

Non-parametric Sensor Noise Modeling and Synthesis

Ali Mosleh¹, Luxi Zhao¹, Atin Singh², Jaeduk Han², Abhijith Punnappurath¹,
Marcus A. Brubaker^{1,3}, Jihwan Choe², and Michael S. Brown¹

¹ Samsung AI Center Toronto, Canada

² Samsung Electronics, South Korea

³ York University, Canada

Abstract. We introduce a novel non-parametric sensor noise model that directly constructs probability mass functions per intensity level from captured images. We show that our noise model provides a more accurate fit to real sensor noise than existing models. We detail the capture procedure for deriving our non-parametric noise model and introduce an interpolation method that reduces the number of ISOs levels that need to be captured. In addition, we propose a method to synthesize noise on existing noisy images when noise-free images are not available. Our noise model is straightforward to calibrate and provides notable improvements over competing noise models on downstream tasks.

Keywords: Sensor Noise · Non-parametric · Model Interpolation

1 Introduction

Modern cameras have improved significantly, but smaller device sizes and higher pixel densities make raw sensor readings more susceptible to noise. Accurate sensor noise models are essential for synthesizing noise on training images for deep neural networks (DNNs) targeting low-level vision tasks (*e.g.*, [3, 26, 36]). Existing noise modeling methods can be categorized into physics-based and DNN-based noise models.

Physics-based methods utilize statistical modeling to represent the distribution of different noise sources. These methods take into account the fact that sensor noise is linked to the physics of light and therefore employ Gaussian-Poisson models [12, 13, 15, 16, 33]. However, it is difficult to precisely model all noise sources that stem from variations in circuit design and signal processing techniques. More recently, DNN-based generative models have been used to learn noise synthesis from real captured data [1, 6, 8, 19]. Such models offer powerful representation capabilities. However, their performance often depends on extensive image capturing of a wide variety of scenes. This inspires combining DNN-based noise models with physics-based ones [11, 37].

We introduce a non-parametric method to model raw sensor noise. Our noise model is based on statistics derived from the image formation process, similar to physics-based models. However, unlike traditional models, our approach does

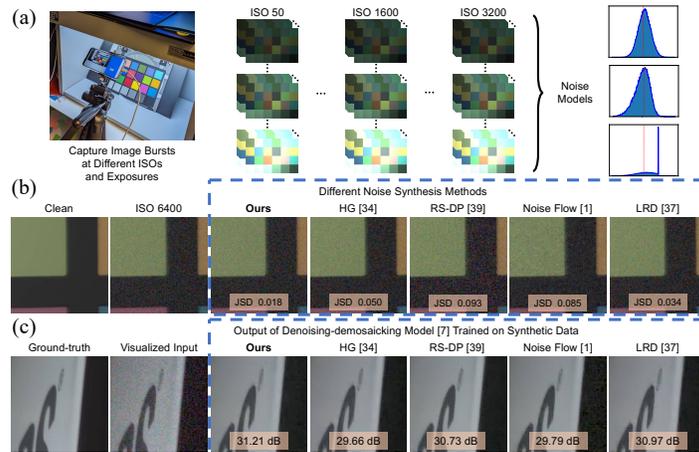


Fig. 1: (a) Our calibration setup and associated image capture. Our approach constructs probability mass functions for discrete intensity levels. (b) An ISO 6400 noise synthesis comparison is shown for a Pixel 6 camera. (c) We demonstrate improved raw denoising-demaicing [7] using our non-parametric noise model over competing ones.

not rely on the typical assumptions about different noise components introduced during the process. Instead, we base our method on the observed distribution of noise at each pixel intensity level. We propose a systematic calibration technique to collect a large sample set of noise. This is achieved by capturing a burst of images of a chart with uniform patches under controlled illumination at a fixed ISO with various exposure values (Fig. 1). The collected noise samples are used to build a probability mass function (PMF) per intensity level, which is then used in an inversion sampling process to synthesize noise.

In real-world applications, noise synthesis needs to cover a wide range of ISO levels. To address such a dense noise calibration more efficiently, we introduce an interpolation procedure that approximates noise PMFs for an arbitrary uncalibrated ISO, given a set of calibrated ISO levels. We show that the high accuracy achieved for the proposed noise model interpolation method is attributed to the accuracy of the underlying proposed non-parametric noise model.

We also examine the impact of a common assumption that is used in existing noise augmentation methods. This assumption presupposes that the original image that noise is added to is noiseless. However, in reality, most images are already contaminated with noise, which can create a domain gap when used for noise synthesis. Obtaining clean raw data may not always be feasible for further training image augmentation due to the difficulty of acquiring it or because the ground-truth training data is produced using black-box and often irreversible image rendering processes [4, 17, 27]. In such cases, clean raw images may not be available. To address this, we propose a solution to transform a pair of source and target noise models into a single model that allows us to accurately synthesize the target noise on existing noisy image (source) data for further training.

Contribution We describe our non-parametric sensor noise model and outline a straightforward calibration procedure to build noise PMFs directly from observations. The noise model is the result of sufficiently collected noise samples to build such PMFs without assumptions regarding the distributions of signal-dependent and signal-independent noise. We later show that obtaining noise models in terms of accurate PMFs helps in resolving two main issues regarding noise modeling: (1) how to infer noise models for uncalibrated ISOs, and (2) how to synthesize noise on already contaminated noisy captures. We show that these issues are best addressed with our non-parametric models with explicit noise PMFs, built accurately from sufficiently large noise sample sets. We demonstrate the effectiveness of our non-parametric noise model by synthesizing training data for a joint denoising-demosaicking network, resulting in significant improvements in PSNR and SSIM evaluated for Samsung S22+ and Google Pixel 6 cameras. We also investigate the approach of synthesizing noise on noisy captures for data augmentation for a raw to sRGB rendering network evaluated for a Samsung S20FE mobile camera.

