摘要

生物视觉系统经过数亿年进化，形成了高效节能且鲁棒性强的信息处理机制。相较于传统计算机视觉系统，其在计算效率、环境适应性和能耗比方面展现出显著优势。近年来，神经科学研究手段的突破为解析这些精细机制提供了新工具，同时视觉假体与视觉修复、脑机接口以及人工智能在自动驾驶、具身智能等领域的应用需求共同推动了基于生物视觉机制的神经编解码研究。尽管已取得重要进展，但在神经信息处理的三个关键环节仍存在重要挑战：神经网络功能结构的精准解析与编码机制阐释、高维神经信号的特征冗余消除与表征优化和神经元群体时序动态特性的有效挖掘与解码精度提升。针对这些挑战，本文聚焦生物视觉神经信号的编解码方法研究，旨在揭示视觉神经网络的功能结构及信息加工原理，建立高效、高精度的神经信号编解码方法。围绕“视觉神经系统网络功能结构精准辨识”、“多尺度神经信号非冗余特征提取”以及“神经信号时序动态特性高效挖掘”三个核心科学问题，本文提出了一系列创新性方法，并在视觉神经信号编解码系统中进行了系统性验证。本文取得的主要研究成果包括：

（1）针对“视觉神经系统网络功能结构精准辨识问题”，本文发展了脉冲激发非负矩阵分解方法（Spike-Triggered Non-negative Matrix Factorization, STNMF），通过辨识视觉神经系统内参与编码的级联计算组件，为构建生物启发性编码方法提供了理论支撑。STNMF 将脉冲激发刺激序列分解为模块矩阵和权重矩阵：通过分析模块矩阵，推断出突触前神经元的空间感受野、时间滤波器及非线性函数；通过分析权重矩阵，推断出突触连接权重和突触前神经元脉冲子集，从而实现对网络功能结构的系统辨识。其中，推断的突触前神经元脉冲子集结合相关性分析，为系统辨识提供了定量评估方法。此外，相继克服了网络传输信号变化、复杂网络连接以及自然图像刺激等难题，成功将 STNMF 推广至更一般的视觉脉冲神经网络的功能结构辨识。实验结果表明，STNMF 在面对复杂刺激和多种细胞类型的网络时表现卓越，能够有效解析网络的非线性计算特性。最后，基于 STNMF 推断的计算组件构建的编码方法，在动态自然视频刺激下，高精度预测了真实生物数据中神经元群体的发放活动。STNMF 不仅为研究记录的神经元活动提供了神经系统功能结构的解析工具，还为设计更合理、更精准的编码方法奠定了坚实基础。

（2）针对“多尺度神经信号非冗余特征提取问题”，本文提出了一种基于小波条件互信息（Wavelet Conditional Mutual Information, WCMI）的表征方法，通过解耦时频分析特征之间的协同性与冗余性，高效提取高维神经信号的时空模式。WCMI 采用

三阶段特征提取流程：首先，利用多分辨率小波变换生成神经信号的时频系数矩阵；其次，结合条件互信息筛选出一组具有代表性的非冗余特征；最后，通过线性解码器验证其表征效果。WCMI 方法不仅充分考虑了特征之间的协同性和冗余性，还能够表征单个神经元信号（如脉冲信号、钙信号）的时序发放模式，同时有效捕捉神经群体活动（如脉冲信号、双光子信号、脑电皮层信号）的时空特征及其高阶依赖性和稀疏编码特性。此外，通过对关键小波系数进行逆小波变换，该方法还能够实现神经信号的降噪重构，进一步提升信号质量。WCMI 方法能够高效表征高维时空神经数据中的重要特征，为时空神经数据的特征提取提供了一种全新且鲁棒的小波分析框架，不仅提高了特征提取的效率和准确性，还为下游的各种解码和信息还原任务提供了可靠的输入和技术支持。

