AnlightenDiff: Anchoring Diffusion Probabilistic Model on Low Light Image Enhancement

Cheuk-Yiu Chan, *Student Member, IEEE*, Wan-Chi Siu, *Life Fellow, IEEE*, Yuk-Hee Chan, *Member, IEEE*, and H. Anthony Chan, *Life Fellow, IEEE*

***Abstract*— Low-light image enhancement aims to improve the visual quality of images captured under poor illumination. However, enhancing low-light images often introduces image artifacts, color bias, and low SNR. In this work, we propose AnlightenDiff, an anchoring diffusion model for low light image enhancement. Diffusion models can enhance the low light image to well-exposed image by iterative refinement, but require anchoring to ensure that enhanced results remain faithful to the input. We propose a Dynamical Regulated Diffusion Anchoring mechanism and Sampler to anchor the enhancement process. We also propose a Diffusion Feature Perceptual Loss tailored for diffusion based model to utilize different loss functions in image domain. AnlightenDiff demonstrates the effect of diffusion models for low-light enhancement and achieving high perceptual quality results. Our techniques show a promising future direction for applying diffusion models to image enhancement.**

***Index Terms*— Low light image enhancement, image processing, deep learning.**

I. Introduction

ADVANCEMENTS in imaging technology have made it possible for people to capture and record memorable moments in their lives with increased ease and convenience. However, one persistent challenge faced by both professional and amateur photographers alike is the degradation of image quality under low-light conditions. Images taken in such environments are often dim and noisy, making it difficult to recognize scenes or objects and compromising the overall visual appeal. In this context, low-light image enhancement has become an area of significant interest, with researchers exploring various techniques to improve visibility and suppress image artifacts while addressing the inherent challenges associated with low-light imaging.

Low-light conditions introduce a range of complexities, including presents of image artifacts, low signal-to-noise ratio (SNR), and the need to balance camera settings, such as ISO, aperture, and exposure time. While increasing ISO or exposure time can improve image brightness, these adjustments often come at the cost of amplifying image artifacts, introducing blur due to camera shake, or overexposing certain areas. Consequently, these trade-offs have motivated researchers to develop novel computational photography techniques for enhancing low-light images, encompassing illumination enhancement.

Traditional approaches to low-light image enhancement have relied on techniques such as histogram equalization [1], [2], retinex-based methods [3], [4], [5], and dehazing theory [6]. These methods aim to improve the dynamic range, separate illumination and reflectance components, or refine refraction maps to enhance the visibility of low-light images. While these approaches have demonstrated some success, they often fall short in capturing the complex interplay of local and global features present in images.

In recent years, researchers have been exploring diffusion probabilistic models [7], [8], [9], [10], which are a class of generative models that can be used for image generation and Image-to-Image synthesis. They model the process of diffusion, where noise perturbation is gradually removed from the input signal over time through a diffusion process. These models define a probability distribution over the clean signal at different points in time, with the variance of the distribution decreasing over time as the signal becomes less noisy. They are able to exploit the gradual reduction in noise perturbation to reconstruct fine details and textures.

Diffusion models have exhibited remarkable performance across various tasks, including super-resolution [11], [12], inpainting [13], [14], [15], and low-light image enhancement (LLIE) [16], [17], [18]. The success of diffusion models can be attributed to their ability to capture the intricate distributions of images and generate high-quality results, making them a promising approach for probabilistic generative modeling. Although a limited number of prior works have investigated the application of diffusion models to LLIE, there remains substantial room for improvement by incorporating domain-specific knowledge. By leveraging the power of diffusion models in conjunction with expertise in LLIE, researchers can unlock new possibilities and push the boundaries of what can be achieved in this particular task, opening up exciting avenues for further exploration in the field.

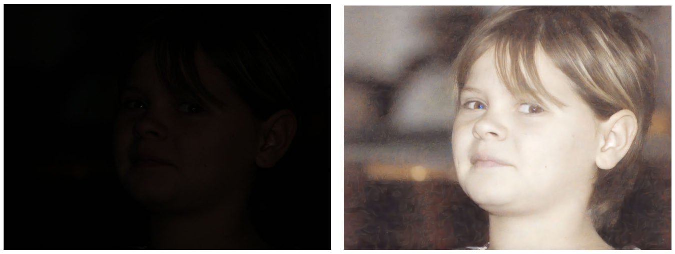
Received 5 August 2023; revised 29 May 2024 and 23 September 2024; accepted 10 October 2024. Date of publication 31 October 2024; date of current version 8 November 2024. This work was supported in part by St. Francis University under Grant ISG200206; and in part by the University Grant Committee (UGC), Hong Kong, SAR, under Grant UGC/IDS(C)11/E01/20. The associate editor coordinating the review of this article and approving it for publication was Prof. Aline Roumy. *(Corresponding author: Wan-Chi Siu.)*

Cheuk-Yiu Chan and Wan-Chi Siu are with the Department of Electrical and Electronic Engineering (EEE), The Hong Kong Polytechnic University (PolyU), Hong Kong, and also with the School of Computing and Information Sciences (SCIS), Saint Francis University (SFU), Hong Kong (e-mail: cy-allan.chan@connect.polyu.hk; enwcsiu@polyu.edu.hk).

Yuk-Hee Chan is with the Department of EEE, PolyU, Hong Kong (e-mail: enyhchan@polyu.edu.hk).

H. Anthony Chan is with SCIS, SFU, Hong Kong (e-mail: hhchan@sfu.edu.hk).

Digital Object Identifier 10.1109/TIP.2024.3486610



(a) Low Light Image

(b) Results of Ours

Fig. 1. Effect of our proposed **AnlightenDiff**. The input image suffers from underexposure and lack of contrast. Our proposed method, AnlightenDiff, is able to enhance the image and reconstruct lost details.

In this work, we propose a method for low-light image enhancement using diffusion based approach that generate remarkable enhancement results for low-light images, as shown in Fig. 1. Specifically, we propose **An**choring En**lighten**ing **Diff**usion Model (AnlightenDiff). The overview is depicted in Fig. 2. The proposed **Dynamical Regulated Diffusion Anchoring** (**DRDA**) mechanism and **Dynamical Regulated Diffusion Sampler** (**DRDS**) aim to address the limitation of existing diffusion-based generative models in incorporating domain knowledge and efficiently exploring complex target distributions. Furthermore, we propose a **Diffusion Feature Perceptual Loss** (**DFPL**) tailored for diffusion models to utilize different loss function developed in the image domain, eg. Learned Perceptual Image Patch Similarity (LPIPS) [19]. Our contributions are summarized as follows:

- We utilize a **Dynamical Regulated Diffusion Anchoring** (**DRDA**) mechanism to dynamically regulate the mean vector of the perturbations $\boldsymbol{\phi}$ to incorporate domain knowledge and match the geometry of the data distribution to explore more complex target distributions, which provide larger flexibility for diffusion-based models.
- We propose **Dynamical Regulated Diffusion Sampler** (**DRDS**), which builds upon the reverse process of diffusion models and dynamically regulates the diffusion process to explore the target distribution. This models more complex distributions compared to existing diffusion-based approaches and enables more efficient exploration of the empirical distribution and thus results in higher-quality sample generation.
- We propose the **Diffusion Feature Perceptual Loss** (**DFPL**), which is a loss function tailored for diffusion models. DFPL leverages the predicted noise perturbation to reconstruct the predicted noisy images $\boldsymbol{x}_t^\theta$ and compares them with the ground truth noisy images $\boldsymbol{x}_t$. This approach allows the use of image-based loss functions and provides image-level supervision, resulting in improved visual quality in generation.

II. Related Work

A. Low Light Image Enhancement (LLIE)

Low-Light Image Enhancement (LLIE) has been an active research area in recent years, with numerous methods proposed. Early approaches in low-light image enhancement includes LIME [20], which estimates an illumination map and applies gamma correction to recover details in dark regions, and NPE [21] which utilizes bright-pass filtering and logarithmic transformation of the illumination to maintain image naturalness while enhancing details for non-uniform illumination images. Recently, deep learning techniques have benefited various computer vision tasks, including low-light enhancement. For example, LLNet [22] simply uses an autoencoder to adaptively enhance images. Other works adopt multi-scale features to improve visual quality [23], [24], [25]. However, these early methods have limited generalization due to their reliance on heuristic illumination models.

Further research has explored the relationship between Retinex theory and deep learning techniques. Some approaches e.g. RetinexNet [5], KinD [26], KinD++ [27] and RUAS [28] employ multiple networks to implement Retinex theory, decomposing and reconstructing images. Other methods, including SCI [29] focus on the calibration of retinex enhanced image to achieve better visual effect. However, Retinex-based methods can be computationally demanding as they require multiple networks to enhance reflectance and illumination separately.

Other techniques have also been applied to the LLIE task. For instance, DRBN [30] proposed a semi-supervised learning approach combining recursive band learning with adversarial techniques. Methods like DLN [31] utilizes Back Projection (BP) [31], [32], [33], [34] to darken and enlighten features (images) repetitively for low-light image enhancement. Zero-reference approaches, such as Zero-DCE [35] estimate light-enhancement curves without reference images. Generative approaches have also been explored. These include EnlightenGAN [36], which employs Generative Adversarial Networks, and GDP [18], uses diffusion models with guided denoising. These diverse methods represent the ongoing innovation in low-light image enhancement techniques. However, these approaches often face challenges with computational efficiency and consistency across diverse conditions.

Various loss functions have been employed in LLIE task, including MSE [22], ℓ_1 loss [37], SSIM [31], smoothness loss [31], [38], and Structural dissimilarity (DSSIM) loss [37], [38]. Cai et al. [37] demonstrated that training the same network with different losses yields varied performance, highlighting the importance of conditional distribution design. As diffusion based models utilize noise predictor network to generate images indirectly, our proposed DFPL that utilizes existing loss function in image domain to train the noise predictor, is able to generate higher quality and less noisy images with fewer artifacts.

B. Diffusion Models

Diffusion models [7], [8], [9], [11], [39], [40], [41] adopt a Markov chain framework to progressively add noise perturbation to images. This process, referred as the forward diffusion process, enables a noise predictor network to learn the imposed noise distribution. Specifically, the forward diffusion process gradually injects noise perturbation into the data and can be

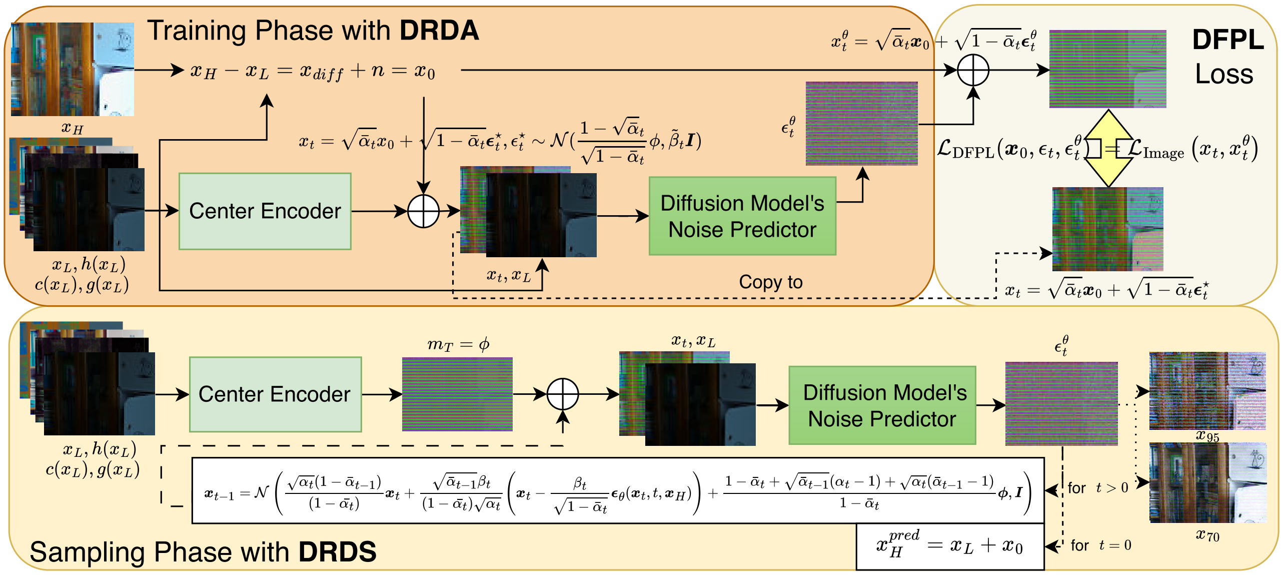

Training Phase with DRDA
x_H
$x_H - x_L = x_{diff} + n = x_0$
$x_t^\theta = \sqrt{\bar\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}_t^\theta$
$x_t = \sqrt{\bar\alpha_t}x_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}_t^\star, \epsilon_t^\star \sim \mathcal{N}(\frac{1-\sqrt{\bar\alpha_t}}{\sqrt{1-\bar\alpha_t}}\phi, \tilde\beta_t\boldsymbol{I})$
ϵ_t^θ
Center Encoder
Diffusion Model's Noise Predictor
$x_L, h(x_L)$
$c(x_L), g(x_L)$
x_t, x_L
Copy to
DFPL Loss
$\mathcal{L}_{\text{DFPL}}(\boldsymbol{x}_0, \epsilon_t, \epsilon_t^\theta) = \mathcal{L}_{\text{Image}}(x_t, x_t^\theta)$
$x_t = \sqrt{\bar\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}_t^\star$
Sampling Phase with DRDS
$x_L, h(x_L)$
$c(x_L), g(x_L)$
Center Encoder
$m_T = \phi$
x_t, x_L
Diffusion Model's Noise Predictor
ϵ_t^θ
x_{95}
x_{70}
$\boldsymbol{x}_{t-1} = \mathcal{N}\left(\frac{\sqrt{\alpha_t}(1-\bar\alpha_{t-1})}{(1-\bar\alpha_t)}\boldsymbol{x}_t + \frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{(1-\bar\alpha_t)\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar\alpha_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t, \boldsymbol{x}_H)\right) + \frac{1-\bar\alpha_t+\sqrt{\bar\alpha_{t-1}}(\alpha_t-1)+\sqrt{\alpha_t}(\bar\alpha_{t-1}-1)}{1-\bar\alpha_t}\boldsymbol{\phi}, \boldsymbol{I}\right)$
for $t > 0$
$x_H^{pred} = x_L + x_0$
for $t = 0$