2 Related Work

Physics-based noise models. The additive white Gaussian noise (AWGN) model is the simplest and arguably the most commonly used physics-based noise model [38]. However, raw sensor noise is signal dependent, and a significant performance drop occurs when AWGN is considered in image reconstruction solutions [2, 25]. The Poisson-Gaussian (PG) noise model [13, 22] was proposed to account for the signal-dependent noise by modeling the shot and read noise components separately using Poisson and Gaussian distributions, respectively. However, replacing the Poisson distribution with a Gaussian distribution having a signal-dependent variance, known as the heteroscedastic Gaussian (HG) model [13, 21, 34], is more widely utilized in practice [3]. A major deviation of the HG model from real noise arises from clipped image intensities, and there have been attempts to account for this effect using a clipped HG model [12]. Recently, more complex noise models have been proposed that carefully examine the sensor processing pipeline and different noise sources [31, 33]. For example, the approach in [33] models shot noise using a Poisson distribution, read noise using a Tukey lambda distribution, row noise using a Gaussian distribution, and quantization noise using a uniform distribution. Note that all these methods are parametric, making them simple and interpretable, whose parameters can be estimated with relatively less amount of data. These models are also typically easy to interpolate to unknown settings (*e.g.*, different ISOs). However, the underlying noise distribution is unknown, and in many cases, complex and simple parametric models both have obvious shortcomings. To address such limitations, the approach presented in [39] employs a Poisson distribution to model signal-dependent noise, while the signal-independent noise is directly sampled from a database of dark frames. Our proposed non-parametric model takes a

different approach and samples from the observed noise distributions for both signal-dependent and signal-independent noise components in a joint manner.

DNN-based noise models. Early work in this area used generative adversarial networks (GANs) to model the noise distribution [6, 8, 18, 19, 30, 35]. However, GANs do not have tractable likelihoods, making the quality of the synthesized noise difficult to assess [23]. More recent methods are based on normalizing flows [1, 23]. Noise Flow [1] employs a conditional normalizing flow for different noise components in the sensor pipeline and synthesizes realistic noisy raw images. Noise2NoiseFlow [23] further extends Noise Flow to relax the requirement of noise-free images for training. However, recent studies [24, 39] indicate that calibrated physics-based models still outperform DNN-based noise models. A few recent works have attempted to combine the strengths of physics-based approaches with DNN methods. A physics-inspired GAN noise model was introduced in [24]. After initial noise is added to a clean image using the physics-inspired parameters, the intermediate noisy image is passed through a convolutional neural network (CNN) that adds a residual noise layer to account for any effects that were not captured by the physics-inspired noise model. The low-light raw denoising (LRD) model in [37] uses a physics-based model to generate signal-dependent noise while signal-independent noise is synthesized using a GAN. A physics-guided noise model is employed by [5], where different noise components are handled by separate normalizing flow models. While DNN-based models demonstrate excellent performance, we show that our non-parametric model synthesizes noise that is statistically more similar to real sensor noise.

3 Image Formation

Image formation in cameras follows the typical processes shown in Fig. 2. The scene radiance modulated by the camera optics appears on the image sensor (*e.g.*, CCD or CMOS), where photons go through a microlens array to improve light collection. Next, the incident light passes through a color filter array (CFA), forming a mosaic of the three RGB stimuli. Finally, photodiodes of the sensor collect the color-filtered light and output a digital single-channel raw image. At this photodiodes layer, a potential well counts photons arriving at the sensor area A during the exposure time t and converts the accumulated photons into electrons. This conversion is affected by the quantum efficiency of the CFA and the detector—that is, η . Then, the electrons are amplified by a gain factor g , controllable by the ISO level of the camera. Finally, an analog-to-digital converter (ADC) converts the modulated electrons to digits.

During the process, electron noise is generated from different sources. Since the gain factor affects the noise distribution, it is crucial to divide the accumulated noise into gained noise as n_g , and read-out noise as n_r . The gained noise n_g mostly includes dark noise, dark current, and fixed pattern noise, while the read-out noise n_r is dominated by thermal noise. Quantization errors n_q of the ADC are also added at the stage before saving the raw image. The quantum nature of light also affects the uncertainty in the collected photons. The number of

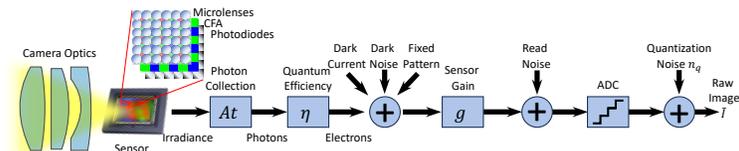


Fig. 2: From the modulation of the scene radiance on the surface of the sensor to the final digital storage, different types of noise are introduced. Some additive noise is photon-related, and some is independent of the incident light. The assumptions regarding different noise components and their distributions can vary for different sensors.

incident photons and the relevant photon noise follow a Poisson distribution with an expected value of μ_p . Thus, the image formation process can be expressed as

$$\tilde{I} = g\left(\eta\mu_p + \eta n_p(\mu_p) + n_g\right) + n_r + n_q, \quad (1)$$

where \tilde{I} is the observed raw pixel intensity, and $n_p(\mu_p)$ denotes the photon noise which depends on the expected number of incident photons μ_p . Considering photo-electrons $I=g\eta\mu_p$ as the clean underlying intensity transforms (1) to

$$\tilde{I} = I + N(I), \quad N(I) = g\eta n_p(I/g\eta) + gn_g + n_r + n_q, \quad (2)$$

where $N(I)$ denotes the overall signal-dependent noise.

