中图法分类号:TP18;TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2026)01-0243-18

论文引用格式: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):0243-0260(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):0243-0260)[DOI:10.11834/jig.250115]

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

罗炜麒[1],高灿[1],刘泓驿[1],夏桂松[2],余磊[2*]

1. 武汉大学电子信息学院,武汉　430072; 2. 武汉大学人工智能学院,武汉　430072

摘　要:**目的**　高时间分辨率的事件相机为传统运动图像去模糊任务提供新的发展思路,但是当前基于事件驱动的运动图像去模糊方法中存在跨模态补偿机制不足、深度特征计算复杂度较高以及缺乏多尺度时空信息关注的问题,在复杂场景中的去模糊泛化性能受限。针对以上挑战,提出一种双通道Mamba去模糊网络(dual channel Mamba network,DCM-Net)。**方法**　使用一种双通道跨模态Mamba模块(dual channel cross-modal Mamba,DCCM),通过线性复杂度的状态空间模型(state space model,SSM)隐状态映射,将事件与模糊图像投影至共享的潜在特征空间中,再通过非线性交叉门控结构,利用低噪声的模糊图像信息抑制事件噪声,并提取事件的清晰边缘特征,将其嵌入到图像特征中,实现事件和模糊图像的跨模态特征互补融合,达到去模糊的效果。此外,提出一种金字塔通道注意力模块(pyramid channel attention,PyCA)对特征的多尺度时空信息进行提取,引导网络聚焦关键时间通道,增强对空间内局部模糊的细节重建,进一步提高潜在清晰图像序列的复原精度。**结果**　实验在合成的REDS(realistic and diverse scenes)数据集与半合成的HQF(high quality frames)数据集上进行,与11种方法进行了比较。与DeMo-IVF方法相比,本文方法在REDS数据集重建序列的峰值信噪比(peak signal-to-noise ratio,PSNR)平均提升了0.16 dB,结构相似性指数(structural similarity,SSIM)平均提升了0.003;在HQF数据集上,PSNR和SSIM分别平均提升约0.11 dB和0.002;在两个数据集上的序列重建结果的学习感知图像块相似度(learned perceptual image patch similarity,LPIPS)达到最优。在与其中5种较先进方法进行比较的主观对比实验中,本文方法取得最佳评分。**结论**　本文方法可以结合模糊图像和事件数据,重建出清晰潜在图像序列,证明了所提网络框架的有效性。

关键词:运动图像去模糊;事件相机;Mamba模型;金字塔通道注意力;跨模态融合

Event-based motion deblurring with dual-channel Mamba and pyramid channel attention

Luo Weiqi[1], Gao Can[1], Liu Hongyi[1], Xia Guisong[2], Yu Lei[2*]

1. School of Electronic Information, Wuhan University, Wuhan　430072, China;

2. School of Artificial Intelligence, Wuhan University, Wuhan　430072, China

Abstract: **Objective** Motion deblurring is an ill-posed problem, because real-world blurred images result from the temporal integration of continuous scene dynamics and often lack information on intraframe motion and texture. Aiming to

收稿日期:2025-03-28;**修回日期**:2025-06-17;**预印本日期**:2025-06-24

*通信作者:余磊　ly.wd@whu.edu.cn

基金项目:中央高校基本科研业务费专项资金资助(204205kf0063);国家自然科学基金项目(62271354)

Supported by: Fundamental Research Funds for the Central Universities (204205kf0063); National Natural Science Foundation of China (62271354)

address the above challenge, recent work has leveraged event cameras, which are bio-inspired vision sensors with extremely high temporal resolution and provide intra-frame clues about motions and intensity textures, to perform the motion deblurring tasks. Despite efforts to enhance deblurring performance, most event-based methods depend on the simple concatenation of primary features, overlooking the substantial modal discrepancy between the spatial sparsity and temporal density of event data and the dense spatial characteristics of blurred images. This oversight restricts effective cross-modal feature interaction and ultimately leads to only marginal improvements in deblurring outcomes. Aiming to facilitate cross-modal fusion of events and blurred images, several methodologies that use Transformer-based cross-modal fusion architectures have been developed. However, the effectiveness of feature extraction and fusion processes within these cross-modal mechanisms depends on the quantity of structural components employed and the associated computational expenses. A limited number of structures may lead to insufficient deblurring performance, while extensive feature extraction often necessitates considerable computational resources. Furthermore, previous methods typically rely on convolutional layers or Transformer-based architectures that model global or windowed contexts within a singular receptive field along the channel dimension. This oversight regarding the integration of multiscale spatiotemporal information in the context of intricate motion scenarios leads to suboptimal reconstruction of local detail textures and a deterioration in generalization performance. Therefore, this study proposes a novel event-based motion deblurring framework called dual-channel Mamba network (DCM-Net) to address the above-mentioned issues. DCM-Net constructs a dual-channel module based on the Mamba model, efficiently achieving feature complementarity between the blurred image and events. Pyramid channel attention is also introduced to supplement multiscale spatiotemporal information in the features, ultimately achieving high-quality reconstruction of sharp latent image sequences. **Method** In this study, a module called dual-channel cross-modal Mamba (DCCM) is established. Specifically, DCCM employs state space model (SSM) with linear complexity to project the blurred image and events into a shared latent feature space. Subsequently, these representations are subjected to processing via nonlinear cross-gating architectures. DCCM integrates the features from both modalities using a nonlinear cross-gate structure that facilitates the learning of complementary features while mitigating redundancy. Aiming to minimize the influence of event noise, DCCM implements event gating thresholds that enable the selective extraction of features from blurred images that are relatively free of noise, which are then incorporated into the event features. Additionally, image features frequently suffer from texture loss due to the degradation caused by blurring. Given that event features retain clear texture information, they can be selectively integrated through the application of blurred image gating thresholds, thereby compensating for the loss and achieving a deblurring effect. Furthermore, to enhance the detailed reconstruction of networks in complex scenarios with non-uniform blur and non-linear motion, this research introduces a pyramid channel attention (PyCA) module based on the channel attention mechanism. Different from traditional channel attention techniques that depend exclusively on global pooling, PyCA not only extracts temporal feature information from vertical channels but also improves the acquisition of local blurred feature information within these channels by integrating multiscale pooling results. This approach enhances the network's focus on intricate details within the local spatiotemporal context of features, thereby facilitating the extraction and enhancement of detailed texture features in that specific region. Consequently, this refinement contributes to the overall improvement of the sequence reconstruction performance of the deblurring network. **Result** Comparative experiments on two public datasets, realistic and diverse scenes (REDS) and high quality frames (HQF), are conducted to validate the effectiveness of the proposed DCM-Net. The original REDS dataset comprises high frame rate videos recorded using the GoPro action camera. This study employs frame interpolation algorithms and event simulators to produce blurred images and simulated event streams. Conversely, the original HQF dataset comprises real events and high-resolution video frames that were captured concurrently using the DAVIS240 event camera. In this study, blurred images are synthesized from the clear video frames. This model is compared with 11 state-of-the-art motion deblurring models, including the frame-based and event-based methods. Three widely recognized metrics in the domain of image processing are employed to evaluate reconstruction quality: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS). The proposed method demonstrates superior performance on the REDS dataset, achieving an average PSNR of 30.166 dB and an average SSIM of 0.909 6, which demonstrates its superiority to other existing methods. Specifically, the proposed approach yields average PSNR and SSIM enhancements of

0.5% and 0.3%, respectively, compared to the leading state-of-the-art technique. Furthermore, on the HQF dataset, the method maintains exceptional performance, exhibiting average improvements of approximately 0.3% in PSNR and SSIM, thereby further confirming its effectiveness in handling real-world events. The proposed method also achieves state-of-the-art performance in average LPIPS for sequence reconstruction on REDS and HQF datasets. Additionally, visual comparisons and subjective experiments (user study) conducted on the two datasets further validate the effectiveness of the proposed method in this study. Considering the effective feature extraction capabilities inherent in the Mamba architecture, DCM-Net attains superior deblurring results while simultaneously exhibiting reduced computational complexity. Specifically, DCM-Net reduces computational overhead by over 40% in comparison to the leading state-of-the-art method that use the Swin-Transformer architecture and 27% compared to its own Transformer variant. Ablation studies on the sequence reconstruction of REDS dataset further validated the contributions of the key modules of DCM-Net. The removal of the DCCM module leads to a notable decline in PSNR and SSIM for DCM-Net. In contrast, the incorporation of PyCA in conjunction with DCCM still yields performance improvements (0.17 dB and 0.003 in PSNR and SSIM, respectively). **Conclusion** The DCM-Net introduced in this study enables effective cross-modal interaction and fusion between blurred image and event data while enhancing the model's focus on multiscale spatiotemporal information. DCM-Net outperforms existing methods, producing reconstructed latent image sequences with clear edge textures and minimal distortion. Furthermore, DCM-Net achieves an improved equilibrium among the number of model parameters, computational complexity, and overall performance.

Key words: motion deblurring; event camera; Mamba model; pyramid channel attention; cross-modal fusion

0 引　言

曝光期间被拍摄对象与镜头间的相对运动会导致运动模糊，并降低成像质量。传统的运动图像去模糊方法通常依赖一定的运动模式先验（Chen 等，2019；Liu 等，2020）或者强度纹理假设（Krishnan 等，2011；Pan 等，2016；Tai 等，2011），但这类人工设计的先验难以匹配现实场景中复杂的非线性运动与多尺度纹理变化，导致算法泛化性能受限。为了突破手工先验的限制，基于学习的方法（Chen 等，2021b；Cho 等，2021；Gong 等，2017；Nimisha 等，2017）被提出，利用卷积神经网络（convolutional neural network，CNN）以端到端的方式恢复潜在的清晰图像。然而，由于这类方法仅依赖单帧模糊图像的内部统计特性，缺乏对运动过程动态信息的显式建模，面对剧烈运动或复杂纹理场景时易出现重建失真。

事件相机作为新型仿生视觉传感器，通过异步检测像素级亮度变化，以微秒级别的时间分辨率输出稀疏事件流（Gallego 等，2022）。与传统相机相比，事件相机能连续捕捉动态场景，很好地弥补运动模糊中缺失的运动信息和边缘纹理信息。得益于上述特性，与传统的基于帧的方法相比，一些基于事件的方法（Jiang 等，2020；Lin 等，2020；Nimisha 等，2017；Shang 等，2021；Xu 等，2021；Zhang 等，2023）获得了更好的去模糊性能。然而，这些方法依赖于简单的初级特征拼接，直接将事件流与模糊图像输入网络。事件数据的空间稀疏性和时间密集性与模糊图像的密集空间特征之间，存在明显的模态不匹配，这限制了跨模态特征的有效交互（Cao 等，2025）。尽管部分工作（Chen 和 Yu，2024；Cheng 等，2023；Sun 等，2022，2025）基于视觉 Transformer（Khan 等，2022）提出跨模态融合注意力机制，以更好地融合事件与模糊图像的特征，但是其计算复杂度随网络深度呈平方级增长，难以平衡网络模型的计算效率与去模糊性能。此外，上述方法依赖于卷积层运算、Transformer 全局或窗口建模在通道内的单一感受野，在面对复杂运动场景时缺乏对多尺度时空信息的关注，导致局部细节纹理重建效果欠佳、泛化性能下降。

