

HVI: A New Color Space for Low-light Image Enhancement

Qingsen Yan^{1*} Yixu Feng^{1*} Cheng Zhang¹
 Guansong Pang^{2†} Kangbiao Shi¹ Peng Wu¹ Wei Dong³ Jinqiu Sun¹ Yanning Zhang^{1†}
¹Northwestern Polytechnical University ²Singapore Management University
³Xi'an University of Architecture and Technology

Abstract

Low-Light Image Enhancement (LLIE) is a crucial computer vision task that aims to restore detailed visual information from corrupted low-light images. Many existing LLIE methods are based on standard RGB (sRGB) space, which often produce color bias and brightness artifacts due to inherent high color sensitivity in sRGB. While converting the images using Hue, Saturation and Value (HSV) color space helps resolve the brightness issue, it introduces significant red and black noise artifacts. To address this issue, we propose a new color space for LLIE, namely Horizontal/Vertical-Intensity (HVI), defined by polarized HS maps and learnable intensity. The former enforces small distances for red coordinates to remove the red artifacts, while the latter compresses the low-light regions to remove the black artifacts. To fully leverage the chromatic and intensity information, a novel Color and Intensity Decoupling Network (CIDNet) is further introduced to learn accurate photometric mapping function under different lighting conditions in the HVI space. Comprehensive results from benchmark and ablation experiments show that the proposed HVI color space with CIDNet outperforms the state-of-the-art methods on 10 datasets. The code is available at <https://github.com/Fediory/HVI-CIDNet>.

1. Introduction

Under low-light conditions, imaging sensors often capture weak light signals with severe noise, resulting in poor visual quality for low-light images. Obtaining high-quality images from such degraded images necessitates Low-Light Image Enhancement (LLIE) that aims at improving the image brightness while simultaneously reducing the impact of noise and color bias [24].

The majority of existing LLIE approaches [10, 15, 18, 35, 51, 55] focus on finding an appropriate image brightness, which is typically done by employing deep neural networks to learn a mapping relationship between low-light images

and normal-light images within the standard RGB (sRGB) space. However, the image brightness exhibits a strong coupling with the color from the three sRGB channels, *a.k.a.* high color sensitivity in [9, 23], causing an obvious color distortion of the restored image in these LLIE methods [3, 18], as shown in Fig. 1 (a).

Inspired by Kubelka-Munk theory [9], recent methods [25, 52, 56] have sought to transform images from the sRGB color space to the Hue, Saturation and Value (HSV) color space. These methods help achieve the brightness enhancement more accurately, but they amplify local **color space noise** [9], introducing severe artifacts in the results. As illustrated in Fig. 1 (b), the transformation from sRGB to the HSV disrupts the continuity of the red (**① Red Discontinuity Noise**) and black (**② Black Plane Noise**) color, resulting in increased Euclidean distances for similar color and the introduction of artifacts in the final images (see the zoomed in images of ① and ②). These two types of noise can cause severe artifacts in the enhancement of red-dominated or extremely dark images.

To address these issues, we introduce a new color space named **Horizontal/Vertical-Intensity (HVI)**, designed specifically for the LLIE task. The key intuition is that minimizing color space noise, by reducing Euclidean distances in similar colors. To this end, we polarize in the Hue and Saturation (HS) plane to enforce smaller distances for similar red point coordinates, which eliminates the red discontinuity noise in the primary HSV space (see Fig. 1 (c)). For the black plane noise issue, we introduce a trainable darkness density parameter k and its corresponding adaptive intensity collapse function C_k , which compresses the radius of low-light regions to zero, with the flexibility to gradually expand to the value of one as the intensity increases, as illustrated in Fig. 1 (d). This helps remove the black noise artifacts while maintaining the primary image appearance.

We further propose an LLIE network, named **Color and Intensity Decoupling Network (CIDNet)**, to leverage the chromatic and intensity information for optimizing the color and intensity in the HVI space. After transforming the image into the HVI space, CIDNet leverages two network

*These authors contributed equally to this work.

†Corresponding authors.

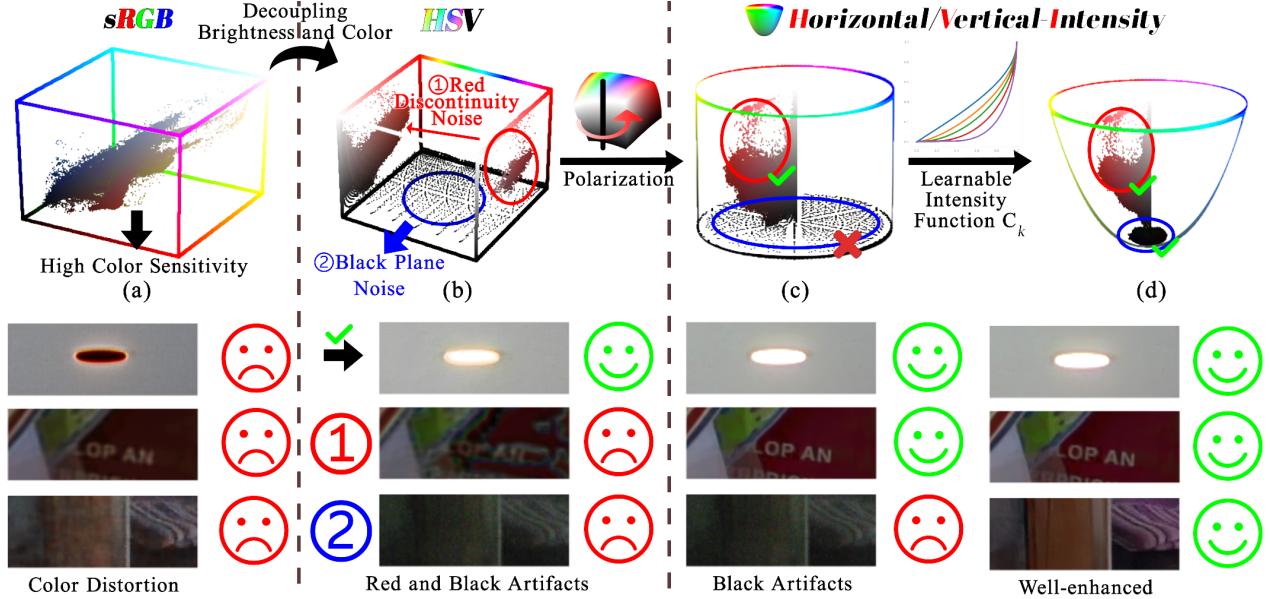


Figure 1. The top row illustrates the process of transforming images from the sRGB color space, via HSV, to the HVI color space. The bottom row presents the corresponding test results. The sRGB color space is known for its high color sensitivity, often causing color distortions in test images. By decoupling brightness and color to obtain the HSV color space, the illumination enhancement appears normalized. However, this transformation introduces varying levels of red discontinuities and black regions, which subsequently cause artifacts in the enhanced images. Introducing polarization to HSV ensures continuity in the red regions. Introducing a learnable intensity function C_k , on the other hand, helps collapse the black regions, yielding the HVI color space with optimal image enhancement.

