Beyond Illumination: Fine-Grained Detail Preservation in Extreme Dark Image Restoration

Tongshun Zhang[1,2], **Pingping Liu**[1,2*], **Zixuan Zhong**[3], **Zijian Zhang**[1,2], **Qiuzhan Zhou**[4]

[1] College of Computer Science and Technology, Jilin University
[2] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University
[3] College of Software, Jilin University
[4] College of Communication Engineering, Jilin University
{tszhang23, zhongzx24}@mails.jlu.edu.cn,{liupp, zhangzijian, zhouqz}@jlu.edu.cn

Abstract

Recovering fine-grained details in extremely dark images remains challenging due to severe structural information loss and noise corruption. Existing enhancement methods often fail to preserve intricate details and sharp edges, limiting their effectiveness in downstream applications like text and edge detection. To address these deficiencies, we propose an efficient dual-stage approach centered on detail recovery for dark images. In the first stage, we introduce a Residual Fourier-Guided Module (RFGM) that effectively restores global illumination in the frequency domain. RFGM captures inter-stage and inter-channel dependencies through residual connections, providing robust priors for high-fidelity frequency processing while mitigating error accumulation risks from unreliable priors. The second stage employs complementary Mamba modules specifically designed for textural structure refinement: (1) Patch Mamba operates on channel-concatenated non-downsampled patches, meticulously modeling pixel-level correlations to enhance fine-grained details without resolution loss. (2) Grad Mamba explicitly focuses on high-gradient regions, alleviating state decay in state space models and prioritizing reconstruction of sharp edges and boundaries. Extensive experiments on multiple benchmark datasets and downstream applications demonstrate that our method significantly improves detail recovery performance while maintaining efficiency. Crucially, the proposed modules are lightweight and can be seamlessly integrated into existing Fourier-based frameworks with minimal computational overhead.

Code — https://github.com/bywlzts/RFGM

Introduction

Images captured in extremely dark conditions often suffer from poor visibility, leading to significant loss of structural and detailed information, which constrains the performance of fine-grained downstream applications (Xiao et al. 2025; Yu et al. 2025; Guo et al. 2026). Traditional restoration techniques, such as histogram equalization (Pizer 1990), Retinex theory (Guo, Li, and Ling 2016), and gamma correction (Rahman et al. 2016), have been explored but struggle in extreme darkness and exhibit limited generalization, leading to their decline. Recently, learning-based methods (Zou et al. 2024b; Weng et al. 2025; Feijoo et al. 2025) have improved reconstruction quality and scene generalization by learning mappings between low-light and normal-light images. However, these approaches primarily focus on global brightness mapping, failing to preserve fine details effectively. They are also hampered by limited noise control and complex architectures with large parameter counts, making it difficult to balance performance and efficiency.

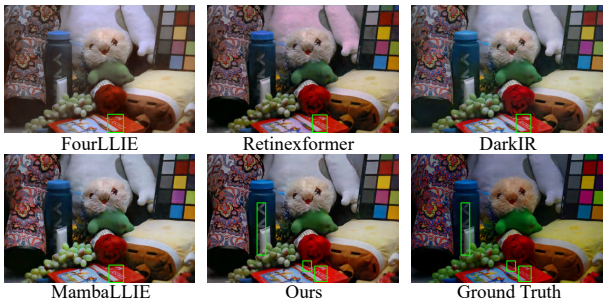

Figure 1: Text detection comparisons on LOL-v1 dataset.

To address the computational complexity and parameter efficiency challenges inherent in spatial-domain methods, frequency-domain approaches (Wang, Wu, and Zhi 2023; Li et al. 2023; Zhang et al. 2024; Feijoo et al. 2025; Cai et al. 2025; Zhang et al. 2025b; Zhao et al. 2025; Liu et al. 2025) have emerged as a promising alternative for dark image restoration. While Fourier-domain methods can achieve effective global information modeling while maintaining compact parameter specifications, these methods commonly rely on sequential simple convolutions or introduce unreliable priors, which lead to redundancy or loss of frequency-domain information. Furthermore, due to the global modeling nature of the Fourier domain, these methods employ additional encoder-decoder structures to enhance spatial detail representation. However, encoder-decoder architectures struggle to capture fine-grained structure and details, and the downsampling process inevitably leads to the loss of critical image detail information.

Concurrently, Mamba (Gu and Dao 2023) based dark im-

*Corresponding author
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

age restoration (Zou et al. 2024a; Bai, Yin, and He 2024; Weng et al. 2025) methods have attracted significant attention due to their linear complexity, with Mamba demonstrating tremendous potential in balancing global receptive fields and computational efficiency. However, these methods unfold 2D images using fixed scanning rules to generate 1D token sequences, introducing redundancy through multiple redundant scans. Moreover, tokens with strong associative properties may be spatially distant in the sequence (Guo et al. 2025b), thereby weakening inter-token modeling capabilities and limiting long-range dependency modeling.

