Unsupervised Low-light Image Enhancement with Lookup Tables and Diffusion Priors

Yunlong Lin[1*] Zhenqi Fu[2*] Kairun Wen[1] Tian Ye[3] Sixiang Chen[3]
Ge Meng[1] Yingying Wang[1] Yue Huang[1] Xiaotong Tu[1] Xinghao Ding[1†]
[1]Xiamen University [2]Tsinghua University
[3]The Hong Kong University of Science and Technology (Guangzhou)

Project page: https://dplut.github.io/

Abstract

*Low-light image enhancement (LIE) aims at precisely and efficiently recovering an image degraded in poor illumination environments. Recent advanced LIE techniques are using deep neural networks, which require lots of low-normal light image pairs, network parameters, and computational resources. As a result, their practicality is limited. In this work, we devise a novel unsupervised LIE framework based on **d**iffusion **p**riors and lookup tables (DPLUT) to achieve efficient low-light image recovery. The proposed approach comprises two critical components: a **l**ight adjustment lookup table (LLUT) and a **n**oise suppression lookup table (NLUT). LLUT is optimized with a set of unsupervised losses. It aims at predicting pixel-wise curve parameters for the dynamic range adjustment of a specific image. NLUT is designed to remove the amplified noise after the light brightens. As diffusion models are sensitive to noise, diffusion priors are introduced to achieve high-performance noise suppression. Extensive experiments demonstrate that our approach outperforms state-of-the-art methods in terms of visual quality and efficiency.*

1. Introduction

The goal of low-light image enhancement (LIE) is to improve the visual quality of images captured in low-light conditions. As a fundamental preprocessing task, LIE algorithms are expected to be effective and efficient, especially on resource-constrained devices and embedded platforms. Over the past few years, prolific algorithms have been proposed, which can be roughly classified into efficiency- and quality-oriented methods.

Efficiency-oriented methods, such as SCI [44] and ZeroDCE++ [33], can generate normal-light images in real-time. However, these methods have limited representative ability in diverse low-light degradation, resulting in suboptimal performance. As shown in Fig. 1, efficiency-oriented methods can meet the real-time requirements of practical applications, but the performance is far from satisfactory. In contrast, quality-oriented approaches, such as Retinexformer [4] and CUE [71] achieve relatively higher performance than efficiency-oriented methods. Their success was largely due to the deep and complex network architectures, massive paired training data, and high computational and memory costs. As presented in Fig. 1, existing quality-oriented approaches cannot process 4K resolution images in real-time.

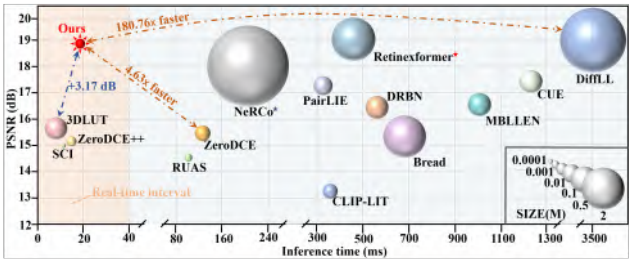

Figure 1. **Comparisons of performance and efficiency.** The average PSNR is evaluated on LSRW [22], and inference time is evaluated on 4K (3840×2160) resolution with a single Titan RTX GPU. Our approach obtains the highest PSNR and can process 4K low-light images in real-time. Note that * and * indicate the maximum size that the models can handle is 480P (640×480) and 1080P (1920×1080), respectively.

The lookup table (LUT) is a crucial component in the image signal processors (ISPs) [31] owing to its high efficiency and practicality. LUTs can be either pre-defined or learned. Pre-defined LUTs are manually tuned by experienced experts, exhibiting limited expressive capacity to adapt to diverse scenes. In contrast, learned LUTs [50, 60, 68] are significantly more expressive and have achieved

*Yunlong Lin and Zhenqi Fu contribute equally to this work.
†Xinghao Ding (dxh@xmu.edu.cn) is the corresponding author.

Figure 2. **Visual performance in real-world scenes with 4K resolution.** Our DPLUT achieves visually pleasing results in terms of brightness, color, contrast, and naturalness across diverse scenes and under various light distributions.

promising outcomes in global exposure adjustment, contrast, and saturation, whereas challenges persist when applying LUTs for LIE. ***First***, the contrast and pixel values of low-light images can be extremely small. Classic LUTs fail to adjust the local contrast of such images and the enhanced results may suffer from color shift and artifacts. This limitation aligns with the findings in [39, 68]. ***Second***, current LUT-based image enhancement methods overlook the inherent sensor noise and artifacts concealed in low-light regions, which is also pointed out in [14, 68]. ***Third***, the paradigm of learning LUTs in an unsupervised fashion remains challenging.

To tackle the aforelisted issues, we introduce a novel LIE framework termed DPLUT, which aims to achieve higher quality and efficiency simultaneously by taking advantage of lookup tables and diffusion priors. The proposed DPLUT consists of two key components: a light adjustment lookup table (LLUT) and a noise suppression lookup table (NLUT). Firstly, LLUT is crafted to generate coarse normal-light images. We treat LIE as a curve mapping issue, adopting LLUT to estimate pixel-wise curve parameters. By explicitly combining the image-specific curve function and LUT, we can effectively perform the mapping within a wide dynamic range, and ensure the enhanced image has a correct local contrast. Secondly, to remove amplified noise and artifacts introduced from LLUT, NLUT is developed that marries the prior knowledge from the diffusion model to achieve real-time and high-quality image enhancement. ***LLUT and NLUT are trained in an unsupervised manner. Notably, the diffusion model is only introduced in NLUT learning phase. With LLUT and NLUT, our solution can cope with diverse light distributions in real-time.*** The enhanced results are more clean and natural compared with existing state-of-the-art (SOTA) approaches.

Our contributions can be summarized as follows:

- We develop a new unsupervised LIE framework with two lookup tables, i.e., a light adjustment lookup table and a noise suppression lookup table.
- We introduce diffusion priors and curve mappings to promote enhancement efficiency and effectiveness.
- Extensive evaluations on three benchmark datasets show that DPLUT achieves state-of-the-art performance and can enhance 4K low-light images in real-time.

