Lighting Every Darkness in Two Pairs:

A Calibration-Free Pipeline for RAW Denoising

Xin Jin\textsuperscript{1}*, Jia-Wen Xiao\textsuperscript{1}*, Ling-Hao Han\textsuperscript{1} Chunle Guo\textsuperscript{1,\dagger} Ruixun Zhang\textsuperscript{2} Xialei Liu\textsuperscript{1} Chongyi Li\textsuperscript{1,\dagger}

\textsuperscript{1}VCIP, CS, Nankai University \hspace{0.5em} \textsuperscript{2}Peking University \hspace{0.5em} \textsuperscript{3}S-Lab, Nanyang Technological University

\{xjin, xiaojw, lhhan\}@mail.nankai.edu.cn, zhangruixun@pku.edu.cn, {guochunle, xialei, lichongyi}@nankai.edu.cn

https://srameo.github.io/projects/led-iccv23

Abstract

Calibration-based methods have dominated RAW image denoising under extremely low-light environments. However, these methods suffer from several main deficiencies: 1) the calibration procedure is laborious and time-consuming, 2) denoisers for different cameras are difficult to transfer, and 3) the discrepancy between synthetic noise and real noise is enlarged by high digital gain. To overcome the above shortcomings, we propose a calibration-free pipeline for Lighting Every Darkness (LED), regardless of the digital gain or camera sensor. Instead of calibrating the noise parameters and training repeatedly, our method could adapt to a target camera only with few-shot paired data and fine-tuning. In addition, well-designed structural modifications during both stages alleviates the domain gap between synthetic and real noise without any extra computational cost. With 2 pairs for each additional digital gain (in total 6 pairs) and 0.5\% iterations, our method achieves superior performance over other calibration-based methods.

1. Introduction

Noise, an unescapable topic for image capturing, has been systematically investigated in recent years [5, 66, 51, 41, 2, 8, 57]. Compared with standard RGB images [56, 21, 54, 34, 33, 32], RAW images enjoy two great potentials for image denoising: tractable, primitive noise distribution [57] and higher bit depth for differentiating signal from noise. Learning-based methods have achieved significant progress on RAW image denoising with paired real datasets [67, 22, 64, 32, 33]. However, it is unfeasible to collect a large-scale real RAW image dataset for each single camera model. Therefore, increasing attention has been drawn from deploying learning-based methods on synthetic dataset [1, 61, 31, 57, 68, 44, 40].

*Equal contribution.
\dagger C. L. Guo is the corresponding author.
Our main contributions are summarized as follows:

- We propose a calibration-free pipeline for lighting every darkness, which avoids all extra costs for calibrating the noise parameters.
- Designed CSA loosens the coupling between the denoising network and camera model, while OMNR enables few-shot transfer by learning the out-of-model noise of different sensors.
- Only 2 raw image pairs for each ratio and 0.5% iterations are required compared with SOTA methods.

2. Related Work

Training with Paired Real Data. Since the pioneering work of SIDD [2], the potential of RAW data for image denoising has been explored. Recent works step aside from normal light image denoising to extremely low-light environment, e.g., SID [8], ELD [57]. Notwithstanding the promising results of real noise-based methods [9, 11, 62, 63], the difficulty in collecting large-scale paired (low-quality and high-quality pairs) real dataset still bottlenecks their deployment. Training with paired low-quality raw images, like Noise2Noise [41] and Noise2NoiseFlow [44], could avoid the labor-intensive collection of noisy-clean image pairs. However, these methods always failed in intensive noise as in terribly dark scenes [8, 57]. Our LED aims to complement the knowledge for real noise removal with new-shot paired images under extremely low-light environments, thus relieving the difficulties in data collection.

