IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 9 July 2022, accepted 24 July 2022, date of publication 28 July 2022, date of current version 8 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3194524

RESEARCH ARTICLE

High-Frequency Attention Residual GAN Network for Blind Motion Deblurring

JITONG ZHANG[1], GUANGMANG CUI[1,2], JUFENG ZHAO[1,2], AND YING CHEN[1]
[1]School of Electronics and Information, Institute of Carbon Neutrality and New Energy, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Zhejiang Provincial Key Laboratory of Equipment Electronics, Hangzhou Dianzi University, Hangzhou 310018, China

Corresponding author: Guangmang Cui (cuigm@hdu.edu.cn)

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LGF20F050003 and Grant LY22F050002, in part by the National Natural Science Foundation of China under Grant 61805063, and in part by the Zhejiang Association for Science and Technology Yucai Project SKX201901.

ABSTRACT The moving image deblurring method based on deep learning has achieved good results. However, some methods are not effective in restoring image texture detail information. Therefore, this paper proposes a High-Frequency Attention Residual Module (HFAR), which is used to guide the network to learn more high-frequency texture information in the image to improve the quality of image detail restoration. The designed attention residual module consists of two sub-modules, Fourier Channel Attention module (FCA) and Edge Spatial Attention module (ESA). The FCA module gives more weight to the feature maps that contain more high-frequency information in multiple channels. While the ESA module gives more weight to the areas in the feature maps which contain more high-frequency information to guide the network to learn image details and texture information. Extensive experiments on different datasets show that our method achieves state-of-the-art performance in motion deblurring.

INDEX TERMS Motion blurred image, image deblurring, Fourier channel attention, edge spatial attention.

I. INTRODUCTION

Images are the main carrier for people to obtain information and play an important and irreplaceable role in every field of modern society. Today, with the highly developed photography and camera technology, image-based related technologies are developing rapidly, such as target detection, image recognition, and target tracking. However, images are often blurred due to the influence of external factors along with the process of imaging acquisition. Common reasons of the image blur include object motion, camera shake, out of focus and so on. Image blur not only affects people's perception, but also seriously decreases the processing efficiency of subsequent target detection, image recognition, and target tracking technologies. Therefore, in the fields of image processing and machine vision, restoring latent images from blurry images has always been a hot topic. The most common image blur is motion blur, which is caused by the relative movement of the camera and the shooting target during the exposure of the camera. Based on the assumption of linear invariant of the image blurring process, uniform blur is the most common problem in many proposed approaches. However, for the imaging objects with complex motion states, non-uniform blur is often formed as the moving directions and speeds of motion targets usually tend to be different in the real world image scene.

Image deblurring technology can be divided into traditional methods and deep learning methods. **For traditional algorithms** [1]–[6], the blur kernel is first calculated and then the latent image is recovered according to the deconvolution of the blur kernel. The prior information of the image is used to constrain the solution space during iteration, such as edge, dark channel, and saliency information. When the restored image is non-uniformly blurred, different image regions have different blur kernels. Image regions are considered to be treated differently and non-uniform regions are segmented to restore a whole latent image. Nevertheless, the result of segmentation method directly affects the quality

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

of the restoration process. **For deep learning methods**, the original intention of the researchers is to use deep learning methods to estimate the blur kernel accurately and deconvolve the input blurred image to get the restored result combined with the traditional deblurring framework subsequently [7]–[10]. With the improvement and development of deep learning technology, various end-to-end deblurring methods have emerged in recent years [11]–[14]. Moreover, a large number of experiments also confirm that compared with the method of estimating the blur kernel through deep learning, the restored quality of deblurring image obtained by end-to-end method is better. While for the above-mentioned end-to-end methods, the restoration effects of image details still need to be further improved.

Tao [12] and Kupyn [15] enhanced the performance of the deblurring approach by improving the network structure at the expenses of large numbers of parameters of the network. However, with the further increase in the number of parameters, the improvement of the restoration quality is limited while the model training time increases rapidly. Accordingly, the attention mechanism [16], [17] is introduced into the model to alleviate the problem and further improve the training speed. However, as the network learns parameters based on the attention mechanism blindly, a lot of redundant information is learned, resulting in that more edge detail information of the restored image is missing.

To solve the above discussed problem, this paper proposes a HFARGAN motion deblurring restoration network for motion blur image restoration. The high-frequency detailed texture information in the image is effectively used to construct an attention mechanism, which improves the model perception of detailed texture information and enhances the image restoration effect. The core part is the design of the High-frequency Attention residual module (HFAR), which is used to adaptively extract the high-frequency information of the feature map. The module is composed of the Fourier Channel Attention module (FCA), the Edge Spatial Attention module (ESA) and the bottleneck. The edge of the feature map is extracted using bidirectional gradient fusion. An example of image deblurring result by the proposed method is shown in Fig.1.

In general, the contributions of the presented deblurring framework are as follows:

1. The HFAR module composed of FCA and ESA is proposed, which is highly sensitive to the high-frequency detail information in the feature map and can be used to extract the high-frequency detail information in the feature map adaptively.

2. The image content and image edge information are comprehensively considered in the construction of the generator loss. On the basis of ensuring the restoration effect, the edge information of the restoration image is enhanced.

3. The bidirectional gradient fusion method is used to extract the edge of the feature map in the construction of ESA module and image edge loss, ensuring the preservation of image edge information to the greatest extent.

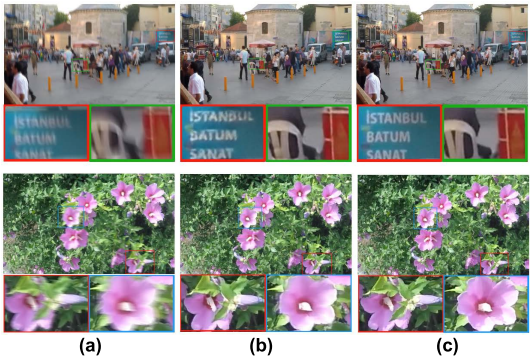


FIGURE 1. An example of the deblurring results of HFARGAN. From left to right: (a)Input blur image,(b)Results of our network, (c)Ground truth clean image.

The sections of this article are arranged as follows: Section I introduces the research background, research significance and summarizes the research innovations of this article. Section II introduces the relevant theoretical basis of this research content. Section III introduces the research content of this article in details. Section IV shows the experimental results and analysis of the propose method. Section V summarizes the research work of this article.

II. RELATED WORKS

A. MOTION DEBLURRING

The image blur model formula [1] is as follows:

$$B = I \otimes K + N \tag{1}$$

where B represents blur image, I represents the latent image, K represents blur kernel and N represents additive noise. For traditional methods, K and B are needed to be solved simultaneously as the blur kernel K is always unknown in the image blur process. It is considered to be an ill-conditioned problem, which could easily lead to unsatisfactory results. Compared with the traditional method, the deep learning method can establish the mapping from the blur image to the blur kernel to estimate the blur kernel and then restore the image accordingly. Moreover, it also could establish the end-to-end mapping from the blurred image to the latent image directly and calculate restore the latent image by deep learning network.

In the early deblurring works [2], [3], [7]–[10], the blur kernel is estimated first and the blur kernel is used for non-blind deconvolution restoration. These methods introduce prior information to regularize the distribution of latent image. Cho *et al.* [8] and Xu *et al.* [3] introduced edge information to regularize the optimal equation. Xu *et al.* [3], [18] proposed L0 norm regularization to extract more significant structures. However, the method has drawbacks that the inaccuracy of the estimated blur kernel would cause ringing artifacts in the image obtained by non-blind deconvolution. With the development of deep learning technology, Chakrabarti [19] proposed deep learning methods to estimate the blur kernel. While the main difficulties and shortcomings of the proposed

method are that it is difficult to produce the training set and test set used for network training from the blur image to the blur kernel. Chang [20] proposed a hyperspectral image restoration method and put forward the WLRTR model, which obtains a good performance on hyperspectral image restoration. This method has a good reference in traditional algorithm deblurring. But it has not been applied to the blur of natural captured images. QuanY [21] introduced a CNN-based image prior defined in the Gabor domain for the noise problem of deblurring tasks, which provides a better performance for noise suppression. Whereas this algorithm is only suitable for the case of known blur kernels, which is somewhat powerless for blind deblurring.

