Polarized Color Image Denoising

Zhuoxiao Li Haiyang Jiang Mingdeng Cao Yinqiang Zheng*
The University of Tokyo

Abstract

Single-chip polarized color photography provides both visual textures and object surface information in one snapshot. However, the use of an additional directional polarizing filter array tends to lower photon count and SNR, when compared to conventional color imaging. As a result, such a bilayer structure usually leads to unpleasant noisy images and undermines performance of polarization analysis, especially in low-light conditions. It is a challenge for traditional image processing pipelines owing to the fact that the physical constraints exerted implicitly in the channels are excessively complicated. In this paper, we propose to tackle this issue through a noise modeling method for realistic data synthesis and a powerful network structure inspired by vision Transformer. A real-world polarized color image dataset of paired raw short-exposed noisy images and long-exposed reference images is captured for experimental evaluation, which has demonstrated the effectiveness of our approaches for data synthesis and polarized color image denoising. The code and data can be found at https://github.com/bandasyou/pcdenoise.

1. Introduction

A beam of light can be considered as a combination of linearly polarized lights oscillating in different planes perpendicular to the imaging plane. The polarized components vary in specific ways up to the object refractive index and incident angle while they are reflected by metallic or dielectric materials. Therefore, polarized reflections convey information about object materials and surface geometries independent of light intensities and surface textures. For this reason, polarization photography plays a crucial role in transparent reflection removal [27, 35], shape-from-polarization [16, 20] and so on.

Formerly, we were only able to capture chromatic or polarimetric information of a scene separately. Thanks to the latest single-chip polarized color sensors (e.g., IMX250MYR), in which an array of directional polarizing filters of four directions (0°, 45°, 90°,135°) is equipped between the Bayer filter and the CMOS sensor, as shown in Fig. 1a, it is now effortless to capture polarization and color information simultaneously. However, the sensors' bilayer layout causes lower photon counts and Signal-to-Noise Ratio (SNR), when compared to typical color imaging. As shown in Fig. 1b, the intensity of a beam will be attenuated by a polarizer. This leads to noisy images (see Fig. 1c, d) and can be harmful to subsequent polarization analysis. This issue motivates us to restore clean signals from noisy polarized color images.

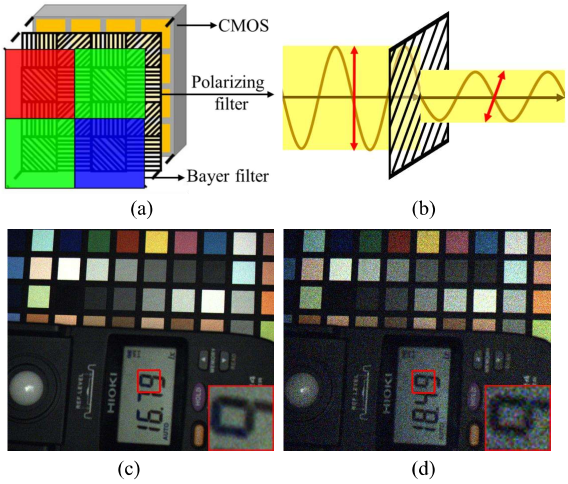


Figure 1. (a) Illustration of the polarized color sensor. (b) The intensity of polarized light decreases after passing through a polarizer. (c) Image produced by a conventional camera (FLIR BFS-U3-51S5C) with a 40x digital magnification. (d) Image captured by a polarized color camera (FLIR BFS-U3-51S5PC) with an 80x digital magnification. The two cameras are equipped with the same type of lens and CMOS sensors, and capture the illuminometer and color checkerboard with the same exposure time and aperture. The comparison demonstrates that even under slightly brighter illumination, polarized color images are noisier.

In the past decades, a number of methods have been proposed for color image denoising [5, 12, 21, 34, 47, 52, 53], which underline the key role of realistic paired data, especially for supervised learning methods. To the best of our knowledge, there does not exist any real-world dataset for denosing polarized color images.

[*] Corresponding author:yqzheng@ai.u-tokyo.ac.jp.

We collect a real-world dataset of noisy polarized color raw data taken under a wide spectrum of low-light settings. The corresponding clean images are captured with low-gain long-exposure settings as ground truth. In spite of the tremendous efforts we payed on data collection, the noise volumes and variations may still be limited. From previous observations, a precise noise model can be used to generate realistic noisy images and benefits learning-based approaches [8, 45] thanks to the infinite noise patterns. It remains unknown how a single-chip polarized color sensor differs from a standard color sensor and to what extend existing physics-based noise model [45] should be adapted. In this paper, we highlight the unique noise characteristics of polarized color sensor via extensive analysis.

Recent researches have shown that Vision Transformer [43] has enormous potentials in exploring global context interactions and outperforms convolutional DNNs in various vision problems [7, 11, 17, 31, 33]. While spatial [44] and channel self-attention [6, 49] take a dominant place in the image restoration realm, we consider them together to boost signal denoising and polarization restoration. Therefore, we propose a Transformer model with hybrid attention mechanisms for effective polarized color image denoising. We have validated the advantages of our method over traditional methods and recently proposed network architectures, as well as its usefulness for down-streaming applications.

2. Related work

2.1. Image Denoising

Image denoising is a well-researched yet still active topic in the computer vision community. Traditional single image denoising methods proposed elaborate models based on total variation [41], SVD [39], sparse coding [18], self-similarity [36], and so on. Recent researches mainly focused on the great capacities and expressiveness of CNNs, which adaptively learn data-driven denoisers from noisy images and their noise-free counterparts [5, 12, 32, 52, 53]. However, in real-world noise image processing scenarios, these methods have been proven to be outperformed by BM3D [15]. The reason is primarily that the learning-based models overfit to the synthetic training data constructed by over-simplified noise models [38]. Therefore, researches captured clean/noisy image pairs to build dataset with real-world noises [2, 9] for both model training and evaluation.

