf{I}_{\alpha_{1,2,3,4}}) \\ & + \lambda_8 R_3(\mathbf{I}_{\alpha_{1,2,3,4}}, \mathbf{I}^{\text{gt}}_{\alpha_{1,2,3,4}}), \end{aligned} \quad (32)$$

where $\lambda_{6,7,8}$ are set to 10.0 and 100.0, and 1000.0, respectively. Here, $R_{2,3}$ stand for another two polarization-based regularization terms, which are defined as:

$$\begin{aligned} R_2(\mathbf{I}_{\alpha_{1,2,3,4}}) &= L_2(\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_3}, \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_4}) \quad \text{and} \\ R_3(\mathbf{I}_{\alpha_{1,2,3,4}}, \mathbf{I}^{\text{gt}}_{\alpha_{1,2,3,4}}) &= L_2(\mathbf{I}_{\alpha_4} - \mathbf{I}_{\alpha_2}, \mathbf{I}^{\text{gt}}_{\alpha_3} - \mathbf{I}^{\text{gt}}_{\alpha_1}) \\ & \quad + L_2(\mathbf{I}_{\alpha_3} - \mathbf{I}_{\alpha_1}, \mathbf{I}^{\text{gt}}_{\alpha_4} - \mathbf{I}^{\text{gt}}_{\alpha_2}), \end{aligned} \quad (33)$$

respectively.

B. Dataset Preparation

Since public polarization datasets (such as [19], [20]) cannot fit our specific setting, and existing fusion methods [4], [38], [63], [64] provide only images without polarization information, we construct a synthetic dataset tailored to our task. The generation process consists of the following steps:

(1) We use the PLIE dataset [70] as the data source. It provides short-exposed noisy polarized snapshots, serving as $\mathcal{L}$, and their corresponding high-quality reference snapshots, serving as $\mathcal{I}$.
(2) We adopt the method from [71] to synthesize polarized snapshots affected by motion blur, denoted as $\mathcal{B}_{\text{temp}}$. During this process, we introduce impulsive variations in the motion trajectories [2], [68] to generate more severe and diverse motion blur patterns.
(3) To simulate saturation, we adjust the exposure of $\mathcal{B}_{\text{temp}}$ using the technique proposed in [66], thereby producing $\mathcal{B}$. Note that our preliminary work PolFusion [67] does not involve this step and directly treat $\mathcal{B}_{\text{temp}}$ as $\mathcal{B}$.
(4) We perform brightness alignment of both $\mathcal{L}$ and $\mathcal{B}$ to match the reference image $\mathcal{I}$.

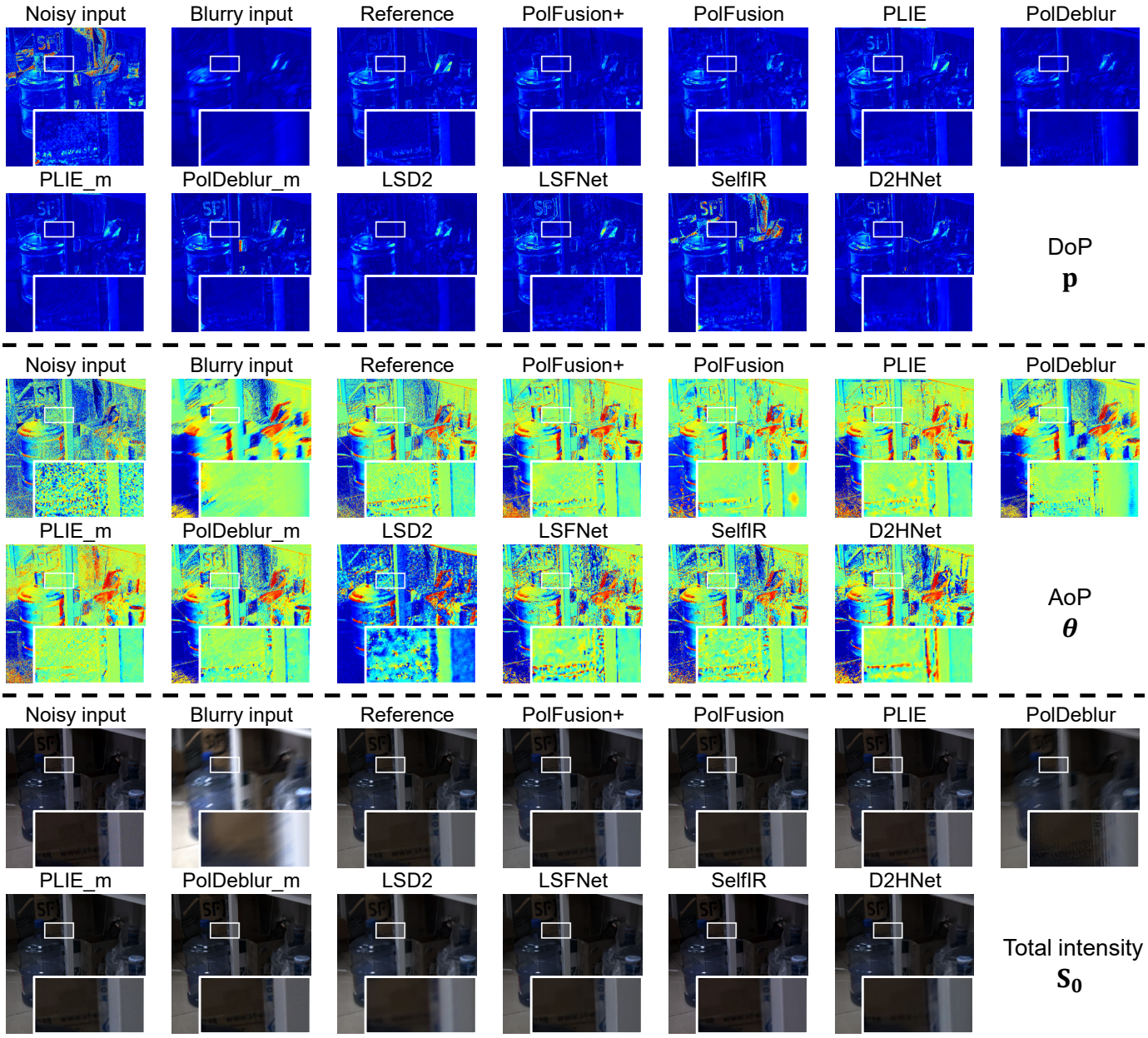


Fig. 6. Qualitative comparisons on synthetic data. See the caption of Table I for explanation.

The training and test splits follow the original partition of the PLIE dataset [70]. For each scene in the training (test) set, we randomly generate 3 (10) different camera shake trajectories. To further increase the diversity of the training data, we apply standard augmentation techniques such as flipping and rotation, resulting in a total of 7500 training images and 300 test images. Note that data augmentation is applied only during training. All images are resized and randomly cropped to 256×256 pixels for training and 512×512 pixels for testing.

C. Training Strategy

We implement our PolFusion+ model using PyTorch and train it on two NVIDIA RTX 4090 GPUs, following a two-stage training strategy. In the first stage, to ensure stable initialization, we independently train the irradiance restoration and polarization reconstruction phases for 300 epochs, using learning rates of 0.01 and 0.0001, respectively. In the second stage, we jointly train all three phases for an additional 300 epochs with a unified learning rate of 0.0001. During this joint training, the loss terms $L_{f_{1,2,3}}$ are weighted by 5.0, 10.0, and 10.0, respectively.

For optimization, we use the Adam optimizer [23] with parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. To introduce non-linearity, each convolutional layer in the network is followed by an instance normalization layer [51] and a $ReLU$ activation function.

