Pattern Recognition 175 (2026) 113060

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Diffusion-based Laplacian frequency-aware network for low-light image enhancement

Le Zhou [a,1], Wenjie Li [b,1], Juncheng Li [c], Guangwei Gao [a,*], Chia-Wen Lin [d]

[a] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China
[b] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
[c] School of Computer Science and Technology, East China Normal University, Shanghai, China
[d] Department of Electrical Engineering, Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan

ARTICLE INFO

Keywords:
LLIE
Diffusion model
Laplacian
Multi-frequency-aware

ABSTRACT

Low-light image enhancement (LLIE) remains two fundamental challenges: accurately recovering fine details and effectively suppressing noise. Existing methods often struggle to strike a balance between these two objectives, resulting in results that are either overly smoothed or contaminated by excessive noise artifacts. Our motivation stems from the fact that during LLIE, noise predominantly manifests in the low-frequency components of the image and tends to propagate uniformly across the image. Leveraging this insight, we propose the Diffusion-based Laplacian Frequency-aware Network (DLFN), which decomposes images into multi-scale high- and low-frequency components using Laplacian decomposition. For low-frequency noise suppression, we exploit the denoising capability of diffusion models to eliminate noise introduced during the enhancement process effectively. To prevent over-smoothing and preserve fine details, we introduce the High-Frequency Enhancement Module (HFEM), which enhances high-frequency details across multiple Laplacian subbands, optimizing texture restoration. Additionally, we design the Multi-Scale Bilateral Fusion Module (MBFM) to facilitate the progressive fusion of high- and low-frequency features. Extensive experiments show that our method outperforms existing LLIE approaches.

1. Introduction

Low-light image enhancement (LLIE) aims to improve the visibility, recover fine details, and suppress noise in images captured under low-light conditions, while maintaining naturalness and minimizing artifacts. LLIE plays a crucial role in embodied perception tasks such as segmentation [1], object detection [2], and image quality improvement [3]. The rapid advancement of deep learning has provided new avenues for LLIE, positioning learning-based methods as the mainstream solutions for addressing complex LLIE challenges [4]. However, low-light images often exhibit high noise levels and lack sufficient high-frequency details, leading to the introduction of noise artifacts or overly smooth results in enhanced images of existing methods.

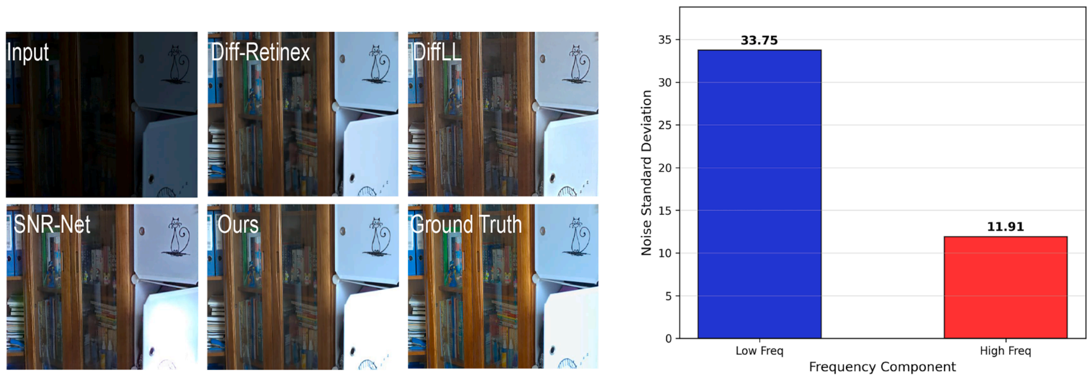
A prominent class of methods, such as Diff-Retinex [5], URetinex [6], and Retinex-Net [7], leverages Retinex decomposition theory to separate images into illumination and reflectance components [8], or applies prior-based decomposition [9] to decompose images into independent components (e.g., contours/details, low-frequency/high-frequency components). These methods aim to enhance high-frequency texture details during restoration. Component-aware enhancement strategies are then applied under physically plausible constraints. However, as shown in Fig. 1(a), these approaches [5,10,11] often lack effective noise suppression mechanisms, resulting in amplified noise, especially in shadow or dark regions, during the brightness enhancement process.

Another category of methods, such as MDMS [12], DiffLL [10], and PyDiff [13], are based on diffusion models [14], which learn to generate new samples by iteratively removing noise. While this denoising mechanism effectively suppresses noise artifacts, it often causes partial loss of fine details in the enhanced low-light images [15]. Furthermore, the noise-addition-and-removal process in these models inherently emphasizes smoothing in low-frequency regions to preserve the overall natural appearance of the restored image. As a result, existing diffusion-based approaches [5,10] frequently struggle to recover intricate structures, producing overly smoothed outputs, as illustrated in Fig. 1(a).

To strike a balance between preserving fine details and suppressing noise artifacts, it is possible to separate most of the noise inherent

* Corresponding author.
E-mail addresses: zhoul5687@163.com (L. Zhou), lewj2408@gmail.com (W. Li), cvjunchengli@gmail.com (J. Li), csggao@gmail.com (G. Gao), cwlin@ee.nthu.edu.tw (C.-W. Lin).
[1] Equal contributions.

https://doi.org/10.1016/j.patcog.2026.113060
Received 22 October 2025; Received in revised form 8 December 2025; Accepted 6 January 2026
Available online 9 January 2026



(a) Visual comparison.

(b) Noise intensity comparisons.

Fig. 1. (a): Visual comparison with existing LLIE methods, including prior-based [11] and diffusion-based [5,10] methods, shows that prior methods amplify noise or fail to recover fine details. In contrast, our method effectively suppresses noise while maximizing detail restoration. (b): The results of noise distribution in low-light images clearly demonstrate that low-frequency components contain significantly more noise compared to high-frequency components.

in low-light images from the high-frequency details and process them independently. This is feasible because, in low-light conditions, weak illumination and high ISO primarily introduce noise in the low-frequency components rather than in the fine high-frequency details. We use the standard deviation of pixel values to quantify noise intensity, measuring the noise levels in low-frequency and high-frequency components, respectively. As shown in Fig. 1(b), the noise distribution in low-light images confirms that the low-frequency component indeed contains the majority of the noise energy. Such a distribution facilitates the separation of low-frequency features from the input, enabling the diffusion denoising process to suppress undesirable noise artifacts to the greatest extent possible. Additionally, compensating for high-frequency details mitigates potential over-smoothing caused by diffusion.

Inspired by the above analysis, we propose Diffusion-based Laplacian Frequency-aware Networks (DLFN) to balance noise suppression and high-frequency detail restoration. The input low-light image is first decomposed into one low-frequency and three high-frequency components via a Laplacian pyramid, where the low-frequency component captures primary structure and illumination, and the high-frequency components represent textures and edges at multiple scales. During the feature enhancement phase, we adopt a dual-branch architecture: for high-frequency components, an efficient High-Frequency Enhancement Module (HFEM) performs adaptive enhancement via spatial-channel co-attention, with spatial attention focusing on critical edges and channel attention adjusting the contribution of each frequency band; for the low-frequency component, a conditional diffusion model reconstructs refined features, using the original low-frequency information as prior guidance to suppress noise while recovering latent details and maintaining structural consistency. In the reconstruction phase, a Multi-scale Bilateral Fusion Module (MBFM) learns adaptive spatial and channel weights to fuse multi-level features effectively. As shown in Fig. 1(a), our DLFN restores fine details while minimizing introduced noise. In summary, our main contributions can be summarized as:

- We propose a spatial–frequency hybrid decomposition strategy that augments Laplacian pyramid decomposition with FFT-based frequency analysis, enabling more complete and robust high-frequency feature representation.
- We design an HFEM that employs a dual spatial-channel mechanism, enabling adaptive adjustment of enhancement intensity across different high-frequency bands under multiple receptive fields.
- We propose an MBFM that integrates multi-level features by learning spatially adaptive weights for the fusion of multi-scale high- and low-frequency features.
- Extensive experiments demonstrate that our DLFN alleviates the issues of inaccurate high-frequency detail recovery or noise artifacts present in existing LLIE methods, surpassing existing approaches.