However, it is labor-intensive to acquire a large volume of labeled high-quality data. Thus, another line of research has tried to generate realistic noise data from clean images. [4] employed a signal-dependent heteroscedastic Gaussian model [19] to simulate both intensity-dependent and intensity-independent noises. Some works [1, 8] applied generative models to learn the latent noise distribution from real noisy images. Beyond the oversimplified models, [45] tried to dig into the physics-based electronic imaging pipeline to formulate and estimate the noise distributions. The calibrated camera-aware parameters help to generate rich noise patterns only from clean images and benefit the denoising performance. [54] bypassed the complex noise modeling process and directly sampled real readout noises from light-free frames. As the first research on polarized color image denoising, we systematically analyze the noise characteristics of the polarized color sensor based on the physics-based noise formation model [45].

There are some restoration methods proposed for monochromatic polarization images [3, 29, 50, 51], but they can not directly handle polarized color images and lack full consideration of polarization properties.

2.2. Vision Transformer

Inspired by the great success of NLP Transformers, Vision Transformer (ViT) [17] utilized the Transformer technique on non-overlapping cropped image patches that achieved accurate yet efficient image classification. Local-window-based Transformer models [33] with various window partition strategies were proposed and achieved considerable improvements on the speed-accuracy trade-off. In SwinIR [31], Swin Transformer [33] was used to build a single-scale architecture for high-quality image restoration. Moreover, a lot of works [44, 49] tried to apply Transformers to build U-shaped networks that make further use of the hierarchical structure of Swin and multi-scale skip connections of U-Net [40].

2.3. Exploiting Polarization

As polarization photography can capture more channels of information exhibited implicitly from environments, it has been extensively studied in both fields of computer vision and computer graphics. A few decades ago, polarization was found to be effective in reflection removal [42], whose performance has been greatly improved via learning based approaches [27, 35]. Furthermore, Shape-from-Polarization (SfP) has developed into a popular research area in recent years. Surface normal information encoded by light polarization was introduced into conventional 3D modeling pipelines [14, 48] and binocular stereo camera systems [20, 55] to produce more precise and complete geometries. Coupling polarization cues with CNNs, 3D object shape and human shape can be easily reconstructed with a single-view polarization image [16, 56].

3. Noise Model

3.1. Model Formation

The CMOS sensor covered by a polarizer suffers from significant degradation in quantum efficiency compared to conventional sensors [26], which leads to massive noises and tends to destroy fragile polarization information, especially in low-light environments. We use the physics-based noise formation model informed by [45] to systematically analyze the noises sources in the physical imaging process, and find some unique characteristics of polarized color senor.

The final signal combined with the physics-based noise model is written as:

$$D = KI + KN_p + N_{read} + N_b + N_q, \tag{1}$$

where I represents the number of photons captured by a CMOS sensor, K is the overall system gain. N_p, N_{read}, N_b and N_q count for photon shot noise, signal readout noise, banding pattern noise, and quantization noise.

Photon shot noise N_p is the major type of noise from incident light. It is caused by the quantum nature of light when the photoelectrons are randomly emancipated from the semiconductor after photon hits. The noise distribution follows Poisson statistics as

$$(I + N_p) \sim \mathcal{P}(I), \tag{2}$$

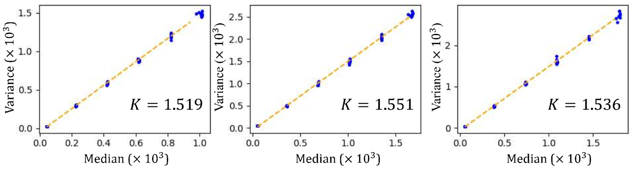
where $\mathcal{P}$ is the Poisson distribution. There are other noise sources when capturing photons, e.g., crosstalk effect [22] of neighboring pixels that cause inaccurate intensity. To alleviate this issue, the most advanced manufacturing technology is to install the polarizer array between the microlens array and pixel array. This is shown to be reliable to reduce crosstalk and promote extinction ratios, thus we ignore this factor since the senor under study is produced like that.

Read noise N_{read} counts for multiple noise sources including dark current noise [22], thermal noise, and source follower noise [28]. Although a zero-mean Gaussian model is mostly used [15,52] and capable of covering a wide range of noises, [22,45] show that real noise contains a long-tailed effect. Thus, a Tukey lambda (TL) distribution [24], consisting of a family of distributions, is able to handle the long-tailed feature:

$$N_{read} \sim \mathcal{T}\left(\lambda; \mu, \sigma_{\mathcal{T}}\right), \tag{3}$$

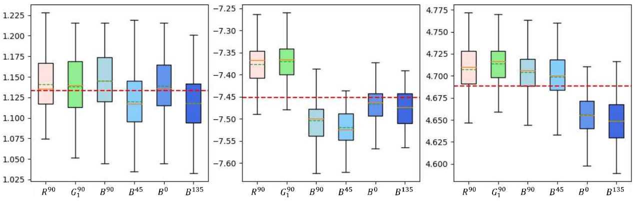
where $\mathcal{T}$ denotes the TL distribution, λ, μ and $\sigma_{\mathcal{T}}$ are shape, location and scale parameters. Specifically, μ is used to model the color-wise biases [45] arising from direct current noise.

Banding pattern noise N_b is caused by the CMOS circuit readout strategy and conveys a specific horizontal line style. The horizontal stripes can be simulated by zero-mean Gaussian samplings with a scale parameter σ_b. A sampled row noise is added to pixels within a same row.



(a) R^{90} (b) G_1^{90} (b) G_1^{45}

Figure 2. Estimation of the overall system gain for three channels, at the gain=12 setting. Given a flat filed frame, its variance and intensity (reprensented by median) satisfy a linear function, and overall system gain equals to the slope of the fitted line. The results indicate photoresponse nonuniformity.



(a) $\boldsymbol{\sigma}_b$ (b) $\boldsymbol{\mu}$ (c) $\boldsymbol{\sigma}_{\mathcal{T}}$

Figure 3. Box plots for parts of $\boldsymbol{\sigma}_b$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_{\mathcal{T}}$ samples at the gain=24 setting. The labels present their channels, the red dash line is their mean value, orange line and green dashed line in the boxes denote the median and mean value of each sample sequence. The comparison illustrates differences of the parameters' value and distribution of the channels.

