Low-Light Image Enhancement using Retinex-based Network with Attention Mechanism

Shaojin Ma[1], Weiguo Pan[2]*, Nuoya Li[3], Songjie Du[4], Hongzhe Liu[5], Bingxin Xu[6], Cheng Xu[7], Xuewei Li[8]*
Beijing Key Laboratory of Information Service Engineering, Beijing Union University, China[1, 2, 3, 4, 5, 6, 7, 8]
College of Robotics, Beijing Union University, Beijing Union University, China[1, 2, 3, 4, 5, 6, 7, 8]

***Abstract*—Images in low-light conditions typically exhibit significant degradation such as low contrast, color shift, noise and artifacts, which diminish the accuracy of the recognition task in computer vision. To address these challenges, this paper proposes a low-light image enhancement method based on Retinex. Specifically, a decomposition network is designed to acquire high-quality light illumination and reflection maps, complemented by the incorporation of a comprehensive loss function. A denoising network was proposed to mitigate the noise in low-light images with the assistance of images' spatial information. Notably, the extended convolution layer has been employed to replace the maximum pooling layer and the Basic-Residual-Modules (BRM) module from the decomposition network has integrates into the denoising network. To address challenges related to shadow blocks and halo artifacts, an enhancement module was proposed to be integration into the jump connections of U-Net. This enhancement module leverages the Feature-Extraction- Module (FEM) attention module, a sophisticated mechanism that improves the network's capacity to learn meaningful features by integrating the image features in both channel dimensions and spatial attention mechanism to receive more detailed illumination information about the object and suppress other useless information. Based on the experiments conducted on public datasets LOL-V1 and LOL-V2, our method demonstrates noteworthy performance improvements. The enhanced results by our method achieve an average of 23.15, 0.88, 0.419 and 0.0040 on four evaluation metrics - PSNR, SSIM, NIQE and GMSD. Those results superior to the mainstream methods.**

Keywords*—*Low-light image enhancement; decomposition network; FEM attention mechanism; denoising network; detail enhancement

I. Introduction

In the field of computer vision, low-light image enhancement has perennially been a focal point of research. Images acquired under low light conditions are often affected by problems like light weakening, increased noise and loss of detail, resulting in degraded image quality and blurred image content. The identified limitations exert a detrimental impact on the efficacy of computer vision applications, presenting challenges across various scenarios, including object detection [1], driverless driving [2], medical imaging. Moreover, these deficiencies introduce inconvenience in routine image capture and sharing within everyday life.

Traditional image enhancement methods often rely on manually adjusting parameters such as brightness and contrast [3], which may not adapt well to changes in different scenes. However, due to the inability to effectively and accurately capture the features of the image, as well as its complex textures, these methods can lead to over-enhancement or the presence of shadow blocks and halo artifacts in the enhanced image. In contrast, methods based on deep learning improve adaptability and generalization by automatically learning image features, making them particularly suitable for complex environments. Specifically, deep learning methods based on Retinex theory, which separate an image's illuminance and reflectance through a decomposition network, allow for detailed adjustments through reflectance recovery and illuminance adjustment networks. These methods then merge the enhanced reflectance and illuminance images to improve brightness, contrast, and maintain natural colors, making them especially suitable for image enhancement in low-light environments. However, the decomposition network in Retinex can be affected by uneven lighting, potentially leading to loss of image detail, especially in dark and highlight areas of the image, thereby affecting the naturalness and realism of the final enhancement effect.

Although there are currently various enhancement methods for low-light images, mainly focusing on improving image contrast, it is important to note that low-light images often contain a significant amount of noise, which can greatly affect the quality and clarity of the image. Many of the current denoising techniques are applied in the pre-processing and post-processing stages of the image. Denoising in the pre-processing stage can cause the image to become blurred, while applying denoising in the post-processing stage can lead to the amplification of noise. Therefore, in the process of enhancing low-light images, how to appropriately balance the suppression of noise with the preservation of image details becomes a key challenge.

To address the aforementioned issues, this paper presents three main contributions, as follows:

1) In this paper, a decomposition network is proposed to obtain illumination and reflection maps through the decomposition of the RM and IM modules, as well as a comprehensive loss function is advanced to maintain the overall structure and consistency of the decomposed images.

2) A denoising network was proposed to remove noise in low-light images, with the assistance of images' spatial information, noise can be efficiently diminished, preserving map details, and consequently elevating the overall quality of enhanced images.

3) To effectively mitigate shadow blocks and halo artifacts in low-light images, this paper introduces an enhancement network featuring the FEM attention mechanism, which can significantly improve the restoration of image details and textures, yielding clear and natural image results.

The rest of this article is organized as follows: Section II discusses the related work. The proposed approach is detailed in Section III. Section IV provides quantitative and qualitative evaluation the method's performance. In Section V, ablation experiments were carried out on the FEM module and the de-noising module, respectively.

II. Related Work

Over the past few decades, various conventional methods have been proposed to address the challenges of low-light image enhancement. Noteworthy among these are traditional image enhancement methods, specifically histogram equalization [4] and methods based on Retinex theory [5]. Among these methods, histogram equalization method is one of the earliest and most extensively utilized methods. It aims to enhance image contrast and brightness by redistributing the gray level of image pixel values. This approach exists the disadvantages such as the limitations of global processing and the sensitivity to noise, which can lead to unnatural effects and information loss.

