

## Research Article

# An Image Deblurring Method Using Improved U-Net Model

Zuozheng Lian , Haizhen Wang , and Qianjun Zhang

*College of Computer and Control Engineering, Qiqihar University, Qiqihar, China*

Correspondence should be addressed to Haizhen Wang; wanghaizhen1976@163.com

Received 20 January 2022; Revised 17 June 2022; Accepted 7 July 2022; Published 31 July 2022

Academic Editor: Hye-Jin Kim

Copyright © 2022 Zuozheng Lian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deblurring methods in dynamic scenes are a challenging problem. Recently, significant progress has been made for image deblurring methods based on deep learning. However, these methods usually stack ordinary convolutional layers or increase convolution kernel size, resulting in limited receptive fields, an unsatisfying deblurring effect, and a heavy computational burden. Therefore, we propose an improved U-Net (U-shaped Convolutional Neural Network) model to restore the blurred images. We first design the model structure, which mainly includes depth-wise separable convolution, residual depth-wise separable convolution, wavelet transform, inverse wavelet transform, and a DMRFC (dense multireceptive field channel) module. Next, a depth-wise separable convolution is designed, which reduces model calculations and the number of parameters when compared with the standard convolution. A residual depth-wise separable convolution is designed, which allows for propagation of detailed information from different layers when compared with standard convolution and a standard residual block. The wavelet transform realizes downsampling by separating the contextual and texture information of the image. It also reduces model training difficulty. The inverse wavelet transform realizes upsampling, which reduces the loss of image information. Finally, by combining an extensional receptive field and channel attention mechanism, a DMRFC module is proposed to extract detailed image information, which further improves the reconstructed image quality via inverse wavelet transform. Experiments on the public dataset GOPRO show that the image deblurring method in this paper has higher-quality visual effects, while the PSNR (peak signal-to-noise ratio) rises to 30.83 and SSIM (structural similarity) rises to 0.948. Fewer model parameters and a shorter recovery time are needed, which provides a more lightweight image deblurring method.

## 1. Introduction

Blurred images are caused by complex reasons in dynamic scenes, such as camera shake, atmospheric turbulence, low light levels, and target motion transformation [1]. The blurred images not only influence the visual experience subjectively, but also influence the following visual tasks such as object detection and object tracking [2–5]. Therefore, image deblurring in dynamic scenes is a challenging problem in computer vision [6].

Most traditional image deblurring approaches use regularization and prior hand-crafted images to assess the blur kernel [7] and the next iterative optimization to increasingly recover the image. The conventional method needs to assess complex blur kernels, which results in a lengthy deblurring process as well as poor real-time and algorithm performance. Recently, image deblurring approaches using deep learning

have obtained a growing application [8]. Li et al. [9] advanced a brand-new convolution architecture, which greatly expanded the reception field. A scale-aware convolutional neural network to restore a clear image was also proposed. Quan et al. [10] advanced a deep-learning-based method to nonblind image deblurring, which combined Gabor-domain and complex-valued CNN-based prior to better handle the noise with unknown parameters or statistical characteristics. Liu et al. [11] advanced a deblurring module to sharpen blur images of dynamic scenes based of high-frequency residual image learning. Zhang et al. [12] advanced an enhanced adversarial network model, which used the weight of feature channels to generate clear images and remove draught board artefacts. Zhang et al. [13] advanced a deep hierarchical multipatch network, which can handle blurred images using a fine-to-coarse hierarchical representation. It can adjust the performance and runtime of a single network to fit different

application scenarios. Chen et al. [14] advanced an attention-adaptive module, which adaptively decided the arrangement of channel and spatial attention modules. A transformable convolutional module was also proposed to process the geometric variations. Zeng and Diao [15] applied a dense network to remove image blur, which may prevent the gradient disappearance problem. By analysing the above deep learning methods, the receptive field of the network is insufficient. Furthermore, the amount of image information obtained is limited while model calculations are extensive and time consuming. This paper proposes an improved method using U-Net to solve these problems.

The overall processing flow of U-Net uses an encoder and decoder, which was proposed by Hinton as early as 2006. At that time, its main function was image compression and denoising. This method of image compression relies on the input of an image and encoding of downsampling to obtain a string of features partially describing the original image. This is equivalent to compression, followed by upsampling for decoding to obtain the original image. Therefore, storing an image needs a feature and a decoder to be stored. This idea can also be used in denoising of the original image. How the method works is to artificially add noise to the original image in the training stage. It is then put into the encoder to train the network based on the difference between the results and the original image followed by the restoration of the original image. Later, this idea was used in image segmentation, so U-Net has been widely used in image segmentation and related fields in recent years [16, 17]. U-Net can integrate different levels of features and has a flexible network structure. As such, a deep network with many parameters can dramatically reduce the number of parameters that lie within the acceptable accuracy range. In recent years, U-Net variants have also been used in other image processing fields. Zhang et al. [18] incorporated domain-specific knowledge to design an attention-based Tri-U-Net architecture which included feature extraction, feature fusion, and image reconstruction; these generated high-quality and high-resolution multispectral images. Fan et al. [19] advanced a SUNET model, which combined Swin Transformer layer and U-Net to denoise the image. Zang et al. [20] proposed a novel cascaded Dense-U-Net structure to fully utilize all hierarchical features for super-resolution single images. A Dense-U-Net block included many short and dense skip pathways, which could profit the flow of information and integrate different receptive fields. Guan et al. [21] advanced an improved convolutional neural network architecture called Fully Dense U-Net to remove artifacts from 2D photoacoustic tomography images reconstructed from sparse data. Then, the reconstructed image quality of Dense U-Net and standard U-Net is compared. Alimjan et al. [22] proposed a remote image change detection algorithm based on U-Net network, which added a multifeature self-attention mechanism between the encoder and decoder to achieve more plentiful context dependence.

