LightenDiffusion: Unsupervised Low-Light Image Enhancement with Latent-Retinex Diffusion Models

Hai Jiang[1,5], Ao Luo[2,5], Xiaohong Liu[4],
Songchen Han[1], and Shuaicheng Liu[3,5,†]

[1] Sichuan University, [2] Southwest Jiaotong University,
[3] University of Electronic Science and Technology of China,
[4] Shanghai Jiao Tong University, [5] Megvii Technology
{jianghai@stu.,hansongchen@}scu.edu.cn, aoluo@swjtu.edu.cn,
xiaohongliu@sjtu.edu.cn, liushuaicheng@uestc.edu.cn
[†] Corresponding Author

Abstract. In this paper, we propose a diffusion-based unsupervised framework that incorporates physically explainable Retinex theory with diffusion models for low-light image enhancement, named LightenDiffusion. Specifically, we present a content-transfer decomposition network that performs Retinex decomposition within the latent space instead of image space as in previous approaches, enabling the encoded features of unpaired low-light and normal-light images to be decomposed into content-rich reflectance maps and content-free illumination maps. Subsequently, the reflectance map of the low-light image and the illumination map of the normal-light image are taken as input to the diffusion model for unsupervised restoration with the guidance of the low-light feature, where a self-constrained consistency loss is further proposed to eliminate the interference of normal-light content on the restored results to improve overall visual quality. Extensive experiments on publicly available real-world benchmarks show that the proposed LightenDiffusion outperforms state-of-the-art unsupervised competitors and is comparable to supervised methods while being more generalizable to various scenes. Our code is available at https://github.com/JianghaiSCU/LightenDiffusion.

Keywords: Image restoration · Low-light image enhancement · Diffusion models · Retinex theroy

1 Introduction

Images captured under weakly illuminated conditions suffer from various degradations such as poor visibility and noise, which leads to adverse impacts on the performance of downstream vision tasks [33,65]. To transform low-light images into high-quality images, numerous works have been proposed in the past decades. Traditional methods [9,14,42,44,53] mainly adopt hand-crafted priors, such as histogram equalization (HE) [2] and Retinex theory [27], to improve
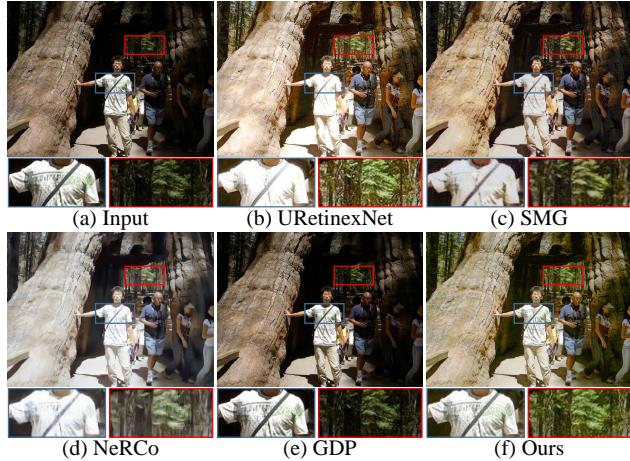
Fig. 1: Visual comparisons of our method with recent state-of-the-art supervised and unsupervised LLIE methods UReinexNet [59], SMG [64], NeRCo [67], and GDP [8]. Previous methods appear incorrect exposure, color distortion, blurred details, or noise amplification to degrade visual quality, while our method properly improves global and local contrast, presents a vivid color, and avoids introducing artifacts.

contrast and restore details. However, it is difficult to adopt a suitable prior for various illumination conditions since low-light image enhancement (LLIE) is an ill-posed problem, thus limiting the practical application of these methods.

These issues have been partially resolved with the development of deep learning, where learning-based methods [5,10,12,13,32,34,54,58,60,63,64] can directly learn the mapping from low-light images to normal-light images through powerful network architectures and sophisticated learning strategies, which present more robustness than traditional methods. While learning-based methods achieve remarkable progress in LLIE, they often suffer from the overfitting problem and struggle with poor generalization ability, resulting in outcomes with unsatisfactory visual fidelity. As shown in Fig. 1(b)-(d), previous state-of-the-art supervised methods URetinexNet [59] and SMG [64], as well as unsupervised method NeRCo [67] present incorrect overexposure, color distortion, blurred details or noise amplification in the highlighted regions.

Recently, generative model-based methods [24,66,75] have emerged for LLIE as promising approaches to obtain better perceptual quality, in which diffusion models [19,49] have gained attention for their impressive generative ability and being free from instability and mode-collapse problems present in previous generative models such as generative adversarial networks (GANs) and variational autoencoders (VAEs). Most diffusion-based methods [20,22,43,47,48,71,76] utilize large-scale paired data with conditional mechanism [6] for supervised learning, which enable favorable contrast enhancement and details reconstruction, while it is challenging to collect paired distorted/sharp images in the real world. To leverage the label-free characteristic of unsupervised learning to improve the gen-

eralization of diffusion models, some methods [8, 25, 35, 55, 78] employ zero-shot solutions that utilize well-established priors from pre-trained diffusion models for restoration without training from scratch. However, these methods are limited by the known degradation modes and thus tend to perform poorly in real-world scenes where distortions are diverse and unknown. As shown in Fig. 1(e), the zero-shot-based method GDP [8] produces an under-enhancement result.

To this end, we propose a diffusion-based learnable unsupervised framework, dubbed LightenDiffusion, which incorporates physically interpretable Retinex theory with diffusion models to learn degradation modes of various scenes. It accomplishes this by training on extensive unpaired real-world data, ultimately achieving visually favorable LLIE. Specifically, we first convert the unpaired low-light and normal-light images into latent space, where the encoded features are decomposed into content-rich reflectance maps that contain abundant content-related details and content-free illumination maps that only represent the lighting conditions through the proposed content-transfer decomposition network. Subsequently, the reflectance map of the low-light feature and the illumination map of the normal-light feature serve as input to the diffusion model for restoration with the guidance of the low-light feature. Moreover, the distribution learned by the diffusion model may be disrupted once the estimated normal-light illumination map still preserves certain content information, leading to the restored result being interfered by the normal-light image content. Therefore, we propose a self-constrained consistency loss to promote the diffusion model to reconstruct images with the same intrinsic content information as input low-light images. As shown in Fig. 1(f), our method properly improves global and local contrast, prevents overcorrection on the well-exposed region, and avoids artifacts or noise amplification. Extensive experiments show that our method outperforms existing state-of-the-art competitors quantitatively and visually. The application for low-light face detection also reveals the potential practical values of our method.

Our contributions can be summarized as follows:

- We propose a diffusion-based framework, termed LightenDiffusion, that leverages the advantages of Retinex theory and the generative ability of diffusion models for unsupervised low-light image enhancement, with a self-constrained consistency loss further proposed to improve visual quality.
- We propose a content-transfer decomposition network that performs decomposition in the latent space, aiming to obtain content-rich reflectance maps and content-free illumination maps to promote unsupervised restoration.
- Extensive experiments demonstrate that our method outperforms existing state-of-the-art unsupervised competitors while being comparable and having better generalization abilities than supervised methods.

