Learnability Enhancement for Low-Light Raw Image Denoising: A Data Perspective

Hansen Feng, Lizhi Wang, *Member, IEEE*, Yuzhi Wang, Haoqiang Fan, and Hua Huang, *Senior Member, IEEE*

Noise
- Shot noise
- Dark shading
  - fixed pattern noise (FPN)
    - iso-dependent (FPN-k, dark current cause by temporal stable heat)
    - iso-independent (FPN-b, offset current)
  - black level error (BLE, also dark current cause by ISO and exposure time t)

lacks learnability 就是很難學的意思

***Abstract*—Low-light raw image denoising is an essential task in computational photography, to which the learning-based method has become the mainstream solution. The standard paradigm of the learning-based method is to learn the mapping between the paired real data, i.e., the low-light noisy image and its clean counterpart. However, the limited data volume, complicated noise model, and underdeveloped data quality have constituted the learnability bottleneck of the data mapping between paired real data, which limits the performance of the learning-based method. To break through the bottleneck, we introduce a learnability enhancement strategy for low-light raw image denoising by reforming paired real data according to noise modeling. Our learnability enhancement strategy integrates three efficient methods: shot noise augmentation (SNA), dark shading correction (DSC) and a developed image acquisition protocol. Specifically, SNA promotes the precision of data mapping by increasing the data volume of paired real data, DSC promotes the accuracy of data mapping by reducing the noise complexity, and the developed image acquisition protocol promotes the reliability of data mapping by improving the data quality of paired real data. Meanwhile, based on the developed image acquisition protocol, we build a new dataset for low-light raw image denoising. Experiments on public datasets and our dataset demonstrate the superiority of the learnability enhancement strategy.**

***Index Terms*—Computational photography, low-light denoising, noise modeling, dataset.**

I. Introduction

COMPUTATIONAL photography, as an efficient way to improve image quality, has long been applied to various cameras. However, inescapable noise significantly limits the performance of computational photography in low-light scenarios. Therefore, low-light raw image denoising plays an important role in computational photography, which has been extensively studied in mobile photography [1], [2], astronomy [3], [4] and microscopy [5]. With the expansion of computing power, learning-based methods have made great progress in recent years and become the mainstream solution to the low-light raw image denoising [6], [7]. The standard paradigm of the learning-based methods is to learn the mapping between the paired real data, i.e., the low-light noisy image and its clean counterpart, via a neural network [8].

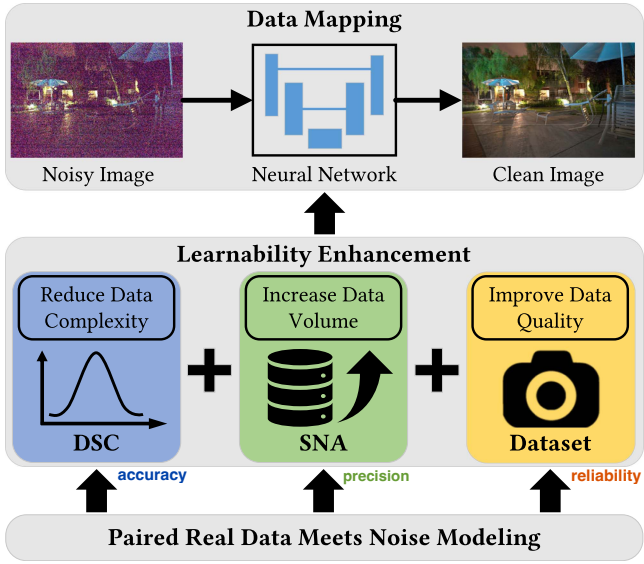


Fig. 1. From the data perspective, image denoising via learning-based methods can be modeled as a data mapping from the noisy image to the clean image. The learnability of data mapping depends on the volume of paired data, the complexity of noise model, and the quality of paired real data. Accordingly, we develop a learnability enhancement strategy for low-light raw image denoising by reforming paired real data according to noise modeling.

Learnability characterizes the difficulty of data mapping being approximated by a neural network [9]. Enhancing the learnability of data mapping is the most efficient way to improve denoising performance. From the data perspective, the learnability of data mapping in image denoising highly depends on the data volume of paired real data, the complexity of noise model, and the quality of paired real data. Most studies on image denoising [10], [11], [12], [13], [14], [15], [16], [17], [18] focus on customizing complicated neural networks to fit the data mapping while neglecting the learnability of the data mapping.

Unfortunately, the mapping between the paired real data is difficult to be learned due to its fragile learnability. First, the data volume is limited by the complex environment, leading to the low precision of data mapping. Second, the real noise follows a complicated distribution related to the sensor imaging process,

Manuscript received 17 January 2023; revised 21 June 2023; accepted 20 July 2023. Date of publication 3 August 2023; date of current version 5 December 2023. This work was supported by the National Natural Science Foundation of China under Grants 62322204, 62072038, and 62131003. Recommended for acceptance by J. Wang. *(Corresponding author: Lizhi Wang.)*

Hansen Feng and Lizhi Wang are with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: fenghansen@bit.edu.cn; wanglizhi@bit.edu.cn).

Yuzhi Wang and Haoqiang Fan are with Megvii Technology, Beijing 100086, China (e-mail: wangyuzhi@megvii.com; fhq@megvii.com).

Hua Huang is with the School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China (e-mail: huahuang@bnu.edu.cn).

Digital Object Identifier 10.1109/TPAMI.2023.3301502

0162-8828 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

leading to the low accuracy of data mapping.[1] Third, the quality of paired real data usually suffers from data flaws due to noise and misalignment in practice, leading to the low reliability of data mapping. The complicated noise model, limited data volume, and underdeveloped data quality contribute to the mapping dilemma of paired real data in low-light raw image denoising.

Recent works attempt to address the mapping dilemma by synthesizing data according to noise modeling instead of employing the paired real data [19], [20], [21]. Such methods can synthesize new noisy and clean image pairs for learning data mapping. However, since some parts of the noise (e.g., read noise) are far from being accurately modeled, the synthetic data would inescapably deviate from the real data, which loses effectiveness in real-world scenarios. As a result, the fragile learnability of paired real data still remains a bottleneck in low-light raw image denoising.

In this paper, we propose a learnability enhancement strategy for low-light raw image denoising. We revisit the denoising task from a data perspective by reforming paired real data according to noise modeling.

Our first observation is that the shot noise is only related to the clean image and can be accurately modeled with the Poisson distribution [22], [23]. We propose the Shot Noise Augmentation (SNA) method to increase the data volume of paired real data. SNA first combines the real data and the shot noise model to synthesize new noisy-clean data pairs and then augment paired real data to promote mapping precision. Benefiting from the increased data volume, the mapping can promote denoised images with clear textures.

Our second observation is that the dark shading keeps temporal stable and can be modeled with a pixel-wise bias [24], [25]. We propose the Dark Shading Correction (DSC) method to decouple the real noise model. DSC first calibrates the dark shading and then corrects it in the noisy image to promote mapping accuracy. Benefiting from the reduced noise complexity, the mapping can promote denoised images with exact colors.

Our third observation is that the existing image acquisition protocol is underdeveloped for low-light raw image denoising datasets, including data flaws due to noise and misalignment. We propose a high-quality image acquisition protocol and a Low-light Raw Image Denoising (LRID) dataset. The protocol first develops the image capture setup and then estimates the clean ground truth to promote mapping reliability. Benefiting from the improved data quality, the mapping can promote denoised images with fewer artifacts.

The ideas originate from the integration of paired real data and noise modeling which are assumed as two parallel directions for low-light raw image denoising. Extensive experiments on public datasets and our dataset demonstrate the superiority of our learnability enhancement strategy.

Our main contributions are summarized as follows:

1) We light the idea of learnability enhancement for low-light raw image denoising by reforming paired real data according to the noise modeling from a data perspective.
2) We increase the data volume of paired real data with a novel shot noise augmentation method, which promotes the precision of data mapping by data augmentation.
3) We reduce the noise complexity with a novel dark shading correction method, which promotes the accuracy of data mapping by noise decoupling.
4) We develop a high-quality image acquisition protocol and build a low-light raw image denoising dataset, which promotes the reliability of data mapping by improving the data quality of paired real data.
5) We demonstrate the superiority of our methods on public datasets and our dataset in both quantitative results and visual quality.

The code and dataset have been released at [26]. We further develop the noise calibration method based on dark shading and provide the data for calibration. We believe the calibration data is meaningful for further studies on low-light raw image denoising.

This work has significant improvements beyond our preliminary work [27].

First, we develop a high-quality image acquisition protocol to improve the data quality of paired real data. Based on the protocol, we build a low-light raw image denoising dataset, which promotes the reliability of data mapping.

Second, we analyze the property of SNA and extend the application strategy based on the real read noise samples. The extended design can promote denoised images with clear details.

Moreover, we develop the linear dark shading model and present an in-depth analysis of the robustness and generalizability of DSC. We further explore the extension of DSC on physics-based noise modeling, which brings huge denoising performance improvements.

Finally, we evaluate our methods on more datasets and conduct more extensive experiments to show the potential widespread usage of our methods.

II. Related Works

A. Low-Light Raw Image Denoising

Classical denoising methods usually rely on image priors such as smoothness [28], [29], self-similarity [30], [31], [32], sparsity [33], [34], and low rank [35]. Instead of pre-setting an image prior, learning-based methods directly learn the mapping from the noisy image to its clean counterpart (paired real data) via deep neural networks. Recent works demonstrate that learning-based methods have been far superior to classical methods in denoising performance [8], [36], [37]. However, these learning-based methods often struggle with fragile learnability due to complicated noise model and limited data volume.

To address the problem of fragile learnability, some studies focus on improving the realism of noise modeling to bypass the need for paired real data. The sensor noise is typically divided into "shot noise" and "read noise'. The Poisson-Gaussian model is often used to model the real noise in low-light raw image denoising [38]. To ensure the noise model can be applied to different cameras under various conditions, various calibration [23], [38], [39] and Variance-Stabilizing Transformation (VST) [40], [41] methods have been proposed. ELD is a physics-based noise

[1]The difference of accuracy and precision can refer to https://en.wikipedia.org/wiki/Accuracy_and_precision

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

modeling method for extreme low-light photography that has achieved results on par with paired real data [20], [21]. However, ELD has limitations due to its assumptions about i.i.d. temporal noise, which differs from real spatial noise, such as fixed pattern noise. As a result, the visual quality of ELD is still affected by residual pattern noise, indicating the importance of real data.

Despite these limitations, noise modeling has shown great potential in breaking through the learnability bottleneck of paired real data. The physical properties of the photoelectric reaction are well-defined, leading to the reliable analysis of shot noise [22], [23], [39]. In contrast, read noise is much more complicated, with no consensus in the imaging community [24], [39], [42], [43], [44], [45]. It is extremely difficult to extract and model all types of noise sources in a camera. One solution is to use real data in conjunction with noise modeling. SFRN uses reliable shot noise for noise modeling and samples complicated read noise from real dark frames, which has liberated the data volume of paired real data [46]. However, SFRN inherits the complexity of real read noise, leading to the challenging learning process of data mapping.

We find that the high complexity of data mapping is mainly due to dark shading [25]. In order to address both the limited data volumes and the complicated noise model, we reform paired real data according to noise modeling. Addressing fragile learnability leads to significant improvements in denoising performance, especially in low-light conditions.

B. Raw Image Denoising Datasets

Raw images have the advantage of being linear and high-bit compared to sRGB images, which drives raw image processing simple and efficient. In recent years, with the increasing demand for computational photography, raw image denoising datasets have also been developed.

