

Article

GLMA: Global-to-Local Mamba Architecture for Low-Light Image Enhancement

Wentao Li ^{1,2} , Xinhao Wu ^{1,3}, Yu Guan ^{1,3} , Sen Lin ^{1,3} , Naida Ding ¹, Qiang Wang ^{4,*}  and Yandong Tang ¹ 

¹ The State Key Laboratory of Robotics and Intelligent System, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; liwentao@sia.cn (W.L.); xinhaowu761@gmail.com (X.W.); guanyu@sylu.edu.cn (Y.G.); linsen@sylu.edu.cn (S.L.); dingnaida@sia.cn (N.D.); ytang@sia.cn (Y.T.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China

⁴ The Key Laboratory of Manufacturing Industrial Integrated Automation, Shenyang University, Shenyang 110044, China

* Correspondence: wangqiang@syu.edu.cn

Abstract

In recent years, Mamba has gained increasing importance in the field of image restoration, gradually outperforming traditional convolutional neural networks (CNNs) and Transformers. However, the existing Mamba-based networks mainly focus on capturing global contextual relationships and neglect the crucial impact of local feature interactions on restoration performance in low-light environments. These environments inherently require the joint optimization of multi-scale spatial dependencies and frequency-domain characteristics. The traditional CNNs and Transformers face challenges in modeling long-range dependencies, while State Space Models (SSMs) in Mamba demonstrate proficiency in sequential modeling yet exhibit limitations in fine-grained feature extraction. To address the limitations of existing methods in capturing global degradation patterns, this paper proposes a novel global-to-local feature extraction framework through systematic Mamba integration. The Low-Frequency Mamba Block (LFMBlock) is introduced first to perform refined feature extraction in the low-frequency domain. The High-Frequency Guided Enhancement Block (HFGBlock) is used, which utilizes low-frequency priors to compensate for texture distortions in high-frequency components. Comprehensive experiments on multiple benchmark datasets show that the Global-to-Local Mamba architecture achieves superior performance in low-light restoration and image enhancement. It significantly outperforms state-of-the-art methods in both quantitative metrics and visual quality preservation, especially in recovering edge details and suppressing noise amplification under extreme illumination conditions. The hierarchical design effectively bridges global structural recovery with local texture refinement, setting a new paradigm for frequency-aware image restoration.

Keywords: low-light environment; state space models; Low-Frequency Mamba Block; High-Frequency Guided Enhancement Block



Academic Editor: Atsushi Mase

Received: 4 September 2025

Revised: 4 October 2025

Accepted: 5 October 2025

Published: 11 October 2025

Citation: Li, W.; Wu, X.; Guan, Y.; Lin, S.; Ding, N.; Wang, Q.; Tang, Y. GLMA: Global-to-Local Mamba Architecture for Low-Light Image Enhancement. *Appl. Sci.* **2025**, *15*, 10931. <https://doi.org/10.3390/app152010931>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Within the realm of image restoration technologies [1,2], low-light enhancement [3,4], together with other crucial research areas like dehazing [5,6], motion deblurring [7], and super-resolution reconstruction [8,9], form the core technological spectrum for enhancing

digital image quality. As a cutting-edge research domain in computer vision, low-light image enhancement primarily confronts technical challenges that stem from the dual constraints of inadequate illumination and the inherent physical limitations of imaging sensors. These constraints collectively lead to nonlinear radiometric distortions, loss of frequency-domain information, and spatially heterogeneous noise patterns in captured images. These degradations not only impair human visual perception but also degrade the performance of downstream vision tasks such as object detection and semantic segmentation. Therefore, effective low-light enhancement is essential for both human-centric and machine vision applications.

The advent of convolutional neural networks [10] (CNNs) has revolutionized the landscape of deep learning, propelling hierarchical feature representation-based methods [11] to become the dominant paradigm. By leveraging multi-scale convolutional kernels to aggregate local receptive fields, CNN architectures have achieved significant improvements in objective metrics like PSNR. Their inherent weight-sharing property effectively reduces the risk of model overfitting. However, the inductive bias imposed by fixed receptive fields in CNNs limits their ability to model global illumination, while the parameter-sharing mechanism restricts their adaptability to diverse degradation patterns, particularly in scenarios requiring spatially variant restoration of non-uniform illumination artifacts and noise distributions. This architectural limitation originates from the inherent conflict between the local operation paradigm of CNNs and the global–local interdependence inherent in low-light enhancement tasks.

Recent advancements in vision enhancement have underscored the transformative potential of Transformer architectures, which transcend spatial constraints through self-attention mechanisms. Although the global attention computation in Transformer-based architectures [12,13] establishes long-range pixel dependencies, the quadratic increase in computational demands for high-resolution images introduces a new trade-off between perceptual quality and model efficiency.

Amid growing interest in efficient long-range modeling, Mamba [14] has emerged as a compelling alternative to attention-based architectures, offering a global receptive field at linear complexity through its selective scanning mechanism. Nonetheless, its inherently unidirectional, sequential prior struggles to capture fine-grained local structures and is easily perturbed by complex noise distributions, particularly when simultaneous multi-scale texture preservation and illumination correction are required. Recent remedies have introduced non-causal scanning strategies, hybrid CNN-SSM blueprints, and frequency-domain gating to inject local inductive biases, yet a systematic framework that seamlessly integrates global context with localized, frequency-aware discrimination remains elusive. In the Mamba framework [15,16], the Visual State Space Module [17] (VSSM) is adept at modeling long-range dependencies with linear complexity, thus serving as a computationally efficient alternative to conventional Transformer architectures. Accordingly, we incorporate the VSSM as the global feature extraction module within our proposed framework. This integration capitalizes on the module's ability to achieve a balance between computational efficiency and extensive contextual modeling. It enables robust hierarchical representation learning across both spatial and frequency domains, while preserving the fidelity of local texture reconstruction.

To address these limitations and enhance Mamba's local feature extraction capabilities, we introduce the Global-to-Local Mamba framework. This framework integrates CNN-based modules to augment local feature extraction, alleviate localized color distortions, and refine restoration quality in fine-grained regions. The Multi-Scale Feedforward Network (MSFFN), which builds on the GDFFN foundation, employs a sequential structure to capture multi-scale features and enable precise recovery of textures and structural details.

This framework proposes a novel approach for low-light image enhancement by combining Mamba, wavelet transforms, and a global-to-local feature extraction strategy within a U-Net architecture. The hierarchical design synergizes Mamba's linear-complexity global modeling with CNN's localized inductive bias, facilitating adaptive illumination correction and artifact suppression across both spatial and frequency domains.

Our main contributions are summarized as follows:

- We propose a Global-to-Local Mamba network for low-light image restoration that effectively captures intricate global and local dependency relationships.
- We incorporate wavelet transforms to avert information loss during downsampling and devise a framework that amalgamates a Global Mamba Block, predicated on the Visual State Space Module (VSSM) for long-range dependency modeling, which can progressively distill multi-level features.
- We integrate a novel Multi-Scale Feedforward Network (MSFFN) to supply complementary structural features across scales, leveraging rich low-frequency feature information to guide the restoration of high-frequency details.
- Extensive experiments conducted across multiple benchmark datasets demonstrate the superior performance of the proposed model, achieving state-of-the-art results in quantitative metrics while maintaining exceptional visual fidelity in restored images.

