

DualPrior-Retinex: Robust Low-Light Image Enhancement via Dual Prior Guided Reflectance and Illumination Decomposition

Author Name

Affiliation

email@example.com

Abstract

Accurate estimation of reflectance and illumination maps is fundamental to Retinex-based low-light image enhancement, yet remains challenging due to inherent information loss. Existing methods struggle to simultaneously achieve high perceptual quality and pixel-level fidelity. We introduce DualPrior-Retinex, a novel framework that leverages a diffusion-based generator for global prior generation and YUV space illumination maps for local prior extraction. These systematically complementary priors provide effective constraints for optimization, guiding the refinement of reflectance and illumination components while reducing decomposition ambiguity. By integrating these refined maps through a Hierarchical Prior Synthesis Module (HPSM), our method produces enhanced images with superior detail preservation and balanced brightness while maintaining both high perceptual quality and pixel-level fidelity. Extensive evaluation on multiple benchmarks demonstrates that our approach establishes new state-of-the-art performance in low-light image enhancement.

1 Introduction

Low-light image enhancement (LLIE) [Li *et al.*, 2022] is a critical yet challenging task in computer vision, aimed at improving the visibility and contrast of images captured under poor lighting conditions. This process not only enhances human visual perception but also plays a pivotal role in downstream tasks such as object detection [Loh and Chan, 2019a], semantic segmentation [Wang *et al.*, 2022a], and tracking in low-light scenarios [Wang *et al.*, 2024]. When capturing images in low-light environments, multiple forms of degradation emerge, including detail loss, contrast reduction, sensor noise amplification, and color distortion. The complex interplay of these factors presents significant technical challenges for achieving high-quality image restoration.

The seminal Retinex theory [Land and McCann, 1971] provides a fundamental theoretical framework for LLIE by decomposing images into reflectance (inherent object properties) and illumination (lighting conditions) components. Traditional Retinex-based methods [Jobson *et al.*, 1997], built

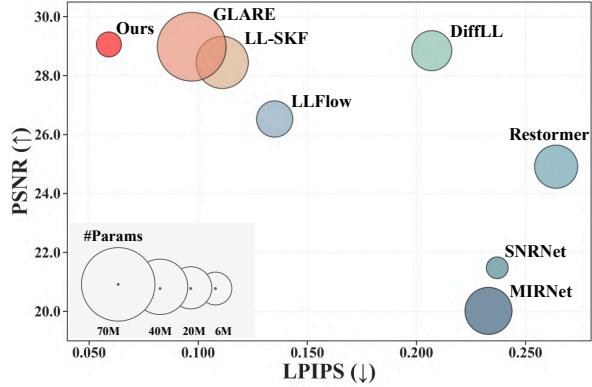


Figure 1: Performance trade-off analysis on the LOLv2-real dataset: LPIPS (perceptual quality, lower is better) versus PSNR (pixel-level fidelity, higher is better). Circle size denotes model parameter count.

on hand-crafted algorithms, often fail to accurately estimate these components, resulting in detail loss and halo artifacts in challenging low-light scenarios.

Deep learning architectures have revolutionized image enhancement methodologies. Convolutional Neural Networks (CNN) [Krizhevsky *et al.*, 2012; Chen *et al.*, 2018] have demonstrated remarkable capabilities in LLIE by learning complex non-linear mappings between degraded inputs and their reflectance-illumination decompositions. The introduction of Transformer architectures[Vaswani, 2017; Dosovitskiy *et al.*, 2021; Wang *et al.*, 2023] has further advanced the field through their self-attention mechanisms, excelling at modeling global dependencies. Nevertheless, both CNN and Transformer-based approaches frequently encounter challenges in artifact suppression and perceptual quality, particularly in regions characterized by significant information loss, severe image degradation, and extreme lighting conditions.

Generative models, particularly diffusion models [Ho *et al.*, 2020], have recently shown promise in image restoration and enhancement tasks [Miao *et al.*, 2023]. These models excel at learning complex data distributions and generating perceptually realistic results through iterative noise manipulation. However, their gradual denoising process poses significant challenges in maintaining pixel-level fidelity and preserving structural details, a critical limitation for low-light image restoration where precise detail preservation directly

