中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2025)10-3187-12

论文引用格式:Qiu Y F, Liu Z Y and Wang M H. 2025. Image deblurring network combining Mamba and snake-like convolution. Journal of Image and Graphics, 30(10):3187-3198(邱云飞,刘则延,王茂华. 2025. 融合Mamba与蛇形卷积的图像去模糊网络. 中国图象图形学报,30(10):3187-3198)[DOI:10.11834/jig.240618]

融合Mamba与蛇形卷积的图像去模糊网络

邱云飞[1],刘则延[1,2*],王茂华[3]

1. 辽宁工程技术大学软件学院,葫芦岛　125105; 2. 辽宁理工学院信息工程学院,锦州　121000;
3. 福建师范大学计算机与网络空间安全学院,福州　350117

摘　要:**目的**　针对Transformer在图像去模糊过程中难以精确恢复图像细节的问题,提出一种结合Mamba模型与蛇形卷积技术的图像去模糊网络MSNet(Mamba snake convolution network)。**方法**　首先,结合Mamba框架与蛇形卷积,提出蛇形状态空间模块(snake state-space module,SSSM)。SSSM通过调整卷积核的形状和路径,动态适应图像局部特征并调整卷积方向,以对齐不同的模糊条纹模式;其次,使用多方向扫描模块(direction scan module,DSM)进行多个方向的扫描,捕捉图像中的长期依赖。再利用离散状态空间方程合并多方向的结构信息,增强模型对全局结构的捕捉能力;最后,引入蛇形通道注意力(snake channel attention,SCA),利用门控设计筛选和调整模糊信息的权重,确保在去除模糊的同时保留关键细节。**结果**　实验在GoPro和HIDE数据集上,与主流的卷积神经网络(convolutional neural network,CNN)和Transformer去模糊方法相比,MSNet的峰值信噪比(peak signal to noise ratio,PSNR)分别提升1.2%和1.9%,结构相似性(structural similarity,SSIM)分别提升0.6%和0.7%。**结论**　本文方法可以有效去除复杂场景下产生的图像模糊,并复原细节。

关键词:图像去模糊;Mamba模型;方向扫描模块(DSM);蛇形卷积;蛇形通道注意力(SCA)

Image deblurring network combining Mamba and snake-like convolution

Qiu Yunfei[1], Liu Zeyan[1,2*], Wang Maohua[3]

1. School of Software, Liaoning Technical University, Huludao　125105, China;
2. College of Information Engineering, Liaoning Institute of Science and Engineering, Jinzhou　121000, China;
3. School of Computer and Cyberspace Security, Fujian Normal University, Fuzhou　350117, China

Abstract: Objective Traditional image deblurring methods, such as those based on convolutional neural networks (CNNs) and Transformers, have achieved substantial advancements in improving deblurring performance. Despite these achievements, these methods are still constrained by high computational demands and limitations in restoring intricate image details. In complex conditions involving motion blur or high-frequency details, existing approaches often rely on fixed convolution kernels or global self-attention mechanisms. Such static designs lack the adaptability to handle diverse types of blur effectively, which leads to suboptimal detail recovery and inadequate reconstruction of global image structures. Moreover, Transformer-based deblurring methods frequently require extensive computational resources, which significantly diminishes their feasibility for deployment on mobile devices or embedded systems. These resource constraints not only restrict their applicability in practical scenarios but also impede their broader adoption in real-world applications.

收稿日期:2024-10-23;**修回日期**:2024-12-23;**预印本日期**:2024-12-30

*通信作者:刘则延　1091661679@qq.com

基金项目:国家自然科学基金项目(62173171);辽宁省自然科学基金项目(2015020095)

Supported by:National Natural Science Foundation of China(62173171);Natural Science Foundation of Liaoning Province, China(2015020095)

To address these challenges, this study proposes a novel image deblurring method, which is termed MSNet. By integrating the efficient state space modeling capabilities of the Mamba framework with snake convolution techniques, MSNet leverages the complementary strengths of these innovations. This approach aims to reduce computational overhead while achieving high-fidelity recovery of fine image details and structural information. With its enhanced adaptability and efficiency, MSNet is better suited for practical applications. It offers robust performance in tackling complex deblurring tasks across diverse scenarios. **Method** To achieve the objective, the MSNet network integrates three key modules: the snake state space module (SSSM), the directional scanning module (DSM), and the snake channel attention module (SCA). Each module is designed for a specific purpose, and their combination effectively tackles local detail recovery and global structure restoration. The SSSM combines the Mamba framework with snake convolution technology, with the aim of enhancing the capability of the model to capture subtle blur features. Unlike traditional CNN-based methods relying on fixed convolution kernels, SSSM dynamically adjusts the shape and path of the convolutional kernels. This way allows them to adapt to local image features and blur stripe patterns. Snake convolution alters the convolution path to effectively capture local blur features. Moreover, the Mamba framework takes advantage of state space models through processing long-range dependencies with linear computational complexity. In contrast to the high computational complexity of Transformer-based models relying on self-attention, Mamba can more efficiently capture long-term dependencies in the image, which avoids the excessive computational burden associated with Transformer models. Simultaneously, snake convolution enhances the precision with which the network adapts to local image features. Thus, it offers notable advantages in capturing complex motion blur and fine detail blur. The DSM module transforms image features into a one-dimensional sequence and scans these features in multiple directions (diagonal, horizontal, and vertical) to capture long-range dependencies. This module effectively improves global structure restoration, particularly in scenes with objects moving simultaneously in multiple directions, which allows for better reconstruction of the overall image structure. The SCA module uses a gating mechanism to filter and adjust the weights of the blurred information. Through combining snake convolution with a channel attention mechanism, this module allows the model to dynamically adjust the weights of different features, which prioritizes key image details while removing irrelevant blur information. Through this selective focus, the SCA module significantly enhances detail recovery and optimizes the overall deblurring performance. **Result** To validate the effectiveness of MSNet, we conducted comparative and ablation experiments on two widely used image deblurring benchmark datasets: GoPro and HIDE. During the experiments, MSNet was compared against several commonly used deblurring methods. The results show that MSNet exhibited outstanding performance in addressing image blur artifacts and restoring fine details. On the GoPro dataset, MSNet achieved significant improvements in PSNR and SSIM compared with Transformer- and CNN-based methods. MSNet demonstrated superior accuracy in restoring blurred regions, which effectively addressed the limitations of existing methods in handling complex scenes. This performance highlights capability of MSNet to process images with intricate details and challenging blur conditions more effectively than its counterparts. On the HIDE dataset, MSNet also outperformed Transformer- and CNN-based methods through achieving higher PSNR and SSIM scores. It showed remarkable accuracy in deblurring fine textual and facial details in blurred images. By leveraging its adaptive convolution design and multi-directional scanning approach, MSNet exhibited strong robustness and generalization capabilities. Thus, it is well suited for complex and dynamic scenarios. Moreover, MSNet demonstrated exceptional computational efficiency. It achieved a computational complexity of 63. 7 GFLOPs on the GoPro dataset, which was significantly lower than those of MIMO-UNet and other comparative methods. This balance of high deblurring performance and low computational cost makes MSNet an ideal solution for real-time deblurring tasks in resource-constrained environments. Ablation studies further validated the contributions of the key modules of MSNet. The removal of the SSSM or the SCA module led to a significant drop in PSNR, with the greatest decrease occurring when both modules were removed. These findings highlight the critical role of these modules in improving deblurring accuracy and restoring fine image details. In addition, network depth analysis revealed that MSNet-28 (28 layers) achieved the best performance, with a PSNR of 33. 51 dB and an SSIM of 0. 97. This result confirms the importance of optimizing network depth and module design to enhance overall performance. **Conclusion** MSNet demonstrates outstanding performance across multiple datasets. It not only showcases its exceptional deblurring accuracy and detail recovery capabilities but also achieves a good balance in computational efficiency. By incorporating the state space model of the Mamba framework and the flexibility of serpentine convolution, MSNet efficiently handles long-range dependencies, particularly exhibiting stronger adaptability in complex blur scenarios. The ablation experiments validate the importance of each module, with the SSSM and SCA modules playing key roles in detail recovery and global structure reconstruction. Overall, MSNet excels in deblurring tasks with its strong generalization capabilities, efficient computation, and superior performance in detail recovery.

