# ExposureDiffusion: Learning to Expose for Low-light Image Enhancement

Yufei Wang[1], Yi Yu[1], Wenhan Yang[2], Lanqing Guo[1], Lap-Pui Chau[3], Alex C. Kot[1], Bihan Wen[1*]
[1]Nanyang Technological University [2]Peng Cheng Laboratory
[3]The Hong Kong Polytechnic University
{yufei001, yuyi0010, lanqing001, eackot, bihan.wen}@ntu.edu.sg
yangwh@pcl.ac.cn lap-pui.chau@polyu.edu.hk

## Abstract

*Previous raw image-based low-light image enhancement methods predominantly relied on feed-forward neural networks to learn deterministic mappings from low-light to normally-exposed images. However, they failed to capture critical distribution information, leading to visually undesirable results. This work addresses the issue by seamlessly integrating a diffusion model with a physics-based exposure model. Different from a vanilla diffusion model that has to perform Gaussian denoising, with the injected physics-based exposure model, our restoration process can directly start from a noisy image instead of pure noise. As such, our method obtains significantly improved performance and reduced inference time compared with vanilla diffusion models. To make full use of the advantages of different intermediate steps, we further propose an adaptive residual layer that effectively screens out the side-effect in the iterative refinement when the intermediate results have been already well-exposed. The proposed framework can work with both real-paired datasets, SOTA noise models, and different backbone networks. We evaluate the proposed method on various public benchmarks, achieving promising results with consistent improvements using different exposure models and backbones. Besides, the proposed method achieves better generalization capacity for unseen amplifying ratios and better performance than a larger feedforward neural model when few parameters are adopted. The code is released at* https://github.com/wyf0912/ExposureDiffusion.

## 1. Introduction

Over the past few years, learning-based methods for low-light image enhancement [22, 21, 17] have gained significant attention and made remarkable progress, and most of them are conducted in the sRGB space. Recently, the enhancement in the raw space is demonstrated to have unique advantages over sRGB spaces [14]. For example, raw images provide a higher dynamic range, leading to better performance in extremely dark environments. Besides, the linear correlation between the low-light and normally-exposed images prevents improper exposure level adjustment in the enhancement process. In addition, the noise modeling in the raw space is more straightforward than that in the sRGB space by ruling out the effect of increasingly complicated image signal processing pipelines. In such space, the domain gap between synthetic and captured data is small and the model trained with paired synthetic images ex-

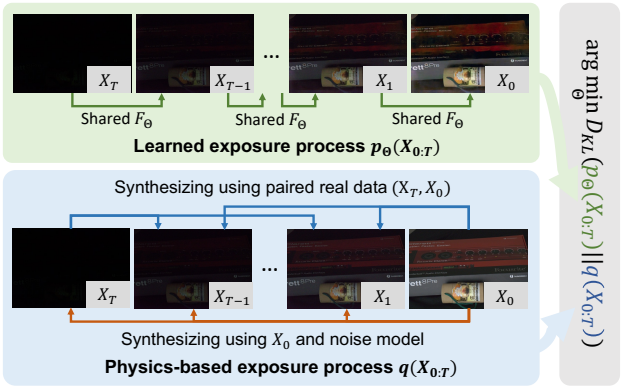



Figure 1: We propose to simulate the physics-based exposure process using a shared neural network $F_\Theta$ in a progressive manner. The learned exposure process is optimized to approximate the physics-based exposure process by minimizing the derived variational upper bound of the KL divergence of their distributions. Besides, the proposed strategy can be applied on real-captured paired data (blue array) and synthetic data with different noise models (orange array), and in different backbone networks. Benefiting from learning a continuous exposure process, the proposed method can be applied to work with an arbitrary amplifying factor, and better performance can be achieved by the iterative refinement process.

*Corresponding author.

hibits comparable or even better performance than that of real-captured data [45, 6]. While promising progress is achieved, the prevailing approach remains to learn a deterministic mapping based on feedforward neural networks. For the images captured in extremely dark environments, this one-step enhancement/denoising process[1] fails to characterize the distribution information and usually obtain undesirable results. For example, there may still exist some residual noise. Besides, existing works mainly pay attention to more accurate noise modeling. The work of effectively incorporating the noise model in the raw space into a learnable model for improved enhancement remains unexplored.

Most recently, generative model-based image restoration methods [25, 41, 34] exhibit appealing performance and pleasing perceptual quality in image restoration tasks. Among these generative models, diffusion models [36, 13] stand out for their capacity to model a complicated distribution with arbitrary neural networks in a progressive manner and exhibit great success in image generation and restoration tasks [34]. Different types of forward processes have been explored in diffusion models, *e.g.*, the unified framework for the Wiener process [49]. Nonetheless, they are inadequate to accurately simulate real exposure processes. First, the low-light image is naturally not an intermediate step of the vanilla diffusion process. Therefore, the reverse (denoising) process needs to start from pure noise and involves a relatively large number of inference steps, which hinders the real applications. Second, since the vanilla diffusion models need to have the capacity of removing Gaussian noise with different noise levels, it usually requires extra model capacity compared with feedforward neural networks.