Deep learning has made remarkable achievements in image super-resolution [22]–[24],video frame synthesis [20], image segmentation [14], [26], image classification [27], image recognition [28] and the same field of image deblurring. Many scholars have proposed end-to-end deblurring methods that directly learn the mapping from blur to clear image without estimating the blur kernel [11]–[13], [15]. Nah *et al.* [11] proposed a deblurring network composed of multi-scale fuzzy input residual blocks. The network uses a scaled structure and a large number of convolutional layers with residual connections to improve the receptive field of the network model. However, due to a large number of parameters, it is difficult to converge. Similarly, Cho *et al.* [29] proposed MIMO-UNet, which uses multi-scale input to extract the information of multiple scales of the image. The network parameters are relatively small and it achieves good results in blurring of natural scenes. While the deblurring performance in text images application is average. Tao *et al.* [12] used the encoder-decoder residual block to reduce the amount of parameters and improved the deblurring network model of Nah *et al.* [11]. Kupyn *et al.* [15] firstly proposed the use of DeblurGAN for image deblurring tasks and proposed DeblurGAN-v2 [30] in subsequent improvements for the network performance, which achieved ideal results in the field of image deblurring.

The research results of Kupyn can be regarded as a milestone breakthrough in the field of image motion deblurring. Currently GAN network deblurring is still the most popular deblurring network framework and many researchers have further improved the model on the basis of DeblurGAN. Chen *et al.* [17] proposed CARGAN and introduced a channel attention mechanism in GAN to extract the difference between channels adaptively. The quality of the deblurring results has been improved effectively. Zhao *et al.* [31] proposed a lightweight network FCLGAN and increased the speed of the deblurring network greatly at the cost of some degree of restoring image quality. Tomosada *et al.* [32] proposed DCTGAN, which introduced the DCT loss function to add frequency domain information to the error evaluation. The restoration effect could be improved to a certain extent. The SharpGAN proposed by Feng *et al.* [33] introduced the receptive field block net (RFBNet), which is beneficial to the extraction of image features. The deblurring effect was effectively improved, but the RFB module increased the number of parameters of the network greatly.

B. GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Network (GAN) is proposed by Goodfellow *et al.* [34] and the training process of this network is a game between the Generator and the Adversarial. The generator takes noise as input and outputs samples that are as similar as possible to the real samples so that the discriminator cannot distinguish the generated samples from the real samples. In image-to-image translation, GANs are known for their ability to generate convincing samples. However, they encounter many problems during training, such as mode collapse, vanishing gradients, etc. Later, Arjovsky *et al.* [35] proposed to refine the GAN training problem using Wasserstein GAN refinement with Earth-Mover distance. But meanwhile, using weight clipping on the criterion can also lead to bad behavior [36]. An alternative to clipping weights was proposed by adding a gradient penalty term, which could train various GANs stably without tuning hyperparameters.

On basis of GAN tools, Isola *et al.* [37] proposed the use of cGAN in image-to-image translation with the known architecture as pix2pix. In the cGAN architecture, the work of the discriminator remains unchanged. While the generator not only allows the discriminator to be fooled, but also minimizes the difference between the generated samples and the underlying input. Kubin *et al.* [9] used some feature extractors in a modified WGAN to obtain perceptual loss, which achieved good results in the deblurring task. Chen *et al.* [17] also used a GAN network for the deblurring task and the effect was remarkable. The above studies show that GAN networks perform well on deblurring tasks.

It can be seen from the above analysis that the following problems still exist in the current deblurring tasks:

1. The amount of parameters of the deblurring network are too large, leading to slow training convergence.

2. The lightweight network tends to obtain a small amount of parameters but sacrifice some degree of restoration quality.

3. The designed deep learning network is always unpleasant for the restoration quality of image texture information.

For most deep learning network frameworks, increasing the amount of network parameters and complexity of the network structure can indeed improve the image restoration effect to a certain extent. However, a lot of redundant information is used in the process of model training due to the large amount of image information, resulting in low training efficiency and unsatisfactory restoration effects. Therefore, more and more researchers pay attention to the attention mechanism and introduce it into the deblurring network model to guide the training process and provide a superior deblurring performance with the balance of model training efficiency.

C. ATTENTION MODULE

Attention model can improve the expressive ability of the network and the image restoration effect [16]. Research by Khan [38] et al showed that blur has an important influence

on visual attention. The attention mechanisms used in this paper mainly include channel attention and spatial attention. **Channel attention** can adaptively adjust the weights of each channel of the feature map to improve the applicability of the network to different images. Chen *et al.* [17] also introduced a channel attention mechanism in the network to learn channel information that is conducive to image deblurring, and adaptively adjusts the weight of the channel. **Spatial attention** mechanism can deal with the blurring effect of different regions by using the spatial features of the image. It can learn regional information that is conducive to image deblurring, and adaptively adjust the weights of different regions of the image.

III. THE PROPOSED METHOD

Currently GAN network deblurring is still the most popular deblurring network framework. In recent years, many deep learning methods have been proposed on the basis of further improved DeblurGAN, such as DeblurGAN-v2, CARGAN, FCLGAN, DCTGAN, SharpGAN, etc. Similarly, inspired by the method presented by QiaoC [39], we proposed an improved DeblurGAN framework and introduced a modified attention mechanism HFAR module into the network to obtain restoration results with high quality. Furthermore, the attention mechanism introduced in the proposed HFAR module is an improvement structure in channel attention and spatial attention, which is helpful to guide the network to learn more high-frequency details and texture information from the features and enhance the restoration effect of the proposed model on texture areas effectively.

The HFARGAN network is composed of a generator and a discriminator. The overall structure of HFARGAN is shown in Fig.3. The generator is composed of encoding module, decoding module and HFAR module, which is used to take the blurred image as input and produce the estimate of the sharp image. The discriminator is utilized to discriminate the restored image and the latent image. Furthermore, it could also be used to guide the generator to optimize the parameters.

A. HFAR GENERATOR

The generator in the proposed deblurring structure is composed of an encoding module, a decoding module and an HFAR module. The generator structure of HFAR GAN is shown in Fig.4. The main function of the encoding module is to extract features. While the designed HFAR module is utilized to analyze the information extracted by the encoding module and the conversion process is performed to calculate a latent image from a blurred image accordingly. The decoding module is applied to reconstruct and restore the feature map information extracted by HFAR module to obtain the final image.

The encoding module is composed of 3 Convolutional layers. The size of the convolution kernel is 7×7, 3×3 and 3×3, respectively. After convolution operation, each layer is output to the next layer after passing through the Normalization layer and the ReLu activation layer. The information of the feature map is extracted by the HFAR module adaptively. Then the conversion operation is carried out to perform the blur image to latent image. Moreover, in order to ensure the texture information of the restored image more accurate, the frequency domain information and edge information of the feature map are utilized comprehensively to guide the designed module and whole deblurring network to pay more attention to the extraction of texture information. Correspondingly, the specific implementations are described in section C of this part.

The decoding module corresponds to the structure of encoding module. Each layer is co mposed of Concat, Convolution, Normalization, Activation function, and Upsampling.The main function of this module is to reconstruct and restore the characteristic information of the image information extracted by the HFAR module. Meanwhile, Skip connection is constructed between the feature map corresponding to the encoder and decoder, which helps to merge multi-scale information and improve the utilization of effective image information. Finally, a Skip connection is constructed between the entire input and output, allowing the model to learn the residual between the input and output and speed up the convergence of the model.

