Efficient Diffusion as Low Light Enhancer

Guanzhou Lan[1,3*] Qianli Ma[2*] Yuqi Yang[1]

Zhigang Wang[1,3] Dong Wang[3] Xuelong Li[1,4] Bin Zhao[1,3†]

[1]Northwestern Polytechnical University [2]Shanghai Jiao Tong University [3]Shanghai AI Lab [4]TeleAI

Abstract

*The computational burden of the iterative sampling process remains a major challenge in diffusion-based Low-Light Image Enhancement (LLIE). Current acceleration methods, whether training-based or training-free, often lead to significant performance degradation, highlighting the trade-off between performance and efficiency. In this paper, we identify two primary factors contributing to performance degradation: fitting errors and the inference gap. Our key insight is that fitting errors can be mitigated by linearly extrapolating the incorrect score functions, while the inference gap can be reduced by shifting the Gaussian flow to a reflectance-aware residual space. Based on the above insights, we design Reflectance-Aware Trajectory Refinement (RATR) module, a simple yet effective module to refine the teacher trajectory using the reflectance component of images. Following this, we introduce **Re**flectance-aware **D**iffusion with **Di**stilled **T**rajectory (**ReDDiT**), an efficient and flexible distillation framework tailored for LLIE. Our framework achieves comparable performance to previous diffusion-based methods with redundant steps in just 2 steps while establishing new state-of-the-art (SOTA) results with 8 or 4 steps. Comprehensive experimental evaluations on 10 benchmark datasets validate the effectiveness of our method, consistently outperforming existing SOTA methods. Code is available at project page.*

1. Introduction

Low-light Image Enhancement (LLIE) is designed to improve the visibility and contrast of images captured in low-light conditions while maintaining natural-looking details, which is crucial for various downstream applications. Diffusion models [9, 22] have achieved remarkable success in this domain, demonstrating significant progress in generating photorealistic normal light images [10, 40, 41]. Previous diffusion-based LLIE methods [10, 40, 41] primarily

---

*Equal Contribution.
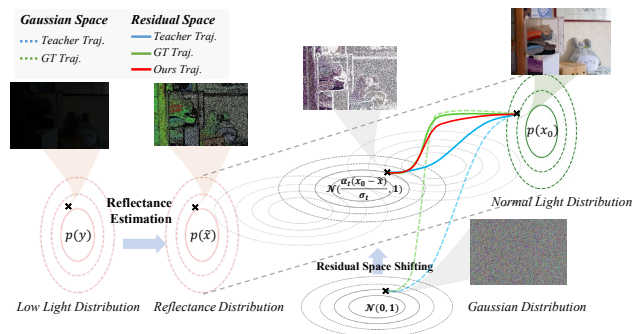
†Corresponding author.

Figure 1. ReDDiT shifts the teacher trajectory from the original Gaussian distribution to a residual space, effectively reducing the sampling gap. Subsequently, it refines the teacher trajectory towards the ground truth trajectory to mitigate fitting errors.

focus on how to condition the low-light image within the diffusion framework. Diff-Retinex [40] applies the Retinex decomposition to the low-light image and condition these components into the diffusion model. PyDiff [48] utilizes pyramid low-light images as conditions and applies DDIM [25] to expedite the sampling stage. In WCDM [11], the diffusion model operates in the wavelet-transformed space, restoring the high-frequency parts of low-light images. GSAD [10] proposes a global structural regularization to enhance structural information learning, enabling smaller noise schedules for the inference stage.

However, the primary challenge to the broader application of diffusion models in LLIE stems from their iterative denoising mechanism. For example, DDPM [9] requires multiple denoising steps, *e.g.*, 1000 steps, to reverse a Gaussian noise into a clean image. The computation overhead conflicts with the demands of LLIE, particularly in computational photography applications for edge devices like mobile phones and surveillance cameras.

Inspired by training-aware acceleration methods for diffusion models [24, 26], we investigate distilling a multi-step diffusion-based LLIE model to improve efficiency. A key observation is that performance degradation is inevitable as the number of sampling steps is reduced, even when using advanced acceleration methods (*e.g.*, consistency distillation [26]). This leads us to pose the question: *Is it possible*

to distill a student diffusion model that surpasses the original diffusion model? If so, even as performance degrades with fewer sampling steps, we could still achieve results comparable to those of the multi-step teacher model.

To address this, we conduct a comprehensive analysis of diffusion model acceleration techniques. Through theoretical analysis, we identify two primary factors contributing to performance degradation: **(1) The fitting error.** It is the unavoidable errors between deep learning models and target fitting data. It will result the additional undesired terms and mismatch during distillation. **(2) The inference gap.** It is the gap between the training target and sampling strategies specialized for the Diffusion models. It is caused by the universal diffusion model often trained and operated on the Gaussian flow for the diversity of generation, while the LLIE requires for more deterministic.

Our key insight is to refine the teacher trajectory by applying linear extrapolation to the score functions of the teacher model, thereby mitigating the adverse effects of fitting errors. Meanwhile, shifting the sampling trajectory to a deterministic space addresses the sampling gap, as shown in Figure 1. A detailed theoretical analysis of this approach is presented in Sec. 3.

Based on the above principles, we design Reflectance-Aware Trajectory Refinement (RATR) module, a simple yet effective module to refine the teacher trajectory for LLIE task. It incorporates the reflectance component of images as a deterministic prior to adjust and refine diffusion trajectories. Following this, we introduce **Re**flectance-aware **Di**ffusion with **Di**stilled **T**rajectory (**ReDDiT**), an efficient and flexible distillation framework tailored for LLIE. This framework perform trajectory matching distillation is conducted between the refined teacher and student models, yields a 2-step diffusion model with comparable performance to previous 10-step diffusion-based methods, as well as 4 and 8-step diffusion models that achieve new SOTA results. our contributions are summarized as follows:

(1) We theoretically analyze the factors contributing to performance degradation and propose targeted improvements: linear extrapolation of the score function to mitigate performance loss due to fitting errors, and residual shifting as a solution to the sampling gap.

