DOI:10. 13382/j. jemi. B2508551

# 双域特征融合的 Mamba 去模糊方法*

高 银[1] 陈晨昕[1,2] 李红云[3] 郭霏霏[3] 李 俊[1,3]
(1. 中国科学院福建物质结构研究所 泉州 362200;2. 福建师范大学 福州 350117;3. 泉州职业技术大学 泉州 362200)

**摘 要**:针对图像去模糊过程中单一域分析的局限和扫描特征分布差异化的问题,提出一种双域特征融合的 Mamba 去模糊方法。通过引入状态空间模型,同步提取模糊图像的空间结构特征与小波变换生成的多尺度频域特征,突破单一域分析的局限,实现空间域上下文信息与小波域高频细节特征在状态空间模型引导下的深度聚焦与自适应融合。设计双分支状态空间模块,分别独立建模空域与频域信息,精准适配空域结构特征与频域高频细节的差异化分布特性,在显著提升特征表征能力的同时,彻底规避扫描特征分布差异化,实现高质量的恢复。实验结果表明,所提方法在 GoPro 数据集上峰值信噪比(PSNR)达到 33. 75 dB,结构相似性(SSIM)为 0. 968;在 HIDE 数据集上 PSNR 为 31. 81 dB,SSIM 为 0. 949;在 RealBlur-J 和 RealBlur-R 数据集上分别取得 PSNR 32. 92/0. 937 和 40. 15/0. 974,显著优于对比方法。提出的方法在模糊去除、结构恢复、边缘保留和视觉效果方面的性能均优于经典去模糊方法,通过该方法设计出的装置能够在实际工程领域实现高精度清晰化处理。

**关键词**: 图像处理;计算装置;Mamba 去模糊;双域特征融合;状态空间
**中图分类号**: TP391; TN01 **文献标识码**: A **国家标准学科分类代码**: 520. 6040

# Mamba deblurring method via dual-domain feature fusion

Gao Yin[1] Chen Chenxin[1,2] Li Hongyun[3] Guo Feifei[3] Li Jun[1,3]
(1. Fujian Institute of Material Structure, Chinese Academy of Sciences, Quanzhou 362200, China;
2. Fujian Normal University, Fuzhou 350117, China; 3. Quanzhou Vocational and Technical University, Quanzhou 362200, China)

**Abstract**: In view of the limitations of single-domain analysis and the differentiated distribution of scanning features in image deblurring, a novel Mamba deblurring method based on dual-domain feature fusion is proposed. By introducing a state-space model, the proposed method simultaneously extracts spatial structural features from blurred images and multi-scale frequency-domain features generated by wavelet transformation. This approach overcomes the constraints of single-domain analysis and enables deep integration and adaptive fusion of spatial-domain contextual information with high-frequency details in the wavelet domain, all under the guidance of the state-space model. A dual-branch state-space module is designed to independently model spatial and frequency-domain information, accurately adapting to the differentiated distribution characteristics of spatial structures and high-frequency details in the frequency domain. While significantly enhancing feature representation capabilities, the method effectively addresses the challenges posed by the differentiated distribution of scanning features and achieves high-quality image restoration. Experimental results demonstrate that the proposed method achieves PSNR of 33. 75 dB and SSIM of 0. 968 on the GoPro dataset, PSNR of 31. 81 dB and SSIM of 0. 949 on the HIDE dataset, and PSNR/SSIM of 32. 92/0. 937 and 40. 15/0. 974 on RealBlur-J and RealBlur-R datasets, respectively, outperforming classical deblurring approaches in terms of blur removal, structural restoration, edge preservation, and overall visual quality. Devices developed based on this method are capable of high-precision image enhancement in practical engineering applications.

**Keywords**: image processing; computing device; Mamba deblurring; dual-domain feature fusion; state space

收稿日期: 2025-07-10 Received Date: 2025-07-10
*基金项目:泉州职业技术大学开放课题(LERIS24-03)、国家自然科学基金(62001452)、福建省科技计划项目(2024T3040)资助

## 0 引 言

计算机视觉是通过摄像设备赋予机器类似于人类的视觉能力,近年来,随着人工智能的爆发式增长,该领域的相关研究取得了长足的进步。但是面对着场景的复杂化,图像模糊成为了计算机视觉研究发展的重要瓶颈之一[1]。在众多的影响图像模糊的因素中,由运动引起的模糊尤为显著。在视觉采集设备工作过程中,视野中的目标由于运动速度大于采集帧率而产生伪影,极大的影响了后续的视觉任务,如无人驾驶,安防监控和遥感等[2]。为了解决这一问题,图像去模糊算法被提出,旨在从模糊图像中恢复原始清晰图像[3]。然而,在去模糊的过程中,通常只能获取模糊图像,模糊过程及原始清晰图像的信息未知,严重制约了去模糊算法的发展[4]。随着智能拍摄终端的大规模普及,对去模糊的需求日益增长,运动去模糊逐渐成为了近年来学者研究热点。