Key words: image deblurring; Mamba model; direction scan module (DSM); snake convolution; snake channel attention (SCA)

0 引　言

图像运动模糊通常源于相机抖动或物体在曝光期间的快速运动。这类模糊不仅严重影响图像的视觉质量，显著增加去模糊任务的难度（Hirsch 等，2011），还会干扰高级计算机视觉任务的执行，如目标检测（Kupyn 等，2018）、人脸识别（Lu 等，2019）。传统的运动去模糊方法主要依赖于数学模型（Helstrom，1967；Krishnan 等，2011）、图像先验知识（Whyte 等，2012；Pan 等，2016；Yan 等，2017）和复杂的优化算法（Ren 等，2016；Jia，2007）。尽管这些方法在一定程度上能够解决运动模糊问题，但在面对复杂场景时去模糊效果较差。

近年来，基于深度学习的去模糊方法（程茹秋等，2022；Tsai 等，2022a；Mao 等，2023a；Mao 等，2023b；陈加保 等，2023；胡张颖 等，2024；Li 等，2023）取得了显著进展。通过端到端方式，直接在模糊图像与清晰图像之间构建非线性映射，涌现了许多优秀的研究工作和创新性结构设计。Nah 等人（2017）提出一种多尺度卷积神经网络，实现端到端的图像去模糊。然而，不同尺度之间的关联性未能得到充分考虑。为此，Tao 等人（2018）提出新的尺度递归网络，通过跨尺度权重共享方案关联不同尺度的特征，但也增加了计算复杂度。Cho 等人（2021）重新评估了单幅图像去模糊中粗到细的方法，采用多输入输出架构有效处理多尺度模糊。Kim 等人（2022）通过在不同尺度上采用深度可变的网络结构，增强了模糊特征提取，从而优化了去模糊效果。Zhang 等人（2020）则将图像复原视为清晰图像的生成，并使用生成对抗网络进行去模糊。随着 Transformer（Xia 等，2024；Yun 和 Ro，2024）在图像领域的普及，Zamir 等人（2022）通过改进自注意机制，在有效恢复图像质量的同时降低了计算复杂度。Wang 等人（2022）引入了局部增强窗口变换器模块，该模块在非重叠窗口上执行自注意力操作，从而大幅降低了计算复杂度。然而，现有去模糊方法虽然有所改进，但仍面临计算开销大、细节处理能力不足的问题。特别是在处理条纹模糊和全局结构信息时，现有方法常因计算复杂度过高或特征捕捉不够细致而导致图像复原效果不足。

针对现有去模糊方法在计算开销和细节处理方面的不足，本文提出一种基于 Mamba（Nguyen 等，2022）框架与蛇形卷积的图像去模糊网络 MSNet（Mamba snake convolution network）。Mamba 框架已应用于视觉任务中，如图像分类（Chen 等，2024a）、视频理解（Wang 等，2023）等，且已有研究引入方向扫描策略以提升其性能。在此基础上，本文方法设计了多方向扫描模块（direction scan module，DSM）来捕捉图像中的长期依赖，以改善全局结构的还原效果。同时，针对细节模糊难以去除的问题，模型引入蛇形卷积（Qi 等，2023）并结合 Mamba 的状态空间模型（state-space module，SSM）（Nguyen 等，2022），提出蛇形状态空间模块（snake state-space module，SSSM），以提升局部细节处理能力。此外，模型嵌入了蛇形通道注意力（snake channel attention，SCA）模块，通过门控设计对模糊信息进行筛选和权重调整，使模型能够在去除模糊的同时保留关键细节，有效提高图像复原精度。实验结果表明，MSNet 在多个基准数据集上表现出更强的去模糊精度与细节恢复能力，同时具有较强的泛化能力。

1　本文网络

基于 Mamba 与蛇形卷积图像去模糊网络如图 1 所示，MSNet 的整体框架由编码器与解码器组成，实现端到端的去模糊处理。其核心组件包括蛇形状态空间模块（SSSM）、方向扫描模块（DSM）和

蛇形通道注意力模块(SCA)。其中,蛇形状态空间模块内部集成了方向扫描模块,二者共同用于去模糊任务。

网络架构采用了逐层细化策略,旨在最大程度地保留高频细节,以最大程度减少去模糊过程中信息的丢失。具体而言,本文方法主要包含两部分:首先,如图1(b)所示,蛇形状态空间模块(SSSM)运用蛇形卷积技术处理条纹模糊特征,并借助DSM构建记忆序列,充分激活图像中的像素信息,捕捉长期依赖;其次,如图1(c)所示,蛇形通道注意力模块(SCA)利用蛇形卷积构建门控机制,并融合通道注意力机制生成用于细节修复的特征图,实现权重的精细调整与优化,进一步提升图像质量。为平衡效果与性能,各模块的层数可根据需求动态调整。同时,通过引入跨尺度互联机制,确保信息流在不同尺度间的顺畅传递,进一步增强去模糊效果。

