# A Physics-based Noise Formation Model for Extreme Low-light Raw Denoising

Kaixuan Wei[1]    Ying Fu[1*]   Jiaolong Yang[2]    Hua Huang[1]
[1]Beijing Institute of Technology     [2]Microsoft Research

## Abstract

*Lacking rich and realistic data, learned single image denoising algorithms generalize poorly to real raw images that do not resemble the data used for training. Although the problem can be alleviated by the heteroscedastic Gaussian model for noise synthesis, the noise sources caused by digital camera electronics are still largely overlooked, despite their significant effect on raw measurement, especially under extremely low-light condition. To address this issue, we present a highly accurate noise formation model based on the characteristics of CMOS photosensors, thereby enabling us to synthesize realistic samples that better match the physics of image formation process. Given the proposed noise model, we additionally propose a method to calibrate the noise parameters for available modern digital cameras, which is simple and reproducible for any new device. We systematically study the generalizability of a neural network trained with existing schemes, by introducing a new low-light denoising dataset that covers many modern digital cameras from diverse brands. Extensive empirical results collectively show that by utilizing our proposed noise formation model, a network can reach the capability as if it had been trained with rich real data, which demonstrates the effectiveness of our noise formation model.*

## 1. Introduction

Light is of paramount importance to photography. Night and low light place very demanding constraints on photography due to limited photon count and inescapable noise. The natural reaction is to gather more light by, *e.g.*, enlarging aperture setting, lengthening exposure time and opening flashlight. However, each method is a tradeoff – large aperture incurs small depth of field, and is unavailable in smartphone cameras; long exposure can induce blur due to scene variations or camera motions; flash can cause color aberrations and is useful only for nearby objects.

A practical rescue for low-light imaging is to use burst capturing [46, 28, 42, 40], in which a burst of images are aligned and fused to increase the signal-to-noise ratio

*Corresponding author: fuying@bit.edu.cn



(a) Input          (b) G          (c) G+P

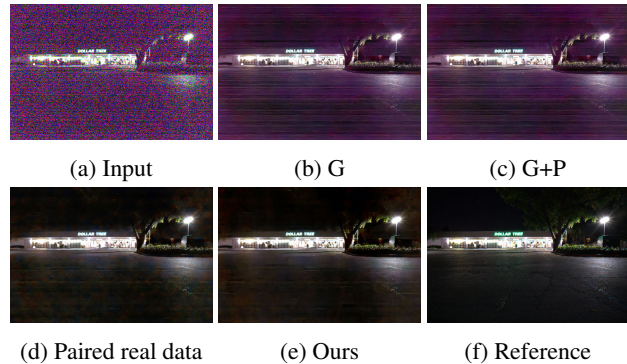(d) Paired real data      (e) Ours      (f) Reference

Figure 1: An image from the See-in-the-Dark (SID) Dataset [9], where we present (a) the short-exposure noisy input image; (f) the long-exposure reference image; (b-e) the outputs of UNets [51] trained with (b) synthetic data generated by the homoscedastic Gaussian noise model (G), (c) synthetic data generated by the signal-dependent heteroscedastic Gaussian noise model (G+P) [22], (d) paired real data of [9], and (e) synthetic data generated by our proposed noise model respectively. *All images were converted from raw Bayer space to sRGB for visualization; similarly hereinafter.*

(SNR). However, burst photography can be fragile, suffering from ghosting effect [28, 56] when capturing dynamic scenes in the presence of vehicles, humans, *etc*. An emerging alternative approach is to employ a neural network to automatically learn the mapping from a low-light noisy image to its long-exposure counterpart [9]. However, such a deep learning approach generally requires a large amount of labelled training data that resembles low-light photographs in the real world. Collecting rich high-quality training samples from diverse modern camera devices is tremendously labor-intensive and expensive.

In contrast, synthetic data is simple, abundant and inexpensive, but its efficacy is highly contingent upon how accurate the adopted noise formation model is. The heteroscedastic Gaussian noise model [22], instead of the commonly-used homoscedastic one, approximates well the real noise occurred in daylight or moderate low-light settings [5, 27, 28]. However, it cannot delineate the full picture of sensor noise under severely low illuminance. An illustrative example is shown in Fig. 1, where the objec-

tionable banding pattern artifacts, an unmodeled noise component that is exacerbated in dim environments, become clearly noticeable by human eyes.

In this paper, to avoid the effect on noise model from the image processing pipeline (ISP) [9, 5, 46] converting raw data to sRGB, we mainly focus on the noise formation model for raw images. We propose a physics-based noise formation model for extreme low-light raw denoising, which explicitly leverages the characteristics of CMOS photosensors to better match the physics of noise formation. As shown in Fig. 2, our proposed synthetic pipeline derives from the inherent process of electronic imaging by considering how photons go through several stages. It models sensor noise in a fine-grained manner that includes many noise sources such as photon shot noise, pixel circuit noise, and quantization noise. Besides, we provide a method to calibrate the noise parameters from available digital cameras. In order to investigate the generality of our noise model, we additionally introduce an extreme low-light denoising (ELD) dataset taken by various camera devices to evaluate our model. Extensive experiments show that the network trained only with the synthetic data from our noise model can reach the capability as if it had been trained with rich real data.

Our main contributions can be summarized as follows:

- We formulate a noise model to synthesize realistic noisy images that can match the quality of real data under extreme low-light conditions.

- We present a noise parameter calibration method that can adapt our model to a given camera.

- We collect a dataset with various camera devices to verify the effectiveness and generality of our model.