The DND dataset [47] includes data from 4 different cameras and 50 scenes and also introduces a new protocol for acquiring raw images, mainly consisting of image pairs captured at high- and low-ISO. However, most of the data in the DND dataset is collected under normal lighting conditions with relatively low noise levels. The SIDD dataset [48] includes data from 5 different smartphone cameras in 10 indoor scenes under four lighting conditions, mainly consisting of multi-frame fusion results and noisy images (reference frames). The SIDD dataset further develops the image acquisition protocol and is currently the most widely used benchmark. However, SIDD has limited scene diversity, especially lacking main application scenarios for low-light denoising such as outdoor scenes, which has been proven essential in recent study [49]. More closely related to our effort is the SID dataset [8], which includes data from 2 different cameras in 424 low-light scenes, mainly consisting of image pairs captured at different exposure times. The SID dataset provides the first large-scale training dataset for low-light raw image denoising with paired real data. However, the image acquisition protocol of the SID dataset actually has many data flaws, such as significant noise in high-ISO and motion blur in the ground truth. These data flaws cause the fragile learnability of the data. The ELD dataset [21] includes data from 4 different cameras in 10 indoor scenes, mainly consisting of image pairs captured at different exposure times and different ISO. The ELD dataset can address the data flaws in indoor scenes without motion, however, the proposed image acquisition protocol cannot be generalized to outdoor scenes due to the lack of alignment. We focus on developing the image acquisition protocol for outdoor scenarios. Our image acquisition protocol addresses various data flaws in previous datasets that affect learnability. We propose a new image acquisition protocol and a dataset for low-light raw image denoising, which promotes the reliability of data mapping by improving the data quality of paired real data.

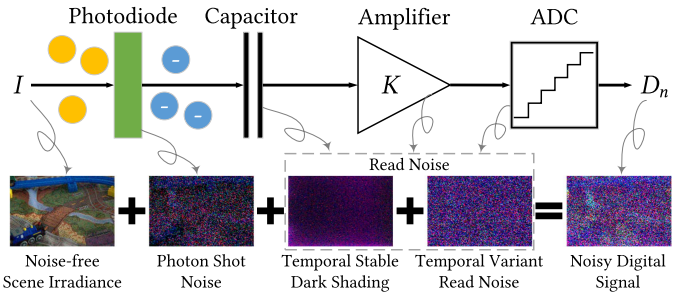


Fig. 2. Overview of a simplified imaging pipeline and our noise modeling. Photons are converted into charge and sequentially voltages, then amplified, and finally quantized into digital signals. We visualize the noise and connect it with the corresponding noise sources. Noise-free scene irradiance suffers inescapable photon shot noise and read noise from various electronic components, thus the output digital signal is noisy.

III. Method

In this section, we first introduce our framework, which includes the principle and procedure of the learnability enhancement strategy. Then, we introduce our Shot Noise Augmentation (SNA) based on the shot noise model, which increases the data volume. Finally, we introduce our Dark Shading Correction (DSC) based on the read noise model, which reduces the noise complexity.

A. Framework

From a data perspective, the mapping between the paired real data lacks learnability, which limits the performance of learning-based denoising methods. The limited data volume and complicated noise model are two of the main culprits that lead to the fragile learnability of paired real data. To enhance the learnability, it is necessary to increase the data volume and reduce the noise complexity while maintaining the real noise model. However, addressing these problems is challenging since the real noise model in paired real data is a "black box". We propose a data augmentation method to increase the data volume. The data augmentation utilizes the additive property of the shot noise model, i.e., Poisson distribution. We propose a noise decoupling method to reduce the noise complexity. The noise decoupling splits the read noise model into simple forms, i.e., temporal stable dark shading and temporal variant read noise. Our learnability enhancement strategy addresses the limited data volume and complicated noise model while preserving the real noise model, thereby enhancing the learnability of data mapping.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

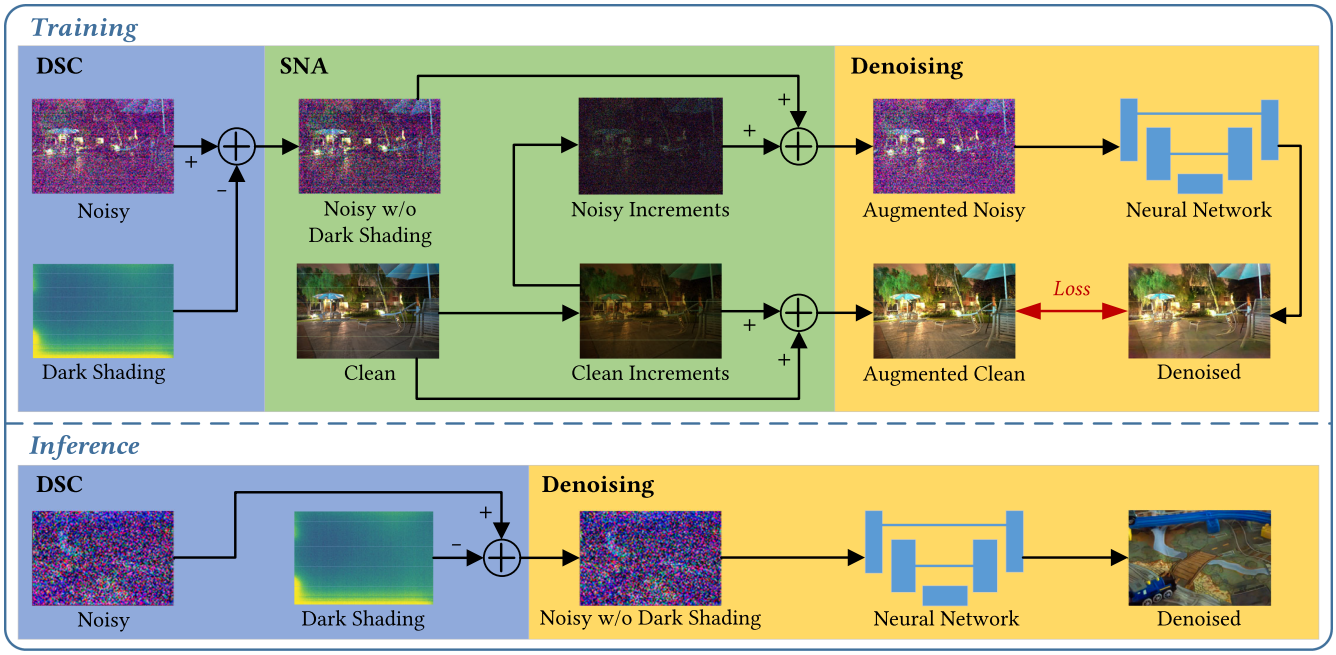


Fig. 3. Overview of our framework. For training, we first correct the dark shading hiding in the noisy raw image by DSC. Then we augment the clean image and the noisy image to obtain new data pairs by SNA. Finally, We use augmented noisy and clean images to train a neural network. For inference, we denoise the noisy image after correcting dark shading with the trained denoising model.

To explain our work clearly, we briefly introduce the imaging model of a sensor. As shown in Fig. 2, in camera electronics, the scene irradiance I is first converted into charges and sequentially voltages, which are then amplified by amplifiers, and finally quantized into a noisy digital signal D_n by an analog-to-digital converter (ADC) [39]. For a raw image captured by a sensor, we model the noise formation process as

$$D_n = K\left(I + N_p\right) + N_{read}, \tag{1}$$

where K represents the overall system gain, I is the number of photo-electrons excited by the scene radiance, N_p is the signal-dependent photon shot noise, and N_{read} is signal-independent read noise.

As we have emphasized, the real noise model in paired real data is a "black box". It is challenging to synthesize new data pairs with the same noise characteristics as the real noise model. In fact, the learning target of the learning-based methods is the data mapping between the paired real data. Thus we can augment and decouple the paired real data as long as these operations maintain reliable data mapping. In other words, as long as the data mapping is consistent, the augmented data can follow the new noise parameters, while the decoupled data can be processed separately.

The framework of learnability enhancement is shown in Fig. 3. SNA is a data augmentation method based on the shot noise model. We continuously synthesize new noisy-clean data pairs by augmenting the shot noise to increase the data volume. DSC is a noise decoupling method based on the read noise model. We first calibrate the dark shading and then correct it in paired real data to reduce the noise complexity. The data augmented by SNA has different noise parameters from the original data while sharing the same noise model. The data decoupled by DSC can be restored to the original data by adding the decoupled data together. Therefore, our strategy can enhance the learnability of data mapping without breaking the real noise model.

Following (1), we suppose a noisy image D_n and the corresponding noise-free clean image D_c. The relationship between them satisfies

$$D_n \sim \mathcal{F}(D_c), \tag{2}$$

where $\mathcal{F}$ donates the noise mode, D_c is equal to KI.

In mathematics, the denoising mapping can be abstracted as the function f, given by

$$f : D_n \rightarrow D_c. \tag{3}$$

Combining (1) and (3), we obtain the denoising mapping expressed as

$$f : K\left(I + N_p\right) + N_{read} \rightarrow KI. \tag{4}$$

By applying the learnability enhancement, the enhanced mapping in our framework can be expressed as

$$f_{LE} : \underbrace{K\left(I + N_p\right) + \Delta N}_{SNA} + \underbrace{\left(N_{read} - N_{ds}\right)}_{DSC} \rightarrow \underbrace{KI + \Delta D}_{SNA}, \tag{5}$$

where f_{LE} is the mapping applying the learnability enhancement, N_{ds} is the dark shading (i.e., temporal stable component of read noise) produced by DSC, ΔN and ΔD are noisy and clean signal increments produced by SNA, respectively. The details of these variables will be introduced in subsequent sections.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

B. Shot Noise Augmentation

1) The Principle of SNA: SNA is a data augmentation method based on the shot noise model. Due to the quantum nature of light and the uncertainty of the collected photon numbers, $(I + N_p)$ for all pixels follow the Poisson distribution

$$(I + N_p) \sim \mathcal{P}(I), \tag{6}$$

where $\mathcal{P}$ denotes the Poisson distribution.

Now we suppose that there is a clean signal increments ΔD, and our target is to find corresponding noisy signal increments ΔN satisfying

$$(D_n + \Delta N) \sim \mathcal{F}(D_c + \Delta D). \tag{7}$$

For paired real data, the clean image D_c is known. According to (1) and (6), the noise model, $\mathcal{F}(D_c)$, can be expressed as

$$\mathcal{F}(D_c) = K\mathcal{P}\left(\frac{D_c}{K}\right) + \mathcal{F}_{read}, \tag{8}$$

where $\mathcal{F}_{read}$ is the noise model of read noise N_{read}, which is unknown.

Variables following Poisson distributions satisfy the additive property, i.e., $X_1 \sim \mathcal{P}(\lambda_1)$ and $X_2 \sim \mathcal{P}(\lambda_2)$ are independent, then $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$. Using the additivity of Poisson distribution, we can derive an expression for $\mathcal{F}(D_c + \Delta D)$ as

$$\begin{aligned}\mathcal{F}(D_c + \Delta D) &= K\mathcal{P}\left(\frac{D_c + \Delta D}{K}\right) + \mathcal{F}_{read} \\ &= K\mathcal{P}\left(\frac{D_c}{K}\right) + K\mathcal{P}\left(\frac{\Delta D}{K}\right) + \mathcal{F}_{read} \\ &= \mathcal{F}(D_c) + K\mathcal{P}\left(\frac{\Delta D}{K}\right).\end{aligned} \tag{9}$$

當clean image 加上data D，Noisy image 再多上這個 term 即可

By substituting this expression into (7), we can obtain

$$\Delta N \sim K\mathcal{P}\left(\frac{\Delta D}{K}\right). \tag{10}$$

Then we can perform data augmentation by adding clean signal increment ΔD and noisy signal increment ΔN to the clean image D_c and noisy image D_n, respectively. Since ΔN only needs to follow the distribution in (10), we can obtain new noisy data every time we randomly sample, even when the clean signal increment ΔD remains constant. As a result, SNA significantly increases the data volume of paired real data, and also improves the density of the data within the manifold, resulting in high precision in the fitting of data mapping.

By increasing the data volume through SNA, the neural network can precisely fit the data mapping between the paired real data, which promotes denoised images with clear texture.

2) The Extension of SNA: Independent variables following Poisson distributions satisfy the additive property, however, their subtraction results no longer conform to the Poisson distribution. As a result, SNA has a property that the signal of base image (i.e., noisy image D_n) can only increase monotonically, causing a practical challenge, i.e., SNA can produce images that are brighter than the base image, but not darker. This limitation restricts the enhancement ability of SNA in dark scenes.

We further propose a solution to simply capture the paired noisy image in the extremely dark environment, i.e., collect the noisy dark frames containing only read noise. It can be regarded as a special case of (1) when the scene irradiance I is 0, where the dark frame is treated as a noisy image D_n, and the noisy signal increments ΔN are sampled according to

$$\Delta N \sim K\mathcal{P}\left(\frac{D_c + \Delta D}{K}\right). \tag{11}$$

However, simply sampling the dark frame is not sufficient due to the lack of information after quantization. To address the lack of information, we adopt the high-bit recovery strategy proposed in SFRN [46] to help dark frames precisely represent the real read noise in this situation.

