



# Enhanced network optimized generative adversarial network for image enhancement

Lingyu Yan<sup>1</sup> · Jiarun Fu<sup>1</sup> · Chunzhi Wang<sup>1</sup> · Zhiwei Ye<sup>1</sup> · Hongwei Chen<sup>1</sup> · Hefei Ling<sup>2</sup>

Received: 3 June 2020 / Revised: 1 November 2020 / Accepted: 22 December 2020 /

Published online: 23 January 2021

© The Author(s) 2021

## Abstract

With the development of image recognition technology, face, body shape, and other factors have been widely used as identification labels, which provide a lot of convenience for our daily life. However, image recognition has much higher requirements for image conditions than traditional identification methods like a password. Therefore, image enhancement plays an important role in the process of image analysis for images with noise, among which the image of low-light is the top priority of our research. In this paper, a low-light image enhancement method based on the enhanced network module optimized Generative Adversarial Networks(GAN) is proposed. The proposed method first applied the enhancement network to input the image into the generator to generate a similar image in the new space, Then constructed a loss function and minimized it to train the discriminator, which is used to compare the image generated by the generator with the real image. We implemented the proposed method on two image datasets (DPED, LOL), and compared it with both the traditional image enhancement method and the deep learning approach. Experiments showed that our proposed network enhanced images have higher PNSR and SSIM, the overall perception of relatively good quality, demonstrating the effectiveness of the method in the aspect of low illumination image enhancement.

**Keywords** Enhanced network · Generative adversarial networks · Low-light · Image enhancement

## 1 Introduction

Image enhancement has become an important direction for image processing research. Their ultimate research goal is to clear low-quality images, enhance detail information, improve contrast, reduce noise, to enrich image content information, improve overall perception quality, and finally meet the image requirements of advanced visual tasks. Light and shadow have the most extensive influence on the picture and the greatest influence on the picture

---

✉ Lingyu Yan  
yanlingyu@hbut.edu.cn

quality. However, in the current research results, there are no perfect means to deal with the above problems. So the study of low light enhancement brings a lot of academic and practical value. Recently, due to the good performance of machine learning algorithm [23, 33], several low-light image enhancement based on machine learning have been proposed [7, 12, 13].

The paper [27] proved that the stacking sparse de-noising self-encoder is based on the training of synthesized data benefits enhancing and de-noising the low-light noisy image. Model training is based on image patches and adopts sparsity regulative reconstruction loss as a loss function. The main content is:(1)A method of training data generation (gamma correction and Gaussian noise addition) is proposed to simulate the low-light environment. (2)The article discusses two different network structures:(a) LLNet, Learn both contrast enhancement and denoising:(b) S-LLNet, Two modules are used to perform contrast enhancement and denoising in stages. (3)Experiments on real low-light images show that the model is effective. (4)It visualizes network weights and provides insights into the characteristics learned. In the first method of using LLNet for two operations, the automatic encoder module is composed of multi-layer hidden units, in which the encoder is trained through unsupervised learning, decoding weights are provided by the compiler and the decoding error is adjusted through backpropagation. In the second method, LLNet with contrast enhancement and noise reduction module and S-LLNET with sequence contrast enhancement and noise reduction module are proposed. The two modules are used to carry out contrast experiment in stages.

In paper [31], CNN is introduced. The traditional multi-scale Retinex (MSR). This method is a feedforward convolutional neural network with different Gaussian convolution kernels, and its construction is analyzed in detail. Then, following the MSR process, they came up with MSR-NET, which acts on the end-to-end mapping from low-light to light images. The training data were high-quality images adjusted by PS and the corresponding composite low-light images (random reduction of brightness, contrast, gamma correction). The loss function is the F- norm square of the error matrix with regular terms, that is, the sum of the error squares. MSR-net includes three modules: multi-scale logarithmic transformation, convolution difference, color recovery.

Paper [3] mainly focuses on single image contrast enhancement (SICE), which for low contrast, underexposure problems. Its main content is as follows :(1) Multi-exposure image data sets were constructed, including a comparison of images with different exposures and corresponding high-performance images. (2) A two-stage enhancement model is proposed, as shown in the following figure. In the first stage, the weighted least square (WLE) filtering method has segmented the image into components of different frequencies, and the low-frequency and high-frequency components are enhanced. The second stage fuses the enhanced low - and high-frequency components, and then enhances them again, and outputs the result. Due to single-phase enhancement result of CNN is not satisfactory, and there is the phenomenon of color, it may be that the single-phase CNN difficult to balance the image smooth component and texture component of enhancement effect, so the network structure design into two stages, the model of the first phase of Decomposition steps using the traditional method, and introduces Retinex-Net using CNN has been realized.

Inspired by the Retinex theory, paper [35] adopts a two-stage decomposition and then enhancement step, which is completely implemented by CNN. For the training of deconvnet, consistency of reflectance, and smoothness of illumination were introduced, which were very easy to reproduce and achieved good experimental results. The article talks about: (1) Constructed the LOL dataset of low-light/normal light of paired, which should also be the first paired dataset to be collected in real scenes. The dataset is divided into two parts: the

image data of real scenes is obtained by changing the camera sensitivity and exposure time; The synthesized image data is obtained by adjusting the Adobe Lightroom interface, and the y-channel histogram of the adjusted image must be as close as possible to the low-light scene in reality. (2) Retinex-net is proposed, which is divided into two sub-networks: decom-net can decouple the image and obtain the illumination map and reflection map; The enhanced lightmap is enhanced by multiplying the enhanced lightmap with the original reflection map. Besides, considering the noise problem, a combined de-noising and enhancement strategy is adopted, and the de-noising method is BM3D [8]. (3) A structure-aware total variation constraint is proposed, which is to use the reflection graph gradient as the weight to weight TV loss, to ensure smooth constraints without destroying the texture details and boundary information. The overall flow of the enhancement is as follows: decomposition. In the decomposition step, the neutron network decomposition network dissolves the input image into a reflection and lightmap. After that, to adjust the steps, the intensifier network based on encoder and decoder enhances the illumination, and introduces a multi-scale cascade to facilitate the adjustment effect from multi-scale. In this step, the noise on the reflection map is also eliminated. Finally, the adjusted lighting and reflection maps are reconstructed for enhanced results. The core idea of paper [28] is the extraction and fusion of features at different levels in the network. Besides, another highlight of this paper is the low-light enhanced network for video. Different from the direct method of frame-by-frame processing, they use 3D convolution to improve the network and effectively improve the performance. Because of the negative side of the video's low-light enhancement, flickering, there may be a jump in brightness from frame to frame that does not meet expectations. This problem can be measured by the AB(Avr) index (mean brightness variance). The network structure contains feature extraction module FE, FEM is a single-flow network with 10 layers of convolution, the output of each layer is input into each EM sub-module, and the hierarchical features are extracted respectively. Finally, these hierarchical features are spliced together and the final result is obtained by the fusion of 1x1 convolution. Besides, the enhancement module EM and fusion module FM are also included. Loss function: this paper does not use the conventional MSE or MAE loss, but proposes a new loss function, including three parts, namely structure loss, content loss, and area loss. The structure loss is in the form of SSIM and m-ssim. Content loss, that is, the features of VGG extraction should be as similar as possible; The area loss makes the low-light area more focused by the network.

