Received September 16, 2020, accepted October 1, 2020, date of publication October 7, 2020, date of current version October 20, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3029356*

# Multi-Frame Blind Restoration for Image of Space Target With FRC and Branch-Attention

**PEIJIAN ZHU, CHUNZHI XIE, AND ZHISHENG GAO**
School of Computer and Software Engineering, Xihua University, Chengdu 610039, China

Corresponding author: Zhisheng Gao (gzs_xihua@mail.xhu.edu.cn)

This work was supported in part by the Ministry of Education, Chunhui Project, under Grant Z2016149, in part by the Key Scientific Research Fund of Xihua University under Grant Z17134, in part by the Xihua University Key Laboratory Development Program under Grant szjj2017-065, in part by the Sichuan Science and Technology Program under Grant 2019YFG0108,2020YFS0351, and in part by the Xihua University Graduate Innovation Fund Research Project under Grant ycjj2018067.

**ABSTRACT** The random noise and anisotropic motion of atmospheric turbulence can cause different degradation patterns, which make images of space targets observed from ground-based stations severely disturbed. In recent years, benefit from the development of convolutional neural networks (CNNs), a large number of effective end-to-end methods were proposed to restore images. However, a single-frame method whose input is just a single image can hardly achieve a further improvement for the restoration image due to the diversified degradation patterns of space-target images. In this paper, we proposed a multi-branch network with a multi-frame input to restore space-target images. The multi-frame input contains space-target images which own different degradation patterns at different moments. In this way, we can fully use the complementary information between input frames. And in this network, two effective technologies are introduced: one is the full resolution convolution module which extracts features by using convolutional layers with different dilation rates to keep feature information complete; the other is the branch-attention module which is used to pass effective information between different branches of the network. Furthermore, we demonstrated the effectiveness of our method by comparing it with those state-of-the-art methods.

**INDEX TERMS** Multi-frame image restoration, multi-branch network, self-attention, branch-attention, full resolution convolution, images of space targets.

## I. INTRODUCTION

Image restoration aims to restore a clean image from a given degraded image. From the perspective of restoration methods, it can be divided into two main classes: the non-blind restoration [1]–[4] and the blind restoration [5]–[8]. It is common to find a de-convolution period which is always used to restore images after determining the point spread function (PSF) and some regularization rules in a non-blind method. Whereas blind methods restore images in processes where the degradation information is unknown. Due to the necessity of priors, the performance of non-blind methods is limited on real degraded images or images with different degradation patterns. With the development of deep convolutional neural networks (CNNs), researchers proposed a lot of effective blind restoration networks, most of which are in end-to-end manner and do not need any prior. But these methods often require a large number of parameters and training images. In the light of degradation manner, it can be divided into motion blur [9]–[12], shaking blur [13], [14], defocus blur [15], [16], physical blur [17], damaged image parts [18], atmospheric turbulence blur [19]–[23] and various kinds of noise [24]–[26]. Among them, atmospheric turbulence is the most complicated factor. The atmospheric turbulence obstructs the space targets from an astronomical telescope, which makes the observed images severely degraded [19], [20], [27]. Moreover, the anisotropic motion of atmospheric turbulence and random noise tend to introduce different degradation patterns to space-target images captured at different moments. On the basis of input data, it can be divided into single-frame input [28]–[32] and multi-frame input [33]–[38]. The methods with single-frame input are often suitable for conditions where degradation obeys a uniform distribution. However, the multi-frame methods are often used in conditions where the input frames own complementary information. The complementary information can be positions and postures of

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

different objects in continuous frames [39] or the relative information of the same object at different moments [36]. Considering the degradation manner of space-target images, it is reasonable to choose a multi-frame input for restoration. Based on this, we proposed a multi-frame network with a multi-frame input. And in this network, multiple branches restore each input frame synchronously. Additionally, the order of the input frames can be arbitrary.

Many popular restoration methods based on CNNs tend to give preference to a U-net structure [40]. A normal part in this structure is the encoder-decoder module [41], [42]. Firstly, the encoder performs downsamplings on input features through pooling operation, and then the decoder does the corresponding up-sample operations to restore features. During a long period, researchers used pooling layers to down-sample image features for large receptive field. But the analysis in [43] showed that unlike clean images, the sharp textures in degraded images change along with the scales of features. Furthermore, Visin *et al.* [44] thought that the pooling layers would drop some detailed information which is important for image restoration. To solve these problems above, Yu and Koltun [45] proposed the dilated convolution. Compared to pooling operation, dilated convolution can not only obtain large receptive fields with different dilation rates, but keep the feature resolution unchanged at the same time [46]. Based on the advantages of dilated convolution, we designed a full resolution convolution (FRC) module to extract image features for each branch in the network.

Recently, self-attention mechanism [47]–[50] shows extraordinary talents in natural language processing (NLP). Compared with traditional recurrent neural networks (RNNs) and CNNs, self-attention mechanism can take part into every input item of the network. At the beginning, the attention mechanisms that CNNs based on are channel attention [51] and spatial attention [52]. After Wang *et al.* [53] proposing to use non-local relation of pixels between image features, the self-attention has been used widely in image restoration [41], [54] till now. In this paper, we proposed a branch-attention mechanism which based on self-attention to share effective information from branch to branch in the network. The main contributions of our method are:

(1) A multi-branch network with a multi-frame input was proposed. The network's input contains multiple degraded image frames of space targets and each frame forms a restoration branch. And all the branches are trained synchronously.

(2) A full resolution convolution module was introduced. This module can get large receptive fields without changing feature scales by using dilated convolutions with different dilation rates.

(3) A modified self-attention module and a branch-attention module were proposed. The branch-attention module is used to share effective information between branches.

(4) Intensive experiments on the degraded images of space targets were conducted and a lot of comparisons with those state-of-the-art methods were analyzed.

The rest of this paper is organized as follows. Section II introduces the researches related to this paper. In section III, we detail the proposed network and modules. In section IV, we show the generation process of multi-frame data. Section V is about experimental details and analyses, including the details about data generation, the architecture details, the ablation studies, the evaluations of the restoration results and the demonstration of our restored images.

## II. RELATED WORKS

### A. RESTORATION APPROACHES

The degradation process of images can be expressed as follows:

$$D = C \otimes k + N \tag{1}$$

where $D$ means the degraded image, $C$ means the clean image, $k$ represents the blur kernel and $N$ represents noise. $\otimes$ and $+$ represent convolution operation and summation operation, respectively. And according to (1), many traditional approaches used priors to predict a blur kernel first and then did the inverse operation like deconvolution to get the clean images [55]–[58]. Thanks to the development of deep learning, many effective methods based on CNNs were proposed. Gao *et al.* [19] proposed a neural network consisting of stacked auto-encoders to restore turbulent blurred images. In order to estimate and remove non-uniform motion blur in images, Sun *et al.* [59] used a CNN to predict the distribution of motion blur first and next removed it by a deconvolution operation. Xu *et al.* [12] predicted the blur kernel by restoring sharp edges of a degraded image and then used a deconvolution operation to obtain a clean image. Inspired by spatial pyramid matching, Zhang *et al.* [42] used a fine-to-coarse manner to construct a multi-patch network for restoring blur images. On the basis of multi-frame input, Sim and Kim [38] used RDU and RUD modules to predict kernels for each pixel and generated a residual image at the same time. The final output was the weighted sum of the locally convolved output and the residual image. Aittala and Durand [37] proposed a multi-branch network with a multi-frame input, each branch of it is an encoder-decoder module. It computes the maximum value of each activation between all tracks, and concatenates these ''global features'' back into the per-frame local features to share information. In the field of video restoration, Pan *et al.* [60] proposed a method which first develops a deep CNN model to estimate optical flow from intermediate latent frames and then restores the latent frames based on the estimated optical flow. To fully recover the high frequency details, Liu *et al.* [61] proposed a two-stage deblurring module to recover the blur images of dynamic scenes based on high frequency residual image learning. In order to explicitly incorporate the blur-kernel in the network's training, Kaufman and Fattal [62] proposed a new architecture which breaks the deblurring network into an analysis network which estimates the blur, and a synthesis network that uses this kernel to deblur the image. To connect maximum posterior and deep models, Ren *et al.* [63]

presented two generative networks for respectively modeling the deep priors of clean image and blur kernel, and propose an unconstrained neural optimization solution to blind deconvolution.