4 Non-parametric Noise Modeling

The image formation model in (1) holds for a variety of different sensor architectures. However, underlying assumptions for the sensor-specific parameters and the distributions of noise components can vary significantly from one sensor to another [10, 14, 20, 29]. Thus, we propose a non-parametric sensor noise model by collecting a sufficiently large number of noise samples and calibrating the overall signal-dependent noise $N(I)$ introduced in (2) rather than modeling noise components individually. Since noise varies for different CFA channels, and is affected by the gain factor, we model it per color channel and per ISO level.

Sensor noise calibration. We make use of raw images captured at specialized controllable settings—namely, capturing innumerable photon flux densities for scenes that contain uniform patches illuminated by a DC-light (Fig. 1a). We first collect a burst of M images of a color checker at a low ISO level denoted by $\bar{\kappa}$ as $\{\tilde{I}_{\bar{\kappa}}^1, \dots, \tilde{I}_{\bar{\kappa}}^M\}$. Assuming that noise follows a zero mean distribution in low ISO captures, the underlying clean signal can be defined as the expected value of the noisy observations. Thus, the low ISO captures are averaged to form a clean image as $I = \text{Mean}(\{\tilde{I}_{\bar{\kappa}}^1, \dots, \tilde{I}_{\bar{\kappa}}^M\})$. Of note, capturing uniform patches rather than natural scenes resolves the slight misalignment issues caused by vibrations during long-exposure photography to generate clean images. We then collect another burst of color checker images denoted by $\{\tilde{I}_{\kappa}^1, \dots, \tilde{I}_{\kappa}^M\}$ at an ISO level

κ which we intend to calibrate. These noisy bursts are captured so that they are spatially aligned with the clean image. Fig. 1a shows examples of clean and noisy bursts extracted from the uniform patches of a color checker. We inspect I for each intensity level and collect corresponding pixels from the noisy bursts. Hence, a noise sample set per intensity level l per ISO level κ is formed as

$$\xi_{\kappa}^l = \left\{ \tilde{I}_{\kappa}^j(i) - l \mid i \in \{1, \dots, H \times W\}, I(i) = l, j \in \{1, \dots, M\} \right\},$$

where $H \times W$ denotes image size, i denotes pixel index, and l denotes sensor intensity level ranging from zero to sensor’s white-level L *i.e.*, $l \in \{0, \dots, L\}$. In order to collect a reasonably large sample set for each intensity level representing various photon flux densities ($\propto At$ as in Fig. 2), we collect burst images from a sweep of various exposure times. This results in extremely dark pixels to almost fully saturated intensities, as shown in Fig. 1a. Our controlled imaging procedure helps automate the process of capturing thousands of images straightforwardly.

The histogram of each noise set ξ_{κ}^l is used to form a PMF $p_{\xi_{\kappa}^l}(n)$ where $n \in \mathbb{R}$ denotes noise value. The PMFs obtained for all the intensity levels of the sensor make our ISO-specific non-parametric noise model as $\{p_{\xi_{\kappa}^0}(n), \dots, p_{\xi_{\kappa}^L}(n)\}$. Since color-filtered channels of sensors have different sensitivities to the incident light, noise modeling is performed per color channel of the sensor.

Sensor noise synthesis. Given a clean image and calibrated noise models, noise is synthesized via inversion sampling [9]. An inversion sampling function denoted by `RandomSampling(\cdot)` generates random samples from a noise distribution through inverting its cumulative distribution functions (CDFs). For every pixel i in the given clean image I , we synthesize the noisy pixel $\tilde{I}(i)$ as

$$\tilde{I}(i) = I(i) + N(I(i)), \quad N(I(i)) \leftarrow \text{RandomSampling}\left(p_{\xi_{\kappa}^{I(i)}}(n)\right). \quad (3)$$

Despite its contribution to a more generalized inference and a better task performance, synthesizing realistic noise as an augmentation strategy for developing camera/sensor-specific application models has two main limitations:

1. Sensor noise modeling is often performed for a few ISO settings—for example, a small set of nominal ISOs as $\{2^m 100 \mid m \in \mathbb{N}, 0 \leq m \leq 6\}$. However, in ubiquitous cameras, the additive system of photographic exposure (APEX) uses a wide range of ISOs to calculate exposure. Thus, realistic applications require noise models for a much larger number of ISO settings. This poses a significant challenge as an accurate noise model per ISO level requires numerous captures, regardless of whether we are using DNN-based or physics-based approaches.

2. Existing noise synthesis methods for generating training data need noise-free raw images to apply Eq. (2). However, in many image restoration applications, ground-truth (clean) data is generated using approaches other than long-exposure photography. For instance, using a secondary high-end DSLR geometrically aligned with the main camera to capture ground-truth data is a common approach [17], especially if the task has to deal with blur effects [27]. Or, for full raw-sRGB rendering applications, using black-box rendering tools like Adobe Photoshop [4] is one approach to obtain ground-truth data. In such cases, we do not have access to clean raw data for noise synthesis. On the other hand,

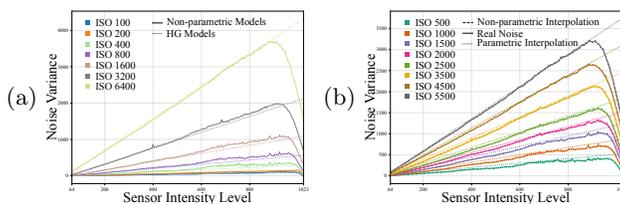


Fig. 3: Plots of noise variance per sensor intensity level for various ISOs of Pixel 6. (a) Variance of noise generated using our calibrated non-parametric models and HG noise models. The calibrated ISOs in (a) are used to interpolate noise models for ISOs shown in (b). The solid-line, dashed-line, and dotted-line plots in (b) correspond to the variance of real noise, noise generated using our interpolated models, and parametric interpolation [34] for the HG noise models, respectively.

such ground-truth data are the results of many non-linear operations. Thus, approximating raw through an inversion process like [3, 36] poses an unavoidable domain gap in the synthesized data. All being said, using a noise model that allows augmenting some calibrated noise on top of existing noisy raw images such that the augmented noise follows sensor characterizations would be valuable.