2. Related Works

2.1. Low-Light Image Enhancement

Low-light image restoration is dedicated to recovering image details and textures in low-light conditions. With the advent of novel paradigms and frameworks, this field has witnessed significant advancements. Traditional low-light image restoration methods [18–20] commonly utilize techniques such as histogram equalization, gamma correction, and Retinex theory [21]. These methodologies enhance image contrast and brightness by either adjusting grayscale distributions or applying nonlinear functions to modify pixel values. In recent years, the progression of deep learning [22] has spurred the continuous proposal of data-driven approaches for low-light image enhancement (LLIE), as illustrated in Figure 1. LLNet is a pioneering CNN-based method in this domain, employing a stacked sparse denoising autoencoder architecture to concurrently achieve noise suppression and illumination enhancement in low-light images. In the realm of LLIE, numerous methods focus on improving illumination through dynamic range expansion and contrast enhancement. Among these, Histogram Equalization (HE)-based approaches are particularly noteworthy, encompassing Global HE (GHE) and Local HE (LHE). Despite their popularity due to simplicity and computational efficiency, these methods often introduce artifacts such as over-enhancement or under-enhancement, which can result in detail loss in the processed images.

Zero-DCE [23] and its enhanced version, Zero-DCE++ [24], approach low-light image enhancement (LLIE) as an image-specific curve estimation task. They investigate unsupervised learning frameworks to enhance the generalization capability of networks. These methods achieve substantial progress in enhancing image brightness and visual appeal by adaptively estimating nonlinear pixel-wise mappings. They effectively circumvent the reliance on paired training data while maintaining natural color consistency. Moreover, diffusion models, such as the DiffLL network [25], have been employed to generate realistic details in low-light tasks through a denoising process. Recently, RetinexMamba [26] and MambaLLIE [27] have explored the use of Mamba-based networks to tackle LLIE challenges. Despite the promising performance of current LLIE methods, they predominantly concentrate on the influence of global features, often overlooking the crucial guiding role of local information in refining global representations. This oversight restricts their ability

to integrate fine-grained details with holistic scene understanding, especially in scenarios that demand precise illumination recovery and texture preservation.

2.2. State Space Models (SSMs)

In recent years, there have been significant advancements in deep learning frameworks based on State Space Models (SSMs) [17]. The SSM architecture exhibits remarkable computational efficiency in modeling long-range dependencies, primarily due to its inherent linear computational complexity in sequence processing. This represents a substantial advantage over traditional Recurrent Neural Networks (RNNs) [28], which exhibit quadratic complexity and are prone to gradient decay in long-sequence scenarios. Given a one-dimensional input sequence $x(t) \in \mathbb{R}$, the system maps it to a new one-dimensional output sequence $y(t) \in \mathbb{R}$ through a latent state transformation $h(t) \in \mathbb{R}$. This dynamical process is formally defined by a linear ordinary differential equation (ODE), where the evolution of the latent state explicitly governs the input–output mapping.

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t) \quad (1)$$

$$y(t) = \mathbf{C}h(t) + \mathbf{D}x(t) \quad (2)$$

Owing to the intrinsically analog characteristics of continuous-time state-space models, discretization is essential for their computational implementation. The discrete analogs of the continuous parameters $\mathbf{A}$ and $\mathbf{B}$, represented as $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$, are obtained via the zero-order hold (ZOH) method [29]. This transformation ensures compatibility with digital systems while maintaining the dynamic properties of the original model. The ZOH-based discretization process can be mathematically formulated as follows:

$$h'_t = \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t \quad (3)$$

$$y_t = \mathbf{C}h_t + \mathbf{D}x_t \quad (4)$$

$$\bar{\mathbf{A}} = e^{\Delta \mathbf{A}} \quad (5)$$

$$\bar{\mathbf{B}} = (\Delta \mathbf{A})^{-1} (e^{\Delta \mathbf{A}} - \mathbf{I}) \cdot \Delta \mathbf{B} \quad (6)$$

where Δ denotes a step size.

This approach innovatively establishes a data-driven dynamic parameter selection mechanism. It not only outperforms Transformer models on natural language processing benchmarks but also retains the inherent linear complexity advantage with respect to input scale. Among these advancements, MambaIR [1] pioneers the first end-to-end optimization framework for image super-resolution tasks based on State Space Models (SSMs), while EVSSM constructs a lightweight SSM-based network for motion deblurring. In recent years, RetinexMamba [26] first reformulates the classical Retinex decomposition as a continuous-time state-space model and solves it with a discretized Mamba backbone, thereby endowing low-light enhancement with a clear physical interpretation and strict linear complexity. To fully develop the potential of Mamba in the low-illumination field, MambaLLIE [27] introduces a Local-Enhanced State Space Block that enriches the original 1-D Mamba scan with 2-D neighborhood residuals, while an implicit Retinex prior generated from max-mean statistics is injected through a lightweight depthwise gating mechanism. Nevertheless, the max-mean prior remains a single, scene-averaged statistic, so it mistakes overlapping sharp shadows for genuine illumination edges; the network then amplifies these false contours, producing noticeable color shifts. Wave-Mamba [30] first explores the role of frequency domain features in low-light enhancement, which casts the feature map into a wavelet packet domain and applies a frequency-selective Mamba scan that skips low-magnitude coefficients, achieving token reduction without sacrificing perceptual

information. These methods based Mamba have successfully extended SSM frameworks to low-light image enhancement (LLIE) by designing multi-scale feature fusion mechanisms. Building upon these technological advancements, our paper proposes a novel global–local feature extraction paradigm to enhance high-order semantic representation learning within SSMs for LLIE tasks. This approach addresses the limitations of existing methods in capturing hierarchical illumination–reflectance relationships.

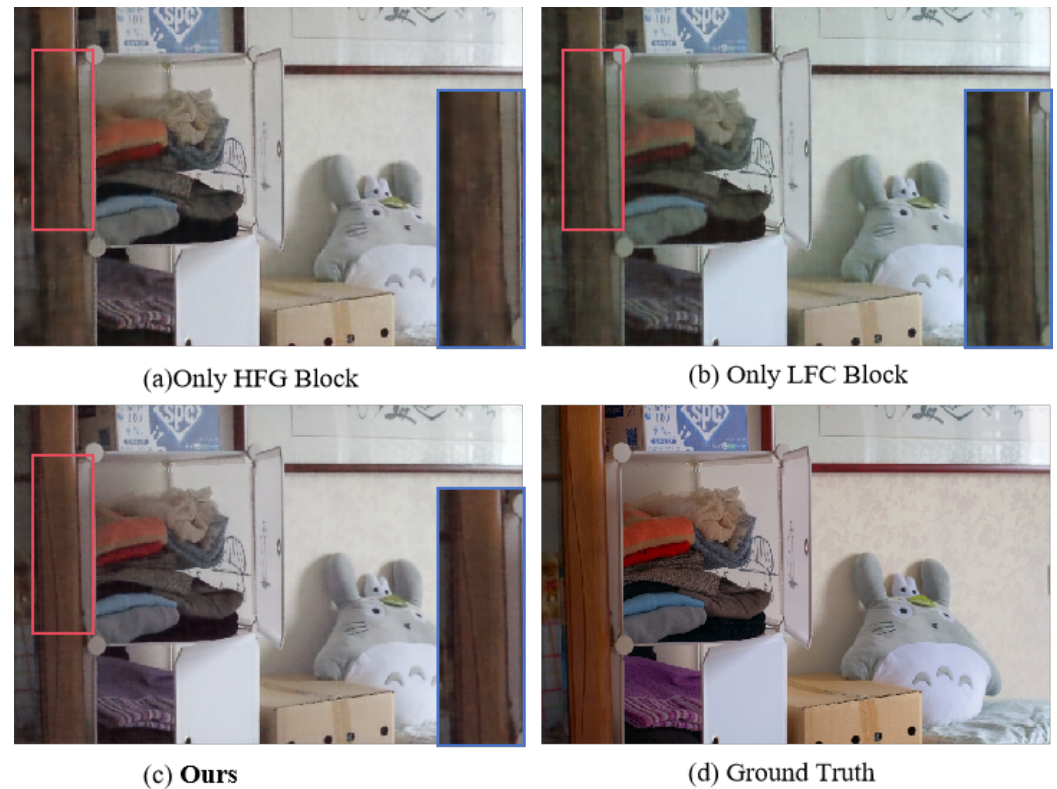


Figure 1. We compare qualitative results on the LOL-v1 dataset using three strategies: (a) Only Retain High-Frequency Blocks, (b) Only Retain Low-Frequency Blocks, and (c) our proposed method. Zoom in for a better view.

