Proceedings of Machine Learning Research 101:852–867, 2019 ACML 2019

Multi-width Activation and Multiple Receptive Field Networks for Dynamic Scene Deblurring

Jinkai Cui JINKAICUI@CQU.EDU.CN
Weihong Li WEIHONGLI@CQU.EDU.CN
Wei Guo GWFEMMA@CQU.EDU.CN
Weiguo Gong WGGONG@CQU.EDU.CN
Key Lab of Optoelectronic Technology and Systems of Education Ministry, College of Optoelectronic Engineering, Chongqing University, Chongqing, China

Editors: Wee Sun Lee and Taiji Suzuki

Abstract

In this paper, we propose an end-to-end multi-width activation and multiple receptive field networks for the large-scale and complicated dynamic scene deblurring. Firstly, we design a multi-width activation feature extraction module, in which a multi-width activation residual block is proposed for making the activation function learn more the nonlinear information and extracting wider nonlinear features. Secondly, we design a multiple receptive field (RF) feature extraction module, in which a multiple RF residual block is proposed for enlarging the RF efficiently and capturing more nonlinear information from distant locations. And then, we design the multi-scale feature fusion module, where a learning fusion structure is designed to adaptively fuse the multi-scale features and complicated blur information from the different modules. Finally, we use a multi-component loss function to jointly optimize our networks. Extensive experimental results demonstrate that the proposed method outperforms the recent state-of-the-art deblurring methods, both quantitatively and qualitatively.

Keywords: multi-width activation, multiple receptive field, multi-scale fusion, nonlinear information, dynamic scene deblurring

1. Introduction

Dynamic scene blur caused by the camera shake, object motions, different scene depths and occlusion in motion boundaries during the exposure time, and it is one of the most common image degradation problems when taking a photo in the wild. The blurry structures of image not only degrade their visual quality seriously, but also directly affect the practical application of image in various fields. Researchers have actively studied this issue in the past decade. However, most existing methods are usually ineffective for large-scale and complicated dynamic scene blurs.

Previous dynamic scene deblurring methods Hyun Kim et al. (2013); Hyun Kim and Mu Lee (2014); Pan et al. (2016) usually rely heavily on non-uniform blur kernels estimation or an accurate image segmentation. And the process of deblurring is time-consuming, since these methods need to solve the highly non-convex optimization problem. Recently, deep learning, especially the deep convolutional neural network (CNN) Krizhevsky et al. (2012), has been proven superior in the field of image processing. The CNN-based methods Sun et al. (2015); Gong et al. (2017); Noroozi et al. (2017); Nah et al. (2017); Kupyn et al. (2018); Tao et al. (2018); Zhang et al. (2018a); Gao et al. (2019) have been proposed for dynamic scene deblurring. Some methods Sun et al. (2015); Gong et al. (2017) estimated pixel-wise blur kernels by utilizing CNN, which have been shown good performance on certain blurred images. These methods can obtain some global nonlinear information, but lead to a high computation cost. And because these methods employ the synthetic blurry images which are generated by convolving the clean natural images with the synthetic motion kernels, they either tackle only several specific type dynamic scene blurs or the simple blur model. Some methods Noroozi et al. (2017); Nah et al. (2017); Kupyn et al. (2018); Tao et al. (2018); Zhang et al. (2018a); Gao et al. (2019) based on end-to-end trainable manner have also been proposed for image deblurring. Nah et al. (2017) proposed a multi-scale CNN, which adopted the coarse-to-fine manner for improving the performance of dynamic scene deblurring. Tao et al. (2018) proposed a scale-recurrent network using a similar multi-scale strategy for dynamic scene deblurring. These methods all required the coarse-to-fine strategy, and adopted a small amount of channels before activation function. The number of the features limits the performance of activation functions. This causes that the activation function may loss some nonlinear information about complex dynamic scene blurs. In addition, the methods usually design a single convolution kernel, which leads to a single RF to learn all nonlinear information. Because the blurs vary from pixel to pixel and from image to image in dynamic scenes, it is difficult to deal with all cases by utilizing the single RF, especially in large-scale and complex blur regions.

In this paper, we propose an end-to-end multi-width activation and multiple receptive field networks for large-scale and complicated dynamic scene deblurring. Firstly, we design a multi-width activation feature extraction module, in which a multi-width activation residual block is proposed for making the activation function learn more the nonlinear information and extracting wider nonlinear features. Because a larger RF is necessary to deal with large-scale blur Jin et al. (2018), a multiple RF residual block is proposed for enlarging the RF efficiently and capturing more nonlinear information from distant locations. Besides, we design the multi-scale feature fusion module to better representation of the features and speeding up the convergence of the training process, where a learning fusion structure is designed to adaptively fuse the multi-scale features and complicated blur information from the different modules. Finally, we apply mean squared error (MSE) losses to help multi-scale features extraction in each module, and we use perceptual loss Johnson et al. (2016) and Wasserstein generative adversarial network (GAN) Arjovsky et al. (2017) with Gradient Penalty Gulrajani et al. (2017) to preserve high texture details and look perceptually more convincing. The main contributions of this paper are summarized as follows:

- We propose an end-to-end multi-width activation and multi-scale fusion networks for large-scale and complicated dynamic scene deblurring.
- We design a multi-width activation residual block for making the activation function learn more the nonlinear information and extracting wider nonlinear features.
- We design a multiple RF residual block for enlarging the RF efficiently and capturing more nonlinear information from distant locations.