2   Related Work

2.1   Low-Light Image Enhancement

Numerous works have been proposed to transform poorly illuminated images into visually pleasant normal-light images. Traditional methods depend on hand-

crafted optimization rules such as Histogram Equalization (HE) [2] and Retinex theory [27]. HE-based methods [42,44] aim to change the histogram distribution of the image to improve the contrast. Retinex-based methods [9,14] first decompose an image into a reflectance map and an illumination map, with the visual quality being improved by changing the dynamic range of the illumination map.

Recently, learning-based methods have achieved remarkable results in the LLIE task and show more robustness than traditional methods, which can be mainly categorized as supervised, semi-supervised, and unsupervised. The former [13,23,34,54,60,63,64,72] leverage powerful network architectures to learn mappings from low-light images to normal-light ones in an end-to-end manner. Some approaches [5,15,58,59,74] combine Retinex theory with deep networks to establish learnable decomposition and adjustment frameworks. However, supervised methods rely on large-scale paired datasets for training and thus suffer from limited generalization ability. To address these issues, unsupervised methods [10,12,24,32,40,67] utilize their characteristics of not requiring paired data to solve the LLIE by employing adversarial learning, curve estimation, or neural architecture search with better generalization in real-world scenes. Semi-supervised methods [29,68] combine the advantages of supervised and unsupervised learning to achieve stable training while maintaining better generalization capability.

2.2   Diffusion-based Image Restoration

With the development of diffusion models (DMs) in low-level vision [19,31,36, 45,49,62,70,77], many works have been conducted to explore their performance in image restoration tasks, such as super-resolution [11,48], inpainting [47,61], weather removal [37,38,43], and low-light image enhancement [18,20,22,56,71, 76]. Most methods utilize the conditional mechanism [6] to train diffusion models from scratch with paired data, where degraded images serve as guidance in the diffusion processes. In contrast, some methods [8,25,35,55,78] employ zero-shot strategies using pre-trained diffusion models to restore degraded images without reference images directly. They leverage the priors from the pre-trained models for restoration, rather than deriving the capability from the training datasets. Although zero-shot approaches provide an attractive alternative, their performance is hampered by the pre-trained models, leading to the restored results with unsatisfactory visual quality. In this paper, we propose to incorporate the physically explainable Retinex theory with diffusion models to achieve visually satisfactory LLIE in an unsupervised manner.

3   Methodology

3.1   Overview

The overall pipeline of our proposed framework is illustrated in Fig. 2. Given an unpaired low-light image $I_{low} \in \mathbb{R}^{H \times W \times 3}$ and normal-light image $I_{high} \in \mathbb{R}^{H \times W \times 3}$, we first employ an encoder $\mathcal{E}(\cdot)$, which consists of k cascaded residual blocks where each block downsamples the input by a scale of 2 using a
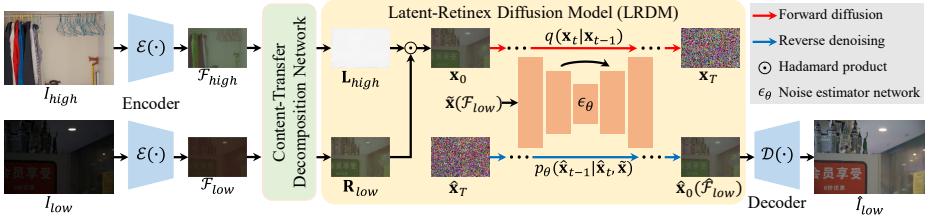
Fig. 2: The overall pipeline of our proposed framework. We first employ an encoder $\mathcal{E}(\cdot)$ to convert the unpaired low-light image I_{low} and normal-light image I_{high} into latent space denoted as $\mathcal{F}_{low}$ and $\mathcal{F}_{high}$. The encoded features are sent to the proposed content-transfer decomposition network (CTDN) to generate content-rich reflectance maps denoted as $\mathbf{R}_{low}$ and $\mathbf{R}_{high}$ and content-free illumination maps as $\mathbf{L}_{low}$ and $\mathbf{L}_{high}$. Then, the reflectance map of the low-light image $\mathbf{R}_{low}$ and the illumination of the normal-light image $\mathbf{L}_{high}$ are taken as the input of the diffusion model to perform the forward diffusion process. Finally, we perform the reverse denoising process to gradually transform the randomly sampled Gaussian noise $\hat{\mathbf{x}}_T$ into the restored feature $\hat{\mathcal{F}}_{low}$ with the guidance of the low-light feature $\mathcal{F}_{low}$ denoted as $\tilde{\mathbf{x}}$, and subsequently send it to a decoder $\mathcal{D}(\cdot)$ to produce the final result $\hat{I}_{low}$.

max-pooling layer, to transform the input images into latent space denoted as $\mathcal{F}_{low} \in \mathbb{R}^{\frac{H}{2^k} \times \frac{W}{2^k} \times C}$ and $\mathcal{F}_{high} \in \mathbb{R}^{\frac{H}{2^k} \times \frac{W}{2^k} \times C}$. Then, we design a content-transfer decomposition network (CTDN) to decompose the features into content-rich reflectance maps $\mathbf{R}_{low}$ and $\mathbf{R}_{high}$ and content-free illumination maps $\mathbf{L}_{low}$ and $\mathbf{L}_{high}$. Subsequently, the $\mathbf{R}_{low}$ and the $\mathbf{L}_{high}$ serve as input for the diffusion model with the guidance of the low-light feature to generate the restored feature $\hat{\mathcal{F}}_{low}$. Finally, the restored feature will be sent to a decoder $\mathcal{D}(\cdot)$ for reconstruction to produce the final restored image $\hat{I}_{low}$.

3.2 Content-Transfer Decomposition Network

The Retinex theory [27] assumes that an image I can be decomposed into a reflectance map $\mathbf{R}$ and an illumination map $\mathbf{L}$ as:

$$I = \mathbf{R} \odot \mathbf{L} \tag{1}$$

where $\odot$ denotes Hadamard product operation. $\mathbf{R}$ represents the inherent content information that should be consistent under diverse illumination conditions, while $\mathbf{L}$ indicates the contrast and brightness information that should be local smoothness. However, existing methods [5,10,58,59,71,74] typically perform decomposition in the image space to obtain the above components, which results in the content information not being fully decomposed into the reflectance map and partially retained in the illumination map, as shown in Fig. 3(a).