2. Related Work

2.1. Low-light Image Enhancement

Enhancing images in low-light conditions has been a long-standing issue and great progress has been made over the decades. They can be roughly categorized into efficiency- and quality-oriented techniques. Efficiency-oriented approaches aim to construct lightweight enhancement models for mobile and source-limited platforms [1, 12, 28, 45]. For example, Wang et al. [55] enhanced the visibility and contrast via gamma correction and dynamic contrast ratio improvement. Guo et al. [21] proposed to refine the initial estimated illumination map by imposing a structure prior. Guo et al. [19] presented a reference-free LIE algorithm based on curve estimation, which can effectively perform mapping within a wide dynamic range. Ma et al. [44] established a cascaded illumination estimation process to achieve fast and robust LIE in complex scenarios. Despite efficiency, the enhancement performance of current efficiency-oriented approaches is greatly inferior to quality-oriented methods [25, 34, 41, 42, 51, 58, 61, 71, 72]. Lore et al. [42] designed a stacked sparse denoising auto-encoder to enhance low-light images. Lv et al. [43] presented a multi-branch network that extracts rich features from different levels to enhance low-light images via multiple sub-networks. Xu et al. [58] incorporated the signal-to-noise ratio (SNR) prior to achieving spatial-varying LIE. Cai et

al. [4] designed a sophisticated transformer-based algorithm for LIE. Hou et al. [25] devised a diffusion-based framework and introduced a global structure-aware regularization to preserve the image's details and textures. Yi et al. [67] combined the diffusion model with Retinex model for low-light image enhancement. Quality-oriented approaches require deep and complex network structures and a huge amount of computational resources. As a classic preposessing task, the practicality of quality-oriented methods is limited.

2.2. LUTs for Image Enhancement

The lookup tables (LUTs) are commonly used in ISPs [13, 16, 31], especially some resource-constrained devices, to accelerate computation. The mapping procedure can be evaluated using only memory access and interpolation without performing the computation again. Due to its portability, various LUT based solutions have been proposed for photo enhancement [39, 59, 60, 68]. For instance, Zeng et al. [68] first leveraged a lightweight CNN to predict the weights for integrating multiple basis LUTs, and the constructed image-adaptive LUT is utilized to enhance photos. Wang et al. [50] proposed a spatially-aware LUT that considers the global and local information. Liang et al. [35] improved the LUTs performance by adjusting learning strategies. Cong et al. [15] embedded the LUT-based sub-module in their network for efficient processing of high-resolution images. Yang et al. [59] focused on improving the sampling strategy for 3D LUTs. Yang et al. [60] combined 1D-LUT and 3D-LUT to promote the enhancement performance while reducing computational costs.

3. Preliminaries

3D-LUT. 3D-LUT is a widely used image enhancement tool that maps the input color values to the corresponding output color values. A classical 3D-LUT is defined as a 3D cube that contains N^3 elements, where N is the number of bins in each color channel. Each element defines a pixel-to-pixel mapping $\mu^c(i,j,k)$, where $i,j,k = 0,1,\ldots,N\text{-}1$ are elements' coordinates and c indicates one of the channels. Given an input image $\left\{I^r_{(i,j,k)}, I^g_{(i,j,k)}, I^b_{(i,j,k)}\right\}$, the mapping procedure can be formulated as follows:

$$O^c_{(i,j,k)} = \mu^c\left(I^r_{(i,j,k)}, I^g_{(i,j,k)}, I^b_{(i,j,k)}\right), \tag{1}$$

where O^c is the output of 3D-LUT, $c \in \{r,g,b\}$, and r,g,b is the color value of the red, green, blue channel, respectively. This mapping contains two basic operations, i.e., lookup and interpolation. The lookup operation is conducted to find its coordinates (i.e., i,j,k) in the 3D-LUT cube. Then, the output can be derived by the trilinear interpolation operation using its nearest eight surrounding elements. More detailed descriptions can be found in the supplementary material. Notably, as the value of N increases, the 3D color transformation space becomes more accurate. Nevertheless, a large N introduces massive parameters, leading to heavy memory burden, high training difficulty, and limited cell utilization. For simplicity, all LUTs mentioned in this paper refer to 3D-LUTs.

Denoising Diffusion Model. Diffusion model has shown remarkable promise in visual generation [2, 5, 6, 20], and it enlightens other tasks like image restoration [9, 64–66], image fusion [27, 38, 52, 53], dehazing [11, 63], and desnowing [7, 8, 10]. Denoising diffusion models generate images by gradually denoising from a gaussian noise $p(x_T) = \mathcal{N}(0,\mathrm{I})$ and transforming into a certain data distribution. The forward diffusion process $q(x_t \mid x_{t-1})$ adds Gaussian noise to the image x_t. The marginal distribution can be written as: $q(x_t \mid x_0) = \mathcal{N}\left(\alpha_t x_0, \sigma_t^2 \mathrm{I}\right)$, where α_t and σ_t are designed to converge to $\mathcal{N}(0,\mathrm{I})$ when t is at the end of the forward process [32, 49]. The reverse diffusion process $p(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)$ learns to denoise. Given infinitesimal timesteps, the reverse diffusion process can be approximated with Gaussian [49] related with an optimal MSE denoiser [47]. The diffusion models are designed as noise estimators $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)$ taking noisy images as input and estimating the noise. They are trained via optimizing the weighted evidence lower bound (ELBO) [24, 32]:

$$\mathcal{L}_{\mathrm{ELBO}}(\theta) = \mathbb{E}\left[w(t)\left\|\boldsymbol{\epsilon}_\theta\left(\alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}; t\right) - \boldsymbol{\epsilon}\right\|_2^2\right], \tag{2}$$

where $\epsilon \sim \mathcal{N}(0,\mathrm{I})$, $w(t)$ is a weighting function. In practice, setting $w(t) = 1$ delivers good performance [24]. Sampling from a diffusion model can be either stochastic [24] or deterministic [48]. After sampling $x_T \sim \mathcal{N}(0,\mathrm{I})$, we can gradually reduce the noise level and reach a clean image with high quality at the end of the iterative process.

4. Methodology

The overall framework of our proposed DPLUT involves two main stages, as illustrated in Fig. 3. In the first stage, we learn a light adjustment lookup table (LLUT) by a set of unsupervised losses, which maps the input RGB values to the corresponding pixel-wise curve parameter. With the LLUT, we can obtain coarse normal-light images. In the second stage, to remove amplified noise and artifacts introduced from LLUT, we learn a noise suppression lookup table (NLUT) through injecting knowledge of a diffusion model. It should be noted that the diffusion model and LLUT remain fixed during the training of NLUT. In the testing phase, with LLUT and NLUT, our solution can cope with diverse light distributions and achieve real-time and high-quality image enhancement. ***Note that LLUT and NLUT are decoupled, i.e., LLUT can yield favorable results without NLUT.*** We provide further details on the key components of our approach below.

