

# Polarization-Aware Low-Light Image Enhancement

Chu Zhou<sup>1</sup>, Minggui Teng<sup>2</sup>, Youwei Lyu<sup>3</sup>, Si Li<sup>3</sup>, Chao Xu<sup>1</sup>, Boxin Shi<sup>2\*</sup>

<sup>1</sup>Key Laboratory of Machine Perception (MOE), School of Intelligence Science and Technology, Peking University

<sup>2</sup>National Engineering Research Center of Visual Technology, School of Computer Science, Peking University

<sup>3</sup>School of Artificial Intelligence, Beijing University of Posts and Telecommunications

## Abstract

Polarization-based vision algorithms have found uses in various applications since polarization provides additional physical constraints. However, in low-light conditions, their performance would be severely degenerated since the captured polarized images could be noisy, leading to noticeable degradation in the degree of polarization (DoP) and the angle of polarization (AoP). Existing low-light image enhancement methods cannot handle the polarized images well since they operate in the intensity domain, without effectively exploiting the information provided by polarization. In this paper, we propose a Stokes-domain enhancement pipeline along with a dual-branch neural network to handle the problem in a polarization-aware manner. Two application scenarios (reflection removal and shape from polarization) are presented to show how our enhancement can improve their results.

## Introduction

Exploring polarimetric properties of light transport has benefited various vision applications, such as reflection removal (Lei et al. 2020), shape from polarization (Deschaintre, Lin, and Ghosh 2021), image dehazing (Zhou et al. 2021), *etc.* Since these applications often need to take full advantage of the physical constraints provided by the unique cues of polarization, their accuracy is closely related to polarization-relevant parameters, such as the degree of polarization (DoP) and the angle of polarization (AoP) of the incoming light to the sensor. With the development of polarization cameras, capturing multiple polarized images of the same scene with different polarizer angles in a snapshot becomes possible, which brings convenience to the acquisition of the DoP and AoP. However, when taking photos in low-light conditions (*e.g.*, capturing with limited illumination or setting a short exposure time at a high frame rate), the signal-to-noise ratio (SNR) degenerates due to low photon counts. In such a situation, since the captured polarized images are noisy, leading to severely degenerated DoP and AoP, the performance of corresponding applications is negatively affected (Hu et al. 2020). Therefore, it is of great interest to enhance multiple polarized low-light images of the same scene for acquiring the DoP and AoP accurately.

Recent advances in single image low-light enhancement (Chen et al. 2018; Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020) have shown effectiveness in imaging in low-light conditions. They adopt neural networks to enhance the visual quality of images in the *intensity domain* by extracting image features and priors from a large amount of training data. However, when it comes to handling multiple polarized low-light images of the same scene, they can only process one of them at a time in a polarization-unaware manner (*i.e.*, without considering the physical constraints among the polarized images), so that the quality of the acquired DoP and AoP are rather unreliable. This situation could be relieved by processing multiple polarized low-light images simultaneously (*e.g.*, using IPLNet (Hu et al. 2020)), however, operating in the *intensity domain* cannot make effective use of the information provided by polarization.

Based on the fact that the polarization characteristics of the incoming light to the sensor are fully encoded in the Stokes parameters, we analyze the average error rates of polarization-relevant variables (including the polarized images, AoP, DoP, and Stokes parameters) w.r.t. the image irradiance reduction factor (caused by decreasing the exposure time or scene radiance) in the polarized low-light image formation model. We observe that the degradation in the Stokes parameters is less severe than the polarized images in low-light conditions, and therefore propose a new pipeline to solve the problem in the *Stokes domain* instead of the intensity domain. Along this pipeline, we design a dual-branch neural network based on the specific properties of the Stokes parameters to perform enhancement in a *polarization-aware* manner. We also present two application scenarios, including reflection removal (Lei et al. 2020) and shape from polarization (Deschaintre, Lin, and Ghosh 2021), to demonstrate the benefits of enhancing polarized low-light images. To summarize, this paper makes contributions by proposing: (1) a Stokes-domain enhancement pipeline for polarized low-light images; (2) a polarization-aware dual-branch network tailored to the pipeline; (3) two applications demonstrating the benefits of enhancing polarized low-light images.

## Related Work

Generally, low-light image enhancement methods could be divided into two categories: traditional methods and

\*Corresponding author: shiboxin@pku.edu.cn

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

learning-based methods. Traditional methods often utilize histogram equalization (Pizer et al. 1987) and Retinex theory (Land 1977) to turn the low-light image enhancement problem into a numerical optimization problem. However, the methods using histogram equalization may cause the problem of over- and under-enhancement since they often do not take the illumination into consideration, and the methods using Retinex theory often ignore noise removal and may amplify the noise. To increase the robustness, learning-based methods have been proposed. Existing learning-based methods are mainly based on supervised learning. They use a large amount of training data to learn the mapping from low-light images to normal-light images (Lore, Akintayo, and Sarkar 2017; Ren et al. 2019; Xu et al. 2020; Li et al. 2020; Lv, Liu, and Lu 2020; Lim and Kim 2020; Wang et al. 2020; Lu and Zhang 2020; Atoum et al. 2020; Ai and Kwon 2020; Li, Feng, and Hua 2021; Zheng, Shi, and Shi 2021; Lv, Li, and Lu 2021), estimate reflectance and illumination maps (Wei et al. 2018; Li et al. 2018; Wang et al. 2019a; Zhang, Zhang, and Guo 2019; Wang et al. 2019b; Fan et al. 2020; Yang et al. 2021b; Zhang et al. 2021b) based on the Retinex theory (Land 1977), or reconstruct enhanced images from raw low-light images directly (Chen et al. 2018; Maharjan et al. 2019; Zhu et al. 2020b; Wei et al. 2020; Lamba and Mitra 2021). Recently, unsupervised learning (Jiang et al. 2021), semi-supervised learning (Yang et al. 2020, 2021a), reinforcement learning (Yu et al. 2018), and zero-shot learning (Zhang et al. 2019; Guo et al. 2020; Zhu et al. 2020a; Liu et al. 2021; Li, Guo, and Chen 2021; Zhao et al. 2021) have also been introduced to solve this challenging problem. To deal with dynamic scenes, learning temporal stability has been considered in some low-light video enhancement methods (Lv et al. 2018; Chen et al. 2019; Jiang and Zheng 2019; Triantafyllidou et al. 2020; Zhang et al. 2021a; Wang et al. 2021). Although these methods have shown effectiveness in a large variety of scenes, they are not suitable for enhancing multiple polarized images since they only focus on enhancing the quality of a single input image and cannot consider the polarization relationship among multiple polarized images. To deal with this problem, Hu et al. (2020) proposed a network, named IPLNet, to enhance multiple polarized low-light images simultaneously. However, it still tried to handle the problem in the intensity domain, which cannot make effective use of the information provided by polarization.

