软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2024,35(5):2220−2234 [doi: 10.13328/j.cnki.jos.007023]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

事件融合与空间注意力和时间记忆力的视频去雨网络*

孙上荃[1,2], 任文琦[3], 操晓春[3]

[1](中国科学院 信息工程研究所, 北京 100085)

[2](中国科学院大学 网络空间安全学院, 北京 100049)

[3](中山大学 网络空间安全学院, 广东 深圳 518107)

通信作者: 任文琦, E-mail: renwq3@mail.sysu.edu.cn

摘　要: 近年来数码视频拍摄设备不断升级, 其感光元件宽容度、快门速率的提升虽然极大程度地丰富了可拍摄景物的多样性, 雨痕这类由于雨滴高速穿过景深范围的退化元素也更容易被记录到, 作为前景的稠密雨痕阻挡了背景景物的有效信息, 从而影响图像的有效采集. 由此视频图像去雨成为一个亟待解决的问题, 以往的视频去雨方法集中在利用常规图像自身的信息, 但是由于常规相机的感光元件物理极限、快门机制约束等原因, 许多光学信息在采集时丢失, 影响后续的视频去雨效果. 由此, 利用事件数据与常规视频信息的互补性, 借助事件信息的高动态范围、时间分辨率高等优势, 提出基于事件数据融合与空间注意力和时间记忆力的视频去雨网络, 利用三维对齐将稀疏事件流转化为与图像大小匹配的表达形式, 叠加输入至集合了空间注意力机制的事件-图像融合处理模块, 有效提取图像的空间信息, 并在连续帧处理时使用跨帧记忆力模块将先前帧特征利用, 最后经过三维卷积与两个损失函数的约束. 在开源视频去雨数据集上验证所提方法的有效性, 同时达到了实时视频处理的标准.

关键词: 视频去雨; 事件数据; 多模态融合; 空间注意力; 时间记忆力

中图法分类号: TP391

中文引用格式: 孙上荃, 任文琦, 操晓春. 事件融合与空间注意力和时间记忆力的视频去雨网络. 软件学报, 2024, 35(5): 2220–2234. http://www.jos.org.cn/1000-9825/7023.htm

英文引用格式: Sun SQ, Ren WQ, Cao XC. Event-fusion-based Spatial Attentive and Temporal Memorable Network for Video Deraining. Ruan Jian Xue Bao/Journal of Software, 2024, 35(5): 2220–2234 (in Chinese). http://www.jos.org.cn/1000-9825/7023.htm

Event-fusion-based Spatial Attentive and Temporal Memorable Network for Video Deraining

SUN Shang-Quan[1,2], REN Wen-Qi[3], CAO Xiao-Chun[3]

[1](Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China)

[2](School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China)

[3](School of Cyber Science and Technology, Sun Yat-sen University, Shenzhen 518107, China)

Abstract: In recent years, digital video shooting equipment has been continuously upgraded. Although the improvement of the latitude of its image sensor and shutter rate has greatly enriched the diversity of the scene that can be photographed, the degraded factors such as rain streaks caused by raindrops passing through the field of view at high speed are also easier to be recorded. The dense rain streaks in the foreground block the effective information of the background scene, thus affecting the effective acquisition of images. Therefore, video image deraining becomes an urgent problem to be solved. The previous video deraining methods focus on using the information of conventional images themselves. However, due to the physical limit of the image sensors of conventional cameras, the constraints of the shutter mechanism, etc., much optical information is lost during video acquisition, which affects the subsequent video deraining effect. Therefore, taking advantage of the complementarity of event data and conventional video information, as well as the high dynamic range

* 基金项目: 国家自然科学基金 (62172409); 深圳市科技计划 (JCYJ20220530145209022)
本文由“多模态协同感知与融合技术”专题特约编辑孙立峰教授、宋新航副研究员、蒋树强教授、王莉莉教授、申恒涛教授推荐.
收稿时间: 2023-04-07; 修改时间: 2023-06-08; 采用时间: 2023-08-23; jos 在线出版时间: 2023-09-11
CNKI 网络首发时间: 2023-11-23

and high temporal resolution of event information, this study proposes a video deraining network based on event data fusion, spatial attention, and temporal memory, which uses three-dimensional alignment to convert the sparse event stream into an expression form that matches the size of the image and superimposes the input to the event-image fusion module that integrates the spatial attention mechanism, so as to effectively extract the spatial information of the image. In addition, in continuous frame processing, the inter-frame memory module is used to utilize the previous frame features, which are finally constrained by the three-dimensional convolution and two loss functions. The video deraining method is effective on the publicly available dataset and meets the standard of real-time video processing.

Key words: video deraining; event data; multi-mode fusion; spatial attention; temporal memory

降雨作为一种常见的恶劣气候, 一定程度上影响着雨天视频拍摄的效果. 密集分布、高速运动、高反射率的雨滴在穿过镜头时, 对背景景物造成大面积的遮挡, 丢失了期望记录的视觉信息. 为解决雨痕遮挡视频图像的问题, 视频去雨技术应运而生. 作为计算机视觉领域中底层视觉、图像复原技术的一种, 视频去雨技术辅助用户去除采集到视频图像中的雨痕, 提升观察者视觉效果的同时, 也为其他下游如目标检测、语义分割等高层视觉算法在降雨天气的应用[1]铺平道路.

现有的视频去雨算法旨在利用常规相机所拍摄到的退化视频图像完成雨痕去除的目的, 然而由于此问题是欠定的, 即受遮挡背景信息已经丢失, 仅依靠常规图像准确还原原始背景信息是困难的. 同时, 常规相机受限于许多物理特性无法完美捕捉物理世界连续的时空信息: (1) 无论是卷帘式快门还是全局快门图像信号均以批量方式记录传输, 两个邻接时间步之间的信息容易模糊丢失[2,3]. (2) 感光元件受限于光度宽容极限无法捕捉超出该极限的视觉信息, 由此造成欠曝光与过曝光问题[4], 此研究项目中的雨痕便是雨滴高反光率导致的过曝光现象. 对此, 尽管研究人员试图利用雨滴稀疏性[4]、雨痕随机性等图像先验缓解欠定性问题, 因其需要具备领域知识的专家对超参数进行人工调整, 其在复杂背景环境与多变降雨条件下的泛化性仍然不够稳定.

针对上述问题, 本文考虑到雨滴在视频中的光学与动力学性质, 引入事件数据[3−5]作为一种极佳、可辅助视频去雨任务的跨模态信息. 事件数据由于其时序分辨率高、时延低、动态范围大的特征, 极好地弥补了常规相机在拍摄雨天视频图像时丢失的视觉信息, 且其具有空间稀疏性、关注运动边缘的特性极其适合捕捉降雨天气视频拍摄过程中的高速运动雨滴. 事件数据以事件信息流的方式被记录和存储, 与常规相机拍摄照片的矩阵表达方式和深度学习中张量特征的表达方式均不相同不适配, 由此本文拟利用空间像素堆叠的方式将稀疏信息流转换为双通道的张量表示方式, 以使其与常规相机图像像素匹配, 同时将事件数据与常规图像进行通道串联作为输入利用事件-图像融合处理模块进行组合编码, 最终获得去雨图像特征.

与此同时, 大部分现有的视频去雨方法无法达到实时处理视频图像的要求, 而视频去雨任务作为视频处理任务的一种, 实时处理能力是相关技术能否落地实用的关键. 最原始的基于模型的方法对输入视频设计目标函数并通过优化该目标获得去雨后的视频图像, 然而该过程需要多步方程运算, 由于大多无法适配 GPU 的矩阵运算加速, 距离实时处理的目标相去甚远; 随着深度学习的发展, 借助 GPU 进行张量计算加速, 轻量级深度神经网络已经可以达到实时处理视频的要求, 但是现有的视频去雨方法要么重度依赖三维卷积, 要么过度利用渐进式自我迭代等技巧, 虽然去雨效果得到了一定提升, 其运算速度却大大降低. 如何使深度网络在融合事件数据、高效关注雨痕建模与去除的同时, 降低运算成本, 提升帧内、帧间信息利用率, 都将是极其棘手而关键的问题.

