

中图法分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2026)06-2045-25

论文引用格式: Zheng Y J, Zhao R, Zhu L, Liu Y J and Huang T J. 2026. SpikeCV: a survey on the hierarchical modeling of continuous-time spike representations and systematic progress in neuromorphic vision. Journal of Image and Graphics, 31(6): 2045-2069 (郑雅菁, 赵睿, 朱林, 刘侯伽, 黄铁军. 2026. SpikeCV 脉冲视觉综述: 连续时间脉冲表征的层级建模与系统化进展. 中国图象图形学报, 31(6): 2045-2069) [DOI: 10.11834/jig.260128]

SpikeCV 脉冲视觉综述: 连续时间脉冲表征的 层级建模与系统化进展

郑雅菁¹, 赵睿², 朱林³, 刘侯伽¹, 黄铁军^{1*}

1. 北京大学计算机学院多媒体信息处理全国重点实验室, 北京 100871; 2. 南洋理工大学计算机与数据科学学院, 新加坡 639798, 新加坡; 3. 北京师范大学人工智能学院, 北京 100875

摘要: 随着神经形态视觉传感器技术的不断发展, 脉冲相机作为一种基于积分-发放机制的连续时间成像设备, 逐渐成为新型视觉感知的重要研究方向。不同于传统帧式图像, 脉冲相机以连续二值脉冲流形式记录光强变化, 具备超高时间分辨率、高动态范围及稀疏编码等特性, 在高速运动、复杂光照以及微弱变化检测等场景中展现出独特优势。然而, 其数据结构与统计分布显著区别于传统视觉数据, 使得经典视觉算法难以直接适配, 如何构建统一的连续时间表征与系统化算法体系成为该领域的重要研究问题。本文围绕连续时间脉冲表征的层级建模思想, 对近年来脉冲视觉领域的代表性工作进行了系统梳理与分类总结, 并提出从“可恢复”到“可理解”的脉冲视觉表征层级框架。该框架按照信息表征能力的提升, 将现有研究划分为多个层次: 层级0: 面向脉冲相机成像机理与噪声特性的物理一致建模; 层级1: 从脉冲流恢复稳定视觉信息的二维光强与纹理重建任务; 层级2: 刻画强度变化与运动信息关系的连续时间时空耦合表征; 层级3: 利用脉冲数据完成目标检测、语义分割与多目标跟踪等视觉理解任务的结构与语义感知; 层级4: 基于神经隐式表示的三维场景建模与动态重建; 层级5: 支撑脉冲视觉研究发展的系统化基础设施与平台化工具链, 包括数据集构建、模拟器设计、开源平台及统一评测协议等。在此基础上, 本文进一步讨论面向实际应用的系统评测问题, 从准确率、时延、吞吐率与能耗等多个维度总结现有方法的评估方式, 并综述脉冲视觉开放生态在数据资源、算法工具与社区协作方面的发展情况。通过统计截至2026年的相关研究成果, 并结合2025年举办的国际竞赛实例, 分析开放基准与社区平台在推动任务标准化与方法演进方面的作用。最后, 针对统一连续时间表征学习、多模态信息融合、系统指标驱动优化以及硬件-算法协同设计等关键问题进行了讨论, 为连续时间视觉智能的发展提供参考。

关键词: SpikeCV; 脉冲视觉; 连续时间表征; 神经形态视觉; 高速运动; 时空耦合建模; 开源社区生态

SpikeCV: a survey on the hierarchical modeling of continuous-time spike representations and systematic progress in neuromorphic vision

Zheng Yajing¹, Zhao Rui², Zhu Lin³, Liu Yujia¹, Huang Tiejun^{1*}

1. State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100871, China;

2. College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore;

3. School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

收稿日期: 2026-03-09; 修回日期: 2026-03-31; 预印本日期: 2026-04-01

* 通信作者: 黄铁军 tjhuang@pku.edu.cn

基金项目: 国家自然科学基金项目(62506009); 北京市自然科学基金项目(4252039); 中国博士后科学基金(2023T160015)

Supported by: National Natural Science Foundation of China (62506009); Beijing Municipal Natural Science Foundation (4252039); China Postdoctoral Science Foundation (2023T160015)

Abstract: With the rapid development of neuromorphic vision sensors, spike cameras have emerged as a promising paradigm for continuous-time visual perception. In contrast with conventional frame-based cameras that sample scenes at fixed frame rates, spike cameras encode luminance variations as asynchronous binary spike streams triggered by intensity accumulation at each pixel. This sensing mechanism changes the manner in which visual information is acquired and represented. Instead of producing discrete image frames, spike cameras generate continuous spike events that directly reflect temporal changes in scene intensity. Consequently, spike cameras provide several unique advantages, including extremely high temporal resolution, wide dynamic range, low motion blur, and sparse event-driven representations. These properties make spike cameras particularly suitable for challenging visual environments, such as high-speed motion analysis, extreme illumination conditions, and subtle temporal change detection, where traditional imaging systems frequently encounter limitations. Spike vision also introduces new challenges for visual computing. The statistical characteristics and data structures of spike streams differ significantly from those of conventional images or videos. Spike cameras produce sparse binary spike sequences in continuous time rather than dense intensity frames. Therefore, many established computer vision algorithms cannot be directly applied to spike data without modifications. Effective processing of spike streams requires new modeling strategies that explicitly consider the temporal dynamics and sparsity of signals. This survey reviews recent progress in spike vision research from the perspective of the hierarchical modeling of continuous-time spike representations. Existing methods are organized into several levels that reflect the progressive expansion of spike vision capabilities, ranging from signal modeling and reconstruction to semantic perception and system deployment. At the lowest level, the physically consistent modeling of spike generation and sensor noise provides a foundation for understanding spike data statistics. Studies in this direction analyze pixel triggering mechanisms, noise characteristics, and spike accumulation processes, forming the basis for reliable signal processing and algorithm design. Low-level visual reconstruction methods aim to recover stable visual signals from spike streams. Representative tasks include intensity reconstruction, high dynamic range imaging, motion deblurring, super-resolution, and low-light enhancement. These approaches convert spike sequences into interpretable intensity representations while preserving the temporal information contained in the spike data. The next level focuses on spatiotemporal modeling. The continuous-time nature of spike streams enables joint modeling of spatial structure and temporal motion. Research in this area addresses problems, such as optical flow estimation, motion segmentation, and dynamic scene analysis. Compared with frame-based methods, spike-based models provide improved temporal fidelity in fast motion scenarios. At the semantic perception level, spike representations are increasingly applied to tasks, such as object detection, recognition, and multi-object tracking. Continuous spike streams are integrated with deep neural networks, Transformer architectures, or spiking neural networks to perform higher-level visual reasoning. These methods utilize the temporal sparsity of spike data while maintaining low-latency processing. Spike cameras have also been introduced into 3D scene modeling. Recent studies combine spike streams with neural implicit representations to reconstruct static or dynamic 3D scenes. Continuous spike measurements provide detailed temporal information that can benefit dynamic scene reconstruction and neural rendering. System-level considerations play an important role in practical spike vision deployment. The evaluation of spike-based methods not only involves accuracy but also system metrics, such as latency, throughput, and energy consumption. These metrics become particularly important in real-time perception systems and edge computing scenarios. Progress in spike vision has also been supported by the development of datasets, simulation tools, and open-source platforms. Collecting spike camera data can be costly, and thus, spike simulators are widely used in algorithm development and validation. Simulation methods attempt to reproduce sensor physics and temporal spike generation processes. Public datasets and benchmarking protocols further support reproducible research. The SpikeCV platform provides a unified open-source framework for spike vision research. It integrates datasets, algorithm implementations, hardware interfaces, and evaluation tools, allowing researchers to prototype and evaluate spike-based algorithms rapidly. The platform has helped facilitate collaborative development and reproducible experiments within the community. Research activity in spike vision has grown rapidly in recent years. Publications, open-source resources, and benchmark datasets have increased steadily. Two international competitions organized in 2025 attracted wide participation from academic institutions and industry teams. These competitions encouraged standardized task definitions and stimulated methodological progress in the field. However, continuous-time representation learning for spike streams is still an active research area. Large-scale

self-supervised learning for spike data remains largely unexplored. Multimodal fusion with complementary sensors introduces additional challenges into temporal alignment and noise modeling. System-level optimization that involves latency, throughput, and energy consumption also requires further investigation. Hardware-algorithm codesign with neuromorphic processors may provide new opportunities for efficient spike-based computation. Spike vision represents an emerging direction that reconsiders visual perception from a continuous-time perspective. Advances in sensor technology, representation learning, and system integration are gradually forming a new framework for visual computing. Continued progress in spike-based sensing, modeling, and deployment may enable high-speed, energy-efficient visual intelligence for future perception systems. Overall, this work provides a structured overview of spike vision research from the perspectives of sensing principles, representation modeling, perception algorithms, and system-level infrastructure. By organizing existing studies within a hierarchical framework of continuous-time spike representations, this study highlights how spike vision methods evolve from low-level signal reconstruction to high-level semantic perception and 3D scene understanding. The discussion of datasets, simulators, evaluation protocols, and open-source platforms further reflects the growing research ecosystem that surrounds spike-based vision. Through this synthesis, we aim to clarify the relationships among different research directions, summarize the current development status of the field, and provide a reference for future work on continuous-time visual computing and neuromorphic perception systems.

Key words: SpikeCV; spike vision; continuous-time representation; neuromorphic vision; high-speed motion; spatiotemporal modeling; open-source ecosystem

0 引言

在高速运动或连续变化的视觉场景中,视觉系统对时间维度的表达能力往往成为影响感知性能的重要因素。传统帧式相机通过在固定曝光时间内对光强进行积分并周期性输出图像帧,从而实现对视觉信息的采样。在动态场景中,这种采样机制可能带来两个典型现象:1)当物体运动速度较高时,曝光时间内的积分过程可能导致纹理细节被平均化,从而产生一定程度的运动模糊;2)受限于相机帧率,连续变化的视觉过程只能以离散时间点进行采样,大量中间信息未被显式记录。尽管通过高速相机、短曝光成像或计算摄影等技术手段可以在一定程度上缓解上述问题,但在实际应用中,这类方案往往伴随着较高的硬件成本、数据带宽需求以及存储与计算负担。因此,如何在保持高时间分辨率的同时实现更加高效的数据表达,仍然是动态视觉感知中的一个重要研究方向。

脉冲相机(spike camera)(Huang等,2023)为上述问题提供了一种新的技术路径。不同于传统帧式采样方式,脉冲相机采用基于阈值触发的信号生成机制:每个像素持续对入射光子进行积分,当累积量达到预设阈值时触发一次脉冲发放,从而形成时间连续的脉冲序列输出。由于各像素以异步方式独立

工作,该机制能够在保持较高动态范围的同时实现极高的时间分辨率。更重要的是,系统输出的数据形式不再是离散图像帧,而是连续时间上的脉冲流,这为视觉信息的时空建模提供了新的表达空间。基于这一特点,脉冲视觉逐渐被视为一种具有潜力的连续时间视觉计算范式,其核心特征可以概括为连续时间表示、稀疏事件表达以及高动态范围感知(黄铁军,2022)。

为了降低脉冲视觉研究的技术门槛并促进相关方法的系统化发展,近年来提出了 SpikeCV 开源平台(Zheng等,2026b)。该平台面向脉冲相机数据特点,构建了包含数据解码、算法模块、数据集管理以及硬件接口在内的完整工具链,旨在为研究者提供统一的实验环境与开发框架。与传统视觉工具库不同,SpikeCV 针对脉冲数据的时序结构和实时处理需求进行了专门设计,在数据封装、数据集接口标准化以及视觉任务模块化等方面提供了较为完善的支持。同时,平台还集成了多种高速场景数据集以及基础算法实现,为脉冲视觉算法的复现与比较提供了便利条件。

在此基础上,随着开源工具链的不断完善,脉冲视觉领域逐渐形成了更加活跃的研究生态。围绕脉冲相机的成像建模、视觉重建以及高层感知任务,近年来已有大量研究工作不断涌现。同时,以 SpikeCV 为基础的技术竞赛与开放挑战也吸引了来

自不同研究机构的团队参与,在一定程度上推动了算法方法、数据集构建以及系统应用的发展。这些进展表明,脉冲视觉正在从早期的单点技术探索逐步走向更加系统化和社区化的发展阶段。在这一背景下,对脉冲视觉近年来的发展脉络进行系统梳理,对于理解该领域的研究进展及其未来方向具有重要意义。

尽管已有研究已经从不同侧面对脉冲视觉进行了总结,但其组织方式大多仍围绕特定应用场景、单一任务链路或原理性介绍展开:例如,面向自动驾驶的综述(张济远等,2024)主要聚焦驾驶场景中的感知任务与系统挑战,强调重建、融合、运动估计、目标感知和三维理解等方向;仿视网膜传感器相关综述(朱林和田永鸿,2023)更多从传感器采样原理与视觉重建的角度切入,同时兼顾事件相机与脉冲相机;而关于脉冲连续摄影与脉冲视觉研究进展的综述(黄铁军,2022;黄铁军等,2022),则更侧重成像机理、硬件实现及整体发展脉络。这类综述为理解脉冲视觉的发展提供了重要基础,但在“任务能力如何分层组织、方法应如何统一建模、性能瓶颈究竟受哪些系统约束主导”这类更面向方法论的问题上,仍缺乏一个兼顾表示层次与系统代价的统一分析框架。

针对上述问题,本文从表征层级与系统指标两个维度对脉冲视觉研究进行系统梳理。一方面,以连续时间脉冲表征为主线,分析脉冲视觉从底层信号恢复、时空信息建模到高层语义感知与三维场景表示的逐级发展脉络;另一方面,将准确性、时延、吞吐率以及能耗等系统指标引入方法分析之中,以讨论不同算法路线背后的性能权衡关系及其系统实现约束。通过这一视角,可以更加清晰地理解脉冲视觉研究在不同任务层级上的演进路径,以及其从算法设计走向系统落地所面临的关键问题。

在结构安排上,本文首先介绍脉冲相机的成像原理、数据表示形式及其在动态视觉建模中的特点。随后,围绕从底层视觉恢复到高层语义理解的任务层次,系统梳理辅助成像、连续时间时空耦合建模、高层感知任务以及三维表示方法等方面的研究进展。在此基础上,进一步总结脉冲视觉在数据集构建、模拟器设计与评测协议方面的发展,并介绍 SpikeCV 平台在工具链构建与开源生态建设中的作用。最后,本文讨论当前脉冲视觉研究仍面临的关键挑战,并对连续时间视觉计算的发展趋势进行