B. SPECTRUM AND EDGE INFORMATION ANALYSIS

Convolutional networks can fuse information in different areas of different channels to construct new channels and feature information. Each feature map reflects certain feature information of the image. For different learning tasks, the useful feature information varies accordingly. When extracting feature maps, each layer of convolutional network would extract a large amount of feature information. Different kinds of feature information occupy different degrees of importance in deblurring tasks. In other words, the feature information of different channels contributes differently to the deblurring task. Similarly, the information on different locations of a certain feature map also provides different contributions to the implementation of deblurring task.

Along with the solution process of image deblurring, the characteristic of image blur is the loss of detailed texture information, which could be reflected in the edge map and saliency map in the spatial domain. Meanwhile, the detailed texture information could also be embodied by spectrum information analysis in the frequency domain.

The HFAR module combines edge information and spectrum information so that the network pays more attention to detailed texture information in the deblurring task.

The following sections analyze the benefits of introducing edge information and spectrum information into the network.

1) SPECTRUM INFORMATION

The detailed texture information of the image can be reflected in the spectrogram in the frequency domain. The frequency-domain information can be obtained in the spectrogram after the image is Fourier transformed. Fig.4 shows the examples of the spectrum map.

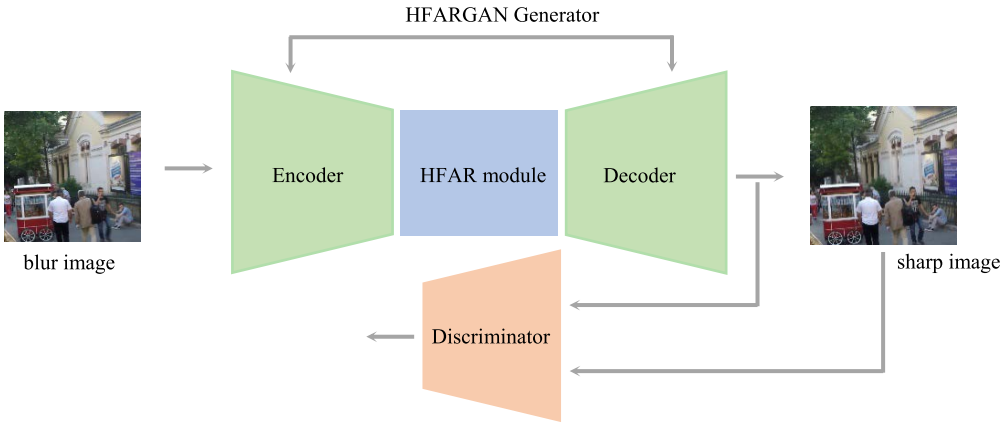


FIGURE 2. The overall structure of HFARGAN.

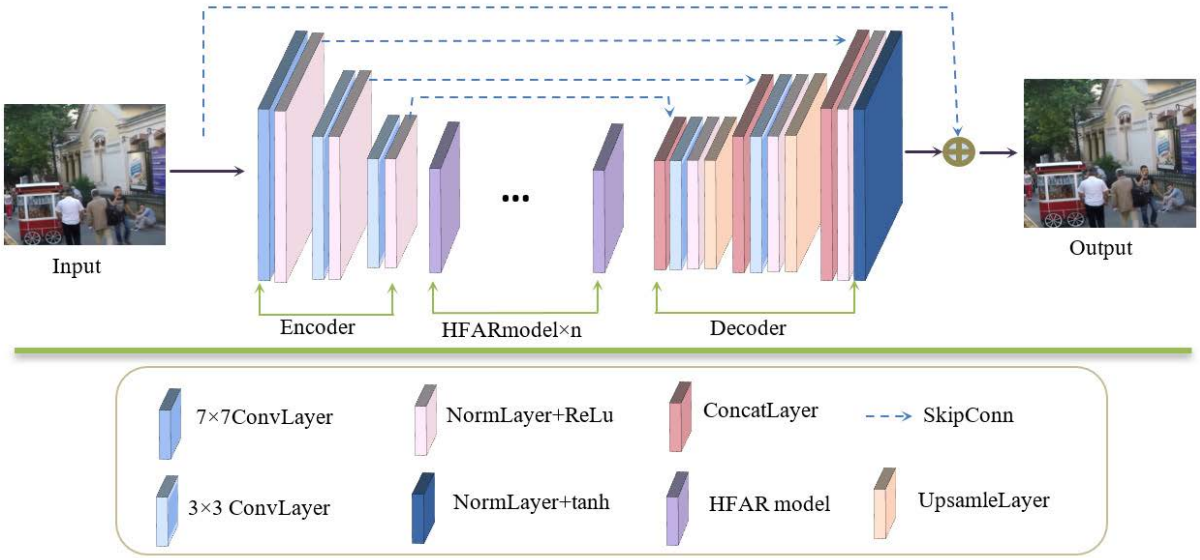


FIGURE 3. The generator structure of HFAR GAN.

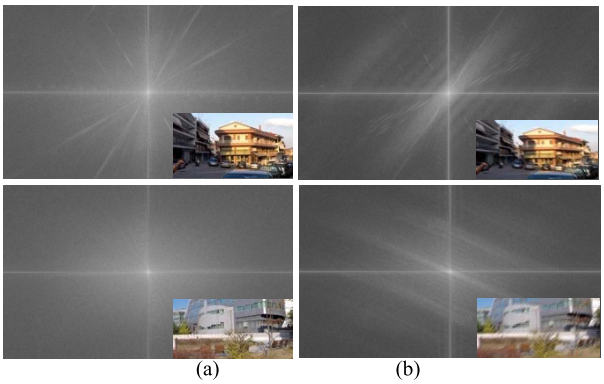


FIGURE 4. Examples of Spectrum map in the dataset. (a)results of sharp image,(b)results of blur image.

It can be seen from the spectrograms in Fig.5, there are obvious differences between the spectrogram of the clear image and the spectrogram of the blurred image. It shows that the loss of high frequency information has a great influence on the spectrum information.

Twenty-five pairs of graphs are selected from the GOPRO and REDs data sets to calculate the spectrogram. The average pixel value of the spectrogram is calculated, and the result is shown in Fig.5.

The average pixel values of the spectrogram of the clear image and the blurred image from GOPRO data set and REDS data set are shown in Fig.5. In the GOPRO dataset and the REDS dataset, the pixel values of the blurred image spectrogram are significantly lower than the corresponding clear images. The conclusion can be drawn that high-frequency information is lost when the image is blurred. It is necessary to introduce spectrogram information as an indicator of channel attention extraction. Therefore, the introduction of frequency domain information can guide the network to pay attention to the influence of frequency domain information on restoration.

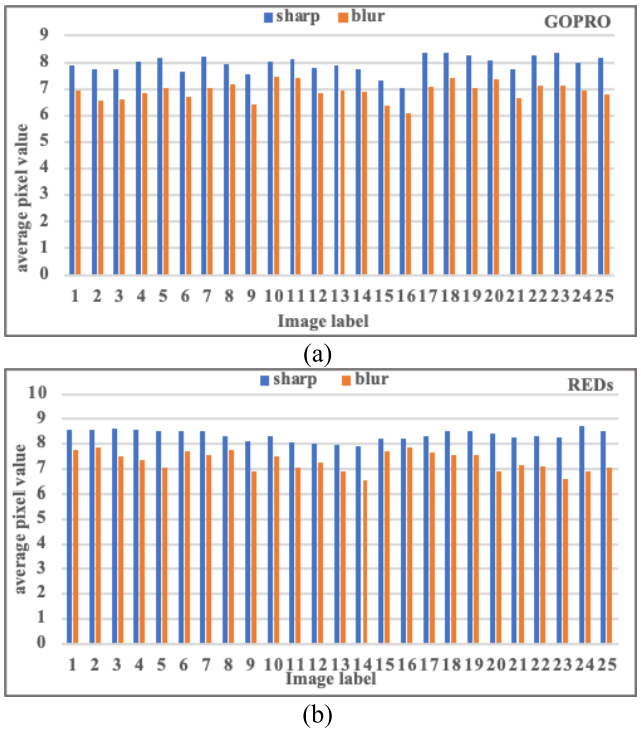


FIGURE 5. Bar chart of the average pixel value of the Spectrum map of 25 pairs of images.(a) GOPRO dataset, (b) REDS dataset.

