IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 30, 2021, accepted July 21, 2021, date of publication July 26, 2021, date of current version August 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099803

PAMSGAN: Pyramid Attention Mechanism-Oriented Symmetry Generative Adversarial Network for Motion Image Deblurring

ZHENFENG ZHANG
Department of Physical Education Training, Zhengzhou University of Aeronautics, Zhengzhou 450015, China
e-mail: publicgj@163.com

This work was supported by the Project of Research and Development of Video Acquisition and Analysis System for Kayak (slalom) Motion, in 2020.

ABSTRACT Motion blur is a common problem in optical imaging, which is caused by the relative displacement between the subject and the camera in the exposure process of the camera. This can result in motion blur of the acquired image, reduce the image resolution and affect the imaging quality. Motion blur image restoration technology uses the existing motion blur image to restore the clear image through the modeling of imaging physical process and mathematical solution without re-photographing the target scene. It has an important application value in the civil and military fields. Solving the problem of motion blur caused by camera jitter and object motion during camera imaging is a very challenging problem. When the popular generative adversarial network model is directly applied to the image blur blind removal task, serious pattern collapse phenomenon will occur. In this paper, we propose a novel motion image deblurring model based on pyramid attention mechanism-oriented symmetry generative adversarial network. This new method does not need to predict the fuzzy kernel of the blurred images, and can directly realize the blind removal of image motion blur. Based on the original CycleGan, the network structure and loss function of the symmetry generative adversarial network are improved. The accuracy of blind removal of motion images is improved, and the stability of the network is greatly enhanced in the case of limited samples. The generative network adopts the encoding and decoding structure, and introduces the feature pyramid attention mechanism. The combination of multi-scale pyramid features and attention mechanism can capture more rich advanced features to improve the model performance. In the experiment, the RMSProp algorithm is used to optimize the network training. Finally, a clear image is obtained through network adversarial training between generative and discriminant network. Experimental results on the related image blur benchmark datasets show that the restoration quality of the proposed method is higher in terms of subjective and objective evaluation. Meanwhile, the restoration results can achieve better results in subsequent object detection tasks.

INDEX TERMS Motion image deblurring, pyramid attention mechanism, symmetry generative adversarial network, RMSProp, CycleGAN.

I. INTRODUCTION

The limitations of the imaging system, the complexity of the environment, the dynamic and uncooperative nature of the object and many other factors will lead to the degradation of the acquired image with strong noise, low quality and distortion [1]. Image blurring is a typical form of image degradation. Factors such as focal length, camera shake, and the movement of the object can all contribute to the image blurring [2].

Research on the restoration of blurred images will help to improve the visual quality of images from the perspective of human vision [3]. From the perspective of practical application, it will help to improve the robustness of visual application, so it has important practical application value.

The traditional image motion blur blind restoration algorithm needs to use the pre-designed calculation method to generate the fuzzy kernel corresponding to the motion field

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

of the fuzzy sample. But the precise motion model of image blind restoration has great uncertainty and it is difficult to predict the precise fuzzy kernel close to the shooting scene [4].

In the process of scene imaging acquisition, for the traditional camera exposure mode, the camera shutter is always open during the exposure process. It is equivalent to defining a box filter in the time domain and convolving with the scene image. This mode can be approximated as a low-pass filter in the frequency domain, which destroys the high-frequency information in the image. At the same time, the external noise makes the motion image blur restoration with deconvolution operation become ill-conditioned. The commonly used fuzzy image restoration methods include Wiener filter, inverse filter, constrained least square method, etc,. In recent years, researchers have proposed various image restoration algorithms and made great progress. However, due to the large amount of information missing when processing image blur restoration, the algorithm is not ideal [5], [6].

The early research on image blur restoration often assumes that the degradation model is known and mainly focuses on the mathematical model [7]. The method of image blur restoration is generally based on the following blur degradation model:

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

where, B and I represent blurred image and clear image respectively. $\otimes$ is the convolution operation. K is the fuzzy kernel or the point spread function (PSF). N is the noise. The common method of image blur restoration is to estimate the fuzzy kernel from the blurred image, and the common methods include regularization method, edge prediction, etc. Then it uses the fuzzy kernel and the blurred image to finish the deconvolution, so as to restore the clear image I.

According to whether the fuzzy kernel is known, motion blur image restoration can be divided into two categories: non-blind image restoration and blind image restoration. Non-blind image restoration is performed on the premise that some prior knowledge of the blur degradation model is known, that is, the fuzzy kernel is known. In the process of imaging, when the camera motion trajectory and other conditions are unknown, the process of motion blur image restoration with unknown fuzzy kernel is called blind image restoration. Due to the lack of prior information, the image restoration has higher morbidity than the non-blind image restoration. In the above image blur degradation modeling, all pixel points in the image are convolved with the same fuzzy kernel. Levin *et al.* [8] proposed the hypothesis of global uniform fuzzy kernel, which could not fully describe the process of image motion blur caused by camera shake and other reasons. In real life, non-uniform blur images are more common. Therefore, this paper focuses on solving the problem of non-uniform blind blur image restoration.

It is a serious ill-posed problem to find the corresponding fuzzy kernel for each pixel value. Therefore, many methods take image statistics information as a priori information and try to build blur degradation model.

Recently, Convolutional Neural Networks (CNNs) [9] have been widely applied to various computer vision problems and have shown promising results, including the motion image blur restoration. Xu *et al.* [10] proposed a non-blind image deconvolution convolutional neural network for deblurring. They constructed the neural network according to the characteristics of the decomposition fuzzy kernel, which made the neural network need to be trained for different fuzzy kernels. Schuler *et al.* [11] proposed a blind deblurring method based on convolutional neural networks. The estimation of the fuzzy kernel and the clear image were obtained by using the iterative method in a Coarse-to-Fine way. They convolved the clear image with the fuzzy kernel generated by the Gaussian model to synthesize the fuzzy and clear image pairs as data sets. However, it was mentioned in this paper that the neural network was not effective in restoring the fuzzy image synthesized by larger fuzzy kernel. Similar to the work of Couzinie-Devy [12], Sun *et al.* [13] proposed an orderly deblurring method. Firstly, they synthesized a pair of fuzzy image blocks (obtained by convolution of 73 linear fuzzy kernels with different sizes and directions) and the corresponding clear image blocks. Secondly, they trained a classified convolutional neural network to measure the motion fuzzy kernel probability distribution corresponding to the fuzzy image blocks. Thirdly, the motion blur kernel was obtained by optimizing the Markov random field model. Finally, based on the above estimated motion blur kernel, the deblurring image was obtained by traditional deconvolution algorithm.

Chen *et al.* [14] proposed a 2-layer convolutional neural network to extract image features, and Tikhonov penalty model was used to predict fuzzy kernel, and then a clear image was restored. However, it did not perform well in dealing with large scale fuzzy samples. Xu *et al.* [15] estimated the fuzzy kernel with the H1-norm as the loss function, and then made the blurred image clear with the hyper-Laplacian penalty. Although this method made reasonable use of the ability of the neural network to obtain the edge information of the image, the restoration effect was obviously reduced when the background of the image was more complex. Wieschollek *et al.* [16] also adopted the idea of fuzzy kernel prediction in convolutional neural network, and made Fourier transform for the image before prediction. So that a clear image could be finally recovered by using the fuzzy kernel prediction. However, this method would occupy more training resources and took longer training time. Yan and Shao [17] proposed a pretrained deep neural network and a general regression neural network (GRNN) to obtain fuzzy classification of images through deep neural network, predict fuzzy related parameters, and finally restored clear images. This method performed well in Gaussian blur, linear motion blur and other fuzzy environments. Li *et al.* [18] skipped the intermediate process of fuzzy kernel prediction and only used convolutional neural network to get clear image samples. The singular values of the fuzzy kernel were decomposed, and the results were input into the deconvolution neural network to recover a clear image. This method has better performance

on image defocus and blur. Chakrabart [19] used CNN to deconvolve the blur image block to get the initial clear image, combined its Fourier coefficient with the discrete Fourier transform of the fuzzy image block. The fuzzy kernel was estimated by averaging the values of all the output image blocks, and the clear image was finally restored. This method could estimate the large scale fuzzy kernel well, but it lacked a more detailed solution to restore the details of the image.

