EL^2NM: Extremely Low-light Noise Modeling Through Diffusion Iteration

Jiahao Qin Pinle Qin* Rui Chai Jia Qin Zanxia Jin
North University of China
Taiyuan City, Shanxi Province, China
q656673477@gmail.com, {qpl, 20160007, jqin, zanxiajin}@nuc.edu.cn

Abstract

Low-light Original Denoising (LOD) is a challenging task in Computational Photography (CP). The low number of photons in low light environments makes imaging very difficult. The most difficult step in LOD is to establish a noise model under low light. Currently, there are numerous approaches aim to noise modeling, however the noise established have significant differences from real noise due to the highly intricate distribution of noise. Towards this goal, this paper proposes an Extremely Low-light Noise Modeling (EL^2NM) approach, which designs an original image condition constraint module and a multi-noise fusion module to generate complex noise consistent with real scenes. In order to satisfy the complex noise distribution in low-light environments instead of just Gaussian noise, we integrate various noises into cold diffusion to establish a realistic noise generation model for extremely low-light environments. At the same time, to avoid the image semantic misinterpret during the reverse diffusion process, we propose to use conditional image to guide noise generation of the diffusion model. Extensive experiments demonstrate that our proposed method EL^2NM exhibits excellent performance in extremely low-light environments and achieves the state-of-the-art on Starlight Dataset.

1. Introduction

With the advancements in imaging sensors and computer technology, computational photography has experienced rapid growth [35]. Certain wildlife species (such as moths) are active in moonless nights, special forces engage in nocturnal operations, and monitoring nighttime sewage discharge requires functioning in extremely low-light conditions. Weak light imaging technology holds significant applications in fields like defense, wildlife observation, mobile photography [20], and astronomy [27]. However, due to limited photon counts and inevitable noise, imaging in low-light and nighttime scenarios becomes exceedingly challenging. A straightforward solution is to widen the aperture or extend exposure time to capture more photons. Nevertheless, these solutions inherently trade off noise, failing to eliminate it entirely.

Deep generative models have achieved success in the realm of weak light denoising. For instance, they have learned to perceive objects in the dark [8], detect motion in the dark [9], and discern moving objects in darkness [24]. Simultaneously, there exist numerous classical denoising algorithms like BM3D [10], VBM4D [26], mean filtering [13], and Kalman filtering [5]. However, these methods are typically built upon a simplistic Gaussian model, which isn't well-suited for noise in extremely low-light conditions. Even sophisticated heteroscedastic Gaussian noise models [16] struggle to accurately capture noise in weak light environments. In situations of extremely low-light, noise often follows non-Gaussian distributions that adhere to sensor characteristics, making their modeling complex. Recently, efforts have been made to study noise characteristics generated during imaging in dim light environments. Among these, Extreme Low-light Denoising ELD [36] approaches noise from a physical imaging perspective, delving into the noise generated at each step of low-light imaging. ELD breaks down the imaging process into four stages: photons, electrons, voltage, and digital signals, thereby reasonably representing noise arising in dim-light environments. Building upon ELD, the dancing in the starlight [29] approach redefines noise types and employs a Wasserstein GAN [19] network to learn parameters for each noise type. This methodology produces high-quality noise images, showcasing for the first time a video of individuals dancing under starlight.

Inspired by recent advancements in conditional diffusion models [11, 22] and cold diffusion models [4], this paper introduces a new noise modeling approach, EL^2NM, building upon the work of Kristina Monakhova [29]. EL^2NM employs a series of refinement steps to convert complex noise distributions into empirical data distributions, akin to Langevin dynamics. At its core lies the U-Net architecture [32], utilized to train the noise model and iteratively produce noise outputs. The U-Net architecture, adapted from

SR3 [34], is modified in this work to accommodate conditional image generation. The sampling technique of the cold diffusion model is applied iteratively to generate the final noise image. In comparison to the GAN-based approach used in Dancing under the Stars, our method doesn't require adversarial training, making noise image generation more straightforward. Moreover, the quality of generated noise images is higher, and the resulting noise image effects are similar to those achieved by dancing under the starlight.

Our main contributions of this paper are summarized as follows:

1: We proposes a novel approach for noise modeling in low-light environments. By combining the conditional diffusion model with the cold diffusion model, the method employs the conditional diffusion model to control the generation of conditional images and integrates the cold diffusion model to introduce a wider range of noise distributions. Through iterative refinement, the approach achieves the creation of high-quality noise images. Importantly, this study represents the first application of diffusion models in the domain of noise modeling for low-light images.

2: We extends the application of the conditional diffusion model to noise image generation. The EL^2NM method involves an iterative subdivision technique to generate noise images in low-light conditions. It departs from an understanding of the physical processes involved, sidestepping the need for adversarial training, yet yielding high-quality noise images for dim-light scenarios.

3: Experimental results affirm that the proposed approach exhibits a high level of advancement in both quantitative and qualitative evaluations. It effectively generates superior quality noise images. Moreover, the method demonstrates remarkable performance on the Starlight Dataset [1], showcasing its competitiveness in the field.

