A lightweight low-light image enhancement network

Yutao Jin
College of Electronic Information and Automation, Tianjin University of Science and Technology, No.1038 Dagu Nanlu, Hexi District, Tianjin, China, 300222

Xiaoyan Chen
College of Electronic Information and Automation, Tianjin University of Science and Technology, No.1038 Dagu Nanlu, Hexi District, Tianjin, China, 300222

Xiwen Liang
College of Electronic Information and Automation, Tianjin University of Science and Technology, No.1038 Dagu Nanlu, Hexi District, Tianjin, China, 300222
E-mail: Charlesjyt@163.com, cxywxr@tust.edu.cn, 1540227453@qq.com
www.tust.edu.cn

Abstract

Existing methods based on deep learning have achieved great success in low-light image enhancement. However, the existing methods generally have a large amount of computation and poor generalization ability for low-light images in different scenes. In order to solve this problem, this paper explores a feasible way to introduce auxiliary blocks in the training process. These blocks can connect the feature map inputs of each stage to the inputs of the first stage to explore whether the model converges. In the prediction phase, due to the superior capabilities of the network, we use only one basic block for inference, so as to greatly reduce the computational cost. Owe to the flexibility of the network, the model can easily handle low-light images in a variety of complex environments.

Keywords: deep learning, low-light image enhancement, auxiliary blocks

1. Introduction

With the increasing availability of devices to capture images and the development of hardware such as impact sensors, people can take pictures in a variety of environments. However, it is difficult to obtain clear and reliable images in poor lighting conditions such as indoors and outdoors at dusk, night, and so on. The pictures taken in these conditions are called low-light images. Therefore, it is necessary to design a method to cope with the brightness of different scenes and enhance the low-light image. In the past decades, many low-light image enhancement algorithms have been proposed, which mainly go through three categories: histogram equalization, Retinex theory and convolutional neural network. Retinex-Net(Retinex Network)[1] is an image enhancement algorithm that combines convolutional neural network with Retinex theory. MBLLEN(Multi-Branch low-light Enhancement Network)[2] achieves low-light image enhancement by extracting the features of multiple branches and fusing them. EnlightenGAN(Enlighten Generative Adversarial Network)[3] is based on unsupervised learning GAN Network to achieve good effect. Compared with traditional algorithms, these methods based on deep learning have made a great leap in performance, but they still have some shortcomings, such as obvious noise and serious distortion, which can not achieve satisfactory results.

In this paper, we extract shallow features through the improved InceptionV2, and then design a sub-network as an auxiliary block to help the network to train. The feature fusion of the auxiliary block is used to guide the network learning and make the training more stable.
After this, we design a joint loss function to further consider the effects of image structure, brightness, detail and noise. In the last section, we prove that the network has excellent performance by comparing subjective and objective evaluation indicators with existing methods.

2. Method

Inspired by the work of Ma et al.[4], our network is composed of shallow feature extraction and three modules in a cascading manner. The image obtained through feature extraction needs to be spliced and transmitted to the middle module. Figure 1 shows the overall framework of our network. The input is a low-light image, from which we first obtain an shallow feature map using an effective strategy(see Section 2.1 for details). In addition, these intermediate modules share the same internal network structure(see Section 2.2 for details) and share weights during the learning process. The method we adopted is based on Retinex theory. This theory holds that the color of an object is determined by the object itself, and has nothing to do with the intensity and color of the light. The original image is obtained by multiplying the reflection and illumination images, as shown in Eq. (1)

$$S = R \times L \tag{1}$$

L is the irradiation component of solar illumination. R is the reflected component of the object itself. S is the imaging result. We modified the Retinex theory and put it at the end of each stage of the network. Furthermore, we define an auxiliary block that accelerates convergence by converging the results of each phase to the same state. The details of the network will be explained next.