Calibration-Based Denoising. Synthetic noise-based methods could avoid the tiosomeness of collecting pairwise datasets, but practical constraints still exist. The widespread noise models, Poisson and Gaussian noises, deviate vigorously from the real noise distribution, especially in extremely low-light environment [8, 57]. Thus, calibration-based methods, which simulate each noise component in the electronic imaging pipelines [4, 24, 20, 29, 39], have flourished due to their reliability. ELD [57] proposed a noise model that fits real noise well, attaining great performance under dark scenarios. Zhang et al. [68] realized that the source of the signal-independent noise is too complicated to model, hence proposed a method that randomly samples signal-independent noise from dark frames. However, it still requires calibration for the parameters of signal-dependent noise, e.g., overall system gain. Kristian et al. [45] build the noise generator combining the physics-based noise model and generative adversarial framework [19]. Zou et al. [69] aims

1“Virtual” cameras do not correspond to any real camera models, but with reasonable noise parameters of the predefined noise model.

2Denosing under extremely low-light scenarios requires applying additional digital gain (up to 300×) to the input, intensifying the domain gap between real and synthetic noise.
for more accurate and concise calibration by using contrastive learning [12, 23] for parameter estimation. Though calibration-based methods achieve superb performance, stable illumination environment (e.g., brightness and temperature), calibration-specialized data collection (e.g., dozens of images for each camera setting), and complicated post-processing (e.g., alignment, locating, and statistics) are required for estimating noise parameters. In addition, repeated calibration and training process is needed for each camera due to the diversity of parameters and nonuniform pre-defined noise model [52, 20, 39, 43]. Also, the domain gap between synthetic noise and real noise is not taken into account. Our LED resolve the above problems with a calibration-free pipeline, a pre-training and fine-tuning framework, and a proposed RepNR block.

**From Synthetic to Real Noise.** The domain gap between real and synthetic noise is an inevitable challenge when training on synthetic data while testing on real data. With the progress of AdaIN [27, 36] and few-shot learning [25, 60, 26], recent works mainly focus on leveraging transfer learning [37] or domain adaptation [47] technique for mitigating the domain gap. However, in extremely dark scenes, these methods would fail in signal reconstruction due to numerical instability caused by extreme noise and the additional digital gain. Our proposed camera-specific alignment avoids numerical instability while still decoupling the camera-specific information and common knowledge of the noise model. Additionally, compared with the instance or layer normalization [50, 3], the alignment operations can be reparameterized into convolution like custom batch normalization [28], thus resulting in no extra computation cost.

### 3. Method

In this section, we start by presenting the whole pipeline for our proposed calibration-free raw image denoising. We then present our reparameterized noise removal (RepNR) block. The whole denoising pipeline is described in Fig. 3.

#### 3.1. Preliminaries and Motivation

In raw image space, captured signals $D$ are always treated as the sum of clean image $I$ and noise components $N$, formulated as Eq.(1).

$$D = I + N,$$

where $N$ is assumed as a noise model,

$$N = N_{\text{shot}} + N_{\text{read}} + N_{\text{row}} + N_{\text{quant}} + \epsilon,$$

where $N_{\text{shot}}, N_{\text{read}}, N_{\text{row}}$ and $N_{\text{quant}}$ denote shot noise, read noise, row noise, and quant noise, respectively. And $\epsilon$ denotes the out-of-model part. Besides the out-of-model noise, other noises is sampled from a certain distribution:

$$N_{\text{shot}} + I \sim \mathcal{P}(\frac{I}{K})K,$$

$$N_{\text{read}} \sim TL(\lambda; \mu_c, \sigma_{TL}),$$

$$N_{\text{row}} \sim \mathcal{N}(0, \sigma_f),$$

$$N_{\text{quant}} \sim U(-\frac{1}{2}, \frac{1}{2}),$$

where $K$ denotes overall system gain. $\mathcal{P}, \mathcal{N}, U$ stand for Poisson, Gaussian, and uniform distributions, respectively. $\text{TL}(\lambda; \mu, \sigma)$ represents Tukey-lambda distribution [35] with
shape λ, mean μ and standard deviation σ. In addition, a linear relationship exists for the joint distribution of \((K, \sigma_{TL})\) and \((K, \sigma_r)\), which can be denoted as:

\[
\log(K) \sim U(\log(K_{\min}), \log(K_{\max})),
\]

\[
\log(\sigma_{TL})|\log(K) \sim \mathcal{N}(a_{TL} \log(K) + b_{TL}, \hat{\sigma}_{TL}),
\]

\[
\log(\sigma_r)|\log(K) \sim \mathcal{N}(a_r \log(K) + b_r, \hat{\sigma}_r),
\]

