

Motion Blur Image Restoration by Multi-Scale Residual Neural Network

Xu Hexin

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 809015737@qq.com;

Jiao Yan

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 1483104508@qq.com

Zhao Li

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: zhaoli1998@163.com;

Abstract—Blind deblurring is a basic subject of computer vision and image processing. Motion image deblurring is divided into non blind deblurring and blind deblurring by whether to estimate the blur kernel. Blind deblurring is easy to produce motion artifacts because of the inaccurate estimation of the blur kernel. Non blind deblurring is the best choice for the current blurred image processing. The purpose of this paper is to further improve the definition of blurred image, restore the edge information of contour, and strengthen the repair of texture details. Based on the multi-scale convolution neural network, a multi-scale residual network is proposed, which can comprehensively extract image features, enhance image feature fusion, and constrain image generation by combining multi-scale loss function with anti loss function. The performance of the algorithm is evaluated by testing the peak signal to noise ratio (PSNR) structure similarity and restoration time of the generated image relative to the clear image. This algorithm improves the average PSNR on GOPRO testset, and reduces the recovery time accordingly. It can successfully recover the detail information lost due to motion blur. This algorithm has simple network structure, strong robustness and good restoration effect, and is suitable for dealing with various image degradation problems caused by motion blur.

Keywords-Image Deblur; Motion Deblur; Multi-Scale Residual Network; Deep Learning

I. THE BACKGROUND OF DEBLURRING

Humans rely on the visual system to obtain a large amount of information. Studies have shown

that about 70% of the information is obtained through the visual system. Therefore, the acquisition, processing and use of image information is particularly important. From the exploration of space 60 years ago, the importance of image restoration technology can be seen. At that time, the images sent back to the earth from space were affected by the imaging technology at that time, the shooting environment was not ideal, the relative movement between objects and the camera shake [1]. And other problems, resulting in degradation of the returned image, such as low image resolution, blurred image, etc. In order to solve the problem of image degradation caused by various reasons, people began to study image restoration algorithms. The two most typical image degradation phenomena are noise and blur. In the process of acquiring images, many factors can cause poor image quality, such as object movement and solar radiation. Alignment is out of focus, optical deviation, atmospheric flow, etc. In the process of image transmission, the image will also be blurred and noise due to the interference of the transmission channel and the shooting of electronic components. These degraded images bring great difficulties for subsequent image processing, such as image feature extraction, target object tracking and other tasks. With the

widespread use of images in various fields, people are also pursuing higher and higher resolution of images to deblur. Therefore, it is necessary to continuously research on image restoration technology to meet human visual requirements and applications in various fields.

There are three main types of blur, Gaussian blur, defocus blur and motion blur. There are three types of blur: Gaussian blur, defocus blur and motion blur. Gaussian blur is caused by the Gaussian distribution of each pixel in the image, which is formed by the external diffusion and superposition. The center image is more blurred, and the edge image is more loose. Defocus blur is caused by different depth of field in the process of photographing, some or all of the objects are not in the plane of the imaging system, and there will be local or global defocus blur in the image. The defocus blur is mainly caused by the camera focusing inaccuracy, which leads to different degrees of degradation of objects in different depths of the image [2]. Defocus blur is similar to a disk, and the influence gradually decreases from the center to the outside. In the process of motion blur shooting, the relative displacement between the camera and the object is caused by the motion blur, which is called motion blur. Motion blur can be solved by two methods, one is to reduce the noise exposure time, which can reduce the phenomenon of image motion blur, but with the decrease of exposure time, the signal-to-noise ratio of the image will decrease, and the quality of the image will also decline. The second is to simulate the gradient distribution of the image through mathematical modeling, and further study the image deblurring. The research object of this paper is motion blur, which is the image blur caused by lens out of focus, object movement, camera shake and other factors during the shooting process [3]. For motion blur, equipment can be avoided by using a sports camera. However, such equipment is generally expensive, ten times or even dozens of times the price of ordinary cameras, and it is difficult to popularize and use it on a large scale. Therefore, using efficient and convenient algorithms designed to restore clear images from motion blurred images is currently the mainstream method to deal with motion blurred. Motion blur is