Paper [5] focus image imaging systems under extreme light conditions. And it has completed the processing from the original image to the RGB image, the experimental results are very excellent. The network structure is based on the fully convolutional network FCN, the loss function adopts L1 loss, and it is directly trained end-to-end. In addition, the see-in-the-dark data set is proposed, which is composed of the short exposure image and the corresponding long-exposure reference image.

For Bayer's original image, they split the input into four channels, reducing the spatial resolution by 1/2 on each dimension. The output is 12 channels with 1/2 spatial resolution. The sub-pixel layer is then processed to restore the original resolution.

Paper [38] proposed three difficulties in low-light enhancement tasks: (1) how to effectively estimate the lightmap components from a single image and flexibly adjust the light level? (2) how to remove degradation such as noise and color distortion after improving the image brightness? (3) how to train the model without ground-truth and with a limited number of samples? The enhancement ideas in this article still follow the two-stage approach of decomposition-then-enhance of retinex-net. The network is divided into three modules: decomposition-net, instance-net, and adjustment-net, which respectively perform image decomposition, reflection map recovery, and illumination map Adjustment. Some

innovations are as follows : (a) for disposition-net, the loss function not only follows the refactor loss and consistent reflection loss of retinex-net but also adds two new loss functions for the regional smoothness and consistency of the lightmap. (b) for Restoration-Net, considering that the reflection map under low illumination often has a degradation effect, the reflection map under good illumination is used as a reference. The distribution of the degradation effect in the reflection map is very complex and highly dependent on the illumination distribution, so the illumination map information is introduced. (c) for adjust-net, a mechanism is implemented to continuously adjust the light intensity (After combining feature map and lightmap, the enhancement ratio is taken as the input). The comparison with gamma correction shows that the adjusted results are better in practice. The KinD network consists of two branches, corresponding to the reflection diagram and the illumination diagram. According to its three different functions, we divided it into different modules, including layer decomposition, reflection map recovery, and illumination map adjustment.

Existing researches improve image enhancement performance to some extent, but there are no perfect means to deal with the extensive influence of light and shadow on image quality. In this paper, we propose a low-light image enhancement method based on the enhanced network module generative antagonism network, which deepens the number of GAN network layers, improves the modeling ability of low illumination image enhancement, and avoid feature loss caused by network deepening. The remainder of this paper is organized as follows. In Section 2, we review the Generative Adversarial Networks and ResNet. Then we present the Enhanced Network Optimized Generative Adversarial Network in Section 3. In Section 4, extensive experiments are conducted to demonstrate the effectiveness of the proposed algorithm. Finally, we conclude Section 5.

## 2 Related work

This chapter will expound several important points in this research.

### 2.1 Generative antagonistic network

Let's start with the original GAN [14]. Ian j. Goodfellow et al, in October 2014 in Generative Adversarial Networks put forward a process through the confrontation in the estimated to generate a new framework of the model. In this framework, the generation model G is trained to capture the distribution of data. In terms of probability estimation, the model trains discriminant model D, which can obtain estimation results from massive training data samples. The specific information is shown in Fig. 1.

As can be seen from Fig. 1, We can divide the generated against the network into two parts: the generation model G (generative model), the discriminant model D (discriminative model). The generated model is used to learn the distribution of real data. There are two types of input data, real data and generated data. The choice of input data type is determined by the discriminant model of binary classifier. X is a real number, which corresponds to the distribution of  $P_r(x)$ . Z is the hidden space variable, also corresponds to the distribution of  $p_z(z)$ , Like a Gaussian or a uniform distribution. Then sample from the assumed hidden space z and generate the data  $x = G(z)$  by generating model G. Then the real data and the generated data are fed into the discriminant model D, and the judgment category is output. In the original generative antagonistic network framework, the discriminant model needs to carry out a dichotomy, so the most basic idea is to use the method of the binary cross-entropy loss function. For the real image, the given label is 1, and for the generated image

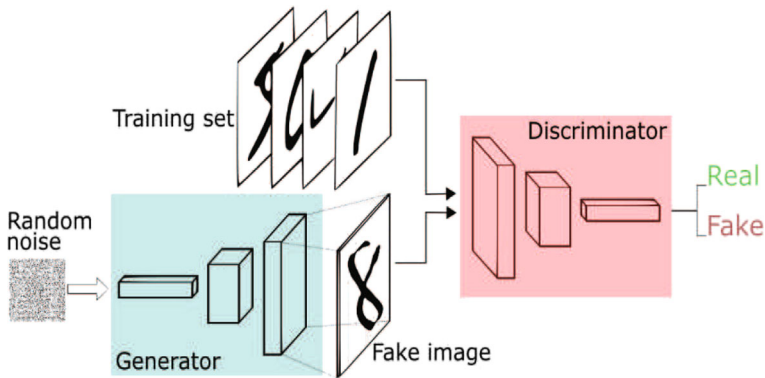


Fig. 1 Standard generation antagonistic network schematic

it is 0. The generated model  $G$  tries to synthesize the picture so that the discriminant model  $D$  can be true. From the comparison of the loss function, we can see that the change is the only very simple addition of the constraint term, and it does not make CGAN (Conditional GAN) [29] adds a constraint term to the generator and discriminator respectively, which is expressed as  $y$  in the formula. After adding the constraint term, it can partially solve the shortcomings such as too free generation process, but it is obviously not the final solution. The loss function is (2).

$$\min_G \min_D V(D, G) = E_{x:P_x} [\log_2(1 - D(G(z|y)))] \tag{1}$$

From the comparison of the loss function, we can see that the change is the only very simple addition of the constraint term, and it does not make any structural change to the final result. Therefore, many problems existing in GAN still exist in CGAN.