<p>68 impacts enhancement quality.</p> <p>69 To address these challenges, we introduce DualPrior- 70 Retinex, abbreviated as DP-Retinex, a framework that com- 71 bines a diffusion-based generator with YUV space illumina- 72 tion maps for both global prior generation and local prior ex- 73 traction. Our framework integrates refined reflectance and il- 74 lumination maps through a Hierarchical Prior Synthesis Mod- 75 ule (HPSM), enabling comprehensive enhancement through 76 dual-prior mechanisms. This integration effectively bridges 77 the gap between pixel-level fidelity and perceptual quality, 78 as demonstrated in Fig. 1, while systematically addressing 79 both low-frequency content recovery and high-frequency de- 80 tails enhancement. The framework achieves superior detail 81 preservation while maintaining balanced brightness distribu- 82 tion. Our main contributions can be summarized as follows:</p> <ul style="list-style-type: none"> • We introduce DP-Retinex, a prior-guided framework that effectively resolves the ambiguity in Retinex de- composition. By leveraging dual-prior for accurate de- composition of reflectance and illumination maps, our method provides a robust solution for low-light image enhancement. • We develop an integrated framework that combines diffusion-based global priors with YUV space local priors through a Hierarchical Prior Synthesis Module (HPSM). This architecture simultaneously achieves low- frequency content recovery through compressed global priors while preserving high-frequency details through local priors, enabling enhanced images that maintain both high perceptual quality and pixel-level fidelity. • Through comprehensive experiments on 10 diverse datasets (5 paired benchmarks and 5 unpaired real-world datasets), our method achieves state-of-the-art perfor- mance in LLIE, consistently outperforming existing ap- proaches in both quantitative and qualitative evaluations. <h2>102 2 Related Work</h2> <h3>103 2.1 Traditional Methods</h3> <p>104 Traditional methods usually enhance images by histogram 105 equalization [Pizer <i>et al.</i>, 1987; Pisano <i>et al.</i>, 1998; Al- 106 Wadud <i>et al.</i>, 2007] and gamma correction [Wang <i>et al.</i>, 107 2009; Huang <i>et al.</i>, 2013; Rahman <i>et al.</i>, 2016]. Another rep- 108 resentative approach [Fu <i>et al.</i>, 2016] relies on the Retinex 109 theory, which decomposes the light into illumination and re- 110 flection components. This framework was further refined by 111 Guo <i>et al.</i> [Guo <i>et al.</i>, 2017], who introduced structural pri- 112 ors for illumination map optimization, substantially improv- 113 ing lighting detail preservation.</p> <h3>114 2.2 CNN and Transformer Based Methods</h3> <p>115 The emergence of deep learning has transformed low-light 116 enhancement approaches. RetinexNet [Wei <i>et al.</i>, 2018] pio- 117 neered CNN-based Retinex decomposition despite challenges 118 with dark region color distortion. Zero-DCE [Guo <i>et al.</i>, 119 2020] advanced the field through zero-reference curve esti- 120 mation learning, eliminating the need for paired training data. 121 URetinex [Wu <i>et al.</i>, 2022] further improved decomposition 122 quality through data-driven implicit prior fitting, achieving 123 enhanced noise suppression and detail preservation. In par- 124 allel, Transformer-based solutions emerged, with SNRNet [Xu 125 <i>et al.</i>, 2022] introducing dynamic SNR-guided enhancement 126 and Retinexformer [Cai <i>et al.</i>, 2023] unifying Retinex prin- 127 ciples within a single-stage architecture for SOTA performance.</p> <h3>128 2.3 Generative Model-Based Methods</h3> <p>129 Generative approaches have emerged as powerful solutions 130 for LLIE. EnlightenGAN [Jiang <i>et al.</i>, 2021] introduced un- 131 supervised enhancement through global-local discriminator 132 architecture. LLFlow [Wang <i>et al.</i>, 2022b] advanced this 133 foundation by reformulating enhancement using MAP es- 134 timation within normalizing flows. Recent diffusion-based 135 methods have shown further progress: DiffLL [Jiang <i>et al.</i>, 136 2023] leveraged wavelet-based conditional diffusion model- 137 ing with dedicated high-frequency recovery for robust detail 138 preservation, Pyd iff [Zhou <i>et al.</i>, 2023] introduced pyramid 139 diffusion sampling for efficient inference while addressing 140 color deviation through a lightweight global corrector, and 141 GLARE [Zhou <i>et al.</i>, 2025] employed generative latent repre- 142 sentation learning for illumination-reflection decomposition.</p> <p>143 Despite recent advances in low-light image enhancement, 144 existing approaches face significant challenges. CNN and 145 Transformer-based methods often introduce undesirable ar- 146 tifacts and struggle with perceptual quality [2022b], while 147 diffusion models, though promising, frequently generate un- 148 realistic results while requiring substantial computational re- 149 sources and inference time [Xia <i>et al.</i>, 2023; Liu <i>et al.</i>, 2024]. 150 These fundamental limitations motivate us to develop a more 151 robust solution that addresses these critical shortcomings.</p> <h2>152 3 Method</h2> <p>153 We present DP-Retinex, a framework that employs multi- 154 scale prior guidance to achieve enhancement of low-light im- 155 ages. Fig. 2 illustrates our architecture, which consists of 156 three main components: (1) diffusion-based global prior gen- 157 erator, (2) YUV-based local prior extractor, and (3) Hierar- 158 chical Prior Synthesis Module (HPSM).</p> <h3>159 3.1 Problem Formulation</h3> <p>160 Low-light image enhancement fundamentally relies on the 161 Retinex theory’s image decomposition principle. However, 162 in low-light scenarios, the captured image suffers from vari- 163 ous degradations, leading to a complex formulation:</p> $I_{low} = (\mathbf{R} + \mathbf{R}_d) \odot (\mathbf{L} + \mathbf{L}_d) \\ = \mathbf{R} \odot \mathbf{L} + \text{Noise} \quad (1)$ <p>164 where \mathbf{R} represents the reflectance component encoding in- 165 trinsic object properties, \mathbf{L} denotes the illumination compo- 166 nent, and \odot indicates element-wise multiplication. \mathbf{R}_d char- 167 acterizes reflectance degradation from sensor noise and com- 168 pression artifacts, while \mathbf{L}_d captures illumination deviations 169 due to ambient light interference and sensor limitations.</p> <p>170 To obtain enhanced normal-light images without noise 171 degradation, we introduce a novel Hierarchical Prior Syn- 172 thesis Module (HPSM) that systematically integrates com- 173 prehensive dual-prior information. Our fundamental key in- 174 sight is that effective enhancement requires both global pri- 175 ors for low-frequency content recovery and local priors for</p>