针对上述问题, 本文设计了基于事件数据融合与空间注意力和时间记忆力的视频去雨网络 (event fusion-based spatial attentive and temporal memorable network for video deraining, EFSATeM), 框架如图 1 所示. 在事件-图像融合处理模块中利用空间注意力机制更好地融合提取单帧事件与图像中的雨痕特征, 从而使事件-图像融合处理模块更好地关注雨痕位置已达到高效学习的效果. 同时利用长短期记忆模块将前序帧的深度注意力特征抽取融合入当前帧的特征, 并记录传承到后序帧, 以此达到前后连续帧之间的交流, 利用前序帧的深度特征提升学习效率, 同时增强帧间连续性. 最后利用三维卷积处理模块对初步去雨的图像串联进行进一步的处理, 以提升输出视频的平滑度. 网络输出以残差连接的方式, 以使网络特征聚焦于形成残差的雨痕区域, 减少学习难度.

本文的主要贡献点可总结为以下 3 点.

(1) 利用事件相机数据辅助常规相机图像, 通过双通道像素堆叠的方式将事件信息流无损失地转换为可以与

常规图像像素配对的张量表示形式, 通过与常规图像通道串联的方式形成输入, 设计事件-图像融合处理模块完成将事件数据融合进视频去雨任务中.

(2) 设计了基于事件数据融合与空间注意力和时间记忆力的视频去雨网络, 在图像融合处理模块中利用空间注意力机制更好地融合、提取单帧事件与图像中的雨痕特征, 利用长短期记忆模块将前序帧的深度注意力特征抽取融合入当前帧的特征, 最后利用三维卷积处理模块对初步去雨的图像串联进行最后的处理.

(3) 在现有数据集上进行了验证性实验, 与多个现有视频图像去雨方法以实时视频处理速度达到了最好效果, 从视觉示例、客观指标、复杂度计算等多个角度验证了提出的方法的可靠性与有效性. 同时对设计模块进行了系统的消融实验, 证明了各个模块的必要性.

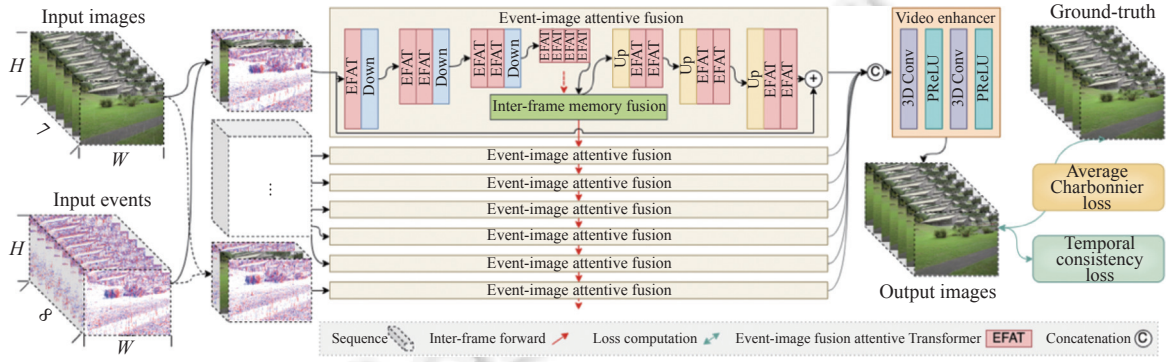


图 1 基于事件数据融合与空间注意力和时间记忆力的视频去雨网络框架图

本文第 1 节介绍视频图像去雨和事件数据融合的现有方法和科研现状. 第 2 节介绍本文相关的基础知识, 包括事件数据、雨痕建模、注意力和长短期记忆机制. 第 3 节介绍本文提出的基于事件数据融合与空间注意力和时间记忆力的视频去雨网络. 第 4 节通过对比实验、消融实验等方式验证所提网络的有效性. 最后总结全文.

1 融合事件数据与视频去雨相关工作

1.1 基于视频的雨痕去除

视频雨痕去除是 Garg 等人提出的单图去雨任务[6]的扩展, 他们基于雨滴动力学和光度学分析将雨痕和非雨区域分离, 从而解决了单图去雨问题. 而后, Barnum 等人[7]系统分析了雨天视频的时空属性, 并在频率空间构建了一个统计模型来检测雨痕区域. 许多工作[8–10]利用雨痕的稀疏性解决视频图像去雨问题. 肖进胜等人提出了一个低秩模型[8], 用稀疏编码处理图像中的雨痕特征. 同样, Kim 等人设计了一种低秩矩阵构建方法[9], 用于去除映射的雨痕. Ren 等人通过多标签马尔可夫随机场域将稀疏雨痕形式化, 以区分其他移动物体和雨痕[10]. 其他一些工作则利用视频中雨的光度学和几何外观和动态来帮助去雨, 例如 Chen 等人将雨检测视为像素级运动分割问题[11], 并考虑雨的光度学和色度学约束. 另一些工作中, 视频中的雨也可以在补丁级别上被视为随机的, 并被高斯分布建模, 例如 Jiang 等人提出的一种基于张量的模型[5], 带有全变分正则化器, 并通过多元交替方向方法来求解模型. 一些最近的工作同时考虑了雨的稀疏性和动态特性, 比如 Li 等人基于单张图像去雨和超分辨率中使用的传统稀疏编码方法[12], 开发了多尺度版本的卷积稀疏编码. Jiang 等人[13]考虑了所有已知的雨的判别特征, 并将它们公式化为 4 个损失项. 然而, 由于复杂的背景和大运动的存在, 真实的雨天视频可能会扭曲雨痕的稀疏性和动态假设, 并降低基于模型的方法的鲁棒性.

近年来, 例如卷积神经网络 (convolutional neural network, CNN) 和循环神经网络 (recurrent neural network, RNN) 等基于深度学习的方法, 已经被证明在视频去雨痕任务中非常有效. Chen 等人[14]利用超像素分割技术将帧分割成有意义的区域, 然后将分割特征传递到 CNN 中以获得增强去雨结果. Liu 等人[15]通过考虑雨痕的叠加重新制定了雨痕遮挡问题, 并提出了一个循环 CNN 来对这个问题进行针对性的求解, 同时他们利用概率模型生成了两

个具有时间不相关雨痕的视频去雨数据集. 根据其重新制定的雨痕遮挡模型, 动态路由残留循环网络 (dynamic routing residue recurrent network, D3R-Net)[16]级联了一个用于空间特征提取的残差 CNN 和一个用于时间特征提取的 RNN, 其 CNN 和 RNN 采用软动态路由机制, 其中每层的输出是多个平行单元的加权总和, 而权重又由可学习的门控制. Yang 等人[17]认为去雨是雨痕模拟合成的逆过程, 并引入一个逆恢复模型, 其参数由相应的神经网络负责估计. 之后 Yang 等人[18]设计了一个无监督的 CNN 架构, 辅助使用光流估计算法进行帧对齐. 张学锋等人利用双注意力残差的循环网络逐阶段地去除雨痕[19]. 在深度学习动态雨生成器的启发下, Yue 等人开发了半监督视频去雨框架 S2VD (semi-supervised video deraining)[20], 该框架由去雨 CNN 和雨痕生成 CNN 组成, 旨在弥合实际雨视频和合成雨视频之间的差距. 孟祥玉等人开发了一个基于运动估计与时空结合的视频去雨网络[21], 引入基于残差连接的编码器解码器结构处理雨痕去除问题. Zhang 等人则设计了一个简单的残差网络[22], 利用帧间信息提取来更好地捕捉时空特征和相邻帧之间的时序相关性. 然而, 传统相机的雨迹缺失信息以及各种迥异的雨痕模式阻碍了现有视频去雨算法的性能, 以上基于神经网络的方法都难以高效对雨痕特征和雨滴运动信息进行建模, 相比之下, 本文方法能借助事件编码雨痕的动态特征, 并通过空间注意力与时间记忆力提升各种雨痕建模效率.