In the next transformation of GAN, LAPGAN [9] introduced the Laplace pyramid into the theoretical structure to generate relatively high-quality pictures. Its essence is to use the Laplace pyramid to learn the residuals between different stages.

Later in the development process, the use of batch normalization (BN) [18] stability and combining the convolution of the neural network training DCGAN [30], successful solution to the corresponding relations between recognition InfoGAN [6], has solved the original GAN different aspects of the different problems, that now use against network has become feasible to generate high-quality images clearly, and efficiency, relatively ever greatly ascend.

## 2.2 ResNet

In the field of deep learning, most of the previous experimental results show that the network depth is positively correlated with the performance of the model. Theoretically, the more layers there are in the network, the more complex feature models can be extracted and the better the overall performance of the model will be. However, subsequent experiments show that there is saturation degradation in the deep network. When the network accuracy tends to saturation, increasing the network depth will reduce accuracy.

To solve the problem of saturation degradation mentioned above, ResNet (Residual Neural Network) [16] was proposed and applied in the field of deep learning. Compared with

other types of networks, Resnet's unique structure accelerates the training of neural networks. For example, compared with VGGNet [32], the number of parameters is much lower, which makes Resnet's training speed significantly superior. At the same time, Resnet also has a significant improvement in accuracy, which makes it very versatile. For example, Resnet is introduced into InceptionNet [34] to improve the overall accuracy. The main idea of ResNet is to add a direct connection channel to the Network, which solves the problem of gradient disappearance and explosion, and improve the performance of the network while training the deeper level network.

### 2.3 Basic theory of low-light image enhancement

Now we turn our attention to the physical model and deep learning, two different research directions of low-light image enhancement.

The theory of low-light enhancement in physical models is the earliest research in this field. At present, it only serves as the theoretical reference of image enhancement with low-light. Histogram equalization (HE) [2], through nonlinear stretching of the image, makes the transformed image histogram evenly distributed and enhances the contrast. This method is widely used and effective as the basic theory in this field. However, due to the poor quality spectral table of low-light images, it is difficult to take into account all kinds of problems, such as texture and color, with such a simple method.

Retinex [24] theory was proposed in 1963. In the experiment, single-scale Retinex theory (SSR) [20] and multi-scale Retinex theory (MSR) [21] were applied, and the experiment [10] successfully proved that this method can improve the color of pictures. For the first time, this method makes it possible to improve the color of the picture. After that, different methods were proposed one after another, and various problems in the research process were gradually improved. And the method based on filtering [36], RGB space is converted to HSV space to achieve the effect of low illumination enhancement [25], the fast adaptive image enhancement algorithm based on wavelet transform for low illumination [25], the definition of dark channel prior in mathematics [15], and the four-dimensional binary tree is introduced to improve the halo in the image [26]. Both have improved different aspects of the problem, but the researchers gradually found that it is very hard to continue to improve the efficiency of the enhancement algorithm using traditional methods. Meanwhile, the advent of ALEXnet has led researchers to focus on deep learning.

The literature published on CVPR in 2018 [4] focuses on the image imaging system under extremely low light conditions and short exposure conditions. It uses CNN to complete the processing of raw images to RGB images, and the experimental results are very impressive. The network structure is based on the fully convolutional network FCN directly trained end-to-end, and the loss function adopts L1 loss.

The literature [37] presented on arxiv in 2019 proposed and solved three long-standing difficulties in the field of low-illumination images: (1) adding two different targeted loss functions for smoothness and consistency, so that the generation of additive can flexibly solve the brightness problem. (2) introduce illumination map information to solve image distortion and other degradation problems. (3) a mechanism that can continuously adjust the light intensity is realized to make different models can be trained efficiently, and the feasibility is proved by experiments.

It is not difficult to see that the introduction of theoretical knowledge related to deep learning makes the field of low-illumination image enhancement enter a new stage and solves many long-standing problems. In the experiment, we also try to introduce different networks based on deep learning to make algorithms feasible.

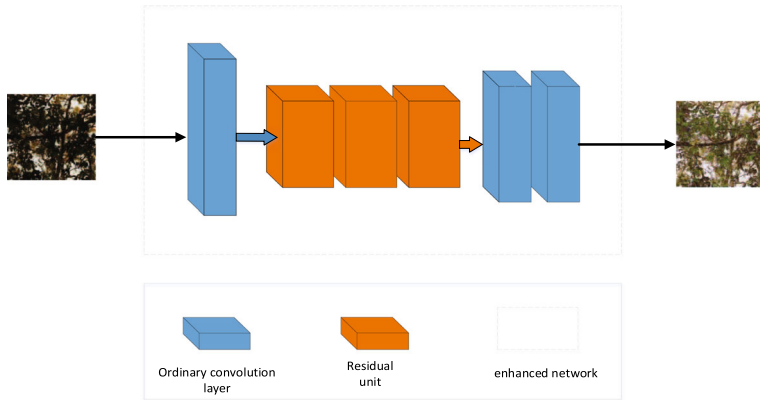


Fig. 2 Schematic diagram of generator network structure

### 3 Enhanced network optimized generative adversarial network

#### 3.1 Algorithm framework

##### 3.1.1 The network structure

The network structure of low-illumination image enhancement is shown in Figs. 2 and 3. In order to enhance the transmission of weak light position information in low-illumination images in the network flow, the generator adds an enhanced network.