based on the image blur mechanism to model, solve and restore the corresponding high-quality clear image. When the fuzzy kernel is unknown, deblurring is a typical ill-posed problem. How to obtain the final clear image of the image with few known variables has brought many difficulties to the research. With the continuous improvement of mathematical theoretical knowledge and the rapid development of computer vision technology, motion deblurring has made great progress and development, and is widely used in astronomical detection, traffic monitoring, industrial detection, target detection and other fields. With the continuous growth of demand and the ever-changing blurring scenes in real applications, this puts forward higher requirements for image deblurring technology, and at the same time brings greater challenges. Image deblurring is an important classification of image restoration technology, and it is also the current research field of computer vision activities [4]. It has important research significance and application value. Image restoration mainly focuses on two aspects. One is to reduce or avoid the blur of the captured image by improving the hardware equipment. The main implementation method is to build the corresponding control system to achieve the purpose of stable shooting, or stable imaging equipment, this method can effectively control the image blur, but it will increase the cost of imaging difficulty. The second is to process the image after imaging, that is, to achieve the purpose of blurred image restoration through the research of image motion blur removal algorithm. Image deblurring is a serious ill posed problem, because in the process of solving, due to the interference of unknown fuzzy kernel and noise and other factors, the difficulty of image motion blur algorithm is also increasing. Therefore, the image motion blur removal algorithm still needs continuous research and improvement [5].

According to the nature of the blur kernel, it is divided into blind deblurring and non-blind deblurring. Non-blind deblurring results in artifacts in the image due to the deviation of the blur kernel estimation, and can only restore limited image blur. Blind deblurring does not rely on the estimation of the blur kernel and achieves

end-to-end deblurring, but due to the illposed nature of blind deblurring, the details of the image are missing, Enhance the color saturation of the image to meet human visual needs. Therefore, this article will focus on restoring the contour edges of the image. A multi-scale residual module is added to the network, and different convolution kernel sizes are used to extract more image features through the information sharing of the shallow network and the deep network. Based on the inspiration of DeblurGAN [6], the method of combining the counter loss function and the multi-scale function is adopted to adjust the network parameters and train a stable network to achieve the research purpose.

II. RELATED INFORMATION

There are many causes of image blurring. It may be affected by the resolution of the capture device, lighting conditions, atmospheric motion, and the photographer's shooting level, etc., resulting in different degrees and types of blurring in the captured pictures. According to the different types of blur, blurred images can be divided into motion blur, defocus blur, Gaussian blur and so on. This article mainly analyzes the image degradation model of motion blur. Motion blur is the blur produced by the relative displacement between the device and the shooting object during the exposure time of the shooting device. There are many uncontrollable factors that cause image motion blur, such as sun exposure, camera shake, atmospheric movement, and so on. Motion blur image restoration can be widely used in various fields, traffic monitoring, medical imaging, target detection, etc. Therefore, restoring clear images is a hot spot in the field of computer vision today..

The degradation model of motion blur is shown below, b is a blurred image, k is a clear image, and l is a blur kernel, also called a point spread function. The blur kernel is a kind of convolution kernel. This convolution kernel will make the image produce special effects, n is additive noise. The research of this paper does not estimate the fuzzy kernel l , and directly outputs clear images from end to end.

$$b = k * l + n \quad (1)$$

Through the mathematical modeling and analysis of the motion blur image, the motion blur removal is to establish a corresponding mathematical model, extract information from the contaminated or distorted image signal, and restore a clear image along the inverse process of image degradation. This topic is to restore a clear image on the basis of blind deblurring. Blind defuzzification is the current mainstream technology. The principle is not to rely on fuzzy kernel estimation, and to adjust the weight parameters and loss function by constructing a neural network to achieve the effect of the convergence of the objective function. In the process of non-blind deblurring, false contours will appear due to the inaccurate estimation of the blur kernel, and a large amount of noise will be present in the image, which will bring great difficulties to the restoration of the image.

