

## Reference

- [1] H. Cho, T. Kim, Y. Jeong, and K.-J. Yoon, "TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 25701–25711.
- [2] J. Dong, K. Ota, and M. Dong, "Video Frame Interpolation: A Comprehensive Survey," ACM Trans. Multimedia Comput. Commun. Appl., vol. 19, no. 2s, May 2023, doi: 10.1145/3556544.
- [3] O. S. Kılıç, A. Akman, and A. A. Alatan, "E-VFIA: Event-Based Video Frame Interpolation with Attention," in 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 8284–8290. doi: 10.1109/ICRA48891.2023.10160276.
- [4] T. Kim, Y. Chae, H.-K. Jang, and K.-J. Yoon, "Event-Based Video Frame Interpolation With Cross-Modal Asymmetric Bidirectional Motion Fields," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 18032–18042.
- [5] L. Sun et al., "Event-Based Frame Interpolation with Ad-hoc Deblurring," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 18043–18052. doi: 10.1109/CVPR52729.2023.01730.
- [6] S. Tulyakov, A. Bochicchio, D. Gehrig, S. Georgoulis, Y. Li, and D. Scaramuzza, "Time Lens++: Event-based Frame Interpolation with Non-linear Parametric Flow and Multi-scale Fusion," IEEE Conference on Computer Vision and Pattern Recognition, 2022.
- [7] S. Tulyakov et al., "TimeLens: Event-based Video Frame Interpolation," IEEE Conference on Computer Vision and Pattern Recognition, 2021.
- [8] S. Lin et al., "Learning Event-Driven Video Deblurring and Interpolation," ECCV, 2020.

## 概述

VFI，全称Video Frame Interpolation，即视频插帧，在两个连续视频帧之间合成不存在的帧。传统的VFI解决方案分为两步：首先估计帧间的运动信息（通常基于光流/卷积核/光流-卷积核混合/相位变化），然后进行像素合成。插帧为很多下游任务（慢动作生成、帧率提升、视频压缩、新视图合成、视频修复、帧内预测等）提供了技术条件和应用基础。

## Event-based VFI方法对比

Citation	Motivations	Model	Strategies	Multimedia Resources	Other Info
----------	-------------	-------	------------	----------------------	------------

Citation	Motivations	Model	Strategies	Multimedia Resources	Other Info
2020 ECCV Learning Event-Driven Video Deblurring and Interpolation	<p>核心目的：捕捉高速运动，传统的相机做不到，只能通过去模糊和插帧解决。</p> <p>大量工作将事件流直接转换为像素视频流。然而重建的视频往往缺乏纹理，并且在没有额外像素强度信息的情况下看起来不逼真。</p> <p>利用强度传感器（传统摄像头）和基于事件的传感器（事件相机）的优点进行高速视频生成是可取的，且即便有一小部分研究注意到了两种信源，这些工作没有把去模糊化考虑进来。</p> <p>前人工作中提出了 EDI (Event-based Double Integral) 模型将给定模糊帧和相应的事件流经过去模糊后，通过估计它们之间的残差，可以从上述初步去模糊的图像中插值出其他潜在视频帧。该方法自然地连接了强度图像和事件数据，在高帧率视频生成方面显示出不错的结果。但是由于事件相机的触发阈值随着硬件和场景条件在空间和时间上变化，会引入强烈的累积噪声。</p>	IntegralNet+GateNet	<span>◆ 残差估计</span> <span>◆ 关键帧去模糊</span> <span>◆ 插帧</span> <span>◆ 帧融合</span>	<span>None</span>	<span>主要聚焦高速视频重建任务，插入帧作为该任务其中一个策略。</span>

