Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.042416

ARTICLE

RF-Net: Unsupervised Low-Light Image Enhancement Based on Retinex and Exposure Fusion

Tian Ma, Chenhui Fu*, Jiayi Yang, Jiehui Zhang and Chuyang Shang

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an, 710054, China

*Corresponding Author: Chenhui Fu. Email: SkyChh@foxmail.com

Received: 29 May 2023 Accepted: 25 August 2023 Published: 31 October 2023

ABSTRACT

Low-light image enhancement methods have limitations in addressing issues such as color distortion, lack of vibrancy, and uneven light distribution and often require paired training data. To address these issues, we propose a two-stage unsupervised low-light image enhancement algorithm called Retinex and Exposure Fusion Network (RF-Net), which can overcome the problems of over-enhancement of the high dynamic range and under-enhancement of the low dynamic range in existing enhancement algorithms. This algorithm can better manage the challenges brought about by complex environments in real-world scenarios by training with unpaired low-light images and regular-light images. In the first stage, we design a multi-scale feature extraction module based on Retinex theory, capable of extracting details and structural information at different scales to generate high-quality illumination and reflection images. In the second stage, an exposure image generator is designed through the camera response mechanism function to acquire exposure images containing more dark features, and the generated images are fused with the original input images to complete the low-light image enhancement. Experiments show the effectiveness and rationality of each module designed in this paper. And the method reconstructs the details of contrast and color distribution, outperforms the current state-of-the-art methods in both qualitative and quantitative metrics, and shows excellent performance in the real world.

KEYWORDS

Low-light image enhancement; multiscale feature extraction module; exposure generator; exposure fusion

1 Introduction

With the rapid development of artificial intelligence, low-light image-enhancement technology has been widely applied for pre-processing in advanced visual tasks. However, low-light images often suffer from detail degradation and color distortion due to the shooting environment and technical limitations. Balancing the image-enhancement effect and maintaining image realism are challenging problems in low-light image enhancement. These problems can significantly affect the performance of advanced downstream vision tasks. Therefore, improving visual quality and recovering image details have become important research topics.

In the process of image enhancement, it is important to strike a balance between preserving image details and maintaining overall image quality. This requires preserving the original details in

well-exposed areas, while appropriately brightening the underexposed areas to achieve a high-quality image. In addition, attention must be paid to balancing the brightness and contrast of the image during enhancement. If only the brightness is increased globally, the texture details in the image may be lost. Therefore, both brightness and contrast changes must be considered when enhancing an image to ensure its quality. Traditional methods [1–3] in the past often required a large amount of manual parameter adjustment to improve image quality; however, this method has significant limitations, as its effectiveness is largely based on assumptions regarding the threshold range. The use of low-light/normal-light images in supervised deep model training has become the main approach in algorithm research owing to advancements in deep learning. The accuracy of supervised learning methods depends on paired training datasets. However, it is technically difficult to obtain paired datasets from the same scene. In addition, the algorithm has poor generalization ability and cannot be effectively applied to real-scene images. In recent years, unsupervised image enhancement algorithms have emerged that eliminate reliance on paired datasets and achieve good enhancement results. For example, Deep Light Enhancement without Paired Supervision (EnlightenGAN) [4] uses unpaired datasets to train and implement low-light image enhancement techniques. Zero-Reference Deep Curve Estimation (Zero-DCE) [5] achieves enhancement using scene images with different illumination intensities. Although these methods eliminate the dependence of deep learning techniques on paired datasets, the quality of enhancement remains a challenge. The EnlightenGAN [4] method may produce artifacts, an overall uneven picture, and color-recovery errors when enhancing dark areas. Images enhanced using the Zero-DCE [5] method may exhibit whitish tones and less vibrant colors. These methods exhibit stronger generalization ability than supervised methods and reduce the requirements for dataset collection.

To address these issues, we propose an unsupervised enhancement network called RF-Net, which combines Retinex with exposure fusion. The network comprises two stages: image decomposition and exposure fusion. In the first stage, to fully consider contextual and global information, we employed the powerful image-generation capabilities of a generative adversarial network and designed a multi-scale feature-extraction module to produce high-quality illumination and reflection images. Specifically, our network uses a multi-scale feature extraction module to perceptively capture different-scale features, preserve more detailed information, and avoid information loss between layers by using residual connections to transmit information from the current layer to the next layer. Most existing Retinex-based image enhancement methods obtain illumination and reflection component information matrices and generate enhanced images through calculations, which not only involve high computational complexity but also result in artifacts when processing shadow parts in dark areas. After obtaining the illumination and reflection images, an exposure image generator with correction coefficients was designed using the camera response function in the second stage to generate the exposure image and fuse it with the original low-light image to complete low-light image enhancement. The results obtained using the proposed method are shown in Fig. 1.

In summary, the main contributions of this paper are as follows:

1. We devised a multi-scale feature-extraction module to produce high-quality illumination and reflection images. We incorporated a Coordinate Attention (CA) module that includes position-encoding information into the Markov discriminator. This module builds on channel attention and pays closer attention to the location information of the generated image, allowing for more accurate discrimination of texture details and improving the quality of the images generated.
2. We improved the original Retinex formulation and designed an exposure image generator module with correction coefficients by referring to the camera response mechanism function.

This module can generate images with different exposure levels while fusing illumination and reflection images.

3. We proposed a novel unsupervised image enhancement method called RF-Net, which exhibits excellent performance in test results on several datasets and can be generalized to real-world low-light conditions.

Figure 1: Representative enhancement results of RF-Net. Which can improve over-enhancement in the high dynamic range and under-enhancement in the low dynamic range

2 Related Work

In this section, we review research on low-light image enhancement using traditional and deep learning methods.

2.1 Traditional Methods

Histogram equalization is a classical image-enhancement method that enhances the contrast of an image by adjusting its brightness distribution. However, histogram equalization tends to cause image noise and over-enhancement problems. Some methods further increase the enhancement effect by setting a threshold to divide the image blocks [6], dividing the clipping points into chunks for processing [7], and combining them with adaptive gamma correction [8] to obtain a more reasonable S-shaped mapping function. However, these methods lead to problems of over-enhancement and amplification artifacts. The contrast was enhanced to an extent, but the details were lost. Retinex theory [9] posits that an image's brightness is composed of two parts, reflection, and illumination, which enhance the image quality by separating these two parts. However, this approach is ineffective for images that are too dark or bright. Accordingly, researchers have proposed various improvement

schemes [1–3], etc. This presupposes that spatial illumination changes slowly during implementation, but the processing is prone to halation and inaccurate color recovery. To reduce the computational cost in Retinex theory, researchers have proposed Low-Light Image Enhancement via Illumination Map Estimation (LIME) [10], the local illumination distribution of the image obtained by analyzing the local information. The local illumination distribution is applied to the reflection component to obtain the enhanced image. Compared with the Retinex method, LIME reduces the occurrence of halo artifacts during processing. However, LIME has a limited ability to distinguish between the foreground and background of an image, which can result in over-enhancement of the foreground and noise in the background.

2.2 Deep Learning Methods

Researchers have widely employed deep learning for image enhancement over the past decade, achieving promising results. In the following section, we review the current state of research on fully supervised, semi-supervised, and unsupervised approaches.