Fuzzy algorithms can be divided into non-blind deblurring and blind deblurring according to whether the fuzzy kernel is known. Non-blind deblurring is performed under the premise that the blur kernel is known, and a clear image is obtained by deconvolution of the blurred image and the blur kernel. Blind deblurring is performed under the premise that the blur kernel is unknown. The traditional blind deblurring method is generally divided into two steps. First, the blur kernel is estimated, and then the blur kernel is deconvolved on the blurred image to obtain a clear image. Fergus et al. [7] discarded the prior assumptions on the image, based on the characteristics of the heavy-tailed distribution of natural images, proposed a deblurring algorithm based on the gradient distribution model, and constructed the original image and the blur under the condition of the known observation image. The joint posterior probability of the kernel, the posterior probability maximizes the combination of the corresponding original image and the convolution kernel. Shan et al. [8] mainly explored the problem of visible artifacts generated by the blind deconvolution problem, and proposed a unified probability model. Through an efficient iterative optimization scheme, the convolution kernel and the restored image are alternately estimated until convergence. Xu et al.[9] introduced a new second order kernel estimation

algorithm. The article introduced a method based on space prior, which can save the memory of potential image edge information and iterative support detection algorithm, which can strengthen the correct preservation of the space constraint kernel. parameter. In the study of unblind deblurring, Sun et al. [10] is a method based on deep learning to estimate uneven motion blur. First, CNN is used to predict the probability of different motion kernels for each image patch, and then image rotation is used. The technology expands the candidate motion kernel set predicted by CNN, thereby significantly improving the performance of motion kernel estimation; Schuler et al. [11] stack multiple convolutional neural networks to simulate the iterative optimization in traditional deblurring methods, and use the kernel estimation module to partially. The estimated value is collected into a single global estimate of the convolution kernel; Gong et al. [12] made the first universal end-to-end mapping of a fully convolutional deep network from a blurred image to a dense motion stream. Unblind deblurring has great limitations in experiments. Blind deblurring has a huge advantage in image restoration, and it has a wide range of application scenarios and can satisfy actual scene deblurring. We use generative countermeasure network structure. The confrontation training mode of generator and discriminator shows great advantages in generating realistic and natural images. The purpose of generator is to learn the distribution of real data, and the purpose of discriminator is to correctly judge whether the data is real data.

Because of this way of confrontation training, Gan network can generate new data based on the original data set. Gan network has powerful image generation ability. The generator uses multi-scale residual module to fully extract image features. In each scale, it uses convolution advantages of different sizes to output to the next scale, and transfers parameters through jump connection, which makes it easy to share data. The discriminator constructs a nine layer network. The generator network and the discriminator play games with each other to restore a clear image. The specific network structure will be introduced in the next section.

III. NETWORK STRUCTURE

The following figure 1. Show the overall structure of the multi-scale residual. The generator network. B, L, and S re-present the blurred image. The clear image and the real clear image are output at the initial scale of each layer. Each layer has 17 residuals. The structure consists of two multi-scale residual modules and two convolutional layers. The subscript represents the scale in the Gaussian pyramid, which is sampled at 1/2 scale. The model uses image pyramids as i-nput, and the output of each intermediate scale is trained into a clear image [13]. The prediction result of the small picture is combined with the original picture of the middle picture as the model input of the middle picture through upsampling, and the prediction of the middle picture is combined with the original picture of the large picture as the model input of the large picture.