### B. WHY CHOOSE DILATED CONVOLUTION ?

It is well known that expanding receptive fields of networks can improve the performance in vision tasks. The ways to achieve large receptive fields can be divided as follows: increasing kernel size [64], stacking convolutional layers, using pooling operations and exploiting dilated convolutions [65]. The first two methods enlarge receptive fields directly but increase the number of parameters at the same time. In the last two methods, the pooling operation enlarges receptive fields relatively by shrinking feature scales. And the dilated convolution can reach the same purpose without changing feature resolutions. Previous studies established by Boureau *et al.* [66] and Scherer *et al.* [67] show that pooling (especially max-pooling) can capture invariant features in some particular tasks and adapting pooling operations carefully can make networks be more competitive. However, in some recent researches, pooling does not seem to be so optimistic. Kauderer-Abrams [68] found that pooling is not very critical by quantifying translation-invariance in CNNs. Ruderman *et al.* [69] confirmed this point even more by showing that pooling does have anti-deforming capability in the early stage of training, but it makes no difference for the final performance no matter using pooling or not. These researches demonstrate shortcomings of max-pooling (the most widely used pooling method): one is the loss of spatial information which explains why we always use skip connections in U-like structures to convey the features before pooling layers; the other is that max-pooling cannot use information from multiple activation features and in the back forward stage, only activations from max pooling would be upgraded even if other activations appear errors. Fortunately, it can effectively avoid those problems by using dilated convolutions. The researches of DeepLab [70]–[72] used dilated convolution to get full resolution features on semantic image segmentation. Li *et al.* [73] used dilated convolutions to extract multi-scale features of input images to replace image pyramid and feature pyramid structures for object detection. Inspired by these works, we stacked dilated convolutional layers with different dilation rates to replace the common encoder-decoder module for feature extraction in an image-restoration task.

### C. ATTENTION MECHANISM

In previous works, Hu *et al.* [51] proposed the SEnet which uses a squeeze module and an exciation module to obtain the importance of features along with channel dimension. Later, Woo *et al.* [53] proposed the CBAM structure which expands SEnet by adding a spatial attention mechanism. On the basis of them, more and more complex attention mechanisms were proposed [74], [75], but the efficiency of these networks is hampered. For pursuing a balance between performance and efficiency, Wang *et al.* [76] proposed the ECANet. ECANet uses an ECA module which only produces $k$ parameters but achieves an obvious improvement in both performance and efficiency. ECA analyzes the relationship between dimension reduction and channels first, and then uses one-dimensional convolution with adaptive kernels to realize information exchange between channels. Considering the structure consistency and detail fineness, Wang *et al.* [18] proposed a multistage attention module which operates in a coarse-to-fine manner for large-scale irregular masks. For removing non-uniform blur, Qi *et al.* [77] proposed a channel attention module and a pixel attention module to capture long-range dependencies. Moreover, they introduced a scale attention module to remove unfavorable features while retaining features that facilitate image recovery. To address the multi-scaled mapping between high-resolution (HR) and low-resolution (LR) images of different scenes, Zhang *et al.* [78] proposed a multiscale attention network (MSAN) to extract the multilevel features of remote sensing images. To address the challenge caused by different levels of blurriness, Wu *et al.* [79] proposed a dual attention mechanism to dynamically aggregate temporal cues for deblurring with an end-to-end trainable network structure. In order to address the limited capability of utilizing the hierarchical features, Zhang *et al.* [80] proposed a memory-based latent attention network (MLANet). In this work, a memory-based latent attention block (MLAB) is stacked in MLANet and makes better use of global and local features through the network. With the success of transformer [47] in NLP, self-attention mechanism has gradually been widely applied to visual tasks. In the task of image generation, Parmar *et al.* [81] used self-attention mechanism to focus on neighborhood pixels and improved texture performance in the generated images. Liu *et al.* [82] merged self-attention mechanism with RNN, and improved the robustness of network by utilizing RNN's characteristics of information broadcast. Purohit and Rajagopalan [41] proposed a motion deblur network with fast execution speed and strong performance ability by combining deformable convolution and self-attention mechanism. In this paper, we used a modified self-attention mechanism as that used in [54] and based on it we designed a branch-attention mechanism to broadcast information between all the restoration branches.

## III. PROPOSED METHODS

### A. AN OVERVIEW OF THE NETWORK

For the purpose of restoring space-target images effectively, we designed a multi-branch neural network which has a multi-frame input (see Fig.1). The network's input consists of degraded frames with different degradation patterns and each frame corresponds to a restoration branch. Note that the input order of these frames can be arbitrary and all the restoration branches in this architecture are trained simultaneously. As shown in Fig.1, each input frame first goes through a full resolution convolution (FRC) module that extracts features with their resolutions unchanged. Next, the output of FRC in

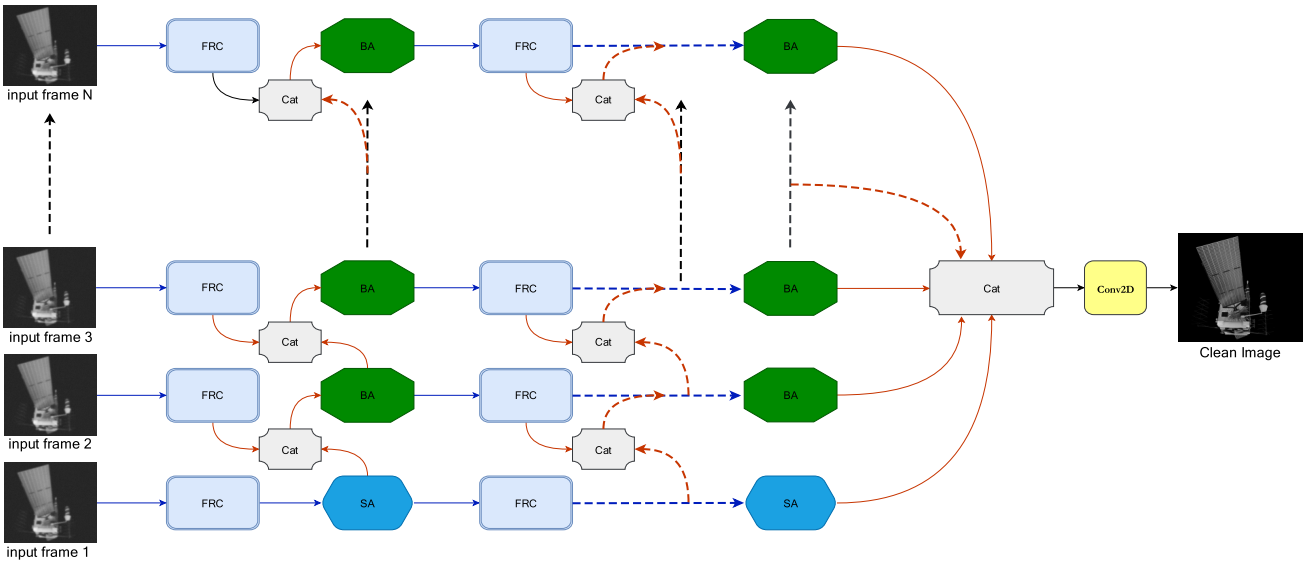


**FIGURE 1.** The architecture of the network. The network has a multi-frame input and each frame corresponds to a restoration branch.The FRC means full resolution convolution module. The SA and BA are self-attention mechanism and branch-attention mechanism respectively. The Cat represents concatenate operation. For simplification, processes the same as previous units are replaced by dotted lines.

