AGLLDiff: Guiding Diffusion Models Towards Unsupervised Training-free Real-world Low-light Image Enhancement

Yunlong Lin*[1], Tian Ye*[2], Sixiang Chen*[2], Zhenqi Fu[4], Yingying Wang[1], Wenhao Chai[5], Zhaohu Xing[2], Lei Zhu[2,3], and Xinghao Ding[1]✉

[1] Xiamen University, China
[2] The Hong Kong University of Science and Technology (Guangzhou), China
[3] The Hong Kong University of Science and Technology, Hong Kong SAR, China
[4] Tsinghua University, China
[5] University of Washington
dxh@xmu.edu.cn
Project page: https://aglldiff.github.io/

Abstract. Existing low-light image enhancement (LIE) methods have achieved noteworthy success in solving synthetic distortions, yet they often fall short in practical applications. The limitations arise from two inherent challenges in real-world LIE: 1) the collection of distorted/clean image pairs is often impractical and sometimes even unavailable, and 2) accurately modeling complex degradations presents a non-trivial problem. To overcome them, we propose the Attribute Guidance Diffusion framework (AGLLDiff), a training-free method for effective real-world LIE. Instead of specifically defining the degradation process, AGLLDiff shifts the paradigm and models the desired attributes, such as image exposure, structure and color of normal-light images. These attributes are readily available and impose no assumptions about the degradation process, which guides the diffusion sampling process to a reliable high-quality solution space. Extensive experiments demonstrate that our approach outperforms the current leading unsupervised LIE methods across benchmarks in terms of distortion-based and perceptual-based metrics, and it performs well even in sophisticated wild degradation.

Keywords: Low-light Image Enhancement · Diffusion Model · Real-world Generalization · Unsupervised Learning · Training-free

1 Introduction

Real-world low-light image enhancement (LIE) aims to ameliorate the quality and brightness of an image suffering from unknown degradation, such as low contrast, multiple artifacts, poor visibility, sensor noise, etc. Great improvement in enhancement quality has been witnessed over the past few years with the exploitation of generative priors [35,58,95]. For instance, Generative Adversarial

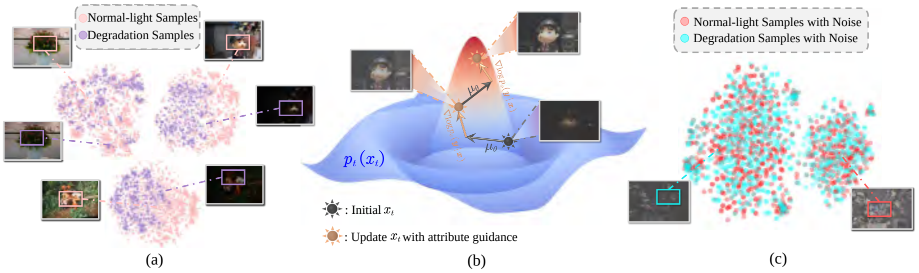

Normal-light Samples
Degradation Samples
(a)
$p_t(x_t)$
μ_θ
$\nabla \log p_\phi(y|x)$
μ_θ
$\nabla \log p_\phi(y|x)$
: Initial x_t
: Update x_t with attribute guidance
(b)
Normal-light Samples with Noise
Degradation Samples with Noise
(c)

Fig. 1: Motivation of our AGLLDiff. (a) represents the data distribution of normal-light samples and degraded samples. It is evident that degradation significantly deviates from normal-light samples. (b) conceptually illustrate the geometries of the proposed attribute guidance sampling algorithm. It shows that, given the initial latent, which lies in the low-probability region, attribute guidance guides the latent to move towards its vicinal high-probability region. (c) presents that imposing gaussian noise on the degraded sample and its corresponding reference sample makes the distributions between them less distinguishable.

Networks (GANs) [19,25] that are trained on extensive datasets of clean images and learn rich knowledge of real-world scenes have succeeded in LIE through GAN inversion. Compared to GANs, Denoising Diffusion Probabilistic Models (DDPMs) [7,8,20,30,36,38,63,79] yield more high-fidelity and realistic details, thereby fostering a surge of interest in adapting diffusion models to LIE [31,34,67,77,89,93].

Recent diffusion-based LIE methods can be roughly classified into two categories: ***1) The common approaches*** [67,77,88,89] are dedicated to accurately modeling degradation process via supervised learning, which show proficiency in synthetic degradation scenes but lack robustness to handle challenging unseen degradations. This inadaptability primarily stems from the inconsistency between the synthetic degradation of training data and the actual degradation in the real world. Enriching the synthetic data for model training would improve the models' generalizability, but it is obviously impractical to simulate every possible degradation in the real world. ***2) The second ones*** [23,49] strive to exploit the diffusion priors in the pre-trained diffusion models, which are effective in adapting to multiple degradations. Yet, despite their versatility, these methods are inevitably constrained in terms of generalizability, as they require prior knowledge of the specific degradation process in advance. In practice, degradations in the wild often include a mixture of multiple types, posing a challenge to accurately model them. In summary, two primary challenges are commonly encountered in real-world LIE: **i)** the collection of distorted/clean image pairs is often impractical and sometimes even unavailable, and **ii)** accurately modeling complex degradations presents a non-trivial problem.

To address the aforementioned challenges, we introduce ***a novel training-free and unsupervised framework, named AGLLDiff***, for real-world low-light enhancement. In contrast to prior works that predefine the degradation process, ***our approach models the desired attributes and incorporates this guidance within the diffusion generative process.*** Concrectly, we leverage a well-performing diffusion model (DM) [22, 61, 90], which generates images through a stochastic iterative sampling process, and the attributes act as classifiers to constrain the generative process to a reliable high-quality (HQ) solution space. As shown in Fig. 1, noisy images are degradation-irrelevant conditions for the DM generative process. Adding extra gaussian noise makes the degradation less distinguishable compared with its corresponding reference distribution. Since diffusion prior can serve as a natural image regularization, one could simply guide the sampling process with easily accessible attributes such as image exposure, structure and color of normal-light images. ***By constraining a reliable HQ solution space, the core of our philosophy is to bypass the difficulty of discerning the prior relationship between low-light and normal-light images, thus improving generalizability.***

Our contributions can be summarized as follows:

- We introduce a novel paradigm, AGLLDiff, a training-free and unsupervised method that requires no degradation of prior knowledge but yields high fidelity and generality towards real-world low-light image enhancement.
- We demonstrate that AGLLDiff suffices to guide the pre-trained diffusion models to a reliable high-quality solution space through easily accessible attributes in the HQ image space.
- Comprehensive experiments reveal that our framework achieves both robustness and high quality on heavily degraded synthetic and real-world datasets.

2 Related Work

Low-light Image Enhancement. To transform low-light images into visually satisfactory ones, numerous efforts have been made over the decades. The conventional approaches are first widely adopted [1, 17, 29, 33, 59, 60]. For example, Wang et al. [73] improved the visibility and contrast by applying gamma correction and enhancing the dynamic contrast ratio. Guo et al. [28] suggested refining the initially estimated illumination map by incorporating a structural prior. Arici et al. [2] introduced penalty terms to avoid the unnatural look and visual artifacts of the enhanced image. Lee et al. [40] applied the layered difference representation of 2D histograms to amplify the gray-level differences between adjacent pixels. The enhancement performance of current conventional methods relies on tedious hand-crafted priors and is only applicable to specific scenarios.

Recently, the paradigm of low-light image enhancement has gradually shifted to data-driven approaches based on deep learning [24, 26, 41, 43, 50, 78, 81, 92, 97]. For instance, Chen et al. [76] combined Retinex theory with a CNN network to estimate and adjust the illumination map. Ma et al. [54] established a cascaded illumination estimation process to achieve fast and robust LIE in complex scenarios. Lore et al. [51] developed a stacked sparse denoising autoencoder framework aimed at improving the quality of low-light images. Lv et al. [52] presented a multi-branch network that extracts rich features from different levels to enhance low-light images via multiple sub-networks. Xu et al. [78] proposed a signal-to-noise (SNR)-aware network that integrates a convolutional short-range branch with a transformer-based long-range branch. Cai et al. [6] designed a novel one-stage Retinex-based framework for LIE. Additionally, in contexts where training images are scarce, the utility of unsupervised [24,81] and zero-shot learning approaches [23, 26, 45] becomes increasingly pronounced.

