摘 要

传统相机成像模式在捕获瞬态过程时面临着高帧率和高动态范围难以兼顾的“两难困境”，而脉冲相机通过模拟哺乳动物视网膜中央凹的“积分—发放”原理与异步采样机制有效化解了短曝光噪声与长曝光模糊之间的矛盾，实现了超高帧率与超宽动态范围。然而，海量二值脉冲数据也给存储和传输带来了巨大压力：一方面，脉冲数据受压缩率要求高、时空相关性弱和失真函数难以定义等因素影响，难以通过传统编码框架实现高效压缩；另一方面，脉冲处理方法在实际部署过程中还面临端侧设备计算资源受限、多任务场景下码流通用性差和复杂场景感知精度不足等问题，严重制约了脉冲信息处理系统的规模化应用。本文针对上述挑战提出了三项面向脉冲相机成像信号的高效编码方法，围绕编码压缩效率、解码信息失真和解码感知质量构建了更优的帕累托平衡。本文的主要创新点总结如下：

第一，提出了一种场景重建引导的脉冲跨模态编码方法，以应对脉冲数据二值离散特性和弱时空相关性导致传统编解码方法难以高效压缩的问题。该方法引入压缩友好性更强的场景模态作为中间媒介，构建了基于重建的上下文脉冲编码方法（Reconstruction-based Contextual Spike Compression，RCSC）；同时，该方法利用脉冲数据中的时空上下文信息建模运动显著性以指导码率自适应分配，并使用场景信号损失近似表示脉冲信息损失以实现端到端优化。实验结果表明，该方法在解码场景和解码脉冲维度上均取得了优于传统编码器和现有神经网络编码器的性能，为脉冲信号高效高保真编码建立了新的技术基准。

第二，提出了一种基于多尺度时空相关性的脉冲预测编码方法，以解决场景模态编码过程中存在的信息瓶颈和解码先验难以利用的问题。该方法将脉冲预测从信号域扩展至语义特征域，基于混合时域滤波提取的多尺度子特征和多尺度时空先验构建了多尺度时空聚合预测单元（Multi-scale Spatio-Temporal Aggregative Prediction Unit，MSTAPU），进而以预测结果为先验提出了基于多尺度聚合预测的脉冲编码方法（Multi-scale Aggregative Prediction-based Spike Compression，MAPSC）。实验结果表明，该方法在脉冲预测和脉冲编码任务中均取得了显著性能提升，验证了多尺度时空相关性建模对提升脉冲编码效率与解码质量的有效

性。

第三，提出了一种面向机器视觉的脉冲感知编码方法，以应对端侧计算复杂度高、复杂场景感知精度不足和多任务码流难以复用等部署应用挑战。该方法将脉冲编码与视觉感知纳入统一优化框架，基于通用特征编码感知策略提出了面向通用视觉感知的脉冲编码方法（Spike Compression for Versatile Visual Perception，SCVVP），通过“腹侧—背侧”双分支架构分别提取空间语义信息与运动动态信息，并通过“显式—隐式”联合对齐策略对解码特征进行时域对齐与融合感知。实验结果表明，该方法能够在保持较高编码效率的同时支撑多项视觉感知任务，并显著降低端侧计算复杂度与推理时间，为脉冲视觉系统在端云协同架构下的实际部署提供了有效支撑。

综上所述，本文针对脉冲数据高效编码的核心挑战进行了系统分析，围绕脉冲高效压缩和系统部署应用等方面分别提出了脉冲跨模态编码方法、脉冲预测编码方法和脉冲感知编码方法，在显著提升脉冲数据编码压缩效率和解码感知质量的同时，推动了模型在复杂场景和多任务场景下的高效部署推理。

关键词：脉冲相机，脉冲高效编码，脉冲信息处理，深度学习

Abstract

Conventional camera imaging paradigms face an inherent dilemma between high frame rate and high dynamic range when capturing transient processes. By emulating the Integrate-and-Fire principle and asynchronous sampling mechanism of the mammalian retinal fovea, spike cameras effectively reconcile short-exposure noise with long-exposure motion blur, thereby achieving ultra-high frame rates and ultra-wide dynamic range. However, massive binary spike signals impose substantial burdens on storage and transmission. On the one hand, stringent compression-ratio requirements, weak spatio-temporal correlations, and the difficulty of defining distortion measurements make spike signals difficult to compress efficiently within conventional coding frameworks. On the other hand, practical spike processing methods remain constrained by limited edge-side computational resources, poor bitstream versatility across multiple tasks, and insufficient perceptual accuracy in complex scenarios, severely hindering the scalable deployment of spike information processing systems. To address these challenges, this thesis proposes three efficient compression methods for spike camera imaging signals, aiming to establish a more favorable Pareto trade-off among compression efficiency, decoded information distortion, and decoded perceptual quality. The main contributions are summarized as follows.

Firstly, this thesis proposes a Cross-modal Spike Compression method guided by scene reconstruction to overcome the inefficient compression caused by the binary discrete nature and weak spatio-temporal correlations of spike signals. By introducing a more compression-friendly scene modality as an intermediate representation, the method develops Reconstruction-based Contextual Spike Compression (RCSC). It further exploits spatio-temporal contextual information to model motion saliency for adaptive bit allocation, and approximates spike information loss with scene signal loss to enable end-to-end optimization. Experimental results demonstrate that RCSC surpasses conventional and neural codecs in both decoded scene and decoded spike quality, establishing a new benchmark for efficient and high-fidelity spike signal compression.

Secondly, this thesis proposes a Spike Predictive Compression method based on multi-scale spatio-temporal correlations to alleviate the information bottleneck in scene-modality compression and facilitate the exploitation of decoded priors. This method extends spike prediction from the signal domain to the semantic feature domain, and constructs a Multi-scale Spatio-Temporal Aggregative Prediction Unit (MSTAPU) based on multi-scale sub-features extracted by hybrid temporal filtering and multi-scale spatio-temporal priors. Using the prediction results as priors, it further proposes Multi-scale Aggregative Prediction-based Spike Compression (MAPSC). Experimental results show that this method achieves significant gains in both spike prediction and spike compression, validating the effectiveness of multi-scale spatio-temporal correlation modeling in improving compression efficiency and decoded quality.

Thirdly, this thesis proposes a Spike Perceptual Compression method towards machine vision to address deployment challenges, including high edge-side computational complexity, insufficient perceptual accuracy in complex scenarios, and limited bitstream reusability across multiple tasks. By integrating spike compression and visual perception into a unified optimization framework, this method proposes Spike Compression for Versatile Visual Perception (SCVVP) based on a general feature-based compression-for-perception strategy. It employs a ventral–dorsal dual-branch architecture to extract spatial semantic information and motion dynamics, respectively, and adopts an explicit–implicit joint alignment strategy to perform temporal alignment and fusion perception on decoded features. Experimental results demonstrate that SCVVP supports multiple visual perception tasks while maintaining high compression efficiency, and substantially reduces edge-side computational complexity and inference latency, providing effective support for the practical deployment of spike vision systems under edge–cloud collaborative architectures.

In summary, this thesis systematically investigates the core challenges of efficient spike compression. Focusing on efficient compression method and practical system deployment, it proposes Cross-modal Spike Compression, Spike Predictive Compression, and Spike Perceptual Compression, which substantially enhance spike compression efficiency and decoded perceptual quality while enabling efficient deployment and

inference in complex and multi-task scenarios.

Keywords: Spike Camera, Efficient Spike Compression, Spike Information Processing, Deep Learning