symmetry

MDPI

Article

Low-Light Image Enhancement Based on Deep Symmetric Encoder–Decoder Convolutional Networks

Qiming Li *, Haishen Wu, Lu Xu, Likai Wang, Yueqi Lv and Xinjie Kang

Department of Computer Science and Technology, Shanghai Maritime University, Shanghai 201306, China; 201830310216@stu.shmtu.edu.cn (H.W.); 201830310117@stu.shmtu.edu.cn (L.X.); 201830310109@stu.shmtu.edu.cn (L.W.); 201830310025@stu.shmtu.edu.cn (Y.L.); 201930310075@stu.shmtu.edu.cn (X.K.)

* Correspondence: qmli@shmtu.edu.cn; Tel.: +86-021-38282823

Received: 7 February 2020; Accepted: 3 March 2020; Published: 11 March 2020

Abstract: A low-light image enhancement method based on a deep symmetric encoder–decoder convolutional network (LLED-Net) is proposed in the paper. In surveillance and tactical reconnaissance, collecting visual information from a dynamic environment and accurately processing that data is critical to making the right decisions and ensuring mission success. However, due to the cost and technical limitations of camera sensors, it is difficult to capture clear images or videos in low-light conditions. In this paper, a special encoder–decoder convolution network is designed to utilize multi-scale feature maps and join jump connections to avoid gradient disappearance. In order to preserve the image texture as much as possible, by using structural similarity (SSIM) loss to train the model on the data sets with different brightness level, the model can adaptively enhance low-light images in low-light environments. The results show that the proposed algorithm provides significant improvements in quantitative comparison with RED-Net and several other representative image enhancement algorithms.

Keywords: deep convolutional neural network; convolution–deconvolution residual network; low-light image enhancement; symmetric encoder–decoder network structure

1. Introduction

The visual system is one of the main ways for humans to obtain information. With the rapid development of modern computer technology and multimedia technology, digital image processing technology came into being and image and video have become important visual mediums for information transmission, and have gradually been widely applied to everyday life, production, work and other fields, promoting the development and progress of these various fields. However, in the practical application process, the data captured by acquisition equipment are often low-quality images with various problems; therefore, a series of image enhancement processing methods must be carried out to output clear and high-quality images that meet the requirements of practical application.

Image enhancement refers to the mathematical operation and logical transformation of the image data obtained by imaging equipment according to the specific application requirements, emphasizing the interesting content in the image, weakening or removing the uninteresting information, and obtaining high-quality images that meet the actual application conditions. Image enhancement will not increase the inherent information content of image data, but will increase the dynamic range of selected features to make them easily detected or more consistent with the visual effect of human eyes and, thus, lay a solid foundation for subsequent image analysis and image understanding. Image enhancement technology has been widely developed and applied in medical research, aerospace, military applications, fingerprint recognition, face recognition, traffic management and other fields,

so the research on image enhancement technology has always been an important part of image processing technology research. For example, Mandić, I. et al. proposed an adaptive algorithm based·on the LPA-RICI algorithm [1] to denoise X-ray images.

This article focuses on the study of low-light image enhancement. During the image acquisition process, due to the poor lighting environment or the reflection of the object surface, the whole image is insufficiently illuminated and dark, and the detail information of the image cannot be distinguished. In addition, due to the inherent physical characteristics of mechanical equipment, image acquisition systems cannot avoid adding noise during the acquisition process or, due to cost constraints, the camera's quality and imaging capability are poor, which usually leads to the overall darkness of the captured image, resulting in image degradation. At last, the limitation of the image display device also causes the decrease in the level sense of image display or the number of colors. Therefore, the research on fast and efficient low-light image enhancement and denoising algorithms has become one of the key contents to promote the development of low-light image analysis and image understanding. Low-light image enhancement processing has been widely applied in many fields, such as security surveillance, road traffic, medical imaging, military reconnaissance, daily photography and so on. In security monitoring, the quality of the images obtained by the monitoring system under low-light conditions has dropped significantly, which brings great challenges to social security video monitoring. Therefore, it is necessary to process the image acquired under this condition to achieve the purpose of observation. Traffic cameras and driving recorders often encounter poor lighting conditions or poor image exposure, which can cause the video and image to be dark and distorted. In medicine, the quality of medical examination images, such as B-ultrasound, will directly affect the diagnosis results and treatment plans given by doctors. In the military aspect, the quality of military reconnaissance images will also affect the commander's orders and decision-making. In daily life, people use cameras to shoot a large number of images and videos recording life and scenery, but due to the limitations of the shooting environment and the low professionalism of photographers, the quality of images will also be low.

2. Related Work

The enhancement of low-light images has always been a research hotspot in the field of image processing, and related image processing algorithms have emerged endlessly.

Low-light image enhancement based on the histogram equalization (HE) was the mainstream method used in the early days of image enhancement. The histogram is the most basic statistical feature of an image, reflecting the distribution of the gray value and the light and dark distribution of the image. For a gray image, the gray statistical histogram can reflect the statistical situation of different gray levels in the image. Generally speaking, the visual effect of an image has a direct relationship with its histogram. By adjusting and transforming the histogram, the visual effect of an image can be greatly affected. The histogram equalization method proposed by Pizer, S. M. et al. [2] in 1987 can change the gray value of the input image point by point, so that each gray level has a similar number of pixel points; that is, to maintain the relative relationship between the pixel values and transform them into a uniform distribution, so as to increase the dynamic range of the pixel values, so as to enhance the overall contrast of the image and obtain a better visual effect. Based on this principle, many extended and improved HE methods were subsequently proposed.