As follows, we address the aforementioned challenges. We first develop an interpolation method to infer noise distributions for uncalibrated ISOs. Next, we discuss how to carefully apply calibrated noise models to existing noisy captures to augment training datasets with new ISOs.

4.1 Noise Model Interpolation

Under certain conditions and using some calibration data, it is possible to approximate the variance of read-out noise and gained noise in (1) for uncalibrated ISOs. Such a parameterized approach was adopted in [34] for the HG model via solving a system of linear equations to estimate noise parameters. We observe a large discrepancy between synthesized noise using such parametric interpolations and actual noise, arguably since some parameters, like g and η in (1), have to be accurately calibrated prior to noise modeling. It is shown using noise variance plots in Fig. 3b obtained via an implementation of the parametric approach [34] for uncalibrated ISOs. Note the gap between the dotted and the solid-line plots.

Our noise modeling approach discussed in Sec. 4 is based on the actual sensor noise measurements. We benefit from the calibrated noise statistics and propose an interpolation approach to estimate uncalibrated PMFs as follows.

We assume that noise distributions per intensity among different ISO levels have similar-shaped PMFs but differ significantly in their scales and means. This assumption allows us to approximate noise distributions for uncalibrated ISO levels through interpolating among calibrated noise variances and means. Let κ denote a calibrated ISO from the set of calibrated ISO levels; $\{\kappa_{\min}, \dots, \kappa_{\max}\}$, for each κ , per intensity level l , we have the following statistics

$$\sigma_{\xi_{\kappa}^l}^2 = \text{Var}(\xi_{\kappa}^l), \mu_{\xi_{\kappa}^l} = \text{Mean}(\xi_{\kappa}^l), p_{\xi_{\kappa}^l}(n), \quad (4)$$

where $\text{Var}(\cdot)$ and $\text{Mean}(\cdot)$ denote the variance and mean of the sample set, respectively. The normalized noise distribution, whose mean and variance are 0 and 1, respectively, can follow a PMF as

$$p_{\xi_{\kappa}^l}(n) = \sigma_{\xi_{\kappa}^l} p_{\xi_{\kappa}^l}(n\sigma_{\xi_{\kappa}^l} + \mu_{\xi_{\kappa}^l}). \quad (5)$$

We assume that this normalized PMF has a similar characteristic function among calibrated ISO levels for each l , *i.e.*, $p_{\xi_{\kappa_{\min}}^l}(n) \approx \dots \approx p_{\xi_{\kappa_{\max}}^l}(n)$. Hence, for each set of calibrated noise distributions, we fit curves to the measured variances $\{\sigma_{\xi_{\kappa_{\min}}^l}^2, \dots, \sigma_{\xi_{\kappa_{\max}}^l}^2\}$ and means $\{\mu_{\xi_{\kappa_{\min}}^l}, \dots, \mu_{\xi_{\kappa_{\max}}^l}\}$ and use them to approximate the variance and mean of noise for an uncalibrated ISO level $j \in \{j | j \in \mathbb{N}, \kappa_{\min} \leq j \leq \kappa_{\max}\}$ as $\sigma_{\xi_j^l}^2$ and $\mu_{\xi_j^l}$, respectively. For each unseen ISO j , the noise PMF is approximated as

$$p_{\xi_j^l}(n) = \frac{1}{\sigma_{\xi_j^l}} p_{\xi_{\kappa}^l}\left(\frac{n - \mu_{\xi_j^l}}{\sigma_{\xi_j^l}}\right), \quad (6)$$

where $p_{\xi_{\kappa}^l}$ is a normalized noise distribution from the set of calibrated ISO levels. The approximated PMF is then used in (3) to synthesize noise for the uncalibrated ISO level j . A use case of such a non-parametric model interpolation is illustrated for a Pixel 6 mobile camera in Fig. 3b, where noise models for some random ISOs are obtained from the set of calibrated ISO levels shown in Fig. 3a. Note the minimal gap between the dashed and the solid-line plots in Fig. 3b.

4.2 Applying Noise on Noisy Images

Given the calibrated noise models in the form of PMFs $p_{\xi_1^l}(n_1)$, $p_{\xi_2^l}(n_2)$ for two different ISO settings and \tilde{I}_1 as an observed image captured at the first ISO level, we need to simulate \tilde{I}_2 as an image at the second ISO level. To the best of our knowledge, current practice is directly sampling noise from $p_{\xi_2^l}(n_2)$ and applying it on \tilde{I}_1 . However, we cannot ignore that \tilde{I}_1 is already contaminated with noise and hope that the image reconstruction network will learn to compensate for the additional non-sensor-specific noise without impacting the task performance.

Hence, we propose to model the noise to be added to the existing noisy image through accounting for the probability of the additive noise on top of existing noise denoted by $n_2 - n_1$ whose distribution characteristic function per intensity level l can be $p_{\xi_2^l}(n_2 - n_1)$. However, for building $p_{\xi_2^l}(n_2 - n_1)$, we need some approximation of the clean intensity image, denoted by \hat{I} . This is needed to pick the right PMFs corresponding to l , *i.e.*, $l \approx \hat{I}(i)$, and also to approximate n_1 as $n_1 \approx \tilde{I}_1(i) - \hat{I}(i)$. Therefore, a denoiser is applied on the observation \tilde{I}_1 to obtain \hat{I} . Thus, we propose to simulate \tilde{I}_2 via

$$\tilde{I}_2(i) = \tilde{I}_1(i) + \text{RandomSampling}\left(p_{\xi_2^l}(n_2 - n_1)\right). \quad (7)$$

In Sec. 5.3 and in Supplemental Material, we show that this approach results in a more accurate noise augmentation compared to naively applying the ISO setting 2 noise models on ISO setting 1 captures—namely, $\tilde{I}_2(i) = \tilde{I}_1(i) +$

RandomSampling($p_{\xi_2^i}(n_2)$), or applying the ISO setting 2 noise models on approximated clean images as $\tilde{I}_2(i) = \hat{I}(i) + \text{RandomSampling}(p_{\xi_2^i}(n_2))$.