Quantization noise is produced when a continuous voltage signal is converted to discrete digits via an AD converter. The quantization noise is simply formulated as:

$$N_q \sim U\left(-\frac{1}{2q}, \frac{1}{2q}\right), \tag{4}$$

where U denotes a uniform distribution and q represents the quantization step.

3.2. Noise Model Calibration

Following the pipeline of [45], we calibrate the parameters of the polarization color camera FLIR BFS-U3-51S5PC equipped with a Sony IMX250MYR sensor. To better reveal parameter distributions and correlations, we calibrate them on individual channels, which leads to 16-dimension parameter vectors. The corresponding parameters include overall system gain $\boldsymbol{K}$ for photon shot noise N_p; shape $\boldsymbol{\lambda}$, location $\boldsymbol{\mu}$ and scale $\boldsymbol{\sigma}_{\mathcal{T}}$ for read noise N_{read}; scale parameter $\boldsymbol{\sigma}_b$ for row banding noise. For better understanding, we apply a combined notation of [R, G_1, B, G_2] and [90, 45, 0, 135], e.g., R^{90}, to denote the channels characterized for colors (Red, Green1, Blue, and Green2) and polarizer angles (90°, 45°, 0°, 135°). We captured flat-field frames to estimate $\boldsymbol{K}$ and bias frames to calibrate other parameters. As an industrial camera, its settings are configurable, such as camera gain, black level, white balance ratios, and so on. Note that different from above mentioned overall system gain, the camera analog gain is a controllable parameter, which is equivalent to ISO widely used in consumer digital cameras.

Flat-field frames are images captured under uniform illumination. They can be used to estimate K via Photon Transfer (PT) method [23]. We capture flat-field frames of a whiteboard placed under natural illumination and the lens focuses to infinity to reduce photon non-uniformity. $\boldsymbol{K}$ has to be computed on 16 individual channels due to their unmatched intensities caused by varied quantum efficiencies [26] and linearly polarized reflected light. The result reveals noticeable photoresponse nonuniformity with the Bayer pattern and slight differences in polarization channels, as shown in Fig. 2.

Given a clean image and estimated K, we are able to convert pixels' intensity D into the number of electrons I by packing the image into 16 channels and dividing corresponding K separately. Then, following the Poisson distribution, random Poisson variate on I generates discrete noisy photon intensity. Finally, I is reversed back to D to simulate a real shot noise formation routine.

Bias frames are images captured in light-free environments where only intensity-independent signals are stored. To fully observe the complete noise distribution, we set the black level as 2% of the maximum signal with black level auto clamping turned off and captured 100 frames for each gain. Following [45], we estimated the parameters of N_{read} and N_b. Specifically, industrial applications require capturing fast-moving objects with high frame rates. Thus the sensor IMX250MYR employs global shutter techniques on a CMOS sensor to avoid the focal plane distortion problem that may be caused by a rolling shutter, and specific devices and technologies are equipped for high-quality imaging, e.g., an analog memory is provided for each pixel. The above techniques may cause specific noise patterns.

As a result, we first observed a global bias decreases rapidly below the black level with lifted camera gains and further variations exhibited in channel-wise biases, which causes a noticeable intensity decline and polarization distortion. Then, we observed that each odd row and its next row suffer the same row noise, which we assume is caused by the sensor's specific readout circuits that process two rows as a group. Significant diversity is finally observed for the parameters of different channels. As illustrated in Fig. 3, box plots indicate the distribution of a channel's parameters, i.e., $\boldsymbol{\sigma}_b$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_\tau$ differs from other color and polarization channels. Such unique properties make the simulation of realistic noises for polarized color sensors much more difficult, compared to typical color sensors.

3.3. Joint Distribution Model

As the above-mentioned parameters are detected for discrete camera gain level samples, e.g., $[0, 6, 12, 18, 24, 30]$, it remains a challenge to reform a set of 16-dimensional

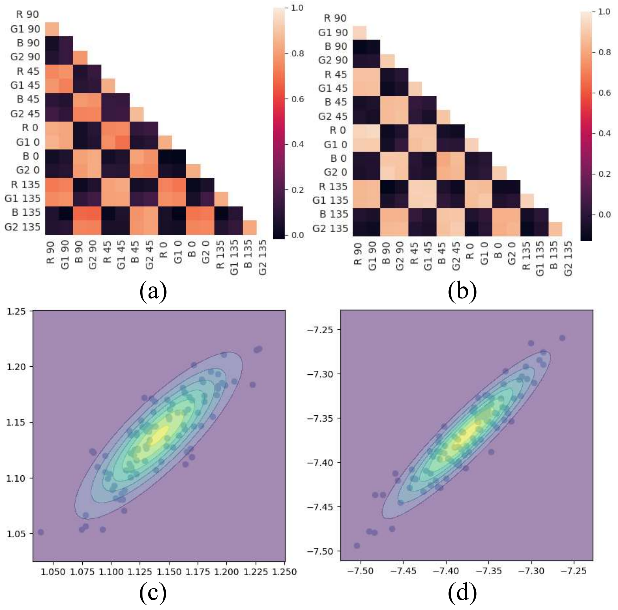


Figure 4. Visualization for parameters estimated at the gain=24 setting (a) and (b) are heatmaps of correlation coefficients of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$. (c) and (d) are scatter plots and fitted multivariate Gaussian distributions of channel R^{90} and G_1^{90} on their $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$.

