

中图法分类号: TP18; TP391.4 文献标识码: A 文章编号: 1006-8961(2026)04-1004-14

论文引用格式: Zhang J, Zhu X Y, Cao J G and Zhai Q J. 2026. Transformer and multi-scale feature fusion for motion image deblurring. Journal of Image and Graphics, 31(4):1004-1017(张冀, 朱晓昀, 曹锦纲, 翟清剑. 2026. Transformer和多尺度特征融合的运动图像去模糊. 中国图象图形学报, 31(4):1004-1017)[DOI:10.11834/jig.250351]

Transformer和多尺度特征融合的运动图像去模糊

张冀^{1,2}, 朱晓昀¹, 曹锦纲^{1,2}, 翟清剑^{1*}

1. 华北电力大学计算机系, 保定 071003; 2. 华北电力大学复杂能源系统智能计算教育部工程研究中心, 保定 071003

摘要: **目的** 运动图像去模糊是图像复原领域中一个具有挑战性的问题,在涉及复杂、非均匀模糊以及高频细节显著丢失的情况下,现有方法通常难以有效处理这些复杂性,导致图像复原结果不理想,表现出明显的模糊和细节缺失。提出了一种基于Transformer和多尺度特征融合的运动图像去模糊方法。**方法** 采用编码器-解码器结构,在编码器部分,设计了结合卷积神经网络(convolutional neural network, CNN)和双重注意力机制的模块,以更好地提取图像特征;在特征融合部分,设计了多尺度特征融合模块,通过门控深度卷积前馈网络和特征增强模块,增强特征表达能力;在解码器部分,引入改进的Transformer模块,在前馈网络层将傅里叶变换融入特征提取过程,增强特征。**结果** 在GoPro数据集和HIDE(human-centric indoor deblurring)数据集上与现有主流方法进行了对比实验,所提方法均取得了最优的去模糊效果。在GoPro数据集上,PSNR(peak signal-to-noise ratio)达到32.70 dB,SSIM(structural similarity index)达到0.954;在HIDE数据集上,PSNR达到30.53 dB,SSIM达到0.922。消融实验进一步验证了本文所提出的各个创新点在运动图像去模糊过程中的积极影响。**结论** 本文提出的基于Transformer和多尺度特征融合的运动去模糊方法,在GoPro数据集和HIDE数据集的实验中,均显著优于现有主流方法,为运动图像去模糊任务提供了一种新的解决方案,展示了其在实际应用中的潜力。代码和模型获取网址:<https://doi.org/10.57760/sciencedb.j00240.00170>和<https://github.com/zh7546/project.git>。

关键词: Transformer;多尺度特征融合;运动去模糊;注意力机制;傅里叶变换

Transformer and multi-scale feature fusion for motion image deblurring

Zhang Ji^{1,2}, Zhu Xiaoyun¹, Cao Jingang^{1,2}, Zhai Qingjian^{1*}

1. Department of Computer Science, North China Electric Power University, Baoding 071003, China;

2. Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education,

North China Electric Power University, Baoding 071003, China.

Abstract: Objective Motion deblurring is an important yet complex task in image restoration. In many practical scenarios, images can exhibit complex motion blur because of camera shake, rapidly moving objects, or changes in shooting conditions. This type of blur often has nonuniform characteristics, resulting in a substantial loss of high-frequency details in the images. High-frequency details are crucial for image clarity, recognition, and visual experience, making their resto-

收稿日期:2025-08-03;修回日期:2025-11-18;预印本日期:2025-11-25

*通信作者:翟清剑 13930856695@139.com

基金项目:国家自然科学基金项目(42577504)

Supported by: National Natural Science Foundation of China (42577504)

ration a research hotspot in image processing. However, existing deblurring methods typically perform poorly when confronted with these complexities, and the restored images often suffer from noticeable blurriness and detail loss. This phenomenon not only affects image quality but also limits performance in high-precision application scenarios, such as video surveillance, autonomous driving, and medical imaging. Therefore, designing a novel, effective motion image deblurring method that can address complex nonuniform blur situations has become an urgent issue that needs resolution. **Method** A motion image deblurring method that is based on a Transformer architecture and multiscale feature fusion is proposed in this study to resolve the challenges of motion deblurring. This method utilizes an encoder-decoder network structure, which is widely used in deep learning because of its excellent feature extraction and restoration capabilities in image tasks. In the encoder part, a module that combines convolutional neural networks (CNNs) and a dual-attention mechanism is designed. The advantage of CNNs is their ability to effectively extract local features from images. The dual-attention mechanism guides the model's focus to important areas in the image. This combination not only enhances the depth and breadth of feature extraction but also substantially improves the model's ability to capture key details, thereby providing a solid foundation for subsequent deblurring tasks. In the feature fusion section, a multiscale feature fusion module is employed to enhance feature expressiveness through a gated deep convolutional feedforward network and a feature enhancement module. The multiscale feature fusion strategy allows the network to integrate feature information from different scales, ensuring that details are preserved during the restoration process and effectively addressing varying degrees of blur across different regions of the image. The feature enhancement module extracts high-frequency information to enhance the clarity and realism of the final image, ensuring that critical details are not lost during reconstruction. In the decoder part, a modified Transformer module is introduced. Fourier transform is integrated into the feature extraction process within the feedforward network layer to enhance frequency-domain information. Fourier transform aids in processing images in the frequency domain, providing rich frequency information and effectively improving the model's ability to recover subtle features in blurred images. This innovative improvement offers a new approach to feature processing, allowing for a high level of restoration performance in actual deblurring tasks. **Result** To evaluate the effectiveness of the proposed method, extensive comparative experiments are conducted on GoPro and Human-Centric Indoor Deblurring (HIDE) datasets, and in-depth comparisons with existing mainstream methods are performed. Experimental results indicate that the proposed method achieves considerable advantages across multiple metrics. In the GoPro dataset, the peak signal-to-noise ratio (PSNR) reaches 32.70 dB, and the structural similarity index measure (SSIM) is 0.954, indicating that the quality of the restored images is very high, and details are largely recovered. In the HIDE dataset, PSNR reaches 30.53 dB, with an SSIM of 0.922, further validating the effectiveness and adaptability of the proposed method. Ablation experiments are performed to provide in-depth analysis of each module's contributions, verifying the positive effect of the proposed innovative points on motion image deblurring. These experiments clearly demonstrate the unique roles of each module and provide quantitative evidence that confirms the effectiveness of multiscale feature fusion and the Transformer module. **Conclusion** The motion deblurring method that is based on the Transformer architecture and multiscale feature fusion substantially outperforms existing mainstream methods in experiments conducted on GoPro and HIDE datasets. This research provides a novel solution to the critical issue of motion image deblurring and showcases the potential applicability of the proposed method in practical scenarios, particularly in image processing, video surveillance, and autonomous driving. Future research efforts could focus on extending the principles underlying this method to other image restoration tasks, such as denoising and super-resolution, further enhancing the model's versatility and robustness. Moreover, exploring the incorporation of domain-specific knowledge and prior information related to the nature of blur into the deblurring process could guide the model toward achieving highly favorable outcomes in real-world applications. Continual optimization and refinement of this innovative approach are expected to create opportunities in the field of image restoration. By doing so, this method has the potential to substantially contribute to the advancement of related technologies and applications, including but not limited to improving image quality in photography and enhancing the performance of image processing systems used in various industries. In summary, the proposed motion deblurring method represents a remarkable advancement in the ongoing effort to overcome the challenges posed by motion blur in images. With the advent of deep learning techniques and modern image processing methodologies, these advancements can be leveraged to develop solutions that directly address the limitations of

existing approaches. Whether utilized to enhance the quality of consumer photography, improve security camera footage, or refine images captured in autonomous driving systems, successful motion deblurring techniques have far-reaching implications. The link is <https://doi.org/10.57760/sciencedb.j00240.00170> and <https://github.com/zh7546/project.git>.

