

PRNet: Low-Light Image Enhancement Based on Fourier Transform

Jiayu Zhang¹, Xiaohua Wang¹, Yingjian Li¹, *Graduate Student Member, IEEE*, and Wenjie Wang¹

Abstract—Low-light image enhancement (LLIE) techniques constitute a significant approach for enhancing image brightness effectively while preserving image details. In this article, PRNet is proposed, which is a novel lightweight LLIE network that leverages the Fourier transform, performing LLIE in two stages. In the first stage, a pixel enhancement network (PENet) enhances the brightness of the low-light image (LLI) through a dense skip-connection structure. This structure incorporates a custom-designed Fourier-based brightness enhancement block (FBEB). In the second stage, a refinement and restoration network (RRNet) processes the output from the first stage, further restoring image details. Detailed refinement is achieved using a dual-branch UNet structure, incorporating a bidirectional frequency-domain cross-attention solver (BFDSC) to optimize image quality. To thoroughly assess the performance of the proposed PRNet, nine well-established benchmark datasets were employed for detailed quantitative and qualitative evaluations. The experimental results show that PRNet achieves high-quality image enhancement at significantly reduced computational complexity.

Index Terms—Amplitude component, cross-attention, Fourier transform, low-light image enhancement (LLIE), phase component.

I. INTRODUCTION

OPTICAL light images captured in low-light environments often exhibit diminished clarity and insufficient contrast. These challenges hinder measurement accuracy in precision-requiring fields, such as scientific imaging [1] and industrial inspection [2], [3]. They also impair the performance of downstream tasks, such as pose estimation [4], action recognition [5], semantic segmentation [6], [7], and object detection [8], [9].

Low-light image enhancement (LLIE) techniques aim to improve brightness, retain image details, and mitigate issues such as noise, underexposure, and texture loss. For this purpose, several LLIE methods have been proposed, which,

however, exhibit notable limitations. Histogram equalization approaches [10] often fail to account for illumination conditions, leading to artifacts and color distortion. Retinex-based methods [11] rely on the assumption that images can be ideally decomposed into reflectance and illumination components, which often fail in complex scenes, resulting in ineffective brightness adjustments. Convolutional neural network (CNN)-based methods [12], [13] perform LLIE by directly learning pixelwise mappings between low-light images (LLIs) and their ground-truth counterparts. These methods have limitations in capturing long-range dependencies, resulting in poor nonlocal self-similarity and reduced color fidelity. Transformer-based approaches [14], [15], [16] present a promising solution to the limitations of CNNs, but their computational complexity increases quadratically with input dimensions, which makes these methods computationally expensive and less efficient for practical applications.

Building on these insights, recent methods [17], [18], [19], [20] incorporate the Fourier transform to model long-range dependencies effectively through the extraction of global image information, while retaining a relatively low computational complexity. In addition, the method presented in [21] computes the similarity between the query and key matrices using elementwise product operation in the frequency domain. This approach reduces the spatial complexity per feature channel from $O(N^2)$ to $O(N)$ and the temporal complexity from $O(N^2C)$ to $O(NC \log N)$, where N is the number of pixels and C represents the number of features. These methods often decompose images into amplitude and phase components, employing intricate processing structures tailored to each component. In the frequency domain, most brightness information is concentrated in the amplitude component, whereas the phase component primarily captures structural or detailed information and is unrelated to brightness. By enhancing the amplitude component in the frequency domain, these methods have demonstrated significant improvements in LLIE performance; however, they may introduce redundant information, which leads to artifacts and color deviations. Additionally, designing independent structures for each component significantly increases the computational overhead, thus reducing their practicality. Thus, there is a pressing need for an efficient and streamlined Fourier-based LLIE method to address these challenges.

In this article, we introduce PRNet, a lightweight LLIE network leveraging the Fourier transform. Experiments and ablation studies across nine datasets demonstrate that PRNet achieves outstanding results with only 1.93M parameters and

Received 1 December 2024; revised 11 February 2025; accepted 1 March 2025. Date of publication 2 April 2025; date of current version 22 April 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 51905405 and in part by the Beilin District 2023 Applied Technology Research and Development Reserve Project under Grant GX2306. The Associate Editor coordinating the review process was Dr. Yu Yang. (Corresponding author: Xiaohua Wang.)

Jiayu Zhang, Xiaohua Wang, and Wenjie Wang are with the School of Electronics and Information and Xi'an Polytechnic University Branch of Shaanxi Artificial Intelligence Joint Laboratory, Xi'an Polytechnic University, Xi'an 710600, China (e-mail: 1252784756@qq.com; wangxiaohua@xpu.edu.cn; wangwenjie@xpu.edu.cn).

Yingjian Li is with the School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China, and also with the Shaanxi Key Laboratory of Clothing Intelligence, Xi'an 710048, China (e-mail: yingjianli@stu.xpu.edu.cn).

Digital Object Identifier 10.1109/TIM.2025.3557114

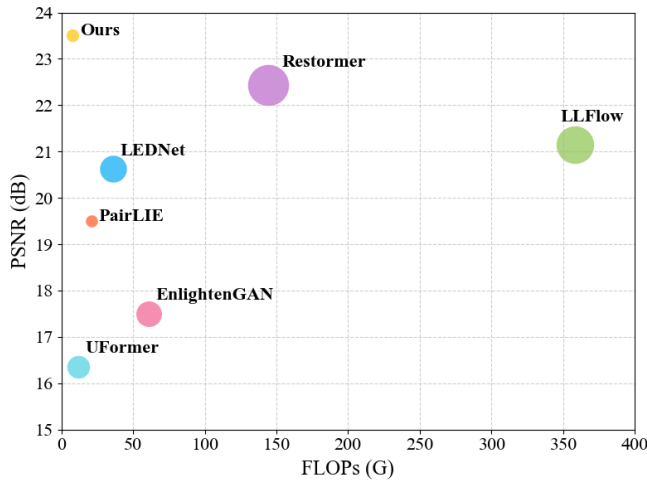


Fig. 1. Comparative analysis of PSNR, FLOPs, and network parameters against leading methods on the LOLv1 dataset. The size of the circles represents the number of network parameters.

8.08G floating point operation (FLOPs), striking an effective balance between performance and efficiency. PRNet provides an efficient and practical LLIE solution for precision applications, such as scientific imaging and industrial inspection through its balanced computational efficiency and performance. In Fig. 1, PRNet achieves superior performance with fewer parameters and lower FLOPs. The main contributions of this work are summarized as follows.

- 1) PRNet, a lightweight two-stage LLIE network based on the Fourier transform, is presented for the first time. The algorithm's first stage focuses on initial brightness enhancement, while the second stage emphasizes detailed restoration.
- 2) In the first stage, a Fourier-based brightness enhancement block (FBEB) is designed, which focuses on extracting brightness-related features from the amplitude component. The FBEB is incorporated into the dense skip-connection-structured pixel enhancement network (PENet), enabling the initial brightness enhancement of LLIs.
- 3) In the second stage, a bidirectional frequency-domain cross-attention solver (BFDCS) is designed and integrated into a dual-branch UNet structure, forming the refinement and restoration network (RRNet). This network combines a fast Fourier transform (FFT) and cross-attention mechanisms to enhance branch interactions and reduce complexity, thereby refining image details.
- 4) The proposed method is evaluated on nine datasets using multiple quantitative metrics, demonstrating high efficiency with fewer parameters and reduced computational requirements.

II. RELATED WORK

In recent years, LLIE has garnered significant attention as a research area due to its wide-ranging applications in industrial inspection, scientific imaging, and pose estimation.

This section provides a concise review of existing LLIE methods and examines their potential limitations, offering a foundation for understanding the contributions and innovations presented in this work.

A. Low-Light Image Enhancement

1) *Traditional Methods:* Traditional LLIE methods often rely on histogram equalization [22] and Retinex theory [23] to improve image quality. These methods often overlook illumination factors, which can lead to artifacts, color distortion, and detailed loss. Moreover, noise amplification further degrades the results, producing inconsistent visuals compared to normal lighting conditions.

2) *Deep Learning Methods:* Since Lore et al. [24] first introduced a CNN for LLIE, deep learning-based methods [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40] have become widely adopted in the field. For instance, Wang et al. [41] proposed DeepUPE, a CNN for enhancing underexposed images. While effective to some extent, this method still has room for improvement in restoring details in extremely dark regions and lacks reliable denoising capabilities. Jiang et al. [42] developed an unsupervised generative adversarial network, EnlightenGAN, which eliminates the need for paired data, offering greater flexibility in terms of training data requirements. EnlightenGAN integrates global-local discriminators, self-regularized perceptual loss, and attention mechanisms to enhance overall brightness and contrast effectively. While effective, EnlightenGAN emphasizes local features, which can result in global coherence inconsistencies after enhancement. Moreover, the network may encounter instability during training. Xu et al. [43] integrated a signal-to-noise ratio prior to the transformer architecture, dynamically enhancing pixels through spatial transformations. The method captures long-range dependencies across different regions, achieving better enhancement performance compared to CNN. However, this method requires substantial memory and computational resources, which presents a challenge.

B. Frequency Information

In the field of computer vision, some methods have been widely applied to various tasks using the Fourier transform, such as person reidentification, semantic segmentation, image super-resolution, and LLIE. Liu et al. [44] employed a 3-D discrete cosine transform to analyze features in the spatiotemporal frequency domain directly, establishing an efficient mapping between the spatiotemporal and frequency domains. This method demonstrated superior performance in person reidentification. Zhang et al. [45] proposed a frequency-domain self-attention network that processes features across different frequency bands effectively, improving the efficiency and performance of semantic segmentation. Jiang et al. [46] introduced FABNet, which enhances single-image super-resolution performance using frequency-aware wavelet decomposition and a dynamic binarization strategy.