近年来，一种名为 Mamba 的新型网络架构在计算机视觉领域引起了广泛关注。与传统的 Transformer 架构相比，Mamba 架构通过线性隐状态映射机制，实现了对输入特征的全局依赖建模与增强，同时保持了线性计算复杂度，这使其在图像处理任务中展现出显著优势。

针对现有基于事件的运动图像去模糊方法在跨模态特征交互、计算开销以及多尺度信息关注面临的挑战，本文受 Mamba（Gu 和 Dao，2024）的启发，提出一种双通道 Mamba 去模糊网络（dual channel

Mamba network, DCM-Net)。网络采用一种先进的双通道跨模态Mamba模块(dual channel cross-modal Mamba, DCCM)作为主体框架,专门用于事件和模糊图像之间的跨模态特征学习与补偿。在Mamba将输入特征映射到潜在特征空间的基础上,通过交叉的门控结构,将模糊图像的绝对强度信息引入事件、利用图像低噪声特征抑制事件噪声,并将事件的尖锐边缘信息有效地嵌入到图像特征中,实现去模糊效果。同时,为了捕获多尺度信息以提升网络在模糊非均匀、运动非线性的复杂场景中的泛化性能,通过嵌入金字塔通道注意力模块(pyramid channel attention, PyCA),提取特征内的多尺度时空信息权重,引导网络增强对关键通道信息的关注以及局部模糊内的边缘纹理细节特征聚合,进一步提高潜在清晰图像的复原精度。

本文的主要贡献如下:1)提出一种双通道Mamba架构的事件驱动运动图像去模糊网络,实现了事件数据与模糊图像的跨模态特征提取与互补融合,重建出高质量潜在清晰图像序列;2)提出一种金字塔通道注意力模块,补充特征局部时空的关注信息,进一步提升运动图像去模糊质量;3)在合成与半合成的数据集上,验证了提出方法在不同运动场景下的去模糊效果,相对于现有去模糊算法具有更好的性能。

1 相关工作

1.1 基于图像帧的运动去模糊

运动去模糊是一个典型的病态逆问题,其核心挑战在于模糊过程对场景动态信息(例如,运动信息和纹理细节)的不可逆压缩特性。传统方法通过建立模糊核卷积模型(Zhang等,2022),将问题转化为正则化优化框架,并引入关于运动或强度纹理的人为先验假设(Chen等,2019;Fergus等,2006;Liu等,2020;Tai等,2011;方贤勇 等,2015;谭海鹏 等,2015;邱枫 等,2019;陈华华和鲍宗袍,2017)以提升求解稳定性。例如,Liu等人(2020)提出一种局部线性模糊核,以从略微模糊的图像中预测潜在的清晰图像;Chan 和 Wong(1998)使用全变分(total variation, TV)(Rudin等,1992)对模糊核进行约束,并采用定点算法对模型进行求解,在抑制噪声干扰的同时保留图像的尖锐边缘信息;Pan等人(2016)提出一种用于图像去模糊的暗通道先验,它利用暗通道的稀疏性来提高模糊核的估计并实现更好的去模糊效果。然而,人工设计的先验假设与真实场景下的统计特性存在本质差异,导致复杂动态场景下传统方法重建质量受限。

近年来,基于学习的方法(Chen等,2021b;Gong等,2017;Nimisha等,2017;吴迪 等,2020;黄彦宁等,2021)利用卷积神经网络(CNN)以端到端的方式预测潜在的清晰图像。尽管上述方法通过卷积神经网络隐式学习模糊到清晰的映射关系,缓解了先验假设的局限性,但其单模态输入特性仍难以应对大动态范围场景的运动模糊问题。近期研究尝试通过从粗略到精细策略(Cho等,2021),以及增强人工或数据驱动的先验(Bigdeli等,2017;Kupyn等,2018)来提高算法潜在锐化图像重建的性能,但未从根本上解决运动信息缺失的问题。

1.2 基于事件的运动去模糊

事件相机可以以极高的时间分辨率记录动态场景,提供丰富的运动信息和纹理细节,为运动去模糊任务提供了新的解决思路(Gallego等,2022)。早期基于事件的运动去模糊研究聚焦于物理模型构建(Pan等,2019;Scheerlinck等,2019;Wang等,2021),其中,Pan等人(2019)提出的基于事件的双积分(event-based double integral, EDI)模型具有里程碑意义。EDI通过积分事件流建立模糊图像与清晰图像间的数学关联,实现了基于事件的非均匀模糊建模。然而,EDI模型忽略了相机电路噪声对积分过程的干扰,导致实际应用中噪声积累显著降低图像重建精度。

为解决模型驱动方法的噪声敏感性问题,研究人员提出基于深度学习的方法,通过神经网络拟合噪声事件的分布(Lin等,2020;Wang等,2020a, c;Xu等,2021)。此外,针对监督学习和事件数据处理的新技术也相继提出,以增强基于事件的去模糊方法的泛化能力。例如,Xu等人(2021)设计了一个包含真实世界事件的半监督框架,以减轻不同数据差异导致的去模糊性能下降;Wang等人(2020a)、Zhang和Yu(2022)采用一种事件重排序策略实现曝光时间内的连续时间恢复。然而,这些方法普遍采用初级的特征拼接或逐元素乘积操作(Kim等,2022),未能充分挖掘事件流与模糊图像的模态差异特性,导致跨模态协同增效不足,性能提升存在

瓶颈。

视觉Transformer（Khan等，2022）的自注意力机制能够将输入信息映射到统一的多模态向量空间，展现出拓展至跨模态特征融合的潜力。最新的研究尝试结合视觉Transformer实现事件与模糊图像的融合。例如，Sun等人（2022）提出一种跨模态注意力结构，将事件特征嵌入到模糊图像中；Chen 和 Yu（2024）进一步使用跨模态交叉注意力将图像特征与事件特征相互嵌合，并通过可学习的权重进行特征融合。但这些方法局限于网络浅层的交互设计，未能构建层次化的跨模态特征融合路径，导致去模糊效果有限。Cheng等人（2023）在Swin-Transformer（Liu等，2021b）的基础上提出一种双通道注意力结构，并以级联形式对事件与模糊图像进行深度特征提取与融合，从而恢复出高质量的清晰潜在图像序列，但是该方法的多层结构使其在计算资源方面的消耗显著。而且上述方法依赖于单一的局部或全局建模机制，也缺乏对于垂直通道上时间信息的依赖建模，导致在复杂场景中对多尺度时空信息的关注不足，关键通道内局部模糊细节纹理复原效果不够理想。

1.3 Mamba与状态空间模型

Mamba模型的核心架构源于状态空间模型（state space model，SSM），其通过隐状态传递机制实现全局依赖建模，突破了传统CNN的局部感受野限制与Transformer的二次复杂度瓶颈。近期研究通过算法创新持续提升SSM的建模能力：Smith等人（2023）提出S5模型使得SSM的计算复杂性降低到线性水平；Mehta等人（2022）提出H3模型，进一步完善和扩展该模型，在语言建模任务中达到与Transformer相当的性能；Mamba（Gu和Dao，2024）则进一步引入一种输入自适应机制来增强SSM模型，与同等规模的Transformer相比，实现了更高的推理速度、数据吞吐量和整体指标。

Mamba架构自提出以来，迅速从自然语言处理领域扩展至计算机视觉任务，展现了其广泛的应用潜力。针对2D图像的空间特性，Liu等人（2025）提出VMamba架构，通过一种四向扫描算法构建视觉主干网络，在目标检测、分割和跟踪方面均展示了良好的性能。在此基础上，Shi等人（2024）则提出VmambaIR用于图像重建；邱云飞等人（2024）提出一种融合Mamba与蛇形卷积的网络框架，对图像进行去模糊操作。在跨模态融合方向，He等人（2024）提出Pan-Mamba用于遥感图像的全色锐化（即全色图像与多光谱图像融合）；Dong等人（2024）提出Fusion-Mamba用以融合RGB图像与红外光谱图像，实现跨模态目标检测。

本文的工作首次将Mamba架构引入事件驱动的运动去模糊任务中，通过设计一个双通道Mamba网络结构，实现事件与模糊图像之间的跨模态特征提取和融合，最终重建出高质量的清晰潜在图像序列。

1.4 通道注意力

通道注意力（channel attention，CA）作为深度神经网络中的重要组成部分，通过自适应地校准通道维度上的特征权重，提升模型在图像分类、目标检测等计算机视觉任务中的性能表现（Guo等，2022）。Hu等人（2018）构建的SENet（squeeze-and-excitation network）是通道注意力机制的代表模型，其主要通过压缩空间维度与激励通道维度，获得通道权重，并使用权重与输入特征进行相乘以更新特征。在SENet的基础上，后续研究从不同角度对通道注意力机制进行了改进和拓展。在特征压缩方面，Gao等人（2019）使用全局二阶池化进行压缩，以提高卷积网络的非线性能力；Qin等人（2021）基于频率分析角度，使用离散余弦转化压缩空间；在激励机制方面，Wang等人（2020b）通过一维卷积实现激励，提升局部通道内的信息交互能力。

尽管通道注意力机制能有效增强垂直通道上的特征表征能力，但在实际应用中往往需要与水平空间上的注意力机制协同使用，使整体网络获得更优性能（Guo等，2022）。为此，研究人员提出多种混合注意力架构：Woo等人（2018）设计了一种串行连接的通道—空间注意力模块，实现特征优化；Zhang等人（2020b）结合金字塔特征网络和通道注意力，实现了多尺度多通道特征的协同优化，以提升图像去雾效果；Si等人（2025）在通道注意力中引入单头自注意力，并构造并联的多语义空间注意力和渐进式通道注意力，在目标检测任务中增强空间引导和减轻语义差异。

2 本文方法

2.1 Mamba基本原理

Mamba模型和状态空间模型SSM受到线性系统

理论的启发不断发展，旨在通过一个隐状态空间$h(t) \in \mathbf{R}^N$将一维序列$\boldsymbol{x}(t) \in \mathbf{R}^N$映射表示为$\boldsymbol{y}(t)$。在此框架下，令$\boldsymbol{A} \in \mathbf{R}^{N \times N}$为状态演化参数矩阵，$\boldsymbol{B} \in \mathbf{R}^{N \times 1}$和$\boldsymbol{C} \in \mathbf{R}^{1 \times N}$分别为输入和输出投影参数矩阵，该系统的数学建模具体为

$$\boldsymbol{h}'(t) = \boldsymbol{A}\boldsymbol{h}(t) + \boldsymbol{B}\boldsymbol{x}(t) \tag{1}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{h}'(t) \tag{2}$$

为了将上述连续时间系统转化为离散计算框架，Mamba模型引入了时间尺度参数Δ以将连续参数$\boldsymbol{A}$和$\boldsymbol{B}$转换为对应的离散参数$\overline{\boldsymbol{A}}$和$\overline{\boldsymbol{B}}$。在此过程中，常用的离散化方法是零阶保持（zero-order hold，ZOH），可以通过下式定义，具体为

$$\overline{\boldsymbol{A}} = \exp(\Delta\boldsymbol{A}) \tag{3}$$

$$\overline{\boldsymbol{B}} = (\Delta\boldsymbol{A})^{-1}(\exp(\Delta\boldsymbol{A}) - \boldsymbol{I}) \cdot \Delta\boldsymbol{B} \tag{4}$$