1.2 融合事件数据的视觉任务

作为一种新型的神经形态相机, 例如 DVS (dynamic vision sensor)[3]、ATIS (asynchronous time-based image sensor)[23]和 DAVIS (dynamic and active pixel vision sensor)[24]等事件相机可以在像素级和微秒级别异步捕捉光照的强度变化, 这使得其相对传统相机在高动态范围和低延迟方面具有极大优势[25]. 由于其优越性, 事件已被广泛研究和利用于许多视觉任务, 包括立体视觉[26]、姿态估计[27]、视觉跟踪[28]、光流估计[29]、图像重建[30]、HDR 重建[31]等. 与此同时, 研究者也开发了许多技术以处理事件数据, 例如脉冲神经网络[32]、确定性滤波器[33]、脉冲时间相关性[34]和脉冲变压器[35]等. 基于事件的视频图像恢复工作目前集中在视频图像去模糊[36], 其主要利用事件数据极高的时间分辨率来补充传统相机缺失的信息. 与处理事件数据的现有方法相比, 我们提出利用事件-图像数据对齐融合技术、空间注意力和时间记忆力来更好地编码融合事件数据的时空特征.

2 基础知识

本文所提算法主要基于事件数据、雨痕建模、注意力机制和长短期记忆机制, 本节将就相关基础知识和有关基本概念予以简单介绍.

2.1 事件数据流

常规相机基于快门的方式将感光元件像素批量地激活生成 RGB 三通道图像 $\boldsymbol{x}\in\mathbb{R}^{3\times W\times H}$, 其中 W 和 H 分别是生成图像的宽度和长度, 快门关闭导致信息记录不完全、目标运动模糊. 与常规相机记录光照的绝对光强相反, 事件相机以光强的相对变化作为主要记录信息, 且以信息流的方式进行记录和存储, 即 $e_i=(x_i,y_i,t_i,p_i)$, 其中 $x_i\in[1,W]$ 和 $y_i\in[1,H]$ 分别为第 i 个被激活像素的空间坐标, 而 t_i 是事件的时刻坐标, 最后 $p_i\in\{-1,+1\}$ 记录事件的极化信息. 当 $p_i=-1$ 时该时空坐标的像素捕捉到光强减弱, 当 $p_i=+1$ 时则反之. 事件数据由此可以有极大的时间分辨率和动态范围, 例如 DAVIS 事件相机的时间分辨率可以达到 3 μs, 其动态范围可达 130 dB 以上[24].

事件数据流的具体生成机理展示在图 2 中. 以雨天的某个雨滴为例, 在雨滴作高速下坠运动时在 T 时刻首次出现在镜头前, 此时由于雨滴运动穿越了多个像素位置, 3 个连续像素因为雨滴的高反光导致的光强变化而被激活, 因此产生了 3 个事件数据流 $(x_1,y_1,T,+1)$、$(x_1,y_2,T,+1)$ 和 $(x_1,y_3,T,+1)$, 其中 x_1 为 3 个像素的横轴坐标, 而 y_1、y_2 和 y_3 是 3 个像素的纵轴坐标, 由于此刻 3 个像素接收光强变大, 故极化信息均为 $+1$. 当时间过了 t 而到了 $T+t$ 时刻时, 雨滴已经下落到了更低的像素位置, 因此先前的 3 个像素光强变小, 而此时雨滴所处 3 个像素位置的接收光强变大, 因此像素产生了 6 个事件流, 分别为 3 个负向极化信息和 3 个极化正向, 在图上分别以蓝色和红色表示. 由于事件相机采集事件数据的时间步长达到了微秒级别的小, 其在常规相机采集连续两帧之间的时间便足以产生数以百万计的事件信息流, 且其与光强绝对强度无关, 可以做到极大动态范围, 同时没有快门关闭的约束, 可以保持不间断持续事件数据采集.

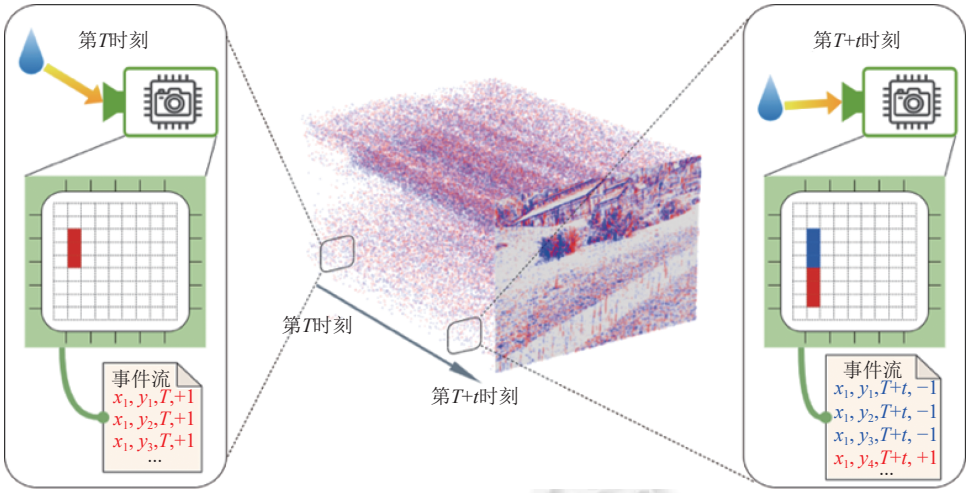


图 2 事件数据流在降雨场景中的生成机理

2.2 视频去雨建模

在单图去雨任务中, 通常将雨图像 O 拆分为雨痕图层 R 与背景图层 B, 即如公式 (1) 所示:

$$O = R + B \tag{1}$$

但是视频需要考虑到多帧图像以及雨痕图层与背景图层的帧间关系, 其物理建模需要改写为以下表达式:

$$O_i = R_i + B_i \tag{2}$$

其中, i 表示为视频中的第 i 帧图像. 由于物体在前后帧图像内的移动往往连续, 可以设定一个掩码扭曲函数可以将前后帧的图像对齐, 即 $O_i = W_o(O_{i+1})$, 在实际中使用光流估计算法来拟合这种对齐函数. 相应地, 雨痕图层和背景图层因其帧间连续的特点也均存在可以估计的扭曲对齐函数, 以 W_R 和 W_B 表示. 由于雨痕图层的遮挡, 背景图层的部分内容丢失, 而雨痕去除算法就是设法仅以雨视频图像作为输入, 估计还原出原始的背景图层.

2.3 注意力机制

注意力机制由 Vaswani 等人[37]提出, 替代了传统的循环神经网络, 使模型能够同时关注输入序列中的所有位置编码, 从而使其更加高效和可并行化, 它也因此成为自然语言处理任务中广泛使用的架构. 而注意力机制也同样可以应用到计算机视觉任务中, 使用注意力层代替或者辅助传统的卷积层, 使模型能够直接关注图像的不同空间或通道区域. 其具体过程如下, 给定长度为 N 的输入序列, 注意力机制需要计算出一组注意力权重矩阵 A, 它表示在预测第 i 个区域时需要关注第 j 个区域的重要程度. 这些注意力权重是由 3 个矩阵计算而来, 即查询矩阵 Q、键矩阵 K 和值矩阵 V, 其均由同一输入序列经过神经网络计算获得. 注意力权重矩阵的计算式如下:

$$A = Softmax\left(QK^{\mathrm{T}} / \sqrt{d_k}\right) \tag{3}$$

其中, d_K 是键向量 K 的维度. 获得注意力权重矩阵后输入序列中上下文矩阵 c 由注意力权重矩阵和值矩阵 V 的矩阵乘法计算而成:

$$c = AV \tag{4}$$

无论是文本序列还是图像, 注意力机制可以捕捉输入的不同区域之间的关系, 允许模型关注相关信息并抑制无关信息, 从而在各种任务上提高了性能.