In conclusion, we extend the application strategy of SNA by introducing dark frames, which can promote denoised images with clear details.

Procedure 1: Shot Noise Augmentation.

Require: D_c, D_n
Ensure: D_c^*, D_n^*
  function parameter samplingμ, σ
    $\epsilon_g \sim \mathcal{N}(\mu + 1, \sigma), \quad \epsilon_r, \epsilon_b \sim \mathcal{N}(1, \sigma)$
    $Gain_g \leftarrow clip(\epsilon_g)_1^{4\mu+1}$
    $Gain_r \leftarrow clip(Gain_g \cdot \epsilon_r)_1^{4\mu+1}$
    $Gain_b \leftarrow clip(Gain_g \cdot \epsilon_b)_1^{4\mu+1}$
    $Gain \leftarrow (Gain_r, Gain_g, Gain_b)$

    return $Gain$
  end Function
  $Gain \leftarrow$ parameter sampling(μ, σ)
  $D_c^* \leftarrow D_c \cdot Gain$
  $\Delta D \leftarrow D_c^* - D_c$
  $\Delta N \sim K\mathcal{P}(\frac{\Delta D}{K})$
  $D_n^* \leftarrow D_n + \Delta N$

3) The Procedure of SNA: The augmentation procedure has been shown in Procedure 1. First, we randomly sample a set of parameters to gain the clean image D_c in order to obtain an augmented clean image D_c^*. Second, we obtain the clean signal increments ΔD by computing the difference of clean image D_c and augmented clean image D_c^*. Then, we synthesize noisy signal increments ΔN according to the clean signal increments ΔD. Finally, we add the synthesized noisy signal increments ΔN to the real noisy image D_n to obtain an augmented noisy image D_n^*. The augmented noisy image D_n^* and augmented clean image D_c^* will constitute a new noisy-clean image pair.

It is free to design the specific parameter sampling strategy for the synthesis of clean signal increments ΔD, however, some basic principles still need to be followed: (1) clean signal increments ΔD should be non-negative to ensure that Poisson sampling can be applied; (2) it is necessary to ensure that SNA will not introduce obvious color bias and large-scale over-exposure.

Based on the above principles, we choose to synthesize clean signal increments ΔD based on clean image D_c. We only augment the ratio of color channels and global brightness in

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

直接拿GT與圖相減的結果
- 那爲什麼告示牌還會出現？不是應該直接被減掉嗎？
  - 可能是因為整張圖像的對比度有差異，導致相減後輪廓出現
- 為什麼 pixel 數值介於 -2~2？
  - 做標準化
  - 所以我們才希望 Dark Shading (high-frequency) 資訊被移除掉
  - 讓背景是 0 (青色部分)

order to simulate the real scene, which refers to the white balance modeling [19]. First, we randomly initialize factors $\epsilon_r, \epsilon_g, \epsilon_b$ that follow a Gaussian distribution, where μ and σ are the parameters of the Gaussian distribution. Then, we randomly sample the gain of the green channel $Gain_g$ based on the clipped ϵ_g. Next, we select the gain of the red channel $Gain_r$ and blue channel $Gain_b$ based on the random factors ϵ_r, ϵ_b and the gain of the green channel $Gain_g$. Finally, the gain parameters $(Gain_r, Gain_g, Gain_b)$ is used to gain different color channels, and they are all clipped to the appropriate value range to meet the basic design principles.

Raw image denoising poses unique challenges in terms of data augmentation. The Bayer pattern of raw images and the risk of breaking the real noise model can create a gap between the training and application domains. As a result, raw image denoising methods must be cautious in their use of limited data augmentation techniques (such as rotation and flipping), which often lack sufficient data volume [50]. Additionally, these methods are insufficient in promoting the diversity of noise signals, which is crucial in fitting the data mapping. In contrast, SNA can promote noise diversity without breaking the real noise model, making SNA novel and important in the field of raw image denoising.

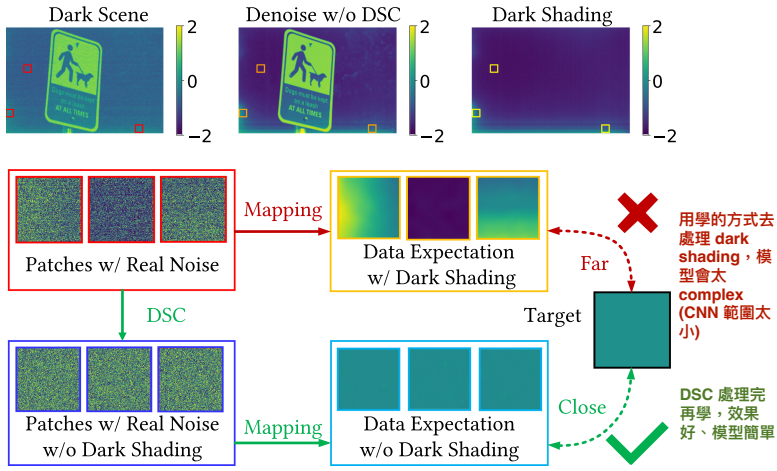


Fig. 4. Example of the mapping dilemma considering dark shading. The dark scene represents an image captured under low-light conditions. The red patches represent regions cropped from the dark scene, while the orange patches represent regions cropped from the data expectation with dark shading (i.e., Denoise w/o DSC). The yellow patches are the corresponding dark shading at the same spatial location. The blue patches are the result of applying DSC to the red patches. The sky blue patches represent the data expectation of the noisy image without dark shading. The black target patch corresponds to the clean counterpart of the dark scene patches, which are all zero values in this specific example. We have converted the color space and limited the value range for viewing.

C. Dark Shading Correction

1) The Motivation of DSC: In this part, we introduce the definition and mechanism of dark shading and clarify the importance of correcting dark shading for low-light raw image denoising.

From a practical perspective, the properties of an array of pixels vary from pixel to pixel in the sensor [39]. We follow the mainstream term "dark shading" in the industry to describe such spatial non-uniformity [25], [45]. In general, dark shading represents the temporal stable component in the read noise, which can be regarded as the union of black level error (BLE) [21], [51] and fixed pattern noise (FPN) [24], [39], [44], [52], [53]. BLE is the spatial invariant bias of dark shading while FPN is a spatial variant pattern.

時間穩定的成分 (Temporal Stable Read Noise)：在不同影像之間幾乎不變的噪聲，例如感測器本身的製造誤差導致的固定圖樣噪聲。

BLE影像所有像素的亮度都有相同的偏移，整體畫面變亮或變暗
FPN不同位置的像素表現不同的噪聲值，形成一種固定的紋理噪聲

Most denoising methods [2], [10], [11], [12], [13], [20], [38], [54] assume that sensor noise is zero-mean in the temporal dimension. In the ideal assumption, the expectation of noisy image $\mathbb{E}(D_n)$ is the same as the mapping target D_c in mathematics. However, for the real sensor noise, the expectation of read noise $\mathbb{E}(N_{read})$ is non-zero dark shading, which breaks the ideal assumption. If spatial variant dark shading is present, the mapping between $\mathbb{E}(D_n)$ and the clean image D_c will become a patch-wise non-injective mapping. Compared with the ideal assumption, this is equivalent to dark shading introducing additional high-frequency information to the data mapping. As per the frequency principle of deep learning, high-frequency information is difficult to be learned by neural networks [55], [56]. Thus the learnability of data mapping is fragile as long as the data includes dark shading. Previous studies have attempted to use convolutional neural networks (CNNs) to remove FPN caused by dark shading, however, CNNs as ensembles of local operators cannot completely remove FPN without global information. Therefore, it is essential to correct dark shading before denoising in order to simplify the data mapping and improve denoising performance.

非一對一映射（多對一）多種 noise 對到一個 clean

To illustrate the mapping dilemma caused by dark shading, consider the example in Fig. 4. In the ideal assumption of zero-mean noise, the mapping target is the expectation of the data. However, when this mapping is applied to real noise, the mapping target includes the corresponding dark shading. If we try to directly force the mapping target to be a clean image (black patch at right), it is equivalent to taking the dark red path, where the mapping dilemma arises. Since the receptive field of the convolutional neural network is limited, we cannot treat different positions differently or use the same mapping to correct all spatial variant patterns caused by dark shading without global position information. Conversely, if we apply DSC to the real noise first, it is equivalent to taking the light green path, where the data mapping is no longer disturbed by dark shading. As a result, DSC significantly reduces the complexity of the noise model, and also addresses the mapping dilemma, resulting in high accuracy in the fitting of data mapping.

By reducing the noise model complexity through DSC, the neural network can accurately fit the data mapping between the paired real data, which promotes denoised images with exact colors.

2) The Principle of Dark Shading: In this part, we introduce and analyze our developed linear dark shading model.

We model dark shading as a linear model based on the physics-based imaging model and experimental experience

$$N_{ds} = N_{FPNk} \cdot iso + N_{FPNb} + BLE(iso, t), \tag{12}$$

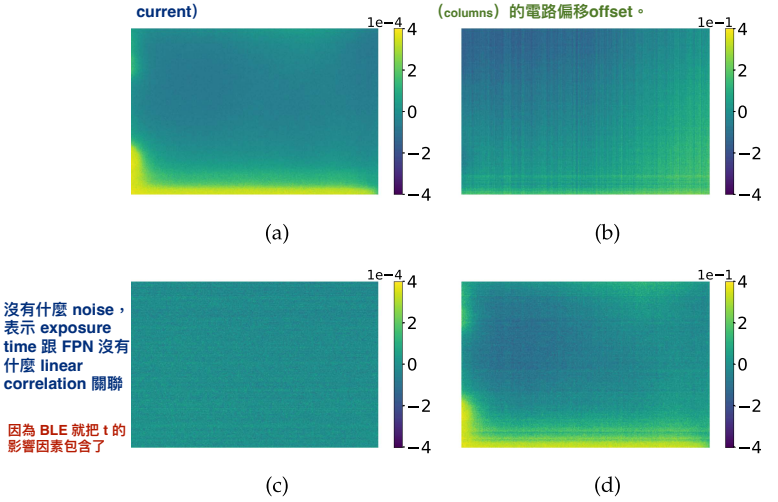
where N_{FPNk} and N_{FPNb} are the coefficient maps of the FPN we need to regress in the dark shading, iso is the ISO value, $BLE(iso, t)$ is the BLE at a specific ISO value iso and exposure time t.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

作者應該是先實驗，才解釋　先透過實驗發現了這種現象，然後用數學模型來描述它，最後再提供物理解釋。

FPN-k 顯示出在高 ISO 下某些區域的 FPN 更強。來自於 temporal stable local heat sources（就是 dark current）

FPN-b 與 ISO 無關，可能來自於感光元件製造時的偏差，例如偏移電壓（Offset Voltage），通常是空間上固定的。顯示出某些感光元件列（rows）或列（columns）的電路偏移offset。



沒有什麼 noise，表示 exposure time 跟 FPN 沒有什麼 linear correlation 關聯

因為 BLE 就把 t 的影響因素包含了

Fig. 5. Linear regression coefficient maps of dark shading obtained according to iso and t, respectively. (a) The iso-dependent component of dark shading N_{FPNk}. (b) The iso-independent component of dark shading N_{FPNb}. (c) The t-dependent component of dark shading. (d) t-independent component of dark shading.

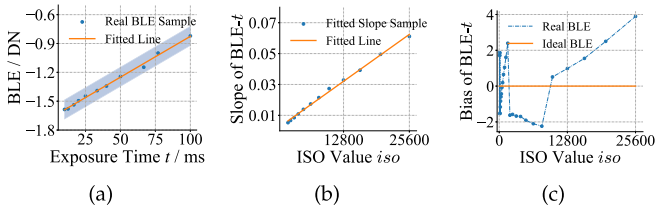


Fig. 6. Linear regression of BLE from real BLE samples at various exposure time t. (a) The calibrated BLE at ISO-3200 based on real BLE samples (blue dots). During training, we randomly sample the BLE in the blue shadow region, whose width is determined by σ_t. During inference, we use the BLE on the fitted line. (b) The calibrated $k_t(iso)$, each blue dot corresponds to a slope value in (a) at an ISO. (c) The calibrated $b_t(iso)$, each blue dot corresponds to a bias value in (a) at an ISO.

