



# MFC-Net: Multi-scale fusion coding network for Image Deblurring

Haiying Xia<sup>1</sup> · Bo Wu<sup>1</sup> · Yumei Tan<sup>2</sup> · Xiaohu Tang<sup>1</sup> · Shuxiang Song<sup>1</sup>

Accepted: 8 November 2021 / Published online: 25 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

The existing image blind deblurring methods mostly adopt the “coarse-to-fine” scheme, which always require a mass of parameters and can not mine the blur information effectively. To tackle the above problems, we design a lightweight multi-scale fusion coding deblurring network (MFC-Net). Specifically, we fuse the multi-resolution features in a single-scale deblurring framework based on Wasserstein generative adversarial network (WGAN). Then we propose a feature fusion module to replace the addition operation in each scale in the skip connection of the encoder-decoder. Besides, we propose a regional attention module to alleviate the inconsistency in non-uniform blurry images and excavate its intrinsic blurry features simultaneously. Plenty of experimental results show that our proposed deblurring model is simple, fast yet robust for image motion deblurring with real-time inference of 10 ms per 720p image, outperforming the state-of-the-art methods in terms of performance-complexity trade-off.

**Keywords** Single image motion blind deblurring · Multi-scale fusion coding · Wasserstein generative adversarial networks · Regional attention module

## 1 Introduction

Image deblurring, that is, restoring a sharp latent image from the blurry image, has long been a vital research for decades. It aims to restore a sharp image with the necessary edge details and the overall structure information in view of the motion blurs caused by camera translation, rotation, and other factors. Conventional deblurring approaches usually depend on different blur kernels approximations by taking various constrained priors (e.g., color, local smoothness, non-local self-similarity, and sparsity) as regularization terms to improve the effect of image deblurring [7, 18, 22–24, 26]. However, it is a daunting task to obtain the accurate blur kernel approximations with kernel-estimation methods, which involve lots of fixed parameters and complicated calculations, especially in the cases of non-uniform motion blurs.

Recently, with the success of deep learning techniques on a wide variety of tasks, the deep neural network has been increasingly applied on image deblurring. Early works utilized Convolution Neural Network (CNN) [19] to replace some modules or steps in traditional framework to facilitate the deblurring process, but they are only applicable to some specific blurry types and they have limitations on variable ones. To this end, some representative works explored solutions for blind motion deblurring in an end-to-end manner without kernel estimations. Nah et al. [25] proposed a new paradigm to tackle dynamic deblurring caused by various sources in an end-to-end manner. They designed a multi-scale convolution neural network to extract features, and improved deblurring performance by Generative adversarial network (GAN) [9]. Afterwards, many works [20, 21, 38, 43] have designed deblurring network based on [25], including single-scale network, multi-scale residual network and deep multi-patch network. In general, existing image deblurring frameworks based on deep learning can be grouped into two categories: 1) one way is to incorporate feature pyramid or residual network based on the single-scale WGAN [1, 2, 20, 21], and add the perceptual loss [16] to enhance deblurring performance, and 2) another way is to leverage the multi-scale residual network or the deep multi-patch network [38, 43] for image deblurring by the deep stacking

---

✉ Shuxiang Song  
songshuxiang@mailbox.gxnu.edu.cn

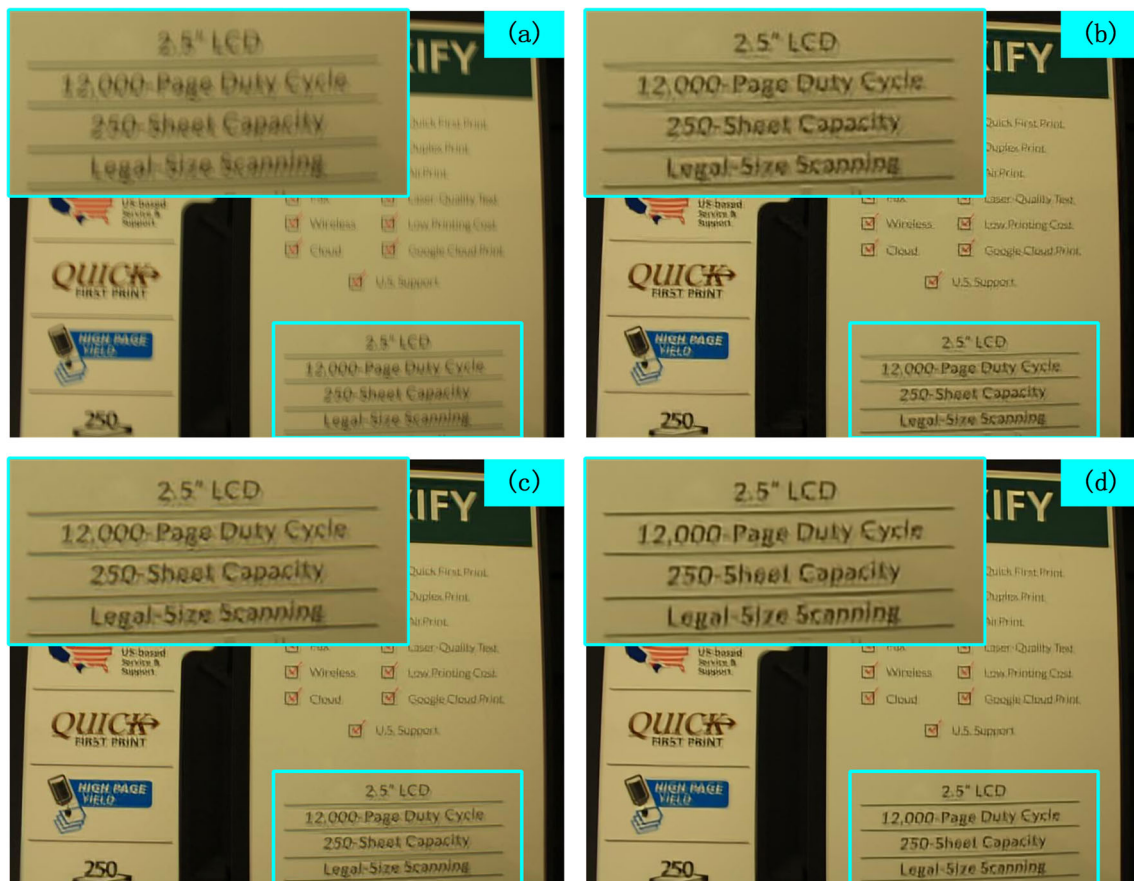
<sup>1</sup> College of Electronic Engineering, Guangxi Normal University, Guilin, 541004, China

<sup>2</sup> School of Computer Science and Engineering, Guangxi Normal University, Guilin, 541004, China

strategies or the “coarse-to-fine” schemes. However, these deblurring methods still suffer from three limitations: 1) The former enables to achieve much faster speed, while the latter enables to achieve better performance, both of them can not make a trade-off between the speed and the performance, and 2) for feature fusion, it is not conducive to make cross-scale fusion for the network learning by simply addition or directly cascade, and 3) they can not perform well on the blurry image with non-uniform blurs and diverse blurs for ignoring the non-uniform nature of blur. As we all know, it is challenging to tackle the non-uniform blurry images due to its diversity, such as the different blur degrees among different images, the inconsistent blurs between different blurry regions of the same image. Therefore, how to effectively and deeply mine the blur characteristics of non-uniform blurred images is very important. However, prior works restore the natural and clear images only by optimizing the GAN-based framework [20, 21, 25], or learning the blur characteristics by transferring the multi-scale blur features with the long-term short-term memory (LSTM) [38], which can't not produce satisfactory results, as seen in Fig. 1(b), there still existed obvious blur in the restored image. For non-uniform

deblurring, we argue that focusing on the blur characteristics of local regions is of vital importance for improving deblurring performance.

Based on the above observations, we propose a multi-scale fusion coding deblurring architecture based on WGAN [1, 2] with its superiority of training stability. Firstly, we design a generator by leveraging multi-scale coding parallel architecture, which not only fuse and construct image information according coding different scale features, but also avoid expensive running time caused by the scale-iteration training. Due to the inconsistent blurs between different blurry regions, a regional attention model is further proposed to enhance the local deblurring performance by focusing on blurring features of regions in image, which strongly manifest the usefulness of mining blurry information on deblurring performance. Additionally, we propose a feature fusion model to make effective fusion between different dimensional features to facilitate the generalization ability of the model. Figure 1d shows that the proposed MFC-Net is able to obtain much better deblurring performance than the methods (i.e. Figure 1b and c). The contributions of our paper can be summarized as follows:



**Fig. 1** A real deblurring example. (a) Input blurred image. (b) Result of SRN [38]. (c) Result of Gao et al. [8]. (d) Our result

1. We propose a powerful and efficient blind image deblurring network named multi-scale fusion coding deblurring network (MFC-Net), which is conveniently applied for image deblurring with different resolutions. The multi-scale fusion coding architecture is adopted to construct the generator by adding multi-scale information to enhance deblurring performance with few weight parameters and less runtime.
2. A regional attention module is introduced to tackle the inconsistent blurs of each region in the blurry image. The blurry images are first divided into regions, then the blurry features of different regions are conducted dynamical deblurring to force the network to focus on local deblurring.
3. A feature fusion module, which can effectively mine the correlation between features and easily transplant to the network, is exploited to further enhance deblurring performance almost without computational overhead.

