photonics

MDPI

Article

Zero-MS: A Zero-Shot Low-Light Image Enhancement Network with Mamba-Inspired Denoising and Structure-Aware Refinement

Lingyun Wang, Minli Tang *, Qun Chai, Guanglin Yang and Nianting Pan

School of Big Data Engineering, Kaili University, Kaili 556011, China; wanglingyun@kluniv.edu.cn (L.W.); chaiqun@kluniv.edu.cn (Q.C.); yangguanglin@kluniv.edu.cn (G.Y.); pnt089@126.com (N.P.)
* Correspondence: tangml@stu.xmu.edu.cn

Abstract

Low-light images often suffer from insufficient illumination and noise, but enhancing dark regions can further amplify noise and degrade texture and structural details. To address these challenges, we propose Zero-MS, a zero-shot method for low-light image enhancement. Zero-MS constructs a noise prior through normalized four-directional cumulative scanning and adaptive selective gating. It then restores illumination through multi-scale information modeling and refines the result using structural and noise responses. Experimental results show that Zero-MS balances information fidelity, visual quality, and noise suppression. The enhanced images exhibit improved brightness, color preservation, and noise suppression, while Zero-MS maintains a lightweight model design.

Keywords: low-light image enhancement; zero-shot; Retinex; multi-scale features; noise suppression

1. Introduction

Under weak or spatially nonuniform illumination, the image sensor receives insufficient photons across the entire scene or in local regions, making signals in dark areas more susceptible to imaging noise. The resulting images typically exhibit low brightness or contrast, along with possible color shifts, texture degradation, and loss of local structural details. Such images are referred to as low-light images. Their degradation arises from the combined effects of scene illumination, exposure settings, sensor response, and imaging noise. During enhancement, directly increasing the brightness of dark regions can amplify hidden noise, while excessive denoising may remove genuine edges and fine textures [1–4]. Therefore, Low-Light Image Enhancement (LLIE) requires not only brightness restoration but also a careful balance between noise suppression and structural preservation. LLIE methods have evolved into several categories, including conventional and learning-based, zero-reference, and zero-shot approaches. Conventional methods improve image brightness through contrast adjustment, gamma correction, and Retinex-based processing [5–9]. Learning-driven methods use CNNs, Transformers, and frequency-domain representations to improve the modeling of complex degradations [3,10–14]. Zero-reference and zero-shot methods further reduce the dependence on training data [4,15–18]. Recent studies on state-space models provide another way to jointly model noise, illumination, color, and spatial dependencies [2,19]. Although existing methods have improved low-light image restoration through illumination decomposition, frequency-domain analysis, physical

Check for updates

Received: 2 September 2026
Revised: 18 October 2026
Accepted: 19 October 2026
Published: 22 October 2026

Copyright: © 2026 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.

noise priors, and long-range dependency modeling, noise, illumination, and structural information remain closely coupled in real-world scenes. Random noise and fine textures can produce similar high-frequency responses, while signal-to-noise characteristics vary with brightness. Uniform noise suppression may therefore leave residual noise or remove important structural details. In addition, texture recovery depends on local details, whereas correcting uneven exposure requires broader spatial context. A single scale is therefore insufficient to address both tasks. Retinex decomposition explicitly models reflectance and illumination but does not inherently suppress imaging noise. Reflectance recovery can also make disturbances in dark regions more visible [3,12]. For zero-shot methods, balancing these factors becomes more difficult without paired-image supervision. Based on the above considerations, we propose Zero-MS, a zero-shot network that combines Mamba-inspired denoising and structure-aware refinement within a noisy Retinex imaging model. Zero-MS consists of three modules: the Mamba-Inspired Denoising Module (MIDM), the Multi-Scale Illumination Restoration Module (MSLM), and the Structure-Aware Refinement Module (SRM). Rather than adopting a full Mamba-based enhancement backbone or explicitly estimating a signal-to-noise ratio (SNR) reliability map from noise statistics, MIDM uses state accumulation to construct a content-adaptive noise prior through four-directional scanning and dark-region, structural, and chromaticity gating. MSLM captures local details and global exposure using full-resolution and low-resolution branches, respectively. SRM then selectively refines the reflectance component based on structural and noise responses. Together, these modules address noise propagation, differences in spatial scale, and the difficulty of distinguishing structure from noise. The main contributions are summarized as follows:

- We propose Zero-MS, a zero-shot low-light image enhancement method based on a noisy Retinex imaging model. It jointly addresses noise suppression, illumination restoration, and structure preservation.
- MIDM constructs a content-adaptive noise prior using normalized four-directional cumulative scanning and selective gating based on dark-region, structural, and chromatic information. This design adapts noise processing to image content and limits noise propagation during subsequent enhancement.
- MSLM combines a full-resolution branch for local details with a low-resolution branch for global exposure. This multi-scale design balances texture preservation and spatially nonuniform illumination restoration.
- SRM refines the reflectance component through structure–noise dual gating. Combined with joint reflectance–illumination denoising, it suppresses residual high-frequency disturbances while preserving structural details.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed Zero-MS framework and its main components. Section 4 reports and analyzes the experimental results. Section 5 discusses the findings and limitations. Finally, Section 6 presents the conclusions.

2. Related Work

2.1. Conventional and Learning-Based Low-Light Enhancement Methods

Conventional LLIE methods mainly rely on hand-crafted priors, illumination models, and optimization strategies to improve image visibility. These methods typically estimate illumination or decompose image components based on predefined assumptions. MSRCR [6] uses multi-scale Retinex to adjust the image dynamic range. LIME [7] improves dark-region visibility by estimating and refining an illumination map. NPLIE [8] uses multi-scale guided filtering to estimate an initial structure-aware illumination map, and then refines it through optimization to preserve image structure and naturalness. Later studies also consider color

and multi-scale structural information. Jeon et al. [9] estimate a transmission map using a gamma prior in a hybrid color space and apply pixel-adaptive gamma correction to reduce under- and over-enhancement. Lang et al. [20] combine guided filtering with side-window guided filtering to separate an image into background, salient-edge, and detail layers. Brightness and details are then adjusted separately for each layer. Although these methods have clear model-based foundations, their performance depends largely on predefined decomposition schemes and enhancement rules. With advances in deep learning, research has shifted from hand-crafted assumptions toward data-driven feature representations. Learning-based methods use neural networks to model interactions among illumination, structure, color, and noise. RetinexNet [10] learns reflectance–illumination decomposition, while Retinexformer [3] captures long-range dependencies through illumination-guided nonlocal interactions. HVI [21] improves color representation, whereas URWKV [22] focuses on modeling long-range dependencies. To reduce computational costs, SCI [23] uses weight sharing and self-calibrated illumination, while IAT [24] combines local and global adjustments for color and exposure correction. DLIENet [14] further applies knowledge distillation to transfer feature representations and content information from a teacher network to a lightweight student network. Recent studies have also explored frequency-domain information and state-space models. CWFAS-Net [13] uses signal-to-noise ratio information to guide the fusion of local and global features. ICDMamba [19] separates illumination and color degradations and models them through frequency–spatial collaborative Mamba. LowLightReward [25] combines Mamba-based enhancement with visual aesthetic constraints. PPMNet [26] incorporates physical noise priors and Visual Mamba to model noise statistics and adjust state propagation according to regional signal-to-noise characteristics. Together, these approaches improve low-light image restoration by decomposing illumination, improving color representation, modeling frequency-domain interactions, and modeling long-range dependencies. However, most still rely on paired training data to learn degradation mappings.

2.2. Zero-Reference and Zero-Shot Methods

To reduce reliance on paired training data, zero-reference and zero-shot methods have become important approaches to low-light image enhancement. Zero-DCE [15] formulates enhancement as pixel-wise high-order curve estimation and uses reference-free losses to constrain exposure, color, and spatial consistency. Zero-DCE++ [16] further simplifies the curve estimation network to reduce computational cost while maintaining enhancement performance. Subsequent studies have incorporated physical models and image-specific priors into the enhancement process. ZRCRN [17] uses a camera response model to constrain image enhancement. QuadPrior [27] combines multiple physical priors derived from a single image to guide restoration, reducing reliance on external training data. More recent zero-shot methods further explore internal image information and degradation priors. ZERO-IG [4] generates downsampled image pairs from the input and jointly performs illumination-guided enhancement and denoising. CoLIE [18] estimates illumination through a coordinate-conditioned implicit function. ZeroIDIR [28] combines adaptive illumination correction with perturbation-consistent diffusion. It is optimized using only the degraded input and restores image details through diffusion-based reconstruction. These studies demonstrate the potential of internal image statistics, physical priors, and generative priors for enhancement without paired supervision. However, illumination, noise, and structural information remain closely coupled in low-light images. In low-signal-to-noise-ratio regions, brightness enhancement can amplify hidden noise. Meanwhile, random noise and fine textures may produce similar high-frequency responses, making them difficult to distinguish. Therefore, balancing noise suppression, illumina-

tion restoration, and detail preservation remains a key challenge for zero-shot low-light image enhancement.

2.3. Noise-Aware Low-Light Image Enhancement

Under low-light conditions, weak image signals make noise more noticeable, while brightness enhancement can further amplify noise in dark regions. LIME [7] introduces a denoising step to reduce noise revealed during enhancement. DI-Retinex [12] extends the Retinex model to account for noise, quantization errors, nonlinearity, and dynamic-range overflow in digital imaging. ZERO-IG [4] jointly addresses enhancement and denoising. It generates downsampled image pairs from a single input for initial denoising and uses illumination information to guide adaptive enhancement and subsequent noise suppression. Other methods incorporate noise estimates directly into the restoration process. Qiao et al. [29] derive a signal-to-noise ratio (SNR) map from the input image and use progressive feature fusion and texture refinement to process regions with different SNR levels. PPMNet [26] constructs a pixel-wise reliability map based on Poisson–Gaussian noise statistics to adjust state propagation in Mamba. Kozlov et al. [30] estimate camera noise components from a single RAW image using scene segmentation, multiple approximations, and weighted averaging. These approaches address noise through joint enhancement and denoising, SNR guidance, and physical noise modeling. However, several rely on explicit SNR or reliability maps, predefined noise statistics, or dedicated estimation procedures. Under unknown noise conditions, distinguishing random disturbances from genuine structures using a single input remains challenging. This requires balancing noise suppression, illumination restoration, and structural preservation.

3. Proposed Method

3.1. Overall Architecture

Based on Retinex theory [31], we model a low-light image as the product of reflectance and illumination, plus an additional noise term. The degradation process is expressed as follows:

$$\boldsymbol{I} = \boldsymbol{R} \odot \boldsymbol{L} + \boldsymbol{N} \tag{1}$$

where $\boldsymbol{I}$ denotes the observed low-light image, $\boldsymbol{R}$ represents the reflectance component containing intrinsic scene information, $\boldsymbol{L}$ denotes the illumination component, and $\boldsymbol{N}$ represents imaging noise. The operator $\odot$ denotes element-wise multiplication. This formulation guides the design of Zero-MS MIDM models and suppresses noise before illumination enhancement. MSLM restores spatially nonuniform illumination, while SRM further refines the reflectance component by distinguishing meaningful structures from residual high-frequency disturbances. The overall architecture is shown in Figure 1. Following this formulation, Zero-MS adopts a sequential restoration strategy consisting of noise suppression, illumination recovery, and structural refinement. As shown in Figure 1, the input image passes sequentially through MIDM, MSLM, and SRM. During training, paired downsampling generates two complementary sub-images to provide additional consistency constraints. MIDM estimates noise residuals from the original image and the paired sub-images. It constructs a noise prior through normalized four-directional cumulative scanning and adjusts its contribution using structural information. Depthwise separable convolutions and residual blocks then extract features and produce an initial denoised image. Next, MSLM takes the denoised image as input. It combines local details from a full-resolution branch with global exposure information from a low-resolution branch to restore illumination. SRM then uses multi-scale average pooling to generate structure and noise response maps. These maps guide reflectance refinement, preserving structural details while suppressing residual high-frequency disturbances. The network

is jointly optimized end-to-end using losses for low-light image denoising, illumination estimation, reflectance denoising, and perceptual quality. The enhanced image is produced through the full-resolution restoration path. The modules and loss functions are described in detail in Sections 3.2–3.5.

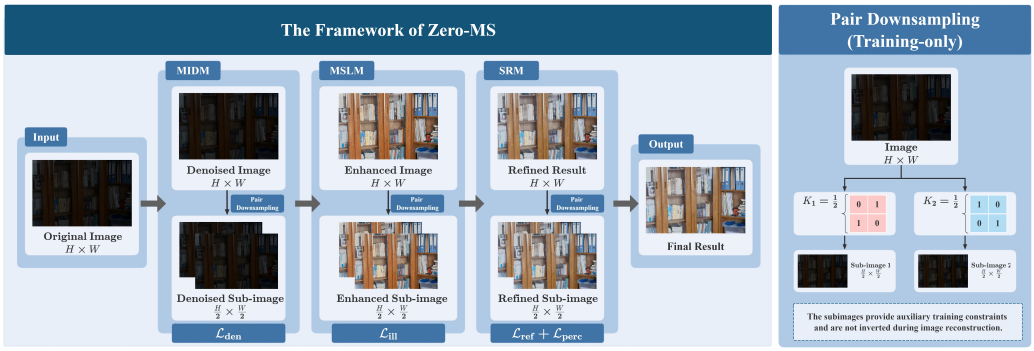


Figure 1. Overall framework of Zero-MS. Zero-MS sequentially employs MIDM for noise suppression, MSLM for illumination restoration, and SRM for reflectance refinement to produce the final enhanced image. During training, the original image and paired sub-images jointly constrain denoising, illumination restoration, and reflectance refinement. The perceptual loss is further applied to the final and paired reflectance predictions. Paired downsampling is not involved in output reconstruction and therefore requires no inverse operation. The details are provided in Sections 3.2–3.5.

3.2. Mamba-Inspired Denoising Module

High-frequency responses in low-light images contain both structural details and random noise. Using these responses directly for noise estimation may lead to the unintended suppression of useful structures. To address this issue, Zero-MS introduces MIDM, a lightweight denoising network, as shown in Figure 2. Inspired by selective state propagation in Mamba, MIDM uses deterministic bidirectional cumulative scanning to aggregate directional information rather than a conventional recurrent state-space model. Since high-frequency responses alone cannot reliably distinguish noise from meaningful structures, MIDM further introduces a content-adaptive modulation mechanism. Noise is often more noticeable in dark regions, whereas genuine structures generally exhibit consistent spatial patterns. Accordingly, selective gates are designed using luminance, structural variation, and chromatic information. These gates strengthen noise-related responses while limiting the suppression of edges and textures. The gated features are then used to construct a noise prior, which is injected into the input through weighted modulation. This provides content-aware noise guidance for subsequent residual prediction.