## Method

### Polarized Low-Light Image Formation Model

In normal-light conditions, assuming the camera response function is linear (Lyu et al. 2019; Hu et al. 2020; Zhou et al. 2021), the formation of an image can be described as

$$\mathbf{I} = \mathbf{R} \cdot t, \quad (1)$$

where  $\mathbf{R}$  denotes the original scene radiance, and  $t$  is the sensor exposure time. Note that since in normal-light conditions the SNR of the captured image is sufficiently high, we ignore the noise term (Chen et al. 2018; Hu et al. 2020). When placing a polarizer with polarizer angle  $\alpha$  in front of the camera, according to the Malus' law (Hecht 2012), the

captured polarized image  $\mathbf{I}_\alpha$  can be calculated as

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

where  $\mathbf{p} \in [0, 1]$  and  $\theta \in [0, \pi]$  denote the DoP and AoP of the incoming light to the sensor respectively. Reformulating Eq. (2) into a polynomial form,  $\mathbf{I}_\alpha$  can be expressed as a linear combination of three parameters  $\mathbf{S}_{0,1,2}$ :

$$\mathbf{I}_\alpha = \frac{1}{2} \mathbf{S}_0 - \frac{1}{2} \cos(2\alpha) \cdot \mathbf{S}_1 - \frac{1}{2} \sin(2\alpha) \cdot \mathbf{S}_2, \quad \text{where} \\ \mathbf{S}_0 = \mathbf{I}, \quad \mathbf{S}_1 = \mathbf{I} \cdot \mathbf{p} \cdot \cos(2\theta), \quad \text{and} \quad \mathbf{S}_2 = \mathbf{I} \cdot \mathbf{p} \cdot \sin(2\theta) \quad (3)$$

are called the Stokes parameters (Können 1985) of the incoming light to the sensor. Once  $\mathbf{S}_{0,1,2}$  are available (from polarized images),  $\mathbf{p}$  and  $\theta$  could be easily acquired by

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

A polarization camera can capture four spatially-aligned and temporally-synchronized polarized images  $\mathbf{I}_{\alpha_{1,2,3,4}}$  with different polarizer angles  $\alpha_{1,2,3,4} = 0^\circ, 45^\circ, 90^\circ, 135^\circ$  in a snapshot<sup>1</sup>, which brings convenience to the acquisition of  $\mathbf{p}$  and  $\theta$  (Fig. 1 (a)). This is because according to the physical meanings of the Stokes parameters<sup>2</sup>,  $\mathbf{S}_{0,1,2}$  can be computed from  $\mathbf{I}_{\alpha_{1,2,3,4}}$  directly:

$$\mathbf{S}_0 = \frac{1}{2} (\mathbf{I}_{\alpha_1} + \mathbf{I}_{\alpha_2} + \mathbf{I}_{\alpha_3} + \mathbf{I}_{\alpha_4}), \quad (5)$$

$$\mathbf{S}_1 = \mathbf{I}_{\alpha_3} - \mathbf{I}_{\alpha_1}, \quad \text{and} \quad \mathbf{S}_2 = \mathbf{I}_{\alpha_4} - \mathbf{I}_{\alpha_2}.$$

However, when it comes to low-light conditions, the SNR of the captured polarized images degenerates due to low photon counts, so that the noise term cannot be ignored anymore. The captured polarized low-light images would be degenerated as (we use  $\hat{\cdot}$  to denote the degenerated variables):

$$\hat{\mathbf{I}}_{\alpha_i} = \frac{1}{\gamma} \mathbf{I}_{\alpha_i} + \mathbf{N}_i \quad (i = 1, 2, 3, 4), \quad (6)$$

where  $\gamma$  ( $\gamma > 1$ ) is a linear scaling factor denoting the image irradiance reduction caused by decreasing the exposure time or scene radiance, and  $\mathbf{N}_i = \mathcal{N}(\frac{1}{\gamma}, \mathbf{I}_{\alpha_i})$  stands for a noise term which is mainly affected by  $\frac{1}{\gamma}$  (Lv, Li, and Lu 2021). Therefore, the computed Stokes parameters and the acquired AoP and DoP would be degenerated correspondingly (Fig. 1 (b)).

**Error Rate Analysis.** To show that the polarized images, Stokes parameters, DoP, and AoP have different sensitivity to  $\gamma$ , we analyze the relationships between their average error rates and  $\gamma$ . Defining the average error rate of a variable  $\mathbf{x}$  ( $\mathbf{x}$  can be a polarized image or a Stokes parameter normalized to  $[0, 1]$ ) as

$$\mathcal{E}_{\mathbf{x}} = \frac{\sum |\gamma \hat{\mathbf{x}} - \mathbf{x}|}{\sum \mathbf{x}}, \quad (7)$$

<sup>1</sup>We do not consider the non-linearity in this paper since a polarization camera usually outputs images with a linear camera response function.

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

where  $\sum$  denotes the pixel-wise sum. We can see that  $\mathcal{E}_x$  can be regarded as the functions of  $\gamma$  given a specific scene (*i.e.*,  $\mathbf{I}_{\alpha_{1,2,3,4}}$ ). Combining Eq. (6) and Eq. (7), the relationship between  $\mathcal{E}_{\mathbf{I}_{\text{avg}}}$  ( $\mathbf{I}_{\text{avg}} = \sum_{i=1}^4 \mathbf{I}_{\alpha_i} / 4$ ) and  $\gamma$  can be written as

$$\begin{aligned} \mathcal{E}_{\mathbf{I}_{\text{avg}}} &= \frac{\gamma \sum |\mathbf{N}_{\text{avg}}|}{\sum \mathbf{I}_{\text{avg}}} = \gamma \cdot \mathcal{K}_0(\gamma), \quad \text{where} \\ \mathbf{N}_{\text{avg}} &= \frac{\sum_{i=1}^4 \mathbf{N}_i}{4} \quad \text{and} \quad \mathcal{K}_0(\gamma) = \frac{\sum |\mathbf{N}_{\text{avg}}|}{\sum \mathbf{I}_{\text{avg}}}. \end{aligned} \quad (8)$$

Similarly, combining Eq. (5) and Eq. (7), we obtain the relationship between  $\mathcal{E}_{\mathbf{S}_i}$  and  $\gamma$ :

$$\begin{aligned} \mathcal{E}_{\mathbf{S}_0} &= \frac{\gamma \sum |2\mathbf{N}_{\text{avg}}|}{\sum 2\mathbf{I}_{\text{avg}}} = \mathcal{E}_{\mathbf{I}_{\text{avg}}} = \gamma \cdot \mathcal{K}_0(\gamma), \\ \mathcal{E}_{\mathbf{S}_1} &= \gamma \cdot \mathcal{K}_1(\gamma), \quad \text{and} \quad \mathcal{E}_{\mathbf{S}_2} = \gamma \cdot \mathcal{K}_2(\gamma), \quad \text{where} \\ \mathcal{K}_1(\gamma) &= \frac{\sum |\mathbf{N}_3 - \mathbf{N}_1|}{\sum \mathbf{S}_1} \quad \text{and} \quad \mathcal{K}_2(\gamma) = \frac{\sum |\mathbf{N}_4 - \mathbf{N}_2|}{\sum \mathbf{S}_2}. \end{aligned} \quad (9)$$