2.1. *Feature extraction*

For the input image, we first extract its shallow features, which is very important for low-light image recovery. The extracted shallow feature contains rich details such as edges and textures of images. The network uses an improved InceptionV2 to extract features(as shown in

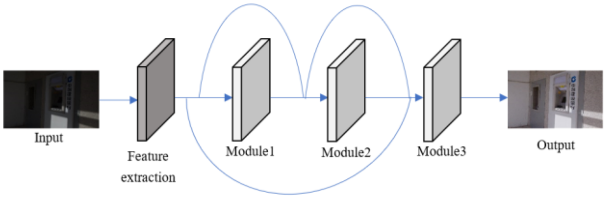

Fig. 1 Overall network framework

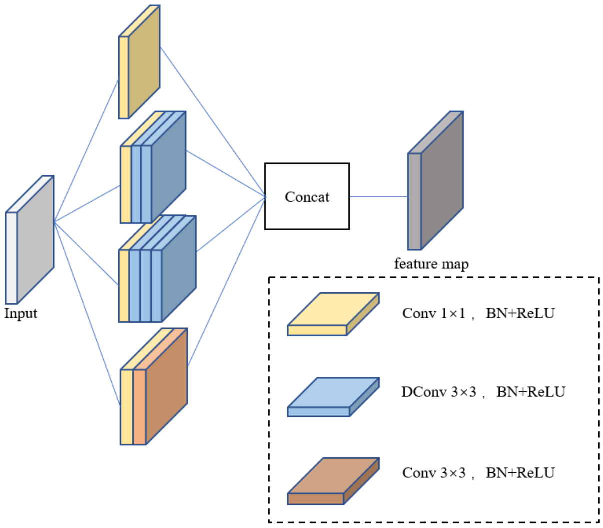

Fig. 2 Feature extraction module

Figure 2). Based on InceptionV2, we delete the average pooling layer and use dilated convolution to avoid the loss of feature information generated in the downsampling process and enlarge the receptive field of the network. The improved InceptionV2 has four branches, and each branch has dilated convolution of different sizes to obtain different scale receptive fields. Finally, shallow features are fused by feature splicing operation.

2.2. *Intermediate module*

In the middle module, part of Retinex theory is used as the illumination adjustment block, and the other part is the auxiliary block. According to previous studies by scholars[4], there is a linear relationship between the irradiation component of solar illumination in Retinex and most regions in low-light images, and the irradiation component L is a core variable for restoring normal illumination. Thus, the exposure component is learned by introducing a model H_θ with a learnable parameter θ, as shown in Eq. (2)

$$F(L^t): \begin{cases} u^t = H_\theta(L^t), L^0 = S, \\ L^{t+1} = L^t + u^t, \end{cases} \quad (2)$$

u^t is the residual difference between the irradiation component of solar illumination at stage t and the low-light image. L^t is the component of the illumination at t stage. Since the weight of the network is shared, the model H_θ will not change significantly in each stage, so that the model H_θ learning residuals does not need to record the number of stages. The network of this model is composed of three convolutional layers, and the structure is shown in Figure 3.

In fact, the enhanced normal light image can be obtained by learning residuals from Retinex theory. However, the reasoning speed of the network will be slowed down because the network is weight sharing. Ideally, the first block will output the desired result, and the subsequent blocks will be similar to the first, so that only one block needs to be used for inference. This is then achieved by building auxiliary block.

The approach here is to link the inputs of each block to the first block for minimizing the differences between stages through a learnable parameter ξ. The auxiliary block is represented by Eq. (3)

$$G(L^t): \begin{cases} R^t = S/L^t, \\ E^t = K_\xi(R^t), \\ V^t = S + E^t, \end{cases} \quad (3)$$

V^t is the input processed by the auxiliary block at each stage. t is always greater than or equal to 1. Then, combining these two blocks (as shown in Figure 4) together as a Module in Figure 1.

2.3. *Joint loss function*

Low-light image has many problems, such as low brightness, color distortion and so on. It is not enough of using simple L1 or L2 loss in regression task to average all pixel differences. Therefore, a joint loss function is introduced to further consider the structure, brightness, detail and noise of the image. The total loss function is defined as Eq. (4)