The details of the enhanced network structure are shown in Table 1, consist of a full convolutional network (FCN) and utilizes the attributes of a convolutional neural network, such as translation invariance and parameter sharing. The network consists of multiple residual blocks and convolution block 2 parts. I start with one convolution block. The middle section contains four residual blocks, with constant height/width, and after each convolution

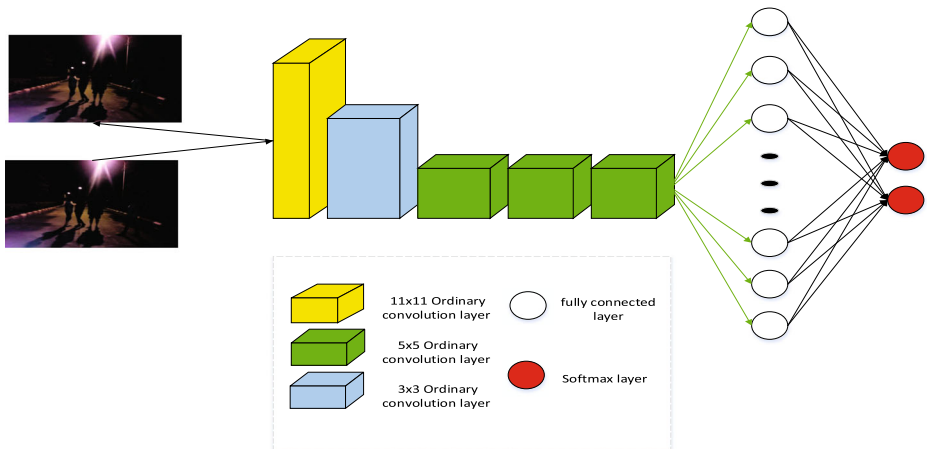


Fig. 3 Schematic diagram of discriminator network structure

**Table 1** The details of the network structure of the enhanced network

(Convolution layer)	(Input)	(convolution kernel/perstep)	(output)
(Conv0)	(I(100x100x3))	(9x9x64/1)	(100x100x64)
(res1)	(conv0)	(3x3x64/1)	(100x100x64)
(res2)	(res1)	(3x3x64/1)	(100x100x64)
(res3)	(res2)	(3x3x64/1)	(100x100x64)
(res4)	(res3)	(3x3x64/1)	(100x100x64)
(Conv1)	(res4)	(3x3x64/1)	(100x100x64)
(Conv2)	(conv1)	(9x9x64/1)	(100x100x3)

is instantiated for regularization and ReLU activation. And then finally we have two convolution blocks. Except for the residual block, the last convolution layer is tanh activated, and only ReLU is activated after each convolution layer.

Table 2 shows the details of the discriminator network, including 5 convolution layers, 1 fully connected layer, and a softmax layer. Multiple convolution layers used for feature extraction step by step input, the size of the convolution kernels from 11 to 3, the characteristics of the channel number increases from 3 to 192, for the low illumination image, due to the uneven illumination distribution and noise impact, such as image showing a large area of dark, weak light, such as single lead to local characteristics, first big receptive field is beneficial to the local characteristic diagram for more information, with the increase of the number of channels, feature-rich gradually, that small receptive field is helpful to extract the image. The full join layer and softmax layer are used to predict the likelihood that the extracted feature map will come from a generator or a real image, resulting in a triple( $batch$ ,  $P_{true}$ ,  $P_{false}$ ),  $P_{true}$ ,  $P_{false}$ . They're all in the range of.

### 3.1.2 Loss function

Since the input and the target photo cannot match closely (i.e., pixel to pixel), that is, different optical elements and sensors can lead to specific local nonlinear distortion and aberration, even after accurate alignment, the pixel number between each image pair will have an unsteady deviation. Therefore, the standard loss per pixel, except for the perceived

**Table 2** The network structure details of discriminator network

(Convolution layer)	(Input)	(convolution kernel/perstep)	(output)
(Conv0)	(I(100x100x3))	(11x11x48/4)	(25x25x48)
(Conv1)	(Conv0)	(5x5x128/2)	(13x13x128)
(Conv2)	(Conv1)	(3x3x192/1)	(13x13x192)
(Conv3)	(Conv2)	(3x3x192/1)	(13x13x192)
(Conv4)	(Conv3)	(3x3x128/2)	(7x7x128)
(Fc)	(conv4)	(6272x1024)	(batch x1024)
(softmax)	(Fc)	(1024x2)	(batch x2)



quality index, does not apply to our case. To enhance the effectiveness of the algorithm from both qualitative and quantitative aspects, we propose a new loss function:

$$Loss = W_{con}L_{con} + W_{tv}L_{tv} + W_{col}L_{col} + W_{adv}L_{adv} \tag{2}$$

$L_{con}, L_{tv}, L_{col}, L_{adv}$  represent content loss, attention loss, total variation loss, color loss, and confrontation loss,  $W_a, W_{adv}, W_{cov}, W_{tv}, W_{col}$ , represent the weight of their losses respectively. The weight of each loss is analyzed according to the specific situation of its picture, determined according to the size (or proportion) of the impact of various losses on the image.

### 3.1.3 Algorithm flow chart

After the training data is obtained, we continued to use high-quality images for repeated training of GAN. In the training discriminator stage, we randomly confused the generated sample of one batch with the real sample of one batch and generated the generated sample of one batch as the discriminator input. The discriminator tries to identify real and fake images, so the discriminator is trained, which is equivalent to the process of maximizing the discrimination loss. The process is to minimize the formula (2), to ensure that the generated picture has the least loss in all aspects compared with the real picture, and the generated effect is realistic. In order to express the whole algorithm flow more concisely and clearly,  $G, D$  are respectively represented as generator network and discriminator network, and the size of a batch is  $m$  during training. Refer to the following Enhanced network module generative adversarial network algorithm flow for details. Data description:  $I_x$  is the low-light image,  $I_y$  is the real image,  $I_{adv}$  is defined as the input to the discriminator.

