forests

MDPI

Article

Forest Fire Image Deblurring Based on Spatial–Frequency Domain Fusion

Xueyi Kong [1], Yunfei Liu [1,*], Ruipeng Han [1], Shuang Li [1] and Han Liu [2]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China; kxy@njfu.edu.cn (X.K.); hrp@njfu.edu.cn (R.H.); lishuang@njfu.edu.cn (S.L.)
[2] College of Letters and Science, University of Wisconsin Madison, Madison, WI 53706, USA; hliu568@wisc.edu
* Correspondence: lyf@njfu.com.cn; Tel.: +86-139-1389-5117

Abstract: UAVs are commonly used in forest fire detection, but the captured fire images often suffer from blurring due to the rapid motion between the airborne camera and the fire target. In this study, a multi-input, multi-output U-Net architecture that combines spatial domain and frequency domain information is proposed for image deblurring. The architecture includes a multi-branch dilated convolution attention residual module in the encoder to enhance receptive fields and address local features and texture detail limitations. A feature-fusion module integrating spatial frequency domains is also included in the skip connection structure to reduce feature loss and enhance deblurring performance. Additionally, a multi-channel convolution attention residual module in the decoders improves the reconstruction of local and contextual information. A weighted loss function is utilized to enhance network stability and generalization. Experimental results demonstrate that the proposed model outperforms popular models in terms of subjective perception and quantitative evaluation, achieving a PSNR of 32.26 dB, SSIM of 0.955, LGF of 10.93, and SMD of 34.31 on the self-built forest fire datasets and reaching 86% of the optimal PSNR and 87% of the optimal SSIM. In experiments without reference images, the model performs well in terms of LGF and SMD. The results obtained by this model are superior to the currently popular SRN and MPRNet models.

Keywords: forest fire; image deblurring; MIMO-UNet; dilated convolution; spatial–frequency domain fusion

Citation: Kong, X.; Liu, Y.; Han, R.; Li, S.; Liu, H. Forest Fire Image Deblurring Based on Spatial–Frequency Domain Fusion. *Forests* **2024**, *15*, 1030. https://doi.org/10.3390/f15061030

Academic Editor: Luis A. Ruiz

Received: 8 May 2024
Revised: 5 June 2024
Accepted: 11 June 2024
Published: 13 June 2024

1. Introduction

Forests are a vital resource on Earth, playing a crucial role in air purification, noise reduction, soil and water conservation, natural oxygen production, and climate regulation [1]. However, forest fires, as frequent natural disasters, not only consume trees and other forest resources but also pose serious threats to humans, animals [2], and the environment. Therefore, early detection and mitigation of forest fires is essential [3]. In recent years, the combination of deep learning techniques with UAV imagery has shown great potential for advancing forest fire identification [4]. Aerial imagery technology, particularly the use of UAVs equipped with optical cameras, has emerged as an important tool for wildfire prevention. These UAVs are capable of real-time monitoring and have gained popularity due to their versatility [5], high speed, and persistence. Their ability to integrate images from different flight altitudes enables wider coverage and the production of detailed images, making them the preferred choice for wildfire monitoring [6]. However, the airborne camera is susceptible to interference from relative motion, attitude changes, atmospheric turbulence and other factors, resulting in motion blur in captured images [7]. This significantly reduces the visibility of forest fires and the accuracy of feature detection, segmentation, and object recognition processes. Thus, research on forest fire image deblurring based on deep learning, combined with UAV imagery, holds significant potential for advancing forest fire recognition [8].

The image deblurring technology has been extensively studied with systematic and mature theories and methods. Image deblurring methods can be categorized into blind deblurring [9–12] and non-blind deblurring [13–16], depending on whether the blur kernel is unknown or known. Traditional image deblurring methods exhibit certain limitations when applied to the practical task of forest fire image deblurring. These drawbacks include the requirement for a significant amount of prior knowledge, the production of low-quality restored images, and the tendency to introduce ringing artifacts. In response to these challenges, many researchers have turned to deep learning techniques for image deblurring. For instance, Schuler et al. [17] employed a deep neural network to estimate the depth features of a blurred image, subsequently transforming these features into the frequency domain to estimate the blurring kernel. This approach allowed for non-blind deblurring using traditional methods. Similarly, Li et al. [18] utilized image a priori information as a dichotomous classifier, trained by a deep convolutional neural network, to achieve image recovery. Despite their potential, these methods are constrained by the accuracy of blurring kernel estimation and exhibit low execution efficiency [19]. While effective for specific scene models, they lack robust generalization capabilities for addressing more challenging real-world scene deblurring tasks [20].

Recently, the deep learning community has shifted its focus to exploring end-to-end blind motion image deblurring strategies [21,22], bypassing explicit blur kernel computation by directly mapping blurred to clear images. Nah et al. [23] first introduced the use of the DeepDeblur algorithm to confront the challenge of blurred images in dynamic scenes. Drawing on the concept of progressing from coarse to fine details, their deep convolutional neural network was intricately designed to operate across multiple scales. While the DeepDeblur algorithm significantly enhances image deblurring capabilities, it is characterized by an exceedingly large number of model parameters [24]. To tackle this issue, Tao et al. [25] proposed a Scale Recurrent Network (SRN) capable of substantially reducing computation time by sharing network weights across different scales. Furthermore, Zhang et al. [22] specialized a spatial pyramid-based multilayer network (DMPHN), which focuses on utilizing cuts instead of downsampling and employing feature map cascading in the encoder-decoder process, leading to a drastic reduction in the amount of model computation. KUPYN et al. [26] successively proposed two deep network models by introducing generative antigrid, DeBlurGAN and DeBlurGANv2, and used the generation ability of generator adversarial networks (GANs) to restore high-quality clear images. On this foundation, Zhang et al. [27] innovated by integrating two GANs. The blur GAN (BGAN) is utilized to generate images that closely resemble real motion blur, while the deblurring GAN (DBGAN) is employed to learn the process of recovering blurred images. Cho et al. [28] proposed a multi-input, multi-output U-Net network and introduced asymmetric feature fusion to effectively merge multi-scale features and gradually improve image clarity from the lower subnet to the upper subnet.

Despite image deblurring algorithms having made significant progress on mainstream datasets, it is still challenging to restore real-world blurred images into clear ones.

- Most current methods use an encoder–decoder structure to learn the features of different receptive fields [29]. However, using many up-sampling and down-sampling will lead to the loss of texture details, which seriously affects image restoration.
- At present, some image deblurring methods use GAN to obtain realistic texture details, but this method will lead to unstable network performance.
- Most deblurring methods do not distinguish the feature information from different spatial and frequency domains [30], resulting in a poor deblurring effect and other problems.

To address the aforementioned issues, this study proposes a forest fire image deblurring model based on the MIMO-UNet algorithm. This model effectively reduces motion blur in UAV images which are captured during forest fire monitoring. The key contributions of this study are as follows:

- We propose a multi-branch dilation convolution to enhance the focus of the residual block. By employing dilated convolutions with different dilated factors, we are able to capture features from various receptive fields. The integration of the residual block with the parallel attention block improved the network's ability to process multi-scale features.
- To further enhance the deblurring effect, we devise a spatial–frequency domain fusion module. This module not only extracts the information in the spatial and frequency domains, but also effectively combines them to reduce information loss.
- We propose a multi-channel convolutional attention residual module, which efficiently captures image details and context information by processing features of different scales in parallel. This approach effectively addresses information loss and insufficient reconstruction quality in the decoder.
- To improve the generalization performance of the model, this study proposes a weighted loss function which contains multi-scale content loss, multi-scale high-frequency information loss and multi-scale structure loss. In this way, the internal texture details of the image and the lost high-frequency information can be recovered, and the deblurring effect can be enhanced comprehensively.

The rest of the paper is organized as follows: Section 2 describes our dataset, and the overall network architecture is selected. Section 3 presents the experimental results and performance analysis, The discussion is provided in Section 4, and finally, Section 5 concludes this paper.