2. Related work

2.1. Multi-scale image processing methods

With the development of low-light image processing tasks, single-scale processing methods [7] have become inadequate for addressing image enhancement requirements under extreme lighting conditions [16]. To handle challenging low-light scenarios, images are typically decomposed into multiple scales. Multi-scale image processing techniques [10] involve more refined operations that focus on detailed textures within images, effectively improving detail representation [17] across different scales to enhance both the hierarchical perception and detailed information of images. The core of this technology lies in decomposing images into sub-band components at different scales, achieving separation between noise and valid signals through frequency-band decoupling mechanisms [18]: high-frequency components focus on enhancing texture details while low-frequency components maintain the overall structural integrity of the image. Meanwhile, several recent innovative studies have broadened the methodological landscape of low-light image enhancement. Notable advancements include event-based sensing [19], targeted artifact removal [20], degradation modeling [21], and domain-adapted enhancement [22], each contributing valuable diversity to the field. However, these approaches generally face limitations in generalization, as they are often tailored to specific scenarios or types of degradation, which may restrict their effectiveness in a wide range of real-world conditions.

Furthermore, Gaussian filter-based multi-scale decomposition methods demonstrate unique advantages. Among these, the Gaussian pyramid [23] serves as a classical multi-scale representation framework that constructs a multi-resolution hierarchical structure ranging from fine to coarse through the linear smoothing characteristics of Gaussian kernels and systematic downsampling operations [24]. The Laplacian pyramid [25] further develops upon the Gaussian pyramid by explicitly extracting detail features at each scale through differential operations between adjacent layers, forming a complementary hierarchical representation system. This collaborative working mechanism not only inherits the multi-scale characteristics of the Gaussian pyramid but also significantly enhances image representation capability through explicit separation of detail features [4]. DLEN [26] integrates Visual Transformers into Laplace Pyramids to explore cross-layer dependencies. However, these methods, due to their inherent noise suppression strategies, inevitably introduce noise artifacts in shadow regions during low-light image enhancement.

2.2. Diffusion-based LLIE methods

In recent years, generative models have made remarkable progress in computer vision [27,28]. Traditional generative-based methods represented by Generative Adversarial Networks [29] have demonstrated powerful generative capabilities: CIGAN [30] achieves precise modeling of complex lighting conditions by learning the characteristics of real low-light images through deep learning; EnlightenGAN [29] significantly improves image quality by employing an unsupervised training strategy combined with a global-local discriminator architecture.

However, with the groundbreaking advances of diffusion models [31] in computer vision, they have shown unique advantages in low-light enhancement tasks. Compared to GANs, which are prone to mode collapse and training instability, diffusion models achieve more stable training dynamics and richer sample diversity through their progressive deterministic denoising process [14]. In contrast to the common blurring or noise artifacts, their iterative refinement mechanism can suppress the generation of noise. These advantages have prompted Diff-Retinex [5], which first combines Retinex theory with diffusion models to mitigate structural loss in low-light scenes effectively, and DiffLL [10], which establishes a robust mapping between low-light and normal-light images through a wavelet-conditional diffusion model. Different from the above methods, this paper proposes an architecture that integrates the Laplacian pyramid with diffusion models, aiming to achieve more efficient and robust low-light image enhancement.

3. Proposed method

3.1. Overview of DLFN

As illustrated in Fig. 2, our DLFN first employs an improved Laplacian pyramid decomposition technique to decompose the input low-light image I_{in} into a low-frequency component $L(\tilde{x})$, which represents the global structure, and three high-frequency components H_1, H_2, and H_3 containing detailed information at different scales. For the obtained high-frequency components, a High-Frequency Enhancement Module (HFEM) based on attention mechanisms is designed to strengthen their detailed representation. As for the low-frequency component, a Conditional Diffusion Model (CDM) is adopted for refined reconstruction. During the training phase, the ground truth image I_{gt} is additionally input, and its bottom-level low-frequency component x_0 is obtained through Laplacian pyramid decomposition and subjected to a forward diffusion process to guide the denoising of the low-light image. Finally, a Multi-scale Bilateral Fusion Module (MBFM) integrates the enhanced results $\hat{H}_1$, $\hat{H}_2$, $\hat{H}_3$ and $\hat{L}(\tilde{x})$ from different scales in a learnable manner, achieving optimal feature fusion and producing the final output image I_{out}.

3.2. The laplacian pyramid

The Laplacian pyramid is a multi-resolution image representation that constructs a series of progressively downsampled and smoothed versions of an image. This pyramid ranges from the full-resolution image at its base to a coarse approximation at its apex, with each level typically halving the resolution of the level below. The Laplacian decomposition produces frequency components where higher pyramid levels capture low-frequency information while lower levels preserve high-frequency details. This multi-scale representation enables separate processing of different frequency components, which is crucial for effective noise suppression and detail enhancement in low-light conditions.

3.2.1. Spatial-domain operations

In the construction process of an image pyramid, spatial-domain operations can be systematically categorized and divided into two distinct computational stages: decomposition and reconstruction. During the decomposition stage, the original image $I_{\text{in}} = G_0 \in \mathbb{R}^{H\times W\times C}$ first undergoes Gaussian filtering, where a 5×5 Gaussian convolution kernel (standard deviation $\sigma = 1.4$) is applied to perform weighted averaging for Gaussian smoothing, resulting in the filtered Gaussian image G_0'. Subsequently, the smoothed image G_0' undergoes dyadic downsampling by removing its even-numbered rows and columns, yielding the downsampled result $G_1 \in \mathbb{R}^{(H/2)\times(W/2)\times C}$with halved resolution.

However, each level of the Gaussian pyramid, being a product of low-pass filtering, couples low-frequency information with high-frequency residuals. This coupling causes two issues: the inability to distinguish meaningful edges from noise, and the lack of selective modulation during global contrast adjustment. The Laplacian pyramid improves upon this foundation by explicitly separating the high-frequency components at the current scale through differential operations between adjacent levels. This is specifically manifested as:

$$Hi = G_i - \text{UpSample}(G_{i+1}), \tag{1}$$

where H_i denotes the ith level of the Laplacian pyramid, G_i represents the ith level of the Gaussian pyramid, and G_{i+1} corresponds to the $(i+1)$th level of the Gaussian pyramid.

This operation explicitly decouples the high-frequency components at each scale, with its physical significance being: (1) H_i precisely characterizes the frequency band features present in G_i but not preserved in G_{i+1}; (2) The top-level $L(\tilde{x})$ serves as a pure low-frequency base, carrying global illumination information. The finally constructed Laplacian pyramid can be expressed as

$$\{H_0, H_1, H_2, \dots, H_{N-1}, L(\tilde{x})\} = \text{LP}_{(D)}(\cdot), \tag{2}$$

where $\text{LP}_{(D)}(\cdot)$ denotes the Laplacian pyramid decomposition function, and N represents the number of decomposition levels in the pyramid,

$$H_k \in \mathbb{R}^{\left\lfloor \frac{H}{2^k} \right\rfloor \times \left\lfloor \frac{W}{2^k} \right\rfloor \times C}, \quad k = 0, 1, \dots, N-1. \tag{3}$$

During the reconstruction phase of the Laplacian pyramid, we reassemble the decomposed multi-scale features through inverse operations to restore the enhanced image. Starting from the top-level low-frequency component $L(\tilde{x})$, we recursively integrate high-frequency details level by level in an upward manner:

$$G_k = H_k + \text{UpSample}(G_{k+1}), \quad k = N-1, N-2, \dots, 0. \tag{4}$$

The final output I_{in} constitutes the enhanced image.

3.2.2. Frequency-domain operations

To more accurately capture the high-frequency detail information lost at different scales in the image, we innovatively introduce frequency-domain analysis as a crucial complement to spatial-domain operations. Specifically, we first perform a Fourier transform on the current scale image G_k to convert it from the spatial domain to the frequency domain, obtaining the frequency-domain representation F_k. Simultaneously, we apply an upsampling operation to the downsampled image G_{k+1} to match its dimensions with G_k, followed by a Fourier transform to derive its frequency-domain representation F_{k+1}:

$$F_k = \text{FFT2}(G_k), F_{k+1} = \text{FFT2}(G_{k+1}), \tag{5}$$

where FFT denotes the two-dimensional Fast Fourier Transform. Subsequently, through frequency-domain differential operations, we can precisely extract the spectral differences between the two scales:

$$\Delta F = F_k - F_{k+1}. \tag{6}$$

This frequency-domain differential result ΔF contains the high-frequency components present in G_k but smoothed out in G_{k+1}. To effectively integrate this information into the conventional Laplacian pyramid, we adopt the following fusion strategy:

$$H_k = G_k - \text{UpSample}(G_{k+1}) + \text{iFFT2}(\Delta F), \tag{7}$$

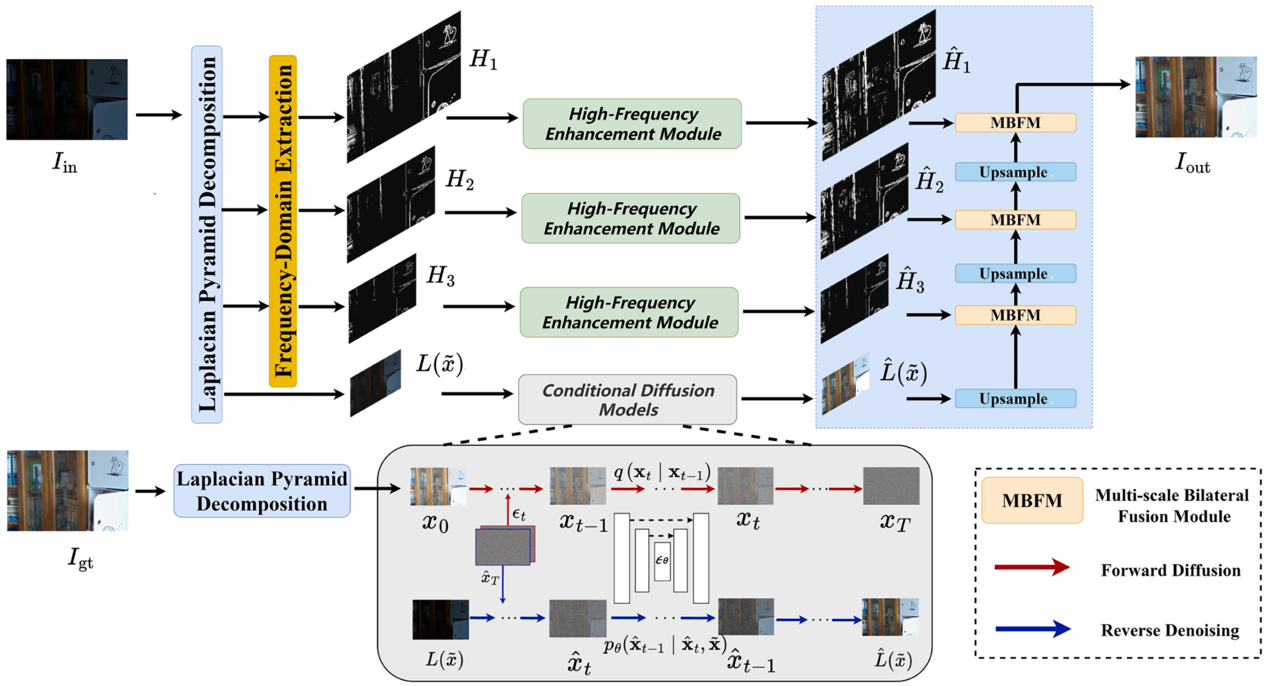


Fig. 2. Overview of our DLFN, which adopts the Laplacian pyramid for the decomposition and construction of low-light images. We employ HFEM to process high-frequency components while utilizing a diffusion model to enhance low-frequency components. The forward diffusion is exclusively executed during training, whereas the reverse denoising operates in both training and inference phases.

where iFFT2(ΔF) converts the frequency-domain differential result back to the spatial domain through inverse Fourier transform. The resulting H_k preserves both spatial-domain local details and supplements precisely extracted high-frequency information from frequency-domain analysis, achieving more complete retention of high-frequency components in texture-rich regions.

3.3. Conditional diffusion models

Inspired by the denoise capabilities of diffusion models [14], we propose a conditional diffusion model to enhance low-frequency components while preserving structural and illumination rationality. The diffusion model consists of two parts: a forward diffusion process and a reverse diffusion process. Both processes leverage the memoryless property of Markov chains to simplify complex distribution learning while optimizing state transition through neural networks.

3.3.1. Forward diffusion

This process adopts a progressive Gaussian noise addition strategy that gradually transforms the data distribution toward a standard normal distribution. Following Markov chain properties, each current state x_t depends solely on its immediate predecessor x_{t-1}, mathematically expressed as:

$$q(x_{1:T} \mid x_0) = \prod_{t=1}^{T} q(x_t \mid x_{t-1}), \tag{8}$$

where $q(x_{1:T} \mid x_0)$ denotes the probability distribution of obtaining the noisy image x_T from the original image x_0 through a T-step noise addition process.

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right), \tag{9}$$

where $q(x_t \mid x_{t-1})$ means that at step t, given the image x_{t-1} from the previous step, we obtain the conditional probability distribution of the current image x_t. This follows a Gaussian distribution with mean $\sqrt{1-\beta_t}x_{t-1}$ and variance β_t, where β_t is a predetermined noise variance coefficient and I denotes the identity matrix.

3.3.2. Reverse diffusion

This process starts from a randomly sampled pure noise $\hat{x}_T$ and progressively denoises it to recover the original data $\hat{x}_0$. Since the true distribution is unknown, a UNet-based network $p_\theta(\hat{x}_{t-1} \mid \hat{x}_t)$ is used to approximate the reverse process.

$$p_\theta(\hat{x}_{t-1} \mid \hat{x}_t) = \mathcal{N}(\hat{x}_{t-1} \mid \mu_\theta(\hat{x}_t, t), \Sigma_\theta(\hat{x}_t, t)), \tag{10}$$

where μ_θ represents the diffusion model's noise predictor by a U-Net network, thereby indirectly obtaining the mean μ_θ for computing $\hat{x}_{t-1}$:

$$\mu_\theta(\mathbf{x}_t, \tilde{\mathbf{x}}, t) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, \tilde{\mathbf{x}}, t)\right). \tag{11}$$

3.3.3. Phases of training and testing

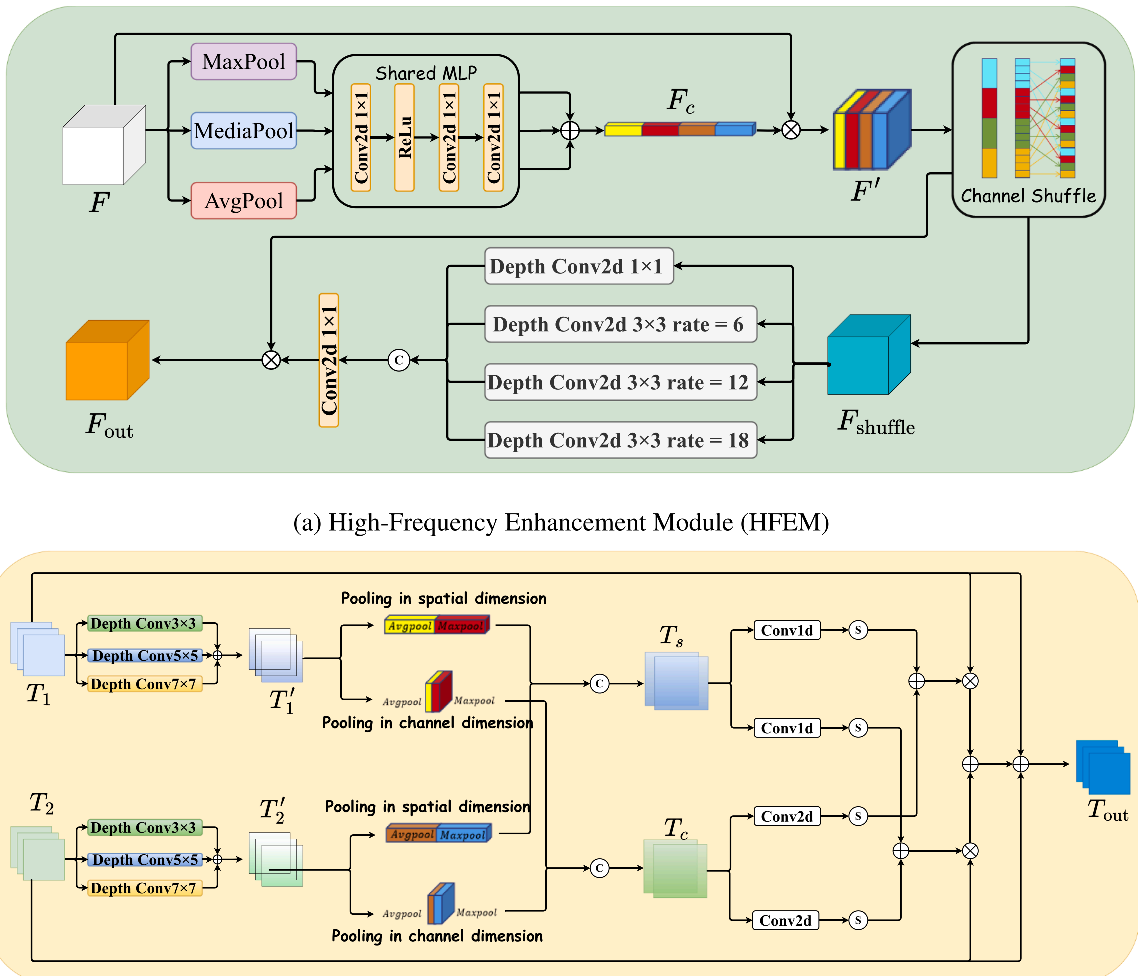
The training phase involves a full diffusion and denoising cycle. Specifically, during training, the forward diffusion process begins with the low-light input image I_{in} and its corresponding ground truth I_{gt}, incorporating their low-frequency components $L(\tilde{x})$ and x_0. Noise is progressively added to the fusion of these components. An inverse denoising process is then applied to the noisy image to train the denoising network's capability in predicting and removing the noise.