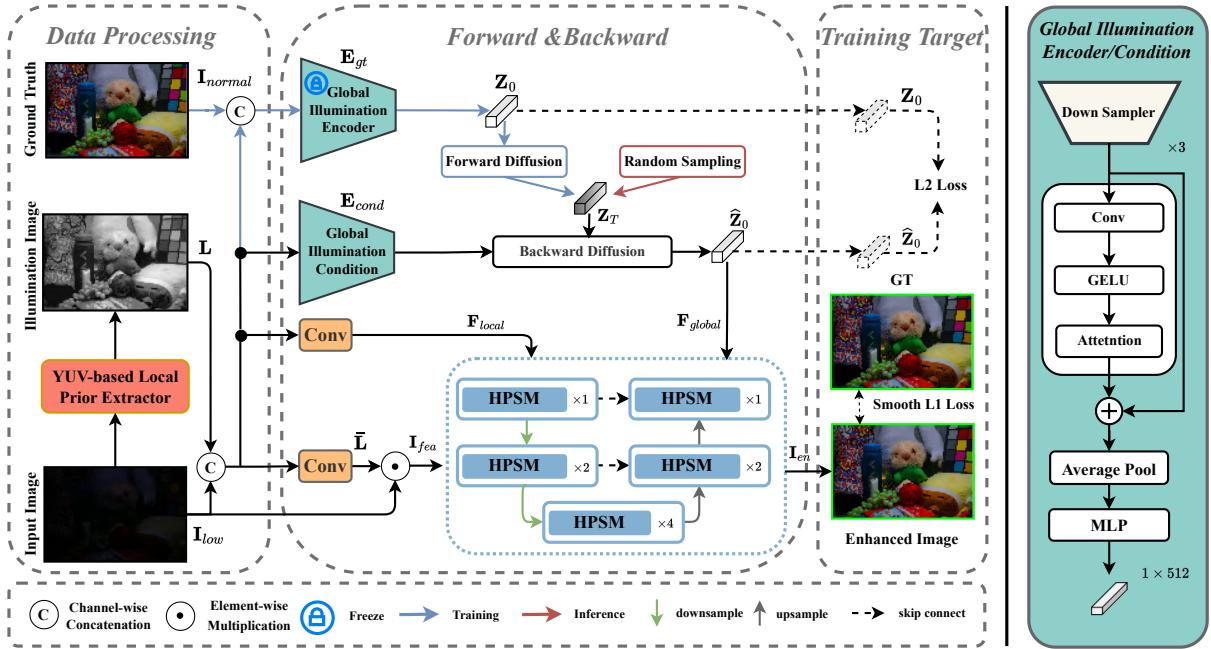


Figure 2: Overview of DP-Retinex architecture. The framework incorporates diffusion-based generator for global prior generation and YUV-based local prior extraction, which optimize reflectance and illumination decomposition via Hierarchical Prior Synthesis Module (HPSM). The dual-prior design enables enhancement by combining low-frequency content recovery with high-frequency details preservation.

high-frequency detail enhancement. This hierarchical synthesis mechanism enables comprehensive enhancement while maintaining both perceptual quality and pixel-level fidelity.

3.2 Local Prior Extraction

Low-light image enhancement requires robust separation of the reflectance component \mathbf{R} from the degraded observation \mathbf{I}_{low} . Given RGB channels representing red, green, and blue intensities, while Y denotes the luminance (brightness) component, we propose a lightweight local prior extraction method that effectively utilizes the YUV color space's natural luminance-chrominance separation properties (detailed structure in supplemental material). The YUV space provides a perceptually linear luminance channel ($Y = 0.299R + 0.587G + 0.114B$) while maintaining illumination-independent chrominance components, enabling more accurate reflectance separation. Our approach focuses on a multiplicative illumination decomposition strategy:

$$\bar{\mathbf{L}} \odot \mathbf{L} = \mathbf{L}_1 \quad (2)$$

where $\bar{\mathbf{L}}$ represents the reciprocal illumination map learned through convolutional operations on \mathbf{L} , and \mathbf{L}_1 is an all-ones matrix. This formulation ensures numerical stability compared to traditional division-based methods. The initial feature representation \mathbf{I}_{fea} is extracted through:

$$\mathbf{I}_{fea} = \mathbf{I}_{low} \odot \bar{\mathbf{L}} = \mathbf{R} + \text{Noise} \quad (3)$$

To integrate with our neural architecture, we employ a convolutional operation that maps the single-channel illumination map \mathbf{L} to a C -dimensional feature space \mathbf{F}_{local} , preserving high-frequency details (e.g., edges, textures, and fine

structural information) while ensuring dimensional compatibility with subsequent processing stages.

3.3 Global Prior Generation

The reflectance component \mathbf{R} , derived from our local prior extraction in Eq. (3), often suffers from missing information and degradation, further compounded by the presence of noise. To address these issues, we propose a diffusion-based prior generation mechanism that focuses on two critical aspects: guiding the recovery of low-frequency degraded content in \mathbf{R} and effectively denoising the Noise term.

As shown in Fig. 2, our approach begins by leveraging a pre-trained Global Illumination Encoder \mathbf{E}_{gt} to generate an initial global prior $\mathbf{Z}_0 \in \mathbb{R}^{C_z}$ that encapsulates the ground truth representation of low-frequency characteristics. This initial prior serves as a foundation for our diffusion process, which progressively refines the representation through careful noise management. During the forward diffusion process, we inject Gaussian noise following a predefined schedule:

$$\mathbf{Z}_T = \sqrt{\bar{\alpha}_T} \mathbf{Z}_0 + \sqrt{1 - \bar{\alpha}_T} \epsilon \quad (4)$$

where $\bar{\alpha}_T$ represents the cumulative product of noise scheduling coefficients, and ϵ is randomly sampled from $\mathcal{N}(0, I)$. This controlled noise injection ensures comprehensive coverage of the content distribution space.

To effectively denoise and recover the degraded content, we introduce a Global Illumination Condition \mathbf{E}_{cond} that shares architectural similarities with \mathbf{E}_{gt} but operates exclusively on the degradation branch. This encoder extracts a conditional embedding $c \in \mathbb{R}^{C_z}$ from low-light images \mathbf{I}_{low} ,

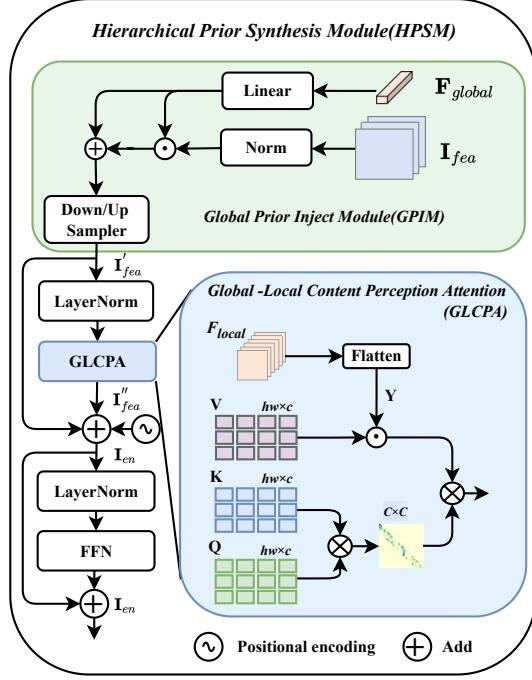


Figure 3: The Hierarchical Prior Synthesis Module (HPSM) consists of two modules: the Global Prior Inject Module (top) and the Global-Local Content Perception Attention (bottom).

