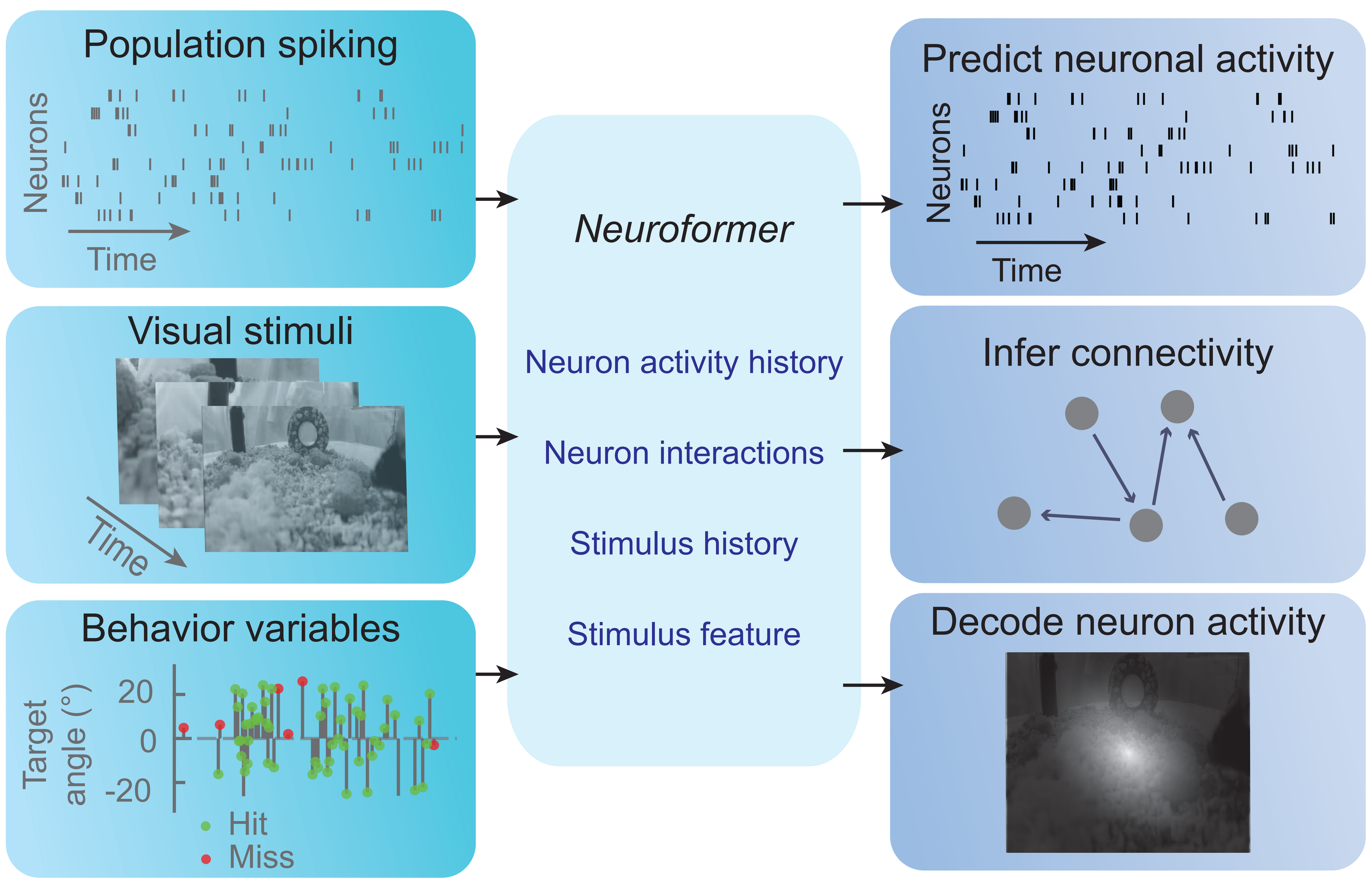
Neuroformer: A Framework for Multimodal Neural Data Analysis

Antonis Antoniades Yiyi Yu, Spencer Smith
UCSB NLP Group, CS, ECE
antonis@ucsb.edu

Introduction

Motivation

- Systems neuroscience experiments are growing in complexity
- Large datasets are acquired with multiple modalities, including visual, neural, reward, pose, eye-movement, environment and more
- No existing tools to unify training and analysis at this scale



Neuroformer

Framework

- Re-frame Neuron IDs as token representations
- Align multiple modalities using contrastive learning
- Model Neural decoding as a sequential autoregressive process
- Optimize using MLE

Architecture

- Iteratively fuse the "Neural State" with all other modalities using a cross-attention transformer that unrolls recurrently in space
- Decode using a causal transformer decoder with two projection outputs, one for temporal prediction, and one for classification

Optimization

- Alignment (contrastive objective)

$$s(F,I) = g_f(F_{p,c})^T g_i(I_c) \qquad s(I,F) = g_i(I_c)^T g_f(F_{p,c})$$

$$(1) \quad p_m^{fi} = \frac{\exp(s(f,i_m)/\tau)}{\sum_{m=1}^{M}\exp(s(f,i_m)/\tau)} \qquad p_m^{if} = \frac{\exp(s(i_m,f)/\tau)}{\sum_{m=1}^{M}\exp(s(i_m,f)/\tau)} \quad (2)$$

$$L_{vnc} = \frac{1}{2}\mathbb{E}_{(F,I)\in d}[H(\boldsymbol{y}^{fi}(F), \boldsymbol{p}^{fi}(F)) + \boldsymbol{y}^{if}(I), \boldsymbol{p}^{if}(I))] \quad (3)$$

