Modeling Neural Population Activity with Spatiotemporal Transformers

Trung Le, Eli Shlizerman

University of Washington, Seattle

NeuroAI

UNIVERSITY of WASHINGTON

INTRODUCTION

- Understanding neural code helps relate neural activity and behaviors.
- Complex behaviors are only interpretable at the level of neural population.
- Neural spiking activities are stochastic, firing rates vary from trials to trials even under the same behavior.

>> requires methods modeling population activity on a single-trial basis to uncover firing mechanism behind noisy spike trains.

- Neural Data Transformers [1]: successful at capturing neural dynamics by characterizing relationship between timesteps of population responses, but ignores rich covariation between neurons.

>> We propose SpatioTemporal Neural Data Transformer (STNDT) to explicitly learn both the spatial coordination between neurons and the temporal progression of population activity.

>> STNDT achieves state-of-the-art performance on modeling neural dynamics.

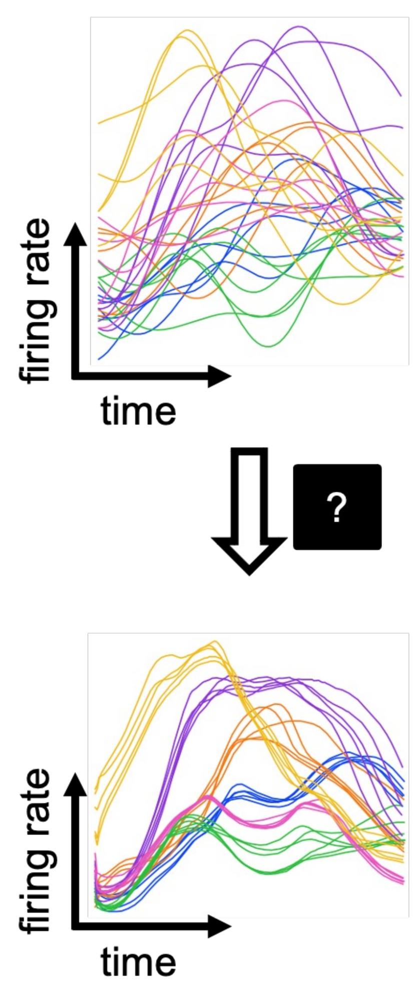


Figure 1: Denoising objective [2]