To address aforementioned issues, we propose a novel approach to effectively inject noise models in raw space into an end-to-end learnable progressive model, named *ExposureDiffusion*. Specifically, we propose to simulate the exposure process using a progressive shared network to minimize the divergence between the simulated process and the real one by optimizing the proposed variational upper bound. Since the intermediate steps of the progressive process all obey the physics-based noise distribution, the restoration process can directly start from a noisy image instead of pure noise. This design significantly benefits the low-light enhancement/denoising in two dimensions. First, the proposed method no longer requires removing Gaussian noises and only needs to learn the process of real-noise denoising, leading to smaller requirements of model capacity. Second, the proposed method greatly reduces the required number of inference steps, which has the potential to significantly benefit real applications. Besides, we further propose an adaptive residual layer to dynamically fuse different denoising strategies for the areas with different noise-to-signal ratios. This strategy effectively screens out the side-effect in the iterative refinement when the intermediate results have been already well-exposed. The proposed method can be applied to both paired real-captured data, synthetic data with different noise models, and different backbone networks. Experimental results demonstrate that the proposed method can achieve significant improvement jointly with both real/synthetic exposure process and backbone networks. The proposed method, which employs the noisy-to-fine strategy, also exhibits superior generalization capability. Our main contributions are summarized as follows:

- We propose the first diffusion-based model for low-light image enhancement in the raw image space. The modeling of the process is inspired and constructed strictly according to the physical noise model. This design enables restoration from any intermediate step of the diffusion process and eliminates the need for the Gaussian denoising process. As a result, the available model capacity and inference efficiency are significantly improved.
- We further propose an adaptive residual layer to dynamically adopt different denoising strategies for areas with different noise-to-signal ratios. This strategy effectively screens out the side-effect in the iterative refinement when the intermediate results have been already well-exposed.
- Extensive experimental results on two public datasets demonstrate the significant performance improvement of the proposed method combined with state-of-the-art noise models/backbones. Besides, the proposed method exhibits better generalization capacity compared with feedforward neural networks and possesses fewer parameters and faster speed to achieve competitive performance.

## 2. Related works

**Low-light image enhancements.** A great number of low-light image enhancement methods based on deep learning have been proposed in the past few years [41, 50, 15, 39, 12, 9, 48, 18, 17]. The mainstream of these methods is based on supervised learning, *i.e.*, training a mapping from low-light images to normally-exposed images. For example, LLNet [23] proposes an autoencoder to enhance the visibility of low-light images. To obtain better perceptual quality, [35, 37, 26, 31] propose to utilize multi-scale features to better learn the global content and salient structures. Retinex theory is also widely used as the prior knowledge to guide the disentanglement of reflection and illumination maps [44, 40, 44, 10]. Unfolding/unrolling-based methods [32, 52, 47] are explored to better utilize priors of

[1] As exposure changes can be approximated with a linear transform in the raw space, low-light image enhancement in the raw image space is regarded as a denoising task in most previous works.

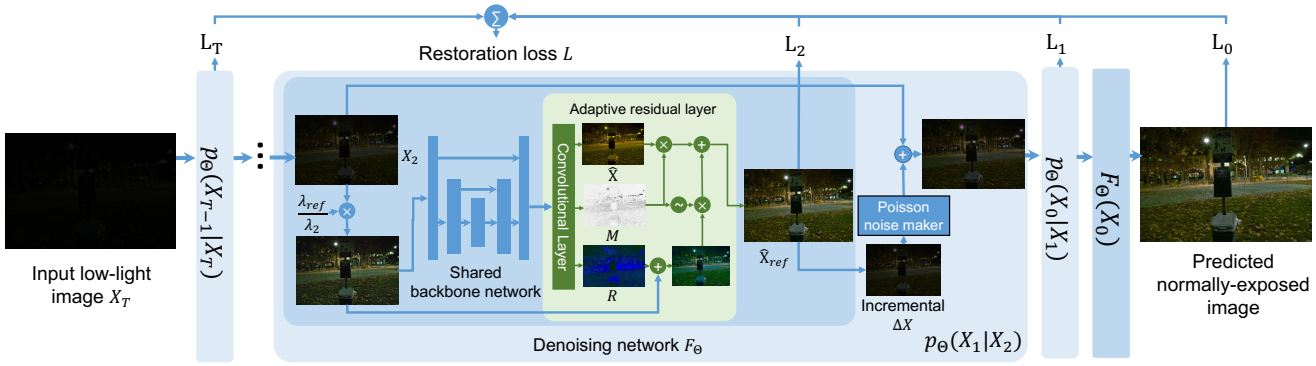



Figure 2: The proposed method follows an overall framework where the final results are achieved through a process of progressive refinement. The adaptive residual layer (in green) can be combined with any backbone network. For both training and inference, the process starts from a low-light image $X_T$, and images with longer exposure time are gradually achieved. The image reconstruction losses $L_t$ of different steps are all used for training, and $F_\theta(X_0)$ is the final result.

low-light image enhancement. Recently, the explicit modeling of the conditional distribution of normally-exposed images is explored in [41], showing superior perceptual quality. Besides the aforementioned enhancement methods in sRGB space, the enhancement in the image space gradually attracts increasing attention recently [14, 1, 30, 5] due to its unique advantages. Current research in the raw image space mainly focuses on the realism of noise modeling [45, 51, 6, 16, 43, 42] so that the synthetic training data can have a smaller domain gap. Specifically, Poisson-Gaussian [7] is a basic and widely used noise model, which assumes the noises include signal-independent Gaussian noise and signal-dependent Poisson noise. The following works mainly improve the modeling of the signal-independent noise, *e.g.*, additionally modeling the row noise [45] and the dark shading [6]. However, the exploration of training strategies to better utilize the clear formulation of the degradation process in the raw space is still lacking.

**Diffusion models.** Recently, diffusion-based [36, 13] image restoration models [46, 38] exhibit remarkable performance by using the degraded image as the conditional input. For example, [3] proposes a diffusion model for different inverse problems by using manifold constraints. By altering the reverse diffusion process, [24] exhibits good performance in free-form inpainting. [19] utilizes the pre-trained diffusion models to conduct multiple restoration tasks, *e.g.*, super-resolution, and deblurring. [34] proposes a conditioned diffusion model for super-resolution that uses the low-resolution image as a part of the input, and learns the whole process in an end-to-end manner. Besides, [33] proposes to handle different image-to-image tasks based on conditional diffusion models, and diffusion models are used as plug-and-play image priors in [8]. While promising results are achieved, they mainly focus on super-resolution, deblurring, inpainting, and colorization, for which the synthetic dataset is relatively easy to be obtained. Diffusion models, especially the physics-based model, for low-light image enhancement are still left to be explored. Besides, some efforts [49, 29, 11, 27] aim to speed up the sampling process of diffusion models. For example, [49] speeds up the inference speed by reducing the number of sampling steps and [29] uses deep nonequilibrium approaches to find the results after convergence. However, the initial states are still usually from pure noise even for the conditional image restoration task [11, 34].