Generative adversarial network (GAN), as a popular generative model, is widely used in the fields of image super resolution and image migration [20]. GAN is composed of generative network G and discriminant network D. Generative network G is used to generate false image close to real image. D is used to distinguish the false image from the real image. Through the alternate training optimization for G and D, the discriminant network D has been unable to distinguish the false image from the real image. The generative network G finally outputs the generated results of GAN.

Reference [21] provided the conditional generative adversarial architecture for image migration, also known as Pix2pix. Conditional generative adversarial networks learn mapping $G: x, z \rightarrow y$ from the observed image x and random noise vector z. In all GAN structures, the CycleGAN structure proposed by Zhu *et al.* [22] did not rely on any task-specific predefined similarity function between input and output, nor did it require the assumption that input and output must be embedded in the same low-dimensional space. CycleGAN performs well on a variety of tasks, including classifying composite photos, object reconstruction based on image edges, and face-age transformation. However, due to the loss of CycleGAN, the structure similarity of the sample generated by CycleGAN is very close to that of the original sample, which makes the performance of CycleGAN in the deblurring task not obvious.

Pan *et al.* [23] proposed a physics-based Generative Adversarial model for image restoration and beyond. The proposed model was trained in an end-to-end fashion and can be applied to a variety of image restoration and low-level vision problems. But for the problem of motion blur caused by camera jitter and object motion during camera imaging, it showed not well.

Kupyn *et al.* [24] applied the conditional GAN to motion blur removal and proposed a DeBlurGAN model. But DeBlurGAN has many problems. Pattern collapse occurs frequently in output samples. A large part of the problem is that the conditional generative adversarial network is not stable enough and its generalization ability is insufficient.

In summary, a new end-to-end symmetry generative adversarial network model based on pyramid attention mechanism (abbreviated to PAMSGAN) is proposed to solve the motion image blur blind removal task to alleviate the irregular texture pattern collapse problems. The PAMSGAN model can directly restore the motion blur image to clear image without predicting the fuzzy kernel [25]. The PAMSGAN consists of two symmetric GAN structures $G \sim D_{I_{clear}}$ and $F \sim D_{I_{blur}}$. Wherein, $G \sim D_{I_{clear}}$ continuously generates clear images by generative network G, while discriminant network D is used to distinguish the samples generated by G from the original clear images. G is a fully convolutional network designed according to the pyramid attention network, and D is a dichotomous network including the convolutional layer and the fully connection layer. Through continuous alternate training of G and $D_{I_{clear}}$, G can generate a clear image close to the real motion blur image. The conditional generative adversarial network $F \sim D_{I_{blur}}$, which has the same structure as $G \sim D_{I_{clear}}$, is used to generate motion blur images of random fuzzy kernels from clear images. This cyclic symmetric structure makes the network less prone to gradient disappearance. And the training of this network is more stable.

Our main contributions are as follows:

1) This paper proposes a novel motion image deblurring model based on pyramid attention mechanism-oriented symmetry generative adversarial network. This new method does not need to predict the fuzzy kernel of the blurred images, and can directly realize the blind removal of image motion blur.

2) Based on the original CycleGAN, the network structure and loss function of the symmetry generative adversarial network are improved. The generative network adopts the encoding and decoding structure, and introduces the feature pyramid attention mechanism. Meanwhile, F and G are encouraged by training mapping G and F simultaneously with adding cyclic consistency loss. This loss is combined with the adversarial loss and the content loss to train the cyclic consistent feature transfer algorithm.

3) The combination of multi-scale pyramid features and attention mechanism can capture more rich advanced features to improve the model performance. The detailed network of FPA and Generative network is optimized. A comprehensive cyclic consistency loss is proposed.

4) In the experiment, the RMSProp algorithm is used to optimize the network training. The experiment results show that the PAMSGAN performs better for motion image de-blurring.

This paper is organized as follows. Section 2 detailed introduces the PAMSGAN model including fuzzy samples generation, symmetry GAN and loss function. Section 3 displays the comparison experiments from objective evaluation and subjective evaluation. There is a conclusion in section 4.

II. PROPOSED PAMSGAN

In order to verify the effect of the proposed de-blurring algorithm in the actual application scene, this paper compares the output results in the scene of object detection. The de-blurring algorithm flow is shown in figure 1, which is mainly divided into three steps: fuzzy sample generation, motion blur removal and object detection.

A. FUZZY SAMPLE GENERATION

The motion blur image can be seen as the result of the convolution between clear image and fuzzy kernel with adding

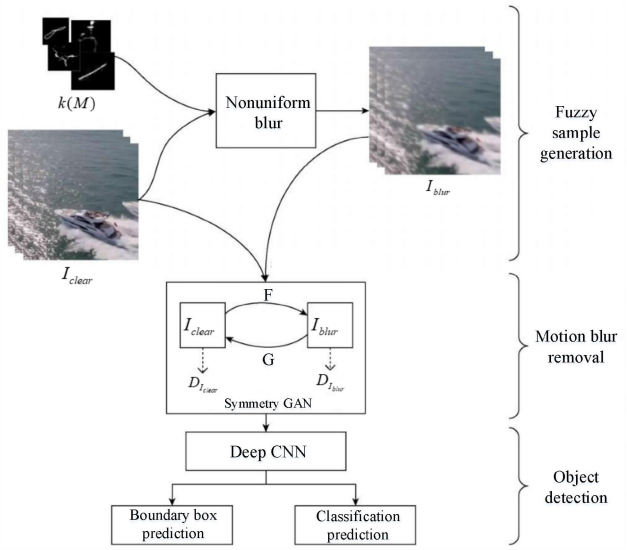

FIGURE 1. Flow chart of proposed method in this paper.

(a) clear image (b) motion blur image (c) corresponding fuzzy kernel (d) Visualized presentation with the removed background for (c).

FIGURE 2. Fuzzy sample image.

noise. Its mathematical expression is:

$$I_{blur} = k(M) * I_{clear} + N \tag{2}$$

where, I_{blur} is a blur image. $k(M)$ is a fuzzy kernel determined by the motion ground M. I_{clear} is clear image. * is the convolution. N is the added random noise. The clear image I_{clear}, the motion blur image I_{blur} and the corresponding fuzzy kernel $k(M)$ are shown in figure 2.