Pizer, S. M. et al. [3] proposed a contrast limited adaptive histogram equalization (CLAHE) and Abdullah, A. W. et al. [4] proposed an intelligent contrast enhancement method based on dynamic histogram equalization (DHE), which controlled the effect of traditional HE methods and enhanced the image without losing the image details. Wu, C. M. et al. [5] put forward a new model of adjustable histogram equalization based on information entropy as a regular term. Using the probability distribution information of the image's gray level, the method of selecting a regular term coefficient in the adjustable histogram equalization is proposed, which improves the traditional HE method and solves the problem of detail loss in image enhancement. Jiang, B. J. et al. [6] combined the characteristics of global HE and local HE, improving the HE algorithm by using incremental HE. This kind of method

can effectively remove the noise in low-light image enhancement, but there are still the problems of local detail information loss and color distortion. Adaptive extended piecewise HE (AEPHE) [7] divides the original histogram into a set of segmented histograms, and then applies adaptive HE to these extended histograms. The final result of AEPHE is a weighted fusion of these equalized histograms.

However, for non-uniformly illuminated images, the histograms of different regions are different. The HE-based method may need to apply different constraints in different local regions, which is difficult to achieve.

Retinex theory has been extensively studied in low-light image enhancement. Land, E. H. et al. [8] first proposed the theory of the retinex algorithm in 1977. The theory holds that the color of the surface of an object that people perceive is closely related to the reflectance factor of the surface of the object. The color change caused by light is generally gentle, which is shown as a smooth lighting gradient, while the color change effect caused by the change in the object surface appears as a sudden change. By distinguishing these two different change forms, one can distinguish the change in the light source and the change in the object's surface. Noting that the color change is caused by the change in the light source means that people's perception of the surface color of objects remains constant. The degree of color constancy is not affected by the changes in the lighting environment, but is only related to the perception of the reflection properties of the object's surface by the human visual system. Therefore, understanding how to estimate the reflection component of the object is the focus of this kind of method. According to the different estimation of the reflection components of the object, several improved algorithms were developed, such as multi scale retinex with color recovery (MSRCR) [9,10], the probabilistic method for image enhancement (PLE) [11] algorithms, simultaneous reflectance and illumination estimation (SRIE) [12] algorithms and robust retinex model algorithms [13]. Although the method based on retinex theory has a good effect on image enhancement, the edge sharpening is still insufficient and sometimes the color of the image after enhancement deviates from the original color.

Guo X. et al. [14] proposed the low-light image enhancement (LIME) algorithm. Red, green and blue (RGB) are the three primary colors of the additive mixing color model, which can can represent almost all the colors that can be perceived by human vision. The main idea of the LIME algorithm is to find the maximum value in the three color channels of RGB, to firstly estimate the illumination of the image, and then to perfect and modify the initial illumination by adding a structure prior to it serving as the final illumination. The LIME algorithm greatly improves the quality of low-light images, but it also brings problems, such as too high brightness and too bright colors in the reconstructed image.

In addition, with He's dark channel prior defogging algorithm [15], the dark channel prior method should also be used in low-light image enhancement. In [16,17], the authors enhance the dark region of the image according to the dark channel prior technology, but this often leads to obvious noise amplification; in addition, Banić, Nikola et al. [18] used a statistical-based method and proposed the green stability assumption method for illumination estimation, which improved the speed of image processing. The fractional-order fusion model (FFM) [19] algorithm was also proposed to extract more invisible content from the darker areas of the image through the fractional order fusion model, which achieved a better effect. Recently, with the improvement of computers' computing powers, the method based on deep learning has achieved great success in the field of image processing, driven by the application of big data sets. For example, Ledig, C. et al. [20] proposed the super-resolution using a generative adversarial network (SRGAN) algorithm to achieve super-resolution reconstruction of images, while, in the field of image deblurring, Kupyn, O. et al. [21] proposed the deblurring using conditional adversarial networks (DeblurGAN) algorithm to remove motion blur from images. In addition, Ai, S. et al. [22] proposed the attention U-shaped neural network algorithm to enhance the image. Lore, K. G. et al. [23] proposed a depth neural network for enhancing low-light images. Firstly, the non-linear method is used to simulate low illumination conditions to darken the natural images which are set as training data. Then, a depth neural network is formed by stacking a sparse denoising auto coder. After training, these encoders can learn the signal features of low-light images and enhance the brightness adaptively. However, the image processed in [23] is a grayscale image, and there are

some problems, such as unclear image texture and loss of image details. To solve these problems, in this paper, a low-light image enhancement method based on the depth encoder–decoder convolution network is proposed. A low-light encoder–decoder network (LLED-Net) is designed to enhance low-light images. Through training on the image sets of different low-light levels, the multi-scale feature maps are combined to generate an enhanced high-quality image. Structural similarity (SSIM) [24] is integrated into LLED-Net to preserve better original features and textures, thereby to complete the end-to-end mapping from a low-light image to high-quality image. The experimental results show that the image enhancement effect of LLED-Net is better than other methods.