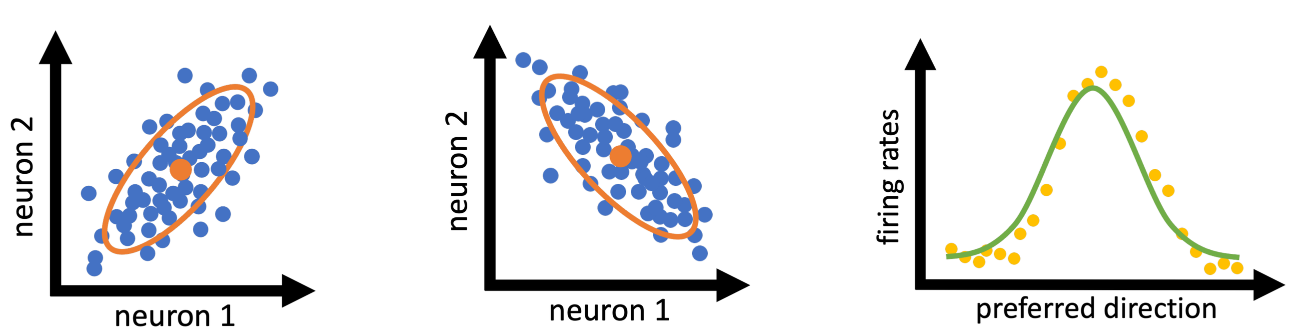
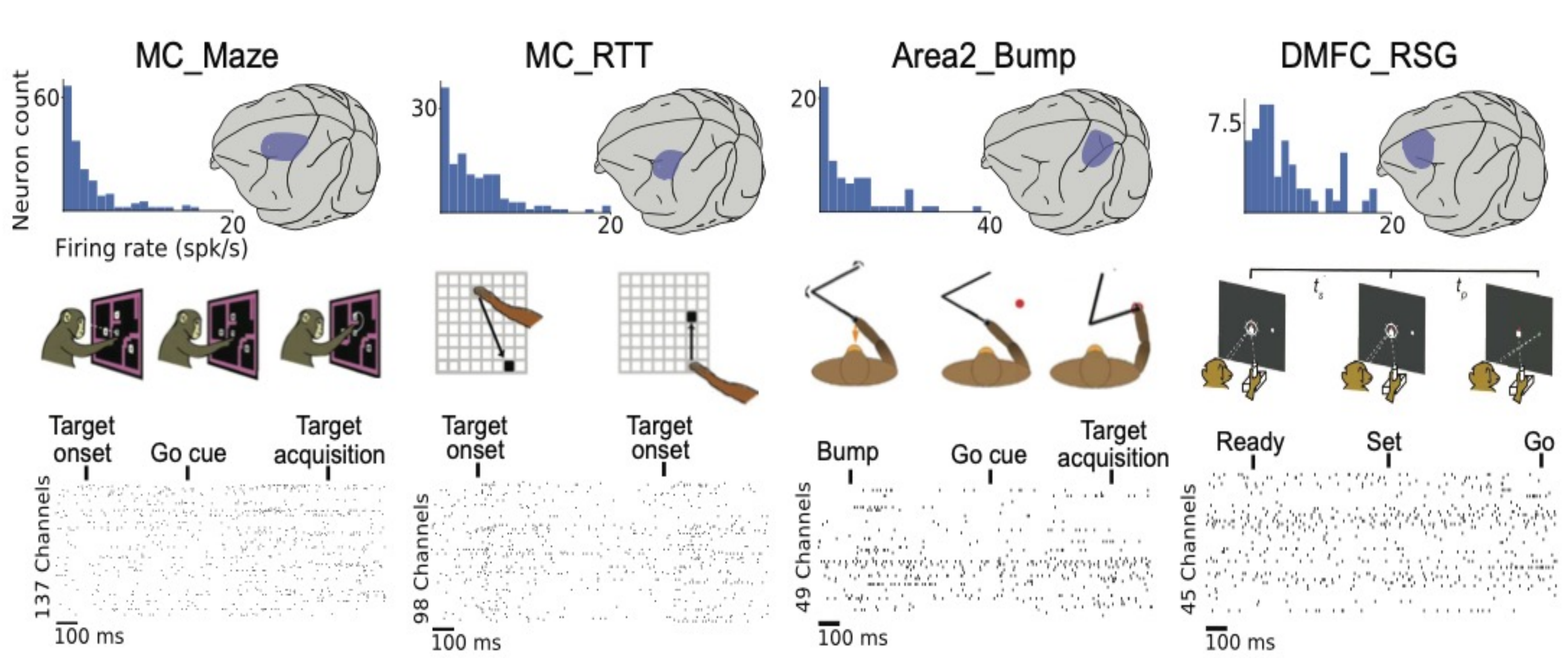


Figure 2: Correlation of firing rates among neurons contains salient information to infer population dynamics.

METHODS

Datasets: Neural Latents Benchmark [2]: electrophysiological recordings from rhesus macaques covering:

- autonomous and non-autonomous neural dynamics.
- variety of behavior tasks (target reaching and time reproduction).
- variety of brain regions (M1, PMd, S2, DMFC).



- Motor cortex (M1) & dorsal premotor cortex (PMd)
- Autonomous dynamics

- Motor cortex
- Naturalistic, unstereotyped behavior

- Somatosensory area 2
- Input-driven dynamics

- Dorsomedial frontal cortex
- No moment-by-moment behavioral correlate

Figure 3: Four datasets covering rich neural dynamics under various scenarios [2].