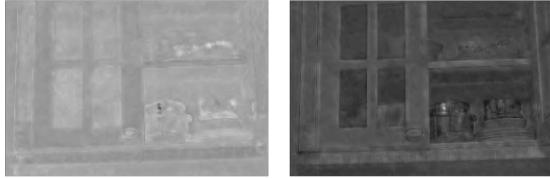


Figure 4: Visual comparison showing the enhancement achieved by GPIM through the incorporation of global statistical priors. (Left: Original image; Right: GPIM enhanced image)

which guides our iterative denoising process:

$$\hat{\mathbf{Z}}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\hat{\mathbf{Z}}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\hat{\mathbf{Z}}_t, t, c) \right) + \sigma_t \epsilon \quad (5)$$

The refined latent representation $\hat{\mathbf{Z}}_0$ serves as our global prior \mathbf{F}_{global} , which provides comprehensive guidance for low-frequency content and illumination recovery while suppressing the noise term \mathbf{Noise} in Eq. (3). To effectively integrate this global information with input image features, we introduce the Global Prior Inject Module (GPIM):

$$I'_{fea} = \text{GPIM}(I_{fea}, \mathbf{F}_{global}) = \mathbf{R} \quad (6)$$

The GPIM, illustrated in Fig. 3, addresses the dimensional mismatch between the one-dimensional global prior and two-dimensional feature space through learned adaptive alignment. This integration enables our model to perform low-frequency content recovery (e.g., smooth regions, gradually varying brightness information, and global illumination consistency), significantly improving the recovery of degraded

image content shown in Fig. 4. Through this approach to prior generation and integration, our method can achieve more coherent and perceptually balanced enhancement results across diverse low-light scenarios.

3.4 Hierarchical Prior Synthesis Module

To effectively integrate the global and local priors introduced in Sections 3.2 and 3.3, we propose HPSM that preserves perceptual quality and structural fidelity. Our HPSM, illustrated in Fig. 3, incorporates the dual-prior approach: it leverages the global prior \mathbf{F}_{global} to guide low-frequency content reconstruction while utilizing the local prior \mathbf{F}_{local} for high-frequency details preservation. The enhanced feature map $I'_{fea} \in \mathbb{R}^{H \times W \times C}$, obtained from the GPIM, serves as the foundation for our attention mechanism.

Following standard transformer architecture principles, we project I'_{fea} into three distinct embedding spaces to obtain query (\mathbf{Q}), key (\mathbf{K}), and value (\mathbf{V}) representations:

$$\mathbf{Q} = I'_{fea} \mathbf{W}_Q, \quad \mathbf{K} = I'_{fea} \mathbf{W}_K, \quad \mathbf{V} = I'_{fea} \mathbf{W}_V, \quad (7)$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{C \times d_k}$ represents learnable transformation matrices, and d_k denotes the dimensionality of the transformed embedding space.

Moreover, to capture spatially varying illumination conditions, we reshape the local prior $\mathbf{F}_{local} \in \mathbb{R}^{H \times W \times C}$ into a matrix $\mathbf{Y} \in \mathbb{R}^{HW \times C}$ and partition it into k attention heads:

$$\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_k]. \quad (8)$$

This partitioning leverages the local prior to enhance high-frequency details and illumination characteristics. Specifically, the fused attention output \mathbf{F}_{out} is then computed by:

$$\mathbf{F}_{out} = (\mathbf{Y} \odot \mathbf{V}) \text{Softmax}\left(\frac{\mathbf{K}^T \mathbf{Q}}{\alpha_i}\right) + \mathbf{P} \quad (9)$$

where α_i is a learnable scaling factor, and \mathbf{P} denotes the positional encoding that preserves spatial structure. Consequently, the product $\mathbf{Y} \odot \mathbf{V}$ injects local prior into the attention process, thereby empowering the network to adaptively focus on regions that require distinct enhancement strategies. The attention output \mathbf{F}_{out} is reshaped to spatial dimensions $I''_{fea} \in \mathbb{R}^{H \times W \times C}$ for feature space consistency.

The above formulations collectively constitute our Global-Local Content Perception Attention (GLCPA) module. The complete attention mechanism can be summarized as follows:

$$I''_{fea} = \text{GLCPA}(I'_{fea}, \mathbf{F}_{local}) = \mathbf{R} \odot \mathbf{L}. \quad (10)$$

Finally, the residual connection combines initial and refined features through a linear summation:

$$I_{en} = I'_{fea} + I''_{fea} \quad (11)$$

where I_{en} represents the enhanced image that combines both the initially enhanced features and the refined features obtained through GLCPA, effectively integrating low-frequency content with high-frequency details.