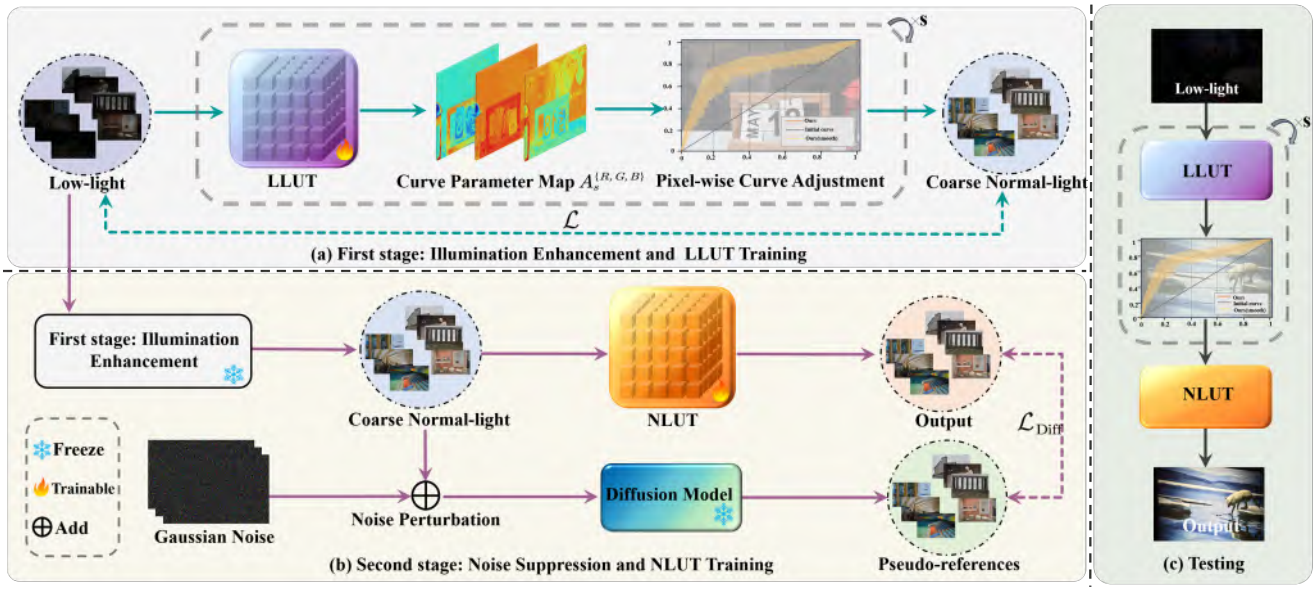

Figure 3. **The overall framework of our proposed DPLUT**. **In the training phase**, DPLUT involves two main stages. (a) In the first stage, we learn a light adjustment lookup table (LLUT), which estimates pixel-wise curve parameters for yielding coarse normal-light images. (b) In the second stage, we learn a noise suppression lookup table (NLUT) by introducing knowledge of a diffusion model, aiming at removing the amplified noise and artifacts introduced from LLUT. **In the testing phase**, with the LLUT and NLUT, DPLUT can robustly recover perceptual-friendly results in real-time.

Table 1. Architecture of the LUT generator, where k is a hyper-parameter that serves as a channel multiplier controlling the width of each convolutional layer. N is the size (number of elements along each dimension) of the LUT.

ID	Layer	Output Shape
0	Bilinear Resize	$3 \times 256 \times 256$
1	Depthwise Separable Conv3x3, LeakyReLU	$k \times 128 \times 128$
2	InstanceNorm	$k \times 128 \times 128$
3	Depthwise Separable Conv3x3, LeakyReLU	$2k \times 64 \times 64$
4	InstanceNorm	$2k \times 64 \times 64$
5	Depthwise Separable Conv3x3, LeakyReLU	$4k \times 32 \times 32$
6	InstanceNorm	$4k \times 32 \times 32$
7	Depthwise Separable Conv3x3, LeakyReLU	$8k \times 16 \times 16$
8	InstanceNorm	$8k \times 16 \times 16$
9	Depthwise Separable Conv3x3, LeakyReLU	$8k \times 8 \times 8$
10	Dropout (0.5)	$8k \times 8 \times 8$
11	AveragePooling	$8k \times 2 \times 2$
12	Reshape	$32k$
13	Fully Connected Layer	M
14	Fully Connected Layer	$3N^3$
15	Reshape	$3 \times N \times N \times N$

4.1. Light Adjustment Lookup Table

Motivation for applying LUTs to predict curve parameters. Existing LUTs mainly focus on RGB-to-RGB mapping. As a result, they always require a relatively large table size (33 or 64 points) to address the diversity of color ranges, leading to a heavy memory burden and high training difficulty. In contrast, ***LLUT implements a more simple and effective mapping, i.e., RGB-to-Curve parameters, which requires a small table size (9 points)***. The superiority of curve mapping is twofold: 1) It is monotonic, ensuring the preservation of contrast between adjacent pixels; 2) It is simple and differentiable, which benefits the gradient back-propagation process and can facilitate convergence.

As depicted in Fig. 3(a), the first stage of our training framework involves the construction of a light adjustment lookup table (LLUT), which estimates the pixel-wise curve parameter for dynamic range adjustment of a specific image. We begin by formulating our setting and introducing our notations. Given a low-light image $I(x) \in \mathbb{R}^{H \times W \times 3}$, LLUT maps the input RGB values to the corresponding curve parameter map $A(x)$. In order to automatically generate an image-adaptive LLUT, as shown in Fig. 4(a), we predict all the N^3 elements in the LUT by the neural network to consider the adaptation to the diversity of various input images. Such an objective formulates a mapping from the image context D to a $3N^3$-dimension parameter space:

$$\text{LLUT} = f_{3D}(D), \tag{3}$$

where LLUT is generated by the LUT generator module $f_{3D}(\cdot)$. The detailed architecture can be found in Tab. 1. Given the predicted LLUT, the estimated pixel-wise curve parameter map $A(x)$ for a specific image $I(x)$ can be formulated as:

$$A(x) = \text{trilinear\_interpolate}(\text{LLUT}, I(x)). \tag{4}$$

Then, we recurrently apply the pixel-wise adjustment curve [19] to obtain the coarse normal-light sample $\hat{I}(x) \in$

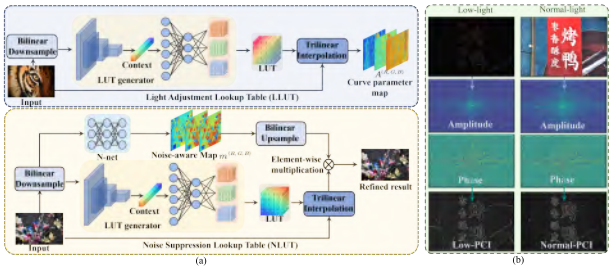

Figure 4. (a) The architecture of our two key components: a light adjustment lookup table (LLUT) and a noise suppression lookup table (NLUT). (b) We applied an inverse discrete Fourier transform to the phase of the low/normal-light image to obtain the phase-only reconstruction image (PCI) in the spatial domain. That means the amplitude of low/normal-light image is set to 1.

$\mathbb{R}^{H\times W\times 3}$. At the s-th step, the intermediate enhanced result $I_{s+1}(x)$ is:

$$I_{s+1}(x) = I_s(x) + \mathcal{A}_s(x) I_s(x)\left(1 - I_s(x)\right), \tag{5}$$

where x denotes the pixel coordinates. The range of $A_s(x)$ is limited to between -1 and 1. Based on Eq. 5, to enable zero-reference learning in LLUT, the following four types of losses are adopted.

