Low-Light Image Enhancement Algorithm Based on the Fusion of Multi-scale Features and Attention Mechanism

Youchen Sun[1,2], Baoju Zhang[1,2(✉)], Bo Zhang[1,2], Cuiping Zhang[1,2], and Jin Zhang[1,2]

[1] Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
wdxyzbj@163.com

[2] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China

Abstract. Aiming at the problems of Retinex-Net such as large noise of reflection component, low brightness of illumination component and insufficient feature extraction, a low-light image enhancement algorithm based on fusion of multi-scale features and attention mechanism is proposed. First, the Retinex-Net network is used as the basic model to decompose the input image, and the atrous convolution and ordinary convolution are fused to achieve multi-scale feature extraction to obtain more detailed information; multi-layer attention is introduced into the enhanced network. The force mechanism module enhances the brightness of the details and illumination components; finally, the denoised reflection components and the enhanced illumination components are fused into a normal illumination image output. Experiments show that the algorithm can effectively improve the details of the image and improve the visual effect of the image.

Keywords: Low-light image enhancement · Retinex-Net · multiscale feature extraction · attention mechanism · convolutional neural network

1 Introduction

In the information age, high-quality images are critical for many computer vision tasks. In reality, good lighting helps to get high-quality images. Conversely, low light results in less useful information being obtained from the image. During the shooting process, due to ambient light intensity and technical limitations, the images obtained by imaging devices such as cameras have problems such as low quality, low contrast, and brightness, so the loss of important image information is not conducive to subsequent feature extraction [1]. To effectively obtain high-quality images from low-light environments, scholars have proposed many low-light image enhancement algorithms. The current research methods for low-light image enhancement can be roughly divided into two categories: traditional methods and deep learning.

This work was supported in part by 2021 Tianjin Postgraduate Research and Innovation Project 2021YJSS209.

Q. Liang et al. (Eds.): CSPS 2022, LNEE 874, pp. 65–72, 2023.
https://doi.org/10.1007/978-981-99-2362-5_9

Compared with traditional methods, deep learning-based methods have better accuracy, robustness, and speed. Therefore, in recent years, researchers have also proposed many methods based on deep learning. Based on the theory of the Retinex algorithm, researchers will combine the Retinex algorithm with convolutional neural networks to improve the visual effect of images, automatically learn the features of images, and solve the problem that Retinex relies on manually setting parameters [2]. Therefore, deep Retinex-based methods have better performance in most cases [3].

Based on the network framework of Retinex-Net, this paper proposes a low-light image enhancement algorithm based on the fusion of multi-scale features and attention mechanisms. Through the fusion of hole convolution and ordinary convolution in the decomposition network, multi-scale feature extraction is realized, and more detailed information of the attention map and reflection map is obtained; the multi-layer attention mechanism module is introduced into the enhancement network to measure the brightness of the illumination component. For enhancement, the improved network improves the details of the image, enhances the visual effect of the image, and improves the overall quality of the image.

2 Related Work

2.1 Retinex-Net Network Improvement

According to the problems of large reflection component noise, low illumination component brightness, and blurred image in the Retinex-Net network, this paper improves based on the Retinex-Net network and proposes a method based on the fusion of multi-scale features and attention mechanism. Low-light image enhancement algorithm. The algorithm mainly includes two sub-networks trained independently of each other, namely the decomposition network and the enhancement network. In the decomposition network, atrous convolution is used to improve the receptive field of the network without increasing the model parameters, and to obtain illumination and reflection components with more detailed information. Secondly, multiple convolutional block attention models (CBAM) are introduced into the enhancement network to enhance the details of the illumination components and reflection components, and guide the network to correct the illumination components; the reflection map is denoised by the Block Matching 3D(BM3D) algorithm. Finally, image reconstruction is performed, and the enhanced image is obtained by multiplying the processed illumination component and the denoised reflection component. The overall network structure of this paper is shown in Fig. 1. Next, the main parts of the model will be introduced in detail.

