中国光学（中英文） Chinese Optics 核心期刊

融合特征增强与轻量级注意力的事件去模糊

顾佳林 吕恒毅 李卓贤 乔善同

Event deblurring via feature enhancement and lightweight attention

GU Jia-lin, LV Heng-yi, LI Zhuo-xian, QIAO Shan-tong

引用本文:
顾佳林, 吕恒毅, 李卓贤, 乔善同. 融合特征增强与轻量级注意力的事件去模糊[J]. 中国光学, 优先发表. doi: 10.37188/CO.2026-0011
GU Jia-lin, LV Heng-yi, LI Zhuo-xian, QIAO Shan-tong. Event deblurring via feature enhancement and lightweight attention[J]. *Chinese Optics*, In press. doi: 10.37188/CO.2026-0011

在线阅读 View online: https://doi.org/10.37188/CO.2026-0011

文章编号　2097-1842(xxxx)x-0001-18

融合特征增强与轻量级注意力的事件去模糊

顾佳林[1,2], 吕恒毅[1*], 李卓贤[1,2], 乔善同[1,2]
（1. 中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033；
2. 中国科学院大学, 北京 100049）

摘要: 针对单帧图像去模糊固有的不适定性, 以及现有扩散模型推理延迟高、状态空间模型跨模态交互不足的问题, 本文提出一种端到端的事件融合多头注意力网络 EFMAN, 利用事件相机的高频时空先验实现高质量复原。首先, 构建跨模态自适应注意力机制, 将异步高频事件流与同步 RGB 特征进行时空维度精确配准, 弥补曝光空缺。接着, 针对传感器固有噪声干扰, 设计特征增强注意力模块 FEA, 通过全局上下文建模强化特征抗噪鲁棒性。然后, 引入轻量级通道-空间注意力模块 LCSA, 在降低计算冗余的同时完成特征响应自适应权重校准。最后, 构建涵盖像素、特征及梯度域的多维联合损失, 协同优化多尺度约束以保证微观纹理与全局结构一致。实验表明, 该方法在保持高效推理的同时显著提升性能。相比基线, 在 GoPro 数据集上 PSNR 和 SSIM 提升 1.19 dB 和 0.005; 在 REBlur 上提升 0.38 dB 和 0.003, 已达先进水平。EFMAN 有效解决了多模态对齐与噪声干扰问题, 在质量与效率间取得平衡, 适用于高动态及剧烈运动场景下的清晰图像重建。

关　键　词: 图像去模糊; 事件相机; 多模态融合; 特征增强; 注意力机制
中图分类号: TP394.1;TH691.9　**文献标志码:** A　**doi:** 10.37188/CO.2026-0011　**CSTR:** 32171.14.CO.2026-0011

Event deblurring via feature enhancement and lightweight attention

GU Jia-lin[1,2], LV Heng-yi[1*], LI Zhuo-xian[1,2], QIAO Shan-tong[1,2]
（1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun* 130033 *China*;
2. *University of Chinese Academy of Sciences, Beijing* 100049 *China*）
* *Corresponding author, E-mail: lvhengyi@ciomp.ac.cn*

Abstract: Single-frame image deblurring remains an inherently ill-posed problem. Furthermore, existing diffusion models suffer from high inference latency, while state space models lack sufficient cross-modal interaction capabilities. To overcome these limitations, we propose an end-to-end Event-fusion Multi-head Attention Network (EFMAN) that exploits high-frequency spatiotemporal priors from event cameras for high-quality image restoration. Specifically, a cross-modal adaptive attention mechanism is designed to precisely align asynchronous high-frequency event streams with synchronous RGB features in both spatial and temporal dimensions, thereby compensating for exposure deficiencies. To mitigate the impact of inherent sensor noise, a Feature Enhancement Attention (FEA) module bolsters feature robustness against noise via global context

收稿日期: xxxx-xx-xx; **修订日期:** xxxx-xx-xx
基金项目: 吉林省科技发展计划(No. 20250201053GX)
Supported by Jilin Province Science and Technology Development Plan (No. 20250201053GX)

modeling. Additionally, a Lightweight Channel-Spatial Attention (LCSA) module is integrated to adaptively recalibrate feature responses while substantially alleviating computational redundancy. These components are optimized by a multidimensional joint loss function—encompassing pixel, feature, and gradient domains—to synergistically enforce multi-scale constraints, ensuring consistency between micro-textures and global topologies. Extensive experiments demonstrate that EFMAN significantly enhances deblurring performance while maintaining efficient inference. Compared to state-of-the-art methods, our approach achieves maximum PSNR and SSIM improvements of 1.19 dB and 0.005 on the GoPro dataset, and 0.38 dB and 0.003 on the REBlur dataset, respectively. By effectively addressing the challenges of multi-modal alignment and noise interference, EFMAN strikes an optimal balance between restoration quality and computational efficiency, making it highly suitable for clear image reconstruction in high-dynamic-range and rapid-motion scenarios.
Key words: image deblurring; event camera; multi-modal fusion; feature enhancement; attention mechanism

1 引 言

运动模糊源于曝光期间相机与场景的相对运动，会导致图像高频信息丢失，至今仍是计算机视觉中一个基础性的不适定难题[1]。早期的卷积神经网络（CNN）通过学习模糊-清晰图像对的映射关系，为此建立了强有力的基线[2]。近年来，该领域经历了以扩散模型和状态空间模型（SSM）为主导的范式转变。基于扩散的方法通过生成式建模捕捉清晰图像的先验分布，取得了卓越的感知质量[3-5]。然而，其反向扩散过程固有的迭代采样机制导致推理延迟极高，难以满足自动驾驶或移动端成像的实时性需求。与此同时，作为 Transformer 的高效替代方案，基于 SSM 的架构（如 Mamba 及其变体）在保持全局感受野的同时实现了线性计算复杂度[6-10]。尽管如此，基于单帧的去模糊方法从根本上受限于物理信息的缺失——从单一积分曝光中恢复连续的时间动态存在理论上的欠约束性，尤其在高动态范围或非线性剧烈运动场景中，极易产生伪影与纹理失真。

为突破传统基于帧成像的物理瓶颈，近年来的研究在多个维度展开了探索。一方面，部分工作致力于通过光子集成计算架构与高维光学特征编码等底层硬件创新来提升信息处理的维度[11-12]；另一方面，在传感器范式的革新上，受生物视觉启发的事件相机（DVS）凭借其微秒级的时间分辨率和超 120 dB 的高动态范围，为解决上述难题提供了新视角[13-15]。此类传感器通过持续编码像素级亮度变化，捕获了丰富的时空先验信息，理论上能够有效填补曝光间隙。鉴于双模态特征融合技术在遥感去云等复杂视觉复原任务中已展现出显著优势[16]，将稀疏的高频事件流与稠密的 RGB 帧相结合的多模态融合方法已成为主流研究方向。针对这一领域，Chen 等人[17] 与 Zhang 等人[18] 探索了不确定性感知与运动插值统一框架，Kim 等人[19] 则提出了事件增强的连续图像恢复方案。近期，Lin 等人[20] 与 Xiao 等人[21] 更是进一步结合了人类视觉特性与状态空间模型，试图挖掘事件流的深层潜力。在时域对齐方面，Tulyakov 等人[22] 提出的 Time Lens 率先验证了事件流在时域对齐上的巨大潜力，随后的研究亦尝试利用 Zamir 等人[23] 改进的 Transformer 或 Yao 等人[24] 设计的脉冲神经网络处理此类异步数据流，试图在时空维度上重建清晰纹理。

然而，在当今基础模型蓬勃发展的背景下，如何有效融合这两种异构模态仍是一个开放性挑战。一方面，异构模态之间的本质差异以及对噪声的敏感性限制了融合的深度。事件数据本质上表征亮度变化的梯度（微分量），而图像则表征绝对强度（积分量）。Sun 等人[25] 的研究指出，这种数据属性上的鸿沟使得直接拼接或简单的交叉注意力机制难以在语义层面实现特征的深度对齐。此外，事件流易受到传感器固有噪声（如热点噪声或背景杂波）的干扰，这种复杂噪声与高频信号的深度耦合在极端视觉恢复任务（如夜间去雾去噪[26]）中同样是极具挑战性的瓶颈。这一点在 Timofte 等人[27] 组织的最新 NTIRE 2025 事件去模糊挑战赛中得到了证实：由于缺乏有效的抗噪机制，多数模型在低光照或极端场景下的鲁棒性均出现了显

著下降。

另一方面, 长序列建模能力与跨模态交互效率之间存在难以调和的瓶颈。尽管 Liu 等人[28]提出的 Video Mamba 在单模态视频理解任务中表现优异, 但其原生的状态空间设计缺乏处理跨模态交互的内在机制, 难以直接适配多模态任务。而在现有的解决方案中, 传统 Transformer 虽然具备强大的全局建模能力, 但在处理高分辨率事件体素时会产生巨大的计算开销; 反之, 当前的轻量级方案往往不得不以牺牲纹理细节为代价换取速度, 难以在计算效率与恢复质量之间取得理想的平衡。

为克服上述局限, 本文提出了事件融合多头注意力网络 EFMAN (Event-Fused Multi-Attention Network)。不同于近期以牺牲推理速度换取感知质量的扩散模型, 也不局限于难以处理多模态交互的纯 SSM 方法, EFMAN 引入了一种新颖的跨模态自适应注意力机制, 能够隐式地将高时间分辨率的事件特征与富含空间语义的 RGB 特征进行精准对齐。具体而言, 我们设计了特征增强注意力模块 FEA (Feature Enhancement Attention), 通过全局上下文建模抑制事件噪声; 并结合 Woo 等人[29]提出的通道-空间解耦思想, 以及视觉任务中高效通道注意力机制的最新演进[30-31], 提出了轻量级通道-空间注意力模块 LCSA (Lightweight Channel-Spatial Attention), 以高效实现特征响应的重校准。此外, 为了进一步提升复原图像的感知质量, 我们设计了包含梯度域和特征域约束的联合损失函数。

本文的主要贡献总结如下:

(1)提出了 EFMAN 融合框架: 这是一个端到端的去模糊架构, 通过创新的注意力机制有效结合了异步事件流与同步模糊帧, 在恢复质量与推理速度之间取得了优异平衡, 性能超越了近期的扩散模型及基于 Mamba 的基线方法。

(2)设计了特征增强注意力模块 FEA: 针对事件相机的噪声特性, 该模块通过鲁棒的全局依赖关系建模, 有效缓解了低光照及复杂场景下的事件噪声干扰, 增强了特征的纯净度。

(3)引入了轻量级通道-空间注意力机制 LCSA: 通过解耦通道与空间维度的注意力计算, 在大幅降低计算冗余的同时实现了特征的自适应重校准, 显著提升了模型的运行效率。

(4)构建了多维联合优化目标: 我们设计了一个涵盖像素级、感知级及高频梯度级的混合损失函数。该策略通过多尺度约束协同优化, 确保模型在保持全局结构一致性的同时, 能够精准恢复微观纹理细节。

(5)实验性能表现: 在多个基准数据集上的大量实验表明, 本方法达到了当前先进水平, 在 PSNR 指标上优于包括 Transformer 与 SSM 变体在内的竞争方法 0.20dB。

2　事件机理与 EFMAN 网络架构

2.1　成像机制与物理去模糊模型

2.1.1　事件相机成像机理

如图 1 所示, 事件相机(DVS)受人眼视网膜神经元结构启发, 其像素电路主要由对数光感受器、差分放大器及阈值比较器三个核心单元构成。与传统有源像素传感器(APS)以固定帧率积分曝光不同, 事件相机采用异步驱动模式, 仅当视场内像素级对数亮度强度发生显著变化时触发信号。这种机制赋予了传感器微秒级的时间分辨率和极高的数据稀疏性。

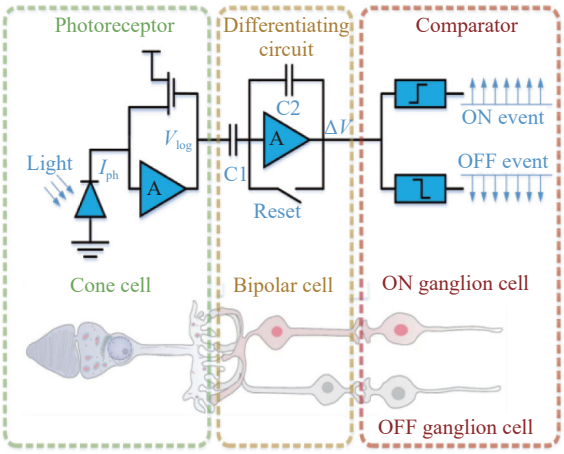


图 1　人眼视网膜三层模型及对应事件相机像素电路

Fig. 1　Three-layer model of the human retina and the corresponding pixel circuitry of an event camera

具体而言, 对于空间坐标 $\mathrm{u} = (x_i, y_i)$ 处的像素, 令 $L(\mathrm{u}, t) = \ln(I(\mathrm{u}, t))$ 表示其在时刻 t 的对数亮度。当当前时刻 t_i 的亮度与上一次触发事件时刻 t_{last} 的亮度偏差超过预设阈值 c 时, 传感器将生成一个离散事件。如图 2 所示, 这一阈值触发机制决定了

事件的极性：当亮度增加超过c时生成正极性事件（ON Event），反之则生成负极性事件（OFF Event）。该过程可形式化定义为：

$$p_i=\begin{cases}+1, & L(u,t_i)-L(u,t_{last})\geqslant c\\ -1, & L(u,t_i)-L(u,t_{last})\leqslant -c\end{cases}\quad , \qquad (1)$$

上式表明，事件流本质上是对场景亮度时空梯度的连续编码。正极性（ON event）对应亮度跃升，负极性（OFF event）对应亮度衰减。

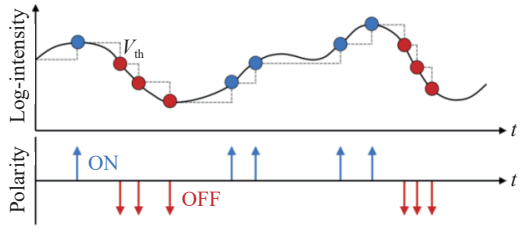


图 2 事件相机生成事件的理论原理

Fig. 2 Theoretical mechanism of event generation in event cameras

2.1.2 基于事件积分的物理约束

运动模糊本质上是相机在曝光时间T内，成像传感器对动态场景辐射强度的积分结果。令$I_{sharp}(x,y,t)$表示t时刻的瞬时清晰潜在图像，则观测到的模糊图像$I_{blur}(x,y)$可建模为：

$$I_{blur}(x,y)=\frac{1}{T}\int_0^T I_{sharp}(x,y,t)dt \quad , \qquad (2)$$

这是一个典型的不适定问题。然而，事件相机记录的高频亮度变化为重建曝光过程中的中间状态提供了关键约束。

遵循 Pan 等人提出的基于事件的双重积分模型 EDI[32]，利用事件流与瞬时亮度间的物理映射约束求解。根据 EDI 理论，对于视场内的任意像素，其在时刻$t\in[0,T]$的瞬时亮度$I(t)$与曝光起始时刻的潜在清晰图像$I(0)$满足如下微分关系：

$$\ln I(t)-\ln I(0)\approx c\cdot E(t) \quad , \qquad (3)$$

其中$E(t)=\int_0^t p(\tau)d\tau$为该像素在曝光时段内的累积事件极性积分。通过指数变换，可将t时刻的潜在状态建模为初始状态与事件累积量的函数：

$$I(t)=I(0)\cdot\exp(c\cdot E(t)) \quad , \qquad (4)$$

将公式 (4) 代入模糊生成模型 (2)，鉴于$I(0)$独立于积分时间t，可得：

$$I_{blur}=\frac{I(0)}{T}\int_0^T\exp(c\cdot E(t))dt \quad , \qquad (5)$$

公式 (5) 揭示了事件去模糊的物理本质：模糊图像I_{blur}实际上是初始清晰图像$I_{sharp}(0)$经由事件流调制的加权积分。理论上，只要能够准确估计事件流的积分项，即可从模糊图像中反解出清晰图像$I_{sharp}(0)$。

然而，在实际应用中，由于传感器噪声、阈值c的非恒定性以及光照变化的复杂性，直接利用公式 (5) 进行逆求解极其困难。鉴于此，本文提出了一种数据驱动的深度学习框架，旨在隐式地学习这一物理映射，通过融合 RGB 的积分特征与 Event 的微分特征，实现对I_{sharp}的鲁棒估计。

2.2 EFMAN 网络整体架构设计

2.2.1 总体流程与多阶段策略

基于前述物理去模糊模型可知，当事件在曝光时间维度上累积时，将模糊观测图像与曝光期间触发的事件流相结合，即可获取反解潜在锐利图像所需的全部信息。受此启发，我们提出的深度学习框架 EFMAN 摒弃了常规的通用“黑盒”映射模式，转而以 EDI 物理模型为引导进行网络构建。

具体而言，EFMAN 中的事件先验提取分支所处理的体素网格在通道维度上天然对应着曝光时间步t。其内部的多层连续卷积操作，隐式地完成了 EDI 模型中对事件极性的非线性指数积分计算（即计算$\int_0^T\exp(c\cdot E(t))dt$。而在多模态交互阶段，LCSA 与 SAM 自监督模块利用事件特征生成跨模态注意力权重，并对 RGB 模糊特征进行自适应的点乘调制。在数学与物理意义上，这一过程恰好等价于执行了物理模型中基于事件积分先验的逆滤波（即除法）操作。

通过上述底层物理机理的驱动，EFMAN 成功构建了从模糊图像 B 和相关事件流 E 到清晰图像的非线性映射关系。为实现这一高质量复原，整个映射过程被精细地建模为两个级联的阶段，其形式化表述如下：

$$F=F_{\theta_1}(B,E) \quad , \qquad (6)$$

$$L=F_{\theta_2}(F,B) \quad , \qquad (7)$$

其中，F表示中间特征表征，θ_1与θ_2分别对应第一阶段与第二阶段的映射参数。延续基准模型

EFNet 的骨干设计, EFMAN 采用两阶段级联架构以逐步恢复清晰图像, 第一阶段与第二阶段分别对应公式(6)和(7)所描述的映射过程。EFMAN 的详细架构如图 3 所示。

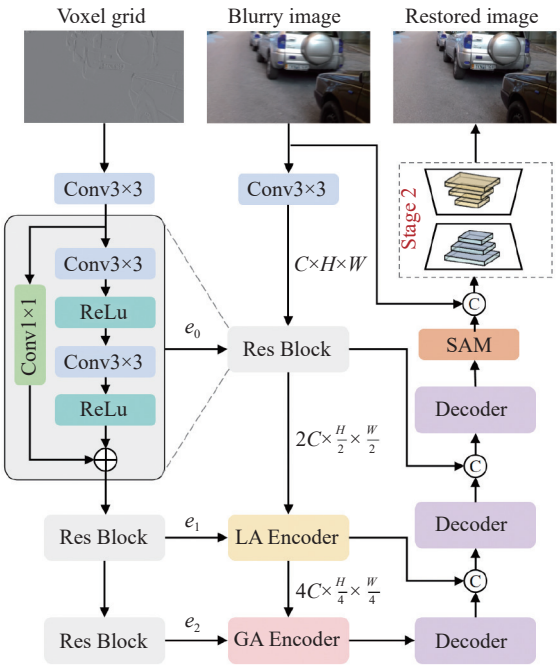


图 3　EFMAN 的整体架构

Fig. 3　Overall architecture of the proposed EFMAN

本框架包含两个并行的分支: 事件先验提取分支与分层图像重建分支。事件先验提取分支负责将体素化网格表示的事件流映射为多尺度运动特征金字塔 $\mathcal{M}=\{e_0,e_1,e_2\}$, 从而为图像分支提供鲁棒的边缘与轨迹引导信息。图像重建分支则采用两阶段级联架构, 其中第一阶段利用事件特征快速提取并恢复图像的主体结构, 第二阶段专注于残差细化。此外, 我们在两个阶段之间引入了监督注意力模块 SAM。作为语义桥接组件, SAM 利用中间监督信号显式地对第一阶段特征进行重校准, 在进入第二阶段前有效抑制了潜在伪影并增强显著特征, 从而实现了“由粗到细”的渐进式复原。这种深度的跨模态协同机制与级联优化策略, 不仅最大化了事件流的高频时空优势, 还确保了在剧烈运动场景下边缘信息的完整性, 极大提升了模型在复杂工况下的鲁棒性。