## 3. Methodology

### 3.1. Preliminary

A raw image $X_t$ can be formulated as follows

$$X_t = \lambda_t K I + N, \tag{1}$$

where $\lambda_t$ represents the exposure time, $K$ is the overall system gain, $I$ is the rate of the photoelectrons which is proportional to the scene irradiation, and $N$ is the summation of all noise sources. The formulation of the noise summation $N$ in the raw image can be simplified as follow

$$N = KN_p + N_{ind}, \tag{2}$$

where $N_p$ is the photon shot noise, and $N_{ind}$ is the signal independent noise. The photon shot noise obeys a Poisson distribution as follows

$$(\lambda_t I + N_p) \sim \mathcal{P}(\lambda_t I). \tag{3}$$

The target of low-light image enhancement is to predict the normally-exposed image $X_0$ given an image $X_T$ with a short exposure time, *i.e.*, $\lambda_0 > \lambda_T$ [2]. The mainstream

[2]We assume that $\lambda_t$ decreases with the increase of $t$. Details are in the supplementary material.

| | Unconditional diffusion [36, 13] | Conditional diffusion [34] | Exposure diffusion (Ours) |
|---|---|---|---|
| Objective | maximizing $p_\Theta(X)$ | maximizing $p_\Theta(X\|Y)$ | minimizing KL divergence with the real exposure process |
| Initial state $X_T$ | $X_T \sim \mathcal{N}(0,1)$ | $X_T \sim \mathcal{N}(0,1)$ | $X_T \sim q(X_T)$ [1] |
| Assumption | $q(X_t\|X_{t-1}) := \mathcal{N}(X_t; \sqrt{1-\beta_t}X_{t-1}, \beta_t\mathbf{I})$ | $q(X_t\|X_{t-1}) := \mathcal{N}(X_t; \sqrt{1-\beta_t}X_{t-1}, \beta_t\mathbf{I})$ | $q(X_{t-1}\|X_t, X_{ref}) := \mathcal{P}(\frac{X_{t-1}-X_t}{K}; \frac{(\lambda_{t-1}-\lambda_t)X_{ref}}{\lambda_{ref}K})$ [2] |
| Reverse process | $p_\Theta(X_{t-1}\|X_t) := \mathcal{N}(X_{t-1}; \mu_\Theta(X_t,t), \sigma_t^2\mathbf{I})$ | $p_\Theta(X_{t-1}\|X_t, Y) := \mathcal{N}(X_{t-1}; \mu_\Theta(X_t,Y,t), \sigma_t^2\mathbf{I})$ | $p_\Theta(X_{t-1}\|X_t) := \mathcal{P}(\frac{X_{t-1}-X_t}{K}; \frac{(\lambda_{t-1}-\lambda_t)F_\Theta(X_t)}{\lambda_{ref}K})$ |
| Training | The expectation over $q(X_t\|X_0)$ | The expectation over $q(X_t\|X_0)$ | The expectation over $p_\Theta(X_t\|X_T)$ [3] |

[1] We start from a low-light image instead of a pure noise sampled from a Gaussian distribution.
[2] The Markov process is formulated based on the physics-based model of the exposure process.
[3] The cumulative error caused by progressive refinement is alleviated by minimizing the loss $L_t$ over the expectation of $p_\theta$.

Table 1: A comparison between the proposed algorithm and vanilla diffusion models.

of the previous works aims to minimize the reconstruction loss between the restored image and the reference image by optimizing a deep network parameterized by $\Theta$ as follows

$$\Theta = \arg\min_\Theta \mathcal{L}(F_\Theta(X_T), X_0), \tag{4}$$

where $\mathcal{L}$ can be any pixel-wised reconstruction loss, *e.g.*, $L1$ and $L2$ loss. However, due to the over-simplified assumption of the distribution of normally-exposed images, such a training paradigm usually leads to the residual of unnatural artifacts in outputs [41]. To integrate the advantages of the learnable conditional distribution of normally-exposed images and the clear formulation of the noise in raw space simultaneously, we propose a novel method to learn to expose which is illustrated in the following section.

## 3.2. Learning to expose

### 3.2.1 The formulation of the training objective

To enhance the visibility of a low-light image $X_T$, we aim to learn a model $F_\Theta$ that can maximize the likelihood of its reference images $X_0$ parameterized by $\Theta$ over the distribution of the training data $q(X_0, X_T)$, *i.e.*,

$$\Theta = \arg\max_\Theta \mathbb{E}_q[p_\Theta(X_0|X_T)]. \tag{5}$$

Due to the difficulty of directly estimating the likelihood of an image, we further formulate $p_\Theta(X_0|X_T)$ as follows,

$$p_\Theta(X_0|X_T) = \frac{\int p_\Theta(X_{0:T})dX_{1:T-1}}{p_\Theta(X_T)}, \tag{6}$$

so that maximizing $\mathbb{E}_q[\log[p_\Theta(X_0|X_T)]]$ is equivalent to maximizing the likelihood of the joint distribution $\mathbb{E}_q[\log[p_\Theta(X_{0:T})]]$, *i.e.*, minimizing the cross entropy between the learned exposure process and the real one. Therefore, we propose to minimize the upper bound of the divergence between $p_\Theta(X_{0:T})$ and $q(X_{0:T})$ as follows

$$\begin{aligned}&\mathcal{D}_{KL}(p_\Theta(X_{0:T})||q(X_{0:T})) \leq \\ &\mathbb{E}_{q(X_{ref})}[\mathcal{D}_{KL}(p_\Theta(X_T)||q(X_T|X_{ref}))+ \\ &\sum_{t=1}^{T}\mathbb{E}_{p_\Theta(X_t)}[\mathcal{D}_{KL}(p_\Theta(X_{t-1}|X_t)||q(X_{t-1}|X_t, X_{ref}))]],\end{aligned} \tag{7}$$

where $X_{ref}$ is the expected clean image, *i.e.*, when $N$ in Eq. 1 is a zero matrix, and $q(X_{0:T})$ is the ground-truth distribution of the exposure process. $X_0$ can be approximately regarded as $X_{ref}$ if its exposure time is long enough. The detailed derivation can be found in the supplement.

