Low-light image enhancement base on brightness attention mechanism generative adversarial networks

Lingyu Yan (yanlingyu@hbut.edu.cn)
Hubei University of Technology

Jiarun Fu
Hubei University of Technology

Yulin Peng
Hunan ProvinciaMaternal and Child Health Care Hospital

KunPeng Zheng
Hubei University of Technology

Rong Gao
Hubei University of Technology

HeFei Ling
Huazhong University of Science and Technology

Research Article

Keywords: Generative Adversarial Networks, low-light image enhancement, attention mechanism

Posted Date: April 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-137966/v2

Low-light image enhancement base on brightness attention mechanism generative adversarial networks

Jiarun Fu[1], Lingyu Yan[1*], Yulin Peng[3], KunPeng Zheng[1], Rong Gao[1] and HeFei Ling[2]

[1]School of computer science, Hubei University of technology.
[2]School of computer science, Huazhong University of Science and Technology.
[3]Hunan ProvinciaMaternal and Child Health Care Hospital Changsha, China.

*Corresponding author(s). E-mail(s): yanlingyu@hbut.edu.cn;

Abstract

With the development of the field of deep learning, image recognition, enhancement and other technologies have been widely used.However, dark lighting environments in reality, such as insufficient light at night, cause or block photographic images in low brightness, severe noise, and a large number of details are lost, resulting in a huge loss of image content and information, which hinders further analysis and use. Such problems not only exist in the traditional deep learning field, but also exist in criminal investigation, scientific photography and other fields, such as the accuracy of low-light image. However, in the current research results, there is no perfect means to deal with the above problems. Therefore, the study of low-light image enhancement has important theoretical significance and practical application value for the development of smart cities. In order to improve the quality of low-light enhanced images, this paper tries to introduce the luminance attention mechanism to improve the enhancement efficiency. The main contents of this paper are summarized as follows: using the attention mechanism, we proposed a method of low-light image enhancement based on the brightness attention mechanism and generative adversarial networks . This method uses brightness attention mechanism to predict

the illumination distribution of low-light image and guides the enhancement network to enhance the image adaptiveness in different luminance regions. At the same time, u-NET network is designed and constructed to improve the modeling process of low-light image. We verified the performance of the algorithm on the synthetic data set and compared it with traditional image enhancement methods (HE, SRIE) and deep learning methods (DSLR). The experimental results show that our proposed network model has relatively good enhancement quality for low-light images, and improves the overall robustness, which has practical significance for solving the problem of low-light image enhancement..

Keywords: Generative Adversarial Networks, low-light image enhancement, attention mechanism

1 Introduction

The Smart City concept mainly refers to the use of various information technologies or innovative concepts to improve the efficiency of resource utilization, optimize urban management and services, and improve the quality of life of citizens.

Image recognition technology plays an important role in this process, because the image contains rich and detailed information of the real scene. By capturing and processing image data, intelligent systems can be developed to perform various tasks such as object detection, classification, segmentation, recognition, scene understanding and 3D reconstructio[1], which can be used for the construction of Smart City.

As shown in Figure 1, such image recognition technology can be applied to automatic driving, video surveillance and virtual augmented reality in smart cities[2].

However, in practical application, the accuracy of the recognition system depends heavily on the quality of the input image. In particular, images were taken in the low-light environment prevalent in smart cities usually suffer from severe degradation, such as poor visibility, low contrast, and unexpected noise [3] (as shown in Figure 2). Therefore, in order to improve the performance of the recognition algorithm widely in the smart city, and because of the limitation of hardware equipment, a special enhancement algorithm for low-light images is needed to solve the problem.

At present, image enhancement technology has been applied in many fields, such as mobile phone shooting, criminal investigation, medical image, remote sensing image, HDTV, digital camera, etc.[4] . In terms of mobile phone shooting, brands such as Xiaomi, Huawei, OPPO, etc. all use the low-light image enhancement function to take good photos in case of insufficient light or night shooting, such as Figure 3.

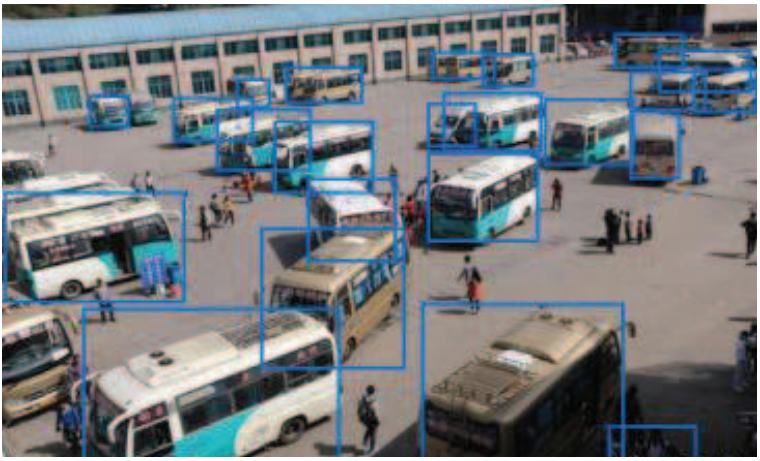

Fig. 1 Smart city image recognition application

Fig. 2 Low light environment photo instance

In the field of the criminal investigation, when a criminal incident occurs, the public security department will first check the surveillance video to identify the criminal suspect. However, when installing monitoring devices, to capture as many scenes as possible, monitoring devices are usually set in wide-angle mode, so the resolution for a single object is often low. Moreover, since images are vulnerable to the influence of weather and light, the phenomenon of uneven illumination, low contrast, blurry and noise will appear in the surveillance video images [5] . This makes it difficult for the public security department to identify the face of the criminal suspect, as shown in Figure 4, thus affecting the detection of the case. The low-light image enhancement technology can process the face images taken by the surveillance equipment to get a clearer face image, which can help the police to accelerate the speed

Fig. 3 Example of image enhancement for mobile phone photography

of case detection to a certain extent. After decades of development, the low

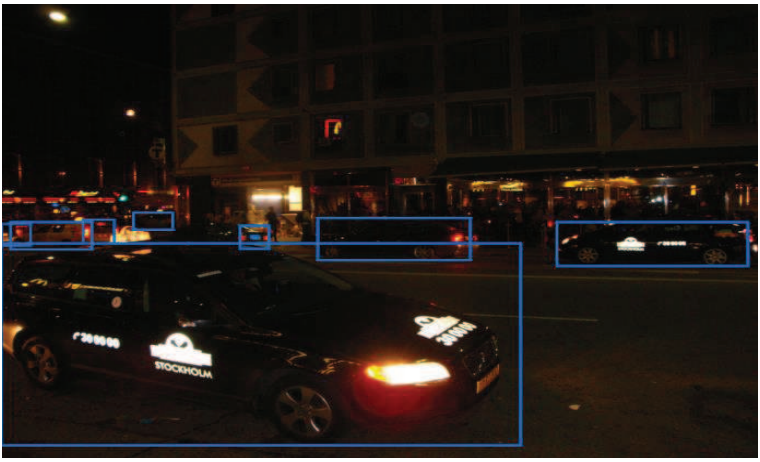

Fig. 4 Low light image dilemma in the field of criminal investigation

light image enhancement has become an important image processing research direction, and their final research goal is to realize their low-quality images, strengthen details, improve contrast, reduce noise, thus enriching the content of image information, improve the quality of the overall awareness, eventually to achieve the image that meets the high-level vision tasks. However, because of the requirements of smart cities, although the corresponding low-illumination enhancement algorithm has been used to restore the image, the quality and details of the enhanced image need to be further improved. In addition, the urban environment is complex and diverse, and the existing algorithms are not

robust enough to adapt to low-light images under different lighting environments. Finally, the excellent response speed is the core factor in the concept of the smart city. Considering computing constraints and other factors, the computational complexity and time consumption of the current algorithm still need to be reduced. With the rapid development of artificial intelligence, more and more researchers begin to try to use new computer vision technologies, such as deep learning to solve image enhancement tasks [6]. It has powerful representativeness and can be used to model more complex and diverse low-illumination enhancement problems.

