

# SPJFNet: Self-Mining Prior-Guided Joint Frequency Enhancement for Ultra-Efficient Dark Image Restoration

Tongshun Zhang<sup>1,2</sup>, Pingping Liu<sup>1,2\*</sup>, Zijian Zhang<sup>1,2</sup>, Qiuzhan Zhou<sup>3</sup>

<sup>1</sup> College of Computer Science and Technology, Jilin University

<sup>2</sup> Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University

<sup>3</sup> College of Communication Engineering, Jilin University

tszhang23@mails.jlu.edu.cn, {liupp, zhangzijian, zhouqz}@jlu.edu.cn

## Abstract

Current dark image restoration methods suffer from severe efficiency bottlenecks, primarily stemming from: computational burden and error correction costs associated with reliance on external priors (manual or cross-modal); redundant operations in complex multi-stage enhancement pipelines; and indiscriminate processing across frequency components in frequency-domain methods, leading to excessive global computational demands. To address these challenges, we propose an Efficient Self-Mining Prior-Guided Joint Frequency Enhancement Network (SPJFNet). Specifically, we first introduce a Self-Mining Guidance Module (SMGM) that generates lightweight endogenous guidance directly from the network, eliminating dependence on external priors and thereby bypassing error correction overhead while improving inference speed. Second, through meticulous analysis of different frequency domain characteristics, we reconstruct and compress multi-level operation chains into a single efficient operation via lossless wavelet decomposition and joint Fourier-based advantageous frequency enhancement, significantly reducing parameters. Building upon this foundation, we propose a Dual-Frequency Guidance Framework (DFGF) that strategically deploys specialized high/low frequency branches (wavelet-domain high-frequency enhancement and Fourier-domain low-frequency restoration), decoupling frequency processing to substantially reduce computational complexity. Rigorous evaluation across multiple benchmarks demonstrates that SPJFNet not only surpasses state-of-the-art performance but also achieves significant efficiency improvements, substantially reducing model complexity and computational overhead.

**Code** — <https://github.com/bywzts/SPJFNet>

## Introduction

Images taken in dark conditions often suffer from insufficient illumination, degrading visual quality and hindering downstream vision applications like autonomous driving and object detection (Du, Shi, and Deng 2024; Li et al. 2024; Zhao et al. 2025; Guo et al. 2026; Yu et al. 2025; Xiao et al. 2025). With the rise of deep learning, learning-based methods (Cai et al. 2023; Zhang et al. 2025a; Gu et al.

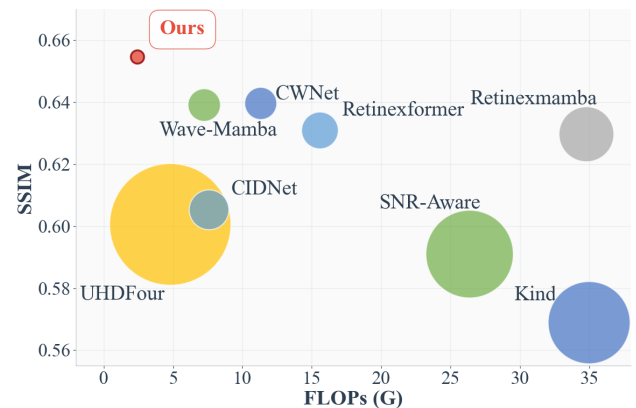


Figure 1: SSIM vs. Computation overhead. The area of each bubble represents the number of parameters.

2025) have surpassed traditional approaches but often come with high computational complexity. Given the limited cues in extremely dark images, many methods (Wu et al. 2023; Zhang et al. 2024; Zou et al. 2024b) introduce pre-trained models or additional prediction networks to generate modal priors. However, pretrained models and the predictive network are likely to suffer from generalization issues, resulting in accumulated errors and high computational overhead.

Recently, Fourier-based methods (Huang et al. 2022; Li et al. 2023; Zhang et al. 2024; Feijoo et al. 2025) have gained significant attention due to their efficiency in image restoration tasks. FourLLIE (Wang, Wu, and Zhi 2023) demonstrated that Fourier domain processing can handle global information as effectively as transformer while significantly reducing parameter overhead by replacing transformer layers in SNR-Aware (Xu et al. 2022). However, the global nature of the Fourier transform (where spatial locality is not preserved) means that downsampling critically damages high-frequency details essential for restoration. This compels Fourier-based methods to maintain full spatial resolution, inevitably forcing a sacrifice in channel dimensions to contain model complexity, thereby restricting the expressive capacity of the model’s feature space. Wavelet transform (Zou et al. 2024a; Jiang et al. 2023; Zhang et al. 2025a)

\*Corresponding author

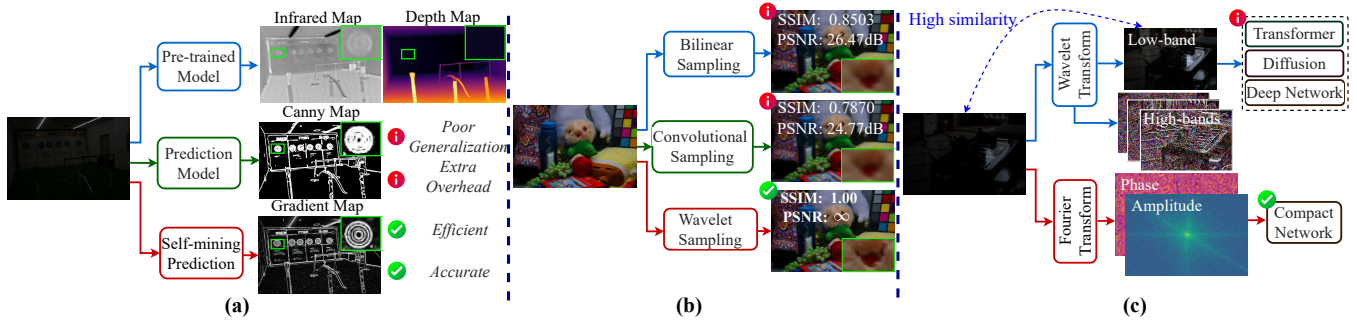


Figure 2: Motivation analysis. (a) Three approaches for prior generation: introducing pre-trained models to generate modal priors, adding prediction networks to predict modal priors, and self-mining priors from the image itself; (b) SSIM and PSNR computed between images after three rounds of down-and-up sampling and the original image, showing that wavelet sampling achieves lossless sampling; (c) Low-frequency information from wavelet transform requires sophisticated model processing, while amplitude and phase in the Fourier domain can be efficiently processed through compact structures.

represents another common frequency domain transformation in LLIE, capable of effectively separating content from noise. However, the low-frequency subbands retain tightly coupled spatial and frequency information, particularly containing both the dominant image content and low-frequency distortions/noise. This residual complexity, akin to a degraded version of the original image, necessitates sophisticated processing modules to effectively decouple content from corruption, offsetting the efficiency gains from wavelet decomposition.