Key words: Transformer; multi-scale feature fusion; motion deblurring; attention mechanism; Fourier transform

0 引言

图像采集时,由于拍摄设备与拍摄对象之间发生相对运动(如设备抖动或对象运动),将产生运动模糊现象,对应的图像即为运动图像(Zhang等,2022)。在计算机视觉领域中,运动图像中存在的模糊是一个常见且极具挑战性的问题。在自动驾驶、安防监控、医学影像和航空航天等众多实际应用场景中,这种模糊会严重影响图像的视觉质量,并阻碍后续基于图像的分析、识别和决策等任务(罗炜麒等,2026)。因此,如何从模糊图像中精准地恢复出清晰的图像内容,成为亟待解决的关键问题。

运动图像中存在的模糊会导致图像细节丢失、边缘模糊,使得图像难以辨认,如图1所示。模糊的运动图像通常具有非均匀性,即图像不同区域的模糊程度不同,这给去模糊算法带来了额外的挑战。尽管近年来涌现出许多去模糊算法,但对于复杂场景、非均匀模糊以及高噪声等情况,现有的方法仍然存在诸多局限性,难以达到令人满意的效果。



图1 运动图像
Fig. 1 Motion images

传统的运动图像去模糊方法主要基于图像退化模型展开。这类方法通常假设图像在成像过程中受到模糊核的作用而发生退化,通过估计模糊核以及清晰图像实现去模糊的目标。

Ge等人(2022)提出了一种基于图像表面高斯曲率的盲图像去模糊算法,通过结合高斯曲率正则

化与图像梯度的范数,有效保留边缘并去除噪声,从而准确估计模糊核。Hsieh和Shao(2021)提出了强加于潜在图像的零补丁最小值约束,并通过将模糊图像中的补丁最小值信息重新分配回潜在图像来增强其结构细节,但对某些复杂场景或噪声较高的图像效果有限。Liu等人(2023)提出了使用显著图和梯度倒谱的快速盲图像去模糊算法,有效提升了去模糊效果并缩短了运行时间,但在复杂场景或面对特定类型的模糊(如运动模糊)时表现受限。这些传统方法往往对复杂场景和高噪声图像的处理效果有限。

深度学习技术的飞速发展运动图像去模糊领域带来了革命性的变化,特别是在Transformer模型和多尺度特征融合策略方面。

Transformer因其独特的自注意力机制(李现国和李滨,2023),能够有效捕捉长距离依赖关系(杨浩等,2024),已被引入到运动图像去模糊任务中。DeblurDiNAT(Liu等,2024)采用紧凑的编解码Transformer架构,通过交替扩张因子结构学习全局局部特征,在图像去模糊方面具有卓越的泛化能力,但在处理极端模糊时仍然难以确保细节的完整性。Waveletformernet(Zhang等,2024b)是基于Transformer的小波网络,通过嵌入离散小波变换和引入并行卷积,解决图像恢复中的细节丢失和颜色失真问题,但对于非常复杂的场景仍然存在失真风险。GTMNet(Li等,2023)结合卷积神经网络(convolutional neural network, CNN)和视觉Transformer(vision Transformer, ViT),通过引入引导传输图和增强模块改善遥感图像去雾效果,但在不同光照条件下鲁棒性不足。IGDFormer(illumination-guided dual Transformer model)(Wen等,2025)是用于低光照图像增强的光照引导双注意力视觉Transformer,通过结合通道注意力和窗口注意力机制,有效解决了现有方法中空间光照不一致、伪影和模糊等问题,但在处理强模糊情况下细节恢复的效果有限。Kacformer(Yang等,2025)结合了CNN和Transformer架构,通过知识迁移方法融合局部细节和全局上下文信息,提升了去雾效果。该方法还引入多尺度注意

力机制优化教师网络,但其多尺度注意力机制的复杂性使得训练过程更加困难,且可能影响实时性能。

在深度学习中,多尺度特征融合策略广泛应用于提升图像恢复效果(谢斌等,2025)。MS2Net(Niu等,2022)结合图像去模糊和超分辨率两个分支,并引入多尺度和多阶段特征融合模块,实现了对模糊图像的有效恢复和高质量重建,但仍需优化两者之间的平衡,防止信息丢失。MSFS-Net(multi-scale frequency separation network)(Zhang等,2023)利用频率分离和对比学习模块分别优化低频与高频信息,并采用了跨尺度特征融合,但融合过程中忽略了各尺度特征之间的相互依赖性。MHU-Net(Deng等,2024)基于U-Net实现非对称多尺度特征融合与高频损失注意力机制,但在高频损失处理上仍有提升空间,特别是在极端运动模糊下效果不佳。DSPF-Net(dual scale parallel spatial fusion network)(An等,2024)通过融合多尺度特征和选择性提取空间信息来增加特征表示,但其并行机制可能会导致信息冗余,降低特征提取的效率。DED-net(Dong等,2025)通过多尺度特征融合和光照引导恢复图像细节,但计算复杂性依然是制约其应用的一个问题。

针对上述研究的局限性,本研究旨在深入探索Transformer模型在运动图像去模糊中的应用潜力,并结合多尺度特征融合策略,提出一种运动图像去模糊网络架构。该方法充分发挥Transformer的全局建模能力和卷积网络的局部特征提取优势,借助特征融合机制,旨在解决非均匀模糊和细节丢失的挑战。具体而言,本文的主要贡献体现在以下几个方面:1)提出了一种结合卷积网络和通道-空间注意力模块(channel and spatial attention module, CSAM)的编码器结构,通过通道-空间注意力机制与深度可分离卷积结合,能够精准地提取多尺度特征。2)设计了一种局部特征Transformer块(local feature Transformer block, LFT-Block),在前馈层引入傅里叶变换和卷积,在自注意力层后添加深度可分离卷积,提升模型的特征表达能力。3)设计了一种多尺度特征融合模块(multiscale fusion module, MSFM),使用门控深度卷积前馈网络(gated-dconv feed-forward network, GDFN)(Zamir等,2022)和自适应特征增强模块(adaptive feature enhancement, AFE),通过多层结构优化输入特征。

1 提出方法

1.1 网络整体架构

本文提出的运动图像去模糊网络采用编码器-解码器结构,这种结构在诸多图像恢复任务中表现出色,能够有效地对图像进行特征提取和重建,其整体结构如图2所示。网络由3层编解码器组成,中间加入多尺度特征融合模块。编码器每层包括一个双注意力深度残差块(dual-attention depth-res block, DADR-Block),解码器前两层使用局部特征Transformer块,最后一层使用残差卷积模块(residual convolutional block, Res-Conv)。

首先输入一幅模糊图像 B ,经过3层编码器提取图像特征,不同层级的编码器分别输出不同尺度的特征图像,将3个层级输出的特征图像经过MSFM融合特征,得到 B_1 和 B_2 。在解码部分,先经过第1层的LFT-Block,第2层LFT-Block的输入为上一层的输出与 B_2 经过MSFM融合后的特征,第3层的残差卷积模块同样如此。Res-Conv的输出通过 3×3 卷积之后,与最开始的模糊图像通过跳跃连接相加,得到最后的去模糊图像 L 。

1.2 双注意力深度残差块

双注意力深度残差块的整体结构如图2中所示,包含卷积(Conv)、双重注意力模块和残差连接。先进行特征提取,再残差增强。采用卷积网络有助于高效提取局部特征,但在复杂任务中可能对冗余信息过度激活。为此,引入通道-空间注意力模块,使编码器能够自适应加权特征,并通过残差连接保留原始特征信息,缓解梯度消失问题。

双重注意力模块将空间注意力模块和通道注意力模块相结合,使得模型能够在两个层次上进行特征调整,其结构如图3所示。

SE(squeeze and excitation)模块(Hu等,2020)是用于增强卷积神经网络通道特征的一个重要组件,它通过自适应地调整每个通道的重要性改善模型性能,其结构如图3(b)所示。通过全局池化操作将每个通道的特征图压缩为一个标量值,这个标量值代表了该通道特征的全局统计信息。然后,通过两个 1×1 卷积组成的瓶颈结构对这些标量值进行非线性变换,学习通道间的注意力权重。最后,将学习到的注意力权重与原始特征图相乘,实现对特征的加