受益于卷积神经网络(convolutional neural network, CNN)的发展,图像去模糊领域取得了显著进展[5]。CNN 在特征提取和图像重建方面表现出较强的能力。然而,作为 CNN 核心运算的卷积操作,由于其固有的空间不变性和局部感受野,导致模型难以捕捉图像的动态变化特性,也无法充分利用有助于去模糊的非局部信息,从而限制了其在去模糊领域的应用。相比之下,Transformer 模型[6]中的自注意力机制通过计算特征图上任意两点之间的相关性,能够有效捕获全局上下文信息,理论上更有利于提取高质量的去模糊特征。然而,标准的缩放点积自注意力机制,其空间和时间复杂度随着输入令牌数量呈二次增长,在处理高分辨率图像时计算代价过高[5]。尽管已有研究提出了基于局部窗口[7-8]、转置注意力[9]或频域近似[10]等方法以降低计算复杂度,但这些方法通常在降低计算复杂度的同时,削弱了对非局部依赖关系或空间细节的建模能力[7],进而影响图像复原质量。因此,亟需开发一种高效的去模糊方法,能够在可控的计算开销下有效整合非局部信息,从而实现高质量的图像复原。

近期,状态空间模型(state space models, SSM)[11-12]在自然语言处理(NLP)任务中展现出强大潜力,能够以线性或近线性计算复杂度有效建模长程依赖关系,尤其在处理长文本或复杂语境时表现突出。改进型 SSM,例如 Mamba[13],通过引入选择性扫描机制(S6),在保持线性计算复杂度的同时,能够动态地选择性保留相关信息并忽略冗余信息。这一点启发我们探索利用 Mamba 有效捕捉对图像去模糊有益的非局部信息。

然而,Mamba 的原始设计是针对一维(1D)序列进行处理的。若直接应用于视觉任务,则需将二维(2D)图像数据转换为一维序列[14]。这种向量化处理不可避免地破坏了原始图像的空间结构信息,从而导致模型难以有效建模局部像素间的空间关联。为缓解这一问题,已有研究尝试在视觉任务中引入多方向扫描机制,以适配状态空间模型[15-17]。然而,该方法显著增加了模型的计算复杂度。

此外,这类基于多尺度的算法还存在一个固有缺陷,制约了其在细节恢复方面的性能。这一缺陷源于其渐进式恢复机制:在从低分辨率到高分辨率的层级处理中,高层的解高度依赖于低层的解作为初始条件[18]。这种依赖策略虽然能够降低高层级恢复的难度,但也引入了关键的信息瓶颈。由于低层空间分辨率较低,其向高层传递的特征主要承载了图像的高层语义信息,而在精确空间定位和细节表征方面存在显著不足。这种空间信息的模糊与损失,直接导致多尺度网络在最终复原结果中难以准确重建高频细节(如锐利的边缘、细微的纹理、清晰的文字等),使得复原图像在视觉上往往显得平滑或缺乏足够的清晰度。因此,迫切需要提升高频细节的复原质量。

针对上述问题,本文提出了基于双域特征融合的 Mamba 去模糊网络。提出双域特征融合模块。该模块突破了传统空间域处理的局限,将频域特征(尤其是通过小波变换获得的特征)与空间特征进行深度融合。离散傅里叶变换(discrete fourier transform, DFT)已被许多最新的去模糊算法用作频率先验,用于识别和保留关键的高低频分量。然而,相较于 DFT,离散小波变换(discrete wavelet transform, DWT)在处理包含更多突变信号的图像时更具优势[19]。因此,本文引入小波变换作为空间特征的补充,其不仅具备有效的全局建模能力,还具有优异的退化分离性能。该模块能够同步提取模糊图像的空间结构特征和小波变换生成的多尺度频域特征,并在状态空间模型的引导下,实现空间域上下文信息与小波域高频细节特征的深度融合与自适应增强。通过该策略,网络能够更全面、精确地解析并增强图像中的关键高频成分,显著提升模糊区域细节特征的恢复能力,为高质量图像复原提供了有力支撑。提出双分支状态空间模块,分别独立建模空间域与频域信息,以解决现有基于 Mamba 方法中通过计算密集型多方向扫描扩展感受野的局限性。与单一路径处理不同,该设计能够精准适配空间域结构特征与频域高频细节的分布差异,在显著提升特征表征能力的同时,彻底规避了冗余的扫描操作,实现了近似线性计算复杂度的优化。通过门控机制与频域特征的选择性增强策略,该模块不仅能够高效抑制低频噪声,还能有针对性地增强高频信息的恢复,从而有效提升图像复原质量。

## 1　本文方法

为了解决图像去模糊这一挑战性问题，本文提出了一种结合小波域与空间域特性的 Mamba 网络。该网络架构采用由粗到细的策略，最大程度减少运动模糊对高频细节的损失，如图 1 所示。该模型由编码器和解码器组成，具备较高的可扩展性。以单幅图像为输入，模型通过自下而上的分层处理，逐步生成清晰的输出结果。在编码与解码阶段，本文引入了一种基于空间域与频域小波特性的 Mamba 模块（SFMamba），并将其应用于图像去模糊过程。该模块以频域分支和空间分支为核心组件，显著提升了算法对图像细节的恢复能力。

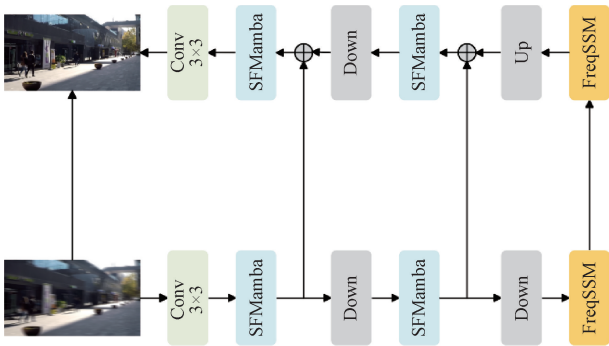



图 1　本文方法网络结构

Fig. 1　The network architecture of our method

### 1.1　双域特征融合模块

频域特征信息对于提升去模糊图像边缘细节的重建效果具有重要作用[10]。为此，本文提出了一种 SFMamba 模块，其网络结构如图 2 所示。SFMamba 主要由空间域特征提取分支和频域特征提取分支组成，同时结合通道拼接与状态空间模型，有效融合空间域与频域特征信息。