The testing phase only involves the denoising process. For an input low-light image I_{in}, its low-frequency component is first extracted via Laplacian decomposition. Starting from a concatenation of random noise $\hat{x}_T$ and I_{in}, the pre-trained denoising model performs the denoising process to ultimately generate the enhanced image with improved illumination (Fig. 3).

3.4. High-frequency enhancement module (HFEM)

To address the limitations of the diffusion model in processing high-frequency details, we propose a high-frequency enhancement module. As illustrated in Fig. 3a, the module integrates channel enhancement and spatial enhancement to enhance high-frequency features [18,32].

For channel enhancement, given an input feature F, we adopt a triple-branch parallel processing architecture. Each branch performs a



(a) High-Frequency Enhancement Module (HFEM)

(b) Multi-scale Bilateral Fusion Module (MBFM)

Fig. 3. Overview of our modules: (a) the architecture of our High-Frequency Enhancement Module (HFEM), (b) the architecture of our Multi-scale Bilateral Fusion Module (MBFM).

different global pooling operation: global average pooling, global max pooling, and global median pooling. These operations capture diverse global statistics, which are then processed by an MLP consisting of two 1×1 convolutional layers with ReLU activation to activate high-frequency global features. By combining the pooled statistics from multiple branches [33], the channel attention map F_c is generated, which is used to weight the feature channels, enhancing the representation of high-frequency details:

$$F_c = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F))) + \sigma(\mathrm{MLP}(\mathrm{MaxPool}(F))) + \sigma(\mathrm{MLP}(\mathrm{MedianPool}(F))), \tag{12}$$

where σ denotes the Sigmoid activation function. To further mix and share information, we employ a channel shuffle operation that divides the enhanced feature maps into multiple groups. By transposing the channels within each group, we increase the mixing of feature information, improving the overall representation and boosting high-frequency feature diversity. This process facilitates better preservation of fine details and textures in the final output:

$$F_{\text{shuffle}} = ChannelShuffle(F_c). \tag{13}$$

For spatial enhancement, we design multiple parallel depthwise separable convolutions with varying dilation rates. These convolutions allow the model to capture multi-scale spatial information, focusing on preserving fine details across different scales. The outputs from these branches are fused and processed through a 1×1 convolution to generate the spatial enhancement map, which is then multiplied with the channel-weighted feature maps to produce the final enhanced output F_{out}. This ensures that high-frequency components, such as edges and textures, are effectively enhanced.

3.5. Multi-scale bilateral fusion module (MBFM)

To effectively fuse enhanced image components across different scales, as shown in Fig. 3b, we design a Multi-scale Bilateral Fusion Module (MBFM). Compared to single-scale convolutional kernels, which exhibit limitations in capturing features of varying scales and complexity, thereby hindering their ability to comprehensively extract detailed information from images. Multi-scale features encompass details across different receptive fields, thereby further enhancing the model's representation.

For two distinct input components T_1 and T_2, our MBFM employs parallel processing using three convolutional kernels of different scales (3×3, 5×5, and 7×7) for each input feature map. This multi-scale operation captures feature information at different levels, allowing the model to better represent both fine details and broader context. The results from each scale are fused by summation, producing expressive multi-scale feature maps T_1' and T_2'.

During the feature fusion stage, we first apply global average pooling and global max pooling along the channel dimension to the multi-scale feature maps T_1' and T_2', while simultaneously performing average pooling and max pooling along the spatial dimension. The pooled channel and spatial features are then concatenated to obtain T_c and T_s, ensuring the preservation of both channel-wise and spatial relationships for improved feature representation:

$$T_c = \mathrm{Cat}\left(\mathrm{Avg}\left(T_1'\right), \mathrm{Max}\left(T_1'\right), \mathrm{Avg}\left(T_2'\right), \mathrm{Max}\left(T_2'\right)\right). \tag{14}$$

The channel features are processed through 1D convolution to obtain channel weights, while the spatial features are processed through 2D convolution to yield spatial weights. Both types of weights undergo Softmax normalization to ensure their sum equals 1. Finally, the final fusion process adopts a combined strategy of weighted addition and weighted multiplication, utilizing the computed channel weights and spatial weights to fuse T_1 and T_2. To maximally preserve original features, we incorporate residual connection mechanisms that add T_1 and T_2 separately to the fusion results T_{out}, ensuring critical features remain intact throughout the fusion process.

3.6. Loss function

In training, our total loss consists of three main components. First, to improve the reconstruction quality, we integrate an L1 loss and an SSIM loss into the content reconstruction term, to minimize the discrepancy between the restored image I_{out} and the ground truth image I_{gt}:

$$\mathcal{L}_{content} = \left| I_{out} - I_{gt} \right|_1 + \left(1 - \mathrm{SSIM}\left(I_{out}, I_{gt}\right)\right), \tag{15}$$

which helps maintain structural similarity while improving pixel-level accuracy. Second, to effectively preserve and reconstruct high-frequency details, we introduce a detail preservation loss $\mathcal{L}_{detail}$ that integrates MSE loss with TV regularization:

$$\mathcal{L}_{detail} = \lambda_1 \sum_{k=1}^{K} \left\| \left\{ \hat{H}_1, \hat{H}_2, \hat{H}_3 \right\} - \left\{ H_1, H_2, H_3 \right\} \right\|^2 + \lambda_2 \sum_{k=1}^{K} \mathrm{TV}\left(\left\{ \hat{H}_1, \hat{H}_2, \hat{H}_3 \right\}\right), \tag{16}$$

where the weights λ_1 and λ_2 are set to 0.1 and 0.01, respectively, aiming to suppress noise and enhance edge continuity. Third, at each step $t \in [0, T]$ of the reverse process of training, the model is optimized by minimizing the difference between the network-predicted noise $\epsilon_\theta\left(\mathbf{x}_t, \tilde{\mathbf{x}}, t\right)$ and the actual noise ϵ_t randomly added during the forward process, with the diffusion loss defined as

$$\mathcal{L}_{diff} = \mathbb{E}_{\mathbf{x}_0, t, \epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\left\| \epsilon_t - \epsilon_\theta\left(\mathbf{x}_t, \tilde{\mathbf{x}}, t\right) \right\|^2 \right]. \tag{17}$$

Moreover, inspired by DiffLL [10], we incorporate Laplacian pyramid decomposition of the ground-truth reference image into training to enhance representational capacity via multi-scale analysis. Specifically, we decompose the reference into a top-level low-frequency component x_0, which serves as structural prior information containing the global structure and coarse features. During the forward diffusion process, we introduce an L2-based constraint that minimizes the Euclidean distance between the recovered low-frequency component $\hat{x}_0$ and the target x_0, thereby enforcing structural consistency in low-light image restoration. This low-frequency component loss $\mathcal{L}_{low}$ is formulated as

$$\mathcal{L}_{low} = \mathcal{L}_{diff} + \left\| \hat{\mathbf{x}}_0 - \mathbf{x}_0 \right\|^2. \tag{18}$$

Finally, the overall loss function $\mathcal{L}_{total}$ is formulated as a combination of the above terms:

$$\mathcal{L}_{total} = \mathcal{L}_{low} + \mathcal{L}_{detail} + \mathcal{L}_{content}. \tag{19}$$

Table 1
Quantitative comparisons on LOL-v2-real [46]. The best and the second-highest result are **bolded** and <u>underlined</u>.

