applied sciences

MDPI

Article

Multi-Feature Fusion Diffusion Post-Processing for Low-Light Image Denoising

Jihui Shi [1,2,3], Jijiang Huang [1,3,*], Lei Guan [1,2,3] and Weining Chen [1,3]

[1] Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China; shijihui23@mails.ucas.ac.cn (J.S.); guanlei21@mails.ucas.ac.cn (L.G.); cwn@opt.ac.cn (W.C.)
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Xi'an Key Laboratory of Aircraft Optical Imaging and Measurement Technology, Xi'an 710119, China
* Correspondence: huangjijiang@opt.ac.cn

Abstract

Various low-light image enhancement techniques inevitably introduce noise to varying degrees while improving visibility, leading to a decline in image quality that adversely affects downstream vision tasks. Existing post-processing denoising methods often produce overly smooth results lacking in detail, presenting the challenge of balancing noise suppression and detail preservation. To address this, this paper proposes a conditional diffusion denoising framework based on multi-feature fusion. The framework utilizes a diffusion model to learn the conditional distribution between underexposed and normally exposed images. Complementary features are extracted in parallel through four dedicated branches. These multi-source features are then concatenated and fused to enrich semantic information. Subsequently, redundant information is compressed via 1×1 convolutional layers, mitigating the issue of information degradation commonly encountered with U-Net skip connections during multi-scale feature fusion. Experimental results demonstrate the method's applicability across diverse scenarios and illumination conditions. It outperforms both traditional methods and mainstream deep learning models in qualitative and quantitative evaluations, particularly in terms of perceptual quality. This research provides significant technical support for subsequent image restoration and denoising within low-light enhancement pipelines.

Keywords: low-light image enhancement; diffusion model; multi-feature fusion; denoising

check for updates

Academic Editor: Thomas Lindner

Received: 7 July 2025
Revised: 6 August 2025
Accepted: 8 August 2025
Published: 11 August 2025

Citation: Shi, J.; Huang, J.; Guan, L.; Chen, W. Multi-Feature Fusion Diffusion Post-Processing for Low-Light Image Denoising. *Appl. Sci.* **2025**, *15*, 8850. https://doi.org/10.3390/app15168850

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Images captured under low-light conditions (e.g., underexposure, uneven illumination, backlighting) often suffer from severe quality degradation. Key manifestations include low visibility, distorted color information, loss of detail, and significantly amplified noise [1]. These issues not only detrimentally impact the visual quality of the images but also pose substantial challenges to subsequent image analysis and understanding tasks, particularly in critical applications such as object detection [2], facial recognition, video surveillance [3] and semantic segmentation [4]. Traditional low-light image enhancement (LLIE) techniques, such as histogram equalization (HE) [5,6], can improve image contrast to some extent but often yield suboptimal results for low-light images. While boosting brightness, they frequently amplify image noise, leading to further quality deterioration. The Retinex algorithm [7], inspired by the human visual system, may induce halo artifacts in highlight regions and reduce image naturalness; it performs poorly on high-noise images. Deep

learning-based approaches can better exploit intrinsic image features, achieving more natural and effective enhancement. However, even state-of-the-art deep learning methods exhibit limitations when confronting complex and diverse real-world scenarios [8]. For instance, while enhancing brightness, these methods may introduce artifacts like color shifts and texture distortions, compromising image fidelity and utility.

Enhancing image visibility—whether through traditional methods (e.g., manual parameter adjustment, denoising) or deep learning techniques that capture spatial features—inevitably introduces varying degrees of noise. This makes it challenging to meet the high-visibility requirements demanded by downstream tasks. Consequently, developing an efficient post-processing denoising algorithm capable of suppressing noise while preserving texture detail and color fidelity has become a central challenge in advancing the practical utility of LLIE.

Existing denoising techniques provide valuable insights for mitigating noise in low-light images. These encompass a broad spectrum, ranging from traditional low-pass filters and classical algorithms likeblock matching and 3D filtering (BM3D) [9]) to advanced deep denoising networks such as New Attention-Free Network [10] (NAFNet) and Multi-scale Inception Residual [11] Network (MIRNet *). While effective at noise reduction, these methods often produce excessively smooth results, leading to loss of detail. This issue is particularly pronounced in high-frequency image regions [12], such as areas rich in edges and textures.

Recently, Diffusion Models (DMs) have demonstrated remarkable potential in image restoration tasks due to their powerful generative capabilities [13]. By simulating a process of iterative noise addition and subsequent denoising, they generate high-fidelity image samples. Compared to traditional generative models [14], DMs exhibit significant advantages in both the quality and diversity of generated outputs. Although their application to LLIE remains nascent, preliminary research indicates that DMs hold great promise for handling complex image distributions and producing perceptually high-quality results. For instance, the Low-light Post-processing Diffusion Model (LPDM) [15] employs a single-step diffusion framework. However, its feature representation capacity is limited, which fails to comprehensively model the multi-scale degradation patterns inherent in low-light enhanced images. Issues like local brightness inconsistencies and insufficient contrast also persist.

Addressing the aforementioned challenges, this paper proposes a **M**ulti-**F**eature **F**usion **D**iffusion (MFF-Diffusion) framework for conditional diffusion-based post-processing, with three primary contributions:

1. Hierarchical Feature Extraction Mechanism: We design a Spatial Gated Recombination Unit (SRU) that employs dynamic thresholds to isolate noise regions from crucial details (e.g., edges and textures). Local contrast is enhanced through cross-scale feature recombination. Additionally, we construct a Channel Refinement Unit (CRU), leveraging group convolution and adaptive fusion strategies to mitigate issues of color distortion and uneven local luminance. An Efficient Channel Attention (ECA) module is further integrated to model global channel dependencies in a lightweight manner, effectively suppressing noise-dominated feature responses.
2. Multimodal Feature Fusion Architecture: Four parallel branches (SRU, CRU, ECA, and Identity) are proposed to extract complementary features: spatial details, refined channel relationships, global semantic information, and original feature representations. Efficient information aggregation is achieved via 1×1 convolutional compression and residual connections. This design effectively overcomes the multi-scale feature degradation problem characteristic of conventional U-Net architectures.
3. Solving local brightness imbalance: Traditional convolutional layers treat RGB channels equally, which cannot correct the noise or over-enhancement of specific channels.

Color correction in different illumination areas lacks synergy. CRU achieves refined color correction and brightness equalization through channel grouping optimization and adaptive feature fusion.

The remainder of this article is structured as follows: Section 2 provides background on LLIE and existing post-processing denoising techniques and their limitations. Section 3 discusses DM fundamentals and details the proposed network architecture. Section 4 elaborates on the experimental setup, results, and ablation studies. Finally, Section 5 concludes the work.

2. Related Work

2.1. Low-Light Image Enhancement

Current methodologies in LLIE are predominantly categorized into traditional techniques and deep learning approaches. Traditional methods primarily encompass histogram equalization-based approaches [5,6] and those grounded in Retinex theory [16,17]. Deep learning methods are chiefly classified into CNN-based methods [16], GAN-based methods [18], and DM-based approaches [19,20].

HE enhances image contrast by broadening the intensity distribution range of pixels. It encompasses both global and local processing variants. Local enhancement, which considers pixel spatial distribution, generally achieves superior results [8]. Implementations like local histogram equalization [21,22] utilize a sliding window strategy for localized processing. However, local HE incurs significantly higher computational costs compared to global HE. Crucially, most HE methods lack the flexibility to adequately restore detailed image information [23], limiting their effectiveness across diverse scenarios. Retinex Theory, proposed to model the human visual system [24,25], relies on a core assumption regarding image composition (reflectance and illumination). Algorithms based on this theory estimate and remove noise at each image position. This necessitates the assumption that illumination is consistent across the three channels of a color image [26]. Designing effective priors and regularization terms for this decomposition is inherently complex. Consequently, the decomposition process is susceptible to color distortion and detail loss, particularly in images with complex illumination or significant noise, where decomposition accuracy becomes unreliable. Furthermore, Retinex-based algorithms often induce halo artifacts in highlight regions [27], diminish image naturalness, and perform poorly on high-noise images by frequently amplifying noise [28].