To alleviate this issue, we introduce a content-transfer decomposition network (CTDN) that performs decomposition within the latent space. By encoding the content information in this latent space, the CTDN facilitates the generation of
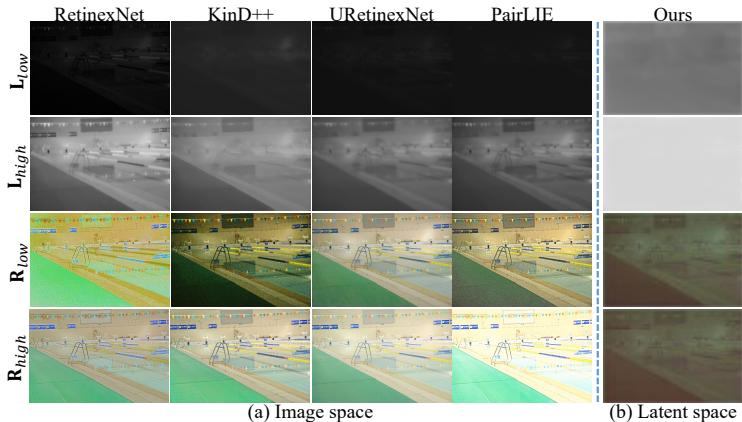
Fig. 3: Illustration of the decomposition results obtained by different methods. (a) shows the results of previous methods, i.e., RetinexNet [58], KinD++ [74], URetinexNet [59], and PairLIE [10], that perform decomposition in image space. (b) presents the results of our CTDN that performs decomposition in latent space. Our method can generate content-rich reflectance maps and content-free illumination maps.

reflectance maps containing abundant content-related details and illumination maps that remain unaffected by content-related influences. As shown in Fig. 4, we first estimate the initial reflectance and illumination maps following [14] as:

$$\tilde{\mathbf{L}}(x) = \max_{c \in [0,C)} \mathcal{F}^c(x), \tilde{\mathbf{R}}(x) = \mathcal{F}(x)/(\tilde{\mathbf{L}}(x) + \tau), \tag{2}$$

for each pixel x, where τ is a small constant to avoid zero denominator. The estimated maps are refined through two branches, in which we first employ several convolutional blocks to obtain the embedded features as $\mathbf{L}' = \text{Convs}(\tilde{\mathbf{L}})$, $\mathbf{R}' = \text{Convs}(\tilde{\mathbf{R}})$. Subsequently, we utilize a cross-attention (CA) [21] module to leverage the illumination map to reinforce the content information in the reflectance map as $\mathbf{R}'' = \text{CA}(\mathbf{R}', \mathbf{L}')$. Moreover, a self-attention module (SA) [50] is adopted to further extract content information in the illumination map, denoted as $\mathbf{L}'' = \text{SA}(\mathbf{L}')$, and complement it to the reflectance map. The final output reflectance map $\mathbf{R}$ and illumination map $\mathbf{L}$ can be expressed as $\mathbf{R} = \text{Convs}(\mathbf{R}'' + \mathbf{L}'')$ and $\mathbf{L} = \text{Convs}(\mathbf{L}' - \mathbf{L}'')$. As shown in Fig. 3(b), our CTDN can generate content-rich reflectance maps that fully represent the intrinsic information of the image, and content-free illumination maps that only reveal the lighting conditions.

3.3   Latent-Retinex Diffusion Models

One straightforward way to obtain the enhanced feature in the ideal case is to multiply the reflectance map of the low-light feature with the illumination map of the normal-light image as $\hat{\mathcal{F}}_{low} = \mathbf{R}_{low} \odot \mathbf{L}_{high}$. However, there are two challenges with the above approach: 1) Retinex decomposition inevitably
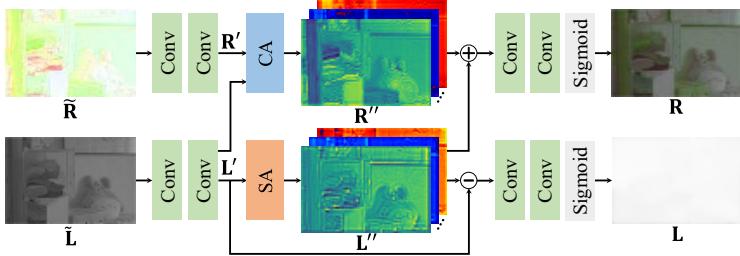
Fig. 4: The detailed architecture of our proposed CTDN.

encounters information loss; 2) the restored image would present artifacts once the illumination map of the reference normal-light image still contains stubborn content information. Although our CTDN is generally effective in most scenes, there may be challenging cases where the accuracy of the estimated illumination map is compromised. To address these problems, we propose a Latent-Retinex diffusion model (LRDM) that leverages the generative ability of diffusion models to compensate for content loss and eliminate potential unexpected artifacts. Our approach follows standard diffusion models [6, 19, 49] that perform forward diffusion and reverse denoising processes to generate restored results.

Forward Diffusion. Given the decomposition components of unpaired images, we take the reflectance map of the low-light image $\mathbf{R}_{low}$ and the illumination map of the normal-light image $\mathbf{L}_{high}$ as input, denoted as $\mathbf{x}_0 = \mathbf{R}_{low} \odot \mathbf{L}_{high}$, to perform the forward diffusion process, and uses a pre-defined variance schedule $\{\beta_1, \beta_2, \ldots, \beta_T\}$ to progressively transform $\mathbf{x}_0$ into Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ through T steps, which can be formulated as:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \tag{3}$$

where $\mathbf{x}_t$ indicates the noisy data at time-step $t \in [0, T]$. By utilizing parameter renormalization, we can merge and refine multiple Gaussian distributions to obtain the $\mathbf{x}_t$ directly from the input $\mathbf{x}_0$ and simplify Eq.(4) into a closed form as $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}_t$, where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Reverse Denoising. By utilizing the editing and data synthesis capabilities offered by conditional diffusion models [6], we aim to gradually denoise a randomly sampled Gaussian noise $\hat{\mathbf{x}}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ into a sharp result $\hat{\mathbf{x}}_0$ with the guidance of the encoded feature $\mathcal{F}_{low} = \mathcal{E}(I_{low})$ of the low-light image denoted as $\tilde{\mathbf{x}}$, which facilitates resulting in high fidelity of restored results to the distribution conditioned on $\tilde{\mathbf{x}}$. The reverse denoising process can be formulated as:

$$p_\theta(\hat{\mathbf{x}}_{t-1} \mid \hat{\mathbf{x}}_t, \tilde{\mathbf{x}}) = \mathcal{N}(\hat{\mathbf{x}}_{t-1}; \boldsymbol{\mu}_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}, t), \sigma_t^2\mathbf{I}), \tag{4}$$

where $\sigma_t^2 = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t}\beta_t$ is the variance and $\boldsymbol{\mu}_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}, t) = \frac{1}{\sqrt{\alpha_t}}(\hat{\mathbf{x}}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}, t))$ is the mean value.

In the training phase, the objective of the diffusion model is to optimize the parameters θ of the network $\boldsymbol{\epsilon}_\theta$ to promote the estimated noise vector $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, \tilde{\mathbf{x}}, t)$

Algorithm 1: LRDM training

input : The decomposition results $\mathbf{R}_{low}$ and $\mathbf{L}_{high}$, low-light feature $\mathcal{F}_{low}$,
time step T, and sampling step S.