展望。

1 脉冲相机原理

1.1 积分发放—成像机理

随着自动驾驶、无人机、工程实验和专业摄影等应用的迅速发展,对超高速、高动态和高清晰的摄影需求日益增长(Sick,2013;Bub等,2010)。传统快门曝光摄影原理采用定时测量方法,即在固定的曝光时间累积光子,并利用曝光时间内累积的光子数与光强的正比关系估计光强(Weckler,1967)。然而,在拍摄高速运动的目标时,由于曝光时间内目标与成像单元之间存在相对位移,图像容易产生运动模糊(Nayar和Ben-Ezra,2004)。为了减轻运动模糊,通常的策略是提高快门速度,使曝光时间远小于相邻两帧的采样间隔。但这一做法不仅造成了非曝光期间的光学信息丢失,也因光子积分时间的急剧缩短而显著限制了动态范围(Zhu等,2019)。由此可见,快门原理对光子信息的采集本质上是离散性的,无法实现对光过程的连续记录(Adelson和Bergen,1991)。这一局限限制了其在超高速、高动态和高清晰成像任务中的应用潜力。

为满足超高速、高动态与高清晰成像的综合需求,黄铁军(2022)发明了脉冲连续摄影原理。该原理采用定量测量方法,通过为光子累积总数设定固定阈值,对成像单元上的入射光子流连续采集。而后在光子累积总数达到设定阈值时立即发放一个脉冲作为事件标记,并同步重置计数器以重新开始累积。其具体表达式为

$$\int_0^t I(\tau) d\tau \geq Q \quad (1)$$

式中, $I(\cdot)$ 为光电流, Q 为预定义的发放阈值, t 为时间变量, τ 为积分变量。当式(1)成立时,即发放一个脉冲。

通过测量相邻脉冲之间的时间间隔(或称脉冲宽度),并利用该间隔与光强之间的反比关系实现对光强的估计。脉冲原理将入射光子流转换为脉冲序列进行表征。由于不再依赖固定的曝光时间窗口,其在原理上避免了信息丢失,从而能够实现对光子动态过程的连续刻画。与此同时,该方法的动态范围仅由系统可测量的最短与最长时间间隔所决定,因而能够轻松实现远超快门原理的高动态范围成

像。目前,脉冲原理已在多个典型的超高速、高动态和高清晰成像场景中展现出显著优势。例如,在高铁交汇、气球爆炸和风扇旋转等瞬态过程的观测中(Zheng等,2026b),该原理能够有效抑制运动模糊,精准捕捉动态细节与瞬时演化行为(Xia等,2023; Fan等,2024)。

1.2 脉冲数据结构与关键时域特征

对于单一像素 p ,其脉冲流可视为一个连续的二值时间序列,在任意时刻 t 的取值为1或0,分别表示该时刻有或者无脉冲发放。进一步将该定义拓展到尺寸为 $H \times W$ 的像素阵列,则每个像素 p 在时刻 t 的脉冲流取值可以由空间坐标 (x, y) 和时间信息 (x, y) 共同描述,构成三元组 (x, y, t) 。所有像素在时间维度上的脉冲序列共同构成一个尺寸为 $H \times W \times T$ 的三维体素,其值可以用集合 $\{(x, y, t)\}$ 表示。

在此基础上,脉冲流中有若干时域特征对于后续算法设计尤为重要。首先是首脉冲发放时间(time to first spike, TTFS),即从计时起点到首个脉冲产生的时间间隔。由于脉冲相机的发放过程与光强累积密切相关,在一些快速感知任务中,仅利用首脉冲时间便可对局部光强进行近似估计,从而降低计算开销。其次是脉冲间隔(inter-Spike interval, ISI),即相邻两个脉冲之间的时间间隔,它能够刻画脉冲序列的局部时域结构,对运动变化和精细时间信息较为敏感。再次是脉冲发放频率,即单位时间窗口内的脉冲计数,它反映了该时间段内像素响应的整体活跃程度,也常用于估计局部亮度强弱。

TTFS、ISI与发放频率并不构成彼此独立的编码体系,而是脉冲流中最常用的几类时域描述量。它们既反映了脉冲相机输出的连续时间特征,也为后续的图像重建、运动估计与高层视觉任务提供了重要基础。特别地,基于ISI与发放频率的建模思想,已在脉冲图像重建等任务中得到广泛应用,相关方法将在第3.1节进一步介绍。

1.3 时间编码带来的计算收益

尽管理论上脉冲流是连续的,但在实际应用中需对其时间维度进行离散化处理,即在每个极小的采样间隔内记录脉冲的发放情况。得益于当前市场时钟精度可达亚微秒级(<https://spikesee.com/product.html>),脉冲流具备了极高的时间分辨率,为感知与计算带来了多重可量化的收益。

首先,在高清摄影领域,脉冲视觉展现出显著的成像优势,其收益直接体现于重构图像的清晰度。极小的脉冲采样间隔等效于极短的曝光时间,使得拍摄高速运动物体或场景时,能够有效避免运动模糊的产生。这种由高时间分辨率换取的图像清晰度,为下游的去模糊等视觉任务提供了更强的先验信息,从而在相同物理条件下,实现比传统摄影原理更优的成像质量。

其次,脉冲流对微弱变化的捕捉能力较强。基于脉冲视觉的积分比较机制,系统对光强变化具有高度敏感性,微小的环境光变化会被“放大”为可观测的脉冲发放差异,进而被及时捕捉。这意味着,对于弱光或红外波段等难以分辨的微弱信号,脉冲视觉也能通过脉冲发放规律的变化而有效感知。这种以时间分辨率换取的灵敏度,显著提升了系统对细微动态的检测能力,使其能够精准捕捉微弱而瞬变的物理过程,在红外、小目标等变化检测领域展现出广阔的应用前景。

最后,脉冲视觉对光强变化的高度敏感性,使其能够有效缩短交互距离,提升系统的安全裕度。基于前述的微弱变化捕捉能力,在同等运动速度下,脉冲摄影相比于传统摄影可对环境光强变化做出更快速的反应。这一特性在自动驾驶领域中,等效于缩短了车辆与潜在危险之间的有效交互距离,从而能够更及时地应对“鬼探头”等高速目标的突然出现。因此,脉冲视觉在自动驾驶、机器人避障等智能系统中具有重要的应用价值。

1.4 关键挑战

脉冲流带来的首要挑战在于其高时间分辨率所衍生的庞大数据量。以 $50 \mu\text{s}$ 的采样间隔计算,单个像素每秒将产生10 000比特的数据。随着像素阵列规模的扩大,总数据量指数级增长,达到十分可观的规模。如此海量的数据对系统的计算吞吐能力构成严峻考验,同时也对数据传输带宽提出了极高要求。因此,如何高效编码脉冲流以压缩数据量,是一个值得深入探索的方向。此外,在极端低光场景下,脉冲发放变得极为稀疏,此时脉冲流的稀疏编码同样具有重要意义。

脉冲数据的获取与标注同样构成一项严峻挑战。作为一种区别于传统成像的全新视觉感知原理,脉冲视觉在数据形式上与常规图像数据存在本质差异,现有的成熟视觉数据集与标注范式难以直

接复用。因此,面向脉冲视觉的研究与应用,必须从零构建专用数据集体系,包括设计适配脉冲特性的模拟或采集方案、制定统一的数据格式与标注规范、建立覆盖不同场景与任务的高质量样本库。这一过程不仅耗时费力,对硬件平台、采集环境与人工标注成本也提出了更高要求,成为脉冲视觉面临的严峻挑战之一。

此外,脉冲相机的噪声分析同样是一项关键挑战。由于脉冲视觉在感知机理上与传统成像存在本质差异,必须设计专用芯片与电路来实现脉冲信号的产生与读取,其噪声来源、分布特性均无法直接沿用传统图像传感器的分析模型。因此,针对脉冲成像硬件的噪声建模与量化分析,直接关系到信号还原精度与最终成像质量。同时,脉冲数据具有高时间分辨率与时序编码特性,光子到达过程本身呈现泊松随机分布,由此带来的ISI波动、累积器重置噪声等时序噪声的建模、评估与抑制,也是脉冲成像系统走向实用化必须重点解决的核心问题。

2 从“可恢复”到“可理解”的表征层级

围绕连续时间脉冲数据的视觉计算研究,近年

来逐渐形成了从低层信号处理到高层语义理解的多层次研究格局。整体来看,相关工作大致沿着一条由低层视觉信息恢复逐步扩展到高层视觉理解的发展路径展开。

早期研究主要关注如何从脉冲流中恢复稳定的视觉信息,例如亮度重建、去模糊以及高动态范围成像等问题。随着对连续时间信息利用能力的提升,研究进一步扩展到运动建模与时空结构分析,例如光流估计与运动分割等任务。在此基础上,越来越多的工作开始探索利用脉冲数据完成目标检测(Liu等,2025c)、目标跟踪以及场景理解等高层视觉任务。近年来,一些研究进一步将脉冲视觉引入三维场景表示与动态重建问题,并逐步形成了围绕数据集构建、模拟器设计以及开源工具链等方面的研究基础设施。

从视觉表征能力逐步提升的角度,现有研究可以概括为6个表征层级(层级0—层级5)。这种层级划分并非严格的任务分类,而是一种用于理解不同研究方向在视觉信息建模能力上逐步扩展关系的分析视角。表1进一步汇总了各层级所涉及的主要研究问题、典型任务以及代表性工作。

层级0:物理一致建模。该层级主要关注脉冲

表1 脉冲视觉研究的表征层级及代表工作概览

Table 1 Representation hierarchy of spike vision research and representative works

表征层级	主要研究问题	典型任务	代表研究方向	代表工作	研究成熟度	对应章节
层级0: 物理一致建模	脉冲相机成像机理与数据生成过程建模	像素触发机制建模、噪声建模、数据表示建模	传感器物理模型、统计建模	(Huang等,2023; Zhao等,2022a;Liu等,2025a)	基础研究方向	第3节
层级1: 光强与纹理重建	从脉冲流恢复稳定视觉信息	图像重建、高动态范围成像、运动去模糊、低照度增强	脉冲辅助成像、多模态重建	(Zhao等,2021, 2024c;Zheng等, 2021;Zhang等, 2023)	较成熟方向	第3节
层级2: 时空耦合表征	连续时间动态信息建模	光流估计、运动补偿、运动分割	连续时间视觉建模、时空对齐	(Hu等,2022;Zhao等,2022d,2024b; Zhao等,2026)	快速发展	第4节
层级3: 结构与语义感知	利用脉冲数据完成高层视觉理解	目标检测、语义分割、多目标跟踪	深度学习感知模型、脉冲神经网络感知	(Zhu等,2024a; Zheng等,2023a, 2026a)	研究热点	第5节
层级4: 三维表示	利用脉冲数据进行三维场景建模	三维重建、新视角合成、动态场景表示	神经渲染、连续时间三维建模	(Zhu等,2024b; Zhang等,2024a; Guo等,2025)	新兴方向	第6节
层级5: 系统化与平台化	构建研究基础设施与工具链	数据集构建、模拟器设计、平台工具链	开源平台、评测体系	(Zheng等,2026b)	逐步发展	第7-8节

相机的成像机理及其数据生成过程,包括像素触发机制、噪声特性以及脉冲数据表示方式等问题。相关研究通常从传感器物理模型或统计建模角度出发,对脉冲数据生成过程进行描述,为后续视觉算法提供更加合理的数据建模基础。

层级1: 光强与纹理重建。在理解脉冲数据生成机制的基础上,大量研究致力于从脉冲流中恢复稳定的视觉信息,例如图像重建、高动态范围成像、运动去模糊以及低照度增强等。这类方法通常将脉冲流视为一种具有高时间分辨率的视觉信号来源,从而与传统图像恢复方法形成互补。

层级2: 时空耦合表征。由于脉冲数据具有连续时间特性,相关研究进一步关注如何在统一框架下对空间结构与时间动态进行联合建模,例如光流估计、运动补偿以及运动分割等问题。该层级的研究重点在于利用脉冲流中的时间信息刻画场景中的动态变化过程。

层级3: 结构与语义感知。随着数据表示与运动建模方法的发展,越来越多的工作开始探索利用脉冲数据完成高层视觉任务,包括目标检测、实例分割以及多目标跟踪等。这类研究尝试将脉冲视觉方法与现有视觉感知框架结合,从而实现了对复杂场景的语义理解。

层级4: 三维表示。近年来,一些研究进一步将脉冲数据引入三维场景建模问题,例如三维重建、新视角合成以及动态场景表示等方向。脉冲数据的高时间分辨率为动态场景的几何建模提供了新的信息来源,在高速运动或光照变化场景中展现出一定潜力。

层级5: 系统化与平台化。随着研究方向的不断拓展,围绕数据集构建、模拟器设计、评测协议以及开源工具链等方面的工作逐渐出现。这些研究为算法开发、实验复现以及社区协作提供了重要基础支撑,有助于推动脉冲视觉研究的系统化发展。

上述层级之间并不存在严格边界,不同研究方向之间往往存在交叉与融合。例如,三维场景建模通常依赖于低层视觉恢复或运动估计结果,而高层视觉任务也可能直接利用脉冲流中的连续时间信息进行建模。因此,这种层级划分更多是一种用于理解研究进展整体图景的分析视角。

在这一总体框架下,后续章节将分别对各研究方向进行系统综述。

3 底层视觉与辅助成像

3.1 经典重建方法

脉冲相机的原始输出是离散二值脉冲序列。由于单个脉冲只表征“累积达到阈值并触发发放”这一事件,而不直接给出连续灰度值,因此早期重建方法的核心任务,是利用脉冲在时间轴上的统计规律反推目标时刻的光强。围绕这一目标,经典方法大致经历了从单像素时域估计到显式运动补偿聚合的演进过程:前者以窗口脉冲计数法(texture from playback, TFP)(Zhu等,2019)和脉冲间隔法(texture from interval, TFI)(Zhu等,2019)为代表,后者则进一步引入运动对齐,将不同时间的有效观测沿目标像素对应的运动轨迹进行汇聚,以减轻噪声与运动模糊之间的冲突。

在像素时域估计方法中,窗口脉冲计数法(TFP)(Zhu等,2019)是最经典的方法之一。其在待重建时刻附近截取一个局部时间窗口,统计该窗口内每个像素的脉冲发放次数,并以此计算脉冲发放频率,进而近似该时刻的亮度。由于窗口内多个脉冲被累加,随机波动会被部分平均,因此该方法对噪声具有一定抑制作用,形式简单、实现直接,常被视为最早期的基线之一。但它的局限也很明显:窗口一旦变长,虽然统计更稳定,却会将不同时间对应的不同场景内容混合起来;在高速运动区域,这种简单累加容易造成明显拖影。换言之,TFP用更强的时间积分换取了更稳的估计,但牺牲了动态区域的瞬时清晰度。

另一个经典的早期像素时域估计方法为脉冲间隔法(TFI)(Zhu等,2019)。其直接利用目标时刻前后最近两次脉冲之间的时间间隔估计光强。其基本出发点是:若某像素接收到的光子更多,则更快达到发放阈值,脉冲间隔通常更短,因此,光强可近似看做与脉冲间隔成反比。相比窗口计数,TFI更贴近瞬时响应,也更不容易因为长窗口平均而引入额外模糊。但由于它高度依赖极少数脉冲的位置,一旦存在脉冲抖动、漏发或误发,重建值则容易出现较大起伏,因此在噪声较强或低照稀疏条件下稳定性往往不足。