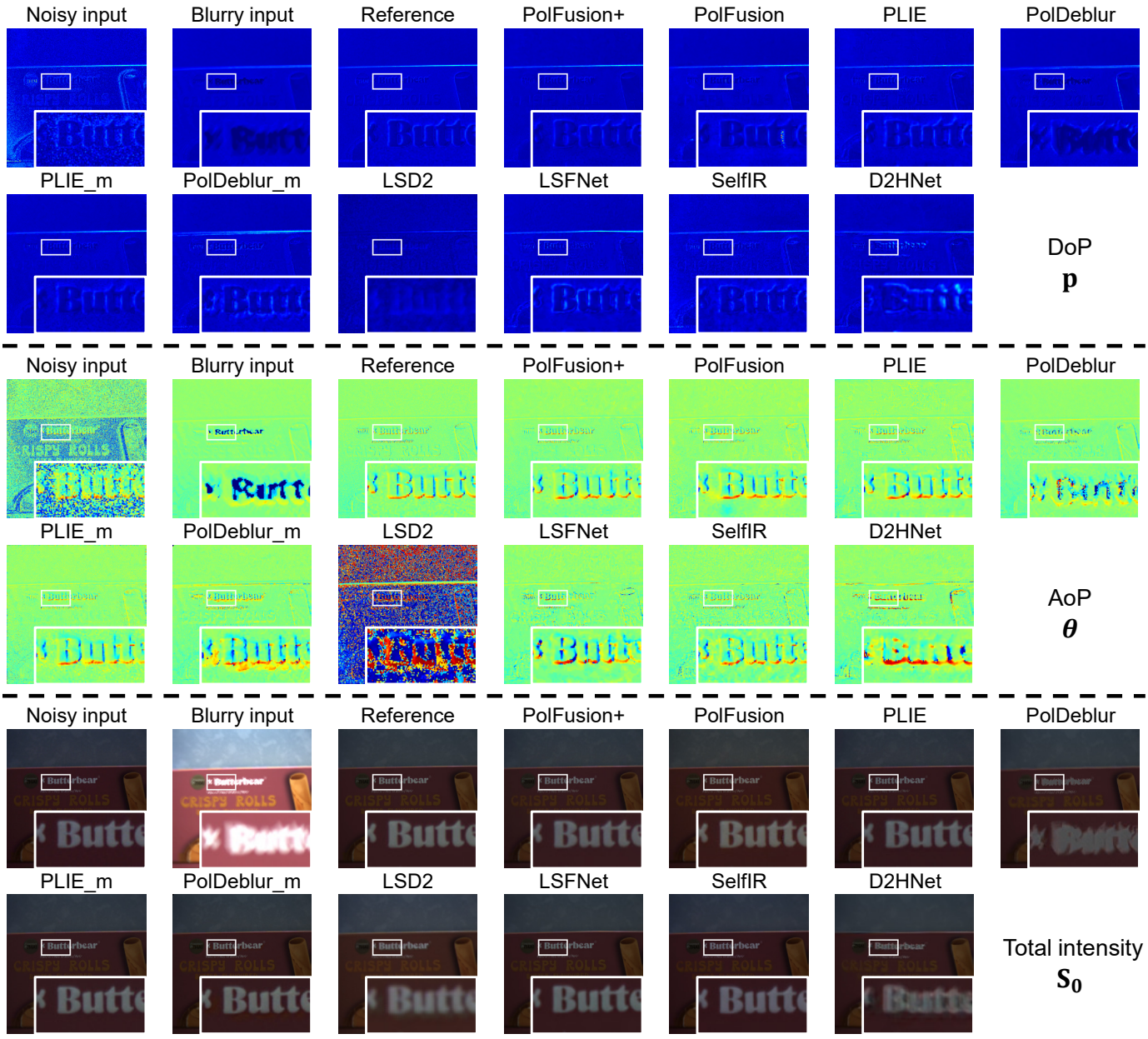


Fig. 7. Qualitative comparisons on real data. See the caption of Table I for explanation.

VII. Experiments

A. Evaluation on Synthetic Data

To evaluate the performance of our PolFusion+, we compare it against the following methods:

(1) Our preliminary work, PolFusion [67].
(2) Two state-of-the-art polarimetric imaging approaches: PLIE [70] for low-light enhancement, and PolDeblur [71] for deblurring.
(3) Modified versions of PLIE [70] and PolDeblur [71], denoted as PLIE_m and PolDeblur_m, respectively. These variants are adapted to fuse a noisy and blurry pair of polarized snapshots by introducing minor modifications to their original input structures.
(4) Four state-of-the-art fusion methods developed for conventional images: LSD2 [38], LSFNet [4], SelfIR [63], and D2HNet [64].

All comparison methods are retrained on our dataset (see Section VI-B) to ensure a fair evaluation. Following our preliminary work [67], we assess not only the quality of the DoP $\mathbf{p}$ and AoP $\boldsymbol{\theta}$, but also the total intensity $\mathbf{S}_0$.

To quantitatively evaluate the results, we adopt two widely used metrics: PSNR and SSIM. The results are reported in Table I, where our PolFusion+ consistently outperforms all compared methods across both metrics. Visual quality comparisons are provided in Figure 6[2], with zoom-in regions selected for detailed analysis. For the DoP $\mathbf{p}$ and AoP $\boldsymbol{\theta}$, our preliminary work PolFusion [67] exhibits noticeable cluster-like artifacts on the right side, while our PolFusion+ produces cleaner and sharper edges, thanks to its polarization-aware

[2]Additional results can be found in the supplementary material.

TABLE II
QUANTITATIVE EVALUATION RESULTS OF ABLATION STUDY.

	PSNR (DoP $\mathbf{p}$)	SSIM (DoP $\mathbf{p}$)	PSNR (AoP $\boldsymbol{\theta}$)	SSIM (AoP $\boldsymbol{\theta}$)	PSNR (total intensity $\mathbf{S}_0$)	SSIM (total intensity $\mathbf{S}_0$)
End2end	25.98	0.802	16.87	0.405	39.87	0.984
Phase1 from PolFusion	29.39	0.817	17.05	0.402	39.62	0.982
SelfIR [63] as Phase1	29.16	0.794	16.95	0.379	39.03	0.984
Phase2 from PolFusion	29.25	0.802	17.09	**0.406**	**40.45**	**0.985**
W/o Phase3	29.37	0.814	17.01	0.394	40.22	0.983
W/o Cartesian	29.24	0.805	16.97	0.391	38.99	**0.985**
W/o R_3	**29.42**	**0.819**	**17.11**	0.399	40.42	**0.985**
Complete PolFusion+	**29.59**	**0.825**	**17.12**	**0.407**	**40.47**	**0.986**

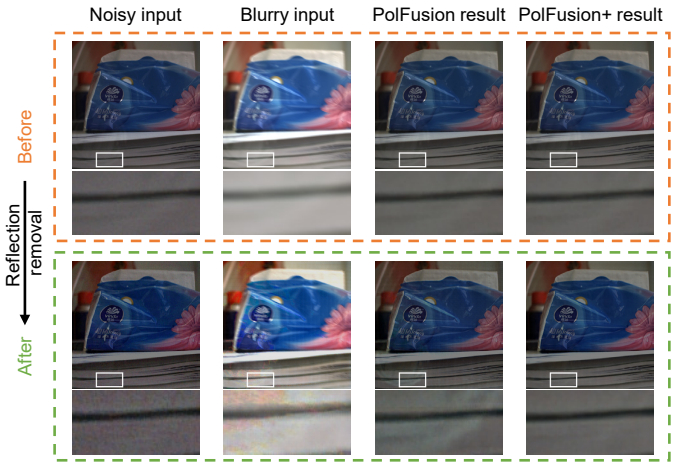


Fig. 8. Results of reflection removal (using RSP [35]). Please zoom-in for better details.

complementary fusion and explicit handling of saturation. In contrast, PLIE [70] suffers from zigzag artifacts on the left side, and PolDeblur [71] fails to fully eliminate blur. Their modified versions (PLIE_m and PolDeblur_m) yield improved results but still fall short compared to PolFusion+. Meanwhile, fusion methods developed for conventional images such as LSD2 [38], LSFNet [4], SelfIR [63], and D2HNet [64] produce less reliable outputs due to their lack of polarization awareness. Regarding the total intensity $\mathbf{S}_0$, results from both PolFusion+, PolFusion [67], PLIE_m, PolDeblur_m, and SelfIR [63] closely resemble the reference. In contrast, PLIE [70] produces darker global tone, PolDeblur [71], LSD2 [38], and D2HNet [64] fail to remove blur effectively, and LSFNet [4] exhibits ringing artifacts. These results collectively demonstrate that our PolFusion+ not only restores superior polarimetric parameters ($\mathbf{p}$ and $\boldsymbol{\theta}$), but also ensures the fidelity of the total intensity $\mathbf{S}_0$.

B. Evaluation on Real Data and Downstream Application

To assess the generalization capability, we collect several pairs of noisy and blurry polarized snapshots across diverse real-world scenes using a Lucid Vision Phoenix polarization camera. Here, for reference acquisition, the camera is mounted on a tripod to avoid motion blur, and the exposure time is carefully adjusted to obtain well-exposed polarized snapshots without saturation. Visual quality comparisons are presented in Figure 7[3], with zoom-in regions selected for detailed analysis. For the DoP $\mathbf{p}$, our preliminary work PolFusion [67] introduces noise-like artifacts between the two characters "t", whereas our PolFusion+ eliminates these artifacts. Other compared methods suffer from either zigzag distortions or noticeable blur. For the AoP $\boldsymbol{\theta}$, PolFusion [67] produces cluster-like artifacts at the bottom of the character "B," while the result from our PolFusion+ closely aligns with the reference. In contrast, the other methods exhibit varying degrees of ringing artifacts. As for the total intensity $\mathbf{S}_0$, our PolFusion+ achieves the most faithful global tone compared to the reference. These results demonstrate that PolFusion+ performs effectively on real-world data, not only in recovering the polarimetric parameters ($\mathbf{p}$ and $\boldsymbol{\theta}$) but also in preserving the total intensity $\mathbf{S}_0$, highlighting its strong generalization capability.

To demonstrate the practical benefits of our PolFusion+ in polarization-based downstream applications, we select a representative task in computer vision: reflection removal. Specifically, we adopt the reflection removal method RSP [35] and evaluate its performance when applied to four different inputs: the short-exposed noisy polarized snapshot, the long-exposed blurry polarized snapshot, the fused result from our preliminary method PolFusion [67], and the fused result produced by PolFusion+. Each input is independently processed by RSP, and the corresponding results before and after reflection removal are shown in Figure 8. As illustrated, the image restored from PolFusion+ result exhibits finer textures and significantly fewer reflection artifacts than those obtained from the noisy or blurry inputs. In contrast, the result from PolFusion [67] suffers from intensified artifacts after reflection removal, further highlighting the superiority of our PolFusion+.

C. Ablation Study

To validate the effectiveness of our overall design as well as the improvements over our preliminary work PolFusion [67], we conduct a series of ablation studies summarized as follows:

(1) End2end: Train the entire network in an end-to-end manner (to evaluate the benefit of our polarization-aware three-phase fusion strategy).
(2) Phase1 from PolFusion: Replace Phase1 with that in our previous work PolFusion [67] (to assess the improvements introduced in the module designs of Phase1).
(3) SelfIR [63] as Phase1: Replace Phase1 with SelfIR [63], a strong fusion method for conventional images that

[3]Please see the supplementary material for additional results.

outperforms other similar methods [4], [63], [64] in $\mathbf{S}_0$ restoration (to highlight the advantage of leveraging $\mathbf{S}_{1,2}$ as guidance in Phase1).

(4) Phase2 from PolFusion: Replace Phase2 with that in our previous work PolFusion [67] (to verify the effectiveness of our redesigned Phase2, particularly its ability to jointly model the individual characteristics and interdependencies of DoP and AoP, which is crucial for improving physical validity).

(5) W/o Phase3: Remove Phase3 (to examine the necessity of detail refinement for artifact suppression).