2.2.1 Fully Supervised Methods

The use of paired datasets to train network models has been widely adopted because of the one-to-one correspondence between the training data. RetinexNet [11], for the first time, combines Retinex theory with Convolutional Neural Networks to implement the low-illumination image enhancement problem by designing a decomposition module, and enhancement module. The feasibility of the Retinex application in deep learning was demonstrated for the first time. Kindling the Darkness (KinD) [12] designed global and local enhancement modules, where the global module extracts global luminance based on Retinex decomposition, and the local module enhances the image texture details. The global and local modules interacted through an adaptive mechanism, and the difference between the image generated by the global enhancement branch and the original image was used to calculate the weight of each pixel. These weights were then passed to the local enhancement branch to achieve contrast enhancement. KinD++ [13] is based on KinD and improves the training speed and accuracy of the model by designing group learning and back-propagation mechanisms. GLobal Illumination-aware and Detail preserving Network (GLADNet) [14] generates a global before light by designing a global illumination estimation module that is then combined with the original input image to produce an enhanced image. Low-Light Image Enhancement with Normalizing Flow (LLFlow) [15] uses adaptive weights to control the effects of optical flow and global constraints; it also uses a deep learning model to learn the optical flow and the image to obtain an enhanced image. Self-Calibrated Illumination (SCI) [16] reduces computational costs by designing an adaptive correction illumination module that ensures the convergence of the results of each training phase to the final one. The literature [17] designs a generative adversarial network containing dual attention units that can effectively inhibit the artifacts and color reproduction bias generated during the enhancement. Transformer Photo Enhancement (TPE) [18] uses a pure transformer architecture to implement image enhancement based on multi-stage curve adjustment. Retinex based deep unfolding network (URetinex-Net) [19] decomposes the input image by designing a continuous optimization model with mutual feedback. To optimize the decomposition results, an implicit a priori regularization model was used, and a data initialization module, specific illumination intensity module, and denoising detail retention module were designed. Illumination Adaptive Transformer (IAT) [20] implements low-light enhancement by designing a lightweight transformer model that uses attention-query techniques to represent and adjust the parameters associated with the image signal processor (ISP).

2.2.2 Semi-Supervised Methods

These methods can learn better feature representations using both paired and unpaired data. First, researchers use paired datasets to train and obtain prior knowledge, and then they use the trained model as pre-training weights for unpaired data training. Based on this subdivision, Deep Recursive Band Network (DRBN) [21] introduces a recursive network architecture that uses the information of the highlight and shadow regions of the image and constructs a low-rank matrix and a sparse matrix to represent the brightness and structural information of the image, then inputs the two matrices into two branches of the network for feature extraction, and finally merges them to obtain an enhanced image. DRBN [22] utilizes a "band representation" technique to enhance low-light images. This method decomposes a low-light image into multiple bands using band representation and trains a neural network with a small amount of labeled data to learn how to enhance each band. Thus, this method retains the detail and texture information of a low-light image, thereby enhancing its quality.

2.2.3 Unsupervised Methods

Obtaining paired datasets can be difficult and using unsupervised methods has become the main approach for accomplishing image-enhancement tasks. This approach improves generality and applicability to many real-world scenarios. Exposure Correction Network (ExCNet) [23] is the first unsupervised enhancement method that uses the powerful learning ability of neural networks to estimate the most suitable "S" curve for low-light images and uses this curve directly to enhance the image. low-light image enhancement network (LEGAN) [24] is enhanced by a cleverly designed light perception module and a loss function that solves the overexposure problem. The EnlightenGAN [4] completed its first unsupervised image enhancement using unpaired datasets. This design overcomes the previous reliance on paired datasets by establishing unpaired mappings between low-light and non-matching images and employing global-local discriminators and feature retention losses to constrain the feature distance between the enhanced and origin images for enhancement. Zero-DCE [5] uses a neural network to match a brightness mapping curve and then generates an enhanced image based on the curve. Unsupervised low-light image enhancement was achieved by designing a multi-stage high-order curve with a pixel-level dynamic range adjustment. Based on this, a lightweight version, Zero-DCE++ [25], was developed. Generative adversarial network and Retinex (RetinexGAN) [26] uses a generative adversarial network based on Retinex to design a decomposition network with a simple two-layer convolution and achieve low-light image enhancement through image fusion. Restoration of Underexposed Images via Robust Retinex Decomposition (RRDNet) [27] achieves enhancement by designing a three-branch decomposition network and iterative loss functions to decompose the three components of reflection, illumination, and noise module. Retinex-inspired Unrolling with Architecture Search (RUAS) [28] used neural structure search to find an effective and lightweight set of networks for low-light enhancement. Retinex Deep Image Prior (RetinexDIP) [29] proposes a Retinex-based generation strategy that reduces the coupling between two components in the decomposition, making it easier to adjust the estimated illumination to perform an enhancement.

3 Proposed Method

In this section, the first module introduces the overall structure of the RF-Net network and provides a hierarchical representation of the first- and second-stage exposure-generation fusion networks. The second module describes the designed exposure image generator, and the third module describes the loss function.

3.1 Network Architecture Design

The proposed RF-Net is a two-stage network with the overall structure shown in Fig. 2. In the first stage, two coupled-generator frameworks were used. The R network generates reflection images, whereas the L network generates illumination images. First, by cascading the maximum, minimum, and mean values of each channel of the original low-light image as inputs to the network, a multi-scale feature extraction module was designed to maintain the global consistency of good illumination and contextual information. After basic feature extraction, the resulting features are concatenated and mapped to high-dimensional information before being reduced in dimension through convolution, thereby improving image quality while learning complex features. For the discriminator, we used a VGG-based network structure in which the original Markov discriminator maps the input to an $N \times N$ matrix, such that each point in the matrix corresponds to the evaluation value of each region. To enhance the decomposition effect and discrimination accuracy of the network, we incorporated a CA attention mechanism with positional information [30]. This further enhances the network's ability to perceive and understand images and spatial information, learn useful features, and suppress irrelevant features, thereby improving the discriminator's ability to accurately judge texture details.

