Extraction and Recovery of Spatio-Temporal Structure in Latent Dynamics Alignment with Diffusion Model (ERDiff)

NeurIPS'23 Spotlight

Authors: Yule Wang, Zijing Wu,
Chengrui Li, Anqi Wu

Presenter: Yule Wang, ML@GT
Advisor: Prof. Anqi Wu

10/19/2023

Georgia Tech

Problem Formulation:
Neural Distribution Alignment

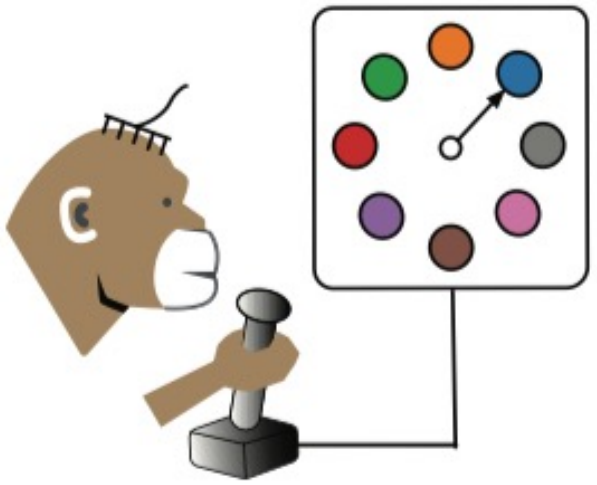
Georgia Tech

Neural Distribution Alignment

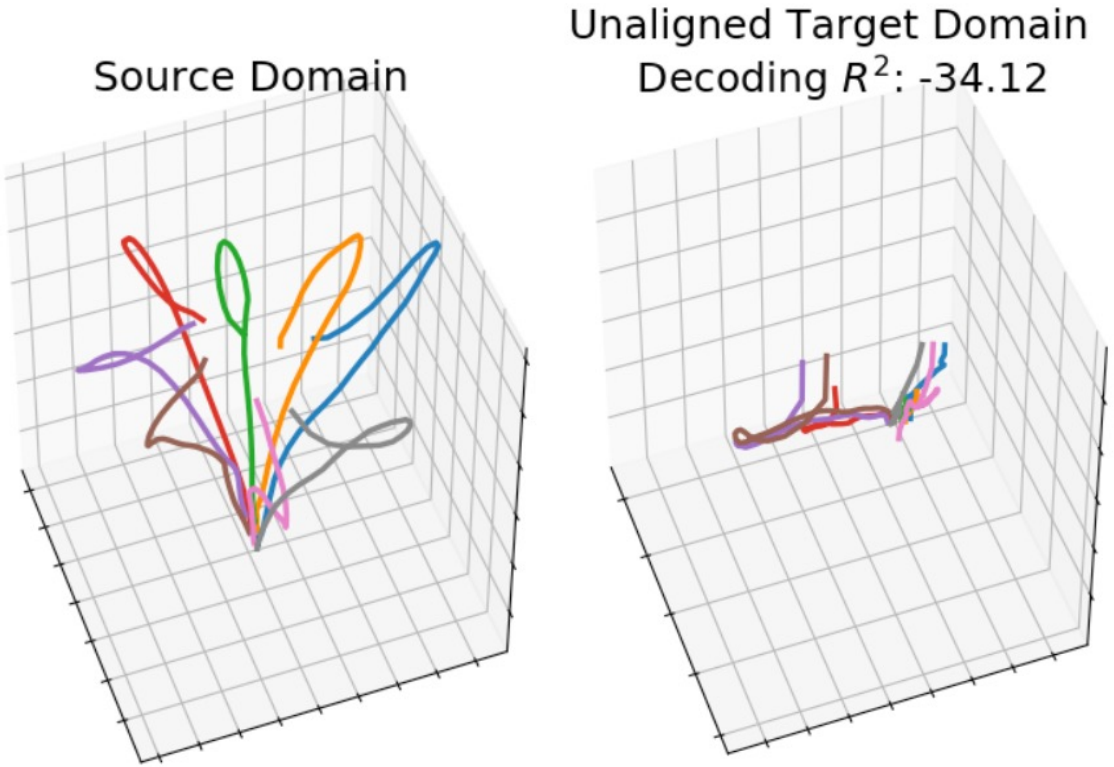
Problem to Solve:

- Perform alignment against the drastic distribution shift between raw neural population activities recorded across-day and inter-subject (individual).

- Task Setting Example



- Neural Latent Dynamics of Single Trials



Georgia Tech

Neural Distribution Alignment

Problem Formulation:

Distribution Alignment in low-dimensional latent space:

- Based on the Manifold hypothesis [1]: a relatively small number of latent dynamical factors contain a large portion of neural activities variability
- Perform alignment in the neural latent space

[1] Gallego, Juan A., et al. "Neural manifolds for the control of movement." Neuron 94.5 (2017): 978-984.

