

Towards General Low-Light Raw Noise Synthesis and Modeling

Feng Zhang¹ Bin Xu² Zhiqiang Li^{1,2} Xinran Liu¹ Qingbo Lu² Changxin Gao¹ Nong Sang^{1†}

¹National Key Laboratory of Multispectral Information Intelligent Processing Technology,
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology

²DJI Technology Co., Ltd

{fengzhangai, lxryx, cgao, nsang}@hust.edu.cn, {mila.xu, cristopher.li, qingbo.lu}@dji.com

Abstract

Modeling and synthesizing low-light raw noise is a fundamental problem for computational photography and image processing applications. Although most recent works have adopted physics-based models to synthesize noise, the signal-independent noise in low-light conditions is far more complicated and varies dramatically across camera sensors, which is beyond the description of these models. To address this issue, we introduce a new perspective to synthesize the signal-independent noise by a generative model. Specifically, we synthesize the signal-dependent and signal-independent noise in a physics- and learning-based manner, respectively. In this way, our method can be considered as a general model, that is, it can simultaneously learn different noise characteristics for different ISO levels and generalize to various sensors. Subsequently, we present an effective multi-scale discriminator termed Fourier transformer discriminator (FTD) to distinguish the noise distribution accurately. Additionally, we collect a new low-light raw denoising (LRD) dataset for training and benchmarking. Qualitative validation shows that the noise generated by our proposed noise model can be highly similar to the real noise in terms of distribution. Furthermore, extensive denoising experiments demonstrate that our method performs favorably against state-of-the-art methods on different sensors.

1. Introduction

Low-light raw denoising is an important yet challenging problem for the increasingly widespread computational photography [53]. Due to the powerful computational capability of deep learning, the learning-based low-light raw denoising algorithms have shown great superiority and become the mainstream manner [10, 21] in recent years. However, since the standard paradigm of deep learning is to learn a mapping from the low-light noisy raw image to its normal-

† Corresponding authors.

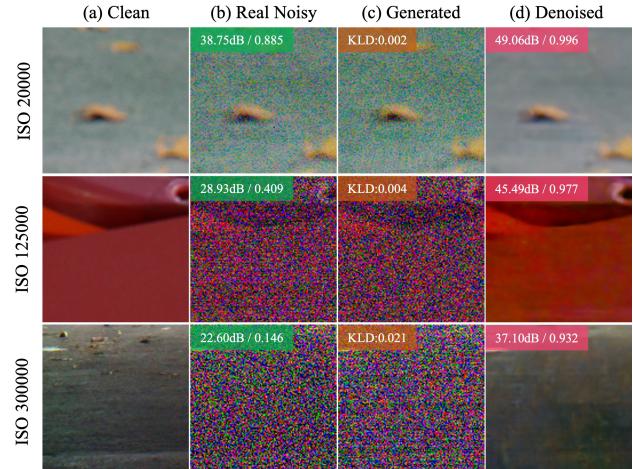


Figure 1. Examples of generated and denoised raw images from our proposed method. (a) Clean raw image, (b) Real noisy raw image, (c) Generated noisy raw image from our proposed noise model, (d) Denoised raw image of the denoise network [10] trained on the image generated by our noise model. Our model can generate noise at different ISO levels by a single learned noise model. The results of the visualization and quantitative metric Kullback-Leibler divergence (KLD) demonstrate that the distribution of the synthetic noise is close to the real noise distribution. Noise at high ISO levels drowns out the signal, resulting in unsatisfactory image denoising.

light counterpart, this leads to a reliance on the large-scale dataset of real-world noisy-clean raw image pairs, which is extremely tedious and labor-intensive to collect.

A naive strategy is to synthesize the low-light noisy raw images to obtain more paired training data. Existing noise synthesis methods can be roughly categorized into physics-based noise models and Deep Neural Network (DNN)-based noise models.

Physics-based noise modeling [23, 54, 52] is the most commonly used noise synthesis method in low-light conditions, which obtains the statistical distribution of different noise sources by analyzing the physical process of cam-

era sensor imaging. However, noise sources on different camera sensors vary widely due to differences in circuit design and signal processing techniques, making it impossible for physics-based methods to extract and model all noise sources accurately. Moreover, each noise source's properties and statistical behavior vary significantly at different exposure times or ISO levels, making physics-based methods tedious and error-prone. All these limitations make it impossible for a physics-based method to achieve accurate noise modeling on multiple camera sensors.

DNN-based noise modeling [1, 9] learns to synthesize noise from real captured datasets with deep generative networks. Although the existing deep models show promising synthetic results on raw images due to their powerful representation capability, some previous studies [58, 44] have revealed that they perform poorly on extremely low-light raw images.

In this paper, we present a new perspective on synthesizing realistic low-light raw noise. Specifically, instead of directly synthesizing noise with generative networks, we separate the noise synthesis process into two components, i.e., signal-dependent and signal-independent, which are implemented through a physics-based manner and a learning-based manner, respectively. We employ a pre-trained denoise network during the training procedure to transfer the synthesized and real noisy raw images into a nearly noise-free image space to perform image domain alignment. Meanwhile, to better distinguish the synthesized and real noise, we present an effective multi-scale discriminator, namely Fourier transformer discriminator (FTD), to perform the noise domain alignment. In addition, we collect a new low-light raw denoising (LRD) dataset for training and benchmarking. Extensive experiments demonstrate that our noise model performs favorably against existing state-of-the-art noise models on different camera sensors. Fig. 1 shows examples of the synthetic noisy raw images and the corresponding denoising results trained on the synthetic noisy raw images.