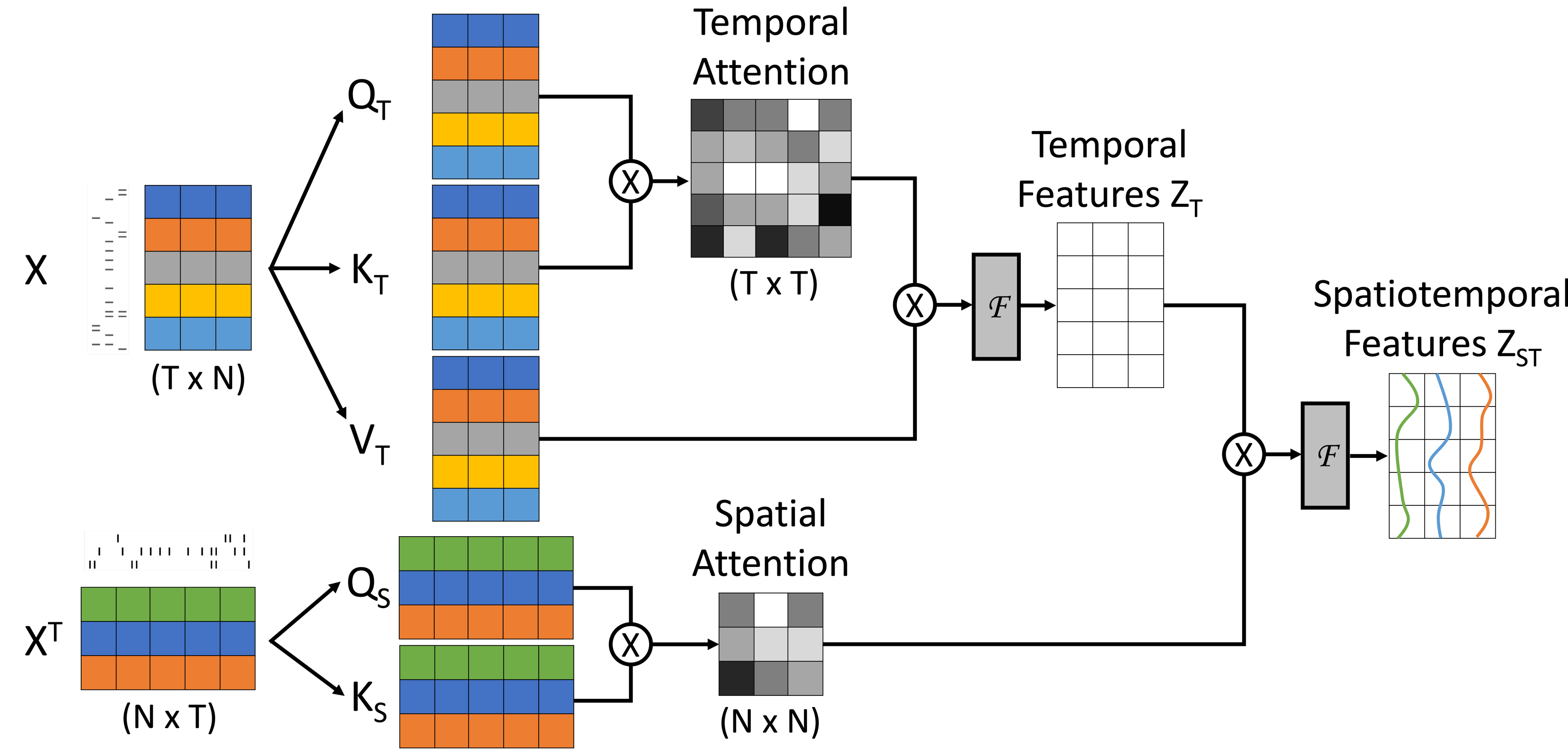


Figure 4: STNDT architecture

Architecture:

