Physics-Guided ISO-Dependent Sensor Noise Modeling for Extreme Low-Light Photography

Yue Cao[1], Ming Liu[1](✉), Shuai Liu, Xiaotao Wang, Lei Lei, Wangmeng Zuo[1,2]
[1]School of Computer Science and Technology, Harbin Institute of Technology, China
[2]Peng Cheng Laboratory, China
cscaoyue@gmail.com, csmliu@outlook.com, 18601200232@163.com, wangxt@126.com,
figoleilei@163.com, wmzuo@hit.edu.cn

Abstract

Although deep neural networks have achieved astonishing performance in many vision tasks, existing learning-based methods are far inferior to the physical model-based solutions in extreme low-light sensor noise modeling. To tap the potential of learning-based sensor noise modeling, we investigate the noise formation in a typical imaging process and propose a novel physics-guided ISO-dependent sensor noise modeling approach. Specifically, we build a normalizing flow-based framework to represent the complex noise characteristics of CMOS camera sensors. Each component of the noise model is dedicated to a particular kind of noise under the guidance of physical models. Moreover, we take into consideration of the ISO dependence in the noise model, which is not completely considered by the existing learning-based methods. For training the proposed noise model, a new dataset is further collected with paired noisy-clean images, as well as flat-field and bias frames covering a wide range of ISO settings. Compared to existing methods, the proposed noise model is equipped with a flexible structure and accurate modeling capabilities, which is beneficial for better denoising performance in extreme low-light scenes. The dataset and code are available at https://github.com/happycaoyue/LLD.

1. Introduction

In recent years, learning-based image denoising methods have achieved tremendous success with pairwise training samples [22, 32]. However, it is still challenging to recover high-quality results in extreme low-light scenarios, mainly due to limited data [9, 31]. Considering the difficulty of collecting enormous pairwise training data, noise modeling [25, 31, 33] becomes an alternative solution by simulating noises that match the extreme low-light distribution.

The noises in extreme low-light scenarios contain severe striping artifacts and color bias, and the damage to image quality is inconsistent across different ISO settings and locations. To model such complicated noises, physics-based methods [9, 29, 31, 33] build cumbersome statistical models according to the physical process from photons (*i.e.*, the light) to digital signals (*i.e.*, the rawRGB image). Nevertheless, the noise parameter calibration relies on a large number of flat-field and bias frames, which is also laborious and expensive. For example, PMN [9] requires 400 bias frames at each ISO setting to calibrate the noise parameters.

For circumventing the tedious parameter calibration process, learning-based methods [1, 5, 25] directly learn the mapping from clean images to their noisy counterparts. Yet the performance is still far inferior to the physics-based statistical methods [31, 33]. To boost the performance of learning-based sensor noise modeling, we delve into such inferiority and attribute the major cause to the inconsistency between the noise models and the imaging process. For example, NoiseFlow [1] utilizes the distribution matching ability of normalizing flow but is unable to model striping artifacts and color bias. Starlight [25] leverages various noise sources like heteroscedastic Gaussian noise, row noise, and fixed-pattern noise. However, it mixes the noises with the clean image and delivers them into a GAN model. In other words, the noises are entangled with each other, which increases the difficulty of describing the noise distributions. Moreover, these methods [1,5,25] either ignore the ISO dependency of the noise or are based on the assumption of a small range of ISO settings, further limiting the performance of learning-based noise modeling methods.

As a remedy, we propose a refined noise model to tap the potential of learning-based sensor noise modeling. As shown in Tab. 1, the proposed noise model covers the most common kind of noises in the imaging process, including shot noise N_{shot}, dark current fixed-pattern noise N_{FP}, black level error noise N_{BLE}, dark current shot noise N_{DCSN}, read noise N_{read}, row noise N_{row}, and quantiza-

Table 1. Comparison between noise modeling methods. $\mathcal{S}$ means that the noise is sampled from real images.

Method	Category	N_{shot}	N_{FP}	N_{BLE}	N_{DCSN}	N_{read}	N_{row}	N_q	Learnability	ISO dependence
ELLE [29]	Physics	✔		✔	✔	✔	✔	✔	None	Incomplete
ELD [31]	Physics	✔		$\mathcal{S}$	✔	✔	✔	✔	None	Incomplete
SFRN [33]	Physics	✔	$\mathcal{S}$	$\mathcal{S}$	$\mathcal{S}$	$\mathcal{S}$	$\mathcal{S}$	✔	None	Incomplete
PMN [9]	Physics	✔	✔	✔				✔	None	Incomplete
NoiseFlow [1]	Learn	✔			✔	✔		✔	Complete	Incomplete
Startlight [25]	Learn	✔	✔		✔	✔	✔	✔	Incomplete	Incomplete
Ours	Learn	✔	✔	✔	✔	✔	✔	✔	Complete	Complete

tion noise N_q. Among them, N_{FP}, N_{BLE}, and N_{DCSN} jointly model the dark current noise N_{DC}. Besides, the ISO dependence is also better considered compared to existing methods. The noise model is implemented in the normalizing flow framework, and each component corresponds to a specific type of noise. Such configuration leverages the explicit distribution modeling ability of normalizing flow models, and the network architecture is also flexible enough to align accurately with our proposed noise model.

Apart from the inconsistency between existing noise models and the imaging process, another key factor impeding accurate noise modeling is the dataset. For obtaining the reference clean images, the SIDD dataset [2] overlays multiple noisy images, which leaves the black level error noise [9, 29, 31] and fixed-pattern noise [4, 21, 23] unable to remove. Another commonly used dataset SID [6] adopts the same ISO setting for pairwise long and short exposure images, which results in the long exposure reference images still containing noises such as fixed-pattern noise. Therefore, for better training the proposed noise model, we have also collected a low-light image denoising (LLD) dataset, which contains pairs of noisy (short exposure, high ISO) and clean (long exposure, low ISO) images. Furthermore, we also provide flat-field and bias frames at various ISO settings in the LLD dataset, hoping it can facilitate the understanding of real noise and image denoising research.