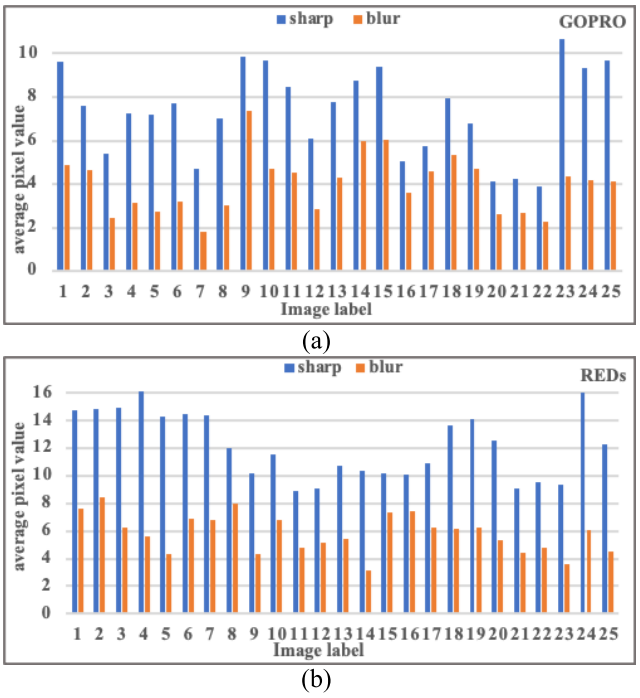


FIGURE 7. Bar chart of the average pixel value of the edge image of 25 pairs of images.(a)GOPRO databases, (b) REDS databases.

2) EDGE INFORMATION

The detailed texture information of the image in the spatial domain can be reflected in the edge map. Fig. 6 shows the examples of edge map.

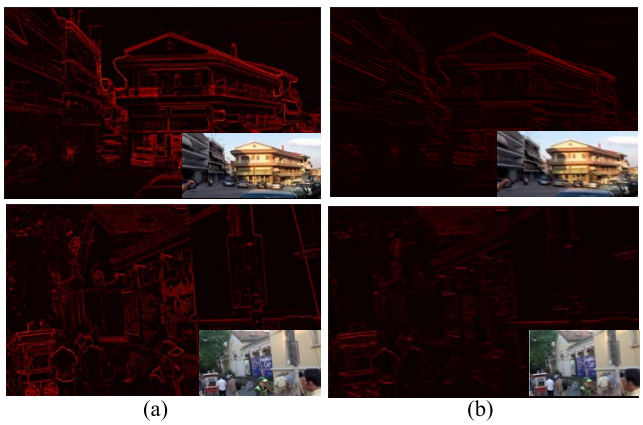


FIGURE 6. examples of edge map extraction in the dataset. (a) results of sharp image,(b)results of blur image.

By analyzing the edge maps of sharp image and blur image, it can be seen that the edge map of the clear image is more clearly described while the edge map of the blurred image has little information. It shows that the edge information is seriously lost after the image is blurred.

Also, another twenty-five pairs of graphs are selected from the GOPR and REDs datasets to calculate the average pixel values of the edge image. The corresponding results of the average pixel values of the edge image for sharp image and blurred image selected from GOPRO and REDS data set are shown in Fig.7.

It can be seen that the average pixel values in the edge image of the blurred image is significantly lower than the corresponding sharp image. Images with low pixel values generally have low energy and darker images, indicating that the high-frequency information of the blurred image is lost due to the imaging blur process. It is necessary to introduce edge map information as the index of spatial attention extraction, which could lead the network to pay more attention to the influence of texture edge information on restoration.

C. HFAR MODULE

The information of the feature map is extracted by the HFAR module adaptively. Fig.8. shows the structure of the HFAR module. The module is composed of bottleneck, FCA and ESA. The bottleneck structure uses two 1×1 and one 3×3 convolution kernels instead of the Res module composed of two 3×3 convolution kernels. As a result, the parameters are greatly reduced while the training effect is equivalent. To obtain more accurate texture information in the restored image, the frequency domain information and edge information of the feature map are introduced to guide the construction of channel attention FCA and spatial attention ESA so that the network pays more attention to the extraction of texture information.

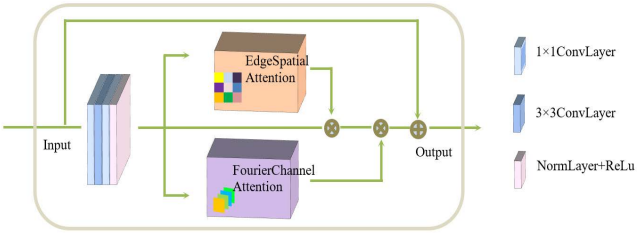


FIGURE 8. HFAR module.

1) FCA MODULE

The FCA module uses the difference in the spectrogram of each channel of the feature map to assign weights to the channels. Fig.9 shows the structure of the FCA module. The $h \times w \times c$ feature map is subjected to Fourier transform to obtain the corresponding spectrogram. Average pooling and maximum pooling operations are performed on the feature map of each channel. The results are added to obtain a preliminary weight of $1 \times 1 \times n$, then each channel is obtained through multi-layer perceptron (MLP). The input feature map is redistributed according to the new channel weight to obtain the output feature map.

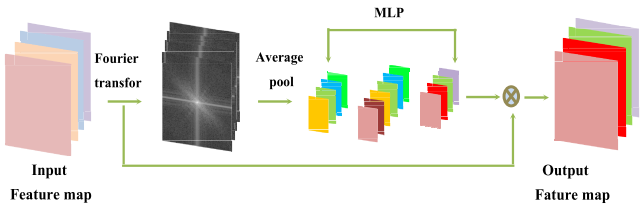


FIGURE 9. Fourier channel attention module (FCA module).

The specific operation is as follows:

$$M_C = MLP(MaxPool(FFT(F)) + AvePool(FFT(F))) = MLP\left(\tilde{F}^c_{avg} + \tilde{F}^c_{\max}\right) \quad (2)$$

$$\tilde{F}^c_{avg} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \tilde{F}_{x,y} \quad (3)$$

$$\tilde{F}^c_{max} = max(\tilde{F}_{x,y}) \quad (4)$$

$$M_C(F) = M_C \times F \quad (5)$$

where M_C represents the channel weight. F represents the input feature map. $\tilde{F}$ represents the spectral feature map. FFT represents the Fourier transform operation. $M_C(F)$ represents the output feature map after channel weight distribution. W, H represents the size of the feature map. x, y represents the pixel of the feature map. MLP stands for multi-layer perceptron layer and consists of two fully connected layers.

2) ESA MODULE

According to the spatial difference of the edge map of the feature map, the ESA module assigns different weights to different regions of the feature map. A larger weight is set for the feature map areas which are conducive to image texture information restoration. Fig.10 shows the structure of the ESA module.

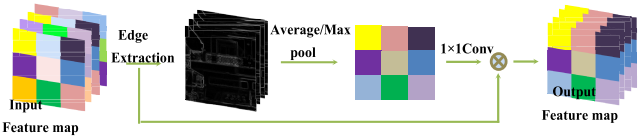


FIGURE 10. Edge spatial attention module (ESA module).

The edge extraction is performed on the $h \times w \times c$ feature map to obtain the edge map and average pooling and maximum pooling operations are conducted on the same pixel points of the c channels. Thereafter, the output Feature map is obtained by adding the results to get the $h \times w \times 1$ weight matrix and assigning the weights to the input feature map.