Inspired by the above literature, we propose an image deblurring method based on U-Net. The key to U-Net is downsampling, upsampling, and skip connections, whose implementation methods are very important to the

performance of U-Net. Wavelet transform can realize the downsampling operation of CNN, while inverse wavelet transform (IWT) can realize the upsampling operation [23]. Wavelet transform may represent contextual dependencies and texture information of the image at different levels. High-frequency sub-bands profit to restore image texture details and constrain reconstruction of low-frequency sub-bands. Consequently, wavelet transform may be used for image processing goals such as super-resolution [24], image reconstruction [25], image defogging [26], and image deblurring [27]. In these processes, the combination of wavelet transforms and deep convolutional neural network is realized to eliminate any image blur influenced by wavelet transform, further removing blur from a noisy image. Considering that U-Net adopts a full CNN structure, we therefore propose to use wavelet transform to realize downsampling and IWT to realize upsampling. However, when the number of conventional convolution parameters is large, this requires a large number of calculations. Furthermore, deepening of network layers causes the gradient to disappear. The depth-wise separable convolution [28] and residual network [29] can solve these problems [30]. Also, the skip connection in U-Net directly connects the features extracted from the encoder to the corresponding decoder layer, which produces a lot of redundant information. The attention mechanism can extract key information [18, 31], which solves this problem. Inspired by the above methods, we aim to increase the receptive field and reduce calculation by combining wavelet transform, depth-wise separable convolution, residual network, and attention mechanism to improve U-Net. As such, we propose a comprehensive image deblurring model and achieve better deblurring performance.

For this paper, an improved U-Net model is proposed to remove image deblur, with the main contributions as follows: (1) We embed a multilevel two-dimensional discrete wavelet transform and IWT into U-Net. The wavelet transform realizes downsampling, which obtains contextual and texture information at different image frequencies and reduces model training difficulty. The IWT realizes upsampling, which reconstructs the processed information to achieve a more accurate image. (2) We propose the DMRFC module to extract semantic features of the image and improve image deblurring performance accordingly. The module consists of four multireceptive field channel blocks and a bottleneck layer. A multireceptive field channel block is composed of an extensional receptive block and a channel attention module. This increases the diversity of extracted features and features are adaptively weighted at each channel according to their importance, while the bottleneck layer reduces the number of feature inputs. (3) We embed a residual depth-wise separable convolution module instead of a standard residual block. This is for propagation of detailed information from different layers to improve blur reconstruction quality. The proposed model is then analysed quantitatively and qualitatively using the GOPRO dataset. The experimental results show that it improves the quality of image restoration. It also reduces information reuse, increases the image receptive field, and results in a good visual effect.

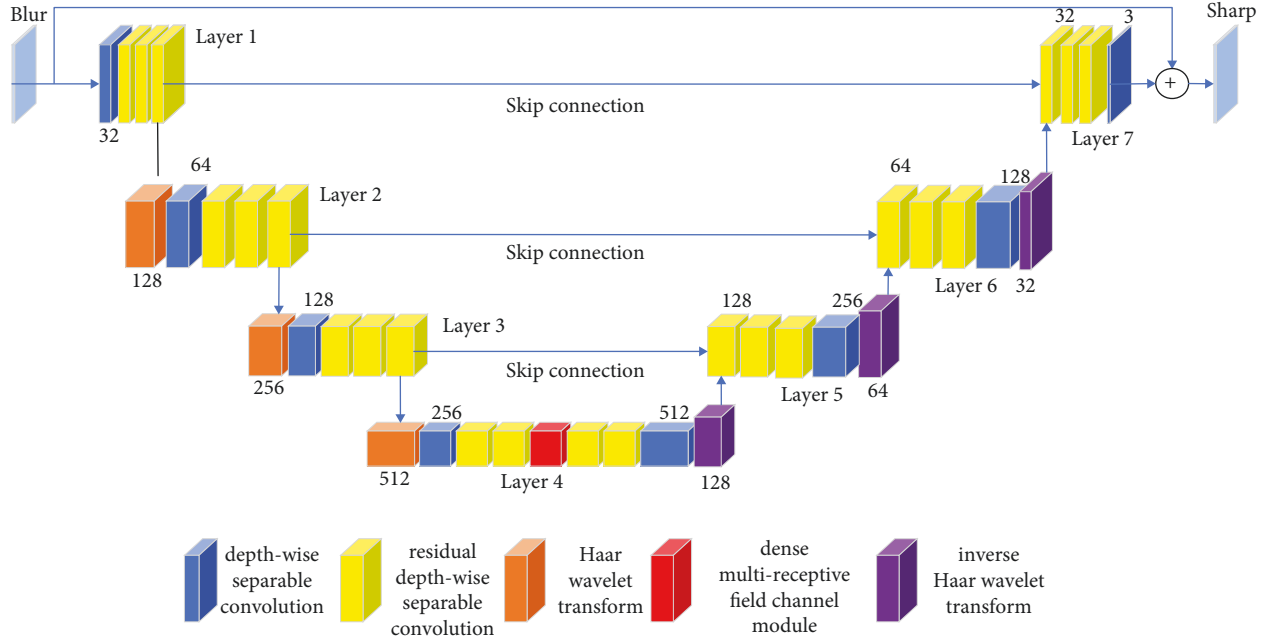


FIGURE 1: The model structure.

## 2. The Proposed Methodology

**2.1. The Model Structure.** We propose an improved U-Net model with the model structure shown in Figure 1.

As shown in Figure 1, the model mainly includes depth-wise separable convolution, residual depth-wise separable convolution, wavelet transform, inverse wavelet transform, and a DMRFC module. The model structure presents a U shape, with the encoder on the left and the decoder on the right. The convolution kernel is  $3 \times 3$  and the activation function in the model is Leaky ReLU. The encoder realizes the extraction of image information through four stages. In the first stage, one 32-channel depth-wise separable convolution and three 32-channel depth-wise separable residual convolutions are used. In the second stage, wavelet transform, one 64-channel depth-wise separable convolution, and three 64-channel residual depth-wise separable convolutions are used. In the third stage, wavelet transform, one 128-channel depth separable convolution, and three 128-channel residual depth-wise separable convolutions are used. The fourth stage uses wavelet transform, one 256-channel depth-wise separable convolution, and two 256-channel residual depth-wise separable convolutions as well as a dense multi-receptive field channel module. The decoder realizes image reconstruction through information fusion, which consists of four stages. The first stage uses two 256-channel residual depth-wise separable convolutions and one 512-channel depth-wise separable convolution as well as IWT. The output of the second stage and the third stage of the encoder are feature-fused through skip connections, three 128-channel residual depth-wise separable convolutions, one 256-channel depth-wise separable convolution, and an IWT. The outputs of the third stage and the second segment of the encoder stage are fused through skip connections, three 64-channel residual depth-wise separable convolutions, one

128-channel depth-wise separable convolution, and an IWT. The output of the fourth stage and the first segment of the encoder stage are feature-fused through skip connections. Three 32-channel residual depth-wise separable convolutions and one 3-channel depth-wise separable convolution restore the feature map to its original resolution. Finally, the restored image is obtained by fusing this with the input data.