基于式（3）和式（4），该线性系统的离散表达可以表示为

$$\boldsymbol{h}_t = \overline{\boldsymbol{A}}\boldsymbol{h}_{t-1} + \overline{\boldsymbol{B}}\boldsymbol{x}_t \tag{5}$$

$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{h}_t \tag{6}$$

式中，$\boldsymbol{h}_t$，$\boldsymbol{x}_t$和$\boldsymbol{y}_t$分别为$\boldsymbol{h}(t)$，$\boldsymbol{x}(t)$和$\boldsymbol{y}(t)$的离散形式。最后，通过全局卷积推导出输出结果$\boldsymbol{y}$，具体为

$$\overline{\boldsymbol{K}} = (\boldsymbol{C}\overline{\boldsymbol{B}}, \boldsymbol{C}\overline{\boldsymbol{A}}\overline{\boldsymbol{B}}, \cdots, \boldsymbol{C}\overline{\boldsymbol{A}}^{L-1}\overline{\boldsymbol{B}}) \tag{7}$$

$$\boldsymbol{y} = \boldsymbol{x} * \overline{\boldsymbol{K}} \tag{8}$$

式中，$\overline{\boldsymbol{K}} \in \mathbf{R}^L$表示结构化的卷积核，$L$表示输入$\boldsymbol{x}$的序列长度。

区别于基于线性运算建模的Mamba，在图像处理领域得到了广泛应用的视觉Transformer（Khan等，2022）则是通过多头注意力机制（multi-head attention），对输入向量独立计算查询（query）、键（key）和值（value）向量，并进行点积操作，实现对全局依赖关系的建模。但这种机制的计算复杂度随输入序列长度L呈平方增长$O(L^2)$，导致其在处理大规模数据时面临显著的计算开销问题。相比之下，Mamba通过对线性运算进行离散化建模，并采用递归方式更新隐藏状态与输出，实现了线性复杂度$O(L)$。特别是在涉及2D图像处理的应用场景中，Mamba的特性使其在减少计算开销方面展现出更大的潜力。

2.2 任务描述

本文提出的DCM-Net旨在利用模糊图像与事件数据，恢复固定数量的清晰潜在帧，实现功能为

$$\{\boldsymbol{I}_n\} = DCM\text{-}Net(\boldsymbol{B}, \boldsymbol{E}), \forall n \in \{1, \cdots, 2N+1\} \tag{9}$$

式中，$\{\boldsymbol{I}_n\}$表示恢复的目标图像序列，$\boldsymbol{B}$表示单幅模糊图像，$\boldsymbol{E}$表示曝光时间内事件流$\boldsymbol{\mathcal{E}}$经过压缩编码后的事件数据。在本文实验中，N设定为3。

2.3 事件表示

与传统光学相机以固定曝光时间进行成像的方式不同，事件相机的感光元件中每个像素独立且异步工作，能够感知场景对应位置的亮度变化，并生成包含时间、位置和极性信息的事件流（Gallego等，2022）。其成像计算为

$$\Delta\boldsymbol{I} = \left|\log(\boldsymbol{I}(\boldsymbol{x}, t)) - \log(\boldsymbol{I}(\boldsymbol{x}, \tau))\right| \geqslant c \cdot \left|p\right| \tag{10}$$

式中，$\Delta\boldsymbol{I}$表示两时刻光强在对数域的变化绝对值；$\boldsymbol{I}(\boldsymbol{x}, t)$与$\boldsymbol{I}(\boldsymbol{x}, \tau)$分别表示像素$\boldsymbol{x}$位置在$t$时刻与$\tau$时刻的瞬时光照强度；$c$为设定的阈值（$c > 0$）；$p \in \{+1, -1\}$表示事件的极性，其中，+1为正极性（亮度增加），−1为负极性（亮度减少），由亮度变化的符号决定，具体为

$$p = \text{sign}(\log(\boldsymbol{I}(\boldsymbol{x}, t)) - \log(\boldsymbol{I}(\boldsymbol{x}, \tau))) \tag{11}$$

式（10）表明，当某个像素在对数域上的亮度变化量超过阈值时，事件相机便会触发生成一个相应极性的事件点。

由于事件集合中的数据是离散且稀疏的，与传统图像的固定张量表示形式不同，因此需要对其进行压缩和特定形式的编码，以转化为适合神经网络处理的张量（Tensor）形式。本文采用一种基于时空表示的方法对事件数据进行处理：首先，将曝光时间内的事件集合$\boldsymbol{\mathcal{E}}$平均划分为K个时间段；然后，在每一个时间段内，分别统计每个像素点上正事件和负事件的总数，并构建各自的时间表面（time surface）（Gallego等，2022）。时间表面是一种二维平面映射，其中每个像素点存储该位置最后一个事件的时间戳。通过上述步骤，每个时间段可生成一个三维数据张量（$4 \times H \times W$），并将这K个时间段的数据张量合并为$4K \times H \times W$的多通道张量$\boldsymbol{E}$。这种处理方式同时保留了事件中的运动信息和时间信息，并减少与模糊图像之间的数据模态差异，有利于网络学习共享的隐藏空间映射与时空特征的提取交互。本文取切片数$K = 6$，其中$H \times W$表示对应模糊图像的空间分辨率。

2.4 基于事件的运动去模糊网络结构

本文提出的双通道Mamba网络DCM-Net结构如图1所示。DCM-Net主要由4种模块组成：浅层特征

提取(shallow feature extraction，SFE)模块、双通道跨模态Mamba(DCCM)模块、金字塔通道注意力(PyCA)模块和全局特征融合(global feature fusion，GFF)模块。其中，核心模块为DCCM模块与PyCA模块。

具体而言，网络的输入包括一帧模糊图像$\boldsymbol{B}$和曝光时间内事件流$\boldsymbol{\mathcal{E}}$经过压缩编码后形成的张量$\boldsymbol{E}$。首先，$\boldsymbol{B}$和$\boldsymbol{E}$分别通过SFE模块生成初始模糊图像特征张量$\boldsymbol{B}_f^0$和事件特征张量$\boldsymbol{E}_f^0$。随后，网络利用M个(本文设置$M=20$)DCCM模块对$\boldsymbol{B}_f^0$和$\boldsymbol{E}_f^0$进行更深层次的特征提取与交互。在每个DCCM模块中，模糊图像特征通道和事件特征通道均引入了PyCA模块，形成残差结构，用以补充多尺度时空信息相关的注意力机制，增强对关键通道内局部模糊的关注，以进一步提升去模糊效果。最后，GFF模块将每一层DCCM输出的、经过金字塔通道注意力修正的特征信息进行融合，并解码生成清晰的潜在图像序列$\{\boldsymbol{I}_n\}$。

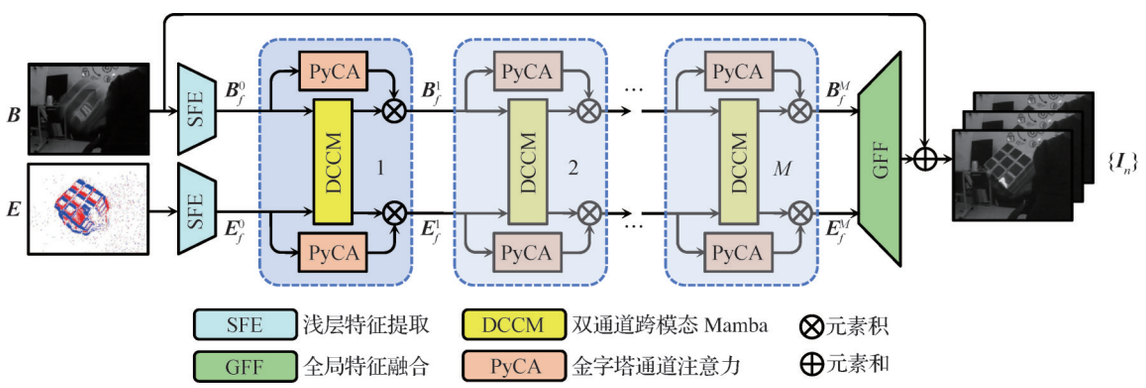


图1　DCM-Net网络架构

Fig. 1　DCM-Net network architecture

2.4.1　双通道跨模态Mamba模块

实现高质量的事件驱动运动图像去模糊，关键在于建立事件数据与模糊图像的特征提取与匹配融合机制(Sun等，2022)。最新的事件驱动运动图像去模糊方法普遍采用跨模态Transformer架构(Chen等，2021a)，通过生成查询(query)、键(key)和值(value)向量实现模态间交互。然而，这类方法存在两个显著局限性：1)特征融合过程多采用单向或浅层交互融合策略，未能充分挖掘跨模态的协同效应，且容易受到事件噪声干扰，去模糊效果欠佳；2)深层次的跨模态注意力机制虽然能够提升特征融合效果，但其计算复杂度显著增长。

针对上述挑战，本文受具有线性计算复杂度的Mamba架构启发，构造双通道跨模态Mamba模块(DCCM)实现深层次跨模态数据提取与融合，同时降低计算成本。DCCM模块架构如图2所示。DCCM中的SSM学习事件与模糊图像的特征分布，并利用隐藏状态映射将二者投影至共享的潜在特征空间，为后续的特征互补融合奠定理论和实现基础。一方面，压缩编码后的事件具有高频率的边缘与纹理信息，但不可避免地含有较多电路固有噪声；另一方面，图像受到模糊退化影响、高频细节信息丢失，但是保留有绝对强度信息且受到的噪声干扰也较少。这些模态特性差异为特征层面的互补融合提供了可能。在此基础上，DCCM基于非线性激活的门限机制(Liu等，2021a)设计交叉门控结构，直接响应事件与图像的模态差异，动态引导事件与图像特征间的相互学习与补偿，同时自适应抑制各模态中的冗余特征，从而实现特征融合。

在第i个($1\leqslant i\leqslant M$)DCCM模块中存在两条并行路径，分别处理输入的模糊图像特征$\boldsymbol{B}_f^{i-1}$和事件特征$\boldsymbol{E}_f^{i-1}$。这两类特征首先通过残差密集块(residual dense block，RDB)进行卷积操作与特征提取，以增强特征表达能力。随后，DCCM模块利用Mamba对$\boldsymbol{B}_f^{i-1}$和$\boldsymbol{E}_f^{i-1}$进行长距离依赖建模，进一步强化其特征表示。具体而言，$\boldsymbol{B}_f^{i-1}$和$\boldsymbol{E}_f^{i-1}$分别经过一个带有层正则化(LayerNorm)功能的线性投影层，将输入标准化。接着，这些特征沿着空间维度展平为一维特征，并输入到可学习的SSM中，投影至共享的潜在特征空间中。上述过程可以表示为

$$\begin{cases}\bar{\boldsymbol{B}}_f^{i-1}=SSM_B^{i-1}(Conv1D(Linear(RDB(\boldsymbol{B}_f^{i-1}))))\\ \bar{\boldsymbol{E}}_f^{i-1}=SSM_E^{i-1}(Conv1D(Linear(RDB(\boldsymbol{E}_f^{i-1}))))\end{cases}\tag{12}$$

