Neural Processing Letters (2024) 56:88
https://doi.org/10.1007/s11063-024-11462-x

Image Deblurring Using Feedback Mechanism and Dual Gated Attention Network

Jian Chen[1] · Shilin Ye[1] · Zhuwu Jiang[2] · Zhenghan Fang[3]

Accepted: 25 December 2023 / Published online: 6 March 2024

Abstract
Recently, image deblurring task driven by the encoder-decoder network has made a tremendous amount of progress. However, these encoder-decoder-based networks still have two disadvantages: (1) due to the lack of feedback mechanism in the decoder design, the reconstruction results of existing networks are still sub-optimal; (2) these networks introduce multiple modules, such as the self-attention mechanism, to improve the performance, which also increases the computational burden. To overcome these issues, this paper proposes a novel feedback-mechanism-based encoder-decoder network (namely, FMNet) that is equipped with two key components: (1) the feedback-mechanism-based decoder and (2) the dual gated attention module. To improve reconstruction quality, the feedback-mechanism-based decoder is proposed to leverage the feedback information via the feedback attention module, which adaptively selects useful features in the feedback path. To decrease the computational cost, an efficient dual gated attention module is proposed to perform the attention mechanism in the frequency domain twice, which improves deblurring performance while reducing the computational cost by avoiding redundant convolutions and feature channels. The superiority of FMNet in terms of both deblurring performance and computational efficiency is demonstrated via comparisons with state-of-the-art methods on multiple public datasets.

Keywords Image deblurring · Encoder-decoder network · Feedback mechanism · Gated attention

✉ Jian Chen
jianchen@fjut.edu.cn

Zhenghan Fang
zfang23@jhu.edu

[1] School of Electronic, Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350118, Fujian, China

[2] School of Ecological Environment and Urban Construction, Fujian University of Technology, Fuzhou 350118, Fujian, China

[3] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

1 Introduction

Image blur is one of the most common artifacts introduced via unpredictable factors such as camera shake, fast-moving objects, etc. Once an image is blurred, post-processing procedures (e.g. object detection [1] and image segmentation [2]) become tougher tasks. Therefore, the image deblurring task has drawn broad attention in the research community, aiming at improving the image quality for various subsequent vision tasks. However, image deblurring is known to be a challenging ill-posed problem. Most classical deblurring methods were based on optimization, using assumptions on image prior distributions to constrain the solution space, such as the L0 sparse prior [3] and dark channel prior [4]. However, severe blur is often difficult to remove with such limited and simple hand-crafted prior assumptions.

In recent years, deep learning-based methods in an end-to-end mode have been introduced in many studies [5–12]. These networks learn the non-linear mapping from the blurred images to the deblurred ones directly without blur kernel estimation beforehand. For example, a deep multi-scale convolutional neural network (CNN) based on a coarse-to-fine strategy was proposed for recovering the blurred images [5]. In another work, an encoder-decoder-based deblurring network was proposed to improve the deblurring result [6]. Furthermore, several improvements were introduced to the encoder-decoder network, including the multi-scale (MS) [7, 8], multi-patch (MP) [9–11], and multi-temporal (MT) [12] modules. These improved encoder-decoder networks obtained encouraging deblurring performance. However, these methods still cannot recover sufficient texture details due to the fact that the decoder path is a one-way procedure for image reconstruction. As a result, if incorrect features are generated in the decoder path, such as checkerboard artifact [13], these errors will be propagated to subsequent layers and cannot be easily corrected in the one-way procedure. Therefore, the methods based on the one-way encoder-decoder network usually produce sub-optimal deblurred results. To solve this issue, we propose to introduce a feedback mechanism into the decoder path.

The feedback mechanism is of great importance in the human visual perception system [14]. It transmits the output information into the input information to optimize the result. The feedback mechanism was already utilized and achieved significant success in many vision tasks [15–18]. Inspired by these methods, we combine the feedback mechanism and the decoder, thereby proposing a feedback-mechanism-based decoder for the proposed image deblurring network. Different from [16], which fed the feedback information directly into the input data, ignoring the fact that incorrect features may occur in any layer of the decoder, we send the feedback information into each layer of the decoder network to improve reconstruction quality. Furthermore, in contrast to the integrated approach proposed in [16], which incorporates the feedback information and input data directly, the proposed feedback attention module (FAM) integrates feedback information and input data via attention mechanism to adaptively exploit useful features in the feedback path.