2. Related Work

2.1. Noise Modeling

Astronomy, wildlife observation, and military sectors, among others, demand noise reduction in computational photography. However, the scarcity of denoising datasets has hindered the advancement of low-light photography. The key to denoising low-light images lies in accurately establishing corresponding noise models. In recent years, numerous video and image denoising methods have emerged. Classical denoising approaches often rely on prior knowledge about images, such as self-similarity [6, 10, 26, 30], smoothness [31, 33], sparsity [3, 12], and low-rank properties [18]. In contrast to pre-defined methods, learning-based [39] techniques obtain image priors by learning the distribution of image data. Recent research indicates that learning-based methods offer significant improvements in image quality compared to classical methods. However, due to the lack of paired datasets, learning-based methods tend to exhibit fragile learnability. These methods often need to make simple assumptions about noise statistics, such as heteroscedastic Gaussian distribution. Nevertheless, such noise models often struggle to accurately fit noise in low-light environments. Consequently, the complex data distribution coupled with limited data availability constrains the development of learning-based methods.

In order to overcome the challenges posed by complex data distributions, a series of studies have focused on enhancing the realism of noise modeling in low-light and nighttime scenarios. Such research stems from the inherent characteristics of sensors, considering noise introduced by the sensors themselves, including photon shot noise and readout noise, to precisely model potential noise under low-light conditions. Examples of these include Poisson-Gaussian noise [16], ELD [36], and Rethink noise [40] methods. Other statistical models including Poisson Mixture model[38], mixed AWGN with Random Value Impulse Noise (RVIN)[42] and Gaussian Mixture Model[44] are also proposed to model real noise. A common approach in low-light photography is to merge multiple images for denoising, known as burst denoising [17, 21, 28]. However, these methods often necessitate substantial paired datasets for training, which can be challenging to obtain in real-world scenarios. To address the limitation of available data, Hansen Feng [14] and others have utilized noise modeling to reconstruct paired real data, effectively augmenting the dataset of original low-light images.

Recently, a multitude of deep learning-based methods have emerged, aiming to address the challenge of extreme environment image denoising by learning the characteristic distributions of each type of noise. Such methods typically start by gathering pairs of clean/noisy images and then employ deep learning networks to approximate the Image Signal Processing (ISP) pipeline for image denoising. Remarkable success has been achieved even in very low-light conditions (0.1 lux), enabling capabilities like learning to see in the dark [8] or seeing motion in the dark [9]. Kristina Monakhova pioneered the collection of datasets in extremely low-light environments. They built upon ELD to precisely model noise in extreme conditions and employed GAN networks to learn parameters for each noise type. This approach yielded exceptionally high-quality noise models. Furthermore, they trained video denoising networks, allowing for the first time the observation of highly realistic videos in extremely dim conditions ($<$0.001 lux, under starlight). In parallel, approaches based on normalized flows [2] have been used to synthesize realistic noise. However, they do not capture sensor-specific noise characteristics. Experimental results suggest that statistically-based methods grounded in physical properties often outperform Neural Network DNN-based methods [40]. But zhang [41] proposed a real time image and video denoising network have good performance.

2.2. Diffusion Model

Recently, Diffusion Models [23] have emerged as potent tools for generative modeling. These models come in various forms, all centered around the concept of removing random noise. Built on the foundation of the fundamental Unet network, DDPM involves an iterative process where Gaussian noise is progressively added to the image during the forward pass. This noisy image is then input to the network, which produces a denoised output. Originating from Langevin dynamics, diffusion models are understood as stochastic processes of the image density function, with each step involving Gaussian noise. With the advent of diffusion models in DDPM, the entire landscape of generative modeling has been reshaped, transcending the dominance of Generative Adversarial Networks (GANs) as the representative generative models. The core process of DDPM comprises two steps: forward noisy addition and backward denoising. During the forward process, Gaussian noise is continually added to a clear image. Subsequently, the reverse process aims to restore the clear image from the noisy one.

Subsequently, Jonathan Ho [22] proposed a conditional diffusion model. During training, a conditional variable 'y' is incorporated to guide the model's generation process. An alternate approach in Conditional Diffusion Models involves introducing the conditional variable 'y' during the reverse process, training a classifier for conditional generation [11]. Arpit Bansal [4] and others suggested that the forward process of diffusion models is not limited to Gaussian noise; it can utilize any form of noise replacement and even invert any image without noise. Diffusion models have found applications in various fields. Chitwan Saharia [34] and colleagues adapted conditional diffusion models for image super-resolution. Yeying Jin [25] and others employed diffusion models for shadow removal in images. Zhou [43] and others employed diffusion models for 3D reconstruction. Yutong Xie[37] and colleagues modified the diffusion model, considering Gaussian, gamma, and Poisson noise, applying it to image denoising. Moreover, diffusion models have shown their prowess in image and text generation, as well as multimodal generation scenarios. For example, Feng [15] and others employed diffusion models for text to image generation.

3. Method

In this section, we first introduced the noise characteristics and principles in low-light environments. Next, we explained the specific implementation and forward process of our proposed method, which uses the conditional diffusion model. Finally, we described the reverse sampling process of our method using cold diffusion. This approach allows us to generate high-quality noise images without the need for adversarial training. The Fig. 1 summarizes our training and inference process.