The innovative work of this article is mainly reflected in the following aspects:

1. An image enhancement network structure is proposed. The network consists of a series of symmetrical convolutional layers and deconvolutional layers. The input image first passes through a convolution layer. The convolution layer acts as a feature extractor, encodes the main features of the image content, and then the deconvolution layer decodes the image to restore image details and enhance image brightness;
2. In order to better train the deep network, a skip connection is added between the corresponding convolutional layer and the deconvolutional layer. These skip connections help to propagate features to the front end of the network and pass image details to the end of the network, making the training of end-to-end mapping from low-light images to high-quality images easier and more effective, accelerating the speed of neural network training, and achieving performance improvement at last;
3. SSIM is introduced into the network model, and the SSIM loss function is constructed instead of the default L2 loss function to pay more attention to the generation of image texture and make the output brightness-enhanced image texture clearer.

3. Low-Light Image Enhancement Based on LLED-Net

3.1. The Network Structure of LLED-Net

In view of the good results achieved by RED-Net [25] in the field of image enhancement, the network structure of LLED-NET proposed in this paper draws on the network structure of RED-Net [25]. As shown in Figure 1, the entire network consists of convolutional layers and deconvolutional layers. After each convolution or deconvolution layer, a linear rectification function (ReLU) is used for activation. It should be noted that the pooling layer is not used in the network model, because the purpose of this model is to enhance the image, reconstruct high-quality images, and restore the texture and details of the image, rather than image classification tasks. The existence of the pooling layer usually results in the loss of image details, which are very important to the image restoration task, and reduces the performance of image reconstruction.

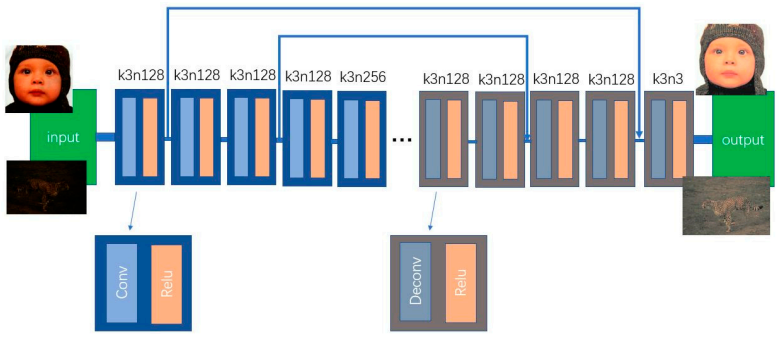

Figure 1. The network structure of low-light encoder–decoder network (LLED-Net).

The size of the convolution kernel of LLED-Net's convolution and deconvolution also refers to the 3×3 size setting that is more common in the areas of image recognition and image enhancement.

As a feature extractor, the convolutional layer extracts the main features of the object in the image and improves the image brightness. After processing by the convolution layer, the input low-light image is converted into a high-quality image. However, the details in the image may be lost during this process, so a deconvolution layer follows to restore the details in the image. In the network model, since the convolutional layer and the deconvolutional layer are symmetrical, the input image is allowed to be of any size. The skip connections are also used in the model, which help to propagate features back to the shallow layer of the network, pass image details to the end of the network, and speed up the training of the network. Meanwhile, the SSIM Loss function is used as the loss function of the network, focusing on the reconstruction of image texture and detail.

3.2. Convolution and Deconvolution Structure

The difference between LLED-Net and traditional convolution neural networks is that the convolution and deconvolution structure is used in the former. The traditional convolutional neural network is a fully convolutional neural network, which is very different in structure from the proposed network. Using a fully convolutional neural network, the image is processed by the convolutional layers, and the image features are extracted and passed to the next layer, which gradually increases the brightness of the image. However, with the continuous increase in convolutional layers, the main content of the image will be recorded, but some small details of the image may be lost, so a deconvolution layer corresponding to the convolutional layer is designed to restore and compensate the details of the image, make up for the deficiency of fully convolutional neural network and, finally, output high-quality images. Among them, the deconvolution layer and the corresponding convolution layer have the same number of layers and the same number of convolution kernels or deconvolution kernels, which makes the proposed neural network form a symmetrical structure.

3.3. Residual Learning

The skip connection is constructed by imitating the residual module [26]. The residual module is shown in Figure 2, where x is the input of the residual block, and F (x) is called residual mapping, which is the output after the first level of linear change and activation by the ReLU activation function, then the second level of linear change and activation. F (x) is activated and output after adding the input value x, which constitutes a residual module. The x added to the output value of the second layer before activation is called a skip connection.

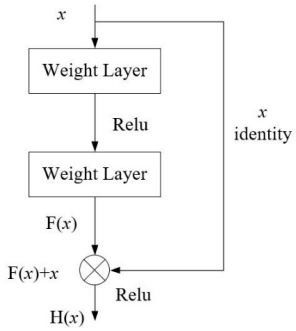

Figure 2. Residual module.

In many computer vision tasks, better performance will be achieved by using deep networks than non-deep networks. According to the description in [26], with the increase in convolution neural network layers, the gradient of backpropagation in the neural network will become unstable, especially large or small, which will often lead to gradient explosion and gradient disappearance. It will hinder the convergence in the training process, and the deeper convolution layer will lead to more and more serious loss of image details. The reason for this phenomenon is that, with the deepening of network layers, the change in features within the network has reached the best situation in one layer, but at this

time, the deep convolution layer of the network will continue to change the features, which will cause the performance of the network to decline; ideally, when the neural network has a large number of layers, the deep convolutional layers of the network should be automatically learned to be the identity mapping form without any change to the features—that is, $H(x) = x$.