branches, namely HV-branch and intensity-branch, to respectively model decoupled color and brightness information for learning accurate photometric mapping under different lighting conditions and restoring more natural colors.

Our contributions can be summarized as follows:

- We introduce a new HVI color space for the LLIE task, which is uniquely defined by polarized HS and trainable intensity. This offers an effective tool that eliminates the color space noise arising from the HSV space, largely enhancing the brightness of the low-light images.
- We further propose a novel LLIE-baseline network, CID-Net, to concurrently model the intensity and chromatic of low-light images in the HVI space. Despite being lightweight and computationally-efficient with relatively small parameters (1.88M) and computational loads (7.57GFLOPs), it enables the learning of accurate photometric mapping under different lighting conditions.
- Comprehensive results from quantitative and qualitative experiments show that our approach outperforms various types of state-of-the-art (SOTA) methods on different metrics across 10 datasets.

2. Related Work

2.1. Low-Light Image Enhancement

Single-stage Methods. Single-stage deep learning approaches [3, 5, 13, 18, 26, 51] has been widely used in LLIE. Existing methods propose distinct solutions to address the

aforementioned issues. For instance, RetinexNet [43] enhances images by decoupling illumination and reflectance based on Retinex theory. Bread [11] decouples the entanglement of noise and color distortion by using YCbCr color space. Furthermore, they designed a color adaption network to tackle the color distortion issue left in light-enhanced images. Still, RetinexNet and Bread can show inaccurate control in terms of brightness and biased color in black areas.

Diffusion-based Methods. With the advancement of Denoising Diffusion Probabilistic Models (DDPMs) [14], diffusion-based generative models have achieved remarkable results in the LLIE task. It has shown the capability to generate more accurate and appropriate images. However, they still exhibit issues such as local overexposure or color shift. Recent LLIE diffusion methods have attempted to address these challenges by incorporating global supervised brightness correction or employing local color correctors [6, 16, 17, 40, 45, 54]. Other methods such as Diff-Retinex [48] rethink the retinex theory with a diffusion-based model in the LLIE task, aiming to decomposed an image to illumination and reflectance from sRGB color space. However, these diffusion models often fail to fully decouple brightness and color information.

2.2. Color Space

RGB. Currently, the most commonly used is the sRGB color space. For the same principle as visual recognition

by the human eye, sRGB is widely used in digital imaging devices [32]. Nevertheless, image brightness and color exhibit a strong interdependence with the three channels in sRGB [9]. A slight disturbance in the color space will cause an obvious variation in both the brightness and color of the generated image. Thus, sRGB is not a desired color space for low-light enhancement.

HSV and YCbCr. The HSV color space represents points in an RGB color model with a cylindrical coordinate system [7]. Indeed, it does decouple brightness and color of the image from RGB channels. However, the red color discontinuity and black plane noise pose significant challenges when we enhance the images in HSV color space, resulting in the emergence of various obvious artifacts. To circumvent issues related to HSV, some methods [1, 11] also transform sRGB images to the YCbCr color space, which has an illumination axis (Y) and reflect-color-plane (CbCr). Although it solved the hue dimension discontinuity problem of HSV, the Y axis is still coupled with the CbCr plane partially, leading to severe color shifts.

3. HVI Color Space

The HVI space is built upon the HSV color space, which is proposed to address the color space noise issues arising from the HSV space. The key intuition in HVI is that the restored images should have good perceptual quality for respective colors, *i.e.*, similar colors have small Euclidean distances. Below we introduce the HVI transformation in detail, where the HSV color space is first applied to decouple the brightness and color information of input images, which could cause color space noise (*e.g.*, red discontinuity and black plane). We then introduce our proposed polarized HS operations and learnable intensity collapse function in HVI to effectively address these issues.

3.1. Color Space Noises in HSV

Intensity Map. In the task of LLIE, one crucial aspect is accurately estimating the illumination intensity map of the scene from a sRGB input image. Previous methods [8, 43, 51] largely rely on the Retinex theory [21], using deep learning to directly generate the corresponding normal-light map. While this approach aligns with statistical principles, it often struggles to fit physical laws and human perception [9], resulting in limited generalizability [40]. Therefore, we instead refer to the Max-RGB theory [21] to estimate the intensity map, *a.k.a.*, Value in HSV, rather than using neural networks to generate it. According to Max-RGB, for each individual pixel x , we can estimate the intensity map of an image $\mathbf{I}_{max} \in \mathbb{R}^{H \times W}$ as follows:

$$\mathbf{I}_{max}(x) = \max_{c \in \{R, G, B\}} (\mathbf{I}_c(x)). \quad (1)$$

The intensity map then goes through the sRGB-HSV

transformation that can lead to various red and black noises. We introduce these separately as follows.