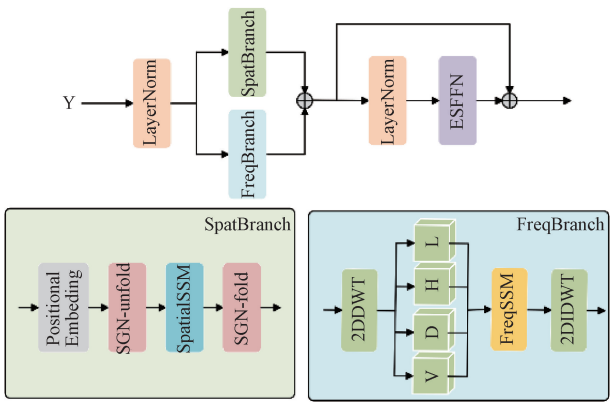



图 2　SFMamba 融合模块

Fig. 2　SFMamba fusion module

在模块的设计过程中，首先对输入张量 $Y$ 进行层归一化（LayerNorm，LN）处理，通过空间域特征提取分支和频域特征提取分支，获得丰富的跨通道上下文信息。该过程可表示为：

$$F_{input} = LN(Y) \tag{1}$$

式中：$LN$ 表示层归一化；$F_{input}$ 同时作为空间域分支和频域分支的输入。

其次，根据空间域特征提取分支对输入进行位置编码，以保留原始结构信息。在此基础上，设计了基于语义引导的邻域（semantic-guided neighborhood，SGN）算法。由图 2 可见，先通过 SGN-unfold 操作将二维图像展开为一维序列，用于后续的注意力状态空间方程建模。此外，采用 SGN-fold 操作作为前一个 SGN 的逆算子，将一维序列重新折叠回二维图像，并进行线性投影，得到模块输出。空间域特征提取分支的输出 $F_{spatial}$ 可表示为：

$$F_{spatial} = unSGN(SSSM(SGN(PosEmbed(F_{input})))) \tag{2}$$

式中：$PosEmbed$ 表示位置编码；$SGN$ 和 $unSGN$ 分别表示 SGN 算法及其逆操作；$SSSM$ 表示空域上的状态空间模型。

最后，针对频域特征提取分支，通过小波变换引入频域特征。与傅里叶变换不同，小波变换具有自适应的时频分辨率，在处理突变信号丰富的数字图像时表现出良好的效果。其与深度神经网络的结合应用已受到越来越多的关注[20-21]。以一维情形为例，定义小波函数 $\psi_{j,k}(t) = 2^{\frac{i}{2}}\psi(2^j t - k)$ 和尺度函数 $\phi_{j,k}(t) = 2^{\frac{i}{2}}\phi(2^j t - k)$，对于离散信号 $t$ 在比例因子为 $j_0$ 处，可以分解为：

$$f(t) = \sum_{j>j_0}\sum_k d_{j,k}\psi_{j,k}(t) + \sum_k c_{j,k}\phi_{j,k}(t) \tag{3}$$

式中：$d_{j,k} = f(t), \psi(t-k)$ 表示细节系数（即高频分量）；$c_{j,k} = f(t), \phi(t-k)$ 表示近似系数，即低频分量。为实现对输入信号的逐步小波分解，引入低通滤波器 $h(n)$ 以提取信号的低频近似分量（即粗尺度信息）；同时，通过高通滤波器 $g(n)$ 提取高频细节分量（即细尺度信息）。一维信号经一次分解后可得到低频（近似）分量与高频（细节）分量；若进行多层分解，则可获得不同尺度下的系数。每一层的分解仅针对前一层的低频部分进行操作，第 $j$ 层的低频近似系数和高频细节系数可以表示为：

$$c_{j,k} = \sum_k h(2k-n)c_{j-1,k} \tag{4}$$

$$d_{j,k} = \sum_k g(2k-n)c_{j-1,k} \tag{5}$$

上述分解过程可视为信号与滤波器进行卷积后再进行下采样。同时，离散小波逆变换则可通过上采样与滤波操作逐层恢复信号。为了确保离散小波变换具有完美重构的性质，即能够无失真地恢复原始信号，需满足以下条件：

$$\tilde{h}(n)=h(-n)$$
$$\tilde{g}(n)=g(-n) \tag{6}$$
$$\sum_n h(n)\tilde{h}(n+2k)=\delta(k)$$
$$\sum_n g(n)\tilde{g}(n+2k)=\delta(k) \tag{7}$$

式中：$\tilde{g}(n)$ 为高通重构滤波器；$\tilde{h}(n)$ 为低通重构滤波器。通过二维离散小波变换，将输入特征图转换到小波域，获得低频子带 $L$、水平方向高频子带 $D$、垂直方向高频子带 $V$ 和对角方向高频子带 $H$ 4 个不同的频域子带特征。在特征提取过程中，采用集成特殊频域扫描机制的状态空间模型进一步进行特征建模。在频域特征输出阶段，通过逆小波变换将小波域特征图恢复为空间域特征图。频域特征提取分支的输出 $F_{freq}$ 可以表示为：

$$F_{freq}=IDWT(FreqSSM(DWT(F_{input}))) \tag{8}$$

式中：$DWT$ 和 $IDWT$ 分别二维离散小波变换及其逆变换；$FreqSSM$ 表示频域上的状态空间模型。

### 1.2　状态空间模型

经典的 Mamba 算法[22] 主要通过离散状态空间方程来建模 token 之间的交互关系：

$$h'(t)=\boldsymbol{A}\cdot h(t)+\boldsymbol{B}\cdot x(t) \tag{9}$$
$$y(t)=\boldsymbol{C}\cdot h(t)+\boldsymbol{D}\cdot x(t) \tag{10}$$