The specific operation is as follows:

$$M_S = \underset{c}{ave}\left(\hat{F}^c_{x,y}\right) + \max_c\left(\hat{F}^c_{x,y}\right) \quad (6)$$

$$M_S(F) = M_S \times F \quad (7)$$

where M_S represents the weight of the pixel of the feature map. F represents the input feature map. $M_S(F)$ represents the output feature map after weight distribution. x, y represent the pixel position of the feature map. represents the edge feature map.

The edge map used in ESA module is extracted by bidirectional gradient fusion method. The introduced fusion method could be implemented to ensure the integrity of the edge information to the greatest extent, which plays a vital role to guarantee the accuracy of edge extraction.

The specific operation is as follows:

$$e_{x,y} = \max\left(|\partial e_h|, |\partial e_v|\right) \quad (8)$$

where $e_{x,y}$ represents the value of the edge map corresponding to the pixel at the position. represents the horizontal gradient of the pixel. represents the vertical gradient of the pixel.

The visual effects of bidirectional gradient extraction and unidirectional edge extraction are shown in Fig.11. Compared with bidirectional extraction, the vertical gradient information tends to be lost when extracting the edge in the horizontal direction. Besides, the horizontal gradient information would be lost when the edge is extracted in the vertical direction.

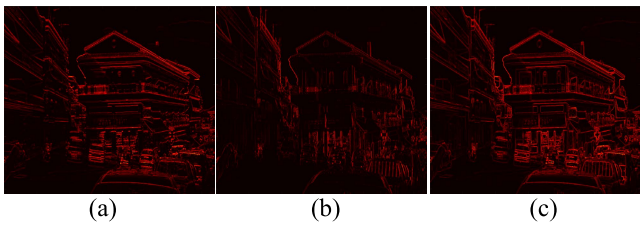


FIGURE 11. Edge extraction image. (a) Horizontal gradient,(b) Vertical gradient, (c) Bidirectional gradient.

It can be seen that the effect of bidirectional gradient extraction achieves a superior performance.

D. LOSS FUNCTION

The loss function is another important aspect in a deep learning deblurring framework, which works in two ways. One is to evaluate the error of the image output by the deblurring network and the other is that the network adjusts the network parameters according to this error adaptively, thereby allowing the network to converge.

The loss function of the HFARGAN generator is designed to be consisted of three parts, including adversarial loss, perceptual loss and edge loss. The role of adversarial loss is to associate the discriminator and the generator to better train the network. Moreover, the perceptual loss and edge loss are introduced to calculate the MSE error of the deblurred image and edge map, respectively. The following is the specific calculation method:

1) ADVERSARIAL LOSS

To improve the stability and convergence speed of GAN network training, WGAN-GP is used as the evaluation function, as expressed as follows:

$$L_{adv} = \underset{\tilde{x}\sim P_g}{E}[D(\tilde{x})] - \underset{x\sim P_r}{E}[D(x)] + \lambda \underset{\hat{x}\sim P_{\hat{x}}}{E}\left[\left(\left\|\nabla_{\hat{x}} D(\hat{x})\right\|_2 - 1\right)^2\right] \tag{9}$$

where,represents the image information generated by the generator $\widehat{x} = \varepsilon x + (1-\varepsilon)\widetilde{x}$. According to experimental results, it can be concluded that the generator could provide better robustness property when λ is selected as 10.

2) PERCEPTUAL LOSS

The MSE error between the VGG-19 convolution feature map of the latent image and the restored image is taken as the perceptual loss, which is expressed as follows:

$$L_p = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I_B)_{x,y} - \phi_{i,j}(G(I_S)_{x,y}))^2 \tag{10}$$

where, I_B represents latent image.represents blur image. $\phi_{i,j}$ represents the output of the i-th layer and j-th channel of VGG-19. $G(I_S)$ represents the restored image generated by the generator based on the blurred image. $W_{i,j}$, $H_{i,j}$ represents the feature map size of the i-th layer and j-th channel of VGG-19. In this paper, the features of the VGG-19 $\phi_{3,3}$ layer is used to calculate the MSE error of the deblurring image, which is confirmed to be a perfect option to guarantee the obtain of high quality restored results in Kupyn's research.

3) EDGE LOSS

The combination of edge loss and the proposed HFAR module can better guide the network to restore the texture information in the image. The calculation method is to used the MSE error of the edge map of the restored map and the clear map. The detailed calculation equation is constructed as follows:

$$L_{edge} = \frac{1}{W_{edge}H_{edge}} \sum_{x=1}^{W_{edge}} \sum_{y=1}^{H_{edge}} (Edge(I_B)_{x,y} - Edge(G(I_S)_{x,y}))^2 \tag{11}$$

where, W_{edge}, H_{edge} represents the size of the edge map. The edge extraction is calculated using the bidirectional gradient fusion method and the specific calculation method is the same as described above.

4) TOTAL LOSS FUNCTION

$$L = L_{adv} + \alpha L_p + \beta L_{edge} \tag{12}$$

where, α, β represent the weight of perceptual loss and edge loss, respectively.

IV. EXPERIMENTS

A. DATASETS

The proposed approach is evaluated on the GOPRO dataset, the REDS dataset and BMVC_OCR_Text dataset. The details of these three datasets are as follow.

The widely used GOPRO dataset is established by Nah *et al* [11], which is utilized GOPRO4 HERO Black camera to shoot multiple videos with a frame rate of 240fps. Then 7 to 13 consecutive frames are merged to obtain images with varying degrees of blur. A total of 3214 pairs of blur-sharp image pairs were generated with a resolution of 1280×720, including 2103 pairs for training and 1111 pairs for testing.

The REDS [40] dataset is a new high-quality (720p) video dataset proposed in the NTIRE19 competition. The dataset is produced with greater motion and complexity compared with GOPRO dataset. REDS contains 240 training segments, 30 verification segments and 30 test segments, with each segment consisted of 100 consecutive frames.

The BMVC_OCR_Text dataset is a textual image dataset proposed by the British Machine Vision Conference (BMVC). A total of 930 pairs of blur-sharp image pairs were generated with a resolution of 512×512. The dataset consists of 93 sharp images processed with 10 different blur kernels.

B. IMPLEMENTATION DETAILS

In our experiment, the TensorFlow deep learning framework tool is used to implement the deblurring network. The model training is conducted on a computer equipped with AMD Ryzen 5 3400G CPU and NVIDIA GeForce GTX 1080 Ti GPU.

For optimization, we use Adam as the optimization for the generator and SGD as the optimization for the discriminator [41]. Furthermore, 5 gradient descent steps are performed on the discriminator and one following step is performed on the generator. All models are trained with a batchsize of 1.

For the learning rate, the learning rate of the discriminator is set to 1×10^{-4}. While the learning rate of the generator is set to 1×10^{-4} after 50 epochs, with the initial parameter selected as 1×10^{-2}.

For the data set, we use 2103 images from GOPRO dataset for training on the training set. The original image size of the GOPRO dataset is 1280×720 and we crop it to 640×360 for input. The test set during training also uses the GOPRO test set.

For training epochs, the maximum value is set to 300. The mean MSE values of the training set and the test set pushed forward by 10 epochs are calculated and the training process is stop when the mean MSE value no longer decreases. Finally training stops at 270epoch. The entire training process would take about 3 days.

C. RESULTS AND COMPARISONS

Several advanced image deblurring methods were selected for performance comparison, including Tao *et al.* [12], Kupun *et al.* [30], Chen *et al.* [16], Chen *et al.* [17] and Cho *et al.* [21]. It is to be noted that the attention-adaptive and deformable convolution modules (AAM, DCM) are proposed in ChenL's method, which are considered to be used in methods such as DebluurGAN, MSCNN and SRN, etc in order to verify the effect of these two modules. The DeblurGAN+AAM+DCM is chosen with almost the same number of parameter as ours for fair comparison. Regarding the experimental results of these methods, the results reported in the literature are quoted as directly as possible. Otherwise, the author's pre-trained model is used to generate results. If only the code is available, try to adjust the parameters to get the best performance of the test data.

