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

Yunlong Lin[1,2*], Zhenqi Fu[3*], Kairun Wen[1,2], Tian Ye[4], Sixiang Chen[4], Ge Meng[1,2], Yingying Wang[1], Chui Kong[5], Yue Huang[1,2], Xiaotong Tu[1,2], and Xinghao Ding[1,2†]

[1]Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, China
[2]School of Informatics, Xiamen University, China
[3]Tsinghua University, China
[4]The Hong Kong University of Science and Technology (Guangzhou), China
[5]Fudan University, China
dxh@xmu.edu.cn

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.

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 (Cai et al. 2023; Yang et al. 2023; He et al. 2023a; Lin et al. 2024b), which can be roughly classified into efficiency- and quality-oriented methods.

Efficiency-oriented methods, such as SCI (Ma et al. 2022) and ZeroDCE++ (Li, Guo, and Loy 2021), 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 (Cai et al. 2023) and CUE (Zheng et al. 2023) 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. In Fig. 1, existing quality-oriented approaches cannot process 4K resolution images in real-time.

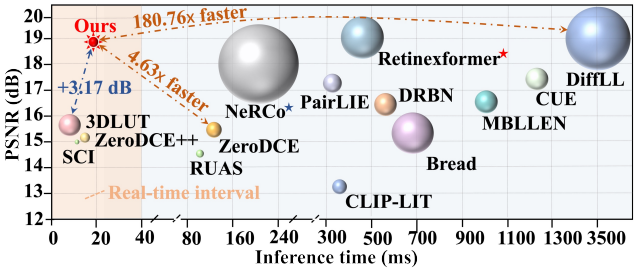

Figure 1: Comparisons of performance and efficiency. The average PSNR is evaluated on LSRW (Hai et al. 2023), and inference time is evaluated on 4K (3840×2160) resolution with a single Titan RTX GPU. Our approach obtains the highest PSNR and can process 4K low-light images in real-time. Note that * and * 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) 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 (Zeng et al. 2020; Wang et al. 2021) are significantly more expressive and

*Equal contribution
†Corresponding author

have achieved 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 (Zeng et al. 2020; Wang et al. 2021). ***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 (Zeng et al. 2020). ***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.

Related Work

Low-light Image Enhancement

Enhancing images in low-light conditions has been a longstanding issue and significant advancements over the decades have made it increasingly effective in supporting downstream tasks (Cai et al. 2023; Yang et al. 2024, 2022; Li et al. 2024). 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 (Rahman et al. 2016; Jin, Yang, and Tan 2022). For example, Wang et al. (Wang, Liang, and Liu 2009) enhanced the visibility and contrast via gamma correction and dynamic contrast ratio improvement. Guo et al. (Guo, Li, and Ling 2017) proposed to refine the initial estimated illumination map by imposing a structure prior. Guo et al. (Guo et al. 2020) presented a reference-free LIE algorithm based on curve estimation, which can effectively perform mapping within a wide dynamic range. Ma et al. (Ma et al. 2022) established a cascaded illumination estimation process to achieve fast and robust LIE in complex scenarios. However, the enhancement performance of current efficiency-oriented approaches is greatly inferior to quality-oriented methods (Lore, Akintayo, and Sarkar 2017; Hou et al. 2023; Xu et al. 2022; Zheng et al. 2023; Yang et al. 2023; Lin et al. 2024b). Lore et al. (Lore, Akintayo, and Sarkar 2017) designed a stacked sparse denoising auto-encoder to enhance low-light images. Xu et al. (Xu et al. 2022) incorporated the signal-to-noise ratio (SNR) prior to achieving spatial-varying LIE. Cai et al. (Cai et al. 2023) designed a sophisticated transformer-based algorithm for LIE. Hou et al. (Hou et al. 2023) devised a diffusion-based framework and introduced a global structure-aware regularization to preserve the image's details and textures. Yi et al. (Yi et al. 2023) 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 prepossessing task, the practicality of quality-oriented methods is limited.

LUTs for Image Enhancement

The lookup tables (LUTs) are commonly used in ISPs, 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 (Zeng et al. 2020; Cong et al. 2022; Wang et al. 2021). For instance, Zeng et al. (Zeng et al. 2020) 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. (Wang et al. 2021) proposed a spatially-aware LUT that considers the global and local information. Cong et al. (Cong et al. 2022) embedded the LUT-based sub-module in their network for efficient processing of high-resolution images.