式中，$\bar{\boldsymbol{B}}_f^{i-1}$和$\bar{\boldsymbol{E}}_f^{i-1}$分别为输出的潜在强化特征；算子$SSM(\cdot)$表示适用于2D图像的线性状态空间运算，由式(8)定义；算子$Conv1D(\cdot)$表示一维卷积操作；算子$Linear(\cdot)$表示线性投影操作；算子$RDB(\cdot)$表示残差密集块卷积操作。

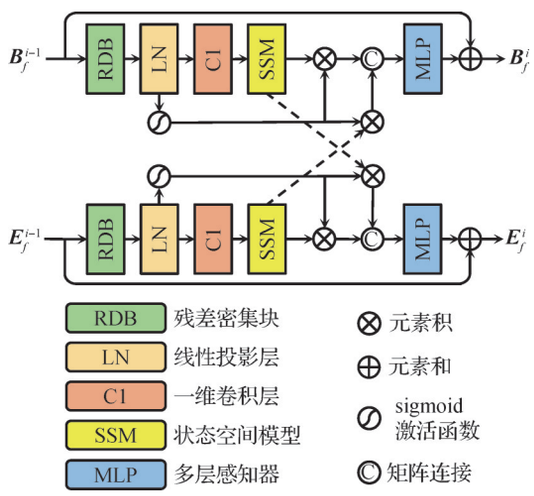


图2　双通道跨模态Mamba模块架构

Fig. 2　Dual channel cross-modal Mamba architecture

为了促进模糊图像特征与事件特征之间的跨模态交互与融合，DCCM模块将两种模态的特征投射到一个共享的特征空间后，通过交叉门控机制鼓励互补特征的学习，同时抑制冗余特征。具体而言，$\boldsymbol{B}_f^{i-1}$和$\boldsymbol{E}_f^{i-1}$在经过线性投影层后，一方面参与下一步的特征强化运算，另一方面被输入到sigmoid激活函数中，生成对应的门控阈值$\boldsymbol{g}_B^{i-1}$和$\boldsymbol{g}_E^{i-1}$。这一过程可描述为

$$\begin{cases}\boldsymbol{g}_B^{i-1}=\sigma(Linear(RDB(\boldsymbol{B}_f^{i-1})))\\ \boldsymbol{g}_E^{i-1}=\sigma(Linear(RDB(\boldsymbol{E}_f^{i-1})))\end{cases}\tag{13}$$

式中，算子$\sigma(\cdot)$表示sigmoid激活函数。门控阈值$\boldsymbol{g}_B^{i-1}$和$\boldsymbol{g}_E^{i-1}$用于选择性地控制模糊图像特征和事件特征的输入强度，从而实现两者的相互补偿与协同优化。

一方面，为了减轻事件噪声对特征提取的干扰，DCCM模块通过引入事件阈值选择性地提取几乎无噪声的模糊图像特征，并将其与事件特征进行融合。具体操作中，事件阈值$\boldsymbol{g}_E^{i-1}$分别与$\bar{\boldsymbol{B}}_f^{i-1}$以及$\bar{\boldsymbol{E}}_f^{i-1}$相乘，随后将相乘结果拼接并输入到多层感知器(multi-layer perceptron，MLP)中，输出更新后的事件特征，具体为

$$\boldsymbol{E}_f^i=MLP_E^{i-1}([\bar{\boldsymbol{B}}_f^{i-1}\cdot\boldsymbol{g}_E^{i-1};\bar{\boldsymbol{E}}_f^{i-1}\cdot\boldsymbol{g}_E^{i-1}])\tag{14}$$

式中，算子$MLP(\cdot)$表示多层感知器操作。

通过上述设计，DCCM模块能够有效减少事件噪声导致的错误特征估计，同时在事件中引入与图像绝对光照强度相关的信息。另一方面，由于模糊退化效应，模糊图像特征通常会因纹理损失而表现不佳，而事件特征则具有更清晰的纹理信息，因此可以利用事件特征对模糊图像特征进行补偿。为此，DCCM模块采用与式(14)类似的操作，通过$\bar{\boldsymbol{B}}_f^{i-1}$，$\bar{\boldsymbol{E}}_f^{i-1}$和门控阈值$\boldsymbol{g}_B^{i-1}$计算输出更新后的模糊图像特征，具体为

$$\boldsymbol{B}_f^i=MLP_B^{i-1}([\bar{\boldsymbol{B}}_f^{i-1}\cdot\boldsymbol{g}_B^{i-1};\bar{\boldsymbol{E}}_f^{i-1}\cdot\boldsymbol{g}_B^{i-1}])\tag{15}$$

由此，本文网络采用DCCM模块级联设计，在充分发挥线性复杂度优势、合理控制计算量的前提下，通过渐进式的跨模态特征交互补偿机制，借助低噪声图像特征抑制事件噪声的同时，保留与融合事件特征中的边缘纹理细节以及模糊图像中的绝对光照强度信息，最终达到图像去模糊效果。

2.4.2　金字塔通道注意力模块

运动模糊在空间分布上通常表现出非均匀特性，不同图像区域的模糊程度呈现明显的差异性(Zhang等，2020a)，需要充分挖掘空间上的多尺度信息，以提升网络对不同模糊程度的场景泛化性能。此外，在曝光时间范围内，物体运动往往表现出非线性动力学特征，这种时序上的非均匀性导致潜在清晰序列中的场景分布存在差异，此时需要提取关键时间维度的信息、引导边缘纹理细节特征聚合，以进一步提升目标潜在图像序列的重建质量。综合上述时空耦合问题，要实现高质量的运动图像去模糊序列重建，网络架构需要精确建模时空域内的信息关联，并自适应地处理不同区域的模糊特性。

前文介绍的DCCM模块对模糊图像与事件进行特征交互与融合，在特征域的空间维度上反映了混合的事件边缘纹理与图像强度信息，而在垂直方向上则反映了连续运动的时间动态信息。考虑到DCCM模块内的Mamba架构专注于通道内的长距离依赖建模与捕获全局特征信息，而缺乏空间多尺度感知与对多通道时间信息的依赖建模，本文在传统的通道注意力(CA)(Woo等，2018)基础上引入了一种类似于空间金字塔池化(He等，2015)的结构，并设计了一个金字塔通道注意力(PyCA)模块，对DCCM模块功能实现互补。PyCA模块架构如图3所

示。PyCA提取垂直通道上的时间特征信息的同时，也通过结合多尺度的池化结果，精细化地捕获通道内的局部模糊特征信息，提升对特征局部时空范围内的细节关注，引导网络对该区域内的细节纹理特征提取与增强，从而整体上提升了去模糊网络的序列重建效果。

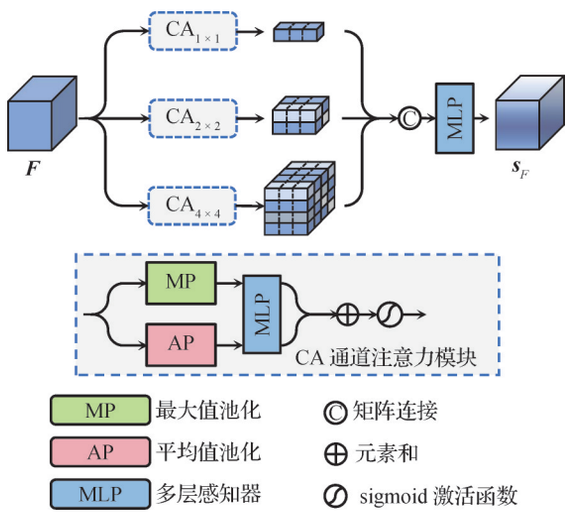


图3　金字塔通道注意力模块架构

Fig. 3　Pyramid channel attention architecture

具体而言，传统的 $n \times n$ 通道注意力机制（n 通常取1）首先对输入特征张量 $\boldsymbol{F}$（形状为 $C \times H \times W$）沿空间维度分别进行最大值池化（maximum pooling，MP）和平均值池化（average pooling，AP），生成两个形状为 $C \times n \times n$ 的特征向量。随后，这两个向量被送入一个权重共享的多层感知器（MLP）中进行学习。MLP的输出结果经过逐元素求和操作后，再通过sigmoid激活函数进行映射处理，最终输出每个通道的统计信息。这一过程可以表示为

$$CA_{n\times n}(\boldsymbol{F}) = \sigma(MLP(AvePool_{n\times n}(\boldsymbol{F})) + MLP(MaxPool_{n\times n}(\boldsymbol{F}))) \tag{16}$$

式中，算子 $CA_{n\times n}(\cdot)$ 表示 $n \times n$ 通道注意力计算；操作算子 $AvePool_{n\times n}(\cdot)$ 和 $MaxPool_{n\times n}(\cdot)$ 分别表示以 $n \times n$ 尺度输出的最大值池化和平均值池化操作。当 n 取1时，通道内进行全局池化，得到空间全局统计信息，但这种单一的池化方式容易导致对局部信息的关注丢失。

区别于传统的CA，PyCA模块在提取通道信息的基础上，为了充分利用输入特征的不同尺度信息，其在多个池化尺度（例如，1×1，2×2 和 4×4）上分别应用CA，提取每个尺度下的通道注意力特征，以确保输入特征向量 $\boldsymbol{F}$ 中的局部与全局信息都能得到有效利用。最后，所有尺度下的通道统计信息通过一个统一的MLP进行组合，计算输出最终的混合通道统计量 $\boldsymbol{s}_F$，具体为

$$\boldsymbol{s}_F = PyCA(\boldsymbol{F}) = MLP(CA_{1\times1}(\boldsymbol{F}); CA_{2\times2}(\boldsymbol{F}); CA_{4\times4}(\boldsymbol{F})) \tag{17}$$

式中，算子 $PyCA(\cdot)$ 表示金字塔通道注意力操作。

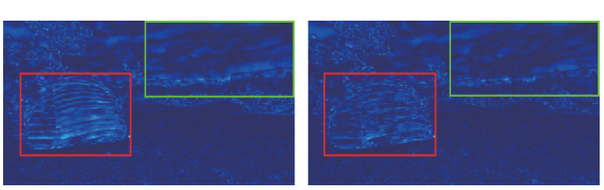
在 $\boldsymbol{B}_f^{i-1}$ 和 $\boldsymbol{E}_f^{i-1}$ 输入到第 i 个（$1 \leqslant i \leqslant M$）DCCM模块进行跨模态特征提取与交互的同时，也被输入到对应层数上的PyCA模块，计算混合通道统计量 $\boldsymbol{s}_B^i$ 和 $\boldsymbol{s}_E^i$，具体为

$$\begin{cases} \boldsymbol{s}_B^i = PyCA_B^{i-1}(\boldsymbol{B}_f^{i-1}) \\ \boldsymbol{s}_E^i = PyCA_E^{i-1}(\boldsymbol{E}_f^{i-1}) \end{cases} \tag{18}$$

随后，包含多尺度信息的混合通道统计量 $\boldsymbol{s}_B^i$ 和 $\boldsymbol{s}_E^i$ 分别与DCCM模块输出的模糊图像特征 $\boldsymbol{B}_f^i$ 和事件特征 $\boldsymbol{E}_f^i$ 相乘，自适应地对特征重新缩放并更新，可表示为

$$\begin{cases} \boldsymbol{B}_f^i \leftarrow \boldsymbol{s}_B^i \cdot \boldsymbol{B}_f^i \\ \boldsymbol{E}_f^i \leftarrow \boldsymbol{s}_E^i \cdot \boldsymbol{E}_f^i \end{cases} \tag{19}$$