3.2.1. Mamba-Inspired Four-Directional Cumulative Scanning

Given a low-light input image $\boldsymbol{I}$, the network first extracts its local low-frequency component using size-preserving average pooling with a 3×3 kernel. The high-frequency residual is then computed as follows:

$$\boldsymbol{N}_h = \boldsymbol{I} - \mathcal{A}_3(\boldsymbol{I}) \tag{2}$$

where $\mathcal{A}_k(\cdot)$ denotes size-preserving average pooling with a $k \times k$ kernel, and k is the kernel size. The high-frequency residual $\boldsymbol{N}_h \in \mathbb{R}^{B \times C \times H \times W}$ contains high-frequency noise, edges, and fine textures. Here, B, C, H, and W denote the batch size, number of channels, image height, and image width, respectively. This residual serves as the input to subsequent directional scanning. Inspired by Mamba-style state aggregation, the network applies

normalized bidirectional cumulative scanning along the horizontal and vertical axes to obtain four directional states:

$$\begin{cases} S_{lr}(i,j) = \dfrac{1}{j}\displaystyle\sum_{t=1}^{j} N_h(i,t) \\ S_{rl}(i,j) = \dfrac{1}{W-j+1}\displaystyle\sum_{t=j}^{W} N_h(i,t) \\ S_{tb}(i,j) = \dfrac{1}{i}\displaystyle\sum_{t=1}^{i} N_h(t,j) \\ S_{bt}(i,j) = \dfrac{1}{H-i+1}\displaystyle\sum_{t=i}^{H} N_h(t,j) \end{cases} \tag{3}$$

where i and j denote the spatial row and column indices, respectively, and t is the accumulation index along the scanning direction. The directional states S_{lr}, S_{rl}, S_{tb}, and S_{bt} correspond to left-to-right, right-to-left, top-to-bottom, and bottom-to-top scanning, respectively. We normalize each cumulative response by its effective path length, yielding the mean response from the corresponding image boundary to the current position. This normalization prevents the state magnitude from increasing systematically with the scanning path length. The four directional states are then combined with equal weights to obtain the aggregated spatial state S:

$$S = \frac{1}{4}(S_{lr} + S_{rl} + S_{tb} + S_{bt}) \tag{4}$$

The aggregated state S is neither the denoised output nor a direct noise estimate. It still contains both genuine structural information and random high-frequency disturbances. Therefore, it serves only as the spatial-state input for subsequent noise-prior construction. Its contribution is further modulated by the dark-region, structural, and chromatic gates.

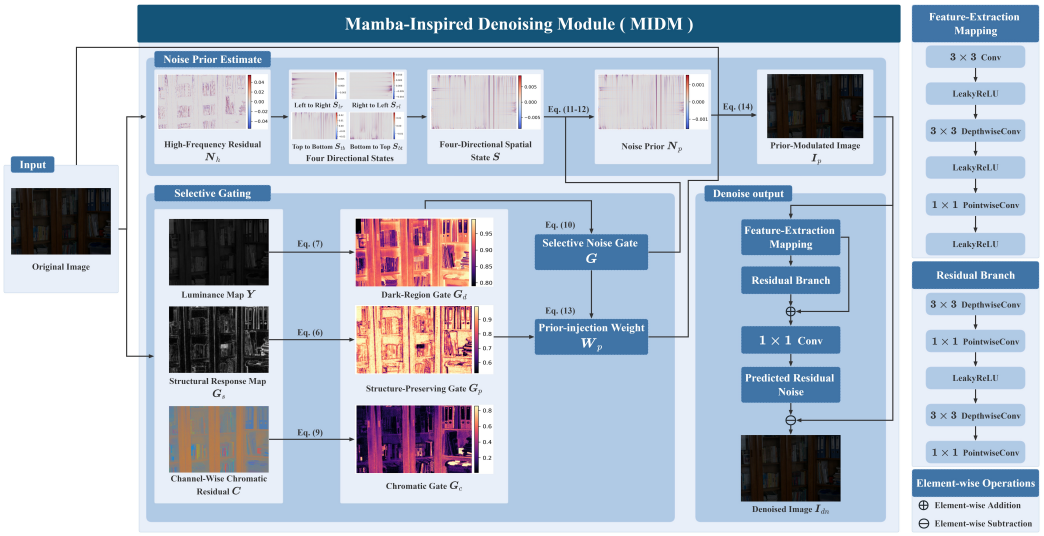


Figure 2. Overview of the MIDM. MIDM constructs a noise prior through four-directional scanning and selective gating. The denoised image is then obtained through feature-extraction mapping, together with a residual branch.

3.2.2. Noise-Aware Modulation

Although four-directional scanning aggregates spatial information over a relatively large region, the resulting directional states still contain genuine structures, fine textures, and random high-frequency disturbances. Propagating these responses without distinction may cause the noise prior to interfere with useful structures in texture-rich regions. Accord-

ingly, the network constructs a content-adaptive selective gate based on local luminance, structural strength, and chromatic variation. This gate modulates the aggregated four-directional state. Specifically, size-preserving average pooling with 3×3 and 7×7 kernels is applied to $\boldsymbol{Y}$, the Y component of the input low-light image. The relative difference between the resulting maps is then used to construct the structural response map $\boldsymbol{G}_s$:

$$\boldsymbol{G}_s = 2\alpha_1 \mathcal{A}_3\left(\frac{|\mathcal{A}_3(\boldsymbol{Y}) - \mathcal{A}_7(\boldsymbol{Y})|}{\mathcal{A}_7(\boldsymbol{Y}) + \varepsilon}\right) \tag{5}$$

where $\alpha_1 = 0.75$, and ε is a small constant that prevents division by zero. A higher $\boldsymbol{G}_s$ indicates a larger relative difference between the two smoothing scales, suggesting an edge or fine texture. The structure-preserving gate $\boldsymbol{G}_p$ is defined as follows:

$$\boldsymbol{G}_p = \boldsymbol{1} - \alpha_1 \boldsymbol{G}_s \tag{6}$$

A stronger structural response produces a smaller $\boldsymbol{G}_p$, reducing the modulation of directional states in regions with prominent structures. This limits the influence of the noise prior on genuine edges and textures. Since noise is generally more pronounced in dark regions, the network further defines a dark-region gate $\boldsymbol{G}_d$ as the complement of the luminance map:

$$\boldsymbol{G}_d = \boldsymbol{1} - \boldsymbol{Y} \tag{7}$$

Lower luminance produces a larger $\boldsymbol{G}_d$, increasing the modulation weight of the directional state in dark regions. Beyond luminance and structure, low-light noise can cause chromatic fluctuations across channels. Here, $\langle \cdot \rangle$ denotes averaging over the specified elements. To characterize these variations, we first compute the per-pixel channel mean $\langle \boldsymbol{I} \rangle_c$ of the input image $\boldsymbol{I}$. The channel-wise chromatic residual $\boldsymbol{C}$ is then defined as follows:

$$\boldsymbol{C} = \boldsymbol{I} - \langle \boldsymbol{I} \rangle_c \tag{8}$$

The chromatic response $\langle |\boldsymbol{C}| \rangle_c$ is computed as the mean absolute residual across channels. To reduce the effect of local luminance, this response is normalized by the 5×5 average-pooled luminance map, yielding the chromatic gate $\boldsymbol{G}_c$:

$$\boldsymbol{G}_c = \frac{\langle |\boldsymbol{C}| \rangle_c}{\boldsymbol{\mathcal{A}}_5(\boldsymbol{Y}) + \epsilon} \tag{9}$$

The structure-preserving, dark-region, and chromatic gates are combined multiplicatively. The coefficient $\alpha_2 = 0.7$ limits the influence of chromatic variations. A 3×3 average-pooling operation then produces the final selective noise gate $\boldsymbol{G}$:

$$\boldsymbol{G} = \boldsymbol{\mathcal{A}}_3\left[\boldsymbol{G}_d \odot \boldsymbol{G}_p \odot (\alpha_2 + (1 - \alpha_2)\, \boldsymbol{G}_c)\right] \tag{10}$$

The selective noise gate $\boldsymbol{G}$ is then applied to the aggregated four-directional state $\boldsymbol{S}$ and the channel-wise chromatic residual $\boldsymbol{C}$, yielding the spatial noise state $\boldsymbol{N}_s$ and chromatic noise state $\boldsymbol{N}_c$, respectively:

$$\begin{cases} \boldsymbol{N}_s = \boldsymbol{G} \odot \boldsymbol{S} \\ \boldsymbol{N}_c = \boldsymbol{G} \odot \boldsymbol{C} \end{cases} \tag{11}$$

where N_s denotes the directional high-frequency response after content-adaptive modulation, while N_c captures complementary inter-channel chromatic variations. These two components are combined with fixed weights to obtain the noise prior $N_p \in [-0.25, 0.25]$:

$$N_p = \alpha_3(3N_s + N_c) \tag{12}$$

where $\alpha_3 = 0.06$. To further limit the influence of the noise prior in regions with prominent structures, the network computes the prior-injection weight W_p using the selective gate G and structural response G_s:

$$W_p = \alpha_2 \cdot \alpha_4 G \odot (1 - \alpha_4 G_s) \tag{13}$$

where $\alpha_4 = 0.5$. The weight W_p increases in regions with strong selective noise responses and weak structural responses, but decreases around prominent structures. The signed noise prior is then injected into the original image through weighted modulation. The resulting prior-modulated image I_p is defined as follows:

$$I_p = I + W_p \odot N_p \tag{14}$$

3.2.3. Noise Reconstruction and Denoising

Given the prior-modulated image I_p, a lightweight convolutional module estimates the residual noise. First, a 3×3 convolution maps I_p into 48 feature channels, followed by LeakyReLU with a negative slope of 0.2. Depthwise separable convolution then extracts spatial and channel information. A 3×3 depthwise convolution captures local spatial features, while a 1×1 pointwise convolution enables inter-channel interactions. Both are followed by LeakyReLU activation. A lightweight residual block further improves feature transformation through two successive depthwise–pointwise convolution pairs and a local residual connection. Finally, a 1×1 convolution predicts the residual noise N_r:

$$N_r = \mathrm{Conv}_{1\times1}\big(F(I_p) + \mathcal{R}(F(I_p))\big) \tag{15}$$

where $F(\cdot)$ denotes the feature extraction mapping consisting of the initial convolution and depthwise separable convolution, while $\mathcal{R}(\cdot)$ represents the residual branch. The denoised image I_{dn} is then obtained from the predicted residual noise N_r as follows:

$$I_{dn} = I_p - N_r \tag{16}$$

Through this process, MIDM aggregates directional information using deterministic bidirectional cumulative scanning and selectively modulates high-frequency states based on luminance, structural, and chromatic cues. This enables noise estimation and suppression while preserving useful structural information. Following the paired-downsampling strategy of ZERO-IG, MIDM processes the original low-light image and two complementary sub-images. The same noise modeling and reconstruction procedures are applied to all three inputs. The resulting denoised image I_{dn} is then passed to MSLM, and its multi-scale outputs are used to enforce a multi-scale consistency constraint. This design helps improve the stability of noise suppression under low-light conditions.

3.3. Multi-Scale Illumination Restoration Module

Low-light images often suffer from insufficient overall exposure and spatially uneven underexposure. Single-resolution convolutional processing makes it difficult to balance local texture preservation and global exposure consistency. To address this challenge, MSLM adopts a dual-branch, multi-scale architecture with local-detail and global-exposure

branches. As shown in Figure 3, the local branch captures fine textures at the original resolution, while the global branch reduces spatial resolution to expand the effective receptive field and model the overall exposure distribution. The complementary features are then combined for illumination restoration.

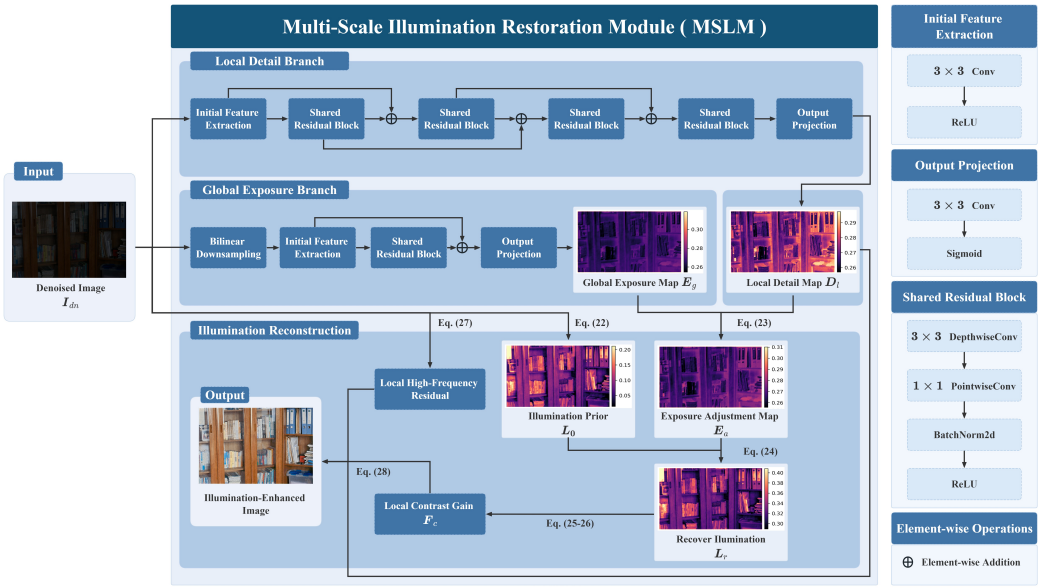


Figure 3. Overview of MSLM. The local-detail branch preserves texture information, while the global-exposure branch captures global illumination information. Their outputs are combined to restore the illumination component of the denoised input image.

3.3.1. Local Detail Branch

The denoised image $\boldsymbol{I}_{dn}$ is first mapped to a 64-channel feature space using a 3×3 standard convolution followed by ReLU activation, yielding the initial feature $\boldsymbol{F}_d^0$:

$$\boldsymbol{F}_d^0 = \mathrm{ReLU}(\mathrm{Conv}_{3\times3}(\boldsymbol{I}_{dn})) \tag{17}$$

The features then pass through three residual transformations with shared parameters. Each transformation consists of a 3×3 depthwise convolution, a 1×1 pointwise convolution, batch normalization, and ReLU activation. The shared residual transformation $\boldsymbol{\mathcal{B}}(\cdot)$ is defined as follows:

$$\boldsymbol{\mathcal{B}}(\boldsymbol{X}) = \mathrm{ReLU}(\mathrm{BN}(\mathrm{PWConv}_{1\times1}(\mathrm{DWConv}_{3\times3}(\boldsymbol{X})))) \tag{18}$$

The three residual updates are expressed as

$$\boldsymbol{F}_d^n = \boldsymbol{F}_d^{n-1} + \boldsymbol{\mathcal{B}}\left(\boldsymbol{F}_d^{n-1}\right), \qquad n = 1, 2, 3 \tag{19}$$

where n denotes the residual refinement stage index. All three stages share the same parameters, enabling successive refinement of full-resolution features without introducing stage-specific parameters. A 3×3 convolution followed by sigmoid activation then produces the local detail map $\boldsymbol{D}_l \in (0,1)^{3\times H\times W}$:

$$\boldsymbol{D}_l = \mathrm{Sigmoid}\left(\mathrm{Conv}_{3\times3}(\boldsymbol{F}_d^3)\right) \tag{20}$$

The sigmoid activation produces bounded adjustment maps through a smooth, differentiable transformation. The local detail map $\boldsymbol{D}_l$ retains high-frequency information at the

original resolution and provides spatial guidance for subsequent exposure adjustment and detail compensation.