Hue/Saturation Plane. Real-world low-light images often contain significant noise, making its identification and removal a key challenge in LLIE. Recent studies [43, 48] indicate that the noise in low-light images is a primary cause to shifts in Hue and Saturation, *a.k.a.* a general case of Reintex theory [40], while having minimal impact on light intensity. Therefore, decoupling the sRGB color space, known for its high color sensitivity, can be advantageous for the LLIE task. By leveraging pixel-based photometric invariance [9] and dichromatic reflection modeling [34], sRGB can be decoupled into illuminance and chromatic components, yielding the HSV (Hue/Saturation-Value) color space. In this representation, the Value map (\mathbf{V}) component corresponds to light intensity map ($\mathbf{V} = \mathbf{I}_{max}$), while the HS plane forms a chromaticity plane independent of illuminance constraints. Specifically, the transformation of sRGB image to Saturation map (\mathbf{S}) is defined as follows

$$\mathbf{s} = \begin{cases} 0, & \mathbf{I}_{max} = \mathbf{0} \\ \frac{\Delta}{\mathbf{I}_{max}}, & \mathbf{I}_{max} \neq \mathbf{0} \end{cases}, \quad (2)$$

where $\Delta = \mathbf{I}_{max} - \min(\mathbf{I}_c)$ and \mathbf{s} is any pixel in \mathbf{S} . The Hue map (\mathbf{H}) is formulated as

$$\mathbf{h} = \begin{cases} 0, & \text{if } \mathbf{s} = 0 \\ \frac{\mathbf{I}_G - \mathbf{I}_B}{\Delta} \bmod 6, & \text{if } \mathbf{I}_{max} = \mathbf{I}_R \\ 2 + \frac{\mathbf{I}_B - \mathbf{I}_R}{\Delta}, & \text{if } \mathbf{I}_{max} = \mathbf{I}_G \\ 4 + \frac{\mathbf{I}_R - \mathbf{I}_G}{\Delta}, & \text{if } \mathbf{I}_{max} = \mathbf{I}_B \end{cases}, \quad (3)$$

where \mathbf{h} is any pixel in \mathbf{H} .

Color Space Noises. Converting an sRGB image to HSV space effectively decouples brightness from color, enabling more accurate color denoising and more natural illuminance recovery. However, this transformation also amplifies noise in the red and dark regions [9], which are critical for LLIE tasks. As illustrated in Fig. 1 (b), enhancing the image brightness within the HSV color space yields a more balanced brightness level. However, excessive noise in the red discontinuities and the black plane introduces significant artifacts, particularly in red and dark areas of output image, which greatly degrade the perceptual quality.

To address this issue, we propose the HVI color space as follows, which effectively preserves the decoupling of brightness and color while minimizing these artifacts.

3.2. Horizontal/Vertical Plane with Polarized HS and Collapsible Intensity

Our primary approach to addressing the color space noise issue is to ensure that more similar colors exhibit smaller Euclidean distances. Along the Hue axis, the color red appears identically at both $\mathbf{h} = 0$ and $\mathbf{h} = 6$, due to the

modular arithmetic of Hue-axis, which splits the same color across two ends of the spectrum. In particular, to address the red discontinuity issue, we apply polarization to the Hue axis (\mathbf{h}) at each pixel in \mathbf{H} , obtaining orthogonal $h = \cos(\frac{\pi\mathbf{h}}{3})$ and $v = \sin(\frac{\pi\mathbf{h}}{3})$. When the Hue axis is polarized, it forms an angle within the orthogonalized $h - v$ plane, with s representing the distance from the origin.

For the black plane noise issue, we aim to collapse regions of low light intensity while preserving those with higher intensity. However, the optimal extent of collapse varies across different datasets and networks. Therefore, it is important to make this region adaptively collapsible through a learning process. To achieve this, we introduce an adaptive intensity collapse function \mathbf{C}_k as follows

$$\mathbf{C}_k(x) = \sqrt[k]{\sin\left(\frac{\pi\mathbf{I}_{max}(x)}{2}\right)} + \varepsilon, \quad (4)$$

where $k \in \mathbb{Q}^+$ is a trainable parameter to control the dark color point density, and a small $\varepsilon = 1 \times 10^{-8}$ is used to avoid gradient explosion. Essentially, \mathbf{C}_k serves a radius mapping function, with smaller \mathbf{C}_k corresponding to smaller radius or lower intensity values. Thus, black points are clustered together as \mathbf{C}_k decreases. We then formalize the Horizontal ($\hat{\mathbf{H}}$) map and Vertical ($\hat{\mathbf{V}}$) map as

$$\begin{aligned} \hat{\mathbf{H}} &= \mathbf{C}_k \odot \mathbf{S} \odot H, \\ \hat{\mathbf{V}} &= \mathbf{C}_k \odot \mathbf{S} \odot V, \end{aligned} \quad (5)$$

where $h \in H$, $v \in V$, and \odot denotes the element-wise multiplication. $\hat{\mathbf{H}}$, $\hat{\mathbf{V}}$, and \mathbf{I}_{max} can be concatenated to form an HVI image.

Thanks to these operations, the HVI space builds a strong color space that maintains the advantages of HSV while removing the red and black noises. More importantly, since HVI is trainable due to k and \mathbf{C}_k , specifically designed neural networks, as discussed in the following, can be created to optimize LLIE upon HVI under various lighting conditions.

4. Color and Intensity Decoupling Network

To more effectively utilize chromatic and brightness information in the HVI space, we introduce a novel dual-branch LLIE network, named Color and Intensity Decoupling Network (CIDNet), to separately model the HV-plane and I-axis information in the HVI space, as shown in Fig. 2. CIDNet employs HV-branch to suppress the noise and chromaticity in the dark regions and utilizes I-branch to estimate the illuminance of the whole images.

The overall framework of CIDNet can be divided into three consecutive main steps. There is an HVI transformation applied before the dual-branch enhancement network. After the enhancement, CIDNet performs perceptual-inverse HVI transformation to map the image to the sRGB space. Below we introduce each step in detail.

4.1. HVI Transformation

As described in Sec. 3.2, the HVI transformation decomposes the sRGB image into two components: an intensity map containing scene illuminance information and an HV color map containing scene color and structure information. Specifically, we first calculate the intensity map using Eq. 1, which is $\mathbf{I}_I = \mathbf{I}_{max}$. Subsequently, we utilize the intensity map and the original image to generate HV color map using Eq. 5. Furthermore, a trainable density- k is employed to adjust the color point density of the low-intensity color plane, as shown in Fig. 2(a).

4.2. Dual-branch Enhancement Network

As illustrated in Fig. 2 (b), the dual-branch network is built upon the UNet architecture, involving an encoder and an decoder with respective three Lighten Cross-Attention (LCA) modules, and multiple skip connections. Two key designs in the network are the dual-branch structure and the LCA module. We explain the key intuition behind these designs as follows.