B. SYMMETRY GENERATIVE ADVERSATIVE NETWORK

There is not a one-to-one correspondence relation between clear image and motion blur image, and there are many approximate solutions to restore a motion blur image. Therefore, it is difficult to learn the features used for removing motion blur in the training process of generative adversative network. So a more stable algorithm is needed to reduce the pattern collapse problem in the GAN [26]. There is a certain mapping relationship between the clear image and blur image. Blur image can be regarded as a rendering of clear image. And this relationship can be learned by GAN. Although there is a lack of sufficient supervised data, the network can still be constrained at the macro level during the training. We select one image from I_{clear} and I_{blur} respectively. The two images are used for training the generative network G: $I_{blur} \rightarrow I_{clear}$, which makes the output clear image $\hat{i}_{clear} = G(i_{blur})$, $i_{blur} \in I_{blur}$. By means of adversarial network training, the generated clear image $\hat{i}_{clear}$ cannot be distinguished from the original clear image $i_{clear} \in I_{clear}$ when it is classified. Finally, this constraint can induce the output distribution of $\hat{i}_{clear}$ to continuously fit the empirical distribution $p_{data}(i_{clear})$. The optimal G can transform the blur image set I_{blur} into the deblurring image set $\hat{I}_{clear}$, which is nearly the same as the clear image set I_{clear}. However, it has been found in experiments that it is difficult to optimize a single adversarial object, standard procedures often lead to pattern collapse problems. And mapping all the input images to the output images cannot be improved significantly. More stability structures need to be added to the model to solve the above problems. Therefore, this paper attempts to utilize the mapping cyclic consistency in image transformation. For example, translating a sentence from English to Japanese and then from Japanese to English should result in a final sentence with similar features to the original sentence.

In abstractive, if there is one translator $G: X \rightarrow Y$ and another translator $F: Y \rightarrow X$, then G and F should be opposite of each other and can map to each other. Therefore, $F(G(i_{blur})) \approx i_{blur}$ and $G(F(i_{clear})) \approx i_{clear}$ are encouraged by training mapping G and F simultaneously with adding cyclic consistency loss [27]. This loss is combined with the adversarial loss and the content loss to train the cyclic consistent feature transfer algorithm. The structure of symmetry generative adversarial network is shown in figure 3.

1) GENERATIVE NETWORK WITH FEATURE PYRAMID ATTENTION

The Feature Pyramid Attention (FPA) module [28] realizes the function of extracting pyramid features at three different scales, and the module can fuse adjacent pyramid features at the three different scales. 3×3, 5×5 and 7×7 convolution kernels are used in the pyramid structure to better extract contextual information from the pyramid features at different scales. Because the size of the high-level feature graph output by the network is usually small, it will need little calculation cost when using the larger convolution kernel. Then, the pyramid structure gradually fuses the feature information of different scales, so that the context features of the three adjacent scales can be more accurately fused. After a 1×1 convolution operation, the original features extracted from a series of convolutional layers are multiplied pixel by pixel with those processed by the pyramid structure. In addition, the global pooling branch is introduced to join with the features after the multiplication mentioned above, which further improves the performance of FPA module. FPA module combines spatial pyramid structure with attention mechanism to extract accurate and dense features. The generated feature map of each layer is enhanced, and it provides better pixel-level attention for high-level feature map, thus effectively improving the performance of the model. Additionally, the context information fused by the module and the original features extracted by the

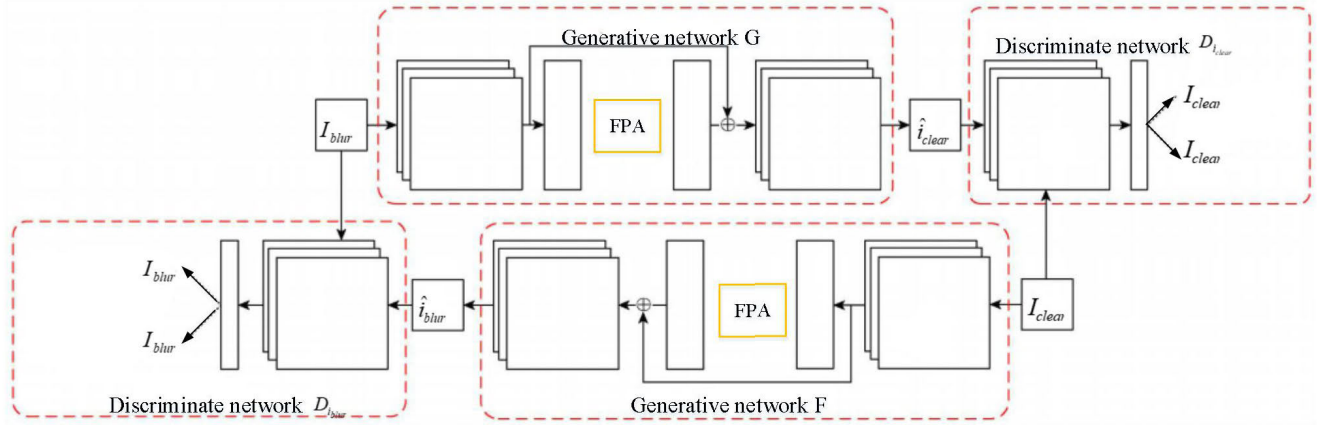

FIGURE 3. Symmetry GAN.

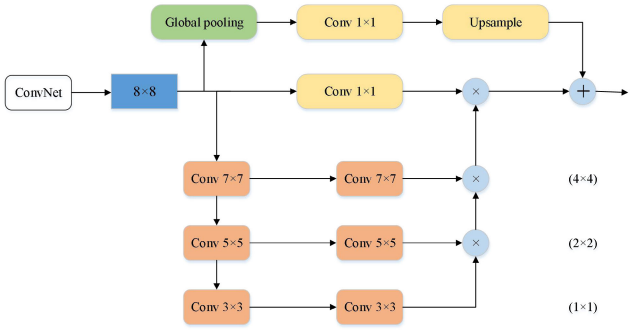

FIGURE 4. FPA module.

convolutional layer are multiplied by pixels, so the amount of calculation for this operation is not too large. The final module structure is shown in figure 4.

The clear image $\hat{i}_{clear}$ is restored with fuzzy kernel information and only the fuzzy image i_{blur} is given as input. The deblurring task is accomplished by a trained convolutional neural network (generative network). The generative network G can predict the corresponding clear image $\hat{i}_{clear}$ for each blur image i_{blur}. Another generative network F can predict the corresponding blur image $\hat{i}_{blur}$ for each clear image i_{clear}. The generative CNN architecture G is shown in figure 5. The generative network F adopts the same structure as G to ensure the symmetry of the two generative samples.

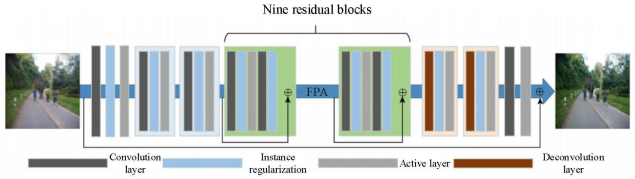

FIGURE 5. Generative network G.

The generative network consists of two strided convolution blocks with step size 0.5, nine ResBlocks, two deconvolution blocks and FPA. Each residual module consists of a convolution layer, instance regularization and ReLU activation. Different from the nine residual modules in CycleGAN, this paper adds two deconvolution modules after the residual module to enhance the detailed representation of the image generated by the generator. In addition, a global residual connection is added, so that CNN can learn the residual correction $i_{restore}$ of blur image i_{blur}. Finally, it outputs the clear sample $i_{clear} = i_{blur} + i_{restore}$. In the experiment, the residual modules are connected in the way shown in figure 5. The residual learning can improve the robustness and the training speed of the model. The adopted FPA is mentioned above.