3.3.2. Global Exposure Branch

The global exposure branch expands the effective receptive field by reducing spatial resolution, allowing the overall exposure distribution to be modeled at a lower computational cost. The input is first downsampled to half its original height and width using bilinear interpolation. Bilinear interpolation provides computationally efficient resizing without introducing learnable parameters, while ensuring smooth transitions between neighboring pixels. Initial features are extracted as in the local branch, but only one shared residual transformation is applied to limit excessive smoothing at low resolution. An output convolution followed by sigmoid activation produces a low-resolution exposure map. This map is then upsampled to the original resolution using bilinear interpolation, yielding the global exposure map $\boldsymbol{E}_g \in (0,1)^{3\times H\times W}$.

$$\begin{cases} \boldsymbol{F}_g^0 = \mathrm{ReLU}(\mathrm{Conv}_{3\times 3}(\mathcal{D}(\boldsymbol{I}_{dn}))) \\ \boldsymbol{E}_g = \mathcal{U}\Big(\mathrm{Sigmoid}\Big(\mathrm{Conv}_{3\times 3}\Big(\boldsymbol{F}_g^0 + \boldsymbol{\mathcal{B}}(\boldsymbol{F}_g^0)\Big)\Big)\Big) \end{cases} \tag{21}$$

where $\mathcal{D}(\cdot)$ denotes bilinear downsampling to half the original height and width, and $\mathcal{U}(\cdot)$ denotes bilinear upsampling to the original spatial resolution. Downsampling expands the effective receptive field, allowing $\boldsymbol{E}_g$ to capture global brightness information and guide subsequent illumination restoration and contrast adjustment.

3.3.3. Illumination Reconstruction

Illumination restoration must correct spatially nonuniform underexposure while preserving local structures. Global exposure correction alone may overlook fine details, whereas excessive local enhancement can amplify residual disturbances or over-brighten well-exposed regions. We therefore adopt a staged reconstruction strategy. An image-derived illumination prior establishes the brightness baseline, the two MSLM branches guide exposure recovery, and subsequent contrast and detail adjustments compensate for information attenuated during restoration. Given the denoised image $\boldsymbol{I}_{dn}$, the initial illumination prior $\boldsymbol{L}_0$ is estimated by combining the maximum and mean responses across RGB channels:

$$\boldsymbol{L}_0 = \mathcal{A}_3\Big[\beta_1 \max_c\Big(\boldsymbol{I}_{dn}^{(c)}\Big) + (1-\beta_1)\langle \boldsymbol{I}_{dn}\rangle_c\Big]. \tag{22}$$

where $\beta_1 = 0.6$, and $\langle\cdot\rangle_c$ denotes the per-pixel channel average. The channel-wise maximum captures the strongest brightness response, while the channel average reduces the influence of individual color channels. Their weighted combination is then smoothed by local average pooling to obtain a stable initial illumination prior. The exposure adjustment map $\boldsymbol{E}_a$ is defined as follows:

$$\boldsymbol{E}_a = \beta_1 \boldsymbol{E}_g + (1-\beta_1)\boldsymbol{D}_l. \tag{23}$$

A larger weight is assigned to $\boldsymbol{E}_g$ to capture the broader spatial context needed for nonuniform exposure correction. The local detail map $\boldsymbol{D}_l$ complements it with full-resolution spatial information. However, uniform exposure adjustment may over-brighten well-exposed regions. We therefore modulate the exposure correction based on the initial illumination:

$$\boldsymbol{L}_r = \boldsymbol{L}_0 + (1-\boldsymbol{L}_0)\odot \boldsymbol{E}_a \odot (1-0.3\boldsymbol{L}_0). \tag{24}$$

The factor $1 - \boldsymbol{L}_0$ emphasizes underexposed regions, while $1 - 0.3\boldsymbol{L}_0$ further reduces the correction as the initial illumination increases. Thus, for a fixed exposure response, the illumination adjustment decreases with increasing $\boldsymbol{L}_0$. Following Retinex theory, the initial reflectance $\boldsymbol{R}_0$ is estimated by dividing the image by the illumination prior $\boldsymbol{L}_0$ pixel-wise. The preliminary enhanced image $\boldsymbol{I}_0$ is then reconstructed by element-wise multiplication of $\boldsymbol{R}_0$ and the restored illumination $\boldsymbol{L}_r$:

$$\boldsymbol{R}_0 = \frac{\boldsymbol{I}_{dn}}{\boldsymbol{L}_0 + \varepsilon}, \qquad \boldsymbol{I}_0 = \boldsymbol{R}_0 \odot \boldsymbol{L}_r \tag{25}$$

Illumination recovery may change the contrast of the preliminary image. To correct local contrast variations without applying a uniform gain, the global exposure map $\boldsymbol{E}_g$ is used to regulate the local contrast gain $\boldsymbol{F}_c$:

$$\boldsymbol{F}_c = \langle \boldsymbol{I}_0 \rangle_s + \left[1 + 0.5\left(\boldsymbol{E}_g - 0.5\right)\right] \odot \left(\boldsymbol{I}_0 - \langle \boldsymbol{I}_0 \rangle_s\right), \tag{26}$$

where $\langle \cdot \rangle_s$ denotes the spatial mean computed separately for each channel. When $\boldsymbol{E}_g > 0.5$, the contrast factor exceeds one, increasing deviations from the spatial mean. When $\boldsymbol{E}_g < 0.5$, these deviations are reduced. Thus, contrast adjustment depends on the estimated exposure response rather than a fixed global gain. Since local smoothing and multi-scale processing may weaken fine details, we further extract the local high-frequency residual $\boldsymbol{F}_{loc}$ from the denoised image $\boldsymbol{I}_{dn}$:

$$\boldsymbol{F}_{loc} = \boldsymbol{I}_{dn} - \mathcal{A}_3(\boldsymbol{I}_{dn}) \tag{27}$$

To avoid uniform compensation of high-frequency responses, we use the local detail map $\boldsymbol{D}_l$ to apply a small, signed modulation to the residual. The modulated residual is then combined with the contrast-adjusted result to obtain the illumination-restored output $\boldsymbol{L}_f$:

$$\boldsymbol{L}_f = \boldsymbol{F}_c + 0.1(\boldsymbol{D}_l - 0.5) \odot \boldsymbol{F}_{loc} \tag{28}$$

The coefficient limits the high-frequency residual contribution, while $\boldsymbol{D}_l$ controls its spatial modulation. The resulting three-channel illumination-restored output $\boldsymbol{L}_f$ is clipped to the valid intensity range before further processing. Thus, the reconstruction combines exposure recovery, contrast adjustment, and selective detail compensation rather than relying solely on the initial illumination estimate.

3.4. Structure-Aware Refinement Module

After multi-scale illumination restoration by MSLM, the reflectance component $\boldsymbol{R}_f$ may still contain residual noise. Genuine edges and fine textures can therefore be difficult to distinguish from high-frequency disturbances. Uniform enhancement or suppression of these responses may remove useful details or amplify noise. As shown in Figure 4, SRM estimates local structural and relative high-frequency noise responses using multi-scale statistics of the reflectance component. It then constructs complementary texture and noise gates to refine the final output.

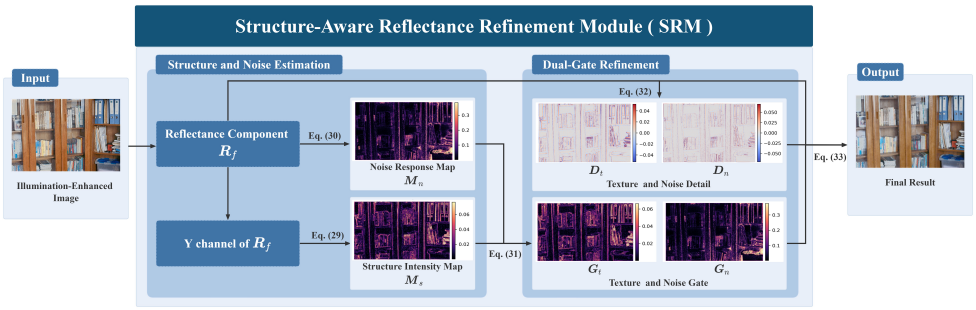


Figure 4. Overview of SRM. SRM estimates a noise response map from the reflectance component of the illumination-restored image and a structural strength map from its grayscale representation. It then combines texture and noise gates with their corresponding detail terms to produce the final output.

3.4.1. Structure and Noise Estimation

Genuine edges and textures tend to exhibit stable structural differences across smoothing scales, while local averaging partially suppresses random high-frequency disturbances. Accordingly, SRM first extracts the Y channel $\boldsymbol{Y}_r$ from the reflectance component $\boldsymbol{R}_f$ before refinement. SRM uses this grayscale reflectance intensity to estimate structural strength without processing the three color channels independently. This provides a common spatial reference for identifying structural variations across different smoothing scales. Size-preserving average pooling with 3×3 and 7×7 kernels is then applied separately. The absolute difference between the resulting maps defines the structure intensity map:

$$\boldsymbol{M}_s = |\mathcal{A}_3(\boldsymbol{Y}_r) - \mathcal{A}_7(\boldsymbol{Y}_r)| \tag{29}$$

A larger value of $\boldsymbol{M}_s$ indicates a greater difference between smoothing scales, suggesting a prominent edge or texture. To characterize local high-frequency disturbances, SRM computes the pixel-wise residual between $\boldsymbol{R}_f$ and its 3×3 average-pooled response. Since the absolute residual may depend on local intensity, it is normalized by the 5×5 average-pooled response to obtain the relative noise response map:

$$\boldsymbol{M}_n = \frac{\left|\boldsymbol{R}_f - \mathcal{A}_3(\boldsymbol{R}_f)\right|}{\mathcal{A}_5(\boldsymbol{R}_f) + \varepsilon} \tag{30}$$

This normalization uses the local average response as a reference, making high-frequency residuals more comparable across regions with different brightness and texture levels. It also provides a basis for the subsequent selective modulation of structural and noise responses.

3.4.2. Dual-Gate Refinement

Based on the structure intensity map $\boldsymbol{M}_s$ and noise response map $\boldsymbol{M}_n$, SRM constructs a texture gate $\boldsymbol{G}_t$ and a noise gate $\boldsymbol{G}_n$. The texture gate emphasizes regions with strong structural responses and weak noise responses, while the noise gate highlights regions with strong high-frequency disturbances and weak structural responses. The two gates are defined as follows:

$$\begin{cases} \boldsymbol{G}_t = \boldsymbol{M}_s \odot (\boldsymbol{1} - \boldsymbol{M}_n) \\ \boldsymbol{G}_n = \boldsymbol{M}_n \odot (\boldsymbol{1} - \boldsymbol{M}_s) \end{cases} \tag{31}$$

When structural responses dominate, the texture gate $\boldsymbol{G}_t$ assigns greater weight to texture compensation. When high-frequency disturbances are strong but structural responses are weak, the noise gate $\boldsymbol{G}_n$ emphasizes noise suppression. These complementary gates guide adaptive refinement based on local structure and noise responses, limiting excessive

processing in ambiguous regions. SRM then extracts two detail terms from the multi-scale smoothed reflectance component. The texture detail term $\boldsymbol{D}_t$ is computed as the difference between the 3×3 and 5×5 smoothed responses, capturing intermediate-scale structural variations. The noise detail term $\boldsymbol{D}_n$ is obtained by subtracting the 3×3 smoothed response from the original reflectance, capturing pixel-level high-frequency variations:

$$\begin{cases} \boldsymbol{D}_t = \mathcal{A}_3(\boldsymbol{R}_f) - \mathcal{A}_5(\boldsymbol{R}_f) \\ \boldsymbol{D}_n = \boldsymbol{R}_f - \mathcal{A}_3(\boldsymbol{R}_f) \end{cases} \tag{32}$$

Finally, the texture gate controls the addition of structural details, while the noise gate regulates the subtraction of high-frequency residuals. The final result is expressed as follows:

$$\boldsymbol{I}_{\text{final}} = \boldsymbol{R}_f + \boldsymbol{G}_{\text{t}} \odot \boldsymbol{D}_{\text{t}} - \boldsymbol{G}_{\text{n}} \odot \boldsymbol{D}_{\text{n}} \tag{33}$$

In regions with prominent structures, $\boldsymbol{G}_t \odot \boldsymbol{D}_t$ moderately compensates for medium-scale textures and edges. Where high-frequency disturbances are strong but structural responses are weak, $\boldsymbol{G}_n \odot \boldsymbol{D}_n$ suppresses local high-frequency disturbances. When both responses are weak, the correction terms have little effect, largely preserving the original structure. Through this dual-gate refinement, SRM balances preserving structural detail and suppressing residual noise. Zero-MS uses the recovered reflectance as its final output, avoiding reintroducing nonuniform illumination while preserving intrinsic scene information.