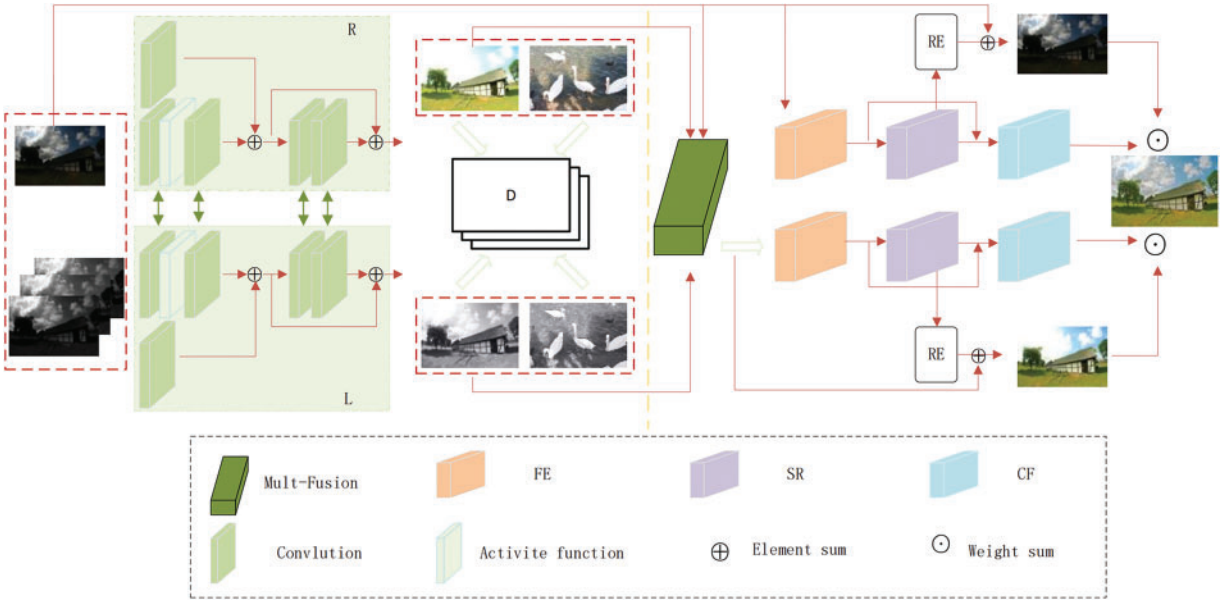

Figure 2: Overview of RF-Net. First, the low-light image and the corresponding maximum, minimum, and average grayscale images are input. The R and L have the same structure and are used to acquire the illumination and reflection components, respectively. Then the exposure image is acquired by the exposure generator. Finally, the low-light image is fused with the exposure image

The second stage also consists of two coupled networks that use the original input image and the output of the first-stage network as inputs. At this stage, the output information from the first stage is processed by the exposure image generator module to create the initial exposure image. The details of this module are discussed in Section 2. The exposed image and original input image were then separately fed into two-branch networks [31]. The two branches consist of a feature extraction module (FE), a super-resolution module (SR), and a feature fusion module (CF). The FE module consists of two convolution layers: SR uses the Convolutional Networks for Biomedical Image Segmentation (U-net) structure to learn more advanced features, and the first two modules are used to extract

advanced features from the input low-dynamic-range images. RE was employed for super-resolution of the original input image before fusion to ensure the accurate extraction of high-level image features. The final module is the image fusion module that combines the super-resolution outputs of the two coupled networks and generates the output by weighting the super-resolution of the original image and the outputs of the two coupled blocks.

3.2 Exposure Generation Module Design

We employed a design that combines Retinex with the camera response mechanism to create an exposure image generator. In the original Retinex theory, the input image S_{output} is represented as the element-wise product of the illumination and reflectance components, which is expressed as Eq. (1):

$$S_{output} = R \times L \tag{1}$$

The input image is denoted as S, the reflected image as R, the illuminated image as L, and $R \times L$ denotes pixel-wise multiplication. However, numerous experiments have shown that the results obtained from the original Retinex equation are over-enhanced and lose detail owing to noise and uneven illumination. Therefore, we improved the original formula by first inverting the source illumination image using Eq. (2) to better utilize the content in the relatively overexposed region.

$$I_{inv} = 1 - L \tag{2}$$

The improved Retinex formula is represented by Eq. (3).

$$S_{ou} = R + L + e^{9(1-\alpha)} \times I_{inv} \times S_{in} \tag{3}$$

where L and R denote the illuminated and reflected images, respectively, S denotes the original input low-light image, and S_{ou} denotes the output result of the improved Retinex formula.

To maintain a balance between brightness and contrast, we re-designed the improved Retinex formula using a camera response mechanism and proposed an exposure image generator. Here, we refer to the camera response function described in [32], where the model parameters of the camera response mechanism are determined by the camera's parameters α, β and k. Parameter k is a correction factor that can be adjusted to obtain images with different exposure levels. As the k value increases, a brighter exposed image is acquired, and the details in the low-light areas become more significant; however, when the k value is too large, more detailed information is lost because of an exposure level that is too high. Therefore, we limited the value of k to a range of 2–6. The equation for generating the initial exposure image by combining the improved Retinex and camera response functions is expressed as Eq. (4).

$$S_{eo} = e^{b\left(1-k^{\alpha}\right)} * S_{ou}{}^{\left(k^{\alpha}\right)} \tag{4}$$

where α and β are fixed parameters suitable for most cameras, with α set to -0.3293 and β set to 1.1258. S_{eo} represents the output exposure image, and different values of k directly affect the resulting output of S_{eo}.

3.3 Loss Function

Adversarial loss: In [11,12], by decomposing a low-light image into illumination and reflection components, the illumination components are approximately the same as those decomposed in a normally exposed image. With only differences in brightness, the reflection component is the same as the reflection component decomposed from the normal exposure image, which can be decomposed into high-quality reflection components by noise reduction. This means that the distribution of

normally exposed images is very similar to that of the original images based on Retinex decomposition. Therefore, the original function [33] was used as an adversarial loss function to train the generator. In practical applications, the generated fake and real input samples are encoded as zero and one, respectively. Discriminators for the illumination and reflection maps were trained using squared error as the objective function. Our adversarial losses are defined by Eqs. (5)–(8).

$$\ell\left(D^L\right) = \frac{1}{2}\mathrm{E}_{g^L}\left[\left(D^L\left(g^L\right)\right)^2\right] + \frac{1}{2}\mathrm{E}_{y^3_{mean}}\left[\left(D^L\left(y^3_{mean}\right) - 1\right)^2\right] \tag{5}$$

$$\ell\left(G^L\right) = \frac{1}{2}\mathrm{E}_{g^L}\left[\left(D^L\left(g^L\right) - 1\right)^2\right] \tag{6}$$

$$\ell\left(D^R\right) = \frac{1}{2}\mathrm{E}_{g^R}\left[D^R(g^R))^2\right] + \frac{1}{2}\mathrm{E}_{y}\left[\left(D^R\left(y\right) - 1\right)^2\right] \tag{7}$$

$$\ell\left(G^R\right) = \frac{1}{2}\mathrm{E}_{g^R}\left[\left(D^L\left(g^R\right) - 1\right)^2\right] \tag{8}$$

where g^L and g^R denote the reflected and illuminated images, respectively, y^3_{mean} denotes the average grey scale value of each channel. y represents the ground-truth reflection map D^L and D^R represent the discriminators for the illumination and reflection maps, respectively.

Perceptual loss: Using non-matching images for unsupervised image enhancement implies that the pixels in the training images are not one-to-one. The same pixel may have different semantics in different images. Therefore, we need a loss function to address the issue of non-corresponding pixel positions, and the perceptual loss function serves this purpose. This function is typically defined in the activation layer of a pre-trained network. By computing the distance between the activation layer features, it can effectively quantify the fundamental attributes of an image as well as the differences between its detailed features and high-level semantic information. This lays the foundation for generating high-quality images. Unlike common perceptual losses, this study adopted the concept of perceptual loss from [31]. This implementation not only maintains the luminance consistency between the original and reference images but also recovers the details better. The perceptual loss function used in this paper is given by Eq. (9).

$$\ell_p = \frac{1}{C_i H_i W_i}\left\|\phi_i\left(y\right) - \phi_i\left(g\right)\right\|_2^2 \tag{9}$$

where denotes the features extracted by the predefined network VGG19. $C_i H_i W_i$ denotes the number of channels and the height and width of the feature map in layer i. y denotes the unpaired real image information learned by the discriminator, g denotes the image generated by the generator.

The total loss function of the network is shown in Eq. (10).

$$L_{total} = \ell\left(G^I\right) + \ell\left(G^R\right) + \ell_p \tag{10}$$

4 Experiment

4.1 Experimental Details

We trained the RF-Net model on 914 randomly selected pairs of asymmetric datasets using the datasets provided in [4] and tested it on various datasets, including NPE [34], DICM [35], LIME [10], MEF [36], and VV. These datasets contain various low-light and unevenly exposed images from both indoor and outdoor settings. To demonstrate the performance of the enhancement algorithm better, we selected images with significant exposure differences from each test set as test set, which made the test

more challenging. The deep learning framework used was PyTorch, and the hardware configuration was Tesla A100.

