PIDSR: Complementary Polarized Image Demosaicing and Super-Resolution

Shuangfan Zhou[1#] Chu Zhou[2#] Youwei Lyu[1] Heng Guo[1*] Zhanyu Ma[1] Boxin Shi[3,4] Imari Sato[2]

[1] School of Artificial Intelligence, Beijing University of Posts and Telecommunications

[2] National Institute of Informatics

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

[4] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

{zhoushuangfan, youweilv, guoheng, mazhanyu}@bupt.edu.cn   zhou_chu@hotmail.com

shiboxin@pku.edu.cn   imarik@nii.ac.jp

Abstract

*Polarization cameras can capture multiple polarized images with different polarizer angles in a single shot, bringing convenience to polarization-based downstream tasks. However, their direct outputs are color-polarization filter array (CPFA) raw images, requiring demosaicing to reconstruct full-resolution, full-color polarized images; unfortunately, this necessary step introduces artifacts that make polarization-related parameters such as the degree of polarization (DoP) and angle of polarization (AoP) prone to error. Besides, limited by the hardware design, the resolution of a polarization camera is often much lower than that of a conventional RGB camera. Existing polarized image demosaicing (PID) methods are limited in that they cannot enhance resolution, while polarized image super-resolution (PISR) methods, though designed to obtain high-resolution (HR) polarized images from the demosaicing results, tend to retain or even amplify errors in the DoP and AoP introduced by demosaicing artifacts. In this paper, we propose **PIDSR**, a joint framework that performs complementary **P**olarized **I**mage **D**emosaicing and **S**uper-**R**esolution, showing the ability to robustly obtain high-quality HR polarized images with more accurate DoP and AoP from a CPFA raw image in a direct manner. Experiments show our PIDSR not only achieves state-of-the-art performance on both synthetic and real data, but also facilitates downstream tasks.*

1. Introduction

Polarization-based vision has benefited various applications, such as shape from polarization [1, 2], reflection removal [19], image dehazing [41], HDR imaging [42], *etc*. By fully utilizing the physical clues encoded in the polarization-

---

# Equal contribution.   * Corresponding author.
Code: https://github.com/PRIS-CV/PIDSR
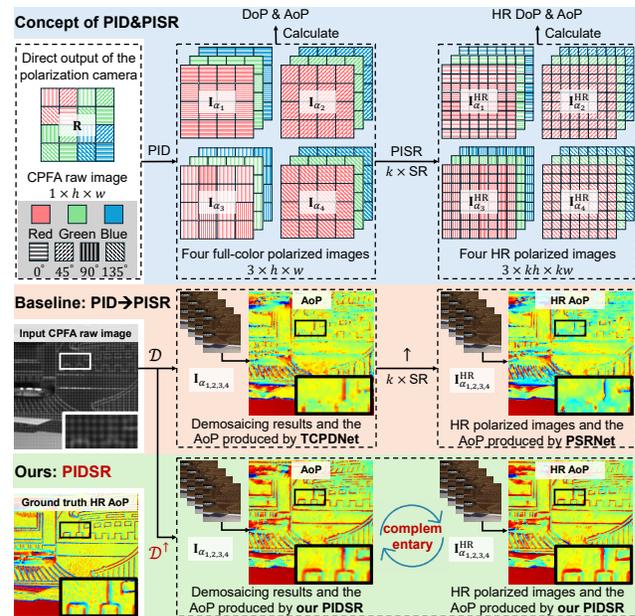
Figure 1. Top: The concept of polarized image demosaicing (PID) and polarized image super-resolution (PISR). Mid: An example shows that the baseline (PID→PISR) works in a sequential manner, where the AoPs calculated from the demosaicing results (produced by TCPDNet [23]) and the HR polarized images (produced by PSRNet [8]) suffer from severe artifacts. Bottom: An example shows that our PIDSR works in a complementary manner, where the calculated AoPs are more accurate. We choose $k = 4$ here.

relevant parameters such as the degree of polarization (DoP) and angle of polarization (AoP), polarization-based methods often achieve higher performance compared with the image-based ones, showing promising potentials. To acquire the DoP and AoP, at least three polarized images with different polarizer angles are required. While a polarizer can be used for this purpose, it demands multiple shots, making

the capture process quite inconvenient. Empowered by the division of focal plane (DoFP) technology, a polarization camera can capture four color polarized images with different polarizer angles ($0°, 45°, 90°, 135°$) in a single shot, bringing convenience to the acquisition of the DoP and AoP.

Since DoFP uses color-polarization filter array (CPFA) to record the color and polarization information simultaneously, the direct output of a polarization camera is a CPFA raw image. As shown in the left part of Fig. 1 (top), each pixel in a CPFA raw image contains information about only one color channel and one polarizer angle, which means that demosaicing is required to reconstruct the corresponding polarized images. Since unavoidable demosaicing artifacts tend to be amplified by the non-linearity of subsequent calculations, the DoP and AoP acquired from a polarization camera usually have a higher level of error than those acquired from a polarizer, making the physical clues less distinctive. Besides, limited by the hardware design, the resolution of a polarization camera is often much lower than that of a conventional RGB camera, restricting the fidelity of the recorded information. Thus, obtaining high-quality high-resolution (HR) polarized images with more accurate DoP and AoP from a polarization camera is of practical significance.

Despite its importance, a practical and reliable approach to simultaneously achieve demosaicing and super-resolution of polarization images has yet to be developed. As shown in Fig. 1 (top), the most straightforward way is to sequentially perform polarized image demosaicing (PID) [20, 22, 23] and polarized image super-resolution (PISR) [8, 37] on the CPFA raw image, *i.e.*, perform "PID→PISR". Defining $\mathbf{R} \in \mathbb{R}^{1 \times h \times w}$ (h and w are the height and width respectively) as the CPFA raw image, $\mathbf{I}_{\alpha_{1,2,3,4}} \in \mathbb{R}^{3 \times h \times w}$ ($\alpha_{1,2,3,4} = 0°, 45°, 90°, 135°$ are the polarizer angles) as the four full-color polarized images, and $\mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{HR}} \in \mathbb{R}^{3 \times kh \times kw}$ (k is the SR scale) as the four HR polarized images respectively, the process of PID→PISR can be written as

$$\mathbf{I}_{\alpha_{1,2,3,4}} = \mathcal{D}(\mathbf{R}) \text{ first, then } \mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{HR}} = \uparrow (\mathbf{I}_{\alpha_{1,2,3,4}}), \quad (1)$$

where $\mathcal{D}$ and $\uparrow$ represent demosaicing and super-resolution (SR) respectively. However, PID→PISR would produce degenerated results, reducing the accuracy of the DoP and AoP, as shown in Fig. 1 (mid). This is because existing PISR methods [8, 37] usually assume that the inputs are free of demosaicing artifacts, while existing PID methods [20, 22, 23] cannot guarantee perfect outputs. Therefore, the essential question needs to be addressed is: *How to robustly obtain* $\mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{HR}}$ *from* $\mathbf{R}$ *in a direct manner?*