Based on the limitations, we present our motivations:

**First**, as shown in Fig. 2(a), there are two primary approaches for introducing prior clues in dark image restoration. The first approach utilizes pre-trained models to generate modal priors, where DMFourLLIE (Zhang et al. 2024) leverages pre-trained models to generate infrared and depth map modal priors for enhanced guidance of phase component recovery, but neglects the generalization error issues introduced by pre-trained models. The second approach introduces prediction modules to generate priors, where SMG-LLE (Xu, Wang, and Lu 2023) designs a dedicated module to generate structural priors, but the predicted results are imprecise and introduce substantial additional overhead. This raises the question: *How can we efficiently and accurately mine high-quality endogenous priors?*

**Second**, as shown in Fig. 2(b), compared to conventional sampling, wavelet sampling possesses the advantage of lossless sampling. However, as shown in Fig. 2(c), the low-frequency subbands of wavelets retain tightly coupled spatial and frequency information that closely resembles the original image (Jiang et al. 2023), still necessitating complex network architectures. In contrast, the Fourier domain can effectively handle these low-frequency components through compact structures, such as simple convolutional activations (Wang, Wu, and Zhi 2023). Therefore, the question arises: *How can we combine Fourier and wavelet transforms to leverage their respective advantages?*

Motivated by the aforementioned questions, we propose an efficient Self-Mining Prior-Guided Joint Frequency Enhancement Network (SPJFNet). First, we utilize loss-

less wavelet downsampling to reduce images to a lower-resolution space, thereby reducing computational overhead. In this lower-resolution space, we develop a Self-Mining Guidance Module (SMGM), which consists of a lightweight gamma enhancement network that mines the image’s intrinsic gradient information to generate high-quality endogenous priors. Subsequently, we introduce a Dual Frequency Guidance Framework (DFGF) that strategically exploits the complementary strengths of wavelet and Fourier domains through a dual-branch architecture. The key innovation lies in our frequency-adaptive processing strategy: wavelet low-frequency components are transformed into the Fourier domain where compact structures can efficiently handle global illumination recovery through precise amplitude and phase reconstruction guided by self-mined priors. Meanwhile, wavelet high-frequency components remain in their native domain, where they are adaptively enhanced through a gradient-prior-guided gating mechanism that preserves fine-grained textural details. This dual-domain approach maximizes the inherent advantages of each transform while mitigating their respective limitations, achieving superior reconstruction efficiency and quality.

Overall, our contributions are summarized as follows:

- We introduce SMGM, which extracts endogenous priors directly from dark images, eliminating external dependencies. Our DFGF enables frequency-adaptive enhancement through prior guidance, ensuring robust reconstruction without added computational burden.
- We propose SPJFNet, a novel dual-domain architecture that leverages the strengths of wavelet and Fourier transforms: efficiently modeling low-frequency components in the Fourier domain while preserving high-frequency details in the wavelet domain, thus achieving superior restoration with reduced complexity.
- Extensive experiments validate our state-of-the-art performance across multiple benchmarks, demonstrating significantly reduced model complexity, as shown in Fig. 1. Our plug-and-play framework seamlessly integrates into existing frequency-domain methods, delivering both performance gains and computational efficiency.

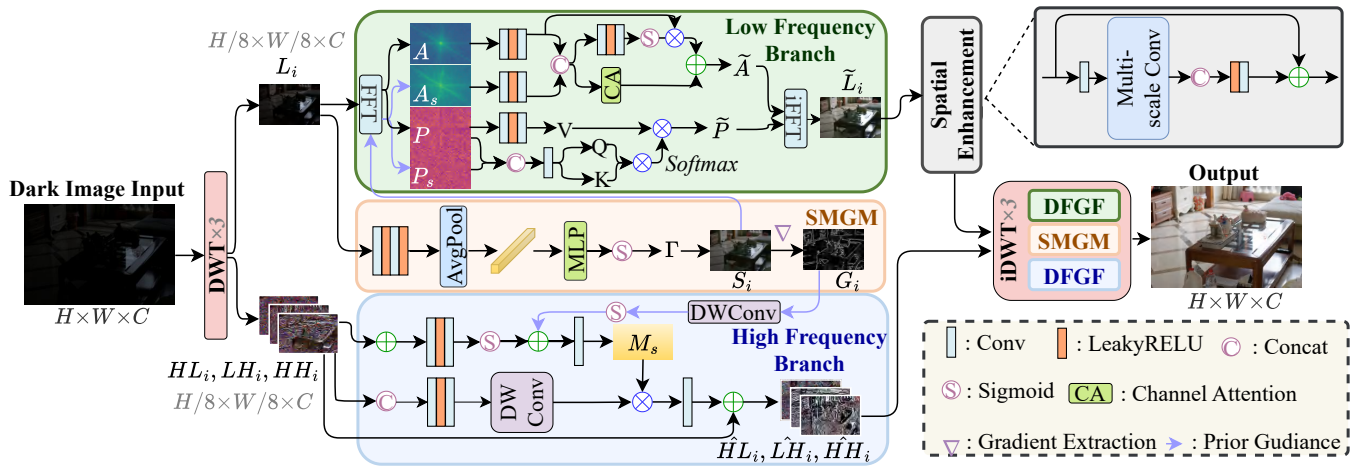


Figure 3: Overall framework of SPJFNet. SPJFNet employs 3-level DWT decomposition and consists of three main modules: SMGM, DFGF low-frequency branch, and DFGF high-frequency branch.