(2) Based on these two design principles, we present ReDDiT, a novel distillation scheme that enhances the efficiency of generative diffusion models for LLIE. Notably, ReDDiT achieves high-quality image restoration in just 2 steps, establishing a new benchmark for efficient diffusion-based models in this domain.

(3) Extensive experiments conducted on 10 benchmark datasets validate that ReDDiT consistently achieves SOTA results, showcasing its superiority in terms of both quality and efficiency, even with a minimal number of steps.

2. Related Work

Low-Light Image Enhancement. Enhancing low-light images is a classical task in low-level vision, with numerous solutions leveraging deep neural networks [7, 12, 36]. Recently, as the diffusion model has demonstrated promising results in image generation tasks, there has been growing interest in leveraging diffusion models for LLIE to achieve faithful restoration results. Diff-Retinex applies Retinex decomposition as the condition of diffusion model [40]. Py-Diff utilizes pyramid low-light images as the condition and applies DDIM to speed up the sampling stage [25, 48]. In WCDM, the diffusion model is constructed in the wavelet-transformed space and restores the high-frequency parts of low-light images [11]. GSAD proposes a global structural regularization to enhance structural information learning, which also helps to train diffusion models with less curvature, enabling smaller noise schedules for the inference stage [10]. Unfortunately, these works only consider how to condition the low-light images, neglecting the efficiency concerns.

Diffusion Model Acceleration. Diffusion models have shown promising performance across various tasks such as style transfer [4], video generation [5], and text-to-image generation [16, 22, 23]. However, the issue of redundant sampling steps remains a significant efficiency concern. Recent research has explored various strategies to accelerate diffusion models. One series of methods involves the use of fast post-hoc samplers [17, 18, 25], which reduces the number of inference steps for pre-trained diffusion models to 20-50 steps. However, most suffer from severe performance degradation when further accelerating sampling within 10 steps. To address this limitation, step distillation [19, 24] is proposed, aiming to distill diffusion models into fewer steps. Progressive distillation (PD) [24] is the first successful practice and produces 2-step unconditional diffusion. Following PD [24], Meng et al. apply PD to the large-scale Stable Diffusion, achieving text-to-image generation with 2 steps. Consistency Distillation (CD) [26] aims to learn consistency among diffusion timesteps of the teacher models [26]. Following this line of research, CTM [13] extends such consistency from individual timesteps to an entire trajectory along the diffusion model, enabling faithful unconditional generation. Despite the booming development of diffusion distillation in other fields, techniques for LLIE are still left blank. In this paper, we present the first distilled diffusion tailored for LLIE, which will be introduced in detail in the following sections.

3. Methods

Let x_0 denote a high-quality image, the forward diffusion process aims to generate a sequence of noisy latent variables $x_1, x_2, ..., x_T$ using a Markovian process, which is defined
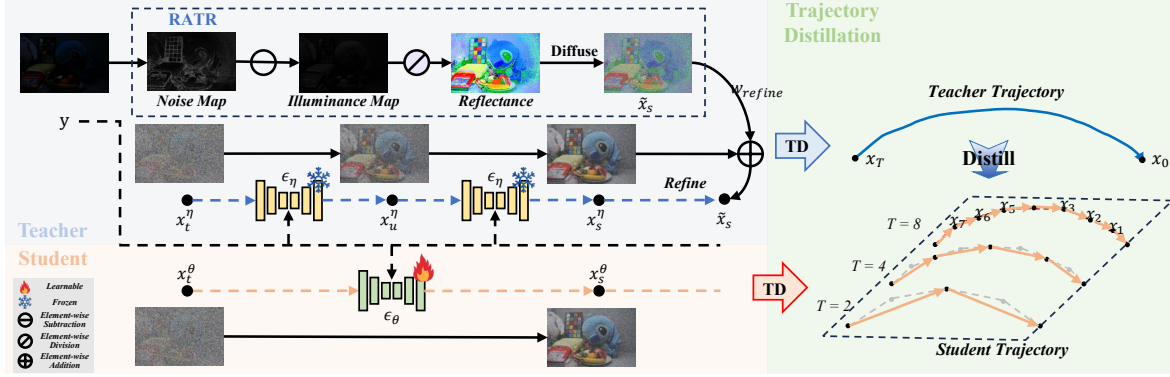
Figure 2. Pipeline of our proposed ReDDiT. The distillation process involves two parts: teacher model leverages the estimated reflectance to refine its trajectory and student model's trajectory is guided by teacher's trajectory, via a distillation loss. **TD** denotes the Trajectory Decoder while **RATR** denotes the Reflectance-Aware Trajectory Refinement.

by the equation:

$$x_t = \alpha_t x_0 + \sigma_t \epsilon, \tag{1}$$

where $\alpha_t \in (0, 1)$ represents the noise schedule, σ_t denotes the covariance at t, and ϵ represents the Gaussian noise. In the reverse process, we incorporate the low-light image y as the conditioning of the score functions $\epsilon_\eta(x_t, y, \alpha_t)$ and predict the clean images iteratively:

$$x_{t-1}^\eta = \alpha_{t-1} \left(\frac{x_t - \sigma_t \epsilon_\eta}{\alpha_t} \right) + \sigma_{t-1} \epsilon_\eta. \tag{2}$$

In this section, we propose ReDDiT, aiming to train a student diffusion model with parameters θ to learn the trajectory of the teacher model η in fewer steps. We begin by introducing trajectory distillation. Next, we theoretically analyze the factors contributing to performance degradation and present a concise formulation for refining the trajectory. Then, we determine that the reflectance component is most effective for refining the teacher trajectory and design a Reflectance-Aware Trajectory Refinement (RATR) module. Finally, all components are integrated into the ReDDiT framework, where distillation is performed.