5 Experiments

We analyze our noise models through statistics and training data synthesis for image reconstruction applications. Our method requires calibration captures at specialized controllable settings. Since we do not have access to the cameras used to build SIDD [2] and ELD [32], we cannot use these datasets to evaluate our noise modeling. Instead, we calibrate three different mobile cameras—Samsung’s S22+ and S20FE, and Google’s Pixel 6—and collect evaluation data using these cameras for noise synthesis analyses.

Noise calibration details. Following the steps in Sec. 4, we first calibrate S22+ and Pixel 6 cameras for a set of seven different ISO levels as $\{2^m 100 | m \in \mathbb{N}, 0 \leq m \leq 6\}$. Since S20FE is used for a set of experiments which require only ISO 1600 and ISO 3200 noise models, we calibrate S20FE for these two ISOs only. For noise sampling, per ISO level, we collect a burst of $M = 100$ raw captures with nine different exposure values—namely, EV -4 to EV 4. At each EV, we capture a burst of ($M = 100$) raw images at ISO 50 ($\bar{\kappa} = 50$) to generate a clean raw image. All of the camera sensors use Bayer CFAs. Thus, we model noise for each of the four Bayer color channels separately. Noise modeling is done for intensity levels ranging from the black level to the white level of the sensor. However, to prevent under-sampling, we assign one noise model to every bin of four intensity levels. For instance, for 10-bit sensors with a black level 64 and white level 1023, we end up with 240 noise PMFs per ISO, per CFA channel. To automate the raw capture sequence, we use the Android Camera2API.

Baseline noise modeling methods. Our comparative evaluations include the following noise modeling methods:

HG [34] as the most typical noise model for raw image denoising. HG models are traditionally calibrated following the procedure described in [13]. We instead calibrate the HG noise model for each camera following the recommendations in [34] that demonstrated improved performance using burst captures with temporal averaging to obtain the expected value for the underlying intensity.

RS-DP [39] as another physics-based method. This model is based on random sampling from dark frames for signal-independent noise and sampling from a Poisson distribution fit to the sensor’s signal-dependent noise (RS-DP for brevity). We perform the gain calibration step using flat-field images and the mean and variance of their pixel values, and capture 10 dark frames per ISO as a signal-independent noise database for this method, as suggested in [39].

Noise Flow [1] as the representative DNN-based method. We train a separate Noise Flow model for each camera to handle all seven ISO levels. Following the setup in [1], we train Noise Flow on real scenes rather than calibration images. We use the evaluation data described in Sec. 5.1 as our training data.

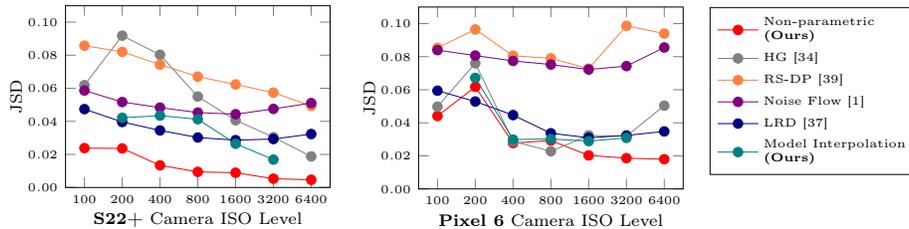


Fig. 4: Statistical analyses of per ISO sensor noise modeling for both S22+ and Pixel 6 cameras. JSD is measured relative to the real noise distribution obtained from homogeneous patches of the color checker chart captures (See Fig. 1b).

LRD [37] is the most recent noise modeling approach that combines physics-based and DNN-based methods. We train LRD on the data described in Sec. 5.1. The work in [37] did not provide source code for noise modelling. We implemented their method according to their descriptions, but with a slight modification to base the noise generator on our calibrated HG model, which provided better performance than the suggested Poisson model used in [37].

We assess our non-parametric noise model and its extensions to noise model interpolation and noise synthesis on top of noisy images separately as follows.

5.1 Analyses of Non-parametric Noise Modeling

Statistical Evaluations Our noise modeling method is evaluated by synthesizing noise on clean images and measuring the difference between the synthetic noise and the noise in real captures. To do so, we capture a sequence of 100 images of the color checker at ISO 50 and average them to generate a clean raw image. We then simulate ISO captures by applying the corresponding noise models on the clean raw image. This is shown in Fig. 1b, where the ISO 6400 calibrated noise model is applied on the clean raw image for the Pixel 6 camera.

We extract the histogram of the subtraction of the clean image from the corresponding noisy image at the homogeneous patches and consider it as the noise distribution. We then quantify the accuracy of noise models by measuring the entropy of the synthetic noise distribution relative to real noise, similarly to [1]. However, we use Jensen-Shannon divergence (JSD) rather than Kullback–Leibler divergence (KLD) since it always has a finite value and is more stable numerically. Fig. 4 shows JSD plots for our non-parametric noise models for different ISO levels for both S22+ and Pixel 6 cameras. The lower JSD indicates more similarity with the actual noise. To account for the randomness, each entry on the plots is the mean JSD of 100 synthetically generated images per ISO.