CNNs have demonstrated significant effectiveness in addressing low-light image enhancement by learning the mapping relationships between input and output images. The convolutional layers are primarily responsible for extracting local features from the input images, while pooling layers contribute to parameter reduction and network efficiency improvement. The fully connected layers subsequently generate the desired output. Through the synergistic integration of these components, CNNs achieve robust performance in low-light image enhancement. Building upon this foundation, Lore et al. [3] proposed LLNet, which employs a dense hierarchical sparse denoising autoencoder to simultaneously perform contrast enhancement and denoising operations. This work established a critical framework for comprehensive network-based approaches to low-light image enhancement. Further advancing this field, Lv et al. [29] developed MBLLEN, an innovative end-to-end multi-branch enhancement network. This architecture significantly improves enhancement performance through three specialized modules: feature extraction, enhancement, and fusion, which collectively extract and process influential feature representations. Li et al. [30] introduced LightenNet, a novel approach that takes blurred illumination images as input and generates corresponding illumination maps as output. These outputs are then processed through a Retinex-based model to produce the final enhanced images. While

this method achieves satisfactory enhancement results for certain images, its performance remains limited in more challenging real-world scenarios with complex lighting conditions.

The GAN-based framework for LLIE consists of a generator and a discriminator. The generator is designed to extract enhanced features from low-light inputs and produce realistic enhanced images, while the discriminator evaluates the authenticity and quality of the generated results. Through adversarial training, the generator and discriminator engage in a competitive optimization process, progressively improving the quality of the synthesized images until the discriminator can no longer reliably distinguish them from real images. To address challenges such as overfitting and limited generalization in paired data training, Jiang et al. [14] proposed EnlightenGAN, which adopts a U-Net-based gene- rator and a dual-branch (global–local) discriminator. Their approach incorporates global and local adversarial losses, along with a self-feature preservation loss, to maintain content consistency between input and enhanced images. This strategy not only stabilizes training but also enhances the perceptual quality of the enhanced low-light images. Further advancing unsupervised learning for low-light enhancement, Zhu et al. [31] deve- loped a GAN framework with cycle-consistent adversarial loss, enabling unpaired image-to-image translation while preserving semantic integrity. Similarly, Fu et al. [32] introduced LE-GAN, an unsupervised enhancement network trained on non-corresponding low-light and normal-light image pairs. To mitigate common artifacts such as noise amplification and color distortion, they integrated an illumination-aware attention mechanism, which adaptively refines feature extraction for improved visual quality. Krishna et al. demonstrated how AI can significantly enhance diagnostic processes, particularly in low-light conditions [33]. The proposed framework in this study could further benefit from such insights.

DMs represent a class of generative models that have recently demonstrated superior sample generation capabilities compared to GANs [34]. These models operate through an iterative denoising process, typically initializing from isotropic Gaussian noise and progressively refining the distribution until high-quality samples are generated. Building upon this framework, Jiang et al. [19] developed an innovative low-light enhancement algorithm that integrates wavelet transform with diffusion modeling. This approach leverages the exceptional generative capacity of DMs to significantly improve the visual qua-lity of enhanced low-light images. Further advancing this research direction, Yi et al. [16] proposed a novel methodology that combines the classical Retinex decomposition theory with DMs for LLIE. This hybrid architecture effectively merges the strengths of traditional physics-based approaches with modern generative modeling techniques.

2.2. Post-Processing Denoising

Post-processing denoising methods can be broadly categorized into two paradigms: traditional prior-based approaches and deep learning-based methods. Among traditional methods, the BM3D [9] algorithm and its variants have emerged as particularly noteworthy solutions. BM3D operates through a distinctive denoising mechanism that involves identifying similar image blocks through pattern matching, constructing three-dimensional data groups, performing transform-domain filtering, and finally reconstructing the denoised image via aggregation processing. This algorithm has been extensively employed in classical low-light enhancement frameworks such as RetinexNet and LIME. However, global application of BM3D to enhanced images often leads to excessive smoothing in bright regions, primarily due to the significant disparity in noise intensity between dark and bright areas. To address this limitation, researchers have developed illumination-adaptive BM3D variants, such as the illumination-weighted denoising approach in LIME, that dynamically adjust denoising strength according to local brightness levels, thereby better preserving

image details. Despite these improvements, BM3D-based methods still face fundamental challenges. First, they demonstrate limited capability in modeling complex noise distributions, particularly signal-dependent noise. Second, they are unable to effectively correct color distortion artifacts that commonly occur during low-light enhancement.

In recent years, deep learning-based denoising methods have gained prominence, with NAFNet and MIRNet* emerging as representative approaches in this field. NAFNet [10] introduces a novel lightweight network architecture that fundamentally differs from conventional models by eschewing attention mechanisms in favor of an innovative feature fusion strategy and efficient structural design. This architecture achieves notable performance on benchmark datasets such as the Smartphone Image Denoising Dataset (SIDD) through two key modifications: the removal of nonlinear activation functions and a streamlined U-Net structure. The model demonstrates particular effectiveness in handling noise variations across different illumination levels, making it well-suited to low-light enhancement tasks. However, several limitations have been observed in practical applications. NAFNet shows sensitivity to characteristic artifacts in low-light enhanced images, particularly halo effects, and struggles with the complex coupling of cross-channel noise. Furthermore, the end-to-end training paradigm tends to produce excessive smoothing of high-frequency details, potentially compromising image quality in scenarios requiring fine detail preservation. These challenges underscore the need for careful consideration when applying NAFNet to real-world low-light enhancement problems.

Recent advances in deep learning-based denoising have led to significant improvements in low-light image enhancement. MIRNet*, an enhanced version of the original MIRNet architecture [11], demonstrates particular effectiveness through fine-tuning with realistic noise data generated by PNGAN. This adaptation significantly improves its robustness against real-world noise patterns. While the model achieves strong performance on mixed datasets containing both synthetic and real noise, two notable limitations persist. First, its generalization capability remains constrained when handling structural degradation patterns produced by various low-light enhancement methods such as EnlightenGAN and ZeroDCE. Second, the computational requirements become prohibitive when processing high-resolution images.

A promising alternative approach emerges in the form of LPDM [15], which introduces an innovative application of conditional diffusion models during the post-processing stage. The framework operates by using low-light images as conditional inputs to model the joint distribution between enhanced images and normally exposed references. This formulation enables comprehensive modeling of diverse degradation patterns including noise, color casts, and artifacts. Experimental validation across multiple datasets (LOL, DICM) and enhancement methods (LLFormer, KinD++) demonstrates LPDM's superior robustness, establishing it as an efficient and reliable solution for post-processing in low-light enhancement pipelines.

3. Method

3.1. Diffusion Model (DM)

DM is a kind of powerful generative model that learns data distribution through gradual denoising and iterative denoising. The core idea is to simulate a Markov chain, in which each step depends on the result of the previous step. The forward diffusion process is a process of gradually adding noise, and its goal is to transform the data from a complex distribution $q(x_0)$ to a simple noise distribution $q(x_T)$. This process can be described by a

Markov chain, where the state x_t of each step t depends on the state x_{t-1} of the previous step. The specific process of noise addition is as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{1}$$

where β_t in $[0,1]$, is to define a predefined noise schedule that controls the noise addition rate. With the reparameterization trick, any time x_t can be directly generated from x_0:

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, \tag{2}$$

where $\overline{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s), \epsilon \sim \mathcal{N}(0,1)$, the diffusion model generates the data by learning the reverse process $p_\theta(x_{t-1}|x_t, c)$ to gradually denoise it. For conditional diffusion models, the reverse process also depends on the conditional input c:

$$p_\theta(x_{0:T}|c) = p(x_T)\prod\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t, c), \tag{3}$$

The distribution of the denoising process at each step is a parameterized Gaussian distribution:

$$p_\theta(x_{t-1}|x_t, c) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t, c), \sum_\theta(x_t, t, c,)\right) \tag{4}$$

In order to avoid learning the variance, let $\sum_\theta(x_t, t, c) = \sigma_t^2\mathbf{I}$ where $\sigma_t^2 = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t$ is a time-dependent constant. Therefore, the only learnable component is μ_0. Instead of directly predicting μ_θ, the DM is parameterized in terms of a denoising autoencoder $\epsilon_\theta(x_t, t, c)$ where $t = 1, \ldots, T$. The number of timesteps T is set to a large number (such as $T = 1000$) in order for the reverse process to better approximate a Gaussian distribution. The corresponding simplified objective is as follows:

$$L_{DM} = E_{x_0,c,\epsilon,t}\left[||\epsilon - \epsilon_\theta\left(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t\epsilon}\right), t, c||^2\right] \tag{5}$$

where t is uniformly sampled from $\{1, \ldots, T\}$ and $\epsilon \sim \mathcal{N}(0,1)$. In simplified terms, L_{DM} guides the DM to predict the underlying ϵ that was involved in sampling x_t. Given x_t and a prediction for ϵ using ϵ_θ, we can calculate an estimate of x_0:

$$x_0 \approx \hat{x}_0 = \frac{1}{\sqrt{\overline{\alpha}_t}}x_t - \left(\sqrt{\frac{1}{\overline{\alpha}_t} - 1}\right)\epsilon_\theta(x_t, t, c) \tag{6}$$

While we keep the standard diffusion preliminaries for completeness, our problem is a post-enhancement denoising task conditioned on both the low-light input and its enhanced counterpart. We therefore define the task condition:

$$c = [I_{low}, I_{enh}] \tag{7}$$

and use it throughout the predictor so that each diffusion equation in this section is directly executable in our architecture. The noise estimator used in the reverse process is:

$$\varepsilon_\theta(x_t, c, t) = UNet_{MFF}(x_t \oplus c, \gamma(t)) \tag{8}$$

where $\oplus$ denotes channel-wise concatenation and $\gamma(\mathrm{t})$ is the sinusoidal time-step embedding. This explicitly binds the reverse dynamics to our $UNet_{MEF}$.