(6) W/o Cartesian: Directly reconstruct the DoP and AoP in Phase2 (to demonstrate the advantage of using the Cartesian coordinate representation).

(7) W/o R_3: Remove R_3 from the refinement loss L_{f_3} (to show the contribution of this additional loss term).

Quantitative comparisons are presented in Table II, which confirm that each proposed component contributes to the final performance, and the full model achieves the best overall results.

D. Computational Complexity Analysis

To quantitatively evaluate the computational complexity of our method, we report the number of parameters, FLOPs, and inference time for different methods using a single NVIDIA RTX 4090 GPU. The results are summarized in Table III. The input image size is set to 512×512 pixels with 3 color channels, and the inference time is measured as the average time required to reconstruct a single scene over all 300 test scenes. For conventional image fusion methods (LSD2 [38], LSFNet [4], SelfIR [63], and D2HNet [64]) that are not designed to process multiple polarized images simultaneously, we report the total time required to process all four polarized snapshots as the effective inference time. As shown in the table, our PolFusion+ exhibits moderate computational complexity, with fewer parameters and lower FLOPs compared to our preliminary method PolFusion [67], thanks to the redesigned network modules. The inference speed demonstrates that PolFusion+ is capable of real-time polarimetric imaging. To further improve efficiency, potential simplification strategies include structured pruning of redundant convolutional filters or channel-wise quantization of feature maps, which have been widely used in deep learning-based image reconstruction to reduce computational load with minimal accuracy loss.

TABLE III
QUANTITATIVE EVALUATION RESULTS OF COMPUTATIONAL COMPLEXITY. SEE THE CAPTION OF TABLE I FOR EXPLANATION. INPUT IMAGES ARE 512×512 PIXELS WITH 3 COLOR CHANNELS, AND INFERENCE TIME REPRESENTS THE AVERAGE TIME TO RECONSTRUCT A SINGLE SCENE.

	Number of parameters	FLOPs	Inference time
PolFusion+	4.72 M	337.24 G	0.021 s
PolFusion [67]	11.01 M	507.31 G	0.022 s
PLIE [70]	1.50 M	453.94 G	0.016 s
PolDeblur [71]	3.14 M	360.20 G	0.015 s
PLIE_m	1.64 M	473.79 G	0.018 s
PolDeblur_m	3.30 M	381.82 G	0.016 s
LSD2 [38]	31.03 M	446.44 G	0.262 s
LSFNet [4]	8.38 M	75.80 G	0.112 s
SelfIR [63]	1.08 M	106.31 G	0.083 s
D2HNet [64]	68.13 M	619.49 G	0.416 s

VIII. CONCLUSION

We propose PolFusion+, a unified complementary fusion framework for robust polarimetric imaging. Building upon and redesigning the polarization-aware three-phase fusion scheme and the degradation-oriented network of the preliminary work PolFusion [67], PolFusion+ introduces several key improvements to the execution of each fusion phase and the corresponding network modules. These improvements enable PolFusion+ to effectively handle saturation by extracting color information from both blurry and noisy polarized snapshots, and to ensure the accuracy of the DoP and AoP by explicitly modeling their individual characteristics and interrelationships. Experimental results demonstrate that our PolFusion+ not only achieves state-of-the-art performance on both synthetic and real data, but also effectively benefits downstream applications.

Limitations. Since our PolFusion+ is specifically designed to reconstruct a single high-quality polarized snapshot from a degraded pair of noisy and blurry polarized inputs, it is not applicable to polarized video reconstruction. In addition, it cannot be used to fuse conventional RGB images, as the first phase of our PolFusion+ relies on the availability of Stokes parameters, which are not accessible in standard RGB settings.

ACKNOWLEDGEMENT

This work was supported in part by JST-Mirai Program (Grant No. JPMJM123G1), National Natural Science Foundation of China (Grant No. 62136001, 62276007), Beijing Natural Science Foundation (Grant No. L233024), and Beijing Municipal Science & Technology Commission, Administrative Commission of Zhongguancun Science Park (Grant No. Z241100003524012).

REFERENCES

[1] Yunhao Ba, Alex Gilbert, Franklin Wang, Jinfa Yang, Rui Chen, Yiqin Wang, Lei Yan, Boxin Shi, and Achuta Kadambi. Deep shape from polarization. In *Proc. of European Conference on Computer Vision*, pages 554–571, 2020.

[2] Giacomo Boracchi and Alessandro Foi. Modeling the performance of image restoration from motion blur. *IEEE Transactions on Image Processing*, 21(8):3502–3517, 2012.

[3] Xu Cao, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Multi-view azimuth stereo via tangent space consistency. In *Proc. of Computer Vision and Pattern Recognition*, pages 825–834, 2023.

[4] Meng Chang, Huajun Feng, Zhihai Xu, and Qi Li. Low-light image restoration with short-and long-exposure raw pairs. *IEEE Transactions on Multimedia*, 24:702–714, 2021.

[5] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proc. of Computer Vision and Pattern Recognition*, pages 3291–3300, 2018.

[6] Byeong-Doo Choi, Seung-Won Jung, and Sung-Jea Ko. Motion-blur-free camera system splitting exposure time. *IEEE Transactions on Consumer Electronics*, 54(3):981–986, 2008.

[7] Akshat Dave, Yongyi Zhao, and Ashok Veeraraghavan. PANDORA: Polarization-aided neural decomposition of radiance. In *Proc. of European Conference on Computer Vision*, pages 538–556, 2022.

[8] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[9] Chunzhi Gu, Xuequan Lu, Ying He, and Chao Zhang. Blur removal via blurred-noisy image pair. *IEEE Transactions on Image Processing*, 30:345–359, 2020.

[10] Yuxuan Guo, Xiaobing Dai, Shaoju Wang, Guang Jin, and Xuemin Zhang. Attention-based progressive discrimination generative adversarial networks for polarimetric image demosaicing. *IEEE Transactions on Computational Imaging*, 10:713–725, 2024.

[11] Chao He, Honghui He, Jintao Chang, Binguo Chen, Hui Ma, and Martin J Booth. Polarisation optics for biomedical and clinical applications: a review. *Light: Science & Applications*, 10(1):194, 2021.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. of Computer Vision and Pattern Recognition*, 2016.

[13] Eugene Hecht. *Optics*. Pearson Education India, 2012.

[14] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[15] Haofeng Hu, Yang Lin, Xiaobo Li, Pengfei Qi, and Tiegen Liu. IPLNet: A neural network for intensity-polarization imaging in low light. *Optics Letters*, 45(22):6162–6165, 2020.

[16] Haofeng Hu, Shiyao Yang, Xiaobo Li, Zhenzhou Cheng, Tiegen Liu, and Jingsheng Zhai. Polarized image super-resolution via a deep convolutional neural network. *Optics Express*, 31(5):8535–8547, 2023.

[17] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proc. of Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.

[18] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proc. of Computer Vision and Pattern Recognition*, 2017.

[19] Inseung Hwang, Kiseok Choi, Hyunho Ha, and Min H Kim. Benchmarking burst super-resolution for polarization images: Noise dataset and analysis. In *Proc. of International Conference on Computer Vision*, pages 24899–24909, 2025.

[20] Yujin Jeon, Eunsue Choi, Youngchan Kim, Yunseong Moon, Khalid Omer, Felix Heide, and Seung-Hwan Baek. Spectral and polarization vision: Spectro-polarimetric real-world dataset. In *Proc. of Computer Vision and Pattern Recognition*, pages 22098–22108, 2024.

[21] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Depth sensing using geometrically constrained polarization normals. *International Journal of Computer Vision*, 125(1):34–51, 2017.

[22] Agastya Kalra, Vage Taamazyan, Supreeth Krishna Rao, Kartik Venkataraman, Ramesh Raskar, and Achuta Kadambi. Deep polarization cues for transparent object segmentation. In *Proc. of Computer Vision and Pattern Recognition*, pages 8602–8611, 2020.

[23] Diederik P Kingma and Jimmy Ba. ADAM: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[24] GP Können. *Polarized light in nature*. CUP Archive, 1985.

[25] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In *Proc. of Computer Vision and Pattern Recognition*, pages 8183–8192, 2018.

[26] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *Proc. of Computer Vision and Pattern Recognition*, pages 1750–1758, 2020.

[27] Ning Li, Yongqiang Zhao, Quan Pan, Seong G Kong, and Jonathan Cheung-Wai Chan. Full-time monocular road detection using zero-distribution prior of angle of polarization. In *Proc. of European Conference on Computer Vision*, pages 457–473, 2020.

[28] Zhuoxiao Li, Haiyang Jiang, Mingdeng Cao, and Yinqiang Zheng. Polarized color image denoising. In *Proc. of Computer Vision and Pattern Recognition*, pages 9873–9882, 2023.

[29] Yupeng Liang, Ryosuke Wakaki, Shohei Nobuhara, and Ko Nishino. Multimodal material segmentation. In *Proc. of Computer Vision and Pattern Recognition*, pages 19800–19808, 2022.

[30] Kaixin Liu, Yingjie Li, Jianguo Yang, Yi Liu, and Yuan Yao. Generative principal component thermography for enhanced defect detection and analysis. *IEEE Transactions on Instrumentation and Measurement*, 69(10):8261–8269, 2020.

[31] Shumin Liu, Jiajia Chen, Yuan Xun, Xiaojin Zhao, and Chip-Hong Chang. A new polarization image demosaicking algorithm by exploiting inter-channel correlations with guided filtering. *IEEE Transactions on Image Processing*, 29:7076–7089, 2020.

[32] Yang Lu, Weihong Ren, Yiming Su, Zhen Zhang, Junchao Zhang, and Jiandong Tian. Polarization image demosaicking based on homogeneity space. *Optics and Lasers in Engineering*, 178:108179, 2024.

[33] Yang Lu, Jiandong Tian, Yiming Su, Yidong Luo, Junchao Zhang, and Chunhui Hao. A hybrid polarization image demosaicking algorithm based on inter-channel correlation. *IEEE Transactions on Computational Imaging*, 10:1400–1413, 2024.