图4为网络未使用PyCA与使用PyCA的去模糊结果的误差图。由图像可以分析，图4（a）（b）在模糊程度较小的背景（绿色框区域）所反映的去模糊性能基本一致，但是在模糊程度较大的局部前景（红色框区域）中，图4（a）对于这部分模糊区域的时空关注度不高，导致细节纹理重建效果欠佳，结果误差较为明显。相比前者，图4（b）通过引入PyCA增强了对局部模糊的时空感知，引导网络增强对于这部分区域的关注与处理，从而进一步提升去模糊效果。



（a）未使用 PyCA 误差图　（b）使用 PyCA 误差图

图4　去模糊结果误差图

Fig. 4　Error maps of deblurred result

（（a）error map without PyCA；（b）error map with PyCA）

2.5　损失函数

在一幅模糊图像由 $2N+1$ 幅清晰尖锐图像合成的基础上，本文选取这 $2N+1$ 幅清晰图像构成的序列 $\{\hat{\boldsymbol{I}}_n\}_{n=1}^{2N+1}$ 作为ground truth参考图像，对重建图像序列 $\{\boldsymbol{I}_n\}_{n=1}^{2N+1}$ 进行约束。本文采用L1范数定义损失

函数$\mathcal{L}$，具体形式为

$$\mathcal{L}=\frac{1}{2N+1}\sum_{k=1}^{2N+1}\left\|\boldsymbol{I}_k-\hat{\boldsymbol{I}}_k\right\|_1 \tag{20}$$

3 实验及结果分析

3.1 实验数据集

为了全面评估本文提出的DCM-Net的运动去模糊性能，本文采用两个公开数据集对网络进行训练与测试，分别为包含合成模糊图像和合成事件的REDS（realistic and diverse scenes）数据集（Nah等，2019）和包含真实世界事件和合成模糊图像的HQF（high quality frames）数据集（Stoffregen等，2020）。

原始的REDS数据集由尺寸为720×1 280像素、帧率为120帧/s的高帧率视频序列组成。为了模拟真实事件相机的输出，先将所有序列转换为灰度图像序列，并使用双线性插值法下采样到180×320像素，再使用插帧算法RIFE（real-time intermediate flow estimation）（Huang等，2022）将帧率提升至480帧/s。然后，使用生成的高帧率序列基于事件模拟器ESIM（event camera simulatiors）（Rebecq等，2018）的默认参数设置生成仿真事件流，对121幅连续帧进行平均以合成单幅模糊图像。每幅模糊图像对应原始视频中的31幅清晰图像。本文取其中第1，6，11，16，21，26和31帧作为ground truth参考图像，并遵循Nah等人（2019）的分配策略，取REDS数据集中240个子序列作为训练集，其余30个子序列作为测试集。

原始的HQF数据集包含DAVIS240事件相机（分辨率180×240像素）捕获的时空同步的真实事件和帧率为25帧/s的清晰视频帧。依照REDS数据集合成模糊图像的方法，将捕获的清晰视频进行插帧，使其帧率提升至200帧/s，然后对49幅连续帧进行平均以生成单幅模糊图像。每幅模糊图像对应于原始视频中的7幅清晰图像，这些图像将作为ground truth参考图像。HQF数据集有14个子数据集，本文从中选取9个作为训练集，其余5个作为测试集。

3.2 训练参数设计

本文网络模型基于PyTorch平台实现，并在两块NVIDIA GeForce RTX 3090 GPU进行训练。为提升训练效率和模型性能，在训练前对输入图像及其对应的事件数据进行预处理。具体而言，图像和事件数据被随机裁剪为128×128像素大小的块，同时采用水平翻转的方式进行数据增强。网络训练数据批次batch设定为4，迭代次数epoch设为200代，其中前50代的学习率固定为1×10^{-4}；之后学习率线性衰减，直到第200代时为1×10^{-5}。此外，本文构建了DCM-Net框架的Transformer版本变体模型（记为DCM-Net-T），并将其纳入对比实验体系以验证模型性能。

3.3 对比实验

3.3.1 定量对比实验

为了定量分析重建图像的质量，本文选用数据集中的参考图像作为评估基准，并采用目前图像处理领域中广泛使用的3种评价指标：峰值信噪比（peak signal-to-noise ratio，PSNR）、结构相似性指数（structural similarity，SSIM）和学习感知图像块相似度（learned perceptual image patch similarity，LPIPS）（Zhang等，2018）。通常情况下，PSNR越高，表明图像的重建质量越好，失真程度越低；SSIM越高，表示重建图像与参考图像在亮度、对比度和结构信息上的一致性越强；LPIPS则利用模型计算两幅图像之间的感知相似性分数，数值越低表示二者之间视觉感知差异越小。

实验将本文方法与当前先进的基于帧和基于事件的去模糊方法进行全面比较。参与对比的方法包括HiNet（half instance normalization network）（Chen等，2021b）、MIMO-UNet（multi-input multi-output U-Net）（Cho等，2021）、EDI（event-based double integral）（Pan等，2019）、eSL-Net（event-enhanced sparse learning network）（Wang等，2020a）、TRMD（two-stage residual-based motion deblurring）（Chen和Yu，2024）、EVDI（event-based video deblurring and interpolation）（Zhang和Yu，2022）、LEDVDI（learning event-driven video deblurring and interpolation）（Lin等，2020）、EFNet（event fusion network）（Sun等，2022）、RED-Net（Xu等，2021）、UEVD（unknown exposure time videos deblurring）（Kim等，2022）和DeMo-IVF（Cheng等，2023）。其中HiNet和MIMO-UNet是基于帧的方法，其余方法则结合事件数据和模糊图像帧进行处理。此外，HiNet、MIMO-UNet、EFNet和UEVD仅能实现曝光时间内的单帧清晰图像重建，而其他方法则支持包含7幅潜在清晰图像的序列重建。

表1和表2分别展示了所有方法在REDS数据集和HQF数据集上的定量对比结果，其中PSNR与SSIM在图像灰度通道上进行计算。实验结果表明，在REDS数据集上，本文方法在PSNR（平均值为30.166 dB），SSIM（平均值为0.909 6）和LPIPS（平均值为0.059 7）3项指标上均优于现有方法。相较于DeMo-IVF，本文方法在PSNR指标上平均提升约0.16 dB，在SSIM指标上平均提升约0.003。在HQF数据集上，本文方法继续表现出良好的性能，PSNR和SSIM分别平均提升约0.11 dB和0.002，进一步验证了其在处理真实世界事件数据时的有效性。

表1　不同方法在REDS数据集上的性能对比

Table 1　Comparison of performance of different methods on REDS dataset

方法	事件数据	单帧重建			序列重建			FLOPs/G	Para/M
		PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS		
HiNet（Chen等，2021b）	×	22.697	0.668 0	0.248 3	–	–	–	80.78	5.41
MIMO-UNet（Cho等，2021）	×	22.978	0.690 9	0.211 0	–	–	–	135.29	16.11
EDI（Pan等，2019）	√	21.836	0.649 0	0.313 3	20.410	0.612 0	0.321 3	–	–
eSL-Net（Wang等，2020a）	√	21.046	0.687 1	0.208 6	21.414	0.687 1	0.202 1	138.23	**0.19**
TRMD（Chen和Yu，2024）	√	24.798	0.748 0	0.168 3	23.037	0.674 2	0.214 4	**29.99**	19.26
EVDI（Zhang和Yu，2022）	√	26.868	0.836 7	0.095 9	26.014	0.812 0	0.107 8	<u>52.91</u>	<u>0.39</u>
LEDVDI（Lin等，2020）	√	28.475	0.878 2	0.068 8	27.771	0.866 5	0.072 1	127.08	5.00
EFNet（Sun等，2022）	√	29.427	0.890 2	0.075 4	–	–	–	94.66	8.46
RED-Net（Xu等，2021）	√	29.598	0.898 9	0.058 3	28.948	0.888 6	0.064 1	140.65	9.76
UEVD（Kim等，2022）	√	30.166	0.911 3	**0.052 8**	–	–	–	285.85	27.88
DeMo-IVF（Cheng等，2023）	√	30.712	0.916 6	0.056 8	30.004	0.907 0	0.060 2	483.39	30.88
DCM-Net-T（本文）	√	<u>31.026</u>	<u>0.920 8</u>	0.054 9	<u>30.139</u>	<u>0.909 3</u>	**0.057 8**	372.62	23.70
DCM-Net（本文）	√	**31.057**	**0.922 1**	<u>0.054 7</u>	**30.166**	**0.909 6**	<u>0.059 7</u>	270.62	23.57

注：加粗、下划线字体分别表示各列最优、次优结果。“–”表示无指标结果。“√”和“×”分别表示使用与不使用。FLOPs表示每秒浮点运算次数，Para表示参数量。

图5和图6分别展示了REDS和HQF数据集上的可视化对比结果。从视觉效果可以看出，本文方法在运动模糊去除和细节保留方面达到了与当前最先进的方法相当的水平，充分体现了其在复杂场景中的优越性能。

模型驱动的EDI（Pan等，2019）仅依赖像素级别的事件积分操作，未能充分利用事件数据中蕴含的空间结构信息，从而导致重建结果仍然存在模糊。在基于CNN的网络模型中，HiNet（Chen等，2021b）和MIMO-UNet（Cho等，2021）在缺乏事件数据提供运动信息的情况下，试图从剧烈运动导致的模糊图像中恢复潜在的清晰图像面临显著挑战，其结果往往不可避免地出现扭曲或变形；虽然eSL-Net（Wang等，2020a）可以通过学习抑制事件噪声，但其网络中的迭代运算机制在推理阶段容易放大重建误差，导致预测图像中出现伪影和显著的噪声残留；EVDI（Zhang和Yu，2022）基于EDI模型构建了一种自监督深度学习框架，但由于缺乏对事件与模糊图像之间跨模态融合的有效机制，其重建结果在细节呈现上缺乏精细；LEDVDI（Lin等，2020）的性能受到网络中动态滤波器设计的限制，导致部分细节纹理的丢失，影响整体重建质量；RED-Net（Xu等，2021）利用大量堆叠的残差密集块（RDB）预测相对清晰的图像序列，但由于缺乏注意力机制，其在细节提取能力上存在一定局限性；UEVD（Kim等，2022）通过较多的计算开销在网络中选择性地使用事件特征，但是忽视了事件和模糊图像之间的跨模态相互补偿，导致网络性能提升有限。