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 $i, j, k = 0, 1, \ldots, N$-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:

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

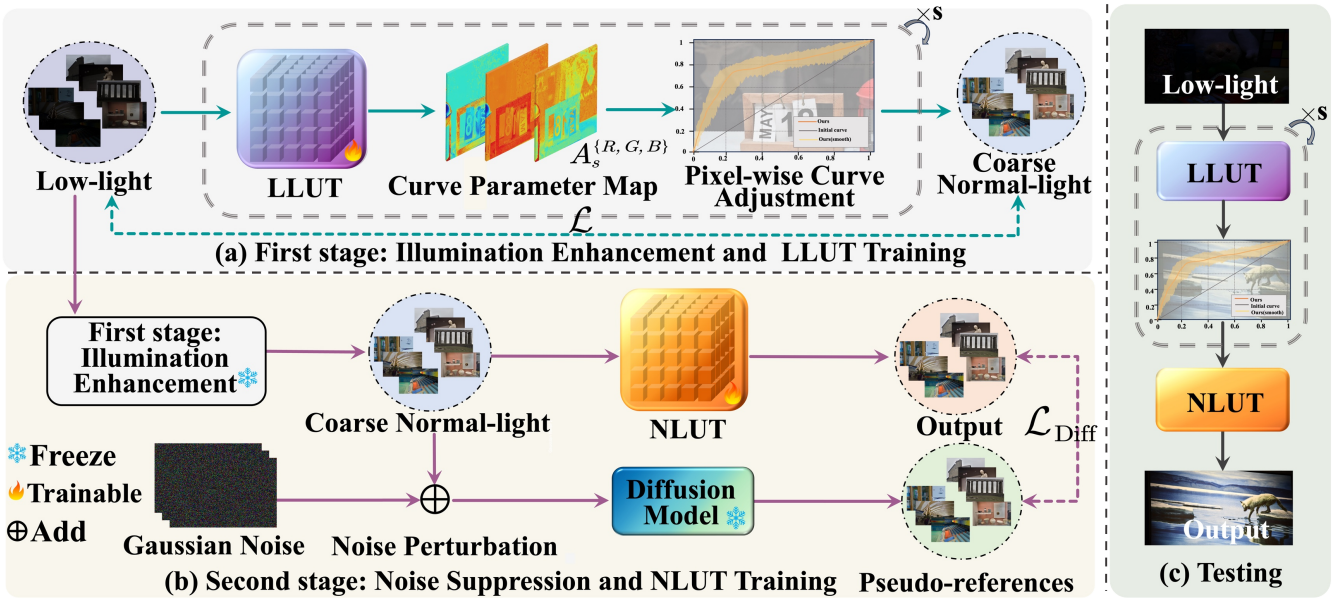

Figure 2: The overall framework of our proposed DPLUT. In the training phase, DPLUT involves two main stages. (a) In the first stage, we learn a light adjustment lookup table (LLUT), which estimates pixel-wise curve parameters for yielding coarse normal-light images. (b) In the second stage, we learn 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.

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 (Guo et al. 2024; Chai et al. 2023), and it enlightens other tasks like image restoration (Ye et al. 2024; Chen et al. 2025), image fusion (Lin et al. 2023; Wang et al. 2023, 2024; He et al. 2023b; Lin et al. 2024a), dehazing (Ye et al. 2022; Jin, Yang, and Tan 2022), and desnowing (Chen et al. 2023). 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 (Song et al. 2020). 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 (Song et al. 2020) related with an optimal MSE denoiser. 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):

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

where $\epsilon \sim \mathcal{N}(0, \mathrm{I})$, $w(t)$ is a weighting function. In practice, setting $w(t) = 1$ delivers good performance. Sampling from a diffusion model can be either stochastic or deterministic. 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.

Methodology

The overall framework of our proposed DPLUT involves two main stages, as illustrated in Fig. 2. 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

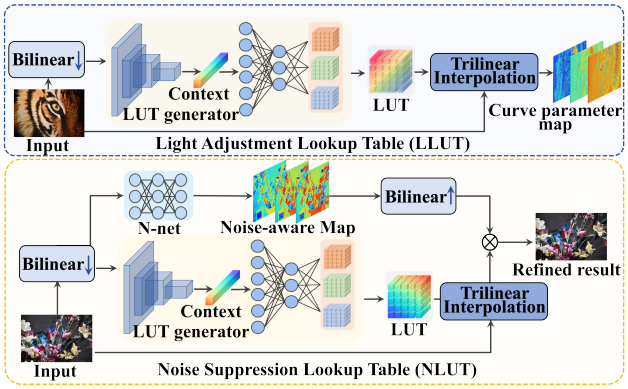

Figure 3: The architecture of our two key components: a light adjustment lookup table (LLUT) and a noise suppression lookup table (NLUT).

distributions and achieve real-time and high-quality image enhancement.