- We design a learning fusion structure to adaptively fuse the multi-scale features and complicated blur information from the different modules.

2. Related Works

Recently, dynamic scene deblurring methods based on deep learning have achieved great progress. The CNN-based methods Sun et al. (2015); Gong et al. (2017) usually require estimating the non-uniform blur kernels, and then restore clear images by using a non-blind deblurring method Zoran and Weiss (2011). Sun et al. (2015) presented a learning method to address non-uniform motion blur through estimating the motion blur of every patch and adopted the Markov random field model to achieve a dense non-uniform motion blur field. This method may lose some the high-level information for some larger regions, since the training process is at the patch-level. Gong et al. (2017) presented a deeper CNN model for estimating motion flow and removing pixel-wise heterogeneous motion blur. However, it was only trained for linear blur examples, and limited to deal with several simple types of blur.

The based on end-to-end trainable manner Noroozi et al. (2017); Nah et al. (2017); Kupyn et al. (2018); Tao et al. (2018); Zhang et al. (2018a); Gao et al. (2019) has been extended to dynamic scene deblurring. Noroozi et al. (2017) proposed a multi-scale CNN to obtain a larger RF and adopted pyramid schemes with skip connections. Each segment of the network only needs to produce a residual image to help image reconstruction. Nah et al. (2017) proposed a multi-scale CNN with 40 convolution layers in each scale for dynamic scene deblurring. The method required 120 convolution layers and added adversarial loss to make sure restore sharp realistic images. Furthermore, Kupyn et al. (2018) proposed a conditional GAN and applied multi-components loss function to preserve high texture details as well as look perceptually more convincing, but their method can only restore blurry image with smaller blur and specific type dynamic scene blurs. Tao et al. (2018) adopted a similar multi-scale strategy and proposed a scale-recurrent network with convolutional long short term memory (LSTM) for dynamic scene deblurring. Moreover, they proposed sharing network weights in each scale to reduce training difficulty. Zhang et al. (2018a) proposed a spatially variant neural network, which consists of three deep CNN and a recurrent neural network. Deblurring step is similar to an infinite impulse response model, which can be approximated by recurrent neural network. Gao et al. (2019) proposed a network with parameter selective sharing and nested skip connections. And the method had achieved some promising results.

3. Proposed Method

3.1. Proposed Network Architecture

The architecture of the proposed multi-width activation and multiple receptive field networks, as shown in Fig. 1, which consists of three modules. The multi-width activation feature extraction module: it aims to make the activation function learn more the nonlinear information and extracting wider nonlinear features. The multiple RF feature extraction module: it is used to efficiently enlarge the RF and capturing more nonlinear information from distant locations. The multi-scale feature fusion module: it is designed to adaptively fuse multi-scale features and reconstruct the final clear image.

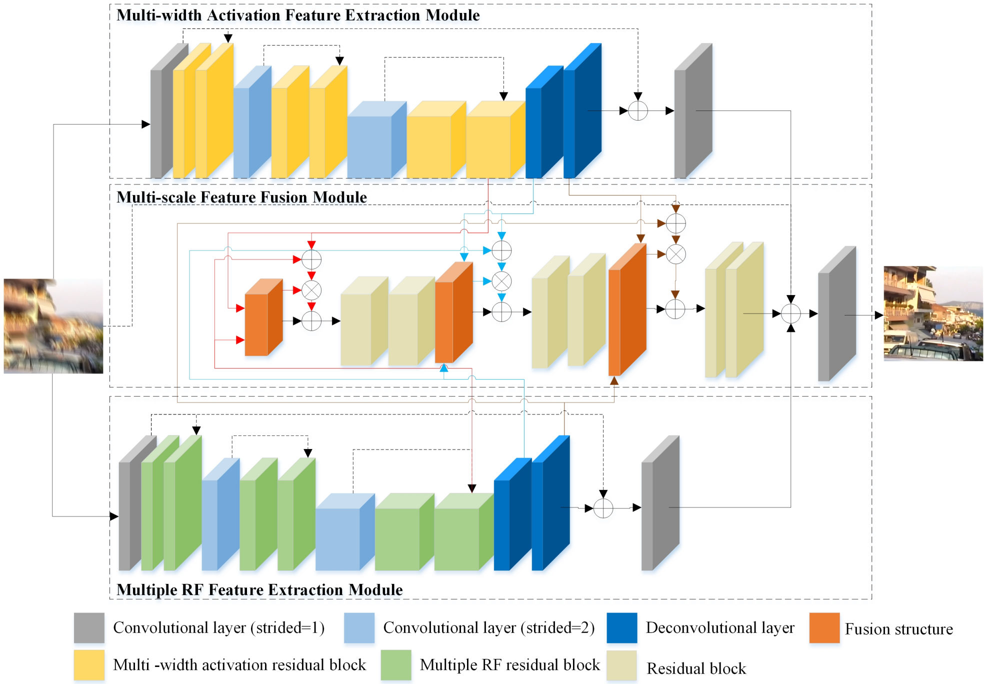

Multi-width Activation Feature Extraction Module
Multi-scale Feature Fusion Module
Multiple RF Feature Extraction Module
Convolutional layer (strided=1)
Convolutional layer (strided=2)
Deconvolutional layer
Fusion structure
Multi -width activation residual block
Multiple RF residual block
Residual block

Figure 1: The architecture of the proposed multi-width activation and multiple receptive field networks.

