摘要

视觉目标检测是通过特定传感器采集目标的视觉信息，并计算感兴趣目标物体的位置信息和类别属性。目标精准检测是目标跟踪、行为理解、视频对象检索等高级视觉任务的基础，并广泛应用于自动驾驶、视频监控和人机交互等领域。现有大多数图像帧相机的采样帧率和动态范围低，在高速运动场景会出现运动模糊或弱光照场景存在目标对比低，从而极大制约了目标精准检测。神经形态视觉传感器（如 DVS 和 Vidar）具有高时域分辨率、高动态范围和数据冗余小等优势，为实现极端场景下目标精准检测提供了一种可行方案。然而，神经形态视觉是以异步时空脉冲信号表示与处理视觉信息，如何从异步脉冲流里精准地检测目标是仍未解决的问题，因此研究神经形态视觉目标检测是具有重大挑战与应用价值。本文构建了神经形态视觉目标检测的评测基准，聚焦于“异步脉冲信号精准表征”、“脉冲流时序信息高效挖掘”和“多路异步视觉流互补融合”三个科学问题，分别从三种视觉流模态组合方式上进行层次递进的研究，并在高速运动目标实时检测系统进行了验证。本文取得的主要研究成果包括：

第一，针对“神经形态视觉目标检测”这一全新研究问题，率先建立了神经形态视觉目标检测的评测基准，为系统性研究目标检测方法奠定了坚实的技术基础。本文首先形式化定义了神经形态视觉目标检测问题，并深入分析了神经形态视觉目标检测的典型特点。为了解决公开数据集稀缺的问题，本文依据不同任务需求构建了多个大规模神经形态视觉目标检测数据集，其包括首个开源的 PKU-DDD17-CAR 数据集、百万级规模标注的 PKU-DAVIS-DVS 数据集、时域连续标注的 PKU-Vidar-DVS 数据集和 KITTI-Vidar-DVS 仿真数据集。在此基础上，本文阐述了神经形态视觉目标检测的评价指标，并对神经形态视觉目标检测数据集进行了基准测试。

第二，针对 DVS 脉冲流时序信息挖掘的问题，提出了一种基于异步时空记忆网络的脉冲流式目标检测方法，实现了具有高效挖掘脉冲流时序信息的全新目标检测框架。为了从信号处理角度来定量评估 DVS 脉冲流的时域相关性，本文提出了一种异步时空脉冲信号度量方法，并定义了脉冲流式目标检测问题。在此基础上，本方法设计了一种时域自适应采样策略，将连续 DVS 脉冲流划分为离散脉冲块，其生成的脉冲块对场景运动速度具有鲁棒性；同时本方法设计了一种时域注意力卷积，将离散脉冲块内的异步脉冲信号编码为可兼容深度网络的脉冲张量。此外，本方法首次提出了一种轻量级循环卷积结构，来高效地挖掘连续 DVS 脉冲流的时序信息，并较好地实现了目标检测精度与推理速度的权衡。实验结果表明，本方法相比国际前沿方法 RED 在 Gen1 Detection 数据集和 1Mpx Detection 数据集分别提升了 6.7% 和 5.4%，尤其比 6 种前馈

式神经形态视觉目标检测方法有更为显著的性能提升。

第三，针对 DVS 脉冲流和图像帧异构融合的问题，提出了一种基于时空 Transformer 的多路流式目标检测方法，实现了两路互补性异构视觉流的联合目标检测框架。本方法首次设计了一种时域 Transformer 结构对 DVS 脉冲流和图像帧序列进行时序建模，来充分挖掘两路视觉流的时域信息。同时，本方法提出了一种证据判决的异步融合模块，以异步方式来融合两路视觉流，并依据任务需求灵活地调整目标检测器的推理频率。实验结果表明，本方法在 PKU-DDD17-CAR 数据集和 PKU-DAVIS-SOD 数据集上分别取得了 0.929 和 0.501 的平均正确率（mAP），并超越 4 种国际先进的目标检测算法，尤其比单一模态方法有更为显著的性能提升。

第四，针对 DVS 和 Vidar 多路异步脉冲流融合的问题，提出了一种动态交互融合网络的仿视网膜目标检测方法，实现了高速运动或极端光照条件下的目标高精度检测框架。本方法设计了一种时域特征聚合模块，来对两路异步脉冲流分别进行精准表征，并充分挖掘异步脉冲流的丰富时域信息。同时，本方法首次提出了一种受视网膜机理启发的动态交互融合结构，其充分融合 DVS 和 Vidar 脉冲流互补特性来提高目标检测精度。实验结果表明，本方法在 PKU-Vidar-DVS 数据集和 KITTI-Vidar-DVS 仿真数据集上分别取得了 0.647 和 0.762 的平均正确率，有效地解决了单一模态方法在高速运动或低光照等极端场景下目标漏检问题。

上述模型与方法**应用到高速运动目标实时检测系统中，验证了采用神经形态视觉方法来实现高速运动目标实时精准检测的可行性。**在系统实现上，本文进一步设计了一种基于神经元动态响应的目标检测方法，并将其部署到 FPGA 硬件实时处理平台，在高速运动弹丸目标检测与躲避场景下，本系统的目标检测率高达 99%，并实现了 4.2 米范围内 150 千米/小时高速弹丸的实时躲避功能。

综上所述，本文构建了神经形态视觉目标检测的评测基准，提出了多项创新性的神经形态视觉目标检测方法，实现了一种高速运动目标的实时检测系统。本文研究工作为神经形态视觉目标检测领域的后续深入研究奠定了基础。