## Related Work

**Dark Image Restoration.** Dark image restoration methods including histogram equalization (Pizer 1990) and gamma correction (Rahman et al. 2016). While traditional methods are computationally efficient, they struggle with limited dynamic range modeling and significant noise amplification under extreme conditions. Recent deep learning methods (Cai et al. 2023; Wu et al. 2023), particularly Mamba-based techniques (Bai, Yin, and He 2024; Weng et al. 2025; Guo et al. 2025), have shown promising results in Low Light Image Enhancement (LLIE). A key limitation in these approaches is the reliance on prior guidance strategies. For example, SMG-LLE (Xu, Wang, and Lu 2023) uses modules to predict structural priors, while SKF (Wu et al. 2023) and DMFourLLIE (Zhang et al. 2024) depend on pre-trained models for semantic or infrared modal priors. However, these methods face significant challenges due to imprecise predicted priors and substantial computational overhead.

**Frequency-Domain Dark Image Restoration.** Recently, frequency-domain methods (Cai et al. 2025; Zhang et al. 2025b) have become prominent for their ability to separate enhancement from noise suppression. FECNet (Huang et al. 2022) and UHDFour (Li et al. 2023) utilize amplitude-phase decomposition for brightness and structural enhancement, while DMFourLLIE (Zhang et al. 2024) employs pre-trained infrared priors for phase correction. Additionally, wavelet-based methods like Wave-Mamba (Zou et al. 2024a) and DiffLL (Jiang et al. 2023) leverage lossless downsampling and noise-content separation for texture preservation, but they require complex modules to effectively manage intricate wavelet components. A fundamental limitation of existing frequency-domain approaches is their single-domain processing: Fourier methods excel in global illumination modeling but struggle with fine details, while wavelet methods maintain local textures but necessitate complex architectures for low-frequency enhancement.

## Method

### Overall Framework

The overall framework of SPJFNet is illustrated in Fig.3. Given an input image  $I \in \mathbb{R}^{H \times W \times C}$ , we use the 2D Discrete Wavelet Transform (2D-DWT) to decompose it into hierarchical frequency sub-bands:

$$\{L_i, HL_i, LH_i, HH_i\} = \text{2D-DWT}^i(I), \quad (1)$$

where  $L_i, HL_i, LH_i, HH_i \in \mathbb{R}^{\frac{H}{2^i} \times \frac{W}{2^i} \times C}$  represent the low-frequency and high-frequency components at the  $i$ -th decomposition level. 2D-DWT enables efficient processing in a lower-dimensional space while preserving comprehensive frequency information. First, the Self-Mining Guidance Module (SMGM) processes the low-frequency component through a lightweight gamma enhancement network to extract endogenous structure prior  $S_i$  and gradient priors  $G_i$ , serving as guidance for both frequency branches. In the Dual Frequency Guidance Framework (DFGF), the low-frequency component  $L_i$  is transformed into the Fourier domain for global illumination recovery using amplitude and phase enhancement guided by structure prior  $S_i$ . The enhanced components are spatially refined after inverse Fourier transform. Meanwhile, the high-frequency components  $\{HL_i, LH_i, HH_i\}$  are processed in their native wavelet domain and enhanced through gradient-prior-guided gating mechanisms to preserve fine details. The enhanced frequency components are progressively reconstructed using multi-scale inverse wavelet transforms, benefiting from dual-domain guidance to yield a restored output image with superior quality and computational efficiency.

### Self-Mining Guidance Module (SMGM)

As illustrated in Fig.2(a), our self-mining prediction approach demonstrates that gradient maps predicted through Gamma transformation of dark images exhibit clear structural details. This endogenous self-mining strategy is both precise and efficient, avoiding the error propagation and

Methods	Venue	LOL-v1			LSRW-Huawei			LSRW-Nikon			#Param (M)	#FLOPs (G)
		PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$		
NPE (Wang et al. 2013)	TIP'13	16.97	0.5928	-	17.08	0.3905	0.2303	14.86	0.3738	0.1464	-	-
LIME (Guo, Li, and Ling 2016)	TIP'16	16.75	0.4440	0.2060	17.00	0.3816	0.2069	13.53	0.3321	0.1272	-	-
SRIE (Fu et al. 2016)	ICCV'16	11.80	0.5000	0.1862	13.42	0.4282	0.2166	13.26	0.3963	0.1396	-	-
Kind (Zhang, Zhang, and Guo 2019)	MM'19	20.87	0.7995	0.2071	16.58	0.5690	0.2259	11.52	0.3827	0.1860	8.02	34.99
MIRNet (Zamir et al. 2020)	ECCV'20	<u>24.14</u>	0.8305	0.2502	19.98	0.6085	0.2154	17.10	0.5022	0.2170	31.79	785.1
Kind++ (Zhang et al. 2021)	IJCV'21	18.97	0.8042	0.1756	15.43	0.5695	0.2366	14.79	0.4749	<u>0.2111</u>	8.27	-
SNR-Aware (Xu et al. 2022)	CVPR'22	23.93	0.8460	0.0813	20.67	0.5911	0.1923	17.54	0.4822	0.0982	39.12	26.35
FourLLIE (Wang, Wu, and Zhi 2023)	MM'23	20.99	0.8071	0.0952	21.11	0.6256	0.1825	17.82	0.5036	0.2150	<b>0.12</b>	4.07
UHDFour (Li et al. 2023)	ICLR'23	22.89	0.8147	0.0934	19.39	0.6006	0.2466	<u>17.94</u>	0.5195	0.2546	17.54	4.78
Retinformer (Cai et al. 2023)	ICCV'23	22.71	0.8177	0.0922	21.23	0.6309	0.1699	17.64	0.5082	0.2784	1.61	15.57
Wave-Mamba (Zou et al. 2024a)	MM'24	22.76	0.8419	0.0791	21.19	0.6391	0.1818	17.34	0.5192	0.2294	1.26	7.22
UHDFormer (Wang et al. 2024)	AAAI'24	22.88	0.837	0.1390	20.64	0.6244	0.1812	17.01	0.5186	0.2793	0.34	<u>3.24</u>
DMFourLLIE (Zhang et al. 2024)	MM'24	22.98	0.8273	0.0792	21.47	0.6331	0.1781	17.04	<u>0.5274</u>	0.2178	0.75	5.81
Retinexmamba (Bai, Yin, and He 2024)	ICONIP'24	23.15	0.8210	0.0876	20.88	0.6298	0.1689	17.59	0.5133	0.2534	3.59	34.76
CWNet (Zhang et al. 2025a)	ICCV'25	<u>23.60</u>	0.8496	<b>0.0648</b>	<u>21.50</u>	<u>0.6397</u>	<u>0.1562</u>	17.38	0.5119	0.2575	1.23	11.3
CIDNet (Yan et al. 2025)	CVPR'25	23.81	<b>0.8574</b>	0.0856	20.30	0.6054	0.3500	17.16	0.4975	0.2816	1.88	7.57
SPJFNet	Ours	<b>24.16</b>	<u>0.8531</u>	<u>0.0694</u>	<b>21.71</b>	<b>0.6547</b>	<b>0.1558</b>	<b>17.98</b>	<b>0.5289</b>	<b>0.2047</b>	<u>0.21</u>	<b>2.43</b>