Diffusion-Based Image Restoration and Low-Light Image Enhancement. Diffusion Models has become increasingly influential in the field of image restoration (IR) tasks [12,23,37,71,84–86], such as super resolution [62,96], blind face restoration [18, 74], image fusion [32, 47, 69, 70], dehazing [16, 87], desnowing [13–15], image enhancement [31, 77, 88, 93]. These methods could be broadly categorized into supervised and unsupervised paradigms. Supervised-based IR solutions usually rely on large-scale, pre-collected paired datasets to train their models with great success. Hou et al. [31] devised a diffusion-based framework, incorporating a global structure-aware regularization to maintain the intricate details and textures within images. Yi et al. [88] integrated the diffusion model alongside the Retinex model to enhance low-light images. Jiang et al. [34] employed wavelet transformation to decrease the input size and a high-frequency restoration module to maintain the details. Wang et al. [68] and Yin et al. [89] directly utilized the color map as an extra conditional control to preserve the color information. A major challenge is that they implicitly assume training and testing data should be identically distributed. As a result, these methods often deteriorate seriously in performance when testing cases deviate from the pre-assumed distribution.

Another prevailing research line is unsupervised-based IR approaches. They adopt a zero-shot approach to leverage a pre-trained diffusion model for restoration without the need for task-specific training. As an early attempt, Kawar et al. [37] hypothesized the linear degradation model and relied on the desirable property of linear formulation to sample from posterior distribution. Wang et al. [71] introduced the range-null space decomposition to further improve the zero-shot image restoration. Fei et al. [23] applied a simultaneous estimation of degradation model to address blind degradation. Yang et al. [80] introduced a partial guidance mechanism for blind face restoration, wherein intermediate outputs of the diffusion model are constrained by a classifier to perform photo restoration. Previous diffusion-based image restoration methods explicitly leveraged a degradation model by solving a maximum posterior problem or a posterior sampling problem to generate solutions. However, for many practical image enhancement problems, the underlying degradation model may not be available. ***In this work, we propose to model the desired attributes of normal-light images. Such a strategy is independent of the degradation process, circumventing the difficulty of modeling the degradation process.***

3 Methodology

3.1 Preliminaries of Diffusion Models

The diffusion models [3, 8, 9, 27, 30, 38, 63] belong to a category of generative models that operate by incrementally incorporating Gaussian noise into training data and subsequently acquiring the denoiser to restore the data distribution $p(\boldsymbol{x})$ by reversing the process of noise injection.

The forward process $q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right)$ transforms an initial image $\boldsymbol{x}_0$ into Gaussian noise $\boldsymbol{x}_T \sim \mathcal{N}(0,1)$ over T iterations. The following equation can express the process of each iteration in the diffusion:

$$q\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}\right)=\mathcal{N}\left(\boldsymbol{x}_t ; \sqrt{1-\beta_t} \boldsymbol{x}_{t-1}, \beta_t \boldsymbol{I}\right), \tag{1}$$

where $\boldsymbol{x}_t$ denotes the noisy image at time-step t, β_t is the pre-determined scaling factor, and $\mathcal{N}$ represents the Gaussian distribution. Under the reparameterization trick, $\boldsymbol{x}_t$ can be written as:

$$\boldsymbol{x}_t=\sqrt{\bar{\alpha}_t} \boldsymbol{x}_0+\sqrt{1-\bar{\alpha}_t} \epsilon, \tag{2}$$

where $\bar{\alpha}_t:=\prod_{i=1}^t\left(1-\beta_i\right)$ and $\epsilon \sim \mathcal{N}(0, I)$. Then $\boldsymbol{x}_T \sim \mathcal{N}(0, I)$ if T is big enough, usually $T=1000$.

The reverse generative process of the inference stage, starting from a Gaussian random noise map $\boldsymbol{x}_T \sim \mathcal{N}(0, I)$ and iteratively performing the denoising step until it attains a high-quality output $\boldsymbol{x}_0$:

$$p_{\boldsymbol{\theta}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t\right)=\mathcal{N}\left(\boldsymbol{x}_{t-1} ; \boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right), \Sigma_\theta \mathrm{I}\right), \tag{3}$$

where variance $\Sigma_\theta \mathrm{I}$ can be either time-dependent constants [30] or learnable parameters [57]. The mean value $\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)$ is generally parameterized by a network $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)$:

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t-\frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)\right). \tag{4}$$

In practice, one can also directly approximate $\hat{\boldsymbol{x}}_0$ from $\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)$:

$$\hat{\boldsymbol{x}}_0=\frac{\boldsymbol{x}_t}{\sqrt{\bar{\alpha}_t}}-\frac{\sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)}{\sqrt{\bar{\alpha}_t}}. \tag{5}$$

Classifier Guidance. The classifier guidance is employed to direct an unconditional diffusion model towards achieving conditional generation. Here, $\boldsymbol{y}$ represents the target and $\boldsymbol{p}_{\boldsymbol{\phi}}(\boldsymbol{y} \mid \boldsymbol{x})$ symbolizes a classifier, the conditional distribution is formulated to resemble a Gaussian distribution akin to its unconditional counterpart, but with the mean shifted by $\Sigma_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right) \boldsymbol{g}$ [22]:

$$\boldsymbol{p}_{\boldsymbol{\theta}, \boldsymbol{\phi}}\left(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t, \boldsymbol{y}\right) \approx \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)+\Sigma_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right) \boldsymbol{g}, \Sigma_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)\right), \tag{6}$$

where $\boldsymbol{g}:=\left.\nabla_{\boldsymbol{x}} \log \boldsymbol{p}_{\boldsymbol{\phi}}(\boldsymbol{y} \mid \boldsymbol{x})\right|_{\boldsymbol{x}=\boldsymbol{\mu}_{\boldsymbol{\theta}}\left(\boldsymbol{x}_t, t\right)}$. The gradient $\boldsymbol{g}$ serves as a guidance that leads the unconditional sampling distribution towards the condition target $\boldsymbol{y}$.

3.2 Overview of the AGLLDiff Framework

Our core motivation is to model the desired attributes of normal-light images and apply them to guide the diffusion generative process into a reliable high-quality (HQ) space. Such a design is agnostic to the degradation process and

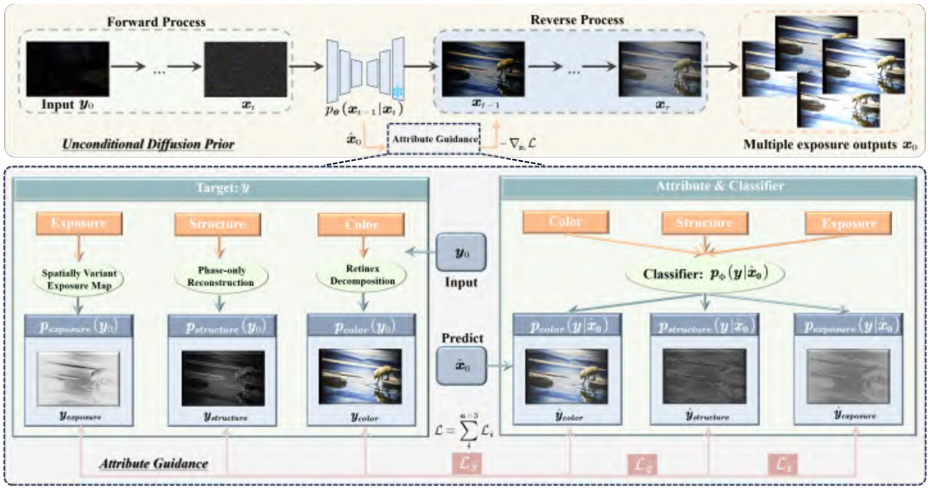


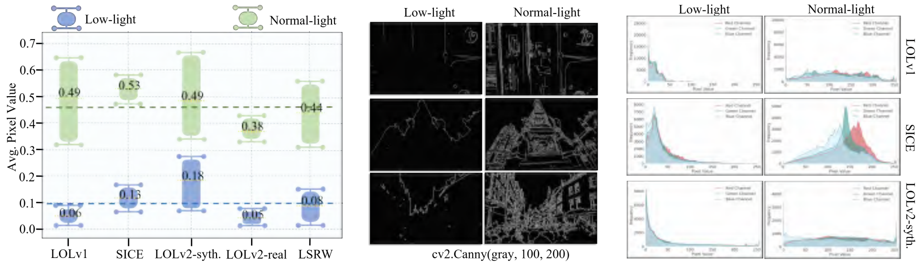
Fig. 2: The overall framework of our proposed AGLLDiff.