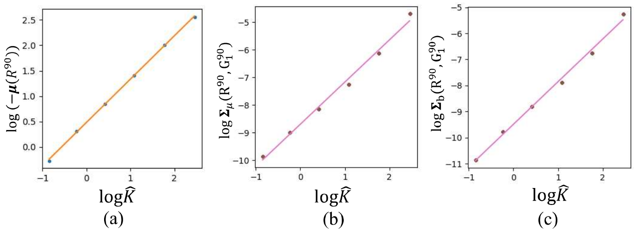
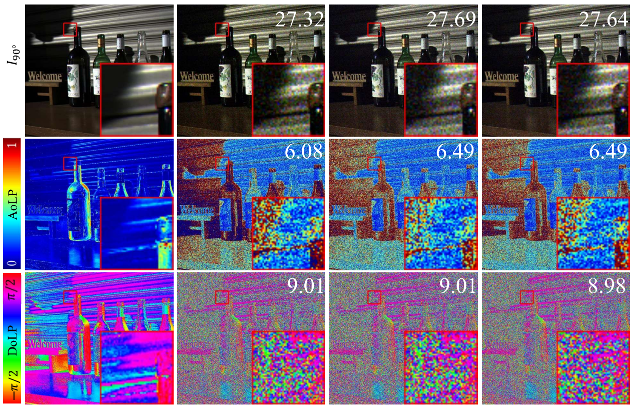


Figure 5. Linear regression results for joint distributions of log-scaled $\left(\hat{K}, \boldsymbol{\mu}\left(R^{90}\right)\right)$, $\left(\hat{K}, \boldsymbol{\Sigma}_\mu\left(R^{90}, G_1^{90}\right)\right)$, and $\left(\hat{K}, \boldsymbol{\Sigma}_b\left(R^{90}, G_1^{90}\right)\right)$.

parameters ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}_b$, $\boldsymbol{\lambda}$, and $\boldsymbol{\sigma}_\tau$) for the channels under an arbitrary camera gain $\boldsymbol{K}$. As all the parameter samples follow a Gaussian distribution, it is an intuitive way to sample them independently for each channel or control them with the same Gaussian sampler. However, our further observations unveil the submerged correlations.

Given calibrated parameters of 100 frames under a gain level, e.g., 24, we employ Pearson correlation coefficients (PCCs) to examine the change tendency of a parameter from different channels. Fig. 4 exhibits the heatmaps of PCCs of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_r$. It is obvious that without considering the polarization channels, specific correlations exist for $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$ samples of the color channels, that R and G_1 channel, B and G_2 channel are highly positively correlated, while the other combinations are nearly uncorrelated. Considering the arrangement of pixels, we can also speculate the im-



Clean image Real noisy Debiased real noisy Debiased synthetic

Figure 6. Visual comparison for image, DoLP and AoLP of a clean image, real noisy image, debiased real noisy image, and our debiased synthetic image. The PSNR to the clean image is shown, which proves that debiased inputs are closer to the reference.

perceptible mechanism that the pixels of two rows are processed by the readout circuit simultaneously as a group. In contrast, the correlation can not be observed for $\boldsymbol{\sigma}_\tau$ and $\boldsymbol{\lambda}$, that all channels are uncorrelated.

From the observation, individual Gaussian models are not sufficient to simulate real channel biases $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$ as the correlation exists. Thus multivariate Gaussian distribution (MGD) is applied to fully formulate the parameter sampling. We estimate 16×16 covariance matrices $\boldsymbol{\Sigma}_\mu$ and $\boldsymbol{\Sigma}_b$ for the parameter $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$ samples of 16 channels, and compute the mean value of the parameter samples along the sample dimension as mean of the MGD. While $\log \boldsymbol{K}$ is uniformly sampled for 16 channels, to reduce the influence of possible calibration error of $\boldsymbol{K}$, the mean of $\boldsymbol{K}$, termed as $\hat{K}$ is used for regression and sampling. Thus the MGD of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$ can be formulated as:

$$\boldsymbol{\mu} \sim \mathcal{N}_{16}\left(\bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_\mu\right), \quad \boldsymbol{\sigma}_b \sim \mathcal{N}_{16}\left(\bar{\boldsymbol{\sigma}}_b, \boldsymbol{\Sigma}_b\right), \tag{5}$$

where $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\sigma}}_b$ are mean value of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$ samples. Fig. 4 illustrates the scatter plot and estimated MGD of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}_b$ samples of R^{90} and G_1^{90}, that shows MGD is proper to model the correlation and simulate real noise level sampling. $\boldsymbol{\sigma}_\tau$ can still be sampled individually for 16 channels via:

$$\boldsymbol{\sigma}_\tau \sim \mathcal{N}\left(\bar{\boldsymbol{\sigma}}_\tau, \hat{\boldsymbol{\sigma}}_\tau\right), \tag{6}$$

where $\bar{\boldsymbol{\sigma}}_\tau$ and $\hat{\boldsymbol{\sigma}}_\tau$ are mean and standard deviation of $\boldsymbol{\sigma}_\tau$ samples.

At last, the covariance of highly correlated channels as well as their variance is observed to increase with lifted gains, so a log-scaled linear model can be used to formulate the variation. With $\mathbf{x} = [-\bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}_b, \boldsymbol{\Sigma}_b, \bar{\boldsymbol{\sigma}}_\tau, \hat{\boldsymbol{\sigma}}_\tau]$, we have:

$$\log(\mathbf{x}) = \mathbf{a}\log(\hat{K}) + \mathbf{b}, \tag{7}$$

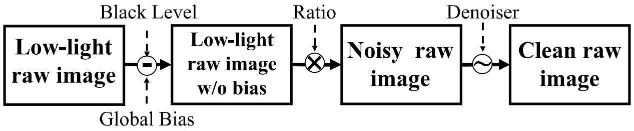


Figure 7. The pipeline of our learning-based polarized color image denoising method.

where $\mathbf{a}$ and $\mathbf{b}$ denote slopes and intercepts of fitted lines.

4. Dataset

To enable the research on polarized color image denoising, we collect a dataset under outdoor low-light environments. Following the image capture strategy of the previous benchmark [45], 824 low-light images and their corresponding references are captured in total. 4 gain settings (6, 12, 18, and 24) and 2 short-exposure ratios (10, 60) are selected for a wider coverage. The conversion of the camera gain and overall system gain, and further detailed capturing protocol can be seen in our supplementary materials. During captures, the camera is mounted on a steady tripod. The exposure time for capturing reference images is adjusted to collect sufficient photons. Moreover, as the SNR of polarization information is extremely sensitive to noises [13], 50 to 100 long-exposure images are averaged to generate a single clean reference image. Also, 3-5 continuous frames are captured for each low-light setting and a noisy image which have the closest intensity to the reference is selected to avoid the flickering effect of manmade AC lamps at night.