According to studies from the electronic imaging community [39], [44], [45], [53], FPN can be classified into dark current FPN and offset FPN. Since the dark current FPN is generated before the amplifier and the offset FPN is generated after the amplifier, FPN can be expressed as $N_{FPNk} \cdot iso + N_{FPNb}$, where N_{FPNk} is iso-dependent FPN and N_{FPNb} is iso-independent FPN. The dark current is also linearly related to the exposure time, however, the dark current FPN is mainly caused by temporal stable local heat sources, which is less affected by the exposure time [45].

To verify the relationship between FPN and these variables (ISO value iso and exposure time t), we design a simple experiment. We collect dark frames with different exposure times at various ISO and then decouple the iso-dependent and t-dependent components via linear regression. As shown in Fig. 5(a) and (b), N_{FPNk} contains the non-uniform pattern noise corresponding to the dark current FPN, and N_{FPNb} contains the banding noise corresponding to the offset FPN, which support our linear dark shading model. As shown in Fig. 5(c) and (d), there is only residual noise in the t-dependent component, and FPN is obviously left in the t-independent component, which indicates that there is no clear linear correlation between the exposure time and the FPN in the dark shading.

The BLE is directly related to the dark current, and therefore it is dependent on both ISO value iso and exposure time t, which can be modeled as $BLE(iso, t)$. To improve the robustness of the noise model, a Gaussian distribution with slight variance σ_t is used to approximate the small temperature fluctuations in practice [51], [57]. We calibrate various sensors to estimate the representation of BLE, as shown in Fig. 6 for the SonyA7S2. Based on the analysis and experiment, we approximate BLE as

$$BLE(iso, t) = k_t(iso) \cdot t + b_t(iso), \tag{13}$$

where $k_t(iso)$ and $b_t(iso)$ represent the slope and the bias of the linear BLE varying with iso, respectively.

kt(iso), bt(iso) 只與 ISO 相關，不受曝光時間 t 影響。實作上應該是用自己建立的 table 查找。

BLE 得出來的是一個 scalar

先做 BLE，再做 FPN

3) The Procedure of DSC: In this part, we introduce the procedure of dark shading correction.

First, we calibrate the linear BLE following (13). To begin with, we collect a dark frame at each ISO and some possible exposure time. Then we calculate the average of the whole image as the BLE under this setting. At last, we estimate the parameters of linear BLE. The error caused by the number of frames is negligible according to our experiments. Second, we collect dark frames in a lightless environment with an exposure time of 1/40 seconds. Then we average multiple dark frames at different ISO to obtain the calibration materials of dark shading $\widehat{N}_{ds}$ based on the zero-mean property of the temporal noise.

Finally, we subtract BLE from calibration materials $\widehat{N}_{ds}$ and estimate the coefficient maps N_{FPNk} and N_{FPNb} according to (12).

The proposed linear dark shading model has two advantages over directly obtaining calibration material for dark shading at all ISO. First, the linear dark shading model acts as a regularization for dark shading, reducing the temporal variant noise that remains in the calibrated dark shading. Second, the linear dark shading model allows us to reduce the amount of data collection required for calibration, meaning that collecting dark frames at partial ISO can still complete the calibration process. In this work, we collect dark frames for calibration only at the ISO from the set $\{100 \times 2^n | n \in \mathbb{N}, 0 \leq n \leq 8\}$.

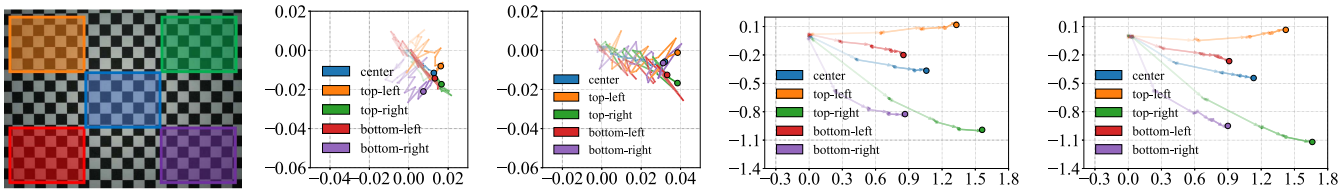
Once the dark shading has been obtained, we can reconstruct it at any ISO and exposure time using (12), and then correct it from the real raw images before denoising. The purpose of DSC is to decouple the read noise model $\mathcal{F}_{read}$ into the temporal stable dark shading and the temporal variant read noise, which will reduce the complexity of noise model, thereby enhancing the learnability of data mapping.

DSC 處理的都是 temporal stable
SNA 才是 temporal variant

IV. Low-Light Raw Image Denoising (LRID) Dataset

The underdeveloped data quality is also one of the culprits for the fragile learnability of data mapping between the paired real data. The underdevelopment is reflected in four data flaws: spatial misalignment, intensity misalignment, noisy ground truth, and insufficient diversity, leading to incorrect data mapping, biased data mapping, poor convergence performance, and overfitting denoising models, respectively.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.



(a) checkboard (b) OIS off, short time (c) OIS on, short time (d) OIS off, long time (e) OIS on, long time

Fig. 7. Checkboard in a closed space without perceivable vibration. Five colored regions correspond to regions for sub-pixel offset estimation. We plot the trajectories of five colored regions compared to the initial state at different OIS states and duration. All trajectories start at the origin (0,0) and end at the dot with a black border.

Existing denoising datasets suffer from significant problems with at least one of these flaws. As a result, the existing datasets are difficult to meet the needs of low-light denoising from a data perspective. Our motivation is to develop the image acquisition protocol and build a high-quality dataset for low-light raw image denoising from a data perspective. Our protocol addresses the data flaws via a well-designed image capture setup. Our dataset will be used to systematically evaluate the performance of low-light raw image denoising methods.

In this section, we first detail the image acquisition protocol of the dataset (Section IV-A), then introduce our noise estimation pipeline considering dark shading (Section IV-B), and finally show the framework of our ground truth estimation method (Section IV-C).

A. Image Acquisition Protocol

We build a high-quality dataset using the Redmi K30 smartphone with the IMX686 sensor. The Low-light Raw Image Denoising (LRID) dataset contains 138 scenes, including 82 indoor and 56 outdoor scenes, with a total of 5754 images. We first captured 25 long-exposure images at ISO-100 and immediately captured several groups of short-exposure images at ISO-6400. Finally, a pair of long-exposure images before and after the original ISP of the smartphone is captured for real-world low-light image enhancement. We use a program to remotely control the smartphone, and the interval of image acquisition is very short (about 0.01s per frame), which means that misalignment between short-exposure frames is negligible.

The indoor scenes are captured in enclosed spaces with various color temperatures and illumination setups. There are five groups of short-exposure images for each scene, and the exposure time ratios of long- and short-exposure images are 64, 128, 256, 512, and 1024, respectively. The total exposure time of the long-exposure images is about 25s.

The outdoor scenes are captured at midnight with the calm wind (within 0.5m/s). There are three groups of short-exposure images for each scene, and the exposure time ratios of the long- and short-exposure images are 64, 128, and 256, respectively. The total exposure time of the long-exposure images is about 64s.

Our image acquisition protocol is superior due to addressing four data flaws that affect learnability. The maximum total exposure time and minimum exposure time are limited by spatial alignment (Section IV-A1) and intensity alignment (Section IV-A2), respectively. The image capture setup of long-exposure images and short-exposure images are designed for clean ground truth (Section IV-A3) and sufficient diversity (Section IV-A4), respectively.

1) Spatial Alignment: Previous works have reported that there will be obvious spatial misalignment during the image acquisition of denoised datasets. SIDD [48] believes that the spatial misalignment is caused by the Optical Image Stabilization (OIS) of the smartphone, and calling the API to disable OIS does not work. Leaving all the spatial alignment to post-processing is risky, thus we reduce spatial misalignment as much as possible during the image acquisition.

To further examine the spatial misalignment, we place the smartphone on a tripod in an enclosed space without perceivable vibration, the same as the image capture environment for indoor scenes. We remotely control the smartphone to collect 5 sets of checkerboard images at one-minute intervals. Each set contains 50 images taken in rapid succession, where 25 images with OIS on and 25 images with OIS off.

We count the spatial misalignment of five regions with a sub-pixel image registration method [1] as shown in Fig. 7(a). A comparison of Fig. 7(b) and (c) shows the spatial misalignment affected by the OIS state for a short time (about 5 seconds). We find that calling the API to turn off OIS is effective. The image misalignment after turning off OIS is noticeably smaller than turning on OIS. The comparison of Fig. 7(d) and (e) shows the spatial misalignment affected by the OIS state for a long time (about 5 minutes). We find that the spatial misalignment of the image obviously exists regardless of the state of OIS. Comparing Fig. 7(b)–(c) and Fig. 7(d)–(e), we find that long-term spatial misalignment is much larger than that of short-term. The above observations indicate:

- Turning off OIS can reduce spatial misalignment during image acquisition, and smartphones can be controlled to switch OIS. The extra movement probably comes from the signal disturbance of OIS.
- As long as the image acquisition takes a long time, there will be obvious spatial misalignment. Unpredictable physical environment vibrations are perhaps the main contributors to spatial misalignment [58]. The frictional damping may prolong the reset process.

In conclusion, the total exposure time needs to be as short as possible to reduce spatial misalignment.

2) Intensity Alignment: We find that too long total exposure time (e.g., 4 minutes) always introduces unintended intensity misalignment, such as light switches in buildings, irradiance fluctuations of the sky, etc. In order to avoid various unintended

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

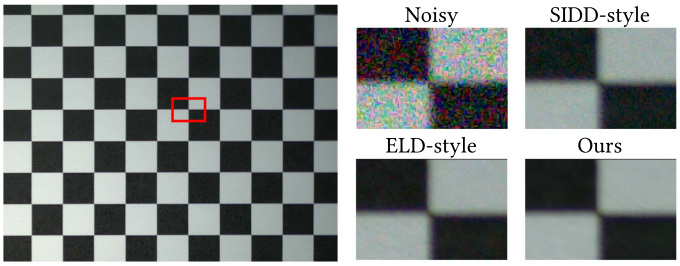


Fig. 8. Comparison of different styles of dataset production methods. "Noisy" is an image captured at ISO-6400. "SIDD-style" is the fusion of 64 images collected at ISO-6400. "ELD-style" is an image captured at ISO-100. "Ours" is the fusion of 25 images captured at ISO-100. **(Best viewed with zoom-in)**

頻閃 (stroboscopic effect) 是指來自交流電 (AC) 光源的亮度變化，導致影像亮度不匹配 (intensity misalignment)，半個週期 (half period) 為 10ms，因此如果曝光時間設置為 10ms，則能夠在曝光期間內平衡光強的變化，減少亮度不匹配的問題。

low-frequency intensity misalignments as well as spatial misalignments, we limited the maximum total exposure time to 64s. Furthermore, existing datasets pay little attention to the stroboscopic problem introduced by AC light sources when collecting outdoor scene data. Outdoor scenes in the real world are full of various AC light sources. The intensity misalignment introduced by the strobe is usually spatial variant, which cannot be removed by post-processing. The integral result is stable when the integral interval is the half period of the periodic function. The line frequency is 50Hz, thus we set the minimum exposure time to 10ms in any scene containing AC light sources.

3) Clean Ground Truth: Previous works have different perspectives for estimating clean ground truth. We classify them into two typical styles: "ELD-style" [21] and "SIDD-style" [48]. "ELD-style" collects long-exposure images at low ISO as clean ground truth. "SIDD-style" fuses multiple frames of noisy images (typically captured at high ISO) to estimate clean ground truth. "ELD-style" is less noisy with the same total exposure time, while "SIDD-style" can handle various misalignments flexibly. In order to obtain a clean and clear ground truth as fast as possible, we integrate the advantages of "ELD-style" and "SIDD-style". Our clean ground truth is the fusion of multiple raw images ("SIDD-style") captured at low ISO ("ELD-style"). As shown in Fig. 8, "ELD-style" is cleaner than "SIDD-style" under the same total exposure time. "Ours" performs the cleanest and clearest result. The comparison demonstrates the necessity of our long-exposure image capture setup.