Therefore, the encoder uses depth-wise separable convolution to implement the standard convolution operation and uses residual depth-wise separable convolution instead of the standard residual block. This reduces the number of parameters in the network. Two-dimensional DWT (discrete wavelet transform) implements downsampling to obtain contextual and texture information at different image frequencies. This reduces the computational complexity and training difficulty. The DMRFC module obtains image information at different scales and reduces gradient disappearance, feature reuse. The decoder implements upsampling process by IWT to reduce image information loss. The image is then reconstructed effectively using depth-wise separable convolution and residual depth-wise separable convolution. This helps to propagate detailed information from different layers and improves the deblurring effect. The main modules are introduced in detail below.

**2.2. Depth-Wise Separable Convolution.** The depth-wise separable convolution improves conventional convolution, reduces the number of parameters, and makes the network lightweight. The depth-wise separable convolution structure in this paper is shown in Figure 2. It consists of depth-wise convolution and point-wise convolution. Depth-wise convolution first divides the multichannel features of the previous layer into the feature map of a single channel. It then uses a  $3 \times 3$  convolution kernel for convolution. This is then

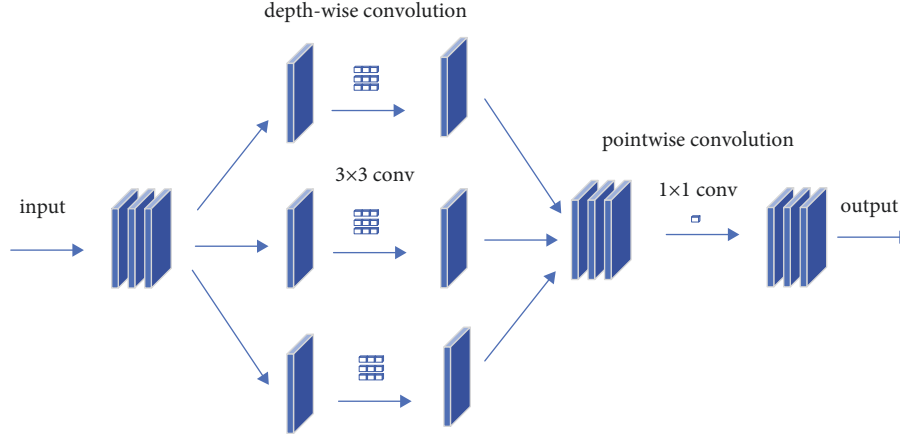


FIGURE 2: The depth-wise separable convolution structure.

TABLE 1: The related parameters of the depth-wise convolution in the model.

Layer	In_channels	Out_channels	Kernel_size	Stride	Padding	Activation function
Layer 1	4	32	$3 \times 3$	1	1	LeakyReLU
Layer 2	128	64	$3 \times 3$	1	1	LeakyReLU
Layer 3	256	128	$3 \times 3$	1	1	LeakyReLU
Layer 4	512	256	$3 \times 3$	1	1	LeakyReLU
Layer 5	128	256	$3 \times 3$	1	1	LeakyReLU
Layer 6	64	128	$3 \times 3$	1	1	LeakyReLU
Layer 7	32	4	$3 \times 3$	1	1	LeakyReLU

followed by re-stacking them together, that is, the size of the feature map of the previous layer is adjusted, while the number of channels remains unchanged. The characteristic image obtained by depth-wise convolution is convoluted using point-wise convolution for the second time using a  $1 \times 1$  convolution kernel followed by fusion of the convolution results for depth-wise convolution and freely changing the number of output channels. The related parameters for depth-wise convolution (Figure 2) in the model are shown in Table 1.

**2.3. Residual Depth-Wise Separable Convolution.** When the number of network layers increases, the gradient will gradually disappear in the back propagation process, resulting in an inability to effectively adjust the weight of the previous network layer. The residual network structure solves the problem of the gradient disappearance. The residual depth-wise separable convolution structure in this paper that is inspired by the residual network structure is shown in Figure 3.

The residual depth-wise separable convolution is based on the residual convolution using two depth-wise separable convolutions and addition of skip connections. The input  $x$  is directly transmitted to the output as the initial result. The training goal is to approximate the residual result to 0 so that the accuracy does not decrease as the network deepens. The related parameters of the residual depth-wise separable convolution in the model are shown in Table 2.

**2.4. Two-Dimensional DWT and IWT.** The Haar wavelet is a wavelet basis function which is easy to implement and operate. Therefore, this paper uses a two-dimensional Haar

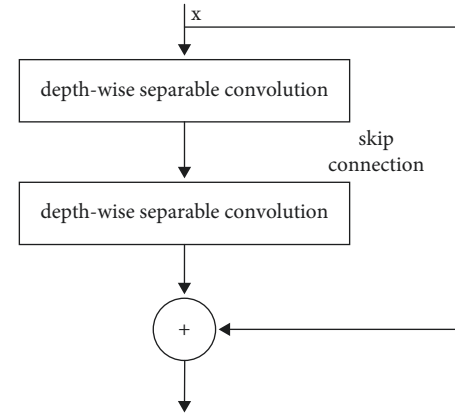


FIGURE 3: Residual depth-wise separable convolution structure.