5. Method

5.1. Pipeline

To synthesize a noisy image based on our model with a clean image, we first divide it by a predetermined low-light ratio, e.g., 10, to simulate a short-exposure capturing. Then, the noise parameters are sampled from a continuous model and Gaussian samplings, via a random $\log \hat{K}$ and above equations 5-7. Using the parameters, a realistic noise map is generated through equations 1-4, and ultimately added to the scaled clean image.

Given a low-light noisy raw polarization color image, following the tradition of learning-based raw image processing pipeline [4, 9, 45], we firstly pack it into 16 individual channels. Then, the input should have been subtracted by a predetermined black level before amplification using the corresponding short-exposure ratio. However, a significant minus global bias is observed under a lifted camera gain setting, which may lead to noticeable brightness reduction, as shown in Fig. 6. Moreover, as a global bias barely affects S_1 and S_2 as it is almost eliminated, it will accumulate in S_0 and finally destroy the DoLP distribution, as shown in Fig. 6. Therefore, for both noisy signals and

Table 1. Polarized color image denoising performance in PSNR(dB)/SSIM calculated on 16-channel images, 4-channel DoLP and 4-channel AoLP. **Bold** values present the best results. ”*” represent the model is trained on our synthetic noises.

		×10				×60			
Training Data	Method	image		DoLP	AoLP	image		DoLP	AoLP
		PSNR	SSIM	PSNR	PSNR	PSNR	SSIM	PSNR	PSNR
	BM3D	26.08	0.901	21.62	8.20	21.48	0.736	19.17	7.14
	MBM3D	25.82	0.878	22.79	13.38	21.19	0.704	19.73	11.79
Real	U-Net	33.25	0.829	19.32	13.66	26.47	0.546	20.13	13.02
	Uformer	38.28	0.968	25.61	15.84	33.88	0.922	23.47	14.89
	Restormer	39.06	0.969	26.48	15.90	34.29	0.923	**23.80**	14.98
	Ours	39.20	0.971	26.57	16.04	**34.42**	**0.927**	23.56	14.89
Synthetic	U-Net*	35.77	0.906	23.64	14.78	27.40	0.611	21.06	13.47
	Uformer*	39.26	0.969	26.14	15.94	33.50	0.900	23.13	14.17
	Restormer*	39.82	0.971	26.65	16.05	33.64	0.918	23.44	14.88
	Ours*	**40.21**	**0.973**	**27.15**	**16.25**	33.93	0.915	23.65	**15.07**

polarization information, the global bias makes the restoration task even more ambiguous and further undermines the denoising performance. Thus, a noisy image is subtracted by the black level and the global bias before amplification.

For a real-world noisy image, the global bias comes from the average of collected bias frames under the corresponding gain. For a synthetic noisy image, its sample location of channel-wise biases $\boldsymbol{\mu}$ are shifted from $\bar{\boldsymbol{\mu}}$ to $(\bar{\boldsymbol{\mu}} - \bar{\mu})$, where $\bar{\mu}$ presents the mean of $\bar{\boldsymbol{\mu}}$, to retain the variation of channel-wise biases, which are difficult to fully estimate and are expected to be fixed by our learning-based approach. Fig. 6 shows that the intensity and DoLP of the noisy image after the offset is closer to the clean image.

5.2. Transformer network

For better polarization color image restoration performance, a Transformer-based neural network is proposed. A 4-stage U-shaped architecture is applied as our backbone, which features multi-resolution processing and low computation consumption. In each stage, successive Transformer blocks are employed, which process deep features. While Transformer-based image restoration algorithms mainly depend on long-term dependencies of spatial-domain [44] or channel-domain [6, 49] self-attention mechanisms, we consider their combination boost both signal denoising and polarization restoration. Thus, Shifted Window based Multi-head Attention (SW-MA) [33], Window based Multi-Shuffled-heads Transposed Attention (W-MSTA) and local-enhanced MLP [30, 44] are used to construct a Transformer block. Please refer to our supplementary material for a more detailed introduction and illustration.

Table 2. Polarized color image denoising performance in PSNR(dB)/SSIM calculated on images, DoLP and AoLP. **Bold** values present the best best results. N_{read} and N^*_{read} represent read out noises sampled with zero-mean and channel-wise biases respectively.

Data	image	DoLP	AoLP
	PSNR/ SSIM	PSNR	PSNR
(a) paired w/ bias	33.01/0.922	23.23	14.75
(b) paired	**34.42/0.927**	23.56	14.89
(c) N_{read}	31.56 /0.849	22.79	13.12
(d) $N_{read} + N_p$	33.81/ 0.907	23.64	14.96
(e) $N_{read} + N_p + N_b + N_q$	33.94/0.914	23.69	15.04
(f) $N^*_{read} + N_p + N_b + N_q$	33.93/0.915	23.65	**15.07**
(g) w/o $\mathcal{L}_S$	33.93/ 0.914	**23.74**	14.92

5.3. Loss Function

Given a noisy input $\mathtt{x}$ and its clean counterpart $\mathtt{y}$, a pixel-wise loss function is generally used to optimize a denoiser network. Here, loss function $\mathcal{L}_1$ is formulated as follows:

$$\mathcal{L}_1 = \parallel \mathtt{y} - \mathcal{F}(\mathtt{x}) \parallel_1, \tag{8}$$

where $\mathcal{F}(\cdot)$ represents our network. The pixel-wise loss function aims to directly minimize the differences of pixels. Moreover, inspired by [46], we also minimize the Stoke parameter loss $\mathcal{L}_S$ of each color channel, which is formulated as follows:

$$\mathcal{L}_S = \frac{1}{4} \sum_{i=0,1,2,3} \parallel \mathcal{S}(\mathtt{y_i}) - \mathcal{S}(\mathcal{F}(\mathtt{x_i})) \parallel_1 \tag{9}$$

where $\mathcal{S}(\cdot)$ denotes computation of Stoke parameters, and $\mathtt{i}$ represents the number of four color channels. The intermediate variables provide a strong yet efficient constraint on polarization information.