Another reason that severe blur is difficult to handle is the small receptive field of existing encoder-decoder networks. Consequently, some networks employ multiple modules, such as self-attention mechanism (SA) [19] to capture the long-range dependencies between distant blur pixels. However, the shortcoming of these modules is that the computation complexity grows quadratically as the image size increases. Therefore, a few computationally efficient variants of SA [20–22] have been proposed. Although the method achieved promising performance, it is still computationally expensive, especially for high-resolution images in the deblurring tasks. We note that the high cost mainly comes from two major computational bottlenecks: (1) the redundant convolutions [23] and (2) the channel redundancy [24]. To

address these bottlenecks, we propose an efficient dual gated attention module (DGAM), where several key improvements were introduced to the network architecture to reduce model redundancies. Tunable hyperparameters are introduced to reduce channel numbers and avoid channel redundancy. Previous works (e.g. [21]) adopt redundant convolutions to capture long-range dependencies in the frequency domain, which is costly. Inspired by [25], a learnable matrix is leveraged to model the long-range dependencies in the frequency domain. It is worth noting that previous works (e.g. [21]) usually capture the dependencies only once, while more than one order of interactions is beneficial to enhancing feature representation ability [26]. Consequently, we propose the dual gated attention module to model long-range dependencies in the frequency domain twice, effectively improving the deblurring performance. Through the integration of a feedback-mechanism-based decoder and DGAM, we present a novel feedback-mechanism-based deblurring network. The main contributions of our work are as follows:

(1) We propose a feedback-mechanism-based decoder, which leverages the feedback attention module to adaptively select useful features in the feedback path to improve the reconstruction quality.

(2) We propose an efficient dual gated attention module, which captures the long-range dependencies in the frequency domain twice, and reduces the computational cost by avoiding redundant convolutions and channels.

(3) We propose a feedback-mechanism-based encoder-decoder deblurring network (namely, FMNet), which equips a feedback-mechanism-based decoder and dual gated attention module. Extensive experiments on public datasets prove the outstanding deblurring performance and excellent computational efficiency of our FMNet.

2 Related Work

In this section, we will give a brief review on the development of image deblurring methods, the self-attention mechanism, and the feedback mechanism.

2.1 Image Deblurring Methods

Image deblurring is known as a challenging low-level vision task. In the early times, most deblurring methods are optimization-based methods. To improve the deblurring performance, deep learning has been introduced to the deblurring task recently. As a pioneering work, Nah et al. [5] proposed a deep convolutional neural network based on a coarse-to-fine strategy to restore a blurred image to its sharp version. After that, Gao et al. [6] presented an encoder-decoder-based network to improve the stability of training by integrating parameter sharing and skip connection. Inspired by [6], many deblurring works adopt the encoder-decoder network as the baseline for the deblurring tasks. For instance, Zhang et al. [9] proposed a neural network with stacked layers, each of which employed a decoder-encoder network to restore a blurred image to its latent sharp image. Zamir et al. [10] used two encoder-decoder networks to learn contextually relevant features, and then combined them with high-resolution branches that preserve local information. Based on an encoder-decoder network, Park et al. [12] introduced the recurrent neural networks (RNN) to perform image deblurring. In recent years, inspired by the great success of transformer [27] in high-level vision tasks, several methods [28, 29] have attempted to introduce transformer for image deblurring in an encoder-decoder network. While transformer enables the network to capture global dependencies to improve

performance, it also brings high computational complexity. Hence, NAFNet [30] builds a more efficient fully convolutional network with gating and attention mechanisms designed to improve the performance, but excessive stacking of attention modules and convolutional layers also imposes a heavy computational burden. Different from [28, 29], the proposed DGAMs are stacked at the lowest image scale in an encoder-decoder network, and then the global dependencies are captured with $O(N)$ complexity in frequency domain to reduce the computational complexity.

2.2 Self-Attention Mechanism

While promising deblurring performance has been demonstrated based on the encoder-decoder-based networks, severe blur is still a challenging task due to the small receptive field of these networks. Therefore, many researchers applied some modules, such as the self-attention mechanism [19], to capture the long-range dependencies for increasing the receptive field of the encoder-decoder network. For instance, Purohit et al. [19] introduced SA to capture the long-range dependencies to improve the deblurring performance for images with severe blur. However, SA also brings high computational cost. Hence, some works attempted to overcome this shortcoming and proposed variants of SA. For example, Zhang et al. [20] proposed the spatial SA and channel SA to capture the global contextual dependence of features. Huang et al, [22] designed a network with parallel processing of convolution and self-attention mechanisms to allow capturing both local and global information at the same time. Rao et al. [25] proposed a learnable global filter to learn long-range dependencies in the frequency domain to reduce computational complexity. Zhou et al. [21] proposed a self-attention module in an encoder-decoder network to model long-range dependencies in the frequency domain for medical image reconstruction. Rao et al. [26] designed a high-order spatial interaction strategy to capture the long-range dependencies efficiently.

2.3 Feedback Mechanism

The feedback mechanism plays a crucial role in the human visual perception system [14]. The mechanism has already been employed in various tasks, such as image segmentation and image super-resolution (SR). For example, Tsuda et al. [15] leveraged the feedback mechanism in an encoder-decoder network for cellular image segmentation. Girum et al. [16] proposed a feedback network to improve the accuracy of medical image segmentation. In image super-resolution tasks, Li et al. [17] designed a negative feedback mechanism to continuously finetune the SR images. Based on [17], Liu et al. [18] introduced the self-attention into the feedback network to obtain better SR images. The feedback mechanism brings significant improvement in those vision tasks.

