

面向自动驾驶场景的脉冲视觉研究

张济远¹, 郑雅菁¹, 余肇飞^{1,2}, 黄铁军^{1,2*}

(1. 视频与视觉技术国家工程研究中心, 北京 100871; 2. 北京大学人工智能研究院, 北京 100871)

摘要: 自动驾驶是计算机视觉重要的研究方向, 具有广阔的应用前景。纯视觉感知方案在自动驾驶场景中具有重要的研究价值。不同于传统相机, 脉冲视觉传感器能更加灵敏地感受光子, 具备比传统视频快千倍以上的成像速度, 具有高时间分辨率、高动态范围、低数据冗余和低功耗等优势。本文面向自动驾驶场景, 总结了脉冲相机的成像原理、感知能力与优势; 围绕自动驾驶相关视觉任务, 详细介绍了脉冲视觉影像重建原理与方法, 讨论了基于传感器融合的影像增强技术路线; 归纳总结了基于脉冲相机的运动光流估计、目标识别检测分割与跟踪, 以及三维场景深度估计算法进展及技术路线; 梳理了脉冲相机数据及感知系统的发展现状, 分析了脉冲视觉的研究挑战; 研究提出了潜在解决方案及未来研究方向。脉冲相机及其算法和系统在自动驾驶领域具有巨大潜力, 是未来计算机视觉的主要研究方向之一。

关键词: 脉冲视觉; 脉冲相机; 自动驾驶; 人工智能

中图分类号: TP752 **文献标识码:** A

Spike-Based Vision for Autonomous Driving Scenarios

Zhang Jiyuan¹, Zheng Yajing¹, Yu Zhaofei^{1,2}, Huang Tiejun^{1,2*}

(1. National Engineering Research Center of Visual Technology, Beijing 100871, China;
2. Institute of Artificial Intelligence, Peking University, Beijing 100871, China)

Abstract: Autonomous driving is an important research direction in computer vision which has broad application prospects. Pure vision perception schemes have significant research value in autonomous driving scenarios. Different from traditional cameras, spike vision sensor offers imaging speeds over a thousand times faster than traditional cameras, possess advantages such as high temporal resolution, high dynamic range, low data redundancy, and low power consumption. This study focuses on autonomous driving scenarios, introducing the imaging principles, perception capabilities, and advantages of the spike camera. Besides, focusing on visual tasks related to autonomous driving, this study elaborates on the principles and methods of spike-based image/video reconstruction, discusses the approach to image enhancement based on sensor fusion with spike cameras, and provides a detailed description of the algorithms and technical routes for motion optical flow estimation, object recognition, detection, segmentation, and tracking, and deep estimation of three-dimensional scenes based on spike cameras. It also summarizes the development of the spike camera data and systems. At last, it analyzes the challenges, potential solutions, and future directions for spike vision research. Spike cameras and their algorithms and systems hold great potentials in the field of autonomous driving and represent one of the future research directions in computer vision.

Keywords: spike-based vision; spike camera; autonomous driving; artificial intelligence

收稿日期: 2023-11-14; 修回日期: 2024-01-05

通讯作者: *黄铁军, 视频与视觉技术国家工程研究中心教授, 主要研究方向为视觉信息处理和类脑智能; E-mail: tjhuang@pku.edu.cn

资助项目: 中国工程院咨询项目“新一代人工智能及产业集群发展战略研究”(2022-PP-07)

本刊网址: www.engineering.org.cn/ch/journal/sscae

一、前言

自动驾驶系统采用先进的通信、计算机、网络和控制技术，对车辆实现实时、连续控制。智能的自动驾驶系统可有效地减少事故发生、缓解拥堵、提高道路运输能力、减少驾驶压力等。走向L4、L5级别的自动驾驶时代是未来发展趋势，近年来，围绕自动驾驶系统的相关研究取得了巨大进展。视觉相机、雷达、激光雷达等传感器在自动驾驶感知系统中均起到重要作用。自动驾驶性能在很大程度上取决于感知系统。目前感知系统的主流技术方案有：以视觉相机为主的多传感器融合方案；以激光雷达为主导、其他传感器为辅助的技术方案；未来可能发展至纯视觉感知的方案。视觉传感器作为直接感知场景的信号源，在辅助理解驾驶场景方面发挥关键作用。

基于传统快门相机的视觉算法研究已取得巨大进展，并达到可观性能。传统相机采用基于曝光的成像原理。曝光时间的长短将导致过曝或欠曝的成像问题，动态范围较低。此外，离散的曝光过程使帧之间缺乏连续性，丢失了时域信息。

神经形态视觉的诞生旨在模拟生物视觉的感知原理及结构，相关硬件、软件、算法及系统成为研究和应用重点。近年来，神经形态视觉不断发展，新型视觉传感器^[1-6]被研发出来。主流传感器按照采样原理可以划分为差分型传感器和积分型传感器。差分型传感器以动态视觉传感器^[1-4,7-10]为代表，感知每个像素的相对光强变化。然而动态视觉传感器只记录像素的相对亮度变化，对运动区域敏感而对静止区域或纹理较弱区域不敏感。积分型传感器以脉冲相机为代表^[6]，其可以有效记录绝对光强，同时提供静态和动态场景信息。脉冲相机的每个像素传感器独立并持续地接收光子，并仅在累积的光子达到阈值时触发脉冲。因此，脉冲相机可以高时间

分辨率产生连续脉冲流。脉冲流可以用来在任何给定时刻重建高质量场景影像。

动态视觉传感器发展较早，研究内容丰富^[11,12]。早期的影像重建研究多使用优化^[13,14]、正则化^[15]及时间滤波^[16,17]等方案，之后深度学习（DL）方案占据主流^[18-28]。在场景理解任务中，使用动态视觉传感器解决光流估计或目标跟踪任务有一定天然优势，事件流的产生往往代表着运动的边缘^[29]，研究多基于此展开^[30-36]。目前也有许多研究针对深度估计展开以加强对传感器三维场景的理解^[37-56]。

脉冲相机的采样机制模拟生物视觉感知系统，其输出数据以脉冲流形式呈现，具有高时域分辨率、低数据冗余、低功耗和高动态范围的优势^[57,58]。这些优势弥补了传统相机感知外部场景的缺陷，从而可以解决自动驾驶场景中的挑战性难题。

本文聚焦自动驾驶场景，介绍脉冲视觉原理，分析脉冲视觉在驾驶场景中的感知能力与优势；总结脉冲影像重建方法，讨论传感器融合技术路线；围绕运动估计、目标感知、三维场景感知论述脉冲视觉场景感知方案。此外，本文梳理了脉冲视觉感知系统进展，分析了自动驾驶场景中面临的挑战，提出了未来研究方向。

二、脉冲视觉传感器的成像原理与优势

（一）脉冲视觉传感器的成像原理

受灵长类动物视网膜中央凹采样机制的启发^[59,60]，脉冲相机直接记录具有时空特征的光强信息。它以脉冲形式输出二值流，其数据仅用0或1表示。脉冲相机主要由感光器、积分器和比较器三个部件组成，如图1所示。

感光像素阵列在空间上排列在脉冲相机的感光器上，连续捕获光子。接着，积分器持续将光信号

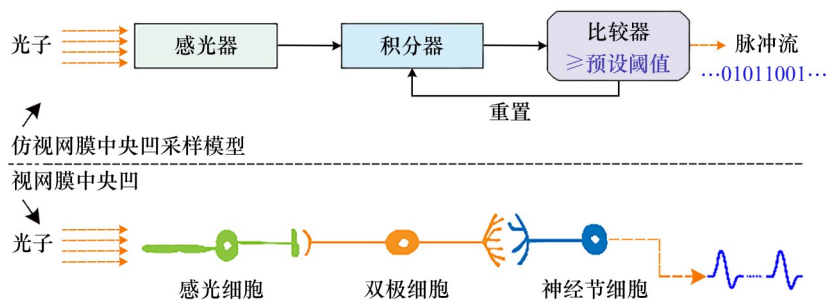


图1 脉冲相机采样原理示意图

转换为电信号，以累积每个单元格的电压。比较器检测累积电压是否达到预设电压阈值 θ 。当达到阈值时，脉冲被触发，电压将重置为预设值。一个像素上的脉冲生成过程可以表述为：

$$\int_{t_{i-1}}^{t_i} \alpha I(t) dt = \theta, \quad (1)$$

式(1)中， $I(t)$ 描述了光强度， t_i 和 t_{i-1} 分别表示第 i 个和第 $(i-1)$ 个脉冲的触发时间， α 是光电转换率。由于硬件电路的限制，输出电路中的单元在固定间隔 $\Delta t = 25 \mu s$ 内周期性地读出脉冲作为离散时间信号 $s(x, y, n)$ 。如果空间坐标 (x, y) 处的像素在时间 t 处触发脉冲，则将读出脉冲 $s(x, y, n) = 1 (n = 1, 2, \dots)$ ，其中 $(n-1)\Delta t < t \leq n\Delta t$ ，否则它将读出 $s(x, y, n) = 0$ 。传感器使用高速轮询在每个离散时间戳 n 处生成大小为 $H \times W$ 的脉冲帧。在固定间隔 $\Delta t \cdot T$ 内，相机将产生一段脉冲流 $S = \{s(x, y, t)\}_{t=1}^T$ ，其尺寸为 $H \times W \times T$ 。

图2所示为脉冲相机相较于动态视觉传感器在相同光变化情况下的成像原理示意图。

(二) 脉冲视觉感知能力与优势

传统相机基于曝光原理成像，无法摆脱帧的概念，图3直接展示了脉冲与图像/视频的成像区别。在传统相机连续摄影过程中，每帧曝光时间小于两帧时间间隔。传统成像原理带来两个问题：第一，在类似驾驶等连续摄影的场景中，曝光时间的长短将影响成像质量进而影响感知能力，曝光时间过长将使场景过曝或引入运动模糊，曝光时间过短则欠曝或引入更多噪声；第二，由于曝光时间小于帧间隔，传统相机的成像过程并不是连续的，丢失了