Light Adjustment Lookup Table

As depicted in Fig. 2(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 $I\left(x\right) \in \mathbb{R}^{H\times W\times 3}$, 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. 3(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 $f_{3D}\left(\cdot\right)$. The detailed architecture can be found in supplementary material. Given the predicted LLUT, the estimated pixel-wise curve parameter map $A(x)$ for a specific image $I\left(x\right)$ can be formulated as:

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

Then, we recurrently apply the pixel-wise adjustment curve (Guo et al. 2020) to obtain the coarse normal-light sample $\hat{I}\left(x\right) \in \mathbb{R}^{H\times W\times 3}$. 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.

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.

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. 2(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. 3(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:

Method	Type	LOL			SICE			LSRW			Params (M)
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
SDD (Hao et al. 2020)	T	13.34	0.63	0.74	15.34	0.73	0.26	14.70	0.49	0.41	-
LECARM (Ren et al. 2018)	T	14.40	0.54	0.32	18.59	0.78	0.26	15.33	0.42	0.32	-
MBLLEN (Lv et al. 2018)	S	15.25	0.70	0.32	18.41	0.73	0.31	16.87	0.51	0.45	0.45
RetinexNet (Wei et al. 2018)	S	17.60	0.64	0.38	19.57	0.78	0.27	15.58	0.41	0.39	0.84
DSLR (Lim and Kim 2020)	S	15.20	0.59	0.32	14.32	0.68	0.38	15.21	0.44	0.38	14.93
DRBN (Yang et al. 2020)	S	19.67	0.82	0.16	18.73	0.78	0.28	16.72	0.51	0.39	0.58
3DLUT (Zeng et al. 2020)	S	16.36	0.64	0.35	15.53	0.64	0.38	15.74	0.48	0.43	0.59
Bread (Hu and Guo 2021)	S	22.95	0.83	0.15	17.28	0.80	0.25	16.06	0.53	0.36	2.12
CUE (Zheng et al. 2023)	S	22.67	0.79	0.20	20.06	0.82	0.24	18.19	0.52	0.33	0.26
DiffLL (Jiang et al. 2023)	S	26.19	0.85	0.11	21.33	0.84	0.22	19.27	0.55	0.30	22.05
Retinexformer (Cai et al. 2023)	S	25.15	0.84	0.13	22.32	0.85	0.20	19.23	0.54	0.31	1.61
EnlightenGAN (Jiang et al. 2021)	U	17.48	0.65	0.32	18.73	0.82	0.23	17.05	0.46	0.33	8.64
ZeroDCE (Guo et al. 2020)	U	14.86	0.55	0.33	18.67	0.80	0.26	15.84	0.45	0.31	0.079
ZeroDCE++ (Li, Guo, and Loy 2021)	U	15.32	0.56	0.33	18.65	0.81	0.28	15.32	0.49	0.33	0.01
RUAS (Liu et al. 2021)	U	16.40	0.49	0.27	13.21	0.72	0.43	14.31	0.48	0.47	0.003
SCI (Ma et al. 2022)	U	14.78	0.52	0.33	15.94	0.78	0.51	15.24	0.42	0.45	0.0003
PairLIE (Fu et al. 2023)	U	19.46	0.73	0.24	21.23	0.83	0.22	17.59	0.49	0.32	0.34
NeRCo (Yang et al. 2023)	U	19.81	0.73	0.24	20.73	0.83	0.23	18.82	0.51	0.32	23.30
CLIP-LIT (Liang et al. 2023)	U	12.39	0.49	0.38	13.70	0.72	0.30	13.46	0.40	0.35	0.28
LightenDiffusion (Jiang et al. 2024)	U	20.44	0.81	0.19	20.59	0.83	0.22	18.54	0.54	0.31	27.81
★ **DPLUT (Ours)**	U	20.66	**0.74**	**0.22**	21.27	0.84	0.21	18.91	**0.53**	0.28	**0.078**

Table 1: 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.