Table 1: Quantitative comparison on LOL-v1, LSRW-Huawei and LSRW-Nikon datasets without using ground truth mean. The best results are highlighted in **bold** and the second-best results are underlined.

computational burden associated with handcrafted or complex prediction modules. Specifically, SMGM operates on the low-frequency component  $L_i$  to extract dual complementary priors: the enhanced structural prior  $S_i$  and gradient prior  $G_i$ , which serve as endogenous guidance for the dual-frequency enhancement framework.  $L_i$  is first processed through convolutional layers to extract feature representations, followed by global average pooling to obtain the global feature vector  $\mathbf{F}_a$ , providing channel-wise global representation.  $\mathbf{F}_a$  is then fed into a multi-layer perceptron (MLP) followed by a Sigmoid activation function to produce the adaptive Gamma mapping  $\Gamma$ :

$$\Gamma = \sigma(\text{MLP}(\mathbf{F}_a)), \quad \Gamma \in \mathbb{R}^{1 \times 1 \times C}, \quad (2)$$

where  $\sigma$  denotes the Sigmoid activation function that constrains output values to  $(0, 1)$ . The Gamma mapping is applied to generate the enhanced structural prior  $S_i$ , from which the gradient prior  $G_i$  is extracted using the Sobel operator (Kanopoulos, Vasanthavada, and Baker 1988):

$$S_i = L_i^\Gamma, \quad G_i = \nabla(S_i). \quad (3)$$

The key insight lies in the dual-guidance mechanism:  $S_i$  serves as the illumination-enhanced prior that guides the DFGF wavelet low-frequency branch for global brightness recovery, while  $G_i$  provides structural gradient information that guides the DFGF wavelet high-frequency branch for detail preservation. This complementary design ensures that when the gradient prior  $G_i$  achieves high accuracy through our supervised learning objective, the corresponding Gamma-enhanced structural prior  $S_i$  inherently possesses reliable illumination guidance capabilities, as both priors originate from the same adaptive Gamma transformation. To ensure the accuracy of  $G_i$ , the learning objective

of SMGM incorporates two complementary losses: gradient fidelity loss and structural consistency loss:

$$\mathcal{L}_s = \lambda_1 \cdot \|G_i - G_i^{gt}\|_1 + \lambda_2 \cdot \text{CE}(\text{Edge}(S_i), \text{Edge}(S_i^{gt})), \quad (4)$$

where  $\lambda_1 = 1.0$  and  $\lambda_2 = 0.1$  are weighting coefficients, CE represents cross-entropy loss for edge maps, and  $\text{Edge}(\cdot)$  extracts edge information.  $\mathcal{L}_s$  ensures that both the gradient prior  $G_i$  and structural prior  $S_i$  maintain high fidelity to ground truth, establishing a robust foundation for the subsequent dual-frequency enhancement process.

## Dual-Frequency Guidance Framework (DFGF)

**Wavelet Low Frequency Branch.** While wavelet decomposition provides efficient multi-scale downsampling, its low-frequency subband inherently retains coupled spatial-frequency information, including residual high-frequency components that impede pure frequency analysis. To leverage both the structural preservation of lossless, we strategically transform wavelet low-frequency components into the Fourier domain.

As illustrated in Fig.3, the low-frequency component  $L_i$  is transformed into the Fourier domain using Fast Fourier Transform (FFT), yielding amplitude  $A$  and phase  $P$ :  $A, P = \text{FFT}(L_i)$ . Simultaneously, the structural prior  $S_i$  encoding precise brightness and structural information is transformed into amplitude prior  $A_s$  and phase prior  $P_s$ :  $A_s, P_s = \text{FFT}(S_i)$ .

Since prior work (Wang, Wu, and Zhi 2023) indicates that brightness information is mainly encoded in the amplitude component, we propose a dual-branch architecture to leverage this. We first concatenate the amplitude prior  $A_s$  with the original amplitude  $A$  to form enriched features,  $A_{concat} = \text{Concat}(A, A_s)$ , which are then processed through complementary branches. The first branch modulates the original

amplitude component using channel attention mechanisms to highlight salient features, while the second branch employs gating mechanisms for precise mapping of the amplitude components. This dual-path processing synergistically enhances global illumination recovery. This process can be formulated as:

$$B_1 = CA(A_{concat}), \quad B_2 = \sigma(\text{Conv}(A_{concat})) \cdot A, \quad (5)$$

where  $CA(\cdot)$  denotes channel attention, and  $\sigma$  is the Sigmoid activation function.  $B_1$  applies channel attention to the concatenated features and modulates the original amplitude, enabling adaptive amplitude adjustment based on both original and prior information.  $B_2$  processes the amplitude prior through convolution and Sigmoid activation to extract refined prior features with values constrained to  $(0, 1)$  for stable enhancement. The enhanced amplitude is obtained through element-wise addition:  $\tilde{A} = B_1 + B_2$ , where the addition operation preserves the original amplitude characteristics while incorporating prior-guided enhancements.

Phase components contain crucial high-frequency structural details that require careful preservation during enhancement. Unlike amplitude components that primarily encode global illumination, phase components capture local structural information that benefits from attention-based refinement. The phase prior  $P_s$  provides structural guidance for enhancing the original phase  $P$  through a self-attention framework:

$$Q, K = \text{Conv}(\text{Concat}(P_s, P)), \quad V = \text{Conv}(P), \quad (6)$$

$$\tilde{P} = \text{Softmax}(Q \cdot K^\top) \cdot V,$$

where the query  $Q$  and key  $K$  are generated from the concatenated phase and phase prior to capture correlations between structural patterns, while the value  $V$  is derived from the original phase to preserve authentic structural information. The self-attention mechanism selectively enhances phase components based on structural similarity, ensuring coherent phase reconstruction. The enhanced amplitude  $\tilde{A}$  and phase  $\tilde{P}$  are then merged and transformed back to the spatial domain via inverse FFT.