For LLIE, Li et al. [47] developed UHDFour, a method that processes images in the frequency domain by decomposing them into amplitude and phase components, thus enhancing

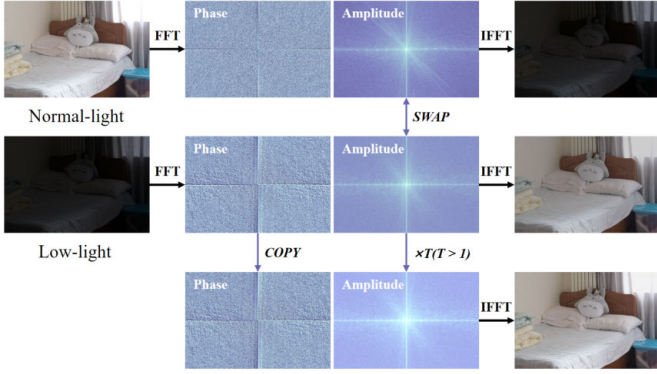


Fig. 2. Exchange of the amplitude components between an LLI and a normal-light image to generate new combined results; increasing the amplitude component of the LLI by a constant factor T results in an enhanced outcome. The intensity of the amplitude component corresponds to the image brightness level. FFT: fast Fourier transform and IFFT: inverse FFT.

brightness and suppressing noise effectively. They also introduced the first real 4k LLI dataset, which provides a foundation for LLIE research in ultrahigh-definition images. Wang et al. [48] presented FourLLIE, an LLIE approach that utilizes the Fourier transform effectively. By integrating frequency-domain information with spatial features, the method demonstrates notable improvements in LLIE performance. The method also incorporates a signal-to-noise ratio map to combine global image information with local spatial details, enabling the incremental refinement of image features.

Current LLIE methods primarily combine the Fourier transform with CNN-based frameworks. These methods often employ complex network structures, which can introduce redundant information and increase computational costs. CNNs are proficient in extracting local features but exhibit limitations in modeling global dependencies. Integrating the global modeling capabilities of transformers with the Fourier transform enables the efficient interaction between frequency and spatial information components. This approach facilitates the development of a streamlined and practical LLIE framework with substantial value for real-world applications.

III. METHODOLOGY

In this section, the theoretical foundations of the Fourier transform are briefly introduced. Then, the architecture of PRNet is presented and the design considerations for its key components are discussed, along with the design of the loss functions and their role in optimizing image quality.

A. Fourier Transform

First, the theoretical foundations of the Fourier transform are introduced to facilitate the understanding of subsequent discussions. Given an input image $f(h, w)$ of shape $H \times W$, where h and w denote the pixel coordinates, its representation in the frequency domain is $X(u, v)$. In the frequency domain, $X(u, v)$ is complex valued and is expressed as follows:

$$X(u, v) = R(X(u, v)) + jI(X(u, v)). \quad (1)$$

The amplitude $A(X(u, v))$ and phase $P(X(u, v))$ can be represented as follows:

$$A(X(u, v)) = \sqrt{R^2(X(u, v)) + I^2(X(u, v))} \quad (2)$$

$$P(X(u, v)) = \tan^{-1}\left(\frac{I(X(u, v))}{R(X(u, v))}\right). \quad (3)$$

To enhance brightness, we primarily focus on the amplitude component $A(X(u, v))$. In Fig. 2, the intensity of the amplitude component is positively correlated with the brightness level of the image. First, new brightness combinations are generated through the swapping of the amplitude components of LLIs and normal-light images. This operation significantly enhances the brightness of LLIs while preserving their original structural details. Subsequently, the brightness of LLIs can be further enhanced by amplifying the amplitude component using a factor T . The experimental results demonstrate that enhancing the amplitude component directly impacts the brightness level of the image. Therefore, in the network designed in this article, in the first stage, PENet extracts features from amplitude components, which can enhance brightness while obtaining the clearest image details.

The image $f(h, w)$ is transformed into its frequency-domain representation $X(u, v)$ through the Fourier transform, which is defined as follows:

$$X(u, v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} f(h, w) e^{-2\pi j \left(\frac{hu}{H} + \frac{wv}{W} \right)} \quad (4)$$

where $e^{-2\pi j \left(\frac{hu}{H} + \frac{wv}{W} \right)}$ is the complex exponential term used to convert the spatial information of the image into frequency information, and $(1/\sqrt{HW})$ is a normalization factor for the Fourier transform.

Finally, the corresponding image in the spatial domain is obtained by applying the inverse Fourier transform to $X(u, v)$, which can be expressed as

$$f(h, w) = \frac{1}{\sqrt{HW}} \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} X(u, v) e^{2\pi j \left(\frac{hu}{H} + \frac{wv}{W} \right)}. \quad (5)$$

In addition to the Fourier transform theory mentioned above, an elementwise product operation is used in the frequency domain to calculate the scaled dot-product attention. This reduces the computational complexity effectively. In Fig. 3, PRNet performs LLIE in a two-stage manner. In the first phase, PENet employs a dense skip-connection structure composed of FBEB to introduce richer global image information and realize the initial pixel-level brightness enhancement of LLIs. In the second phase, RRNet uses a two-branch UNet architecture composed of BFDCS, which enhances the information interaction between branches and reduces the computational complexity. RRNet further refines and merges the two outputs of the first phase, resulting in a clear and naturally enhanced image. The detailed implementation is discussed in Sections III-B–III-D.

B. PENet

Fig. 3 shows that PENet takes LLIs as input and extracts features using a dense skip-connection structure consisting of

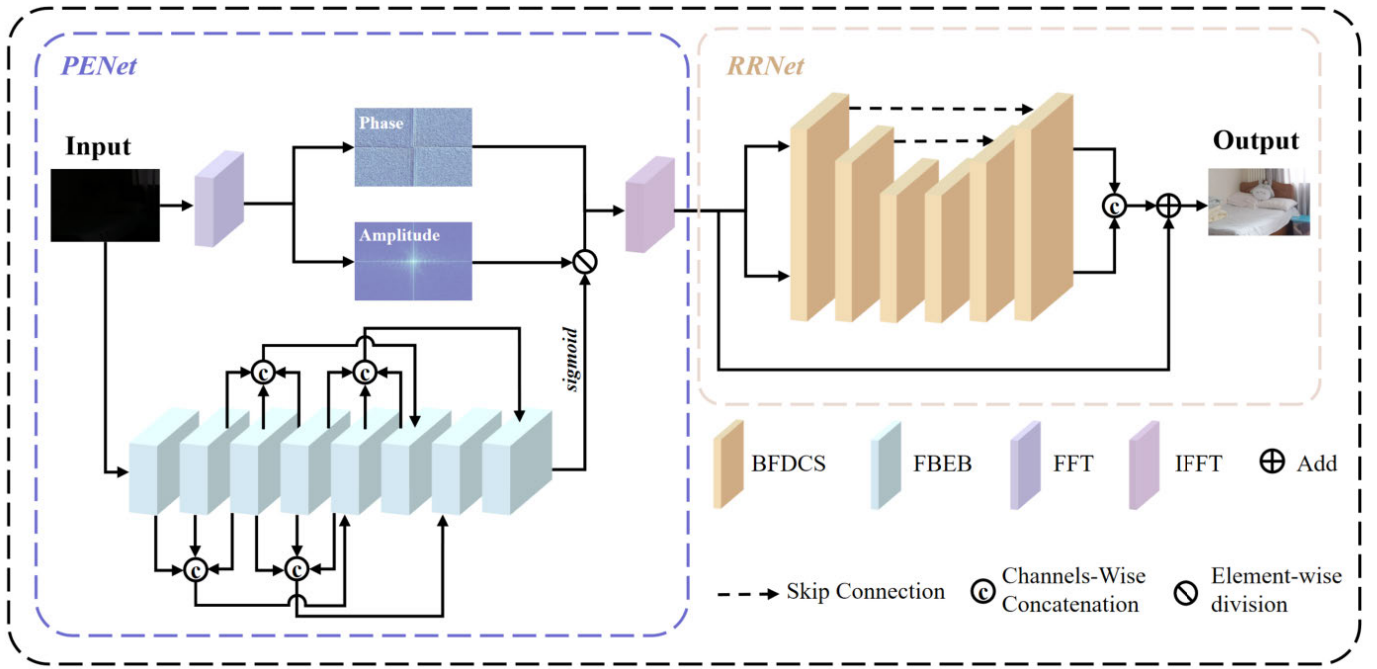


Fig. 3. PRNet consists of two main parts. In the first stage, PENet employs a dense skip-connection structure to extract features related to brightness. The output of dense skip-connection structure is applied to modify the amplitude component of the original input image by division. The adjusted amplitude is combined with the original phase component to perform the IFFT and realize the initial brightness enhancement. In the second stage, RRNet adopts a dual-branch UNet architecture, integrated with the proposed BFDCS to process the two channel-split outputs from PENet. This step refines and restores the image with greater precision.

eight FBEB blocks. This structure integrates features from multiple layers, which enables information fusion and feature propagation to be performed more closely. For example, the fifth FBEB receives outputs from the first, second, third, and fourth blocks, which are concatenated through channel merging to further enrich the frequency-domain information.

In Fig. 4, the input of each FBEB is first subjected to initial feature extraction with a 1×1 convolutional layer. Then, the input is transformed into the frequency domain via FFT and decomposed into its amplitude and phase components. For the amplitude component, a 1×1 deformable convolutional layer is used to further extract the features related to the brightness information. Compared with standard convolution, deformable convolution improves the flexibility of feature extraction and significantly improves the brightness enhancement performance. In addition, the phase component's original form is maintained without additional processing, as it mainly represents the structure and spatial details of the image and has no direct relationship with the brightness information. The processed amplitude components and unaltered phase components are converted back into the spatial domain using the inverse FFT (IFFT) and are fused with the original input features using residual connections, which further enhances the network's ability to retain detail. Finally, the output of the dense skip-connection structure is mapped to the range $[0, 1]$ via the sigmoid function, which is divided by element into the amplitude components of the original input image. The amplitude component is then combined and transformed with the phase component of the original input image, and the output-enhanced brightness is obtained using the IFFT.