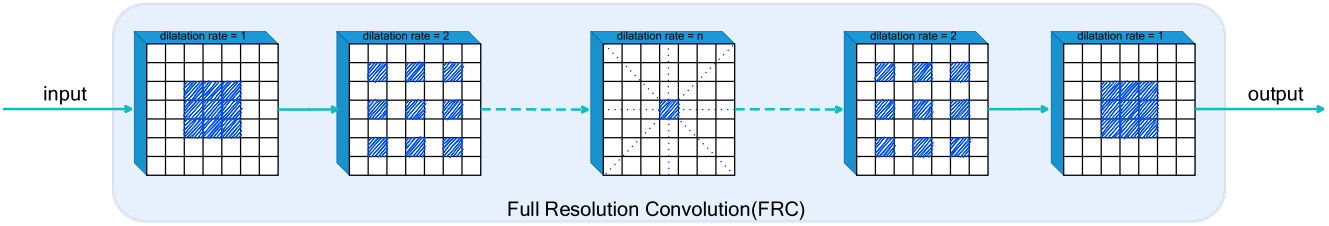


**FIGURE 2.** The full resolution convolution module.

each branch enters an attention mechanism. It is worth noting that the first branch uses a self-attention (SA) module but other branches use branch-attention (BA) modules instead. In the same period of all the branches, we can see that the attention features of the previous branch will concatenate with FRC's output in the current branch and then the concatenated features will be fed into BA (Fig.4). Each branch repeats the above process until the end of the network. The outputs of all branches concatenate together at last and the concatenated feature will be fed into a convolution layer to generate the final restoration image.

### B. FRC

Full resolution convolution (FRC) module is a very concise module. As shown in Fig.2, compared to encoder-decoder module, FRC increases receptive fields by stacking convolution layers, which is the same as that in encoder. The difference is that FRC uses dilated convolutions with increasing dilation rates to achieve large receptive fields whereas encoder tends to implement this by using pooling layers. Moreover, the dilated convolutions used in FRC can keep feature information more complete by maintaining feature resolutions unchanged. However, pooling layers may produce some drawbacks as proved in [69]. From Fig.2 we can see that the dilation rate of FRC first increases from 1 to $n$, and then decreases from $n$ to 1. It brings two benefits: one is that different kinds of dilated convolutions help to extract features from different regions which can enrich the diversity of features [73]; the other is that using different dilation rates can help all pixels participate in the convolution process which can avoid losing information when pixels come across the holes in dilated kernels. Note that, FRC has no skip connections and samplings which are common components in encoder-decoder structure. So it's easy to build a module like this and it can be used as a plug-in in CNNs.

### C. SELF-ATTENTION AND BRANCH-ATTENTION

In the tasks of computer vision, self-attention (SA) mechanism has introduced a non-local idea which can compute the relationship between two distant positions directly. Given an input tensor of shape $(H, W, C)$, it will be first flattened into two matrices in the shape of $(HW, \hat{C})$ and $(\hat{C}, HW)$ $(\hat{C} \leq C)$ respectively. And then a matrix multiplication $(HW, \hat{C}) \times (\hat{C}, HW)$ is conducted to get a $(HW, HW)$ matrix which will be activated by a softmax function in the next step.

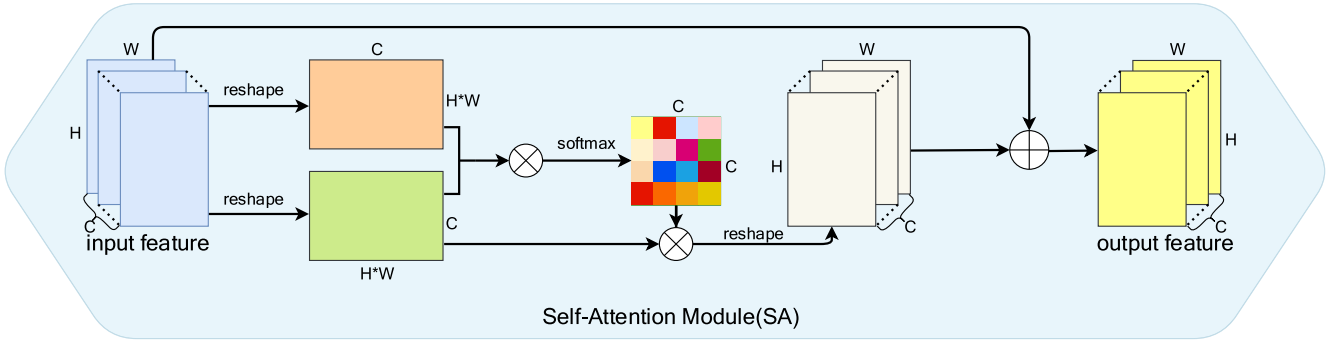


**FIGURE 3.** The modified self-attention module.

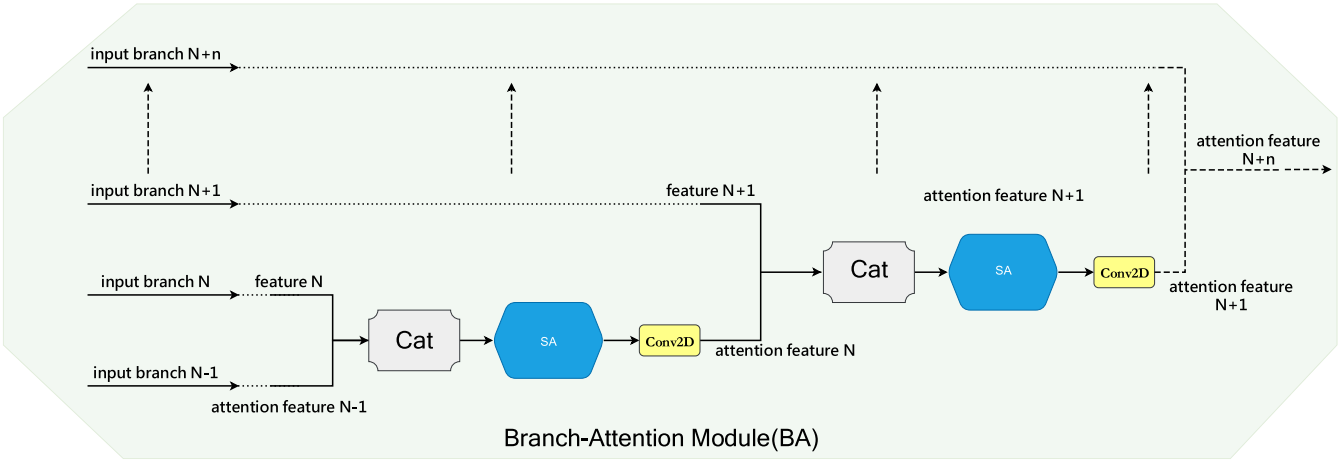


**FIGURE 4.** The branch-attention module.

The main drawbacks of this approach are high memory and computation of matrix $(HW, HW)$. In order to solve this problem, we do another matrix multiplication $(\hat{C}, HW)\times(HW, \hat{C})$ to obtain a matrix in the shape of $(\hat{C}, \hat{C})$ and in this paper $\hat{C} = C$ (see Fig.3). The final step is to get the output feature after adding the input feature to the reshaped attention feature generated by $(HW, \hat{C}) \times (\hat{C}, \hat{C})$.