bypasses the difficulty of modeling the degradation process, making it more suitable for real-world LIE. The enhanced image should simultaneously satisfy: i) it is faithful to the degraded image, and ii) it conforms to the model distribution of pre-trained diffusion models that incorporate a vast repository of prior knowledge about HQ natural images. The overview of AGLLDiff is summarized in Fig. 2 and Algorithm 1. Given a degraded low-light image $\boldsymbol{y}_0$ in the wild domain, the diffusion forward process adds a few steps of slight Gaussian noise to the $\boldsymbol{y}_0$, aiming to narrow the distribution between the degraded image and its potential counterpart, i.e., the HQ image. After obtaining the noisy image $\boldsymbol{x}_t$, we implement the reverse generative process through a pre-trained diffusion denoiser and attribute guidance to generate the enhancement result $\boldsymbol{x}_0$. The inherent attributes of normal-light samples, such as image exposure, structure and color, can be readily derived from their degraded counterparts. Further elaboration on the pivotal components of our method, including attributes, classifiers, and targets, is provided in subsequent sections.

3.3 Attribute Guidance

Our attribute guidance eschews any assumptions about degradation. Instead, with diffusion prior acting as a regularization, we provide guidance only on the desired attributes of HQ images. The key to AGLLDiff is to construct proper guidance on the generative process.

Attribute and Classifier. The initial step of AGLLDiff is to determine the desired attributes that the normal-light output possesses. Each of these attributes corresponds to a specific classifier $\log \boldsymbol{p}_{\boldsymbol{\phi}}(\boldsymbol{y} \mid \boldsymbol{x}_0)$, and the intermediate outputs $\boldsymbol{x}_t$ are updated by back-propagating the gradient computed on the loss between the classifier output and the target $\boldsymbol{y}$. Through this mechanism, enhanced results can be obtained via an iterative refinement. The significance



(a) Average exposure levels for different datasets (b) Visualization structure via Canny operator (c) Average color distribution for different datasets

Fig. 3: Statistics and Visualization. (a) The average exposure values of the low- and normal-light subsets of the five LIE datasets. (b) Visualization of the structure of low- and normal-light images by the Canny operator. (c) Histogram of the color distribution of the low- and normal-light images.

of attributes in achieving the desired final result cannot be overstated. This raises a fundamental question: ***What attributes are possessed by HQ images? Through comprehensive observation and statistical analysis, we conclude that the following three fundamental attributes are typically found in HQ images: 1) well exposure, 2) clear structure, and 3) vivid colors.*** As presented in Fig. 3, we analyzed the average exposure levels and color distributions of five LIE datasets, visualizing the structure of low- and normal-light images. The observations are as follows: 1) there exists a significant discrepancy in average exposure values, with low-light images at around 0.1 and normal-light images at approximately 0.46, 2) normal-light images have clearer and richer structures compared to low-light images, and 3) the color distribution of images in normal light is both more colorful and homogeneous than in low-light images.

Exposure Guidance Formulation. To guide the output exposure toward that of normal-light images, we utilize the spatially variant exposure map [42] to constrain the exposure of the output $\hat{\boldsymbol{x}}_0$. The loss is formulated as follows:

$$\mathcal{L}_1 = \|\mathrm{Mean}\left(\hat{\boldsymbol{x}}_0\right) - \mathrm{Mean}\left(E\right)\|_2^2 , \tag{7}$$

where E denotes the spatially variant exposure map. Specifically, the exposure map is set with non-uniform exposure values in different regions, e.g., the underexposed region is assigned a large exposure value while the overexposed region is allocated a small exposure value. To achieve the spatially variant exposure map, we first obtain the luminance channel Y by color space decomposition (e.g., YCbCr, YUV) and its average value Y_{avg} of a low-light image $\boldsymbol{y}_0$. Then we calculate the spatially variant condition exposure map by:

$$E = A \times \mathrm{Norm}\left(Y - Y_{\mathrm{avg}}\right) + B, \tag{8}$$

where B is the base exposure value, A is the adjustment amplitude, and Norm operation normalizes its input to the range of $[-1, 1]$. Based on extensive statistics in Fig. 3(a), B and A are empirically set to 0.46 and 0.25, respectively. Fig. 4(a) presents three outcomes of our method utilizing spatially variant expo-

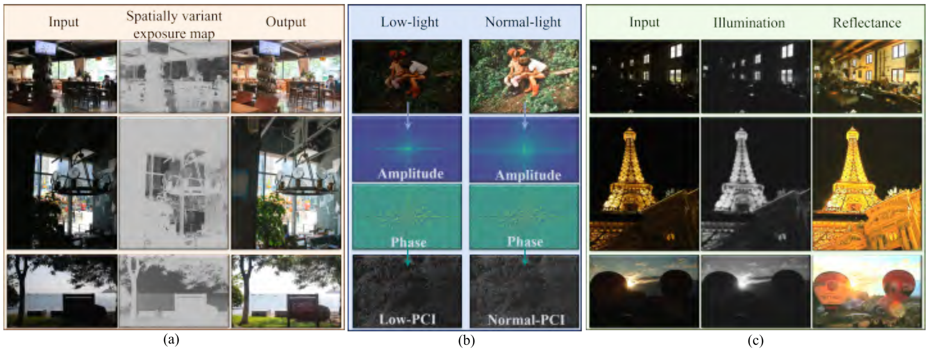


Fig. 4: (a) Visualization of the spatially variant exposure maps. Based on Eq. 8, we automatically assign the underexposed regions large exposure values (light gray) and wellexposed/overexposed regions small exposure values (dark gray). (b) Visualization of the phase-only reconstruction image (PCI) in the spatial domain. We applied an inverse discrete Fourier transform to the phase of the low/normal-light image to obtain the phase-only reconstruction image. That means the amplitude of low/normal-light image is set to 1. (c) Visualization of the Retinex decomposition. We employ a pre-trained decomposition network, RNet, to decompose the input into a reflectance map R and an illumination map L.

sure maps, where underexposed areas receive higher exposure values and well-exposed or overexposed areas receive lower ones. Such a spatially variant setting enables precise exposure adjustments and allows for generating multiple results at varying exposure levels by adjusting the exposure values.

Structure Guidance Formulation. For constraining the structure of the output faithful to the degraded image, we minimize the phase error between the degraded image $\boldsymbol{y}_0$ and the output $\hat{\boldsymbol{x}}_0$:

$$\mathcal{L}_2 = \left\| \mathcal{P}\left(\hat{\boldsymbol{x}}_0\right) - \mathcal{P}\left(\boldsymbol{y}_0\right) \right\|_2^2 , \tag{9}$$

where $\mathcal{P}(\cdot)$ indicates the phase in the Fourier domain. In Fig. 4(b), we perform the inverse discrete Fourier transform to obtain phase-only reconstruction images in the spatial domain. As observed, the phase-only reconstruction versions of low-light and normal-light exhibit structural consistency. This is because most illumination information is expressed as amplitudes, and structural information is revealed in phases [41]. We find that this simple phase constraint is sufficient to produce reliable results.