With the LLD dataset, we train the proposed noise model in a two-stage manner. Specifically, the noises are divided into two groups, *i.e.*, the fixed-pattern noise and the random noise. Then the noise model is first trained to describe the random noise and then fitted to the fixed-pattern noise. Thanks to the flexible structure and accurate modeling capabilities, the proposed noise model can better capture the characteristics of real noise. The image denoising methods can also benefit from our noise model and achieve superior performance in extreme low-light scenes.

To sum up, the main contributions of this work include:

- We investigate the noise formation process in extreme low-light scenarios and propose a novel physics-guided noise model. The ISO dependence is taken into consideration in the proposed method.
- We collect a dataset for extreme low-light image denoising. The dataset contains pairwise noisy-clean images captured by two cameras (*i.e.*, Sony A7S2 and Nikon D850). We also provide flat-field and bias frames covering a wide range of ISO settings.
- While the learning-based nature eliminates the labor-intensive parameter hand-calibration process, our proposed method can achieve superior noise modeling accuracy and boost the image denoising performance.

2. Related Work

2.1. Physics-based Statistical Noise Modeling

Physics-based statistical noise modeling approaches generally divide the noises into signal-dependent and signal-independent components. Shot noise is the dominant source of signal-dependent noises, which is generally modeled as a Poisson distribution and is determined by the signal and the camera gain. To obtain the camera gain, some methods [9, 31, 33] capture flat-filed frames when the sensor is uniformly illuminated, and derive the system gain via the photon transfer method [15]. Similarly, PMRID [30] utilizes a burst of grayscale chart images to calibrate the system gain by means of the photon transfer method.

Signal-independent noises typically are composed of dark current noise, read noise, row noise, and quantization noise. Generally, the parameters of signal-independent noises can be obtained from bias frames captured in a light-free environment. To model the read noise and dark current noise, ELD [31] employs the Tukey lambda distribution [17], which has a longer tail and can mitigate the chrominance artifacts in low-light conditions. By contrast, PMN [9] only models dark current fixed-pattern noise. Besides dark current fixed pattern noise, the thermal effects of the circuit also lead to black level error noise, which is modeled by a uniform distribution [29], randomly sampling from real data [31], or averaging hundreds of bias frames [9]. Row noise is typically modeled by a vector following Gaussian distribution [12, 29, 31], while quantization noise mathematically follows a uniform distribution.

Although current physics-based noise modeling methods have achieved decent performance, they generally require laborious calibration and still cannot model the color bias in extreme low-light scenarios, which motivates us to develop more accurate and flexible methods for noise modeling.

2.2. Learning-based Noise Modeling

Recently, several learning-based noise models [1,5,7,16, 18, 24, 25] have adopted normalizing flow-based [20, 28] or GAN-based [11, 13] frameworks to model real-world noises in a data-driven manner. For example, NoiseFlow [1] applies a flow-based model that maximizes the likelihood of sampled noise. CA-GAN [5] feeds the clean image and a random Poisson-Gaussian noise into a noise generator to produce the final synthetic noise. The above pioneering works are based on normal light or ordinary low-light scenes ($\sim 10^3$ photons per pixel or > 1 lux). However, due to the ignorance of the complex physical noise properties, existing learning-based methods are still far inferior to physics-based statistical methods, especially in extreme low-light scenarios [31, 33]. Starlight [25] adopts a more demanding setting under starlight ($\sim$1 photon per pixel or ≤ 0.001 lux), yet the performance is still limited and the fixed-pattern noise needs to be hand-calibrated. To this end, there is still a lack of learning-based noise models to simulate extreme low-light noise, which can give full play to the potential of neural networks while circumventing specific scenario captures like no lighting [9, 33] or flat-field [31] to obtain certain parameters in physics-based methods.

2.3. Image Denoising Dataset

To obtain sufficient paired noisy-clean data, SIDD [2] employs a straightforward approach to collect burst images and then average them as ground-truth images. In addition to being sensitive to misalignment problems, such a method is unable to remove the fixed-pattern noise and black level error noise. Chen *et al*. [6] have also collected a dataset termed SID, consisting of short-exposure noisy images and the corresponding clean images exposed for a long time. However, with the same ISO settings for both short- and long-exposure images, fixed-pattern dark current noise is still observed in the long-exposure image when a high ISO is adopted. To sum up, there are still some challenges inextricable with existing rawRGB image denoising datasets.

Some commonly used benchmarks (*e.g*., DND [27] and ELD [31]) adopt high-ISO short-exposure settings for noisy images and low-ISO long-exposure settings to obtain the ground-truth clean images. Such a configuration is more reasonable, and we follow them to collect our low-light image denoising (LLD) dataset since they are reserved for testing only. Besides, to the best of our knowledge, our LLD dataset is also the first public dataset with flat-field and bias frames available for various ISO settings.

3. Proposed Method

In this section, we first discuss the noise formation process and present our noise model, which points out the key noise sources in extreme low-light environments. Then, the noise model is implemented in a normalizing flow framework under the guidance of the physical process. Finally, we give the details of noisy image synthesis and the proposed low-light image denoising (LLD) dataset.

3.1. Physics-guided Noise Model

There are multiple stages in the rawRGB image acquisition process, and various noises can be induced in these operations [3]. Inspired by existing physics-based methods [9, 29, 31], we initialize the noise model as follows,

$$D = I + N_{shot} + N_{DC} + N_{read} + N_{row} + N_q, \quad (1)$$

where D and I are the observed noisy image and potential clean rawRGB image, N_{shot}, N_{DC}, N_{read}, N_{row}, and N_q represent the shot noise, dark current noise, read noise, row noise, and quantization noise, respectively. In the following, we will expand on these noises to finalize our noise model.

Shot Noise. Shot noise arises due to the uncertainty generated when photons are incident, which is an unavoidable physical phenomenon [12, 21]. Typically, the shot noise is modeled by a Poisson distribution controlled by the signal and the system gain. Following [10, 14], we approximate the Poisson distribution by a Gaussian,

$$N_{shot} \sim \mathcal{N}(0, \beta_{shot} \cdot I), \quad (2)$$

where $N_{shot} \in \mathbb{R}^{H \times W}$ and β_{shot} is relevant to the ISO.