## 2. Related Work

Noise removal from a single image is an extensively-studied yet still unresolved problem in computer vision and image processing. Single image denoising methods generally rely on the assumption that both signal and noise exhibit particular statistical regularities such that they can be separated from a single observation. Crafting an analytical regularizer associated with image priors (*e.g.* smoothness, sparsity, self-similarity, low rank), therefore, plays a critical role in traditional design pipeline of denoising algorithms [52, 48, 18, 16, 43, 15, 6, 26]. In the modern era, most single image denoising algorithms are entirely data-driven, which consist of deep neural networks that implicitly learn the statistical regularities to infer clean images from their noisy counterparts [53, 12, 45, 61, 24, 57, 10, 27]. Although simple and powerful, these learning-based approaches are often trained on synthetic image data due to practical constraints. The most widely-used additive, white, Gaussian
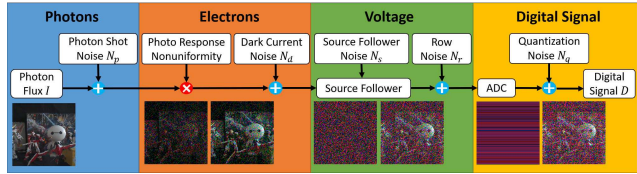


Figure 2: Overview of electronic imaging pipeline and visualization of noise sources and the resulting image at each stage.

noise model deviates strongly from realistic evaluation scenarios, resulting in significant performance declines on photographs with real noise [49, 2].

To step aside the domain gap between synthetic images and real photographs, some works have resorted to collecting paired real data not just for evaluation but for training [2, 9, 54, 8, 33]. Notwithstanding the promising results, collecting sufficient real data with ground-truth labels to prevent overfitting is exceedingly expensive and time-consuming. Recent works exploit the use of paired (Noise2Noise [38]) or single (Noise2Void [37]) noisy images as training data instead of paired noisy and noise-free images. However, they can not substantially ease the burden of labor requirements for capturing a massive amount of real-world training data.

Another line of research has focused on improving the realism of synthetic training data to circumvent the difficulties in acquiring real data from cameras. By considering both photon arrival statistics ("shot" noise) and sensor readout effects ("read" noise), the works of [46, 5] employed a signal-dependent heteroscedastic Gaussian model [22] to characterize the noise properties in raw sensor data. Most recently, Wang *et al.* [59] proposes a noise model, which considers the dynamic streak noise, color channel heterogeneous and clipping effect, to simulate the high-sensitivity noise on real low-light color images. Concurrently, a flow-based generative model, namely Noiseflow [1] is proposed to formulate the distribution of real noise using latent variables with tractable density[1]. However, these approaches oversimplify the modern sensor imaging pipeline, especially the noise sources caused by camera electronics, which have been extensively studied in the electronic imaging community [36, 29, 25, 3, 17, 19, 30, 31, 4, 58, 14]. In this work, we propose a physics-based noise formation model stemming from the essential process of electronic imaging to synthesize the noisy dataset and show that sizeable improvements of denoising performance on real data, particularly under extremely low illuminance.

---

[1]Note that Noiseflow requires paired real data to obtain noise data (by subtracting the ground truth images from the noisy ones) for training.

## 3. Physics-based Noise Formation Model

The creation of a digital sensor raw image $D$ can be generally formulated by a linear model

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

where $I$ is the number of photoelectrons that is proportional to the scene irradiation, $K$ represents the overall system gain composed by analog and digital gains, and $N$ denotes the summation of all noise sources physically caused by light or camera. We focus on the single raw image denoising problem under extreme low-light conditions. In this context, the characteristics of $N$ are formated in terms of the sensor physical process beyond the existing noise models. Deriving an optimal regularizer to tackle such noise is infeasible, as there is no analytical solver for such a noise distribution[2]. Therefore, we rely on a learning-based neural network pipeline to implicitly learn the regularities from data. Creating training samples for this task requires careful considerations of the characteristics of raw sensor data. In the following, we first describe the detailed procedures of the physical formation of a sensor raw image as well as the noise sources introduced during the whole process. An overview of this process is shown in Fig. 2.

### 3.1. Sensor Raw Image Formation

Our photosensor model is primarily based upon the CMOS sensor, which is the dominating imaging sensor nowadays [50]. We consider the electronic imaging pipeline of how incident light is converted from photons to electrons, from electrons to voltage, and finally from voltage to digital numbers, to model noise.

**From Photon to Electrons.** During exposure, incident lights in the form of photons hit the photosensor pixel area, which liberates photon-generated electrons (photoelectrons) proportional to the light intensity. Due to the quantum nature of light, there exists an inevitable uncertainty in the number of electrons collected. Such uncertainty imposes a Poisson distribution over this number of electrons, which follows

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

where $N_p$ is termed as the *photon shot noise* and $\mathcal{P}$ denotes the Poisson distribution. This type of noise depends on the light intensity, *i.e.*, on the signal. Shot noise is a fundamental limitation and cannot be avoided even for a perfect sensor. There are other noise sources introduced during the photon-to-electron stage, such as photo response nonuniformity and dark current noise, reported by many previous literatures [29, 25, 58, 3]. Over the last decade, technical advancements in CMOS sensor design and fabrication, *e.g.*, on-sensor dark current suppression, have led to

a new generation of digital single lens reflex (DSLR) cameras with lower dark current and better photo response uniformity [23, 41]. Therefore, we assume a constant photo response and absorb the effect of dark current noise $N_d$ into read noise $N_{read}$, which will be presented next.

**From Electrons to Voltage.** After electrons are collected at each site, they are typically integrated, amplified and read out as measurable charge or voltage at the end of exposure time. Noise present during the electrons-to-voltage stage depends on the circuit design and processing technology used, and thus is referred to as pixel circuit noise [25]. It includes thermal noise, reset noise [36], source follower noise [39] and banding pattern noise [25]. The physical origin of these noise components can be found in the electronic imaging literatures [36, 25, 58, 39]. For instance, source follower noise is attributed to the action of traps in silicon lattice which randomly capture and emit carriers; banding pattern noise is associated with the CMOS circuit readout pattern and the amplifier.