Methods	Reference	PSNR↑	SSIM↑	LPIPS↓
EnlightenGAN [29]	TIP'21	18.676	0.678	0.364
RUAS [34]	CVPR'21	15.351	0.626	0.395
SNR-Net [11]	CVPR'22	21.482	0.849	0.237
URetinex-Net [6]	CVPR'22	21.093	0.853	0.208
CDEF [35]	TMM'22	19.757	0.630	0.349
SCI [36]	CVPR'22	17.304	0.540	0.345
Uformer [37]	CVPR'22	18.442	0.759	0.347
UHDFour [38]	ICLR'23	21.785	<u>0.854</u>	0.292
LLFormer [39]	AAAI'23	20.485	0.792	**0.192**
GDP [40]	CVPR'23	14.294	0.493	0.435
WeatherDiff [41]	TPAMI'23	20.009	0.829	0.253
DeVigNet [42]	AAAI'24	19.813	0.706	0.256
SGF [43]	TMM'24	<u>22.052</u>	0.850	0.228
QuadPrior [44]	CVPR'24	20.483	0.811	0.302
AGLLDiff [45]	AAAI'25	19.306	0.783	0.291
DLFN (Ours)	–	**22.412**	**0.859**	<u>0.201</u>

Table 2
Quantitative results on the LOL dataset [7]. The best and the second-highest result are **bolded** and <u>underlined</u>.

Methods	Reference	PSNR↑	SSIM↑	LPIPS↓
SNR-Net [11]	CVPR'22	24.613	0.842	0.233
SCI [36]	CVPR'22	14.784	0.526	0.392
CDEF [35]	TMM'22	16.335	0.585	0.407
WeatherDiff [41]	TPAMI'23	17.913	0.811	0.272
UHDFour [38]	ICLR'23	23.093	0.821	0.259
GDP [40]	CVPR'23	15.904	0.546	0.431
LLFLOW [47]	AAAI'22	24.998	**0.873**	<u>0.152</u>
Retinexformer [48]	ICCV'23	25.152	0.843	**0.121**
Uformer [37]	CVPR'22	19.001	0.741	0.354
Diff-Retinex [5]	ICCV'23	21.751	0.812	0.372
DiffLL [10]	TOG'23	**26.336**	0.844	0.219
QuadPrior [44]	CVPR'24	20.301	0.808	0.232
LightenDiff [17]	ECCV'24	20.453	0.803	0.192
IGDFormer [49]	PR'25	24.112	0.821	0.235
DLFN (Ours)	/	<u>25.632</u>	<u>0.853</u>	0.216

4. Experiments

4.1. Implementation settings

4.1.1. Implementation details

All experiments are conducted on an NVIDIA GeForce RTX 2080Ti GPU using the PyTorch framework. The network is optimized with the Adam optimizer, configured with an initial learning rate of 1×10^{-4}, batch size of 16, and patch size of 256×256 pixels. The total number of training epochs is set to 500 for complete model convergence. The training process for 500 epochs takes approximately 20 hours.

4.1.2. Datasets

To validate the generalization capability and robustness of the proposed method, we conducted comprehensive experiments on multiple widely-used benchmark datasets for low-light image enhancement: LOL [7], LOL-v2-real [46], LOL-v2-synthetic [46], and LSRW [50]. The LOL dataset contains 500 carefully aligned low/normal-light image pairs, following the standard data partitioning scheme with 485 pairs for training and 15 pairs for testing. The more challenging LOLv2-real dataset consists of real-world captured images from diverse complex scenarios, comprising 689 training images and 100 test images. The LOL-v2-synthetic includes 900 paired images for training and the 100 pairs for testing. The LSRW dataset comprises 5650 image pairs captured under various conditions. A random sample of 5600 pairs was selected for training, with the remaining 50 pairs reserved for evaluation. This experimental setup not only ensures fair comparison with existing

Table 3
Quantitative results on LOL-v2-synthetic [46] and LSRW [50] datasets. The best and the second-highest result are **bolded** and underlined.

Methods	Reference	LOL-v2-syn [46]			LSRW [50]		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RUAS [34]	CVPR'21	16.550	0.652	0.335	14.271	0.461	0.501
SCI [36]	CVPR'22	22.203	0.801	0.301	15.242	0.419	0.404
SNR-Net [11]	CVPR'22	24.142	0.912	**0.187**	16.499	0.505	0.419
CDEF [35]	TMM'22	23.067	0.865	0.298	16.758	0.465	0.399
Uformer [37]	CVPR'22	22.037	0.871	0.330	16.591	0.494	0.435
PyDiff [13]	IJCAI'23	**25.334**	0.904	0.193	17.264	0.510	0.335
GDP [40]	CVPR'23	18.550	0.652	0.400	12.887	0.362	0.412
UHDFour [38]	ICLR'23	24.733	0.887	0.223	17.300	0.529	0.443
WeatherDiff [41]	TPAMI'23	23.796	0.902	0.214	16.507	0.487	0.431
QuadPrior [44]	CVPR'24	16.112	0.758	0.306	14.538	0.476	0.425
LightenDiff [17]	ECCV'24	18.550	0.707	0.334	17.314	0.512	**0.310**
AGLLDiff [45]	AAAI'25	21.112	0.870	0.256	16.213	0.498	0.423
DLFN (Ours)	/	25.132	**0.920**	0.198	**17.562**	**0.530**	0.390

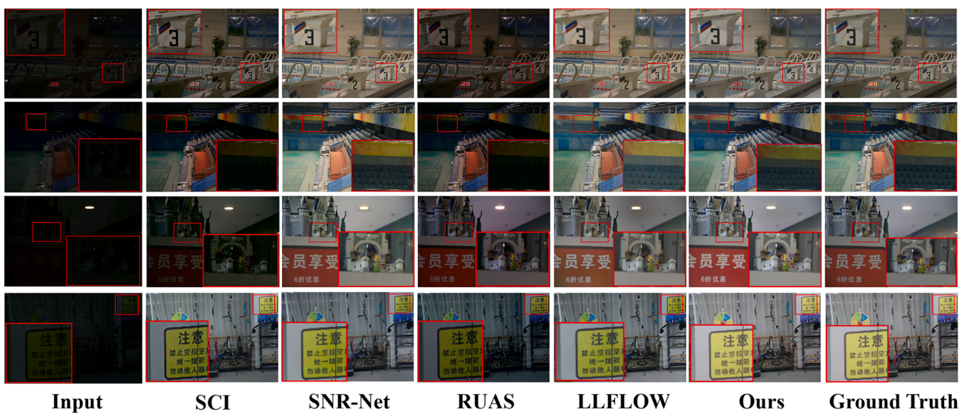


Fig. 4. Qualitative Comparisons on LOL [7] (row 1,2) and LOL-v2-real [46] (row 3,4) dataset.

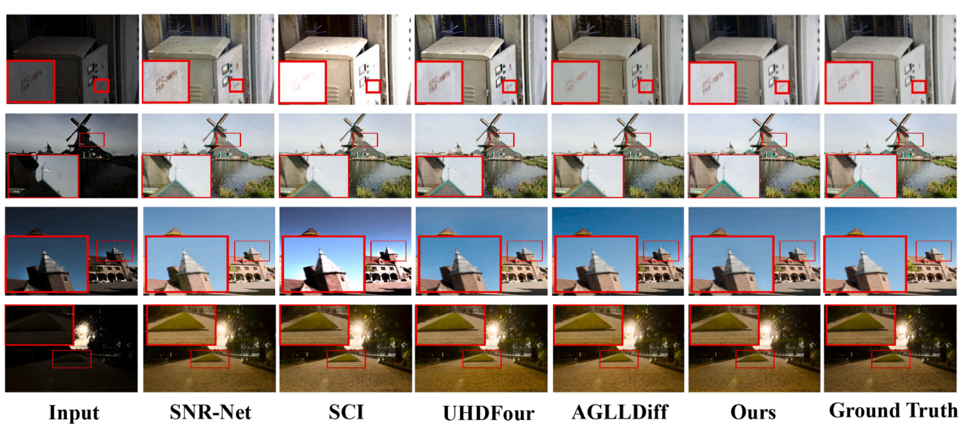


Fig. 5. Qualitative Comparisons on LSRW [50] (row 1) and LOL-v2-synthetic [46] (row 2,3,4) dataset.

studies but also thoroughly evaluates the adaptability of our method across different imaging conditions.

4.1.3. Metrics

For comprehensive performance evaluation, we employ three widely-adopted metrics: Peak Signal-to-Noise Ratio (PSNR) [51] for quantifying pixel-level fidelity, Structural Similarity Index (SSIM) [51] for assessing structural information preservation, and Learned Perceptual Image Patch Similarity (LPIPS) [52] for evaluating high-level perceptual consistency. This multi-dimensional assessment framework enables rigorous analysis of both low-level reconstruction accuracy and human visual perception quality.