2.3. Wavelet Transformation

The wavelet transform serves as a time–frequency analysis tool that captures localized signal features by decomposing signals into “wavelet basis functions” across various scales. Unlike the Fourier transform, which solely provides global frequency information, the wavelet transform offers localized information in both the time (or spatial) and frequency domains simultaneously. Within this context, DiffLL introduces the first enhancement framework based on a wavelet-domain diffusion model.

The wavelet transformation process can be illustrated as follows:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt. \quad (7)$$

In this formulation, a denotes the scale parameter that governs frequency resolution, b represents the translation parameter that controls spatial localization, and ψ corresponds to the mother wavelet function.

The Discrete Wavelet Transform (DWT) [31] is the discretized version of the wavelet transform, utilizing filter banks to achieve multi-scale decomposition. This process employs low-pass and high-pass filters to decompose a signal into low-frequency components (approximation coefficients) and high-frequency components (detail coefficients). The

decomposition results in a low-frequency subband (LL) and three high-frequency subbands (LH, HL, HH). These subbands correspond to horizontal, vertical, and diagonal directional details, respectively. The DWT represents the two-dimensional Discrete Wavelet Transform operation. Subsequently, the decomposed frequency subbands can be reconstructed into the original signal via the Inverse Wavelet Transform (IWT) without information loss. This property enables effective downsampling and upsampling of images while preserving critical structural and textural details, which is the principal rationale for the iterative application of the DWT in such tasks.

3. Method

In this section, we delineate our methodology through a tripartite exposition. First, we elaborate on the theoretical motivation underpinning the adopted approach. Second, we present the Local Feature Modeling (LFM) module, which constitutes the methodological cornerstone of the framework. Third, we detail the High Frequency Guidance (HFG) Block module, designed to refine feature interactions through spectral prioritization.

3.1. Framework Overview

The proposed network architecture utilizes the Discrete Wavelet Transform (DWT) to decompose input images into high-frequency and low-frequency components. A specialized Low-Frequency Mamba (LFM) Block is designed to extract semantically rich features from the low-frequency subbands. This configuration enables the network to simultaneously enhance and optimize both global contextual information and local structural details while maintaining linear computational complexity, a critical advantage inherited from the State Space Model (SSM) framework.

A high-frequency enhancement module is proposed, based on a low-frequency guided compensation strategy. This module constructs a frequency-domain attention constraint through the fusion of global and local low-frequency components. This mechanism reconstructs the spatial gradient distribution of high-frequency subbands, thereby establishing a cross-frequency feature coupling framework. This framework optimizes inter-band interactions while preserving edge coherence and spectral consistency.

The proposed method effectively addresses the high-frequency texture distortion inherent in conventional approaches. It leverages a constrained deconvolutional network to achieve precise reconstruction of local microstructural features. This framework ensures that multi-scale neural networks can accurately extract physically meaningful edge oscillation characteristics and gradient-direction-consistent texture patterns.

In the following sections, we will detail the comprehensive workflow of our methodology and provide an in-depth exposition of its core components.

Our network architecture is constructed based on a multi-scale U-Net [32] framework. For downsampling, we employ the Discrete Wavelet Transform (DWT) [31] instead of conventional methods, effectively avoiding information loss typically caused by traditional downsampling operations. This is attributed to the inherent ability of DWT to preserve texture details during resolution reduction. Leveraging the characteristic of decomposing images into high-frequency and low-frequency components, we apply a global-to-local feature extraction strategy to the low-frequency subbands, enabling the derivation of more precise feature representations. These representations are then processed through a Multi-Scale Feedforward Network to hierarchically aggregate structural information from the image features, thereby enhancing the model's capability to recover fine-grained details and maintain spatial-spectral consistency.

For the high-frequency components, we refrain from performing dedicated feature extraction operations and instead leverage low-frequency components to correct

and restore high-frequency information. This strategy effectively reduces computational costs while preserving critical edge details and spectral fidelity through cross-frequency interaction mechanisms.

3.2. Low-Frequency Mamba Block

The LFMBlock is designed to restore low-frequency features in low-light images. It extracts low-frequency information flows from the spatial domain of feature embeddings and models them through the LFMBlock architecture, as illustrated in Figure 2. Given the low-frequency input features $F_n \in \mathbb{R}^{H \times W \times C}$, we first apply Layer Normalization to stabilize feature distributions, followed by the Visual State Space Module (VSSM) to capture global contextual relationships and long-range dependencies. This hierarchical paradigm extracts a robust representation by progressively distilling global contextual cues, while simultaneously preserving the linear complexity and memory efficiency that are intrinsic to state-space formulations.

Building upon the global semantic relationships established by the VSSM, we provide contextual guidance for subsequent local feature extraction via the Local Feature Module (LFM). To preclude mutual gradient interference between the VSSM and LFM, we instantiate a principled sequential topology that operationalizes a global-first-local-later inductive bias. By enforcing a unidirectional flow of supervisory signals, the scheme guarantees temporal coherence of gradient update directions across successive modules, thereby eradicating stationary-point conflicts and assuring stable, joint optimization.

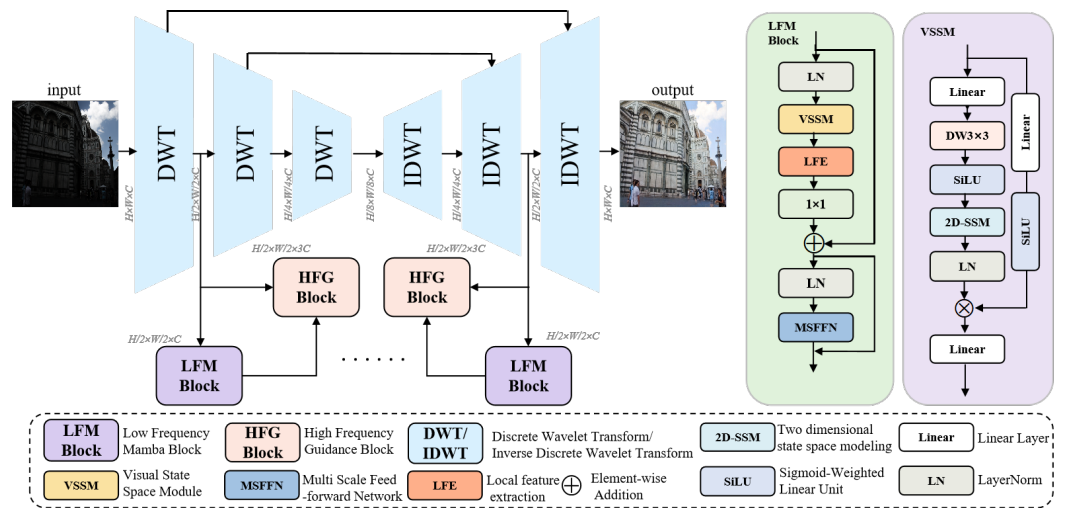


Figure 2. The proposed Global-to-Local Mamba architecture constructs a hierarchical framework that integrates wavelet transform-based upsampling and downsampling operations. Specifically, the Low-Frequency Mamba Block (LFMBlock) performs multi-scale feature extraction on frequency-decoupled components, while the High-Frequency Guided Enhancement Block (HFGBlock) refines details through gradient-aware attention mechanisms.