Method	Reference	LOLv1			LOLv2-real			LOLv2-synthetic			Params(M)
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	
LIME	TIP16	17.54	0.531	0.387	17.48	0.505	0.428	17.34	0.520	0.471	-
RetinexNet	BMVC 18	16.77	0.462	0.417	17.71	0.652	0.436	14.80	0.610	0.206	0.62
MIRNet	ECCV 20	24.13	0.830	0.250	20.02	0.820	0.233	21.94	0.876	0.169	31.76
SNRNet	CVPR 22	24.61	0.842	0.233	21.48	0.849	0.237	24.14	0.928	0.189	4.01
Restormer	CVPR 22	20.61	0.797	0.288	24.91	0.851	0.264	21.41	0.830	-	26.13
LLFlow	AAAI 22	25.19	0.870	0.113	26.53	0.892	0.135	26.23	0.943	-	17.42
Retinexformer	ICCV 23	27.18	0.850	0.082	27.70	0.856	0.108	29.03	0.938	0.040	1.61
DiffLL	SIG Asia 23	26.33	0.845	0.217	28.85	0.876	0.207	-	-	-	22.08
LL-SKF	CVPR 23	26.80	0.879	0.105	28.45	0.905	0.111	29.11	0.953	-	39.91
GSAD	NeurIPS 23	27.83	0.877	0.091	28.82	0.895	0.095	28.67	0.944	0.047	17.36
GLARE	ECCV 24	27.35	0.883	0.083	28.98	0.905	0.097	29.84	0.958	-	71.71
Ours	—	26.60	0.866	0.059	29.06	0.894	0.059	29.93	0.948	0.029	6.45

Table 1: Method comparison table on LOLv1, LOLv2-real, and LOLv2-synthetic datasets. The best results are highlighted in bold and red. ‘-’ denotes the result or implementation is not available.

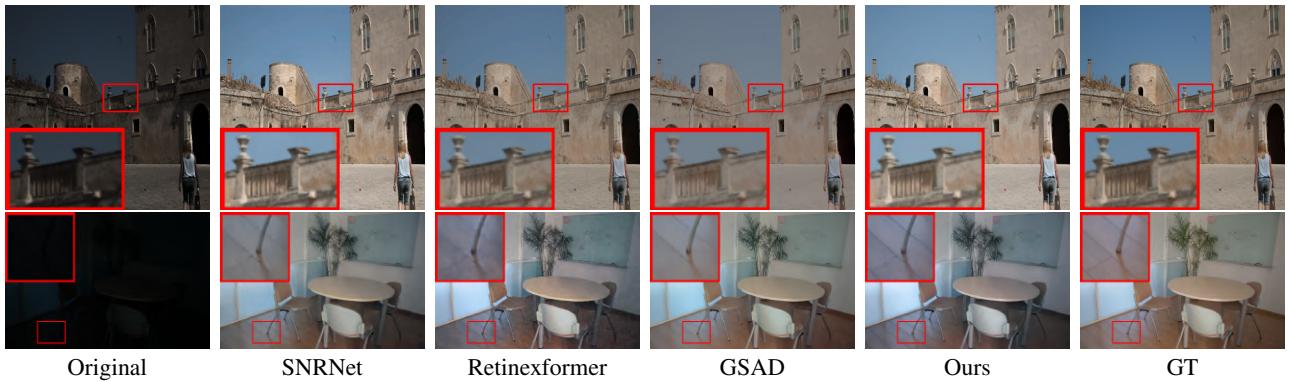


Figure 5: Qualitative comparison of different enhancement methods on LOLv2-synthetic (top) and LOLv2-real (bottom) dataset.

4 Experiments

4.1 Datasets and Settings

Datasets and Metrics. We evaluate our method on multiple benchmark datasets: LOL-v1 (485/15 train/test split) [Wei *et al.*, 2018] and LOL-v2 [Yang *et al.*, 2021], which comprises real-captured (689/100 split) and synthetic (900/100 split) subsets. For SDSD evaluation [Wang *et al.*, 2021], we assess performance on indoor (62/6 train/test split) and outdoor (116/10 train/test split) scene collections.

The evaluation metrics include Peak Signal-to-Noise Ratio (PSNR) for pixel-wise accuracy, Structural Similarity Index (SSIM) [Wang *et al.*, 2004] for structural fidelity, and Learned Perceptual Image Patch Similarity (LPIPS) [Zhang *et al.*, 2018] for perceptual quality assessment. We employ PSNR, SSIM, and LPIPS metrics for the LOL dataset evaluation, while PSNR and SSIM are used for SDSD assessment.

Additionally, we conduct cross-dataset evaluation on five non-reference datasets (DICM [Lee *et al.*, 2013], LIME [2017], MEF [Ma *et al.*, 2015], NPE [Wang *et al.*, 2013], and VV [Vonikakis *et al.*, 2017]) using no-reference metrics: Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [Mittal *et al.*, 2012] and Natural Image Quality Evaluator (NIQE) [Mittal *et al.*, 2013].

Implementation Details. We pre-train E_{gt} on paired ground-truth samples for content representation. The training combines smooth L1 loss for enhanced image and local prior extraction, with L2 loss for global prior constraint (Detailed formulations are provided in supplementary). Our PyTorch implementation uses Adam [Kinga *et al.*, 2015] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ for 1.5×10^5 iterations. The learning rate follows cosine annealing [Abelson *et al.*, 1985] from 3×10^{-4} to 5×10^{-6} . Training uses 192×192 random patches with rotation and flipping augmentations at batch size 8.

4.2 Experimental Results

Quantitative Evaluation. Our approach achieves state-of-the-art results across multiple metrics on the LOL dataset series, as shown in Table 1. We attain the best LPIPS scores of 0.059, 0.059, and 0.029 on LOLv1, LOLv2-real, and LOLv2-synthetic respectively, substantially outperforming recent methods including GSAD [Hou *et al.*, 2024] and GLARE. The method also achieves leading PSNR scores of 29.06dB on LOLv2-real and 29.93dB on LOLv2-synthetic, while maintaining competitive SSIM metrics across all variants. Comprehensive experiments across multiple benchmark datasets demonstrate the superior performance in low-light image enhancement.

Methods	DICM		LIME		MEF		NPE		VV	
	BRI↓	NIQE↓	BRI↓	NIQE↓	BRI↓	NIQE↓	BRI↓	NIQE↓	BRI↓	NIQE↓
KinD	48.72	5.15	39.91	5.03	49.94	5.47	36.85	4.98	50.56	4.30
ZeroDCE	27.56	4.58	20.44	5.82	17.32	4.93	20.72	5.45	34.66	4.81
LLFlow	26.36	4.06	23.12	4.59	30.27	4.70	28.86	4.67	31.67	4.04
SNRNet	37.35	4.71	39.22	5.74	31.28	4.18	26.65	4.32	78.72	9.87
PairLIE	33.31	4.03	25.23	4.58	27.53	4.06	28.27	4.18	39.13	3.57
Ours	17.62	3.83	16.01	4.55	12.70	3.63	17.24	3.85	30.56	4.95

Table 2: A comparison of BRISQUE and NIQE scores across various methods was conducted on the DICM, LIME, MEF, NPE, and VV datasets. The evaluation was performed by testing these no-reference samples using a model pre-trained on the LOL dataset. The best results are marked in bold and red.