Following the two-stage processing paradigm of current Fourier-based methods, we employ a Spatial Enhancement Module to further refine local spatial details. The core component consists of two complementary convolution modules with different receptive field characteristics: dilated convolution (Yu and Koltun 2015) with dilation rate 4 for expanded receptive field coverage, and wavelet transform convolution (WTConv) (Finder et al. 2025) with 3 decomposition levels for multi-scale feature representation. The dilated convolution captures long-range spatial dependencies through enlarged receptive fields, while the WTConv leverages wavelet decomposition to extract multi-resolution features across different frequency bands. This dual-branch design achieves comprehensive spatial detail enhancement while maintaining computational efficiency, effectively recovering fine-grained spatial information.

**Wavelet High Frequency Branch.** High-frequency components primarily encode image edges, textures, and fine-grained details that require precise local enhancement to preserve structural integrity while avoiding noise amplification.

Methods	DICM	LIME	MEF	NPE	VV
Kind (Zhang, Zhang, and Guo 2019)	<b>3.61</b>	4.77	4.82	4.18	3.84
MIRNet (Zamir et al. 2020)	4.04	6.45	5.50	5.24	4.74
SGM (Yang et al. 2021)	4.73	5.45	5.75	5.21	4.88
FECNet (Huang et al. 2022)	4.14	6.04	4.71	4.50	3.75
HDMNet (Liang et al. 2022)	4.77	6.40	5.99	5.11	4.46
Bread (Guo and Hu 2023)	4.18	4.72	5.37	4.16	<b>3.30</b>
Retinexformer (Cai et al. 2023)	4.01	<u>3.44</u>	3.73	3.89	3.71
UHDFormer (Wang et al. 2024)	4.42	4.35	4.74	4.40	4.28
Wave-Mamba (Zou et al. 2024a)	4.56	4.45	4.76	4.54	4.71
CWNet (Zhang et al. 2025a)	3.92	3.58	<u>3.66</u>	<u>3.61</u>	3.74
Ours	<u>3.85</u>	<b>3.38</b>	<b>3.69</b>	<b>3.52</b>	<u>3.62</u>

Table 2: NIQE scores on DICM, LIME, MEF, NPE, and VV datasets. All methods are pre-trained on LSRW-Huawei.

Our gradient-guided enhancement strategy leverages the extracted gradient prior  $G_i$  to provide structural guidance for adaptive high-frequency enhancement.

As illustrated in Fig.3, we first fuse the gradient prior  $G_i$  with the high-frequency components through element-wise addition to create structurally-informed features:  $\mathbf{F}_{hf} = (HL_i + LH_i + HH_i) + G_i$ . Simultaneously,  $G_i$  is processed through depthwise separable convolution to extract efficient spatial features:  $\mathbf{F}_{dw} = \text{DWConv}(G_i)$ .

To generate precise enhancement weights, the extracted gradient features are transformed into a spatial gating map through  $1 \times 1$  convolution and Sigmoid activation:

$$M_s = \sigma(\text{Conv}_{1 \times 1}(\mathbf{F}_{dw})), \quad (7)$$

where  $M_s \in (0, 1)$  serves as a gradient-guided gating map that indicates enhancement confidence at each spatial location. Regions with strong gradient responses receive higher enhancement weights, while smooth regions are preserved with minimal modification.

The high-frequency components  $\{HL_i, LH_i, HH_i\}$  are concatenated and processed through depthwise separable convolution to extract spatial features:  $\mathbf{F}_{spatial} = \text{DWConv}(\text{Concat}(HL_i, LH_i, HH_i))$ . The gradient-guided gating mechanism is applied for selective enhancement:  $\mathbf{F}_{enhanced} = \mathbf{F}_{spatial} \odot M_s + \mathbf{F}_{spatial}$ , where the residual connection preserves original information while enabling adaptive enhancement. Finally, the enhanced features are processed through convolution and residual blocks, then decomposed back into refined high-frequency components  $\{\hat{H}L_i, \hat{L}H_i, \hat{H}H_i\}$  through learned projection layers, ensuring structurally-consistent detail enhancement for subsequent wavelet reconstruction.

## Experiments

### Datasets and Experimental Details

**Datasets.** We train and evaluate our method on four widely recognized dark image datasets: LOL (Yang et al. 2021), LSRW-Huawei (Hai et al. 2023) and LSRW-Nikon (Hai et al. 2023). The LOL dataset includes 485 training pairs and 15 testing pairs. LSRW-Huawei contains 3,150 training pairs and 20 testing pairs, while LSRW-Nikon pro-

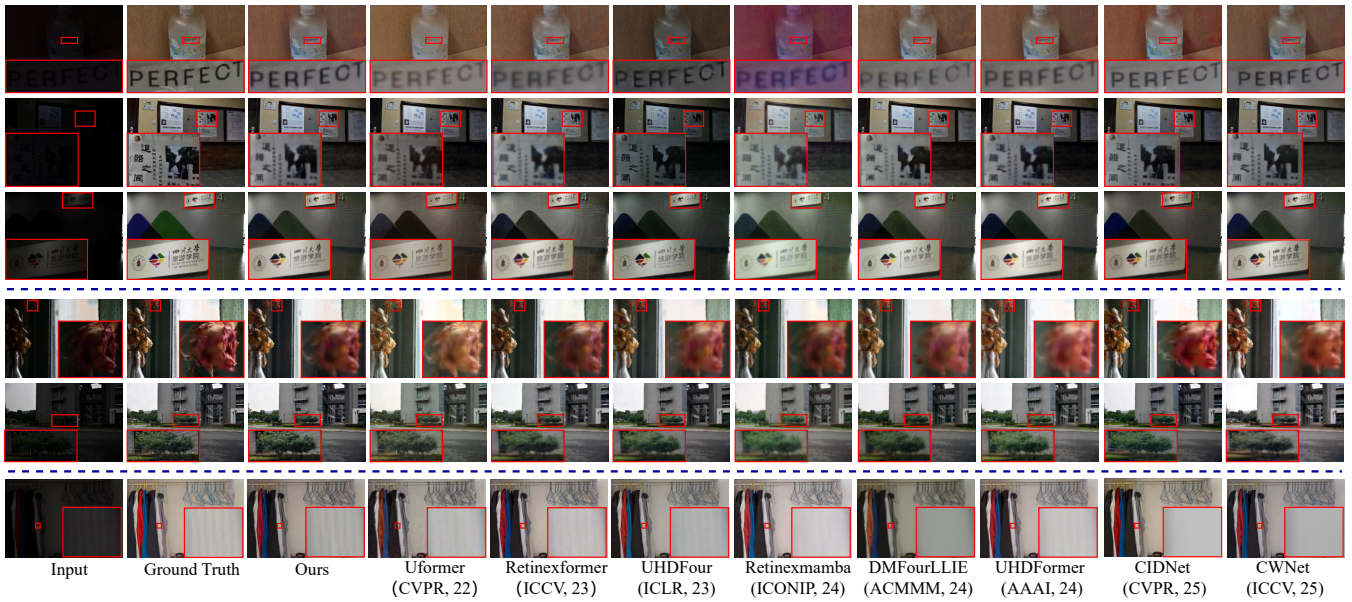


Figure 4: Visual comparison with state-of-the-art methods across three datasets. From top to bottom: three groups of comparisons are conducted on LSRW-Huawei, LSRW-Nikon, and LOL-v1 datasets, respectively.