2.2.2　事件先验提取分支

尽管事件流具备极高的时间分辨率, 但其空间稀疏性给特征提取带来了挑战。为了提取时空连续且具有判别力的运动表征, 我们构建了一个深度残差编码器。其基础构建单元采用改进的残差卷积块 RCB。为了解决深层网络中的梯度流问题并增强特征的非线性表达能力, RCB 被设计为包含两条路径: 主路径由两个级联的3×3卷积层与 Leaky ReLU 激活函数构成, 负责捕获局部时空上下文; 而残差连接路径则引入1×1卷积进行特征投影与通道变换。这种设计不仅实现了特征维度的自适应对齐, 还起到了特征重筛选的作用。

2.2.3　图像多尺度重建分支

图像重建分支采用 U-Net 风格的拓扑结构。考虑到运动模糊通常是由局部物体运动和全局相机抖动共同引起的复杂退化, 单一的卷积或注意力机制难以同时应对。因此, 我们提出了一种基于物理特性的混合编码策略, 在不同层级针对性地部署计算模块。

在第一层编码器 (H×W) 上, 主要任务是重建图像的边缘与纹理细节, 鉴于卷积算子在提取高频特征方面的优势, 我们复用了上述的深度残差编码器。在第二层编码器 $(\frac{H}{2}\times\frac{W}{2})$ 上, 我们引入了局部注意力编码器, 即本文提出的 FEA 模块。该模块将自注意力计算限制在局部窗口内, 有效地聚合了局部运动上下文, 并利用中层事件特征 e_1 引导局部特征的去模糊, 同时保持了较低的计算复杂度。在第三层编码器 $(\frac{H}{4}\times\frac{W}{4})$ 上, 我们引入了全局注意力编码器, 即 LCSA 模块, 利用全局感受野将高层语义事件特征 e_2 深度嵌入图像特征中, 从而实现对全局运动轨迹的建模与校正。这种差异化的设计确保了网络在有效处理多尺度模糊的同时, 实现了感受野与计算效率的最佳平衡。

2.3　特征增强注意力模块构建

针对当前特征提取过程中存在的特征局部化及增强效果欠佳的问题, 我们提出了一种基于结构化全局上下文建模的特征增强注意力模块 FEA。该模块以自注意力为核心机制, 通过对输入特征进行全局相关性建模, 显著提升模型对长程依赖的捕获能力。

如图 4 所示, 该模块旨在通过多尺度特征交互与注意力机制来增强特征表示能力。整体架构包含两个主要部分: 主干特征提取路径与嵌入其中的特征增强注意力模块 FEA Block。

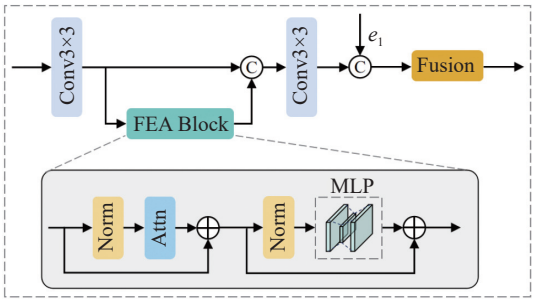


图 4 LA 编码器的详细结构(基于 FEA 模块)

Fig. 4 Detailed structure of the LA Encoder (based on the FEA module)

2.3.1 全局特征交互流程

输入特征首先经过一个3×3卷积层(Conv3×3)进行浅层特征提取。随后，特征流被分流为两路：一路作为恒等映射保留原始局部特征，另一路则输入至 FEA Block 以捕获长距离依赖或进行全局特征精炼。这两路特征在通道维度上进行拼接，实现了原始特征与增强特征的融合。融合后的特征再次经过一个3×3卷积层进行降维或特征平滑，随后与经过事件先验提取分支处理得到的先验信息e_1进行第二次拼接。最后，通过借鉴 EFNet 中的 Fusion 模块，将多源信息整合并输出最终特征。假设输入特征为X_{in}，第一层卷积操作为f_{conv1}，FEA 模块为$\mathcal{F}_{FEA}$，整体流程可表示为：

$$F_{inter} = \mathrm{Concat}(f_{conv1}(X_{in}), \mathcal{F}_{FEA}(f_{conv1}(X_{in}))) \quad ,（8）$$

$$X_{out} = \mathcal{F}_{fusion}(\mathrm{Concat}(f_{conv2}(F_{inter}), e_1)) \quad ,（9）$$

其中，Concat(·)表示沿通道维度的拼接操作，f_{conv2}为第二层卷积，$\mathcal{F}_{fusion}$代表最终的融合操作。

此外，由于事件相机在高速运动下易产生高频的“热噪点”与背景杂波。传统的自注意力机制由于倾向于增强显著性高频响应，在缺乏约束时往往会错误地放大这些孤立噪点。为了克服这一固有缺陷，本文设计的 FEA 模块并未直接作用于单一的事件流，而是作用于经过初步跨模态融合的混合特征。在这一计算机制下，RGB 图像保留的浅层低频物理结构充当了天然的“结构锚点”。在计算自注意力矩阵时，网络被强制依赖于 RGB 图像中真实存在的物理几何连通性，从而在全局感受野内自适应地过滤并孤立了与图像语义不相关的事件热噪点。这一跨模态结构锚点机制，从物理特征层面上有效强化了模型的抗噪鲁棒性。

2.3.2 上下文聚合内部结构

FEA Block 采用了经典的 Pre-Norm 残差结构和前馈神经网络 MLP 级联组成。输入特征z在内部结构中依次通过两个残差子层。第一个子层利用层归一化 Norm 和注意力机制 Attn 来聚合上下文信息；第二个子层则通过归一化和 MLP 进行特征变换。每个子层后均包含一个逐元素相加操作。具体计算过程如下：

$$z' = z + \mathrm{Attn}(\mathrm{Norm}(z)) \quad ,（10）$$

$$z_{out} = z' + \mathrm{MLP}(\mathrm{Norm}(z')) \quad ,（11）$$

其中，Norm(·)代表 Layer Normalization，Attn(·)为自注意力机制。

2.4 轻量级空间通道注意力设计

针对当前特征提取过程中存在的特征不足且推理速度欠佳的问题，我们提出了一种轻量化空间通道注意模块 LCSA。该模块以我们设计的 MFFM (Multi-branch Feature Fusion Module) 多分支特征融合模块和 PFR (Planar Feature Refinement) 平量特征细化模块为核心机制，通过对输入特征进行全局相关性建模，显著提升模型对长程依赖的捕获能力。

如图 5 所示，该模块旨在利用高效的注意力机制对特征进行自适应增强与多源融合。给定输入特征F_{in}，首先通过一个3×3卷积层进行初步特征提取。为了增强特征的判别力，特征流被分为两路：一路作为恒等映射保留原始信息，另一路送入 LCSA 模块以捕获通道与空间维度的上下文依赖。增强后的特征与原始特征在通道维度进行拼接，随后经过第二个3×3卷积层进行特征平滑与降维。最后，模块引入事件先验提取分支处理得到的先验信息e_2，再次进行拼接并通过融合层生成最终输出F_{out}。在此，我们将核大小为k的卷积操作记为$\mathcal{F}_k(\cdot)$，沿通道维度的拼接操作记为$[\cdot]$，则该过程表示为：

$$F_{enh} = \Phi_{\mathrm{LCSA}}(\mathcal{F}_3(F_{in})) \quad ,（12）$$

$$F_{inter} = \mathcal{F}_3([\mathcal{F}_3(F_{in}), F_{enh}]) \quad ,（13）$$

$$F_{out} = \mathcal{F}_{\mathrm{fusion}}([F_{inter}, e_2]) \quad ,（14）$$

其中Φ_{LCSA}表示 LCSA 模块的映射函数。

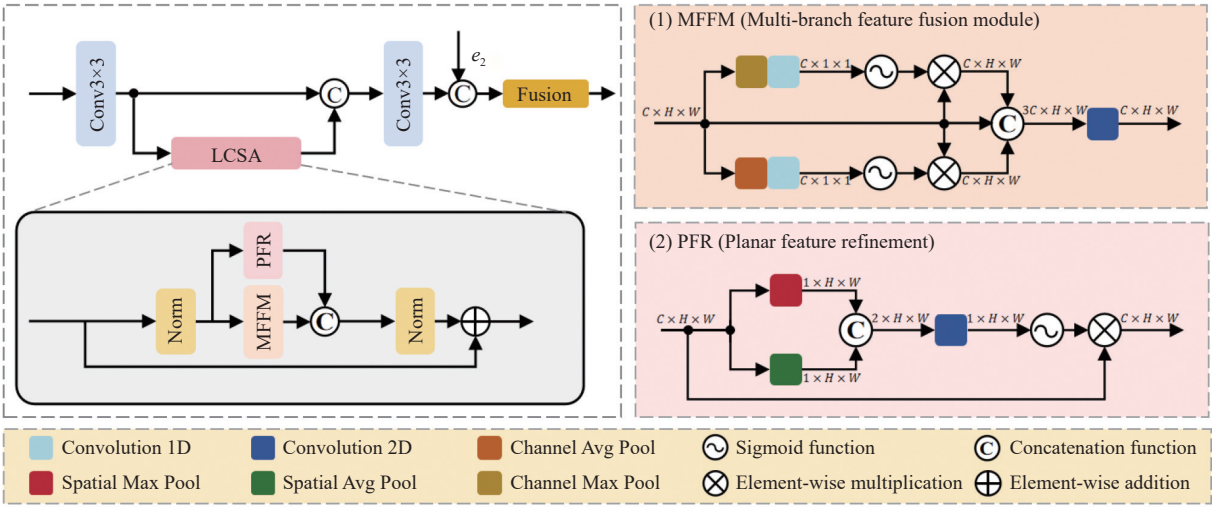


图 5　GA 编码器中 LCAS 模块的架构示意图

Fig. 5　Schematic architecture of the LCSA module within the GA Encoder

LCSA 内部采用并行结构以同时捕捉跨通道交互与空间区域显著性。输入特征X首先经过层归一化得到$\tilde{X}$，随后并行进入我们所设计的 MFFM 和 PFR 分支。在我们所设计的 MFFM 多分支特征融合模块分支中，我们同时采用全局平均池化$\mathcal{P}_{avg}$和全局最大池化$\mathcal{P}_{max}$将输入特征压缩，随后利用一维卷积$\mathcal{F}_{1d}$捕捉跨通道的局部依赖关系，并经过 Sigmoid 激活函数σ生成权重w。该分支输出X_{ca}定义为：

$$w_{\phi}=\sigma\left(\mathcal{F}_{1d}(\mathcal{P}_{\phi}(\tilde{X}))\right),\ \phi\in\{avg,max\}\quad,\qquad(15)$$

$$X_{ca}=\mathcal{F}_{1}\left(\left[\tilde{X},\tilde{X}\otimes w_{avg},\tilde{X}\otimes w_{max}\right]\right)\quad,\qquad(16)$$

