A Latent Neural ODE-VAE for Modeling Hippocampal Population Activity on Low-Dimensional Manifolds

Anonymous Author(s)
Affiliation
Address
email

Abstract

Neural activity can be represented as a trajectory in a high-dimensional space where each axis corresponds to one neuron. A central question in systems neuroscience is whether these trajectories are constrained to low-dimensional manifolds that jointly encode external variables and trial-to-trial internal variability. Building on this question, we develop and analyze a latent Neural ODE variational autoencoder (ODE-VAE) for calcium population recordings. The model encodes trial onset activity into a stochastic latent initial condition, evolves it through a continuous-time mixture-of-experts neural ODE, and decodes latent trajectories back to neural activity. Relative to earlier ODE-VAE versions, `v5` adds transition-aware reconstruction regularization and a soft locally linear embedding (LLE) latent constraint. Using repository-grounded experiments, we report strong reconstruction on synthetic random-foraging data (best seed $R^2 = 0.9789$) and mixed outcomes on the E65 hippocampal dataset, including an archived ODE-VAE run at $R^2 = 0.4368$, strict raw-space `v5` sweeps with best-epoch $R^2 \leq 0.0569$, and a no-PCA `v6` variant at $R^2 = 0.0956$. These results show that continuous-time latent manifold models can capture structured dynamics, but evaluation protocol, preprocessing, and regularization balance strongly affect performance on real neural data.

1 Introduction

Neural activity can be described as a point in a high-dimensional coordinate system, where each coordinate axis represents a single neuron's activity. Prior work shows that population trajectories often occupy low-dimensional manifolds, in motor cortex, hippocampus, and entorhinal circuits [Gallego et al., 2017, Russo et al., 2018, O'Keefe and Dostrovsky, 1971, Aronov et al., 2017, Nieh et al., 2021]. This geometric view supports the broader "cognitive map" hypothesis: neural circuits organize both physical and abstract knowledge into structured latent spaces [Tolman, 1948, O'Keefe and Nadel, 1978, Stachenfeld et al., 2017, Bellmund et al., 2018].

Recent manifold analyses of hippocampal recordings report intrinsic dimensionalities around 4–6 and show that manifold geometry can explain both task variables and trial-level variability [Low et al., 2018, Chaudhuri et al., 2019, Nieh et al., 2021]. However, many pipelines are multi-stage: fit local transition models, build pairwise distances, embed points, and separately map back to activity [Low et al., 2018, Tenenbaum et al., 2000, Yu et al., 2009]. This decomposition gives useful geometry but does not explicitly enforce globally smooth continuous-time dynamics in a single end-to-end objective.

This paper develops a continuous-time generative alternative tailored to the repository's ODE-VAE codebase. Starting from an initial latent ODE-VAE, we focus on `v5`, which adds (i) a mixture-of-experts (MoE) latent vector field, (ii) transition-aware reconstruction regularization, and (iii) soft LLE

regularization in latent space. We evaluate this family on synthetic and real hippocampal calcium data and document where the method works and where it fails.

Contributions.

- We present a mathematically explicit latent Neural ODE-VAE formulation for trial-based neural population sequences, including MoE latent dynamics and local-geometric regularizers.
- We instantiate this formulation using the repository's `v1`–`v6` implementations, with `v5` as the main model and `v6` as a no-PCA extension.
- We provide artifact-grounded empirical analysis on synthetic and E65 datasets, including seed sensitivity and configuration-dependent instability.
- We identify concrete protocol factors (split strategy, landmark subsampling, and metric space) that materially influence reported R^2.

2 Related Work

Our approach sits at the intersection of hippocampal manifold neuroscience and deep latent dynamical systems. In neuroscience, place cells and cognitive map theories motivate low-dimensional geometric structure in population activity [O'Keefe and Dostrovsky, 1971, O'Keefe and Nadel, 1978, Eichenbaum and Cohen, 2014]. More recent work extends these ideas to abstract and non-spatial variables [Constantinescu et al., 2016, Schuck and Niv, 2019, Park et al., 2020, Aronov et al., 2017, Nieh et al., 2021]. Population-level manifold analyses in hippocampus and other circuits show that latent geometry can explain behavior and internal state variability [Low et al., 2018, Chaudhuri et al., 2019, Gallego et al., 2017].

In machine learning, variational autoencoders [Kingma and Welling, 2014] and neural ODEs [Chen et al., 2018] enable continuous-time latent-variable models, including latent ODEs for irregular trajectories [Rubanova et al., 2019]. Our model adopts this framework but targets neuroscientific interpretability: trialized sequences, manifold visualizations, and explicit regularizers for smoothness, transition consistency, and local linearity.