（3）针对“神经信号时序动态特性高效挖掘问题”，本文提出了一种基于小波引导延迟信号增强（Wavelet-informed Delayed Signal Augmentation, WIDSA）的视觉重建方法，能够高效挖掘神经信号时序动态特性，并从多种神经信号中高精度重建动态自然场景。该方法突破了传统脉冲计数方法的局限，通过结合离散小波变换和深度神经网络，高效提取并整合神经信号中的时序信息，显著提升了视觉场景重建的准确性。此外，本文提出了一种新的神经响应延迟计算方法，不仅能够筛选出对神经信号响应更优的神经元，还能精确捕捉不同神经元对视觉刺激的响应时间，从而优化了刺激-响应对的匹配。实验结果表明，该方法能够更准确地捕捉神经元对外部视觉刺激的响应延迟和发放模式，从而大幅提高了重建图像的质量。进一步研究表明，WIDSA 在不同空间分辨率下展现出良好的重建能力，能够从低分辨率输入中实现 2 倍超分重建，但在 4 倍超分重建上仍面临挑战。最后，本文还验证了 WIDSA 方法在处理彩色视觉刺激下宽场钙信号的表现，尽管其时间分辨率较低，本方法在单个试验数据上仍能有效重建视觉场景，展示了其在处理不同类型神经信号时的鲁棒性和广泛适用性。

最后，基于对生物视觉神经信号编码、表征与重建机制的系统性研究，本研究创新性地构建了一套完整的视觉神经信息处理系统，该系统由编码模块、表征模块和重建模块三大核心组件构成，不仅支持各模块独立运行，更能协同工作形成”编码-表征-重建”的闭环处理通路，前向信息处理反向反馈验证。该系统的创新性的实现了从神经网络结构解析与信号编码，到多尺度特征表征，再到视觉场景重建的全流程整合，揭示了视觉信息从局部神经元响应到全局场景理解的转换机制，建立了结构解析-信号编码-表征优化-功能实现的多层次研究框架。在应用层面，本研究系统对比了各模块的核心技术指标，提出了基于不同应用场景的方法选择策略，为生物视觉启发的神经信息处理研究提供了兼具理论深度和应用灵活性的系统解决方案。

关键词：生物视觉系统，神经信号编码方法，神经信号表征方法，视觉刺激重建方法

Research on Encoding and Decoding Methods of Biological Visual Neural Signals

Shanshan Jia ((Computer application technology) Major)
Supervised by Asst. Prof. Zhaofei Yu

ABSTRACT

The biological visual system has evolved over hundreds of millions of years, forming an efficient, energy-saving, and robust information processing mechanism. Compared with traditional computer vision systems, it shows significant advantages in computational efficiency, environmental adaptability, and energy consumption ratio. In recent years, breakthroughs in neuroscience research methods have provided new tools for parsing these delicate mechanisms. Meanwhile, the application demands of visual prosthetics and visual repair, brain-computer interfaces, and artificial intelligence in fields such as autonomous driving and embodied intelligence have jointly propelled the research on neural encoding and decoding based on biological visual mechanisms. Despite important progress, there are still significant challenges in three key aspects of neural information processing: precise parsing of the functional structure of neural networks and elucidation of encoding mechanisms, elimination of feature redundancy and optimization of representation of high-dimensional neural signals, and effective mining of temporal dynamic characteristics of neuronal population and improvement of decoding accuracy. Focusing on these challenges, this study focuses on the research of encoding and decoding methods of biological visual neural signals, aiming to reveal the functional structure and information processing principles of visual neural networks and establish efficient and high-precision neural signal encoding and decoding methods. Focusing on three core scientific issues—”precise identification of functional structures in visual neural networks”, ”multi-scale neural signal non-redundant feature extraction”, and ”efficient mining of temporal dynamic characteristics in neural signals”—this study proposes a series of innovative methods and systematically validates them in a visual neural signal encoding-decoding system. The main research achievements of this study include:

(1) To address the challenge of ”precise identification of functional structures in visual neural networks,” this study develops Spike-Triggered Non-negative Matrix Factorization,

