Neuronal Identity as an Organizational Basis for Analyzing Neural Population Dynamics

Seungjae Han[1*] **Joshua Yedam You**[1*] **Soi Kim**[1*] **Minho Eom**[1] **Jinho Park**[1]
Seunghee Park[1] **Byeongwook Lee**[2,3,4†] **Young-Gyu Yoon**[1,5,6†]

[1]School of Electrical Engineering, KAIST, Daejeon, Republic of Korea
[2]Department of Brain and Cognitive Sciences, KAIST, Daejeon, Republic of Korea
[3]Graduate School of Data Science, KAIST, Daejeon, Republic of Korea
[4]Department of AX, KAIST, Daejeon, Republic of Korea
[5]Department of Semiconductor System Engineering, KAIST, Daejeon, Republic of Korea
[6]KAIST Institute for Human Augmentation Convergence, KAIST, Daejeon, Republic of Korea
[*]Co-first authors [†]Co-corresponding authors: {bwlee, ygyoon}@kaist.ac.kr

Abstract

Individual neurons exhibit diverse molecular, morphological, and electrophysiological properties that shape their distinct roles in neural circuits. While these multimodal features are largely shared across individuals, they remain an underutilized source of information in analyzing neural population dynamics. We posit that these shared biological profiles provide a common reference for relating neural dynamics across subjects and recording sessions. Our framework leverages this structure by routing neurons into latent groups based on their biological profiles and analyzing population dynamics through group-level representations. Because these assignments are anchored in biological properties that are consistent across subjects, the learned organization supports transfer to unseen animals without retraining or neuron-wise correspondence. We evaluate our framework on whole-brain calcium imaging in larval zebrafish and large-scale electrophysiology in the mouse visual cortex. Across both species and recording modalities, the framework delivers state-of-the-art performance in both neural decoding and manifold analysis: it achieves superior decoding of behavior and stimulus labels while better preserving the temporal consistency and local geometric structure of latent trajectories on unseen data. Our experiments demonstrate that not only does providing biological features improve baseline performance, but our proposed framework achieves even stronger results by using these properties to explicitly structure the population rather than treating them as auxiliary inputs. Together, these results suggest that organizing neural populations by their multimodal biological features offers a promising framework for generalizing population dynamics across subjects and recording sessions.

1 Introduction

Recent advances in large-scale neural recording and analysis have opened new opportunities for studying how neural populations give rise to perception, action, and cognition, and have spurred a growing body of deep learning approaches for analyzing population dynamics [1–8]. Yet existing approaches still struggle with generalization across individuals. Every recording captures a different set of neurons, and activity patterns shift with recording conditions, internal states, and individual

Project page and code: https://nicalab.github.io/identity-group-routing

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

variability [9, 10]. Because the recorded populations themselves differ across animals, there is no guaranteed correspondence between the neural populations, presenting a significant challenge for the direct transfer of learned representations to unseen individuals.

The multimodal biological features of individual neurons, or neuronal identity, represent a complementary source of information that has remained largely underutilized in population dynamics analysis. Neurons can be characterized along multiple biological axes, including transcriptomic, regional, and electrophysiological features, which techniques like Patch-seq integrate at single-cell resolution to link molecular identity with distinct functional roles in neural circuits [11]. For example, in the mouse visual cortex, Pvalb, Sst, and Vip interneurons play specialized, non-redundant roles in gain control, novelty detection, and state-dependent modulation [12], illustrating how neuronal identity shapes individual contributions to circuit computation [13]. Large-scale transcriptomic studies further demonstrate that neurons close in molecular profile participate in related circuit computations [14–17]. Importantly, these features are largely shared across individuals of the same species [15], suggesting they may offer a more stable basis for analyzing population dynamics across individuals.

We propose to exploit this structure through an identity-guided, permutation-invariant framework for analyzing neural population dynamics. Neurons are routed into latent groups based on their biological profiles, including transcriptomic, regional, and electrophysiological features, and population dynamics are analyzed at the group level through shared parameters conditioned on these features. Because neuron ordering in a neural recording dataset is arbitrary and varies across recordings, the framework should also be permutation-invariant. Standard permutation-invariant architectures such as DeepSets [18] achieve this by treating all neurons as interchangeable units through shared element-wise transformations followed by symmetric global aggregation. Our framework instead maintains permutation invariance by encoding each neuron's identity and activity through shared neuron-wise networks, then routing these representations into latent groups via assignment-weighted aggregation. Because these group assignments are determined by biological properties that are stable across individuals, the learned organization can be applied to unseen animals without retraining or neuron-wise correspondence.

This design contrasts with existing approaches for analyzing neural population dynamics. Recent large-scale models, such as POYO [9] and NEDS [19], achieve cross-session transfer by learning session-specific embeddings and relying on statistical regularities in observed dynamical patterns to bridge recordings. Neural manifold analysis methods such as CEBRA [3] and MARBLE [4] learn low-dimensional embeddings of population dynamics that reveal how neural circuits represent and transform information, providing an approach for relating population activity to behavior and perception. However, because these approaches rely primarily on activity patterns, they often struggle to generalize to new animals or sessions. Our framework instead grounds each neuron in a common biological coordinate, thereby inducing a shared group structure that supports transfer to held-out animals and sessions without any neuron-wise correspondence.

We evaluate our framework on two large-scale datasets spanning distinct species, recording modalities, and available biological features: a whole-brain larval zebrafish calcium imaging dataset that jointly acquires neural activity and single-cell transcriptomics [20], and a large-scale mouse visual cortex Neuropixels electrophysiology dataset [21]. Across both settings, our framework achieves state-of-the-art cross-session behavior and stimulus decoding when applied directly to held-out recordings without retraining or neuron-wise correspondence. While supplying baselines with these same neuronal identity features improves their performance, our identity-guided routing framework consistently outperforms all baselines. These results demonstrate both the inherent value of multimodal biological features of individual neurons and the critical importance of explicitly using them to organize population dynamics rather than just treating them as auxiliary inputs. Our contributions are as follows:

- We introduce, to our knowledge, the first neural population analysis framework that leverages multimodal biological features of individual neurons to support direct transfer across animals and recording sessions without retraining or neuron-wise correspondence.
- We propose an identity-guided routing method that uses multimodal biological features as the organizing basis to assign neurons to latent groups, thereby analyzing population dynamics at the group level through shared parameters.

- Across whole-brain zebrafish calcium imaging recordings and mouse visual cortex electrophysiology, our framework achieves state-of-the-art neural decoding of behavior and stimulus labels and enables robust neural manifold analysis by preserving the temporal and local geometric structure of latent trajectories in held-out animals.

2 Method

We seek a framework for analyzing neural population dynamics that transfers across sessions without neuron-wise correspondence. The core challenge is that neurons cannot be matched across animals by recording index because neuron order in a neural recording dataset is arbitrary and changes across animals and sessions. Neuronal identity provides a shared description of each neuron across individuals. Neurons with high overlap in neuronal identity tend to support similar circuit roles even across individuals, and can serve as a stable basis for organizing the population. This leads to the following two properties:

- **Permutation invariance.** The model output should be independent of how recorded neurons are ordered.
- **Identity-based organization.** Neurons with shared properties should be considered together, reflecting the group structure observed in biological neural circuits.

Permutation-invariant models treat the neural population as an unordered set, requiring aggregation across neurons in a way that does not depend on their ordering [22, 18]. A global sum or mean over all neuron representations achieves this, but treats all neurons identically regardless of their neuronal identity, discarding the organizational structure that identity encodes. This motivates an aggregation organized by neuronal identity, where neurons with overlapping identity profiles are pooled into a shared group representation, contributing to population dynamics through a common functional organization. We therefore analyze population dynamics through two components:

1. **Identity-guided routing.** Each neuron is assigned to a latent group on the basis of its neuronal identity features.
2. **Group-level dynamics analysis.** Neural activity is aggregated into latent group representations, which are then processed to obtain the final population state.

Because group assignments are determined by neuronal identity, which is shared across individuals, the learned organizational structure can be applied across subjects and sessions without retraining or neuron-wise correspondence. Section 2.1 specifies each stage in detail.

2.1 Framework

We represent the neural population activity at a given time point t as a collection of identity–activity pairs

$$\mathcal{S} = \{(\mathbf{f}_i, \mathbf{x}_i)\}_{i=1}^{N},$$

where N denotes the number of recorded neurons in the corresponding session. For the i-th neuron, $\mathbf{f}_i \in \mathbb{R}^F$ is a neuronal identity vector encoding neuron-specific information, and $\mathbf{x}_i \in \mathbb{R}^W$ is a segment of its neural activity centered at time t. The goal is to learn a mapping

$$\mathbf{z} = \mathcal{F}(\mathcal{S}) \in \mathbb{R}^D, \tag{1}$$

that produces a population state representation $\mathbf{z}$ transferable across sessions and subjects without neuron-wise correspondence. Since neuron ordering is arbitrary and differs across recordings, $\mathcal{F}$ must satisfy permutation invariance: for every permutation π of $\{1, \ldots, N\}$,

$$\mathcal{F}\big(\{(\mathbf{f}_{\pi(i)}, \mathbf{x}_{\pi(i)})\}_{i=1}^{N}\big) \;=\; \mathcal{F}\big(\{(\mathbf{f}_i, \mathbf{x}_i)\}_{i=1}^{N}\big). \tag{2}$$

We now formalize each stage of $\mathcal{F}$ in turn and establish that it satisfies permutation invariance.

Identity-Guided Routing. The first stage maps each neuron's identity features into a soft group membership vector. Neuronal identity comprises M modalities $\mathbf{f}_i = (\mathbf{f}_i^{(1)}, \ldots, \mathbf{f}_i^{(M)})$ with $\mathbf{f}_i^{(m)} \in \mathbb{R}^{F_m}$, which are heterogeneous in type and measurement scale. Direct concatenation of raw modality features would allow modalities with larger magnitudes to dominate downstream computation irrespective of biological relevance. Each modality is encoded by a dedicated MLP $\phi_m : \mathbb{R}^{F_m} \to \mathbb{R}^{d_m}$, and the resulting embeddings are layer-normalized and concatenated before being projected by a shared MLP $\psi : \mathbb{R}^{\sum_m d_m} \to \mathbb{R}^{d_s}$:

$$\mathbf{s}_i \;=\; \psi\Big(\big[\mathrm{LN}\big(\phi_1(\mathbf{f}_i^{(1)})\big);\; \cdots\;;\; \mathrm{LN}\big(\phi_M(\mathbf{f}_i^{(M)})\big)\big]\Big) \;\in\; \mathbb{R}^{d_s}. \tag{3}$$

Given the identity embedding $\mathbf{s}_i$, neuron i is assigned a soft membership vector $\mathbf{a}_i \in \Delta^{K-1}$ over K latent groups, allowing each neuron to partially contribute to multiple groups in proportion to its similarity in the learned identity embedding space:

$$\mathbf{a}_i = \sigma\left(\frac{\mathbf{W}_{\text{route}}\mathbf{s}_i}{\tau}\right), \qquad \mathbf{W}_{\text{route}} \in \mathbb{R}^{K\times d_s}, \qquad \tau > 0, \tag{4}$$

where $\sigma : \mathbb{R}^K \to \Delta^{K-1}$ denotes the softmax operator.