3 The Proposed Model

Figure 1 illustrates the overall architecture of FMNet. The network consists of three main components: an encoder, a latent space module, and a feedback-mechanism-based decoder. We first use the encoder to extract multi-scale features from the image. Specifically, the encoder consists of two encoder blocks (EBs). Each block is composed of a convolutional layer followed by a residual block [5]. The first EB block (named EB_1 in Fig. 1) uses 3×3 convolution to extract low-level features from the input image, while the second EB block

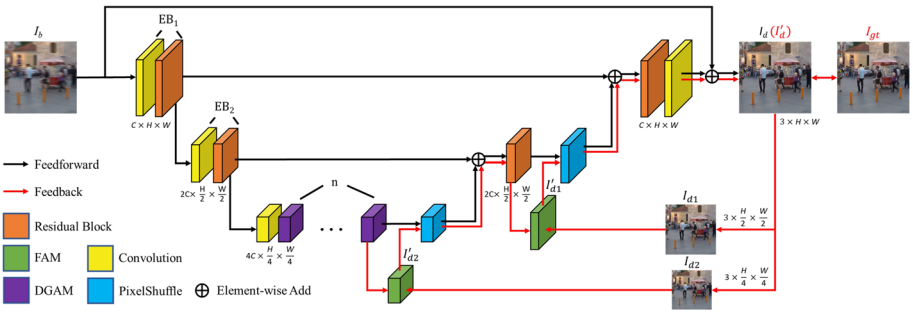

Fig. 1 The architecture of proposed FMNet. The decoder contains a structure with a feedback mechanism (marked as red line), which is named feedback-mechanism-based decoder

(named EB$_2$ in Fig. 1) uses 2×2 convolution with a stride of 2 for down-sampling and feature extraction. Then, the features from the encoder are fed into a latent space module consisting of cascades of multiple DGAMs to obtain critical deblurring cues from the encoder features. Finally, a feedback-mechanism-based decoder is employed to reconstruct the deblurred image.

3.1 Feedback-Mechanism-Based Decoder

To improve the performance of image reconstruction, we propose to integrate the feedback mechanism into the decoder path and propose the feedback-mechanism-based decoder. More precisely, our decoder includes two reconstruction processes: the feedforward process and the feedback process. Since image reconstruction is performed twice in the decoder via a feedback loop, more texture details can be recovered and incorrect predictions from the preceding feedforward path can be suppressed.

As shown in Fig. 1, the feedforward path of the decoder is illustrated with black lines. The up-sampling operations in the decoder are implemented via PixelShuffle [31], followed by residual blocks for feature reconstruction. The output of the decoder after the feedforward process, dubbed I_d, is obtained by progressive reconstruction from the output of DGAM. Next, the feedback process is used to improve the reconstruction result, as illustrated by red lines in Fig. 1. To improve the reconstruction quality at each layer of the decoder, we down-sample I_d by 2 and 4 times, respectively, to obtain feedback information (I_{d1} and I_{d2}) for the decoder at two different scales. Specifically, the feedback information (I_{d1} and I_{d2}) are fed into the feedback attention module to estimate attention maps used to reweight the features obtained in the feedforward path. Accordingly, FAM can efficiently utilize the feedback information to extract informative features and suppress the artifacts in the feedforward feature maps, leading to an improvement in image reconstruction quality. Then, the reweighted features, I'_{d1} and I'_{d2}, are fed into the corresponding decoder layers to generate improved deblurred image I'_d (see red lines in Fig. 1). It is worth noting that the weights of residual blocks and convolution layers in the decoder are shared between the feedforward and feedback processes.

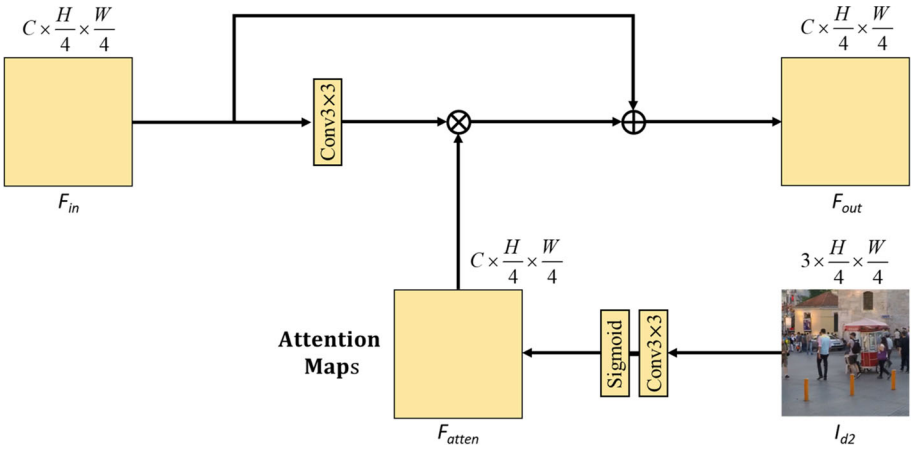

Fig. 2 The architecture of the proposed feedback attention module (FAM)