2. Materials and Methods

2.1. Datasets

The blurred and clear images used for training and testing are the basis for the deblurring study. The blurred image and the clear image must be geometrically aligned; both images should be taken at the same camera position. This is difficult because the camera has to be shaken to get a blurry image. In fact, it is hard to obtain motion-blurred image datasets. Current public datasets in the field of motion blur image restoration include GoPro [31], Lai, and Kohler datasets.

In this study, we created a forest fire dataset containing motion-blurred images, inspired by the generation method of GoPro and Lai datasets in the field of deblurring. To simulate the Lai dataset creation method, we used 20,000 clear forest fire images and applied a linear motion-blurred kernel and random noise to them to generate blurred images. In addition, following the idea of creating the GoPro dataset, we recorded the forest scene and treated each frame as a clear image. By averaging the clear image of consecutive frames, we obtained the blurred image and formed clear and blurred image pairs. The forest fire dataset that we created included a training set of 3000 pairs and a test set of 1000 pairs for deblurring image testing. Figure 1 shows the forest fire image pairs in different scenarios.

Figure 1. The forest fire image pairs in different scenarios.

2.2. Overall Network Architecture

Image deblurring networks such as U-Net [32] and DeblurGan are known for their good deblurring performance, but these networks also have certain drawbacks and will be limited in practical application scenarios. This study is to develop a new deblurring network, which can effectively improve the deblurring performance. MIMO-UNet (multi-input multi-output UNet) is an improved and extended network based on U-Net structure, which can better deal with complex image blurring by introducing multi-module and multi-input and multi-output channels. On the public datasets, MIMO-UNet performs well in image deblurring, showing high accuracy and robustness in image restoration and enhancement tasks. Therefore, we selected MIMO-UNet as our base network architecture.

3. The Proposed Method

In this section, we will describe the structure of the network and elaborate the details of the preprocessing module. There are mainly the multi-branch dilated convolution attention residual module in the encoder module, the spatial–frequency domain fusion module, and the multi-channel convolution attention residual module in the decoder module.

3.1. Structural Description of the Network

The network proposed in this study adopts multi-scale input and output with coarse-to-fine structure strategy. We divide the network into four parts: a preprocessing module (PM) for shallow feature extraction, an encoder module (EM) for deep information extraction, a spatial–frequency domain fusion module (SFFM), and a decoder module (DM) for image restoration and reconstruction, where B_k (k = 1, 2, 3) represents the input blurred image with multi-scale and S_k (k = 1, 2, 3) represents the output restored image with multi-scale, as shown in Figure 2.

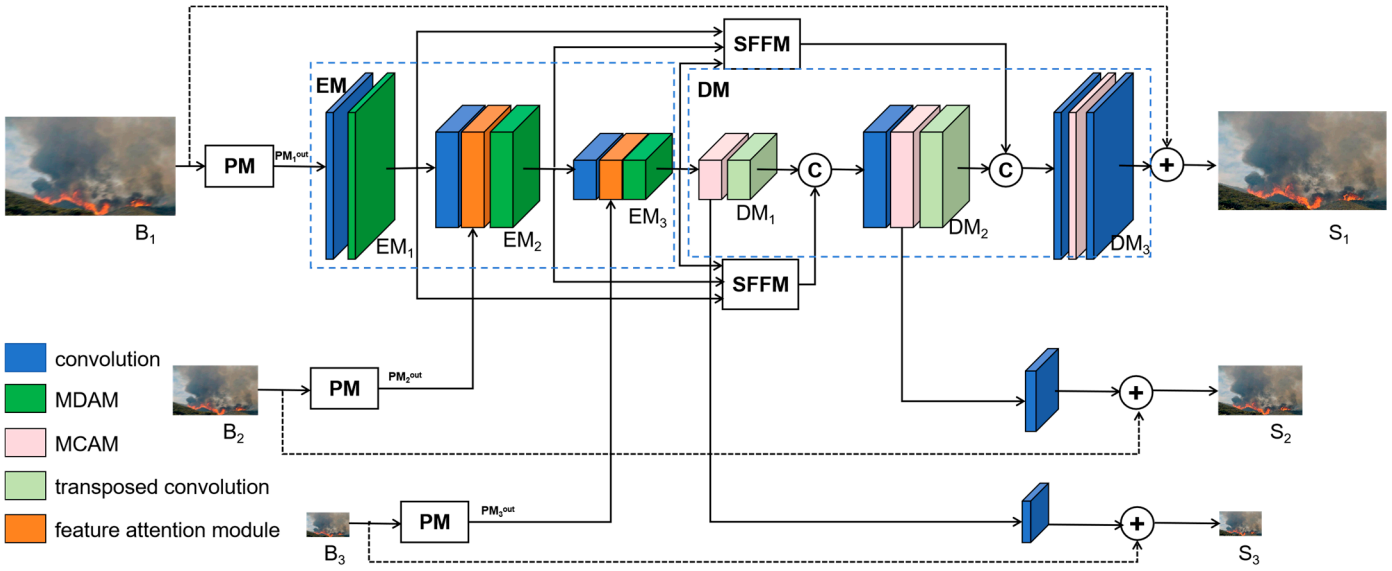

Figure 2. The structure of the proposed model.

EM consists of three sub-encoder modules, i.e., EM_1, EM_2, and EM_3, and it is used to extract blurred images at different scales. Initially, the input blurred image is scaled to obtain three blurry images with different scales and resolutions, namely B_1, B_2, and B_3. Then, EM employs multi-branch dilated convolution attention residual modules (MDAMs) for feature extraction, enhancing the capture of detailed information. In EM_2 and EM_3, feature fusion is optimized using the feature attention module, which integrates features EM_k^{out} (k = 1, 2) and PM_k^{out} (k = 2, 3). The SFFM merges features across spatial and frequency domains of different encoder scales before passing them to the decoder, thereby improving feature utilization and reducing information loss. DM consists of DM_1, DM_2,

and DM_3. The inputs of DM_1 and DM_2 are the result of the fusion of the decoder output of the previous layer with the SFFM module. In addition, we designed a multi-channel convolution attention residual module (MCAM) in the DM to extract multi-scale feature information.

3.2. Preprocessing Module

Due to the continuous smoothness and sparsity of the motion-blurred images, it is essential to employ receptive fields of varying sizes for effective feature extraction. To address this problem, PM uses multiple convolution modules connected in series and parallel for shallow feature extraction before EM, where different receptive fields are captured using different convolution kernel sizes, namely 3×3 and effective 5×5 (achieved by two cascaded 3×3 convolutions). Then, 1×1 convolutions are used to subtly integrate these extracted features. This not only simplifies the output channel but also enhances the effectiveness of back propagation while mitigating the risk of vanishing gradients. Integrating local connections that jump between the input and output layers ensures accurate synthesis of the final output, as shown in Figure 3.

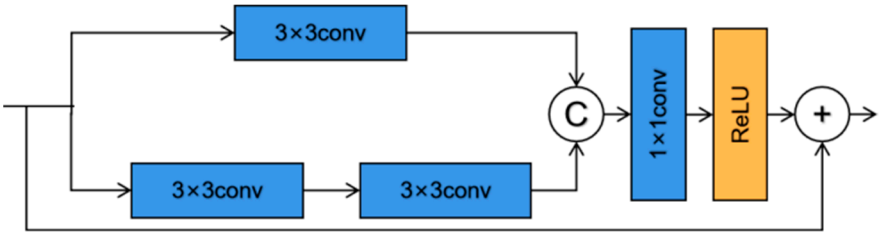

Figure 3. Preprocessing Module.

3.3. The Multi-Branch Dilated Convolution Attention Residual Module

Residual modules in deep neural networks often overlook image blur caused by limited receptive fields, leading to an irreversible loss of resolution and edge details. To address this problem, our study proposes an effective multi-branch dilated convolution attention residual module (MDAM) in EM, in which dilated convolution uses a multi-branch structure to further enhance image feature expression [33], thereby mitigating blur and preserving feature information. Multiple MDAM can be interconnected to realize feature reuse and maximize the utilization feature information. This module comprises a multi-branch dilated convolution residual module (MDCM) and a parallel attention module (AM), as shown in Figure 4.