2) DISCRIMINANT NETWORK

The discriminant network adopts the same architecture as PatchGAN [29], [30]. All convolutional layers except the last one use instance regularization and the LeakyRelu activation function. During the training period, the corresponding discriminators $D_{I_{clear}}$ and $D_{I_{blur}}$ are introduced and trained in an adversarial way.

3) LOSS FUNCTION

Due to the content loss of CycleGAN, the generated samples are very similar to the original samples in terms of the structure, as shown in figure 6.

FIGURE 6. Clear sample i_{clear}, blur sample generated by CycleGAN and the method in this paper.

The image is hardly changed after generative network F, and the blurring effect is not obvious. This suggests that

ultimately the entire network is not adequately trained. Therefore, the content loss of generative network F is removed in this paper to improve the diversity of generated fuzzy samples and make the whole network get out of the under-fitting state.

Loss function L_{total} includes adversarial loss $L_{adversarial}$, content loss $L_{G-content}$ and cyclic consistency loss $L_{cyclic-consistency}$, namely,

$$L_{total} = L_{adversarial} + \lambda_1 \cdot L_{G-content} + \lambda_2 \cdot L_{cycle-consistency} \tag{3}$$

In subsequent rich experiments, the evaluation indexes of the new model reach the highest when the super parameter $\lambda_1 = 100$ and $\lambda_2 = 10$.

(1) Adversarial loss $L_{adversarial}$. Adversarial loss is applied to the two mapping functions G and F to conduct adversarial training, that is,

$$L_{adversarial} = L_{GAN}(G, D_{I_{clear}}, I_{blur}, I_{clear}) + L_{GAN}(F, D_{I_{blur}}, I_{clear}, I_{blur}) \tag{4}$$

where, $L_{GAN}(G, D_{I_{clear}}, I_{blur}, I_{clear})$ is the adversarial loss of the mapping function $G: I_{blur} \rightarrow I_{clear}$ and its discriminator $D_{I_{clear}}$. In the training, $L_{GAN}(G, D_{I_{clear}}, I_{blur}, I_{clear})$ is shown in the red line in figure 7, which will become smaller with the continuous iteration training. And $L_{GAN}(F, D_{I_{blur}}, I_{clear}, I_{blur})$ is the adversarial loss of the mapping function $F: I_{clear} \rightarrow I_{blur}$ and its discriminator $D_{I_{blur}}$. The trend of $L_{GAN}(F, D_{I_{blur}}, I_{clear}, I_{blur})$ during training is shown in the blue line in figure 7. The loss value becomes smaller. The blur image i_{blur} generated by F has a wide distribution range. It is difficult to find a reasonable blur quality evaluation method to optimize F within the distribution range. Therefore, $L_{GAN}(F, D_{I_{blur}}, I_{clear}, I_{blur})$ appears oscillation phenomenon in the middle and late period of the training.

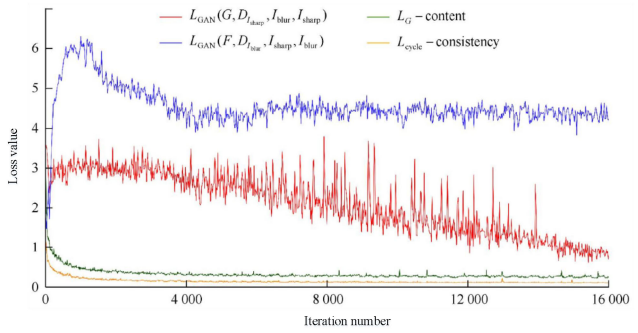

FIGURE 7. The trend of loss value during training.

(2) Content loss $L_{G-content}$. Using L1 loss or L2 loss on the original pixel as the only optimization objective will result in blurred artifacts on the generated image [31]. Therefore, this paper adopts the recently proposed perceived loss [32]. Perceived loss is a simple MSE loss based on the difference of CNN features between the generated image and the target image, as shown the in green line of figure 7. The definition of content loss is:

$$L_{G-content} = \frac{1}{P_{i,j} Q_{i,j}} \sum_{x=1}^{P_{i,j}} \sum_{y=1}^{Q_{i,j}} (\phi_{i,j}(I_{clear})_{x,y} - \phi_{i,j}(G_{\theta_G}((I_{blur}))_{x,y})^2 \tag{5}$$

where, $\phi_{i,j}$ is the feature mapping of the j-th transformation of the i-th max-pooling layer in the VGG19 network. It is pretrained on the ImageNet. $P_{i,j}$ and $Q_{i,j}$ are feature graphs of corresponding dimensions respectively. In this paper, VGG 3×3 convolution layer is used as activation function. Perceived loss is generally focused on restoring content details, whereas adversarial loss is focused on restoring texture details. Without perceived loss, it is difficult to converge to a meaningful state for the network trained by simple MSE on the image.

(3) Cyclic consistency loss $L_{cyclic-consistency}$. If G and F are random functions, adversarial training can learn mapping $G: I_{blur} \rightarrow I_{clear}$ and $F: I_{clear} \rightarrow I_{blur}$, which produces the outputs that are as similar as possible as the targets I_{clear} and I_{blur}, respectively. But if the network capacity is large enough, the network can map the same set of input images to any image in the target domain randomly. Any of these learned mappings can be fitted to the output that matches the target. To further reduce the possible space of mapping functions, the learned mapping functions should be cyclic consistency.

As shown in figure 8, the proposed model in this paper contains two mapping functions $G(I_{blur}) \sim I_{clear}$, $F(I_{clear}) \sim I_{blur}$, and the related symmetry discriminators $D_{I_{clear}}$ and $D_{I_{blur}}$. I_{blur} represents an input motion blur image. $D_{I_{clear}}$ is used to distinguish the i_{blur} through the clear image I_{clear} generated by G and the original clear image I_{clear}. Similarly, $D_{I_{blur}}$, $F(I_{clear}) \sim I_{blur}$ have similar structures. In order to further standardize the mapping, two cyclic consistency losses are introduced. If the blur image I_{blur} is converted to a clear image I_{clear} and back again, an original blur image I_{blur} can be obtained. As shown in figure 8(b), for each image i_{blur} of I_{blur}, two conversions can make i_{blur} return to the initial image, that is, $i_{blur} \rightarrow G(i_{blur}) \rightarrow F(G(i_{blur})) \approx i_{blur}$, which is called forward cyclic consistency. Similarly, as shown in figure 8(c), for each image i_{clear} of I_{clear}, G and F should also satisfy the backward cyclic consistency: $i_{clear} \rightarrow F(i_{clear}) \rightarrow G(F(i_{clear})) \approx i_{clear}$ as shown the orange line in figure 7. It makes constraint with cyclic consistency loss:

$$\begin{aligned} &L_{cyclic-consistency}(G, F) \\ &= E_{i_{blur} \sim p_{data}(i_{blur})}[||F(G(i_{blur})) - i_{blur}||_1] \\ &\quad + E_{i_{clear} \sim p_{data}(i_{clear})}[||F(G(i_{clear})) - i_{clear}||_1] \end{aligned} \tag{6}$$

C. OBJECT DETECTION

Based on the original Fast R-CNN [33], the object detection task in this experiment combines the regional proposal network (RPN) structure, which greatly accelerates the speed of boundary box prediction. Non-maximum suppression (NMS) [33] is applied to the object scores, and the