Fig. 2. AnlightenDiff overview. AnlightenDiff consists of a Dynamical Regulated Diffusion Anchoring (DRDA) mechanism, Dynamical Regulated Diffusion Sampler (DRDS) and Diffusion Feature Perceptual Loss (DFPL) design. DRDA anchors the diffusion process to the target distribution with domain knowledge feature $\boldsymbol{\phi}$, which is computed by center encoder (see Fig. 3), by $\mathcal{N}(\boldsymbol{m}_t := \frac{1-\sqrt{\bar\alpha_t}}{\sqrt{1-\bar\alpha_t}}\boldsymbol{\phi}, \tilde\beta_t\boldsymbol{I})$ rather than the standard $\mathcal{N}(0, \boldsymbol{I})$ to *conditional diffusion model's noise predictor* $\boldsymbol{\epsilon}_\theta$ (see Fig. 3). Collaboratively, DRDS utilizes anchor information in reverse diffusion. In addition, DFPL tailored for diffusion models, which effectively processes perceptual features to calculate gradients for back-propagation and outperforms ℓ_1 or ℓ_2 loss.

expressed as a Markov chain:

$$q(\boldsymbol{x}_{1:T} \mid \boldsymbol{x}_0) = \prod_{t=1}^{T} q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) \tag{1}$$

$$q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) = \sqrt{\alpha_t}\boldsymbol{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}_t \tag{2}$$

where $\boldsymbol{x}_t$ denotes the data at time step t, and α_t and β_t represent the *noise perturbation schedule* such that $\alpha_t + \beta_t = 1$ and $\boldsymbol{\epsilon}_t$ is the *noise perturbation* sampled from the standard normal distribution $\mathcal{N}(0, \boldsymbol{I})$ at time t. The forward process of an arbitary t can be further simplified [8] as:

$$q(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \sqrt{\bar\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon} \tag{3}$$

where $\bar\alpha_t = \prod_{s=1}^{t}\alpha_s$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$. Thus, the learning process can be formulated as a noise perturbation prediction task. Specifically, a noise predictor network $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ is employed to learn and to estimate the conditional probability $p_\theta(\boldsymbol{x}_t|_{t-1})$, which is used in the reverse diffusion process to reconstruct the clean data $\boldsymbol{x}_0$ from $\boldsymbol{x}_T$ by minimizing a noise perturbation prediction objective:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon}} \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right\|_2^2, \text{where } t \sim \mathcal{U}(1, T) \tag{4}$$

The noise predictor network $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ takes the noisy data $\boldsymbol{x}_t$ and time step t as input, and predicts the noise perturbation $\boldsymbol{\epsilon}$ that is added to $\boldsymbol{x}_t$ according to the forward process. To invert the noise perturbation injection (forward) process and reconstruct the image, referred to as the reverse process, the following reverse equation has been proposed in [8], [40]:

$$\boldsymbol{x}_{t-1} = \mathcal{N}\left(\frac{\sqrt{\alpha_t}\,(1-\bar\alpha_{t-1})}{(1-\bar\alpha_t)}\boldsymbol{x}_t + \frac{\sqrt{\bar\alpha_{t-1}}\beta_t}{(1-\bar\alpha_t)}\boldsymbol{x}_0, \tilde\beta_t\boldsymbol{I}\right) \tag{5}$$

$$\tilde\beta_t = \frac{1-\bar\alpha_{t-1}}{1-\bar\alpha_t}\beta_t \tag{6}$$

Utilizing the forward equation Eq. (3), the predicted mean $\bar{\boldsymbol{\mu}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ is formulated to approximate the original data $\boldsymbol{x}_0$ according to:

$$\bar{\boldsymbol{\mu}}_\theta(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar\alpha_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right) \tag{7}$$

By inserting Eq. (7) s.t. $\boldsymbol{x}_0 := \bar{\boldsymbol{\mu}}_\theta(\boldsymbol{x}_t, t)$ to Eq. (5), we can obtain the final reverse equation:

$$\boldsymbol{x}_{t-1} = \mathcal{N}\left(\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{(1-\bar\alpha_t)}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\right), \tilde\beta_t\boldsymbol{I}\right) \tag{8}$$

By applying the reverse process, the diffusion model can recover the clean data $\boldsymbol{x}_0$ from the pure Gaussian noise $\boldsymbol{x}_T \sim N(0, \boldsymbol{I})$. The whole process can be optimized end-to-end with neural networks that parameterize the forward and reverse chains.

Compared to previous models that require a separate inference network [36], this learning process is more straightforward and stable [7]. As a result, diffusion models have achieved state-of-the-art results in various image generation tasks [9], [11] and generate high-quality and coherent samples without the mode collapse issue.

To learn conditional diffusion models [7], [11], [41], the *conditional information* $\boldsymbol{c}$ can be concatenated with the input for the noise prediction objective:

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{c},\boldsymbol{\epsilon}} \left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, \boldsymbol{c})\right\|_2^2, \tag{9}$$

and the reverse equation is defined as:

$$\boldsymbol{x}_{t-1} = \mathcal{N}\left(\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{(1-\bar\alpha_t)}}\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t, \boldsymbol{c})\right), \tilde\beta_t\boldsymbol{I}\right) \tag{10}$$

III. Our Proposed Approach: Anchoring Enlightening Diffusion Model (AnlightenDiff)

A. Motivation of Diffusion Model in LLIE and Residual Learning

Low-light image enhancement (LLIE) is a challenging task that aims to improve the quality and visibility of images captured under low-light conditions by enhancing their brightness, contrast, and overall visual appeal while preserving important details and minimizing artifacts. However, the inherent difficulty of LLIE lies in its one-to-many nature, as there may exist multiple well-exposed images with different configurations, such as white balance and color temperature, for a given underexposed input. This lack of a unique ground truth makes it challenging to define a clear mapping between underexposed images and their corresponding ideally exposed counterparts. To address this challenge, diffusion models have shown great potential as a promising approach for handling the one-to-many nature of LLIE, as they can generate diverse outputs by learning the underlying data distribution. By capturing the inherent variability in well-exposed images, diffusion models enable the generation of multiple enhancements for a given underexposed input, accommodating different artistic preferences and subjective perceptions of ideal exposure.

In AnlightenDiff, LLIE is formulated as a residual learning problem, where a normal light RGB image $\boldsymbol{x}_H$ is derived from a low light input image $\boldsymbol{x}_L$. Instead of directly learning a mapping, their difference is decomposed into a residual component $\boldsymbol{x}_{\text{diff}}$ and an inherent noise term $\boldsymbol{n}$ (Eq. (11)). The $\boldsymbol{n}$ represents artifacts from various sources, e.g., dark current noise and CMOS image sensor limitations. To simplify the task, the inherent noise term is considered subsumed within the initial input image $\boldsymbol{x}_0$, used in the diffusion model's forward and reverse processes.

$$\boldsymbol{x}_H - \boldsymbol{x}_L = \boldsymbol{x}_{\text{diff}} + \boldsymbol{n} = \boldsymbol{x}_0 \tag{11}$$

Residual learning plays a crucial role in enabling the model to explicitly focus on capturing the essential information needed for enhancement by learning the residual component, which represents the difference between the underexposed and well-exposed images. This targeted approach simplifies the learning problem, allows the model to more effectively capture the necessary adjustments, and reduces the risk of generating artifacts or unstable results while preserving spatial information and coherence from the underexposed input image. The combination of residual learning and diffusion models in LLIE provides a powerful framework to handle the one-to-many problem by generating diverse and high-quality outputs. By leveraging the diffusion model's capability to capture the underlying data distribution and explicitly focusing on the residual component, the proposed approach can produce diverse enhanced images that align with human perception and preferences, while preserving the spatial information and coherence of the original underexposed images, mitigating the possibility of generating multiple outputs that may not be perceptually satisfying, resulting in better performance compared to the direct learning method (Section VI-A).

B. Dynamical Regulated Diffusion Anchoring (DRDA) Mechanism

Diffusion models have recently gained much attention for their ability to learn and generate complex empirical distributions by transforming intricate data distributions into simpler parametric forms, typically $\mathcal{N}(0, \boldsymbol{I})$, through a series of Markov chain steps optimized via machine learning. However, this conventional approach often lacks the capacity to integrate domain-specific prior knowledge directly into the generative process. To overcome this limitation, we introduce Dynamical Regulated Diffusion Anchoring (DRDA), a novel guidance mechanism that enhances diffusion models by incorporating a flexible, learned mean vector $\boldsymbol{\phi}$ into the forward diffusion process. Unlike the standard Denoising Diffusion Probabilistic Model (DDPM), DRDA progressively injects noise perturbations centered around a non-zero target mean, effectively anchoring the diffusion trajectory to align with prior knowledge. This anchoring technique not only increases the model's flexibility and control over the generation process but also ensures that the final samples are more faithful to the input data's underlying structure, thereby enhancing the model's ability to produce high-quality, domain-specific outputs.

The target mean vector $\boldsymbol{\phi}$ is a critical component of the DRDA method, encoding the desired attributes or characteristics of the output samples and guiding the diffusion process. The flexibility of $\boldsymbol{\phi}$ lies in its ability to be designed as either image-dependent or image-independent. In the image-dependent setting, $\boldsymbol{\phi}$ is computed specifically for each input image based on its unique features, allowing the DRDA process to be guided by the specific content of the individual input, such as introducing a center encoder to design the specific center via backpropagation (see Section III-D). Conversely, when $\boldsymbol{\phi}$ is image-independent, it represents a learned or manually set fixed target during training and applied consistently across all input samples. Setting $\boldsymbol{\phi} = 0$ reduces the equation to the same form as DDPM [8], which is beneficial when generating samples that adhere to a particular common style, regardless of the specific input image.

Specifically, in contrast with DDPM where the additive noise perturbations are centered at the origin in the forward diffusion process, DRDA gradually steers the noise distribution mean towards a target mean $\boldsymbol{\phi}$ as the diffusion time step increases. This targets the final diffusion model sample $\boldsymbol{x}_T$ to anchor around the desired target mean $\boldsymbol{\phi}$, enhancing the flexibility and controllability of guiding the diffusion model. We demonstrate that by manipulating the target mean $\boldsymbol{\phi}$, DRDA can steer diffusion models to generate samples with desired attributes. The proposed method achieves superior performance on various benchmark datasets compared to existing baselines.

As illustrated in Fig. 2, at time step t, the *anchoring noise perturbation* $\boldsymbol{\epsilon}_t^\star$ is sampled from $\mathcal{N}(\boldsymbol{m}_t, \tilde{\beta}_t \boldsymbol{I})$, where $\boldsymbol{m}_t$ is the *dynamically regulated mean vector* which is proportional to $\boldsymbol{\phi}$. More precisely, during the forward process, $\boldsymbol{m}_t$ is adaptively adjusted at the current timestep t so that each anchoring noise perturbation $\boldsymbol{\epsilon}_t^\star \sim \mathcal{N}(\boldsymbol{m}_t, \tilde{\beta}_t I)$ and $\boldsymbol{m}_T \approx \boldsymbol{\phi}$ eventually, where T is the maximum timestep of diffusion. We thus propose the following iterative process in Eq. (12) and 13 and appendix

appendix A accordingly:

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t^\star \tag{12}$$

where $\boldsymbol{\epsilon}_t^\star \sim \mathcal{N}(\boldsymbol{m}_t, \tilde{\beta}_t \boldsymbol{I})$,

$$\boldsymbol{m}_t = \frac{1-\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\phi} \quad \text{and} \quad \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t. \tag{13}$$

The two equations described above allow the diffusion model progressively to map complex empirical distributions to simple parametric distribution with a flexible, learned mean vector that incorporates prior knowledge. As the training phase of AnlightenDiff with DRDA, as illustrated in Figure Fig. 2, employs two distinct strategies, the training from scratch and the two-step training approach will be elucidated in Section IV-B and IV-C respectively.