The remainder of this paper is organized as follow. Section 2 introduces the related work. Section 3 details the proposed MFC-Net. In Section 4, we evaluate the proposed MFC-Net and make discussions. The conclusions are given in Section 5.

## 2 Related work

In this section, we briefly review three main topics that are related to our work: Image Deblurring, Wasserstein Generative Adversarial Networks (WGAN) and Attention mechanism.

### 2.1 Image Deblurring

Image deblurring can be divided into two categories: blind and non-blind deblurring. Early works mainly focus on non-blind deblurring, most of them utilized probabilistic prior model to work out the regularization items for the given blurred image and blurred core [6, 14, 31, 42]. However, these non-blind deblurring methods would yield over-smooth details, which significantly affect the visual result. While the Blind deblurring, of which the blur kernel is unknown, is much more challenging to recover both the latent sharp image and the blur kernel by only observation, but it is crucial in practical application. With the recent advancement of deep learning, many works exploit deep CNN models to improve the performance of blind image deblurring. Schuler et al. [32] proposed a blind deblurring method by stacking multiple CNNs in a coarse-to-fine manner to simulate iterative optimization, kernel estimation, and the latent image estimation steps. In addition, they applied a Gaussian Process to generate

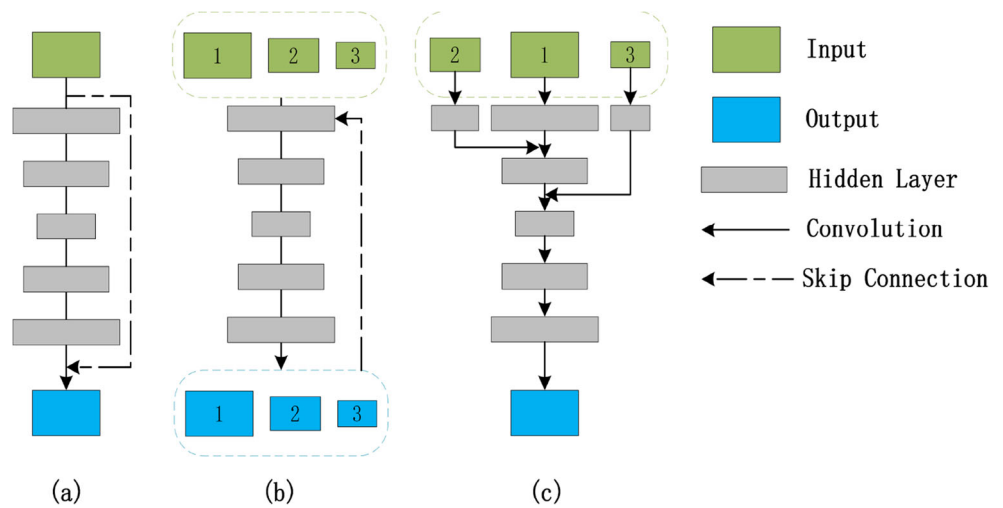
uniform blur kernels and blurry/sharp image pairs, and synthesized abundant blurry images. Bai et al. [4] proposed a two phases blind deblurring method by leveraging a multi-scale latent structure prior. They restored sharp images from the coarsest scale to the finest scale on a blurry image pyramid, and progressively updated the prior image by using the newly restored sharp image. However, the blind deblurring algorithms that estimate both the potential sharp image and the blurring kernel for finding out the unknown blurring function is ill-posed.

Therefore, contrary to the kernel-estimation methods for blind deblurring, Nah et al. [25] developed a multi-scale architecture that progressively restored the sharp images at different resolutions without blur kernel estimation in an end-to-end manner. They preserved fine-grained detail information from coarser scales by training model with multi-scale losses, and achieved better performance. Since then, more works have been explored deep learning methods for image deblurring without blur kernel estimation. There are two representative frameworks on image deblurring: single-scale network and multi-scale network.

**Single-scale network** A simplified single-scale network is shown in Fig. 2a. For instance, Kupyn et al. [20] proposed a DeblurGAN framework for blind motion deblurring based on Conditional Adversarial Network with optimized multi-component loss function, Firstly, they introduced multi-layer residual blocks (ResBlock) in the deblurring network, and made the global skip connection between the input and the output, which reduced restoration difficulty and ensured color consistency. Later, Kupyn et al. [21] proposed DeblurGAN-v2 by introducing the feature pyramid network into deblurring generator, and by adopting light-weight backbone to speed up the deblurring process. What's more, a double-scale discriminator and the perceptual loss [16] were utilized to facilitate the deblurring performance. Similar to the deblurring framework of [20], Zhou et al. [44] added a context module to integrate more rich hierarchical context information, which was proved to be beneficial to blur elimination and parallax estimation. Yuan et al. [9] designed a spatially variant deconvolution network by leveraging deformable convolutions to guide the network learning blurring information, and by utilizing bi-directional optical flows to supervise sampling points of deformable convolution for better deblurring performance. However, it led to complex computation and abundant inference time on the deformable convolutions.

**Multi-scale network** Tao et al. [38] proposed a Scale-Recurrent CNN (SRN), for blind motion deblurring by sharing weights across different scales based on LSTM, as shown in Fig. 2b. They extended the work [25] by adopting a scale-recurrent model and a LSTM, passing the

**Fig. 2** Different CNNs for image processing. (a) U-Net [30] or Kupyn et al. encoder-decoder [20] network. (b) Tao et al. scale-recurrent network [38]. (c) Our proposed network



hidden intermediate results and blur patterns to next scale, so as to progressively recover the latent image from lower resolutions to the full resolution. Zhang et al. [43] proposed a deep hierarchical multi-patch network to restore sharp images via a fine-to-coarse hierarchical representation. The results showed that the training strategy of stacking image regions had limitations: it achieved the best effect when the depth reached (1-2-4-8), but the increase was not obvious compared with (1-2-4), and the performance declined when the depth deepened. Suin et al. [36] proposed a deep patch-hierarchical attentive architecture by leveraging an effective content-aware global-local filtering module that significantly improved performance.

We observe that the state-of-the-art end-to-end methods above all make full use of the blurry information by optimizing the network strategy to enhance the nonlinear representation, no matter in multi-scale or in patch, which demonstrated that the full use of the blurry image information in the blind image deblurring is more helpful than the stacked convolution learning method. Inspired by [14, 20, 38], but differently, we integrate the multi-scale coding information into the encoder of the single-scale deblurring network by layers to enhance the nonlinear representation of the encoder, as shown in Fig. 2c. Our proposed network can improve the deblur performance within multi-scale inputs and multi-scale fusion coding in a single scale network, so as to avoid the redundant calculation cost in the multi-stage [38, 43].

**2.2 Wasserstein generative adversarial networks**

In recent years, GAN networks have been widely used in many tasks due to the superior generated performance, such as image style transformer and deblurring. Wasserstein GAN (WGAN) [1, 2] is one of the popular variants of GAN. Arjovsky et al. [1] used the Wasserstein-1 distance

to replace the JS divergence as as the objective function to tackle the training difficulties in traditional GAN. The value function for WGAN is constructed by using Kantorovich-Rubinstein duality:

$$\min_G \max_{D \in \mathcal{D}} E_{x \sim P_r} [D(x)] - E_{\tilde{x} \sim P_g} [D(\tilde{x})] \tag{1}$$

where  $D$  is the set of 1-Lipschitz functions and  $P_g$  is once again the model distribution.  $W(P_r, P_\theta)$ , where  $K$  is a Lipschitz constant and  $W(P_r, P_\theta)$  is a Wasserstein distance. Simultaneously, for improving the robustness of the model, Arjovsky et al. [1] enforced Lipschitz constraint, and added weighted clipping to  $[-c, c]$  in WGAN. Gulrajani et al. [10] propose to add a gradient penalty term instead:

$$\lambda E_{\tilde{x} \sim P_{\tilde{x}}} \left[ (\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2 \right] \tag{2}$$

to the value function as an alternative way to enforce the Lipschitz constraint. This approach is robust to the choice of generator architecture and requires almost no hyperparameter, which has been widely applied in deblurring tasks. Based on WGAN, DeblurGAN [20] achieved excellent results for motion deblurring, which yielded more realistic images. Hong et al. [13] exploited the WGAN network to restore more realistic information of the edge and texture. Chen et al. [5] utilized the WGAN to restore the tiny feature details of cell images, which achieved great success. To further restore more natural details for blurring images, DeblurGANv2 [21] and perceptual-DualGAN [29] introduced the perceptual loss [16] to generate high restoration quality. Thus, we also adopted the WGAN and the perceptual loss for better restoration performance in our work.