We observe that the spatial resolution often correlates negatively with the severity of demosaicing artifacts, suggesting that enhancing resolution can benefit PID, while suppressing demosaicing artifacts could, in turn, improve the performance of PISR. The observation indicates that PID and PISR may be complementary, *i.e.*, optimizing both of

them in a single framework can potentially enhance each other's performance. This motivates us to propose **PIDSR**, a joint framework that performs complementary **P**olarized **I**mage **D**emosaicing and **S**uper-**R**esolution. As shown in Fig. 1 (bottom), given a CPFA raw image $\mathbf{R}$, our PIDSR can not only output the demosaicing results $\mathbf{I}_{\alpha_{1,2,3,4}}$ with fewer artifacts, but also output the HR polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{HR}}$ with higher quality, which can be described as

$$\mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{HR}}, \mathbf{I}_{\alpha_{1,2,3,4}} = \mathcal{D}^{\uparrow}(\mathbf{R}), \quad (2)$$

where $\mathcal{D}^{\uparrow}$ denotes complementary demosaicing and SR. Here, it is non-trivial to carefully design the formulation of $\mathcal{D}^{\uparrow}$, since naively formulating $\mathcal{D}^{\uparrow}$ as a combination of $\mathcal{D}$ and $\uparrow$ would result in error accumulation. To reduce the level of error, we propose to formulate $\mathcal{D}^{\uparrow}$ as a series of polarized pixel reconstruction sub-problems, and introduce a two-stage pipeline to handle the intra-resolution and cross-resolution components of each sub-problem in a recurrent manner, fully exploiting the complementary aspects of $\mathcal{D}$ and $\uparrow$ to optimize each other jointly. Tailored to the pipeline, we design a neural network to explicitly inject the physical clues into both two stages to preserve the polarization properties, making full use of the Stokes-domain information of the polarized images. To summarize, this paper makes contributions by demonstrating: (1) **PIDSR, a complementary polarized image demosaicing and super-resolution framework**, including: (2) **a two-stage recurrent pipeline** to fundamentally reduce the level of error; and (3) **a Stokes-aided neural network** to preserve the polarization properties.

2. Related work

Polarized image demosaicing (PID). Unlike the demosaicing methods designed for conventional RGB images [7, 10, 24] that handle the mosaic generated from color filter array (CFA), the methods designed for PID aim to deal with the mosaic from color-polarization filter array (CPFA). Maeda developed Polanalyser [20], an open-source software that provides an interpolation-based basic PID tool, which is widely adopted in polarization-based vision tasks [1, 2]. For higher performance, some methods attempted to adopt numerical optimization based on handcrafted priors to suppress the demosaicing artifacts [5, 13, 16, 21, 22, 25, 34–36]. Some works adopted learning-based approaches to solve this challenging problem, including convolutional neural network (CNN) [14, 15, 23, 28, 31, 39], generative adversarial network (GAN) [4], dictionary learning [17, 18, 32, 40], *etc*. Li *et al*. [12] proposed a no-reference physics-based quality assessment metric and show that it can be used to address the PID problem. *However, these methods can only restore full-color polarized images from a CPFA raw image, and cannot further enhance the resolution.*

Polarized image super-resolution (PISR). Unlike the SR methods designed for conventional RGB images [3, 27, 29,

30] that focus solely on resolution enhancement, the methods designed for PISR aim to not only improve the resolution of multiple polarized images but also preserve polarization properties in them, making the task much more challenging. Hu *et al.* [8] proposed two polarized image degradation models to simulate real image degradation, and designed a network named PSRNet to perform polarization-aware SR on monochrome polarized images along with a loss function to refine the DoP and AoP in a direct manner. Yu *et al.* [37] proposed a network named CPSRNet to perform polarization-aware SR on color polarized images, which incorporated a cross-branch activation module (CBAM) [33] to leverage high-frequency information contained in the DoP and AoP for preserving the polarization properties explicitly. *However, they are largely based on an assumption that the demosaicing artifacts are not that significant, and ignore the errors in the DoP and AoP of the input polarized images.*

3. Method

3.1. Background

CPFA raw image formation model. Since polarization cameras have a linear camera response function (*i.e.*, the pixel values linearly relate to the input irradiance), here we follow other works [19, 41] by not applying any special adjustments for non-linearity. As shown in Fig. 1 (top), a full-color polarized image $\mathbf{I}_{\alpha_i} \in \mathbb{R}^{3 \times h \times w}$ can be regarded as a collection of single-channel polarized images, *i.e.*, $\mathbf{I}_{\alpha_i} = \{\mathbf{I}_{\alpha_i}^{c_j}\}$, where $i = 1, 2, 3, 4$ and $\alpha_{1,2,3,4} = 0°, 45°, 90°, 135°$ denote the polarizer angles, $j = r, g, b$ and $c_{r,g,b}$ denote RGB color channels. Here, each single-channel polarized image $\mathbf{I}_{\alpha_i}^{c_j} \in \mathbb{R}^{1 \times h \times w}$ can be written as

$$\mathbf{I}_{\alpha_i}^{c_j} = \mathcal{C}_j(\mathcal{P}_i(\mathbf{E})), \qquad (3)$$

where $\mathbf{E} \in \mathbb{R}^{1 \times h \times w}$ denotes the input irradiance sampled by an $h \times w$ pixel array, $\mathcal{C}_j$ and $\mathcal{P}_i$ denote the color and polarization filtering operations at c_j and α_i performed by the CPFA respectively. A CPFA raw image $\mathbf{R} \in \mathbb{R}^{3 \times h \times w}$ captured by a polarization camera can be regarded as the weighted sum of each single-channel polarized image $\mathbf{I}_{\alpha_i}^{c_j}$:

$$\mathbf{R} = \sum_{\substack{i \in \{1,2,3,4\} \\ j \in \{r,g,b\}}} \mathbf{M}_{ij} \cdot \mathbf{I}_{\alpha_i}^{c_j} = \sum_{\substack{i \in \{1,2,3,4\} \\ j \in \{r,g,b\}}} \mathbf{M}_{ij} \cdot (\mathcal{C}_j(\mathcal{P}_i(\mathbf{E}))), \qquad (4)$$

where $\mathbf{M}_{ij} \in \mathbb{R}^{1 \times h \times w}$ denotes the weight of each summation term whose pixel value at coordinates (x, y) satisfies

$$\mathbf{M}_{ij}(x, y) = \begin{cases} 1 & \text{if } \mathbf{I}_{\alpha_i}^{c_j}(x, y) \text{ is in the CPFA pattern} \\ 0 & \text{otherwise} \end{cases}. \qquad (5)$$

Combing Eq. (3) and Eq. (4), we can see that PID is similar to performing interpolation on the missing pixels (*i.e.*, the

pixels at coordinates (x, y) satisfying $\mathbf{M}_{ij}(x, y) = 0$) from one out of twelve necessary intensity measurements, and it is an ill-posed problem without closed form solution.