Exposure loss. This loss function ensures the enhanced image has a reasonable exposure level by penalizing the gray-scale intensity deviation from the mid-tone value:

$$\mathcal{L}_e = \frac{1}{z}\sum_{i=1}^{z} \|V_i - v\|_2^2, \tag{6}$$

where z represents the number of non-overlapping local regions of size 16×16, and V_i is the is the average intensity value of a local region in $\hat{I}$. We set $v = 0.65$ in our experiments.

Structural consistency loss. This loss function encourages spatial coherence of the enhanced image by minimizing phase error between the input image and its enhanced version:

$$\mathcal{L}_p = \left\| \mathcal{P}(I) - \mathcal{P}\left(\hat{I}\right) \right\|_1, \tag{7}$$

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. This conclusion is consistent with that in [34].

Color loss. This loss is based on the gray-world assumption, endeavoring to minimize the mean value difference between each color channel pair to correct the potential color deviations in the enhanced image:

$$\mathcal{L}_c = \sum_{(i,j)\in\xi} \left(\hat{I}^i - \hat{I}^j\right)^2, \xi \in \{(R,G),(G,B),(B,R)\}. \tag{8}$$

Smoothing loss. This loss function is calculated in pixel-wise of each curve parameter map A, which preserves the monotonicity between neighboring pixels:

$$\mathcal{L}_s = \frac{1}{n}\sum_{s=1}^{n}\sum_{c\in\delta} \left(|\nabla_x \mathcal{A}_s^c| + \nabla_y \mathcal{A}_s^c\ |\right)^2, \delta = \{R,G,B\}, \tag{9}$$

where n is the number of curve parameter maps. ∇_x and ∇_y represent the horizontal and vertical gradient operations, respectively.

The full objective function for LLUT is a weighted sum of all sub-loss terms:

$$\mathcal{L} \ = \ \lambda_1\mathcal{L}_e + \mathcal{L}_p + \lambda_2\mathcal{L}_c + \lambda_3\mathcal{L}_s, \tag{10}$$

where λ_1, λ_2 and λ_3 are the weights of the losses, which are empirically set to 10, 5 and 1600 in all experiments.

4.2. Noise Suppression Lookup Table

LLUT learns the curve parameter mapping based on a set of unsupervised losses, its results might remain undesired noise and artifacts. Recently, the diffusion model has garnered considerable attention for its powerful generative capability and remarkable performance across various vision tasks. ***In this paper, we introduce the powerful prior knowledge of the pre-trained diffusion model (PTDM) to facilitate noise suppression lookup table learning.*** Specifically, as presented in Fig. 3(b), we feed the coarse normal-light sample $\hat{I}$ to NLUT and PTDM, which generate final results $\hat{Y}$ and pseudo-references Y, respectively. As shown in Fig. 4(a), NLUT has a similar architecture to LLUT. The only difference is that we employ an additional lightweight network to estimate the pixel-wise noise-aware map for adjusting the output value. The output of NLUT can be formulated as:

$$\hat{Y}\left(x\right) = \ \text{trilinear\_interpolate}\left(\text{NLUT}, \hat{I}\left(x\right)\right) \odot m, \tag{11}$$

where NLUT is generated by the LUT generator module $f_{3D}\left(\cdot\right)$, as shown in Tab. 1. $m = \{m_{h,w} \mid h \in \mathbb{R}^{H-1}, w \in \mathbb{R}^{W-1}\}$ is a noise-aware pixel-wise weight map for NLUT at location (h, w), which is estimated by a lightweight network $\gamma\left(\cdot\right)$. Specifically, $\gamma\left(\cdot\right)$ contains six convolutional layers, and the first five layers are followed by a ReLU function to increase the nonlinear mapping ability.

Meanwhile, we use a PTDM to refine coarse normal-light sample $\hat{I}$ to pseudo-reference Y through the forward

Table 2. Quantitative comparison on LOL, SICE and LSRW. "T", "S", and "U" represent "Traditional", "Supervised" and "Unsupervised" methods, respectively. The best results of "S" and "U" are marked in blue and red, respectively.

Method	Type	LOL			SICE			LSRW			Params (M)
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
SDD [23]	T	13.34	0.63	0.74	15.34	0.73	0.26	14.70	0.49	0.41	-
LECARM [46]	T	14.40	0.54	0.32	18.59	0.78	0.26	15.33	0.42	0.32	-
MBLLEN [43]	S	15.25	0.70	0.32	18.41	0.73	0.31	16.87	0.51	0.45	0.45
RetinexNet [56]	S	17.60	0.64	0.38	19.57	0.78	0.27	15.58	0.41	0.39	0.84
DSLR [37]	S	15.20	0.59	0.32	14.32	0.68	0.38	15.21	0.44	0.38	14.93
DRBN [62]	S	19.67	0.82	0.16	18.73	0.78	0.28	16.72	0.51	0.39	0.58
3DLUT [68]	S	16.36	0.64	0.35	15.53	0.64	0.38	15.74	0.48	0.43	0.59
Bread [26]	S	22.95	0.83	0.15	17.28	0.80	0.25	16.06	0.53	0.36	2.12
CUE [71]	S	22.67	0.79	0.20	20.06	0.82	0.24	18.19	0.52	0.33	0.26
DiffLL [29]	S	26.19	0.85	0.11	21.33	0.84	0.22	19.27	0.55	0.30	22.05
Retinexformer [4]	S	25.15	0.84	0.13	22.32	0.85	0.20	19.23	0.54	0.31	1.61
EnlightenGAN [30]	U	17.48	0.65	0.32	18.73	0.82	0.23	17.05	0.46	0.33	8.64
ZeroDCE [19]	U	14.86	0.55	0.33	18.67	0.80	0.26	15.84	0.45	0.31	0.079
ZeroDCE++ [33]	U	15.32	0.56	0.33	18.65	0.81	0.28	15.32	0.49	0.33	0.01
RUAS [40]	U	16.40	0.49	0.27	13.21	0.72	0.43	14.31	0.48	0.47	0.003
SCI [44]	U	14.78	0.52	0.33	15.94	0.78	0.51	15.24	0.42	0.45	0.0003
PairLIE [18]	U	19.46	0.73	0.24	21.23	0.83	0.22	17.59	0.49	0.32	0.34
NeRCo [61]	U	19.81	0.73	0.24	20.73	0.83	0.23	18.82	0.51	0.32	23.30
CLIP-LIT [36]	U	12.39	0.49	0.38	13.70	0.72	0.30	13.46	0.40	0.35	0.28
★ **DPLUT (Ours)**	U	20.66	0.74	0.22	21.27	0.84	0.21	18.91	0.53	0.28	**0.078**