- Two separate self-attention modules for temporal and spatial dimensions:

$$Z_{\mathcal{T}} = \text{Attention}(Q_{\mathcal{T}}, K_{\mathcal{T}}, V_{\mathcal{T}}) = \mathcal{F}\left(\text{softmax}\left(\frac{Q_{\mathcal{T}}K_{\mathcal{T}}^{\top}}{\sqrt{N}}\right)V_{\mathcal{T}}\right)$$

$$A_{\mathcal{S}} = \text{softmax}\left(\frac{Q_{\mathcal{S}}K_{\mathcal{S}}^{\top}}{\sqrt{T}}\right)$$

$$Z_{\mathcal{ST}} = \mathcal{F}(A_{\mathcal{S}}Z_{\mathcal{T}}^{\top})$$

- Mask language modeling loss:

$$\mathcal{L}_{mask} = \sum_{i=1}^{N}\sum_{j=1}^{T}\exp(\tilde{z}_{ij}) - \tilde{x}_{ij}\tilde{z}_{ij}$$

- Contrastive loss:

$$\mathcal{L}_{contrastive} = \sum_{ij} l_{ij} = \sum_{ij} -\log\frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N}\mathbf{1}_{k\neq i}\exp(\text{sim}(z_i, z_k)/\tau)}$$