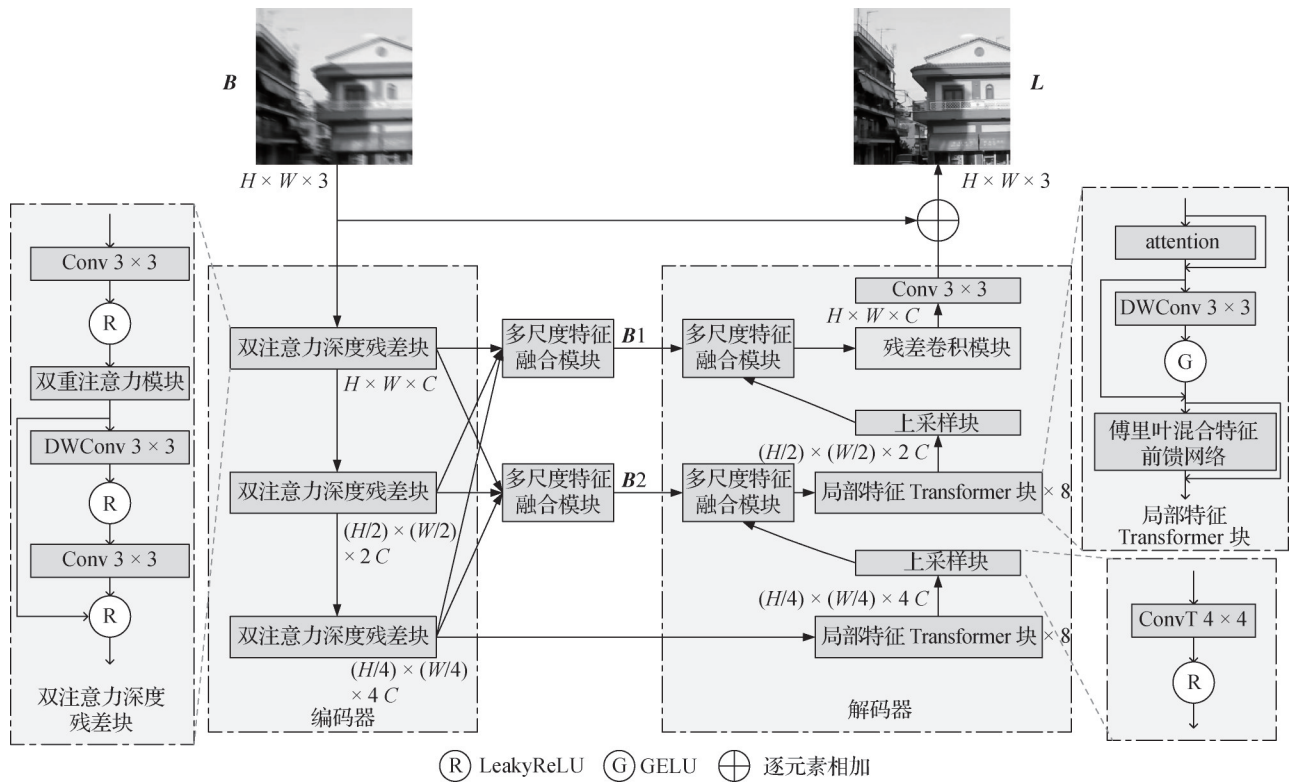


图2 网络结构

Fig. 2 Network architecture

权,加上原始特征(残差连接),从而增强重要通道的特征表达,具体为

$$y = \sigma \left(Conv_{1 \times 1} \left(Conv_{1 \times 1} \left(GlobalP(x) \right) \right) \right) \quad (1)$$

$$x_{ca} = x \times y + x \quad (2)$$

式中, \times 是逐元素相乘, σ 是 sigmoid 函数, $Conv_{1 \times 1}$ 是 1×1 的卷积层, $GlobalP$ 是全局池化。

CSAM 模块同时考虑了通道和空间两个维度的注意力,其结构如图 3(a)所示。它首先引入 SE 模块,在通道维度上通过全局池化操作,进行通道重标定。接着,在空间维度上,将经过通道注意力处理后的特征图进行处理,通过平均池化和最大池化得到两个新的特征图,随后将这两个特征图拼接并通过一个 7×7 的卷积生成空间注意力权重。这一过程允许模型突出重要的空间区域,从而增强特征的表达能力。并通过残差连接将处理后的特征图与原始特征图相加。具体为

$$x_{ca} = SE(x) \quad (3)$$

$$x_c = Cat(AvgP(x_{ca}), MaxP(x_{ca})) \quad (4)$$

$$x_{sa} = \sigma(Con_{7 \times 7}(x_c)) \quad (5)$$

$$x_{out} = x_{ca} \times x_{sa} + x \quad (6)$$

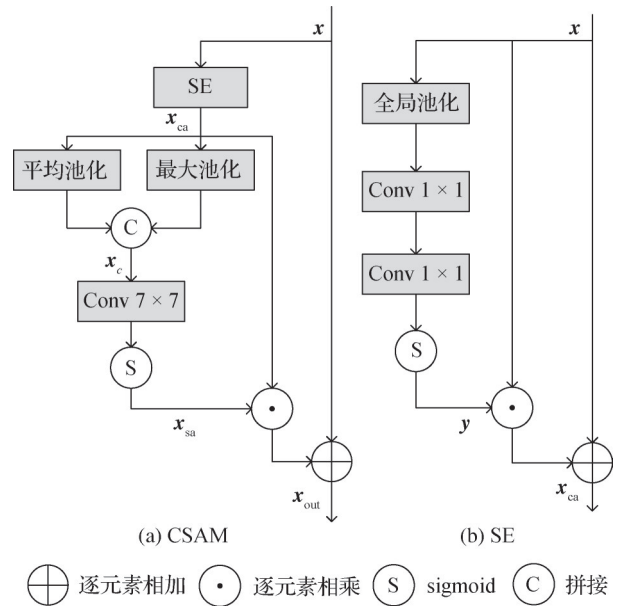


图3 双重注意力机制

Fig. 3 Channel and spatial attention module((a)CSAM;(b)SE)

式中, SE 是SE模块, Cat 是拼接操作, $AvgP$ 是平均池化, $MaxP$ 是最大池化, σ 是 sigmoid 函数, $Con_{7 \times 7}$ 是 7×7 卷积。

1.3 局部特征 Transformer 块

LFT-Block 的整体结构如图 2 中所示,包含注意

力层(attention)、卷积模块(DWConv+GELU(Gaussian error linear unit))和傅里叶混合特征前馈网络(Fourier mixed feature feedforward network, FM-FFN)。该模块使用双重局部增强设计,具体为空间域方面在自注意力层后插入 3×3 的深度可分离卷

积与GELU,频域方面通过傅里叶混合特征前馈网络的频域特征增强实现。

傅里叶混合特征前馈网络通过引入傅里叶变换以及多通道特征交互,有效提升特征表达能力,其结构如图4所示。

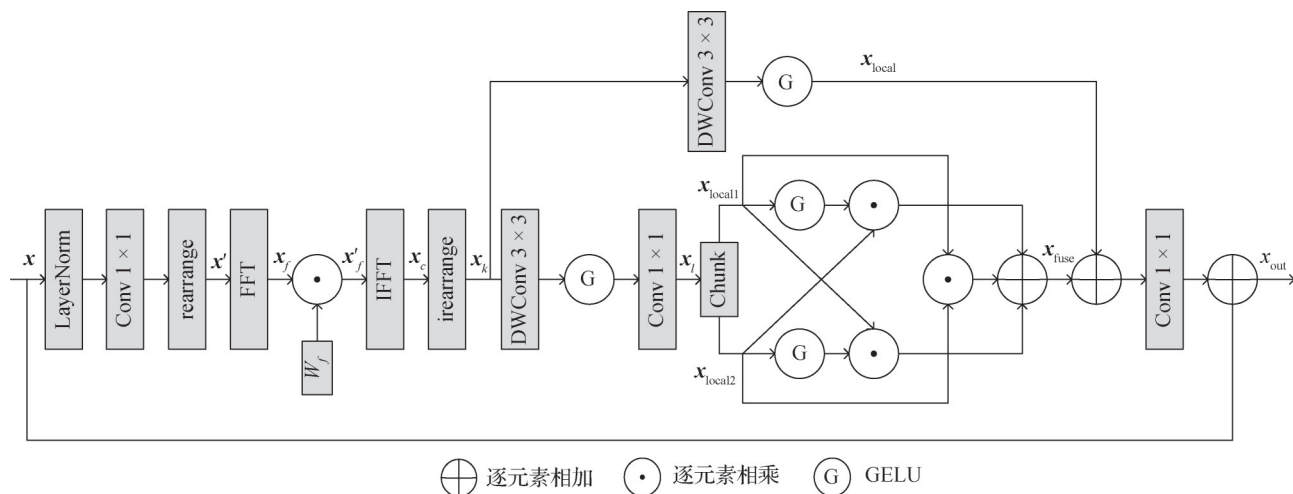


图4 傅里叶混合特征前馈网络

Fig. 4 Fourier mixed feature feedforward network

首先,输入特征图 x 进行层归一化处理,经过 1×1 的卷积进行通道重标定,生成隐藏特征。接下来,为了提取频域信息,特征图被分割为 8×8 的局部块,通过rearrange操作重塑为6D张量 x' ,以便在频域中进行处理。输入张量的形状为 $[B, C, H, W]$,分块后张量形状为 $[B, C, H/8, W/8, 8, 8]$ 。对重塑后的张量进行快速傅里叶变换(fast Fourier transform, FFT),利用得到的频域表示,可以捕捉到更丰富的频域特征。通过可训练的频域权重 W_f 对频域特征进行加权,以增强重要的频域成分。