Finally, a Multi-Scale Feedforward Network (MSFFN) is employed to hierarchically learn and refine structural representations. The overall process can be formulated as follows:

$$F'_i = W_{1 \times 1}(VSSM(LN(F_i)) + LFE(LN(F_i))) + F_i, \quad (8)$$

$$F_o = MSFFN(LN(F'_i)) + F'_i, \quad (9)$$

where $VSSM(\cdot)$, $MSFFN(\cdot)$, and $LFM(\cdot)$ denote the functional operations of the Visual State Space Module, MSFFN, and LFM, respectively. $W_{1 \times 1}(\cdot)$ represents a convolutional layer with a 1×1 kernel, while LN indicates the layer normalization applied to the input features. The advantages of integrating VSSM into the LSM module can be summarized as follows:

- **Computational Efficiency:** Using state-space models like VSSM allows for linear complexity with respect to sequence length, which is beneficial for high-resolution images.
- **Improved Optimization:** The sequential structure avoids the common issue of gradient cancellation when parallel global and local paths are used.
- **Contextual Coherence:** Global context informs local processing, which is especially useful in low-light scenarios where local features might be noisy or ambiguous.

Vision State Space Module: Drawing inspiration from the linear computational complexity demonstrated by the Mamba architecture in long-range dependency modeling, we integrate the Visual State Space Module (VSSM) into the framework of low-light image enhancement (LLIE) tasks. The proposed model systematically establishes inter-regional feature correlations through discrete state-space equations and achieves efficient low-light image restoration via a multi-phase synergistic processing mechanism. The architecture of the VSSM is illustrated in Figure 2.

The input features first undergo channel dimension adjustment via a linear projection layer, which projects the raw data into a high-dimensional representation space to enhance expressive capacity. Subsequently, a depthwise separable convolution (DW3 × 3) performs lightweight local feature extraction through a decoupled strategy of spatial filtering and cross-channel fusion.

Following this, the two-dimensional state space modeling (2D-SSM) unfolds the image bidirectionally along the primary and secondary diagonals, as shown in Figure 3. It dynamically fuses multi-path features via learnable directional gating mechanisms. The processed features are stabilized through layer normalization to calibrate their distribution and are ultimately mapped to the target space via a linear output layer. In parallel, another branch incorporates a linear layer followed by a SiLU activation. The outputs from this branch are additively integrated into the target space to further refine feature representations.

$$X_1 = \text{SS2D} - \text{SSM}(\text{SiLU}(\text{DWConv}(\text{Linear}(X)))) \quad (10)$$

$$X_{\text{out}} = \text{Linear}(\text{LN}(X_1) \odot \text{SiLU}(\text{Linear}(X))) \quad (11)$$

where $\odot$ denotes the Hadamard product, which generates an output with the same dimensionality as the input. $\text{Linear}(\cdot)$ and $\text{DWConv}(\cdot)$ represent the linear projection and depthwise convolution operations, respectively.

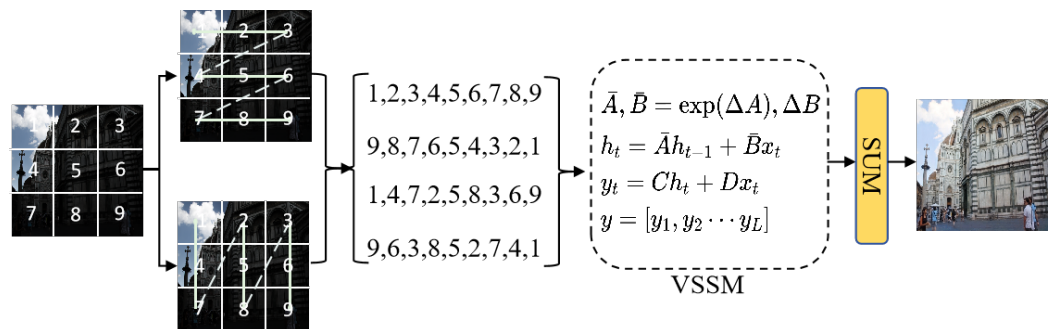


Figure 3. The 2D-SSM framework functions by decomposing the input image via bidirectional diagonal scanning mechanisms. Specifically, the spatial tensor is disentangled along two principal axes: the primary diagonal (extending from the top-left to the bottom-right) and the secondary diagonal (extending from the bottom-left to the top-right). These scans are executed in both forward and reverse orientations, thereby generating four distinct scanning trajectories.

Local Feature Extraction: Given that the Visual State Space Module (VSSM) leverages self-attention mechanisms to excel in capturing global contextual semantic relationships, we serially connect the Low-Frequency Mamba Block (LFM) after the VSSM's global semantic

modeling to avoid conflicting optimization directions between global and local feature extraction branches. This sequential architecture enables progressive stage-wise feature refinement: the LFM focuses on extracting fine-grained texture details and short-range features, thereby refining low-frequency components with higher precision. The enhanced low-frequency representations subsequently guide the High-Frequency Guided Enhancement Block (HFG Block) to amplify high-frequency features through targeted spectral adjustments, ensuring synergistic interaction between global and local information flows.

As illustrated in Figure 4a, the input features are first processed through Global Average Pooling (GAP), which aggregates spatial information to generate a channel-wise global descriptor capturing the overall response intensity of each channel. Subsequently, two 1×1 convolutional operations are employed to establish a cross-channel interaction pathway: the first convolution reduces the channel dimensionality while incorporating a ReLU activation function to model inter-channel dependency patterns; the second convolution restores the original dimensionality, followed by a Sigmoid function to generate normalized channel attention weights $\mathcal{W}$. This sequential design enables adaptive recalibration of channel-wise feature importance through learned nonlinear interactions. The resulting attention weights are then element-wise multiplied with the original features along the channel dimension, enabling adaptive feature enhancement of critical channels and dynamic suppression of redundant information. Consequently, we formalize the Lightweight Feature Module (LFM) as follows:

$$\mathcal{W} = \sigma(\mathbf{W}_{1 \times 1}(\text{ReLU}(\mathbf{W}_{1 \times 1}(\text{GAP}(\mathbf{F}_n)))))) \quad (12)$$

$$\mathbf{F}_{\text{local}} = \mathcal{W} \odot \mathbf{F}_n \quad (13)$$

where $\sigma(\cdot)$ denotes the Sigmoid function, $\text{ReLU}(\cdot)$ represents the Rectified Linear Unit activation function, and $\text{GAP}(\cdot)$ corresponds to Global Average Pooling.