To ensure that the training of the network could fully utilize the computing and storage resources of the computer and obtain better training results, the size of the training image was 640 × 400, and we randomly cropped the training data into a face slice of size 300 × 300. The batch size was set to one. To increase the data diversity, data expansion was performed, including random flipping, rotation, and cropping. This allowed the network to adapt better to various image scenes. The Adam optimizer was used to optimize the network, and the learning rate was set to 1e-4. Our network achieved better enhancement results with no more than 50 training iterations.

4.2 Performance Evaluation

To demonstrate the advantages of our proposed RF-Net, we compared our method with 10 other advanced methods: RetinexNet [11], EnlightenGAN [4], KinD [12], KinD++ [13] RUAS [28], Zero-DCE [5], LLFlow [15], SCI [16], IAT [20], and GLADNet [14]. Among these, RetinexNet, KinD, KinD++, RUAS, LLFlow, SCI, IAT, and GLADNet are supervised enhancement methods, whereas [4] and [5] are unsupervised enhancement methods. To ensure fairness, tests were conducted using the network parameters recommended in each study. Because we were unable to train the supervised models on unpaired datasets, we evaluated them using pre-trained models saved in the original papers. For unsupervised methods, if the original study used unpaired datasets for training, we used the datasets provided by [4]. If the method used images with different exposures for training, we used the datasets provided by [5]. Finally, the optimal model was selected for testing. These comparisons enabled us to evaluate the performance of the RF-Net method and demonstrate its competitiveness in image enhancement.

4.2.1 Qualitative Comparison

Figs. 3–5 show a visual comparison of the different algorithms for several special scenes. These source images contain information on well-characterized scenes. However, the results of other competing methods, particularly supervised methods, tend to lose the consistency of global information after enhancement, resulting in over-enhancement. For example, in Fig. 3, we highlight the regions of low-light and partial exposure. The results show that, in (b), the restoration of a person's face produces severe color deviation, whereas in (e), there are obvious artifacts on the face. Some task faces in (f), (i), (j), and (k) are moderately enhanced, but details in a few of the exposed regions are completely lost. In the other partially exposed image regions, the results were normal, but the enhancement was weak in the low-light regions. In (c) and (g), the enhancement results are good, but artifacts appear on the face of the person in (c), and the task face in (g) appears overexposed and has unsaturated colors. In Figs. 4f, 4i, and 4j, the sky details are severely lost, and the enhancement is not evident for the dark areas. In contrast, (d), (e), and (h) managed to enhance the dark areas, but the noise in these areas was also amplified. Several artifacts were produced; in (b), the image color was restored accurately, but there was significant noise when zoomed in. For the unsupervised methods, Fig. 4c also exhibited a significant color deviation, while (g) produced good results. In Fig. 5, the sky information was lost entirely in (f), (i), and (j), and the color restoration was inaccurate in (h) and (k). In a comprehensive comparison, our enhancement results successfully eliminated artifacts, enhanced the details, and achieved more natural illumination and dark tones.

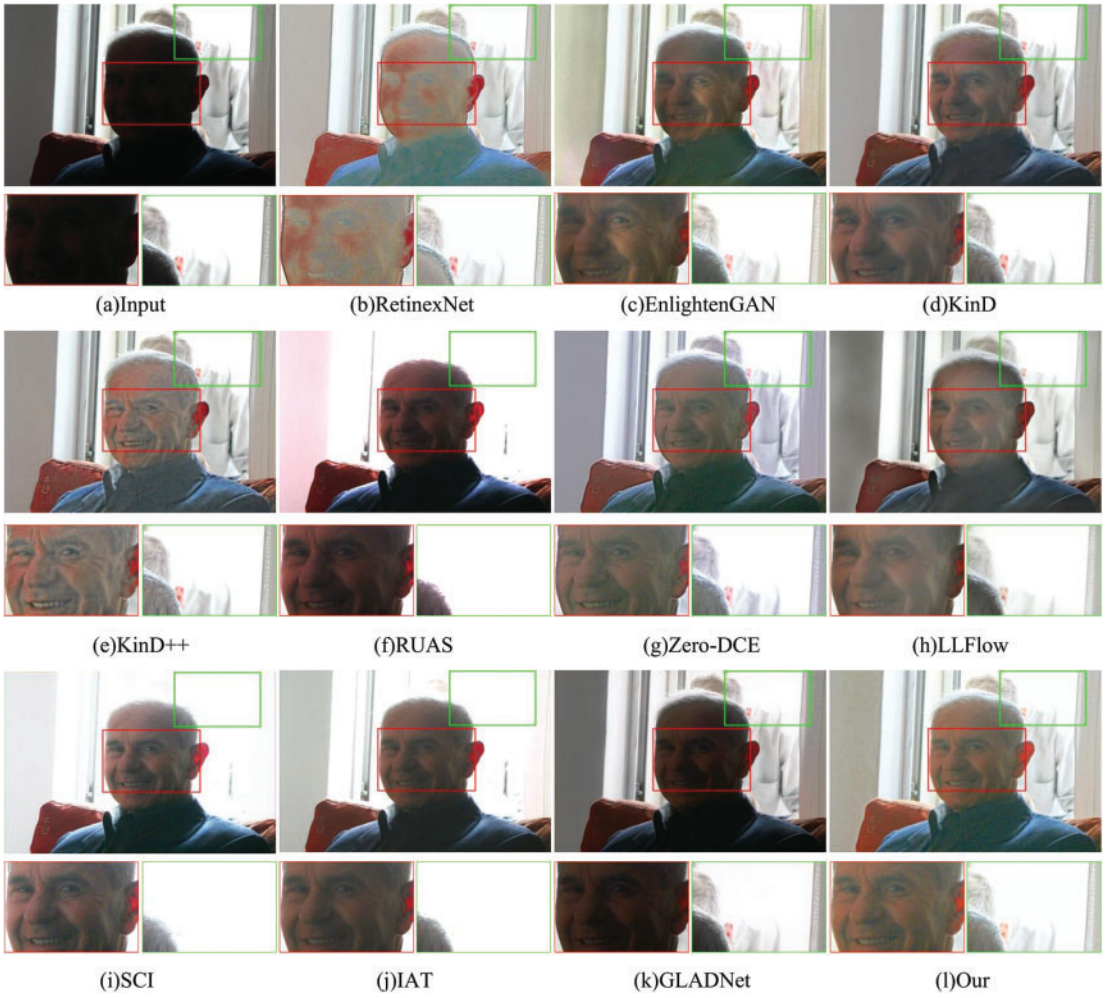

Figure 3: Qualitative comparison of RF-Net with other advanced algorithms. See the patch area for more detailed information