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

where NLUT is generated by the LUT generator module $f_{3D}(\cdot)$, as shown in supplementary material. $m = \{m_{h,w} \mid h \in \mathbb{R}^{H-1}, w \in \mathbb{R}^{W-1}\}$ is a noise-aware pixel-wise weight map for NLUT at location (h, w), which is estimated by a lightweight network $\gamma(\cdot)$. Specifically, $\gamma(\cdot)$ 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 and reverse steps. As illustrated in Fig. 2(b), we first apply the diffusion forward process on $\hat{I}$ to sample I_t, which can be described as:

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

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:

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

where $t = T, T$-1, ..., 1, $\epsilon_\theta(\cdot)$ 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:

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

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

Experiments

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 and employ the implicit sampling strategy (DDIM). 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.

Datasets

In order to validate the effectiveness of the proposed method, we use low-light images from LOL (Wei et al. 2018) and SICE-Part2 (Cai, Gu, and Zhang 2018) 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 (Hai et al. 2023), which includes 1000 pairs for training and 50 ones for testing.

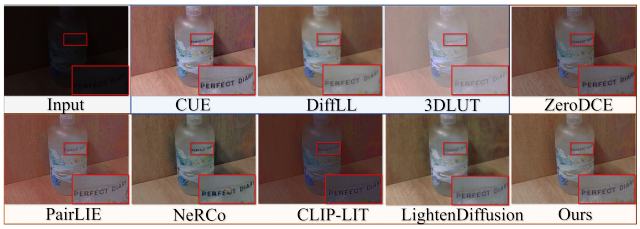

Figure 4: Visual comparisons on LSRW. The proposed method achieves visually pleasing results in terms of brightness, color, contrast, and naturalness.

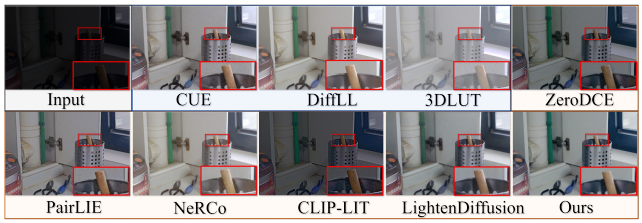

Figure 5: Performance comparisons in noise suppression. Our method achieves remarkably higher quality among unsupervised methods with less noise and artifacts.

Resolution	640 × 480	1920 × 1080	3840 × 2160
MBLLEN (Lv et al. 2018)	34.7	259.4	1045.1
RetinexNet (Wei et al. 2018)	20.4	139.1	550.4
DRBN (Yang et al. 2020)	15.3	127.5	550.4
3DLUT (Zeng et al. 2020)	0.4	0.9	3.9
Bread (Hu and Guo 2021)	16.9	148.8	683.5
CUE (Zheng et al. 2023)	31.3	228.5	1222.8
DiffLL (Jiang et al. 2023)	152.7	1172.6	3579.1
Retinexformer (Cai et al. 2023)	54.8	383.4	OOM
EnlightenGAN (Jiang et al. 2021)	5.1	38.9	OOM
ZeroDCE (Guo et al. 2020)	1.9	18.9	91.7
ZeroDCE++ (Li, Guo, and Loy 2021)	0.6	1.6	10.5
RUAS (Liu et al. 2021)	2.2	16.8	85.2
SCI (Ma et al. 2022)	0.4	1.4	10.1
PairLIE (Fu et al. 2023)	11.3	78.9	316.5
NeRCo (Yang et al. 2023)	236.4	OOM	OOM
CLIP-LIT (Liang et al. 2023)	9.6	86.1	347.6
LightenDiffusion (Jiang et al. 2024)	184.4	1019.7	4107.2
⋆ **DPLUT (Ours)**	**6.3**	**6.6**	**19.8**

Table 2: The runtime (ms) comparison at different resolutions. OOM means "Out of Memory".