3.2. Multi-width Activation Feature Extraction Module

Since the blur situations of the same dynamic scene image are usually extremely complicated, we design a multi-width activation feature extraction module. And recently the encoder-decoder networks have been shown to produce great results for various generative tasks Pathak et al. (2016); Su et al. (2017). Therefore, we utilize an asymmetric residual encoder-decoder structure in this module. The encoder comprises of three scales, where each scale contains two multi-width activation residual blocks as shown in Fig. 2 and two strided convolutional layers with stride 1/2 to downsample the feature maps. The decoder includes two transposed convolution layers to enlarge the spatial resolution of the feature maps. In addition, the skip connections are used for avoiding the issue of vanishing gradients and accelerating convergence. Moreover, wider features before ReLU activation have been proven significantly better performance Yu et al. (2018). The proposed multi-width activation residual block contains two parts: multi-width activation and local residual learning.

Multi-width activation We design a dual-branch structure in this block and different branches use 3 × 3 and 5 × 5 convolution layer, respectively. In this way, the information between those branches can be shared with each other so that able to extract the wider image features at different scales. Moreover, in order to greatly increase the nonlinear characteristics without losing the any image information and obtain wider activation, we expand channel numbers by using 1 × 1 convolution layer and add the instance normalization (InstanceNorm) Ulyanov et al. (2016) and Leaky Rectified Liner Unit (LeakyReLU) Xu et al. (2015) with $\alpha = 0.2$ after the convolution layer. Then it is a stack of one 1 × 1 convolution layer to reduce number of channels. The expansion and reduction factors of the number of channels are 8 and 0.8 respectively. And then two convolution layers with the filter size of 3 × 3 and 5 × 5 to perform spatial-wise feature extraction. Finally, all of these feature maps are concatenated and sent to a 1 × 1 convolution layer for final features fusion.

Local residual learning We adopt local residual learning in each block for making the network more effective. The multi-width activation residual block is described as:

$$F_n = M' + F_{n-1} \tag{1}$$

where F_n and F_{n-1} represent the input and output of the multi-width activation residual block, respectively. The operation $M' + F_{n-1}$ is performed by a skip connection and element-wise addition. It is worth noting that the computational complexity is greatly reduced by using local residual learning. Meanwhile, the performance of the proposed network is improved.

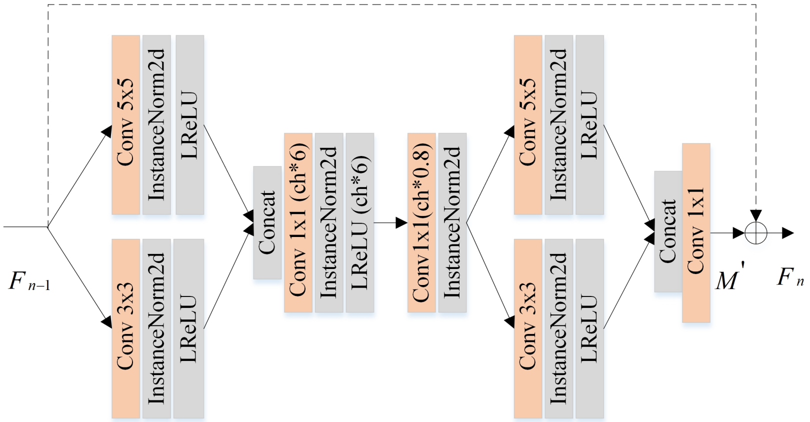


Figure 2: Multi-width activation residual block.

3.3. Multiple Receptive Field Feature Extraction Module

The residual block He et al. (2016) is proposed to ease the training difficulty of the deeper networks, as shown in Fig. 3(a). And the inception block Szegedy et al. (2015) takes different scale features simply concatenate together as shown in Fig. 3(b), which leads to the loss of some local features. In addition, dealing with large-scale blur usually requires a larger RF Jin et al. (2018), meanwhile to capture more nonlinear information, we design a multiple RF residual block with different receptive fields by using convolution kernels of different sizes to make better use of the advantages of the above structure, as shown in Fig. 3(c). The encoder in this module is constructed using three scales, where each scale contains two multiple RF residual blocks and two strided convolutional layers with stride 1/2 to downsample the feature maps. The block consists of two 3 × 3 convolution layers with the same dilation rate, two convolution layers with the filter size of 3 × 3 and 5 × 5, and a 1 × 1 output layer for multi-scale feature fusion. Furthermore, we add InstanceNorm layer and employ the LeakyReLU with $\alpha = 0.2$ as the activation function of multiple RF residual block instead of ReLU. It can void the "dead features" caused by zero gradients in ReLU. Simultaneously, the skip connection is used between different scales so that the feature information can be shared and complemented with each other. The decoder includes two transposed convolution layers to enlarge the spatial resolution of the feature maps. Moreover, we adopt local residual learning in each multiple RF residual block for strengthening feature propagation and reducing its complexity.