Dark Current Noise. For converting the light into digital signals, the camera leverages the photoelectric effect to generate current, whose value represents the light intensity. However, apart from the current yielded by the photons, the thermal effect in the sensor will also cause a current named dark current, which results in the dark current noise [8, 21]. Ideally, we can subtract the dark current generated by the thermal effect, but the randomness of the thermal effect, the ISO dependence, and the spatial non-uniformity also need to be considered [8, 9, 21]. Therefore, we model the dark current noise with three components according to the physical properties,

$$N_{DC} = k \cdot N_{FP} + N_{BLE} + N_{DCSN}, \quad (3)$$

where k and N_{BLE} are two numbers linearly and non-linearly related to the ISO configuration respectively, and N_{BLE} is called black level error noise. $N_{FP} \in \mathbb{R}^{H \times W}$ denotes the dark current fixed-pattern noise, which is a fixed non-uniform variable. $N_{DCSN} \sim \mathcal{N}(0, \sigma^2_{DCSN})$ is the dark current shot noise representing the randomness of the thermal effect.

Read Noise and Row Noise. In the readout procedure of the sensors, the uncertainty of the electronic readout will cause inevitable read noise, which is a device-dependent noise [12]. Besides, in CMOS sensors, each column shares one analog-to-digital converter (ADC), *i.e*., the readout pro-

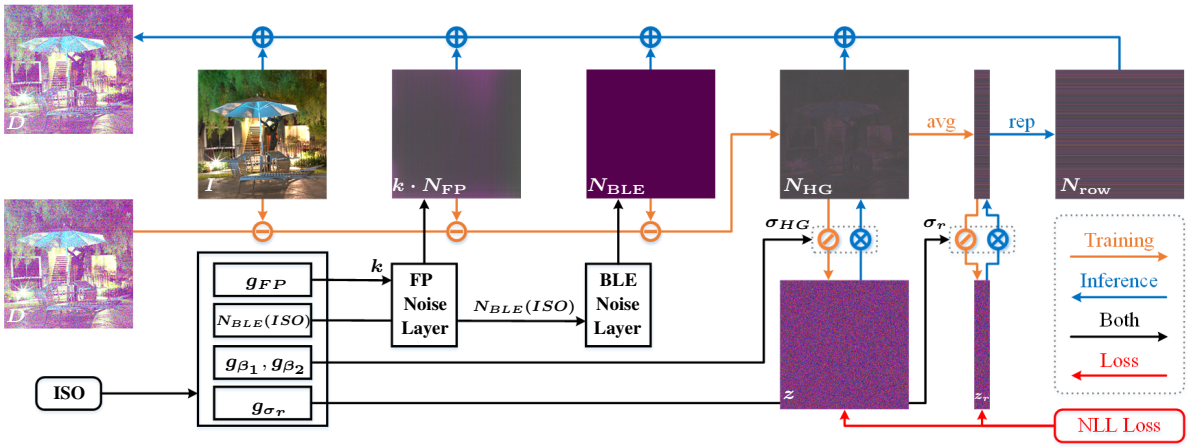

Figure 1. The pipeline of the proposed physics-guided ISO-dependent sensor noise model. Note that $\boldsymbol{z_r} \in \mathbb{R}^{H\times 1}$.

cess is performed in a row-by-row manner. The randomness of ADCs between different rows will lead to row noise, which contributes significantly to the strip artifacts in the horizontal direction. Following previous works [12, 29], we model these two noises with the following zero-mean Gaussian distribution, *i.e.*,

$$N_{read} \sim \mathcal{N}(0, \sigma_{read}^2), N_{row} \sim \mathcal{N}(0, \sigma_{row}^2), \tag{4}$$

where $N_{read} \in \mathbb{R}^{H\times W}$ and $N_{row} \in \mathbb{R}^{H\times 1}$.

Quantization Noise. Considering that the ADC has a finite accuracy, the signals should be quantified before being stored as the rawRGB image. The probability distribution of the quantization noise is usually considered to be uniform, *i.e.*,

$$N_q \sim U(-1/2q, 1/2q), \tag{5}$$

where q denotes the quantification step. Since the bit number of the rawRGB digital image is known, q does not need to be estimated.

Heteroscedastic Gaussian Noise. Considering that the shot noise N_{shot}, dark current shot noise N_{DCSN}, and read noise N_{read} are all Gaussian distributions with the size of $H \times W$, we introduce a new term N_{HG} following a heteroscedastic Gaussian distribution [10] to model them, *i.e.*,

$$N_{HG} \sim \mathcal{N}\left(0, \beta_1 \cdot I + \beta_2\right), \tag{6}$$

where β_1 is the β_{shot} of shot noise, and β_2 is determined by the dark current shot noise and read noise.

To sum up, the final noise model can be formulated as,

$$D = I + N_{HG} + k \cdot N_{FP} + N_{BLE} + N_{row} + N_q. \tag{7}$$

3.2. Noise Model Implementation

As shown in Fig. 1, our proposed deep noise model is implemented in a normalizing flow-based framework, and four components are deployed to model different noises. We start with a brief introduction to the normalizing flow-based framework and then introduce each component in detail.