$\mathbf{x}_0 = \mathbf{R}_{low} \odot \mathbf{L}_{high}$, $\tilde{\mathbf{x}} = \mathcal{F}_{low}$

while *Not converged* **do**

    $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $t \sim \text{Uniform}\{1, \cdots, T\}$

    Perform gradient descent steps on $\nabla_\theta \|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}_t, \tilde{\mathbf{x}}, t)\|^2$

    $\hat{\mathbf{x}}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$

    for $i = S : 1$ **do**

        $t = (i - 1) \cdot T/S + 1$

        $t_{\text{next}} = (i - 2) \cdot T/S + 1$ if $i > 1$, else 0

        $\hat{\mathbf{x}}_t \leftarrow \sqrt{\bar{\alpha}_{t_{\text{next}}}}(\frac{\hat{\mathbf{x}}_t - \sqrt{1-\bar{\alpha}_t}\cdot\boldsymbol{\epsilon}_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}, t)}{\sqrt{\bar{\alpha}_t}}) + \sqrt{1 - \bar{\alpha}_{t_{\text{next}}}} \cdot \boldsymbol{\epsilon}_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}, t)$

    end

    Perform gradient descent steps on $\nabla_\theta \|\mathbf{R}_{low} \odot \mathbf{L}_{low}^\gamma - \hat{\mathbf{x}}_0\|^2$

end

output: θ

---

close to Gaussian noise like [19], which is formulated as:

$$\mathcal{L}_{diff} = \|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, \tilde{\mathbf{x}}, t)\|_2. \tag{5}$$

During inference, we obtain the restored feature $\hat{\mathcal{F}}_{low}$ from the distribution learned by the diffusion model through reverse denoising process with implicit sampling strategy [49], and subsequently send it to the decoder to produce the final result $\hat{I}_{low}$. However, as mentioned above, the input $\mathbf{x}_0$ would present artifacts once the estimated illumination map still contains content information, which may affect the learned distribution and result in the $\hat{\mathcal{F}}_{low}$ being disrupted.

Therefore, we propose a self-constrained consistency loss $\mathcal{L}_{scc}$ to enable the restored feature to share the same intrinsic information as the input low-light image. Specifically, we first perform the reverse denoising process in the training phase following [20,22,76] to generate the restored feature and construct a pseudo label $\tilde{\mathcal{F}}_{low}$ from decomposition results of the low-light image as a reference based on traditional Gamma correction approaches as $\tilde{\mathcal{F}}_{low} = \mathbf{R}_{low} \odot \mathbf{L}_{low}^\gamma$, where γ is the illumination correction factor. Thus, the $\mathcal{L}_{scc}$ aims to constrain the feature similarity to prompt the diffusion model to reconstruct $\hat{I}_{low}$ as:

$$\mathcal{L}_{scc} = \|\tilde{\mathcal{F}}_{low} - \hat{\mathcal{F}}_{low}\|_1. \tag{6}$$

Overall, the training strategy of our LRDM is summarized in Alg. 1 and the objective function used for optimization is rewritten as $\mathcal{L} = \mathcal{L}_{diff} + \lambda_1 \mathcal{L}_{scc}$.

3.4   Network Training

Our approach adopts a two-stage strategy for network training. In the first stage, we follow [10] that utilizes paired low-quality images, denoted as I_{low}^1 and I_{low}^2, from the SICE dataset [3] to optimize the encoder $\mathcal{E}(\cdot)$, CTDN, and decoder $\mathcal{D}(\cdot)$,

while freezing the parameters of the diffusion model. The encoder and decoder are optimized with the content loss $\mathcal{L}_{con}$ as:

$$\mathcal{L}_{con} = \sum_{i=1}^{2} \|I_{low}^i - \mathcal{D}(\mathcal{E}(I_{low}^i))\|_2. \tag{7}$$

The CTDN is optimized with the decomposition loss $\mathcal{L}_{dec}$ as [58, 71, 74] that consist of the reconstruction loss $\mathcal{L}_{rec}$, the reflectance consistency loss $\mathcal{L}_{ref}$, and the illumination smoothing loss $\mathcal{L}_{ill}$. The $\mathcal{L}_{rec}$ aims to guarantee the decomposed components can reconstruct the input features, which is expressed as:

$$\mathcal{L}_{rec} = \sum_{i=1}^{2} \sum_{j=1}^{2} \|\mathcal{F}_{low}^j - \mathbf{R}_{low}^i \odot \mathbf{L}_{low}^j\|_1. \tag{8}$$

The $\mathcal{L}_{ref}$ aims to enforce the network to produce invariant reflectance maps and the $\mathcal{L}_{ill}$ is adopted to guarantee the illumination map to be local smoothness, which can be expressed respectively as:

$$\mathcal{L}_{ref} = \|\mathbf{R}_{low}^1 - \mathbf{R}_{low}^2\|_1, \mathcal{L}_{ill} = \sum_{i=1}^{2} \|\nabla \mathbf{L}_{low}^i \cdot \exp(-\lambda_g \nabla \mathbf{R}_{low}^i)\|_2, \tag{9}$$

where ∇ denotes the horizontal and vertical gradients, and λ_g is the coefficient to balance the perceived strength of the structure. The overall decomposition loss used to optimize the CTDN is formulated as $\mathcal{L}_{dec} = \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{ref} + \lambda_3 \mathcal{L}_{ill}$.

In the second stage, we collect ~180k unpaired low/normal-light image pairs to optimize the diffusion model while freezing the parameters of other modules.

4   Experiments

4.1   Experimental Settings

Implementation Details. We implement the proposed method with PyTorch on four NVIDIA RTX 2080Ti GPUs, where the batch size and patch size are set to 12 and 512 ×512. The networks can be converged after training in two stages with 1×10^5 and 4×10^5 iterations, respectively. We employ the Adam optimizer [26] for optimization with the initial learning rate set to 1×10^{-4} in the first stage and decays by a factor of 0.8 while reinitializing it to a fixed value of 2×10^{-5} in the second stage. The feature downsampling scale k and the illumination correction factor γ are set to 3 and 0.2, respectively. The hyperparameters λ_1, λ_2, λ_3, and λ_g are empirically set to 0.01, 0.1, 0.01, and 10, respectively. For our LRDM, the U-Net [46] architecture is adopted as the noise estimator network with the time step T and sampling step S set to 1000 and 20 for the forward diffusion and reverse denoising processes, respectively.

Datasets and Metrics. To evaluate the performance of the proposed method, we conduct experiments on the test sets of two paired datasets that contain

Table 1: Quantitative comparisons on the paired LOL [58] and LSRW [16] datasets, and unpaired DICM [28], NPE [53], and VV [51] datasets. The best results are highlighted in **bold**. 'T', 'SL', 'SSL', and 'UL' indicate that the methods belong to traditional, supervised, semi-supervised, and unsupervised methods, respectively.