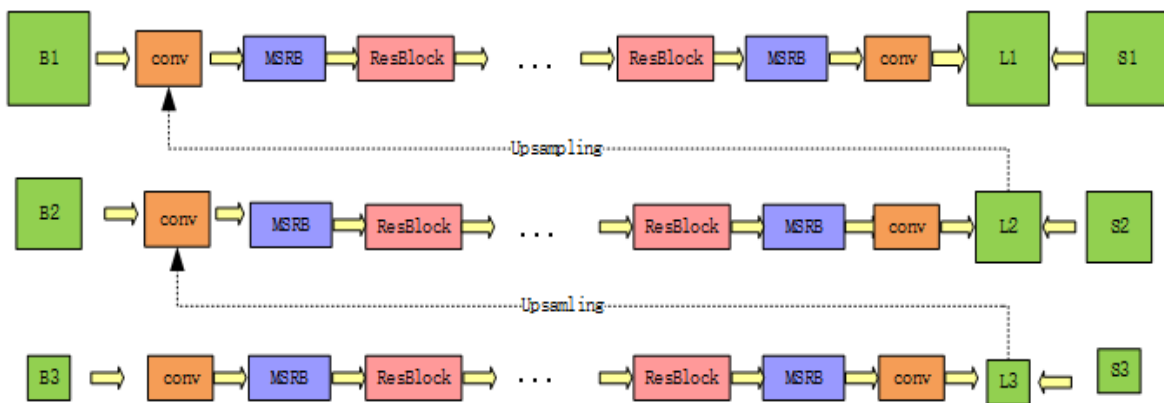


Figure 1. Generator network structure diagram

A. Residual network

ResNet uses the input of one layer and the output of another layer as the output of a block. Assuming that x is the input of a block and one block is composed of two layers, then he first passes through a convolutional layer and activates relu to obtain $F(x)$, and then the result of $F(x)$ after the convolutional layer is added to the previous input x to obtain a result, and the result is activated by relu as the output of the block. For ordinary convolutional networks, our output is $F(x)$, but in ResNet, our output is $H(x) = F(x) + x$. This changes the learning goal and turns the original learning into the goal the function is equal to a known constant value and changed to make the residual between the output and the input 0, which is the identity mapping. After the residual is introduced, the mapping is more sensitive to the change of the output. $H(x)$ is regarded as an underlying mapping fitted by partially stacked layers (not necessarily all networks), where x is the input of these layers. Assuming that multiple nonlinear layers can approximate complex functions, this is equivalent to that these layers can approximate complex residual functions, for example, $H(x) - x$ (assuming that the dimensions of input and output are the same). So we explicitly let these layers estimate a residual function: $F(x) = H(x) - x$ instead of $H(x)$. So the original function becomes: $F(x) + x$. Although these two forms should be able to approximate the required function (as assumed), the learning difficulty is not the same. The motivation of this re expression is caused by the abnormal phenomenon of degradation. If the added layer can be constructed by identity mapping, the training error rate of a deeper model should not be higher than that of its corresponding shallow model. The degeneracy problem shows that it may be difficult for the solver to estimate the identity map through multiple nonlinear layers. With the re expression of residual learning, if the identity map is optimal, the solver drives the weights of multiple nonlinear layers to zero to approximate the identity map. In practice, identity mapping is unlikely to be optimal, but our re expression is helpful for the preprocessing of this problem. If the optimal function is closer to the identity map than to the

zero Map, it is much easier for the solver to find the disturbance about the identity map than to learn a new function. Experiments show that the residual function usually has a small response, which shows that identity mapping provides a reasonable preprocessing.

The formula $F(x)+x$ can be realized by the "shortcut connection" of the feedforward neural network. Shortcut connection is to skip one or more layers. In our example, the shortcut connection simply performs identity mapping, and then superimposes their output with the output of the stacked layer. Identical shortcut connection does not increase additional parameters and computational complexity. The complete network can still be trained through end-to-end SGD backpropagation, and can be implemented simply through the public library without modifying the solver.

Show in Figure 2. Residual network, (1) is the original residual structure, (2) is the modified application in the residual structure, removing the batch standardization layer. The main task of this article is to restore the details of the image without generating noise. It is found through experiments, The BN layer is not sensitive to noise, and key image details cannot be found through a large number of image features. When training the modified residual module, it is found that the convergence speed is improved.