Acquisition of the DoP and AoP. Given a CPFA raw image $\mathbf{R}$, one can perform PID on it to obtain four full-color polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ and use them to acquire the DoP $\mathbf{p} \in [0, 1]$ and AoP $\boldsymbol{\theta} \in [0, \pi]$ for downstream tasks by

$$\mathbf{p} = \frac{\sqrt{\mathbf{S}_1^2 + \mathbf{S}_2^2}}{\mathbf{S}_0} \text{ and } \boldsymbol{\theta} = \frac{1}{2} \arctan(\frac{\mathbf{S}_2}{\mathbf{S}_1}), \qquad (6)$$

where $\mathbf{S}_{0,1,2}$[1] are called the Stokes parameters [6, 11] that can be computed as

$$\begin{cases} \mathbf{S}_0 = 2\bar{\mathbf{I}}_{\alpha_i} = \mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_3} = \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_4} \\ \mathbf{S}_1 = \mathbf{I}_{\alpha_3} - \mathbf{I}_{\alpha_1}, \text{ and } \mathbf{S}_2 = \mathbf{I}_{\alpha_4} - \mathbf{I}_{\alpha_2} \end{cases}, \qquad (7)$$

where $\bar{\mathbf{I}}_{\alpha_i} = \sum_{i=1}^{4} \mathbf{I}_{\alpha_i}/4$ is the average polarized image.

Acquisition of the HR counterparts. As the polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$ become available, one can perform PISR on them to acquire their HR counterparts $\mathbf{I}_{\alpha_{1,2,3,4}}^{\text{HR}}$. Similarly, the HR counterparts of the Stokes parameters $\mathbf{S}_{0,1,2}^{\text{HR}}$, DoP $\mathbf{p}^{\text{HR}}$ and AoP $\boldsymbol{\theta}^{\text{HR}}$ can also be acquired by substituting $\mathbf{I}_{\alpha_{1,2,3,4}}$ with $\mathbf{I}_{\alpha_{1,2,3,4}}^{\text{HR}}$ in Eq. (7) and Eq. (6). It is important to note existing PISR methods [8, 37] cannot directly perform super-resolution on CPFA raw images, and they require PID as a pre-processing step to generate the polarized images first.

3.2. Motivation and overall framework

As indicated in Eq. (7) and Eq. (6), $\mathbf{p}$ and $\boldsymbol{\theta}$ exhibit non-linear relationships with $\mathbf{I}_{\alpha_{1,2,3,4}}$. This non-linearity would exacerbate demosaicing artifacts, meaning errors arising from imperfections in PID (*e.g.*, inaccurate interpolation, failure to handle sensor noise, *etc.*) are more noticeable in $\mathbf{p}$ and $\boldsymbol{\theta}$ than in $\mathbf{I}_{\alpha_{1,2,3,4}}$. To verify it, we design an experiment on our test dataset to evaluate the average error rates of $\mathbf{p}$, $\boldsymbol{\theta}$, and $\mathbf{S}_0$[2] acquired from the demosaicing results. Here, we choose Polanalyser [20], IGRI2 [22], and TCPDNet [23] as the PID methods, and define the error rate of a variable $\mathbf{v}$ (normalized to $[0, 1]$) similar to the one in [43]: $\text{ER}_{\mathbf{v}} = \frac{\sum_{\text{p/w}} |\mathbf{v} - \mathbf{v}_{\text{gt}}|}{\sum_{\text{p/w}} \mathbf{v}}$, where $\sum_{\text{p/w}}$ denotes the pixel-wise sum, the subscript gt stands for the ground truth throughout this paper. As shown in Fig. 2 (a), the average error rates of $\mathbf{p}$ and $\boldsymbol{\theta}$ are much larger than $\mathbf{S}_0$ for all PID methods. Besides, obtaining high-quality HR polarized images is challenging because the performance of PISR is constrained by the effectiveness of the pre-processing step, PID. To verify it, we design another experiment on our test dataset to evaluate the performance of

---

[1]$\mathbf{S}_0$ describes the total intensity (which can be regarded as the unpolarized image), and $\mathbf{S}_1$ ($\mathbf{S}_2$) describes the difference between the intensity of the vertical and horizontal (135° and 45°) polarized light.

[2]Since $\mathbf{S}_0$ has a linear relationship with $\mathbf{I}_{\alpha_{1,2,3,4}}$ (see Eq. (7)), we can use $\mathbf{S}_0$ to represent $\mathbf{I}_{\alpha_{1,2,3,4}}$ in such a proof-of-concept experiment.
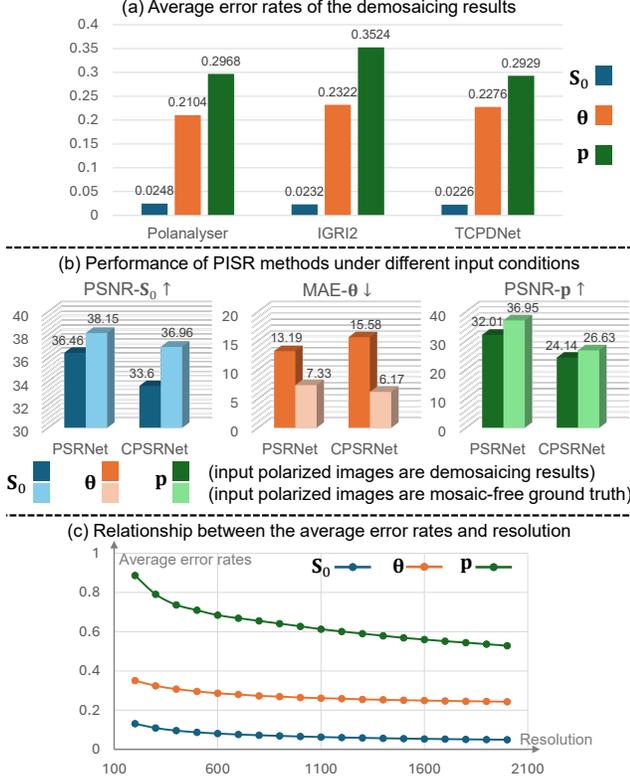
Figure 2. (a) The average error rates of $\mathbf{p}$ and $\boldsymbol{\theta}$ are much larger than the one of $\mathbf{S}_0$ for PID methods (Polanalyser [20], IGRI2 [22], and TCPDNet [23]). (b) The performance of PISR methods (PSRNet [8] and CPSRNet [37]) on both $\mathbf{p}$, $\boldsymbol{\theta}$, and $\mathbf{S}_0$ is much better when using mosaic-free ground truth polarized images compared with using demosaicing results. (c) The average error rates of both $\mathbf{p}$, $\boldsymbol{\theta}$, and $\mathbf{S}_0$ decrease as the resolution increases.

two PISR methods (PSRNet [8] and CPSRNet [37]) under different input conditions. Specifically, we use polarized images generated by an existing PID method (Polanalyser [20]) and their corresponding mosaic-free ground truth as inputs respectively for comparison. As shown in Fig. 2 (b), the performance of PISR methods using polarized images generated by Polanalyser [20] as input is inferior to that using the ground truth as input for both $\mathbf{p}$, $\boldsymbol{\theta}$, and $\mathbf{S}_0$.