TFP与TFI均将各像素视为相互独立,缺乏对跨时刻、跨像素关联关系的建模。为此,后续方法通

过显式估计运动,将相邻时刻中对应同一场景点的观测进行对齐后再聚合。运动对齐滤波(motion aligned filtering, MAF)(Zhao等,2020)是该思路的早期代表:首先通过光流估计建立待重建时刻与相邻时刻间的对应关系,再沿估计的运动轨迹将相邻时刻信号迁移至当前时刻对应位置,并通过加权滤波完成融合。该方式使时间维度多帧信息不再直接叠加,而是经运动补偿后再累积,在保持较强去噪能力的同时,可有效缓解高速场景下的错位模糊。从方法论角度来看,MAF的价值不仅在于提升重建质量,更在于将脉冲重建从单像素统计方法推进至基于运动对齐的时空聚合范式。

在复杂场景下,仅依赖短邻域滤波仍存在局限:有效脉冲数量显著减少,单次局部聚合难以实现稳定估计。针对该问题,后续方法MAHTF(motion-aligned hierarchical temporal filtering)(Zhao等,2022c)沿“先对齐、再聚合”思路进一步发展:通过光流估计稀疏脉冲间的运动关系,利用递归时间滤波实现更长时间范围的信息累积,并结合亮度自适应空间滤波实现噪声抑制。相较于早期MAF方法,此类方法不再局限于固定邻域的一次性聚合,而是通过自回归方式将历史信息逐步传递至当前时刻,更适用于低光照条件下观测信息稀缺、单次观测不稳定的场景。从方法脉络来看,其仍属于经典“显式对齐+显式聚合”范式,仅将聚合范围从短时扩展至长时范畴。

经典重建方法的主线可概括为:先通过像素级时间统计实现光强初步估计,再借助运动补偿引入时序一致性约束,逐步提升时空信息利用效率。其中,TFP与TFI奠定了脉冲强度恢复的基本框架;MAF明确了沿运动轨迹聚合的核心思路;递归长时聚合则将经典方法从简单滤波推向更充分的时空上下文建模。后续鲁棒重建与学习类方法虽采用更复杂网络结构,本质上仍延续了对齐—聚合—恢复的核心逻辑。

除基于时间统计与显式运动补偿的经典灰度重建路线外,研究者还从“神经元机制模拟”和“端到端学习”两条方向拓展了基础重建范式:前者借助生物突触可塑性与神经元动力学实现时间自适应积分与动静分离,后者则直接以数据驱动方式学习脉冲流到光强图像的非线性映射与时空对齐策略。

在生物神经元机制模拟方法方面,Zhu等人

(2020)提出的脉冲神经元模型(spiking neural model, SNM)以脉冲时间依赖可塑性(spike-timing-dependent plasticity, STDP)规则(Bi和Poo,1998)为核心,通过“运动局部兴奋—脉冲时序优化—视觉重建”的分层结构刻画脉冲时序相关性:其先进行动静区域划分,再基于LIF(leaky integrate-and-fire)神经元的积分特性自适应调节不应期,使静态区域更强调噪声抑制、运动区域更侧重纹理细节保留,最后以动态阈值机制融合得到重建结果。Zheng等人(2021)提出的TFSTP(texture from short-time plasticity)引入短时可塑性(short-time plasticity, STP)机制(Tsodyks等,1998),将突触传递效率对脉冲频率的短时依赖显式纳入建模:当输入呈高频快速变化时产生适应性衰减,而在缓慢变化段保持更稳定的响应,从而形成对时空信息的自适应编码与聚合。随着深度脉冲神经网络的发展,Zhao等人(2024a)将基于时序反向传播与代理梯度法进行训练的深度神经网络引入脉冲相机的图像重建中,实现了时间连续的重建。

在端到端深度学习方面,Spk2ImgNet(Zhao等,2021)是较早将“光强推断—特征抽取—对齐与融合”纳入统一网络的代表,其框架包含面向光强推断的前端模块、深层特征抽取,以及基于金字塔可变形卷积的跨时刻特征对齐与可靠性引导融合,从而在同一网络内兼顾短时细节与长时上下文信息的利用。WGSE(wavelet-guided spike enhancing)(Zhang等,2023)将多尺度离散小波变换引入脉冲表征学习:先将脉冲流分解为刻画瞬态变化的高频系数与体现整体趋势的低频系数,再用时序残差卷积网络分别优化各尺度系数,最终通过逆小波变换合成兼具时域与频域特征的特征,以更细致地刻画离散脉冲的时序结构。另一类工作则针对脉冲噪声的统计特性进行专门建模,如BSF(boosting spike camera image reconstruction from a perspective of dealing with fluctuations)(Zhao等,2024c)讨论了在光照近似稳定时脉冲发放间隔仍会波动的问题,并由此构造基于脉冲发放的时间差分(differential of spike firing time, DSFT)的无偏光强估计;同时通过多阶DSFT融合在动态敏感性与长时稳定性之间取得平衡,并在对齐阶段结合块级交叉注意力与可变形卷积的多粒度校正以提升匹配鲁棒性。

除此之外,研究者也对基于颜色滤镜阵列

(color filter array, CFA)的彩色脉冲相机开展了图像重建研究。由于CFA使每个像素仅记录单一颜色通道的脉冲响应,彩色重建通常需要将去马赛克与噪声抑制,以及在动态场景下进一步结合时序对齐进行联合处理。围绕该问题,Dong等人(2024a,b,c)提出了一系列代表性方法:从三阶段彩色重建框架CSpkNet(Dong等,2024b),到将去马赛克与去噪耦合并采用优化展开的多阶段网络SJDD-Net(spike-based joint demosaicing and denoising)(Dong等,2024a),再到联合超分辨率以增强空间细节CSCSR(color spike camera super-resolution)(Dong等,2024c)。进一步地,其还探索了面向真实数据域差异的自监督学习策略(Dong等,2025b)和无需预训练的零样本重建方案(Dong等,2025a),以提升跨场景适应性。

3.2 复杂条件下的鲁棒重建

脉冲相机独特的采样机制使其能够根据光照强度自适应地调整等效曝光时间。如图1所示,在高光条件下,密集的脉冲流能够记录更丰富的场景信息,从而实现更可靠的重建。然而,真实环境并不总是具备理想照明条件,这使得低光照条件下鲁棒的脉冲重建具有重要的现实意义。随着光强降低,脉冲流逐渐变得稀疏,可用于表征场景的信息量显著减少,在有限的时间窗口内甚至出现信息缺失。同时,不同脉冲噪声成分的相对影响也随着光照变化,高光条件下以量化噪声为主,而在低光条件下固定模式噪声(fixed pattern noise, FPN)和截断噪声的影响更加突出。因此,有必要针对低光脉冲流的稀疏性和噪声等问题,设计专门的脉冲鲁棒重建方法。已有的方法按光照条件区分可以分为如下3类:

1)较低光照场景。针对较低光照场景的脉冲重建方法主要通过循环来聚合稀疏的脉冲信息和去噪。Dong等人(2022)通过光流预测稀疏脉冲之间的运动,并结合循环时间滤波和亮度自适应空间滤波实现噪声抑制。Hu等人(2025a)提出光强鲁棒表示来聚合稀疏脉冲信息,并利用双向循环网络融合时域特征。

2)通用光照场景。针对通用光照场景的脉冲重建方法通过构建统一的脉冲噪声模型,实现跨高光与低光条件的通用去噪。RSIR(recurrent spike-based image restoration)(Zhu等,2023)对通用光照下的脉冲噪声进行统一的建模,综合考虑散粒噪声、暗

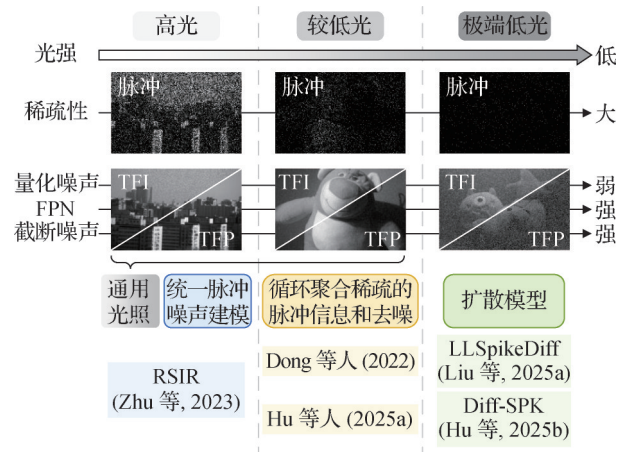


图1 不同光照条件下脉冲流的鲁棒重建

Fig. 1 Robust reconstruction of spike streams under different illumination conditions

电流噪声、响应不一致噪声、量化噪声以及截断噪声,并在此基础上设计循环网络,其中的去噪模块分别针对特定噪声成分设计,从而实现覆盖高光和低光条件的通用复原。

3)极端低光照场景。针对极端低光照场景的脉冲重建方法主要利用扩散模型的生成能力,从低信噪比且存在信息缺失的脉冲流中恢复出干净、完整的图像。LLSpikeDiff(Liu等,2025a)将扩散过程建模为低光脉冲退化过程,以帮助模型捕捉脉冲噪声的结构特性,并引入退化提示作为模型条件引导生成。Diff-SPK(Hu等,2025b)设计自监督纹理增强模块以获得聚合稀疏信息的脉冲表示,并设计编码器和特征融合模块,将编码后的脉冲表示作为Control-Net的输入,同时利用融合后的特征对生成过程进行调控。

3.3 复杂条件下的鲁棒重建

在基础重建与鲁棒重建之外,如图2所示,脉冲相机还常与帧式相机等传感器形成协同成像系统,以满足更高阶的成像目标。其动机在于不同传感器在信息维度上的互补性:帧式相机能够提供较高空间分辨率与完整纹理/颜色信息,但在强弱光并存及高速运动条件下易受曝光积分与动态范围限制;脉冲相机以连续时间脉冲流记录光强变化过程,能够为运动过程刻画、跨时刻一致性约束以及极端光照条件下的有效观测提供线索。基于这种互补关系,研究者提出了若干典型的辅助成像任务,包括高动态范围(high dynamic range, HDR)辅助成像、去模糊辅助成像等方向。

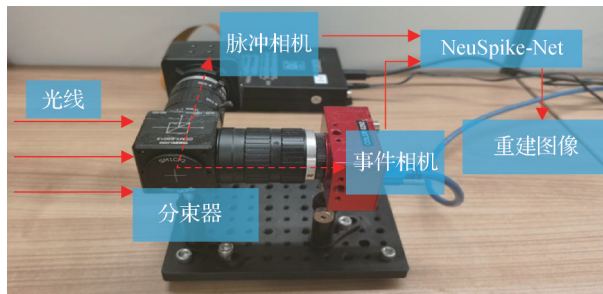


图2 脉冲相机与事件相机的协同成像系统
(Zhu等, 2021)

Fig. 2 Collaborative imaging system composed of a spike camera and an event camera (Zhu et al., 2021)

3.3.1 高动态范围(HDR)辅助成像

高动态范围成像旨在同时保持高亮区域的细节不过曝与暗部区域的结构可辨识度。脉冲相机的阈值触发机制与短时间尺度的响应特性,使其在高照度条件下更不易出现长曝光导致的显著饱和和拖尾;同时,暗部区域可通过更长时间范围的有效脉冲积累获得一定观测支撑。然而,在实际读出机制与阈值设置下,强光区域仍可能出现脉冲过密导致的有效信息截断,而低照区域则可能因脉冲稀疏使估计方差增大,从而难以在单一链路内兼顾亮暗两端的稳定重建。为进一步扩展动态范围并提升细节保真度,现有研究多采用多传感器联合策略:一方面,将脉冲相机与帧式相机配合,由脉冲流提供宽亮度变化下的时序证据,再以帧式图像补足高分辨率纹理与颜色信息(Chang等, 2024);另一方面,将脉冲相机与事件相机结合,利用事件在极端亮度变化下的响应优势扩展可观测范围,同时由脉冲相机提供更完整的结构与纹理约束(Zhu等, 2021)。

3.3.2 去模糊辅助成像

去模糊任务关注在高速运动与曝光积分条件下恢复清晰结构细节,或是涉及离焦模糊造成的空间信息丢失。在多传感器协同框架中,脉冲流通常用于提供更细粒度的时间证据与运动先验:相较于仅依赖相邻帧式图像估计运动,脉冲流能够更连续地刻画运动过程,从而为运动补偿对齐与模糊形成过程的推断提供额外约束。该类方法常将脉冲信息用于辅助估计更可靠的运动场/轨迹,并与帧式图像的高分辨率纹理互补,以实现清晰重建(Chen等, 2023, 2024; Zhang等, 2024b)。此外,针对离焦模糊情形,若系统能够在成像过程中提供焦平面变化或

多深度信息,则脉冲记录的连续光强变化可用于辅助构建更丰富的聚焦线索,从而提升对离焦区域结构的可恢复性(Teng等, 2024)。总体而言,协同去模糊方法通常遵循“运动/焦散线索提取—对齐补偿—细节恢复”的处理链路,其中脉冲流主要承担运动相关信息补充与时间一致性约束的角色。

4 连续时间时空耦合建模

4.1 光强度与光流联合建模

时域连续的脉冲流信息具有时空耦合性,即场景运动与强度变化在时空维度上相互关联、相互制约,因此充分利用这种耦合关系有助于同时提升重建质量和运动估计精度。如图3所示,光流可以引导翘曲操作,使脉冲流沿运动方向对齐,从而获得运动伪影更少的重建结果;反之,不同时刻的重建图像又能够提供更可靠的结构信息,以支持更准确的光流估计。因此,脉冲流的强度和光流之间存在天然的互促关系,可以通过联合优化实现同步学习与共同提升。

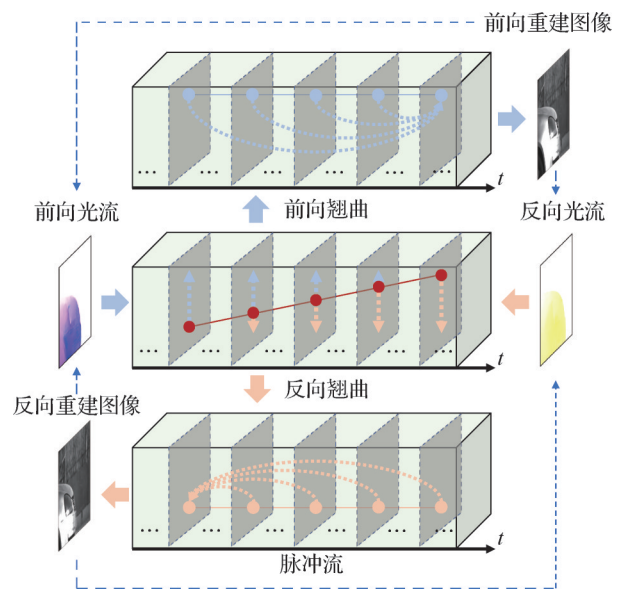


图3 脉冲流光强与光流的时空耦合性

Fig. 3 Spatio-temporal coupling between spike intensity and optical flow

基于这一思想, SLIO (simultaneous learning intensity and optical flow)(Zhu等, 2025)提出了一种同时估计高速脉冲流强度和光流的统一框架。该方法采用 SNN-ANN (spiking neural network—artificial neural network)混合网络,将编码器设计为一种密集