Since  $\mathcal{K}_1(\gamma)$  and  $\mathcal{K}_2(\gamma)$  have similar formulations (the numerator is the difference between two noise terms following the same distribution, and the denominator is the difference (normalized to  $[0, 1]$ ) between the intensity of two polarized images), and the numerators of  $\mathcal{K}_1(\gamma)$  and  $\mathcal{K}_2(\gamma)$  are close to zero, we could approximately derive  $\mathcal{K}_1(\gamma) \approx \mathcal{K}_2(\gamma) < \mathcal{K}_0(\gamma)$ , which means

$$\mathcal{E}_{\mathbf{I}_{\text{avg}}} = \mathcal{E}_{\mathbf{S}_0} > \mathcal{E}_{\mathbf{S}_1} \approx \mathcal{E}_{\mathbf{S}_2}. \quad (10)$$

Similar to Eq. (7), we define the average error rates of the DoP and AoP (also normalized to  $[0, 1]$ ) as  $\mathcal{E}_{\mathbf{p}} = \frac{\sum |\hat{\mathbf{p}} - \mathbf{p}|}{\sum \mathbf{p}}$  and  $\mathcal{E}_{\theta} = \frac{\sum |\hat{\theta} - \theta|}{\sum \theta}$  respectively<sup>3</sup>. According to Eq. (4), the relationship between  $\mathcal{E}_{\mathbf{p}}$  (or  $\mathcal{E}_{\theta}$ ) and  $\gamma$  could be quite complicated. We therefore suspect the DoP and AoP are more sensitive to  $\gamma$  than the polarized images and the Stokes parameters. To verify it, we perform simulation on 6000 synthetic scenes to quantitatively obtain the relationships between the average error rates ( $\mathcal{E}_{\mathbf{I}_{\text{avg}}}$ ,  $\mathcal{E}_{\mathbf{S}_{0,1,2}}$ ,  $\mathcal{E}_{\mathbf{p}}$ ,  $\mathcal{E}_{\theta}$ ) and  $\gamma$ , as show in Fig. 1 (c)<sup>4</sup>. We can see that in low-light conditions the degradation in the DoP and AoP is quite noticeable, which can lead to degenerated performance of applications of polarization-based vision (Lei et al. 2020; Deschaintre, Lin, and Ghosh 2021).

### Stokes-Domain Enhancement Pipeline

We aim to enhance multiple polarized images captured by a polarization camera in low-light conditions to acquire the DoP and AoP with high accuracy. Directly denoising  $\hat{\mathbf{p}}$  and  $\hat{\theta}$  seems a straightforward way to achieve our goal. However, the methods designed for image denoising cannot handle this problem since the noise distributions of the DoP and

<sup>3</sup>Note that there is no need to multiply  $\gamma$  to  $\hat{\mathbf{p}}$  or  $\hat{\theta}$  like Eq. (7), since according to Eq. (4)  $\frac{1}{\gamma}$  will be canceled in the division operation.

<sup>4</sup>Details of this simulation experiment can be found in the supplementary material.

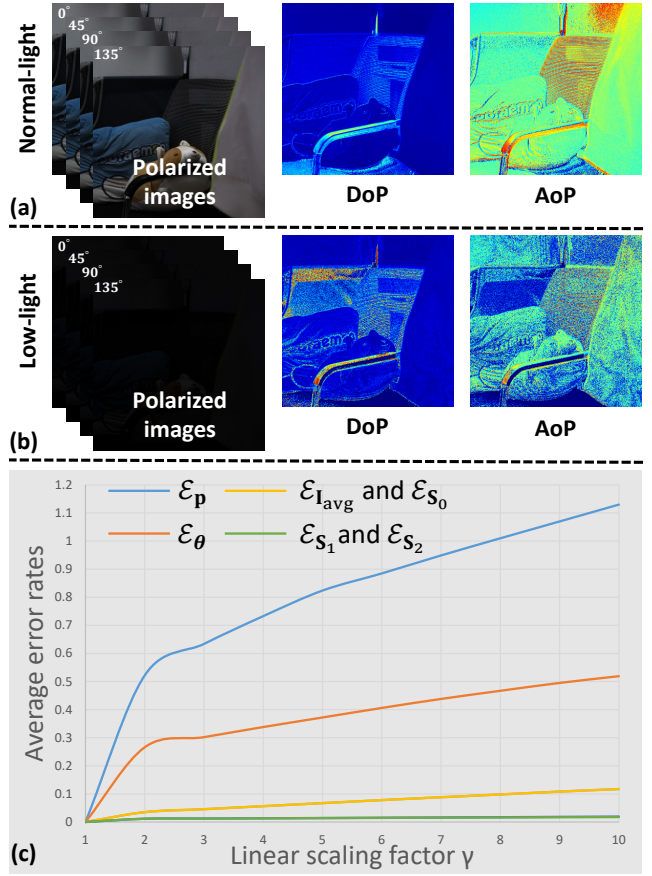


Figure 1: (a) Normal-light polarized images with the computed DoP and AoP. (b) Low-light polarized images with the degenerated DoP and AoP. (c) Relationships between the average error rates ( $\mathcal{E}_{\mathbf{I}_{\text{avg}}}$ ,  $\mathcal{E}_{\mathbf{S}_{0,1,2}}$ ,  $\mathcal{E}_{\mathbf{p}}$ ,  $\mathcal{E}_{\theta}$ ) and the linear scaling factor  $\gamma$ . We visualize the DoP and AoP (normalized to  $[0, 1]$ ) using color maps after averaging their RGB channels (as Hu et al. (2020) do) throughout this paper.

AoP are inherently different from those in the image intensity domain (Hu et al. 2020). According to Eq. (5), a possible solution could be adopting single-image low-light enhancement methods (Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020) to reconstruct  $\mathbf{I}_{\alpha_{1,2,3,4}}$  from  $\hat{\mathbf{I}}_{\alpha_{1,2,3,4}}$ . However, these methods cannot consider the polarization relationship among the polarized images and can only process them in a frame-by-frame manner, leaving the acquired DoP and AoP unreliable. Another solution is implicitly considering the physical constraints of polarization by enhancing  $\hat{\mathbf{I}}_{\alpha_{1,2,3,4}}$  simultaneously (Hu et al. 2020). However, the information provided by polarization cannot be effectively exploited since it still operates in the intensity domain.

From Eq. (3) we could know that the polarization characteristics of the incoming light to the sensor are fully encoded in the Stokes parameters  $\mathbf{S}_{0,1,2}$ , *i.e.*, one can render a polarized image with an arbitrary polarizer angle from them. Besides, from Eq. (10) and Fig. 1 (c) we can see that the level of degradation in  $\mathbf{S}_0$  is similar to the polarized images,