Group-level dynamics analysis. Using the identity-guided routing weights defined in Eq. (4), we aggregate neuron-level activity embeddings into latent group-level representations. Each neuron's activity window $\mathbf{x}_i \in \mathbb{R}^W$ is first encoded by a shared dynamics encoder $\rho : \mathbb{R}^W \to \mathbb{R}^{d_r}$, which is applied independently to every neuron:

$$\mathbf{v}_i = \rho(\mathbf{x}_i) \in \mathbb{R}^{d_r}. \tag{5}$$

Because the same encoder ρ is used for all neurons, the activity representation does not depend on the arbitrary neuron index. For each latent group $k \in \{1, \ldots, K\}$, we form a group-level activity representation by taking a routing-weighted aggregate of the neuron-level activity embeddings:

$$\mathbf{c}_k = \frac{1}{\sqrt{\max(m_k, \epsilon)}} \sum_{i=1}^{N} a_{ik}\mathbf{v}_i, \qquad m_k = \sum_{i=1}^{N} a_{ik}. \tag{6}$$

Here, a_{ik} denotes the contribution of neuron i to latent group k, while the normalization by $\sqrt{\max(m_k, \epsilon)}$ stabilizes the scale of group representations across groups, with $\epsilon = 10^{-6}$ included for numerical stability under sparse routing. Stacking the K group representations yields the group matrix $\mathbf{C} \in \mathbb{R}^{K\times d_r}$. To model interactions across latent groups, $\mathbf{C}$ is processed by a group interaction module $\Gamma : \mathbb{R}^{K\times d_r} \to \mathbb{R}^D$, operating along the group axis:

$$\mathbf{z} = \Gamma(\mathbf{C}) \in \mathbb{R}^D. \tag{7}$$

The resulting population state $\mathbf{z}$ can be used for behavioral decoding, manifold learning, or other downstream objectives, without modifying the upstream identity-guided routing and group-level aggregation framework.

The construction is permutation-invariant by design. Since the identity encoder, dynamics encoder, and routing map are shared across neurons, any permutation of the input neurons only permutes the corresponding neuron-level embeddings and membership weights. The group representations are then formed by summing over the neuron index, so they are unchanged by such a permutation. Consequently, the group matrix $\mathbf{C}$ and the final population state $\mathbf{z} = \Gamma(\mathbf{C})$ satisfy the invariance requirement in Eq. (2). A formal proof of the permutation invariance is provided in Appendix A.2.

3 Experiments

We evaluate on two datasets spanning different species, recording modalities, and degrees of richness of available neuronal identity information (Section 3.1). We compare against three decoding baselines and four manifold methods (Section 3.2), and assess performance under direct transfer to held-out animals and sessions without retraining using decoding and geometric metrics (Section 3.3).

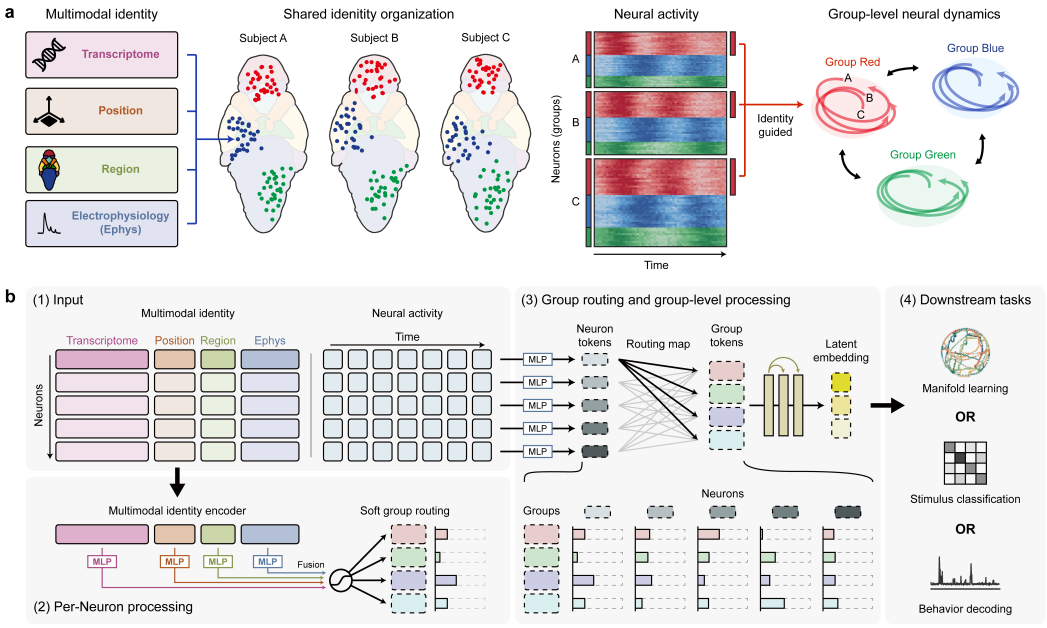

Figure 1: Overview of identity-guided neural population analysis. (a) Conceptual overview. Multimodal biological features organize neurons into corresponding groups across sessions, even when exact individual neurons cannot be matched across subjects. Population dynamics are described at the level of these identity-defined groups and their interactions. (b) Identity-guided routing framework. Identity-related features are encoded to softly route neuron-level activity tokens into latent group tokens, which are processed into a population embedding. Manifold learning, stimulus classification, and behavior decoding are evaluated as separate downstream objectives.

3.1 Datasets and Experimental Setup

Whole-brain Neuronal Activity and RNA Profiling (WARP). The WARP dataset [20] provides whole-brain calcium imaging and single-cell transcriptomics from three larval zebrafish at 6 days post-fertilization, each with approximately 77,000 neurons registered to the mapzebrain atlas [23]. Recordings span approximately 70 minutes at 2 Hz and include structured visual stimuli such as grating in four directions, bright and dark flashes, and bilateral looming. Stimulus labels are derived from the visual stimulus conditions and used as targets for stimulus classification, whereas behavioral labels are derived from fictive swimming signals recorded through extracellular motor nerve measurements and used as targets for behavior decoding (Appendix A.1, Figure A3).

To construct multimodal biological features, we use each neuron's transcriptomic expression profile, three-dimensional position, anatomical region, and electrophysiological features motivated by met-type profiling [11]. Because jointly acquired gene expression and neural activity data are difficult to obtain in practice, we additionally evaluate our framework using independently acquired, public gene expression data from the mapzebrain atlas. This atlas provides hybridization chain reaction (HCR) in situ RNA profiles [23, 24], which we assign to each neuron based on its atlas coordinates. Feature construction details are provided in Appendix A.3.

Allen Visual Coding Neuropixels. The Allen Visual Coding Neuropixels dataset [21] provides large-scale extracellular electrophysiology recordings from the mouse visual cortex. We use both available datasets: the Brain Observatory dataset with 26 sessions and the Functional Connectivity dataset with 25 sessions, comprising over 80,000 units in total. Each session uses six Neuropixels probes to record spiking activity, with unit locations registered to the Allen Common Coordinate Framework [25]. Recordings include high-frequency spiking activity over varied durations and cover structured visual stimuli such as drifting gratings, static images, and natural scenes, grouped into five functional classes per dataset. These stimulus categories are used as targets for stimulus classification, and behavioral data consist of running speed, which serves as the target for behavior decoding

(Appendix A.1, Figure A3). This dataset allows us to test whether our framework improves held-out transfer performance even when molecular profiles are unavailable, utilizing only the remaining neuronal identity features described in Appendix A.3.

3.2 Baselines

First, we evaluate neural decoding performance by comparing our method to three decoding baselines: POYO [9], NEDS [19], and DeepSets [18]. POYO uses cross-attention over tokenized neural events, relying on learned session and unit embeddings to aggregate activity across recordings. NEDS uses a transformer with a multi-task masking strategy over neural activity and behavior. Both methods operate across sessions without requiring neuron-wise correspondence, but rely on fine-tuning for transfer to new sessions. DeepSets serves as the permutation-invariant baseline, aggregating neuron-wise activity representations by summation without identity-guided routing. Full details of the baseline adaptation and identity feature integration are provided in Appendix A.6. To isolate the effect of the group-routing architecture from access to multimodal biological features, each baseline is evaluated in two variants. The first is its original form, and the second is an augmented form that receives the same per-neuron identity features as our method, with all other training details unchanged. Concretely, neuronal identity features are incorporated into neuron-level embeddings (Appendix A.10).

Second, we evaluate neural manifold analysis by comparing the manifolds learned from our method with those produced by CEBRA [3], MARBLE [4], LFADS [26], and Joint PCA. CEBRA learns low-dimensional embeddings of neural population activity through contrastive learning. MARBLE models local flow fields over the manifold structure using geometric deep learning. LFADS infers single-trial population dynamics through a sequential variational autoencoder. Joint PCA provides the linear baseline, projecting jointly concatenated population activity from source animals into a shared low-dimensional space. Full details of this adaptation are provided in Appendix A.8.

3.3 Evaluation metrics and protocol

To evaluate decoding performance, we report the Pearson correlation coefficient (PCC) for behavior regression and accuracy scores for stimulus classification, both evaluated on held-out data.

To assess the neural manifold analysis, we quantify how the latent representation preserves the temporal geometry of population dynamics on held-out data. We compute three metrics on the latent trajectory $\{z_t\}_{t=1}^{T}$ that evaluate global temporal organization and local structure (Appendix A.4).

- **Temporal Distance Correlation** ($\uparrow$) measures the Spearman correlation between pairwise latent distances $\|z_t - z_s\|_2$ and temporal distances $|t - s|$. Higher values indicate that states nearby in time remain close in latent space while temporally distant states separate.
- **Normalized Latent-Neighbor Time Gap** ($\downarrow$) measures the temporal distance to each point's nearest neighbor in latent space, normalized by recording length and averaged over time. Lower values indicate higher temporal consistency in local latent neighborhoods.
- **Local Trajectory Linearity** ($\uparrow$) measures the fraction of variance explained by the first principal component within short sliding windows of the latent trajectory [27]. Higher values indicate smoother local dynamics.

Held-out transfer protocol. We evaluate transfer at the level of complete held-out recordings. All methods are trained exclusively on source animals or sessions and then applied directly to the held-out recordings. During inference, learned parameters are strictly frozen. At evaluation time, the entire held-out recording is analyzed, leveraging all available data to construct the per-neuron identity features prior to inference (see Appendix A.3 for construction details). All behavioral and stimulus labels from the held-out recordings are used only to compute the reported evaluation metrics. To ensure fair comparison, this exact transfer protocol is maintained across all baseline evaluations.

For the WARP dataset, we train on two fish and evaluate on the remaining fish. We average these results across all three held-out combinations and five random seeds. For the Allen mouse visual cortex electrophysiology datasets, we train on a fixed set of 20 sessions and evaluate independently on the remaining held-out sessions (six for the Brain Observatory and five for the Functional Connectivity

Table 1: Behavior and stimulus decoding under direct held-out transfer across three datasets (mean $\pm$ std). Bold indicates the best result per column within each dataset. * denotes models adapted to calcium signals.

Dataset	Method	Behavior (PCC) ↑	Stimulus (Balanced Accuracy) ↑	Stimulus (Accuracy) ↑
WARP	**Ours**	$\mathbf{0.793 \pm 0.015}$	$\mathbf{0.835 \pm 0.029}$	$\mathbf{0.839 \pm 0.029}$
	POYO* [9]	0.304 ± 0.120	0.153 ± 0.032	0.156 ± 0.034
	NEDS* [19]	0.248 ± 0.207	0.195 ± 0.034	0.196 ± 0.033
	DeepSets [18]	0.160 ± 0.081	0.161 ± 0.011	0.160 ± 0.011
Allen Brain Observatory	**Ours**	$\mathbf{0.734 \pm 0.008}$	$\mathbf{0.893 \pm 0.012}$	$\mathbf{0.899 \pm 0.010}$
	POYO [9]	0.526 ± 0.043	0.220 ± 0.022	0.258 ± 0.114
	NEDS [19]	0.500 ± 0.042	0.424 ± 0.043	0.432 ± 0.078
	DeepSets [18]	0.615 ± 0.022	0.684 ± 0.027	0.612 ± 0.034
Allen Functional Connectivity	**Ours**	$\mathbf{0.694 \pm 0.058}$	$\mathbf{0.896 \pm 0.017}$	$\mathbf{0.932 \pm 0.007}$
	POYO [9]	0.554 ± 0.080	0.211 ± 0.008	0.288 ± 0.117
	NEDS [19]	0.454 ± 0.036	0.398 ± 0.020	0.408 ± 0.091
	DeepSets [18]	0.588 ± 0.096	0.730 ± 0.044	0.663 ± 0.095

datasets), averaging across three random seeds. All identity feature ablations follow the same held-out protocol, differing only in the specific subsets of identity features provided to the model.