其中⊗表示逐元素相乘。考虑到事件流数据的特殊性，LCSA 改进了传统 CBAM 处理 RGB 图像时基于全连接层(MLP)的通道注意力机制，转而采用一维卷积(1D Conv)。在事件体素张量中，通道维度实际对应于离散的时间步(即 Channel=Time)。由于传统 MLP 会破坏特征的时间局部连续性，LCSA 利用 1D 卷积显式地对时间维度进行时序相关性建模。这一设计不仅契合事件流的物理本质，更能精准捕获极速运动物体在微秒级曝光时刻下的连续运动轨迹。PFR 分支则主要关注特征在空间维度的分布。首先沿通道轴分别进行最大池化和平均池化(记为$\mathcal{P}^{c}$)，生成两个空间特征图。将两者拼接后，利用一个二维7×7卷积层生成空间注意力掩膜M_s，经激活后与输入特征相乘得到X_{sa}：

$$M_s=\sigma\left(\mathcal{F}_{7}\left(\left[\mathcal{P}^{c}_{avg}(\tilde{X}),\mathcal{P}^{c}_{max}(\tilde{X})\right]\right)\right)\quad,\qquad(17)$$

$$X_{sa}=\tilde{X}\otimes M_s\quad,\qquad(18)$$

最后，两个分支的输出特征被拼接后，再次经过归一化处理，并与原始输入通过残差连接相加，形成最终输出Y：

$$Y=X+\mathcal{N}([X_{sa},X_{ca}])\quad,\qquad(19)$$

在此结构中，池化操作被用于高效地聚合全局特征描述，在大幅降低计算复杂度的同时保留了关键的通道与空间语义信息。结合并行的注意力机制，LCSA 模块能够有效抑制冗余噪声并增强显著性特征。通过这种轻量化的设计，该模块在极低的计算开销下实现了对局部细节与长程依赖的精准捕捉，为系统提供了一种高效且鲁棒的特征细化方案。

2.5　多维联合优化目标函数

为了从退化的模糊输入中恢复出兼具高保真度与视觉锐度的清晰图像，我们构建了一个多维联合损失函数。该函数在像素域、特征域及梯度域同时施加约束，旨在确保模型在保持全局结构一致性的同时，能够精准重建微观纹理细节。

我们的多维联合损失函数包含四项单独损失，但级联这四项损失并非简单的经验性叠加，而是针对事件相机数据“高频边缘极度敏锐，但伴随稀疏热噪声”的独特物理性质所进行的针对性设计。在图像复原领域(如 DeblunGAN-v2[33] 与 MPRNet[34])，联合使用多尺度约束以解耦优化频

率分量已成为常见范式。在本研究中，这四项损失在频域上实现了正交与解耦：$\mathcal{L}_{pix}$与$\mathcal{L}_{per}$构成了低中频的结构与色彩基底；而在极高频段，由于事件流的引入会同时增强边缘与热噪声，必须通过$\mathcal{L}_{hf}$（强化高频突变）与$\mathcal{L}_{tv}$（惩罚孤立噪声尖峰）形成闭环对抗。移除任何一项，都会为模型带来边缘模糊或者噪声的放大干扰。

令I_B为输入的模糊图像，$\hat{I}=G(I_B)$表示生成网络输出的复原图像，I_{GT}为对应的地面真值。总优化目标$\mathcal{L}_{total}$定义为以下四项的加权线性组合：

$$\mathcal{L}_{total}=\lambda_{pix}\mathcal{L}_{pix}+\lambda_{per}\mathcal{L}_{per}+\lambda_{hf}\mathcal{L}_{hf}+\lambda_{tv}\mathcal{L}_{tv}, \quad (20)$$

其中，$\lambda_{pix},\lambda_{per},\lambda_{hf},\lambda_{tv}$为平衡各约束项权重的超参数。

2.5.1 像素级重建损失

为了约束预测图像$\hat{I}$与真实图像I_{GT}在色彩与亮度上的低频一致性，我们采用ℓ_1范数作为基础重建项。相较于ℓ_2损失，ℓ_1范数对离群值具有更强的鲁棒性，且能有效抑制由均方误差诱导的纹理过度平滑问题：

$$\mathcal{L}_{pix}=\frac{1}{CHW}\sum_{c,h,w}|\hat{I}_{c,h,w}-I_{GT,c,h,w}| \quad , \quad (21)$$

2.5.2 深度感知损失

考虑到传统的像素级距离（如ℓ_1或ℓ_2）往往导致生成图像纹理过度平滑，且无法充分反映人眼对图像语义结构的敏感性，我们引入感知损失以提升图像的视觉真实感。我们采用在 ImageNet 数据集上预训练的 VGG-19 网络Φ作为固定的特征提取器。

具体而言，我们计算生成图像$\hat{I}$与真实图像I_{GT}在 VGG-19 网络第5个最大池化层之前的最后一个卷积层（记为 conv5_4）特征映射之间的欧氏距离。该层特征富含高层语义及结构信息，能够强制模型学习抽象的特征一致性：

$$\mathcal{L}_{per}=\frac{1}{C_jH_jW_j}\|\Phi_{5,4}(\hat{I})-\Phi_{5,4}(I_{GT})\|_2^2 \quad , \quad (22)$$

其中$\Phi_{5,4}$表示 VGG-19 网络中 conv5_4 层的激活输出，该网络的参数在训练阶段保持冻结。通过最小化该损失，模型被迫关注图像的结构合理性而非单纯的像素值匹配，从而有效恢复出更符合人类感知的锐利边缘与纹理。

2.5.3 高频梯度一致性损失

由于模糊过程本质上是对高频信息的滤除，为了显式恢复边缘锐度与微观纹理，我们设计了基于拉普拉斯算子的梯度域约束。该损失项通过惩罚二阶梯度的差异，强制模型关注图像的高频突变区域：

$$\mathcal{L}_{hf}=\frac{1}{CHW}\sum_{c,h,w}\|\Delta(\hat{I})_{c,h,w}-\Delta(I_{GT})_{c,h,w}\|_1 \quad , \quad (23)$$

其中$\Delta(\cdot)$表示拉普拉斯滤波操作（与固定核K_Δ的卷积）。

2.5.4 全变分正则化

为了平衡高频细节增强与伪影抑制，避免在去模糊过程中引入非自然的噪声，我们引入全变分损失作为平滑先验。该项约束了相邻像素间的过度跳变，从而保证了平坦区域的纯净度：

$$\mathcal{L}_{tv}=\sum_{h,w}\sqrt{(\hat{I}_{h+1,w}-\hat{I}_{h,w})^2+(\hat{I}_{h,w+1}-\hat{I}_{h,w})^2} \quad , \quad (24)$$

在本文的多模态联合优化框架中，全变分正则化扮演着关键的“噪声抑制器”角色。由于 FEA 等注意力模块在聚合全局事件特征时存在放大局部热噪声的风险，$\mathcal{L}_{tv}$能够对孤立的高频尖峰（异常噪点）施加梯度惩罚。这种频域惩罚机制迫使注意力模块学习出更平滑的权重掩膜，在确保边缘锐化的同时，有效抑制了背景杂波的放大现象。像素级重建、深度感知、高频梯度与全变分正则化协同作用，共同解决了去模糊任务中“细节恢复”与“噪声抑制”的平衡难题。

3 实验结果与分析

为了系统地验证所提方法的有效性，我们在两个行业主流的基准数据集上进行了广泛的实验。本章首先介绍了具体的实验设置，包括训练细节与超参数配置；随后明确了所采用的图像复原质量评价指标，并将本方法与当前最先进的算法进行了对比。实验结果涵盖了定量的客观指标分析与定性的视觉效果展示，从多个维度充分证明了本方法的先进性能。

3.1 数据集介绍

我们采用当前最主流的 GoPro[2] 和 REBlur[25] 数据集作为我们的对比基准数据集，同时对

GoPro 和 REBlur 数据集的微调遵循 Sun 等人所提出的 EFNet 的设置, 为了方便并公平的比较, 在后续试验中对所有算法进行同样的设置。

GoPro 数据集: 最初由 Seungjun Nah 等人发布, 是用 GoPro Hero 4 Black 采集 240 fps 的超高帧率视频。后续 Sun 等人及相关研究者通过 ESIM 模拟器重构了新的 GoPro 数据集, 因为模拟器重构的数据集可以得到高帧率、高质量的图像, 符合我们对地面训练真实(GT)图像的要求, 因此我们应用此数据集进行我们的算法验证。该数据集包含 3 214 对模糊图像和清晰图像, 我们严格遵循 EFNet 预设划分协议, 使用了 2 103 张图片进行训练, 1 111 张图片进行测试, 该数据集包含室外车辆, 行进中行人, 路边玻璃标识, 餐厅广告牌等多方面场景, 对于算法验证具有极强的价值。