### 3.2.2 Training strategy

We do not need to optimize the first term in the proposed upper bound in Eq. 7 since we aim to learn a restoration model instead of a generative model. For the second term in Eq. 7, it calculates the divergence between the distribution of predicted image $p_\Theta(X_{t-1}|X_t)$, *i.e.*, the image with slightly longer exposure time and higher signal-noise-ratio (SNR) than $X_t$, and the real exposure process. The real exposure process $q(X_{t-1}|X_t, X_{ref})$ is well-defined based on the noise model in Eq. 2 as follows

$$q(X_{t-1}|X_t, X_{ref}) = \mathcal{P}\left(\frac{X_{t-1}-X_t}{K}; \frac{(\lambda_{t-1}-\lambda_t)X_{ref}}{\lambda_{ref}K}\right), \tag{8}$$

where $\frac{(\lambda_{t-1}-\lambda_t)X_{ref}}{\lambda_{ref}K}$ is the rate of the Poisson distribution $\mathcal{P}$. Namely, the increment part in the count of photons obeys Poisson distribution. For the design of $p_\Theta(X_{t-1}|X_t)$, the increment part is assumed to obey the following Poisson distribution

$$p_\Theta(X_{t-1}|X_t) = \mathcal{P}\left(\frac{X_{t-1}-X_t}{K}; \frac{(\lambda_{t-1}-\lambda_t)F_\Theta(X_t)}{\lambda_{ref}K}\right), \tag{9}$$

**Algorithm 1** Training (default $T = 2$)

**while** not converged **do**
  sample $(X_T, \lambda_T, X_{ref}, \lambda_{ref})$ from either synthetic dataset or real-captured paired dataset
  $L = 0$
  **for** $t = T, T-1, ..., 0$ **do**
    $\hat{X}_{ref} = F_\Theta(X_t)$
    $L = L + L_t(\hat{X}_{ref}, X_{ref})$ following Eq. 11
    **if** $t > 0$ **then**
      $X_{t-1} \sim p_\Theta(X_{t-1}|X_t)$ following Eq. 10.
    **end if**
  **end for**
  Perform a gradient descent step on $\nabla_\Theta L$
**end while**

**Algorithm 2** Inference (default $T = 1$)

**Input:** model $F_\theta$, desired exposure time $\lambda_{ref}$, low-light image $X_T$, and its exposure time $\lambda_T$
**for** $t = T, T-1, ..., 0$ **do**
  $\hat{X}_{ref} = F_\Theta(X_t)$
  $X_{t-1} \sim p_\Theta(X_{t-1}|X_t)$ following Eq. 10.
**end for**
return $F_\Theta(X_0)$

which can be trivially sampled as follows,

$$\begin{aligned} X_{t-1} &= X_t + \Delta X + K\Delta N_p, \\ \text{where } \Delta X &= \frac{(\lambda_{t-1} - \lambda_t)F_\Theta(X_t)}{\lambda_{ref}} \\ \text{and } &\left(\frac{\Delta X}{K} + \Delta N_p\right) \sim \mathcal{P}\left(\frac{\Delta X}{K}\right). \end{aligned} \tag{10}$$

After defining the formulation of two terms in $\mathcal{D}_{KL}(p_\Theta(X_{t-1}|X_t)||q(X_{t-1}|X_t, X_{ref}))$, the pixel-wise reconstruction loss can be optimized as follows

$$L_t = \mathbb{E}_{p_\Theta}[F_\Theta(X_t) \cdot \log \frac{F_\Theta(X_t)}{X_{ref}} + X_{ref} - F_\Theta(X_t)], \tag{11}$$

and the derivation can be found in the supplement. It is worth noting that the expectation of $L_t$ is calculated over the distribution of $p_\Theta$, *i.e.*, the input image with shorter exposure time should be sampled from the distribution parameterized by $\Theta$. The details of the proposed training and inference procedures are in Algorithms 1, 2 and Fig. 2.

#### 3.2.3 Adaptive residual layer

Although there are no inherent restrictions on the network design imposed by the proposed algorithm, making specific modifications to the network can further enhance its performance. Specifically, we find that although the proposed inference algorithm can overall improve the quality of restored images, it may increase the error in bright areas, *e.g.*, light bulbs. Namely, due to the high signal-noise ratio in the bright area, the reconstructed result of the initial step may be the most accurate one and may be further degraded by subsequent refinements. To solve this problem, we propose an adaptive residual layer. Specifically, the network $F_\Theta$ is designed to predict normally-exposed image $X_{ref}$, the noise residual $R = X_{ref} - \frac{\lambda_{ref} X_t}{\lambda_t}$, and a soft mask $M$ simultaneously, and the final output $F_\Theta(X_t)$ is as follows

$$F_\Theta(X_t) = M \cdot \lfloor \hat{X} \rceil + (1 - M) \cdot \lfloor \frac{\lambda_{ref}}{\lambda_t} X_t + \hat{R} \rceil, \tag{12}$$

where $\hat{X}$ and $\hat{R}$ are the predicted reference image and the residual respectively, and $\lfloor \rceil$ is the $[0, 1]$ clip operation[3]. More specifically, the only change in the architecture is the increase in the number of output channels. For example, if the number of raw image channels is 4, the proposed network will have 9 channels for output, which include 4 channels for $\hat{X}$, 4 channels for $R$, and one channel for $M$.