Method	SDSD-indoor		SDSD-outdoor	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
KinD	21.95	0.672	21.97	0.654
MIRNet	24.38	0.864	27.13	0.837
DRBN	24.08	0.868	25.77	0.841
Retinexformer	29.77	0.896	29.84	0.877
SNRNet	29.44	0.894	28.66	0.866
GLARE	30.10	0.896	30.85	0.884
Ours	30.69	0.902	29.57	0.851

Table 3: Method comparison on SDSD-indoor and SDSD-outdoor datasets. The best results are marked in bold and red.

To assess model generalization ability, we conducted comprehensive experiments on five diverse datasets without reference samples. As shown in Table 2, our method demonstrates superior performance in BRISQUE metrics, achieving optimal scores across all datasets (DICM: 17.62, LIME: 16.01, MEF: 12.70, NPE: 17.24, VV: 30.56). The NIQE evaluation further validates our approach, yielding best performance in four datasets (DICM: 3.83, LIME: 4.55, MEF: 3.63, NPE: 3.85), with competitive results on VV (4.95). Notably, compared to existing methods, our approach reduces BRISQUE values by 20-36%, while maintaining natural enhancement quality across varying lighting conditions.

Experiments on the challenging SDSD dataset further validate the method’s effectiveness, as shown in Table 3. The approach achieves superior performance on indoor scenes with PSNR/SSIM scores of 30.69/0.902, surpassing recent state-of-the-art methods GLARE and Retinexformer. While GLARE exhibits marginally better performance on outdoor scenes, our method maintains competitive results across diverse lighting conditions and scene types. Experiments across multiple benchmark datasets demonstrate our method’s superior performance in low-light image enhancement.

Qualitative Analysis. We evaluate our method through visual comparisons on the LOLv2-synthetic and LOLv2-real datasets, as shown in Fig. 5. The results demonstrate superior performance across controlled and real-world scenarios. In the architectural scene, our method achieves exceptional low-frequency content recovery while avoiding the noise artifacts present in SNRNet and color distortions seen in GSAD’s results. The indoor scene validates our method’s high-frequency details preservation capabilities, particularly in floor texture rendition, producing results that align closest

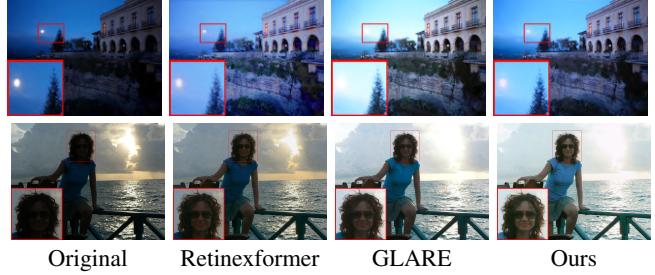


Figure 6: Visual results of cross-dataset evaluation on unpaired real-world datasets. These four images are from LIME VV.

with the ground truth (GT). These observations confirm our method’s effectiveness across diverse lighting conditions.

The robust generalization capability of our approach for reference-free scenarios is clearly demonstrated in Fig. 6. In the twilight architectural scene (top), our method achieves balanced enhancement while preserving the natural evening ambiance, effectively avoiding the over-brightening artifacts observed in GLARE’s results. Additionally, The backlit seaside portrait (bottom) further validates our method’s superior performance through its preservation of facial details and natural contrast between the subject and bright background, whereas competing approaches like Retinexformer and GLARE show significant limitations in handling such extreme lighting conditions. More qualitative results are provided in the supplementary material.

4.3 Object Detection in Low-Light Conditions

We evaluate our method’s effectiveness as a preprocessing step for object detection using the ExDark [Loh and Chan, 2019b] dataset, which comprises 7,363 low-light images across 12 object categories. The evaluation protocol employs YOLOv3 [Redmon, 2018] as the detector, trained from scratch on 5,896 images enhanced by various low-light enhancement methods.

Quantitative Analysis. Our method achieves state-of-the-art performance (mAP: 77.94%), surpassing MBLLEN, KinD, and Retinexformer, as shown in Table 4. It excels in detecting challenging objects, with superior results for Bicycle (82.6%), Boat (80.0%), Bus (93.1%), and Cat (75.0%). This performance across diverse categories, from large objects (buses, cars) to smaller ones (bottles, cups), demon-

Methods	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motor	People	Table	Mean↑
KIND	81.4	75.2	77.6	92.0	84.9	71.3	73.3	79.6	78.1	78.1	81.2	54.1	77.23
MBLLEN	81.7	76.5	78.2	92.2	84.8	71.8	71.0	78.4	76.4	75.5	80.8	56.3	76.97
Zero-DCE	82.2	77.0	75.3	91.6	83.5	72.1	72.8	78.2	80.0	78.9	82.0	56.5	77.52
IAT	80.6	79.0	78.8	92.2	84.7	68.9	71.3	79.1	77.5	77.9	82.4	58.4	77.58
Retinexformer	82.2	78.1	78.3	91.8	84.3	70.7	69.5	78.8	81.4	78.7	82.0	58.5	77.86
GLARE	82.6	77.2	75.1	91.1	85.0	71.5	72.5	78.1	79.6	78.3	82.0	54.5	77.31
Ours	82.6	80.0	77.1	93.1	83.9	75.0	70.1	78.0	79.4	78.1	82.0	56.0	77.94

Table 4: Enhancement Methods for YOLOv3 Detection on ExDark Dataset (mAP↑). Best results marked in red.