In conclusion, our contributions can be summarized into three aspects:

- We propose a general noise model with separated synthesis processes to express the noise terms of according characteristics, enabling the noise model to imitate accurate low-light raw noise on different sensors.
- We establish an effective multi-scale discriminator framework, namely Fourier transformer discriminator (FTD), which encourages the generator to favor solutions that reside on the manifold of real low-light raw noise distributions.
- We collect a new large-scale dataset for low-light raw denoising benchmarking and researching.

2. Related Work

Deep Image Denoising. Image denoising is an extensively-studied yet still unresolved issue in computational photography. In the designing of traditional denoising algorithms, making an analytical regularizer related to image priors (e.g., sparsity [19, 3], smoothness [46, 49], self-similarity [7, 16, 41], low rank [27]) plays a critical role. In the modern era, deep learning-based algorithms [57, 6, 29, 55] have demonstrated their powerful superiority. However, most of them are based on the assumption that the noise conforms to a Gaussian distribution, but the noise captured in the real world is much more complex than Gaussian noise, which makes these methods even inferior to the traditional method BM3D [15]. To address this problem, several works have established a database of noisy and clean image pairs taken by real cameras as a benchmark [2, 45], thus improving the denoising performance of learning-based algorithms [13, 11, 24] in real-world scenes. Although this line of work is promising, the burden of acquiring real image pairs is heavy, and the collected pairs suffer from pixel misalignment, luminance misalignment, limited data volume, and lack of diversity.

Physics-based Noise Model. The additive white Gaussian (AWGN) noise model is the most widely-used physics-based noise model. However, it deviates strongly from the realistic noise distribution, which leads to significant performance degradation on images with real noise [45, 2]. The classical Poisson-Gaussian (P-G) noise model [23, 22, 30, 6, 42, 52] is proposed to handle the domain gap between synthetic images and real images, which considers both photon shot noise and sensor readout noise. Most recently, Wei *et al.* [54] has developed a novel noise model by analyzing the noise generation process in the image processing pipeline and obtaining the distribution of noise sources by using statistical methods. Nonetheless, the noise sources vary dramatically on different sensors, making it impractical to extract and model all kinds of noise sources accurately. Zhang *et al.* [58] proposes a novel strategy for synthesizing noisy raw images, which samples directly from a Poisson distribution and a database of dark frame images to obtain synthetic images. However, the dark frames vary with the exposure time and ISO level, which makes it challenging to obtain the dark frame database for all exposure times and ISO levels.

DNN-based Noise Model. Early works [12, 35] attempt to synthesize realistic noisy images by utilizing Generative Adversarial Networks (GANs), but they provide limited improvement in real-world denoising performance. Abdelhamed *et al.* [1] proposes a novel generative noise model, Noise-Flow, based on normalization flow to synthesize realistic noisy raw image, which is the first generative model that provides powerful noise modeling capabilities. Chang *et al.* [9] presents a camera-aware generative noise

model termed CA-NoiseGAN to generate multiple noises for different camera sensors. Monakhova *et al.* [44] introduces a physics-inspired generative noise model to learn the noise distribution of the camera sensor in low-light conditions, but this approach is limited to synthesizing noise that mimics a single ISO level, which is cumbersome in practical applications. In this work, we introduce a generative noise model to synthesize noise for different ISO levels and generalize it to various camera sensors.

3. Methodology

3.1. Sensor Noise Formation

To model more realistic low-light raw noise, it is necessary to understand the noise formation of Complementary Metal-Oxide-Semiconductor (CMOS) [47]. For a raw image produced by a CMOS sensor, the raw image formation process from incident photons to digital values can be modeled as follows:

$$D = KI + N, \quad (1)$$

where K presents the gain of the whole system, I stands for the number of photoelectrons stimulated by the scene radiation, N denotes the summation of all noise sources physically caused by light or camera.

Due to several physical limitations of CMOS, noise occurs from various sources. According to the relation between noise and light intensity [31, 18, 20], we analyze the raw image formation in Eq. (1) and categorize the raw image noise into signal-dependent and signal-independent components:

$$D = K(I + N_{dep}) + N_{indep}, \quad (2)$$

where N_{dep} represents the signal-dependent noise, and N_{indep} represents the signal-independent noise. Generally, the signal-dependent noise includes photon-shot noise and photo response non-uniformity. For the signal-independent noise, it includes read-out noise, fix pattern noise [51], dark current noise [4], quantization noise, flicker noise [5], thermal noise, reset noise [37], and so on.

3.2. Separated Noise Modeling

According to the noise formation process described in Eq.(2), we can separate the noise synthesis process into signal-dependent and signal-independent components. In well-lit environments, these two noise components can be accurately synthesized by physics-based models [58]. However, in low-light conditions, the signal-independent noise becomes more complex and varies significantly among camera sensors. Existing physics-based methods cannot accurately model this noise component, and building a noise model for each sensor is laborious. Therefore, we adopt a generative model to synthesize the signal-independent

noise. For the signal-dependent noise, we still utilize a physics-based model for synthesis since most previous works [22, 54, 58] have shown that the physics-based model can synthesize it accurately at a lower price. The framework of our proposed noise model is shown in Fig 2.

Signal-Dependent Noise Synthesis. From the noise formation process of CMOS sensors discussed above, we can observe that the signal-dependent noise consists of photon shot noise and photo response non-uniformity. However, according to previous reserches [26, 33], the photo response non-uniformity contribute less than 3% in the signal-dependent noise, which has a minimal impact. Therefore, the photon shot noise can be considered the only signal-dependent noise source.