3.1. Trajectory Distillation

The essence of diffusion distillation differs from traditional knowledge distillation is that distilled diffusion models learn the trajectory from the teacher, preserving the iterative sampling characteristics unique to diffusion models. We define $G(x_t, y, t, s)$ as the trajectory decoder that transition from time step t to s ($s < t$) to represent the diffusion model trajectory. Draw inspiration from the denoising diffusion implicit model, we represent the decoder with the student score function ϵ_θ as:

$$G_\theta(x_t, y, t, s) = \frac{\alpha_s}{\alpha_t} x_t + (\sigma_s - \frac{\alpha_s}{\alpha_t} \sigma_t) \epsilon_\theta(x_t, y, t). \tag{3}$$

With this trajectory decoder, the teacher model's trajectory can be decoded from time step s to t. For a more precise

estimation of the entire teacher trajectory, we also utilize the intermediate step $u \in [s, t]$ to estimate the trajectory, which formulates the trajectory decoder in the second order as:

$$
\begin{aligned}
x_{s,u,t}^\eta &= G_\eta(G_\eta(x_t, y, t, u), y, u, s) \\
&= \frac{\alpha_s}{\alpha_t} x_t + \sigma_s \epsilon_\eta(x_u, y, u) \\
&+ \frac{\alpha_s}{\alpha_u} \sigma_u (\epsilon_\eta(x_t, y, t) - \epsilon_\eta(x_u, y, u)) \\
&- \frac{\alpha_s}{\alpha_t} \sigma_t \epsilon_\eta(x_t, y, t),
\end{aligned}
\tag{4}
$$

To facilitate trajectory learning during distillation, the student model should match the teacher trajectory from t to s. Denoting the student trajectory as $x_{s,t}^\theta = G_\eta(x_t, y, t, s)$, the distillation regularization is formulated as:

$$\theta = \arg\min_\theta \|x_{s,t}^\theta - x_{s,u,t}^\eta\|_2^2. \tag{5}$$

With such regularization, the information intermediate step $u \in [s, t]$ will be distilled in student model. In practice, we utilize the clean image predicted by the trajectory both on the student and teacher to perform distillation for stable training. The predicted clean images of the teacher and student, denoted as x_{target} and x_{est}, are formulated as follows:

$$x_{target} = \frac{x_{s,u,t}^\eta - (\sigma_s/\sigma_t)x_t}{\bar{\alpha}_s - (\sigma_s/\sigma_t)\bar{\alpha}_t}, \quad x_{est} = \frac{x_{s,t}^\theta - (\sigma_s/\sigma_t)x_t}{\bar{\alpha}_s - (\sigma_s/\sigma_t)\bar{\alpha}_t}, \tag{6}$$

and the distillation regularization is modified as:

$$\theta = \arg\min_\theta \lambda(t)\|x_{target} - x_{est}\|_2^2, \tag{7}$$

where $\lambda(t)$ is an adaptive weight and is set as $\max(1, \frac{\alpha_t^2}{\sigma_t^2})$.

3.2. On the Refinement of the Teacher Trajectory

Direct application of trajectory distillation often leads to significant performance degradation. In this section, we theoretically analyze the core reasons behind this degradation and propose a strategy to mitigate its effects.

21279

On the fitting error of the teacher trajectory. To understand this, let us revisit Equation (3) and Equation (4), and consider the ideal condition where $\mathcal{L}_{distill} = 0$, meaning that the student's trajectory perfectly matches the teacher's trajectory. Under this condition, we have:

$$(\sigma_s - \frac{\alpha_s}{\alpha_t}\sigma_t)\epsilon_\theta(x_t, y, t) = \sigma_s\epsilon_\eta(x_u, y, u) - \frac{\alpha_s}{\alpha_t}\sigma_t\epsilon_\eta(x_t, y, t)$$
$$+ \frac{\alpha_s}{\alpha_u}\sigma_u\Big(\epsilon_\eta(x_t, y, t) - \epsilon_\eta(x_u, y, u)\Big).$$
$$(8)$$

Since the teacher model is trained using the vanilla diffusion loss function, the ideal teacher trajectory satisfies the condition $\epsilon_\eta(x_t, y, t) = \epsilon_\eta(x_u, y, u) = \tilde{\epsilon}$. Under this condition, the Equation (8) can be simplified as:

$$(\sigma_s - \frac{\alpha_s}{\alpha_t}\sigma_t)\epsilon_\theta(x_t, y, t) = (\sigma_s - \frac{\alpha_s}{\alpha_t}\sigma_t)\tilde{\epsilon}. \quad (9)$$

However, the existence of fitting errors makes it impossible to achieve this goal. The presence of guidance with undesired components inevitably leads to suboptimal results.

On the mitigation of the fitting error. Fortunately, we find that these undesired terms can be mitigated by a scaling parameter $\omega \in (0, 1]$. For the term $\frac{\alpha_s}{\alpha_u}\sigma_u\Big(\epsilon_\eta(x_t, y, t) - \epsilon_\eta(x_u, y, u)\Big)$, we can further reduce its impact by multiplying it by the scaling parameter ω. The mismatch between $\epsilon_\eta(x_u)$ and $\epsilon_\eta(x_t)$ introduces a more significant optimization conflict. Due to the presence of fitting error, this term is not always consistent with $\epsilon_\eta(x_t, y, t)$. Our key insight is to refine this term through linear interpolation toward the ideal value, expressed as $\sigma_t\Big(\omega\epsilon_\eta(x_t, y, t) + (1 - \omega)\tilde{\epsilon}\Big)$. This approach not only maintains consistency with previous terms, but also provides linear trajectory guidance. After applying these operations, the refined teacher trajectory can be expressed as follows:

$$\tilde{x}^\eta_{s,u,t} = \frac{\alpha_s}{\alpha_t}x_t + \sigma_s\epsilon_\eta(x_u, y, u)$$
$$+ \omega\frac{\alpha_s}{\alpha_u}\sigma_u\Big(\epsilon_\eta(x_t, y, t) - \epsilon_\eta(x_u, y, u)\Big) \quad (10)$$
$$- \frac{\alpha_s}{\alpha_t}\sigma_t\Big(\omega\epsilon_\eta(x_t, y, t) + (1 - \omega)\tilde{\epsilon}\Big).$$