连接的脉冲神经元模块,在保持低功耗特性的同时有效建模短期与长期脉冲信息,并构建两个相互协同的解码器分别用于光流和强度估计。其中,基于光流特征的时间感知翘曲模块用于对齐强度解码器的时序特征,而改进后的强度特征又能够促进更准确的光流预测,从而在网络中形成相互促进的优化关系,最终实现更优的强度重建和光流估计性能。

4.2 去遮挡与连续视角补全

去遮挡问题旨在通过移除前景遮挡物提取清晰的背景图像,这在实际视觉感知与应用中具有重要意义。传统的帧式相机成像机制受限于固定的离散曝光时间与可能引发的运动模糊,在处理密集遮挡与高速运动场景时,往往难以获取充足的视角信息与清晰的背景观测。相比之下,脉冲相机凭借其超高的时间分辨率与连续积分发放机制,仅需单个处于快速移动状态的相机即可实现连续的多视角成像。这种高频采样的连续脉冲流能够提供更为密集且连贯的视角与运动线索,为解决极端遮挡条件下的背景结构恢复与视角补全提供了基础。

基于脉冲流中连续视角与时间的一致性,Zhang等人(2024c)首次探索了基于脉冲相机的去遮挡任务,并提出了端到端的时空耦合建模网络SpkOccNet。该网络首先将输入连续脉冲流沿时间轴划分为中心与两端共3个片段。随后利用长短时间窗口特征提取器,采用密集短窗口将两端片段转换为保留精细时序特征的局部表示;该网络同时采用较长的时间窗口将中心片段转换为类似长曝光的全局模糊表示。由于在相机快速移动时,前景遮挡物相对于背景的运动视差会随着视角的改变而发生显著变化,因此不同时间片段的脉冲在空间结构上形成了天然的互补关系。最后,网络引入交叉视角互注意力模块,在空间与通道维度上对来自不同视角的特征进行引导与融合,从而精准细化并重建出被遮挡的背景纹理。为推动该领域的规范化研究,研究者还构建了首个真实的脉冲视觉去遮挡数据集S-OCC,进一步验证了连续时间表征在复杂遮挡场景下的泛化能力与应用潜力。

5 高层感知任务

5.1 结构感知

结构感知任务关注从脉冲流中提取区域级显著

性与边界结构,以获得对目标轮廓、显著区域及其时序一致性的稳定刻画。与“先重建强度图像再执行视觉任务”的串行范式不同,该方向更强调在二值脉冲的时空域内进行端到端结构建模,以降低中间强度重建误差对下游结构预测的传递,并发挥连续脉冲流在时间密度与运动过程刻画方面的优势。以显著性检测为例,其输出为稠密掩码并天然包含边界信息,因此可视为从脉冲流中获取区域边缘结构的代表性任务。

Zhu等人(2024a)围绕连续脉冲流的显著性检测构建了系统化的模型,并提出基于循环脉冲Transformer的架构。该框架由脉冲时空特征提取、递归特征聚合、多尺度细化与多步监督构成。该方法配套构建了用于训练与评估的真实脉冲显著性检测数据集。在结构层面,该方法的关键在于保持脉冲通信二值性:直接沿用部分脉冲Transformer的残差相加会破坏二值脉冲序列属性,不利于纯脉冲信息传递;因此引入递归式聚合,增强时序信息提取并维持全脉冲传播。另外,文中采用脉冲间隔等时间统计量作为输入表征,并给出强度与脉冲计数相关的形式化表达,为“从脉冲流时间结构到区域显著性”的建模提供了表征的设计思路。

Liu等人(2025b)进一步指出,真实脉冲成像在长序列与复杂光照条件下容易出现复合噪声与域偏置,导致显著性预测产生不连续与系统性偏移,从而削弱结构图(区域与边界)的稳定性。针对该问题,作者提出SOTA(spike-navigated optimal transport)框架,将结构稳定化显式分解为“时间去偏”和“空间去偏”两条路径:时间去偏侧重利用SNN的多尺度特征关系刻画跨时间步的细微变化,并通过轻量卷积增强局部动态建模以缓解置信度漂移;空间去偏则以最优传输(optimal transport, OT)构造映射,最小化脉冲显著分布与真实图像显著分布之间的距离,并结合对抗学习以保持结构一致性。在形式化层面,文中将显著分布对齐刻画为Kantorovich最优传输问题,为“从脉冲域结构预测到跨域分布校正”的建模提供了明确的优化视角。

5.2 目标检测、语义分割及识别

针对脉冲相机的目标检测、语义分割与目标识别研究仍处于探索阶段。早期研究主要尝试将脉冲流信息引入深度学习框架以提升检测性能。例如, Li等人(2022)首次提出基于深度学习的脉冲相机目

标检测框架 Tar-Spikes + YOLO, 并设计了基于固定时间窗口和固定脉冲数量采样的两种重建基线方法 (TFI + YOLO 和 TFP + YOLO)。相比于依赖固定时间窗口的重建方法, 该方法通过可学习的时间聚合机制直接对脉冲流进行特征建模, 从而在一定程度上保留了脉冲流的时空特性。但在目标与背景反射特性相近的场景中, 脉冲信号之间仍存在较高相似性, 从而给目标检测带来一定挑战。

近期研究开始尝试直接在脉冲流上构建端到端检测模型。例如, Liu 等人 (2025a) 提出的 EASD (entropy-attention spike detector) 检测框架, 在自动驾驶场景中实现了基于脉冲流的目标检测, 并基于 DSEC (driving scenarios event camera dataset) (Gehrig 等, 2021) 数据集构建了面向驾驶场景的 DSEC-Spike 检测数据集。在该数据集上, 该方法在检测性能上超过了现有 RGB + 事件相机融合方法以及传统脉冲检测基线, 表明脉冲相机在高速动态场景中具有潜在优势。

在目标识别方面, Zhao 等人 (2022a, 2023, 2024d) 围绕高速运动目标识别开展了一系列研究。其首先提出脉冲视觉仿真平台 SpikeSim, 为脉冲视觉算法提供统一的仿真环境 (Zhao 等, 2022a); 随后针对高速运动目标识别问题提出相应识别方法, 并进一步通过去噪、运动建模以及知识蒸馏等策略提升识别性能和效率 (Zhao 等, 2023, 2024d)。

在语义分割方面, Zhang 等人 (2023) 首次探索了在脉冲流上开展语义分割任务, 并提出基于可学习小波变换的脉冲表征方法, 以充分利用脉冲流中的时序与频谱信息。在不改变下游图像分割网络结构的情况下, 该方法即可在多个语义分割模型上获得性能提升, 表明有效的脉冲表示对于高层视觉任务具有重要作用。

总体而言, 相较于脉冲视觉在低层视觉任务中的研究, 高层视觉任务仍处于发展初期。如何充分利用脉冲流的连续时间特性构建统一的高层感知框架, 并在检测、识别与语义理解等任务中发挥脉冲视觉的优势, 仍是未来的重要研究方向。

5.3 在线多目标追踪

多目标跟踪 (multi-object tracking, MOT) 是视觉感知系统中的重要任务, 在自动驾驶、机器人感知以及无人系统等应用中具有广泛需求。然而, 传统基于图像序列的跟踪方法通常依赖离散帧数据进行目

标检测与关联。当场景中存在高速运动时, 有限帧率会导致时间信息缺失, 而曝光积分过程则容易产生明显的运动模糊, 从而影响检测与跟踪性能。脉冲相机以连续时间方式记录光强变化, 并以异步脉冲流形式输出视觉信息。这种数据形式能够在极高时间分辨率下捕获运动变化, 为高速目标的连续跟踪提供新的信息来源。因此, 如何直接利用脉冲流的时空结构设计在线跟踪算法, 逐渐成为脉冲视觉研究中的一个重要方向。

图4展示了传统RGB相机方法与脉冲相机方法在高速多目标跟踪任务中的差异。传统RGB相机输出离散图像序列, 在高速场景中容易受到运动模糊和帧率限制的影响; 相比之下, 脉冲相机能够生成连续、低时延的脉冲流, 从而在更细粒度的时间尺度上刻画目标运动过程。现有方法大致可以分为3类: 第1类是直接基于离散图像序列的多目标跟踪方法, 在高速运动场景下容易出现目标关联失败; 第2类方法通过脉冲重建获得清晰图像并结合检测框进行跟踪, 例如利用YOLOv5 (you only look once) 等检测模型, 但该路线通常依赖大量标注数据, 并引入额外的重建计算开销; 第3类方法则直接在原始脉冲流上进行计算, 例如基于脉冲神经网络 (Maass, 1997) 的跟踪方法, 可以在无需图像重建和标注训练的情况下实现低时延的在线多目标跟踪 (Zheng 等, 2026a)。

在这一方向上, Huang 等人 (2023) 首先提出了 SpikeSORT 算法, 这是较早面向脉冲相机的在线多目标检测与跟踪框架。该方法首先通过时空滤波对脉冲数据进行预处理, 以减少背景区域产生的冗余脉冲信号。随后利用空间连通区域搜索方法对运动目标进行检测, 并通过匈牙利匹配算法进行目标关联 (Kuhn, 1955), 结合卡尔曼滤波器进行轨迹预测与更新 (Kalman, 1960)。由于算法结构相对简单, 该方法能够在连续脉冲流上实现在线多目标跟踪, 并为后续研究奠定了基础。

在 SpikeSORT 算法基础上, Zhu 等人 (2022) 进一步实现了基于 FPGA (field-programmable gate array) 的硬件加速系统。该系统针对脉冲数据的高时间分辨率特点, 对检测与跟踪模块进行了硬件级优化, 实现了面向高速场景的实时目标检测与跟踪。实验结果表明, 该系统在高速旋转风扇数据集上能够实现超过 20 kHz 的处理速度, 同时保持较高的跟踪准确

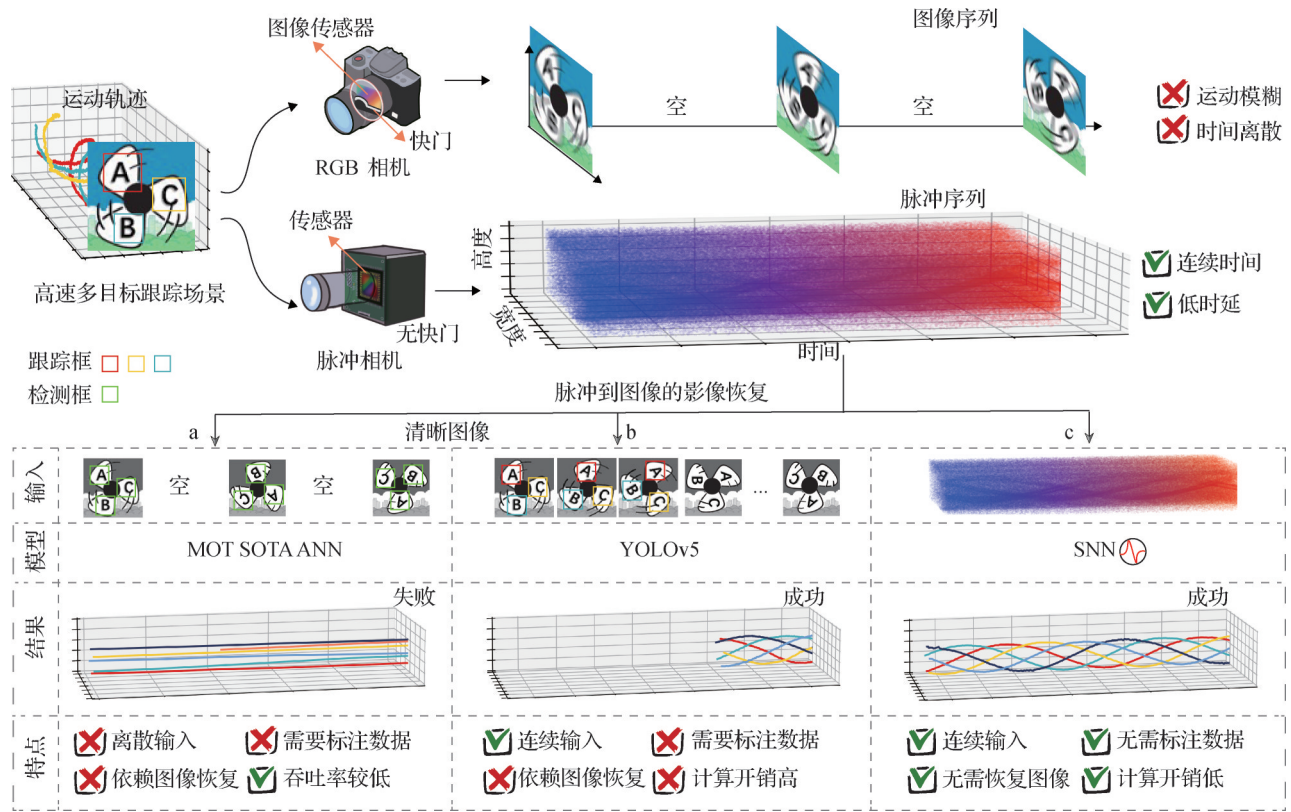


图4 传统RGB相机与脉冲相机在高速多目标跟踪任务中的比较(Zheng等,2026a)

Fig. 4 Comparison between traditional RGB cameras and spike cameras for high-speed multi-object tracking(Zheng et al. ,2026a)

率。这一工作从系统实现角度验证了脉冲视觉方法在超高速目标跟踪任务中的应用潜力。

在此基础上, Zheng 等人 (2023a) 提出了 ODTNet (spiking network for object detection and tracking) 方法, 在检测阶段引入脉冲神经网络运动估计模块, 以增强对动态场景的感知能力。与 SpikeSORT 不同, 该方法在目标检测阶段采用基于密度的聚类算法 DBSCAN (density-based spatial clustering of applications with noise) 对脉冲事件进行空间聚类 (Ester 等, 1996), 从而能够更加稳定地提取运动目标区域。通过结合运动信息与空间位置, 该方法能够在存在相机运动或复杂动态变化的场景中实现更加稳健的目标检测与跟踪。

随后, Zheng 等人 (2026a) 进一步提出 SVS2 (spike vision system) 方法, 对 SpikeSORT 框架中的目标检测模块进行了改进。该方法将原有的八连通区域搜索替换为基于动态神经场 (dynamic neural field, DNF) 的注意力检测机制, 从而能够更加有效地从脉冲流中提取运动目标区域。与此同时, SVS2 仍然保留了传统的目标关联与跟踪模块, 即通过匈牙利匹配算法进行数据关联, 并结合卡尔曼滤波器

实现轨迹更新。实验结果表明, 这种基于神经注意机制的检测方式能够有效提升目标检测的稳定性与跟踪性能。

进一步地, Zheng 等人 (2026a) 提出了 SNN-Tracker 方法, 实现了一个完全基于脉冲神经网络的在线多目标跟踪框架。该方法利用动态神经场机制实现目标检测, 并结合胜者为王 (winner-take-all, WTA) 机制以及基于脉冲时序依赖可塑性 (spike-timing-dependent plasticity, STDP) 的在线学习模块, 实现对目标运动轨迹的持续建模。与依赖图像重建或传统聚类算法的方法不同, SNNTracker 能够直接处理连续脉冲流数据, 在连续时间域中完成检测与跟踪过程, 从而减少图像重建带来的计算开销与延迟。