6. Experiments

We implement a learning-based pipeline for polarization color image denoising. The training procedure continues up to 3000 epochs and the network is optimized by Adam [25] optimizer with batch size 16. While the initial learning rate is set as $2e-4$, warm-up and cosine annealing

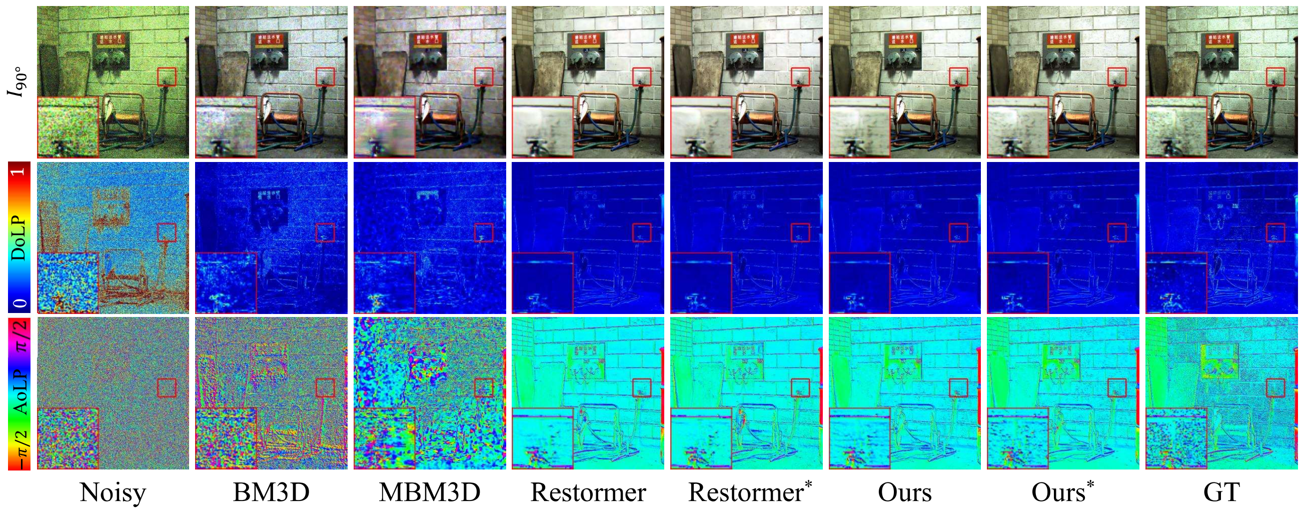


Figure 8. Visual comparison for polarized color image denoising. I_{90°, DoLP and AoLP are exhibited, and "*" represents the model is trained on synthetic noisy images generated via our noise model. Please refer to more results in our supplementary material.

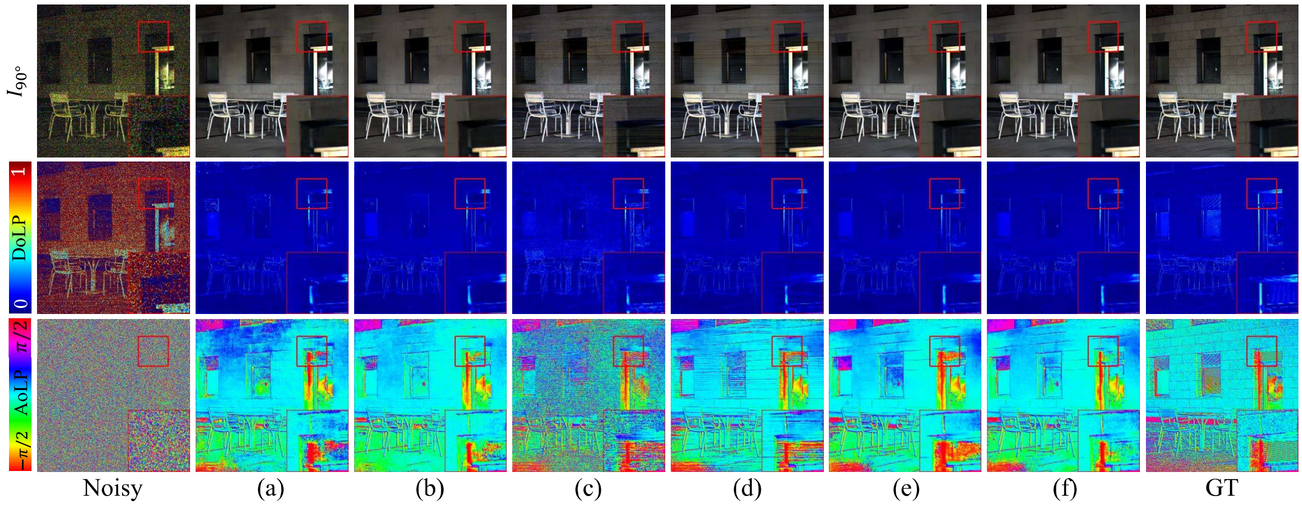


Figure 9. Visual comparison for ablation study. I_{90°, DoLP and AoLP are visualized and the labels represent settings presented in Table 2.

strategies are applied for better regression. In each epoch, a 128 × 128 × 16-sized patch is randomly cropped from a packed input. Random rotation and flipping are applied for data augmentation. We compare our approach with classic non-deep single-image denoising methods, and further train SOTA neural networks on our real-world paired noisy images and synthetic noisy images to validate the accuracy of our estimated noise properties as well as the performance of our proposed Transformer model.

6.1. Comparison

Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index Measure (SSIM) are used for quantitative evaluations of restored images, and PSNR is applied to evaluate DoLP and AoLP restoration. We compare existing algorithms, including BM3D [37], U-Net [40], Uformer [44] and Restormer [49] on data of two low-light ratios ×10 and ×60, respectively. Specifically, BM3D [37] is able to handle multi-channel inputs, so we apply it on 4-channel (BM3D, taking each color channel containing 4 polarization channels as a grayscale input) and 16-channel packed pattern (MBM3D). Given an estimated noise level [10] for each input image, (M)BM3D approach is used directly on a real-world test set. Neural networks are trained on real noisy and our synthetic noisy images and are evaluated on real noises.