and reverse steps. As illustrated in Fig. 3(b), we first apply the diffusion forward process on $\hat{I}$ to sample I_t, which can be described as:

$$q\left(I_t \mid \hat{I}\right) = \mathcal{N}\left(I_t; \sqrt{\bar{\alpha}_t}\hat{I}, \left(1-\bar{\alpha}_t\right)\hat{I}\right), \tag{12}$$

where $t = 0, 1, ..., T$-1, T is the total number of iterations, and I_t is the noisy image at time-step t. $\mathcal{N}$ represents the Gaussian distribution. $\bar{\alpha}_t = \prod_{t=0}^{T} \alpha_i$, where $\alpha_t = 1 - \beta_t$ and β_t is the predefined scale factor. After obtaining I_t, the reverse process infers a noise-free sample Y via iterative refinement, expressed as:

$$I_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{I_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta\left(I_t\right)}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\epsilon_\theta\left(I_t\right), \tag{13}$$

where $t = T, T$-1, ..., 1, $\epsilon_\theta\left(\cdot\right)$ is the noise estimator. Note that, benefiting from the strong generative prior of the diffusion model, the recovered sample Y exhibits less interference of noise and artifacts. Hence, we use Y as the pseudo-reference to supervise the NLUT learning. The introduced diffusion prior can be expressed as:

$$\mathcal{L}_{Diff} = \left\| Y - \hat{Y} \right\|_1, \tag{14}$$

where $\|\cdot\|_1$ is the ℓ_1 regularization term. Notably, we only employ the diffusion model in the NLUT learning stage.

5. Experiments

5.1. Implementation Details

All experiments are conducted on a single Titan RTX GPU, and the PyTorch framework is used to construct our networks. We employ an Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$, batch size is set to 1. The training iterations of LLUT and NLTU are set to 200 and 300, respectively. The learning rates of LLUT and NLUT are $1e^{-4}$ and $1e^{-5}$, respectively. The total number of curve steps for illumination enhancement is set to $n = 8$. We utilize the pre-trained diffusion model on ImageNet [17] and employ the implicit sampling strategy (DDIM) [48]. The total number of DDIM iteration steps is set to 100. We select the final 4 steps to implement the noise addition and removal process. The sizes of LLUT and NLUT are set to 9 and 17, respectively.

5.2. Datasets

In order to validate the effectiveness of the proposed method, we use low-light images from LOL [56] and SICE-Part2 [3] to train and test the network. The LOL dataset is officially divided into two parts, i.e., 485 low-light images for training and 15 low-light images for testing. SICE consists of 224 normal-light images and 783 low-light images. Each normal-light image corresponds to 2∼4 low-light images. We use the first 50 normal-light images and corresponding 150 low-light images for testing and the rest 633 low-light images for training. For a more convincing comparison, we further extend evaluations on the LSRW dataset [22], which includes 1000 pairs for training and 50 ones for testing.

5.3. Comparison with the State-of-the-Art

For a more comprehensive analysis, DPLUT is compared with 19 state-of-the-art LIE methods, which can be divided into the following three categories: traditional methods (SDD [23], LECARM [46]), supervised approaches (MBLLEN [43], RetinexNet [56], DSLR [37],

Table 3. Runtime (ms) comparison between our approach and SOTA methods on different resolutions. All methods are tested on one Titan RTX GPU. OOM means "Out of Memory".

Resolution	640 × 480	1920 × 1080	3840 × 2160
MBLLEN [43]	34.7	259.4	1045.1
RetinexNet [56]	20.4	139.1	550.4
DRBN [62]	15.3	127.5	550.4
3DLUT [68]	0.4	0.9	3.9
Bread [26]	16.9	148.8	683.5
CUE [71]	31.3	228.5	1222.8
DiffLL [29]	152.7	1172.6	3579.1
Retinexformer [4]	54.8	383.4	OOM
EnlightenGAN [30]	5.1	38.9	OOM
ZeroDCE [19]	1.9	18.9	91.7
ZeroDCE++ [33]	0.6	1.6	10.5
RUAS [40]	2.2	16.8	85.2
SCI [44]	0.4	1.4	10.1
PairLIE [18]	11.3	78.9	316.5
NeRCo [61]	236.4	OOM	OOM
CLIP-LIT [36]	9.6	86.1	347.6
⋆ **DPLUT (Ours)**	**6.3**	**6.6**	**19.8**

DRBN [62], 3DLUT [68], Bread [26], CUE [71], DiffLL [29], Retinexformer [4]), and unsupervised methods (EnlightenGAN [30], ZeroDCE [19], ZeroDCE++ [33], RUAS [40], SCI [44], PairLIE [18], NeRCo [61], and CLIP-LIT [36]). Note that the results of all those methods are reproduced by using the official codes with recommended parameters.

Quantitative Comparisons. We employ three full-reference metrics, i.e., peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [54], and learned perceptual image patch similarity (LPIPS) [70] to objectively evaluate the performance of each method. A higher PSNR/SSIM score indicates the result is closer to the reference. A lower LPIPS value denotes better enhancement performance. Tab. 2 reports the average assessment metrics on three datasets. The best results of supervised and unsupervised ones are highlighted in blue and red, respectively. The results show that DPLUT achieves the best performance among all unsupervised methods. Note that DPLUT performs on par with some supervised approaches, which demonstrates the effectiveness of our solution.

Visual Comparisons. For a more intuitive comparison, we further provide visual comparisons with other advanced algorithms in Figs 8-10. One can observe that DPLUT achieves visually pleasing results in terms of brightness, color, contrast, and naturalness. While other methods fail to cope with the extreme black light. Additionally, DPLUT can successfully suppress sensor noise in dark regions, and the result is clear and natural. In contrast, the competitors either amplify noise or are unable to correct color and contrast, leading to poor visual quality.

Inference Time Comparisons. Apart from the superior enhancement performance, another important advantage of our method is its efficiency. In this subsection, we report the inference time of three different resolutions, including 480P (640 × 480), 1080P (1920 × 1080), and 4K (3840 × 2160). For fair comparisons, we run all inference steps on a single Titan RTX GPU. Notably, we use the API call $torch.cuda.synchronize()$ to obtain precise feed-forward runtime. For each resolution, we record the average inference time on 100 images. In Tab. 3, as can be seen, DPLUT can handle all resolutions and the inference speed is considerably fast, especially for 4K images. EnlightenGAN [30] and ZeroDCE [19] are faster than DPLUT at 480P resolution. However, as the image resolution increases, their inference speed decreases dramatically. Significantly, EnlightenGAN and ZeroDCE cannot handle 4K low-light images. Although 3DLUT [68], SCI [44] and ZeroDCE++ [33]) outperform our method in speed, their enhancement performance is inferior to DPLUT. The above analysis validates the superb performance and practicality of our approach.