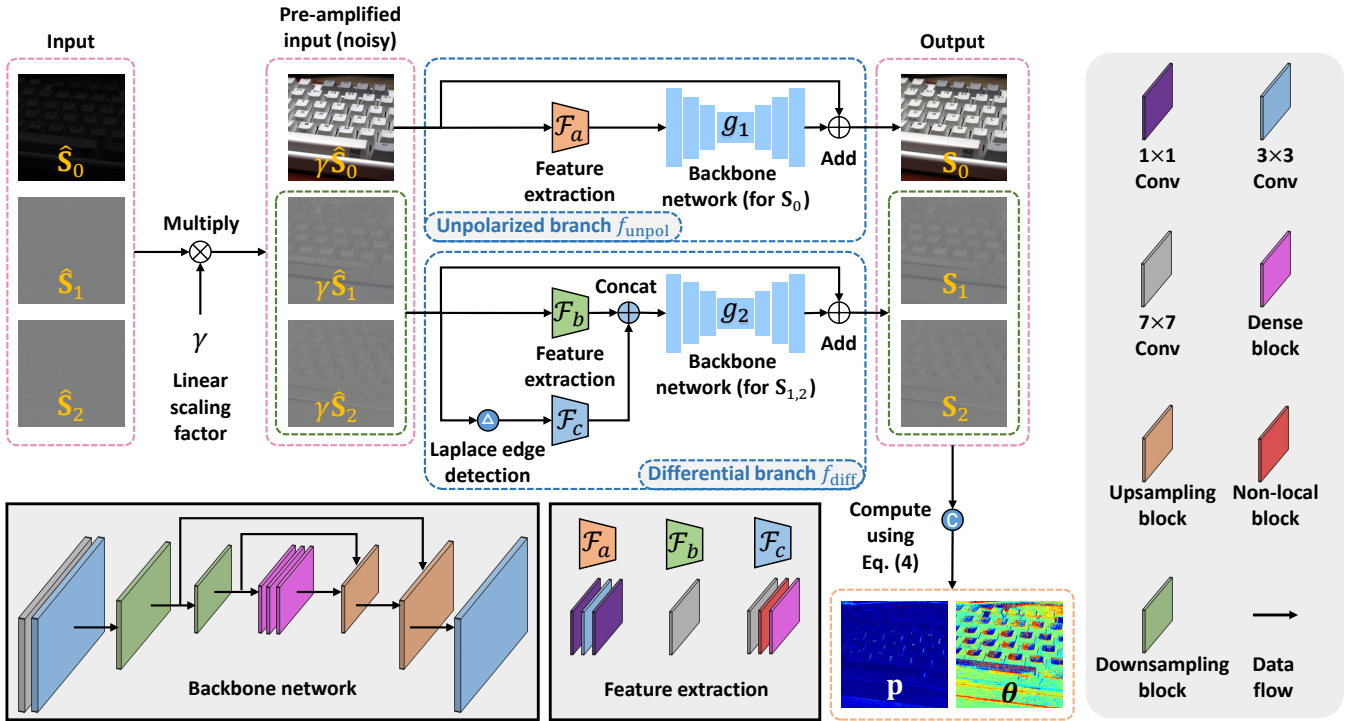


Figure 2: We design a network tailored to our Stokes-domain enhancement pipeline, which recovers the Stokes parameters  $\mathbf{S}_{0,1,2}$  from their degenerated counterparts  $\hat{\mathbf{S}}_{0,1,2}$ . It consists of two branches for enhancing  $\hat{\mathbf{S}}_0$  and  $\hat{\mathbf{S}}_{1,2}$  respectively based on their different properties. Then, the DoP  $\mathbf{p}$  and AoP  $\theta$  with high accuracy can be computed from  $\mathbf{S}_{0,1,2}$  using Eq. (4).

while  $\mathbf{S}_{1,2}$  are less degenerated, which suggests that enhancing the Stokes parameters may have chance to provide more reliable results. Therefore, we propose to enhance  $\hat{\mathbf{S}}_{0,1,2}$  instead of enhancing  $\hat{\mathbf{I}}_{\alpha_{1,2,3,4}}$  directly. Here, it is not wise to process  $\hat{\mathbf{S}}_{0,1,2}$  simultaneously since the properties of  $\mathbf{S}_0$  and  $\mathbf{S}_{1,2}$  are essentially different: (1)  $\mathbf{S}_0$  is the unpolarized image (Eq. (3)); (2)  $\mathbf{S}_{1,2}$  are two similar “differential signals”, which can be expressed as the differences between two polarized images (Eq. (5)). It might be a better idea to deal with  $\hat{\mathbf{S}}_0$  and  $\hat{\mathbf{S}}_{1,2}$  separately based on their different properties.

Therefore, we propose a Stokes-domain enhancement pipeline, which adopts two network branches to perform enhancement on  $\hat{\mathbf{S}}_0$  and  $\hat{\mathbf{S}}_{1,2}$  independently:

$$\mathbf{S}_0 = f_{\text{unpol}}(\hat{\mathbf{S}}_0) \quad \text{and} \quad \mathbf{S}_{1,2} = f_{\text{diff}}(\hat{\mathbf{S}}_{1,2}), \quad (11)$$

where  $f_{\text{unpol}}$  denotes the *unpolarized branch* for enhancing the unpolarized image  $\hat{\mathbf{S}}_0$  and  $f_{\text{diff}}$  denotes the *differential branch* for enhancing the “differential signals”  $\hat{\mathbf{S}}_{1,2}$ , which will be detailed in the next subsection. As  $\mathbf{S}_{0,1,2}$  become available, the DoP  $\mathbf{p}$  and AoP  $\theta$  with high accuracy can be calculated using Eq. (4), and the polarized images  $\mathbf{I}_{\alpha_{1,2,3,4}}$  can be calculated using Eq. (3).

### Polarization-Aware Dual-Branch Network

Tailored to our Stokes-domain enhancement pipeline, we design a dual-branch network to enhance  $\hat{\mathbf{S}}_0$  and  $\hat{\mathbf{S}}_{1,2}$  indepen-

dently in a polarization-aware manner based on their different properties, as shown in Fig. 2.

**Unpolarized Branch.** As shown in the upper branch of Fig. 2, it consists of a feature extraction block ( $\mathcal{F}_a$ ) to extract high-level features in the latent space, and a backbone network ( $g_1$ ) to perform blind noise suppression and detail enhancement on the pre-amplified input  $\gamma\hat{\mathbf{S}}_0$ . By adopting Tanh as the output activation, it learns the residual between  $\gamma\hat{\mathbf{S}}_0$  and  $\mathbf{S}_0$ , which can be described as

$$\mathbf{S}_0 = f_{\text{unpol}}(\hat{\mathbf{S}}_0) = g_1(\mathcal{F}_a(\gamma\hat{\mathbf{S}}_0)) + \gamma\hat{\mathbf{S}}_0. \quad (12)$$

**Differential Branch.** We observe that despite most regions in  $\gamma\hat{\mathbf{S}}_{1,2}$  are degenerated by noise, their edges are less affected, which could provide abundant structure information. So, we propose to explicitly extract their edges (denoted as  $\mathbf{E}_{1,2}$ ) using Laplace kernels as priors. As shown in the lower branch of Fig. 2, it consists of two feature extraction blocks ( $\mathcal{F}_b$  and  $\mathcal{F}_c$ ) to extract features from  $\gamma\hat{\mathbf{S}}_{1,2}$  and their edges respectively and a backbone network ( $g_2$ ) to complete the reconstruction. This branch can be described as ( $i = 1, 2$ )

$$\mathbf{S}_i = f_{\text{diff}}(\hat{\mathbf{S}}_i) = g_2(\text{concat}(\mathcal{F}_b(\gamma\hat{\mathbf{S}}_i), \mathcal{F}_c(\mathbf{E}_i))) + \gamma\hat{\mathbf{S}}_i. \quad (13)$$

**Layer Details.** The feature extraction blocks  $\mathcal{F}_a$  and  $\mathcal{F}_b$  consist of several convolution layers, which are quite simple since they only extract features directly from inputs.  $\mathcal{F}_c$  consists of a convolution layer, a dense block (Huang et al.