Due to the intrinsic stochastic nature of photons reaching the CMOS sensor, the photon shot noise can be directly sampled from the Poisson distribution \mathcal{P} :

$$(N_{dep} + I) \sim \mathcal{P}(I). \quad (3)$$

For the synthesis of signal-dependent noise, extensive studies [54, 58] have demonstrated that the incident photon numbers strictly follow the Poisson distribution. Thus we can accurately synthesize the signal-dependent noise in a physics-based manner, which can be modeled as follows:

$$\hat{Y} = (N_{dep} + I) = \mathcal{P}\left(\frac{Y}{K}\right)K, \quad (4)$$

where Y is the clean image, \hat{Y} is the sampled image contaminated by signal-dependent noise. The overall system gain K can be easily obtained from the meta information of DNG file [32].

Signal-Independent Noise Synthesis. Previous studies [54, 58] have demonstrated that signal-independent noise is the dominant component in low-light conditions. Since the noise sources in the signal-independent component are extraordinarily complicated and vary significantly with different exposure times, ISO levels, and camera sensors. Instead of adopting a physics-based approach to model the signal-independent noise, we exploit the powerful learning capabilities of the Generative Adversarial Networks (GAN) to model it. Fig. 2 shows the overall framework of our proposed noise model.

To synthesize signal-dependent noise, we first sample a random initial noise map to reflect the stochastic noise behavior according to the conditions of each ISO level. Then, we fed the random initial noise map into a noise generator, which we utilize a typical residual 2D U-shape architecture [48]. (More details of the generator architecture can be found in the supplementary material.) Formally, this synthesis process can be formulated as follow:

$$N_{indep} = G(N_{init} \sim \mathcal{N}(0, \sigma_r^2)), \quad (5)$$

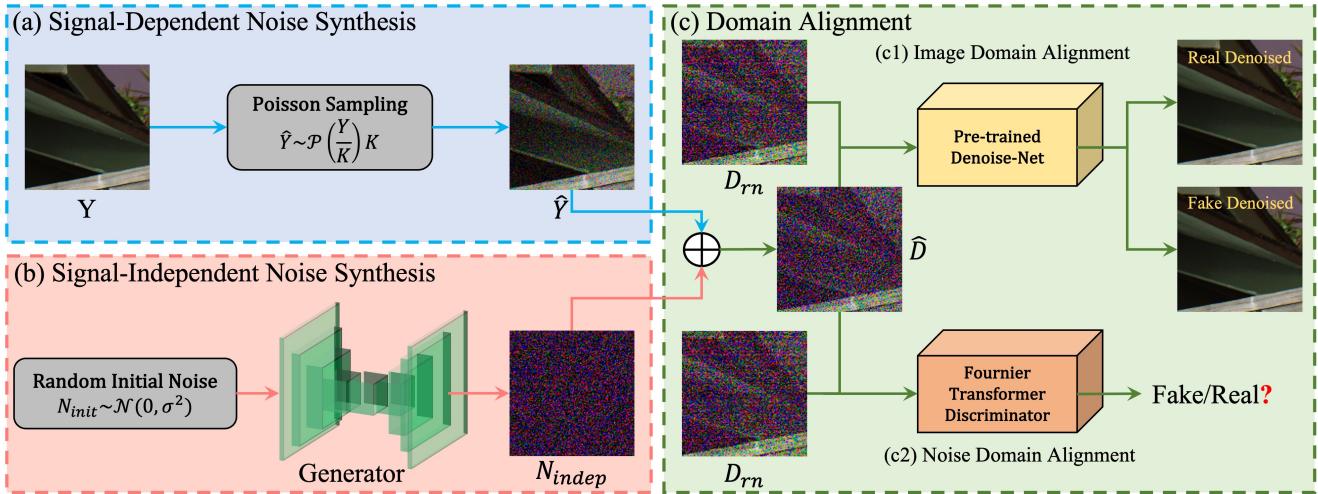


Figure 2. Overview of the framework. The proposed noise model is divided into three components: (a) signal-dependent noise synthesis, (b) signal-independent noise synthesis, and (c) domain alignment. Please refer to Sec. 3 for more detailed descriptions.

where N_{init} is the sampled random initial noise map, G is the proposed U-shape noise generator. We sample the random initial noise map from $\mathcal{N}(0, \sigma_r^2)$ and spatially replicate through all pixel positions of the clean raw image Y . σ_r is the noise parameter of the signal-independent component in the in-camera noise profile, which is related to the camera ISO levels. Similar to the conditional GAN [43], this parameter can be specified as a condition to control the noise level of N_{init} , which enables the network to generate different ISO levels of signal-independent noise.

Following the synthesis of signal-dependent and signal-independent noise by the two methods described above. Given a clean raw image Y , we can produce a pseudo-noisy raw image \hat{D} as follows:

$$\hat{D} = K(I + N_{dep}) + G(N_{init}) = \hat{Y} + N_{indep}. \quad (6)$$

In the training phase, we begin by taking paired raw images as inputs and subsequently extract essential parameters, including ISO levels, exposure times, and in-camera noise profiles from the noisy raw images, employing ExifTool. The generation of noisy raw images is accomplished by utilizing both the in-camera noise profile and the clean raw images as inputs to the generator.

During the inference phase, we adopt a different approach. Here, we leverage the in-camera noise profile of the target ISO level, in combination with the clean raw image, to serve as inputs. These inputs are then jointly fed into the generator, facilitating the synthesis of the noisy raw images corresponding to the desired target ISO level.

Domain Alignment. The most common strategy in image generation to construct image domain alignment is to minimize the distance between synthetic and real images directly. However, since the noise generator should produce different noise samples at each forward pass, it is incompat-

ible with deploying the \mathcal{L}_1 loss directly between the synthesized noisy raw image \hat{D} and the real noisy raw image D_{rn} . Besides, this strategy may drastically damage the quality of \hat{D} due to the stochastic characteristics of noise [8]. Therefore, the interference of noise should be excluded in the process of constructing image domain alignment.