---

#### Algorithm 1 Enhanced network optimized generative adversarial network algorithm.

---

**Require:**

- the low-light image:  $I_x$
- the real image:  $I_y$
- the input to the discriminator:  $I_{adv}$

**Ensure:** – The Enhanced image

**repeat**

**Training generator network**

- 1:  $M$  low-light image pairs were randomly selected in the data set  $(I_x^1), (I_x^2, I_y^2)!$   $(I_x^m, I_y^m)$ .
- 2: The input of the fixed discriminant network is  $I_{adv}=0,0;0$ , length of  $m$
- 3: The total loss of the generator network:  $Loss_{gen} = W_a L_a + W_{adv} L_{adv} + W_{con} L_{con} + W_{tv} L_{tv} + W_{col} L_{col}$

**Train the discriminator network**

- 4: The input of the randomly initialized discriminant network is  $I_{adv}=1,0;0$ , and  $m$  stands for length.
- 5:  $M$  low-light image pairs were randomly selected in the data set  $(I_x^1), (I_x^2, I_y^2)!$   $(I_x^m, I_y^m)$ .
- 6: Maximizes the overall loss of the discriminator network:

$$L_{adv} = \sum_{i=1}^n \log D(G(I_x), I_y)$$

**until** Maximum number of iterations do

---

### 3.2 Enhanced network

As a backbone network, this module is used to overcome the disadvantage of low contrast and improve the details of the picture, so as to achieve the effect of enhancing the picture. Considering that there is little feature information of the low-light region in low-illumination pictures, and there may be noise interference, we use the residual connection to build the basic residual module, which is used to deepen the number of network layers, improve the modeling ability of the network to enhance low-illumination pictures, and avoid the feature loss caused by network deepening. Therefore, the residual module is used as the feature transformation layer in the enhanced network, as shown in Fig. 4.

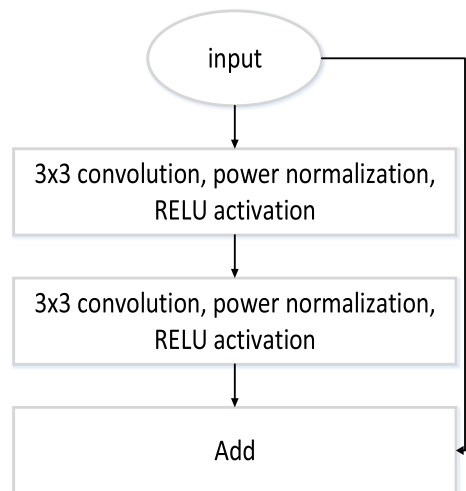
As shown in Fig. 4, the module contains 2 layers of 3x3 ordinary convolution. The concatenation of two 3x3 convolution layers has the same perception as the 5x5 convolution layers, with the former having fewer parameters. Multiple 3\*3 convolution layers have more levels of nonlinear functions than larger convolution layers, which increases the nonlinear expression and makes the decision function more decisive. After each convolution, the instance normalization and Relu activation are carried out, and we get the final result by adding input. The residual network is effective in image segmentation and target monitoring. This operation USES residual connections to train deeper networks and avoid feature loss. By introducing residual touch block, we connect multiple residual cells to carry out complex feature transformation, the network can improve the modeling ability of low-illumination pictures.

To improve the overall image perception quality and constrain the enhancement network, the following four-loss functions are designed in the next section:

#### 3.2.1 Content loss

Firstly, the definition of the content loss function is given. We use the activation diagram of the VGG-19 network as the basis of the function. This loss makes them have similar feature representations, rather than measuring per-pixel differences between images, but perceived quality. In our vision, it is mainly used to preserve the semantics of the image, and it is not defined to take into account other aspects of loss. Let be the feature graph obtained by the

**Fig. 4** This chapter residual module network structure diagram



VGG-19 network after the  $i$ th convolutional layer, then we define the content loss as the distance between the enhanced image and the target image features in Euclidean distance:

$$L_{con} = \frac{x1}{C_i H_I W_I} \|\phi(G(I_x, A) - \phi_i(I_y))\| \tag{3}$$

$C, H, W$  are the enhanced image’s numbers of channels, height and width.  $I_e$  is low-light image and  $I_e$  is real image.

### 3.2.2 Total variation loss

This loss can enhance the image’s spatial smoothness, operate on the pixels of the generated composite image, meanwhile promote the generated image’s spatial continuity, so as to avoid major differences between adjacent pixels and prevent the appearance of a checkerboard pattern in the image.

$$L_{tv} = \frac{x1}{C H W} \|\nabla_x I_e - \nabla_y I_e\| \tag{4}$$

$C, H, W$  are the enhanced image’s numbers of channels, height, and width.  $\nabla_x I_e, \nabla_y I_e$  is the enhanced image’s gradient in the opposite direction of X and Y. Because of its relatively low weight, it can eliminate noise without damaging the high-frequency part of the picture.

### 3.2.3 Color loss

Color loss aims to avoid color distortion and to evaluate the color difference between the enhanced and target image, we introduce color loss. It first applies a gaussian blur to images and then calculates their Euclidean distance. The color loss is defined as follows:

$$L_{col} = \|\delta(G(I_x)) - \delta(I_y)\|^2 \tag{5}$$

$\delta$  represents the Gaussian blur function, which preserves the global information of the image, such as color, while removing the local details.

### 3.2.4 Confrontation loss

This loss encourages the generation of network conversions from low-illumination images to natural images, facilitating generators to learn the features of natural images, including texture, contrast, and so on. At the same time, we use gradient punishment to stabilize the training of the discriminator. The combat loss is defined as follows:

$$L_{adv} = - \sum_{i=1}^m \log D(G(I_x), I_y) \tag{6}$$

$D$  represents the discriminant network,  $G$  represents the generating network,  $I_x, I_y$  represent the low-illumination picture and the natural illumination picture respectively.

## 4 Experimental results and analysis

### 4.1 Experiment setting

(1) The experiment data set

First of all, in terms of data set selection, we chose low light paired data set (LOL), which contains 500 pairs of low light/normal light images. It is a dataset of images used for low-light enhanced shots in real scenes. Most low-light images are collected by changing the exposure time and ISO and using a three-step method to align the image pairs. The dataset contains images captured from various scenarios, such as houses, campuses, clubs, and streets. Since it is difficult to obtain low illumination data sets, we use the LOL dataset to synthesize low illumination data sets and uses gamma correction and random parameters to add low brightness to the image, yielding 11,592 images of the training set and 360 images of the test set.

We also select another data set DPED to continue our experiment. The DSLR algorithm's low-light image enhanced data set (DPED) contains a large real-world data set collected by three smartphones and a DSLR camera. The final data set contains 4,549 photos from the Sony smartphone, 5,727 photos from the iPhone, and 6,015 photos from the blackberry camera. We selected the iPhone alignment image data set provided by the author, which contains about 160,000 images.

We completed the experiment of this paper through Tensorflow [1]. The network proposed in this paper converges rapidly and has been trained for 20,000 generations on NVIDIA GeForce GTX1080 using the synthesized data set. To prevent overfitting, we used rollover and rotation for data enhancement. We set the batch size to 32 and the input image value is scaled to [0,1]. We use layer 4 of the fifth convolution module in the vgg-19 network as the perception loss extraction layer.