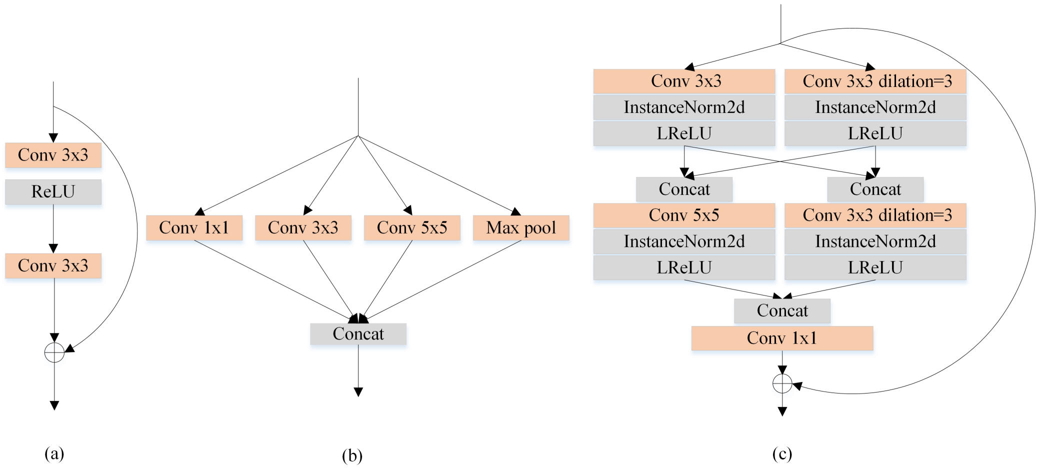


Figure 3: (a)Residual block. (b)Inception block. (c)The proposed multiple RF residual block.

3.4. Multi-scale Feature Fusion Module

It has been observed that fusing convolutional features obtained from different scales can lead to a better representation of an object in the image and its surrounding context Zhang and Patel (2018); Zhang et al. (2018b). In order to adaptively fuse the features and facilitate the convergence of the training process, we design the multi-scale feature fusion module, which contains three scale fusion structures, and each fusion structure followed three residual blocks as shown in Fig. 3(a), two transposed convolution layers, a convolution layer with ReLU and a convolution layer with Tanh. The most important part of this module is the fusion structure, which mainly consists of two convolutional layers with the filter size of 3 × 3 and 1 × 1 and three pixel by pixel operations. The fusion structure F generates a pixel-wise weight maps by blending ψ_{f1_i}, which represent features of three different scales from the multi-width activation feature extraction module, and ψ_{f2_i}, which represent the features of three different scales from multiple RF feature extraction module. Furthermore, our fusion structure takes the set of ψ_{f1_i}, ψ_{f2_i} and $\psi_{fusion_{i-1}}$, as the input, where $\psi_{fusion_{i-1}}$ is obtained via a transposed convolution layers, in addition to F only takes as input the set of ψ_{f1_i} and ψ_{f2_i}, when i equal to 1. It is expressed as:

$$C = F_i(cat[\psi_{f1_i}, \psi_{f2_i}, \psi_{fusion_{i-1}}]) \tag{2}$$

$$A = (\psi_{f1_i} \oplus \psi_{f2_i}) \tag{3}$$

$$\psi_{fusion_i} = (C \otimes A) \oplus C \tag{4}$$

where cat represents concatenation, $\otimes$ represents the element-wise multiplication and $\oplus$ represents the element-wise addition. The values of i are from 1 to 3. The fusion structure mainly consists of two convolutional layers with the filter size of 3 × 3 and 1 × 1.

Finally, to further refine the estimated three coarse deblurring images and make sure better details well preserved, we concatenate three coarse deblurring image, and then feed into a convolutional layer with ReLU and a convolutional layer with Tanh as the final clean image refinement.

3.5. Loss Function

We train our network by jointly optimizing multi-module loss L_{MM}, perceptual loss L_P and adversarial loss L_A. The loss function of our entire network is designed as:

$$L = L_{MM} + L_P + L_A \tag{5}$$

The multi-scale loss has been proven to achieve good deblurring effects Nah et al. (2017); Tao et al. (2018). So we introduce a multi-module loss to extract more available features and texture details of dynamic scene blur image in each feature extraction module. We use MSE loss in each module of output and the ground truth. The proposed multi-module loss function is calculated as the following:

$$L_{MM} = \sum_{i=1}^{3} L_{MSE}(I^{B_i}, I^S) \tag{6}$$

where I^S is a sharp dynamic scene image, I^{B_i} are latent sharp images of the corresponding modules.

Recently, perceptual and adversarial loss Kupyn et al. (2018); Isola et al. (2017) are proven to generate higher quality images indistinguishable from real images. We introduce the perceptual loss and adversarial loss to constrain the final output of multi-scale feature fusion module. Our perceptual loss function is expressed as:

$$L_P = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\Phi_{i,j}(I^S)_{x,y} - \Phi_{i,j}(G_{\Theta_G}(I^{B_3}))_{x,y})^2 \tag{7}$$

where I^{B_3} is final restored image, $\Phi_{i,j}$ is the feature achieved from the $j-th$ convolution layer before the $i-th$ maxpooling layer within the $VGG19$ network, which is a well-trained on ImageNet Deng et al. (2009). $W_{i,j}$, $H_{i,j}$ are the width and height of the feature maps. In this paper, we utilize activations from $VGG_{3,3}$ convolution layer.