In many deep learning methods, SA is mainly used to enhance important pixels in the current input features. For the multi-frame network in this paper, we could design an SA module for each branch independently. But with SA alone we cannot effectively use the complementary information between the input frames. Thus, in order to broadcast useful information between different branches, we designed a novel branch-attention (BA) module that based on SA in Fig.4. Different from delivering max pooling features to each branch in [37], we deliver features enhanced by BA from previous branch to the next branch. In Fig.4, the (N-1)-th branch's attention features will concatenate with the output features of FRC from the N-th branch. And then the concatenated features will go through an SA module and a convolutional layer to get the N-th attention features. The following convolutional layer is used to change the feature channels so as to match the (N+1)-th features from next branch. Every output of SA and BA in each branch will enter into the next FRC simultaneously to start a new round of feature extraction until the end of the network. The process of FRC and SA/BA can be expressed by formula (2) and algorithm1.

$$AF_{BA}^{N} = \begin{cases} SA(F_{FRC}^{N}), & N = 1 \\ Conv2D(SA(Cat(AF_{BA}^{N-1}, F_{FRC}^{N}))), & N > 1 \end{cases} \quad (2)$$

where $AF_{BA}^{N}$ represents the attention features generated by BA module of the $N$-th branch, $SA$ is the self-attention mechanism, $FRC$ is the full resolution convolution, $F_{FRC}^{N}$ means the output features of FRC module in the $N$-th branch, $Conv2d$ represents the 2D convolutional layer and $Cat$ means the concatenation operation.

## IV. DATA GENERATION

### A. DEGRADATION MODEL

Thanks to a lot of meaningful researches about atmospheric turbulence, including researches on optical astronomy imaging under the influence of atmospheric turbulence [83], experiments of turbulence degradation by simulating Kolomogorov phase screen [84], etc. Based on the above researches, in this

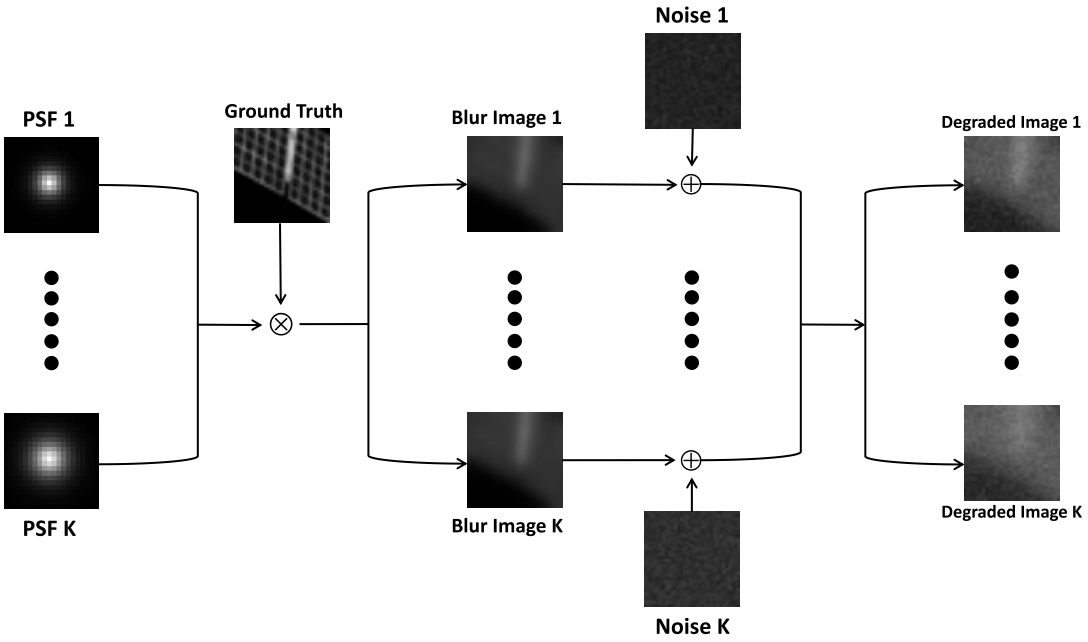


**FIGURE 5.** The process of generating degraded images.

paper the structure function of atmospheric turbulence can be descried by the PSF [19]:

$$PSF = \exp\{-3.44(\alpha f U/r)^{5/3}\} \tag{3}$$

where $U = \sqrt{u^2 + v^2}$ is the frequency, $(u, v)$ is the unit pulse. $\alpha$ is the wavelength, $f$ represents a focal length of the optical system and $r$ is the Fried parameter. Combined with (1), the degradation model on the images of space targets can be expressed as:

$$D = C \otimes PSF + N(\lambda) \tag{4}$$

where $N$ is the poisson noise in this paper and its distribution parameter is $\lambda$.

### B. SIMULATION FOR DATA SET

Unlike other kinds of synthesized degradation images which are easier to be generated from clean images due to the availability of high-frame-rate cameras. It is hard to obtain clean space-target images because of the irregularly motion of atmospheric turbulence and the limitation of observation equipment. In this paper, we obtain clean space-target images from the STK toolkit. STK is a simulation toolkit produced by the American Analytical Graphics company. It can not only calculate the orbits and attitudes of satellites, but the positions and actual illuminations of each space target depending on real information. It also provides a lot of space targets' texture information and aerospace models, so it can simulate the space scenes. In practice, we selected dozens of representative space targets and finally collected more than 1,200 clean images.

### C. HOW TO GENERATE TRAINING DATA

In order to comply with the motion property of atmospheric turbulence and the randomness of noise, we need to generate degradation patterns those conform to the real observed space-target images. According to (3)(4), we can find that the degradation process of space-target images mainly depends on the parameters $r$ and $\lambda$. The $r$ determines blur mode and the $\lambda$ determines noise intensity. Assuming that the number of input frames is $K$ and there're $n$ clean images need to be degraded in set $C$. Thus we need to generate $K$ pairs $(r, \lambda)$ for each clean image $C_i (i = 1, 2, \ldots, n)$. In order to determine the degradation range of $r$ and $\lambda$, we constantly compared the evaluation scores between the generated degraded images and the real observed images by using two non-reference image quality evaluation methods (Variance and Brenner) [85]. Eventually, the ranges of these two variables are set to $(r_x, r_y)$ and $(\lambda_x, \lambda_y)$. As shown in Fig.5, the clean images (groundtruth) are first convolved with $K$ PSFs which are randomly generated in range $(r_x, r_y)$ through (3). Next, the blurred images of each clean frame are added with the corresponding noises which are randomly generated in range $(\lambda_x, \lambda_y)$. So that the final output frames are degraded images with different degradation patterns.

## V. EXPERIMENTS AND RESULTS

### A. DETAILS ABOUT EXPERIMENTAL DATA

During the course of experiments, we set the ranges of $r$ and $\lambda$ in (3)(4) to $[0.01, 0.015]$ and $[35, 42]$ respectively. Different PSFs and noise are generated randomly from these two ranges and the clean images we selected from STK would go through the process showed in Fig.5 to produce the multi-frame