To enhance the visual alignment of images with human perception, Land et al. [6] proposed Retinex theory. At the core of the theory lies the concept that an object's color is determined not only by the intensity of the reflected light but also by its ability to reflect light waves. However, this method exhibits limitations when applied to images with complex lighting conditions and strong contrasts. To address these issues, researchers introduced the Multiscale Retinex (MSR) algorithm [7], incorporating multi-scale Gaussian filters to more accurately estimate the illumination components at various scales, and suppressed the halo effect through weighted summation. Furthermore, in pursuit of preserving the natural and authentic visual characteristics of images, Gao et al. [8] proposed an improved Retinex algorithm based on the traditional Retinex algorithm, which introduced color correction and multiscale processing technology to improve image details at different scales more precisely, thereby improving the visual effect and quality of images. However, most RetineX-based methods can cause severe color distortion and struggle to effectively enhance images with relatively high dynamic range.

In recent years, the remarkable adaptive capabilities of deep learning in low-light image enhancement have established it as an effective method, contributing to the improvement of image quality and finding widespread application in various computer vision tasks. Numerous scholars have extended their efforts to constructing learning-based models based on Retinex theory. For instance, RetinexNet [9] integrates Retinex theory with deep convolutional neural networks, enhancing image contrast through brightness maps estimation and adjustment, with subsequent post-processing using Block-Matching and 3D Filtering (BM3D) for denoising. Zhang et al. [10] designed an efficient network based on Retinex theory to enhance low-light images. Lim et al. [11] introduced the Deep-Stacked Laplacian Restorer (DSLR), capable of recovering global brightness and local detail from the original input, achieving notable success in contrast improvement and noise reduction. Moreover, several non-Retinex-based methods have been proposed. Li et al. [12] developed LightenNet, a convolutional neural network employing a stacked sparse denoising autoencoder structure to learn the nonlinear transformation function for adaptive brightness and contrast enhancement in low light images. However, this method faces challenges in effectively addressing noise in low-light images conditions. Ma et al. [13] proposed a low-light image enhancement method based on a fast, flexible and robust strategy. The method combines adaptive enhancement and parameter adjustment to efficiently enhance images while preserving detail and quality. EnlightenGAN [14] employs an unsupervised deep learning approach within a Generative Adversarial Network (GAN) framework to address low-light image enhancement challenges. Depth-Aware Decomposition and Restoration Network (DA-DRN) [15] introduces a self-sensing depth Retinex network, directly restores degraded reflectance and preserves the detail information in the decomposition stage by using the dependence between reflectance and illumination pattern. Although these methods can significantly improve the brightness and contrast of images, challenges persist in noise removal and image detail recovery. Some methods may result in overly enhanced image, leading to potential distortions.

The essence of the Retinex method lies in the estimation of luminance and reflectance maps. Traditional methods, with their limited decomposition ability, often result in over-enhancement or under-enhancement. In contrast, the learning-based approach demonstrates enhanced decomposition results and effectively improves contrast. It is noteworthy that many learning-based methods primarily utilize spatial information from low-light images to generate high-quality normal-light images, often neglecting the recovery of detailed information. Therefore, the enhancement module proposed in this paper leverages the FEM attention module, embedding it into U-Net's jump connection to augment the network's learning capacity for meaningful features. This augmentation is achieved by integrating image features and spatial attention mechanisms in the channel dimension. This design not only utilizes spatial information to strengthen contrast but also prioritizes the recovering of finer detail from the image.

In low-light conditions, image quality is often constrained by the optical signal attenuation, resulting in a significant degradation of the signal-to-noise ratio and a notable increase in noise. Some methods use the denoising model as preprocessing methods for low-light image enhancement; however, this preprocessing result in the loss of details in the low-light image. Chen et al. [16] introduced a model employing two parallel CNN branches: one for extracting brightness information and the other for extracting residual noise information. Low-light Photo Denoising via a Diffusion Model (LPDM) [17] is a denoising method for low-light images that utilizes a diffusion process. Initially, low-light images are enhanced to improve their brightness and contrast, followed by noise reduction through a diffusion process.

Although this method is effective in reducing noise, it introduces certain issues, such as the loss of some image details. To mitigate the loss of detail information during the denoising process, this paper introduces a denoising network to suppress noise in the reflection map. However, eliminating noise by restraining high-frequency signals in reflectivity map may lead to the loss of inherent details.

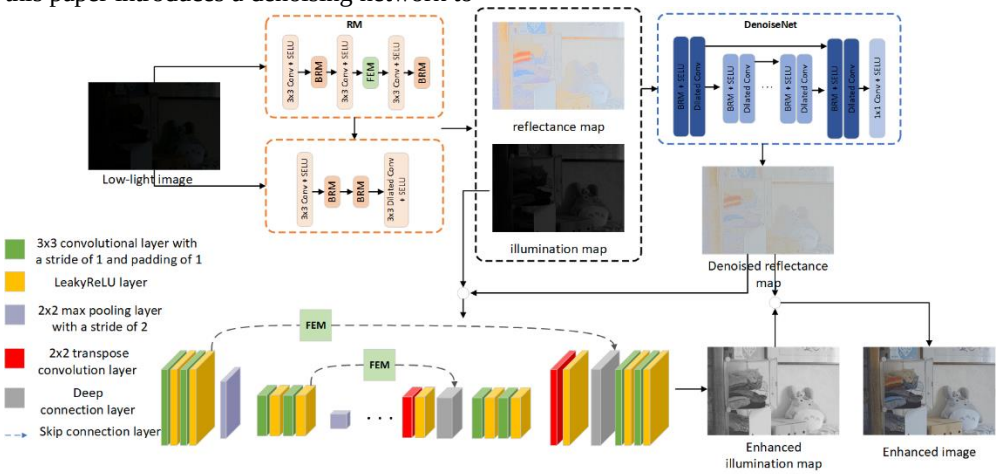

Fig. 1. The framework of the proposed method.