4) Sufficient Diversity: Data diversity can be seen in two aspects: scene diversity and noise diversity. However, different datasets collected with different cameras may have different noise models, therefore it is difficult to compare them quantitatively. To further demonstrate the sufficient data diversity of our dataset, we show a subset of our dataset in Fig. 9 and design a series of ablation studies as shown in Fig. 10. We simulate different image acquisition protocols by reducing the data volume of the dataset in different dimensions. We train neural networks with paired real data under different data schedules and regard the denoising performance of neural networks as an evaluation index of data quality.

Our ablation studies of noise diversity and scene diversity show that data quality improves with an increasing number of noisy images per scene and an increasing percentage of scenes. The increase of data quality gradually saturates when the number of noisy images is close to our value, indicating that our dataset has sufficient noise diversity and scene diversity. Most existing image denoising datasets pay little attention to the number of noisy images. The limited data volume leads to the fragile learnability of paired real data in these datasets, but not in our dataset. The ablation studies demonstrate the necessity of our short-exposure image capture setup.

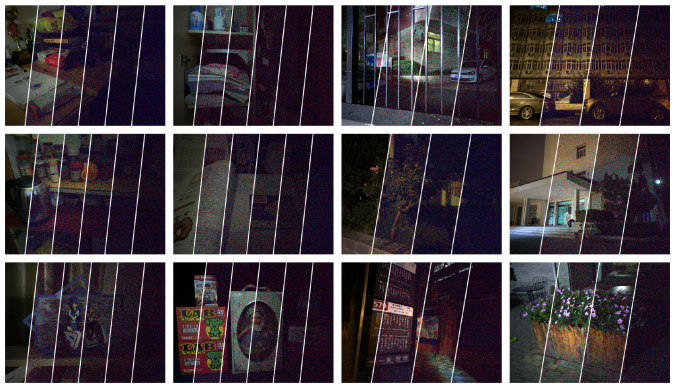


Fig. 9. Examples in our LRID dataset. Indoor images are in the left two columns, and outdoor images are in the right two columns. Clean ground truths are shown in the upper left. Short-exposure input images with different ratios are shown in the bottom right of each image. **(Best viewed with zoom-in)**

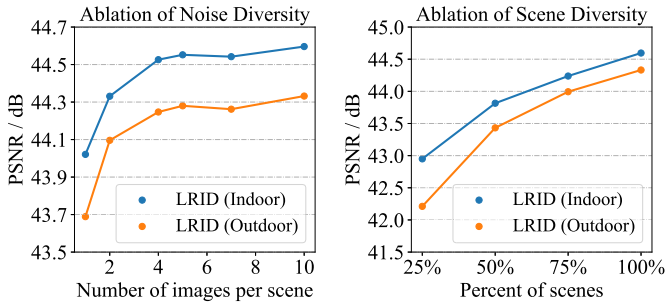


Fig. 10. Ablation study of different image acquisition protocols.

B. Noise Estimation

Previous noise estimation work [21] ignores the influence of temporal stable dark shading during the calibration of temporal variant noise. We propose a developed noise calibration method.

We record flat-field frames to estimate system gain K via the Photon Transfer method [23] and record dark frames to estimate other noise parameters. We need to calibrate dark shading with dark frames before estimating other noise parameters. The details have already been introduced in Section III-C3. Next, in order to avoid calibration errors caused by dark shading, we need to apply DSC to each dark frame before the noise parameter calibration. Finally, we apply the existing calibration methods [2], [21], [38] to estimate the noise parameters. We recorded the noise parameters K and σ_{read} according to P-G [38] at ISO-100 for the later ground truth estimation.

We further develop the method of defective pixel detection [48] based on dark shading. We use the median filter to remove the high-frequency component of dark shading to obtain the low-frequency component, where exists no outliers including defective pixels. Then we subtract the low-frequency

上述的 BLE, FPN

預測 SNA 會用到的參數

對 Dark shading 做偵測 (為了後續的 GT estimation)

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

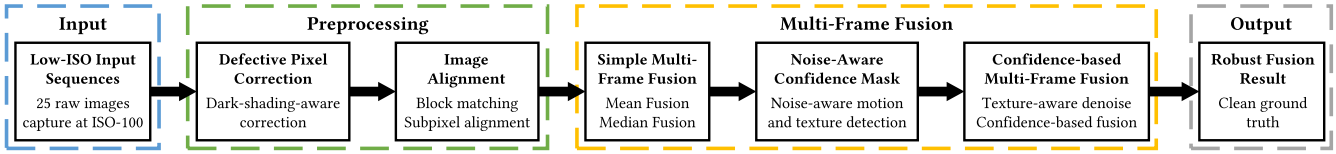


Fig. 11. Our ground truth estimation pipeline. We first preprocess each input raw image, and then fuse multiple frames as output.

component from dark shading to obtain the high-frequency component, which hardly contains low-frequency spatial bias. Finally, we consider all pixels having values outside $6\sigma_{read}$ range in high-frequency components as defective pixels and record their positions for defective pixel correction.

The developed method corrects the zero-mean assumption of read noise in previous works [2], [21], [38], [48] and takes into account the low-frequency spatial bias introduced by dark shading, which can accurately estimate the noise parameters of the sensor.

C. Ground Truth Estimation

Our ground truth estimation pipeline is shown in Fig. 11. In Section IV-C1, we introduce the preprocessing details in ground truth estimation, including defective pixel correction and image alignment. In Section IV-C2, we introduce the details of the robust multi-frame fusion, including simple multi-frame fusion, noise-aware confidence mask, and confidence-based multi-frame fusion.

1) Preprocessing: Defective Pixel Correction: The details of our dark-shading-aware defective pixel detection have shown in Section IV-B. The defective pixels will be removed by a median filter at the recorded position. The correction of defective pixels not only improves the image quality but also enhances the accuracy of image alignment and confidence mask, which is important for ground truth estimation.

Image Alignment: Indoor scenes are typically highly stable, with only minor sub-pixel shifts under our image acquisition protocol. However, outdoor scenes are often subject to significant global and local spatial misalignment due to inevitable environmental vibrations. Therefore, spatial alignment algorithms are necessary for ground truth estimation.

We regard the last frame in the sequence of raw images as the reference frame. Since the image acquisition interval is very short, the reference frame can be assumed to be perfectly aligned with the noisy image. All other frames in the sequence are then aligned to the reference frame using the block-matching-based subpixel alignment algorithm [1], which is robust to noise.

Fig. 12 compares the impact of alignment on ground truth estimation. In the result of multi-frame mean fusion, the leaves are clearer than the leaves before alignment, which demonstrates that alignment can effectively compensate for obvious spatial misalignment. However, we also observe that for complex misalignment with occlusions, alignment may introduce extra errors. Therefore, we introduce confidence masks in the multi-frame fusion stage to further improve the quality of the ground truth.

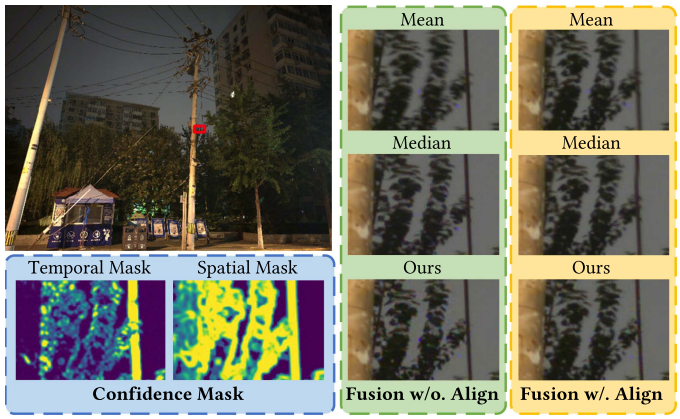


Fig. 12. Results in the confidence-based multi-frame fusion processing. The blue region represents the confidence masks. The green region represents the fusion results without alignment. The yellow region represents the fusion results with alignment. The final result of our ground truth estimation is "Ours" with alignment. **(Best viewed with zoom-in)**

2) Multi-Frame Fusion: Simple Multi-frame Fusion: We temporally filter the input images and sequentially obtain multi-frame mean fusion and multi-frame median fusion. The former is less noisy in static regions, and the latter is clearer in motion regions.

Noise-Aware Confidence Mask: The noise-aware confidence mask consists of two components: a temporal mask for motion detection and a spatial mask for spatial frequency detection.

一個是運動 mask，另一個是紋理與背景mask

These masks are detected using a soft threshold and the calibrated noise parameters K and σ_{read} from the Variance-Stabilizing Transform (VST) [40], [41]. The temporal mask is obtained by comparing the difference between multi-frame mean fusion and multi-frame median fusion (both after VST). The spatial mask is determined by calculating the spatial variance of the ground truth after VST. Both the temporal mask and spatial mask are Gaussian blurred and clipped to ensure smooth transitions.

Poisson 噪聲(原始圖像)的變異數與訊號強度（像素值）成正比。

使用 VST 後，整體的Variance比較 stable，可以避免因為訊號區域的變異數較大，而誤判 mask。

Confidence-based Multi-frame Fusion: We use multi-frame median fusion as the guide image to denoise the reference frame, where the parameters of the guided filter [59] vary with the spatial mask. This results in the denoised reference frame.

We follow the principles below to fuse different results according to the weights of confidence masks to balance noise and misalignment:

1) In regions with high confidence mask weights, the reference frame and multi-frame median fusion will have high weights.
2) In regions with low confidence mask weights, denoised image and multi-frame mean fusion will have high weights.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

TABLE I
QUANTITATIVE RESULTS (PSNR/SSIM) OF DIFFERENT METHODS ON THE ELD DATASET, SID DATASET, AND OUR LRID DATASET

Dataset	Ratio	Input PSNR / SSIM	P-G [38] PSNR / SSIM	ELD [21] PSNR / SSIM	SFRN [46] PSNR / SSIM	Paired PSNR / SSIM	Ours PSNR / SSIM
ELD	×100	30.85 / 0.5045	42.05 / 0.8721	45.45 / 0.9754	46.38 / 0.9793	44.47 / 0.9676	46.99 / 0.9840
	×200	25.92 / 0.2607	38.18 / 0.7827	43.43 / 0.9544	44.38 / 0.9651	41.97 / 0.9282	44.85 / 0.9686
	Average	28.38 / 0.3826	40.12 / 0.8274	44.44 / 0.9649	45.38 / 0.9722	43.22 / 0.9479	45.92 / 0.9763
SID	×100	29.10 / 0.5266	39.44 / 0.8995	41.95 / 0.9530	42.81 / 0.9568	42.06 / 0.9548	43.47 / 0.9606
	×250	23.95 / 0.3595	34.32 / 0.7681	39.44 / 0.9307	40.18 / 0.9343	39.60 / 0.9380	41.04 / 0.9471
	×300	22.00 / 0.2752	30.66 / 0.6569	36.36 / 0.9114	37.09 / 0.9175	36.85 / 0.9227	37.87 / 0.9344
	Average	24.81 / 0.3793	34.52 / 0.7666	39.05 / 0.9303	39.82 / 0.9349	39.32 / 0.9374	40.59 / 0.9465
LRID-Indoor	×64	32.81 / 0.6728	46.14 / 0.9872	48.19 / 0.9898	47.94 / 0.9899	48.77 / 0.9906	49.24 / 0.9916
	×128	29.10 / 0.4621	44.98 / 0.9809	46.55 / 0.9836	46.52 / 0.9854	47.00 / 0.9860	47.47 / 0.9868
	×256	25.07 / 0.2380	43.31 / 0.9682	44.39 / 0.9730	44.74 / 0.9789	44.74 / 0.9786	45.36 / 0.9804
	×512	20.53 / 0.0872	40.80 / 0.9429	41.56 / 0.9452	42.46 / 0.9652	42.40 / 0.9647	43.09 / 0.9671
	×1024	15.43 / 0.0241	37.74 / 0.8905	37.50 / 0.8915	40.10 / 0.9453	40.07 / 0.9437	40.20 / 0.9453
	Average	24.59 / 0.2968	42.59 / 0.9539	43.64 / 0.9566	44.35 / 0.9729	44.60 / 0.9727	45.07 / 0.9743
LRID-Outdoor	×64	33.25 / 0.7255	42.16 / 0.9796	45.00 / 0.9841	45.05 / 0.9850	45.84 / 0.9876	46.27 / 0.9884
	×128	29.49 / 0.5100	41.48 / 0.9709	43.48 / 0.9734	43.67 / 0.9766	44.50 / 0.9821	44.86 / 0.9834
	×256	25.26 / 0.2557	40.36 / 0.9525	41.31 / 0.9450	41.89 / 0.9591	42.66 / 0.9709	42.99 / 0.9703
	Average	29.33 / 0.4971	41.33 / 0.9677	43.26 / 0.9675	43.54 / 0.9736	44.33 / 0.9802	44.71 / 0.9807

The red color indicates the best results and the blue color indicates the second-best results.

