摘要

随着工业检测、自动驾驶等领域对高速成像技术需求的不断增长，传统基于帧的传感器在高速动态场景中暴露出一系列固有缺陷：同步曝光机制导致运动模糊与噪声的权衡困境，动态范围受限引发过曝与欠曝问题；而传统高速相机则具有较高的成本，且其高时间分辨率带来的冗余与存储带宽之间的矛盾制约了连续场景记录能力。在此背景下，神经形态相机通过异步采样机制突破传统成像限制，其中脉冲相机作为积分型神经形态传感器的代表，通过光子累积与脉冲发放机制，能够以 40,000 Hz 的超高采样率捕获具有更多纹理细节的运动场景，实现了神经形态相机从“运动轮廓感知”到“全纹理感知”的跨越。然而，脉冲相机的实用化进程仍面临挑战：其异步二值稀疏脉冲流难以直接适配传统视觉算法，且超高时间分辨率与低空间分辨率、无色彩感知的特性限制了其在复杂场景中的应用。本文的研究课题是面向脉冲相机的图像重建与轻量化方法研究，本文的研究力求拓宽脉冲相机的应用空间，释放脉冲相机在高速独立成像和高速辅助成像下的潜力。

具体来说，脉冲相机在独立成像与辅助成像上面临多重挑战。在独立成像方面，脉冲相机面临：（1）模拟场景与真实场景之间的领域偏移问题，现有基于插帧或物理仿真的数据集受噪声建模不完善、场景覆盖不足的制约，相关基于模拟数据集训练的有监督重建算法在真实场景泛化性不足；（2）重建质量与计算效率的两难权衡，脉冲相机的超高时间分辨率对算法效率有着较高要求，而现有算法往往难以在重建质量和效率之间取得平衡。辅助成像方面，脉冲相机与传统相机的互补性优势使其在辅助成像领域具有显著潜力，然而两种相机存在空时分辨率严重失配问题，跨模态信息互补机制尚未被充分探索。针对上述挑战，本文围绕脉冲相机在独立成像与辅助成像中的图像重建任务展开系统性研究，做出以下贡献：

第一，针对有监督脉冲重建算法受限于模拟数据集且存在领域偏移、在真实场景上表现不佳的问题，提出首个面向脉冲相机的自监督高质量重建算法，通过引入盲点网络与构造合适的伪标签实现了重建任务的自监督式学习，并提出自适应运动推理模块来更好地进行脉冲表征，设计了互学习框架以进一步提升计算效率和算法性能。实验表明，所提自监督范式明显优于无监督方法，且在真实数据上进行微调后展现出了显著优于有监督算法的效果，具有更高的灵活性和实用价值。

第二，针对脉冲相机和传统相机之间的信息互补特性和严重的空时分辨率失配问题，提出首个基于脉冲辅助的动态场景重建算法，构建了脉冲域和图像域双向信息互补框架，提出了高效的信息融合模块来实现两种模态信息的高效协同交互，利用模糊

图像中的高分辨率信息辅助脉冲流的超分辨率重建，并进一步通过细化脉冲特征辅助彩色清晰图像的重建。其次，进一步提出基于零值域分解的学习方式，解耦了基于脉冲辅助的动态场景重建任务中的一致性学习和真实性学习，在增强可解释性的同时提升了网络性能。针对脉冲相机和传统相机空间分辨率严重失配带来的纹理信息缺失问题，引入高效扩散模型增强重建结果，实现高质量重建。

第三，针对现有脉冲重建算法网络结构复杂、计算复杂度高而难以实时应用的问题，提出基于纯 1×1 卷积的高度共享化的轻量化脉冲重建网络设计方案，通过一系列的特征拆分、特征偏移、分组卷积的操作有效地扩展了感受野，并通过一系列通道混洗、分组卷积的操作高效地实现了特征融合。所提方案高度模块化，可以有效适配到有监督和自监督脉冲重建及基于脉冲辅助的动态场景重建任务上，在自监督模式下以 4.3% 的参数量和 7.4% 的乘法-累加运算量超越之前的自监督脉冲重建网络，在有监督模式下以 26.5% 的参数量和 19.7% 的乘法-累加运算量领先目前最先进的有监督脉冲重建方法。

最后，本文集成了上述研究方案，构建了一个交互式脉冲影像重建系统，直观展示了本文所提各种重建方案对于真实数据高效、高质量的处理过程，并提供了在线微调和基于高效扩散模型的生成式增强的功能，使算法能够更好地适配到真实场景上、获得更好的可视化效果。

综上所述，本文从学习范式、模态融合、计算效率三个维度在脉冲相机实用化关键问题上取得突破。提出的自监督重建框架实现了真实场景下的高质量脉冲重建，跨模态互补机制为混合成像任务提供了新思路，而轻量化方案则为脉冲相机的实时应用奠定基础。