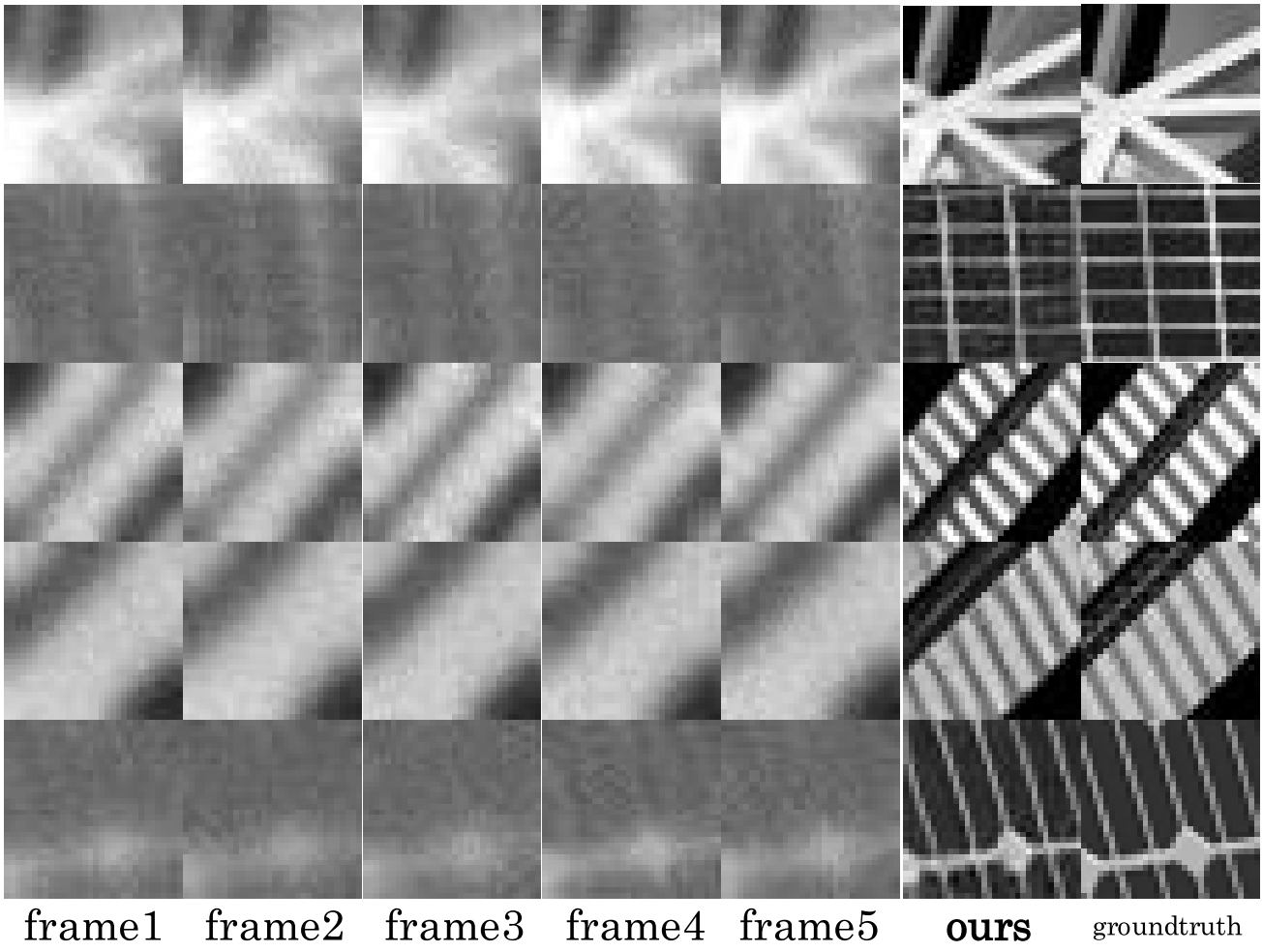


**FIGURE 6.** The degraded patches, restoration patches and clean patches.

**TABLE 1.** Average PSNR and SSIM values of various methods on the test set.

| methods | PSNR | SSIM |
|---|---|---|
| CBDnet | 25.9593 | 0.9116 |
| stack(4)-DMPHN | 26.2110 | 0.9146 |
| fastMBD | 16.1321 | 0.6179 |
| Sim | 27.0416 | 0.9325 |
| Aittala | 27.4222 | 0.9386 |
| **Ours** | **27.5172** | **0.9405** |

degraded inputs for the network. Specially, there were more than 1200 big and clean images of space targets. In experiments, the number of input frames was 5 which means we created 5 pairs of (*PSF*, *noise*) randomly and applied each of them to a clean image to generate 5 different degraded frames for the network in Fig.1. To increase the diversity of features, we did a data augmentation and cropped training images into patches which were in the size of (32, 32). Finally, we got more than 577, 000 training images and more than 58, 000 images for validation. Moreover, we reserved 2000 images for the test set.

### B. DETAILS OF ARCHITECTURE AND TRAINING PROCESS

As described above, the network in Fig.1 receives 5 frames of degraded images as its input, so the number of branches of this network is 5 as well. The maximum dilation rate we set for each FRC (Fig.2) is 4, so the order of dilation rates in each FRC block is (1, 2, 3, 4, 3, 2, 1). Note that when the dilation rate equals 1, the corresponding size of the convolution kernel is (3, 3). The output of FRC will enter into an SA or BA module. The SA (Fig.3) is a modified self-attention mechanism that uses less memory than the traditional version and the BA is a branch-attention mechanism based on SA to deliver important information between all branches. In practice, the maximum execution number of BA (algorithm 1) is 5 and the channels of each convolution layer in FRC will be doubled after each SA or BA. Thus channels of FRCs are (32, 64, 128, 256) which means each branch in the network will go through 4 FRCs and 4 SAs/BAs as performed in algorithm 1. The channel number of the last convolution layer in the network is 1. In training process, we used mean square error (MSE) as the loss function and Adam as the optimizer.

**TABLE 2.** The ablation study on proposed modules.

| Variations | Model1 | Model2 | Model3 | Model4 | Model5 |
|---|---|---|---|---|---|
| FRC | | | | √ | √ |
| SA | | √ | √ | √ | √ |
| BA | | | √ | | √ |
| PSNR | 23.8403 | 24.2294 | 25.6013 | 25.6591 | **27.5172** |
| SSIM | 0.8441 | 0.8532 | 0.9223 | 0.9037 | **0.9405** |
| parameters | 11.6725M | 11.6725M | 14.4018M | 11.6725M | 14.4018M |

**Algorithm 1** The Process of FRC and SA/BA

**Require:**
FRC: the function of full resolution convolution;
SA/BA: the function of self-attention mechanism or branch-attention mechanism;
DF: one output feature of FRC;
DFs: a list to store output features of FRC;
AF: one output features of SA/BA;
AFs: a list to store output features of SA/BA;
InputFrames: a list storing the input frames;
length: the number of input frames;
turns: the execution number of SA/BA;

**Ensure:**
DFs = []; * initialize the DFs*\
**for** *i* in range(length) **do**
* get the initial features of input frames*\
DF = FRC(InputFrames[i]);
* do FRC for each input frame in InputFrames*\
DFs.append(DF);
* store the output of FRC for each input frame *\
**end for**
**for** *j* in range(turns) **do**
AFs = []; * initialize the AFs*\
**for** *k* in range(len(DFs)) **do**
* do SA/BA for features from FRC *\
**if** j == 0 **then**
AFs.append(SA(DFs[j]))
**else**
AFs.append(BA(DFs[j], AFs[j-1]))
**end if**
**end for**
**if** *k* < len(turns-1) **then**
**for** *t* in range(len(AFs)) **do**
DFs[t] = FRC(AFs[t])
**end for**
**end if**
**end for**

The learning rate was set to $1e-4$ and it reduced half after every 50000 iterations. The batch size of training set was 16 and in validation process it was 64. The training epochs of the network was 100. To get the real generalization ability of the proposed network and ensure fairness with other comparative approaches, we only saved the best results on validation data. At last, we applied these results of all the approaches to the test set to analyze the final restoration effectiveness of them.

### C. ABLATION STUDY

In order to reflect the effectiveness of the proposed modules, we conducted an ablation study on them and the results are demonstrated in Table 2. In Table 2, there're 5 models with different proposed modules and for the models (model1, model2 and model3) without FRC modules, we have designed an encoder-decoder module for each of them. Note that, the encoder-decoder module has the same number of parameters as FRC module because they have the same number of convolutional layers whose kernels are in the shape of (3, 3). From model1 and model2 we can find that the SA module can improve the performance of the network without producing extra parameters. From model2 and model3 (or model4 and model5) we can see that the BA module improves the network's performance to a great extent, which proves the existence of complementary information between the multiple input frames. And from model3 and model 5 (or model2 and model4) we find that compared to the encoder-decoder module, the FRC module can achieve better performance with the same number of parameters.

