

Article

Not peer-reviewed version

---

# Frequency-Domain Feature Enhanced Convolutional Neural Network for Image Deblurring Algorithm

---

[Yecai Guo](#), [Lixiang Ma](#)<sup>\*</sup>, [Yangyang Zhang](#)

Posted Date: 6 February 2026

doi: 10.20944/preprints202602.0532.v1

Keywords: image deblurring; frequency domain feature; fourier transform; convolutional neural network



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Frequency-Domain Feature Enhanced Convolutional Neural Network for Image Deblurring Algorithm

Yecai Guo <sup>1,3</sup>, Lixiang Ma <sup>1,2,\*</sup> and Yangyang Zhang <sup>3</sup>

<sup>1</sup> School of Electrical and Electronic Engineering, Anhui Institute of Information Technology, China

<sup>2</sup> State Key Laboratory of Millimeter Waves, China;

<sup>3</sup> College of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Jiangsu, China;

\* Correspondence: Ma Lixiang, e-mail: lxma@iflytek.com

## Abstract

To address the issues of insufficient restoration of texture details in deblurred images and inadequate learning of frequency domain features, an image deblurring algorithm based on frequency domain feature enhancement and convolutional neural networks is proposed. First, a Fourier residual module with a parallel structure is constructed to achieve collaborative learning and modeling of spatial and frequency domain features. By introducing the Fourier transform, the frequency domain feature learning is enhanced to improve the restoration of texture details. Second, after a gated feed-forward unit is applied to Fourier residual module, its nonlinear representation capability is further improved. In addition, a supervised attention module is introduced at the decoder stage to promote more effective extraction of key features essential for image reconstruction. Experimental results have demonstrated that the proposed algorithm effectively removes blur while better preserving image details.

**Key words:** image deblurring; frequency domain feature; fourier transform; convolutional neural network

---

## 1. Introduction

In the digital information era, images, as intuitive and efficient information carriers, play a key role in many fields such as medical care<sup>[1]</sup>, security, remote sensing images<sup>[2]</sup>, etc. However, in the process of image acquisition and transmission, images are often blurred due to many factors such as human operation, jitter of imaging equipment, and rapid movement of objects, which seriously affects visual perception and both the acquisition and recognition of information in images. Blurring is a phenomenon in which the quality suffers from distortion, and the images lose the original edge information and detail information, seriously affecting the accurate transmission and analysis of information. There are many reasons for blurring, such as camera shake, rapid movement of the subject or slow shutter speeds can cause blurring artefacts in the image.

With the advancement of technology and the urgent need for clear images in various industries, various image deblurring methods have emerged. Image deblurring is often viewed as a degenerate problem based on an image blurring model that models a blurred image as a convolution of a clear image with a blurring kernel and estimates a potentially clear image by inverse convolution. In early deblurring tasks, some deblurring methods such as filtering algorithms<sup>[3,4]</sup>, least squares<sup>[5]</sup> and equalization filters<sup>[6]</sup>, rely on various constraints and a priori knowledge, and may model to the blurring degradation. Their strong dependence on priors or blur kernels also directly leads to their ability to handle blur only under specific conditions, and reduce adaptability to real complex scenes.

With the development of computer technology, some researches began to use Convolutional Neural Networks(CNNs) for image deblurring task, and leverage their powerful fitting capabilities

and combine them with traditional methods to improve the image restoration effect. Early approaches often relied on two-stage processes involving kernel estimation and non-blind deconvolution. However, estimating accurate blur kernels for dynamic scenes remains challenging. To overcome this, research shifted towards end-to-end Convolutional Neural Networks (CNNs). Tao et al.<sup>[7]</sup> proposed a recursive network that shares network weights across scales, which requires fewer training parameters compared to previous multiscale deblurring methods. Gao et al.<sup>[8]</sup> proposed parameter-selective sharing and nested skip connections, which incorporate an encoder-decoder structure in the sub-network with skip connections, which are able to share the network parameters among themselves while reducing the memory requirement. Zhang et al.<sup>[9]</sup> proposed a multi-patch hierarchical network where the input blurred image is divided into more available patches and processed independently by each sub-network, and then the features are fused step by step in order to recover a clear image. Cho et al.<sup>[10]</sup> proposed a multi-input multi-output UNet (MIMO-UNet), which utilizes the different scales of blur as network inputs to provide multilevel information to solve the image blurring problem. While early methods often relied on small convolution kernels, which limited the effective receptive field and struggled to capture global motion blur. To address this limitation without abandoning the efficiency of CNNs, recent studies have revitalized CNN architectures. Ding et al.<sup>[11]</sup> demonstrated that by scaling up convolutional kernels to large sizes, CNNs can achieve effective receptive fields comparable to Transformers, effectively recovering large-scale motion structures. Building on this, Zhou et al.<sup>[12]</sup> proposed an efficient cascaded multiscale adaptive network that leverages hierarchical feature extraction to handle diverse degradation patterns with significantly lower computational cost than self-attention based models. To better handle the global properties of blur, Cui et al.<sup>[13]</sup> introduced frequency-domain learning into CNNs. By integrating Fast Fourier Transform (FFT) modules with spatial convolutions, these methods can explicitly model global blur patterns in the spectral domain while preserving local details in the spatial domain. These modern CNN-based approaches prove that with optimized structural designs, CNNs can achieve state-of-the-art deblurring performance with better inference efficiency than heavy Transformer architectures.

Existing image deblurring algorithms mainly focus on feature learning in the spatial domain to estimate potentially clear images. However, in the image deblurring task, frequency domain feature information also plays an important role in modelling the nonlinear mapping relationship between blurred images and clear ones. The low-frequency part of an image contains global information that carries the overall contour, structure and color distribution, while the high-frequency part contains edges, textures and rich local details. In recent years, some studies have begun to introduce frequency domain feature learning into image deblurring networks to enhance the texture structure and clarity of the recovered image through frequency domain feature learning. Zou et al.<sup>[14]</sup> designed a wavelet reconstruction module to complement the spatial domain by recovering the information in the frequency domain so that the recovered image contains more details in the high frequency. Huang et al.<sup>[15]</sup> used a parallel learning mechanism to separately extract the features and subsequently fused them in the frequency domain, while using local details and global information of the image to remove blur. Based on the above studies, this paper proposes Frequency Domain feature Enhanced convolutional neural Network for image deblurring (FDENet) in conjunction with Fourier Transform, which aims to efficiently establish the model on high frequency and low frequency feature information in images through frequency domain feature learning. The main work is as follows: (1) The frequency domain enhancement module is proposed to integrate the inverted residuals and Fourier residuals modules, and the parallel branching structure is adopted to effectively integrate the spatial and frequency domain features, so as to enhance the ability of the model to capture and express different features. (2) A gated feedforward unit is introduced after the frequency domain enhancement module to improve the feature representation capability of the network. In addition, a supervised attention module is introduced at the decoder side to

<https://doi.org/10.3390/xxxxx>

effectively enhance the delivery of key features in the decoding process. (3) Through ablation and comparison experiments, it is verified that the proposed FDENet removes image blur while the recovered image retains more image detail information.