3.1. Physically Induced Noise Model

The noise D generated by digital image sensors can be represented using a simple linear model [36]:

$$D = KI + N \tag{1}$$

where I represents the number of photo electrons, K signifies the sum of analog and digital gains, and N accounts for the total noise components. Previous research indicates that the imaging process of a camera can be divided into four stages: photon, electron, voltage, and digital signal. Across these stages, distinct types of noise are generated, influencing the overall image quality. In well-lit environments, where the sensor collects a substantial number of photons, noise has a minor impact on the image and is mainly attributed to photon shot noise, hence modeled as heteroscedastic Gaussian noise. However, in low-light conditions like during nighttime, where the sensor receives a limited number of photons, noise proportionally occupies a larger part of the image. The previous noise models cannot adequately describe the intricate noise characteristics in such dim environments.

Therefore, we approximate the distribution of shot noise and readout noise to Gaussian noise, so that they can be learned by the network:

$$N_sI + N_r \sim \mathcal{N}(x, \lambda_r + \lambda_s x) \tag{2}$$

Line stripe noise usually exists between each frame, and we model it as a Gaussian random distribution with a mean of 0. The quantization noise N_q is the rounding error between the input voltage and output voltage of the Analog-to-Digital Converter (ADC), which follows a uniform distribution:

$$N_q \sim \mathcal{U}\left(-1/2q, 1/2q\right) \tag{3}$$

In addition, we have added fixed mode noise N_p, which is a specific distribution that needs to be separately introduced into the forward process.

At this point, we introduce all the distribution types that all noise follows in weak light environments to ensure that the network can learn all the noise distribution types, thus establishing a complete noise model. The noise model in this paper primarily encompasses shot noise, row noise, read noise, quantization noise, and fixed pattern noise. For the original image captured by the sensor, this paper models the formation process of noise as follows:

$$N = N_s + N_r + N_q + N_p, \tag{4}$$

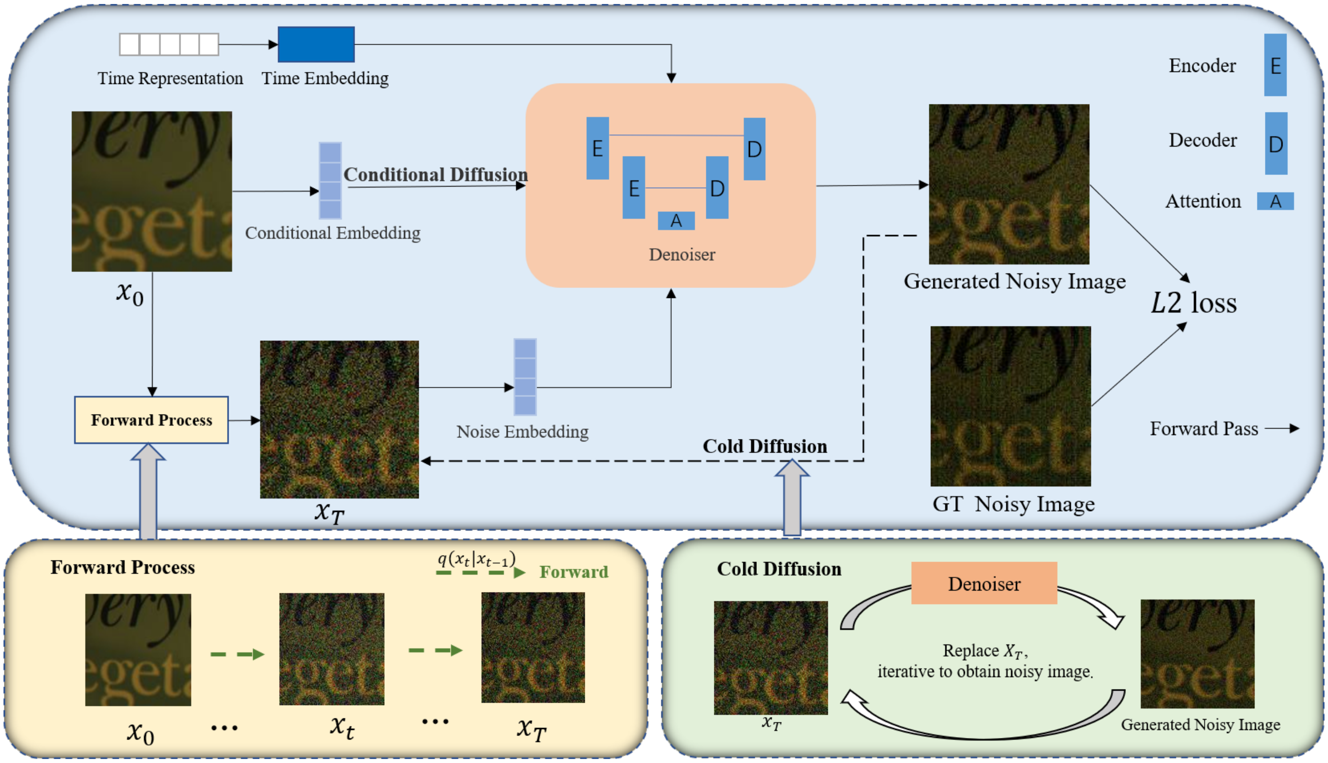

Figure 1. The framework of our method EL^2NM. Given a clear imageX_0, our method gradually introduces diverse noise distributions to obtain a blurry image X_T in the Forward Process. During the training phase, we fed both the clear image X_0 and blurry image X_T into the conditional diffusion model to learn the noise characteristics within the noisy image. During the inference phase, we introduces the inference method of the cold diffusion model to generate the noisy image in an iterative manner, as illustrated by the dashed lines.

where N_s is the shot noise, N_r is the read noise, N_q is the quantization noise and N_p is the fixed pattern noise.