关键词：神经形态视觉传感器，目标检测，类脑视觉，机器学习，神经形态工程

Neuromorphic Object Detection in Asynchronous Visual Streams

Jianing Li (Computer Application Technology)
Directed by Prof. Yonghong Tian

ABSTRACT

Vision-based object detection is to obtain the spatial location and the category of the interest objects in visual streams. High-accuracy object detection is the basis for high-level vision tasks (e.g., object tracking, human behavior understanding, and video object retrieval). It has been widely used in autonomous driving, video surveillance, human-computer interaction, etc. Due to the low sampling frame rate and low dynamic range of conventional frame-based cameras, there will be motion blur in high-speed scenes or low object contrast in low-light scenarios, which results in a sharp drop in performance using unusable images. Neuromorphic vision sensors (e.g., DVS and Vidar), offering high temporal resolution, high dynamic range, and low redundancy, have brought a new perspective to overcoming these object detection challenges. However, spatiotemporal spikes are used to represent and process visual information in neuromorphic vision. How to accurately detect objects in asynchronous spike streams is a challenging and valuable research topic. Therefore, this thesis focuses on the research of object detection using neuromorphic vision sensors, and aims at addressing three key issues including "designing neuromorphic representation for asynchronous spike streams", "efficient leveraging temporal cues from spike streams" and "developing complementary fusion models for asynchronous visual streams". This thesis explores neuromorphic object detectors from three perspectives of input modalities. On this basis, the optimized method is verified in a real-world high-speed object detection system. The main contributions of this thesis are summarized as follows:

Firstly, this thesis builds a benchmark for the emerging research topic of neuromorphic object detection, which provides a solid technical foundation for the systematic study of the corresponding algorithms. This thesis first formally defines the novel problem setting of neuromorphic object detection, and then analyzes the unique attributes of neuromorphic object detection compared with conventional frame-based object detection. To solve the lack of related public datasets and support model training, this thesis builds four large-scale neuromorphic object detection datasets, which include the first open-source PKU-DDD17-CAR dataset, the PKU-DAVIS-SOD dataset with millions of labels, the PKU-Vidar-DVS with temporal continuous labels, and the KITTI-Vidar-DVS simulated dataset. Besides, this thesis elaborates on the evaluation metrics for neuromorphic object detection and benchmarks the related object detection datasets.

Secondly, this thesis proposes a streaming object detection method via an asynchronous spatiotemporal memory network, which realizes a novel object detection framework with efficient leveraging of temporal cues from DVS spike streams. To introduce the concept of streaming object detection, an asynchronous spatiotemporal spike metric scheme is proposed to quantitatively evaluate the temporal correlation in a continuous spike stream. This method designs a temporal adaptive sampling strategy to split the continuous stream into discrete spike bins robust to motion speed. Meanwhile, this method adopts a temporal attention convolutional network to encode each spike bin into a spike tensor in combination with deep learning techniques. Besides, a lightweight recurrent convolutional network is designed to efficiently use rich temporal cues between spike bins, which achieves a good trade-off between detection performance and inference speed. Empirically, it shows that this method improves the mean average precision (mAP) by 6.7% and 5.4% than the state-of-the-art method RED in the Gen1 Detection dataset and the 1Mpx Detection dataset, and it significantly outperforms the six feed-forward neuromorphic object detectors.

Thirdly, this thesis proposes a multi-streaming object detection method via a spatiotemporal transformer, which realizes a joint object detection framework to make complementary use of DVS spike streams and frames. This method develops a temporal transformer, which models long-term dependencies among DVS spike streams and adjacent frames to fully leverage rich temporal cues from two visual streams. Besides, the Demster-Shafer theory is used to asynchronously fuse two visual streams and flexibly set the output frequency of object detection. The experimental results on the PKU-DDD17-CAR dataset and the PKU-DAVIS-SOD dataset show that this method achieves the mAP of 0.929 and 0.501 respectively. It outperforms four state-of-the-art methods, especially compared to the single-modality methods.

Fourthly, this thesis proposes a retinomorphic object detection method to address common challenges in high-speed motion or low-light scenarios, which develops a dynamic interaction fusion network to integrate DVS and Vidar spike streams. In this method, a temporal aggregation representation strategy is designed to encode two asynchronous streams into spike tensors respectively, and it can fully utilize rich temporal cues from continuous spike streams. Inspired by the interaction between foveal and peripheral signals, a bio-inspired unifying framework to fuse two streams via dynamic interactions between sub-networks. The experimental results show that this method obtains the mAP of 0.647 and 0.762 in the PKU-Vidar-DVS dataset and the KITTI-Vidar-DVS simulated dataset respectively. It has significant improvements over the single-modality methods, especially in high-speed motion and low-light scenarios.

The above-mentioned models and methods are applied to the real-time high-speed object detection system, which verifies the feasibility of the neuromorphic vision scheme for real-time high-accuracy object detection. In this system, a high-speed object detection algorithm is designed via the spiking neural model, and it is deployed on the FPGA for real-time processing. For high-speed projectile object detection and avoidance tasks, the detection accuracy of high-speed projectiles is close to 99%, and this system is capable of evading 150 km/h high-speed projectiles within a range of 4.2 meters.

In conclusion, this thesis builds a comprehensive benchmark and proposes several innovative neuromorphic object detection methods. On this basis, this thesis develops a real-time high-speed object detection system. Moreover, this thesis builds a solid technical foundation for the follow-up research work in the field of neuromorphic object detection.

KEYWORDS: Neuromorphic Vision Sensors, Object Detection, Brain-Inspired Vision，Machine Learning, Neuromorphic Engineering