Notably, we observe that the spatial resolution at which the input irradiance $\mathbf{E}$ is sampled (*i.e.*, the same scene sampled at different $h \times w$ resolutions) often correlates negatively with the severity of demosaicing artifacts. As a proof of concept, we use a virtual camera with varying resolutions to sample the input irradiance from rendered scenes (using Mitsuba 3[3]), and adopt Eq. (4) to obtain the CPFA raw images at different resolutions; then, a PID method (Polanalyser [20]) is adopted to produce the corresponding demosaicing results. Results are shown in Fig. 2 (c), which demonstrate the

---

[3]https://www.mitsuba-renderer.org/
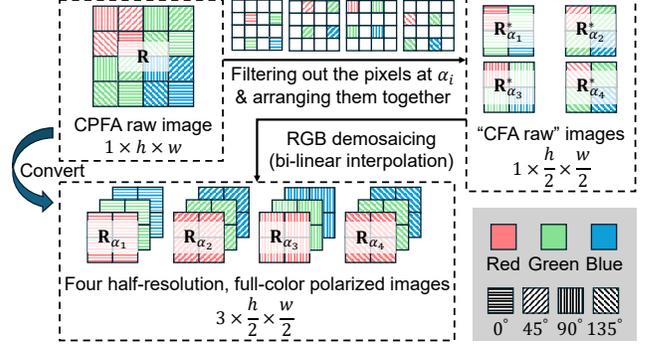
Figure 3. A CPFA raw image $\mathbf{R}$ can be approximately converted to four half-resolution, full-color polarized image $\mathbf{R}_{\alpha_{1,2,3,4}}$.

severity of demosaicing artifacts (measured using average error rates of both $\mathbf{p}$, $\boldsymbol{\theta}$, and $\mathbf{S}_0$) decreases as the resolution increases. We can see these results align with the fact that existing PID methods [20, 22, 23] typically leverage inter-actions among neighboring pixels for interpolation, where preserving finer details could lead to more accurate interpolated pixel values (*i.e.*, at higher resolutions, artifacts such as blurring or jagged edges (aliasing) are less likely to occur). This observation suggests enhancing resolution can benefit PID. Combining the fact that suppressing demosaicing artifacts could improve the performance of PISR, we can deduce that PID and PISR may be complementary.

Based on the above analysis, we propose to design a joint framework that performs complementary polarized image demosaicing and super-resolution, named PIDSR. As shown in Eq. (2), given a CPFA raw image $\mathbf{R}$ as input, our PIDSR aims to perform complementary demosaicing and SR ($\mathcal{D}^{\uparrow}$) on it to output not only demosaicing results $\mathbf{I}_{\alpha_{1,2,3,4}}$ but also HR polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}^{\text{HR}}$ with more accurate DoP and AoP. Thus, the overall process of PIDSR can be regarded as maximizing a posteriori of the outputs $\mathbf{I}_{\alpha_{1,2,3,4}}$ and $\mathbf{I}_{\alpha_{1,2,3,4}}^{\text{HR}}$ conditioned on the inputs $\mathbf{R}$ along with the complementary demosaicing and SR function $\mathcal{D}^{\uparrow}$ parameterized by Ψ:

$$\underset{\Psi}{\arg\max} \; \mathcal{D}^{\uparrow}(\mathbf{I}_{\alpha_{1,2,3,4}}, \mathbf{I}_{\alpha_{1,2,3,4}}^{\text{HR}} | \mathbf{R}, \Psi). \quad (8)$$

3.3. Two-stage recurrent PIDSR pipeline

The most straightforward way to solve the maximum a posteriori estimation problem in Eq. (8) is directly formulating the complementary demosaicing and SR function $\mathcal{D}^{\uparrow}$ as a cascade of two stages, *i.e.*, demosaicing function $\mathcal{D}$ and SR function $\uparrow$, as shown in Eq. (1). However, such a pipeline has two main drawbacks that limit its overall performance. First, errors tend to accumulate across stages because $\mathcal{D}$ and $\uparrow$ are independent, preventing the formation of negative feedback loops to stabilize the level of error. Second, the sequential nature of these stages fails to leverage the comple-
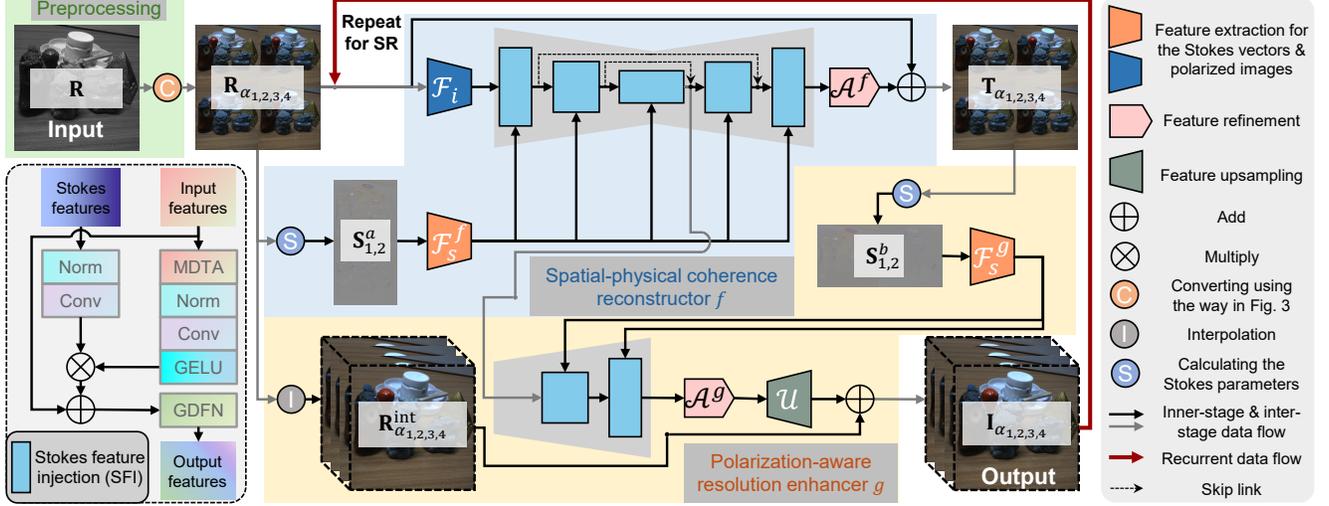
Figure 4. The workflow and network design of our PIDSR framework, consisting of two stages: spatial-physical coherence reconstructor $f(\cdot)$ and polarization-aware resolution enhancer $g(\cdot)$. Here we only illustrate the demosaicing workflow, and the SR one is in a repeated manner.

mentary aspects of $\mathcal{D}$ and $\uparrow$, missing the opportunity for joint optimization. Therefore, a more robust pipeline is required.