5.4. Ablation study

To understand the role of different components of our approach, we conduct several ablation studies on LOL [56].

Size of Lookup Tables. (1) We first explore the influence of the size of LLUT. As shown in Fig. 6(a), increasing the size of LLUT will improve the performance continuously until the size is close to 9-point. Compared with existing learning-based LUTs (e.g., 33-point or 64-point LUTs [39, 50, 68]), the LUT size of our solution is relatively small. One potential explanation is that combining curve mapping and LUTs can promote the effectiveness of the mapping function. (2) We investigate the impacts of the NLUT size. As shown in Fig. 6(b), enlarging the size of the NLUT yields limited performance gains but significantly increases the number of parameters, especially as the size exceeds 17. Such a phenomenon suggests the capacity redundancy of the NLUT. Consequently, the size of NLUT is set to 17 to better balance the performance and computational efficiency.

Impact of the generator. The generator of LUT is responsible for providing a coarse analysis of the input image. Here, we investigate the impact of the generator via ablating the network width, i.e., the hyper-parameter k. In Fig. 6(c-d), the ablation results indicate that increasing the width of the generator network does not always improve the performance. In contrast, it might increase capacity redundancy and training difficulties. Considering the trade-off between performance and memory footprint, we set $k = 8$ for LLUT and NLUT.

Effectiveness of the Noise-aware Map. The noise-aware weight map is estimated by a lightweight predictor network. Its architecture is listed in supplementary material. We analyze the impact of the noise-aware weight map by adjusting the network width. As shown in Fig. 6(e), enlarging the width of the predictor network will improve its capability in noise removal. Meanwhile, the number of parameters also increases. These ablation results confirm

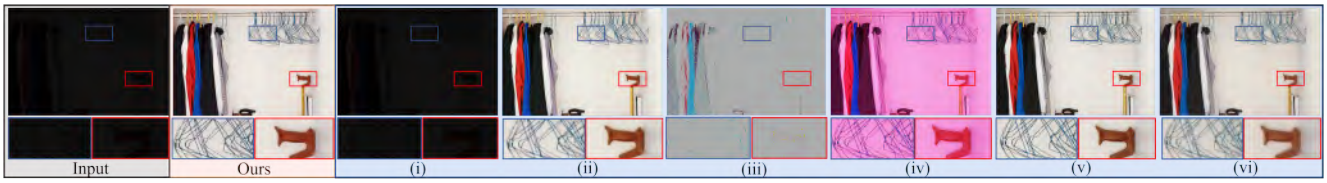

Figure 5. Visual comparisons of the ablation study. The full model achieves the best performance.

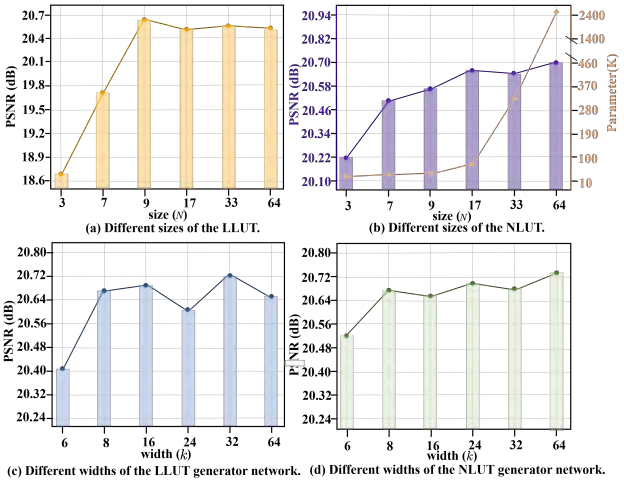

Figure 6. Ablation studies on different sizes of the LLUT and NLUT (a-b) and different widths of the LLUT and NLUT generator network (c-d).

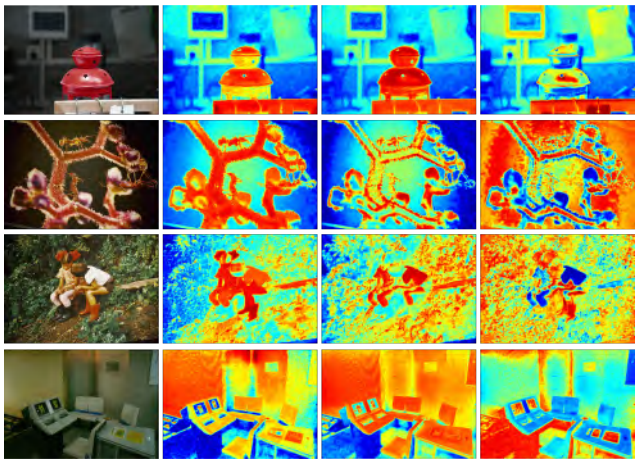

Figure 7. Visual results of the noise-aware weight map. In each row, the first is the coarse normal-light image, and the other three are visualizations for different channels. Red pixels indicate more activation, and blue pixels indicate less activation.

the effectiveness of the noise-aware weight map, which can assist the NLUT in suppressing noise. Given the trade-off between memory footprint and denoising performance, the width of the predictor network is configured to 16. Furthermore, we also provide the visualization of the noise-aware weight map in Fig. 7.

Impact of Iterations Steps. We investigate the influence of the iteration steps of adding and removing noise. In Tab. 4, we perform different numbers of iterations on the coarse enhanced samples to generate different versions of pseudo-references. One can see that fewer iteration steps lead to higher distortion metrics (i.e., PSNR and SSIM), whereas more iteration steps yield improved perception results (i.e., LPIPS) at the cost of increased processing time. This observation aligns with the findings in [57]. By carefully balancing these metrics and the sampling time, we set the iteration steps as 4 to generate pseudo-references for the NLUT learning.

Variants of the supervision. We further conduct ablation studies to verify the effectiveness of the loss functions. Concretely, we have tested the following four variations over the original setting: (i) without the exposure loss. (ii) without the structural consistency loss. (iii) without the smoothing loss. (iv) without the color loss. (v) without the NLUT, i.e., only the LLUT. (vi) without prior knowledge of the pre-trained diffusion model and replace it with existing denoising method [69]. Results are listed in Tab. 5 and Fig. 5. We have the following observations: 1) removing the exposure loss fails to recover the low-light regions and the objective measures degrade significantly. 2) The structural consistency loss can promote the naturalness of the enhanced image. 3) Removing the smoothness loss hampers the correlations between neighboring regions, leading to obvious artifacts. 4) The absence of color loss results in serious color distortion. 5) Without the NLUT, i.e., only the LLUT, the enhanced image exhibits obvious noise and artifacts. In contrast, our full model generates clean and natural predictions, demonstrating the effectiveness of our approach. 6) Compared with the advanced diffusion model, training the NLUT with the existing denoising method [69] is inevitably constrained in terms of robustness and generalizability, yielding only suboptimal results as they either depend on synthetic datasets or necessitate hand-crafted assumptions.