式中：$h(t)$ 是系统的状态；$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 是描述状态转移和输入影响的矩阵；$x(t)$ 是输入信号；$\boldsymbol{C}$ 是状态到输出的矩阵；$\boldsymbol{D}$ 是直接从输入到输出的参数，通常作为一个可学习的跳跃连接存在。标准的 Mamba 模型主要面向一维序列数据设计，因此在处理二维图像数据时存在一定局限性。将图像数据转换为一维序列后，可能导致局部像素信息的丢失，即原本在空间域中相邻的像素，在一维序列中被视为远距离像素。因此，现有基于 Mamba 的方法通常采用多方向扫描以扩展感受野，这不可避免地增加了计算复杂度。针对频域与空间域信息，本文分别提出了不同的状态空间模块，可有效解决两者特征分布的差异，同时也显著降低了计算复杂度。

鉴于标准的二维 Mamba 扫描无法有效对齐离散的频率层级，本文提出了一种新的频域状态空间模型用于特征建模，如图 3 所示。对于拼接后的频域特征图，采用 1×1 普通卷积和 3×3 深度卷积以提取细节模糊特征，经过 SiLU 激活函数后，将特征按照频域层级划分为 4 个固定大小的窗口（从低频到高频分别为 $LL$、$LH$、$HL$ 和 $HH$），并在每个窗口内部独立进行四向扫描建模。由于频域成像的特殊性，Frequency Scan 模块能够在低频成分中侧重于分析图像的整体结构，而在高频部分更注重细节特征的建模，如边缘和纹理，通过逐步融合各频段特征，实现更优的去模糊效果。

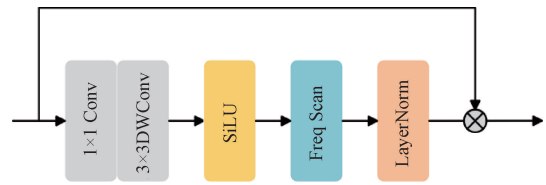



图 3　频域状态空间模型

Fig. 3　Frequency domain state space model

空间域的状态空间建模如图 4 所示。首先构建提示池 $P\in\boldsymbol{R}^{T\times d}$，其中 $T$ 是提示池中提示的数量。该提示池用于查询图像中尚未扫描的像素，从而更好地捕捉图像中的细粒度空间特征与关系。为提升参数效率，对 $P$ 进行低秩分解：

$$P=MN, M\in\boldsymbol{R}^{T\times r}, N\in\boldsymbol{R}^{r\times d} \tag{11}$$

式中：$N$ 在不同的网络块中共享；$M$ 用于特定的网络块；$r$ 为矩阵内秩，满足 $r\ll min\{T,d\}$。式(11)矩阵采用低秩分解方法，使得不同的网络块可以共享相似的特征空间（即 $N$ 为共享部分），同时通过特定的组合系数 $M$ 赋予每个网络块独立的特性。

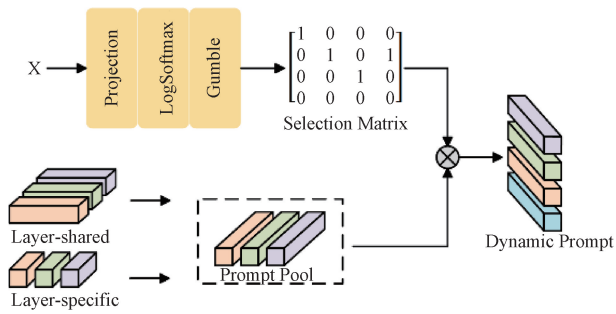



图 4　空域状态空间模型

Fig. 4　Spatial domain state space model

其次，从提示池 $P$ 中选择 $L$ 个实例特定的提示向量 $\boldsymbol{P'}\in\boldsymbol{R}^{T\times d}$。这些提示向量将被加入到 $C$ 中以补充未扫描像素的信息。给定展平后的输入特征 $x'\in T\times d$，通过一个线性层将通道维度从 $C$ 投影到 $T$，利用 LogSoftmax 计算每个提示向量被 $x_i, i=1,2,\cdots,L$ 采样的概率。对这些对数概率应用 Gumbel-Softmax，实现可微分的提示选择，并生成单热矩阵 $\boldsymbol{P'}\in\boldsymbol{R}^{L\times T}$。实例特定的提示向量通过矩阵乘法生成 $\boldsymbol{P'}=\boldsymbol{RP}$，通过残差加法将 $\boldsymbol{P'}$ 融入 $C$ 中，从而得到最终的注意力状态空间方程：

$$h_i=\bar{A}h_{i-1}+\bar{B}x_i \tag{12}$$
$$y_i=(C+P)h_i+Dx_i \tag{13}$$

通过引入提示向量，空间分支具备了类似注意力机制的能力，能够在全图范围内对像素进行查询。此外，借助能够表示相似像素集合的提示向量，所提出的空间分支有效缓解了未扫描像素的感知受限问题。同时，空间分支仅需单方向扫描即可完成操作，从而避免了现有方