3 Problem Setup and Data

Let $x_b(t_\ell) \in \mathbb{R}^N$ denote population activity for trial $b \in \{1, \ldots, B\}$ at resampled time index $\ell \in \{1, \ldots, L\}$.

E65 dataset. The repository's main real dataset is `E65_data.npz`, containing calcium activity ($N = 375$, $T = 7434$), trial IDs, timestamps, and aligned behavioral covariates. In the `v5` preprocessing path: (i) PCA retains 95% variance, producing $K = 129$ components; (ii) trials are resampled to $L = 120$ bins using `trial_len_s=12` and `fps=10`; (iii) first 10 trials are dropped, yielding 180 usable trials; (iv) default validation holds out 3 trials (train 177 / val 3).

Synthetic benchmark. The repository also includes `synthetic_rat_data.npz` (4000 frames, 300 neurons, 20 trials), used to test recoverability of smooth low-dimensional dynamics under controlled conditions.

4 Model: Latent Neural ODE-VAE

4.1 Stochastic encoder

For each trial, the encoder uses only $x_b(t_1)$ and outputs a diagonal Gaussian posterior on the latent initial state:

$$q_\phi(z_{0,b} \mid x_b(t_1)) = \mathcal{N}\left(\mu_b, \mathrm{diag}(\sigma_b^2)\right), \tag{1}$$

with reparameterization

$$z_{0,b} = \mu_b + \sigma_b \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{2}$$

2

4.2 Continuous-time latent dynamics

Latent trajectories are generated by a neural ODE:

$$\frac{dz_b(t)}{dt} = f_\theta(z_b(t), t), \qquad z_b(t_1) = z_{0,b}. \tag{3}$$

In v5, f_θ is a mixture of experts:

$$f_\theta(z, t) = \sum_{e=1}^{E} \pi_e(z) f_e(z), \qquad \pi(z) = \text{softmax}(g(z)), \tag{4}$$

with $E = 4$ latent experts by default and Dormand–Prince integration (dopri5).

4.3 Decoder family

A decoder maps latent states back to observations:

$$\hat{x}_b(t_\ell) = g_\psi(z_b(t_\ell)). \tag{5}$$

The codebase supports MLP, neuron-aware, local-attention, and MoE decoders; v5 default is MoE
decoder with 8 decoder experts.

5 Training Objective and Regularization

The base objective combines reconstruction and KL terms:

$$\mathcal{L}_{\text{base}} = \mathcal{L}_{\text{rec}} + \beta \, \mathcal{L}_{\text{KL}}, \tag{6}$$

where

$$\mathcal{L}_{\text{rec}} = \frac{1}{BLK} \sum_{b,\ell} \|\hat{x}_b(t_\ell) - x_b(t_\ell)\|_2^2, \tag{7}$$

$$\mathcal{L}_{\text{KL}} = \frac{1}{B} \sum_b D_{\text{KL}}\big(q_\phi(z_{0,b} \mid x_b(t_1)) \,\|\, \mathcal{N}(0, I)\big). \tag{8}$$

Smoothness regularization.

$$\mathcal{L}_{\text{smooth}} = \frac{1}{B(L-1)D} \sum_{b,\ell} \left\| \frac{z_b(t_{\ell+1}) - z_b(t_\ell)}{t_{\ell+1} - t_\ell} \right\|_2^2. \tag{9}$$

Transition-aware regularization (v5).

$$\mathcal{L}_{\text{trans}} = \frac{1}{B(L-1)K} \sum_{b,\ell} \left\| \big(\hat{x}_b(t_{\ell+1}) - \hat{x}_b(t_\ell)\big) - \big(x_b(t_{\ell+1}) - x_b(t_\ell)\big) \right\|_2^2. \tag{10}$$

This term is linearly warmed up for the first 30 epochs.

Soft LLE latent regularization (v5). For flattened latent points $\{z_i\}_{i=1}^M$, with k-NN set $\mathcal{N}_k(i)$:

$$\mathcal{L}_{\text{LLE}} = \frac{1}{M} \sum_{i=1}^{M} \left\| z_i - \sum_{j \in \mathcal{N}_k(i)} w_{ij} z_j \right\|_2^2, \quad w_{ij} \propto \exp\left(-\frac{\|z_i - z_j\|_2}{\tau} \right). \tag{11}$$

Default parameters: $k = 8$, $M \leq 256$, $\tau = 0.1$.