To tackle this issue, inspired by [8], we introduce a pre-trained denoise network [10] (Denoise-Net) to transfer \hat{D} and D_{rn} into virtually noise-free image space. Then, perform \mathcal{L}_1 loss between the denoised versions of \hat{D} and D_{rn} . Since the denoised image is relatively stable, minimizing the \mathcal{L}_1 loss enables the synthesized noise to converge to the real noise distribution while preserving the stochastic characteristics of the noise. The proposed \mathcal{L}_1 loss can be formulated as follows:

$$\mathcal{L}_1 = \| P(\hat{D}) - P(D_{rn}) \|_1, \quad (7)$$

where P present the Denoise-Net. Note that we utilize the same paired data to train the Denoise-Net and the noise model, thus eliminating potential domain gap issues.

We adopt adversarial learning for noise domain alignment to make the generated noise distribution fit real noise distribution as closely as possible. General discriminators perform superiorly in discriminating noise-free or noisy images with low noise levels. However, we find that they are insufficient to discriminate noisy raw images with a high noise level, especially for low-light noisy raw images. (See Table 4.) To tackle this issue, we introduce a new discriminator. See more details in the following subsection.

3.3. Fourier Transformer Discriminator

Based on the spectral transform theory, it is postulated that noise can be categorized as high-frequency information, whereas the content information of an image is typ-

生出來的噪音要真實，denoise model 才會好的。這樣就可以保留噪音會出現的隨機性。

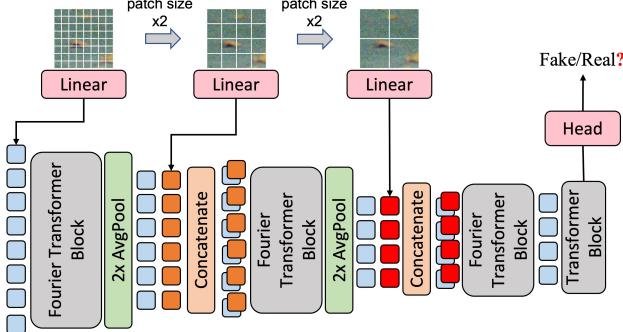


Figure 3. Framework of the proposed Fourier Transformer Discriminator. The architecture comprises three Fourier transformer blocks and one vanilla transformer block.

ically associated with low-frequency components. Consequently, the operation performed in the spectral domain can more accurately differentiate between the noise and content information. Therefore, we present a transformer block called Fourier transformer block inspired by fast Fourier convolution (FFC) [14]. The structure of this block is constructed by imitating the FFC, which replaces the convolution layer with the transformer block. (See supplementary material for the detailed structure.) Like the FFC, our proposed Fourier transformer block also takes two interlinked paths: a spectral path that conducts operation in the spectral domain with half of the input sequences and a spatial path that operates with another half. Each path receives complementary information and then performs an exchange internally to fuse the information.

Motivated by TransGAN [34], we establish an effective multi-scale discriminator, namely Fourier transformer discriminator (FTD), based on the proposed Fourier transformer block. The overall framework of FTD is shown in Fig 3. The 4-channel input raw image is firstly divided into three sequences by choosing different patch sizes. Then, these three sequences are fed into different Fourier transformer blocks through a linear transformation. The most extended sequence is combined with the learnable position encoding to serve as the input of the first block. The second and third sequences are concatenated into the second and third blocks. To reduce the resolution of the feature map between each block, we convert the 1D sequences to 2D feature maps and feed them into the Average Pooling layer. At last, a vanilla transformer block and a classification head are applied to output the prediction score.

3.4. Training Objective

To achieve the training objective, we optimize the generator and discriminator in an adversarial way [25]. Among adversarial frameworks, we select WGAN-GP [28] to calculate the adversarial loss \mathcal{L}_{adv} , which minimizes the Wasserstein distance to stabilize the training. The loss can



Figure 4. Example images of the LRD dataset. First column: long exposure reference (ground truth) images. Second column: low-light images with -1EV. Third column: low-light images with -2EV. Fourth column: low-light images with -3EV.

be defined as follows:

$$\begin{aligned} \mathcal{L}_{adv} = & \mathbb{E}_{\hat{D} \sim \mathbb{P}_g} [D_F(\hat{D})] - \mathbb{E}_{D_{rn} \sim \mathbb{P}_r} [D_F(D_{rn})] \\ & + \lambda \mathbb{E}_{\tilde{x} \sim \mathbb{P}_{\tilde{x}}} [\|(\nabla_{\tilde{x}} D_F(\tilde{x}))\|_2 - 1]^2, \end{aligned} \quad (8)$$

where D_F is our proposed discriminator FTD, \mathbb{P}_g is the synthetic noisy data distribution defined by the generator, \mathbb{P}_r is the real noisy data distribution, D_{rn} is the real noisy raw image.

In addition to the aforementioned \mathcal{L}_1 loss, we employ the perceptual loss to improve the perceptual quality:

$$\mathcal{L}_{per} = \|\phi(P(\hat{D})) - \phi(P(D_{rn}))\|_2^2, \quad (9)$$

where $\phi(\cdot)$ denotes the feature map extracted from a VGG-19 [50] model pre-trained on ImageNet [17].

To summarize, the full objective of our proposed noise model is combined as follows:

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{per}, \quad (10)$$

where λ_1 and λ_2 are two hyper-parameters to control the balance of loss functions. In our experiment, these parameters are set to $\lambda_1 = 0.1$, $\lambda_2 = 0.01$.

4. Low-light Raw Denoising (LRD) Dataset

Creating low-light raw image datasets is essential for standardizing low-light raw denoising techniques. The existing dataset for benchmarking low-light raw denoising is the See-in-the-Dark (SID) dataset [10]. However, since it is designed to produce perceptually good images in low-light conditions, there are some limitations in benchmarking low-light raw denoising. First, the long-exposure reference images in this dataset still contain some noise, which may disorient the generative noise model. Second, the varying number of images for each ISO level and exposure time

Table 1. Quantitative comparisons on the SID, ELD and LRD datasets in terms of PSNR and SSIM. The results are conducted on different exposure ratios. The red color indicates the best results, and the blue color indicates the second-best results.