法中多方向扫描所带来的高计算成本与冗余。

### 1.3　损失函数

除新增模块与扫描方式外,本文还引入了新的损失函数以优化网络训练过程,从而在空间域与频率域均获得良好性能。所提出的损失函数由像素损失、感知损失和小波损失 3 部分组成。首先,采用像素损失以确保图像的高保真度(低失真),其定义如下:

$$L_{pixel} = \| x - \hat{x} \|_1 \tag{14}$$

式中:$f(y)=\hat{x}$ 和 $x$ 分别是去模糊输出图像和干净(GT)图像;$L_{pixel}$ 是 L1 范数损失。

其次,为了确保高保真度,本文使用 L1 范数损失,而针对感知相似性,引入感知损失项。其中,感知损失采用基于 VGG19 的 LPIPS[23] 指标,用于衡量图像特征之间的距离:

$$L_{percep} = LPIPS(x, \hat{x}) \tag{15}$$

通过该损失项,可以促使网络生成更符合人类视觉感知的图像。最后,进一步提高约束,引入小波损失,以引导模型在多尺度(高频与低频)层次上对预测结果进行优化,其定义如下:

$$L_{wavelet} = \frac{1}{J}\sum_{k=1}^{J}(W(x)_k - W(\hat{x})_k)^2 \tag{16}$$

式中:$W(x)_k$ 和 $W(\hat{x})_k$ 分别是原始图像 $x$ 和重建图像 $\hat{x}$ 的小波系数;$J$ 是小波分解的层数。

因此,损失函数的总体形式可表示为:

$$L_{total} = L_{pixel} + \alpha L_{percep} + \beta L_{wavelet} \tag{17}$$

在本文实验中,$\alpha$ 和 $\beta$ 根据经验分别被设置为 0.01 和 0.05。权重系数 $\alpha$ 和 $\beta$ 的选取是基于一系列初步实验确定的,目的是平衡不同损失项的数量级及其对最终性能的影响。$L_{pixel}$ 主导像素级保真度,而较小的 $\alpha$ 和 $\beta$ 值可在不过度干扰主要优化目标的前提下,引入感知相似性约束和频域多尺度约束。增大 $\alpha$ 会使复原结果更符合人类主观感知但可能略微降低峰值信噪比/结构相似性(PSNR/SSIM)指标;增大 $\beta$ 则会加强对高频细节的优化力度。

## 2　实验与分析

为验证所提出方法的性能优势,本文与多种先进方法进行对比实验,并基于多个客观指标对各方法进行评估。对比方法包括 SRN[1]、DMPHN[24]、MIMO-UNet[3]、MPRNet[4]、Uformer[7]、Restormer[9]、FFTformer[10]、MAXIM[25]、Stripformer[26]、DeepRFT+[27]、CU-mamba[28]、BANet[29]、DeblurGAN-v2[30] 和 Loformer[31]。在客观分析实验中,采用 PSNR 和 SSIM 作为主要评估指标。众所周知,PSNR 用于衡量图像与原始图像之间的相似程度,其值越高表明图像质量越接近原图;SSIM 衡量图像结构信息的一致性,其值越高表明结构信息保留效果越好。在数据集方面,本文根据 Nah 等[32]、Shen 等[33] 和 Rim 等[34] 提出的常用数据集(包括 GoPro、HIDE 和 RealBlur 数据集),对所提方法进行了评估。GoPro 数据集包含 2 103 张训练图像和 1 111 张测试图像。HIDE 数据集包含 2 025 张以人物为主体的测试图像。RealBlur 数据集由不同后处理策略生成的 RealBlur-J 和 RealBlur-R 两个子集组成,其中包含 182 个训练场景和 50 个测试场景。为确保公平性,严格遵循各数据集的评估协议对方法进行测试。

在训练过程中,采用带有默认参数的 ADAM 优化器[35],并结合翻转和旋转操作的数据增强方法生成训练数据。训练初始采用 128 pixel×128 pixel 的补丁大小和 64 的批量大小,进行 300 000 次迭代,学习速率从 $1\times10^{-3}$ 逐渐降低到 $1\times10^{-7}$。之后,将补丁尺寸扩大至 256 pixel×256 pixel,进行 16 个批次,每批 300 000 次迭代,学习速率从初始值 $5\times10^{-4}$,降低至 $1\times10^{-7}$。学习速率的更新基于余弦退火方案完成。所有实验均在 Ubuntu 20.04.5 操作系统、Pytorch 1.13 和 Python 3.8 环境下进行,硬件环境包括 4 张 NVIDIA A100 显卡、Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz,以及多光谱视觉实验平台。

### 2.1　不同数据集的主客观实验

基于 GoPro 数据集的定量与定性评估结果如图 5 所示,其中图 5(a)和(b)分别为模糊图像和清晰图像;(c)~(g)分别为 MIMO-Unet、stripform-er、MPRNet、Restormer 和 CU-Mamba;(h)为本文结果。基于 CNN 的方法在全局信息建模方面存在局限性[3-4],所生成的图像仍存在明显的模糊残差(图 5(c)和(e))。尽管基于 Transformer 的方法[9,26] 具备对全局上下文的建模能力,为降低计算复杂度而引入多种近似策略,但在一定程度上削弱了对全局信息的建模效果,导致部分关键结构(如轮胎和车轮)未能得到有效恢复(图 5(d)和(f))。基于视觉的 Mamba 模型[28] 虽在多项视觉恢复任务中表现优异,但在去模糊任务中对细节的恢复效果仍有不足。例如,车厢上的文字未能完全清晰呈现(图 5(g))。与以上方法相比,本文进一步考虑了小波域与空域信息分布的差异性,在保持低计算开销的同时,显著提升了对非局部信息的建模能力。由图 5(h)可见,所提方法在图像细节恢复方面表现出色,尤其在文本和轮胎区域,去模糊效果明显优于现有方法。