REBlur 数据集: 由 Sun 等人专门为事件驱动的去模糊任务设计。在高度受控的光学实验环境中, 通过搭建工业相机+事件相机+电动平移台的光学系统, 同步采集真实场景下的事件流、高分辨率清晰图像及相应运动模糊图像所形成的真实世界数据集。涵盖 12 类典型的线性和非线性运动模式, 系统性地模拟了多样化运动退化过程。该数据集共包含 36 个独立序列, 总计 1,469 组严格对齐的模糊-清晰图像对及其对应的事件流数据。我们严格遵循 EFNet 预设划分协议, 其中 486 组用于训练, 983 组用于测试, 该数据集补全了 GoPro 数据集中缺乏的室内运动模糊的场景, 通过集成高精度的光学系统, 该平台能够在高速运动条件下生成具有精确像素级对齐的模糊图像, 并借助时间戳同步的事件流提供可靠的物理对应关系, 从而为基于事件的运动去模糊算法提供具备严格真实标签的基准数据支持。

3.2　实验环境设置

本实验在一台配备 8 张 NVIDIA GeForce RTX 3 090 显卡的服务器上进行, 该显卡搭载了 24 GB 显存, 足以支持大规模深度学习模型的训练与推理。系统运行于 Ubuntu 22.04 操作系统, 为实验提供了稳定且高效的 Linux 环境。在软件环境方面, 实验采用 PyTorch 2.1.1 作为核心深度学习框架, 并基于 Python 3.11.5 进行算法实现与流程控制。为了充分发挥 GPU 的计算性能, 我们配置了 CUDA 11.5 并行计算平台, 并配合 cuDNN 8.2.0 深度神经网络加速库, 以保障模型训练过程中张量运算的高效执行。本实验的实验环境配置如表 1 所示。

表 1　实验环境配置

Tab. 1　Configuration of the experimental environment

参数	配置
系统环境	Ubuntu 22.04
显卡	NVIDIA GeForce RTX 3 090
深度学习框架	PyTorch 2.1.1
加速环境	CUDA 11.5 和 CuDNN 8.2.0
语言	Python 3.11.5

为全面评估所提方法的性能并进行公平比较, 本研究采用统一的训练与测试配置。所有实验均在分辨率为 256×256 像素的图像上进行。训练过程中, 采用批梯度下降法, 批次大小设置为 4。此设置是在充分考虑模型复杂度与 GPU 显存限制后确定的平衡选择, 以确保训练过程的稳定性。数据加载器配置了 8 个工作线程, 以优化数据预处理与 I/O 效率, 缓解训练瓶颈。模型训练总计进行 200,000 次迭代。整个训练周期耗时约 41 小时, 体现了所提模型在给定硬件条件下的实际收敛效率。我们采用 AdamW 优化器, 其初始学习率设置为 1×10^{-4}, 并配合余弦退火调度策略进行动态调整。权重衰减系数设定为 5×10^{-4}, 以有效正则化模型复杂度, 防止过拟合。

为确保实验结果严格可复现, 所有实验均在固定随机种子下进行。为避免全空间高维盲目网格搜索带来的高昂成本, 本文采用“物理启发式定界与局部精细搜索”策略设定联合损失函数的超参数。首先, 进行定性物理分析与量级边界粗筛: 基于各损失项频段解耦与梯度量级的差异, 我们确立了降阶调节的搜索区间。以保障全局低频信息的像素重建损失为主导(锚定权重为 1.0); 负责中频语义的感知损失降阶至 10^{-1} 量级(锁定区间 [0.05, 0.3])以防止色彩偏移; 而作用于极高频的高频梯度与 TV 正则化项, 因其对梯度响应极度敏感, 被严格框定在 10^{-2} 至 10^{-3} 的极小量级区间内。随后, 开展局部精细搜索: 在划定上述具备物理意义的狭小搜索区间后, 我们设定 0.005 的搜索步长进行遍历。结果显示: 当感知损失权重

在 0.15 附近时色彩与语义保真度最佳；当高频梯度权重提升至 0.045 时边缘锐化收益达到峰值，超逾此阈值则诱发振铃伪影；而 TV正则化权重设置为 0.008 时，恰好能在较好抹平事件热噪点的同时保留微观纹理。最终，我们锁定了客观指标与主观视觉达到最优平衡的精确参数组合：$\lambda_{pix}=1.0,\lambda_{per}=0.15,\lambda_{hf}=0.045,\lambda_{tv}=0.008$。这一策略既保证了参数设定的物理合理性，又极大地降低了算力开销与调试成本。

此外，本研究融入了自动混合精度（AMP）训练策略，在保持数值精度的同时，显著提升了计算效率与训练吞吐量，具体参数设置如表 2 所示。

表 2　实验参数配置

Tab. 2　Configuration of the experimental parameters

参数	数值
图片大小 (Image size)	256×256
训练迭代次数 (Iterations)	200,000
每次更新的批次数 (Batch size)	4
线程数 (Workers)	8
时间	41h

3.3　评价指标

为客观评价算法模型的效果，本文采用两个行业内公认指标，结构相似性 SSIM、峰值信噪比 PSNR 作为标准来评估模型的性能表现。

3.3.1　峰值信噪比 PSNR

PSNR 是衡量图像失真程度最常用的定量指标，它基于误差敏感度，通过计算对应像素点之间的差异来评估质量。其公式推导如下所示：

在计算 PSNR 之前，必须先定义均方误差 MSE (Mean Squared Error)。假设参考图像为 I，待评价的复原图像为 K，图像尺寸为 $m\times n$：

均方误差 MSE：

$$MSE=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2 \quad , \qquad (25)$$

H 和 W 分别表示图像的高度与宽度；$\mathbf{I}$ 代表原始清晰地面真值图像（Ground Truth）；$\hat{\mathbf{I}}$ 代表经由网络模型复原的去模糊图像；(i,j) 表示像素的空间坐标索引。MSE 越小，表示复原图像与真值越接近。

PSNR 通过 MSE 的对数形式表示，单位为分贝（dB），计算公式为：

$$PSNR=10\cdot\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad , \qquad (26)$$

其中，MAX_I 是图像像素的最大可能值。对于常见的 8-bit 图像，$MAX_I=255$，PSNR 虽然能直观的表示出图片质量的差异来迅速分析出去模糊后的图像质量提升但也有其局限性，PSNR 仅关注像素值的数值差异，并不考虑人类视觉系统（HVS）的特性。可能出现 PSNR 高，但视觉上存在明显的伪影或过度平滑。

3.3.2　结构相似性 SSIM

结构相似性 SSIM（Structural Similarity Index）并非指传统物理学中的力、热、光、电，而是指其模仿人类视觉系统（HVS）对物理世界中物体结构属性的感知机理。因此，它不再局限于像素点的逐点对比，而是从亮度、对比度、结构三个维度进行综合建模。

SSIM 从亮度、对比度、结构三个维度进行综合建模：

$$SSIM(x,y)=[l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma} \qquad (27)$$

其中 $l(x,y)$ 为亮度比较，$c(x,y)$ 为对比度比较，$s(x,y)$ 为结构比较，α 为亮度权重因子，控制人眼对平均亮度差异的敏感程度；β 为对比度权重因子，控制人眼对信号幅值/纹理强度差异的敏感程度；γ 为结构权重因子，控制人眼对空间拓扑/几何关联差异的敏感程度。在标准的 SSIM 实现中，通常设定 $\alpha=\beta=\gamma=1$。SSIM 从亮度、对比度与结构保持三个维度建模，有效地表征了复原图像对原始场景物理几何特征的恢复能力，对于纯能量指标 PSNR，它很好的补充评价算法在保持物体边缘锐度与拓扑完整性方面的性能。

3.4　基准实验对比

为系统性地评估 EFMAN 在复杂动态场景下的复原效能，我们在上述两个基准数据集上，选取峰值信噪比 PSNR 和结构相似性 SSIM 作为量化指标，与当前最具代表性的行业先进算法展开严格的对比实验。考虑到目前开源的事件相机去模糊方案相对匮乏，为保证对比的全面性，我们将实验对象划分为基于纯强度图像的传统算法与基于事件辅助的多模态算法两大类。

GoPro 数据集：表 3 的定量评估显示，EFMAN

在显著降低参数冗余的同时, 实现了性能的突破。在传统方法中, 尽管 FFTformer 取得了 34.21 dB 的最优成绩, 但由于缺乏曝光时间内的帧间运动信息, 其在应对非线性高速运动时仍面临严重的病态反演难题。相较之下, 本方法通过引入具备微秒级分辨率的事件流先验, 有效补全了丢失的时空流形, 促使 PSNR 显著提升了 2.44 dB。在多模态融合赛道, EFMAN 以 36.65 dB / 0.977 的成绩稳步超越了 STCNet。但是, 网络参数规模小并不一定代表实际运行时间最短或计算复杂度最低。为了全面验证模型在实际应用中的部署潜力, 我们在单张 NVIDIA RTX 3090 GPU 上, 统一测试了各模型在 256×256 分辨率下的计算复杂度(FLOPs)与单帧推理时间。测试表明, EFMAN 的参数量为 6.59M, 计算复杂度可以达到 15.2G FLOPs, 单帧平均推理时间仅为 28.5 ms。这一成绩在计算开销与运行速度上领先于 HINet (85.4 ms) 和 NAFNet (48.6 ms) 等大规模网络。这直观且严谨地证明了, 得益于 LCSA 模块对通道和空间特征计算的解耦设计, EFMAN 成功规避了高分辨率特征图下的全局密集运算, 在轻量化设计与高保真重建之间取得了较好的平衡。

表 3　各模型在 GoPro 数据集上的性能对比

Tab. 3　Quantitative performance comparison of various models on the GoPro dataset

