

Research on Motion Image Deblurring via Spatial-Frequency Dual-Domain Information Fusion

Jiahao Wang

Qinghai University for Nationalities

Qing Qi

qiqing@tju.edu.cn

Qinghai University for Nationalities

Kaiqiang Zhang

Qinghai University for Nationalities

Research Article

Keywords: Spatial domain information, Frequency domain information, Multi domain feature fusion, Motion image deblurring

Posted Date: August 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10074836/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Research on Motion Image Deblurring via Spatial–Frequency Dual-Domain Information Fusion

Jiahao Wang¹, Qing Qi^{1,2*}, Kaiqiang Zhang¹

^{1*}College of Intelligent Science and Engineering, Qinghai Minzu University, Xining, 810007, Qinghai, China.

²National Experimental Teaching Demonstration Center of Communication Engineering, Qinghai Minzu University, Xining, 810007, Qinghai, China.

*Corresponding author(s). E-mail(s): qiqing@tju.edu.cn;

Abstract

Motion image deblurring remains challenging due to the inherent trade-off between global structure modeling and local detail preservation. Spatial-domain methods excel at capturing fine-grained textures but struggle with long-range dependencies, whereas frequency-domain approaches effectively model global blur patterns yet often introduce artifacts such as ringing in complex motion scenarios. This reveals a fundamental tension between global frequency modeling and spatial detail fidelity. To address this issue, we propose a spatial-frequency dual-domain framework that explicitly exploits and coordinates the complementary properties of both domains. Specifically, we construct a balanced dual-branch architecture to investigate their representational differences, revealing that frequency-domain features inherently suit long-range dependency modeling but lack spatial consistency. Based on this insight, we design a constrained dynamic fusion mechanism that adaptively integrates spatial and frequency features while preventing either branch from dominating. This enables effective coordination between global structure reconstruction and local detail preservation. Extensive experiments demonstrate the effectiveness of the proposed approach. On GoPro, our method achieves 31.08 dB PSNR and 0.915 SSIM; on RealBlur-J and RealBlur-R, it achieves 29.98 dB/0.879 and 35.93 dB/0.959, respectively, while effectively suppressing ringing artifacts and producing visually consistent results.

Keywords: Spatial domain information, Frequency domain information, Multi domain feature fusion, Motion image deblurring

1 Introduction

Motion-induced image blur is a common phenomenon in real-world scenarios. It is primarily caused by the relative motion between the camera and the scene during exposure, which results in blurred light trajectories on the imaging sensor [1]. Such degradation not only reduces visual quality but also poses significant challenges for downstream image analysis and vision-based applications.

The core objective of image deblurring is to reverse the blur formation process, which involves both the restoration of fine-grained textures and the reconstruction of global structural information [2]. The former corresponds to high-frequency components in the spatial domain, while the latter depends on global geometric consistency [3]. Spatial-domain methods typically exhibit low computational complexity; however, they are constrained by limited receptive fields. In contrast, frequency-domain approaches transform images into global frequency representations, enabling efficient modeling of long-range dependencies and effective characterization of blur kernels. Nevertheless, purely frequency-based methods often struggle to maintain spatial consistency in regions with severe or complex motion and are prone to ringing artifacts.

In summary, relying solely on either spatial-domain or frequency-domain techniques makes it difficult to simultaneously achieve high fidelity in local details and accurate global structure reconstruction. Effectively integrating the complementary advantages of both domains—preserving local detail representation while leveraging global modeling capability and suppressing artifacts—remains a critical challenge in motion deblurring.

1.1 Spatial Domain Image Deblurring Methods

Traditional image deblurring approaches [4–6] are primarily based on physical modeling and deconvolution theory. These methods reconstruct sharp images by estimating the blur kernel and performing inverse operations. However, their performance heavily depends on the accuracy of kernel estimation and degrades significantly in the presence of complex motion or nonlinear blur. To alleviate this issue, Gong et al. [7] proposed a motion flow estimation method that indirectly models the blur kernel by estimating pixel-wise motion. Despite this improvement, the final restoration process still relies on conventional deconvolution techniques.

With the rapid development of deep learning, image deblurring has shifted toward end-to-end learning paradigms. Spatial-domain deep learning methods have become dominant due to their ability to directly learn mappings from blurry to sharp images. These methods can be broadly categorized into three groups:

(1) Convolutional Neural Network(CNN) based methods [8–11] feature simple architectures and efficient computation, making them effective at capturing local texture information. However, their inherently limited receptive fields restrict their ability to model long-range dependencies and global context.

(2) Generative Adversarial Network(GAN) based methods [12–14] improve perceptual quality through adversarial training and exhibit strong generalization capability. Nevertheless, they often suffer from training instability and mode collapse, and still struggle to fully exploit global information.

(3) Transformer [15] based methods [16–18] leverage self-attention mechanisms to achieve powerful global modeling capabilities, alleviating the limitations of CNNs [19] in capturing long-range dependencies [20]. However, the quadratic computational complexity of self-attention leads to high computational cost for high-resolution images, and these methods may still fall short in modeling local blur patterns and recovering fine details.

Overall, spatial-domain approaches face challenges in balancing global modeling and local detail preservation, as well as achieving a trade-off between performance and computational efficiency.

1.2 Frequency Domain Image Deblurring Methods

To address the limitations of spatial-domain networks in global modeling, frequency-domain-based deblurring methods have been extensively investigated. Early approaches [21] relied on Fourier transform and hand-crafted filtering strategies for deconvolution, but they lack robustness and generalization in real-world scenarios.

With the integration of deep learning, modern frequency-enhanced networks abandon manual priors and incorporate frequency transformations into neural architectures. For example, FFTformer [22] performs attention computation in the frequency domain to reduce complexity and adaptively filter multi-frequency components. LoFormer [23] further enhances frequency feature representation by combining discrete cosine transform with window-based self-attention, leading to improved detail restoration.

In addition, some studies incorporate physical priors into frequency learning to improve cross-scene generalization [24, 25]. Wavelet-based methods [26, 27] decompose images into different frequency bands, enabling targeted learning of high- and low-frequency features. Despite these advances, most existing approaches adopt a spatial-dominant framework where frequency information serves as auxiliary enhancement. As a result, spatial and frequency domains are not treated as equally important branches, limiting the effectiveness of cross-domain interaction and often leading to ringing artifacts in complex scenarios.

1.3 Multi-Domain Image Deblurring Methods Combining Spatial and Frequency Information

Recent advances in image deblurring have highlighted the limitations of single-domain methods in handling complex real-world scenarios. To overcome these challenges,

multi-domain approaches that integrate spatial and frequency information have been proposed.

Given that spatial-domain methods excel at local detail modeling while frequency-domain methods are effective for global representation, constructing a dual-branch architecture that processes both domains in parallel has emerged as a promising direction. This paradigm enables complementary feature learning and has become a key focus of recent research.

Existing works introduce techniques such as multi-scale architectures [28], frequency gating mechanisms [3, 29], and dual-stream learning frameworks [30] to enhance cross-domain interaction and improve restoration performance. Furthermore, adaptive frequency decomposition [31] and lightweight parallel spatial-frequency structures [32] have been proposed to improve fusion flexibility and computational efficiency.

Although existing dual-branch approaches have demonstrated promising spatial–frequency collaboration, several critical limitations remain. Most studies primarily focus on architectural design, while lacking a systematic understanding of the underlying mechanisms of frequency-domain modeling, particularly its role in capturing long-range dependencies. Moreover, existing fusion strategies often employ fixed weighting, lacking the ability to dynamically adjust the contribution of spatial and frequency features across network depth and spatial locations. In addition, insufficient attention has been paid to the artifacts introduced by frequency-domain processing, such as ringing effects in complex motion regions.

To address these issues, this paper investigates both the mechanism of frequency-domain long-range dependency modeling and the design of effective cross-domain fusion strategies for motion image deblurring. The main contributions are summarized as follows:

(1) We construct a balanced dual-branch framework with strictly comparable single-branch baselines to systematically analyze the representational differences between spatial and frequency domains. The study reveals that frequency-domain features are inherently effective in capturing global dependencies, while exhibiting limitations in maintaining spatial consistency, thereby providing a principled foundation for cross-domain fusion.

(2) We propose a Dual-Domain Dynamic Fusion Block (DDFB) that adaptively integrates spatial and frequency features through a learnable gating mechanism with explicit constraints. By dynamically adjusting the contribution of each domain across different network stages, the proposed method effectively balances local detail restoration and global structure reconstruction, leading to robust and high-quality deblurring performance in complex motion scenarios.

2 Proposed method

2.1 Overall Structure of the GAN-Based Model

This paper adopts a GAN [33] framework, where the proposed dual-branch deblurring network serves as the generator. This dual-branch architecture shows great potential for handling complex motion blur [34]. As the core component of this study,

the generator is designed based on a U-Net [35] backbone, within which two parallel branches are embedded: a spatial-domain convolution branch and a learnable Fourier transform (FFT) branch. The spatial branch focuses on capturing local texture information, while the frequency branch models global motion spectra. The discriminator is employed solely for adversarial training, and all subsequent experiments are conducted based on the generator to validate the effectiveness of the proposed frequency-domain mechanisms and fusion strategies.