W_f 初始化为全1的张量,意味着模型在最初阶段对快速傅里叶变换特征的处理是均匀的。 W_f 是一个可学习的参数,通过反向传播和梯度更新,可以自适应调整其值,以实现不同频域特征的选择性增强和抑制。 W_f 的更新规则为

$$W_f' = W_f - \eta \frac{\partial L}{\partial W_f} \quad (7)$$

式中, L 是L1损失函数,通过预测值和模型参数 W_f 之间的关系进行连接, η 是学习率。

当损失函数是L1损失时,损失函数与权重 W_f 的关系并不是直接的,而是通过预测值连接两者。具体而言,L1损失函数定义为

$$L = \sum_{i=1}^N |y_i' - y_i| \quad (8)$$

式中, y_i' 是模型针对第 i 个样本的预测值。 y_i 是第 i 个样本的真实标签, N 是样本总数。

在前向传播过程中,模型的输入通过网络各层后得到输出。此输出取决于模型的参数,包括 W_f 。因此,当模型参数发生变化时,预测输出会改变,从而影响损失,即

$$L = L(y', y; W_f) \quad (9)$$

式中, y' 是模型的预测, y 是真实标签, W_f 是可学习的权重参数。

然后,对转换后的特征块进行逆傅里叶变换(inverse fast Fourier transform, IFFT),以获取经过频域处理的特征图,具体为

$$x_f = FFT(x') \quad (10)$$

$$x_f' = W_f \times x_f \quad (11)$$

$$x_c = IFFT(x_f') \quad (12)$$

式中, x_f 为频域特征,FFT为快速傅里叶变换, W_f 为可训练的频域权重,IFFT为逆傅里叶变换。

使用逆傅里叶变换将处理后的频域特征转换回空间域,并重新排列成原始特征图的维度,得到特征图 x_k 。该特征图一方面依次通过深度可分离卷积、

GELU 激活函数、 1×1 卷积和特征拆分操作,生成两个输出 $\mathbf{x}_{\text{local1}}$ 和 $\mathbf{x}_{\text{local2}}$,通过非线性组合实现特征增强;另一方面, \mathbf{x}_k 通过 3×3 的深度卷积和 GELU 激活函数,得到输出特征图 $\mathbf{x}_{\text{local}}$ 。与经过频域增强的特征图相加,通过卷积模块映射到原始维度,再加上原始的输入特征 \mathbf{x} ,得到最终输出特征图。具体为

$$\mathbf{x}_l = \text{Conv}_{1 \times 1} \left(\text{Gelu} \left(\text{DWConv}_{3 \times 3} (\mathbf{x}_k) \right) \right) \quad (13)$$

$$\mathbf{x}_{\text{local1}}, \mathbf{x}_{\text{local2}} = \text{Chunk} (\mathbf{x}_l) \quad (14)$$

$$\mathbf{x}_{\text{fuse}} = \text{Gelu} (\mathbf{x}_{\text{local1}}) \times \mathbf{x}_{\text{local2}} + \text{Gelu} (\mathbf{x}_{\text{local2}}) \times \mathbf{x}_{\text{local1}} + \mathbf{x}_{\text{local2}} \times \mathbf{x}_{\text{local1}} \quad (15)$$

$$\mathbf{x}_{\text{local}} = \text{Gelu} \left(\text{DWConv}_{3 \times 3} (\mathbf{x}_k) \right) \quad (16)$$

$$\mathbf{x}_{\text{out}} = \text{Conv}_{1 \times 1} (\mathbf{x}_{\text{fuse}} + \mathbf{x}_{\text{local}}) + \mathbf{x} \quad (17)$$

式中, Chunk 表示将特征图拆分成两个相等的部分, $\text{DWConv}_{3 \times 3}$ 表示 3×3 的深度卷积, $\text{Conv}_{1 \times 1}$ 表示 1×1 的卷积, Gelu 表示 GELU 激活函数。

1.4 多尺度特征融合模块

MSFM 使用门控深度卷积前馈网络、自适应特征增强和残差连接,其结构如图 5 所示。

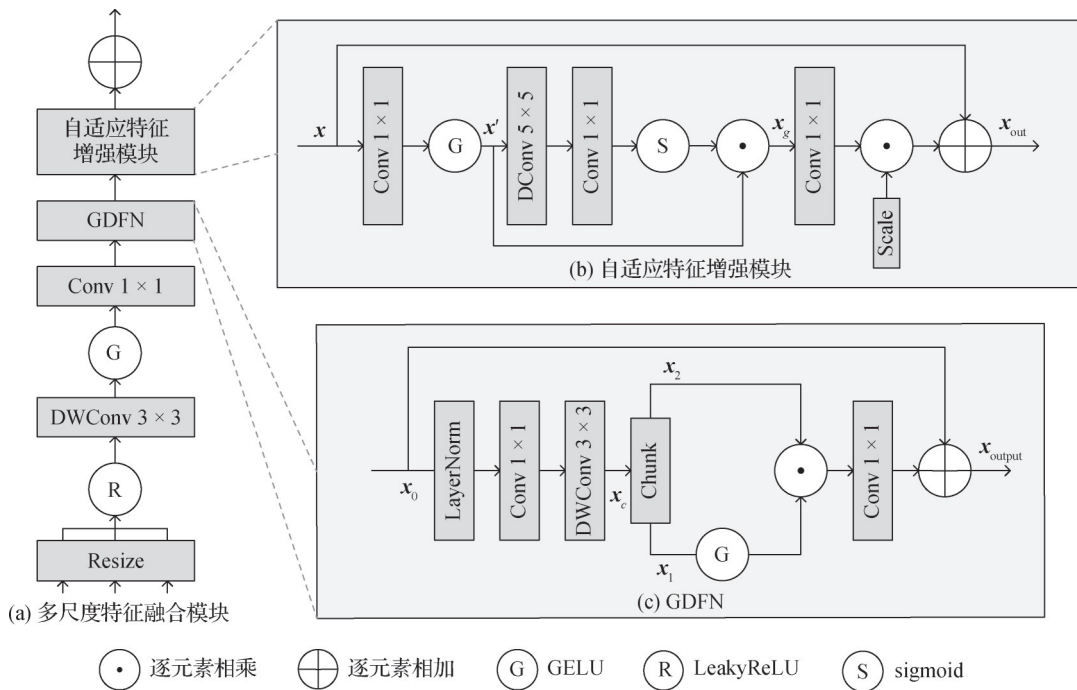


图 5 多尺度特征融合模块

Fig. 5 Multiscale feature fusion module((a)MSFM;(b)AFE;(c)GDFN)

MSFM 首先对输入特征图进行特征压缩,通过深度可分离卷积(3×3 的深度卷积 + 1×1 的卷积)将每个通道的特征保留在空间结构的信息中。使用前馈网络(GDFN)对压缩后的特征进行进一步的非线性变换,接着通过自适应特征增强,增强特征表达能力。在特征增强后,采用残差连接将原始输入特征与经过处理的特征进行相加。

1.4.1 GDFN

GDFN 结构如图 5(c)所示。它首先对输入 \mathbf{x}_0 进行归一化处理,通过层归一化(layer normalization)实现,接着通过一个 1×1 卷积将输入投影到更高维的特征空间。然后,使用 3×3 深度卷积提取特征并将