Table 1 shows quantitative results of polarized color image denoising. It is obvious that non-deep approaches are far from being able to handle such heavy noises as well as polarization distortion. Interestingly, BM3D outperforms MBM3D on image denoising, but fails to restore polarization information. The reason is that packing different po-

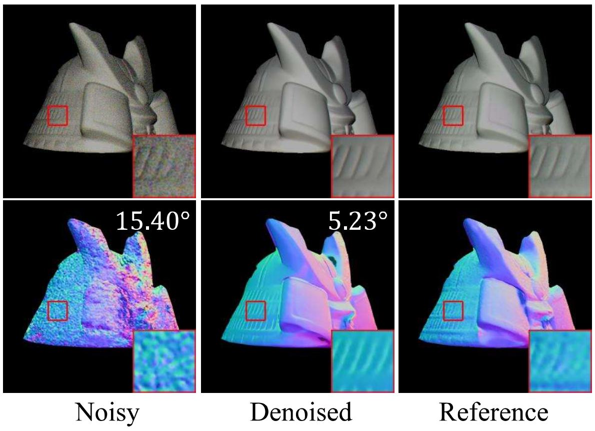


Figure 10. Visual results of SfP [56] with MAE of normalized accumulated angle to the reference.

larization pixels in the same patch without considering their variances do harm polarization restoration. Focusing on the comparison of training sets under ×10 low-light scenes, our synthetic noisy training set significantly improves performance of every learning-based models. We give credits to the infinite noise patterns in simulation. Under a shorter exposure setting, ×60, a small deviation of noise model could be extremely enlarged, and makes the synthetic noisy image's polarization distribution distorted. The results show our synthetic noisy training set still have comparative performance. Regarding the comparison of different neural network denoising models, our network performs clearly better on both restoring clean images and precise polarization information in most scenarios.

The visual comparison is shown in Fig. 8. The BM3D approaches fail to address such heavy noises and barely restore polarization, while Restormer and our model can effectively eliminate noises. Restormer* and Ours*, trained on synthetic noisy images, restore more vivid details. Moreover, due to the window-based hybrid attention mechanism, our denoising model is able to remove noises and restore sharp details for both images and polarization information.

6.2. Ablation Study

All ablation experiments are conducted based on our proposed Transformer model with noisy images at ×60 low-light ratio. See more ablation analysis in our supplementary material.

Global bias. In our pipeline, a global bias is observed under lifted camera gains and subtracted from input before it is amplified. Table 2(a), (b) show a dramatic performance degradation without this pretreatment. Moreover, Fig. 9 (a), (b) show that by subtracting the global bias, more vivid image and polarization details can be restored.

Ablation on noise models. To verify the effectiveness of the introduced noise model, we conduct ablation experiments on synthetic training set of degenerated noise models. As shown in the Table 2, a full set of noise components deliver significant and stable performance improvements. Visual comparison is shown in Fig. 9 (c)-(f). The visual results of (c) and (d) indicate the photon shot noise plays a great role as a major noise source. However, the noticeable banding stripes in (c), (d), especially for AoLP, are eliminated until banding pattern noise is considered. At last, the network trained on our final noise model shows sharp yet rich details in images and polarization, and far outperforms the real noisy training set, as shown in Fig. 9 (b) and (f).

Stoke parameter loss. A loss function combining pixelwise differences and polarization information errors are applied to optimize the network. (f) and (g) in Table 2 show the performance of the model trained with and without $\mathcal{L}_S$. The comparison demonstrates that $\mathcal{L}_S$ can help restore precise polarization information.

6.3. Polarization Application

Here, we show our denoising approach can be beneficial to downstream polarization-based algorithms, e.g., shape-from-polarization. We feed a noisy and denoised polarization image into a 3D shape reconstruction network [16]. As illustrated in Fig. 10, compared with the noisy input, the 3D normal reconstructed with our denoising process contains a more detailed structure and accurate shape. Quantitative results on MAE of normalized accumulated angle to reference normal map are shown as well.

7. Conclusion

In this paper, we address the low SNR issue of polarized color sensors. We propose a learning-based pipeline to simultaneously restore clean signals and polarization information. A real-world polarized color image dataset of paired raw short-exposure noisy images and long-exposure reference images is captured to support the learning-based pipeline. Furthermore, we systematically analyze the noise sources in the physical imaging pipeline and try to generate realistic noisy images for our learning-based approach. Spatial and channel domain self-attention mechanisms are applied to construct a Transformer model for better denoising performance. Experimental results validate the effectiveness of proposed noise model as well as Transformer-based denoising model. Extensive ablation studies justify our contributions. We also demonstrate that our denoising process benefits downstream polarization applications.

Acknowledgement

This research was supported in part by JSPS KAKENHI Grant Numbers 22H00529, 20H05951, and ROIS NII Open Collaborative Research 2023-23S1201.

References

[1] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *ICCV*, pages 3165–3173, 2019. 2

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, pages 1692–1700, 2018. 2

[3] Abubakar Abubakar, Xiaojin Zhao, Shiting Li, Maen Takruri, Eesa Bastaki, and Amine Bermak. A block-matching and 3-d filtering algorithm for gaussian noise in dofp polarization images. *IEEE Sensors Journal*, 18(18):7429–7435, 2018. 2

[4] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *CVPR*, pages 11036–11045, 2019. 2, 5

[5] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *CVPR*, pages 2392–2399. IEEE, 2012. 1, 2

[6] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *CVPR*, pages 17502–17511, 2022. 2, 6

[7] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021. 2

[8] Ke-Chi Chang, Ren Wang, Hung-Jin Lin, Yu-Lun Liu, Chia-Ping Chen, Yu-Lin Chang, and Hwann-Tzong Chen. Learning camera-aware noise models. In *ECCV*, pages 343–358. Springer, 2020. 2

[9] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, pages 3291–3300, 2018. 2, 5

[10] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *ICCV*, pages 477–485, 2015. 7

[11] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *CVPR*, pages 12299–12310, 2021. 2

[12] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE TPAMI*, 39(6):1256–1272, 2016. 1, 2