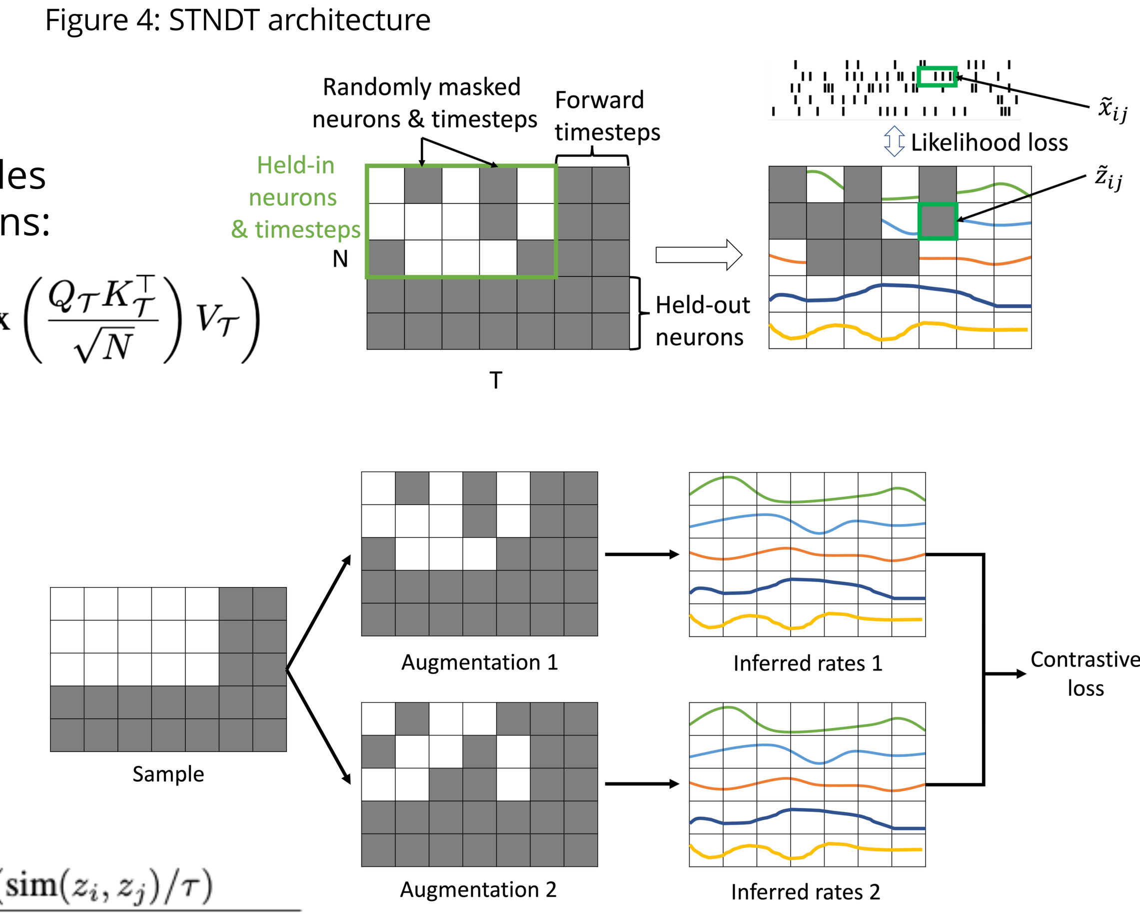


Figure 5: Mask loss and contrastive loss

RESULTS

- STNDT achieves state-of-the-art prediction of firing activities for unobserved neurons and timesteps.
- STNDT reveals stereotyped features of neural activity.
- STNDT helps accurate decoding of behaviors.

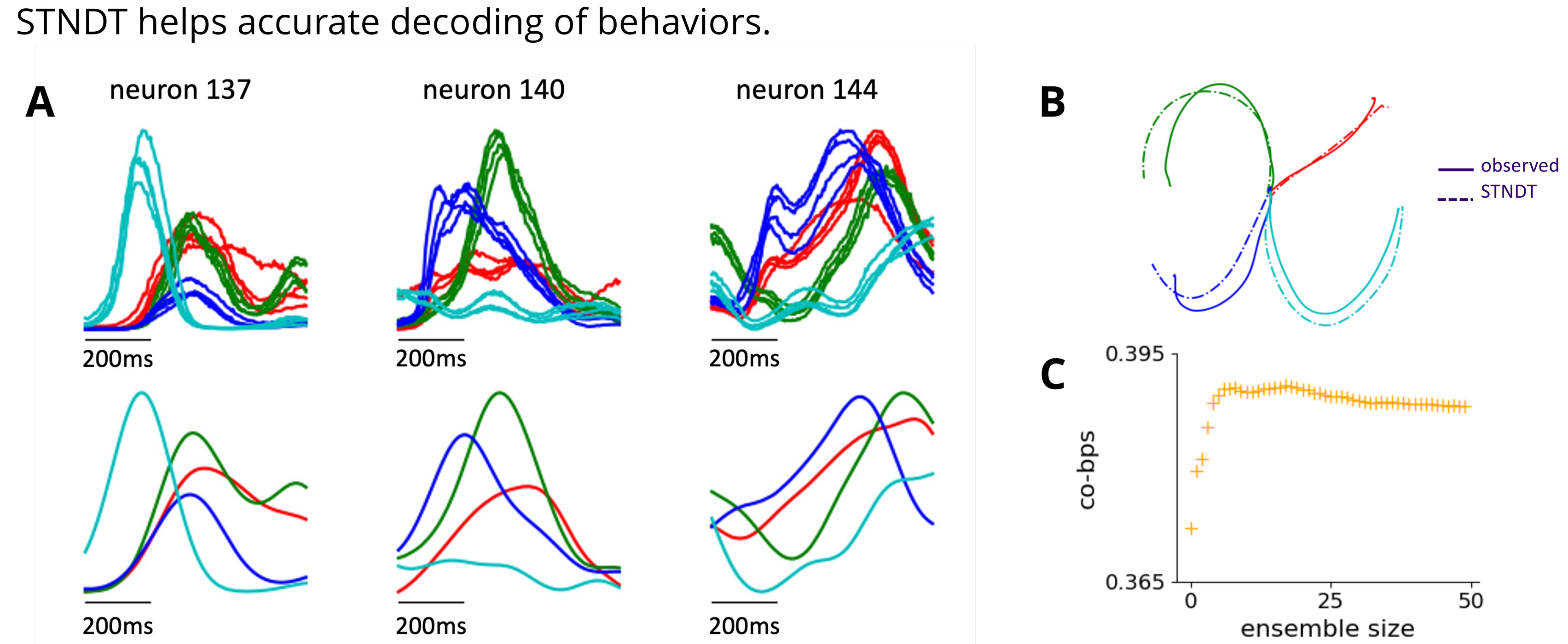


Figure 6: A) Inferred firing rates recover PSTH structures. B) Behavior decoded from STNDT-inferred rates match the ground truth behavior. C) Model performance improves as multiple models are ensembled.

