Unsupervised Identification of Cortical Activity Patterns in Mouse Brain for Understanding Neurological Disorders  
 **Name:** Zhiyao Zhang  
 **Major:** Computer Science  
 **Research Descriptive Keywords:** fpCNMF, Deep Learning, Brain Imaging  
 **Mentor:** Dr. Murat Yildirim   
 **Department:** Neuroscience Department, Cleveland Clinic Lerner Research Institute

## **Project Background**

Understanding **cortical activity patterns** in large-scale neural recordings is essential for **deciphering brain function** and identifying dysfunctions associated with **neurological disorders**. The brain operates as a **highly dynamic and interconnected system**, where certain activity motifs provide critical insights into cognitive processes and disease mechanisms. **Traditional methods for analyzing neural data** often fail to capture the **complex and dynamic nature** of these signals, necessitating **advanced deep learning approaches** for more effective pattern recognition1.

Recent **advancements in computational neuroscience and deep learning** have transformed how we extract meaningful information from **high-dimensional neural datasets**. One powerful technique is **Non-negative Matrix Factorization (NMF)**, widely used for dimensionality reduction and feature extraction. However, standard NMF lacks **temporal modeling** capabilities, limiting its effectiveness for analyzing time-dependent neural signals. **Convolutional NMF (CNMF)** enhances this approach by incorporating **spatial and temporal constraints**, making it particularly useful for **calcium imaging**—a critical technique for monitoring neural activity in real time2.

To push the boundaries further, this project applies **fpCNMF (factorized, point-process Convolutional Non-negative Matrix Factorization)**, an advanced deep learning framework designed to extract **neural motifs from calcium imaging data**. This method improves **signal decomposition** by leveraging a **factorized point-process framework**, enhancing our ability to link structured neural motifs with **behavioral states**. The findings of this study will contribute to **computational neuroscience and neuroinformatics** by providing a powerful framework for understanding **how neural motifs relate to behavioral and cognitive functions**, with implications for disorders such as **autism spectrum disorder (ASD), glioblastoma (GBM), and schizophrenia3**.

## **Project Overview**

This research leverages **deep learning models** to analyze **large-scale calcium imaging data** and extract meaningful **neural motifs** in transgenic mouse models expressing **calcium indicators with different temporal resolutions (GCaMP6s - slow, GCaMP6f - medium, jGCaMP8m - fast)4-6. I** will use **custom-built optical microscopes**, developed by the **Yildirim Lab**, which allow for **high-speed imaging (20–100 Hz) of large-scale cortical activity (10 mm diameter, 60 brain regions, >10 million neurons).**

To establish links between neural activity and behavior, we will synchronize neural recordings with **movement patterns, pupil dilation, and orofacial features** as mice engage in **virtual reality-based cognitive tasks**. These tasks, developed in our lab, provide an **ecologically relevant framework** for studying **adaptive behaviors**. By correlating extracted neural motifs with behavioral data, we aim to identify **activity signatures associated with cognitive and motor functions in neurological disorders**. The findings could contribute to **AI-driven diagnostic tools** capable of predicting disease-related changes in brain function7.

## **Methodology**

YildirimLAB has already developed custom-made optical microscopes and behavioral rigs to perform longitudinal brain and behavioral recordings in awake mice (**Fig. 1a-b**). These systems enable simultaneous recording and manipulation of pupil dynamics, orofacial features, locomotion, and large-scale brain activity in virtual reality environments. For this project, we will utilize three transgenic mouse lines optimized for calcium imaging at varying temporal resolutions: **GCaMP6s (slow), GCaMP6f (medium), and jGCaMP8m (fast)**4-6. Calcium imaging measures neural activity across brain regions by monitoring calcium transients normalized to compute ΔF/F values, providing a quantifiable measure of neuronal firing (**Fig. 1c-d**).

A collage of a computer lab

Description automatically generated**Figure 1**. (**A**) The custom-made optical system to perform brain recordings and manipulations. (**B**) Close lookup for awake mouse recordings during virtual reality experiments. (**C**) Segmentation of brain regions on the Allen Brain Atlas. (**D**) Example calcium imaging neural recordings from Motor (**Mos1**) and Somatosensory (**SSp\_ll1**) regions.