Normalizing Flow-based Framework. Normalizing flows are composed of several differentiable and bilateral mapping functions that learn a transformation $\mathbf{z} = f(\mathbf{x}|\Theta)$ with parameter Θ. Following this scheme, the proposed normalizing flow-based noise model transforms a data sample $\mathbf{x}_i$ from a real extreme low-light noise distribution $p_{\mathcal{X}}$ to the $\mathcal{Z}$ space which follows standard Gaussian distribution. Based on the change of variables formula, the probability density function in real extreme low-light noise space can be expressed as,

$$p_{\mathcal{X}}(\mathbf{x}) = p_{\mathcal{Z}}(f(\mathbf{x}|\Theta)) \left|\det \mathbf{J} f(\mathbf{x}|\Theta)\right|, \tag{8}$$

where $\mathbf{J}f(\mathbf{x})$ denotes the Jacobian matrix of function f at $\mathbf{x}$, det denotes the determinant of the input matrix, and $|\cdot|$ denotes the absolute value. Equipped with real noise samples $\{\mathbf{x}_i\}_{i=1}^{M}$, the mapping function f can be optimized by minimizing the negative log-likelihood (NLL) function,

$$\mathcal{L}_{NLL} = -\sum_{i=1}^{M} \log p_{\mathcal{Z}}\left(f\left(\mathbf{x}_i|\Theta\right)\right) + \log\left|\det \mathbf{J} f\left(\mathbf{x}_i|\Theta\right)\right|. \tag{9}$$

When the training is finished, we can sample a random noise $\mathbf{z}$ from the Gaussian distribution $p_{\mathcal{Z}}$. Then the inverse transformation operation f^{-1} can synthesize the real noise by $\mathbf{x} = f^{-1}(\mathbf{z})$. In the remaining part, we will introduce the four components in detail, which are designed following the noise model presented in Sec. 3.1.

Dark Current Fixed-pattern Noise Layer. For modeling the dark current fixed-pattern noise that stays constant throughout all images, we define a learnable tensor $N_f \in \mathbb{R}^{H\times \mathrm{W}}$. Then a multilayer perceptron (MLP) g_{FP} is deployed to learn the linear parameter k in Eq. (3), *i.e.*, $k = g_{FP}(ISO)$. Therefore, the dark current fixed-pattern noise layer f_{FP} is defined as,

$$f_{FP}(x) = x - g_{FP}(ISO) \odot N_f, \tag{10}$$

where $\odot$ denotes point-wise multiplication. The corresponding inverse mapping can be obtained by $f_{FP}^{-1}(x) = x + g_{FP}(ISO) \odot N_f$.

Black Level Error Noise Layer. As shown in Eq. (3), the black level error noise is non-linearly related to the ISO. Given that there are a limited number of ISO configurations, we choose to model the N_{BLE} for each ISO configuration respectively. Specifically, $N_{BLE}(ISO)$ generates the N_{BLE} for a given ISO, and the bilateral mapping in the black level error noise layer can be written as,

$$\begin{cases} f_{BLE}(x) = x - N_{BLE}(ISO) \\ f_{BLE}^{-1}(x) = x + N_{BLE}(ISO). \end{cases} \tag{11}$$

HG Noise Layer. The HG noise layer is based on the heteroscedastic Gaussian noise model in Eq. (6). To map the heteroscedastic Gaussian noise to a standard Gaussian distribution, the operation can be formulated as,

$$z = f_{HG}(x) = x/\sigma_{HG}, \quad \text{where } \sigma_{HG} = \sqrt{\beta_1 \odot I + \beta_2}, \tag{12}$$

where β_1 and β_2 indicate the noise parameter terms of the signal-dependent and signal-independent parts, respectively, which are fitted by the MLPs g_{β_1} and g_{β_2}, *i.e.*,

$$\beta_1 = g_{\beta_1}(ISO), \beta_2 = g_{\beta_2}(ISO). \tag{13}$$

The absolute value of the determinant of the Jacobian matrix at this layer needs to be calculated when optimizing the negative log-likelihood, *i.e.*,

$$\log|\det \mathbf{J} f_{HG}| = \sum_{j=1}^{H\times W} -\log(\sigma_{HG}[j]), \tag{14}$$

where $\sigma_{HG}[j]$ denotes the j-th element of σ_{HG} and $H \times W$ denotes the total dimension of the clean image I. During the test phase, the operation of the HG noise layer can be represented as,

$$f_{HG}^{-1}(z) = \sigma_{HG} \odot z, \tag{15}$$

where z is a random noise sampled from the standard Gaussian distribution.

Row Noise Layer. The row noise layer is based on a Gaussian noise model in Eq. (4), which is further applied to the network as a flow-based layer and can be expressed as,

$$z_r = f_{row}(x) = h(x)/\sigma_r, \quad \text{where } \sigma_r = g_{\sigma_r}(ISO), \tag{16}$$

where $h(\cdot)$ indicates mean operation by row and MLP g_{σ_r} is used to fit the relationship between noise parameter σ_r and the ISO. The absolute value of the determinant of the Jacobian matrix at row noise layer can be calculated as:

$$\log|\det \mathbf{J} f_{row}| = -\log \sigma_r. \tag{17}$$

During the test phase, the row noise layer can be represented as,

$$f_{row}^{-1}(z_r) = \sigma_r \odot z_r, \tag{18}$$

where z_r is also a random noise sampled from a standard Gaussian distribution.

3.3. Training Strategy

Following previous studies [21,23], noise sources can be broadly classified as either fixed-pattern (time-invariant) or random (time-variant) noises, *i.e.*,

$$D = I + \underbrace{k \cdot N_{FP}}_{\text{Fixed-pattern Noise}} + \underbrace{N_{HG} + N_{row} + N_{BLE} + N_q}_{\text{Random Noise}}. \tag{19}$$

Note that the linear dependence of fixed-pattern noise (*i.e.*, $k \cdot N_{FP}$) holds only when N_{BLE} and N_{DCSN} are removed. Therefore, to stabilize the training process, we adopt a two-stage training strategy to train the proposed noise model, *i.e.*, the random noise modeling phase and the fixed-pattern noise modeling phase.