Motivated by these critical limitations, we propose an efficient dual-stage approach specifically designed for dark image detail recovery. Our method strategically addresses the aforementioned challenges through the synergistic combination of frequency-domain global modeling and spatial-domain detail refinement. **In the first stage,** we introduce a Residual Fourier-Guided Module (RFGM). In the Fourier domain, the amplitude component represents the brightness of an image, while the phase component encodes its structural details. Amplitude recovery requires precise amplitude mapping, whereas phase components necessitate robust adaptive adjustment. Therefore, we leverage inter-stage and inter-channel correlations in a residual manner to provide robust prior guidance. We identify optimally matched amplitude components as residual Fourier channels, serving as prior guidance for amplitude mapping in subsequent stages, while phase components provide additional structural prior compensation for later stages in a residual fashion. This achieves efficient and robust recovery of global frequency-domain information. **In the second stage,** we advance beyond illumination to prioritize fine-grained detail preservation. To achieve this, we propose complementary dual-branch Mamba modules that work synergistically: Patch Mamba specializes in pixel-level fine detail enhancement, while Grad Mamba targets the reconstruction of structural textures. Patch Mamba functions on channel-concatenated, non-downsampled patches, avoiding the pitfalls of encoding-decoding sampling losses and meticulously modeling pixel-level correlations to enhance fine details without sacrificing resolution or increasing computational load. In contrast, Grad Mamba concentrates on high-gradient regions, utilizing gradient score prediction to enhance interactions among tokens that are closely associated with gradients. Inspired by MambaIRv2 (Guo et al. 2025a), we further integrate gradient prediction scores with state space models to alleviate state decay, ensuring a concentrated effort on reconstructing sharp edges and boundaries.

In Fig. 1, we validate the effectiveness of our method for text detection, showcasing its capacity to restore fine details in extremely dark images. In summary, our main contributions are as follows:

- We propose an efficient dark image restoration framework focused on fine-grained detail preservation.
- We present a Residual Fourier-Guided Module (RFGM), which utilizes inter-stage and inter-channel correlations to enhance prior guidance and mitigate issues of redundancy and error propagation.
- We overcome Mamba's inherent limitations by designing dual-branch modules: Patch Mamba for fine detail enhancement without resolution loss, and Grad Mamba for gradient-driven structural boundary reconstruction.
- Extensive experiments demonstrate significant improvements in restoration quality for dark images and downstream tasks (text detection, edge detection) while maintaining minimal computational overhead and compatibility with existing Fourier-based methods.

Related Work

Frequency-Based Dark Image Restoration Methods. Frequency-domain approaches (Wang, Wu, and Zhi 2023; Huang et al. 2022; Feijoo et al. 2025) have proven effective by distinguishing high-frequency from low-frequency information, enhancing brightness while minimizing noise. FourLLIE (Wang, Wu, and Zhi 2023) leverages Fourier transforms for efficient global feature extraction, replacing Transformer modules in SNR-Aware (Xu et al. 2022) and significantly reducing parameter counts. UHDFour (Li et al. 2023) enhances ultra-high-definition images by utilizing consistent amplitude patterns across resolutions but suffers from information loss. DMFourLLIE (Zhang et al. 2024) enhances frequency-domain information by introducing infrared priors, but fails to consider the generalization limitations of pretrained models. Wavelet-based methods, such as Wave-Mamba (Zou et al. 2024a), apply wavelet transforms in ultra-high-definition enhancement but face challenges due to complexity in low-frequency processing architectures (Jiang et al. 2023). CWNet (Zhang et al. 2025a) combines causal and wavelet methods for brightness restoration but neglects detailed modeling of image nuances.

Mamba-Based Dark Image Restoration Methods. Mamba (Gu and Dao 2023) introduced input-dependent state space models (SSMs) with selective mechanisms, applied successfully across various tasks like super-resolution (Guo et al. 2025a), classification (Xiao et al. 2024), and restoration (Li et al. 2025). In low-light enhancement, RetinexMamba (Bai, Yin, and He 2024) utilized Mamba with Retinex theory for improved efficiency. Wave-Mamba (Zou et al. 2024a) integrated Mamba with wavelet transforms, while MambaLLIE (Weng et al. 2025) introduced implicit Retinex-aware mechanisms in a state space model. However, these methods do not resolve the limitations of state space models in 2D applications, which hampers effective token modeling. Our work not only addresses the decay of state space models but also pioneers pixel-level fine-grained modeling with Mamba.

Method

Overview

The overall architecture of our proposed dual-stage approach is illustrated in Fig. 2. Given a dark image $I \in \mathbb{R}^{H\times W\times 3}$, we first apply a 3×3 convolutional layer to extract shallow feature embeddings of size $\mathbb{R}^{H\times W\times C}$, where H, W, and C denote height, width, and channel dimensions, respectively. **First Stage - Frequency-Domain**