Dataset	Ratio	Physics-based		Real-noise-based		DNN-based	
		Poisson-Gaussian		Paired data		Noise Flow	Ours
		PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
SID	×100	37.51 / 0.856	41.21 / 0.952	41.39 / 0.954	36.75 / 0.787	41.95 / 0.956	
	×250	31.67 / 0.765	38.54 / 0.929	38.90 / 0.937	33.98 / 0.739	39.25 / 0.931	
	×300	28.53 / 0.667	35.35 / 0.908	36.55 / 0.922	31.82 / 0.713	36.03 / 0.909	
ELD	×100	39.46 / 0.785	45.06 / 0.975	43.80 / 0.963	38.68 / 0.793	44.95 / 0.979	
	×200	33.81 / 0.615	43.21 / 0.954	41.54 / 0.918	36.30 / 0.713	43.32 / 0.966	
LRD	-1EV	33.77 / 0.895	38.31 / 0.968	38.80 / 0.970	35.19 / 0.874	38.89 / 0.971	
	-2EV	32.99 / 0.856	37.35 / 0.959	37.88 / 0.961	34.55 / 0.842	37.95 / 0.962	
	-3EV	31.44 / 0.770	36.49 / 0.950	36.92 / 0.951	33.72 / 0.826	37.01 / 0.953	

為什麼 low-light 也做得好？
- 應該是訓練 noise 時，也把暗部一併處理了？

may cause class imbalance issues. All these limitations hinder applications for low-light raw denoising and low-light raw image synthesis.

To address this issue, we collected a new low-light raw denoising (LRD) dataset for training and benchmarking. In contrast to the SID dataset, which sets a fixed exposure time to capture long and short exposure images, we captured long and short exposure images based on the exposure value (EV). Motivated by multi-exposure image fusion [40, 39], the exposure value for long exposure images was set to 0, and the exposure value for short exposure was set to the commonly used parameters -1, -2, and -3. The dataset is designed for application to low-light raw image denoising and low-light raw image synthesis.

The dataset contains both indoor and outdoor scenes. For each scene instance, we first captured a long-exposure image at ISO 100 to get a noise-free reference image. Then we captured multiple short-exposure images using different ISO levels and EVs, with a 1-2 second interval between subsequent images to wait for the sensor to cool down, thus avoiding unexpected noise introduced by sensor heating.

We captured 6 different ISO levels ranging from 200 to 6400 to obtain various noise levels. We captured 100 images at each ISO and EV setting to preserve a balanced training sample. Therefore, the total number of images in our dataset is 1800 (100 images × 6 ISO levels × 3 exposure value). Images were captured using a typical camera sensor: IMX586, which has a full-frame Bayer filter. The image resolution is 4000 × 3000. We mounted the camera on a sturdy tripod to ensure the sensor would not wobble. An example of long-short image pairs at different exposure values is shown in Fig. 4. We selected 10% of the images in each exposure value and ISO levels to form the test set and selected another 5% of the images as the validation set.

5. Experiments

5.1. Experimental Setup

Dataset. We first utilize the SID Sony set [10] or our proposed LRD dataset to train the pre-trained denoise net-

Table 2. Average Kullback-Leibler divergence (AKLD) [56] evaluation of different noise models. Our proposed noise model outperforms state-of-the-art methods on both SID and LRD datasets.

Dataset	Ratio	P-G	ELD	Noise Flow	Ours
SID	×100	0.179	0.117	0.162	0.075
	×250	0.254	0.177	0.249	0.113
	×300	0.325	0.231	0.293	0.119
LRD	-1EV	0.459	0.135	0.274	0.099
	-2EV	0.592	0.163	0.338	0.109
	-3EV	0.701	0.188	0.375	0.123

work. Then we fix the pre-trained denoise network to train the generator and discriminator on the same set. Subsequently, the generator takes the clean raw images from the SID Sony set or LRD dataset as input images to generate realistic pseudo-noisy raw images. The generated noisy-clean raw image pairs are utilized to evaluate the denoising performance on three benchmarks: the Sony set of SID and ELD datasets, and the LRD dataset. The images in the SID Sony set are collected using Sony α7SII in 231 static scenes. There are 1865 images for training, 234 for validation, and 598 for testing. The ELD Sony set [54] consists of 60 low-light noisy raw images for testing, which are also captured using the same camera as the SID.

Implementation Details. To optimize the noise generator and discriminator, we augment all training samples by randomly cropping and horizontal flipping to construct a mini-batch of size 128. The Adam [36] optimizer is adopted for training, while the initial parameters β_1 and β_2 to 0.5 and 0.999. The initial learning rate is set to 2×10^{-4} , and the patch size is set to 64×64 . The cosine annealing strategy is employed to steadily decrease the learning rate from the initial value to 10^{-6} during the training procedure, where the model is trained over 100 epochs. Denoising models are optimized using the generated training pairs from the trained generator with a mini-batch of size 1. The patch size is set to 512×512 . The Adam [36] optimizer is utilized with an initial learning rate of 2×10^{-4} , followed by halving at epoch 100 and finally to 5×10^{-5} at epoch 180.