6. Conclusions

In this paper, we introduce a novel unsupervised LIE framework based on lookup tables and diffusion priors (DPLUT) to achieve effective and efficient low-light image recovery. Two core components are devised to equip the framework, i.e., a light adjustment lookup table (LLUT) and a noise suppression lookup table (NLUT). Concretely, LLUT is de-

Table 4. Ablation study on different iteration numbers.

Iterations	PSNR↑	SSIM↑	LPIPS↓	Time (s)
2	20.01	0.712	0.242	0.764
3	20.21	0.733	0.231	1.095
4	**20.66**	**0.744**	**0.222**	**1.465**
5	20.57	0.739	0.215	1.795
10	20.19	0.712	0.211	3.389
25	19.90	0.683	0.208	8.709
30	19.64	0.673	0.204	10.482

Table 5. Quantitative results of ablation studies on LOL.

Variants	PSNR↑	SSIM↑	LPIPS↓
(i)	8.31	0.24	0.56
(ii)	18.21	0.60	0.38
(iii)	17.79	0.59	0.39
(iv)	17.49	0.56	0.40
(v)	19.89	0.61	0.34
(vi)	20.11	0.64	0.30
⋆ **DPLUT (Full)**	**20.66**	**0.74**	**0.22**

signed to predict pixel-wise curve parameters for the dynamic range adjustment. NLUT is employed to remove the amplified noise after light brightening. Both LLUT and NLUT are trained in an unsupervised manner with a set of unsupervised losses and prior knowledge from a pretrained diffusion model, respectively. Extensive experimentation validates that our novel framework can enhance 4K low-light images in real-time and surpasses contemporary methods on three challenging benchmark datasets. In the future, we plan to optimize two lookup tables jointly to further promote the performance. Besides, we intend to apply our solution for different vision tasks.

References

[1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE transactions on consumer electronics*, 53(2):593–600, 2007. 2

[2] Jinbin Bai, Zhen Dong, Aosong Feng, Xiao Zhang, Tian Ye, Kaicheng Zhou, and Mike Zheng Shou. Integrating view conditions for image synthesis. *arXiv preprint arXiv:2310.16002*, 2023. 3

[3] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 6

[4] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12504–12513, 2023. 1, 3, 6, 7

[5] Shidong Cao, Wenhao Chai, Shengyu Hao, Yanting Zhang, Hangyue Chen, and Gaoang Wang. Difffashion: Reference-based fashion design with structure-aware transfer by diffusion models. *IEEE Transactions on Multimedia*, 2023. 3

[6] Wenhao Chai, Xun Guo, Gaoang Wang, and Yan Lu. Stablevideo: Text-driven consistency-aware diffusion video editing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23040–23050, 2023. 3

[7] Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2, 2022. 3

[8] Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Yi Ye, and Erkang Chen. Msp-former: Multi-scale projection transformer for single image desnowing. *arXiv preprint arXiv:2207.05621*, 2022. 3

[9] Sixiang Chen, Tian Ye, Jinbin Bai, Erkang Chen, Jun Shi, and Lei Zhu. 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*, pages 13106–13117, 2023. 3

[10] Sixiang Chen, Tian Ye, Yun Liu, Jinbin Bai, Haoyu Chen, Yunlong Lin, Jun Shi, and Erkang Chen. Cplformer: Cross-scale prototype learning transformer for image snow removal. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 4228–4239, 2023. 3

[11] Sixiang Chen, Tian Ye, Jun Shi, Yun Liu, JingXia Jiang, Erkang Chen, and Peng Chen. 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)*, pages 1–5. IEEE, 2023. 3

[12] Heng-Da Cheng and XJ Shi. A simple and effective histogram equalization approach to image enhancement. *Digital signal processing*, 14(2):158–170, 2004. 2

[13] Marcos V Conde, Steven McDonagh, Matteo Maggioni, Ales Leonardis, and Eduardo Pérez-Pellitero. Model-based image signal processors via learnable dictionaries. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 481–489, 2022. 3

[14] Marcos V Conde, Javier Vazquez-Corral, Michael S Brown, and Radu Timofte. Nilut: Conditional neural implicit 3d lookup tables for image enhancement. *arXiv preprint arXiv:2306.11920*, 2023. 2

[15] Wenyan Cong, Xinhao Tao, Li Niu, Jing Liang, Xuesong Gao, Qihao Sun, and Liqing Zhang. High-resolution image harmonization via collaborative dual transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18470–18479, 2022. 3

[16] Mauricio Delbracio, Damien Kelly, Michael S Brown, and Peyman Milanfar. Mobile computational photography: A tour. *Annual Review of Vision Science*, 7:571–604, 2021. 3

[17] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 6

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

[19] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1777–1786, 2020. 2, 4, 6, 7

[20] Jianshu Guo, Wenhao Chai, Jie Deng, Hsiang-Wei Huang, Tian Ye, Yichen Xu, Jiawei Zhang, Jenq-Neng Hwang, and Gaoang Wang. Versat2i: Improving text-to-image models with versatile reward. *arXiv preprint arXiv:2403.18493*, 2024. 3

[21] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 2

[22] Jiang Hai, Zhu Xuan, Ren Yang, Yutong Hao, Fengzhu Zou, Fang Lin, and Songchen Han. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712, 2023. 1, 6

[23] Shijie Hao, Xu Han, Yanrong Guo, Xin Xu, and Meng Wang. Low-light image enhancement with semi-decoupled decomposition. *IEEE Transactions on Multimedia*, 22(12):3025–3038, 2020. 6

[24] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3

[25] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *arXiv preprint arXiv:2310.17577*, 2023. 2, 3

[26] Qiming Hu and Xiaojie Guo. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131:48–66, 2021. 6, 7

[27] Jingjia Huang, Ge Meng, Yingying Wang, Yunlong Lin, Yue Huang, and Xinghao Ding. 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)*, pages 268–279. Springer, 2023. 3

[28] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012. 2

[29] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 6, 7

[30] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 6, 7

[31] Hakki Can Karaimer and Michael S Brown. A software platform for manipulating the camera imaging pipeline. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 429–444. Springer, 2016. 1, 3

[32] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021. 3

[33] Chongyi Li, Chunle Guo, and Chen Change Loy. 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. 1, 6, 7

[34] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embeddingfourier for ultra-high-definition low-light image enhancement. In *ICLR*, 2023. 2, 5

[35] Jie Liang, Hui Zeng, Miaomiao Cui, Xuansong Xie, and Lei Zhang. Ppr10k: A large-scale portrait photo retouching dataset with human-region mask and group-level consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 653–661, 2021. 3

[36] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8094–8103, 2023. 6, 7

[37] Seokjae Lim and Wonjun Kim. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23:4272–4284, 2020. 6