3.5. Loss Function

The network is optimized end-to-end using a composite unsupervised loss. All training signals are derived from the input low-light image and its paired downsampled observations. The total loss $\mathcal{L}_{\text{total}}$ consists of four terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{den}} + \mathcal{L}_{\text{ill}} + \mathcal{L}_{\text{ref}} + \mathcal{L}_{\text{perc}} \tag{34}$$

The low-light denoising loss $\mathcal{L}_{\text{den}}$, illumination estimation loss $\mathcal{L}_{\text{ill}}$, and reflectance denoising loss $\mathcal{L}_{\text{ref}}$ follow the formulations in ZERO-IG. Together, they constrain denoising consistency, illumination restoration, and structural preservation. Zero-MS further introduces a perceptual loss $\mathcal{L}_{\text{perc}}$, applied to the final reflectance $\boldsymbol{R}_{\text{ref}}$ and the reflectance maps obtained from paired downsampled sub-images. This loss combines four constraints on natural-scene statistics, noise in flat dark regions, pixel clipping, and saturation in dark regions:

$$\mathcal{L}_{\text{perc}} = \mathcal{L}_{\text{mscn}} + \omega_1 \mathcal{L}_{\text{noise}} + \omega_2 \mathcal{L}_{\text{clip}} + \omega_3 \mathcal{L}_{\text{sat}} \tag{35}$$

where $\omega_1 = 0.8$, $\omega_2 = 0.2$, $\omega_3 = 0.1$. The natural-scene-statistics constraint $\mathcal{L}_{\text{mscn}}$ uses mean-subtracted, contrast-normalized (MSCN) coefficients, which are widely used in image quality assessment. The luminance component $\boldsymbol{Y}_{rm}$ of the evaluated image is normalized by its local mean and standard deviation computed within a 7×7 neighborhood:

$$\text{MSCN}(\boldsymbol{Y}_{rm}) = \frac{\boldsymbol{Y}_{rm} - \mathcal{A}_7(\boldsymbol{Y}_{rm})}{\sqrt{\mathcal{A}_7\left[(\boldsymbol{Y}_{rm} - \mathcal{A}_7(\boldsymbol{Y}_{rm}))^2\right]} + \varepsilon} \tag{36}$$

The MSCN coefficients of natural images typically have a mean near zero and a standard deviation near one. Based on this property, $\mathcal{L}_{\text{mscn}}$ constrains their mean, standard deviation, skewness, and kurtosis:

$$\mathcal{L}_{\text{mscn}} = (\mu)^2 + 0.5(\sigma - 1)^2 + 0.05|\gamma_1| + 0.02|\gamma_2 - 3| \tag{37}$$

where μ, σ, γ_1, and γ_2 denote the mean, standard deviation, skewness, and kurtosis of the MSCN coefficients, respectively. This term penalizes deviations from natural-image statistics and helps suppress spurious textures. The flat dark-region noise constraint $\mathcal{L}_{\text{noise}}$ mainly targets low-luminance regions with weak structural responses. The pixel-level high-frequency residual is represented by $\boldsymbol{Y}_{rm} - \mathcal{A}_3(\boldsymbol{Y}_{rm})$, while the difference between the 3×3 and 7×7 smoothed responses captures mid-scale structural variations. The term $1 - \boldsymbol{Y}_{rm}$ emphasizes dark regions, while the weak-structure weight reduces the contribution of prominent structural regions:

$$\mathcal{L}_{\text{noise}} = \langle |\boldsymbol{Y}_{rm} - \mathcal{A}_3(\boldsymbol{Y}_{rm})| \odot (1 - \boldsymbol{Y}_{rm}) \odot [1 - 8|\mathcal{A}_3(\boldsymbol{Y}_{rm}) - \mathcal{A}_7(\boldsymbol{Y}_{rm})|] \rangle \tag{38}$$

This constraint penalizes local high-frequency fluctuations more strongly in dark regions, while weak-structure weighting limits its effect on genuine edges and textures. The pixel-clipping constraint $\mathcal{L}_{\text{clip}}$ discourages large areas with intensity values near the valid range limits. A quadratic penalty is applied only to luminance values below 0.02 or above 0.98:

$$\mathcal{L}_{\text{clip}} = \text{mean}\Big(\text{ReLU}(0.02 - \boldsymbol{Y}_{rm})^2\Big) + \text{mean}\Big(\text{ReLU}(\boldsymbol{Y}_{rm} - 0.98)^2\Big) \tag{39}$$

The dark-region saturation constraint $\mathcal{L}_{\text{sat}}$ computes pixel-wise saturation Sat as the difference between the maximum and minimum RGB channel values. It penalizes saturation values above $\theta = 0.55$ in dark regions, with $1 - \boldsymbol{Y}_{rm}$ weighting the penalty according to local luminance:

$$\begin{cases} Sat = \frac{\max_C(\boldsymbol{Y}^C_{rm}) - \min_C(\boldsymbol{Y}^C_{rm})}{\max_C(\boldsymbol{Y}^C_{rm}) + \epsilon} \\ \mathcal{L}_{\text{sat}} = \text{mean}\big(\text{ReLU}(Sat - \theta)^2 \odot (1 - \boldsymbol{Y}_{rm})\big) \end{cases} \tag{40}$$

Thus, this constraint primarily suppresses excessive chromatic responses in dark regions while having little effect on well-lit regions. Overall, the proposed perceptual loss combines complementary reflectance constraints, including natural-scene statistics, dark-region noise, dynamic-range clipping, and dark-region saturation. Together with the low-light image denoising, illumination estimation, and reflectance denoising losses, it supports joint optimization of illumination recovery, structure preservation, noise suppression, and perceptual quality without reference images.

4. Experimental Result Analysis

4.1. Experimental Settings

Datasets and Compared Methods: To evaluate Zero-MS under real-world low-light conditions, we train the model on LOL-v1 [10] and test it on LOL-v1, LOL-v2-Real [32], and NPE [33]. LOL-v1 contains 500 pairs of low-light and normal-light images captured in real scenes, including 485 training pairs and 15 test pairs. The low-light images were mainly obtained by adjusting exposure time and ISO settings. LOL-v2-Real contains 689 training pairs and 100 test pairs, covering more diverse scenes and complex real-world degradations. NPE is an unpaired test dataset containing real images with spatially nonuniform illumination and no corresponding normal-light reference images. Together, these datasets cover a range of scenes and illumination conditions. Their main characteristics are summarized in Table 1.

Table 1. Characteristics of the low-light datasets used in this study.

Dataset	Type	Pairing	Low-Light Characteristics	Role
LOL-v1	Real scenes	Paired	Low brightness, exposure variation, real-world noise	Training & testing
LOL-v2-Real	Real scenes	Paired	Diverse illumination and complex real-world degradation	Testing
NPE	Real scenes	Unpaired	Low brightness and nonuniform illumination	Testing

The comparison includes conventional methods, namely CLAHE [5], MSRCR [6], LIME [7], NPLIE [8], and Jeon et al.'s method [9], as well as deep learning-based methods, including Zero-DCE [15], Zero-DCE++ [16], ZRCRN [17], ZERO-IG [4], RetinexNet [10], Pie [11], CoLIE [18], and DI-Retinex [12]. All experiments and metric calculations were performed using Python 3.12 and PyTorch 2.5.1 with CUDA 12.1 support. PyCharm Community Edition 2024.3.1.1 was used as the development environment. For fair comparison, each baseline is evaluated using the original authors' recommended settings and available pretrained models under the same experimental environment as Zero-MS. **Evaluation Metrics**: We evaluate Zero-MS in terms of information fidelity, visual quality, noise suppression, and model complexity. Peak signal-to-noise ratio (PSNR) [34], structural similarity index measure (SSIM) [35], and CIEDE2000 color difference [36] assess pixel fidelity, structural similarity, and color consistency between enhanced and reference images, respectively. The Perception-based Image Quality Evaluator (PIQE) [37] and Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [38] assess perceptual quality and image naturalness. The estimated noise standard deviation $\hat{\sigma}_n$ [39], residual noise standard deviation σ_{R_n} [40], and mean Noise Power Spectrum (NPS) power $\overline{P}_{\mathrm{NPS}}$ [41] measure the noise level, residual noise, and average noise power, respectively. Parameter count (Params.), model size, and mean runtime are used to assess model complexity, storage requirements, and inference efficiency, respectively. Image quality metrics supported by PyIQA are computed using the pyiqa library, while the remaining metrics and noise statistics are implemented in Python according to their definitions. **Implementation Settings**: All experiments were conducted on a Lenovo ThinkPad T14p laptop (Lenovo Group Limited, Beijing, China) running Windows 11 Home (version 25H2; Microsoft Corporation, Redmond, WA, USA) and equipped with an Intel® Core™ Ultra 9 185H processor (2.30 GHz; Intel Corporation, Santa Clara, CA, USA) and an NVIDIA GeForce RTX 4050 Laptop GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 6 GB of dedicated memory. The initial learning rate was set to 3×10^{-4}, with a batch size of 4. The model was trained for 2,000 epochs. Training and evaluation were conducted under the same hardware and software environment to ensure consistency and reproducibility.

4.2. Comparison with State-of-the-Art Methods

4.2.1. Quantitative Comparison

We compare Zero-MS with 13 representative low-light image enhancement methods on the LOL-v1, LOL-v2-Real, and NPE datasets to assess information fidelity, visual quality, and denoising performance. In the tables, ↑ and ↓ indicate that higher and lower values are better, respectively. The top three results are highlighted in bold red, blue, and brown, respectively. As shown in Table 2, Zero-MS ranks first on seven of the eight LOL-v1 metrics. Its PSNR and SSIM reach 18.6513 dB and 0.7918, outperforming the second-best DI-Retinex by 9.35% and 3.68%, respectively. These results indicate improved reconstruction fidelity and structural similarity. Zero-MS also achieves the lowest CIEDE2000 value of 0.0874, 12.77% lower than DI-Retinex, indicating better color consistency with the reference images. For perceptual quality, Zero-MS obtains the best PIQE and BRISQUE scores of 4.3103 and 17.0363, improving upon the second-best results by 8.16% and 14.13%, respectively. For denoising performance, $\hat{\sigma}_n$ and $\overline{P}_{\mathrm{NPS}}$ reach 2.4504 and 6.2530, respectively, both ranking

first. Its σ_{R_n} is 4.6650, ranking third and differing from the second-ranked DI-Retinex by only 1.09%. As shown in Table 3, Zero-MS ranks first across all eight metrics on the LOL-v2-Real dataset. It achieves a PSNR of 17.2403 dB and an SSIM of 0.6611, exceeding DI-Retinex by 8.51% and 7.62%, respectively. Its CIEDE2000 score of 0.1068 is 12.96% lower than that of DI-Retinex, indicating better color fidelity. For perceptual quality, Zero-MS obtains PIQE and BRISQUE scores of 6.1294 and 32.2282, improving on the second-best results by 48.39% and 1.26%, respectively. For denoising performance, $\hat{\sigma}_n$, σ_{R_n}, and $\overline{P}_{\mathrm{NPS}}$ reach 4.0418, 5.9119, and 20.4770, respectively, all ranking first. These results demonstrate Zero-MS's overall effectiveness in image reconstruction, structural preservation, perceptual quality, and noise suppression. The NPE dataset contains real low-light images with nonuniform illumination. Because reference images are unavailable, we evaluate no-reference visual quality and denoising performance. As shown in Table 4, Zero-MS ranks among the top three methods across all five metrics. It achieves PIQE and BRISQUE scores of 12.0764 and 13.6440, ranking third and second, respectively. For denoising performance, its estimated noise standard deviation $\hat{\sigma}_n$ is 1.8886, ranking second and 22.43% higher than that of the top-ranked Zero-DCE++. Zero-MS achieves the lowest residual noise standard deviation σ_{R_n} of 6.8890, 4.46% lower than the second-ranked MSRCR. Its mean NPS power $\overline{P}_{\mathrm{NPS}}$ is 2.6555, ranking third. These results indicate that Zero-MS maintains a favorable balance between perceptual quality and noise suppression under nonuniform illumination, with relatively low residual noise. Overall, Zero-MS achieves strong information fidelity on the two full-reference datasets, LOL-v1 and LOL-v2-Real. It also maintains favorable visual quality and denoising performance across all three datasets. To further examine how different low-light enhancement methods handle noise, we use LOL-v1 as a representative benchmark. The analysis covers dark-region high-frequency residuals, local residual fluctuations, luminance-dependent residuals, the noise–structure trade-off, and the enhancement–noise trade-off. Here, high-frequency residuals refer to image-domain high-frequency disturbances in enhanced images rather than specific components of physical sensor noise. As shown in Figure 5a, the methods differ considerably in their control of dark-region high-frequency disturbances. Jeon et al., RetinexNet, ZERO-IG, and MSRCR generally exhibit higher residuals. Jeon et al.'s method has a median near 0.03 and a relatively wide distribution. CoLIE, NPLIE, Zero-DCE, Zero-DCE++, and ZRCRN fall within the middle range, while CLAHE, DI-Retinex, and LIME show lower residuals. Zero-MS achieves a median of approximately 0.008–0.009 with a compact distribution, indicating relatively low residuals and limited variation across test images. A similar pattern appears in Figure 5b. Jeon et al., ZERO-IG, and RetinexNet show larger local residual fluctuations, whereas Zero-MS maintains a median of approximately 0.012–0.013 with a narrower distribution. These results indicate that Zero-MS provides relatively stable control of high-frequency disturbances in dark regions and local residual fluctuations. As shown in Figure 5c, CLAHE and Pie exhibit relatively high residuals across several luminance intervals, particularly at higher luminance levels. Zero-DCE++ also shows some dependence on luminance. LIME maintains low residuals overall, while the residual of DI-Retinex decreases rapidly with increasing luminance, approaching zero above approximately 0.55. However, a very low high-frequency response alone cannot distinguish noise suppression from the loss of fine details. In comparison, Zero-MS maintains low and relatively stable residuals across low-to-medium luminance levels without a noticeable increase in high-frequency disturbances. Because excessive smoothing can also reduce high-frequency responses, a low residual does not necessarily indicate better enhancement. Therefore, Figure 5d examines dark-region residuals alongside structure preservation. Zero-MS achieves a relatively low residual while maintaining a high structure-preservation score. Compared with DI-Retinex, it has a slightly higher residual but better

structure preservation. Compared with ZERO-IG, RetinexNet, and Jeon et al., it produces lower residuals while retaining a favorable structure score. Although LIME achieves both a lower residual and a higher structure-preservation score, Zero-MS still maintains low residuals without evident structural degradation. These findings suggest a favorable balance between controlling high-frequency disturbances and preserving structure. Limited brightness enhancement can also result in low residuals.