直接用模型合成的 paired data (各自的資料集各自生), 不用 real world 的資料, 最後還有跟真實世界的比較

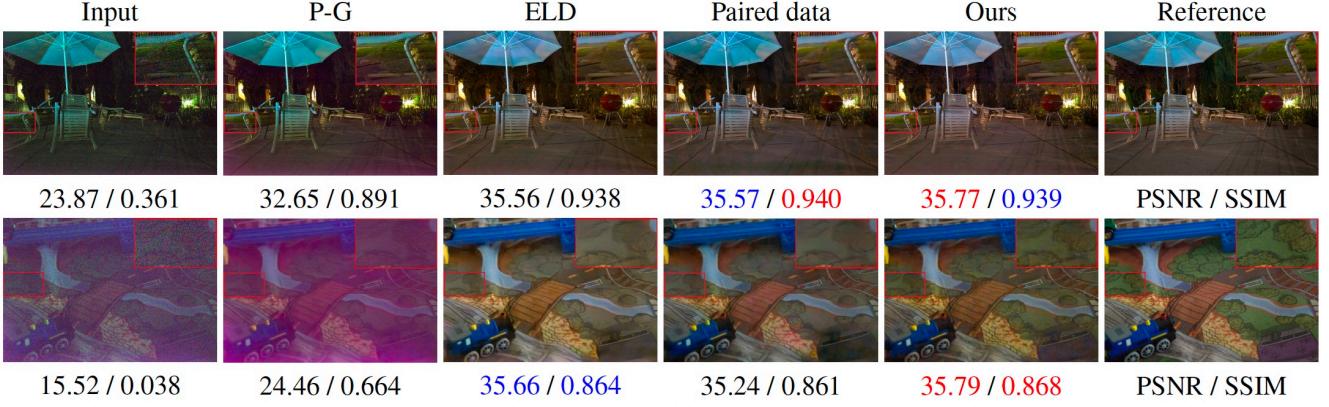


Figure 5. Raw image denoising comparison with state-of-the-art methods on low-light noisy raw images from the SID dataset [10]. Best viewed in color and by zooming in.

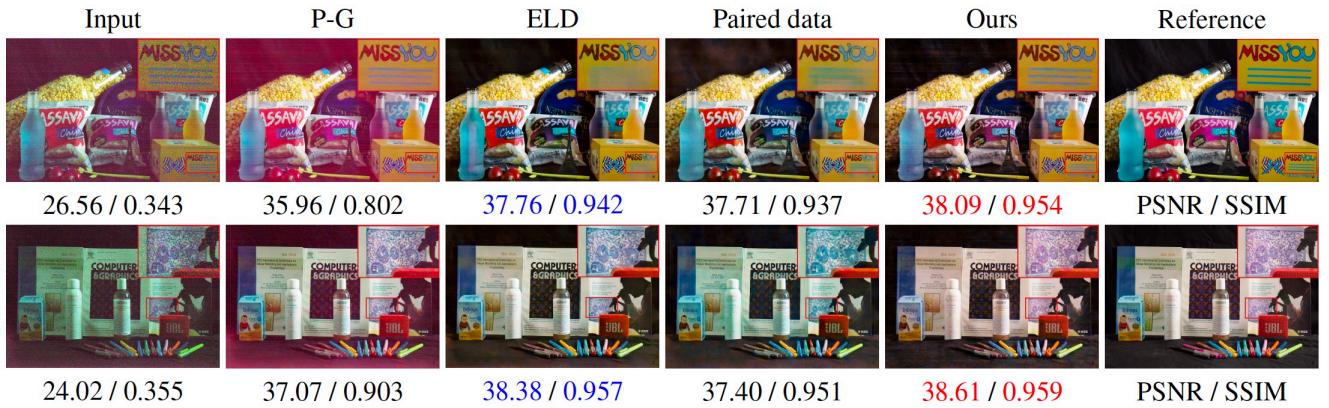


Figure 6. Raw image denoising comparison with state-of-the-art methods on low-light noisy raw images from the ELD dataset [54]. Best viewed in color and by zooming in.

The training runs for 200 epochs. We select the same U-shape architecture in [10] as our denoising baseline. All of our experiments are conducted on four-channel raw images. The implementation is conducted on the Pytorch framework with a single GeForce 2080Ti GTX GPU.

5.2. Model Analysis on Synthetic Noise

We first analyze the realism of the synthesized pseudo-noise raw images generated by our proposed generator. For quantitative comparison, we utilize the widely applied metric, Average Kullback-Leibler divergence (AKLD) [56], to measure the discrepancy between the real noise and the synthetic noise patches generated by different noise formation models in the SID dataset and LRD dataset. The results are depicted in Table. 2. Our proposed noise model achieves the minimum AKLD on both datasets. These results demonstrate that the distribution of synthetic noise generated by our proposed noise model more closely matches the real noise distribution.

5.3. Denoising Results on SID and ELD datasets

In order to demonstrate the reliability of our proposed noise model, we conduct raw image denoising experiments on SID and ELD datasets and compare them

with some state-of-the-art methods: physics-based noise model Poisson-Gaussian (P-G), ELD, and DNN-based noise model Noise Flow. Moreover, we also compare our noise model with the model trained with paired real data.