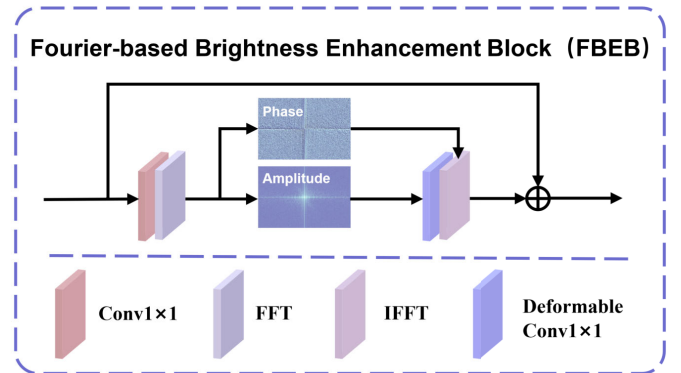


Fig. 4. Structural diagram of the FBEB. The 1×1 convolutional layer is used for initial feature extraction, and the 1×1 deformable convolutional layer extracts features related to brightness from the amplitude component. The phase component remains unchanged and is fused with the processed amplitude component through a residual connection to preserve structural details.

In addition, the number of FBEBs directly affects the magnitude of the amplitude component and the degree of brightness enhancement. Too many FBEBs can result in a halo effect or excessive smoothing of the details. Given this, we chose eight FBEB blocks as a design choice.

C. RRNet

UNet is the model on which this article's RRNet was based. The encoder consists of three BFDCS blocks and a down-sampling layer, while the decoder consists of three BFDCS blocks and an upsampling layer. At this stage, the input features are split into two separate branches that serve as inputs

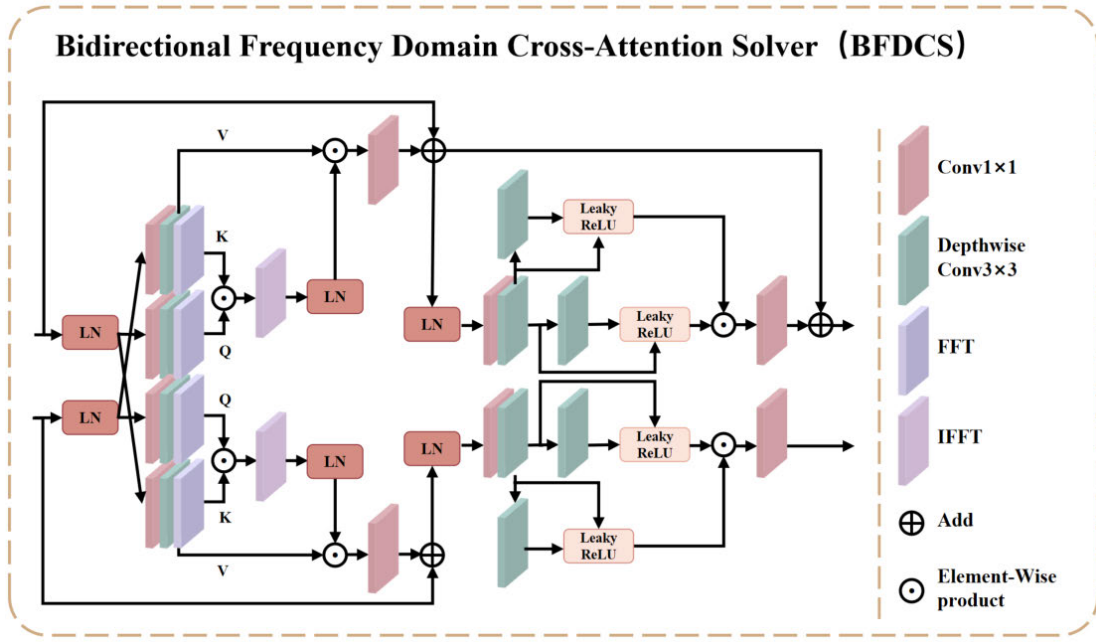


Fig. 5. BFDCS module consists of two branches. The upper branch includes an additional residual connection, and the remaining structure is identical to the lower branch. The two branches use the FFT to calculate the feature correlation and reconstruct the processed features in the frequency domain to the spatial domain through the IFFT. Each branch then refines its features independently using a postprocessing structure consisting of a 1×1 convolutional layer, a 3×3 depthwise convolutional layer, and leaky ReLU activation.

to RRNet. This dual-path design facilitates enhanced feature interaction and supports better refinement during subsequent processing. The two feature branches feed into the first BFDCS module, which performs initial image refinement and restoration. The second BFDCS module processes the output of the previous module and applies upsampling or downsampling to accommodate multiscale feature improvements. Finally, the two outputs from the sixth BFDCS are concatenated and combined with the original RRNet input image through a residual connection.

Given the input feature X with a spatial resolution of $H \times W$ pixels and C channels, a typical visual transformer applies linear transformations W_q , W_k , and W_v to the input feature, mapping it to the query (Q), key (K), and value (V) vectors for information interaction. The transformation matrices W_q , W_k , and $W_v \in R^{D \times D_k}$ are the learnable weight matrices designed to map input features into different vector spaces required for the attention mechanism. Based on these mappings, the scaled dot-product attention is defined as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)V \quad (6)$$

where K^T represents the transpose of K , which is used to compute the correlation matrix between the Q and K vectors. The term $\sqrt{D_k}$ acts as a scaling factor to prevent excessively large dot-product values, which can lead to vanishing or unstable gradients when the dimension of K is large. The $\text{softmax}(\cdot)$ function normalizes the correlation matrix along each row and ensures that the sum of the weights is one, which helps in the probabilistic interpretation of the weights. The computation of the attention map relies on the matrix multiplication QK^T ,

which incurs spatial and temporal complexities of $O(N^2)$ and $O(N^2C)$, respectively.

According to the convolution theorem, the convolution of two signals in the spatial domain is equivalent to an element-wise product operation in the frequency domain. This property allows us to not only reduce computational complexity, but also decrease memory consumption. Specifically, operations in the frequency domain reduce the spatial complexity of each feature channel from $O(N^2)$ to $O(N)$ and the temporal complexity from $O(N^2C)$ to $O(NC \log N)$. In Fig. 5, the BFDCS module is proposed for dual-path input features. This solver designates one feature branch as Q and the other as both K and V , leveraging their complementary information to enhance the interaction of frequency-domain features. For example, consider the lower branch shown in Fig. 5. We first calculate Q , K , and V using a 1×1 convolutional layer and a 3×3 depthwise convolutional layer. Next, the FFT is applied to Q and K to estimate their correlation in the frequency domain, which is expressed as follows:

$$P = \text{FFT}^{-1}(\text{FFT}(Q)\overline{\text{FFT}(K)}) \quad (7)$$

where $\overline{\text{FFT}(\cdot)}$ indicates the conjugate transpose. After applying the IFFT, the features are further processed as follows:

$$\text{Att}_{\text{BFDCS}} = \text{LN}(P)V \quad (8)$$

where $\text{LN}(\cdot)$ represents the layer normalization. Let the initial input for a single branch be X . After extracting features via a 1×1 convolutional layer, X is combined with the original input via a residual connection to produce X_1

$$X_1 = \text{Conv1} \times 1(\text{Att}_{\text{BFDCS}}) + X. \quad (9)$$

Subsequently, X_1 undergoes the following steps:

$$X_2 = \text{LN}(X_1) \quad (10)$$

$$X_3 = \text{DC Conv3} \times 3(\text{Conv1} \times 1(X_2)) \quad (11)$$

$$X_4 = \text{Leaky}(\text{DC Conv3} \times 3(X_3)) + X_3 \quad (12)$$

$$X_5 = \text{Leaky}(\text{DC Conv3} \times 3(X_3)) + X_3 \quad (13)$$

$$X_6 = \text{Conv1} \times 1(X_4 \cdot X_5) \quad (14)$$

where $\text{Conv1} \times 1$ refers to the 1×1 convolutional layer used for initial feature extraction and channel dimension reduction, while $\text{DC Conv3} \times 3$ represents the 3×3 depthwise convolutional layer, designed to significantly reduce computational complexity while enhancing feature representation. Additionally, $\text{Leaky}(\cdot)$ denotes the leaky ReLU activation function, which alleviates the vanishing gradient problem by allowing the flow of small gradients. After completing the above operations, the final output of the lower branch is X_6 . The only difference between the upper and lower branches lies in an additional residual connection in the upper branch, with all other operations remaining identical. Compared to traditional cross-attention solvers, BFDSCS integrates richer frequency-domain information, enabling simultaneous image detail recovery and computational complexity reduction, thereby improving the network's efficiency in handling complex LLIE.

D. Loss Function

The goal of LLIE is to perform complex image operations such as brightness enhancement, detailed restoration, and color correction, without introducing noise. For this purpose, a joint loss function is adopted to improve the visual quality of the output image more effectively.

1) L_1 Loss: To achieve accurate reconstruction, L_1 loss is employed to measure the per-pixel absolute difference between the enhanced image and the ground-truth image. This loss ensures the accuracy of brightness and color information in the enhanced image, serving as a fundamental metric for optimizing global consistency. L_1 loss is defined as follows:

$$L_1 = \frac{1}{N} \sum_{i=1}^N \|I_{\text{pred},i} - I_{\text{gt},i}\| \quad (15)$$

where $I_{\text{pred},i}$ and $I_{\text{gt},i}$ represent the pixel values at position i in the enhanced and ground-truth images, respectively; N is the total number of pixels in the image; and $\|\cdot\|$ represents the absolute value operation.

2) *Edge Loss*: Edge loss extracts edge features from images using Laplacian filtering and measures the structural consistency between the enhanced and ground-truth images. This loss is highly sensitive to detailed variations and focuses on edge information in the image, thereby improving detailed restoration performance. The equation is defined as

$$L_e = \|\nabla I_{\text{pred}} - \nabla I_{\text{gt}}\|_2 \quad (16)$$

where ∇I_{pred} and ∇I_{gt} denote the gradient features of the enhanced and ground-truth images obtained via Laplacian filtering. The operator $\|\cdot\|_2$ refers to the L2-norm, which quantifies the Euclidean distance between the gradient feature maps.