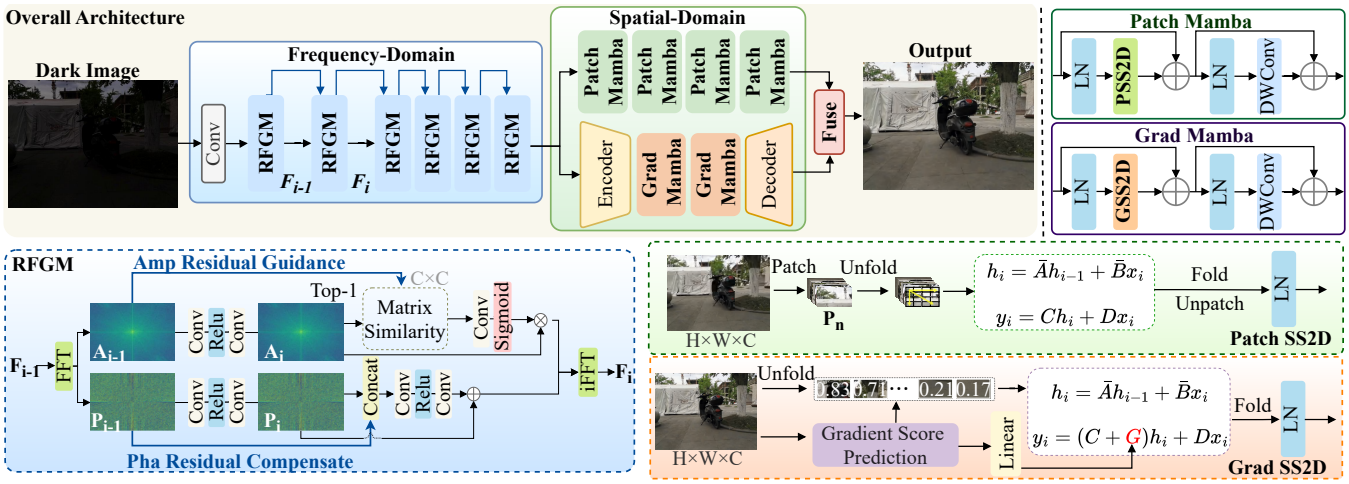

Figure 2: Overall architecture of our proposed dual-stage framework. The First Stage Frequency Domain comprises six RFGMs, while the Second Stage Spatial Domain consists of four Patch Mambas and encoder-decoder with two Grad Mambas.

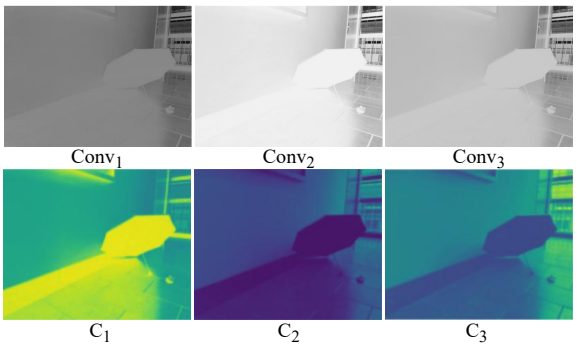

Figure 3: Visualization of DMFourLLIE across different stages and channels. Conv_i denotes stage-specific convolutional features, while C_i represents distinct feature channels.

Global Modeling: Global information is processed through six identical RFGMs, which leverage inter-stage and inter-channel correlations in a residual manner to provide robust prior guidance. This stage focuses on global illumination recovery and overall structural restoration. **Second Stage - Spatial-Domain Detail Refinement:** Decoupled from illumination adjustment, the second stage employs complementary dual-branch Mamba modules that work synergistically to preserve fine-grained textural and structural details. One branch consists of four Patch Mamba modules operating on channel-concatenated non-downsampled image patches, ensuring computational efficiency while precisely modeling pixel-level correlations. The other branch comprises an encoder-decoder and two Grad Mamba modules, specifically designed for structural texture reconstruction with emphasis on sharp edge and boundary recovery.

First Stage - Frequency-Domain Global Modeling

Motivation Analysis. As visualized in Fig. 3, current methods (Huang et al. 2022; Zhang et al. 2024) typically handle amplitude and phase components sequentially through a series of convolutional blocks ($\text{Conv}_1 \rightarrow \text{Conv}_2 \rightarrow \text{Conv}_3$). However, our feature analysis uncovers inconsistencies in feature representations across these stages. Notably, the features from Conv_2 exhibit significantly brighter activations compared to those from Conv_3. This inconsistency signals potential information degradation inherent in a strictly sequential processing paradigm, where vital cues from earlier stages (Conv_2) are insufficiently preserved or leveraged in subsequent stages (Conv_3), leading to less effective illumination adjustments. Moreover, a channel-wise examination of features (C_1, C_2, C_3) within a single stage reveals that essential structural contours manifest with varying prominence across different channels (C_1 vs. C_2 vs. C_3). This disparity emphasizes the limitation of treating channels in isolation, suggesting a failure to synthesize the complementary structural information that is distributed across the channels. Collectively, Fig. 3 elucidates two critical shortcomings: (1) the risk of progressive information loss due to sequential stage processing, and (2) the oversight of comprehensive structural synthesis resulting from isolated channel processing. These deficiencies underscore the urgent need for strategies that explicitly foster inter-stage information preservation and inter-channel information fusion.

Motivated by the observation, and to eliminate error accumulation from pre-trained model (Wu et al. 2023; Zhang et al. 2024) and manual priors (Bai, Yin, and He 2024) while reducing computational overhead from additional modules (Xu, Wang, and Lu 2023), we propose the Residual Fourier-Guided Module (RFGM). This module captures the most valuable amplitude channel priors from the previous stage, avoiding redundant processing while providing precise guidance for amplitude component mapping in subsequent stages. Meanwhile, phase components provide robust structural compensation through residual connections, enabling adaptive structural information reconstruction.