Comparison with the State-of-the-Art

For a more comprehensive analysis, DPLUT is compared with 20 state-of-the-art LIE methods, which can be divided into the following three categories: traditional methods (SDD (Hao et al. 2020), LECARM (Ren et al. 2018)), supervised approaches (MBLLEN (Lv et al. 2018), RetinexNet (Wei et al. 2018), DSLR (Lim and Kim 2020), DRBN (Yang et al. 2020), 3DLUT (Zeng et al. 2020), Bread (Hu and Guo 2021), CUE (Zheng et al. 2023), DiffLL (Jiang et al. 2023), Retinexformer (Cai et al. 2023)), and unsupervised methods (EnlightenGAN (Jiang et al. 2021), ZeroDCE (Guo et al. 2020), ZeroDCE++ (Li, Guo, and Loy 2021), RUAS (Liu et al. 2021), SCI (Ma et al. 2022), PairLIE (Fu et al. 2023), NeRCo (Yang et al. 2023), CLIP-LIT (Liang et al. 2023)) and LightenDiffusion (Jiang et al. 2024). 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), and learned perceptual image patch similarity (LPIPS) 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. 1 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 Figures 4 and 5. In Fig. 4, 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. We show an example of noise suppression in Fig. 5. As can be seen, 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. For each resolution, we record the average inference time on 100 images. In Tab. 2, as can be seen, DPLUT can handle all resolutions and the inference speed is considerably fast, especially for 4K images. EnlightenGAN (Jiang et al. 2021) and ZeroDCE (Guo et al. 2020) 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 (Zeng et al. 2020), SCI (Ma et al. 2022) and ZeroDCE++ (Li, Guo, and Loy 2021)) outperform our method in speed, their enhancement performance is inferior to DPLUT. These analyses validate the superb performance and practicality of our approach.

Ablation study

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

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 (Zeng et al. 2020; Wang et al. 2021)), the LUT size of our solution is relatively small. One potential explanation is that combining curve mapping and LUTs can promote the effectiveness

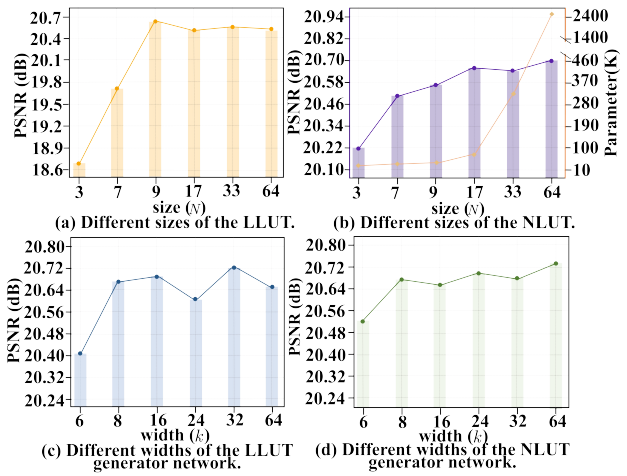

Figure 6: Ablation studies on different sizes of the LLUT and NLUT (a-b) and different widths of the LLUT and NLUT generator network (c-d).

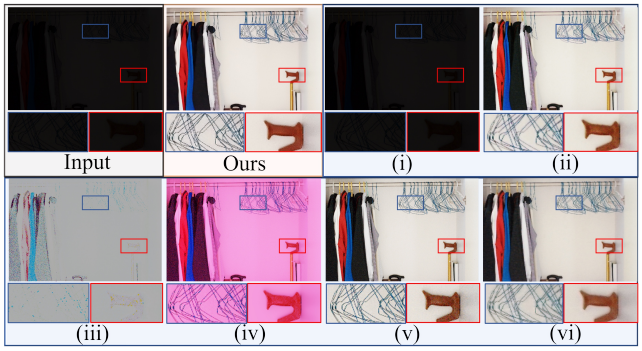

Figure 7: Visual comparisons of the ablation study. The full model achieves the best performance.

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.

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 (Zhang et al. 2023). Results are listed in Tab. 3 and Fig. 7. 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 (Zhang et al. 2023) 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.

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**

Table 3: Quantitative results of ablation studies on LOL.

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 designed 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 pre-trained 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.

Acknowledgments

The work was supported in part by the National Natural Science Foundation of China under Grant 82172033, U19B2031, 61971369, 52105126, 82272071, 62271430, and the Fundamental Research Funds for the Central Universities 20720230104

References

Cai, J.; Gu, S.; and Zhang, L. 2018. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4): 2049–2062.

Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 12504–12513.

Chai, W.; Guo, X.; Wang, G.; and Lu, Y. 2023. Stablevideo: Text-driven consistency-aware diffusion video editing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 23040–23050.