The LLIE task can be decomposed into two sub-tasks: noise removal in low-light regions and brightness enhancement. By converting the image to the HVI color space, where luminance and color are decoupled, we can apply brightness mapping to the Intensity map and denoising to the HV color map. Since these two sub-tasks follow distinct statistical patterns [24], inspired by Retinex-based methods [8, 43, 44], we use separate branches, the I-branch and HV-branch, to address each sub-task individually. Additionally, the input to the HV branch is formed by concatenating the Intensity map with the HV color map, as we observed that severely low-light images contain a small amount of noise in the luminance component as well.

The cross-attention between the two branches is used, rather than using self-attention individually to each branch. One main reason is that the illumination intensity is inversely proportional to image noise intensity. A low-light image may also contain high-illumination regions requiring only minimal denoising and enhancement. Therefore, using the intensity features to guide the HV-branch in denoising can reduce global color shifts and achieve more effective noise suppression. For another reason, the noisy intensity information, after being denoised in the HV-branch, is transferred to the I-branch through cross-attention, resulting in smoother enhancement outcomes.

4.3. Perceptual-inverse HVI Transformation

To convert HVI back to the HSV color space, we perform a Perceptual-inverse HVI Transformation (PHVIT), which is a surjective mapping while allowing for the independent adjustment of the image's saturation and brightness.

The PHVIT sets \hat{h} and \hat{v} as an intermediate variable as $\hat{h} = \frac{\hat{\mathbf{H}}}{\mathbf{C}_k + \varepsilon}$, $\hat{v} = \frac{\hat{\mathbf{V}}}{\mathbf{C}_k + \varepsilon}$, where $\varepsilon = 1 \times 10^{-8}$ is used to avoid

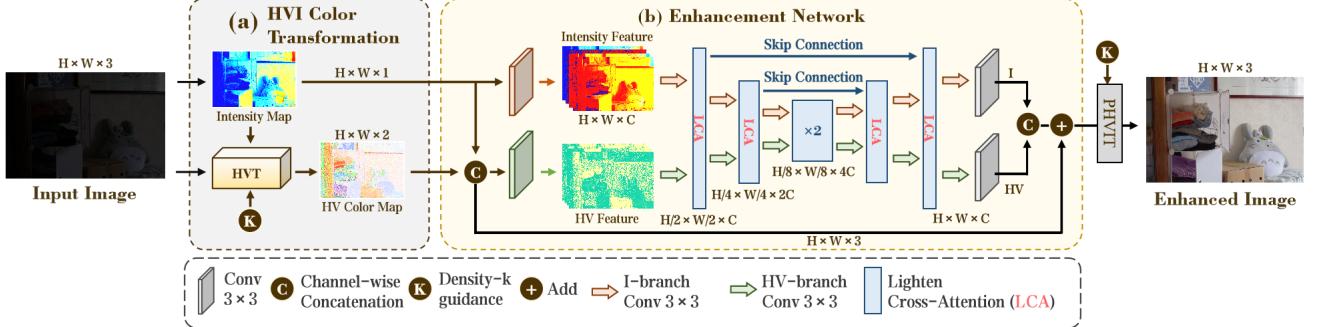


Figure 2. The overview of the proposed CIDNet. (a) HVI Color Transformation (HVIT) takes an sRGB image as input and generates HV color map and intensity map as outputs. (b) Enhancement Network performs the main processing, utilizing a dual-branch UNet architecture with six Lighten Cross-Attention (LCA) blocks. Lastly, we apply Perceptual-inverse HVI Transform (PHVIT) to take a light-up HVI map as input and transform it into an sRGB-enhanced image.

gradient explosion. Then, we convert \hat{h} and \hat{v} to HSV color space. The Hue (**H**), Saturation (**S**) and Value (**V**) map can be estimated as

$$\begin{aligned} \mathbf{H} &= \arctan\left(\frac{\hat{v}}{\hat{h}}\right) \bmod 1, \\ \mathbf{S} &= \alpha_S \sqrt{\hat{h}^2 + \hat{v}^2}, \\ \mathbf{V} &= \alpha_I \hat{\mathbf{I}}_I, \end{aligned} \quad (6)$$

where α_S, α_I are the customizing linear parameters to change the image color saturation and brightness. Finally, we will obtain the sRGB image with HSV image [7].

4.4. Loss Function

To provide comprehensive supervision for training CIDNet, we guide the enhancement from two key perspectives, including the GroundTruth in the sRGB space and the HVI map in the HVI space. Specifically, given the enhanced HVI map $\hat{\mathbf{I}}_{\text{HVI}}$ and the restored sRGB image $\hat{\mathbf{I}}$ output from CIDNet, we aim to minimize their difference to the primary sRGB GroundTruth \mathbf{I} and its corresponding HVI map \mathbf{I}_{HVI} :

$$L = \lambda \cdot l(\hat{\mathbf{I}}_{\text{HVI}}, \mathbf{I}_{\text{HVI}}) + l(\hat{\mathbf{I}}, \mathbf{I}), \quad (7)$$

where λ is a weighting hyperparameter to balance the losses in the two different color spaces. This helps achieve not only more closely with the probabilistic distribution of sRGB in the HVI space, especially the red and black ones, due to the optimized k and \mathbf{C}_k , but also inheritance of pixel-level structure detail in the sRGB space.

5. Experiments

5.1. Datasets and Settings

We employ seven commonly-used LLIE benchmark datasets for evaluation, including LOLv1 [43], LOLv2 [47], DICM [22], LIME [12], MEF [28], NPE [38], and VV [36].

We also conduct further experiments on two extreme datasets, SICE [2] (containing mix and grad test sets [53]) and SID (Sony-Total-Dark) [4].

LOL. The LOL dataset has v1 [43] and v2 [47] versions. Compared to LOL-v1 that contains both real and synthetic data, LOL-v2 is divided into real and synthetic subsets. For LOLv1 and LOLv2-Real, we crop the training images into 400×400 patches and train CIDNet for 1,500 epochs with a batch size of 8. For LOLv2-Synthetic, we set the batch size to 1 and train CIDNet for 500 epochs without cropping.

SICE. The original SICE dataset [2] contains 589 low-light and overexposed images, with the training, validation, and test sets divided into three groups according to 7:1:2. CIDNet is optimized using 160×160 cropped images from the training set for 1,000 epochs with a batch size of 10 and tested on the datasets SICE-Mix and SICE-Grad [53].