To this end, the distillation loss, $\mathcal{L}_{distill}$, is equivalent to the following, disregarding the coefficients of each term:

$$\mathcal{L}_{distill} = ||\epsilon_\theta(x_t, y, t) - \epsilon_\eta(x_u, y, u)||$$
$$+ ||\epsilon_\theta(x_t, y, t) - (\omega\epsilon_\eta(x_t, y, t) + (1 - \omega)\tilde{\epsilon})||. \quad (11)$$

On the mitigation of the inference gap. Considering the vanilla diffusion are trained on the Gaussian flow, we determine the $\tilde{\epsilon}$ not the pure Gaussian noise but shift it into a residual space as:

$$\tilde{\epsilon} = \frac{x_t - \alpha_t\tilde{x}_0}{\sigma_t} = \frac{\alpha_t(x_0 - \tilde{x}_0)}{\sigma_t} + \epsilon, \quad (12)$$

where $\tilde{x}_0$ should lie between the low-light and clean image spaces, serving as an intermediate initial space for the student model to learn. This positioning ensures a closer initial distribution for the student model compared to the Gaussian distribution.

Additionally, our investigation shows that the refinement of the teacher trajectory in Equation (10) can be implemented through a straightforward approach, as outlined in the Corollary 1.

Corollary 1 (Proof in the supplementary material). *Given the refinement component $\tilde{x}_s = \alpha_s\tilde{x}_0 + \sigma_s\epsilon_\eta$, the Equation (10) is equivalent to : $\tilde{x}^\eta_{s,u,t} = \omega x^\eta_{s,u,t} + (1 - \omega)\tilde{x}_s$.*

3.3. Reflectance-Aware Trajectory Refinement

On the determination of the component $\tilde{x}_0$, using the ground truth clean images x_0 will cause $\tilde{\epsilon}$ degrade to ϵ. A natural approach is to use the low-light images y as $\tilde{x}_0$. In practice, the reflectance, which shares characteristics with both the clean images and the low-light images y, serves as a better component for refining the trajectory. Building on this observation, we propose the RATR module to refine the teacher's trajectory, thereby reducing inference gap.

Given an illumination map h and the noise z, the reflectance can be obtained through $x = \frac{y-z}{h}$ based on the Retinex theory. For illumination estimation, we employ the maximum channel of the low-light image y to represent the estimated illumination map h', as the common practice. Regarding ISO noise estimation, similar to the previous non-learning-based denoising method [1], the noise can be modeled as the distance between the noisy image and the clean image. We use this distance to estimate the noise map of the input low-light images:

$$z' = |y - \psi(y)|, \quad (13)$$

where ψ represents a non-learning-based denoising operation, allowing for flexible distillation. With these estimations, we can obtain a latent clean image $\tilde{x}_0 = \frac{y-z'}{h'}$. The $\tilde{x}_0$ is then employed to the trajectory refinement to perform distillation. Extensive experiments in ablation studies show that such refinement reduce the inference gap and greatly improves performance.

3.4. Auxiliary Loss

In knowledge distillation for classification, direct training signal derived from the data label will help student classifier outperforms the teacher classifier. In ReDDiT, we extend the principles to our model by introducing direct signals from both pixel and feature spaces. In the pixel space, we employ L2 loss on the pixel space:

$$\mathcal{L}_{pix} = \lambda_{pix}||x_0 - x_{est}||^2_2. \quad (14)$$

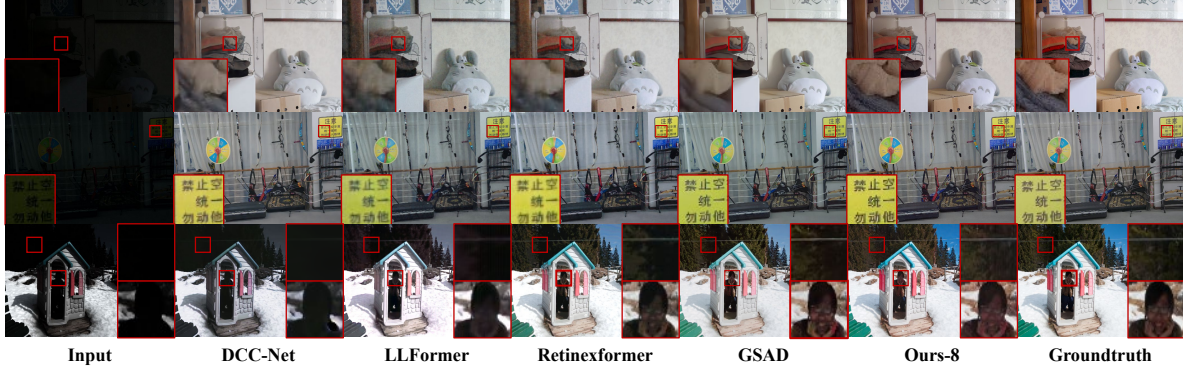| Input | DCC-Net | LLFormer | Retinexformer | GSAD | Ours-8 | Groundtruth |

Figure 3. Qualitative results on LOLv1 (top) , LOLv2-real (middle), and LOLv2-synthetic (bottom). Patches highlighted in each image by the red box indicate that ReDDiT effectively enhances the visibility, preserves the color and generates finer details in normal light images. Zoom in to better observe the image details.