## 2. Materials and Methods

### 2.1. Network Structure

The overall structure of the Frequency Domain feature-Enhanced image deblurring Network(FDENet) is shown in Figure 1, which adopts a U-shape network structure and realizes the information transmission between the encoder and the decoder through hopping connections. The blurred image is the input of the FDENet, and the shallow features are firstly obtained by  $3 \times 3$  convolution with LeakyReLU activation function. The encoder consists of a Frequency Enhancement Block(FEB), which can improve the deblurring effect through the enhancement of the frequency domain feature information. The shallow features are gradually extracted into deep features  $F_{enc,i}$  through the three-level encoder, which is expressed as follows:

$$F_{enc,i} = \text{Encoder}_i(F_{enc,i-1}) \quad (1)$$

where  $F_{enc,i}$  denotes the output feature of the  $i$ th encoder, and  $\text{Encoder}_i$  stands for the composite function operation of the FEB. After the features  $F_{enc,i}$  extracted by the current encoder are passed to the next encoder layer through downsampling, the spatial size of the feature map is halved and the number of channels is expanded to twice. After the feature is extracted by the encoder at each scale, the resulting feature map is passed through the corresponding decoder. The decoder also consists of three stages, each of which consists of a Frequency Domain Augmentation Block(FDAB) and a Supervised Attention Block(SAB). The FDENet adopts an asymmetric structure, where the decoder contains additional SABs compared to the encoder to improve the feature learning capability in the decoding stage. The features after the SAB are upsampled, and the spatial size of the upsampled feature map is doubled and the number of channels is halved. The feature map after upsampling is aggregated with the corresponding encoder stage by skip connection map, which is expressed as follows:

$$F_{dec,j} = H_{FEB}(F_{SAB,j} + F_{enc,i}) \quad (2)$$

where  $F_{dec,j}$  denotes the output feature of the  $i$ th decoder, and  $H_{FEB}$  represents the composite function operation of the FEB.

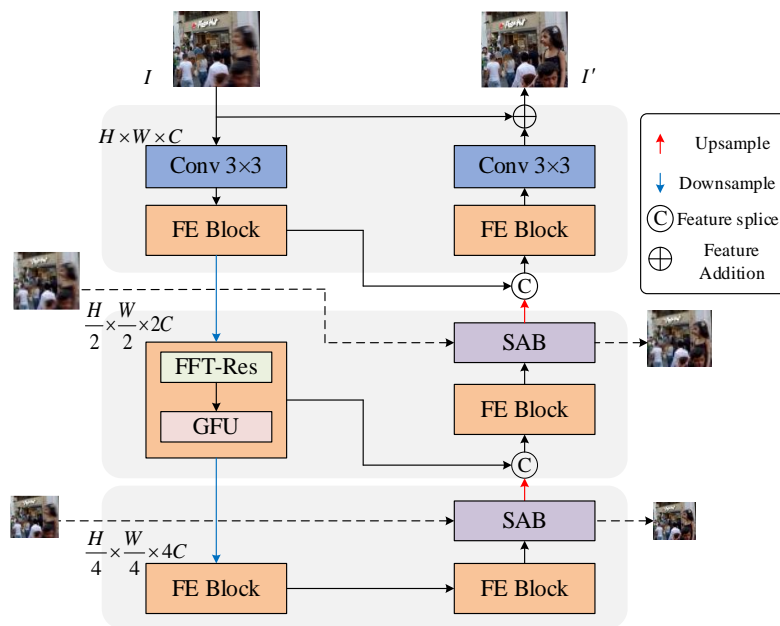


Figure 1. Network architecture

Skip connections directly transmit the features from the shallow layers in the encoder to the corresponding layers in the decoder, the spatial details and texture information that play an important role in the restoration of the blurred image are preserved. After the last decoder, a  $3 \times 3$  convolution is applied to the feature map to obtain the residual image, which is added to the original image to get hold of the final restored image.

## 2.2. Fast Fourier Transform Residual Block

The structure of Fast Fourier Transform Residual Block (FFT-Res) is shown in Figure 2, which consists of two branches, that is, the spatial domain branch and the frequency domain branch. By jointly learning features in the spatial and frequency domains, the ability of network to model image details and structures is effectively enhanced. In the frequency domain branch, firstly, the input feature map  $X$  undergoes a two-dimensional fast Fourier variation to obtain its imaginary part  $f_i(x)$  and real part  $f_r(x)$ , and subsequently, these two parts are spliced in the channel dimension. The splicing result  $f$  constructs the relationship between high-frequency and low-frequency information of the image through cascaded convolutional layers, which consist of two  $1 \times 1$  convolutions and a GELU activation function.

Compared to global feature learning in the spatial domain, which relies on deeper networks or larger receptive fields, frequency domain learning can efficiently capture the global structural information of the blurred images with lower structural complexity. At the same time, the high-frequency features in the frequency domain are more obvious, which helps the model to efficiently extract texture and edge information, and restore image details. By transferring the feature map to the frequency domain, the aim is to more effectively model the high-frequency and low-frequency feature information in the image, so as to enhance the restoration effect of texture details and overall visual quality of the deblurred image.

The feature information learned in the frequency domain needs to be transferred back to the spatial domain and fused with another branch to perform a splitting operation to obtain new real and imaginary parts. Then, the inverse Fourier transform is used to transform the features from the frequency domain back to the spatial domain to obtain the output  $Y_{fft}$  of the frequency domain branch. The above computational process can be represented as:

$$f' = \text{Conv}(\sigma(\text{Conv}(f))) \quad (3)$$

$$(f'_r, f'_i) = \text{chunk}(f') \quad (4)$$

$$Y_{\text{fft}} = \text{ifft}(f'_r + j * f'_i) \quad (5)$$

where  $\text{fft}$  is the fast Fourier variation,  $\text{ifft}$  represents its inverse variation,  $\text{chunk}$  denotes splitting,  $\text{Conv}$  stands for  $1 \times 1$  convolution, and  $\sigma$  denotes the ReLU activation function.