For an input feature map S, each residual block applies our MFF design and outputs

$$F = Conv_{1\times1}(ECA(S), SRU(S), CRU(S), Id(S)) + S \tag{9}$$

Here, ECA models lightweight channel dependencies, SRU performs dual-threshold spatial gating, CRU reorganizes channels and fuses them adaptively, and Id preserves the original pathway.

3.2. MFF-Diffusion

MFF-Diffusion presents an innovative framework for low-light image post-processing, employing a conditional DM enhanced by multi-branch feature fusion to achieve precise noise and artifact removal in enhanced images. The architecture builds upon a modified U-Net structure that maintains the classical encoder-decoder paradigm while incorporating several key adaptations specifically designed for DM characteristics. The encoder component comprises four hierarchical down-sampling stages, each containing dual residual blocks that progressively compress spatial dimensions while extracting increasingly abstract feature representations. Correspondingly, the decoder implements four symmetrical up-sampling stages that meticulously reconstruct image details, with skip connections from the encoder preserving crucial multi-scale information throughout the process. At the architectural core, the Middle Block integrates two residual blocks with an 8-head scaled dot-product attention mechanism, effectively capturing global contextual relationships that guide the diffusion process.

Figures 1 and 2 illustrate the training workflow of the diffusion model incorporating a conditional input mechanism. During the noise addition process, noise is progressively added via the forward diffusion process denoted as $q(x_t|x_0)$, thereby generating a time-series noise sequence as shown in Equation (2), where $\overline{\alpha}_t$ represents the cumulative noise attenuation coefficient and $t \in \{1, 2, 3, \ldots, T\}$ denotes a discrete time step. The conditional information (low-light image) is incorporated, and the noisy image x_t is concatenated with the conditional input c to serve as the input information for the diffusion model. The prediction model employs a U-Net architecture to estimate the original noise $\hat{\epsilon}$ based on the combined input $[x_t, c]$. Subsequently, the mean squared error loss function L_{DM} is utilized to minimize the discrepancy between the predicted noise $\hat{\epsilon}$ and the true noise ϵ, thereby progressively enhancing the accuracy of noise estimation through the optimization objective.

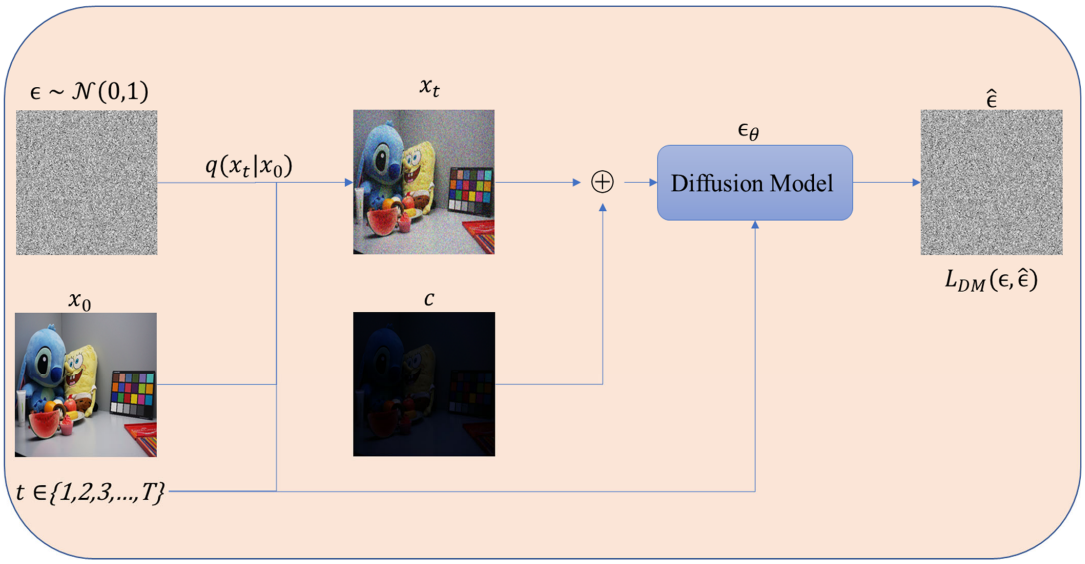


Figure 1. The training phase of MFF-Diffusion.

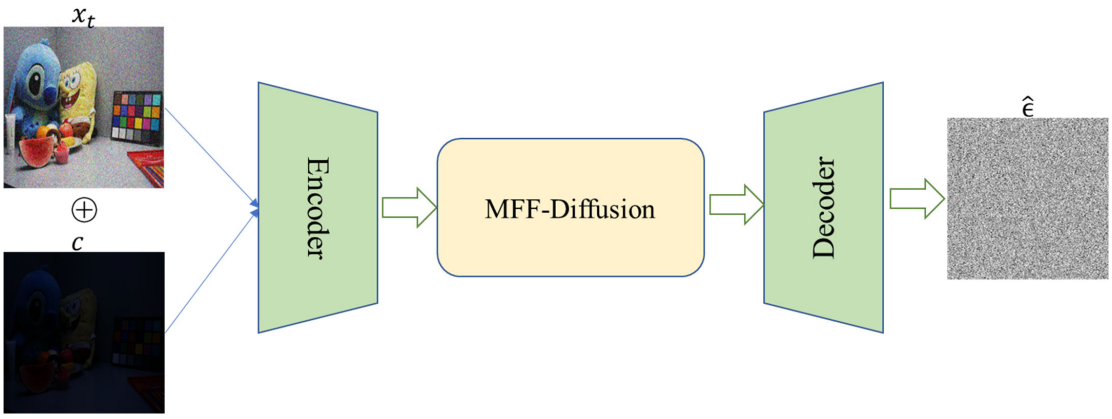


Figure 2. The architecture of Diffusion Model in training phase.

By introducing the low-light image c as the conditional input, the model dynamically perceives the structural difference between the enhanced noise and the real scene and improves the pertinence of denoising. The time step t is embedded into the network by sinusoidal encoding, which enables the model to distinguish the feature distribution of different noise levels.

Figures 3 and 4 show the noise prediction stage. After the low-light image c is enhanced by LLIE, $\hat{x}_0^{\eta}$ is obtained. This image is spliced with the low-light image c and input into the trained noise prediction network ϵ_θ to obtain the predicted noise η_φ, where $\eta_\varphi = \epsilon_\theta\left(\hat{x}_0^{\eta}, \varphi, c\right)$ and where φ = 300 corresponds to the time step parameter for adding noise, corresponding to different noise levels. The final denoised image $\hat{x}_0^{DM}$ can be further controlled by the parameter s.

$$x_0^{DM} = \frac{1}{\sqrt{\bar{\alpha}_s}}\hat{x}_0^{\eta} - \sqrt{\frac{1}{\bar{\alpha}_s} - 1} \cdot \eta_\varphi \tag{10}$$