In the feature space, we employ the perceptual loss for enhancing the student learning in structure and texture details:

$$\mathcal{L}_{per} = \lambda_{per}\|\phi(x_0) - \phi(x_{est})\|_2^2. \quad (15)$$

The final loss in the distillation stage integrates all the components into a single and unified training framework:

$$\mathcal{L}_{total} = \mathcal{L}_{distill} + \mathcal{L}_{per} + \mathcal{L}_{pix}. \quad (16)$$

4. Experiments

4.1. Dataset and Implementation Details

Datasets. We evaluate the performance of ReDDiT across various datasets in low-light image regions, including LOLv1 [35], LOLv2 [39], SID [3], SDSD [29], DICM [14], LIME [8], MEF [15], NPE [30], and VV [27].

Evaluation metrics. We comprehensively evaluate various Low-Light Image Enhancement (LLIE) methods, employing both full-reference and non-reference image quality metrics. In cases with paired data, we measure the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), along with the learned perceptual image patch similarity (LPIPS). For datasets such as DICM, LIME, MEF, NPE, and VV, which lack paired data, we rely on the Naturalness Image Quality Evaluator (NIQE) for assessment.

4.2. Comparison with State-of-the-Art Methods

Results on LOLv1 and LOLv2. On LOLv1 and LOLv2, we compare ReDDiT against LIME [8], RetinexNet [35], KinD [45], Zero-DCE [7], DRBN [38], KinD++ [46], EnlightenGAN [12], MIRNet [43], LLFlow [33], DCC-Net [47], SNR-Net [37], LLFormer [32], PyDiff [48], WCDM [31], and GSAD [10]. Tab. 1 presents the quantitative results of various LLIE methods, showcasing our method's superior performance across all compared metrics, including PSNR, SSIM, and LPIPS. New SOTA PSNR results of 28.090, 31.250 and 30.166 can be observed on LOLv1,

LOLv2-real and LOLv2-synthetic individually. Notably, ReDDiT consistently outperforms other methods, exhibiting substantial improvements in LPIPS scores, indicative of enhanced perceptual quality. Remarkably, ReDDiT achieves SOTA performance on LOLv2-real/LOLv2-synthetic datasets across all distilled models (8, 4, 2 steps). On LOLv1 datasets, our method attains SOTA results with 8 and 4 steps sampling and remains comparable to previous methods with 2 steps. The visual comparison presented in Figure 3 further highlights the effectiveness of ReDDiT in mitigating artifacts and enhancing image details. Notably, as depicted in the red box of Figure 3, ReDDiT excels in restoring clear structures and intricate details. This underscores our method's ability to leverage generative modeling to capture natural image distributions and retain such characteristics in the distilled models, resulting in superior visual effects.

Results on SID and SDSD. On SID and SDSD, we compare ReDDiT with SID [3], DeepUPE [28], Uformer [34], RetinexNet [35], KinD [45], DRBN [38], EnlightenGAN [12], MIRNet [43], SNR-Net [37], and Retinexformer [2]. Tab. 2 presents the quantitative results on the SID and SDSD, indicating its capability of ReDDiT to handle more complex low-light conditions effectively. Specifically, our method establishes new SOTA PSNR values of 25.32 dB/29.95 dB on SID and SDSD dataset. The visual comparison is provided in supplementary materials.

Results on DICM, LIME, MEF, NPE, and VV. We directly apply the model trained on the LOLv2-synthetic dataset to these unpaired real-world datasets. From Tab. 3, it is evident that ReDDiT outperforms all competitors in terms of NIQE scores, demonstrating its robust generalization ability. As shown in Figure 4, our model adeptly adjusts illumination conditions, effectively enhancing visibility in low-light areas while avoiding over-exposure.

Methods	NFE	LOLv1			LOLv2-real			LOLv2-synthetic		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
LIME [8] (TIP 16)		16.760	0.560	0.350	15.240	0.470	0.415	16.880	0.776	0.675
Zero-DCE [7] (CVPR 20)		14.861	0.562	0.335	18.059	0.580	0.313	-	-	-
EnlightenGAN [12] (TIP 21)		17.483	0.652	0.322	18.640	0.677	0.309	16.570	0.734	-
RetinexNet [35] (BMVC 18)		16.770	0.462	0.474	18.371	0.723	0.365	17.130	0.798	0.754
DRBN [38] (CVPR 20)		19.860	0.834	0.155	20.130	0.830	0.147	23.220	0.927	-
KinD [45] (ACM MM 19)	1	20.870	0.799	0.207	17.544	0.669	0.375	16.259	0.591	0.435
KinD++ [45] (IJCV 20)		21.300	0.823	0.175	19.087	0.817	0.180	-	-	-
MIRNet [43] (TPAMI 22)		24.140	0.842	0.131	20.357	0.782	0.317	21.940	0.846	-
LLFlow [33] (AAAI 22)		25.132	0.872	0.117	26.200	0.888	0.137	24.807	0.9193	0.067
SNR-Net [37] (CVPR 22)		26.716	0.851	0.152	27.209	0.871	0.157	27.787	0.941	0.054
LLFormer [32] (AAAI 23)		25.758	0.823	0.167	26.197	0.819	0.209	28.006	0.927	0.061
Retinexformer [2] (ICCV 23)		27.180	0.850	-	27.710	0.856	-	29.040	0.939	-
PyDiff [48] (IJCAI 23)	4	27.090	**0.930**	0.100	29.629	0.876	0.194	20.659	0.879	0.275
WCDM [11] (ToG 23)	10	26.947	0.847	0.088	30.461	0.879	0.070	29.852	0.908	0.077
GSAD [10] (NIPS 23)	10	27.839	<u>0.877</u>	0.091	28.818	0.895	0.095	28.670	**0.944**	0.047
Ours-8	8	**28.090**	0.861	**0.052**	<u>30.919</u>	<u>0.895</u>	<u>0.040</u>	**30.166**	<u>0.943</u>	**0.028**
Ours-4	4	<u>27.981</u>	0.872	<u>0.053</u>	**31.250**	**0.904**	**0.038**	<u>30.034</u>	0.933	<u>0.031</u>
Ours-2	2	27.397	0.848	0.062	30.613	0.890	0.045	29.346	0.913	0.040