Method	Input		PSNR	SSIM	Params(M)	FLOPs(G)	Inference Time (ms)
	Events	Image					
DeblunGAN-v2[33]	×	√	29.55	0.934	60.90	41.5	35.2
SRN[35]	×	√	30.26	0.934	10.25	52.8	45.1
MPRNet[34]	×	√	32.66	0.959	20.10	145.5	78.3
HINet[36]	×	√	32.71	0.959	88.67	170.5	85.4
Restormer[17]	×	√	32.92	0.961	26.13	140.7	82.1
IRNeXt[37]	×	√	33.16	0.962	13.21	32.5	30.8
ChaIR[38]	×	√	33.28	0.963	15.02	38.6	36.4
NAFNet[39]	×	√	33.69	0.967	67.89	63.2	48.6
FFTformer[40]	×	√	34.21	0.969	16.6	45.8	42.3
EFNet[25]	√	√	35.46	0.972	8.47	25.3	24.5
DiffEvent[41]	√	√	35.55	0.972	35.41	210.6	450.5
STCNet[42]	√	√	36.45	0.975	14.35	48.2	38.6
MAENet[17]	√	√	36.07	0.976	12.80	42.1	35.4
EFMAN(ours)	√	√	**36.65**	**0.977**	**6.59**	**15.2**	**28.5**

此外, 图 6 的可视化结果提供了直观佐证。在极端模糊区域, FFTformer 等传统方法难以重构清晰的几何边缘, 导致黑车车牌数字出现严重的伪影残留与字符粘连, 关键的高频纹理信息大量丢失。反观本文方法, 得益于对事件极性变化的精准捕捉, 成功消除了运动模糊, 使得数字轮廓锐利、清晰, 显著提升了图像的可读性与结构完整性。

同时, 在旁边白车车灯处的细节对比中可以看出, MAENet 虽然在一定程度上解决了运动残影问题, 但导致了部分色彩信息的丢失。而本文算法不仅完整清晰地复原了指示牌细节, 在第二组花朵图片的对比中, 更精确地恢复了边缘信息。在保证去模糊精度的前提下, 我们的算法恢复出了色彩更鲜艳、饱和度更高的图像, 显著增强了整体视觉对比度, 呈现出更接近真实环境(GT)的色彩深度。综上所述, 本算法在细节还原、纹理恢复及去除运动模糊方面均已达到先进水准。

REBlur 数据集: 为了进一步验证模型在处理真实物理环境下的非线性运动模糊及复杂感光噪

声时的泛化能力，我们在 REBlur 数据集上实施了对比实验。如表 4 所示，实验结果进一步印证了 EFMAN 在真实场景下的技术优越性：实验数据显示，所有引入事件流的方法在 PSNR 指标上均显著优于纯图像驱动的传统算法（如 HINet 和 Restormer），平均涨幅超过 2.5dB。这证明了在真实物理退化过程中，事件流所提供的亚毫秒级时空细节是重建高保真图像的关键。而在具备强竞争力的事件驱动方法中，我们的 EFMAN 算法 达到了 38.50dB 的 PSNR 和 0.978 的 SSIM，相比于此前的领先算法 MAENet，本方法在保持最高 SSIM 的同时，进一步提升了峰值信噪比，不仅如此，本方法在实际推理效率上也展现出了极大的优势。正如前文所述，复杂的网络架构往往难以兼顾速度，例如采用迭代采样机制的多模态扩散模型（如 DiffEvent）伴随着极高的推理延迟（约 450.5 ms），而近期最新的多模态网络 STCNet (38.6 ms) 与 MAENet (35.4 ms) 的耗时也相对较高。相比之下，EFMAN 在参数量为 6.59M、单帧推理时间为 28.5 ms 的条件下，依然取得了最优的复原效果。这进一步验证了我们所提出的事件与图像高效融合机制的卓越性，表明该设计不仅能够以轻量级的架构捕捉更深层次的物理约束，更能满足了高动态场景对算法实时性的严苛要求。

(a) 模糊图像 (a) Blurred image　(b) FFTformer效果 (b) FFTformer　(c) MAENet效果 (c) MAENet　(d) EFMAN效果 (d) EFMAN (ours)　(e) 真值图像 (e) GT

图 6　各模型在 Gopro 数据集上的实验效果图

Fig. 6　Visual comparison of deblurring results obtained by various models on the GoPro dataset

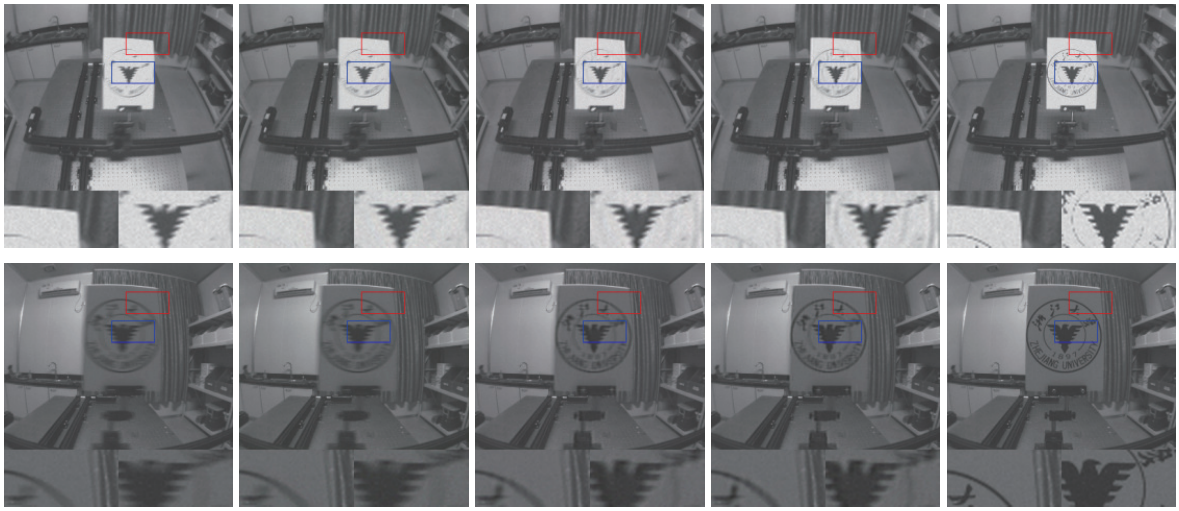
表 4　各模型在 REBlur 数据集上对比效果

Tab. 4　Quantitative performance comparison of various models on the REBlur dataset

Method	Input		PSNR	SSIM	Params(M)	FLOPs(G)	Inference Time (ms)
	Events	Image					
SRN	×	√	35.10	0.961	10.25	52.8	45.1
NAFNet	×	√	35.48	0.962	67.89	63.2	48.6
Restormer	×	√	35.50	0.959	26.13	140.7	82.1
HINet	×	√	35.58	0.965	88.67	170.5	85.4
EFNet	√	√	38.12	0.975	8.47	25.3	24.5
DiffEvent	√	√	38.23	0.974	35.41	210.6	450.5
STCNet	√	√	37.78	0.976	14.35	48.2	38.6
MAENet	√	√	38.46	0.978	12.80	42.1	35.4
EFMAN(ours)	√	√	**38.50**	**0.978**	**6.59**	**15.2**	**28.5**

同样的，我们在 REBlur 数据集对公开代码进行对比实验，将传统和基于事件的最新算法与我们网络进行可视化对比，在图 7 展示的视觉对比中，我们针对复杂运动路径下的校徽细节进行了局部放大分析。如第一组放大图所示，在存在旋转模糊的挑战下，对比算法普遍出现了严重的边

缘弥散与伪影叠加。相比之下, 本方法(EFMAN)几乎完全消除了校徽边缘的弥散阴影, 成功恢复了具有极高对比度的阶跃边缘, 使徽标的轮廓界限清晰可见。在第二组对比图中, 本算法表现出卓越的细节重构能力。在恢复校徽外部圆形几何轮廓的同时, 对上方汉字与下方英文标识进行了深度还原。观察可见, 对比算法在处理字符信息时存在明显的笔画粘连与模糊重影, 而本方法恢复的文字形态端正、笔画可辨, 显著提升了图像的语义可读性。



(a) 模糊图像 (a) Blurred image　(b) FFTformer效果 (b) FFTformer　(c) MAENet效果 (c) MAENet　(d) EFMAN效果 (d) EFMAN (ours)　(e) 真值图像 (e) GT

图 7　各模型在 ReBlur 数据集上的实验效果图

Fig. 7　Visual comparison of deblurring results obtained by various models on the REBlur dataset

在基准模型基础上加入 FEA 模块, 虽然自注意自监督的过程会使得模型整体的参数增加, 但同时也会使 PSNR 和 SSIM 分别提升 3.11dB 和 0.019, 这表明我们通过自注意力机制构建全局依赖的做法对基准模型整体进行了算法提升, 使得网络不再限制于局部感受野, 有效捕捉模糊图像的整体细节关系, 全局信息关联, 从而增加了大尺度非均匀模糊的恢复能力。在可视化中, 则是可以直观地感受到图片的细节部分得到大大增强, 相比于基准模型可以直观地看到更多的细节信息如窗帘的花纹和桌子的纹理细节等; 加入 LCSA 模块后, 模型的通道和空间信息得到全面的增强, 使得模糊图像中更多的遗漏信息经过这个模块被同步采集, 在数据上看更为直观 PSNR 和 SSIM 分别提升 3.43dB 和 0.031, 同时通过轻量化设计有效降低了模型冗余并大大加快了推理速度。这种“高性能-低计算成本”的特性显著提升了算法的部署效率。在可视化中, 则是可以直观地感受到我们的色彩, 纹理信息相比于基线模型得到了很好的恢复; 加入我们的损失函数后, 通过强制约束了事件流与图像梯度的时空一致性, 解决了模糊图像高频细节丢失的问题, 这一策略在推理阶段零成本的前提下, 带来了 3.23dB 的性能提升, 成功引导模型恢复出更锐利的边缘和纹理, 在可视化图像中, 则是可以清晰地看到, 图片质量变得更加高清, 更多的噪点被清除, 重建出具有更丰富信息的图像。当同时引入 FEA 和 LCSA 时, 模型在丰富的语义信息基础上实现了跨模态特征的深度融合。在此基础上进一步结合优化目标损失函数, 最终模型可以达到 PSNR 和 SSIM 分别达到了最优的 38.50 dB 和 0.978。此外, 为了从物理机理上进一步验证本文针对事件流特性设计的 LCSA 模块的不可替代性, 我们在消融实验中特别增加了一组结构替换对照组如表 5 中的实验 D 所示: 在保持基础网络、FEA 模块与联合损失函数完全一致的前提下, 将本文的 LCSA 模块等效替换为经典的 CBAM 注意力模块。实验结果显示, 使用 CBAM 的模型 PSNR 仅为 37.85 dB, 而使用本文 LCSA 模块的模型(EFMAN)达到了 38.50 dB。高达 0.65 dB 的性能落差与 0.006 的 SSIM 优势有力地证明了: 单纯的空间-通道注意力(如 CBAM)在处理常规 RGB 图像时固然有效, 但由于其内部采用了全连接层(MLP), 严重破坏了事件流体素在时间维度上的局部连续性, 难