III. Proposed Method

As illustrated in Fig. 1, the whole pipeline is based on Retinex. Initially, a comprehensive loss function is incorporated into the decomposition network to obtain light and reflection maps. In order to mitigate the loss of detailed information, a denoising network is introduced to suppress the high frequency information in the reflection image. Finally, the FEM attention module is proposed in the enhancement network to process the light image, aiming to better preserve object details, create smoother color transitions, and yield a clearer, more natural image.

A. Decomposition Module

The loss function of decomposition network consists of perception loss function, illumination consistency loss function and global consistency loss function.

$$L = \lambda_1 \cdot L_{perceptual} + \lambda_2 \cdot L_{\text{consistency}} + \lambda_3 \cdot L_{\text{global}} \quad (1)$$

The values of λ_1, λ_2, λ_3 and are 0.3, 0.4, and 0.3.

The perception loss function is designed to preserve the perceived quality of the image. Traditional pixel level loss function often falls short in accurately capturing the perceived quality of the image. Hence, we choose a pre-trained convolutional neural network to extract image features and subsequently calculate the feature difference between the generated low-light image and the target image.

$$\mathrm{L}_{perceptual} = \frac{1}{N}\sum_{k=1}^{N} \| \phi(I)_k - \phi(L)_k \|_2^2 \quad (2)$$

Where, the input low-light image is I, the target light map is L, $\phi(I)$ represents the feature map extracted by the pre-trained convolutional neural network (such as VGG16), and N is the number of feature maps, $\|.\|_2$ stands the $L2$ norm.

The illumination consistency loss function ensures the structural information's consistency between the generated low-light image and the target image.

$$L_{perceptual} = \frac{1}{N}\sum_{k=1}^{N} \| \mathrm{G}(I)_k - \mathrm{G}(L)_k \|_2^2 \quad (3)$$

The $\mathrm{G}()$ represents the gradient of the image.

The global consistency loss function elevates the overall consistency of the generated low-light image and the target image.

$$L_{global} = \frac{1}{N}\sum_{k=1}^{N} \| \mathrm{mean}(I)_k - \mathrm{mean}(L)_k \|_2^2 \quad (4)$$

The *mean*() represents the mean of the image.

B. BRM Module

The BRM module (Fig. 2) adopts the concept of a residual network as a reference and comprises 5 convolution layers. The convolution kernel size is {1, 3, 3, 1}, and the corresponding number of the convolution kernel is {64, 128, 128, 64}. For the activation function, SELU is assigned to correspond to the convolution kernel 3, and LeakyReLu to the convolution kernel 1. Finally, a 64×1×1 convolution layer added to the jump junction.

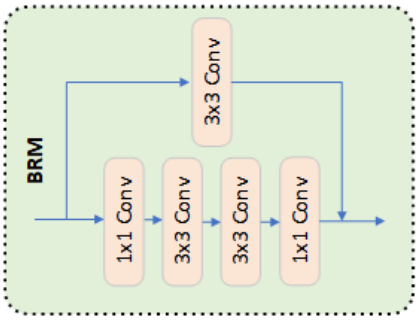

Fig. 2. The framework of BRM.

C. Denoising Module

The heavy noise in low-light images obscures essential details, structure, and other valuable information, burying useful features beneath irrelevant ones. It is these severe degradations that make the training process challenging for network learning and the recovery of useful features (such as details, structure, and corrected color information).

Previous methods often eliminate noise by suppressing high-frequency signals in the reflection map. However, those signals frequently encompass critical image details and texture information. The inhibition of high-frequency signals can lead to detail loss or blurring, resulting in visually smooth or less sharp enhanced image. In this paper, a denoising network is posed to remove noise in low-light images by considering spatial information from images. U-Net [18] has demonstrated excellent results in numerous computer vision tasks, and it is frequently employed in low-light image enhancement networks. Nevertheless, U-Net's use of multiple max pooling layers has resulted in loss of feature information. In our network, the maximum pooling layer is replaced by the extended convolution layer, enabling a broader context information range by increasing the receptive field size of the convolution kernel without reducing the feature map resolution. The BRM module from the decomposition network is integrated into the denoising network, with each sub-module of U-Net being replaced with BRM. In the denoising network's encoder part, subsampling is achieved by adding an average pooling layer with a pool core size and step size of 2 to the BRM module at the end. In the decoder part, up-sampling is realized by incorporating a deconvolution layer with convolution kernel size and step size of 2 to the BRM module. The spatial information in images often contains rich details and structural information. The denoising network proposed in this paper leverages spatial information for denoising, enhancing its ability to preserve details and resulting in clearer and more natural images after denoising.