2.4 长短时记忆模块

长短时记忆模型 (long-short term memory model, LSTM) 由 Hochreiter 等人提出[38], 是一种常用的循环神经网

络变体. 相较于传统的 RNN, LSTM 通过引入门机制 (gate mechanism) 来有效解决了长序列依赖问题, 广泛应用于序列预测、语音识别、自然语言处理等领域. LSTM 模型中的一个关键组件是记忆单元 (memory cell), 它具有自我更新和存储状态的能力. 同时, LSTM 通过门控单元 (gate unit) 来控制记忆单元中信息的输入、输出和遗忘, 使得模型可以选择性地保留和遗忘先前的信息. 具体来说, LSTM 的记忆单元可以定义如下:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{5}$$

其中, c_t 表示时刻 t 的记忆单元状态, f_t 表示遗忘门 (forget gate) 的输出, 控制着上一时刻的记忆单元状态被保留的程度, i_t 表示输入门 (input gate) 的输出, 控制着当前信息被记忆单元更新的程度, g_t 表示当前时刻的候选状态. 遗忘门、输入门和候选状态可以通过公式 (6)–公式 (8) 计算:

$$f_t = Sigmoid(W_f[h_{t-1}, x_t] + b_f) \tag{6}$$

$$i_t = Sigmoid(W_i[h_{t-1}, x_t] + b_i) \tag{7}$$

$$g_t = \tanh(W_g[h_{t-1}, x_t] + b_g) \tag{8}$$

其中, h_{t-1} 表示上一时刻的隐藏状态, x_t 表示当前时刻 t 的输入, W_f 、 W_i 和 W_g 分别表示对应的权重矩阵, b_f 、 b_i 和 b_g 分别表示对应的偏置. 最后, LSTM 的输出可以通过公式 (9) 计算:

$$h_t = o_t \odot \tanh(c_t) \tag{9}$$

其中, o_t 表示输出门 (output gate) 的输出, 控制着当前时刻的输出是基于哪些记忆单元状态. LSTM 模型通过引入门机制来控制信息的输入、输出和遗忘, 从而解决了传统 RNN 中长序列依赖问题, 提高了模型的性能. 在本文中, 我们将权重矩阵和偏置的线性层计算改为了卷积层计算, 以适应视频图像数据特征.

3 基于事件数据融合与空间注意力和时间记忆力的视频去雨网络

本文利用深度神经网络的强大表征能力和特征融合能力, 提出一种基于事件数据融合与空间注意力和时间记忆力的视频去雨网络, 该模型由以下 3 个部分组成.

首先, 针对事件数据流与常规图像的表达方式不一致的问题, 我们采用了像素极化信息双通道求和的方法. 具体来说, 首先取出两个连续常规视频帧之间的所有事件信息流, 生成与常规视频图像长宽一致的双通道空白画布, 随后将每个事件数据流的空间位置所对应的像素处累加权值, 遇到正向极化信息的事件时对第 1 通道的对应像素值加一, 遇到反向极化的事件时对第 2 通道的对应像素值加一, 如此一来便将稀疏事件数据流转换成了与常规图像表征类似的图像表达形式, 为后续的事件-图像融合铺平道路.

其次, 为达到实时处理视频的目标, 利用二维图像卷积设计了事件-图像注意力融合模块. 在事件-图像注意力融合模块中, 主要利用二维逐点卷积、逐深度卷积与注意力机制将串联输入的事件-图像数据三联体进行深度融合, 经过 3 次下采样、上采样与多层事件-图像注意力融合 Transformer 的处理, 将雨视频图像初步复原得到去雨图像. 同时在经过 3 次下采样得到深度特征的阶段, 利用跨帧记忆力特征融合模块将前序帧的深度特征进行记忆融合, 此处主要利用到了长短期记忆机制, 以强化去雨效果.

最后, 使用视频增强模块对连续多帧初步处理好的图像进行几层三维卷积和 PReLU 激活层搭配的增强处理, 而后使用 Charbonnier 损失函数对去雨效果进行有力监督, 同时添加上将前后帧图像扭曲对齐的损失函数对视频时序连续性进行增强约束, 最终达到雨痕图层的去除与背景图层的估计还原. 最终模型的运算速度与去雨效果达到了较优平衡.

3.1 事件图像数据配对预处理

首先, 为了使事件与图像得到更好融合, 事件数据需要被转换为与图像数据相匹配的表示方式, 即将事件从稀疏信息流的表达方式转变为二维矩阵的表达方式. 为此需要进行事件数据预处理, 其过程如图 3 所示, 将事件信息逐个取出排列在三维空间中, 并对两帧图像之间相同像素位置的事件进行双通道合并, 正向极化信息的事件累加并入第 1 通道, 在图例中以红色表示, 负向极化信息的事件累加并入第 2 通道, 在图例中以蓝色表示, 最终形成与

视频图像长宽一致的双通道事件图像表达方式. 事件数据以其高动态范围和高时间分辨率的特点, 记录下以雨线为代表的移动反光物体, 为后续的视频去雨任务提供雨线位置的掩码信息和尺度信息. 且事件数累加的形式提供了视频帧间隔内像素位置的光强变化次数, 提供了反光物体的移动幅度等信息.

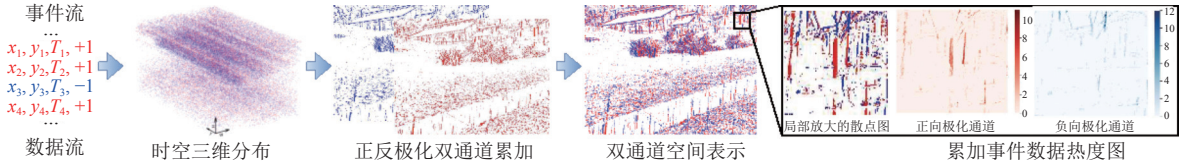


图 3 事件数据从信息流表示转化为双通道空间表示的预处理

接着, 多帧事件图像需要与多帧视频图像进行交错合并, 形成可以输入神经网络的输入数据. 具体而言, 如图 4 所示, 一串 7 帧的视频图像有 8 帧与之相邻的事件图像, 得到这 15 帧图像之后, 可以将其交错对齐, 使相邻的事件图像与视频图像在序列维度邻接排列, 而后在输入至基于事件数融合与空间注意力和时间记忆力的视频去雨网络时, 为减少过多输入数据带来的信息冗余和计算负担, 仅取出单帧视频图像和与其相邻的前后两帧的事件图像, 再将该三联体序列其在通道维度串联成通道数为 7 的二维矩阵, 为后续的事件-图像融合铺平道路.