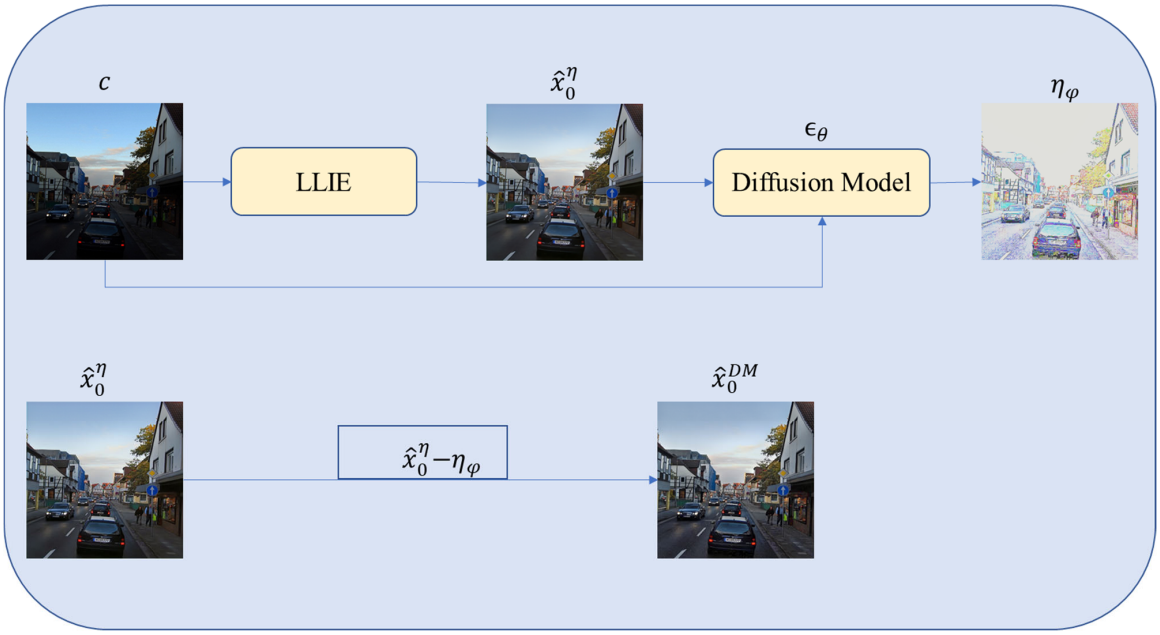


Figure 3. The inference phase of MFF-Diffusion.

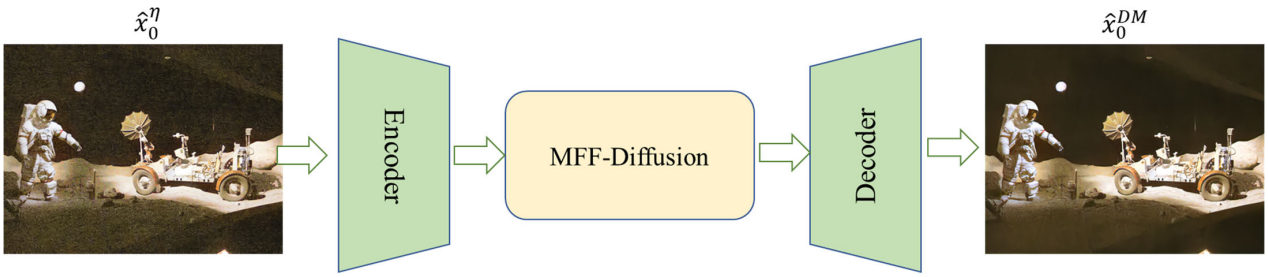


Figure 4. The architecture of the Diffusion Model in the inference phase.

3.2.1. Residual Block

As shown in Figure 5, the residual block design incorporates several optimization techniques: (1) Group Normalization replaces traditional batch normalization to improve the stability of mini-batch training; (2) the SiLU activation function balances nonlinear expression ability and computational efficiency; (3) the time-step embedding is converted into a continuous vector by sinusoidal position encoding and the residual block is input together with the channel input after concatenation so that the network can dynamically adapt to different noise levels. In addition, the attention mechanism is only applied to deep low-resolution feature maps, which balances the computational overhead and modeling ability. The architecture handles all time steps of the diffusion process through a parameter sharing mechanism and combines conditional input to achieve accurate noise estimation, which provides an efficient feature extraction basis for single-step denoising.

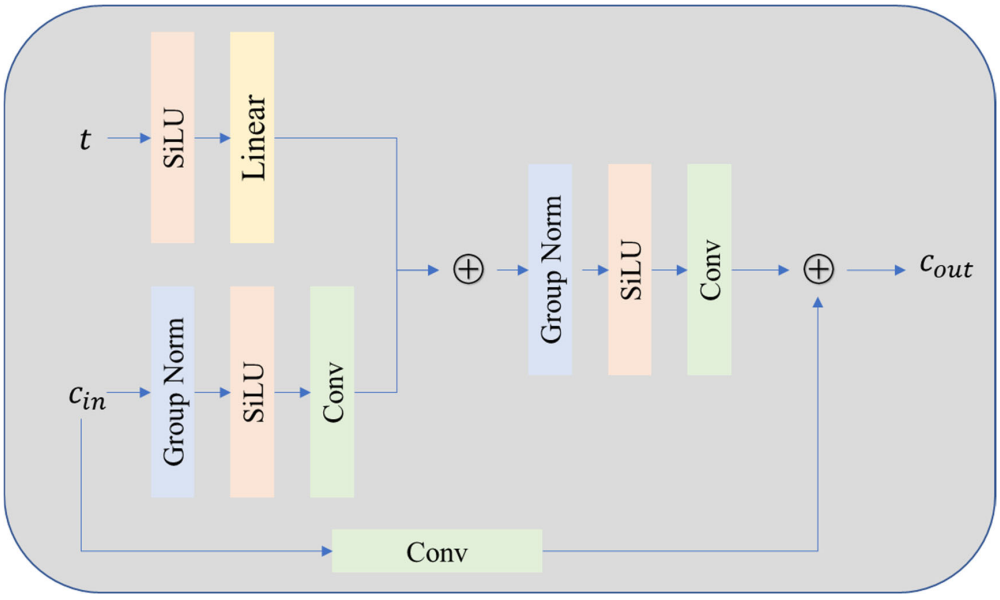


Figure 5. Residual block used throughout the DM architecture.

3.2.2. Feature Fusion Module

The multimodal feature fusion module is the core innovation of the MFF-Diffusion framework, which aims to solve the limitations of traditional U-Net in feature expression and cross-scale information fusion through hierarchical feature interaction. As shown in Figure 6, the module consists of four parallel branches:

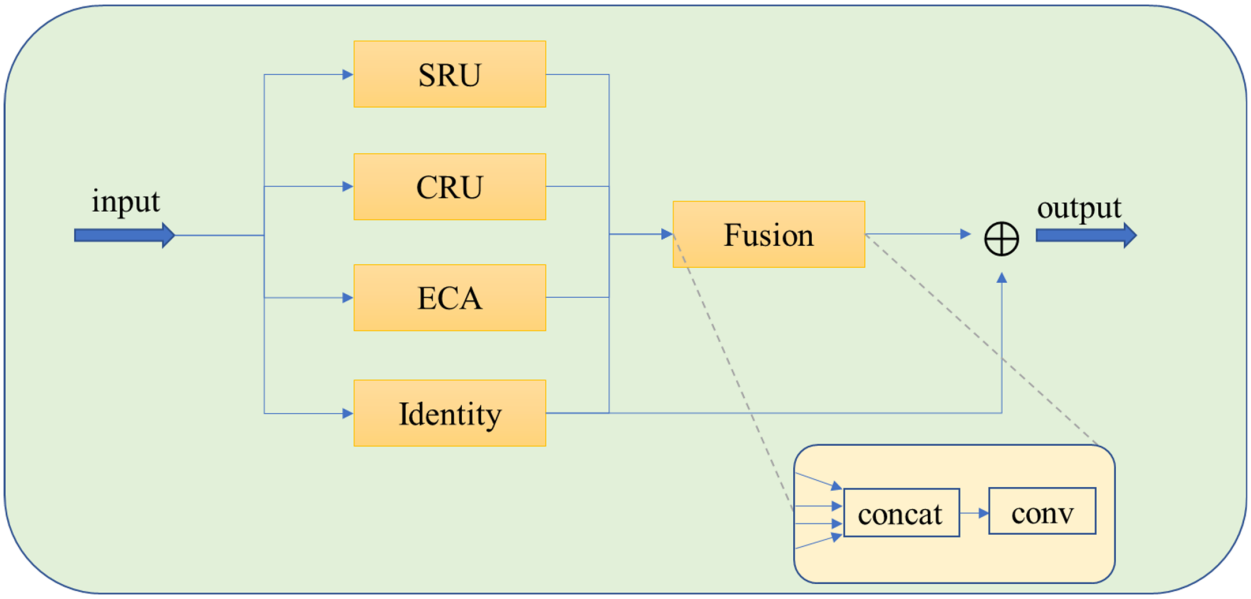


Figure 6. The framework of Multi-feature fusion module.

The ECAAttention branch (ECA) is based on channel attention mechanism and captures inter-channel dependencies through adaptive global average pooling. Specifically, after compressing the spatial dimension of the input feature map, one-dimensional convolution is used to learn the channel weights and the attention mask is generated by sigmoid activation. This branch strengthens the characteristic response of important channels.