Methods	MC_Maze co-bps↑	MC_Maze vel R^2↑	MC_Maze psth R^2↑	MC_Maze fp-bps↑	MC_RTT co-bps↑	MC_RTT vel R^2↑	MC_RTT fp-bps↑
GPFA	0.1872	0.6399	0.5150	–	0.1548	0.5339	–
Smoothing	0.2109	0.6238	0.1853	–	0.1468	0.4142	–
SLDS	0.2249	0.7947	0.5330	1.1579	0.1649	0.5206	0.0620
MINT	0.3304	**0.9121**	**0.7496**	0.2076	0.1676	0.5953	0.1012
AutoLFADS	0.3364	0.9097	0.6360	0.2349	0.1868	0.6167	0.1213
iLQR-VAE	0.3559	0.8840	0.6062	0.1480	–	–	–
AESMTE1 (single)	0.3599	0.9105	0.6641	0.2470	0.1927	**0.6627**	0.1229
AESMTE3 (ensemble)	0.3676	0.9114	0.6683	0.2589	0.2053	0.6334	**0.1344**
STNDT single (ours)	0.3632	0.8997	0.6749	0.2486	0.1865	0.5988	0.0964
STNDT ensemble (ours)	**0.3835**	0.9103	0.6704	**0.2682**	**0.2095**	0.6270	0.1244

Methods	Area2_Bump co-bps↑	Area2_Bump vel R^2↑	Area2_Bump psth R^2↑	Area2_Bump fp-bps↑	DMFC_RSG co-bps↑	DMFC_RSG tp-corr↓	DMFC_RSG psth R^2↑	DMFC_RSG fp-bps↑
GPFA	0.1680	0.5975	0.5289	–	0.1176	−0.3763	0.2142	–
Smoothing	0.1544	0.5736	0.2084	–	0.1202	−0.5139	0.2993	–
SLDS	0.1960	0.7385	0.5740	0.0242	0.1243	−0.5412	0.3372	−0.0418
MINT	0.2735	0.8877	**0.9135**	0.1483	0.1821	−0.6929	**0.7013**	0.1650
AutoLFADS	0.2569	0.8492	0.6318	0.1505	0.1829	**−0.8248**	0.6359	0.1844
iLQR-VAE	–	–	–	–	–	–	–	–
AESMTE1 (single)	0.2801	0.8675	0.6367	0.1523	0.1733	−0.6189	0.5267	0.1511
AESMTE3 (ensemble)	0.2860	**0.8999**	0.7109	**0.1603**	0.1886	−0.7601	0.6064	0.1828
STNDT single (ours)	0.2690	0.8593	0.7191	0.1353	0.1859	−0.5205	0.6051	0.1601
STNDT ensemble (ours)	**0.2896**	0.8912	0.7358	0.1464	**0.1940**	−0.4857	0.6452	**0.1910**

Table: STNDT achieves state-of-the-art performance on modeling neural population dynamics, most notably on the primary metric co-bps.

- Ensembling multiple models improves STNDT performance.
- Incorporating contrastive loss improves STNDT performance.
- STNDT identifies consistent subsets of important neurons whose activities contain salient information representing the response of the population to behavior task.
- Predictive performance decreases slightly when random neurons are removed but deteriorates significantly when important neurons are removed.