3) *Perceptual Loss*: Perceptual loss computes the feature discrepancies between the enhanced image and the ground-truth image by leveraging high-level feature representations extracted from a pretrained VGG19 network. This loss results in generated images of significantly enhanced visual quality, aligning them more closely with human perception. Perceptual loss is defined as follows:

$$L_p = \sum_l w_l \cdot \|\phi_l(I_{\text{pred}}) - \phi_l(I_{\text{gt}})\|_1 \quad (17)$$

where I_{pred} denotes the enhanced image generated by the network, and I_{gt} represents the ground-truth image. $\phi_l(\cdot)$ indicates the feature mapping of the l th layer in the VGG network, which captures high-level semantic features. w_l is the weight assigned to the l th layer, and $\|\cdot\|_1$ refers to the L_1 -norm, which calculates the sum of absolute differences elementwise between the feature maps.

4) *SSIM Loss*: Structural similarity (SSIM) loss evaluates the similarity between the enhanced image and the ground-truth image in terms of brightness, contrast, and structure. It is used to optimize the visual fidelity of the generated image. The equation for SSIM loss is as follows:

$$L_{\text{SSIM}} = 1 - \text{SSIM}(I_{\text{pred}}, I_{\text{gt}}) \quad (18)$$

where $\text{SSIM}(I_{\text{pred}}, I_{\text{gt}})$ is defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (19)$$

where μ_x and μ_y are the local means of images x and y , respectively, σ_x^2 and σ_y^2 are their local variances, σ_{xy} is the local covariance between x and y , and C_1 and C_2 are the stability constants to avoid division by zero.

The total loss function is formulated by combining the four aforementioned functions

$$L_{\text{total}} = \lambda_1 L_1 + \lambda_e L_e + \lambda_p L_p + \lambda_{\text{SSIM}} L_{\text{SSIM}} \quad (20)$$

where λ_1 , λ_e , λ_p , and λ_{SSIM} are the weight parameters used to balance the contributions of each loss term. In this study, these weights were empirically set to $\lambda_1 = 0.6$, $\lambda_e = 50.0$, $\lambda_p = 0.01$, and $\lambda_{\text{SSIM}} = 1.0$. The carefully designed combined loss function accounts for various degradation factors in LLIs, enabling significant improvements in the overall visual quality of generated images under uncertain imaging conditions.

IV. EXPERIMENTS

In this section, PRNet's performance is evaluated on multiple datasets and compared to that of state-of-the-art methods.

A. Dataset and Implementation Details

The proposed framework was evaluated on nine datasets: LOLv1 [49], LOLv2 [50], DICM [51], LIME [23], MEF [52], NPE [53], VV [54], SICE [55] (including the Mix and Grad test sets [56]), and Sony-Total-Dark.

1) *LOL*: The LOL dataset is available in two versions, v1 and v2. LOLv1 contains 485 pairs of low-/normal-light images for training and 15 pairs for testing. LOLv2 is divided into two subsets: LOLv2-real and LOLv2-synthetic. LOLv2-real includes 689 pairs of low-/normal-light images for training and 100 pairs for testing. LOLv2-synthetic contains 900 pairs of images for training and 100 pairs for testing. PRNet was trained and tested separately on the three datasets in the LOL series.

2) *SICE*: SICE [55] (including the Mix and Grad test sets [56]) contains 589 pairs of low- and normal-light images. The 70% of the dataset is the training set; 10% is the validation set; and 20% is the test set. PRNet was trained on the SICE training set, while testing was conducted on the SICE-Mix and SICE-Grad datasets.

3) *Sony-Total-Dark*: Sony-Total-Dark is a subset of the SID dataset [57], captured using Sony cameras. This subset contains 2697 pairs of short- and long-exposure RAW images. To simulate a more realistic and challenging low-light environment, RAW images were converted to sRGB format for training. The conversion followed the scripts provided by the SID dataset. Gamma correction was not applied during the conversion process.

4) *Unpaired Datasets*: Five benchmark datasets were utilized for evaluation: DICM, LIME, MEF, NPE, and VV. These datasets are unpaired, meaning that they do not provide corresponding normal-light images as ground truth. Instead, they serve as benchmarks to assess the generalization capability of LLIE methods in real-world low-light scenarios.

5) *Implementation Details*: PRNet was trained and tested on a Linux-based personal computer equipped with an NVIDIA 3090 GPU. The model was implemented using PyTorch and optimized with the Adam optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). The initial learning rate was set to 1×10^{-4} and gradually reduced to 1×10^{-7} following a cosine annealing schedule. For the LOLv1 and LOLv2 datasets, the patch and batch sizes were set to 256 and 10, respectively, and PRNet was trained for 1000 epochs. Additionally, PRNet was trained and tested on the SICE and Sony-Total-Dark datasets with the same patch and batch size settings for 1000 epochs. The DICM, LIME, MEF, NPE, and VV datasets consisted of several unpaired real-world LLIs and were used exclusively for testing.

6) *Evaluation Metrics*: The visual quality of the resulting enhanced images on the paired datasets LOLv1, LOLv2, SICE, and SID was evaluated using the following metrics: peak signal-to-noise ratio (PSNR), SSIM index, learned perceptual image patch similarity (LPIPS), normalized mean squared error (NMSE), mean absolute error (MAE), and universal image quality index (UIQI). PSNR measures the improvement in signal-to-noise ratio after enhancement, with higher values indicating better performance. SSIM evaluates the similarity between the enhanced and reference images in terms of brightness, contrast, and structure, and values closer to one indicate better quality. LPIPS measures the perceptual similarity between images, and lower values indicate better perceptual fidelity of the enhanced image.

NMSE assesses pixel-level differences between the enhanced and reference images, and smaller values indicate smaller errors. MAE calculates the mean absolute difference between the pixels of two images, and lower values indicate smaller pixel deviations. UIQI quantifies the consistency between the enhanced and reference images in terms of brightness, contrast, and structure, with values closer to one indicating better enhancement.

For the unpaired datasets (DICM, LIME, MEF, NPE, and VV), we used the blind/referenceless image spatial (BRIS) quality evaluator and the natural image quality evaluator (NIQE) to assess image quality. BRIS evaluates whether an image exhibits texture and noise characteristics consistent with natural images, and lower values indicate better quality. NIQE measures the naturalness of image quality by analyzing natural scene statistics; lower values indicate better alignment with natural image characteristics.

B. Evaluation on Image Enhancement

1) *LOL Dataset Results*: In Tables I and II, the results of the comparison of PRNet with the latest LLIE methods on the LOLv1 and LOLv2 datasets are summarized. These data were either directly cited from related articles or obtained by reproducing the results using publicly available codes on the same device. The best and second-best results are marked in bold and underlined, respectively. PSNR, SSIM, and LPIPS are used as the primary evaluation metrics. Higher PSNR and SSIM values and lower LPIPS values generally indicate superior similarity and restoration quality of the enhanced images.

Compared with other recent methods, PRNet achieved the best results in most cases and was close to the second-best in others. Notably, PRNet accomplished this with only 8.08G FLOPs of computation and 1.93M parameters. The latest method, LLFormer, which has $12\times$ the number of parameters as PRNet, only showed a slight advantage in the PSNR metric under the non-GT-Mean condition on the LOLv1 dataset. In contrast, PRNet consistently demonstrated superior performance across most evaluation metrics, significantly surpassing LLFormer.

Moreover, compared with the recent Fourier-based method FourLLIE, PRNet exhibited significant advantages in overall performance through its efficient utilization of frequency information. FourLLIE only showed a slight advantage in the PSNR metric under the non-GT-Mean condition on the LOLv2-Real dataset. In most cases, our method improves PSNR and SSIM indexes to varying degrees. For example, PSNR increased by 0.926 dB in the normal case of LOLv1 and by 0.823 dB in the GT-Mean case of LOLv1. It also increased by 0.511 dB in the normal case of LOLv2-Syn and by 0.536 dB in the GT-Mean case of LOLv2-Syn. Additionally, PRNet achieved a significant improvement in the LPIPS metric compared with recent methods such as SWANet and RetinexFormer, further highlighting its leadership in enhancing perceptual quality. SWANet performed better than PRNet on PSNR, probably because it employed more sophisticated strategies in the enhancement process of low-light regions, such as finer brightness adjustment and noise suppression techniques. The results

TABLE I

QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON LOLV1 AND LOLV2. HIGHER PSNR AND SSIM VALUES INDICATE BETTER RESULTS. BEST VALUES ARE IN BOLD; SECOND-BEST ARE UNDERLINED. THIS NOTATION IS CONSISTENT ACROSS ALL TABLES IN THIS ARTICLE