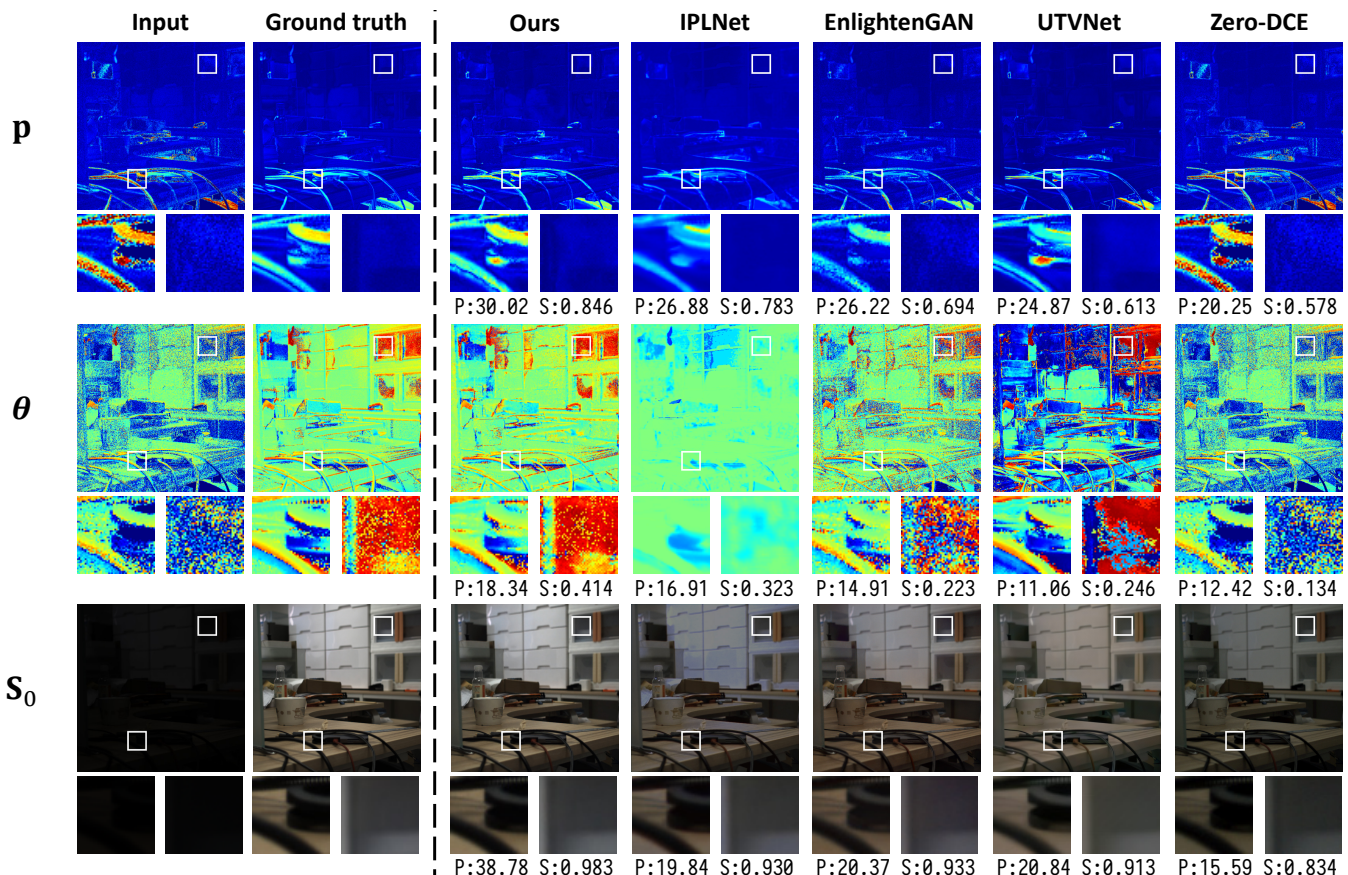


Figure 3: Qualitative evaluation results on the PLIE dataset among our method, IPLNet (Hu et al. 2020), EnlightenGAN (Jiang et al. 2021), UTVNet (Zheng, Shi, and Shi 2021), and Zero-DCE (Guo et al. 2020). Quantitative results evaluated using PSNR (P) and SSIM (S) are displayed below each image. Please zoom-in for better details.

2017), and a non-local block (Wang et al. 2018) since it extracts features from sparse edges, which requires large receptive fields and long-range dependencies. As for the backbone network  $g_i$  ( $i = 1, 2$ ), we design it as a modified autoencoder architecture (Hinton and Salakhutdinov 2006), by virtue of its excellent context generalization ability for enriching detail contents. We set the number of downsampling/upsampling blocks to 2 for multi-scale observations, embed 3 dense blocks (Huang et al. 2017) in the coarsest layer for more fine-grained contextual information, and add skip-connections to make full use of the shallow features. The downsampling block is a residual bottleneck block (He et al. 2016) enhanced with channel shuffle operations (Zhang et al. 2018) to help the information flow across feature channels. The upsampling block is similar to the corresponding one in Attention U-Net architecture (Oktay et al. 2018). Note that we add an instance normalization (Ulyanov, Vedaldi, and Lempitsky 2016) layer and a ReLU activation function after each convolution layer.

**Loss Function and Training Strategy.** The total loss function of our network is defined as  $\mathcal{L}(\mathbf{S}_{0,1,2}) = \lambda_1 \cdot \mathcal{L}_1(\mathbf{S}_{0,1,2}) + \lambda_2 \cdot \mathcal{L}_2(\mathbf{S}_{0,1,2}) + \lambda_3 \cdot \mathcal{L}_{tv}(\mathbf{S}_{0,1,2}) + \lambda_4 \cdot \mathcal{L}_{grad}(\mathbf{S}_{1,2})$ , where  $\mathcal{L}_1$  is the  $\ell_1$  loss,  $\mathcal{L}_2$  is the  $\ell_2$  loss,  $\mathcal{L}_{tv}$

is the total variation loss to enforce smoothness, and  $\mathcal{L}_{grad}$  is the gradient loss ( $\ell_2$  loss in the gradient domain) to ensure the structure invariance of  $\mathbf{S}_{1,2}$ .  $\lambda_i$  ( $i = 1, 2, 3, 4$ ) are empirically set to be 10.0, 100.0, 1.0, 100.0 respectively. We implement the network using PyTorch on an NVIDIA 2080Ti GPU, and train it for 400 epochs using ADAM optimizer (Kingma and Ba 2014) with a batch size of 8. The learning rate is set to 0.01.

## Experiments

### Comparisons with Existing Methods

There is no public dataset for such a polarization-aware low-light image enhancement task. Besides, existing single-image low-light enhancement benchmark datasets (*e.g.*, (Lv, Li, and Lu 2021; Chen et al. 2018)) do not contain any polarization information, which cannot be used to generate polarized images. Therefore, we propose to build a real-world dataset, named PLIE (Polarization-aware Low-light Image Enhancement) dataset<sup>5</sup>, which contains pairwise low- and normal-light polarized images to train our network and test it quantitatively and qualitatively.