表 1 为各方法在 GoPro 测试集上基于 PSNR 和 SSIM 指标的具体表现。在所有对比方法中,本文提出的 SFmamba 在 PSNR(33.75)和 SSIM(0.968)两项指标上均取得最佳表现,进一步验证了其在运动模糊图像处理任务中的卓越性能。同时,表 1 也包含所提出方法与其

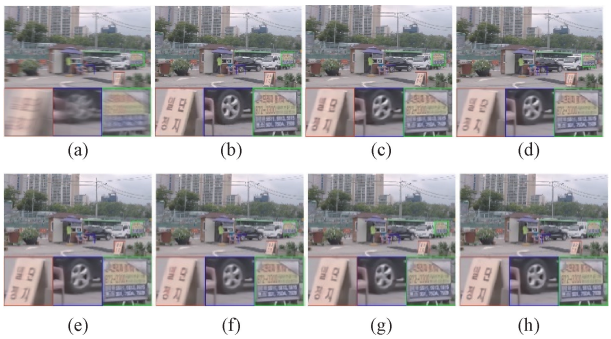



图5　GoPro数据集主观实验对比。

Fig. 5　GoPro dataset subjective experiment comparison

他SOTA方法在GoPro数据集计算量的对比分析。与如MIMO-UNet+的高性能多阶段网络相比,本文方法极其高效,与如Stripformer等基于Transformer的先进方法相比,本文实现了更优的权衡。

**表1　不同算法在GoPro数据集上的性能指标**

**Table 1　Performance metrics of different algorithms on the GoPro dataset**

| 算法 | PSNR/dB | SSIM | 计算量/FLOPs |
|---|---|---|---|
| SRN | 30.26 | 0.934 | - |
| DMPHN | 31.20 | 0.945 | 648 |
| MIMO-UNet+ | 32.45 | 0.956 | 1 235 |
| MPRNet | 32.66 | 0.958 | 760 |
| MAXIM | 32.86 | 0.961 | 169.5 |
| Uformer | 33.06 | 0.967 | 89.5 |
| Stripformer | 33.08 | 0.962 | 170 |
| Restormer | 33.57 | 0.966 | 140 |
| DeepRFT+ | 33.23 | 0.963 | 187 |
| CU-mamba | 33.53 | 0.965 | 95 |
| 本文 | 33.75 | 0.968 | 92.3 |

注:'-'表示原文献未提供该数据,下同

所提方法与现有SOTA模型的主观对比结果如图6所示。现有SOTA模型方法在人脸和瓶子的细节恢复方面都出现不同程度的扭曲。与以上方法相比,所提方法能够有效消除近距离人脸的模糊,恢复更多面部轮廓细节;在饮料瓶等细节区域,也表现出更优的还原能力,在恢复人脸细节方面取得了更优的效果。

此外,本文在HIDE数据集[33]上对所提方法进行了客观评估,结果如表2所示。测试过程中,直接采用在GoPro数据集上训练的模型进行推理。所提方法与多种最新算法进行了对比。本文方法在这HIDE数据集上获得1个最高指标1个次高指标,在PSNR指标上较对比方法高出0.19 dB,进一步验证了其优异的泛化能力。

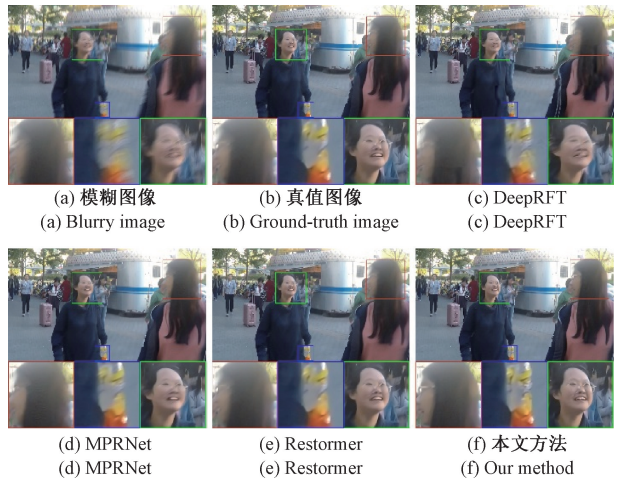



图6　HIDE数据集去模糊结果对比

Fig. 6　HIDE dataset deblurring results comparison

**表2　不同算法在HIDE数据集上的性能指标**

**Table 2　Performance metrics of different algorithms on the HIDE dataset**

| 算法 | PSNR/dB | SSIM |
|---|---|---|
| MIMO-UNet | 29.99 | 0.930 |
| MPRNet | 30.32 | 0.939 |
| Restormer | 31.22 | 0.942 |
| Stripformer | 31.03 | 0.962 |
| BANet | 30.16 | 0.930 |
| DeepRFT+ | 31.42 | 0.944 |
| FFTformer | 31.62 | 0.945 |
| 本文 | 31.81 | 0.949 |

针对真实世界模糊数据集[34],本文通过多种算法对所提方法进行了评估。表3为本文方法与现有方法在RealBlur数据集上的性能对比。其中,在这2个数据集上(RealBlur-J和RealBlur-R),本文方法获得了3个最高指标,1个次高指标,远远优于其他对比方法,整体表现显著优于现有方法。