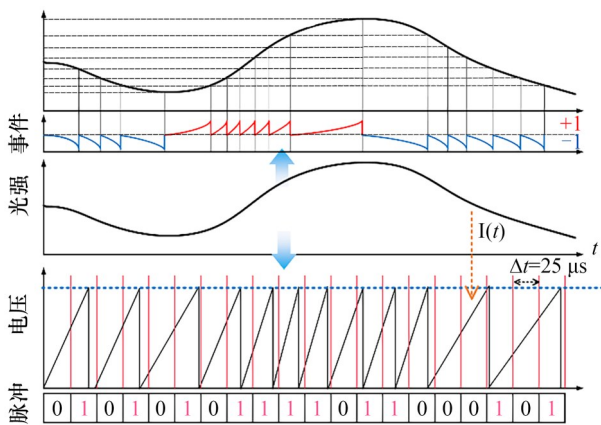


图2 脉冲视觉传感器脉冲产生原理图

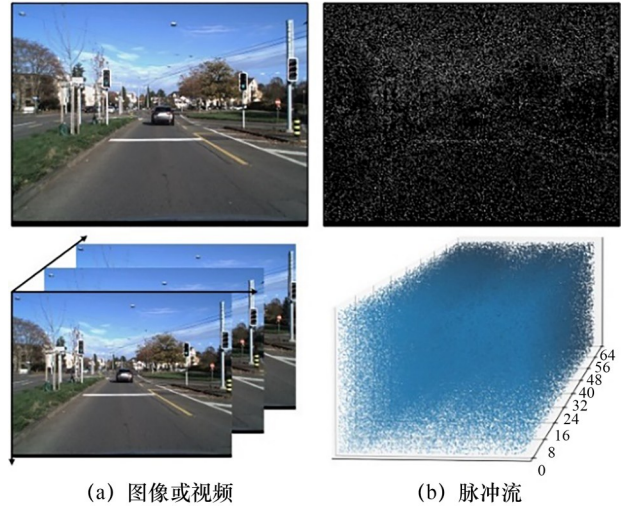


图3 传统相机与脉冲相机的成像效果图

时间连续性，致使感知算法无法充分利用时间上的特征。自动驾驶中许多复杂场景为传统视觉感知算法带来挑战，在路况复杂、光线变化剧烈、光线不足或存在高速运动的场景中，传统相机的感知会出现错误，存在数据采样冗余大、动态范围小和时间分辨率低等问题。

在适应极端环境能力方面，脉冲相机相较传统相机存在感知优势。传统相机因存在上述成像上的特点，在驾驶场景中会导致感知能力上的局限性。然而，图4所示的复杂驾驶场景中脉冲相机可以更好地感知环境，图中三个示例分别呈现车辆高速驶过时类似鬼探头的场景、车辆颠簸场景及光线较强时的场景。所示三种场景均使传统相机感知质量下降，产生模糊或过曝效应；而脉冲相机可以较好地恢复这些挑战性驾驶场景的高清、高动态影像。

图4所示的室外交通环境中，街道监控或路口车辆交汇等具体场景常常出现高速车辆。对于此类车辆，传统相机产生的剧烈模糊效应使其无法有效感知环境中的对象（如图5所示，传统相机拍摄的



(a) 车辆高速驶过 (b) 车辆颠簸模糊 (c) 光线过曝

图4 车外交通环境中脉冲相机与传统相机成像对比

车辆等)。利用脉冲相机的成像优势，高质量影像可从脉冲流中恢复出来，从而辅助高速驾驶场景中的视觉感知。图5展示了直接利用脉冲流可以准确预测高速车辆的运动状态，而传统相机在剧烈模糊下无法有效估计车辆运动。

利用脉冲相机和传统相机的传感器融合方案，可以有效将脉冲流的高时间分辨率和传统相机高空间分辨率的优势结合起来，实现更优的感知能力。例如，传统视觉和脉冲视觉融合可以恢复高动态范围影像^[61]，使感知过程消除过曝、欠曝等极端问题；图6利用脉冲相机消除视频影像中的模糊效应等^[62]。实际应用中，常规速度情况下的驾驶感知可以只使用传统相机，高速或光线极端情况下可以采用传感器融合方案补偿感知能力。

三、脉冲视觉影像重建

自动驾驶场景中快速且高质量的重构算法是非常重要的。传统视觉领域中视觉算法研究丰富，脉冲视觉影像重建算法可以有效地将脉冲流转化为视

频流并输入现有视觉模型中，完成对环境的感知。

脉冲流是连续且不规则的，只包含0和1，人眼或机器无法直接观测和理解场景信息。脉冲流的影像重建是最基础且最重要的视觉任务，也是实现人眼从此类视觉信号中观测场景的基本需求。本章将从两方面阐述此类影像重建任务：直接从传感器中恢复等分辨率、超分辨率或高动态范围影像；借助和传统相机融合方案实现更优影像重建效果。目前脉冲视觉影像重建研究工作丰富，图7展示了脉冲影像重建算法沿时间的发展过程。

(一) 脉冲影像重建算法研究

1. 脉冲流直接影像重建算法

脉冲相机影像重建算法分为直接分析脉冲时序信息、基于优化的方案与基于深度学习的方案。早期研究直接分析脉冲信息，如统计脉冲数量或相邻脉冲间关系等，早期重建算法在此基础上被提出^[63]。对于高速运动，感受视野中的亮度变化很快。为捕捉瞬时亮度，只使用相邻脉冲间隔（ISI）估计一个像素值，可以通过以下等式表示：

$$P_i(r, c) = \frac{C}{ISI(r, c)} \quad (2)$$

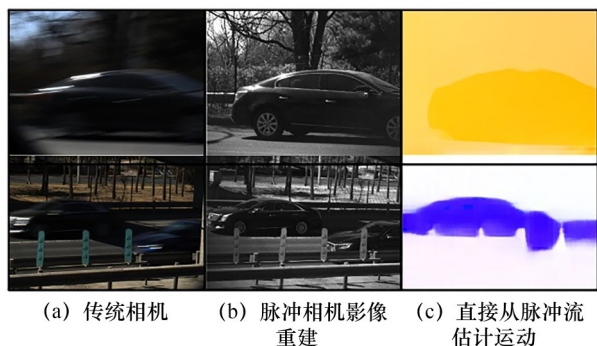


图5 室外交通车辆高速行驶场景中脉冲相机与传统相机运动感知能力对比

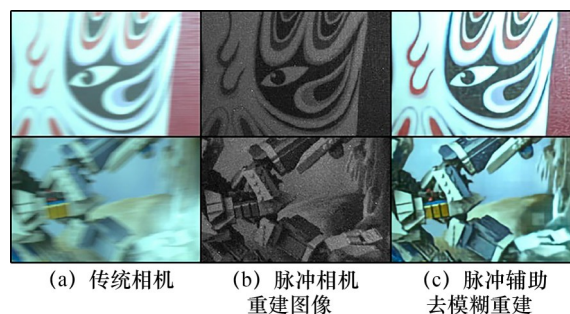


图6 脉冲相机辅助传统相机进行去模糊重建

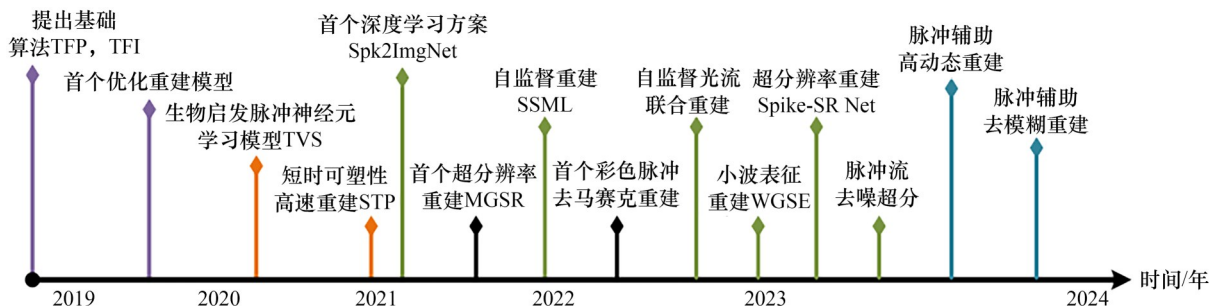


图7 脉冲视觉影像重建研究发展时间轴

注：TFP为窗口重建纹理法；TFI为间隔重建纹理法；TVS为脉冲神经网络重建法；STP为短时可塑性重建法；Spk2ImgNet为“脉冲-图像”网络；MGSR为运动引导的脉冲超分法；SSML为自监督互学习重建法；WGSE为小波引导的脉冲增强算法；Spike-SR Net为“脉冲-超分”网络。

式(2)中, $P_i(r, c)$ 是像素 (r, c) 中 t 时刻的估计像素值。 C 表示最大动态范围。 $ISI(r, c)$ 表示像素 (r, c) 处的脉冲间隔。此方法被称为TFI。对于静止场景, 主要考虑噪声影响。在重建中利用更多脉冲有利于降低噪声。采用较大时间窗口来记录特定时间段内的脉冲信号。TFP方法可用以下公式描述:

$$P_i(r, c) = \frac{C \cdot N_w(r, c)}{w} \quad (3)$$

式(3)中, w 表示回放窗口的尺寸, $N_w(r, c)$ 表示在本时间窗口内 (r, c) 位置像素上的总脉冲发放数量。

然而方法TFI和TFP都存在局限性。图8所示为一个高速运动场景下的TFI(左)和TFP(右)重建结果。TFI方法重建的结果通常会受到脉冲流中的泊松噪声、暗电流噪声等影响而无法输出稳定的信号。TFP方法重建的结果通常会因目标或场景运动幅度较大而产生模糊效应。使用手动超参数部署时间窗口存在局限性, 窗口尺寸可选择直接影像重建图像质量。基于脉冲响应模型, 可为纹理重建过程增加自适应性^[63], 通过动态阈值和脉冲发放后电位的自动调整以适应输入的光电流。动态阈值可以看作是对输入电流特征的学习过程, 因此适用于描述纹理。

影像重建也可建模为优化的过程^[64], 考虑物体的运动并利用光流来对齐不同时刻的脉冲。在运动对齐之后, 可以对信号采用沿运动轨迹的滤波, 以利用时间相关性, 同时不引入运动模糊。

基于DL方案在目前及未来将作为脉冲视觉重建的最佳解决方案之一。目前利用脉冲流直接进行影像重建可归纳为如图9所示的四种模型: 基于卷积神经网络(CNN)的重建模型^[65-67], 基于循环神经网络(RNN)的重建模型^[20,21], 基于Transformer的重建模型^[23], 基于脉冲神经网络(SNN)的重建模型^[24,68,69]。基于CNN的算法考虑输入一段脉冲流并重建中间时刻的清晰图像。端到端模型选择一个参考帧并将其周围的一段连续脉冲流作为输入, 经计



图8 TFI与TFP影像重建效果图

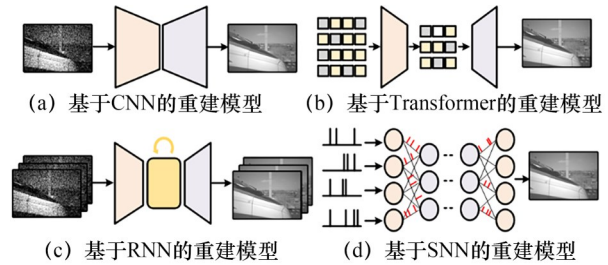


图9 基于深度学习的脉冲流直接影像重建算法方案

算输出参考帧的影像。采用可变性卷积处理脉冲流的相关性并推断光强是处理脉冲流时间相关性的有效策略^[65]。为了适应雨天、黄昏等不同照明条件场景, 结合脉冲相机的物理噪声模型及成像原理, 并结合循环建模方案^[66], 能够普适地将一般光照条件下的脉冲数据重构为高清影像。除此之外, 增强脉冲影像重建普适性也可通过结合小波变换等频域分析策略^[67]实现, 将脉冲流转换为鲁棒的表征形式, 以适应不同场景的重建。