Citation	Motivations	Model	Strategies	Multimedia Resources	Other Info
2021 CVPR Time Lens: Event-based Video Frame Interpolation	<p>所有情节基于帧的VFI由于缺乏帧间盲区时间的视觉信息而必须假设简化的运动模型（例如线性）【根本限制】这些简化假设依赖于帧间亮度和外观的恒定性，限制了它们在高度动态场景中的适用性，例如 (i) 输入关键帧之间的非线性运动，(ii) 光照变化或运动模糊的情况，以及 (iii) 关键帧之间场景中出现的新物体和非刚性运动。</p> <p>具有不同的时空折衷 (spatio-temporal trade-offs) 的多相机方法克服这一限制，但存两份或多分同一个视频流（可能是高帧低分辨率和低帧高分辨率）会占用巨大的存储资源，视频流融合也需要更大的功耗</p> <p>单纯依赖事件相机(Events-only)的数据进行图像强度恢复不够完美，表现为1.插值帧的整体对比度通常会计算错误；2.插值结果强依赖于物体的运动情况（换言之对不运动的物体无法插值，也就无法完成VFI）</p> <p>依赖事件数据和强度帧数据(Events-plus-frames)的方法虽然能有效估计非线性运动，但由于对比度阈值和传感器噪声的固有不稳定性，并非所有的亮度变化都能准确地被记录为事件。</p>	Time Lens(CNN-based)	<ul style="list-style-type: none"> <li>从事件数据出发的基于变形(Warping-based)的插值方法</li> <li>引入“变形优化”(warping refinement)步骤，提高这些初始插值帧的质量</li> <li>基于合成的方法，利用两种信源(像素强度和事件流)融合生成新的帧</li> <li>最后利用基于注意力机制的平均化(attention-based averaging)模块组合两种方法的结果</li> </ul>	Project Demo	<p>每一个单独的插帧策略不执行，就两个一起用（有点大力出奇迹的味道）</p>

Citation	Motivations	Model	Strategies	Multimedia Resources	Other Info
2022 CVPR Time Lens++: Event-based Frame Interpolation with Parametric Non-linear Flow and Multi-scale Fusion	<p>结合帧和事件相机的视频帧插值 方法仍存在如下问题  <math>\curvearrowleft</math> 输入中存在伪影时，合成插值 的“残影”会被传递到最终输出中  <math>\curvearrowleft</math> 依赖于非参数运动估计，这一 过程独立运行于每个插入的帧， 计算成本为 <math>O(N)</math>，其中 <math>N</math> 是 插入的帧数，如果时间和对不上 就会产生估计偏差  <math>\curvearrowleft</math> 为利用非线性运动的信息，依 赖于仅基于事件的运动估计，这 在低对比度区域（没有事件）中 会导致伪影</p>	<b>❖ Multi-scale Feature Fusion Decoder</b> (CNN-based) <b>❖ Motion spline estimator</b> (Encoder-Decoder)	<ul style="list-style-type: none"> <li>◆ 新引入的运动估计算法 (Motion spline estimator), 除了利用事件外还使用边界图像</li> <li>◆ 改进事件尺度特征融合解码器，引入新的门控机制</li> <li>◆ Softmax Splatting 变型插帧</li> </ul>	<a href="#">Project Demo</a>	程序源代码未开源，仅数据集开源
2023 CVPR Event-Based Frame Interpolation with Ad-hoc Deblurring	<p>前人方法虽认识到了异步事件数据的好，但忽略如下问题  <math>\curvearrowleft</math> 运动是否会导致输入视频出现模糊，取决于帧的曝光时间长短 和运动速度；  <math>\curvearrowleft</math> 之前的工作要么假设输入视频是清晰的，仅限于帧插值，要么假设输入视频是模糊的，在插值之前包括一个显式的去模糊阶段；  <math>\curvearrowleft</math> 与帧不同，事件不受运动模糊的影响。无论帧是清晰还是模糊，相应的事件都是相同的。事件驱动的VFI的理想状态是能够在清晰图像插值和模糊图像插值上均表现出稳健的性能。  <math>\curvearrowleft</math> 我们提出将上述两个任务统一为一个问题：给定两个输入图像和对应的事件流，恢复在输入图像之间任意时间点的潜在清晰图像。输入图像可以是模糊的或清晰的。</p>	<b>❖ REFID</b> (CNN&RNN-based)	<ul style="list-style-type: none"> <li>◆ EGACA(event-guided adaptive channel attention)</li> <li>◆ bidirectional event recurrent (EVR) blocks</li> </ul>	<a href="#">Github Repo</a>	数据集未开源，仅程序开源