[34] Yidong Luo, Junchao Zhang, Jianbo Shao, Jiandong Tian, and Jiayi Ma. Learning a non-locally regularized convolutional sparse representation for joint chromatic and polarimetric demosaicking. *IEEE Transactions on Image Processing*, 33:5029–5044, 2024.

[35] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Physics-guided reflection separation from a pair of unpolarized and polarized images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2151–2165, 2022.

[36] Haiyang Mei, Bo Dong, Wen Dong, Jiaxi Yang, Seung-Hwan Baek, Felix Heide, Pieter Peers, Xiaopeng Wei, and Xin Yang. Glass segmentation using intensity and spectral polarization cues. In *Proc. of Computer Vision and Pattern Recognition*, pages 12622–12631, 2022.

[37] Miki Morimatsu, Yusuke Monno, Masayuki Tanaka, and Masatoshi Okutomi. Monochrome and color polarization demosaicking based on intensity-guided residual interpolation. *IEEE Sensors Journal*, 21(23):26985–26996, 2021.

[38] Janne Mustaniemi, Juho Kannala, Jiri Matas, Simo Särkkä, and Janne Heikkilä. LSD2 - joint denoising and deblurring of short and long exposure images with CNNs. In *Proc. of British Machine Vision Conference*, 2020.

[39] Vy Nguyen, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi. Two-step color-polarization demosaicking network. In *Proc. of International Conference on Image Processing*, pages 1011–1015, 2022.

[40] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[41] Taishi Ono, Yuhi Kondo, Legong Sun, Teppei Kurita, and Yusuke Moriuchi. Degree-of-linear-polarization-based color constancy. In *Proc. of Computer Vision and Pattern Recognition*, pages 19740–19749, 2022.

[42] Simeng Qiu, Qiang Fu, Congli Wang, and Wolfgang Heidrich. Linear polarization demosaicking for monochrome and colour polarization focal plane arrays. In *Computer Graphics Forum*, pages 77–89, 2021.

[43] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.

[44] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Instant dehazing of images using polarization. In *Proc. of Computer Vision and Pattern Recognition*, pages I–I, 2001.

[45] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[46] Chang-Hwan Son, Hyunseung Choo, and Hyung-Min Park. Image-pair-based deblurring with spatially varying norms and noisy image updating. *Journal of Visual Communication and Image Representation*, 24(8):1303–1315, 2013.

[47] Chang-Hwan Son and Hyung-Min Park. A pair of noisy/blurry patches-based PSF estimation and channel-dependent deblurring. *IEEE Transactions on Consumer Electronics*, 57(4):1791–1799, 2011.

[48] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proc. of Computer Vision and Pattern Recognition*, pages 8174–8182, 2018.

[49] Chaoran Tian, Weihong Pan, Zimo Wang, Mao Mao, Guofeng Zhang, Hujun Bao, Ping Tan, and Zhaopeng Cui. DPS-Net: Deep polarimetric stereo depth estimation. In *Proc. of International Conference on Computer Vision*, pages 3569–3579, 2023.

[50] K-H Tsai, K-S Kim, and TF Morse. General solutions for stress-induced polarization in optical fibers. *Journal of lightwave technology*, 9(1):7–17, 1991.

[51] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[52] Sijia Wen, Yinqiang Zheng, and Feng Lu. A sparse representation based joint demosaicing method for single-chip polarized color sensor. *IEEE Transactions on Image Processing*, 30:4171–4182, 2021.

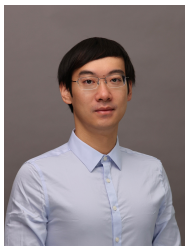
[53] Sijia Wen, Yinqiang Zheng, Feng Lu, and Qinping Zhao. Convolutional demosaicing network for joint chromatic and polarimetric imagery. *Optics Letters*, 44(22):5646–5649, 2019.

[54] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In *Proc. of European Conference on Computer Vision*, pages 3–19, 2018.

[55] Xuesong Wu, Hong Zhang, Xiaoping Hu, Moein Shakeri, Chen Fan, and Juiwen Ting. HDR reconstruction based on the polarization camera. *IEEE Robotics and Automation Letters*, 5(4):5113–5119, 2020.

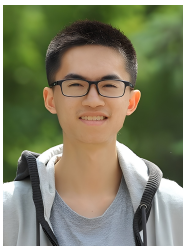
[56] Xiuyu Xu, Minjie Wan, Junyu Ge, Hao Chen, Xingcheng Zhu, Xiaojie

Zhang, Qian Chen, and Guohua Gu. ColorPolarNet: Residual dense network-based chromatic intensity-polarization imaging in low-light environment. *IEEE Transactions on Instrumentation and Measurement*, 71:1–10, 2022.
[57] Mingde Yao, Menglu Wang, King-Man Tam, Lingen Li, Tianfan Xue, and Jinwei Gu. PolarFree: Polarization-based reflection-free imaging. In *Proc. of Computer Vision and Pattern Recognition*, pages 10890–10899, 2025.
[58] Naifu Yao, Yang Guo, Yongqiang Zhao, and Seong G Kong. Subsurface defect detection in carbon fiber-reinforced plastics using infrared polarization imaging. *IEEE Transactions on Instrumentation and Measurement*, 74:1–16, 2025.
[59] Dabing Yu, Qingwu Li, Zhiliang Zhang, Guanying Huo, Chang Xu, and Yaqin Zhou. Color polarization image super-resolution reconstruction via a cross-branch supervised learning strategy. *Optics and Lasers in Engineering*, 165:107469, 2023.
[60] Lu Yuan, Jian Sun, Long Quan, and Heung-Yeung Shum. Image deblurring with blurred/noisy image pairs. *ACM Transactions on Graphics (Proc. of ACM SIGGRAPH)*, 26(3):1–es, 2007.
[61] Xianglong Zeng, Yuan Luo, Xiaojing Zhao, and Wenbin Ye. An end-to-end fully-convolutional neural network for division of focal plane sensors to reconstruct $\mathbf{S}_0$, DoLP, and AoP. *Optics Express*, 27(6):8566–8577, 2019.
[62] Junchao Zhang, Jianlai Chen, Hanwen Yu, Degui Yang, Buge Liang, and Mengdao Xing. Polarization image demosaicking via nonlocal sparse tensor factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–10, 2021.
[63] Zhilu Zhang, RongJian Xu, Ming Liu, Zifei Yan, and Wangmeng Zuo. Self-supervised image restoration with blurry and noisy pairs. In *Proc. of Advances in Neural Information Processing Systems*, pages 29179–29191, 2022.
[64] Yuzhi Zhao, Yongzhe Xu, Qiong Yan, Dingdong Yang, Xuehui Wang, and Lai-Man Po. D2HNet: Joint denoising and deblurring with hierarchical network for robust night image restoration. In *Proc. of European Conference on Computer Vision*, pages 91–110, 2022.
[65] Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. In *Proc. of International Conference on Computer Vision*, pages 4439–4448, 2021.
[66] Chu Zhou, Yufei Han, Minggui Teng, Jin Han, Si Li, Chao Xu, and Boxin Shi. Polarization guided HDR reconstruction via pixel-wise depolarization. *IEEE Transactions on Image Processing*, 32:1774–1787, 2023.
[67] Chu Zhou, Yixing Liu, Chao Xu, and Boxin Shi. Quality-improved and property-preserved polarimetric imaging via complementarily fusing. In *Proc. of Advances in Neural Information Processing Systems*, pages 136018–136036, 2024.
[68] Chu Zhou, Minggui Teng, Jin Han, Jinxiu Liang, Chao Xu, Gang Cao, and Boxin Shi. Deblurring low-light images with events. *International Journal of Computer Vision*, 131(5):1284–1298, 2023.
[69] Chu Zhou, Minggui Teng, Yufei Han, Chao Xu, and Boxin Shi. Learning to dehaze with polarization. In *Proc. of Advances in Neural Information Processing Systems*, pages 11487–11500, 2021.
[70] Chu Zhou, Minggui Teng, Youwei Lyu, Si Li, Chao Xu, and Boxin Shi. Polarization-aware low-light image enhancement. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 3742–3750, 2023.
[71] Chu Zhou, Minggui Teng, Xinyu Zhou, Chao Xu, and Boxin Sh. Learning to deblur polarized images. *International Journal of Computer Vision*, 133(9):5976–5991, 2025.
[72] Chu Zhou, Chao Xu, and Boxin Shi. Polarization guided mask-free shadow removal. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 10716–10724, 2025.
[73] Chu Zhou, Yixin Yang, Junda Liao, Heng Guo, Boxin Shi, and Imari Sato. Polarimetric neural field via unified complex-valued wave representation. In *Proc. of International Conference on Computer Vision*, pages 25660–25669, 2025.
[74] Shuangfan Zhou, Chu Zhou, Youwei Lyu, Heng Guo, Zhanyu Ma, Boxin Shi, and Imari Sato. PIDSR: Complementary polarized image demosaicing and super-resolution. In *Proc. of Computer Vision and Pattern Recognition*, pages 16081–16090, 2025.
[75] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable ConvNets v2: More deformable, better results. In *Proc. of Computer Vision and Pattern Recognition*, pages 9308–9316, 2019.



Chu Zhou received the B.E. degree from Huazhong University of Science and Technology in 2019 and the Ph.D. degree from School of Intelligence Science and Technology, Peking University in 2024. He is currently an assistant professor by special appointment at Digital Contents and Media Sciences Research Division, National Institute of Informatics. His research interest includes computational photography and computer vision, with a focus on unconventional camera-based vision.

Yixing Liu is currently pursuing the B.S. degree in Computer Science at Peking University, Beijing, China. He is an undergraduate researcher at the National Engineering Research Center of Video Technology, School of Computer Science, Peking University. His research interests include computational imaging, polarimetric imaging, and event-based vision.