### D. COMPARATIVE EXPERIMENT AND ANALYSIS

In order to better analyze and evaluate the performance of the proposed algorithm, we selected five representative baseline methods for experimental comparisons. Among them, there are three multi-frame methods and two sing-frame methods. Aittala and Durand [37] is a deep-learning based method that put multiple frames into a neural network, and then performing image restoration. Sim and Kim [38] is a deep-learning based method as well, it takes multiple frames as input and makes each pixel learn an adaptive kernel. fastMBD [33] is a traditional multi-frame deblurring method. CBDnet [26] is a single-frame method that can restore the real degraded image well. Stack(4)-DMPHN [43] is a single-frame method using a deep stacked hierarchical multi-patch network for image deblurring. For a fair comparison, all multi-frame methods used the same training set, validation set and test set as those used by our network. For all single-frame methods, the least degraded images (with $r = 0.01$ and $\lambda = 35$) are used as training data, validation data and test data. The baseline methods that need training are all re-trained by using the source code and parameters provided by the corresponding authors.

Fig.6 shows the degraded patches, our restoration patches and clean patches of space-target images. Columns 1 to 5

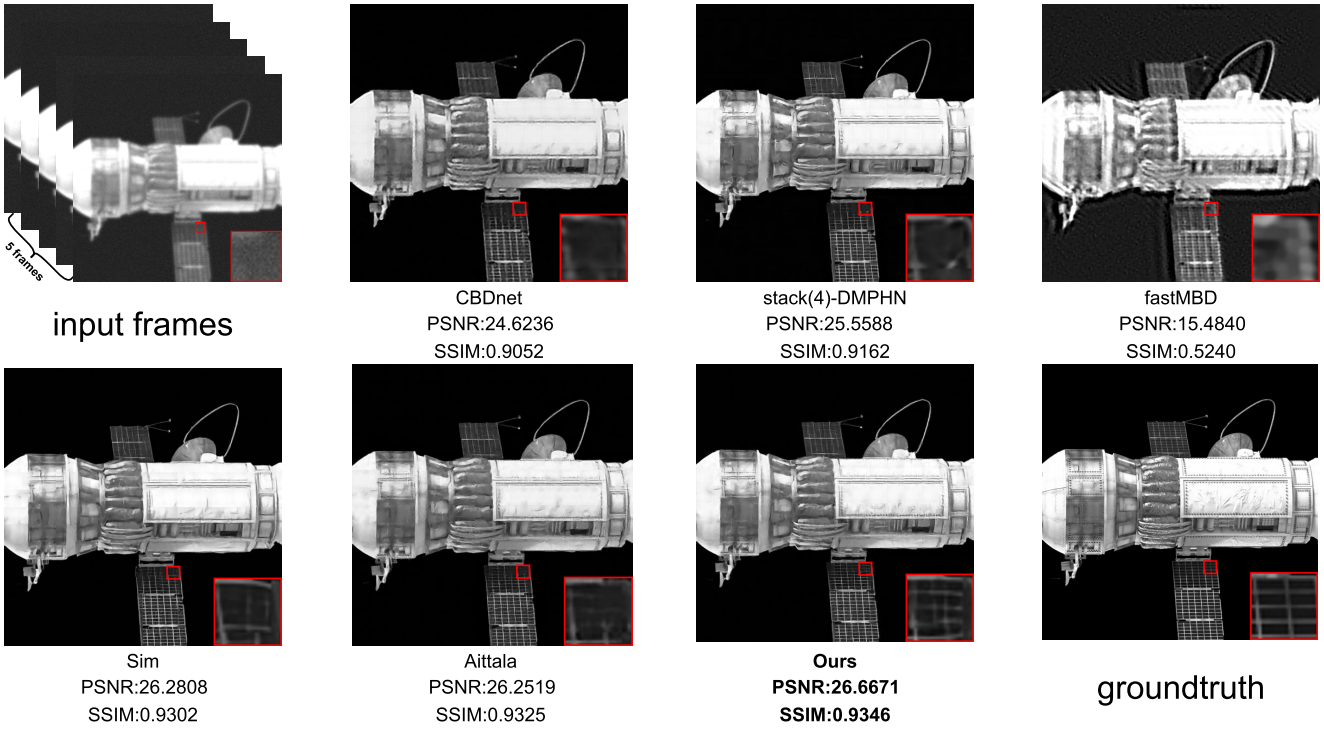


**FIGURE 7.** The restoration images of a whole space target.

**TABLE 3.** Models with different input frames.

| Metrics | Model(2frames) | Model(3frames) | Model(4frames) | Model(5frames) | Model(6frames) |
|---|---|---|---|---|---|
| PSNR | 25.1159 | 25.9014 | 26.9393 | 27.5172 | 27.5665 |
| SSIM | 0.8911 | 0.9106 | 0.9340 | 0.9405 | 0.9431 |
| parameters | 5.3515M | 8.3682M | 11.3851M | 14.4018M | 17.4186M |

show the degraded patches in the training process, the 6-th column shows the restoration results of our network and the last column shows the clean images used as labels. And it also shows that our method can restore the textures effectively from degraded images. Fig.7 shows the restored images from a perspective that can observe the whole space target. Fig.8 demonstrates the restoration results generated by different approaches on a local part of a satellite from the close view. In Fig.9 we chose the representative part that has a lot of textures of a space-target. It can be seen that the image restored by our method has more clean details.

Table 1 shows the average PSNR and SSIM values of our method and the other five state-of-the-art methods on the test set. From the table and the restored images showed in this paper, it can be seen that in the condition of severely degradation of blur and noise, the traditional method fastMBD can effectively restore the outline and some textures of the degraded images, but it cannot remove noise completely and it brought many overlapped shadows as well. The result also shows that the single-frame methods are inferior to multi-frame methods even if they used the least degraded training data. This phenomenon also suggests that the complementary information indeed exists in multiple frames with different degraded patterns. In these two single-frame methods, the stack(4)-DMPHN cut the image into different parts to restore them separately and then spliced them together to get the clean image. Compared with the CBDnet who used the whole image and residual operations, stack(4)-DMPHN got a better performance. For the three multi-frame methods, the Sim method used the adjacent frames to restore the target frame, but in the condition where each frame was degraded randomly it is inferior to the Aittala method and our method that can accept arbitrary order of the input frames and prepare a restoration branch for each frame. The Aittala method used encoder-decoder for each branch which may course loss of information. And the strategy Aittala used was using the maximum features to deliver information between branches, whereas we use an SA/BA module which is more targeted on feature selection to extract important features for each branch. So compared to Aittala method, our method has an advantage. In general, our method can reach a better performance with higher resolution and more details compared to these similar methods.

We have conducted experiments on our network by using different number of input frames as well and the restoration results of these different models are provided in Table 3.