Georgia Tech

Neural Distribution Alignment

Problem Formulation:

Source-Domain

- We denote the source-domain single-trial neural population activities as $\mathbf{X}^{(s)} = \left[\mathbf{x}_1^{(s)}, \ldots, \mathbf{x}_l^{(s)}\right]^\top \in \mathbb{R}^{l \times n}$, where l is the trial length and n is the number of neurons. $\mathbf{Z}^{(s)} = \left[\mathbf{z}_1^{(s)}, \ldots, \mathbf{z}_l^{(s)}\right]^\top \in \mathbb{R}^{l \times d}$ are their low-dimensional latent dynamics inferred by a latent variable model (LVM), where d is the latent dimension size.

- $\boldsymbol{\phi}_s, \boldsymbol{\psi}_s = \underset{\boldsymbol{\phi},\boldsymbol{\psi}}{\operatorname{argmax}} \left[\mathbb{E}_{q\left(\mathbf{Z}^{(s)} \mid \mathbf{X}^{(s)}; \boldsymbol{\phi}\right)}\left[\log p\left(\mathbf{X}^{(s)} \mid \mathbf{Z}^{(s)}; \boldsymbol{\psi}\right)\right] - \mathbb{D}_{\mathrm{KL}}\left(q\left(\mathbf{Z}^{(s)} \mid \mathbf{X}^{(s)}; \boldsymbol{\phi}\right) \| \bar{q}\left(\mathbf{Z}^{(s)}\right)\right)\right]$

- We denote the probabilistic encoder and decoder of the LVM trained on source-domain as $q\left(\mathbf{Z}^{(s)} \mid \mathbf{X}^{(s)}; \boldsymbol{\phi}_s\right)$ and $p\left(\mathbf{X}^{(s)} \mid \mathbf{Z}^{(s)}, \boldsymbol{\psi}_s\right)$, respectively.

Georgia Tech

Neural Distribution Alignment

Problem Formulation:

Target-Domain

- Given the target-domain single-trial neural population activities as $\mathbf{X}^{(t)} = \left[\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_l^{(t)}\right]^\top \in \mathbb{R}^{l \times n}$, we perform distribution alignment by probing the probabilistic encoder $q\left(\mathbf{Z}^{(t)} \mid \mathbf{X}^{(t)}; \boldsymbol{\phi}\right)$

- The alignment is conducted by minimizing certain probability divergence $\mathbb{D}(\cdot|\cdot)$ between the two posterior distributions:

$$\min_{\phi_t} \mathbb{D}\left(q\left(\mathbf{Z}^{(s)} \mid \mathbf{X}^{(s)}; \boldsymbol{\phi}_s\right) \| \, q\left(\mathbf{Z}^{(t)} \mid \mathbf{X}^{(t)}; \boldsymbol{\phi}_t\right)\right)$$

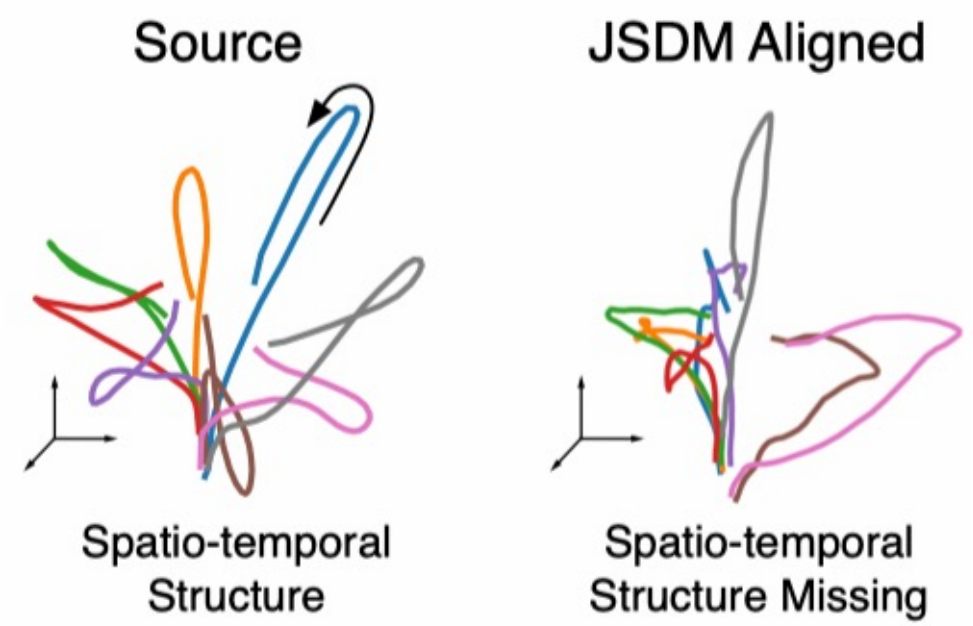
Georgia Tech

Neural Distribution Alignment

Challenges:

- The underlying spatio-temporal structures of neural latent dynamics are both non-linear and complex.