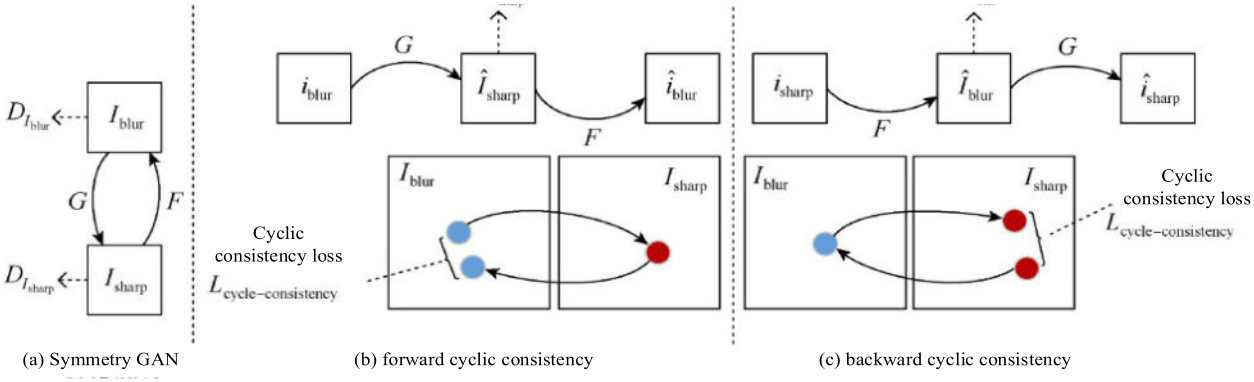

FIGURE 8. Cyclic consistency of feature transfer.

proposal regions with the output scores are finally output to the Fast R-CNN. Finally, two output vectors are obtained through the fully connection layer, one for Softmax classification and the other for boundary regression. It fine-tunes the regional proposals from the RPN.

III. EXPERIMENTS AND ANALYSIS

A. EXPERIMENTS SETTINGS

The PAMSGAN network is implemented using the PyTorch 0.4 framework. The experiment environment is: Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz, RAM 16GB, Windows 10, NVIDIA GTX1060 TI graphics card.

In this paper, RMSProp algorithm [35] is used to optimize the loss function. The global learning rate is 0.0001, the attenuation momentum is 0.9, and the batch size is 32. The network trains 25 epoch. The attenuation rate is 0.1, every 10 epoch occurs one attenuation.

We compare the PAMSGAN with three state-of-the-art motion image deblurring methods including DGAN [36], SharpGAN [37], GCGAN [38].

B. DATA SETS

(1) GoPro data set [39]: The GoPro data set disclosed by Nah *et al.* It simulates the generation process of real blurred images by adding video frames in high-speed videos shot by GoPro. The images produced in this way are closer to the real situation. The whole data set contains 3214 pairs of blur-clear images, the training set contains 2103 pairs, while the rest are used as the test set. This training set is used to train the new model in this paper, and the remaining test sets are selected to show some results.

(2) Lai data set [40]: Real and synthetic images of uniform and non-uniform fuzzy kernels are included, and some common image degradation methods are used. However, the non-uniform fuzzy kernel used to synthesize the blur image is generated according to the Kohlerl *et al.* They replay the motion tracks generated by a 6D camera to create a fuzzy kernel.

(3) ILSVRC2015 VID: The detailed information can be download from reference [41].

C. EVALUATION INDEX

Peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and running time are used to evaluate the emotion image deblurring.

(1) PSNR is often used to measure the quality of the processed image. It is expressed as formula (6).

$$PSNR = 10 \times \log_{10}(\frac{(2^i - 1)^2}{MSE}) \tag{7}$$

where i is the bit number per sample value. MSE is the mean square error between the original image and the processed image. The higher PSNR denotes the less distortion.

(2) SSIM is an index to measure the similarity of two images and it is often used as a method to measure video quality. It is widely used in super-resolution and image deblurring. Given two images A and B, the structural similarity between A and B is:

$$SSIM(A, B) = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)} \tag{8}$$

where μ_A and μ_B are the average values of image A and B respectively. σ_A^2 and σ_B^2 are the variance of image A and B respectively. σ_{AB}^2 is the covariance of A and B. $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are the stability constants. L is the dynamic range of pixel values. $k_1 = 0.01$, $k_2 = 0.03$. The range of structural similarity is 0∼1. When the two images are the same, the value of SSIM is equal to 1.

First, we make comparison with FPA and without FPA as shown in table 1. From table 1, we can know that the FPA plays an important role in feature extraction.

D. EXPERIMENT ON GoPro DATA SET

The results of experiment on GoPro data set are shown in figures 9-11. The objective evaluation indexes are shown in tables 2-4.

It can be noted that the DGAN and SharpGAN methods are not very successful in restoring sharp edges or areas where the blur is nonlinear, while the results of GCGAN contain very serious ringing phenomena. The results with PAMSGAN show that our method has the better subjective visual effect.

TABLE 1. Comparison with/without fpa (average value).

Method	GoPro data set		Lai data set		ILSVRC2015 VID	
	PSNR /dB	SSIM	PSNR /dB	SSIM	PSNR/ dB	SSIM
Without FPA	28.64 71	0.897 4	27.66 47	0.883 7	20.963 3	0.897 4
With FPA	30.25 87	0.916 3	29.79 42	0.901 24	22.810 9	0.925 5

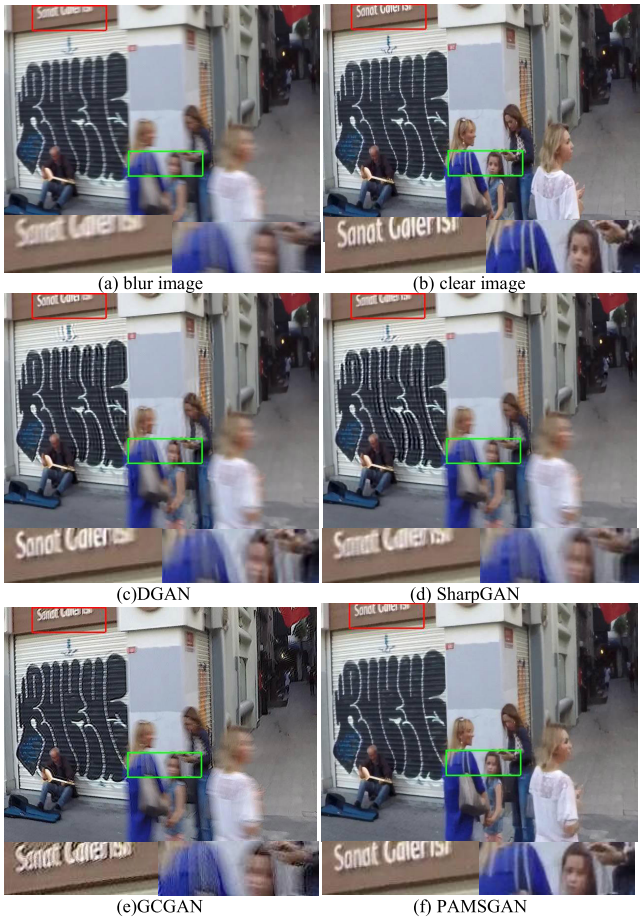

FIGURE 9. Comparison results with different methods for image1 in GoPro data set.

Fig. 10 demonstrates the deblurred results of five methods on a test image. In Fig. 10(b-e), the resulted image deblurring contains ghosts in the surrounding area because the input blurred image has a non uniform blur property. According to Fig. 10(f), those ghosts are removed with the proposed method. It verifies that the pyramid attention mechanism can remove the ghost phenomena and gradually restore the texture details.