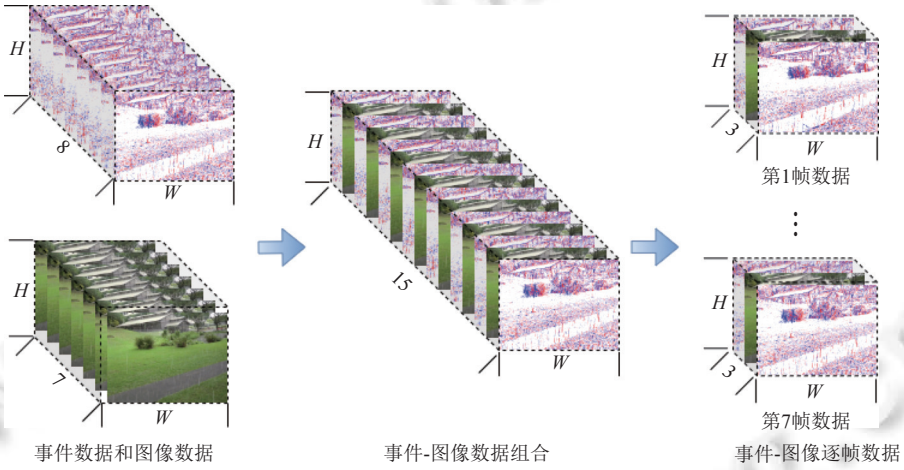


图 4 事件图像与视频图像的排列并对齐合并为输入形式

3.2 空间注意力与时间记忆力

获得可以输入模型的数据表达之后, 我们设计了基于事件数据融合与空间注意力和时间记忆力的视频去雨网络, 如图 1 所示. 其由若干个事件-图像注意力融合 Transformer 模块组成, 并且利用 3 次上下采样形成 U-Net 形式. 事件-图像注意力融合 Transformer 模块的结构则在图 5 中展示, 它是由空间注意力模块和前馈模块组成的 Transformer 网络, 中间利用层标准化 (layer normalization) 将深度特征进行标准化. 注意力模块利用逐点卷积层和逐深度卷积层分别提取特征中的空间与通道信息, 并拆分为注意力机制中所需的 Q 、 K 和 V 这 3 个张量特征, 将 Q 和 K 转换维度, 再进行 $Softmax(QK^{\mathrm{T}})$ 的操作获得空间通道注意力, 注意此处的注意力在通道维度展开, 3 次下采样与多层卷积将空间信息压缩混合至通道维度, 因此这样的通道注意力同样可以学习到空间注意力, 以此达到计算资源与注意力范围的平衡. 而后前馈模块则同样使用逐点卷积层与逐深度卷积层的串联混合提取空间与通道特征, 并沿通道拆分为两支, 一支通过 GeLU 激活层得到一个门控, 逐像素乘以另一支特征之后通过另一层逐点卷积层, 最后与输入特征相加输出残差特征.

连续输入两帧数据的事件-图像注意力融合模块之间也有特征的传递与交互, 由跨帧记忆力融合模块完成. 如图 6 所示, 该模块以卷积长短期记忆模块结构展开, 前序帧的特征 $[h_{t-1}, c_{t-1}]$ 和传至后序帧的输出特征 $[h_t, c_t]$ 的传

递以红色箭头显示, 先将前序输出特征 h_{t-1} 和当前特征串联, 通过 4 层卷积和激活层分别获得遗忘门、输入门、记忆单元和候选状态, 而后通过公式 (5) 和公式 (9) 获得输入至下一帧计算的深度特征.

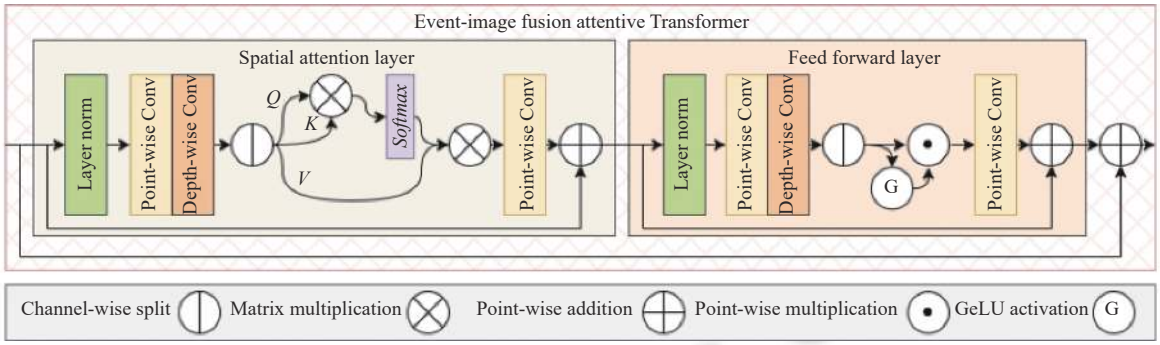


图 5 事件-图像注意力融合 Transformer 的结构

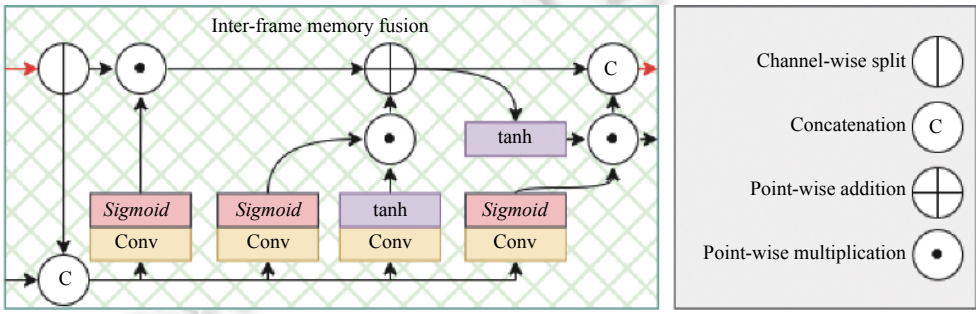


图 6 跨帧记忆力融合模块的结构

3.3 视频三维卷积与损失函数

通过事件-图像融合处理模块获得每一帧图像的初步去雨图像之后, 利用视频增强模块对去雨图像帧进行三维视频处理, 增强其去雨效果. 其内容在图 1 右侧展示, 由两层三维卷积和两层 PReLU 激活层组成, 由于其卷积通道数均为 RGB 图像的通道维度 3, 其计算负担较小, 为实时视频处理的目标做足准备, 同时由于其三维卷积的计算处理, 并且与下文将介绍的时序连续性损失函数相配合, 视频的帧间连续性可以得到有力保证和有效提升, 最终达到视频实时处理与视频去雨效果的有机结合和高效均衡.

而后我们设计了两个损失函数分别约束视频图像去雨效果和视频帧间连续性. 约束去雨效果的损失函数由 Charbonnier loss 构成, 给定输出的 7 帧去雨视频片段 $\left\{\tilde{B}_i\right\}_{i=1}^{7}$ 和已有的对应干净视频片段 $\{B_i\}_{i=1}^{7}$, 其表达式由公式 (10) 给出:

$$\mathcal{L}_1\left(\left\{\tilde{B}_i\right\}_{i=1}^{7},\{B_i\}_{i=1}^{7}\right)=\mathcal{L}_{\text{Charbonnier}}\left(\left\{\tilde{B}_i\right\}_{i=1}^{7},\{B_i\}_{i=1}^{7}\right)=\frac{1}{7}\sum_{i=1}^{7}\sqrt{\left(\tilde{B}_i-B_i\right)^2+\epsilon} \tag{10}$$

而后, 利用 LightFlowNet3[39]作为光流估计算法, 得到扭曲对齐函数 W_B, 我们可以对输出的去雨视频片段的前后连续帧进行互相对齐, 由此进行帧间连续性增强, 其表达式由公式 (11) 给出:

$$\mathcal{L}_2\left(\left\{\tilde{B}_i\right\}_{i=1}^{7}\right)=\frac{1}{6}\sum_{i=1}^{6}W_B\left(\tilde{B}_i\right)-\tilde{B}_{i+1}+\frac{1}{6}\sum_{i=1}^{6}W_B\left(\tilde{B}_{i+1}\right)-\tilde{B}_i \tag{11}$$

最终, 所用损失函数为前述两个损失函数的求和, 由公式 (12) 给出:

$$\mathcal{L}=\alpha\mathcal{L}_1+\beta\mathcal{L}_2 \tag{12}$$

其中, α 和 β 分别是两个损失函数权重, 其选值在后文给出. 至此, 算法的模型设计和损失函数设计均介绍完毕.