The mathematical formulation of the CRU Adaptive Fusion Mechanism is as follows: Let $Y_1, Y_2 \in R^{N\times C\times H\times W}$ be the two subspace features. We first concatenate them:

$$Z = Concat(Y_1, Y_2) \in R^{N\times 2C\times H\times W} \tag{11}$$

Channel scores are obtained by global average pooling (GAP):

$$s = GAP(Z) \in R^{N\times 2C\times 1\times 1} \tag{12}$$

We convert the scores into normalized, non-negative channel weights using a temperatured softmax:

$$\alpha = soft\max(s/\tau), \tau > 0 \tag{13}$$

The fused representation is computed via channel-wise reweighting and half-split summation:

$$\hat{Z} = \alpha \odot Z \tag{14}$$

$$\hat{Z}^{(1)}, \hat{Z}^{(2)} = Split(\hat{Z}; C, C) \tag{15}$$

$$CRU(X) = \hat{Z}^{(1)} + \hat{Z}^{(2)} \in R^{N\times C\times H\times W} \tag{16}$$

The softmax implements competitive, convex normalization across channels (non-negativity and sum-to-one). Moreover, α is the closed-form solution to the entropy-regularized linear utility maximization on the probability simplex:

$$\max_{\alpha\in\Delta^{2C-1}} [\alpha, S] + \tau H(\alpha), \tag{17}$$

where $H(\alpha) = -\sum_c \alpha_c \log \alpha_c$ is the Shannon entropy and τ controls the sharpness (competition strength). We use $\tau = 1$ by default.

The Gated Residual branch (SRU) introduces a group normalization and adaptive gating mechanism. By grouping and normalizing the feature maps and learning the channel weights, the dual-threshold gating strategy is used to separate the salient features from the redundant noise:

- Features above the threshold ϕ are kept directly (hard thresholding).
- Features below the threshold are suppressed by soft weighting. The output achieves cross-group information interaction via a channel reorganization operation to enhance the robustness of local features.

The Channel Reorganization branch (CRU) adopts the channel divide-reassemble strategy to divide the input features into two subspaces, high and low. High-level features are extracted by grouping convolution to extract local details, and low-level features are modeled by point convolution to model the global context. The outputs of the two branches are fused by adaptive weights, and the cross-scale feature complementarity is realized by softmax driven channel recalibration.

The Identity branch retains the original feature information, avoids the gradient disappearance problem, and provides the basic features for the subsequent fusion.

In the feature fusion strategy, the outputs of the four branches are concatenated along the channel dimension and then fused with the original features using weighted addition after dimensionality reduction via 1×1 convolution.

4. Experiments

4.1. Dataset Selection

To correctly model the conditional distribution between underexposed and normally exposed images, the training dataset was trained with the original paired LOL dataset and the synthetic dataset was avoided. The dataset contains 1000 images, of which 970 (485 pairs) are in the training set, and each pair contains one low-light image and one normal exposure image.

The paired dataset LOL (15 pairs), unpaired test sets DICM (64 images), MEF (17 images), and NPE (7 images) were used to evaluate the MFF-Diffusion model in the test set. This dataset contains low-light images in different environments such as urban, outdoor, and indoor, which can effectively evaluate the performance of the model in different scenarios.

4.2. Experimental Details

In the training process, the time step T was set to 1000 steps, the number of training rounds was 1000, the batch size was 8, and eight processes were loaded in parallel to match the GPU computing power and avoid a data bottleneck. The AdamW optimizer was trained on the LOL dataset with the AdamW parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\lambda = 0.01$. The learning rate was set to 1×10^{-6}; a lower learning rate is suitable for stable training of the diffusion model to prevent gradient explosion in noisy prediction tasks. The loss function uses the L_{DM} loss function. The images were randomly cropped to 256×256 size for training and randomly flipped. The linear variance table range was [0.00085,0.0120], grew linearly, and covered mild to severe noise levels. The number of input channels was 6, which corresponds to the stitching of the low-light image (3 channels) and the enhanced image (3 channels). The initial number of channels was then extended to [128, 256, 512, 512] by channel_mult, where each layer contained two residual blocks. The generated effects (e.g., noise prediction, repair results) were recorded every 500 batches to facilitate visual verification. It was verified every ten epochs to balance training efficiency and early stop detection requirements.

ECAAttention controls the local cross-channel interaction range of 1D convolutions with a kernel size of 3. The groups in the SRU are normalized to 16 groups. In the dual-

threshold gating strategy, the hard threshold is set to 0.5 to directly retain the feature responses that exceed this value and avoid the gradient attenuation of high-frequency details. The CRU channel reorganization module divides the input channel into high and low layer features by $\alpha = 0.5$. The compression ratio squeeze_radio = 2 reduces the number of channels in the high- and low-level subspaces to reduce the amount of computation. The group_size is set to 2 and the kernel size is 3×3 to extract the local texture. The number of groups in the group normalization module group_num = 16 and the numerical stability constant is set to 10^{-10}. Currently, we have measured the total number of parameters in our model, which is 97.39 million parameters. Given the resource-intensive nature of diffusion models, one of the challenges in their deployment is ensuring computational efficiency. As demonstrated by recent work on edge computing for cardiac monitoring on the Internet of Medical Things (IoMT) [35], strategies for deploying complex models in latency-sensitive medical systems can offer important guidance.

4.3. Benchmark for MFF-Diffusion

We combined the non-learning methods BIMEF and LIME with deep learning methods such as GAN, KinD, KinD ++, LLFlow, LLFormer, RetinexNet, Retinex-Net, ZeroDCE++.

For paired datasets, such as LOL, we selected four widely used performance metrics—PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index Measure), MAE (Mean Absolute Error), and LPIPS (Learned Perceptual Image Patch Similarity)—to evaluate the performance of the proposed model.

PSNR measures the pixel-level difference between the enhanced and reference images. A higher PSNR value indicates better image quality with fewer pixel errors. It is widely used for evaluating overall image fidelity but does not always align with human visual perception, especially in terms of fine details. SSIM evaluates the structural similarity between two images, considering luminance, contrast, and structural information. It better reflects human perception of image quality, making it useful for assessing how well the model preserves important details such as edges and textures. MAE calculates the average absolute difference between corresponding pixels in the enhanced and reference images. This metric provides a straightforward measure of global image accuracy, helping us assess how closely the enhanced image matches the original. LPIPS is a perceptual metric that evaluates image similarity based on deep learning features. It focuses on higher-level image features and is more aligned with human perception, making it effective for assessing the perceptual quality of enhanced images.

By combining these four metrics, we can evaluate both fidelity (via PSNR and MAE) and perceptual quality (via SSIM and LPIPS), providing a comprehensive assessment of the model's performance across different dimensions of image quality.

For unpaired datasets, such as DICM, we used three no-reference quality metrics: NIQE, BRISQUE, and SPAQ. These metrics assess image quality without needing a reference image, which is useful in real-world scenarios where ground truth images are not available.

NIQE estimates the naturalness of an image by comparing its statistical features to those of a large database of natural images. A lower NIQE score indicates better naturalness and perceptual quality. This is particularly useful for assessing the realism of images, especially in tasks where human-like fidelity is crucial. BRISQUE evaluates the spatial quality of an image, including its perceptual quality by modeling the local structural and luminance distortions. It is specifically designed to correlate with human judgments of image quality, thus providing a robust metric for perceptual assessment in non-reference scenarios. SPAQ evaluates the perceptual quality of images in terms of their viewing experience on screens. It combines image quality metrics with human perceptual responses

to assess how likely an image is to be perceived as clear and natural when viewed on display devices.

Figures 7 and 8 show the enhanced results on the LOL validation set. When combined with the MFF-Diffusion model for low-light image enhancement, not only was the brightness of the original image improved, but the image details and color restoration were also significantly improved. It can provide clearer and more natural enhancement effects in various low-light scenes.