Figure 11 demonstrates the deblurring results of images3. It can be seen in figure 11(c) that the DGAN produces small amount of ringing artifacts. It cannot recover sharp content because it does not consider the relative motion and depth of different objects in this image. Figure 11(d) and figure 11(f) show that SharpGAN and GCGAN significantly improve

FIGURE 10. Comparison results with different methods for image2 in GoPro data set.

TABLE 2. Quality assessment results for image1 in GoPro data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	23.6169	0.8017	3.22
SharpGAN	25.8398	0.8895	3.05
GCGAN	26.4354	0.8996	2.76
PAMSGAN	**29.9778**	**0.9285**	**1.53**

TABLE 3. Quality assessment results for image2 in GoPro data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	28.7215	0.8587	2.53
SharpGAN	28.8614	0.9146	1.67
GCGAN	29.0583	0.9215	1.57
PAMSGAN	**29.3268**	**0.9335**	**0.96**

the image degradation caused by ringing effect and noise. However, it is obvious that the fine details seem to be failed recovered. For example, the stitch in the bricks and the text on the license plate still lost some details. Finally, the results of PAMSGAN are shown in figure 11(f). It not only provided almost zero ringing artifacts while retaining satisfactory sharp details.

FIGURE 11. Comparison results with different methods for image3 in GoPro data set.

TABLE 4. Quality assessment results for image3 in GoPro data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	25.0935	0.8873	2.78
SharpGAN	27.2382	0.8916	1.96
GCGAN	28.2419	0.9135	1.75
PAMSGAN	**31.2325**	**0.9306**	**1.14**

TABLE 5. Quality assessment results for image1 in Lai data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	13.6198	0.5834	2.52
SharpGAN	13.8298	0.6075	1.85
GCGAN	13.9223	0.6263	1.47
PAMSGAN	**14.5516**	**0.6328**	**0.96**

E. EXPERIMENT ON LAI DATA SET

The results of experiment on Lai data set are shown in figures 12-14. The objective evaluation indexes are shown in tables 5-7.

We can see the discrepancy between the different methods. In Fig. 12(b-e), it can be found that the results obtained by the compared methods contain certain degrees of sharpness and ghosting. For instance, the results of DGAN and SharpGAN provide sharp image while suffering from the most serious

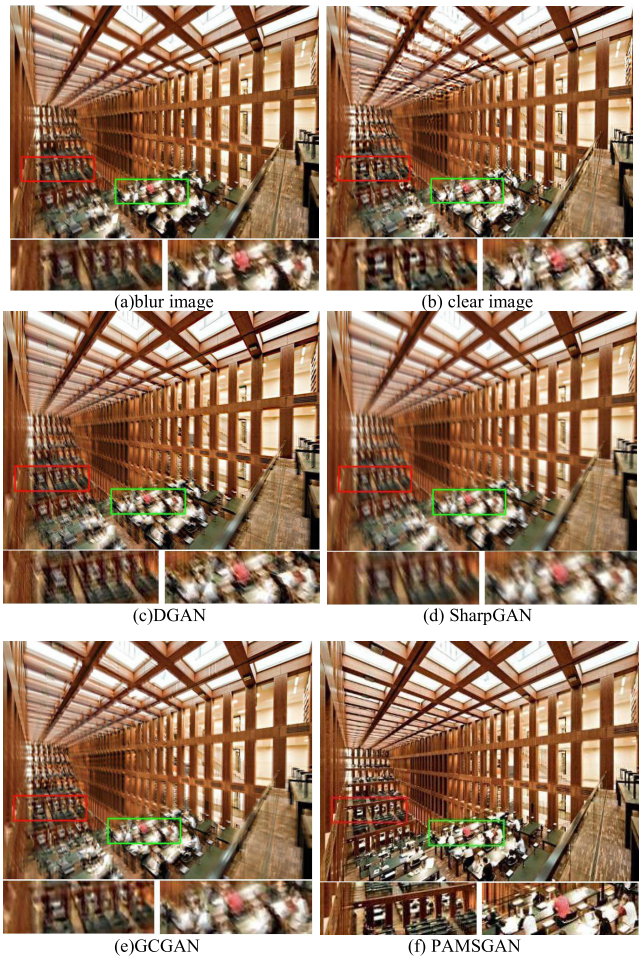

FIGURE 12. Comparison results with different methods for image1 in Lai data set.

TABLE 6. Quality assessment results for image2 in Lai data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	25.7209	0.7465	2.37
SharpGAN	26.2316	0.7613	1.94
GCGAN	27.9331	0.7921	1.58
PAMSGAN	**28.6542**	**0.8024**	**1.13**

ringing effect, while the result of GCGAN has less ringing effect but loses sharpness. As shown in Fig. 12(f), our method not only eliminates the ringing artifacts but retains reasonable level of sharpness. It again confirms the superior restoration capability of our method.

Table 5 shows the results of image quality evaluation. It can be seen from the quantization indexes in the table that the PAMSGAN is superior to the other three algorithms in terms of PSNR and SSIM. Therefore, the PAMSGAN achieves good results in subjective visual effects and objective evaluation results.

Figs. 13-14 shows the deblurring results of all the methods implemented on two selected test images including landscape

FIGURE 13. Comparison results with different methods for image2 in Lai data set.

FIGURE 14. Comparison results with different methods for image3 in Lai data set.

and portrait, respectively. Fig. 13 shows the superior deblurring ability of our method. In addition, Fig. 14 shows a qualitative comparison conducted on a portrait photo. It can be seen that the restored images of DGAN contain significifant ringing artifacts around the foreground region (i.e., the human face). This is probably due to the inaccurate kernel estimation caused by the interference of background. On the other hand, methods SharpGAN and GCGAN do not suffer this issue. Although end-to-end methods have the advantages of suppressing the artifacts and noise, the clarity of the restored images is still limited, particularly for the restoration of the edges of individual objects. In contrast to the state-of-arts, our method focuses on the restoration of high-frequency components. According to Fig. 14(f), our method successfully recovers the details of foreground in acceptable sharpness level while suppressing artifacts.

Observing the data, we find that the processing effect of our method is better than the other algorithms. Compared with other algorithms, our algorithm can effectively remove the blur caused by the motion of the object, and effectively restore the reduced edge information in the image, and restore a clear image with higher quality.

Figure 13 shows a subjective comparison conducted on a portrait photo. It can be seen that the restored images of SharpGAN has significant ringing artifacts in the human face.

TABLE 7. Quality assessment results for image3 in Lai data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	25.1966	0.7258	2.52
SharpGAN	25.8478	0.7437	1.76
GCGAN	26.9127	0.7855	1.44
PAMSGAN	**27.8836**	**0.8194**	**1.05**

DGAN and GCGAN still lose the overall sharpness. PAMSGAN method successfully recovers the details of foreground in acceptable sharpness level while suppressing artifacts.

F. EXPERIMENT ON ILSVRC2015 VID DATA SET

The results of experiment on ILSVRC2015 VID data set are shown in figures 15-17. The objective evaluation indexes are shown in tables 8-10.

It can be seen from figure 17 and Table 10 that in the three randomly selected video scenes, the PAMSGAN algorithm in this paper is superior to other algorithms in all indicators. In the process of learning image features, DGAN attaches too much importance to texture effect, resulting in wavy texture in the generated results and serious pattern collapse in some samples. SharpGAN is statistically effective at deblurring,

FIGURE 15. Comparison results with different methods for image1 in ILSVRC2015 VID data set.