基于RNN的算法致力于输入流式数据并流式地重建连续影像。基础模型设计可以使用如E2VID模型^[20]的U-Net结构及循环模块结合的方案^[70]。为实现序列信息的有效利用, 网络的输入为当前时刻脉冲流表征及此时此刻前若干预测所得的重建图像。然而此方法的参数量与推理速度均不理想, 为此, 可采用全程无下采样顺序连接的卷积层^[21]实现。循环结构的短板是推理过程中的早期重建图像质量偏低, 可通过改进E2VID的循环计算策略解决此问题^[22]。

加入Transformer结构的算法致力于使用非局部的注意力机制深层次挖掘时空关系。结合Transformer和CNN, ET-Net等模型^[23]因有效提取并融合了CNN的局部特征及Transformer的全局特征而获得出众的性能。

SNN是受生物学启发的新一代人工神经网络模型, 是处理视觉任务的新工具, 特别是在脉冲视觉领域有着巨大潜力, 可探索生物系统中智能操作的本质, 并通过数学模型推广到人工智能模型, 融合了受生物启发的高效神经信息处理机制。此外, 其网络架构、配置可更好地匹配脉冲数据(神经元模型、学习规则、仿生结构等)以及拥有理想的低功耗特性, 因此可以解决脉冲视觉感知任务。相比于传统的人工神经网络, 以SNN为基础设计的类视网膜图像重建框架具备更好的生物合理性。采用时间上

连续的脉冲数据，可以在任意采样时刻重建图像，利用SNN中的脉冲时间相关可塑性（STDP）学习规则^[71]驱动网络学习动态与静态区域，使得重建影像同时保留了高速运动和静态背景的细节。

然而，类似上述基于能量的优化方式与神经元模型相结合是耗时的，从而削弱了脉冲相机低延迟的优势。大脑的短时可塑性机制（STP）^[72,73]与脉冲流的动态过程有着强相关性，高速图像重建提供可能性^[69]。使用脉冲流作为输入，该方法推导出每个像素的时变放电频率与突触后神经元动力学之间的关系，并进一步推断场景光强，重建图像的像素值。此外，由于STP模型的动态过程会在脉冲发放率发生变化时在一个稳定值附近波动，因此在STP中引入了运动估计方法来增强重建结果。

为了克服脉冲相机有标签数据的缺少及在真实环境的泛化性问题，自监督新范式^[74]将运动估计与影像重构联合优化，实现无需标签的强泛化性网络框架。此外，彩色脉冲相机的去马赛克算法^[75]，为脉冲相机重建真彩世界打开新的大门。

为了对比不同脉冲影像重建的方法，表1展示了目前脉冲相机主流影像重建方法在不同数据集上的结果对比。其中，真实动态场景数据集（REDS）模拟数据^[65]为利用现有图像数据集模拟生成的脉冲数据，供深度学习模型训练和验证使用，PKU-Spike-High-Speed^[63]为真实世界中所拍摄的高速场景脉冲数据，用于模型的高速重建性能验证。其中模拟数据集使用了峰值信噪比（PSNR）及平均结构相似性（SSIM）作为评价指标，真实数据集使用自然度图像质量评估器（NIQE）及无参考图像空间质量评估器（BRISQUE）作为主观评价指标。

表1 脉冲流直接影像重建算法质量评价对比

重建算法	REDS 模拟数据		PKU-Spike-High-Speed 模拟数据	
	PSNR	SSIM	NIQE	BRISQUE
TFP	22.37	0.580 1	4.01	19.41
TFI	24.94	0.715 0	3.84	22.28
TVS	19.03	0.745 2	36.16	43.46
STP	22.37	0.730 0	10.90	43.35
SSML	34.26	0.971 8	7.39	19.42
Spk2ImgNet	38.44	0.976 7	4.94	28.66
WGSE	38.88	0.977 4	7.45	33.57

2. 脉冲超分辨率影像重建算法

近年来，脉冲超分辨率影像重建研究主要集中在重建相同空间维度的强度图像上，然而与传统相机相比，脉冲相机的时间分辨率要高出三个数量级以上，由于数据存储与传输效率的限制，感光空间分辨率往往不能很大。通过设计算法来从低分辨率（LR）数据重建高分辨率（HR）图像，使此类脉冲流式数据可以更好地还原高分辨率的场景细节，弥补空间分辨率上的不足。

目前利用脉冲流直接进行超分辨率影像重建可归纳为如图10所示的三种：基于优化方案的超分重建、基于RNN的超分重建模型、在脉冲域上直接进行超分。

在运动场景中，通过分析脉冲相机每个脉冲的产生原理，可以求解超空间分辨率下的光强^[76]。一个传感器像素发放的脉冲不仅仅携带真实场景同一点的信息，而是携带一个局部空间上的多个位置的光强信息。这是由相机和物体之间的相对运动引起的，即每个脉冲都可以通过某种权重关系映射到场景中的任意位置上。该工作建模了所有脉冲对每个HR图像上像素位置的权重方程，并优化求解。

近年来的研究多基于深度学习开展。由于场景或相机自身运动的存在，连续的像素值变化给超分提供了更多线索，大部分DL算法都基于RNN建模。利用光流计算不同时刻与参考时刻的像素运动关系

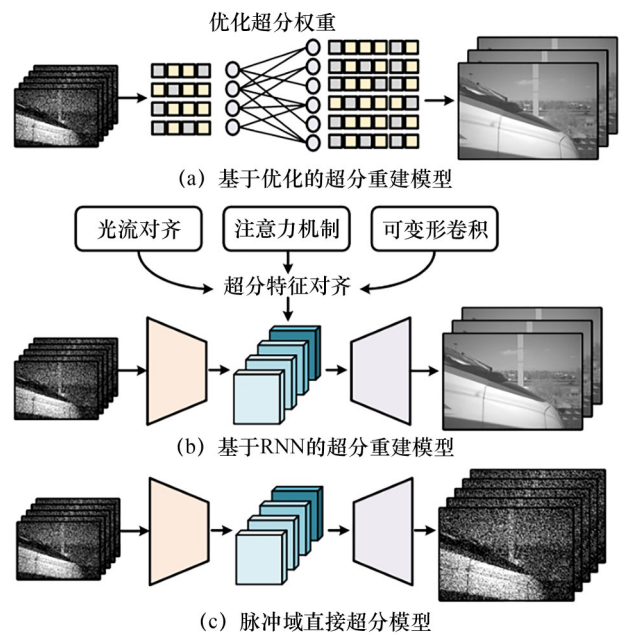


图10 脉冲流超分影像重建算法方案

后，可使用循环网络优化超分结果^[77]。后续研究^[78]使用注意力机制代替光流计算，并加入可变形卷积进行特征对齐，取得了更优的性能。新的超分辨率框架 Spike-SRNet^[79]，采用可变性卷积与循环神经网络结合，建模了脉冲时序上的长程相关性。Xiang 等^[80]直接利用光流估计提取了相邻脉冲流之间的时空相关性。另外，部分研究^[81,82]将脉冲直接转换为高空间分辨率的脉冲流，进而使重构、跟踪等视觉任务均获得性能的提升。也有研究使用无监督方案并利用对抗训练实现超分重建^[27]。

3. 脉冲视觉高动态范围影像重建算法

普通相机受到传统成像原理的限制，其影像或图片的动态范围十分有限。在使用普通摄影设备对存在过暗、过亮部分的极端场景拍摄时，将会产生欠曝和过曝效果，从而损失画面细节。在驾驶场景中，因剧烈光线变化或环境光条件限制，车辆视觉传感设备或因为拍摄细节损失而导致感知能力下降。与传统相机相比，脉冲相机的连续积分采样原理使其具有高动态范围的成像优势，这为解决在黑暗等极端环境中的感知问题提供了有效工具^[26,83]。在连续脉冲流中，脉冲的稠密程度以及脉冲之间的间隔可以直接反映场景的光强信息，对光线较强或较暗部分均可进行较高质量地记录。利用简单 U-Net 网络及对抗训练获得可观效果。借助无监督域自适应方案，可完成低光下增强^[83]。利用注意力机制、循环网络等深度学习算法也可以有效综合时域中的高动态信息^[84,85]。

(二) 视觉传感器融合的重建算法研究

脉冲相机与传统卷帘快门相机的成像原理不同，各自具备互补的优势。传统相机虽然无法以超高速完成摄像，但由于其发展时间长，更好的工艺而具备更优的成像质量（包括色彩质量、空间分辨率等）。在实际应用中，利用脉冲相机辅助进行影像重建及增强是重要的研究方向。传感器融合的方案目前已在去模糊、超分辨率、高动态范围重建、视频插帧等任务取得进展。在自动驾驶场景中，高速场景往往需要应对模糊效应，而光线过强或过弱则需要高动态范围成像解决，本节将分别从去模糊及高动态范围重建展开介绍。

1. 脉冲流辅助去模糊算法

本节总结归纳在脉冲相机上的去模糊技术路线

及基本思路，如图 11 所示。首先，解决如何使用算法融合模糊图像与脉冲流进行去模糊；其次，解决真实影像中每帧图像的模糊情况。

在建模模糊图像重建过程后，可以通过深度神经网络提取脉冲流中的运动纹理特征并在特征域融合两个模态信息^[62]。此外，脉冲流的稠密时序提供了物体运动方向的先验，可以通过脉冲流学习额外的运动线索以对齐特征。运动估计也可以结合模糊一致性和光度一致性实现自监督方案^[86]。

在实际应用中，视频影像每帧的模糊程度是不同的，且场景不一定总是存在模糊，这导致将模糊程度一致的图像作为输入的方案失效。为此，可独立设计模糊检测模块以驱动模型自动寻找序列中不模糊的图像帧^[87]，用以辅助训练网络。除了不同帧直接的模糊程度不同之外，同一帧不同区域像素具备不同模糊程度的情况^[88]。后续研究继续将输入条件松弛，可以做到输入任意程度模糊的视频影像^[89]，同时考虑到训练数据和测试数据之间的时空分辨率不同时的泛化性问题^[90]。大多数研究都基于两种传感器坐标系是对齐的假设，因为供训练的数据集是基于图像集合成的。可通过学习空间对齐机制以支持两种模态输入的坐标系不同轴的情况^[91]。