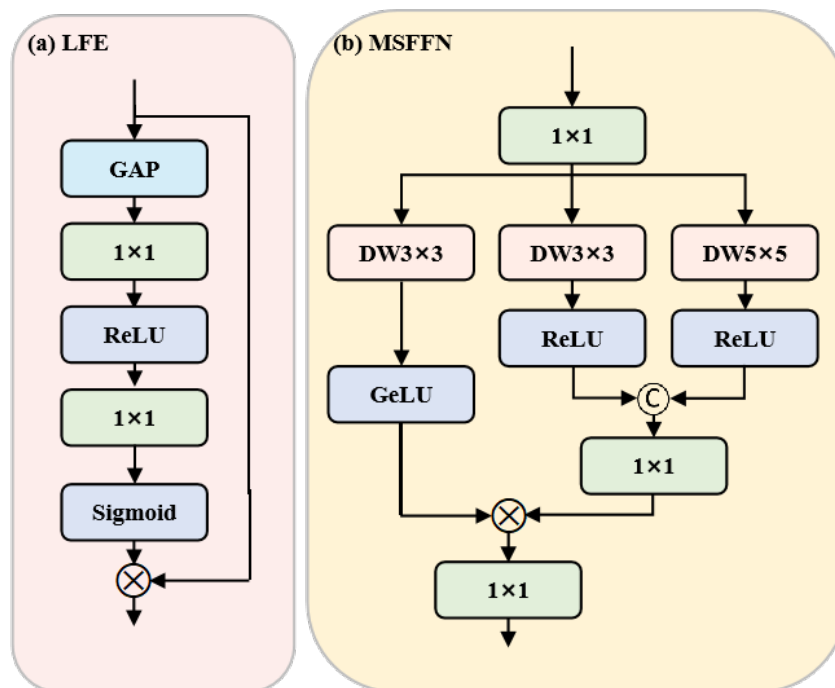


Figure 4. (a) Local Feature Extraction (LFM). (b) Multi-Scale Feed-Forward Network (MSFFN).

Multi-Scale Feedforward Network: To augment the capacity for extracting structural information from image features, this study introduces a Multi-Scale Hybrid Feedforward Network (MSFFN), depicted in Figure 4b. Initially, the module promotes cross-channel

interaction of the input features via a 1×1 convolutional layer, which is succeeded by the formation of a multi-branch parallel processing architecture.

In the main branch, a 3×3 depthwise separable convolution coupled with a GELU activation function is utilized to conduct fundamental feature extraction. The gated branch extends the conventional GDFFN framework by incorporating 3×3 and 5×5 depthwise separable convolutional layers, which facilitates the acquisition of multi-granularity spatial features.

Subsequently, features from all branches are activated by ReLU, concatenated along the channel dimension, and then compressed back to the original dimensionality through a 1×1 convolutional layer. The concluding step involves element-wise multiplication to adaptively fuse the primary branch features with gating weights, resulting in an optimized output that retains both local fine-grained details and global structural coherence. The operational workflow of the MSFFN is formulated as follows:

$$F_1, F_2, F_3 = \text{Chunk}(W_{1 \times 1}(F_n)) \quad (14)$$

$$F_a = \text{GeLU}(DW_{3 \times 3}(F_1)) \quad (15)$$

$$F_b = W_{1 \times 1}(\text{ReLU}(DW_{3 \times 3}(F_2)) \otimes \text{ReLU}(DW_{5 \times 5}(F_3))) \quad (16)$$

$$F' = W_{1 \times 1}(F_a \odot F_b) \quad (17)$$

where $\text{Chunk}(\cdot)$ denotes the chunk function along the channel dimension.

3.3. High Frequency Guidance Block

To enhance the feature representation capability of high-frequency components, this study introduces a low-frequency-guided cross-frequency feature migration mechanism. By leveraging the semantic consistency of low-frequency components, this approach performs collaborative optimization on high-frequency features, addressing the issue of weakened high-frequency details often seen in conventional methodologies.

The core component of this mechanism is the High-Frequency Guided Block (HFG-Block), which employs a dual-path collaborative architecture comprising two critical components: the Frequency-Matching Attention Module (FMAM) and the Frequency-Correction Feedforward Network (FCFN).

The FMAM performs frequency-domain alignment operations to identify cross-frequency semantic cues within low-frequency components that exhibit strong correlations with high-frequency features. Concurrently, the FCFN applies nonlinear mapping and distribution calibration to the fused cross-frequency features, effectively mitigating phase misalignment artifacts introduced during inter-band integration.

Through iterative interaction between low-frequency and high-frequency features, this framework dynamically fills information-deficient regions in high-frequency sub-bands while preserving edge sharpness and suppressing structural artifacts. The detailed workflow is illustrated as follows:

$$\mathbf{F}'_H = \text{FMAM}(\text{LN}(\mathbf{F}_H^{ln}), \mathbf{F}_L) + \mathbf{F}_H^{ln} \quad (18)$$

$$\mathbf{F}_H^{out} = \text{FCFN}(\text{LN}(\mathbf{F}'_H), \mathbf{F}_L) + \mathbf{F}' \quad (19)$$

where F_H^{ln} and F_H^{out} represent the input and output high-frequency features of the HFGBlock, respectively. The notation $\text{FMAM}(\cdot)$ and $\text{FCFN}(\cdot)$ corresponds to the operational functions of the FMAM and FCFN.

Frequency Matching Attention Module: Building on prior research into query effectiveness in attention mechanisms, this study proposes a low-frequency feature-guided semantic enhancement strategy to optimize the attention weight generation mechanism.

Frequency Correction Forward Network: Subsequent to the processing through the Frequency-Matching Attention Module (FMAM), a hierarchical cross-domain feature refinement architecture is implemented to further facilitate the enhancement of high-frequency components. This architecture achieves joint spatial-frequency domain enhancement via multi-stage collaborative processing, as illustrated in Figure 5a. The processing pipeline is formulated as follows: Initially, Layer Normalization (LN) is applied to the output features of the FMTA to stabilize channel-wise distributions and accelerate model convergence. Subsequently, a 1×1 convolution operation performs cross-channel information interaction and dimensional adaptation, constructing an efficient feature representation space for subsequent operations. Following this, a local spatial context extraction module based on 3×3 convolutions captures fine-grained texture and edge features. The Frequency Matching Transformation (FMT) is then executed to establish cross-band semantic correlations between low- and high-frequency features, thereby reinforcing the synergistic representation of structural patterns and detailed components. The pipeline ultimately outputs optimized results with enhanced visual consistency, as formally described below:

$$FCFN(F_H^l, F_L^e) = FMT(W_3 W_1 (LN(F_H^l)) F_L^e) \quad (23)$$

Frequency Matching Transformation: As illustrated in Figure 5b, the Frequency Matching Transformation (FMT) primarily functions to transfer low-frequency features into enhanced high-frequency representations through the following computational workflow: Initially, a similarity matrix between high-frequency and low-frequency components is computed to identify the optimal feature vector D . Subsequently, the most semantically aligned low-frequency components are selected as channel-specific outputs based on D . These selected features are then concatenated with the original high-frequency features via parallel processing branches.

$$Y_s = \text{Select}(F_L | \text{Indices}(\text{Top}_1(\text{Sim}(F_L, F_H)))) \quad (24)$$

$$Y_s^c = \text{Concat}(Y_s, F_H), \quad (25)$$

where the notation $\text{Sim}(\cdot, \cdot)$ denotes the similarity computation measured by Euclidean distance, $\text{Select}(\cdot | \cdot)$ represents the feature selection operation with conditional filtering, and $\text{Indices}(\cdot)$ indicates the index retrieval operation.