2.2 Decomposing the Network

In the decomposition network (Dceom-Net), the input low-light S_{low} and normal-light images S_{normal} are decomposed into reflection components R_{low} and R_{normal} and illumination components I_{low} and I_{normal}. In order to obtain more feature information, the decomposition network combines the atrous convolution with the ordinary convolution, and the atrous convolution can expand the characteristics of the receptive field without increasing the number of effective units of the convolution kernel. By combining

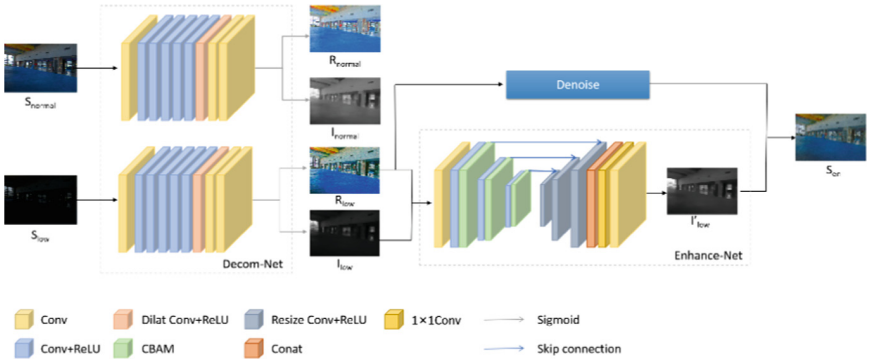

Fig. 1. Overall network structure

ordinary convolutions, feature extraction at different scales can be achieved. Therefore, while atrous convolution is used to perform sparse feature extraction on image features, ordinary convolution with the same size of receptive field is used to densely extract the features of the input image, and finally the features output by two different convolution methods are processed in the channel dimension. Stitching.The decomposition network module is shown in Fig. 1.

The loss function LDecom of the decomposition network consists of three parts: reconstruction loss Lrecon, reflection consistency loss Lir and illumination smoothing loss Lis:

$$L_{Decom} = L_{recon} + \lambda_{ir} L_{ir} + \lambda_{is} L_{is} \tag{1}$$

where λ_{ir} and λ_{is} are the coefficients for reflection consistency and lighting smoothness, respectively. When L1, L2, and SSIM losses are chosen, the L2 norm does not correlate well with the perception of human visual image quality and tends to be locally minimized during training. Although SSIM can better understand the structural features of the image, it is less sensitive to errors in smooth regions, resulting in chromatic aberration [4]. Therefore, this paper chooses to use the L1 norm as a constraint to constrain the loss.

Since each of the decomposition networks R_{low} and R_{normal} can be reconstructed with their respective light maps, the specific form of the reconstruction loss is:

$$L_{recon} = \|R_{low} \cdot I_{low} - S_{low}\|_1 + \|R_{normal} \cdot I_{normal} - S_{normal}\|_1 \tag{2}$$

The reflection consistency loss is used to make the reflection components of the low-illuminance image and the normal-illumination image as consistent as possible. The specific performance of the reflection consistency loss is::

$$L_{ir} = \|R_{low} - R_{noemal}\|_1 \tag{3}$$

To ensure that the illumination component keeps the overall smoothness and the structure and local details of the image, the structure-aware smoothing loss [5] is used as the illumination smoothing loss Lis, which is specifically expressed as:

$$L_{is} = \left\|\nabla I_{low} \cdot \exp\left(-\lambda_g \nabla R_{low}\right)\right\|_1 + \left\|\nabla I_{noemal} \cdot \exp\left(-\lambda_g \nabla R_{normal}\right)\right\|_1 \tag{4}$$

Among them, ∇ represents the gradient in the horizontal and vertical directions of the image, and λ_g represents the coefficient of balancing the perceived strength of the structure. The weights $\left(-\lambda_g \nabla R\right)$ allow Lis to relax the smoothness constraints when the gradient of the reflection component is large, even if the smoother areas in the reflection component match the areas corresponding to the lighting component, ensuring that texture details and boundary information in the smoothness constraints are not destroyed.

2.3 Enhanced Network

The enhancement network uses the overall framework of the encoder-decoder structure, combined with the U-Net network structure, to enhance the illumination component Ilow obtained by the decomposition network, and perform BM3D noise suppression processing for the reflection component R_{low}. The specific structure is shown in Fig. 1.

Aiming at the problems of color distortion, low brightness, and insufficient details in the illumination images obtained by the enhancement network, an attention mechanism Convolutional Block Attention Module(CBAM) is introduced to enhance and correct the details of the illumination images.The structure diagram of CBAM is shown in Fig. 2.