Table 1. Quantitative comparisons are conducted on the LOLv1 and LOLv2 (real and synthetic) datasets. The best result is highlighted in bold, while the second-best result is underlined. Our method significantly outperformed SOTA methods.
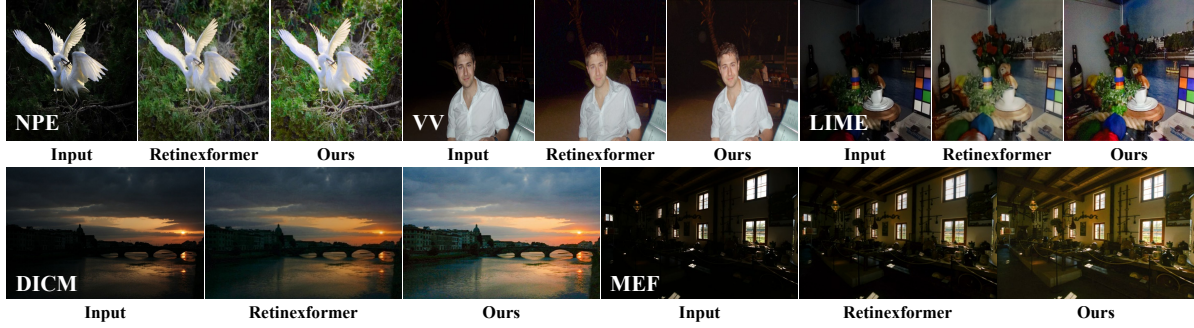

Figure 4. Qualitative results on NPE, VV, LIME, DICM, and MEF. Compared to Retinexformer, our method effectively enhances the visibility and preserves the color in normal light images. Zoom in to better observe the image details.

Methods	SID		SDSD	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
SID [3] (CVPR 18)	16.97	0.591	23.29	0.703
RetinexNet [35] (BMVC 18)	16.48	0.578	20.84	0.617
DeepUPE [28] (CVPR 19)	17.01	0.604	21.70	0.662
KinD [45] (ACM MM 19)	18.02	0.583	21.95	0.672
3DLUT [44] (TPAMI 20)	20.11	0.592	21.66	0.655
DRBN [38] (CVPR 20)	19.02	0.577	24.08	0.868
EnGAN [12] (TIP 21)	17.23	0.543	20.02	0.604
Uformer [34] (CVPR 22)	18.54	0.577	23.17	0.859
Restormer [42] (CVPR 22)	22.27	0.649	25.67	0.827
MIRNet [43] (TPAMI 22)	20.84	0.605	24.38	0.864
SNR-Net [37] (CVPR 22)	22.87	0.625	29.44	<u>0.894</u>
Retinexformer [2] (ICCV 23)	24.44	0.680	29.77	**0.896**
Ours-8	25.32	**0.802**	**29.95**	0.883
Ours-4	<u>25.85</u>	<u>0.792</u>	<u>29.52</u>	0.857
Ours-2	**25.88**	0.745	28.10	0.731

Table 2. Quantitative comparisons on SID and SDSD datasets. The best result is highlighted in bold, while the second-best result is underlined.

Methods	DICM	LIME	MEF	NPE	VV
LIME [8]	5.42	3.84	3.73	4.37	3.63
Zero-DCE [7]	3.77	3.61	4.12	6.19	4.81
EnGAN [12]	3.98	4.49	4.95	3.42	3.93
KinD++ [46]	6.19	4.24	3.93	3.44	3.72
SCI [20]	3.73	3.52	4.07	6.44	4.05
DCC-Net [47]	3.84	6.27	6.04	7.97	4.66
Retinexformer [2]	4.47	3.96	4.12	4.28	4.53
WCDM [11]	6.38	4.62	4.86	4.65	5.39
GSAD [10]	4.35	3.97	4.08	5.35	3.18
Ours	**3.62**	**3.45**	**3.93**	**3.24**	**3.00**

Table 3. Quantitative comparisons of different methods on DICM, LIME, MEF, NPE, and VV in terms of NIQE. Lower values indicate better performance, with the best result highlighted in bold.

4.3. Comparison with Other Accelerate Methods

We conduct a comprehensive evaluation of various acceleration techniques applied to our pre-trained diffusion model.

The quantitative results of this comparative analysis are presented in Figure 5. Notably, our investigation highlights the critical importance of employing distillation strategies in enhancing model performance.

Among the evaluated techniques, our distillation strategy emerges as the top performer across all datasets and all sampling steps. We achieve superior results compared to
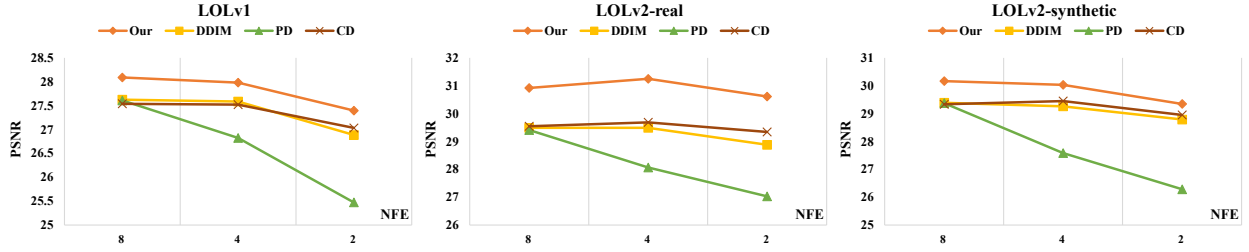
Figure 5. Quantitative comparisons on the LOLv1 and LOLv2 (Real and Synthetic) datasets with other accelerate methods. Popular acceleration methods experience a significant decline in performance as the number of steps decreases. Our method outperforms other acceleration techniques as well as **the original teacher model (in terms of DDIM performance)**.
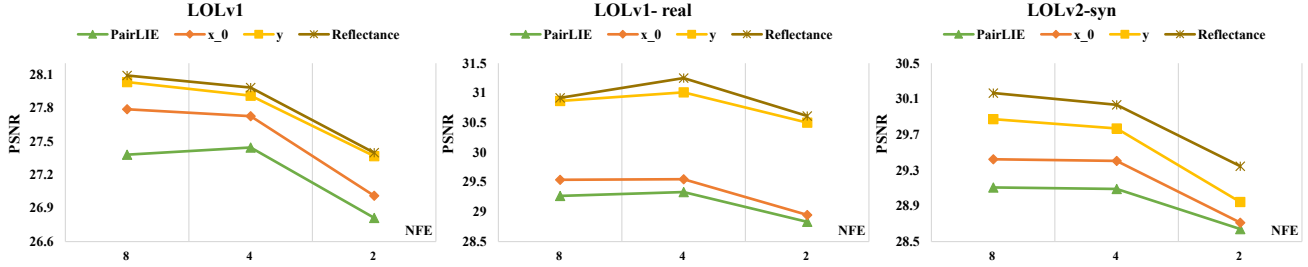