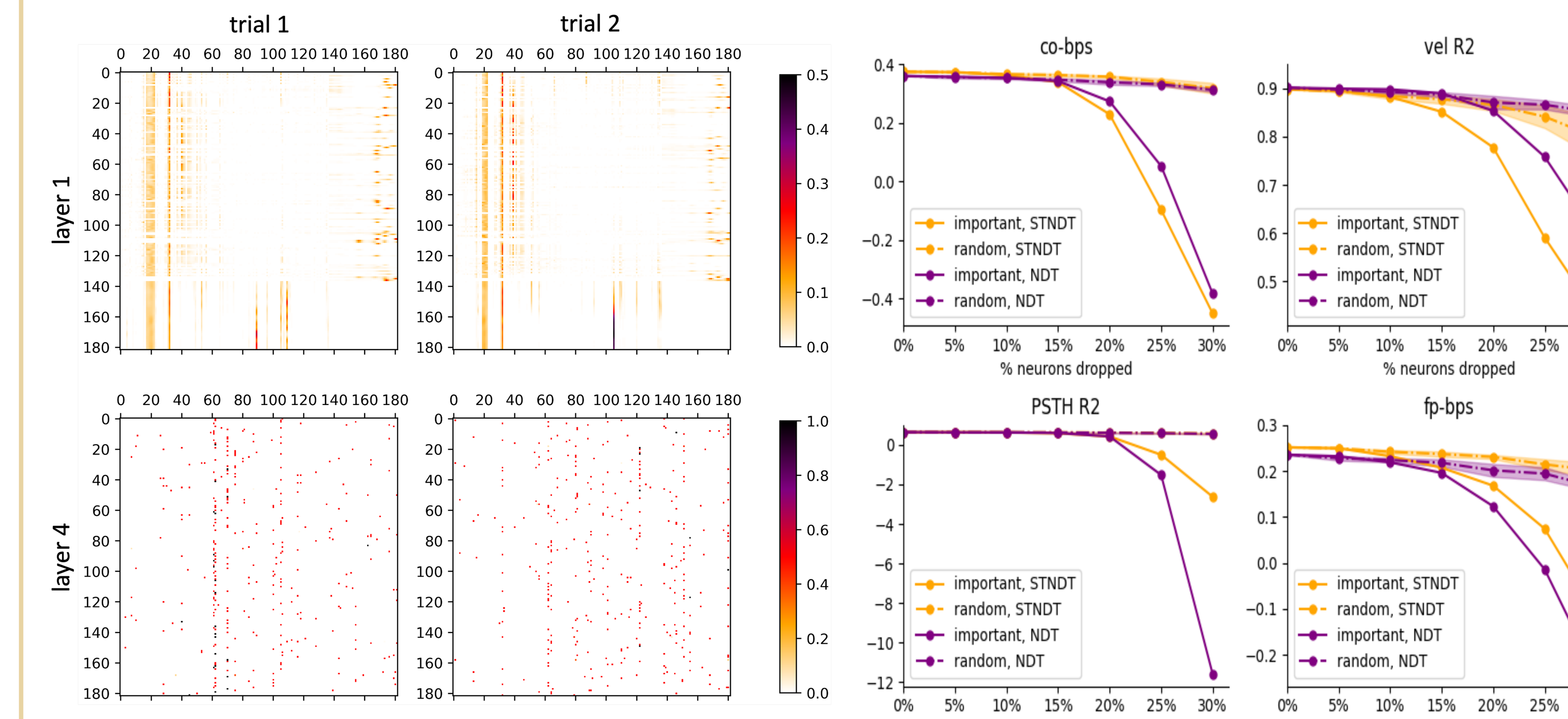


Figure 7: Left: Spatial attention weights reveals important neurons. Right: Performance decreases significantly when these important neurons are dropped from the population.

REFERENCES

[1] Ye, Joel, and Chethan Pandarinath. "Representation learning for neural population activity with Neural Data Transformers." *Neurons, Behavior, Data analysis, and Theory* (2021).

[2] Pei, F., et al. "Neural Latents Benchmark'21: Evaluating latent variable models of neural population activity." *Advances in Neural Information Processing Systems (NeurIPS), Track on Datasets and Benchmarks* 34 (2021).

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation grant OAC-2117997.