Table 2. Quantitative results on the LOL-v1 dataset. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	Information Fidelity			Visual Quality		Denoising Performance		
	PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\text{NPS}}$ ↓
CLAHE [5]	9.5831	0.4091	0.2107	**5.7630**	22.4353	3.2693	**4.5251**	**6.6742**
MSRCR [6]	15.3270	0.6560	0.1235	11.5596	28.0863	7.2302	9.2861	20.5511
LIME [7]	15.8091	**0.7609**	0.1173	10.9505	**19.8405**	3.1737	8.9368	11.8773
Jeon et al. [9]	**15.8813**	0.4767	0.1272	20.5806	45.8188	13.5549	17.0641	95.1148
NPLIE [8]	12.2643	0.5475	0.1589	11.2561	29.4081	5.4591	7.5279	18.1817
ZRCRN [17]	15.0900	0.6244	0.1232	11.4221	27.8511	5.7701	7.7206	17.2167
ZERO-IG [4]	15.3517	0.5309	0.1278	18.0261	41.6178	9.7309	12.9538	57.4309
Zero-DCE [15]	14.1621	0.6106	0.1318	11.5056	29.9345	5.6498	7.3310	18.3755
Zero-DCE++ [16]	14.0932	0.5262	0.1334	7.2786	**19.9397**	**3.0242**	6.3374	10.0014
RetinexNet [10]	15.7416	0.5448	**0.1127**	20.7505	42.9578	10.2056	11.1633	45.4369
Pie [11]	10.9588	0.5101	0.1781	8.4620	25.0714	4.0555	5.8206	10.2966
DI-Retinex [12]	**16.9067**	**0.7627**	**0.1002**	**4.6935**	22.1282	**2.8508**	**4.6146**	**6.3317**
CoLIE [18]	12.8364	0.5462	0.1543	11.3177	29.4730	5.6864	7.6368	22.5803
Zero-MS	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**	**2.4504**	**4.6650**	**6.2530**

Table 3. Quantitative results on the LOL-v2-Real dataset. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

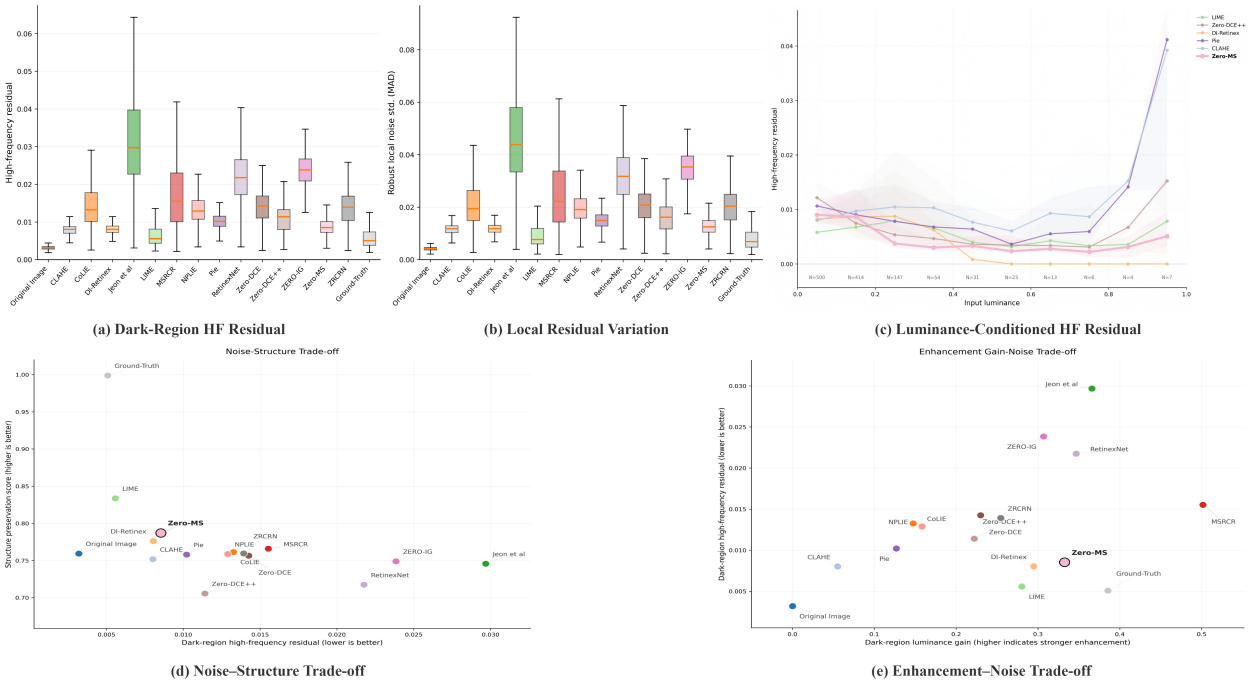
Methods	Information Fidelity			Visual Quality		Denoising Performance		
	PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\text{NPS}}$ ↓
CLAHE [5]	12.0536	0.3846	0.1865	22.4440	50.4241	16.5950	16.9635	263.7445
MSRCR [6]	14.0580	0.5188	0.1499	18.3699	37.4947	12.3736	13.3296	88.5480
LIME [7]	14.4202	**0.5861**	0.1535	18.2577	**33.4342**	10.5326	13.9811	143.0365
Jeon et al. [9]	14.3681	0.3664	0.1632	27.6070	54.2559	19.4381	20.9075	229.6046
NPLIE [8]	11.9850	0.4253	0.1774	19.2719	40.3574	10.7299	11.4899	109.1455
ZRCRN [17]	13.8512	0.4813	0.1576	18.8534	40.2548	11.5839	12.1474	108.3774
ZERO-IG [4]	13.3520	0.4020	0.1809	25.3910	51.1564	19.3731	20.3563	345.1712
Zero-DCE [15]	13.4158	0.4718	0.1556	19.1918	41.5718	10.4810	10.9959	95.5221
Zero-DCE++ [16]	13.8725	0.4112	0.1403	**13.7994**	34.8750	**6.8770**	9.6489	**63.8539**
RetinexNet [10]	**14.7420**	0.4220	**0.1305**	28.1725	52.6804	16.1986	15.5762	176.3323
Pie [11]	11.0645	0.3977	0.1872	16.6554	36.9293	8.3735	**9.2326**	73.6547
DI-Retinex [12]	**15.7740**	**0.6107**	**0.1227**	**11.8753**	**32.6411**	**6.5344**	**7.4781**	**53.6183**
CoLIE [18]	12.4628	0.4260	0.1741	19.1865	40.4989	11.1432	11.7994	129.5286
Zero-MS	**17.2403**	**0.6611**	**0.1068**	**6.1294**	**32.2282**	**4.0418**	**5.9119**	**20.4770**

Therefore, Figure 5e further considers dark-region luminance gain. CLAHE and Pie produce low residuals but offer limited brightness recovery. LIME and DI-Retinex maintain low residuals while providing greater luminance gains. ZERO-IG, RetinexNet, and Jeon et al. achieve stronger enhancement but exhibit higher high-frequency residuals. Zero-MS provides a dark-region luminance gain of approximately 0.33 with a residual near 0.009. Although its residual is slightly higher than those of LIME and DI-Retinex, it achieves

greater brightness recovery. Compared with ZERO-IG, RetinexNet, Jeon et al., and MSRCR, which provide comparable or greater luminance gains, Zero-MS maintains substantially lower residuals. Overall, these results show that Zero-MS achieves a favorable balance among dark-region brightness recovery, high-frequency disturbance control, and structure preservation. Its low residuals, together with its structure-preservation performance, suggest that noise suppression is not achieved solely through excessive smoothing.

Table 4. Quantitative results on the NPE dataset. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	Visual Quality		Denoising Performance		
	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\mathrm{NPS}}$ ↓
CLAHE [5]	14.5490	23.4263	2.8783	9.8481	12.2276
MSRCR [6]	12.2045	19.3031	2.1164	**7.2109**	**0.7985**
LIME [7]	14.6989	20.1091	2.8976	13.5124	7.9475
Jeon et al. [9]	12.2578	**15.3135**	3.1648	12.1476	4.9068
NPLIE [8]	12.1418	17.4079	2.9371	11.2903	6.5845
ZRCRN [17]	14.0335	21.8698	2.5151	9.4575	4.5302
ZERO-IG [4]	12.6498	26.2115	3.9067	14.2973	6.0745
Zero-DCE [15]	13.0088	23.0987	2.5198	10.2996	5.4996
Zero-DCE++ [16]	12.6710	21.6166	**1.5426**	**8.2163**	3.5749
RetinexNet [10]	**11.1549**	**12.5517**	3.9531	12.6122	8.3565
Pie [11]	13.6224	24.4788	2.5164	10.7490	7.4119
DI-Retinex [12]	14.9845	43.2812	**2.0274**	8.8064	**1.3164**
CoLIE [18]	**11.6032**	17.8360	2.2818	8.9772	4.9297
Zero-MS	**12.0764**	**13.6440**	**1.8886**	**6.8890**	**2.6555**



(a) Dark-Region HF Residual (b) Local Residual Variation (c) Luminance-Conditioned HF Residual

(d) Noise–Structure Trade-off (e) Enhancement–Noise Trade-off

Figure 5. Analysis of noise-related characteristics and enhancement trade-offs on the LOL-v1 dataset. (**a**) Dark-region high-frequency residual distribution. (**b**) Local residual variation distribution. (**c**) Luminance-conditioned high-frequency residual. Solid curves represent the median residual in each bin, while shaded regions indicate the interquartile range (25th–75th percentiles) across valid test images. N denotes the number of valid images contributing to each bin. Lower residual values

indicate weaker image-domain high-frequency fluctuations. (**d**) Noise–structure preservation trade-off. (**e**) Enhancement–noise trade-off. The compared methods include CLAHE [5], MSRCR [6], LIME [7], NPLIE [8], Jeon et al. [9], Zero-DCE [15], Zero-DCE++ [16], ZRCRN [17], ZERO-IG [4], RetinexNet [10], Pie [11], CoLIE [18], and DI-Retinex [12]. Please zoom in for clearer visualization.

4.2.2. Qualitative Comparison

Although quantitative metrics help evaluate low-light image enhancement, they do not fully capture perceptual quality. We therefore compare the visual results of Zero-MS with competing methods on the LOL-v1, LOL-v2-Real, and NPE datasets. As shown in Figure 6, CLAHE, NPLIE, Pie, Zero-DCE, and Zero-DCE++ improve visibility in some dark regions but leave much of the image underexposed. MSRCR increases brightness but introduces noticeable color shifts and reduces contrast. RetinexNet shows color distortion and a slightly washed-out appearance. LIME, Jeon et al., and ZERO-IG provide stronger brightness enhancement, although some yellow and red regions appear overly saturated compared with the ground truth. DI-Retinex achieves relatively uniform exposure but produces a somewhat dull result. CoLIE and ZRCRN retain recognizable structural details, although some shadows remain dark. In contrast, Zero-MS improves dark-region visibility without obvious overexposure in bright areas. The toy, color chart, and right-side textures are more clearly revealed, with relatively natural transitions between bright and dark regions. However, some dark areas remain slightly under-enhanced, and minor color differences persist compared with the ground truth. The corresponding vectorscopes further illustrate differences in chroma distribution. NPLIE, Pie, Zero-DCE++, and DI-Retinex show relatively concentrated distributions, while CLAHE exhibits a broader spread in some chromatic directions. In contrast, LIME, Jeon et al., RetinexNet, and ZERO-IG achieve mean chroma values of 25.30, 27.09, 27.24, and 26.84, respectively, exceeding the ground-truth value of 17.17. Although MSRCR's mean chroma is close to the reference, its mean V value of -11.83 differs substantially from the ground-truth value of 9.78. Zero-MS achieves mean U and V values of -6.42 and 10.21, close to the reference values of -8.81 and 9.78. Its mean and median chroma values are 14.78 and 10.44, compared with 17.17 and 13.34 for the ground truth. Its chroma standard deviation is also lower, indicating a narrower distribution. Overall, the visual comparison and vectorscopes suggest that Zero-MS improves exposure while maintaining relatively consistent color characteristics, although its chroma distribution remains narrower than that of the reference image. As shown in Figure 7, CLAHE, NPLIE, and Pie provide insufficient enhancement. MSRCR and Zero-DCE++ introduce noticeable color shifts, while Jeon et al. and ZERO-IG improve brightness but overexpose and oversaturate. RetinexNet produces oversaturated vegetation. CoLIE, Zero-DCE, and ZRCRN yield relatively stable results, although their illumination or contrast recovery remains limited. LIME and DI-Retinex achieve more natural exposure. However, LIME shows excessive saturation and less-smooth sky regions, while DI-Retinex introduces visible noise. In contrast, Zero-MS produces smoother transitions in the sky and building regions without noticeable color shifts. Its overall tone is also closer to the reference image. The enlarged regions reveal grainy artifacts in the results of MSRCR, CLAHE, Jeon et al., NPLIE, RetinexNet, Pie, ZERO-IG, Zero-DCE, and CoLIE. These artifacts are less apparent in Zero-MS, which avoids noticeable noise amplification and over-sharpening. Overall, Zero-MS achieves a favorable balance among illumination recovery, color preservation, and noise suppression. As shown in Figure 8, the lower panels show spatial brightness maps derived from the enhanced images using the same guided-filtering procedure, rather than method-specific illumination estimates. CLAHE, NPLIE, Pie, and CoLIE leave some shadowed regions relatively dark, while MSRCR produces strong overall brightening but reduces contrast. ZRCRN and RetinexNet show noticeable

color or tonal shifts. DI-Retinex and ZERO-IG produce extensive bright regions in the sky and around light sources, with reduced tonal variation. In comparison, Zero-MS improves vehicle and road visibility while preserving structural details and moderate brightness transitions. Its brightness map also shows less extensive high-brightness regions than methods with pronounced over-enhancement. These results suggest that Zero-MS achieves a reasonable balance between shadow enhancement and highlight preservation in the illustrated scene.

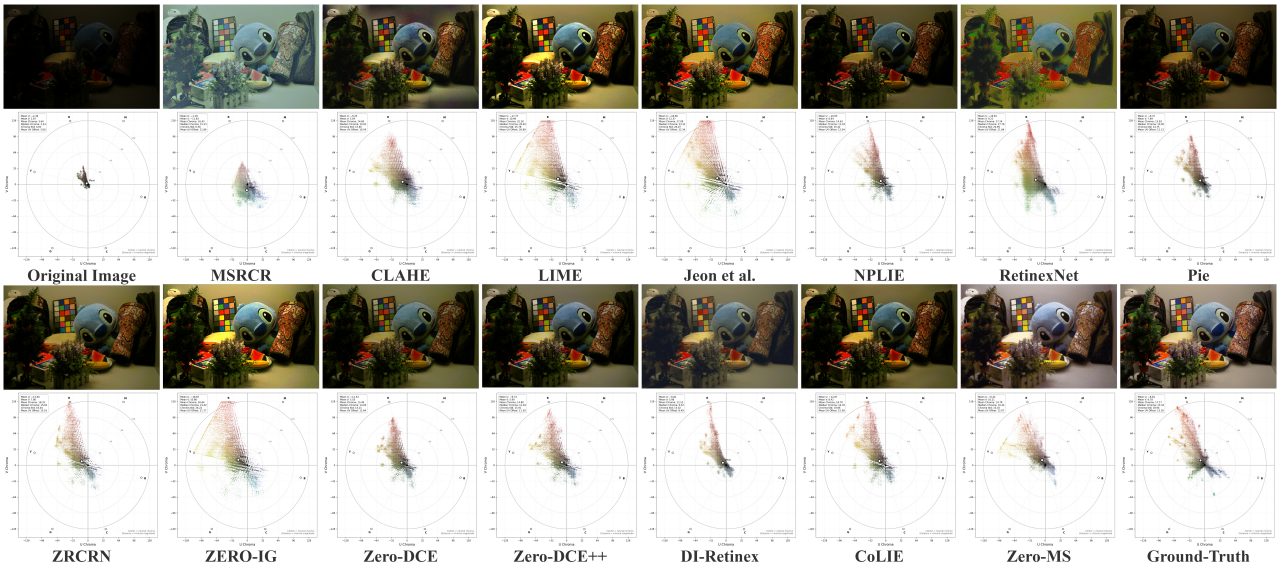


Figure 6. Qualitative comparison of Zero-MS and other methods on the LOL-v2-Real dataset. For each method, the enhanced image is shown in the upper panel, with the corresponding vectorscope in the lower panel. The vectorscope illustrates color direction and chroma strength through the U–V distribution. Points closer to the center indicate lower chroma, while those farther away indicate higher chroma. An overall shift in the distribution may suggest a color bias. Mean U/V values describe the average chroma direction. Mean Chroma and Median Chroma measure chroma strength, while Chroma Std indicates the distribution's spread. Please zoom in for clearer visualization.