Total loss.

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \beta_t \mathcal{L}_{\text{KL}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}} + \lambda_{\text{trans},t} \mathcal{L}_{\text{trans}} + \lambda_{\text{LLE}} \mathcal{L}_{\text{LLE}}. \tag{12}$$

The KL coefficient β_t is warmed up over 30 epochs to a final value $\beta = 0.02$.

Table 1: Synthetic random-foraging benchmark from `seed_sweep_results.txt`.

Seed	Final R^2	Best validation loss
1	0.9789	0.06642
42	0.6757	0.40521
1337	0.9116	0.16298
2025	0.7467	0.31145
777	0.0000 (NaN collapse)	∞
Mean (all seeds)	0.6626	–
Mean (non-collapsed seeds)	0.8282	–

Table 2: E65 run artifacts from saved metadata/checkpoints.

Artifact (path)	Variant summary	Best reported R^2	Final rep
`runs/ode_vae_E65/run_metadata.json`	archived ODE-VAE	–	0.43
`runs/ode_vae_E65/ld5_b0.02_ls0.0005/run_metadata.json`	v5, $d = 5$	0.0354	-0.23
`runs/ode_vae_E65/ld5_b0.02_ls0.0002/run_metadata.json`	v5, $d = 5$	0.0353	-0.13
`runs/ode_vae_E65/ld8_b0.02_ls0.0005/run_metadata.json`	v5, $d = 8$	0.0569	-0.20
`src/pt_files/final_metrics.pt`	v6, no PCA	–	0.09

6 Experimental Protocol

6.1 Configurations

Main v5 settings from `config.txt`: latent dimension 5 (with sweep to 8), batch size 8, 150 epochs, Adam optimizer (learning rate 0.002, weight decay 10^{-5}), $\lambda_{\text{smooth}} \in \{5 \times 10^{-4}, 2 \times 10^{-4}\}$, $\lambda_{\text{trans}} = 0.01$, $\lambda_{\text{LLE}} = 0.01$, landmark count 100, and baseline correction enabled.

6.2 Metrics

The primary training metric is coefficient of determination,

$$R^2 = 1 - \frac{\sum (X - \hat{X})^2}{\sum (X - \bar{X})^2}. \tag{13}$$

The repository contains two evaluation styles: (i) PCA-space R^2 and (ii) strict raw-neuron-space R^2 via inverse PCA and de-normalization. We report artifact values as stored in each run metadata file.

7 Results

7.1 Synthetic benchmark: high ceiling with seed sensitivity

Table 1 summarizes the repository seed sweep on synthetic data. Best seed reaches $R^2 = 0.9789$, but one seed collapses to NaN/0.0, indicating optimizer sensitivity.

7.2 E65 hippocampal data: mixed performance across variants

Table 2 reports all E65 artifact metrics found in the repository for the ODE-VAE family. An archived ODE-VAE run reports $R^2 = 0.4368$. Strict raw-space v5 sweep artifacts are substantially lower (best-epoch $R^2 \leq 0.0569$, negative final R^2 in three runs), while v6 (no PCA) records $R^2 = 0.0956$.

7.3 Manifold interpretability

The codebase saves latent manifold projections (MDS) and reconstruction diagnostics for each run. Figure 1 shows an example latent trajectory embedding from the trained model artifacts.
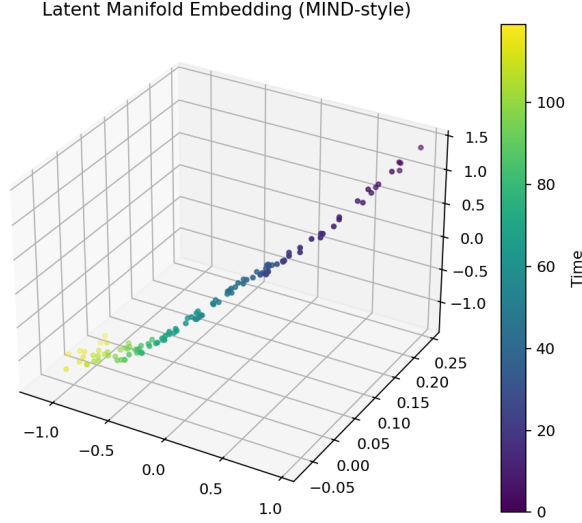
Figure 1: Latent manifold embedding produced by the ODE-VAE analysis pipeline.

8 Discussion

The model captures the intended inductive bias: low-dimensional continuous latent trajectories with explicit geometric regularization. On synthetic data, this bias is highly effective. On real E65 recordings, however, results are sensitive to implementation and evaluation choices.