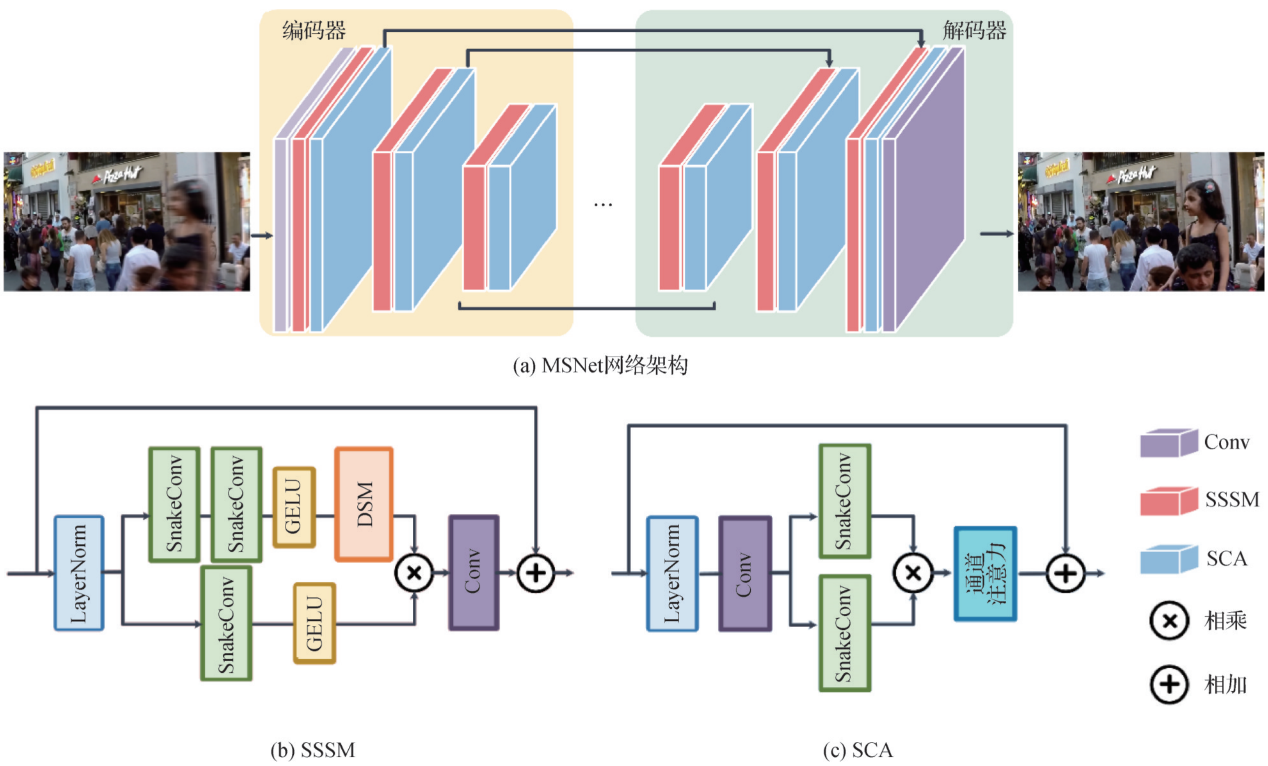


图1 MSNet网络总体架构

Fig. 1 MSNet network overall architecture ((a) MSNet network architecture; (b) SSSM; (c) SCA)

1.1 蛇形状态空间模块

基于Transformer的复原网络通常将输入分割成小块(Chen等,2021a),或采用位移窗口注意力机制(Liang等,2021),这限制了全局图像级别的捕获,也会导致计算复杂度显著增加。相比于Transformer的高计算消耗,Mamba状态空间模型在处理长序列时仅需线性计算复杂度。基于这一特性,研究中引入状态空间模型用于图像复原。在此基础上,本文提出蛇形状态空间模块(SSSM)。区别于传统的SSM模块,所提模块通过一系列改进显著增强了模型的去模糊效果。具体地,在空间扫描模块基础上加入蛇形卷积,增强对细微模糊特征的捕获能力。如图2所示,蛇形卷积通过动态调整卷积核的形状和扫描路径,以更好地捕捉图像中的局部结构。

卷积过程中,首先使用LayerNorm对输入$\boldsymbol{X}$进行层归一化,将输入标准化,并沿通道均等分为$\boldsymbol{Y}_1$和$\boldsymbol{Y}_2$,分别执行蛇形卷积操作。紧接着,在第一个分支中,对$\boldsymbol{Y}_1$进行两层3×3蛇形卷积SConv以提取细节模糊特征,经过GeLU激活函数,最后通过DSM捕获状态空间方程中的长程依赖$\boldsymbol{Y}_{\mathrm{dsm}}$,计算为

$$\boldsymbol{Y}_{\mathrm{dsm}} = DSM(Gelu(SConv(\boldsymbol{Y}_1))) \tag{1}$$

对于第2个分支,对$\boldsymbol{Y}_2$使用3×3的$SConv$获取清晰图像,使用GeLU激活函数激活得到$\boldsymbol{Y}_g$,具体为

$$\boldsymbol{Y}_g = Gelu(SConv(\boldsymbol{Y}_2)) \tag{2}$$

最后,将$\boldsymbol{Y}_{\mathrm{dsm}}$与$\boldsymbol{Y}_g$进行哈达玛乘积,并将结果进行1×1的卷积层降低通道数后与输入$\boldsymbol{X}$相加,生成