以胜任跨模态的高频对齐；而本文 LCSA 中创新的 1D 卷积时序建模机制，从根本上抓住了事件数据特有的时间通道物理属性，是实现无伪影、高保真多模态特征融合的核心关键。

表 5　消融实验结果

Tab. 5　Quantitative results of the ablation study

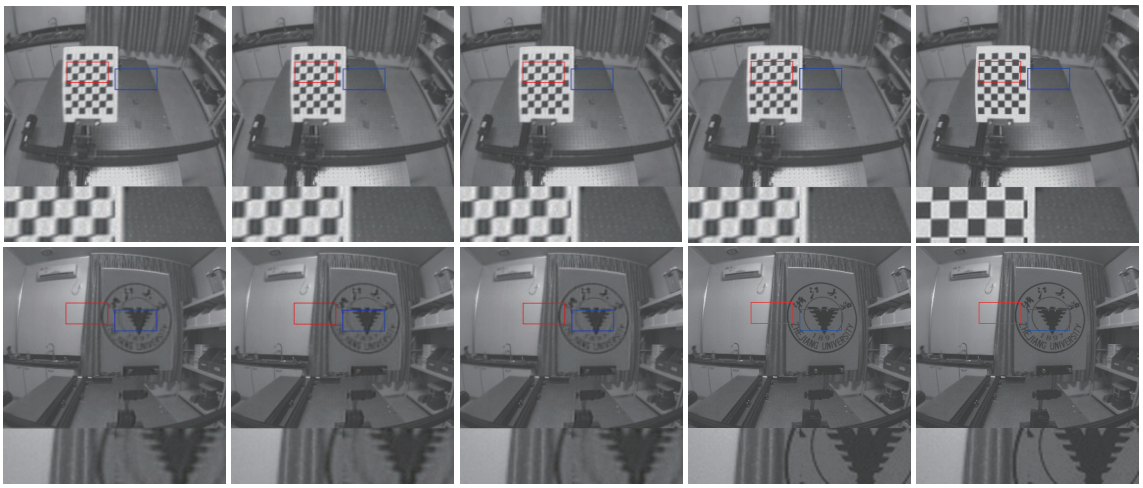
Model	Structure				PSNR	SSIM
	FEA	LCAS	Loss	CBAM		
Base	×	×	×	×	34.20	0.936
实验A	√	×	×	×	37.31	0.955
实验B	×	√	×	×	37.63	0.967
实验C	×	×	√	×	37.43	0.968
实验D	√	×	√	√	37.85	0.972
EFMAN	√	√	√	×	38.50	0.978

同时为了最纯粹地验证多维联合优化目标函数中各项附加损失的非冗余性与协同效应，我们在剔除了 FEA 与 LCSA 模块的基础网络（Base）上，开展了损失函数的渐进式消融实验（如表 6 所示）。在深度学习中，模型训练必须依赖基础误差函数，因此这里的 Base 模型代表仅使用基础像素级重建损失 ($\mathcal{L}_{pix}$ L_{pix}) 进行训练的基线状态。在此状态下，网络仅能恢复图像的基础低频轮廓，边缘表现平滑且缺乏高频细节，其 PSNR 为 34.20 dB，SSIM 为 0.936。在此基础上引入感知损失 ($\mathcal{L}_{per}$ L_{per}) 后，图像的色彩与深层结构语义得到了明显的视觉矫正，性能随之提升至 35.68dB。当继续叠加高频梯度损失 ($\mathcal{L}_{hf}$ L_{hf}) 时，事件流的高频物理先验被强制激活，图像边缘变得极为锐利（PSNR 提升至 36.95 dB），但在视觉上也伴随着背景中明显的离散热噪声的放大。最后，全变分正则化 ($\mathcal{L}_{tv}$ L_{tv}) 的引入起到了关键的平滑先验作用，它在视觉上抹平了孤立的噪声尖峰，同时保留了微观纹理的完整性。在推理阶段零额外参数成本的前提下，本文提出的完整多维联合损失函数为基础模型带来了 3.23 dB 的显著净提升，最终达到 37.43 dB (SSIM 0.968)。这一视觉质量与量化指标的双重递进过程，充分证明了级联这四项损失的物理必要性与科学性。

表 6　损失函数渐进式消融实验结果

Tab. 6　Progressive Ablation Study of Loss Functions

优化目标	PSNR	SSIM
Base(仅含 $\mathcal{L}_{pix}$)	34.20	0.936
Base + $\mathcal{L}_{per}$	35.68	0.948
Base + $\mathcal{L}_{per}$ + $\mathcal{L}_{hf}$	36.95	0.961
Base + $\mathcal{L}_{per}$ + $\mathcal{L}_{hf}$ + $\mathcal{L}_{tv}$	37.43	0.968



(a) 基础模型 (a) Base　(b) 添加FEA (b) FEA　(c) 添加LCSA (c) LCSA　(d) 添加损失函数 (d) Loss　(e) 完整模型 (e) Full

图 8　各模块消融效果图

Fig. 8　Visual results of the ablation study

3.5　对目标检测性能的影响

为了评估所提图像去模糊方法对下游任务的增益，我们在 EMRS 数据集[44] 上使用 YOLOv8 进行了目标检测实验。定量结果如表 7 所示，我

们的方法取得了 75.4% 的 mAP@0.5:0.95, 较模糊图像显著提升了 1.1%, 且仅比完全清晰的参考图像低 0.4%。在所有对比方法中, 我们的模型取得了最高的平均精度。此外, 基于深度学习的去模糊方法在检测性能上普遍优于传统方法, 这也进一步证明了其在恢复有助于目标识别的高层语义特征方面的优势。

表 7　目标检测任务上的定量对比

Tab. 7　Quantitative performance comparison on the object detection task

Model	Car	Bus	Truck	Two-wheel	Pedestrian	mAP@0.5: 0.95
Blur	0.855	0.877	0.723	0.488	0.485	0.743
DiffEvent	0.861	0.874	0.727	0.505	0.478	0.746
STCNet	0.872	0.882	0.726	0.525	0.477	0.747
MAENet	0.868	0.897	0.731	0.529	0.489	0.750
EFMAN (Ours)	0.887	0.905	0.740	0.533	0.505	0.754
Reference	0.895	0.913	0.746	0.536	0.510	0.758

图 9 的视觉对比进一步直观地展现了去模糊对检测结果的改善。所提方法有效消除了运动模糊的干扰, 显著提升了模型对密集小目标(如高架桥上的高速车辆)的感知能力, 大幅减少了误检和漏检现象。实验表明, 我们的模型能够有效提升挑战性场景下的目标检测精度, 并可直接应用于各类复杂环境下的智能分类与监测等下游任务。

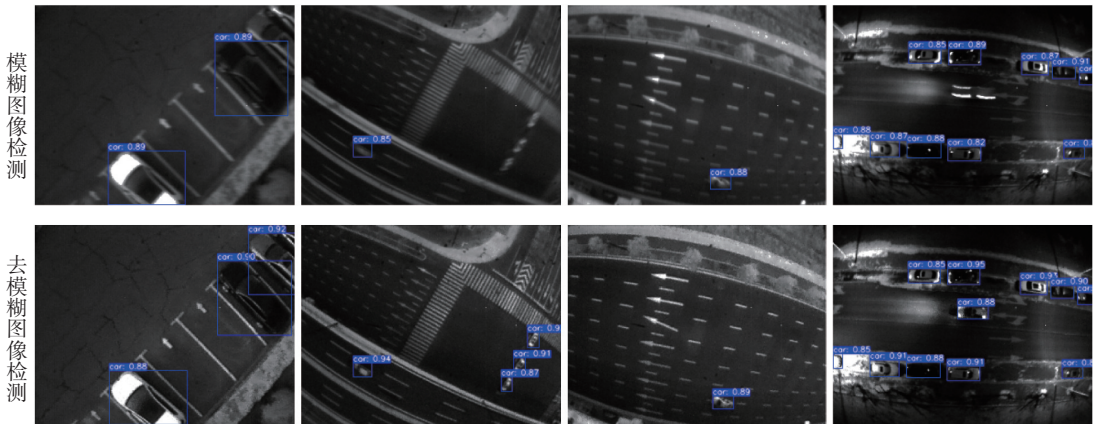


图 9　去模糊前后检测效果对比图

Fig. 9　Comparison of object detection performance before and after applying the deblurring method

4　结　论

本文针对单帧图像去模糊任务中固有的不适定性, 以及现有扩散模型推理延迟高、状态空间模型跨模态交互能力不足的问题, 提出了一种端到端的事件融合多头注意力网络 EFMAN。该网络通过引入事件相机的高频时空先验, 构建了跨模态自适应注意力机制, 实现了异步事件流与同步 RGB 特征在时空维度的精确对齐, 有效弥补了单一曝光时间内的信息缺失。

针对多模态融合过程中面临的传感器噪声干扰与计算效率瓶颈, 本文分别设计了特征增强注意力模块 FEA 与轻量级通道-空间注意力模块 LCSA。FEA 通过全局上下文建模显著增强了特征的抗噪鲁棒性, 有效解决了低光照及复杂场景下的噪声问题; LCSA 则通过解耦通道与空间维度的计算, 在大幅降低模型计算冗余的同时完成了特征响应的自适应重校准。此外, 本文构建的涵盖像素域、特征域及梯度域的多维联合损失函

数，协同优化了多尺度约束，确保了恢复图像在微观纹理细节与全局结构一致性上的平衡。

在 GoPro 与 REBlur 基准数据集上的广泛实验表明，EFMAN 在主观视觉质量与客观评价指标上均达到了当前先进水平。定量分析显示，相较于现有的 Transformer 及 SSM 基线方法，本方法在 GoPro 数据集上的峰值信噪比 PSNR 和结构相似度 SSIM 分别提升了 1.19 dB 和 0.005；在 REBlur 数据集上，PSNR 与 SSIM 则分别提升了 0.38 dB 和 0.003。综上所述，EFMAN 算法不仅有效克服了多模态特征对齐困难与传感器噪声干扰，还在去模糊质量与模型运行效率之间取得了显著平衡，为高动态及剧烈运动场景下的清晰图像重建提供了一种高效、鲁棒的解决方案。

参考文献：

[1] ZHANG K H, REN W Q, LUO W H, *et al*.. Deep image deblurring: a survey[J]. *International Journal of Computer Vision*, 2022, 130(9): 2103-2130.