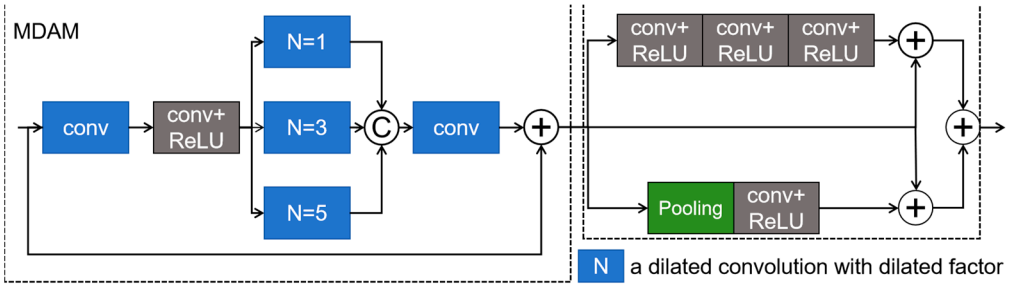

Figure 4. The multi-branch dilated convolution attention residual module.

MDCM consists of two convolution blocks and a dilated convolution block. The dilated convolution block is composed of multiple dilated convolutions with different dilated rates in parallel, which can be expressed as:

$$F_{\mathrm{re}} = \delta\left(H_{\mathrm{conv}}\left(H_{\mathrm{conv}}\left(F_{\mathrm{input}}\right)\right)\right) \quad (1)$$

$$F_{\mathrm{dr\_1}} = \delta\left(H_{dr\_1}\left(F_{\mathrm{re}}\right)\right) \quad (2)$$

$$F_{\text{dr\_3}} = \delta(H_{\text{dr\_3}}(F_{\text{re}})) \tag{3}$$

$$F_{\text{dr\_5}} = \delta(H_{\text{dr\_5}}(F_{\text{re}})) \tag{4}$$

$$F_{dr\_cat} = Concat\left(F_{dr\_1}, F_{dr\_3}, F_{dr\_5}\right) \tag{5}$$

where H_{dr_1}, $H_{\text{dr\_3}}$, and $H_{\text{dr\_5}}$ represent the dilated convolution with the dilated factor of 1, 3, 5, respectively, δ represents the ReLU activation function, and $F_{\text{dr\_1}}$, $F_{\text{dr\_3}}$, and $F_{\text{dr\_5}}$ represent the output of the dilated convolution with different dilated factors.

The dilated convolution network introduces the "dilation rate" parameter into the traditional convolution operation, so that the sampling points inside the convolution kernel are no longer continuous, but sampled at certain intervals, so as to expand the receptive field. The dilation rate determines the interval at which the convolution kernel performs the sampling. A larger dilation rate allows the convolution kernel to span a larger area, thus expanding the receptive field.

In the last layer of the dilated convolution module, a dilated convolution with a dilated factor of 1 combines features from different receptive fields, while a 1×1 convolution reduces the number of channels for integration. Finally, we superimpose the fused features onto the input features to get the output. The output characteristics can be written as:

$$F_{\text{conv}} = H_{\text{conv}}(F_{\text{dr}\_cat}) \tag{6}$$

$$F_{\text{out}} = F_{conv} + F_{\text{input}} \tag{7}$$

H_{conv} represents the 1×1 convolution layers for information integration. F_{conv} and F_{out} represent convolution feature and output feature, respectively.

To address the issue of non-uniform blur distribution in images and to leverage the varying importance of information across different spatial and channel dimensions, a novel parallel attention module (AM) is proposed. This module encompasses three distinctive branches: one specifically designed for spatial attention, another for preserving the original image features, and a third branch dedicated to implementing channel attention.

The spatial attention mechanism consists of three cascaded modules composed of a convolution layer and an activation layer, and the channel attention mechanism consists of one pooling layer, one activation layer, and one convolution layer, utilizing a 3×3 convolution kernel. The formula is as follows:

$$A_s = \delta(c\text{onv}(\delta(c\text{onv}(\delta(c\text{onv}(F)))))) \tag{8}$$

$$A_c = \delta(c\text{onv}(P\text{ool}(F))) \tag{9}$$

$$A_p = (A_s * F) + (A_c * F) \tag{10}$$

where F is the input feature of the parallel attention module, δ represents the ReLU activation function, *conv* represents the convolution, *P*ool represents the pooling operation, A_s is the output of the spatial attention mechanism, A_c is the output of the channel attention mechanism, and A_p is the final output feature of the parallel attention module.

3.4. Spatial–Frequency Domain Fusion Module

In the traditional U-Net architecture, skip connections directly transfer encoder features to the decoder, which will lead to the decoder not being able to make full use of the multi-scale features generated in EM. Furthermore, based on the multi-scale frequency reconstruction (MSFR) loss function of MIMO-UNet to recover reduced high-frequency elements, this study designs a novel multi-scale feature fusion module within the skip connections, referred to as the Spatial and Frequency Feature Module (SFFM). The EM outputs three multi-scale features, each divided into two branches. One branch undergoes 2D real-time fast Fourier transform, followed by feature extraction in the frequency domain with 3×3 convolution and ReLU activation. The other branch conducts feature extraction

in the spatial domain using 3 × 3 convolution and ReLU activation. These branches are then fused, resized for combining different scale features, and fed into the DM, as shown in Figure 5.

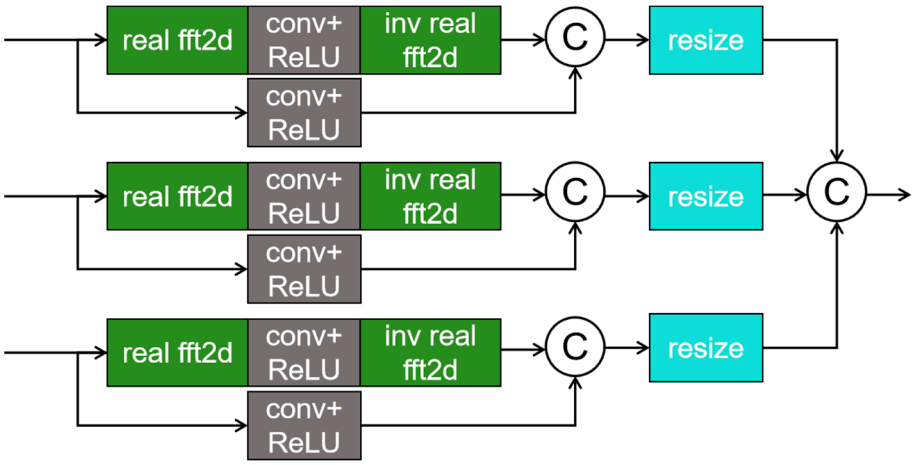

Figure 5. Spatial–frequency domain fusion module.

3.5. Multi-Channel Convolution Attention Residual Module

The basic task of the decoder is to reconstruct a clear, high-quality image from the feature representation. However, the problems, such as information loss and suboptimal reconstruction quality, often occur with conventional decoder modules. To mitigate these challenges, we integrate an innovative multi-channel convolutional attention residual module into DM, as shown in Figure 6, which effectively captures image details and context information by processing features of different scales in parallel through a multi-channel structure.