Specifically, as shown in the bottom-left of Fig. 2, features $F_{i\text{-}1} \in \mathbb{R}^{H\times W\times C}$ from the previous stage are first

transformed to the Fourier domain using Fast Fourier Transform (FFT), yielding amplitude $A_{i\text{-}1}$ and phase $P_{i\text{-}1}$ components. These components are processed separately: $A_{i\text{-}1}$ and $P_{i\text{-}1}$ undergo convolution followed by ReLU activation to obtain A_i and P_i, respectively. For amplitude components, both A_{i-1} and A_i are flattened to $\mathbb{R}^{HW\times C}$ and subjected to matrix similarity computation $MS(\cdot,\cdot)$, yielding a similarity matrix $M \in \mathbb{R}^{C\times C}$. From M, we select a Top-1 vector $V \in \mathbb{R}^{C\times 1}$ as an index corresponding to the most similar channel between A_{i-1} and A_i. The objective is to select the most reliable brightness prior from channels with varying brightness distributions in A_{i-1}. The selected prior information is then expanded through 1×1 convolution and processed with a Sigmoid activation function to generate prior guidance P_a. Subsequently, amplitude A_i is multiplied by P_a and combined with residual connections to complete the prior guidance fusion from the previous stage. This process can be formulated as:

$$\begin{aligned} M = MS(A_{i-1}, A_i), \quad V = \text{Top-1}(M), \\ P_a = \text{Sigmoid}(\text{Conv}(A_{i-1}|\text{Index}(V))), \\ \tilde{A}_i = A_i \times P_a + A_i. \end{aligned} \tag{1}$$

For phase components, P_{i-1} and P_i are concatenated along the channel dimension, where phase information from the previous stage serves as structural compensation and is adaptively fused through convolutional layers. Subsequently, through ReLU activation and convolution, the phase is restored to its original scale $\tilde{P}_i \in \mathbb{R}^{H\times W\times C}$:

$$\tilde{P}_i = \text{Conv}(\text{Concat}(P_{i-1}, P_i)) + P_i. \tag{2}$$

Finally, the processed amplitude $\tilde{A}_i$ and phase $\tilde{P}_i$ are combined through inverse Fast Fourier Transform (iFFT) to generate the output features F_i for the next stage.

Second Stage - Spatial-Domain Detail Refinement

Motivation Analysis. While frequency-domain illumination recovery effectively tackles low-frequency brightness challenges in dark images, it is essential to focus on fine-grained structural detail and sharp edge preservation, surpassing mere illumination adjustment. Therefore, we introduce complementary dual-branch Mamba aimed at fine-grained detail recovery. Mamba-based restoration methods employ discrete state space equations to model interactions between tokens:

$$h_i = \overline{\mathbf{A}} h_{i-1} + \overline{\mathbf{B}} x_i, \quad y_i = \mathbf{C} h_i + \mathbf{D} x_i, \tag{3}$$

where the i-th token depends entirely on the preceding i-1 tokens, creating direct causal relationships between neighboring pixels. However, this scanning mechanism disrupts correlations among distant features, rendering it less effective for vision tasks. Furthermore, the causal modeling may induce long-range decay effects, while multiple fixed-direction scanning strategies add unnecessary complexity and information redundancy (Guo et al. 2025a).

Based on these observations, rather than being constrained by Mamba's limitations, we reverse the approach by fully leveraging Mamba's unique characteristics and transforming them into advantages for 2D image modeling.

Patch Mamba. Image details are manifested in pixel-level variations and correlations. However, direct pixel-level operations incur substantial memory overhead. While encoder-decoder architectures can enhance spatial performance, sampling layers inevitably lose crucial pixel-level content, which is devastating for fine-grained detail preservation. Since we focus primarily on relationships between adjacent pixels, capturing global contextual dependencies is unnecessary. Mamba's robust scanning mechanism precisely satisfies pixel adjacency requirements while weakening connections between distant pixels. However, processing without downsampling creates computational burden and poor parallelization. Therefore, we propose Patch Mamba, which operates on channel-concatenated non-downsampled patches. As illustrated in the top-right of Fig. 2, Patch Mamba consists of two residual blocks: the first comprising LayerNorm and Patch SS2D, and the second comprising LayerNorm and depthwise separable convolution.

Patch SS2D performs patch-wise selective scanning. Given an input feature map $F \in \mathbb{R}^{H\times W\times C}$, it is uniformly divided into (n, m) non-overlapping patches $P_{i,j} \in \mathbb{R}^{h\times w\times C}$, where $h = \frac{H}{\sqrt{n}}$ and $w = \frac{W}{\sqrt{m}}$. These patches are then stacked along the channel dimension to form $\mathbb{R}^{h\times w\times n\cdot m\cdot C}$, significantly reducing scanning dimensions and alleviating computational burden while improving parallel efficiency. Note that the Patch SS2D scanning strategy encompasses horizontal, vertical, diagonal, and their respective reverse directions to comprehensively enhance inter-pixel correlations. After scanning all patches, results are recombined into $\mathbb{R}^{H\times W\times C}$ for subsequent processing.

Grad Mamba. Previous methods often fail to preserve edge sharpness and textural integrity, especially in areas with complex gradient distributions. To address this, we propose Grad Mamba, a gradient-guided state space model that explicitly targets high-gradient regions and prioritizes the reconstruction of sharp edges and boundaries.

As shown in Fig. 2, Grad Mamba operates within an encoder-decoder framework, complementing Patch Mamba. Its overall structure is similar to Patch Mamba. Specifically, given input features $F \in \mathbb{R}^{H\times W\times C}$, we first perform gradient score prediction using three complementary operators: Sobel operators in x and y directions to capture directional gradients, and a Laplacian operator to detect edge transitions. The extracted gradient magnitudes are then converted into priority scores through adaptive normalization and learnable scaling parameters:

$$P_{grad} = \sigma(\cdot \frac{G_{mag} - G_{min}}{G_{max} - G_{min} + \epsilon} + \beta), \tag{4}$$

where $P_{grad} \in \mathbb{R}^{H\times W}$ represents gradient priority scores, β is learnable offset parameter, σ denotes the sigmoid function, and G_{min}, G_{max} are the minimum and maximum gradient magnitudes for each sample.