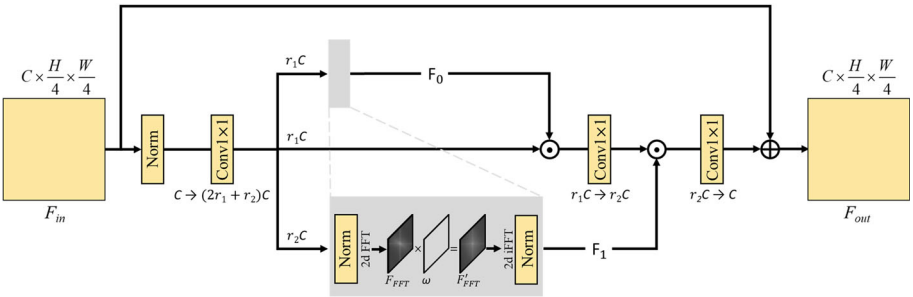

Fig. 3 The architecture of the proposed DGAM. $\odot$ is the element-wise multiplication operation to perform the gated attention. We perform the attention to capture the long-range dependencies in the frequency domain twice

3.2 Feedback Attention Module

The structure of the proposed feedback attention module is shown in Fig. 2. For illustrative purposes, we describe the FAM at the bottom layer of the decoder in detail, while the FAMs at other layers are constructed similarly. As shown in Fig. 2, we denote the input of FAM (i.e., output from DGAM) as F_{in} and the feedback information as I_{d2}. First, a 3×3 convolution followed by sigmoid activation is applied on I_{d2} to derive the attention map F_{atten}. Then, F_{in} is fed into a 3×3 convolution layer, and multiplied by F_{atten} to adaptively select informative features and generate residual information. Finally, we add the residual information to the original F_{in} to obtain the final reweighted output F_{out}.

3.3 Dual Gated Attention Module

It has been shown that self-attention module is beneficial in reducing severe blur [19] as it can capture the long-range dependencies to increase the receptive field. However, these SA modules often come with high computational complexity, especially when processing high-resolution images that are common in deblurring tasks. To reduce computational burden and

improve the efficiency of the deblurring process, we propose dual gated attention module, the structure of which is shown in Fig. 3. As the figure shows, the input of DGAM (denoted F_{in}) is first fed into a LayerNorm layer for normalization. Then, a 1×1 convolution is used to expand the number of channels from C to $(2r_1 + r_2)C$ where $r_1 \leq 1$ and $r_2 \leq 1$ are hyperparameters that control the reduction factor for removing channel redundancy. The $(2r_1 + r_2)C$ channels are divided into three streams, each with r_1C, r_1C and r_2C channels, to capture the long-range dependencies.

Next, we capture the long-range dependencies in the frequency domain in two streams (i.e., F_0 and F_1 in Fig. 3.). The long-range dependencies are captured twice to improve the deblurring performance. The existing work [21] used vanilla convolution to capture long-range dependencies, which resulted in redundant convolutions and increased computational cost. In our work, we propose to use a learnable $H \times W \times C$ matrix ω (of the same size as the feature map) to model the dependencies in the frequency domain. Taking the F_1 stream in Fig. 3 as an example, the feature map is first transformed to the frequency domain using 2D fast Fourier transform (2D FFT) to obtain F_{FFT} (see Fig. 3). Then, the learnable matrix ω is multiplied (entrywise) with F_{FFT} to capture the dependencies in the frequency domain and obtain the feature $F^{'}_{\text{FFT}}$. Afterwards, the $F^{'}_{\text{FFT}}$ acts as a gating signal for the feature map from r_2C channel to perform gated attention. Compared with vanilla convolution [21], the computational complexity is reduced from $O(N^2)$ for convolutional operations to $O(N)$ for the element-wise matrix multiplication, where N denotes all pixel points in a feature map.

3.4 Loss Function

We adopt the L1 loss to minimize the distance between the deblurred image $I^{'}_{d}$ and the ground truth I_{gt}. Note that the final deblurred image $I^{'}_{d}$ is the result of the feedback process. We further use the Fourier loss [32] as an auxiliary loss. The total loss is given by:

$$\ell_{total} = \left\| I^{'}_{d} - I_{\text{gt}} \right\|_1 + \lambda \left\| \mathcal{F}(I^{'}_{d}) - \mathcal{F}(I_{\text{gt}}) \right\|_1 \tag{1}$$

where $\mathcal{F}$ denotes the Fourier transform and λ is the weight of the auxiliary loss, The value of λ is empirically set to 0.1 according to reference [32].