Prominently, we found that a CPFA raw image $\mathbf{R}$ can be approximately converted to four half-resolution, full-color polarized images $\mathbf{R}_{\alpha_{1,2,3,4}}$. As shown in Fig. 3, first, by filtering out the pixels corresponding to a specific polarizer angle α_i from a given CPFA raw image $\mathbf{R}$ and arranging them together, we could form an image $\mathbf{R}^*_{\alpha_i} \in \mathbb{R}^{1 \times h/2 \times w/2}$ with a format similar to a CFA raw image (*i.e.*, the mosaic pattern produced by a color filter array); then, by applying a simple RGB demosaicing method (*e.g.*, bi-linear interpolation) to $\mathbf{R}^*_{\alpha_i}$, we could generate a half-resolution, full-color polarized image $\mathbf{R}_{\alpha_i}$, though it would exhibit spatial discontinuities between neighboring pixels. This suggests that $\mathcal{D}$ can be formulated into two sub-problems: spatial discontinuity alleviation and resolution enhancement, equivalent to performing intra-resolution and cross-resolution polarized pixel reconstruction. Similarly, $\uparrow$ can also be formulated into two sub-problems: physical correlation restoration (the intra-resolution one) and resolution enhancement (the cross-resolution one). This decoupled formulation ensures that the disrupted physical correlation among multiple polarized images, caused by demosaicing artifacts from the pre-processing step (PID), has minimal negative impact on resolution enhancement, facilitating the accurate acquisition of DoP and AoP. To this end, we could unify $\mathcal{D}$ and $\uparrow$ into a recurrent structure, which can not only avoid error accumulation but also make full use of the complementary aspects of $\mathcal{D}$ and $\uparrow$ to optimize each other jointly.

We design a two-stage recurrent PIDSR pipeline to implement $\mathcal{D}^{\uparrow}$, as illustrated in Fig. 4. The first stage f is a *spatial-physical coherence reconstructor* that performs intra-resolution pixel reconstruction, aiming to alleviate the spa-

tial discontinuities between neighboring pixels, restore the physical correlation among multiple polarized images, and deal with the potential sensor noise; the second stage g is a *polarization-aware resolution enhancer* that performs cross-resolution pixel reconstruction, with a focus on both SR and preserving polarization properties. Starting with a CPFA raw image $\mathbf{R}$ as the initial input, we first approximately convert it into four half-resolution, full-color polarized images $\mathbf{R}_{\alpha_{1,2,3,4}}$ using the way shown in Fig. 3 as a pre-processing step, then send $\mathbf{R}_{\alpha_{1,2,3,4}}$ to f and g in a sequential manner to finish the first round of iteration to obtain four full-color polarized images $\mathbf{I}_{\alpha_{1,2,3,4}}$; after that, we can repeat the iteration for n additional rounds to produce four HR polarized images $\mathbf{I}^{\mathrm{HR}}_{\alpha_{1,2,3,4}}$ with an SR factor of $k = 2^n \times$.

3.4. Stokes-aided PIDSR network

Spatial-physical coherence reconstructor (f). As shown in the first stage of Fig. 4, it aims to solve the intra-resolution polarized pixel reconstruction sub-problem, which alleviates the inherent spatial discontinuity in $\mathbf{R}_{\alpha_{1,2,3,4}}$ and restores the imperfect physical correlation in $\mathbf{I}_{\alpha_{1,2,3,4}}$ during the demosaicing and SR workflows, respectively. Taking the demosaicing workflow as an example, this stage learns the residual between $\mathbf{R}_{\alpha_{1,2,3,4}}$ and $\mathbf{T}_{\alpha_{1,2,3,4}}$ (which are the spatially continuous intermediate results). First, two feature extraction heads $\mathcal{F}_i$ and $\mathcal{F}^f_s$ are used to extract the image and polarization features from $\mathbf{R}_{\alpha_{1,2,3,4}}$ and their corresponding Stokes parameters $\mathbf{S}^a_{1,2}$ respectively. Then, a backbone network is adopted to process the extracted features to compensate the missing spatial information in the high-dimensional feature space. Here, we should not directly concatenate the extracted features and send them into the backbone network, since the domain gap between the features of $\mathbf{R}_{\alpha_{1,2,3,4}}$ and $\mathbf{S}^a_{1,2}$ could

Table 1. Quantitative comparisons on synthetic data. The comparisons involve our PIDSR, three state-of-the-art PID methods (Polanalyser [20], IGRI2 [22], and TCPDNet [23]), and the only existing two PISR methods (PSRNet [8] and CPSRNet [37]).

Metric	PSNR↑/SSIM↑						MAE↓
Demosaicing	$\mathbf{I}_{\alpha_1}(0°)$	$\mathbf{I}_{\alpha_2}(45°)$	$\mathbf{I}_{\alpha_3}(90°)$	$\mathbf{I}_{\alpha_4}(135°)$	$\mathbf{S}_0$	$\mathbf{p}$	$\boldsymbol{\theta}$
Polanalyser [20]	31.95/0.8955	32.08/0.8968	32.44/0.8989	32.14/0.8973	33.28/0.9138	26.68/0.7164	17.8666
IGRI2 [22]	34.25/0.9349	34.33/0.9359	34.64/0.9369	34.46/0.9365	35.50/0.9450	27.78/0.7486	16.5830
TCPDNet [23]	37.26/0.9585	37.60/0.9602	37.90/0.9609	37.81/0.9612	38.65/0.9669	32.26/0.8262	13.1881
PIDSR	**38.90/0.9717**	**38.98/0.9719**	**39.11/0.9721**	**39.21/0.9732**	**40.24/0.9780**	**33.33/0.8447**	**12.2383**
Super-resolution	$\mathbf{I}_{\alpha_1}^{\mathrm{HR}}(0°)$	$\mathbf{I}_{\alpha_2}^{\mathrm{HR}}(45°)$	$\mathbf{I}_{\alpha_3}^{\mathrm{HR}}(90°)$	$\mathbf{I}_{\alpha_4}^{\mathrm{HR}}(135°)$	$\mathbf{S}_0^{\mathrm{HR}}$	$\mathbf{p}^{\mathrm{HR}}$	$\boldsymbol{\theta}^{\mathrm{HR}}$
PSRNet ($2\times$) [8]	35.66/0.9309	35.49/0.9301	35.65/0.9306	35.65/0.9319	36.46/0.9439	32.01/0.8298	13.1884
CPSRNet ($2\times$) [37]	32.97/0.8936	33.11/0.8944	33.35/0.8947	33.17/0.8958	33.60/0.9021	24.14/0.7649	15.5811
PIDSR ($2\times$)	**36.55/0.9488**	**36.64/0.9493**	**36.85/0.9502**	**36.77/0.9505**	**37.44/0.9553**	**32.97/0.8438**	**12.3520**
PSRNet ($4\times$) [8]	35.15/0.9227	35.41/0.9247	35.74/0.9264	35.57/0.9257	36.13/0.9311	31.95/0.8305	13.7751
CPSRNet ($4\times$) [37]	30.82/0.8599	30.75/0.8596	30.98/0.8600	30.76/0.8608	31.16/0.8677	22.52/0.7325	16.5469
PIDSR ($4\times$)	**35.48/0.9297**	**35.58/0.9307**	**35.83/0.9321**	**35.70/0.9319**	**36.31/0.9371**	**32.43/0.8379**	**13.0520**

be very large, *i.e.*, the features of $\mathbf{R}_{\alpha_{1,2,3,4}}$ contain mainly low-frequency structures, while the features of $\mathbf{S}_{1,2}^a$ contain mainly high-frequency structures. To handle this issue, we design the backbone network as a modified U-Net [26] architecture, where in each scale the original convolution block is substituted with a Stokes feature injection (SFI) block to explicitly utilize the physical clues encoded in the Stokes parameters to provide guidance for bridging the domain gap. The SFI block contains two different branches for processing the input and Stokes features respectively, which learn a bias by multiplying the processed features to adjust the input features. To effectively capture long-range feature interactions, we design the SFI block to incorporate a multi-Dconv head transposed attention (MDTA) module [38] at the beginning of the branch for input features along with a gated-Dconv feed-forward network (GDFN) module [38] before output. After the backbone network, a feature refinement block $\mathcal{A}^f$ (containing an MDTA and a GDFN module [38]) is used to reconstruct the residual between $\mathbf{R}_{\alpha_{1,2,3,4}}$ and $\mathbf{T}_{\alpha_{1,2,3,4}}$.