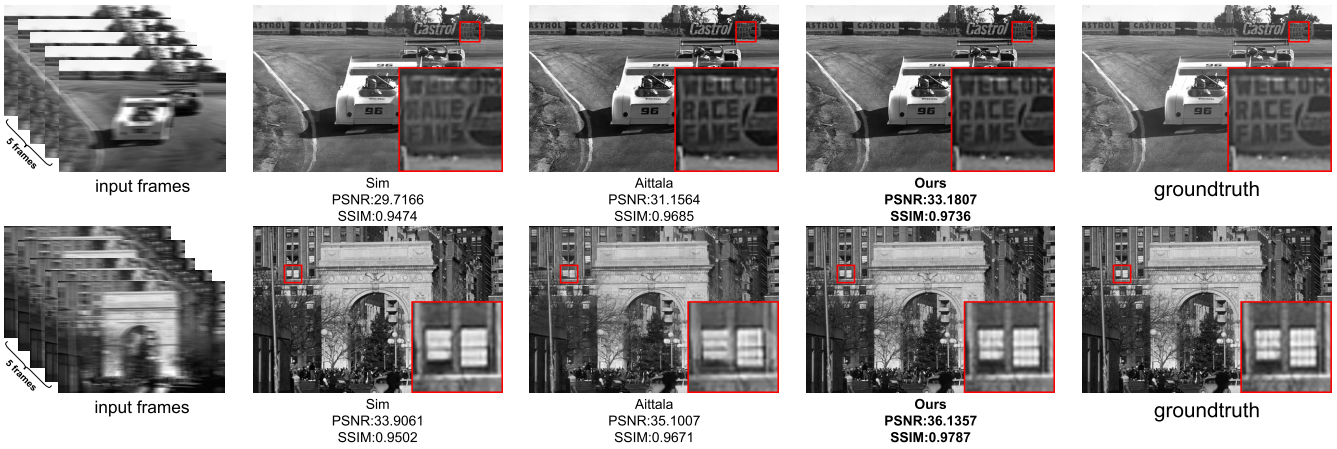


**FIGURE 8.** The deblurred images of different multi-frame methods.


**FIGURE 9.** The restoration part of a satellite from a close view.

From Table 3 we can find that the PSNR and SSIM values increased obviously in the first three models, whereas the performance improved slightly of the last two models. And the parameters changed with the input frames from less to more because one more input frame means one more restoration branch in the network. We can conclude that the improvement of the network's performance will be less and less with the increasing number of input frames and parameters. It's not hard to understand this phenomenon: in Section IV.C we determined the ranges of $r$ and $\lambda$, and in these ranges we generated degraded images randomly. With the increasement of input frames (over 6 frames), the differences in degradation patterns between them are being smaller, which means the effective complementary information brought by these extra frames are less. Thus, for the balance of effectiveness and performance, we chose 5 as the final number of input frames for our network.

Moreover, for verifying the robustness and generalization of the proposed method, we have conducted the motion deblurring experiments. Instead of utilizing the existing motion-blur datasets in the form of continuous video frames, we chose to generate motion-blur frames with random


**FIGURE 10.** The textures restored by various methods.

**TABLE 4.** Average PSNR and SSIM values of multi-frame methods for motion deblurring.

| methods | PSNR | SSIM |
|---|---|---|
| Sim | 29.5164 | 0.9535 |
| Aittala | 29.7312 | 0.9664 |
| **Ours** | **31.2420** | **0.9772** |

differences which can reflect the characteristics of our method better. Like the process in Section IV.C, we first determined two degradation ranges to control the motion-blur patterns and then generated motion-blur images in these two ranges randomly. The key factors to generate motion-blur images are angle and degree which represent the motion direction and motion speed respectively. In practice, these two ranges were set to [40°,50°] and [15p, 30p] (n-p means n pixels). The datasets we used to generate training data are Urban100, Set5, Set14, T91 and we used dataset BSD100 to generate test data. We compared our method with Sim method [38] and Aittala method [37], because they are preeminent multi-frame methods on motion deblurring in the two years. The results of motion deblurring methods are shown in Table 4 and Fig. 8. From the results we can find that our method has an advantage.

## VI. CONCLUSION

In this paper, we have proposed a multi-branch network with multi-frame input. In this network, a full resolution convolution (FRC) module is used to extract image features. Compared to pooling layers and other downsampling methods, FRC keeps feature resolutions and scales unchanged. Thus, it can keep feature information more complete. Moreover, a modified self-attention (SA) module which has less memory usage and computation is proposed to enhance features. Based on SA, a branch-attention (BA) mechanism is proposed to broadcast effective information between all the branches in the network. In addition, intensive experiments were conducted to evaluate our method. At last, the experimental results and analysis approve that our method performs better than other similar methods.

## REFERENCES

[1] S. Tang, W. Gong, W. Li, and W. Wang, "Non-blind image deblurring method by local and nonlocal total variation models," *Signal Process.*, vol. 94, pp. 339–349, Jan. 2014.

[2] S. Vasu, V. R. Maligireddy, and A. N. Rajagopalan, "Non-blind deblurring: Handling kernel uncertainty with CNNs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3272–3281.

[3] J. Dong, J. Pan, D. Sun, Z. Su, and M.-H. Yang, "Learning data terms for non-blind deblurring," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 748–763.

[4] J. Zhang, J. Pan, W.-S. Lai, R. W. H. Lau, and M.-H. Yang, "Learning fully convolutional networks for iterative non-blind deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3817–3825.

[5] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical J.*, vol. 79, p. 745, Jun. 1974.

[6] M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 1, pp. 58–63, Feb. 1976.

[7] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Scholkopf, "Learning to deblur," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, Jul. 2016.