By leveraging this knowledge, we consider the thermal noise $N_t$, source follower noise $N_s$ and banding pattern noise $N_b$ in our model. The noise model of $N_b$ will be presented later. Here, we absorb multiple noise sources into a unified term, *i.e.* read noise

$$N_{read} = N_d + N_t + N_s. \qquad (3)$$

*Read noise* can be assumed to follow a Gaussian distribution, but the analysis of noise data (in Section 3.2) tells a long-tailed nature of its shape. This can be attributed by the flicker and random telegraph signal components of source follower noise [25], or the dark spikes raised by dark current [36]. Therefore, we propose using a statistical distribution that can better characterize the long-tail shape. Specifically, we model the read noise by a Tukey lambda distribution ($TL$) [34], which is a distributional family that can approximate a number of common distributions (*e.g.*, a heavy-tailed Cauchy distribution):

$$N_{read} \sim TL(\lambda; 0, \sigma_{TL}), \qquad (4)$$

where $\lambda$ and $\sigma$ indicate the shape and scale parameters respectively, while the location parameter is set to be zero given the zero-mean noise assumption.

Banding pattern noise $N_b$ appears in images as horizontal or vertical lines. We only consider the row noise component (horizontal stripes) in our model, as the column noise component (vertical stripes) is generally negligible when measuring the noise data (Section 3.2). We simulate the *row noise* $N_r$ by sampling a value from a zero-mean Gaussian distribution with a scale parameter $\sigma_r$, then adding it as an offset to the whole pixels within a single row.

**From Voltage to Digital Numbers.** To generate an image that can be stored in a digital storage medium, the analog voltage signal read out during last stage is quantized
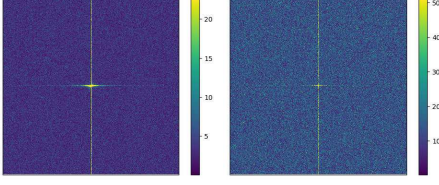
---

[2]Even if each noise component has an analytical formulation, their summation can generally be intractable.

Figure 3: Centralized Fourier spectrum of bias frames captured by SonyA7S2 (*left*) and (*right*) NikonD850 cameras

into discrete codes using an ADC. This process introduces *quantization noise* $N_q$ given by

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

where $U(\cdot, \cdot)$ denotes the uniform distribution over the range $[-1/2q, 1/2q]$ and $q$ is the quantization step.

**To summarize**, our noise formation model consists of four major noise components:

$$N = KN_p + N_{read} + N_r + N_q, \tag{6}$$

where $K$, $N_p$, $N_{read}$, $N_r$ and $N_q$ denotes the overall system gain, photon shot noise, read noise, row noise and quantization noise, respectively.

### 3.2. Sensor Noise Evaluation

In this section, we present a noise parameter calibration method attached to our proposed noise formation model. According to Eq. (2) (4) (6), the necessary parameters to specify our noise model include overall system gain $K$ for photon shot noise $N_p$; shape and scale parameters ($\lambda$ and $\sigma_{TL}$) for read noise $N_{read}$; scale parameter $\sigma_r$ for row noise $N_r$. Given a new camera, our noise calibration method consists of two main procedures, *i.e.* (1) estimating noise parameters at various ISO settings[3], and (2) modeling joint distributions of noise parameters.

**Estimating noise parameters.** We record two sequences of raw images to estimate $K$ and other noise parameters: *flat-field frames* and *bias frames*.

Flat-field frames are the images captured when sensor is uniformly illuminated. They can be used to derive $K$ according to the Photon Transfer method. [32]Once we have $K$, we can firstly convert a raw digital signal $D$ into the number of photoelectrons $I$, then impose a Poisson distribution on it, and finally revert it to $D$ – this simulates realistic photon shot noise.

Bias frames are the images captured under a lightless environment with the shortest exposure time. We took them at a dark room and the camera lens was capped on. Bias frames delineate the read noise picture independent of light, blended by the multiple noise sources aforementioned. The banding pattern noise can be tested via performing discrete Fourier transform on a bias frame. In Fig.3, the highlighted

---

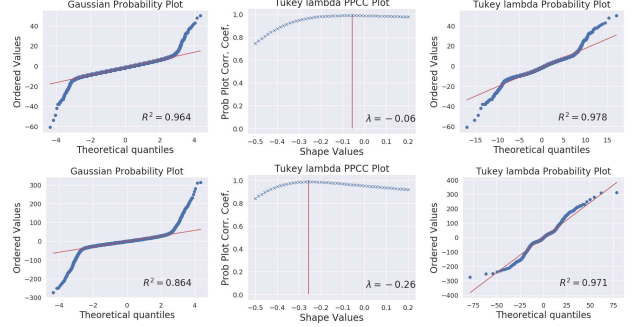[3]Noise parameters are generally stationary at a fixed ISO.



Figure 4: Distribution fitting of read noise for SonyA7S2 (*top*) and NikonD850 (*bottom*) cameras. *Left:* probability plot against the Gaussian distribution; *Middle:* Tukey lambda PPCC plot that determines the optimal $\lambda$ (shown in red line); *Right:* probability plot against the Tukey Lambda distribution. A higher $R^2$ indicates a better fit. (**Best viewed with zoom**)

vertical pattern in the centralized Fourier spectrum reveals the existence of row noise component. To analyze the distribution of row noise, we extract the mean values of each row from raw data. These values, therefore, serve as good estimates to the underlying row noise intensities, given the zero-mean nature of other noise sources. The normality of the row noise data is tested by a Shapiro-Wilk test [55]: the resulting $p$-value is higher than $0.05$, suggesting the null hypothesis that the data are normally distributed cannot be rejected. The related scale parameter $\sigma_r$ can be easily estimated by maximizing the log-likelihood.

After subtracting the estimated row noise from a bias frame, statistical models can be used to fit the empirical distribution of the residual read noise. A preliminary diagnosis (Fig. 4 Left) shows the main body of the data may follow a Gaussian distribution, but it also unveils the long-tail nature of the underlying distribution. In contrast to regarding extreme values as outliers, we observe an appropriate long-tail statistical distribution can characterize the noise data better.