### 3.3. Comparison with diffusion models

In this section, we compare the proposed algorithm with diffusion models [36, 13, 34] since they all involve density estimation and progressive refinement. The differences are summarized in Table 1. As shown in the table, the proposed method has a different motivation and formulation, leading to different inductive biases and model performance. The main advantages of the proposed method are as follows: first, each intermediate step $X_t$ obeys the physics-based noise distribution in the proposed method while it is not satisfied in previous diffusion models. This consistency makes the proposed model a better generalization ability towards different noise levels and does not need to spend model capacity to learn the Gaussian denoising. In addition, fewer inference steps can be achieved by staring from a low-light image instead of pure noise. Besides, explicitly integrating the accumulative error, *i.e.*, the divergence between $q(X_{t-1}|X_T, X_{ref})$ and $p_\Theta(X_{t-1}|X_t)$, into the training process enables the proposed method to achieve higher fidelity results compared to vanilla diffusion models. More details can be found in the supplementary material.

## 4. Experiment

### 4.1. Experimental setting

**Implementation details.** We evaluate the performance of the proposed method on two widely-used raw image low-light enhancement datasets: ELD [45] and SID [1]. Specifically, for the commonly used Bayer array, SID [1] contains 2697 pairs of raw images under dark environments, which are captured under different ISO and amplification ratios, *e.g.*, $\times 100$, $\times 250$, and $\times 300$. We use the same split as [45] for SID [1] dataset, and train all the models using its training set. In this work, ELD [45] is used for additional evaluation

[3] To stabilize the training process, the input to the network is amplified to match the brightness of the reference images. However, before sampling $p_\Theta(X_{t-1}|X_t)$, the images are converted to photon counts.

(a) Input (a) BM3D (a) N2N (a) P+G (a) Ours (a) Reference

Figure 3: Low-light image enhancement results on both indoor and outdoor environments after the same ISP pipeline for better visualization. The results in *Ours* are obtained by using the same noise model and backbone as P+G.

of the generalization ability of models under different scenarios and devices. For the modeling of the real noise distribution, we adopt the widely-used P+G model [7, 45] as the baseline, in which the distribution of the signal-dependent noise is modeled as Poisson distribution, and the signal-independent noise is set to Gaussian noise as default. The impact of the choice of signal-independent noise is further explored in Sec. 4.3. For all experiments, we use the real-captured paired data for evaluation. More details can be found in the supplementary material.

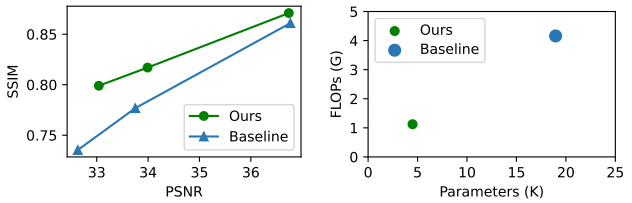
### 4.2. Comparison with different methods.

Since there are few works exploring the network design of the low-light image enhancement under raw space, we include the following competitors for comparison: the typical non-deep methods which do not require paired data for training, *e.g.*, BM3D [4], and A-BM3D [28]. The model that utilizes paired noise image for training, *i.e.*, Noise2Noise (N2N) [20]. The model trained with the synthetic paired data, *i.e.*, P+G [7, 45]. Specifically, P+G [7, 45] means we train the model using the synthetic image with the Possion-Gaussian noise model. Ours uses the same settings with P+G [7, 45], *e.g.*, the same noise model, and almost the same network architecture (the only difference is the number of input/output channels), except that ours is trained/evaluated by the proposed algorithm. Since involving the proposed adaptive residual layer has little impact on the model complexity[4], we claim that they are the same architecture/backbone for consistency. The evaluation results on SID [1] are reported in Table 2 and the results on ELD [45] are in Table 3. As we can see in these tables, the deep learning-based methods tend to achieve better performance than the non-deep methods even if trained without using paired-real data. Besides, the proposed method achieves better performance than P+G [7] under the same noise model and backbone network. Some visual results are shown in Fig. 3.

### 4.3. The results with different noise models

To explore whether the proposed strategy is compatible with different modeling of the exposure process, we further evaluate the performance of models trained on synthetic data synthesized by SOTA noise models [45, 6] and paired real data. Specifically, compared with P-G noise

[4]The number of parameters/FLOPs increases from 7.761M/54.83G to 7.762M/55.17G after involving the proposed adaptive residual layer.




(a) The performance of models under amplifying ratios of $[100, 250, 300]$.

(b) Comparisons of models in the number of parameters, FLOPs, and inference time.

Figure 4: The comparison on SID [1] dataset between a small model (ours) with the proposed method, and a baseline larger model. The proposed method can achieve better performance using around 25% of parameters and FLOPs of the larger model. Even taken iterations into consideration, the inference time (represented as the point size in (b)) of the proposed method is still shorter than the larger model.

| Model | ×100 PSNR / SSIM | ×250 PSNR / SSIM | ×300 PSNR / SSIM |
|---|---|---|---|
| BM3D [4] | 32.92/0.758 | 29.56/0.686 | 28.88/0.674 |
| A-BM3D [28] | 33.79/0.743 | 27.24/0.518 | 26.52/0.558 |
| N2N [20] | 37.42/0.853 | 33.48/0.725 | 32.37/0.686 |
| P+G [7, 45] | 38.31/0.884 | 34.39/0.765 | 33.37/0.730 |
| Ours | **38.89 / 0.902** | **36.02 / 0.832** | **35.00 / 0.808** |

Table 2: Quantitative results on Sony subset of SID.