Minggui Teng received the B.S. degree from Peking University, Beijing, China, in 2021. He is currently working toward the Ph.D. degree with the National Engineering Research Center of Video Technology, School of Computer Science, Peking University. His research interests are focused on neuromorphic camera and image enhancement. He has served as a reviewer for CVPR, ICCV, ECCV, *etc*.

Chao Xu received the B.E. degree from Tsinghua University in 1988, the M.S. degree from University of Science and Technology of China in 1991 and the Ph.D. degree from Institute of Electronics, Chinese Academy of Sciences in 1997. Between 1991 and 1994 he was employed as an assistant researcher by University of Science and Technology of China. Since 1997 Dr. Xu has been at Peking University, and since 2005 he has become a Professor. His research interests are in image processing and computer vision. He has authored or co-authored around 200 publications which have been cited over 18 thousand times in terms of Google Scholar.

Boxin Shi received the B.E. degree from the Beijing University of Posts and Telecommunications, the M.E. degree from Peking University, and the Ph.D. degree from the University of Tokyo, in 2007, 2010, and 2013. He is currently a Boya Young Fellow Associate Professor (with tenure) and Research Professor at Peking University, where he leads the Camera Intelligence Lab. Before joining PKU, he did research with MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University, National Institute of Advanced Industrial Science and Technology, from 2013 to 2017. His papers were awarded as Best Paper, Runners-Up at CVPR 2024, ICCP 2015, and selected as Best Paper candidate at ICCV 2015. He is an associate editor of TPAMI/IJCV and an area chair of CVPR/ICCV/ECCV. He is a senior member of IEEE.

Imari Sato received the B.S. degree in policy management from Keio University, Tokyo, Japan, in 1994, and the M.S. and Ph.D. degrees in interdisciplinary Information Studies from the University of Tokyo, Tokyo, in 2002 and 2005, respectively. She was a visiting scholar with the Robotics Institute of Carnegie Mellon University, Pittsburgh, PA, USA. In 2005, she joined the National Institute of Informatics, where she is currently a professor/director of the Digital Contents and Media Sciences Research Division. She is a professor at the University of Tokyo and a visiting professor at the Tokyo Institute of Technology, Tokyo, Japan. Her primary research interests include physics-based vision, spectral analysis, image-based modeling, and medical image analysis. She received various research awards, including the Young Scientists' Prize from the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2009, and the Microsoft Research Japan New Faculty Award in 2011.

Supplementary Material: Towards a Unified Complementary Fusion Framework for Robust Polarimetric Imaging

Chu Zhou*, Yixing Liu, Minggui Teng, Chao Xu, Boxin Shi*, *Senior Member, IEEE* and Imari Sato, *Member, IEEE*

IX. Failure Case Analysis

Although PolFusion+ performs robustly across diverse scenarios, certain challenging conditions can still lead to noticeable failures. As shown in Figure 9, the left zoom-in region suffers from fast non-rigid motion accompanied by severe color bleeding, where the white and blue parts become heavily blended. In such cases, the long-exposure snapshot, which is normally expected to provide relatively reliable color information, becomes overly degraded. Although the short-exposure snapshot still provides structurally accurate edges, the strongly degraded color information in the long exposure limits the fidelity of the recovered DoP. In contrast, the right zoom-in region contains extremely dark materials, where both the short- and long-exposed inputs exhibit very low intensity. Under such conditions, the Stokes parameters become strongly noise-dominated, causing unstable estimation of both DoP and AoP. This issue stems from inherent signal limitations rather than the fusion strategy itself.

Possible remedies. For degradation caused by fast non-rigid motion, one potential remedy is to incorporate lightweight motion-aware consistency constraints that mitigate color bleeding without relying on explicit optical-flow estimation. For degradation in extremely low-intensity regions, a practical solution is to introduce noise-aware regularization or integrate a physics-based denoising prior to stabilize the estimation under near-zero illumination. These strategies are orthogonal to our fusion framework and could be incorporated in future work without altering its core design.

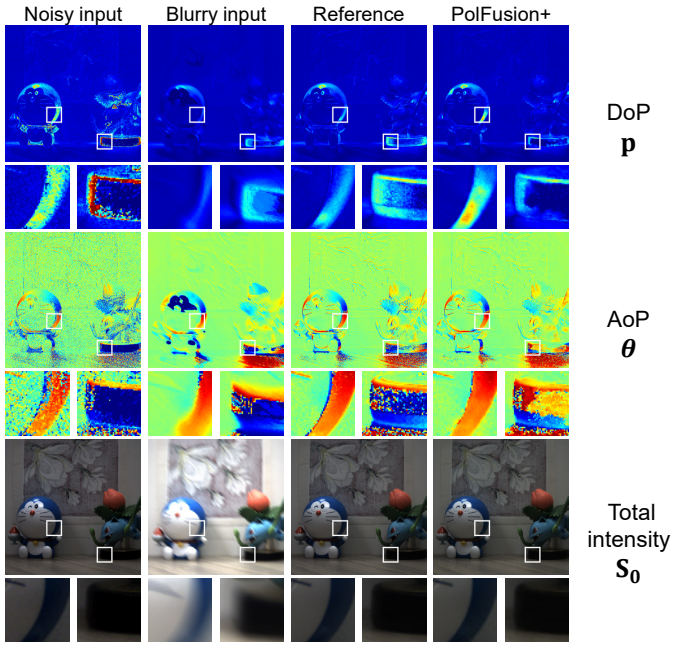


Fig. 9. A failure example of our PolFusion+, illustrating two representative challenging scenarios: (1) fast non-rigid motion with severe color bleeding (left zoom-in), and (2) extremely low-intensity regions (right zoom-in). Both conditions result in inaccurate DoP and AoP reconstruction.

X. Cross-Device and Cross-Scene Generalization

To further demonstrate the cross-device and cross-scene generalization capability of our PolFusion+, we provide an outdoor scene captured using a FLIR BFS-U3-51S5PC-C polarization camera, as shown in Figure 10. This camera model was entirely unseen during training and differs from our synthetic training data not only in sensor characteristics but also in scene type, since the synthetic dataset is largely composed of indoor scenes captured using a Lucid Vision Phoenix polarization camera. Despite these substantial differences, PolFusion+ consistently outperforms all compared methods, including our preliminary work PolFusion [7], two state-of-the-art polarimetric imaging approaches (PLIE [8] for low-light enhancement and PolDeblur [9] for deblurring) and their corresponding modified versions (PLIE_m and PolDeblur_m), four state-of-the-art fusion methods developed for conventional images (LSD2 [3], LSFNet [1], SelfIR [5], and D2HNet [6]) in reconstructing the DoP and AoP.

The generalization capability of PolFusion+ mainly stems from two aspects: (1) Physics-grounded modeling: all reconstruction stages are based on polarimetric imaging principles, which hold for all linear-response polarization cameras, thereby reducing reliance on device-specific appearance statistics. (2) Degradation-oriented design: rather than learning scene-specific textures, the network focuses on universal degradations,

* Corresponding authors: zhou_chu@hotmail.com and shiboxin@pku.edu.cn
Chu Zhou and Imari Sato are with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan.
Yixing Liu, Minggui Teng, and Boxin Shi are with the State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100080, China, and also with the National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing 100080, China.
Chao Xu is with the National Key Laboratory of General Artificial Intelligence, School of Intelligence Science and Technology, Peking University, Beijing 100080, China.