Assuming that the convolutional layer before a certain convolutional layer has optimally extracted and changed the image features, then the latter convolutional layer is redundant. Before the introduction of residual learning, it is very difficult to make the parameters learned in this layer realize identity mapping; that is, when the input data is x, the output is still x after the convolution layer and activation by the activation function. However, after the residual learning is introduced, the network avoids directly fitting the parameters to realize the identity mapping at this layer, but, instead, the network is modified to the structure shown in Figure 2 by introducing skip connections to make $H(x) = F(x) + x$. It will be easy to know from the network structure shown in Figure 2 that, in order to make the convolution layer achieve identity mapping, it is only necessary to let $F(x) = 0$. Obviously, it is easier to let $F(x) = 0$ than $H(x) = x$, because when the network is just trained, the parameter initialization in each layer of the network is often close to zero, which makes the value of $F(x)$ approach zero, and the parameters of this layer can be perfectly fitted without too much change. In this way, skip connection is introduced to the network module to realize the residual learning. Compared with updating the parameters of the network layer in the process of training to get $H(x) = x$, using the redundancy layer to get the update parameters to achieve $F(x) = 0$ can make the network convergence faster.

Inspired by residual learning, LLED-Net connects the convolutional layer to the corresponding mirrored deconvolution layer by using a skip connection, so that not only the convolutional feature maps transmitted by the layers are summed up, item by item, as the input of the deconvolution layer, but also the detailed information of the image can be transmitted to the corresponding mirrored deconvolution layer through the skip connection, which will help the deconvolution layer to better recover the high-quality image. In the process of backpropagation in the network, the skip connection also provides a simpler path for backpropagation parameters. Skip connection not only solves the problem of gradient explosion and gradient disappearance of LLED-Net during the training process, but also accelerates the convergence speed of the network and reduces the training time of the network.

It is easy to train a deep convolution neural network and improve the quality of image reconstruction by using this residual learning strategy.

3.4. SSIM Loss

The SSIM loss function is adopted in the network instead of the L2 loss function of RED-Net [25]. L2 loss function is also known as least square error loss function. It minimizes the quadratic sum of the difference between the target value and the estimated value. The L2 loss function can easily amplify the gap between the maximum error and the minimum error, and easily fall into the local optimal solution. SSIM is a measure of similarity between two images. Considering human visual perception, the result of SSIM loss function is more detailed than that of L2 loss function.

For a clear image captured in normal light, the visual perception of the image can be changed by increasing or reducing all pixel values. However, in this case, the image structure hardly changes, and the peak signal to noise ratio (PSNR) [27] before and after the image change will be very different. PSNR is the most widely used objective image evaluation index, which is used to reflect the degree of image distortion. The higher the value, the less the distortion. Therefore, the L2 loss function may not be suitable for the task of this paper. For low-light image enhancement, there will be better protection for the texture of the image by using SSIM loss, so using SSIM loss is more suitable for the task of this paper.

Given two images, x and y, the structural similarity of the two images can be obtained as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

where μ_x is the mean of x, μ_y is the mean of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, and σ_{xy} is the covariance of x and y. $c_1 = (k_1 l)$ and $c_2 = (k_2 l)$ are constants used to maintain stability and prevent the error of denominator equal to 0, so take a small value of k_1 and k_2 to avoid affecting the final result. $l \in L$, where L is the dynamic range of pixel values.

The value of SSIM is in the range [0,1]. The value of one means that the two images are exactly the same. Therefore, 1-SSIM (X, Y) is used to calculate the loss. The loss function of SSIM is defined as follows:

$$l_{ssim} = \frac{1}{N} \sum_{x \in X, y \in Y} 1 - SSIM(x, y) \quad (2)$$

where N represents the number of samples in the data set.

4. Experimental Process and Result Analysis

4.1. Data Set

The biggest difficulty of using deep learning to enhance low-light images is the collection of training data, because it is very difficult to collect a natural low-light image and a high-quality image of normal lighting in the same scene at the same time. The general solution in the field of image enhancement is to synthesize the low-light image as the training data. The public data set BSD500 [28] is selected as our training data. There are 500 images of various people, houses, streets, animals, natural landscapes, etc. in BSD500 [28]. The number of images is expanded to 1200 by using the data enhancement method of rotation and flipping, then the images are enlarged by two, three, and four times, and 720,000 image blocks with the size of 41 × 41 pixels are randomly selected. Then, non-linear adjustment is performed on these 720,000 image blocks to make them become low-light images. The value of γ is randomly selected in the interval [2,5]. When the value of γ is 2, the image appears darker, which is in line with the image taken when the light conditions are slightly weak, and when the value of γ is 5, the image becomes very dark, which is consistent with an image taken under extremely weak light conditions. Therefore, the value of γ in the interval [2,5] can simulate the common natural environment of insufficient light when taking images, so as to make the generated images more closer to the images obtained naturally, Among them, 648,000 images are used as training images, and 72,000 images are used as verification images. The experimental environment is Ubuntu 16.04 operating system, the deep learning framework used is Pytorch, and GTX780 is used as the GPU for accelerated training.