Sony-Total-Dark. This dataset is a customized version of a SID subset [4]. To make this dataset more challenging, we convert the raw format images to sRGB images with *no gamma correction*, resulting in images with extreme darkness. We crop the training images into 256×256 patches and train CIDNet for 1,000 epochs with a batch size of 4.

Experiment Settings. We implement our CIDNet by PyTorch. The model is trained with the Adam [19] optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) by using a single NIVIDA 2080Ti or 3090 GPU. The learning rate is initially set to 1×10^{-4} and then steadily decreased to 1×10^{-7} by the cosine annealing scheme [27] during the training process.

Evaluation Metrics. For the *paired* datasets, we adopt the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) [42] as the distortion metrics. To comprehensively evaluate the perceptual quality of the restored images, we report Learned Perceptual Image Patch Similarity (LPIPS) [50] with AlexNet [20] as the reference. For the *unpaired* datasets, we evaluate single recovered images using BRISQUE [29] and NIQE [30] perceptually.

Table 1. Quantitative results of PSNR/SSIM↑ and LPIPS↓ on the LOL (v1 and v2) datasets. Due to the limited number of test set in LOLv1, we use GT mean method during testing to minimize errors. This approach will be explained in supplementary. The FLOPs is tested on a single 256×256 image. The best performance is in red color and the second best is in cyan color.

Methods	Color Model	Complexity		LOLv1			LOLv2-Real			LOLv2-Synthetic		
		Params/M	FLOPs/G	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexNet [43]	Retinex	0.84	584.47	18.915	0.427	0.470	16.097	0.401	0.543	17.137	0.762	0.255
KinD [51]	Retinex	8.02	34.99	23.018	0.843	0.156	17.544	0.669	0.375	18.320	0.796	0.252
ZeroDCE [10]	RGB	0.075	4.83	21.880	0.640	0.335	16.059	0.580	0.313	17.712	0.815	0.169
RUAS [33]	Retinex	0.003	0.83	18.654	0.518	0.270	15.326	0.488	0.310	13.765	0.638	0.305
LLFlow [41]	RGB	17.42	358.4	24.998	0.871	0.117	17.433	0.831	0.176	24.807	0.919	0.067
EnlightenGAN [18]	RGB	114.35	61.01	20.003	0.691	0.317	18.230	0.617	0.309	16.570	0.734	0.220
SNR-Aware [46]	SNR+RGB	4.01	26.35	26.716	0.851	0.152	21.480	0.849	0.163	24.140	0.928	0.056
Bread [11]	YCbCr	2.02	19.85	25.299	0.847	0.155	20.830	0.847	0.174	17.630	0.919	0.091
PairLIE [8]	Retinex	0.33	20.81	23.526	0.755	0.248	19.885	0.778	0.317	19.074	0.794	0.230
LLFormer [39]	RGB	24.55	22.52	25.758	0.823	0.167	20.056	0.792	0.211	24.038	0.909	0.066
RetinexFormer [3]	Retinex	1.53	15.85	27.140	0.850	0.129	22.794	0.840	0.171	25.670	0.930	0.059
GSAD [16]	RGB	17.36	442.02	27.605	0.876	0.092	20.153	0.846	0.113	24.472	0.929	0.051
QuadPrior [40]	Kubelka-Munk	1252.75	1103.20	22.849	0.800	0.201	20.592	0.811	0.202	16.108	0.758	0.114
CIDNet(Ours)	HVI	1.88	7.57	28.201	0.889	0.079	24.111	0.871	0.108	25.705	0.942	0.045

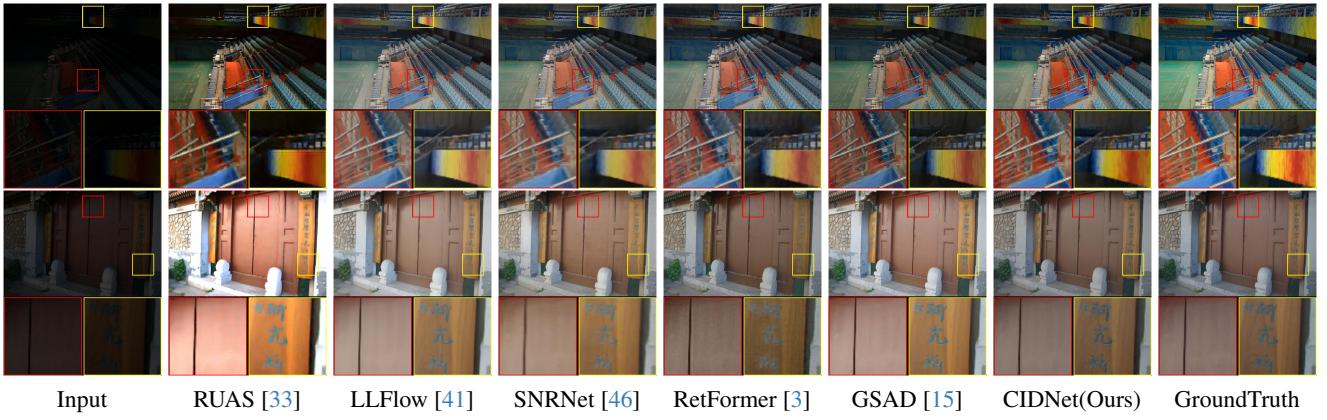


Figure 3. Visual comparison of the enhanced images yielded by different SOTA methods on LOLv1 (top row) and LOLv2 (bottom row).

Table 2. Quantitative result on SICE, Sont-Total-Dark, and the five unpaired datasets (DICM [22], LIME [12], MEF [28], NPE [38], and VV [36]). The top-ranking score is in red.

Methods	SICE		Sony-Total-Dark		Unpaired	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	BRIS↓	NIQE↓
RetinexNet [43]	12.424	0.613	15.695	0.395	23.286	4.558
ZeroDCE [10]	12.452	0.639	14.087	0.090	26.343	4.763
URetinexNet [44]	10.899	0.605	15.519	0.323	26.359	3.829
RUAS [33]	8.656	0.494	12.622	0.081	36.372	4.800
LLFlow [41]	12.737	0.617	16.226	0.367	28.087	4.221
CIDNet (Ours)	13.435	0.642	22.904	0.676	23.521	3.523

5.2. Main Results

Results on LOL Datasets. In Tab. 1, it can be found that our method is optimal on all metrics for both LOLv1 and LOLv2 datasets with 1.88M parameters and 7.57 GFLOPs. We outperform the best RGB-based method GSAD (difusion) in terms of all PSNR, SSIM, and LPIPS metrics, while utilizing only 10.8% parameters of GSAD. Compared to RetinexFormer, a SOTA method based on Retinex theory, CIDNet delivers higher image quality while reducing computational cost by 8.28 GFLOPs. Subjective results are illustrated in Fig. 3, which demonstrates that our method not

only more accurately recovers multi-color regions compared to GroundTruth but also achieves stable brightness enhancement, thanks to the HVI color space. More visualization comparison can be found in supplementary materials.