Color Guidance Formulation. According to the Retinex theory [39], a low-light image can be decomposed into illumination L and reflectance R. As shown in Fig. 4(c), the reflectance map R represents the physical properties of the objects, which contain abundant color information. Therefore, we could guide the color synthesis process with the reflectance map R. Equivalently, the loss is formulated as follows:

$$\mathcal{L}_3 = \left\| \mathcal{F}\left(\hat{\boldsymbol{x}}_0\right) - \mathcal{F}\left(\boldsymbol{y}_0\right) \right\|_2^2 , \tag{10}$$

where $\mathcal{F}(\cdot)$ denotes the pre-trained Retinex-based decomposition network [24], termed RNet. It takes a low-light image and generates the reflectance map R

Algorithm 1 Sampling with attribute guidance

```
Require: A pre-trained diffusion model $(\boldsymbol{\mu_\theta}(\boldsymbol{s}_t, t), \Sigma_\theta(\boldsymbol{x}_t, t))$, classifier $\boldsymbol{p_\theta}(\boldsymbol{y}|\boldsymbol{x}_0)$,
target $\boldsymbol{y}$, gradient scale $\boldsymbol{s}$, the number of gradient steps $\boldsymbol{N}$ and the iteration steps $\boldsymbol{\omega}$
of adding and removing noise.
Input: A low-light image $\boldsymbol{y}_0$
Output: Output image $\boldsymbol{x}_0$
$\boldsymbol{x}_\omega \leftarrow \sqrt{\bar{\alpha}_t}\boldsymbol{y}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$
for $t = \omega$ to $1$ do
    $\boldsymbol{\mu}_t, \Sigma \leftarrow \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)$
    $\hat{\boldsymbol{x}}_0 \leftarrow \frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t - \frac{\sqrt{1-\alpha_t}}{\alpha_t}\boldsymbol{\epsilon_\theta}(\boldsymbol{x}_t, t)$
    $\hat{\boldsymbol{s}} = \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_2}{\|\nabla_{\hat{\boldsymbol{x}}_0}\mathcal{L}\|_2} \cdot \boldsymbol{s}$        ▷ Dynamic guidance scale
    $\hat{\boldsymbol{N}} \leftarrow max\left(1, \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_2}{\|\nabla_{\hat{\boldsymbol{x}}_0}\mathcal{L}\|_2} \cdot \boldsymbol{N}\right)$        ▷ Dynamic gradient steps
    repeat
        $\boldsymbol{x}_t \leftarrow$ sample from $\mathcal{N}(\boldsymbol{\mu}_t - \hat{\boldsymbol{s}}\Sigma\nabla_{\hat{\boldsymbol{x}}_0}\log \boldsymbol{p_\theta}(\boldsymbol{y}|\hat{\boldsymbol{x}}_0), \Sigma)$
        $\hat{\boldsymbol{x}}_0 \leftarrow \frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t - \frac{\sqrt{1-\alpha_t}}{\alpha_t}\boldsymbol{\epsilon_\theta}(\boldsymbol{x}_t, t)$
    until $\hat{\boldsymbol{N}} - 1$ times
    $\boldsymbol{x}_{t-1} \leftarrow \mathcal{N}(\boldsymbol{\mu}_t - \hat{\boldsymbol{s}}\Sigma\nabla_{\hat{\boldsymbol{x}}_0}\log \boldsymbol{p}_\theta(\boldsymbol{y}|\hat{\boldsymbol{x}}_0), \Sigma)$
end for
return $\hat{\boldsymbol{x}}_0$
```

and illumination map L. RNet learns adaptive physical-based constraints from low-light image pairs in a self-supervised manner, significantly reducing the dependence on hand-crafted priors. Such an effective constraint ensures that our method can generalize well to various exposure scenes, which is why we chose it.

Our attribute guidance controls only the attributes of HQ outputs, and therefore composing the classifiers and summing the loss corresponding to each attribute can easily guide the diffusion model to generate HQ results:

$$\mathcal{L} \;=\; \lambda_1\mathcal{L}_1 + \lambda_2\mathcal{L}_2 + \lambda_3\mathcal{L}_3, \tag{11}$$

where λ_1, λ_2 and λ_3 are constants controlling the relative importance of the different losses, which are empirically set to 1000, 10 and 0.03 in all experiments, respectively.

Dynamic Guidance Scheme. As illustrated in Fig. 2, given the desired attributes, we construct the corresponding classifier and apply classifier guidance during the generative process. Similarly to the external classifier gradient guidance in [22], we evaluate the anti-gradient $-\nabla_{\hat{\boldsymbol{x}}_0}\mathcal{L}$ to bring the attribute-guidance to the generative process. ***However, our observations indicate that the traditional guidance scheme often leads to suboptimal outcomes.*** Concretely, the traditional guidance scheme, which adpots a constant gradient scale $\boldsymbol{s}$, often falls short in guiding the output towards the target value. Additionally, it executes merely one single gradient step per denoising step, which may not sufficiently steer the output towards the intended target, especially when early-phase denoising process intermediate outputs are significantly affected by noise. Such

Table 1: Quantitative comparison on LOLv1 [76], SICE [5] and LOLv2-synthetic [83]. "T", "S" and "U" represent "Traditional", "Supervised" and "Unsupervised" methods, respectively. The best results of "S" and "U" are marked in blue and orange, respectively.

Method	Type	LOLv1			SICE			LOLv2-synthetic		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SDD [29]	T	13.34	0.63	0.74	15.34	0.73	0.26	16.46	0.73	0.35
LECARM [60]	T	14.40	0.54	0.32	18.59	0.78	0.26	17.44	0.76	0.37
MBLLEN [52]	S	15.25	0.70	0.32	18.41	0.73	0.31	18.16	0.80	0.28
RetinexNet [76]	S	17.60	0.64	0.38	19.57	0.78	0.27	17.41	0.67	0.34
DSLR [46]	S	15.20	0.59	0.32	14.32	0.68	0.38	15.80	0.72	0.25
DRBN [82]	S	19.67	0.82	0.16	18.73	0.78	0.28	21.51	0.82	0.27
DiffLL [34]	S	26.19	0.85	0.11	21.33	0.84	0.22	29.46	0.92	0.09
PyDiff [93]	S	27.56	0.87	0.10	21.18	0.83	0.23	26.13	0.92	0.08
CUE [92]	S	22.67	0.79	0.20	20.06	0.82	0.24	24.47	0.90	0.12
Retinexformer [6]	S	25.15	0.84	0.13	22.32	0.85	0.20	25.66	0.95	0.05
EnlightenGAN [35]	U	17.48	0.65	0.32	18.73	0.82	0.23	16.79	0.76	0.31
RUAS [48]	U	16.40	0.49	0.27	13.21	0.72	0.43	16.31	0.65	0.38
SCI [54]	U	14.78	0.52	0.33	15.94	0.78	0.45	18.07	0.77	0.27
PairLIE [24]	U	19.46	0.73	0.24	21.23	0.83	0.22	19.12	0.77	0.23
NeRCo [81]	U	19.81	0.73	0.24	20.73	0.82	0.23	19.14	0.74	0.26
ZeroDCE [26]	U	14.86	0.55	0.33	18.67	0.80	0.26	17.75	0.83	0.16
ZeroDCE++ [43]	U	15.32	0.56	0.33	18.65	0.81	0.27	17.55	0.83	0.18
RRDNet [94]	U	11.38	0.51	0.36	13.27	0.68	0.32	14.85	0.65	0.24
CLIP-LIT [45]	U	12.39	0.49	0.38	13.70	0.73	0.30	16.18	0.79	0.20
GDP [23]	U	15.83	0.61	0.34	14.12	0.67	0.31	13.21	0.49	0.36
AGLLDiff (Ours)	U	21.81	0.84	0.15	22.12	0.84	0.21	21.11	0.87	0.13

limitations are particularly adverse within LIE tasks that demand high similarity to the target. To mitigate this issue, we introduce a dynamic guidance scheme that consists of two distinct components, i.e., the dynamic guidance scale $\hat{\boldsymbol{s}}$ and the dynamic gradient steps $\hat{\boldsymbol{N}}$ at each denoising step. Specifically, we calculate the $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{N}}$ based on the magnitude change of the intermediate image [10,65]:

$$\hat{\boldsymbol{s}} = \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_2}{\|\nabla_{\hat{\boldsymbol{x}}_0}\mathcal{L}\|_2} \cdot \boldsymbol{s} \text{ and } \hat{\boldsymbol{N}} = max\left(1, \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_2}{\|\nabla_{\hat{\boldsymbol{x}}_0}\mathcal{L}\|_2} \cdot \boldsymbol{N}\right), \tag{12}$$

where $\boldsymbol{x}_{t-1} \sim \mathcal{N}(\boldsymbol{\mu_\theta}, \Sigma_\theta)$, $\boldsymbol{s}$ and $\boldsymbol{N}$ are empirically set to 1.8 and 3 in all experiments, respectively. Such a dynamic guidance scheme affords users the flexibility to adjust the strength of guidance for attributes as per their unique requirements, thereby improving overall controllability.