## 2.3 Attention mechanism

Attention mechanism plays an important role in stimulating the deep learning of human visual perception. Bahdanau et al. [3] first tried to apply attention mechanism to machine translation and achieved great success. Since then, a variety of attention models had been proposed for different tasks. Hu et al. [15] proposed a channel attention module by assigning different weights according to the contributions of each feature channel to effectively exploit the inter-channel relationship. Based on the channel attention, Woo et al. [40] proposed a convolutional block attention module (CBAM), which exploited both spatial and channel-wise attention to induce the network to focus on the target properly. For the task of blind image deblurring, Qi et al. [28] proposed a dense feature fusion block by joining a channel attention module and a pixel attention module to enhance the performance of image deblurring. By combining non-local features and local features to model the rich context correlations, the uncertainty maps can be learned effectively to guide the pixel loss for more robust optimization. Purohit et al. [27] exploited a self-attention model to capture non-local spatial relationships and enhanced the spatially varying shifts for non-uniform blur. Suin et al. [36] adopted the self-attention block in both encoder and decoder at each level and the cross-attention by combining the global dependencies and the neighboring pixel information to help the feature propagation across layers and levels. Different from those aforementioned approaches, we represent the non-uniform blurry properties by means of adaptive method, and learn the non-uniform blurry properties in different regions effectively. More concretely, inspired by the spatial pyramid pooling

strategy [11], we design a regional attention model to represent blurry information by conducting features spatial pooling and fusion, and combining the context module which has large receptive field, so as to enhance attention to local blur.

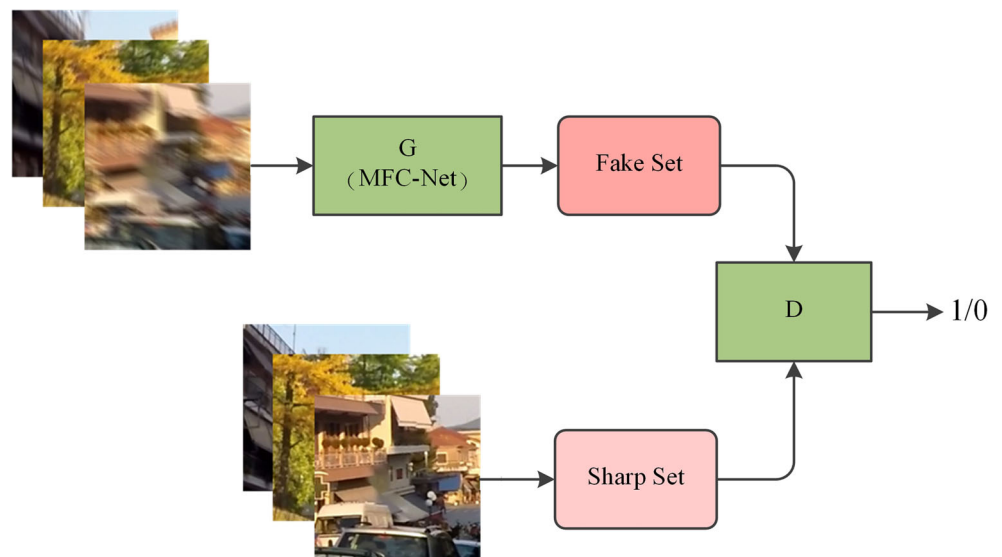
## 3 Our proposed method

The pipeline of the whole architecture is shown in Fig. 3. It consists of a generator G and a discriminator D. Specifically, we take a sequence of blurry images downsampled from an input image at different scales, and next feeds them to the generator G. Next the G extracts latent encodings of the inputs by a multi-scale encoder and generates new fake sharp images by a decoder. The discriminator D constructed by stacked deep convolutions is trained to distinguish between the corresponding real sharp images and the generated fake sharp samples until the generator G can generate the real-like sharp images that we need. To further enhance the restoration quality, a regional attention model and a feature fusion module are proposed for the generator G. Finally, to obtain much higher perceptual quality, we add a perceptual loss into the whole losses for training.

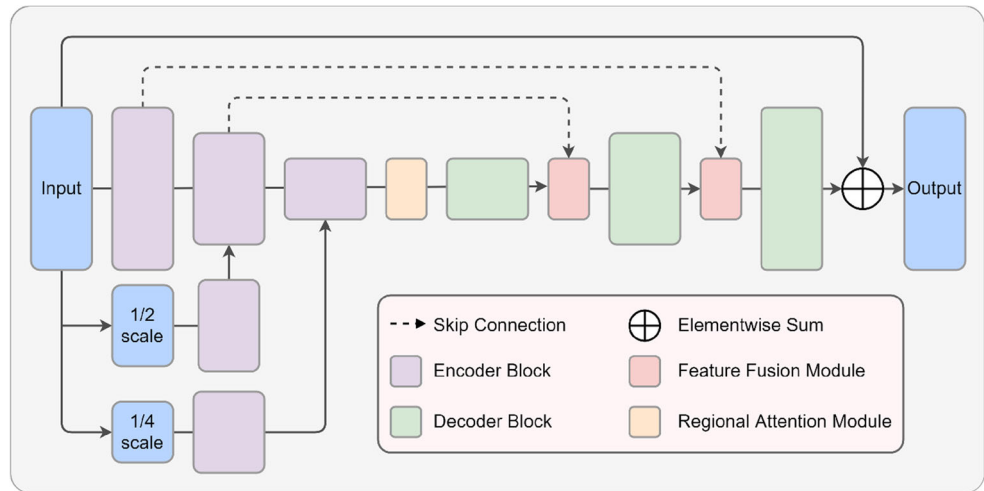
### 3.1 The proposed generator G

We design a MFC-Net as the generator G, as illustrated in Fig. 4. It contains three parts: Multi-scale Fusion Coding Architecture, Regional Attention Module (RA) and Feature Fusion Model (FF). We make detailed description in the following sections.

**Fig. 3** The pipeline of the whole architecture



**Fig. 4** The proposed generator G



### 3.1.1 Multi-scale fusion coding architecture

Since the seminal work of U-Net [30], skip connections between the corresponding encoder and decoder stages have been widely used in image/video restoration and raindrop removal owing to its superior pixel-wise regression ability. Benefit from this, similar U-Net architectures are exploited as the basic image deblurring framework, which are composed of convolution modules and multi-layer cascaded residual blocks [12]. After deep research on the representative deblurring works [20, 21, 25, 38], we find that it is essential to explore the pixel mapping relationship between the input blurry image and the output clear image according to different blur features, and utilize the multi-scale information of images as much as possible.

Additionally, based on the typical encoder-decoder structure [30], we need to consider the following factors for the deblurring task: 1) It is difficult to comprehensively learn the blur information by using only single-scale image in the encoder. 2) More encoder-decoder modules are stacked for large receptive field to deal with pretty blurry image, which will lead to a sharp increase of model parameters and the size of the intermediate feature map is too small to maintain the spatial distribution for reconstruction information. 3) More convolution layers at each level of the encoder-decoder module will slow down the convergence speed of the model and cause redundant parameters. Taking into account all these factors, we propose a multi-scale fusion coding framework by merging the multiscale feature information of the blurry image into the single scale deblurring network, which making full advantages of the multi-scale feature information and fast inference time of single scale deblurring framework. Especially, we adopt convolutions with different dilated rates to extract features from the 2 times and 4 times down sampled images, so as to enhance the representation ability of the encoder layer.

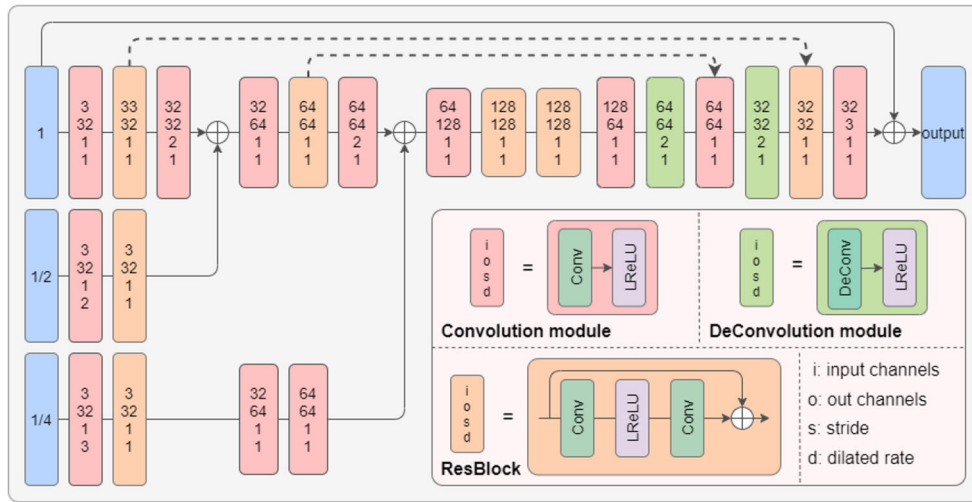
Since the Batch Normalization layer will disrupt the distribution among feature maps [25], we remove the Batch Normalization layer of the resblock of the encoder-decoder to make the model fit for our deblurring task. The proposed framework consists of convolution module and ResBlock. We construct the convolution module by cascading convolution layer and LeakyReLU activation functions. ResBlock is cascaded by convolution layer, LeakyReLU activation function, and convolution layer. Figure 5 illustrates the proposed multi-scale fusion coding network in detail. Firstly, we adopt twice subsamplings by residual learning to enhance the restoration quality. Then, we add multi-scale feature information for fusion coding to obtain large receptive field for deblurring. Meanwhile, we apply dilated convolutions with different dilated rates (i.e., 2 and 3) to 1/2 and 1/4 of the input image respectively.