$$L_{total} = \gamma_f L_f + \gamma_s L_s + \gamma_v L_v \quad (4)$$

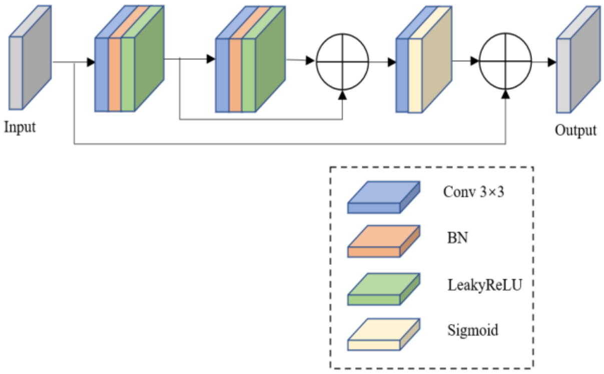

Fig. 3 Model $\boldsymbol{H_\theta}$

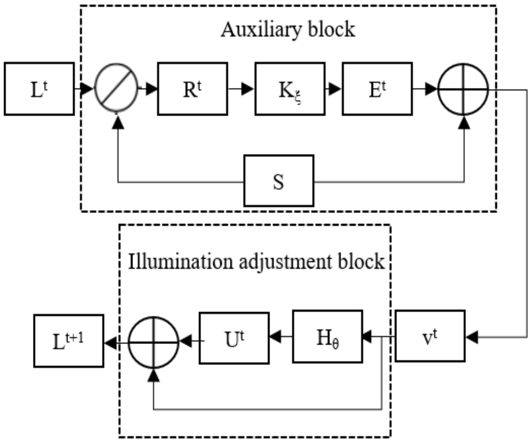

Fig. 4 An intermediate block

Where, L_f、L_s and L_t are fidelity loss, smoothing loss and perceptual loss respectively. γ_f, γ_s and γ_v are the weight balance coefficients, which are 0.8, 0.65 and 0.5 respectively. The fidelity loss is to ensure pixel-level consistency between the enhanced image and the input of each stage, as shown in Eq. (5)

$$L_f = \sum_{t=1}^{T} ||L^t - (s + E^{t-1})||^2 \quad (5)$$

T is the total number of stages. In the formula, the illumination component L^t is limited by $s + E^{t-1}$ to represent the normal illumination image. Real normal light images are not used in order to enhance the generalization of the model.

In addition, we used smoothing losses in our study, as shown in Eq. (6). The enhanced image noise is reduced and the image quality is better.

$$L_s = \sum_{i=1}^{H}\sum_{j=1}^{W}\sqrt{(p_{i,j} - p_{i+1,j})(p_{i,j} - p_{i,j+1})} \quad (6)$$

Where H and W are the width and height of the image, and $p_{i,j}$ is the pixel value of the image.

In order to measure whether the enhanced image is real, image perception loss is introduced here. The basic method is to use the trained VGG16[5] as the future extractor. If the enhanced illuminance component is similar to $s + E^T$, the output from the VGG16 feature extractor should also be similar. The definition is shown in Eq.(7)

$$L_v = \frac{1}{W_{i,j}+H_{i,j}+C_{i,j}} \cdot \sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}\sum_{z=1}^{C_{i,j}} ||\varphi_{i,j}(s + E^T)_{x,y,z} - \varphi_{i,j}(L^T)_{x,y,z}|| \quad (7)$$

$\varphi_{i,j}$ is the feature graph of the j-th convolution layer of block i in VGG16 network. $W_{i,j}$, $H_{i,j}$ and $C_{i,j}$ are the dimensions of each feature graph in VGG16 network, respectively.

3. Experimental results and analysis

3.1. *Experimental detail*

The hardware platform of this experiment is configured with AMD Ryzen 7 5800H CPU and NVIDIA GeForce RTX 3060 GPU. Adam optimizer was used in the training stage, the parameters were $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\alpha = 10^{-8}$, and the learning rate was 0.0001. Iteration epoch is set to 800 times. The LOL-v2 dataset[6] is selected for training in this paper, which is larger and more diverse than the LOL dataset[1]. The dataset consists of 1589 image pairs, which are divided into two parts: synthetic image pairs and real image pairs. Most low-light images are obtained by changing the camera exposure time and ISO.

3.2. *Experimental result*

To test the effect of the model, MBLLEN and EnlightenGAN are selected as a comparison, and the subjective and objective evaluation index are analyzed. In order to show the algorithm effect more clearly, the original diagram is attached.

The subjective effects of vision are evaluated. The comparison results of different algorithms are shown in Figure 5. We selected 5 pictures of indoor and outdoor with different scenes. From left to right, they are original low-light images, MBLLEN, EnlightenGAN, our method and the normal light images.