Specifically, one branch computes attention maps using 1×1 convolutional layers coupled with Sigmoid activation to emphasize critical spatial locations, while the other branch employs 3×3 convolutions to capture contextual attention patterns. Finally, the outputs from both branches are multiplied and integrated through an additional 1×1 convolutional layer to produce the refined output features. The complete procedural pipeline is formally described as follows:

$$F_H^{\text{out}} = W_1(\text{Sigmoid}(W_1(Y_s^c))) \odot W_3(Y_s^c) \quad (26)$$

Compared with conventional attention mechanisms, the proposed Frequency-Matching Transformation (FMT) offers three distinct advantages:

- Frequency-Matching Transformation module ensures that only semantically coherent low-frequency cues are injected into the high-frequency, eliminating semantic misalignment.
- Frequency-Matching Transformation module offers statistically reliable cues for fine-grained detail recovery, resulting in perceptually faithful edge sharpening.
- The residual pathway ensures unimpeded gradient flow, facilitating stable end-to-end back-propagation.

4. Experiment

The proposed Global-to-Local Mamba (GL-Mamba) framework is subjected to rigorous evaluation across a variety of vision tasks, utilizing several widely adopted datasets for low-light enhancement benchmarking. Furthermore, comprehensive ablation studies are conducted to systematically quantify the efficacy of individual architectural components. The first two comparative results are analytically highlighted in red and blue, respectively, to delineate performance variations under distinct configurations.

4.1. Implementation Details

The network architecture adopts a hierarchical configuration, comprising [1, 2, 4] LFMBlocks and [1, 1, 1] HFGBlocks per layer within the encoder–decoder framework. The model employs eight attention heads and maintains a channel dimension C of 32 throughout. Training is conducted using the AdamW optimizer, initialized with a learning rate of 5×10^{-4} , which is progressively decayed to 1×10^{-7} via cosine annealing over 100k iterations. Data augmentation techniques include random geometric transformations (90, 180, 270), random flips, and random cropping to 512×512 patches. The Global-to-Local Mamba framework is optimized under an L1 loss constraint. All experiments are executed on a NVIDIA RTX 3090 GPU (24 GB) setup.

4.2. Datasets

The experimental validation utilizes two benchmark datasets: LOL-v1 and LOL-v2-synthetic. The original Low-Light Dataset version 1, specifically curated for low-light image enhancement research, consists of 500 aligned low/normal-light image pairs with a fixed resolution of 400×600 pixels. This dataset comprises 485 training pairs and 15 testing pairs, primarily capturing indoor scenarios. The enhanced LOL-v2 version introduces two distinct subsets: LOL-v2-real, captured under authentic low-light conditions, and LOL-v2-synthetic, generated through illumination distribution analysis of RAW images. Our experiments specifically employ the LOL-v2-synthetic subset, which contains 1000 synthetically generated low- or normal-light image pairs. These pairs are divided into 900 training pairs and 100 testing pairs through systematic data partitioning.

4.3. Comparisons with State-of-the-Art Methods

In this section, we provide a comprehensive evaluation of the proposed Global-to-Local Mamba framework through quantitative and qualitative comparisons with state-of-the-art methods. The assessment utilizes three well-established metrics: PSNR, SSIM, and LPIPS. PSNR quantifies reconstruction quality by calculating the pixel intensity differences between images, with higher values indicating greater reconstruction fidelity. SSIM evaluates perceptual quality by analyzing luminance, contrast, and structural similarity, where values closer to 1 signify better preservation of image characteristics. LPIPS measures perceptual similarity using deep neural networks, with lower scores indicating better alignment with human visual perception. Collectively, these metrics offer a holistic view of both pixel-level accuracy and perceptual authenticity across different enhancement paradigms.

Quantitative Evaluation: The quantitative evaluation results on the LOL-v1 dataset are systematically summarized in Table 1. The proposed Global-to-Local Mamba framework is benchmarked against 12 state-of-the-art (SOTA) low-light enhancement methodologies, including HVIet, KinD [33], Zero-DCE [23], RUAS [34], EnlightenGAN [35], UFormer [36], IAT [37], PairLIE [38], SCI [39], LLFormer [40], Wave-Mamba [30], and HVI [41]. As demonstrated quantitatively, the Global-to-Local Mamba achieves superior PSNR performance, outperforming all comparative methods by a significant margin of 0.7 dB. This empirical

evidence substantiates the method’s exceptional capability in preserving structural fidelity during illumination recovery processes.

Table 1. Quantitative comparisons on LOL-v1 datasets. The best result is in red color while the second best result is in blue color.

Methods	Venue	PSNR	SSIM	LPIPS
RetinexNet [26]	BMVC 2018	16.77	0.56	0.47
KinD [33]	MM 2019	17.65	0.72	0.17
ZeroDCE [23]	CVPR 2020	14.86	0.58	0.33
RUAS [34]	CVPR 2021	16.40	0.50	0.27
EnlightenGAN [35]	TIP 2021	17.48	0.76	0.32
UFormer [36]	CVPR 2022	16.36	0.77	0.32
IAT [37]	BMVC 2022	21.30	0.81	0.32
PairLIE [38]	CVPR 2023	19.51	0.73	0.24
SCI [39]	CVPR 2022	14.78	0.52	0.33
LLFormer [40]	AAAI 2023	23.65	0.84	0.16
WaveMamba [30]	ACM MM 2024	23.27	0.84	0.14
HVI [41]	CVPR 2025	23.80	0.85	0.08
Ours	-	23.91	0.84	0.14

The quantitative evaluation results on the LOL-v2-synthetic dataset are meticulously detailed in Table 2. This table juxtaposes the proposed Global-to-Local Mamba framework against 11 state-of-the-art low-light image enhancement methods, namely RetinexNet [26], KinD [33], ZeroDCE [23], RUAS [34], UFormer [36], Bread [42], PairLIE [38], LLFormer [40], QuadPrior [43], and Wave-Mamba [30]. As evidenced in Table 2, the Global-to-Local Mamba attains optimal performance metrics in both PSNR and SSIM on this benchmark, thereby establishing new advancements in synthetic low-light image restoration.

Table 2. Quantitative comparisons on LOL-v2-synthetic datasets. The best result is in red color while the second best result is in blue color.

Methods	VENUE	PSNR	SSIM	LPIPS
RetinexNet [26]	BMVC 2018	17.13	0.76	0.25
KinD [33]	MM 2019	18.32	0.79	0.25
RUAS [34]	CVPR 2021	13.76	0.63	0.30
ZeroDCE [23]	CVPR 2020	17.71	0.81	0.17
Uformer [36]	CVPR 2022	19.66	0.87	-
PairLIE [38]	CVPR 2023	19.07	0.79	0.23
LLFormer [40]	AAAI 2023	24.03	0.90	0.06
Bread [42]	JVC 2023	17.63	0.91	0.09
GSAD [44]	NeurIPS 2023	24.47	0.92	0.05
QuadPrior [43]	CVPR 2024	16.10	0.75	0.11
Wave-Mamba [30]	ACMM 2024	24.76	0.92	0.06
Ours	-	24.87	0.93	0.06

Qualitative Evaluation: Figure 6 presents a visual comparison between our method and existing approaches. Current methods often exhibit insufficient illumination, failing to restore fine details effectively. Additionally, color distortion and image degradation further compromise the enhancement outcomes of prior techniques. In contrast, our Global-to-Local Mamba framework not only effectively enhances brightness but also re-constructs intricate details with high precision. It demonstrates superior capability in amplifying low-visibility and low-contrast regions. The proposed method reliably eliminates noise without introducing artificial artifacts and robustly preserves original chromatic

information throughout the enhancement process. This dual-scale architecture achieves balanced performance by integrating global contextual modeling with local feature refinement, ensuring both radiometric consistency and structural fidelity in challenging illumination conditions.