4.2. Results and analysis

To comprehensively validate the effectiveness of our proposed method, we conducted extensive experiments to evaluate DLFN against state-of-the-art image enhancement approaches in recent years, including RUAS [34], EnlightenGAN [29], SNR-Net [11], URetinex-Net [6], CDEF [35], SCI [36], Uformer [37], UHDFour [38], LLFormer [39], GDP [40], WeatherDiff [41], SGF [43], DeVigNet [42], QuadPrior [44],

Table 4
Ablation studies of frequency-domain extraction and MBFM on LOL dataset [7].

Methods	FDE	MBFM	PSNR↑	SSIM↑	LPIPS↓	Params (M)
Case1	✗	✓	25.154	0.849	0.256	20.809
Case2	✓	✗	24.883	0.846	0.235	20.503
Case3	✗	✗	24.586	0.835	0.289	20.412
Ours	✓	✓	**25.632**	**0.853**	**0.216**	20.902

AGLLDiff [45], through quantitative metrics and qualitative visual assessments.

4.2.1. Quantitative comparison

To comprehensively assess the effectiveness of the proposed method, we evaluated it on several standard benchmark datasets: LOL [7], LOL-v2-real [46], LOL-v2-synthetic [46], and LSRW [50]. The experiments employed three widely recognized objective evaluation metrics: PSNR, SSIM, and LPIPS, and performed comparative analysis with state-of-the-art low-light image enhancement methods. The test results demonstrate that our method achieves significant advantages across all three key metrics, fully validating its outstanding performance in low-light image enhancement tasks. Specifically, as shown in Table 1, on the LOLv2-real dataset [46], our method ranked first in PSNR and SSIM evaluations and achieved the second-best LPIPS score, thoroughly verifying its capability to generate detailed images. For diffusion model-based methods such as GDP [40], WeatherDiff [41], and QuadPrior [44], our model also outperforms them in PSNR, SSIM, and LPIPS metrics. Furthermore, as indicated in Table 2, our method demonstrates strong competitiveness on the LOL dataset [7], highlighting the model's excellent generalization ability to restore high-quality images with satisfactory results. Additionally, as illustrated in Table 3, our method achieves superior performance on both the LOL-v2-synthetic [46] and LSRW [50] datasets, with its metrics ranking at the forefront. This consistently high performance across diverse benchmarks validates the robustness and exceptional LLIE capability of our approach.

4.2.2. Qualitative comparison

Quantitative comparisons across four benchmark datasets-LOL [7], LOL-v2-real [46], LOL-v2-synthetic [46], and LSRW [50]-as presented in Figs. 4 and 5, demonstrate the effectiveness of our method. The visual results indicate that existing LLIE methods [11,34,36,45,47] often struggle with inadequate exposure control, color distortion, and noise amplification. When processing low-light images, these approaches commonly introduce visual artifacts such as localized over- or underexposure, unnatural color casts, and amplified noise patterns. Specifically, RUAS [34] introduces noticeable block artifacts in extremely dark regions, SCI [36] yields an unnatural bluish-green color tint across the scene, and SNR-Net [11] produces severe noise grains in smooth areas. In contrast, our method delivers more stable and visually reliable enhancement results. In terms of exposure control, it achieves balanced illumination adjustment, effectively avoiding regions of excessive overexposure or underexposure. Regarding color restoration, our enhanced images exhibit natural and faithful color reproduction without perceptible color shifts. Furthermore, our method suppresses noise amplification while improving brightness, thereby preserving fine structural details and visual realism. These qualitative observations are consistent with the quantitative results, jointly validating the superiority of our approach in producing enhanced low-light images with a more natural appearance, realistic tone, and overall better visual perception.

4.3. Ablation studies

4.3.1. Effectiveness of frequency-domain extraction and MBFM

To validate the improvements made to the traditional Laplacian pyramid algorithm in this study, we conducted systematic ablation experiments on the LOL dataset to evaluate the contributions of the Frequency-Domain Extraction Method and MBFM. As shown in Table 4, when removing the Frequency-Domain Extraction Method in Case 1, the PSNR decreased by 0.26dB, respectively, confirming the critical role of this module in high-frequency component extraction. In Case 2, replacing MBFM with the traditional Laplacian pyramid upsampling and additive fusion operation led to a comprehensive decline in performance metrics, demonstrating the irreplaceable role of MBFM in feature fusion. Finally, version Case 3 completely reverted to the traditional Laplacian pyramid algorithm, resulting in noticeable drops across all metrics, which fully proves the overall effectiveness of our proposed improvement strategy. Together, these components break through the performance limitations of traditional pyramid algorithms, providing a superior solution for LLIE.

Table 5
Ablation studies of HFEM on the LOL dataset [7]. HFEM_v1 uses the SE Block framework, HFEM_v2 employs the spatial attention mechanism, and HFEM_v3 adopts the standard convolutional architecture.

Versions	PSNR↑	SSIM↑	LPIPS↓	Params (M)
HFEM_v1	24.683	0.849	0.245	20.824
HFEM_v2	24.465	0.851	0.238	20.836
HFEM_v3	23.879	0.842	0.307	20.892
HFEM	**25.632**	**0.853**	**0.216**	20.902

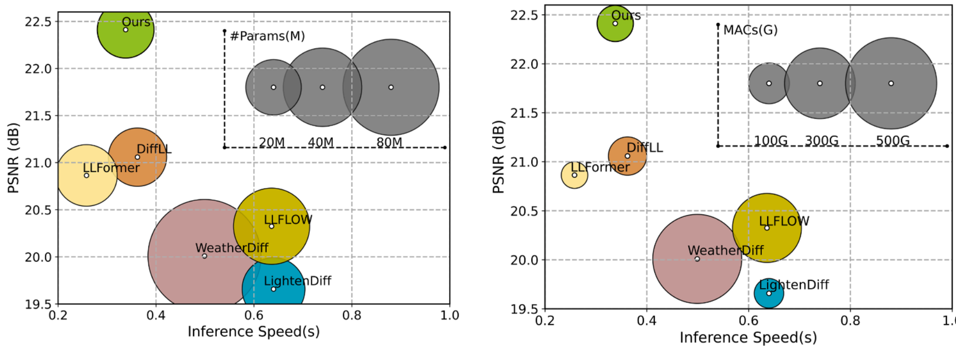
4.3.2. Effectiveness of HFEM

To comprehensively evaluate the synergistic mechanisms of HFEM components, we conducted systematic ablation studies on the LOL dataset [7]. As shown in Table 5, $HFEM\_v1$, which uses a basic SE Block [53] instead of the full structure, exhibited significant performance degradation. This confirms that single-channel attention cannot effectively capture spatial-domain detail features. $HFEM\_v2$, which implements feature enhancement using a basic spatial attention mechanism, showed reduced performance when relying solely on spatial attention, highlighting the inherent limitations of modeling inter-channel feature correlations with unidimensional spatial attention. $HFEM\_v3$, further simplified to a standard convolutional structure without any attention mechanism, resulted in a marked performance drop, conclusively validating the essential role of our HFEN. These experiments clearly demonstrate the complementary nature of our HFEN in combining channel and spatial information enhancement, with their synergistic interaction being indispensable for achieving optimal high-frequency enhancement.

Additionally, for the functional verification of internal components within HFEM, we conducted detailed performance evaluations, as shown in Table 6. The experimental results demonstrate that removing the Channel Shuffle (CS) operation leads to significant performance degradation, validating the importance of cross-channel information interaction for feature enhancement. Replacing the Dilated Convolutions (DC) in spatial dimensions with standard convolutions also reduces performance, indicating the critical role of multi-scale receptive fields in spatial feature extraction. Notably, the most severe performance deterioration occurs when the Median Pooling (MP) operation in the channel attention mechanism is removed, which conclusively proves that the synergistic fusion of global statistical information and local feature representation is the core factor ensuring enhancement effectiveness.

4.3.3. Effectiveness of the diffusion for low-frequency reconstruction

The core objective of employing a diffusion model in our framework is to address the challenging task of low-frequency reconstruction, which involves holistic illumination adjustment and color correction. To unequivocally demonstrate its necessity, we conducted a critical ablation study where the diffusion model was replaced with the high-frequency enhancement module (HFEM) for processing the



(a) Trade-off of Params, PSNR, and Speed.   (b) Trade-off of MACs, PSNR, and Speed.

Fig. 6. Model effiency trade-off on the LOL-v2-real dataset [46].

Table 6
Ablation studies of internal components in HFEM (CS: Channel Shuffle, DC: Dilated Convolutions, MP: Median Pooling).