We generate a probability plot correlation coefficient (PPCC) plot [20] to identify a statistical model from a Tukey lambda distributional family [34] that best describes the data. The Tukey lambda distribution is a family of distributions that can approximate many distributions by varying its shape parameter $\lambda$. It can approximate a Gaussian distribution if $\lambda = 0.14$, or derive a heavy-tailed distribution if $\lambda < 0.14$. The PPCC plot (Fig. 4 Middle) is used to find a good value of $\lambda$. The probability plot [60] (Fig. 4 Right) is then employed to estimate the scale parameter $\sigma_{TL}$. The goodness-of-fit can be evaluated by $R^2$ – the coefficient of determination *w.r.t.* the resulting probability plot [47]. The $R^2$ of the fitted Tukey Lambda distribution is much higher than the Gaussian distribution (*e.g.*, $0.972$ *vs.* $0.886$), indicating a much better fit to the empirical data.

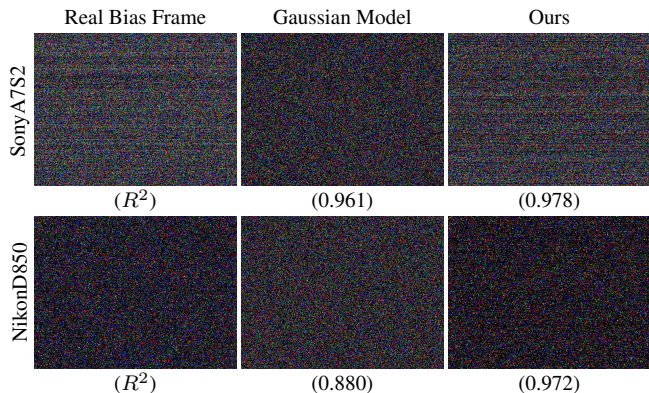Although we use a unified noise model for different cam-

Figure 5: Simulated and real bias frames of two cameras. A higher $R^2$ indicates a better fit quantitatively. **(Best viewed with zoom)**
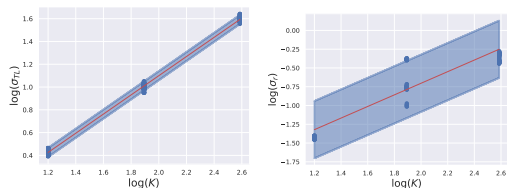


Figure 6: Linear least squares fitting from estimated noise parameter samples (blue dots) from a NikonD850 camera. Left and right figures show the joint distributions of $(K, \sigma_{TL})$ and $(K, \sigma_r)$ respectively, where we sample the noise parameters from the blue shadow regions.

eras, the noise parameters estimated from different cameras are highly diverse. Figure 4 shows the selected optimal shape parameter $\lambda$ differs camera by camera, implying distributions with varying degree of heavy tails across cameras. The visual comparisons of real and simulated bias frames are shown in Fig. 5. It shows that our model is capable of synthesizing realistic noise across various cameras, which outperforms the Gaussian noise model both in terms of the goodness-of-fit measure (*i.e.*, $R^2$) and the visual similarity to real noise.

**Modeling joint parameter distributions.** To choose noise parameters for our noise formation model, we infer the joint distributions of $(K, \sigma_{TL})$ and $(K, \sigma_r)$, from the parameter samples estimated at various ISO settings. As shown in Fig. 6, we use the linear least squares method to find the line of best fit for two sets of log-scaled measurements. Our noise parameter sampling procedure is

$$\log(K) \sim U\left(\log(\hat{K}_{min}), \log(\hat{K}_{max})\right),$$

$$\log(\sigma_{TL}) \,|\, \log(K) \sim \mathcal{N}\left(a_{TL}\log(K) + b_{TL}, \hat{\sigma}_{TL}\right), \quad (7)$$

$$\log(\sigma_r) \,|\, \log(K) \sim \mathcal{N}\left(a_r\log(K) + b_r, \hat{\sigma}_r\right),$$

where $U(\cdot, \cdot)$ denotes a uniform distribution and $\mathcal{N}(\mu, \sigma)$ denotes a Gaussian distribution with mean $\mu$ and standard



(a) Image capture setup     (b) example images

Figure 7: Capture setup and example images from our dataset.

Table 1: Quantitative Results on Sony set of the SID dataset. The noise models are indicated as follows. $G$: the Gaussian model for read noise $N_{read}$; $G^*$: the tukey lambda model for $N_{read}$; $P$: the Gaussian approximation for photon shot noise $N_p$; $P^*$: the true Poisson model for $N_p$; $R$: the Gaussian model for row noise $N_r$; $U$: the uniform distribution model for quantization noise $N_q$. The best results are indicated by red color and the second best results are denoted by blue color.

| Model | ×100 PSNR / SSIM | ×250 PSNR / SSIM | ×300 PSNR / SSIM |
|---|---|---|---|
| BM3D | 32.92 / 0.758 | 29.56 / 0.686 | 28.88 / 0.674 |
| A-BM3D | 33.79 / 0.743 | 27.24 / 0.518 | 26.52 / 0.558 |
| Paired real data | 38.60 / 0.912 | 37.08 / 0.886 | 36.29 / 0.874 |
| Noise2Noise | 37.42 / 0.853 | 33.48 / 0.725 | 32.37 / 0.686 |
| $G$ | 36.10 / 0.800 | 31.87 / 0.640 | 30.99 / 0.624 |
| $G+P$ | 37.08 / 0.839 | 32.85 / 0.697 | 31.87 / 0.665 |
| $G+P^*$ | 38.31 / 0.884 | 34.39 / 0.765 | 33.37 / 0.730 |
| $G^*+P^*$ | 39.10 / 0.911 | 36.46 / 0.869 | 35.69 / 0.855 |
| $G^*+P^*+R$ | 39.23 / 0.912 | 36.89 / 0.877 | 36.01 / 0.864 |
| $G^*+P^*+R+U$ | 39.27 / 0.914 | 37.13 / 0.883 | 36.30 / 0.872 |

deviation $\sigma$. $\hat{K}_{min}$ and $\hat{K}_{max}$ are the estimated overall system gains at the minimum and maximum ISO of a camera respectively. $a$ and $b$ indicate the fitted line's slope and intercept respectively. $\hat{\sigma}$ is an unbiased estimator of standard deviation of the linear regression under the Gaussian error assumption. For shape parameter $\lambda$, we simply sample it from the empirical distribution of the estimated parameter samples.