Variants	Components		Metrics		Params (M)
	Global	Local	PSNR↑	LPIPS↓	
Baseline	×	×	28.44	0.041	1.59
(1)	✓	△	28.89	0.034	4.75
(2)	✓	✗	28.89	0.036	4.71
(3)	△	✓	29.06	0.034	3.38
(4)	✗	✓	28.45	0.037	1.65
Full Model	✓	✓	29.93	0.029	6.45

Table 5: Ablation study on global and local priors in DP-Retinex. Best results are marked in **bold**. ✓ denotes component inclusion, ✗ signifies removal, and △ represents partial modification. 1: Substitute YUV-based Local Prior extraction with an average grayscale map of input image. 2: Eliminate Local Prior module entirely. 3: Reduce parameters in GCPIM. 4: Omit Global Prior module completely. Params indicate parameters in millions.



Figure 7: Visual comparison of object detection in low-light (left) and enhanced (right) scenes by our method on the Exdark dataset.

The crucial role of global prior integration is demonstrated through variants (3) and (4). While reducing GCPIM’s alignment capacity (variant 3) shows moderate performance decline to 29.06dB PSNR, completely eliminating the global prior module (variant 4) leads to substantial degradation to 28.45dB PSNR and 0.037 LPIPS. These results highlight two critical findings: the significant parameters in GCPIM are necessary for effective alignment between one-dimensional global priors and two-dimensional feature maps, while the global prior module proves indispensable for comprehensive scene content reconstruction, as substantiated by the significant performance deterioration in its absence.

These comprehensive ablation results substantiate that the proposed dual-prior approach effectively leverages complementary information from both global content and local details, empirically validating our architectural design choices for robust low-light image enhancement.

5 Conclusion

This paper introduces DualPrior-Retinex, a novel framework addressing fundamental challenges in Retinex-based low-light image enhancement. By combining diffusion-based global priors with YUV-space extracted local priors, the method achieves superior performance in balancing perceptual quality and pixel-level fidelity. Extensive experimental results demonstrate state-of-the-art performance across multiple benchmarks, validating the effectiveness of the dual-prior design for low-light image enhancement tasks.

References

- [Abelson *et al.*, 1985] Harold Abelson, Gerald Jay Sussman, and Julie Sussman. *Structure and Interpretation of Computer Programs*. MIT Press, Cambridge, Massachusetts, 1985.
- [Al-Wadud *et al.*, 2007] Mohammad Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE transactions on consumer electronics*, 53(2):593–600, 2007.
- [Cai *et al.*, 2023] 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 *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12504–12513. IEEE, 2023.
- [Chen *et al.*, 2018] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3291–3300. IEEE, 2018.
- [Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Jakob Uszkoreit, Neil Houlsby, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *2021 International Conference on Learning Representations (ICLR)*, 2021.
- [Fu *et al.*, 2016] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2782–2790. IEEE, 2016.
- [Guo *et al.*, 2017] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing (TIP)*, 26(2):982–993, 2017.
- [Guo *et al.*, 2020] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1777–1786. IEEE, 2020.
- [Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems (NeurIPS)*, 33:6840–6851, 2020.
- [Hou *et al.*, 2024] 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 (NeurIPS)*, 36, 2024.
- [Huang *et al.*, 2013] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing (TIP)*, 22(3):1032–1041, 2013.
- [Jiang *et al.*, 2021] 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 (TIP)*, 30:2340–2349, 2021.
- [Jiang *et al.*, 2023] 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.
- [Jobson *et al.*, 1997] D.J. Jobson, Z. Rahman, and G.A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing (TIP)*, 6(7):965–976, 1997.
- [Kinga *et al.*, 2015] D Kinga, Jimmy Ba Adam, et al. A method for stochastic optimization. In *International conference on learning representations (ICLR)*, 2015.
- [Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems (NeurIPS)*, 25, 2012.
- [Land and McCann, 1971] Edwin Herbert Land and John J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61 1:1–11, 1971.
- [Lee *et al.*, 2013] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE Transactions on Image Processing (TIP)*, 22(12):5372–5384, 2013.
- [Li *et al.*, 2022] 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.
- [Liu *et al.*, 2024] Qingguo Liu, Chenyi Zhuang, Pan Gao, and Jie Qin. Cdformer: When degradation prediction embraces diffusion model for blind image super-resolution. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7455–7464. IEEE, 2024.
- [Loh and Chan, 2019a] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding (CVIU)*, 178:30–42, 2019.
- [Loh and Chan, 2019b] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding (CVIU)*, 178:30–42, 2019.
- [Ma *et al.*, 2015] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing (TIP)*, 24(11):3345–3356, 2015.
- [Miao *et al.*, 2023] Yuchun Miao, Lefei Zhang, Liangpei Zhang, and Dacheng Tao. Dds2m: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12086–12096. IEEE, 2023.
- [Mittal *et al.*, 2012] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing (TIP)*, 21(12):4695–4708, 2012.
- [Mittal *et al.*, 2013] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013.
- [Pisano *et al.*, 1998] Etta D. Pisano, Shuquan Zong, Bradley M. Hemminger, Marla DeLuca, R. Eugene Johnston, Keith Muller, M. Patricia Braeuning, and Stephen M. Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital Imaging*, 11(4):193–200, nov 1998.
- [Pizer *et al.*, 1987] Stephen M. Pizer, E. Philip Amburn, John D. Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, sep 1987.