[2] NAH S, KIM T H, LEE K M. Deep multi-scale convolutional neural network for dynamic scene deblurring[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2017: 257-265.

[3] LUO Z W, GUSTAFSSON F K, ZHAO ZH, *et al*.. Image restoration with mean-reverting stochastic differential equations[C]. *Proceedings of the 40th International Conference on Machine Learning*, JMLR, 2023: 957.

[4] LIN X Q, HE J W, CHEN Z Y, *et al*.. DiffBIR: toward blind image restoration with generative diffusion prior[C]. *18th European Conference on Computer Vision – ECCV 2024*, Springer, 2024: 430-448.

[5] WU H J, ZHANG M Q, HE L CH, *et al*.. Enhancing diffusion model stability for image restoration via gradient management[C]. *Proceedings of the 33rd ACM International Conference on Multimedia*, Association for Computing Machinery, 2025: 10768-10777.

[6] GUO H, LI J M, DAI T, *et al*.. MambaIR: a simple baseline for image restoration with state-space model[C]. *18th European Conference on Computer Vision – ECCV 2024*, Springer, 2024: 222-241.

[7] LIU Y, TIAN Y J, ZHAO Y ZH, *et al*.. VMamba: visual state space model[C]. *Proceedings of the 38th International Conference on Neural Information Processing Systems*, Curran Associates Inc., 2024: 3273.

[8] ZHU L H, LIAO B CH, ZHANG Q, *et al*.. Vision mamba: efficient visual representation learning with bidirectional state space model[C]. *Proceedings of the 41st International Conference on Machine Learning*, JMLR, 2024: 2584.

[9] LI B Y, ZHAO H Y, WANG W X, *et al*.. MaIR: a locality- and continuity-preserving mamba for image restoration[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2025: 7491-7501.

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

[11] WANG Y F, LIAO K, ZHANG K, *et al*.. Reconfigurable versatile integrated photonic computing chip[J]. *eLight*, 2025, 5(1): 20.

[12] FANG X Y, HU X N, LI B L, *et al*.. Orbital angular momentum-mediated machine learning for high-accuracy mode-feature encoding[J]. *Light*: *Science & Applications*, 2024, 13(1): 49.

[13] GALLEGO G, DELBRUCK T, ORCHARD G, *et al*.. Event-based vision: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 154-180.

[14] LICHTSTEINER P, POSCH C, DELBRUCK T. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor[J]. *IEEE Journal of Solid-State Circuits*, 2008, 43(2): 566-576.

[15] 方应红, 徐伟, 朴永杰, 等. 事件视觉传感器发展现状与趋势[J]. 液晶与显示, 2021, 36(12): 1664-1673.
FANG Y H, XU W, PIAO Y J, *et al*.. Development status and trend of event-based vision sensor[J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(12): 1664-1673.

[16] 柳长源, 曹青, 刘金凤. 遥感图像双模态融合去云方法[J]. 光学 精密工程, 2025, 33(18): 2996-3007.
LIU CH Y, CAO Q, LIU J F. A bimodal fusion method for remote sensing images to cloud removal[J]. *Optics and Precision Engineering*, 2025, 33(18): 2996-3007.

[17] CHEN C, SHI H, YANG Y, *et al*.. Uncertainty-aware fusion for event-based deblurring[C]. ACM Multimedia, 2023. **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[18] ZHANG X, YU L, YANG W. Unifying motion deblurring and frame interpolation with events[C]. *2022 IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2022: 17744-17753.

[19] SONG CH, HUANG Q X, BAJAJ C. E-CIR: event-enhanced continuous intensity recovery[C]. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2022: 7793-7802.

[20] LIN X P, HUANG Y L, REN H W, *et al*.. ClearSight: human vision-inspired solutions for event-based motion deblurring[C]. *IEEE/CVF International Conference on Computer Vision*, IEEE, 2025: 7462-7471.

[21] XIAO Z Y, WANG X CH. Event-based video super-resolution via state space models[C]. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2025: 12564-12574.

[22] TULYAKOV S, GEHRIG D, GEORGOULIS S, *et al*.. Time lens: event-based video frame interpolation[C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2021: 16150-16159.

[23] ZAMIR S W, ARORA A, KHAN S, *et al*.. Restormer: efficient transformer for high-resolution image restoration[C]. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2022: 5718-5729.

[24] YAO M, HU J K, ZHOU ZH K, *et al*.. Spike-driven transformer[C]. *Proceedings of the 37th International Conference on Neural Information Processing Systems*, Curran Associates Inc. , 2023: 2798.

[25] SUN L, SAKARIDIS C, LIANG J Y, *et al*.. Event-based fusion for motion deblurring with cross-modal attention[C]. *17th European Conference on Computer Vision – ECCV 2022*, Springer, 2022: 412-428.

[26] 吕建威, 钱锋, 韩昊男, 等. 结合光源分割和线性图像深度估计的夜间图像去雾[J]. 中国光学, 2022, 15(1): 34-44.
LV J W, QIAN F, HAN H N, *et al*.. Nighttime image dehazing with a new light segmentation method and a linear image depth estimation model[J]. *Chinese Optics*, 2022, 15(1): 34-44.

[27] SUN L, ALFARANO A, DUAN P Q, *et al*.. NTIRE 2025 challenge on event-based image deblurring: methods and results[C]. *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 2025: 1315-1332.

[28] LI K CH, LI X H, WANG Y, *et al*.. VideoMamba: state space model for efficient video understanding[C]. *18th European Conference on Computer Vision – ECCV 2024*, Springer, 2024: 237-255.

[29] WOO S, PARK J, LEE J Y, *et al*.. CBAM: convolutional block attention module[C]. *15th European Conference on Computer Vision – ECCV 2018*, Springer, 2018: 3-19.

[30] 王慧, 曹召良, 王军. 改进丰富卷积特征算法的液滴边缘检测模型[J]. 中国光学 (中英文), 2024, 17(4): 886-895.
WANG H, CAO ZH L, WANG J. Improved droplet edge detection model based on RCF algorithm[J]. *Chinese Optics*, 2024, 17(4): 886-895.

[31] 贺兴, 王磊, 张鹏超, 等. 基于多维注意力网络的图像超分辨率重建[J]. 液晶与显示, 2025, 40(7): 1056-1066.
HE X, WANG L, ZHANG P CH, *et al*.. Image super-resolution reconstruction based on multidimensional attention network[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(7): 1056-1066.

[32] PAN L Y, SCHEERLINCK C, YU X, *et al*.. Bringing a blurry frame alive at high frame-rate with an event camera[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2019: 6813-6822.

[33] KUPYN O, MARTYNIUK T, WU J R, *et al*.. DeblurGAN-v2: deblurring (orders-of-magnitude) faster and better[C]. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, 2019: 8877-8886.

[34] ZAMIR S W, ARORA A, KHAN S, *et al*.. Multi-stage progressive image restoration[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2021: 14821-14831.

[35] TAO X, GAO H Y, SHEN X Y, *et al*.. Scale-recurrent network for deep image deblurring[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, 2018: 8174-8182.

[36] CHEN L Y, LU X, ZHANG J, *et al*.. HINet: half instance normalization network for image restoration[C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 2021: 182-192.

[37] CUI Y N, REN W Q, YANG S N, *et al*.. IRNeXt: rethinking convolutional network design for image restoration[C]. *Proceedings of the 40th International Conference on Machine Learning*, PMLR, 2023: 6545-6564.

[38] CUI Y N, KNOLL A. Exploring the potential of channel interactions for image restoration[J]. *Knowledge-Based Systems*, 2023, 282: 111156.

[39] CHEN L Y, CHU X J, ZHANG X Y, *et al*.. Simple Baselines for Image Restoration[C]. *17th European Conference on Computer Vision – ECCV 2022*, Springer, 2022: 17-33.

[40] KONG L, DONG X, ZHANG J, *et al*.. FFTformer: toward efficient image restoration via frequency domain learning[C]. CVPR, 2023. **(查阅网上资料, 未找到本条文献信息, 请确认)**.

[41] WANG P, HE J M, YAN Q S, *et al*.. DiffEvent: event residual diffusion for image deblurring[C]. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024: 3450-3454.

[42] YANG W, WU J J, MA J P, *et al*.. Motion deblurring via spatial-temporal collaboration of frames and events[C]. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, AAAI, 2024: 726.

[43] SUN ZH J, FU X Y, HUANG L ZH, *et al*.. Motion aware event representation-driven image deblurring[C]. *18th European Conference on Computer Vision – ECCV 2024*, Springer, 2024: 418-435.

[44] JING SH L, LV H Y, ZHAO Y CH, *et al*.. Hyper lightweight neural networks towards spike-driven deep residual learning[J]. *Knowledge-Based Systems*, 2025, 327: 114099.

作者简介：

顾佳林（1999—），男，吉林长春人，硕士研究生，2021 年于长春理工大学获得学士学位，主要从事深度学习、图像处理及动态视觉传感器应用研究。E-mail: gujialin23@mails.ucas.ac.cn

吕恒毅（1984—），男，辽宁大连人，博士，研究员，2018 年于中国科学院大学获得博士学位，主要从事空间相机电子学、航天遥感成像技术、动态视觉传感器应用以及人工智能先进成像技术研究。E-mail: lvhengyi@ciomp.ac.cn