C. Anchoring Mechanism in AnlightenDiff and LLIE

The Dynamical Regulated Diffusion Anchoring (DRDA) mechanism in AnlightenDiff significantly enhances Low-Light Image Enhancement (LLIE) performance by imposing task-specific constraints on the diffusion process. DRDA incorporates domain knowledge through a designed mean vector $\boldsymbol{\phi}$ in the noise perturbation $\boldsymbol{\epsilon}_t^\star$, encoding pixel-level enhancement information. By introducing a new initial noise perturbation x_T that includes a color map (see Section IV-C), DRDA embeds domain-specific priors directly into the diffusion trajectory. This color information acts as a constraint that guides the generative process, ensuring that the enhanced images maintain accurate color representations and realistic lighting adjustments essential for high-quality LLIE.

Unlike other diffusion-based approaches such as RetinexDiff [16], which utilizes a dual DDPM setup to separately enhance reflectance and illumination maps, AnlightenDiff employs the DRDA mechanism to integrate color information directly into the diffusion trajectory. By embedding the color map within the diffusion process, DRDA provides more direct and efficient control over the enhancement process, ensuring that color accuracy and realistic lighting adjustments are consistently maintained throughout the generation steps. This direct incorporation of a color map as a domain-specific prior allows AnlightenDiff to produce superior performance and more realistic outcomes compared to methods that handle different aspects of image enhancement independently.

Fig. 12 demonstrates DRDA's effectiveness by comparing the initial noise perturbation x_T and the resulting enhanced image x_H^{pred} with and without anchoring. The results clearly show that DRDA achieves superior preservation of image details and color mapping, significantly improving lighting and details, while enhancement without anchoring produces less detailed results with limited color information. This comparison underscores how DRDA guides the diffusion process towards realistic enhancements by maintaining a strong connection to injected pixel-level color constraints in noise perturbation x_t.

The rationale behind DRDA's effectiveness is its integration of color maps as domain-specific priors, which guide the diffusion process to accurately adjust color balance and natural light distribution. This direct incorporation helps prevent the introduction of color artifacts and noise, while ensuring that enhancements preserve fine image details and maintain a realistic appearance. By embedding color information as a constraint via the anchored x_T, DRDA effectively imposes domain-specific priors that lead to more realistic and high-quality image enhancements.

D. Architecture of AnlightenDiff

Figure 3 illustrates the architecture of AnlightenDiff. As determining a suitable representative feature for the perturbation is challenging, we utilize a trainable center encoder network $\boldsymbol{\phi}_e$ to obtain the non-zero mean perturbation vector $\boldsymbol{\phi}$. In this work, we provide $\boldsymbol{\phi}_e$ with the low-light input image $\boldsymbol{x}_L$ and multiple illumination-invariant components, including:

- histogram equalized image $h(\boldsymbol{x}_L)$,
- channel weighted mapped image $c(\boldsymbol{x}_L)$ to normalize or weight the contribution of a specific color channel based on the overall brightness or intensity of the pixel, and
- the maximum gradient map $g(\boldsymbol{x}_L)$ that considers high frequency components in the image.

The channel weighted map $c(\boldsymbol{x}_L)$ is defined as:

$$c(x_{i,j}) = \frac{x_{i,j}}{(R_{i,j}+G_{i,j}+B_{i,j})/3} \tag{14}$$

where the variables $R_{i,j}$, $G_{i,j}$, and $B_{i,j}$ represent the red, green, and blue channel values, respectively, for the pixel at row i and column j in the image.

Similarly, the maximum gradient map is defined as:

$$g(x_{i,j}) = \max\left\{\left|\nabla_x c(x_{i,j})\right|, \left|\nabla_y c(x_{i,j})\right|\right\} \tag{15}$$

where ∇_x, and ∇_y are the image gradients in horizontal and vertical direction. Therefore, the perturbations in this work is computed by a trainable encoder network $\boldsymbol{\phi}_e$ as:

$$\boldsymbol{\phi} = \boldsymbol{\phi}_e\left(\boldsymbol{x}_L, h(\boldsymbol{x}_L), c(\boldsymbol{x}_L), g(\boldsymbol{x}_L)\right) \tag{16}$$

When selecting these components, we strike a balance between their computational efficiency and their ability to represent important aspects in LLIE. By using simple mathematical equations, we ensure that the components are easy to process and formulate, freeing up computing power for model training and making them efficient to implement within the proposed framework. During forward propagation, the input features $\left(\boldsymbol{x}_L, h(\boldsymbol{x}_L), c(\boldsymbol{x}_L), g(\boldsymbol{x}_L)\right)$ are concatenated into a 12-dimensional vector. This concatenated input is passed through a U-shaped convolutional neural network architecture for further processing.

Each 2D convolutional block consists of a 2D convolutional layer followed by a Mish activation function [42] to introduce non-linearity. The 2D convolutional layers extract salient features from the input while the residual connections facilitate efficient training of deep networks. Two such 2D convolutional blocks with a skip connection [43] constitute a residual block. Similarly, two residual blocks with a downsampling layer form a level in the U-shaped network. The downsampling layers are 2D convolutional layers with stride 2. Analogous to the U-Net [44], the U-shaped network has 3 levels. Finally, the

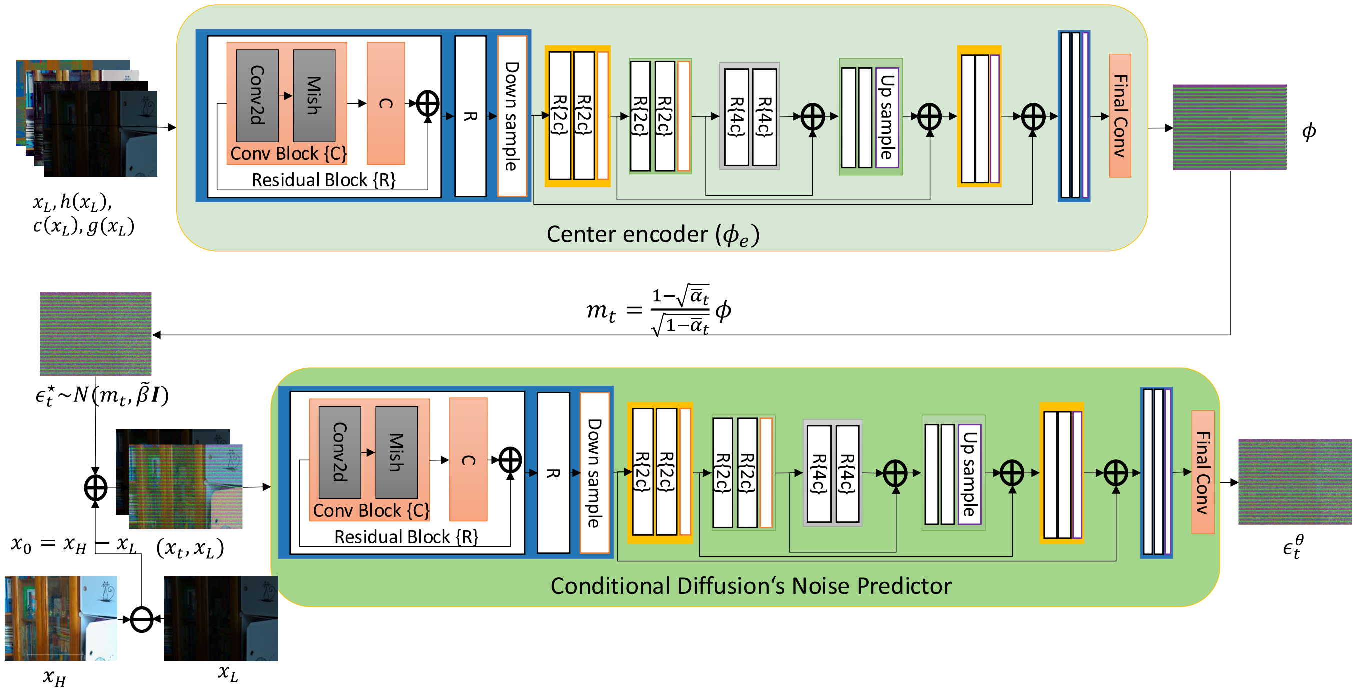


Fig. 3. The architecture of AnlightenDiff conditional diffusion noise predictor $\boldsymbol{\epsilon}_\theta$ and center encoder $\boldsymbol{\phi}_e$. The notation c, 2c and 4c after the block name means the channel size of each block w.r.t c. "Conv Block", "Res Block", "Downsample" and "Upsample" denote 2D-Convolution block, residual block, downsampling layer and upsampling layer respectively.

features are passed through a final convolutional block to generate the output $\boldsymbol{\phi}$.

The center output $\boldsymbol{\phi}$ is used to compute the dynamically regulated mean vector $\boldsymbol{m}_t$ in Eq. (13). The mean vector $\boldsymbol{m}_t$ then allows calculation of the anchoring noise perturbation $\boldsymbol{\epsilon}_t^\star$ and the input $\boldsymbol{x}_t$ using Eq. (12) and 11 respectively. The input $\boldsymbol{x}_t$ and conditional information $\boldsymbol{c} := \boldsymbol{x}_L$ are concatenated and passed through the conditional diffusion model's noise predictor. The noise predictor has a similar architecture to the center encoder described previously. It is trained to predict the anchoring noise perturbation $\boldsymbol{\epsilon}_t^\star$ added to $\boldsymbol{x}_t$, denoted as the predicted noise perturbation $\boldsymbol{\epsilon}_t^\theta$.

E. Dynamical Regulated Diffusion Sampler (DRDS)

The diffusion model builds a link between the empirical data distribution and the simpler parametric distribution by progressively adding noise perturbations at each iteration in the forward process and progressively removing noise perturbations at each iteration in the reverse process. At each iteration, the diffusion model, based on $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$, samples the previous image $\boldsymbol{x}_{t-1}$ conditioned on the current image $\boldsymbol{x}_t$.

In reverse process, the generated samples exhibit progressive improvements in quality, ultimately getting closer to the ground truth. As shown in Fig. 2, as more iterations are performed, the generated samples become progressively refined, achieving enhanced quality, thereby approaching the empirical data distribution.

Many properties of the diffusion model also apply to the proposed Dynamical Regulated Diffusion Sampler (DRDS). The DRDS introduces the non-zero mean vector $\boldsymbol{\phi}$ to effectively incorporate prior knowledge and better match the geometry of the data distribution. We thus propose the reverse process in Eq. (17) to (19) and appendix appendix B accordingly:

$$\boldsymbol{x}_{t-1} = \mathcal{N}\left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)}\boldsymbol{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{(1-\bar{\alpha}_t)}\boldsymbol{\mu}_\theta^\star(\boldsymbol{x}_t, t), \tilde{\beta}_t \boldsymbol{I}\right) \tag{17}$$

$$\boldsymbol{\mu}_\theta^\star(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right) + \tilde{\boldsymbol{\phi}} \tag{18}$$

$$\tilde{\boldsymbol{\phi}} = \frac{1-\bar{\alpha}_t + \sqrt{\bar{\alpha}_{t-1}}(\alpha_t - 1) + \sqrt{\alpha_t}(\bar{\alpha}_{t-1} - 1)}{1-\bar{\alpha}_t}\boldsymbol{\phi} \tag{19}$$

The inference phase utilizes the proposed equations to iteratively denoise the input image by incorporating prior knowledge through the non-zero mean vector $\boldsymbol{\phi}$, as illustrated in Figure 2. At each timestep, the equations are applied to progressively refine the estimate. Figure 4 depicts the intermediate denoising results obtained using the proposed DRDS. Further details on the inference procedure and the reverse diffusion process can be found in Section IV-D and Algorithm 3 respectively. Notably, setting $\boldsymbol{\phi} = 0$ reduce the equation to the same form as DDPM [8].