### 3.1.2 Regional attention module

For most image deblurring tasks [21, 25, 38], lots of weights usually are utilized to learn the mapping relationship between the blurred image and the real image, which leads to heavy learning burden. So, we propose a regional attention module to adjust the feature weights dynamically according to the blurry representation of different regions, as illustrated in Fig. 6. Concretely, we train our model in patch manner. Firstly, the feature maps are sliced into patches closed to the size of the patch. Then, each patch is sent to our RA module for inference. Finally, we concatenate all the inference results. The whole RA model can be described as:

$$U_{out} = Conv\left[\sum_{i=1,2,3,4}^C (FS_i(S(f_x)) \cdot dConv_i(f_x))\right] + f_x \quad (3)$$

Where  $f_x$  represents the feature map of the output of the Encoder in Fig. 4, denoting the spatial pooling operation,



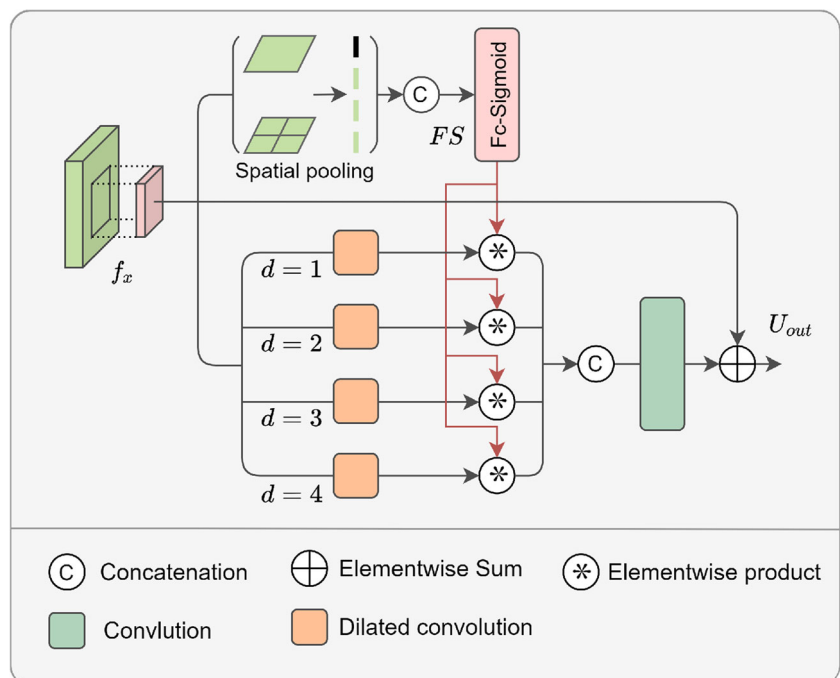
**Fig. 5** Our Multi-scale Fusion Coding Architecture. Including the input blurred image on the left (1, 1/2, 1/4 scale), the output deblurring image on the right (output), convolution modules, deconvolution modules, Resblocks. The four numbers of each module represent the channels of the input feature maps, the channels of output feature maps, the

convolution/deconvolution strides and the convolution/deconvolution dilated rates respectively. We set the size as 3 for all convolution kernels.  $\oplus$  represents element-wise summation, and the dotted lines represent skip connections with element-wise summation

$C(\cdot)$  denotes the concatenation operation,  $Conv$  denotes the convolution operation. Inspired by SPPNet [11], we divide the feature map  $f_x$  into  $1 \times 1$  and  $2 \times 2$  regions with a total of five groups of feature maps followed by the global average pooling and the concatenation operation to form the representation information.  $FS_i$  represents the  $i$ -th group of full connection (FC) layer and sigmoid function.  $dConv_i(\cdot)$  represents the  $i$ -th parallel of dilated convolutions with the

$i$ -th dilated rates (i.e.,  $i=1,2,3$  and 4), which can effectively cover different surrounding receptive field.  $FS_i$  outputs the  $i$ -th blurry attention value, and then weight it to the output feature vector of  $dConv_i(\cdot)$ . The weighted feature vectors are concatenated and fused through the  $Conv$ . We introduce skip connection between the inputs and outputs of fused features to conduct residual learning, which is capable of improving deblurring performance effectively. In general,

**Fig. 6** The structure of Regional Attention Module



the proposed RA module allows the network to focus on the blur information of local regions, which can significantly improve the performance of local deblurring.

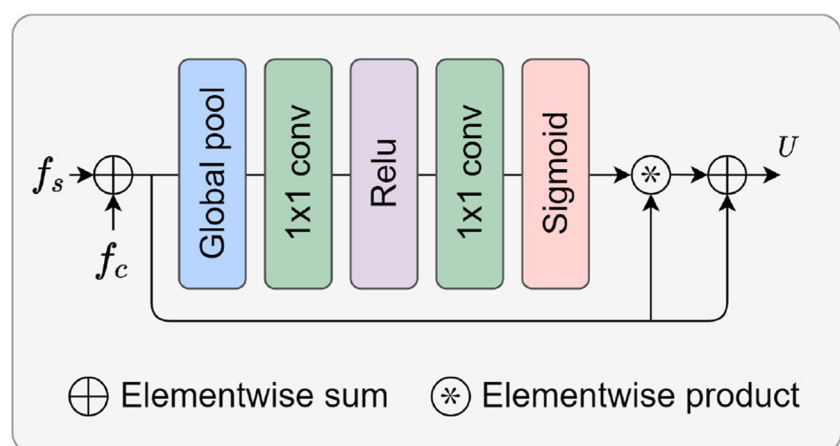
### 3.1.3 Feature fusion module

Recently, multi-scale blurry feature information are utilized to restore more details via different structures in deblurring tasks, which yielded more accurate results. Prior works in [9, 25] aggregated multi-scale feature information by connecting the shallow features with the global skip connection to obtain sufficient feature information to get better restored images. In [21], authors exploited feature pyramid for integrating multi-scale feature coding information to obtain clear images with more local textures and colors. Scale-recurrent network [38] restored sharp images from blurring images by using the “coarse-to-fine” scheme in mutiple scales. Deep hierarchical multi-patch network [43] restored sharp images from fine-to-coarse by stacking multi-patch. Although these methods are effective by making full use of multi-scale feature information of blurry images, most of them only adopt simple fused methods by concatenation or addition. In other words, they do not find out the correlation between features only through a simple way to integrate the high-level semantic information and low skip-level feature information in essence. To tackle these limitations above, we proposed an effective feature fusion model, as illustrated in Fig. 7.  $f_s$  represents the skip feature map of the encoder,  $f_c$  represents the high-level semantic feature map,  $f_s, f_c \in R^{H \times W \times C}$ . We first merge  $f_s$  and  $f_c$  as the input of the FF model, as shown in formula (4). Then we obtain the output  $U$  of the FF model. can be calculated by formula (5),where  $\cdot$  represents element-wise product.

$$f = f_s + f_c \quad (4)$$

$$U = F_{sq}(f) \cdot f + f \quad (5)$$

**Fig. 7** The structure of Feature fusion Module



In our FF model, the channel average response is generated by the global average pool, then it is weighted by convolution and sigmoid activation function, finally, the residual connection is added to make effective fusion of channel features. By this way, our proposed network is capable of improving the generalization ability with FF model without increased parameters.

### 3.2 Discriminator

The discriminator D is used to distinguish the generated clear image from the real clear image. It can not only judge whether the image is real or not, but also help to generate a clear image closer to the real image. If it is a real image, the output is a clear image generated by the generation network, otherwise, the output is 0. In both cases, for the image that cannot be judged, the generated image is very similar to the real image, and then the network model stops training. All convolution layers in the network are identified, followed by the InstanceNorm layer layer and LeakyRelu activation layer. In order to avoid the output of LeakyRelu function being 0, sigmoid function is used in the last layer.

Figure 8 illustrates the framework of our discriminator, where Conv, LeakyRelu, INorm, Fully Connect and Sigmoid refer to convolution layer, leaky correction, Instance Norm, full-connected layer and Sigmoid active function, respectively. All of them use a convolution layer with a step size of 2, a filter size of  $4 \times 4$ , and a leaky correction linear unit with a slope of 0.2.

## 4 Experiments and discussions

In this section, we first report the databases and the implementation details in Section 4.1. Then, we make extensive qualitative and quantitative evaluations of the proposed MFC-Net and the representative non-uniform blind image deblurring works in Section 4.2. Next, we



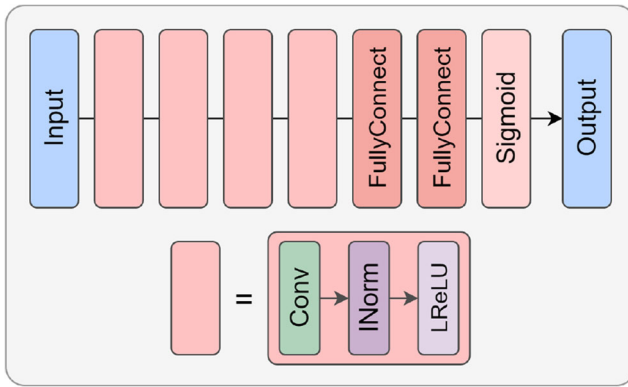


Fig. 8 Our discriminator D

conduct ablation studies to verify the effectiveness of the different proposed components in Section 4.3. Finally, we conduct model analysis in Section 4.4.