Chen, S.; Ye, T.; Liu, Y.; Bai, J.; Chen, H.; Lin, Y.; Shi, J.; and Chen, E. 2023. CPLFormer: Cross-scale Prototype Learning Transformer for Image Snow Removal. In *Proceedings of the 31st ACM International Conference on Multimedia*, 4228–4239.

Chen, S.; Ye, T.; Zhang, K.; Xing, Z.; Lin, Y.; and Zhu, L. 2025. Teaching Tailored to Talent: Adverse Weather Restoration via Prompt Pool and Depth-Anything Constraint. In *European Conference on Computer Vision*, 95–115. Springer.

Cong, W.; Tao, X.; Niu, L.; Liang, J.; Gao, X.; Sun, Q.; and Zhang, L. 2022. High-resolution image harmonization via collaborative dual transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18470–18479.

Fu, Z.; Yang, Y.; Tu, X.; Huang, Y.; Ding, X.; and Ma, K.-K. 2023. 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*, 22252–22261.

Guo, C.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1777–1786.

Guo, J.; Chai, W.; Deng, J.; Huang, H.-W.; Ye, T.; Xu, Y.; Zhang, J.; Hwang, J.-N.; and Wang, G. 2024. Versat2i: Improving text-to-image models with versatile reward. *arXiv preprint arXiv:2403.18493*.

Guo, X.; Li, Y.; and Ling, H. 2017. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Transactions on Image Processing*, 26(2): 982–993.

Hai, J.; Xuan, Z.; Yang, R.; Hao, Y.; Zou, F.; Lin, F.; and Han, S. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712.

Hao, S.; Han, X.; Guo, Y.; Xu, X.; and Wang, M. 2020. Low-Light Image Enhancement With Semi-Decoupled Decomposition. *IEEE Transactions on Multimedia*, 22(12): 3025–3038.

He, C.; Fang, C.; Zhang, Y.; Ye, T.; Li, K.; Tang, L.; Guo, Z.; Li, X.; and Farsiu, S. 2023a. Reti-diff: Illumination degradation image restoration with retinex-based latent diffusion model. *arXiv preprint arXiv:2311.11638*.

He, C.; Li, K.; Xu, G.; Zhang, Y.; Hu, R.; Guo, Z.; and Li, X. 2023b. Degradation-resistant unfolding network for heterogeneous image fusion. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12611–12621.

Hou, J.; Zhu, Z.; Hou, J.; Liu, H.; Zeng, H.; and Yuan, H. 2023. Global Structure-Aware Diffusion Process for Low-Light Image Enhancement. *arXiv preprint arXiv:2310.17577*.

Hu, Q.; and Guo, X. 2021. Low-light Image Enhancement via Breaking Down the Darkness. *International Journal of Computer Vision*, 131: 48–66.

Jiang, H.; Luo, A.; Fan, H.; Han, S.; and Liu, S. 2023. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6): 1–14.

Jiang, H.; Luo, A.; Liu, X.; Han, S.; and Liu, S. 2024. LightenDiffusion: Unsupervised Low-Light Image Enhancement with Latent-Retinex Diffusion Models. In *European Conference on Computer Vision*.

Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; and Wang, Z. 2021. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Transactions on Image Processing*, 30: 2340–2349.

Jin, Y.; Yang, W.; and Tan, R. T. 2022. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, 404–421. Springer.

Li, C.; Guo, C.; and Loy, C. C. 2021. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8): 4225–4238.

Li, C.; Liu, X.; Li, W.; Wang, C.; Liu, H.; Liu, Y.; Chen, Z.; and Yuan, Y. 2024. U-kan makes strong backbone for medical image segmentation and generation. *arXiv preprint arXiv:2406.02918*.

Liang, Z.; Li, C.; Zhou, S.; Feng, R.; and Loy, C. C. 2023. Iterative prompt learning for unsupervised backlit image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8094–8103.

Lim, S.; and Kim, W. 2020. DSLR: Deep stacked Laplacian restorer for low-light image enhancement. *IEEE Transactions on Multimedia*, 23: 4272–4284.

Lin, H.; Lin, Y.; Xia, J.; Fan, L.; Li, F.; Wang, Y.; and Ding, X. 2024a. Fusion2Void: Unsupervised Multi-focus Image Fusion Based on Image Inpainting. *IEEE Transactions on Circuits and Systems for Video Technology*.