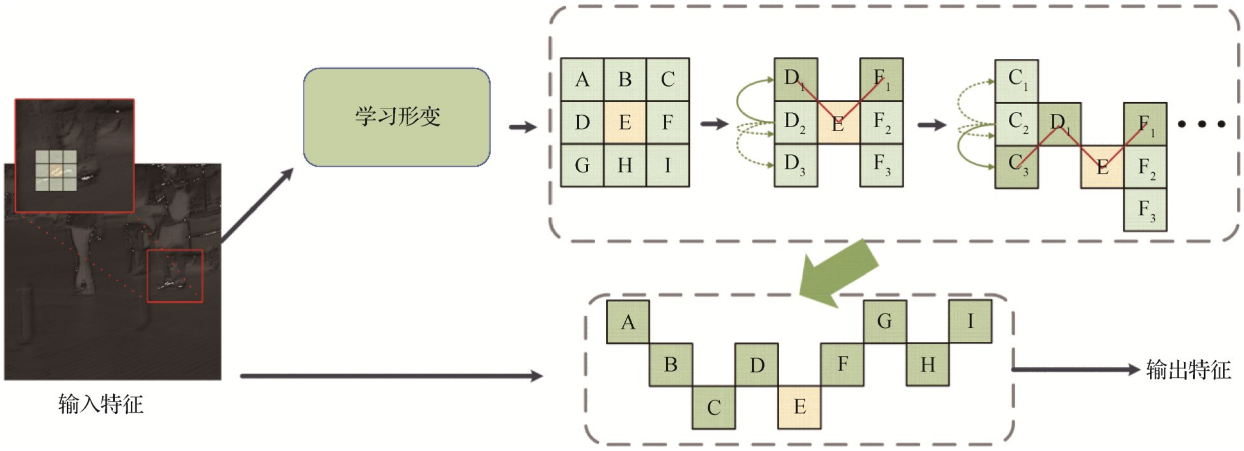


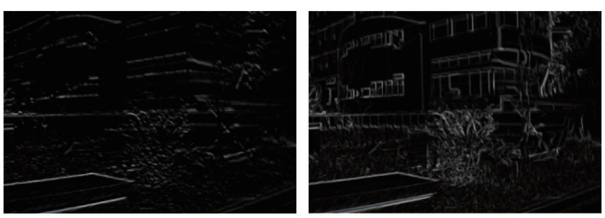
图 2　蛇形卷积

Fig. 2　Snake convolution

恢复图像 Y，计算为

$$Y = X + Conv(Y_{dsm} \odot Y_g) \tag{3}$$

图 3 展示了模糊图像经过 SSSM 处理后生成的内部特征图。具体而言，图 3(a) 与图 3(b) 相比，前者的模糊特征未能有效捕捉细节模糊信息，尤其在纹理复杂和局部细节较为丰富的区域表现不佳，而后者通过蛇形状态空间模型的引入，显著增强了对细微模糊特征的感知能力。使模型能够更好地捕捉和处理图像中的局部模糊现象，从而整体上提升了模型的去模糊效果。



(a) 未使用 SSSM 结果图　　(b) 使用 SSSM 结果图

图 3　模糊特征图

Fig. 3　Blur feature map

((a) result without SSSM; (b) result with SSSM)

1.2　方向扫描模块

结构化状态空间序列模型(Nguyen 等，2022)受到连续线性时变系统的启发，通过将输入序列映射到隐含状态并生成输出，从而高效处理长程依赖性。具体而言，通过隐含的潜在状态 $\boldsymbol{h}(t) \in \mathbf{R}^N$ 来将一维输入序列 $\boldsymbol{x}(t) \in \mathbf{R}^N$ 转换为输出 $t \in \mathbf{R}$。其数学形式为

$$\boldsymbol{h}'(t) = \boldsymbol{A}\boldsymbol{h}(t) + \boldsymbol{B}\boldsymbol{x}(t) \tag{4}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{h}(t) + \boldsymbol{D}\boldsymbol{x}(t) \tag{5}$$

式中，$\boldsymbol{A} \in \mathbf{R}^{N \times N}$，$\boldsymbol{B} \in \mathbf{R}^{N \times 1}$，$\boldsymbol{C} \in \mathbf{R}^{1 \times N}$，$\boldsymbol{D} \in \mathbf{R}$，$N$ 为状态大小，$\boldsymbol{h}'(t)$ 作为下一次的潜在状态，$\boldsymbol{y}(t)$ 为当前状态的输出。接下来，通常会采用离散化过程将式(4)集成到深度学习算法中。具体而言，将 $\boldsymbol{\Delta}$ 表示为时标参数，将连续参数 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 使用零阶保持算法 exp 转换为离散参数用于二维图像，其定义为

$$\begin{cases} \bar{\boldsymbol{A}} = \exp(\boldsymbol{\Delta}\boldsymbol{A}) \\ \bar{\boldsymbol{B}} = (\boldsymbol{\Delta}\boldsymbol{A})^{-1}(\exp(\boldsymbol{A}) - \boldsymbol{I}) \cdot \boldsymbol{\Delta}\boldsymbol{B} \end{cases} \tag{6}$$

最新提出的高级状态空间模型 Mamba 在 S4 的基础上进行了改进，动态调整了参数 $\bar{\boldsymbol{B}}$，$\boldsymbol{C}$ 和 $\boldsymbol{\Delta}$，使其能够根据输入数据进行自适应调整，从而实现更具灵活性的特征表示。在图像复原任务中，Mamba 延续了 S4 模型的递归结构(如式(6))，使其能够记忆较长的序列，从而激活更多像素，增强图像复原能力。同时，并行扫描算法赋予了 Mamba 并行处理能力，从而提高训练效率。为充分利用二维图像中的空间信息，本文提出方向扫描模块(DSM)，如图 4 所示。首先，将二维图像特征扁平化为一维序列，并沿 4 个方向扫描：左上至右下、右下至左上、左下至右上、右上至左下。接着，通过离散状态空间方程捕获每个序列中的长程依赖性。最后，将所有序列相加，通过重塑操作恢复其二维结构。

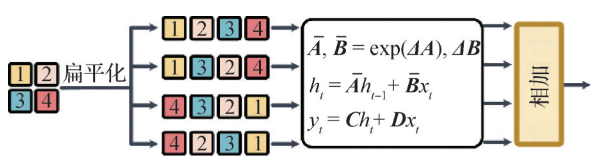


图 4　方向扫描模块

Fig. 4　Direction scan module

1.3 蛇形通道注意力

为筛选关键的模糊信息,通道注意力机制方法(Hu等,2018)通过对每个通道执行池化操作来确定其权重。在此基础上,本文提出一种基于蛇形卷积的蛇形通道注意力机制(SCA),如图5所示。该机制能够更精确地调整特征权重,使模型能够在多维度上更灵活地筛选和调整特征细节。与Chen等人(2021b)提到的SCA模块仅进行单一通道注意操作不同,本文所提蛇形通道注意力模块(SCA)通过蛇形卷积引导注意力机制,使得模型可以更加细致地选择性保留关键模糊特征的细节信息,有效弥补传统通道注意力机制在去模糊任务中对条纹细节的关注不足,从而提升整体去模糊效果。