For adversarial loss, following the architecture of method Kupyn et al. (2018), we adopt WGAN-GP Gulrajani et al. (2017) as the adversarial loss function, which has been verified to be robust to the choice of generator architecture Arjovsky et al. (2017). We build discriminator as in Table 1. The adversarial loss function is given as follows:

$$L_A = \sum_{n=1}^{N} -D(G(I^B)) \tag{8}$$

where G represents the generator, and D represents the discriminator. I^B is the input blurred image.

Table 1: The structure of discriminator. Except for the last layer, all the convolution layers are followed by InstanceNorm layer and LeakyReLU with α= 0.2.

Layer	Channel	Kernel size	Stride
Conv1	3 × 64	4 × 4	2
Conv2	64 × 128	4 × 4	2
Conv3	128 × 256	4 × 4	2
Conv4	256 × 512	4 × 4	1
Conv5	512 × 1	4 × 4	1

4. Experimental Setup

4.1. Datasets

GoPro dataset Nah et al. (2017) is used for training and testing our method, which includes 3214 paris of blurred and ground truth images with the size of 1280 × 720. Following the same ways as in Nah et al. (2017), we utilize 2103 pairs as training dataset and the remaining 1111 pairs for test dataset. Notably, the blur image provided by this paper are more realistic because it can simulate complex camera and spatially varying blurs caused by objects motion that are common in real scenarios and the static background. We further evaluated the state-of-the-art methods and our method on other datasets, which are Köhler dataset Köhler et al. (2012), which contains 48 blurred images with the size of 800 × 800, Kupyn dataset Kupyn et al. (2018) that contains 1151 images with size of 720 × 720 and Su dataset Su et al. (2017) that includes 6708 images from multiple devices including Canon $7D$, GoPro Hero 4, and iPhone $6s$.

4.2. Training Details

For the training phase, we utilize the PyTorch deep learning framework for training and testing our network on a single NVidia Titan GPU, and MATLAB for all peak signal-to-noise ratio (PSNR) and self-similarity measure (SSIM) evaluation. Different from these methods, we employ the image patches of size 320 × 320 as input instead of 256 × 256 , which will be beneficial to learn more about effective information of blurs. We perform 5 gradient descent steps on D, then one step on G. In addition, our model was trained with a batch size = 1. And we use ADAM optimizer Kingma and Ba (2014) with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The learning rate is set initially to 10^{-4} for both generator and discriminator. All trainable variables are initialized with the method Kupyn et al. (2018) is same. During the fine-tuning, after the first 40 epochs we linearly decay the learning rate to zero over the next 40 epochs.

5. Experimental Results

5.1. Quantitative Evaluations

Table 2 shows the average PSNR and SSIM values of the restored images on the GoPro and Köhler datasets. We can observe from the quantitative measures in Table 2 that our method perform favorably against with state-of-the-art methods in terms of PSNR and SSIM. These generated results reveal that our method can effectively improve the quality of restored images and have much higher PSNR and SSIM values. In addition, we also evaluate the average test time (s) for images with the size of 1280 × 720 pixels on GoPro dataset. In contrast to end-to-end method Kupyn et al. (2018), our method take slightly more time, but our method can generate much clearer images with higher PSNR and SSIM values.

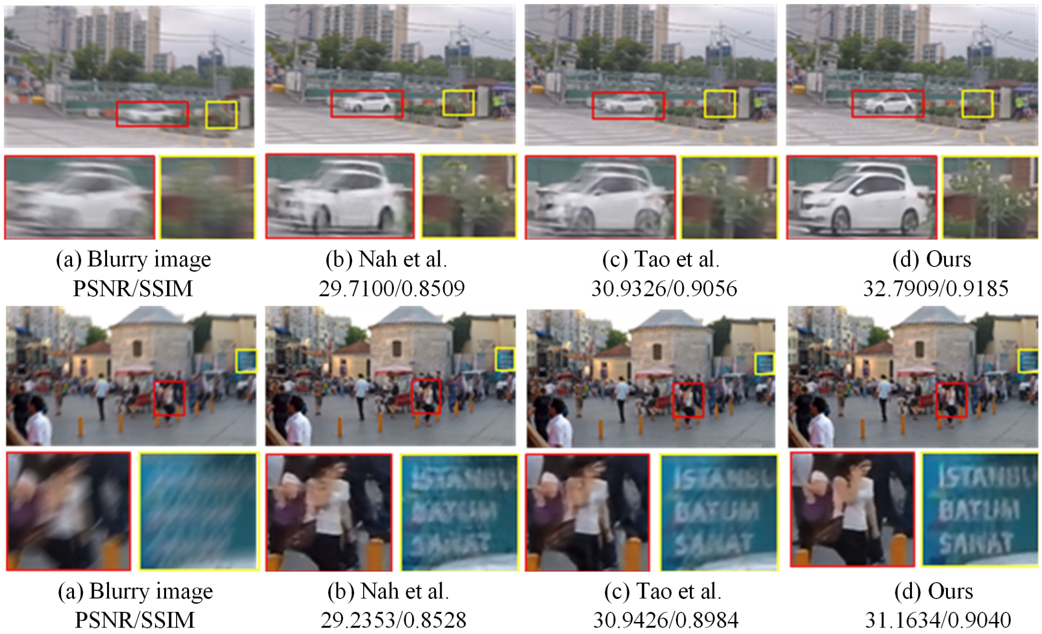
Table 2: Quantitative results comparison with state-of-the-art methods on GoPro and Köhler dataset.