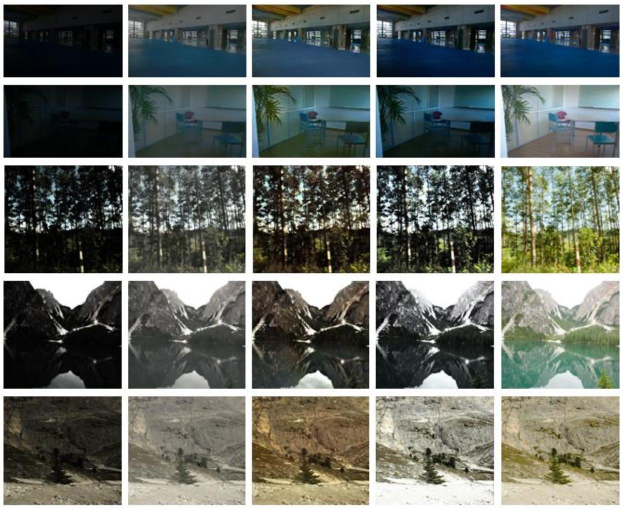

Fig. 5 A comparison of different enhancement algorithms

Tab. 1 Performance index

Methods	PSNR	SSIM
MBLLEN	16.88	0.719
EnlightenGAN	14.04	0.667
Ours	**17.02**	**0.732**

Tab. 2 Parameter quantity

Methods	Parameters(M)
MBLLEN	5.9527
EnlightenGAN	8.636
Ours	**0.0013**

As can be seen from the Figure 5, the overall clarity of the image enhanced by MBLLEN is insufficient, and it loses the image color. Besides, the visual effect is poor, and there are a lot of noise. EnlightenGAN image color distortion is serious, and enhanced boundaries look not very true. The processing result of the algorithm in this paper has clear outline, less noise and moderate brightness. Comparison results show that the improved algorithm in this paper is more reliable and the overall visual effect is close to natural.

As can be seen from Table 1 and Table 2, on the LOL-v2 test set, the algorithm in this paper achieves optimal values in both PSNR and SSIM indexes. It indicates that the quality of low-light images enhanced by the algorithm in this paper is improved. It also shows that the algorithm parameters in this paper are small and it is a lightweight network.

4. Conclusion

In this paper, we design a low-light image enhancement algorithm using auxiliary block to help network training. The Retinex theory makes the reconstructed image better in brightness, and prominent in detail with less color distortion and closer to nature. The enhanced network is not only efficient but also lightweight with excellent performance in terms of image quality. In the follow-up study, we will continue to optimize the algorithm to further improve its ability to maintain color enhancement and other aspects.

Acknowledgment

This work was supported by "Tianjin University of Science and Technology – Softsz Technology Intelligent edge computing Joint Laboratory".

References

1. Wei, Chen et al. "Deep Retinex Decomposition for Low-Light Enhancement." *British Machine Vision Conference* (2018).
2. LV F, LU F, WU J, et al. MBLLEN : Low-light image/video enhancement using CNNs ［C］ //BMVC. 2018, 220（1） : 4.
3. JIANG Y, GONG X, LIU D, et al. Enlightengan : Deep light enhancement without paired supervision ［J］ . IEEE Transactions on Image Processing, 2021, 30 : 2340-2349.
4. Ma, Long et al. "Toward Fast, Flexible, and Robust Low-Light Image Enhancement." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022): 5627-5636.
5. Simonyan, Karen and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." CoRR abs/1409.1556 (2014): n. pag.
6. Yang, W. , et al. "Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement." IEEE Transactions on Image Processing 30(2021):2072-2086.

Authors Introduction

Mr. Yutao Jin

He received his bachelor's degree from the School of Information and intelligence Engineering of Ren'ai College of Tianjin University in 2022. He is acquiring for his master's degree at Tianjin University of science and technology.

Dr. Xiaoyan Chen

She, professor of Tianjin University of Science and Technology, graduated from Tianjin University with PH.D (2009), worked as a Post-doctor at Tianjin University (2009.5-2015.5). She had been in RPI, USA with Dr. Johnathon from Sep.2009 to Feb.2010 and in Kent, UK with Yong Yan from Sep-Dec.2012. She has researched electrical impedance tomography technology in monitoring lung ventilation for many years. Recently, her research team is focus on the novel methods through deep learning network models.

Mr. Xiwen Liang

He received his bachelor's degree from the school of electronic information and automation of Tianjin University of science and technology in 2021. He is acquiring for his master's degree at Tianjin University of science and technology.