**Noisy image synthesis.** To synthesize noisy images, clean images are chosen and divided by low light factors sampled uniformly from $[100, 300]$ to simulate low photon count in the dark. Noise is then generated and added to the scaled clean samples, according to Eq. (6) (7). The created noisy images are finally normalized by multiplying the same low light factors to expose bright but excessively noisy contents.

## 4. Extreme Low-light Denoising (ELD) Dataset

To systematically study the generality of the proposed noise formation model, we collect an extreme low-light denoising (ELD) dataset that covers 10 indoor scenes and 4 camera devices from multiple brands (SonyA7S2, NikonD850, CanonEOS70D, CanonEOS700D). We also

record bias and flat field frames for each camera to calibrate our noise model. The data capture setup is shown in Fig. 7. For each scene and each camera, a reference image at the base ISO was firstly taken, followed by noisy images whose exposure time was deliberately decreased by low light factors $f$ to simulate extreme low light conditions. Another reference image then was taken akin to the first one, to ensure no accidental error (*e.g.* drastic illumination change or accidental camera/scene motion) occurred. We choose three ISO levels (800, 1600, 3200)[4] and two low light factors (100, 200) for noisy images to capture our dataset, resulting in 240 ($3\times2\times10\times4$) raw image pairs in total. The hardest example in our dataset resembles the image captured at a "pseudo" ISO up to 640000 ($3200\times200$).

## 5. Experiments

### 5.1. Experimental setting

**Implementation details.** A learning-based neural network pipeline is constructed to perform low-light raw denoising. We utilize the same U-Net architecture [51] as [9]. Raw Bayer images from SID Sony training dataset [9] are used to create training data. We pack the raw Bayer images into four channels (R-G-B-G) and crop non-overlapped $512 \times 512$ regions augmented by random flipping/rotation. Our approach only use the clean raw images, as the paired noisy images are generated by the proposed noise model on-the-fly. Besides, we also train networks based upon other training schemes as references, including training with paired real data (short exposure and long exposure counterpart) and training with paired real noisy images (*i.e.*, Noise2Noise [38]).

Our implementation[5] is based on PyTorch. We train the models with 200 epoch using $L_1$ loss and Adam optimizer [35] with batch size 1. The learning rate is initially set to $10^{-4}$, then halved at epoch 100, and finally reduced to $10^{-5}$ at epoch 180.

**Competing methods.** To understand how accurate our proposed noise model is, we compare our method with:

1. The approaches that use real noisy data for training, *i.e.* "paired real data" [9][6] and Noise2Noise [38];
2. Previous noise models, *i.e.* homoscedastic (G) and heteroscedastic Gaussian noise models (G+P) [22, 21];
3. The representative non-deep methods, *i.e.* BM3D [15] and Anscombe-BM3D (A-BM3D) [44][7].

---

[4]Most modern digital cameras are ISO-invariant when ISO is set higher than 3200 [13].

[5]Code is released at https://github.com/Vandermode/NoiseModel

[6][9] used paired real data to perform raw-to-sRGB low-light image processing. Here we adapt its setting to raw-to-raw denoising.

[7]The noise level parameters required are provided by the off-the-shelf image noise level estimators [22, 11].
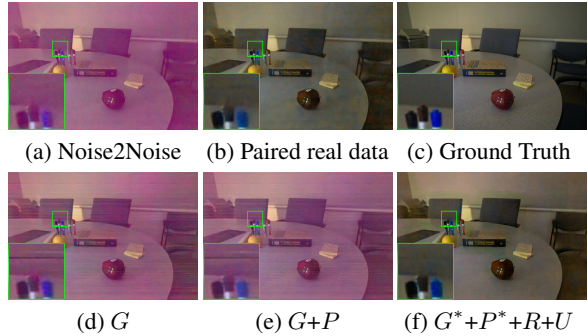


(a) Noise2Noise  (b) Paired real data  (c) Ground Truth

(d) $G$  (e) $G+P$  (f) $G^*+P^*+R+U$

Figure 8: Visual result comparison of different training schemes. Our final model ($G^*+P^*+R+U$) suppresses the "purple" color shift, residual bandings and chroma artifacts compared to other baselines.

### 5.2. Results on SID Sony dataset

Single image raw denoising experiment is firstly conducted on images from SID Sony validation and test sets. For quantitative evaluation, we focus on indoor scenes illuminated by natural lights, to avoid flickering effect of alternating current lights [2] [8]. To account for the imprecisions of shutter speed and analog gain [2], a single scalar is calculated and multiplied into the reconstructed image to minimize the mean square error evaluated by the ground truth.

**Ablation study on noise models.** To verify the efficacy of the proposed noise model, we compare the performance of networks trained with different noise models developed in Section 3.1. All noise parameters are calibrated using the ELD dataset, and sampled with a process following (or similar to) Eq. (7). The results of the other methods described in Section 5.1 are also presented as references.

As shown in Table 1, the domain gap is significant between the homoscedastic/heteroscedastic Gaussian models and the de facto noise model (characterized by the model trained with paired real data). *This can be attributed to (1) the Gaussian approximation of Possion distribution is not justified under extreme low illuminance; (2) horizontal bandings are not considered in the noise model; (3) long-tail nature of read noise is overlooked.* By taking all these factors into account, our final model, *i.e.* $G^*+P^*+R+U$ gives rise to a striking result: the result is comparable to or sometimes even better than the model trained with paired real data. Besides, training only with real low-light noisy data is not effective enough, due to the clipping effects (that violates the zero-mean noise assumption) and the large variance of corruptions (that leads to a large variance of the Noise2Noise solution) [38]. A visual comparison of our final model and other methods is presented in Fig. 8, which

---

[8]Alternating current light is not noise, but a type of illumination that breaks the irradiance constancy between short/long exposure pairs, making the quantitative evaluation inaccurate.