Figure 5: Visual comparison on UHD-LL dataset. (UHD) is a designed for UHD images.

vides 2,450 training pairs and 30 testing pairs. Additionally, we evaluate on three unpaired datasets: NPE (Wang et al. 2013), DICM (Lee, Lee, and Kim 2012), LIME (Guo, Li, and Ling 2016), MEF (Ma, Zeng, and Wang 2015) and VV (Vonikakis, Kouskouridas, and Gasteratos 2018).

**Implementation Details.** Our method is implemented in PyTorch and trained end-to-end. Images are randomly cropped to  $256 \times 256$  and augmented with random flipping. We use the ADAM optimizer with an initial learning rate of  $4.0 \times 10^{-4}$  and a multi-step scheduler. Training is conducted for  $1.5 \times 10^5$  iterations with a batch size of 8 on two NVIDIA 4090 GPUs. Loss function utilizes  $\mathcal{L}_1$  loss for training.

**Evaluation.** Our network performance is primarily evaluated using peak signal-to-noise ratio (PSNR (Hore and Ziou 2010)), structural similarity index (SSIM (Wang et al. 2004)) and LPIPS (Zhang et al. 2018). We conduct comprehensive comparisons with 16 state-of-the-art dark image restoration methods, including traditional approaches (NPE (Wang et al. 2013), LIME (Guo, Li, and Ling 2016), SRIE (Fu et al.

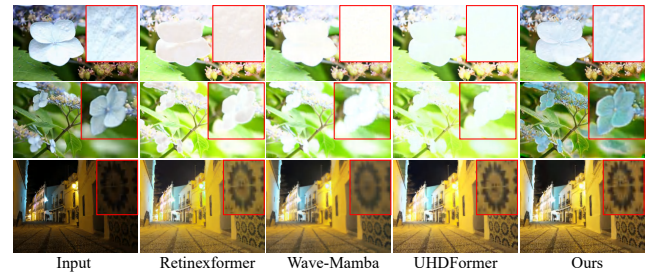


Figure 6: Visual comparison on no-reference datasets.

2016)), deep learning-based methods (Kind (Zhang, Zhang, and Guo 2019), MIRNet (Zamir et al. 2020), Kind++ (Zhang et al. 2021), SNR-Aware (Xu et al. 2022), FourLLIE (Wang, Wu, and Zhi 2023), UHDFour (Li et al. 2023), Retinexformer (Cai et al. 2023), Wave-Mamba (Zou et al. 2024a), UHDFormer (Wang et al. 2024), DMFourLLIE (Zhang et al. 2024), Retinexmamba (Bai, Yin, and He 2024), CWNet (Zhang et al. 2025a) and CIDNet (Yan et al. 2025)).

### Comparisons with State-of-the-Art Methods

The quantitative results are presented in Tab. 1. On the LOL-V1 dataset, our method achieves the best PSNR while maintaining competitive SSIM (0.8531) and LPIPS (0.0694) scores. For the LSRW-Huawei dataset, our method obtains the highest PSNR (21.71 dB) and SSIM (0.6547), with the best LPIPS score (0.1558). Similarly, on the LSRW-Nikon dataset, our method achieves the best performance across all metrics. **Notably**, our method excels in computational efficiency while maintaining superior performance. With only 0.21M parameters and 2.43G FLOPs, our approach is significantly more lightweight compared to competing methods.

Method	DWT <sub>i</sub>	$\mathcal{L}_s$	SMGM	D-L	D-H	PSNR $\uparrow$	SSIM $\uparrow$
A	2	✓	✓	✓	✓	21.13	0.64
B	4	✓	✓	✓	✓	<b>21.73</b>	<b>0.6535</b>
C	3	×	✓	✓	✓	<i>Nan</i>	<i>Nan</i>
D	3	✓	×	✓	✓	20.98	0.6384
E	3	✓	✓	×	✓	20.84	0.6371
F	3	✓	✓	✓	×	21.30	0.6523
G	3	✓	×	✓	×	20.65	0.6346
H	3	✓	✓	×	×	20.47	0.6331
I	3	✓	×	×	✓	20.82	0.6405
J	3	✓	✓	✓	✓	<u>21.71</u>	<b>0.6547</b>

Table 3: Ablation study on core architectural components. ✓ indicates the component is used, × indicates it is not used.

Amp	D-L			D-H			$\mathcal{L}_s$ $\lambda_1, \lambda_2$	Metrics	
	<i>Pha</i>	<i>Spa</i>	<i>M<sub>s</sub></i>	<i>F<sub>hf</sub></i>	<i>F<sub>s</sub></i>	PSNR $\uparrow$		SSIM $\uparrow$	
×	✓	✓	✓	✓	✓	✓	1.0, 0.1	21.23	0.6497
✓	×	✓	✓	✓	✓	✓	1.0, 0.1	21.58	0.6532
✓	✓	×	✓	✓	✓	✓	1.0, 0.1	21.28	0.6488
✓	✓	✓	×	✓	✓	✓	1.0, 0.1	21.42	0.6527
✓	✓	✓	✓	×	×	✓	1.0, 0.1	21.54	0.6543
✓	✓	✓	✓	✓	✓	×	1.0, 0.1	21.42	0.6527
✓	✓	✓	✓	✓	✓	✓	1.0, 0.1	21.30	0.6495
✓	✓	✓	✓	✓	✓	✓	1.0, 0.1	21.41	0.6502
✓	✓	✓	✓	✓	✓	✓	0.8, 0.1	21.69	0.6539
✓	✓	✓	✓	✓	✓	✓	1.0, 0.2	21.64	0.6528
✓	✓	✓	✓	✓	✓	✓	1.0, 0.1	<b>21.71</b>	<b>0.6547</b>

Table 4: Ablation study on fine-grained components.