Compared to the original diffusion model, the DRDS has two key benefits. Domain expertise can be incorporated to inform the generative process, providing guidance for enhanced model performance. For instance, in the context of image generation, the incorporation of domain knowledge such as segmentation maps enables the synthesis of perceptually realistic samples. By leveraging information that constrains the output space to semantically and structurally coherent images, the model is able to generate higher-fidelity samples that more

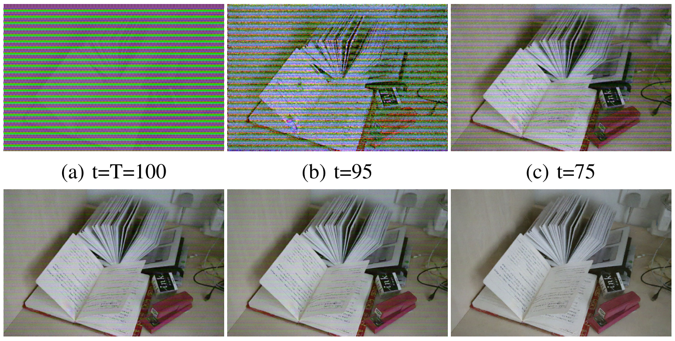

(a) t=T=100 (b) t=95 (c) t=75
(d) t=50 (e) t=30 (f) t=0

Fig. 4. Iterative denoising results for the LOL dataset image "547.png" [5] obtained using the Dynamical Regulated Diffusion Sampler (DRDS) method. The predicted outputs exhibit a gradual reduction in noise perturbation over decreasing time steps t. The final output at $t = 0$ is a denoised image.

closely adhere to the manifold of normal light images. Furthermore, dynamically regulating the reverse diffusion process enables the progressive embedding of the geometry of the data distribution. This allows for more efficient exploration of the empirical distribution and, consequently, the generation of higher quality samples compared to the original diffusion model.

F. Diffusion Feature Perceptual Loss (DFPL)

Diffusion Feature Perceptual Loss (DFPL) is a loss function tailored for diffusion models that focus on perceptual feature. Typically, for optimizing the noise predictor network, we apply ℓ_1 or ℓ_2 loss between *random sampled noise perturbation* $\boldsymbol{\epsilon}_t$ and *predicted noise perturbation* $\boldsymbol{\epsilon}_t^\theta := \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)$ from the noise predictor network, i.e. $\nabla_\theta \|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_t^\theta\|^2$.

The key innovation of DFPL lies in its transformation of the loss calculation from the noise domain to the image domain. Instead of directly comparing noise perturbations, DFPL utilizes the predicted noise perturbation $\boldsymbol{\epsilon}_t^\theta$ to reconstruct the predicted noisy image $\boldsymbol{x}_t^\theta$ in the image domain. By comparing $\boldsymbol{x}_t^\theta$ with the ground truth noisy image $\boldsymbol{x}_t$, DFPL leverages well-established perceptual image-based loss functions e.g. [19], [45], and [46], denoted as $L_{\text{Image}}(.)$, which have been developed for measuring human perception in the image domain. As shown in Fig. 2, the predicted noisy perturbation $\boldsymbol{\epsilon}_t^\theta$ from the noise predictor network is used to reconstruct $\boldsymbol{x}_t^\theta$ by applying the forward process in Eq. (3):

$$\boldsymbol{x}_t^\theta = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t^\theta \tag{20}$$

where $\boldsymbol{x}_0$ is the original image, α_t is the noise perturbation schedule and $\boldsymbol{\epsilon}_t^\theta$ is the predicted noise perturbation. The image loss is then calculated between the predicted noisy image $\boldsymbol{x}_t^\theta$ and the ground truth noisy image $\boldsymbol{x}_t$ as follows:

$$\mathcal{L}_{\text{DFPL}}(\boldsymbol{x}_0, \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t^\theta) = \mathcal{L}_{\text{Image}}\Big(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t^\theta\Big) \tag{21}$$

The primary contribution of DFPL lies in its ability to bridge the gap between the noise-based optimization of diffusion models and the common practice of comparing output images in other image-to-image models. By optimizing Diffusion Feature Perceptual Loss (DFPL), the noise predictor network learns to generate noise perturbations that accurately reconstruct noisy images, promoting the incorporation of semantically meaningful predicted noise perturbation and the generation of high-quality images. This approach provides image-level supervision for the diffusion model at each timestep, enhancing visual quality and coherence, and guiding the diffusion model's back-propagation to align with human perception. In contrast to optimizing noise prediction in isolation, the DFPL loss offers image-level supervision for the diffusion model, which consequently enhances visual quality and coherence. As demonstrated in Section VI-C, DFPL has shown promising results in the task of low-light image enhancement and potentially may be applicable in other image restoration tasks that make use of diffusion models: a point warrants for further exploration.

IV. Experiments

A. Dataset

Several publicly available datasets were employed for optimizing and assessing the model in this work. The LOL [5], VE-LOL [47], LOLv2 [48], LIME [20], NPE [21], and VV datasets were selected for this purpose. The full set of real-world images from LOL and VE-LOL datasets were utilized in the corresponding phases of the model development, while LOLv2, LIME, NPE, and VV datasets were used for testing the model's generalization ability. More details of comparison can be found in Section V.

These datasets were partitioned into training and testing subsets as per the publisher's default separation. The training images were leveraged to tune the model parameters, and the testing sets were subsequently utilized for final performance analysis. By amalgamating multiple datasets for training, the model was exposed to a more diverse and challenging range of low-light conditions, enabling it to learn more robust and generalizable features for low-light image enhancement.

B. Training From Scratch (FS)

As illustrated in Fig. 2, the training phase involved jointly optimizing the center encoder $\boldsymbol{\phi}_e$ and the diffusion model's noise predictor $\boldsymbol{\epsilon}_\theta$ with a maximum timestep of $T = 100$. The full model was trained to minimize the DPFL loss with an LPIPS loss backbone [19]. The training process is outlined in Algorithm 1.

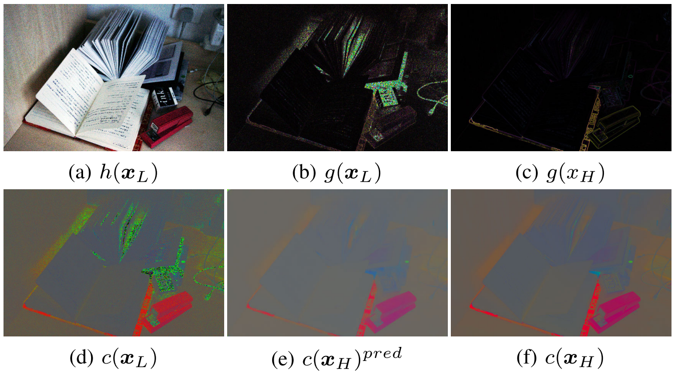
Training was performed on an NVIDIA RTX 3090 GPU system. The Lion optimizer [49] was used with a learning rate of 0.0004 and a batch size of 24 for 1000 epochs. The total training time for the full model was approximately 22 hours.

C. Two-Step (TS) Training

The encoder model $\boldsymbol{\phi}_e$ was trained separately from the diffusion model's noise predictor $\boldsymbol{\epsilon}_\theta$ with the manually designed target mean as $c(x_H)^{pred} = \boldsymbol{\phi} = \boldsymbol{\phi}_e(.)$ to enrich the color accuracy of final output. The encoder $\boldsymbol{\phi}_e$ with illumination invariant features in Fig. 5 was optimized to minimize

Algorithm 1 Training From Scratch (With Pretrained $\boldsymbol{\phi}_e$)

```
Input: Low Light (LL) image and its corresponding
  Normal Light (NL) image pairs $\boldsymbol{P} = \{\boldsymbol{x}_L^i, \boldsymbol{x}_H^i\}_{i=1}^I$,
  total diffusion step $T$
Initialize: noise predictor $\boldsymbol{\epsilon}_\theta$ with center encoder $\boldsymbol{\phi}_e$
  randomly (with pretrained center encoder $\boldsymbol{\phi}_e$);
repeat
    Sample $(\boldsymbol{x}_L, \boldsymbol{x}_H) \sim \boldsymbol{P}$; $\boldsymbol{x}_0 = \boldsymbol{x}_H - \boldsymbol{x}_L$
    Encode target mean $\boldsymbol{m}_t = \frac{1-\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\phi}_e(\boldsymbol{x}_L, \dots)$ as
      Eq. (13) and (16)
    Sample $\boldsymbol{\epsilon}_t^\star \sim \mathcal{N}(\boldsymbol{m}_t, \boldsymbol{I})$, where $t \sim \mathcal{U}(\{1, \cdots, T\})$
    Compute $\boldsymbol{\epsilon}_t^\theta = \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t, \boldsymbol{c} := \boldsymbol{x}_L)$, where
      $\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t^\star$ as Eq. (12)
    Take gradient step on
      $\mathcal{L}_{\text{LPIPS}}(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t^\star + \boldsymbol{x}_L,$
        $\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t^\theta + \boldsymbol{x}_L)$ with respect to $\boldsymbol{\epsilon}_\theta$
      and $\boldsymbol{\phi}_e$ ($\boldsymbol{\epsilon}_\theta$ only) as Section III-F
until converged;
```


(a) $h(\boldsymbol{x}_L)$ (b) $g(\boldsymbol{x}_L)$ (c) $g(\boldsymbol{x}_H)$

(d) $c(\boldsymbol{x}_L)$ (e) $c(\boldsymbol{x}_H)^{pred}$ (f) $c(\boldsymbol{x}_H)$

Fig. 5. Illustration of illumination invariant feature for image "547.png" [5]. The predicted $c(\boldsymbol{x}_H)^{pred}$ exhibit the center encoder $\boldsymbol{\phi}_e$ accurately removes the inherent high frequency components, as highlighted in $g(\boldsymbol{x}_L)$, of low-light image.

the $\ell 1$ loss between the predicted channel weighted map $c(x_H)^{pred}$ and the ground truth channel weighted map $c(x_H)$, as outlined in Algorithm 2. Subsequently, using the pretrained encoder $\boldsymbol{\phi}_e$, the diffusion model $\boldsymbol{\epsilon}_\theta$ with a maximum of 100 timesteps ($T = 100$) was trained to minimize the DPFL loss with an LPIPS backbone, as shown in Algorithm 1 with red color highlight.

Training for both models was performed on an NVIDIA RTX 3090 GPU system. The lion optimizer [49] was used with a learning rate of 0.0004 and a batch size of 32 for 1000 epochs. The training time for $\boldsymbol{\phi}_e$ and $\boldsymbol{\epsilon}_\theta$ was approximately 2 hours and 20 hours, respectively.

D. Inference

The AnlightenDiff model produces an NL image $\hat{\boldsymbol{x}}_N$ from an LL input $\boldsymbol{x}_L$ over T timesteps, as outlined in Algorithm 3.

The LL image x_L was first encoded into the target latent mean $\boldsymbol{m}_t$ by the pretrained encoder network ϕ_e, as expressed in Eq. (13) and (16). The noise predictor network $\boldsymbol{\epsilon}_\theta$ then estimated the noise perturbation $\boldsymbol{\epsilon}_t^\theta$ at each timestep t, as shown in Eq. (18).

Algorithm 2 Training of Center Encoder $\boldsymbol{\phi}_e$ in Two-Step (TS) Training

```
Input: Low Light (LL) image and its corresponding
  Normal Light (NL) image pairs $\boldsymbol{P} = \{\boldsymbol{x}_L^i, \boldsymbol{x}_N^i\}_{i=1}^I$
Initialize: center encoder $\boldsymbol{\phi}_e$ randomly
repeat
    Sample $(\boldsymbol{x}_L, \boldsymbol{x}_H) \sim \boldsymbol{P}$;
    Create illumination-invariant components $h(\boldsymbol{x}_L)$,
      $c(\boldsymbol{x}_L)$, and $g(\boldsymbol{x}_L)$ as Eq. (14) and (15)
    Compute $\boldsymbol{\phi} = \boldsymbol{\phi}_e(\boldsymbol{x}_L, h(\boldsymbol{x}_L), c(\boldsymbol{x}_L), g(\boldsymbol{x}_L))$ as
      Eq. (16)
    Take gradient step on $\mathcal{L}_{\ell 1}(\boldsymbol{\phi}, c(\boldsymbol{x}_H))$ with respect
      to $\boldsymbol{\phi}_e$
until converged;
```

Algorithm 3 Inference

```
Input: LL image $\boldsymbol{x}_L$, total diffusion step $T$
Load: pretrained noise predictor $\boldsymbol{\epsilon}_\theta$ and center encoder
  $\boldsymbol{\phi}_e$;
Calculate $\boldsymbol{m}_T = \boldsymbol{\phi} = \boldsymbol{\phi}_e(\boldsymbol{x}_L, \dots)$ as Eq. (16)
Sample $\boldsymbol{x}_T \sim N(\boldsymbol{m}_T, \sigma^2\boldsymbol{I})$, where $\sigma^2 = 1$
for t = T, ..., 1 do
    $\boldsymbol{\mu}_\theta^\star(\boldsymbol{x}_t, t, \boldsymbol{c} := \boldsymbol{x}_L) =$
      $\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t, \boldsymbol{x}_L)\right) +$
      $\frac{1-\bar{\alpha}_t+\sqrt{\bar{\alpha}_{t-1}}(\alpha_t-1)+\sqrt{\alpha_t}(\bar{\alpha}_{t-1}-1)}{1-\bar{\alpha}_t}\boldsymbol{\phi}$ as Eq. (18)
    $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ if $t > 1$, else $\boldsymbol{z} = \boldsymbol{0}$
    $\boldsymbol{x}_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)}\boldsymbol{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{(1-\bar{\alpha}_t)}\boldsymbol{\mu}^\star(\boldsymbol{x}_t) + \sigma_t\boldsymbol{z}$
      using Eq. (17)
end
return $\boldsymbol{x}_H^{pred} = \boldsymbol{x}_0 + \boldsymbol{x}_L$ as the enhanced image
```

The predicted noise perturbation $\boldsymbol{\epsilon}_t^\theta$ was applied to the reverse process expressed in Eq. (17) and repeated for T timesteps to obtain the predicted $\boldsymbol{x}_0$. Finally, the predicted $\boldsymbol{x}_0$ was added to the input $\boldsymbol{x}_L$ to produce the predicted NL image $\boldsymbol{x}_H^{pred}$.