In the denoising network, a novel loss function is proposed in this paper. By integrating the Mean Square Error (MSE) loss term with the smoothing loss term, the image's noise suppression and smoothing effect can be optimized simultaneously. The MSE loss term aids in minimizing the pixel-level difference between the real image and the low-light image, thereby mitigating the impact of noise. The smoothing loss item promotes the smooth characteristics of the low-light image, enhancing the clarity of edges and details. The specific process is as follows:

$$L = \lambda Lsmooth + Lmse \tag{5}$$

The tradeoff between smoothness and the difference at the pixel level in the denoising loss function can be controlled by adjusting the weighting factor λ for the smoothing loss term, which defaults to 0.2.

$$Lmse = ||Igi - \tilde{} Igi||^2 \tag{6}$$

Where, Igi represents the grayscale image relative to the low-light image, and $\tilde{} Igi$ represents the grayscale image relative to the real image.

$$Lsmooth = ||\nabla \tilde{} Igi||^2 \tag{7}$$

Where, $\tilde{} Igi$ represents the image processed by the denoising network, and ∇ represents the gradient operator, which is used to calculate the gradient of the image. In this paper, the Prewitte operator is arranged to represent the value of the gradient operator, which can be represented by the following two matrices:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, G_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{8}$$

D. Enhancement Module

After denoising, residual shadow blocks and halo artifacts persist in low light images. In this paper, an enhancement module is devised to remove these artifacts while improving the quality of low-light images. Notably, the generation of shadow blocks and halo artifacts can be attributed to the U-Net's jump connection, wherein severely degraded features are directly conveyed to the up-sampled stage by linking up-sampled features with previous down-sampled features, leading to the retention of degraded features.

Inspired by SENet [19] in image recognition, the FEM attention module is incorporated into U-Net's jump connection to enhance noise removal and facilitate detail recovery. This inclusion is particularly effective in eliminating shadow blocks and halo artifacts. The FEM attention module operates by integrating image features in the channel dimension, assigning higher weights to valuable features (such as the correct color, detail, and texture features). This enabling the network to better learn these crucial features, while assigning lower or zero weights to less important features (such as noise, distorted colors, shadow blocks, and halo artifacts) or even giving no weight at all.

The loss function of the enhancement module is shown as follows:

$$L_{Re} = L_{Ren-con} + \lambda L_{Re-per} \tag{9}$$

Where λ is the weight used to balance different loss terms, the default value is 0.1. $L_{Ren-con}$ the content loss value obtained by calculating the absolute value of the difference between the enhanced image and the real enhanced image at each position of pixel, and adding the absolute value of all differences. The presented loss metric quantifies the holistic disparity between the generated enhanced image and the authentic enhanced image. L_{Re-per} is the perception loss, which is gained by computing the square difference in the perception space between the generated enhanced image and the actual enhanced image, summing across all pixel positions. To maintain a balanced consideration of the various layers within the feature map, normalization is conducted by dividing the dimensions of the feature map.

Content loss is defined as follows:

$$L_{Re-con} = \sum_{i}^{N} |\tilde{} S_{low} - S_{en}| \tag{10}$$

Where i is the index of the pixel, $\square S_{low}$ is the generated enhanced image, S_{en} is the real enhanced image. $\|$ is the absolute value symbol. $\sum$ means adding the difference of each pixel, that is, adding the difference between the generated enhanced image and the real enhanced image at each pixel position.

Perceived loss means:

$$LRe-per=\frac{1}{(CjHjWj)}*\Sigma\|\varphi j(\square S_{low})-\varphi j(S_{en})\|^2 \quad (11)$$

Where Cj represents the number of channels in the feature map of the *J-th* layer and the number of channels in the feature map of different layers used in the perception loss. Hj is the height of the feature map of the *J-th* layer, which represents the height of the feature map of the different layers used in perception loss. Wj represents the width of the feature map of the *J-th* layer, representing the width of the feature map of the different layers used in perception loss. φj represents a function that maps the image to the *J-th* layer feature map, which is used to extract the representation of the features of image on the perceptual space. $\square S_{low}$ represents the enhanced image generated. S_{en} is a true enhanced image. $\|\cdot\|^2$ represent the square of the Euclidean norm of two vectors used to calculate the square of the difference in perceptual space between the generated and real enhanced images.

E. FEM Module

In recent years, a multitude of attention modules have been proposed to incorporate learnable weights in information processing, facilitating dynamic adjustments and the assignment of significance to various parts of the input data. This approach draws inspiration from human perceive and cognitive processes, enabling models to concentrate on information that significantly contributes to a given task or problem. For instance, Hu et al. [19] propose a Squeeze-and-Excitation (SE) block, which effectively performs feature recalibration by modeling the interrelationships between channels. Recognizing the importance of positional relationships between pixels, Non-Local Network (NLNet) [20] explored a nonlocal operation to capture the interactions between any two positions, irrespective of their spatial distance. Subsequently, the Cross Partial attention Volume Transformer (CPVT) module introduces an attention mechanism that crosses partial channels and divides the input feature map into several subgroups, each encompassing a subset of the channels. The attention mechanism is applied to each subgroup, focusing exclusively on channel relationships within the current subgroup and not considering channels in other subgroups. This approach achieves a balance between computational efficiency and performance.