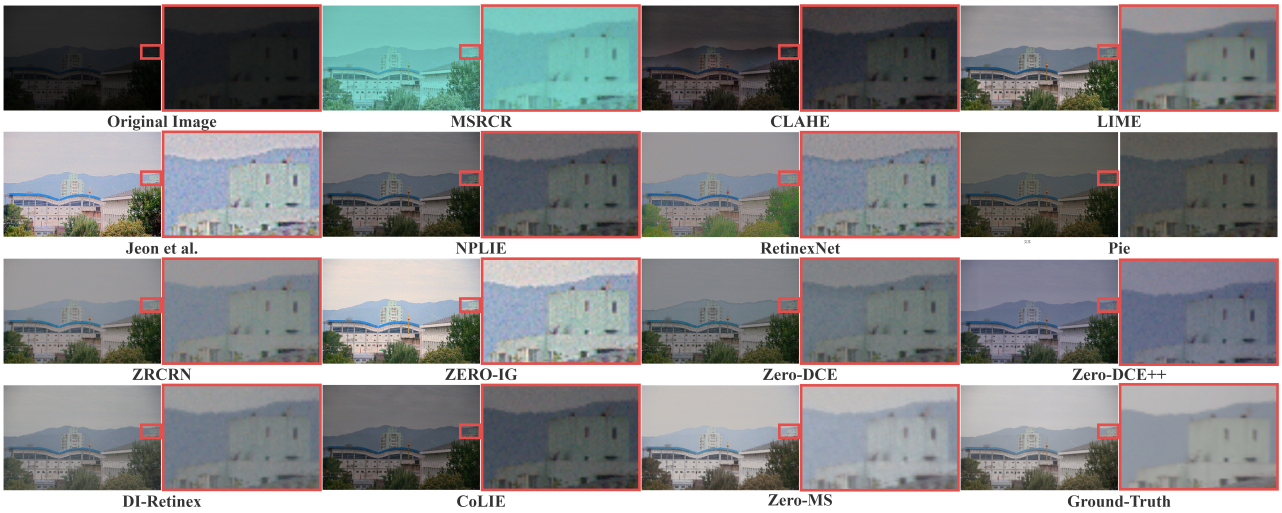


Figure 7. Qualitative comparisons between Zero-MS and other methods on the LOL-v1 dataset. For each method, the left panel shows the full, enhanced image, while the right panel provides a magnified view of the region marked by a red box. Please zoom in for better visualization.

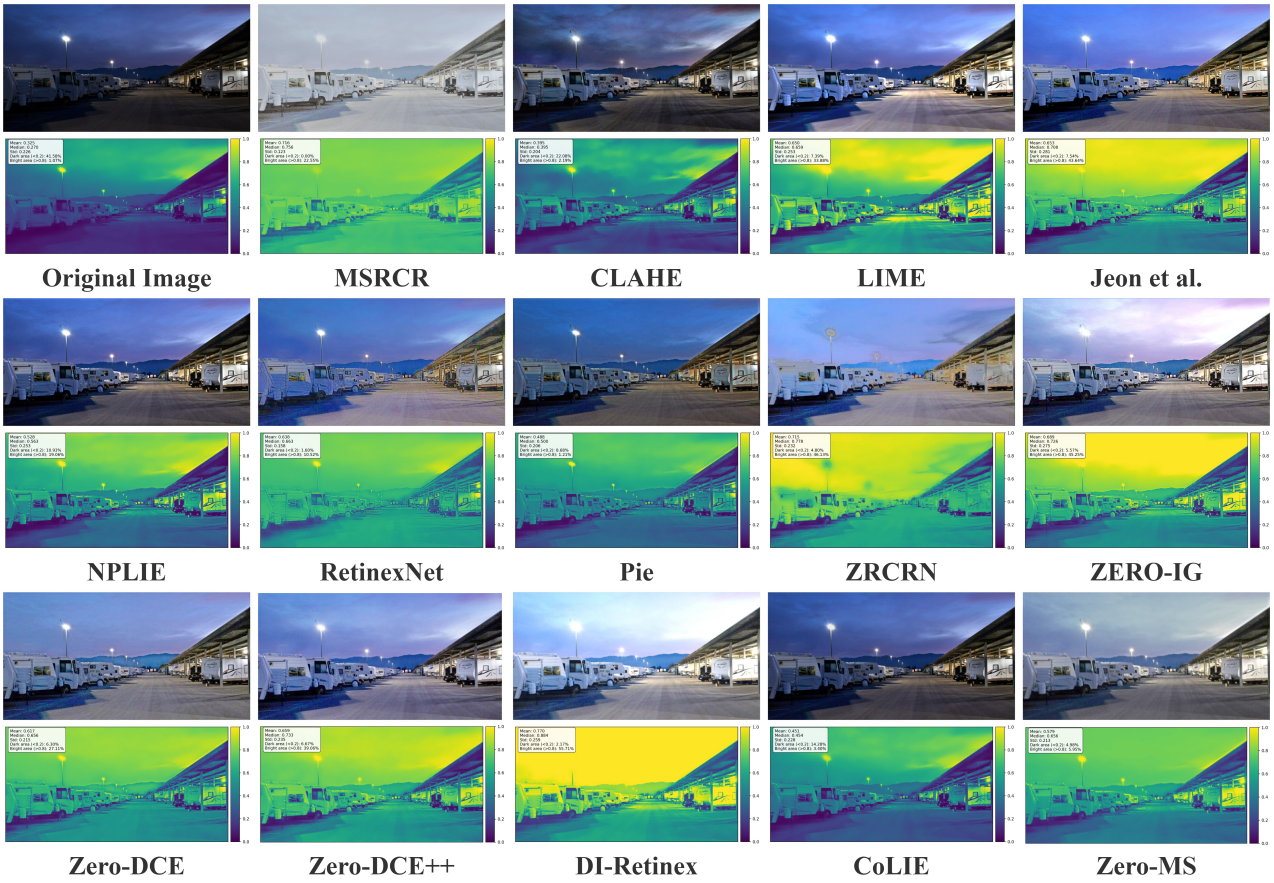


Figure 8. Qualitative comparison of low-light image enhancement results and corresponding brightness distributions on the NPE dataset. The upper panels show the input or enhanced images, while the lower panels display brightness maps derived from the maximum RGB channel values and refined using grayscale-guided filtering. All maps use a fixed [0, 1] scale for consistent comparison. The mean, median, and standard deviation describe the brightness distribution. Dark and bright areas indicate the proportions of pixels with values below 0.20 and above 0.80, respectively. These maps represent spatial brightness distributions rather than illumination components estimated by individual methods. Please zoom in for clearer visualization.

4.2.3. Efficiency and Complexity Comparison

As shown in Table 5, Zero-MS has the third-lowest parameter count and model size among the compared models with reported complexity data, requiring only 0.029740 M parameters and 0.138001 MB of storage. Compared with DI-Retinex, ZERO-IG, CoLIE, and RetinexNet, its parameter count is reduced by 29.50%, 65.65%, 77.66%, and 93.31%, respectively. The corresponding reductions in model size are 19.56%, 61.45%, 74.34%, and 92.24%. However, its mean runtime of 0.838589 ms is longer than that of the other eight models with reported parameter counts. The sequential denoising, multi-scale illumination recovery, and structure-aware refinement stages may contribute to this additional processing overhead. Despite the longer runtime, Zero-MS achieves the highest PSNR of 18.6513 dB and SSIM of 0.7918 on LOL-v1. It also obtains the lowest CIEDE2000 of 0.0874, PIQE of 4.3103, and BRISQUE of 17.0363. These results highlight Zero-MS's advantages in enhancement quality and model compactness, although its inference speed remains a limitation.

Table 5. Comparison of model complexity and inference-time stability on the LOL-v1 dataset in terms of parameter count, model size, and runtime standard deviation. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	Params. (M) ↓	Model Size (MB) ↓	Mean Runtime (ms) ↓
CLAHE [5]	-	-	**0.002989**
MSRCR [6]	-	-	3.478907
LIME [7]	-	-	50.986181
Jeon et al. [9]	-	-	0.158808
NPLIE [8]	-	-	1.455616
ZRCRN [17]	**0.000075**	**0.003184**	0.049109
ZERO-IG [4]	0.086572	0.358013	0.113388
Zero-DCE [15]	0.079416	0.322804	0.225698
Zero-DCE++ [16]	**0.010561**	**0.052226**	**0.010319**
RetinexNet [10]	0.444613	1.778452	0.257137
Pie [11]	0.079416	0.322804	0.224405
DI-Retinex [12]	0.042182	0.171568	**0.023920**
CoLIE [18]	0.133121	0.537714	0.052782
Zero-MS	**0.029740**	**0.138001**	0.838589

4.3. Ablation Study

To evaluate the contribution of each component of Zero-MS, we conduct ablation studies at three levels: modules, internal components, and loss functions. The module-level study examines the individual and combined effects of MIDM, MSLM, and SRM. The component-level study evaluates the directional states and selective gates in MIDM; the local detail, global exposure, and contrast components in MSLM; and the texture and noise gates and detail terms in SRM. Loss ablations assess the contribution of each perceptual constraint. In the tables, ✓ indicates that the corresponding module, component, or loss term is included, whereas ✗ indicates that it is excluded. The notation w/o X indicates the removal of component X while keeping all other settings unchanged. Only X indicates that only component X is retained, with the remaining components in the same group removed. Full Model denotes the complete configuration with all proposed components enabled.

4.3.1. Ablation on Module Combinations

As shown in Table 6, the full model ranks first in four of the eight metrics and second in the remaining four, indicating balanced overall performance. Compared with w/o MIDM, the full model improves PSNR and SSIM by 1.67% and 14.62%, respectively, while reducing PIQE and BRISQUE by 61.59% and 41.47%. The marked reduction in $\overline{P}_{\mathrm{NPS}}$ further supports MIDM's contribution to noise suppression. Removing MSLM substantially degrades PSNR, SSIM, and perceptual quality, highlighting its role in preserving local textures and restoring spatially nonuniform illumination. Compared with w/o SRM, the full model improves PSNR and SSIM by 1.85% and 3.08%, respectively. It also reduces CIEDE2000, PIQE, and all three noise-related metrics, with $\overline{P}_{\mathrm{NPS}}$ decreasing by 38.43%. These results demonstrate the complementary roles of MIDM, MSLM, and SRM in noise suppression, illumination recovery, and reflectance refinement.

Table 6. Ablation study on module combinations. The symbols ✗ and ✓ indicate that the corresponding module is excluded and included, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	MIDM	MSLM	SRM	Information Fidelity			Visual Quality		Denoising Performance		
				PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\mathrm{NPS}}$ ↓
w/o MIDM	✗	✓	✓	**18.3402**	**0.6760**	**0.0867**	**11.2227**	29.1052	5.5749	7.6276	26.9117
w/o MSLM	✓	✗	✓	5.3725	0.4712	0.3058	19.9124	**27.1819**	**0.9513**	**1.2321**	**0.2190**
w/o SRM	✓	✓	✗	**18.3063**	**0.7674**	**0.0916**	**5.4737**	**19.9521**	**3.0394**	**6.3768**	**10.1559**
Full Model	✓	✓	✓	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**	**2.4504**	**4.6650**	**6.2530**

4.3.2. Ablation on Module Components

Ablation on MIDM: As shown in Table 7, the full model improves PSNR and SSIM by 0.79% and 0.45%, respectively, while reducing CIEDE2000 by 0.91% compared with w/o $\boldsymbol{S}$. The three noise-related metrics, $\hat{\sigma}_n$, σ_{R_n}, and $\overline{P}_{\mathrm{NPS}}$, decrease by 4.74%, 1.92%, and 0.84%, respectively. Compared with the single-direction configurations (only $\boldsymbol{S}_{lr}$, $\boldsymbol{S}_{rl}$, $\boldsymbol{S}_{tb}$, and $\boldsymbol{S}_{bt}$), the full model achieves the best PSNR, SSIM, CIEDE2000, σ_{R_n}, and $\overline{P}_{\mathrm{NPS}}$, along with the second-best $\hat{\sigma}_n$. Despite a 2.37% increase in BRISQUE, the full model provides a more balanced overall performance. These results suggest that combining four directional states is more effective than relying on a single direction to balance reconstruction quality and noise suppression. As shown in Table 8, the full model ranks among the top three across all eight metrics. It achieves the best PSNR and ranks second in SSIM, CIEDE2000, and all three noise-related metrics. Compared with w/o $\boldsymbol{G}$, PSNR and SSIM increase by 0.21% and 0.48%, respectively, while PIQE decreases by 1.04%. The values of $\hat{\sigma}_n$, σ_{R_n}, and $\overline{P}_{\mathrm{NPS}}$ decrease by 1.06%, 5.86%, and 5.97%, respectively. Although CIEDE2000 and BRISQUE increase by 0.45% and 4.00%, the full model achieves better information fidelity and noise suppression. Compared with only $\boldsymbol{G}_d$, it improves SSIM by 1.63% and reduces PIQE by 9.12%, with reductions of 5.36%, 10.54%, and 20.15% in the three noise-related metrics. Compared with only $\boldsymbol{G}_c$, PSNR increases by 0.84%, while CIEDE2000, σ_{R_n}, and $\overline{P}_{\mathrm{NPS}}$ decrease by 1.45%, 2.43%, and 4.28%, respectively. Although only $\boldsymbol{G}_r$ performs better on some individual metrics, its PSNR is 1.08% lower and CIEDE2000 is 4.33% higher than those of the full model. Overall, the full model achieves a favorable balance between information fidelity, visual quality, and noise suppression. **Ablation on MSLM:** As shown in Table 9, the full model performs well across most metrics. Compared with w/o $\boldsymbol{D}_l$, the full model improves PSNR and SSIM by 0.77% and 0.47%, respectively, while reducing PIQE, BRISQUE, and $\hat{\sigma}_n$ by 9.66%, 32.82%, and 5.11%. These results support the role of $\boldsymbol{D}_l$ in preserving local details. Compared with w/o $\boldsymbol{E}_g$, the full model improves PSNR by 1.35% and reduces PIQE, BRISQUE, and $\overline{P}_{\mathrm{NPS}}$ by 4.18%, 23.03%, and 1.32%, respectively, highlighting the contribution of $\boldsymbol{E}_g$ to global illumination recovery. Performance declines most when $\boldsymbol{F}_c$ is removed. Compared with this variant, the full model improves PSNR and SSIM by 5.95% and 7.03%, while reducing CIEDE2000, PIQE, and BRISQUE by 7.32%, 30.05%, and 32.43%, respectively. The three noise-related metrics also decrease by 16.67%, 11.43%, and 30.76%. Overall, these results demonstrate the complementary roles of the MSLM components in balancing illumination recovery and detail preservation. **Ablation on SRM**: As shown in Table 10, the full model achieves a favorable balance among information fidelity, perceptual quality, and noise suppression. Compared with w/o $\boldsymbol{G}_t$, the full model improves PSNR and SSIM by 3.64% and 3.36%, respectively, while reducing CIEDE2000, BRISQUE, and $\overline{P}_{\mathrm{NPS}}$ by 14.23%, 32.76%, and 42.01%. Compared with w/o $\boldsymbol{D}_t$, PSNR and SSIM increase by 3.46% and 7.81%, while PIQE, $\hat{\sigma}_n$, and $\overline{P}_{\mathrm{NPS}}$ decrease by 33.20%, 21.97%, and 35.16%, respectively. The largest performance degradation occurs when $\boldsymbol{D}_n$ is removed. Although w/o $\boldsymbol{G}_n$ achieves lower noise-related metrics, its PIQE and BRISQUE scores

are substantially worse. Compared with this variant, the full model reduces PIQE and BRISQUE by 88.59% and 59.77%, respectively. These results indicate that the complete SRM provides a better balance among structural preservation, perceptual quality, and residual noise suppression.

