摘要

视觉皮层作为大脑感知和认知的核心枢纽，其神经群体活动的计算机制始终是计算神经科学界的研究焦点。针对视觉皮层神经群体活动的计算建模，表征与动力学构成了两个互补维度：“表征”定义了从视觉刺激映射至神经活动的编码模式，“动力学”则刻画了神经活动跨时间的演化规律。随着神经采集技术的发展，可记录的视觉神经活动的规模日益庞大且具有高度非线性。面对复杂大规模的神经数据，深度神经网络凭借其强大的特征提取和非线性拟合能力，在建模神经表征与神经动力学方面均展现出优势，已逐渐更迭了依赖人工先验的传统计算模型，成为计算神经科学研究的主流工具。然而，现有的深度神经网络在结构和机制上缺乏一定的生物合理性，导致其对于复杂视觉神经活动的表征模拟精度和动力学可解释性等方面仍存在局限。围绕“揭示视觉神经群体活动的表征编码与动力学演化的规律和机制”这一核心科学问题，本文在融合深度网络优势的基础上，进一步引入生物启发的结构和机制，沿“表征建模”与“动力学建模”双路径，提出并验证了一系列新型算法，构建了面向建模的系统性评测与分析框架，为理解视觉皮层的神经群体计算机制提供更具生物可解释性的见解。主要研究贡献如下：

第一，为克服传统人工神经网络在生物合理性方面的局限，本研究在国内外首次利用深度脉冲神经网络（Spiking Neural Network, SNN）进行生物神经表征的建模，实现了大规模跨物种的表征相似性评估实验。通过集成多种神经数据集和度量指标，本研究全方位对比了深度脉冲神经网络与传统卷积神经网络、视觉 Transformer 在表征相似性上的性能。实验证实，通过引入模拟了膜电位积分和阈值发放机制的脉冲神经元，深度脉冲神经网络在跨物种神经表征建模方面展现出显著优于传统人工神经网络的性能。在整体架构和网络层数一致的条件下，脉冲神经网络的表征相似性得分的增幅达 6.6%。基于表征相似性的分析，研究进一步探讨了跨物种视觉皮层的功能层级结构和信息加工机制。结果表明了猕猴与小鼠之间的显著差异：猕猴视觉皮层呈现清晰的递进式顺序结构且对异质刺激有着特化的加工机制，而小鼠则表现出更为扁平化的组织特征且皮层间有着并行的信息处理流。

第二，鉴于现有视觉皮层表征建模研究主要聚焦于静态图像刺激，导致对自然动态刺激下神经时空整合机制的理解仍较为有限，本研究提出了一种集成脉冲动力学和反馈连接的长程反馈脉冲网络（Long-Range Feedback Spiking Network, LoRaFB-SNet），旨在实现对视觉皮层动态表征的有效模拟与定量解析。基于前述研究确立的脉冲动力学在静态表征建模中的优势，本研究进一步引入模拟生物皮层自上而下循环机制的长程反馈结构，构建了更具生物合理性的深度神经网络。为评估蕴含时间依赖性的动态表

征，本研究开发了时序表征相似性分析（Time-Series Representational Similarity Analysis, TSRSA）方法。实验结果显示，LoRaFB-SNet 在处理自然电影刺激时表现出优越的时空特征整合能力，揭示了长程反馈连接和视频预训练任务的协同效应。更为重要的是，其表征相似性优势随刺激时序的延长而显著增强，在长时程刺激下较主流类脑模型实现了 26.8% 的性能增幅。研究进一步通过定向刺激扰动实验解析了动态信息和静态信息对表征相似性的差异化影响，一方面说明视觉皮层在动态刺激下的神经群体编码具有显著的上下文相关性，另一方面为生物视觉系统在复杂环境下的功能稳定性提供了潜在的计算解释。

第三，针对现有神经动力学隐变量模型多局限于运动脑区，而在解析复杂视觉神经动力学方面仍显不足的现状，本研究提出了一种时变视觉动力系统模型（Time-Evolving Visual Dynamical System, TE-ViDS），旨在挖掘视觉神经群体活动的动力学演化规律并提取刺激相关的神经成分。前两项研究重点采用仿生结构建模视觉神经静动态表征，本研究则在此基础上，深入解析驱动表征跨时间演化的底层动力学。视觉皮层在处理高维外部输入的同时，也受到唤醒度、注意力等内源性状态波动的影响，这种多源信号的非线性耦合形成了视觉神经动力学的复杂演化规律。为此，TE-ViDS 通过引入显式时序结构和专门设计的隐变量解耦策略，成功捕捉了神经群体活动中受外部刺激驱动和内部状态波动影响的协同演化规律。在合成数据和真实小鼠神经数据的双重验证下，TE-ViDS 在隐变量重建和跨受试个体视觉刺激解码等任务中表现卓越，其解码准确率相较现有高性能隐变量模型平均提升达 16.1%。实验进一步证实了该模型提取的隐变量具有显著的生物学意义，包括识别不同类别刺激下的表征敏感度差异、解析个体间变异性的来源以及捕捉长时程刺激下的神经适应特征。这些结果说明了 TE-ViDS 不仅能从视觉神经活动中提取出与刺激高度相关的特征，更为深入理解神经群体活动的底层动力学规律提供了重要见解。