Table. 1 shows the comparison results on the SID and ELD datasets over different exposure ratios. For the physics-based methods, the ELD noise model significantly outperforms the P-G noise model in terms of both PSNR and SSIM scores. As described in the previous work [58], the DNN-based method Noise Flow performs poorly in low-light conditions, which indicates that this method is unsuitable for synthesizing low-light noisy raw images. Our noise model outperforms all existing low-light noise synthesis methods over different exposure ratios and even partially outperforms the denoiser trained with real paired data. This is because the real captured low-light raw image pairs still suffer from issues such as luminance misalignment and pixel misalignment [38], which may lead to unsatisfactory denoising results. A visual comparison of our noise model and other noise models is shown in Fig. 5 and Fig. 6. The P-G noise model is far from the real noise distribution. Thus it fails to remove the noise. The ELD noise model pro

真實世界收集的不太好，因為光線或是兩張照片本身就沒對齊

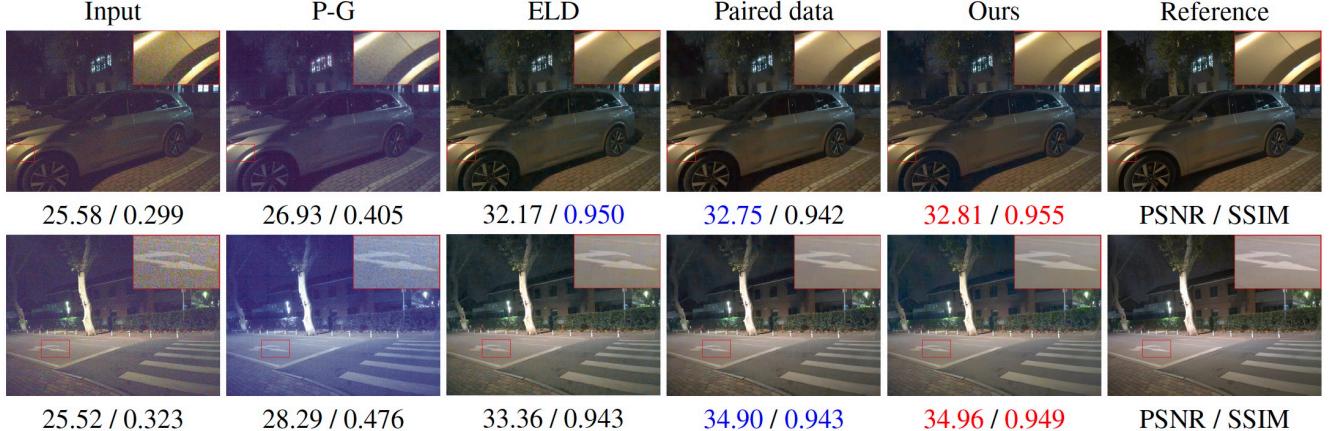


Figure 7. Raw image denoising comparison with state-of-the-art methods on low-light noisy raw images from our proposed LRD dataset. Best viewed in color and by zooming in.

Table 3. Ablation study of the contribution of each component in the model framework in terms of PSNR and SSIM. The best results have been shown in bold.

Components	Index	SID		
		$\times 100$	$\times 250$	$\times 300$
w/o Noise Separation	PSNR	39.91	37.53	33.05
	SSIM	0.902	0.883	0.867
w/o Denoise-Net	PSNR	41.09	38.83	35.66
	SSIM	0.937	0.913	0.899
Full framework	PSNR	41.95	39.25	36.03
	SSIM	0.956	0.931	0.909

vides better denoising results but suffers from color bias and residual noise. The result of denoising with paired real data obtains favorable visual quality but still suffers from imperfections. Our proposed noise model can provide competitive denoising results with visual pleasurable.

5.4. Denoising Results on LRD dataset

To further validate the generalization ability of our proposed noise model, we assess model performance on our LRD dataset. Table. 1 and Fig. 7 summarize all competing methods' quantitative and qualitative results. The results are consistent with the results on the SID and ELD datasets. Our method can still achieve comparable results with state-of-the-art methods. The P-G noise model still fails to remove the real noise, resulting in unsatisfactory visual results. ELD noise model provides satisfactory noise removal but still suffers from color bias issues. The results of our noise model are visually pleasing without significant residual noise or color bias, demonstrating our noise model's generalization ability across different camera devices.

5.5. Ablation Study

Impact of Each Components. We perform break-down ablations to evaluate each component's effects in the model framework. The comparison results evaluated on the SID dataset are reported in Table. 3. First, we remove the noise

Table 4. Ablation study of the contribution of Fourier transformer discriminator in terms of PSNR, SSIM. The best results have been shown in bold.

Discriminator	Index	SID		
		$\times 100$	$\times 250$	$\times 300$
Transformer	PSNR	40.23	37.36	33.22
	SSIM	0.931	0.910	0.892
Fourier Transformer	PSNR	41.95	39.25	36.03
	SSIM	0.956	0.931	0.909

separation strategy and directly synthesize the noise using the generative model, we follow the scheme of CA-NoiseGAN [9] and take the noise parameter instead of the initial Gaussian noise map as the condition information. The PSNR and SSIM values show a significant decrease. Secondly, after deploying the pre-trained denoise network (Denoise-Net), the performance of the noise model is improved moderately, suggesting that the pre-trained denoising network successfully performs image domain alignment and improves the quality of synthetic images. These results convincingly demonstrate the superiority of our proposed noise model in synthesizing low-light raw noise.

Effectiveness of Fourier Transformer Discriminator. A significant advantage of our proposed noise model over existing generative noise models is that the FTD can effectively fuse information through operations in spectral and spatial domains. To verify the effectiveness of the FTD, we employ a conventional discriminator with vanilla transformer blocks for comparison. Table. 4 shows the comparison results on the SID dataset, and our proposed FTD can discriminate the noise distribution more accurately, thus encouraging the generator to synthesize more accurate noise.

6. Conclusion

In this paper, we present a new perspective for realistic low-light raw noise synthesis. Specifically, we syn-

thesize the signal-dependent and signal-independent noise in a physics- and learning-based manner. We employ a pre-trained denoise network during the training procedure to transfer the synthesized and real noisy raw images into a nearly noise-free image space for image domain alignment. Meanwhile, we introduce an effective discriminator, namely Fourier transformer discriminator (FTD), to perform noise domain alignment. Our method is general for different ISO levels and different camera sensors. Subsequently, we collect a new low-light raw denoising (LRD) dataset for training and benchmarking. Both qualitative and quantitative experiments on the public datasets and our dataset collectively demonstrate the superiority of our method over state-of-the-art methods.