4 实验分析

4.1 实验数据

我们在公开视频去雨数据集 NTURain[14]上进行定量实验. 表 1 给出了 NTURain 数据集所对应的详细信息, 其中, 视频帧数量×3 表示对同一段干净视频模拟了 3 种雨痕视频, 其视频图像分为模拟雨痕的训练集、模拟雨痕的测试集以及真实雨痕的测试集, 图像长宽分别为 640 和 480, 视频雨痕模拟由视频编辑软件 Adobe Effects 完成, 在训练集上每一段干净视频均对应 3–4 种随机生成的雨痕视频, 训练集由此有共 3 123 帧图像, 模拟雨痕测试集有 1 682 帧, 真实雨痕测试集有 658 帧视频图像.

表 1 实验数据集 NTURain 的视频数据统计

数据集组别	雨痕真实性	镜头运动幅度	视频命名	视频帧数量
训练集	模拟雨痕视频	轻微镜头运动	t1	80×3
			t7	138×4
		大幅镜头运动	t2	112×3
			t3	128×3
			t4	136×3
			t5	128×3
			t6	138×3
			t8	135×3
测试集	模拟雨痕视频	轻微镜头运动	a1	168
			a2	116
			a3	125
			a4	298
		大幅镜头运动	b1	256
			b2	250
			b3	219
			b4	250
	真实雨痕视频	轻微镜头运动	ra1	60
			ra2	90
			ra3	80
			ra4	108
		大幅镜头运动	rb1	120
			rb2	60
			rb3	140

4.2 评价指标及基准模型

在本文中, 我们使用两个视频图像复原任务中常用的指标: 峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似性 (structural similarity, SSIM) 作为评价视频去雨算法的主要指标, 二者均为指标越大效果越好. 同时提供视频去雨的图像样例, 作为定性评价的依据.

我们将本文视频去雨算法与其他最先进的模型进行了比较, 包括多尺度卷积稀疏编码算法 (MSCSC)[12]、渐进式循环网络工作 (PReNet)[40]、细节计算的超像素对齐 CNN (SpacCNN)[14]、自学习去雨网络 (SLDNet)[18]、多阶段渐进式网络 (MPRNet)[41]、半监督视频去雨 (S2VD)[20]和增强时空交互网络 (ESTINet)[22]. 其中, MSCSC 是基于模型的方法, 其余的是基于深度学习的方法. PReNet 和 MPRNet 是单图去雨网络, 其余是视频去雨模型.

4.3 实验方法

我们将所有训练视频均匀裁剪为 7 帧 128×128 大小的图像块, batch size 选为 4. 学习率设置为 0.000 5, 当损失函数的值在连续两个 epoch 减小幅度小于 1E–6 时, 学习率缩小 10 倍. 优化器选用 Adam[42]. 公式 (12) 中的 α 和 β 分别设置为 1 和 0.1. 训练过程不采用任何数据增强或后处理. 训练总 epoch 数为 50. 对于对比的现有方法, 我们

保留其原始配置, 并根据它们发布的代码运行训练和测试实验. 我们使用 ESIM[43]来模拟视频序列中的事件. 所有实验在一台搭载 Nvidia Tesla V100 PCIe 32 GB GPU 和 Intel® Silver 4214 @ 2.20 GHz CPU 的服务器上进行, 服务器系统为 Linux Ubuntu 16.04.7 LTS. 实验使用的深度学习框架为 PyTorch 1.9.1, 使用了深度学习开源工具 Torchvision 0.10.1. 代码已发布在 https://github.com/sunsean21/EFSATem/.

4.4 实验结果与分析

我们首先对比了本文方法与现有方法的 PSNR 指标的差异, 它们在 NTURain 测试集的实验结果见表 2. 可以看出本文方法在所有测试视频中均获得了最高的评分. 而且相较于相机运动幅度较小的前 4 个视频, 本文方法在相机运动幅度较大的后 4 个视频的优势更为明显, 例如在 b4 视频中本文方法相较于第 2 名的 S2VD 算法取得了 3.63 dB 的显著提升, 这得益于我们对于视频图像帧间事件数据融合的有效利用, 且所提跨帧记忆力融合模块也有一定帮助, 具体分析会在消融实验给出.

表 2　本文方法与现有视频图像去雨算法的 PSNR 性能比较 (dB)

Clip	Rainy	MSCSC	PReNet	SpacCNN	SLDNet	MPRNet	S2VD	ESTINet	Ours
a1	29.71	25.10	32.13	30.57	33.72	35.80	36.39	36.99	**37.78**
a2	29.30	26.77	30.41	31.29	33.82	31.83	33.06	34.48	**34.71**
a3	29.08	24.71	30.73	30.63	33.12	34.38	35.75	36.09	**36.71**
a4	32.62	31.65	35.77	35.30	37.35	37.71	39.53	40.00	**41.56**
b1	30.03	26.35	32.66	32.26	34.21	36.75	37.34	37.15	**38.76**
b2	30.69	28.84	33.74	35.11	35.80	37.20	40.55	40.01	**40.69**
b3	32.31	26.63	35.34	34.69	36.34	39.30	38.83	38.06	**41.63**
b4	29.41	26.61	33.17	34.87	33.85	35.93	37.53	36.81	**41.16**
Mean	30.41	27.08	32.99	33.11	34.89	36.11	37.37	37.48	**39.12**

作者而后又比较了现有视频去雨算法与本文方法之间 SSIM 的差距, 结果展示在表 3 中. 其取值范围为 0–1, 当前大多数方法均已接近最优值. 尽管如此, 还是可以看出本文方法仍在大多数测试视频中取得了最佳的 SSIM 指标数值, 且相机运动幅度大的视频中本文方法的优势仍然更加明显.

表 3　本文方法与现有视频图像去雨算法的 SSIM 性能比较

Clip	Rainy	MSCSC	PReNet	SpacCNN	SLDNet	MPRNet	S2VD	ESTINet	Ours
a1	0.9149	0.7635	0.9511	0.9334	0.9508	0.9649	0.9658	0.9698	**0.9745**
a2	0.9284	0.8242	0.9375	0.9356	0.9512	0.9495	0.9519	0.9611	**0.9636**
a3	0.8964	0.7326	0.9316	0.9247	0.9404	0.9539	0.9564	0.9649	**0.9665**
a4	0.9381	0.9327	0.9700	0.9620	0.9722	0.9756	0.9779	0.9795	**0.9843**
b1	0.8956	0.7954	0.9491	0.9454	0.9482	0.9646	0.9712	0.9683	**0.9758**
b2	0.8874	0.8860	0.9557	0.9677	0.9595	0.9659	**0.9821**	0.9752	0.9811
b3	0.9299	0.8142	0.9681	0.9566	0.9614	0.9740	0.9754	0.9740	**0.9838**
b4	0.8933	0.8029	0.9526	0.9536	0.9469	0.9613	0.9657	0.9653	**0.9805**
Mean	0.9108	0.8189	0.9520	0.9475	0.9540	0.9637	0.9683	0.9700	**0.9763**

为了更清晰地进行视觉对比, 我们展示了几幅去雨图像案例, 其对比图在图 7 中显示. 在第 1 张对比图中, 可以看出红色花朵处有很多密集的雨痕, 已有的 3 种方法均无法将它们完全去除, 只有本文方法可以将这些与背景图案中的红色花朵混杂在一起的雨痕完全清除. 而在第 2 个示例对比图中, 柏油马路上有很多雨痕, 已有算法中 MPRNet 和 S2VD 均有几处明显的雨痕未能去除, 而 ESTINet 虽然能基本去除掉雨痕, 但是在原本雨痕处留下了很多条纹状伪影, 使本该平滑的柏油路面变得质地不一, 与之对比, 本文方法达到了与参考图像最为一致的复原效果, 将所有雨痕完全去除.