Polarization-aware resolution enhancer (g). As shown in the second stage of Fig. 4, it aims to solve the cross-resolution polarized pixel reconstruction sub-problem, which focuses on resolution enhancement during both the demosaicing and SR workflows. Also taking the demosaicing workflow as an example, this stage learns the residual between $\mathbf{R}_{\alpha_{1,2,3,4}}^{\mathrm{int}}$ (the interpolated version of $\mathbf{R}_{\alpha_{1,2,3,4}}$) and $\mathbf{I}_{\alpha_{1,2,3,4}}$. Since $\mathbf{T}_{\alpha_{1,2,3,4}}$ are spatially continuous, their corresponding Stokes parameters $\mathbf{S}_{1,2}^b$ could offer robust physical clues to facilitate the SR process with polarization-awareness. Besides, since the backbone network in f already encodes fine-grained multiscale features in the image domain, we do not need to extract features from $\mathbf{T}_{\alpha_{1,2,3,4}}$ additionally. Therefore, in this stage, we choose to directly grab the features from the coarsest level of the backbone network in f

and send them into a decoder (sharing the same architecture with the decoder part of the backbone network in f), under the guidance of the features of $\mathbf{S}_{1,2}^b$ extracted by another feature extraction head $\mathcal{F}_s^g$. Then, the output features of the decoder are fed into another feature refinement block $\mathcal{A}^g$ and a feature upsampling block $\mathcal{U}$ in a sequential manner to form the residual between $\mathbf{R}_{\alpha_{1,2,3,4}}^{\mathrm{int}}$ and $\mathbf{I}_{\alpha_{1,2,3,4}}$.

Loss function. We design the loss function for both demosaicing and SR rounds as $L = \lambda_1 L_{\mathrm{img}} + \lambda_2 L_{\mathrm{Stokes}} + \lambda_3 L_{\mathrm{pol}}$, where L_{img} is the image loss aiming to ensure the pixel accuracy in the image domain, L_{Stokes} is the Stokes loss aiming to preserve the continuity in the Stokes domain, and L_{pol} is the polarization loss aiming to enforce the physical correctness of the DoP and AoP, $\lambda_{1,2,3}$ are set to be 1.0, 10.0, and 10.0 respectively. Here, we only detail each loss term in the demosaicing round, and the SR round could be similar (just replace the variables with the corresponding HR counterparts). For the second stage (g), L_{img} can be written as $L_{\mathrm{img}} = L_1(\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_3}, \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_4}) + L_{\mathrm{grad}}(\mathbf{I}_{\alpha_{1,2,3,4}}, \mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{gt}})$, where L_1 and L_{grad} denote the ℓ_1 loss and gradient loss respectively, the superscript gt labels the ground truth throughout this paper. Here, $L_1(\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_3}, \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_4})$ aims to adjust the numerical relationship among $\mathbf{I}_{\alpha_{1,2,3,4}}$ since $\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_3} = \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_4}$ always holds for polarized images. L_{Stokes} can be written as $L_{\mathrm{Stokes}} = L_{\mathrm{grad}}(\mathbf{S}_0, \mathbf{S}_0^{\mathrm{gt}}) + L_1(\mathbf{S}_{1,2}, \mathbf{S}_{1,2}^{\mathrm{gt}})$. L_{pol} can be written as $L_{\mathrm{pol}} = L_1(\mathbf{p}, \mathbf{p}^{\mathrm{gt}}) + L_1(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{gt}})$. For the first stage ($f$), we use $\mathbf{T}_{\alpha_{1,2,3,4}}^{\mathrm{gt}}$ (the half-resolution version of $\mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{gt}}$) along with the corresponding Stokes parameters, DoP, and AoP for supervision.

Training strategy. Our PIDSR is implemented using PyTorch and trained on an NVIDIA A800 GPU. For both demosaicing and SR, we train the two stages f and g for 100 epochs simultaneously in total, with a learning rate of 0.005. We use Adam optimizer [9] for optimization.
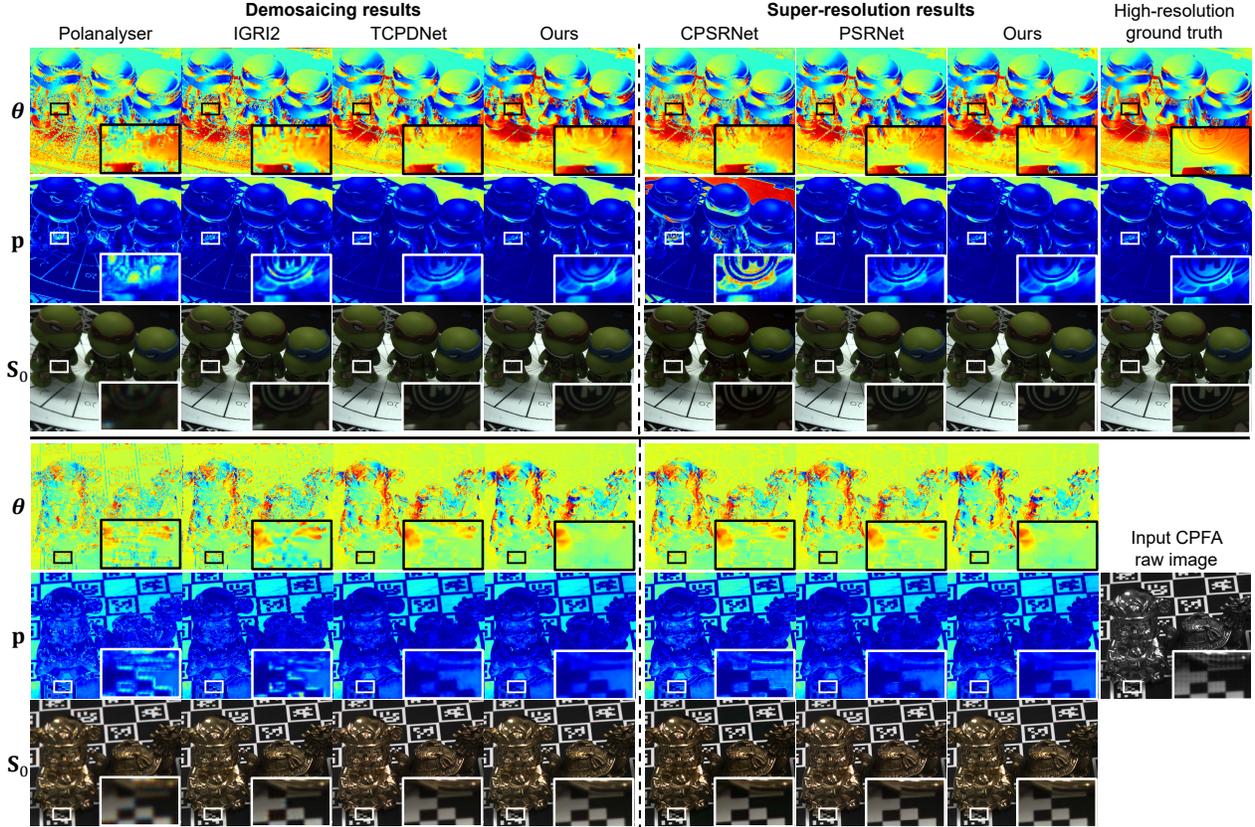
Figure 5. Qualitative comparisons on both synthetic (the top group) and real data (the bottom group) of both demosaicing and $4\times$ SR tasks.