4 Results

4.1 Behavior and Stimulus Decoding Under Direct Held-Out Transfer

Table 1 reports behavior and stimulus decoding performance under held-out transfer to unseen animals or sessions without retraining across all three datasets. Our framework achieves the best performance across all datasets and tasks, outperforming prior methods in all cases. On the WARP dataset, our method achieves a PCC value of 0.793. On the Allen Brain Observatory and Functional Connectivity datasets, our method yields PCC values of 0.734 and 0.694, respectively, marking the highest performance in each setting.

Next, we investigate whether the performance gains stem simply from the inclusion of neuronal features, or from how our framework explicitly utilizes them. Accordingly, we provide POYO, NEDS, and DeepSets with the same multimodal biological features, leaving their original implementation details unchanged (Appendix A.6). Across datasets, these identity-augmented baselines show consistent improvements over their original counterparts. Behavior decoding on the WARP dataset increases from 0.304 to 0.489 for POYO, from 0.248 to 0.574 for NEDS, and from 0.160 to 0.466 for DeepSets. Similar trends are observed for stimulus decoding and for the Allen datasets, where identity features are based mainly on position and anatomical region (Appendix A.10). However, their performance remains below that of our method across all datasets. This pattern demonstrates that our framework's superior performance cannot be fully explained by just the availability of these biological features. Instead, the key advantage lies in how they are incorporated through identity-guided organization.

We further assess our framework's robustness to the choice and availability of identity modalities by evaluating ablations where we remove one modality at a time, reported in Appendix A.11. Across both datasets, all ablated variants remain competitive and continue to outperform baseline methods, showing that the performance gains are not attributable to any single modality. The model without transcriptomic descriptors achieves a behavior PCC of 0.727 and a stimulus balanced accuracy of 0.711, compared with 0.793 and 0.835 for the full model. This suggests that the framework remains competitive even when costly jointly acquired transcriptomic measurements are unavailable.

4.2 Neural Manifold Geometry Under Direct Held-Out Transfer

Next, we evaluate the performance of our framework in neural manifold analysis, specifically testing its ability to preserve temporal and local geometric structure on unseen recordings without retraining, a capability that has remained largely beyond the reach of existing methods. This capability is crucial for rigorous comparative analysis, allowing researchers to investigate how neural population dynamics vary or remain invariant across different subjects under identical computational constraints.

Table 2: Geometric structure of latent trajectories evaluated on held-out animals or sessions across three datasets (mean $\pm$ std). Bold indicates the best result per column.

Dataset	Method	Temporal Distance Correlation ↑	Normalized Latent-Neighbor Time Gap ↓	Local Trajectory Linearity ↑
WARP	**Ours**	$\mathbf{0.633 \pm 0.078}$	$\mathbf{0.145 \pm 0.009}$	$\mathbf{0.911 \pm 0.006}$
	CEBRA [3]	0.086 ± 0.010	0.328 ± 0.003	0.684 ± 0.010
	MARBLE [4]	0.045 ± 0.017	0.332 ± 0.002	0.512 ± 0.006
	LFADS [26]	0.141 ± 0.020	0.297 ± 0.012	0.666 ± 0.015
	Joint PCA	0.151 ± 0.045	0.312 ± 0.007	0.887 ± 0.032
Allen Brain Observatory	**Ours**	$\mathbf{0.355 \pm 0.033}$	$\mathbf{0.063 \pm 0.021}$	$\mathbf{0.896 \pm 0.005}$
	CEBRA [3]	0.029 ± 0.004	0.298 ± 0.005	0.497 ± 0.006
	MARBLE [4]	0.049 ± 0.004	0.190 ± 0.009	0.654 ± 0.004
	LFADS [26]	0.179 ± 0.011	0.134 ± 0.009	0.545 ± 0.004
	Joint PCA	0.071 ± 0.019	0.244 ± 0.017	0.699 ± 0.029
Allen Functional Connectivity	**Ours**	$\mathbf{0.317 \pm 0.036}$	$\mathbf{0.059 \pm 0.019}$	$\mathbf{0.893 \pm 0.013}$
	CEBRA [3]	0.025 ± 0.006	0.298 ± 0.015	0.494 ± 0.007
	MARBLE [4]	0.047 ± 0.005	0.196 ± 0.014	0.663 ± 0.021
	LFADS [26]	0.150 ± 0.012	0.126 ± 0.019	0.507 ± 0.009
	Joint PCA	0.052 ± 0.017	0.226 ± 0.041	0.682 ± 0.030

Table 2 reports temporal and local geometric metrics on held-out animals under the same held-out transfer protocol. On the WARP dataset, our framework achieves a temporal distance correlation of 0.633 ± 0.078, indicating that neural states nearby in time remain close in latent space. It also obtains the lowest normalized nearest-neighbor temporal gap, 0.145 ± 0.009, and the highest local trajectory linearity, 0.911 ± 0.006. For the manifold baselines, we additionally test $1,000$ random neuron orderings at inference time and find that our method outperforms the best sampled baseline results in most cases (Appendix A.12). A consistent pattern is observed on the Allen mouse visual cortex datasets, where our framework again outperforms all baselines across the evaluated metrics. Figure 2 shows that held-out trajectories remain temporally coherent and avoid the fragmentation or collapse observed in several baseline embeddings. Trajectories colored by behavior and stimulus labels are shown in Figs. A4 and A5, respectively.

Together, these results indicate that our framework produces transferred latent trajectories with more consistent temporal neighborhoods and smoother local structure in held-out recordings under the evaluated metrics. This advantage is consistent with a key difference in how the methods handle unseen animals. Baseline methods learn representations tied to the specific set of neurons recorded in the source animals and lack a way to align these representations when the neuron population changes across individuals. Our framework instead grounds each neuron in a common identity coordinate, so the model induces a shared identity-defined group coordinate that supports transfer to held-out animals and sessions.

5 Related Work

Neural population dynamics analysis methods. Recent work in neural population dynamics has increasingly explored the use of deep neural networks [28, 29]. Methods such as POYO, POCO, and NEDS learn representations that aggregate neural activity across recordings through tokenization or latent mapping, enabling cross-session and cross-task decoding without requiring neuron-wise correspondence [9, 19, 30, 31]. Across these methods, neurons are represented through learned, session-specific representations derived from observed dynamics and implicit neuron indexing, without explicit grounding in biological identity. While some recent work has begun to explore learning neuron identity from population context [10], cross-session generalization still primarily relies on statistical regularities in observed dynamics rather than on molecular or spatial information that makes neurons comparable across individuals.

Neural manifold methods. A central observation in neuroscience is that the high-dimensional activity of neural populations often lies on low-dimensional manifolds whose geometry reflects the structure of the underlying dynamics [3, 4, 26]. Methods for characterizing these manifolds include sequential variational autoencoders, contrastive embedding approaches that align neural activity with behavioral or stimulus variables [3], and geometric deep learning methods that analyze local

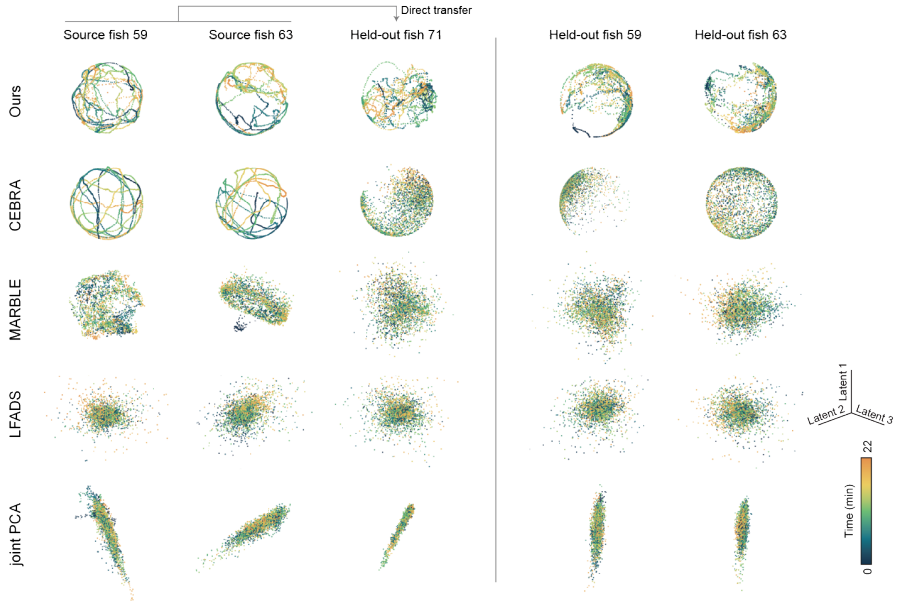

Figure 2: Latent trajectories for two source animals (Fish 59 and Fish 63) and held-out animals across all methods. Points are colored by time within each recording. Our method preserves structured temporal trajectories in held-out animals, whereas baseline methods yield fragmented or collapsed trajectories under direct held-out transfer.

flow fields over manifold structure [4]. These approaches have become standard tools for relating population-level dynamics to cognition, perception, and movement [28, 29].

Multimodal neuron profiling and cell-type atlases. Recent work in systems neuroscience shows that neurons can be characterized along multiple biological axes. Multimodal profiling approaches, including Patch-seq and spatial transcriptomics, enable large-scale measurement of molecular and spatial properties [11, 17, 32–35]. Large-scale atlas efforts have produced spatially resolved gene expression maps and anatomical reference frameworks at cellular resolution across whole-brain volumes [14–16, 23, 24, 36]. More recent datasets jointly profile or align neural activity with transcriptomic identity within the same brain [17, 20]. Across species, multimodal measurements show that molecular and spatial identity are associated with functional circuit organization [13, 37]. This growing availability of per-neuron identity information provides a natural foundation for incorporating biological structure into methods for analyzing population dynamics.

6 Discussion

We propose multimodal biological features as an organizational basis for analyzing neural population dynamics. Our method incorporates these features to route neurons into latent groups and analyze dynamics at the group level. This design preserves invariance to neuron order while enabling neurons with similar biological context to contribute through a shared functional organization. As a result, populations from new animals or sessions can be mapped into a common identity-defined structure without requiring neuron-wise correspondence or retraining.

Across zebrafish and mouse visual cortex datasets, our method achieves strong behavior and stimulus decoding while producing geometrically consistent neural manifolds under direct held-out transfer. Baseline methods also improve when provided with the same neuronal identity features, yet our routing-based framework consistently outperforms them. This suggests that effective generalization depends not only on the availability of neuronal identity information, but also on how it is used to impose a shared organization on population dynamics.

Our method also remains effective when identity features are not jointly recorded with neural activity, performing comparably when using independently acquired atlas data. This expands applicability to

existing datasets where neural activity is available but neuron-level biological feature measurements are not collected in the same animals.

The performance of the method is shaped by the availability and quality of identity features. Joint measurements of neural activity and transcriptomic identity remain limited [17, 20, 32, 33, 35], and using independently acquired atlas data introduces additional challenges in registration [38–40]. Variability across individuals and imperfections in alignment can introduce noise into identity features, affecting group assignments. Improvements in multimodal registration and joint data acquisition are therefore likely to further strengthen identity-based population analysis.

More broadly, these results highlight shared biological structure as an underexplored signal for neural population analysis. Neuronal identity provides a complementary axis for organizing population representations across individuals, suggesting a path toward methods that can integrate data across subjects without requiring exact neuron-wise correspondence.

Acknowledgements. This research was supported by the National Research Foundation of Korea (NRF) (RS-2026-25475246, RS-2026-25504129, RS-2026-25475217, RS-2023-00209473, RS-2023-00264409, RS-2026-25498280, RS-2022-NR068758), Korea Basic Science Institute (National Research Facilities and Equipment Center) grant funded by the Ministry of Science and ICT (RS-2024-00401676), Bio & Medical Technology Development Program through the NRF funded by the Ministry of Science and ICT (RS-2021-NR056586), the BK21 plus program through the NRF funded by the Ministry of Education of Korea, a Research Grant from HFSP (RGP003/2024) with the award DOI (https://doi.org/10.52044/HFSP.RGP0032024.pc.gr.194148), the KAIST Quantum+X Convergence R&D Project, the KAIST C2 (Creative & Challenging) Project, the KAIST Jang Young Sil Fellow Program, and KAIST National Research and Development Project (N10260124), and KAIST AI-Science Hub Project (N10260124).