These methods prove advantageous in addressing complex tasks like object detection and scene segmentation. A notable example is Axial-Deeplab [21], a deep learning model designed for image segmentation tasks. It employs an Axial attention mechanism for processing large-scale images and incorporates a segmented attention mechanism to enhance the capture of relationships between objects. However, these approaches may exhibit limited impact on low-level tasks, such as image enhancement. To address this issue, we propose an FEM (Fig. 3) attention module in this paper. The module effectively removes shadow blocks and halo artifacts by integrating image features and spatial attention mechanisms in the channel dimension. Furthermore, optimized jump connections are introduced, enabling adaptive exploration of contrast information within the image and facilitating the recovery of potentially fine details in low-brightness areas.

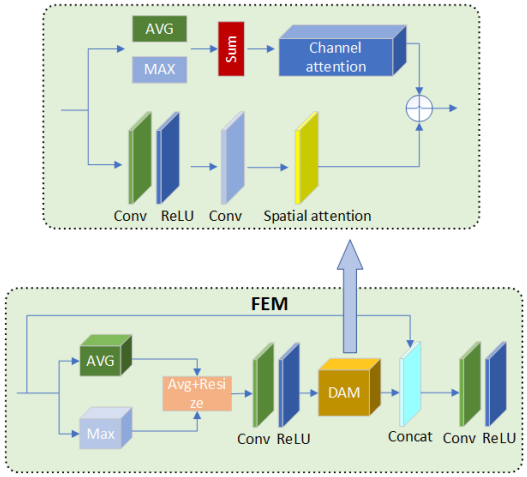

Fig. 3. The framework of FEM.

The FEM module comprises two Conv+ReLU layers, AdaptiveAvgPool2d, AdaptiveMaxPool2d, an up-sampling module and a Dual Attention Module (DAM). Specifically, for an input feature graph X with dimensions H × W × C, AdaptiveAvgPool2d and AdaptiveMaxPool2d are employed to extract representative information. The average of these two operations generates a global information feature map with dimensions 1 × 1 × C. Then, the feature map with global information undergoes amplified through up-sampling, and the number of channels is compressed using 1×1 Conv to obtain a global feature map with dimensions H × W × C1. Following this, a DAM module is introduced to extract global features from spatial and channel dimensions. The DAM consists of two input branches: channel attention branch, and spatial attention branch. For an input feature graph X with dimensions of H × W ×C, the channel attention branch employes global average pooling and global maximum pooling to generate the global average pooling feature map Cavg and the global maximum pooling feature map Cmax in spatial dimension, respectively. Their purpose is to emphasize the information regions, and these results are combined to produce the output Fc (R1×1×C) of the attention branch of the channel. The spatial attention branch aims to generate a space-based attention map. Similar to the channel attention branch, it computes Savg (RH×W ×1) and Smax (RH×W ×1) through global average pooling and maximum pooling in the channel dimension respectively. The output spatial attention map Fs (RH×W ×1) is then obtained through a convolution layer. Finally, Fc and FS are combined to rescale and optimize the global feature map, resulting in the output. The input feature map (encoding local information) and the optimized global feature map (encoding global information) are combined using the

concatenate function and the Conv+ReLU function to generate an output feature map with dimensions H × W × C.

IV. EXPERIMENTAL

A. Parameter Setting

The experiment was conducted using PyTorch 1.8.0, with network training confined to a 256×256 patch on a single NVIDIA GTX 1080Ti GPU. The batch size was set to 2 and a total of 1×10^5 iterations were performed. Data enhancement involved random horizontal and vertical flips. Employing the Adam optimizer, the initial learning rate was set to 10^{-4}, gradually reduced to 10^{-6} through a cosine annealing strategy.

B. Data Set

The low light image dataset comprises a curated collection of specialized images tailored for the examination and evaluation of image processing algorithms in low light conditions. Typically, these datasets contain images captured in settings with insufficient illumination, offering researchers a challenging assortment for the development and assessment of algorithms aimed at enhancing the quality of low-light images. Derived real-world low-light scenes, these datasets encompass derived environments, both indoor and outdoor, capturing a range of shooting conditions and objects. These images within these datasets commonly exhibit characteristics such as low contrast, indistinct light and dark details, and elevated noise levels.

In the experiment of this paper, we utilized the LOL-v1 [9] and LOL-v2-real [22] datasets. The LOL dataset consists of 500 pairs of images, each with low and normal illumination, totaling 1,000 images. The images are captured with the same camera in varying lighting conditions, these images span a diverse array of scenes, both indoor and outdoor, and cover various shooting conditions and subjects. The dataset's diversity ensures comprehensive coverage of low-light scenes, which enhance the generalization and robustness of the evaluation algorithm. The LOL v2 dataset serves as a sequel to LOL v1, consists of two pairs of training and validation images, involving the actual shot and composite-generated images. Specifically, LOL-v2 is divided into two subsets: LOL-v2-REAL and LOL-v2-synthetic. The former includes 689 pairs of low-light/normal-light images for training and 100 pairs for testing, primarily adjusted by modifying camera parameters like exposure time and ISO. The latter is generated on an illumination distribution analysis of RAW format images.

C. Evaluation Index

1) Subjective evaluation: The method presented in this paper is compared with several advanced low-light image enhancement methods, including KIND [10], KIND+ + [23], NE [24], SCI [13] and RetinexNet [9]. Experiments are conducted using publicly available source code provided by the authors of these methods.