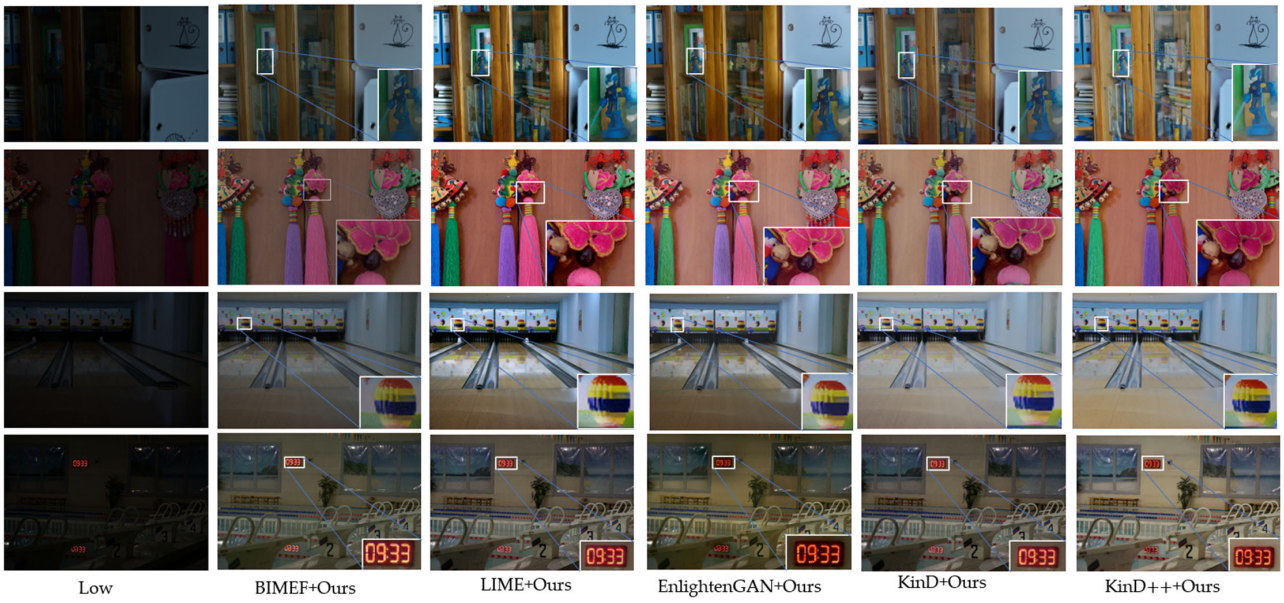


Figure 7. Visualization of different LLIE methods with MFF-Diffusion on the LOL test set.



Figure 8. Other LLIE methods with MFF-Diffusion on the LOL test set.

As shown in Table 1, both traditional methods and deep learning methods combined with the MFF-Diffusion model have improved the full reference indicators on the LOL dataset. Among them, PSNR is improved by 0.8 dB, 1.1 dB, and 1.5 dB, respectively, after combining LIME, high-level GAN, and RetinexNet. SSIM is significantly improved in

traditional methods, and the improvement of deep learning methods such as GAN and Reti- nexNet is more than 15%. MAE decline presents a small decrease, about 5%. The parameters of LPILPS are significantly improved, and some methods are increased by 50%.

Table 1. Results on the LOL test set for different LLIE methods, with and without MFF-Diffusion. The results for LPDM and MFF-Diffusion are averaged for three runs.

Methods	PSNR↑	SSIM↑	MAE↓	LPIPS↓
BIMEF	13.875	0.595	0.206	0.306
BIMEF + NAFNet	13.356	0.652	0.203	0.344
BIMEF + MIRNet*	13.946	0.669	0.204	0.313
BIMEF + LPDM	13.873	0.674	0.207	0.217
BIMEF + Ours	13.894	0.696	0.208	0.201
LIME	17.181	0.484	0.124	0.372
LIME + NAFNet	18.216	0.746	0.118	0.253
LIME + MIRNet	18.125	0.772	0.113	0.248
LIME + LPDM	17.692	0.652	0.121	0.239
LIME + Ours	17.977	0.735	0.118	0.176
EnlightenGAN	17.483	0.652	0.135	0.322
EnlightenGAN + NAFNet	17.622	0.739	0.143	0.251
EnlightenGAN + MIRNet	17.863	0.764	0.134	0.231
EnlightenGAN + LPDM	18.524	0.745	0.118	0.193
EnlightenGAN + Ours	18.607	0.769	0.117	0.177
KinD	17.648	0.771	0.123	0.175
KinD + NAFNet	17.566	0.763	0.124	0.249
KinD + MIRNet	17.605	0.770	0.126	0.248
KinD + LPDM	17.733	0.795	0.123	0.144
KinD + Ours	17.745	0.796	0.123	0.140
KinD^{++}	17.752	0.758	0.113	0.198
KinD^{++} + NAFNet	17.759	0.755	0.126	0.251
KinD^{++} + MIRNet	17.820	0.777	0.116	0.236
KinD^{++} + LPDM	17.955	0.790	0.115	0.171
KinD^{++} + Ours	18.039	0.799	0.111	0.177
LLFlow	21.133	0.852	0.084	0.119
LLFlow + NAFNet	20.662	0.810	0.089	0.221
LLFlow + MIRNet	20.867	0.814	0.089	0.196
LLFlow + LPDM	21.421	0.844	0.084	0.115
LLFlow + Ours	21.385	0.845	0.083	0.114
LLFormer	23.346	0.819	0.067	0.168
LLFormer + NAFNet	22.799	0.807	0.069	0.234
LLFormer + MIRNet	22.956	0.818	0.068	0.234
LLFormer + LPDM	23.734	0.866	0.065	0.113
LLFormer + Ours	23.750	0.872	0.065	0.109
RetinexNet	16.774	0.425	0.126	0.474
RetinexNet + NAFNet	17.521	0.719	0.124	0.309
RetinexNet + MIRNet	17.826	0.722	0.117	0.329
RetinexNet + LPDM	18.051	0.693	0.113	0.240
RetinexNet + Ours	18.191	0.761	0.111	0.211
URetinexNet	19.842	0.824	0.099	0.128
URetinexNet + NAFNet	19.482	0.774	0.100	0.222
URetinexNet + MIRNet	19.667	0.803	0.105	0.197
URetinexNet + LPDM	20.115	0.817	0.097	0.129
URetinexNet + Ours	20.119	0.822	0.096	0.123
ZeroDCE^{++}	15.357	0.573	0.176	0.335
ZeroDCE^{++} + NAFNet	15.628	0.704	0.172	0.288
ZeroDCE^{++} + MIRNet	15.630	0.707	0.172	0.297
ZeroDCE^{++} + LPDM	15.486	0.682	0.177	0.219
ZeroDCE^{++} + Ours	15.532	0.712	0.177	0.204

Table 2 shows the effects on the DICM, MEF, and NPE datasets. The model's performance was evaluated using the NIQE, MEF, and NPE parameters. It can be seen that

the parameters were significantly improved on the DICM dataset, NIQE improved by about 20%, BRISQUE improved by 40%, and SPAQ improved by about 5%. However, the parameter improvement was not significant on the MEF and NPE datasets.

Table 2. Results on other test sets for different LLIE methods with and without MFF-Diffusion.

Methods	DICM			MEF			NPE		
	NIQE↓	BRISQUE↓	SPAQ↑	NIQE↓	BRISQUE↓	SPAQ↑	NIQE↓	BRISQUE↓	SPAQ↑
BIMEF	3.827	23.588	64.375	3.330	15.262	67.352	3.330	15.262	67.352
BIMEF + Ours	3.015	9.323	68.372	3.981	22.591	66.471	3.981	22.591	66.471
LIME	3.812	23.356	63.301	3.558	16.337	70.911	3.558	16.337	70.911
LIME + Ours	2.726	9.011	69.782	3.768	19.356	70.632	3.768	19.356	70.632
EnlightenGAN	3.561	18.808	63.124	3.221	14.332	68.701	3.221	14.332	68.701
EnlightenGAN + Ours	2.925	10.038	70.479	3.894	22.712	69.556	3.894	22.712	69.556
KinD^{++}	3.786	26.314	68.086	3.737	28.635	70.250	3.737	28.635	70.250
KinD^{++} + Ours	3.063	17.728	71.441	4.126	29.774	68.937	4.126	29.774	68.937
LLFormer	3.850	14.610	63.080	3.682	17.043	63.139	3.682	17.043	63.139
LLFormer + Ours	3.491	11.363	66.839	4.242	20.813	69.072	4.242	20.813	69.072
RetinexNet	4.451	29.339	68.805	4.416	20.068	74.386	4.416	20.068	74.386
RetinexNet + Ours	3.076	9.344	73.354	3.757	14.877	73.510	3.757	14.877	73.510

As demonstrated in Figure 9, the MFF-Diffusion model delivered significant enhancements on the DICM validation set for low-light image processing. Beyond substantially improving the brightness of underexposed originals, the model achieved *breakthrough performance* in detail reconstruction and color fidelity restoration. Its unique fusion mechanism and diffusion process meticulously recovered textures and edge information typically lost in low-light conditions while accurately preserving authentic scene colors. This approach effectively mitigates common artifacts seen in conventional methods such as noise amplification, color shifts, and unnatural tonal rendering. Experimental results confirm that MFF-Diffusion consistently produces visually clear, detail-rich, and naturally colored outputs across diverse challenging scenarios, including extreme darkness, uneven illumination, and mixed indoor/outdoor lighting conditions.