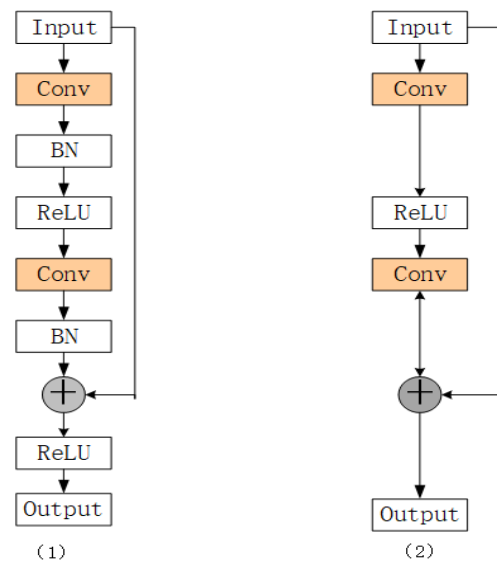


Figure 2. Residual network structure diagram before and after modification

B. Multi-scale residual block

The figure 3 below is based on the multi-scale residual module designed by Li et al. [14] MSRB is mainly divided into two parts, multi-scale fusion and residual learning. These two parts will be described in detail below. Use MSRB to acquire image features of different scales and treat them as local multi-scale features. Secondly, the output of each MSRB is combined for global feature fusion. Finally, combining local multi-scale features with global features can maximize the use of low resolution image features and completely solve the problem of features disappearing during transmission. In addition, we introduce a convolutional layer with a 1×1 convolution kernel as the bottleneck layer to obtain global feature fusion. In addition, we used a well-designed reconstruction structure that is simple but efficient and can be easily migrated to any upward scaling factor. At present, many feature extraction blocks have been proposed. The main idea of the inception block is to find out how the optimal local sparse structure in the convolutional network works. However, these different scale features are simply connected together, leading to underutilization of local features. In 2016, Kim et al. proposed a residual learning framework to simplify the training of the network and enable it to obtain more competitive results. Later, Huang et al. introduced dense blocks. The residual block and the dense block use a single size convolution kernel, and the computational complexity of the dense block increases at a higher growth rate [15]. In order to solve these problems, we propose a multi-scale residual block. Based on the residual structure, we introduce convolution kernels of different sizes to adaptively detect the features of images of different scales. At the same time, jump connections are applied between features of different scales to realize the sharing and reuse of feature information. This helps to make full use of the local features of the image. In addition, a 1×1 convolutional layer at the end of the block can be used as a bottleneck layer, which helps feature fusion and reduces computational complexity.

Multi-scale fusion: This part uses different convolution kernels, 1×1 , 3×3 , 5×5 . Through different convolution kernel sizes, different levels

of information can be extracted, and different scales of information are transmitted to the next layer of network, feature map. The elephants can share and pass on each other. Each part is combined by jump connections to construct a double bypass network. In this way, the information between these bypasses can be shared with each other, so that image features of different scales can be detected. The operation can be defined as:

$$S_1 = \sigma(w_{3 \times 3}^1 * M_{n-1} + b^1) \quad (2)$$

$$P_1 = \sigma(w_{5 \times 5}^1 * M_{n-1} + b^1) \quad (3)$$

$$S_2 = \sigma(w_{2 \times 3}^1 * [S_1, P_1] + b^2) \quad (4)$$

$$P_2 = \sigma(w_{5 \times 5}^2 * [P_1, S_1] + b^2) \quad (5)$$

$$S' = w_{1 \times 1}^3 * [S_2, P_2] + b^3 \quad (6)$$

Among them, w and b represent the weight and bias terms, the superscript represents the number of layers, and the subscript represents the size of the convolution kernel used in the layer. $\sigma(x) = \max(0, x)$ represents the ReLU activation function, $[S_1, P_1]$, $[P_1, S_1]$, $[S_2, P_2]$ represents the connection operation. Let M denote the number of feature maps sent to the MSRB.