Three factors emerge from the repository artifacts:

1. **Metric-space mismatch.** PCA-space training can look favorable while strict raw-space R^2 may degrade.

2. **Data-efficiency tradeoff.** Landmark subsampling (100 selected sequences from 180 usable trials) accelerates training but may reduce generalization.

3. **Optimization stability.** Strong regularization with small validation sets (3 trials) and stiff latent dynamics can produce unstable or negative final R^2, despite early high points.

These observations suggest that future gains likely require protocol-level changes in addition to architectural changes: larger and randomized holdout splits, early stopping on a stable cross-validated objective, trial-level (not frame-level) landmark selection, and direct raw-space reconstruction losses.

9 Limitations and Reproducibility

This study is bounded by available run artifacts in the repository and inherits version-specific logging differences. In particular, some run files report "best" and "final" R^2 under different conditions, and not all checkpoints include identical metadata fields. We therefore report values exactly as saved in each artifact path. The codebase also indicates unresolved training fragility (including occasional NaN collapse), which should be addressed before definitive biological claims.

10 Conclusion

We presented a mathematically grounded latent Neural ODE-VAE framework for neural manifold modeling and analyzed the repository's `v1-v6` trajectory with `v5` as the primary model. The method can recover smooth low-dimensional dynamics and high synthetic reconstruction quality, but real-data performance remains sensitive to preprocessing and evaluation protocol. This work provides a formalized foundation and a clear set of engineering directions for turning ODE-VAE manifold modeling into a robust neuroscience analysis tool.

Societal impact. This work is basic research on neural representation learning from animal neuroscience data and has no immediate direct societal deployment. Potential long-term impact is improved scientific understanding of memory and cognition.

References

Dmitriy Aronov, Rachel Nevers, and David W Tank. Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature*, 543:719–722, 2017.

Jacob L S Bellmund, Peter Gardenfors, Edvard I Moser, and Christian F Doeller. Navigating cognition: Spatial codes for human thinking. *Science*, 362:eaat6766, 2018.

Rishidev Chaudhuri, Burak Gercek, Bikash Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22:1512–1520, 2019.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *NeurIPS*, 2018.

Alexandra O Constantinescu, Jill X O'Reilly, and Timothy EJ Behrens. Organizing conceptual knowledge in humans with a grid-like code. *Science*, 352:1464–1468, 2016.

Howard Eichenbaum and Neal J Cohen. Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83:764–770, 2014.

Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. Neural manifolds for the control of movement. *Neuron*, 94:978–984, 2017.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *ICLR*, 2014.

Ryan J Low, Sean Lewallen, Dmitriy Aronov, Rachel Nevers, and David W Tank. Probing variability in a cognitive map using manifold inference from neural dynamics. *bioRxiv*, 2018.

Christopher J MacDonald, Kyle Q Lepage, Uri T Eden, and Howard Eichenbaum. Hippocampal time cells bridge the gap in memory for discontiguous events. *Neuron*, 71:737–749, 2011.

Edward H Nieh et al. Geometry of abstract learned knowledge in the hippocampus. *Nature*, 595:80–84, 2021.

John O'Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map. *Brain Research*, 34:171–175, 1971.

John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map*. Clarendon Press, 1978.

Sang Ah Park, David S Miller, Hamed Nili, Charan Ranganath, and Erie D Boorman. Map making: Constructing, combining, and inferring on abstract cognitive maps. *Neuron*, 107:1226–1238.e8, 2020.

Eftychios A Pnevmatikakis et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 89:285–299, 2016.

Eftychios A Pnevmatikakis and Andrea Giovannucci. NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *Journal of Neuroscience Methods*, 291:83–94, 2017.

Stefano Recanatesi et al. Predictive learning as a network mechanism for extracting low-dimensional latent space representations. *Nature Communications*, 12:1417, 2021.

Yulia Rubanova, Ricky TQ Chen, and David Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. *NeurIPS*, 2019.

Abigail A Russo et al. Motor cortex embeds muscle-like commands in an untangled population response. *Neuron*, 97:953–966.e8, 2018.

Nicolas W Schuck and Yael Niv. Sequential replay of nonspatial task states in the human hippocampus. *Science*, 364:eaaw5181, 2019.

Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hippocampus as a predictive map. *Nature Neuroscience*, 20:1643–1653, 2017.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

Edward C Tolman. Cognitive maps in rats and men. *Psychological Review*, 55:189–208, 1948.

Byron M Yu et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102:614–635, 2009.