在基于Transformer的网络模型中，EFNet（Sun等，2022）引入了具有注意力机制的跨模态融合结构

表2　不同方法在HQF数据集上的性能对比

Table 2　Comparison of performance of different methods on HQF dataset

方法	事件数据	单帧重建			序列重建			FLOPs/G	Para/M
		PSNR/dB	SSIM	LPIPS	PSNR/dB	SSIM	LPIPS		
HiNet(Chen等,2021b)	×	24.071	0.727 6	0.230 5	–	–	–	60.58	5.41
MIMO-UNet(Cho等,2021)	×	24.456	0.744 4	0.206 0	–	–	–	101.47	16.11
EDI(Pan等,2019)	√	19.665	0.629 5	0.295 0	19.014	0.611 1	0.313 8	–	–
eSL-Net(Wang等,2020a)	√	20.576	0.618 1	0.206 6	20.011	0.605 1	0.208 5	103.67	**0.19**
TRMD(Chen和Yu,2024)	√	26.128	0.783 0	0.161 4	24.606	0.739 7	0.199 2	**22.49**	19.26
EVDI(Zhang和Yu,2022)	√	26.410	0.801 6	0.147 5	25.288	0.777 2	0.155 4	39.68	0.39
LEDVDI(Lin等,2020)	√	28.208	0.847 3	0.111 9	27.698	0.841 5	0.115 8	95.31	5.00
EFNet(Sun等,2022)	√	29.886	0.879 5	0.101 8	–	–	–	70.99	8.46
RED-Net(Xu等,2021)	√	30.336	0.882 5	0.096 5	29.498	0.875 3	0.098 1	105.48	9.76
UEVD(Kim等,2022)	√	30.316	0.886 9	0.092 0	–	–	–	214.38	27.88
DeMo-IVF(Cheng等,2023)	√	31.299	0.900 8	0.082 6	30.484	0.894 2	0.083 3	386.30	30.88
DCM-Net-T(本文)	√	31.374	0.902 0	0.083 6	30.512	0.895 8	**0.082 5**	297.26	23.70
DCM-Net(本文)	√	**31.461**	**0.904 1**	**0.081 8**	**30.590**	**0.896 5**	0.083 2	202.97	23.57

注:加粗、下划线字体分别表示各列最优、次优结果。“–”表示无指标结果。“√”和“×”分别表示使用与不使用。FLOPs表示每秒浮点运算次数,Para表示参数量。

将事件和图像特征融合,但缺乏深度提取特征能力,导致在处理剧烈运动模糊时,其结果仍然可能出现变形和扭曲;TRMD(Chen和Yu,2024)采用两阶段网络设计,通过事件数据从模糊图像中学习有限数量的残差帧,但该方法主要适用于模糊程度较低的场景,在复杂环境下仍难以完全消除模糊伪影;DeMo-IVF(Cheng等,2023)基于Swin-Transformer构建深度网络,实现事件与模糊图像特征的跨模态交互与融合,能够有效提取特征并实现图像去模糊。然而,受限于Transformer架构本身较高的计算复杂度,该方法在实际应用中面临着较高的计算成本压力。

相较于上述方法,本文提出的基于双通道Mamba架构的DCM-Net网络模型不仅实现事件与模糊图像特征的跨模态交互与融合,也通过构建金字塔注意力机制,补充多尺度的时空局部信息,提升对局部模糊的关注,重建出较高质量的清晰图像,验证了本文框架的有效性。

除了图像质量评价指标(如PSNR),网络的参数量与反映计算复杂度的每秒浮点运算次数(floating point operations per second,FLOPs)同样是衡量模型性能的重要维度。以REDS数据集单帧重建结果为例分析,如图7所示。得益于Mamba架构高效的特征提取能力,DCM-Net通过深度多层网络设计实现高质量去模糊效果的同时,也将计算复杂度控制在合理范围内。尽管DCM-Net的计算量相较于基于CNN和Transformer架构的基准模型(RED-Net和EFNet)翻倍,但在PSNR指标上分别提升了4.9%和5.5%,展现出显著的性能优势。此外,与当前最优的Transformer架构方法(DeMo-IVF)相比,DCM-Net的计算量减少超过40%,参数量也显著降低;而与最优的CNN架构方法(UEVD)相比,DCM-Net不仅在PSNR上提升了约3%,其计算量和参数量均低于UEVD。

为进一步验证所提出的双通道跨模态Mamba(DCCM)在计算效率上的优势,本文对DCM-Net进行了架构对比实验:将DCCM模块中的Mamba结构替换为标准的自注意力机制(self-attention),构建Transformer版本的DCM-Net-T。从表1、表2与图7的数据分析得出,尽管DCM-Net-T在去模糊性能(如PSNR、SSIM)上接近原始Mamba架构的DCM-Net,但

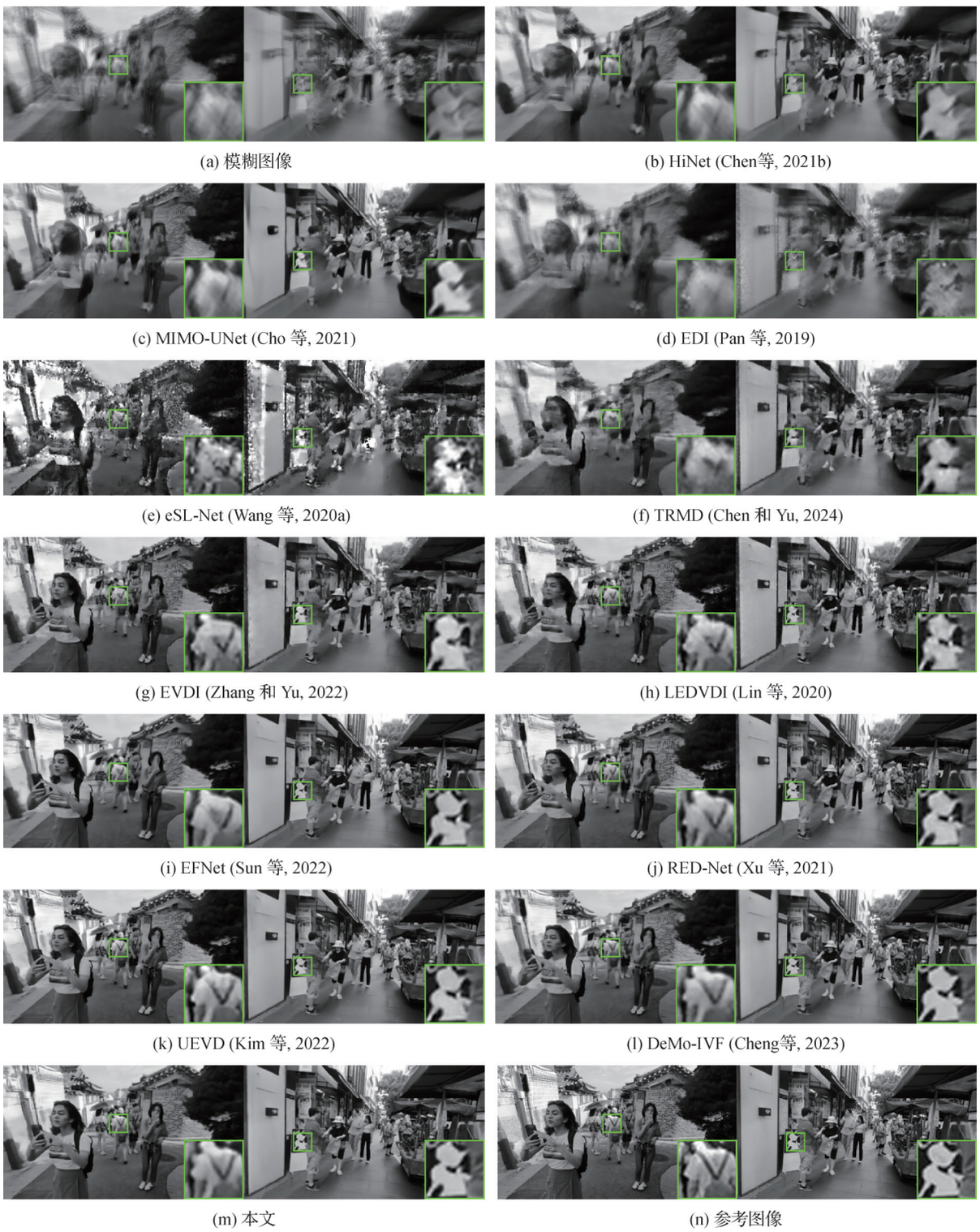


图 5　REDS 数据集中去模糊可视化结果对比

Fig. 5　Comparison of deblurring visualization results on the REDS dataset ((a) blurred images; (b) Chen et al. (2021b); (c) Cho et al. (2021); (d) Pan et al. (2019); (e) Wang et al. (2020a); (f) Chen and Yu(2024); (g) Zhang and Yu(2022); (h) Lin et al. (2020); (i) Sun et al. (2022); (j) Xu et al. (2021); (k) Kim et al. (2022); (l) Cheng et al. (2023); (m) ours; (n) reference images)

其计算代价显著增加。前者比后者在 REDS 数据集上的计算量增加近 37%，在 HQF 数据集上的计算量增加近 50%。这一对比实验充分验证了 Mamba 架构的线性计算复杂度优势。

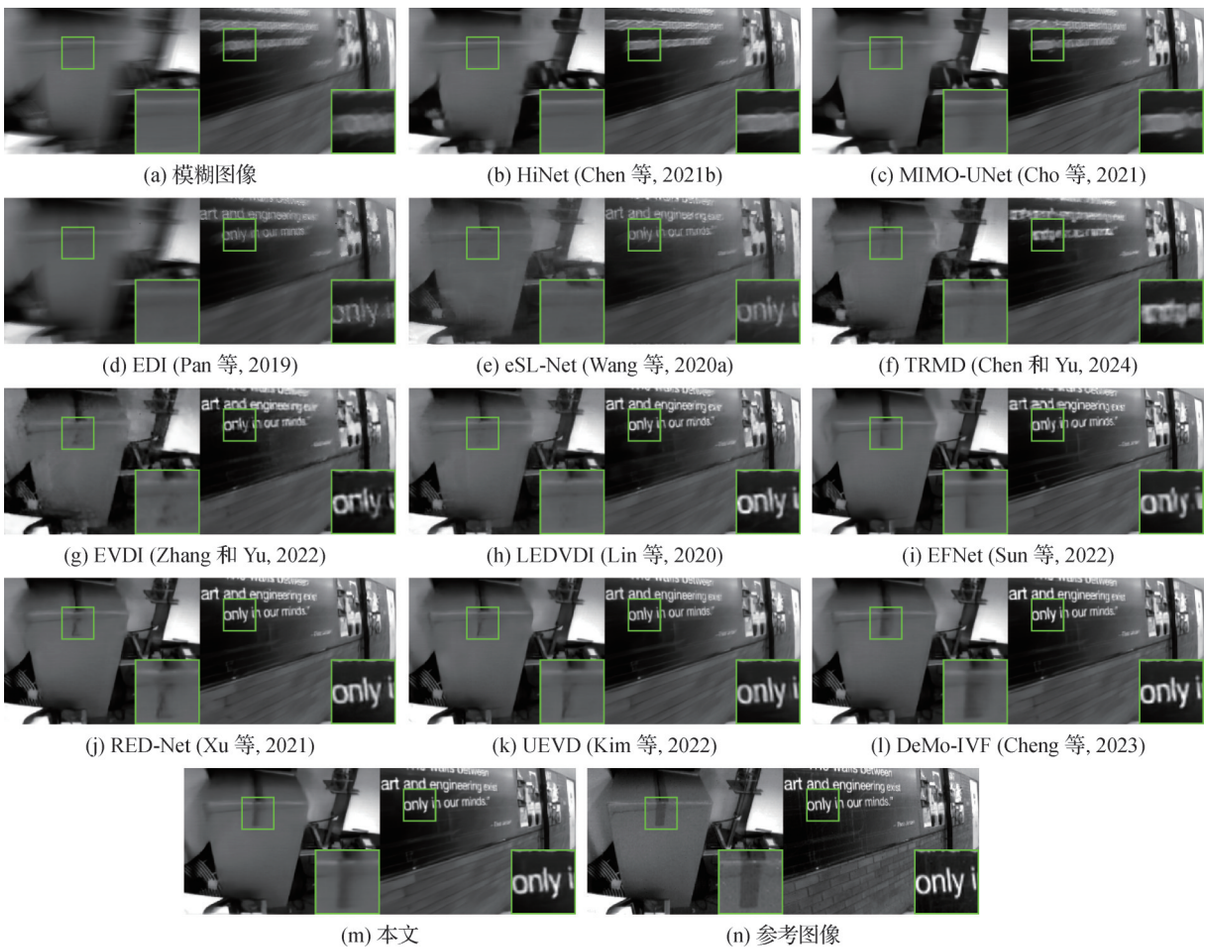


图6　HQF数据集中去模糊可视化结果对比

Fig. 6　Comparison of deblurring visualization results on the HQF dataset ((a) blurred images; (b) Chen et al. (2021b); (c) Cho et al. (2021); (d) Pan et al. (2019); (e) Wang et al. (2020a); (f) Chen and Yu(2024); (g) Zhang and Yu(2022); (h) Lin et al. (2020); (i) Sun et al. (2022); (j) Xu et al. (2021); (k) Kim et al. (2022); (l) Cheng et al. (2023); (m) ours; (n) reference images)