Figure 6. An ablation study conducted to evaluate the choice of $\tilde{x}$. The results indicate that reflectance contributes the most to performance during distillation, while low-light images y offer the second-best performance. Both reflectance and low-light images y effectively explore a better residual space, leading to significant improvements in performance compared to the original Gaussian space (x_0 as $\tilde{x}$).

the teacher model, as demonstrated by the performance improvements over the straightforward application of DDIM. However, it is noteworthy that recent advancements in consistency distillation techniques showcase slight superiority over traditional ODE solver methods, further underscoring the significance of distillation approaches in optimizing model performance.

Most notably, ReDDiT significantly outperforms alternative techniques, demonstrating the substantial benefits of investing in additional training costs to achieve superior results. This emphasizes the pivotal role of distillation strategies in effectively transferring knowledge from a well-prepared teacher model to its student counterpart, thereby elevating the overall performance of ReDDiT. These findings underscore the significance of ReDDiT's superior performance, reinforcing its efficacy and the value of employing sophisticated distillation techniques in advancing the SOTA in low-light image enhancement.

4.4. Ablation Study

The ablation on refinement module. The performance degradation without the refinement module is evident in Figure 7, where the absence of this module results in the most significant decline compared to the SOTA performance. Further analysis in Figure 8 delves into the impact of the refinement module in more detail.

The refinement module's influence becomes more pronounced as the number of sampling steps decreases. In the case of 2-step distillation, Figure 8 illustrates a consistent performance improvement with increasing strength of the

refinements. On LOLv2-real, the 4-step distillation results even surpass the performance of the 8-step diffusion model and achieve new SOTA results with $\omega = 0.8$. This trend is attributed to the heightened significance of the refinement module in addressing the increasing fitting error of the teacher model as the number of steps decreases. Therefore, the refinement module becomes crucial for accurately estimating the trajectory under such conditions.

In our previous experiments, we select the refinement strength that yielded the best performance in the previous ablation studies. This decision ensures that the refinement module effectively mitigates the fitting error of the teacher model, thereby enhancing the performance of the student model during distillation.

The ablation on the choice of $\tilde{x}$. We compare the choice of $\tilde{x}$ with refinement using clean images x_0, low-light images y, the enhanced result from the learnable Retinex method (*e.g.* PairLIE [6]), and our reflectance component, as shown in Figure 6. Both reflectance and low-light images y effectively explore a better residual space, leading to significant improvements in performance compared to the original Gaussian space (x_0 as $\tilde{x}$). This suggests that shifting the residual space is crucial for enhancing the performance of the student model in LLIE tasks.

Furthermore, the PairLIE results, when used for refinement, perform even poorly compared to refinements using clean images x_0 and low-light images y. Although its prediction is closer to clean images x_0, it fails to identify a suitable residual space for the student model to learn from, ultimately failing to address the inference gap.

Figure 7. The ablation study on the main components of our method. It can be observed that the refinement module significantly contributes to the final performance, with its impact becoming more pronounced as the number of sampling steps decreases.
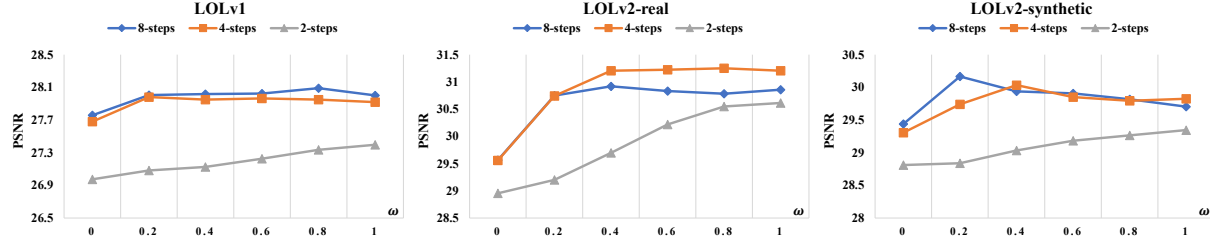

Figure 8. The ablation study on the weights of the reflectance-aware trajectory refinement. The horizontal axis represents the strength of the refinement and the vertical axis represents PSNR values.

The ablation on the auxiliary loss. As depicted in Figure 7, conducting ablation studies on $\mathcal{L}_{pix}$ and $\mathcal{L}_{per}$ results in a slight decrease in distillation performance. This observation underscores the effectiveness of directly obtaining supervision signals from ground truth data. Interestingly, $\mathcal{L}_{per}$ exhibits a more pronounced influence on performance compared to $\mathcal{L}_{pix}$, emphasizing the significance of leveraging supervision signals in the feature space.

Furthermore, the degradation in performance resulting from ablation of either loss term is minor compared to the substantial degradation observed when the RATR module is removed. This finding highlights the critical role of the RATR module within our framework.