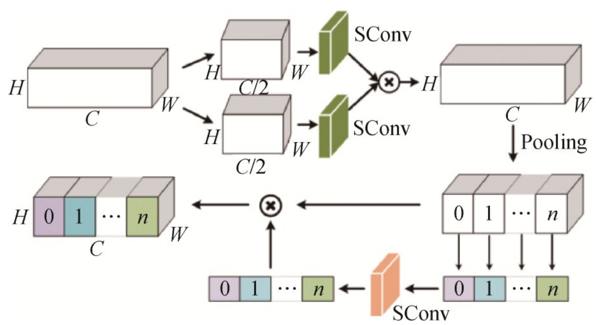


图5 蛇形通道注意力
Fig. 5 Snake channel attention

提取过程中,首先对输入$\boldsymbol{Y}$进行LayerNorm处理,并通过3×3卷积层提取模糊信息$\boldsymbol{X}_1$。接下来,沿着通道维度将$\boldsymbol{X}_1$一分为二,得到通道数为$C/2$的模糊特征,再分别执行3×3的SConv,以捕捉细节模糊特征。受Chen等人(2021b)启发,将两个3×3的SConv进行门控设计,从而实现$\boldsymbol{X}_1$的非线性映射$\boldsymbol{X}_2$,丰富模糊信息。最后利用通道注意力生成的模糊残差,与输入$\boldsymbol{Y}$进行相加,计算为

$$\boldsymbol{Out} = \boldsymbol{Y} + CA(\boldsymbol{X}_2) \tag{7}$$

式中,CA表示通道注意力计算,$\boldsymbol{Out}$表示所恢复的图像。本模块通过丰富细节模糊信息和筛选高层语义信息,模型的去模糊效果得到显著提升。

1.4 损失函数

实验采用了峰值信噪比(peak signal to noise ratio, PSNR)(Wang等,2004)为网络优化过程中的主要损失函数指标。PSNR通过衡量重建图像与原始图像之间的差异来评估恢复效果,对重建质量高度敏感。为了防止训练过程中模型对异常值的过度拟合,使用基于L2范数的损失函数L_{psnr},该范数可以平滑异常值的影响,减少网络过拟合问题,从而提升模型的泛化能力。为了实现这一优化,MSNet模型进行端到端优化,具体为

$$L_{\text{psnr}} = -10\lg\left(\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\boldsymbol{Y}_{\text{gt}}(i,j) - \boldsymbol{Y}_{\text{pred}}(i,j)\right)^2\right) \tag{8}$$

式中,H和W表示图像的高度和宽度,用i和j表示像素坐标。$\boldsymbol{Y}_{\text{gt}}$和$\boldsymbol{Y}_{\text{pred}}$分别代表真值清晰图像和网络恢复图像。通过这种损失函数的设计,能够有效评估模型在图像复原中的效果,确保模型不仅在训练集上表现良好,且在测试集上具有较强的泛化能力。

2 实验及结果分析

实验环境采用Linux系统,配备了4个NVIDIA A100 GPU显卡和一个Intel Xeon Gold 6248R CPU。采用PyTorch框架训练网络,同时使用Cuda 11.6和Cudnn加速网络训练。

2.1 数据集及网络参数设计

为了全面评估提出的MSNet模型的性能,本研究选取了两个公开的图像去模糊基准数据集进行综合实验。

1)GoPro数据集(Nah等,2017)。该数据集由连续帧视频平均生成的模糊数据集组成,所有图像都有清晰的对应图像作为真值图像(ground truth, GT)。共有3 214对1 280×720像素分辨率的视频序列,其中2 103对图像用于训练,1 111对图像用于测试。

2)HIDE数据集(Shen等,2019)。包含2 025对高清图像,均用于测试。参照Yuan等人(2020)的方法,采用旋转、翻转和随机裁剪等方法进行数据扩增。训练时使用经过数据增强的256×256像素分辨率的图像对,测试时使用1 280×720像素作为输入。模型使用Adam优化器进行训练,参数为β_1 = 0.9和β_2 = 0.9。初始学习率设置为10^{-3},并根据余弦退火计划衰减到10^{-6}。

2.2 对比实验

采用峰值信噪比(PSNR)和结构相似性(structural similarity, SSIM)(Wang等,2004)为主要评价指标。通常,PSNR值越高,表示复原图像与原始图像越接近,图像质量越佳;SSIM值越高,则表示结构

信息的保持效果越好。在定量分析中，使用PSNR和SSIM对所提方法进行客观评价。表1展示了各方法在GoPro测试集上的对比结果，包括DeepDeblur（deep multi-scale convolutional neural network for dynamic scene deblurring）（Nah 等，2017）、MIMO-UNet（multi-input multi-output U-Net）（Cho等，2021）、HINet（half instance normalization network for image restoration）（Chen 等，2021b）、MPRNet（multi-stage progressive image restoration network）（Zamir 等，2021）、Hi-Diff（hierarchical integration diffusion model）（Chen等，2024b）、Restormer（Zamir等，2022）和Stripformer（Tsai等，2022a）。实验结果表明，所提方法取得了最佳性能，PSNR为33.51 dB，SSIM为0.966。与基于CNN的最优结果相比，PSNR提升了0.7%，SSIM提升了0.2%；与基于Transformer的最优结果相比，PSNR提升了1.3%，SSIM提升了0.3%。实验结果证明了所提方法在处理运动模糊任务中的有效性。

除了图像质量的评价指标外，模型计算需求也是评估模型性能的重要因素。FLOPs（floating point operations）表示模型在推理或训练过程中所需的浮点运算量，较低的FLOPs值能够减少计算资源需求。表1展示了各方法的FLOPs，本文方法在达到最佳PSNR和SSIM的同时，其计算开销为63.7 GFLOPs，显著低于HINet、Stripformer等其余方法。对比结果表明，所提模型在保持高质量去模糊效果的同时，实现了较低的计算复杂度，具有良好的实用性和部署潜力。

表1 不同方法在GoPro数据集上的性能指标结果对比

Table 1 Comparison of performance of different methods on GoPro dataset

方法	基准模型	PSNR/dB	SSIM	FLOPs/G
DeepDeblur（Nah等，2017）	CNN-Based	30.40	0.807	110.0
MIMO-UNet（Cho等，2021）	CNN-Based	31.73	0.951	67.0
HINet（Chen等，2021b）	CNN-Based	32.71	0.959	170.7
MPRNet（Zamir等，2021）	CNN-Based	32.66	0.959	760.0
Hi-Diff（Chen等，2024b）	CNN-Based	33.28	0.964	125.5
Stripformer（Tsai等，2022a）	Transformer-Based	33.08	0.962	170.0
Restormer（Zamir等，2022）	Transformer-Based	32.92	0.963	140.0
本文	Mamba-Based	**33.51**	**0.966**	**63.7**

注：加粗字体表示各列最优结果。

为验证所提网络的泛化性能，使用在GoPro数据集上训练的模型对HIDE数据集进行测试。表2展示了所提方法与其他方法在HIDE数据集上的对比结果。实验结果显示，所提方法在PSNR（31.82 dB）和SSIM（0.949）指标上均取得了最佳表现。相比基于CNN的方法，PSNR提升了1.3%，SSIM提升了0.5%；相比基于Transformer的方法，PSNR提升了1.9%，SSIM提升了0.7%。这些结果表