Due to the increased computational consumption caused by the use of Fourier transform in the frequency domain branch, the number of the inverted residual parameters is adopted in the spatial domain branch instead of ordinary residual parameters. Compared to ordinary residuals, the inverted residuals utilize depthwise separable convolution and effectively reduce the number of the parameters and computational overhead by increasing the dimension first and then decreasing it. Finally, the spatial domain branch and frequency domain branch are added element by element to obtain the final output  $Y$ . The above process can be represented as:

$$Y = Y_{\text{fft}} + Y_{\text{res}} + X \quad (6)$$

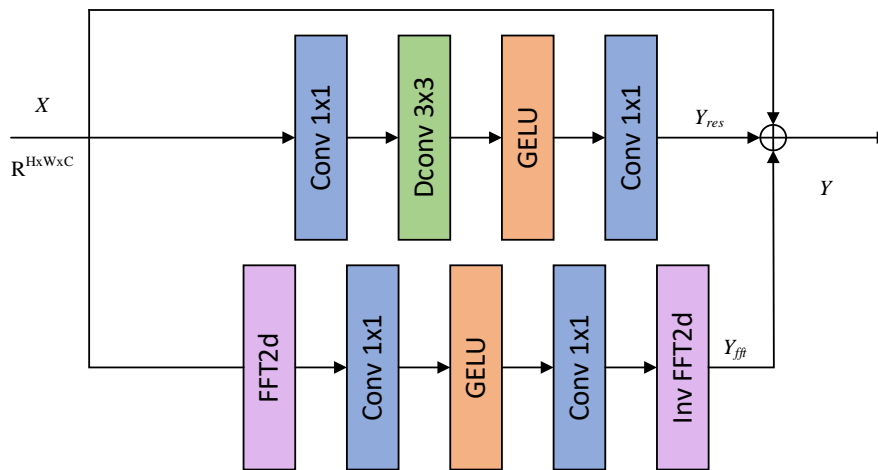


Figure 2. Fast Fourier Transform residual block

### 2.3. Gated Feed-Forward Unit

In order to improve the feature representation capability and the modelling ability of the network, a Gated Feed-forward Unit (GFU) is cascaded after Fast Fourier Transform (FFT) residual block, as shown in Figure 3. The feature information extracted by FFT residual block contains both spatial and frequency domain information. Therefore, adaptively enhancing important feature information and suppressing redundant feature information through a gating mechanism helps improve the deblurring effect. Firstly, the input features  $X$  are subjected to layer-normalization to stabilize the training process and speed up the convergence of the model.

The layer-normalized features are expanded in channels through  $1 \times 1$  convolution, which doubles the number of channels of the original features, and then the local spatial modelling capability is introduced through a  $3 \times 3$  depthwise separable convolution.

The convolution output is split into two branches according to the number of channels, and the lower half of the branch is subjected to a GELU activation function to form a gating unit, which is then multiplied element-by-element with the features in the upper half of the branch to achieve adaptive modulation of the feature information. The gating mechanism can strengthen the important feature information, suppress the redundant features, and effectively improve the expression ability of the features.

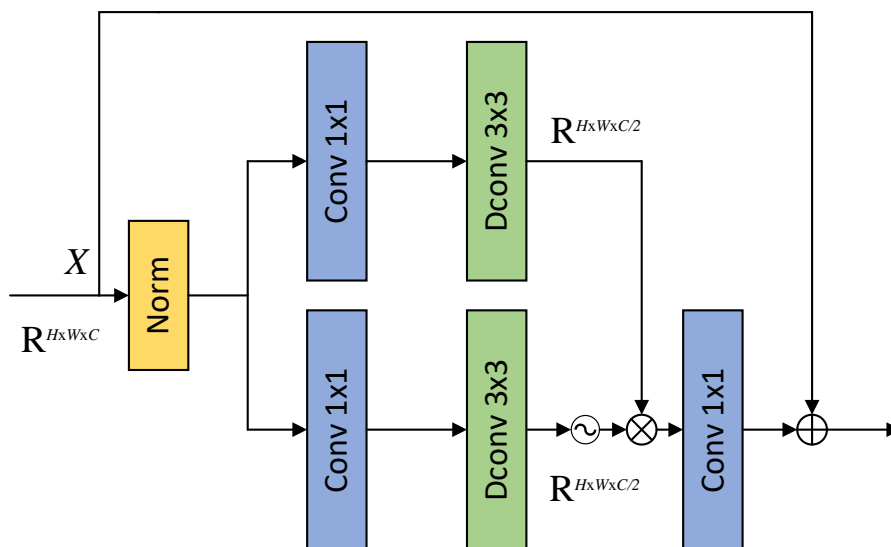


Figure 3. Gated feed-forward unit

Finally, the number of channels is varied to the initial dimension via using  $1 \times 1$  convolution. By combining the depthwise convolution with the gating mechanism, the GFU enhances mapping capability of the nonlinear feature in the network, which in turn improves the ability of the model to extract the features from the FFT residual block. The mathematical expression for the GFU is defined as follows:

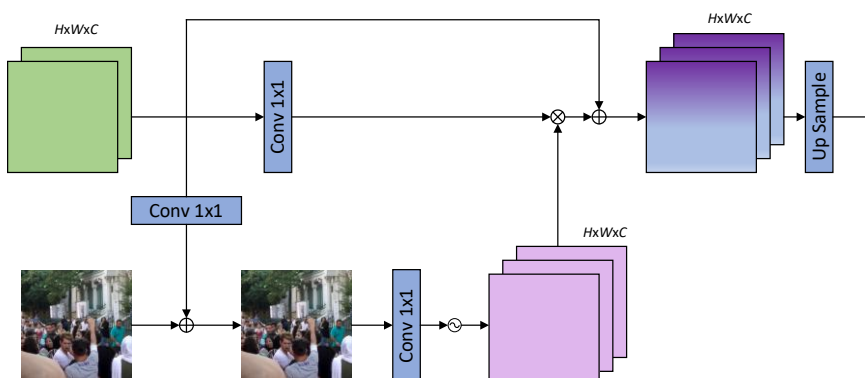
$$GFU = Conv(Gating(X)) + X \tag{7}$$

$$X = \sigma(DConv(Conv(LN(X)))) \otimes DConv(Conv(LN(X))) \tag{8}$$

where  $\otimes$  denotes element-by-element multiplication,  $Conv$  represents  $1 \times 1$  convolution,  $DConv$  stands for depthwise convolution,  $LN$  signifies layer normalization, and  $\sigma$  is the GELU nonlinear activation function.