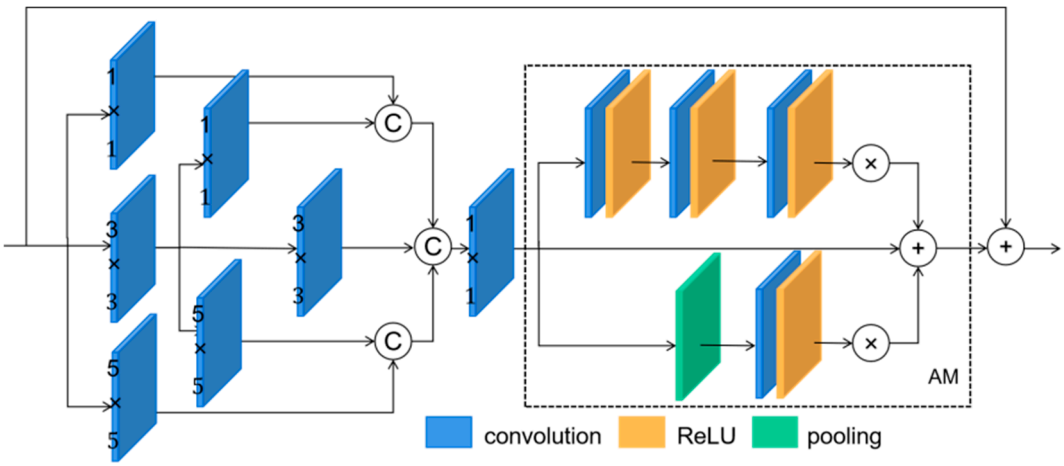

Figure 6. Multi-channel convolution attention residual module.

Each channel consists of a different convolution layer, which is specifically used to extract the features of the corresponding layer. This approach ensures that irrelevant information is eliminated, enabling the extraction of deeper and more comprehensive information through the parallel attention module, as in the encoder. This method, which combines multi-channel feature extraction and attention mechanisms, not only significantly reduces the information loss in the reconstruction process, but also improves the overall quality of the reconstructed image.

3.6. Loss Function

Based on the structure from coarse to fine, the whole model is divided into three stages, and each stage can output the restored image. During deblurring, there are losses in the space and frequency domains as well as structural losses due to the unstable model training. Therefore, the weighted loss strategy based on multi-scale content loss, multi-scale high-frequency information loss, and multi-scale structure loss [34] is adopted to ensure supervision and improve the blurring effect. It is assumed that S_k (k = 1, 2, 3) represents the output multi-scale restored image, and R_k (k = 1, 2, 3) represents the corresponding real clear image.

1. Multi-scale content loss

The L1 distance between the real clear picture of different scales and the model restoration map is used as the multi-scale content loss:

$$L_c = \sum_{k=1}^{3} \|R_k - S_k\|_1 \tag{11}$$

The L1 distance does not excessively punish large error values, which is conducive to preserving the edge features of the image.

2. Multi-scale high-frequency information loss

This study utilizes Fast Fourier Transform (FFT) to quantify the high-frequency information loss between the blurred image and the reference clear image:

$$L_F = \sum_{k=1}^{3} \|FFT(R_k) - FFT(S_k)\| \tag{12}$$

3. Structural loss

Structural losses can be expressed as:

$$L_S = L_{SSIM} + L_{MS-SSIM} \tag{13}$$

L_{SSIM} can be expressed as:

$$L_{SSIM} = 1 - SSIM(P) \tag{14}$$

Multi-scale structural similarity loss ($MS - SSIM$) can be expressed as:

$$L_{MS-SSIM} = 1 - MS - SSIM(P) \tag{15}$$

P is the middle pixel value of the pixel block.

4. Total loss function

The formula used to calculate the total loss function is as follows:

$$L_{\text{total}} = \lambda_A L_A + \lambda_F L_F + \lambda_S L_S \tag{16}$$

where L_C means absolute error (MAE) and L_F is multi-scale frequency reconstruction (MSFR). L_S is structural loss (SL), where λ_C, λ_F, and λ_S are set 0.1, 0.01 and 0.08, respectively. The allocation of proportions to each loss is based on the variability of values obtained for each loss during the training process. Consequently, losses with lower volatility are assigned a smaller proportion of the model optimization impact. The weighting coefficients in the equations mentioned above are derived from this approach and determined experimentally.

4. Results

4.1. Training Parameters and Environment

The details of the model proposed in this paper and the training hyperparameters are shown in Tables 1 and 2. Based on our self-built forest fire dataset, several data augmentation methods are used, including random horizontal flipping, vertical flipping, and 90-degree rotation. We use the Adam optimizer [35] to optimize the network model; the training time is set to 800 epochs, and the initial learning rate is set to 0.001. The learning rate scheduling strategy is the cosine annealing strategy.

Table 1. Experimental environments.

Experimental Environments	Details
Program Language	Python 3.8
Framework	Pytorch 1.13.1
Operating System	Windows 11
GPU Type	Nvidia RTX 4060 (manufactured by Nvidia Corporation, Santa Clara, CA, USA)
Acceleration Tool	CUDA 11.6

Table 2. Training parameters.

Training Parameters	Details
Epochs	800
Batch Size	8
Learning Rate	1×10^{-3}
Optimizer	Adam
Betas	(0.9, 0.99)
Eps	1×10^{-8}

4.2. Evaluation Index

Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as evaluation indexes to verify the effectiveness of the proposed model. PSNR reflects the degree of distortion before and after image processing. The larger the value, the smaller the distortion and the better the deblurring effect. A higher SSIM indicates that the processed image is more similar to the original image structure information.

However, in practical situations, assessing the effectiveness of deblurring methods using PSNR and SSIM proves difficult because it is difficult to obtain paired blurred and correspondingly clear images at the same time. Therefore, auxiliary evaluation indexes such as Laplacian gradient function (LGF) [36] and the sum of modulus of gray difference (SMD) without reference images are introduced as complementary indices to facilitate the evaluation of the image definition and deblurring result. Their calculation formula is as follows:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [I(i,j) - K(i,j)]^2 \tag{17}$$

$$PSNR = 10log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{18}$$

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{19}$$

where μ_x and μ_y respectively represent the average value of x and y. σ_x and σ_y respectively represent the standard deviation of x and y, σ_{xy} represents the covariance of x and y, c_1 and c_2 are constant to avoid systematic error due to the zero numerator.

LGF is an image blur evaluation index without a reference image. The higher the value, the clearer the image. The formula of Laplacian is as follows:

$$D(f) = \sum_{y}\sum_{x}|G(x,y)| \quad G(x,y) > T \tag{20}$$

where $G(x,y)$ is the convolution of the Laplacian operator at pixel point (x,y).

When fully focused, the image is the clearest, and the high-frequency component in the image is also the most, so the gray change can be used as the basis for image clarity. The formula of SMD is as follows:

$$G = |f(x,y) - f(x+1,y)| + |f(x,y) - f(x,y+1)| \tag{21}$$

$$D_0 = MAX\sum_{x}\sum_{y} G \tag{22}$$

D_0 corresponds to the focus position.

4.2.1. Forest Fire Dataset Test

The quality of the image restored by the proposed method and the current popular deblurring method is evaluated. The validity of the model is assessed using the self-build forest fire dataset. Test results indicate that, relative to Table 3 and Figure 7 our model outperforms the other six models in PSNR, and the SSIM [37] is also better, except for being slightly lower than the MPRNet model. In addition, LGF and SMD are also superior to other models, indicating that they perform well in defining image edges and texture details.

Table 3. The performance comparison using the self-built forest fire dataset.