- 567 [Rahman *et al.*, 2016] Shanto Rahman, Md Mostafijur Rahman,
 568 M. Abdullah-Al-Wadud, Golam Dastegir Al-Quaderi, and Mo-
 569 hammad Shoyaib. An adaptive gamma correction for image en-
 570 hancement. *EURASIP Journal on Image and Video Processing*,
 571 2016(1):1–13, oct 2016.
- 572 [Redmon, 2018] Joseph Redmon. Yolov3: An incremental im-
 573 provement. *arXiv preprint arXiv:1804.02767*, 2018.
- 574 [Vaswani, 2017] A Vaswani. Attention is all you need. *Advances in
 575 Neural Information Processing Systems (NeurIPS)*, 2017.
- 576 [Vonikakis *et al.*, 2017] Vassilios Vonikakis, Rigas Kouskouridas,
 577 and Antonios Gasteratos. On the evaluation of illumination
 578 compensation algorithms. *Multimedia Tools and Applications*,
 579 77(8):9211–9231, may 2017.
- 580 [Wang *et al.*, 2004] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P.
 581 Simoncelli. Image quality assessment: from error visibility to
 582 structural similarity. *IEEE Transactions on Image Processing
 (TIP)*, 13(4):600–612, 2004.
- 583 [Wang *et al.*, 2009] Zhi-Guo Wang, Zhi-Hu Liang, and Chun-Liang
 584 Liu. A real-time image processor with combining dynamic con-
 585 trast ratio enhancement and inverse gamma correction for pdp.
 586 *Displays*, 30(3):133–139, 2009.
- 587 [Wang *et al.*, 2013] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and
 588 Bo Li. Naturalness preserved enhancement algorithm for non-
 589 uniform illumination images. *IEEE Transactions on Image Pro-
 590 cessing (TIP)*, 22(9):3538–3548, 2013.
- 591 [Wang *et al.*, 2021] Ruixing Wang, Xiaogang Xu, Chi-Wing Fu,
 592 Jiangbo Lu, Bei Yu, and Jiaya Jia. Seeing dynamic scene in the
 593 dark: A high-quality video dataset with mechatronic alignment.
 594 In *2021 IEEE/CVF International Conference on Computer Vision
 (ICCV)*, pages 9700–9709. IEEE, 2021.
- 595 [Wang *et al.*, 2022a] Hai Wang, Yanyan Chen, Yingfeng Cai, Long
 596 Chen, Yicheng Li, Miguel Angel Sotelo, and Zhixiong Li. Sfnet-
 597 nn: An improved sfnet algorithm for semantic segmentation of
 598 low-light autonomous driving road scenes. *IEEE Transactions
 599 on Intelligent Transportation Systems*, 23(11):21405–21417, nov
 600 2022.
- 601 [Wang *et al.*, 2022b] Yufei Wang, Renjie Wan, Wenhan Yang, Hao-
 602 liang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhance-
 603 ment with normalizing flow. In *Proceedings of the AAAI Confer-
 604 ence on Artificial Intelligence (AAAI)*, volume 36, pages 2604–
 605 2612, 2022.
- 606 [Wang *et al.*, 2023] Tao Wang, Kaihao Zhang, Tianrun Shen, Wen-
 607 han Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-
 608 light image enhancement: A benchmark and transformer-based
 609 method. In *Proceedings of the AAAI Conference on Artificial In-
 610 telligence (AAAI)*, volume 37, pages 2654–2662, 2023.
- 611 [Wang *et al.*, 2024] Xinzhe Wang, Kang Ma, Qiankun Liu, Yunhao
 612 Zou, and Ying Fu. Multi-object tracking in the dark. In *2024
 613 IEEE/CVF Conference on Computer Vision and Pattern Recog-
 614 nition (CVPR)*, pages 382–392. IEEE, 2024.
- 615 [Wei *et al.*, 2018] Chen Wei, Wenjing Wang, Wenhan Yang, and Ji-
 616 aying Liu. Deep retinex decomposition for low-light enhance-
 617 ment. In *2018 British Machine Vision Conference (BMVC)*, 2018.
- 618 [Wu *et al.*, 2022] Wenhui Wu, Jian Weng, Pingping Zhang,
 619 Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net:
 620 Retinex-based deep unfolding network for low-light image en-
 621 hancement. In *2022 IEEE/CVF Conference on Computer Vision
 622 and Pattern Recognition (CVPR)*, pages 5891–5900. IEEE, 2022.
- 623 [Xia *et al.*, 2023] Bin Xia, Yulun Zhang, Shiyan Wang, Yitong
 624 Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc
 625 Van Gool. Diffir: Efficient diffusion model for image restora-
 626 tion. In *2023 IEEE/CVF International Conference on Computer
 627 Vision (ICCV)*, pages 13095–13105. IEEE, 2023.
- 628 [Xu *et al.*, 2022] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and
 629 Jiaya Jia. Snr-aware low-light image enhancement. In *2022
 630 IEEE/CVF Conference on Computer Vision and Pattern Recog-
 631 nition (CVPR)*, pages 17714–17724. IEEE, 2022.
- 632 [Yang *et al.*, 2021] Wenhan Yang, Wenjing Wang, Haofeng Huang,
 633 Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep
 634 retinex network for robust low-light image enhancement. *IEEE
 635 Transactions on Image Processing (TIP)*, 30:2072–2086, 2021.
- 636 [Zhang *et al.*, 2018] Richard Zhang, Phillip Isola, Alexei A Efros,
 637 Eli Shechtman, and Oliver Wang. The unreasonable effectiveness
 638 of deep features as a perceptual metric. In *2018 IEEE/CVF Con-
 639 ference on Computer Vision and Pattern Recognition (CVPR)*,
 640 pages 586–595. IEEE, 2018.
- 641 [Zhou *et al.*, 2023] Dewei Zhou, Zongxin Yang, and Yi Yang. Pyra-
 642 mid diffusion models for low-light image enhancement. In *2023
 643 Thirty-Second International Joint Conference on Artificial Intel-
 644 ligence (IJCAI)*, pages 1795–1803. International Joint Confer-
 645 ences on Artificial Intelligence Organization, 2023.
- 646 [Zhou *et al.*, 2025] Han Zhou, Wei Dong, Xiaohong Liu,
 647 Shuaicheng Liu, Xiongkuo Min, Guangtao Zhai, and Jun
 648 Chen. Glare: Low light image enhancement via generative latent
 649 feature based codebook retrieval. In *2024 European Conference
 650 on Computer Vision (ECCV)*, pages 36–54. Springer, 2025.
- 651