#### 2.4. Supervised Attention Block

In the Multiple-Input Multiple-Output(MIMO) structure[10], downsampled blurr images are passed into the network as inputs to enrich feature information. Inspired by this, the supervised attention module is introduced into the decoder, information exchange between consecutive decoders is achieved through ground truth supervision and assigning attention weights to feature maps. The structure is shown in Figure 4. Under the ground truth supervision of the predicted image, features with less information can be suppressed, and effective features can be propagated to the next decoding layer.



<https://doi.org/10.3390/xxxxx>

**Figure 4.** Supervised attention block

Firstly, the input features are added to the low-scale blurred image to generate a ground-truth supervised image, and then the attention weights are extracted from this image through convolution for supervised attention. The enhanced image features are fed into the next decoder by means of ground-truth supervision, which reduces redundant features to improve network efficiency and enhance network deblurring capability. The input features  $X$  are first obtained through  $1 \times 1$  convolution to produce residual features and then added to the blurred image  $I$  to obtain a low-scale predicted image  $I'$ . The obtained prediction image serves as ground-truth supervision to help the network reduce the pixel differences between the final predicted image and the ground-truth image. The predicted image is sequentially passed through  $1 \times 1$  convolution and the Sigmoid activation function to generate an attention feature map with weights. Subsequently, the attention feature map is multiplied element-wise with the features from another branch, and finally, the output of the SAB is obtained through a residual connection. The output is upsampled and used as the feature input for the next decoder. The specific process is represented as follows:

$$I' = Conv(X) \quad (9)$$

$$SAB(X) = Conv(Sigmoid(I')) \otimes Conv(X) + X \quad (10)$$

where  $Conv$  is the  $1 \times 1$  convolution,  $\otimes$  denotes element-by-element multiplication, and  $Sigmoid$  stands for the activation function.

The traditional upsampling operations are typically implemented through some methods such as transposed convolution or bilinear interpolation, but there are certain limitations in practical applications, such as the tendency to produce artefacts, insufficient detail restoration, or structural distortion, etc. Therefore, the CARAFE(Content-Aware ReAssembly of FEatures) [16] is used instead of the traditional upsampling method to obtain a high-quality feature map recovery, and its structure is shown in Figure 5.

The CARAFE, during upsampling, adaptively generates reconstruction weights by incorporating content information from the current feature map to achieve more refined feature reconstruction. This module primarily consists of two parts: the kernel prediction module and the feature recombination module. The kernel prediction module comprises three sub-modules: channel compression, content encoding, and kernel normalization, and is used to generate upsampling kernels. Given an input feature map of size  $H \times W \times C$  and an upsampling rate  $\sigma$ , the channels of the input feature map are first reduced using a  $1 \times 1$  convolution to reduce subsequent computational complexity. Content encoding involves predicting upsampling kernels for the compressed feature map using a convolutional layer. The convolved feature map is then expanded along the channel dimension to obtain a recombined upsampling kernel of size  $\sigma H \times \sigma W \times k_{up} \times k_{up}$ . Kernel normalization involves normalizing the upsampling kernel using the Softmax function before applying the feature recombination module.

The feature reorganization module is responsible for mapping each position of the output feature map back to the input feature map and performing a dot product operation between the  $k_{up} \times k_{up}$  region centered on the output feature map and the upsampling kernel of the prediction points to obtain the output. The content-aware reorganization module allows more attention to be paid to the relevant feature information in the local region, so richer feature information can be extracted by using CARAFE as the upsampling method.

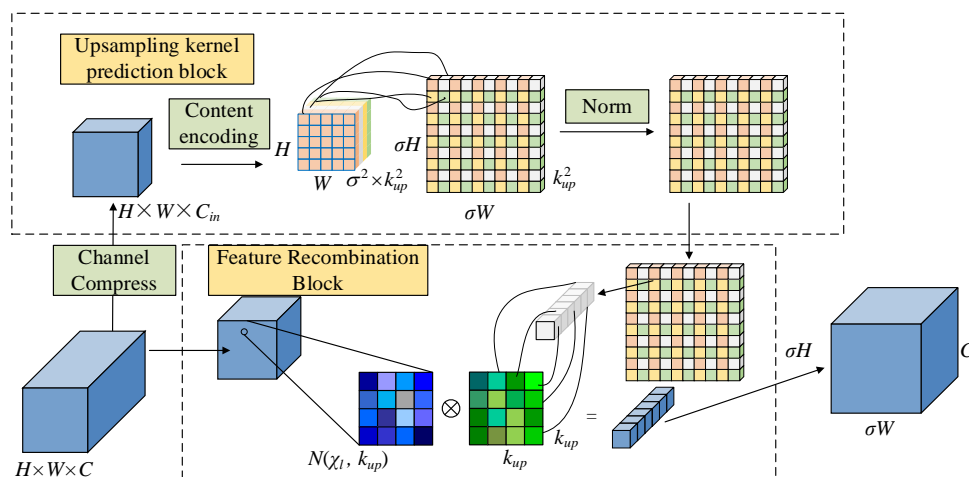


Figure 5. Content-aware ressembly of features architecture

### 2.5. Loss Function

The total loss  $L_{total}$  defined in this paper includes the content loss  $L_{charbonnier}$  and the loss function auxiliary term  $L_{freq}$ , whose expression is written as follows:

$$L_{total} = L_{charbonnier} + \alpha \times L_{freq} \quad (11)$$

where  $L_{charbonnier}$  represents Charbonnier loss,  $L_{freq}$  denotes frequency loss,  $\alpha$  is the loss function weight coefficient set to 0.1.

Compared to the L1 loss function, the presence of a constant term ensures that the gradient does not become too small when the loss function approaches zero. The advantage of this loss function is that its curve is smoother, and it has better tolerance for small errors, so it leads to better convergence during training. As a result, it is more stable in handling minor errors, avoids the vanishing gradient phenomenon, and thus improves the efficiency and stability of model optimization. The expression for the Charbonnier loss is defined as follows:

$$\mathcal{L}_{charbonnier} = \sqrt{(I_{pred} - I_{gt})^2 + \eta^2} \quad (12)$$

where  $I_{pred}$  and  $I_{gt}$  are the restored image and the clear image, and  $\eta$  is the regularization constant and set to 0.001.

The frequency loss<sup>[17]</sup> is introduced as an auxiliary term in the loss function. The frequency loss restores the high-frequency details of the blurred image by minimizing the difference between the blurred image and the true clear image in the frequency domain. The expression is defined as follows :

$$\mathcal{L}_{freq} = \sqrt{(\mathcal{F}(I_{pred}) - \mathcal{F}(I_{gt}))^2 + \eta^2} \quad (13)$$

where  $\mathcal{F}$  denotes the FFT, and  $\eta$  is the regularization constant term.