References

[1] Anne E Urai, Brent Doiron, Andrew M Leifer, and Anne K Churchland. Large-scale neural recordings call for new insights to link brain and behavior. *Nature neuroscience*, 25(1):11–19, 2022.

[2] Carsen Stringer and Marius Pachitariu. Analysis methods for large-scale neuronal recordings. *Science*, 386(6722):eadp7429, 2024.

[3] Steffen Schneider, Jin Hwa Lee, and Mackenzie Weygandt Mathis. Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 617(7960):360–368, 2023.

[4] Adam Gosztolai, Robert L Peach, Alexis Arnaudon, Mauricio Barahona, and Pierre Vandergheynst. Marble: interpretable representations of neural population dynamics using geometric deep learning. *Nature Methods*, 22(3):612–620, 2025.

[5] Seungjae Han, Joshua Yedam You, Minho Eom, Sungjin Ahn, Eun-Seo Cho, and Young-Gyu Yoon. From pixels to information: Artificial intelligence in fluorescence microscopy. *Advanced Photonics Research*, 5(9):2300308, 2024.

[6] Mackenzie Weygandt Mathis, Adriana Perez Rotondo, Edward F Chang, Andreas S Tolias, and Alexander Mathis. Decoding the brain: From neural representations to mechanistic models. *Cell*, 187(21):5814–5832, 2024.

[7] Joshua Yedam You, Minho Eom, Tae-Ik Choi, Eun-Seo Cho, Jieun Choi, Minyoung Lee, Changyeop Shin, Jieun Moon, Eunji Kim, Pilhan Kim, Cheol-Hee Kim, and Young-Gyu Yoon. Preserving spatial and quantitative information in unpaired biomedical image-to-image translation. *Cell Reports Methods*, 5(6), 2025.

[8] Jongmin Yoon, Soi Kim, Seungjae Han, Minho Eom, Eun-Seo Cho, Fatemeh Dehghan Nezhad, Soojung Hong, Eunjee Kim, In-Hyun Park, Euiheon Chung, Keunyoo Shin, Young-Gyu Yoon, and Myunghwan Choi. Event-based optical imaging and reconstruction of in vivo neuronal and vascular dynamics. *PhotoniX*, 7(1):21, 2026.

[9] Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, Santosh Nachimuthu, Michael Mendelson, Blake Richards, Matthew Perich, Guillaume Lajoie, and Eva Dyer. A

unified, scalable framework for neural population decoding. *Advances in Neural Information Processing Systems*, 36:44937–44956, 2023.

[10] Vinam Arora, Divyansha Lachi, Ian J Knight, Mehdi Azabou, Blake Richards, Cole Hurwitz, Joshua H Siegle, and Eva L Dyer. Know thyself by knowing others: Learning neuron identity from population context. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[11] Nathan W Gouwens, Staci A Sorensen, Fahimeh Baftizadeh, Agata Budzillo, Brian R Lee, Tim Jarsky, Lauren Alfiler, Katherine Baker, Eliza Barkan, Kyla Berry, Darren Bertagnolli, Kris Bickley, Jasmine Bomben, Thomas Braun, Krissy Brouner, Tamara Casper, Kirsten Crichton, Tanya L Daigle, Rachel Dalley, Rebecca A de Frates, Nick Dee, Tsega Desta, Samuel Dingman Lee, Nadezhda Dotson, Tom Egdorf, Lauren Ellingwood, Rachel Enstrom, Luke Esposito, Colin Farrell, David Feng, Olivia Fong, Rohan Gala, Clare Gamlin, Amanda Gary, Alexandra Glandon, Jeff Goldy, Melissa Gorham, Lucas Graybuck, Hong Gu, Kristen Hadley, Michael J Hawrylycz, Alex M Henry, DiJon Hill, Madie Hupp, Sara Kebede, Tae Kyung Kim, Lisa Kim, Matthew Kroll, Changkyu Lee, Katherine E Link, Matthew Mallory, Rusty Mann, Michelle Maxwell, Medea McGraw, Delissa McMillen, Alice Mukora, Lindsay Ng, Lydia Ng, Kiet Ngo, Philip R Nicovich, Aaron Oldre, Daniel Park, Hanchuan Peng, Osnat Penn, Thanh Pham, Alice Pom, Zoran Popović, Lydia Potekhina, Ramkumar Rajanbabu, Shea Ransford, David Reid, Christine Rimorin, Miranda Robertson, Kara Ronellenfitch, Augustin Ruiz, David Sandman, Kimberly Smith, Josef Sulc, Susan M Sunkin, Aaron Szafer, Michael Tieu, Amy Torkelson, Jessica Trinh, Herman Tung, Wayne Wakeman, Katelyn Ward, Grace Williams, Zhi Zhou, Jonathan T Ting, Anton Arkhipov, Uygar Sümbül, Ed S Lein, Christof Koch, Zizhen Yao, Bosiljka Tasic, Jim Berg, Gabe J Murphy, and Hongkui Zeng. Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. *Cell*, 183(4):935–953, 2020.

[12] Carsten K Pfeffer, Mingshan Xue, Miao He, Z Josh Huang, and Massimo Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature Neuroscience*, 16(8):1068–1076, 2013.

[13] Anton Arkhipov, Nuno da Costa, Saskia de Vries, Trygve Bakken, Corbett Bennett, Amy Bernard, Jim Berg, Michael Buice, Forrest Collman, Tanya Daigle, Marina Garrett, Nathan Gouwens, Peter A. Groblewski, Julie Harris, Michael Hawrylycz, Rebecca Hodge, Tim Jarsky, Brian Kalmbach, Jerome Lecoq, Brian Lee, Ed Lein, Boaz Levi, Stefan Mihalas, Lydia Ng, Shawn Olsen, Clay Reid, Joshua H. Siegle, Staci Sorenson, Bosiljka Tasic, Carol Thompson, Jonathan T. Ting, Cindy van Velthoven, Shengin Yao, Zizhen Yao, Christof Koch, and Hongkui Zeng. Integrating multimodal data to understand cortical circuit architecture and function. *Nature Neuroscience*, 28(4):717–730, 2025.

[14] Amit Zeisel, Hannah Hochgerner, Peter Lönnerberg, Anna Johnsson, Fatima Memic, Job van der Zwan, Martin Häring, Emelie Braun, Lars E. Borm, Gioele La Manno, Simone Codeluppi, Alessandro Furlan, Kawai Lee, Nathan Skene, Kenneth D. Harris, Jens Hjerling-Leffler, Ernest Arenas, Patrik Ernfors, Ulrika Marklund, and Sten Linnarsson. Molecular architecture of the mouse nervous system. *Cell*, 174(4):999–1014, 2018.

[15] Bosiljka Tasic, Zizhen Yao, Lucas T. Graybuck, Kimberly A. Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, Emma Garren, Michael N. Economo, Sarada Viswanathan, Osnat Penn, Trygve Bakken, Vilas Menon, Jeremy Miller, Olivia Fong, Karla E. Hirokawa, Kanan Lathia, Christine Rimorin, Michael Tieu, Rachael Larsen, Tamara Casper, Eliza Barkan, Matthew Kroll, Sheana Parry, Nadiya V. Shapovalova, Daniel Hirschstein, Julie Pendergraft, Heather A. Sullivan, Tae Kyung Kim, Aaron Szafer, Nick Dee, Peter Groblewski, Ian Wickersham, Ali Cetin, Julie A. Harris, Boaz P. Levi, Susan M. Sunkin, Linda Madisen, Tanya L. Daigle, Loren Looger, Amy Bernard, John Phillips, Ed Lein, Michael Hawrylycz, Karel Svoboda, Allan R. Jones, Christof Koch, and Hongkui Zeng. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78, 2018.

[16] Zizhen Yao, Cindy T. J. van Velthoven, Thuc Nghi Nguyen, Jeff Goldy, Adriana E. Sedeno-Cortes, Fahimeh Baftizadeh, Darren Bertagnolli, Tamara Casper, Megan Chiang, Kirsten Crichton, Song-Lin Ding, Olivia Fong, Emma Garren, Alexandra Glandon, Nathan W. Gouwens, James Gray, Lucas T. Graybuck, Michael J. Hawrylycz, Daniel Hirschstein, Matthew Kroll,

Kanan Lathia, Changkyu Lee, Boaz Levi, Delissa McMillen, Stephanie Mok, Thanh Pham, Qingzhong Ren, Christine Rimorin, Nadiya Shapovalova, Josef Sulc, Susan M. Sunkin, Michael Tieu, Amy Torkelson, Herman Tung, Katelyn Ward, Nick Dee, Kimberly A. Smith, Bosiljka Tasic, and Hongkui Zeng. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*, 184(12):3222–3241, 2021.

[17] BRAIN Initiative Cell Census Network. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*, 598(7879):86–102, 2021.

[18] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

[19] Yizi Zhang, Yanchen Wang, Mehdi Azabou, Alexandre Andre, Zixuan Wang, Hanrui Lyu, The International Brain Laboratory, Eva Dyer, Liam Paninski, and Cole Hurwitz. Neural encoding and decoding at scale. In *Forty-second International Conference on Machine Learning*, 2025.

[20] Emmanuel Marquez-Legorreta, Greg M Fleishman, Luuk W Hesselink, Mark Eddison, Kasper Smeets, Carsen Stringer, Philipp J Keller, Sujatha Narayan, Alex B Chen, Brett D Mensh, Scott M Sternson, Bernhard Englitz, Paul W Tillberg, and Misha B Ahrens. Whole-brain co-mapping of gene expression and neuronal activity at cellular resolution in behaving zebrafish. *bioRxiv*, 2026.

[21] Joshua H. Siegle, Xiaoxuan Jia, Séverine Durand, Sam Gale, Corbett Bennett, Nile Graddis, Greggory Heller, Tamina K. Ramirez, Hannah Choi, Jennifer A. Luviano, Peter A. Groblewski, Ruweida Ahmed, Anton Arkhipov, Amy Bernard, Yazan N. Billeh, Dillan Brown, Michael A. Buice, Nicolas Cain, Shiella Caldejon, Linzy Casal, Andrew Cho, Maggie Chvilicek, Timothy C. Cox, Kael Dai, Daniel J. Denman, Saskia E. J. de Vries, Roald Dietzman, Luke Esposito, Colin Farrell, David Feng, John Galbraith, Marina Garrett, Emily C. Gelfand, Nicole Hancock, Julie A. Harris, Robert Howard, Brian Hu, Ross Hytnen, Ramakrishnan Iyer, Erika Jessett, Katelyn Johnson, India Kato, Justin Kiggins, Sophie Lambert, Jerome Lecoq, Peter Ledochowitsch, Jung Hoon Lee, Arielle Leon, Yang Li, Elizabeth Liang, Fuhui Long, Kyla Mace, Jose Melchior, Daniel Millman, Tyler Mollenkopf, Chelsea Nayan, Lydia Ng, Kiet Ngo, Thuyahn Nguyen, Philip R. Nicovich, Kat North, Gabriel Koch Ocker, Doug Ollerenshaw, Michael Oliver, Marius Pachitariu, Jed Perkins, Melissa Reding, David Reid, Miranda Robertson, Kara Ronellenfitch, Sam Seid, Cliff Slaughterbeck, Michelle Stoecklin, David Sullivan, Ben Sutton, Jackie Swapp, Carol Thompson, Kristen Turner, Wayne Wakeman, Jennifer D. Whitesell, Derric Williams, Ali Williford, Rob Young, Hongkui Zeng, Sarah Naylor, John W. Phillips, R. Clay Reid, Stefan Mihalas, Shawn R. Olsen, and Christof Koch. Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852):86–92, 2021.

[22] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pages 3744–3753. PMLR, 2019.