Results on SICE and Sony-total-Dark. To validate the performance on large-scale datasets, we evaluate CIDNet on SICE (including Mix and Grad) and SID-Total-Dark. The results are presented in Table 2, where it is clear that CIDNet is the best performer in both PSNR and SSIM metrics on the two datasets. Notably, on Sony-Total-Dark, our model surpasses the second-best method by 6.678 dB in PSNR. This improvement is due to the extreme darkness of the dataset images, which substantially increases the difficulty of distinguishing details from noise. However, CIDNet leverages the intensity collapse function C_k to effectively maintain an optimal signal-to-noise ratio during training, enabling better detail recovery.

Results on Unpaired Datasets. We evaluate the effectiveness of models trained on LOLv1 or LOLv2-Syn using various methods, and report their performance using BRISQUE and NIQE metrics in Tab. 2. Our method exhibits a substantial improvement in the NIQE metric compared

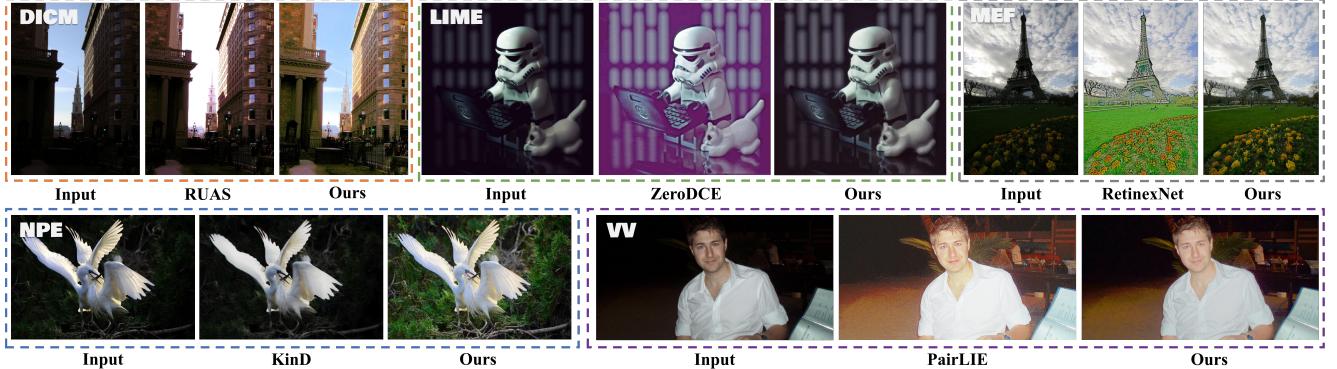


Figure 4. Visual comparison on the five unpaired datasets. Follow RetinexFormer [3], we select one image in each dataset to compare our method with the other methods. More visual comparison can be found in the supplementary materials.

Table 3. Results of applying HVI transformation as a plug-in to various LLIE methods on LOLv2-Real. Values in brackets represent the absolute improved performance gain. The best PSNR/SSIM \uparrow , LPIPS \downarrow , and inference time \downarrow are in **bolded**.

Methods	FourLLIE [37]	LEDNet [57]	SNR-Aware [46]	LLFormer [39]	GSAD [15]	DiffLight [6]	CIDNet
PSNR \uparrow	22.730(+0.381)	23.394(+3.456)	22.251(+0.771)	22.671(+2.615)	23.715(+3.562)	23.969(+1.364)	24.111
SSIM \uparrow	0.856(+0.009)	0.837(+0.010)	0.840(-0.009)	0.852(+0.060)	0.876(+0.030)	0.859(+0.003)	0.871
LPIPS \downarrow	0.125(+0.011)	0.115(-0.005)	0.117(-0.054)	0.117(-0.094)	0.103(-0.010)	0.109(-0.012)	0.108
GPU Time/s \downarrow	0.075	0.054	0.070	0.139	0.315	0.578	0.053
Model Type	CNN	CNN	Transformer	Transformer	Diffusion	CNN+Diffusion	Transformer

Table 4. Model ablation. Dual denotes the dual-branch network.

Metrics	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Color Space	sRGB	20.062	0.825
	HSV	21.349	0.801
	HVI (w/ Polarization Only)	21.558	0.821
	HVI (w/ C _k Only)	21.536	0.825
Structure	UNet Baseline [31]	19.306	0.778
	SelfAttn [49]	22.313	0.835
	Dual+SelfAttn [49]	23.159	0.856
Loss	HVI Only	23.221	0.854
	sRGB Only	23.319	0.857
Full Model (HVI-CIDNet)		24.111	0.871
		0.108	

to other approaches. As shown in Fig. 4, while CIDNet does not outperform RetinexNet in the BRISQUE metric in Tab. 2, its recovered perceptual results are closer to realistic appearances than RetinexNet. This may be attributed to the fact that building upon the HSV space, the HVI color space is derived from real-world perceptual models [9].

Generalizing HVI to Other LLIE Models. To verify the effectiveness of the HVI color space, we further evaluate its performance when it is used with different LLIE models. In particular, HVIT, together with its inverse mapping PHVIT, is used as a plug-and-play module into six SOTA methods that use sRGB images as input and are independent of specific color space characteristics. The results are reported in Tab. 3. It is clear that transforming to the HVI color space improves PSNR, SSIM, and LPIPS metrics across various methods compared to the results in the sRGB color space. Notably, the GSAD method demonstrates the most signifi-

cant improvement, with a PSNR increase of 3.562 dB. This demonstrates not only the generalizability of HVI to various sRGB-based methods but also its general effectiveness as a color space for the LLIE task.

For the inference time results in Tab. 3, it is evident that the diffusion-based methods require longer GPU time but achieve better enhancement results. In contrast, CIDNet shows the most efficient inference while achieving the highest PSNR and the second-best SSIM and LPIPS scores. This highlights the strong ability of CIDNet in balancing the efficiency and effectiveness within the HVI color space.