Method	GoPro dataset		Köhler dataset		Time
	PSNR(dB)	SSIM	PSNR(dB)	SSIM	
Nah et al. (2017)	29.5762	0.8708	26.4581	0.8084	6.59
Kupyn et al. (2018)	28.8248	0.8507	26.1088	0.8163	0.85
Tao et al. (2018)	31.0659	0.9085	26.7597	0.8371	1.55
Zhang et al. (2018a)	30.1978	0.9013	25.7146	0.8000	1.45
Ours	31.5027	0.9118	29.0208	0.8957	1.30

5.2. Qualitative Evaluations

5.2.1. Results on GoPro evaluation dataset

To fairly compare with other methods, we further qualitatively evaluate the method on GoPro evaluation dataset. In Fig. 4, we show two visual comparison examples generated by the proposed method and state-of-the-art methods. These blurry examples caused by scene depth variations, object motions and camera shake, and contain local large scale blur. From Fig. 4, in addition to achieving the highest PSNR and SSIM values, the presented method restores the clear images with much clearer structures and finer texture details than existing state-of-the-art methods.



(a) Blurry image PSNR/SSIM

(b) Nah et al. 29.7100/0.8509

(c) Tao et al. 30.9326/0.9056

(d) Ours 32.7909/0.9185

(a) Blurry image PSNR/SSIM

(b) Nah et al. 29.2353/0.8528

(c) Tao et al. 30.9426/0.8984

(d) Ours 31.1634/0.9040

Figure 4: Deblurring results on GoPro dataset. The proposed method generates much clearer images with higher PSNR and SSIM values.

5.2.2. Results on other real datasets

We further qualitatively evaluate our method on other blurred images. Fig. 5 shows two different examples from Kupyn dataset (1 row) and Su dataset (2-3 rows). The results generated by the proposed method and state-of-the-art methods. Visual comparisons clearly show the superior performance of the proposed method over previous state-of-the-art methods, especially for fine details such as numbers, moving objects and texts.

6. Analysis and Discussions

6.1. Effectiveness of the Multi-width Activation and Multiple RF Feature Extraction Module

To demonstrate the effectiveness of the designed modules, we perform four related experiment on GoPro dataset. To demonstrate the effectiveness of the designed activ-module and RF-module, we remove activ-module and RF-module, respectively. To prove the effectiveness of the designed activ-residual and RF-residual blocks, we replace active-residual and RF-residual by the residual block. We call two modules and two blocks as **activ-module**, **RF-module** and **activ-residual**, **RF-residual**, respectively. We show the deblurring results generated by these four networks and the proposed method in Fig. 6. From Fig. 6(b, c), the deblurring results without multi-width activation feature extraction module and multi-width activation residual block still contain blur residual, especially in the detail

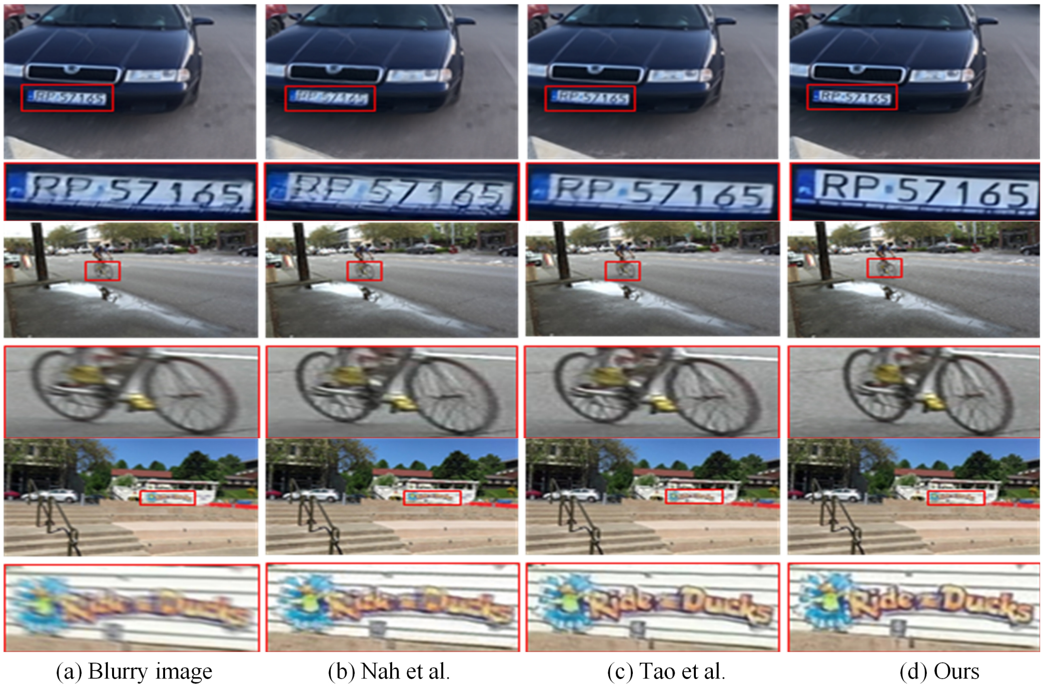
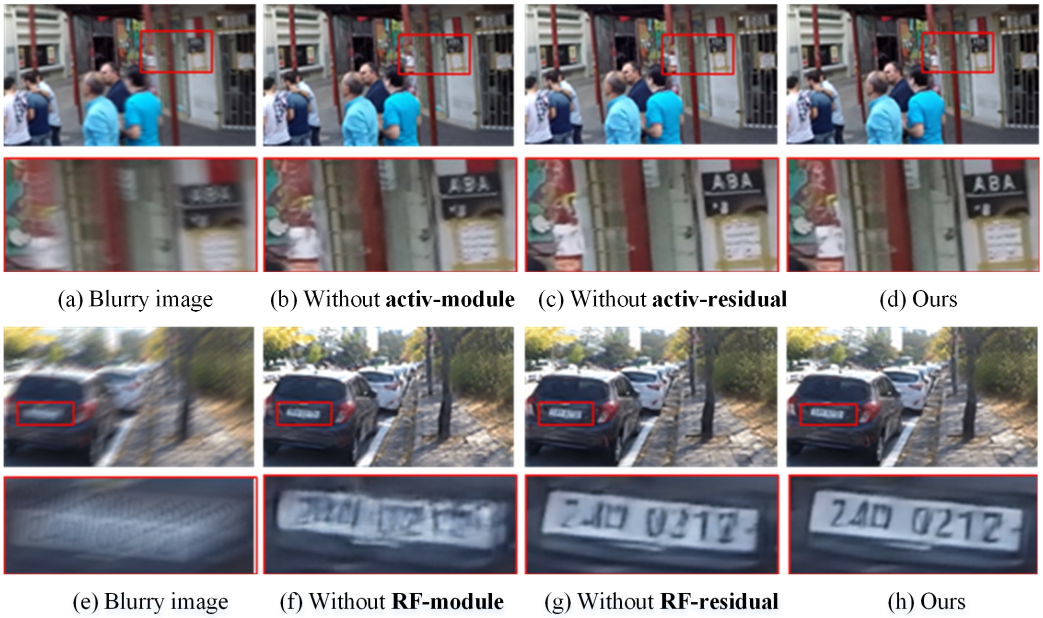