[23] Michael Kunst, Eva Laurell, Nouwar Mokayes, Anna Kramer, Fumi Kubo, António M Fernandes, Dominique Förster, Marco Dal Maschio, and Herwig Baier. A cellular-resolution atlas of the larval zebrafish brain. *Neuron*, 103(1):21–38, 2019.

[24] Inbal Shainer, Enrico Kuehn, Eva Laurell, Mariam Al Kassar, Nouwar Mokayes, Shachar Sherman, Johannes Larsch, Michael Kunst, and Herwig Baier. A single-cell resolution gene expression atlas of the larval zebrafish brain. *Science Advances*, 9(8):eade9909, 2023.

[25] Quanxin Wang, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, Maitham Naeemi, Benjamin Facer, Anh Ho, Tim Dolbeare, Brandon Blanchard, Nick Dee, Wayne Wakeman, Karla E. Hirokawa, Aaron Szafer, Susan M. Sunkin, Seung Wook Oh, Amy Bernard, John W. Phillips, Michael Hawrylycz, Christof Koch, Hongkui Zeng, Julie A. Harris, and Lydia Ng. The allen mouse brain common coordinate framework: a 3d reference atlas. *Cell*, 181(4):936–953, 2020.

[26] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg,

Jamie M. Henderson, Krishna V. Shenoy, L. F. Abbott, and David Sussillo. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, 15(10):805–815, 2018.

[27] Christian K Machens. Demixing population activity in higher cortical areas. *Frontiers in computational neuroscience*, 4:126, 2010.

[28] Krishna V Shenoy, Maneesh Sahani, and Mark M Churchland. Cortical control of arm movements: a dynamical systems perspective. *Annual review of neuroscience*, 36(1):337–359, 2013.

[29] Krishna V Shenoy and Jonathan C Kao. Measurement, manipulation and modeling of brain-wide neural population dynamics. *Nature Communications*, 12(1):633, 2021.

[30] Mehdi Azabou, Krystal Xuejing Pan, Vinam Arora, Ian Jarratt Knight, Eva L Dyer, and Blake Aaron Richards. Multi-session, multi-task neural decoding from distinct cell-types and brain regions. In *The Thirteenth International Conference on Learning Representations*, 2025.

[31] Yu Duan, Hamza Tahir Chaudhry, Misha B Ahrens, Christopher D Harvey, Matthew G Perich, Karl Deisseroth, and Kanaka Rajan. Poco: Scalable neural forecasting through population conditioning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

[32] Cathryn R Cadwell and Andreas S Tolias. Patch-seq: multimodal profiling of single-cell morphology, electrophysiology, and gene expression. In *Single Cell Analysis: Methods and Protocols*, pages 227–243. Springer, 2024.

[33] János Fuzik, Amit Zeisel, Zoltán Máté, Daniela Calvigioni, Yuchio Yanagawa, Gábor Szabó, Sten Linnarsson, and Tibor Harkany. Integration of electrophysiological recordings with single-cell rna-seq data identifies neuronal subtypes. *Nature biotechnology*, 34(2):175–183, 2016.

[34] Mingting Shao, Wei Zhang, Ye Li, Lei Tang, Zhao-Zhe Hao, and Sheng Liu. Patch-seq: advances and biological applications. *Cellular and Molecular Neurobiology*, 44(1):8, 2024.

[35] Haruya Yagishita and Takuya Sasaki. Integrating physiological and transcriptomic analyses at the single-neuron level. *Neuroscience Research*, 214:69–74, 2025.

[36] Lei Han, Zhen Liu, Zehua Jing, Yuxuan Liu, Yujie Peng, Huizhong Chang, Junjie Lei, Kexin Wang, Yuanfang Xu, Wei Liu, et al. Single-cell spatial transcriptomic atlas of the whole mouse brain. *Neuron*, 113(13):2141–2160, 2025.

[37] Lijuan Liu, Zhixi Yun, Linus Manubens-Gil, Hanbo Chen, Feng Xiong, Hongwei Dong, Hongkui Zeng, Michael Hawrylycz, Giorgio A Ascoli, and Hanchuan Peng. Connectivity of single neurons classifies cell subtypes in mouse brains. *Nature Methods*, 22(4):861–873, 2025.

[38] Kalen Clifton, Manjari Anant, Gohta Aihara, Lyla Atta, Osagie K Aimiuwu, Justus M Kebschull, Michael I Miller, Daniel Tward, and Jean Fan. Stalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nature Communications*, 14(1):8123, 2023.

[39] Junmo Cho, Seungjae Han, Eun-Seo Cho, Kijung Shin, and Young-Gyu Yoon. Robust and efficient alignment of calcium imaging data through simultaneous low rank and sparse decomposition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1939–1948, 2023.

[40] Biao Zhang, Shuqin Zhang, and Shihua Zhang. Whole brain alignment of spatial transcriptomics between humans and mice with brainalign. *Nature Communications*, 15(1):6302, 2024.

[41] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

A Supplementary Results and Implementation Details

A.1 Dataset overview

Multimodal information across recordings. Figures A1 and A2 show representative WARP fish and Allen visual cortex sessions. Although the exact recorded neurons or units differ across animals and sessions, both datasets exhibit broadly comparable spatial and regional organization. Fish in the WARP dataset show similar whole-brain distributions of transcriptomic markers and brain-region labels, while Allen sessions cover comparable visual cortical and associated areas. This shared structure motivates our held-out transfer evaluation, where models are applied to held-out recordings without neuron-wise correspondence.

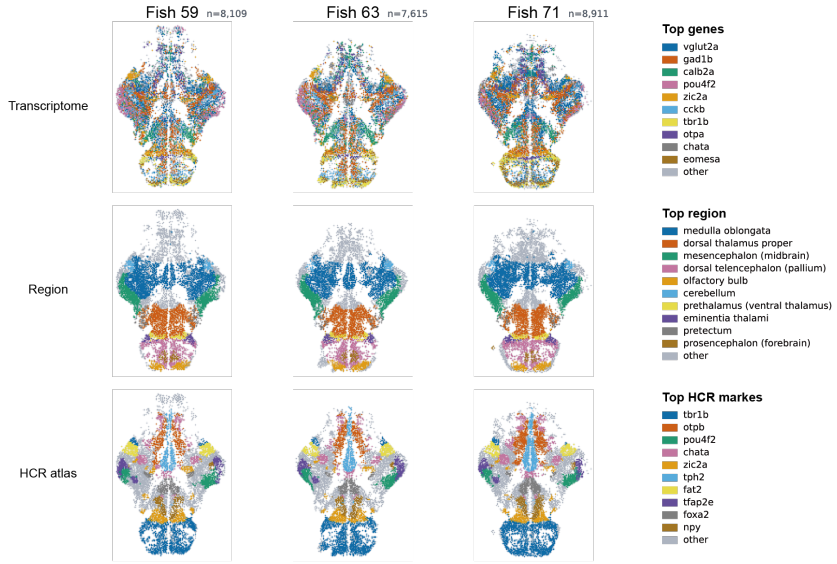

Figure A1: Data from three fish in the WARP dataset. Neurons are shown in the registered zebrafish brain coordinate space and colored by transcriptomic marker identity (top), anatomical region label (middle), and selected marker-gene categories (bottom).

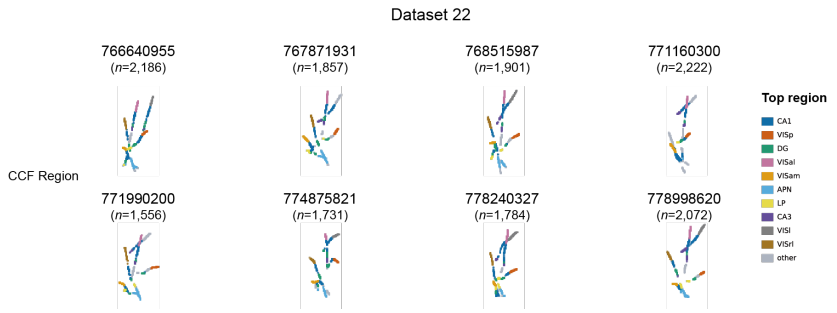

Figure A2: Representative Allen visual cortex sessions shown in a shared anatomical coordinate space and colored by visual area label.

Decoding Targets. Figure A3 summarizes the supervised decoding targets used in each dataset. Behavior decoding is formulated as scalar regression and evaluated with Pearson correlation coefficient. Stimulus decoding is formulated as multi-class classification and evaluated with accuracy. In WARP, the behavioral target is swim power, and the stimulus target is the type of visual stimuli. In the Allen datasets, the behavioral target is running speed, and the stimulus target is the type of visual stimuli. These targets are used consistently for the proposed method and all supervised decoding baselines.

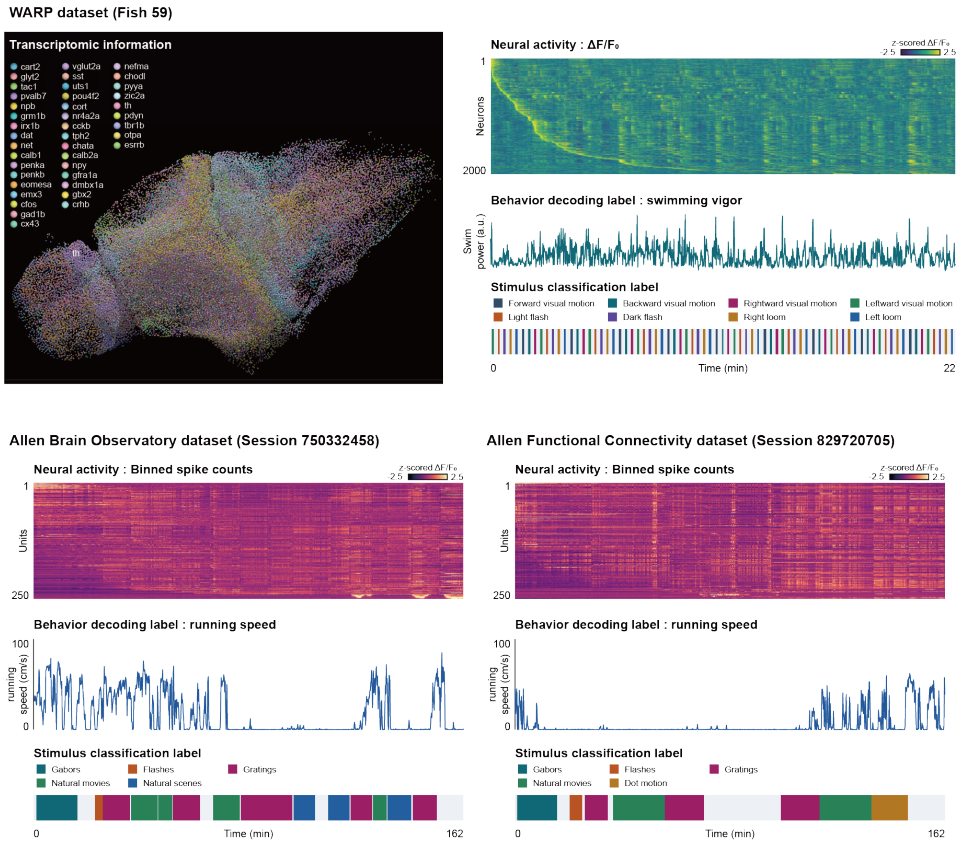

Figure A3: Decoding targets used in WARP and Allen experiments. For WARP, we show a representative fish with neural activity from selected neurons, the swim-power behavioral target, and the visual-stimulus classification target. For Allen, we show representative sessions from Allen Brain Observatory dataset and Allen Functional Connectivity dataset with binned spike counts from selected units, running-speed behavioral targets, and grouped stimulus-family labels. Behavior decoding is evaluated with Pearson correlation coefficient, and stimulus decoding is evaluated with accuracy.