<sup>5</sup>More information can be found in the supplementary material.

	PSNR- $\mathbf{p}$	SSIM- $\mathbf{p}$	PSNR- $\boldsymbol{\theta}$	SSIM- $\boldsymbol{\theta}$	PSNR- $\mathbf{S}_0$	SSIM- $\mathbf{S}_0$
Ours	<b>27.15</b>	<b>0.765</b>	<b>16.42</b>	<b>0.336</b>	<b>39.19</b>	<b>0.977</b>
IPLNet (Hu et al. 2020)	25.32	0.715	16.21	0.276	22.84	0.930
EnlightenGAN (Jiang et al. 2021)	24.55	0.652	13.55	0.190	22.14	0.887
UTVNet (Zheng, Shi, and Shi 2021)	24.14	0.636	12.18	0.271	18.45	0.821
Zero-DCE (Guo et al. 2020)	19.34	0.527	12.09	0.134	17.48	0.815

Table 1: Quantitative evaluation results on the PLIE dataset among our method, IPLNet (Hu et al. 2020), EnlightenGAN (Jiang et al. 2021), UTVNet (Zheng, Shi, and Shi 2021), and Zero-DCE (Guo et al. 2020). Bold font indicates the best performance.

	PSNR- $\mathbf{p}$	SSIM- $\mathbf{p}$	PSNR- $\boldsymbol{\theta}$	SSIM- $\boldsymbol{\theta}$	PSNR- $\mathbf{S}_0$	SSIM- $\mathbf{S}_0$
Intensity-domain enhancement	24.98	0.702	16.38	0.286	38.00	0.972
Single-branch network	26.34	0.756	15.87	0.328	37.19	0.928
Without edge priors	26.13	0.751	16.26	0.330	39.17	0.975
Without gradient loss	20.63	0.485	14.67	0.181	37.36	0.973
Without total variation loss	15.70	0.576	14.35	0.250	37.40	0.970
Our complete model	<b>27.15</b>	<b>0.765</b>	<b>16.42</b>	<b>0.336</b>	<b>39.19</b>	<b>0.977</b>

Table 2: Quantitative evaluation results of ablation study.

We compare our method to IPLNet<sup>6</sup> (Hu et al. 2020) (the only existing method designed for enhancing polarized low-light images as far as we know), and three state-of-the-art single-image low-light enhancement methods including EnlightenGAN (Jiang et al. 2021), UTVNet (Zheng, Shi, and Shi 2021), and Zero-DCE (Guo et al. 2020) on the PLIE dataset. We do not compare to image denoising methods in this paper since IPLNet (Hu et al. 2020) has made such comparisons with Polarization-BM3D (Tibbs et al. 2018) (the state-of-the-art polarized image denoising method). Note that comparing with single-image low-light enhancement methods might be a bit unfair because of the difference in the way of processing the input data, and we just attempt to show the significance of polarization-awareness.

As Hu et al. (2020) do, we not only evaluate the accuracy of the enhanced DoP  $\mathbf{p}$  and AoP  $\boldsymbol{\theta}$ , but also evaluate the quality of the enhanced unpolarized image  $\mathbf{S}_0$ , since one can render a polarized image with an arbitrary polarizer angle using Eq. (3) when  $\mathbf{p}$ ,  $\boldsymbol{\theta}$ , and  $\mathbf{S}_0$  are available. Note that we only re-train IPLNet (Hu et al. 2020) on the PLIE dataset, while directly adopting the pre-trained models for those single-image low-light enhancement methods (Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020), since the performance of those single-image methods degenerates after re-training. This is because those methods rely strongly on the semantic information they extracted for enhancement; when training on the PLIE dataset (which is not as large as the dataset used for obtaining the pre-trained models) the semantic information is limited.

Visual quality comparisons are shown in Fig. 3<sup>7</sup>. As for  $\mathbf{p}$  and  $\boldsymbol{\theta}$ , our method achieves much better performance than other methods, thanks to our polarization-aware network.

<sup>6</sup>The code of IPLNet (Hu et al. 2020) is not available and the demonstrated results are based on our own implementation according to the descriptions in the paper.

<sup>7</sup>Additional results can be found in the supplementary material.

Taking the left box of  $\mathbf{p}$  as an example, the results of single-image low-light enhancement methods (Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020) suffer severely from artifacts, and IPLNet (Hu et al. 2020) tends to generate over-smooth results. This is because the single-image low-light enhancement methods (Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020) process the polarized images in a frame-by-frame manner so that they are not aware of the polarization relationship among the polarized images, while IPLNet (Hu et al. 2020) still handles the problem in the intensity domain so that the information provided by polarization cannot be effectively exploited. As for  $\mathbf{S}_0$ , our results resemble the ground truth more closely with less color distortion. This is because the methods operating in the intensity domain (Hu et al. 2020; Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020) need to compute  $\mathbf{S}_0$  from  $\mathbf{I}_{\alpha_{1,2,3,4}}$ , while our unpolarized branch focuses on enhancing  $\mathbf{S}_0$  directly, which can avoid error accumulation during the computation. To evaluate the results quantitatively, as Hu et al. (2020) do, we adopt two frequently-used metrics including PSNR and SSIM. Results are shown in Tab. 1 (also below corresponding examples in Fig. 3). Our model consistently outperforms the compared methods on all metrics.

## Ablation Study

We conduct a series of ablation studies and show comparisons in Tab. 2. We first show the contribution of our Stokes-domain enhancement pipeline by comparing with a model that performs enhancement in the intensity domain. We find that our Stokes-domain pipeline is better since it can make effective use of the information provided by polarization. We further verify the effectiveness of our dual-branch network structure (Fig. 2) by comparing with a model that uses only a single branch to estimate  $\mathbf{S}_{0,1,2}$  simultaneously. From the results we can see that our network design is more reasonable and robust. Then, we demonstrate the necessity of extracting

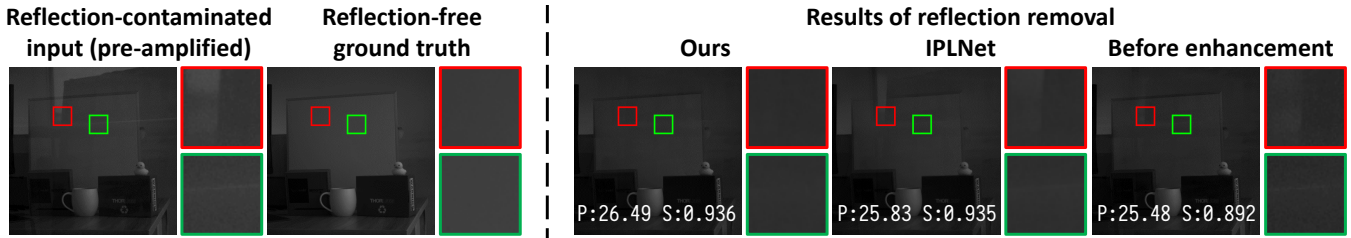


Figure 4: Results of reflection removal (using PRRPAW (Lei et al. 2020)) before and after enhancement by our method and IPLNet (Hu et al. 2020). Quantitative results evaluated using PSNR (P) and SSIM (S) are displayed in each image.