4 Experiments

4.1 Implementation Details

We introduce two publicly available datasets for comparison: the GoPro [5] and RealBlur [33] datasets. The GoPro dataset [5] consists of 3214 pairs of blurred and sharp images, 2103 of which are leveraged for training, and the rest 1111 images are used for testing. The RealBlur dataset [33] consists of 1960 paired images, all of which are used for testing. We adopt the AdamW optimizer (β_1=0.9, β_2= 0.99, $weight\ decay = 1 \times 10^{-3}$) and train our model on NVIDIA RTX 4000 GPU. All test experiments are implemented under the same computer. The training is carried out in two stages. In the first stage, we randomly crop the input image into 256×256 patches and set the batch size as 4 for 3000 epochs. The initial learning rate is set as 1×10^{-3} and decays to 1×10^{-7} with the cosine annealing strategy. Then, the second training stage increases the input size to 384×384 with a batch size of 1 for 500 epochs, with an initial learning rate of 1×10^{-4}. Additionally, random rotation

Table 1 Performance comparison on the GoPro and RealBlur test dataset

Methods	GoPro PSNR	SSIM	RealBlur PSNR	SSIM	Param	FLOPs	Time
MSCNN	29.08	0.914	30.19	0.834	11.72	336	2702
DFG	29.81	0.937	N/A	N/A	3.1	N/A	N/A
SRN	30.26	0.934	32.11	0.907	6.8	167	5150
DG	28.70	0.858	30.88	0.868	6.06	35	1671
DGv2	29.55	0.934	31.98	0.905	60.9	42	178
DBGAN	31.10	0.942	29.35	0.827	11.58	759	651
MTRNN	31.15	0.945	32.11	0.906	**2.63**	**27**	74
DMPHN	31.20	0.940	32.06	0.904	21.7	235	29
SAPHN	32.02	0.953	N/A	N/A	N/A	N/A	N/A
RADN	31.76	0.953	N/A	N/A	N/A	N/A	N/A
MIMO+	32.45	0.957	31.585	0.892	16.1	154	**25**
MPRNet	32.66	0.959	**32.34**	0.912	20.1	760	206
BANet	32.54	0.957	31.91	0.896	18	263	**25**
KiT	32.70	0.959	N/A	N/A	N/A	N/A	N/A
DGUNet	32.71	0.960	31.83	0.9136	17.3	865	254
IPT	32.58	N/A	N/A	N/A	114	N/A	N/A
MAXIM	32.86	**0.961**	32.30	0.911	22.2	339	17373
FMNet	**32.95**	**0.961**	**32.34**	**0.916**	10.4	71	143

The best results are highlighted in bold. Param (the number of parameters in the network) is measured in millions (M). FLOPs represent the floating point operations and are measured in giga (G). Time denotes the running time and is measured in milliseconds (m's)

and vertical flipping are utilized for data augmentation to improve generalization capability. Note that the comparison methods included in the experiments follow their original training strategies.

4.2 Method Comparison

4.2.1 Quantitative Comparison

To further evaluate our method, we compared FMNet with state-of-the-art deblurring methods, including MSCNN [5], DFG [34], SRN [35], DG [36], DGv2 [37], MTRNN [12], DMPHN [9], RADN [19], DBGAN [38], MIMO+ [32], BANet [39], MPRNet [10], HINet [40], DGUNet [41], KiT [42], SAPHN [43] and MAXIM [44], IPT [45]. The average PSNR and SSIM for the GoPro dataset are shown in Table 1. Among these methods, results of DFG [34], SAPHN [43], RADN [19], KiT [42], and IPT [45] are derived from the original papers, as the source codes of those methods are not available, results of other methods are obtained via retraining the corresponding models.

FMNet achieves higher PSNR and SSIM on the GoPro test dataset compared with the encoder-decoder-based network. For example, FMNet increases the PSNR by 1.8 dB, 0.29 dB and 0.09 dB compared with MTRNN [12], MPRNet [10] and MAXIM [44], respectively. This can be attributed to the introduction of DGAM which enlarges the receptive field and the

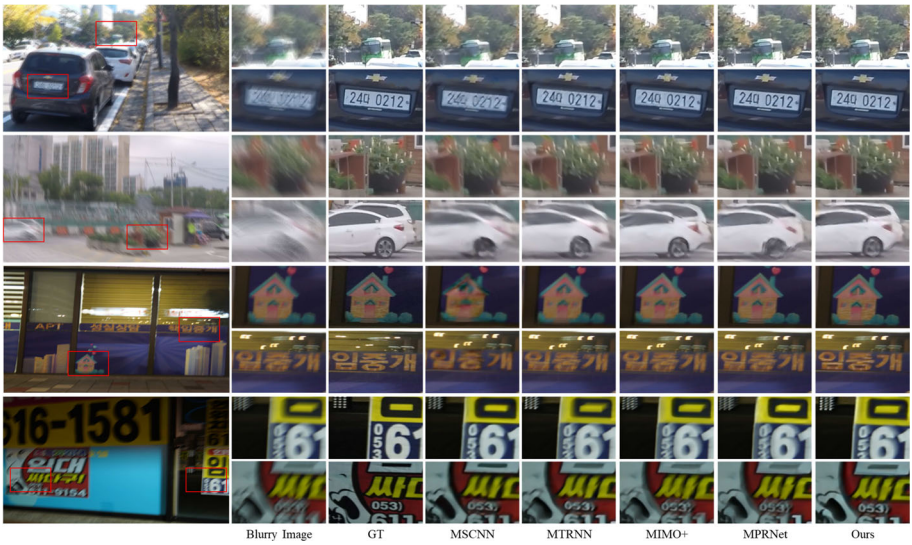

Fig. 4 Visual comparison of different deblurring methods on the GoPro dataset and RealBlur dataset. The top two rows show results on the GoPro dataset, and the bottom two rows show results on the RealBlur dataset