wavelet for wavelet transform operations to divide the image signal into directional sub-bands in order to achieve different frequency image information. Filter is an effective method to realize DWT. Let the one-dimensional high-pass filter be represented by  $\varphi(\cdot)$  and the low-pass filter be represented by  $\psi(\cdot)$ . The detailed transformation process is as follows: first,  $\varphi(\cdot)$  is used to filter and vertically down-sample each column of the image.  $\varphi(\cdot)$  and  $\psi(\cdot)$  are then used to filter and horizontally downsample each row. The sub-frequency information for  $I_{HH}$  and  $I_{HL}$  is obtained. Second,  $\psi(\cdot)$  is used to filter and vertically downsample each column of the image.  $\varphi(\cdot)$  and  $\psi(\cdot)$  are then used to filter and horizontally downsample each column. The sub-frequency information of  $I_{LL}$  and  $I_{LH}$  is obtained. Sub-frequency information for the four parameters is shown as follows:

TABLE 2: The related parameters of the residual depth-wise separable convolution in the model.

Layer	In_channels	Out_channels	Kernel_size	Stride	Padding	Activation function
Layer 1, layer 7	32	32	$3 \times 3$	1	1	LeakyReLU
Layer 2, layer 6	64	64	$3 \times 3$	1	1	LeakyReLU
Layer 3, layer 5	128	128	$3 \times 3$	1	1	LeakyReLU
Layer 4	256	256	$3 \times 3$	1	1	LeakyReLU

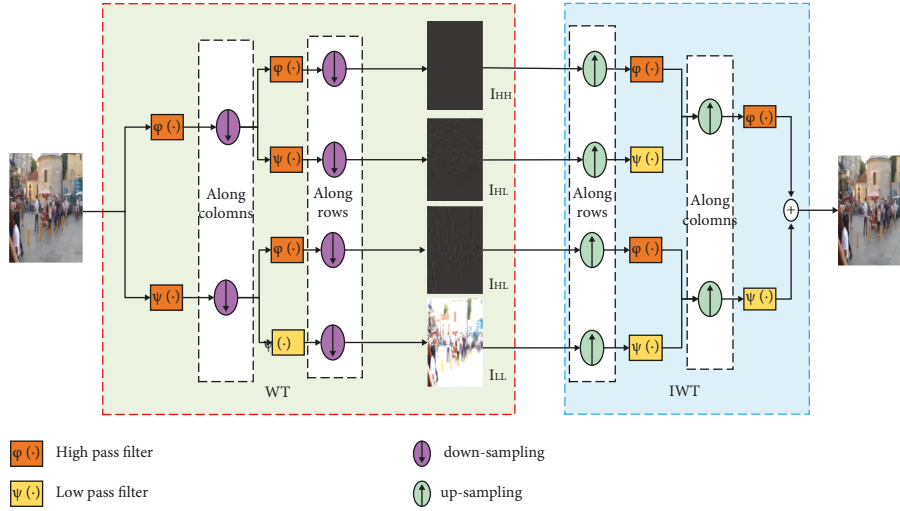


FIGURE 4: Two-dimensional discrete Haar wavelet transform and inverse transform.

$$I_{HL}(x, y) = \varphi(x)\psi(y), \quad (1)$$

$$I_{HH}(x, y) = \varphi(x)\varphi(y), \quad (2)$$

$$I_{LH}(x, y) = \psi(x)\varphi(y), \quad (3)$$

$$I_{LL}(x, y) = \psi(x)\psi(y). \quad (4)$$

The  $x$  and  $y$  in equations (1)–(4) denote rows and columns for the information of the image.  $I_{HL}$  expresses the horizontal high-frequency and vertical low-frequency information of the image.  $I_{HH}$  expresses the horizontal and vertical high-frequency information of the image.  $I_{LL}$  expresses the horizontal and vertical low-frequency information of the image.  $I_{LH}$  denotes the horizontal low-frequency and vertical high-frequency image information. In contrast, the IWT performs inverse operations on the four sub-images using the above filter. Thus,  $I_{HL}$ ,  $I_{HH}$ ,  $I_{LL}$ , and  $I_{LH}$  are used to fuse into the original image. Consequently, as shown in Figure 4, the original image is decomposed by DWT and then reconstructed by IWT without loss of information. And so on, multilevel wavelet transform can be implemented by further processing of  $I_{HL}$ ,  $I_{HH}$ ,  $I_{LL}$ , and  $I_{LH}$  according to the above method. For the two-dimensional Haar wavelet transform, the sum mean value is used for low-frequency information, regarded as  $\psi(\cdot)$ , and the difference mean value is used for high-frequency information, regarded as  $\varphi(\cdot)$ . We take a  $4 \times 4$  two-dimensional matrix  $A$  (Figure 5) as an example.

$A$  is divided into 4 sub-matrices. The sub-matrix composed of odd rows and odd columns is represented by  $x1$ , the sub-matrix composed of odd rows and even columns is represented by  $x2$ , the sub-matrix composed of even rows and odd columns is represented by  $x3$  and the sub-matrix composed of even rows and even columns is represented by  $x4$ . They are shown in equations (5)–(8) while equations (1)–(4) can be expressed by equations (9)–(12).  $I_{HL}$ ,  $I_{HH}$ ,  $I_{LH}$ , and  $I_{LL}$  are the results of the real-time Haar wavelet transform.

$$x1 = \begin{bmatrix} a & a \\ a & a \end{bmatrix}, \quad (5)$$

$$x2 = \begin{bmatrix} b & b \\ b & b \end{bmatrix}, \quad (6)$$

$$x3 = \begin{bmatrix} c & c \\ c & c \end{bmatrix}, \quad (7)$$

$$x4 = \begin{bmatrix} d & d \\ d & d \end{bmatrix}, \quad (8)$$

$$I_{HL} = \frac{(-x1 - x2 + x3 + x4)}{2}, \quad (9)$$

$$I_{HH} = \frac{(x1 - x2 - x3 + x4)}{2}, \quad (10)$$



a	b	a	b
c	d	c	d
a	b	a	b
c	d	c	d

FIGURE 5:  $4 \times 4$  two-dimensional matrix A.