Type	Method	LOL [58]			LSRW [16]			DICM [28]		NPE [53]		VV [51]	
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE ↓	PI ↓	NIQE ↓	PI ↓	NIQE ↓	PI ↓
T	LIME [14]	17.546	0.531	0.290	17.342	0.520	0.416	4.476	4.216	4.170	3.789	3.713	3.335
	SDDLLE [17]	13.342	0.634	0.261	14.708	0.486	0.382	4.581	3.828	4.179	3.315	4.274	3.382
	CDEF [30]	16.335	0.585	0.351	16.758	0.465	0.314	4.142	4.242	3.862	2.910	5.051	3.272
	BrainRetinex [4]	11.063	0.475	0.327	12.506	0.390	0.374	4.350	3.555	3.707	3.044	4.031	3.114
SL	RetinexNet [58]	16.774	0.462	0.390	15.609	0.414	0.393	4.487	3.242	4.732	3.219	5.881	3.727
	KinD++ [74]	17.752	0.758	0.198	16.085	0.394	0.366	4.027	3.399	4.005	3.144	3.586	2.773
	LCDPNet [52]	14.506	0.575	0.312	15.689	0.474	0.344	4.110	3.250	4.106	3.127	5.039	3.347
	URetinexNet [59]	19.842	0.824	0.128	18.271	0.518	0.295	4.774	3.565	4.028	3.153	3.851	2.891
	SMG [64]	**23.814**	0.809	0.144	17.579	0.538	0.456	6.224	4.228	5.300	3.627	5.752	3.757
	PyDiff [76]	23.275	**0.859**	**0.108**	17.264	0.510	0.335	4.499	3.792	4.082	3.268	4.360	3.678
	GSAD [20]	22.021	0.848	0.137	17.414	0.507	**0.294**	4.496	3.593	4.489	3.361	5.252	3.657
SSL	DRBN [68]	16.677	0.730	0.252	16.734	0.507	0.376	4.369	3.800	3.921	3.267	3.671	3.117
	BL [39]	10.305	0.401	0.382	12.444	0.333	0.384	5.046	4.055	4.885	3.870	5.740	4.030
UL	Zero-DCE [12]	14.861	0.562	0.330	15.867	0.443	0.315	3.951	3.149	3.826	2.918	5.080	3.307
	EnlightenGAN [24]	17.606	0.653	0.319	17.106	0.463	0.322	3.832	3.256	3.775	2.953	3.689	2.749
	RUAS [32]	16.405	0.503	0.257	14.271	0.461	0.455	7.306	5.700	7.198	5.651	4.987	4.329
	SCI [40]	14.784	0.525	0.333	15.242	0.419	0.321	4.519	3.700	4.124	3.534	5.312	3.648
	GDP [8]	15.896	0.542	0.337	12.887	0.362	0.386	4.358	3.552	4.032	3.097	4.683	3.431
	PairLIE [10]	19.514	0.731	0.254	17.602	0.501	0.323	4.282	3.469	4.661	3.543	3.373	2.734
	NeRCo [67]	19.738	0.740	0.239	17.844	0.535	0.371	4.107	3.345	3.902	3.037	3.765	3.094
	Ours	20.453	0.803	0.192	**18.555**	**0.539**	0.311	**3.724**	**3.144**	**3.618**	**2.879**	**2.941**	**2.558**

paired low-light and normal-light images, including LOL [58] and LSRW [16], as well as three real-world unpaired benchmarks that contain low-light images only, including DICM [28], NPE [53], and VV [51]. For paired datasets, we adopt two distortion metrics PSNR and SSIM [57], and a full-reference perceptual metric LPIPS [73] for evaluation. For unpaired datasets, we use two non-reference perceptual metrics NIQE [41] and PI [1] to measure the visual quality.

4.2   Comparison with Existing Methods

Comparison Methods. We compare our method with four categories of existing LLIE methods: 1) traditional methods including LIME [14], SDDLLE [17], BrainRetinex [4], and CDEF [30], 2) supervised methods including RetinexNet [58], KinD++ [74], LCDPNet [52], URetinexNet [59], SMG [64], PyDiff [76], and GSAD [20], 3) semi-supervised methods DRBN [68] and BL [39], 4) unsupervised methods including Zero-DCE [12], EnlightenGAN [24], RUAS [32], SCI [40], GDP [8], PairLIE [10], and NeRCo [67]. Note that supervised methods are trained on the LOL training set, and the reported performance of GDP and our method are the mean values for five times evaluation.

   Quantitative Comparison. We first compare the proposed method with all comparison methods on the LOL [58] and LSRW [16] test sets. As shown in Table 1, our LightenDiffusion outperforms all unsupervised competitors on both two benchmarks. The reason we cannot surpass supervised approaches on the LOL dataset is that they are typically trained on it and can therefore achieve
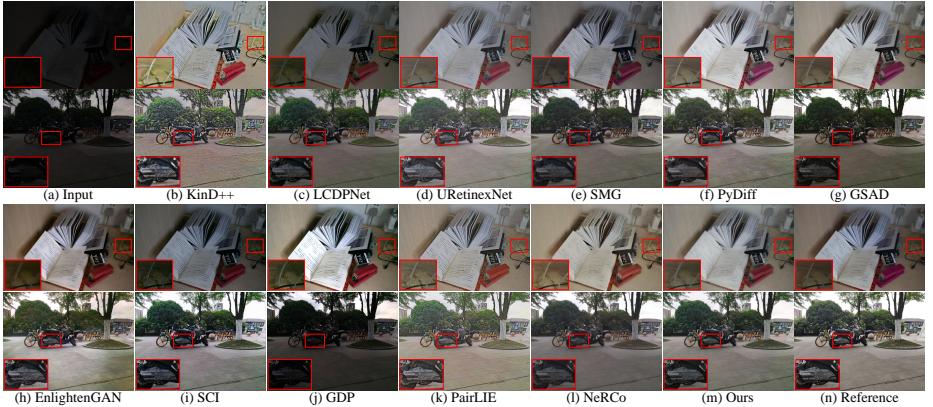
Fig. 5: Qualitative comparison of our method and competitive methods on the LOL [58] and LSRW [16] test sets. Best viewed by zooming in.
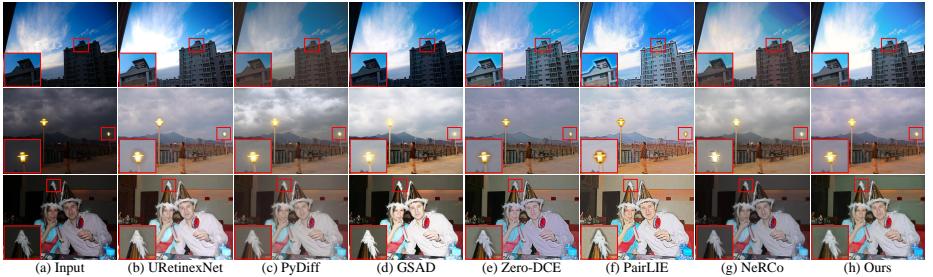

Fig. 6: Qualitative comparison of our method and competitive methods on the DICM [28], NPE [53], and VV [51] datasets. Best viewed by zooming in.

satisfactory performance. However, our method outperforms supervised methods on the LSRW dataset, achieving the highest PSNR and SSIM with slightly inferior in terms of LPIPS. To further validate the effectiveness of our method, we also compare the proposed LightenDiffusion with comparison methods on three unpaired benchmarks DICM [28], NPE [53], and VV [51]. As shown in Table 1, unsupervised methods present better generalization ability than supervised ones on these unseen datasets, where our method obtains the best results on all three datasets. It indicates that our method is able to generate visually satisfactory images and can generalize well to various scenes.