Variants	PSNR↑	SSIM↑	LPIPS↓	Params (M)
w/o CS	25.349	0.849	0.223	20.896
w/o DC	25.562	0.851	0.219	20.892
w/o MP	25.076	0.845	0.227	20.897
Ours	**25.632**	**0.853**	**0.216**	20.902

Table 7
Why we employ diffusion models for the reconstruction of low-frequency features?.

Methods	LOL [7]			LOL-v2-real [46]		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
HFEM for Low-Freq	21.325	0.701	0.432	18.987	0.734	0.392
Diffusion for Low-Freq	**25.632**	**0.853**	**0.216**	**22.412**	**0.859**	**0.201**

Table 8
Ablation studies on loss functions.

Configurations	PSNR↑	SSIM↑	LPIPS↓
Full Model	25.632	0.853	0.216
w/o $\mathcal{L}_{low}$	24.325	0.821	0.325
w/o $\mathcal{L}_{detail}$	24.843	0.841	0.265
w/o $\mathcal{L}_{content}$	23.676	0.815	0.271

Table 9
Model complexity (Parameters (M) and MACs (G)) of existing diffusion-based methods. The best and the second-highest result are **bolded** and <u>underlined</u>.

Method	Params (M)	MACs (G)
LLFormer [39]	24.523	**70.134**
LLFLOW [47]	37.682	285.364
Restormer [16]	26.097	140.990
WeatherDiff [41]	82.963	474.934
DiffLL [10]	<u>22.058</u>	87.829
LightenDiff [17]	25.704	108.553
AGLLDiff [45]	99.329	170.238
DLFN (Ours)	**20.902**	<u>78.808</u>

low-frequency component. The results, quantitatively presented in Table 7, substituting the diffusion model with the HFEM leads to a substantial degradation in performance across all metrics on both LOLv1 and LOLv2 datasets. This demonstrates the necessity of employing diffusion models to process low-frequency information.

4.3.4. *Effectiveness of loss functions*

To assess the contribution of each loss function, we performed an ablation study comparing the full model to its variants, as summarized in Table 8. The results show that the full model achieves the best performance across all evaluation metrics. Removing the low-frequency component loss ($\mathcal{L}_{low}$) causes a significant drop in performance, with PSNR decreasing by approximately 1.3dB, highlighting its critical role in preserving structural coherence. Omitting the detail preservation loss ($\mathcal{L}_{detail}$) leads to a noticeable decline in perceptual quality, as indicated by an increase of about 0.05 in LPIPS. Excluding the content loss ($\mathcal{L}_{content}$) substantially impacts semantic consistency, resulting in a PSNR decrease of nearly 2dB. These findings demonstrate that each loss function targets a unique aspect of the enhancement task, and their combination is crucial for achieving optimal performance.

4.3.5. *Efficiency trade-off*

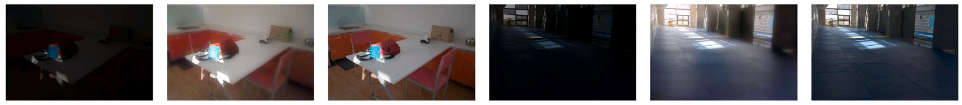
To comprehensively evaluate the efficiency of different LLIE methods, we analyze the efficiency trade-off between computational overhead and restoration performance. This analysis primarily considers three key metrics: PSNR as the quality indicator, inference speed as the time metric, and computational complexity represented by either the Parameters or the number of MACs.

As illustrated in Fig. 6, the trade-off analysis based on the parameter count (as shown in Fig. 6(a)) demonstrates the superior performance of our method. It achieves the highest PSNR while maintaining competitive inference speed. As described in Table 9, our method accomplishes this with 20.902M parameters, which is significantly smaller than other high-performing methods. For instance, WeatherDiff [41] has the largest bubble, requiring 82.963M parameters, yet delivers an inferior PSNR, indicating poor parameter efficiency. Similarly, LLFLOW [47] requires 37.682M parameters but achieves only 20.325dB PSNR. This comparison highlights the exceptional parameter efficiency and performance balance of our method.

The computational complexity analysis based on MACs, as shown in Fig. 6(b), further validates the efficiency advantages of our method. Our method has 78.8G MACs while maintaining the best performance-efficiency trade-off. In contrast, WeatherDiff [41] requires 474.9G MACs-approximately 6 times higher than our method-for significantly worse performance. LLFLOW [47] requires 285.4G MACs yet achieves lower PSNR, demonstrating inefficient computational utilization. LightenDiff [17], despite having moderate MACs, suffers from noticeable quality degradation.

5. Limitations and future works

While the proposed method demonstrates compelling performance on standard benchmarks, our analysis reveals certain limitations that warrant discussion and present clear pathways for future research.



Input Ours Ground Truth Input Ours Ground Truth

Fig. 7. Analysis of visual artifacts in challenging cases.

The main limitation of our approach lies in handling scenes with extremely high dynamic range, especially those containing both very dark regions and intense light sources. In these challenging cases-illustrated in Fig. 7-our method may occasionally produce minor over-enhancement or halo artifacts near sharp illumination transitions. These visual imperfections arise from two primary factors: first, the current fixed-frequency band decomposition is not always optimal for handling abrupt brightness changes, which may result in imperfect separation of signal and noise in transitional regions; second, our enhancement parameters are applied uniformly across the entire image, lacking spatial adaptation to local characteristics, which makes it difficult to achieve balanced enhancement in areas with extreme dynamic ranges.

These observations naturally suggest several promising avenues for future research. First, we plan to explore adaptive band selection mechanisms that dynamically adjust the decomposition strategy based on local image content, which would better accommodate abrupt changes in illumination. Second, investigating spatially variant enhancement strategies-modulating enhancement strength according to local luminance and texture-could help prevent over-enhancement in bright regions while still sufficiently amplifying dark areas. Finally, extending our framework to explicitly model and process scenes with extremely high dynamic range could greatly enhance robustness in the most challenging scenarios.

6. Conclusions

In this paper, we introduce the DLFN for low-light image enhancement, which addresses challenges such as detail loss and excessive noise by performing multi-scale enhancement and fusion. By enhancing the traditional Laplacian pyramid approach with frequency-domain feature extraction and the Multi-Scale Bilateral Fusion Module (MBFM), we achieve more comprehensive feature decomposition and fusion. The High-Frequency Enhancement Module (HFEM) restores high-frequency details using a hybrid attention mechanism, while a generative conditional diffusion model enhances low-frequency components, ensuring both detailed recovery and structural consistency. Experiments demonstrate that our framework achieves significant improvements in low-light image enhancement, preserving natural appearance and fine details across varying levels of degradation, showing strong stability and generalization capability. Nonetheless, our method occasionally produces artifacts near sharp illumination transitions, motivating future work on adaptive band selection to better handle abrupt brightness changes and extend the flexible framework to other related restoration tasks.

CRediT authorship contribution statement

Le Zhou: Writing – original draft, Methodology, Data curation; **Wenjie Li:** Writing – original draft, Investigation, Data curation; **Juncheng Li:** Writing – review & editing, Validation, Methodology, Investigation; **Guangwei Gao:** Writing – review & editing, Visualization, Supervision, Project administration, Formal analysis, Conceptualization; **Chia-Wen Lin:** Writing – review & editing, Investigation, Formal analysis.

Data availability

Data will be made available on request.