We also obtain JSD plots for the baseline noise models. As Fig. 4 shows, except for Pixel 6 for ISO 800, where HG has a lower JSD, and ISO 200 where LRD has a lower JSD, in all other cases, the proposed non-parametric method results in more realistic noise compared to other methods. See Supplemental Material for details about JSD and corresponding analyses using classic KLD.

Table 1: The performance of the raw denoising-demaicking model [7] trained on synthetically generated noisy raw images for S22+ and Pixel 6 cameras. Average PSNR and SSIM are measured on S22+ and Pixel 6 evaluation sets.

Noise-synthesis Method	S22+		Pixel 6	
	PSNR	SSIM	PSNR	SSIM
Non-parametric	39.74	0.954	38.49	0.950
HG [34]	39.62	0.942	38.00	0.949
RS-DP [39]	39.58	0.947	37.99	0.944
Noise Flow [1]	39.45	0.951	37.17	0.948
LRD [37]	39.63	0.949	38.11	0.945
Non-parametric ISO 1600 Interpolation	39.73	0.960	38.38	0.949
Non-parametric ISO 400 Interpolation	39.72	0.955	38.42	0.951

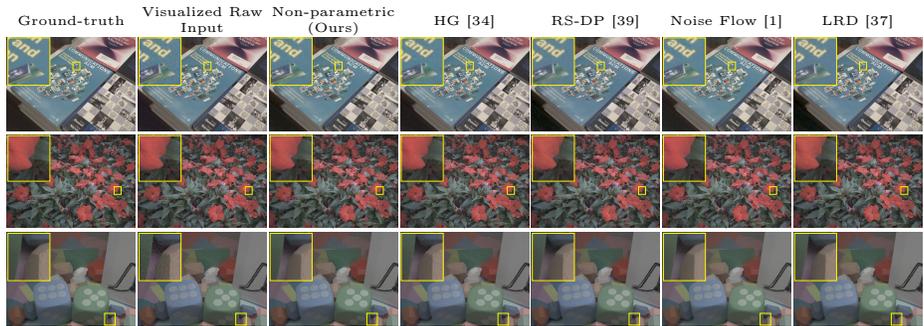


Fig. 5: Qualitative results on S22+ real raw captures. We apply different noise synthesis methods on EXR images [28] to create training images for the denoising-demaicking model [7]. Our non-parametric noise models help the image reconstruction model remove more noise whilst preserving more details. Raw inputs are bilinearly demosaicked and all images are white-balanced and gamma-corrected for better visualization.

Evaluation via Image Reconstruction We also evaluate our non-parametric noise models by using them to synthesize training data for the raw denoising-demaicking model introduced in [7] and evaluate it on real captures.

Synthesizing training data. For generating synthetic raw data, we use the graphics dataset introduced in [28] that contains 292 high-dynamic-range images in EXR format. We treat EXR data as photo-electrons and first convert them into the same bit format of the corresponding sensor, considering the sensor’s black and white levels. We then sub-sample the images according to the CFA pattern of the sensor and apply noise models per color channel per ISO level to simulate sensor raw measurements. We also account for the unbalanced color channels in the sensor. Prior to noise synthesis, we apply an inversion of the white-balance factors obtained during the calibration on the images. The ground-truth data in this training setup are the EXR images converted to 16-bit RGB format.

Evaluation data. For each camera, we collect a set of evaluation images using the Camera2API application. For each image pair in the evaluation set, we first

capture a burst of 50 frames at ISO 50, followed by 1 frame per calibrated ISO. For any given ISO, the exposure time is fixed such that the image’s EV is metered at 0. The images at ISO 50 are averaged to generate a clean raw image, followed by DCRaw’s demosaicker to generate an RGB image. The corresponding raw captures at higher ISO levels are used as noisy input images. After removing the misaligned captures, we end up with 64 test image pairs for S22+ and 72 pairs for Pixel 6, a mix of indoor and outdoor scenes.

Fig. 5 shows qualitative results of image reconstruction model [7] to handle S22+ raw images. The model is trained with synthetically generated data using our proposed noise models and the baseline noise models for S22+. These images show that using a more accurate noise model such as ours leads to a significant noise reduction along with much better detail preservation compared to using the other noise modeling methods. We list the peak-signal-to-noise ratio (PSNR) and structural similarity index (SSIM) measured on the entire evaluation sets for all the denoising-demosaicking [7] experiments in Table 1 for both cameras. Table 1 suggests that using the proposed non-parametric noise models for noise synthesis results in a higher quality compared to the other models. This is evident from the higher PSNR and SSIM metrics for the proposed method. A qualitative result is shown in Fig. 1c for Pixel 6. See Supplemental Material for more.

5.2 Analyses of Noise Model Interpolation

Noise modeling has to be done for a large number of ISO levels for realistic applications. This needs many captures per ISO calibration, *e.g.*, $2 \times 100 \times 9 = 1800$, as in our experiments. Fig. 3b shows the noise variance measured per intensity level of Pixel 6 sensor for some random ISO levels. Using the calibrated noise models obtained in the previous subsection, whose plots are shown in Fig. 3a, we obtain noise models for the random ISO levels using the approach proposed in Sec. 4.1. As Fig. 3b shows, our interpolated noise models result in synthetic noise quite similar to real noise in terms of measured variance per intensity level. We still go through a systematic analysis similar to Sec. 5.1 to validate the accuracy of the interpolated noise models.

We mask out the calibrated models obtained for ISO 200 to ISO 3200, one at a time, and approximate them through the proposed interpolation method given the rest of the calibrated models. We run the JSD analyses for each interpolated model and show the results in Fig. 4 for S22+ and Pixel 6 cameras, respectively. These plots indicate that the distributions of the synthetic noise generated using these interpolated models are quite similar to those of the calibrated models. In all cases except ISO 800 for Pixel 6, the approximated noise models show lower JSD compared to all the baselines calibrated or modelled on actual data.