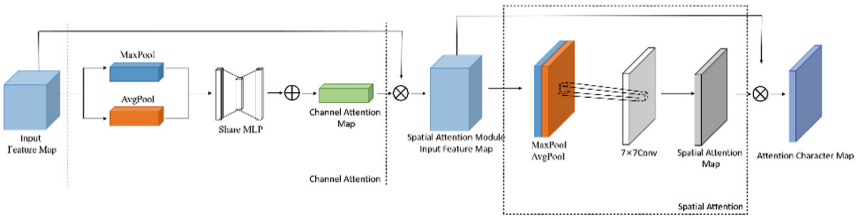

Fig. 2. CBAM structure diagram

The enhancement network first performs a connection operation on the illumination components and reflection components obtained in the decomposition stage and then performs a 3 × 3 convolution operation, performs channel selection on the input low-illuminance map illumination component Ilow, and passes the input feature map through two parallel The global maximum connection layer and the intermediate layer are transmitted, and then through the shared network (Share MLP) module, through the activation function ReLU to obtain two activated results. The two output results are added together, and then a sigmoid activation function is used to obtain the output result of the channel module, and then the output result is multiplied by the original image to obtain the output result of the channel module. The output of the channel module is obtained by max pooling and average pooling to obtain two feature maps, splicing the two feature maps, and convolution through a standard convolution layer to obtain a spatial attention map. After another sigmoid, the final attention feature map is obtained. After passing through the attention module, the image is enhanced by the nearest neighbor interpolation method, and up-sampling is performed to ensure that the size of the combined feature map is consistent, and the corresponding sums are added, and then the feature fusion is performed to obtain a feature map with more complete details, and

finally, the network is fine-tuned end-to-end using stochastic gradient descent to obtain augmented images [6].

The loss Len of the augmentation network includes the reconstruction loss L_{recon} and the illumination smoothing loss L_{is}. The reconstruction loss is expressed as the multiplication of the denoised reflection component and the corrected illumination component to reconstruct the image to obtain an enhanced image.

The reconstruction loss in the augmented network is $L_{recon} = \left\| \hat{R}_{low} \cdot \hat{I}_{low} - S_{normal} \right\|_1$. $\hat{R}_{low}$ is the reflection component after denoising, $\hat{I}_{low}$ is the illumination component weighted by the R_{low} gradient map, and is the normal illumination image. The calculation of the illumination smoothing loss L_{is} is the same as that of Eq. (4).

While enhancing the illumination image, denoise the decomposed reflection image, and use the BM3D algorithm to suppress the amplified noise in R_{low}. Finally, the enhanced illumination image and the denoised reflection image are multiplied element by element to achieve image reconstruction and form the final enhanced image.

3 Experimental Design and Results

3.1 Dataset

The dataset in this paper is divided into two parts: the real scene dataset LOL and the artificial synthetic dataset Brighting Train [5]. The LOL dataset is a dataset of image pairs collected from real scenes for low-light enhancement. The dataset captures images from multiple real scenes, including 500 pairs of images of the furniture, streets, buildings, etc., of which 485 pairs of images are used as a training set and 15 pairs of images are used as a test set. In RAISE [7], 1000 original images were used to synthesize low-light images, forming 1000 pairs of artificially synthesized datasets as the training set, and named Brighting Train. To make the experimental results more accurate, the training samples are rotated, translated, cropped, and other operations to perform data enhancement preprocessing on the images of the data set, to obtain more data image sets.

3.2 Experimental Environment and Parameter Settings

The training and testing experiments of the overall network in this paper are completed based on the PyTorch framework on the NVIDIA GeForce 1080Ti GPU device, using CUDA11.6 to accelerate it, and using the Adam optimizer to optimize the loss function. The loss function coefficients λ_{ir} and λ_{is} are set to 0.001 and 0.1, respectively. The balanced structure perception strength factor λ_g is 10. During network training, the batch size batch_size is set to 16 and the image patch size is set to 48×48. The training batch epoch is set to 100. The initial value of the learning rate is 0.001, and the number of iterations is selected to perform a learning rate decay every 20 iterations, that is, a decay of 5 times every 20 epochs.

3.3 Comparative Experiment

To verify the performance and effectiveness of the algorithm in this paper, a comparison experiment was conducted with other low-light image enhancement methods, including the EnlightenGAN algorithm [8], the MBLLEN algorithm [9], the Retinex-Net algorithm [5], and the RRDNet algorithm [10], Zero-DCE algorithm [11], 4 low-light images were selected in the LOL dataset and compared with five algorithms to reflect the superiority of the improved algorithm in this paper. The comparison results are shown in Fig. 3.