[8] Y. Zhang, Y. Lau, H.-W. Kuo, S. Cheung, A. Pasupathy, and J. Wright, "On the global geometry of sphere-constrained sparse blind deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4894–4902.
[9] A. Levin, "Blind motion deblurring using image statistics," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 841–848.
[10] M. Jin, G. Meishvili, and P. Favaro, "Learning to extract a video sequence from a single motion-blurred image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6334–6342.
[11] T. M. Nimisha, A. K. Singh, and A. N. Rajagopalan, "Blur-invariant deep learning for blind-deblurring," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4752–4760.
[12] 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.
[13] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, "Non-uniform deblurring for shaken images," *Int. J. Comput. Vis.*, vol. 98, no. 2, pp. 168–186, Jun. 2012.
[14] V. Khare, P. Shivakumara, P. Raveendran, and M. Blumenstein, "A blind deconvolution model for scene text detection and recognition in video," *Pattern Recognit.*, vol. 54, pp. 128–148, Jun. 2016.
[15] C. Zhou, S. Lin, and S. K. Nayar, "Coded aperture pairs for depth from defocus and defocus deblurring," *Int. J. Comput. Vis.*, vol. 93, no. 1, pp. 53–72, May 2011.
[16] J. Lee, S. Lee, S. Cho, and S. Lee, "Deep defocus map estimation using domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12222–12230.
[17] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Blind correction of optical aberrations," in *Proc. Eur. Conf. Comput. Vis.* Heidelberg, Germany: Springer-Verlag, 2012, pp. 187–200.
[18] N. Wang, S. Ma, J. Li, Y. Zhang, and L. Zhang, "Multistage attention network for image inpainting," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107448.
[19] Z. Gao, C. Shen, and C. Xie, "Stacked convolutional auto-encoders for single space target image blind deconvolution," *Neurocomputing*, vol. 313, pp. 295–305, Nov. 2018.
[20] R. Mourya, L. Denis, J.-M. Becker, and E. Thiebaut, "A blind deblurring and image decomposition approach for astronomical image restoration," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 1636–1640.
[21] L. Yan, M. Jin, H. Fang, H. Liu, and T. Zhang, "Atmospheric-Turbulence-Degraded astronomical image restoration by minimizing second-order central moment," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 672–676, Jul. 2012.
[22] X. Zhu and P. Milanfar, "Removing atmospheric turbulence via space-invariant deconvolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 157–170, Jan. 2013.
[23] Y. Xie, W. Zhang, D. Tao, W. Hu, Y. Qu, and H. Wang, "Removing turbulence effect via hybrid total variation and deformation-guided kernel regression," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4943–4958, Oct. 2016.
[24] M. Lebrun, "An analysis and implementation of the BM3D image denoising method," *Image Process. Line*, vol. 2, pp. 175–213, Aug. 2012.
[25] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising prior driven deep neural network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, Oct. 2019.
[26] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1712–1722.
[27] J. Shi, R. Zhang, S. Guo, Y. Yang, R. Xu, W. Niu, and J. Li, "Space targets adaptive optics images blind restoration by convolutional neural network," *Opt. Eng.*, vol. 58, no. 9, 2019, Art. no. 093102.
[28] G. Chen, Z. Gao, Q. Wang, and Q. Luo, "U-net like deep autoencoders for deblurring atmospheric turbulence," *J. Electron. Imag.*, vol. 28, no. 5, 2019, Art. no. 053024.
[29] D. Li, R. M. Mersereau, and S. Simske, "Atmospheric turbulence-degraded image restoration using principal components analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 340–344, Jul. 2007.
[30] M. H. Furhad, M. Tahtali, and A. Lambert, "Restoring atmospheric-turbulence-degraded images," *Appl. Opt.*, vol. 55, no. 19, pp. 5082–5090, 2016.
[31] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, and M.-H. Yang, "A comparative study for single image blind deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1701–1709.
[32] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Learning a discriminative prior for blind image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6616–6625.
[33] F. Sroubek and P. Milanfar, "Robust multichannel blind deconvolution via fast alternating minimization," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1687–1700, Apr. 2012.
[34] T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 10, no. 5, pp. 1064–1073, 1993.
[35] A. Yang, X. Jiang, and D. Day-Uei Li, "Multi-frame blind deconvolution of atmospheric turbulence degraded images with mixed noise models," *Electron. Lett.*, vol. 54, no. 4, pp. 206–208, Feb. 2018.
[36] F. A. G. Pena, P. D. Marrero Fernandez, T. Ing Ren, J. de Jesus Gomes Leandro, and R. M. Nishihara, "Burst ranking for blind multi-image deblurring," *IEEE Trans. Image Process.*, vol. 29, pp. 947–958, 2020.
[37] M. Aittala and F. Durand, "Burst image deblurring using permutation invariant convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 731–747.
[38] H. Sim and M. Kim, "A deep motion deblurring network based on per-pixel adaptive kernels with residual down-up and up-down modules," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–10.
[39] S. Zhou, J. Zhang, J. Pan, W. Zuo, H. Xie, and J. Ren, "Spatio-temporal filter adaptive network for video deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2482–2491.
[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer-Verlag, 2015, pp. 234–241.
[41] K. Purohit and A. Rajagopalan, "Region-adaptive dense network for efficient motion deblurring," in *Proc. AAAI*, 2020, pp. 11882–11889.
[42] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5978–5986.
[43] H. Gao, X. Tao, X. Shen, and J. Jia, "Dynamic scene deblurring with parameter selective sharing and nested skip connections," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3848–3856.
[44] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. Courville, and Y. Bengio, "ReNet: A recurrent neural network based alternative to convolutional networks," 2015, *arXiv:1505.00393*. [Online]. Available: http://arxiv.org/abs/1505.00393
[45] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*. [Online]. Available: http://arxiv.org/abs/1511.07122
[46] S. A. Kamran and A. S. Sabbir, "Efficient yet deep convolutional neural networks for semantic segmentation," in *Proc. Int. Symp. Adv. Intell. Informat. (SAIN)*, Aug. 2018, pp. 123–130.
[47] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[48] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI Blog*, vol. 1, no. 8, p. 9, 2019.
[49] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*. [Online]. Available: http://arxiv.org/abs/1810.04805
[50] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. V. Le, and R. Salakhutdinov, "Transformer-XL: Attentive language models beyond a fixed-length context," 2019, *arXiv:1901.02860*. [Online]. Available: http://arxiv.org/abs/1901.02860
[51] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[52] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
[53] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.
[54] M. Suin, K. Purohit, and A. N. Rajagopalan, "Spatially-attentive patch-hierarchical network for adaptive motion deblurring," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3606–3615.
[55] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. CVPR*, Jun. 2011, pp. 233–240.

[56] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1628–1636.
[57] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4003–4011.
[58] L. Chen, F. Fang, T. Wang, and G. Zhang, "Blind image deblurring with local maximum gradient prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1742–1750.
[59] 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.
[60] J. Pan, H. Bai, and J. Tang, "Cascaded deep video deblurring using temporal sharpness prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3043–3051.
[61] K.-H. Liu, C.-H. Yeh, J.-W. Chung, and C.-Y. Chang, "A motion deblur method based on multi-scale high frequency residual image learning," *IEEE Access*, vol. 8, pp. 66025–66036, 2020.
[62] A. Kaufman and R. Fattal, "Deblurring using analysis-synthesis networks pair," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5811–5820.
[63] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo, "Neural blind deconvolution using deep priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3341–3350.
[64] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 1790–1798.
[65] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 4898–4906.
[66] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 111–118.
[67] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proc. Int. Conf. Artif. Neural Netw.* Heidelberg, Germany: Springer-Verlag, 2010, pp. 92–101.
[68] E. Kauderer-Abrams, "Quantifying translation-invariance in convolutional neural networks," 2017, *arXiv:1801.01450*. [Online]. Available: http://arxiv.org/abs/1801.01450
[69] A. Ruderman, N. C. Rabinowitz, A. S. Morcos, and D. Zoran, "Pooling is neither necessary nor sufficient for appropriate deformation stability in CNNs," 2018, *arXiv:1804.04438*. [Online]. Available: http://arxiv.org/abs/1804.04438
[70] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.
[71] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587
[72] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[73] Y. Li, Y. Chen, N. Wang, and Z.-X. Zhang, "Scale-aware trident networks for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6054–6063.
[74] J. Hu, L. Shen, S. Albanie, G. Sun, and A. Vedaldi, "Gather-excite: Exploiting feature context in convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 9401–9411.
[75] I. Bello, B. Zoph, Q. Le, A. Vaswani, and J. Shlens, "Attention augmented convolutional networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3286–3295.
[76] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11534–11542.
[77] Q. Qi, J. Guo, and W. Jin, "Attention network for non-uniform deblurring," *IEEE Access*, vol. 8, pp. 100044–100057, 2020.
[78] S. Zhang, Q. Yuan, J. Li, J. Sun, and X. Zhang, "Scene-adaptive remote sensing image super-resolution using a multiscale attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4764–4779, Jul. 2020.
[79] J. Wu, X. Yu, D. Liu, M. Chandraker, and Z. Wang, "DAVID: Dual-attentional video deblurring," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2376–2385.
[80] X. Zhang, P. Gao, K. Zhao, S. Liu, G. Li, and L. Yin, "Image restoration via deep memory-based latent attention network," *IEEE Access*, vol. 8, pp. 104728–104739, 2020.
[81] N. Parmar, A. Vaswani, J. Uszkoreit, Ł. Kaiser, N. Shazeer, A. Ku, and D. Tran, "Image transformer," 2018, *arXiv:1802.05751*. [Online]. Available: http://arxiv.org/abs/1802.05751
[82] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration" in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1673–1682.
[83] M. Carbillet, "Astronomical imaging... Atmospheric turbulence? Adaptive optics!," *EAS Publications*, vol. 59, pp. 59–76, Mar. 2013.
[84] M. C. Roggemann, B. M. Welsh, and B. R. Hunt, *Imaging Through Turbulence*. Boca Raton, FL, USA: CRC Press, 1996.
[85] N. M. Burns and J. Watson, "A study of focus metrics and their application to automated focusing of inline transmission holograms," *Imag. Sci. J.*, vol. 59, no. 2, pp. 90–99, Apr. 2011.

**PEIJIAN ZHU** received the B.S. degree from the School of Computer and Software Engineering, Changzhou Institute of Technology, in 2018. He is currently pursuing the master's degree with the School of Computer and Software Engineering, Xihua University. His research interests include deep learning for image processing and computer vision.

**CHUNZHI XIE** received the Ph.D. degree in computer science from Sichuan University, in 2017. She is currently an Associate Professor with Xihua University. She is the author of more than 20 journal articles. Her current research interests include machine learning, image processing, and uncertain reasoning.

**ZHISHENG GAO** received the Ph.D. degree in computer science from Sichuan University, in 2012. He is currently an Associate Professor with Xihua University. He is the author of more than 30 journal articles. His current research interests include machine learning and visual analysis.