Seven images of a bird, a cheetah, a parrot, a woman, a boat, an elephant, and a baby in the public data set ImageNet [29] were selected for the test data, and these images were also adjusted non-linearly into low-light images. When the value of γ is 3, it is in line with images taken in dark environments in most cases, so the value of γ is set to 3 to simulate low-light images taken under poor lighting conditions.

4.2. Selection of Parameters

The network configuration of LLED-Net is shown in Table 1. Since the network structure of super-resolution using very deep convolutional networks (VDSR) [30] has achieved great success in the field of super-resolution reconstruction, the network structure and parameter settings of VDSR [30] are taken as references in this paper. The total number of network layers is set to 20, and the size of each convolution kernel is 3 × 3. The number of feature maps of the final deconvolution layer is three, which outputs the RGB three-channel image and completes the reconstruction of the color image. In the experiment, the stochastic gradient descent (SGD) method [31] and the Adam optimizer were used. The number of batch training samples is 128, the initial learning rate is 0.01, and the learning rate becomes one tenth of the original after every three epochs.

Table 1. LLED-Net configuration.

Convolution Type	Convolution Kernel Size	Number of Feature Maps	Convolutional Layers
convolution	3 × 3	128	4
convolution	3 × 3	256	3
convolution	3 × 3	512	3
deconvolution	3 × 3	512	2
deconvolution	3 × 3	256	3
deconvolution	3 × 3	128	4
deconvolution	3 × 3	3	1

4.3. Selection of Loss Function

The SSIM loss was compared with the L2 loss in the experiment. If L2 norm loss is used, the network structure of LLED-NET will be the same as the 20-layer RED-Net [25] network structure. Taking PSNR as the evaluation standard, using the same LLED-Net network structure, the bird, elephant, cheetah, baby and boat images to test, and taking the average PSNR of the five images after reconstruction, the results are shown in Figure 3. It can be clearly seen that SSIM loss function is better than L2 loss function in the quality of reconstructed images in the network. Therefore, the SSIM loss function is adopted in this article.

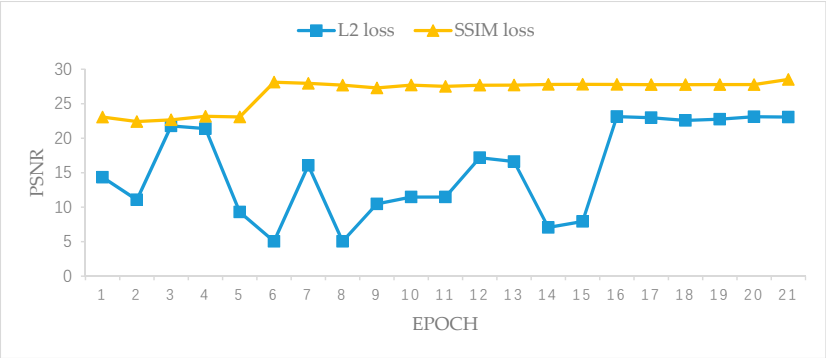

Figure 3. Comparison using different loss functions.

4.4. Impact of Using Skip Connections on Network Training

In the experiment, the results of network training with and without a skip connection are compared. Taking SSIM loss as an indicator, using the same loss function and the same network parameters, the bird, elephant, cheetah, baby and boat images to test and recording the training situation of the two network structures, the results are shown in Figure 4. During the training process, the SSIM losses of the two network structures have different convergence conditions. It can be seen that the network structure using skip connections obviously converges faster than the network structure without skip connections. Therefore, in the experiment, skip connections are used to speed up the training of the network.

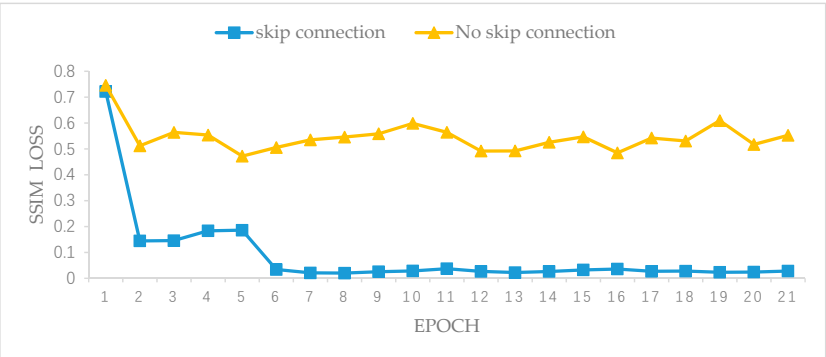

Figure 4. Impact of skip connections on network training.

4.5. Experimental Results and Analysis

When $\gamma = 3$, the five pictures in [23] are processed by using LLED-Net, the PSNR and SSIM values are shown in Table 2, and the processed pictures are shown in Figure 5. It can be seen that the processing results of five pictures with LLED-Net are better than the algorithm proposed in [23] in both PSNR and SSIM indicators.

Table 2. Peak to signal noise ratio (SNR) and structural similarity (SSIM) indicators of [23] and LLED-Net on test images.