输出分成两部分,其中一部分经过 GELU 激活后与另一部分相乘以增强非线性特征。最后,经过另一个 1×1 卷积将处理后的特征投影回原始通道数,再与原始输入相加,生成最终输出。

$$\mathbf{x}_c = \text{DWConv}_{3 \times 3} \left(\text{Conv}_{1 \times 1} \left(\text{Norm} (\mathbf{x}_0) \right) \right) \quad (18)$$

$$\mathbf{x}_1, \mathbf{x}_2 = \text{Chunk} (\mathbf{x}_c) \quad (19)$$

$$\mathbf{x}_{\text{output}} = \text{Conv}_{1 \times 1} \left(\text{Gelu} (\mathbf{x}_1) \times \mathbf{x}_2 \right) + \mathbf{x}_0 \quad (20)$$

式中, Norm 表示层归一化。

1.4.2 自适应特征增强模块

AFE 的结构如图 5(b)所示,是一种集成了动态特征增强和残差连接的特征处理方法,旨在增强输

人特征的表达能力,以便为后续网络提供更高质量的特征表示。

首先,通过 1×1 的卷积对输入特征图 \mathbf{x} 进行通道压缩,降低模型的复杂性,再经过GELU激活得到 \mathbf{x}' 。然后,通过 5×5 的深度卷积捕捉输入特征图的局部空间特征,并使用 1×1 卷积生成特征图,特征图经过sigmoid激活生成一个加权系数。通过将特征图 \mathbf{x}' 与加权系数相乘,得到加权后的特征图 \mathbf{x}_g ,具体为

$$\mathbf{x}' = \text{Gelu}(\text{Conv}_{1 \times 1}(\mathbf{x})) \quad (21)$$

$$\mathbf{x}_g = \mathbf{x}' \times \sigma(\text{Conv}_{1 \times 1}(\text{DWConv}_{5 \times 5}(\mathbf{x}')))) \quad (22)$$

式中, σ 表示sigmoid函数。

随后,模块通过 1×1 的卷积恢复特征图的通道数,使其与输入特征图兼容。通过乘以可学习的缩放因子Scale,调整细化特征 \mathbf{x}_g 在最终输出中的影响力。Scale初始化为零,意味着一开始模型完全依赖于原始特征,具体为

$$f_{\text{Scale}} = f_{\text{Scale}}^{\text{old}} - \eta \frac{\partial L}{\partial f_{\text{Scale}}^{\text{old}}} \quad (23)$$

式中, L 是L1损失函数,它与模型参数 f_{Scale} 并非直接连接,而是通过模型的预测值连接, η 是学习率。 $f_{\text{Scale}}^{\text{old}}$ 为Scale参数上一次更新前的值。

Scale是一个可学习的标量,在模型训练时,PyTorch会计算这个参数的梯度,使得它可以通过反向传播算法自适应地调整,使得模型能够选择性地融合原始特征和细化特征。将经过缩放的细化特征与原始特征图 \mathbf{x} 相加,形成新的特征表示 \mathbf{x}_{out} ,具体为

$$\mathbf{x}_{\text{out}} = \mathbf{x} + f_{\text{Scale}} \times \text{Conv}_{1 \times 1}(\mathbf{x}_g) \quad (24)$$

2 实验与结果分析

2.1 数据集与实验设置

为了评估本文提出的运动图像去模糊方法的性能,实验在GoPro(Nah等,2017)和HIDE(human-centric indoor deblurring)(Shen等,2019)两个公开数据集上进行。

GoPro数据集包含了大量由GoPro相机拍摄的视频序列,其中的模糊图像是由于相机的运动和物体的快速移动所产生的。该数据集涵盖了多种类型的运动模糊,包括线性模糊、旋转模糊等,并且模糊程度各不相同,为评估方法在实际场景中的性能提

供了丰富的样本。

HIDE数据集着重模拟了复杂室内场景下的运动模糊情况,包含了办公室、客厅等多种室内环境的图像。在这些场景中,由于光线条件的变化、物体的遮挡等因素,图像的模糊情况更加复杂,对去模糊方法的鲁棒性提出了更高的要求。

实验的GoPro数据集,训练集包含2103对模糊—清晰图像对,测试集包含1111对图像。在GoPro数据集上训练得到的模型用于验证其在HIDE数据集图像上的泛化能力,测试集中包含2025对模糊图像和清晰图像。

实验使用PyTorch框架进行模型的搭建和训练,在硬件环境方面,实验在单张NVIDIA GeForce RTX V100上进行训练。优化器采用AdamW,初始学习率设为 1×10^{-4} ,使用余弦退火策略调整学习率,训练300轮,训练过程中采用L1损失函数作为优化目标。

2.2 评价指标

为了准确评估去模糊方法的性能,采用峰值信噪比(peak signal-to-noise ratio,PSNR)和结构相似性(structural similarity index,SSIM)作为主要评价指标。

PSNR用于衡量去模糊图像与真实清晰图像之间的像素级差异,值越高,说明去模糊图像与真实清晰图像之间的差异越小,去模糊效果越好。

SSIM评估图像结构信息的保持程度,它考虑了图像的亮度、对比度和结构3个方面的相似性。SSIM的取值范围在-1到1之间,值越接近1,表示去模糊图像与真实清晰图像的结构越相似,去模糊效果越好。

2.3 对比实验

为了评估本文方法在去模糊效果上的性能,在GoPro数据集和HIDE数据集上与8种模型进行了对比,包括DeepDeblur(Zhang等,2017)、DeblurGAN-v2(Kupyn等,2019)、DMPHN(deep multi-patch hierarchical network)(Zhang等,2019)、DBGAN(learning-to-DeBlur GAN)(Zhang等,2020)、MIMOUNet(multi-input multi-output U-Net)(Cho等,2021)、CODE(Zhao等,2023)、MRDNet(Zhang等,2024a)和DSANet(Jang等,2025)。各方法在GoPro数据集的定量比较结果见表1。从表1可以看出,本文方法在PSNR和SSIM指标上均取得了最优结果,PSNR达到了32.70 dB,与次优方法相比,PSNR提高了0.76 dB。

这表明本文方法在去模糊效果上具有明显的优势,能够更准确地恢复图像的细节。

为更直观地评估本文方法在去模糊视觉质量上的表现,从GoPro数据集中挑选了3幅模糊图像,分别与DeblurGAN-v2、DMPHN、MIMOUNet、CODE、MRDNet以及DSANet方法的去模糊结果进行了视觉对比。为突出局部细节差异,图6展示了这些去模糊结果的放大细节区域。

从图6的对比结果来看,对于模糊图a中的人物头发,本文方法得到的去模糊图像几乎没有伪影,并且形状基本还原。而除MIMOUNet外的其他5种方法都存在明显的伪影,MIMOUNet虽然头发的轮廓清晰,但是头发形状与清晰图像有很大偏差。对于

表1 GoPro数据集的测试结果

Table 1 Testing results of the GoPro dataset

方法	PSNR/dB	SSIM
DeepDeblur	29.23	0.916
DeblurGAN-v2	29.55	0.934
DMPHN	31.20	0.940
DBGAN	31.10	0.942
MIMOUNet	31.58	0.950
CODE	31.94	0.954
MRDNet	31.79	0.951
DSANet	31.65	0.949
本文	32.70	0.954