2. 脉冲流辅助高动态范围重建算法

在自动驾驶场景中，在夜间或正午时段行驶，

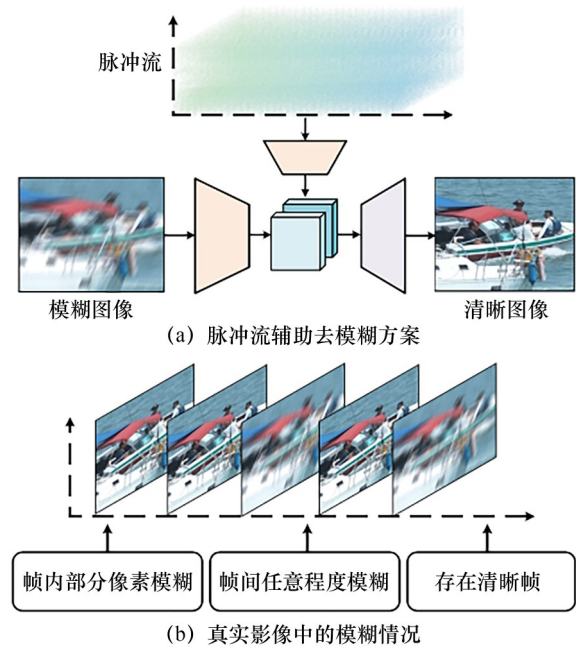


图 11 脉冲辅助去模糊技术路线图

或在诸如存在强光反射、进出隧道等情况均会造成传统相机过曝或欠曝。脉冲相机采用积分型原理相比动态视觉传感器更加适于直接地记录场景的光强信息。为了解决此问题, Han 等^[92]通过累积一段脉冲序列作为模拟的光强图像并融合低动态范围 (LDR) 图像恢复高动态。其方案将 LDR 图像转换至 YUV 域并将动态有源像素视觉传感器 (DAVIS) 图像通过可学习权重融合并最后通过色度补偿融合为高动态范围 (HDR) 图像。也可输入基于多次曝光的红绿蓝 (RGB) 图像及脉冲流, 并利用脉冲中的光流信息指导长时间曝光的图像所丢失的时间信息, 进而实现 HDR^[61]。利用多次曝光的 LDR 图像结合脉冲流的高频和动态范围信息, 二者通过可变性卷积、注意力机制等深度学习模块可以实现互补而可达到更优的 HDR 重建效果^[93,94]。在实际应用中, 难以实现多次曝光, 利用视频序列作为输入以替代多次曝光 HDR 作为输入的方案^[95]对落地自动驾驶更有价值。对于易出现的暗光场景, Liang 等^[96]通过构建跨模态在时空上的相关性矩阵, 建立了不同模态和分辨率之间的时空一致性。

四、脉冲视觉场景感知算法

(一) 脉冲视觉高速运动估计算法

物体与场景的运动是在三维空间发生的, 而光流估计是在无三维几何和运动先验的情况下计算成像平面上像素位移或物体速度的问题。光流反映了物体最基本的运动方向和幅度, 在研究目标检测、跟踪、三维场景深度估计及场景重建等其他感知任务中可发挥至关重要的作用, 有效准确估计物体及场景的运动状态是自动驾驶领域众多视觉算法的关键研究基础^[11,97]。对于传统相机, 这个问题可被描述为寻找不同时刻两个连续图像之间的像素对应关系。然而, 由于传统图像传感器的动态范围有限, 且当存在运动模糊时基于图像的方法将遇到巨大挑战。

脉冲相机具有高动态范围与高时间分辨率的优势, 为应对高速运动等极端挑战性场景下的光流估计问题带来新的解决方案。脉冲流由于其不规则的数据形式, 无法直接提供绝对空间亮度信息以及空间上的连续信息, 给光流估计算法带来新的挑战。目前相关研究一般需要借助一段脉冲流的累积信息

以求解光流。相比于传统相机, 我们可以认为, 脉冲相机产生的脉冲流是时间上连续的, 因此理论上脉冲流中包含了整个三维空间上物体的连续运动信息。然而, 在实践中求解速度无法匹配流式数据产生的速度且对连续运动信息的求解将给硬件带来巨大的计算量^[98], 因此现有研究普遍会在特定时刻计算光流, 如在脉冲发放位置或在人工选择的时刻, 其中后者占据多数。

脉冲相机通过异步发放脉冲来编码每个像素的绝对光强, 携带丰富的纹理信息。在脉冲流中, 一个基本的求解运动的思路是借助运动补偿的思想^[29,99-101]。如图 12 所示, 在将脉冲沿着运动轨迹即光流对齐后, 经过简单 TFP 的重构后即可获得消除运动的清晰影像, 此图像的对比度或清晰度在光流估计准确时应为最大的。对比度最大化框架可进一步扩展, 设计目标函数将脉冲对齐至多个参考时刻以防止过度拟合, 将轨迹模型优化以更好地处理遮挡, 以及通过可通过空间进行多尺度变换来提高收敛性^[102]。

为保证理想性能, 采用 DL 方案的脉冲光流估计将更加有效。图 13 为采用 DL 方案的脉冲光流的四种技术路线。

基于脉冲流对齐的方案。第一项使用脉冲相机探索高速场景中的光流估计研究提出了一种新颖且面向脉冲流的输入表征模块的定制神经网络架构 SCFlow^[103], 允许自适应时间窗口选择, 以用于处理脉冲流中的固定方向卷积导致运动模糊无法去除的问题。时间窗口的选择应该是根据运动信息动态变化的和有方向的。考虑有关动态视觉传感器的基于对比度最大化的研究思路, 在脉冲流上, 如果时间窗口的方向与像素的运动轨迹一致, 则时间窗口中的平均亮度会更接近像素在该时刻的亮度。为此, 引入了脉冲流的一种新颖表征方式——光流引导的

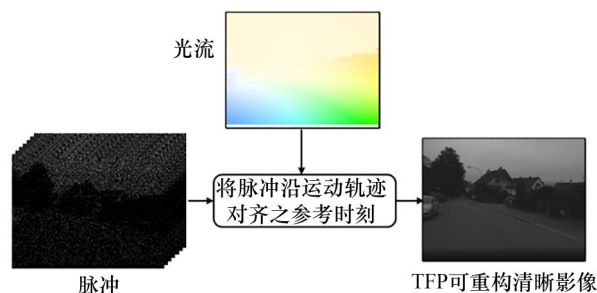


图 12 基于运动补偿的光流估计框架

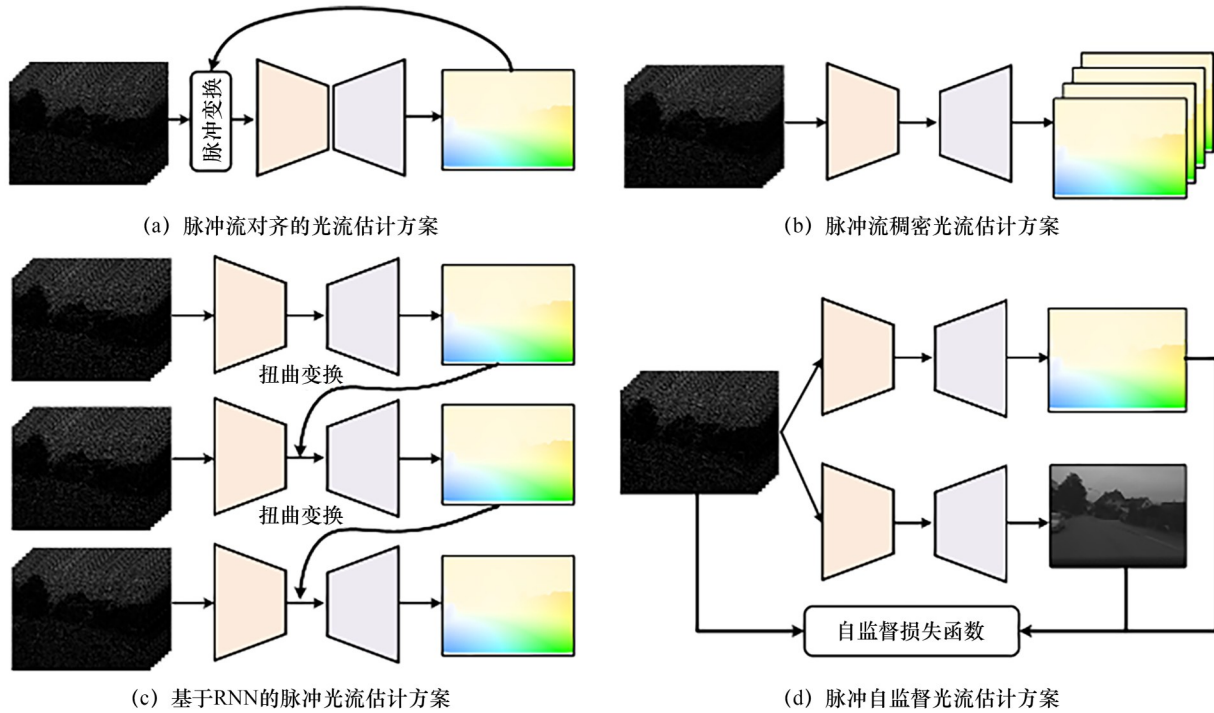


图 13 基于深度学习的脉冲光流估计技术路线

自适应窗口 (FAW)。光流信息被迭代地用作窗口方向选择的引导信息。该工作的网络骨架采用了传统相机光流估计的经典网络PWC-Net。实验结果证明，该方法不仅可以在真实脉冲数据上获得不错的结果，并领先于其他经典光流估计网络架构。该研究对未来脉冲相机光流估计的研究提供了基准，同时提出了可用于大规模训练的脉冲光流数据集以供脉冲视觉社区使用。

脉冲流的稠密光流估计方案。在目前的光流数据集上，序列数据中光流的真值限制在相同时间间隔上，且通常真值较为稀疏。前序光流估计算法以此建立训练过程，无法充分利用脉冲数据的时域优势。为解决此问题，可以将数据流分割为更加精细的片段并对齐时域上精细的运动特征^[104]，或采用序列学习的思路^[105]。Gehrig重新定义了光流估计任务^[106]，旨在计算每个像素在连续时间上的密集光流。Spike2Flow^[107]将脉冲光流估计在时域上细化并使用循环解码聚合不同时刻上的空间特征，最后估计出一段脉冲流中的稠密光流。

基于RNN的脉冲光流估计方案。Gehrig等^[108]首次提出专为密集光流估计设计的CNN模型E-RAFT，利用匹配代价并引入循环结构以结合在时序上的先验信息。

脉冲自监督光流估计方案。基于模糊一致性和光度一致性假设，将光流估计和影像重建的联合优化方案^[74]，克服了脉冲数据缺乏光流真值的难题。

传统的计算机视觉方法以及ANN架构不适合与脉冲数据的不规则特性匹配。SNN是处理脉冲的理想架构，采用SNN进行模型的设计是可行的^[109,110]，同时可将SNN模型在神经形态硬件上实现，满足高速推理的实时性需求^[111]。同时，为了克服SNN在网络规模较大时的性能下降，可采用ANN与SNN混合建模的方案^[112,113]。