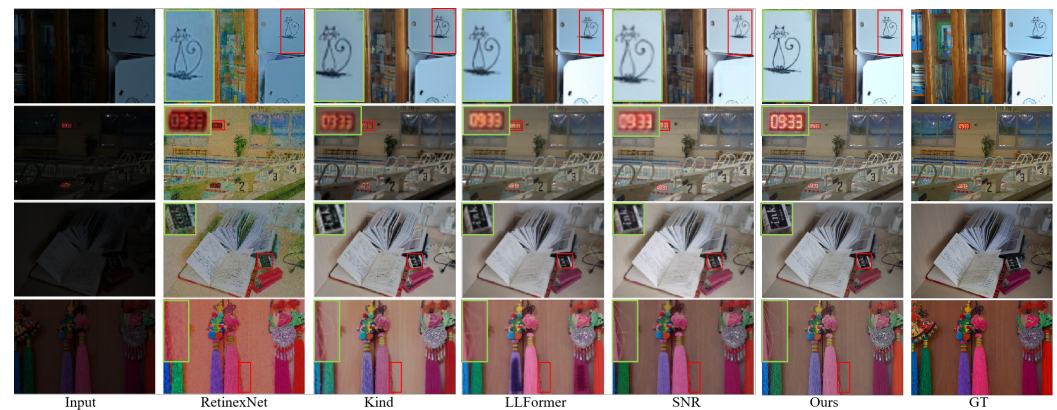


Figure 6. Visualization of the low-light enhancement model. Each column is a different image example, and each row is the prediction of the models. Comparative visualization style is an updated version based on [30].

Model Parameters and Efficiency: We also conduct experiments on the parameters and efficiency of the model. The experimental results are shown in Table 3. It can be seen that the proposed GLMA achieves the best balance between model complexity and computational efficiency. With only 1.9 M parameters and 9.02 G FLOPs, it significantly outperforms most competing methods, including both lightweight models (e.g., PairLIE, KinD) and large-scale models (e.g., GSAD, QuadPrior). Despite its compact size, GLMA delivers superior performance in terms of PSNR and SSIM, demonstrating the effectiveness of the Global-to-Local Mamba design and wavelet-based frequency decomposition strategy. This lightweight yet powerful architecture makes GLMA highly suitable for deployment on resource-limited devices and real-time applications.

Table 3. Calculation of model parameters and efficiency.

Method	Param (M)	Flops (G)
KinD	8.02	34.99
PairL	0.33	20.81
LLFormer	24.55	22.52
Uformer	50.88	45.90
GSAD	217.36	442.02
QuadPrior	1252.7	1103.2
Ours	1.9	9.02

4.4. Ablation Study

Low-frequency Mamba Block: As the core component of the Global-to-Local Mamba framework, we conduct ablation experiments to validate the functional roles of individual modules. The results in Table 4 show that the wave-mamba architecture, which prioritizes global feature extraction followed by local refinement, achieves a 0.64 dB improvement in PSNR. These findings substantiate the practical significance of enhancing low-frequency components for subsequent high-frequency refinement, while also confirming the intrinsic value of low-frequency features in low-light image restoration. Compared to the LFMBBlock, the High-Frequency Guidance Blocks (HFGBlocks) impose a more substantial impact on

model parameters but yield a relatively minor performance gain. This underscores the efficiency superiority of our low-frequency-centric design paradigm.

Table 4. Ablation study on the proposed modules. The highest scores are indicated in red.

Method	PSNR	SSIM
w/o LFMBlock	22.81	0.82
w/o HFGBlock	22.95	0.83
w/o LFE	23.06	0.83
w/o MSFFN	23.12	0.83
w/o FMT	23.58	0.84
Full Model	23.91	0.84

5. Conclusions

In this study, we introduce an effective framework for low-light image enhancement (LLIE) based on a Global-to-Local Mamba Architecture. Motivated by the ability of wavelet transforms to decompose high- and low-frequency features, we employ wavelet-based downsampling to mitigate information loss during feature decomposition. The framework adopts a global-to-local feature extraction strategy to holistically capture structural information in the low-frequency domain. These low-frequency features further guide the refinement of high-frequency details through cross-domain interaction. Extensive experiments on multiple datasets demonstrate the compelling performance of our method in enhancing low-light images.

Nevertheless, several limitations of the proposed approach deserve further discussion. First, the current single-level Discrete Wavelet Transform (DWT) may fall short in representing mid-frequency components, which can lead to aliasing artifacts in highly textured regions. Second, the Frequency Matching Transformation (FMT) relies on a hand-crafted similarity metric, which can be sensitive to noise and complex illumination variations. Third, the lack of an explicit noise modeling module limits the model's robustness under real-world noisy conditions. Additionally, while the method performs well on static images, its behavior on video sequences—such as temporal consistency and processing speed—has not been evaluated, which is critical for deployment in video applications.

In future work, we will explore multi-level wavelet decomposition to enhance mid-frequency representation. We also plan to design a learnable frequency matching mechanism to replace the fixed similarity metric in FMT, and integrate an explicit noise modeling branch to improve robustness. Extending the current framework to video-based low-light enhancement and restoration will be another important direction, with the goal of maintaining temporal coherence and enabling real-time processing. We will also investigate model compression and acceleration strategies to facilitate practical deployment on resource-constrained devices.

Author Contributions: Conceptualization, W.L.; data curation, S.L. and Y.G.; formal analysis, W.L. and Y.T.; funding acquisition, W.L.; investigation, Q.W.; methodology, W.L. and N.D.; project administration, Y.T.; resources, Y.G. and X.W.; software, W.L. and N.D.; supervision, Q.W.; validation, S.L.; visualization, W.L., X.W. and N.D.; writing—review editing, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by 14th Five Year National Key R & D Program Project (No. 2023YFB3211001), the National Natural Science Foundation of China (Grant No. 62073205), the General Talents Project for Scientific Research grant of the Educational Department of Liaoning Province (LJ212410144062), and the Science and Technology Program of Liaoning Province (2023JH26/10300015).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The LOL-v1 dataset is an open dataset accessed on 18 February 2020 and can be downloaded at <https://arxiv.org/abs/1808.04560>. The LOL-v2 dataset is an open dataset accessed on 28 May 2018 and can be downloaded at <https://doi.org/10.1109/TIP.2021.3050850>.