We also evaluate the noise model interpolation method by using the interpolated models to synthesize training data for the raw denoising-demosaicking task [7]. We pick ISO 400 and ISO 1600 as relatively low ISO and high ISO, respectively, and use their corresponding interpolated noise models in the training experiment explained in Sec. 5.1. We replace the calibrated models with the interpolated ones and retrain the raw denoising-demosaicking model. As Table 1

shows, the performance of the raw denoising-demosaicking model trained using the interpolated noise models is similar to that of using the original calibrated noise models for both S22+ and Pixel 6. This validates the accuracy of the proposed noise model interpolation method that allows for a large set of ISOs in synthesizing data without going through the full calibration.

5.3 Validating Our Noise-on-noise Synthesis

We choose the night-time raw to sRGB image rendering task discussed in [26] for this set of experiments. The authors published a set of low-light raw images captured at ISO 1600 and ISO 3200 with a Samsung S20FE mobile camera along with their corresponding photo-finished clean sRGB images for training a neural image signal processor (ISP). This is an example of applications where the ground-truth data is the result of a cascade of black-box image processing steps, including non-linear processes such as contrast stretching and tone mapping, which cannot be truly inverted. Thus, the actual raw image is not accessible for further data augmentation through noise synthesis.

We choose this baseline over other popular datasets (*e.g.* [2]) since the S20FE camera is relatively more recent, and we can access the device to perform our non-parametric noise modeling. We train the neural ISP with the ISO 3200 sub-set of the training data as in [26]. The average PSNR of 39.38 dB and SSIM of 0.948 measured on the S20FE evaluation data set an upper bound for comparisons.

We remove ISO 3200 images from training data and need to synthesize ISO 3200 raw images for neural ISP training, assuming that ISO 1600 captures are the only available data. Typical noise synthesis is not an option here as clean raw captures are not available, and the target (clean) images are in sRGB. A naive solution to this data augmentation problem would be to use the ISO 3200 noise models directly on ISO 1600 images. Despite the fact that the underlying image is already noise-contaminated, this type of augmentation for training is very common in image reconstruction [36]. Table 2 shows that such an application of noise models on existing noisy captures leads to 1 dB lower PSNR, implying a low performance for this approach. The high JSD also shows that the synthetic noise is very different from the actual noise in this experiment.

We thus carefully account for the underlying noise using the approach discussed in Sec. 4.2; The calibrated noise models for the two ISO settings are used to build a single noise model to synthesize noise on top of ISO 1600 raw captures using Eq. (7). To obtain an estimate of a clean intensity raw image (*i.e.*, \hat{I}) for determining an appropriate noise model per sensor intensity level, we train a raw-to-raw denoising model for ISO 1600 captures. We use the same approach of using graphics data discussed in Sec. 5.1 along with the calibrated ISO 1600 noise models to synthesize data and train a raw-to-raw UNet network. This denoiser is then used to obtain an estimate of the clean intensity. Table 2 reports a relatively low JSD for this experiment. It also shows the raw-to-sRGB rendering performance of the neural ISP trained using the data generated through our proposed noise synthesis approach given ISO 1600 noisy captures. The average PSNR and SSIM on the S20FE evaluation set are very close to those of the model

Table 2: JSD analysis for the proposed noise-on-noise model, and evaluation via synthesizing S20FE ISO 3200 raw images for neural ISP. Different data sets; ISO 3200 real captures, naively applying ISO 3200 noise model on ISO 1600 raw images, and ISO 3200 noise synthesis using the proposed noise-on-noise model.

Training Data for Neural ISP [26]	PSNR	SSIM	JSD
Real Captures (Upper Bound)	39.38	0.948	-
Naive Noise-on-Noise Synthesis	38.03	0.929	0.049
Proposed Noise-on-Noise Synthesis	39.05	0.939	0.008



Fig. 6: Neural ISP qualitative results on S20FE real raw captures at ISO 3200. The upper bound corresponds to training with real captures at ISO 3200. We use different noise synthesis methods applied to ISO 1600 images to create ISO 3200 training images for raw to sRGB rendering. Our noise-on-noise model to synthesize training data helps the neural ISP preserve more details compared to naively applying calibrated ISO 3200 noise models to ISO 1600 images. (See Supplemental Material for more results.)

trained using real captures. For ablation studies on different choices of denoiser, underlying calibrated noise model, and directly applying the calibrated ISO 3200 noise model on the denoised ISO 1600 images, see Supplemental Material.

6 Limitation and Conclusion

We have presented a novel non-parametric sensor noise model consisting of probability mass functions per intensity level that are constructed directly from captured images. We show that this straightforward noise model provides state-of-the-art performance in fitting real sensor noise compared to existing noise models. We also outline an image capture procedure to derive our non-parametric noise model with an interpolation technique that effectively reduces the necessary number of capturing ISOs. Moreover, we have proposed a method for synthesizing noise on pre-existing noised images in the absence of noise-free counterparts.

We do acknowledge that fixed-pattern noise and random band-pattern noise are not addressed explicitly in our work. Such signal-independent noise components, however, can be modeled as efficiently as in [33, 39] before obtaining our non-parametric noise PMFs. Moreover, the need for a probability mass function at dense intensity levels results in an inherently larger footprint over conventional parametric models. Future work that targets a mixture of parametric and non-parametric modeling over the sensor’s intensity range could help reduce capture and memory size while maintaining accuracy.