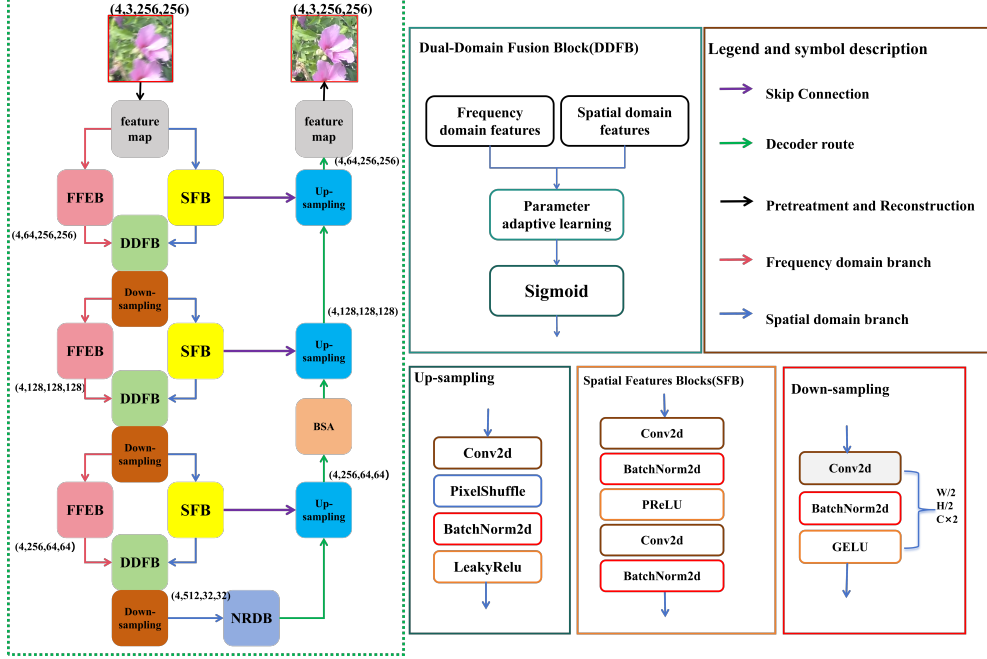


Fig. 1 Generator structure diagram

The overall architecture of the generator is illustrated in Fig. 1. It follows a U-Net encoder-decoder structure, leveraging its effectiveness in image generation and restoration tasks. The dimensional annotations in Fig. 1 indicate the output feature sizes of each module.

The encoder is responsible for extracting deep features from blurry input images through stacked convolutional layers and nonlinear activation functions. It consists of three down-sampling modules, which progressively reduce the spatial resolution of feature maps while capturing increasingly abstract and global representations. These operations effectively enlarge the receptive field, enabling the network to learn broader contextual information that is essential for motion blur understanding and removal.

Correspondingly, the decoder restores spatial resolution via a series of up-sampling operations. Specifically, three up-sampling modules are employed, with PixelShuffle serving as the primary up-sampling technique. This operation rearranges feature maps

from the channel dimension to the spatial dimension, thereby increasing resolution and facilitating the recovery of high-frequency details.

The generator adopts a dual-branch design comprising two parallel feature processing streams. The spatial feature processing branch (blue solid lines in Fig. 1) focuses on extracting spatial-domain information, including textures, edges, and structural details, which are effectively captured by convolutional operations and provide a strong foundation for image restoration. In parallel, the frequency feature processing branch (red solid lines in Fig. 1) extracts frequency-domain representations through Fourier transform operations, enabling the modeling of global blur patterns.

By jointly leveraging spatial and frequency representations, the proposed dual-branch architecture captures complementary information from both domains, leading to more comprehensive feature representations for subsequent fusion and reconstruction, and ultimately improving motion image deblurring performance.

2.2 Spatial Feature Extraction Module

The spatial-domain feature block (SFB) is designed to capture local textures and structural information by performing convolution directly in the pixel domain. This enables effective modeling of edges, textures, and fine-grained local patterns.

As illustrated in Fig. 1, the SFB consists of stacked residual blocks, where each block serves as a fundamental unit for feature extraction. Each residual block contains two convolutional layers, each followed by Batch Normalization and Parametric ReLU activation. Through residual connections—where the input is directly added to the output—the blocks effectively alleviate the vanishing gradient problem and enhance feature propagation and reuse. The formulation of a residual block can be expressed as:

$$\mathcal{B}(x) = x + \text{BN}(\text{Conv}(\text{PReLU}(\text{BN}(\text{Conv}(x_0)))) \quad (1)$$

$$\mathcal{S}(x) = \mathcal{B}^5(x) \quad (2)$$

Where $\mathcal{B}(\{x\})$ represents a residual block. BN denotes Batch Normalization. PReLU is the activation function. Conv indicates the convolution operation, specifically 3×3 convolution. $\mathcal{S}(\{x\})$ represents the output of the SFB module. This output is formed by stacking five residual blocks. This stacking number of five is empirically determined through extensive experiments to be optimal for feature extraction.

2.3 Frequency Domain Feature Extraction Module

The frequency-domain feature extraction block (FFEB), illustrated in Fig. 2, is designed to model global dependencies and motion blur patterns in the frequency domain. The input features are first transformed into the frequency domain via a 2D Fourier transform.

In the frequency domain, a learnable Fourier kernel is introduced to adaptively reweight frequency components through element-wise multiplication. Specifically, the complex spectrum of each channel is decomposed into polar form, and two learnable matrices are employed to modulate magnitude and phase, respectively:

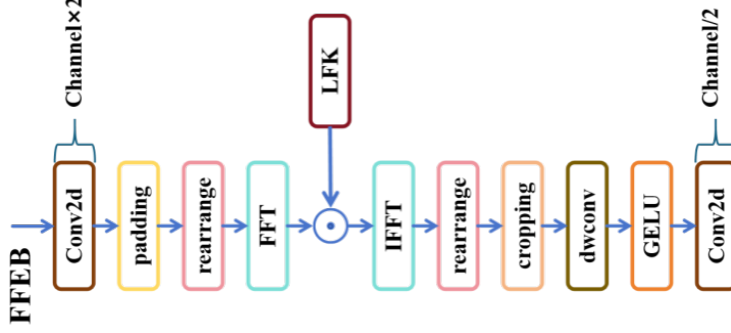


Fig. 2 Frequency Feature Extraction Block

$$\tilde{F}_c = W_{amp} \odot |F_c| \cdot e^{i(\angle F_c + W_{phase})} \quad (3)$$

Because complex weights must adjust both magnitude and phase, updating a single parameter easily couples the two dimensions and causes gradient fluctuation; hence W_{amp} is initialized to an all-one matrix and W_{phase} to an all-zero matrix, so the block starts as an identity mapping. During training, the loss that optimizes edge sharpness drives W_{amp} toward 0 in low-frequency regions and above 1 in high-frequency regions, achieving low-frequency suppression and high-frequency amplification. The phase matrix W_{phase} aligns the phases of textures with the same orientation, so that after the inverse transform edge structures add coherently while random phase scatter that would blur edges is cancelled. To compress the parameter count from $\mathcal{O}(H \cdot W)$ to $\mathcal{O}(H + W)$, W_{amp} and W_{phase} are factorized into separable outer products of vectors:

$$\begin{cases} W_{amp,c} = u_c v_c^T \quad (u_c \in \mathbb{R}^H, v_c \in \mathbb{R}^W, W_{amp,c} \in \mathbb{R}^{H \times W}) \\ W_{phase,c} = p_c q_c^T \quad (p_c \in \mathbb{R}^H, q_c \in \mathbb{R}^W, W_{phase,c} \in \mathbb{R}^{H \times W}) \end{cases} \quad (4)$$

Via the outer-product factorization of Eq.(4), each channel reconstructs the entire $H \times W$ frequency mask using only two vectors of length H and W , markedly reducing memory and computation. During training, u, v, p, q are updated end-to-end together with other convolution weights; no extra optimizer is required. After the inverse transform, boosting high-frequency energy and suppressing low-frequency energy in the image spectrum highlights edges and textures while suppressing smooth backgrounds and random noise. This achieves the goal of preserving global structure while enhancing spatial details. After processing, FFEB converts the features back to the spatial domain by inverse Fourier transform, thereby accomplishing image feature extraction and enhancement. The mathematical expressions are as shown in equations (5) and (6):

$$x_{fft} = \mathcal{F}^{-1}(\mathcal{F}(x) \odot \mathcal{K}) \quad (5)$$

$$x = \mathcal{P}_{out}(\text{GELU}(\mathcal{P}(\mathcal{D}(x_{fft})))) \quad (6)$$

Where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote 2D fast Fourier transform and inverse transform, respectively. $X \odot Y$ represents element-wise multiplication for weighting frequency

components. $\mathcal{K}$ is the learnable Fourier kernel. $\mathcal{D}$ denotes depthwise separable convolution. $\mathcal{P}$ denotes pointwise convolution. $\mathcal{P}_{out}$ denotes output projection. GELU is the activation function.

Through this structure, the frequency-domain processing module effectively captures global features and frequency information.

2.4 Dual-Domain Dynamic Fusion Block

The dual-domain dynamic fusion block (DDFB) is designed to dynamically fuse spatial and frequency features. For motion image deblurring, such fusion is crucial for restoring image details and global information while leveraging the complementary advantages of both domains to enhance deblurring performance.

As shown in the DDFB part of Fig. 1, the module structure is designed as follows. The core of DDFB is a dynamic fusion mechanism that adaptively adjusts the weights of spatial and frequency features through a learnable gating parameter. Specifically, given spatial feature map X_1 and frequency feature map X_2 as inputs, the module produces a fused feature map by controlling the fusion ratio via the gate. The specific formulation is given below:

$$\alpha = \sigma(\text{gate}) \cdot 0.8 + 0.1 \quad (7)$$

$$\text{output} = \alpha \cdot X_1 + (1 - \alpha) \cdot X_2 \quad (8)$$

Where σ represents a function. It maps the gate to the (0,1) interval. It constrains the value within a range through scaling and shifting. Empirical results from multiple experiments show that the range [0.1, 0.9] works best. This prevents any single feature from completely dominating the fusion process. During training, the gating network is supervised only by loss of $\mathcal{L}_1$ and perceptual loss; no extra labels or branch losses are required. The specific process is as follows: At initialization, the weight tensor of the 1×1 convolution is set to zero and the bias to zero, ensuring $\text{gate} \equiv 0.5$. Both branches then contribute equally, preventing initial bias that would block gradients; During each forward pass, the mean and standard deviation of gate are recorded in real-time. If the global mean deviates from 0.5 by more than 0.3 for five consecutive epochs, the Sigmoid temperature coefficient t is linearly increased from 1.0 to 2.0, forcing the gated curve to tend towards smoothness and preventing one branch from being continuously suppressed. During back-propagation, the gating gradient is computed by pixel-wise differencing, as given in Eq. (9):

$$\frac{\partial L}{\partial \text{gate}} = (X_1 - X_2) \odot \frac{\partial L}{\partial F_{out}} \cdot 0.8 \cdot \sigma(\text{gate}) \cdot (1 - \sigma(\text{gate})) \quad (9)$$

This formulation introduces a competitive relationship between the two branches, where gradients are propagated in opposite directions based on feature differences. As a result, the network dynamically adjusts the contribution of each domain without requiring explicit supervision.