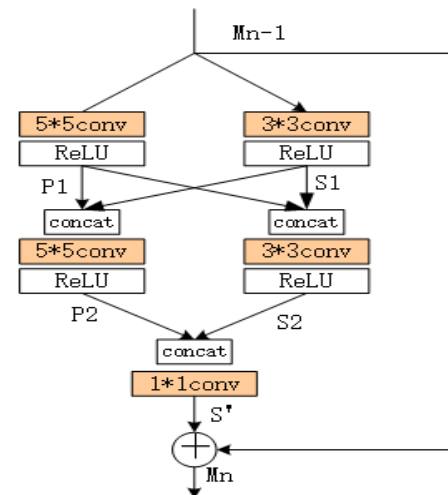


Figure 3. Multi-scale residual structure diagram

So the input and output of the first convolutional layer have M feature maps. The input or output of the second convolutional layer has $2M$ feature maps. All these feature maps are connected and sent to a 1×1 convolutional layer. This layer reduces the number of these feature maps to M , so the input and output of the MSRB have the same number of feature maps. The unique architecture allows multiple MSRBs to be used together.

Local residual learning: inspired by residual blocks. Multi-scale residual blocks introduce residual ideas to improve the expressive ability of the network. The expression of the local residual is as follows:

$$M_n = S + M_{n-1} \quad (7)$$

Among them, M_n and $M_{(n-1)}$ represent the input and output of MSRB respectively. The operation $S+M_{(n-1)}$ is executed by shortcut connection and

adding in order of elements. The use of local residual learning greatly reduces the computational complexity and improves the performance of the network.

With the increase of depth, the spatial expression ability of the network gradually decreases, while the semantic expression ability gradually increases [16]. In addition, the output of each MSRB contains different features. Therefore, how to make full use of these hierarchical features will directly affect the quality of the reconstructed image. In this paper, we use a simple hierarchical feature fusion structure. We send all the output of MSRB to the end of the network for reconstruction. On the one hand, these feature maps contain a lot of redundant information. On the other hand, directly using them for refactoring will greatly increase the computational complexity. In order to adaptively extract useful information from these hierarchical features.

TABLE I. DISCRIMINATOR NETWORK STRUCTURE DIAGRAM

#	Layer	Weight demension	stride
1	conv	$32 \times 32 \times 5 \times 5$	2
2	conv	$64 \times 32 \times 5 \times 5$	1
3	conv	$64 \times 64 \times 5 \times 5$	2
4	conv	$128 \times 64 \times 5 \times 5$	1
5	conv	$128 \times 128 \times 5 \times 5$	4
6	conv	$256 \times 128 \times 5 \times 5$	1
7	conv	$256 \times 256 \times 5 \times 5$	4
8	conv	$512 \times 256 \times 5 \times 5$	1
9	conv	$512 \times 512 \times 4 \times 4$	4
10	fc	$512 \times 1 \times 1 \times 1$	-
11	sigmoid		-

C. Loss function

Inspired by the loss function of GAN, the loss function used in this paper is a combination of multi-scale loss functions. The multi-scale loss function can be extracted from the features of different scales, and deblurring from coarse to fine; the adversarial loss function uses the idea of generating against each other to generate a clear image that is closest to the real image [17]. The specific definition is as follows:

$$L_{total} = L_{multi} + \lambda \times L_{adv} \quad (8)$$

1) *Multi-scale loss function*: This method is coarse to fine method and the output of each intermediate layer is a clear image of the corresponding scale. Therefore, the use of a multi-scale loss function can form a clear Gaussian pyramid into each clear image in the middle.

The MSE standard applies to each level of the pyramid. Therefore, the loss function is defined as follows:

$$L_{cont} = \frac{1}{2K} \sum_{k=1}^k \frac{1}{c_k w_k h_k} L_k - S_k^2 \quad (9)$$

Among them, L_k and S_k respectively represent the model output and the real image on the scale level k . The loss of each scale is normalized by the number of channels c_k , with w_k and height h_k .