The results depicted in Fig. 4, The RetinexNet method overly smoothens details and even causes color deviations, making the image look unnatural. Moreover, the results of RetinexNet still contain a lot of noise. Although it uses BM3D to remove noise from the decomposed reflectance component, it cannot clearly remove the noise. The main reason is that BM3D is designed to remove Gaussian noise with a fixed noise level. However, the noise in the reflectance component is more diverse and complex than Gaussian noise. Although KIND and KIND + + introduced a recovery network to recovery color and remove noise from reflected images, the results were still inconsistent. For instance, in the first row of the Fig. 4, the KIND++ enhanced image still displays color deviation when compared to the reference image. In the second row, KIND treats a small light source as noise and removes it. In the sixth row, it is evident that KIND has unevenly enhanced the image. In the third image, the enhanced KIND ++ image still has color bias compared to the reference image. The SCI based method produces visually appealing results, it carries some undesirable artifacts (such as white walls). In contrast, upon observing the visualization results, particularly the outline of the teddy bear in the first row of Fig. 4 and the texture details of the book in the third row, our proposed method outperforms in terms of enhancement, reduced noise, and more accurate detail recovery. Conversely, other methods yield a fuzzy recovery effect due to noise interference in low light images, with the recovered images retaining significant noise. By comparing the results of different methods in the fifth line of images, our proposed method performs superior performance in restoring color saturation, presenting a more natural and vivid color enhancement effect. Moreover, in contrast to the restored color of the window in the second row of Fig. 4 and the floor in the last row, it can be concluded that our proposed method excels in maintaining color fidelity, suppressing noise, and removing artifacts.

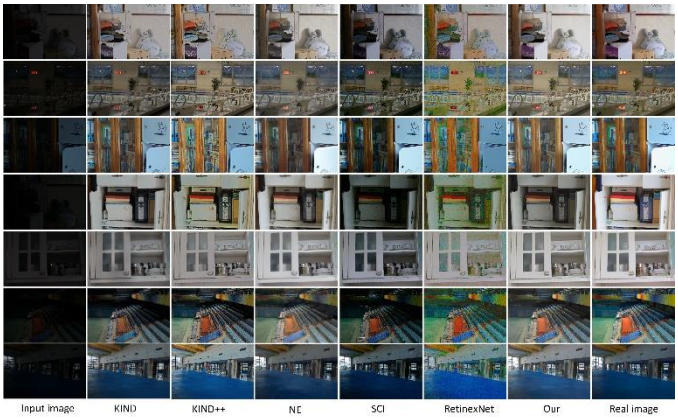

Fig. 4. Subjective comparison on the LOL-V1 dataset.

Fig. 5 illustrates the impact of the experiment detailed in this paper, along with comparisons to other experiments on the LOL V2 dataset. Although RetinexNet can enhance the low-light areas in images, it exhibits severe color distortion and halo artifacts (a significant amount of halo artifacts can be observed around the contours of the mountains in the second row of images). In contrast, our method effectively brightens the low-light areas reasonably while maintaining the realistic visibility of the result. Besides enhancing brightness, our method also successfully reduces color distortion and halo artifacts. SCI introduces and even amplifies noise after enhancement, and therefore suffers from severe noise

distortion and color degradation when brightening dark areas. Conversely, our experimental results have superior color-recovery performance. KIND tends to exhibit some under-enhanced areas and apparent color distortions. Although the KIND ++ can remove noise, excessive sharpening (the surface of the mountain in the second row and the intersection of branches in the sixth picture) may introduce other problems such as loss of detail, blurring, and causing the image to unnatural. Compared to all these methods, our method can effectively enhance the brightness and display details of the image while suppressing noise, and it presents the most abundant and reasonable color information.

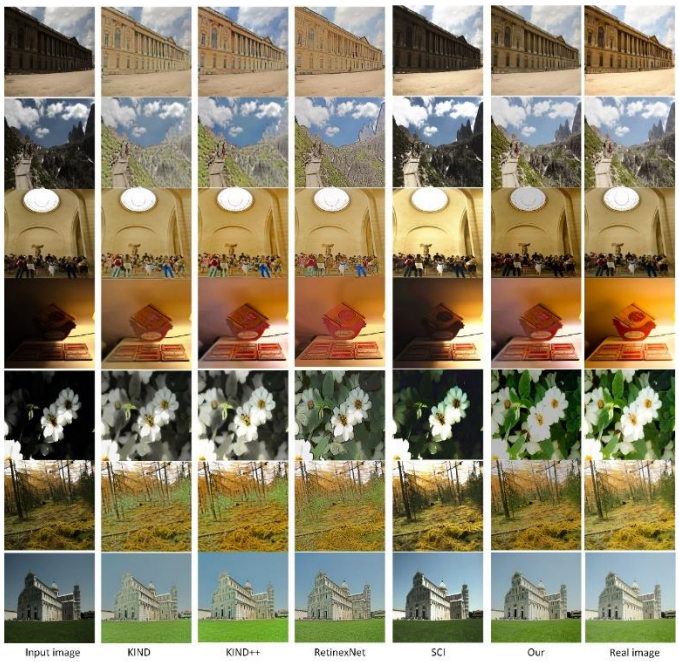

Fig. 5. Subjective comparison on the LOL-V2dataset.