Overall, the DDFB enables pixel-wise adaptive fusion, allowing the network to emphasize spatial details in texture-rich regions while relying more on frequency information in globally blurred areas.

2.5 Nested Residual Dense Block

The Nested Residual Dense Block (NRDB) is designed to enhance feature reuse and representation capacity. It consists of three cascaded residual dense blocks, each employing dense connections to facilitate multi-level feature aggregation.

Within each block, features pass through five sequential 3×3 convolutional layers, where outputs are concatenated along the channel dimension and compressed via a 1×1 convolution. This design promotes hierarchical feature reuse and enriches local receptive fields.

An outer residual connection integrates shallow and deep features, further improving information flow. Notably, the NRDB operates without additional downsampling, preserving spatial resolution and ensuring consistent feature representation for subsequent upsampling.

2.6 Bi-Stripe Attention

To address the directionally variant nature of real-world motion blur, we introduce a Bi-Stripe Attention (BSA) module in the upsampling stage. The goal is to recover missing directional structures and enhance long-range dependency modeling along dominant blur trajectories.

The BSA consists of two parallel attention branches operating along horizontal and vertical directions, respectively. Each branch reshapes the feature map into stripe sequences and applies multi-head self-attention to model long-range dependencies along a single spatial axis.

This design enables efficient directional modeling without increasing channel complexity, effectively restoring structured textures that are difficult to recover using isotropic convolution alone.

2.7 Spatial only branch model and frequency-domain only branch model

To quantitatively evaluate the contribution of each domain, two single-branch baseline models are constructed based on the proposed framework.

As shown in Fig. 3, Baseline-S retains only the spatial branch, removing all frequency-domain operations, resulting in a standard U-Net architecture. Baseline-F, in contrast, removes spatial convolutional components and relies solely on frequency-domain processing.

Both models share identical backbone structures, channel configurations, and training settings, ensuring a fair comparison. These baselines provide a controlled experimental setup for analyzing the individual effectiveness of spatial and frequency mechanisms.

2.8 Loss Function

The proposed model is trained using a combination of reconstruction, perceptual, edge, and adversarial losses. The overall objective is defined as:

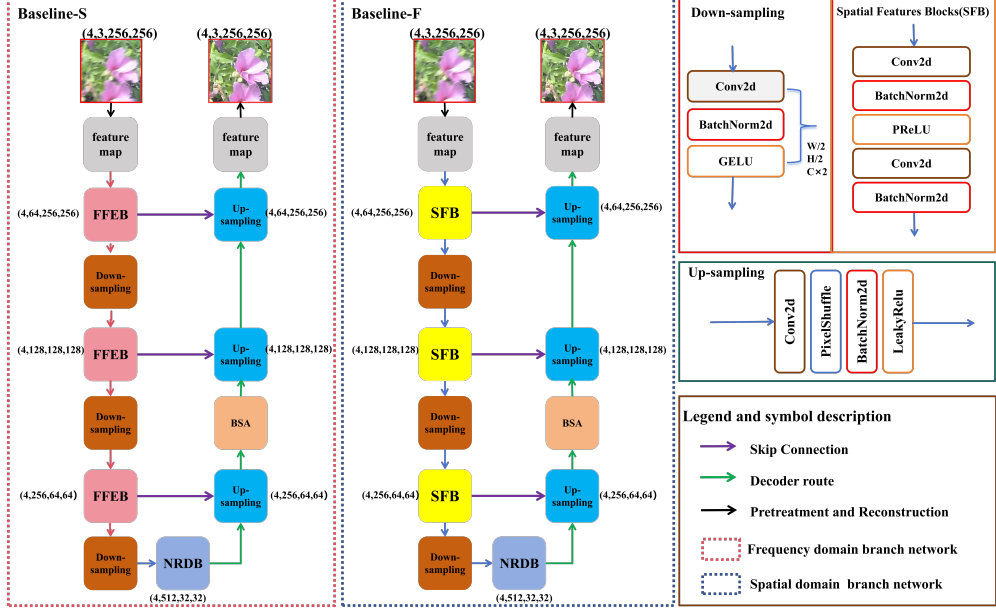


Fig. 3 Network structure diagram of only spatial domain branch and only frequency domain branch

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{rec}} + \lambda_2 \mathcal{L}_{\text{prec}} + \lambda_3 \mathcal{L}_{\text{edge}} + \lambda_4 \mathcal{L}_{\text{adv}} \quad (10)$$

where $\mathcal{L}_{\text{rec}}$ denotes the pixel-wise reconstruction loss ($\mathcal{L}_1/\mathcal{L}_2$) between the generated image and the ground truth, $\mathcal{L}_{\text{prec}}$ is the perceptual loss computed on feature maps extracted from a pre-trained VGG network [13], and $\mathcal{L}_{\text{edge}}$ enforces consistency of image gradients to preserve edge sharpness. The adversarial loss $\mathcal{L}_{\text{adv}}$ is implemented based on the WGAN-GP [36] formulation to improve visual realism and stabilize training.

The weighting coefficients λ_1 , λ_2 , λ_3 , and λ_4 are empirically determined to balance reconstruction accuracy and perceptual quality. Specifically, we set $\lambda_1 = 10$, $\lambda_2 = 10$, $\lambda_3 = 12$, and $\lambda_4 = 1$.

3 Experimental Results and Analysis

3.1 Training Strategy and Implementation Detail

The experimental environment is summarized in Table 1. All models are implemented in PyTorch and trained on an NVIDIA RTX 4070Ti Super GPU (16 GB VRAM) under Ubuntu 20.04.

Training hyperparameters are listed in Table 2. The model is trained for 300 epochs using the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate is set to 1×10^{-4} for the first 150 epochs and then linearly decayed to zero over

Table 1 Experimental Environment

Basic structure	Version
CPU	Intel i5-12600K (16 cores, 3.7GHz)
GPU	NVIDIA RTX 4070Ti Super (16GB VRAM)
operating platform	Ubuntu 20.04
development language	Python 3.10
Training Framework	PyTorch 2.5.0
Parallel computing mechanism	CUDA 12.4

the remaining epochs. The batch size is set to 4, and the input images are randomly cropped to 256×256 patches.

Table 2 Experimental Parameters

Experimental parameters	Parameter value
epoch	300
batch	4
image size	256×256
Adam parameter β_1	0.5
Adam parameter β_2	0.999

During training, the generator is updated once per iteration, while the discriminator is updated five times to ensure stable adversarial learning. The model is trained on the GoPro dataset in a supervised manner, where the objective is to minimize the discrepancy between the restored images and corresponding ground-truth sharp images.

3.2 Comparison between the Model and the Classic Model

To evaluate the effectiveness of the proposed method, we report PSNR and SSIM results on GoPro [8], RealBlur-J [37], and RealBlur-R [37], as summarized in Table 3. All compared models are trained and tested using official dataset splits. Params and FLOPs are measured at an input resolution of 1280×720 . We adopt a task-adaptive configuration: the Full Model is employed for GoPro, and the Full Model with BSA is employed for RealBlur-J and RealBlur-R (see Sec. 3.3.5 for ablation justification).

Quantitatively, our method achieves 31.08 dB / 0.915 on GoPro, 29.98 dB / 0.879 on RealBlur-J, and 35.93 dB / 0.959 on RealBlur-R. On the real-world RealBlur-J dataset, the proposed method attains the highest PSNR among all compared approaches. It surpasses the transformer-based methods DeblurDiNAT-L [18] by 1.00 dB, Restormer [43] by 1.02 dB, and FFTformer [22] by 2.23 dB. It also exceeds the CNN-based method MPRNet [40] by 1.28 dB and all GAN-based competitors by a considerable margin. This demonstrates that the spatial-frequency collaborative design generalizes effectively to real-world scenes with complex, non-uniform blur patterns.

Table 3 Quantitative Comparison of the Latest Methods Using PSNR and SSIM

Methods	Category	GOPRO		RealBlur-J		RealBlur-R		Performance Params(M)	Efficiency FLOPs(G)
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM		
DeblurGAN [12]	GAN	28.23	0.859	27.97	0.834	33.79	0.903	44.0	678.29
DeblurGAN-v2 [13]	GAN	29.55	0.934	28.70	0.866	35.26	0.944	54.0	411.34
DBGAN [38]	GAN	31.10	0.942	24.93	0.745	33.78	0.909	—	—
Ghost-DeblurGAN [39]	GAN	28.75	0.919	—	—	—	—	6.1	20.51
FCL-GAN [14]	GAN	24.84	0.771	25.35	0.736	28.37	0.663	24.6	—
MPRNet [40]	CNN	32.66	0.959	28.70	0.873	35.99	0.952	23.0	777.01
MIMOU-Net+ [11]	CNN	32.45	0.957	27.63	0.873	35.54	0.947	16.1	—
MT-RNN [41]	RNN	31.15	0.945	28.44	0.862	35.79	0.951	—	—
RNN-MBP [42]	RNN	33.32	0.963	—	—	—	—	—	—
Restormer [43]	Transformer	32.92	0.961	28.96	0.879	36.19	0.957	26.1	140.99
FFTformer [22]	Transformer	34.21	0.969	27.75	0.853	35.87	0.953	16.6	131.45
DeblurDiNAT-L [18]	Transformer	33.63	0.967	28.98	0.885	36.09	0.955	16.1	152.61
Ours*	GAN	31.08	0.915	29.98	0.879	35.93	0.959	19.5	112.04

*We adopt a task-adaptive configuration: the Full Model is employed for GoPro, and the Full Model with BSA is employed for RealBlur-J and RealBlur-R.