采用多模态的方案解决光流估计是目前新兴的方向。结合图像，可进行密集光流估计并可有效解决在只依赖脉冲数据难以估计可靠的密集光流的问题^[114]。最近也有研究探索了神经形态视觉传感器同时与图像、点云融合的新方案，将模态优势互补，提高光流估计和场景流估计的性能^[115]。

(二) 脉冲视觉目标检测与跟踪算法

目标检测、跟踪及分割是自动驾驶中的关键课题，这些高层视觉任务依赖于算法对场景高层语义的分析和提取能力，任务难度更大。传统相机是自动驾驶中的关键感知元件，基于纯视觉的自动驾驶方案在逐渐流行。在基于视觉的智能系统中准确、

及时地检测行人和车辆、正确地分析场景语义，对于无人驾驶发展和预防交通事故极为重要。由于脉冲相机具备高时间分辨率，其优势可直接体现在高速目标或高速场景的检测、跟踪及分割任务上。对于目标检测、跟踪、分割等高层视觉任务，为保证在自动驾驶场景中的时效性及准确性，采用深度学习的方案更合理。

1. 目标检测与跟踪

在自动驾驶场景中，使用脉冲相机进行目标检测与跟踪的优势是可以利用数据的时间连续性进行连续跟踪且不容易丢失目标。脉冲视觉目标检测跟踪人物的关键点是如何保证性能与计算效率。结合脉冲相机发展与研究现状，我们给出如图 14 所示的几种技术方案。

(1) 采用脉冲流直接训练神经网络的方案。通过建立新的脉冲相机目标检测跟踪数据集，可以通过 DL 方案对脉冲流直接训练。一种思路是每次输入一小段脉冲，并使用现有网络或设计新的网络对脉冲训练^[116]。通过循环建模的方式可提高模型对时域信息的建模能力，采用循环结构将脉冲流紧凑地表示到模型内部中的能力对于实现高精度检测至关重要，同时可根据序列结构限制时间的一致性^[117,118]。

(2) 采用脉冲域变换的训练方案。由于脉冲的不规则性，现有计算机视觉生态中的网络架构难以直接训练达到最优性能，可以将脉冲域经过简单高效的设计变换至图像域并微调下游的图像域目标检测跟踪网络^[119]。

(3) 图像至脉冲的蒸馏训练方案。在不建造新数据集的情况下，可以利用现有图像数据集生成模拟脉冲，并利用现有基于图像的网络中间特征蒸馏至基于脉冲的网络。也可采用网络嫁接思路^[120]，使用专门处理脉冲流的前端网络取代处理图像帧的预训练模型的前端网络，使模型借助图像信息完成自监督训练，并且在推理阶段只需要单模态的脉冲流。蒸馏方案的优势是可保证脉冲在检测任务上的性能。

(4) SNN 方案。直接使用代理梯度可以训练深度 SNN 实现目标检测跟踪^[121]，或使用 SNN 的生物启发学习规则训练。SNN 中的生物启发学习规则可以和脉冲流数据天然结合，实现无监督训练框架^[6,69]，利用由脉冲相机捕捉的时空信息来捕获固有的运动模式。该模型可以使用短时可塑性机制来过滤冗余信号，并使用 STDP 学习规则来提取运动模式，实现首个基于脉冲相机的检测跟踪模型。该团队同时在硬件上实现了模型的部署，实现了实时高速检测跟踪^[122]。

(5) 多模态融合方案，即将脉冲流与图像帧或其他模态结合，完成目标检测跟踪任务^[123-125]。Li 等^[123]提出了一种结合模态融合的用于车辆检测的联合框架，启发了对于脉冲流的目标检测，可通过 SNN 接收脉冲流作为输入生成视觉注意力图后与图像流合并输送至 CNN 中进行融合处理。脉冲相机和动态视觉相机 (DVS) 两种脉冲流也可以有机结合，实现优势互补^[124]，是脉冲视觉目标检测的新范式。

目前基于脉冲相机的目标跟踪研究还处于发展

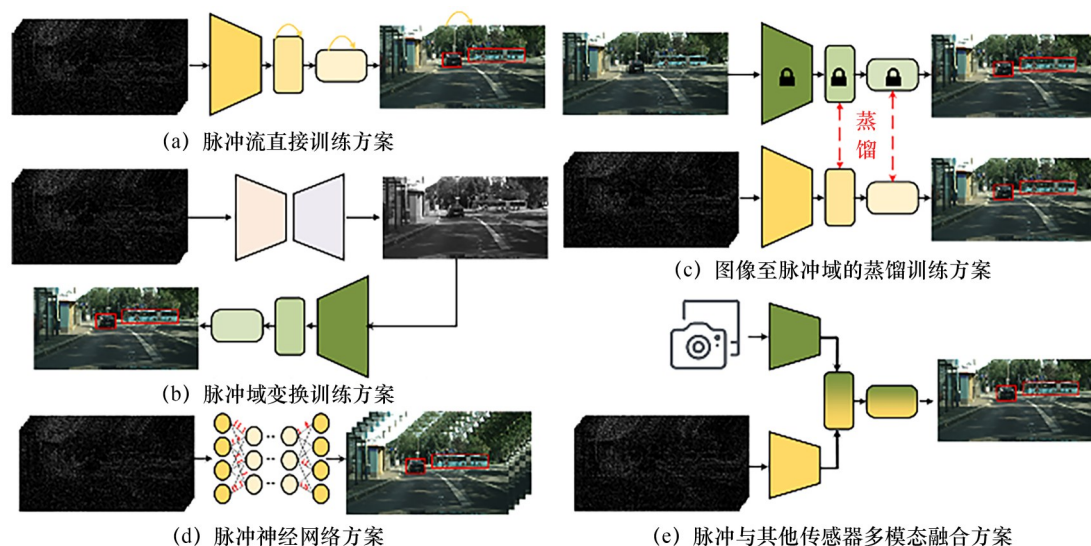


图 14 脉冲相机目标检测与跟踪技术路线

初期。未来基于脉冲相机的目标检测跟踪可以继续挖掘SNN对脉冲建模的潜力，并尝试在神经形态硬件上实现，真正实现对高效、高性能的脉冲目标进行检测、跟踪。

2. 目标分割

目标分割的目的是通过分析并提取视觉高层特征，将像素在空间上分成若干区域，并辅助定位目标位置或对场景进行分析，是自动驾驶中导航和场景理解的关键过程。目前基于脉冲相机的目标分割研究工作处于发展初期。

Zhang等^[67]首次探索了在脉冲流上完成语义分割任务。发掘充分表达时序和空间信息的脉冲表征是一个普遍需求，一个通用且有效的脉冲流表征可使脉冲视觉算法性能得到普遍提升。脉冲流包含丰富的时频信息，而小波正是有效的频谱分析工具。通过设计在时域上进行可学习的脉冲小波变换过程，脉冲被变换成为更加鲁棒的视觉表示。在不重新设计下游语义分割网络的情况下，其方法在多个现有的基于图像的模型上获得性能提升。

（三）脉冲视觉三维场景深度估计

深度估计是三维场景感知、理解的最关键任务之一，基于脉冲相机完成深度估计任务有着巨大的潜力。脉冲相机捕获的脉冲流包含了更加紧密的时空相关性、连续的运动信息以及更加丰富的边缘和纹理信息。因此，对基于异步的、高时间分辨率、高动态的脉冲数据流进行深度估计是一项完全不同于传统相机的视觉任务。本节只考虑使用纯视觉信号而不借助其他信号辅助（如结构光、激光雷达等）进行的研究工作。深度估计任务可分为单目深度估计和多目深度估计。对于单目深度估计，常使用多视角方案或DL算法完成；对于多目深度估计，可通过匹配算法计算视差或直接采用DL算法。

脉冲相机由于其积分型原理，比DVS系列相机更适合估计稠密深度图。Zhang等首次提出了基于Transformer编码器-解码器架构的脉冲相机单目深度估计算法^[126]，将脉冲流进行时空块划分并融合时域上的浅层局部特征和全局特征，并将编码器使用有时空注意力机制的三维Transformer块结构以从不同时空区域融合特征，辅助理解场景的结构信息。Wang等^[127]提出了第一个端到端学习脉冲相机立体深度估计网络，命名为SSDEFormer，其利用双目之间

的交叉注意力，建立了双目直接脉冲流特征的关系。该项研究构建了一个混合摄像机平台，提供了一个新的立体深度估计数据集（即PKU-Spike-Stereo）用于脉冲相机的双目深度估计研究。

五、面向自动驾驶的脉冲感知系统

在自动驾驶等应用中，完整的感知与计算过程，即先使用脉冲相机并结合传统相机等其他视觉传感器作为信息采集装置获取场景信息；通过软硬件接口对视觉数据进行预处理，使其转换为计算设备可处理的形式；最后将数据输入视觉处理算法，进行对场景的感知计算并进一步发送到驾驶控制系统等。

在数据方面，目前已有成熟软件接口实现将Windows/Linux等多种处理系统直接连接脉冲相机以实现数据采集。对于面向脉冲相机的视觉算法研究，影像重建相关任务大多需要同步且标定的脉冲和图像数据供模型训练使用，目前大多数的研究采用分光镜装置使脉冲相机和传统相机同时采集场景数据^[61]。对于光流估计、目标检测跟踪、深度估计等需要复杂标签的视觉任务，研究大多采用通过现有带标签的视频数据集生成高帧率视频后模拟生成脉冲供模型训练和验证。表2汇总了目前脉冲相机相关的所有开源数据集，其中包含了多种影像重建、目标识别检测跟踪、深度估计、光流估计、语义分割等视觉任务所需的数据。在所列数据集中，半数以上为驾驶场景数据，可为室外驾驶环境感知算法的设计提供便利。

计算方面的难点在于实际应用中，即如何将脉冲相机数据采集、数据处理及感知计算的全过程统一。目前针对脉冲相机的开源平台SpikeCV^[128]已建立，主要针对脉冲相机数据、算法、应用的一体化需求开发。SpikeCV框架对脉冲相机数据集进行了封装设计，并针对相机设计了设备接口并整合了多种面向脉冲相机的视觉处理算法，便于开发者轻松构建脉冲视觉应用。框架对离线数据集以及在线采集过程进行了标准化设计，使用户可以简易地使用脉冲相机部署于视觉应用开发系统中。在接口层面，提供了对硬件友好的C++接口和用户友好的Python接口^[128]供用户调用。为了实现感知计算的统一，该平台支持各型号脉冲相机硬件采集、数据处理与下游视觉任务模型计算的同步且实时的流水线。