## (2) Implementation Detail

In the experiment, the Adam optimizer [22] was used for training, and we also used the learning rate attenuation strategy. When the loss index stopped improving, we reduced the learning rate by 50 percent. Meanwhile, to stabilize Gan training, we used spectral normalization and gradient punishment to constrain the discriminator.

To verify the performance of the image enhancement algorithm, the following typical image enhancement algorithms are set up as control tests: histogram equalization (HE), reflection illumination estimation (SRIE) [11] Efficient and unsupervised generative confrontation network(EnlightenGAN) [19], and DSLR [17].

In order to evaluate the quality of the enhanced image, we use PSNR and SSIM for quantitative analysis. Their definitions are as follows:

Peak signal-to-noise ratio (PSNR): It is an objective standard for evaluating images, the unit is dB, and the calculation formula is as follows:

$$PSNR = 10 * \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (7)$$

Where n represents the number of bits of each pixel value, MSE represents the mean square error, and the calculation formula is as follows:

$$MSE = \frac{1}{w * h} \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} ||X(i, j) - Y(i, j)||^2 \quad (8)$$

Among them, X, Y represents the source image and target image, respectively, w, h is the width and height of the image, if the PSNR value is greater, it means that the distortion is less, the higher the quality of the restored target image.

Structural similarity (SSIM) is a comprehensive image brightness, contrast, and structural difference to evaluate the similarity of the structure of two images. The mathematical expression is as follows:

$$SSIM(X, Y) = 1(X, Y)^\alpha c(X, Y)^\beta s(X, Y)^\gamma \tag{9}$$

In the formula, X, Y represent the source image and the target image, respectively,  $1(X, Y)$ ,  $c(X, Y)$ ,  $s(X, Y)$  represent the image brightness, contrast and structure difference, the calculation formula is as follows:

$$1(X, Y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{10}$$

$$c(X, Y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{11}$$

$$s(X, Y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{12}$$

Where  $\mu_x, \mu_y$  represent the average brightness of the source image x and the target image  $\sigma_x, \sigma_y$  represent the brightness standard deviation of the image x, y, and  $\sigma_{xy}$  represent the covariance between the image x and y.  $C_1, C_2, C_3$  are constants set to avoid the denominator being 0, generally set  $\alpha = \beta = \gamma = 1$ , the SSIM value is usually [0,1], the larger the value, the better the restored image effect.

SSIM-GC. In the low-light enhancement task, the average luminance level is hard to be predicted. Therefore, the detailed fidelity might be not well captured by PSNR and SSIM. Therefore, SSIM-GC is introduced, where a global illumination is corrected first via the Gamma transformation and then SSIM values are calculated.

### 4.2 Data set experiment

On LOL data set, we compared four methods of HE, SRIE, EnlightenGAN and DSLR. Because some methods could not achieve the denoising function, we combined with BM3D method to achieve the denoising, and then produced the final result. Table 3 shows the quantitative results, while Figs. 5 and 6 shows the qualitative results.

Contrast Figs. 6 and 7 of each algorithm in the subjective visual effect can be seen, HE appeared on the algorithm in LOL data set more content of distortion and color distortion, such as enhancing the first picture, HE’s in the picture on the wall there is an obvious artifact that is independent of the background and overall image grey, enhance the second image, light color floor brunet floor, and real photos is an apparent discrepancy; The images enhanced by SRIE algorithm have more brightness distortion. It can be seen from the first and second images that the improvement of the brightness of low-illumination images is not enough to identify the image content.

**Table 3** Our method was compared with the experimental results of HE, SRIE, ENlightenGAN and DSLR in the LOL data set

	HE	SRIE	EnlightenGAN	DSLR	our
PSNR	15.4271	11.8453	18.23	19.1426	20.9983
SSIM	0.573)	0.5347	0.615	0.9982	0.9986
SSIM-GC	0.6037	0.5928	0.6874	0.9985	0.9988

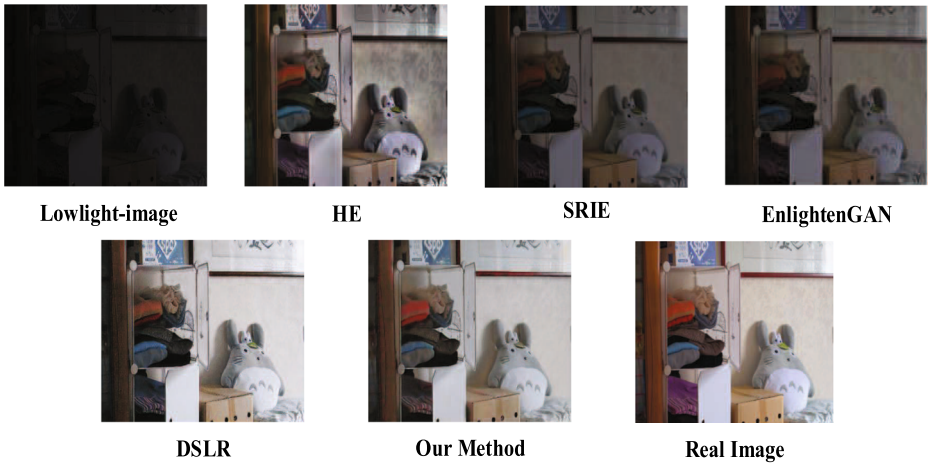


Fig. 5 Our methods work with HE, SRIE, DSLR and EnlightenGAN in the LOL dataset for visual effects

The EnlightenGAN algorithm has the problem of excessive noise in the generated pictures in the experiment, and at the same time, the increase in brightness is only similar to the SRIE algorithm. For example, in the comparison of the enhancement results of the first and second pictures, the brightness of the generated pictures is similar to the result of the SIRE algorithm, but there is too much noise, which causes the pictures to be much blurred compared with the results of the SIRE algorithm; D SLR can enhance the image illumination and retain a certain content, but there are still a noise exposure and local problems, such as enhanced in the two pictures, the first white of bright, closet door, floor and wall of the second picture brightness is too high, enlarge images can be found at the same time, the local has certain noise and details is not smooth; However, the brightness of the enhanced images is moderate, which avoids the problem of overexposure, while the content remains intact and the details are less lost.