4 Experiments

4.1 Implementation and Datasets

Inference Requirements. The pre-trained diffusion model we employ is a 256×256 denoising network trained on the ImageNet dataset [21] provided by [22]. The total number of iteration steps is defaulted to 1000. We select the final 10 steps to implement the noise addition and attribute guidance. The inference process is carried out on the NVIDIA RTX 3090 GPU.

Table 2: Quantitative comparison on DICM [40], MEF [53], LIME [44], NPE [66] and VV [64]. "T", "S" and "U" represent "Traditional", "Supervised", "Unsupervised" methods, respectively. The best results of "S" and "U" are marked in blue and orange, respectively. BRI. denotes BRISQUE.

Method	DICM			MEF			LIME			NPE			VV		
	NIQE↓	BRI.↓	PI ↓	NIQE↓	BRI.↓	PI ↓	NIQE↓	BRI.↓	PI ↓	NIQE↓	BRI.↓	PI ↓	NIQE↓	BRI.↓	PI ↓
SDD(T)	4.64	31.74	4.18	4.52	38.90	4.12	4.58	29.75	3.84	4.64	37.10	3.72	3.62	23.46	3.42
LECARM (T)	4.24	28.70	4.34	4.54	33.60	4.47	4.92	31.64	4.12	9.61	38.70	5.92	3.68	23.66	3.31
MBLLEN (S)	4.54	36.18	4.15	5.03	38.75	4.38	4.70	32.87	3.84	4.13	30.72	3.48	4.68	43.49	5.06
RetinexNet (S)	4.19	23.42	3.14	4.56	35.91	3.91	5.54	36.58	4.19	4.76	33.51	3.16	5.34	46.80	5.18
DSLR (S)	3.81	26.97	3.57	4.18	27.96	3.84	4.17	24.09	3.34	4.55	33.82	3.40	4.18	30.59	4.44
DRBN (S)	4.25	31.72	4.18	4.18	32.67	3.66	4.42	31.64	3.84	3.61	24.34	3.24	3.75	31.48	4.16
DiffLL (S)	3.70	18.08	3.13	3.46	23.27	2.99	3.60	19.44	3.06	3.46	14.97	2.52	2.75	18.53	3.03
PyDiff (S)	3.98	29.79	3.55	4.12	29.19	3.65	4.58	32.82	3.93	3.66	26.60	2.82	3.74	31.23	4.04
CUE (S)	3.76	17.87	3.32	3.63	25.81	3.21	3.83	16.90	3.08	3.53	19.82	2.75	3.54	21.83	3.96
Retinexformer (S)	3.72	17.22	3.08	3.44	22.01	3.07	3.86	17.41	2.96	3.39	20.51	2.56	2.97	16.24	3.35
EnlightenGAN (U)	3.89	29.23	3.41	3.56	25.31	3.25	4.21	25.34	3.34	3.66	24.10	2.95	3.63	27.79	3.82
RUAS (U)	6.58	45.06	5.22	5.40	41.68	4.57	5.36	31.62	4.34	7.12	46.89	5.61	4.86	34.03	4.24
SCI (U)	4.12	31.64	3.74	3.63	24.57	3.23	4.38	24.85	3.32	4.12	27.31	3.41	5.13	21.45	3.49
PairLIE (U)	4.13	28.59	3.69	4.18	29.54	3.10	4.51	25.21	3.26	4.17	26.22	3.03	3.66	25.88	3.69
NeRCo (U)	3.86	28.34	3.45	3.53	23.22	2.97	3.68	24.49	2.98	3.56	25.21	2.80	3.70	32.18	3.12
ZeroDCE (U)	3.91	24.05	3.13	3.51	26.63	3.13	4.34	26.48	3.21	3.80	22.34	2.92	4.12	25.13	3.33
ZeroDCE++ (U)	3.87	23.40	3.21	3.49	23.59	3.12	4.29	27.24	3.23	3.81	22.60	2.95	3.96	26.11	3.32
RRDNet (U)	3.81	23.95	3.21	3.55	25.92	3.35	4.35	35.23	3.81	3.69	25.51	3.24	3.65	29.37	3.28
CLIP-LIT (U)	3.71	25.78	3.24	3.57	27.72	3.22	3.99	24.41	3.07	3.62	24.14	2.74	3.33	28.66	3.09
GDP (U)	4.08	30.11	3.58	4.10	28.94	3.31	4.65	27.05	3.61	3.72	25.38	3.09	3.61	28.29	3.17
AGLLDiff (U)	3.57	19.13	3.07	3.44	23.21	3.11	3.64	19.13	2.98	3.50	15.13	2.58	3.50	21.13	2.77

Testing Datasets. We construct one synthetic dataset and seven real-world datasets for testing. The LOLv1 [76] dataset is composed of 500 low-light and normal-light image pairs and divided into 485 training pairs and 15 testing pairs. The LOLv2-synthetic [76] dataset is officially divided into two parts, i.e., 900 low-light images for training and 100 low-light images for testing. The SICE benchmark collects 224 normal-light images and 783 low-light images. Each normal-light image corresponds to 2∼4 low-light images. We adopt the first 50 normal-light images and the corresponding 150 low-light images for testing, and the rest (633 low-light images) for training. Moreover, we further assess our method on five commonly used real-world unpaired benchmarks: LIME [44], NPE [66], MEF [53], DICM [40], and VV [64]. Notably, we only utilize the testing sets to evaluate our approach. Additionally, unlike some existing methods such as LLFlow [72] that adjust brightness using reference images, potentially causing biases, we follow the approaches [24, 41] and compute metrics without using any reference information to ensure fairness.

Metrics. For the paired datasets, we adopt two distortion-based metrics: PSNR and SSIM [75] to evaluate the performance of the proposed method, and also the perceptual-based metric LPIPS [91] to measure the visual quality of the enhanced results. For the other five unpaired datasets, we use three non-reference perceptual-based metrics: NIQE [55], BRISQUE [56], and PI [4] for evaluation.

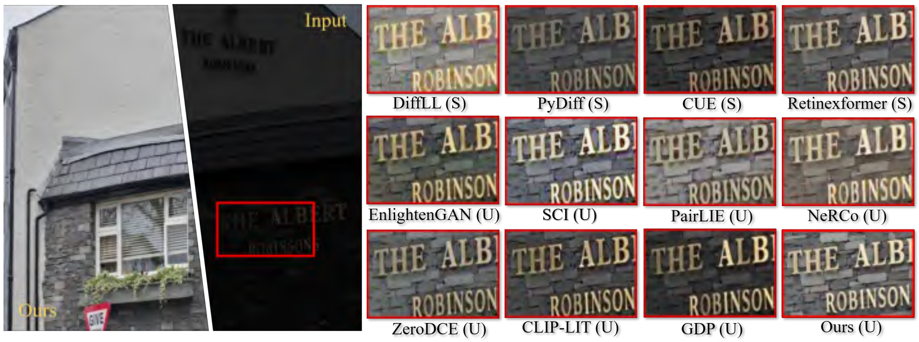


Fig. 5: Visual comparisons of various LIE methods on SICE. The proposed method achieves visually pleasing results in terms of brightness, color, contrast, and naturalness.

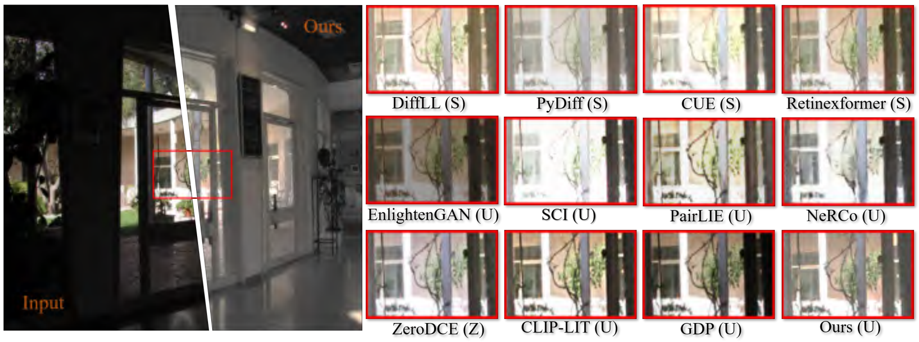


Fig. 6: Visual comparisons of various LIE methods on MEF. Our method achieves remarkably higher quality among unsupervised methods with less noise and artifacts.