| Methods | Complexity | | LOLv1 | | | | LOLv2-Real | | | | LOLv2-Syn | | | |
|-------------------|------------|---------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|---------------|--------------|
| | Params/M | FLOPs/G | Normal | | GT-Mean | | Normal | | GT-Mean | | Normal | | GT-Mean | |
| | | | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| ZeroDCE [29] | 0.075 | 4.83 | 14.861 | 0.559 | 21.880 | 0.640 | 16.059 | 0.580 | 19.771 | 0.671 | 17.712 | 0.815 | 21.463 | 0.848 |
| DeepUPE [41] | 1.020 | 21.10 | 14.380 | 0.446 | — | — | 13.270 | 0.452 | — | — | 15.080 | 0.623 | — | — |
| RetinexNet [49] | 0.840 | 584.47 | 16.774 | 0.419 | 18.915 | 0.427 | 16.097 | 0.401 | 18.323 | 0.447 | 17.137 | 0.762 | 19.099 | 0.774 |
| 3DLUT [30] | 0.590 | 7.67 | 14.350 | 0.445 | 21.350 | 0.585 | 17.590 | 0.721 | 20.190 | 0.745 | 18.040 | 0.800 | 22.173 | 0.854 |
| Sparse [50] | 2.330 | 53.26 | 17.200 | 0.640 | — | — | 20.060 | 0.816 | — | — | 22.050 | 0.905 | — | — |
| KinD [31] | 8.020 | 34.99 | 17.650 | 0.775 | 20.860 | 0.802 | 14.740 | 0.641 | 17.544 | 0.669 | 13.290 | 0.578 | 16.259 | 0.591 |
| DRBN [32] | 5.270 | 48.61 | 16.290 | 0.617 | 19.550 | 0.746 | 20.130 | 0.830 | — | — | 23.220 | 0.927 | — | — |
| EnlightenGAN [42] | 8.640 | 61.01 | 17.480 | 0.651 | 20.003 | 0.691 | 18.230 | 0.617 | — | — | 16.570 | 0.734 | — | — |
| RUAS [33] | 0.003 | 0.83 | 16.405 | 0.500 | 18.654 | 0.518 | 15.326 | 0.488 | 19.061 | 0.510 | 13.765 | 0.638 | 16.584 | 0.719 |
| UFormer [34] | 5.290 | 12.00 | 19.270 | 0.796 | 23.829 | 0.812 | 18.619 | 0.769 | 24.213 | 0.806 | 20.104 | 0.879 | 25.735 | 0.917 |
| LLFlow [35] | 17.420 | 358.40 | 21.149 | 0.854 | 24.998 | 0.871 | 17.433 | 0.831 | 25.421 | 0.877 | 24.807 | 0.919 | 27.961 | 0.930 |
| Restormer [36] | 26.130 | 144.25 | 22.430 | 0.823 | <u>26.682</u> | 0.853 | 19.940 | 0.827 | 26.116 | 0.853 | 21.413 | 0.830 | 25.428 | 0.859 |
| SNR-Aware [43] | 4.010 | 26.35 | 23.129 | 0.834 | 26.497 | 0.841 | 20.671 | 0.833 | 27.103 | 0.869 | 24.582 | 0.917 | 27.618 | 0.927 |
| LEDNet [37] | 7.070 | 35.92 | 20.627 | 0.823 | 25.570 | 0.846 | 19.938 | 0.827 | <u>27.614</u> | 0.870 | 23.709 | 0.914 | 27.367 | 0.928 |
| LLFormer [38] | 24.550 | 22.52 | 23.529 | 0.815 | 26.097 | 0.827 | 21.300 | 0.801 | 27.019 | 0.824 | 24.038 | 0.909 | 28.006 | 0.927 |
| PairLIE [39] | 0.330 | 20.81 | 19.510 | 0.736 | 23.526 | 0.755 | 19.885 | 0.778 | 24.025 | 0.803 | 21.351 | 0.857 | 25.326 | 0.873 |
| FourLLIE [48] | 0.120 | 2.57 | 22.526 | 0.831 | 26.037 | 0.849 | 22.127 | <u>0.843</u> | 27.591 | 0.871 | 24.763 | 0.924 | <u>28.121</u> | 0.926 |
| MBPNet [62] | 14.191 | 16.73 | 22.917 | 0.829 | 25.946 | 0.843 | 21.037 | 0.832 | 27.231 | 0.873 | 24.319 | 0.907 | 27.831 | 0.921 |
| Ours | 1.930 | 8.08 | <u>23.452</u> | <u>0.837</u> | 26.860 | <u>0.856</u> | <u>21.314</u> | 0.848 | 27.661 | 0.883 | 25.274 | 0.931 | 28.657 | 0.943 |

TABLE II

QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON THE LOLV1 AND LOLV2 DATASETS. LOWER LPIPS VALUES INDICATE BETTER RESULTS. SMALLER FLOPS VALUES REPRESENT LOWER COMPUTATIONAL COMPLEXITY

| Methods | LOLv1 | LOLv2-Real | LOLv2-Syn | FLOPs/G |
|--------------------|--------------|--------------|--------------|--------------|
| RetinexNet [49] | 0.474 | 0.543 | 0.255 | 587.47 |
| KinD [31] | 0.207 | 0.375 | 0.435 | 34.99 |
| EnlightenGAN [42] | 0.322 | 0.309 | 0.220 | 61.01 |
| LLFlow [35] | 0.117 | 0.137 | 0.067 | 358.40 |
| MIRNet [59] | 0.131 | 0.144 | 0.057 | 785.00 |
| LEDNet [37] | 0.118 | <u>0.121</u> | 0.061 | 35.92 |
| LLFormer [38] | 0.175 | 0.211 | 0.066 | 22.52 |
| SNR-Aware [43] | 0.152 | 0.157 | <u>0.055</u> | 26.35 |
| RetinexFormer [60] | 0.131 | 0.171 | <u>0.059</u> | <u>15.85</u> |
| MBPNet [62] | 0.119 | 0.132 | 0.063 | 16.73 |
| SWANet [19] | <u>0.107</u> | 0.124 | 0.056 | 18.28 |
| Ours | 0.098 | 0.119 | 0.053 | 8.08 |

shown in Fig. 6 present a comparison of the visual results between PRNet and other methods. For instance, RetinexNet often produced overexposed or underexposed regions and failed to suppress noise effectively. The enhancement results of SNR-Aware were often blurry, while LLFormer resulted in noticeable color distortions. In contrast, PRNet significantly enhanced the visibility of low-contrast and low-light regions, effectively removing artifacts and noise, maintaining more realistic colors, and accurately reconstructing the original scene. This leads to remarkable improvements in perceptual quality and visual effects.

In Fig. 7, several quantitative metrics, namely, NMSE, MAE, and UIQI, were further compared. Among them, lower NMSE and MAE values are better, while UIQI values closer to one are preferable. The results indicate that PRNet achieved significantly lower NMSE and MAE values than other methods, demonstrating smaller deviations between the enhanced and normal-light images. Additionally, PRNet achieved a UIQI value closer to one, indicating superior performance

in terms of brightness, contrast, and structural consistency. These results collectively validate the significant advantages of PRNet in LLIE.

2) *Unpaired Dataset Results:* The models trained on the LOLv1 or LOLv2-Syn datasets were directly applied to the real-world unpaired datasets DICM, LIME, MEF, NPE, and VV, and their performance was evaluated using the NIQE and BRIS metrics. Notably, lower NIQE and BRIS values generally indicate higher image quality. In the comparison with the latest methods, the best and second-best results are highlighted in bold and underlined, respectively.

Compared to FourLLIE, our method achieved lower NIQE scores across multiple datasets, indicating improved perceptual quality in the enhanced images. Specifically, on the MEF dataset, PRNet achieved an NIQE reduction of 0.94 relative to FourLLIE, while on the LIME dataset, the reduction was 0.56. Likewise, for the NPE and VV datasets, PRNet achieved reductions of 0.17 and 0.07, respectively, further confirming its effectiveness in preserving natural image structures in low-light conditions. FourLLIE showed a marginal advantage only on the DICM dataset.

For BRIS, PRNet consistently outperformed FourLLIE across all five datasets. Notably, it achieved a significant BRIS reduction of 8.12 on the MEF dataset, underscoring its effectiveness in enhancing contrast and reducing artifacts.

In Table III, except for the NIQE metric on the DICM dataset and the BRIS metric on the VV dataset, our method consistently achieved the best results across nearly all metrics, demonstrating its excellent generalization capability. Furthermore, as illustrated in Fig. 8, our model effectively enhanced the visibility of low-light regions while preventing overexposure, resulting in higher quality image enhancement.

3) *SICE and Sony-Total-Dark Dataset Results:* To evaluate our model's performance on large-scale datasets, experiments were conducted on two highly challenging datasets: SICE

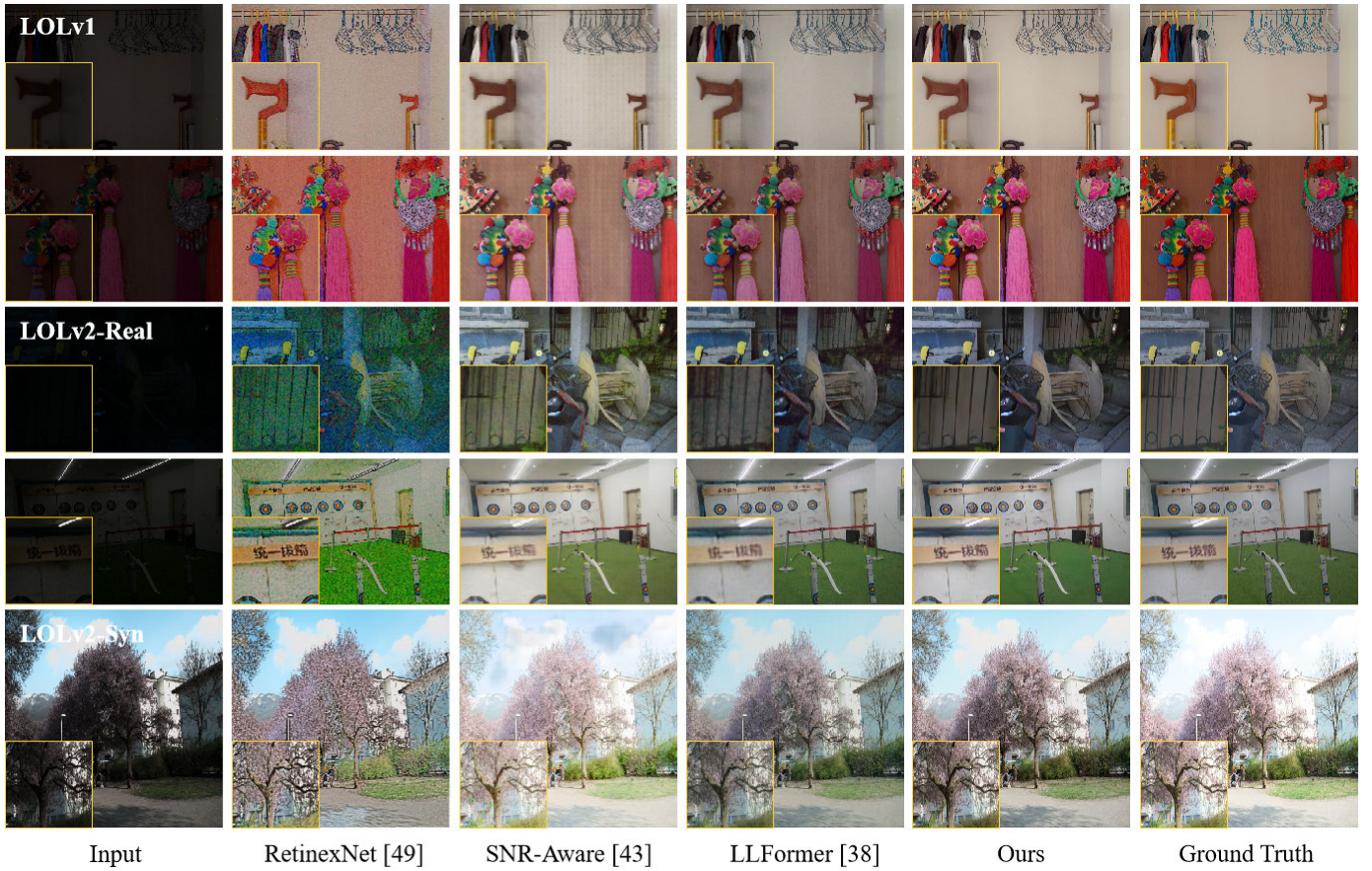


Fig. 6. Visual comparison of enhancement results on the LOLv1 and LOLv2 datasets using different methods. RetinexNet typically produces overexposed or underexposed areas and is not effective at suppressing noise, which results in artifacts that are inconsistent with the ground-truth image. The enhancement results of SNR-Aware may appear blurry. LLFormer produces significant color distortion in the enhanced effect.