3) In the regions with the largest temporal mask weight, only the reference frame is used.
4) In the regions with the smallest spatial mask weight, only the denoised frame is used.

To demonstrate the robustness of our methods, we use an outdoor scene from our dataset and magnify the region with the most significant motion. Fig. 12 shows the intermediate results of our confidence-based multi-frame fusion process. The blue region represents the confidence mask, which forms the basis for our robust multi-frame fusion. The response of the temporal mask is concentrated on the edges of swaying branches and swaying lines, effectively detecting regions of motion. The response of the spatial mask is concentrated on high-contrast regions, effectively detecting texture regions. Benefiting from the noise-aware confidence mask, our methods produce cleaner and sharper results than mean fusion and median fusion, demonstrating superior performance with fewer artifacts.

V. EXPERIMENT

In this section, we first introduce the experimental setting including implementation details and compared methods. Then, we compare our methods against prior art with quantitative results and visual quality on public datasets and our dataset. Finally, we conduct comprehensive ablation studies for an in-depth analysis of our methods.

A. Experimental Setting

1) Implementation Details: We use the same UNet architecture [60] as SID [8] and ELD [21]. On public datasets, raw images from the SID Sony training set are used to create training data. The quantitative results are reported on the ELD Sony dataset and the whole SID Sony dataset, including validation and test sets. On our dataset, 90% of the scenes are used for training, and 10% of the scenes are used for validation. Since the image acquisition protocols of outdoor and indoor scenes are different, we separately illustrate their results. We tabulate the performance of denoising models at different exposure ratios with the averaged results in Table I.

DSC is applied to the noisy raw images, and SNA is only used for training. There is a 75% probability that the training data pairs will be augmented by SNA. The calibration of the SID dataset and ELD dataset are based on a SonyA7S2 camera, which has the same sensor as the public datasets but not the same camera. We collect 400 dark frames per ISO for dark shading calibration[2] and set σ_t to 0.1. We pack the raw Bayer images into four channels and sample non-overlapped 512×512 patches of each image, then randomly rotate and flip them as a batch. We visualize the raw images as sRGB images for viewing through Rawpy (a Python wrapper for LibRaw) with the metadata of ground truth following the existing works [8], [21]. The quantitative results are computed on the raw domain.

Our implementation is based on PyTorch. We train denoising models with 600 epochs using Adam optimizer [61] and L_1 loss. Each epoch contains data pairs of different scenarios at different ratios. The learning rate will vary with iterations like [62]. The base learning rate is set to 2×10^{-4} and the minimum learning rate is set to 10^{-5}. The optimizer restarts every 200 epochs and the learning rate is halved on restarts.

2) Compared Methods: In order to demonstrate the reliability of our strategy, we compare our methods with:

- The denoising model trained with synthetic data based on different noise models, including Poisson-Gaussian (P-G) [38] and ELD [21], which is the classical method for low-light raw image denoising.
- The denoising model trained with the half-real data proposed by SFRN [46], which is the state-of-the-art method before our work.

[2]We empirically find the performance varies less than 0.02% when the number of dark frames is more than 64 for dark shading calibration.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

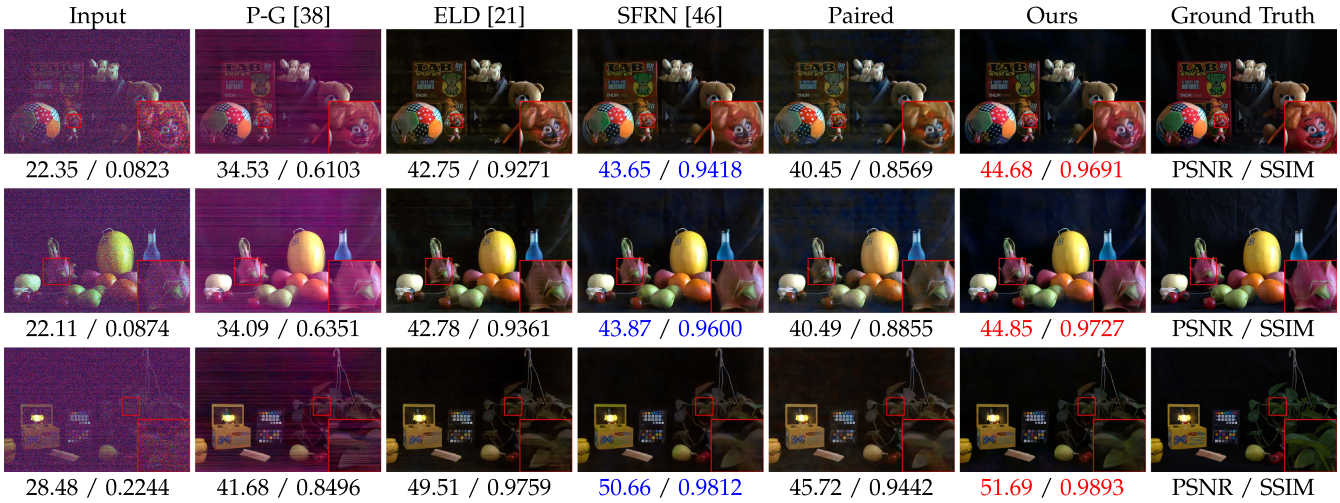


Fig. 13. Low-light raw image denoising results on images from the ELD dataset. **(Best viewed with zoom-in)**

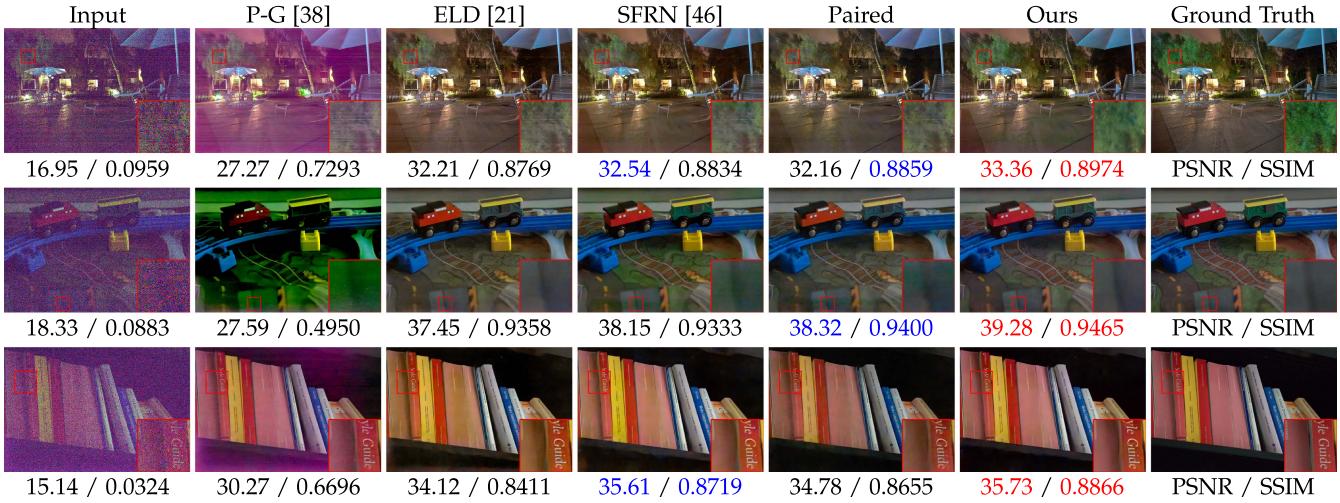


Fig. 14. Low-light raw image denoising results on images from the SID dataset. **(Best viewed with zoom-in)**

- The denoising model trained with paired real data (i.e., Paired), for which the learnability enhancement strategy needs to serve.

The results of P-G, ELD, and Paired are implemented from the code and weights released by the open-source project. Since the training code and weights of SFRN have never been released, we reproduce the SFRN with the help of the authors, and the results of the experiments have been confirmed by the authors. The SFRN in this paper utilizes much more dark frames and thus performs better than the records in our preliminary work [27].

B. Results

Table I summarizes the performances of denoising models trained with the data coming from different data schedules. The denoising results on the public datasets are shown in Figs. 13 and 14. Denoising models trained with synthetic data are unable to completely remove complicated real noise. P-G is far from the real noise model, resulting in limited performance. ELD considers more noise sources but still deviates from the real noise model, resulting in color bias and residual noise. Although SFRN sampled real read noise, the patch-wise method cannot inherently address the mapping dilemma caused by dark shading, resulting in residual FPN. The paired real data, despite containing real noise, is so fragile in learnability that the denoising model cannot learn the precise and accurate data mapping, resulting in blurry results and color bias. By applying the learnability enhancement strategy to the paired real data, the denoising performance is significantly improved in both quantitative results and visual quality. Our work performs clean denoising results with the clearest texture and the most exact colors.

As shown in Fig. 15, our methods demonstrate superior denoising performance on our dataset, with the clearest texture and most exact colors. The performance aligns with our results on public datasets, indicating the high generalizability of our methods.

It is worthwhile to highlight that both the SID dataset and ELD dataset are collected using the SonyA7S2, yet the results

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

Fig. 15. Low-light raw image denoising results on images from our LRID dataset. **(Best viewed with zoom-in)**

of training with paired real data show significant differences in performance on the two datasets. On the SID dataset, the denoising model trained with paired real data performs close to SFRN. However, on the ELD dataset, the denoising model trained with paired real data performs even worse than ELD. Overall, our experiments draw a conclusion for this observation. The data flaws caused by the underdeveloped image acquisition protocol of the SID training set indeed lead to the underdeveloped data quality of paired real data. The paired real data with fragile learnability further leads to unreliable data mapping, which lowers both the denoising performance and generalizability of the denoising model.

On our dataset, paired real data generally outperforms the best synthetic data (SFRN), which indicates that the data quality of synthetic data is inferior to that of paired real data on our dataset. This result differs from the findings of other noise modeling studies on public datasets. We attribute this discrepancy to the fact that our image acquisition protocol has addressed various data flaws. When data is well-annotated (i.e., the ground truth is clean and well-aligned) and sufficient in quantity, the learnability of data mapping is well-guaranteed. Under these conditions, the data quality of paired real data should surpass that of synthetic data. The observation indicates the superiority of our dataset.

In conclusion, our methods can achieve state-of-the-art performance on various datasets and significantly outperform previous methods. Extensive experiments demonstrate the essence of our learnability enhancement strategy, which increases the data volume via SNA, reduces the noise complexity via DSC and improves the data quality via our image acquisition protocol. By reforming paired real data according to noise modeling, our methods significantly improve the denoising performance and generalizability.

C. Ablation Study

To understand the individual contributions of each module, we compare the performance of neural networks trained with or without SNA and DSC. The comparison also includes preliminary versions of these modules, denoted as SNA* and DSC*, respectively. Fig. 16 summarizes the quantitative results of the neural networks trained with these different data schemes, while Fig. 17 shows a representative visual comparison of different data schemes.