V. Results

A. Quantitative Results

The proposed generative low-light image enhancement method is thoroughly evaluated on multiple datasets using both full-reference (FR) metrics, including PSNR, SSIM [50], and LPIPS [19], which assess the quality of the enhanced images by comparing them with their corresponding ground truth references, and non-reference (NR) metrics, including HyperIQA [52], NIMA [53], and TReS [54], which evaluate the perceptual quality of the enhanced images for datasets where normal-light reference images are unavailable. As presented in Table I and II, our method consistently achieves state-of-the-art performance, particularly in perceptually-driven metrics including SSIM, LPIPS, HyperIQA, and NIMA, which better align with human visual quality perception. These metrics are calculated using well-established packages: scikit-image [55] for PSNR and SSIM, IQA-Pytorch [56] for HyperIQA, NIMA,

TABLE I

Quantitative Full-Reference Comparison on LOL [5], VELOL [47] and LOLv2 [48] Datasets in Terms of PSNR, SSIM [50], and LPIPS [19]. ↑ (↓) Denotes That, Larger (Smaller) Values Lead to Better Quality.
(RED: BEST; BLUE: THE 2nd BEST, **PURPLE**: THE 3rd BEST)

Dataset		LOL [5]			VE-LOL [47] / LOLv2 [48] (Real)		
Type	Method	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Non-generative	NPE [21]	16.97	0.484	0.400	17.333	0.464	0.396
	LIME [20]	17.546	0.531	0.387	17.483	0.505	0.428
	RUAS [28]	11.309	0.435	0.377	13.975	0.469	0.329
	SCI [29]	14.784	0.525	0.333	17.304	0.540	0.307
	Zero-DCE [35]	14.861	0.562	0.330	18.059	0.580	0.308
	SGZ [51]	14.546	0.438	0.353	16.992	0.359	0.338
	KinD [26]	17.648	0.771	0.174	**20.588**	**0.818**	**0.143**
	KinD++ [27]	17.752	0.758	0.198	17.660	0.761	0.218
	RetinexNet [5]	17.559	0.645	0.381	17.676	0.642	0.441
	DRBN [30]	16.777	0.730	0.345	18.466	0.768	0.352
	DLN [31]	**21.946**	**0.807**	**0.148**	17.878	0.693	0.300
Generative	EnlightenGAN [36]	17.606	0.653	0.372	18.676	0.678	0.364
	GDP [18]	15.821	0.541	0.338	14.412	0.497	0.363
	Ours (FS)	**19.661**	**0.747**	**0.176**	**20.45**	**0.772**	**0.16**
	Ours	**21.726**	**0.814**	**0.141**	**20.657**	**0.837**	**0.146**

TABLE II

Quantitative Non-Reference Comparison on LIME [20], NPE [21] and VV Datasets in Terms of HyperIQA [52], NIMA [53] and TReS [54]. ↑ (↓) Denotes That, Larger (Smaller) Values Lead to Better Quality.(**BOLD** Represents the BEST)

Dataset	LIME [5]			NPE [47]			VV		
Method	HyperIQA↑	NIMA↑	TReS↑	HyperIQA↑	NIMA↑	TReS↑	HyperIQA↑	NIMA↑	TReS↑
LIME [20]	0.5549	4.548	71.76	0.5639	4.725	79.128	0.3286	4.113	43.688
RUAS [28]	0.521	4.623	64.977	0.5219	4.301	72.632	0.3373	3.798	35.2
SCI [29]	0.5263	4.55	65.273	0.5463	4.379	77.639	0.3422	3.895	40.433
Zero-DCE [35]	0.5538	4.587	72.138	0.5718	4.698	80.072	0.3242	3.758	**44.745**
SGZ [51]	0.5461	4.647	71.894	0.5719	4.726	81.465	0.3211	3.774	41.581
RetinexNet [5]	0.5054	4.049	72.928	0.5196	3.894	80.551	-	-	-
DRBN [30]	0.5071	4.231	63.21	0.5319	4.397	73.611	0.3117	3.694	37.514
DLN [31]	0.544	4.684	74.126	0.5465	4.728	78.63	-	-	-
EnlightenGAN [36]	0.5331	4.661	70.372	0.5595	4.667	79.552	0.2924	3.865	34.293
Ours	**0.5832**	**4.686**	**77.17**	**0.5974**	**4.824**	**81.944**	**0.3486**	**4.147**	44.677

and TReS, and TorchMetrics [57] for the LPIPS, ensuring a fair and standardized comparison with state-of-the-art approaches. Notably, we compare our method with other generative low-light image enhancement models, including EnlightenGAN [36] which applies a Generative Adversarial Network (GAN) architecture, and GDP [18] which employs a diffusion model. Results of the comparison show that our approach is superior.

For the FR evaluation, we compare model performances on the LOL [5], VE-LOL (Real) [47], and LOLv2 (Real) [48] datasets (Table I), where VE-LOL and LOLv2 share the same testing dataset. Our method consistently achieves state-of-the-art performance across all datasets, surpassing both traditional and deep learning-based approaches. On the LOL dataset, our two-step training approach yields the best results in SSIM and LPIPS among all models, while maintaining highly competitive performance in PSNR, closely following the top-performing DLN [31]. Although the PSNR results of our method are slightly lower compared to one or two other approaches, the difference is expected as PSNR depends strongly on luminance changes for which perception can vary subjectively between individuals. SSIM and LPIPS are more perceptually-driven metrics better reflecting visual quality perception. Our superior SSIM and LPIPS demonstrate compelling enhanced images with preserved details.

When compared to other generative models, our method significantly outperforms both EnlightenGAN [36], which employs a GAN-based architecture, and GDP [18], another diffusion-based model, by a substantial margin in all metrics. Similarly, on the VE-LOL/LOLv2 (Real) dataset, our two-step training approach demonstrates superior performance across almost all metrics, achieving the highest PSNR and

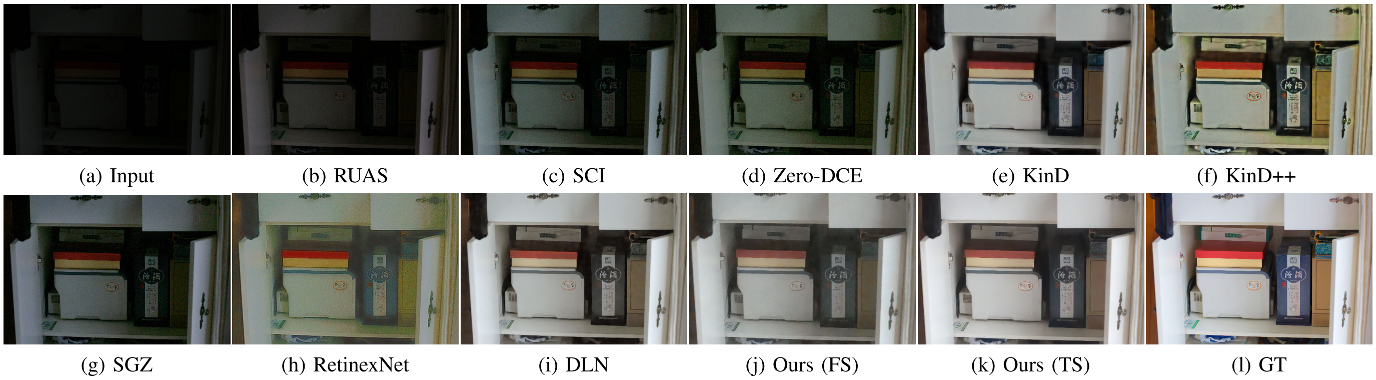


Fig. 6. Visual comparison of 55.png on LOL dataset [5], where FS and TS stand for "from scratch" and "two step" respectively.

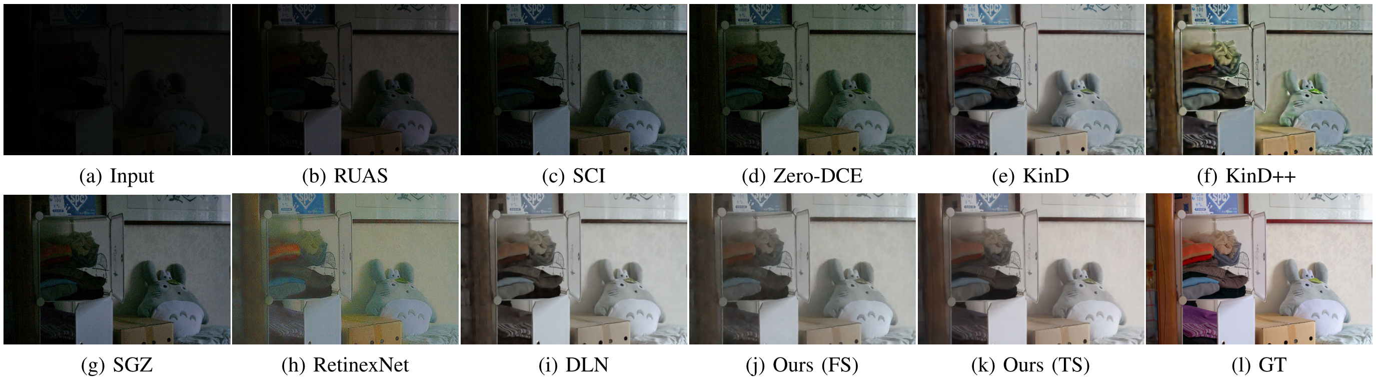


Fig. 7. Visual comparison of 23.png on LOL dataset [5], where FS and TS stand for "from scratch" and "two step" respectively.

SSIM among all models, and a very competitive LPIPS score slightly behind KinD [26]. Compared to EnlightenGAN and GDP, our method showcases a significant improvement in all metrics, further validating the effectiveness of our diffusion-based approach in enhancing low-light images across different datasets. Moreover, even our from-scratch model surpasses both EnlightenGAN and GDP by a considerable margin, highlighting the robustness and generalizability of our method. These results demonstrate the state-of-the-art performance of our diffusion-based generative model in low-light image enhancement, showcasing its superiority over existing approaches, including both non-generative and generative models. The substantial improvements over other generative models, particularly GDP, which is also a diffusion-based model, underscore the effectiveness of our proposed work.

For the NR evaluation, we also make use of the most challenging datasets, including LIME [5], NPE [47], and VV (Table II). These datasets only provide low-light images without their normal-light counterparts, making it impossible to train a model directly on them. As a result, the evaluation on these datasets is inherently zero-shot, requiring the use of pre-trained models without any further fine-tuning. As shown in Table II, our approach consistently achieves the best results across all datasets, outperforming both traditional and deep learning-based approaches, as well as other generative models such as EnlightenGAN or GDP. Our model attains the highest scores in all three NR metrics (HyperIQA, NIMA, and TReS) on the LIME and NPE datasets, and the best performance in HyperIQA and NIMA on the VV dataset, while remaining competitive in TReS. These results highlight our model's ability to generate visually appealing enhanced images with better perceptual quality, aesthetics, and overall image quality in this challenging zero-shot setting, validating its strong generalization capability and effectiveness in producing high-quality enhanced images that align with human perception and aesthetic preferences.

B. Qualitative Results

This section presents a visual comparison of various low-light image enhancement methods on the LOL and VE-LOL/LOLv2 (Real) datasets. As observed in Fig. 6 to 11, our proposed method, AnlightenDiff, significantly enhances the brightness and details of the input low-light images while maintaining a natural appearance and preserving the original color scheme. In contrast, other methods suffer from various issues, such as insufficient brightness enhancement, loss of details, or unnatural color shifts.