注:加粗字体为每列最优值。

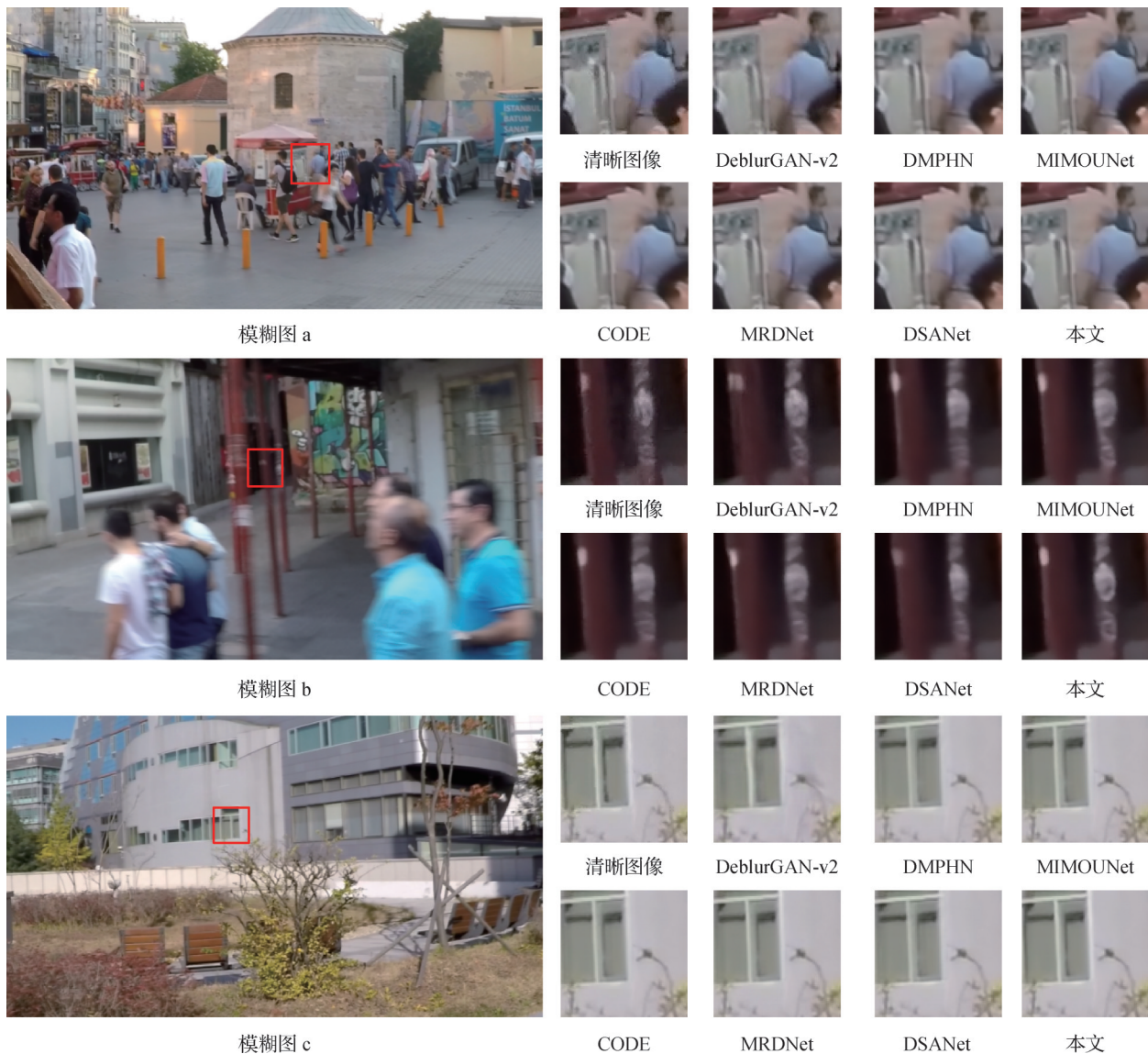


图6 不同算法在GoPro数据集上的视觉结果对比图

Fig. 6 Visual comparison of different algorithms on the GoPro dataset

模糊图b中的红色柱子,本文方法虽然与清晰图像有一定出入,但所恢复的锈迹细节相对锐利,图像更加清晰,其他6种方法得到的图像都没有清晰复原出红色柱子表面的锈迹。在模糊图c中的窗户里的阴影问题上,本文方法则能够更清晰地恢复阴影部分的形状,与清晰图像几乎一致。其他5种方法复原出的图像均与清晰图像里的阴影部分有着一定出入,并且相对于本文方法更加模糊。

为进一步验证本文方法的泛化能力,在覆盖更多场景和不同拍摄距离的HIDE数据集上进行了去模糊效果的对比实验。实验结果如表2所示,从中可以观察到,与其他对比方法相比,本文方法在PSNR指标上表现最佳,达到了30.53 dB。同时,本文方法在SSIM指标上取得了0.922。实验结果表明,本文方法具有较强的泛化能力。

表2 HIDE数据集的测试结果

Table 2 Testing results of the HIDE dataset

方法	PSNR/dB	SSIM
DeepDeblur	25.73	0.874
DeblurGAN-v2	26.61	0.875
DMPHN	29.10	0.918
DBGAN	28.94	0.915
MIMOUNet	29.22	0.917
CODE	29.67	0.927
MRDNet	29.36	0.921
DSANet	26.15	0.897
本文	30.53	0.922

注:加粗字体为每列最优值。

为了进一步验证本文方法的去模糊视觉效果,选取了HIDE数据集中的3幅模糊图像,并与DeblurGAN-v2、DMPHN、MIMOUNet、CODE、MRDNet和DSANet方法的去模糊结果进行了对比。为了使对比结果更为直观,对去模糊图像的细节部分进行了放大处理,展示出更清晰的细节如图7所示。

从图7的对比结果来看,对于模糊图a中的地板,本文方法成功恢复了地板纹路,并且相比较起来更加清晰。而其他6种方法都没有得到像清晰图像那样密集的地板纹路,得到的纹路很模糊并且与实际形状不符。对于模糊图b中的电梯扶手,本文方法则能够清晰地恢复扶手形状,几乎没有伪影。在

这方面,其他6种方法得到的去模糊图像的扶手上面的黑色长条偏窄,与清晰图像的扶手形状有一定出入。在模糊图c中的人脸模糊问题上,本文方法虽然留下了伪影,但与另外6种方法相比,所恢复的人脸细节更加锐利。DeblurGAN-v2、CODE和DSANet未能有效去除人脸区域的运动模糊。尽管另外3种在一定程度上减轻了面部模糊,但仍导致面部表情信息的丢失,与清晰图像之间存在差异。

为评估所提方法的模型效率,在GoPro测试集上进行了多个评价指标对比,结果见表3。实验衡量了各图像去模糊方法的峰值信噪比(PSNR)、模型参数量(Params)、浮点运算次数(floating point operations, FLOPs)及推理速度(运行时间)。表3数据显示,尽管DeblurGAN-v2在运行速度上表现最佳,但其他指标均显著落后于其他方法,去模糊效果欠佳。CODE在计算复杂度上表现最佳,但运行时间处于中下水平。相比之下,本文方法在Params和PSNR上具有最优性能,在FLOPs和运行时间上也优于大多数对比方法。

2.4 消融实验

为了深入验证本文方法中所提出的双重注意力模块、多尺度特征融合模块和傅里叶混合特征前馈网络的有效性,本文进行了消融实验。表4展示了不同配置下的性能对比。

1)方法1。去掉CSAM后,模型的PSNR下降到32.41 dB,SSIM下降到0.9516。

2)方法2。去掉CSAM中的SE,模型的PSNR下降到32.40 dB,SSIM下降到0.9517。

3)方法3。用SE替换CSAM,模型的PSNR下降到32.37 dB,SSIM下降到0.9514。这表明双重注意力模块能够有效地增强模型对重要特征的关注能力,提高特征提取的效果,从而提升去模糊性能。在CSAM模块中,首先SE模块通过全局信息调整通道权重,而后通过局部空间信息调整空间权重。两者的结合使得网络既能从全局(通道)层面调整特征的重要性,又能从局部(空间)层面选择关键区域进行强化。这两个注意力模块共同使网络在训练过程中能够自适应地调整关注的焦点,不仅在空间上聚焦重要的区域,而且在通道上筛选最具判别性的特征。