A.2 Proof of Permutation Invariance

In Section 2.1, the neural population is represented as the unordered collection

$$\mathcal{S} = \{(\mathbf{f}_i, \mathbf{x}_i)\}_{i=1}^{N}.$$

For the proof, we make the arbitrary enumeration of this collection explicit and write

$$\mathbf{S} = \big((\mathbf{f}_1, \mathbf{x}_1), \ldots, (\mathbf{f}_N, \mathbf{x}_N)\big).$$

For any permutation $\pi \in \mathfrak{S}_N$, define the reindexed input

$$\mathbf{S}^{\pi} = \big((\mathbf{f}_{\pi(1)}, \mathbf{x}_{\pi(1)}), \ldots, (\mathbf{f}_{\pi(N)}, \mathbf{x}_{\pi(N)})\big).$$

Proposition 1 (Permutation invariance)**.** *Let $\mathcal{F}$ be the population encoder defined by the identity-guided routing in Eq.* (4)*, the shared dynamics encoder in Eq.* (5)*, the group-level aggregation in Eq.* (6)*, and the group interaction module Γ. If the identity encoder, dynamics encoder, and routing map are shared across neurons, then for any permutation $\pi \in \mathfrak{S}_N$,*

$$\mathcal{F}(\mathbf{S}^{\pi}) = \mathcal{F}(\mathbf{S}).$$

Proof. Let the input neuron ordering be permuted by π. Since the identity encoder, dynamics encoder, and routing map are shared across neurons, the corresponding neuron-level quantities are permuted in the same way:

$$\mathbf{s}_i^{\pi} = \mathbf{s}_{\pi(i)}, \qquad \mathbf{v}_i^{\pi} = \mathbf{v}_{\pi(i)}, \qquad a_{ik}^{\pi} = a_{\pi(i)k}.$$

For each latent group k, the effective group mass under the permuted input is

$$m_k^\pi = \sum_{i=1}^{N} a_{ik}^\pi = \sum_{i=1}^{N} a_{\pi(i)k} = \sum_{j=1}^{N} a_{jk} = m_k,$$

where the third equality follows by reindexing $j = \pi(i)$. Similarly,

$$\sum_{i=1}^{N} a_{ik}^\pi \mathbf{v}_i^\pi = \sum_{i=1}^{N} a_{\pi(i)k} \mathbf{v}_{\pi(i)} = \sum_{j=1}^{N} a_{jk} \mathbf{v}_j.$$

Therefore,

$$\mathbf{c}_k^\pi = \frac{1}{\sqrt{\max(m_k, \epsilon)}} \sum_{i=1}^{N} a_{ik}^\pi \mathbf{v}_i^\pi = \frac{1}{\sqrt{\max(m_k, \epsilon)}} \sum_{j=1}^{N} a_{jk} \mathbf{v}_j = \mathbf{c}_k.$$

Thus $\mathbf{C}^\pi = \mathbf{C}$. Since the final representation is computed as a deterministic function of the group matrix,

$$\mathbf{z} = \Gamma(\mathbf{C}),$$

we have

$$\mathbf{z}^\pi = \Gamma(\mathbf{C}^\pi) = \Gamma(\mathbf{C}) = \mathbf{z}.$$

Hence, $\mathcal{F}(\mathbf{S}^\pi) = \mathcal{F}(\mathbf{S})$, and $\mathcal{F}$ is permutation-invariant. □

A.3 Neuronal Identity Feature Construction

We designed the identity encoder framework to serve as a versatile interface for any available neuron-level features, allowing the set of input modalities to vary across datasets and experimental conditions. Table A1 summarizes all feature groups that we used and their availability across datasets. For modality-ablation experiments, we use the same construction procedure and vary only which groups are provided. For the WARP dataset, spatial coordinates are registered to the mapzebrain reference spaces [23], and jointly acquired transcriptomic marker features are available alongside atlas-derived HCR expression features. For the Allen dataset, joint transcriptomic measurements are not available, and spatial coordinates are registered to the Allen Common Coordinate Framework [25].

mapzebrain HCR gene expression atlas features. To test whether independently acquired, public atlas information could provide useful spatial gene-expression context for WARP neurons, we derived atlas-based features from the mapzebrain single-cell-resolution gene-expression atlas [23, 24], which uses multiplexed fluorescent in situ RNA hybridization chain reaction (HCR) to generate spatial expression maps across the larval zebrafish brain. The atlas contains 236 HCR marker volumes in the mapzebrain reference space. For each WARP neuron, marker intensity was sampled at its registered position using a local Gaussian-weighted average (std. 8 voxels, truncated at 2 std., with boundary normalization), yielding a 236-dimensional feature vector used as independently derived spatial gene-expression covariates.

Table A1: Identity feature groups for each dataset.

Feature Group	Description	WARP	Allen
Transcriptomics	41-dimensional gene-expression marker vector	✓	
Spatial position	3D registered coordinate in reference space	✓	✓
Anatomical region	Multihot vector over brain-region labels	✓	✓
HCR expression	236-dimensional atlas-derived gene-expression marker vector as documented in Appendix A.3	✓	
Electrophysiological features	Features from observed dynamics including mean, standard deviation, extrema, quartiles, positive fraction, mean squared value, and the standard deviation of first differences	✓	✓

A.4 Geometric Trajectory Metrics

We provide the details of the geometric metrics used to evaluate whether transferred latent trajectories preserve temporal organization and local trajectory structure on held-out recordings. Let $\{\mathbf{z}_t\}_{t=1}^{T}$, with $\mathbf{z}_t \in \mathbb{R}^d$, denote the latent trajectory for a held-out recording. Each metric is computed separately for each held-out trajectory and then averaged across held-out animals or sessions.

Temporal distance correlation. This metric measures whether distances in latent space increase monotonically with temporal separation. We sample $20{,}000$ frame pairs by drawing a frame index i and an offset $\Delta \in \{1, \ldots, \min(100, T-1)\}$, retaining valid pairs (i, j) with $j = i + \Delta \leq T$. We then compute the Spearman correlation between the latent distances $\|\mathbf{z}_i - \mathbf{z}_j\|_2$ and the temporal gaps $|i - j|$. Higher values indicate that temporal separation is more consistently reflected in the latent geometry.

Normalized latent-neighbor time gap. This metric measures whether local neighborhoods in latent space correspond to nearby time points. For each frame t, we find its $k = 10$ nearest neighbors in latent space, excluding the point itself, and compute the average temporal gap to these neighbors:

$$\frac{1}{Tk} \sum_{t=1}^{T} \sum_{j \in \mathcal{N}_{10}(t)} |t - j|.$$

We then divide this mean gap by T to report the metric on a normalized $[0, 1]$ scale. Lower values indicate that nearby latent states tend to correspond to nearby time points.

Local trajectory linearity. This metric quantifies the local one-dimensionality of short latent trajectory segments. We use sliding windows of length 25 frames with stride 5. For each window W, we center the latent states in the window, compute their covariance matrix, and let $\lambda_1 \geq \lambda_2 \geq \cdots$ denote its positive eigenvalues. The window-level linearity score is

$$\ell(W) = \frac{\lambda_1}{\sum_r \lambda_r}.$$

The reported local trajectory linearity is obtained by averaging $\ell(W)$ over all sliding windows. Larger values indicate that local trajectory variance is concentrated along a single dominant principal direction.

A.5 Implementation Details for the Proposed Method

All experiments use the identity-guided population method described in Sec. 2.1. At each evaluation time point, neuronal activity and neuronal identity features are provided as input. Routing weights are computed only from neuronal identity features, and neuron-level activity representations are aggregated into latent groups using the resulting identity-guided routing weights. The group representations are processed by the group interaction module Γ. Here, Γ is implemented as an MLP mixer over the latent group axis, whose output dimension is set by the corresponding objective, following the common contrastive manifold learning setup of CEBRA [3].

The same architecture is used for supervised decoding and manifold learning, with task-specific output dimensions, training objectives, and hyperparameters (Tables A2-A3). For behavior decoding, Γ outputs one scalar. For stimulus decoding, Γ outputs a vector whose dimension is the number of stimulus classes. For manifold learning, Γ outputs a three-dimensional latent vector $\mathbf{z}_t \in \mathbb{R}^3$, which is ℓ_2-normalized before applying the contrastive objective.

When a recording contains more neurons than the specified neurons-per-subset value in Tables A2–A3, we apply the model to uniformly sampled neuron subsets. During training, one subset is sampled per minibatch. During evaluation, we use a fixed set of sampled subsets and aggregate their outputs: predictions are averaged for behavior regression, class predictions are combined by majority vote for stimulus classification, and latent states are averaged for manifold analysis. All reported metrics are computed from these aggregated outputs. The same subsampling and evaluation ensembling protocol is used for all decoding baseline methods.

Table A2: Hyperparameters for supervised decoding with the proposed method.

Parameter	Value
Latent groups K	256
Activity encoder hidden size	128
Group activity dimension	384
Identity hidden size	192
Group mixer hidden size	64
Routing temperature	0.35
Routing sparsity	Top-5 soft assignment
Batch size	64
Optimizer	AdamW
Learning rate	2×10^{-4}
Weight decay	0
Maximum training epochs	180
WARP neurons per subset	3000 train / 3000 eval
Allen neurons per subset	250 train / 250 eval
Evaluation ensemble subsets	21

Neural decoding. Neural decoding follows the held-out transfer protocol in Section 3.3.

For WARP, each method is trained on source fish and evaluated on the held-out fish. For the Allen datasets, the model is trained on source sessions and evaluated on held-out sessions. For each Allen dataset, we randomly select the source and held-out sessions once and use the same fixed split for all methods. Model parameters are fixed during held-out evaluation, and no held-out behavioral or stimulus labels are used for training, fine-tuning, or model selection.

For behavior decoding, the model predicts the standardized target and is trained with the Huber loss

$$\mathcal{L}_{\text{beh}} = \frac{1}{|\mathcal{B}|} \sum_{t \in \mathcal{B}} h_\delta(\hat{y}_t - y_t), \qquad h_\delta(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta, \\ \delta\left(|e| - \frac{1}{2}\delta\right), & |e| > \delta, \end{cases}$$

with $\delta = 1.0$. Here, $\mathcal{B}$ denotes the training batch, $\hat{y}_t$ is the predicted standardized behavioral target, y_t is the standardized ground-truth target, and $e = \hat{y}_t - y_t$ is the prediction error. For stimulus decoding, the model is trained with the standard cross-entropy loss over stimulus classes. We use AdamW [41] and early stopping based on the validation score, with exponential moving average smoothing coefficient $\alpha = 0.5$. Hyperparameters are listed in Table A2.

Manifold learning. For manifold learning, the model is trained on source recordings using a time-contrastive objective following the CEBRA implementation [3]. For each anchor time t, a positive time point $t + \Delta$ is chosen within a fixed temporal offset and negative time points $\mathcal{N}_t$ are drawn from other time points in the training batch. Let $\mathbf{z}_t$ denote the normalized latent embedding of the population state at time t, and let $s(\cdot, \cdot)$ denote cosine similarity. We use the CEBRA-style generalized InfoNCE objective, in which the positive pair contributes an attraction term and the log-partition term is estimated over negative samples:

$$\mathcal{L}_{\text{manifold}} = -s(\mathbf{z}_t, \mathbf{z}_{t+\Delta})/\tau_{nce} + \log \sum_{u \in \mathcal{N}_t} \exp(s(\mathbf{z}_t, \mathbf{z}_u)/\tau_{nce})$$

where τ_{nce} is the contrastive temperature and $\mathcal{N}_t$ denotes the negative set only. This follows the CEBRA convention in which the positive sample is not included in the log-sum-exp term. After training on the source fish, the learned mapping is applied directly to the held-out fish without retraining or neuron-wise alignment. Hyperparameters are listed in Table A3.

A.6 Baseline methods for decoding

We evaluated DeepSets [18], POYO [9], and NEDS [19]. We evaluated a DeepSets baseline using the standard permutation-invariant form $\rho(\sum_i \phi(\mathbf{x}_i))$, where each neuron is treated as one set element. Here, $\mathbf{x}_i$ denotes the input feature vector for neuron i, ϕ is a shared element encoder, and ρ is a set-level MLP applied after summation. The input $\mathbf{x}_i$ contains the neuron's activity segment, with the neuronal identity vector concatenated when identity features are used. The element MLP ϕ produces

Table A3: Hyperparameters for manifold learning with the proposed method.