表2总结了上述几种面向脉冲相机的在线多目标检测与跟踪方法的主要特点。总体来看, 该研究方向正在从早期基于时空滤波与传统数据关联的方法, 逐步发展到结合神经网络模型的连续时间跟踪框架, 并进一步向完全基于脉冲神经网络的端到端方法演进。这些研究表明, 直接在脉冲流上进行计算不仅能够充分利用脉冲数据的时间连续性, 还为

表2 面向脉冲相机的在线多目标检测与跟踪方法比较

Table 2 Comparison of spike-camera-based online multi-object tracking frameworks

方法	检测方法	关联与跟踪	是否脉冲计算	主要特点
SpikeSORT(Huang等,2023)	八连通区域检测	匈牙利匹配+卡尔曼滤波	否	结构简单,适用于固定相机场景
ODTSNet(Zheng等,2023a)	DBSCAN聚类+运动估计	匈牙利匹配+卡尔曼滤波	部分	引入运动信息,提高复杂场景鲁棒性
SVS2(Zheng等,2026a)	DNF注意力检测	匈牙利匹配+卡尔曼滤波	部分	改进检测模块
SNNTracker(Zheng等,2026a)	DNF注意力检测	WTA+STDP在线学习	是	全SNN框架,直接处理脉冲流

实现低时延、高效率的实时视觉感知系统提供了新的技术路径。

6 三维表示方法

近年来,三维场景表示方法取得了快速发展。如图5所示,从早期基于多视图几何与体素网格的显式建模,到以隐式函数为核心的连续表示,研究者不断探索更高效、精细的场景表达方式。尤其是神经辐射场(neural radiance field, NeRF)与三维高斯泼溅(3D Gaussian splatting, 3DGS)的提出,为高精度三维表示与新视角渲染提供了新的技术范式。在脉冲相机的采样机制下,三维重建的建模方式发生了显著变化。传统NeRF或多视图方法通常依赖高信噪

比强度图像进行光度一致性优化,而脉冲视觉需要从连续二值脉冲序列中恢复稳定的几何与辐射属性。这要求模型不仅具备连续辐射场表达能力,还需建立辐射强度到脉冲编码之间的可微映射,并在长时间脉冲累积过程中保持几何与光度一致性。因此,脉冲视觉下的三维建模本质上是在时间冗余与瞬时稀疏之间建立平衡,从连续脉冲流中恢复稳定的空间结构。在神经三维表示框架中,该问题进一步转化为:如何在辐射场建模过程中嵌入脉冲成像物理机制,使渲染结果与脉冲观测在统计和物理层面保持一致。

围绕这一目标,现有研究主要沿两条技术路线展开:隐式连续表示的NeRF(Mildenhall等,2020)以及显式参数表示的3DGS(Kerbl等,2023),并逐渐从

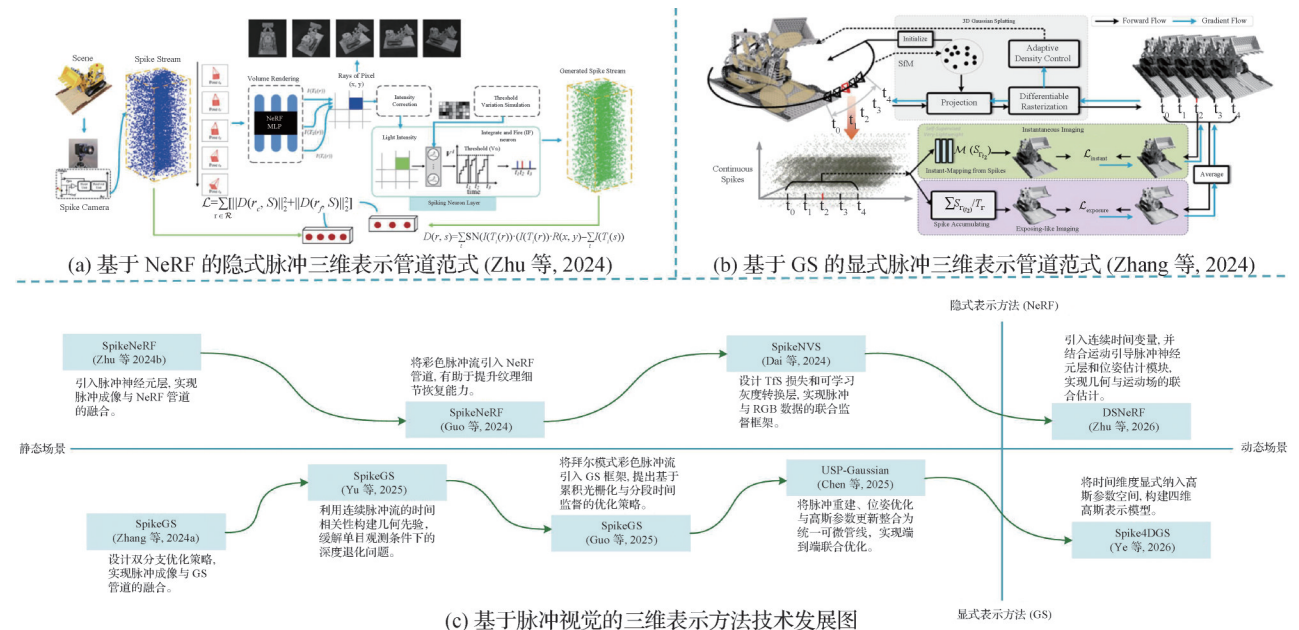


图5 基于脉冲视觉三维表示方法与技术发展图

Fig. 5 Development of spike-based 3D representation methods and technologies ((a) NeRF-based implicit spike 3D representation pipeline; (b) GS-based explicit spike 3D representation pipeline; (c) technical roadmap of spike-vision-based 3D representation methods)

静态场景扩展到动态建模。

6.1 静态场景下的脉冲三维重建

在静态场景中,空间结构随时间保持不变,核心任务是从连续脉冲流中估计稳定的几何与辐射分布。由于脉冲数据具有二值编码和高时间分辨率,其统计特性与传统图像存在显著差异,因此三维模型需要显式考虑成像物理过程。

在隐式表示方向,NeRF通过学习空间位置与视角方向到颜色和体密度的映射,实现连续辐射场建模。SpikeNeRF(Zhu等,2024b)在NeRF渲染链路中引入可微脉冲神经元层,模拟脉冲相机的积分与触发机制,从而直接在脉冲域进行监督。随后,SpikeNeRF(Guo等,2024)将拜耳模式彩色脉冲流引入隐式辐射场,通过显式建模滤波结构与时间积分过程,使模型能够适配彩色脉冲数据。SpikeNVS(Dai等,2024)进一步在脉冲与RGB联合监督框架下优化,通过“TfS(texture from spike)”损失将传统脉冲重建算法中的时域信息转化为学习约束,并引入可学习灰度转换层以提升脉冲重建与NeRF渲染之间的对齐精度。总体而言,隐式表示具有良好的物理一致性和表达能力,但训练与渲染开销较高。

随着3DGS的提出,研究开始探索脉冲视觉与显式表示的结合。3DGS通过可优化的高斯基元表示场景结构,并利用可微光栅化实现高效渲染。SpikeGS(Zhang等,2024a)首次将脉冲监督引入3DGS框架,通过“瞬时损失”和“曝光损失”联合约束高斯参数,使模型在保持高效优化的同时适应脉冲观测特性。随后,SpikeGS(Yu等,2025)利用连续脉冲流的时间相关性构建几何先验,以缓解单目观测下的深度退化问题。针对彩色脉冲数据,SpikeGS(Guo等,2025)通过累积光栅化与分段监督策略,使高斯参数在时间与颜色维度上同时受到约束,从而提升几何初始化稳定性和纹理恢复能力。此外,USP-Gaussian(unifying spike-based image reconstruction, pose correction and Gaussian splatting)(Chen等,2025)将脉冲重建、位姿优化和高斯更新整合为统一可微管线,实现端到端联合优化,从而显著提升快速运动条件下的位姿稳定性和重建精度。

总体来看,静态场景研究经历了从可行性验证到复杂数据建模,再到端到端联合优化的发展过程。NeRF的隐式精细表达与3DGS的显式高效建模在目标上相互补充,共同推动了脉冲三维重建的发展。

6.2 动态NeRF

当相机或物体发生高速运动时,运动本身成为场景的重要组成部分。脉冲相机的高时间分辨率能够提供无运动模糊的瞬时观测,并记录连续运动轨迹,因此动态场景重建的核心问题转变为在高速运动中同时恢复几何结构与运动场。在隐式表示方向,DSNeRF(dynamic view synthesis NeRF)(Zhu等,2026)在辐射场模型中引入连续时间变量,并结合运动引导脉冲神经元层,实现几何结构与场景流的联合估计,同时联合优化相机位姿,从而保持辐射场在时间维度上的连续变化;在显式表示方向,Spike4DGS(Ye等,2025)将时间维度引入高斯参数空间,构建四维高斯表示模型。该方法利用多目脉冲阵列进行稠密初始化,并提出“脉冲—像素”协同监督策略,通过像素渲染损失与时空脉冲损失联合优化四维高斯参数,从而实现动态场景的端到端重建。相比隐式方法,四维高斯表示在计算效率和实时扩展方面更具潜力。

6.3 脉冲视觉与三维表示融合的讨论与思考

脉冲视觉与神经三维表示的结合,不仅是数据形式或监督信号的变化,更是一种建模范式的转变。与传统帧式成像依赖曝光积分生成高信噪比图像不同,脉冲相机通过高时间分辨率的连续采样,将光强信息编码为二值脉冲序列。理解这种融合的关键,在于分析观测机制如何重塑表示结构和优化策略。

NeRF与3DGS在脉冲视觉融合中体现了不同建模侧重:前者强调连续表达和物理一致性,适合复杂光照和高噪声场景;后者强调结构参数与优化效率,适合大规模动态场景。研究趋势正在向混合表示演进,通过在显式结构中加入物理约束提升稳定性,同时为隐式方法探索加速结构以降低计算成本。二者融合可能成为未来重要方向,实现连续表达与高效优化的统一框架。NeRF-GS(Fang等,2025)通过共享空间特征、残差建模与联合优化策略,将连续体表示与显式高斯结构进行系统融合,在保持实时渲染能力的同时显著提升重建质量。该工作进一步验证了这一观点,表明NeRF与3DGS并非竞争关系,而具有明确的互补性,为混合三维表示框架提供了实证支持。总体而言,脉冲视觉不仅改变了数据形式,也重塑了三维表示的建模方式。高时间分辨率提供了动态建模优势,但对表示结构和优化策略提出更高要求。NeRF与3DGS分别承担物理建模和效

率优化角色,未来融合有望实现既物理一致又高效的统一脉冲三维表示。

7 数据、模拟与评测体系

7.1 模拟器发展

在早期脉冲视觉研究中,由于真实脉冲数据的匮乏,研究者往往需要依据脉冲原理,在现有图像数据集的基础上合成模拟脉冲流数据以支持相关研究,由此催生了脉冲模拟器这一研究分支。

早期的脉冲数据模拟器通常将图像像素值线性映射为光电流强度,并结合实际应用中的最小时间分辨率对时间轴进行离散化处理,随后依据给定的脉冲发放阈值生成等间隔的脉冲序列(Zhao等,2021)。在模拟视频对应的脉冲流数据时,为了获得具有更高时间分辨率的脉冲表示,相关工作(Zhang等,2022)采用了一种视频插值方法,在相邻的30帧/s视频帧之间生成RGB中间帧(Sim等,2021),并在此基础上按照Zhao等人(2021)工作中的像素值与脉冲流映射机制生成高时间分辨率的脉冲流。尽管上述脉冲数据模拟方法能够高效地生成脉冲流数据,但其通常未充分考虑光子的物理统计特性以及实际电路中不可避免的噪声因素。

事实上,光子的到达过程并非均匀分布,已有研究表明其可建模为泊松过程(Nelson,2017)。这意味着,即便在不存在额外噪声的理想条件下,恒定光强所对应的脉冲流在时间上也不应呈现严格的等间隔分布。此外,实际硬件系统中还普遍存在诸如暗电流等时序噪声,使得传感器即便在完全无光的环境下仍可能产生脉冲发放(Chamberlain,1969)。针对上述问题,Zhu等人(2021)同时考虑了光子到达的物理统计特性以及电路中时序噪声的影响,通过对光强引入随机扰动,并采用高斯分布对模拟脉冲流的脉宽进行约束,从而提升了模拟结果的物理合理性。

随着真实拍摄脉冲数据集的逐步出现(Zhu等,2019;Zheng等,2026b),模拟脉冲数据的生成目标应当尽可能逼近真实脉冲数据的统计特性与时序行为。Zhao等人(2022b)利用真实脉冲数据对像素值与光强之间的线性系数进行了拟合,并依据实际系统中的最小可测时间单元对强光条件下产生的极短脉冲间隔进行截断约束,其模拟结果与真实脉冲流

达到了约99%的相似度。

总体而言,脉冲模拟器正从无真实数据输入的理想仿真,向有真实数据参考的高逼真度模拟演进。为实现这一目标,模拟器需精确建模真实场景中的各类噪声,这推动了其内部噪声模型的复杂化。现有模拟器的分类如图6所示。

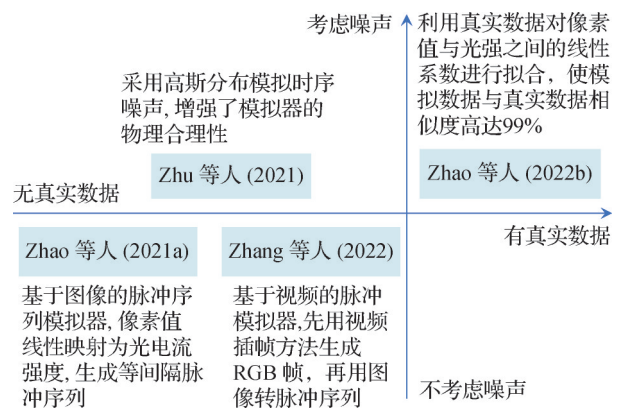


图6 脉冲模拟器的分类

Fig. 6 Classification of spike simulators

7.2 数据集与标注

自脉冲相机(Dong等,2017)问世以来,基于其真实采集数据的脉冲视觉数据集不断涌现。其中,recVidarReal2019作为最早且最具代表性的真实场景脉冲数据集,涵盖了物体高速运动(如旋转风扇)与相机高速运动(如置于高铁上拍摄)两类典型场景,已成为图像重建任务中广泛使用的基准数据集。随后,面向不同视觉任务的数据集相继发布:motVidarReal2020专注于多目标检测与跟踪,提供了带有边界框标注的脉冲序列,适用于动态场景下的多目标感知;motVidarReal2021则进一步扩展,包含51段真实场景脉冲数据,支持图像重建、目标检测、目标跟踪及运动分析等多种任务,为脉冲视觉算法的综合评估提供了更丰富的平台。PKU-Vidar-DVS(Li等,2022)作为首个多模态神经形态目标检测数据集,引入了Vidar脉冲与DVS(dynamic vision sensor)事件的双模态融合。该数据集利用混合相机系统在9个室内外挑战性场景中采集了490个序列,侧重高速与低光照条件下的目标检测任务,为神经形态视觉在复杂环境下的应用提供了重要支撑。