TABLE 8. Quality assessment results for image1 in ILSVRC2015 VID data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	25.7123	0.8245	2.55
SharpGAN	28.7421	0.8574	1.72
GCGAN	29.6967	0.9052	1.43
PAMSGAN	**31.2532**	**0.9377**	**0.88**

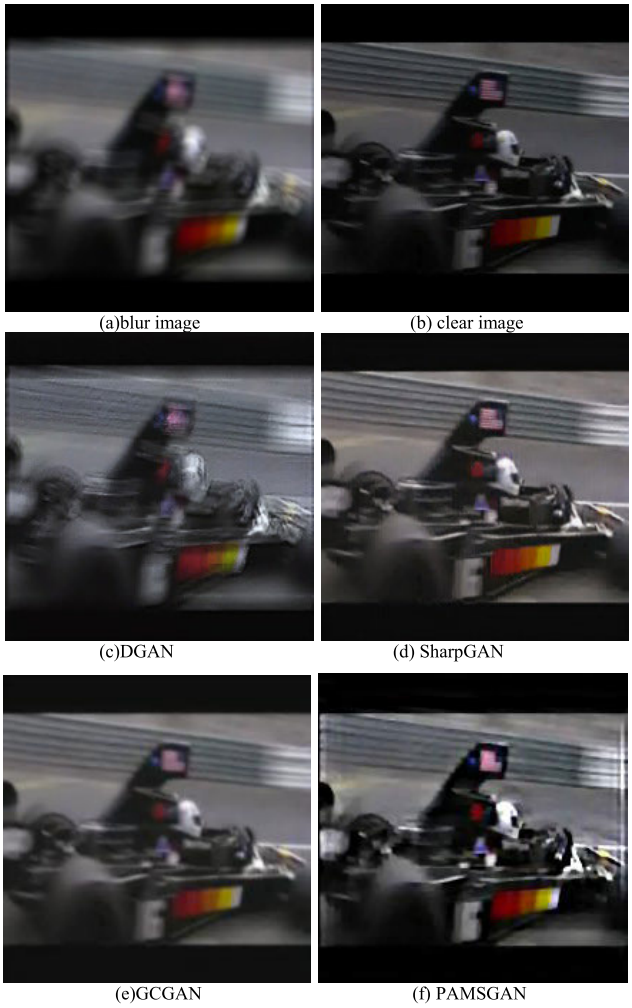

FIGURE 16. Comparison results with different methods for image2 in ILSVRC2015 VID data set.

TABLE 9. Quality assessment results for image2 in ILSVRC2015 VID data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	17.9635	0.6854	1.82
SharpGAN	19.8522	0.6931	1.59
GCGAN	22.4713	0.7254	1.37
PAMSGAN	**25.6417**	**0.7893**	**0.96**

and it is better than DGAN, but it is still not impressive. The SharpGAN algorithm only calculates the content loss of the image. When generating clear images, the local blurring of a single object is easy to affect the global image, which reduces the performance of the algorithm when there are multiple motion objects in the generated deblurring images. The GCGAN method can solve the problem of image defocus and blur. The recovered samples can be sharpened to a certain degree visually, so it has a more stable performance in high-speed motion scenes. However, GCGAN method needs to predict the fuzzy kernel in advance, which will consume more time and produce unstable results in a specific environment. However, the performance of the PAMSGAN algorithm in this paper is ideal. Because it introduces the cyclic consistency loss, which makes it easier to adjust the image to a clear state similar to the original image. From the above experiments, the time of PAMSGAN algorithm is the lowest, which can accelerate the deblurring effect.

FIGURE 17. Comparison results with different methods for image3 in ILSVRC2015 VID data set.

TABLE 10. Quality assessment results for image3 in ILSVRC2015 VID data set.

Method	PSNR/dB	SSIM	Time/s
DGAN	25.1107	0.8367	1.66
SharpGAN	25.4967	0.8543	1.37
GCGAN	29.8253	0.9135	0.92
PAMSGAN	**32.6987**	**0.9493**	**0.76**

G. COMPARISON OF OBJECT DETECTION PERFORMANCE IN MOTION BLUR ENVIRONMENT

The proposed algorithm in this paper can reduce the influence of motion blur on object detection to a certain extent. The object detection algorithm is optimized from the perspective of image enhancement to improve the average accuracy. Compared with the current popular object detection algorithms, the performance is shown in table 11.

It can be seen that Proposed+VGG16 has the higher mAP with 72.5% average precision. After deblurring with the proposed algorithm, then YOLOv3 is used for object detection, which can improve the precision to a certain extent. However, in the environment of motion blur+random noise, the mAP of YOLOv3(darknet53) and Faster R-CNN(VGG16) are greatly reduced. After the optimization of the proposed algorithm in this paper, the Proposed+VGG16 and Proposed+darknet53 have considerable improvement when dealing with the motion blur and random noise environment.

TABLE 11. Map comparison/%.

Method	Average precision	Motion blur sample	Random noise sample	Blur+noise
YOLOv3(darknet53)	52.7	31.3	39.9	29.7
Faster R-CNN(VGG16)	65.4	36.5	48.7	36.5
Proposed+darknet53	66.3	48.7	52.3	40.2
Proposed+VGG16	**72.5**	**59.6**	**55.8**	**49.5**

IV. CONCLUSION

In this paper, a novel motion image deblurring model based on pyramid attention mechanism-oriented symmetry generative adversarial network is proposed to solve the problems of motion blur caused by camera phase movement and object moving depth. In this paper, the cyclic symmetry principle is adopted to provide a stable solution to the problem of image motion blur. Conditional generative adversarial network can not only generate expected samples, but also improve the quality of samples by using the cyclic consistency. The results of this novel algorithm are applied to the object detection task and other relevant interference environments. The results show that the proposed symmetry GAN in this paper improves the robustness and accuracy of the motion deblurring to a great extent.

REFERENCES

[1] L. Teng, H. Li, and S. Yin, "Modified pyramid dual tree direction filter-based image denoising via curvature scale and nonlocal mean multigrade remnant filter," *Int. J. Commun. Syst.*, vol. 31, no. 16, p. e3486, Nov. 2018, doi: 10.1002/dac.3486.

[2] T. Goto, H. Senshiki, M. Sawada, H. Chen, and R. Aoki, "Image restoration method for non-uniform blurred images," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2018, pp. 1–6, doi: 10.1109/IST.2018.8577084.

[3] B. Yan, Y. Xiang, and G. Hua, "Improving the visual quality of size-invariant visual cryptography for grayscale images: An analysis-by-synthesis (AbS) approach," *IEEE Trans. Image Process.*, vol. 28, no. 2, pp. 896–911, Feb. 2019, doi: 10.1109/TIP.2018.2874378.

[4] H.-L. Yang, P.-H. Huang, and S.-H. Lai, "A novel gradient attenuation Richardson–Lucy algorithm for image motion deblurring," *Signal Process.*, vol. 103, pp. 399–414, Oct. 2014, doi: 10.1016/j.sigpro.2014.01.023.

[5] K. Ohkoshi, T. Goto, S. Hirano, and M. Sakurai, "Blind image restoration based on total variation regularization of blurred images," in *Proc. 1st IEEE Global Conf. Consum. Electron.*, Oct. 2012, pp. 621–622, doi: 10.1109/GCCE.2012.6379933.

[6] H. Su, H.-J. Feng, Z.-H. Xu, Q. Li, and Y.-T. Chen, "Restoration method of TDI remote sensing image based on optimization of numerical fidelity term," *J. Zhejiang Univ. Eng. Sci.*, vol. 52, no. 4, pp. 674–679, 2018, doi: 10.3785/j.issn.1008-973X.2018.04.009.