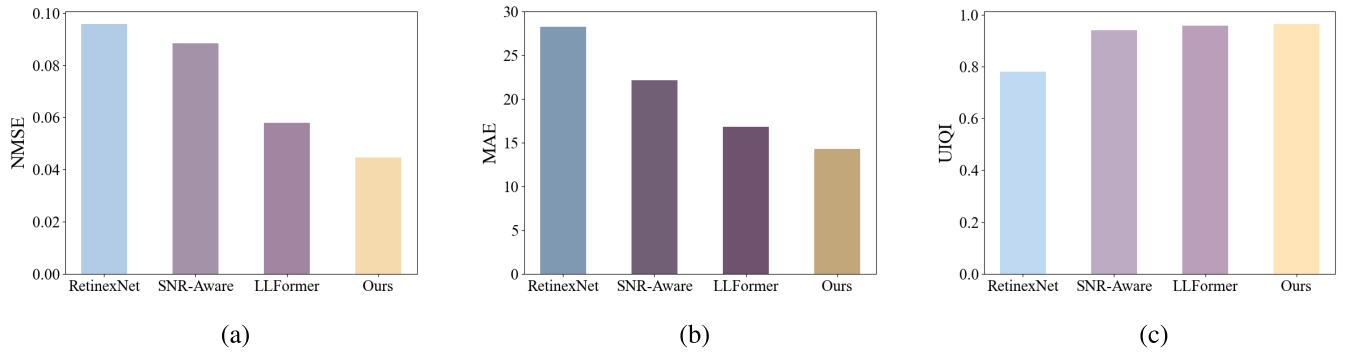


Fig. 7. Quantitative evaluation of different methods using NMSE, MAE, and UIQI metrics. (a) PRNet has the lowest NMSE, which indicates that it has higher pixel-level accuracy. (b) PRNet shows the lowest MAE, which reflects the smallest absolute error between the enhanced image and the reference image. (c) PRNet achieves the highest UIQI near one, which demonstrates its superior performance in maintaining brightness, contrast, and structural consistency.

(including the Mix and Grad test sets) and Sony-Total-Dark. Table IV shows that PRNet outperformed other methods across three key metrics: PSNR, SSIM, and LPIPS, demonstrating superior generalization capabilities. On the SICE-Mix test set, PRNet achieved a PSNR of 12.985. On the SICE-Grad test set, the PSNR reached 13.171, which indicates a high perceived quality. From a visual perspective, as shown in Fig. 9, PRNet significantly enhanced low-light regions while effectively suppressing noise and artifacts, restoring image

details and structural consistency. In contrast, RetinexNet introduced some noise and showed limited performance in color recovery, and LLFlow struggled to recover dark details. Although LEDNet showed advantages in brightness enhancement, it often exhibited significant color distortion. PRNet, by contrast, achieved an optimal balance between brightness enhancement and detailed preservation. Its images were more closely aligned with the actual scene and presented more natural color reproduction.

TABLE III
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS ON DICM, LIME, MEF, NPE, AND VV DATASETS.
LOWER NIQE AND BRIS VALUES INDICATE BETTER PERFORMANCE

| Methods | DICM | | MEF | | LIME | | NPE | | VV | | AVG | |
|----------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|
| | NIQE | BRIS | NIQE | BRIS | NIQE | BRIS | NIQE | BRIS | NIQE | BRIS | NIQE | BRIS |
| LIME [23] | 4.48 | 27.38 | 4.74 | 39.10 | 5.05 | 32.84 | 4.17 | 28.94 | 3.71 | 18.93 | 4.43 | 29.44 |
| KinD [31] | 5.15 | 48.72 | 5.47 | 49.94 | 5.03 | 39.91 | 4.98 | 36.85 | 4.30 | 50.56 | 4.99 | 45.20 |
| DRBN [32] | 4.37 | 30.71 | 4.86 | 44.67 | 4.56 | 34.56 | 3.92 | 25.34 | 3.67 | 24.95 | 4.42 | 32.05 |
| RUAS [33] | 7.31 | 46.88 | 5.44 | 42.12 | 5.32 | 34.88 | 7.20 | 48.98 | 4.99 | 35.88 | 6.05 | 41.75 |
| SCI [58] | 4.52 | 27.92 | <u>3.61</u> | 26.72 | 4.46 | 25.17 | 4.12 | 28.89 | 5.31 | <u>22.80</u> | 4.40 | 26.30 |
| LLFlow [35] | 4.06 | 26.36 | 4.70 | 30.27 | 4.59 | 27.06 | 4.67 | 28.86 | 4.04 | 31.67 | 4.41 | 28.84 |
| SNR-Aware [43] | 4.71 | 37.35 | 4.18 | 31.28 | 5.74 | 39.22 | 4.32 | 26.65 | 9.87 | 78.72 | 5.76 | 42.64 |
| PairLIE [39] | <u>4.03</u> | 33.31 | 4.06 | 27.53 | 4.58 | 25.23 | 4.18 | 28.27 | 3.57 | 39.13 | 4.08 | 30.69 |
| FourLLIE [48] | 3.71 | 27.41 | 4.41 | 20.37 | 4.27 | 24.38 | <u>3.87</u> | 22.37 | <u>3.21</u> | 30.14 | 3.89 | 24.93 |
| MBPNet [62] | 4.09 | 28.37 | 4.27 | 21.15 | 3.67 | <u>19.73</u> | 4.15 | 22.67 | 3.88 | 31.15 | 4.01 | 24.61 |
| SWANet [19] | 3.97 | 27.39 | 3.74 | <u>18.79</u> | 3.98 | 21.38 | 3.97 | <u>21.56</u> | 3.29 | 29.87 | <u>3.79</u> | <u>23.80</u> |
| Ours | 4.04 | <u>27.30</u> | 3.47 | 12.25 | <u>3.71</u> | 18.64 | 3.70 | 16.74 | 3.14 | 29.84 | 3.61 | 20.95 |

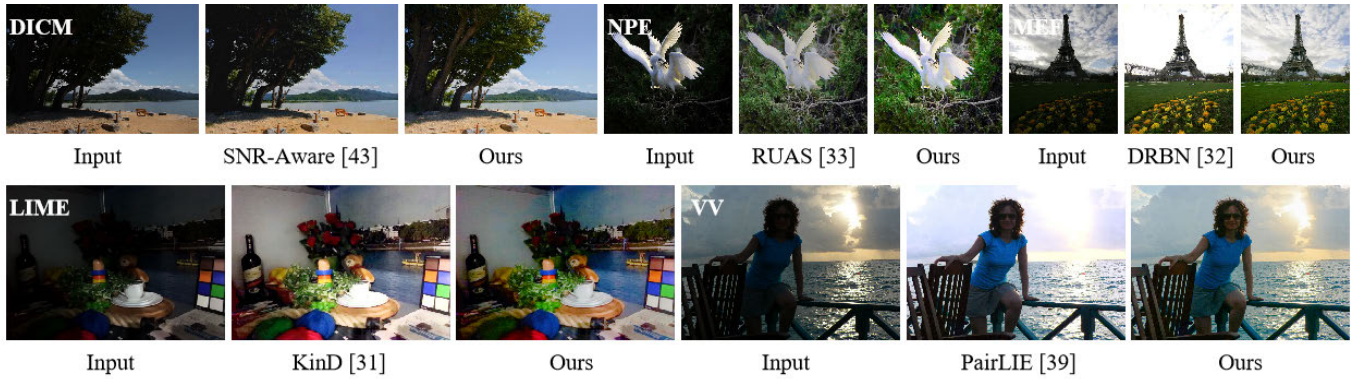


Fig. 8. Visual comparison was conducted on the DICM, LIME, MEF, NPE, and VV datasets, where one image was randomly selected from each dataset and compared with results from other methods. PRNet delivers more realistic visual results.

TABLE IV
QUANTITATIVE COMPARISON ON THREE HARD DATASETS, THE HIGHER THE PSNR AND SSIM VALUES, THE BETTER THE RESULTS. LOWER LPIPS VALUES INDICATE BETTER RESULTS

| Methods | SICE-Mix | | | SICE-Grad | | | Sony-Total-Dark | | |
|------------------|---------------|--------------|--------------|---------------|--------------|--------------|-----------------|--------------|--------------|
| | PSNR | SSIM | LPIPS | PSNR | SSIM | LPIPS | PSNR | SSIM | LPIPS |
| RetinexNet [49] | 12.397 | 0.606 | 0.407 | 12.450 | 0.619 | 0.364 | 13.241 | 0.367 | 0.769 |
| ZeroDCE [29] | 12.428 | <u>0.633</u> | <u>0.382</u> | 12.475 | <u>0.644</u> | <u>0.334</u> | 14.087 | 0.090 | 0.813 |
| URetinexNet [61] | 10.903 | 0.600 | 0.402 | 10.894 | 0.610 | 0.356 | 15.519 | 0.323 | 0.599 |
| RUAS [33] | 8.684 | 0.493 | 0.525 | 8.628 | 0.494 | 0.499 | 12.622 | 0.081 | 0.920 |
| LLFlow [35] | <u>12.737</u> | 0.617 | 0.388 | <u>12.819</u> | 0.629 | 0.361 | 18.317 | 0.571 | 0.499 |
| LEDNet [37] | 12.668 | 0.579 | 0.412 | 12.551 | 0.576 | 0.383 | <u>20.830</u> | <u>0.648</u> | <u>0.471</u> |
| Ours | 12.985 | 0.635 | 0.371 | 13.171 | 0.647 | 0.329 | 22.954 | 0.685 | 0.427 |

C. Computational Efficiency Analysis

To evaluate the computational efficiency of PRNet, experiments were conducted on the LOL dataset with an image size of 600×400 pixels. The runtime and parameter comparisons between PRNet and other state-of-the-art methods are shown in Table V. The results highlight PRNet's ability to achieve an optimal balance between processing speed and parameter efficiency. LLFormer, while slightly superior in PSNR,

requires substantially more parameters and longer processing times. On the other hand, Zero-DCE demonstrated faster execution and a more compact architecture but at the cost of significantly inferior PSNR performance. PRNet achieved a notable advantage by offering competitive PSNR values while maintaining a lightweight architecture and efficient runtime, underscoring its practicality for resource-constrained environments.