To sum up, image recognition technology plays an important role in smart cities, and the accuracy of image recognition is seriously affected by the existing low-light environment. Therefore, the research on algorithms related to low-light image enhancement is of great scientific significance and practical application value for the development of smart cities. In view of the above considerations, this paper sets out to study the efficient low-light image enhancement algorithm, which is expected to make up for the deficiency of imaging equipment in hardware, to improve the sharpness of photos taken in a low-light environment and comprehensively improve the image quality.

This paper, starting from the Generative Adversarial Networks(GAN), combines with the attention mechanism and is applied to the enhancement of low-light images. It mainly involves the application research of the brightness attention mechanism in low-light image enhancement, compares its traditional method with the existing deep learning algorithm, and finally carries out experimental verification. The main contributions are as follows:

In order to solve the problem of uneven illumination and noise in low-light images, a method of image enhancement based on the brightness attention mechanism Generative Adversarial Networks was proposed. By learning the mapping between low-light images and normal light images, the images under enhanced normal illumination environment in the corresponding scene were obtained. To restore the low-light image, especially its low-light region information, a generator was constructed by combining the attention mechanism and U-NET concept. Based on the backbone network structure, a branch network of brightness attention was added to strengthen the transmission of light distribution information in the network flow. To enhance the recognition effect of the low-light image.

In this paper, the background of the research is described in chapter 2, the principle of the algorithm is explained in Chapter 3, the experimental results are proved in Chapter 4, and the conclusions obtained are illustrated in Chapter 5.

2 Research background

The purpose of this chapter is to introduce the relevant research background, including the generative adversarial networks, low-illumination image enhancement, and the research background of attention mechanism in recent years, so as to provide reference for our experiment.

2.1 Development status of the generative adversarial networks

Since the original GAN[7] was proposed, many gan-based variants have been proposed in subsequent studies to solve the training instability of GAN, generate sample richness and apply it in unsupervised learning. Against the original GAN for generator almost without any constraint, the generation process is too free, cause it is difficult to control in case of large image model, CGAN (Conditional GAN) [8] on the basis of the original GAN respectively in the generator and the discriminant of input one more constraint y, as shown in figure 5, the network generated in the direction of the given sample, CGAN objective function into a formula 1:

$$\min_G \min_D V(D,G) = E_{x:P_x}[log_2(1 - D(G(z\mathrm{y})))]+\mathrm{E\_–Z:P\_Z''[log\_2(1\text{-}D(G(zy)))]} \tag{1}$$

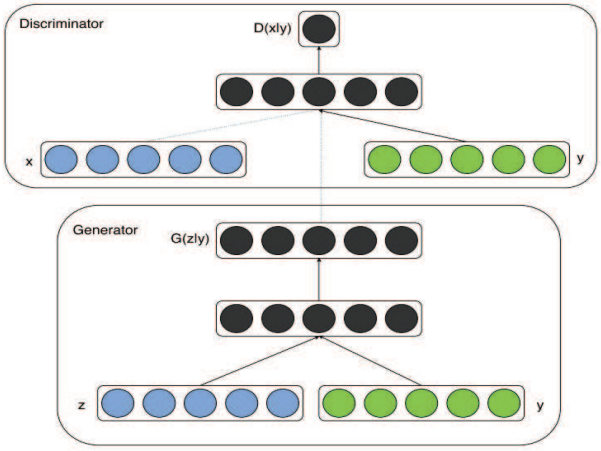

Fig. 5 Network structure diagram of CGAN

The disadvantage of CGAN is that its model training is unstable. As can be seen from the loss function, CGAN only adds additional constraints to generate the specified image, but does not solve the problem of training instability.

LAPGAN[9] makes improvement based on CGAN, and then seeks to improve the effect of image enhancement. LAPGAN combined the concepts of Gaussian gold pyramid and Laplace pyramid with GAN, and then used The

Gaussian pyramid for image down-sampling, and The Laplace Gold pyramid for up-sampling to reconstruct the image. The specific process is shown in Figure 6.

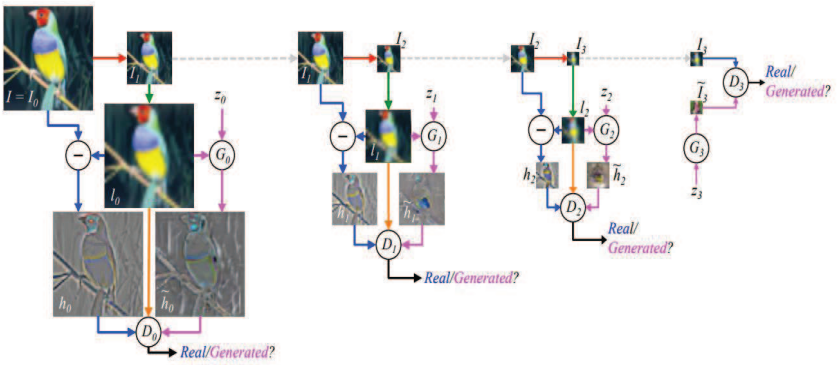

Fig. 6 Network structure diagram of LAPGAN

DCGAN[10] combined the convolutional neural network (CNN) with GAN, as shown in Figure 7, which improved the diversity of the generated images. DCGAN uses batch normalization (BN)[11] for stable training, and applies ReLU activation function to reduce the risk of gradient disappearance. At the same time, the pooling layer is eliminated and the characteristic information is retained effectively by using step convolution and micro-step convolution. DCGAN also has many problems. For example, although it can generate rich diversity, the image quality is not optimal, and there is also the problem of unstable training.

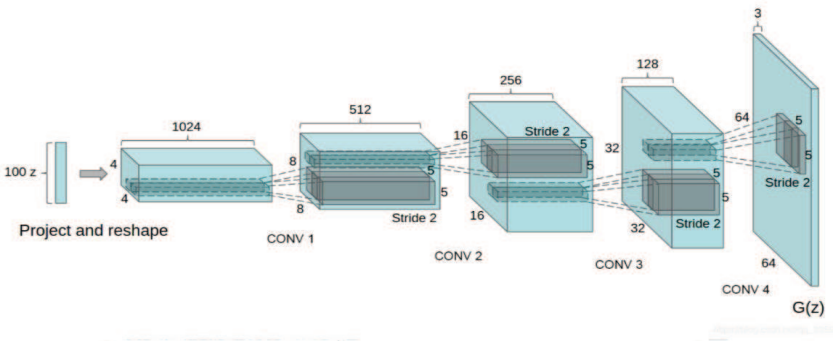

Fig. 7 Network structure diagram of DCGAN

SA-GAN [12] introduced the Attention mechanism in GAN, which enables the generator and discriminator to automatically learn important targets in the image, and enables us to know which parts of the image should be learned from the task focus (similar to significance detection), in order to submit the image quality of the generated image. Compared to Baseline, SA-Gan raised the Inception Score from 36.8 to 52.52 on the ImageNet dataset and reduced Frechet Inception Distance from 27.62 to 18.65. This network construction combining attention mechanism and GAN has great reference value for our experiment. As shown in the figure above, f(x), g(x), and H (x) are all ordinary

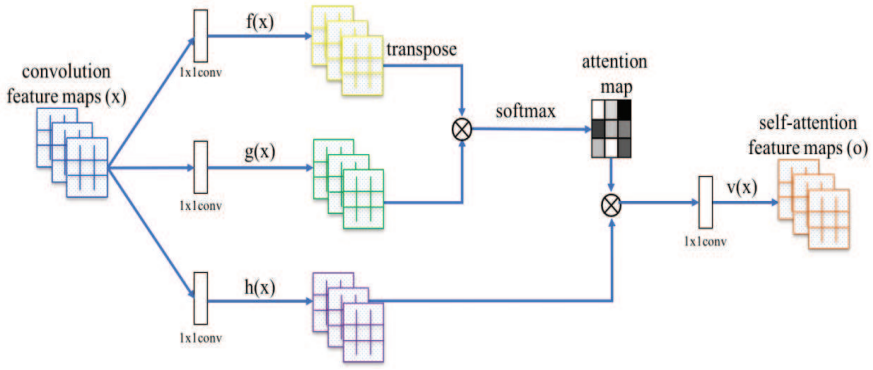

Fig. 8 Self-attention module network structure diagram

1x1 convolution, the only difference being the size of the output channel; Take the output of F (x) and multiply it by the output of g(x). After softmax normalization, you get an attention Map. Multiply the obtained Attention Map and H (x) pixel by pixel to obtain adaptive Attention feature maps.