Model	PSNR	SSIM	LGF	SMD
Deepdeblur	29.04 dB	0.924	8.1	27.99
DeblurGAN-v2	29.30 dB	0.934	8.33	28.10
SRN	30.01 dB	0.941	8.96	30.05
MIMO-UNet	30.35 dB	0.936	8.98	30.13
SDWNet	31.65 dB	0.942	9.91	32.15
MPRNet [38]	32.04 dB	0.958	10.10	33.26
NAFNet	32.13 dB	0.951	10.50	33.90
Ours	32.26 dB	0.955	10.93	34.31

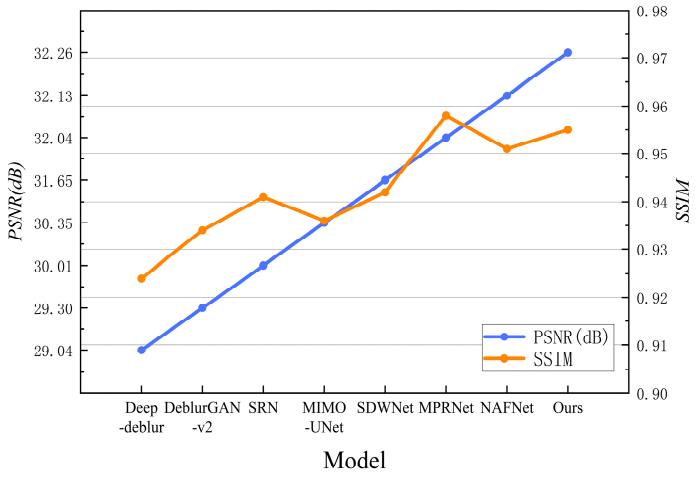

Figure 7. Comparison of deblurring effect in different models.

Upon comparing the results of each image within the dataset, we observed that our model achieved 86% of the optimal PSNR, 87% of the optimal SSIM, 84% of the optimal LGF,

and 90% of the optimal SMD. These outcomes collectively indicate the excellent deblurring effect achieved by our model.

We also analyzed the subjective effect of image deblurring. Figure 8 illustrates the subjective visual comparison before and after deblurring for diverse forest fire types, such as surface fires, crown fires, and trunk fires. It is obvious that the image texture of the proposed model is more clearly visible after deblurring. Furthermore, as showed in Table 4 and Figure 9, our model has demonstrated exceptional performance in handling diverse forest fire scenarios, and all four quantitative measures are optimal or suboptimal.

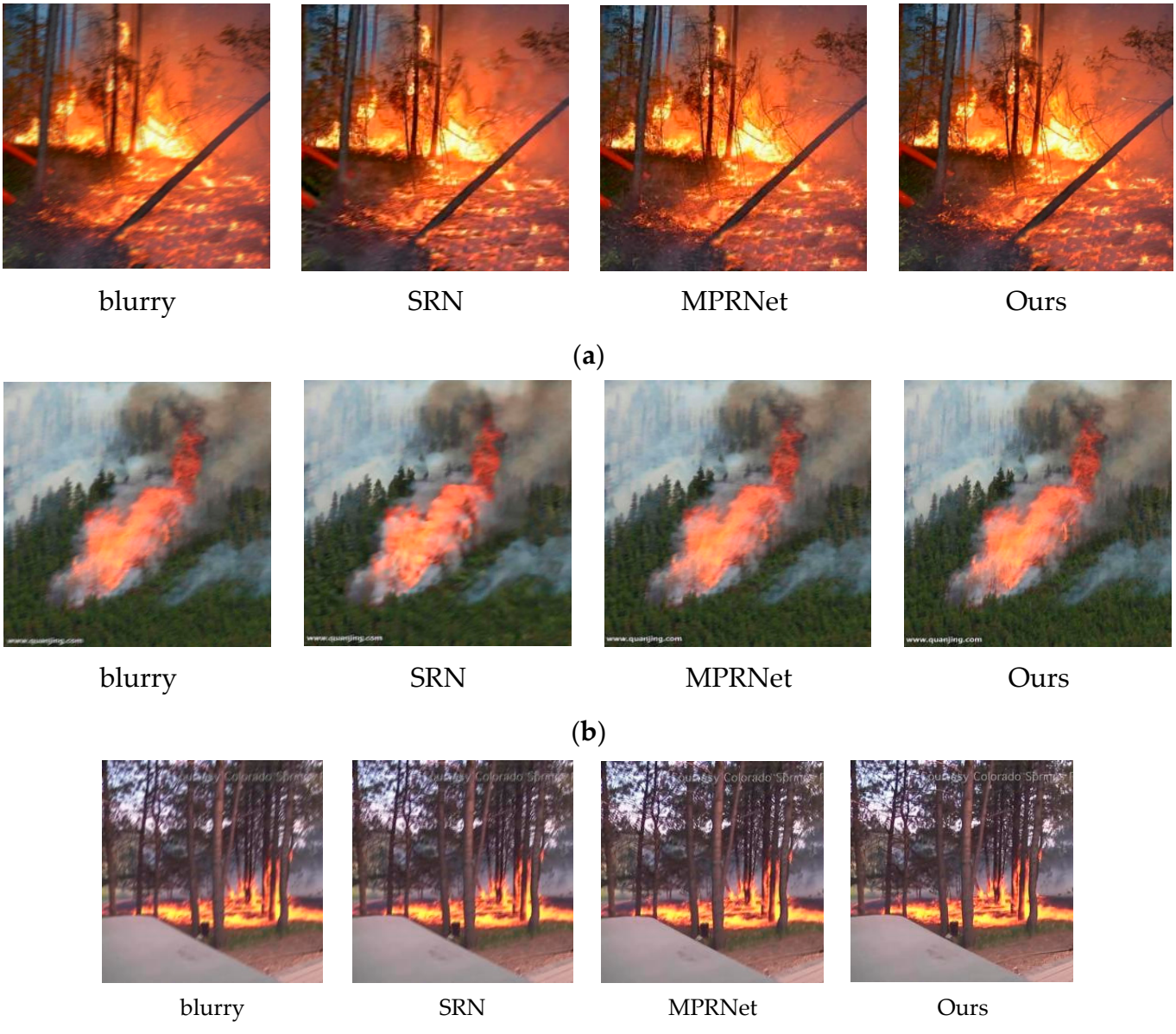

Figure 8. The subjective visual comparison results before and after deblurring. (**a**) surface fire; (**b**) crown fire; (**c**) trunk fire.

Table 4. Comparison of indicators under different forest fire scenarios.

Image	Model	PSNR	SSIM	LGF	SMD
Surface Fire	Blurry	26.06 dB	0.6671	8.17	21.19
	SRN	27.24 dB	0.7266	10.74	25.88
	MPRNet	29.06 dB	0.7863	11.33	27.36
	Ours	29.28 dB	0.7922	12.73	28.25
Crown Fire	Blurry	20.31 dB	0.8222	6.72	18.32
	SRN	21.14 dB	0.8427	7.37	17.73
	MPRNet	23.50 dB	0.8628	8.27	19.7
	Ours	25.49 dB	0.8523	9.03	22.76
Trunk Fire	Blurry	23.49 dB	0.8136	7.13	30.76
	SRN	24.49 dB	0.8369	12.25	32.17
	MPRNet	26.32 dB	0.8805	14.8	34.65
	Ours	26.44 dB	0.8839	13.98	35.25

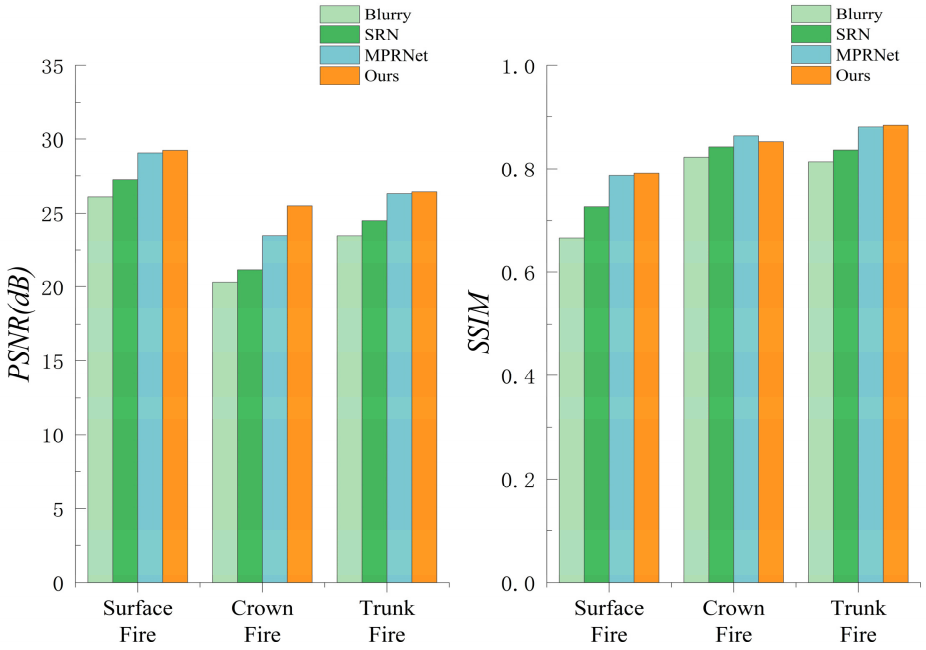

Figure 9. Comparison of deblurring effect under different forest fire scenarios.