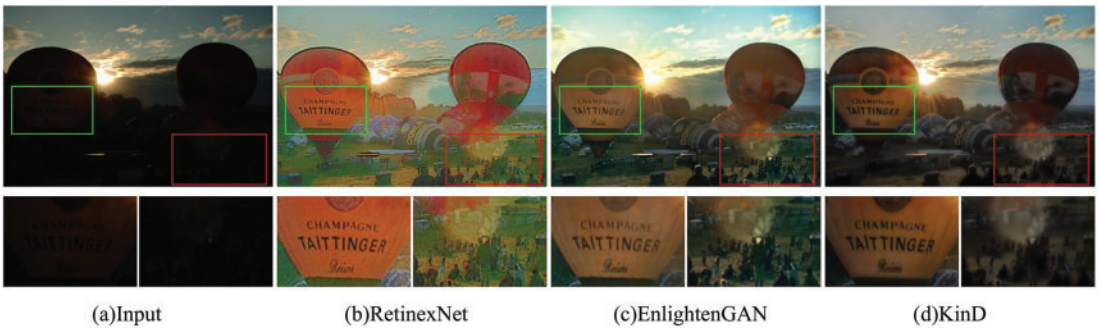

Figure 4: (Continued)

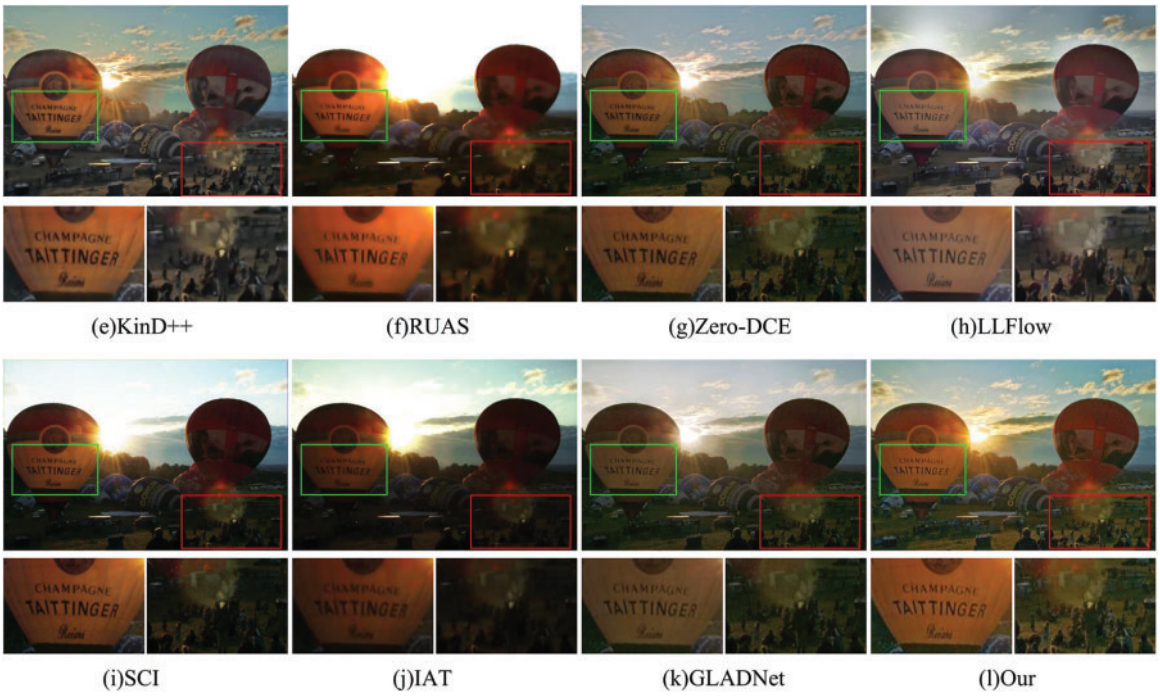

Figure 4: Qualitative comparison of RF-Net with other advanced algorithms. See the patch area for more detailed information

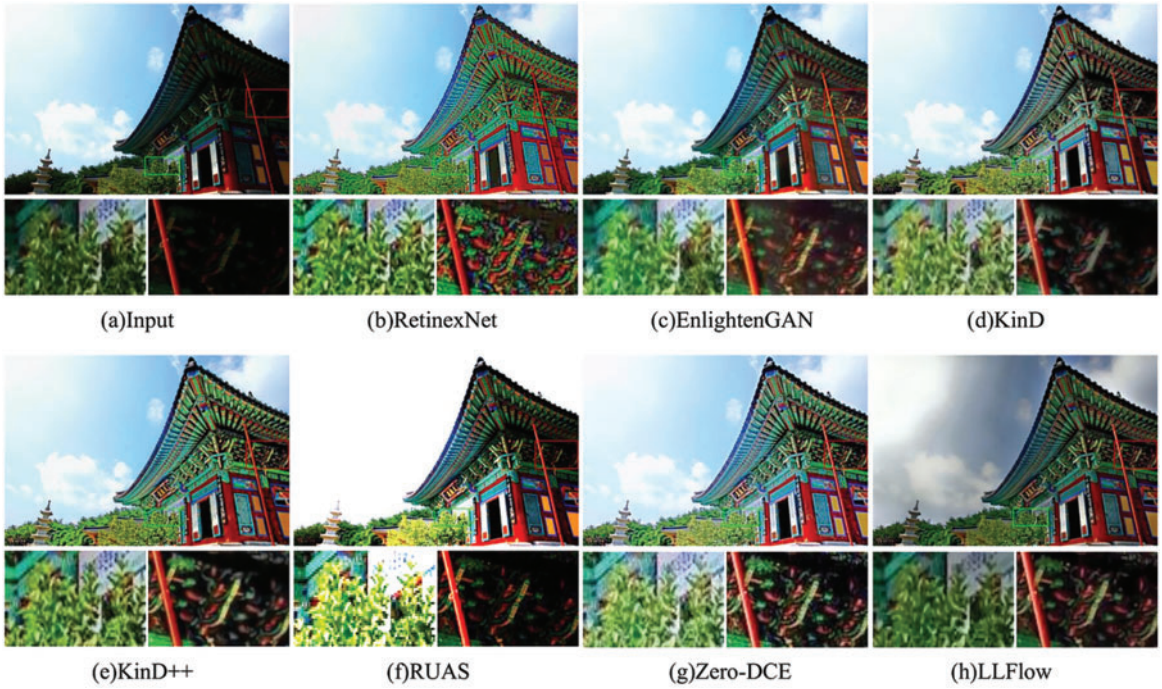

Figure 5: (Continued)

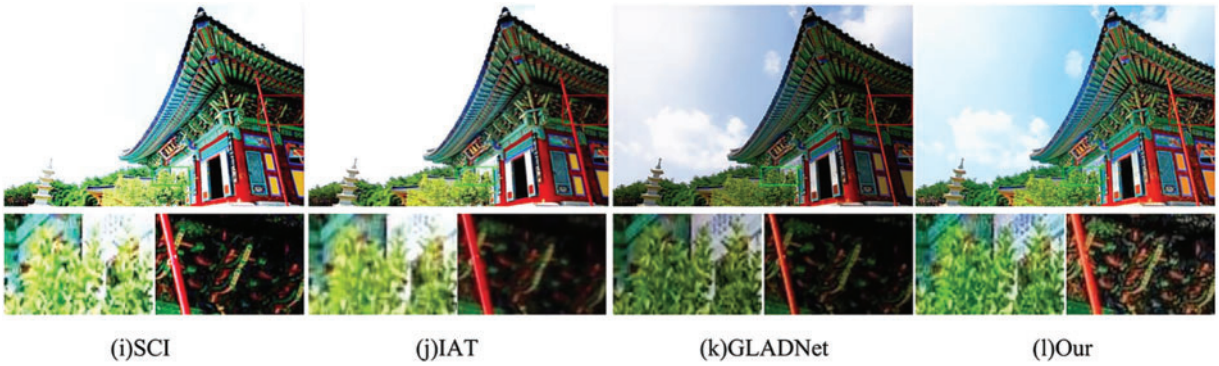

Figure 5: Qualitative comparison of RF-Net with other advanced algorithms. See the patch area for more detailed information