feedback-mechanism-based decoder which improves the reconstruction quality. Compared with the methods equipped with self-attention or its variants, FMNet also achieves better performance and lower computational cost. For instance, FMNet increases the PSNR by 0.93 dB and 1.19 dB compared with SAPHN [43] and RADN [19], respectively. Furthermore, FMNet reduces the number of parameters by a factor of ten and has higher PSNR compared with IPT [45]. Meanwhile, MAXIM [44] cost 2 times more parameters and 4 times more FLOPs than FMNet, but obtain lower PSNR by 0.09 dB. In addition, FMNet also obtains the best average PSNR and SSIM in the RealBlur dataset, as summarized in Table 1.

4.2.2 Visual Comparison

The visual results are depicted in Fig. 4. The test images are selected from the GoPro dataset and the Realblur test dataset, respectively. We compare FMNet with the state-of-the-art methods, including MSCNN [5], MTRNN [12], MIMO+ [32], and MPRNet [10]. Among those methods, MSCNN [5] is the classical method for image deblurring, while other methods are all encoder-decoder-based methods. The deblurring details for different scenes are zoomed in for visual comparison. As shown in Fig. 4, the earlier CNN-based methods (such as MSCNN [5]) have poor deblurring results, with significant artifacts remaining on the license plate. The encoder-decoder-based methods (such as MTRNN [12]) obtain a clearer outline of the license plate. MIMO+ [32] and MPRNet [10] generate much better deblurring results on license plates and letters. However, the detailed information from these methods is still not clear enough, and with a few blur artifacts remaining in deblurred results. By contrast, FMNet shows fewer blur artifacts and better image quality, which demonstrates the superiority of the proposed method.

Table 2 Ablation studies for the feedback-mechanism-based decoder

Baseline	One-P	Two-P	FAM	PSNR	Params	FLOPs	Time (ms)
✓				32.74	9.4	51	142.43
✓	✓			32.79	11.3	71	143.06
✓	✓	✓		32.83	11.3	77	146.34
✓	✓	✓	✓	32.95	10.4	71	143.30

Table 3 Ablation studies for the DGAM

	One-T	Two-T	PSNR	Params	FLOPs	Time
Conv		✓	31.87	10.4	80	169
ω	✓		32.49	10.4	67	103
ω		✓	32.95	10.4	71	143

Table 4 Ablation studies for hyperparameters in DGAM

r_1	r_2	PSNR	Params	FLOPs	Time
0.5	0.5	32.59	7.5	62	121
0.5	1	32.95	10.4	71	143
1	1	32.44	13.6	81	179

4.3 Ablation Studies

We also demonstrate the effectiveness of FMNet by validating each component through ablation experiments. All experiments are implemented on the GoPro dataset.

4.3.1 Ablation study of feedback-mechanism-based decoder

To illustrate the effectiveness of the feedback-mechanism-based decoder, we implement the original decoder without feedback mechanism in our FMNet as the baseline decoder (named Baseline in Table 2). First, the feedback mechanism and integrated method proposed in [16] are embedded in the bottom layer of the original decoder to establish one feedback path (named One-P in Table 2). It can be observed that the feedback mechanism increases the value of PSNR by 0.05 dB. To ensure that each layer of the decoder has a feedback path, we continue to embed the feedback mechanism into the second layer of the decoder to establish a second feedback path (named Two-P in Table 2), which further increases the value of PSNR from 32.79 dB to 32.83 dB. It proves that the second layer of the decoder is also important to improve the reconstruction quality, due to the fact that incorrect features (such as checkerboard artifact [13]) may occur in any layer of the decoder. Finally, we replace the integrated method from [16] with the proposed FAM for both of the feedback paths, and the PSNR significantly increases by 0.12 dB while the number of parameters (Params in Table 2) decreases by 0.9 M, which demonstrates the advantage of FAM in both deblurring performance and computational efficiency.

4.3.2 Ablation Study of DGAM

Experiments are also conducted to evaluate the effectiveness of DGAM. To avoid redundant convolutions, we introduce the learnable $H \times W \times C$ matrix ω to replace conventional convolution, which increases PSNR from 31.87 dB to 32.95 dB while using the same number of parameters, as shown in Table 3. When capturing the long-range dependencies twice in the frequency domain (named Two-T in Table 3), the PSNR further increases by 0.46 dB with the same number of parameters compared with the dependencies captured only once (named One-T in Table 3), demonstrating the benefit of the dual interactions to capture long-range dependencies. Meanwhile, to study the effect of number of channels, we tune the two hyperparameters r_1 and r_2. As shown in Table 4, the best PSNR is obtained when $r_1 = 0.5$ and $r_2 = 1$. When $r_1 = 1$ and $r_2 = 1$, the PSNR decreases by 0.51 dB, which proves that increasing the number of channels degrades the performance because of the redundancy in features at different channels.