Table 2: Quantitative results (PSNR/SSIM) of different methods on our ELD dataset containing four representative cameras.

| Camera | $f$ | Index | Non-deep | | Training with real data | | Training with synthetic data | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | BM3D [15] | A-BM3D [44] | Paired data [9] | Noise2Noise [38] | G | G+P [22] | Ours |
| SonyA7S2 | ×100 | PSNR | 37.69 | 37.74 | 44.50 | 41.63 | 42.35 | 42.46 | 45.36 |
| | | SSIM | 0.803 | 0.776 | 0.971 | 0.856 | 0.893 | 0.889 | 0.972 |
| | ×200 | PSNR | 34.06 | 35.26 | 42.45 | 37.98 | 38.93 | 38.88 | 43.27 |
| | | SSIM | 0.696 | 0.721 | 0.945 | 0.775 | 0.813 | 0.812 | 0.949 |
| NikonD850 | ×100 | PSNR | 33.97 | 36.60 | 41.28 | 40.47 | 39.57 | 40.29 | 41.79 |
| | | SSIM | 0.725 | 0.779 | 0.938 | 0.848 | 0.823 | 0.845 | 0.912 |
| | ×200 | PSNR | 31.36 | 32.59 | 39.44 | 37.98 | 36.68 | 37.26 | 39.69 |
| | | SSIM | 0.618 | 0.723 | 0.910 | 0.820 | 0.757 | 0.786 | 0.875 |
| CanonEOS70D | ×100 | PSNR | 30.79 | 31.88 | 40.10 | 38.21 | 40.59 | 40.94 | 40.62 |
| | | SSIM | 0.589 | 0.692 | 0.931 | 0.826 | 0.925 | 0.934 | 0.937 |
| | ×200 | PSNR | 28.06 | 28.66 | 37.32 | 34.33 | 37.49 | 37.64 | 38.17 |
| | | SSIM | 0.540 | 0.597 | 0.867 | 0.704 | 0.871 | 0.873 | 0.890 |
| CanonEOS700D | ×100 | PSNR | 29.70 | 30.13 | 39.05 | 38.29 | 39.77 | 40.08 | 39.84 |
| | | SSIM | 0.556 | 0.640 | 0.906 | 0.859 | 0.884 | 0.897 | 0.921 |
| | ×200 | PSNR | 27.52 | 27.68 | 36.50 | 34.94 | 37.67 | 37.86 | 37.59 |
| | | SSIM | 0.537 | 0.579 | 0.850 | 0.766 | 0.870 | 0.879 | 0.879 |
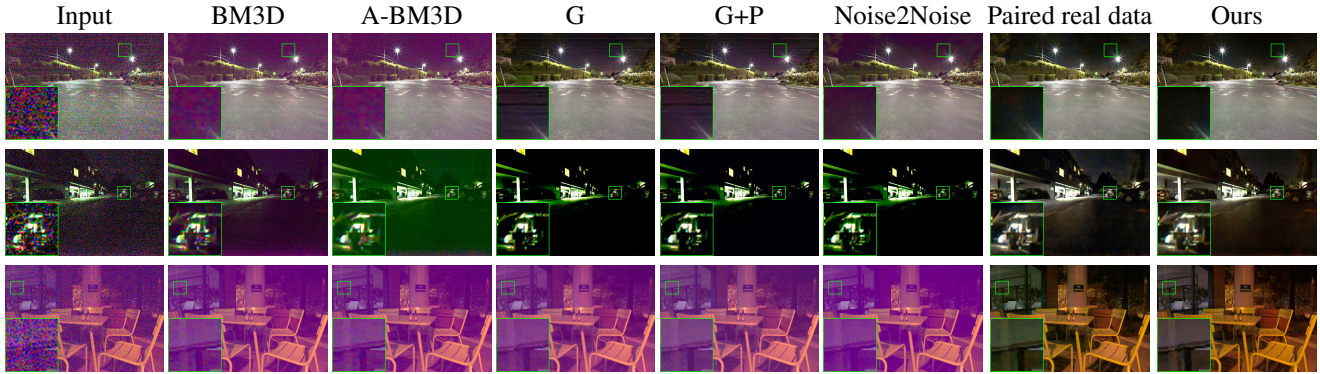


Figure 9: Raw image denoising results on both indoor and outdoor scenes from SID Sony dataset. (**Best viewed with zoom**)
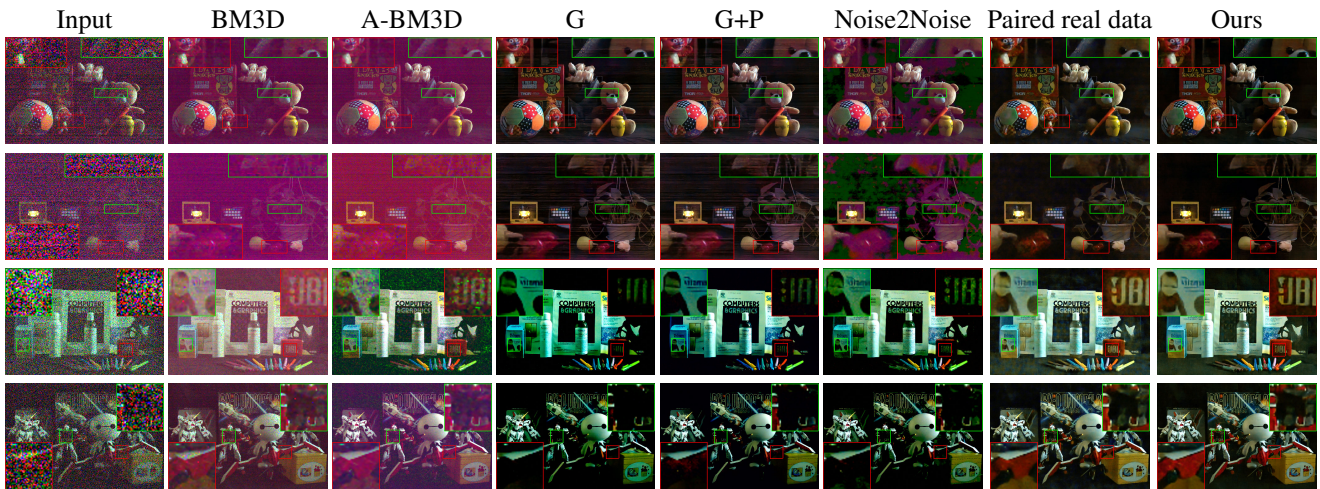