5.3. Ablation Study

We validate our HVI color space and the key modules in CIDNet using both quantitative (Tab. 4) and qualitative results (Figs. 5 and 6). The experiments are all performed on LOLv2-Real for fast convergence and stable performance.

HVI Color Space. It can be seen in Tab. 4 that the enhancement in the sRGB color space leads to chromatic aberration and luminance bias, as demonstrated by the difference between Figs. 5(b) and (g). Compared to sRGB, using HSV yields images aligned more closely with the GroundTruth in both luminance and color due to its effectiveness on decoupling brightness from color, which can be observed the enhancement from Fig. 5(b) to Fig. 5(c). This results in improved PSNR in Tab. 4. However, it significantly introduces more noise due to the red discontinuity in HSV, leading to the noisy black spots in the red regions in Fig. 5(c) and the degraded SSIM and LPIPS in Tab. 4.

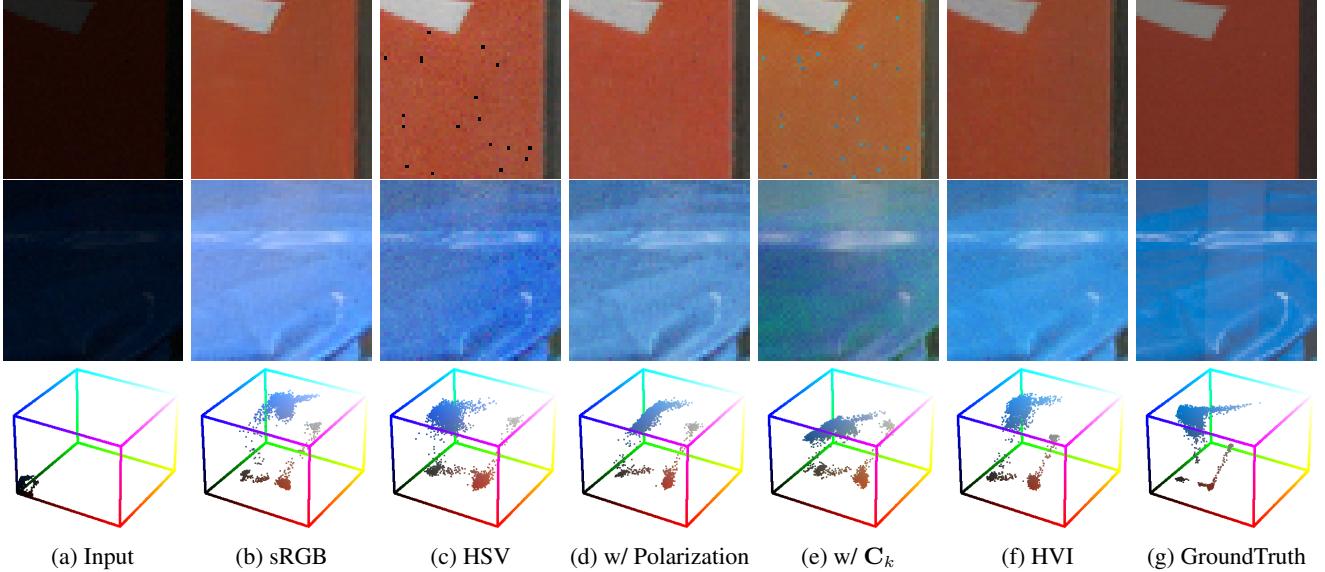


Figure 5. Top and middle rows are ablation results on LOLv2-Real for five different color spaces used by CIDNet. The bottom row provides a visual comparison by mapping the pixel values of the results to sRGB. Note that due to the dual-branch and the cross-attention mechanism are specifically designed for HVI, we only use UNet [31] with self attentions [49] for a fair comparison.

Using the polarization or the C_k solely in the HVI space can lead to similar image quality in PSNR and SSIM compared to those in the HSV color space. Qualitatively, as shown in Fig. 5(d), using polarization only helps cluster similar red tones, avoiding the red discontinuity. Relying solely on the intensity function C_k helps adjust the brightness, but leads to confusion between red and other colors. Consequently, dot-like artifacts appear not only in the red regions but also color shifts in other areas, as shown in Fig. 5(e). These issues are all effectively mitigated when the polarization and the C_k function are applied together, as shown by consistent improvement of the full model in all three metrics in Tab. 4 and the better image enhancement in Fig. 5(f).

Dual-branch Network Structure. In Tab. 4, adding self-attention to the baseline noticeably improves all three metrics, indicating that transformer-based models hold potential for application in the HVI color space. We then modified the architecture from a single-branch to a dual-branch structure without cross-attention, resulting in a PSNR increase of 0.846 dB, while SSIM and LPIPS showed minimal change. Further incorporating the cross-attention into the I-branch and HV-branch (Full Model) obtains the best color restoration, light enhancement, and optimal metric performance, as shown in Tab. 4. This can also be observed in Fig. 6.

Loss Function. As shown in Tab. 4, compared to using both HVI and sRGB losses, relying solely on the HVI loss lacks pixel-level spatial consistency constraints, leading to a loss of structural detail in the image and thus lower performance across the three metrics, especially in the LPIPS metric. On the other hand, using only sRGB loss is focused on pixel-space enhancement, neglecting the low-light



Figure 6. Results of using different structures on LOLv2-Real.

probability distribution in the HVI color space, resulting in undesired color imbalance.

6. Conclusion

In this work, we introduce the HVI color space and the CIDNet approach to address the color bias and brightness artifact issues that occur to current sRGB-based LLIE approaches. By encapsulating polarized HS maps and a learnable intensity component, HVI shows strong robustness to both issues. To further enhance LLIE, CIDNet is designed to model decoupled chromatic and intensity information in the HVI space for achieving precise photometric adjustments under varying lighting conditions. Experimental results on 10 datasets demonstrate that the HVI color space, combined with CIDNet, outperforms SOTA LLIE methods, establishing it as a robust solution for low-light enhancement.

Acknowledgement

NSFC of China under Grant 62301432 and Grant 6230624, the Natural Science Basic Research Program of Shaanxi under Grant 2023-JC-QN-0685 and Grant QCYRCXM-2023-057, the Fundamental Research Funds

for Central Universities, and Guangdong Basic and Applied Basic Research Foundation 2025A1515011119.