## 3. Results

### 3.1. Dataset and Parameter Settings

The dataset used in this section is derived from video sequences captured by Zhang et al.[9] using GoPro cameras, and blurred images are obtained by averaging consecutive exposure frames. The GoPro dataset consists of 3,214 pairs of blurred and clear images at a resolution of  $1,280 \times 720$ , 2,103 pairs of which are used for training set and 1,111 pairs of which are used for testing set.

<https://doi.org/10.3390/xxxxx>

Additionally, to analyze the generalization ability of the network, the HIDE dataset proposed by Shen et al.<sup>[18]</sup> is used for comparative experiments, which covers a wide range of scenes and various types of non-uniform blurring, consists of 8,422 pairs of blurred and clear images, where 2,025 pairs of blurred and clear images are employed the testing set.

The experiments were conducted on the Windows 11 OS, with a Python 3.6 compilation environment and PyTorch 1.7.1 as the deep learning framework. All results were obtained by running the experiment on an NVIDIA GeForce RTX 3060 Ti. During data preprocessing, the input images were first randomly cropped to 256×256 pixels, and then horizontally flipped with a 50% probability to enhance data diversity and model generalization capabilities. The experiment was trained for a total of 3,000 epochs, with a batch size of 4. The Adam optimizer with a warm-up strategy was used for optimization, where the momentum decay exponents of the Adam optimizer were set to  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . The learning rate adjustment strategy employed a more robust cosine annealing strategy, with an initial learning rate of  $1 \times 10^{-4}$  and a minimum learning rate of  $1 \times 10^{-6}$ .

### 3.2. Quantitative Comparative Analyses

To validate the effectiveness of the designed module, quantitative comparison experiments were conducted in this section. Peak Signal-to-Noise Ratio(PSNR) and Structured Similarity Index Measurement(SSIM)<sup>[19]</sup> were used as image quality evaluation metrics for quantitative evaluation. Higher PSNR and SSIM values indicate that the model has better deblurring capabilities and higher image reconstruction quality. The performance of model was evaluated on the benchmark datasets GoPro and HIDE, and compared with the SRN<sup>[7]</sup>, PSS-NSC<sup>[8]</sup>, DMPHN<sup>[9]</sup>, SDWNet<sup>[14]</sup>, VDN<sup>[20]</sup>, MSAN<sup>[21]</sup>, BANet<sup>[22]</sup>, MPRNet<sup>[23]</sup>, and MIMO-UNet<sup>[10]</sup>, as shown in Table 1.

**Table 1.** Test results on GoPro and HIDE datasets

Model	GOPRO		HIDE	
	PSNR	SSIM	PSNR	SSIM
SRN	30.26	0.934	28.36	0.903
PSS-NSC	30.92	0.942	29.10	0.913
DMPHN	31.20	0.940	29.10	0.917
SDWNet	31.26	0.966	28.99	0.957
VDN	31.65	0.951	29.27	0.923
MIMO-UNet	31.73	0.951	29.28	0.920
MSAN	32.24	0.956	30.04	0.935
BANet	32.54	0.957	30.16	0.930
MPRNet	32.66	0.959	30.96	0.939
FDENet	32.85	0.961	30.93	0.941

As shown in the Table 1, the FDENet achieved a PSNR value of 32.85 dB on the GoPro dataset. The SSIM value reached 0.961 and was second only to the SDWNet, which used wavelet transformation, and achieved the second-best value. Compared to the MPRNet, which stacks multiple subnetworks, its PSNR has an improvement of about 0.19 dB comparison with the MPRNet, and SSIM has an improvement of about 0.02 comparison with the MPRNet. On the HIDE dataset, the FDENet achieved a PSNR value of 30.93 dB and an SSIM value of 0.941, it also demonstrates excellent performance of the FDENet. Compared to other algorithms, the FDENet exhibits outstanding performance in image deblurring, and ability to effectively fuse spatial and frequency domain features, and utilize global information through the frequency domain enhancement module. Additionally, the FDENet employs the CARAFE to alleviate the limitations of transposed convolutions during upsampling. Experimental results have indicated that the FDENet can achieve

outstanding performance on two commonly used datasets while also validating its generalization capability in complex blurring scenarios.

### 3.3. Qualitative Comparative Analysis

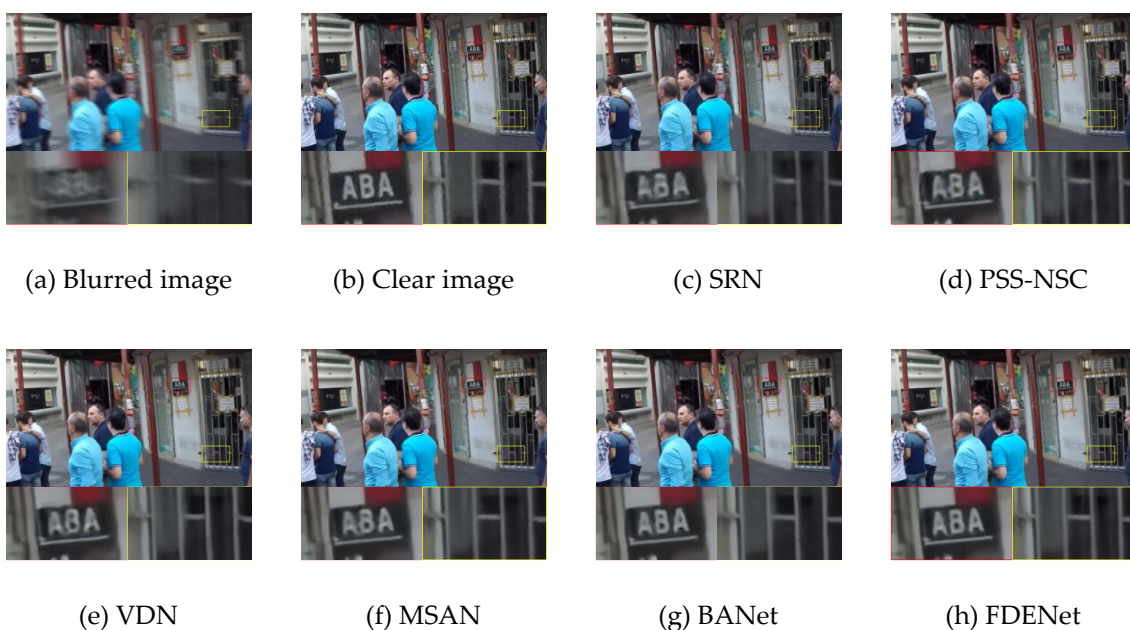
#### (1) Benchmark dataset

To demonstrate the effectiveness of the proposed FDENet, a qualitative analysis was conducted by comparing it with existing mainstream image deblurring algorithms. Figures 6 and 7 visually present the comparative visual effects of the SRN, PSSNSC, VDN, MSAN, BANet, and FDENet on the GoPro dataset and the HIDE dataset, respectively.