2) Objective evaluation: According to Table I, our method achieves superior results, which manifests that the enhanced images obtained by this pipeline are more visually satisfactory.

To objectively evaluate the enhancement results, we employed four classical indicators, and the results are presented in Table I. Among these metrics, PSNR [25] is utilized to measure the noise level and degree of distortion between the original image and the processed image. A higher PSNR value indicates a closer result to the reference image at the pixel-level. SSIM [26] is employed to evaluate structural similarity, considering perceptual properties such as image structure and content. A higher SSIM value indicates a greater similarity in structure to the reference image. The method proposed in this paper achieves excels in both PSNR and SSIM, highlighting its advantages in lighting restoration and structural restoration. KIND and NE also achieved high PSNR values, indicating their effectiveness in restoring global illumination. GMSD [27] assesses image quality by comparing gradient amplitude difference between the original and the processed images. NIQE [28] quantifies the effect of the image enhancement algorithm by analyzing the statistical characteristics of the image and generating a continuous value score. A lower NIQE score indicates higher quality detail, brightness, and tone, indicative of a more natural appearance devoid of artifacts or pseudo-details. Consistently, our method achieves superior performance in both PSNR and SSIM. The enhanced images generated by our pipeline exhibit a more natural and vivid appearance, showcasing enhanced global and local contrast.

TABLE I. COMPARISON OF OBJECTIVE EVALUATION INDICATORS OF DIFFERENT MODELS

	KIND	KIND++	NE	SCI	RetinexNet	Our
PSNR	21.38	19.21	22.61	20.80	18.4	23.15
SSIM	0.85	0.79	0.82	0.72	0.62	0.88
NIQE	0.610	0.556	0.526	0.470	0.831	0.419
GMSD	0.060	0.106	0.091	0.063	0.137	0.040

D. Ablation Experiment

1) Comparison of the effectiveness of FEM module: To validate the efficacy of the FEM module, a model was trained with the FEM module replaced by an ordinary convolution. Fig. 6 illustrates the impact of FEM module on image enhancement. The results indicate that the model incorporating the FEM module effectively preserve object details and achieves smoother color transition in the image. Notably, in the third and fourth images, the enhanced results closely approximate the real image, aligning with the characteristics of the natural landscape. Analysis of the data presented in Table I reveals a notable improvement when utilizing models with FEM module compared to those without. Specifically, the PSNR increases by 3.74 (=20.22-16.48) and SSIM increases by 0.18 (=0.81-0.63). These finding lead to conclusion that models incorporating FEM modules exhibit superior performance in image enhancement.

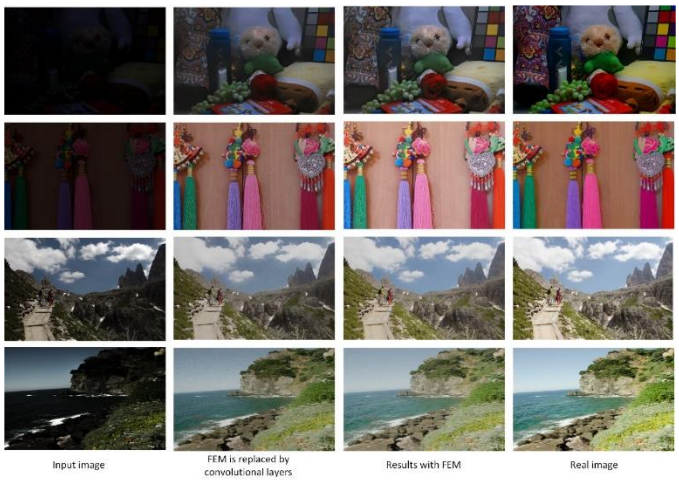

Fig. 6. Ablation experiments to verify the effectiveness of the FEM model.

2) The effectiveness of the denoising network: To validate the efficacy of the denoising module, a comparison was made between a method trained without the denoising network and the original method. As depicted in the Fig. 7, the

experimental results employing the denoising network effectively remove the noise from the image, and the results after the removal of noise contain finer details and more vivid colors. *It is* illustrate*d* in the Table *II* that the method incorporating the denoising network improves the PSNR ratio by 2.88 (22.77-19.29) and the SSIM ratio by 0.17 (0.84-0.67). These results demonstrate the effectiveness of the denoising network in this experiment.

Fig. 7. Ablation experiments to verify the effectiveness of the FEM module.

TABLE II. RESULTS OF THE OBJECTIVE EVALUATION OF THE ABLATION EXPERIMENTS

Different situations	PSNR	SSIM	NIQE	GMSD
No FEM module	16.48	0.63	0.51	0.083
FEM module	20.22	0.80	0.46	0.066
No denoising network	19.29	0.67	0.58	0.073
denoising network	22.17	0.84	0.43	0.042

V. CONCLUSIONS

This paper proposes a low light image enhancement method based on Retinex. We propose an efficient network for decomposing a low light image, incorporating an innovative loss function that integrates perceptual, light consistency and global consistency to obtain high quality light and reflection maps. The decomposition network comprises a reflection image extraction module (RM) and an illumination image extraction module (IM). Additionally, we integrate a denoising network and an enhancement module to further improve image quality. Our method not only enhances image color smoothness, reduce artifacts, but also effectively remove noise to restore image details under low-light conditions. By employing the FEM attention module instead of the convolution layer, our method successfully preserves object details. This results in a smoother color transition, yielding clearer and more natural images. Experimental results demonstrate that the proposed method achieves significant performance improvement across various low-light scenes. When compared to other existing methods, our method excels in image enhancement and detail recovery, showcasing superior noise removal and artifact suppression. In future work, we aim to extend the application of this method to other computer vision tasks, substantiating its versatility and performance advantages in different domains through comparisons with other state-of-the-art methods.