尽管真实场景脉冲数据集陆续问世,但其数据规模仍然有限,且人工标注成本高昂。考虑到深度网络训练通常依赖大规模数据,在实际训练中仍需

大量借助模拟脉冲数据进行补充。

Hu 等人(2022)针对光流估计任务,提出了一种与渲染引擎紧密耦合的脉冲相机模拟器,并生成了模拟数据集 SPIFT(spikingly flying things)和 PHM(photo-realistic high-speed motion)。该模拟器首先由渲染引擎根据预设参数合成连续图像序列,再通过采样处理函数将其转换为对应的脉冲流数据。在此基础上,相关工作(Hu等,2025a)进一步在模拟场

景中引入0—1区间的随机衰减因子以降低整体光照强度,从而构建了一组面向低光照条件下脉冲图像重建任务的模拟数据集 LLR(low-light reconstruction)。此外,Zhang 等人(2023)基于 Cityscapes 数据集(Cordts等,2016),采用Zhang 等人(2022)提出的模拟方法,构建了首个面向脉冲图像分割任务的模拟数据集。本文将开源的脉冲数据集汇总于表3。

表3 脉冲数据集汇总
Table 3 Spike-based datasets

数据集	真实或模拟	主要任务
recVidarReal2019	真实	脉冲图像重构
motVidarReal2020	真实	多目标运动物体的检测跟踪
motVidarReal2021	真实	脉冲图像重构、高速目标检测跟踪、运动分析(运动分割、光流估计)等任务
PKU-Vidar-DVS	真实	高速与极端光照场景下的目标检测
PKU-Spike-Stereo	真实	深度估计
SPIFT	模拟	脉冲图像重构、高速目标检测跟踪、运动分析(光流估计)等任务
PHM	模拟	脉冲图像重构、高速目标检测跟踪、运动分析(光流估计)等任务
LLR	模拟	低光脉冲图像重构
Spike-Cityscapes	模拟	脉冲语义分割

7.3 统一评测指标体系

本节按照目前脉冲视觉中常见任务所利用的信息层级,分别阐述不同任务下的评价指标。

首先阐述底层视觉任务,主要包括脉冲信号模拟与脉冲图像重建,这类任务旨在从原始脉冲流中恢复出模拟信号的物理特性或可视化图像。对于脉冲信号模拟任务而言,主要的评价指标包含模拟属于与真实数据的脉冲发放率相似性(Zhao等,2022a),以及二者分布的差异。具体而言,假设脉冲流的表示为一个 $M \times N \times T$ 的体素,像素 p 的脉冲发放频率为

$$f_p(\cdot) = \frac{n}{T} \quad (2)$$

式中, n 为发放的脉冲数量,则模拟脉冲流 \hat{S} 与真实脉冲流 S 之间的发放率相似性可计算为

$$F_{\text{sim}}(S, \hat{S}) = \frac{1}{MN} \sum_p \left(1 - \frac{|f_p(S) - f_p(\hat{S})|}{f_p(S)} \right) \quad (3)$$

而模拟脉冲流与真实脉冲流之间的分布相似性可以用KL(Kullback-Leibler)散度表示,假设给定脉冲流 S ,像素 p 发放与不发放脉冲的概率分别记为

$P(S_p = 1)$ 和 $P(S_p = 0)$,则模拟脉冲流 \hat{S} 与真实脉冲流 S 之间的分布相似性可计算为

$$KL(S, \hat{S}) = \sum_p \sum_{i=0,1} P(S_p = i) \log_2 \frac{P(S_p = i)}{P(\hat{S}_p = i)} \quad (4)$$

脉冲图像重建任务需要着重考虑重建图像的视觉质量。当真实的参考图像可获得时,需通过对比重建图像与参考图像之间的相似度衡量重建质量,常用的全参考图像质量评价指标包括结构相似性(structural similarity index measure, SSIM)(Wang等,2004)。当真实的参考图像无法获取时,则需采用无参考图像质量评价指标,例如峰值信噪比(peak signal-to-noise ratio, PSNR)和学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)(Zhang等,2018)等,以评估重建图像的感知真实度与失真程度。

中层视觉任务主要关注从脉冲流中恢复场景的几何结构与运动信息,为高层语义理解提供基础视觉表征。典型任务包括深度估计与光流估计等。在深度估计任务中,常用评价指标包括预测深度与真实深度之间的绝对相对误差(absolute relative error,

Abs Rel)、平方相对误差(squared relative error, Sq Rel)、平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)以及阈值准确度等,用于衡量预测深度与真实深度之间的偏差程度。在光流估计任务中,则主要采用平均端点误差(end-point error, EPE)以及光流角度误差(angular error)等指标,对预测光流场与真实光流场之间的差异进行量化评估。这类指标能够直接反映脉冲视觉在连续时间运动建模方面的性能表现。

高层视觉任务则关注从脉冲流中提取更高级的语义信息,实现对场景中目标及其行为的理解。以目标检测和多目标跟踪任务为例,常用评价指标包括平均精度均值(mean average precision, mAP)、多目标跟踪精度(multiple object tracking accuracy, MOTA)、多目标跟踪精确度(multiple object tracking precision, MOTP)、误检数(false positives, FP)、漏检数(false negatives, FN)以及标识切换次数(identity switch, ID Switch)等(Bernardin 和 Stiefelhagen, 2008)。其中,误检数表示将背景误判为目标的检测错误数量,漏检数表示真实目标未被检测到的数量,标识切换次数则反映跟踪过程中目标标识错误切换的情况,这些指标共同衡量检测与跟踪算法的准确性与稳定性。

除了上述任务级指标外,面向实际应用的脉冲视觉系统还需要综合考虑系统层面的性能评价。系统指标通常包括算法推理延迟(latency)、系统吞吐率(throughput)、能耗效率(energy efficiency)以及实时处理能力(real-time capability)等。例如,在自动驾驶或机器人等应用场景中,系统延迟直接影响感知与决策的响应速度,而能耗指标则决定算法在嵌入式平台或神经形态硬件上的部署可行性。此外,在一些任务驱动型应用中,还可以通过任务成功率或任务完成精度等综合指标评价系统整体性能。需要指出的是,目前针对脉冲视觉系统的系统级评测指标尚未形成统一标准,如何建立覆盖感知性能、计算效率与系统可靠性的统一评测体系,仍然是未来需要进一步研究和完善的重要方向。

8 脉冲视觉平台化与开放生态

随着脉冲相机在高速视觉感知中的研究不断深入,仅依赖单个算法或孤立数据集难以支撑该领域

的持续发展。由于脉冲视觉数据在表示形式、时序结构以及处理方式上与传统帧式图像存在显著差异,现有视觉算法工具链和评测框架难以直接适配。因此,构建统一的数据处理工具、算法模块库以及标准化评测流程,逐渐成为推动脉冲视觉研究生态发展的重要基础设施。

近年来,脉冲视觉研究逐渐从早期的算法探索阶段,发展为涵盖数据获取、算法设计与系统应用的完整研究体系。与此同时,开源平台与社区协作机制开始在该领域发挥重要作用,通过整合数据资源、算法模块以及评测工具,为研究者提供统一的研究环境。本节将从平台工具链、研究流程以及社区生态3个方面,对脉冲视觉研究基础设施的发展进行总结。

8.1 SpikeCV 平台与工具链

为促进脉冲视觉研究的系统化发展,北京大学多媒体信息处理全国重点实验室于2022年推出了SpikeCV 开源平台(Zheng 等, 2026b)。该平台旨在为脉冲视觉研究提供从数据处理、算法开发到性能评测的一体化工具链,使研究者能够在统一框架下开展算法研究与应用开发。

SpikeCV 平台整合了多种核心功能模块,包括脉冲数据加载与解码、数据集管理、视觉算法模块、训练与测试脚本以及统一评测工具等,从而显著降低了脉冲视觉算法开发与实验复现的门槛。通过标准化接口,研究者可以在同一平台上调用不同数据集、算法模型以及评测指标,实现跨任务和跨算法的统一比较。

如图7所示,SpikeCV 平台旨在构建一个面向脉冲视觉研究的开放生态系统。平台通过统一工具链将脉冲相机、算法研究与应用开发连接起来,在研究者与传感器之间建立桥梁。通过提供标准化数据资源、算法模块以及统一的接口与工具,SpikeCV 使研究者能够更加便捷地开展脉冲视觉相关研究,并降低了实验复现和系统开发的门槛。在平台设计上,SpikeCV 强调模块化与标准化的工具体系。平台通过封装化接口与统一的数据结构,将数据读取、算法调用和结果评测等环节整合到同一框架之中,使不同研究任务能够在统一环境下进行开发与测试。同时,平台还支持实时数据处理与系统接口扩展,使研究成果能够更容易地从离线算法验证过渡到真实场景应用。

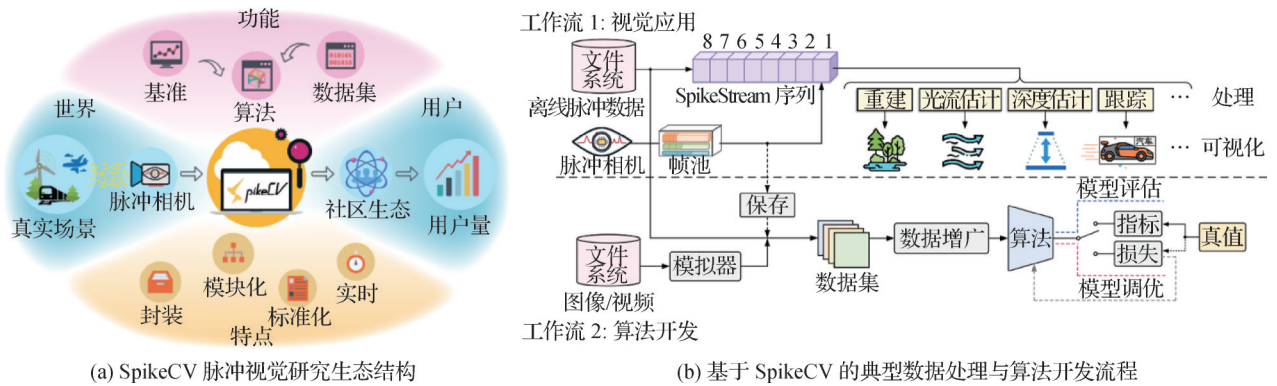


图 7 SpikeCV 脉冲视觉平台生态示意图(Zheng 等, 2026b)

Fig. 7 SpikeCV platform and workflow for spike vision research(Zheng et al. , 2026b)

((a)the SpikeCV ecosystem;(b)the typical processing pipeline)

在此基础上, SpikeCV 不仅作为算法开发工具, 也逐渐形成了一个面向脉冲视觉研究的社区平台。通过共享数据、算法实现和实验基准, 平台促进了不同研究团队之间的协作与资源共享, 并推动脉冲视觉技术从单一算法研究逐步发展为一个开放的研究生态。在数据层面, 平台集成了多种真实采集数据与模拟生成数据, 并提供统一的数据读取与转换接口; 在算法层面, 平台收录并实现了多类典型脉冲视觉任务的算法模块, 包括图像重建、光流估计、深度估计以及目标检测与跟踪等任务; 此外, 在系统层面, SpikeCV 还提供了面向脉冲相机设备与神经形态计算平台的接口支持, 为算法向实际系统部署提供了基础条件。

通过这些模块的协同作用, SpikeCV 逐渐形成了连接数据资源、算法模型与硬件系统的统一研究工具链。目前, SpikeCV 平台已在 OpenI 启智开源社区上线并持续维护, 截止目前累计下载量已超过 15 万次, 并持续集成新的数据资源与算法模块, 为脉冲视觉研究社区提供重要的基础支撑(Zheng 等, 2022)。

8.2 基于 SpikeCV 的典型研究流程

在统一工具链的支持下, 研究者可以基于 SpikeCV 快速构建完整的脉冲视觉研究流程。从原始脉冲数据获取到算法评测, 平台提供了一套标准化的数据处理与算法开发流程, 如图 8 所示。

在数据获取阶段, 脉冲数据可以来源于真实脉冲相机采集, 也可以通过脉冲模拟器从传统图像或视频生成。随后数据通过统一的 SpikeStream 数据格式进行管理, 并进一步构建标准化数据集。在算法开发阶段, 研究者可以利用平台提供的数据增强

模块、训练脚本以及模型评测工具, 对算法进行训练与验证。最终, 通过统一评测指标与基准测试流程, 可以对不同算法在多个任务上的性能进行系统比较。

基于该工具链, SpikeCV 平台提供了多种典型的端到端研究范式流程, 帮助研究者快速复现实验并开发新算法。例如: 1) “脉冲流—运动/深度估计—高质量图像重建”。通过提取脉冲流中的时空信息估计场景中的光流或深度信息, 结合 Transformer 等神经网络架构实现高速场景下的高质量图像重建, 例如高动态范围图像或去遮挡等。2) “脉冲流—时空特征估计—高层次视觉任务”。直接从脉冲流中估计光强信息或光流等时空特征, 从而实现高速视觉任务, 例如目标检测、识别或跟踪。3) “脉冲流—三维表示—场景重建”。通过基于多层感知机或高斯泼溅等的神经渲染三维重建方法, 从连续脉冲数据中恢复动态场景结构, 并进一步实现新视角合成。这些示例流程为脉冲视觉算法开发提供了可复现的实验框架, 同时也促进了不同研究方法之间的公平比较。

8.3 开放竞赛与社区生态

除了算法工具链之外, SpikeCV 平台还通过开放数据与竞赛机制推动研究社区的协同发展。近年来, 随着脉冲相机在高速运动、高动态范围以及复杂光照环境中的优势逐渐显现, 脉冲视觉研究正在迅速增长。然而, 由于脉冲数据表示形式与传统图像存在本质差异, 现有视觉算法与评测体系难以直接适用于脉冲数据, 因此构建开放数据集、统一评测协议以及共享算法平台成为推动该领域发展的关键。

为促进脉冲视觉生态建设, SpikeCV 平台在国

际学术会议上组织了多项开放竞赛与基准评测任务。例如,在国际人工智能联合会议(International Joint Conference on Artificial Intelligence, IJCAI) 2025 上举办的第 1 届 SpikeCV 竞赛吸引了来自全球 96 支队伍、200 余名研究者参与,参赛单位覆盖 30 所高校、科研机构以及企业团队。随后在第 8 届中国模式识别与计算机视觉学术会议(The 8th Chinese Conference on Pattern Recognition and Computer Vision, PRCV) 2025 年举办的第 2 届竞赛中,共有 38 支队伍、100 余名研究者参与,参赛单位来自 22 所高校、3 家科研机构以及 4 家企业团队。两届竞赛的连续举办,使 SpikeCV 逐渐从单一算法平台发展为连接高校、科研机构与产业团队的开放研究社区。