References

- [1] A. Brateanu, R. Balmez, A. Avram, and C. Orhei. Lyt-net: Lightweight yuv transformer-based network for low-light image enhancement, 2024. 3
- [2] 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. 5
- [3] 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 ICCV*, pages 12504–12513, 2023. 1, 2, 6, 7
- [4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, 2018. 5
- [5] Guo-Dong Fan, Bi Fan, Min Gan, Guang-Yong Chen, and C. L. Philip Chen. Multiscale low-light image enhancement network with illumination constraint. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11):7403–7417, 2022. 2
- [6] Yixu Feng, Shuo Hou, Haotian Lin, Yu Zhu, Peng Wu, Wei Dong, Jinqui Sun, Qingsen Yan, and Yanning Zhang. Diff-flight: Integrating content and detail for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6143–6152, 2024. 2, 7
- [7] James D Foley and Andries Van Dam. *Fundamentals of interactive computer graphics*. Addison-Wesley Longman Publishing Co., Inc., 1982. 3, 5
- [8] 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. 3, 4, 6
- [9] Theo Gevers, Arjan Gijsenij, Joost Van de Weijer, and Jan-Mark Geusebroek. *Color in computer vision: fundamentals and applications*. John Wiley & Sons, 2012. 1, 3, 7
- [10] Chunle Guo 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 *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1780–1789, 2020. 1, 6
- [11] Xiaojie Guo and Qiming Hu. Low-light image enhancement via breaking down the darkness. *International Journal of Computer Vision*, 131(1):48–66, 2023. 2, 3, 6
- [12] X. Guo, Y. Li, and H. Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 5, 6
- [13] Zhiqian He, Wu Ran, Shulin Liu, Kehua Li, Jiawen Lu, Changyong Xie, Yong Liu, and Hong Lu. Low-light image enhancement with multi-scale attention and frequency-domain optimization. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4):2861–2875, 2024. 2
- [14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, pages 6840–6851. Curran Associates, Inc., 2020. 2
- [15] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 6, 7
- [16] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, 36, 2024. 2, 6
- [17] 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. 2
- [18] 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. 1, 2, 6
- [19] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. 5
- [20] Alex Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25(2), 2012. 5
- [21] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 3
- [22] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Transactions on Image Processing*, 22(12): 5372–5384, 2013. 5, 6
- [23] Jaerin Lee, JoonKyu Park, Sungyong Baik, and Kyoung Mu Lee. Rethinking RGB color representation for image restoration models, 2024. 1
- [24] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9396–9416, 2022. 1, 4
- [25] Zhi Li, Zhenhong Jia, Jie Yang, and Nikola Kasabov. Low illumination video image enhancement. *IEEE photonics journal*, 12(4):1–13, 2020. 1
- [26] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement, 2016. 2
- [27] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts, 2017. 5
- [28] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, 2015. 5, 6
- [29] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12): 4695–4708, 2012. 5

- [30] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. 5
- [31] Olivier Petit, Nicolas Thome, Clément Rambour, and Luc Soler. U-net transformer: Self and cross attention for medical image segmentation, 2021. 7, 8
- [32] Charles Poynton. 18 - image digitization and reconstruction. In *Digital Video and HDTV*, pages 187–194. Morgan Kaufmann, San Francisco, 2003. 3
- [33] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 6
- [34] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985. 3
- [35] Chunyan She, Fujun Han, Lidan Wang, Shukai Duan, and Tingwen Huang. Mpc-net: Multi-prior collaborative network for low-light image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2024. 1
- [36] Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77:1–21, 2018. 5, 6
- [37] Chenxi Wang, Hongjun Wu, and Zhi Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7459–7469, 2023. 7
- [38] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 22, 2013. 5, 6
- [39] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2654–2662, 2023. 6, 7
- [40] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. Zero-reference low-light enhancement via physical quadruple priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26057–26066, 2024. 2, 3, 6
- [41] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C Kot. Low-light image enhancement with normalizing flow. *arXiv preprint arXiv:2109.05923*, 2021. 6
- [42] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. 5
- [43] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2, 3, 4, 5, 6
- [44] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5901–5910, 2022. 4, 6
- [45] Yuhui Wu, Guoqing Wang, Zhiwen Wang, Yang Yang, Tianyu Li, Peng Wang, Chongyi Li, and Heng Tao Shen. Recodiff: Explore retinex-based condition strategy in diffusion model for low-light image enhancement. *arXiv preprint arXiv:2312.12826*, 2023. 2
- [46] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17693–17703, 2022. 6, 7
- [47] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 30:2072–2086, 2021. 5
- [48] 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. 2, 3
- [49] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022. 7, 8
- [50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 5
- [51] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, New York, NY, USA, 2019. ACM. 1, 2, 3, 6
- [52] Yu Zhang, Xiaoguang Di, Bin Zhang, Ruihang Ji, and Chun-hui Wang. Better than reference in low-light image enhancement: Conditional re-enhancement network. *IEEE Transactions on Image Processing*, 31:759–772, 2021. 1
- [53] Shen Zheng, Yiling Ma, Jinjian Pan, Changjie Lu, and Gaurav Gupta. Low-light image and video enhancement: A comprehensive survey and beyond. *arXiv preprint arXiv:2212.10772*, 2022. 5
- [54] Deweい Zhou, Zongxin Yang, and Yi Yang. Pyramid diffusion models for low-light image enhancement. *arXiv preprint arXiv:2305.10028*, 2023. 2
- [55] Han Zhou, Wei Dong, Xiaohong Liu, Shuaicheng Liu, Xiongkuo Min, Guangtao Zhai, and Jun Chen. Glare: Low light image enhancement via generative latent feature based codebook retrieval. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. 1
- [56] Libing Zhou, Xiaojing Chen, Baisong Ye, Xueli Jiang, Sheng Zou, Liang Ji, Zhengqian Yu, Jianjian Wei, Yixin Zhao, and Tianyu Wang. A low-light image enhancement method based on hsv space. *The Imaging Science Journal*, pages 1–14, 2023. 1
- [57] Shangchen Zhou, Chongyi Li, and Chen Change Loy. Lednet: Joint low-light enhancement and deblurring in the dark. In *ECCV*, 2022. 7