[13] Yingkai Chen, Zhongmin Zhu, Zuodong Liang, Leanne E Iannucci, Spencer P Lake, and Viktor Gruev. Analysis of signal-to-noise ratio of angle of polarization and degree of polarization. *OSA Continuum*, 4(5):1461–1472, 2021. 5

[14] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz. Polarimetric multi-view stereo. In *CVPR*, pages 1558–1567, 2017. 2

[15] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE TIP*, 16(8):2080–2095, 2007. 2, 3

[16] Valentin Deschaintre, Yiming Lin, and Abhijeet Ghosh. Deep polarization imaging for 3d shape and svbrdf acquisition. In *CVPR*, pages 15567–15576, 2021. 1, 2, 8

[17] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 2

[18] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE TIP*, 15(12):3736–3745, 2006. 2

[19] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE TIP*, 17(10):1737–1754, 2008. 2

[20] Yoshiki Fukao, Ryo Kawahara, Shohei Nobuhara, and Ko Nishino. Polarimetric normal stereo. In *CVPR*, pages 682–690, 2021. 1, 2

[21] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising. In *ECCV*, pages 538–554, 2018. 1

[22] Ryan D Gow, David Renshaw, Keith Findlater, Lindsay Grant, Stuart J McLeod, John Hart, and Robert L Nicol. A comprehensive tool for modeling cmos image-sensor-noise performance. *IEEE Transactions on Electron Devices*, 54(6):1321–1329, 2007. 3

[23] James Janesick, Kenneth Klaasen, and Tom Elliott. Ccd charge collection efficiency and the photon transfer technique. In *Solid-state imaging arrays*, volume 570, pages 7–19. SPIE, 1985. 4

[24] Brian L Joiner and Joan R Rosenblatt. Some properties of the range in samples from tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66(334):394–399, 1971. 3

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[26] LUCID VISION LABS. Phoenix 5.0 mp polarization model. https://thinklucid.com/product/phoenix-5-0-mp-polarized-model/. 3, 4

[27] Chenyang Lei, Xuhua Huang, Mengdi Zhang, Qiong Yan, Wenxiu Sun, and Qifeng Chen. Polarized reflection removal with perfect alignment in the wild. In *CVPR*, pages 1750–1758, 2020. 1, 2

[28] Cedric Leyris, Alain Hoffmann, Matteo Valenza, J-C Vildeuil, and F Roy. Trap competition inducing rts noise in saturation range in n-mosfets. In *Noise in Devices and Circuits III*, volume 5844, pages 41–51. SPIE, 2005. 3

[29] Xiaobo Li, Haiyu Li, Yang Lin, Jianhua Guo, Jingyu Yang, Huanjing Yue, Kun Li, Chuan Li, Zhenzhou Cheng, Haofeng Hu, et al. Learning-based denoising for polarimetric images. *Optics express*, 28(11):16309–16321, 2020. 2

[30] Yawei Li, Kai Zhang, Jiezhang Cao, Radu Timofte, and Luc Van Gool. Localvit: Bringing locality to vision transformers. *arXiv preprint arXiv:2104.05707*, 2021. 6

[31] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCVW*, 2021. 2

[32] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *CVPR*, pages 5802–5811, 2022. 2

[33] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *ICCV*, 2021. 2, 6

[34] Ziwei Liu, Lu Yuan, Xiaoou Tang, Matt Uyttendaele, and Jian Sun. Fast burst images denoising. *ACM TOG*, 33(6):1–9, 2014. 1

[35] Youwei Lyu, Zhaopeng Cui, Si Li, Marc Pollefeys, and Boxin Shi. Reflection separation using a pair of unpolarized and polarized images. *NeurIPS*, 32:14559–14569, 2019. 1, 2

[36] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *ICCV*, pages 2272–2279. IEEE, 2009. 2

[37] Ymir Mäkinen, Lucio Azzari, and Alessandro Foi. Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching. *IEEE TIP*, 29:8339–8354, 2020. 7

[38] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *CVPR*, pages 1586–1595, 2017. 2

[39] Ajit Rajwade, Anand Rangarajan, and Arunava Banerjee. Image denoising using the higher order singular value decomposition. *IEEE TPAMI*, 35(4):849–862, 2012. 2

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2, 7

[41] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 2

[42] Yoav Y Schechner, Joseph Shamir, and Nahum Kiryati. Polarization-based decorrelation of transparent layers: The inclination angle of an invisible surface. In *ICCV*, volume 2, pages 814–819. IEEE, 1999. 2

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pages 5998–6008, 2017. 2

[44] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *arXiv preprint arXiv:2106.03106*, 2021. 2, 6, 7

[45] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *IEEE TPAMI*, 2021. 2, 3, 4, 5

[46] 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. 6

[47] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *CVPR*, pages 11844–11853, 2020. 1

[48] Luwei Yang, Feitong Tan, Ao Li, Zhaopeng Cui, Yasutaka Furukawa, and Ping Tan. Polarimetric dense monocular slam. In *CVPR*, pages 3857–3866, 2018. 2

[49] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *arXiv preprint arXiv:2111.09881*, 2021. 2, 6, 7

[50] Junchao Zhang, Haibo Luo, Rongguang Liang, Wei Zhou, Bin Hui, and Zheng Chang. Pca-based denoising method for division of focal plane polarimeters. *Optics express*, 25(3):2391–2400, 2017. 2

[51] Junchao Zhang, Jianbo Shao, Haibo Luo, Xiangyue Zhang, Bin Hui, Zheng Chang, and Rongguang Liang. Learning a convolutional demosaicing network for microgrid polarimeter imagery. *Optics letters*, 43(18):4534–4537, 2018. 2

[52] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7):3142–3155, 2017. 1, 2, 3

[53] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *CVPR*, July 2017. 1, 2

[54] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *ICCV*, pages 4593–4601, 2021. 2

[55] Dizhong Zhu and William AP Smith. Depth from a polarisation+ rgb stereo pair. In *CVPR*, pages 7586–7595, 2019. 2

[56] Shihao Zou, Xinxin Zuo, Yiming Qian, Sen Wang, Chi Xu, Minglun Gong, and Li Cheng. 3d human shape reconstruction from a polarization image. In *ECCV*, pages 351–368. Springer, 2020. 2, 8