Once the noise model was established, we proceeded by simulating the forward process of the conditional diffusion model to introduce noise into clear images, thereby causing them to become blurred. Subsequently, these images were fed through a Unet network alongside the original clear images. This approach aimed to learn the characteristics of the noise. Further, through reverse inference, the study synthesized noise images.

3.2. Conditional Diffusion

Given a dataset consisting of input images and corresponding output images, it is denoted as: $\mathcal{D} = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{N}$, which represent samples drawn from an unknown conditional distribution $p(y|x)$. This constitutes a many-to-one mapping, where multiple images y might correspond to a single image x. However, this article is more concerned with the mapping where image y corresponds to the same image x, and it aims to learn its stochastic iterative optimization process. This paper employs a denoising diffusion probabilistic model for conditional image generation to address this issue.

The conditional diffusion model can control the output results through conditions. Therefore, this article modifies the DDPM model by modifying the input of the model to two images, introducing a reference image as the condition to guide the output of the diffusion model. The other image inputted by the model is a blurred image that undergoes a forward process, which gradually adds noise to the clear image. The goal of the proposed method in this paper is to iteratively restore the noisy image by using a reverse process conditional on the reference image. Finally, the noise image is obtained using the method proposed in this paper.

During the forward noise addition process, this paper does not apply noise randomly. Instead, the approach involves an analysis inspired by the work conducted under starlight conditions. Specifically, the paper examines the proportions of various types of noise in the resulting image. The method employed entails calculating the changes in Kullback-Leibler divergence upon the addition of each type of noise. This process reveals the extent to which each type of noise influences the noise image. Subsequently, based on the determined noise proportions, the different categories of noise are incorporated into the forward process in ascending order of their calculated proportions. The specific approach

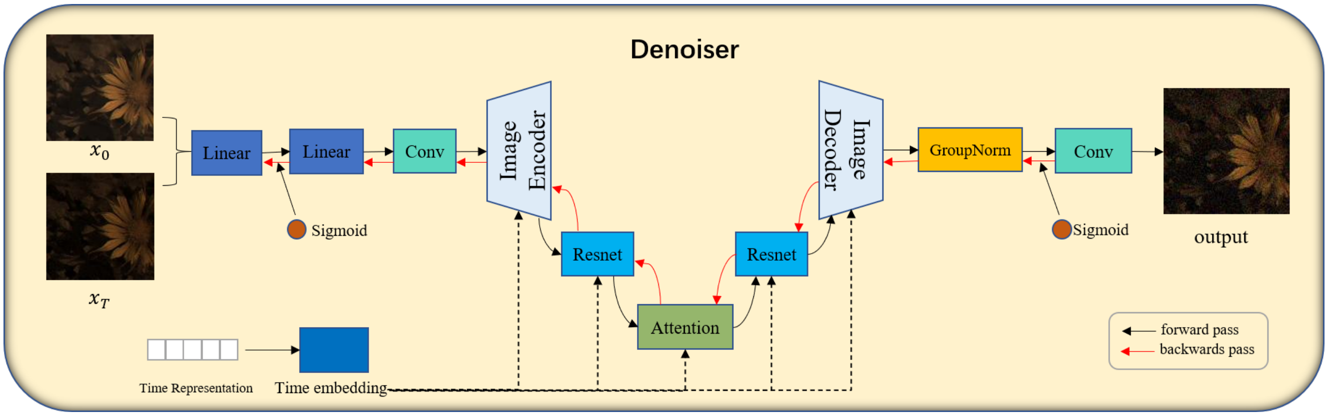

Figure 2. Denoiser: The clear image and the denoised image generated by the forward process are input into the network, and finally the noisy image is output through a loop.

is to add shot noise and readout noise in the first 40 steps, add quantization noise in the next 6 steps, and finally add fixed mode noise. The reason is that shot noise and readout noise account for the largest proportion in weak light environments, while quantization noise and fixed mode noise decrease sequentially.

Given an image $x_0 \in R_N$, considering an operator S that blurs x_0 with a level of t , meaning $t = S(x_0, t)$. The operator S also satisfies $S(x_0, 0) = x_0$. This S corresponds to the forward process described in this paper, which is depicted in Fig. 3.

The architecture of $\mathrm{E}L^2\mathrm{NM}$ is similar to the Unet structure in DDPM. The input image goes through two linear layers and a convolutional layer. It then passes through four encoder layers, with the middle layer consisting of two residual blocks and an attention module. Afterward, it goes through four decoder layers and a GroupNorm layer before producing the final output. The network's structural diagram is illustrated in Fig. 2.

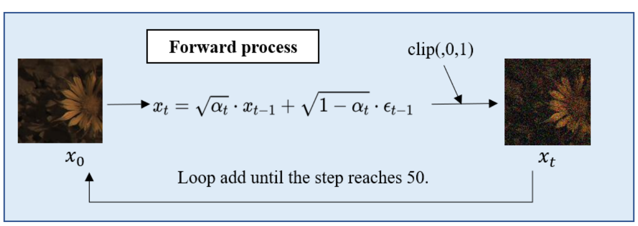

Figure 3. forward process: During the forward process, different distributions of noise are sequentially added to the clear image using the above formula, resulting in an over noisy image.