4. Experiment

4.1. Evaluation

Since existing public datasets are insufficient for the setting of our PIDSR, we generate a synthetic dataset[4] for evaluation. As for the demosaicing performance evaluation, we compare our PIDSR with three state-of-the-art PID methods Polanalyser [20], IGRI2 [22], and TCPDNet [23]; as for the SR performance evaluation, we compare our PIDSR with the only existing two PISR methods PSRNet [8] and CPSR-Net [37]. Here, since PSRNet [8] is initially designed for grayscale polarized images, we make slight modifications on it to allow it to accept the color polarized images. Besides, since the compared PISR methods [8, 37] can only take the polarized images (instead of the CPFA raw image) as input, we provide them the demosaicing results from TCPDNet [23] (which achieves the best performance among the compared PID methods [20, 22, 23]). Note that all compared methods based on deep-learning [8, 23, 37] are retrained on our dataset for a fair comparison. As the compared methods do, we not only evaluate the quality of $\mathbf{I}_{\alpha_{1,2,3,4}}$ (for demosaicing task) and $\mathbf{I}_{\alpha_{1,2,3,4}}^{\mathrm{HR}}$ (for SR task), but also $\mathbf{p}$, $\boldsymbol{\theta}$, $\mathbf{S}_0$

(for demosaicing task) and $\mathbf{p}^{\mathrm{HR}}$, $\boldsymbol{\theta}^{\mathrm{HR}}$, $\mathbf{S}_0^{\mathrm{HR}}$ (for SR task).

We evaluate the results quantitatively on synthetic data using: Mean Angular Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). Here, MAE (lower values indicating better performance) is exclusively used to evaluate angular variables ($\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{\mathrm{HR}}$), while PSNR and SSIM are applied to the remaining variables. Results are shown in Tab. 1, where our framework consistently outperforms the compared methods on all metrics in both demosaicing and SR tasks. Visual quality comparisons on both synthetic and real data are shown in Fig. 5[5]. From the results, we can see that our PIDSR can produce more accurate DoP and AoP, while the compared methods suffer from severe artifacts (*e.g.*, broken edges and discontinuity).

4.2. Ablation study

We conduct several ablation studies in Tab. 2 to verify the validity of each design choice. First, we show the significance of our PIDSR framework design that formulates $\mathcal{D}^\uparrow$ as complementary demosaicing and SR, by comparing with an alternative design that naively formulates $\mathcal{D}^\uparrow$ as a combination of $\mathcal{D}$ and $\uparrow$ using the same network architecture

---

[4]Details about our dataset can be found in the supplementary material.

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

Table 2. Quantitative evaluation results of ablation study.

Metric	PSNR↑/SSIM↑		MAE↓
Demosaicing	S_0	p	θ
Sequential $\mathcal{D}$ and $\uparrow$	32.32/0.9134	23.78/0.6661	19.2907
Single-stage pipeline	34.61/0.9426	27.95/0.7406	38.2174
Without SFI blocks	37.18/0.9612	32.73/0.8392	13.1242
Ours (demosaicing only)→PSRNet [8]	**40.24/0.9780**	**33.33/0.8447**	**12.2383**
TCPDNet [23]→ ours (SR only)	38.65/0.9669	32.26/0.8262	13.1881
Our complete PIDSR	**40.24/0.9780**	**33.33/0.8447**	**12.2383**
Super Resolution	S_0^{HR} (2×)	p^{HR} (2×)	θ^{HR} (2×)
Sequential $\mathcal{D}$ and $\uparrow$	32.16/0.8965	20.07/0.6225	21.8965
Single-stage pipeline	34.35/0.9278	28.10/0.7584	38.2203
Without SFI blocks	36.35/0.9458	31.65/0.8322	14.1309
Ours (demosaicing only)→PSRNet [8]	36.81/0.9513	32.68/0.8412	12.5958
TCPDNet [23]→ ours (SR only)	36.83/0.9457	32.19/0.8308	13.1530
Our complete PIDSR	**37.44/0.9553**	**32.97/0.8438**	**12.3520**

(Sequential $\mathcal{D}$ and $\uparrow$). The performance degenerates, since such a naive pipeline would result in error accumulation. Next, we verify the necessity of our two-stage pipeline, by comparing to a single-stage pipeline that does not explicitly reconstruct spatial-physical coherence under the same PIDSR framework (Single-stage pipeline). The results are not that good since the still remaining spatial discontinuity and disrupted physical correlation would have negative impact on resolution enhancement. Then, we validate the effectiveness of our Stokes-aided neural network, by substituting the SFI blocks with original convolution blocks (Without SFI blocks). We find that it does not perform well since it cannot make full use of the Stokes-domain information to preserve the polarization properties. Finally, we also compare with two different hybrid baselines that feed our demosaicing results into PSRNet [8] for SR (Ours (demosaicing only)→PSRNet [8]), and feed the demosaicing results of TCPDNet [23] into our PIDSR for SR (TCPDNet [23]→ ours (SR only)), respectively. We can see that our complete PIDSR achieves the first performance.