4.2 Comparison with the State-of-the-Art

For a more comprehensive analysis, we compare our proposed AGLLDiff with three categories of existing state-of-the-art methods, including: 1) traditional methods SDD [29] and LECARM [60], 2) supervised approaches MBLLEN [52], RetinexNet [76], DSLR [46], DRBN [82], DiffLL [34], PyDiff [93], CUE [92] and Retinexformer [6], and 3) unsupervised methods EnlightenGAN [35], RUAS [48], SCI [54], PairLIE [24], NeRCo [81], ZeroDCE [26], ZeroDCE++ [43], RRD-Net [94], CLIP-LIT [45] and GDP [23]. Note that the results of all those methods are reproduced by using the official codes with recommended parameters. The metrics are recalculated with the *pyiqa* [11].

Quantitative Comparisons. Tables 1 and 2 report the quantitative performance of three paired datasets and five unpaired datasets, respectively. The best results of supervised and unsupervised methods are highlighted in blue and orange, respectively. Compared with recent competitive unsupervised approaches, AGLLDiff achieves the most advanced quantitative performance in terms of distortion-based and perceptual-based metrics across all benchmarks. Note that AGLLDiff even outperforms partial supervised approaches, which confirms the superiority of our solution.

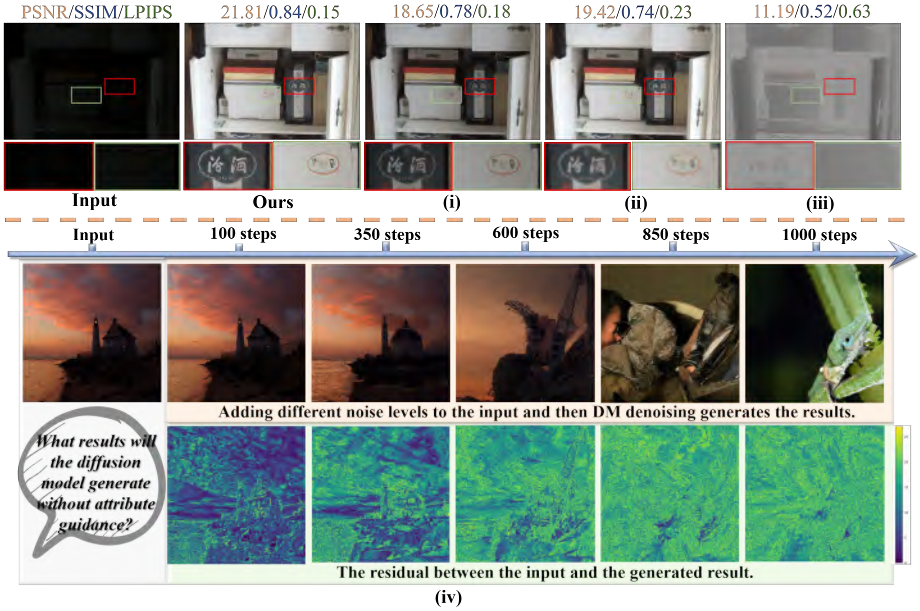

PSNR/SSIM/LPIPS
21.81/0.84/0.15
18.65/0.78/0.18
19.42/0.74/0.23
11.19/0.52/0.63
Input
Ours
(i)
(ii)
(iii)
Input
100 steps
350 steps
600 steps
850 steps
1000 steps
Adding different noise levels to the input and then DM denoising generates the results.
What results will the diffusion model generate without attribute guidance?
The residual between the input and the generated result.
(iv)

Fig. 7: Visual and quantitative results of ablation studies on LOL. The full model achieves the best performance.

Visual Comparisons. For a more comprehensive comparison, we further provide visual comparisons with leading algorithms in Figs 5 and 6. Our observations are twofold: 1) the proposed method distinctly surpasses other approaches in delivering aesthetically superior enhancements in terms of brightness, color fidelity, contrast, and natural appearance, especially under extremely low-light conditions where others falter; and 2) despite supervised approaches like Retinexformer, DiffLL, and CUE exhibiting notable efficacy on LOLv1, SICE and LOLv2-synthetic datasets, their generalization capabilities may be limited as supervised learning is sensitive to the data distribution. For more visual comparisons, please refer to the supplementary material.

4.3 Ablation study

To assess the impact of our approach's key components: attributes, gradient scale $\boldsymbol{s}$, number of gradient steps $\boldsymbol{N}$, and noise addition iteration steps $\boldsymbol{\omega}$, we conduct several ablation studies on the LOLv1 dataset [76].

Impact of the attributes. We undertake ablation studies to verify the effectiveness of the mentioned attributes in Sec. 3.3. Concretely, we have tested the following three variations over the original setting: (i) without the exposure attribute guidance. (ii) without the structure attribute guidance. (iii) without the color attribute guidance. (iv) using only the pre-trained diffusion model without the attribute guidance. Results are shown in Fig. 7. We have the following observations: 1) The removal of exposure attribute guidance limits users' control over

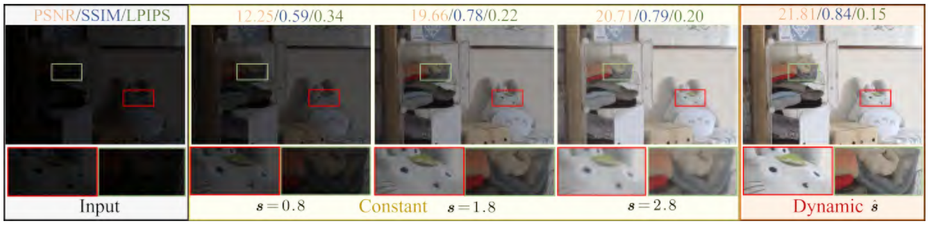
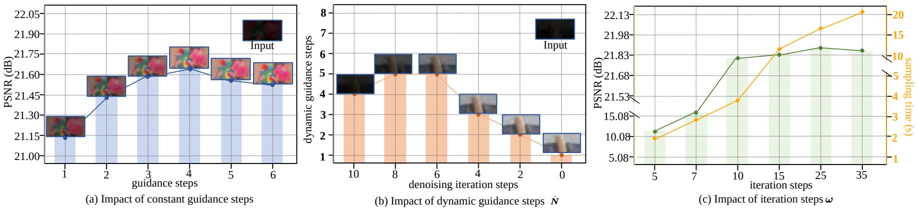


Fig. 8: Ablation studies on the dynamic gradient scale $\hat{\boldsymbol{s}}$. The comparison results verify its effectiveness over the conventional constant guidance scale.



(a) Impact of constant guidance steps

(b) Impact of dynamic guidance steps $\hat{N}$

(c) Impact of iteration steps ω

Fig. 9: Ablation studies on dynamic gradient steps $\hat{\boldsymbol{N}}$ (a-b) and different iteration steps $\boldsymbol{\omega}$ (c). The blue box in (a-b) is the enhanced result.

exposure adjustments. 2) The lack of structure attribute guidance leads to blurring in the structure. 3) The absence of color attribute guidance causes severe color distortions, and the objective measures degrade significantly. 4) Without attribute guidance, relying solely on the pre-trained diffusion model, the LIE task will gradually degenerate into an unconditional image generation task as the level of noise added to the input increases. In contrast, our full model yields clear and natural outputs, validating the efficacy of the introduced attributes.

Effectiveness of Dynamic Guidance Scale. The effectiveness of the dynamic guidance scale $\hat{\boldsymbol{s}}$ is evaluated quantitatively and qualitatively. As illustrated in Fig. 8, without the dynamic guidance scale, although plausible results can be generated, the guidance scale must be manually adjusted for specific scenes, and the fidelity and clarity of the content cannot be guaranteed. In contrast, with the dynamic guidance scale $\hat{\boldsymbol{s}}$ replacing the constant guidance scale $\hat{\boldsymbol{s}}$, high quality and fidelity results can be robustly delivered. The results highlight the indispensable role of our dynamic guidance scale in ensuring high fidelity to the target during the guidance process.