表2 不同方法在HIDE数据集上的性能指标结果对比

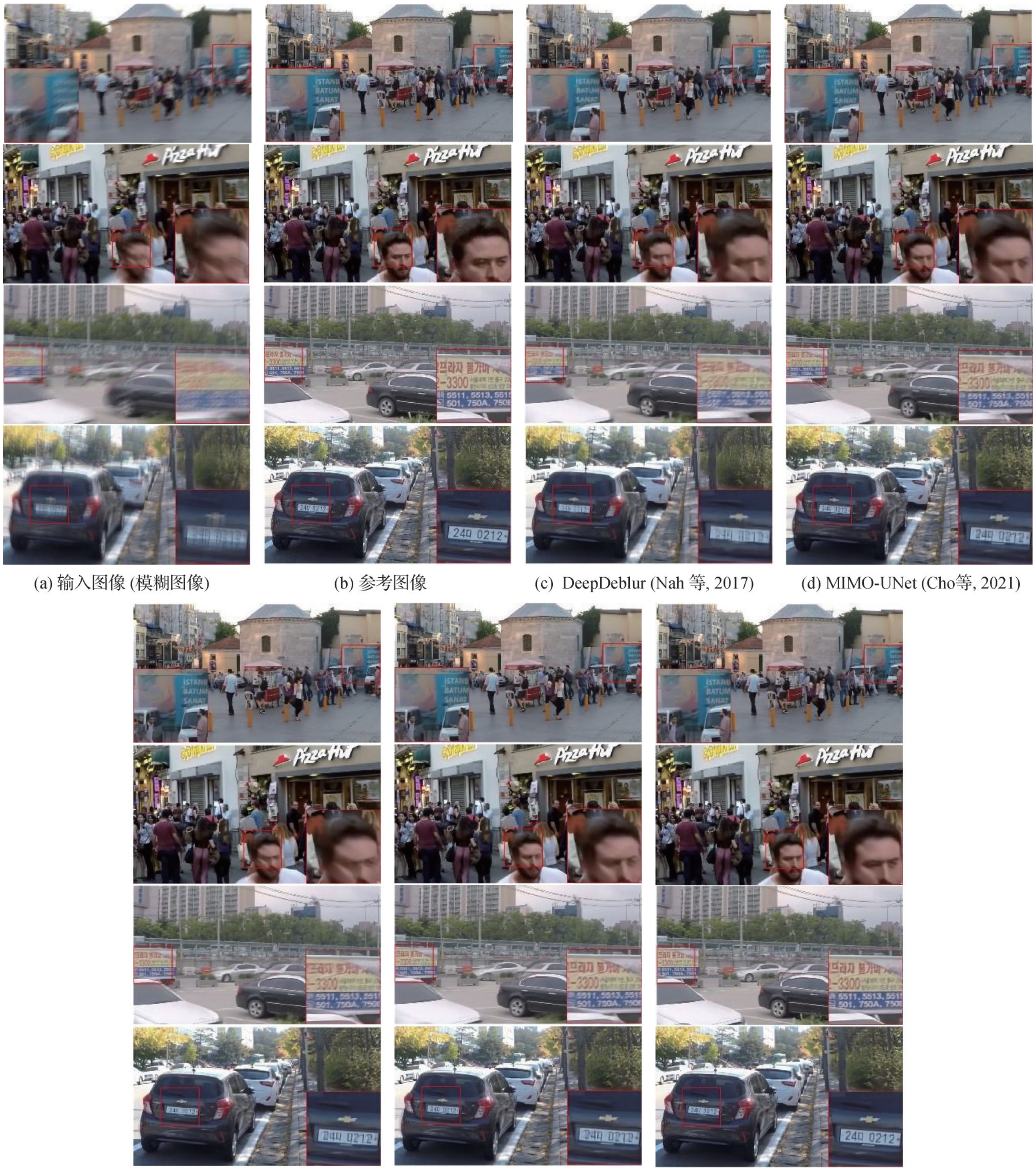
Table 2 Comparison of performance of different methods on HIDE dataset

方法	基准模型	PSNR/dB	SSIM	FLOPs/G
MIMO-UNet（Cho等，2021）	CNN-Based	29.99	0.930	67.0
HINet（Chen等，2021b）	CNN-Based	30.32	0.932	170.7
MPRNet（Zamir等，2021）	CNN-Based	30.96	0.939	760.0
DeepRFT+（Zamir等，2021）	CNN-Based	31.42	0.944	187.0
Hi-Diff（Chen等，2024b）	CNN-Based	31.10	0.941	125.5
Restormer（Zamir等，2022）	Transformer-Based	31.22	0.942	140.0
Stripformer（Tsai等，2022a）	Transformer-Based	31.03	0.940	170.0
本文	Mamba-Based	**31.82**	**0.949**	**63.7**

注：加粗字体表示各列最优结果。

明,所提方法在不同数据集上具有良好的泛化性能。此外,在计算开销方面,所提方法仅为63.7 GFLOPs,显著低于其他对比方法。在计算开销大幅降低的同时,所提模型仍在PSNR和SSIM上取得了领先性能,使得本文方法能够在保持高效图像去模糊效果的同时,显著降低了计算开销。

接下来,在GoPro数据集上,对所提方法与其他4种经典去模糊方法进行了对比。图6对比了不同



(a) 输入图像 (模糊图像)　(b) 参考图像　(c) DeepDeblur (Nah 等, 2017)　(d) MIMO-UNet (Cho等, 2021)

(e) MPRNet (Zamir 等, 2021)　(f) Restormer (Zamir 等, 2022)　(g) 本文

图6　GoPro数据集中去模糊可视化结果对比

Fig. 6　Comparison of deblurring visualization results on the GoPro dataset ((a) blurred images; (b) reference images; (c) DeepDeblur (Nah et al., 2017); (d) MIMO-UNet (Cho et al., 2021); (e) MPRNet (Zamir et al., 2021); (f) Restormer (Zamir et al., 2022); (g) ours)

去模糊方法在处理垂直和水平抖动引起的模糊效果时的表现。第1幅图像（最下面1行）中的车牌信息由于相机垂直抖动而模糊不清。相比其他方法，所提方法更为接近地恢复了车牌信息，准确还原了车牌数字，同时有效地去除了伪影。第2幅图像中的模糊源于目标的快速水平移动，导致文字和数字出现明显的模糊现象。其他方法在恢复过程中出现了虚化，无法清晰重现细节，而所提方法在细节还原和结构保留方面表现更为出色。第3幅图像中，在复杂背景下的人脸模糊得到更为完整的恢复，头发上的残影也成功去除。最后1幅图像（最上面1行）中，远处文字的模糊问题明显改善，其他方法在去模糊过程中未能保持背景的清晰度，而所提方法则有效恢复出更加干净且清晰的图像。