Among the compared methods, KinD [26] and DLN [31] produce relatively better results, but they still introduce some color distortions and fail to restore some details. EnlightenGAN [36], a generative adversarial network-based method, improves the brightness but generates unnatural artifacts and color deviations. GDP [18], another diffusion-based generative model, enhances the overall brightness but introduces an unnatural yellowish tint and fails to restore fine details. Other methods, such as RUAS [28], SCI [29], Zero-DCE [35], RetinexNet [5], and SGZ [51], also exhibit various limitations in their enhanced results.

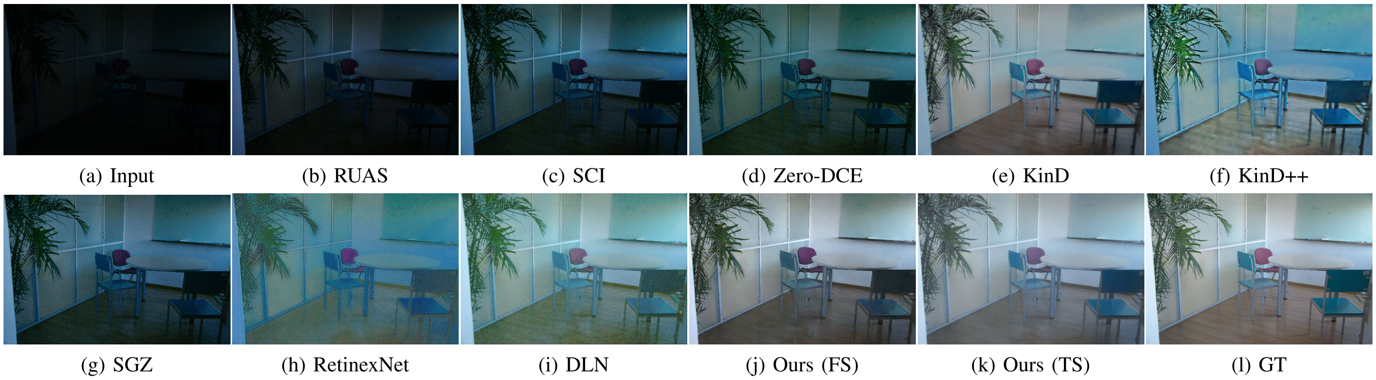


Fig. 8. Visual comparison of low00702.png on VE-LOL/LOLv2 (Real) dataset [47], [48], where FS and TS stand for "from scratch" and "two step" respectively.

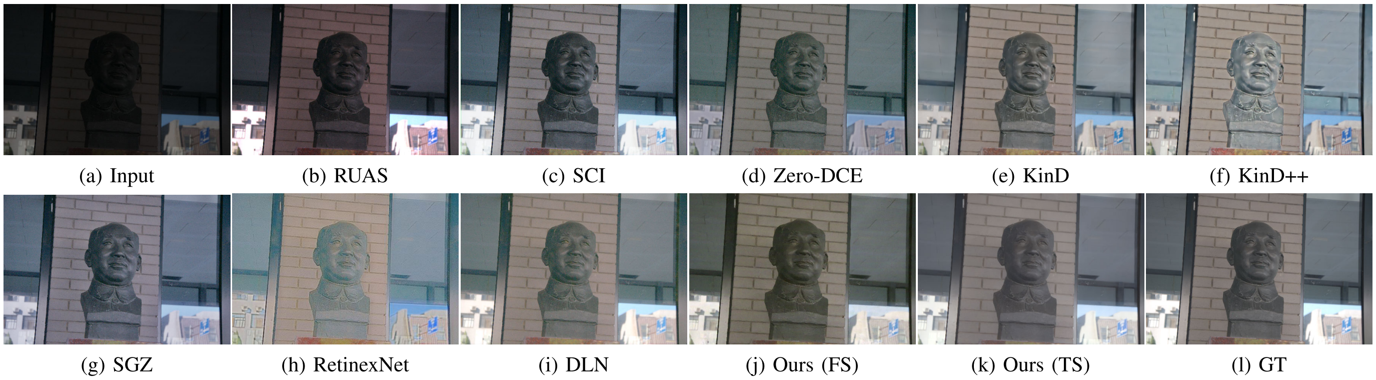


Fig. 9. Visual comparison of low00716.png on VE-LOL/LOLv2 (Real) dataset [47], [48], where FS and TS stand for "from scratch" and "two step" respectively.

The superior performance of our AnlightenDiff method can be attributed to its ability to preserve fine details and textures, maintain color accuracy, provide balanced brightness enhancement, and effectively reduce image artifacts. These advantages stem from the key technical contributions of our method. The DRDA (Section III-B) and DRDS (Section III-E) anchor the diffusion process to the incorporated prior feature of LLIE as the center, altering the way of noise perturbation sampling and contributing to a more complex domain mapping between the low-light and normal-light domains. Additionally, the DFPL (Section III-F), a tailored loss function that combines the ideas of human perception, image-based loss functions, and time-step wise diffusion loss, guides the diffusion process to generate high-quality images with well-preserved details and natural appearance by providing an explicit connection between noise perturbation in the noise domain and image-based perceptual loss in the image domain. These technical contributions work together to enable the generation of more realistic and visually appealing results, outperforming both traditional and deep learning-based approaches, as well as other generative models. The qualitative results across multiple datasets and image examples (Fig. 6 to 11) demonstrate the superiority of our AnlightenDiff method in enhancing low-light images while preserving their natural appearance and details.

VI. Analysis of Network Structure

To rigorously validate the effectiveness of each component in our proposed model, we have performed ablation studies on the LOL dataset [5]. Specifically, we conducted control experiments by removing one component at a time from the full model to examine its impact. Furthermore, to isolate the efficacy of the diffusion module itself, the two-step training procedure with a pretrained central encoder $\boldsymbol{\phi}_e$ as described in Section IV-C was employed in this ablation study.

TABLE III
Comparison Between Direct and Residual Learning

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Direct Learning	20.662	0.795	0.172
Residual Learning	**21.726**	**0.814**	**0.141**

A. Effect of Residual Learning

Our proposed AnlightenDiff model utilizes residual learning for low-light image enhancement, where the residual is defined in Section III-A. To validate the effectiveness of this residual learning approach, we compare against a baseline model that directly estimates $\boldsymbol{x}_0 = \boldsymbol{x}_H^{pred}$. As shown in Table III, our residual learning model outperforms the direct learning baseline across all three evaluation metrics. This demonstrates that modeling the enhancement residual is more effective for low-light image enhancement compared to directly estimating the normal-light image. The key advantage of residual learning is that the model only needing to estimate the enhancement residual. In contrast, the direct learning approach has to completely reconstruct the normal-light image, which is more

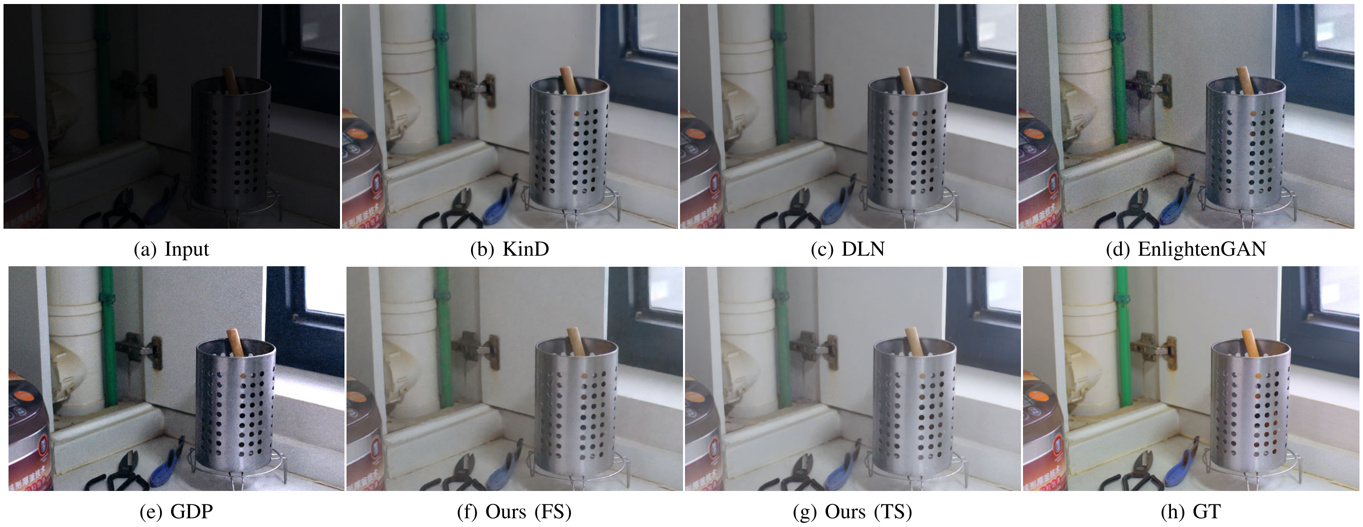

(a) Input (b) KinD (c) DLN (d) EnlightenGAN
(e) GDP (f) Ours (FS) (g) Ours (TS) (h) GT

Fig. 10. Enlarged Visual comparison of 111.png on LOL dataset [5], where FS and TS stand for “from scratch” and “two step” respectively.

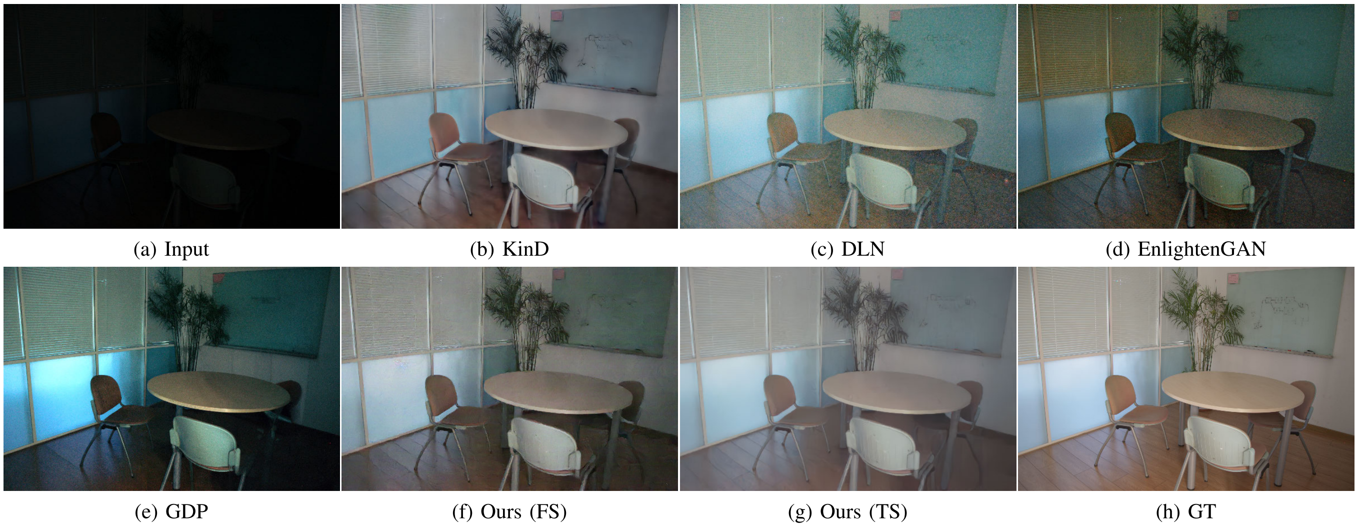

(a) Input (b) KinD (c) DLN (d) EnlightenGAN
(e) GDP (f) Ours (FS) (g) Ours (TS) (h) GT

Fig. 11. Enlarged Visual comparison of low00706.png on VE-LOL/LOLv2 (Real) dataset [47], [48], where FS and TS stand for “from scratch” and “two step” respectively.

difficult to optimize. As a result, optimizing the residual is much easier than the original direct learning problem, allowing our residual learning approach to achieve superior performance.

B. Effect of Dynamical Regulated Diffusion Anchoring (DRDA) and Sampler (DRDS)

To validate the effectiveness of the diffusion modules (DRDA and DRDS) in our proposed model, we conducted an ablation study by removing each diffusion module separately and jointly.

Dynamical Regulated Diffusion Anchoring (DRDA): The model without DRDA (denoted as “Ours w/o DRDA” in Table IV) achieves a PSNR of 8.143 dB, SSIM of 0.289, and LPIPS of 0.609. By incorporating the proposed DRDA module (denoted as “Ours”), the model achieves significant performance gains, improving PSNR to 21.726 dB (an increase of 13.583 dB), SSIM to 0.814 (an increase of 0.525), and reducing LPIPS to 0.141 (a decrease of 0.468).