Lin, Y.; Fu, Z.; Meng, G.; Wang, Y.; Dong, Y.; Fan, L.; Yu, H.; and Ding, X. 2023. Domain-irrelevant Feature Learning for Generalizable Pan-sharpening. In *Proceedings of the*

31st ACM International Conference on Multimedia, 3287–3296.

Lin, Y.; Ye, T.; Chen, S.; Fu, Z.; Wang, Y.; Chai, W.; Xing, Z.; Zhu, L.; and Ding, X. 2024b. AGLLDiff: Guiding Diffusion Models Towards Unsupervised Training-free Real-world Low-light Image Enhancement. *arXiv preprint arXiv:2407.14900*.

Liu, R.; Ma, L.; Zhang, J.; Fan, X.; and Luo, Z. 2021. Retinex-Inspired Unrolling With Cooperative Prior Architecture Search for Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10561–10570.

Lore, K. G.; Akintayo, A.; and Sarkar, S. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61: 650–662.

Lv, F.; Lu, F.; Wu, J.; and Lim, C. 2018. MBLLEN: Low-light image/video enhancement using CNNs. In *British Machine Vision Conference (BMVC)*, 1–13.

Ma, L.; Ma, T.; Liu, R.; Fan, X.; and Luo, Z. 2022. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5637–5646.

Rahman, S.; Rahman, M. M.; Abdullah-Al-Wadud, M.; Al-Quaderi, G. D.; and Shoyaib, M. 2016. An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(1): 1–13.

Ren, Y.; Ying, Z.; Li, T. H.; and Li, G. 2018. LECARM: Low-light image enhancement using the camera response model. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(4): 968–981.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Wang, T.; Li, Y.; Peng, J.; Ma, Y.; Wang, X.; Song, F.; and Yan, Y. 2021. Real-time image enhancer via learnable spatial-aware 3d lookup tables. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2471–2480.

Wang, Y.; He, X.; Dong, Y.; Lin, Y.; Huang, Y.; and Ding, X. 2024. Cross-Modality Interaction Network for Pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*.

Wang, Y.; Lin, Y.; Meng, G.; Fu, Z.; Dong, Y.; Fan, L.; Yu, H.; Ding, X.; and Huang, Y. 2023. Learning high-frequency feature enhancement and alignment for pan-sharpening. In *Proceedings of the 31st ACM International Conference on Multimedia*, 358–367.

Wang, Z.-G.; Liang, Z.-H.; and Liu, C.-L. 2009. A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP. *Displays*, 30(3): 133–139.

Wei, C.; Wang, W.; Yang, W.; and Liu, J. 2018. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference (BMVC)*, 1–12.

Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. SNR-aware low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17714–17724.

Yang, S.; Ding, M.; Wu, Y.; Li, Z.; and Zhang, J. 2023. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12918–12927.

Yang, W.; Wang, S.; Fang, Y.; Wang, Y.; and Liu, J. 2020. 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)*, 3060–3069.

Yang, X.; Mi, M. B.; Yuan, Y.; Wang, X.; and Tan, R. T. 2022. Object detection in foggy scenes by embedding depth and reconstruction into domain adaptation. In *Proceedings of the Asian Conference on Computer Vision*, 1093–1108.

Yang, X.; Yan, W.; Yuan, Y.; Mi, M. B.; and Tan, R. T. 2024. Semantic Segmentation in Multiple Adverse Weather Conditions with Domain Knowledge Retention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 6558–6566.

Ye, T.; Chen, S.; Chai, W.; Xing, Z.; Qin, J.; Lin, G.; and Zhu, L. 2024. Learning Diffusion Texture Priors for Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2524–2534.

Ye, T.; Zhang, Y.; Jiang, M.; Chen, L.; Liu, Y.; Chen, S.; and Chen, E. 2022. Perceiving and Modeling Density for Image Dehazing. In *European Conference on Computer Vision*, 130–145. Springer.

Yi, X.; Xu, H.; Zhang, H.; Tang, L.; and Ma, J. 2023. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12302–12311.

Zeng, H.; Cai, J.; Li, L.; Cao, Z.; and Zhang, L. 2020. 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.

Zhang, D.; Zhou, F.; Jiang, Y.; and Fu, Z. 2023. 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*, 4188–4197.

Zheng, N.; Zhou, M.; Dong, Y.; Rui, X.; Huang, J.; Li, C.; and Zhao, F. 2023. Empowering Low-Light Image Enhancer through Customized Learnable Priors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12559–12569.