Acknowledgments: This work is supported by 14th Five Year National Key R & D Program Project (No. 2023YFB3211001), the National Natural Science Foundation of China (Grant No. 62073205), the General Talents Project for Scientific Research grant of the Educational Department of Liaoning Province (LJ212410144062), and the Science and Technology Program of Liaoning Province (2023JH26/10300015).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Guo, H.; Li, J.; Dai, T.; Ouyang, Z.; Ren, X.; Xia, S.T. Mambair: A simple baseline for image restoration with state-space model. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; pp. 222–241.
- Shi, Y.; Xia, B.; Jin, X.; Wang, X.; Zhao, T.; Xia, X.; Xiao, X.; Yang, W. VmambaIR: Visual State Space Model for Image Restoration. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 5560–5574. [\[CrossRef\]](#)
- Li, J.; Li, B.; Tu, Z.; Liu, X.; Guo, Q.; Juefei-Xu, F.; Xu, R.; Yu, H. Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 15205–15215.
- Liang, D.; Xu, Z.; Li, L.; Wei, M.; Chen, S. PIE: Physics-inspired low-light enhancement. *Int. J. Comput. Vis.* **2024**, *132*, 3911–3932. [\[CrossRef\]](#)
- Goyal, B.; Dogra, A.; Lepcha, D.C.; Goyal, V.; Alkhayyat, A.; Chohan, J.S.; Kukreja, V. Recent advances in image dehazing: Formal analysis to automated approaches. *Inf. Fusion* **2024**, *104*, 102151. [\[CrossRef\]](#)
- Zhang, Y.; Zhou, S.; Li, H. Depth information assisted collaborative mutual promotion network for single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 2846–2855.
- Xiang, Y.; Zhou, H.; Li, C.; Sun, F.; Li, Z.; Xie, Y. Deep learning in motion deblurring: Current status, benchmarks and future prospects. *Vis. Comput.* **2024**, *41*, 3801–3827. [\[CrossRef\]](#)
- Su, H.; Li, Y.; Xu, Y.; Fu, X.; Liu, S. A review of deep-learning-based super-resolution: From methods to applications. *Pattern Recognit.* **2024**, *157*, 110935. [\[CrossRef\]](#)
- Lei, X.; Zhang, W.; Cao, W. Dvmsr: Distillated vision mamba for efficient super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 6536–6546.
- Zhao, X.; Wang, L.; Zhang, Y.; Han, X.; Deveci, M.; Parmar, M. A review of convolutional neural networks in computer vision. *Artif. Intell. Rev.* **2024**, *57*, 99. [\[CrossRef\]](#)
- Shou, Y.; Cao, X.; Liu, H.; Meng, D. Masked contrastive graph representation learning for age estimation. *Pattern Recognit.* **2025**, *158*, 110974. [\[CrossRef\]](#)
- Feng, K.; Ma, Y.; Wang, B.; Qi, C.; Chen, H.; Chen, Q.; Wang, Z. Dit4edit: Diffusion transformer for image editing. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 2969–2977.
- Xia, C.; Wang, X.; Lv, F.; Hao, X.; Shi, Y. Vit-comer: Vision transformer with convolutional multi-scale feature interaction for dense predictions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 5493–5502.
- Ma, X.; Zhang, X.; Pun, M.O. Rs 3 mamba: Visual state space model for remote sensing image semantic segmentation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6011405. [\[CrossRef\]](#)
- Zhao, H.; Zhang, M.; Zhao, W.; Ding, P.; Huang, S.; Wang, D. Cobra: Extending mamba to multi-modal large language model for efficient inference. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 10421–10429.
- Shen, H.; Wan, Z.; Wang, X.; Zhang, M. Famba-v: Fast vision mamba with cross-layer token fusion. In Proceedings of the European Conference on Computer Vision, Dublin, Ireland, 22–23 October 2025; pp. 268–278.
- Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. Vmamba: Visual state space model. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 103031–103063.
- Chang, M.; Feng, H.; Xu, Z.; Li, Q. Low-light image restoration with short-and long-exposure raw pairs. *IEEE Trans. Multimed.* **2021**, *24*, 702–714. [\[CrossRef\]](#)

19. Xu, K.; Yang, X.; Yin, B.; Lau, R.W. Learning to restore low-light images via decomposition-and-enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2281–2290.
20. Wu, X.; Lai, Z.; Yu, S.; Zhou, J.; Liang, Z.; Shen, L. Coarse-to-fine low-light image enhancement with light restoration and color refinement. *IEEE Trans. Emerg. Top. Comput. Intell.* **2023**, *8*, 591–603. [\[CrossRef\]](#)
21. Land, E.H.; McCann, J.J. Lightness and retinex theory. *J. Opt. Soc. Am.* **1971**, *61*, 1–11. [\[CrossRef\]](#)
22. Zhao, L.; Wang, K.; Zhang, J.; Wang, A.; Bai, H. Learning deep texture-structure decomposition for low-light image restoration and enhancement. *Neurocomputing* **2023**, *524*, 126–141. [\[CrossRef\]](#)
23. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.
24. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4225–4238. [\[CrossRef\]](#)
25. Jiang, H.; Luo, A.; Fan, H.; Han, S.; Liu, S. Low-light image enhancement with wavelet-based diffusion models. *ACM Trans. Graph. (TOG)* **2023**, *42*, 238. [\[CrossRef\]](#)
26. Bai, J.; Yin, Y.; He, Q.; Li, Y.; Zhang, X. Retinexmamba: Retinex-based mamba for low-light image enhancement. In Proceedings of the ICONIP 2024, Auckland, New Zealand, 2–6 December 2024; pp. 427–442.
27. Weng, J.; Yan, Z.; Tai, Y.; Qian, J.; Yang, J.; Li, J. Mamballie: Implicit retinex-aware low light enhancement with global-then-local state space. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 27440–27462.
28. Medsker, L.R.; Jain, L. Recurrent neural networks. *Des. Appl.* **2001**, *5*, 64–67.
29. Mai, H.; Yin, Z. A new zero-order algorithm to solve the maximum hands-off control. *IEEE Trans. Autom. Control* **2023**, *69*, 2761–2768. [\[CrossRef\]](#)
30. Zou, W.; Gao, H.; Yang, W.; Liu, T. Wave-mamba: Wavelet state space model for ultra-high-definition low-light image enhancement. In Proceedings of the 32nd ACM International Conference on Multimedia, Melbourne, VIC, Australia, 28 October–1 November 2024; pp. 1534–1543.
31. Osadchiy, A.; Kamenev, A.; Saharov, V.; Chernyi, S. Signal processing algorithm based on discrete wavelet transform. *Designs* **2021**, *5*, 41. [\[CrossRef\]](#)
32. Azad, R.; Aghdam, E.K.; Rauland, A.; Jia, Y.; Avval, A.H.; Bozorgpour, A.; Karimijafarbigloo, S.; Cohen, J.P.; Adeli, E.; Merhof, D. Medical Image Segmentation Review: The Success of U-Net. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10076–10095. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Zhang, Y.; Zhang, J.; Guo, X. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1632–1640.
34. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.
35. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693.
37. Cui, Z.; Li, K.; Gu, L.; Su, S.; Gao, P.; Jiang, Z.; Qiao, Y.; Harada, T. You only need 90k parameters to adapt light: A light weight transformer for image enhancement and exposure correction. *arXiv* **2022**, arXiv:2205.14871. [\[CrossRef\]](#)
38. Fu, Z.; Yang, Y.; Tu, X.; Huang, Y.; Ding, X.; Ma, K.K. Learning a simple low-light image enhancer from paired low-light instances. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 22252–22261.
39. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward fast, flexible, and robust low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5637–5646.
40. Jie, H.; Zuo, X.; Gao, J.; Liu, W.; Hu, J.; Cheng, S. Llformer: An efficient and real-time lidar lane detection method based on transformer. In Proceedings of the 2023 5th International Conference on Pattern Recognition and Intelligent Systems, Shenyang, China, 28–30 July 2023; pp. 18–23.
41. Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; Zhang, Y. Hvi: A new color space for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 10–17 June 2025; pp. 5678–5687.
42. Guo, X.; Hu, Q. Low-light image enhancement via breaking down the darkness. *Int. J. Comput. Vis.* **2023**, *131*, 48–66. [\[CrossRef\]](#)

43. Wang, W.; Yang, H.; Fu, J.; Liu, J. Zero-reference low-light enhancement via physical quadruple priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 26057–26066.
44. Hou, J.; Zhu, Z.; Hou, J.; Liu, H.; Zeng, H.; Yuan, H. Global structure-aware diffusion process for low-light image enhancement. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 79734–79747.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.