**Quantitative Comparison on No-reference Datasets.** As shown in Tab.2, our method obtains the best NIQE scores on three out of five datasets (LIME, MEF, NPE) and ranks second on the remaining two datasets (DICM, VV), consistently outperforming existing state-of-the-art methods.

**Visual Comparisons.** To demonstrate the effectiveness of our approach, we present visual comparisons with state-of-the-art methods across three challenging datasets in Fig. 4. The results clearly illustrate the superior enhancement quality achieved by our method in various challenging scenarios. Fig. 5 shows the visual comparisons on the UHD-LL (Li et al. 2023) dataset. Compared to other methods, our approach produces results that are closer to the ground truth in terms of brightness, image quality, and detail preservation. Fig. 6 shows visual comparisons on the unpaired DICM and LIME datasets, which demonstrates our method’s superior performance in detail preservation and exposure control.

## Ablation Study

**Core Component Analysis.** Tab. 3 presents the ablation study on core architectural components. We investigate the optimal wavelet decomposition levels (rows A-B) and validate the necessity of loss  $\mathcal{L}_s$  (row C), whose removal causes training instability (NaN values). Individual module analy-

Methods	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
FourLLIE (Baseline)	21.11	0.6256	0.1825
FourLLIE + SMGM	<u>21.37</u>	<u>0.6401</u>	<u>0.1688</u>
FourLLIE + SMGM + DFGF	<b>21.49</b>	<b>0.6448</b>	<b>0.1632</b>
DMFourLLIE (Baseline)	21.47	0.6331	0.1781
DMFourLLIE + SMGM	<u>21.52</u>	<u>0.6429</u>	<u>0.1597</u>
DMFourLLIE + SMGM + DFGF	<b>21.70</b>	<b>0.6457</b>	<b>0.1601</b>

Table 5: Plug-and-play validation on Fourier-based methods.

sis (rows D-I) shows SMGM contributes 0.73 dB PSNR improvement, while DFGF low-frequency and high-frequency (D-L, D-H) branches provide 0.87 dB and 0.41 dB improvements respectively. The combination analysis (rows G-I) confirms their synergistic effect.

**Fine-grained Component Analysis.** Tab. 4 analyzes sub-components within each module. In the DFGF low-frequency branch (D-L), amplitude guidance contributes most significantly (0.48 dB), followed by spatial enhancement (0.43 dB) and phase guidance (0.13 dB). For the high-frequency branch (D-H), gradient map guidance ( $M_s$ ) provides 0.29 dB improvement, while feature processing components  $F_{hf}$  and  $F_s$  contribute 0.17 dB and 0.29 dB respectively. In the multi-scale module, WTConv and dilated convolution contribute 0.41 dB and 0.30 dB respectively, demonstrating their complementary roles. Additionally, we conducted an ablation study to verify the weight of  $\mathcal{L}_s$ .

**Plug-and-Play Validation.** To further demonstrate the generalizability and effectiveness of our proposed modules, we conduct plug-and-play experiments by integrating our components into existing Fourier-based low-light enhancement methods, specifically FourLLIE and DMFourLLIE. As shown in Tab. 5, both methods achieve consistent performance improvements with our modules.

## Limitation and Conclusion

**Limitation.** As illustrated in the second row of Fig. 6, our method occasionally exhibits color distortion where the restored image shows slight color shifts compared to natural appearance. This limitation stems from our frequency-domain processing strategy, which emphasizes amplitude and phase information while color consistency constraints could be strengthened. The gradient-guided enhancement, though effective for spatial detail recovery, may prioritize structural information over color accuracy in challenging scenarios, suggesting room for improvement in color fidelity preservation, particularly in complex lighting conditions.

**Conclusion.** In this paper, we present SPJFNet, an efficient self-mining prior-guided joint frequency enhancement network. By introducing the SMGM to eliminate dependence on external priors and the DFGF to decouple frequency processing, our method achieves significant efficiency improvements while surpassing state-of-the-art performance. In the future, we will focus our research efforts on how to efficiently combine frequency domain processing with ensuring color consistency in dark image restoration.

## Acknowledgements

This work was supported by Jilin Province Industrial Key Core Technology Tackling Project (20230201085GX).