To ensure the robustness of our analysis, we will implement a structured computational pipeline that involves data preprocessing, normalization, dimensionality reduction, and motif extraction using the fpCNMF1 algorithm. The first step in this process is data preprocessing, which includes reducing the dimensionality of the raw imaging data, masking non-neural pixels such as vasculature, and applying normalization techniques before passing the data into the fpCNMF model. First, we apply **masking** to remove non-neural pixels and artifacts. This step excludes regions such as blood vessels and areas with low fluorescence variance, which do not contribute meaningful neural signals. The masked dataset initially consists of **300×300×6000** spatial-temporal points, which we reduce to **75×75×6000** by systematically selecting every fourth pixel in both the x and y dimensions while preserving the full time series length. This dimensionality reduction step ensures computational efficiency while retaining essential spatial information for motif detection.

The remaining neural data undergoes **normalization**, ensuring uniform scaling of fluorescence intensities across different regions. To standardize neural signal intensities and remove inactive regions, pixels with near-zero variance are set to NaN. The data is then reshaped into a structured format that preserves spatial and temporal dependencies. Histograms are generated to visualize the distribution of fluorescence changes, and sample frames are plotted to inspect data integrity and detect any artifacts. This preprocessing pipeline ensures that only high-quality data is passed into the model, improving motif extraction performance.

Next, we apply **deconvolution** to better estimate the underlying neural dynamics. Deconvolution removes the temporal smoothing effects inherent to calcium imaging, allowing for sharper isolation of neural activity patterns. This step enhances the ability of fpCNMF to detect meaningful motifs by reducing noise and improving the interpretability of extracted features. The dataset is then split into **training and testing chunks** to enable internal cross-validation, ensuring that motifs discovered in the training set generalize to withheld data.

After data normalization and splitting, we proceed with motif extraction using **fpCNMF**. This function is a streamlined adaptation of the seqNMF framework optimized for extracting spatiotemporal motifs from large-scale neural datasets. The fitting process begins with defining key hyperparameters, including the number of latent motifs, sparsity constraints, and the number of iterative updates for factorization. The optimization process iteratively updates the weight matrix (W) and activation matrix (H) to decompose the dataset into meaningful components. Multiple fits are performed to improve robustness. To ensure reproducibility, the fitting process is repeated across multiple initialization settings, and the best result is chosen based on variance explained and motif stability.

Once motifs have been extracted, their **spatial and temporal structures** are analyzed. Each motif is examined for its variance contribution, and motifs with minimal impact on overall variance are pruned. The final set of motifs is then ranked by their ability to explain variance in the dataset. A comparison with **Principal Component Analysis (PCA)** is conducted to assess the efficiency of motif extraction relative to traditional dimensionality reduction techniques2. This step validates the distinctiveness of extracted motifs and ensures that they capture meaningful neural activity patterns.

The extracted motifs are then **correlated with behavioral data**, including pupil, speed and orofacial features. Behavioral data is synchronized with neural recordings, allowing us to determine whether specific motifs correspond to distinct behavioral states. Statistical tests and regression models are applied to quantify the strength of these correlations. The final step involves cross-validating the motifs by fitting the fpCNMF model to the withheld test dataset and evaluating how well the discovered motifs generalize to unseen data. The **percent variance explained (PVE)** is computed for both training and test sets, serving as a key metric for motif quality and model performance1.

The computational pipeline concludes with extensive **visualization** of the extracted motifs and their corresponding weightings. The spatial organization of each motif is plotted alongside its temporal activation pattern. Additionally, peak activity points in the motifs are identified and displayed in relation to behavioral time series, providing insight into how different motifs align with cognitive and motor processes. A comparison is made between the discovered motifs and principal components derived from PCA, allowing for validation of the uniqueness and interpretability of the fpCNMF motifs (Fig.2).

**A diagram of a brain activity

Description automatically generated**

**Figure 2**. (**Left**) Custom-made optical system to record brain activity, locomotion, pupil and orofacial movement simultaneously. (Middle) Raw cortical brain activity snapshot across whole experiment. (Right) Identified motif activities at different time points.

By implementing this methodology, we ensure that our approach is both rigorous and interpretable, allowing us to gain deeper insights into neural activity motifs and their relationship with behavior. This study not only advances motif detection in neuroscience but also establishes a scalable framework for future applications of deep learning in neural data analysis.