On RealBlur-R, our method achieves 35.93 dB PSNR and 0.959 SSIM, ranking among the top-performing methods. It exceeds all GAN-based approaches and most CNN-based and RNN-based competitors. Specifically, it surpasses MIMO-UNet+ [11] by 0.39 dB and MT-RNN [41] by 0.14 dB. While Restormer [43], DeblurDiNAT-L [18], and MPRNet [40] achieve slightly higher PSNR values, our method maintains competitive accuracy with substantially lower computational cost.

On the synthetic GoPro dataset, our method achieves 31.08 dB PSNR, ranking second among all GAN-based methods and slightly below DBGAN [38] by 0.02 dB. However, DBGAN exhibits severe performance degradation on real-world datasets, dropping to 24.93 dB on RealBlur-J and 33.78 dB on RealBlur-R, which indicates poor cross-dataset generalization. In contrast, our method maintains consistent performance across both synthetic and real-world benchmarks. Compared with MT-RNN [41], RNN-MBP [42], Restormer [43], FFTformer [22], and DeblurDiNAT-L [18], our GoPro performance is lower. This is attributed to the relatively uniform global blur patterns in GoPro, where the dual-branch frequency-spatial coordination introduces marginal redundancy compared to pure spatial-domain transformers or RNNs optimized for sequential modeling.

The performance gains on real-world benchmarks can be attributed to the complementary nature of spatial and frequency domain modeling. Motion blur inherently suppresses high-frequency components and disturbs directional structures. The proposed frequency branch explicitly compensates for frequency-domain attenuation through learnable Fourier kernels, enabling effective recovery of global blur patterns and long-range dependencies. Simultaneously, the spatial branch preserves localized

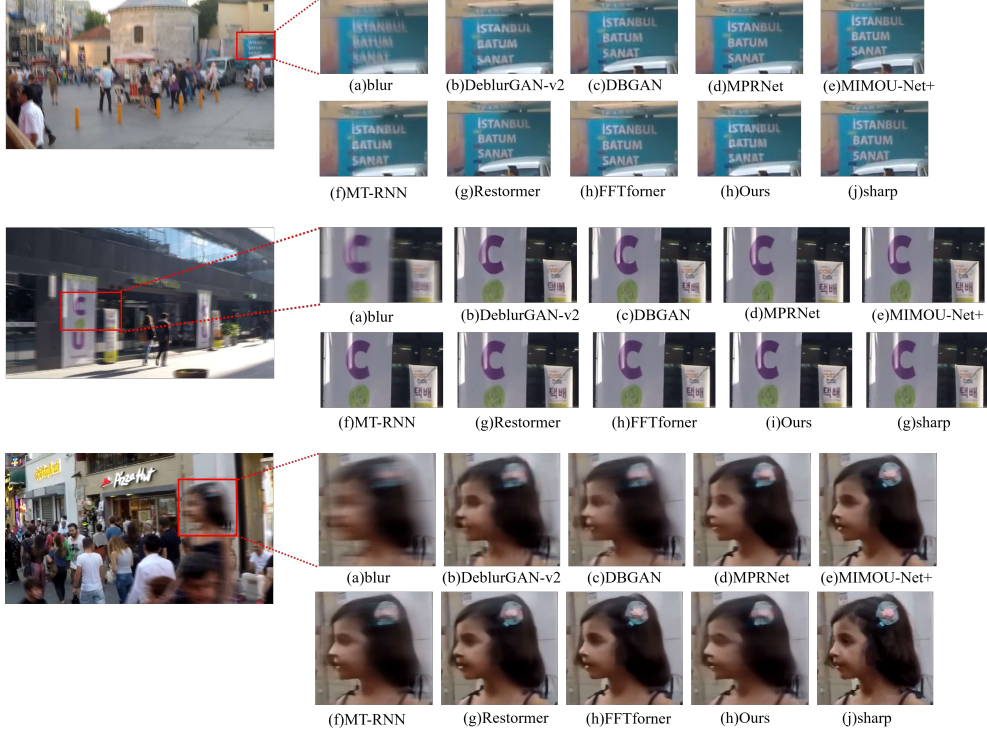


Fig. 4 Comparison of deblurring methods on the GOPRO dataset

textures and sharp edges through residual dense connections. The constrained dynamic fusion mechanism adaptively coordinates these dual-domain representations, preventing either branch from dominating and ensuring that global structure reconstruction and local detail preservation mutually reinforce rather than interfere. Consequently, the method achieves robust generalization across diverse blur patterns.

Qualitative comparisons in Figs. 4–6 further demonstrate that the proposed method preserves sharper edges in dynamic regions, avoids ringing artifacts in smooth areas, and produces more visually natural results than GAN-based competitors and several transformer-based methods.

Overall, the experimental results validate that controlled spatial–frequency collaboration provides an effective and generalizable solution for motion image deblurring, particularly for real-world scenarios with complex, spatially varying blur.

3.3 Ablation Experiment

3.3.1 Module ablation experiment

To verify the effectiveness of each module, we conduct systematic ablation studies on the GoPro training and testing sets. PSNR and SSIM are used as quantitative metrics. Results are reported in Table 4 and Fig. 7.

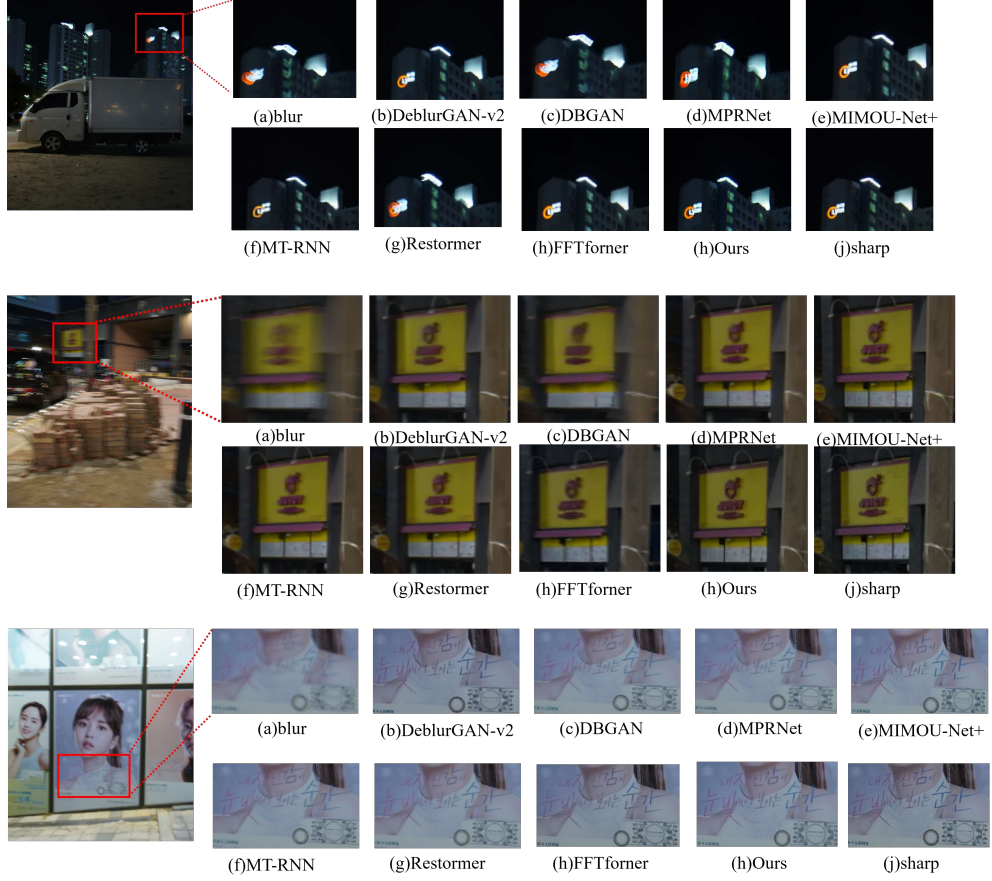


Fig. 5 Comparison of deblurring methods on the RealBlur-J dataset

(a)Baseline. We start from a plain encoder–decoder without any proposed modules (SFB, FFEB, DDFB, NRDB, and BSA disabled), serving as the lower bound at 27.12 dB PSNR / 0.782 SSIM.

(b)+NRDB. Equipping the baseline with the Nested Residual Dense Block (NRDB) improves performance to 28.45 dB, confirming the representational gain of hierarchical dense residual connections.

(c)Baseline-S. Retaining only the spatial branch (SFB + NRDB, without FFEB or fusion) yields 29.46 dB / 0.867 SSIM. This indicates that spatial features effectively capture local details but are limited in modeling global blur patterns.

(d)Baseline-F. Retaining only the frequency branch (FFEB + NRDB, without SFB or fusion) achieves 29.37 dB / 0.864 SSIM, comparable to the spatial branch but 0.09 dB lower. This demonstrates that frequency features provide crucial global cues yet suffer from loss of spatial structural details when used alone.