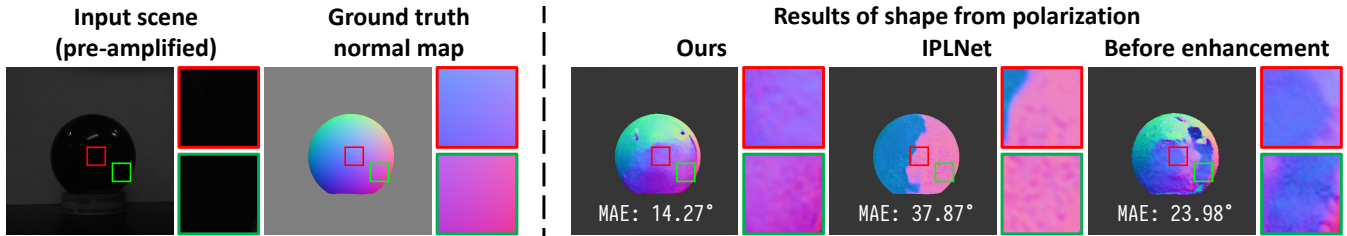


Figure 5: Results of shape from polarization (using DP3I (Deschaintre, Lin, and Ghosh 2021)) before and after enhancement by our method and IPLNet (Hu et al. 2020). Quantitative results evaluated using mean angle error (MAE) are displayed in each image (lower means better).

the edges of  $S_{1,2}$  as priors by removing them, and validate the significance of adopting gradient loss and total variation loss by removing them respectively. These results show that our complete model achieves the first performance with the proposed specific designs.

## Applications

To demonstrate the benefits of enhancing polarized low-light images, we choose two typical applications including reflection removal (Lei et al. 2020) and shape from polarization (Deschaintre, Lin, and Ghosh 2021) and show that the enhancement can improve their performance. Note that at this point we only compare to IPLNet (Hu et al. 2020) due to the inferior performance of single-image low-light enhancement methods (Jiang et al. 2021; Zheng, Shi, and Shi 2021; Guo et al. 2020) on the PLIE dataset according to Tab. 1.

**Reflection Removal.** We choose the state-of-the-art polarization-based reflection removal method PRRPAW (Lei et al. 2020) for validation. First, we use a Lucid Vision Phoenix polarization camera to capture reflection-contaminated polarized low-light images behind a piece of glass with a short exposure time  $t_{\text{short}}$  and adopt our method and IPLNet (Hu et al. 2020) to enhance them respectively. Then, we take the enhanced images (converted to grayscale) as the input of PRRPAW (Lei et al. 2020) to obtain the reflection removed results. For reference, we also capture the reflection-free ground truth images with a long exposure time  $t_{\text{long}} = 10t_{\text{short}}$  by removing the glass. Comparisons of the reflection-removed grayscale unpolarized images<sup>8</sup> are

<sup>8</sup>PRRPAW (Lei et al. 2020) takes polarized images as input and outputs unpolarized results in grayscale.

shown in Fig. 4<sup>9</sup>. We can see that PRRPAW (Lei et al. 2020) cannot remove the reflection adequately in low-light conditions, and our method can improve its performance by providing the enhanced polarized images, DoP, and AoP as its input. Our method outperforms IPLNet (Hu et al. 2020) both quantitatively and qualitatively.

**Shape from Polarization.** We choose the state-of-the-art shape from polarization method DP3I (Deschaintre, Lin, and Ghosh 2021) for validation. We directly capture the polarized low-light images as the input of our method and IPLNet (Hu et al. 2020). Comparisons of the estimated normal maps are shown in Fig. 5<sup>10</sup> (the ground truth normal map is computed analytically according to the contour of the sphere). We can see that our method could improve the performance of DP3I (Deschaintre, Lin, and Ghosh 2021) by a large margin, while IPLNet (Hu et al. 2020) brings negative effects since it generates over-smooth DoP and AoP, providing unreliable physical constraints for such an application.

## Conclusion

We presented a learning-based solution to enhance multiple polarized low-light images for enhancing the accuracy of the DoP and AoP. To make effective use of the information provided by polarization, we proposed a Stokes-domain enhancement pipeline along with a dual-branch neural network, handling the problem in a polarization-aware manner. We also demonstrated that our method can improve the performance of applications of polarization-based vision in low-light conditions, including reflection removal and shape from polarization.

<sup>9</sup>Additional results can be found in the supplementary material.

<sup>10</sup>Additional results can be found in the supplementary material.

## Acknowledgements

This work is supported by National Key R&D Program of China (2021ZD0109803), and National Natural Science Foundation of China under Grant No. 62136001, 62088102, 62276007, and 61876007.

## References

- Ai, S.; and Kwon, J. 2020. Extreme low-light image enhancement for surveillance cameras using attention U-Net. *Sensors*, 20(2): 495.
- Atoum, Y.; Ye, M.; Ren, L.; Tai, Y.; and Liu, X. 2020. Color-wise attention network for low-light image enhancement. In *Proc. of Computer Vision and Pattern Recognition Workshops*.
- Chen, C.; Chen, Q.; Do, M. N.; and Koltun, V. 2019. Seeing motion in the dark. In *Proc. of Computer Vision and Pattern Recognition*.
- Chen, C.; Chen, Q.; Xu, J.; and Koltun, V. 2018. Learning to see in the dark. In *Proc. of Computer Vision and Pattern Recognition*.
- Deschaintre, V.; Lin, Y.; and Ghosh, A. 2021. Deep polarization imaging for 3D shape and SVBRDF acquisition. In *Proc. of Computer Vision and Pattern Recognition*.
- Fan, M.; Wang, W.; Yang, W.; and Liu, J. 2020. Integrating semantic segmentation and retinex model for low-light image enhancement. In *Proc. of ACM MM*.
- Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proc. of Computer Vision and Pattern Recognition*.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proc. of Computer Vision and Pattern Recognition*.
- Hecht, E. 2012. *Optics*. Pearson Education India.
- Hinton, G. E.; and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504–507.
- Hu, H.; Lin, Y.; Li, X.; Qi, P.; and Liu, T. 2020. IPLNet: A neural network for intensity-polarization imaging in low light. *Optics Letters*, 45(22): 6162–6165.
- Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proc. of Computer Vision and Pattern Recognition*.
- Jiang, H.; and Zheng, Y. 2019. Learning to see moving objects in the dark. In *Proc. of International Conference on Computer Vision*.
- Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Wang, Z. 2021. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30: 2340–2349.
- Kingma, D. P.; and Ba, J. 2014. ADAM: A method for stochastic optimization. arXiv:1412.6980.
- Können, G. 1985. *Polarized light in nature*. CUP Archive.
- Lamba, M.; and Mitra, K. 2021. Restoring extremely dark images in real time. In *Proc. of Computer Vision and Pattern Recognition*.
- Land, E. H. 1977. The retinex theory of color vision. *Scientific American*, 237(6): 108–129.
- Lei, C.; Huang, X.; Zhang, M.; Yan, Q.; Sun, W.; and Chen, Q. 2020. Polarized reflection removal with perfect alignment in the wild. In *Proc. of Computer Vision and Pattern Recognition*.
- Li, C.; Guo, C.; and Chen, C. L. 2021. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Li, C.; Guo, J.; Porikli, F.; and Pang, Y. 2018. LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104: 15–22.
- Li, J.; Feng, X.; and Hua, Z. 2021. Low-light image enhancement via progressive-recursive network. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(11): 4227–4240.
- Li, J.; Li, J.; Fang, F.; Li, F.; and Zhang, G. 2020. Luminance-aware pyramid network for low-light image enhancement. *IEEE Transactions on Multimedia*, 23: 3153–3165.
- Lim, S.; and Kim, W. 2020. DSLR: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23: 4272–4284.
- Liu, R.; Ma, L.; Zhang, J.; Fan, X.; and Luo, Z. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proc. of Computer Vision and Pattern Recognition*.
- Lore, K. G.; Akintayo, A.; and Sarkar, S. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61: 650–662.
- Lu, K.; and Zhang, L. 2020. TBEFN: A two-branch exposure-fusion network for low-light image enhancement. *IEEE Transactions on Multimedia*, 23: 4093–4105.
- Lv, F.; Li, Y.; and Lu, F. 2021. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7): 2175–2193.
- Lv, F.; Liu, B.; and Lu, F. 2020. Fast enhancement for non-uniform illumination images using light-weight CNNs. In *Proc. of ACM MM*.
- Lv, F.; Lu, F.; Wu, J.; and Lim, C. 2018. MBLLEN: Low-light image/video enhancement using CNNs. In *Proc. of British Machine Vision*.
- Lyu, Y.; Cui, Z.; Li, S.; Pollefeys, M.; and Shi, B. 2019. Reflection separation using a pair of unpolarized and polarized images. In *Proc. of Advances in Neural Information Processing Systems*.
- Maharjan, P.; Li, L.; Li, Z.; Xu, N.; Ma, C.; and Li, Y. 2019. Improving extreme low-light image denoising via residual learning. In *Proc. of International Conference on Multimedia and Expo*.
- Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N. Y.; Kainz, B.; Glocker, B.; and Rueckert, D. 2018. Attention U-Net: Learning where to look for the pancreas. arXiv:1804.03999.