Citation	Motivations	Model	Strategies	Multimedia Resources	Other Info
2023 CVPR Event-based Video Frame Interpolation with Cross-Modal Asymmetric Bidirectional Motion Fields	<p>现有方法仅通过事件或近似方法估计双向帧间运动场，这无法考虑现实世界场景中的复杂运动。</p> <p>大多数基于 CNN 的帧合成方法由于受限的感受野大小，在长距离像素相关性上存在弱点。</p>	<b>EIF-BiOFNet</b> (直接估计帧间运动场，而无需任何近似方法) <b>基于交互注意的帧合成网络</b> (Transformer Decoder-based)	♦ 锚帧特征 预处理(Anchor Frame Feature Pre-processing) ♦ E-BiOF+I-BiOF+F-BiOF	<a href="#">Github</a> <a href="#">Repo</a>	相比前几篇的CVPR, 这一篇的含金量就略显逊色，只做一个数据集的结果
2023 ICRA Event-Based Video Frame Interpolation with Attention	<p>可变卷积 (deformable convolutions, DCN) 在合成中间帧方面的有效性也丰富了基于核的视频帧插值方法</p> <p>DCN+注意力机制可能有奇效</p>	<b>SmoothNet.</b> <b>Spatio-temporal absPooling layers.</b> <b>Parallel SynBlocks</b>	♦ Synblock主要负责事件流和RGB融合。 ♦ MSA用在计算时间维度的注意力值	<a href="#">Github</a> <a href="#">Repo</a>	没有关注插帧网络本身，侧重实际应用中面对不同场景的插针算法的适应性和鲁棒性；程序源代码未开源，仅数据集开源
2024 CVPR TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation	<p>事件相机数据的分布会根据相机设置和场景条件发生显著变化，弹性差</p> <p>在线适应网络环境在实际应用中至关重要-&gt;测试时适应 (Test-Time Adaptation, TTA)</p> <p>无高帧率视频的适应：网络在目标域中需要适应场景，同时从低帧率数据中学习高帧率生成的知识。</p> <p>在连续提供具有相似分布的数据时，网络可能会过拟合。</p>	<b>TTA-EVF Framework</b>	♦ Reliable Pixel Sampling (RPS) ♦ Patch-Mixed Sampling (PMS)	<a href="#">Github</a> <a href="#">Repo</a>	没有关注插帧网络本身，侧重实际应用中面对不同场景的插针算法的适应性和鲁棒性；程序源代码未开源，仅数据集开源

## Event-based VFI数据集

Citation	Multimedia Resources	Other Info
2017 CVPR Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring	DeepDeblur	GoPro相机拍摄, 无事件流
2020 CVPR Learning Event-Based Motion Deblurring	Blur-DVS	未公开
2021 CVPR Time Lens: Event-based Video Frame Interpolation	HS-ERGB	
2022 CVPR Time Lens++: Event-based Frame Interpolation with Parametric Non-linear Flow and Multi-scale Fusion	BS-ERGB	
2023 CVPR Event-Based Frame Interpolation with Ad-hoc Deblurring	HighREV	Because of the commercial reason, dataset download is allowed only with the authority of Alpsentek(锐思智芯®). 还在等作者回邮件中
2023 CVPR Event-based Video Frame Interpolation with Cross-Modal Asymmetric Bidirectional Motion Fields	ERF-X170FPS_train ERF-X170FPS_test	
2024 CVPR TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation	ERDS	

## Event-based VFI评价指标

1. PSNR:Peak Signal-to-Noise Ratio, 中文译为**峰值信噪比**

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

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right)$$

2. SSIM:Structural Similarity, 中文译为**结构相似性**

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$$

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