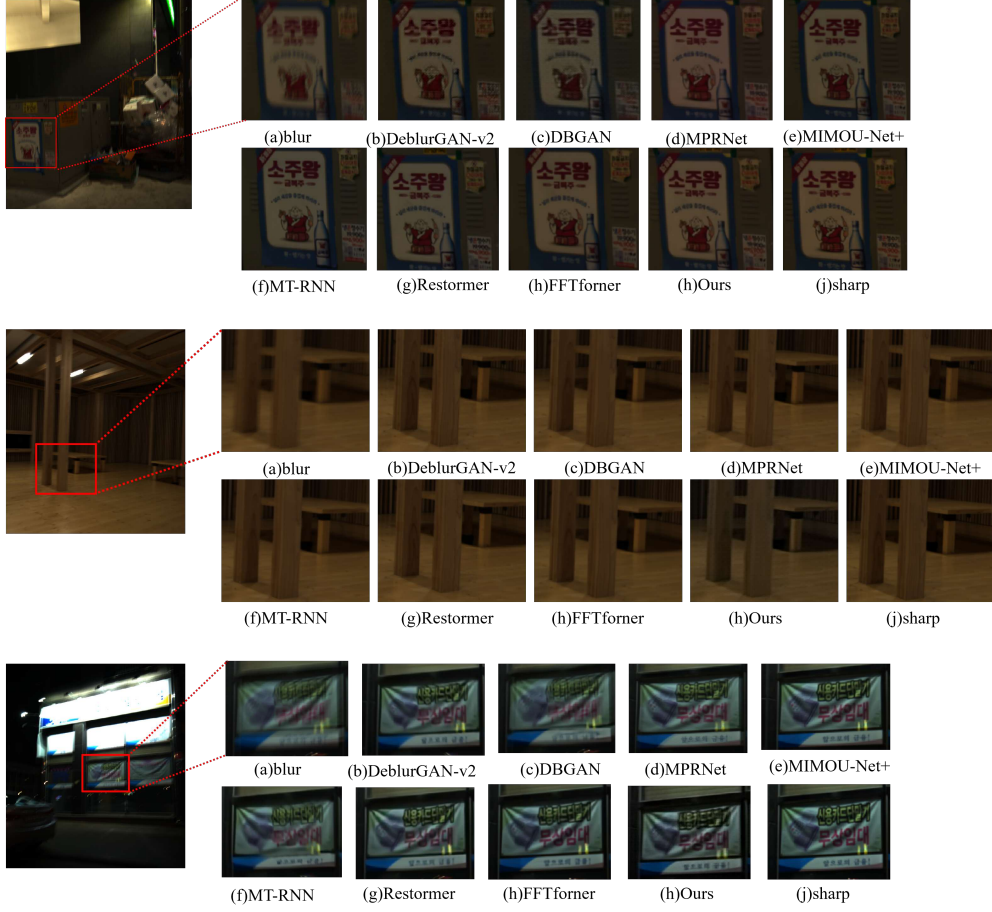


Fig. 6 Comparison of deblurring methods on the RealBlur-R dataset

(e) Simple Fusion. Enabling both SFB and FFEB but replacing DDFB with simple channel-wise concatenation (followed by NRDB) produces 29.48 dB, nearly identical to single-branch results. This reveals that naïvely stacking dual-branch features without effective interaction fails to achieve cross-domain synergy and may induce feature redundancy.

(f) Full Model. Incorporating the complete cross-domain fusion framework (SFB + FFEB + DDFB + NRDB) without BSA achieves 31.08 dB PSNR and 0.915 SSIM on GoPro, yielding gains of 1.62 dB and 0.048 over the naïve fusion variant. This verifies the necessity of the proposed DDFB for dynamic cross-domain coordination. These results demonstrate that spatial and frequency features are inherently complementary: the spatial branch preserves local structure, while the frequency branch suppresses periodic blur. Notably, the performance gap between individual branches is merely 0.09 dB, yet our fusion strategy achieves a superadditive gain exceeding 1.6 dB over the

Table 4 Ablation Study Results on the GoPro Dataset

Model	SFB	FFEB	DDFB	NRDB	BSA	PSNR (dB)	SSIM
(a) Baseline	×	×	×	×	×	27.12	0.782
(b) +NRDB	×	×	×	✓	×	28.45	0.815
(c) Baseline-S	✓	×	×	✓	×	29.46	0.867
(d) Baseline-F	×	✓	×	✓	×	29.37	0.864
(e) Simple Fusion	✓	✓	×	✓	×	29.48	0.867
(f) Full Mode	✓	✓	✓	✓	×	31.08	0.915
(g) Full Model+BSA	✓	✓	✓	✓	✓	30.26	0.891

baseline. This confirms that the DDFB dynamically integrates both representations to produce more discriminative features rather than performing simple feature stacking.

(g) Full Model+BSA. Further integrating the Bi-Stripe Attention (BSA) on top of the full model produces divergent effects across datasets, results are reported in Table 5 and Fig. 8. On GoPro, the PSNR decreases to 30.26 dB and SSIM to 0.891, as the directional stripe attention introduces additional parameters that are not fully compatible with globally uniform synthetic motion blur. Conversely, on RealBlur-J and RealBlur-R, BSA brings substantial gains to 29.98 dB / 0.879 and 35.93 dB / 0.959, respectively, confirming its effectiveness for real-world scenarios with complex, directionally variant blur and texture loss. We therefore adopt a task-adaptive configuration in the final evaluation: the Full Model is used for GoPro, and the Full Model + BSA is used for RealBlur-J and RealBlur-R.

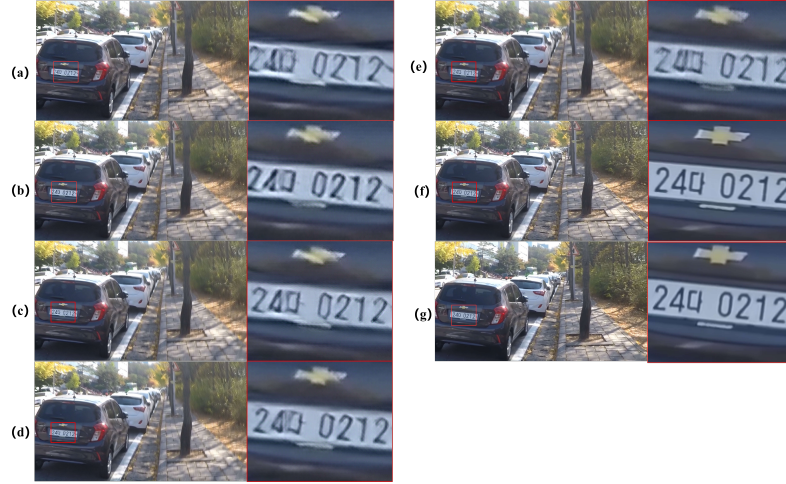
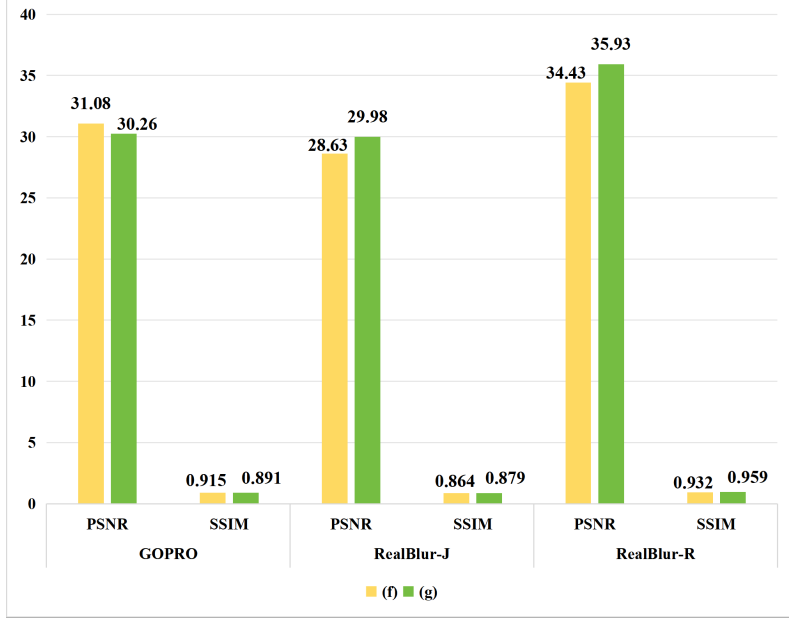
**Fig. 7** Qualitative comparison of ablation configurations

Table 5 Effect of BSA across benchmark datasets

Methods	GOPRO		RealBlur-J		RealBlur-R	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
(f) Full Model	31.08	0.915	28.63	0.864	34.43	0.932
(g) Full Model+BSA	30.26	0.891	29.98	0.879	35.93	0.959

**Fig. 8** Effect of the BSA module across benchmark datasets

3.3.2 Mathematical analysis of the synergistic effect of dual-branch fusion

To quantitatively analyze the collaborative effect of the dual-branch fusion mechanism, we provide an information-theoretic interpretation of the proposed Dual-Domain Dynamic Fusion Block (DDFB). The fusion process is modeled as a gated feature coupling system. Let X_s and X_f denote the spatial-domain and frequency-domain feature representations, respectively, and let $\alpha \in [0, 1]$ denote the dynamically learned gating weight. The fused representation is expressed as:

$$I(F_{out}; Y) \geq \alpha^2 I_{spa} + (1 - \alpha)^2 I_{freq} + \alpha(1 - \alpha) \ln(1 - \rho^2)^{-1} \quad (11)$$

where α controls the contribution of each branch. In this analysis, α is set to 0.5 for simplicity.

This formulation indicates that DDFB performs content-adaptive integration rather than simple concatenation or fixed-weight fusion, enabling effective coordination between spatial detail preservation and frequency-based global structure modeling.

To further characterize the information preservation capability of the fusion process, we adopt a correlation-based approximation inspired by mutual information under Gaussian assumptions. Specifically, the mutual information between the fused feature X_{out} and the reference feature Y is approximated as:

$$I(F_{out}; Y) \geq \alpha^2 I_{spa} + (1 - \alpha)^2 I_{freq} + 2\alpha(1 - \alpha) \cdot \frac{1}{2} \ln(1 - \rho^2)^{-1} \quad (12)$$

where I_{spa} and I_{freq} denote the mutual information between each branch and the reference, and ρ is the Pearson correlation coefficient between X_s and X_f . The first two terms represent the independent contributions of each branch, while the third term captures their interaction, which increases as the statistical redundancy between the two representations decreases.

In practice, the Pearson correlation coefficient is computed after flattening feature maps along spatial and channel dimensions. The resulting correlation is $\rho = 0.818$, indicating a moderate level of redundancy between the two branches.

Based on this approximation, we define the synergistic increment:

$$\Delta I = \alpha(1 - \alpha) \ln(1 - \rho^2)^{-1} \quad (13)$$

which quantifies the additional information gain beyond linear combination.

As shown in Table 6, the simple concatenation baseline (e) achieves a synergistic increment of 0.11 nats, whereas the proposed DDFB increases it to 0.28 nats, accounting for approximately 23% of the total information. Meanwhile, the PSNR improves from 29.49 dB to 31.08 dB.

The consistent increase in both ΔI and PSNR indicates that DDFB effectively reduces redundancy and enhances complementary information integration between spatial and frequency representations, demonstrating a non-trivial collaborative effect beyond simple feature aggregation.