表 2 脉冲相机针对各项视觉任务已建立的数据库汇总

数据库	视觉任务	脉冲类型	标签类型	数据来源
PKU-Spike-Recon	影像重建	真实	无	[63]
PKU-Spike-High-Speed	影像重建、光流估计	真实	无	[63]
REDS (Spike)	影像重建	模拟	图像	[65]
Spk-X4K1000FPS	去模糊重建	模拟	高帧率视频	[62]
Spk-GoPro	去模糊重建	模拟	高帧率视频	[62]
Spike-RGB (HDR)	高动态重建	真实	图像	[61]
Vimeo90k (Spike)	视频插帧	模拟	高帧率视频	[80]
Spike-Vimeo	超分重建	模拟	超分辨率图像	[80]
Spike-SR	超分重建	真实	超分辨率图像	[128]
S-MNIST	目标识别	真实	类别	[102]
S-CIFAR	目标识别	真实	类别	[102]
S-CALTECH	目标识别	真实	类别	[106]
HSSR (S-ImgNet, S-Cars)	目标识别	真实	类别	[123]
SPIFT	光流估计	模拟	类别	[121]
PHM	光流估计	模拟	光流	[126]
RSSF	光流估计	模拟	光流	[122]
PKU-Vidar-DVS	目标检测、目标跟踪	真实	包围框	[122]
motVidarReal2020	目标检测、目标跟踪	真实	无	[67]
PKU-Spike-Stereo	双目深度估计	模拟	深度图	[63]
DENSE-Spike	单目深度估计	模拟	深度图、图像	[63]
Outdoor-Spike	单目深度估计	真实	无	[65]
Spike-Cityscapes	语义分割	模拟	语义图、图像	[62]

六、面向自动驾驶的脉冲视觉研究挑战

目前,基于脉冲的视觉感知任务研究已经具备一定规模且处于不断上升期。本节分析自动驾驶场景中脉冲视觉感知的研究挑战并提供未来研究方向,为未来脉冲相机与自动驾驶的高度融合赋能。

第一,脉冲流相较于图像格式是不规则的二值形式,且往往因硬件等原因数据易掺杂噪声,从而引起算法的失效。如何有效消除噪声或设计算法对噪声鲁棒是一项难题。在硬件层面,通过额外的电路设计统计在无光照条件下传感器的暗电流噪声分布,并将此部分噪声从脉冲流中减去。在算法层面,可以针对不同噪声类型手工设计去噪算法或通过学习的方式建模使模型学习噪声分布或自适应地抗噪特性。大多数面向脉冲相机的深度学习模型均在模拟脉冲数据集上训练,而模拟数据欠缺考虑真实脉冲噪声,可对数据集加入噪声进行数据增强以实现模型对噪声的适应性。

第二,脉冲视觉算法在真实自动驾驶场景中的泛化性。目前面向脉冲相机的深度学习模型多使用模拟数据集进行训练,模拟数据集除欠缺考虑真实噪声分布外,其脉冲流与真实脉冲相机所拍摄脉冲流依然存在一定差别。① 模拟数据中的生成脉冲过程较为简单,难以建模真实世界光子到达的过程,且其模拟的光强与真实世界光强有一定区别;② 脉冲相机的硬件电路较为复杂,模拟过程无法完全复现电路层面传感的过程。此外,供模型训练的数据集场景可能不是道路、街道等驾驶场景。上述两种原因将导致目前脉冲视觉算法在真实自动驾驶环境中的泛化有一定局限性。针对该挑战性问题,可对大型数据集的脉冲模拟过程做更精细化的建模,如更加细粒度地模拟不同时刻的光照变化,将脉冲流的产生增加随机性。同时,也可通过扩大模型、扩展数据集等方式使模型的泛化能力提高;也可对模型采用大数据预训练、小数据微调的策略以针对不同场景、不同任务增强性能。

第三，脉冲相机往往可在高速、过曝等极端场景中发挥最大优势，目前单独使用脉冲相机作为传感器做到在所有视觉任务上超越传统相机的性能具有挑战性。如何突破脉冲相机目前的感知算法性能上限是一个大课题。从脉冲视觉算法模型设计考虑，未来的模型设计可聚焦于时序建模，将算法侧重于利用脉冲时域，连续性挖掘其对动态场景感知的优势和潜力。例如，目前采用RNN/SNN的建模方案或许可以突破相机对单一时刻的感知能力，增强对场景的连续感知能力。

第四，传统相机尽管在动态范围及成像速度上不及脉冲相机，但其制造工艺发展时间长，成像质量好，且相关视觉感知算法及系统性能优秀、成熟，如何更好地融合或借助传统相机、图像域算法是一个重要机遇。可行方案有两种：一是蒸馏算法，二是域适应算法。脉冲流虽只包含0、1数据，其与图像之间的模态距离相较于点云、语音等数据更近，可采用蒸馏算法设计，将应用于图像或视频的大型模型蒸馏至脉冲相机视觉模型中，从而使模型快速学习已有知识。也可使用域适应思路，通过对抗驯良等方式将脉冲及图像在特征域上的距离拉近，从而对齐两种数据域。

第五，脉冲相机的低延迟和视觉感知算法速度的匹配十分困难。脉冲相机以40 000 Hz的频率输出脉冲流，一些算法需充分利用脉冲流的时间连续性展开计算，但大部分深度学习模型的推理速度最高只能达到每秒数百次推理的速度，使得数据和算法之间的速度无法匹配，进而无法真正用于超高速场景。在算法层面，通过模型剪枝、量化等方法可以实现推理加速；在硬件层面，通过并行或特殊电

路设计可实现加速。

第六，自动驾驶系统中的脉冲相机与多种车载传感器协同感知是目前系统层面主要的技术难题之一。其关键点在于如何使超高时间分辨率脉冲流与其他低时间分辨率模态数据同步，且采用何种算法融合多模态数据并有效提高感知能力。针对数据同步问题，一个潜在的解决方案是，将车载系统的精确时钟信号与脉冲相机在硬件电路层面融合，使脉冲流可获得额外精确绝对时间戳信息，以达到与其他车载系统中的传感器同步。车辆主要传感器包括传统相机、激光雷达等，设计新的多模态算法融合脉冲流信息增强感知能力是目前未探索的方向。

七、结语

针对上述几项挑战性问题及脉冲视觉感知算法的研究进展，我们提出五项具有建设性的未来研究方向如图15所示，均对自动驾驶场景有重要价值。研究认为探索脉冲相机的预训练模型、脉冲视觉适配器、多模态模型、SNN、脉冲三维视觉这五个方向的研究具有重要价值。第一，视觉预训练模型是统一地解决多种下游视觉任务的重要路线。通过生成大型脉冲数据集（类似于N-ImageNet等^[129]）、构造脉冲友好的自监督预训练模型，可以有效解决现有小型单一模型的精确率和泛化性问题。同时大型的预训练模型可加强模型对脉冲流的特征提取能力，有效消除数据中噪声带来的影响。第二，研究轻量化的、有效的脉冲视觉适配器。脉冲流和RGB图像有着一定的数据域差别，建立脉冲视觉适配器的目标是以一个小型的网络或模块与图像域视觉模

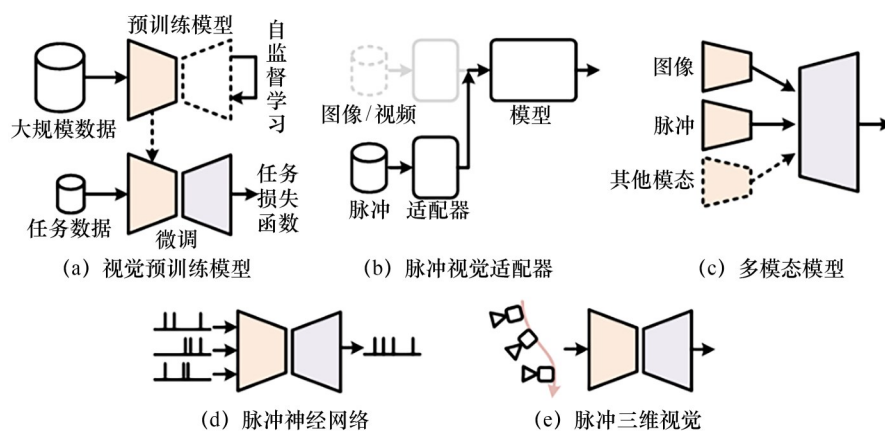


图15 脉冲视觉研究新方向示意图

型连接, 通过设计学习算法使得脉冲可以快速与图像在特征域上对齐进而达到最优性能。第三, 结合脉冲相机建立多模态视觉模型。目前部分研究已经证明了使用脉冲流作为辅助并结合图像、点云等模态数据可以有效提升性能以及在极端场景的感知能力。图像和脉冲之间存在互补关系, 分别占据了时间和空间上的不同优势。另外, 在自动驾驶中, 与激光雷达、深度相机、多目环视相机等传感器设置有机融合均有一定潜力。第四, 探索深层的、复杂的 SNN 模型。SNN 已被证实有着低能耗的优势, 且天然适配脉冲数据。结合神经形态硬件的部署实现, 可以显著加速模型的推理, 从而缓解数据和算法之间速度无法匹配的难题。第五, 发展脉冲视觉三维模型。在自动驾驶场景中, 车辆或场景总是在运动的, 在运动过程中感知三维世界是困难的。脉冲相机的高时间分辨率可以为三维世界提供连续时间线索, 即提供连续的视角变换, 为三维视觉感知能力突破赋能。

利益冲突声明

本文作者在此声明彼此之间不存在任何利益冲突或财务冲突。

Received date: November 14, 2023; **Revised date:** January 5, 2024

Corresponding author: Huang Tiejun is a professor from National Engineering Research Center of Visual Technology. His major research fields include visual information processing and brain like intelligence; E-mail: tjhuang@pku.edu.cn

Funding project: Chinese Academy of Engineering project “Strategic Research on New Generation of Artificial Intelligence and Industrial Clusters” (2022-PP-07)