2.2 Development status of low-light image enhancement

In the past two years, low illumination image enhancement mainly focuses on the realization of algorithm structure by using CNN.

In a paper published in 2018 [13], the convolutional neural network was used to complete Raw image processing to RGB image processing to solve the problems in image imaging under extremely low light and short exposure conditions. The results were stunning. The network structure is based on the fully convolutional network FCN, and the efficiency of the algorithm is greatly improved through end-to-end training.

As shown in figure 9, for the original Bayer image, they split the input into four channels, reducing the spatial resolution by half on each dimension; The output is a 12-channel image with a 1/2 spatial resolution; It is then processed by the sub-pixel layer to restore the original resolution.

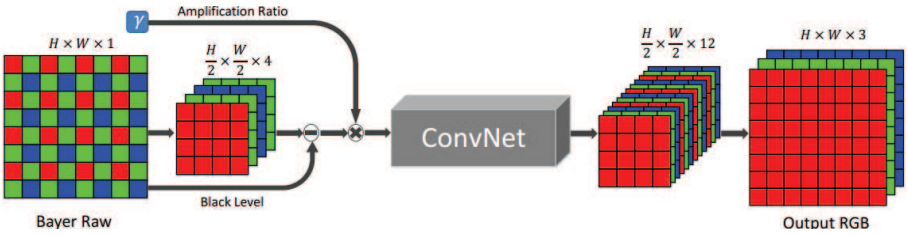

Fig. 9 The structure diagram of FCN

The literature [14] published in 2019 puts forward three problems existing in low-light enhancement tasks :(1) how to estimate illumination map components from a single image and adjust illumination level flexibly? (2) After increasing image brightness, how to remove degradation such as noise and color distortion? (3) In the absence of ground-truth, how to train the model with a limited number of samples? The network built in the article is divided into three modules: decomposing image, restoring reflection image, and adjusting the light image, Restoration and Adjustment Net respectively. For the three problems mentioned above, the solution is as follows (a) for the Decomposition-Net, for the region of the lightmap smoothness and mutual consistency, also added two new loss function. (b) For Restoration-Net, considering the degradation effect of the reflection map under low light, we used the reflection map under good light as a reference. The illumination map information is introduced to solve the complex problem of the distribution of the degradation effect in the reflection map (c) For Adjustment Net, a mechanism is realized to continuously adjust the illumination intensity (the enhancement ratio is used as the input after the combination of the feature map and the illumination map).

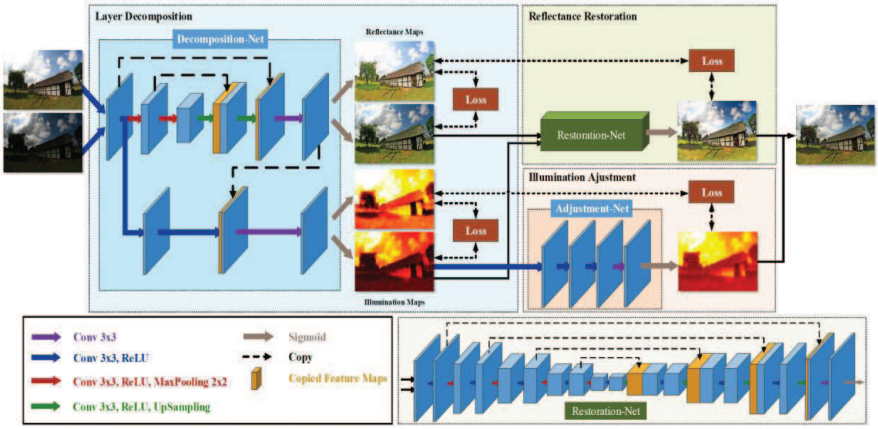

Fig. 10 The structure diagram of KinD network

As shown in Figure 10, the KinD network mentioned in the literature consists of two branches, corresponding to the reflection diagram and the lighting

diagram. From the perspective of function, it can be divided into three modules, namely layer decomposition, reflection map restoration and lighting map adjustment.

2.3 Development status of low-light image enhancement

The visual attention mechanism is a brain signal processing mechanism unique to human vision. Human vision can quickly scan the global image to obtain the target area that needs to be paid attention to, and then devote more attention resources to this area, to obtain more detailed information about the target that needs to be paid attention to and suppress other useless information. This is a means for human beings to quickly screen out high-value information from a large amount of information with limited attention resources. It is a survival mechanism formed in the long-term evolution of human beings. The human visual attention mechanism greatly improves the efficiency and accuracy of visual information processing. The attention mechanism in deep learning is similar to the selective visual attention mechanism of human beings in essence, and the target is also to select more critical information for the current task target from a lot of information.

The mechanisms of visual attention can be divided into two categories: soft attention and hard attention. The key point of soft attention is to pay more attention to the region or channel, and have certainty, after the completion of learning can be generated directly through the network. The crucial point is that soft attention is differentiable, which means that it can compute the gradient through a neural network and learn to get the weight of attention by propagating forward and feedback backward. The difference between strong attention and soft attention is more focused, that is, every point in the image is likely to extend attention. It is also a random prediction process, with more emphasis on dynamic changes. Contrary to soft attention, strong attention is not reinforcement and the training process is often completed by reinforcement learning.

This paper focuses on three types of attention domains under soft attention: spatial domain, channel domain, and mixed domain.

The spatial domain transforms the spatial information in the original image into another space and retains the key information. The pooling layer in the common convolutional neural network directly uses some methods of Max pooling or average pooling to compress the image information, reduce the amount of calculation, and improve the accuracy.

To solve the problem that key information cannot be recognized when information is merged directly, researchers have proposed a module called Spatial Transformer, which transforms the spatial domain information in the picture into corresponding spatial transformation to extract the key information [15].

An example of the transformation process shown in Figure 11:

Column (a) is the original picture information, in which the first handwritten number 7 does not make any transformation, the second handwritten

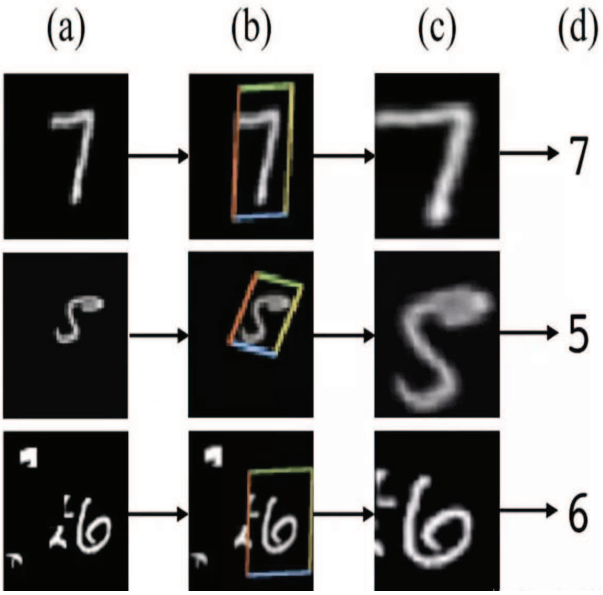

Fig. 11 Spatial Transformer application example diagram

number 5 makes a certain rotation change, and the third handwritten number 6 adds some noise signals;

Bounding The colored bounding boxes in B are spatial Transformer learned, and each bounding box is actually a spatial transformer learned from corresponding pictures.