Effectiveness of Dynamic Gradient Steps. The dynamic gradient steps $\hat{\boldsymbol{N}}$ serve to adaptively adjust the strength of guiding the output toward the intended target. As depicted in Fig. 9(a), employing constant gradient steps yields suboptimal results, either under- or over-enhancement, with artifacts and noise. Conversely, in Fig. 9(b), artifacts and noise are progressively removed and finer details are generated. During the early phases of the denoising process, the $\hat{\boldsymbol{N}}$ is larger, while in the later stages, $\hat{\boldsymbol{N}}$ is smaller. Such a phenomenon suggests that the intermediate outputs are laden with noise in the early phases, and hence the gradient step should be increased to effectively steer the outputs towards the intended target. Whereas in the later phases, the gradient step should be decreased to produce refined results.

Impact of Iteration Steps. We explore the influence of the iteration steps $\boldsymbol{\omega}$ of adding noise and removing noise. In Fig. 9(c), we perform different numbers of iterations on the input to generate multiple noisy image outcomes. One can see that enlarging the iteration steps yields limited performance gains but significantly increases the sampling time, especially as the $\boldsymbol{\omega}$ exceeds 10. Consequently, we set the $\boldsymbol{\omega} = 10$ for the trade-off between performance and sampling time.

5 Conclusion and Limitations

This manuscript introduces an Attribute Guidance Diffusion (AGLLDiff) framework to alleviate the challenges in real-world low-light image enhancement (LIE). AGLLDiff innovatively focuses on modeling desired high-quality image attributes such as exposure, structure and color, which do not depend on specific assumptions about the degradation process. This attribute-based guidance facilitates the diffusion sampling process towards achieving high-quality image recovery. Despite outstanding quantitative and qualitative performance achieved in eight challenging LIE benchmarks, there remain areas for improvement, such as accelerating sampling via advanced techniques and exploring more underlying high-quality attributes. Furthermore, future work will extend the application of this framework to various restoration challenges.

References

1. Abdullah-Al-Wadud, M., Kabir, M.H., Dewan, M.A.A., Chae, O.: A dynamic histogram equalization for image contrast enhancement. IEEE transactions on consumer electronics **53**(2), 593–600 (2007) 3
2. Arici, T., Dikbas, S., Altunbasak, Y.: A histogram modification framework and its application for image contrast enhancement. IEEE Transactions on image processing **18**(9), 1921–1935 (2009) 3
3. Bai, J., Dong, Z., Feng, A., Zhang, X., Ye, T., Zhou, K., Shou, M.Z.: Integrating view conditions for image synthesis. arXiv preprint arXiv:2310.16002 (2023) 4
4. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 pirm challenge on perceptual image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018) 11
5. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing **27**(4), 2049–2062 (2018) 10
6. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 12504–12513 (October 2023) 4, 10, 12
7. Cao, S., Chai, W., Hao, S., Wang, G.: Image reference-guided fashion design with structure-aware transfer by diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3524–3528 (2023) 2

8. Cao, S., Chai, W., Hao, S., Zhang, Y., Chen, H., Wang, G.: Difffashion: Reference-based fashion design with structure-aware transfer by diffusion models. IEEE Transactions on Multimedia (2023) 2, 4
9. Chai, W., Guo, X., Wang, G., Lu, Y.: Stablevideo: Text-driven consistency-aware diffusion video editing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 23040–23050 (2023) 4
10. Chefer, H., Alaluf, Y., Vinker, Y., Wolf, L., Cohen-Or, D.: Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models. ACM Transactions on Graphics (TOG) **42**(4), 1–10 (2023) 10
11. Chen, C., Mo, J.: IQA-PyTorch: Pytorch toolbox for image quality assessment (2022) 12
12. Chen, S., Ye, T., Bai, J., Chen, E., Shi, J., Zhu, L.: Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13106–13117 (2023) 4
13. Chen, S., Ye, T., Liu, Y., Bai, J., Chen, H., Lin, Y., Shi, J., Chen, E.: Cplformer: Cross-scale prototype learning transformer for image snow removal. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 4228–4239 (2023) 4
14. Chen, S., Ye, T., Liu, Y., Chen, E., Shi, J., Zhou, J.: Snowformer: Scale-aware transformer via context interaction for single image desnowing. arXiv preprint arXiv:2208.09703 **2** (2022) 4
15. Chen, S., Ye, T., Liu, Y., Liao, T., Ye, Y., Chen, E.: Msp-former: Multi-scale projection transformer for single image desnowing. arXiv preprint arXiv:2207.05621 (2022) 4
16. Chen, S., Ye, T., Shi, J., Liu, Y., Jiang, J., Chen, E., Chen, P.: Dehrformer: Real-time transformer for depth estimation and haze removal from varicolored haze scenes. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023) 4
17. Cheng, H.D., Shi, X.: A simple and effective histogram equalization approach to image enhancement. Digital signal processing **14**(2), 158–170 (2004) 3
18. Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S.: Ilvr: Conditioning method for denoising diffusion probabilistic models. arXiv preprint arXiv:2108.02938 (2021) 4
19. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A.: Generative adversarial networks: An overview. IEEE signal processing magazine **35**(1), 53–65 (2018) 2
20. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023) 2
21. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009) 10
22. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021) 3, 5, 9, 10
23. Fei, B., Lyu, Z., Pan, L., Zhang, J., Yang, W., Luo, T., Zhang, B., Dai, B.: Generative diffusion prior for unified image restoration and enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9935–9946 (2023) 2, 4, 10, 12
24. Fu, Z., Yang, Y., Tu, X., Huang, Y., Ding, X., Ma, K.K.: 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. pp. 22252–22261 (2023) 3, 4, 8, 10, 11, 12

25. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020) 2
26. Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R.: Zero-reference deep curve estimation for low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1777–1786 (2020) 3, 4, 10, 12
27. Guo, J., Chai, W., Deng, J., Huang, H.W., Ye, T., Xu, Y., Zhang, J., Hwang, J.N., Wang, G.: Versat2i: Improving text-to-image models with versatile reward. arXiv preprint arXiv:2403.18493 (2024) 4
28. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing **26**(2), 982–993 (2017). https://doi.org/10.1109/TIP.2016.2639450 3
29. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. IEEE Transactions on Multimedia **22**(12), 3025–3038 (2020). https://doi.org/10.1109/TMM.2020.2969790 3, 10, 12
30. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020) 2, 4, 5
31. Hou, J., Zhu, Z., Hou, J., Liu, H., Zeng, H., Yuan, H.: Global structure-aware diffusion process for low-light image enhancement. arXiv preprint arXiv:2310.17577 (2023) 2, 4
32. Huang, J., Meng, G., Wang, Y., Lin, Y., Huang, Y., Ding, X.: Dp-innet: Dual-path implicit neural network for spatial and spectral features fusion in pan-sharpening. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 268–279. Springer (2023) 4
33. Huang, S.C., Cheng, F.C., Chiu, Y.S.: Efficient contrast enhancement using adaptive gamma correction with weighting distribution. IEEE transactions on image processing **22**(3), 1032–1041 (2012) 3
34. Jiang, H., Luo, A., Fan, H., Han, S., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM Transactions on Graphics (TOG) **42**(6), 1–14 (2023) 2, 4, 10, 12
35. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing **30**, 2340–2349 (2021) 1, 10, 12
36. Jiang, Z., Zhou, Z., Li, L., Chai, W., Yang, C.Y., Hwang, J.N.: Back to optimization: Diffusion-based zero-shot 3d human pose estimation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6142–6152 (2024) 2
37. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. Advances in Neural Information Processing Systems **35**, 23593–23606 (2022) 4
38. Kingma, D., Salimans, T., Poole, B., Ho, J.: Variational diffusion models. Advances in neural information processing systems **34**, 21696–21707 (2021) 2, 4
39. Land, E.H., McCann, J.J.: Lightness and retinex theory. Josa **61**(1), 1–11 (1971) 8
40. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation. In: 2012 19th IEEE international conference on image processing. pp. 965–968. IEEE (2012) 3, 11
41. Li, C., Guo, C.L., Zhou, M., Liang, Z., Zhou, S., Feng, R., Loy, C.C.: Embedding-fourier for ultra-high-definition low-light image enhancement. In: ICLR (2023) 3, 8, 11