4)方法4。移除MSFM后,PSNR降低到32.41 dB,SSIM下降到0.9516。实验结果表明,MSFM对模型

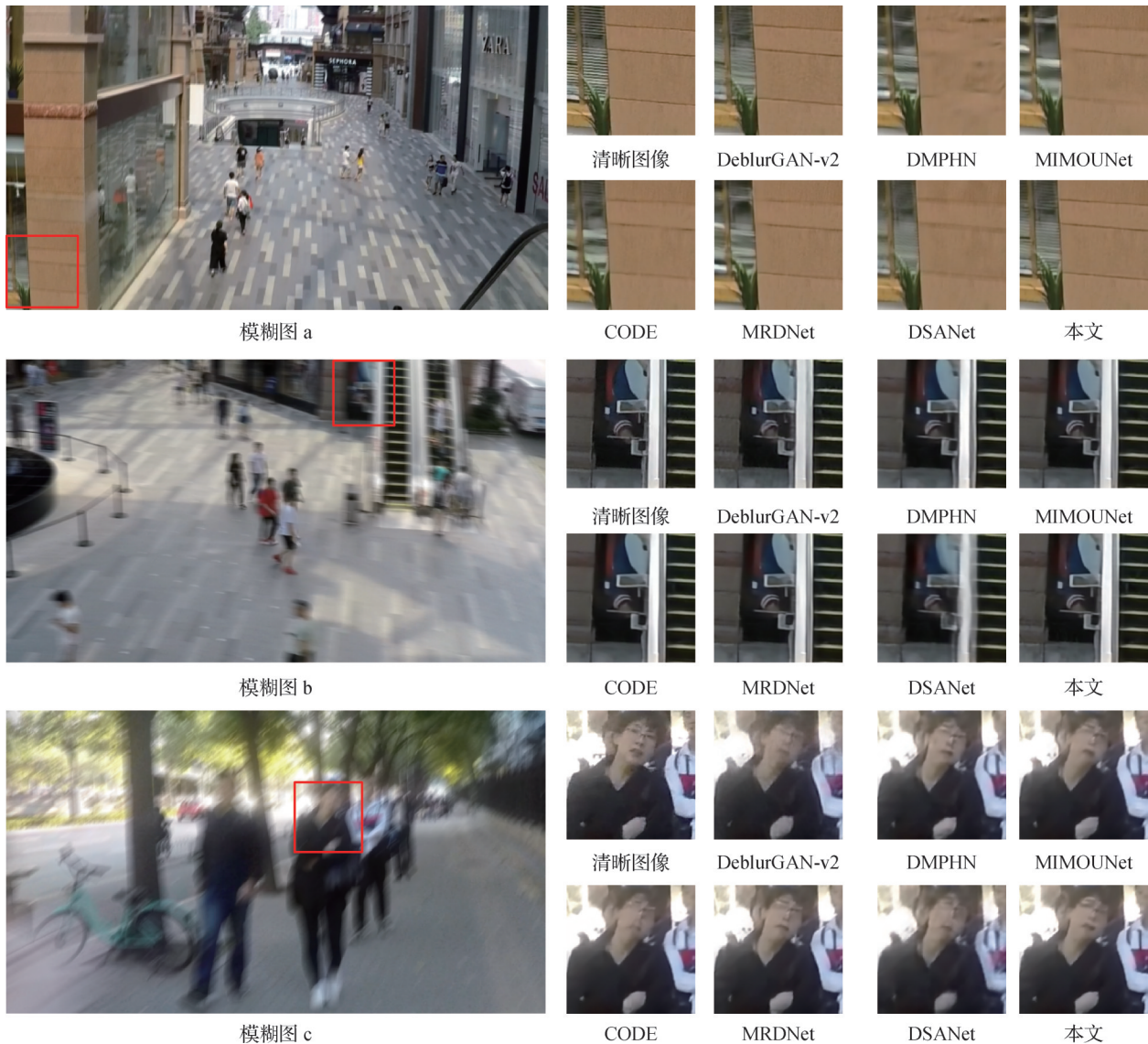


图7 不同算法在HIDE数据集上的视觉结果对比图

Fig. 7 Visual comparison of different algorithms on the HIDE dataset

表3 GoPro数据集的其他参数对比

Table 3 Comparison of other parameters on the GoPro dataset

方法	Params/M	FLOPs/G	运行时间/s
DeepDeblur	11.70	336.00	4.330
DeblurGAN-v2	60.90	42.00	0.041
DMPHN	21.70	235.00	0.315
DBGAN	11.60	759.85	1.298
MIMOUNet	6.80	62.87	0.230
CODE	12.18	22.52	1.775
MRDNet	7.10	72.66	0.677
DSANet	32.20	37.59	0.254
本文	6.11	45.86	0.043

注:加粗字体为每列最优值。

表4 消融实验

Table 4 Ablation experiments

方法	CSAM	SE	CSAM (无SE)	MSFM	FM-FFN	PSNR /dB	SSIM
1	-	-	-	√	√	32.41	0.951 6
2	-	-	√	√	√	32.40	0.951 7
3	-	√	-	√	√	32.37	0.951 4
4	√	-	-	-	√	32.41	0.951 6
5	√	-	-	√	-	32.21	0.949 8
本文	√	-	-	√	√	32.70	0.954 1

注:加粗字体为每列最优值,“√”表示该方法使用的模块,“-”表示未采用。

性能的提升起到了关键作用,通过结合 GDFN 和 AFE,有效处理输入特征,同时具备了动态学习与融合的能力。AFE 通过结合原始特征和细化特征,利用可学习的缩放因子实现了一种动态的特征增强策略。这个过程有效地提高了特征图的表示能力,使得网络能够更好地捕捉重要的特征信息,同时降低了模型的复杂性,增强了计算效率。特别是在对视觉任务的处理上,使得 MSFM 在特征融合方面变得更加高效和灵活。最终,这使得模型在复杂的图像理解和分析任务中表现得更加出色,能够更准确地捕捉关键而显著的特征信息。

5)方法5。在 Transformer 结构中,用 GDFN 替换 FM-FFN,相当于去掉了 FM-FFN 里的 FFT(快速傅里叶变换)和多通道特征交互,PSNR 为 32.21 dB,SSIM 下降到 0.949 8。傅里叶变换在这一过程中发挥了重要作用,能够有效提取图像的频率信息,使模型更深入地理解图像中的细节和结构。通过将图像转换到频域,模型不仅能捕捉到局部特征,还能在全局范围内调制这些特征,从而实现更精细的图像处理。

特征融合通过非线性组合强调了重要特征信息,使得模型在处理复杂图像内容时表现更加出色。同时,深度卷积增强机制进一步提升了对局部细节的敏感性,使得模型在处理复杂图像内容时表现得更加出色。在所有模块中,FM-FFN 对 PSNR 和 SSIM 的提升最大,分别达到了 0.49 dB 和 0.004 3。

为了更直观地验证 3 个模块的作用,进行了可视化分析。图 8 显示了消融实验的可视化结果,由于消融实验的数值差异较小,各结果图呈现出一定的相似性。其他 5 种方法虽然在一定程度上恢复了门框的形状,但与清晰图像仍存在差异。对于黑色门框旁边的银条,方法 1、方法 4 和方法 5 恢复的图像相对模糊;而方法 2 和方法 3 虽然较为清晰,但银条的形状表现过于平滑,与清晰图像存在出入。相较而言只有本文方法在清晰度和形状保真度方面与清晰图最相近。

综上所述,CSAM、MSFM 和 FM-FFN 均对模型性能提升有显著贡献,它们相互协作,共同提升了本文方法的去模糊性能。



图 8 消融实验的不同方法的视觉结果对比图

Fig. 8 Visual comparison of ablation study results of different methods

2.5 提出算法对后续视觉任务的影响

图像去模糊作为一种基础的图像处理技术,常用于更高级视觉任务的预处理步骤,例如目标检测。本文研究了去模糊处理对后续目标检测的影响,以 YOLOv5(you only look once version 5)算法为例进行了分析。研究发现,当输入 YOLOv5 算法的图像存在不均匀模糊时,检测的准确性会显著下降,并出现较大的误检和漏检情况。如图 9 所示,未经去模糊处理的图像,目标检测的识别率较低,且存在支架被误识别为碗、碗被错误识别成杯子以及书本被漏检

等问题。但经过本文提出的去模糊方法处理后, YOLOv5 算法对图像中检测目标的识别率和精确度都得到了显著提升。因此,本文方法可以有效减少图像模糊,提高后续视觉任务的处理性能。

3 结 论

本文提出了一种基于 Transformer 和多尺度特征融合的运动图像去模糊方法。通过结合卷积网络和双重注意力机制的编码器结构,能够有效提取多尺

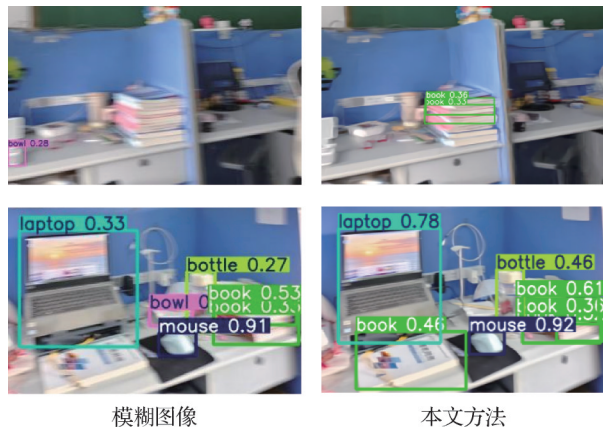


图9 目标检测对比

Fig. 9 Comparison of results of target detection

度特征;改进的Transformer解码器增加了局部增强,并引入傅里叶变换,增强了特征表达能力;多尺度特征融合模块增强特征表达。实验结果表明,本文方法在PSNR和SSIM指标均优于现有主流方法。消融实验进一步验证了各模块的有效性。