Figure 5: Deblurring results on other datasets.

regions. Fig. 6(e) is a large-scale example, and our method is able to generate a clearer image by adding the multiple RF feature extraction module and multiple RF residual block, as presented in Fig. 6(f-h). Moreover,we visualize the output of these modules as shown in Fig. 7(a, b). It is obvious that the output of these module contains more effective activation maps and complex nonlinear information, which further verifies the effectiveness of our modules. Moreover, the average PSNR and SSIM comparison with these four models on the GoPro test datasets are shown in Table 3. From the Table 3, it is clearly that the proposed method yielded significantly better PSNR / SSIM than the corresponding four models. In view of the above, the designed multi-width activation and multiple RF features extraction module can effectively improve the quality of large-scale and complex blur images.

Table 3: Quantitative results comparison with different models on on GoPro dataset.

Method	Real GoPro		Synthetic GoPro	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Without **activ-module**	29.4089	0.8860	27.7811	0.8704
Without **activ-residual**	30.7619	0.9001	28.6984	0.8896
Without **RF-module**	29.1619	0.8751	27.3984	0.8689
Without **RF-residual**	30.5039	0.8907	28.4284	0.8790
Ours	31.5027	0.9118	29.0208	0.8957


(a) Blurry image (b) Without **activ-module** (c) Without **activ-residual** (d) Ours

(e) Blurry image (f) Without **RF-module** (g) Without **RF-residual** (h) Ours

Figure 6: Visual comparison of our method with other verification models on GoPro dataset.

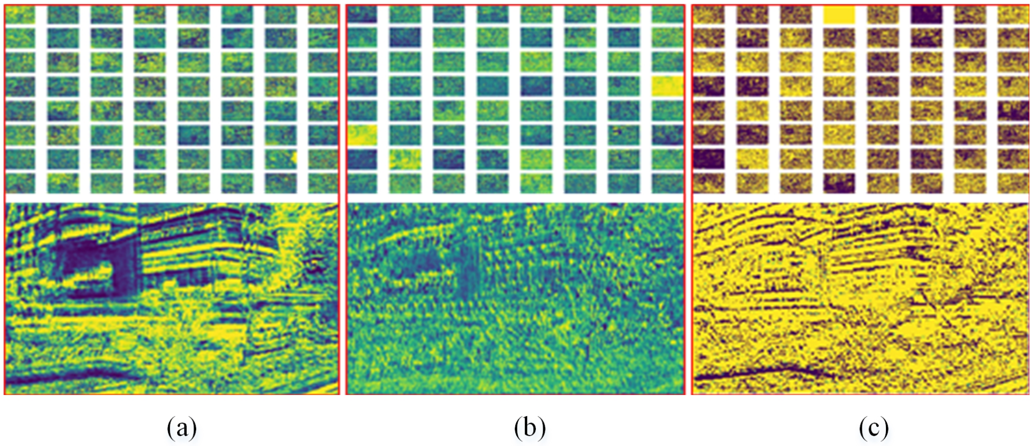

(a) (b) (c)

Figure 7: The visualization features (a) and (b) from **activ-module** and **RF- module**. (c) is fused by our designed fusion structure.