In that case, a camera can be approximately represented as a coordinate \(C\) of ten dimensions:

\[
C = (\hat{K}_{\min}, \hat{K}_{\max}, \lambda, \mu_{c}, a_{TL}, b_{TL}, \hat{\sigma}_{TL}, a_r, b_r, \hat{\sigma}_r).
\]

Previous methods focus on calibration to adjust the coordinate \(C\), suffering from intensive labor and huge domain gap (i.e., gap between simulated noise and real noise). In addition, a repeated training process is necessary due to the entanglement between neural networks and cameras. Our aim is to abandon the complicated calibration process and impair the strong coupling between networks and cameras. Furthermore, we fully account for the out-of-model noise, which can be alleviated by the structural modifications of our RepNR block. In general, our motivation is to force the network to become a fast adapter [48, 18].

### 3.2. Pre-train with Camera-Specific Alignment

**Preprocessing.** In order to promote the network to become a fast adapter, we first pre-train our network utilizing virtual cameras. Given the number of virtual cameras \(m\) and parameter space (formulated as \(\mathcal{S}\)), for the \(k\)-th camera, we select the \(k\)-th \(m\) bisection points of each parameter range and combine them to obtain a virtual camera. With the data augmented by the synthetic noise, we can pre-train our network based on several virtual cameras, forcing the network to learn the common knowledge.

**Camera-Specific Alignment.** As shown in Fig. 3, during the pre-training process, we introduce our Camera-Specific Alignment (CSA) module, which focuses on adjusting the distribution of input features. In the baseline model, a \(3 \times 3\) convolution followed by leaky-ReLU [59] is the main component. To reflect features from different virtual cameras into a shared space, a multi-path alignment layer is inserted before each convolution. Each path is the CSA corresponding to the \(k\)-th camera, aligning the distribution of the \(k\)-th camera-specific feature into a shared space.

Let feature of the \(k\)-th virtual camera be \(F = (f_1, ..., f_c) \in \mathcal{R}^{B \times C \times H \times W}\). Formally, the \(k\)-th branch contains a weight \(W^k = (w_1^k, ..., w_c^k) \in \mathcal{R}^{C}\) and a bias \(b^k = (b_1^k, ..., b_c^k) \in \mathcal{R}^{C}\), operating channel-wise linear projection to \(F\), denoted by \(Y = W^k F + b^k\). \(W^k (k = 1, ..., m)\) are initialized as \(1\) and \(b^k (k = 1, ..., m)\) are initialized as \(0\), with no effect on the \(3 \times 3\) convolution at the beginning. During training, data augmented by the noise of the \(k\)-th virtual camera will be fed into the \(k\)-th path for aligning, and into a shared \(3 \times 3\) convolution for further processing. The detailed pre-training pipeline is described in Algorithm 1.

### 3.3. Fine-tune with Few-shot RAW Image Pairs

After the pre-training process, the model is suspected to be used in realistic denoising tasks. We propose to use a few-shot strategy and, in particular, only \(6\) pairs (2 pairs for each of the 3 ratios) of raw images to fine-tune the pre-trained model. \(3 \times 3\) convolutions are assumed to have learned enough to deal with features aligned by CSAs. In order to make better use of the model parameters obtained from pre-training, the convolutions are kept frozen for further fine-tuning. To deal with real noise, we replace the multi-branch CSA with a new CSA layer, denoted as CSA\(^T\) (CSA for the target camera). Unlike the multi-branch CSA during pre-training, the CSA\(^T\) layer is initialized by averaging the pre-trained CSAs, which can be viewed as model ensembling. (d) The reparameterization process during deployment. Rep. denotes reparameterize.