Given gradient priority scores P_{grad}, we sort tokens according to their gradient importance, prioritizing high-gradient tokens for processing. This enables preferential interactions between high-gradient associated regions while avoiding long-range decay limitations. Additionally, inspired by MambaIRv2 (Guo et al. 2025a), since state C in

Methods	LOL-v1			LOL-v2-Real			LOL-v2-Syn		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Kind (Zhang, Zhang, and Guo 2019)	20.87	0.7995	0.2071	20.01	0.8412	0.0813	22.62	0.9041	0.0515
MIRNet (Zamir et al. 2020)	24.14	0.8305	0.2502	22.11	0.7942	0.1448	22.52	0.8997	0.0568
Kind++ (Zhang et al. 2021)	18.97	0.8042	0.1756	20.59	0.8294	0.0875	21.17	0.8814	0.0678
SNR-Aware (Xu et al. 2022)	23.93	0.8460	0.0813	21.48	0.8478	0.0740	24.13	0.9269	0.0318
FourLLIE (Wang, Wu, and Zhi 2023)	20.99	0.8071	0.0952	22.34	0.8403	0.0573	24.65	0.9192	0.0389
UHDFour (Li et al. 2023)	22.89	0.8147	0.0934	19.42	0.7896	0.1151	23.64	0.8998	0.0341
Retinexformer (Cai et al. 2023)	22.71	0.8177	0.0922	22.79	0.8397	0.0724	25.67	0.9295	0.0273
DMFourLLIE (Zhang et al. 2024)	22.98	0.8273	0.0792	22.71	0.8583	0.0539	25.74	0.9308	0.0251
UHDFormer (Wang et al. 2024)	22.88	0.8370	0.1390	19.71	0.8320	0.0758	24.48	0.9277	0.0306
Wave-Mamba (Zou et al. 2024a)	22.76	0.8419	0.0791	20.35	0.8379	0.1908	24.69	0.9271	0.0584
RetinexMamba (Bai, Yin, and He 2024)	23.15	0.8210	0.0876	21.73	0.8290	0.1164	25.89	0.9346	0.0389
MambaLLIE (Weng et al. 2025)	22.80	0.8315	0.0907	21.85	0.8276	0.1673	<u>25.87</u>	<u>0.9400</u>	0.0467
CWNet (Zhang et al. 2025a)	<u>23.60</u>	0.8496	<u>0.0648</u>	23.31	<u>0.8641</u>	<u>0.0532</u>	25.74	0.9365	<u>0.0241</u>
CIDNet (Yan et al. 2025)	23.81	**0.8574**	0.0856	**23.43**	0.8622	0.1691	25.70	0.9419	0.0437
URetienxNet++ (Yan et al. 2025)	23.83	0.8390	0.2310	21.97	0.8360	0.2030	24.60	0.9270	0.1020
Ours	**24.11**	<u>0.8517</u>	**0.0557**	<u>23.38</u>	**0.8662**	**0.0527**	**25.97**	**0.9408**	**0.0195**

Table 1: Quantitative comparison on LOL-v1, LOL-v2-Real, and LOL-v2-Syn (Yang et al. 2021) datasets without using ground truth mean. The best and second-best results are highlighted in **bold** and <u>underlined</u>, respectively.

Methods	LSRW-Huawei		LSRW-Nikon		Efficiency	
	PSNR	SSIM	PSNR	SSIM	#Param	#FLOPs
Kind	16.58	0.5690	11.52	0.3827	8.02	34.99
MIRNet	19.98	0.6085	17.10	0.5022	31.79	785.1
SNR-Aware	20.67	0.5911	17.54	0.4822	39.12	26.35
UHDFour	19.39	0.6006	<u>17.94</u>	0.5195	17.54	<u>4.78</u>
Retinexformer	21.23	0.6309	17.64	0.5082	1.61	15.57
Wave-Mamba	21.19	0.6391	17.34	0.5192	1.26	7.22
DMFourLLIE	21.47	0.6331	17.04	<u>0.5274</u>	<u>0.75</u>	5.81
Retinexmamba	20.88	0.6298	17.59	0.5133	3.59	34.76
CWNet	<u>21.50</u>	<u>0.6397</u>	17.38	0.5119	1.23	11.3
CIDNet	20.30	0.6054	17.16	0.4975	1.88	7.57
MambaLLIE	20.98	0.6388	17.25	0.5084	2.28	20.85
Ours	**21.59**	**0.6441**	**17.96**	**0.5342**	**0.37**	**4.39**

Table 2: Quantitative comparison on LSRW-Huawei and LSRW-Nikon datasets.

state space models resembles the query Q in attention mechanisms, it acquires attention-like capabilities to query high-gradient regions throughout the image, making the network more focused on edge information and structural detail reconstruction. Therefore, the enhanced state space formulation is expressed as:

$$h_i = \overline{\mathbf{A}} h_{i-1} + \overline{\mathbf{B}} x_i, \quad y_i = (\mathbf{C} + \mathbf{G}) h_i + \mathbf{D} x_i, \tag{5}$$

where $\mathbf{G} \in \mathbb{R}^{N \times D}$ represents the gradient guidance matrix derived from P_{grad}, and is computed as:

$$\mathbf{G} = \text{Linear}(P_{grad}) \cdot \mathbf{W}_{\mathbf{G}}, \tag{6}$$

where $\mathbf{W}_{\mathbf{G}}$ is a learnable projection matrix that transforms gradient scores into the same dimensional space as the output projection matrix $\mathbf{C}$. In high-gradient regions, the enhanced projection $(\mathbf{C}+\mathbf{G})$ amplifies the hidden state contributions, effectively strengthening the model's focus on edge and structural information. Through this mechanism, Grad Mamba achieves adaptive enhancement of structural details while maintaining the computational efficiency and sequential modeling capabilities inherent to state space models.