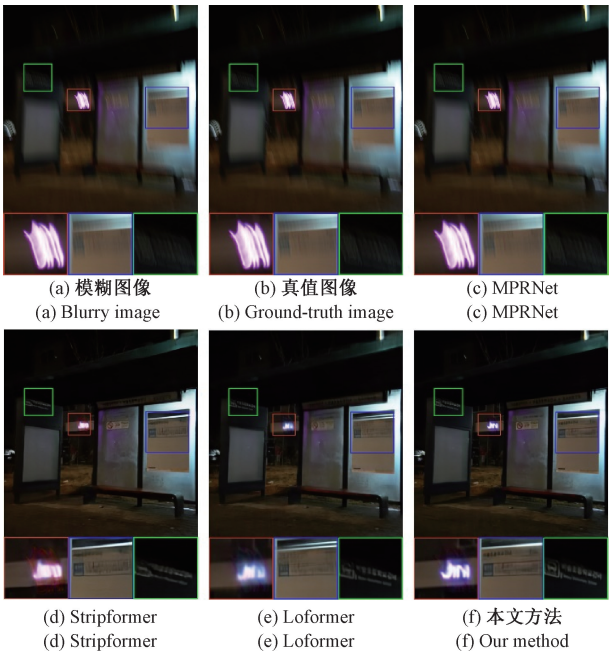
**表3　不同算法在RealBlur数据集上的性能指标**

**Table 3　Performance metrics of different algorithms on the RealBlur dataset**

| 算法 | RealBlur-J | | RealBlur-R | |
|---|---|---|---|---|
| | PSNR/dB | SSIM | PSNR/dB | SSIM |
| DeblurGAN-v2 | 29.69 | 0.870 | 36.44 | 0.935 |
| MIMO-Unet+ | 31.92 | 0.909 | | - |
| MPRNet | 31.76 | 0.922 | 39.31 | 0.972 |
| DeepRFT+ | 32.19 | 0.930 | 39.84 | 0.972 |
| Stripformer | 32.48 | 0.929 | 39.94 | 0.974 |
| Loformer | 32.90 | 0.933 | 40.23 | 0.974 |
| 本文 | 32.92 | 0.937 | 40.15 | 0.974 |

在RealBlur数据集上的不同算法处理的视觉效果如图7所示。在夜晚和极端的模糊情况下,对比方法其残

影比较明显,其中的字母模糊不清,而本文的算法较好的恢复了局部细节,使得图片的信息较为清晰可识别,整体具有较好的细节恢复能力。



(a) 模糊图像 (a) Blurry image　(b) 真值图像 (b) Ground-truth image　(c) MPRNet (c) MPRNet

(d) Stripformer (d) Stripformer　(e) Loformer (e) Loformer　(f) 本文方法 (f) Our method

图 7　RealBlur 数据集去模糊结果对比

Fig. 7　RealBlur dataset deblurring results comparison

### 2.2　消融实验

进一步分析所提方法,并探讨其主要组件的作用。实验中,采用批量大小为 64、图块大小为 128 × 128 pixels,在 GoPro 数据集上训练所提方法及其消融模型。为验证 Mamba 块的重要性,首先以用标准卷积层替换 Mamba 块作为 1 号实验基线;在此基础上,分别单独引入空间 Mamba 分支和频率 Mamba 分支,分别作为 2 号和 3 号实验对照组,并保持其他部分不变。4 号实验为所提出的完整方法。实验结果如表 4 所示。相比基线模型,PSNR 分别提升了 0.76 和 0.68 dB,表明空间 Mamba 分支和频率 Mamba 分支能够显著增强去模糊性能。同时,包含空间 Mamba 分支和频率 Mamba 分支的模型在 PSNR 和 SSIM 指标上均取得最优表现,进一步验证了所提组件联合使用能够显著提升去模糊效果,并具有更强的鲁棒性。

**表 4　消融实验结果**

**Table 4　Ablation study results**

| 算法 | Spatial | Wavelet | PSNR/dB | SSIM |
|---|---|---|---|---|
| 1 | × | × | 32.37 | 0.930 |
| 2 | √ | × | 33.13 | 0.959 |
| 3 | × | √ | 33.05 | 0.953 |
| 4 | √ | √ | 33.75 | 0.968 |

## 3　结　论

本文提出了一种基于双域特征融合的 Mamba 去模糊方法。通过状态空间模型,提取模糊图像的空间结构特征与小波变换生成的多尺度频域特征,解决单一域分析的局限问题,实现空间域上下文信息与小波域高频细节特征在状态空间模型引导下的深度聚焦与自适应融合。构造双分支状态空间模块,独立建模空域与频域信息,在显著提升特征表征能力的同时,规避扫描特征分布差异化,实现高质量的恢复。通过多个数据主客观实验验证,提出的方法在模糊恢复等方面的性能均比优于现有经典去模糊方法。未来将聚焦于该方法的实际应用。

### 参考文献


[1] TAO X, GAO H, SHEN X, et al. Scale-recurrent network for deep image deblurring[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:8174-8182.

[2] GAO H, TAO X, SHEN X, et al. Dynamic scene deblurring with parameter selective sharing and nested skip connections[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:3848-3856.

[3] CHO S J, JI S W, HONG J P, et al. Rethinking coarse-to-fine approach in single image deblurring[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:4641-4650.

[4] ZAMIR S W, ARORA A, KHAN S, et al. Multi-stage progressive image restoration[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:14821-14831.