| Camera | Ratio | Metrics | BM3D | N2N | P+G | Ours |
|---|---|---|---|---|---|---|
| Sony A7S2 | ×100 | PSNR | 37.69 | 41.63 | 42.46 | 43.29 |
| | | SSIM | 0.803 | 0.856 | 0.889 | 0.929 |
| | ×200 | PSNR | 34.06 | 37.98 | 38.88 | 40.39 |
| | | SSIM | 0.696 | 0.775 | 0.812 | 0.873 |
| Nikon D850 | ×100 | PSNR | 33.97 | 40.47 | 40.29 | 40.89 |
| | | SSIM | 0.725 | 0.848 | 0.845 | 0.897 |
| | ×200 | PSNR | 31.36 | 37.98 | 37.26 | 38.51 |
| | | SSIM | 0.618 | 0.820 | 0.786 | 0.856 |
| Canon EOS70D | ×100 | PSNR | 30.79 | 38.21 | 40.94 | 40.99 |
| | | SSIM | 0.589 | 0.826 | 0.934 | 0.944 |
| | ×200 | PSNR | 28.06 | 34.33 | 37.64 | 37.90 |
| | | SSIM | 0.540 | 0.704 | 0.873 | 0.874 |
| Canon EOS700D | ×100 | PSNR | 29.70 | 38.29 | 40.08 | 40.19 |
| | | SSIM | 0.556 | 0.859 | 0.897 | 0.918 |
| | ×200 | PSNR | 27.52 | 34.94 | 37.86 | 37.71 |
| | | SSIM | 0.537 | 0.766 | 0.879 | 0.878 |

Table 3: The quantitative results on ELD [45] dataset.

model [7], [45] additionally models the row noise and further refines the modeling of signal-independent noise. Most recently, [6] collects a set of dark frames to correct the dark shading. By subtracting the dark shading in the preprocessing pipeline, the performance is further improved. As shown in Table 4, the more accurate noise models we use, the better performance we achieved. The models based on SOTA noise models even achieved slightly better performance than that trained on the real paired data. By training models using the proposed method, we achieve consistent improvements in performance across all models. the results serve as a strong indication of the effectiveness and generality of the proposed model.

| | Model | Baseline PSNR / SSIM | w/ Ours PSNR / SSIM |
|---|---|---|---|
| ×100 | P+G [7] | 38.31 / 0.884 | **38.89 / 0.902** |
| | Paired data | 38.60 / 0.912 | **38.98 / 0.915** |
| | ELD [45] | 39.27 / 0.914 | **39.37 / 0.917** |
| | PMN [6] | 39.77 / 0.919 | **39.80 / 0.920** |
| ×250 | P+G [7] | 34.39 / 0.765 | **36.02 / 0.832** |
| | Paired data | 37.08 / 0.886 | **37.45 / 0.895** |
| | ELD [45] | 37.13 / 0.883 | **37.47 / 0.889** |
| | PMN [6] | 37.68 / 0.892 | **37.90 / 0.896** |
| ×300 | P+G [7] | 33.37 / 0.730 | **34.99 / 0.808** |
| | Paired data | 36.29 / 0.874 | **36.82 / 0.888** |
| | ELD [45] | 36.30 / 0.872 | **36.78 / 0.878** |
| | PMN [6] | 37.01 / 0.881 | **37.27 / 0.888** |

Table 4: The performance of models trained with paired real data and different noise models on SID [1].

### 4.4. The results with different backbone models

To further explore whether the proposed method is compatible with different model sizes, we conduct experiments utilizing different backbone networks and model sizes. For the backbone networks, the widely used UNet [45] backbone and a most recent SOTA backbone NAFNet [2] are used for evaluation. We also evaluate NAFNet [2] with different model sizes to explore the effect of different model capacities. The performance of different backbones and model sizes are reported in Table 5. As shown in the table, the proposed method can stably improve the image quality of enhancement results under different backbones/model sizes. Besides, the most significant improvement is achieved for small models, which makes it possible to deploy a small model on mobile devices and introduce the proposed inference algorithm to boost performance for extreme low-light cases. The verification can be seen in Fig. 4, in which we find that the proposed method can achieve better performance using only around 25% parameters and FLOPs of the larger feedforward model. It is still faster than the larger one for inference even if we use an iteration number of 3, and can be further sped up by using a smaller number of iteration steps or no iteration for cases that are not very dark.

### 4.5. Generalization ability

One challenge of the existing methods is that better performance on benchmarking datasets may lead to worse out-of-distribution (O.O.D) performance. To evaluate the gen-

(a) Low light image (b) Reference image (c) Adaptive mask $M$ (d) w/o ARL (PSNR 32.57) (e) w/ ARL (PSNR 35.44)

Figure 5: An illustration of the proposed adaptive residual layer (ARL). (d) and (e) are the maps of error magnitude change of models w/o and w/ ARL after iterative refinement. For the highlight areas, *e.g.*, the area in the bounding box, the iterative refinement tends to lead to a larger value of the absolute error if w/o the proposed ARL.