Parameter	Value
Output dimension	3
Latent groups K	32
Activity encoder hidden size	16
Group activity dimension	32
Identity hidden size	16
Group mixer hidden size	32
Routing temperature	0.35
Routing sparsity	Top-4 soft assignment
Contrastive similarity	Cosine
Contrastive temperature τ_{nce}	1.0
Positive temporal offset Δ	5 frames
Batch size	128
Optimizer	AdamW
Learning rate	1×10^{-3}
Weight decay	0
Training epochs	3000
WARP neurons per subset	3000 train / 3000 eval
Allen neurons per subset	250 train / 250 eval
Evaluation ensemble subsets	11

neuron-level embeddings, which are summed across neurons and passed through ρ to produce the decoding representation. We used the same temporal input construction and task outputs as the other supervised decoding baselines.

POYO [9] and NEDS [19] were originally developed for spike-based neural recordings. In this work, we evaluate these baselines on datasets with different signal modalities, including calcium signals in the WARP dataset and spiking activity in the Allen datasets. For each dataset, we adapt only the input representation to match the signal modality, while preserving the original architectural design of each model as much as possible.

Calcium-based Input for the WARP dataset. For NEDS, we extend the input stage to process continuous calcium traces. Calcium activity is first divided into fixed-length temporal windows, yielding a neuron-by-time representation within each window. Each neuron's calcium signal is embedded into a latent space and concatenated with the corresponding neuron identity embedding. Since this representation includes both neuron and temporal dimensions, we use an attention-based adapter to aggregate neuron-level representations into temporally indexed tokens aligned with the input structure of NEDS. The resulting tokens are then passed to the original NEDS transformer encoder and task-specific decoding heads, yielding a calcium-adapted NEDS.

For POYO, we use a POYO+ [30] style calcium input representation. Specifically, neuron identity embeddings are added to the original unit embeddings, and calcium value embeddings at each time point are added to these representations. This allows POYO to process continuous calcium observations while retaining its token-based architecture, resulting in a calcium-adapted POYO.

Spike-based Input for the Allen Datasets. The Allen datasets consist of spike recordings, and thus both POYO and NEDS are used in their original spike-based formulations without modifying the input representation.

To incorporate neuronal identity information, we extend each model in a manner consistent with its original design. For NEDS, neuron identity embeddings are combined with neural activity embeddings before being processed by the model. For POYO, identity embeddings are included with the unit embeddings from the original implementation. These unit embeddings were initialized following the official implementation. These modifications allow both models to incorporate neuronal identity information while preserving their original architectural assumptions.

A.7 Hyperparameter Settings for Decoding Baselines

POYO is implemented as a transformer with cross-attention and self-attention, using a token dimension of 64, 6 layers, 2 cross-attention heads, and 8 self-attention heads. NEDS is implemented as a transformer encoder with a hidden size of 128, 3 layers, and 4 attention heads. In both models, neuronal identity features are encoded as 32-dimensional embeddings produced by an MLP comprising linear layers, an activation function, and layer normalization.

For training, the batch size is set to 64 across all experiments. POYO uses a learning rate of 3.125×10^{-5} for both WARP and Allen datasets, while NEDS uses 3×10^{-4} for WARP and 3×10^{-5} for Allen.

A.8 Manifold Baseline Implementation Details

We evaluated CEBRA, MARBLE, joint PCA, and LFADS under the same held-out transfer protocol defined in Section 3.3. Unless otherwise stated, all baseline embeddings were three-dimensional and were evaluated on the same held-out animals or sessions as the proposed model.

For CEBRA, we used the official CEBRA Python implementation. We trained a time-conditional CEBRA encoder on each source recording and applied the trained encoder directly to the held-out recording without refitting. We used the `offset1-model` architecture with `time_offsets=1`, output dimension 3, learning rate 3×10^{-4}, and temperature 1.0. WARP transfer runs used batch size 512 and 10,000 optimization iterations; Allen runs used batch size 256 and 2,000 iterations.

For MARBLE, we used the public MARBLE implementation. Neural activity was first normalized and projected to 10 principal components. We used the following setup: $k = 60$, 64 eigenvectors, and `inner_product_features=False`. MARBLE was trained for 160 epochs on the source recording. Held-out activity was transformed using the source PCA basis and the trained source MARBLE checkpoint.

For joint PCA, we concatenated source activity along time, fit PCA on the source recordings only, and applied the resulting projection to held-out activity. LFADS baselines used precomputed LFADS latent trajectories from the corresponding LFADS training environment; no LFADS parameters were refit during manifold comparison. For source-encoder baselines on Allen, metrics were averaged over source-session encoders.

A.9 Manifold visualizations

We additionally visualize latent trajectories from source and held-out fish across methods, colored by behavior (Fig. A4) and stimulus (Fig. A5) labels.

A.10 Effect of Neuronal Identity on Decoding Baselines

To evaluate the contribution of neuronal identity, we incorporate multimodal biological features into baseline models. Details of the implementation and input construction are provided in Appendix A.6. For each baseline, we compare the original model with a variant that incorporates per-neuron identity features. Training is performed only on source data, and models are evaluated on held-out recordings without any fine-tuning.

Baselines with identity features are distinguished based on the composition of multimodal biological features used. For the WARP dataset, we utilize transcriptomic profile (T), spatial position (S), HCR gene expression marker (H), anatomical region (R), and electrophysiological features (P), denoted as POYO + (T, S, H, R, P) and NEDS + (T, S, H, R, P). In contrast, for the Allen datasets, we use spatial position (S), anatomical region (R), and electrophysiological features (P), denoted as POYO + (S, R, P) and NEDS + (S, R, P).

As shown in Table A4, incorporating neuronal identity generally leads to improved baseline performance across datasets. Notably, performance gains are observed for NEDS and POYO even under the held-out transfer setting. This suggests that neuronal identity provides information that is useful for transfer without requiring retraining or neuron-wise correspondence.

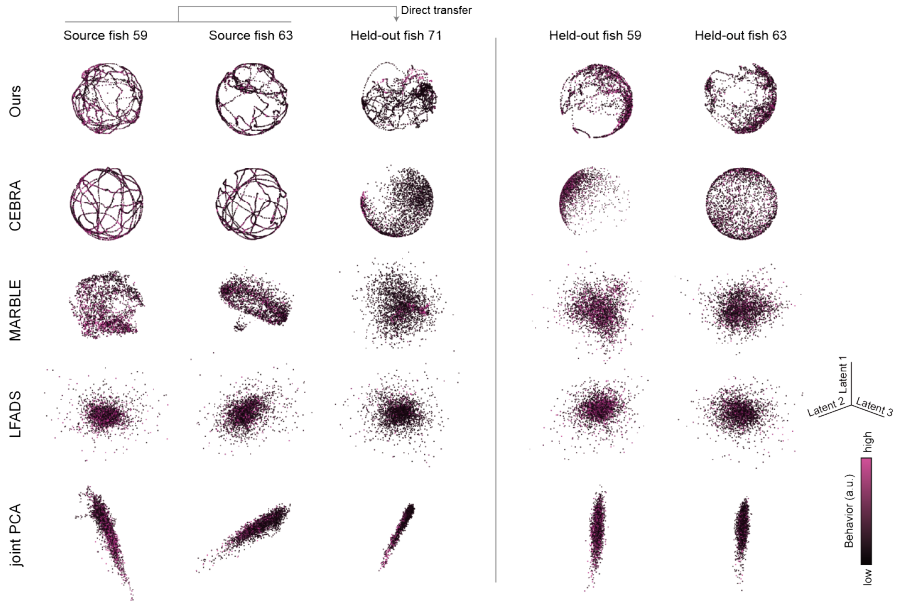

Figure A4: Latent trajectories for two source fish (Fish 59 and Fish 63) and held-out fish across all methods. Points are colored by behavior label within each recording.

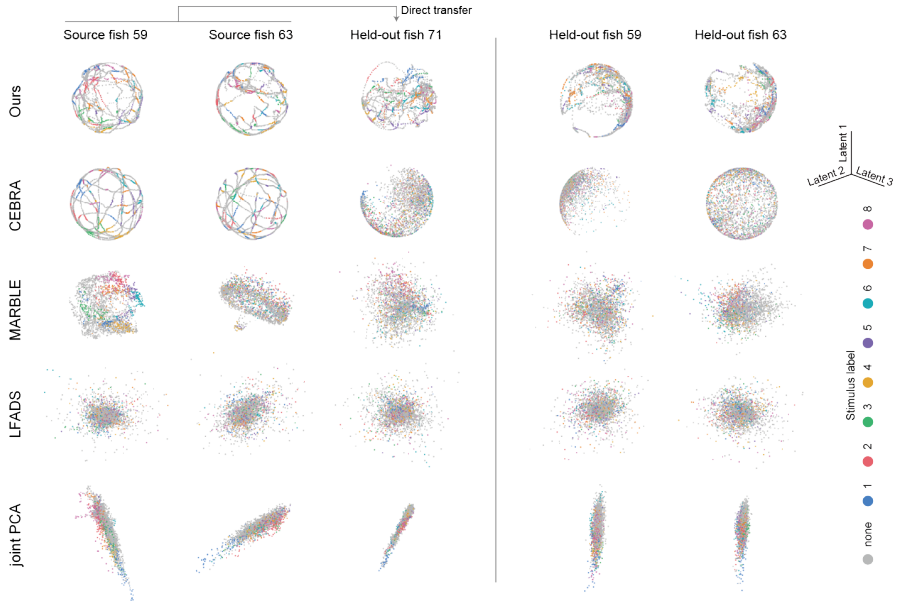

Figure A5: Latent trajectories for two source fish (Fish 59 and Fish 63) and held-out fish across all methods. Points are colored by stimulus label within each recording.

A.11 Ablation Studies

We conduct ablations to separate the contribution of identity information from the contribution of the group-routing architecture. Throughout this section, identity modalities are denoted by T for transcriptomic profile, S for spatial position, H for atlas-derived HCR gene markers, R for anatomical region, and P for electrophysiological features.

Identity-modality ablations. Table A5 reports WARP modality ablations, removing one modality at a time from the complete neuronal descriptor set (T, S, H, R, P). Across ablations, performance

Table A4: Behavior and stimulus decoding under direct held-out transfer across three datasets (mean $\pm$ std). * denotes models adapted to calcium signals.

Dataset	Method	Behavior (PCC) ↑	Stimulus (Balanced accuracy) ↑
WARP	POYO* + (T,S,H,R,P)	0.489 ± 0.096	0.164 ± 0.042
	POYO* [9]	0.304 ± 0.120	0.153 ± 0.032
	NEDS* + (T,S,H,R,P)	0.574 ± 0.108	0.374 ± 0.074
	NEDS* [19]	0.248 ± 0.207	0.195 ± 0.034
	DeepSets + (T,S,H,R,P)	0.466 ± 0.051	0.163 ± 0.023
	DeepSets [18]	0.160 ± 0.081	0.161 ± 0.011
Allen Brain Observatory	POYO + (S,R,P)	0.695 ± 0.003	0.439 ± 0.051
	POYO [9]	0.526 ± 0.043	0.220 ± 0.022
	NEDS + (S,R,P)	0.498 ± 0.055	0.475 ± 0.025
	NEDS [19]	0.500 ± 0.042	0.424 ± 0.043
	DeepSets + (S,R,P)	0.623 ± 0.060	0.727 ± 0.063
	DeepSets [18]	0.615 ± 0.022	0.684 ± 0.027
Allen Functional Connectivity	POYO + (S,R,P)	0.622 ± 0.046	0.494 ± 0.068
	POYO [9]	0.554 ± 0.080	0.211 ± 0.008
	NEDS + (S,R,P)	0.470 ± 0.029	0.449 ± 0.043
	NEDS [19]	0.454 ± 0.036	0.398 ± 0.020
	DeepSets + (S,R,P)	0.565 ± 0.089	0.708 ± 0.065
	DeepSets [18]	0.588 ± 0.096	0.730 ± 0.044

remains comparable to the all-modality configuration for most removals, indicating that the framework uses complementary identity features without depending on a single descriptor group.