- Prior works ignoring such crucial spatio-temporal structure information, resulting in comparably inferior alignment performance.



Georgia Tech

Methodology of ERDiff

Georgia Tech

Neural Distribution Alignment

Key Insight:

- **E**xtraction and **R**ecovery of Spatio-Temporal Structure in Latent Dynamics Alignment with **Diff**usion Model **(ERDiff)**

- How to precisely extract the spatio-temporal structures of the source domain neural latent dynamics?

Strong Density Estimation Capability

- How to precisely recover such extracted spatio-temporal structures to the aligned target domain neural latent dynamics?

Proper Alignment Algorithm

Georgia Tech

Neural Distribution Alignment

Key Insight:

- How to precisely extract the spatio-temporal structures of the source domain neural latent dynamics?

Strong Density Estimation Capability

(Score-based) Diffusion Model

- How to precisely recover such extracted spatio-temporal structures to the aligned target domain neural latent dynamics?

Proper Alignment Algorithm

Georgia Tech

Methodology of ERDiff

Source-domain Diffusion Model (DM) Training:

- The LVM alone focuses on the point-to-point underlying mapping between $\mathbf{X}^{(s)}$ and $\boldsymbol{Z}^{(s)}$. The overall distribution of the source domain latent dynamics $p_s(\mathbf{Z})$, abbreviation for $q\left(\mathbf{Z}^{(s)} \mid \mathbf{X}^{(s)}; \boldsymbol{\phi}_s\right)$, is still inaccessible by building a LVM alone

- We propose to learn $p_s(\mathbf{Z})$ via a Diffusion Model (DM) by taking the entire trials of latent dynamics on source-domain $\mathbf{Z}^{(s)} \sim q\left(\cdot \mid \mathbf{X}^{(s)}; \phi_s\right)$ as input to the DM for training

- Specifically, the DM fits $p_s(\mathbf{Z})$ through training of denoiser $\boldsymbol{\epsilon}(\mathbf{Z}, t; \boldsymbol{\theta}_s)$: $\left(\mathbb{R}^{l\times d} \times \mathbb{R}\right) \rightarrow \mathbb{R}^{l\times d}$ through the denoising score matching (DSM) loss:

- $$\boldsymbol{\theta}_s = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \mathbb{E}_{t\sim\mathcal{U}[0,T]} \mathbb{E}_{\mathbf{Z}_0^{(s)}\sim q\left(\cdot\mid\mathbf{X}^{(s)};\boldsymbol{\phi}_s\right),\boldsymbol{\epsilon}\sim\mathcal{N}\left(\mathbf{0},\mathbf{I}_{l\times d}\right)} \left[w(t)^2 \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}\left(\mathbf{Z}_t^{(s)}, t; \boldsymbol{\theta}\right) \right\|_2^2 \right]$$

  - where $\mathbf{Z}_t = \mu_t \mathbf{Z}_0 + \boldsymbol{K}_t \boldsymbol{\epsilon}$, and $\boldsymbol{\epsilon}(\mathbf{Z}_t, t; \boldsymbol{\theta}) = -\boldsymbol{K}_t^{-T} \boldsymbol{s}(\mathbf{Z}_t, t; \boldsymbol{\theta})$, $\boldsymbol{K}_t \boldsymbol{K}_t^T = \Sigma_t$

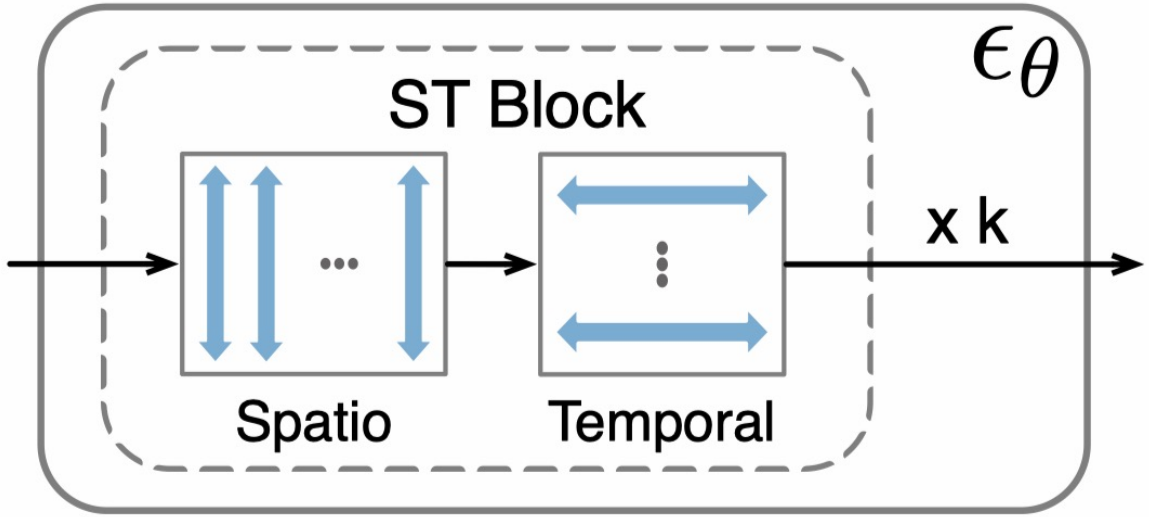
[1] Pascal Vincent. A connection between score matching and denoising autoencoders.