| | Noise model | Model | Baseline PSNR / SSIM | w/ Ours PSNR / SSIM |
|---|---|---|---|---|
| ×100 | P+G [7] | UNet | 38.31 / 0.884 | **38.88 / 0.901** |
| | | NAFNet-1 | 32.37 / 0.698 | **36.74 / 0.871** |
| | | NAFNet-2 | 39.32 / 0.909 | **39.71 / 0.917** |
| ×100 | ELD [45] | UNet | 39.27 / 0.914 | **39.37 / 0.917** |
| | | NAFNet-1 | 35.92 / 0.848 | **36.87 / 0.879** |
| | | NAFNet-2 | 39.52 / 0.917 | **39.75 / 0.918** |
| ×250 | P+G [7] | UNet | 34.39 / 0.765 | **36.02 / 0.832** |
| | | NAFNet-1 | 30.33 / 0.594 | **33.99 / 0.817** |
| | | NAFNet-2 | 36.57 / 0.849 | **37.58 / 0.886** |
| ×250 | ELD [45] | UNet | 37.13 / 0.883 | **37.47 / 0.889** |
| | | NAFNet-1 | 32.34 / 0.726 | **34.07 / 0.821** |
| | | NAFNet-2 | 37.39 / 0.884 | **37.90 / 0.889** |
| ×300 | P+G [7] | UNet | 33.37 / 0.730 | **34.59 / 0.798** |
| | | NAFNet-1 | 29.52 / 0.549 | **33.04 / 0.794** |
| | | NAFNet-2 | 35.83 / 0.835 | **36.90 / 0.877** |
| ×300 | ELD [45] | UNet | 36.30 / 0.872 | **36.78 / 0.878** |
| | | NAFNet-1 | 31.05 / 0.668 | **33.12 / 0.800** |
| | | NAFNet-2 | 36.53 / 0.871 | **37.23 / 0.884** |

(a) Performance with different models.

| Model | Parameters | FLOPs | Inference time |
|---|---|---|---|
| UNet | 7.762M | 55.17G | 0.1243s |
| NAFNet-1 | 4.697K | 1.124G | 0.0468s |
| NAFNet-2 | 6.871M | 15.32G | 0.3514s |

(b) Computational cost of each model.

Table 5: Performance of models w/ and w/o the proposed method under different noise models and backbones.

eralization ability of the proposed model, we train models on the ×100 task of SID dataset [1] and evaluates their performance on the ×250 and ×300 tasks. The SOTA noise model and network architecture are utilized, *i.e.*, PMN [6] and NAFNet [2]. The results are reported in Table 6. As shown in the table, the proposed method can alleviate the performance degradation caused by the domain gap. We conjecture the reason is that even if a single step of denoising is not accurate enough, by involving slightly more denoising steps, the performance gap can be alleviated.

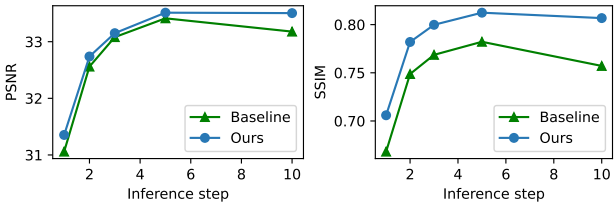



Figure 6: The performance of models with different numbers of iterations for $F_\Theta$. *Baseline* represents the model w/o the proposed training paradigm and improved architecture. All the models are trained based on the NAFNet-1 backbone in Table 4 (b) and ELD [45] noise model.

| | Method | PSNR | SSIM |
|---|---|---|---|
| ×250 | Baseline | 37.37 | 0.8741 |
| | Ours | **37.73** | **0.8846** |
| ×300 | Baseline | 36.31 | 0.8512 |
| | Ours | **36.83** | **0.8650** |

Table 6: The generalization ability of methods on O.O.D tasks. The models are trained on the ×100 task and evaluated on ×250 and ×300 tasks. All the models use the same SOTA noise model PMN [6] and backbone NAFNet-2 [2]. *Baseline* is not equipped with the improved architecture and the proposed training paradigm.

### 4.6. Ablation study

**The impact of different inference steps.** Similar to previous works that involve progressive refinement, the number of inference steps is a key hyper-parameter that we need to set manually. Therefore, we evaluate the effect of different inference steps and the results can be seen in Fig. 6. As shown in the figure, in the first few steps, the performance of all the models increases monotonously. However, the baseline method has an obvious performance degradation when the number of inference steps increases to 10, due to the mismatch between the inference distribution and the training one. Benefiting from the better matching of the training

and testing distribution and the improved architecture, the proposed method has better initial results and performs relatively stable at a relatively large inference step. It is worth noting that we evaluate the metrics on $\hat{X}_{ref}$ of each step instead of $X_t$, which are more accurate to evaluate the effectiveness of the proposed method.

**Adaptive residual layer.** To better understand the role of the proposed adaptive residual layer, an example is provided in Fig. 5. As we can see in the figure, the highlight areas of the low-light image and reference image are almost the same after multiplying the amplifying ratio to the low-light image [14]. Progressively refining the highlight areas may make their values deviate from the real values due to the inductive bias of neural networks if we directly predict a clean image from a noisy one. By introducing the proposed adaptive residual layer, the model relies more on the results by residual-based denoising, *i.e.*, the second term in Eq. 12, which tends to predict noises with zero mean so that the intensity level is less affected in the highlight areas. As shown in the figure, the results after iteration without the proposed ARL degrade the fidelity in highlights while the proposed method greatly solves this problem.

## 5. Conclusion

In this paper, we propose a novel strategy for raw-image enhancement. Specifically, we propose to utilize a shared-weight network to simulate the physics-based exposure process by minimizing their KL divergence in an iterative end-to-end manner. An adaptive residual layer is further proposed to alleviate the fidelity deterioration caused by the iterative refinement in the highlight areas. We evaluate the effectiveness of the proposed algorithm on two benchmarks and the results demonstrate that the proposed method can stably improve the performance combined with real-paired data, different noise models, and different backbones. Besides, the proposed method also achieves better generalization ability in unseen amplifying ratios.

**Acknowledgement.** This work was done at Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University. This research is supported in part by the NTU-PKU Joint Research Institute (a collaboration between the Nanyang Technological University and Peking University that is sponsored by a donation from the Ng Teng Fong Charitable Foundation), the Basic and Frontier Research Project of PCL, the Major Key Project of PCL, the MOE AcRF Tier 1 (RG61/22) and Start-Up Grant, and Hong Kong Jockey Club Charities Trust - JC STEM Lab Project.