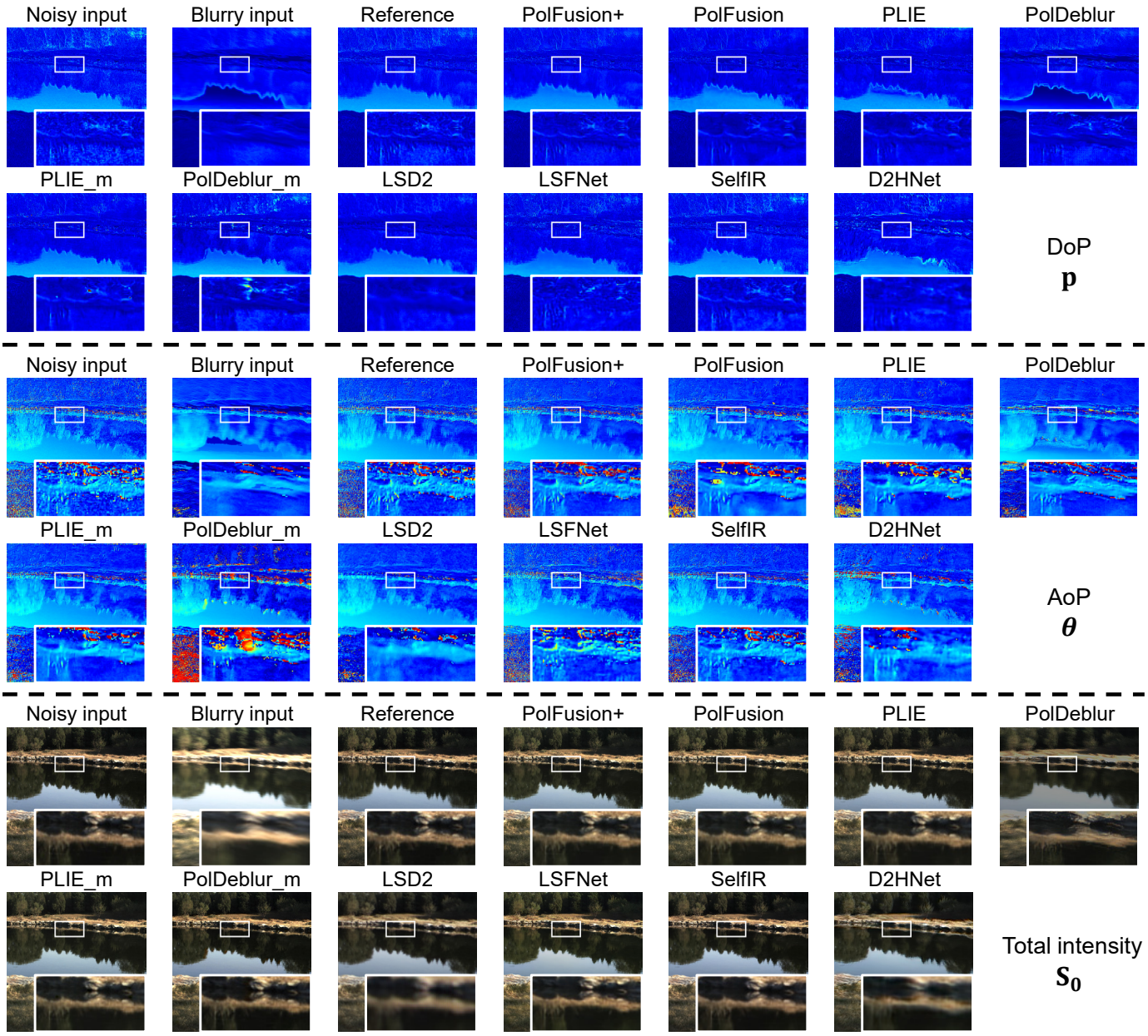


Fig. 10. Visual quality comparisons on an outdoor scene captured using a FLIR BFS-U3-51S5PC-C polarimetric camera, demonstrating the cross-device and cross-scene generalization capability of our PolFusion+.

producing representations that naturally transfer across different environments. Together, these factors enable robust performance across unseen devices and scenes.

XI. Additional Results of Downstream Application

In addition to the reflection removal results presented in the main paper using RSP [2], we further provide additional reflection removal results using a recent state-of-the-art approach, PolarFree [4]. As shown in Figure 11, the outcomes are consistent with those reported in the main paper.

XII. Additional Results on Synthetic Data

In this section, we provide additional visual quality comparisons on synthetic data, as shown in Figure 12, Figure 13, and Figure 14, corresponding to Footnote 2 of the paper. The comparisons involve our PolFusion+, our preliminary work PolFusion [7], two state-of-the-art polarimetric imaging approaches (PLIE [8] for low-light enhancement and PolDeblur [9] for deblurring) and their corresponding modified versions (PLIE_m and PolDeblur_m), four state-of-the-art fusion methods developed for conventional images (LSD2 [3], LSFNet [1], SelfIR [5], and D2HNet [6]).

XIII. Additional Results on Real Data

In this section, we provide additional visual quality comparisons on real data, as shown in Figure 15, Figure 16, and Figure 17, corresponding to Footnote 3 of the paper. The comparisons involve our PolFusion+, our preliminary

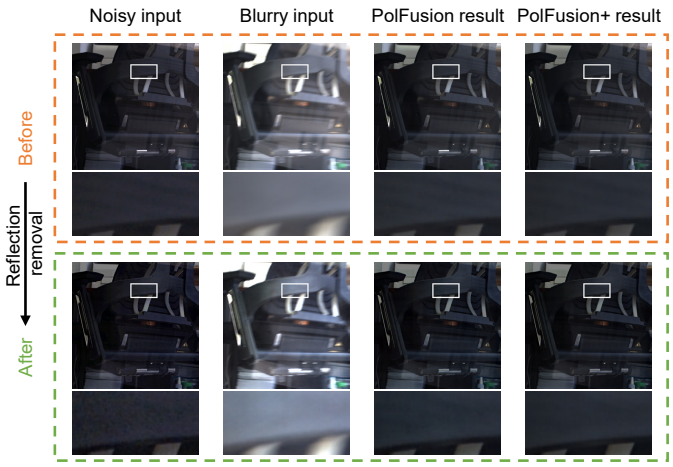


Fig. 11. Additional results of reflection removal (using PolarFree [4]). Please zoom-in for better details.

work PolFusion [7], two state-of-the-art polarimetric imaging approaches (PLIE [8] for low-light enhancement and PolDeblur [9] for deblurring) and their corresponding modified versions (PLIE_m and PolDeblur_m), four state-of-the-art fusion methods developed for conventional images (LSD2 [3], LSFNet [1], SelfIR [5], and D2HNet [6]).

REFERENCES

[1] Meng Chang, Huajun Feng, Zhihai Xu, and Qi Li. Low-light image restoration with short-and long-exposure raw pairs. *IEEE Transactions on Multimedia*, 24:702–714, 2021.

[2] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Physics-guided reflection separation from a pair of unpolarized and polarized images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2151–2165, 2022.

[3] Janne Mustaniemi, Juho Kannala, Jiri Matas, Simo Särkkä, and Janne Heikkilä. LSD2 - joint denoising and deblurring of short and long exposure images with CNNs. In *Proc. of British Machine Vision Conference*, 2020.

[4] Mingde Yao, Menglu Wang, King-Man Tam, Lingen Li, Tianfan Xue, and Jinwei Gu. PolarFree: Polarization-based reflection-free imaging. In *Proc. of Computer Vision and Pattern Recognition*, pages 10890–10899, 2025.

[5] Zhilu Zhang, RongJian Xu, Ming Liu, Zifei Yan, and Wangmeng Zuo. Self-supervised image restoration with blurry and noisy pairs. In *Proc. of Advances in Neural Information Processing Systems*, pages 29179–29191, 2022.

[6] Yuzhi Zhao, Yongzhe Xu, Qiong Yan, Dingdong Yang, Xuehui Wang, and Lai-Man Po. D2HNet: Joint denoising and deblurring with hierarchical network for robust night image restoration. In *Proc. of European Conference on Computer Vision*, pages 91–110, 2022.

[7] Chu Zhou, Yixing Liu, Chao Xu, and Boxin Shi. Quality-improved and property-preserved polarimetric imaging via complementarily fusing. In *Proc. of Advances in Neural Information Processing Systems*, pages 136018–136036, 2024.

[8] Chu Zhou, Minggui Teng, Youwei Lyu, Si Li, Chao Xu, and Boxin Shi. Polarization-aware low-light image enhancement. In *Proc. of the AAAI Conference on Artificial Intelligence*, pages 3742–3750, 2023.

[9] Chu Zhou, Minggui Teng, Xinyu Zhou, Chao Xu, and Boxin Sh. Learning to deblur polarized images. *International Journal of Computer Vision*, 133(9):5976–5991, 2025.

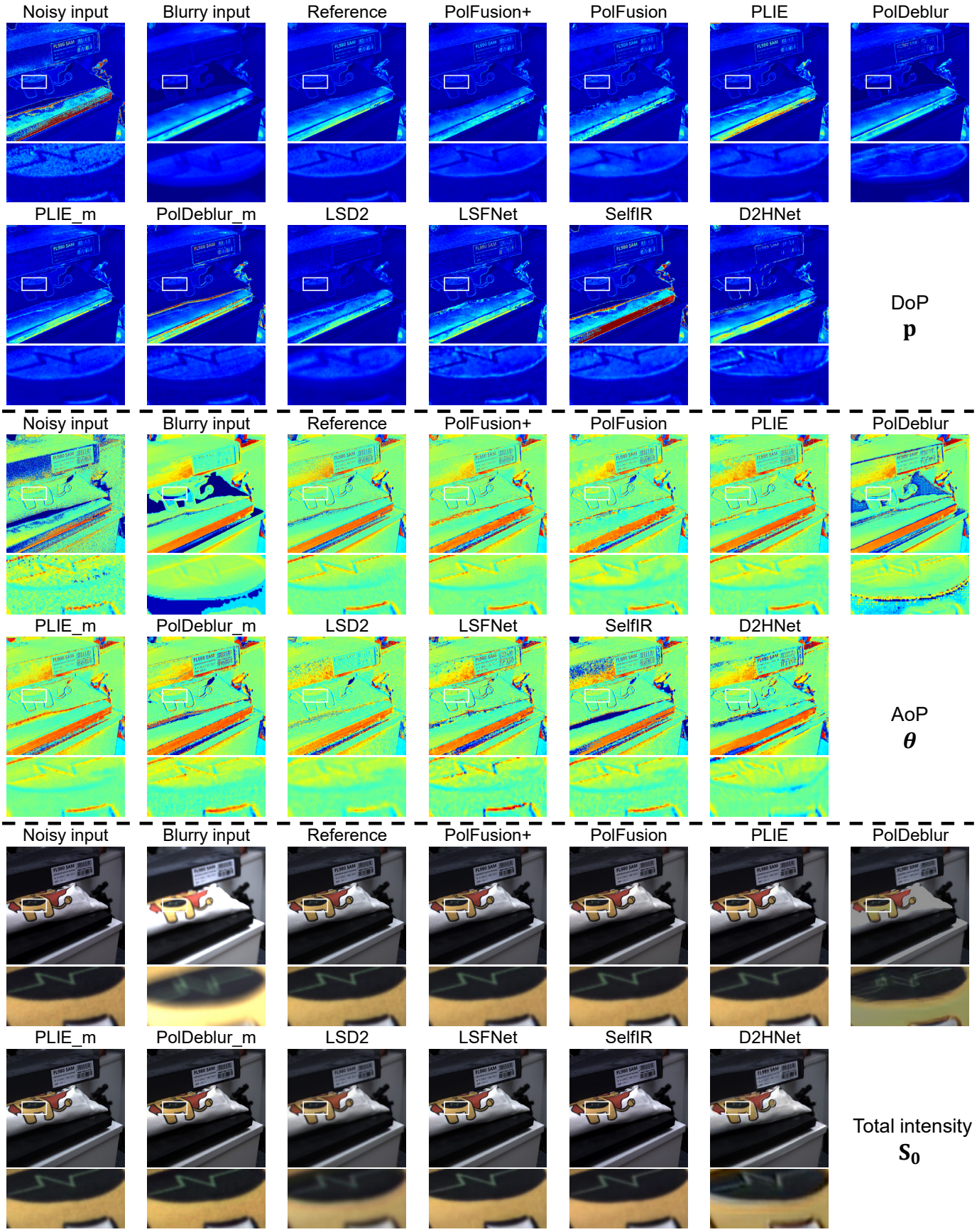


Fig. 12. Additional visual quality comparisons on synthetic data (part1).