$$I_{HL} = \frac{(-x_1 + x_2 - x_3 + x_4)}{2}, \quad (11)$$

$$I_{LL} = \frac{(x_1 + x_2 + x_3 + x_4)}{2}. \quad (12)$$

The Haar inverse wavelet transform calculates  $x_1, x_2, x_3, x_4$  using equations (13)–(16) to reconstruct matrix A.

$$x_1 = \frac{(I_{LL} - I_{HL} - I_{LH} + I_{HH})}{4}, \quad (13)$$

$$x_2 = \frac{(I_{LL} - I_{HL} + I_{LH} - I_{HH})}{4}, \quad (14)$$

$$x_3 = \frac{(I_{LL} + I_{HL} - I_{LH} - I_{HH})}{4}, \quad (15)$$

$$x_4 = \frac{(I_{LL} + I_{HL} + I_{LH} + I_{HH})}{4}. \quad (16)$$

**2.5. DMRFC Module.** In order to obtain deep semantic information from the image and improve image deblurring performance, this paper proposes the DMRFC module, which is shown in Figure 6.

The DMRFC module is composed of four multireceptive field channel blocks and a bottleneck layer. The semantic features of the image are extracted through multireceptive field channel blocks. The bottleneck layer is used to reduce the number of feature inputs to improve the compactness and computational efficiency of the model. A dense connection is used to strengthen the transmission of image features and make more effective use of image features. The DMRFC is shown as follows:

$$X_{\text{out}} = G\{(H_i[x_0, x_1, \dots, x_{i-1}]); \varepsilon\}, \quad (17)$$

where  $[x_0, x_1, \dots, x_{i-1}]$  represents the feature map generated by the multireceptive field channel blocks of 0, 1,  $\dots$ ,  $i-1$  layers in series;  $H_i$  means converting multiple input tensors into a single tensor;  $G(\cdot)$  means the output of the bottleneck layer,  $\varepsilon$  is the super-parameter of the bottleneck layer, and the filter size used in the bottleneck layer is  $1 \times 1$ .

The multireceptive channel blocks used by the DMRFC module are shown in Figure 7, which is a combination of the extensional receptive field blocks and the channel attention module. To increase the diversity of extracted feature in the extensional receptive field, four feature extraction branches are used that utilize  $3 \times 3$  convolution kernels with extensional rates of 1, 3, 5, and 7. The connection operation fuses the parallel feature maps of the four branches as shown in equation (18). The channel attention module from CBAM

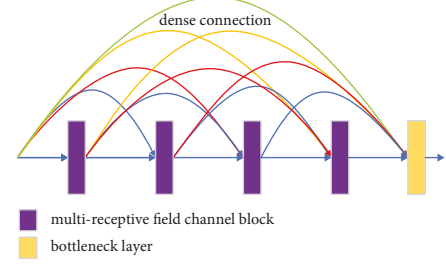


FIGURE 6: DMRFC module.

[32] is used for learning the weights of each channel, which make the nonlinear features be adaptively weighted at each channel. Merging the average and maximum pooling features improves the nonlinear representation of the network and its deblurring ability. The output of the multireceptive channel block is shown in

$$R = \text{cat}[L_K\{(x * w_{1 \times 1}) * w_{3 \times 3}^{d=1}\}, L_K\{(x * w_{1 \times 1}) * w_{3 \times 3}^{d=3}\}, \\ L_K\{(x * w_{1 \times 1}) * w_{3 \times 3}^{d=5}\}, L_K\{(x * w_{1 \times 1}) * w_{3 \times 3}^{d=7}\}]. \quad (18)$$

$$\text{Out} = (R \cdot \delta(\text{FC}(L_k(\text{FC}(\text{Maxpool}(C)))))) \\ + \text{FC}(L_k(\text{FC}(\text{Avgpool}(C)))) * w_{1 \times 1}. \quad (19)$$

In equations (18) and (19),  $x$  means the input features,  $*$  means the convolution operation, and  $w$  means the convolution layer. The superscript means the extensional convolution size and the subscript means the convolution kernel size.  $L_K$  denotes the Leaky ReLU activation function, the  $\text{cat}[\cdot]$  denotes the connection operation, and  $R$  means the fused features. The expression  $\text{Avgpool}$  means the average pooling and  $\text{FC}$  means the fully connection layer.  $\delta$  means the Sigmoid activation function and  $\text{Out}$  means the output of multireceptive channel blocks.

**2.6. Loss Function.** The mean square error (MSE) can obtain the difference between predicted and actual values; it is widely used as the loss function of the model evaluation. Because MSE does not consider the distribution characteristics of the image, which makes the model to be too smooth for restored image, while the SSIM can obtain the image edge details. The calculation methods of MSE and SSIM are shown in equations (20) and (21). Thus, in order to achieve a clearer image, we fuse MSE and SSIM to design the loss function, which is shown in equation (22). Consequently, we train the network by combining MSE and SSIM loss functions.

$$L_{\text{MSE}} = R - S^2, \quad (20)$$

$$L_{\text{SSIM}} = 1 - \text{SSIM}(R, S), \quad (21)$$

$$L_{\text{total}} = L_{\text{MSE}} + w_1 L_{\text{SSIM}}. \quad (22)$$

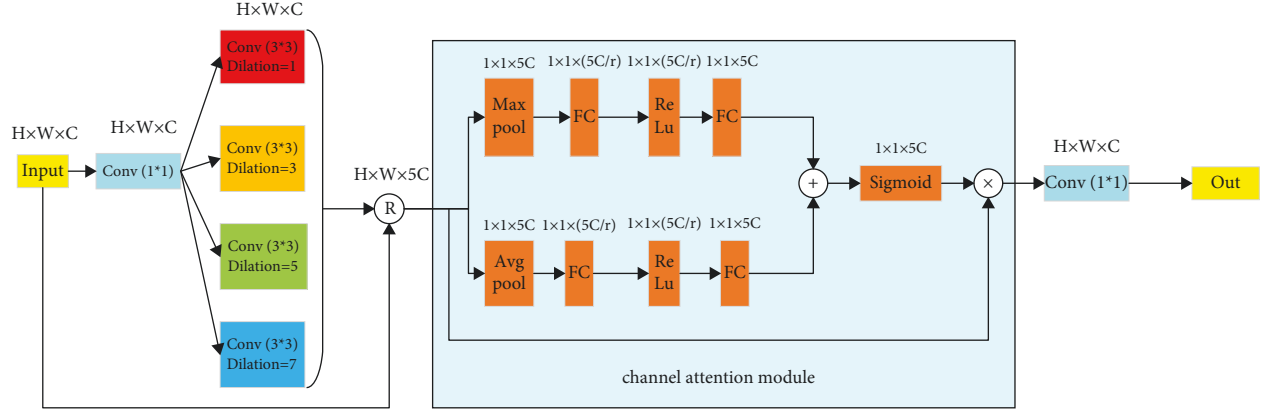


FIGURE 7: Multireceptive channel blocks.