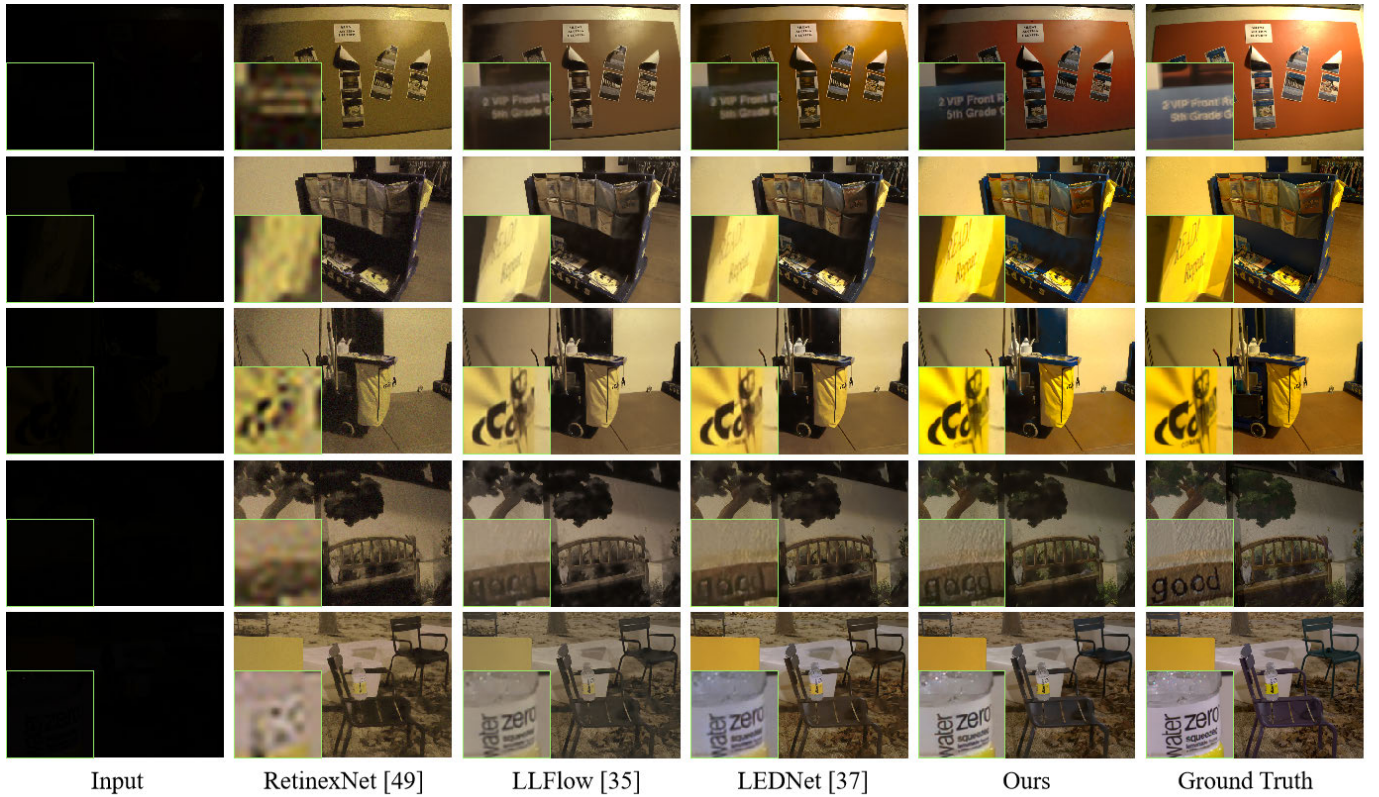


Fig. 9. Enhancement results of PRNet on the Sony-Total-Dark dataset. Our method produces enhanced images with color similarity closer to the ground truth, while other methods suffer from color distortion, varying degrees of blurriness, noise, or artifacts. Compared to other approaches, our method more accurately restores the ground truth.

TABLE V

COMPARISON OF RUNNING TIME AND MODEL PARAMETERS BETWEEN PRNET AND THE OTHER MODELS ON LOL DATASET

| Methods | Running time/s | Parameter/M | PSNR/dB |
|-------------------|----------------|-------------|---------|
| LIME [23] | 0.0783 | — | 16.76 |
| ZeroDCE [29] | 0.0047 | 0.075 | 14.86 |
| RetinexNet [49] | 0.5441 | 0.840 | 16.77 |
| KinD [31] | 0.7255 | 8.020 | 17.65 |
| EnlightenGAN [42] | 0.2278 | 8.640 | 17.48 |
| URetinexNet [61] | 0.2157 | 0.400 | 21.33 |
| LLFormer [38] | 2.0680 | 24.550 | 23.65 |
| Ours | 0.1306 | 1.930 | 23.45 |

TABLE VI

QUANTITATIVE RESULTS OF THE ABLATION STUDIES ON LOLV1. HIGHER PSNR AND SSIM VALUES AND LOWER LPIPS VALUES INDICATE BETTER RESULTS

| | (a) | (b) | (c) | (d) | PSNR | SSIM | LPIPS |
|-----|-----|-----|-----|-----|---------------|--------------|--------------|
| (1) | ✓ | | | | 21.867 | 0.805 | 0.131 |
| (2) | ✓ | ✓ | | | 22.582 | 0.821 | 0.119 |
| (3) | ✓ | ✓ | ✓ | | 22.892 | 0.829 | 0.106 |
| (4) | ✓ | ✓ | ✓ | ✓ | 23.452 | 0.837 | 0.098 |

D. Ablation Study

A more in-depth analysis of PRNet was conducted to demonstrate the effectiveness of its key components. For the

ablation study presented in this section, the model was trained on the LOLv1 dataset with a batch size of 10 for 1000 epochs, and a patch size of 256 to illustrate the impact of each component in our approach.

As illustrated in Table VI, for PENet, “✓” in (c) denotes the use of a dense skip-connection structure; otherwise, a standard cascade configuration was employed for the experiments. Similarly, “✓” in (d) indicates the application of FBEB, while its absence implies the use of a basic brightness enhancement module based solely on a 1×1 convolutional layer. For RRNet, “✓” in (a) signifies the implementation of a dual-branch structure, whereas a “✓” in (b) reflects the incorporation of the BFDCS; in its absence, a standard cross-attention solver is utilized for the experiments.

As demonstrated in the ablation study in Table VI, for PENet, both the separate and combined utilization of the dense skip-connection structure and the FBEB yield notable improvements in PSNR, SSIM, and LPIPS. As depicted in Fig. 10, on the LOLv1 dataset, the integration of the proposed dense skip-connection structure and FBEB enhances the brightness of LLIs markedly, facilitating the initial phase of image restoration.

For RRNet, the proposed BFDCS leverages the properties of the FFT, helping reduce computational costs. It achieves spatial and temporal complexities of $O(N)$ and $O(NC \log N)$, respectively, which are significantly lower than the $O(N^2)$ and $O(N^2C)$ complexities associated with the original scaled dot-product attention calculation, where N



Fig. 10. Visual results of the four settings in the ablation study in Table VI. (a) Dual-branch structure only. (b) Dual-branch structure + BFDCS module. (c) Dual-branch structure + BFDCS + dense skip-connection structure. (d) Dual-branch structure + BFDCS + dense skip-connection structure + FBEB (full PRNet).

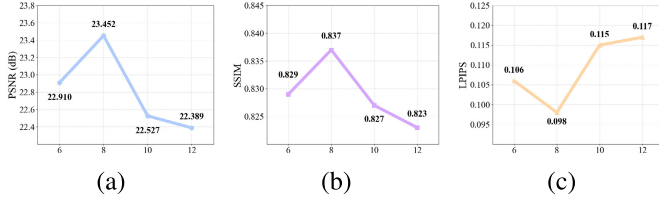


Fig. 11. Quantitative analysis of the impact of the number of FBEB modules on (a) PSNR, (b) SSIM, and (c) LPIPS metrics. The results show that the performance of FBEB increases with the number of modules and reaches a peak value at eight modules. Additional modules introduce redundancy, which degrades performance across all metrics. As a result, the eight FBEB block modules can effectively balance brightness, contrast, and structural consistency.

denotes the number of pixels and C represents the number of features. Concurrently, the BFDCS improves the LLIE performance substantially. BFDCS was compared against a baseline method operating within the spatial domain. Replacing the conventional cross-attention solver with the BFDCS results in notable gains in PSNR, SSIM, and LPIPS, as evidenced by the comparison between configurations (1) and (2) in Table VI. As depicted in Fig. 10, the proposed BFDCS generates significantly clearer visual outcomes on the LOLv1 dataset.

The impact of the number of FBEB modules on performance was analyzed for module counts of 6, 8, 10, and 12. The quality of the generated images was evaluated using PSNR, SSIM, and LPIPS metrics. As shown in Fig. 11, eight FBEB blocks obtained the best PSNR, SSIM, and LPIPS values. This shows that the right number of FBEB modules can effectively balance brightness, contrast, and structural consistency. However, too many modules can lead to redundancy, resulting in performance degradation.