In Figure 6, residual blurring still exists after processing via using the SRN and MSAN, which results in a certain degree of deformation in the letter structure. The image processed by the BANet appears overly smooth in the edge areas of the letters. In contrast, the proposed FDENet performs better in restoring clear edges and fine details. In Figure 7, the SRN still exhibits severe blurring artifacts, making it impossible to identify specific text information. The image processed by the PSSNSC also suffers from color distortion. The proposed FDENet restores letters and faces more clearly without severe artifacts, which are closer to the actual clear image. Through frequency domain feature enhancement, the proposed FDENet effectively removes blurring while preserving the texture and structure of the original image, and making the deblurred image more natural and realistic.

#### (2) Actual shot motion-blurred image

The blurred image in Figure 8 was captured by a smartphone with a fixed camera position while a car was moving rapidly. Such blurs are extremely common in real-life scenarios. To verify the deblurring performance on real-world blurred images, qualitative comparative experiments were conducted between the proposed FDENet and different algorithms. As can be seen from the Figure 8, the SRN and VDN have poor deblurring effects in practical applications, and results in a very blurry restored image. The PSSNSC and DMPHN have shown some improvement in deblurring effects, but they still fail to effectively restore image details, and the license plate in the Figure 8 remains blurry. The proposed FDENet can obtain clearer image details for blurred images caused by fast-moving objects, and the effect is closer to the real world.