Figure 9. Visualization of different LLIE methods with MFF-Diffusion on the DICM test set.

4.4. Comparison with Alternative Denoisers

In order to comprehensively evaluate the performance of the proposed improved model, we systematically compared it with the current mainstream methods NAFNet, MIRNet, and the optimal model LPDM. NAFNet and MIRNet* were evaluated separately as post-processing steps after LLIE. LPDM sets the optimal value of s = 15 or s = 30. Experiments were carried out on multiple low-light enhancement tasks (BIMEF, LIME, GAN, KinD, KinD^{++}, LLFlow, LLFormer, RetinexNet, URetinexNet, and ZeroDCE++). The quantitative analysis was performed from four dimensions: SSIM, PSNR, MAE, and LPIPS. From the overall trend, MFF-Diffuison outperformed NAFNet and MIRNet in most tasks and achieved comprehensive performance improvement compared with LPDM.

As shown in Table 1, for PSNR, on the BIMEF network, MFF-Diffusion (13.893 dB) was 0.5 dB higher than NAFNet (13.356 dB) and better than MIRNet (13.946 dB) and LPDM (13.869 dB). In other deep learning methods, it was almost equal to the optimal model LPDM, and the overall improvement was 0.1 dB. It shows that the improved model can better retain high-frequency information while suppressing noise. For the SSIM index, DFF-Diffusion (0.689) is better than the optimal model BIMEF + LPDM (0.675), an increase of 2%. It shows that the improved model is more robust to structure recovery in areas with uneven illumination. For the MAE index in the network of EnlightenGAN, compared with NAFNet and MIRNet, it significantly improved by 18%, 13%, and slightly higher than the optimal model, namely, EnlightenGAN + LPDM (0.118). It was on par with the optimal model on both the KinD and Reti- nexNet networks. The difference between MFF-Diffusion and LPDM in all tasks was small ($\pm$0.001), but it was significantly better than NAFNet, which verifies the stability of the pixel-level accuracy. In terms of LPIPS parameters, both traditional methods and deep learning methods achieved optimal or suboptimal results in all tasks, especially in low-light complex scenes. Compared with the optimal solution LPDM (0.237) in LIME network, LPIPS increased by 28.5%. The LPIPS of RetinexNet + Ours was 0.209, which is significantly lower than that of LPDM (0.244) and MIRNet (0.329), which proves that its ability to suppress noise and artifacts is stronger. In URetinexNet, it was 44% and 37% higher than NAFNet and MIRNet, 6% higher than the optimal model, and 8% higher than LPDM (0.217) in ZeroDCE++.

4.5. Ablation Study

To verify the effectiveness of the proposed SRU and CRU modules, for this paper, systematic ablation experiments were carried out on several mainstream low-light enhancement models, including BIMEF, LIME, High-Level GAN, KinD, LLFlow, LLFormer, RetinexNet, URetinexNet, and ZeroDCE. The experiment compared the performance changes of the model by gradually removing the SRU and CRU modules, and the evaluation indicators include SSIM, PSNR, MAE, and LPIPS. The experimental results are shown in Table 3:

Table 3. Performance comparison between MFF-Diffusion and its ablated variants on LOL test set.

Methods	PSNR↑	SSIM↑	MAE↓	LPIPS↓
BIMEF + Ours	13.893	0.689	0.208	0.201
BIMEF + Remove_SRU	13.871	0.688	0.208	0.207
BIMEF + Remove_GRU	13.934	0.698	0.208	0.181
BIMEF + Remove_Both	13.868	0.689	0.208	0.208
LIME + Ours	17.971	0.739	0.118	0.171
LIME + Remove_SRU	17.841	0.684	0.119	0.222
LIME + Remove_GRU	17.765	0.684	0.120	0.208
LIME + Remove_Both	17.848	0.695	0.119	0.217

Table 3. *Cont.*

Methods	PSNR↑	SSIM↑	MAE↓	LPIPS↓
EnlightenGAN + Ours	18.613	0.767	0.117	0.171
EnlightenGAN + Remove_SRU	16.771	0.734	0.159	0.189
EnlightenGAN + Remove_GRU	16.794	0.736	0.159	0.176
EnlightenGAN + Remove_Both	16.728	0.736	0.159	0.192
KinD + Ours	17.745	0.796	0.123	0.143
KinD + Remove_SRU	17.697	0.794	0.123	0.143
KinD + Remove_GRU	17.787	0.797	0.123	0.141
KinD + Remove_Both	17.714	0.792	0.123	0.148
KinD^{++} + Ours	18.042	0.801	0.110	0.153
KinD^{++} + Remove_SRU	17.982	0.796	0.111	0.154
KinD^{++} + Remove_GRU	18.028	0.797	0.110	0.160
KinD^{++} + Remove_Both	18.025	0.799	0.111	0.155
LLFlow + Ours	21.185	0.845	0.083	0.112
LLFlow + Remove_SRU	21.175	0.844	0.084	0.108
LLFlow + Remove_GRU	21.316	0.847	0.083	0.108
LLFlow + Remove_Both	21.063	0.837	0.085	0.120
LLFormer + NAFNet	23.742	0.857	0.065	0.107
LLFormer + MIRNet	23.714	0.850	0.066	0.111
LLFormer + LPDM	23.836	0.858	0.065	0.109
LLFormer + Ours	23.550	0.848	0.067	0.116
RetinexNet + Ours	18.172	0.761	0.111	0.209
RetinexNet + Remove_SRU	17.841	0.634	0.115	0.294
RetinexNet + Remove_GRU	17.838	0.635	0.115	0.281
RetinexNet + Remove_Both	17.856	0.641	0.115	0.288
URetinexNet + Ours	20.115	0.819	0.096	0.124
URetinexNet + Remove_SRU	19.926	0.826	0.098	0.123
URetinexNet + Remove_GRU	20.015	0.830	0.097	0.121
URetinexNet + Remove_Both	19.879	0.821	0.098	0.133
ZeroDCE^{++} + Ours	15.539	0.714	0.177	0.199
ZeroDCE^{++} + Remove_SRU	15.499	0.707	0.177	0.206
ZeroDCE^{++} + Remove_GRU	15.519	0.713	0.177	0.186
ZeroDCE^{++} + Remove_Both	15.500	0.711	0.177	0.210

The SRU module enhances the expression ability of spatial local features through group normalization and a dynamic gating mechanism. Experiments show that after removing the SRU, the LPIPS of most models increase significantly (perceptual quality decreases), especially in complex scenes (e.g., RetinexNet + Ours LPIPS from 0.209 to 0.294). This indicates that the SRU has a key role in detail recovery and local structure consistency. The removal of SRU results in the SSIM decreasing from 0.767 to 0.734 and PSNR decreasing from 18.613 to 16.771 in the case of <s:2> GAN + Ours, indicating that the degradation of spatial features affects the overall structure and brightness recovery. In RetinexNet + Ours, LPIPS deteriorates most significantly (+40.7%), verifying the robustness of SRU to noise and texture in low-light images.

The CRU module optimizes the correlation between feature channels using channel splitting and adaptive fusion. Experiments show that after removing the CRU, the PSNR and SSIM of some models slightly decreased, but the LPIPS performance was differentiated: In BIMEF + Ours and ZeroDCE, LPIPS decreased from 0.201 to 0.181 and 0.199 to 0.186 after removing the CRU, respectively, indicating that the CRU may suppress redundant channel noise. In LIME + Ours, removing CRU caused LPIPS to rise from 0.171 to 0.208, indicating that it is crucial for modeling channel relationships in complex lighting.

In LLFlow + Ours, the removal of SRU and CRU resulted in LPIPS increasing from 0.112 to 0.120 and SSIM decreasing from 0.845 to 0.837, indicating that the joint optimization of spatial and channel features can significantly improve the generation quality. In URetinexNet, removing SRU or CRU alone improves performance (e.g., SSIM from 0.819

to 0.830), which may be due to the redundant design of local and global features in the model, but removing both of them still degrades performance (LPIPS from 0.124 to 0.133). This further confirms the necessity of modules. In KinD + Ours, LPIPS improved after removing CRU (0.143→0.141), but SSIM and PSNR were almost unchanged, indicating that CRU tends to improve perceptual quality rather than absolute accuracy. In all experiments, MAE had the smallest variation (±0.002), indicating that SRU and CRU mainly affect higher-order features rather than pixel-level errors. Our multi-feature fusion (MFF) strategy demonstrates strong robustness, particularly in tasks involving sparse data. This is consistent with other domain-specific approaches such as the vSegNet architecture [36].