In particular, the network architecture for the random noise modeling phase can be represented as,

$$\begin{cases} f_1(\mathbf{x}_i|\Theta_1) = f_{HG}(f_{BLE}(\mathbf{x}_i)) \\ f_2(\mathbf{x}_i|\Theta_2) = f_{\text{row}}(f_{BLE}(\mathbf{x}_i)). \end{cases} \tag{20}$$

The loss function consists of two components, for the two branches respectively, *i.e.*,

$$\begin{aligned} -\sum_{i=1}^{M} &\log p_{\mathcal{Z}}(f_1(\mathbf{x}_i|\Theta_1)) + \log|\det \mathbf{J} f_1(\mathbf{x}_i|\Theta_1)| \\ &+ \log p_{\mathcal{Z}}(f_2(\mathbf{x}_i|\Theta_2)) + \log|\det \mathbf{J} f_2(\mathbf{x}_i|\Theta_2)| \end{aligned} \tag{21}$$

Then, the parameters for the random noise modeling phase can be fixed, and the network architecture of the fixed-pattern noise modeling phase becomes,

$$f_3(\mathbf{x}_i|\Theta_3) = f_{HG}(f_{BLE}(f_{FP}(\mathbf{x}_i))), \tag{22}$$

with the following optimization loss function,

$$-\sum_{i=1}^{M} \log p_{\mathcal{Z}}(f_3(\mathbf{x}_i|\Theta_3)) + \log|\det \mathbf{J} f_3(\mathbf{x}_i|\Theta_3)|. \tag{23}$$

3.4. Noisy Image Synthesis Schemes

After training, our noise model can be adopted in two noisy image synthesis schemes, *i.e.*, *clean image-based* synthesis and *real noise-based* synthesis.

Clean Image-based Synthesis. Clean image-based synthesis is the most intuitive way to generate noisy images. When given enough clean images I, the corresponding synthetic noisy images $\widehat{D}$ can be obtained by,

$$\widehat{D} = I + f_{FP}^{-1}\left(f_{BLE}^{-1}\left(f_{row}^{-1}(\mathrm{z}_r) + f_{HG}^{-1}(\mathrm{z})\right)\right) + N_q, \tag{24}$$

where $\widehat{D}$ and I are respectively utilized as the input and the learning target of denoising networks.

Real Noise-based Synthesis. In addition to clean image-based synthesis, real noise-based synthesis strategies are attracting upsurging attention [9, 26, 33].

Darkshading Correction Strategy. According to the noise model in Sec. 3.1, for a given camera, the fixed pattern noise and black level error noise can be regarded as deterministic. Thus, PMN [9] proposes subtracting the fixed

pattern noise and black level error noise from the noisy images, in both the training and inference phases of the denoiser. Specifically, the noises are divided into darkshading and zero-mean noises, *i.e.*,

$$D = I + \underbrace{k \cdot N_{FP} + N_{BLE}}_{\text{Darkshading}} + \underbrace{N_{HG} + N_{row} + N_q}_{\text{Zero-mean Noise}}. \quad (25)$$

Then PMN utilizes the $(D - k \cdot N_{FP} - N_{BLE}(ISO), I)$ pair for training the denoiser, and the denoiser takes $D^{test} - k \cdot N_{FP} - N_{BLE}(ISO)$ as input for noisy testing image D^{test}. The zero-mean nature of the remaining noises makes the denoising task much easier, leading to better performance. Since the darkshading is also modeled by our method, we can follow PMN to synthesize such noisy-clean pairs.

Zero-mean Noise Strategy. Furthermore, all the zero-mean noises in Eq. (25) have been modeled by our noise model, thus we can also generate more synthetic data from the opposite direction, resulting in the $(I + N_{HG} + N_{row} + N_q, I)$ training pairs. It is worth noting that PMN is unable to generate such synthetic noisy images due to their incomplete noise model.

3.5. Low-light Image Denoising Dataset

In our low-light image denoising dataset (LLD), we provide noisy-clean pairs for two cameras, *i.e.*, Sony A7S2 and Nikon D850. To guarantee that the noisy-clean pairs have consistent illumination and minimal misalignment issues, we choose to collect the LLD dataset in indoor scenes, and the cameras are controlled by the official remote software. Our LLD follows similar strategies to ELD [31], *i.e.*, the noisy images are captured with high ISO and shorter exposure time, while the reference images are captured with low ISO and longer exposure time. In specific, the base ISO (*i.e.*, 100) is used to capture the reference image, and 24 other different ISO settings are adopted for obtaining the noisy images. There are 30 scenes in total, resulting in 720 rawRGB noisy-clean image pairs for each camera. More details about our LLD dataset can be found in the supplementary material. Besides, to facilitate the real noise understanding and image denoising research, the flat-field and bias frames are also provided in our LLD dataset. There are 20 flat-field frames and 400 bias frames for each ISO setting.

3.6. Discussion

In this part, we have a brief discussion about the differences between our approach and existing physics-based or learning-based approaches separately. For physics-based methods, ELD [31] and ELLE [29] consider only image-specific black level error noise and lack modeling of fixed-pattern noise. Although PMN [9] takes into account fixed-pattern noise and ISO-specific black level error noise, the lack of modeling of random signal-independent noise still limits its modeling capabilities. Inspired by the above-mentioned approach, we present a refined noise model, which guides us in designing the modeling network and training strategy. Further, the learning-based method can estimate noise parameters and their ISO dependencies more accurately than the statistical calibration process.

As for learning-based methods, the structure design in NoiseFlow [1] is relatively simple, ignoring complex low-light noise components such as row noise and fixed-pattern noise. Starlight [25] only considers a single ISO setting and ignores ISO-related noise parameters, which greatly reduces the accuracy and generalization ability of the noise model. In addition, they still require manually calibrating the fixed-pattern noise, which limits the flexibility of starlight [25]. In contrast, our approach takes into account strip artifacts and color bias without manual parameter calibration and is flexible enough to incorporate the emerging real noise-based denoising process.

4. Experiments

4.1. Implementation Details

We use 4-channel packed rawRGB images, *i.e.*, R, Gr, B, Gb, and the Adam optimizer [19] for training. g_{FP}, g_{β_1}, g_{β_2}, and g_{σ_r} are implemented by one-layer MLP. Assuming that there are v different sets of ISO settings, $N_{BLE}(ISO)$ is then achieved by a v-dimensional learnable parameter.