	[23]		LLED-Net	
	PSNR	**SSIM**	**PSNR**	**SSIM**
Bird	24.01	0.72	24.07	0.89
Girl	23.16	0.68	31.01	0.98
Pepper	20.54	0.69	26.23	0.91
Town	23.56	0.62	28.05	0.97
House	17.26	0.34	24.00	0.82

Figure 5. LLED-Net processing results of five images in [23].

By using HE, CLAHE, MSRCR [9,10], PLE [11], SRIE [12] algorithms, the LIME algorithm based on retinex theory and the LLED-Net model in this paper, the bird, cheetah, parrot, woman, boat, elephant, and baby images were processed and compared. The difference between the reconstructed image and the original image was used to evaluate the algorithm performance. The results are shown in Table 3.

As shown in Table 3, the best results of the seven algorithms on the test images are labeled in bold. It can be seen that, in most cases, whether using PSNR or SSIM as the evaluation index, LLED-Net is superior to other algorithms.

Table 3 shows the PSNR and SSIM values of several image enhancement algorithms after processing the test image when $\gamma = 3$, and when the γ value of the test image is one, three and five, the average values of PSNR and SSIM of the image processed by several image enhancement algorithms are shown in Table 4.

It can be seen from Table 4 that the LLED-Net algorithm has a higher processing effect on low-light images than other algorithms and has a higher robustness.

Table 3. PSNR and SSIM indicators of different algorithms on test images.

		HE	CLAHE	PLE	SRIE	MSRCR	LIME	LLED-Net
bird	PSNR	11.34	14.79	16.39	15.88	23.49	18.96	31.75
	SSIM	0.5677	0.7689	0.8454	0.8446	0.9655	0.8885	0.9825
cheetah	PSNR	17.56	20.23	15.22	14.33	16.73	13.08	27.37
	SSIM	0.5940	0.8048	0.7812	0.7332	0.8732	0.5212	0.8608
parrot	PSNR	21.06	20.59	15.55	13.62	23.25	15.13	23.34
	SSIM	0.7643	0.8493	0.5758	0.4701	0.8553	0.6020	0.8608
woman	PSNR	13.69	15.78	14.04	13.03	22.18	18.54	26.51
	SSIM	0.7184	0.7615	0.6554	0.6086	0.8412	0.7309	0.9263
boat	PSNR	21.19	17.48	14.68	13.52	24.46	13.66	27.88
	SSIM	0.8218	0.7468	0.6283	0.5867	0.9217	0.6717	0.9635
elephant	PSNR	22.98	17.89	13.66	13.37	23.37	15.07	27.93
	SSIM	0.8480	0.7665	0.5860	0.5894	0.9324	0.6545	0.9512
baby	PSNR	12.33	17.40	13.43	12.63	27.36	16.91	27.68
	SSIM	0.6891	0.8121	0.6679	0.6139	0.9083	0.7692	0.9324

Table 4. Average values of PSNR and SSIM indicators of different algorithms on test images with different γ values.

	HE	CLAHE	PLE	SRIE	MSRCR	LIME	LLED-Net
PSNR	16.17	14.93	14.62	13.46	16.70	14.73	**27.89**
SSIM	0.6891	0.6670	0.6701	0.6456	0.7518	0.6914	**0.9453**

The processing results of the five images of the elephant, bird, boat, cheetah, and baby are shown in Figure 6. It can be seen that the LLED-Net proposed in this paper has the best effect, while the HE, CLAHE, MSRCR [9,10] and LIME algorithms based on retinex theory cannot restore the image to a higher quality state. Among them, there is almost no enhancement in brightness for the image processed by HE, and the whole image is still in a dark situation. The enhanced brightness is very uneven, resulting in many irregular white spots and color distortion, which is quite different from the original image. The image processed by CLAHE is much better than HE, but the enhanced brightness still cannot reach a satisfactory level. The image becomes more blurred, and many non-existent textures and stains appear in the background. The image processed by the LIME algorithm has a very high quality, but the image color is too gorgeous and deviates from the original color seriously. Although the image processed by MARCR [9,10] is better in color restoration, the various evaluation values are lower, the detail restoration of the image is poor, and there are some cases where the background color of some images are very different from the original images. As to the images processed by PLE [11] or SRIE [12], not only the evaluation indicators of SSIM and PSNR are low, but also there is no high improvement in image brightness, and most of the details in the image are hidden in the dark. The color of the image processed by LLED-Net is more vivid, and the texture details in the image are clearer. The original colors and features of the image are better retained, and they are closer to the original image. In short, by using all the algorithms listed above, the image processed by LLED-Net is closest to the original high-quality image.

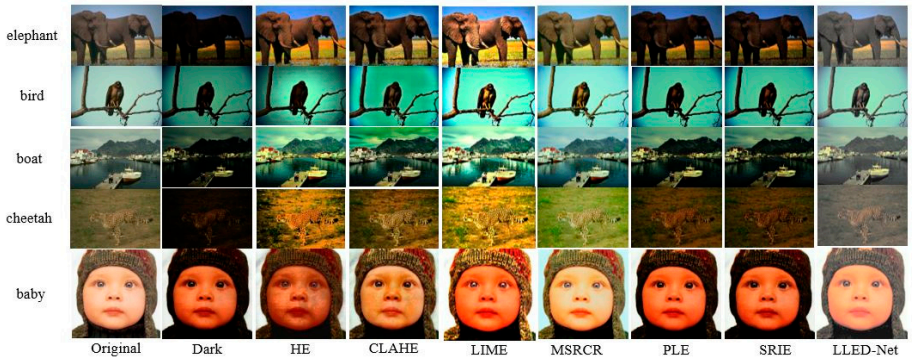

Figure 6. Effects of 7 algorithms on five types of images.