Table 6 Comparison of Information Theory Indicators

Fusion Method	Synergistic Increment ΔI (nats)	Percentage of Total Mutual Information	PSNR(dB)
(e)	0.11 nats	9%	29.48 dB
(f)	0.28 nats	23%	31.08 dB
Improvement	+0.17 nats	+14%	+1.60 dB

3.3.3 Comparison of Long Distance Dependence Extraction Ability between Spatial and Frequency Domain Feature Extraction Methods

To evaluate the ability of different branches to capture global structures, we construct a controlled experiment using low-pass filtered images as global structural targets.

As shown in Table 8, the frequency branch consistently outperforms the spatial branch across all scales, with an average PSNR gain of approximately 1.3 dB. This demonstrates that frequency-domain representations are more effective in preserving large-scale structural information and capturing long-range dependencies.

Table 7 Comparison of Ablation Study Results

Model	Generator Params	Discriminator Params	Total
Baseline-S	14.66 M	2.76 M	≈ 17.4 M
Baseline-F	14.66 M	2.76 M	≈ 17.4 M

Table 8 Comparison of Long-Distance Dependency Extraction Capabilities between Spatial Domain and Frequency Domain at Different Scales

Scale	Spatial Domain	Frequency Domain
256×256	29.1 dB	30.2 dB
128×128	27.9 dB	29.4 dB
64×64	26.2 dB	27.0 dB

3.3.4 Analysis of Frequency Domain Advantage Mechanism

This subsection reveals the unique value of the frequency domain branch in motion image deblurring. It uses both visualization and quantitative methods. To visually compare high-frequency detail recovery quality, in Fig. 8, we apply high-pass filtering to model-restored images. This generates black-and-white texture maps with local magnification. (a) shows results from the pure spatial domain branch. (b) shows results from the pure frequency domain branch. (c) shows output from our dual-branch fusion. Corresponding texture structure maps are presented below with local magnification for microscopic comparison. Magnified observation shows that (b) has flat edges on license plate digits with fewer ripples. This exhibits typical frequency-domain characteristics of "low energy-high fidelity." (a), though brighter overall, shows slightly chaotic micro-textures. Since the overall PSNR difference between them is only 0.09

dB, visual differences are subtle. Therefore, we use three metrics for quantitative comparison: HF-PSNR, Edge density, and Local-std (see Table 9). This further verifies the objective advantage of the frequency branch in high-frequency detail recovery.

HF-PSNR (High-Frequency PSNR) is the peak signal-to-noise ratio calculated only on DoG high-frequency maps. The mathematical formula is as follows:

$$\left\{ \begin{array}{l} \text{DoG}(I) = G_{\sigma=1} * I - G_{\sigma=3} * I \\ \text{MSE}_{\text{HF}} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \left[\text{DoG}(I^{\text{pred}})_{i,j} - \text{DoG}(I^{\text{gt}})_{i,j} \right]^2 \\ \text{HF-PSNR} = 20 \log_{10} \frac{255}{\sqrt{\text{MSE}_{\text{HF}}}} \end{array} \right. \quad (14)$$

Where I^{pred} represents the test image, I^{gt} represents the clear image; $G_{\sigma} * I$ represents Gaussian convolution (low-pass) with standard deviation of σ ; DoG: Difference of Gaussians, the difference between $\sigma = 1$ and $\sigma = 3$, equivalent to a band-pass filter, retaining only high-frequency textures at scales of 3-15 px; H, W are the image height and width.

Edge density is the ratio of Canny edge points per thousand pixels ($\%$). The mathematical formula is as follows:

$$\left\{ \begin{array}{l} \text{EdgeMap} = \text{Canny}(I; 50, 150) \\ \text{EdgeDensity} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \text{EdgeMap}_{i,j} \times 1000 (\%) \end{array} \right. \quad (15)$$

Where $\text{Canny}(I; 50, 150)$ represents Canny edge detection with low threshold 50 and high threshold 150; $\text{EdgeMap}_{i,j}$ represents a binary map where edge pixels are 1 and non-edge pixels are 0; $H \times W$ represents the total number of pixels (for ratio calculation), $\times 1000$ indicates the result is in per mille ($\%$), i.e., the number of edge points per thousand pixels.

Local-std is the mean of local standard deviations. The image is divided into non-overlapping $B \times B$ blocks. The standard deviation of each block is calculated and then averaged. The mathematical formula is as follows:

$$\left\{ \begin{array}{l} \mathcal{P}_{i,j} = I[iB : (i+1)B, jB : (j+1)B] \\ \mu_{i,j} = \frac{1}{B^2} \sum_{(x,y) \in \mathcal{P}_{i,j}} I(x,y) \\ \sigma_{i,j} = \sqrt{\frac{1}{B^2} \sum_{(x,y) \in \mathcal{P}_{i,j}} (I(x,y) - \mu_{i,j})^2} \\ \text{Local-std} = \frac{1}{N} \sum_{i,j} \sigma_{i,j} \end{array} \right. \quad (16)$$

Where B represents block side length (default 32 px); $\mathcal{P}_{i,j}$ represents the pixel set of block (i, j) ; $\mu_{i,j}$ represents the average grayscale within the block; $\sigma_{i,j}$ represents the intra-block standard deviation (a measure of texture variation); N represents the

total number of blocks, $N = \lfloor H/B \rfloor \cdot \lfloor W/B \rfloor$; a larger value of Local-std indicates richer local texture.

Table 9 Comparison of Generation Results between Baseline-S and Baseline-F

Comparison metrics	Specific indicators
HF-PSNR (Freq vs GT)	35.60 dB
HF-PSNR (Spat vs GT)	35.51 dB
Freq: Edge density	27237.91‰
Spat: Edge density	27939.60‰
Freq: Local-std	25.17
Spat: Local-std	25.57

Table 9 analysis reveals that frequency-domain methods yield smaller high-frequency pixel errors. The frequency branch achieves a 0.09 dB higher HF-PSNR compared to the pure spatial branch. Images generated by the frequency branch exhibit approximately 1.6% lower local standard deviation and about 2% lower edge density than those from the pure spatial branch. These differences are not visible to the naked eye. However, the higher HF-PSNR by 0.09 dB demonstrates that the frequency branch does not simply "accumulate" high-frequency energy. Instead, it redistributes the spectrum to bring high-frequency structures closer to the real image. This achieves detail enhancement at the error level rather than at the energy level.

3.3.5 Spatially Varying Deblurring Comparison

To validate the robustness of the proposed method against spatially varying motion blur, we conduct specialized evaluations on the Köhler dataset [44]. In this dataset, images are convolved region-wise with twelve distinct motion blur kernels, resulting in blur characteristics that vary with spatial location, which differs fundamentally from the globally linear blur present in datasets such as GoPro. However, the Köhler dataset was released relatively early and is limited in image resolution and scene diversity. Consequently, recent mainstream methods predominantly focus on high-resolution real-world motion blur (e.g., GoPro, RealBlur), and few contemporary approaches report results on this benchmark. Moreover, the restoration performance of existing methods on this dataset is generally low with marginal differences among them, rendering the absolute metrics of limited discriminative value. Therefore, we select only representative earlier methods with publicly available test results for comparison, with the aim of verifying the generalization capability of our model to spatially varying blur kernels rather than pursuing absolute metric superiority.

As reported in Table 10, our method achieves 22.14 dB PSNR and 0.696 SSIM on the Köhler dataset, attaining the best performance among the compared approaches. The visualizations in Fig. 9 reveal that, due to the increased complexity of the blur patterns in this dataset, the restoration results of all compared methods remain relatively

limited. Nevertheless, our results exhibit fewer ringing artifacts and color distortions near image edges and color-block boundaries, further corroborating the generalization ability of the proposed model to spatially varying kernels.

Table 10 Comparison of PSNR/SSIM trends of motion deblurring algorithm on Köhler datasets

Model	Köhler	
	PSNR	SSIM
Xu et al. [4]	19.16	0.553
Sun et al. [45]	20.85	0.634
Pan et al. [6]	20.28	0.613
Gong et al. [7]	21.24	0.649
DeblurGAN-v2 [12]	20.85	0.634
Ours	22.14	0.696

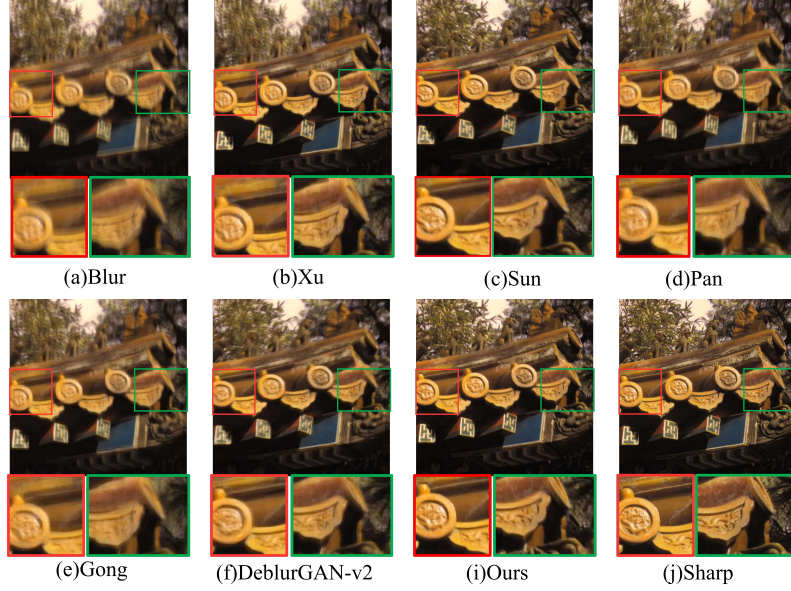


Fig. 9 Comparison of deblurring methods on the Köhler dataset

3.3.6 Limitations of Single Frequency Domain Branch in Complex Scenarios

To investigate the limitations of a single frequency branch when handling complex blur scenes, we conduct a dedicated test on the Köhler dataset [44]. This dataset is generated by convolving images with twelve different motion-blur kernels on a per-region basis, producing spatially varying blur strengths. Unlike GoPro and other globally linear-blur data, the kernel changes with location. Fig. 10 clearly exposes the challenges that frequency-only methods face on images rich in texture and fine detail.