4.2.2 Quantitative Comparison

The subjective evaluation may not be sufficient for determining the degree of detail retention during image enhancement. To demonstrate the feasibility of the proposed method further, we conducted quantitative comparisons. As we used unsupervised methods for model training, we could not evaluate the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) of the enhanced images to the ground truth, as with other supervised methods. Therefore, we used a non-referenced image quality assessment metric to compare the RF-Net method with other competitors. The metrics evaluated were Natural Image Quality Evaluator (NIQE) [37] and Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [38]. NIQE is a natural-image-based evaluation metric that compares algorithm processing results with a model calculated based on natural scenes. BRISQUE is an image-based no-reference quality score that is calculated based on natural scene images with similar distortions. The metrics used to evaluate the performance of RF-Net compared with the other algorithms are listed in Table 1. Based on the qualitative evaluation, the following issues were observed: RetinexNet [11] resulted in inaccurate color restoration with color bias; KinD [12] did not significantly enhance dark areas; KinD++ [13] introduced artifacts while enhancing dark areas; RUAS [28] over-enhanced the image, causing loss of information; LLFlow [15] produced incorrect color restoration; SCI [16] and IAT [20] over-enhanced the image and had insignificant enhancement in dark areas; Zero-DCE [5] had lower metrics compared to other unsupervised methods, possibly due to the whitish image produced by its enhancement results, as noted in the qualitative analysis. As shown in the graphs, the enhancement results of RF-Net validated this. In summary, because both NIQE and BRISQUE are methods based on local image statistical information, the impact of different algorithms on the small differences in the achieved metrics after image enhancement is better illustrated in the quantitative evaluation.

4.2.3 Subjective Evaluation of People

We also conducted a human subjective visual evaluation to further quantify the subjective quality of RF-Net compared to the other methods. We randomly selected 20 original low-light images from the test set (NPE [34], DICM [35], LIME [10], MEF [36], and VV) and applied four state-of-the-art methods (EnlightenGAN [4], Zero-DCE [5], SCI [16], and IAT [20]) to each image separately. We invited twelve reviewers to independently score the results of the five algorithms, including RF-Net. The reviewers primarily observed the following aspects: 1. Whether the results contained artifacts in

the over- or under-enhanced areas; 2. Whether the color restoration in the results was accurate (e.g., whether the colors were distorted); and 3. Whether the noise in dark areas was amplified, and whether there was an obvious loss of texture details in the results. As can be observed from the statistics in Fig. 6, RF-Net achieved a higher subjective evaluation score for the reviewed images.

Table 1: The NIQE (↓) and BRISQUE (↓) scores are shown, with lower scores indicating better image quality and richer information contained. The averages of the test image metrics are taken for each of the five datasets, and the five averages are eventually averaged again. The best result is shown in red and the second-ranked result is shown in blue

Method	DICM	LIME	MEF	NPE	VV	Average
RetinexNet	4.31/26.50	4.83/24.33	4.90/26.04	4.14/29.23	2.90/23.38	4.21/25.89
EnGAN	2.79/23.71	3.49/23.03	3.01/25.69	3.00/24.88	2.25/23.04	2.90/24.01
KinD	3.45/30.60	2.91/26.72	3.37/30.43	3.13/23.73	2.67/27.79	3.10/27.85
KinD++	3.42/28.10	3.55/27.49	3.48/30.03	3.36/25.42	2.74/24.89	3.31/27.19
RUAS	4.56/36.63	4.26/30.76	4.08/33.92	4.42/36.32	4.29/31.78	4.32/33.88
DCE-Net	3.26/31.40	4.10/23.60	3.31/25.48	3.22/24.58	2.66/18.97	3.31/24.86
LLFlow	3.10/27.91	4.04/26.68	3.52/27.07	3.29/27.84	2.80/24.98	3.35/26.88
SCI	3.70/32.58	4.26/22.74	3.43/26.48	3.42/30.38	3.01/21.36	3.56/26.71
IAT	3.29/35.91	3.47/35.14	3.04/33.83	3.22/32.59	2.86/35.68	3.17/34.63
GLADNet	3.28/28.23	3.01/23.88	3.17/22.91	2.95/25.63	2.51/23.40	2.98/24.81
Our	2.98/27.88	2.91/20.72	2.69/27.37	3.00/26.49	2.23/21.08	2.77/24.68

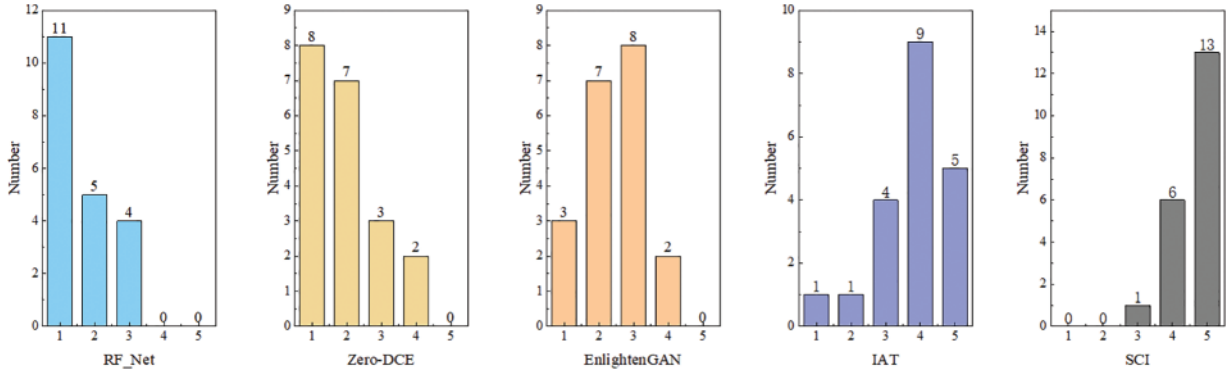

Figure 6: Overview of people's objective evaluation, in each graph, the x-axis indicates the observed quality of the five algorithms (1 for the best and 5 for the worst) and the y-axis indicates the number of good and bad images corresponding to m each algorithm. RF-Net shows the best performance

4.3 Ablation Study

To demonstrate the effectiveness of the modules used, the following ablation studies were conducted separately: three studies were designed in terms of CA removal, separation of inception [39] from residual connectivity [40], and direct fusion of illuminated and reflected components.

In Ablation Study 1, we experimented with the multiscale module, residual connection, and CA attention separately, as shown in Fig. 7, we can observe from the visual results that zooming in on the person above the house in the first image reveals that using only inception and the Markov discriminator leads to a blurred task, whereas using both inception and residual linking with a Markov discriminator produces more vivid colors and preserves more detailed information owing to residual linking. Furthermore, using inception, residual linking, and an improved Markov discriminator leads to a more realistic restoration of texture details, because the Markov discriminator has a clearer ability to distinguish true from false. The reddish hue in our image is due to the reflected image generated by the first stage of the network, which does not include knowledge of the illumination image or fusion module.