综上,DCM-Net不仅性能优于现有方法,在模型参数量、计算复杂度与性能之间也实现了更优的平衡。

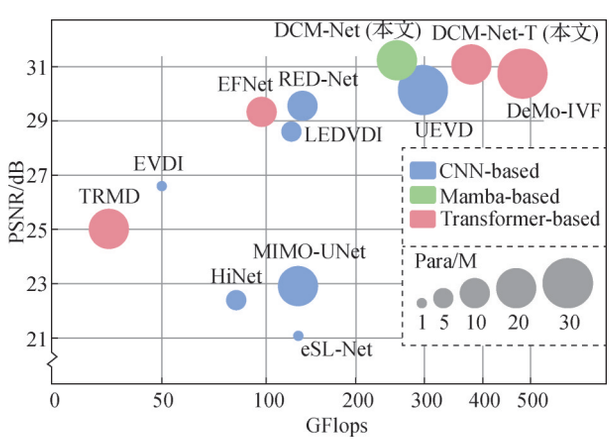


图7　不同方法的参数量、计算量与重建指标结果对比

Fig. 7　Comparison of different methods on numbers of parameters, FLOPs and reconstruction performance

3.3.2　主观对比实验

为全面评估本文提出的DCM-Net在视觉感知质量上的性能,本文选取了其他5种当前先进的去模糊方法进行主观对比实验,包括MIMO-UNet(Cho等,2021)、eSL-Net(Wang等,2020a)、LEDVDI(Lin等,2020)、UEVD(Kim等,2022)和DeMo-IVF(Cheng等,2023)。实验采用双盲测试策略,通过专业调查问卷软件向48名参与者展示不同方法去模糊后的图像样本。参与者根据整体视觉质量对样本进行从6(最好)到1(最差)的评分,最后统计平均分,结果如图8所示。主观评价结果与客观定量指标(如PSNR、SSIM)的评估结果基本一致,进一步验证了DCM-Net在运动图像去模糊任务中的优越性能,特别是在处理复杂运动模糊和细节恢复方面,DCM-Net展现出了更强的视觉感知质量。

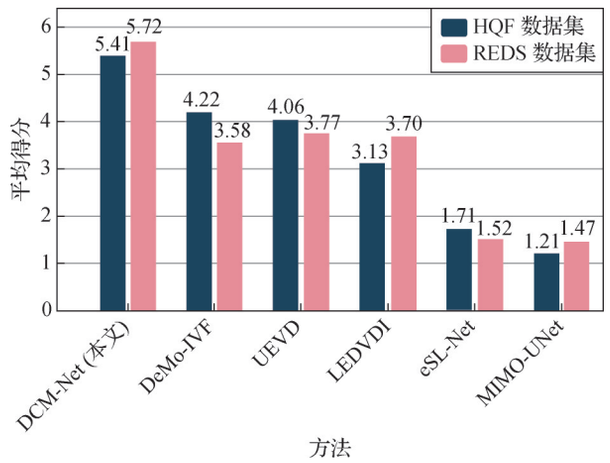


图8　主观对比实验结果

Fig. 8　Results of subjective comparison

3.4　消融实验

为了分析不同模块对DCM-Net提升网络鲁棒性及图像去模糊效果的性能贡献，本文在REDS数据集上进行消融实验，以REDS数据集7帧序列重建结果的指标作为分析依据，结果如表3所示。

首先，在实验Ⅰ中，移除网络中的双通道跨模态Mamba模块（DCCM）与金字塔通道注意力模块（PyCA），同时取消双通道结构（Dual）。此时，网络退化为一个由大量残差密集块堆叠而成的单通道框架，直接输入连接起来的事件张量$\boldsymbol{E}$和图像张量$\boldsymbol{B}$并进行特征提取。实验Ⅱ在实验Ⅰ的基础上引入双通道结构，并使用由MLP简单构成的普通融合结构。从结果可以分析，尽管实验Ⅱ使用双通道的结构设计，但由于图像特征与事件特征之间未能实现有效的相互补偿，图像重建质量显著低于实验Ⅰ。

表3　消融实验结果

Table 3　Results of ablation experiment

实验	Dual	DCCM	PyCA	PSNR/dB	SSIM
Ⅰ	×	×	×	29.068	0.891 5
Ⅱ	√	×	×	26.385	0.832 3
Ⅲ	√	√	×	29.997	0.906 9
Ⅳ	√	√	√	**30.166**	**0.909 6**

注：加粗字体表示各列最优结果。“√”和“×”分别表示使用和未使用对应模块。

实验Ⅲ在实验Ⅱ基础上引入DCCM实现特征相互补偿。该模块学习事件与模糊图像的特征分布，引导图像低噪强度信息嵌入事件中、抑制事件噪声的同时，也利用事件边缘纹理信息增强图像特征，相较于实验Ⅰ，实验Ⅲ在PSNR指标上提升约0.93 dB，在SSIM指标上提升约0.015。此外，实验Ⅳ在实验Ⅲ的基础上引入PyCA以补充特征多尺度的时空注意力，引导网络增强对于关键通道局部模糊的细节重建，在PSNR指标上相比实验Ⅲ进一步提升约0.17 dB，在SSIM指标上提升约0.003。实验结果验证了本文所提出的DCCM和PyCA模块的有效性。

4　结　论

本研究提出一种结合双通道Mamba架构与金字塔通道注意力的事件驱动运动图像去模糊网络DCM-Net，旨在解决传统方法在事件与模糊图像跨模态交互融合、多尺度时空信息关注方面存在的不足，同时降低深度特征计算的复杂度，从而实现从模糊图像中高效恢复高质量的清晰潜在图像序列。DCM-Net使用一种双通道跨模态Mamba模块（DCCM），用于事件和模糊图像的跨模态补偿，将图像绝对强度信息引入事件、利用低噪的图像特征抑制事件中的噪声，同时提取事件的尖锐边缘信息嵌入到图像特征中，达到去模糊的效果。此外，网络嵌入了一种金字塔通道注意力模块（PyCA），提取特征的多尺度时空信息，引导模型提升对关键通道和局部运动模糊区域的关注，进一步提高潜在清晰图像的重建精度。实验结果表明，本文提出的DCM-Net在合成与半合成数据集上去模糊效果显著，重建的潜在图像序列边沿纹理清晰、失真程度低，整体性能优于目前先进的方法。然而，现有实验基于低分辨率合成模糊数据展开，对高分辨率真实模糊场景数据的适应性仍有待验证，这在一定程度上限制了网络模型在实际应用场景中的泛化能力。未来的研究工作将重点从以下两个方面展开：1）进一步优化网络架构，提升模型在真实复杂场景下的性能；2）探索轻量化策略，使其能够适配移动端和边缘计算设备的部署需求。

参考文献（References）

Bigdeli S A, Jin M G, Favaro P and Zwicker M. 2017. Deep mean-shift priors for image restoration//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long

Beach, USA: Curran Associates Inc.: 763-772

Cao C Z, Fu X Y, Zhu Y R, Sun Z J and Zha Z J. 2025. Event-driven video restoration with spiking-convolutional architecture. IEEE Transactions on Neural Networks and Learning Systems, 36(1): 866-880 [DOI: 10.1109/TNNLS.2023.3329741]

Chan T F and Wong C K. 1998. Total variation blind deconvolution. IEEE Transactions on Image Processing, 7(3): 370-375 [DOI: 10.1109/83.661187]

Chen C F R, Fan Q F and Panda R. 2021a. CrossViT: cross-attention multi-scale vision transformer for image classification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 347-356 [DOI: 10.1109/ICCV48922.2021.00041]

Chen H H and Bao Z P. 2017. Strong edge-oriented blind deblurring algorithm. Journal of Image and Graphics, 22(8): 1034-1044 (陈华华, 鲍宗袍. 2017. 强边缘导向的盲去模糊算法. 中国图象图形学报, 22(8): 1034-1044) [DOI: 10.11834/jig.170020]

Chen K and Yu L. 2024. Motion deblur by learning residual from events. IEEE Transactions on Multimedia, 26: 6632-6647 [DOI: 10.1109/TMM.2024.3355630]

Chen L, Fang F M, Wang T T and Zhang G X. 2019. Blind image deblurring with local maximum gradient prior//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1742-1750 [DOI: 10.1109/CVPR.2019.00184]

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]

Cheng Z Y, Zhang X, Yu L, Liu J Z, Yang W and Xia G S. 2023. Recovering continuous scene dynamics from a single blurry image with events [EB/OL]. [2025-03-10]. https://arxiv.org/pdf/2304.02695.pdf

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]

Dong W H, Zhu H D, Lin S H, Luo X Y, Shen Y H, Liu X H, et al. 2024. Fusion-Mamba for cross-modality object detection [EB/OL]. [2025-03-10]. https://arxiv.org/pdf/2404.09146.pdf

Fang X Y, Kan W R, Zhou J, Li L and Guo Y W. 2015. Deblurring of motion blurred images based on the irradiance. Journal of Image and Graphics, 20(3): 395-407 (方贤勇, 阚未然, 周健, 李黎, 郭延文. 2015. 基于辐照度的运动模糊图像去模糊. 中国图象图形学报, 20(3): 395-407) [DOI: 10.11834/jig.20150311]

Fergus R, Singh B, Hertzmann A, Roweis S T and Freeman W T. 2006. Removing camera shake from a single photograph//ACM SIGGRAPH 2006 Papers. Boston, USA: ACM: 787-794 [DOI: 10.1145/1179352.1141956]

Gallego G, Delbrück T, Orchard G, Bartolozzi C, Taba B, Censi A, et al. 2022. Event-based vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1): 154-180 [DOI: 10.1109/TPAMI.2020.3008413]