Besides, the deblurring performance is measured by the Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM) and Visual Information Fiedity (VIF). PSNR and SSIM are popular image quality evaluation metrics, which are widely used in deblurring result quality evaluation. While VIF [42] better reflects the human eye's perception mechanism of image quality in principle. But the implementation of its bionic model is very complicated, here is just a brief introduction.

We train our model on the GoPro dataset, and test the trained model on all images in the GOPRO test dataset, REDS test dataset, and TEXT dataset. When testing these three datasets, the models were trained using the GOPRO train dataset.PSNR, SSIM and VIF are used as indicators to measure performance.

1) RESULTS ON GOPRO DATASET

Firstly, the proposed model is trained on the GOPRO dataset and tested on test set, respectively. A group of recovery examples are shown in Fig.12. It can be seen that the restored images could obtain improved visual effects compared with the input blurred image. Furthermore, by observing the details in the enlarged local image region, our proposed method and Chos's method could achieve better recovery effect compared to other deblurring methods. Especially, for the deblurring result of image (3), the license plate numbers in the enlarged areas are sharper than other compared approaches, with higher image resolution and richer detail information.

Moreover, in order to evaluate the objective deblurring quality, all of the GOPRO test set images are deblurred to calculate the mean evaluation values. The corresponding PSNR, SSIM and VIF mean metrics for these deblurred images are calculated respectively, which are shown in Table 1.

TABLE 1. Quantitative results on GoPro test dataset.The best results are boldfaced and the second best results are underlined.

Model	PSNR(dB)	SSIM	VIF	RunningTime(s)
Tao	30.26	0.9050	0.4872	3.32
ChenL	29.14	0.8971	0.4931	0.31
ChenY	30.02	0.8924	0.5782	0.33
Kupun	30.98	0.8912	0.5056	**0.14**
ChoS	**31.63**	**0.9144**	**0.6235**	<u>0.21</u>
Ours	<u>31.26</u>	<u>0.9131</u>	<u>0.6147</u>	0.45

It can be seen that the proposed approach has excellent results in all of PSNR, SSIM and VIF evaluation values, which reach the second best assessment results compared with other deblurring techniques. As the most advanced deblurring method nowadays, ChoS' method is a little bit better than our method. The last column in Table 1 illustrates the running time. And the proposed approach could achieve a balance performance of delubrring quality and algorithm efficiency. The performance improvement proves the effectiveness of the proposed approach.

2) RESULTS ON REDS DATASET

Besides, the REDs dataset was also tested by using the training model of the GOPRO dataset. Another group of deblurring results are illustrated in Fig.13. Similarly, the local detail regions are enlarged below the restored image to show the performances of the deblurring methods more clearly. It can be seen that for the 4 selected deblurred image from REDs dataset, the deblurring results obtained by the proposed method could have better visual perception with higher detail resolution and edge definition owing to the constructed HFAR module and designed generator loss. Compared with other methods, it could obtain superior subjective image deblurring evaluation effects.

Moreover, the PSNR and SSIM values of the selected 4 deblurring images in Fig. 13 are calculated, which are shown in Table 2 and Table 3. To provide more intuitive visual experience, the bar charts of PSNR and SSIM values in Table2 and Table 3 are illustrated in Fig.14.

It can be seen that the objective assessment values of ChoS' method and our proposed method are more prominent than other approaches with a certain equivalent level. For the average evaluation value, ChoS' method is a little bit better than our method and the proposed method is the second best

FIGURE 12. Visual comparisons of different methods on some blurred images from the GOPRO test datasets,from top to bottom: (a) Input, (b) Tao *et al.*,(c) ChenL *et al.*, (d) ChenY *et al.*,(e) Kupun *et al.*, (f) ChoS *et al.*, (g) ours and (h) Groundtruth.

FIGURE 13. Visual comparisons of different methods on some blurred images from the REDs test datasets, from top to bottom: (a) Input, (b) Tao *et al.*,(c) ChenL *et al.*, (d) ChenY *et al.*,(e) Kupun *et al.*, (f) ChoS *et al.*, (g) ours and (h) Groundtruth.

TABLE 2. PSNR value comparison of comparative images of REDs test dataset. The best results are boldfaced and the second best results are underlined.

image	Tao	ChenL	ChenY	Kupun	ChoS	Ours
(1)	27.76	28.71	28.78	27.41	29.06	29.38
(2)	28.67	29.33	28.47	26.25	30.27	30.06
(3)	27.45	28.10	27.99	26.72	28.70	28.72
(4)	29.21	29.01	28.26	27.89	29.39	29.17
ave	28.27	28.79	28.38	27.07	**29.36**	29.33

TABLE 3. SSIM value comparison of comparative images of REDs test dataset. The best results are boldfaced and the second best results are underlined.

image	Tao	ChenL	ChenY	Kupun	ChoS	Ours
(1)	0.794	0.803	0.820	0.787	0.865	0.858
(2)	0.828	0.818	0.846	0.813	0.879	0.872
(3)	0.815	0.836	0.827	0.810	0.859	0.867
(4)	0.864	0.854	0.871	0.861	0.913	0.907
ave	0.825	0.828	0.841	0.818	**0.879**	0.876

TABLE 4. Quantitative results on REDs test dataset.The best results are boldfaced and the second best results are underlined.

Model	PSNR(dB)	SSIM	VIF
Tao	28.31	0.8301	0.4469
ChenL	28.84	0.8348	0.4501
ChenY	29.19	0.8517	0.5292
Kupun	27.66	0.8293	0.4418
ChoS	**29.90**	**0.8832**	0.5875
Ours	29.85	0.8720	**0.5906**

value. It confirms the effectiveness and universality of the proposed FCA and ESA module models.

Not for a general discussion, in order to further test the applicability of the deblurring framework, 1000 more blurred images from the REDS dataset are restored to calculate the mean PSNR, SSIM and VIF evaluation values of all deblurred images, respectively. The results are shown in Table 4.

For the test on REDs dataset, the methods of the presented technique in this paper and ChoS are the best two deblurring approaches on objective assessment. The ChoS' method performs slightly better in PSNR and SSIM values than our method. However, the VIF value of our proposed method performs the best. Experimental results have demonstrated that the proposed method has excellent applicability and validity on various different images.

3) RESULTS ON TEXT DATASET

In addition, the TEXT dataset was also tested by utilizing the training model of the GOPRO dataset. Compared with natural image, text image has a single color and a relatively simple structure. The restored images of different methods are shown

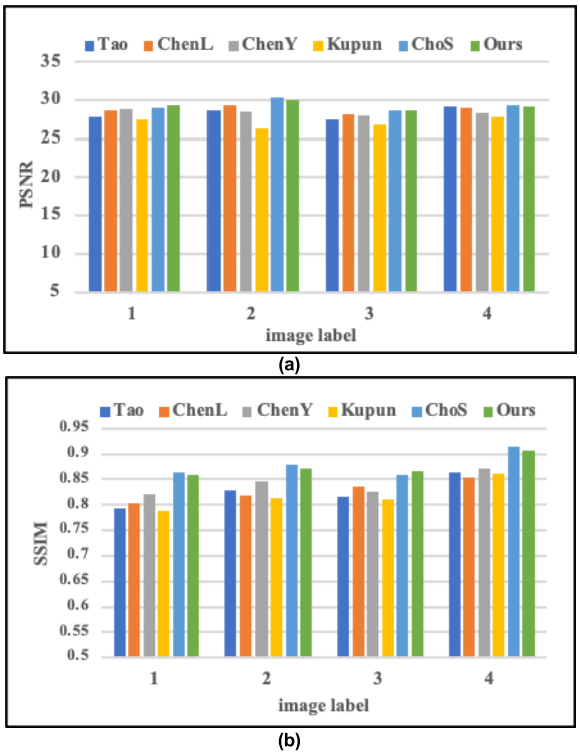


FIGURE 14. Bar chart of PSNR and SSIM values of comparative images. (a) PSNR, (b) SSIM. Test on the REDS data set.

in Fig.15. Most of the deblurring results tend to be unpleasant and still blurred, which indicate that the corresponding training models are ineffective on the deblurring application on text character image. The test results of our model on the text dataset are optimal and obtain a superior deblurring performance, especially for the restoration of edge regions of printed English characters owing to construction of ESA module and the FCA module. These two introduced modules could extract texture details and high frequency information effectively even for the blurred text images.