Figure 10: Raw image denoising results on our ELD dataset. (**Best viewed with zoom**)

shows the effectiveness of our noise formation model.

Though we only quantitatively evaluate the results on indoor scenes of the SID Sony set, our method can be applied

to outdoor scenes as well. The visual comparisons of both indoor and outdoor scenes from SID Sony set are presented in Fig. 9. It can be seen that the random noise can be sup-
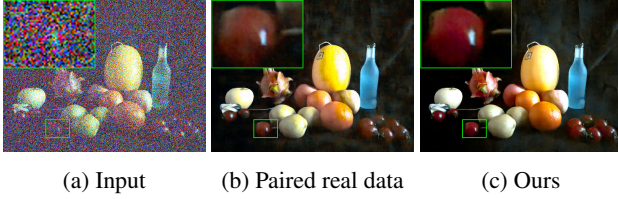
(a) Input     (b) Paired real data     (c) Ours

Figure 11: Denoising results of a low-light image captured by a Huawei Honor 10 camera.



(a) SID only     (b) SID + MIT5K     (c) Ground Truth

Figure 13: Denoising results of a low-light image captured by a NikonD850 camera.
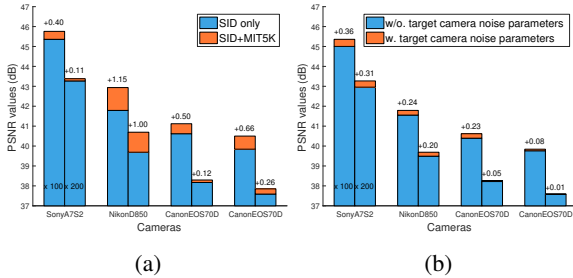


(a)               (b)

Figure 12: (a) Performance boost when training with more synthesized data. (b) Noise parameter sensitivity test.

pressed by the model learned with heteroscedastic Gaussian noise (G+P) [22], but the resulting colors are distorted, the banding artifacts become conspicuous, and the image details are barely discernible. By contrast, our model produces visually appealing results as if it had been trained with paired real data.

## 5.3. Results on our ELD dataset

**Method comparisons.** To see whether our noise model can be applicable to other camera devices as well, we assess model performance on our ELD dataset. Table 2 and Fig. 10 summarize the results of all competing methods. It can be seen that the non-deep denoising methods, *i.e.* BM3D and A-BM3D, fail to address the banding residuals, the color bias and the extreme values presented in the noisy input, whereas our model recovers vivid image details which can be hardly perceived on the noisy image by human observers. *Moreover, our model trained with synthetic data even often outperforms the model trained with paired real data.* We note the finding here conforms with the evaluation of sensor noise presented in Section 3.2, especially in Fig. 4 and 5, where we show the underlying noise distribution varies camera by camera. Consequently, training with paired real data from SID Sony camera inevitably overfits to the noise pattern merely existed on the Sony camera, leading to sub-optimal results on other types of cameras. In contrast, our model relies on a very flexible noise model and a noise calibration process, making it adapts to noise characteristics of other (calibrated) camera models as well. Additional evidence can be found in Fig. 11, where we apply these two models to an image captured by a smartphone camera. Our reconstructed image is clearer and cleaner than what is re-

stored by the model trained with paired real data.

**Training with more synthesized data.** A useful merit of our approach against the conventional training with paired real data, is that our model can be easily incorporated with more real clean samples to train. Fig. 12(a) shows the relative improvements of our model when training with the dataset synthesized by additional clean raw images from MIT5K dataset [7]. We find the major improvements, as shown in Fig. 13, are owing to the more accurate color and brightness restoration. By training with more raw image samples from diverse cameras, the network learns to infer picture appearances more naturally and precisely.

**Sensitivity to noise calibration.** Another benefit of our approach is we only need clean samples and a noise calibration process to adapt to a new camera, in contrast to capturing real noisy images accompanied with densely-labeled ground truth. Besides, the noise calibration process can be simplified once we already have a collection of parameter samples from various cameras. Fig. 12(b) shows models can reach comparable performance on target cameras without noise calibration, by simply sampling parameters from other three calibrated cameras instead.

## 6. Conclusion

We have presented a physics-based noise formation model together with a noise parameter calibration method to help resolve the difficulty of extreme low-light denoising. We revisit the electronic imaging pipeline and investigate the influential noise sources overlooked by existing noise models. This enables us to synthesize realistic noisy raw data that better match the underlying physical process of noise formation. We systematically study the efficacy of our noise formation model by introducing a new dataset that covers four representative camera devices. By training only with our synthetic data, we demonstrate a convolutional neural network can compete with or sometimes even outperform the network trained with paired real data.

# References

[1] Abdelrahman Abdelhamed, Marcus A. Brubaker, and Michael S. Brown. Noise flow: Noise modeling with conditional normalizing flows. In *The IEEE International Conference on Computer Vision (ICCV)*, 2019.

[2] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[3] Richard L. Baer. A model for dark current characterization and simulation. *Proceedings of SPIE - The International Society for Optical Engineering*, 6068:37–48, 2006.