图7展示了不同方法在HIDE数据集上的去模糊效果。在图7(a)中，所提方法有效消除了字体模糊，不仅准确还原了文字细节，还有效减少了抖动和失真带来的影响。图中红色背景在复原过程中未受黑色部分干扰，保持了色彩和边缘的准确性。图7(b)中，所提方法有效地去除了近距离的人脸模糊，显著还原了人脸的细节。与其他方法相比，本文方法在处理人脸特征时，特别是在细节还原和纹理复原方面，表现出明显的优势。

(a) 文字为目标中心的模糊图像

(b) 人物为目标中心的模糊图像

图7　HIDE数据集中去模糊可视化结果对比

Fig. 7　Comparison of deblurring visualization results on the HIDE dataset

（(a) a blurred image focused on text; (b) a blurred image focused on a person）

2.3　消融实验

为了验证不同模块对图像去模糊效果的影响，在GoPro数据集上进行消融实验。实验中对SSSM和SCA进行了消融分析，并探讨了不同层数对网络性能的影响，如表3所示。实验结果表明，在MSNet-16中加入SSSM可使PSNR提高1.09 dB。引入SCA可使PSNR提高2.79 dB，同时引入SSSM和SCA可使PSNR提升3.54 dB，实验结果验证了所提出模块的有效性。

实验对SSSM和SCA的层数进行了分析，结果显示在28层时效果最佳。随后，针对MSNet-28进行消融分析。在去除SSSM模块后，PSNR值降至33.10 dB，SSIM为0.95。去除SCA模块时，PSNR降至33.05 dB，SSIM为0.95。当两个模块均移除时，则模型的性能显著下降。消融实验结果进一步验证SSSM和SCA在MSNet中的必要性与有效性。

表3　消融实验结果

Table 3　Results of ablation experiment

网络	SSSM/SCA数量	SSSM	SCA	PSNR/dB	SSIM
baseline	16	×	×	29.11	0.82
MSNet-16	16	×	√	31.90	0.10
MSNet-16	16	√	×	30.20	0.50
MSNet-16	16	√	√	32.65	0.96
MSNet-24	24	√	√	33.39	0.96
MSNet-28	28	×	×	32.75	0.94
MSNet-28	28	×	√	33.10	0.95
MSNet-28	28	√	×	33.05	0.95
MSNet-28	28	√	√	**33.51**	**0.97**

注：加粗字体表示各列最优结果。

3 结 论

提出一种基于Mamba和蛇形卷积的图像复原网络MSNet,旨在解决传统Transformer方法在高计算消耗和全局图像交互建模能力方面的局限性,从而实现对模糊图像的高精度细节恢复。模型通过引入蛇形状态空间模块(SSSM),有效捕捉条纹模糊特征。引入多方向扫描模块(DSM),捕获多个方向的长期依赖并整合多方向全局信息,从而增强对图像整体结构的复原能力。此外,嵌入蛇形通道注意力模块(SCA),对模糊信息进行筛选和优化,进一步提升图像复原的精度与质量。实验结果表明,MSNet在多个基准数据集上的去模糊效果显著,复原图像边缘清晰、细节丰富,并在降低计算开销的同时整体性能优于现有方法。然而,本文方法在极端复杂场景下的表现仍存在优化空间。未来工作将集中于提升模型对复杂模糊细节的处理能力,进一步优化其计算效率,并探索该方法在超分辨率、图像增强等其他计算机视觉任务中的应用潜力。

参考文献(References)

Chen H T, Wang Y H, Guo T Y, Xu C, Deng Y P, Liu Z H, Ma S W, Xu C J, Xu C and Gao W. 2021a. Pre-trained image processing transformer//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 12294-12305 [DOI: 10.1109/CVPR46437.2021.01212]

Chen J B, Xiong B S, Kuang F and Zhang Z Z. 2023. Motion deblurring based on deep feature fusion attention and double-scale. Journal of Image and Graphics, 28(12): 3731-3743 (陈加保, 熊邦书, 况发, 章照中. 2023. 深度特征融合注意力与双尺度的运动去模糊. 中国图象图形学报, 28(12): 3731-3743) [DOI: 10.11834/jig.220931]

Chen K Y, Chen B W, Liu C Y, Li W Y, Zou Z X and Shi Z W. 2024a. RSMamba: remote sensing image classification with state space model. IEEE Geoscience and Remote Sensing Letters, 21: 1-5 [DOI: 10.1109/LGRS.2024.3407111]

Chen L Y, Lu X, Zhang J, Chu X J and Chen C P. 2021b. HINet: half instance normalization network for image restoration//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 182-192 [DOI: 10.1109/CVPRW53098.2021.00027]

Chen Z, Zhang Y, Liu D, Xia B, Gu J J, Kong L H and Yuan X. 2024b. Hierarchical integration diffusion model for realistic image deblurring//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #1267

Cheng R Q, Yu Y, Shi D Z and Cai W. 2022. The critical review of image and video quality assessment methods. Journal of Image and Graphics, 27(5): 1410-1429 (程茹秋, 余烨, 石岱宗, 蔡文. 2022. 图像与视频质量评价综述. 中国图象图形学报, 27(5): 1410-1429) [DOI: 10.11834/jig.210314]

Cho S J, Ji S W, Hong J P, Jung S W and Ko S J. 2021. Rethinking coarse-to-fine approach in single image deblurring//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4621-4630 [DOI: 10.1109/ICCV48922.2021.00460]

Helstrom C W. 1967. Image restoration by the method of least squares. Journal of the Optical Society of America, 57(3): 297-303 [DOI: 10.1364/JOSA.57.000297]

Hirsch M, Schuler C J, Harmeling S and Schölkopf B. 2011. Fast removal of non-uniform camera shake//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 463-470 [DOI: 10.1109/ICCV.2011.6126276]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Hu Z Y, Zhou Q, Chen M J, Cui J C, Wu X F and Zheng B Y. 2024. Survey of image deblurring. Journal of Image and Graphics, 29(4): 841-861 (胡张颖, 周全, 陈明举, 崔景程, 吴晓富, 郑宝玉. 2024. 图像去模糊研究综述. 中国图象图形学报, 29(4): 841-861) [DOI: 10.11834/jig.230555]

Jia J Y. 2007. Single image motion deblurring using transparency//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE: 1-8 [DOI: 10.1109/CVPR.2007.383029]

Kim K, Lee S and Cho S. 2022. MSSNet: multi-scale-stage network for single image deblurring//Proceedings of 2022 European Conference on Computer Vision. Tel Aviv, Israel: Springer: 524-539 [DOI: 10.1007/978-3-031-25063-7_32]