Acknowledgements. This work was supported by the National Natural Science Foundation of China No.62176097, Hubei Provincial Natural Science Foundation of China No.2022CFA055.

References

- [1] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3165–3173, 2019. 2
- [2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1692–1700, 2018. 2
- [3] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. 2
- [4] Richard L Baer. A model for dark current characterization and simulation. In *Sensors, Cameras, and Systems for Scientific/Industrial Applications VII*, volume 6068, pages 37–48. SPIE, 2006. 3
- [5] JA Barnes and DW Allan. A statistical model of flicker noise. *Proceedings of the IEEE*, 54(2):176–178, 1966. 3
- [6] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11036–11045, 2019. 2
- [7] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 2, pages 60–65. Ieee, 2005. 2
- [8] Yuanhao Cai, Xiaowan Hu, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, and Donglai Wei. Learning to generate realistic noisy images via pixel-level noise-aware adversarial training. *Advances in Neural Information Processing Systems*, 34:3259–3270, 2021. 4
- [9] Ke-Chi Chang, Ren Wang, Hung-Jin Lin, Yu-Lun Liu, Chia-Ping Chen, Yu-Lin Chang, and Hwann-Tzong Chen. Learn-ing camera-aware noise models. In *European Conference on Computer Vision*, pages 343–358. Springer, 2020. 2, 8
- [10] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 1, 4, 5, 6, 7
- [11] Chang Chen, Zhiwei Xiong, Xinmei Tian, and Feng Wu. Deep boosting for image denoising. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–18, 2018. 2
- [12] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3155–3164, 2018. 2
- [13] Yunjin Chen, Wei Yu, and Thomas Pock. On learning optimized reaction diffusion processes for effective image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5261–5269, 2015. 2
- [14] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 33:4479–4488, 2020. 5
- [15] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising with block-matching and 3d filtering. In *Image processing: algorithms and systems, neural networks, and machine learning*, volume 6064, pages 354–365. SPIE, 2006. 2
- [16] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 2
- [17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5
- [18] Abbas El Gamal and Helmy Eltoukhy. Cmos image sensors. *IEEE Circuits and Devices Magazine*, 21(3):6–20, 2005. 3
- [19] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006. 2
- [20] Joyce Farrell, Michael Okincha, and Manu Parmar. Sensor calibration and simulation. In *Digital Photography IV*, volume 6817, pages 249–257. SPIE, 2008. 3
- [21] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1436–1444, 2022. 1
- [22] Alessandro Foi. Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 89(12):2609–2629, 2009. 2, 3
- [23] 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. 1, 2

- [24] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 2
- [25] 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. 5
- [26] 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
- [27] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 2
- [28] 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. 5
- [29] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 2
- [30] Samuel W Hasinoff. Photon, poisson noise. *Computer Vision, A Reference Guide*, 4, 2014. 2
- [31] Glenn E Healey and Raghava Kondepudy. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, 1994. 3
- [32] ADOBE SYSTEMS INCORPORATED. Digital negative (dng) specification. https://www.adobe.com/content/dam/acom/en/products/photoshop/pdfs/dng_spec_1.4.0.0.pdf, 2012. 3
- [33] James R Janesick, Tom Elliott, Stewart Collins, Morley M Blouke, and Jack Freeman. Scientific charge-coupled devices. *Optical Engineering*, 26(8):692–714, 1987. 3
- [34] Yifan Jiang, Shiyu Chang, and Zhangyang Wang. Transgan: Two pure transformers can make one strong gan, and that can scale up. *Advances in Neural Information Processing Systems*, 34:14745–14758, 2021. 5
- [35] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2
- [36] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [37] 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. 3
- [38] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *International Conference on Machine Learning*, pages 2965–2974. PMLR, 2018. 7
- [39] Kede Ma, Zhengfang Duanmu, Hanwei Zhu, Yuming Fang, and Zhou Wang. Deep guided learning for fast multi-exposure image fusion. *IEEE Transactions on Image Processing*, 29:2808–2819, 2019. 6
- [40] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 6
- [41] Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. *IEEE Transactions on image processing*, 21(9):3952–3966, 2012. 2
- [42] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2502–2510, 2018. 2
- [43] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 4
- [44] Kristina Monakhova, Stephan R Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: video denoising in starlight. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16241–16251, 2022. 2, 3
- [45] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1586–1595, 2017. 2
- [46] Javier Portilla, Vasily Strela, Martin J Wainwright, and Eero P Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image processing*, 12(11):1338–1351, 2003. 2
- [47] Grand View Research. Image sensors market analysis. <http://www.grandviewresearch.com/industry-analysis/imagesensors-market>, 2016. 3
- [48] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18, pages 234–241. Springer, 2015. 3
- [49] 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
- [50] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5
- [51] Martijn F Snoeij, Albert JP Theuwissen, Kofi AA Makinwa, and Johan H Huijsing. A cmos imager with column-level adc using dynamic column fixed-pattern noise reduction. *IEEE Journal of Solid-State Circuits*, 41(12):3007–3015, 2006. 3
- [52] Wei Wang, Xin Chen, Cheng Yang, Xiang Li, Xuemei Hu, and Tao Yue. Enhancing low light videos by exploring high sensitivity camera noise. In *Proceedings of the IEEE/CVF*

- International Conference on Computer Vision*, pages 4111–4119, 2019. 1, 2
- [53] 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. 1
 - [54] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 1, 2, 3, 6, 7
 - [55] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang. Variational denoising network: Toward blind noise modeling and removal. *Advances in neural information processing systems*, 32, 2019. 2
 - [56] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 41–58. Springer, 2020. 6, 7
 - [57] 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. 2
 - [58] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4593–4601, 2021. 2, 3, 7