In Column C is the feature map after spatial Transformer conversion, it can be seen that the key area of number 7 is selected, the handwritten number 5 is rotated into a positive picture, and the noise information of number 6 is not recognized.

The realization of spatial Transformer operation can be regarded as the realization of attention mechanism because after training, spatial Transformer can find out the areas that need to be paid attention to in the picture information. At the same time, the transformer has the functions of rotation and scaling transformation, which enables the important information of the picture to be extracted by the box through transformation.

In a traditional convolution neural network, every picture (R, G, and B) by linking the initial three-channel said to come out, after different convolution kernels, each channel will generate a new signal, such as image features of each channel using the 64 nuclear convolutions, can produce 64 new channel matrix (H, W, 64), H, W, respectively characteristics the height and width of the picture.

Similarly, attention mechanisms require similar channels to assign attention weights to images.

Above is a schematic of the channel attention module [16].

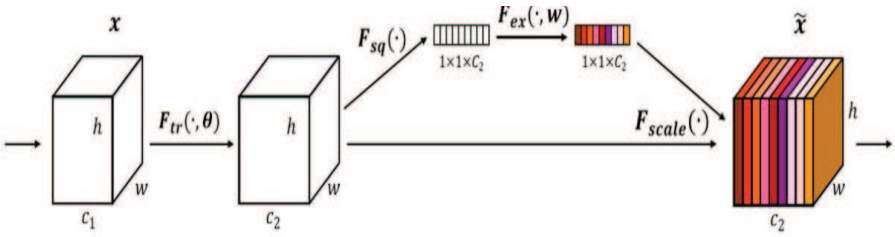

Fig. 12 Conceptual diagram of attention mechanism

As shown in the figure, given an input x, the characteristic channel number is C_1, and a character with the characteristic channel number is C_2 after a series of convolution and other general transformations. Corresponding to the traditional CNN, three operations were finally used to re-calibrate the previously obtained features.

The first is the Squeeze operation, which compresses the features along the spatial dimensions and turns each two-dimensional feature channel into a real number, which has a global receptive field, and the output dimension matches the input number of feature channels. It represents the global distribution of responses on the feature channel and enables the layer near the input to obtain the global receptive field.

The second is the Gate operation, which is similar to the portal mechanism for cyclic neural networks. The weight of each feature channel is generated by the parameter W to show the correlation between the modeled feature channels.

The final is a Reweight operation. The output weights are used for the importance of each feature channel after feature selection. The re-calibration of the original feature on the channel dimension is done using the per-channel multiplication weighting to the previous features.

2.4 U-net

In 2015, Olaf Ronneberger, Philipp Fischer and Thomas Brox proposed the u-NET network structure [17]. U-net is based on the expansion and modification of full convolutional network. It is mainly composed of two parts: contracting Path for obtaining context information, and expanding path for precise positioning. As shown in Figure 13, in the structure of U-NET, the left side is a downsampling process, which is divided into 4 groups of convolution operations (blue arrow). After each set of convolution operations, the MaxPool operation (red arrow) is performed to reduce the image further to 1/21/2. Through four groups of operations, the input image of size 572x572x1572x572x1 was transformed into 32x32x1024.

The upper sampling process is shown on the right. The upsampling process uses four sets of deconvolution (light green arrows), each upsampling expands the image 22 times, then clipping and copying the image (feature map) of the corresponding layer, and is concat to the result of the convolution (gray arrow).

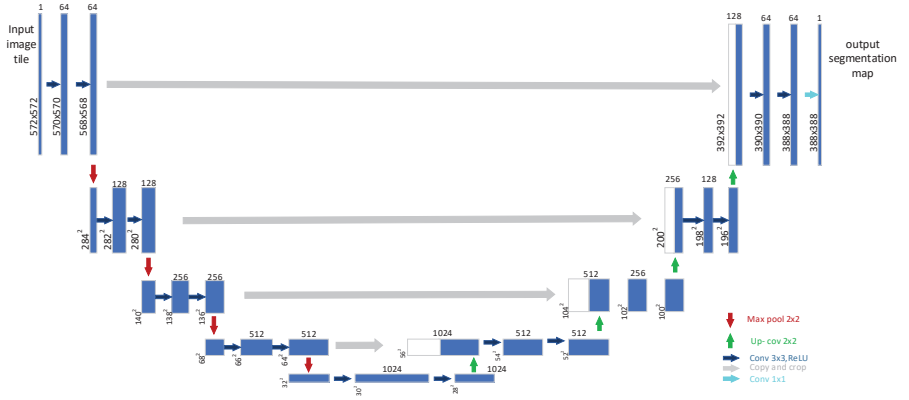

Fig. 13 Standard U-net structure diagram

After the upsampling process, a graph of 388x388x64 size is obtained. Finally, a convolution kernel of 1x11x1 is used to reduce the number of channels to 2 (dark green arrow), namely the two colors on the label.

It can be found that the feature of U-NET structure design is the fusion of low-level features and high-level features, which can make full use of features of all levels of the image, and the constructed feature vector can describe the region more accurately.

3 Research background

3.1 Algorithm Framework

3.1.1 Network structure

The network structure of low-light image enhancement is shown in figures 14 and 15. In order to enhance the transmission of low-light location information in the network flow, the generator has an attentional branch network proposed in this paper. The attention-net predicts the position mask of the weak light, which is the same size as the input image, with each pixel being a probability value between 0 and 1. Finally, we combine the input image, the attention diagram, and the converted input to form the final enhanced image. The discriminator can receive the image generated by the generator and the real image at the same time, and finally, produce the predicted value of true or false.

The details of the network structure of the attention network are shown in Table 1. It is a fully convolutional network structure similar to U-NET. The design feature of this network structure is to fuse the features of the low layer and the high layer and make full use of the features of all levels of the image. The network consists of three parts: a contraction path to obtain multi-scale context information, an asymmetric expansion path to restore multi-stage feature map, and the last two convolutional layers to restore the attention diagram with the same size as the input. The contraction path has four lower sampling