Krishnan D, Tay T and Fergus R. 2011. Blind deconvolution using a normalized sparsity measure//Proceedings of 2011 Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, USA: IEEE: 233-240 [DOI: 10.1109/CVPR.2011.5995521]

Kupyn O, Budzan V, Mykhailych M, Mishkin D and Matas J. 2018. DeblurGAN: blind motion deblurring using conditional adversarial networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8183-8192 [DOI: 10.1109/CVPR.2018.00854]

Li Z Y, Gao Z, Yi H, Han Yi, Fu Y and Chen B A. 2023. Image deblurring with image blurring. IEEE Transactions on Image Processing, 32: 5595-5609 [DOI: 10.1109/TIP.2023.3321515]

Liang J Y，Cao J Z，Sun G L，Zhang K，Gool L V and Timofte R. 2021. SwinIR：image restoration using swin transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：1833-1844［DOI：10.1109/ICCVW54120.2021.00210］

Lu B Y，Chen J C，Castillo C D and Chellappa R. 2019. An experimental evaluation of covariates effects on unconstrained face verification. IEEE Transactions on Biometrics，Behavior，and Identity Science，1(1)：42-55［DOI：10.1109/TBIOM.2018.2890577］

Mao X T，Liu Y M，Liu F Z，Li Q L，Shen W and Wang Y. 2023a. Intriguing findings of frequency selection for image deblurring//Proceedings of the 37th AAAI Conference on Artificial Intelligence and 35th Conference on Innovative Applications of Artificial Intelligence and 13th Educational Advances in Artificial Intelligence. Washington，USA：AAAI Press：#212［DOI：10.1609/aaai.v37i2.25281］

Mao Y X，Wan Z X，Dai Y C and Yu X. 2023b. Deep idempotent network for efficient single image blind deblurring. IEEE Transactions on Circuits and Systems for Video Technology，33(1)：172-185［DOI：10.1109/TCSVT.2022.3202361］

Nah S，Kim T H and Lee K M. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，USA：IEEE：257-265［DOI：10.1109/CVPR.2017.35］

Nguyen E，Goel K，Gu A，Downs G W，Shah P，Dao T，Baccus S A and Ré C. 2022. S4ND：modeling images and videos as multidimensional signals using state spaces//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans，USA：Curran Associates Inc.：#206

Pan J S，Sun D Q，Pfister H and Yang M H. 2016. Blind image deblurring using dark channel prior//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas，USA：IEEE：1628-1636［DOI：10.1109/CVPR.2016.180］

Qi Y L，He Y T，Qi X M，Zhang Y and Yang G Y. 2023. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris，France：IEEE：6047-6056［DOI：10.1109/ICCV51070.2023.00558］

Ren W Q，Cao X C，Pan J S，Guo X J，Zuo W M and Yang M H. 2016. Image deblurring via enhanced low-rank prior. IEEE Transactions on Image Processing，25(7)：3426-3437［DOI：10.1109/TIP.2016.2571062］

Shen Z Y，Wang W G，Lu X K，Shen J B，Ling H B，Xu T F and Shao L. 2019. Human-aware motion deblurring//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul，Korea (South)：IEEE：5571-5580［DOI：10.1109/ICCV.2019.00567］

Tao X，Gao H Y，Shen X Y，Shen X Y，Wang J and Jia J Y. 2018. Scale-recurrent network for deep image deblurring//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City，USA：IEEE：8174-8182［DOI：10.1109/CVPR.2018.00853］

Tsai F J，Peng Y T，Lin Y Y，Tsai C C and Lin C W. 2022a. Stripformer：strip transformer for fast image deblurring//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv，Israel：Springer：146-162［DOI：10.1007/978-3-031-19800-7_9］

Tsai F J，Peng Y T，Tsai C C，Lin Y Y and Lin C W. 2022b. BANet：a blur-aware attention network for dynamic scene deblurring. IEEE Transactions on Image Processing，31：6789-6799［DOI：10.1109/TIP.2022.3216216］

Wang J，Zhu W T，Wang P C，Yu X，Liu L D，Omar M and Hamid R. 2023. Selective structured state-spaces for long-form video understanding//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver，Canada：IEEE：6387-6397［DOI：10.1109/CVPR52729.2023.00618］

Wang Z，Bovik A C，Sheikh H R and Simoncelli E P. 2004. Image quality assessment：from error visibility to structural similarity. IEEE Transactions on Image Processing，13(4)：600-612［DOI：10.1109/TIP.2003.819861］

Wang Z D，Cun X D，Bao J M，Zhou W G，Liu J Z and Li H Q. 2022. Uformer：a general U-shaped transformer for image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，USA：IEEE：17662-17672［DOI：10.1109/CVPR52688.2022.01716］

Whyte O，Sivic J，Zisserman A and Ponce J. 2012. Non-uniform deblurring for shaken images. International Journal of Computer Vision，98(2)：168-186［DOI：10.1007/s11263-011-0502-7］

Xia C L，Wang X L，Lyu F，Hao X and Shi Y F. 2024. ViT-CoMer：vision transformer with convolutional multi-scale feature interaction for dense predictions//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：5493-5502［DOI：10.1109/CVPR52733.2024.00525］

Yan Y Y，Ren W Q，Guo Y F，Wang R and Cao X C. 2017. Image deblurring via extreme channels prior//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu，USA：IEEE：4003-4011［DOI：10.1109/CVPR.2017.738］

Yuan Y，Su W and Ma D D. 2020. Efficient dynamic scene deblurring using spatially variant deconvolution network with optical flow guided training//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：3552-3561［DOI：10.1109/CVPR42600.2020.00361］

Yun S and Ro Y. 2024. SHViT：single-head vision transformer with memory efficient macro design//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle，USA：IEEE：5756-5767［DOI：10.1109/CVPR52733.2024.00550］

Zamir S W，Arora A，Khan S，Hayat M，Khan F S and Yang M H. 2022. Restormer：efficient transformer for high-resolution image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]

Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H and Shao L. 2021. Multi-stage progressive image restoration//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 14816-14826 [DOI: 10.1109/CVPR46437.2021.01458]

Zhang K H, Luo W H, Zhong Y R, Ma L, Stenger B, Liu W and Li H D. 2020. Deblurring by realistic blurring//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2737-2746 [DOI: 10.1109/CVPR42600.2020.00281]

作者简介

邱云飞,男,教授,主要研究方向为机器视觉和图像处理。E-mail:7415575@qq.com

刘则延,通信作者,男,硕士研究生,主要研究方向为机器视觉和图像复原。E-mail:1091661679@qq.com

王茂华,男,硕士研究生,主要研究方向为图像增强。E-mail:wmh10238578@163.com