Declaration of competing 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] Z. Ying, Z. Lin, Z. Li, X. Huang, W. Ding, SCDF: seeing clearly through dark and fog, an adaptive semantic segmentation scheme for autonomous vehicle, IEEE Trans. Intell. Transp. Syst. 26 (7) (2025) 9640–9653.
[2] N. Yu, J. Wang, H. Shi, Z. Zhang, Y. Han, Degradation-removed multiscale fusion for low-light salient object detection, Pattern Recognit. 155 (2024) 110650.
[3] S. Sun, W. Ren, X. Gao, R. Wang, X. Cao, Restoring images in adverse weather conditions via histogram transformer, in: ECCV, Springer, 2024, pp. 111–129.
[4] M. Wang, J. Li, C. Zhang, Low-light image enhancement by deep learning network for improved illumination map, Comput. Vision Image Understanding 232 (2023) 103681.
[5] X. Yi, H. Xu, H. Zhang, L. Tang, J. Ma, Diff-Retinex: rethinking low-light image enhancement with a generative diffusion model, in: ICCV, 2023, pp. 12302–12311.
[6] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, J. Jiang, Uretinex-Net: retinex-based deep unfolding network for low-light image enhancement, in: CVPR, 2022, pp. 5901–5910.
[7] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-Light enhancement, in: BMVC, 2018, pp. 1–12.
[8] L. Ma, R. Liu, Y. Wang, X. Fan, Z. Luo, Low-light image enhancement via self-reinforced retinex projection model, IEEE Trans. Multimed. 25 (2022) 3573–3586.
[9] H. Zhou, W. Dong, X. Liu, Y. Zhang, G. Zhai, J. Chen, Low-light image enhancement via generative perceptual priors, in: AAAI, vol. 39, 2025, pp. 10752–10760.
[10] H. Jiang, A. Luo, H. Fan, S. Han, S. Liu, Low-light image enhancement with wavelet-based diffusion models, ACM Trans. Graph. (TOG) 42 (6) (2023) 1–14.
[11] X. Xu, R. Wang, C.-W. Fu, J. Jia, SNR-aware low-light image enhancement, in: CVPR, 2022, pp. 17714–17724.
[12] K. Shang, M. Shao, C. Wang, Y. Cheng, S. Wang, Multi-domain multi-scale diffusion model for low-light image enhancement, in: AAAI, vol. 38, 2024, pp. 4722–4730.
[13] D. Zhou, Z. Yang, Y. Yang, Pyramid diffusion models for low-light image enhancement, in: IJCAI, 199, 2023, pp. 1795–1803.
[14] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: NeurIPS, vol. 33, 2020, pp. 6840–6851.
[15] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, L. Van Gool, Repaint: inpainting using denoising diffusion probabilistic models, in: CVPR, 2022, pp. 11461–11471.
[16] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, Restormer: efficient transformer for high-resolution image restoration, in: CVPR, 2022, pp. 5728–5739.
[17] H. Jiang, A. Luo, X. Liu, S. Han, S. Liu, Lightendiffusion: unsupervised low-light image enhancement with latent-retinex diffusion models, in: ECCV, Springer, 2024, pp. 161–179.
[18] W. Li, H. Guo, X. Liu, K. Liang, J. Hu, Z. Ma, J. Guo, Efficient face super-resolution via wavelet-based feature enhancement network, in: ACM MM, 2024, pp. 4515–4523.
[19] L. Sun, Y. Bao, J. Zhai, J. Liang, Y. Zhang, K. Wang, D.P. Paudel, L. Van Gool, Low-light image enhancement using event-based illumination estimation, in: ICCV, 2025, pp. 6667–6677.
[20] W. Wu, W. Wang, Z. Wang, K. Jiang, Z. Li, For overall nighttime visibility: integrate irregular glow removal with glow-aware enhancement, IEEE Trans. Circuits Syst. Video Technol. 35 (1) (2024) 823–837.
[21] K. Jiang, Z. Wang, Z. Wang, C. Chen, P. Yi, T. Lu, C.-W. Lin, Degrade is upgrade: learning degradation for low-light image enhancement, in: Proceedings of the AAAI Conference on Artificial Intelligence, 36, 2022, pp. 1078–1086.
[22] C. Wang, J. Jiang, K. Jiang, X. Liu, Low-light face super-resolution via illumination, structure, and texture associated representation, in: Proceedings of the AAAI Conference on Artificial Intelligence, 38, 2024, pp. 5318–5326.
[23] Z. Li, H. Shu, C. Zheng, Multi-scale single image dehazing using Laplacian and Gaussian pyramids, IEEE Trans. Image Process. 30 (2021) 9270–9279.
[24] S. Li, Q. Hao, X. Kang, J.A. Benediktsson, Gaussian pyramid based multiscale feature fusion for hyperspectral image classification, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (9) (2018) 3312–3324.
[25] S. Lim, W. Kim, DSLR: deep stacked Laplacian restorer for low-light image enhancement, IEEE Trans Multimed. 23 (2021) 4272–4284.
[26] X. Wei, K. Chang, G. Li, M. Huang, Q. Qin, DLEN: deep Laplacian enhancement networks for low-light images, in: ICIP, IEEE, 2023, pp. 2120–2124.
[27] W. Li, X. Wang, H. Guo, G. Gao, Z. Ma, Self-supervised selective-guided diffusion model for old-photo face restoration, in: NeurIPS, 2025.

[28] W. Li, J. Shi, J. Han, H. Guo, Z. Ma, Seeing through the rain: resolving high-frequency conflicts in deraining and super-resolution via diffusion guidance, in: AAAI, 2026.

[29] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang, Enlightengan: deep light enhancement without paired supervision, IEEE Trans. Image Process. 30 (2021) 2340–2349.

[30] Z. Ni, W. Yang, H. Wang, S. Wang, L. Ma, S. Kwong, Cycle-interactive generative adversarial network for robust unsupervised low-light enhancement, in: ACM MM, 2022, pp. 1484–1492.

[31] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: CVPR, 2022, pp. 10684–10695.

[32] W. Li, H. Guo, Y. Hou, G. Gao, Z. Ma, Dual-domain modulation network for lightweight image super-resolution, IEEE Trans Multimed. (2025).

[33] W. Li, J. Li, G. Gao, W. Deng, J. Yang, G.-J. Qi, C.-W. Lin, Efficient image super-resolution with feature interaction weighted hybrid network, IEEE Trans. Multimed. 27 (2025) 2256–2267.

[34] R. Liu, L. Ma, J. Zhang, X. Fan, Z. Luo, Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement, in: CVPR, 2021, pp. 10561–10570.

[35] X. Lei, Z. Fei, W. Zhou, H. Zhou, M. Fei, Low-light image enhancement using the cell vibration model, IEEE Trans. Multimed. 25 (2022) 4439–4454.

[36] L. Ma, T. Ma, R. Liu, X. Fan, Z. Luo, Toward fast, flexible, and robust low-light image enhancement, in: CVPR, 2022, pp. 5637–5646.

[37] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, H. Li, Uformer: a general u-shaped transformer for image restoration, in: CVPR, 2022, pp. 17683–17693.

[38] C. Li, C.-L. Guo, M. Zhou, Z. Liang, S. Zhou, R. Feng, C.C. Loy, Embedding fourier for ultra-high-definition low-light image enhancement, in: ICLR, 2023, pp. 1–27.

[39] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, T. Lu, Ultra-high-definition low-light image enhancement: a benchmark and transformer-based method, in: AAAI, 37, 2023, pp. 2654–2662.

[40] B. Fei, Z. Lyu, L. Pan, J. Zhang, W. Yang, T. Luo, B. Zhang, B. Dai, Generative diffusion prior for unified image restoration and enhancement, in: CVPR, 2023, pp. 9935–9946.

[41] O. Özdenizci, R. Legenstein, Restoring vision in adverse weather conditions with patch-based denoising diffusion models, IEEE Trans. Pattern Anal. Mach. Intell. 45 (8) (2023) 10346–10357.

[42] S. Luo, X. Chen, W. Chen, Z. Li, S. Wang, C.-M. Pun, Devignet: high-resolution vignetting removal via a dual aggregated fusion transformer with adaptive channel expansion, in: AAAI, vol. 38, 2024, pp. 4000–4008.

[43] G. Li, B. Zhao, X. Li, Low-light image enhancement with sam-based structure priors and guidance, IEEE Trans. Multimed. 26 (2024) 10854–10866.

[44] W. Wang, H. Yang, J. Fu, J. Liu, Zero-reference low-light enhancement via physical quadruple priors, in: CVPR, 2024, pp. 26057–26066.

[45] Y. Lin, T. Ye, S. Chen, Z. Fu, Y. Wang, W. Chai, Z. Xing, W. Li, L. Zhu, X. Ding, Aglldiff: guiding diffusion models towards unsupervised training-free real-world low-light image enhancement, in: AAAI, 39, 2025, pp. 5307–5315.

[46] W. Yang, W. Wang, H. Huang, S. Wang, J. Liu, Sparse gradient regularized deep retinex network for robust low-light image enhancement, IEEE Trans. Image Process. 30 (2021) 2072–2086.

[47] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, A. Kot, Low-light image enhancement with normalizing flow, in: AAAI, vol. 36, 2022, pp. 2604–2612.

[48] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, Y. Zhang, Retinexformer: one-stage retinex-based transformer for low-light image enhancement, in: ICCV, 2023, pp. 12504–12513.

[49] Y. Wen, P. Xu, Z. Li, W.X. ATO, An illumination-guided dual attention vision transformer for low-light image enhancement, Pattern Recognit. 158 (2025) 111033.

[50] J. Hai, Z. Xuan, R. Yang, Y. Hao, F. Zou, F. Lin, S. Han, R2rnet: Low-light image enhancement via real-low to real-normal network, J. Vis. Commun. Image Represent. 90 (2023) 103712.

[51] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[52] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: CVPR, 2018, pp. 586–595.

[53] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: CVPR, 2018, pp. 7132–7141.