- Spatio-temporal Decoding (MLE)

$$(4) \quad L_{ce(I)} = \frac{1}{2}\mathbb{E}_{(I)\sim d}\, H(\boldsymbol{y}_I, \boldsymbol{p}_I) \qquad L_{ce(dt)} = \frac{1}{2}\mathbb{E}_{(dt)\sim d}\, H(\boldsymbol{y}_{dt}, \boldsymbol{p}_{dt}) \quad (5)$$

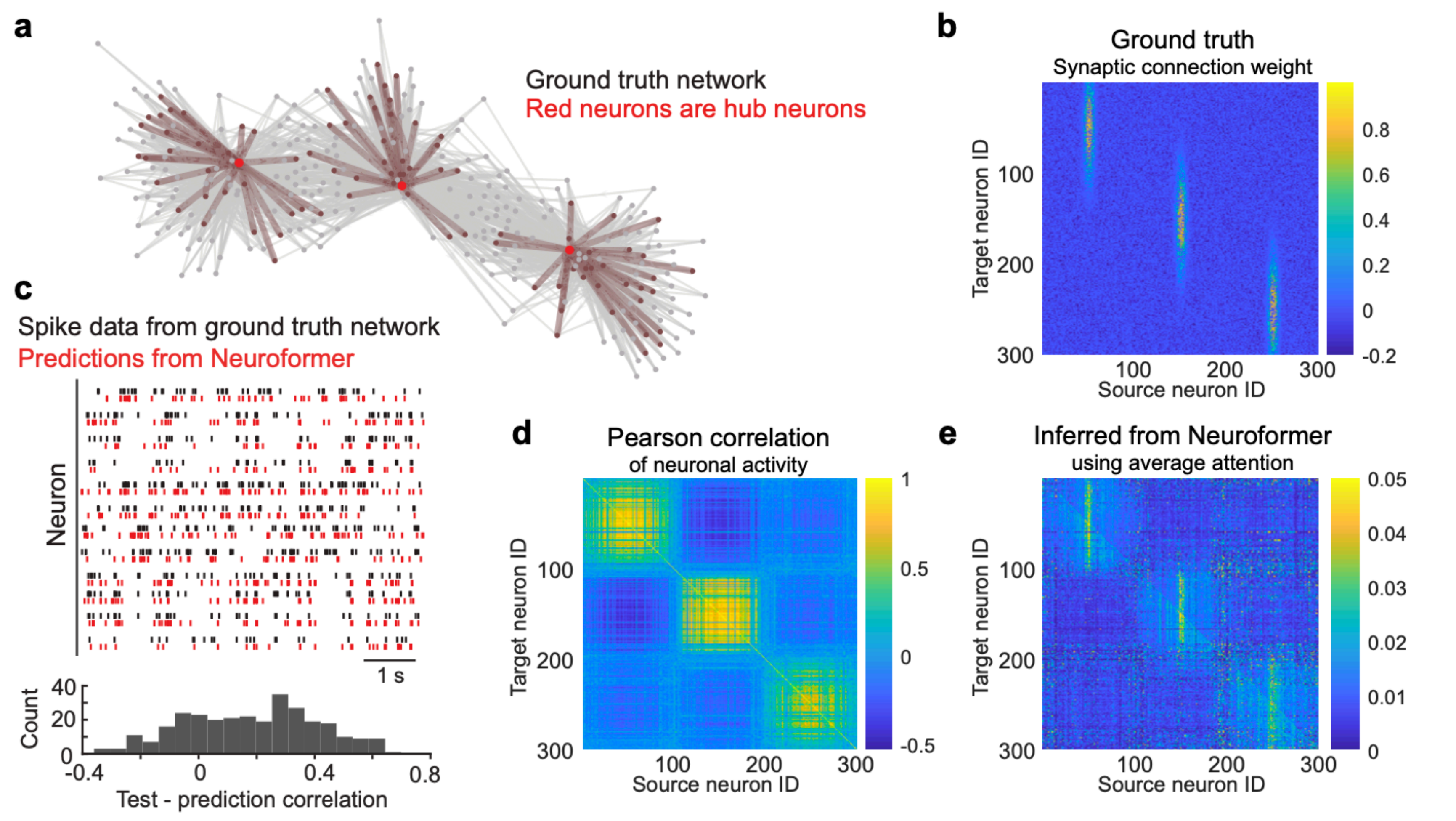
- Weighted sum

$$L = (\gamma)L_{vnc} + (\mu)L_{ce(I)} + (1-\gamma-\mu)L_{ce(dt)} \quad (6)$$

Experiments

Uncovering Ground-truth Connectivity

- Simulated dataset of Hub Neurons ("Neurons that fire together, wire together")
- Attention can uncover ground-truth connectivity (20% variability, compared to 13% for Pearson correlation)



Multi-region Mouse Cortex Recordings

- Wide-field-of-view 2-photon imaging of V1 + AL brain areas
- Mouse watching a naturalistic video
- Neuroformer can generate high-precision simulations of ground-truth trials over 32 seconds
- Cross-Attention between Neurons and Video reveals salient features