Also, the mean values of PSNR, SSIM, and VIF metrics of the blurred images in Fig.15 are calculated, respectively. The evaluation results are shown in Table 5. As can be seen from the data in Table 5, several methods have high SSIM value while the corresponding PSNR and VIF values are relatively low. This is related to the characteristics of text images, which are very different from natural images. Our proposed method achieves the best performance results for all of the three assessment metrics.

D. ABLATION STUDY

1) ABLATION STUDY ON NUMBER OF ATTENTION RESIDUAL BLOCKS

The number of attention residual blocks in the model will affect the speed and training effect of the model training. The residual number ablation study results of the GOPRO dataset are shown in Table 6 and the corresponding visual curve images are shown in Fig.16.

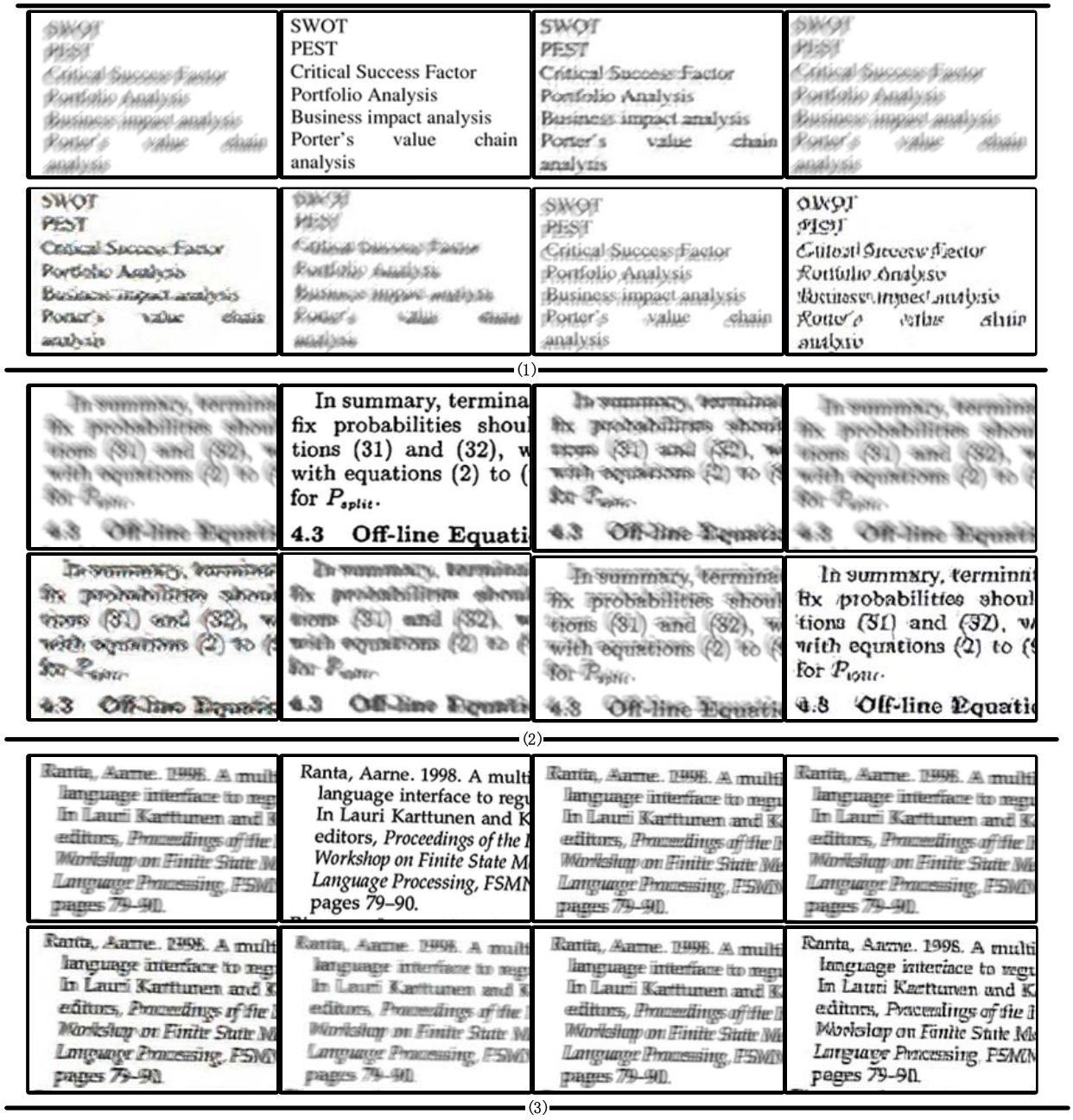


FIGURE 15. Visual comparisons of different methods on some blurred images from the Text test datasets, From left-top to right-bottom: Blurry images, ground-truth images, and the resultant images obtained by Tao, ChenL, ChenY, Kupun, Chos, and Ours, respectively.

The ablation study results show the changes of PSNR, SSIM value and running time as the number of attention residuals changes. It can be seen that as the number of attention residual blocks increases, there is no obvious change for running time. However, the values of PSNR and SSIM increased significantly as the number of attention residual blocks changed from 8 to 10 and tended to be stable when the block number is more than 10. Considering the amount of model parameters and model performance comprehensively, 10 attention residual blocks are selected as the optimal number.

2) ABLATION STUDY ON ATTENTION MECHANISMS

To verify the necessity of the attention mechanisms, different ablated versions of the propose model are formed by removing some of the attention modules:

'Nothing':removeing all attention modules;

'FCAGAN':removing all except the ESA modules;

'ESAGAN': removing all except the FCA modules;

'HFARGAN': the proposed model with both FCA and ESA modules.

The results in PSNR value of the ablated versions on the GoPro dataset are listed in Table 7. The 'HFARGAN' refers

TABLE 5. Quantitative results on Text test dataset.The best results are boldfaced and the second best results are underlined.

Model	PSNR(dB)	SSIM	VIF
Tao	<u>25.63</u>	<u>0.9214</u>	<u>0.5354</u>
ChenL	24.30	0.9014	0.5288
ChenY	24.13	0.8939	0.4482
Kupun	24.86	0.9168	0.4873
ChoS	24.01	0.9033	0.4769
Ours	**28.78**	**0.9795**	**0.5753**

TABLE 6. Ablation study on number of attention residual blocks.

Number	PSNR(dB)	SSIM	RunningTime(s)
8	29.93	0.8692	0.409
9	30.81	0.8975	0.436
10	31.26	0.9133	0.447
11	31.32	0.9142	0.457
12	31.28	0.9167	0.476
13	31.35	0.9131	0.482

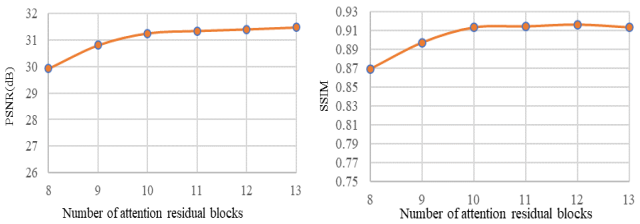


FIGURE 16. PSNR and SSIM line graphs of different numbers of attention residual blocks.

to the proposed model with all attention modules. It can be seen that the introduction of marginal space attention or Fourier channel attention has significantly improved the restoration effect. Compared with 'Nothing', both 'FCA-GAN' and 'ESAGAN' have improved by more than 0.9dB. Combining the two kinds of attention, the restoration effect can be further improved. However, the gain benefit of the two attention mechanisms cannot be superimposed and the further improvement is about 0.5dB. One plausible cause of such a minor performance improvement when combining both attention mechanisms is that there exist redundancy between the ESA and FCA modules in terms of their functions.