## 4.1 Datasets and implementation details

### 4.1.1 Datasets

The **GoPro** dataset [25] is the large realistic blurred image benchmark by averaging consecutive frames in a high-speed video. It contains 3,214 blurry/clear image pairs with  $1280 \times 720$  resolutions. We adopted 2,103 pairs of pixel-aligned blurry/clear images for training and the rest 1,111 pairs for evaluation in our work.

The **DVD** dataset [35] contains 71 videos, each with 3-5s average running time, which are split into 61 training videos and 10 testing videos. All of them are captured at 240fps with multiple devices, such as iPhone 6s, GoPro Hero 4 Black and Nexus 5x mobile phone for well generalization avoiding bias towards specific capturing device. In total, 71 videos are generated 6,708 synthetic blurry/clear pairs according average continuous short exposure frames [39], of which 5,708 blurry/clear pairs are used for training and the remaining 1,000 pairs were used for testing in our work.

Remote Sensing dataset [34]. The Remote Sensing dataset contains 800 remote sensing images. We adopt DeblurGAN [20] to generate random motion blur kernel to generate the blurry/clear pairs of the test dataset of the Remote Sensing dataset for direct test without training, aimed to evaluate the generalization ability of the proposed model in different scenarios.

In order to improve the generalization ability of image deblurring during training, we conducted training by combining the training sets of the GoPro dataset and the DVD dataset, and then tested and evaluated them respectively. In addition, we enhanced the datasets by applying data augmentation, including random flippings and rotations ( $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ$ ). We randomly cropped each

original image to a  $256 \times 256$  pixels as the input of the training in each iteration.

### 4.1.2 Implementation details

We trained our MFC-Net by utilizing PyTorch framework on a PC equipped with Ubuntu 16.04 system, an GTX 2080Ti GPU and Intel Xeon(R) Gold 6152 CPU. We conducted comprehensive evaluations on different network structures and models. In order to quantitatively evaluate the performance of our proposed model, we adopted the commonly used evaluation indicators, namely, Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), runtime (i.e., the average inference time on a single GPU) and model parameter. Note that, a higher SSIM value implies a deblurred image that is closer to the sharp image in terms of structural similarity, while a higher PSNR value indicates the similarity in terms of pixel-wise values. For all compared models, which have provided their source codes, to compare testing runtime of different models ran under different deep learning framework for fairness.

In this paper, we adopted the Adam optimizer with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  to train the proposed model. We set the batch size as 16, initial learning rate as  $10^{-4}$ , which degraded by multiplying by 0.5 after every 150K iterations, and the whole training process continued to 1000K iterations. To be fair, we reproduced these methods by conducting their official implementations with default settings and parameters.

The loss function  $L_G$  of our proposed MFC-Net consists of content loss [36]  $L_{content}$ , perception loss [21, 29]  $L_{perceptual}$  and WGAN loss [5]  $L_{WGAN}$ , while we adopted the loss function  $L_D$  of the discriminative network in WGAN-GP, we need to minimize the objective function as below:

$$L_G = \alpha \frac{1}{2CHW} \|X_{sharp} - X_g\|^2 + \beta \|VGG_{5-3}(X_{sharp}) - VGG_{5-3}(X_g)\| + \gamma [-E[D(X_g)]] \quad (6)$$

$$L_D = E[D(X_g)] - E[D(X_{sharp})] + \lambda E[(\nabla D(\hat{X}))_2 - 1]^2, \text{ where } \hat{X} = \varepsilon X_{sharp} + (1-\varepsilon)X_g \quad (7)$$

Where  $X_g$  and  $X_{sharp}$  represent the image generated by the generator and the real image respectively,  $C$ ,  $H$ , and  $W$  are dimensions,  $VGG_{5-3}$  represents the output of the active layer (ReLU) of the third convolutional layer of the fifth convolutional block in the pre-trained VGG-16 network [33].  $E$  denotes expectations, and  $D$  denotes discriminator,  $\nabla$  denotes the gradient,  $\hat{X}$  denotes random

**Table 1** Deblurring results on GoPro dataset

Method type	Method	PSNR/SSIM	Size	Runtime
Machine Learning	Kim et al. [17]	23.64/0.8239	-	3600
	Xu et al. [24]	25.10/0.8300	-	3800
	Sun et al. [37]	24.61/0.8129	54.1	1500
Deep Learning	Nah et al. [25]	28.49/0.8543	303	3.09
	DeblurGAN [20]	28.70/0.8419	37	0.05
	Tao et al. [38]	30.19/0.9024	80	0.72
	Zhou et al. [44]	30.55/0.9072	<b>18</b>	0.06
	DeblurGANv2 [21]	29.55/0.9010	238	0.35
	Zhang et al. [43]	30.25/0.9051	29	<b>0.03</b>
	Gao et al. [8]	<b>30.92/0.9153</b>	50	0.83
	Qi et al. [28]	28.81/0.8664	32	1.3
	Ye et al. [41]	30.22/0.9052	25	0.25
	Ours	<b>31.04/0.9160</b>	<b>16</b>	<b>0.01</b>

Size and Runtime are expressed in MB and second. The Best Performance is shown in Red and the Second-Best is in Blue. A “-” indicates that the result is not available

difference sampling on  $X_{sharp}$  and  $X_g$ . The trade-off parameters of  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\lambda$  are set to be 0.5, 0.005, 0.001 and 10, respectively.

## 4.2 Quantitative evaluations

### 4.2.1 Quantitative evaluation on GoPro dataset

We compare our proposed model with several state-of-the-art works: Kim et al. [17], Xu et al. [24], Sun et al. [37], Nah et al. [25], DeblurGAN [20], Tao et al. [38], Zhou et al. [44], DeblurGANv2 [21], Zhang et al. [43], Gao et al. [8], Qi et al. [28] and Ye et al. [41] on **GoPro** dataset. Unless otherwise noted, all images we evaluated in the experiment are RGB images.

As illustrated in Table 1, our proposed model achieves better performance compared with the methods listed in Table 1, including the machine learning based methods and the end-to-end deep learning based methods. Compared with the machine learning algorithms with Kim et al. [17], Xu et al. [24] and Sun et al. [37], our proposed method achieves the best result with improvements of 7.4dB/0.0921, 5.94dB/0.086, and 6.43dB/0.1031, respectively. Compared with the end-to-end deep learning methods, our MFC-Net

again achieves the best PSNR/SSIM results, which is improved by 4.11dB/0.0741, 1.49dB/0.0150, 0.49dB/0.0088 and 2.23dB/0.0496 compared with the single scale deblurring network DeblurGAN [20], DeblurGANv2 [21], Zhang et al. [43], Ye et al. [41] and Qi et al. [28], respectively. We can observe that, our MFC-Net also outperforms the the multi-scale images or stacked images based models, including Nah et al. [25], Tao et al. [38], Ye et al. [41] and Zhang et al. [43], with increases of 2.25dB/0.0617, 0.85dB/0.0136, 0.82dB/0.0108 and 0.79dB/0.0109, respectively. Although our proposed MFC-Net achieves comparable PSNR/SSIM result with Gao et al. [8] with only an increase of 0.12dB/0.0007, we have absolute advantages in terms of running speed and model size. To further verify the comparison results in Table 1 and compare performances, we obtain the performance for various test subsets chosen randomly in GoPro dataset, and perform statistical tests on the results. Specifically, we randomly split the GoPro test set into 5 subsets of the same size for testing with our proposed MFC-Net and methods of Gao et al. [8], Zhang et al. [43], Ye et al. [41] and DeblurGANv2 [21], respectively. The results are shown in Table 2. Moreover, we conduct T-test, as shown in Table 3. It is obviously to see that our proposed MFC-Net outperforms the methods including