Previous diffusion models [23] required 1 to 2 thousand diffusion steps during the forward process, which significantly slowed down the generation of noise images. To address this issue, this paper adopts the approach from the cold diffusion model to achieve more efficient forward. This allows the proposed method to utilize fewer forward steps. Inspired by the snowflake operator forward steps in the cold diffusion model, this paper sets the forward steps to 50 and incorporates various types of noise within these 50 steps.

3.3. Refined Inference with Cold Diffusion

In cases of cold diffusion's smooth/differentiable scenarios (non-Gaussian random noise), we have added various types of noise, standard diffusion model inference performs poorly. Hence, the approach from cold diffusion's algorithm is adopted for inference in this work. After selecting the blur operator S and training the model R for restoration, these operators can be used in a sequential manner. By employing methods borrowed from diffusion model literature, severe blurring can be reversed.

For low values of t, a single application of R suffices to achieve the desired outcome. However, for larger t values, iterative application of R is necessary due to intensified blurring effects. Specifically, through an iterative process, each step computes $S(x_0, t)$ based on the previous state. As the goal of this work is noise generation, the number of iterations is determined experimentally. The algorithmic workflow is detailed as follows.

algorithm: Cold diffusion sampling

input: sample x_t, time series T
for s = t, t-1, ,,, 1 do
$x_0 = R(x_s, s)$
$x_{s-1} = x_s - S(x_0, s) + S(x_0, s-1)$
end for

Table 1. Inference Sampling Algorithm

For an iterative behavior R, it is exactly the same as the degenerate behavior D of standard diffusion, which can be proven as follows:

$$
\begin{aligned}
x_{s-1} &= x_s - S(R(x_s, s), s) + S(R(x_s, s), s-1) \\
&= S(x_0, s) - S(R(x_s, s), s) + S(R(x_s, s), s-1) \\
&= x_0 + s \cdot e - R(x_s, s) - s \cdot e + R(x_s, s) + (s-1) \cdot e \\
&= x_0 + (s-1) \cdot e \\
&= S(x_0, s-1)
\end{aligned} \tag{5}
$$

This iterative behavior can avoid errors caused by standard diffusion if R is not a perfect inverse process of S. Therefore, this method can enable the model to incorporate noise from other non Gaussian distributions and achieve an inverse process. Through induction, it was discovered that using the sampling algorithm from the cold diffusion model can achieve the expectations of this method. In other words, this approach can obtain high-quality noise images without the need for adversarial training. The inference process is illustrated in Fig. 4.

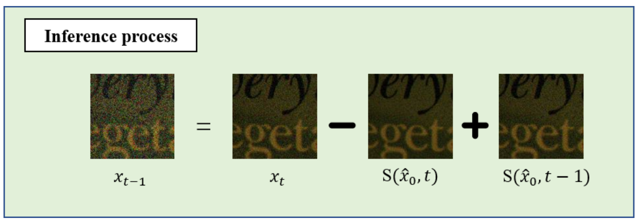

Figure 4. Inference process: The reasoning process is calculated using the above formula, and the images above correspond to the formulas in the algorithm in sequence.

Therefore, the cold diffusion model can be used to iteratively decode the noise to generate realistic noisy image. And it also benefits the forward process adding diverse noise, including shot noise, row noise, quantization noise and fixed pattern noise.

4. Experiment

In this section, the paper begins by introducing the dataset used in the study and the experimental setup, including details of the experiments. Next, the paper analyzes the proportions of various types of noise in the dataset. Then, the proposed method is compared with baselines of existing methods. The results are subsequently analyzed and discussed.

4.1. Data and experimental details

The paper utilizes a publicly available dataset from Star Light Dance for training and validation purposes. Specifically, a paired dataset containing grayscale images of clean/noisy static scenes is employed to train the proposed noise generator. Additionally, another paired dataset comprising clean/noisy images from natural scenes is used for validation. This validation dataset consists of 67 pairs of images, with each clean image accompanied by 16 noise variations. All images were captured under extremely low-light conditions, which aligns with the task requirements for noise modeling in the paper.

Our implementaion is based on PyTorch. All experiments in this paper were conducted using PyTorch 1.12.1 on a system equipped with four NVIDIA V100 32GB GPUs. The paper's model was trained using the Adam optimizer with an initial learning rate of 0.00001, and a batch size of 1. All images were in raw format. The training was performed for 700 epochs using the mean squared error (MSE) loss. During training, the learning rate was adjusted, with the lowest rate being 10^{-6}. The lowest loss was achieved at the 616th epoch of training.

To ensure experimental fairness, this study cropped the dataset into 128×128 video patches. After subtracting the clean images, the paper computed the Kullback-Leibler (KL) divergence between the synthesized noise and real noise. It is used to calculate the similarity of two probability distributions. This metric was used to assess the authenticity of the noise images generated by the method proposed in this paper.