Fig. 10(a) shows the original clear image. Fig. 10(b), (c), and (d) present blurry images with different levels of severity. After processing by the frequency branch model, Fig. 10(e), (f), and (g) show the corresponding reconstruction results. In cases with low blur severity (as shown in Fig. 10(e)), the frequency branch can effectively improve image quality. However, as blur severity increases (as shown in Fig. 10(f) and (g)), reconstruction quality does not improve. Instead, a counterproductive effect of increased blurriness occurs. This suggests that although the frequency branch can theoretically capture global motion features, it tends to over-enhance high-frequency components during reconstruction. This leads to unnatural color block artifacts in local regions.

This phenomenon is particularly evident in edge and texture transition regions. It manifests as color discontinuity and visual abruptness. This color block effect can be attributed to the insensitivity of frequency domain processing to local image details. Since frequency domain operations mainly focus on overall frequency distribution, they may lose or incorrectly reconstruct local details during restoration. This is especially true in high-frequency regions. This indicates that although frequency domain methods can theoretically provide global consistency, they may fail to adequately capture and restore local details and texture continuity in practice. This affects final visual quality.

Furthermore, frequency domain methods may introduce additional artifacts when processing images with complex textures and details. This is shown as the colored block regions in Fig. 10(g). These artifacts affect not only visual quality but also reduce restoration accuracy. Therefore, although frequency domain methods can effectively restore global image structure in some cases, their limitations cannot be ignored in complex scenarios.

In summary, the single frequency domain branch has clear limitations when processing heavily blurred images in complex scenes. This is particularly true in preserving local details and texture continuity. These findings emphasize that both frequency and spatial domain information should be considered comprehensively when designing image deblurring algorithms. This achieves better image restoration results.

3.4 Cross-Task Generalization: Deraining and Gaussian Denoising

To evaluate the generalization capability of the proposed spatial-frequency dual-domain framework, we extend the model to two representative image restoration tasks: single-image deraining and Gaussian denoising. Although these tasks share the

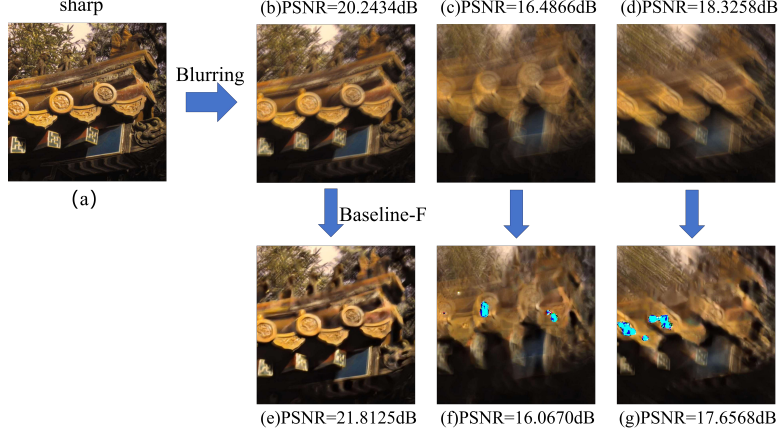


Fig. 10 Comparison of reconstructed images of Baseline-F model under different degrees of ambiguity

common objective of recovering clean images from degraded observations, their degradation characteristics differ significantly from motion blur, providing a challenging testbed for cross-task transferability.

3.4.1 Training Strategy and Implementation Details

For deraining, we adopt the Rain1400 dataset [46], which contains diverse rain streak patterns with varying density and orientation. For Gaussian denoising, we follow a blind denoising setting by adding Gaussian noise with $\sigma \in [0, 50]$ to clean images during training. Evaluation is conducted on the CBSD68 dataset [47] without providing noise level information to the network.

To ensure a fair assessment of generalization, the network architecture and all training settings remain identical to those used in motion deblurring (Section 3.1), without any task-specific modifications. This zero-adaptation protocol explicitly evaluates whether the proposed DDFB module can automatically adjust spatial-frequency contributions under different degradation types.

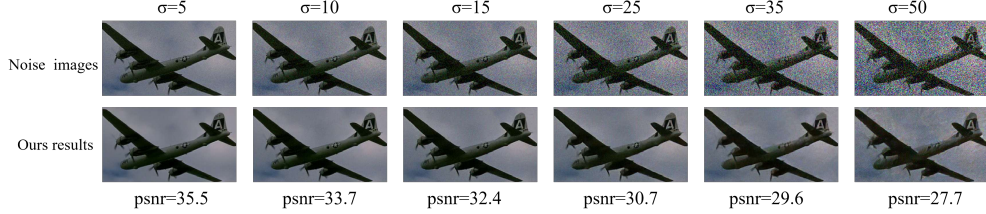
3.4.2 Quantitative and Qualitative Results

To intuitively verify the restoration effectiveness of the dual-domain mechanism in cross-task scenarios, this section conducts quantitative and qualitative evaluations on the Gaussian blind denoising and image deraining tasks, respectively. It is worth noting that the proposed method is not specifically designed or optimized for denoising or deraining tasks. Therefore, instead of comparing with task-specific state-of-the-art methods, this experiment focuses on validating the generalization capability of the proposed dual-domain mechanism under unseen degradation types.

The quantitative results are shown in Table 11 and Table 12. In the Gaussian blind denoising task, the proposed method demonstrates stable noise suppression capability on the CBSD68 test set. As the noise level σ increases from 5 to 50, the model maintains

Table 11 Display of reconstructed images of the model under different Gaussian noise levels σ

Dataset	$\sigma = 5$		$\sigma = 10$		$\sigma = 15$		$\sigma = 25$		$\sigma = 35$		$\sigma = 50$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CBSD68	35.643	0.957	33.832	0.927	32.314	0.8753	30.747	0.815	29.6975	0.797	27.332	0.735

**Fig. 11** Display of reconstructed images of the model under different Gaussian noise levels σ

continuously stable PSNR and SSIM outputs, achieving 27.332 dB PSNR and 0.735 SSIM even under heavy noise ($\sigma = 50$), indicating robust noise suppression capability.

For the deraining task, the model achieves 28.27 dB PSNR and 0.896 SSIM on Rain1400. Despite the absence of task-specific design, the model effectively removes directional rain streaks while preserving background structures. These results suggest that the proposed dual-domain framework is capable of adapting to different degradation characteristics without architectural modification.

Table 12 Comparison of Rain Removal Experiment Results

Dataset	PSNR	SSIM
Rain1400	28.27	0.896

Qualitative results are presented in Figs. 11–12. In the denoising task, the frequency-only branch tends to produce over-smoothed results, while the spatial-only branch struggles to fully suppress noise. In contrast, the proposed fusion model achieves a better balance, producing clean yet detail-preserving outputs.

In the deraining task, the frequency branch effectively suppresses global rain streak patterns but sacrifices fine details, whereas the spatial branch preserves textures but fails to remove rain artifacts completely. The proposed DDFB fusion mechanism successfully integrates both advantages, generating visually natural results without noticeable artifacts. The above results demonstrate that the proposed

spatial-frequency collaborative mechanism generalizes beyond motion deblurring. Without any task-specific adaptation, the model can automatically adjust feature contributions according to different degradation patterns.

This indicates that the effectiveness of the proposed method does not solely rely on motion blur characteristics, but rather stems from a more general capability of jointly modeling global and local information across domains.



Fig. 12 Shows the rain removal reconstruction image of the model on the Rain1400 dataset

3.5 Comparison and discussion of consensus conclusions with existing methods and model deblurring methods

Based on the experimental results in Sections 3.2–3.4, three key observations can be drawn from controlled ablation studies:

- (1) Spatial- and frequency-domain features exhibit strong complementarity in motion deblurring, while single-domain methods fail to jointly achieve accurate global structure modeling and high-fidelity local detail restoration.
- (2) Frequency-domain representations are effective in capturing global motion patterns and long-range dependencies.
- (3) Pure frequency-domain reconstruction tends to introduce ringing artifacts and detail degradation in complex scenes, highlighting the necessity of spatial-domain constraints.

These findings are consistent with the design motivations of existing spatial-frequency methods [3],[31],[32], as well as the effectiveness of frequency-domain modeling demonstrated in FFTformer [22]. However, the empirical foundation of this work differs in that we construct strictly parameter-matched single-branch baselines (Baseline-S and Baseline-F), enabling a fair and quantitative evaluation of both the strengths (Section 3.3.3) and limitations (Section 3.3.5) of frequency-domain modeling. This provides a more objective justification for dynamic cross-domain fusion.

Specifically, Xing et al. [3] employ dual parallel subnetworks for spatial and frequency feature extraction and perform fusion at the output level. Gao et al. [31] introduce a learnable low-pass filtering mechanism for adaptive frequency decomposition and adopt a gating fusion module (GFM) for feature weighting. Guo et al. [32] design a lightweight parallel structure with FFT-Res modules and utilize residual fusion for frequency enhancement. While these methods share the motivation of

exploiting spatial–frequency complementarity, they differ from our approach in fusion granularity, architectural symmetry, and mechanism interpretability. In contrast, FFT-former [22] incorporates frequency-domain operations into Transformer attention via the convolution theorem. Although not a dual-branch framework, it reflects a similar principle of frequency-aware adaptive processing.

Compared with the above methods, the proposed approach differs in three key aspects. First, in terms of fusion strategy, the proposed dynamic dual-domain fusion block (DDFB) adopts element-wise adaptive gating with an anti-dominance constraint, enabling pixel-level dynamic weight allocation. In contrast, [31] performs channel-wise soft gating without explicit constraints, [32] relies on fixed residual fusion, and [3] conducts fusion at the output level. Second, in architectural design, our method maintains symmetric parallel branches throughout the entire U-Net encoder–decoder framework, whereas [31] operates within a single-stream decomposition paradigm, [32] treats the frequency branch as a residual refinement, and [22] introduces frequency modules only in specific stages. Third, from a mechanism perspective, this work explicitly analyzes both the advantages and inherent limitations (e.g., ringing artifacts) of frequency-domain modeling through controlled ablations, while existing methods primarily focus on performance gains without systematic failure-mode investigation.