References

1. Abdelhamed, A., Brubaker, M.A., Brown, M.S.: Noise flow: Noise modeling with conditional normalizing flows. In: ICCV (2019)
2. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: CVPR (2018)
3. Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: CVPR (2019)
4. Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input / output image pairs. In: CVPR (2011)
5. Cao, Y., Liu, M., Liu, S., Wang, X., Lei, L., Zuo, W.: Physics-guided ISO-dependent sensor noise modeling for extreme low-light photography. In: CVPR (2023)
6. Chang, K.C., Wang, R., Lin, H.J., Liu, Y.L., Chen, C.P., Chang, Y.L., Chen, H.T.: Learning camera-aware noise models. In: ECCV (2020)
7. Chen, C., Chen, Q., Xu, J., Koltun, V.: Learning to see in the dark. In: CVPR (2018)
8. Chen, J., Chen, J., Chao, H., Yang, M.: Image blind denoising with generative adversarial network based noise modeling. In: CVPR (2018)
9. Devroye, L.: Non-Uniform random variate generation. Springer New York, NY (1986)
10. El Gamal, A., Fowler, B.A., Min, H., Liu, X.: Modeling and estimation of FPN components in CMOS image sensors. In: Solid State Sensor Arrays: Development and Applications II. vol. 3301, pp. 168–177 (1998)
11. Feng, H., Wang, L., Wang, Y., Fan, H., Huang, H.: Learnability enhancement for low-light raw image denoising: A data perspective. IEEE TPAMI pp. 1–18 (2023)
12. Foi, A.: Clipped noisy images: Heteroskedastic modeling and practical denoising. Signal Processing **89**(12), 2609–2629 (2009)
13. Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. IEEE TIP **17**(10), 1737–1754 (2008)
14. Gow, R.D., Renshaw, D., Findlater, K., Grant, L., McLeod, S.J., Hart, J., Nicol, R.L.: A comprehensive tool for modeling CMOS image-sensor-noise performance. IEEE Transactions on Electron Devices **54**(6), 1321–1329 (2007)
15. Hasinoff, S.W.: Photon, poisson noise. Computer Vision, A Reference Guide **4**(16), 1 (2014)
16. Hwang, Y., Kim, J.S., Kweon, I.S.: Difference-based image noise modeling using skellam distribution. IEEE TPAMI **34**(7), 1329–1341 (2011)
17. Ignatov, A., Van Gool, L., Timofte, R.: Replacing mobile camera ISP with a single deep learning model. In: CVPRW (2020)
18. Jang, G., Lee, W., Son, S., Lee, K.M.: C2N: Practical generative noise modeling for real-world denoising. In: ICCV (2021)
19. Kim, D.W., Ryun Chung, J., Jung, S.W.: GRDN: Grouped residual dense network for real image denoising and GAN-based real-world noise modeling. In: CVPRW (2019)
20. Konnik, M., Welsh, J.: High-level numerical simulations of noise in CCD and CMOS photosensors: Review and tutorial. arXiv (2014)
21. Liu, X., Tanaka, M., Okutomi, M.: Practical signal-dependent noise parameter estimation from a single noisy image. IEEE TIP **23**(10), 4361–4371 (2014)
22. Makitalo, M., Foi, A.: Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise. IEEE TIP **22**(1), 91–103 (2012)

23. Maleky, A., Kousha, S., Brown, M.S., Brubaker, M.A.: Noise2NoiseFlow: Realistic camera noise modeling without clean images. In: CVPR (2022)
24. Monakhova, K., Richter, S.R., Waller, L., Koltun, V.: Dancing under the stars: Video denoising in starlight. In: CVPR (2022)
25. Ploetz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: CVPR (2017)
26. Punnappurath, A., Abuolaim, A., Abdelhamed, A., Levinshtein, A., Brown, M.S.: Day-to-night image synthesis for training nighttime neural ISPs. In: CVPR (2022)
27. Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: ECCV (2020)
28. Seo, D., Punnappurath, A., Zhao, L., Abdelhamed, A., Tedla, S.K., Park, S., Choe, J., Brown, M.S.: Graphics2RAW: Mapping computer graphics images to sensor raw images. In: ICCV (2023)
29. Suh, S., Itoh, S., Aoyama, S., Kawahito, S.: Column-parallel correlated multiple sampling circuits for CMOS image sensors and their noise reduction effects. *Sensors* **10**(10), 9139–9154 (2010)
30. Tran, L.D., Nguyen, S.M., Arai, M.: GAN-based noise model for denoising real images. In: ACCV (2020)
31. Wang, W., Chen, X., Yang, C., Li, X., Hu, X., Yue, T.: Enhancing low light videos by exploring high sensitivity camera noise. In: ICCV (2019)
32. Wei, K., Fu, Y., Yang, J., Huang, H.: A physics-based noise formation model for extreme low-light raw denoising. In: CVPR (2020)
33. Wei, K., Fu, Y., Zheng, Y., Yang, J.: Physics-based noise modeling for extreme low-light photography. *IEEE TPAMI* **44**(11), 8520–8537 (2021)
34. Yoshimura, M., Otsuka, J., Irie, A., Ohashi, T.: Rawgment: Noise-accounted raw augmentation enables recognition in a wide variety of environments. In: CVPR (2023)
35. Yue, Z., Zhao, Q., Zhang, L., Meng, D.: Dual adversarial network: Toward real-world noise removal and noise generation. In: ECCV (2020)
36. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: CycleISP: Real image restoration via improved data synthesis. In: CVPR (2020)
37. Zhang, F., Xu, B., Li, Z., Liu, X., Lu, Q., Gao, C., Sang, N.: Towards general low-light RAW noise synthesis and modeling. In: ICCV (2023)
38. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE TIP* **26**(7), 3142–3155 (2017)
39. Zhang, Y., Qin, H., Wang, X., Li, H.: Rethinking noise synthesis and modeling in raw denoising. In: ICCV (2021)