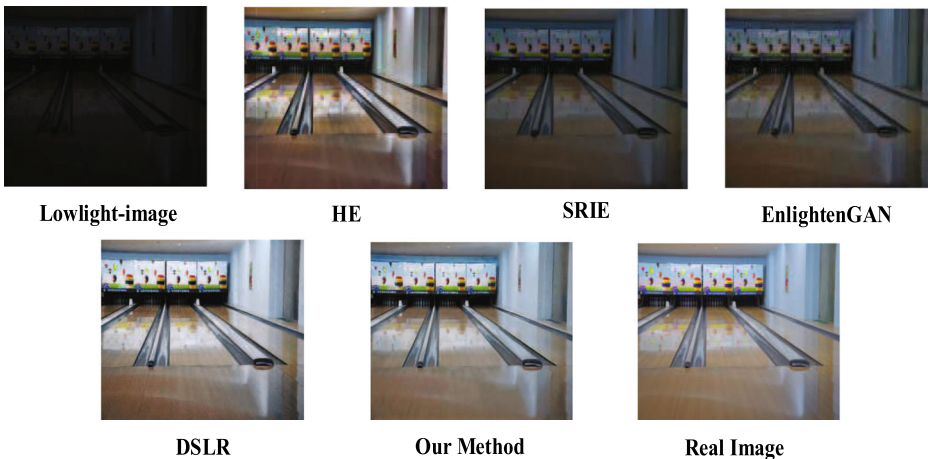


Fig. 6 Our methods work with HE, SRIE, DSLR and EnlightenGAN in the LOL dataset for visual effects



**Fig. 7** Our methods work with HE, SRIE, DSLR in the DPED data set for visual effects

In the comparison of Table 3, we can see that our algorithm is superior to other methods in terms of both indicators, indicating that our method also has excellent performance and good robustness in real data sets. Compared with the traditional methods of HE and SRIE, we improved 36.11 percent and 77.27 percent on PSNR, 74.06 percent, and 86.76 percent on SSIM, respectively. On SSIM-GC, it increased by 39.51 percent and 40.6 percent. Compared with the current deep learning method DSLR and EnlightenGAN, we also improved 9.69 percent and 2.768 percent on PSNR. And we improved 0.04 percent and 38.3 percent on SSIM. On SSIM-CG, it increased by 0.003 percent and 31.14 percent respectively. It can be seen that our algorithm has certain advantages over traditional and deeper learning methods.



**Table 4** visual effects of our methods with HE, SRIE, EnlightenGAN and DSLR in DPED data sets

	HE	SRIE	EnlightenGAN	DSLR	our
PSNR	14.3	15.2	19.16	20.42	20.46
SSIM	0.54	0.64	0.6823	0.9192	0.9156
SSIM-GC	0.5832	0.6973	0.7245	0.9187	0.9312

After that, we select another data set DPED to continue our experiment. The DSLR algorithm's low-light image enhanced data set (DPED) contains a large real-world data set collected by three smartphones and a DSLR camera. The final data set contains 4,549 photos from the Sony smartphone, 5,727 photos from the iPhone, and 6,015 photos from the blackberry camera. We selected the iPhone alignment image data set provided by the author, which contains about 160,000 images.

On the DPED data set, we compared the four methods of HE, SRIE, DSLR, and EnlightenGAN. Since some methods could not achieve the denoising function, we combined with the BM3D method for denoising to produce the final result. The quantitative results are shown in Table 4, qualitative results are shown in Fig. 7.

By comparing the subjective visual effects of each algorithm in Fig. 7, it can be seen that the overall brightness of the enhanced image is moderate, and the visual effect is better and more natural. In contrast, the color difference between the HE method is too bright, and the contrast is oversaturated. The whole image of the SRIE method is green, and there is more color distortion. The EnlightenGAN method generates too much noise, making the generated image too blurry. The DSLR image's brightness is darker than that in our method. The experimental results show that our method is effective in low illumination images. In Table 4, our method is superior to other methods in two indexes, which shows the superiority of this method. Compared with the traditional methods of HE and SRIE, we improved 43.08% and 34.61% on PSNR and 69.56% and 43.06% on SSIM, and On SSIM-GC, it increased by 34.8% and 23.39% respectively. Compared with the current deep learning method DSLR and EnlightenGAN, we also improved by 0.20% and 1.3% on PSNR, we also improved by 0.29% and 23.23% on SSIM, and On SSIM-GC, it increased by 20.67% and 1.23% respectively. Compared with the result in the LOL dataset, our algorithm decreased by 97.9% on PNSR. The main reason is that the DSLR method is especially proposed for DPED data, so it performs better, which leads to a relative reduction in the promotion range. It can be seen that our method is much more effective than the traditional method in the DPED data set, and the deeper learning method also has certain advantages.

## 5 Conclusion

In this paper, we propose to use the enhanced network module to optimize Generative Adversarial Networks for low illumination image enhancement, aiming at tackling the problems of low illumination image, such as low noise and low brightness. The enhanced network uses the residual connection to build the residual module to deepen the network, which improves the ability to model the image enhancement with low illumination. The method in this paper verifies the performance of the algorithm on two image dataset (DPED, LOL), and compares with the traditional image enhancement methods (HE, SRIE) and the deep learning methods (EnlightenGAN, DSLR), the experimental results show that the effectiveness of the proposed algorithm.