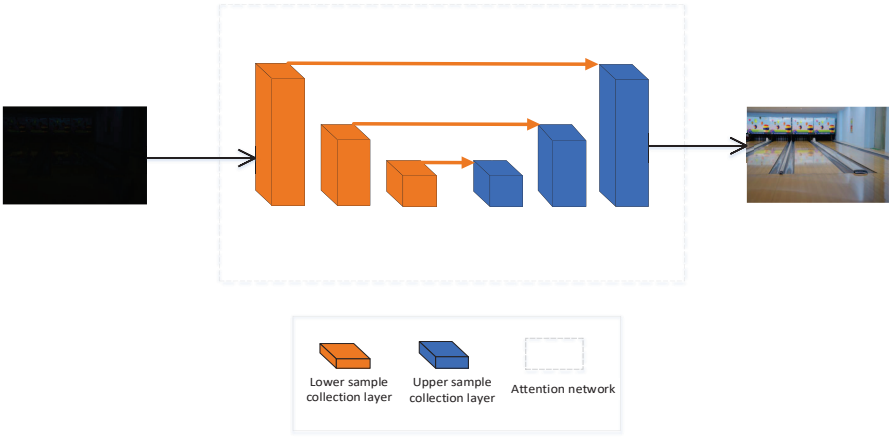

Fig. 14 Schematic diagram of generator network structure

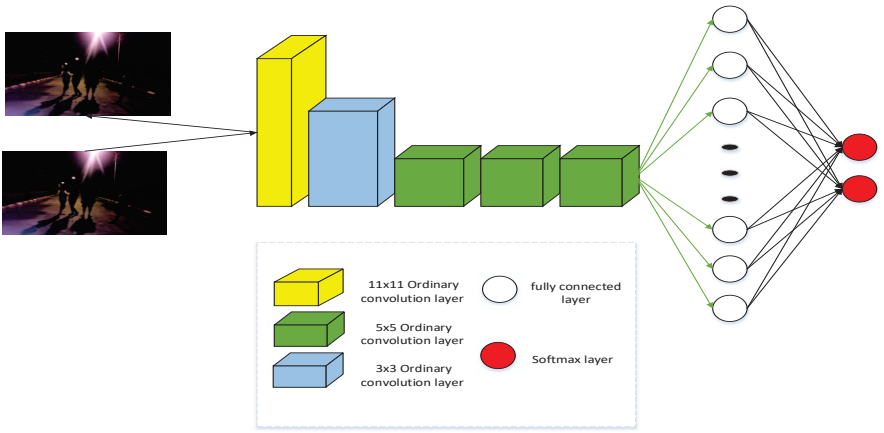

Fig. 15 Schematic diagram of discriminator network structure

convolution blocks, each of which contains two convolutional layers with step size 1 and one "pooling layer" with step size 2. The expansion path has four upsampled deconvolution blocks, and each deconvolution block contains two layers of convolution layer with step size 1, one deconvolution layer, and one layer feature splicing. And then the last 2 convolution layers, one deconvolution layer, one ordinary convolution, but the activation function of the last convolution layer is tanh. All the convolution kernels have a size of 3x3. Except for the last layer, the convolution is activated by the lrelu function after the convolution.

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

Table 1 Structural details of the attention network

Convolution layer	Input	convolution kernel/perstep	output
Conv0	I(100 x 100 x 3)	3 x 3 x 16/2	50 x 50 x 16
Conv1	Conv0	3 x 3 x 16 / 1,3 x 3 x 16 / 2	25 x 25 x 16
Conv2	Conv1	3 x 3 x 32 / 1,3 x 3 x 32 / 2	13 x 13 x 32
Conv3	Conv2	3 x 3 x 64 / 1,3 x 3 x 64 / 2	7 x 7 x 64
Conv4	Conv3	3 x 3 x 128 / 1,3 x 3 x 128 / 2	4 x 4 x 128
Conv5	Conv4	3 x 3 x 256 / 1	4 x 4 x 256
DeConv0	Conv5,Conv4	3 x 3 x 128 / 1	7 x 7 x 128
DeConv1	DConv0,Conv3	3 x 3 x 64 / 1	13 x 13 x 64
DeConv2	DConv1,Conv2	3 x 3 x 32 / 1	25 x 25 x 32
DeConv3	DConv2,Conv1	3 x 3 x 16 / 1	50 x 50 x 16
DeConv4	DConv3,Conv0	2 x 2 x 3 / 2	100 x 100 x 3
Conv6	DeConv4	3 x 3 x 1 / 1	100 x 100 x 1

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($batch, P_{ture}, P_{false}$),$P_t rue$,$P_f alse$.They're all in the range of[0,1].

Table 2 Structural details of discriminator network

Convolution layer	Input	convolution kernel/perstep	output
Conv0	I(100 x 100 x 3)	11 x 11 x 48/4	25 x 25 x 48
Conv1	Conv0	5 x 5 x 128/2	13 x 13 x 128
Conv2	Conv1	3 x 3 x 192/1	13 x 13 x 19
Conv3	Conv2	3 x 3 x 192/1	13 x 13 x 192
Conv4	Conv3	3 x 3 x 128/2	7 x 7 x 128
Fc	conv4	6272 x 1024	batch x 1024
softmax	Fc	1024x2	batch x 2

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 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:

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

$L, L_adv, L_con, L_tv, L_col$represent attention loss, confrontation loss, content loss, total variation loss, and color loss,$W_a, W_adv, W_cov, W_tv, W_col$,represent the weight of their losses respectively. 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, so as to ensure that the generated picture has the least loss in all aspects compared with the real picture, and the generated effect is realistic.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. **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

M low-light image pairs were randomly selected in the data set $(I_x^1), (I_x^2, I_y^2)...(I_x^m, I_y^m)$.

The input of the fixed discriminant network isI_adv=0,0,...0,length of m

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

The input of the randomly initialized discriminant network isI_adv=1,0...0,and m stands for length.

M low-light image pairs were randomly selected in the data set $(I_x^1), (I_x^2, I_y^2)...(I_x^m, I_y^m)$.

Maximizes the overall loss of the discriminator network:

$$L_adv = \sum_{i=1}^{n} logD(G(I_x), I_y)$$

until Maximum number of iterations do

end

3.2 Brightness attention mechanism module

To solve the uneven distribution of brightness in low-illumination images, traditional image enhancement methods generally enhance the image as a whole, but ignore the inconsistency of brightness in each area of the image, which will easily lead to overexposure in high-brightness areas and underexposure in low-brightness areas. In this paper, a u-NET-like attention-branching network is designed to predict the distribution of low-light regions in low-light images, so as to promote the network to pay more attention to the low-light regions in images.

We take the brightness attention branch network as an auxiliary network, and combine the brightness attention diagram obtained by training with the output of the main network to enhance the enhancement effect of the low-light region in the low-light image. Figure 15 shows the brightness attention branch network. The white area of the output image represents the low illumination area of the input image, and the black area represents the brighter area of the input image. From the input image to the bottom of U-NET is the under-sampling process of illumination information, from the bottom to the right of U-NET is the fusion of multi-scale illumination information, so as to fully model the illumination information in low-illumination images, and finally generate the luminance attention diagram.

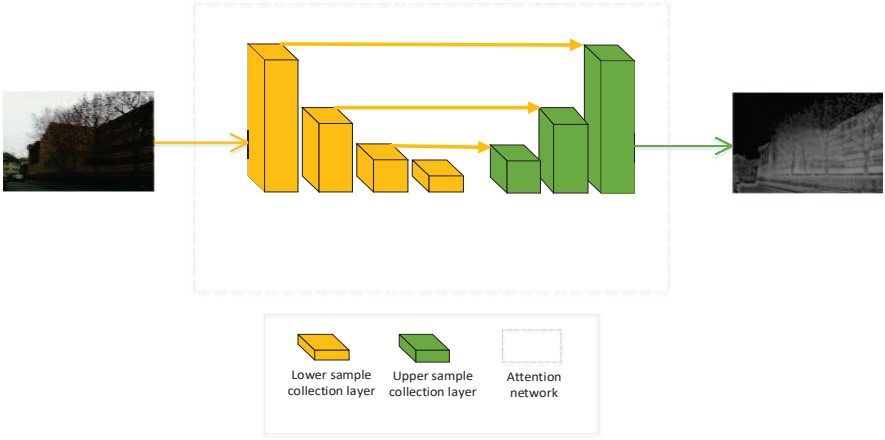

Fig. 16 Schematic diagram of discriminator network structure

In order to better constrain the module's modeling of light distribution, the following attention loss function is used in this paper:

$$L_a = F_a(I_x) - A \tag{3}$$

$I_x$represents the input image,Arepresents the expected brightness attention diagram,$F_a(I_x)$ and represents the predicted brightness attention. The expected luminance attention diagram A is calculated by formula 4 below.

$$A = \frac{(max_c(I_y) - max_c(I_x))}{max_c(I_y)} \tag{4}$$

Where I_x represents low-light image, I_y represents true-light image, and $max_C()$ represents the maximum pixel value on the image channel.