Random Noise Modeling Stage. We randomly crop the image patches with size $64 \times 64 \times 4$ and train the random noise model with batch size 64 for 200 epochs. The learning rate is initialized to 1×10^{-4} and decreased to 1×10^{-5} after 50 epochs.

Fixed-pattern Noise Modeling Stage. The network parameters in the random noise part (*i.e.*, g_{β_1}, g_{β_2}, g_{σ_r}, and $N_{BLE}(ISO)$) are frozen in this stage. We use the rawRGB images with original size $(H/2) \times (W/2) \times 4$ and train the fixed-pattern network parameters for 200 epochs with batch size 1. The initialized learning rate is initially set to 1×10^{-4} and then set to 1×10^{-5} after 100 epochs.

Denoising Stage. As shown in Sec. 3.4, we use our noise model in two noisy image synthesis schemes, *i.e.*, *clean image-based* and *real noise-based*, which are denoted by Ours and Ours*, respectively. For the former, all noisy images are synthesized from the clean image. For the latter, in each mini-batch, half data comes from the darkshading correction strategy (*i.e.*, $(D - k \cdot N_{FP} - N_{BLE}(ISO), I)$ pairs) and the other comes from the zero-mean noise strategy (*i.e.*, $(I + N_{HG} + N_{row} + N_q, I)$ pairs). More training details can be found in the supplementary material.

4.2. Comparison with State-of-the-Art

We compare our physics-guided ISO-dependent noise model with 3 learning-based methods, 3 physics-based

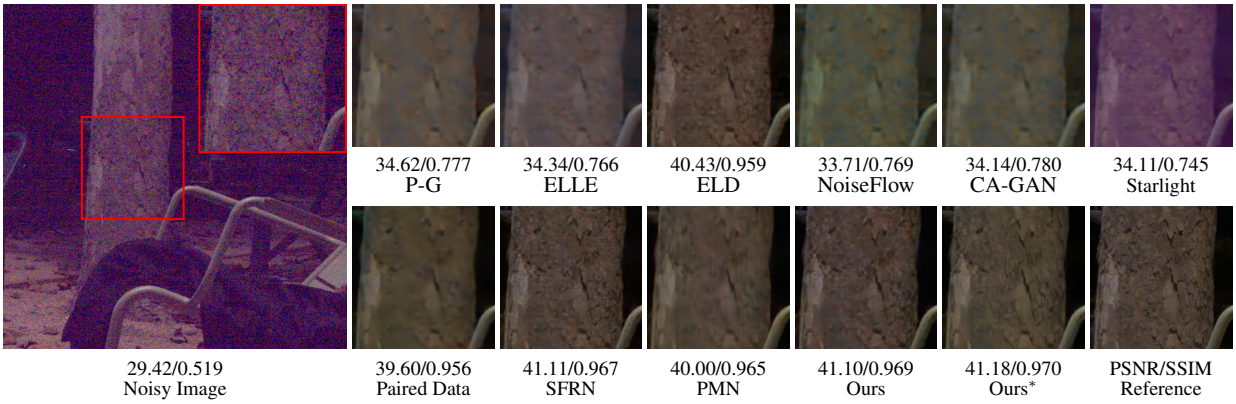

Figure 2. Denoising results of different methods on real noisy images from SID [6].

Table 2. Quantitative results (PSNR/SSIM) of different noise modeling methods on Sony A7S2 Camera (SID and ELD dataset). Red and blue are utilized to indicate the best and second best results, respectively.

Dataset	Ratio	Index	Physics-based			Learning-based				Real noise-based			
			P-G [10]	ELLE [29]	ELD [31]	NoiseFlow [1]	CA-GAN [5]	Starlight [25]	Ours	Paired Data (SID)	SFRN [33]	PMN [9]	Ours*
SID [6]	×100	PSNR	39.03	40.09	41.95	38.89	38.66	40.47	42.10	42.06	42.31	43.16	43.36
		SSIM	0.926	0.931	0.953	0.929	0.921	0.926	0.955	0.955	0.955	0.960	0.961
	×250	PSNR	35.57	36.13	39.44	35.80	35.30	36.25	39.76	39.60	40.18	40.92	41.02
		SSIM	0.861	0.863	0.931	0.867	0.846	0.858	0.933	0.938	0.937	0.947	0.948
	×300	PSNR	32.26	32.54	36.36	32.29	32.02	32.99	36.76	36.85	36.97	37.77	37.80
		SSIM	0.781	0.782	0.911	0.801	0.768	0.780	0.912	0.923	0.915	0.934	0.935
ELD [31]	×100	PSNR	41.76	43.11	45.45	41.05	41.48	43.80	45.61	44.47	45.74	46.50	46.74
		SSIM	0.930	0.940	0.975	0.925	0.933	0.936	0.977	0.968	0.976	0.985	0.986
	×200	PSNR	39.33	40.30	43.43	39.23	39.26	40.86	43.84	41.97	43.84	44.51	44.95
		SSIM	0.872	0.884	0.954	0.889	0.877	0.884	0.959	0.928	0.955	0.973	0.977

Table 3. Quantitative results (PSNR/SSIM) of different noise modeling methods on Nikon D850 Camera (ELD). Red and blue are utilized to indicate the best and second best results, respectively.