2) *Adversarial loss function*: The image restoration is constrained by the game between the generator and the discriminator, where G and D represent the generator respectively, that is, the multi-scale deblurring network and discriminator (classifier) in Table I $\log(x)$ is the probability that the discriminator judges the real data as the real data, and $\log(1-D(G(B)))$ is the probability that the discriminator judges the false data generated by the generator as the opposite of the real data, that is, the probability that the false data is still judged as the false data. The total loss of the discriminator is the sum of the two, which should be maximized. As the ability of discriminator to distinguish the true from the false becomes higher and higher with the increase of training times, the generator has to compete with it, and the generator has to improve its technology accordingly. Therefore, the two improve each other or reduce their own losses, in order to constantly confront each other by combining multi-scale content loss and adversarial loss, the generation network and the discrimination network are jointly trained. The expression of the loss function is as follows:

$$L_{adv} = E_{S \sim P_{sharp}(S)} [\log D(S)] + E_{S \sim P_{blurr}(S)} [\log(1 - D(G(B)))] \quad (10)$$

Without the generation confrontation network, the generated image will have some improvement compared with the original image, but most of the images are fuzzy, the transition of the object edge will be smooth, and the gap is obvious compared with the real image. After joining the generative countermeasure network, the network can further explore the gap between the generated samples

and the real samples, and further improve the visual effect of the generated image. In addition, the network also improves the robustness of the algorithm [18].

IV. EXPERIMENTS

All experiments adopt Pytorch deep learning architecture, and process training images before each batch of training. Firstly, the blurred image and the clear image are randomly placed in the same position, and the image is cropped to 256×256 pixels. The fuzzy image after cutting is used as the input of the generator, the discriminator is used as the input of the discriminator, and the clear cutting image is used as the output of the generator [19]. It should be noted that only when the discriminating ability of the discriminator is strong enough, can it guarantee the optimal result. The dataset and model are introduced respectively [20].

A. Dataset

Neural network training needs a large number of data sets, and the early blurred image is obtained by convolution of fuzzy kernel and clear image. However, the blurred image produced by this simple method is quite different from the real image collected by the camera. Nah et al. Proposed a new image generation method, which uses a high-speed moving camera to capture video, and extracts the connected short exposure frames for averaging, so as to get the blurred image. For example, GOPRO hero 4 black is used to obtain a long exposure blurred image. This method can simulate complex camera jitter and object motion, and the image is closer to the real image. GOPRO dataset is generated by this method. In this experiment, GOPRO dataset is used to train the network. The dataset contains 3214 pairs of blurred and clear images, 2103 pairs of images are selected for training, and 1111 pairs of images are tested.

B. Model training

This article uses the Pytorch deep learning framework. The neural network requires a large number of datasets to train and test the network, shows in Table II. The figure below is a simple discriminator built with 9 layers of convolutional

layers, the activation function uses the sigmoid function, and the volume can all be 5×5 size, as shown in the following table. The weight constant $\lambda = 1 \times 10^4$. We use ADAM optimizer and mini batch size 4 for training. The learning rate is adjusted adaptively from 5×10^{-5} . After 1.5×10^5 iterations, the learning rate is reduced to 1/10 of the previous one. The overall training requires 4.5×10^5 iterations.

C. Test results

The blurred images in the testset are sent to the generating network for processing, and the

deblurring images are obtained. No discriminator is needed in this process. Through the analysis of the original author's data, it can be concluded that PSNR and SSIM have been significantly improved, and the running time has been significantly shortened. In the image shown in the Figure 4. The clarity of the image is significantly improved. The algorithm in this paper can restore the details of the image clearly and meet the basic visual requirements.

TABLE II. QUANTITATIVE COMPARISON OF FUZZY PERFORMANCE OF GOPRO DATASET

	PSNR	SSIM	Runtime
<i>Nah et al.</i>	26.64	0.9142	0.93s
<i>ours</i>	27.33	0.9324	0.72s



Blurred image



Nah et al.