Table A5: Modality ablation on WARP (mean $\pm$ std over 5 seeds). Each row removes one modality from the full identity set. We denote static identity features by T for transcriptomic profile, S for spatial position, H for atlas-derived HCR gene markers, and R for anatomical region. We denote the electrophysiological features by P.

Method	Behavior (PCC) ↑	Stimulus (Balanced accuracy) ↑
Ours (T,S,H,R,P)	0.793 ± 0.015	0.835 ± 0.029
Ours (S,H,R,P), w/o T	0.727 ± 0.013	0.711 ± 0.016
Ours (T,H,R,P), w/o S	0.721 ± 0.014	0.573 ± 0.010
Ours (T,S,R,P), w/o H	0.717 ± 0.029	0.693 ± 0.032
Ours (T,S,H,P), w/o R	0.716 ± 0.020	0.682 ± 0.030
Ours (T,S,H,R), w/o P	0.704 ± 0.026	0.722 ± 0.025

Table A6 reports the corresponding Allen Brain Observatory modality ablations using the available identity set (S, R, P). Both behavior decoding and stimulus classification remain effective under single-modality removals in this dataset. We interpret the Allen ablation conservatively: the compact (S, R, P) identity set is sufficient for transfer when molecular identity features are unavailable, while no single modality alone explains the full behavior-decoding performance.

Table A6: Modality ablation on the Allen Brain Observatory dataset (mean $\pm$ std over 3 seeds). Each row removes one modality from the full identity set. Here S denotes spatial position, R anatomical region, and P electrophysiology features.

Method	Behavior (PCC) ↑	Stimulus (Balanced accuracy) ↑
Ours (S,R,P)	0.734 ± 0.008	0.893 ± 0.012
Ours (R,P), w/o S	0.686 ± 0.037	0.786 ± 0.066
Ours (S,P), w/o R	0.711 ± 0.007	0.800 ± 0.040
Ours (S,R), w/o P	0.696 ± 0.004	0.787 ± 0.007

Routing-design ablations. Table A7 isolates an architectural choice in the group-routing framework on behavior decoding. Replacing soft routing with hard top-1 assignment reduces performance,

consistent with the need to represent soft identity-to-group assignments rather than forcing each neuron into a single discrete group.

Table A7: Architectural ablations on behavior decoding (mean ± std over 5 seeds for WARP and 3 seeds for Allen). Each row varies one design choice from the full model. Bold indicates the best result within each dataset.

Dataset	Configuration	Changed component	Behavior (PCC) ↑
WARP	**Full model**	Soft routing	$\mathbf{0.793 \pm 0.015}$
	Hard top-1 routing	Soft routing removed	0.708 ± 0.022
Allen Brain Observatory	**Full model**	Soft routing	$\mathbf{0.734 \pm 0.008}$
	Hard top-1 routing	Soft routing removed	0.678 ± 0.003
Allen Functional Connectivity	**Full model**	Soft routing	$\mathbf{0.694 \pm 0.058}$
	Hard top-1 routing	Soft routing removed	0.626 ± 0.078

A.12 Best-Ordering Analysis for Manifold Baselines

Several manifold baselines operate on fixed input coordinates: their learned encoders, projection matrices, or latent mappings are tied to the neuron indices used during source training. In held-out transfer across animals, the target recording contains a different set of neurons, and no neuron-wise correspondence to the source recording is available. As a result, applying these baselines to a held-out animal requires assigning an arbitrary order to the target neurons.

To make this comparison more favorable to the index-dependent baselines, we do not rely on a single target-neuron ordering. Instead, for each held-out recording and each baseline, we sample 1000 random permutations of the target neuron order, apply the source-trained baseline under each ordering, and recompute the three geometric trajectory metrics. We then examine the full distribution of metric values as well as the best value obtained across sampled orderings.

Figures A6–A8 show the resulting distributions. Across baselines and metrics, our method generally outperforms the best sampled ordering. This indicates that the observed gap is not due to choosing a single unfavorable target ordering for the baselines. Rather, the index-dependent baselines lack a mechanism for mapping an unaligned held-out population into a stable shared coordinate system.

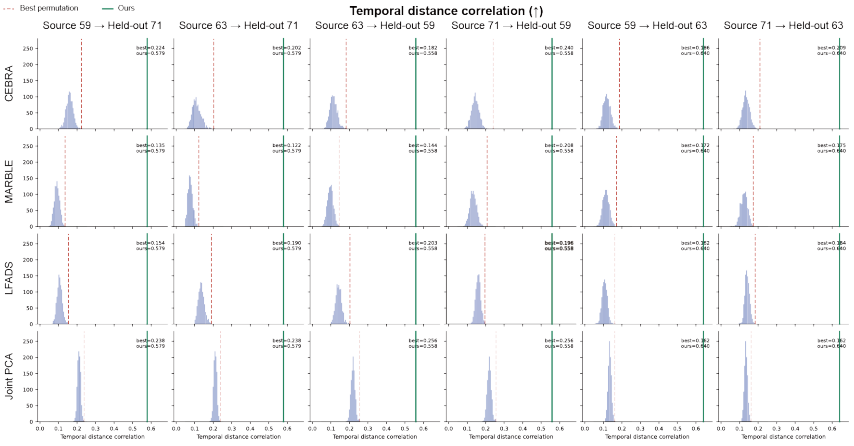

Figure A6: Target-ordering analysis for temporal distance correlation. Each histogram shows baseline performance over 1000 sampled target-neuron orderings. The vertical green line indicates our method and the vertical red dash indicates the best performance among the sampled orderings.

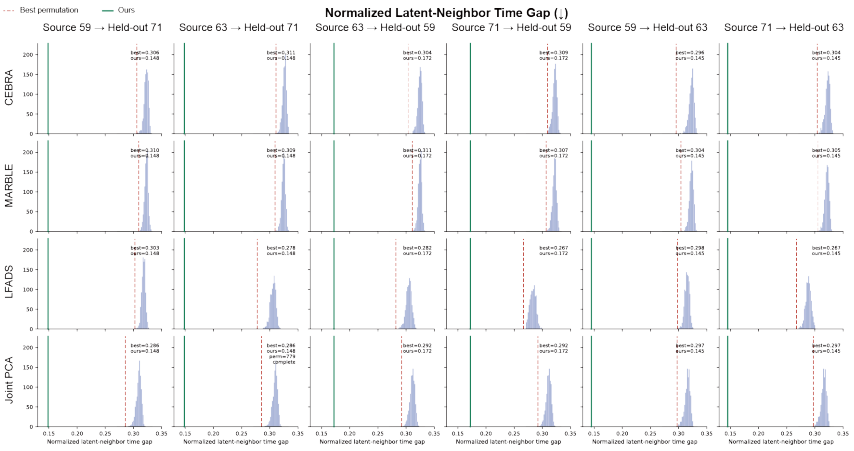

Figure A7: Target-ordering analysis for normalized latent-neighbor time gap. Each histogram shows baseline performance over 1000 sampled target-neuron orderings. The vertical green line indicates our method and the vertical red dash indicates the best performance among the sampled orderings.

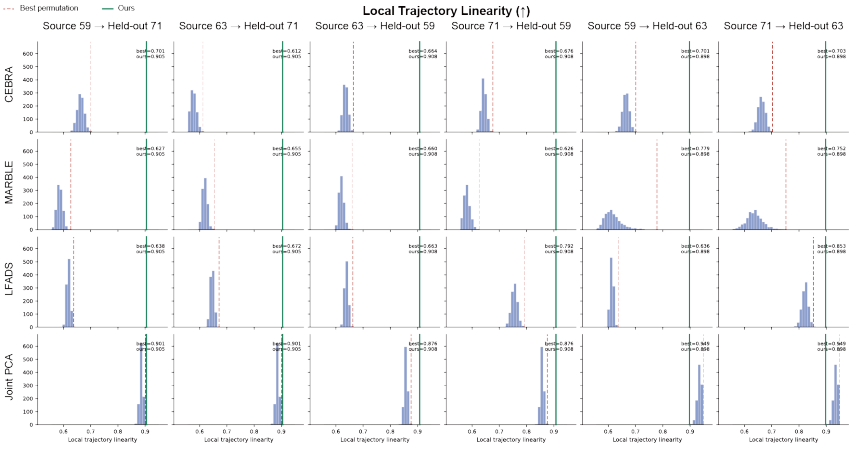

Figure A8: Target-ordering analysis for local trajectory linearity. Each histogram shows baseline performance over 1000 sampled target-neuron orderings. The vertical green line indicates our method and the vertical red dash indicates the best performance among the sampled orderings.

A.13 Compute Resources

All experiments were run on a Linux workstation with two Intel Xeon Silver 4214R CPUs, 440 GB system RAM, and three NVIDIA GeForce RTX 3090 GPUs with 24 GB memory per GPU. Each training or evaluation job used a single GPU. All reported experiments fit within the memory of a single RTX 3090 GPU.

Table A8 reports the wall-clock time required to reproduce the main reported experiments using our run configurations.

Table A8: Compute required to reproduce the reported experiments. GPU-hours are estimated as wall-clock time multiplied by the number of GPUs used by the launcher.

Experiment group	GPUs used	Wall-clock time	Approx. GPU-hours
WARP supervised decoding experiments	3 RTX 3090 GPUs	1.5 h	4.6
Allen supervised decoding experiments	3 RTX 3090 GPUs	7.8 h	23.3
WARP latent trajectory experiments	3 RTX 3090 GPUs	0.5 h	1.5
Allen latent trajectory experiments	3 RTX 3090 GPUs	6.4 h	19.2
Total for reported experiments	–	–	≈ 49 GPU-hours

The reported compute covers the runs used for the main decoding and latent trajectory tables, including the proposed method, ablations, and evaluated baselines. Additional exploratory runs used during development are not required to reproduce the reported results and are not included in Table A8.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction claim that multimodal per-neuron descriptors can provide an organizing basis for modeling neural population dynamics across animals and sessions without neuron-wise correspondence. The experiments support this claim by showing improved held-out behavior decoding, stimulus decoding, and latent trajectory structure compared with the evaluated baselines on WARP and Allen datasets.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The discussion notes limitations related to descriptor availability, atlas-registration noise, and variability across animals and sessions.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Appendix A.2 states the assumptions for permutation invariance and provides the full proof. The result applies when the identity encoder, activity encoder, and routing map are shared across neurons.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide the code as supplementary material. Sections 3 and Appendix A.1–A.9 describe the datasets, feature construction, train/test protocols, objectives, baseline adaptations, and hyperparameters. These details provide the information needed to reproduce the reported experiments.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide anonymized code and instructions for reproducing the proposed method, feature construction, training, evaluation, and baseline adaptations. The experiments use existing datasets cited in the paper.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 3 describes the datasets, evaluation protocol, metrics, and baselines. Appendix A provides feature construction, optimization details, hyperparameters, and baseline adaptation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The main tables report mean $\pm$ standard deviation across repeated runs.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix A.13 reports the compute resources used for the experiments, including CPU, GPU, memory, runtime, and approximate GPU-hour estimates.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses existing non-human animal neuroscience datasets and does not involve human subjects, crowdsourcing, or personally identifiable information. The research is intended for scientific analysis of neural population dynamics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: The paper does not include a dedicated broader impacts discussion. The work is foundational research on non-human animal neural recordings, with potential benefits for neuroscience data integration and limited immediate societal risk.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release a high-risk dual-use model, generative system, or scraped dataset. The method is evaluated on existing non-human animal neural recordings.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites the existing datasets and baseline methods used in the experiments, and explicitly list all dataset and code licenses.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The released code is documented with setup instructions, training and evaluation commands, and descriptions of the main scripts. No new dataset or pretrained model is released.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing, user studies, or human-subject experiments. All experiments use existing non-human animal neural recording datasets.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve human-subject research. Human-subject IRB approval is therefore not applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: LLMs were used for manuscript writing, editing, and code-development assistance. They were not used as a scientific component of the proposed model, and all experiments, analyses, and reported results were reviewed and validated by the authors.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.