Dataset	Ratio	Index	Physics-based			Learning-based				Real noise-based			
			P-G [10]	ELLE [29]	ELD [31]	NoiseFlow [1]	CA-GAN [5]	Starlight [25]	Ours	Paired Data (LLD)	SFRN [33]	PMN [9]	Ours*
ELD [31]	×100	PSNR	41.71	42.04	42.86	41.64	41.55	42.12	42.94	43.01	43.04	43.28	43.58
		SSIM	0.910	0.916	0.949	0.881	0.886	0.927	0.949	0.951	0.949	0.960	0.963
	×200	PSNR	40.05	39.94	41.08	38.96	38.72	40.33	41.11	41.12	41.28	41.32	41.63
		SSIM	0.890	0.886	0.938	0.820	0.822	0.907	0.938	0.936	0.930	0.941	0.945

methods, and 3 real noise-based methods. For a fair comparison, we follow all settings of PMN [9] except the noise model and noisy image synthesis strategy. For the Sony A7S2 camera, we use the LLD dataset for training all the noise models. When training the denoising network, all noise models are applied to SID [6] data and the same denoising network from SID is adopted. For the Nikon D850 camera, we retrain the learning-based model using the noisy-clean image pairs in our LLD dataset. P-G [10], ELLE [29], ELD [31], NoiseFlow [1], CA-GAN [5], Starlight [25] and Ours only require long-exposure clean images from paired data for noise synthesis. SFRN [33] requires clean images and real bias frames. As mentioned in Sec. 3.4, PMN [9] and Ours* require paired noisy-clean images for noise synthesis. For a fair comparison, the training data for the learning-based noise models are consistent, and the implementation details and training strategies follow the official settings. Since SFRN [33] and ELLE [29] do not release pre-trained models and code, we use bias and flat-field frames from our LLD to calibrate the noise parameters as described in their paper.

4.2.1 Evaluation on Sony A7S2 Camera

The results of the comparison between the 7 methods in the clean image-based synthesis strategy and the 4 methods in the real noise-based synthesis strategy are summarised in Tab. 2. Early learning-based methods [1, 5] are even inferior in performance to the Poissonian-Gaussian noise and ELLE [29], due to that their network design and training strategies ignore the physical properties of the noise. Despite being inspired by physical models, Starlight [25] still has a significant performance gap over the classical low-light noise modeling (*e.g.*, ELD [31]), since the ISO dependence and BLE noise are ignored. Our proposed

Table 4. PSNR/SSIM with ablation studies about noise components and training strategy performed on the SID and ELD.

Noise Components					Training Mode		SID			ELD	
N_{HG}	N_{BLE}	N_q	N_{row}	$k \cdot N_{FP}$			100	250	300	100	200
✔					Two Stage	PSNR	39.62	35.28	31.98	43.08	40.02
						SSIM	0.913	0.821	0.739	0.931	0.871
✔	✔	✔				PSNR	41.65	39.56	36.37	44.86	42.99
						SSIM	0.951	0.928	0.906	0.968	0.944
✔		✔	✔			PSNR	40.87	37.00	33.77	44.75	42.50
						SSIM	0.934	0.872	0.812	0.955	0.923
✔	✔	✔	✔			PSNR	41.89	39.54	36.43	45.28	43.33
						SSIM	0.953	0.931	0.910	0.973	0.951
✔	✔	✔	✔	✔		PSNR	**42.10**	**39.76**	**36.76**	**45.61**	**43.84**
						SSIM	**0.955**	**0.933**	**0.912**	**0.977**	**0.959**
✔	✔	✔	✔	✔	One Stage	PSNR	41.23	37.11	33.80	45.03	42.72
						SSIM	0.940	0.885	0.829	0.965	0.934

physical-guided ISO-related noise model (Ours) can outperform ELD [31] since we have considered the ISO dependence for black level error noise and FPN.

For real noise-based methods, PMN [9] adopts the dark-shading correction strategy, which outperforms Paired Data and SFRN [33] and inspires us to adopt similar strategies. Even though PMN adopts the Poisson noise augmentation strategy, our proposed zero-mean noise model can augment signal-dependent and signal-independent noise simultaneously. As shown in Fig. 2, our method provides the clearest texture and most accurate color denoising result.

4.2.2 Evaluation on Nikon D850 Camera

The results on the Nikon D850 camera between different noise modeling methods are summarised in Tab. 3. Compared to the physics-based method, our proposed data-driven noise model performs better and eliminates the laborious noise parameter calibration procedure. Owing to the use of real noise data, real noise-based synthesis methods are superior to clean image-based synthesis methods. It is worth noting that PMN [9] requires not only pairs of noisy-clean images, but also a large number of bias and flat-field frames. By contrast, our method relies only on paired noisy-clean data and achieves the best performance. More qualitative results are given in the supplementary material.

4.3. Ablation Study

All the ablation experiments are performed on the Sony A7S2 camera for both SID [6] and ELD [31].

Noise Components and Training Strategy. We first study the effectiveness of our five noise components and two-stage training strategy, *i.e.*, random and fixed-pattern noise modeling stages. In Tab. 4, one can see that, removing the noise components N_{BLE} or $k \cdot N_{FP}$ results in great performance degradation, due to that the color bias introduced by dark currents greatly damages visual quality in extreme low-light scenes. When the two-stage training strategy is eliminated, the learning-based noise model fails to model black level error and dark current fixed-pattern noise accurately, which therefore results in a significant degradation in performance. The heteroscedastic Gaussian and row noise also contribute to the final performance.

Table 5. PSNR/SSIM performance with ablation studies about different black level error (BLE) noise.

BLE	Index	SID			ELD	
		100	250	300	100	200
ELLE [29]	PSNR	41.78	39.66	36.65	45.22	43.34
	SSIM	0.954	0.932	0.907	0.975	0.958
ELD [31]	PSNR	42.02	39.73	36.59	45.11	43.32
	SSIM	0.955	0.933	0.903	0.973	0.956
Ours	PSNR	**42.10**	**39.76**	**36.76**	**45.61**	**43.84**
	SSIM	**0.955**	**0.933**	**0.912**	**0.977**	**0.959**

Table 6. PSNR/SSIM performance with ablation studies about different real noise-based synthesis strategies.