## **Connection with Education and Career Goals**

This research directly aligns with my academic and professional aspirations in **computational neuroscience and artificial intelligence**, providing an unparalleled opportunity to apply **cutting-edge machine learning techniques** to the study of large-scale neural data. By working with **advanced deep learning architectures, signal processing methodologies, and high-dimensional data analysis**, I will develop the technical expertise essential for careers in **AI-driven neuroscience, neuroinformatics, and biomedical research**.

Beyond technical proficiency, this project will **deepen my understanding of how AI can be leveraged to decode complex neural activity patterns** and uncover mechanisms underlying neurological disorders. The ability to extract **meaningful motifs from large-scale neural recordings** will not only advance our understanding of brain function but also lay the groundwork for **machine learning models capable of diagnosing and predicting disease progression**. Through this study, I will strengthen my ability to **develop computational models with real-world applications in neurobiology, neuroengineering, and personalized medicine7**.

My long-term goal is to **integrate AI and neuroscience to develop transformative healthcare solutions**, particularly in **neurological disorder diagnostics and intervention strategies**. This project provides a **critical foundation** for achieving that vision by offering hands-on experience at the intersection of **machine learning, large-scale brain imaging, and behavioral neuroscience**. Furthermore, engaging in this interdisciplinary research will prepare me for **doctoral studies in computational neurobiology**, equipping me with the skills necessary to **bridge cutting-edge AI advancements with medical applications**. By merging **neuroscience, engineering, and artificial intelligence**, I aim to **push the boundaries of neuroinformatics and contribute to AI-powered medical breakthroughs that enhance patient outcomes**.

## **Time Commitment and Budget Summary**

This project will follow a **structured and efficient timeline** to ensure the systematic collection, processing, and analysis of large-scale neural data. **During the first two weeks**, I will focus on **preprocessing neural imaging data**, optimizing data quality, and refining inputs for deep learning models. **Over the following four weeks**, I will implement and fine-tune the **fpCNMF algorithm**, extracting key neural motifs and aligning them with behavioral data to uncover meaningful brain-behavior relationships. **Weeks seven and eight** will involve conducting **comparative analyses between wild-type and ASD-model mice**, applying rigorous statistical techniques to identify correlations between neural motifs and behavioral patterns. The **final two weeks** will be dedicated to synthesizing results, documenting key findings, and preparing a **comprehensive research presentation** to communicate the impact of this work effectively.

To support this research, I am requesting **$4,000 in stipend support**. These funds will be instrumental in facilitating **data processing and computational analyses**, ensuring the necessary resources for **machine learning model implementation and behavioral data synchronization**. Securing this financial support will allow me to focus on generating **high-impact insights** from the extracted neural motifs, contributing to **a deeper understanding of cortical function and its alterations in neurological disorders**. Beyond this specific project, the methodologies and insights developed will lay the foundation for **AI-driven brain activity analysis**, with potential applications in **neuromedical diagnostics and therapeutic innovations**. This opportunity is not only a critical step toward my personal career goals in **computational neuroscience and AI-powered healthcare** but also a **transformative contribution to advancing neuroinformatics and translational neuroscience research**.

# Reference

1 MacDowell, C. J. & Buschman, T. J. Low-Dimensional Spatiotemporal Dynamics Underlie Cortex-wide Neural Activity. *Curr Biol* **30**, 2665-2680 e2668 (2020). <https://doi.org:10.1016/j.cub.2020.04.090>

2 Mackevicius, E. L. *et al.* Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *Elife* **8** (2019). <https://doi.org:10.7554/eLife.38471>

3 Degleris, A. & Gillis, N. A Provably Correct and Robust Algorithm for Convolutive Nonnegative Matrix Factorization. *Ieee T Signal Proces* **68**, 2499-2512 (2020). <https://doi.org:10.1109/Tsp.2020.2984163>

4 Zhang, Y. *et al.* Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature* **615**, 884-891 (2023). <https://doi.org:10.1038/s41586-023-05828-9>

5 Dana, H. *et al.* High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat Methods* **16**, 649-657 (2019). <https://doi.org:10.1038/s41592-019-0435-6>

6 Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295-300 (2013). <https://doi.org:10.1038/nature12354>

7 Levine, J. A factorization of the Conway polynomial. *Comment Math Helv* **74**, 27-52 (1999). <https://doi.org:DOI> 10.1007/s000140050075