4.2. Analysis of Noise Contribution Ratios

Due to the traditional DDPM's characteristic of increasing the hyperparameter β_t as the diffusion steps progress, the straightforward superimposition of noise during the forward diffusion process in this study is not viable. Instead, it is necessary to adapt the noise addition process to align with the progressive increase in noise characteristic of traditional DDPM. To address this, the paper analyzes the proportions of various noise types in the overall noise accumulation within low-light environments.

The methodology involves using the hyperparameters obtained from the starry dance model in the low-light environment. Starting from a clean image, each type of noise is successively added to the image. The paper computes the Kullback-Leibler Divergence (KLD) between the image after each addition and the noisy image in a low-light environment. The change in KLD values is observed, with the assumption that noise types contributing significantly to KLD values also hold a substantial proportion in the overall noise. Conversely, noise types with smaller KLD values contribute less proportionally. The KLD values calculated by sequentially adding various types of noise in low-light environments are presented in Table 2, where N_s, N_r, N_q, N_{row} and N_p approximate shot noise, read noise, quantization noise, row and fixed pattern noise.

By calculating, it was found that quantization noise contributes the most to the overall noise pattern in low-light environments and has a significant impact on the metrics. Therefore, it is placed at the end of the forward noise ad-

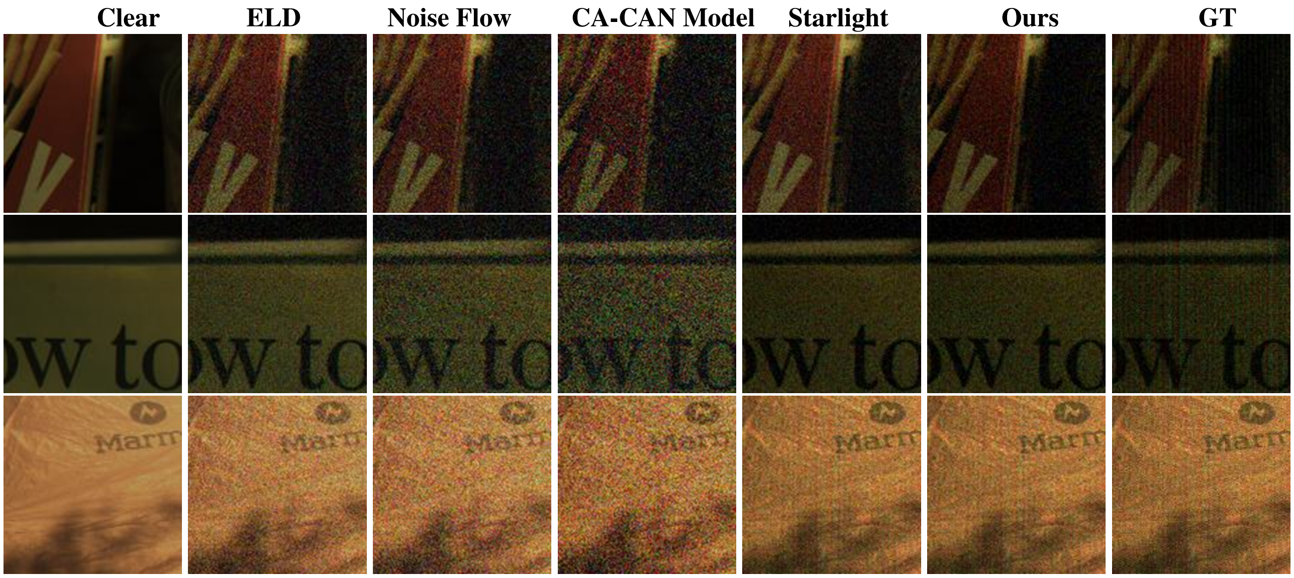

Figure 5. Method Comparison: This study compared commonly used noise modeling methods and demonstrated the noise images generated by the proposed method and baseline methods. We found that our method EL^2NM is more stable than the starlight and can generate realistic noisy image in low light environments.

Noise Type	KLD
$N_s + N_r$	0.402
$N_s + N_r + N_q$	0.125
$N_s + N_r + N_q + N_{row}$	0.117
$N_s + N_r + N_q + N_{row} + N_p$	**0.113**

Table 2. The KL divergence value is calculated by stacking various types of noise in turn

Noise Model	KLD
ELD[36]	1.360
Noise Flow[2]	0.386
CA-GAN model[7]	0.513
starlight[29]	0.069
ours	**0.068**

Table 3. Method Comparison: This study compared the proposed noise model with previous works, with each row representing a different noise modeling method. The experiments demonstrated that the proposed method is capable of generating noise images similar to the Starlight method, yielding higher quality noise images.

dition process. Granular noise, readout noise, fixed pattern noise, and quantization noise are sequentially added to the forward process in proportion. The added noise includes Poisson distribution, Gaussian distribution, and uniform distribution, covering various types of distributions that noise in low-light environments can exhibit. As a result, this approach can iteratively generate high-quality noise images for extremely low-light conditions.