**Acknowledgements** This work is funded by the National Natural Science Foundation of China under Grant No.61772180 & No. 61701173, Technological innovation project of Hubei Province 2019(2019AAA047), Green Industry Science and Technology Leadership Program of Hubei University of Technology(No.CPYF2018005) and Hubei province graduate education innovation plan.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Abadi M, Barham P, Chen J, et al. (2016) Tensorflow: A system for large-scale machine learning[C]. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) 265–283
2. bdullah-Al-Wadud M, Kabir MH, Dewan MAA et al (2007) A dynamic histogram equalization for image contrast enhancement[J]. *IEEE Trans Consumer Electron* 53(2):593–600
3. Cai J, Gu S, Zhang L (2018) Learning a deep single image contrast enhancer from multi-exposure images[J]. *IEEE Trans Image Process* 27(4):2049–2062
4. Chen C, Chen Q, Xu J et al (2018) Learning to see in the dark[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3291–3300
5. Chen C, Chen Q, Xu J et al (2018) Learning to see in the dark[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3291–3300
6. Chen X, Duan Y, Houthoof R et al (2016) Infogan: Interpretable representation learning by information maximizing generative adversarial nets[C]. *Adv Neural Inform Process Syst* 2172–2180
7. Chen J, Ying H, Liu X, Gu J, Feng R, Chen T, Gao H, Wu J (2020) A transfer learning based Super-Resolution microscopy for biopsy slice images: The joint methods perspective *IEEE/ACM transactions on computational biology and bioinformatics (TCBB)* <https://doi.org/10.1109/TCBB.2020.2991173>
8. Dabov K, Foi A, Katkovnik V et al (2007) Image denoising by sparse 3-D transform-domain collaborative filtering[J]. *IEEE Trans Image Process* 16(8):2080–2095
9. Denton EL, Chintala S, Fergus R (2015) Deep generative image models using a laplacian pyramid of adversarial networks[C]. *Adv Neural Inform Process Syst* 1486–1494
10. Fu Q, Jung C, Xu K (2018) Retinex-based perceptual contrast enhancement in images using luminance adaptation[J]. *IEEE Access* 6:61277–61286
11. Fu X, Zeng D, Huang Y et al (2016) A weighted variational model for simultaneous reflectance and illumination estimation[C]. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2782–2790
12. Gao H, Xu Y, Yin Y, Zhang W, Li R, Wang X (2020) Context-aware QoS prediction with neural collaborative filtering for internet-of-things services. *IEEE Int Things J (IoT-J)* 7(5):4532–4542
13. Gao W, Zhu Y, Zhang W, Zhang K, Gao H (2019) A hierarchical recurrent approach to predict scene graphs from a visualion-oriented perspective. *Comput Intell (COIN)* 35(3):496–516
14. Goodfellow IJ, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets[C]. In: Proceedings of the 27th international conference on neural information processing systems, Montreal, Dec 8–13, 2014. MIT Press, Cambridge, pp 2672–2680
15. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior[J]. *IEEE Trans Pattern Anal Machine Intell* 33(12):2341–2353
16. He K, Zhang X, Ren S et al (2016) Identity mappings in deep residual networks[C]. In: European conference on computer vision. Springer, Cham, pp 630–645
17. Ignatov A, Kobyshev N, Timofte R et al (2017) DSLR-quality photos on mobile devices with deep convolutional networks[C]. In: Proceedings of the IEEE International Conference on Computer Vision, pp 3277–3285
18. Ioffe S, Szegedy C. (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. *arXiv:1502.03167*
19. Jiang Y, Gong X, Liu D et al (2019) EnlightenGAN: Deep light enhancement without paired supervision[J]
20. Jobson DJ, Rahman Z, Woodell GA (1997) Properties and performance of a center/surround retinex[J]. *IEEE Trans Image Process* 6(3):451–462

21. Jobson DJ, Rahman Z, Woodell GA (1997) A multiscale retinex for bridging the gap between color images and the human observation of scenes[J]. *IEEE Trans Image Process* 6(7):965–976
22. Kingma DP, Ba JL (2014) Adam optimizer[J]. arXiv:1412.6980 1-15
23. Lan M., Zhang Y., Zhang L. (2020) Global context based automatic road segmentation via dilated convolutional neural network. *B. D. u. Inform Sci* 535:156–171
24. Land EH, McCann JJ (1971) Lightness and retinex theory[J]. *Josa* 61(1):1–11
25. Li ZQ (2015) Liu Q.Low-light image adaptive enhancement algorithm based on wavelet transform[J]. *Chinese J Lasers* 42(02):280–286
26. Li C, Fan T, Ma X et al (2017) An improved image defogging method based on dark channel prior[C]. In: 2017 2nd international conference on image, vision and computing (ICIVC) IEEE, pp 414–417
27. Lore KG, Akintayo A, Sarkar S (2017) LLNet: A deep autoencoder approach to natural low-light image enhancement[J]. *Pattern Recognit* 61:650–662
28. Lv F, Lu F, Wu J et al (2018) MBLEN: Low-Light Image/Video Enhancement Using CNNs[C]/BMVC 220
29. Mirza M, Osindero S. (2014) Conditional generative adversarial nets[J]. arXiv:1411.1784
30. Radford A, Metz L, Chintala S. (2015) Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv:1511.06434
31. Shen L, Yue Z, Feng F et al (2017) Msr-net: Low-light image enhancement using deep convolutional network[J]. arXiv:1711.02488
32. Simonyan K, Zisserman A. (2014) Very deep convolutional networks for large-scale image recognition[J]. arXiv:1409.1556
33. Song L, Wang C, Zhang L, Du B, Zhang Q, Huang C, Wang X (2020) Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognit* 102:1–11
34. Szegedy C, Ioffe S, Vanhoucke V et al (2016) Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[C]. *National Conference on Artificial Intelligence* 4278–4284
35. Wei C, Wang W, Yang W et al (2018) Deep retinex decomposition for low-light enhancement[J]. arXiv:1808.04560
36. Yuan L, Sun J (2012) Automatic exposure correction of consumer photographs[C]. In: *European conference on computer vision*. Springer, Berlin, pp 771–785
37. Zhang Y, Zhang J, Guo X (2019) Kindling the Darkness: A Practical Low-light Image Enhancer[C]. In: *Proceedings of the 27th ACM International Conference on Multimedia*, pp 1632–1640
38. Zhang Y, Zhang J, Guo X (2019) Kindling the Darkness: A Practical low-light image enhancer[C]. In: *Proceedings of the 27th ACM international conference on multimedia*, pp 1632–1640

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Lingyu Yan<sup>1</sup>  · Jiarun Fu<sup>1</sup> · Chunzhi Wang<sup>1</sup> · Zhiwei Ye<sup>1</sup> · Hongwei Chen<sup>1</sup> · Hefei Ling<sup>2</sup>

Jiarun Fu  
jiarunfu@163.com

Chunzhi Wang  
chunzhiwang@163.com

Zhiwei Ye  
weizhiye@163.com

Hongwei Chen  
chw@sina.com

Hefei Ling  
hefei.ling@hust.edu.cn

<sup>1</sup> School of Computer Science, Hubei University of Technology, Wuhan, China

<sup>2</sup> School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China