4.2.2. Real Forest Fire Scene Test

Taking forest fire images captured by drones in Tangzu Village (101°3′29.2068″ E, 29°53′59.28″ N), Malangcuo Town, Yajiang County, Sichuan Province, China on 31 May 2023, as an example, the deblurring effects of different models at multiple shooting points were compared, as shown in Figure 10. It was found that our model can highlight the edge and texture characteristics of the fire while deblurring the image, which is conducive to detection and segmentation.

Table 5 and Figure 11 show the comparison of objective quality between the proposed method and two other blind moving image deblurring models. Because there is no clear image for reference, auxiliary indexes, such as LGF and SMD, are used to evaluate deblurring performance. Tests show that our model is superior on LGF and SMD.

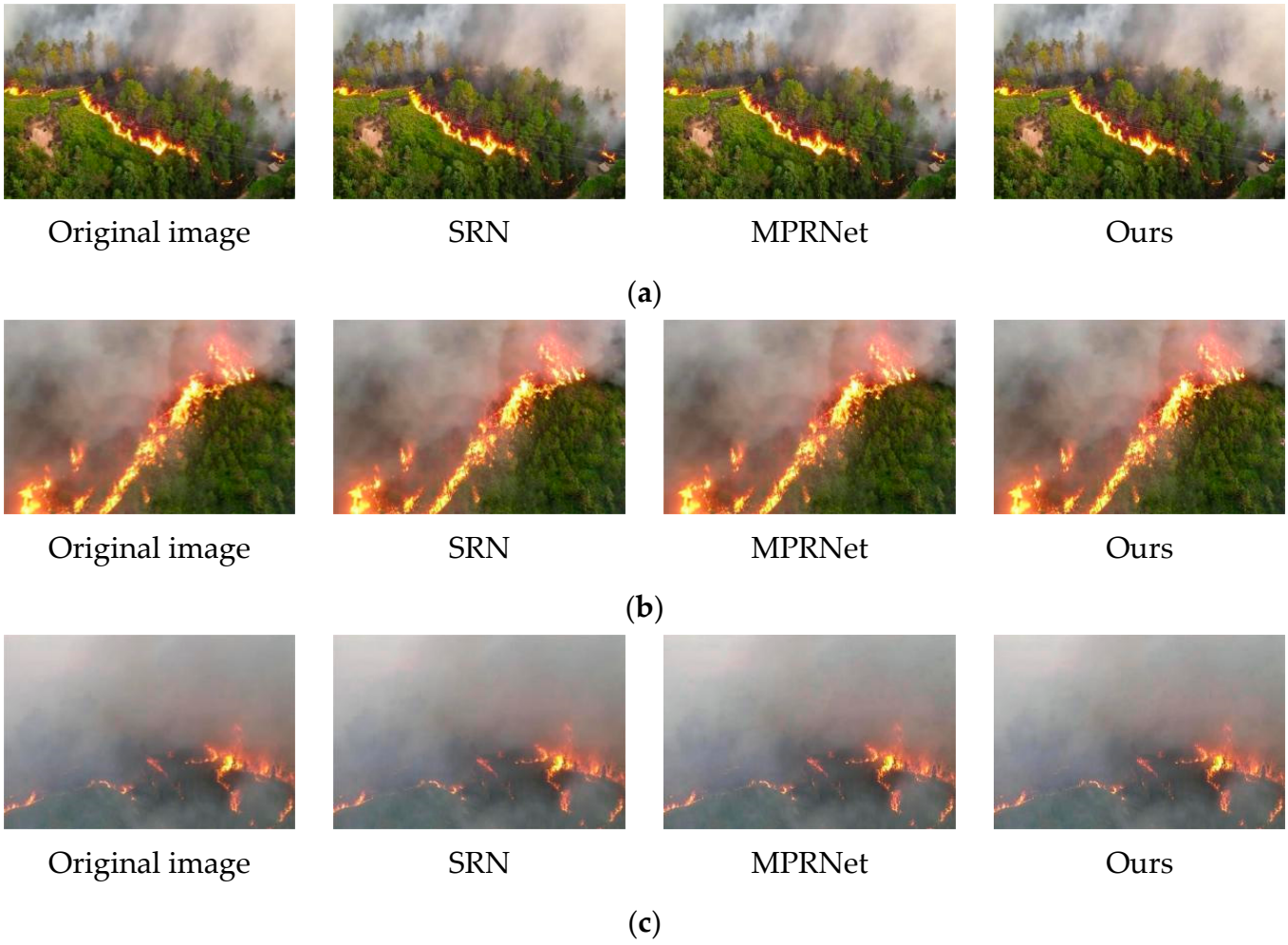

Figure 10. Comparison of deblurring effect. (**a**) The first forest fire scene; (**b**) the second forest fire scene; (**c**) the third forest fire scene.

Table 5. Forest fire scenes are used to test LGF and SMD.

Title 1	Model	LGF	SMD
The first forest fire scene	Original image	11.50	38.32
	SRN	13.27	44.17
	MPRNet	14.31	47.75
	Ours	14.50	50.80
The second forest fire scene	Original image	6.16	21.57
	SRN	8.24	27.16
	MPRNet	9.27	31.71
	Ours	10.04	36.49
The third forest fire scene	Original image	2.90	8.42
	SRN	3.03	9.01
	MPRNet	3.54	11.15
	Ours	3.89	13.11

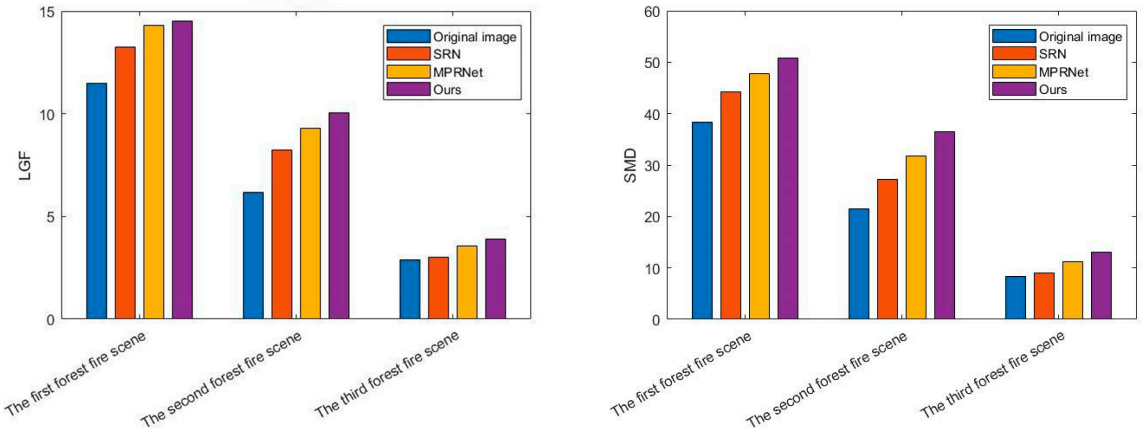

Figure 11. Forest fire scenes are used to test LGF and SMD.