Qualitative Comparison. We present visual comparisons of our method and competitive methods on the paired datasets in Fig. 5, where the images in rows 1-2 are selected from LOL [58] and LSRW [16] test sets, respectively. We can see that previous methods yield results with underexposure, color distortion, or noise amplification, while our method properly improves global and local contrast, reconstructs sharper details, and suppresses noise, resulting in visually pleasing results. We also provide results on the unpaired benchmarks in Fig. 6,
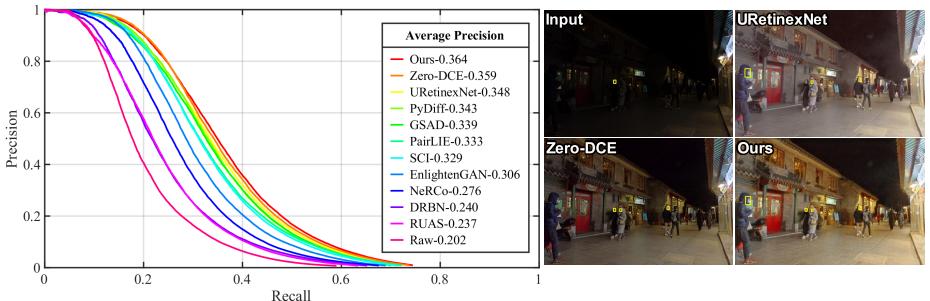
Fig. 7: Comparison of low-light face detection results on the DARK FACE dataset [69].

where the images in rows 1-3 are selected from DICM [28], NPE [53], and VV [51] datasets, respectively. Previous methods fail to generalize well to these scenes, especially in row 2, where some methods present artifacts around the light or produce overexposed results. In contrast, our method presents correct exposure and vivid color, which proves the superiority of our generalization ability.

4.3   Low-Light Face Detection

In this section, we conduct experiments on the DARK FACE dataset [69], which consists of 6,000 images captured under weakly illuminated conditions with annotated labels for evaluation, to investigate the impact of LLIE methods as a pre-processing step in improving the low-light face detection task. Following [12, 22, 40], we employ our method and 10 competitive LLIE methods to restore the images, followed by the well-known detector RetinaFace [7] for evaluation under the IoU threshold of 0.3 to depict the precision-recall (P-R) curves and calculate the average precision (AP). As illustrated in Fig. 7, our method effectively improves the precision of RetinaFace from 20.2% to 36.4% compared to the raw images without enhancement and outperforms other methods in the high recall area, which reveals the potential practical values of our method.

4.4   Ablation Study

In this section, we conduct a series of ablation studies to validate the impact of different component choices. We use the implementation details described in Sec. 4.1 for training and quantitative results on the LOL [58] and DICM [28] datasets are illustrated in Table 2. Detailed settings are discussed below.

Latent Space v.s. Image Space. To validate the effectiveness of our latent-Retinex decomposition strategy, we conduct experiments by performing the decomposition in image space, i.e., $k = 0$, and in various scales of latent space, i.e., $k \in [1, 4]$. As shown in Fig. 8(a), it is difficult to achieve satisfactory decomposition in the image space, where the illumination map would exhibit certain content information thus making the restored image present artifacts. Conversely, as shown in Fig. 8(b)-(d), performing decomposition in the latent space can yield

Table 2: Quantitative results of ablation studies. The results using default settings are underlined. 'w/o' denotes without and 'Time' denotes the inference speed when performing inference on RTX 2080Ti for an image with 400×600×3 resolution.

		LOL [58]			DICM [28]		Time (s) ↓
	Method	PSNR ↑	SSIM ↑	LPIPS ↓	NIQE↓	PI↓	
1)	$k = 0$ (Image Space)	17.054	0.715	0.372	4.519	4.377	4.733
2)	$k = 1$ (Latent Space)	19.228	0.728	0.355	4.101	3.457	0.872
3)	$k = 2$ (Latent Space)	20.097	0.798	0.210	4.021	3.402	0.411
4)	$k = 4$ (Latent Space)	20.104	0.785	0.195	3.906	3.332	0.256
5)	RetinexNet [58]	16.616	0.563	0.579	5.859	6.056	0.296
6)	URetinexNet [59]	17.916	0.703	0.391	4.371	4.561	0.293
7)	PairLIE [10]	17.089	0.605	0.568	6.017	6.349	0.295
8)	w/o $\mathcal{L}_{scc}$ ($S = 20$)	19.184	0.785	0.213	4.045	3.408	0.314
9)	w/o $\mathcal{L}_{scc}$ ($S = 50$)	19.473	0.791	0.209	3.998	3.392	0.687
10)	w/o $\mathcal{L}_{scc}$ ($S = 100$)	20.255	0.801	0.209	3.831	3.228	1.208
11)	Default	20.453	0.803	0.192	3.724	3.144	0.314

illumination maps that represent only the lighting conditions, which facilitates the diffusion model to generate restored images with visual fidelity. Moreover, as reported in rows 1-4 of Table 2, increasing k improves the overall performance and inference speed, while showing slight performance degradation at $k = 4$ due to the substantial reduction in feature information richness, which adversely affects the generative ability of the diffusion model. For a trade-off between performance and efficiency, we choose $k = 3$ as the default setting.

Retinex Decomposition Network. To validate the effectiveness of our proposed CTDN, we replace it with the decomposition network of three previous Retinex-based methods, including RetinexNet [58], URetinexNet [59], and
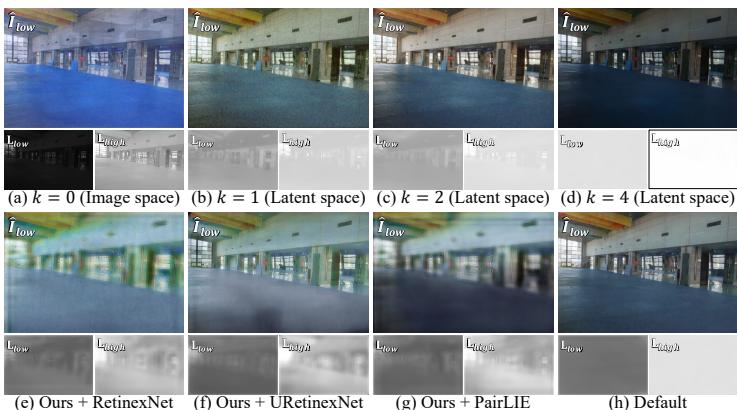

(a) $k = 0$ (Image space)   (b) $k = 1$ (Latent space)   (c) $k = 2$ (Latent space)   (d) $k = 4$ (Latent space)

(e) Ours + RetinexNet   (f) Ours + URetinexNet   (g) Ours + PairLIE   (h) Default

Fig. 8: Visual results of the ablation study about our employed latent-Retinex decomposition strategy and the proposed content-transfer decomposition network. The first row shows the restored results with different settings, and the second row presents estimated illumination maps of low/normal-light images.
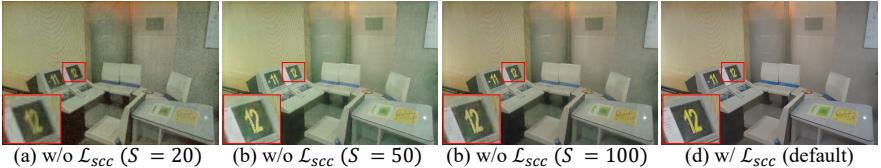
| (a) w/o $\mathcal{L}_{scc}$ ($S = 20$) | (b) w/o $\mathcal{L}_{scc}$ ($S = 50$) | (b) w/o $\mathcal{L}_{scc}$ ($S = 100$) | (d) w/ $\mathcal{L}_{scc}$ (default) |

Fig. 9: Visual results of the ablation study about our proposed $\mathcal{L}_{scc}$.