- Pizer, S. M.; Amburn, E. P.; Austin, J. D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J. B.; and Zuiderveld, K. 1987. Adaptive histogram equalization and its variations. *Computer vision, Graphics, and Image Processing*, 39(3): 355–368.
- Ren, W.; Liu, S.; Ma, L.; Xu, Q.; Xu, X.; Cao, X.; Du, J.; and Yang, M.-H. 2019. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9): 4364–4375.
- Tibbs, A. B.; Daly, I. M.; Roberts, N. W.; and Bull, D. R. 2018. Denoising imaging polarimetry by adapted BM3D method. *Journal of the Optical Society of America*, 35(4): 690–701.
- Triantafyllidou, D.; Moran, S.; McDonagh, S.; Parisot, S.; and Slabaugh, G. 2020. Low light video enhancement using synthetic data produced with an intermediate domain mapping. In *Proc. of European Conference on Computer Vision*.
- Ulyanov, D.; Vedaldi, A.; and Lempitsky, V. 2016. Instance normalization: The missing ingredient for fast stylization. arXiv:1607.08022.
- Wang, L.-W.; Liu, Z.-S.; Siu, W.-C.; and Lun, D. P. 2020. Lightening network for low-light image enhancement. *IEEE Transactions on Image Processing*, 29: 7984–7996.
- Wang, R.; Xu, X.; Fu, C.-W.; Lu, J.; Yu, B.; and Jia, J. 2021. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Proc. of International Conference on Computer Vision*.
- Wang, R.; Zhang, Q.; Fu, C.-W.; Shen, X.; Zheng, W.-S.; and Jia, J. 2019a. Underexposed photo enhancement using deep illumination estimation. In *Proc. of Computer Vision and Pattern Recognition*.
- Wang, X.; Girshick, R.; Gupta, A.; and He, K. 2018. Non-local neural networks. In *Proc. of Computer Vision and Pattern Recognition*.
- Wang, Y.; Cao, Y.; Zha, Z.-J.; Zhang, J.; Xiong, Z.; Zhang, W.; and Wu, F. 2019b. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In *Proc. of ACM MM*.
- Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. In *Proc. of British Machine Vision*.
- Wei, K.; Fu, Y.; Yang, J.; and Huang, H. 2020. A physics-based noise formation model for extreme low-light raw denoising. In *Proc. of Computer Vision and Pattern Recognition*.
- Xu, K.; Yang, X.; Yin, B.; and Lau, R. W. 2020. Learning to restore low-light images via decomposition-and-enhancement. In *Proc. of Computer Vision and Pattern Recognition*.
- Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proc. of Computer Vision and Pattern Recognition*.
- Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2021a. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Transactions on Image Processing*, 30: 3461–3473.
- Yang, W.; Wang, W.; Huang, H.; Wang, S.; and Liu, J. 2021b. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30: 2072–2086.
- Yu, R.; Liu, W.; Zhang, Y.; Qu, Z.; Zhao, D.; and Zhang, B. 2018. DeepExposure: Learning to expose photos with asynchronously reinforced adversarial learning. In *Proc. of Advances in Neural Information Processing Systems*.
- Zhang, F.; Li, Y.; You, S.; and Fu, Y. 2021a. Learning temporal consistency for low light video enhancement from single images. In *Proc. of Computer Vision and Pattern Recognition*.
- Zhang, L.; Zhang, L.; Liu, X.; Shen, Y.; Zhang, S.; and Zhao, S. 2019. Zero-shot restoration of back-lit images using deep internal learning. In *Proc. of ACM MM*.
- Zhang, X.; Zhou, X.; Lin, M.; and Sun, J. 2018. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In *Proc. of Computer Vision and Pattern Recognition*.
- Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; and Zhang, J. 2021b. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4): 1013–1037.
- Zhang, Y.; Zhang, J.; and Guo, X. 2019. Kindling the darkness: A practical low-light image enhancer. In *Proc. of ACM MM*.
- Zhao, Z.; Xiong, B.; Wang, L.; Ou, Q.; Yu, L.; and Kuang, F. 2021. RetinexDIP: A unified deep framework for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3): 1076–1088.
- Zheng, C.; Shi, D.; and Shi, W. 2021. Adaptive unfolding total variation network for low-light image enhancement. In *Proc. of International Conference on Computer Vision*.
- Zhou, C.; Teng, M.; Han, Y.; Xu, C.; and Shi, B. 2021. Learning to dehaze with polarization. In *Proc. of Advances in Neural Information Processing Systems*.
- Zhu, A.; Zhang, L.; Shen, Y.; Ma, Y.; Zhao, S.; and Zhou, Y. 2020a. Zero-shot restoration of underexposed images via robust retinex decomposition. In *Proc. of International Conference on Multimedia and Expo*.
- Zhu, M.; Pan, P.; Chen, W.; and Yang, Y. 2020b. EEMEFN: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proc. of the AAAI Conference on Artificial Intelligence*.