Figure 7: Ablation study for each module. (a) Input, (b) w/o Multiscale, (c) w/o Residual, (d) w/o CA

In Ablation Study 2, we compared our method with the direct integration of light and reflection components, as shown in Fig. 8. The color of the sky in the second column of the first row was inaccurately restored, and the face of the person in the second column of the second row was overexposed. The proposed method in the third column performs significantly better than the direct fusion method.

In Ablation Study 3, we tested the designed exposure generator module by adjusting the value of k to obtain images with different exposure levels and achieved exposure fusion. As shown in Fig. 9, for different input images, we acquired images with varying exposure levels for fusion, with the k values set differently in different scenarios. For example, in images without exposed areas, k values are

usually set between 4–6, while in images with exposed areas, k values are typically set between 2–4. This approach achieved the best enhancement when the exposed images were fused with the original low-light images. Finally, we conducted a quantitative evaluation of the three sets of ablation studies. The results in Table 2 show that adding the residual block to the inception and residual block combinations produced superior performance. In addition, exposure fusion in the second stage of RF-Net exhibited advantages. A comparison of the second and fourth experimental rows demonstrates the effectiveness of RF-Net.

Figure 8: Direct fusion of illuminated and reflected components with RF-Net ablation study

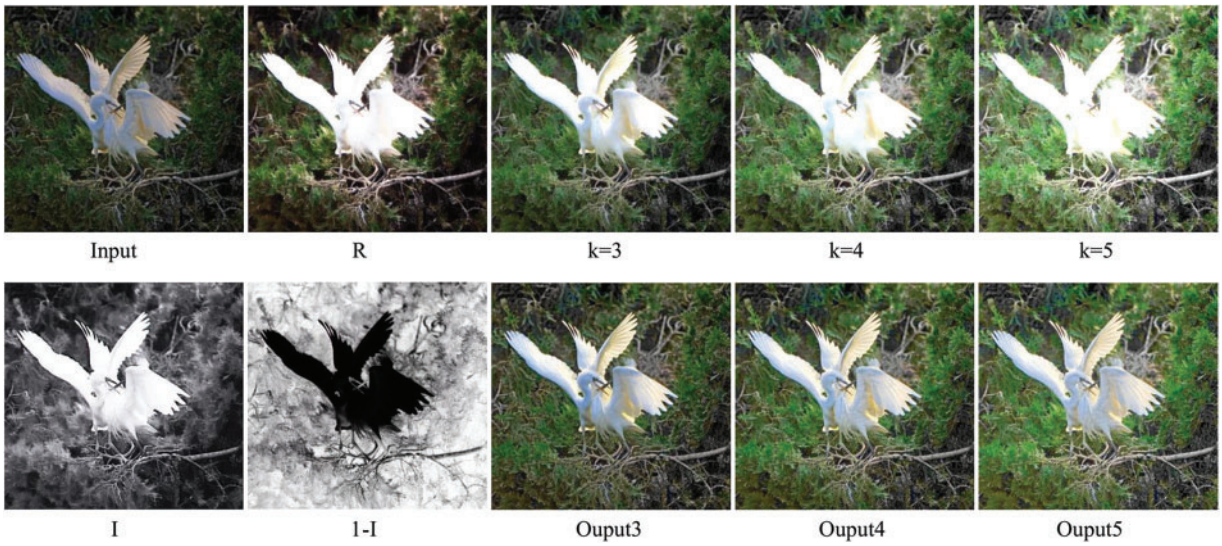

Figure 9: Image decomposition and exposure generation fusion ablation study

Table 2: Quantitative review of ablation studies for each module

Inception	Res	CA	I*R	Ex-Fusion	NIQE	BRISQUE
√	×	√	√	×	3.21	27.68
√	√	√	√	×	3.03	27.50
√	×	√	×	√	2.97	26.75
√	√	√	×	√	**2.77**	**24.68**

In Ablation Study 4, we use the low-light image and its corresponding Red, Green, Blue (RGB) channels with the maximum, minimum, and average values as inputs to our network. The final network input consists of six channels, with the first three channels used to generate the reflection map and the last three channels used to generate the illumination map. In the generated illumination images, the method used in this paper retains more details in the final generated illumination map compared with the input of only the original low illumination image. The qualitative comparison in Fig. 10 shows that in the first column, there are artifacts near the shoulders of the person in the illumination map generated by the network with only low illumination input. In the second column, there is a noticeable loss of detail in the wall of the house on the right. In the third column, there are artifacts near the candle flame, and the grayscale boundary near the cup in the upper left corner is not clear; In the fourth column, there is an excessive amount of detail in the cave, while the detail in the trees outside the cave is lacking or insufficient. In Fig. 11, the grayscale histogram results show that our method can generate a light map with smoother contrast and luminance changes, making the light and dark parts more visible and details more prominent.

Figure 10: Qualitative comparison of illuminated images. (a) Input, (b) Low-light images as input, (c) Ours

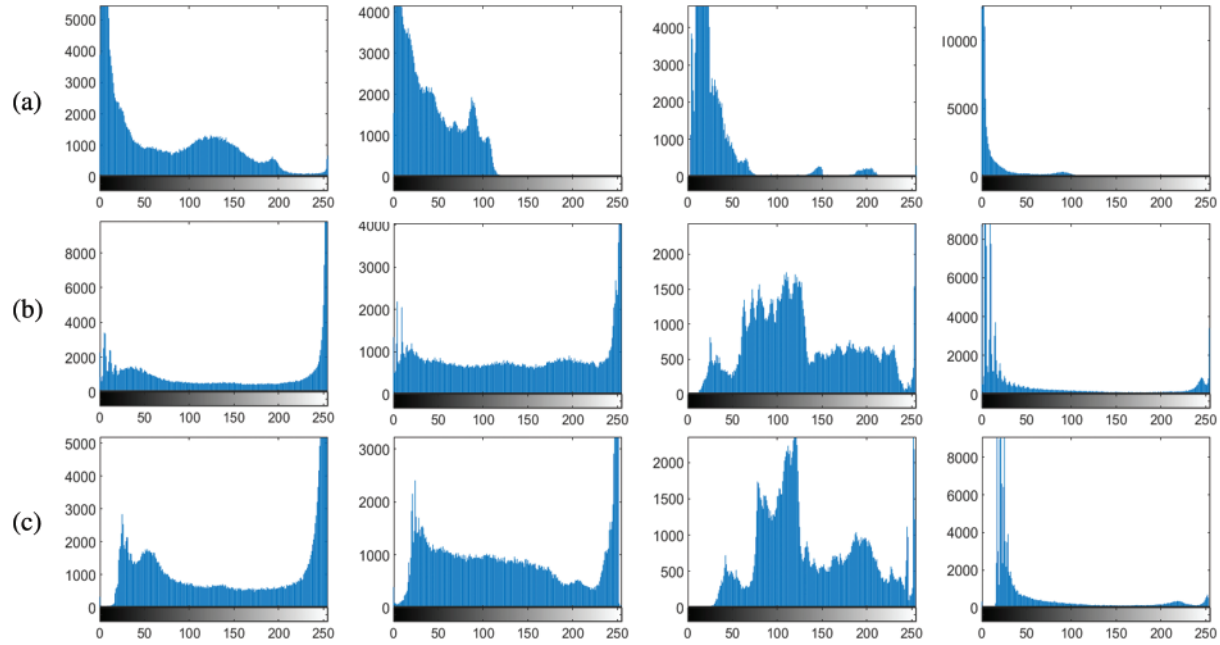

Figure 11: Grayscale histogram. (a) Low-light image grayscale histogram, (b) Grayscale histogram after low-illumination image as network input, (c) Ours