参考文献

- [1] Lichtsteiner P, Posch C, Delbruck T. A $128 \times 128 \times 120$ dB $15 \mu\text{s}$ latency asynchronous temporal contrast vision sensor [J]. *IEEE Journal of Solid-State Circuits*, 2008, 43(2): 566–576.
- [2] Posch C, Matolin D, Wohlgenannt R. An asynchronous time-based image sensor [C]. Seattle: 2008 IEEE International Symposium on Circuits and Systems, 2008.
- [3] Brandli C, Berner R, Yang M H, et al. A $240 \times 180 \times 130$ dB $3 \mu\text{s}$ latency global shutter spatiotemporal vision sensor [J]. *IEEE Journal of Solid-State Circuits*, 2014, 49(10): 2333–2341.
- [4] Moeys D P, Corradi F, Li C H, et al. A sensitive dynamic and active pixel vision sensor for color or neural imaging applications [J]. *IEEE Transactions on Biomedical Circuits and Systems*, 2018, 12(1): 123–136.
- [5] Huang J, Guo M H, Chen S S. A dynamic vision sensor with direct logarithmic output and full-frame picture-on-demand [C]. Baltimore: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), 2017.
- [6] Huang T J, Zheng Y J, Yu Z F, et al. $1000 \times$ faster camera and machine vision with ordinary devices [J]. *Engineering*, 2023, 25: 110–119.
- [7] 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 [J]. *IEEE Journal of Solid-State Circuits*, 2011, 46(1): 259–275.
- [8] Chen D G, Matolin D, Bermak A, et al. Pulse-modulation imaging—Review and performance analysis [J]. *IEEE Transactions on Biomedical Circuits and Systems*, 2011, 5(1): 64–68.
- [9] Son B, Suh Y, Kim S, et al. 4.1 A 640×480 dynamic vision sensor with a $9 \mu\text{m}$ pixel and 300 Meps address-event representation [C]. San Diego: 2017 IEEE International Solid-State Circuits Conference (ISSCC), 2017.
- [10] Culurciello E, Etienne-Cummings R, Boahen K A. A biomorphic digital image sensor [J]. *IEEE Journal of Solid-State Circuits*, 2003, 38(2): 281–294.
- [11] 李家宁, 田永鸿. 神经形态视觉传感器的研究进展及应用综述 [J]. *计算机学报*, 2021, 44(6): 1258–1286.
Li J N, Tian Y H. Recent advances in neuromorphic vision sensors: A survey [J]. *Chinese Journal of Computers*, 2021, 44(6): 1258–1286.
- [12] Gallego G, Delbrück T, Orchard G, et al. Event-based vision: A survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 154–180.
- [13] Bardow P, Davison A J, Leutenegger S. Simultaneous optical flow and intensity estimation from an event camera [C]. Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [14] Li H M, Li G Q, Shi L P. Super-resolution of spatiotemporal event-stream image [J]. *Neurocomputing*, 2019, 335: 206–214.
- [15] Munda G, Reinbacher C, Pock T. Real-time intensity-image reconstruction for event cameras using manifold regularisation [J]. *International Journal of Computer Vision*, 2018, 126(12): 1381–1393.
- [16] Scheerlinck C, Barnes N, Mahony R. Continuous-time intensity estimation using event cameras [C]. Los Angeles: Asian Conference on Computer Vision, 2019.
- [17] Scheerlinck C, Barnes N, Mahony R. Asynchronous spatial image convolutions for event cameras [J]. *IEEE Robotics and Automation Letters*, 2019, 4(2): 816–822.
- [18] Barua S, Miyatani Y, Veeraraghavan A. Direct face detection and video reconstruction from event cameras [C]. Lake Placid: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.
- [19] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311–4322.
- [20] Rebecq H, Ranftl R, Koltun V, et al. Events-to-video: Bringing modern computer vision to event cameras [C]. Long Beach: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [21] Scheerlinck C, Rebecq H, Gehrig D, et al. Fast image reconstruction with an event camera [C]. Snowmass Village: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 2020.

- [22] Cadena P R G, Qian Y Q, Wang C X, et al. SPADE-E2VID: Spatially-adaptive denormalization for event-based video reconstruction [J]. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 2021, 30: 2488–2500.
- [23] Weng W M, Zhang Y Y, Xiong Z W. Event-based video reconstruction using transformer [C]. Beijing: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [24] Zhu L, Wang X, Chang Y, et al. Event-based video reconstruction via potential-assisted spiking neural network [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [25] Pini S, Borghi G, Vezzani R. Learn to see by events: Color frame synthesis from event and RGB cameras [C]. Beijing: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2020.
- [26] Wang L, Mohammad Mostafavi I S, Ho Y S, et al. Event-based high dynamic range image and very high frame rate video generation using conditional generative adversarial networks [C]. Long Beach: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [27] Wang L, Kim T K, Yoon K J. EventSR: From asynchronous events to image reconstruction, restoration, and super-resolution via end-to-end adversarial learning [C]. Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [28] Paredes-Vallés F, de Croon G C H E. Back to event basics: Self-supervised learning of image reconstruction for event cameras via photometric constancy [C]. Online: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [29] Liu S C, Rueckauer B, Ceolini E, et al. Event-driven sensing for efficient perception: Vision and audition algorithms [J]. *IEEE Signal Processing Magazine*, 2019, 36(6): 29–37.
- [30] Barranco F, Fermüller C, Ros E. Real-time clustering and multi-target tracking using event-based sensors [C]. Madrid: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018.
- [31] Piątkowska E, Belbachir A N, Schraml S, et al. Spatiotemporal multiple persons tracking using Dynamic Vision Sensor [C]. Providence: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012.
- [32] Lagorce X, Meyer C, Ieng S H, et al. Asynchronous event-based multikernel algorithm for high-speed visual features tracking [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(8): 1710–1720.
- [33] Rodríguez-Gomez J P, Eguíluz A G, Martínez-de Dios J R, et al. Asynchronous event-based clustering and tracking for intrusion monitoring in UAS [C]. Paris: 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020.
- [34] Mitrokhin A, Fermüller C, Parameshwara C, et al. Event-based moving object detection and tracking [C]. Madrid: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018.
- [35] Chen H S, Wu Q Q, Liang Y J, et al. Asynchronous tracking-by-detection on adaptive time surfaces for event-based object tracking [C]. New York: The 27th ACM International Conference on Multimedia, 2019.
- [36] Chen H S, Suter D, Wu Q Q, et al. End-to-end learning of object motion estimation from retinal events for event-based object tracking [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 10534–10541.
- [37] Rebecq H, Gallego G, Mueggler E, et al. EMVS: Event-based multi-view stereo—3D reconstruction with an event camera in real-time [J]. *International Journal of Computer Vision*, 2018, 126(12): 1394–1414.
- [38] Rebecq H, Gallego G, Scaramuzza D. EMVS: Event-based multi-view stereo [C]. York: The British Machine Vision Conference, 2016.
- [39] Hidalgo-Carrió J, Gehrig D, Scaramuzza D. Learning monocular dense depth from events [C]. Fukuoka: 2020 International Conference on 3D Vision (3DV), 2020.
- [40] Gehrig D, Rügge M, Gehrig M, et al. Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction [J]. *IEEE Robotics and Automation Letters*, 2021, 6(2): 2822–2829.
- [41] Schraml S, Belbachir A N, Milosevic N, et al. Dynamic stereo vision system for real-time tracking [C]. Paris: 2010 IEEE International Symposium on Circuits and Systems, 2010.
- [42] Kogler J, Sulzbachner C, Humenberger M, et al. Address-event based stereo vision with bio-inspired silicon retina imagers [C]. Rijeka: *Advances in Theory and Applications of Stereo Vision*, 2011.
- [43] Kogler J, Humenberger M, Sulzbachner C. Event-based stereo matching approaches for frameless address event stereo data [C]. Las Vegas: The 7th International Conference on Advances in Visual Computing, 2011.
- [44] Rogister P, Benosman R, Ieng S H, et al. Asynchronous event-based binocular stereo matching [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(2): 347–353.
- [45] Lee J, Delbruck T, Park P K J, et al. Live demonstration: Gesture-based remote control using stereo pair of dynamic vision sensors [C]. Seoul: 2012 IEEE International Symposium on Circuits and Systems (ISCAS), 2012.
- [46] Camuñas-Mesa L A, Serrano-Gotarredona T, Ieng S H, et al. On the use of orientation filters for 3D reconstruction in event-driven stereo vision [J]. *Frontiers in Neuroscience*, 2014, 8: 48.
- [47] Tulyakov S, Fleuret F, Kiefel M, et al. Learning an event sequence embedding for dense event-based deep stereo [C]. Seoul: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [48] Ahmed S H, Jang H W, Uddin S N, et al. Deep event stereo leveraged by event-to-image translation [C]. Vancouver: The AAAI Conference on Artificial Intelligence, 2021.
- [49] Nam Y, Mostafavi M, Yoon K J, et al. Stereo depth from events cameras: Concentrate and focus on the future [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [50] Zhang K X, Che K W, Zhang J G, et al. Discrete time convolution for fast event-based stereo [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [51] Cho H, Cho J, Yoon K J. Learning adaptive dense event stereo from the image domain [C]. Vancouver: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

- [52] Rañon U, Cuadrado-Anibarro J, Cottreau B R, et al. StereoSpike: Depth learning with a spiking neural network [J]. *IEEE Access*, 2022, 10: 127428–127439.
- [53] Hadviger A, Marković I, Petrović I. Stereo dense depth tracking based on optical flow using frames and events [J]. *Advanced Robotics*, 2021, 35(3–4): 141–152.
- [54] Wang Z W, Pan L Y, Ng Y, et al. Stereo hybrid event-frame (SHEF) cameras for 3D perception [C]. Prague: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021.
- [55] Zuo Y F, Cui L, Peng X, et al. Accurate depth estimation from a hybrid event-RGB stereo setup [C]. Prague: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021.
- [56] Mostafavi I S M, Yoon K J, Choi J. Event-intensity stereo: Estimating depth by the best of both worlds [C]. Montreal: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [57] 黄铁军. 脉冲连续摄影原理与超高速高动态成像验证 [J]. *电子学报*, 2022, 50(12): 2919–2927.
Huang T J. Spiking continuous photographing principle and demonstration on ultrahigh speed and high dynamic imaging [J]. *Acta Electronica Sinica*, 2022, 50(12): 2919–2927.
- [58] 黄铁军, 余肇飞, 李源, 等. 脉冲视觉研究进展 [J]. *中国图象图形学报*, 2022, 27(6): 1823–1839.
Huang T J, Yu Z F, Li Y, et al. Advances in spike vision [J]. *Journal of Image and Graphics*, 2022, 27(6): 1823–1839.
- [59] Masland R H. The neuronal organization of the retina [J]. *Neuron*, 2012, 76(2): 266–280.
- [60] Wässle H. Parallel processing in the mammalian retina [J]. *Nature Reviews Neuroscience*, 2004, 5: 747–757.
- [61] Chang Y K, Zhou C, Hong Y C, et al. 1000 FPS HDR video with a spike-RGB hybrid camera [C]. Vancouver: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- [62] Chen S Y, Zhang J Y, Zheng Y J, et al. Enhancing motion deblurring in high-speed scenes with spike streams [C]. New Orleans: Thirty-Seventh Conference on Neural Information Processing Systems, 2023.
- [63] Zhu L, Dong S W, Huang T J, et al. A retina-inspired sampling method for visual texture reconstruction [C]. Shanghai: 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019.
- [64] Zhao J, Xiong R Q, Huang T J. High-speed motion scene reconstruction for spike camera via motion aligned filtering [C]. Online: 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020.
- [65] Zhao J, Xiong R Q, Liu H F, et al. Spk2ImgNet: learning to reconstruct dynamic scene from continuous spike stream [C]. Nashville: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [66] Zhu L, Zheng Y L, Geng M Y, et al. Recurrent spike-based image restoration under general illumination [EB/OL]. (2023-08-06)[2023-11-01]. <http://arxiv.org/abs/2308.03018.pdf>.
- [67] Zhang J, Jia S, Yu Z, et al. Learning temporal-ordered representation for spike streams based on discrete wavelet transforms [C]. Washington, DC: The AAAI Conference on Artificial Intelligence, 2023.
- [68] Zhu L, Dong S W, Li J N, et al. Retina-like visual image reconstruction via spiking neural model [C]. Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [69] Zheng Y J, Zheng L X, Yu Z F, et al. High-speed image reconstruction through short-term plasticity for spiking cameras [C]. Nashville: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [70] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]. Munich: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.
- [71] Bi G Q, Poo M M. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type [J]. *The Journal of Neuroscience*, 1998, 18(24): 10464–10472.
- [72] Tsodyks M, Pawelzik K, Markram H. Neural networks with dynamic synapses [J]. *Neural Computation*, 1998, 10(4): 821–835.
- [73] Tsodyks M V, Markram H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1997, 94(2): 719–723.
- [74] Chen S, Yu Z, Huang T. Self-supervised joint dynamic scene reconstruction and optical flow estimation for spiking camera [C]. Washington, DC: The AAAI Conference on Artificial Intelligence, 2023.
- [75] Dong Y C, Zhao J, Xiong R Q, et al. 3D residual interpolation for spike camera demosaicing [C]. Bordeaux: 2022 IEEE International Conference on Image Processing (ICIP), 2022.
- [76] Zhao J, Xie J Y, Xiong R Q, et al. Super resolve dynamic scene from continuous spike streams [C]. Montreal: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [77] Mostafavi I S M, Choi J, Yoon K J. Learning to super resolve intensity images from events [C]. Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [78] Weng W M, Zhang Y Y, Xiong Z W. Boosting event stream super-resolution with a recurrent neural network [C]. Tel Aviv: Computer Vision-ECCV 2022: 17th European Conference, 2022.
- [79] Zhao J, Xiong R, Zhang J, et al. Learning to super-resolve dynamic scenes for neuromorphic spike camera [C]. Washington, DC: The AAAI Conference on Artificial Intelligence, 2023.
- [80] Xiang X J, Zhu L, Li J N, et al. Learning super-resolution reconstruction for high temporal resolution spike stream [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(1): 16–29.
- [81] Duan P Q, Ma Y, Zhou X Y, et al. NeuroZoom: Denoising and super resolving neuromorphic events and spikes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(12): 15219–15232.
- [82] Duan P Q, Wang Z W, Zhou X Y, et al. EventZoom: Learning to denoise and super resolve neuromorphic events [C]. Nashville: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [83] Zhang S, Zhang Y, Jiang Z, et al. Learning to see in the dark with events [C]. Cham: European Conference on Computer Vision, 2020.
- [84] Rebecq H, Ranftl R, Koltun V, et al. High speed and high dynamic range video with an event camera [J]. *IEEE Transactions on Pat-*