Methods	**PSNR ↑**	**SSIM ↑**	**LPIPS ↓**
w/o First Stage	20.73	0.6331	0.1962
w/o Second Stage	20.95	0.6387	0.1827
w/o RFGM	21.24	0.6374	0.1696
w/o Patch Mamba	21.54	0.6427	0.1612
w/o Grad Mamba	21.21	0.6392	0.1831
Full Model	**21.59**	**0.6441**	**0.1607**

Table 3: Ablation study on each component.

Experiments

Datasets and Experimental Setting

Datasets. We conduct comprehensive experiments on five benchmark datasets for extreme dark image restoration: LOL-v1 (Yang et al. 2021) (485 training/15 testing pairs), LOL-v2-Real (689 training/100 testing pairs), LOL-v2-Syn (900 training/100 testing pairs), LSRW-Huawei (Hai et al. 2023) (3,150 training/20 testing pairs) and LSRW-Nikon (2,450 training/30 testing pairs). These datasets feature severely underexposed images captured in extremely low-light conditions, presenting significant challenges for meaningful visual content recovery.

Implementation Details. Our method is implemented end-to-end on the PyTorch platform. Images are randomly cropped to 256×256 resolution and augmented with random horizontal/vertical flips and rotations. We use the ADAM optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$, initialized at a learning rate of 4.0×10^{-4}. Learning rate scheduling follows a MultiStepLR strategy with decay steps at 5×10^4

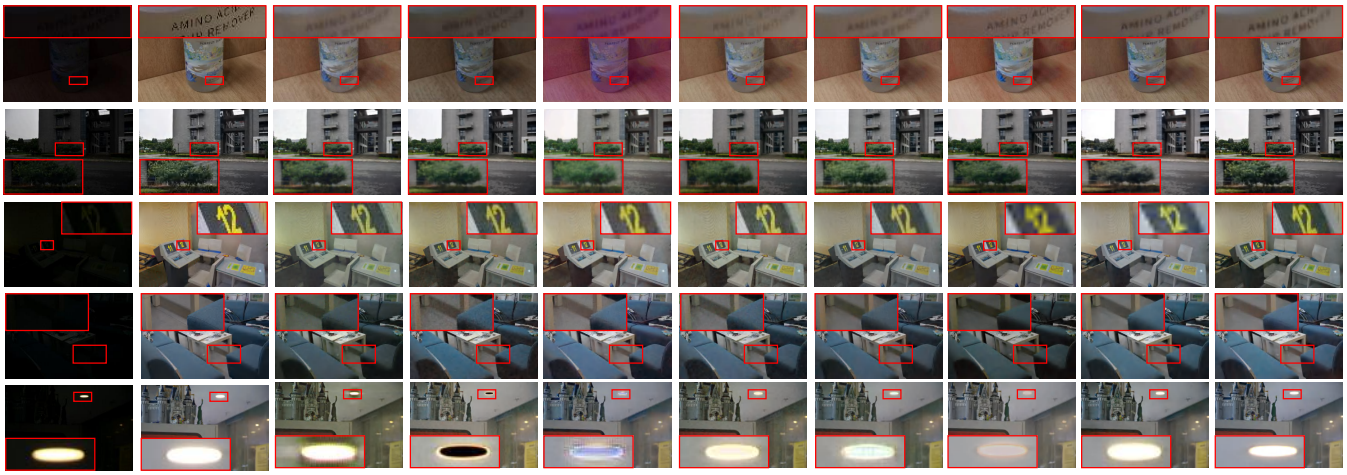

Figure 4: Visual comparisons with state-of-the-art methods. From top to bottom: LSRW-Huawei (row 1), LSRW-Nikon (row 2), LOL-v1 (row 3), and LOL-v2-Real (rows 4-5) datasets.

Methods	PSNR	SSIM	LPIPS	Param
FourLLIE (Baseline)	21.11	0.6256	0.1825	0.12
FourLLIE + RFGM	21.39	0.6425	0.1683	0.126
FourLLIE + RFGM + PatchSS2D	**21.47**	**0.6433**	**0.1622**	0.14
DMFourLLIE (Baseline)	21.47	0.6331	0.1781	0.75
DMFourLLIE + RFGM	21.50	0.6408	0.1593	0.756
DMFourLLIE + RFGM + PatchSS2D	**21.52**	**0.6439**	**0.1492**	0.77

Table 4: Plug-and-play validation on Fourier-based methods.