5 Conclusion

In this study, we combined Retinex theory with exposure fusion for the first time to achieve unpaired low-light image enhancement. In the first stage, we designed a multi-scale generator by combining a residual network and inception. We also added a CA attention mechanism with position information to the discriminator network to obtain high-quality illumination and reflection components. By improving the original Retinex and camera response mechanism functions, we designed an exposure image generator with correction coefficients to solve the problems of illumination, reflection image fusion, and exposure image generation. Based on this, we realized low-light image enhancement with second-stage exposure fusion and proved the superiority of our method by comparing it with state-of-the-art methods. However, the algorithm has some limitations. On the one hand, it requires manual adjustment of the correction parameters for the exposure image generator based on the scene's exposure level. On the other hand, Compared to other lightweight networks, RF-Net processes 640×400 images at a rate of 5 frames per second. In the future, our research will focus on lightweight the network structure and developing an adaptive low-light image enhancement method based on negative feedback control to solve the manual tuning problem of existing methods. We also aim to apply this model to enhance specific scenes, which will not only improve the generalization of the algorithm but also enhance the accuracy of other vision tasks.

Acknowledgement: The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

Funding Statement: This work was supported by the National Key Research and Development Program Topics (Grant No. 2021YFB4000905), the National Natural Science Foundation of China

(Grant Nos. 62101432 and 62102309), and in part by Shaanxi Natural Science Fundamental Research Program Project (No. 2022JM-508).

Author Contributions: Study conception and design: Tian Ma, Jiayi Yang; data collection: Chenhui Fu; analysis and interpretation of results: Chenhui Fu, Jiehui Zhang, Chuyang Shang; draft manuscript preparation: Chenhui Fu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this paper can be requested from the corresponding author upon request.

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

References

[1] D. J. Jobson, Z. Rahman and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 451–462, 1997.
[2] Z. Rahman, D. J. Jobson and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *3rd IEEE Image Processing Conf.*, Lausanne, Switzerland, vol. 3, pp. 1003–1006, 1996.
[3] D. J. Jobson, Z. Rahman and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.
[4] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang *et al.,* "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Transactions on Image Processing*, vol. 30, pp. 2340–2349, 2021.
[5] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou *et al.,* "Zero-reference deep curve estimation for low-light image enhancement," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 1780–1789, 2020.
[6] Y. T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1–8, 1997.
[7] Y. Wang, Q. Chen and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 1, pp. 68–75, 1999.
[8] C. Liu, X. Sui, Y. Liu, X. Kuang, G. Gu *et al.,* "Adaptive contrast enhancement based on histogram modification framework," *Journal of Modern Optics*, vol. 66, no. 15, pp. 1590–1601, 2019.
[9] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108–129, 1977.
[10] X. Guo, L. Yu and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2016.
[11] C. Wei, W. Wang, W. Yang and J. Liu, "Deep retinex decomposition for low-light enhancement," arXiv preprint arXiv:1808.04560, 2018.
[12] Y. Zhang, J. Zhang and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. of the ACM Conf. on Multimedia*, Ottawa, ON, Canada, pp. 1632–1640, 2019.
[13] Y. Zhang, X. Guo, J. Ma, W. Liu and J. Zhang, "Beyond brightening low-light images," *International Journal of Computer Vision*, vol. 129, no. 4, pp. 1013–1037, 2021.
[14] W. Wang, C. Wei, W. Yang and J. Liu, "Gladnet: Low-light enhancement network with global awareness," in *2018 Automatic Face & Gesture Recognition Conf. (FG 2018)*, Xi'an, China, pp. 751–755, 2018.
[15] Y. Wang, R. Wan, W. Yang, H. Li, L. P. Chau *et al.,* "Low-light image enhancement with normalizing flow," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 36, no. 3, pp. 2604–2612, 2022.
[16] L. Ma, T. Ma, R. Liu, X. Fan and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 5637–5646, 2022.

[17] T. Ma, C. Fu, M. Guo, J. Yang and J. Liu, "Dual attention unit-based generative adversarial networks for low-light image enhancement," in *2022 IEEE Signal Processing, Communications and Computing Conf. (ICSPCC)*, Xi'an, China, pp. 1–5, 2022.
[18] T. Ma, J. An, R. Xi, J. Yang, J. Lyu *et al.,* "TPE: Lightweight transformer photo enhancement based on curve adjustment," *IEEE Access*, vol. 10, pp. 74425–74435, 2022.
[19] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang *et al.,* "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 5901–5910, 2022.
[20] Z. Cui, K. Li, L. Gu, S. Su, P. Gao *et al.,* "Illumination Adaptive Transformer," arXiv preprint arXiv:2205.14871, 2022.
[21] 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. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 3063–3072, 2020.
[22] W. Yang, S. Wang, Y. Fang, Y. Wang and J. Liu, "Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality," *IEEE Transactions on Image Processing*, vol. 30, pp. 3461–3473, 2021.
[23] L. Zhang, L. Zhang, X. Liu, Y. Shen, S. Zhang *et al.,* "Zero-shot restoration of back-lit images using deep internal learning," in *Proc. of the ACM Conf. on Multimedia*, Ottawa, ON, Canada, pp. 1623–1631, 2019.
[24] Y. Fu, Y. Hong, L. Chen and S. You, "LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss," *Knowledge-Based Systems*, vol. 240, no. 6, pp. 108010, 2022.
[25] C. Li, C. Guo and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4225–4238, 2021.
[26] T. Ma, M. Guo, Z. Yu, Y. Chen, X. Ren *et al.,* "RetinexGAN: Unsupervised low-light enhancement with two-layer convolutional decomposition networks," *IEEE Access*, vol. 9, pp. 56539–56550, 2021.
[27] A. Zhu, L. Zhang, Y. Shen, Y. Ma, S. Zhao *et al.,* "Zero-shot restoration of underexposed images via robust retinex decomposition," in *Proc. of the ICME Conf. on Multimedia and Expo*, London, UK, pp. 1–6, 2020.
[28] 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. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 10561–10570, 2021.
[29] Z. Zhao, B. Xiong, L. Wang, Q. Qu, L. Yu *et al.,* "RetinexDIP: A unified deep framework for low-light image enhancement," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1076–1088, 2021.
[30] Q. Hou, D. Zhou and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 13713–13722, 2021.
[31] X. Deng, Y. Zhang, M. Xu, S. Gu and Y. Duan, "Deep coupled feedback network for joint exposure fusion and image super-resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 3098–3112, 2021.
[32] Z. Ying, G. Li and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," arXiv preprint arXiv:1711. 00591, 2017.
[33] X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang *et al.,* "Least squares generative adversarial networks," in *Proc. of the ICCV Conf. on Computer Vision*, Cambridge, MA, USA, pp. 2794–2802, 2017.
[34] S. Wang, J. Zheng, H. M. Hu and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3538–3548, 2013.
[35] C. Lee, C. Lee and C. S. Kim, "Contrast enhancement based on layered difference representation," in *Proc. of the ICIP Conf. on Image Processing*, Orlando, FL, USA, pp. 965–968, 2012.
[36] K. Ma, K. Zeng and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3345–3356, 2015.
[37] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a "completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.

[38] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed *et al.,* "Going deeper with convolutions," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 1–9, 2015.
[40] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, Massachusetts, USA, pp. 770–778, 2016.