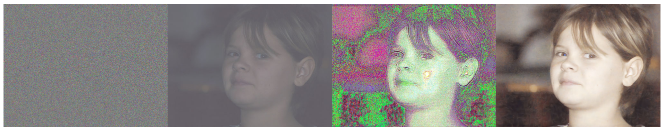
Dynamical Regulated Diffusion Sampler (DRDS): The model without DRDS (denoted as “Ours w/o DRDS” in Table IV) achieves a PSNR of 13.145 dB, SSIM of 0.411, and LPIPS of 0.434. By incorporating the DRDS module (denoted as “Ours”), the model gains significant improvements, with PSNR increasing to 21.726 dB (an improvement of 8.581 dB), SSIM increasing to 0.814 (an increase of 0.403), and LPIPS decreasing to 0.141 (a decrease of 0.293).

Joint Effect: When the model is trained without the DRDA and DRDS modules, it applies the forward and reverse diffusion processes of DDPM [8] without the support of the center feature. The absence of these modules (denoted as “Ours w/o DRDA & DRDS” in Table IV) results in a PSNR of 16.602 dB, an SSIM of 0.726, and an LPIPS of 0.254. In comparison with Fig. 12, the full proposed model achieves substantial performance improvements, with the PSNR increasing to 21.726 dB (a gain of 5.124 dB), the SSIM increasing to 0.814 (an improvement of 0.088), and the LPIPS decreasing to 0.141 (a reduction of 0.113).

TABLE IV

ABLATION STUDY FOR DRDA AND DRDS

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Ours w/o DRDA	8.143	0.289	0.609
Ours w/o DRDS	13.145	0.411	0.434
Ours w/o DRDA & DRDS	16.602	0.726	0.254
Ours	**21.726**	**0.814**	**0.141**



(a) $x_T \sim \mathcal{N}(0, I)$ (b) x_H^{pred} (c) $x_T \sim \mathcal{N}(\phi, I)$ (d) x_H^{pred}

Ours w/o DRDA & DRDS Ours

Fig. 12. Comparison between our method without DRDA & DRDS and our proposed method with anchoring. (a) and (c) show x_T, the initial noise perturbation. (b) and (d) show x_H^{pred}, the enhanced image. With anchoring (DRDA) via x_T, our proposed method (right) demonstrates superior preservation of image details and color mapping, achieving significant improvement in lighting and detail. In contrast, the enhanced image without anchoring (left) produces a less detailed result with limited color information, tending towards a white filter effect.

TABLE V

ABLATION STUDY FOR DFPL

Method	PSNR ↑	SSIM ↑	LPIPS ↓
ℓ_1 Loss	19.161	0.677	0.409
ℓ_2 Loss	18.55	0.685	0.385
DFPL (Ours)	**21.726**	**0.814**	**0.141**

These ablation studies clearly demonstrate the synergistic effects of the DRDA and DRDS modules, both individually and jointly. The proposed full model achieved a significant performance gains over the model without these modules, affirming that the DRDA and DRDS modules have complementary advantages for denoising that are enhanced when used together.

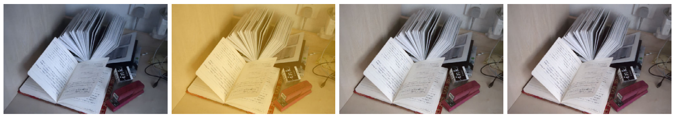
C. Effectiveness of Diffusion Feature Perceptual Loss (DFPL)

We have evaluated the effectiveness of our proposed diffusion feature perceptual loss (DFPL) by comparing against two common losses: ℓ_1 and ℓ_2. As shown in Table V, models trained with either ℓ_1 or ℓ_2 loss obtain inferior performance compared to our model trained with DFPL loss. Specifically, the DFPL loss leads to improvements of 2.565 dB and 3.176 dB in PSNR, 0.139 and 0.131 in SSIM and 0.268 and 0.244 in LPIPS over the ℓ_1 and ℓ_2 respectively.

The considerable improvements validate the efficacy of DFPL for enhancing perceptual quality and global consistency of reconstructed images. DFPL effectively preserves the image structural similarity and perceptual information, thus achieving superior performance compared to the baselines.

D. Effect on Illumination Invariant Feature on Center Encoder

An ablation study was conducted to evaluate the impact of illumination invariant features on the center encoder $\boldsymbol{\phi}_e$ by individually removing the histogram equalized image $h(x_L)$, channel weighted mapped image $c(x_L)$, and maximum gradient map $g(x_L)$. As shown in Fig. 13 and Table VI, removing any of these components leads to a noticeable degradation in the enhanced image quality and a decrease in PSNR, SSIM, and LPIPS scores. The absence of $h(x_L)$ results in a loss of contrast and brightness balance, removing $c(x_L)$ causes color distortions and an unnatural appearance, and the lack of $g(x_L)$ leads to a loss of fine details and textures. These findings emphasize the importance of each illumination invariant feature in enabling the center encoder to extract robust center, which is invariant to changes in illumination, resulting in high-quality enhanced images with well-preserved details, natural colors, and balanced brightness.



(a) $\boldsymbol{\phi}_e$ w/o $h(x_L)$ (b) $\boldsymbol{\phi}_e$ w/o $c(x_L)$ (c) $\boldsymbol{\phi}_e$ w/o $g(x_L)$ (d) *Ours*

Fig. 13. Illustration of the impact of removing illumination invariant features from the center encoder $\boldsymbol{\phi}_e$ for image "547.png" [5]. (a) $\boldsymbol{\phi}_e$ without the histogram equalized feature $h(x_L)$, (b) $\boldsymbol{\phi}_e$ without the channel weighted mapped feature $c(x_L)$, (c) $\boldsymbol{\phi}_e$ without the maximum gradient map $g(x_L)$, and (d) our complete model. The center $\phi = \phi_e(.)$ demonstrates the importance of each illumination invariant feature in preserving image details and maintaining natural appearance.

TABLE VI

ABLATION STUDY FOR CENTER ENCODER $\boldsymbol{\phi}_e$

Method	PSNR ↑	SSIM ↑	LPIPS ↓
ϕ_e w/o $h(x_L)$	19.375	0.806	0.183
ϕ_e w/o $c(x_L)$	14.59	0.747	0.327
ϕ_e w/o $g(x_L)$	21.343	0.809	0.152
Ours	**21.726**	**0.814**	**0.141**

VII. CONCLUSION

In conclusion, AnlightenDiff leverages Dynamical Regulated Diffusion Anchoring and Sampling to incorporate prior knowledge and to match the data distribution. The proposed Diffusion Feature Perceptual Loss further improves perceptual quality. Experimental results demonstrate state-of-the-art performance on perceptual metrics, producing enhanced images aligning with human perception. AnlightenDiff shows the potential of anchoring diffusion models for low light enhancement through high perceptual quality results matching human perception. This provides a promising direction for applying diffusion models to image enhancement. Future work will explore anchoring for other tasks like super resolution. Code is available at https://github.com/allanchan339/AnlightenDiff.

APPENDIX A
DERIVATION OF THE DRDA

Given $\boldsymbol{x}_0$ and a mean vector $\boldsymbol{\phi}$, inductively we define two sequences

$$\boldsymbol{x}_t = \sqrt{\alpha_t}\boldsymbol{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_t; \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, I) \tag{A.1}$$

$$\boldsymbol{\mu}_t \triangleq \frac{1-\sqrt{\alpha_t}}{\sqrt{1-\alpha_t}}\boldsymbol{\phi} \tag{A.2}$$

and via solving Eq. (A.1) we obtain a closed form

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sum_{j=1}^{t}\sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_j}}\sqrt{1-\alpha_j}\boldsymbol{\epsilon}_j \tag{A.3}$$

where $\boldsymbol{\epsilon}_j \sim \mathcal{N}(\boldsymbol{\mu}_j, I)$ is a random perturbation. Taking expectation conditional on $\boldsymbol{x}_0$, we have

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{x}_t \mid \boldsymbol{x}_0\right] &= \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sum_{j=1}^{t}\sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_j}}\sqrt{1-\alpha_j}\cdot\frac{1-\sqrt{\alpha_j}}{\sqrt{1-\alpha_j}}\boldsymbol{\phi}\\ &= \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sum_{j=1}^{t}\left(\sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_j}} - \sqrt{\frac{\alpha_t}{\alpha_{j-1}}}\right)\boldsymbol{\phi}\\ &= \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \left(1-\sqrt{\bar{\alpha}_t}\right)\boldsymbol{\phi} \to \boldsymbol{\phi} \text{ as } t \to +\infty\end{aligned}$$

Moreover, by the law of total variance, we have

$$\operatorname{Var}\left(\boldsymbol{x}_t \mid \boldsymbol{x}_0\right) = \sum_{j=1}^{t}\left(\frac{\bar{\alpha}_t}{\bar{\alpha}_j} - \frac{\bar{\alpha}_t}{\bar{\alpha}_{j-1}}\right)I = \left(1-\bar{\alpha}_t\right)I \tag{A.4}$$

Let us denote $\boldsymbol{m}_t := \frac{1-\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\phi}$ and define a sequence of random perturbation by

$$\boldsymbol{\epsilon}_t^\star := \frac{\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0}{\sqrt{1-\bar{\alpha}_t}} \tag{A.5}$$

From the above, we can see that $\boldsymbol{\epsilon}_t^\star$ is normally distributed where

$$\mathbb{E}[\boldsymbol{\epsilon}_t^\star] = \mathbb{E}\left[\frac{\mathbb{E}[\boldsymbol{x}_t \mid \boldsymbol{x}_0] - \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0}{\sqrt{1-\bar{\alpha}_t}}\right] = \mathbb{E}\left[\frac{1-\sqrt{\bar{\alpha}_t}}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\phi}\right] = \boldsymbol{m}_t$$

and

$$\operatorname{Var}\left(\boldsymbol{\epsilon}_t^\star\right) = \frac{1}{1-\bar{\alpha}_t}\operatorname{Var}\left(\boldsymbol{x}_t \mid \boldsymbol{x}_0\right) = I.$$

That is to say, this means $\boldsymbol{\epsilon}_t^\star \sim \mathcal{N}\left(\boldsymbol{m}_t, I\right)$

Appendix B
Derivation of the DRDS

Now let us discuss the reverse process of our proposed AnlightenDiff, for which we call DRDS. According to Bayes Theorem, the conditional distribution of $\boldsymbol{x}_{t-1}$ given $\boldsymbol{x}_t$ and $\boldsymbol{x}_0$ is given by

$$p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0\right) = \frac{p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{x}_0\right)p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_0\right)}{p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_0\right)}$$

Since $p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{x}_0\right)$ and $p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_0\right)$ are both density functions of Gaussian distributions, $\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0$ is also normally distributed. Thus, we can let $\boldsymbol{\mu}_t^\star := \boldsymbol{\mu}_t^\star(\boldsymbol{x}_t, \boldsymbol{x}_0)$ and $\tilde{\beta}_t := \tilde{\beta}_t(\boldsymbol{x}_t, \boldsymbol{x}_0)$ be functions such that $\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0 \sim \mathcal{N}\left(\boldsymbol{\mu}_t^\star\left(\boldsymbol{x}_t, \boldsymbol{x}_0\right), \tilde{\beta}_t\left(\boldsymbol{x}_t, \boldsymbol{x}_0\right)I\right)$.