4 Experiment

4.1 Experimental data and pre-processing

As for the selection of experiment images, we chose a HIGH-quality image data set with a resolution of DIVerse 2K (DIV2K), which included a training set with 800 pictures and a verification set with 100 pictures. This dataset has been used in NTIRE [17, 18] and @PIRm [19]competitions [20]. In order to test the clear results of the algorithm for low illumination image recognition accuracy, this experiment preprocessed div2K with low illumination, that is, used it to synthesize low light image data set.

Low-light images have two significant characteristics: low brightness and noise. In the pre-processing process, gamma correction and random parameters are used to adjust the image's low brightness. The formula is as follows:

$$I_L = rand * (I_H)^y \tag{5}$$

Where, I_L is the low light image, $rand$ is the random number between (0,1), $I_H$is the high-resolution image, γ is the gamma coefficient, obeying the uniform distribution between [1.1,2]. Considering the noise caused by low light, we also added gaussian noise with uniform distribution of variance [0.01,0.05] into the images treated with low light. After pre-processing, we made a training set with 30,744 images and a test set with 1080 images on the DIV2K data set, among which the size of the images is 100*100, as shown in Figure 17,18.

4.2 Experimental setting

To verify the performance of the image enhancement algorithm proposed in this paper, the experiment was compared with the following representative image enhancement algorithms: histogram equalization (HE)[21], reflected illumination estimation (SRIE)[22], and deep learning method DSLR[23].

Among them, HE and SIRE are traditional low-light image enhancement methods, and the advantages and disadvantages of deep learning and traditional methods can be obtained by comparing them. By comparing the results with DSLR, which is also a deep learning method, the effect of the brightness attention mechanism on the enhancement of low-light images can be clearly obtained.

Fig. 17 Original DIV2K image set

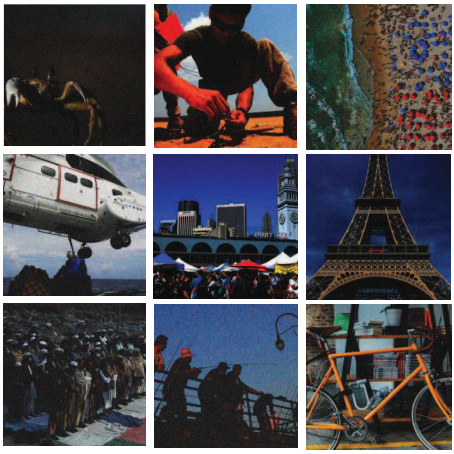

Fig. 18 Div2k image set after low light preprocessing

For the experimental platform, Tensorflow[24] was used to complete the experiment in this paper. The proposed network converges rapidly and has been trained on NVIDIA GeForce GTX1080 for 20,000 generations using synthetic data sets. To prevent overfitting, we use roll off and rotation for data enhancement. For the specific experimental parameters, we set the batch size to 32, and the input image value is scaled to [0,1]. we also added Adam optimizer [25] to optimize the experimental process. In order to stabilize Gan training, we use spectral normalization and gradient penalty to constrain the discriminator.

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:

$$PSNR = 10 * log_{10}\frac{(2^n - 1)^2}{MSE} \tag{6}$$

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

$$MSE = \frac{1}{w * h}\sum_{i=0}^{w-1}\sum_{j=0}^{h-1} X(i,j) - Y(i,j)^2 \tag{7}$$

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:

$$SSIM(X,Y) = 1(X,Y)^{\alpha}c(X,Y)^{\beta}s(X,Y)^{\gamma} \tag{8}$$

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:

$$1(X,Y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{9}$$

$$c(X,Y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{10}$$

$$s(X,Y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{11}$$

Where μ_x,$\mu_y$represent the average brightness of the source image x and the target image σ_x, σ_y represent the brightness standard deviation of the image x, y, and σ_{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.

4.3 Experimental results and discussion

On the data set pre-process with low light, we conducted a comparison experiment with HE, SRIE and DSLR in accordance with the experimental setup in Section 4.2. The quantitative results are shown in table 3, while the qualitative results are shown in figure 19,20.

As can be seen from Table 3, our method is superior to other methods in both indexes, which proves its universal superiority. Compared with the

Table 3 Quantitative results

	HE	SRIE	DSLR	Our
PSNR	13.83	14.50	19.09	22.16
SSIM	0.36	0.58	0.998	0.999

traditional methods HE and SRIE, we improved the PSNR by 60.2%, 52.8%, and SSIM by 177.5% and 72.2%, respectively, indicating that the performance of our method in the synthetic data set is far superior to that of the traditional algorithm. In particular, compared with the HE algorithm, superiority is the most obvious.

Compared with the current deep learning method, DSLR, we also improved by 16.1% on PSNR and slightly improved on SSIM, indicating that our method effectively reduced image noise. Thus, it can be seen that the combination of attention mechanism and GAN for low-light image enhancement is greatly improved compared with the traditional enhancement algorithm, and also has certain advantages compared with the deep learning method.

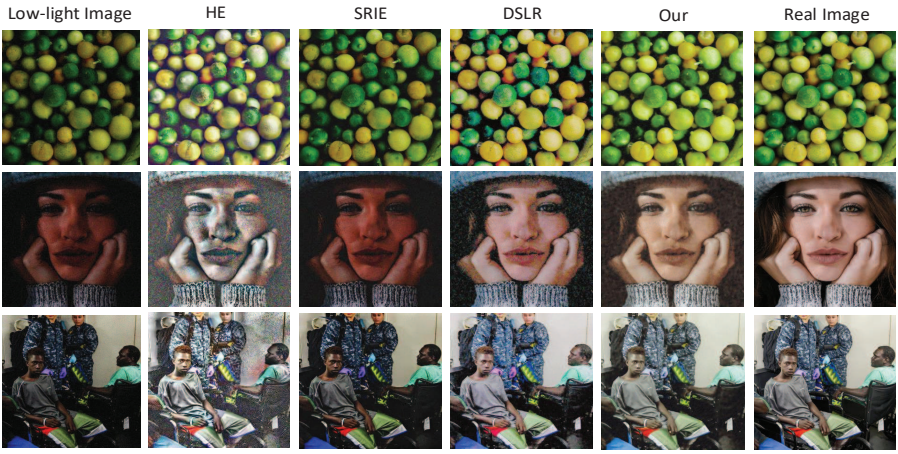

Fig. 19 Image results of the experiment

By comparing the subjective visual effects of the algorithms in FIG. 19, it can be seen that our algorithm effectively enhances the brightness of the image and improves the overall perceived quality of the image. Images enhanced by HE and DSLR algorithm are significantly different in color from real images. Images enhanced by HE and DSLR algorithm tend to be white on the whole, while those enhanced by DSLR algorithm tend to be of cool color, while the color of images enhanced by our algorithm is the closest to the real image. Although the color deviation after SRIE enhancement is not large, the illumination recovery effect is not ideal and the brightness is too dark. Only after the enhancement of our method, the image brightness recovery is relatively high, while the contrast of the image is relatively moderate, there is basically

no color difference, the overall visual effect is almost consistent with the real image.