which identifies cascaded computational components involved in neural coding within the visual system, providing theoretical support for biologically inspired encoding methods. STNMF decomposes spike-triggered stimulus sequences into a module matrix and a weight matrix. By analyzing the module matrix, it infers the spatial receptive fields, temporal filters, and nonlinear functions of presynaptic neurons.By analyzing the weight matrix, it deduces synaptic connection weights and a subset of presynaptic neuron spikes, enabling systematic identification of network functional structures. The inferred subset of presynaptic spikes, combined with correlation analysis, provides a quantitative evaluation method for system identification. Furthermore, STNMF overcomes challenges such as signal variability in network transmission, complex synaptic connectivity, and natural image stimuli, extending its applicability to general functional structure identification in spiking neural networks. Experimental results demonstrate that STNMF excels in decoding nonlinear computational properties even under complex stimuli and diverse cell types. Finally, an encoding model based on STNMF-derived computational components accurately predicts neuronal population activity in response to dynamic natural videos, closely matching real biological recordings. Beyond serving as a functional mapping tool for recorded neural activity, STNMF lays a solid foundation for designing more biologically plausible and precise neural encoding models.

(2) To address the challenge of "multi-scale neural signal non-redundant feature extraction," this study proposes a Wavelet Conditional Mutual Information-based representation method, which efficiently extracts spatiotemporal patterns from high-dimensional neural signals by decoupling the synergy and redundancy among time-frequency analysis features. The WCMI method follows a three-stage feature extraction pipeline: first, it generates a time-frequency coefficient matrix of neural signals using multi-resolution wavelet transform; second, it selects a set of representative non-redundant features by leveraging conditional mutual information; and finally, it validates their representational efficacy through a linear decoder. WCMI not only comprehensively accounts for feature synergy and redundancy but also captures the temporal firing patterns of single-neuron signals (e.g., spikes, calcium signals) while effectively characterizing the spatiotemporal features, higher-order dependencies, and sparse coding properties of neural population activity (e.g., spike trains, two-photon imaging signals, electrocorticography). Moreover, by applying inverse wavelet transform on selected wavelet coefficients, the method enables noise-reduced signal reconstruction, further enhancing signal quality. WCMI provides an efficient and robust wavelet-based framework for extracting critical features from high-dimensional spatiotemporal neural data, significantly improving

feature extraction efficiency and accuracy while offering reliable input and technical support for downstream decoding and information recovery tasks.

(3) To address the challenge of "efficient mining of temporal dynamic characteristics in neural signals", this study proposes a Wavelet-informed Delayed Signal Augmentation-based visual reconstruction method, which effectively extracts temporal dynamics from neural signals and achieves high-precision reconstruction of dynamic natural scenes from diverse neural recordings. This approach overcomes the limitations of traditional spike-counting methods by integrating discrete wavelet transforms with deep neural networks to efficiently extract and integrate temporal information, significantly improving visual scene reconstruction accuracy. Furthermore, we introduce a novel neural response delay computation method that not only identifies neurons with superior response properties but also precisely captures stimulus-response latencies across different neurons, thereby optimizing stimulus-response pairing. Experimental results demonstrate that WIDSA more accurately captures neuronal response delays and firing patterns to external visual stimuli, substantially enhancing reconstructed image quality. Additional studies reveal WIDSA's robust reconstruction capability across varying spatial resolutions, achieving successful 2× super-resolution reconstruction from low-resolution inputs while facing challenges at 4× magnification. Finally, we validate WIDSA's performance on wide-field calcium signals under colored visual stimulation, demonstrating its ability to reliably reconstruct visual scenes from single-trial data despite lower temporal resolution, highlighting the method's robustness and broad applicability across diverse neural signal modalities.

Finally, based on systematic research into the encoding, representation, and reconstruction mechanisms of biological visual neural signals, this study has innovatively developed a complete visual neural information processing system consisting of three core components: an encoding module, a representation module, and a reconstruction module. These modules not only function independently but also work synergistically to form a closed-loop "encoding-representation-reconstruction" processing pathway that integrates forward information processing with reverse feedback verification. The system innovatively achieves end-to-end integration from neural network structure analysis and signal encoding to multi-scale feature representation and visual scene reconstruction, revealing the transformation mechanism from local neuronal responses to global scene understanding in visual information processing, while establishing a multi-level research framework encompassing structural analysis, signal encoding, representation optimization, and functional implementation. At the application level, the study systematically compares the core technical indicators of each module and proposes method selection strategies for different application scenarios, providing a systematic solution for biologically inspired neural information processing research that combines theoretical depth with practical flexibility.

KEY WORDS: Biological vision systems, Neural signal encoding methods, Neural signal representation methods, Visual stimulus reconstruction methods