[7] X. Shi, L. Wang, X. Shao, H. Wang, and Z. Tao, "Accurate estimation of motion blur parameters in noisy remote sensing image," *Proc. SPIE*, vol. 9501, May 2015, Art. no. 95010X, doi: 10.1117/12.2176893.

[8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1964–1971, doi: 10.1109/CVPR.2009.5206815.

[9] S. Yin and H. Li, "Hot region selection based on selective search and modified fuzzy C-means in remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5862–5871, Sep. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9204442, doi: 10.1109/JSTARS.2020.3025582.

[10] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Dec. 2014, pp. 1790–1798.

[11] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, Jul. 2016, doi: 10.1109/TPAMI.2015.2481418.

[12] F. Couzinié-Devy, J. Sun, K. Alahari, and J. Ponce, "Learning to estimate and remove non-uniform image blur," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1075–1082, doi: 10.1109/CVPR.2013.143.

[13] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777, doi: 10.1109/CVPR.2015.7298677.

[14] C. Haoyu, T. Minggui, S. Boxin, W. YIzhou, and H. Tiejun, "Learning to deblur and generate high frame rate video with an event camera," 2020, *arXiv:2003.00847*. [Online]. Available: https://arxiv.org/abs/2003.00847

[15] X. Xu, J. Pan, Y.-J. Zhang, and M.-H. Yang, "Motion blur kernel estimation via deep learning," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 194–205, Jan. 2018, doi: 10.1109/TIP.2017.2753658.

[16] P. Wieschollek, B. Schölkopf, H. P. A. Lensch, and M. Hirsch, "End-to-end learning for image burst deblurring," in *Computer Vision* (Lecture Notes in Computer Science), vol. 10114, S. H. Lai, V. Lepetit, K. Nishino, and Y. Sato, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-54190-7_3.

[17] R. Yan and L. Shao, "Blind image blur estimation via deep learning," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1910–1921, Apr. 2016, doi: 10.1109/TIP.2016.2535273.

[18] S. Li, R. Fan, G. Lei, G. Yue, and C. Hou, "A two-channel convolutional neural network for image super-resolution," *Neurocomputing*, vol. 275, pp. 267–277, Jan. 2018, doi: 10.1016/j.neucom.2017.08.041.

[19] A. Chakrabarti, "A neural approach to blind motion deblurring," in *Computer Vision* (Lecture Notes in Computer Science), vol. 9907, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46487-9_14.

[20] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680, doi: 10.1145/3422622.

[21] P. Isola, J. 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. 5967–5976, doi: 10.1109/CVPR.2017.632.

[22] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[23] J. Pan, J. Dong, Y. Liu, J. Zhang, J. Ren, J. Tang, Y.-W. Tai, and M.-H. Yang, "Physics-based generative adversarial models for image restoration and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2449–2462, Jul. 2021, doi: 10.1109/TPAMI.2020.2969348.

[24] 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.*, Jun. 2018, pp. 8183–8192, doi: 10.1109/CVPR.2018.00854.

[25] Q. Luo and Q. Cai, "Blind motion image deblurring using two-frame generative adversarial network," (in Chinese), *J. Graph.*, vol. 40, no. 6, pp. 1056–1063, 2019, doi: 10.11996/JG.j.2095-302X.2019061056.

[26] W. Li, R. Chen, S. Xu, and W. Gong, "Blind motion image deblurring using nonconvex higher-order total variation model," *J. Electron. Imag.*, vol. 25, no. 5, Oct. 2016, Art. no. 053033, doi: 10.1117/1.JEI.25.5.053033.

[27] W. Tan, B. Yan, C. Lin, and X. Niu, "Cycle-IR: Deep cyclic image retargeting," *IEEE Trans. Multimedia*, vol. 22, no. 7, pp. 1730–1743, Jul. 2020, doi: 10.1109/TMM.2019.2959925.

[28] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," 2018, *arXiv:1805.10180*. [Online]. Available: https://arxiv.org/abs/1805.10180

[29] C. Li and M. Wand, "Precomputed real-time texture synthesis with Markovian generative adversarial networks," in *Computer Vision* (Lecture Notes in Computer Science), vol. 9907, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46487-9_43.

[30] A. Alanov, M. Kochurov, D. Volkhonskiy, D. Yashkov, E. Burnaev, and D. Vetrov, "User-controllable multi-texture synthesis with generative adversarial networks," 2019, *arXiv:1904.04751*. [Online]. Available: https://arxiv.org/abs/1904.04751

[31] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114, doi: 10.1109/CVPR.2017.19.

[32] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision* (Lecture Notes in Computer Science), vol. 9906, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46475-6_43.

[33] S. Yin, H. Li, and L. Teng, "Airport detection based on improved faster RCNN in large scale remote sensing images," *Sens. Imag.*, vol. 21, no. 1, pp. 1–13, Dec. 2020, doi: 10.1007/s11220-020-00314-2.

[34] S. Yin, H. Li, L. Teng, M. Jiang, and S. Karim, "An optimised multi-scale fusion method for airport detection in large-scale optical remote sensing images," *Int. J. Image Data Fusion*, vol. 11, no. 2, pp. 201–214, Apr. 2020, doi: 10.1080/19479832.2020.1727573.

[35] H. Zhang, N. Yang, W. Huangfu, K. Long, and V. C. M. Leung, "Power control based on deep reinforcement learning for spectrum sharing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4209–4219, Jun. 2020, doi: 10.1109/TWC.2020.2981320.

[36] B. Zhao, W. Li, and W. Gong, "Deep pyramid generative adversarial network with local and nonlocal similarity features for natural motion image deblurring," *IEEE Access*, vol. 7, pp. 185893–185907, 2019, doi: 10.1109/ACCESS.2019.2956947.

[37] H. Feng, J. Guo, and S. Shuzhi Ge, "SharpGAN: Receptive field block net for dynamic scene deblurring," 2020, *arXiv:2012.15432*. [Online]. Available: https://arxiv.org/abs/2012.15432

[38] H. Zhao, D. Wu, H. Su, S. Zheng, and J. Chen, "Gradient-based conditional generative adversarial network for non-uniform blind deblurring via DenseResNet," *J. Vis. Commun. Image Represent.*, vol. 74, Jan. 2021, Art. no. 102921, doi: 10.1016/j.jvcir.2020.102921.

[39] J. Li, D. H. Ye, T. Chung, M. Kolsch, J. Wachs, and C. Bouman, "Multi-target detection and tracking from a single camera in unmanned aerial vehicles (UAVs)," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 4992–4997, doi: 10.1109/IROS.2016.7759733.

[40] W. Lai, J. Huang, Z. Hu, N. Ahuja, and M. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1701–1709, doi: 10.1109/CVPR.2016.188.

[41] ImageNet. [Online]. Available: https://image-net.org/index.php

[42] S. Karim, Y. Zhang, S. Yin, A. A. Laghari, and A. A. Brohi, "Impact of compressed and down-scaled training images on vehicle detection in remote sensing imagery," *Multimedia Tools Appl.*, vol. 78, no. 22, pp. 32565–32583, Nov. 2019, doi: 10.1007/s11042-019-08033-x.

ZHENFENG ZHANG was born Hanan, China, in 1974. He is currently an Associate Professor with the Department of Physical Education Training, Zhengzhou University of Aeronautics, Zhengzhou, China. His research interests include physical education teaching and training, and physical image processing. He had published more than ten papers on the above research fields.