SpikeCV 平台及相关数据资源、算法实现和竞赛信息均通过官方网站持续发布与维护(Zheng 等, 2023b)。该网站通过标准化数据共享、算法开源以及竞赛机制的结合, SpikeCV 逐渐形成了连接研究者、数据资源与算法工具的开放生态系统。当前,脉冲视觉相关研究成果已在多个主流人工智能与计算机视觉会议和期刊上发表。这些进展表明,脉冲视觉研究正在逐步从早期探索阶段发展为具有广泛关注度的新兴研究方向。

总体而言, SpikeCV 平台展示了脉冲视觉研究正在从单一算法探索阶段,逐步迈向可复现研究基础设施与开放生态系统建设的新阶段。通过统一工具链、开放数据资源以及社区竞赛机制,该平台为脉冲视觉算法研究与实际应用之间建立了重要桥梁,也为未来该领域的持续发展提供了关键支撑。

9 结语及未来挑战

如图 8 所示,脉冲视觉的未来发展正在由多模传感输入、类脑视觉算法、神经形态计算芯片以及多类实际应用场景协同驱动。相较于传统帧式视觉,脉冲视觉在数据形态、建模范式与系统实现层面均具有显著差异。尽管近年来相关算法与应用快速扩展,但若要进一步支撑更大规模、跨场景的稳定部署,仍需在统一表征、数据与学习范式、多模态鲁棒性、系统级优化以及闭环应用等方面形成更加系统的方法体系。下面围绕若干关键挑战及其潜在研究方向进行总结。

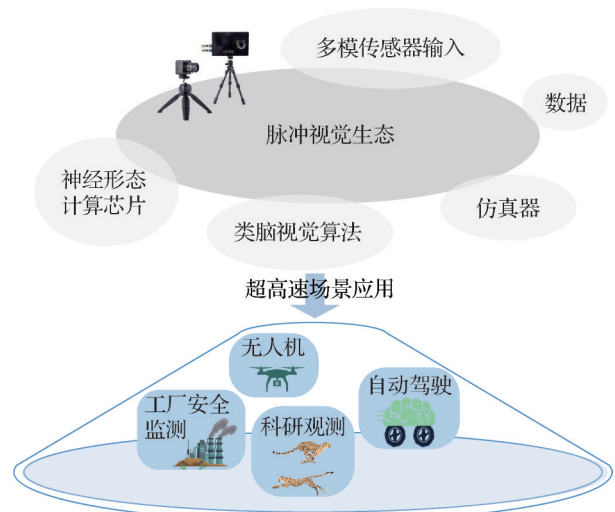


图 8 脉冲视觉未来研究生态示意图

Fig. 8 Future ecosystem of spike vision research

9.1 连续时间表征统一

现有方法往往在离散时间窗内对脉冲流进行体系化、计数或时间间隔统计,再将得到的表征输入卷积网络或 Transformer 等模型(Zhao 等, 2022b, 2024b, 2026)。这类离散化处理在工程实现上较为便利,但不可避免地引入了时间量化误差,也可能削弱脉冲相机对场景进行连续记录的优势。未来的一个重要方向是探索更贴近连续时间本质的统一表征与算子设计,例如基于时间戳的连续核积分、神经微分方程模型、状态空间模型以及可微分的连续时间滤波与对齐机制等,使模型能够在不依赖固定时间窗的情况下处理不同速率和不同运动模式的输入。同时,统一表征还应具备跨任务可迁移性,即同一类连续时间表示不仅能够支撑低层场景恢复,也能够服务于高层语义理解,并具备可解释的时序一致性约束。

9.2 大规模真实数据与自监督学习

脉冲相机数据的采集与标注成本较高,现有公开实拍数据规模仍然有限,这使得不少方法容易在小规模数据集上过拟合,并在跨场景测试中表现出泛化能力不足的问题。未来需要进一步推动大规模真实脉冲数据集建设,覆盖更加丰富的光照条件、运动模式和应用环境,并在跨传感器同步、标注规范以及数据格式方面逐步建立统一标准,以支撑更可比较的基准评测与可复现实验。与此同时,自监督与弱监督学习值得进一步关注。可利用跨模态一致性、序列时间一致性以及脉冲生成机理约束构造训

练信号,从而在弱标注甚至无标注条件下学习稳健表征。更长远目标是发展适用于脉冲视觉的大规模预训练体系,使模型具备更强的迁移能力和跨任务泛化能力。

9.3 多模态协同与对齐鲁棒性

多传感器协同是脉冲视觉的重要实现形式之一,但对齐误差往往成为融合性能的上限决定因素(Zhang等,2024b)。实际系统中常见问题包括时间同步误差、空间外参漂移、视差与遮挡引起的跨模态不一致,以及不同模态噪声统计差异导致的融合偏置等。未来研究可从两个方面推进:1)发展显式的时空校准与不确定性建模,将对齐误差作为可估计变量并纳入端到端优化;2)构建对对齐误差更鲁棒的融合机制,例如引入可学习的对齐置信度、局部一致性筛选或分层融合策略,使系统在对齐不完美时仍能稳定利用互补信息。此外,对齐鲁棒性也有必要在评测层面被显式度量,而不仅仅通过最终任务精度间接反映。

9.4 系统指标驱动的算法优化

脉冲视觉方法的评价不应仅停留在重建质量或任务精度,还应进一步纳入时延、吞吐、功耗与存储等系统指标,尤其是在高速感知、边缘部署和实时控制场景中(Zhu等,2022)。当前不少研究仍以离线精度为主要目标,较少系统分析窗口长度、时序聚合范围和网络复杂度对端到端时延与资源消耗的影响。未来可探索“系统指标驱动”的联合优化范式,将延迟约束、算力预算、带宽预算与精度目标共同纳入训练目标或架构搜索过程;在算法侧发展动态计算、时间步自适应、分辨率自适应等机制;在评测侧建立统一的端到端基准,更真实地反映部署代价与实际应用收益。

9.5 硬件与算法协同设计

脉冲相机的阈值设置、读出机制、噪声特性以及像素响应非一致性会直接影响算法可利用的信息形态。如果仅在固定硬件输出上进行后处理,算法往往需要为硬件缺陷付出较高复杂度代价。未来的重要趋势是硬件—算法协同设计:一方面,算法应更显式地建模脉冲生成机制与器件噪声,并支持针对不同硬件参数的快速适配;另一方面,硬件设计也可考虑为下游算法提供更友好的输出形式,例如多阈值或可变阈值读出、更灵活的时间编码方式以及可编程的片上预处理模块,以降低后端模型负担。协同

设计的关键在于建立可微分或可学习的硬件代理模型,使硬件参数能够在任务目标与系统约束下被联合优化。

9.6 面向闭环系统的感知与决策融合

当前多数脉冲视觉工作仍停留在“开环”评测阶段,即以重建质量或感知精度作为终点。然而在具身智能、自动驾驶以及高速机器人控制等应用中,感知结果往往会直接进入预测、决策与控制闭环,感知误差可能被策略放大,也可能被系统动力学吸收。因此,未来需要从闭环系统视角重新审视脉冲视觉方法,将“感知—预测—决策”作为协同问题进行研究。例如,可以任务成功率、轨迹稳定性以及安全约束满足率等系统级目标作为优化目标,推动评测体系从像素级或特征级指标进一步扩展到系统级指标。同时,脉冲相机的低时延特性为快速反馈控制提供了天然优势,这也使其在机器人与具身智能场景中具有重要潜力。相应地,闭环应用对鲁棒性与不确定性估计提出更高要求,未来模型需要能够输出可信度信息,并支持风险敏感的决策融合。

总体来看,脉冲视觉的未来发展不应仅理解为某一类算法性能的持续改进,而更应视为一个由传感器、表征、算法、硬件和应用共同演进的研究生态。只有在连续时间统一表征、大规模数据与学习范式、多模态协同、系统指标驱动优化以及闭环应用等方面形成协同突破,脉冲视觉才有望从当前的新兴研究方向进一步发展为面向真实世界复杂任务的连续时间视觉计算范式。

参考文献 (References)

- Adelson E H and Bergen J R. 1991. The plenoptic function and the elements of early vision//Landy M, Movshon J A, eds. *Computational Models of Visual Processing*. Cambridge, USA: The MIT Press: 3-20 [DOI: 10.7551/mitpress/2002.003.0004]
- Bernardin K and Stiefelhagen R. 2008. Evaluating multiple object tracking performance: the clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008(1): #246309 [DOI: 10.1155/2008/246309]
- Bi G Q and Poo M M. 1998. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience*, 18(24): 10464-10472 [DOI: 10.1523/JNEUROSCI.18-24-10464.1998]
- Bub G, Tecza M, Helmes M, Lee P and Kohl P. 2010. Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging.

- Nature Methods, 7(3): 209-211 [DOI: 10.1038/nmeth.1429]
- Chamberlain S G. 1969. Photosensitivity and scanning of silicon image detector arrays. *IEEE Journal of Solid-State Circuits*, 4(6): 333-342 [DOI: 10.1109/JSSC.1969.1050032]
- Chang Y K, Xiaokaiti Y, Liu Y J, Fan B, Huang Z J, Huang T J, et al. 2024. Towards HDR and HFR video from rolling-mixed-bit splittings//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 25117-25127 [DOI: 10.1109/CVPR52733.2024.02373]
- Chen K, Chen S Y, Zhang J Y, Zhang B Y, Zheng Y J, Huang T J, et al. 2024. SpikeReveal: unlocking temporal sequences from real blurry inputs with spike streams//*Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 62673-62696
- Chen K, Zhang J Y, Hao Z C, Zheng Y J, Huang T J and Yu Z F. 2025. USP-Gaussian: unifying spike-based image reconstruction, pose correction and Gaussian splatting//*Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 16609-16618 [DOI: 10.1109/CVPR52734.2025.01548]
- Chen S Y, Zhang J Y, Zheng Y J, Huang T J and Yu Z F. 2023. Enhancing motion deblurring in high-speed scenes with spike streams//*Proceedings of the 37th International Conference on Neural Information Processing Systems*. New Orleans, USA: Curran Associates Inc.: 70279-70292
- Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, et al. 2016. The cityscapes dataset for semantic urban scene understanding//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 3213-3223 [DOI: 10.1109/CVPR.2016.350]
- Dai G L, Wang Z Y, Xu Q W, Lu M, Chen W, Shi B X, et al. 2024. SpikeNVS: enhancing novel view synthesis from blurry images via spike camera[EB/OL]. [2026-02-28]. <https://arxiv.org/pdf/2404.06710.pdf>
- Dong S W, Huang T J and Tian Y H. 2017. Spike camera and its coding methods//*Proceedings of the 2017 Data Compression Conference*. Snowbird, USA: IEEE: #437 [DOI: 10.1109/DCC.2017.69]
- Dong Y C, Xiong R Q, Fan X P, Yu Z F, Tian Y H and Huang T J. 2025b. Self-supervised learning for color spike camera reconstruction//*Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 6231-6240 [DOI: 10.1109/CVPR52734.2025.00584]
- Dong Y C, Xiong R Q, Fan X P, Zhu S Y, Wang J and Huang T J. 2025a. Dynamic scene reconstruction for color spike camera via zero-shot learning. *IEEE Transactions on Computational Imaging*, 11: 129-141 [DOI: 10.1109/TCL.2025.3527156]
- Dong Y C, Xiong R Q, Zhang J, Yu Z F, Fan X P, Zhu S Y, et al. 2024c. Super-resolution reconstruction from Bayer-pattern spike streams//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 24871-24880 [DOI: 10.1109/CVPR52733.2024.02349]
- Dong Y C, Xiong R Q, Zhao J, Zhang J, Fan X P, Zhu S Y, et al. 2024a. Joint demosaicing and denoising for spike camera//*Proceedings of the 38th AAAI Conference on Artificial Intelligence*. Vancouver, Canada: AAAI: 1582-1590 [DOI: 10.1609/aaai.v38i2.27924]
- Dong Y C, Xiong R Q, Zhao J, Zhang J, Fan X P, Zhu S Y, et al. 2024b. Learning a deep demosaicing network for spike camera with color filter array. *IEEE Transactions on Image Processing*, 33: 3634-3647 [DOI: 10.1109/TIP.2024.3403050]
- Dong Y C, Zhao J, Xiong R Q and Huang T J. 2022. High-speed scene reconstruction from low-light spike streams//*Proceedings of 2022 IEEE International Conference on Visual Communications and Image Processing*. Suzhou, China: IEEE: 1-5 [DOI: 10.1109/VCIP56404.2022.10008850]
- Ester M, Kriegel H P, Sander J and Xu X W. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise//*Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*. Portland, Oregon, USA: AAAI Press: 226-231 [DOI: 10.5555/3001460.3001507]
- Fan B, Yin J Y, Dai Y C, Xu C, Huang T J and Shi B X. 2024. Spatio-temporal interactive learning for efficient image reconstruction of spiking cameras// *Proceedings of the 38th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc.: 21401-21427
- Fang S K, Shen I C, Igarashi T, Wang Y F, Wang Z S, Yang Y, et al. 2025. NeRF is a valuable assistant for 3D Gaussian splatting//*Proceedings of IEEE/CVF International Conference on Computer Vision*. Hawaii, USA: IEEE: 26230-26240
- Gehrig M, Aarents W, Gehrig D and Scaramuzza D. 2021. DSEC: a stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 6(3): 4947-4954 [DOI: 10.1109/LRA.2021.3068942]
- Guo Y J, Bai Y X, Hu L W, Liu M Z, Guo Z Y, Ma L, et al. 2024. Spike-NeRF: neural radiance field based on spike camera//*Proceedings of 2024 IEEE International Conference on Multimedia and Expo (ICME)*. Niagara Falls, Canada: IEEE: 1-6 [DOI: 10.1109/ICME57554.2024.10687382]
- Guo Y J, Hu L W, Bai Y X, Yao J W, Ma L and Huang T J. 2025. SpikeGS: reconstruct 3D scene captured by a fast-moving bio-inspired camera//*Proceedings of the 39th AAAI Conference on Artificial Intelligence*. Philadelphia, USA: AAAI: 3293-3301 [DOI: 10.1609/aaai.v39i3.32340]
- Hu L W, Ding Z L, Liu M Z, Ma L and Huang T J. 2025a. Learning to robustly reconstruct dynamic scenes from low-light spike streams// *Proceedings of the 18th European Conference on Computer Vision*. Milan, Italy: Springer: 88-105 [DOI: 10.1007/978-3-031-72643-9_6]