![Algorithm 1 Pre-training pipeline in LED](image)

**Algorithm 1** Pre-training pipeline in LED

**Require:** model \(\Phi\), \(m\), \(\mathcal{S}\), clean dataset \(D\)

\[
\Phi_{\text{pre}} \leftarrow \text{insert-multi-CSA}(\Phi) \\
\{c_k\}_{k=1}^m \leftarrow \text{generate-virtual-camera}(\mathcal{S})
\]

while not converged do

Sample mini-batch \(x_i \sim D\)

\(k \leftarrow \text{random}(1, m)\)

\(\tilde{x}_i \leftarrow \text{augment}(c_k, x_i)\)

\(\text{train}(\Phi_{\text{pre}}, \{\tilde{x}_i, x_i\})\)

end while
Nevertheless, real noise includes not only the modeled part, but also some out-of-model noise. Since our CSA layer is only designed for aligning features augmented by synthetic noise, there is still a gap between real noise and the one IMNR can handle (i.e., ϵ in Eqn. (2)). Thus, we propose to add a new branch, named the out-of-model noise re- 

Algorithm 2 Fine-tuning and deploy pipeline in LED

Require: pre-trained model Φ_{pre}, real dataset D_{real} 

Φ_{0} ← freeze-3×3(Φ_{pre})
Φ_{f} ← average-CSA(Φ_{B})
while not converged do
Sample mini-batch pairs \( \{x_i, y_i\} \sim D_{real} \)
train(Φ_{B}, \{x_i, y_i\})
end while
Φ_{f} ← freeze(Φ_{f})
Φ_{f} ← add-OMNR(Φ_{f})
while not converged do
Sample mini-batch pairs \( \{x_i, y_i\} \sim D_{real} \)
train(Φ_{B}, \{x_i, y_i\})
end while
Φ_{final} ← deploy(Φ_{f})

4. Experiments and Analysis

In this section, we detailed our implementation, stated the datasets and evaluation metrics, provided comparison experiments and demonstrated ablation studies.

4.1. Implementation Details

Like most denoising methods [61, 13], we use \( L_1 \) loss function as the training objectives. We use the same UNet [49] architecture as previous methods for a fair comparison, and the difference is that we replace the convolution blocks inside the UNet with our proposed RepNR block. As stated in Sec. 3.4, the RepNR block can be structurally reparameterized into a simple convolution block without any extra computational cost. Same data preprocessing and optimization strategy as ELD [57] is used during pre-training. The raw images with long exposure time in SID [8] train subset are used for noise synthesis. As for the data preprocessing, we pack the Bayer images into 4 channels, then crop the long exposure data with patch size 512 × 512, non-overlap, enlarging the iterations of one epoch from 161 to 1288. Our implementation is based on PyTorch [46] and MindSpore. We train the models with 200 epochs (257.6K iter.) and Adam optimizer [38] with \( \beta_1 = 0.9 \) and \( \beta_2 = 0.999 \) for optimization, where no weight decay is applied. The initial learning rate is set to \( 10^{-4} \) and then halved at the 100th epoch (128.8K iter.) before finally reduced to \( 10^{-5} \) at the 180th epoch (231.84K iter.).

During fine-tuning, we first freeze the 3 × 3 convolution and average the multi-branch CSA as the initialization of CSA\(^T\). After training the CSA\(^T\) for 1K iterations with \( 10^{-4} \) learning rate, we add the out-of-model noise removal branch (a parallel 3 × 3 convolution) and freeze all the left parameters in our network. Finally, we train the OMNR branch for 500 iterations with a learning rate of \( 10^{-5} \). After the entire training process, we deploy our model by reparameterizing the RepNR blocks into convolutions.
Calibration-Based

- Calibration-P-G
  - 300 calibration data
  - PSNR: 39.1576, SSIM: 0.8963
- ELD [57]
  - 300 calibration data
  - PSNR: 41.8271, SSIM: 0.9538
- Zhang et al. [68]
  - ~150/~150 for calib./database
  - PSNR: 37.6866, SSIM: 0.7818