---

Algorithm 1: The main steps of our method

---

Input: Sharp and blurred image pairs; Adam optimizer parameters  $lr$ , number of training rounds and epochs  $t$ ,  $num$

---

Output: Average PSNR, average SSIM

---

Define the U-Net model structure shown in Figure 1;

Initialization the model as described in Section 2.1;

Model training

Set the network training method to Adam;

For  $t=0,1,\dots,1000$

For  $num = 0,1,K$ , number of images to be trained

Forward propagation to calculation model output;

Algorithm 2

Calculate the Mean Square Error  $LMSE$  by equation (20);

Calculate the SSIM loss  $LSSIM$  by equation (21);

Calculate the  $L_{total}$  by equation (22);

Calculate the PSNR, SSIM by equations (23);

Calculate the SSIM by equations (24);

Calculate the gradient by back propagation;

if  $num = 2000$

$lr = lr * 0.5$ ;

endif

Endfor

Endfor

Calculate the average PSNR, average SSIM;

---

FIGURE 8: The main steps of our method.

In equations (20) and (21),  $R$  denotes an image to be blurred,  $S$  denotes a clear image, and  $SSIM(\cdot)$  denotes that it is calculated using structural similarity.  $w_1$  is the trade-off parameter and is set to 0.001.

Therefore, the main steps of our method are shown in Figure 8, where Algorithm 2 implements the forward propagation, which is shown in Figure 9.

### 3. Results

**3.1. Dataset.** In this paper, a GOPRO dataset was used to train the model, which consists of 3214 clear and blurred image pairs including 22 different scenes. Using 2103 image pairs as the training set and 1111 pairs of images as the test set, to enhance the generalization ability of the model, a data enhancement operation was carried out on the training set. This was conducted through random rotation including randomly flipping left, right, up, and down. Gaussian noise

was added as well. The rotation angles are 90, 180, and 270 degrees with the average value of the noise being 0 and the variance being 0.0001.

**3.2. Training Details.** In order to avoid model overfitting, images from the training set were randomly cropped to a size of  $256 \times 256$  pixels. The training period was set to 4000 rounds and the initial learning rate was set to  $1e-4$ , which was halved every 1000 rounds. The network optimization method used Adam [33], with parameters  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The network model was realized based on PyTorch framework and a GTX 2080 GPU.

**3.3. Quantitative Evaluation.** We use PSNR and SSIM to quantitatively evaluate the model, the larger the value, the better the image quality. Their calculation expressions are shown in

Algorithm 2: Forward propagation

---

Layer 1:  
 First Two-dimensional Haarwavelet transform;  
 One 32-channel depth-wise separable convolution;  
 Three 32-channel depth-wise separable residualconvolutions;

Layer 2:  
 Two-dimensional Haarwavelet transform;  
 One 64-channel depth-wise separable convolution;  
 Three 64-channel depth-wise separable residualconvolutions;

Layer 3:  
 Two-dimensional Haarwavelet transform;  
 One 128-channel depth separable convolution;  
 Three 128-channel residual depth-wise separable convolutions;

Layer 4:  
 Two-dimensional Haarwavelet transform;  
 One 256-channel depth-wise separable convolution;  
 Two 256-channel residual depth-wise separable convolutions;  
 Calculate DMRFC Moduleoutput by equations(18)-(19);  
 Two 256-channel residual depth-wise separable convolutions;  
 One 512-channel depth-wise separable convolution;  
 Two-dimensional inverse Haarwavelet transform;

Layer 5:  
 The output of theLayer 3are feature-fused through skip connections;  
 Three 128-channel residual depth-wise separable convolutions;  
 One 256-channel depth-wise separable convolution;  
 Two-dimensional inverse Haar wavelet transform;

Layer 6:  
 The output of theLayer 2are feature-fused through skip connections;  
 Three 64-channel residual depth-wise separable convolutions;  
 One 128-channel depth-wise separable convolution;  
 Two-dimensional inverse Haar wavelet transform;

Layer 7:  
 The output of theLayer 1 are feature-fused through skip connections;  
 Three 32-channel residual depth-wise separable convolutions;  
 One 3-channel depth-wise separable convolution;

---

FIGURE 9: The forward propagation algorithm.

TABLE 3: Comparison of quantitative evaluation on the GOPRO dataset.

Method	Literature [12]	Literature [13]	Literature [15]	Literature [34]	Literature [35]	Literature [36]	Literature [37]	Literature [38]	Literature [39]	Ours
PSNR	28.67	30.25	28.77	29.21	29.17	30.24	29.46	26.48	29.73	30.83
SSIM	0.969	0.9351	0.917	0.931	0.891	0.934	0.9365	0.839	0.902	0.948

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}^2}{\text{RMSE}^2}\right), \quad (23)$$

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (24)$$

where MAX is the maximum value for the image and the value of MAX is 255. RMSE denotes the root mean square error of the restored and blurred images.  $\mu_x$  and  $\mu_y$  denote the average values of the images  $x$  and  $y$ , respectively,  $\sigma_x^2$ ,  $\sigma_y^2$  denote the variance of images  $x$  and  $y$ , respectively, and  $\sigma_{xy}$  denotes the covariance of images  $x$  and  $y$ .  $c_1$  and  $c_2$  denote small constants to prevent division by zero.