4.6. Processing of Real Low-Light Images

Five real low-light images are downloaded from the network and processed by the HE, CLAHE, MSRCR [9,10], PLE [11] and SRIE [12] algorithms based on retinex theory, the LIME algorithm and the LLED-Net model in this paper, and the results are shown in Figure 7. It can be seen that the image processing by the HE and CLAHE algorithms is prone to high color sharpness and color spots. For example, in the fourth image, the contour of the sun in the sky is too obvious and the edge is sharp after the HE algorithm processing, while rasters that do not exist in the original image appear in the sky after the CLAHE algorithm processing. MSRCR [9,10] produces the phenomenon of color deviation and too gorgeous color. After the MSRCR [9,10] algorithm processes the second and third pictures, the color of the image is too strong and gorgeous, and the entire image tends to a certain hue. After the PLE [11] and SRIE [12] algorithms process these five images, the brightness of the processed images does not increase. For example, in the fourth image, the person in the figure can be recognized by an approximate outline, but the body and the clothing are still not clear enough, and the details of the doors and windows on the building in the distance are hidden in the dark. The image processed by the LIME algorithm appears to be too bright and obviously over exposed. All the five real low-light images are processed well by the LLED-Net model. In subjective visual evaluation, the five images generated by the model have better enhancements in brightness, the details hidden in the dark in the original image are clearly visible in the processed image, and the color and texture of the image are kept relatively complete. It can be seen that the LLED-Net algorithm has a very good performance on real low-light images.

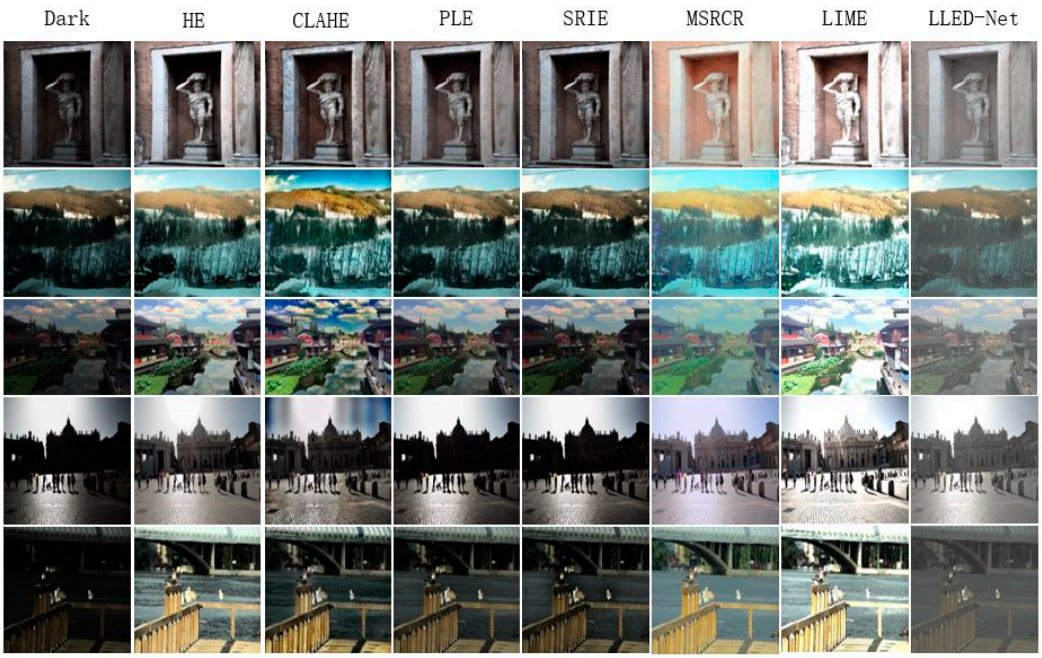

Figure 7. Processing results of various algorithms on real low-light images.

5. Conclusions

Image brightness enhancement plays a very important role in many applications based on computer vision, and is the basis for subsequent image analysis and processing. However, the existing algorithms have problems, such as unclear image texture and loss of image details. Aiming to combat this, a convolutional network with a 20-layer convolution–deconvolution network structure and a SSIM loss function is proposed. This network learns the ability to brighten the image from a large number of synthesized low-brightness images, then applies it to enhance low-light images to restore clarity and achieve end-to-end image enhancement. It is not constrained by various existing algorithms, such as the need to estimate various lighting levels, reflection and other parameters. At the same

time, the skip connection is used to solve the problem of gradient descent in the training process, which makes the network easier to optimize and avoids problems, such as gradient disappearance and gradient explosion.

Experimental results show that the method proposed in this paper is very suitable for enhancing low-light images with different light intensities. On the synthesized low-light images, the images processed by this method are better than the previous several image enhancement algorithms in SSIM and PSNR evaluation indicators. Even if the subjective experience of the naked eye is used, its color and detail restoration effect is better than other image enhancement algorithms; in the processing of real low-light images, the quality of the image processed by this algorithm is higher and clearer, and the texture of the image and the details are also richer. However, in the generated image, there are still problems such as loss of image texture and insufficient color restoration. The next step is to conduct in-depth research on the issue of image detail preservation and further improve the quality of images.