**Figure 6.** Visual comparison of different algorithms on GoPro dataset

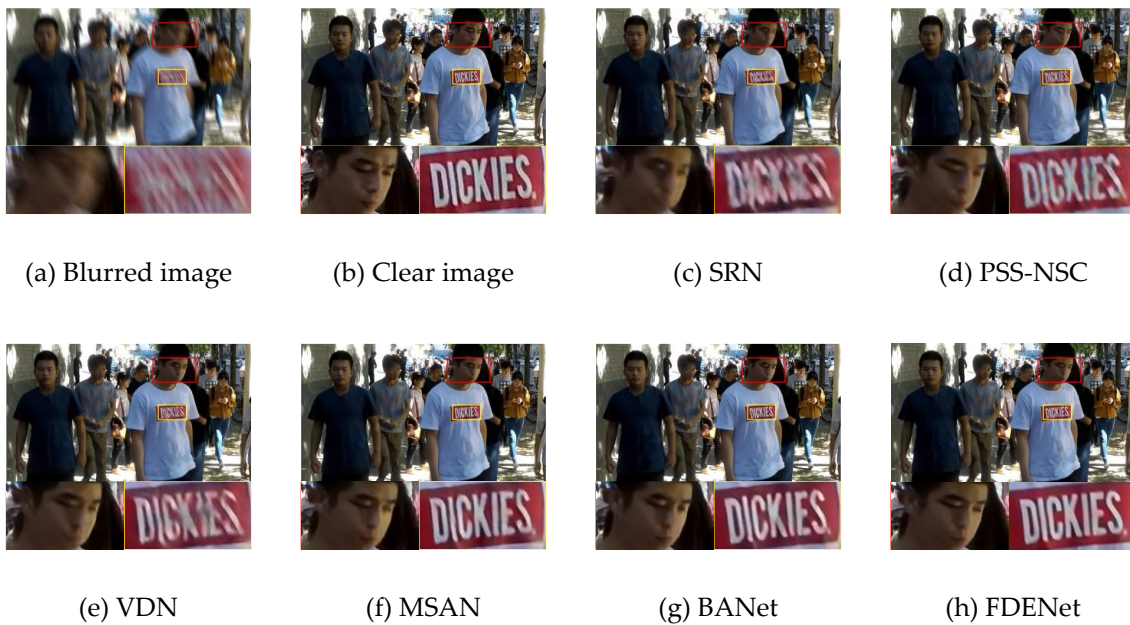


Figure 7. Visual comparison of different algorithms on HIDE dataset

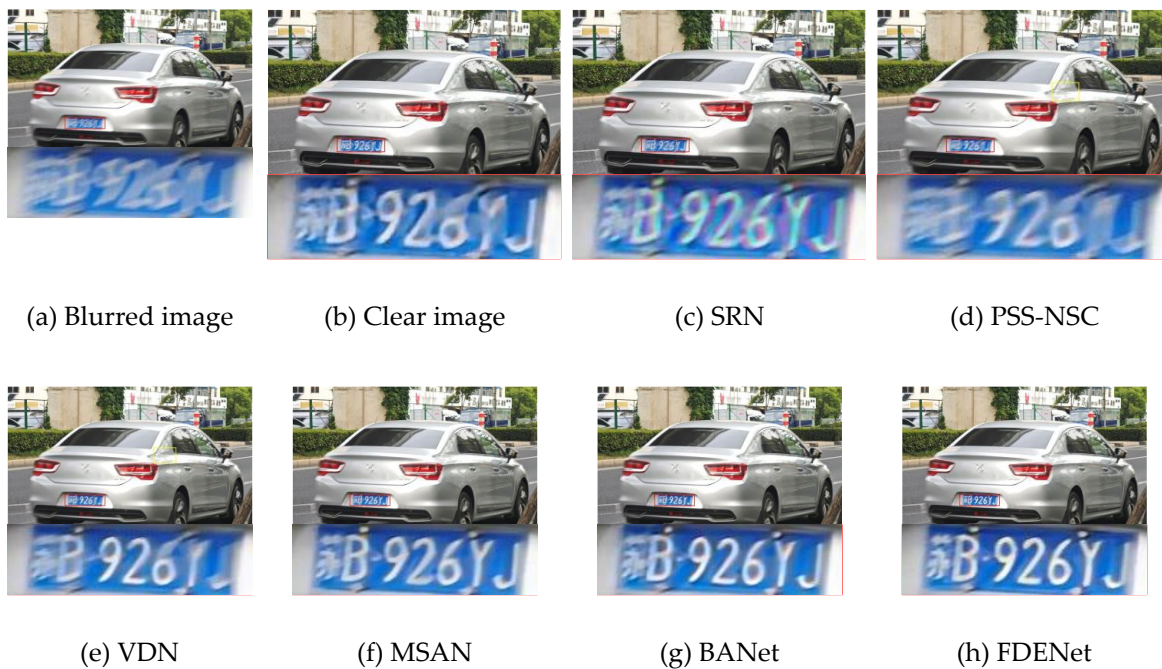


Figure 8. Visual comparison of different algorithms on real blurred images

### 3.4. Impact on Subsequent Visual Task

Image deblurring, as a low-level visual task, can be applied to the preprocessing stage of advanced computer vision applications such as object detection. When the input to the detection algorithm is a blurred image, it can lead to false positives and missed detections during the detection process, thereby this will affect detection accuracy. In this section, in order to verify the effectiveness and feasibility of the proposed FDENet in processing blurred images, the verification experiments based on the YOLOv7 object detection algorithm were conducted. As shown in Figure 9, the visualization results of the object detection for blurred images, the SRN-processed images, and the proposed FDENet-processed images are presented, respectively. As can be seen from the Figure 9, blurred images seriously affect the effectiveness of the object detection, with obvious missed detections. The deblurring results of the proposed FDENet retain the texture and edge details in the image, thus playing an important role in improving detection accuracy. Experimental results show that the proposed FDENet enhances the removal effect of blur on image texture details through frequency domain feature enhancement, and can improve the performance and robustness of subsequent visual tasks.

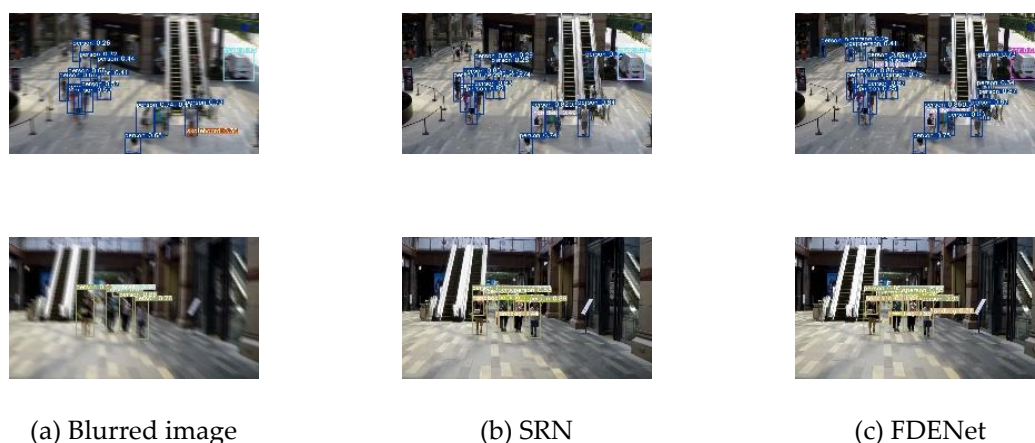


Figure 9. Comparison of results of target detection

### 3.5. Runtime and Parameter Analyses

The running time and model parameter counts of the algorithm in this paper on the GoPro testing set with a resolution of 1280×720 pixels are shown in Table 2. To enhance comparability, the parameter counts and runtimes of the models such as the SRN, DMPHN, MSAN, and MPRNet were also compared.

As shown in the Table 2, the parameter count of the SRN is  $6.8 \times 10^6$  with a runtime of 0.814 s. The parameter count of the MPRNet is  $20.1 \times 10^6$  with a processing time of 1.002s. The parameter count of the proposed FDENet is  $9.15 \times 10^6$  with a running time of 0.347s. Among all models, the SRN has a relatively small parameter and runtime. The MPRNet achieves excellent performance by stacking multi-stage networks, but this comes with an increase in parameters and runtime. The proposed FDENet maintains high speed while having relatively small parameter count.

Table 2. Runtime and parameter comparison

Model	Runtime/s	Params
SRN	0.814	$6.8 \times 10^6$
DMPHN	0.424	$21.7 \times 10^6$
MSAN	0.419	$35.9 \times 10^6$
MPRNet	1.002	$20.1 \times 10^6$
FDENet	0.347	$9.15 \times 10^6$

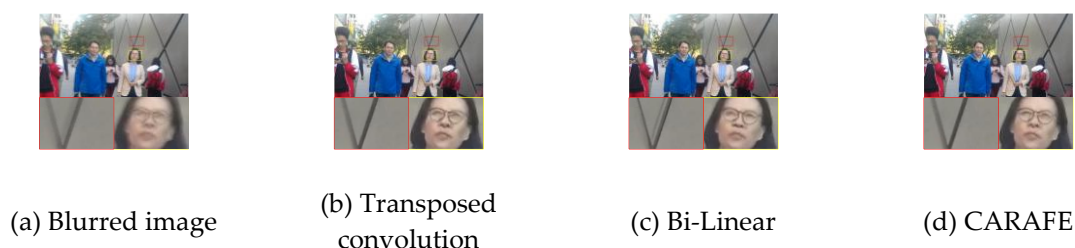
### 3.6. Ablation Experiment

To verify the effectiveness of upsampling method of the CARAFE in the proposed FDENet, it was replaced with the commonly used transposed convolution and bilinear interpolation, respectively, and comparison experiments in performance were conducted. The comparison results are shown in Table 3. When using the CARAFE as the upsampling module, the PSNR has an improvement of about 0.14 dB compared to transposed convolution. Compared to bilinear interpolation method, the PSNR also has an improvement of about 0.08dB. Additionally, the CARAFE introduces only a small number of parameters and computational costs. Therefore, the CARAFE was selected as the upsampling method in the proposed FDENet.

**Table 3.** Ablation experiments with different upsampling methods

Upsampling method	PSNR	SSIM
Transposed convolution	32.71	0.958
Bi-linear	32.77	0.959
CARAFE	32.85	0.961

Figure 10 shows a comparison of the visual effects of three upsampling methods, namely transposed convolution, bilinear interpolation, and the CARAFE, in image deblurring tasks. All three methods demonstrate excellent overall deblurring effects, but there are significant differences in their handling of details. The CARAFE can enhance feature learning capabilities by expanding the local receptive field, so it will result in clearer detail restoration and effectively reduces the generation of artifacts.