4.3. Result comparison

After training the EL^2NM model, this study validated the feasibility and superiority of the proposed EL^2NM method using paired data from natural scenes in Star Light. The test dataset was cropped into 128×128 image patches. By subtracting clean images, the Kullback-Leibler (KL) divergence between synthesized noise and real noise was computed. We compared the baseline Star Light model based on GAN networks [29], the non-deep low-light noise model ELD [36], as well as two deep learning-based noise models, CA-GAN [7] and Noise Flow [2]. Both Noise Flow and CA-GAN miss the significant banding noise present in real noisy clips. ELD miss the quantizaion noise. The EL^2NM method exhibited good performance on this dataset. Qualitative performance indicators are presented in Table 3, indicating that KL divergence computed by EL^2NM was comparable to the baseline. The quantitative results are illustrated in Fig. 5, showing that EL^2NM effectively learned noise under extremely low lighting conditions, closely resembling the baseline's performance. Furthermore, the proposed method does not require adversarial training, offering better interpretability. EL^2NM achieves optimal results mainly because we decouple noise step by step through iterative loops in the reverse process, which can better approximate real noisy images. Meanwhile, due to the support of formulas, EL^2NM has interpretability.

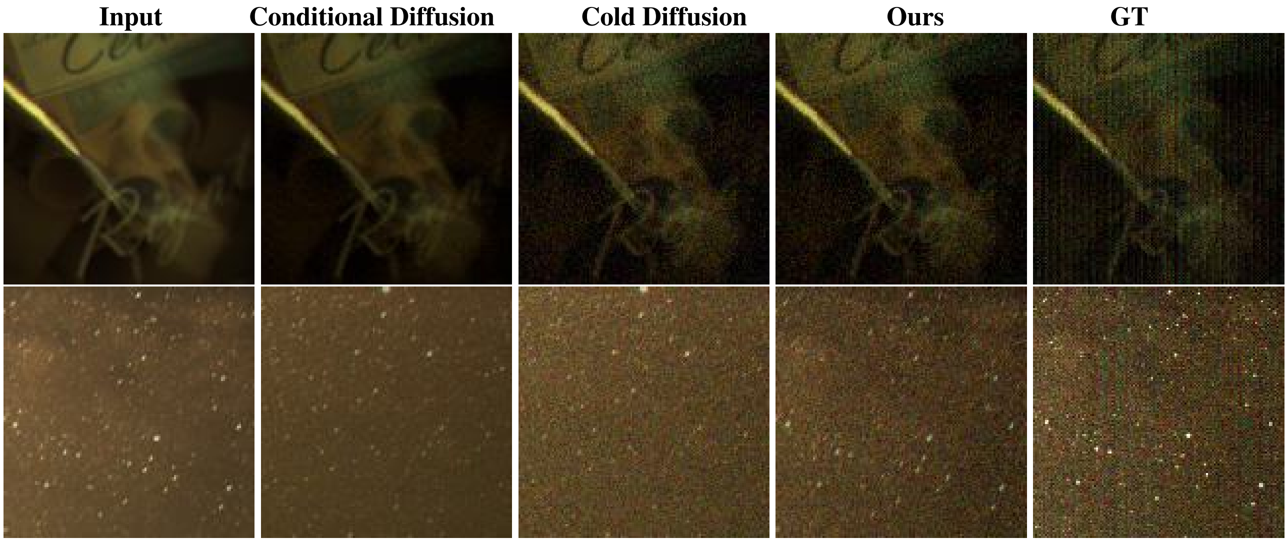

Figure 6. Ablation Experiment: We can see that when only the conditional diffusion model is used, the image does not recover noise information and some image information is lost. When there is only a cold diffusion model, the noise in image restoration is not complete enough, and the visual effect quality is not high. When combined, it can generate noise images in weak light environments more completely while ensuring the stability of the generation.

4.4. Ablation Experiment

We compared the effects of only the Conditional Diffusion and only the Cold Diffusion 6. As before, we calculate the KL divergence between synthetic and real noisy patches 4, and we find that when only using the conditional diffusion model, it is not possible to generate noise images with non Gaussian distributions due to the noise added in this paper following a non Gaussian distribution. When using only the cold diffusion model, the loss of information from the original image during model training prevents the restoration of the original image, resulting in the inability to generate noisy images. When combined, complementary advantages can be achieved, ultimately generating better noisy images.

Noise Model	KLD
only Conditional Diffusion	0.473
only Cold Diffusion	0.106
ours	**0.068**

Table 4. Ablation Comparsion: We compared the method with only conditional diffusion model and the method with only cold diffusion model with our method, and the results show that our method can better establish the noise model.

Finally, we use a combination of conditional diffusion model and cold diffusion model to establish the noise model. We use the conditional diffusion model to ensure that the structure of the image is not lost, and the cold diffusion model to add diverse noise.

5. Conclusion

We presents a novel approach that combines Conditional Diffusion Models with Cold Diffusion Models to perform noise modeling in low-light conditions, marking the first application of diffusion models in noise modeling under such environments. By employing the Conditional Diffusion Model, our work generates desired noisy images, while the Cold Diffusion Model introduces noise following various distributions. This combined approach ultimately produces high-quality noisy images. We hope that this work will inspire further developments in denoising for low-light images/videos and the advancement of diffusion models.

Our method offers opportunities for further exploration. Currently, we add noise sequentially during the forward noise addition process. However, the effects of parallel addition of noise distributions with different characteristics remain uncertain and could provide interesting outcomes. Hence, there is significant room for the diffusion model to make even greater contributions in the field of noise modeling in low-light conditions.