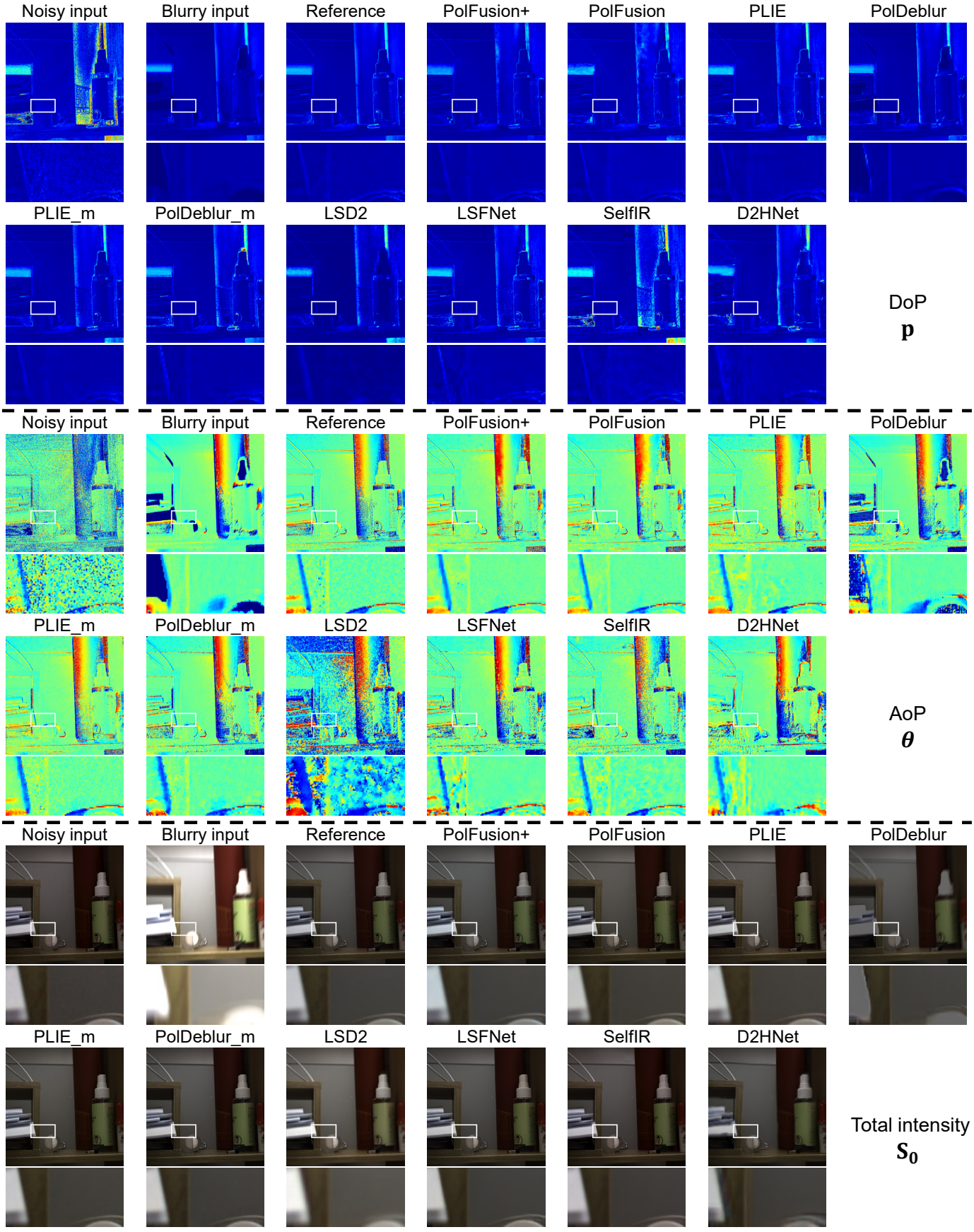


Fig. 13. Additional visual quality comparisons on synthetic data (part2).

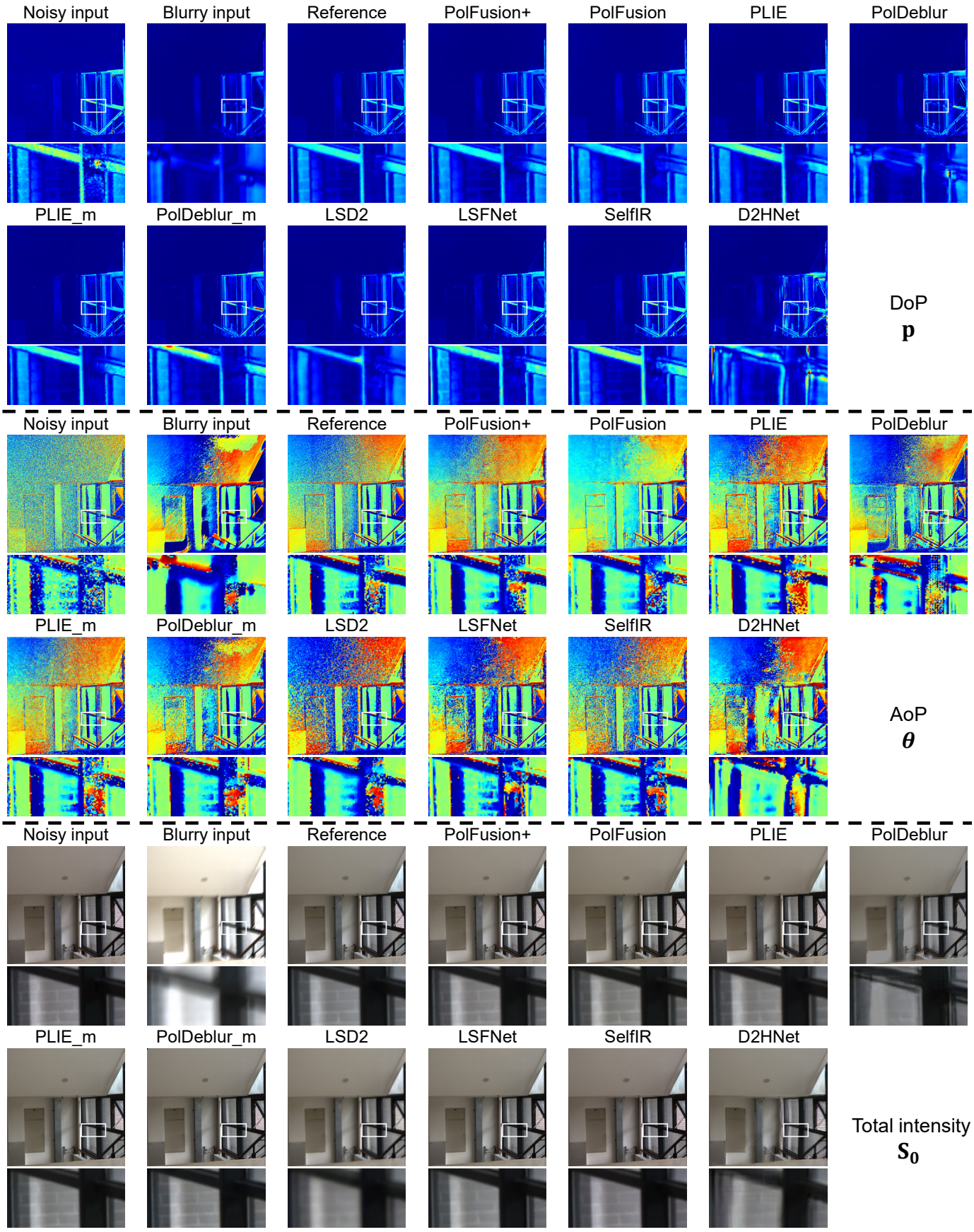


Fig. 14. Additional visual quality comparisons on synthetic data (part3).