Table 7. Ablation results for the directional states in MIDM. The symbols ✗ and ✓ indicate that the corresponding module is excluded and included, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	S_{lr}	S_{rl}	S_{tb}	S_{bt}	Information Fidelity			Visual Quality		Denoising Performance		
					PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\text{NPS}}$ ↓
w/o S	✗	✗	✗	✗	18.5040	0.7882	0.0882	4.3120	**16.6422**	2.5723	**4.7561**	**6.3061**
only S_{lr}	✓	✗	✗	✗	18.5105	0.7877	**0.0875**	**4.1659**	**15.3429**	**2.3620**	4.7800	**6.3058**
only S_{rl}	✗	✓	✗	✗	**18.5592**	**0.7884**	0.0908	4.3404	**16.5378**	**2.4877**	4.8838	6.5094
only S_{tb}	✗	✗	✓	✗	**18.5946**	0.7841	**0.0876**	4.3195	16.6712	2.6494	5.0369	7.4401
only S_{bt}	✗	✗	✗	✓	18.5329	**0.7892**	0.0879	**4.1297**	16.6629	2.5389	**4.7134**	6.4690
Full Model	✓	✓	✓	✓	**18.6513**	**0.7918**	**0.0874**	**4.3103**	17.0363	**2.4504**	**4.6650**	**6.2530**

Table 8. Ablation results for the noise gate in MIDM. The symbols ✗ and ✓ indicate that the corresponding module is excluded and included, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	G_r	G_d	G_c	Information Fidelity			Visual Quality		Denoising Performance		
				PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\text{NPS}}$ ↓
w/o G	✗	✗	✗	**18.6113**	0.7880	**0.087036**	4.3555	**16.3806**	**2.4767**	4.9553	6.6498
only G_r	✓	✗	✗	18.4497	**0.7931**	0.091385	**4.1929**	17.1579	**2.2437**	**4.5170**	**5.6077**
only G_d	✗	✓	✗	**18.5925**	0.7789	**0.087435**	4.7430	**16.0458**	2.5891	5.2148	7.8313
only G_c	✗	✗	✓	18.4954	**0.7890**	0.088710	**4.1975**	17.0725	2.4792	**4.7814**	**6.5327**
Full Model	✓	✓	✓	**18.6513**	**0.7918**	**0.087425**	**4.3103**	**17.0363**	**2.4504**	**4.6650**	**6.2530**

Table 9. Ablation results for the components of MSLM. The symbols ✗ and ✓ indicate that the corresponding module is excluded and included, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	D_l	E_g	F_c	Information Fidelity			Visual Quality		Denoising Performance		
				PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\text{NPS}}$ ↓
w/o D_l	✗	✓	✓	**18.5086**	**0.7881**	**0.0900**	**4.7713**	25.3609	**2.5824**	**4.8640**	**6.2252**
w/o E_g	✓	✗	✓	**18.3989**	**0.7890**	**0.0862**	**4.4982**	**22.1344**	**2.4921**	**4.5409**	**6.3364**
w/o F_c	✓	✓	✗	17.5423	0.7361	0.0943	6.1623	**25.2116**	2.9407	5.2671	9.0311
Full Model	✓	✓	✓	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**	**2.4504**	**4.6650**	**6.2530**

Table 10. Ablation results for the components of SRM. The symbols ✗ and ✓ indicate that the corresponding module is excluded and included, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	G_t	G_n	D_t	D_n	Information Fidelity			Visual Quality		Denoising Performance		
					PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓	$\hat{\sigma}_n$ ↓	σ_{R_n} ↓	$\overline{P}_{\text{NPS}}$ ↓
w/o G_t	✗	✓	✓	✓	17.9715	**0.7652**	0.1019	**4.4083**	**25.3356**	**2.8578**	6.4519	10.7837
w/o G_n	✓	✗	✓	✓	**18.4567**	**0.7931**	**0.0856**	37.7736	42.3501	**0.3854**	**2.7002**	**2.1076**
w/o D_t	✓	✓	✗	✓	**18.0059**	0.7300	**0.0953**	**6.4521**	**18.6239**	3.1402	**5.4907**	**9.6433**
w/o D_n	✓	✓	✓	✗	15.0300	0.4817	0.1318	10.9573	32.5732	7.0613	8.0119	23.6310
Full Model	✓	✓	✓	✓	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**	**2.4504**	**4.6650**	**6.2530**

4.3.3. Ablation on Loss Function

As shown in Table 11, the full model ranks among the top three across all eight metrics, achieving the highest PSNR of 18.6513 dB and the lowest CIEDE2000 of 0.0874. The ablation results also reveal differences in the effects of individual loss terms. Removing $\mathcal{L}_{\text{mscn}}$ yields a slightly higher SSIM of 0.7958, but reduces PSNR and degrades the visual quality and noise-related metrics. Without $\mathcal{L}_{\text{noise}}$, all three noise-related metrics increase, indicating that it limits residual image-domain fluctuations. Notably, removing $\mathcal{L}_{\text{sat}}$ achieves the best results for PIQE, BRISQUE, and all three noise-related metrics. However, PSNR and SSIM decrease to 18.4200 dB and 0.7897, respectively, while CIEDE2000 increases to 0.0892. Removing $\mathcal{L}_{\text{clip}}$ also improves selected no-reference metrics but reduces reconstruction and color fidelity. These findings highlight the trade-off between information fidelity, perceptual quality, and noise suppression. The full loss configuration achieves better reconstruction and color fidelity while maintaining competitive performance on the remaining metrics, rather than optimizing each metric independently.

Table 11. Ablation results for the perceptual loss terms. The symbols ✗ and ✓ indicate that the corresponding module is excluded and included, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Methods	$\mathcal{L}_{\text{mscn}}$	$\mathcal{L}_{\text{noise}}$	$\mathcal{L}_{\text{clip}}$	$\mathcal{L}_{\text{sat}}$	**Information Fidelity**			**Visual Quality**		**Denoising Performance**		
					PSNR ↑	**SSIM ↑**	**CIEDE2000 ↓**	**PIQE ↓**	**BRISQUE ↓**	$\hat{\sigma}_n$ **↓**	σ_{R_n} **↓**	$\overline{P}_{\text{NPS}}$ **↓**
w/o $\mathcal{L}_{\text{perc}}$	✗	✗	✗	✗	18.4769	0.7475	**0.0889**	7.3667	25.3760	3.7628	5.6161	13.7544
w/o $\mathcal{L}_{\text{mscn}}$	✗	✓	✓	✓	18.4431	**0.7958**	0.0907	5.7838	19.0323	3.0074	5.5889	11.6442
w/o $\mathcal{L}_{\text{noise}}$	✓	✗	✓	✓	**18.5302**	0.7731	**0.0891**	4.8824	19.1448	3.1827	5.2485	9.0256
w/o $\mathcal{L}_{\text{clip}}$	✓	✓	✗	✓	**18.5012**	**0.7901**	0.0895	**4.2843**	**16.5390**	**2.4217**	**4.7641**	**6.4185**
w/o $\mathcal{L}_{\text{sat}}$	✓	✓	✓	✗	18.4200	0.7897	0.0892	**4.2265**	**15.1818**	**2.2517**	**4.4996**	**5.4937**
Full Model	✓	✓	✓	✓	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**	**2.4504**	**4.6650**	**6.2530**

4.4. Hyperparameter Sensitivity Analysis

4.4.1. Training Hyperparameter Sensitivity

To evaluate hyperparameter sensitivity and training convergence, Figure 9 presents the normalized performance metrics and training loss curves of Zero-MS. As shown in Figure 9a, a batch size of 4 achieves favorable PSNR, SSIM, and CIEDE2000 results, balancing reconstruction fidelity, structural preservation, and color consistency. Although larger batch sizes yield better PIQE or BRISQUE scores in some cases, their performance on other metrics is less favorable. Figure 9b shows that a learning rate of 0.003 achieves the best PSNR, SSIM, and BRISQUE results among the tested settings. In contrast, a learning rate of 0.001 improves CIEDE2000 and PIQE but yields lower PSNR and higher BRISQUE. As shown in Figure 9c, the threshold $\theta = 0.55$ achieves the best PSNR, SSIM, and CIEDE2000 results, while $\theta = 0.75$ produces lower PIQE and BRISQUE scores. Based on overall performance, a batch size of 4, a learning rate of 0.003, and $\theta = 0.55$ are selected as the default settings. Finally, Figure 9d shows that the training loss decreases rapidly at first, and then declines more slowly with occasional fluctuations. Beyond approximately 2,000 epochs, further reductions are modest. Therefore, 2,000 epochs are selected to balance convergence and computational cost.

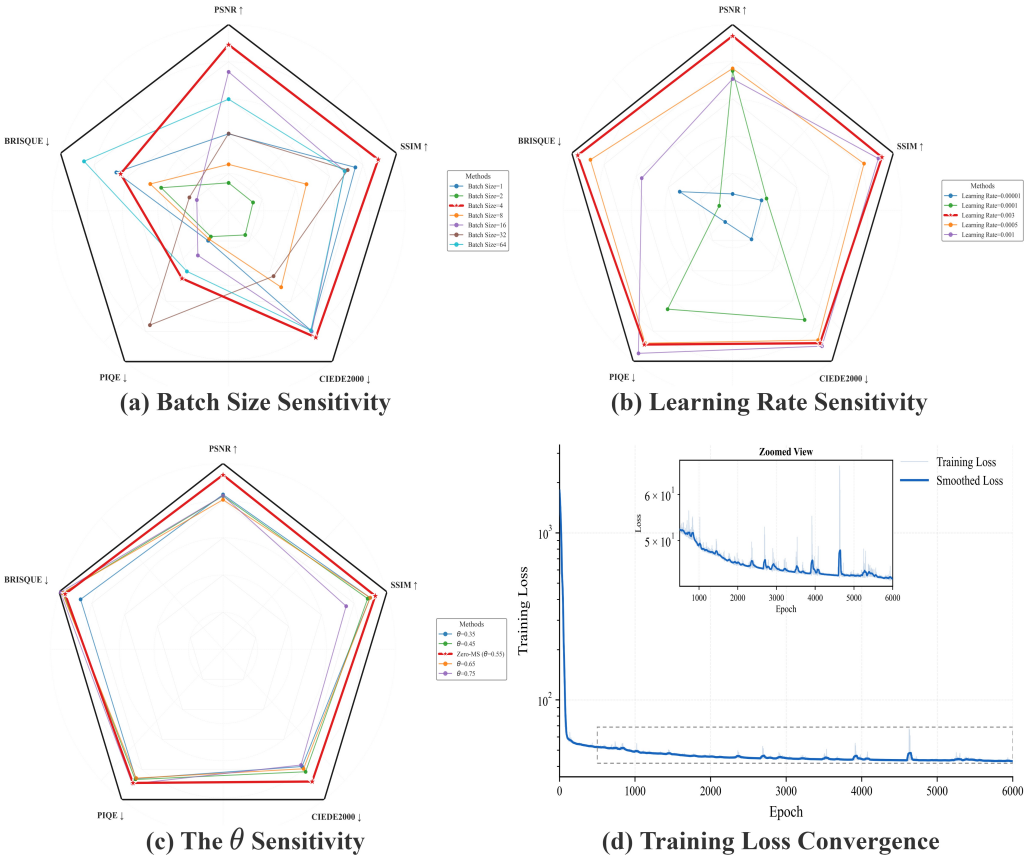


Figure 9. Hyperparameter sensitivity and training convergence of Zero-MS. (**a**) Performance comparison across different batch sizes. (**b**) Performance comparison across different learning rates. (**c**) Performance comparison across different threshold values θ. (**d**) Training loss convergence, with a zoomed-in view of the later training stage. The evaluation metrics in (**a**–**c**) are normalized to a common scale, with values closer to the outer boundary indicating better performance. In (**d**), the light-blue and dark-blue curves represent the original and smoothed training losses, respectively. Please zoom in for clearer visualization.

4.4.2. Method Hyperparameter Sensitivity

To evaluate the selected hyperparameter settings, we jointly vary several parameters around their default values. The results are presented in Table 12. Among the tested configurations, the default setting achieves the best overall performance. Increasing or decreasing the parameter values leads to varying degrees of performance degradation across the evaluated metrics. These results suggest that the default configuration provides a favorable balance among reconstruction accuracy, structural preservation, color consistency, and perceptual quality within the tested range. We therefore adopt this configuration in the final model. To examine how pooling scales affect low-light image enhancement, we compare different single-scale and multi-scale average pooling configurations. The results are presented in Table 13. For single-scale pooling, smaller kernels better preserve local structural details but capture large-scale illumination variations less effectively. Larger kernels expand the receptive field but may weaken fine textures. Thus, a single pooling scale struggles to balance preserving local detail and modeling global illumination. In contrast, multi-scale pooling combines information across spatial scales to better represent complex illumination patterns. Among the tested configurations, the proposed setting achieves the best overall performance, providing a favorable balance between local detail recovery

and global illumination modeling. Therefore, Zero-MS adopts this multi-scale pooling configuration for feature extraction.