- tern Analysis and Machine Intelligence, 2021, 43(6): 1964–1980.
- [85] Zou Y H, Zheng Y Q, Takatani T, et al. Learning to reconstruct high speed and high dynamic range videos from events [C]. Nashville: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [86] Xu F, Yu L, Wang B S, et al. Motion deblurring with real events [C]. Montreal: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [87] Shang W, Ren D W, Zou D Q, et al. Bringing events into video deblurring with non-consecutively blurry frames [C]. Montreal: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [88] Sun L, Sakaridis C, Liang J Y, et al. Event-based fusion for motion deblurring with cross-modal attention [C]. Tel Aviv: Computer Vision-ECCV 2022: 17th European Conference, 2022.
- [89] Sun L, Sakaridis C, Liang J Y, et al. Event-based frame interpolation with ad-hoc deblurring [C]. Vancouver: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- [90] Zhang X, Yu L, Yang W, et al. Generalizing event-based motion deblurring in real-world scenarios [EB/OL]. (2023-08-11)[2023-11-01]. <http://arxiv.org/abs/2308.05932.pdf>.
- [91] Cho H, Jeong Y, Kim T, et al. Non-coaxial event-guided motion deblurring with spatial alignment [C]. Montreal: The IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
- [92] Han J, Zhou C, Duan P Q, et al. Neuromorphic camera guided high dynamic range imaging [C]. Seattle: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [93] Shaw R, Catley-Chandar S, Leonardis A, et al. HDR reconstruction from bracketed exposures and events [EB/OL]. (2022-03-28)[2023-11-01]. <http://arxiv.org/abs/2203.14825.pdf>.
- [94] Messikommer N, Georgoulis S, Gehrig D, et al. Multi-bracket high dynamic range imaging with event cameras [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022.
- [95] Yang Y X, Han J, Liang J X, et al. Learning event guided high dynamic range video reconstruction [C]. Vancouver: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- [96] Liang J, Yang Y, Li B, et al. Coherent Event Guided Low-Light Video Enhancement [C]. Paris: The IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
- [97] Janai J, Güneş F, Behl A, et al. Computer vision for autonomous vehicles: Problems, datasets and state of the art [J]. Foundations and Trends® in Computer Graphics and Vision, 2020, 12(1–3): 1–308.
- [98] Gallego G, Rebecq H, Scaramuzza D. A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation [C]. Salt Lake City: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- [99] Gallego G, Gehrig M, Scaramuzza D. Focus is all you need: Loss functions for event-based vision [C]. Long Beach: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [100] Gallego G, Gehrig M, Scaramuzza D. Focus is all you need: Loss functions for event-based vision [C]. Long Beach: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [101] Stoffregen T, Kleeman L. Simultaneous Optical Flow and Segmentation (SOFAS) using dynamic vision sensor [EB/OL]. (2018-05-31)[2023-11-01]. <http://arxiv.org/abs/1805.12326.pdf>.
- [102] Shiba S, Aoki Y, Gallego G. Secrets of event-based optical flow [C]. Cham: European Conference on Computer Vision, 2022.
- [103] Hu L W, Zhao R, Ding Z L, et al. Optical flow estimation for spiking camera [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [104] Ponghiran W, Liyanagedera C M, Roy K. Event-based temporally dense optical flow estimation with sequential learning [EB/OL]. (2022-10-03)[2023-11-01]. <http://arxiv.org/abs/2210.01244.pdf>.
- [105] Liu H T, Chen G, Qu S Q, et al. TMA: Temporal motion aggregation for event-based optical flow [EB/OL]. (2023-03-21)[2023-11-01]. <http://arxiv.org/abs/2303.11629.pdf>.
- [106] Gehrig M, Muglikar M, Scaramuzza D. Dense continuous-time optical flow from events and frames [EB/OL]. (2022-03-25)[2023-11-01]. <http://arxiv.org/abs/2203.13674.pdf>.
- [107] Zhao R, Xiong R, Zhao J, et al. Learning optical flow from continuous spike streams [J]. Advances in Neural Information Processing Systems, 2022, 35: 7905–7920.
- [108] Gehrig M, Millhäusler M, Gehrig D, et al. E-RAFT: Dense optical flow from event cameras [C]. London: 2021 International Conference on 3D Vision (3DV), 2021.
- [109] Orchard G, Benosman R, Etienne-Cummings R, et al. A spiking neural network architecture for visual motion estimation [C]. Rotterdam: 2013 IEEE Biomedical Circuits and Systems Conference (BioCAS), 2013.
- [110] Paredes-Valles F, Scheper K Y W, de Croon G C H E. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2051–2064.
- [111] Haessig G, Cassidy A, Alvarez R, et al. Spiking optical flow for event-based sensors using IBM’s TrueNorth neurosynaptic system [J]. IEEE Transactions on Biomedical Circuits and Systems, 2018, 12(4): 860–870.
- [112] Lee C, Kosta A K, Zhu A Z, et al. Spike-FlowNet: Event-based optical flow estimation with energy-efficient hybrid neural networks [C]. Glasgow: Computer Vision—ECCV 2020: 16th European Conference, 2020.
- [113] Lee C, Kosta A K, Roy K. Fusion-FlowNet: Energy-efficient optical flow estimation using sensor fusion and deep fused spiking-analog network architectures [C]. Philadelphia: 2022 International Conference on Robotics and Automation (ICRA), 2022.
- [114] Wan Z X, Dai Y C, Mao Y X. Learning dense and continuous optical flow from an event camera [J]. IEEE Transactions on Image Processing, 2022, 31: 7237–7251.
- [115] Wan Z X, Mao Y X, Zhang J, et al. RPEFlow: Multimodal fusion of RGB-PointCloud-event for joint optical flow and scene flow estimation [EB/OL]. (2023-09-26)[2023-11-01]. <http://arxiv.org/abs/2309.15082.pdf>.

- [116] Chen N F Y. Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion [C]. Salt Lake City: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018.
- [117] Perot E, de Tournemire P, Nitti D, et al. Learning to detect objects with a 1 megapixel event camera [C]. Vancouver: The 34th International Conference on Neural Information Processing Systems, 2020.
- [118] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]. Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [119] Cannici M, Ciccone M, Romanoni A, et al. Asynchronous convolutional networks for object detection in neuromorphic cameras [C]. Long Beach: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019.
- [120] Hu Y H, Delbruck T, Liu S C. Learning to exploit multiple vision modalities by using grafted networks [C]. Glasgow: Computer Vision-ECCV 2020: 16th European Conference, 2020.
- [121] Zhang J Q, Dong B, Zhang H W, et al. Spiking transformers for event-based single object tracking [C]. New Orleans: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- [122] Zhu Y Y, Zhang Y, Xie X D, et al. An FPGA accelerator for high-speed moving objects detection and tracking with a spike camera [J]. *Neural Computation*, 2022, 34(8): 1812–1839.
- [123] Li J N, Dong S W, Yu Z F, et al. Event-based vision enhanced: A joint detection framework in autonomous driving [C]. Shanghai: 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019.
- [124] Li J N, Wang X, Zhu L, et al. Retinomorphic object detection in asynchronous visual streams [C]. Vancouver: The AAAI Conference on Artificial Intelligence, 2022.
- [125] Zhang J, Wang Y, Liu W, et al. Frame-event alignment and fusion network for high frame rate tracking [C]. Vancouver: The IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023.
- [126] Zhang J Y, Tang L L, Yu Z F, et al. Spike transformer: Monocular depth estimation for spiking camera [C]. Tel Aviv: Computer Vision-ECCV 2022: 17th European Conference, 2022.
- [127] Wang Y X, Li J N, Zhu L, et al. Learning stereo depth estimation with bio-inspired spike cameras [C]. Taiwan: 2022 IEEE International Conference on Multimedia and Expo (ICME), 2022.
- [128] Zheng Y J, Zhang J Y, Zhao R, et al. SpikeCV: Open a continuous computer vision era [EB/OL]. (2023-05-21)[2023-11-01]. <https://arxiv.org/abs/2303.11684>.
- [129] Kim J, Bae J, Park G, et al. N-imagenet: Towards robust, fine-grained object recognition with event cameras [C]. Montreal: The IEEE/CVF International Conference on Computer Vision (ICCV), 2021.