Method	UHDFour	FourLLIE	WaveMamba	MambaLLIE	Ours
CRAFT	0.4103	0.4211	0.4390	0.3684	**0.4500**
PAN	0.1760	0.1290	0.1760	0.1880	**0.2350**

Table 5: H-Mean comparison of text detection methods.

and 1×10^5 iterations, applying a decay factor of 0.5. All experiments are conducted on dual NVIDIA RTX 4090 GPUs (24GB) and an Intel Core i9-14900K processor, with training performed using a batch size of 8 for 2×10^5 iterations. The experiments utilize $\mathcal{L}_1$ loss for training.

Comparative Methods and Evaluation Metrics. We compare our method against various SOTA approaches, including deep learning methods (Kind (Zhang, Zhang, and Guo 2019), Kind++ (Zhang et al. 2021), MIRNet (Zamir et al. 2020), SNR-Aware (Xu et al. 2022), CWNet (Zhang et al. 2025a), CIDNet (Yan et al. 2025), URetienxNet++ (Yan et al. 2025)), Fourier-based methods (FourLLIE (Wang, Wu, and Zhi 2023), UHDFour (Li et al. 2023), DMFourLLIE (Zhang et al. 2024)), Transformer-based methods (Retinexformer (Cai et al. 2023), UHDFormer (Wang et al. 2024)), and Mamba-based methods (RetinexMamba (Bai, Yin, and He 2024), Wave-Mamba (Zou et al. 2024a), MambaLLIE (Weng et al. 2025)). All methods are evaluated using the same testing protocols for fairness, employing Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) (Wang et al. 2004), and Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018) as full-reference metrics.

Quantitative and Qualitative Results

Comparison on LOL-v1, LOL-v2-Real, and LOL-v2-Syn Datasets. Quantitative results are in Tab. 1. On the LOL-v1 dataset, our method achieves the highest PSNR and LPIPS while maintaining competitive SSIM. For the LOL-v2-Real dataset, we attain the second-best PSNR, the highest SSIM, and the best LPIPS score. On the LOL-v2-Syn dataset, our method outperforms all metrics.

Comparison on LSRW-Huawei and LSRW-Nikon Datasets. Results in Tab. 2 show that we achieve the highest PSNR and SSIM on both datasets. **Notably**, our method is computationally efficient, requiring just **0.37M** parameters and **4.34G** FLOPs.

Visual Comparisons. Fig. 4 demonstrate our method's effectiveness against state-of-the-art techniques across four challenging extreme dark datasets, highlighting its superior enhancement quality in recovering fine details and preserving natural color balance in extremely dark conditions.

Ablation Study

Component Ablation. To evaluate the contribution of each component in our proposed method, we conduct an ablation study by sequentially removing essential modules. The results are summarized in Tab. 3. The removal of the second stage also leads to a noticeable decrease in metrics, indicating the importance of spatial information reconstruction. Each component has a measurable impact on performance, demonstrating the robustness of our approach.

Component Count and Patch SS2D Efficiency Validation. Fig. 5 presents an analysis of component counts and

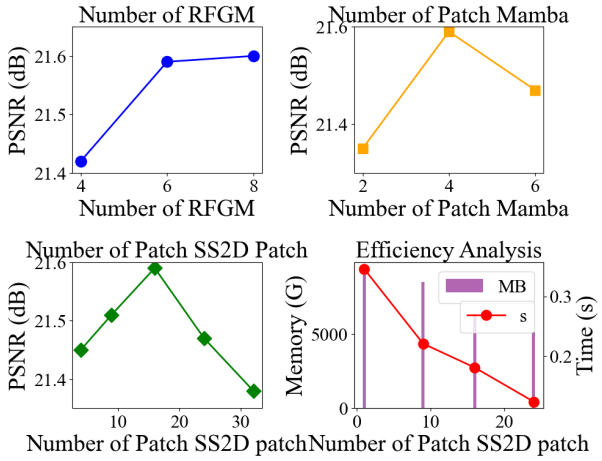

Figure 5: Analysis of component counts and the efficiency of Patch SS2D. The plots illustrate the effects of varying the number of RFGM, Patch Mamba, and Patch SS2D patch on PSNR. Additionally, the efficiency analysis highlights the relationship between the number of patches, GPU memory usage, and computation time.

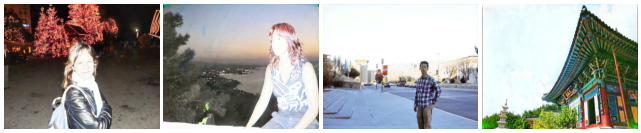

Figure 6: Examples of failure cases.

the efficiency of the Patch SS2D module. The plots demonstrate the effects of varying the number of RFGMs, Patch Mamba blocks, and Patch SS2D patches on PSNR. Specifically, the leftmost plot indicates that six RFGMs yield optimal PSNR results, while the second plot shows that four Patch Mamba blocks effectively balance performance. The third plot reveals that PSNR is maximized with 16 SS2D patches. The efficiency analysis in the bottom right plot illustrates the relationship between the number of patches, GPU memory usage, and computation time. As the number of patches increases from 1 to 24, both GPU memory consumption and computation time decrease significantly. Notably, using 16 patches strikes an optimal balance, reducing computation time by 47.7% (from 0.346s to 0.181s) and GPU memory usage by 33.2% (from 9362MB to 6252MB).

Plug-and-Play Effectiveness of Core Modules. RFGM and Patch SS2D can be seamlessly integrated into existing Fourier-based methods with minimal additional parameters. Tab. 4 shows that incorporating our modules into FourLLIE and DMFourLLIE significantly enhances performance while maintaining high efficiency, validating their effectiveness.