## References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 3, 5, 6, 7, 8

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *Computer Vision–ECCV 2022: 17th European Conference, October 23–27, 2022, Proceedings, Part VII*, pages 17–33, 2022. 7, 8

[3] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *arXiv preprint arXiv:2206.00941*, 2022. 3

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 6, 7

[5] Xingbo Dong, Wanyan Xu, Zhihui Miao, Lan Ma, Chao Zhang, Jiewen Yang, Zhe Jin, Andrew Beng Jin Teoh, and Jiajun Shen. Abandoning the bayer-filter to see in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17431–17440, 2022. 3

[6] Hansen Feng, Lizhi Wang, Yuzhi Wang, and Hua Huang. Learnability enhancement for low-light raw denoising: Where paired real data meets noise modeling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1436–1444, 2022. 2, 3, 6, 7, 8

[7] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing*, 17(10):1737–1754, 2008. 3, 6, 7, 8

[8] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. In *Advances in Neural Information Processing Systems*. 3

[9] Lanqing Guo, Renjie Wan, Guan-Ming Su, Alex C Kot, and Bihan Wen. Multi-scale feature guided low-light image enhancement. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 554–558. IEEE, 2021. 2

[10] Lanqing Guo, Renjie Wan, Wenhan Yang, Alex Kot, and Bihan Wen. Enhancing low-light images in real world via cross-image disentanglement. *arXiv preprint arXiv:2201.03145*, 2022. 2

[11] Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. 2023. 3

[12] Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, pages 1–19, 2022. 2

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 2, 3, 4, 5

[14] Haofeng Huang, Wenhan Yang, Yueyu Hu, Jiaying Liu, and Ling-Yu Duan. Towards low light enhancement with raw

images. *IEEE Transactions on Image Processing*, 31:1391–1405, 2022. 1, 3, 9

[15] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *Computer Vision–ECCV 2022: 17th European Conference, 2022, Proceedings, Part XIX*, pages 163–180. Springer, 2022. 2

[16] Xin Jin, Jia-Wen Xiao, Ling-Hao Han, Chunle Guo, Ruixun Zhang, Xialei Liu, and Chongyi Li. Lighting every darkness in two pairs: A calibration-free pipeline for raw denoising. 2023. 3

[17] Yeying Jin, Beibei Lin, Wending Yan, Wei Ye, Yuan Yuan, and Robby T. Tan. Enhancing visibility in nighttime haze images using guided apsf and gradient adaptive convolution, 2023. 1, 2

[18] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVII*, pages 404–421. Springer, 2022. 2

[19] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*. 3

[20] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 6, 7

[21] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):9396–9416, 2021. 1

[22] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *International Journal of Computer Vision*, 129(4):1153–1184, 2021. 1

[23] Kin Gwn Lore et al. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2

[24] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022. 3

[25] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. Srflow: Learning the super-resolution space with normalizing flow. In *European Conference on Computer Vision*, pages 715–732. Springer, 2020. 2

[26] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, page 220, 2018. 2

[27] Hengyuan Ma, Li Zhang, Xiatian Zhu, and Jianfeng Feng. Accelerating score-based generative models with preconditioned diffusion sampling. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIII*, pages 1–16. Springer, 2022. 3

[28] Markku Makitalo and Alessandro Foi. Optimal inversion of the anscombe transformation in low-count poisson image denoising. *IEEE transactions on Image Processing*, 20(1):99–109, 2010. 6, 7

[29] Ashwini Pokle, Zhengyang Geng, and J Zico Kolter. Deep equilibrium approaches to diffusion models. In *Advances in Neural Information Processing Systems*. 3

[30] Abhijith Punnappurath, Abdullah Abuolaim, Abdelrahman Abdelhamed, Alex Levinshtein, and Michael S Brown. Day-to-night image synthesis for training nighttime neural isps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10769–10778, 2022. 3

[31] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019. 2

[32] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

[33] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. 3

[34] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 2, 3, 4, 5

[35] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. MSR-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017. 2

[36] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 2, 3, 4, 5

[37] Li Tao, Chuang Zhu, Guoqing Xiang, Yuan Li, Huizhu Jia, and Xiaodong Xie. LLCNN: A convolutional neural network for low-light image enhancement. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4, 2017. 2

[38] Chulin Wang, Kyongmin Yeo, Xiao Jin, Andres Codas, Levente J Klein, and Bruce Elmegreen. S3rp: Self-supervised super-resolution and prediction for advection-diffusion process. *arXiv preprint arXiv:2111.04639*, 2021. 3

[39] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. *arXiv preprint arXiv:2212.11548*, 2022. 2

[40] Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, Zhiwei Xiong, Wei Zhang, and Feng Wu. Progressive retinex: Mutually reinforced illumination-noise perception network for

low-light image enhancement. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2015–2023, 2019. 2

[41] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2604–2612, 2022. 2, 3, 4

[42] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. Beyond learned metadata-based raw image reconstruction. *arXiv preprint arXiv:2306.12058*, 2023. 3

[43] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. Raw image reconstruction with learned compact metadata. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18206–18215, 2023. 3

[44] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2

[45] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2758–2767, 2020. 2, 3, 5, 6, 7, 8

[46] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022. 3

[47] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5901–5910, 2022. 2

[48] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17714–17724, 2022. 2

[49] Qinsheng Zhang, Molei Tao, and Yongxin Chen. gddim: Generalized denoising diffusion implicit models. *arXiv preprint arXiv:2206.05564*, 2022. 2, 3

[50] Rongkai Zhang, Lanqing Guo, Siyu Huang, and Bihan Wen. Rellie: Deep reinforcement learning for customized low-light image enhancement. In *Proceedings of the 29th ACM international conference on multimedia*, pages 2429–2437, 2021. 2

[51] Yi Zhang, Hongwei Qin, Xiaogang Wang, and Hongsheng Li. Rethinking noise synthesis and modeling in raw denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4593–4601, 2021. 3

[52] Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4439–4448, 2021. 2