未来的研究方向包括进一步优化模型效率,通过剪枝、量化等技术减少模型的计算复杂度和参数量,使其能够在资源受限的设备上更高效地运行。此外,将进一步优化模型结构,提升运动去模糊的性能。

参考文献

- An H, Chen X X, Wang L, Hou B P, Jin Z C, Meng N, et al. 2024. Deblurring method for remote sensing image via dual scale parallel spatial fusion network. *Journal of Applied Remote Sensing*, 18(1): #014527 [DOI: 10.1117/1.JRS.18.014527]
- 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: 4621-4630 [DOI: 10.1109/ICCV48922.2021.00460]
- Deng Z J, Li G M, Yang H, Jiang P, Jiang H and Yu D L. 2024. The visual motion blur elimination method for silicon nitride bearing roller fissures based on U-Net asymmetric multi-scale feature fusion. *AIP Advances*, 14(5): #055227 [DOI: 10.1063/5.0212675]
- Dong H S, Cheng S H, Fang T Y, Chen J and Tu Z G. 2025. DED-net: multi-scale fusion and illumination-guided enhancement for low-light image restoration. *The Visual Computer*, 41(11): 9243-9255 [DOI: 10.1007/S00371-025-03924-7]
- Ge X Y, Tan J Q, Zhang L, Liu J and Hu D D. 2022. Blind image deblurring with Gaussian curvature of the image surface. *Signal Processing: Image Communication*, 100: #116531 [DOI: 10.1016/J.IMAGE.2021.116531]
- Hsieh P W and Shao P C. 2021. Blind image deblurring based on the sparsity of patch minimum information. *Pattern Recognition*, 109: #107597 [DOI: 10.1016/j.patcog.2020.107597]
- Hu J, Shen L, Albanie S, Sun G and Wu E H. 2020. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8): 2011-2023 [DOI: 10.1109/TPAMI.2019.2913372]
- Jang S W, Yan L M and Kim G Y. 2025. Deep supervised attention network for dynamic scene deblurring. *Sensors*, 25(6): #1896 [DOI: 10.3390/S25061896]
- Kupyn O, Martyniuk T, Wu J R and Wang Z Y. 2019. DeblurGAN-v2: deblurring (orders-of-magnitude) faster and better//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 8877-8886 [DOI: 10.1109/ICCV.2019.00897]
- Li H Q, Zhang Y P, Liu J T and Ma Y J. 2023. GTMNet: a vision transformer with guided transmission map for single remote sensing image dehazing. *Scientific Reports*, 13(1): #9222 [DOI: 10.1038/S41598-023-36149-6]
- Li X G and Li B. 2023. Image deblurring based on transformer and multi-scale CNN. *Computer Engineering*, 49(9): 226-233, 245 (李现国, 李滨. 2023. 基于Transformer和多尺度CNN的图像去模糊. *计算机工程*, 49(9): 226-233, 245) [DOI: 10.19678/j.issn.1000-3428.0065513]
- Liu H Z, Li B H, Liu C K and Lu M. 2024. Deblurdinat: a lightweight and effective transformer for image deblurring [EB/OL]. [2025-08-03]. <https://arxiv.org/pdf/2403.13163.pdf>
- Liu J, Tan J Q and He L. 2023. A fast blind image deblurring method using saliency map and gradient cepstrum. *The Visual Computer*, 39(7): 3091-3107 [DOI: 10.1007/S00371-022-02515-0]
- Luo W Q, Gao C, Liu H Y, Xia G S and Yu L. 2026. Event-based motion deblurring with dual-channel Mamba and pyramid channel attention. *Journal of Image and Graphics*, 31(1): 243-260 (罗炜麒, 高灿, 刘泓驿, 夏桂松, 余磊. 2026. 结合双通道Mamba与金字塔通道注意力的事件驱动运动图像去模糊. *中国图象图形学报*, 31(1): 243-260) [DOI: 10.11834/jig.250115]
- 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]
- Niu A X, Zhu Y, Zhang C N, Sun J Q, Wang P, Kweon I S, et al. 2022. MS2Net: multi-scale and multi-stage feature fusion for blurred image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8): 5137-5150 [DOI: 10.1109/TCSVT.2022.3153390]
- Shen Z Y, Wang W G, Lu X K, Shen J B, Ling H B, Xu T F, et al. 2019. Human-aware motion deblurring//Proceedings of 2019 IEEE/

- CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South) : IEEE: 5571-5580 [DOI: 10.1109/ICCV.2019.00567]
- Wen Y J, Xu P, Li Z H and Xu W T. 2025. An illumination-guided dual attention vision transformer for low-light image enhancement. *Pattern Recognition*, 158: #111033 [DOI: 10.1016/j.patcog.2024.111033]
- Xie B, Li Y X, Shao X and Dai B Q. 2025. Multiscale image deblurring based on prompt learning and gated feedforward networks. *Journal of Image and Graphics*, 30(3): 755-768 (谢斌, 黎彦先, 邵祥, 戴邦强. 2025. 提示学习与门控前馈网络的多尺度图像去模糊. *中国图象图形学报*, 30(3): 755-768) [DOI: 10.11834/jig.240315]
- Yang B Q, Wang M L, Fan J H and Liu J L. 2025. Kacformer: a hybrid CNN-transformer framework for image dehazing via knowledge transfer. *The Journal of Supercomputing*, 81(8): #922 [DOI: 10.1007/S11227-025-07429-X]
- Yang H, Zhou D M and Zhao Q. 2024. Image deblurring model combining gradient guidance and local enhancement transformer. *Small Microcomputer Systems*, 45(1): 216-223 (杨浩, 周冬明, 赵倩. 2024. 结合梯度指导和局部增强Transformer的图像去模糊网络. *小型微型计算机系统*, 45(1): 216-223) [DOI: 10.20009/j.cnki.21-1106/TP.2022-0344]
- 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]
- Zhang B, Sun J, Sun F M, Wang F S and Zhu B. 2024a. Image deblurring method based on self-attention and residual wavelet transform. *Expert Systems with Applications*, 244: #123005 [DOI: 10.1016/J.ESWA.2023.123005]
- Zhang H G, Dai Y C, Li H D and Koniusz P. 2019. Deep stacked hierarchical multi-patch network for image deblurring//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 5971-5979 [DOI: 10.1109/CVPR.2019.00613]
- Zhang K, Zuo W M, Chen Y J, Meng D Y and Zhang L. 2017. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7): 3142-3155 [DOI: 10.1109/TIP.2017.2662206]
- Zhang K H, Luo W H, Zhong Y R, Ma L, Stenger B, Liu W, et al. 2020. Deblurring by realistic blurring//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: 2734-2743 [DOI: 10.1109/CVPR42600.2020.00281]
- Zhang K H, Ren W Q, Luo W H, Lai W S, Stenger B, Yang M H, et al. 2022. Deep image deblurring: a survey. *International Journal of Computer Vision*, 130(9): 2103-2130 [DOI: 10.1007/s11263-022-01633-5]
- Zhang S L, Tao Z Y and Lin S. 2024b. WaveletFormerNet: a transformer-based wavelet network for real-world non-homogeneous and dense fog removal. *Image and Vision Computing*, 146: #105014 [DOI: 10.1016/j.imavis.2024.105014]
- Zhang Y N, Li Q, Qi M, Liu D, Kong J and Wang J Z. 2023. Multi-scale frequency separation network for image deblurring. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(10): 5525-5537 [DOI: 10.1109/TCSVT.2023.3259393]
- Zhao H Y, Gou Y B, Li B Y, Peng D Z, Lyu J C and Peng X. 2023. Comprehensive and delicate: an efficient transformer for image restoration//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 14122-14132 [DOI: 10.1109/CVPR52729.2023.01357]

作者简介

张冀,男,副教授,主要研究方向为计算机测控、故障诊断、智能信息处理、电力信息化。E-mail:72zhangji@163.com

翟清剑,通信作者,男,工程师,主要研究方向为图像处理、数字孪生、智能机器人。E-mail:13930856695@139.com

朱晓昀,女,硕士研究生,主要研究方向为图像处理。

E-mail:zh6_163o@163.com

曹锦纲,男,讲师,主要研究方向为数字图像处理、模式识别和智能系统。E-mail:caojg168@126.com