Author Contributions: Conceptualization, Q.L. and H.W.; methodology, Q.L. and H.W.; software, H.W., L.X. and L.W.; validation, Q.L.,H.W. and L.X.; formal analysis, Q.L. and H.W.; investigation, Q.L., H.W., Y.L. and X.K.; resources, Q.L.; data curation, H.W. and L.W.; writing—original draft preparation, H.W.; writing—review and editing, Q.L.; visualization, H.W; supervision, Q.L.; project administration, Q.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Science Foundation of Shanghai, China, grant number 14ZR1419700.

Acknowledgments: The authors gratefully thank the science team of Information and Communication System Engineering Lab of College of Information Engineering, Shanghai Maritime University for their general support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mandić, I.; Peić, H.; Lerga, J.; Štajduhar, I. Denoising of X-ray Images Using the Adaptive Algorithm Based on the LPA-RICI Algorithm. *J. Imag.* **2018**, *4*, 34. [CrossRef]
2. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; Zuiderveld, K. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Imag. Process.* **1987**, *39*, 355–368. [CrossRef]
3. Pizer, S.M.; Johnston, R.E.; Ericksen, J.P.; Yankaskas, B.C.; Muller, K.E. Contrast-limited adaptive histogram equalization: Speed and effectiveness. In Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, GA, USA, 22–25 May 1990; pp. 337–345.
4. Abdullah, A.W.; Kabir, M.H.; Dewan, M.A.; Chae, O.A. Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 593–600. [CrossRef]
5. Wu, C.M. Regularization explanation of adjustable histogram equalization and its improvement. *Tien Tzu Hsueh Pao Acta Electron. Sin.* **2011**, *39*, 1278–1284.
6. Jiang, B.J.; Zhong, M.X. Improved histogram equalization algorithm in the image enhancement. *Laser Infrared* **2014**, *44*, 702–706.
7. Ling, Z.; Liang, Y.; Wang, Y.; Shen, H.; Lu, X. Adaptive extended piecewise histogram equalisation for dark image enhancement. *IET Image Process.* **2015**, *9*, 1012–1019. [CrossRef]
8. Land, E.H. The Retinex Theory of Color Vision. *Sci. Am.* **1977**, *237*, 108–128. [CrossRef] [PubMed]
9. Rahman, Z.U.; Jobson, D.J.; Woodell, G.A. Retinex processing for automatic image enhancement. *J. Electron. Imag.* **2004**, *13*, 100–110.
10. Hanumantharaju, M.C.; Ravishankar, M.; Rameshbabu, D.R.; Ramachandran, S. Color image enhancement using multiscale retinex with modified color restoration technique. In Proceedings of the Second International Conference on Emerging Applications of Information Technology, Kolkata, India, 19–20 February 2011; pp. 93–97.
11. Fu, X.; Liao, Y.; Zeng, D.; Huang, Y.; Zhang, X.; Ding, X. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Trans. Image Process.* **2015**, *24*, 4965–4977. [CrossRef]
12. Fu, X.; Zeng, D.; Huang, Y.; Zhang, X.; Ding, X. A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2782–2790.

13. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-Revealing Low-Light Image Enhancement via Robust Retinex Model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef]
14. Guo, X.; Li, Y.; Ling, H. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. Image Process.* **2017**, *26*, 982–993. [CrossRef] [PubMed]
15. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 2341–2353. [PubMed]
16. Zhang, L.; Shen, P.; Peng, X.; Zhu, G.; Song, J.; Wei, W.; Song, H. Simultaneous enhancement and noise reduction of a single low-light image. *IET Image Process.* **2016**, *10*, 840–847. [CrossRef]
17. Tang, C.; Wang, Y.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. Low-light image enhancement with strong light weakening and bright halo suppressing. *IET Image Process.* **2019**, *13*, 537–542. [CrossRef]
18. Banić, N.; Lončarić, S. Green Stability Assumption: Unsupervised Learning for Statistics-Based Illumination Estimation. *J. Imag.* **2018**, *4*, 127. [CrossRef]
19. Dai, Q.; Pu, Y.-F.; Rahman, Z. Fractional-order fusion model for low-light image enhancement. *Symmetry* **2019**, *11*, 574. [CrossRef]
20. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Shi, W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
21. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 8183–8192.
22. Ai, S.; Kwon, J. Extreme Low-Light Image Enhancement for Surveillance Cameras Using Attention U-Net. *Sensors* **2020**, *20*, 495. [CrossRef]
23. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A Deep Autoencoder Approach to Natural Low-light Image Enhancement. *Pattern Recognit.* **2015**, *61*, 650–662. [CrossRef]
24. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
25. Mao, X.J.; Shen, C.; Yang, Y.B. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections. In Proceedings of the IEEE Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–8 December 2016; pp. 2810–2818.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Winkler, S.; Mohandas, P. The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics. *IEEE Trans. Broadcast.* **2008**, *54*, 660–668. [CrossRef]
28. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [CrossRef] [PubMed]
29. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE conference on computer vision and pattern recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
30. Kim, J.; Kwon Lee, J.; Mu Lee, K. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
31. Wijnhoven, R.G.; de With, P.H.N. Fast training of object detection using stochastic gradient descent. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 424–427.