Real Data Based

- SID [8]
  - ~1800 noisy-clean pairs
  - PSNR: 41.7273, SSIM: 0.9531
- Noise2Noise [41]
  - ~12000 noisy pairs
  - PSNR: 39.0084, SSIM: 0.8391
- AINDNet [37]
  - ~300 noisy-clean pairs
  - PSNR: 40.5636, SSIM: 0.9194
- AINDNet* [37]
  - ~300 noisy-clean pairs
  - PSNR: 39.8052, SSIM: 0.9350
- LED (Ours)
  - 6 noisy-clean pairs
  - PSNR: 41.9842, SSIM: 0.9359

Table 2. Quantitative results on two camera models, Sony A7S2 and Nikon D850, of ELD [57] dataset. The best result is denoted as bold.

4.2. Datasets and Evaluation Metrics

- We have benchmarked our proposed LED on two RAW-based denoising datasets, i.e., SID [8] and ELD [57]. Four different camera models: Sony A7S2, Nikon D850, Canon EOS70D, and Canon EOS700D, and 7 varying additional digital gains from ~1 to ~300 are included in these two datasets. As for the SID dataset, we randomly choose two pairs of data for each additional gain (~100, ~250, and ~300) as the few-shot training datasets. For the ELD dataset, the paired raw images of the first two scenarios are used for fine-tuning the pre-trained network. After the entire training process, the test set of the SID [8] Sony subset and the left scenes of the ELD [57] dataset are used to validate the effectiveness of our proposed LED. LED is also evaluated on Canon cameras (Canon EOS70D and Canon EOS700D), on which we also achieve state-of-the-art performance. Results will be released in future version.

- We regard PSNR and SSIM [55] as the quantitative evaluation metrics for pixel-wise and structural assessment. Notice that, the pixel value of low-light raw images usually lies in a smaller range than sRGB images, i.e., [0, 0.5] after normalization, resulting in a lower mean square error and higher PSNR.

4.3. Comparison with State-of-the-art Methods

- We evaluate our LED on two datasets, the Sony subset of SID [8] and the ELD dataset [57], to assess the generalization capabilities of LED on outdoor and indoor scenes, respectively. The state-of-the-art raw denoising methods under extremely low-light environments are compared with LED, including:
  - **DNN model based methods**: Kristina et al. [45] and NoiseFlow [1]. These methods are first trained on paired real raw images to learn how to generate noise for a specific camera, resulting in more iterations when deployed on a new camera model.
  - **Calibration-based methods**: ELD [57], Zhang et al. [68], and Calibration-P-G. These methods require a time-consuming and laborious calibration process.
  - **Real data based methods**: training with noisy-clean pairs (SID [8]), noisy-noisy pairs (Noise2Noise [41]) and transfer learning (AINDNet [37]).

The denoising network of all the above methods is trained with the same setting as ELD [57], as stated in Sec. 4.1, for a fair comparison.
Quantitative Evaluation. As shown in Tab. 1 and Tab. 2, our method outperforms previous calibration-based methods under extremely low-light environments. The domain gap between synthetic noise and real noise would be magnified with a large ratio (×250 and ×300), leading to a performance drop on training with synthetic noise, as shown in the comparison between ELD [57] and SID [8]. In addition, DNN model based methods often yield more discrepancies than calibration-based methods. In particular, different system gains are not taken into consideration by Kristina et al. [45]. However, our method alleviates this discrepancy by fine-tuning with few-shot real data, thus achieving better performance under ×100 and ×250 digital gain, as shown in Tab. 1. AINDNet [37] would also achieve better performance under extremely dark scenes with a noise model of less discrepancy. The noise model deviation does not affect the denoising ability under small additional digital gain, as shown in Tab. 2. Nevertheless, our method shows superiority under extremely low-light scenes, also in different camera models. Notice that, LED introduces less training cost, both in data requirement and training iterations, compared with other methods.