第四，针对当前视觉皮层神经群体活动建模领域缺乏集成化工具的现状，本研究构建了一套标准化评测与分析框架，旨在系统性地度量神经表征和神经动力学的建模性能，进而解析视觉皮层的信息加工机制。本研究深度整合了前三项研究的建模算法和分析范式，构建了自动化的评估体系和友好便捷的交互接口，以此提升神经群体活动建模的分析效率和工具实用性。通过构建涵盖数据集成、核心度量以及交互展示的分层架构，该框架实现了从标准化数据接入、多维指标自动计算到交互式可视化分析的端到端自动化流程。应用示例表明，该框架支持自定义数据集和多样化模型架构的量化评估，从而有效提升复杂视觉神经数据集上的建模研究迭代效率，并为视觉皮层计算机制分析提供必要支撑。

综上所述，本文开展了涵盖神经表征模拟和神经动力学解构的系统性神经群体活动建模研究。通过在多范式神经数据上的广泛验证，本文不仅证实了生物启发深度网

络相比于传统深度网络的显著优势，更成功揭示了视觉神经活动的表征机制和动力学演化规律。研究成果不仅在科学认知层面深化了对视觉皮层功能层级结构、时序整合机制及群体活动动力学的理解，也在工程层面为高维视觉神经数据的分析提供了高性能建模工具和标准化实验框架。在深度学习与计算神经科学的交叉探索中，本文为开发具有高度生物合理性的视觉计算模型提供了支持。

关键词：视觉皮层，神经表征，神经动力学，任务驱动的计算模型，隐变量模型

Representational and Dynamical Modeling of Visual Neural Population Activity

Liwei Huang (Computer Applied Technology)
Supervised by Prof. Yonghong Tian

ABSTRACT

The visual cortex serves as a core hub for brain perception and cognition, and the computational mechanisms of its neural population activity have always been a research focus within the computational neuroscience community. In the computational modeling of visual neural population activity, representation and dynamics constitute two complementary dimensions: 'representation' defines the encoding patterns mapped from visual stimuli to neural activity, while 'dynamics' characterizes the evolutionary laws of neural activity over time. With advancements in neural recording technologies, the acquired data of visual neural activity have become increasingly large-scale and highly nonlinear. To address such complex data, deep neural networks, empowered by their superior feature extraction and nonlinear fitting capabilities, have demonstrated distinct advantages in modeling both neural representations and dynamics. They have progressively superseded traditional models reliant on manual priors, emerging as a dominant tool in the field. However, current deep neural networks still face challenges regarding biological plausibility in their architectures and mechanisms, leading to limitations in the accuracy of representational simulation and the interpretability of neural dynamics for complex visual activities. Centering on the core scientific question of "uncovering the patterns and mechanisms underlying the representational encoding and dynamical evolution of visual neural activity", this thesis merges the advantages of deep neural networks with biologically inspired structures and mechanisms. By pursuing a dual pathway of 'representational modeling' and 'dynamical modeling,' a series of novel algorithms are proposed and validated, and a systematic benchmarking and analysis framework for neural modeling is established. This thesis provides more biologically interpretable insights into the population-level computational mechanisms of the visual cortex. The primary contributions are as follows:

First, to overcome the limitations of traditional artificial neural networks (ANNs) in biological plausibility, this work pioneers the use of deep Spiking Neural Networks (SNNs) for modeling biological neural representations and conducts large-scale cross-species represen-

tational similarity evaluation experiments. By integrating diverse neural datasets and metrics, this work comprehensively compares the representational similarity performance of deep SNNs against traditional Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The results demonstrate that, by incorporating spiking neurons that simulate membrane potential integration and threshold-firing mechanisms, deep SNNs significantly outperform traditional ANNs in cross-species neural representation modeling. Under the condition of identical overall architecture and network depth, SNNs achieve a 6.6% increase in representational similarity scores. Furthermore, analysis based on representational similarity reveals distinct differences in functional hierarchy and information processing between macaques and mice: the macaque visual cortex exhibits a clear progressive deep hierarchy with specialized processing for heterogeneous stimuli, whereas the mouse visual cortex displays flatter organizational features with parallel processing streams.