为了进一步分析视频去雨算法在真实雨痕视频上的效果, 我们展示了一张 NTURain 的真实雨痕视频帧上各

个去雨算法的效果对比图, 在图 8 中显示. 在图 8 中可以看出草地、地砖路面和柏油马路均有深浅不一较为密集的雨痕, PReNet 和 MPRNet 由于没有考虑到帧间时序信息, 均无法有效去除掉雨痕, 而 S2VD 虽然可以去除掉画面中间最为明显的雨滴, 但是对于草地和路面的雨痕却没有处理, 而 ESTINet 则虽然基本去除了多数雨痕, 但是却留有了一些浅浅的痕迹, 只有本文方法完全将所有雨痕去除, 证实了其在真实雨痕视频图像上的去雨效果.

图 7 视频去雨算法的雨痕去除效果案例展示

图 8 视频去雨算法的真实雨痕去除效果案例展示

4.5 消融实验

为了验证设计的各种模块的有效性, 我们进行了不同网络结构和损失函数设置下的消融实验. 除了最终版本

的模型, 我们定义了另外 5 个模型设置, 在 Model 0 中完全去除事件数据的编码; 在 Model 1 中将事件-图像注意力融合 Transformer 模块全部替换成了参数量相近的卷积层模块; 在 Model 2 中将跨帧记忆力融合模块完全抛弃; 在 Model 3 中将跨帧记忆力融合模块替换成了参数量相近的卷积层模块; 在 Model 4 中将 Charbonnier 损失函数替换成 MSE (mean square error) 损失函数. 最终结果展示在表 4 中, 根据对比结果可以发现事件-图像注意力融合 Transformer 模块最为重要, 而跨帧记忆力融合同样也有较大影响.

表 4 不同设置下本文方法性能比较

Metric	Model 0	Model 1	Model 2	Model 3	Model 4	Ours
PSNR (dB)	36.39	37.48	37.66	38.35	39.05	**39.12**
SSIM	0.967 1	0.968 9	0.969 5	0.972 9	0.975 5	**0.976 3**

4.6 复杂度对比

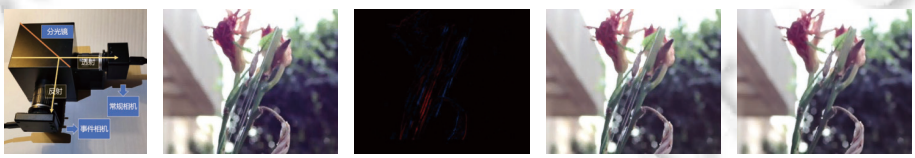
我们将本文方法与现有算法在 NTURain 测试集上进行了运算速度和模型复杂度的对比, 对比结果展示在表 5 之中. 由于本文方法不严重依赖三维卷积, 本方法运算速度相较于其他视频处理算法快, 同时我们使用逐点卷积层和逐深度卷积层串联的方式[44,45]替换常规卷积层, 最终使本文算法达到了视频实时去雨增强效果.

表 5 本文方法与现有视频图像去雨算法的每帧运行时间和模型复杂度比较

Metric	MSCSC	PReNet	SpacCNN	SLDNet	MPRNet	S2VD	ESTINet	Ours
Time (s)	15.33	0.18	8.23	2.65	0.15	0.14	0.27	**0.03**
Param (M)	—	0.17	—	4.00	3.64	0.53	**0.44**	**0.44**
FLOPs (G)	—	2 174	—	4 782	18 003	336	153	**64**

4.7 真实事件数据

我们使用一台事件图像采集设备拍摄了一段包含了真实事件数据的雨天视频, 结果如图 9 所示, 采集到的事件数据背景色为黑色, 可以看出一簇密集的雨线和植物的茎杆重合, 由此 ESTINet 无法有效去除这种真实的雨线, 而所提方法则成功去雨.



事件采集设备 Rainy Event ESTINet Ours

图 9 采集到包含真实事件的一张视频帧的去雨效果展示

5 总 结

针对当前视频去雨算法仅利用常规视频图像而丢失关键信息和特征提取困难的问题, 我们使用事件数据高动态范围、时间分辨率高的特点, 将其从信息流的表达形式转换为双通道图像形式, 并与常规视频帧进行空间像素对齐组合, 使用它对常规视频帧之间的时序信息和空间特征进行补充辅助. 同时设计了事件-图像注意力融合 Transformer 模块和跨帧记忆力融合模块, 组合构成基于事件数据融合与空间注意力和时间记忆力的视频去雨网络, 更加高效地提取和融合事件与视频图像帧之中的时空特征, 且这些模块主要依赖于二维卷积、逐点卷积与逐深度卷积的串联、长短期记忆机制这些高效模块, 实现了特征提取与高速运算的平衡, 最后经过两层三维卷积后输出去雨视频图像. 此方法的视频处理速度由此达到了实时运算的要求.

References:

[1] Li SY, Araujo IB, Ren WQ, Wang ZY, Tokuda EK, Junior RH, Cesar-Junior R, Zhang JW, Guo XJ, Cao XC. Single image deraining: A comprehensive benchmark analysis. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3838–3847. [doi: 10.1109/CVPR.2019.00396]

[2] Le T, Le NT, Jang YM. Performance of rolling shutter and global shutter camera in optical camera communications. In: Proc. of the 2015 Int'l Conf. on Information and Communication Technology Convergence. Jeju: IEEE, 2015. 124–128. [doi: 10.1109/ICTC.2015.7354509]

[3] Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor. IEEE Journal of Solid-state Circuits, 2008, 43(2): 566–576. [doi: 10.1109/JSSC.2007.914337]

[4] Fowles GR. Introduction to Modern Optics. 2nd ed., New York: Dover Publications, 1989.

[5] Jiang TX, Huang TZ, Zhao XL, Deng LJ, Wang Y. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2818–2827. [doi: 10.1109/CVPR.2017.301]

[6] Garg K, Nayar SK. Vision and rain. Int'l Journal of Computer Vision, 2007, 75(1): 3–27. [doi: 10.1007/s11263-006-0028-6]

[7] Barnum PC, Narasimhan S, Kanade T. Analysis of rain and snow in frequency space. Int'l Journal of Computer Vision, 2010, 86(2): 256–274. [doi: 10.1007/s11263-008-0200-2]

[8] Xiao JS, Wang W, Zou WT, Tong L, Lei JF. An image rain removal algorithm via depth of field and sparse coding. Chinese Journal of Computers, 2019, 42(9): 2024–2034 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2019.02024]

[9] Kim JH, Sim JY, Kim CS. Video deraining and desnowing using temporal correlation and low-rank matrix completion. IEEE Trans. on Image Processing, 2015, 24(9): 2658–2670. [doi: 10.1109/TIP.2015.2428933]

[10] Ren WH, Tian JD, Han Z, Chan A, Tang YD. Video desnowing and deraining based on matrix decomposition. In: Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 2838–2847. [doi: 10.1109/CVPR.2017.303]

[11] Chen J, Chau LP. A rain pixel recovery algorithm for videos with highly dynamic scenes. IEEE Trans. on Image Processing, 2014, 23(3): 1097–1104. [doi: 10.1109/TIP.2013.2290595]

[12] Li MH, Xie Q, Zhao Q, Wei W, Gu SH, Tao J, Meng DY. Video rain streak removal by multiscale convolutional sparse coding. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6644–6653. [doi: 10.1109/CVPR.2018.00695]

[13] Jiang TX, Huang TZ, Zhao XL, Deng LJ, Wang Y. FastDeRain: A novel video rain streak removal method using directional gradient priors. IEEE Trans. on Image Processing, 2019, 28(4): 2089–2102. [doi: 10.1109/TIP.2018.2880512]

[14] Chen J, Tan CH, Hou JH, Chau LP, Li H. Robust video content alignment and compensation for rain removal in a CNN framework. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6286–6295. [doi: 10.1109/CVPR.2018.00658]