ACKNOWLEDGMENTS

This work was supported by Beijing Natural Science Foundation (4232026), National Natural Science Foundation of China (Grant Nos. 62272049, 62171042, 61871039, 62102033, 62006020); Key project of science and technology plan of Beijing Education Commission (KZ202211417048); The Project of Construction and Support for high-level Innovative Teams of Beijing Municipal Institutions (No. BPHR20220121); the Collaborative Innovation Center of Chaoyang (No. CYXC2203); Scientific research projects of Beijing Union University (Grant Nos. ZK10202202, BPHR2020DZ02, ZK40202101, ZK120202104).

REFERENCES

[1] Shrikhande, Saachi, Siddhesh Borse, and Shripad Bhatlawande. Face Recognition Based Attendance System. No. 10070. EasyChair, 2023.

[2] Li, X., Chen, S., Hu, X., & Wang, J. (2021). Autonomous driving using deep learning: A survey. IEEE Transactions on Intelligent Transportation Systems, 23(10), 3789-3813.

[3] Ma, S., Pan, W., Liu, H., Dai, S., Xu, B., Xu, C., ... & Guan, H. (2023). Image Dehazing Based on Improved Color Channel Transfer and Multiexposure Fusion. Advances in Multimedia, 2023.

[4] Pizer S M, Amburn E P, Austin J D, et al. Adaptive histogram equalization and its variations[J]. Computer vision, graphics, and image processing, 1987, 39(3): 355-368.

[5] Abdullah-Al-Wadud, M., Kabir, M. H., Dewan, M. A. A., & Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. IEEE transactions on consumer electronics, 53(2), 593-600.

[6] Land, Edwin H. "The retinex theory of color vision." Scientific american 237.6 (1977): 108-129.

[7] Jobson, D. J., Rahman, Z. U., & Woodell, G. A. (1997). Multiscale Retinex for color image enhancement. IEEE Transactions on Image Processing, 6(7), 965-976.

[8] Gao, Yuhang, Chuhao Su, and Zhaoheng Xu. "Color image enhancement algorithm based on improved Retinex algorithm." 2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML). IEEE, 2022.

[9] Wei, C., Wang, W., Yang, W., & Liu, J. (2018). Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560.

[10] Zhang, Yonghua, Jiawan Zhang, and Xiaojie Guo. "Kindling the darkness: A practical low-light image enhancer." Proceedings of the 27th ACM international conference on multimedia. 2019.

[11] Lim, Seokjae, and Wonjun Kim. "DSLR: Deep stacked Laplacian restorer for low-light image enhancement." IEEE Transactions on Multimedia 23 (2020): 4272-4284.

[12] Li C, Guo J, Porikli F, et al. LightenNet: A convolutional neural network for weakly illuminated image enhancement[J]. Pattern recognition letters, 2018, 104: 15-22.

[13] Ma, Long, et al. "Toward fast, flexible, and robust low-light image enhancement." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[14] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., ... & Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing, 30, 2340-2349.

[15] X. Wei, X. Zhang, S. Wang, C. Cheng, Y . Huang, K. Yang, and Y . Li,"Da-drn: Degradation-aware deep retinex network for low-light image enhancement," arXiv preprint arXiv:2110.01809, 2021.

[16] Chen, C., Chen, Q., Xu, J., & Koltun, V. (2018). Learning to see in the dark. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3291-3300).

[17] Panagiotou, Savvas, and Anna S. Bosman. "Denoising Diffusion Post-Processing for Low-Light Image Enhancement." arXiv preprint arXiv:2303.09627 (2023).

[18] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015.

[19] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[20] Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7794-7803).

[21] Wang, Huiyu, et al. "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation." European conference on computer vision. Cham: Springer International Publishing, 2020.

[22] Yang, W., Wang, W., Huang, H., Wang, S., & Liu, J. (2021). Sparse gradient regularized deep retinex network for robust low-light image enhancement. IEEE Transactions on Image Processing, 30, 2072-2086.

[23] Zhang, Y., Guo, X., Ma, J., Liu, W., & Zhang, J. (2021). Beyond brightening low-light images. International Journal of Computer Vision, 129, 1013-1037.

[24] Jin, Yeying, Wenhan Yang, and Robby T. Tan. "Unsupervised night image enhancement: When layer decomposition meets light-effects suppression." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.

[25] Huynh-Thu, Quan, and Mohammed Ghanbari. "Scope of validity of PSNR in image/video quality assessment." Electronics letters 44.13 (2008): 800-801.

[26] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

[27] Xue, W., Zhang, L., Mou, X., & Bovik, A. C. (2013). Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE transactions on image processing, 23(2), 684-695.

[28] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. "No-reference image quality assessment in the spatial domain." IEEE Transactions on image processing 21.12 (2012): 4695-4708.