Training (Input, Target)	Inference Input	Index	SID			ELD	
			100	250	300	100	200
(D, I)	D^{test}	PSNR	42.06	39.60	36.85	44.47	41.97
		SSIM	0.955	0.938	0.923	0.968	0.928
$(D - k \cdot N_{FP} - N_{BLE(ISO)}, I)$	$D^{test} - k \cdot N_{FP} - N_{BLE(ISO)}$	PSNR	42.84	40.53	37.57	46.12	44.34
		SSIM	0.960	0.946	0.933	0.982	0.974
$(I + N_{HG} + N_{\text{row}} + N_q, I)$	$D^{test} - k \cdot N_{FP} - N_{BLE(ISO)}$	PSNR	42.75	40.41	37.15	46.24	44.51
		SSIM	0.956	0.932	0.916	0.984	0.973
$(D - k \cdot N_{FP} - N_{BLE(ISO)}, I)$, $(I + N_{HG} + N_{\text{row}} + N_q, I)$	$D^{test} - k \cdot N_{FP} - N_{BLE(ISO)}$	PSNR	**43.36**	**41.02**	**37.80**	**46.74**	**44.95**
		SSIM	**0.961**	**0.948**	**0.935**	**0.986**	**0.977**

ISO-Dependence of Black Level Error Noise. In addition to our more complete modeling of noise, the main difference between our noise model and existing methods is that we regard the black level error (BLE) noise as ISO-dependent. Such an operation is contrary to the assumption in ELLE [29] (which adopts image-specific BLE parameters and simulates the black level error noise with a uniform distribution) and ELD [31] (which randomly samples from real black level errors). For verifying our noise model, we conduct experiments with different black level error noise sources. From Tab. 5, our method has better performance, showing the rationality of our noise model.

Real noise-based Synthesis Strategy. As mentioned in Secs. 3.4 and 4.2, the real noise-based synthesis strategy can greatly improve the performance of denoising networks compared to training on paired data. Thanks to the flexibility of our noise model, we have also conducted an experiment to evaluate different strategies for utilizing the real noise-based synthesis method. The results in Tab. 6 show the effectiveness of our proposed strategy in Sec. 3.4.

5. Conclusion

In this paper, we proposed a novel physics-guided ISO-dependent sensor noise model for modeling real rawRGB noises in extreme low-light scenes and collected an LLD dataset under a more reasonable data collection scheme. The noise model implemented in the normalizing flow-based framework achieved appealing accuracy and marvelous flexibility without any manual calibrations, and the denoiser trained with our noise model also achieved superior performance. In the future, we will consider more noise types in our noise model such as the periodic noise.

Acknowledgement This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. U19A2073.

References

[1] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *IEEE International Conference on Computer Vision*, pages 3165–3173, 2019. 1, 2, 3, 6, 7

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 2, 3

[3] European Machine Vision Association et al. Standard for characterization of image sensors and cameras. *EMVA Standard*, 1288, 2010. 3

[4] Stephen C Cain, Majeed M Hayat, and Ernest E Armstrong. Projection-based image registration in the presence of fixed-pattern noise. *IEEE Transactions on Image Processing*, 10(12):1860–1872, 2001. 2

[5] 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 *European Conference on Computer Vision*, pages 343–358. Springer, 2020. 1, 3, 7

[6] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 2, 3, 7, 8

[7] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3155–3164, 2018. 3

[8] Yuan Chen, Steven M Guertin, Mihail Petkov, Duc N Nguyen, and Frank Novak. A chip and pixel qualification methodology on imaging sensors. In *IEEE International Reliability Physics Symposium. Proceedings*, pages 435–439. IEEE, 2004. 3

[9] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *ACM International Conference on Multimedia*, pages 1436–1444, 2022. 1, 2, 3, 5, 6, 7, 8

[10] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008. 3, 4, 7

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 3

[12] 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. 2, 3, 4

[13] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. *Advances in Neural Information Processing Systems*, 30, 2017. 3

[14] Yeonseong Hwang, Seongjoo Lee, and Minkyu Song. Design of a cmos image sensor with a 10-bit two-step single-slope a/d converter and a hybrid correlated double sampling. In *2014 10th Conference on Ph. D. Research in Microelectronics and Electronics (PRIME)*, pages 1–4. IEEE, 2014. 3

[15] 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. 2

[16] Geonwoon Jang, Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. C2N: Practical generative noise modeling for real-world denoising. In *IEEE International Conference on Computer Vision*, pages 2350–2359, 2021. 3

[17] 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. 2

[18] Dong-Wook Kim, Jae Ryun Chung, and Seung-Won Jung. GRDN: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 3

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

[20] Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in Neural Information Processing Systems*, 31, 2018. 3

[21] Mikhail Konnik and James Welsh. High-level numerical simulations of noise in ccd and cmos photosensors: review and tutorial. *arXiv preprint arXiv:1412.4031*, 2014. 2, 3, 5

[22] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 773–782, 2018. 1

[23] Matteo Maggioni, Enrique Sánchez-Monge, and Alessandro Foi. Joint removal of random and fixed-pattern noise through spatiotemporal video filtering. *IEEE Transactions on Image Processing*, 23(10):4282–4296, 2014. 2, 5

[24] Ali Maleky, Shayan Kousha, Michael S. Brown, and Marcus A. Brubaker. Noise2NoiseFlow: Realistic camera noise modeling without clean images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 17632–17641, June 2022. 3

[25] Kristina Monakhova, Stephan R Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: video denoising in starlight. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 16241–16251, 2022. 1, 2, 3, 6, 7

[26] Ben Moseley, Valentin Bickel, Ignacio G López-Francos, and Loveneesh Rana. Extreme low-light environment-driven image denoising over permanently shadowed lunar regions with a physical noise model. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6317–6327, 2021. 5

[27] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1586–1595, 2017. 3

[28] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pages 1530–1538. PMLR, 2015. 3

[29] Jing Wang, Yitong Yu, Songtao Wu, Chang Lei, and Kuanhong Xu. Rethinking noise modeling in extreme low-light environments. In *IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2021. 1, 2, 3, 4, 6, 7, 8

[30] Yuzhi Wang, Haibin Huang, Qin Xu, Jiaming Liu, Yiqun Liu, and Jue Wang. Practical deep raw image denoising on mobile devices. In *European Conference on Computer Vision*, pages 1–16. Springer, 2020. 2

[31] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1, 2, 3, 6, 7, 8

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

[33] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *IEEE International Conference on Computer Vision*, pages 4593–4601, 2021. 1, 2, 3, 5, 7, 8