[15] Liu JY, Yang WH, Yang S, Guo ZM. Erase or fill? Deep joint recurrent rain removal and reconstruction in videos. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 3233–3242. [doi: 10.1109/CVPR.2018.00341]

[16] Liu JY, Yang WH, Yang S, Guo ZM. D3R-Net: Dynamic routing residue recurrent network for video rain removal. IEEE Trans. on Image Processing, 2019, 28(2): 699–712. [doi: 10.1109/TIP.2018.2869722]

[17] Yang WH, Liu JY, Feng JS. Frame-consistent recurrent video deraining with dual-level flow. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 1661–1670. [doi: 10.1109/CVPR.2019.00176]

[18] Yang WH, Tan RT, Wang SQ, Liu JY. Self-learning video rain streak removal: When cyclic consistency meets temporal correspondence. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1717–1726. [doi: 10.1109/CVPR42600.2020.00179]

[19] Zhang XF, Li JJ. Single image de-raining using a recurrent dual-attention-residual ensemble network. Ruan Jian Xue Bao/Journal of Software, 2021, 32(10): 3283–3292 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6018.htm [doi: 10.13328/j.cnki.jos.006018]

[20] Yue ZS, Xie JW, Zhao Q, Meng DY. Semi-supervised video deraining with dynamical rain generator. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 642–652. [doi: 10.1109/CVPR46437.2021.00070]

[21] Meng XY, Xue XW, Li WL, Wang Y. Motion-estimation based space-temporal feature aggregation network for multi-frames rain removal. Computer Science, 2021, 48(5): 170–176 (in Chinese with English abstract). [doi: 10.11896/jsjkx.210100104]

[22] Zhang KH, Li DX, Luo WH, Ren WQ, Liu W. Enhanced spatio-temporal interaction learning for video deraining: Faster and better. IEEE

Trans. on Pattern Analysis and Machine Intelligence, 2023, 45(1): 1287–1293. [doi: 10.1109/TPAMI.2022.3148707]

[23] Posch C, Matolin D, Wohlgenannt R. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. IEEE Journal of Solid-state Circuits, 2011, 46(1): 259–275. [doi: 10.1109/JSSC.2010.2085952]

[24] Brandli C, Berner R, Yang MH, Liu SC, Delbruck T. A 240×180 130 dB 3 μs latency global shutter spatiotemporal vision sensor. IEEE Journal of Solid-state Circuits, 2014, 49(10): 2333–2341. [doi: 10.1109/JSSC.2014.2342715]

[25] Gallego G, Delbrück T, Orchard G, Bartolozzi C, Taba B, Censi A, Leutenegger S, Davison AJ, Conradt J, Daniilidis K, Scaramuzza D. Event-based vision: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2022, 44(1): 154–180. [doi: 10.1109/TPAMI.2020.3008413]

[26] Andreopoulos A, Kashyap HJ, Nayak TK, Amir A, Flickner MD. A low power, high throughput, fully event-based stereo system. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7532–7542. [doi: 10.1109/CVPR.2018.00786]

[27] Gallego G, Lund JEA, Mueggler E, Rebecq H, Delbruck T, Scaramuzza D. Event-based, 6-DOF camera tracking from photometric depth maps. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2018, 40(10): 2402–2412. [doi: 10.1109/TPAMI.2017.2769655]

[28] Mitrokhin A, Fermüller C, Parameshwara C, Aloimonos Y. Event-based moving object detection and tracking. In: Proc. of the 2018 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems. Madrid: IEEE, 2018. 1–9. [doi: 10.1109/IROS.2018.8593805]

[29] Wan ZX, Dai YC, Mao YX. Learning dense and continuous optical flow from an event camera. IEEE Trans. on Image Processing, 2022, 31: 7237–7251. [doi: 10.1109/TIP.2022.3220938]

[30] Wang L, Kim TK, Yoon KJ. EventSR: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 8312–8322. [doi: 10.1109/CVPR42600.2020.00834]

[31] Han J, Zhou C, Duan PQ, Tang YH, Xu C, Xu C, Huang TJ, Shi BX. Neuromorphic camera guided high dynamic range imaging. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1727–1736. [doi: 10.1109/CVPR42600.2020.00180]

[32] Orchard G, Benosman R, Etienne-Cummings R, Thakor NV. A spiking neural network architecture for visual motion estimation. In: Proc. of the 2013 IEEE Biomedical Circuits and Systems Conf. Rotterdam: IEEE, 2013. 298–301. [doi: 10.1109/BioCAS.2013.6679698]

[33] Brosch T, Tschechne S, Neumann H. On event-based optical flow detection. Frontiers in Neuroscience, 2015, 9: 137. [doi: 10.3389/fnins.2015.00137]

[34] Paredes-Vallés F, Scheper KYW, de Croon GCHE. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2051–2064. [doi: 10.1109/TPAMI.2019.2903179]

[35] Zhang JQ, Dong B, Zhang HW, Ding JC, Heide F, Yin BC, Yang X. Spiking transformers for event-based single object tracking. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 8791–8800. [doi: 10.1109/CVPR52688.2022.00860]

[36] Shang W, Ren DW, Zou DQ, Ren JS, Luo P, Zuo WM. Bringing events into video deblurring with non-consecutively blurry frames. In: Proc. of the 2021 IEEE/CVF Int'l Conf. on Computer Vision. Montreal: IEEE, 2021. 4511–4520. [doi: 10.1109/ICCV48922.2021.00449]

[37] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

[38] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: 10.1162/neco.1997.9.8.1735]

[39] Hui TW, Loy CC. LiteFlowNet3: Resolving correspondence ambiguity for more accurate optical flow estimation. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 169–184. [doi: 10.1007/978-3-030-58565-5_11]

[40] Ren DW, Zuo WM, Hu QH, Zhu PF, Meng DY. Progressive image deraining networks: A better and simpler baseline. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3932–3941. [doi: 10.1109/CVPR.2019.00406]

[41] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH, Shao L. Multi-stage progressive image restoration. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 14816–14826. [doi: 10.1109/CVPR46437.2021.01458]

[42] Kingma DP, Ba JL. Adam: A method for stochastic optimization. arXiv:1412.6980, 2015.

[43] Rebecq H, Gehrig D, Scaramuzza D. ESIM: An open event camera simulator. In: Proc. of the 2nd Annual Conf. on Robot Learning. Zürich: PMLR, 2018. 969–982.

[44] Tan MX, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc. of the 36th Int'l Conf. on Machine

Learning. Long Beach: PMLR, 2019. 6105–6114.

[45] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH. Restormer: Efficient transformer for high-resolution image restoration. In: Proc. of the 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 5728–5739. [doi: 10.1109/CVPR52688.2022.00564]

附中文参考文献:

[8] 肖进胜, 王文, 邹文涛, 童乐, 雷俊锋. 基于景深和稀疏编码的图像去雨算法. 计算机学报, 2019, 42(9): 2024–2034. [doi: 10.11897/SP.J.1016.2019.02024]

[19] 张学锋, 李金晶. 基于双注意力残差循环单幅图像去雨集成网络. 软件学报, 2021, 32(10): 3283–3292. http://www.jos.org.cn/1000-9825/6018.htm [doi: 10.13328/j.cnki.jos.006018]

[21] 孟祥玉, 薛昕惟, 李汶霖, 王祎. 基于运动估计与时空结合的多帧融合去雨网络. 计算机科学, 2021, 48(5): 170–176. [doi: 10.11896/jsjkx.210100104]

孙上荃(1997－), 男, 博士生, CCF 学生会员, 主要研究领域为底层视觉, 图像复原, 图像增强, 轻量化模型.

操晓春(1980－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为人工智能安全, 网络空间内容安全.

任文琦(1987－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为人工智能, 计算机视觉, 图像处理, 网络空间内容安全.