V. CONCLUSION

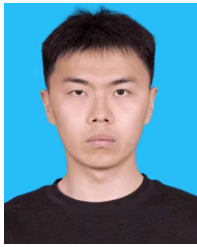
In this article, PRNet is presented, which is an efficient network for LLIE that leverages frequency-domain information in a dual-stage framework. In the first stage, a dense skip-connection structure, based on the FBEB, performs pixel-level brightness enhancement, significantly increasing the luminance of LLIs. In the second stage, a dual-branch UNet architecture utilizing the BFDCS further enhances brightness and detail restoration. Extensive experiments on nine public datasets demonstrate PRNet's superior performance with low computational complexity, consistently surpassing existing methods across multiple metrics. However, the proposed method still has some drawbacks. Sometimes, there will be a problem of local overexposure in regions with only slightly low light levels. Future work will focus on optimizing the inference speed of PRNet so that it can be deployed in

resource-constrained environments, such as mobile devices. In addition, we will further explore its potential in fog removal and super-resolution missions.

REFERENCES

- [1] S. P. Mallick et al., "Real-time, automated, multiobjective, cloud computing whole slide imaging device," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–9, 2023.
- [2] Y. Hui, J. Wang, and B. Li, "WSA-YOLO: Weak-supervised and adaptive object detection in the low-light environment for YOLOV7," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–12, 2024.
- [3] X. Li, S. Wang, P. Chen, and L. Wang, "3-D inspection method for industrial product assembly based on single X-ray projections," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.
- [4] Z. Liu, B. Guan, Y. Shang, Q. Yu, and L. Kneip, "Line-based 6-DoF object pose estimation and tracking with an event camera," *IEEE Trans. Image Process.*, vol. 33, pp. 4765–4780, 2024.
- [5] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human action recognition from various data modalities: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3200–3225, Mar. 2023.
- [6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.
- [7] W. Shen, A. Ma, J. Wang, Z. Zheng, and Y. Zhong, "Adaptive self-supporting prototype learning for remote sensing few-shot semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5634116.
- [8] L. Zhang, W. Zhou, H. Fan, T. Luo, and H. Ling, "Robust domain adaptive object detection with unified multi-granularity alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9161–9178, Dec. 2024.
- [9] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
- [10] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
- [11] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2828–2841, Jun. 2018.
- [12] K. Xu et al., "HFMNet: Hierarchical feature mining network for low-light image enhancement," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.
- [13] W. Wu, W. Wang, K. Jiang, X. Xu, and R. Hu, "Self-supervised learning on a lightweight low-light image enhancement model with curve refinement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 1890–1894.
- [14] S. Yang, D. Zhou, J. Cao, and Y. Guo, "LightingNet: An integrated learning method for low-light image enhancement," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 29–42, 2023.
- [15] S. Zhang, N. Meng, and E. Y. Lam, "LRT: An efficient low-light restoration transformer for dark light field images," *IEEE Trans. Image Process.*, vol. 32, pp. 4314–4326, 2023.
- [16] D. Ye, Z. Ni, W. Yang, H. Wang, S. Wang, and S. Kwong, "Glow in the dark: Low-light image enhancement with external memory," *IEEE Trans. Multimedia*, vol. 26, pp. 2148–2163, 2023.
- [17] J. Huang et al., "Deep Fourier-based exposure correction network with spatial-frequency interaction," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer*, 2022, pp. 163–180.

- [18] Z. Yao, G. Fan, J. Fan, M. Gan, and C. L. Philip Chen, "Spatial-Frequency dual-domain feature fusion network for low-light remote sensing image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4706516.
- [19] Z. He et al., "Low-light image enhancement with multi-scale attention and frequency-domain optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2861–2875, Apr. 2024.
- [20] H. Yu, N. Zheng, M. Zhou, J. Huang, Z. Xiao, and F. Zhao, "Frequency and spatial dual guidance for image dehazing," in *Proc. Eur. Conf. Comput. Vision. (ECCV)*, Jan. 2022, pp. 181–198.
- [21] L. Kong, J. Dong, J. Ge, M. Li, and J. Pan, "Efficient frequency domain-based transformers for high-quality image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 5886–5895.
- [22] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 889–896, May 2000.
- [23] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
- [24] K. G. Lore, A. Akinlayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.
- [25] Y. Zhang, X. Di, B. Zhang, R. Ji, and C. Wang, "Better than reference in low-light image enhancement: Conditional re-enhancement network," *IEEE Trans. Image Process.*, vol. 31, pp. 759–772, 2022.
- [26] K. Lu and L. Zhang, "TBEFN: A two-branch exposure-fusion network for low-light image enhancement," *IEEE Trans. Multimedia*, vol. 23, no. 11, pp. 4093–4105, Nov. 2020.
- [27] J. Li, J. Li, F. Fang, F. Li, and G. Zhang, "Luminance-aware pyramid network for low-light image enhancement," *IEEE Trans. Multimedia*, vol. 23, pp. 3153–3165, 2021.
- [28] H. Cui, J. Li, Z. Hua, and L. Fan, "Progressive dual-branch network for low-light image enhancement," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–18, 2022.
- [29] C. Guo et al., "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1780–1789.
- [30] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, "Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2058–2073, Apr. 2022.
- [31] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia (ACM MM)*, Oct. 2019, pp. 1632–1640.
- [32] W. Yang, S. Wang, Y. Fang, Y. Wang, and J. Liu, "From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3060–3069.
- [33] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10556–10565.
- [34] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general U-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17662–17672.
- [35] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. C. Kot, "Low-light image enhancement with normalizing flow," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 2604–2612.
- [36] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5718–5729.
- [37] S. Zhou, C. Li, and C. C. Loy, "LEDNet: Joint low-light enhancement and deblurring in the dark," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 573–589.
- [38] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 3, 2023, pp. 2654–2662.
- [39] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 22252–22261.
- [40] J. Tao, Y. Li, B. Guan, Y. Shang, and Q. Yu, "Simultaneous enhancement and noise suppression under complex illumination conditions," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.
- [41] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6842–6850.
- [42] Y. Jiang et al., "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.
- [43] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "SNR-aware low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17693–17703.
- [44] L. Liu, X. Yang, N. Wang, and X. Gao, "Frequency information disentanglement network for video-based person re-identification," *IEEE Trans. Image Process.*, vol. 32, pp. 4287–4298, 2023.
- [45] F. Zhang, A. Panahi, and G. Gao, "FsaNet: Frequency self-attention for semantic segmentation," *IEEE Trans. Image Process.*, vol. 32, pp. 4757–4772, 2023.
- [46] X. Jiang et al., "FABNet: Frequency-aware binarized network for single image super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 6234–6247, 2023.
- [47] C. Li et al., "Embedding Fourier for ultra-high-definition low-light image enhancement," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Jan. 2023, pp. 1–11.
- [48] C. Wang, H. Wu, and Z. Jin, "FourLLIE: Boosting low-light image enhancement by Fourier frequency information," in *Proc. 31st ACM Int. Conf. Multimedia*, Oct. 2023, pp. 7459–7469.
- [49] W. Chen, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Jan. 2018, pp. 1–21.
- [50] W. Yang, W. Wang, H. Huang, S. Wang, and J. Liu, "Sparse gradient regularized deep retinex network for robust low-light image enhancement," *IEEE Trans. Image Process.*, vol. 30, pp. 2072–2086, 2021.
- [51] C. Lee, C. Lee, and C. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.
- [52] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [53] S. Wang, J. Zheng, H. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
- [54] V. Vonikakis, R. Kouskouridas, and A. Gasteratos, "On the evaluation of illumination compensation algorithms," *Multimedia Tools Appl.*, vol. 77, no. 8, pp. 9211–9231, May 2017.
- [55] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 2049–2062, Apr. 2018.
- [56] S. Zheng, Y. Ma, J. Pan, C. Lu, and G. Gupta, "Low-light image and video enhancement: A comprehensive survey and beyond," 2022, *arXiv:2212.1077*.
- [57] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.
- [58] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 5627–5636.
- [59] S. W. Zamir et al., "Learning enriched features for real image restoration and enhancement," in *Proc. Eur. Conf. Comput. Vis.*, vol. 12370, 2020, pp. 492–511.
- [60] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12470–12479.
- [61] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5891–5900.
- [62] K. Zhang, C. Yuan, J. Li, X. Gao, and M. Li, "Multi-branch and progressive network for low-light image enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 2295–2308, 2023.



Jiayu Zhang was born in Shaanxi, China, in 2001. He received the B.S. degree from Xi'an Polytechnic University, Xi'an, China, in 2023, where he is currently pursuing the M.E. degree with the School of Electronic Information.

His current research interests include artificial intelligence, deep learning, computer vision, mobile robotics, and low-light image enhancement.



Yingjian Li was born in Zhangjiakou, China, in 1999. He received the B.S. degree from Xi'an Polytechnic University, Xi'an, China, in 2023, where he is currently pursuing the M.E. degree with the School of Computer Science.

His current research interests include artificial intelligence, deep learning, computer vision, and low-light image enhancement.



Xiaohua Wang received the Ph.D. degree in control science and engineering from Xi'an University of Technology, Xi'an, China, in 2012.

She is currently the Leader of robot engineering with the School of Electronic Information, Xi'an Polytechnic University, Xi'an, where she is also a Member of the Ninth Academic Committee, and a Master's Tutor and a Visiting Scholar with the University of Göttingen, Göttingen, Germany. Her research interests include mobile robot autonomous navigation and industrial robot autonomous operation.

In these fields, she has published more than 100 academic articles, of which more than 70 have been included in the three major searches, and trained more than 30 graduate students.



Wenjie Wang (Graduate Student Member, IEEE) received the B.S. and Ph.D. degrees in mechanical engineering from the College of Mechanical and Electrical Engineering, Harbin Engineering University, Harbin, China, in 2012 and 2018, respectively.

He is currently an Associate Professor of control science and engineering with Xi'an Polytechnic University, Xi'an, China. He has led and participated in over ten projects. He has authored more than 20 articles and more than ten inventions. His current research interests include artificial intelligence,

mechatronics, surgical robots, and intelligent equipment.