Furthermore, the proposed dual-domain mechanism is validated on additional restoration tasks, including image deraining and Gaussian denoising, demonstrating its generalization capability without architectural modification.

Overall, building upon the established consensus of spatial–frequency complementarity [3],[22],[31],[32], this work achieves more effective cross-domain collaboration through fine-grained and constrained dynamic fusion, providing a principled and efficient solution for motion image deblurring.

4 Conclusion

This paper systematically investigates the role of frequency-domain modeling in motion image deblurring. Through controlled experiments with strictly parameter-matched architectures, we demonstrate that frequency-domain features are highly effective for capturing global motion patterns and long-range dependencies, while also revealing their inherent tendency to introduce ringing artifacts in complex scenarios.

To address this limitation, a lightweight cross-domain dynamic fusion module is proposed to adaptively balance spatial and frequency representations. The proposed design preserves the global modeling capability of the frequency branch while effectively mitigating artifacts through spatial-domain constraints. Extensive experiments on multiple benchmark datasets validate the effectiveness and robustness of the proposed method.

Beyond performance improvements, this study provides a clearer understanding of the strengths and limitations of frequency-domain mechanisms, offering a practical foundation for future multi-domain image restoration research.

In future work, we will investigate adaptive model simplification strategies to reduce redundancy in dual-branch architectures. In particular, an elastic mechanism

will be explored to enable automatic branch selection or collapse under simple motion conditions, thereby improving efficiency without compromising performance.

Supplementary information. Not applicable.

Acknowledgements. The authors would like to thank the anonymous reviewers for their constructive comments and suggestions. This work was supported by the Science Research Project of Qinghai Minzu University (No. 2024XJMA06) and the Science and Technology Department of Qinghai Province (No. 2023-ZJ-950Q).

Declarations

Funding

This work was supported by the Science Research Project of Qinghai Minzu University (No. 2024XJMA06) and the Science and Technology Department of Qinghai Province (No. 2023-ZJ-950Q).

Conflicts of interest/Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The GoPro, RealBlur and Köhler datasets used in this study are publicly available.

Materials availability

Not applicable.

Code availability

The source code will be made publicly available on GitHub upon acceptance of this manuscript.

Author contribution

JW: Conceptualization, Methodology, Software, Writing – original draft. QQ: Supervision, Writing – review and editing. ZK: Validation, Investigation.

References

- [1] Hu, Z., Zhou, Q., Chen, M., Cui, J., Wu, X., Zheng, B.: Survey of image deblurring. *Journal of Image and Graphics*. 29(4), 841–861 (2024)
- [2] Li, S., Li, J., Cao, F., Zheng, M.: Fusing local and global features for deep multi-view clustering networks. *Frontiers of Computer Science and Technology*. 19(8), 2085–2098 (2025) (in Chinese)
- [3] Xing, Y., Chen, X., Xu, Q., Huang, R.: Fusing spatial and frequency domain information for image deblurring. *Computer Systems and Applications*. 33(2), 1–12 (2024) (in Chinese)
- [4] Xu, L., Zheng, S., Jia, J.: Unnatural l0 sparse representation for natural image deblurring. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1107–1114 (2013)
- [5] Hu, Z., Cho, S., Wang, J., Yang, M.H.: Deblurring low-light images with light streaks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3382–3389 (2014)
- [6] Pan, J., Sun, D., Pfister, H., Yang, M.H.: Blind image deblurring using dark channel prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1628–1636 (2016)
- [7] Gong, D., Yang, J., Liu, L., Zhang, Y., Reid, I., Shen, C., Van Den Hengel, A., Shi, Q.: From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2319–2328 (2017)
- [8] Nah, S., Kim, T.H., Lee, K.M.: 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 (2017)
- [9] 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*. pp. 8174–8182 (2018)
- [10] Wu, T., Wan, S.: Multi-scale feature extraction and fusion network for single image deblurring. *Journal of Nanjing University of Posts and Telecommunications (Natural Science)*. 45(5), 57–65 (2025) (in Chinese)
- [11] 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*. pp. 4641–4650 (2021)
- [12] Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: DeblurGAN: Blind motion deblurring using conditional adversarial networks. In: *Proceedings*

- of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8183–8192 (2018)
- [13] 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. pp. 8878–8887 (2019)
 - [14] Zhao, S., Zhang, Z., Hong, R., Xu, M., Yang, Y., Wang, M.: FCL-GAN: A lightweight and real-time baseline for unsupervised blind image deblurring. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6220–6229 (2022)
 - [15] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Advances in Neural Information Processing Systems*. 30 (2017)
 - [16] Kim, K., Lee, S., Cho, S.: MSSNet: Multi-scale-stage network for single image deblurring. In: European Conference on Computer Vision. pp. 524–539. Springer (2022)
 - [17] Zhang, Y., Jiang, S., Zhu, S.: Single image deblurring via multi-scale feature fusion and context hybrid attention mechanism. *Journal of Chinese Computer Systems*. 46(11), 2633–2642 (2025) (in Chinese)
 - [18] Liu, H., Li, B., Liu, C., Lu, M.: DeblurDiNAT: A compact model with exceptional generalization and visual fidelity on unseen domains. *arXiv preprint arXiv:2403.13163* (2024)
 - [19] Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 25 (2012)
 - [20] Guo, W., Tan, J., Mao, L., Hu, N., Sun, S.: Dual-branch network with adaptive rational nonlinear function for image deblurring. *Signal, Image and Video Processing*. 19(15), 1290 (2025)
 - [21] Wiener, N.: *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications*. The MIT Press (1949)
 - [22] Kong, L., Dong, J., Ge, J., Li, M., Pan, J.: Efficient frequency domain-based transformers for high-quality image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5886–5895 (2023)
 - [23] Mao, X., Wang, J., Xie, X., Li, Q., Wang, Y.: LoFormer: Local frequency transformer for image deblurring. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 10382–10391 (2024)

- [24] Sun, J., Lei, F., Zhang, J., Sun, W., Yang, Y.: Frequency-domain learning with kernel prior for blind image deblurring. arXiv preprint arXiv:2504.14664 (2025)
- [25] Zhao, K., Chen, Y., Li, J.: DNEFNET: Denoising and frequency domain feature enhancement event fusion network for image deblurring. *Computers, Materials & Continua*. 84(1) (2025)
- [26] Lim, J., Yu, C., Kwak, J.T.: HiFi-Deblur: High-frequency intense image deblurring with frequency-decoupled U-Net and discrete wavelet transform. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 984–993 (2026)
- [27] Liu, Q., Wang, J., Li, Z., Zhu, F., Gao, Z., Shen, B.: Enhancing pavement disease classification through wavelet frequency features and attention mechanisms. *The Visual Computer*. 42(7), 296 (2026)
- [28] Chen, J., Xiong, B., Kuang, F., Zhang, Z.: Deep feature fusion attention and dual-scale motion deblurring. *Journal of Image and Graphics*. 28(12), 3731–3743 (2023) (in Chinese)
- [29] Chen, Z., Zhang, H., Miao, D., Cai, K.: Deep fusion of frequency and spatial domain features for multi-granularity dynamic scene image deblurring network. *Pattern Recognition and Artificial Intelligence*. 37(6), 557–569 (2024) (in Chinese)
- [30] Xu, Z., Zeng, H.: Deblurring network based on deep-shallow dual-branch feature fusion. *Computer Engineering and Applications*. 62(3), 254–264 (2026) (in Chinese)
- [31] Gao, H., Dang, D.: Spatial and frequency domain adaptive fusion network for image deblurring. arXiv preprint arXiv:2502.14209 (2025)
- [32] Guo, Y., Ma, L., Zhang, Y.: Image deblurring via frequency-domain feature enhanced convolutional neural networks. *Sensors*. 26(6), 1784 (2026)
- [33] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in Neural Information Processing Systems*. 27 (2014)
- [34] Zhao, S., Zhang, Z., Wei, Y., Fan, J., Zhao, Y., Yan, S., Wang, M.: Individual/joint deblurring and low-light image enhancement in one go via unsupervised deblurring paradigm. *Science China Information Sciences*. 68(12), 1–17 (2025)
- [35] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 234–241. Springer (2015)

- [36] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. *Advances in Neural Information Processing Systems*. 30 (2017)
- [37] Rim, J., Lee, H., Won, J., Cho, S.: Real-world blur dataset for learning and benchmarking deblurring algorithms. In: *European Conference on Computer Vision*. pp. 184–201. Springer (2020)
- [38] Zheng, S., Zhu, Z., Zhang, X., Liu, Z., Cheng, J., Zhao, Y.: Distribution-induced bidirectional generative adversarial network for graph representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7224–7233 (2020)
- [39] Liu, Y., Haridevan, A., Schofield, H., Shan, J.: Application of Ghost-DeblurGAN to fiducial marker detection. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 6827–6832. IEEE (2022)
- [40] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 14821–14831 (2021)
- [41] Park, D., Kang, D.U., Kim, J., Chun, S.Y.: Multi-temporal recurrent neural networks for progressive non-uniform single image deblurring with incremental temporal training. In: *European Conference on Computer Vision*. pp. 327–343. Springer (2020)
- [42] Zhu, C., Dong, H., Pan, J., Liang, B., Huang, Y., Fu, L., Wang, F.: Deep recurrent neural network with multi-scale bi-directional propagation for video deblurring. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 36, no. 3, pp. 3598–3607 (2022)
- [43] 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*. pp. 5728–5739 (2022)
- [44] Köhler, R., Hirsch, M., Mohler, B., Schölkopf, B., Harmeling, S.: Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In: *European Conference on Computer Vision*. pp. 27–40. Springer (2012)
- [45] Sun, J., Cao, W., Xu, Z., Ponce, J.: Learning a convolutional neural network for non-uniform motion blur removal. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 769–777 (2015)

- [46] Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision (ICCV). vol. 2, pp. 416–423. IEEE (2001)
- [47] Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3855–3863 (2017)