3) ABLATION STUDY ON LOSS FUNCTION

The edge loss is added in the construction of the loss function and the following experiments are performed to verify the effect of the part with or without Edge Loss.

'Loss': Without Edge Loss;

'Loss+':With Edge Loss;

The loss function ablation study results of the GOPRO dataset are shown in Table 8. The mean values of PSNR, SSIM, and VIF of the blurred image are calculated,

TABLE 7. Ablation study on attention mechanisms. Best results are boldfaced.

model	FCA	ESA	PSNR(dB)
Nothing	✗	✗	29.72
FCAGAN	✓	✗	30.70
ESAGAN	✗	✓	30.64
HFARGAN	✓	✓	**31.26**

TABLE 8. Ablation study on number of attention residual blocks.

Loss function	PSNR	SSIM	VIF
Loss	30.35	0.8938	0.4982
Loss+	31.26	0.9131	0.6147

FIGURE 17. Visual comparisons of with or without Edge Loss on some blurred images, From left-top to right-bottom: Blurry images, ground-truth images, and the resultant images obtained by Loss, Loss+, respectively.

respectively. It can be seen that the introduce of edge loss into the whole loss function contribute to the improvement of the deblurring result significantly.

Some examples of deblurring results of Loss and loss+ are given below and shown in Fig.17. It can be seen that the texture of the deblurred image with edge loss is better on the visual effects of edge sharpness and detail richness.

V. CONCLUSION

This paper proposes a HFARGAN network with a novel attention module for motion blur image restoration. The edge information and spectrum information of the feature map is integrated into the attention module to form the ESA module and the FCA module, which are used to guide the network to learn more image texture information to improve the restoration effect of the blurred image. The feasibility and necessity of introducing edge information and frequency domain information into the network are described in the paper. Meanwhile, two modules of ESA and FCA are introduced in details. Unlike most deblurring networks, the generator loss

of the HFARGAN network adds edge loss to further improve the effective of the restored image. The model training results are compared with multiple methods and the experimental results show that the network proposed in this paper achieves better performance both subjectively and objectively. ESA and FCA ablation experiments are carried out to demonstrate the effectiveness of the proposed ESA and FCA modules. We believe that the designed ESA and FCA modules can also be used for other image processing and visual tasks, which we will explore in the future.

REFERENCES

[1] C. Chun-Lei, Y. Dong-Yi, and C. Zhao-Jiong, "Blind image deblurring via multi-local kernels' fusion," *Acta Photonica Sinica*, vol. 47, no. 10, 2018, Art. no. 1010002.

[2] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *Proc. ACM SIGGRAPH Papers*, 2006, pp. 787–794.

[3] L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1107–1114.

[4] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2010, pp. 157–170.

[5] D. Perrone and P. Favaro, "Total variation blind deconvolution: The devil is in the details," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2909–2916.

[6] C. Sun, J. Mao, Y. Hu, and W. Sheng, "Research on image deblurring based on RGB three-channel adaptive Wiener filter," *Comput. Meas. Control*, 2019, pp. 215–219.

[7] S. Cho and S. Lee, "Fast motion deblurring," in *Proc. ACM SIGGRAPH Asia Papers*, 2009, pp. 1–8.

[8] S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 495–502.

[9] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1964–1971.

[10] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, 2008.

[11] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.

[12] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8174–8182.

[13] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. H. Lau, and M.-H. Yang, "Dynamic scene deblurring using spatially variant recurrent neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 2521–2529.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[15] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 8183–8192.

[16] L. Chen, Q. Sun, and F. Wang, "Attention-adaptive and deformable convolutional modules for dynamic scene deblurring," *Inf. Sci.*, vol. 546, pp. 368–377, Feb. 2021.

[17] Y. Chen, G. Cui, J. Zhang, and J. Zhao, "A deep motion deblurring network using channel adaptive residual module," *IEEE Access*, vol. 9, pp. 65638–65649, 2021.

[18] L. Xu, S. Zheng, and J. Jia. *Unnatural L_0 Sparse Representation for Natural Image Deblurring Supplementary Material*. Accessed: Aug. 2021. [Online]. Available: http://www.cse.cuhk.edu.hk/~leojia/projects/l0deblur/l0deblur_supp.pdf

[19] A. Chakrabarti, "A neural approach to blind motion deblurring," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2016, pp. 221–235.

[20] Y. Chang, L. Yan, X. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.

[21] Y. Quan, P. Lin, Y. Xu, Y. Nan, and H. Ji, "Nonblind image deblurring via deep learning in complex field," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 14, 2021, doi: 10.1109/TNNLS.2021.3070596.

[22] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 517–532.

[23] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 2472–2481.

[24] X. Tao, H. Gao, R. Liao, J. Wang, and J. Jia, "Detail-revealing deep video super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4472–4480.

[25] Z. Liu, R. A. Yeh, X. Tang, Y. Liu, and A. Agarwala, "Video frame synthesis using deep voxel flow," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4463–4471.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Cham, Switzerland: Springer, 2015, pp. 234–241.

[27] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556v6*.

[29] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4641–4650.

[30] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8878–8887.

[31] S. Zhao, Z. Zhang, R. Hong, M. Xu, Y. Yang, and M. Wang, "FCL-GAN: A lightweight and real-time baseline for unsupervised blind image deblurring," 2022, *arXiv:2204.07820*.

[32] H. Tomosada, T. Kudo, T. Fujisawa, and M. Ikehara, "GAN-based image deblurring using DCT loss with customized datasets," *IEEE Access*, vol. 9, pp. 135224–135233, 2021.

[33] H. Feng, J. Guo, H. Xu, and S. S. Ge, "SharpGAN: Dynamic scene deblurring method for smart ship based on receptive field block and generative adversarial networks," *Sensors*, vol. 21, no. 11, p. 3641, May 2021.

[34] I. J. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[35] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*.

[36] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of Wasserstein GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[37] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

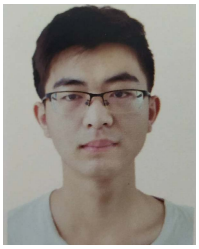
[38] R. A. Khan, E. Dinet, and H. Konik, "Visual attention: Effects of blur," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3289–3292.

[39] C. Qiao, D. Li, Y. Guo, C. Liu, T. Jiang, Q. Dai, and D. Li, "Evaluation and development of deep neural networks for image super-resolution in optical microscopy," *Nature Methods*, vol. 18, no. 2, pp. 194–202, Feb. 2021.

[40] S. Nah, S. Baik, S. Hong, G. Moon, S. Son, R. Timofte, and K. M. Lee, "NTIRE 2019 challenge on video deblurring and super-resolution: Dataset and study," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–10.

[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[42] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.



JITONG ZHANG received the B.S. degree in electronic information science and technology from China Jiliang University, in 2020. He is currently pursuing the master's degree in electronic science and technology with Hangzhou Dianzi University. His research interest includes image restoration techniques, such as image deblurring.

GUANGMANG CUI received the B.S. degree from Zhejiang University, Hangzhou, China, in 2011, and the Ph.D. degree in optical engineering from the State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, in 2016. He is currently a Lecturer with Hangzhou Dianzi University. His research interests include optical imaging systems, image processing, and deep learning applications in computer vision.

JUFENG ZHAO received the B.S. degree from Zhejiang University, Hangzhou, China, in 2008, and the Ph.D. degree from the State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, in 2013.

Since 2013, he has been an Associate Professor with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou. His research interests include the representation and analysis of visual images, including digital image processing, such as image restoration and quality assessment.

YING CHEN received the B.S. degree from the Huaiyin Institute of Technology, Hangzhou, China, in 2019. She is currently pursuing the master's degree in electronic science and technology with Hangzhou Dianzi University. Her research interests include image processing and deep learning applications in computer vision.