Table 3 shows the average PSNR and average SSIM for the GOPRO test dataset, which are quantitatively compared with those for other methods. Through comparison, it was

TABLE 4: Comparison of running time and network model parameters size.

Method	Literature [33]	Literature [39]	Ours
Time (s)	1.87	1.25	0.56
Size (MB)	41.3	26.5	22.6

found that our method is superior to other methods when comparing PSNR and SSIM.

Table 4 shows the comparison of running time and model parameters size between different algorithms on the GOPRO test dataset. The running time is labelled as “Time.” The model parameters size is labelled using “Size.” The proposed model requires less time and fewer model parameters than in [33, 39].





FIGURE 10: Visual effect comparison. (a) Blurred picture. (b) Literature[15]. (c) Literature [37]. (d) Literature [38]. (e) Literature [39]. (f) The restored image in the paper.

## 4. Discussion

**4.1. Visual Effect.** Figure 10 shows a visual comparison example generated by this method and other methods on the GOPRO test dataset. [15, 37, 38] all show a relatively limited receptive field. Although high-quality images can be obtained, they are lacking in the restoration of texture details. Deblurring is realized by introducing a high-quality reference image into the deep network [39]. It is found that this method has a significant deblurring effect [39], but high-quality reference images are not always easy to obtain, which prevent recovery of the blurred image with spatial variation. The proposed model uses wavelet transform to retain image details and uses the DMRFC module to learn nonlinear image features. Compared with the details and structures generated by other methods, deblurring using this network maintains clear texture details, has a stronger effect, and leads to higher-quality visual effects.

**4.2. Effectiveness of Each Module.** To prove the effectiveness of the designed module, we conduct three model experiments to verify the effectiveness of wavelet transform and the DMRFC module on the model in this paper. In model 1, only depth-wise separable convolution and depth-wise separable residual convolution are included. The upsampling and downsampling of the image is controlled by the convolution step size, with five depth-wise separable residual blocks replacing the DMRFC module. Model 2 uses wavelet transform instead of image upsampling and downsampling

TABLE 5: Quantitative evaluation of model 3.

Method	Model 1	Model 2	Model 3 (ours)
PSNR	26.21	28.17	30.83
SSIM	0.871	0.891	0.948

and is based on model 1. Model 3 further introduces DMRFC based on model 2, with model 3 being the proposed model. Their results are compared in Table 5.

As shown in Table 5, in model 1, the image evaluation metric PSNR is 26.21 while that for SSIM is 0.871. After introducing wavelet transform in model 2, PSNR rises to 28.17 and SSIM rises to 0.891. This shows the introduction of wavelet.

Transform can improve model performance because wavelet transform can obtain sub-frequency information in four directions and learn more image details. When the DMRFC module is introduced in model 3, PSNR rises to 30.83 and SSIM rises to 0.948. It shows that the DMRFC module profits to improve model restoration quality, because they reduce information reuse and increase the image receptive field.

## 5. Conclusion

In this paper, an image deblurring method using an improved U-Net model is proposed, in which a two-dimensional discrete Harr wavelet is introduced and a DMRFC module is designed. A wavelet transform is used instead of the downsampling in the model and the inverse wavelet

transform is used  $t$  instead of the upsampling. Through the wavelet transform, more image details are obtained and computational complexity is reduced. The MRFC module is used to connect the multireceptive channel blocks using a dense connection, which reduces the number of multireceptive field channel block parameters and strengthens the feature transmission. Our method significantly reduces the number of model parameters. It also reduces the running time for the clear image restored by the model and achieves good visual results when deblurring the image. The average PSNR and average SSIM for the GOPRO dataset is 30.83 and 0.948, respectively, which is higher than that for other image deblurring methods. Therefore, our proposed method achieves a better deblurring effect and is more lightweight.

In future work, by further reducing model parameters and calculations, the ability of the model in restoring the blurred images can be further improved. This will make the model more lightweight and allow for exploration of its application in resource-constrained and widely used mobile devices.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publishing of this paper.

## Acknowledgments

This work was supported by the Heilongjiang Education Department Basic Scientific Research Business Research Innovation Platform "Scientific Research Project Funding of Qiqihar University" (Grant: 135409421).