Table 12. Quantitative evaluation of joint hyperparameter configurations. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Configuration	α_1	α_2	α_3	α_4	β_1	ω_1	ω_2	ω_3	PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓
Parameter 1	0.55	0.50	0.02	0.30	0.40	0.60	0.10	0.050	18.2858	**0.7465**	0.0909	**6.4263**	28.8807
Parameter 2	0.65	0.60	0.04	0.40	0.50	0.70	0.15	0.075	**18.4084**	0.7282	**0.0885**	8.5408	**26.5232**
Parameter 4	0.85	0.80	0.08	0.60	0.70	0.90	0.25	0.125	**18.3054**	**0.7294**	0.0907	**8.3858**	**25.7531**
Parameter 5	0.95	0.90	0.10	0.70	0.80	1.00	0.30	0.150	18.2269	0.7210	**0.0896**	9.3643	36.0572
Zero-MS	0.75	0.70	0.06	0.50	0.60	0.80	0.20	0.100	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**

Table 13. Sensitivity analysis of different pooling-scale configurations. The parameters k_f, k_m, and k_c denote fine-, intermediate-, and coarse-scale average pooling kernel sizes, respectively. The symbols ↑ and ↓ indicate that higher and lower values are better, respectively. The best, second-best, and third-best results are highlighted in bold red, blue, and brown, respectively.

Configuration	k_f	k_m	k_c	PSNR ↑	SSIM ↑	CIEDE2000 ↓	PIQE ↓	BRISQUE ↓
Single-scale (3)	3	3	3	18.4986	**0.7914**	0.0878	5.6032	21.3570
Single-scale (5)	5	5	5	**18.6357**	0.7786	0.0895	4.6666	23.8200
Single-scale (7)	7	7	7	18.2912	0.7647	0.0888	5.1158	25.0668
Single-scale (9)	9	9	9	17.9203	0.7438	0.0914	5.6164	26.2616
Multi-scale (3, 5, 9)	3	5	9	**18.6309**	**0.7897**	**0.0875**	**4.4926**	**17.3057**
Multi-scale (3, 7, 9)	3	7	9	18.3521	0.7882	**0.0876**	**4.1679**	**17.6260**
Multi-scale (5, 7, 9)	5	7	9	18.4128	0.7636	0.0883	5.1808	26.0993
Zero-MS	**3**	**5**	**7**	**18.6513**	**0.7918**	**0.0874**	**4.3103**	**17.0363**

5. Discussion and Limitations

Low-light image enhancement requires balancing illumination recovery, noise suppression, and detail preservation. Zero-MS addresses these challenges through a noise-aware Retinex framework that integrates denoising, illumination restoration, and structural refinement. Ablation studies demonstrate the complementary roles of these modules in improving overall image quality. Perceptual constraints further help balance reconstruction fidelity and visual quality. Comparisons with existing methods illustrate these trade-offs. Although LIME and DI-Retinex achieve lower high-frequency residuals in some cases, Zero-MS provides greater brightness recovery in dark regions. In contrast, ZERO-IG and RetinexNet achieve stronger illumination enhancement but exhibit more high-frequency disturbances. These findings suggest that lower noise residuals alone do not guarantee better visual quality. By combining structural refinement with perceptual constraints, Zero-MS maintains a favorable balance among illumination recovery, noise suppression, and texture preservation. Nevertheless, some limitations remain. Certain dark regions are still slightly under-enhanced compared with the reference images, suggesting room for improvement in local illumination modeling. Although Zero-MS aims to suppress image-domain high-frequency noise, the datasets used in this study do not provide separate measurements of individual sensor-noise components. Despite its compact design, Zero-MS exhibits greater runtime than most competing methods. Future work will explore more flexible local illumination modeling, systematic approaches to the targeted suppression of specific noise components, and methods to reduce inference time for practical deployment.

6. Conclusions

This paper proposes Zero-MS, a lightweight zero-shot low-light image enhancement method based on a noise-aware Retinex framework. It integrates MIDM for noise modeling and suppression, MSLM for multi-scale illumination recovery, and SRM for reflectance structure refinement. Experimental comparisons and ablation studies show that these modules jointly improve reconstruction fidelity, visual quality, and noise suppression while preserving structural details. Its low parameter count and small model size further demonstrate its compact design. However, some dark regions remain slightly under-enhanced. The specific types of noise it suppresses remain unclear, and its runtime is higher than most competing methods. Future work will focus on more flexible local illumination modeling, targeted suppression of specific noise, and reduced runtime. Overall, Zero-MS provides a compact approach that balances illumination recovery, noise suppression, and detail preservation.

Author Contributions: L.W.: conceptualization, methodology, software, validation, supervision, writing—original draft, and resources. M.T.: writing—review and editing, formal analysis, and resources. Q.C.: writing—review and editing, and data curation. G.Y.: investigation and software. N.P.: visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Foundation Research Project of Kaili University (grant No. 2025YB028), the Guizhou Provincial Science and Technology Projects (grant No. QKHJC [2024] Youth 408), the Science and Technology Innovation Talent Team Project of Data Science and Computing Intelligence of Guizhou Province (grant No. QKHRC-CXTD2025038), the Scientific Research Platform (Key Laboratory) Project of Kaili University (grant No. YTH-PT202501), the Big Data and Pattern Recognition Research Team Project of Kaili University (grant No. YTH-TD20252I), and the Specialized Fund for the Doctoral of Kaili University (grant No. BS20240102).

Informed Consent Statement: Informed consent was not required for this study because it used public image datasets and did not involve personal or private information or biomedical imaging data.

Data Availability Statement: The datasets used in this study are publicly available from the sources cited in the manuscript. The experimental results supporting the findings of this study are presented in the article. The source code of the proposed method is not publicly available.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. 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*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1632–1640. [CrossRef]
2. Xu, X.; Wang, R.; Fu, C.W.; Jia, J. SNR-Aware Low-light Image Enhancement. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 17693–17703. [CrossRef]
3. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-stage Retinex-based Transformer for Low-light Image Enhancement. In *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 12470–12479. [CrossRef]
4. Shi, Y.; Liu, D.; Zhang, L.; Tian, Y.; Xia, X.; Fu, X. ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2024; pp. 3015–3024. Available online: https://openaccess.thecvf.com/content/CVPR2024/html/Shi_ZERO-IG_Zero-Shot_Illumination-Guided_Joint_Denoising_and_Adaptive_Enhancement_for_Low-Light_CVPR_2024_paper.html (accessed on 30 October 2025).
5. Pisano, E.D.; Zong, S.; Hemminger, B.M.; DeLuca, M.; Johnston, R.E.; Muller, K.; Braeuning, M.P.; Pizer, S.M. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging* **1998**, *11*, 193. [CrossRef] [PubMed]

6. Jobson, D.J.; Rahman, Z.u.; Woodell, G.A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.* **1997**, *6*, 965–976. [CrossRef] [PubMed]
7. Guo, X.; Li, Y.; Ling, H. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. Image Process.* **2017**, *26*, 982–993. [CrossRef] [PubMed]
8. Singh, K.; Parihar, A.S. Illumination estimation for nature preserving low-light image enhancement. *Vis. Comput.* **2024**, *40*, 121–136. [CrossRef]
9. Jeon, J.J.; Park, J.Y.; Eom, I.K. Low-light image enhancement using gamma correction prior in mixed color spaces. *Pattern Recognit.* **2024**, *146*, 110001. [CrossRef]
10. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep Retinex Decomposition for Low-Light Enhancement. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 3–6 September 2018. Available online: https://daooshee.github.io/BMVC2018website/ (accessed on 18 November 2024).
11. 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]
12. Sun, S.; Ren, W.; Peng, J.; Song, F.; Cao, X. DI-Retinex: Digital-imaging Retinex model for low-light image enhancement. *Int. J. Comput. Vis.* **2025**, *133*, 8293–8314. [CrossRef]
13. Zhou, H.; Cao, C.; Tong, J.; Zheng, K. CWFAS-Net: Low-light image enhancement using curvelet wavelet Attention and Fourier Transform. *Expert Syst. Appl.* **2025**, *278*, 127263. [CrossRef]
14. Zhang, L.; Li, Z.; Cheng, L.; Zhang, Q.; Liu, Z.; Zhang, X.; Xiao, C. DLIENet: A lightweight low-light image enhancement network via knowledge distillation. *Pattern Recognit.* **2026**, *169*, 111777. [CrossRef]
15. 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*; IEEE: New York, NY, USA, 2020; pp. 1780–1789. Available online: https://openaccess.thecvf.com/content_CVPR_2020/html/Guo_Zero-Reference_Deep_Curve_Estimation_for_Low-Light_Image_Enhancement_CVPR_2020_paper.html (accessed on 18 July 2025).
16. 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] [PubMed]
17. Wang, X.; Huang, L.; Li, M.; Han, C.; Liu, X.; Nie, T. Fast, Zero-Reference Low-Light Image Enhancement with Camera Response Model. *Sensors* **2024**, *24*, 5019. [CrossRef] [PubMed]
18. Chobola, T.; Liu, Y.; Zhang, H.; Schnabel, J.A.; Peng, T. Fast context-based low-light image enhancement via neural implicit representations. In *Proceedings of the Computer Vision–ECCV 2024*; Springer Nature: Cham, Switzerland, 2025; pp. 413–430. [CrossRef]
19. Li, J.; Tu, L.; Ni, Y.; Cheng, S.; Mei, Y.; Zhou, G.; Wang, C. ICDMamba: Illumination-color decoupling and frequency-spatial synergistic mamba for low-light image enhancement. *Neurocomputing* **2026**, *707*, 135075. [CrossRef]
20. Lang, Y.; Qian, Y. A low-light image enhancement framework based on hybrid multiscale decomposition and adaptive brightness adjustment model. *Opt. Laser Technol.* **2025**, *185*, 112621. [CrossRef]
21. 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 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2025; pp. 5678–5687. [CrossRef]
22. Xu, R.; Niu, Y.; Li, Y.; Xu, H.; Liu, W.; Chen, Y. URWKV: Unified RWKV Model with Multi-state Perspective for Low-light Image Restoration. In *Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2025; pp. 21267–21276. [CrossRef]
23. Ma, L.; Ma, T.; Liu, R.; Fan, X.; Luo, Z. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 5627–5636. [CrossRef]
24. 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. In Proceedings of the 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, 21–24 November 2022. Available online: https://bmvc2022.mpi-inf.mpg.de/0238.pdf (accessed on 1 August 2023).
25. Zhang, M.; Yin, J.; Han, H.; Zhong, J.; Li, P.; Wang, Y.; Yang, C.; Zheng, Y.; Liu, H.; Tan, J.; et al. LowLightReward: A unified framework for low-light enhancement across spatial, channel, and aesthetic domains. *Neurocomputing* **2026**, *697*, 134083. [CrossRef]
26. Wang, Z.; Shang, Z.; Zhang, L.; Ghayvat, H.; He, S.; Ning, X. Fusing physical priors and visual mamba: An SNR-Aware framework for low-Light enhancement in HVI space. *Inf. Fusion* **2026**, *134*, 104350. [CrossRef]
27. Wang, W.; Yang, H.; Fu, J.; Liu, J. Zero-Reference Low-Light Enhancement via Physical Quadruple Priors. In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2024; pp. 26057–26066. [CrossRef]

28. Jiang, H.; Liu, Z.; Lei, Y.; Han, S.; Zeng, B.; Liu, S. ZeroIDIR: Zero-Reference Illumination Degradation Image Restoration with Perturbed Consistency Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2026; pp. 1320–1330. Available online: https://openaccess.thecvf.com/content/CVPR2026/html/Jiang_ZeroIDIR_Zero-Reference_Illumination_Degradation_Image_Restoration_with_Perturbed_Consistency_Diffusion_CVPR_2026_paper.html?utm_source=chatgpt.com (accessed on 29 July 2026).
29. Qiao, S.; Chen, R. Progressive feature fusion for SNR-aware low-light image enhancement. *J. Vis. Commun. Image Represent.* **2024**, *100*, 104148. [CrossRef]
30. Kozlov, A.V.; Cheremkhin, P.A.; Volkov, A.A.; Svistunov, A.S.; Starikov, R.S.; Nebavskiy, V.A.; Zlokazov, E.Y.; Rodin, V.G. Camera Noise Extraction From a Raw Single Image by Segmentation, Multiple Approximations, and Weighted Averaging. *IEEE Access* **2026**, *14*, 51768–51782. [CrossRef]
31. Land, E.H. The Retinex Theory of Color Vision. *Sci. Am.* **1977**, *237*, 108–129. Available online: http://www.jstor.org/stable/24953876 (accessed on 27 March 2020). [CrossRef] [PubMed]
32. Yang, W.; Wang, W.; Huang, H.; Wang, S.; Liu, J. Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement. *IEEE Trans. Image Process.* **2021**, *30*, 2072–2086. [CrossRef] [PubMed]
33. Wang, S.; Zheng, J.; Hu, H.M.; Li, B. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Trans. Image Process.* **2013**, *22*, 3538–3548. [CrossRef] [PubMed]
34. Horé, A.; Ziou, D. Image Quality Metrics: PSNR vs. SSIM. In *Proceedings of the 2010 20th International Conference on Pattern Recognition*; IEEE: New York, NY, USA, 2010; pp. 2366–2369. [CrossRef]
35. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
36. Sharma, G.; Wu, W.; Dalal, E.N. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.* **2005**, *30*, 21–30. [CrossRef]
37. N, V.; D, P.; Bh, M.C.; Channappayya, S.S.; Medasani, S.S. Blind image quality evaluation using perception based features. In *Proceedings of the 2015 Twenty First National Conference on Communications (NCC)*; IEEE: New York, NY, USA, 2015; pp. 1–6. Available online: https://ieeexplore.ieee.org/abstract/document/7084843 (accessed on 17 May 2021).
38. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
39. Immerkær, J. Fast Noise Variance Estimation. *Comput. Vis. Image Underst.* **1996**, *64*, 300–302. [CrossRef]
40. Rabie, T. Robust estimation approach for blind denoising. *IEEE Trans. Image Process.* **2005**, *14*, 1755–1765. [CrossRef] [PubMed]
41. Boedeker, K.L.; Cooper, V.N.; McNitt-Gray, M.F. Application of the noise power spectrum in modern diagnostic MDCT: Part I. Measurement of noise power spectra and noise equivalent quanta. *Phys. Med. Biol.* **2007**, *52*, 4027. [CrossRef] [PubMed]

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.