[38] Yunlong Lin, Zhenqi Fu, Ge Meng, Yingying Wang, Yuhang Dong, Linyu Fan, Hedeng Yu, and Xinghao Ding. Domain-irrelevant feature learning for generalizable pan-sharpening. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3287–3296, 2023. 3

[39] Chengxu Liu, Huan Yang, Jianlong Fu, and Xueming Qian. 4d lut: learnable context-aware 4d lookup table for image enhancement. *IEEE Transactions on Image Processing*, 32: 4742–4756, 2023. 2, 3, 7

[40] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10561–10570, 2021. 6, 7

[41] Yun Liu, Zhongsheng Yan, Sixiang Chen, Tian Ye, Wenqi Ren, and Erkang Chen. Nighthazeformer: Single nighttime haze removal using prior query transformer. *arXiv preprint arXiv:2305.09533*, 2023. 2

[42] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 2

[43] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *British Machine Vision Conference (BMVC)*, pages 1–13, 2018. 2, 6, 7

[44] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image

enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5637–5646, 2022. 1, 2, 6, 7

[45] Shanto Rahman, Md Mostafijur Rahman, Mohammad Abdullah-Al-Wadud, Golam Dastegir Al-Quaderi, and Mohammad Shoyaib. An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(1):1–13, 2016. 2

[46] Yurui Ren, Zhenqiang Ying, Thomas H Li, and Ge Li. Lecarm: Low-light image enhancement using the camera response model. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(4):968–981, 2018. 6

[47] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015. 3

[48] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 3, 6

[49] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 3

[50] Tao Wang, Yong Li, Jingyang Peng, Yipeng Ma, Xian Wang, Fenglong Song, and Youliang Yan. Real-time image enhancer via learnable spatial-aware 3d lookup tables. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2471–2480, 2021. 1, 3, 7

[51] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2604–2612, 2022. 2

[52] Yingying Wang, Yunlong Lin, Ge Meng, Zhenqi Fu, Yuhang Dong, Linyu Fan, Hedeng Yu, Xinghao Ding, and Yue Huang. Learning high-frequency feature enhancement and alignment for pan-sharpening. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 358–367, 2023. 3

[53] Yingying Wang, Xuanhua He, Yuhang Dong, Yunlong Lin, Yue Huang, and Xinghao Ding. Cross-modality interaction network for pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 3

[54] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7

[55] Zhi-Guo Wang, Zhi-Hu Liang, and Chun-Liang Liu. A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for pdp. *Displays*, 30(3):133–139, 2009. 2

[56] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference (BMVC)*, pages 1–12, 2018. 6, 7

[57] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022. 8

[58] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17714–17724, 2022. 2

[59] Canqian Yang, Meiguang Jin, Xu Jia, Yi Xu, and Ying Chen. Adaint: Learning adaptive intervals for 3d lookup tables on real-time image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17522–17531, 2022. 3

[60] Canqian Yang, Meiguang Jin, Yi Xu, Rui Zhang, Ying Chen, and Huaida Liu. Seplut: Separable image-adaptive lookup tables for real-time image enhancement. In *European Conference on Computer Vision*, pages 201–217. Springer, 2022. 1, 3

[61] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12918–12927, 2023. 2, 6, 7

[62] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3060–3069, 2020. 6, 7

[63] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. In *European Conference on Computer Vision*, pages 130–145. Springer, 2022. 3

[64] Tian Ye, Sixiang Chen, Jinbin Bai, Jun Shi, Chenghao Xue, Jingxia Jiang, Junjie Yin, Erkang Chen, and Yun Liu. Adverse weather removal with codebook priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12653–12664, 2023. 3

[65] Tian Ye, Sixiang Chen, Yun Liu, Wenhao Chai, Jinbin Bai, Wenbin Zou, Yunchen Zhang, Mingchao Jiang, Erkang Chen, and Chenghao Xue. Sequential affinity learning for video restoration. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 4147–4156, 2023.

[66] Tian Ye, Sixiang Chen, Wenhao Chai, Zhaohu Xing, Jing Qin, Ge Lin, and Lei Zhu. Learning diffusion texture priors for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2524–2534, 2024. 3

[67] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12302–12311, 2023. 3

[68] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2020. 1, 2, 3, 6, 7

[69] Dan Zhang, Fangfang Zhou, Yuwen Jiang, and Zhengming Fu. Mm-bsn: Self-supervised image denoising for real-world

with multi-mask based on blind-spot network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4188–4197, 2023. 8

[70] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018. 7

[71] Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Feng Zhao. Empowering low-light image enhancer through customized learnable priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12559–12569, 2023. 1, 2, 6, 7

[72] Wenbin Zou, Hongxia Gao, Tian Ye, Liang Chen, Weipeng Yang, Shasha Huang, Hongsheng Chen, and Sixiang Chen. Vqcnir: Clearer night image restoration with vector-quantized codebook. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7873–7881, 2024. 2

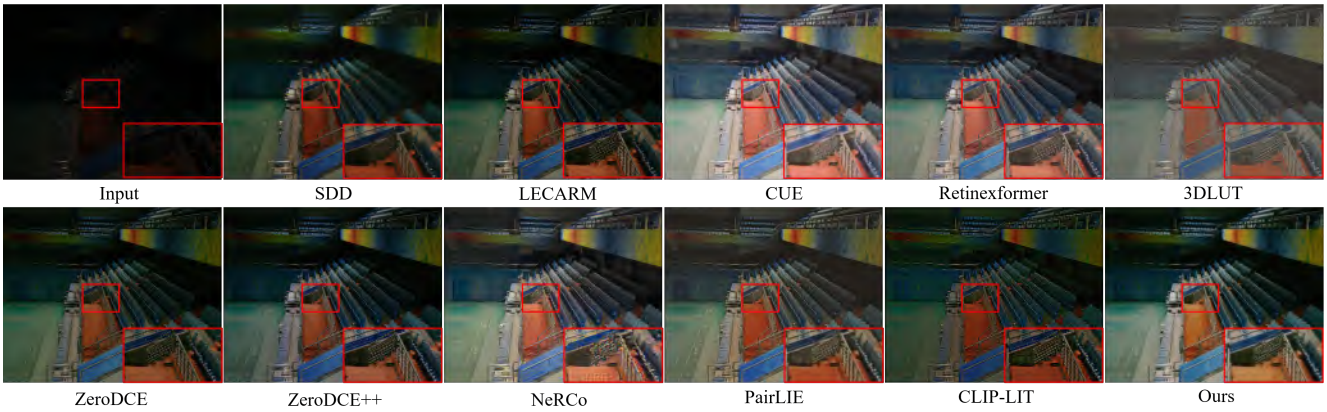

Figure 8. Qualitative comparison of our method and competitive methods on the LOL dataset.

Figure 9. Qualitative comparison of our method and competitive methods on the LSRW dataset.

Figure 10. Qualitative comparison of our method and competitive methods on the MEF dataset.