- Hu L W, Li Y, Liu M Z, Guo Y J, Xie S H, Ding Z L, et al. 2025b. Seeing the unseen in low-light spike streams [EB/OL]. [2026-02-28]. <https://arxiv.org/pdf/2509.23304.pdf>
- Hu L W, Zhao R, Ding Z L, Ma L, Shi B X, Xiong R Q, et al. 2022. Optical flow estimation for spiking camera//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 17823-17832 [DOI: 10.1109/CVPR52688.2022.01732]
- Huang T J. 2022. Spiking continuous photographing principle and demonstration on ultrahigh speed and high dynamic imaging. *Acta Electronica Sinica*, 50(12): 2919-2927 (黄铁军. 2022. 脉冲连续摄影原理与超高速高动态成像验证. *电子学报*, 50(12): 2919-2927) [DOI: 10.12263/DZXB.20221075]
- Huang T J, Yu Z F, Li Y, Shi B X, Xiong R Q, Ma L, et al. 2022. Advances in spike vision. *Journal of Image and Graphics*, 27(6): 1823-1839 (黄铁军, 余肇飞, 李源, 施柏鑫, 熊瑞勤, 马雷, 等. 2022. 脉冲视觉研究进展. *中国图象图形学报*, 27(6): 1823-1839) [DOI: 10.11834/jig.220175]
- Huang T J, Zheng Y J, Yu Z F, Chen R, Li Y, Xiong R Q, et al. 2023. 1000× faster camera and machine vision with ordinary devices. *Engineering*, 25: 110-119 [DOI: 10.1016/j.eng.2022.01.012]
- Kalman R E. 1960. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1): 35-45 [DOI: 10.1115/1.3662552]
- Kerbl B, Kopanas G, Leimkühler T and Drettakis G. 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4): #139 [DOI: 10.1145/3592433]
- Kuhn H W. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1/2): 83-97 [DOI: 10.1002/nav.3800020109]
- Li J N, Wang X, Zhu L, Li J, Huang T J and Tian Y H. 2022. Retinomorphic object detection in asynchronous visual streams//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual Event: AAAI Press: 1332-1340 [DOI: 10.1609/AAAI.V36I2.20021]
- Liu R N, Zhu L, Xiang X J, Wang L Z and Huang H. 2025a. Noise-modeled diffusion models for low-light spike image restoration//Proceedings of 2025 IEEE/CVF International Conference on Computer Vision. Hawaii, USA: IEEE: 4080-4089
- Liu W X, Deng Y, Chen K, Zhong X, Yu Z F and Huang T J. 2025b. SOTA: spike-navigated optimal transport saliency region detection in composite-bias videos//Proceedings of the 34th International Joint Conference on Artificial Intelligence. Montreal, Canada: Morgan Kaufmann: 1621-1629 [DOI: 10.24963/ijcai.2025/181]
- Liu Z Y, Su Q, Tang L L, Yu Z F and Huang T J. 2025c. Driving in spikes: an entropy-guided object detector for spike cameras [EB/OL]. [2026-02-28]. <https://arxiv.org/pdf/2511.15459.pdf>
- Maass W. 1997. Networks of spiking neurons: the third generation of neural network models. *Neural Networks*, 10(9): 1659-1671 [DOI: 10.1016/S0893-6080(97)00011-7]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2020. NeRF: representing scenes as neural radiance fields for view synthesis//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 405-421 [DOI: 10.1007/978-3-030-58452-8_24]
- Nayar S K and Ben-Ezra M. 2004. Motion-based motion deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6): 689-698 [DOI: 10.1109/TPAMI.2004.1]
- Nelson P. 2017. From Photon to Neuron: Light, Imaging, Vision. Princeton, USA: Princeton University Press
- Sick V. 2013. High speed imaging in fundamental and applied combustion research. *Proceedings of the Combustion Institute*, 34(2): 3509-3530 [DOI: 10.1016/j.proci.2012.08.012]
- Sim H, Oh J and Kim M. 2021. XVFI: extreme video frame interpolation//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14469-14478 [DOI: 10.1109/ICCV48922.2021.01422]
- Teng M G, Lou H Y, Yang Y X, Huang T J and Shi B X. 2024. Hybrid all-in-focus imaging from neuromorphic focal stack. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 10124-10137 [DOI: 10.1109/TPAMI.2024.3433607]
- Tsodyks M, Pawelzik K and Markram H. 1998. Neural networks with dynamic synapses. *Neural Computation*, 10(4): 821-835 [DOI: 10.1162/089976698300017502]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Weckler G P. 1967. Operation of p-n junction photodetectors in a photon flux integrating mode. *IEEE Journal of Solid-State Circuits*, 2(3): 65-73 [DOI: 10.1109/JSSC.1967.1049795]
- Xia L J, Zhao J, Xiong R Q and Huang T J. 2023. SVFI: spiking-based video frame interpolation for high-speed motion//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI Press: 2910-2918 [DOI: 10.1609/AAAI.V37I3.25393]
- Ye Q H, Chang Y Q, Li J N, Xu H R, Wang X, Zhang W, et al. 2025. Spike4DGS: towards high-speed dynamic scene rendering with 4D Gaussian splatting via a spike camera array//Proceedings of the 39th Conference on Neural Information Processing Systems. San Diego, USA: Curran Associates Inc.: 1-14
- Yu J Z, Peng X, Lu Z D, Kneip L and Wang Y Q. 2025. SpikeGS: learning 3D Gaussian fields from continuous spike stream//Proceedings of the 17th Asian Conference on Computer Vision. Hanoi, Vietnam: Springer: 159-177 [DOI: 10.1007/978-981-96-0972-7_10]
- Zhang J Y, Chen K, Chen S Y, Zheng Y J, Huang T J and Yu Z F. 2024a. SpikeGS: 3D Gaussian splatting from spike streams with

- high-speed camera motion//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia; ACM: 9194-9203 [DOI: 10.1145/3664647.3680996]
- Zhang J Y, Chen S Y, Zheng Y J, Yu Z F and Huang T J. 2024b. Spike-guided motion deblurring with unknown modal spatiotemporal alignment//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 25047-25057 [DOI: 10.1109/CVPR52733.2024.02366]
- Zhang J Y, Chen S Y, Zheng Y J, Yu Z F and Huang T J. 2024c. Transient glimpses: unveiling occluded backgrounds through the spike camera//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 637-645 [DOI: 10.1609/aaai.v38i1.27820]
- Zhang J Y, Jia S S, Yu Z F and Huang T J. 2023. Learning temporal-ordered representation for spike streams based on discrete wavelet transforms//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 137-147 [DOI: 10.1609/AAAI.V37I1.25085]
- Zhang J Y, Tang L L, Yu Z F, Lu J W and Huang T J. 2022. Spike transformer: monocular depth estimation for spiking camera//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel; Springer: 34-52 [DOI: 10.1007/978-3-031-20071-7_3]
- Zhang J Y, Zheng Y J, Yu Z F and Huang T J. 2024. Spike-based vision for autonomous driving scenarios. *Strategic Study of CAE*, 26(1): 160-177 (张济远, 郑雅菁, 余肇飞, 黄铁军. 2024. 面向自动驾驶场景的脉冲视觉研究. *中国工程科学*, 26(1): 160-177) [DOI: 10.15302/J-SSCAE-2024.01.012]
- 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]
- Zhao J, Xiong R Q and Huang T J. 2020. High-speed motion scene reconstruction for spike camera via motion aligned filtering//2020 IEEE International Symposium on Circuits and Systems. Seville, Spain; IEEE: 1-5 [DOI: 10.1109/ISCAS45731.2020.9181055]
- Zhao J, Xiong R Q, Liu H F, Zhang J and Huang T J. 2021. Spk2ImgNet: learning to reconstruct dynamic scene from continuous spike stream//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA; IEEE: 11991-12000 [DOI: 10.1109/CVPR46437.2021.01182]
- Zhao J, Xiong R Q, Xie J Y, Shi B X, Yu Z F, Gao W, et al. 2022c. Reconstructing clear image for high-speed motion scene with a retina-inspired spike camera. *IEEE Transactions on Computational Imaging*, 8: 12-27 [DOI: 10.1109/TCL.2021.3136446]
- Zhao J W, Ye J M, Zhang S L, Yu Z F and Huang T J. 2023. Recognizing high-speed moving objects with spike camera//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada; ACM: 7657-7665 [DOI: 10.1145/3581783.3612054]
- Zhao J W, Yu Z F, Ma L, Ding Z L, Zhang S L, Tian Y H, et al. 2022a. Modeling the detection capability of high-speed spiking cameras//Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore, Singapore: IEEE: 4653-4657 [DOI: 10.1109/ICASSP43922.2022.9747018]
- Zhao J W, Zhang S L, Ma L, Yu Z F and Huang T J. 2022b. Spiking-SIM: a bio-inspired spiking simulator//Proceedings of 2022 IEEE International Symposium on Circuits and Systems. Austin, USA: IEEE: 3003-3007 [DOI: 10.1109/ISCAS48785.2022.9937811]
- Zhao J W, Zhang S L, Yu Z F and Huang T J. 2024d. SpiReco: fast and efficient recognition of high-speed moving objects with spike camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7): 5856-5867 [DOI: 10.1109/TCSVT.2023.3272375]
- Zhao R, Xiong R Q, Wang D K, Xuan S Y, Zhang J, Fan X P, et al. 2026. Spike camera optical flow estimation based on continuous spike streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4): 4756-4770 [DOI: 10.1109/TPAMI.2025.3649050]
- Zhao R, Xiong R Q, Zhang J, Yu Z F, Zhu S Y, Ma L, et al. 2024a. Spike camera image reconstruction using deep spiking neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(6): 5207-5212 [DOI: 10.1109/TCSVT.2023.3326294]
- Zhao R, Xiong R Q, Zhang J, Zhang X F, Yu Z F and Huang T J. 2024b. Optical flow for spike camera with hierarchical spatial-temporal spike fusion//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 7496-7504 [DOI: 10.1609/aaai.v38i7.28581]
- Zhao R, Xiong R Q, Zhao J, Yu Z F, Fan X P and Huang T J. 2022d. Learning optical flow from continuous spike streams//Proceedings of the 36th Conference on Neural Information Processing Systems. New Orleans, USA; Curran Associates Inc.: 7905-7920 [DOI: 10.52202/068431-0574]
- Zhao R, Xiong R Q, Zhao J, Zhang J, Fan X P, Yu Z F, et al. 2024c. Boosting spike camera image reconstruction from a perspective of dealing with spike fluctuations//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 24955-24965 [DOI: 10.1109/CVPR52733.2024.02357]
- Zheng Y J, Li C E, Zhang J Y, Yu Z F and Huang T J. 2026a. SNN-Tracker: online high-speed multi-object tracking with spike camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(1): 624-638 [DOI: 10.1109/TPAMI.2025.3610696]
- Zheng Y J, Yu Z F, Wang S and Huang T J. 2023a. Spike-based motion estimation for object tracking through bio-inspired unsupervised learning. *IEEE Transactions on Image Processing*, 32: 335-349 [DOI: 10.1109/TIP.2022.3228168]
- Zheng Y J, Zhang J Y, Zhao R, Ding J H, Chen S Y, Wu W J, et al. 2026b. SpikeCV: open a continuous computer vision era. *Science*

- China Information Sciences, 69(3): #132106 [DOI: 10.1007/s11432-023-4565-6]
- Zheng Y J, Zhao R, Zhang J Y, Ding J H, Chen S T and Huang T J. 2022. SpikeCV open-source framework repository [EB/OL]. <https://openi.pcl.ac.cn/Cordium/SpikeCV>
- Zheng Y J, Zhao R, Zhang J Y, Ding J H, Chen S Y and Huang T J. 2023b. SpikeCV: an open platform for spike vision research [EB/OL]. <https://spikecv.github.io/>
- Zheng Y J, Zheng L X, Yu Z F, Shi B X, Tian Y H and Huang T J. 2021. High-speed image reconstruction through short-term plasticity for spiking cameras//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 6354-6363 [DOI: 10.1109/CVPR46437.2021.00629]
- Zhu L, Chen X Z, Wang X and Huang H. 2024a. Finding visual saliency in continuous spike stream//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 7757-7765 [DOI: 10.1609/aaai.v38i7.28610]
- Zhu L, Dong S W, Huang T J and Tian Y H. 2019. A retina-inspired sampling method for visual texture reconstruction//Proceedings of 2019 IEEE International Conference on Multimedia and Expo. Shanghai, China: IEEE: 1432-1437 [DOI: 10.1109/ICME.2019.00248]
- Zhu L, Dong S W, Li J N, Huang T J and Tian Y H. 2020. Retina-like visual image reconstruction via spiking neural model//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1435-1443 [DOI: 10.1109/CVPR42600.2020.00151]
- Zhu L, Jia K M, Zhao Y F, Qi Y S, Wang L Z and Huang H. 2024b. SpikeNeRF: learning neural radiance fields from continuous spike stream//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 6285-6295 [DOI: 10.1109/CVPR52733.2024.00601]
- Zhu L, Jia K M, Zhao Y F, Qi Y S, Wang L Z and Huang H. 2026. DSNeRF: dynamic view synthesis for ultra-fast scenes from continuous spike streams. IEEE Transactions on Pattern Analysis and Machine Intelligence: 1-18 [DOI: 10.1109/TPAMI.2026.3656825]
- Zhu L, Li J N, Wang X, Huang T J and Tian Y H. 2021. NeuSpikeNet: high speed video reconstruction via bio-inspired neuromorphic cameras//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2380-2389 [DOI: 10.1109/ICCV48922.2021.00240]
- Zhu L and Tian Y H. 2023. Review of visual reconstruction methods of retina-like vision sensors. Scientia Sinica Informationis, 53(3): 417-436 (朱林, 田永鸿. 2023. 仿视网膜传感器视觉重建算法研究综述. 中国科学: 信息科学, 53(3): 417-436) [DOI: 10.1360/SSI-2021-0397]
- Zhu L, Yan W Q, Chang Y, Tian Y H and Huang H. 2025. Simultaneous learning intensity and optical flow from high-speed spike stream. IEEE Transactions on Circuits and Systems for Video Technology, 35(6): 5126-5139 [DOI: 10.1109/TCSVT.2024.3516478]
- Zhu L, Zheng Y L, Geng M Y, Wang L Z and Huang H. 2023. Recurrent spike-based image restoration under general illumination//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 8251-8260 [DOI: 10.1145/3581783.3611829]
- Zhu Y Y, Zhang Y, Xie X D and Huang T J. 2022. An FPGA accelerator for high-speed moving objects detection and tracking with a spike camera. Neural Computation, 34(8): 1812-1839 [DOI: 10.1162/neco_a_01507]

作者简介

郑雅菁,女,博雅博士后,主要研究方向为类脑视觉计算、计算机视觉。E-mail: yj.zheng@pku.edu.cn

黄铁军,通信作者,男,教授,主要研究方向为视觉信息处理、类脑计算、脉冲视觉。E-mail: tjhuang@pku.edu.cn

赵睿,男,博士后研究员,主要研究方向为计算摄像、计算机视觉。E-mail: ruizhao26@gmail.com

朱林,男,副教授,主要研究方向为计算机视觉。

E-mail: linzhu@bnu.edu.cn

刘俣伽,女,博雅博士后,主要研究方向为神经网络的对抗鲁棒性、脉冲视觉。E-mail: yujia_liu@pku.edu.cn