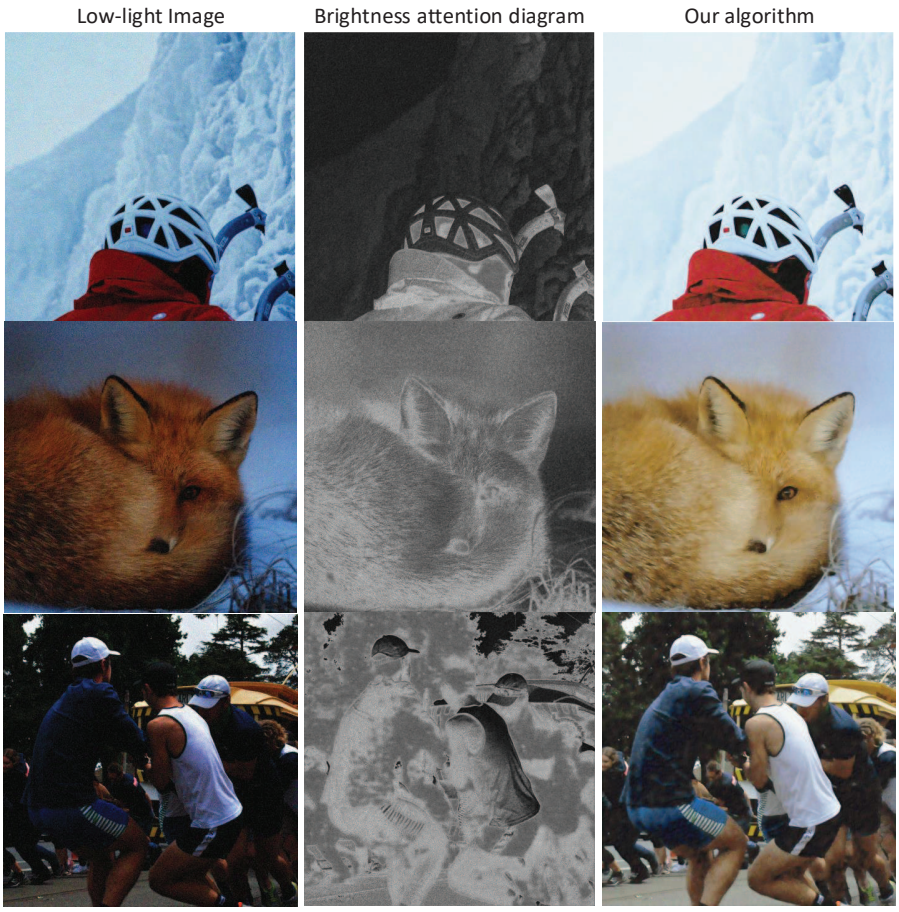

Fig. 20 Image results of the experiment

For the brightness diagram in FIG. 20, the white area corresponds to the lower brightness of the original image, and the black area corresponds to the higher brightness of the original image. The distribution of the white and black regions in the brightness attention diagram in the figure above is basically consistent with that in the original image, indicating that the brightness attention diagram generated by the algorithm in this paper can effectively correspond to the original image. Based on the detailed analysis of the effect diagram in the third line of FIG. 20, although the brightness of the white vest on the right of the original low-light image is relatively high, some areas are still dark, while the overall brightness of the gray vest on the left is extremely low, resulting in serious detail loss. According to the attention diagram of the

brightness predicted by our network, the enhanced gray vest corresponded to the lower illumination part of the image, and the darker area of the white vest was enhanced to a certain extent, but the excessive enhancement of the high brightness part was also avoided.

The above results show that combined with the attention diagram of brightness, the image enhancement algorithm is able to restore the corresponding regions in accordance with the different degrees of brightness in the image. Especially for low-light images, the luminance attention diagram plays an excellent auxiliary effect, so that the enhancement algorithm can effectively identify the low-light region in the image, and then restore the luminance. Therefore, the algorithm has higher robustness, can deal with a variety of complex image conditions, and avoid damage to the appropriate brightness of the low-light image.

conclusion

This article first analyzes the problems faced by low-light image enhancement during the development of image recognition technology in smart cities and then proposes the brightness attention generative adversarial networks designed for the corresponding problems. For photos taken in real low-light or extremely dark environments, the overall brightness of the image is low and the brightness distribution is uneven, and a certain amount of noise is also introduced. Therefore, the network uses the brightness attention branch network to predict the light distribution in the low-illuminance image. Improve the network's sensitivity to different brightness areas in the image. The brightness attention mechanism module adopts a U-net-like structure, which is conducive to multi-scale feature extraction and fusion of features of different brightness areas in the image, and the generated brightness attention map is more accurate.

In the course of the experiment, we conducted comparison experiments with traditional algorithms HE, SIRE, and deep learning algorithm DSLR, and analyzed the experimental results. The following conclusions are drawn: due to the use of the brightness attention mechanism, compared with previous methods, our algorithm may be able to adapt to image enhancement in a variety of low-light environments, and the introduced attention mechanism greatly improves the algorithm's Sensitivity to the different brightness. These advantages can be appropriately enhanced for low-light images produced by complex and changeable environments in smart cities. However, in addition to the above challenges, low-light image enhancement still has computational complexity and excessive model problems on computing-constrained platforms such as mobile terminals, which are waiting for our further research to solve.

References

References

[1] Abdullah-Al-Wadud M, Kabir M H, Dewan M A A, et al,A dynamic histogram equalization for image contrast enhancement, IEEE Transactions on Consumer Electronics, 2007, 53(2): 593-600.

[2] Land E H, McCann J,Lightness and retinex theory,Josa, 1971, 61(1): 1-11.

[3] Jobson D J, Rahman Z, Woodell G A,Properties and performance of a center/surround retinex,IEEE transactions on image processing, 1997, 6(3): 451-462.

[4] Jobson D J, Rahman Z, Woodell G A,A multiscale retinex for bridging the gap between color images and the human observation of scenes,IEEE Transactions on Image processing, 1997, 6(7): 965-976.

[5] Fu Q, Jung C, Xu K,Retinex-based perceptual contrast enhancement in images using luminance adaptation, IEEE Access, 2018, 6: 61277-61286.

[6] Yuan L, Sun J. Automatic exposure correction of consumer photographs[C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012: 771-785.

[7] He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 33(12): 2341-2353.

[8] Li C, Fan T, Ma X, et al. An improved image defogging method based on dark channel prior[C]//2017 2nd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2017: 414-417.

[9] Lore K G, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement[J]. Pattern Recognition, 2017, 61: 650-662.

[10] Shen L, Yue Z, Feng F, et al. Msr-net: Low-light image enhancement using deep convolutional network[J]. arXiv preprint arXiv:1711.02488, 2017.

[11] Cai J, Gu S, Zhang L. Learning a deep single image contrast enhancer from multi-exposure images[J]. IEEE Transactions on Image Processing, 2018, 27(4): 2049-2062.

[12] Wei C, Wang W, Yang W, et al. Deep retinex decomposition for low-light enhancement[J]. arXiv preprint arXiv:1808.04560, 2018.

[13] Lv F, Lu F, Wu J, et al. MBLLEN: Low-Light Image/Video Enhancement Using CNNs[C]//BMVC. 2018: 220.

[14] Chen C, Chen Q, Xu J, et al. Learning to see in the dark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3291-3300.

[15] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2017: 472-480.

[16] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization[C]//arXiv preprint.2016:1607.08022.

[17] Howard A G , Zhu M , Chen B , et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.

[18] Sandler, Mark, Andrew G. Howard, Menglong Zhu, Andrey Zhmoginov and Liang-Chieh Chen.MobileNetV2: Inverted Residuals and Linear Bottlenecks.2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018): 4510-4520

[19] Zhang X , Zhou X , Lin M , et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[J]. 2017.

[20] Ma N , Zhang X , Zheng H T , et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[J]. 2018.

[21] Lingyu Yan,Fu Jairun,Chunzhi Wang,Zhiwei Ye,HongWei Chen,Hefei Ling.Enhanced network optimized generative adversarial network for image enhancement[J].Multimedia Tools and Applications.2021.1.

[22] Youmei Zhang.Research on population Counting Algorithm based on attention convolutional neural network[J].Shandong university.2019.

[23] Zheng Zhang.Study on weather classification and Low-quality Image Enhancement for outdoor monitoring scenes[J].Beijing University of Posts and Telecommunications.2017.