5 Conclusion

In this study, we propose a feedback-mechanism-based network for image deblurring, which contains two novel components: a feedback-mechanism-based decoder and a dual gate attention module. To improve the image reconstruction quality, we introduce a feedback-mechanism-based decoder, which efficiently extracts useful features in the feedback path via the proposed feedback attention module. To reduce the computational cost, the proposed DGAM introduces tunable hyperparameters to reduce channel numbers and avoid channel redundancy. Furthermore, a learnable matrix is leveraged to replace the costly convolutions and capture long-range dependencies in the frequency domain, and gated attention is performed twice to improve the performance. Experiments on public datasets show that our method achieves better deblurring performance while requiring less computational resource compared with the state-of-the-art methods.

Despite the excellent deblurring performance, our approach is still sub-optimal in terms of running time as the feedback-mechanism-based decoder requires reconstructing the deblurred image twice. In the future, we will try to design more efficient network structures for the feedback-mechanism-based decoder. In addition, we will also apply the feedback-mechanism-based decoder to other vision tasks, such as image dehazing and deraining, to explore its generalization capability to other image processing problems.

Acknowledgements This research was supported in part by Natural Science Foundation of Fujian Province (No. 2022J01952, 2023J01953), Open Fund Project of Fujian Key Laboratory of Spatial Information Perception and Intelligent Processing (Yango University) (No. FKLSIPIP1005). Scientific research development fund of Fujian University of Technology (No. GY-Z23078, GY-Z23048).

Data Availability The datasets that support the findings of this study are available in the public domain: 1. GoPro, https://seungjunnah.github.io/Datasets/gopro/ 2. RealBlur, http://cg.postech.ac.kr/research/realblur/

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Zheng A, Zhang Y, Zhang X, et al (2022) Progressive end-to-end object detection in crowded scenes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 857–866, https://doi.org/10.1109/cvpr52688.2022.00093
2. Kim N, Kim D, Lan C, et al (2022) Restr: Convolution-free referring image segmentation using transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 18145–18154, https://doi.org/10.1109/cvpr52688.2022.01761
3. Xu L, Zheng S, Jia J (2013) Unnatural l0 sparse representation for natural image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1107–1114, https://doi.org/10.1109/cvpr.2013.147
4. Pan J, Sun D, Pfister H, et al (2016) Blind image deblurring using dark channel prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1628–1636, https://doi.org/10.1109/cvpr.2016.180
5. Nah S, Hyun Kim T, Mu Lee K (2017) Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3883–3891, https://doi.org/10.1109/cvpr.2017.35
6. Gao H, Tao X, Shen X, et al (2019) Dynamic scene deblurring with parameter selective sharing and nested skip connections. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3848–3856, https://doi.org/10.1109/cvpr.2019.00397
7. Wu Y, Hong C, Zhang X et al (2021) Stack-based scale-recurrent network for face image deblurring. Neural Process Lett 53:4419–4436. https://doi.org/10.1007/s11063-021-10604-9
8. Zhao Q, Zhou D, Yang H (2022) Cdmc-net: context-aware image deblurring using a multi-scale cascaded network. Neural Process Lett. https://doi.org/10.1007/s11063-022-10976-6
9. Zhang H, Dai Y, Li H, et al (2019) Deep stacked hierarchical multi-patch network for image deblurring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5978–5986, https://doi.org/10.1109/cvpr.2019.00613
10. Zamir SW, Arora A, Khan S, et al (2021) Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 14821–14831, https://doi.org/10.5201/ipol.2023.446
11. Tang K, Xu D, Liu H et al (2021) Context module based multi-patch hierarchical network for motion deblurring. Neural Process Lett 53:211–226. https://doi.org/10.1007/s11063-020-10370-0
12. Park D, Kang DU, Kim J, et al (2020) Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In: European conference on computer vision. Springer, pp 327–343, https://doi.org/10.1007/978-3-030-58539-6_20
13. Bhat G, Danelljan M, Van Gool L, et al (2021) Deep burst super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 9209–9218, https://doi.org/10.1109/cvpr46437.2021.00909
14. Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17(3):455–462. https://doi.org/10.1038/nn.3635
15. Tsuda H, Shibuya E, Hotta K (2020) Feedback attention for cell image segmentation. In: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16. Springer, pp 365–379, https://doi.org/10.1007/978-3-030-66415-2_24
16. Girum KB, Créhange G, Lalande A (2021) Learning with context feedback loop for robust medical image segmentation. IEEE Trans Med Imaging 40(6):1542–1554. https://doi.org/10.1109/tmi.2021.3060497
17. Li Z, Yang J, Liu Z, et al (2019) Feedback network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3867–3876, https://doi.org/10.1109/access.2022.3142510
18. Liu X, Chen S, Song L et al (2022) Self-attention negative feedback network for real-time image super-resolution. J King Saud Univ-Comput Inf Sci 34(8):6179–6186. https://doi.org/10.1016/j.jksuci.2021.07.014