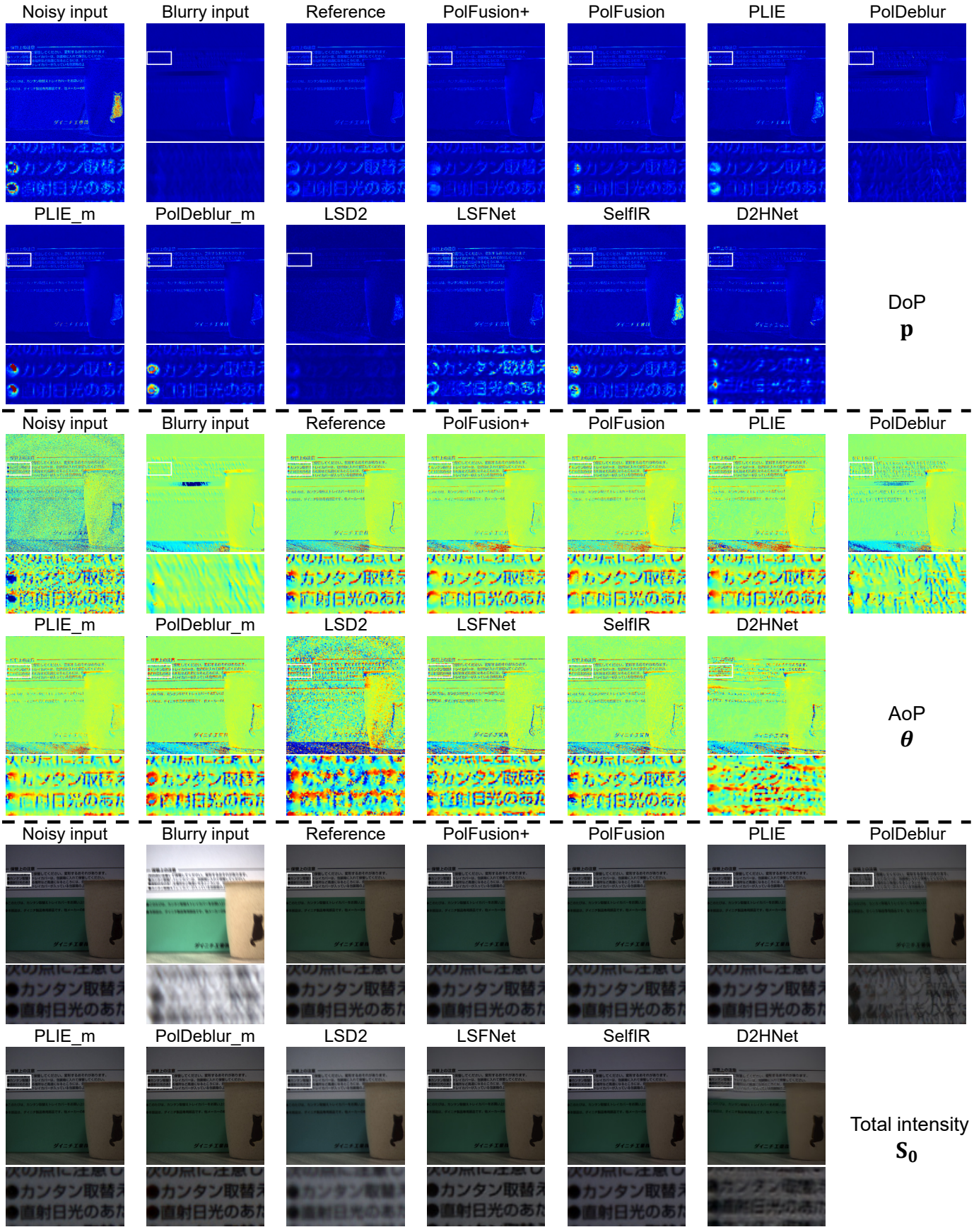


Fig. 15. Additional visual quality comparisons on real data (part1).

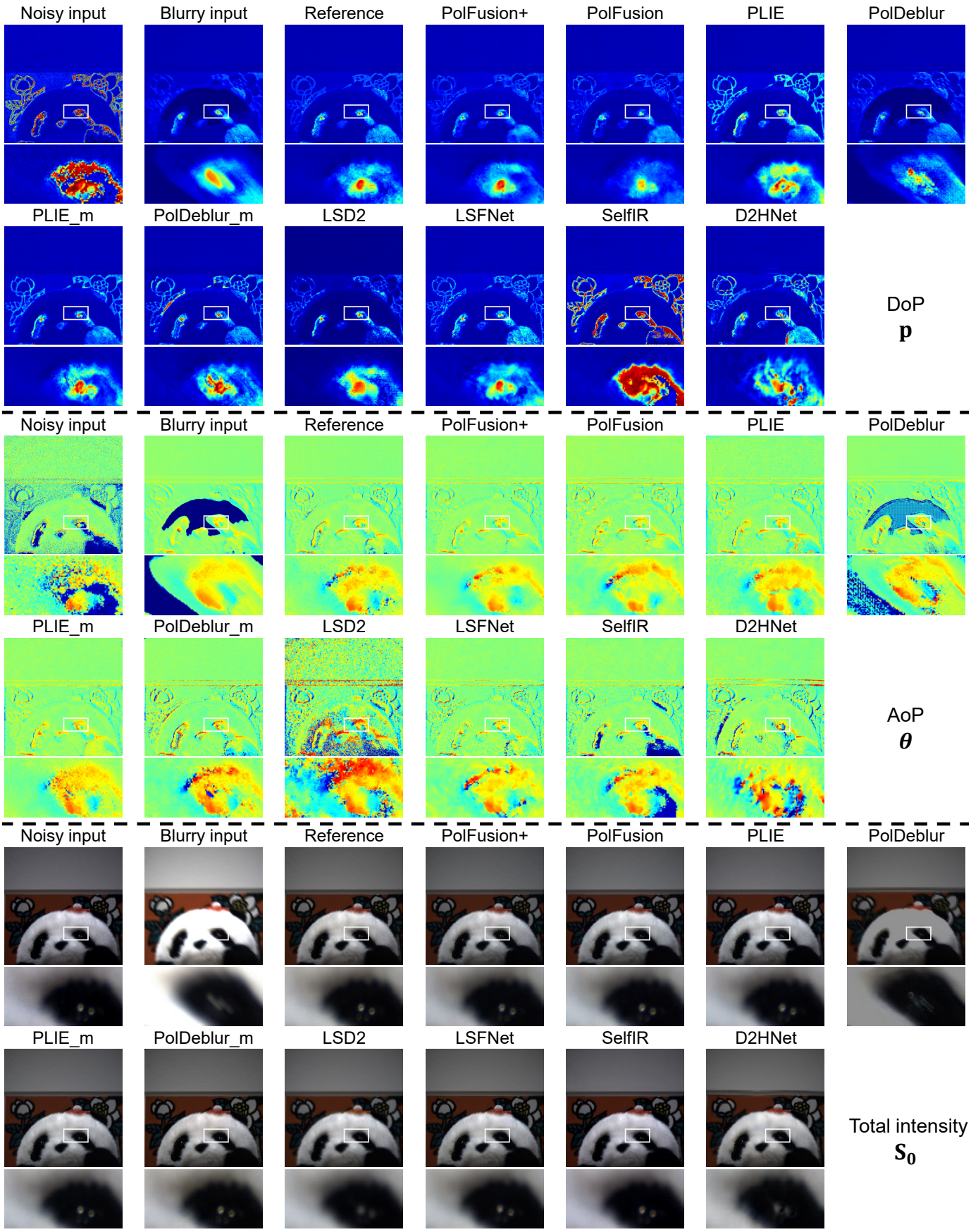


Fig. 16. Additional visual quality comparisons on real data (part2).

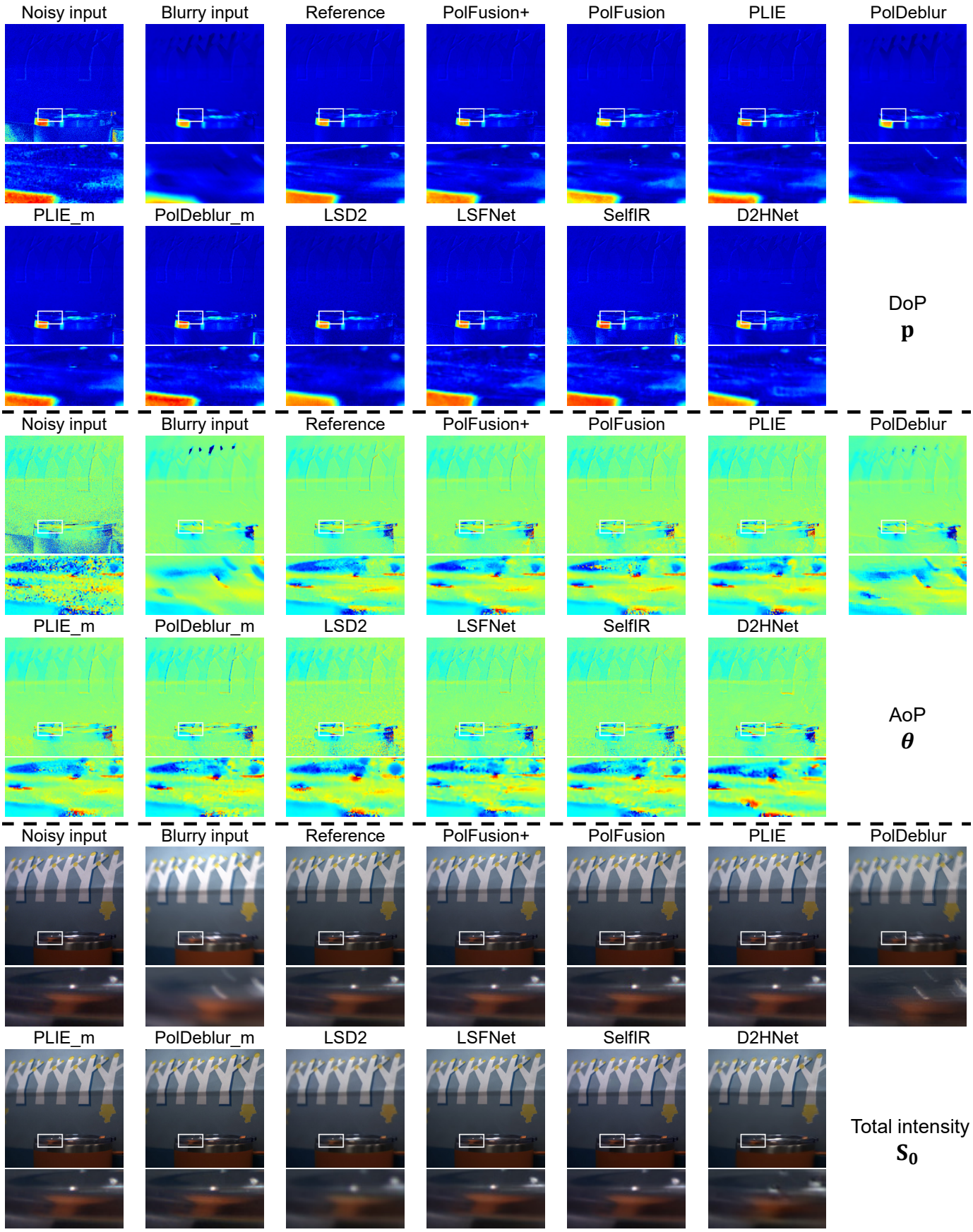


Fig. 17. Additional visual quality comparisons on real data (part3).