PairLIE [10], to estimate the reflectance and illumination maps. As shown in Fig. 8(e)-(g), previous decomposition networks are unable to obtain content-free illumination maps, resulting in the restored results with blurry details and artifacts. In contrast, our method benefits from the well-designed network architecture of CTDN that enables the generation of content-rich reflectance maps and content-free illumination maps, resulting in remarkable performance superiority in comparison, as reported in Table 2.

Loss Function. To validate the effectiveness of the proposed self-constrained consistency loss $\mathcal{L}_{scc}$, we conduct an experiment to remove it from the object function utilized to optimize the diffusion model. As reported in row 8 of Table 2, removing $\mathcal{L}_{scc}$ results in decreased overall performance. Moreover, we increase the sampling step S to 50 and 100 to evaluate the performance of the diffusion model trained with vanilla diffusion loss, i.e., Eq.(5), since the quality of generated results from diffusion models would improve with increasing S [49], as shown in Fig. 9. Compared to the default setting in row 11, while increasing the sampling step size to $S = 100$ yields comparable performance to the model trained with $\mathcal{L}_{scc}$, it results in almost 4 times slower inference speed, which proves our loss can facilitate the model to achieve efficient and robust restoration.

5   Conclusion

We have presented LightenDiffusion, a diffusion-based framework that incorporates Retinex theory with diffusion models for unsupervised LLIE. Technically, we propose a content-transfer decomposition network that performs decomposition within the latent space to obtain content-rich reflectance maps and content-free illumination maps to facilitate subsequently unsupervised restoration. The reflectance map of the low-light image and the illumination map of the normal-light image captured in different scenes serve as inputs to the diffusion model for training. Moreover, we propose a self-constrained consistency loss to further constrain the restored result to have the same inherent content information as the low-light input. Experimental results show that our method outperforms state-of-the-art competitors both quantitatively and visually.

References

1. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The 2018 pirm challenge on perceptual image super-resolution. In: ECCV (2018) 10
2. Bovik, A.C.: Handbook of image and video processing. Academic press (2010) 1, 4
3. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE TIP **27**(4), 2049–2062 (2018) 8
4. Cai, R., Chen, Z.: Brain-like retinex: A biologically plausible retinex algorithm for low light image enhancement. PR **136**, 109195 (2023) 10
5. Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., Zhang, Y.: Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In: ICCV. pp. 12504–12513 (2023) 2, 4, 5
6. Chung, H., Sim, B., Ye, J.C.: Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In: CVPR. pp. 12413–12422 (2022) 2, 4, 7
7. Deng, J., Guo, J., Ververas, E., Kotsia, I., Zafeiriou, S.: Retinaface: Single-shot multi-level face localisation in the wild. In: CVPR. pp. 5203–5212 (2020) 12
8. 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: CVPR. pp. 9935–9946 (2023) 2, 3, 4, 10
9. Fu, X., Zeng, D., Huang, Y., Zhang, X.P., Ding, X.: A weighted variational model for simultaneous reflectance and illumination estimation. In: CVPR. pp. 2782–2790 (2016) 1, 4
10. 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: CVPR. pp. 22252–22261 (2023) 2, 4, 5, 6, 8, 10, 13, 14
11. Gao, S., Liu, X., Zeng, B., Xu, S., Li, Y., Luo, X., Liu, J., Zhen, X., Zhang, B.: Implicit diffusion models for continuous super-resolution. In: CVPR. pp. 10021–10030 (2023) 4
12. 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: CVPR. pp. 1780–1789 (2020) 2, 4, 10, 12
13. Guo, X., Hu, Q.: Low-light image enhancement via breaking down the darkness. IJCV **131**(1), 48–66 (2023) 2, 4
14. Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illumination map estimation. IEEE TIP **26**(2), 982–993 (2016) 1, 4, 6, 10
15. Hai, J., Hao, Y., Zou, F., Lin, F., Han, S.: Advanced retinexnet: a fully convolutional network for low-light image enhancement. Signal Processing: Image Communication **112**, 116916 (2023) 4
16. Hai, J., Xuan, Z., Yang, R., Hao, Y., Zou, F., Lin, F., Han, S.: R2rnet: Low-light image enhancement via real-low to real-normal network. Journal of Visual Communication and Image Representation **90**, 103712 (2023) 10, 11
17. Hao, S., Han, X., Guo, Y., Xu, X., Wang, M.: Low-light image enhancement with semi-decoupled decomposition. IEEE TMM **22**(12), 3025–3038 (2020) 10
18. He, C., Fang, C., Zhang, Y., Li, K., Tang, L., You, C., Xiao, F., Guo, Z., Li, X.: Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. arXiv preprint arXiv:2311.11638 (2023) 4
19. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. NeurIPS **33**, 6840–6851 (2020) 2, 4, 7, 8

20. Hou, J., Zhu, Z., Hou, J., Liu, H., Zeng, H., Yuan, H.: Global structure-aware diffusion process for low-light image enhancement. NeurIPS **36** (2023) 2, 4, 8, 10

21. Hou, R., Chang, H., Ma, B., Shan, S., Chen, X.: Cross attention network for few-shot classification. NeurIPS **32** (2019) 6

22. Jiang, H., Luo, A., Han, S., Fan, H., Liu, S.: Low-light image enhancement with wavelet-based diffusion models. ACM TOG **42**(6), 1–14 (2023) 2, 4, 8, 12

23. Jiang, H., Ren, Y., Han, S.: Revisiting coarse-to-fine strategy for low-light image enhancement with deep decomposition guided training. Computer Vision and Image Understanding **142**, 103952 (2024) 4

24. 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 TIP **30**, 2340–2349 (2021) 2, 4, 10

25. Kawar, B., Elad, M., Ermon, S., Song, J.: Denoising diffusion restoration models. NeurIPS **35**, 23593–23606 (2022) 3, 4

26. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015) 9

27. Land, E.H.: The retinex theory of color vision. Scientific American **237**(6), 108–129 (1977) 1, 4, 5

28. Lee, C., Lee, C., Kim, C.S.: Contrast enhancement based on layered difference representation of 2d histograms. IEEE TIP **22**(12), 5372–5384 (2013) 10, 11, 12, 13

29. Lee, S., Jang, D., Kim, D.S.: Temporally averaged regression for semi-supervised low-light image enhancement. In: CVPR. pp. 4207–4216 (2023) 4

30. Lei, X., Fei, Z., Zhou, W., Zhou, H., Fei, M.: Low-light image enhancement using the cell vibration model. IEEE TMM (2022) 10

31. Li, H., Jiang, H., Luo, A., Tan, P., Fan, H., Zeng, B., Liu, S.: Dmhomo: Learning homography with diffusion models. ACM TOG **43**(3), 1–16 (2024) 4