[4] Robert A. Boie and Ingemar J. Cox. An analysis of camera noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6):671–674, 1992.

[5] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11036–11045, 2019.

[6] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[7] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[8] Chen Chen, Qifeng Chen, Minh N. Do, and Vladlen Koltun. Seeing motion in the dark. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[9] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[10] Chang Chen, Zhiwei Xiong, Xinmei Tian, and Feng Wu. Deep boosting for image denoising. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[11] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[12] Yunjin Chen, Wei Yu, and Thomas Pock. On learning optimized reaction diffusion processes for effective image restoration. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[13] Roger N. Clark. Exposure and digital cameras, part 1: What is iso on a digital camera? when is a camera isoless? iso myths and digital cameras. http://www.clarkvision.com/articles/iso/, 2012.

[14] Roberto Costantini and Sabine Susstrunk. Virtual sensor design. *Proceedings of SPIE - The International Society for Optical Engineering*, 5301:408–419, 2004.

[15] 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.

[16] Weisheng Dong, Xin Li, Lei Zhang, and Guangming Shi. Sparsity-based image denoising via dictionary learning and structural clustering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 457–464. IEEE, 2011.

[17] Abbas El Gamal and Helmy Eltoukhy. Cmos image sensors. *IEEE Circuits and Devices Magazine*, 21(3):6–20, 2005.

[18] 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.

[19] Joyce Farrell and Manu Parmar. Sensor calibration and simulation. *Proceedings of SPIE - The International Society for Optical Engineering*, 2008.

[20] James J. Filliben. The probability plot correlation coefficient test for normality. *Technometrics*, 17(1):111–117, 1975.

[21] Alessandro Foi. Clipped noisy images: Heteroskedastic modeling and practical denoising. *Signal Processing*, 89(12):2609–2629, 2009.

[22] 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.

[23] Eric R. Fossum and Donald B. Hondongwa. A review of the pinned photodiode for ccd and cmos image sensors. *IEEE Journal of the Electron Devices Society*, 2(3):33–43, 2014.

[24] Michaël Gharbi, Gaurav Chaurasia, Sylvain Paris, and Frédo Durand. Deep joint demosaicking and denoising. *ACM Transactions on Graphics*, 35(6):191:1–191:12, Nov. 2016.

[25] 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.

[26] Shuhang Gu, Zhang Lei, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[27] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[28] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics*, 35(6):192, 2016.

[29] 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.

[30] Kenji Irie, Alan E. Mckinnon, Keith Unsworth, and Ian M. Woodhead. A model for measurement of noise in ccd digital-video cameras. *Measurement Science and Technology*, 19(4):334–340, 2008.

[31] Kenji Irie, Alan E. Mckinnon, Keith Unsworth, and Ian M. Woodhead. A technique for evaluation of ccd video-camera noise. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(2):280–284, 2008.

[32] James Janesick, Kenneth Klaasen, and Tom Elliott. Ccd charge collection efficiency and the photon transfer technique. *Proceedings of SPIE - The International Society for Optical Engineering*, 570:7–19, 1985.

[33] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

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

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

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

[37] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void - learning denoising from single noisy images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[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 (ICML)*, pages 2971–2980, 2018.

[39] Cedric Leyris, Alain Hoffmann, Matteo Valenza, J.-C. Vildeuil, and F. Roy. Trap competition inducing r.t.s noise in saturation range in n-mosfets. *Proceedings of SPIE - The International Society for Optical Engineering*, 5844:41–51, 2005.

[40] Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T Barron, Dillon Sharlet, Ryan Geiss, et al. Handheld mobile photography in very low light. *ACM Transactions on Graphics (TOG)*, 38(6):1–16, 2019.

[41] Wensheng Lin, Guoming Sung, and Jyunlong Lin. High performance cmos light detector with dark current suppression in variable-temperature systems. *Sensors*, 17(1):15, 2016.

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

[43] Julien Mairal, Michael Elad, and Guillermo Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, 2008.

[44] Markku Makitalo and Alessandro Foi. Optimal inversion of the anscombe transformation in low-count poisson image denoising. *IEEE Transactions on Image Processing*, 20(1):99–109, 2011.

[45] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2802–2810, 2016.

[46] Ben Mildenhall, Jonathan T. Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[47] Eugene C. Morgan, Matthew Lackner, Richard M. Vogel, and Laurie G. Baise. Probability distributions for offshore wind speeds. *Energy Conversion and Management*, 52(1):15 – 26, 2011.

[48] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling and Simulation*, 4(2):460–489, 2005.

[49] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[50] Grand View Research. Image sensors market analysis, 2016. [online]. http://www.grandviewresearch.com/industry-analysis/imagesensors-market, 2016.

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

[52] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms . *Physica D Nonlinear Phenomena*, 60(14):259–268, 1992.

[53] Uwe Schmidt and Stefan Roth. Shrinkage fields for effective image restoration. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[54] Eli Schwartz, Raja Giryes, and Alex M. Bronstein. Deep-isp: Learning end-to-end image processing pipeline. *IEEE Transactions on Image Processing*, PP(99):1–1, 2018.

[55] S. S. Shapiro and R. S. Francia. An approximate analysis of variance test for normality. *Biometrika*, 67(337):215–216, 1975.

[56] Ziyi Shen, Wenguan Wang, Xiankai Lu, Jianbing Shen, Haibin Ling, Tingfa Xu, and Ling Shao. Human-aware motion deblurring. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[57] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[58] Hans Wach and Edward R. Dowski Jr. Noise modeling for design and simulation of computational imaging systems. *Proceedings of SPIE - The International Society for Optical Engineering*, 5438:159–170, 2004.

[59] Wei Wang, Xin Chen, Cheng Yang, Xiang Li, Xuemei Hu, and Tao Yue. Enhancing low light videos by exploring high sensitivity camera noise. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

[60] Martin B. Wilk and Ram Gnanadesikan. Probability plotting methods for the analysis of data. *Biometrika*, 55(1):1–17, 1968.

[61] 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*, 2017.