**Figure 10.** Visual comparison of different up-sampling methods

Table 4 presents the ablation experiment results of each module on the proposed FDENet model using the GoPro testing dataset. As shown in the Table 4, The introducing GFU alone does not significantly improve the performance of model, with only an improvement of about 0.04dB in PSNR, this indicates a weak enhancement effect on deblurring performance in the absence of other structural synergies. When the FFT-Res and GFU work together, the PSNR has an improvement of about 0.83 dB compared to the ordinary residual. The modeling of frequency domain information and the nonlinear enhancement of features through the FE Block effectively enhance the ability of network to model image details and blur structures. The experimental results have verified the effectiveness of each module in improving deblurring performance. Furthermore, under the synergistic effect of various modules, the restoration capability of model in complex dynamic blur scenarios is further enhanced.

**Table 4.** Results of ablation experiments

FFT-Res	GFU	SAB	PSNR	SSIM
×	×	×	31.78	0.951
×	√	×	31.82	0.953
√	×	×	32.49	0.955

√	√	×	32.61	0.956
√	×	√	32.74	0.958
√	√	√	32.85	0.961

To intuitively validate the effectiveness of the FFT-Res in proposed FDENet, it is visually compared with a model using a standard residual structure. As shown in Figure 11, the model using the FFT-Res can achieve better deblurring effects for text edges and structural components in the processed images. By enhancing the ability of model to learn frequency-domain features, it can further improve its capacity to represent global image information and high/low-frequency feature information, thereby demonstrating superior performance in texture detail restoration.



**Figure 11.** Visual comparison of different residual methods

#### 4. Discussion

Aim to the problem that existing image deblurring models often suffer from the loss of detailed structure and texture information in the restored images, Frequency Domain Feature Enhanced convolutional neural Network for image deblurring(FDENet) is proposed by incorporating Fast Fourier Transform(FFT). The learning of spatial and frequency domain feature information is realized by the frequency domain enhancement module, the spatial domain features are learned by the inverted residual block; The frequency domain feature is captured by the FFT residual block, and the gated feed-forward unit is responsible for the enhancement of the feature mapping capability. Effective propagation within the network is realized by the supervised attention module. The network achieved a PSNR of 32.85 dB and 30.93 dB, and a SSIM of 0.961 and 0.941 on the GoPro and HIDE datasets, respectively. The experiment results have demonstrated that the proposed FDENet, by combining with frequency-domain feature enhancement, outperforms current image deblurring algorithms in improving the quality of blurred images, and has robustness and generalization capabilities.

**Author Contributions:** Conceptualization, Guo.Yecai. and Ma.Lixiang.; methodology and software, Guo.Yecai. and Ma.Lixiang.; validation, Guo.Yecai.,Ma.Lixiang. and Zhang.Yangyang.; formal analysis, Ma.Lixiang.; investigation, Guo.Yecai.; resources, Guo.Yecai.; data curation, Guo.Yecai.; writing—original draft preparation, Ma.Lixiang.; writing—review and editing, Zhang.Yangyang.; supervision, Zhang.Yangyang.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research work was funded by the Open Project of the State Key Laboratory of Millimeter Wave, Project No.:K202530

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Wu Y F. The Medical Image Fusion Algorithm Based on Fuzzy Transformation and Maximum Entropy [J]. *Journal of Southwest China Normal University (Natural Science Edition)*, 2018, 43(11): 49-56.
2. Zhang W B, Jiang Z. Research on deblurring of remote sensing images based on deep learning [J]. *Journal of Yunnan Minzu University (Natural Sciences Edition)*, 2023, 32(4), 480-484.
3. Li Y, An H, Zhang T, et al. Omni-Deblurring: Capturing Omni-Range Context for Image Deblurring [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(8): 7543-53.
4. Liang P, Jiang J, Liu X, et al. Decoupling Image Deblurring Into Twofold: A Hierarchical Model for Defocus Deblurring [J]. *IEEE Transactions on Computational Imaging*, 2024, 10: 1207-20.
5. Zhang J, Zhang Y, Yang F, et al. Improving the deblurring method of D2Net network for infrared videos [J]. *Journal of Electronic Imaging*, 2024, 33(4).
6. Zzhao S, Zhang Z, Wei Y, et al. Individual/joint deblurring and low-light image enhancement in one go via unsupervised deblurring paradigm [J]. *Science China-Information Sciences*, 2025, 68(12).
7. Tao X, Gao H, Wang J, et al. Scale-recurrent network for deep image deblurring[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 8174-8180.
8. Gao H, Tao X, Shen X, et al. Dynamic scene deblurring with parameter selective sharing and nested skip connections[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 3848-3856.
9. Zhang H, Dai Y, Li H, et al. Deep stacked hierarchical multi-patch network for image deblurring[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 5978-5986.
10. Cho S J, Ji S W, Hong J P, et al. Rethinking coarse-to-fine approach in single image deblurring[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 4641-4650.
11. Ding X, Zhang Y, Ge Y, et al. UniRepLKNet: A universal perception large-kernel convnet for audio, video, point cloud, time-series and image recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 5518-5528.
12. Zhou Y, Zhou P, Ng T K. Efficient cascaded multiscale adaptive network for image restoration[C]//*European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024: 92-110.
13. Cui Y, Ren W, Cao X, et al. Selective frequency network for image restoration[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023: 12549-12559.
14. Zou W, Jiang M, Zhang Y, et al. Sdwnet: A straight dilated network with wavelet transformation for image deblurring[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021: 1895-1904.
15. Huang X, He J S. Fusing convolution and self-attention parallel in frequency domain for image deblurring[J]. *Neural Processing Letters*, 2023, 55(7): 9811-9829.
16. Wang J, Chen K, Xu R, et al. Carafe: Content-aware reassembly of features[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 3007-3016.
17. Zamir S W, Arora A, Khan S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022: 5728-5739.
18. Shen Z, Wang W, Lu X, et al. Human-aware motion deblurring[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 5572-5581.
19. Neji H, Ben Halima M, Noguera-iso J, et al. Deep architecture for super-resolution and deblurring of text images [J]. *Multimedia Tools and Applications*, 2023.
20. Guo C, Wang Q, Dai H N, et al. VDN: Variant-depth network for motion deblurring[J]. *Computer Animation and Virtual Worlds*, 2022, 33(3-4): e2066.

21. Guo C, Chen X N, Chen Y H, et al. Multi-stage attentive network for motion deblurring via binary cross-entropy loss[J]. *Entropy*, 2022, 24(10):1-14.
22. Tsai F J, Peng Y T, Tsai C C, et al. Banet: a blur-aware attention network for dynamic scene deblurring[J]. *IEEE Transactions on Image Processing*, 2022, 31: 6789-6799.
23. Zamir S W, Arora A, Khan S, et al. Multi-stage progressive image restoration[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021: 14821-14831.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.