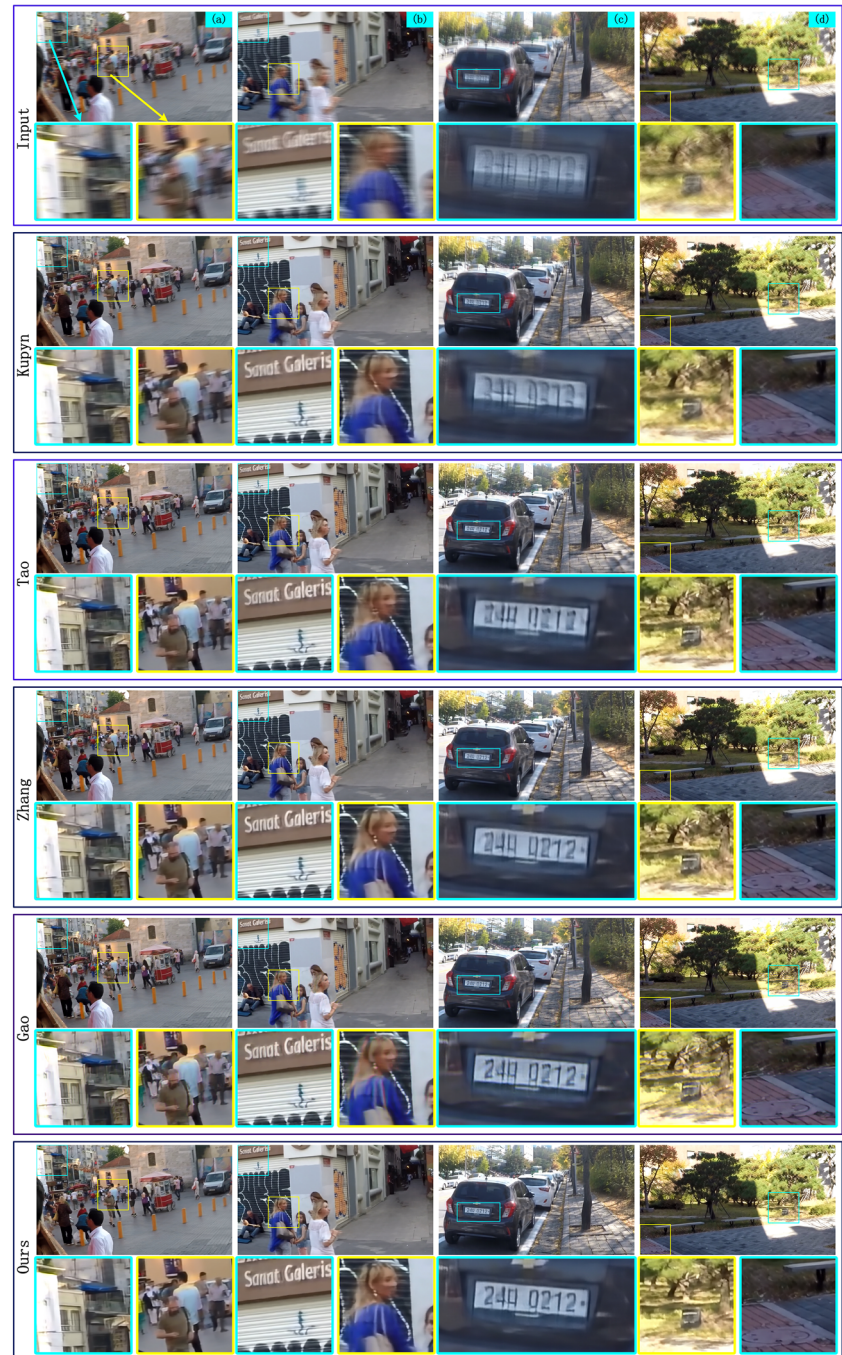
**Table 2** PSNR results of GoPro test set

Test dataset	Our	Gao et al. [8]	DeblurGANv2 [21]	Zhang et al. [43]	Ye et al. [41]
Test set subset 1	<b>32.54</b>	32.23	31.58	31.71	31.88
Test set subset 2	<b>32.16</b>	31.98	31.15	31.33	31.46
Test set subset 3	<b>29.68</b>	29.51	28.02	29.00	28.79
Test set subset 4	29.49	<b>29.58</b>	27.60	28.58	28.65
Test set subset 5	<b>31.32</b>	31.29	29.44	30.62	30.30

**Table 3** T-test results of Table 2

Pairing number	Method	Mean	Standard deviation	Mean difference	$t$	$p$
1	Our vs DeblurGANv2	31.04 vs 29.56	1.4 vs 1.79	1.48	7.171	0.002
2	Our vs Zhang et al.	31.04 vs 30.25	1.4 vs 1.4	0.79	18.172	0.000
3	Our vs Gao et al.	31.04 vs 30.92	1.4 vs 1.3	0.12	1.747	0.156
4	Our vs Ye et al.	31.04 vs 30.22	1.4 vs 1.48	0.82	12.588	0.000

**Fig. 9** Visual comparisons on testing dataset from the GoPro dataset. From top to down are the blurry input, deblurring results of Kupyn et al. [21], Tao et al. [38], Zhang et al. [43], Gao et al. [8] and our results





Zhang et al. [43], Ye et al. [41] and DeblurGANv2 [21] with  $p < 0.05$ , which indicates that there is significant differences. Our proposed MFC-Net achieves PSNR of 31.04, while Gao et al. [8] achieves PSNR of 30.92, in terms of the indicator of PSNR, the performance of the both is comparable ( $p > 0.05$ ). Although our model has a similar result on PSNR/SSIM with Gao's but achieves running time with 0.1s, which is more faster than Gao's with 83 times and 34 MB reduction of the model parameters. Compared to Zhang et al. [43], our model has a similar result on runningtime with 0.1s, but it has advantage in PSRN indicator ( $p < 0.05$ ), and it is more faster than Zhang et al with 3 times and 13MB reduction of the model parameters. On the whole, our proposed method makes trade-off in terms of performance-complexity, achieving the effect of real-time use. In general, most of the deep learning based methods have great advantages over the traditional machine learning based methods for image deblurring. And we find that the performance of the proposed MFC-Net is greatly improved compared with the same single-scale networks, including DeblurGAN [20], DeblurGANv2 [21] and Qi et al. [28], which verifies the effectiveness of MFC-Net. Zhou et al. [44] adopted the traditional U-net framework and removed the Batch Normalization layer, which greatly improved the performance (PSNR = 30.55). However, even if we adopted the same strategy by removing the Batch Normalization layer, the MFC-Net still achieves better performance. Differently from Zhou et al. [44], Gao et al. [8] utilized the cyclic parameter sharing network to learn blur information repeatedly to obtain good performance, but the cyclic scheme would caused serious time-consuming. While we carefully design the regional attention module, multi-scale fusion coding architecture and feature fusion module, without cyclic parameter, which still achieve better performance than the former and make balance between performance and speed.

As illustrated in Fig. 9, our MFC-Net also achieves the best visual restoration quality, and generates more explicit details and sharper structures in terms of restoration quality. The visual restoration results from Gao et al. still suffer from artifacts, as illustrated in Fig. 9a, artifacts appear near the person's hand, and also artifacts appear near the woman's glasses in Fig. 9b. Similar to the basic framework of Tao et al. [38], but differently, Gao et al. [39] adopted a nested skip connection structure for the nonlinear transformation modules to enhance the sharp details. However, it may result in magnifying the tiny blurry information in the blurry image and creating artifacts around it. Differs from Gao et.al, the proposed MFC-Net can effectively transfer the main features information avoid paying too much attention to the local details by exploiting the ResBlocks in Fig. 5. Simultaneously, the adversarial training, the adversarial loss and the perceived loss all

**Table 4** Deblurring results on DVD dataset

Method	PSNR/SSIM
DVD(single) [35]	25.44/0.8412
MoblieNet [21]	28.54/0.8654
Qi et al. [28]	28.72/0.8702
Tao et al. [38]	29.55/0.8822
Zhang et al. [43]	29.01/0.8732
Ye et al. [41]	<b>30.81/0.9045</b>
Ours	<b>30.92/0.9057</b>

The Best Performance is shown in Red and the Second-Best is in Blue

help to generate more natural images. Compared to the DeBlurGANv2, the proposed MFC-Net produces excellent deblur performance, which benefits from the proposed multi-scale fusion coding framework, the feature fusion module and the regional attention module.

#### 4.2.2 Quantitative evaluation on DVD dataset

Table 4 shows the comparison results on DVD dataset. Among all the results, our MFC-Net again achieves the best PSNR/MSSIM result, with improvements of 5.48dB/0.0645, 2.38dB/0.0103, 2.20dB/0.0355, 1.37dB/0.0235, 1.91dB/0.0325 and 0.11dB/0.0012 compared with DVD [35], MoblieNet [21], Qi et al. [28], Tao et al. [38], Zhang et al. [43] and Ye et al. [41], respectively. All these results furtherly demonstrates that our proposed model has better generalization in different scenarios. The visual examples is shown in Fig. 10. Obviously, the deblurring performance of the proposed MFC-Net is far more superior to the existing methods. In addition, the proposed MFC-Net is capable of restoring better local details, such as the car, boat pole and leaves in Fig. 10. Consequently, all the images shown for visual comparisons strongly demonstrate that the proposed light-weight MFC-Net can generatemore explicit details and sharper structures.

#### 4.2.3 Quantitative evaluation on remote sensing dataset

The comparison results on the Remote Sensing dataset are shown in Table 5, which further verify the generalization ability of our model. We can observe that our MFC-Net again achieves the best PSNR/SSIM result, with improvements of 2.66dB/0.0523, 0.57dB/0.0102 and 0.42dB/0.0085 compared DeblurGAN [20], Gao et al. [8] and Zhang et al. [43], respectively. DeblurGAN [20] adopts a single-scale scheme, and Zhang et al. [43] and Gao et al. [8] employed multistage parameter sharing schemes respectively in building their deep networks. Visual comparison is shown in Fig. 11. Obviously, our model generally produces



**Fig. 10** Visual comparisons on DVD dataset. From top to bottom are the blurry input, deblurring results of Zhang et al. [43], Ye et al. [41], and our results



better results than those of [20, 43] and [8]. For example, the clarity of houses and edges is restored better, the deblurring effect of grass is more natural, as can be seen from Fig. 11a, b and c, respectively.