4.3. Application

To show that our PIDSR can be beneficial to downstream polarization-based vision applications, we take polarization-based reflection removal (PRR, which takes reflection-contaminated polarized images as input and outputs reflection-removed unpolarized images) as an example, and try to obtain a reflection-removed unpolarized image from a reflection-contaminated CPFA raw image captured by a polarization camera. To achieve it, the following approaches could be used: (1) "PID→PISR→PRR": performing PID and PISR sequentially on the CPFA raw image, then performing reflection removal; (2) "PID→PRR→SISR": performing PID on the CPFA raw image first, then performing reflection removal, and performing single image super-resolution (SISR) in the end; (3) "PIDSR→PRR": performing our PIDSR on the CPFA raw image first, then performing reflection removal. Here, the SR scale is 2, and the

Input reflection-contaminated CPFA raw image

Reflection-free CPFA raw image (for reference)

(1) PID→PISR→PRR (2) PID→PRR→SISR (3) PIDSR→PRR (ours)
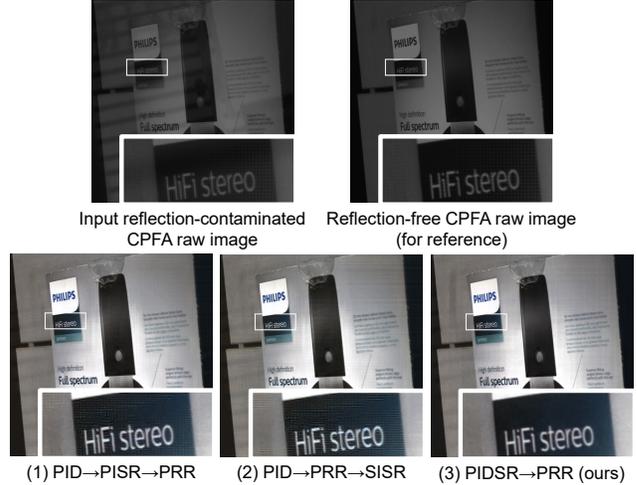
Figure 6. Results of polarization-based reflection removal. Our method is not influenced by the zigzag artifacts. Please zoom-in for better details.

involved PRR, PID, PISR, and SISR methods are selected to be RSP [19], TCPDNet [23], PSRNet [8], and OmniSR [29] respectively. Visual comparisons are shown in Fig. 6, where we can see the result from PIDSR→PRR (ours) contains more detailed textures and less reflection contamination.

5. Conclusion

We propose PIDSR, a joint framework that performs complementary polarized image demosaicing and super-resolution. By carefully designing a two-stage recurrent pipeline to fundamentally reduce the level of error and a Stokes-aided neural network to preserve the polarization properties, our PIDSR can robustly obtain HR polarized images with more accurate polarization-related parameters such as the DoP and AoP from a CPFA raw image in a direct manner.

Limitations. Since our PIDSR is specifically designed to process a single CPFA raw image, it is unsuitable for reconstructing a polarized video. Additionally, it cannot handle CFA raw images, as it requires the Stokes parameters as input, which are unavailable in this setting.

Acknowledgment

References

[1] 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. 1, 2

[2] 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. 1, 2

[3] Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang. Implicit diffusion models for continuous super-resolution. In *Proc. of Computer Vision and Pattern Recognition*, pages 10021–10030, 2023. 2

[4] 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*, 2024. 2

[5] Nathan Hagen, Thijs Stockmans, Yukitoshi Otani, and Prathan Buranasiri. Fourier-domain filtering analysis for color-polarization camera demosaicking. *Applied Optics*, 63(9):2314–2323, 2024. 2

[6] Eugene Hecht. *Optics*. Pearson Education India, 2012. 3

[7] Jingchao Hou, Garas Gendy, Guo Chen, Liangchao Wang, and Guanghui He. DTDeMo: A deep learning-based two-stage image demosaicing model with interpolation and enhancement. *IEEE Transactions on Computational Imaging*, 2024. 2

[8] 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. 1, 2, 3, 4, 6, 7, 8

[9] Diederik P Kingma and Jimmy Ba. ADAM: A method for stochastic optimization, 2014. 6

[10] Filippos Kokkinos and Stamatios Lefkimmiatis. Deep image demosaicking using a cascade of convolutional residual denoising networks. *Proc. of European Conference on Computer Vision*, 2018. 2

[11] GP Können. *Polarized light in nature*. CUP Archive, 1985. 3

[12] Ning Li, Benjamin Le Teurnier, Matthieu Boffety, François Goudail, Yongqiang Zhao, and Quan Pan. No-reference physics-based quality assessment of polarization images and its application to demosaicking. *IEEE Transactions on Image Processing*, 30:8983–8998, 2021. 2

[13] 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. 2

[14] Xiangbo Liu, Xiaobo Li, and Shih-Chi Chen. Enhanced polarization demosaicking network via a precise angle of polarization loss calculation method. *Optics Letters*, 47(5):1065–1068, 2022. 2

[15] 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. 2

[16] 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*, 2024. 2

[17] Yidong Luo, Junchao Zhang, and Di Tian. Sparse representation-based demosaicking method for joint chromatic and polarimetric imagery. *Optics and Lasers in Engineering*, 164:107526, 2023. 2

[18] 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*, 2024. 2

[19] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. In *Proc. of Advances in Neural Information Processing Systems*, 2019. 1, 3, 8

[20] Ryota Maeda. Polanalyser: Polarization image analysis tool, 2019. 2, 3, 4, 6, 7

[21] Miki Morimatsu, Yusuke Monno, Masayuki Tanaka, and Masatoshi Okutomi. Monochrome and color polarization demosaicking using edge-aware residual interpolation. In *Proc. of International Conference on Image Processing*, pages 2571–2575, 2020. 2

[22] 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. 2, 3, 4, 6, 7

[23] 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. 1, 2, 3, 4, 6, 7, 8

[24] Guocheng Qian, Yuanhao Wang, Jinjin Gu, Chao Dong, Wolfgang Heidrich, Bernard Ghanem, and Jimmy S Ren. Rethink the pipeline of demosaicking, denoising, and super-resolution. *Proc. of International Conference on Computational Photography*, 2022. 2

[25] 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. 2

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proc. of International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 234–241, 2015. 6

[27] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022. 2

[28] Yuanyuan Sun, Junchao Zhang, and Rongguang Liang. Color polarization demosaicking by a convolutional neural network. *Optics Letters*, 46(17):4338–4341, 2021. 2

[29] Hang Wang, Xuanhong Chen, Bingbing Ni, Yutian Liu, and Jinfan Liu. Omni aggregation networks for lightweight image super-resolution. In *Proc. of Computer Vision and Pattern Recognition*, pages 22378–22387, 2023. 2, 8

[30] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. SinSR: diffusion-based image super-resolution in a single step. In *Proc. of Computer Vision and Pattern Recognition*, pages 25796–25805, 2024. 3

[31] 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. 2

[32] 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. 2

[33] 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. 3

[34] Rongyuan Wu, Yongqiang Zhao, Ning Li, and Seong G Kong. Polarization image demosaicking using polarization channel difference prior. *Optics Express*, 29(14):22066–22079, 2021. 2

[35] Jianqiao Xin, Zheng Li, Shiguang Wu, and Shiyong Wang. Demosaicking DoFP images using edge compensation method based on correlation. *Optics Express*, 31(9):13536–13551, 2023.

[36] Yanji Yi, Peng Zhang, Zhiyu Chen, Hui Zhang, Zhendong Luo, Guanglie Zhang, Wenjung Li, and Yang Zhao. A demosaicking method based on an inter-channel correlation model for DoFP polarimeter. *Optics and Lasers in Engineering*, 181: 108388, 2024. 2

[37] 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. 2, 3, 4, 6, 7

[38] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proc. of Computer Vision and Pattern Recognition*, pages 5728–5739, 2022. 6

[39] 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 S_0, DoLP, and AoP. *Optics Express*, 27(6):8566–8577, 2019. 2

[40] 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. 2

[41] 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*, 2021. 1, 3

[42] 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. 1

[43] 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. 3