Gao Z L, Xie J T, Wang Q L and Li P H. 2019. Global second-order pooling convolutional networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3019-3028 [DOI: 10.1109/CVPR.2019.00314]

Gong D, Yang J, Liu L Q, Zhang Y N, Reid I, Shen C H, et al. 2017. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 3806-3815 [DOI: 10.1109/CVPR.2017.405]

Gu A and Dao T. 2024. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2025-03-10]. https://arxiv.org/pdf/2312.00752.pdf

Guo M H, Xu T X, Liu J J, Liu Z N, Jiang P T, Mu T J, et al. 2022. Attention mechanisms in computer vision: a survey. Computational Visual Media, 8(3): 331-368 [DOI: 10.1007/s41095-022-0271-y]

He K M, Zhang X Y, Ren S Q and Sun J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1904-1916 [DOI: 10.1109/TPAMI.2015.2389824]

He X H, Cao K, Yan K Y, Li R, Xie C J, et al. 2024. Pan-Mamba: effective pan-sharpening with state space model [EB/OL]. [2025-03-10]. https://arxiv.org/pdf/2402.12192.pdf

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]

Huang Y N, Li W H, Cui J K and Gong W G. 2021. Strong edge extraction network for non-uniform blind motion image deblurring. Acta Automatica Sinica, 47(11): 2637-2653 (黄彦宁, 李伟红, 崔金凯, 龚卫国. 2021. 强边缘提取网络用于非均匀运动模糊图像盲复原. 自动化学报, 47(11): 2637-2653) [DOI: 10.16383/j.aas.c190654]

Huang Z W, Zhang T Y, Heng W, Shi B X and Zhou S C. 2022. Real-time intermediate flow estimation for video frame interpolation//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 624-642 [DOI: 10.1007/978-3-031-19781-9_36]

Jiang Z, Zhang Y, Zou D Q, Ren J, Lyu J C and Liu Y B. 2020. Learning event-based motion deblurring//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3317-3326 [DOI: 10.1109/CVPR42600.2020.

00338］

Khan S，Naseer M，Hayat M，Zamir S W，Khan F S and Shah M. 2022. Transformers in vision：a survey. ACM Computing Surveys，54(10s)：#200［DOI：10.1145/3505244］

Kim T，Lee J，Wang L and Yoon K J. 2022. Event-guided deblurring of unknown exposure time videos//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv，Israel：Springer：519-538［DOI：10.1007/978-3-031-19797-0_30］

Krishnan D，Tay T and Fergus R. 2011. Blind deconvolution using a normalized sparsity measure//Proceedings of the CVPR 2011. 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］

Lin S N，Zhang J W，Pan J S，Jiang Z，Zou D Q，Wang Y T，et al. 2020. Learning event-driven video deblurring and interpolation//Proceedings of the 16th European Conference on Computer Vision. Glasgow，UK：Springer：695-710［DOI：10.1007/978-3-030-58598-3_41］

Liu H X，Dai Z H，So D R and Le Q V. 2021a. Pay attention to MLPs//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtual Event：Curran Associates Inc.：9204-9215

Liu P D，Janai J，Pollefeys M，Sattler T and Geiger A. 2020. Self-supervised linear motion deblurring. IEEE Robotics and Automation Letters，5（2）：2475-2482［DOI：10.1109/LRA. 2020. 2972873］

Liu Y，Tian Y J，Zhao Y Z，Yu H T，Xie L X，Wang Y W，et al. 2025. VMamba：visual state space model//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver，Canada：Curran Associates Inc.：103031-103063

Liu Z，Lin Y T，Cao Y，Hu H，Wei Y X，Zhang Z，et al. 2021b. Swin transformer：hierarchical vision transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：9992-10002［DOI：10.1109/ICCV48922.2021.00986］

Mehta H，Gupta A，Cutkosky A and Neyshabur B. 2022. Long range language modeling via gated state spaces［EB/OL］.［2025-03-10］. https://arxiv.org/pdf/2206.13947.pdf

Nah S，Baik S，Hong S，Moon G，Son S，Timofte R，et al. 2019. NTIRE 2019 challenge on video deblurring and super-resolution：dataset and study//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach，USA：IEEE：1996-2005［DOI：10.1109/CVPRW.2019.00251］

Nimisha T M，Singh A K and Rajagopalan A N. 2017. Blur-invariant deep learning for blind-deblurring//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice，Italy：IEEE：4762-4770［DOI：10.1109/ICCV.2017.509］

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］

Pan L Y，Scheerlinck C，Yu X，Hartley R，Liu M M and Dai Y C. 2019. Bringing a blurry frame alive at high frame-rate with an event camera//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，USA：IEEE：6813-6822［DOI：10.1109/CVPR.2019.00698］

Qin Z Q，Zhang P Y，Wu F and Li X. 2021. FcaNet：frequency channel attention networks//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：763-772［DOI：10.1109/ICCV48922.2021.00082］

Qiu F，Hou F，Yuan Y and Wang W C. 2019. Blind image deblurring with reinforced use of edges. Journal of Image and Graphics，24(6)：847-858（邱枫，侯飞，袁野，王文成. 2019. 加强边缘感知的盲去模糊算法. 中国图象图形学报，24(6)：847-858）［DOI：10.11834/jig.180543］

Qiu Y F，Liu Z Y and Wang M H. 2024. Image deblurring network combining Mamba and snake-like convolution. Journal of Image and Graphics（邱云飞，刘则延，王茂华. 2024. 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］

Rebecq H，Gehrig D and Scaramuzza D. 2018. ESIM：an open event camera simulator//Proceedings of the 2nd Conference on Robot Learning. Zurich，Switzerland：PMLR：969-982

Rudin L I，Osher S and Fatemi E. 1992. Nonlinear total variation based noise removal algorithms. Physica D：Nonlinear Phenomena，60(1/4)：259-268［DOI：10.1016/0167-2789(92)90242-F］

Scheerlinck C，Barnes N and Mahony R. 2019. Asynchronous spatial image convolutions for event cameras. IEEE Robotics and Automation Letters，4(2)：816-822［DOI：10.1109/LRA.2019.2893427］

Shang W，Ren D W，Zou D Q，Ren J S，Luo P and Zuo W M. 2021. Bringing events into video deblurring with non-consecutively blurry frames//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal，Canada：IEEE：4511-4520［DOI：10.1109/ICCV48922.2021.00449］

Shi Y，Xia B，Jin X Y，Wang X，Zhao T Y，Xia X，et al. 2024. VmambaIR：visual state space model for image restoration［EB/OL］.［2025-03-10］. https://arxiv.org/pdf/2403.11423.pdf

Si Y Z，Xu H Y，Zhu X Z，Zhang W H，Dong Y，Chen Y X，et al. 2025. SCSA：exploring the synergistic effects between spatial and channel attention. Neurocomputing，634：#129866［DOI：10.1016/j.neucom.2025.129866］

Smith J T H，Warrington A and Linderman S W. 2023. Simplified state space layers for sequence modeling［EB/OL］.［2025-03-10］. https://arxiv.org/pdf/2208.04933.pdf

Stoffregen T, Scheerlinck C, Scaramuzza D, Drummond T, Barnes N, Kleeman L, et al. 2020. Reducing the Sim-to-Real gap for event cameras//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 534-549 [DOI: 10.1007/978-3-030-58583-9_32]

Sun L, Sakaridis C, Liang J Y, Jiang Q, Yang K L, Sun P, et al. 2022. Event-based fusion for motion deblurring with cross-modal attention//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 412-428 [DOI: 10.1007/978-3-031-19797-0_24]

Sun Z J, Fu X Y, Huang L Z, Liu A P and Zha Z J. 2025. Motion aware event representation-driven image deblurring//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 418-435 [DOI: 10.1007/978-3-031-72952-2_24]

Tai Y W, Tan P and Brown M S. 2011. Richardson-lucy deblurring for scenes under a projective motion path. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(8): 1603-1618 [DOI: 10.1109/TPAMI.2010.222]

Tan H P, Zeng X J, Niu S J, Chen Q and Sun Q S. 2015. Remote sensing image multi-scale deblurring based on regularization constraint. Journal of Image and Graphics, 20(3): 386-394 (谭海鹏, 曾炫杰, 牛四杰, 陈强, 孙权森. 2015. 基于正则化约束的遥感图像多尺度去模糊. 中国图象图形学报, 20(3): 386-394) [DOI: 10.11834/jig.20150310]

Wang B S, He J W, Yu L, Xia G S and Yang W. 2020a. Event enhanced high-quality image recovery//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 155-171 [DOI: 10.1007/978-3-030-58601-0_10]

Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020b. ECA-Net: efficient channel attention for deep convolutional neural networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]

Wang Z W, Duan P Q, Cossairt O, Katsaggelos A, Huang T J and Shi B X. 2020c. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1606-1616 [DOI: 10.1109/CVPR42600.2020.00168]

Wang Z W, Ng Y, Scheerlinck C and Mahony R. 2021. An asynchronous Kalman filter for hybrid event cameras//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 438-447 [DOI: 10.1109/ICCV48922.2021.00050]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Wu D, Zhao H T and Zheng S B. 2020. Motion deblurring method based on DenseNets. Journal of Image and Graphics, 25(5): 890-899 (吴迪, 赵洪田, 郑世宝. 2020. 密集连接卷积网络图像去模糊. 中国图象图形学报, 25(5): 890-899) [DOI: 10.11834/jig.190400]

Xu F, Yu L, Wang B S, Yang W, Xia G S, Jia X, et al. 2021. Motion deblurring with real events//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2563-2572 [DOI: 10.1109/ICCV48922.2021.00258]

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 L M, Zhang H G, Chen J H and Wang L. 2020a. Hybrid deblur net: deep non-uniform deblurring with event camera. IEEE Access, 8: 148075-148083 [DOI: 10.1109/ACCESS.2020.3015759]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 586-595 [DOI: 10.1109/CVPR.2018.00068]

Zhang X and Yu L. 2022. Unifying motion deblurring and frame interpolation with events//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 17744-17753 [DOI: 10.1109/CVPR52688.2022.01724]

Zhang X, Yu L, Yang W, Liu J Z and Xia G S. 2023. Generalizing event-based motion deblurring in real-world scenarios//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 10700-10710 [DOI: 10.1109/ICCV51070.2023.00985]

Zhang X Q, Wang T, Wang J X, Tang G Y and Zhao L. 2020b. Pyramid channel-based feature attention network for image dehazing. Computer Vision and Image Understanding, 197-198: #103003 [DOI: 10.1016/j.cviu.2020.103003]

作者简介

罗炜麒,男,硕士研究生,主要研究方向为数字图像处理和计算机视觉。E-mail:wikyluo@whu.edu.cn

余磊,通信作者,男,教授,主要研究方向为类脑视觉感知和成像。E-mail:ly.wd@whu.edu.cn

高灿,男,硕士研究生,主要研究方向为数字图像处理和计算机视觉。E-mail:gaocan9920@whu.edu.cn

刘泓驿,男,硕士研究生,主要研究方向为数字图像处理和计算机视觉。E-mail:mcclhy@whu.edu.cn

夏桂松,男,教授,主要研究方向为计算视觉、机器学习和时空智能。E-mail:guisong.xia@ieee.org