[24] Zhou, Zhiyuan and Feng, Zhuang and Liu, Jilong and Hao, Shijie.Single-image low-light enhancement via generating and fusing multiple sources[J].Neural Computing and Applications.2018.

[25] Li and Huafeng and He and Xiaoge and Tao and Dapeng and Tang and Yuanyan and Wang and Ruxin, Joint medical image fusion, denoising

and enhancement via discriminative low-rank sparse dictionaries learning[J].Pattern Recognition the Journal of the Pattern Recognition Society.2018.

[26] Agustsson E, Timofte R,Ntire 2017 challenge on single image super-resolution: Dataset and study[J].Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.2017: 126-135.

Appendix

Figure legends

This section will introduce the images used in the article.

Figure 1:Smart city image recognition application.

Figure 1 detailed legend:This picture shows an example of the application of image recognition technology in smart cities, specifically the detection and recognition of vehicles by the traffic monitoring system. The blue box is the recognized vehicle. We can see that the quality of the image is slightly blurred, which corresponds to the point that the image quality of smart city recognition described in the text is limited by hardware conditions.

Figure 2:Low light environment photo instance.

Figure 2 detailed legend:This picture shows a specific example of a low-light image that needs to be enhanced. The picture shows the result of taking pictures of several people under street lights with mobile phone cameras in the dark. It can be seen that although there is a light source, due to the lack of light in the overall environment, the characters in the picture are basically completely black. And this is the dilemma that low-light image enhancement algorithms are trying to solve.

Figure 3:Example of image enhancement for mobile phone photography.

Figure 3 detailed legend:This picture shows the actual image enhancement algorithm to enhance the results of mobile phone camera photos. The specific content on the way is a busy street, and we can see that various colors have become more vivid due to enhanced enhancement algorithms. And the influence of the light source on the image quality has also been well controlled. This picture is also the goal pursued by the low-light enhancement algorithm.

Figure 4:Low light image dilemma in the field of criminal investigation.

Figure 4 detailed legend:This picture shows the dilemma of image recognition technology in the field of criminal investigation photography. As you can see in the picture, although the blue frame contains all vehicle targets, because the lighting environment is too dark, only a group of "dark" results can be seen. Obviously, this is fatal to the technical field of criminal investigation, which requires precision. The problem in this figure is also the goal our algorithm aims to solve.

Figure 5:Network structure diagram of CGAN.

Figure 5 detailed legend:This picture is a specific explanation of CGAN in the reference. We can see that the figure includes two parts: generator and

discriminator. In each part, a different prototype part is used to explain the generation process of the generation result X and the discrimination result Z respectively. Supplementary explanation is provided for the above formula.

Figure 6:Network structure diagram of LAPGAN.

Figure 6 detailed legend:This picture is a specific explanation of Laplace GAN in the cited references. This picture shows how to use Laplace's theorem in combination with GAN to complete the recognition of image targets. From left to right is the display of the cycle process.

Figure 7:Network structure diagram of DCGAN.

Figure 7 detailed legend:This picture is a concrete display of the DCGAN structure in the reference. From left to right, it shows how each convolutional layer processes the input image and finally generates the output result. From the detailed figures indicated in the figure, we can see that the input picture is gradually converted from 1024x4x4 to the final 64x64x3.

Figure 8:Self-attention module network structure diagram.

Figure 8 detailed legend:This picture is a specific explanation of SA-GAN in the reference. The attention mechanism in this picture is the focus of this article. The f(x) and g(x) in the upper part of the figure are the specific functions of the attention mechanism, and h(x) is the normal convolutional neural network. Its specific action process is described in detail in the text.

Figure 9:The structure diagram of FCN.

Figure 9 detailed legend:This picture is a detailed description of the FCN in the reference. The specific content in the figure is how the RGB image is disassembled and reorganized by the convolutional layer, and how to reassemble it into three separate color layers of red, green and blue.

Figure 10:The structure diagram of KinD network.

Figure 10 detailed legend:This picture is a detailed description of the KinD network cited in the literature. The color squares in the figure correspond to the three modules of layer decomposition from the perspective of function, reflection image restoration and lighting image adjustment.

Figure 11:Spatial Transformer application example diagram.

Figure 11 detailed legend:This picture shows a concrete example of low-light image enhancement. The content in the figure is the specific process of the enhancement of the handwritten digit set. The specific process explanation has been described in detail in the text.

Figure 12:Spatial Transformer application example diagram.

Figure 12 detailed legend:This picture is a detailed description of the attention channel. The content in the figure is to explain the role of attention channels with squares and functions. The specific process description has been introduced in detail in the text.

Figure 13:Standard U-net structure diagram.

Figure 13 detailed legend:This picture is a specific description of U-net. Because U-net is the prototype of the network used in this article. Therefore, the various parts of U-net are marked in different colors in the local area, and

their principles and names are explained in detail in the figure. Other detailed descriptions have been given in the text.

Figure 14:Schematic diagram of generator network structure.

Figure 14 detailed legend:This picture is an explanation of the structure of the generator in this article. The specific content in the figure is that the input low-light image is processed by the attention network composed of the orange and blue sampling layers to generate a clearer image result.

Figure 15:Schematic diagram of discriminator network structure.

Figure 15 detailed legend:This picture is an explanation of the structure of the discriminator in the network constructed in this article. The specific content in the figure is that after the input low-light image is processed by the U-net-like network, a clearer image result is generated after inputting the fully connected layer and the Softmax layer.

Figure 16:Schematic diagram of generator network structure.

Figure 16 detailed legend:The content of this figure is basically the same as in figure 14. The purpose is to further explain the same problem with different colors.

Figure 17:Original DIV2K image set.

Figure 17 detailed legend:This picture shows the Original DIV2K image set. The specific content in the figure is nine images that have not been preprocessed, which is explained with the preprocessed images below.

Figure 18:Div2k image set after low light preprocessing.

Figure 18 detailed legend:This picture shows the DIV2K image set that has undergone low-light processing. The specific content in the picture is nine preprocessed images. We can see from the figure that after preprocessing, the image is relatively dark compared to Fig18. This result accomplishes our purpose of artificially constructing low-light images.

Figure 19:Image results of the experiment.

Figure 19 detailed legend:This picture is a concrete display of the experimental results. The figure includes three comparison algorithms, our algorithm, the original image and the image after low-light preprocessing. The specific contents are human faces, oranges and an oil painting. This picture intuitively shows the visual effects of various images.

Figure 20:Image results of the experiment.

Figure 20 detailed legend:This picture shows the effect of the brightness attention mechanism in the experiment. The figure includes three types of low-light images, the brightness attention map generated by the algorithm, and the final image generated by the algorithm. After comparison, we can find that the image of the brightness attention mechanism plays a very good auxiliary effect on the enhancement of low-light images, and finally helps our algorithm generate excellent results.

Funding

Project supported by the National Natural Science Foundation (61502155,61772180);

Abbreviations

CGAN - Conditional Generative adversarial networks
GAN - Generative adversarial networks
HE - Histogram equalization
SRIE - Reflected illumination estimation
PSNR - Peak signal-to-noise ratio
SSIM - Structural similarity

Availability of data and materials

No suitable materials are available.

Competing interests

The authors do not have any possible conflicts of interest.All authors read and approved the final manuscript.

Authors' contributions

Jiarun.Fu and Lingyu.Yan conceived the study, participated in the design of its experiments, and drafted the manuscript. Yulin.Peng participated in the experimental part of the research, assisted in completing the experiment and made statistical analysis. Kunpeng.Zheng participated in the synthesis of the low-light image set and counted the experimental results. Gao.Rong and Hefei.Ling provided ideas for the use of the attention mechanism in the paper and participated in the construction of the experimental network. Lingyu.Yan provided financial support for the research of this paper and participated in the improvement of the paper manuscript. All authors read and approved the final manuscript.