32. Liu, R., Ma, L., Zhang, J., Fan, X., Luo, Z.: Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: CVPR. pp. 10561–10570 (2021) 2, 4, 10

33. Loh, Y.P., Chan, C.S.: Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding **178**, 30–42 (2019) 1

34. Lore, K.G., Akintayo, A., Sarkar, S.: Llnet: A deep autoencoder approach to natural low-light image enhancement. PR **61**, 650–662 (2017) 2, 4

35. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: Inpainting using denoising diffusion probabilistic models. In: CVPR. pp. 11461–11471 (2022) 3, 4

36. Luo, A., Li, X., Yang, F., Liu, J., Fan, H., Liu, S.: Flowdiffuser: Advancing optical flow estimation with diffusion models. In: CVPR. pp. 19167–19176 (2024) 4

37. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Image restoration with mean-reverting stochastic differential equations. In: ICML (2023) 4

38. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In: CVPRW. pp. 1680–1691 (2023) 4

39. Ma, L., Jin, D., An, N., Liu, J., Fan, X., Luo, Z., Liu, R.: Bilevel fast scene adaptation for low-light image enhancement. IJCV pp. 1–19 (2023) 10

40. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: CVPR. pp. 5637–5646 (2022) 4, 10, 12

41. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Sign. Process. Letters **20**(3), 209–212 (2012) 10

42. Ooi, C.H., Isa, N.A.M.: Quadrants dynamic histogram equalization for contrast enhancement. IEEE TCE **56**(4), 2552–2559 (2010) 1, 4

43. Özdenizci, O., Legenstein, R.: Restoring vision in adverse weather conditions with patch-based denoising diffusion models. IEEE TPAMI **45**(8), 10346–10357 (2023) 2, 4

44. Park, J., Vien, A.G., Kim, J.H., Lee, C.: Histogram-based transformation function estimation for low-light image enhancement. In: ICIP. pp. 1–5 (2022) 1, 4

45. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10684–10695 (2022) 4

46. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241 (2015) 9

47. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH. pp. 1–10 (2022) 2, 4

48. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE TPAMI **45**(4), 4713–4726 (2022) 2, 4

49. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: ICLR (2021) 2, 4, 7, 8, 14

50. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. NeurIPS **30** (2017) 6

51. Vonikakis, V., Kouskouridas, R., Gasteratos, A.: On the evaluation of illumination compensation algorithms. Multimedia Tools and Applications **77**, 9211–9231 (2018) 10, 11, 12

52. Wang, H., Xu, K., Lau, R.W.: Local color distributions prior for image enhancement. In: ECCV. pp. 343–359 (2022) 10

53. Wang, S., Zheng, J., Hu, H.M., Li, B.: Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE TIP **22**(9), 3538–3548 (2013) 1, 10, 11, 12

54. Wang, Y., Liu, Z., Liu, J., Xu, S., Liu, S.: Low-light image enhancement with illumination-aware gamma correction and complete image modelling network. In: ICCV. pp. 13128–13137 (2023) 2, 4

55. Wang, Y., Yu, J., Zhang, J.: Zero-shot image restoration using denoising diffusion null-space model. In: ICLR (2023) 3, 4

56. Wang, Y., Yu, Y., Yang, W., Guo, L., Chau, L.P., Kot, A.C., Wen, B.: Exposurediffusion: Learning to expose for low-light image enhancement. In: ICCV. pp. 12438–12448 (2023) 4

57. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE TIP **13**(4), 600–612 (2004) 10

58. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: BMVC (2018) 2, 4, 5, 6, 9, 10, 11, 12, 13

59. Wu, W., Weng, J., Zhang, P., Wang, X., Yang, W., Jiang, J.: Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In: CVPR. pp. 5901–5910 (2022) 2, 4, 5, 6, 10, 13

60. Wu, Y., Pan, C., Wang, G., Yang, Y., Wei, J., Li, C., Shen, H.T.: Learning semantic-aware knowledge guidance for low-light image enhancement. In: CVPR. pp. 1662–1671 (2023) 2, 4

61. Xie, S., Zhang, Z., Lin, Z., Hinz, T., Zhang, K.: Smartbrush: Text and shape guided object inpainting with diffusion model. In: CVPR. pp. 22428–22437 (2023) 4

62. Xu, H., Li, H., Wang, Y., Liu, S., Fu, C.W.: Handbooster: Boosting 3d hand-mesh reconstruction by conditional synthesis and sampling of hand-object interactions. In: CVPR. pp. 10159–10169 (2024) 4
63. Xu, X., Wang, R., Fu, C.W., Jia, J.: Snr-aware low-light image enhancement. In: CVPR. pp. 17714–17724 (2022) 2, 4
64. Xu, X., Wang, R., Lu, J.: Low-light image enhancement via structure modeling and guidance. In: CVPR. pp. 9893–9903 (2023) 2, 4, 10
65. Xu, X., Wang, S., Wang, Z., Zhang, X., Hu, R.: Exploring image enhancement for salient object detection in low light images. ACM TOMM **17**(1s), 1–19 (2021) 1
66. Yang, S., Zhou, D., Cao, J., Guo, Y.: Rethinking low-light enhancement via transformer-gan. IEEE Sign. Process. Letters **29**, 1082–1086 (2022) 2
67. Yang, S., Ding, M., Wu, Y., Li, Z., Zhang, J.: Implicit neural representation for cooperative low-light image enhancement. In: ICCV. pp. 12918–12927 (2023) 2, 4, 10
68. 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: CVPR. pp. 3063–3072 (2020) 4, 10
69. Yang, W., Yuan, Y., Ren, W., Liu, J., Scheirer, W.J., Wang, Z., Zhang, T., Zhong, Q., Xie, D., Pu, S., et al.: Advancing image understanding in poor visibility environments: A collective benchmark study. IEEE TIP **29**, 5737–5752 (2020) 12
70. Yang, Z., Li, H., Hong, M., Zeng, B., Liu, S.: Single image rolling shutter removal with diffusion models. arXiv preprint arXiv:2407.02906 (2024) 4
71. Yi, X., Xu, H., Zhang, H., Tang, L., Ma, J.: Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In: ICCV. pp. 12302–12311 (2023) 2, 4, 5, 9
72. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: ECCV. pp. 492–511 (2020) 4
73. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR. pp. 586–595 (2018) 10
74. Zhang, Y., Guo, X., Ma, J., Liu, W., Zhang, J.: Beyond brightening low-light images. IJCV **129**, 1013–1037 (2021) 4, 5, 6, 9, 10
75. Zheng, D., Zhang, X., Ma, K., Bao, C.: Learn from unpaired data for image restoration: A variational bayes approach. IEEE TPAMI **45**(5), 5889–5903 (2022) 2
76. Zhou, D., Yang, Z., Yang, Y.: Pyramid diffusion models for low-light image enhancement. In: IJCAI (2023) 2, 4, 8, 10
77. Zhou, T., Li, H., Wang, Z., Luo, A., Zhang, C.L., Li, J., Zeng, B., Liu, S.: Recdiffusion: Rectangling for image stitching with diffusion models. In: CVPR. pp. 2692–2701 (2024) 4
78. 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: CVPRW. pp. 1219–1229 (2023) 3, 4