4.3. Ablation Experiment

4.3.1. Analysis of Layers of Stacked Residuals in Encoder–Decoder

The encoder–decoder module designed in this study is composed of multiple stacked residuals. To evaluate the influence of the number of residuals on the network performance, an ablation experiment was conducted on the self-built forest fire dataset for the number of residuals M, whose values are set as 2, 4, 8, 16, and 20, respectively. The experimental results are listed in Table 6, and Figure 12 illustrates ablation studies on different numbers of residual modules. At M = 4, the PSNR and SSIM values were lower. As M increased, PSNR and SSIM also increased. When M = 16, PSNR and SSIM were corresponding. If the value was greater than 16, PSNR and SSIM increased smoothly. Considering the running speed of the model and the defusing performance, the M value used in the model was 16.

Table 6. Ablation studies on different numbers of residual modules.

M	2	4	8	16	20
PSNR	28.91	29.86	30.65	32.26	32.31
SSIM	0.921	0.933	0.941	0.955	0.958

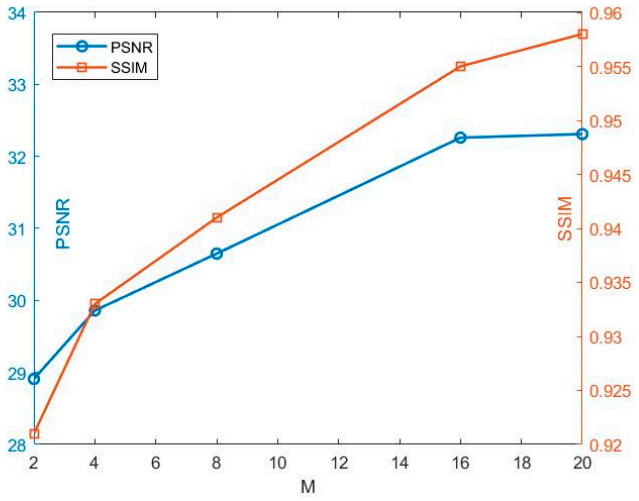

Figure 12. Ablation studies on different numbers of residual modules.

4.3.2. Network Module Combination

In this study, the model designs different modular combinations of self-built forest fire datasets to prove the validity of SFFM, MDAM, MCAM, and PM. Each combination scheme maintains the same training parameter environment and adopts the same PSNR

and SSIM indicators to evaluate the performance of each module, where the baseline is a multi-input and multi-output U-Net network (MIMO-UNet), as shown in Table 7 and Figure 13. Ablation studies with different module combinations list the index comparison data of different module combination structures. Experiments have demonstrated that the utilization of SFFM has led to a significant improvement in PSNR and SSIM. Building upon this foundation, MDAM and MCAM in the encoder–decoder component also improve the deblurring ability of the model. Additionally, PM also effectively contributed to the improvement of the deblurring effect. The results show that the above modules can improve the deblurring performance of the model and help restore the image with more detailed information.

Table 7. Ablation studies with different module combinations.

Baseline	SFFM	MDAM + MCAM	PM	PSNR	SSIM
√				30.28 dB	0.936
√	√			31.23 dB	0.941
√	√	√		31.65 dB	0.950
√	√	√	√	32.26 dB	0.955

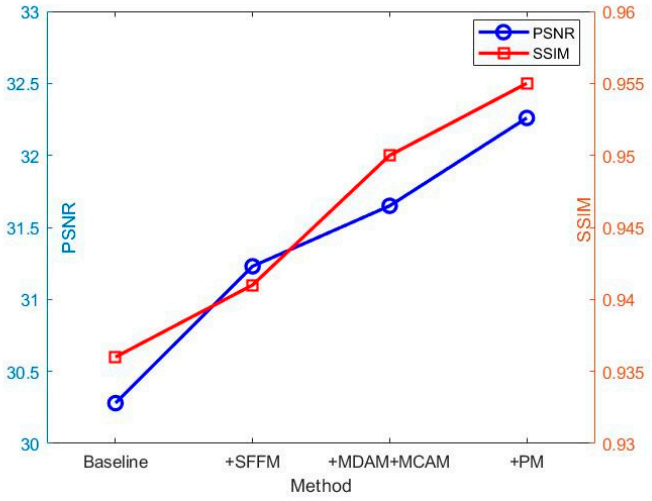

Figure 13. Ablation studies with different module combinations.

5. Discussion

Forest resources are crucial ecological assets essential for the survival of human society. Detection of forest fires holds significant importance in safeguarding the ecological security of a country. With the advancement of information technology, the utilization of drones for forest fire detection is on the rise [39], accompanied by escalating demands for high-quality aerial imagery. The issue of motion blur often arises during UAV image capture while in flight. Currently, numerous deep learning-based methods exist for image deblurring, but they exhibit certain limitations. Therefore, this study proposes a comprehensive investigation into the deblurring challenge encountered in forest fire images. Leveraging the inherent strengths of deep learning in end-to-end image deblurring, we propose novel image deblurring models based on the MIMO-UNet algorithm.

To address the challenge of insufficient clear-blurred image pairs in forest fire scenes, this study generated a dataset comprising such pairs. The comparative experiments are shown in Table 7 and Figure 13. Ablation studies with different module combinations demonstrate that the enhanced SFFM, MDAM, MCAM, and PM methods notably enhance both PSNR and SSIM metrics in our self-built forest fire dataset. Our model demonstrated exceptional performance compared to other models, with a slightly lower SSIM of 0.03 compared to the MPRNet model. We attribute this difference to MPRNet's utilization of a multi-stage progressive

restoration strategy. Specifically, the original-resolution subnetwork (ORSNet) in the final stage is believed to enhance image quality without sacrificing image structure and semantic information. Therefore, in our future endeavors, we aim to prioritize the reduction of semantic information loss and enhancement of deblurring performance.

In this study, it was found that there was still a certain gap between the subjective visual perception of the deblurred image and the effect reflected by the objective evaluation index. How to establish an image restoration quality evaluation index more in line with subjective perceptions is a problem that needs to be solved. In future work, we intend to use machine learning techniques to assess image quality without reference to enabling a more comprehensive assessment of blurring results. The model presented in this paper still requires extensive training time on high-configuration computers. However, computational complexity and training duration can be diminished through optimization of the network structure. Given that edge computing devices typically possess lower computational power compared to conventional computers, the model can be further enhanced through techniques such as quantization and pruning. These methods compress the model size or employ a more lightweight network structure, consequently reducing computational demands. As a result, the network model can be efficiently deployed and recovered on edge computing devices.

Based on the forest fire detection system made by our team [40,41] (the whole system consists of a UAV, a Raspberry Pi controller, an OAK-D camera, and a GPS module), we aim to build a forest fire detection model based on multi-task learning, consisting of 3 tasks (a deblurring task, a detection task, and a segmentation task). In order to enhance the user experience and ease of operation, we intend to build a cross-platform HMI using PyQt5. This design not only improves the utility of the system, but also ensures reliability and stability in the field environment. By deploying our models on UAVs, we can achieve true real-time recognition and response capabilities to more effectively monitor and deal with emergencies such as forest fires. This integrated solution will unlock new potential for forest fire prevention and improve the ability to respond to disasters in a timely manner, thereby enhancing the protection of human life and property.

6. Conclusions

In this study, we successfully built and validated an innovative spatial–frequency domain fusion network model with significant improvements over MIMO-UNet to optimize image deblurring tasks. By introducing an advanced MDAM in EM, we not only increased the receptive field of the model, but also effectively suppressed redundant information, thereby improving the overall performance of the network. In addition, in the multi-scale feature fusion module, we abandoned the traditional U-Net jump connection module and adopted a strategy of combining spatial domain and frequency domain information, which significantly reduced the information loss in the feature fusion process and improved the recovery with more detailed characteristics. MCAM in DM improved the reconstruction of local and contextual information. During model training, we used the weighted loss function, which not only improved the stability of the model but also optimized the performance of image blurring.