Qualitative Evaluation. Fig. 5 and Fig. 6 show the comparison with other state-of-the-art methods on the SID [8] and the ELD [57] dataset, respectively. When imaging under extremely low-light conditions, the intensive noise would disturb the color tone seriously. As shown in Fig. 5, the input images exhibit green or purple color shifts, and most comparison methods could not restore the correct color tone. Benefiting from the implicit noise modeling and the diverse sampling space, the LED efficiently restores signals with severe noise interference, yielding accurate color rendering and rich texture detail. Besides, comparison methods are hard to recognize the enlarged out-of-model noises, which corrupt the resulting image in fixed patterns or certain positions. While during the fine-tuning stage, LED additionally learns to remove these camera-specific noises, thus achieving superior visual quality and strong robustness.

4.4. Ablation Studies
Reparameterized Noise Removal Block. We conduct experiments for the ablation of different components in the Reparameterized Noise Removal Block (RepNR). As shown in Tab. 3, our RepNR achieves better performance in three different ratios, and each component in the RepNR block contributes positively to the whole pipeline.
Why 2 pairs for each ratio? As shown in Eqn. (4), the variance of noise \( \log(\sigma) \) is linearly related to overall system gain \( \log(K) \). With only one pair of data, it is impossible to find the correct linear relationship, thus resulting in the worst performance, as shown in Tab. 6. Plus, utilizing two or more pairs with similar system gains can’t model the linear relationship precisely due to a non-negligible error of the sampling scope \( \langle \hat{\sigma} \rangle \) (as shown in Fig. 8). With the principle of using two points to determine a straight line, we adapt 2 pairs of marginally different system gains to model the linearity, greatly improving the capability of denoising. Furthermore, as shown in Fig. 7, with the number of the pairs increasing, linearity can be fitted more accurately, leading to further elimination of the regression error.

5. Conclusion

To relieve the inherent defects of calibration-based methods, we propose a calibration-free pipeline for lighting every darkness. Benefiting from the camera-specific alignment, we replace the explicit calibration procedure with an implicit learning process. CSA enables fast adaptation to the target camera by decoupling the camera-specific information and common knowledge of the noise model. Plus, a parallel convolution mechanism is designed for learning to remove the out-of-model noise. With 2 pairs for each ratio (in total 6 pairs) and 1.5K iterations, we achieve superior performance than existing methods.
Acknowledgement. This research was supported by the NSFC (NO. 62225604) and the Fundamental Research Funds for the Central Universities (Nankai University, 070-63233089), China Postdoctoral Science Foundation (NO.2021M701780). The Supercomputing Center of Nankai University supports computation. We are also sponsored by CAAI-Huawei MindSpore Open Fund (CAAIXSJLJJ-2022-024A).

References

[8] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In CVPR, 2018. 1, 2, 5, 6, 7, 8
[27] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In ICCV, 2017. 3
[32] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In CVPR, 2018. 1, 2, 5, 6, 7, 8
[27] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In ICCV, 2017. 3
[32] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In CVPR, 2018. 1, 2, 5, 6, 7, 8
images using guided apsf and gradient adaptive convolution. arXiv:2308.01738, 2023. 1

[34] Ypaging Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In ECCV, 2022. 1


[37] Yoonis Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In CVPR, 2020. 1, 3, 6, 7


[42] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. ICLR, 2017. 8


[54] Yufei Wang, Renjie Wen, Wenhan Yang, Haoiang Li, Lappui Chau, and Alex Kot. Low-light image enhancement with normalizeing flow. In AAAI, 2022. 1


[57] Kaixuan Wei, Ying Fu, Yiqiang Zheng, and Jialong Yang. Physics-based noise modeling for extreme low-light photography. 2021. 1, 2, 5, 6, 7, 8


[60] Han-Jie Ye, Lu Ming, De-Chuan Zhan, and Wei-Lun Chao. Few-shot learning with a strong teacher. TPAMI, 2022. 3


[64] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shabbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. IEEE TPAMI, 2022. 1


[68] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis. In ICCV, 2021. 1, 2, 6, 7