4.5. Efficiency Comparison

We evaluate the efficiency of our method in terms of inference time, frames per second (FPS), and number of parameters, comparing it with recent diffusion-based LLIE methods. The quantitative results of this comparative analysis are presented in Tab. 4. Our approach, particularly in its 2-step variant, demonstrates superior performance regarding inference speed, FPS, and parameter efficiency. It achieves an optimal balance between computational efficiency and model performance, surpassing other methods in both speed and resource utilization. This comparison highlights the effectiveness of our method in delivering high-performance results with minimal computational overhead.

4.6. Limitation and Future work

While ReDDiT has excelled in restoration within 2 steps, single-step restoration is not yet optimal, presenting a notable limitation. In single-step restoration, inefficient de-

	Teacher-16	PyDiff	GSAD	WCDM	Ours-8	Ours-4	Ours-2
Time (s)	0.611	0.254	0.381	_0.117_	0.323	0.162	**0.076**
FPS	1.64	3.92	2.62	_8.51_	3.09	6.15	**13.1**
Param (M)	17.43	97.89	17.43	_22.08_	17.43	17.43	**17.43**

Table 4. The efficiency comparison with other methods. Our method resulted in a fast and lightweight diffusion-based low-light enhancer with excellent performance.

noising results in low PSNR and artifacts in detail. Failure cases in the single step will be demonstrated in the supplementary materials. Additionally, the exploration of a lightweight denoising network was not addressed in this work. In our future work, we will continue exploring efficient diffusion-based methods for LLIE. Our focus will be on investigating the potential of a single-step diffusion model and developing a lightweight denoising network.

5. Conclusions

In this paper, we theoretically analyze the two main factors contributing to the performance degradation of the diffusion distillation technique. Building on this, we introduce ReDDiT, a significant advancement in efficient diffusion models for LLIE. Central to ReDDiT is the use of linear exploration in the reflectance-aware residual space, which reduces trajectory fitting errors and the sampling gap. These innovations enable ReDDiT to preserve the intrinsic structural integrity of images while minimizing sampling steps. ReDDiT achieves performance comparable to previous methods with just 2 steps and sets new SOTA results at 4 and 8 steps. Across 10 benchmark datasets, ReDDiT consistently outperforms existing methods. This marks a promising step toward real-time diffusion models for LLIE, and our research in this area will continue.

References

[1] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pages 60–65. Ieee, 2005. 4

[2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 5, 6

[3] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 5, 6

[4] Jingwen Chen, Yingwei Pan, Ting Yao, and Tao Mei. Controlstyle: Text-driven stylized image generation using diffusion priors. In *Proceedings of the 31st ACM International Conference on Multimedia*, page 7540–7548. Association for Computing Machinery, 2023. 2

[5] Zijun Deng, Xiangteng He, Yuxin Peng, Xiongwei Zhu, and Lele Cheng. Mv-diffusion: Motion-aware video diffusion model. In *Proceedings of the 31st ACM International Conference on Multimedia*, page 7255–7263. Association for Computing Machinery, 2023. 2

[6] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22252–22261, 2023. 7

[7] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 2, 5, 6

[8] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 5, 6

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, pages 6840–6851. Curran Associates, Inc., 2020. 1

[10] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2, 5, 6

[11] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 1, 2, 6

[12] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 2, 5, 6

[13] Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ode trajectory of diffusion. *arXiv preprint arXiv:2310.02279*, 2023. 2

[14] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation. In *2012 19th IEEE international conference on image processing*, pages 965–968. IEEE, 2012. 5

[15] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 5

[16] Yanyu Li, Huan Wang, Qing Jin, Ju Hu, Pavlo Chemerys, Yun Fu, Yanzhi Wang, Sergey Tulyakov, and Jian Ren. Snapfusion: Text-to-image diffusion model on mobile devices within two seconds. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[17] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022. 2

[18] C Lu, Y Zhou, F Bao, J Chen, C Li, and J Dpmsolver+ Zhu. Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022. 2

[19] Eric Luhman and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021. 2

[20] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5637–5646, 2022. 6

[21] Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14297–14306, 2023. 2

[22] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1, 2

[23] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023. 2

[24] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2021. 1, 2

[25] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2020. 1, 2

[26] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proceedings of the*

40th International Conference on Machine Learning, pages 32211–32252, 2023. 1, 2

[27] Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77:9211–9231, 2018. 5

[28] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6849–6857, 2019. 5, 6

[29] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu, Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the dark: A high-quality video dataset with mechatronic alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9700–9709, 2021. 5

[30] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013. 5

[31] Tao Wang, Kaihao Zhang, Ziqian Shao, Wenhan Luo, Bjorn Stenger, Tae-Kyun Kim, Wei Liu, and Hongdong Li. Lld-iffusion: Learning degradation representations in diffusion models for low-light image enhancement. *arXiv preprint arXiv:2307.14659*, 2023. 5

[32] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2654–2662, 2023. 5, 6

[33] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2604–2612, 2022. 5, 6

[34] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 5, 6

[35] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 5, 6

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

[37] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 5, 6

[38] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3063–3072, 2020. 5, 6

[39] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex

network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 5

[40] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12302–12311, 2023. 1, 2

[41] Yuyang Yin, Dejia Xu, Chuangchuang Tan, Ping Liu, Yao Zhao, and Yunchao Wei. Cle diffusion: Controllable light enhancement diffusion model. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 8145–8156, 2023. 1

[42] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 6

[43] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):1934–1948, 2022. 5, 6

[44] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2020. 6

[45] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 5, 6

[46] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129:1013–1037, 2021. 5, 6

[47] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1899–1908, 2022. 5, 6

[48] Dewei Zhou, Zongxin Yang, and Yi Yang. Pyramid diffusion models for low-light image enhancement. *arXiv preprint arXiv:2305.10028*, 2023. 1, 2, 5, 6