Second, given that current modeling of visual cortex representations predominantly focuses on static image stimuli, the understanding of neural spatio-temporal integration mechanisms under natural dynamic stimuli remains limited. To address this, this work proposes the Long-Range Feedback Spiking Network (LoRaFB-SNet). This model integrates spiking dynamics with feedback connections to achieve effective simulation and quantitative analysis of dynamic neural representations. Building upon the advantages of spiking dynamics in static representation modeling established by the aforementioned work, this work incorporates a long-range feedback structure to mimic top-down recurrent mechanisms, creating a more biologically plausible deep neural network. To evaluate dynamic representations with inherent temporal dependencies, a Time-Series Representational Similarity Analysis (TSRSA) method is developed. Experimental results demonstrate that LoRaFB-SNet exhibits superior spatiotemporal feature integration capabilities when processing natural movie stimuli, revealing a synergistic effect between long-range feedback connections and video pre-training tasks. Importantly, its representational similarity advantage significantly scales as the stimulus duration increases, achieving a 26.8% performance gain over mainstream bio-inspired models under long-duration stimulus conditions. Through directional stimulus perturbation experiments, this work further deciphers the differential impacts of dynamic and static information on representational similarity. These findings indicate that neural population coding in the visual cortex exhibits significant context-dependence under dynamic stimuli, while providing a potential computational explanation for the functional stability of biological visual systems in complex environments.

Third, addressing the fact that existing latent variable models for neural dynamics are

predominantly restricted to motor cortex and insufficient for parsing complex visual neural dynamics, this work introduces the Time-Evolving Visual Dynamical System (TE-ViDS). This model is designed to unearth the dynamical evolution of visual neural activity and extract stimulus-related components. While the previous two studies focused on bio-inspired structural modeling for static and dynamic representations, this work delves deeper into the underlying dynamics driving the temporal evolution of these representations. While processing high-dimensional external visual inputs, the visual cortex is simultaneously modulated by fluctuations in internal states such as arousal and attention. The nonlinear coupling of these multi-source signals manifests as the complex evolutionary laws of visual neural dynamics. To this end, through explicit temporal structures and specialized latent disentangling strategies, TE-ViDS successfully captures the synergistic evolution of neural population activity driven by external stimuli and internal state fluctuations. Validated on both synthetic and real mouse neural data, TE-ViDS shows exceptional performance in latent reconstruction and cross-subject visual decoding tasks, achieving an average 16.1% improvement in decoding accuracy over existing high-performance models. Experiments further confirm that the extracted latent variables by TE-ViDS are highly biologically significant, enabling the identification of representational sensitivity differences across stimulus categories, the analysis of sources of inter-individual variability, and the capturing of neural adaptation features under long-term stimulation. These results demonstrate that TE-ViDS not only extracts features highly correlated with visual inputs but also provides crucial insights into the underlying dynamical laws of visual neural population coding.

Fourth, addressing the lack of integrated tools in the field of visual cortical neural population activity modeling, this work constructs a standardized benchmarking and analysis framework. This framework is designed to systematically measure the modeling efficacy of both neural representations and dynamics, thereby deciphering the information processing mechanisms of the visual cortex. By deeply integrating the modeling algorithms and analytical paradigms established in the three aforementioned works, this work constructs an automated evaluation system and a user-friendly interactive interface, significantly enhancing the analytical efficiency and practical utility of neural population activity modeling. By employing a hierarchical architecture encompassing data integration, core metrics, and interactive visualization, the framework enables an automated, end-to-end workflow from standardized data access and multi-dimensional metric calculation to interactive visual analysis. Application examples demonstrate that the framework supports the quantitative evaluation of custom datasets and diverse model architectures, effectively accelerating the iteration efficiency of modeling

research on complex visual neural datasets and providing essential methodological support for analyzing the computational mechanisms of the visual cortex.

In summary, this thesis conducts a systematic study of neural population activity modeling, encompassing both neural representation simulation and neural dynamics deconstruction. By extensive validation on multi-paradigm neural data, this thesis not only proves the significant advantages of bio-inspired deep networks over traditional deep networks but also successfully reveals the mechanisms of population representation and dynamical evolution in visual neural activity. These findings deepen the scientific understanding of functional hierarchy, temporal integration, and population activity dynamics, while providing high-performance modeling tools and a standardized framework for the analysis of high-dimension visual neural data. By exploring the intersection of deep learning and computational neuroscience, this thesis offers valuable support for the development of highly biologically plausible visual computing models.

KEY WORDS: Visual Cortex, Neural Representations, Neural Dynamics, Task-Driven Computational Models, Latent Variable Models