42. Li, C., Guo, C., Feng, R., Zhou, S., Loy, C.C.: Cudi: Curve distillation for efficient and controllable exposure adjustment. arXiv preprint arXiv:2207.14273 (2022) 7
43. Li, C., Guo, C., Loy, C.C.: Learning to enhance low-light image via zero-reference deep curve estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(8), 4225–4238 (2021) 3, 10, 12
44. Li, M., Liu, J., Yang, W., Sun, X., Guo, Z.: Structure-revealing low-light image enhancement via robust retinex model. IEEE Transactions on Image Processing **27**(6), 2828–2841 (2018). https://doi.org/10.1109/TIP.2018.2810539 11
45. Liang, Z., Li, C., Zhou, S., Feng, R., Loy, C.C.: Iterative prompt learning for unsupervised backlit image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8094–8103 (2023) 4, 10, 12
46. Lim, S., Kim, W.: Dslr: Deep stacked laplacian restorer for low-light image enhancement. IEEE Transactions on Multimedia **23**, 4272–4284 (2020) 10, 12
47. Lin, Y., Fu, Z., Meng, G., Wang, Y., Dong, Y., Fan, L., Yu, H., Ding, X.: Domain-irrelevant feature learning for generalizable pan-sharpening. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 3287–3296 (2023) 4
48. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10561–10570 (2021) 10, 12
49. Liu, Y., Liu, F., Ke, Z., Zhao, N., Lau, R.W.: Diff-plugin: Revitalizing details for diffusion-based low-level tasks. In: CVPR (2024) 2
50. Liu, Y., Yan, Z., Chen, S., Ye, T., Ren, W., Chen, E.: Nighthazeformer: Single nighttime haze removal using prior query transformer. arXiv preprint arXiv:2305.09533 (2023) 3
51. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition **61**, 650–662 (2017) 3
52. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: Low-light image/video enhancement using cnns. In: British Machine Vision Conference (BMVC). pp. 1–13 (2018) 3, 10, 12
53. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE Transactions on Image Processing **24**(11), 3345–3356 (2015) 11
54. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5637–5646 (2022) 3, 10, 12
55. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing **21**(12), 4695–4708 (2012) 11
56. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters **20**(3), 209–212 (2012) 11
57. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021) 5
58. Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C.C., Luo, P.: Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 7474–7489 (2021) 1
59. Rahman, S., Rahman, M.M., Abdullah-Al-Wadud, M., Al-Quaderi, G.D., Shoyaib, M.: An adaptive gamma correction for image enhancement. EURASIP Journal on Image and Video Processing **2016**(1), 1–13 (2016) 3
60. Ren, Y., Ying, Z., Li, T.H., Li, G.: Lecarm: Low-light image enhancement using the camera response model. IEEE Transactions on Circuits and Systems for Video Technology **29**(4), 968–981 (2018) 3, 10, 12

61. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022) 3
62. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(4), 4713–4726 (2022) 4
63. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020) 2, 4
64. Vonikakis, V., Kouskouridas, R., Gasteratos, A.: On the evaluation of illumination compensation algorithms. Multimedia Tools and Applications **77**, 9211–9231 (2017) 11
65. Voynov, A., Aberman, K., Cohen-Or, D.: Sketch-guided text-to-image diffusion models. In: ACM SIGGRAPH 2023 Conference Proceedings. pp. 1–11 (2023) 10
66. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE transactions on image processing **22**(9), 3538–3548 (2013) 11
67. Wang, T., Zhang, K., Shao, Z., Luo, W., Stenger, B., Kim, T.K., Liu, W., Li, H.: Lldiffusion: Learning degradation representations in diffusion models for low-light image enhancement. arXiv preprint arXiv:2307.14659 (2023) 2
68. Wang, T., Zhang, K., Shao, Z., Luo, W., Stenger, B., Kim, T.K., Liu, W., Li, H.: Lldiffusion: Learning degradation representations in diffusion models for low-light image enhancement. arXiv preprint arXiv:2307.14659 (2023) 4
69. Wang, Y., He, X., Dong, Y., Lin, Y., Huang, Y., Ding, X.: Cross-modality interaction network for pan-sharpening. IEEE Transactions on Geoscience and Remote Sensing (2024) 4
70. Wang, Y., Lin, Y., Meng, G., Fu, Z., Dong, Y., Fan, L., Yu, H., Ding, X., Huang, Y.: Learning high-frequency feature enhancement and alignment for pan-sharpening. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 358–367 (2023) 4
71. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. arXiv preprint arXiv:2212.00490 (2022) 4
72. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.P., Kot, A.: Low-light image enhancement with normalizing flow. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 2604–2612 (2022) 11
73. Wang, Z.G., Liang, Z.H., Liu, C.L.: A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for pdp. Displays **30**(3), 133–139 (2009) 3
74. Wang, Z., Zhang, Z., Zhang, X., Zheng, H., Zhou, M., Zhang, Y., Wang, Y.: Dr2: Diffusion-based robust degradation remover for blind face restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1704–1713 (2023) 4
75. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004) 11
76. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference (BMVC). pp. 1–12 (2018) 3, 10, 11, 12, 13
77. Wu, Y., Wang, G., Wang, Z., Yang, Y., Li, T., Wang, P., Li, C., Shen, H.T.: Reco-diff: Explore retinex-based condition strategy in diffusion model for low-light image enhancement. arXiv preprint arXiv:2312.12826 (2023) 2, 4

78. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17714–17724 (2022) 3
79. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.H.: Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys **56**(4), 1–39 (2023) 2
80. Yang, P., Zhou, S., Tao, Q., Loy, C.C.: Pgdiff: Guiding diffusion models for versatile face restoration via partial guidance. Advances in Neural Information Processing Systems **36** (2024) 4
81. Yang, S., Ding, M., Wu, Y., Li, Z., Zhang, J.: Implicit neural representation for cooperative low-light image enhancement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12918–12927 (2023) 3, 4, 10, 12
82. Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J.: From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3060–3069 (2020) 10, 12
83. Yang, W., Wang, W., Huang, H., Wang, S., Liu, J.: Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing **30**, 2072–2086 (2021) 10
84. Ye, T., Chen, S., Bai, J., Shi, J., Xue, C., Jiang, J., Yin, J., Chen, E., Liu, Y.: Adverse weather removal with codebook priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12653–12664 (2023) 4
85. Ye, T., Chen, S., Chai, W., Xing, Z., Qin, J., Lin, G., Zhu, L.: Learning diffusion texture priors for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2524–2534 (2024) 4
86. Ye, T., Chen, S., Liu, Y., Chai, W., Bai, J., Zou, W., Zhang, Y., Jiang, M., Chen, E., Xue, C.: Sequential affinity learning for video restoration. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 4147–4156 (2023) 4
87. Ye, T., Zhang, Y., Jiang, M., Chen, L., Liu, Y., Chen, S., Chen, E.: Perceiving and modeling density for image dehazing. In: European Conference on Computer Vision. pp. 130–145. Springer (2022) 4
88. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12302–12311 (2023) 2, 4
89. Yin, Y., Xu, D., Tan, C., Liu, P., Zhao, Y., Wei, Y.: Cle diffusion: Controllable light enhancement diffusion model. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 8145–8156 (2023) 2, 4
90. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023) 3
91. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 586–595 (2018) 11
92. Zheng, N., Zhou, M., Dong, Y., Rui, X., Huang, J., Li, C., Zhao, F.: Empowering low-light image enhancer through customized learnable priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12559–12569 (2023) 3, 10, 12
93. Zhou, D., Yang, Z., Yang, Y.: Pyramid diffusion models for low-light image enhancement. arXiv preprint arXiv:2305.10028 (2023) 2, 4, 10, 12

94. Zhu, A., Zhang, L., Shen, Y., Ma, Y., Zhao, S., Zhou, Y.: Zero-shot restoration of underexposed images via robust retinex decomposition. In: 2020 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2020) 10, 12
95. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017) 1
96. Zhu, Y., Zhang, K., Liang, J., Cao, J., Wen, B., Timofte, R., Van Gool, L.: Denoising diffusion models for plug-and-play image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1219–1229 (2023) 4
97. Zou, W., Gao, H., Ye, T., Chen, L., Yang, W., Huang, S., Chen, H., Chen, S.: Vqcnir: Clearer night image restoration with vector-quantized codebook. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 7873–7881 (2024) 3