[5] CHEN L, CHU X, ZHANG X, et al. Simple baselines for image restoration[C]. European Conference on Computer Vision, 2022:17-33.

[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:6000-6010.

[7] WANG Z, CUN X, BAO J, et al. Uformer:A general u-shaped transformer for image restoration[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:17683-17693.

[8] LIANG J, CAO J, SUN G, et al. Swinir: Image restoration using swin transformer[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:1833-1844.

[9] ZAMIR S W, ARORA A, KHAN S, et al. Restormer:

Efficient transformer for high-resolution image restoration[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:5728-5739.

[10] KONG L, DONG J, GE J, et al. Efficient frequency domain-based transformers for high-quality image deblurring [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023:5886-5895.

[11] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces [J]. ArXiv preprint arXiv:2111.00396, 2021.

[12] GU A, JOHNSON I, GOEL K, et al. Combining recurrent, convolutional, and continuous-time models with linear state space layers[J]. Advances in Neural Information Processing Systems, 2021, 34:572-585.

[13] GU A, DAO T. Mamba:Linear-time sequence modeling with selective state spaces[J]. ArXiv preprint arXiv: 2312.00752, 2023.

[14] GAO H, DANG D. Aggregating local and global features via selective state spaces model for efficient image deblurring [J]. ArXiv e-prints arXiv: 2403.20106, 2024.

[15] LIU Y, TIAN Y, ZHAO Y, et al. Vmamba:Visual state space model [J]. Advances in Neural Information Processing Systems, 2024, 37:103031-103063.

[16] GUO H, LI J, DAI T, et al. Mambair:A simple baseline for image restoration with state-space model [C]. European Conference on Computer Vision, 2024: 222-241.

[17] SHI Y, XIA B, JIN X, et al. Vmambair:Visual state space model for image restoration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(6):5560-5574.

[18] KIM K, LEE S, CHO S. Mssnet: Multi-scale-stage network for single image deblurring [C]. European Conference on Computer Vision, 2022:524-539.

[19] DAUBECHIES I. The wavelet transform, time-frequency localization and signal analysis[J]. IEEE Transactions on Information Theory, 2002, 36(5):961-1005.

[20] HA W, SINGH C, LANUSSE F, et al. Adaptive wavelet distillation from neural networks through interpretations[J]. Advances in Neural Information Processing Systems, 2021, 34:20669-20682.

[21] LIU P, ZHANG H, LIAN W, et al. Multi-level wavelet convolutional neural networks[J]. IEEE Access, 2019, 7:74973-74985.

[22] ZHANG H, ZHU Y, WANG D, et al. A survey on visual mamba [J]. Applied Sciences, 2024, 14(13):5683.

[23] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:586-595.

[24] ZHANG H, DAI Y, LI H, et al. Deep stacked hierarchical multi-patch network for image deblurring[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:5978-5986.

[25] TU Z, TALEBI H, ZHANG H, et al. Maxim:Multi-axis mlp for image processing[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:5769-5780.

[26] TSAI F J, PENG Y T, LIN Y Y, et al. Stripformer:Strip transformer for fast image deblurring [C]. European Conference on Computer Vision, 2022:146-162.

[27] MAO X, LIU Y, LIU F, et al. Intriguing findings of frequency selection for image deblurring [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023:1905-1913.

[28] DENG R, GU T. Cu-mamba: Selective state space models with channel learning for image restoration[C]. 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval, 2024:328-334.

[29] TSAI F J, PENG Y T, TSAI C C, et al. BANet:A blur-aware attention network for dynamic scene deblurring[J]. IEEE Transactions on Image Processing, 2022, 31:6789-6799.

[30] KUPYN O, BUDZAN V, MYKHAILYCH M, et al. Deblurgan: Blind motion deblurring using conditional adversarial networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:8183-8192.

[31] MAO X, WANG J, XIE X, et al. Loformer: Local frequency transformer for image deblurring [C]. Proceedings of the 32nd ACM International Conference on Multimedia, 2024:10382-10391.

[32] NAH S, HYUN KIM T, MU LEE K. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:3883-3891.

[33] SHEN Z, WANG W, LU X, et al. Human-aware motion deblurring [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 5572-5581.

[34] RIM J, LEE H, WON J, et al. Real-world blur dataset

for learning and benchmarking deblurring algorithms[C]. Computer vision-ECCV 2020:16th European Conference, 2020:184-201.

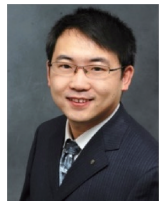
[35] KINGMA D P. Adam: A method for stochastic optimization [J]. ArXiv preprint arXiv: 1412.6980, 2014.

## 作者简介

**高银**,2012 年于云南师范大学获得硕士学位,现为中科院大学博士研究生,主要研究方向视频图像清晰化。

E-mail:yingao@ fjirsm. ac. cn

**Gao Yin** received the M. Sc. degree from Yunnan Normal University in 2012. He is now a Ph. D. degree at the University of Chinese Academy of Sciences. His main research research interet includes video image clarity.



**李俊**(通信作者),2012 年于德国慕尼黑大学获博士学位,现为中国科学院大学教授,主要研究方向为机器人自适应交互控制相关理论与技术研究。

E-mail:iunli@ fiirsm. ac. cn

**Li Jun** (Corresponding author) received his Ph. D. from the University of Munich in 2012. He is now a professor at the Chinese Academy of Sciences University. His main research interests include the theory and technology of adaptive interactive control in robots.