19. Purohit K, Rajagopalan AN (2020) Region-adaptive dense network for efficient motion deblurring. In: Proceedings of the AAAI conference on artificial intelligence, pp 11882–11889, https://doi.org/10.1609/aaai.v34i07.6862
20. Zhang Y, Li W, Li Z et al (2021) Dual attention per-pixel filter network for spatially varying image deblurring. Digital Signal Process 113:103008. https://doi.org/10.1016/j.dsp.2021.103008
21. Zhou L, Zhu M, Xiong D et al (2023) RNLFNet: residual non-local Fourier network for undersampled MRI reconstruction. Biomed Signal Process Control 83:104632. https://doi.org/10.1016/j.bspc.2023.104632
22. Huang X, He J (2023) Fusing convolution and self-attention parallel in frequency domain for image deblurring. Neural Process Lett. https://doi.org/10.1007/s11063-023-11228-x
23. Yu C, Xiao B, Gao C, et al (2021) Lite-hrnet: A lightweight high-resolution network. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 10440–10450, https://doi.org/10.1109/cvpr46437.2021.01030
24. Han K, Wang Y, Tian Q, et al (2020) Ghostnet: More features from cheap operations. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1580–1589, https://doi.org/10.1109/cvpr42600.2020.00165
25. Rao Y, Zhao W, Zhu Z et al (2021) Global filter networks for image classification. Adv Neural Inf Process Syst 34:980–993. https://doi.org/10.35925/j.multi.2020.1.7
26. Rao Y, Zhao W, Tang Y et al (2022) Hornet: efficient high-order spatial interactions with recursive gated convolutions. Adv Neural Inf Process Syst 35:10353–10366. https://doi.org/10.1080/03050629.2022.2031182
27. Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need. Adv Neural Inf Process Syst. https://doi.org/10.48550/arXiv.1706.03762
28. Zamir SW, Arora A, Khan S, et al (2022) Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5728–5739, https://doi.org/10.48550/arXiv.2111.09881
29. Wang Z, Cun X, Bao J, et al (2022) Uformer: A general u-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 17683–17693, https://doi.org/10.48550/arXiv.2106.03106
30. Chen L, Chu X, Zhang X, et al (2022) Simple baselines for image restoration. In: European conference on computer vision. Springer, pp 17–33, https://doi.org/10.48550/arXiv.2204.04676
31. Guo J, Zou X, Chen Y, et al (2023) AsConvSR: Fast and Lightweight Super-Resolution Network with Assembled Convolutions. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1582–1592
32. Cho SJ, Ji SW, Hong JP, et al (2021) Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4641–4650, https://doi.org/10.1109/iccv48922.2021.00460
33. Rim J, Lee H, Won J, et al (2020) Real-world blur dataset for learning and benchmarking deblurring algorithms. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16. Springer, pp 184–201, https://doi.org/10.1007/978-3-030-58595-2_12
34. Yuan Y, Su W, Ma D (2020) Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3555–3564, https://doi.org/10.1109/cvpr42600.2020.00361
35. Tao X, Gao H, Shen X, et al (2018) Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8174–8182, https://doi.org/10.1109/cvpr.2018.00853
36. Kupyn O, Budzan V, Mykhailych M, et al (2018) Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8183–8192, https://doi.org/10.1109/cvpr.2018.00854
37. Kupyn O, Martyniuk T, Wu J, et al (2019) Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 8878–8887, https://doi.org/10.1109/iccv.2019.00897
38. Zhang K, Luo W, Zhong Y, et al (2020) Deblurring by realistic blurring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2737–2746, https://doi.org/10.1109/cvpr42600.2020.00281
39. Tsai FJ, Peng YT, Tsai CC et al (2022) Banet: a blur-aware attention network for dynamic scene deblurring. IEEE Trans Image Process 31:6789–6799. https://doi.org/10.1109/tip.2022.3216216
40. Chen L, Lu X, Zhang J, et al (2021) Hinet: Half instance normalization network for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 182–192, https://doi.org/10.1109/cvprw53098.2021.00027

41. Mou C, Wang Q, Zhang J (2022) Deep generalized unfolding networks for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 17399–17410, https://doi.org/10.1109/cvpr52688.2022.01688
42. Lee H, Choi H, Sohn K, et al (2022) KNN local attention for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2139–2149, https://doi.org/10.1109/cvpr52688.2022.00218
43. Suin M, Purohit K, Rajagopalan AN (2020) Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3606–3615, https://doi.org/10.1109/cvpr42600.2020.00366
44. Tu Z, Talebi H, Zhang H, et al (2022) Maxim: Multi-axis mlp for image processing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5769–5780, https://doi.org/10.1109/cvpr52688.2022.00568
45. Chen H, Wang Y, Guo T, et al (2021) Pre-trained image processing transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 12299–12310, https://doi.org/10.18653/v1/2020.sdp-1.38