关键词：脉冲相机，图像重建，自监督算法，运动去模糊，轻量化网络

Research on Image Reconstruction and Lightweight Methods for Spike Cameras

Shiyan Chen (Computer Science)
Directed by Assistant Prof. Zhaofei Yu

ABSTRACT

With the increasing demand for high-speed imaging technology in fields such as industrial inspection and autonomous driving, traditional frame-based sensors have revealed a series of inherent defects in high-speed dynamic scenes: the synchronous exposure mechanism leads to the dilemma of balancing motion blur and noise, and the limited dynamic range causes overexposure and underexposure issues. Traditional high-speed cameras, on the other hand, are costly, and the contradiction between the redundancy brought by their high temporal resolution and the storage bandwidth constrains their ability to record continuous scenes. Against this backdrop, neuromorphic sensors break through traditional imaging limitations through asynchronous sampling mechanisms. Among them, the spike camera, as a representative of integrating-type neuromorphic sensors, can capture motion scenes with more texture details at an ultra-high sampling rate of 40,000 Hz through the photon accumulation and spike firing mechanism, achieving the leap from "motion contour perception" to "full texture perception" in neuromorphic sensors. However, the practical application of spike cameras still faces challenges: their asynchronous binary sparse spike streams are difficult to directly adapt to traditional vision algorithms, and their characteristics of ultra-high temporal resolution, low spatial resolution, and lack of color perception limit their application in complex scenes. The research topic of this paper is the research on image reconstruction and lightweight methods for spike cameras, aiming to expand the application space of spike cameras and unleash their potential in high-speed independent imaging and high-speed assisted imaging.

Specifically, spike cameras face multiple challenges in both independent and assisted imaging. In terms of independent imaging, spike cameras are confronted with: (1) the domain shift problem between synthetic and real scenes. Existing datasets based on frame interpolation or physical simulation are limited by imperfect noise modeling and insufficient scene coverage, and supervised reconstruction algorithms trained on synthetic datasets have

poor generalization in real scenes; (2) the dilemma of balancing reconstruction quality and computational efficiency. The ultra-high temporal resolution of spike cameras demands high algorithm efficiency, while existing algorithms often fail to strike a balance between reconstruction quality and efficiency. In terms of assisted imaging, the complementary advantages of spike cameras and traditional cameras give them significant potential in this field. However, the severe mismatch in spatiotemporal resolution between the two types of cameras has not been fully explored for cross-modal information complementarity. To address the above challenges, this paper systematically studies the image reconstruction tasks of spike cameras in both independent and assisted imaging and makes the following contributions:

Firstly, to address the issue that supervised spike reconstruction algorithms are limited by synthetic datasets and perform poorly in real scenes due to domain shift, the first self-supervised high-quality reconstruction algorithm for spike cameras is proposed. By introducing a blind-spot network and constructing suitable pseudo-labels, self-supervised learning for the spike reconstruction task is achieved. An adaptive motion inference module is also proposed to better represent spikes and a mutual learning framework is designed to further enhance computational efficiency and algorithm performance. Experiments show that the proposed self-supervised paradigm significantly outperforms unsupervised methods and, after fine-tuning on real data, demonstrates significantly better results than supervised algorithms, with higher flexibility and practical value.

Secondly, to address the complementary characteristics and severe spatiotemporal resolution mismatch between spike cameras and traditional cameras, the first dynamic scene reconstruction algorithm based on spike-assisted imaging is proposed. A bidirectional information complementary framework between the spike domain and the image domain is constructed and an efficient information fusion module is proposed to achieve efficient collaborative interaction between the two modalities. The high-resolution information in blurred images is used to assist the super-resolution reconstruction of spike streams and spike features are further refined to assist the reconstruction of color clear images. In addition, a learning method based on range-null space decomposition is proposed, which decouples the consistency learning and authenticity learning in spike-assisted dynamic scene reconstruction tasks. This enhances interpretability while improving network performance. To address the texture information loss caused by the severe spatial resolution mismatch between spike cameras and traditional cameras, an efficient diffusion model is introduced to enhance the reconstruction results and achieve high-quality reconstruction.

Thirdly, to address the issue that existing spike reconstruction algorithms have complex network structures and high computational complexity, making them difficult to apply in real-time, a lightweight spike reconstruction network design scheme based on pure 1×1 convolution with high sharing is proposed. Through a series of operations including feature split, feature shift, and grouped convolution, the receptive field is effectively expanded. A series of channel shuffle and grouped convolution operations efficiently achieve feature fusion. The proposed scheme is highly modular and can be easily adapted to both supervised and self-supervised reconstruction tasks. In self-supervised mode, it outperforms previous self-supervised networks with only 4.3% of the parameters and 7.4% of the multiply–accumulate operations. In supervised mode, it outperforms the state-of-the-art supervised solution with 26.5% of the parameters and 19.7% of the multiply–accumulate operations.

Finally, an interactive spike image reconstruction system is constructed by integrating the aforementioned research solutions. This system intuitively demonstrates the efficient and high-quality processing of real data by the various reconstruction methods proposed in this paper. It also provides the functions of online fine-tuning and generative enhancement based on an efficient diffusion model, enabling the algorithm to better adapt to real scenes and achieve superior visualization results.

In summary, this paper makes breakthroughs in key practical challenges of spike cameras from three dimensions: learning paradigms, modality fusion, and computational efficiency. The proposed self-supervised reconstruction framework achieves high-quality spike reconstruction in real scenes. The cross-modal complementary mechanism provides new ideas for hybrid imaging tasks, and the lightweight solution lays the foundation for real-time applications of spike cameras.

KEYWORDS: Spkie Camera, Image Reconstruction, Self-Supervised, Motion Deblurring, Lightweight Network