## References

- Bai, J.; Yin, Y.; and He, Q. 2024. Retinexmamba: Retinex-based Mamba for Low-light Image Enhancement. *arXiv preprint arXiv:2405.03349*.
- Cai, M.; Zhang, T.; Liu, P.; and Zhou, Q. 2025. APMoE-Net: Fourier Amplitude-Phase Joint Enhancement and MoE Compensation for Low-Light Image Enhancement. *Expert Systems with Applications*, 129664.
- Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; and Zhang, Y. 2023. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *ICCV*.
- Du, Z.; Shi, M.; and Deng, J. 2024. Boosting Object Detection with Zero-Shot Day-Night Domain Adaptation. In *ICCV*, 12666–12676.
- Feijoo, D.; Benito, J. C.; Garcia, A.; and Conde, M. V. 2025. Darkir: Robust low-light image restoration. In *CVPR*, 10879–10889.
- Finder, S. E.; Amoyal, R.; Treister, E.; and Freifeld, O. 2025. Wavelet Convolutions for Large Receptive Fields. In *ECCV*, 363–380. Springer.
- Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.-P.; and Ding, X. 2016. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, 2782–2790.
- Gu, Y.; Wang, H.; Ling, P.; Wei, Z.; Chen, H.; Jin, Y.; and Chen, E. 2025. Improving Visual and Downstream Performance of Low-Light Enhancer with Vision Foundation Models Collaboration. In *CVPR*, 16071–16080.
- Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2025. Mambair: A simple baseline for image restoration with state-space model. In *ECCV*, 222–241. Springer.
- Guo, Q.; Wang, Y.; Zhang, Y.; Qi, H.; Hu, Y.; and Jiang, Y. 2026. Hyper-BTS: Brain tumor segmentation based on hypergraph guidance. *Pattern Recognition*, 169: 111926.
- Guo, X.; and Hu, Q. 2023. Low-light image enhancement via breaking down the darkness. *IJCV*, 131(1): 48–66.
- Guo, X.; Li, Y.; and Ling, H. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE TIP*, 26(2): 982–993.
- Hai, J.; Xuan, Z.; Yang, R.; Hao, Y.; Zou, F.; Lin, F.; and Han, S. 2023. R2rnet: Low-light image enhancement via real-low to real-normal network. *Journal of Visual Communication and Image Representation*, 90: 103712.
- Hore, A.; and Ziou, D. 2010. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition*, 2366–2369. IEEE.
- Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; and Xiong, Z. 2022. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*, 163–180. Springer.
- Jiang, H.; Luo, A.; Fan, H.; Han, S.; and Liu, S. 2023. Low-light image enhancement with wavelet-based diffusion models. *ACM Transactions on Graphics (TOG)*, 42(6): 1–14.
- Kanopoulos, N.; Vasanthavada, N.; and Baker, R. L. 1988. Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits*, 23(2): 358–367.
- Lee, C.; Lee, C.; and Kim, C.-S. 2012. Contrast enhancement based on layered difference representation. In *ICIP*, 965–968. IEEE.
- Li, C.; Guo, C.-L.; Zhou, M.; Liang, Z.; Zhou, S.; Feng, R.; and Loy, C. C. 2023. EmbeddingFourier for Ultra-High-Definition Low-Light Image Enhancement. In *ICLR*.
- Li, J.; Li, B.; Tu, Z.; Liu, X.; Guo, Q.; Juefei-Xu, F.; Xu, R.; and Yu, H. 2024. Light the Night: A Multi-Condition Diffusion Framework for Unpaired Low-Light Enhancement in Autonomous Driving. In *ICCV*, 15205–15215.
- Liang, Y.; Wang, B.; Ren, W.; Liu, J.; Wang, W.; and Zuo, W. 2022. Learning hierarchical dynamics with spatial adjacency for image enhancement. In *ACM MM*, 2767–2776.
- Ma, K.; Zeng, K.; and Wang, Z. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE TIP*, 24(11): 3345–3356.
- Pizer, S. M. 1990. Contrast-limited adaptive histogram equalization: Speed and effectiveness. In *Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia*, volume 337, 2.
- Rahman, S.; Rahman, M. M.; Abdullah-Al-Wadud, M.; Al-Quaderi, G. D.; and Shoyaib, M. 2016. An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016: 1–13.
- Vonikakis, V.; Kouskouridas, R.; and Gasteratos, A. 2018. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, 77(8): 9211–9231.
- Wang, C.; Pan, J.; Wang, W.; Fu, G.; Liang, S.; Wang, M.; Wu, X.-M.; and Liu, J. 2024. Correlation Matching Transformation Transformers for UHD Image Restoration. In *AAAI*, 5336–5344.
- Wang, C.; Wu, H.; and Zhi, J. 2023. FourLLIE: Boosting Low-Light Image Enhancement by Fourier Frequency Information. In *ACM MM*.
- Wang, S.; Zheng, J.; Hu, H.-M.; and Li, B. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP*, 22(9): 3538–3548.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4): 600–612.
- Weng, J.; Yan, Z.; Tai, Y.; Qian, J.; Yang, J.; and Li, J. 2025. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. *NeurIPS*, 37: 27440–27462.
- Wu, Y.; Pan, C.; Wang, G.; Yang, Y.; Wei, J.; Li, C.; and Shen, H. T. 2023. Learning Semantic-Aware Knowledge Guidance for Low-Light Image Enhancement. In *CVPR*, 1662–1671.
- Xiao, J.; Chen, Y.; Feng, X.; Wang, R.; and Wu, Z. 2025. RecNet: Optimization for Dense Object Detection in Retail Scenarios Based on View Rectification. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. SNR-aware low-light image enhancement. In *CVPR*, 17714–17724.

Xu, X.; Wang, R.; and Lu, J. 2023. Low-light image enhancement via structure modeling and guidance. In *CVPR*, 9893–9903.

Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; and Zhang, Y. 2025. Hvi: A new color space for low-light image enhancement. In *CVPR*, 5678–5687.

Yang, W.; Wang, W.; Huang, H.; Wang, S.; and Liu, J. 2021. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE TIP*, 30: 2072–2086.

Yu, F.; and Koltun, V. 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Yu, J.; Ma, Z.; Ma, Y.; Liu, K.; Wang, Y.; and Li, J. 2025. MILD: Multi-Layer Diffusion Strategy for Complex and Precise Multi-IP Aware Human Erasing. *arXiv preprint arXiv:2508.06543*.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2020. Learning enriched features for real image restoration and enhancement. In *ECCV*, 492–511. Springer.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 586–595.

Zhang, T.; Liu, P.; Lu, Y.; Cai, M.; Zhang, Z.; Zhang, Z.; and Zhou, Q. 2025a. CWNNet: Causal Wavelet Network for Low-Light Image Enhancement. *arXiv preprint arXiv:2507.10689*.

Zhang, T.; Liu, P.; Zhao, M.; and Lv, H. 2024. DM-FourLLIE: Dual-Stage and Multi-Branch Fourier Network for Low-Light Image Enhancement. In *ACM MM*, 7434–7443.

Zhang, T.; Liu, P.; Zhong, Z.; Zhang, Z.; and Zhou, Q. 2025b. Beyond Illumination: Fine-Grained Detail Preservation in Extreme Dark Image Restoration. *arXiv preprint arXiv:2508.03336*.

Zhang, Y.; Guo, X.; Ma, J.; Liu, W.; and Zhang, J. 2021. Beyond brightening low-light images. *IJCV*, 129: 1013–1037.

Zhang, Y.; Zhang, J.; and Guo, X. 2019. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, 1632–1640.

Zhao, M.; Liu, P.; Zhang, T.; and Zhang, Z. 2025. ReF-LLE: Personalized Low-Light Enhancement via Reference-Guided Deep Reinforcement Learning. *arXiv preprint arXiv:2506.22216*.

Zou, W.; Gao, H.; Yang, W.; and Liu, T. 2024a. Wave-Mamba: Wavelet State Space Model for Ultra-High-Definition Low-Light Image Enhancement. In *ACM MM*, 1534–1543.

Zou, W.; Gao, H.; Ye, T.; Chen, L.; Yang, W.; Huang, S.; Chen, H.; and Chen, S. 2024b. VQCNIR: Clearer Night Image Restoration with Vector-Quantized Codebook. In *AAAI*, 7873–7881.