Limitation. Through testing on an unsupervised dataset DICM (Lee, Lee, and Kim 2012) and VV (Vonikakis, Kouskouridas, and Gasteratos 2018).), we observed that our method exhibits suboptimal performance in suppressing exposure, as illustrated in Fig.6. This issue is particularly evident in facial and sky regions.

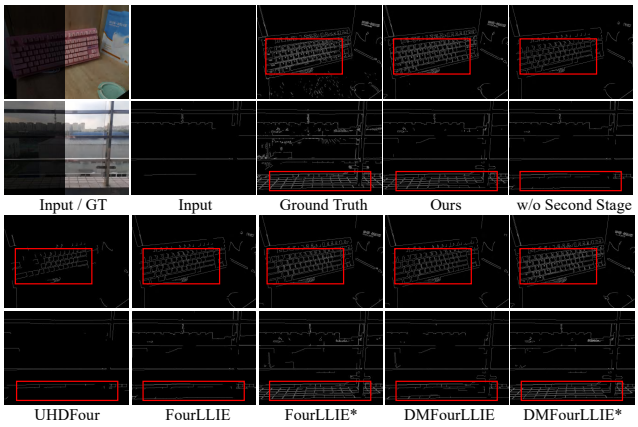

Figure 7: Canny edge detection visualization. The first column contains input and ground truth RGB images, followed by Canny edge results. We also assessed results without the second stage, confirming its role in enhancing edges and details. Special attention is given to the details in the red-boxed area; zooming in provides better visual contrast. In contrast to frequency-domain methods, which often struggle with fine detail preservation, we replaced the second stage of both FourLLIE* and DMFourLLIE* with our dual-branch Mamba, validating its efficacy in enhancing edge clarity.

Downstream Application

Edge Detection. To validate our approach in edge detection, we refer to the results in Fig. 7. The enhanced Canny edge outputs show significant improvement in preserving fine details and sharp boundaries, particularly in the red-boxed areas highlighting intricate structures. Comparing these results with the original images demonstrates that our dual-stage method effectively enhances edge definition, recovering details lost in extremely dark conditions.

Text Detection. As shown in Tab.5, we conducted text detection on the LOL-Text dataset using the Text-CP (Lin et al. 2025), alongside CRAFT (Baek et al. 2019) and PAN (Wang et al. 2019) detectors. The H-Mean, representing the mean of precision and recall, indicates a substantial improvement in our method's performance. This highlights our ability to recover details effectively in dark image conditions.

Conclusion

In this work, we presented an efficient dual-stage approach for recovering fine-grained details in extremely dark images, significantly improving edge and text detection performance. Our method utilizes a Residual Fourier-Guided Module and complementary Mamba modules, demonstrating robust enhancements in detail preservation. Looking ahead, we will explore further avenues to ensure not only the recovery of intricate details but also the restoration of color semantic consistency in dark images, enhancing overall image quality and usability in practical applications.

Acknowledgements

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

References

Baek, Y.; Lee, B.; Han, D.; Yun, S.; and Lee, H. 2019. Character region awareness for text detection. In *CVPR*, 9365–9374.

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*.

Feijoo, D.; Benito, J. C.; Garcia, A.; and Conde, M. V. 2025. Darkir: Robust low-light image restoration. In *CVPR*, 10879–10889.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Guo, H.; Guo, Y.; Zha, Y.; Zhang, Y.; Li, W.; Dai, T.; Xia, S.-T.; and Li, Y. 2025a. Mambairv2: Attentive state space restoration. In *CVPR*, 28124–28133.

Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; and Xia, S.-T. 2025b. 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.; 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.

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.

Lee, C.; Lee, C.; and Kim, C.-S. 2012. Contrast enhancement based on layered difference representation. In *ICIP*, 965–968. IEEE.

Li, B.; Zhao, H.; Wang, W.; Hu, P.; Gou, Y.; and Peng, X. 2025. Mair: A locality-and continuity-preserving mamba for image restoration. In *CVPR*, 7491–7501.

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*.

Lin, C.-T.; Ng, C. C.; Tan, Z. Q.; Nah, W. J.; Wang, X.; Kew, J. L.; Hsu, P.; Lai, S. H.; Chan, C. S.; and Zach, C. 2025. Text in the dark: Extremely low-light text image enhancement. *Signal Processing: Image Communication*, 130: 117222.

Liu, P.; Wang, X.; Zhang, T.; and Yin, L. 2025. Multi-modal Fusion Guided Retinex-based Low-Light Image Enhancement. *Expert Systems with Applications*, 128653.

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, W.; Xie, E.; Song, X.; Zang, Y.; Wang, W.; Lu, T.; Yu, G.; and Shen, C. 2019. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. In *ICCV*, 8440–8449.

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. *Advances in Neural Information Processing Systems*, 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, C.; Li, M.; Zhang, Z.; Meng, D.; and Zhang, L. 2024. Spatial-mamba: Effective visual state space models via structure-aware state fusion. *arXiv preprint arXiv:2410.15091*.

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, 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. CWNet: Causal Wavelet Network for Low-Light Image Enhancement. *arXiv preprint arXiv:2507.10689*.

Zhang, T.; Liu, P.; Zhang, Z.; and Zhou, Q. 2025b. SPJFNet: Self-Mining Prior-Guided Joint Frequency Enhancement for Ultra-Efficient Dark Image Restoration. *arXiv preprint arXiv:2508.04041*.

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, 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.