Fig. 3. Experimental comparison of each algorithm in the LOL dataset

As can be seen from Fig. 3, in the EnlightenGAN algorithm, through the fourth low-light image, it can be found that there are obvious shadows in the image, such as the shadow at the intersection of the book on the desk and the desk is more obvious. The MBLLEN algorithm is bright as a whole, which makes the picture overexposed. For example, the brightness of the stainless steel kitchen utensils in the second picture is high, which is more obvious. The Retinex-Net algorithm produces a blurring phenomenon in the comparison image, the image is noisy, the color is slightly distorted, the edges of some objects are too prominent, and artifacts appear in some areas. The RRDNet algorithm has insufficient brightness enhancement, and the overall image is dark, resulting in the phenomenon of underexposure and serious problems in image denoising, and image sharpening is serious. In the Zero-DCE algorithm, the brightness of the enhanced image is slightly lower and the color saturation is weaker. The overall brightness of the image enhanced by the algorithm in this paper is more in line with the visual perception of the human eye be improved. Therefore, from a subjective point of view, the experimental results of our algorithm are better than other algorithms.

Since people have different visual and sensory preferences, comparisons using only visual quality are one-sided. Therefore, this paper adopts two evaluation indicators, PSNR and SSIM, to objectively evaluate the image quality. The experimental results are shown in Table 1. As can be seen from Table 1, the algorithm in this paper has achieved good results in the PSNR and SSIM scores, with the highest PSNR value and the second highest SSIM value. The PSNR value of the algorithm in this paper is 0.72dB higher than that of the MBLLEN algorithm with the second highest PSNR value, and the SSIM value is only 0.036 lower than the highest value. In summary, the results of the algorithm in this

paper have been significantly improved, which proves the effectiveness of the algorithm in low-light image enhancement, and the improved network can effectively improve the processing capability of low-light images.

Table 1. Average objective evaluation of different algorithms on the LOL dataset

Algorithm	PSNR↑	SSIM↑
EnlightenGAN [8]	17.483	0.677
MBLLEN [9]	17.902	**0.715**
Retinex-Net [5]	16.774	0.462
RRDNet [10]	11.392	0.468
Zero-DCE [11]	16.796	0.589
Ours	**18.622**	0.679

4 Conclusion

Aiming at the problems of Retinex-Net such as large noise of reflection component, low brightness of illumination component and insufficient feature extraction, a low-light image enhancement algorithm based on fusion of multi-scale features and attention mechanism is proposed. The algorithm in this paper combines atrous convolution with ordinary convolution in the decomposition network to achieve multi-scale feature extraction, so as to obtain illumination components and reflection components with more detailed information. The attention mechanism CBAM module is introduced into the enhancement network to enhance and correct the brightness of the illumination components. Finally, the denoised reflection component and the enhanced illumination component are constructed into a normal illumination image. On the LOL data set, the algorithm in this paper is compared with some other advanced algorithms from both subjective and objective aspects. The experimental results show that the algorithm in this paper improves the brightness of the image, the sharpness and texture details are significantly improved, and finally the quality of the image is improved.

References

1. Li, C., Guo, C., Han, L.H., et al.: Low-light image and video enhancement using deep learning: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **01**, 1 (2021)
2. Ou, J.M., Hu, X., Yang, J.X.: Low-Light Image enhancement algorithm based on improved retinex-net. Pattern Recogn. Artif. Intell. **34**(01), 77–86 (2021)
3. Wang, M.M., Peng, D.L.: Retinex-ADNet: a low-light image enhancement system. J. Chin. Comput. Syst. **43**(02), 367–371 (2022)
4. Zhao, H., Gallo, O., et al.: Loss functions for image restoration with neural networks. IEEE Trans. Comput. Imaging **3**, 27–47 (2017)

5. Wei, C., Wang, W., Yang, W., et al.: Deep retinex decomposition for low-light enhancement (2018). arXiv preprint arXiv:1808.04560
6. Liu, J.M., He, N., Yin, X.J.: Low lllumination image enhancement based on retinex-UNet algorithm. Comput. Eng. Appl. **56**(22), 211–216 (2020)
7. Dang-Nguyen, D.T., Pasquini, C., Conotter, V., et al.: RAISE: a raw images dataset for digital image forensics. In: Proceedings of the 6th ACM Multimedia Systems Conference, pp. 219–224 (2015)
8. Jiang, Y., et al.: Enlightengan: deep light enhancement without paired supervision (2019). arXiv preprint arXiv:1906.06972
9. Lv, F., Lu, F., Wu, J., Lim, C.: Mbllen: low-light image/video enhancement using cnns. In: BMVC (2018)
10. Zhu, A., Zhang, L., Shen, Y., Ma, Y., Zhao, S., Zhou, Y.: Zero-shot restoration of underexposed images via robust retinex decomposition. In: ICME, pp. 1–6 (2020)
11. Guo, C., et al.: Zeroreference deep curve estimation for low-light image enhancement. In: CVPR, pp. 1780–1789 (2020)