Figure 4. Comparison of GOPRO dataset test results

V. SUMMARY AND PROSPECT

The main content of this paper is motion image deblurring. In view of the poor effect of motion deblurring, the idea of end-to-end blind deblurring is proposed. Unblinded deblurring is a method that does not depend on the estimation of fuzzy kernel, and it can restore the image directly by constructing neural network. The network structure of this paper Multi-scale fusion is through the use of different convolution kernel, which can extract image features in multiple directions and process the texture details of the image; the function of local residual should be used to fuse the extracted different feature images, and the second function can reduce the load of neural network. The combination of multi-scale loss function and adversarial loss function constrains the generation effect of clear image, making the final image closer to the real image. The whole network structure of this paper is simple, and it is suitable to deal with image degradation caused by motion blur. The future work will focus on the following aspects:

1) The algorithm is still lacking in the restoration of the details of the blurred image, and it is necessary to further modify the network design to improve the clarity of the image.

2) The multi-scale residual network structure extracts feature maps of different scales through convolution kernels of different sizes, and will be considered for application in the field of restoring blurred videos in the future.

REFERENCES

- [1] K. Zhang, W. Zuo, and L. Zhang. Learning a single convolutional super-resolution network for multiple degradations. In CVPR, 2018.
- [2] Lai W S, Huang J B, Hu Z, et al. A Comparative Study for Single Image Blind Deblurring[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016.
- [3] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia. Detail-revealing deep video super-resolution. In ICCV. IEEE, 2017.
- [4] Y.-W. Tai, P. Tan, and M. S. Brown. Richardson-lucy deblurring for scenes under a projective motion path. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2011:1603–1618.
- [5] Wang F, Jiang M, Qian C, et al. Residual Attention Network for Image Classification [J]. arXiv: Computer Vision and Pattern Recognition, 2017
- [6] Kupyn O, Budzan V, Mykhailych M, Mishkin D and Matas J. DeblurGAN: blind motion deblurring using conditional adversarial networks/ /Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.
- [7] Fergus R, Singh B, Hertzmann A, Roweis S T and Freeman W T. Removing camera shake from a single photograph. ACM Transactions on Graphics, 2006:787-794.
- [8] Shan Q, Jia J, Agarwala A. High-quality motion deblurring from a single image[J].Acm transactions on graphics (tog), 2008:73.
- [9] Xu L, Ren J S J, Liu C, et al. Deep convolutional neural network for image deconvolution[C]//Advances in neural information processing systems. 2014:1790-1798.
- [10] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2015:769–777.
- [11] Schuler, H. Christopher Burger, S. Harmeling, and B. Scholkopf. A machine learning approach for non-blind image deconvolution. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013:1067–1074.

- [12] Gong D, Yang J, Liu L Q, Zhang Y N, Reid L, Shen C H, Van Den Hengel A and Shi Q F. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu: IEEE, 2017:3806-3815.
- [13] Nah S, Kim T H and Lee K M. Deep multi-scale convolutional neural network for dynamic scene deblurring//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:257-265.
- [14] Li J, Fang F, Mei K, et al. Multi-scale Residual Network for Image Super-Resolution: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VIII[C]// European Conference on Computer Vision. Springer, Cham, 2018.
- [15] Yan Y, Ren W, Guo Y, et al. Image deblurring via extreme channels prior[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:4003-4011.
- [16] Xingjian S H I, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Advances in neural information processing systems. 2015:802-810.
- [17] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning [J]. arXiv preprint arXiv:1603.07285, 2016.
- [18] Xu X, Pan J, Zhang Y J, et al. Motion blur kernel estimation via deep learning [J]. IEEE Transactions on Image Processing, 2018, 27(1):194-205.
- [19] Li Y, Liu S, Yang J, et al. Generative face completion[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017:5892-5900.
- [20] Xingjian S H I, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Advances in neural information processing systems. 2015:802-810.