Without any learnability enhancement, the denoising model trained with paired real data seems blurry and noisy. It is difficult for the neural network to learn a precise and accurate mapping from the paired real data with fragile learnability. SNA typically brings a small improvement in quantitative results but promotes denoised images with clear texture. The improvement in visual quality is attributed to the precise fitting benefiting from the increased data volume. Compared to the previous version of SNA, denoted as SNA*, our updated version, SNA, has a more complete range of augmentation, resulting in high mapping precision. However, the mapping dilemma still exists, causing the neural network to overfit a biased mapping augmented by SNA. The overfitted biased mapping sometimes leads to worse results. DSC typically brings a significant improvement in quantitative results and promotes denoised images with exact colors. The improvement in quantitative results is attributed to the accurate fitting benefiting from the reduced noise complexity.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

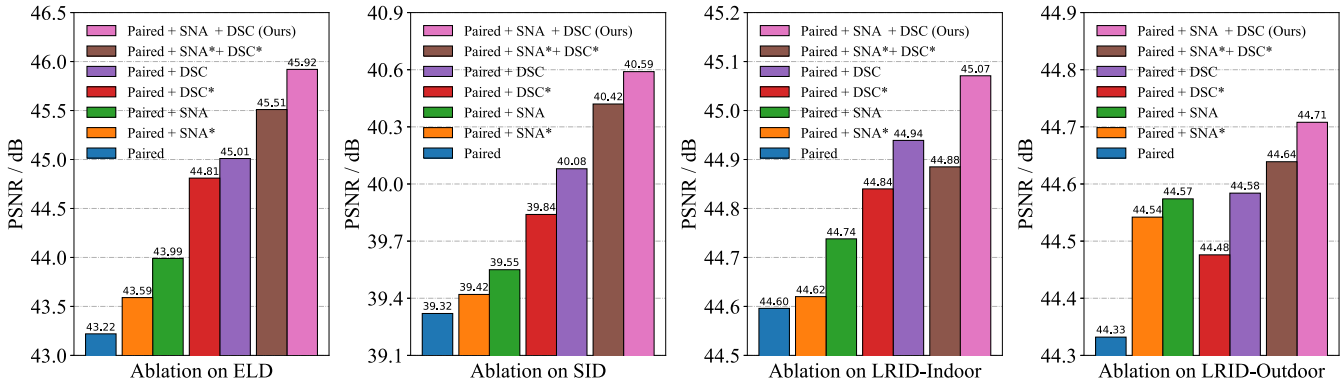


Fig. 16. Ablation study of different learnability enhancement modules on the ELD dataset, SID dataset, and our LRID Dataset. "*" indicates that the module uses the implementation from the preliminary version.

Fig. 17. Representative visual result comparison of different data schemes. "*" indicates that the module uses the implementation from the preliminary version. **(Best viewed with zoom-in)**

Compared to the previous version of DSC, denoted as DSC*, our updated version, DSC, has a more accurate dark shading model, resulting in high mapping accuracy. However, DSC does not significantly promote denoising precision, which relies on SNA. Overall, the best performance is achieved using the complete learnability enhancement strategy. SNA focuses on promoting mapping precision, while DSC focuses on promoting mapping precision. The combination of these two modules results in the best performance in both quantitative and visual quality. Benefiting from the development of SNA and DSC, our learnability enhancement strategy successfully refreshes the state-of-the-art in our preliminary work [27].

VI. Discussion of Dark Shading Correction

Various practical problems encountered in the application of DSC will be discussed in this section. We believe these discussions are helpful for implementing learnability enhancement flexibly and efficiently. Section VI-A shows the dark shading calibrated by different cameras of the same sensor and reports the denoising results based on these dark shading. Section VI-B shows dark shading on more different sensors. Section VI-C shows the results of developing noise models with DSC. Section VI-D discusses the limitation of DSC.

A. Dark Shading on Cameras With Same Sensor

Our dark shading is calibrated by a different camera with the same sensor as public datasets, thus it is necessary to verify the consistency of dark shading calibrated by different cameras with the same sensor.

We collect four different SonyA7S2 cameras for dark shading calibration, and the quantitative results have been shown in Fig. 18. The camera used in this paper is SonyA7S2-4. Although the dark shading of different cameras has a little difference, the pattern of dark shading is stable within the normal operating temperature range.

Table II shows the denoising results of the neural networks trained based on the dark shading of different cameras. Dark shading calibrated by different cameras leads to different

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

different cameras with the same sensor quantitative results are close

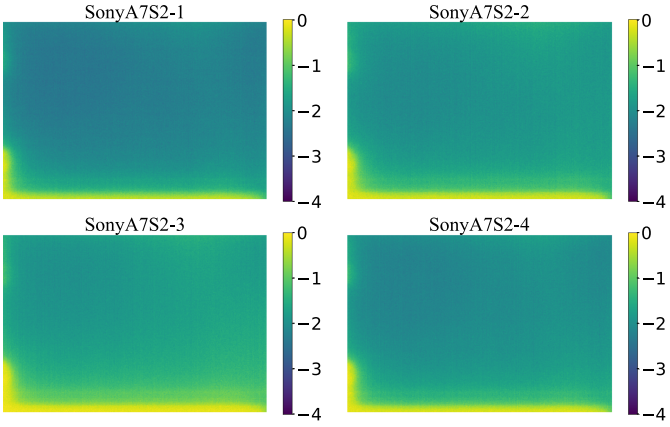


Fig. 18. Visual comparison of dark shading at ISO-3200 on different SonyA7S2 cameras. The camera used in this paper is SonyA7S2-4.

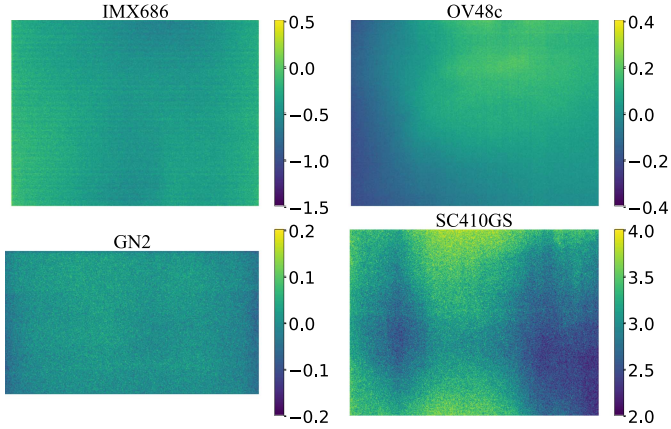


Fig. 19. Visual comparison of dark shading on different sensors.

TABLE II
QUANTITATIVE RESULTS (PSNR/SSIM) OF DIFFERENT CAMERAS ON THE ELD DATASET AND SID DATASET WITH DIFFERENT EXPOSURE RATIOS

Camera	Index	ELD		SID		
		×100	×200	×100	×250	×300
SonyA7S2-1	PSNR	46.63	44.90	43.20	40.83	37.68
	SSIM	0.982	0.976	0.959	0.945	0.931
SonyA7S2-2	PSNR	47.04	45.00	43.47	41.10	37.93
	SSIM	0.985	0.971	0.961	0.947	0.934
SonyA7S2-3	PSNR	46.98	44.96	43.32	40.97	37.84
	SSIM	0.985	0.973	0.960	0.947	0.933
SonyA7S2-4	PSNR	46.99	44.85	43.47	41.04	37.87
	SSIM	0.984	0.969	0.961	0.947	0.934

The camera used in this paper is SonyA7S2-4.

denoising performances, however, their quantitative results are close, which is still significantly higher than previous works. This comparison demonstrates the high consistency of dark shading calibrated by different cameras with the same sensor, which indicates that our DSC is feasible under the above configuration.

B. Dark Shading on Different Sensors

To demonstrate our DSC is indispensable (不可或缺), we verify that dark shading is widespread on different sensors.

According to our observation, dark shading with noticeable patterns widely exists in sensors of various mainstream sensors. We list the dark shading of more sensors in Fig. 19, among which IMX686, OV48c, and GN2 are the mainstream sensors on smartphones in recent two years and SC410GS is used in surveillance cameras. Since the noise characteristics of different sensors are different, we only show the dark shading at maximum system gain and limit the range of values for viewing. The sensors widely used on smartphones and surveillance cameras also contain noticeable dark shading, which indicates that our DSC is indispensable.

C. Extension of DSC on Noise Modeling

Dark shading, the comprehensive modeling of temporal stable noise, is also an extension of physical-based noise modeling. P-G [38] only considers temporal variant noise. Some studies [21], [51] consider the BLE but pay little attention to the temporal stable FPN. SFRN [46] implicitly covers dark shading via real read noise samples. Some studies [4], [63] have even modeled simplified dark shading, however, they do not realize that it is essential to remove the global fixed pattern noise before denoising. Therefore, the learnability of synthetic data based on these noise modeling methods is still fragile. The extended application of DSC on noise modeling can significantly improve the performance of existing noise modeling methods.

We find that the lack of learnability can be solved by extending the existing physics-based noise model with DSC:

- For noise modeling methods where temporal stable read noise has never been considered (e.g., P-G [38]), DSC can be applied directly in the inference stage without changing the training strategy.
- For noise modeling methods where BLE has been considered (e.g., ELD [21]), their training strategies need not be changed. DSC can be applied directly in the inference stage with dark shading without BLE.
- For noise modeling methods where dark shading has been considered (e.g., SFRN [46]), only temporal noise in the noise model should be used during training, and DSC should be applied during inference.

It is worthwhile to be highlighted that SFRN requires High-Bit Recovery (HBR) to work, however, HBR and DSC conflict in implementation. The efficient implementation of HBR relies on the assumption that each pixel of the dark frame is zero-mean in the temporal dimension, but the assumption does not hold due to dark shading. If dark shading is considered, each pixel needs to fit an independent distribution according to the definition of HBR, which would increase the computational complexity to an unacceptable level. A simple solution is to apply DSC first, then quantize the signal, and finally apply HBR, so that the original computational complexity can be maintained. Unfortunately, since dark shading is also high-bit data, quantization will introduce extra errors and reduce the benefits of DSC.

To address this conflict between DSC and HBR, we propose a novel approximation algorithm. We find that HBR is somewhat robust to Probability Density Function (PDF) selection

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

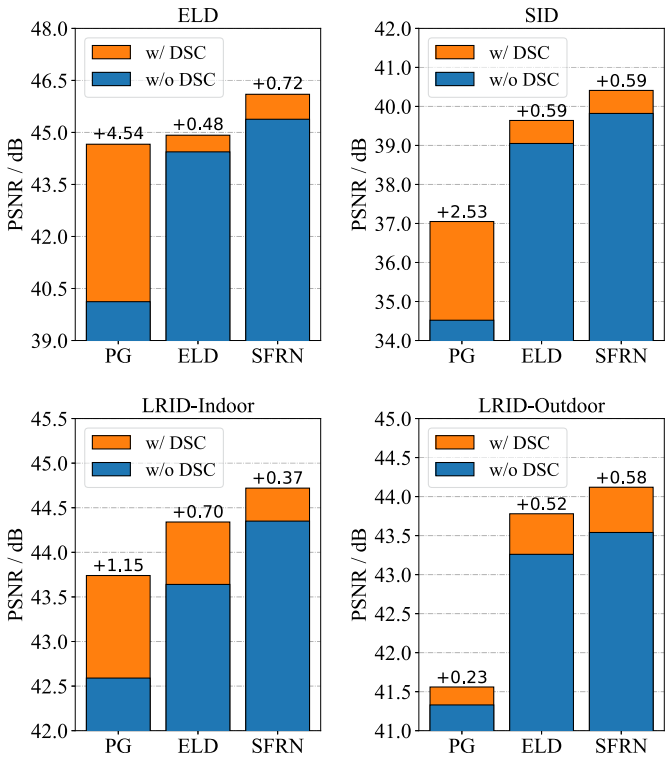


Fig. 20. Quantitative comparison of noise models with and without DSC on various datasets.

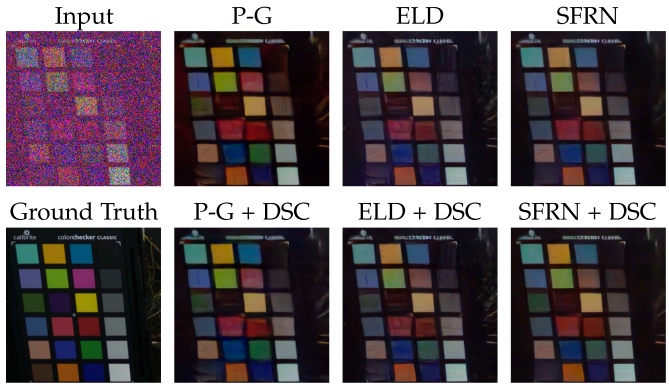


Fig. 21. Representative comparison of noise models with and without DSC on our LRID dataset. **(Best viewed with zoom-in)**

for low-bit to high-bit mapping. Since the signals of the real dark frame are constant, a slight deviation of the PDF has a limited impact on denoising as long as the monotonicity of the PDF is consistent. Therefore, we cache the quantization error before applying HBR, and then compensate for it after applying HBR. This approximation algorithm can effectively avoid extra errors in the quantization process while maintaining the original computational complexity.