## References

- [1] Y. Chen, V. Phonevilay, J. Tao et al., "The face image super-resolution algorithm based on combined representation learning," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30839–30861, 2021.
- [2] Y. Yuan, W. Su, and D. D. Ma, "Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13–19, IEEE, Seattle, WA, USA, June 2020.
- [3] J. Zhang, J. Sun, J. Wang, Z. Li, and X. Chen, "An object tracking framework with recapture based on correlation filters and Siamese networks," *Computers & Electrical Engineering*, vol. 98, Article ID 107730, 2022.
- [4] R. Xia, Y. Chen, and B. Ren, "Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 2, pp. 1319–1578, 2022.
- [5] J. M. Zhang, W. J. Feng, T. Y. Yuan, J. Wang, and A. K. Sangaiah, "SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking," *Applied Soft Computing*, vol. 118, Article ID 108485, 2022.
- [6] Y. Xu, Y. Zhu, Y. H. Quan, and H. Ji, "Attentive deep network for blind motion deblurring on dynamic scenes," *Computer Vision and Image Understanding*, vol. 205, no. 11, Article ID 103169, 2021.
- [7] Z. Z. Lian, H. Zhao, Q. J. Zhang, H. Z. Wang, and E. Erdun, "Enhancement of biomass material characterization images using an improved U-net," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1515–1528, 2022.
- [8] J. Pan, "Research progress on deep learning-based image deblurring," *Computer Science*, vol. 48, no. 3, pp. 9–13, 2021.
- [9] J. Li, Ke Li, and Bo Yan, "Scale-aware deep network with hole convolution for blind motion deblurring," in *Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 8–12, IEEE, Shanghai China, July 2019.
- [10] Y. H. Quan, P. K. Lin, Y. Xu, Y. S. Nan, and H. Ji, "Nonblind image deblurring via deep learning in complex field," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1–14, 2021.
- [11] K. Liu, C. Yeh, J. Chung, and C. Chang, "A motion deblur method based on multi-scale high frequency residual image learning," *IEEE Access*, vol. 8, pp. 66025–66036, 2020.
- [12] Y. Zhang, S. Ma, Xi Zhang, L. Li, W. H. Ip, and K. L. Yung, "EDGAN: motion deblurring algorithm based on enhanced generative adversarial networks," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 8922–8937, 2020.
- [13] H. Zhang, Y. Dai, H. Li, and P. Koniusz, "Deep stacked hierarchical multi-patch network for image deblurring," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15–20, IEEE, Long Beach, CA, USA, June 2019.
- [14] L. Chen, Q. Sun, and F. Wang, "Attention-adaptive and deformable convolutional modules for dynamic scene deblurring," *Attention-adaptive and deformable convolutional modules for dynamic scene deblurring Information Sciences*, vol. 546, pp. 368–377, 2021.
- [15] T. Zeng and C. Diao, "Single image motion deblurring based on modified denseNet," in *Proceedings of the 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDI)*, pp. 521–525, IEEE, Taiyuan, TY, China, October 2020.
- [16] T. Feng, C. Wang, X. Chen, H. Fan, Z. ZengLi, URNet: a U-Net based residual network for image dehazing," *Applied Soft Computing*, vol. 102, no. 12, Article ID 106884, 2021.
- [17] D. Lin, Y. Li, T. L. Nwe, S. Dong, and Z. M. Oo, "RefineU-Net: improved U-Net with progressive global feedbacks and residual attention guided local refinement for medical image segmentation," *Pattern Recognition Letters*, vol. 138, pp. 267–275, 2020.
- [18] W. Zhang, J. Li, and Z. Hua, "Attention-based tri-UNet for remote sensing image pan-sharpening," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, pp. 3719–3732, 2021.
- [19] C. M. Fan, T. J. Liu, and K. H. Liu, "SUNet: swin transformer UNet for image denoising," 2022, <https://arxiv.org/abs/2202.14009>.
- [20] H. Zang, L. Zhu, Z. Ding, and S. LiZhan, "Cascaded dense-UNet for image super-resolution," *Journal of Circuits Systems and Computers*, vol. 29, no. 8, Article ID 2050121, 2019.
- [21] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 568–576, 2020.

- [22] G. Alimjan, Y. Jiaermuhamaiti, H. Jumahong, and P. ZhuNurmat, "An image change detection algorithm based on multi-feature self-attention fusion mechanism UNet network," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, no. 14, Article ID 2159049, 2021.
- [23] Y. Li, Y. Wang, T. Leng, and Z. J. Zhijie, "Wavelet U-net for medical image segmentation," *Artificial Neural Networks and Machine Learning - ICANN 2020*, Springer, Cham, Switzerland, 2020.
- [24] F. A. Dharejo, F. Deeba, Y. Zhou et al., "TWIST-GAN: towards wavelet transform and transferred gan for spatio-temporal single image super resolution," *ACM Transactions on Intelligent Systems and Technology*, vol. 12, no. 6, pp. 1–20, 2021.
- [25] H. Yang and Y. Wang, "An effective and comprehensive image super resolution algorithm combined with A novel convolutional neural network and wavelet transform," *IEEE Access*, vol. 9, pp. 98790–98799, 2021.
- [26] M. Sarkar, U. Mondal, and D. Nandi, "A novel wavelet-based image defogging using dark channel prior and guided filter," *Advances in Communication, Devices and Networking*, vol. 537, pp. 239–246, 2019.
- [27] A. Khmag and N. Kamarudin, "Natural image deblurring using recursive deep convolutional neural network (R-DbCNN) and second-generation wavelets," in *Proceedings of the 2019 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 17–19, IEEE, Kuala Lumpur, Malaysia, September 2019.
- [28] Z. Y. Khan and Z. Niu, "CNN with depthwise separable convolutions and combined kernels for rating prediction," *Expert Systems with Applications*, vol. 170, no. 5, Article ID 114528, 2021.
- [29] H. Gao, Y. Yang, C. Li, L. Gao, and B. Zhang, "Multiscale residual network with mixed depthwise convolution for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3396–3408, 2021.
- [30] Y. N. Wu, K. B. Jia, and Z. H. Sun, "Facial expression recognition based on multi-scale feature fusion convolutional neural network and attention mechanism," *Pattern Recognition and Computer Vision*, vol. 13020, pp. 324–335, 2021.
- [31] Z. Q. Qin, P. Y. Zhang, F. Wu, and X. Li, "FcaNet: frequency channel attention networks," in *Proceedings of the 2021 IEEE International Conference on Computer Vision (ICCV 2021)*, October 2021.
- [32] S. H. Wang, S. Fernandes, Z. Q. Zhu, and Y. D. Zhang, "AVNC: attention-based VGG-style network for COVID-19 diagnosis by CBAM," *IEEE Sensors Journal*, vol. 99, p. 1, 2021.
- [33] H. Liu and X. Tian, "AEGD: adaptive gradient decent with energy," 2020, <https://arxiv.org/pdf/2010.05109.pdf>.
- [34] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: blind motion deblurring using conditional adversarial networks," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [35] J. Zhang, J. Pan, J. Ren et al., "Dynamic scene deblurring using spatially variant recurrent neural networks," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [36] X. Tao, H. Gao, J. Wang, X. Shen, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8174–8182, IEEE, Salt Lake City, UT, USA, June 2018.
- [37] W. Ren, J. Zhang, J. Pan et al., "Deblurring dynamic scenes via spatially varying recurrent neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, p. 1, 2021.
- [38] Z. Jiang, Y. Zhang, D. Zou, J. Ren, J. Lv, Y. Liu, Learning event-based motion deblurring," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3317–3326, IEEE, Seattle, Washington, USA, June 2020.
- [39] Y. W. Li, Y. Luo, and J. W. Lu, "Reference-guided deep deblurring via a selective attention network," *Applied Intelligence*, vol. 52, no. 4, pp. 3867–3879, 2022.