6.2. Effectiveness of the Multi-scale Feature Fusion Module

In order to further illustrate the effectiveness of the designed multi-scale feature fusion module, an ablation experiment is performed on GoPro dataset. We compare the proposed network with the network without fusion structure (FS). We show the visual features extracted by the **activ-module** and **RF- module** and FS in Fig. 7(a-c). From Fig. 7(c), we can see that it mainly includes more effective activation maps and rich details. We also show the visual results generated by these two methods in Fig. 8. As show in Fig. 8(b), the deblurring result without using FS still contains blur residual, and many details are still not clear enough. Compared to the proposed method without FS, a clearer image with clearer details can be recovered by the proposed method as show in Fig. 8(c).

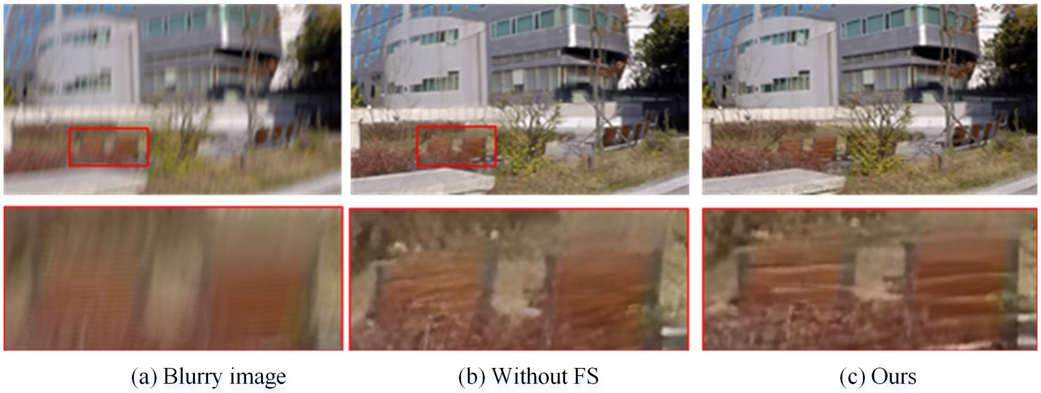


Figure 8: Visual comparison with the proposed method without FS and with FS on GoPro dataset.

Correspondingly, quantitative results comparison with the proposed method with FS and without FS on real dynamic scene and synthetic dynamic scene dataset are shown in Table 4. It is obvious that the proposed method outperforms the proposed method without FS. Concretely, the average PSNR and SSIM gain of the proposed method over the proposed method without FS are $1.0997dB$, 0.0166 and $0.9151dB$, 0.0193, respectively, in two dataset experiments. Thus, the experimental results well prove the effectiveness of the fusion structure.

Table 4: Quantitative results comparison with the proposed method with FS and without FS on GoPro dataset.

Method	Real GoPro		Synthetic GoPro	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Without FS	30.4030	0.8952	28.1057	0.8764
Ours	31.5027	0.9118	29.0208	0.8957

7. Conclusion

We propose an end-to-end multi-width activation and multiple receptive field networks for large-scale and complicated dynamic scene blurs. In the proposed method, a multi-width activation feature extraction module is designed, where a multi-width activation residual block is proposed for making the activation function learn more the nonlinear information and extracting wider nonlinear features. Furthermore, a multiple RF feature extraction module is designed, in which a multiple RF residual block is proposed for enlarging the RF efficiently and capturing more nonlinear information from distant locations. And then, the multi-scale features fusion module is designed to adaptively fuse the multi-scale features and complicated blur information from the different modules. Extensive experiments are performed on GoPro and other datasets. The experimental results have demonstrated the proposed method significantly outperforms state-of-the-art methods.

References

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3848–3856, 2019.

Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2319–2328, 2017.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777, 2017.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Tae Hyun Kim and Kyoung Mu Lee. Segmentation-free dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2766–2773, 2014.

Tae Hyun Kim, Byeongjoo Ahn, and Kyoung Mu Lee. Dynamic scene deblurring. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3160–3167, 2013.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

Meiguang Jin, Michael Hirsch, and Paolo Favaro. Learning face deblurring fast and wide. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 745–753, 2018.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European conference on computer vision*, pages 27–40. Springer, 2012.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8183–8192, 2018.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3883–3891, 2017.

Mehdi Noroozi, Paramanand Chandramouli, and Paolo Favaro. Motion deblurring in the wild. In *German conference on pattern recognition*, pages 65–77. Springer, 2017.

Jinshan Pan, Zhe Hu, Zhixun Su, Hsin-Ying Lee, and Ming-Hsuan Yang. Soft-segmentation guided object motion deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 459–468, 2016.

Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.

Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1279–1288, 2017.

Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 769–777, 2015.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

Jiahui Yu, Yuchen Fan, Jianchao Yang, Ning Xu, Zhaowen Wang, Xinchao Wang, and Thomas Huang. Wide activation for efficient and accurate image super-resolution. *arXiv preprint arXiv:1808.08718*, 2018.

He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018.

Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2529, 2018a.

Xinyi Zhang, Hang Dong, Zhe Hu, Wei-Sheng Lai, Fei Wang, and Ming-Hsuan Yang. Gated fusion network for joint image deblurring and super-resolution. *arXiv preprint arXiv:1807.10806*, 2018b.

Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, pages 479–486. IEEE, 2011.