Ablation experiments show that the SRU module has a significant effect on spatial detail recovery and anti-noise interference, especially in low light conditions (such as Reti- nexNet). The CRU module can dynamically optimize channel interactions, and its effect depends on the characteristics of the dataset (e.g., suppressing noise in BIMEF and enhancing semantic consistency in LIME). The cooperation of the two models can cover the space and channel dimensions and achieve a balanced improvement of SSIM, PSNR, and LPIPS in most models.

5. Conclusions

In this paper, we presented a novel Multi-Feature Fusion Diffusion (MFF-Diffusion) framework for post-processing low-light image enhancement. The proposed method integrates advanced feature fusion techniques and a conditional diffusion model to effectively remove noise and artifacts while preserving image details and perceptual quality. Our approach demonstrates significant improvements over existing methods, achieving superior results in both qualitative and quantitative evaluations.

The experimental results show that MFF-Diffusion outperforms traditional techniques and modern deep learning models, particularly in terms of perceptual quality, structural integrity, and noise reduction across various datasets. The multi-branch feature fusion and dynamic thresholding mechanisms play a key role in enhancing the robustness and performance of the model, making it suitable for real-world low-light enhancement tasks.

While we instantiate MFF as a lightweight post-processing block for modularity, fair cross-backbone comparison, and feasible training cost, an end-to-end variant that jointly optimizes the LLIE backbone and the diffusion model is a promising direction. We expect joint training to further improve global–local consistency and denoising, and we plan to investigate this in future work.

Author Contributions: Conceptualization, J.H. and J.S.; methodology, J.S. and L.G.; software, J.S.; validation, J.S. and W.C.; formal analysis, L.G. and W.C.; investigation, J.S.; resources, J.H.; data curation, J.S.; writing—original draft preparation, J.S.; writing—review and editing, J.H., L.G. and W.C.; visualization, J.S.; supervision, J.H.; project administration, J.H.; funding acquisition, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting this study's findings are available from the author upon reasonable request.

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

References

1. 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]
2. Liu, J.; Xu, D.; Yang, W.; Fan, M.; Huang, H. Benchmarking low-light image enhancement and beyond. *Int. J. Comput. Vis.* **2021**, *129*, 1153–1184. [CrossRef]
3. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]
4. Zheng, S.; Gupta, G. Semantic-guided zero-shot learning for low-light image/video enhancement. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 581–590.
5. Thepade, S.D.; Pardhi, P.M. Contrast enhancement with brightness preservation of low light images using a blending of CLAHE and BPDHE histogram equalization methods. *Int. J. Inf. Technol.* **2022**, *14*, 3047–3056. [CrossRef]
6. Han, Y.; Chen, X.; Zhong, Y.; Huang, Y.; Li, Z.; Han, P.; Li, Q.; Yuan, Z. Low-Illumination Road Image Enhancement by Fusing Retinex Theory and Histogram Equalization. *Electronics* **2023**, *12*, 990. [CrossRef]
7. Li, X.; Wang, W.; Feng, X.; Li, M. Deep parametric Retinex decomposition model for low-light image enhancement. *Comput. Vis. Image Underst.* **2024**, *241*, 103948. [CrossRef]
8. Li, L.; Wang, R.; Wang, W.; Gao, W. A low-light image enhancement method for both denoising and contrast enlarging. In Proceedings of the 2015 IEEE International Conference on Image Processing, Quebec City, QC, Canada, 27–30 September 2015; pp. 3730–3734.
9. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process* **2007**, *16*, 2080–2095. [CrossRef]
10. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In Proceedings of the Computer Vision–ECCV 2022, Proceedings of the 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; Springer Nature: Cham, Switzerland, 2022; pp. 17–33.
11. Cai, Y.; Hu, X.; Wang, H.; Zhang, Y.; Pfister, H.; Wei, D. Learning to generate realistic noisy images via pixel-level noise-aware adversarial training. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 3259–3270.
12. Zhou, Y.; Jiao, J.; Huang, H.; Wang, Y.; Wang, J.; Shi, H.; Huang, T. When awgn-based denoiser meets real noises. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13074–13081.
13. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
14. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [CrossRef] [PubMed]
15. Panagiotou, S.; Bosman, A.S. Denoising diffusion post-processing for low-light image enhancement. *Pattern Recognit.* **2024**, *156*, 110799. [CrossRef]
16. Yi, X.; Xu, H.; Zhang, H.; Tang, L.; Ma, J. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12302–12311.
17. Hou, J.; Zhu, Z.; Hou, J.; Liu, H.; Zeng, H.; Yuan, H. Global structure-aware diffusion process for low-light image enhancement. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 79734–79747.
18. Rahman, Z.; Aamir, M.; Gulzar, K.; Bhutto, J.A.; Ishfaq, M.; Dayo, Z.A.; Mohammadani, K.H. A Non-Reference Low-Light Image Enhancement Approach Using Deep Convolutional Neural Networks. In *Deep Learning for Multimedia Processing Applications*; CRC Press: Boca Raton, FL, USA, 2024; pp. 146–159.
19. Fu, J.; Yan, L.; Peng, Y.; Zheng, K.; Gao, R.; Ling, H. Low-light image enhancement base on brightness attention mechanism generative adversarial networks. *Multimed. Tools Appl.* **2024**, *83*, 10341–10365. [CrossRef]
20. Jiang, H.; Luo, A.; Fan, H.; Han, S.; Liu, S. Low-light image enhancement with wavelet-based diffusion models. *ACM Trans. Graph.* **2023**, *42*, 1–14. [CrossRef]
21. Wang, C.; Ye, Z. Brightness preserving histogram equalization with maximum entropy: A variational perspective. *Consum. Electron. IEEE Trans.* **2005**, *51*, 1326–1334. [CrossRef]
22. Lee, C.H.; Shih, J.L.; Lien, C.C.; Han, C.C. Adaptive Multiscale Retinex for Image Contrast Enhancement. In Proceedings of the International Conference on Signal-image Technology & Internetbased Systems, Kyoto, Japan, 2–5 December 2013.
23. Li, C.; Zhu, J.; Bi, L.; Zhang, W.; Liu, Y. A low-light image enhancement method with brightness balance and detail preservation. *PLoS ONE* **2022**, *17*, e0262478. [CrossRef]
24. Land, E.H. The Retinex Theory of Color Vision. *Sci. Am.* **1978**, *237*, 108–128. [CrossRef] [PubMed]
25. Land, E.H.; Mccann, J.J. Lightness and Retinex theory. *Josa* **1971**, *61*, 1–11. [CrossRef] [PubMed]
26. 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]
27. Rasheed, M.T.; Guo, G.; Shi, D.; Khan, H.; Cheng, X. An Empirical Study on Retinex Methods for Low-Light Image Enhancement. *Remote Sens.* **2022**, *14*, 4608. [CrossRef]

28. Chen, Y.; Wen, C.; Liu, W.; He, W. A depth iterative illumination estimation network for low-light image enhancement based on retinex theory. *Sci. Rep.* **2023**, *13*, 19709. [CrossRef]
29. Lv, F.; Lu, F.; Wu, J.; Lim, C. *MBLLEN: Low-Light Image/Video Enhancement Using CNNs*; BMVC: Newcastle, UK, 2018; Volume 220, p. 4.
30. Li, C.; Guo, J.; Porikli, F.; Pang, Y. LightenNet: A Convolutional Neural Network for Weakly Illuminated Image Enhancement. *Pattern Recognit. Lett.* **2018**, *104*, 15–22. [CrossRef]
31. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
32. Fu, Y.; Hong, Y.; Chen, L.; You, S. LE-GAN: Unsupervised Low-Light Image Enhancement Network Using Attention Module and Identity Invariant Loss. *Knowl. Based Syst.* **2022**, *240*, 108010. [CrossRef]
33. Krishna, T.G.; Mahboub, M.A.A. Improving Breast Cancer Diagnosis with AI Mammogram Image Analysis. *Medinformatics* **2024**. [CrossRef]
34. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In Proceedings of the Machine Learning Research, Proceedings of the 38th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Volume 139, pp. 8162–8171.
35. Murugesan, S.; Naganathan, A. Edge Computing-Based Cardiac Monitoring and Detection System in the Internet of Medical Things. *Medinformatics* **2024**. [CrossRef]
36. Dheivya, I.; Kumar, G.S. VSegNet–A Variant SegNet for Improving Segmentation Accuracy in Medical Images with Class Imbalance and Limited Data. *Medinformatics* **2025**, *2*, 36–48. [CrossRef]

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.