By training and testing the self-built forest fire dataset, our model outperformed the comparison model in various performance indices and achieved excellent results in experimental comparison. Especially when processing forest fire images, our model could highlight the texture details of recovered images, which is of great importance for wildfire monitoring and management. The LGF and SMD were used to evaluate the deblurring effect of forest fire images in real scenes, and our model has achieved the best performance in comparative experiments. The conclusive experimental findings show that the proposed forest fire image deblurring model has a PSNR of 32.26 dB, SSIM of 0.955, LGF of 10.93, and SMD of 34.31. In experiments without reference images, the model performs well in terms of LGF and SMD. It is worth noting that, compared with other baseline models and

other commonly used image deblurring models, this model is generally improved in terms of indicators.

Author Contributions: Conceptualization, X.K. and Y.L.; methodology, X.K.; software, X.K.; validation, R.H. and S.L.; formal analysis, X.K.; investigation, Y.L.; resources, Y.L.; data curation, X.K.; writing—original draft preparation, X.K.; writing—review and editing, X.K.; visualization, R.H.; supervision, Y.L.; project administration, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (grant number KYCX22_1056) and National Key R&D Program of China (grant number 2017YFD0600904).

Data Availability Statement: The data supporting this study will be shared on reasonable request sent to the corresponding author.

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

References

1. Jin, L.; Yu, Y.; Zhou, J.; Bai, D.; Lin, H.; Zhou, H. SWVR: A Lightweight Deep Learning Algorithm for Forest Fire Detection and Recognition. *Forests* **2024**, *15*, 204. [CrossRef]
2. Oishi, Y.; Yoshida, N.; Oguma, H. Detecting Moving Wildlife Using the Time Difference between Two Thermal Airborne Images. *Remote Sens.* **2024**, *16*, 1439. [CrossRef]
3. Chen, B.; Bai, D.; Lin, H.; Jiao, W. Flametransnet: Advancing forest flame segmentation with fusion and augmentation techniques. *Forests* **2023**, *14*, 1887. [CrossRef]
4. Peruzzi, G.; Pozzebon, A.; Van Der Meer, M. Fight fire with fire: Detecting forest fires with embedded machine learning models dealing with audio and images on low power iot devices. *Sensors* **2023**, *23*, 783. [CrossRef] [PubMed]
5. Duangsuwan, S.; Klubsuwan, K. Accuracy Assessment of Drone Real-Time Open Burning Imagery Detection for Early Wildfire Surveillance. *Forests* **2023**, *14*, 1852. [CrossRef]
6. Ahmed, Z.E.; Hashim AH, A.; Saeed, R.A.; Saeed, M.M. Monitoring of Wildlife Using Unmanned Aerial Vehicle (UAV) with Machine Learning. In *Applications of Machine Learning in UAV Networks*; IGI Global: Hershey, PA, USA, 2024; pp. 97–120.
7. Huihui, Y.; Daoliang, L.; Yingyi, C. A state-of-the-art review of image motion deblurring techniques in precision agriculture. *Heliyon* **2023**, *9*, e17332. [CrossRef] [PubMed]
8. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 22–27 October 2022; Springer Nature Switzerland: Cham, Switzerland, 2022; pp. 17–33.
9. Mou, C.; Wang, Q.; Zhang, J. Deep Generalized Unfolding Networks for Image Restoration. *arXiv* **2022**, arXiv:2204.13348.
10. Rim, J.; Kim, G.; Kim, J.; Lee, J.; Lee, S.; Cho, S. Realistic blur synthesis for learning image deblurring. In *European Conference Computer Vision LNCS*; Springer Nature Switzerland: Cham, Switzerland, 2022; Volume 13667, pp. 487–503.
11. Shyam, P.; Kim, K.-S.; Yoon, K.-J. GIQE: Generic Image Quality Enhancement via Nth Order Iterative Degradation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 2067–2077. [CrossRef]
12. Kaufman, A.; Fattal, R. Deblurring Using Analysis-Synthesis Networks Pair. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 5810–5819. [CrossRef]
13. Fang, Y.; Zhang, H.; Wong, H.S.; Zeng, T. A robust non-blind deblurring method using deep denoiser prior. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), New Orleans, LA, USA, 18–24 June 2022; pp. 734–743. [CrossRef]
14. Guan, Z.; Tsai, E.H.R.; Huang, X.; Yager, K.G.; Qin, H. Non-Blind Deblurring for Fluorescence: A Deformable Latent Space Approach with Kernel Parameterization. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2022; pp. 101–109. [CrossRef]
15. Dong, J.; Pan, J.; Sun, D.; Su, Z.; Yang, M.H. Learning Data Terms for Non-blind Deblurring. In *Computer Vision—ECCV 2018: ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 11215.
16. Vasu, S.; Maligireddy, V.R.; Rajagopalan, A.N. Non-blind Deblurring: Handling Kernel Uncertainty with CNNs. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3272–3281. [CrossRef]
17. Schuler, C.J.; Hirsch, M.; Harmeling, S.; Schölkopf, B. Learning to Deblur. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1439–1451. [CrossRef]
18. Li, L.; Pan, J.; Lai, W.S.; Gao, C.; Sang, N.; Yang, M.H. Learning a Discriminative Prior for Blind Image Deblurring. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6616–6625.

19. Zhuang, Z.; Li, T.; Wang, H.; Sun, J. Blind image deblurring with unknown kernel size and substantial noise. *Int. J. Comput. Vis.* **2024**, *132*, 319–348. [CrossRef]
20. Wang, H.; Hu, C.; Qian, W.; Wang, Q. RT-Deblur: Real-time image deblurring for object detection. *Vis. Comput.* **2024**, *40*, 2873–2887. [CrossRef]
21. Zhao, S.; Xing, Y.; Xu, H. WTransU-Net: Wiener deconvolution meets multi-scale transformer-based U-net for image deblurring. *Signal Image Video Process.* **2023**, *17*, 4265–4273. [CrossRef]
22. Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5971–5979. [CrossRef]
23. Nah, S.; Kim, T.H.; Lee, K.M. IEEE: Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 257–265.
24. Zheng, S.; Wu, Y.; Jiang, S.; Lu, C.; Gupta, G. Deblur-yolo: Real-time object detection with efficient blind motion deblurring. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Virtual, 18–22 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–8.
25. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8174–8182.
26. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8878–8887.
27. Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Stenger, B.; Liu, W.; Li, H. Deblurring by realistic blurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2020; pp. 2737–2746.
28. Cho, S.J.; Ji, S.W.; Hong, J.P.; Jung, S.W.; Ko, S.J. Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4641–4650.
29. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693.
30. Huang, X.; He, J.S. Fusing Convolution and Self-Attention Parallel in Frequency Domain for Image Deblurring. *Neural Process. Lett.* **2023**, *55*, 9811–9829. [CrossRef]
31. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
32. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.
33. Zhang, Z.; Wang, X.; Jung, C. DCSR: Dilated convolutions for single image super-resolution. *IEEE Trans. Image Process.* **2018**, *28*, 1625–1635. [CrossRef]
34. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57. [CrossRef]
35. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2017**, arXiv:1711.05101.
36. He, L.; Huang, J. A visual SLAM algorithm based on demotion blur. *Geo Spat. Inf.* **2019**, *21*, 31–35. (In Chinese)
37. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
38. Mehri, A.; Ardakani, P.B.; Sappa, A.D. MPRNet: Multi-path residual network for lightweight image super resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 2704–2713.
39. Chen, Y.; Wang, T.; Lin, H. Research on Forest Flame Detection Algorithm Based on a Lightweight Neural Network. *Forests* **2023**, *14*, 2377. [CrossRef]
40. Lu, K.; Huang, J.; Li, J.; Zhou, J.; Chen, X.; Liu, Y. MTL-FFDET: A Multi-Task Learning-Based Model for Forest Fire Detection. *Forests* **2022**, *13*, 1448. [CrossRef]
41. Lu, K.; Xu, R.; Li, J.; Lv, Y.; Lin, H.; Liu, Y. A Vision-Based Detection and Spatial Localization Scheme for Forest Fire Inspection from UAV. *Forests* **2022**, *13*, 383. [CrossRef]