The noise modeling method corrected by DSC results in a more realistic noise model, leading to a high denoising performance. DSC brings significant improvements to noise modeling methods in quantitative results on various datasets, as shown in Fig. 20. In general, if the original noise model does not adequately consider the temporal stable noise model, the quantitative results of the developed noise model will be significantly improved. A representative comparison is shown in Fig. 21. Compared with the noise model without DSC, DSC promotes denoised images with fewer artifacts and more exact colors. Our extensive experiments demonstrate the potential widespread usage of our methods.

D. Limitations

Although DSC is efficient and essential for low-light raw image denoising, there are some uncertainties of dark shading due to the variations in complex environments and sensor circuits. Dark shading is mainly affected by dark current, which increases exponentially with temperature [39], [53]. If the temperature of the application scenario is not within the normal operating temperature (within 50°C), especially in the case of overheating, there may be a difference between the actual dark shading and the calibrated dark shading, which will cause the DSC to be less effective. Fortunately, the dark shading of modern sensors is less sensitive to temperature thanks to the dark current suppression techniques [64], [65]. As a result, dark shading can be considered a temporal stable component of read noise. For example, the dark shading of four different cameras, as presented in Section VI-A, is calibrated under varying environmental conditions with different temperatures. Nevertheless, the consistent pattern of dark shading illustrated in Fig. 18 and the comparable denoising performance demonstrated in Table II indicate that the neural network exhibits robustness to dark shading variations within the range of normal operating temperatures.

Meanwhile, we find that some sensor manufacturers apply special processing in imaging, which may cause the calibrated dark shading to be inaccurate. Dark shading, which is closely tied to the layout of the sensor circuit [45], may be deformed if the circuit is switched. The sensor circuit may be switched when the sensor triggers some special conditions. An example is that SonyA7S2, a sensor with the dual-native ISO technique, has two sets of circuits for high ISO and low ISO, respectively. Our linear dark shading model is proposed according to the noise characteristics of electronic components and has strong applicability. Circuit switching usually does not break the linear dark shading model but does change the model parameters. Therefore, the above problems can generally be addressed by additional dark shading calibration for the circuit after switching.

VII. Conclusion

In this paper, we introduce a learnability enhancement strategy for low-light raw image denoising from a data perspective. Our learnability enhancement strategy inherently breaks through the learnability bottleneck of paired real data. Our strategy integrates three efficient methods: SNA, DSC, and a developed image acquisition protocol, which help the neural network efficiently learn the data mapping by increasing the data volume, reducing the noise complexity, and improving the data quality, respectively. Based on the developed image acquisition protocol, we build a new dataset for low-light raw image denoising. Extensive experiments on public datasets and our dataset collectively demonstrate the superiority of our methods on low-light raw image denoising.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] S. W. Hasinoff et al., "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, 2016.

[2] Y. Wang, H. Huang, Q. Xu, J. Liu, Y. Liu, and J. Wang, "Practical deep raw image denoising on mobile devices," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 1–16.

[3] I. S. McLean, *Electronic Imaging in Astronomy: Detectors and Instrumentation*. Berlin, Germany: Springer Science & Business Media, 2008.

[4] B. Moseley, V. Bickel, I. G. López-Francos, and L. Rana, "Extreme low-light environment-driven image denoising over permanently shadowed lunar regions with a physical noise model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 6317–6327.

[5] M. S. Joens et al., "Helium ion microscopy (HIM) for the imaging of biological samples at sub-nanometer resolution," *Sci. Rep.*, vol. 3, no. 1, pp. 1–7, 2013.

[6] C. Li et al., "Low-light image and video enhancement using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9396–9416, Dec. 2022.

[7] H. Huang, W. Yang, Y. Hu, J. Liu, and L.-Y. Duan, "Towards low light enhancement with raw images," *IEEE Trans. Image Process.*, vol. 31, pp. 1391–1405, 2022.

[8] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3291–3300.

[9] H. White, "Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings," *Neural Netw.*, vol. 3, no. 5, pp. 535–549, 1990.

[10] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[11] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[12] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1712–1722.

[13] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, "Variational denoising network: Toward blind noise modeling and removal," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 1690–1701.

[14] S. W. Zamir et al., "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 14821–14831.

[15] S. Cheng, Y. Wang, H. Huang, D. Liu, H. Fan, and S. Liu, "NBNet: Noise basis learning for image denoising with subspace projection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 4896–4906.

[16] Z. Tu et al., "MAXIM: Multi-axis MLP for image processing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5769–5780.

[17] L. Chen, X. Lu, J. Zhang, X. Chu, and C. Chen, "HINet: Half instance normalization network for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 182–192.

[18] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5728–5739.

[19] T. Brooks, B. Mildenhall, T. Xue, J. Chen, D. Sharlet, and J. T. Barron, "Unprocessing images for learned raw denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11036–11045.

[20] K. Wei, Y. Fu, J. Yang, and H. Huang, "A physics-based noise formation model for extreme low-light raw denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2755–2764.

[21] K. Wei, Y. Fu, Y. Zheng, and J. Yang, "Physics-based noise modeling for extreme low-light photography," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 8520–8537, Nov. 2022.

[22] J. Janesick, K. Klaasen, and T. Elliott, "CCD charge collection efficiency and the photon transfer technique," in *Proc. Solid-State Imag. Arrays*, 1985, pp. 7–19.

[23] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 3, pp. 267–276, Mar. 1994.

[24] L. Song and H. Huang, "Fixed pattern noise removal based on a semi-calibration method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 11842–11855, Oct. 2023.

[25] A. J. Theuwissen, "How to measure the dark shading?," 2011 [Online]. Available: https://harvestimaging.com/blog/?p=866

[26] H. Feng, L. Wang, Y. Wang, H. Fan, and H. Huang, "The project of learnability enhancement for low-light raw image denoising: A data perspective," 2023. [Online]. Available: https://fenghansen.github.io/publication/PMN/

[27] H. Feng, L. Wang, Y. Wang, and H. Huang, "Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 1436–1444.

[28] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[29] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 14, pp. 259–268, 1992.

[30] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 60–65.

[31] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[32] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3952–3966, Sep. 2012.

[33] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[34] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[35] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2862–2869.

[36] C. Chen, Q. Chen, M. N. Do, and V. Koltun, "Seeing motion in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3185–3194.

[37] H. Yue, C. Cao, L. Liao, R. Chu, and J. Yang, "Supervised raw video denoising with a benchmark dataset on dynamic scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2298–2307.

[38] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[39] B. Jähne, "EMVA 1288 standard for machine vision: Objective specification of vital camera data," *Optik Photonik*, vol. 5, no. 1, pp. 53–54, 2010.

[40] M. Makitalo and A. Foi, "A closed-form approximation of the exact unbiased inverse of the anscombe variance-stabilizing transformation," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2697–2698, Sep. 2011.

[41] M. Makitalo and A. Foi, "Optimal inversion of the generalized anscombe transformation for Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 91–103, Jan. 2013.

[42] H. Wach and E. R. Dowski Jr, "Noise modeling for design and simulation of computational imaging systems," in *Proc. Vis. Inf. Process.*, 2004, pp. 159–170.

[43] R. Gow et al., "A comprehensive tool for modeling CMOS image-sensor-noise performance," *IEEE Trans. Electron Devices*, vol. 54, no. 6, pp. 1321–1329, Jun. 2007.

[44] M. Konnik and J. Welsh, "High-level numerical simulations of noise in CCD and CMOS photosensors: Review and tutorial," pp. 1–21, 2014, *arXiv:1412.4031*.

[45] J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton, FL, USA: CRC, 2017.

[46] Y. Zhang, H. Qin, X. Wang, and H. Li, "Rethinking noise synthesis and modeling in raw denoising," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4573–4581.

[47] T. Plötz and S. Roth, "Benchmarking denoising algorithms with real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2750–2759.

[48] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1692–1700.

[49] A. Punnappurath, A. Abuolaim, A. Abdelhamed, A. Levinshtein, and M. S. Brown, "Day-to-night image synthesis for training nighttime neural ISPs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10759–10768.

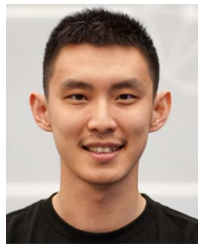
Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.

[50] J. Liu et al., "Learning raw image denoising with bayer pattern unification and bayer preserving augmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 2070–2077.
[51] J. Wang, Y. Yu, S. Wu, C. Lei, and K. Xu, "Rethinking noise modeling in extreme low-light environments," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2021, pp. 1–6.
[52] M. Maggioni, E. Sánchez-Monge, and A. Foi, "Joint removal of random and fixed-pattern noise through spatiotemporal video filtering," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4282–4296, Oct. 2014.
[53] G. C. Holst and T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*, 2nd Ed., Bellingham, WA, USA: SPIE, 2011.
[54] J. Lehtinen et al., "Noise2noise: Learning image restoration without clean data," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2971–2980.
[55] Z.-Q. J. Xu, Y. Zhang, T. Luo, Y. Xiao, and Z. Ma, "Frequency principle: Fourier analysis sheds light on deep neural networks," *Commun. Comput. Phys.*, vol. 28, no. 5, pp. 1746–1767, 2020.
[56] Z.-Q. J. Xu, Y. Zhang, and Y. Xiao, "Training behavior of deep neural network in frequency domain," in *Proc. Int. Conf. Neural Inf. Process.*, 2019, pp. 264–274.
[57] W. C. Porter, B. Kopp, J. C. Dunlap, R. Widenhorn, and E. Bodegom, "Dark current measurements in a CMOS imager," in *Proc. Sensors, Cameras, Syst. Industrial/Scientific Appl. IX*, 2008, pp. 98–105.
[58] X. Sun, W. Liu, H. Zhai, D. Ding, and J. Guo, "Study on vibration effects upon precise instruments due to metro train and mitigation measures," in *Environmental Vibrations: Prediction, Monitoring, Mitigation and Evaluation*. Boca Raton, FL, USA: CRC, 2021, pp. 439–444.
[59] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.
[60] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.- Assist. Intervention*, 2015, pp. 234–241.
[61] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.
[62] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–16.
[63] K. Monakhova, S. R. Richter, L. Waller, and V. Koltun, "Dancing under the stars: Video denoising in starlight," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16241–16251.
[64] W.-S. Lin, G.-M. Sung, and J.-L. Lin, "High performance CMOS light detector with dark current suppression in variable-temperature systems," *Sensors*, vol. 17, no. 1, pp. 1–15, 2016.
[65] D. Sander and P. Abshire, "Mismatch reduction for dark current suppression," in *Proc. Sensors*, 2010, pp. 1696–1700.

Hansen Feng received the BS degree from the University of Science and Technology Beijing, China, in 2020. He is currently working toward the PhD degree with the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include computational photography and image processing. He received the Best Paper Runner-Up Award of ACM MM 2022.

Lizhi Wang (Member, IEEE) received the BS and PhD degrees from Xidian University, Xi'an, China, in 2011 and 2016, respectively. He is currently an associate professor with the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include computational photography and image processing. He received the Best Paper Award of IEEE VCIP 2016.

Yuzhi Wang received the BS degree from the School of Telecommunication Engineering, Xidian University, Xi'an, China, in 2012, and the PhD degree with the Department of Electronic Engineering, Tsinghua University, Beijing, China, under the supervision of Prof. H. Yang. His research interests include wireless sensor networks, computational photography, machine learning, and deep neural networks



Haoqiang Fan received the BS degree from the Institute for Interdisciplinary Information Sciences "Yao Class," Tsinghua University, Beijing, China. He is currently the director of the AI Algorithm Research, MEGVII Institute. His current research interests include deep leaning and the applications in different fields.

Hua Huang (Senior Member, IEEE) received the BS and PhD degrees from Xi'an Jiaotong University, in 1996 and 2006, respectively. He is currently a professor with the School of Artificial Intelligence, Beijing Normal University. He is also an adjunct professor with Xi'an Jiaotong University and Beijing Institute of Technology. His main research interests include image and video processing, computational photography, and computer graphics. He received the Best Paper Award of ICML2020 / EURASIP2020 / PRCV2019 / ChinaMM2017.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on January 16,2025 at 14:01:46 UTC from IEEE Xplore. Restrictions apply.