From $p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{x}_0\right)p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_0\right)$, we consider

$$\begin{aligned}&\tfrac{1}{2(1-\alpha_t)}\left\|\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1} - \left(1-\sqrt{\alpha_t}\right)\boldsymbol{\phi}\right\|^2\\ &\quad + \tfrac{1}{2(1-\bar{\alpha}_{t-1})}\left\|\boldsymbol{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0 - \left(1-\sqrt{\bar{\alpha}_{t-1}}\right)\boldsymbol{\phi}\right\|^2\\ &= \tfrac{\alpha_t(1-\bar{\alpha}_{t-1})+(1-\alpha_t)}{2(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\left\|\boldsymbol{x}_{t-1}\right\|^2\\ &\quad - \left\langle\tfrac{\sqrt{\alpha_t}\boldsymbol{x}_t - \sqrt{\alpha_t}(1-\sqrt{\alpha_t})\boldsymbol{\phi}}{1-\alpha_t} + \tfrac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0 + \left(1-\sqrt{\bar{\alpha}_{t-1}}\right)\boldsymbol{\phi}}{1-\bar{\alpha}_{t-1}}, \boldsymbol{x}_{t-1}\right\rangle\\ &\quad + \text{const.}\end{aligned}$$

Since $p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{x}_0\right) \propto p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \boldsymbol{x}_0\right)p\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_0\right)$, we compare the above equation with

$$\frac{1}{2\tilde{\beta}_t}\left\|\boldsymbol{x}_{t-1} - \boldsymbol{\mu}_t^\star\right\|^2 = \frac{1}{2\tilde{\beta}_t}\left\|\boldsymbol{x}_{t-1}\right\|^2 - \left\langle\frac{\boldsymbol{\mu}_t^\star}{\tilde{\beta}_t}, \boldsymbol{x}_{t-1}\right\rangle + \text{const.}$$

Therefore, we obtain

$$\begin{aligned}\tilde{\beta}_t\left(\boldsymbol{x}_t, \boldsymbol{x}_0\right) &= \tfrac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{\alpha_t(1-\bar{\alpha}_{t-1})+(1-\alpha_t)}\\ &= \tfrac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{\alpha_t-\bar{\alpha}_t+1-\alpha_t} = \tfrac{(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\beta_t\end{aligned}$$

as in Eq. (6), and

$$\begin{aligned}&\boldsymbol{\mu}_t^\star\left(\boldsymbol{x}_t, \boldsymbol{x}_0\right)\\ &= \left(\tfrac{\sqrt{\alpha_t}\boldsymbol{x}_t - \sqrt{\alpha_t}(1-\sqrt{\alpha_t})\boldsymbol{\phi}}{1-\alpha_t} + \tfrac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0 + \left(1-\sqrt{\bar{\alpha}_{t-1}}\right)\boldsymbol{\phi}}{1-\bar{\alpha}_{t-1}}\right)\tilde{\beta}_t\\ &= \tfrac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 + \tfrac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t\\ &\quad + \tfrac{\left(1-\sqrt{\bar{\alpha}_{t-1}}\right)(1-\alpha_t) - \sqrt{\alpha_t}(1-\sqrt{\alpha_t})(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{\phi}\\ &= \tfrac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\boldsymbol{x}_0 + \tfrac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_t\\ &\quad + \tfrac{1-\bar{\alpha}_t+\sqrt{\bar{\alpha}_{t-1}}(\alpha_t-1)+\sqrt{\alpha_t}(\bar{\alpha}_{t-1}-1)}{1-\bar{\alpha}_t}\boldsymbol{\phi}\end{aligned} \tag{B.1}$$

By letting $\boldsymbol{x}_0 := \bar{\boldsymbol{\mu}}_\theta\left(\boldsymbol{x}_t, t\right)$ in Eq. (7), we have

$$\begin{aligned}&\boldsymbol{\mu}_t^\star\left(\boldsymbol{x}_t, \bar{\boldsymbol{\mu}}_\theta\left(\boldsymbol{x}_t, t\right)\right)\\ &= \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right)\\ &\quad + \frac{1-\bar{\alpha}_t+\sqrt{\bar{\alpha}_{t-1}}(\alpha_t-1)+\sqrt{\alpha_t}(\bar{\alpha}_{t-1}-1)}{1-\bar{\alpha}_t}\boldsymbol{\phi}\end{aligned}$$

as in Eq. (18) and (19).

References

[1] K. Singh, R. Kapoor, and S. K. Sinha, "Enhancement of low exposure images via recursive histogram equalization algorithms," *Optik*, vol. 126, no. 20, pp. 2619–2625, Oct. 2015.

[2] Q. Wang and R. Ward, "Fast image/video contrast enhancement based on weighted thresholded histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 757–764, May 2007.

[3] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Org. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, Jan. 1971, doi: 10.1364/JOSA.61.000001.

[4] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proc. 3rd IEEE Int. Conf. Image Process.*, vol. 3, Sep. 1996, pp. 1003–1006.

[5] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," presented at the Brit. Mach. Vis. Conf., 2018.

[6] Q. Tang, J. Yang, X. He, W. Jia, Q. Zhang, and H. Liu, "Nighttime image dehazing based on Retinex and dark channel prior using Taylor series expansion," *Comput. Vis. Image Understand.*, vol. 202, Jan. 2021, Art. no. 103086, doi: 10.1016/j.cviu.2020.103086.

[7] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 8780–8794.

[8] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[9] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8162–8171.

[10] C.-Y. Chan, W.-C. Siu, Y.-H. Chan, and H. A. Chan, "Generative strategy for low and normal light image pairs with enhanced statistical fidelity," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, vol. 33, Jan. 2024, pp. 1–3, doi: 10.1109/ICCE59016.2024.10444437.

[11] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10674–10685.

[12] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4713–4726, Apr. 2023, doi: 10.1109/TPAMI.2022.3204461.

[13] C.-C. Hui, W.-C. Siu, N.-F. Law, and H. A. Chan, "Intelligent painter: New masking strategy and self-referencing with resampling," in *Proc. 24th Int. Conf. Digit. Signal Process. (DSP)*, Jun. 2023, pp. 1–5, doi: 10.1109/DSP58604.2023.10167925.

[14] B. Xia et al., "DiffIR: Efficient diffusion model for image restoration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 13049–13059, doi: 10.1109/ICCV51070.2023.01204.

[15] Y. Zhu et al., "Denoising diffusion models for plug-and-play image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 1219–1229, doi: 10.1109/cvprw59228.2023.00129.

[16] X. Yi, H. Xu, H. Zhang, L. Tang, and J. Ma, "Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12268–12277, doi: 10.1109/ICCV51070.2023.01130.

[17] J. Hou, Z. Zhu, J. Hou, H. Liu, H. Zeng, and H. Yuan, "Global structure-aware diffusion process for low-light image enhancement," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds. Red Hook, NY, USA: Curran Associates, pp. 79734–79747.

[18] B. Fei et al., "Generative diffusion prior for unified image restoration and enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 9935–9946.

[19] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

[20] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017, doi: 10.1109/TIP.2016.2639450.

[21] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013, doi: 10.1109/TIP.2013.2261309.

[22] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017, doi: 10.1016/j.patcog.2016.06.008.

[23] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNNs," in *Proc. Brit. Mach. Vis. Conf.*, 2018, p. 4.

[24] W. Ren et al., "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019, doi: 10.1109/TIP.2019.2910412.

[25] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4, doi: 10.1109/VCIP.2017.8305143.

[26] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 1632–1640, doi: 10.1145/3343031.3350926.

[27] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1013–1037, Apr. 2021, doi: 10.1007/s11263-020-01407-x.

[28] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.

[29] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5637–5646.

[30] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3060–3069.

[31] L.-W. Wang, Z.-S. Liu, W.-C. Siu, and D. P. K. Lun, "Lightening network for low-light image enhancement," *IEEE Trans. Image Process.*, vol. 29, pp. 7984–7996, 2020.

[32] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1664–1673.

[33] Z.-S. Liu, L.-W. Wang, C.-T. Li, and W.-C. Siu, "Hierarchical back projection network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2041–2050.

[34] Z.-S. Liu, L.-W. Wang, C.-T. Li, W.-C. Siu, and Y.-L. Chan, "Image super-resolution via attention based back projection networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3517–3525, doi: 10.1109/ICCVW.2019.00436.

[35] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. CVPR*, Jun. 2020, pp. 1777–1786.

[36] Y. Jiang et al., "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[37] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018, doi: 10.1109/TIP.2018.2794218.

[38] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6842–6850, doi: 10.1109/CVPR.2019.00701.

[39] H. Li et al., "SRDiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, Mar. 2022.

[40] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," 2020, *arXiv:2010.02502*.

[41] J. Ho and T. Salimans, "Classifier-free diffusion guidance," in *Proc. NeurIPS Workshop Deep Generative Models Downstream Appl.*, Dec. 2021, pp. 1–8.

[42] D. Misra, "Mish: A self regularized non-monotonic activation function," in *Proc. 31st Brit. Mach. Vis. Conf.*, 2020, pp. 1–14.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham, Switzerland: Springer, pp. 234–241.

[45] S. Lao et al., "Attentions help CNNs see better: Attention-based hybrid image quality assessment network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1139–1148, doi: 10.1109/CVPRW56347.2022.00123.

[46] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, "Image quality assessment: Unifying structure and texture similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2567–2581, May 2022, doi: 10.1109/TPAMI.2020.3045810.

[47] J. Liu, D. Xu, W. Yang, M. Fan, and H. Huang, "Benchmarking low-light image enhancement and beyond," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1153–1184, Apr. 2021, doi: 10.1007/s11263-020-01418-8.

[48] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 2072–2086, 2021, doi: 10.1109/TIP.2021.3050850.

[49] X. Chen et al., "Symbolic discovery of optimization algorithms," 2023, *arXiv:2302.06675*.

[50] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Nov. 2003, pp. 1398–1402.

[51] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 581–590.

[52] S. Su et al., "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3664–3673.

[53] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3998–4011, Aug. 2018, doi: 10.1109/TIP.2018.2831899.

[54] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 3209–3218.

[55] S. van der Walt et al., "Scikit-image: Image processing in Python," *PeerJ*, vol. 2, Jun. 2014, Art. no. e453.

[56] C. Chen and J. Mo, "IQA-PyTorch: PyTorch toolbox for image quality assessment," 2022. [Online]. Available: https://github.com/chaofengc/IQA-PyTorch

[57] N. Detlefsen et al., "TorchMetrics—Measuring reproducibility in PyTorch," *J. Open Source Softw.*, vol. 7, no. 70, p. 4101, Feb. 2022, doi: 10.21105/joss.04101.

Cheuk-Yiu Chan (Student Member, IEEE) received the B.Eng. degree (Hons.) in electronic and information engineering from The Hong Kong Polytechnic University in 2021, where he is currently pursuing the M.Phil. degree in electrical and electronic engineering (EEE). Concurrently, he is a Research Assistant with the School of Computing and Information Sciences, Saint Francis University, Hong Kong. His research interests include computer vision, deep learning, and image/video enhancement.

Wan-Chi Siu (Life Fellow, IEEE) received the M.Phil. degree from The Chinese University of Hong Kong in 1977 and the Ph.D. degree from Imperial College London in 1984. He is currently an Emeritus Professor (formerly a Chair Professor, the HoD of EIE, and the Dean of Engineering Faculty) with The Hong Kong Polytechnic University and a Research Professor of Saint Francis University, Hong Kong. He has been a keynote speaker and an invited speaker of many conferences. He has published over 500 research papers (200 appeared in international journals, such as IEEE TRANSACTIONS ON IMAGE PROCESSING) in DSP, transforms, fast algorithms, machine learning, deep learning, super-resolution imaging, 2D/3D video coding, and object recognition and tracking. He is an outstanding scholar with many awards, including the Distinguished Presenter Award, the Best Teacher Award, the Best Faculty Researcher Award (twice), and the IEEE Third Millennium Medal in 2000. He is the Vice President, the Chair of Conference Board, and a Core Member of Board of Governors of the IEEE SP Society (2012–2014); and the President of APSIPA (2017-2018). He has organized IEEE Society-sponsored flagship conferences as the TPC Chair (ISCAS1997) and the General Chair (ICASSP2003 and ICIP2010). He was an independent non-executive Director (2000–2015) of a publicly-listed video surveillance company and chaired the First Engineering/IT Panel of the RAE(1992/93) in Hong Kong. Recently, he has been a member of the IEEE Educational Activities Board, the IEEE Fourier Award for Signal Processing Committee (2017–2020), the Hong Kong RGC Engineering-JRS Panel (2020–2026), Hong Kong ASTRI Tech Review Panel (2006–2024) and some other IEEE technical committees. He has been a Guest Editor/a Subject Editor/an AE of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Electronics Letters*. He was an APSIPA Distinguished Lecturer (2021–2022) and an Advisor and a Distinguished Scientist of the European Research Project SmartEN (offered by European Commissions).

Yuk-Hee Chan (Member, IEEE) received the B.Sc. degree (Hons.) in electronics from The Chinese University of Hong Kong in 1987 and the Ph.D. degree in signal processing from The Hong Kong Polytechnic University in 1992. From 1987 to 1989, he was an Research and Development Engineer with Elec & Eltek Group, Hong Kong. He joined The Hong Kong Polytechnic University in 1992, where he is currently an Associate Professor with the Department of Electrical and Electronic Engineering. He has published over 165 research papers in various international journals and conferences. His research interests include image processing and deep learning. He was the Chair of the IEEE Hong Kong Section in 2015. He is the Treasurer of Asia–Pacific Signal and Information Processing Association (APSIPA) Headquarters.

H. Anthony Chan (Life Fellow, IEEE) received the B.Sc. degree from The University of Hong Kong, the M.Phil. degree from The Chinese University of Hong Kong, and the Ph.D. degree in physics from University of Maryland. He is currently the Dean of Yam Pak Charitable Foundation, School of Computing and Information, Saint Francis University. He conducted industry research with former AT&T Bell Labs, where he was the Lead AT&T Delegate at 3GPP network standards. He was a Professor with the University of Cape Town, and then joined Huawei Technologies, USA, to conduct standards and research in 5G Wireless and IETF standards. He has authored/co-authored 30 USA and international patents, over 260 journal/conference papers, and a book and five book chapters; and edited/authored/contributed to four network standards documents at IEEE and IETF. He has presented over 20 keynotes/invited talks and 40 conference tutorials. He has been a Distinguished Speaker of IEEE ComSoc, IEEE CMPT Society, and IEEE Reliability Society.