### 4.3 Ablation study and analysis

Finally, we conduct ablation experiments on GoPro dataset to verify the effectiveness of different components in our

proposed MFC-Net, including baseline, RA module and FF module. It is note that, for all experiments, we also train these networks using the identical training strategy and parameters as in Sections 4.1 and 4.2.

**Validation on Basic Network Architecture** We set our multi-scale fusion coding network(i.e.,removed the FF ans RA models) as the baseline, and make comparison with its incomplete versions,i.e., the single-scale network with only

**Table 5** Deblurring results on remote sensing dataset

Method	PSNR/SSIM
blurred images	19.85/0.6042
Kupyn et al. [20]	20.64/0.6285
Gao et al. [8]	21.94/0.6463
Zhang et al. [43]	<b>22.09/0.6480</b>
Ours	<b>22.51/0.6565</b>

The Best Performance is shown in Red and the Second-Best is in Blue

the original input, as reported in Table 6. From the results, our baseline model exhibits significant superiority over its single-scale network, surpassing by 0.63dB/0.0048 in terms of PSNR/MSSIM indicators. It is evident that the

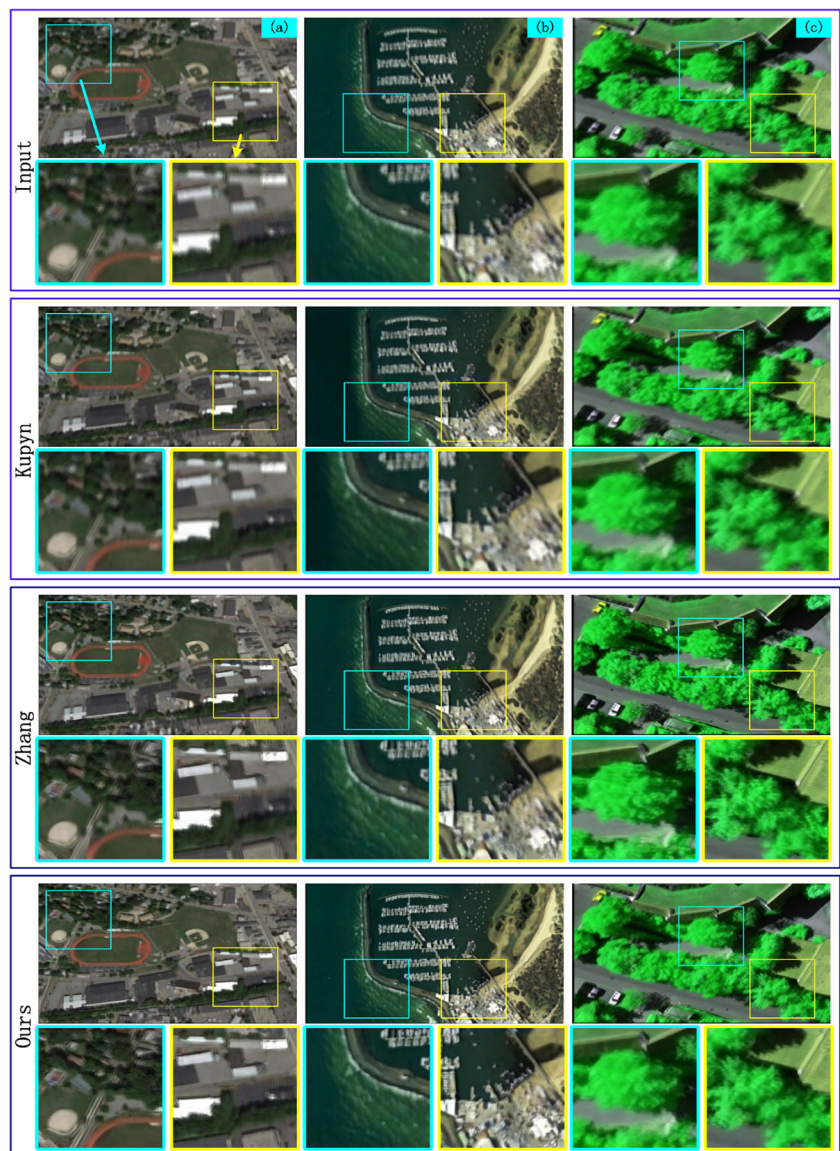
**Table 6** Evaluation of the basic network architecture on GoPro dataset

Method	PSNR/SSIM
Single-scale network	29.54/0.8952
Baseline	<b>31.17/0.9010</b>

multi-scale fusion coding framework boosts the deblurring performance significantly.

**Validation on FF and RA Models** For further testing and verifying the effectiveness of our FF model and RA model, each of them was added gradually to the baseline network. As shown in Table 7, compared with the baseline, an increase of 0.23dB/0.0030 is achieved with the addition of the FF module, almost without any extra computation.

**Fig. 11** Visual comparisons on Remote Sensing dataset. From top to bottom are the blurry input, deblurring results of Kupyn et al. [20], Zhang et al. [43], and our results





**Table 7** Ablation analysis of FF and RA models on GoPro dataset

Method	PSNR/SSIM
Baseline	30.17/0.9010
Baseline+FF model	30.40/0.9040
Baseline+FF+RA model	<b>31.04/0.9160</b>

Additionally, it only needs 0.008s for image deblurring with  $1280 \times 720$  pixels in testing stage, which indicates that our FF model is beneficial to real-time deblurring while maintaining good performance. Afterwards, we add the RA model based on the FF model, the PSNR/MSSIM again improves with 0.64dB/0.0120, which demonstrates that our proposed FF model is effective for deblurring. All these results demonstrate that our proposed FF and RA model is prone to enhance the deblurring performance.

**Validation on Regional Feature Weighting Strategy** We further verify the effectiveness on focusing the local blurry feature of the regional feature weighting strategy in RA module. It is noted that, in training stage, we trained images cropped randomly with size of  $256 \times 256$ . After twice subsampling by the network encoder, we obtained feature maps with size of  $64 \times 64$ , and then sent them into the RA module. In test stage, the resolution of the test images is  $1280 \times 720$ , so we need to divide each test image into sub-regions firstly, and then send each sub-region into the RA module one by one. We adopt six methods for regional division, i.e.,  $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ ,  $5 \times 3$ ,  $4 \times 4$  and  $5 \times 5$ . As shown in Table 8, the PSNR/SSIM result improves along with the increasing sub-regions. Especially the PSNR/SSIM achieves the best result with the size of  $64 \times 60$  pixels when adopt  $5 \times 3$  regional division method, which is close to the training size of  $64 \times 64$  pixels. After that, as the number of sub-regions increase, that is, the smaller the size of the feature map sent into RA module, the performance of PSNR/SSIM gradually decreases. These results show that the weight value obtained by the RA module of different regional sizes can effectively focus on the local blurring

**Table 8** Evaluation of the proposed regional feature weighting strategy on GoPro dataset

Regionalization	Size of input module	PSNR/SSIM
$1 \times 1$	$320 \times 180$	30.79/0.9119
$2 \times 2$	$160 \times 90$	30.85/0.9127
$3 \times 3$	$106 \times 60$	30.95/0.9142
$5 \times 3$	$64 \times 60$	<b>31.04/0.9160</b>
$4 \times 4$	$80 \times 45$	31.00/0.9148
$5 \times 5$	$64 \times 36$	30.73/0.9115

features by weighting the regional features, which helps to improve the deblurring performance.

#### 4.4 Model analysis

From the above quantitative and qualitative results, we can clearly observe the effectiveness of MFC-Net. Here we make a deeper discussion on MFC-Net.

**Performance** As illustrated in Tables 1, 4 and 5, our MFC-Net is superior to other competitive methods without exploiting complex computation and multi-stage network structure. Moreover, as illustrated in Table 5 and Fig. 11, we show the advantages of our model when evaluated in different scenarios. It is obviously to see more clearer details in different scenarios produced by our MFC-Net compared with other methods. Combined with the analysis of the experimental results, we think that the performance obtained by proposed MFC-Net is mainly attributed to three points: 1) removing the Batch Normalization layer. Intuitively, the essential task for image deblurring is to learn the direct absolute difference between the input and the output. Batch normalization layer changes the data distribution, which is not conducive low-level task learning. 2) proposing a multi-scale fusion coding framework. It is difficult for the single-scale network to learn blurry image information, while the multi-scale network is able to pay more attention to learn the blur information. Consequently, we adopt the multi-scale coding fusion strategy to make balance for the speed and performance. 3) proposing a regional attention module. The stacked residual blocks can not effectively learn and reconstruct the sharp image information. So, we exploit the fusion of blur representation and atrous spatial pyramid pooling to make the network learn the image blur information independently, so as to achieve better deblurring performance.