Acknowledgments. This work was supported by the Research on Key Technologies of Intelligent Mining Cloud Platform Science and Technology Major Special Plan of Shanxi Province (No.202101010101018), National Natural Science Foundation of China (62302466), Natural Science Basic Research Program of Shanxi Province (202303021212188, 202303021212193), Fund Program for the Scientific Activities of Selected Returned Overseas Professionals in Shanxi Province (20230017).

References

[1] Dancing under the stars: video denoising in starlight dataset. 2022. 2

[2] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3165–3173, 2019. 2, 7

[3] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. 2

[4] Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. In *Advances in Neural Information Processing Systems*, pages 41259–41282. Curran Associates, Inc., 2023. 1, 3

[5] Tamer Basar. *A New Approach to Linear Filtering and Prediction Problems*, pages 167–179. 2001. 1

[6] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 60–65 vol. 2, 2005. 2

[7] Ke-Chi Chang, Ren Wang, Hung-Jin Lin, Yu-Lun Liu, Chia-Ping Chen, Yu-Lin Chang, and Hwann-Tzong Chen. Learning camera-aware noise models. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, pages 343–358. Springer, 2020. 7

[8] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3291–3300, 2018. 1, 2

[9] Chen Chen, Qifeng Chen, Minh Do, and Vladlen Koltun. Seeing motion in the dark. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3184–3193, 2019. 1, 2

[10] 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. 1, 2

[11] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34:8780–8794, 2021. 1, 3

[12] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 2

[13] Uğur Erkan, Dang N. H. Thanh, Serdar Enginoğlu, and Samet Memiş. Improved adaptive weighted mean filter for salt-and-pepper noise removal. In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pages 1–5, 2020. 1

[14] 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

[15] Zhida Feng, Zhenyu Zhang, Xintong Yu, Yewei Fang, Lanxin Li, Xuyi Chen, Yuxiang Lu, Jiaxiang Liu, Weichong Yin, Shikun Feng, Yu Sun, Li Chen, Hao Tian, Hua Wu, and Haifeng Wang. Ernie-vilg 2.0: Improving text-to-image diffusion model with knowledge-enhanced mixture-of-denoising-experts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10135–10145, 2023. 3

[16] 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. 1, 2

[17] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising. In *Proceedings of the European conference on computer vision (ECCV)*, pages 538–554, 2018. 2

[18] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2862–2869, 2014. 2

[19] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 5769–5779, Red Hook, NY, USA, 2017. Curran Associates Inc. 1

[20] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph.*, 35(6), 2016. 1

[21] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 2

[22] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 1, 3

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

[24] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7324–7333, 2019. 1

[25] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Shadowdiffusion: Diffusion-based shadow removal using classifier-driven attention and structure preservation. *arXiv preprint arXiv:2211.08089*, 2022. 3

[26] Matteo Maggioni, Giacomo Boracchi, Alessandro Foi, and Karen Egiazarian. Video denoising, deblocking, and enhancement through separable 4-d nonlocal spatiotemporal transforms. *IEEE Transactions on Image Processing*, 21(9): 3952–3966, 2012. 1, 2

[27] Ian S. McLean and Gordon Walker. Electronic Imaging in Astronomy: Detectors and Instrumentation. *Physics Today*, 51(6):66–66, 1998. 1

[28] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2502–2510, 2018. 2

[29] Kristina Monakhova, Stephan R. Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: video denoising in starlight. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16220–16230, 2022. 1, 7

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

[31] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*, 12(11):1338–1351, 2003. 2

[32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 1

[33] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 2

[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

[35] Wei Wang, Xin Chen, Cheng Yang, Xiang Li, Xuemei Hu, and Tao Yue. Enhancing low light videos by exploring high sensitivity camera noise. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4110–4118, 2019. 1

[36] Kaixuan Wei, Ying Fu, Yinqiang Zheng, and Jiaolong Yang. Physics-based noise modeling for extreme low-light photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8520–8537, 2022. 1, 2, 3, 7

[37] Yutong Xie, Minne Yuan, Bin Dong, and Quanzheng Li. Diffusion model for generative image denoising. *arXiv preprint arXiv:2302.02398*, 2023. 3

[38] Jiachao Zhang and Keigo Hirakawa. Improved denoising via poisson mixture modeling of image sensor noise. *IEEE Transactions on Image Processing*, 26(4):1565–1578, 2017. 2

[39] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image Processing*, 26(7):3142–3155, 2017. 2

[40] 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. 2

[41] Zhaoyang Zhang, Yitong Jiang, Wenqi Shao, Xiaogang Wang, Ping Luo, Kaimo Lin, and Jinwei Gu. Real-time controllable denoising for image and video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14028–14038, 2023. 2

[42] Yuqian Zhou, Jianbo Jiao, Haibin Huang, Y. Wang, Jue Wang, Humphrey Shi, and Thomas S. Huang. When awgn-based denoiser meets real noises. In *AAAI Conference on Artificial Intelligence*, 2019. 2

[43] Zhizhuo Zhou and Shubham Tulsiani. Sparsefusion: Distilling view-conditioned diffusion for 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12588–12597, 2023. 3

[44] Fengyuan Zhu, Guangyong Chen, and Pheng Ann Heng. From noise modeling to blind image denoising. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 420–429, 2016. 2