**Runtime** In addition to the excellent generalization performance and lightweight characteristic, it only takes 10 milliseconds to process a  $720 \times 1280$  image by MFC-Net. Compared with other models [8, 38, 43], the following factors may be contributed to the rapid operation of our MFC-Net: 1) multi-scale fusion coding network helps to avoid the unnecessary computation cost at the multi-stage; 2) adopting small-size convolutional filters (kernel size equals 3); 3) adopting simple residual structure to reduce the unnecessary complex calculation process.

**Deblurring method** Different from the existing similar Unet [30] networks and multi-scale networks [8, 38, 43], our proposed MFC-Net adopts simple multi-scale fusion coding method. As shown in Table 1, the mainstream multi-scale patch network [8, 43] and multi-scale recursive

network [38] achieve excellent deblurring performance at the cost of time due to the multi-scale information of the input image. Differently, we fuse the multi-scale information in a single-scale network to achieve excellent deblurring performance under the premise of ensuring fast speed. Ablation experiments can be proved that our proposed RAM can independently learn the blurry features of different regions, which helps to enhance the interpretability of model deblurring and improve the generalization ability of the model. Although our MFC-Net has a much better overall performance, we believe there is still room for it to make improvements. A possible solution is to further deeply mining the blurry information of blurry image, such as edge information, spatial transformation, etc, which will be explored in our future work. Another possible solution is to design network that is able to disentangle the blur information, and take the fusion results of the blur information with the shrap image as the training data to enhance effectively training.

## 5 Conclusions

In this paper, we propose a novel powerful MFC-Net for image deblurring with kernel-free in an end to end manner. The proposed framework can obtain enhanced blurry feature representation essentially by encoding the multi-scale information of the blurring image, aiming to improve deblurring ability of the network. Furthermore, we propose regional attention model by adopting regional representation weighting to focus on the local blurry information so as to induce the network to learn better. We also design a special feature fusion model to mining correlations between multi-scale features. Additionally, for making the deblurred image as close as possible to the real one, we introduce context loss and perceptual loss to enhance the model's recovery at local texture, color, edge and other details, effectively, avoiding ringing artifacts while preserving details. Extensive experiments demonstrate that our proposed network outperforms those existing methods with fewer parameters and high restoration quality.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China under Grant 61762014 and 62106054, and in part by the Science and Technology Project of Guangxi under Grant 2018GXNSFAA281351.

## References

- Arjovsky M, Bottou L (2017) Towards principled methods for training generative adversarial networks. *Stat*:1050
- Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan
- Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. *Computer Science*
- Bai Y, Jia H, Jiang M, Liu X, Xie X, Gao W (2020) Single-image blind deblurring using multi-scale latent structure prior. *IEEE Transactions on Circuits and Systems for Video Technology*
- Chen S, Shi D, Sadiq M, Cheng X (2020) Image denoising with generative adversarial networks and its application to cell image enhancement. *IEEE Access* 8(99):82819–82831
- Cho S, Wang J, Lee S (2011) Handling outliers in non-blind image deconvolution. In: 2011 International Conference on Computer Vision
- Cho S, Lee S (2009) Fast motion deblurring. *Acm Trans Graph* 28(5):1–8
- Gao H, Tao X, Shen X, Jia J (2019) Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: 2019 IEEE/CVF Conference on computer vision and pattern recognition (CVPR)
- Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial networks. *Adv Neural Inf Process Syst* 3:2672–2680
- Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A (2017) Improved training of wasserstein gans
- He K, Zhang X, Ren S, Sun J (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 37(9):1904–16
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE Conference on computer vision & pattern recognition
- Hong M, Choe Y (2018) Wgan based edge preserved deblurring using perceptual style similarity. In: 2018 International symposium on computer, consumer and control (IS3c)
- Jiang K, Wang Z, Yi P, Chen C, Huang B, Luo Y, Ma J, Jiang J (2020) Multi-scale progressive fusion network for single image deraining. In: 2020 IEEE/CVF Conference on computer vision and pattern recognition (CVPR)
- Jie H, Li S, Gang S, Albanie S (2017) Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* PP(99)
- Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision
- Kim TH, Ahn B, Lee KM (2013) Dynamic scene deblurring. In: IEEE International conference on computer vision
- Krishnan D, Tay T, Fergus R (2011) Blind deconvolution using a normalized sparsity measure. In: Computer vision & pattern recognition
- Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural networks. In: NIPS
- Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) Deblurgan: Blind motion deblurring using conditional adversarial networks. In: 2018 IEEE/CVF Conference on computer vision and pattern recognition (CVPR)
- Kupyn O, Martyniuk T, Wu J, Wang Z (2019) Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. 2019 IEEE/CVF International Conference on Computer Vision (ICCV)
- Lai WS, Ding JJ, Lin YY, Chuang YY (2015) Blur kernel estimation using normalized color-line priors. In: Computer vision & pattern recognition
- Li X, Zheng S, Jia J (2010) Two-phase kernel estimation for robust motion deblurring. In: Lecture notes in computer science(including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)



24. Li X, Zheng S, Jia J (2013) Unnatural 10 sparse representation for natural image deblurring. In: IEEE Conference on computer vision & pattern recognition
25. Nah S, Kim TH, Lee KM (2016) Deep multi-scale convolutional neural network for dynamic scene deblurring. IEEE Computer Society
26. Pan J, Sun D, Pfister H, Yang MH (2016) Blind image deblurring using dark channel prior. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR)
27. Purohit K, Rajagopalan AN (2019) Region-adaptive dense network for efficient motion deblurring
28. Qi Q, Guo J, Jin W (2020) Attention network for non-uniform deblurring. IEEE Access PP(99):1–1
29. Qu X, Wang X, Wang Z, Wang L, Zhang L (2018) Perceptual-dualgan: Perceptual losses for image to image translation with generative adversarial nets. pp 1–8
30. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. Springer, Cham
31. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D Nonlinear Phenom* 60(1-4):259–268
32. Schuler CJ, Hirsch M, Harmeling S, Schlkopf B (2016) Learning to deblur. *IEEE Trans Pattern Anal Mach Intell* 38(7):1439–1451
33. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *Computer Science*
34. Su H, Wei S, Yan M, Wang C, Zhang X (2019) Object detection and instance segmentation in remote sensing imagery based on precise mask r-cnn. In: IGARSS 2019 - 2019 IEEE International geoscience and remote sensing symposium
35. Su S, Delbracio M, Wang J, Sapiro G, Wang O (2017) Deep video deblurring for hand-held cameras. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR)
36. Suin M, Purohit K, Rajagopalan AN (2020) Spatially-attentive patch-hierarchical network for adaptive motion deblurring
37. Sun J, Cao W, Xu Z, Ponce J (2015) Learning a convolutional neural network for non-uniform motion blur removal. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR)
38. Tao X, Gao H, Wang Y, Shen X, Wang J, Jia J (2018) Scale-recurrent network for deep image deblurring. In: 2018 IEEE/CVF Conference on computer vision and pattern recognition
39. Telleen J, Sullivan A, Yee J, Wang O, Davis J (2010) Synthetic shutter speed imaging. In: Wiley, pp. 591–598
40. Woo S, Park J, Lee JY, Kweon IS (2018) Cbam: Convolutional block attention module. Springer, Cham
41. Ye M, Dong L, Chen G (2020) Scale-iterative upscaling network for image deblurring. *IEEE Access* PP(99):1–1
42. Yuan L, Sun J, Quan L, Shum HY (2008) Progressive inter-scale and intra-scale non-blind image deconvolution. *Acm Transa Graph* 27(3):1–10
43. Zhang H, Dai Y, Li H, Koniusz P (2019) Deep stacked hierarchical multi-patch network for image deblurring. *IEEE*
44. Zhou S, Zhang J, Zuo W, Xie H, Ren JS (2019) Davanet: Stereo deblurring with view aggregation. In: 2019 IEEE/CVF Conference on computer vision and pattern recognition (CVPR)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Haiying Xia** received the M.S. and Ph.D. degrees from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2007 and 2011, respectively. She is currently a Full Professor with Guangxi Normal University. Her current research interests include pattern recognition, medical image analysis, and neural networks.



**Bo Wu** is currently pursuing the M.S. degree with the College of Electronic Engineering, Guangxi Normal University. His research interests include deep learning and image deblurring.



**Yumei Tan** is currently pursuing the Ph.D. degree with School of Computer Science and Engineering, Guangxi Normal University, Guilin, China. She received the M.S. degree from the College of Electronic Engineering, Guangxi Normal University, in 2013. Her research interests include deep learning and facial emotion recognition.



**Xiaohu Tang** received the B.S. degree in electronics and information engineering, and the M.S. degree in electronic circuit and system from Guangxi Normal University, Guilin, P.R. China, in 2009, 2012 respectively. He is an lecturer in Guangxi Normal University, Guilin, China. His research interests include light field sampling, and feature extraction and identification of moving target contours in image signal processing.



**Shuxiang Song** received the Ph.D. degree from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China. He is currently a Full Professor with Guangxi Normal University. His current research interests include intelligent detection, automatic control, and signal and image processing.