Georgia Tech

Methodology of ERDiff

Extract the spatio-temporal structures of the source domain:

- We use Spatio-Temporal Transformer Blocks (STBlocks)
- Each STBlock is composed of Spatio-Transformer layers followed by Temporal – Transformer layers
- Spatio-Transformer layer takes latent states of each time bin as inputs to extract spatial structure
- Temporal-Transformer layer takes the entire latent trajectory of each latent space dimension as inputs to extract temporal structure



Georgia Tech

Neural Distribution Alignment

Key Insight:

- How to precisely extract the spatio-temporal structures of the source domain neural latent dynamics?

Strong Density Estimation Capability

- How to precisely recover such extracted spatio-temporal structures to the aligned target domain neural latent dynamics?

Proper Alignment Algorithm

Maximum Likelihood Alignment with Diffusion Model

Georgia Tech

Methodology of ERDiff

Maximum Likelihood Alignment with Diffusion Model:

- Given the target-domain neural activities $\mathbf{X}^{(t)}$, we propose to perform distribution alignment via maximum likelihood estimation (MLE):

$$\underset{\boldsymbol{\phi}}{\operatorname{argmax}} \mathbb{E}_{\mathbf{X}\sim p(\mathbf{X}^{(t)})}[\log p_s(q(\mathbf{Z} \mid \mathbf{X}; \boldsymbol{\phi}))] = \underset{\boldsymbol{\phi}}{\operatorname{argmax}} \mathbb{E}_{\mathbf{Z}\sim q(\mathbf{Z}|\mathbf{X}^{(t)};\boldsymbol{\phi})}[\log p_s(\mathbf{Z})],$$

- We use the marginal distribution $p_0(\mathbf{Z}; \boldsymbol{\theta}_s)$ learnt by the DM to approximate $p_s(\mathbf{Z})$, the maximum likelihood estimation can thus be written as:

$$\underset{\boldsymbol{\phi}}{\operatorname{argmax}} \mathbb{E}_{\mathbf{Z}\sim q(\mathbf{Z}|\mathbf{X}^{(t)};\boldsymbol{\phi})}[\log p_0(\mathbf{Z}; \boldsymbol{\theta}_s)]$$

- We note that the alignment is conducted by probing the parameter set $\boldsymbol{\phi}$ of the probabilistic encoder while keeping the DM $p_0(\mathbf{Z}; \boldsymbol{\theta}_s)$ fixed

Georgia Tech

Methodology of ERDiff

Maximum Likelihood Alignment with Diffusion Model:

- Consequently, we could obtain an upper bound of the maximum likelihood loss function, as follows:

$$-\mathbb{E}_{\mathbf{Z}\sim q(\mathbf{Z}|\mathbf{X}^{(t)};\boldsymbol{\phi})}[\log\, p_0(\mathbf{Z};\boldsymbol{\theta}_s)] \leqslant \underbrace{\mathbb{D}_{\mathrm{KL}}(p_T(\mathbf{Z};\boldsymbol{\theta}_s)\parallel \pi(\mathbf{Z}))}_{\text{Constant Term}}$$

$$+\mathbb{E}_{t\sim\mathcal{U}[0,T]}\mathbb{E}_{\mathbf{Z}_0\sim q(\mathbf{Z}|\mathbf{X}^{(t)};\boldsymbol{\phi}),\boldsymbol{\epsilon}\sim\mathcal{N}(0,\boldsymbol{I}_{l\times d})}[\underbrace{w(t)^2\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}(\mathbf{Z}_t,t;\boldsymbol{\theta}_s)\|_2^2}_{\text{Denoising Score Matching}} - \underbrace{2\nabla_{\mathbf{Z}}\cdot\boldsymbol{f}(\mathbf{Z}_t,t)}_{\text{Divergence}}]$$

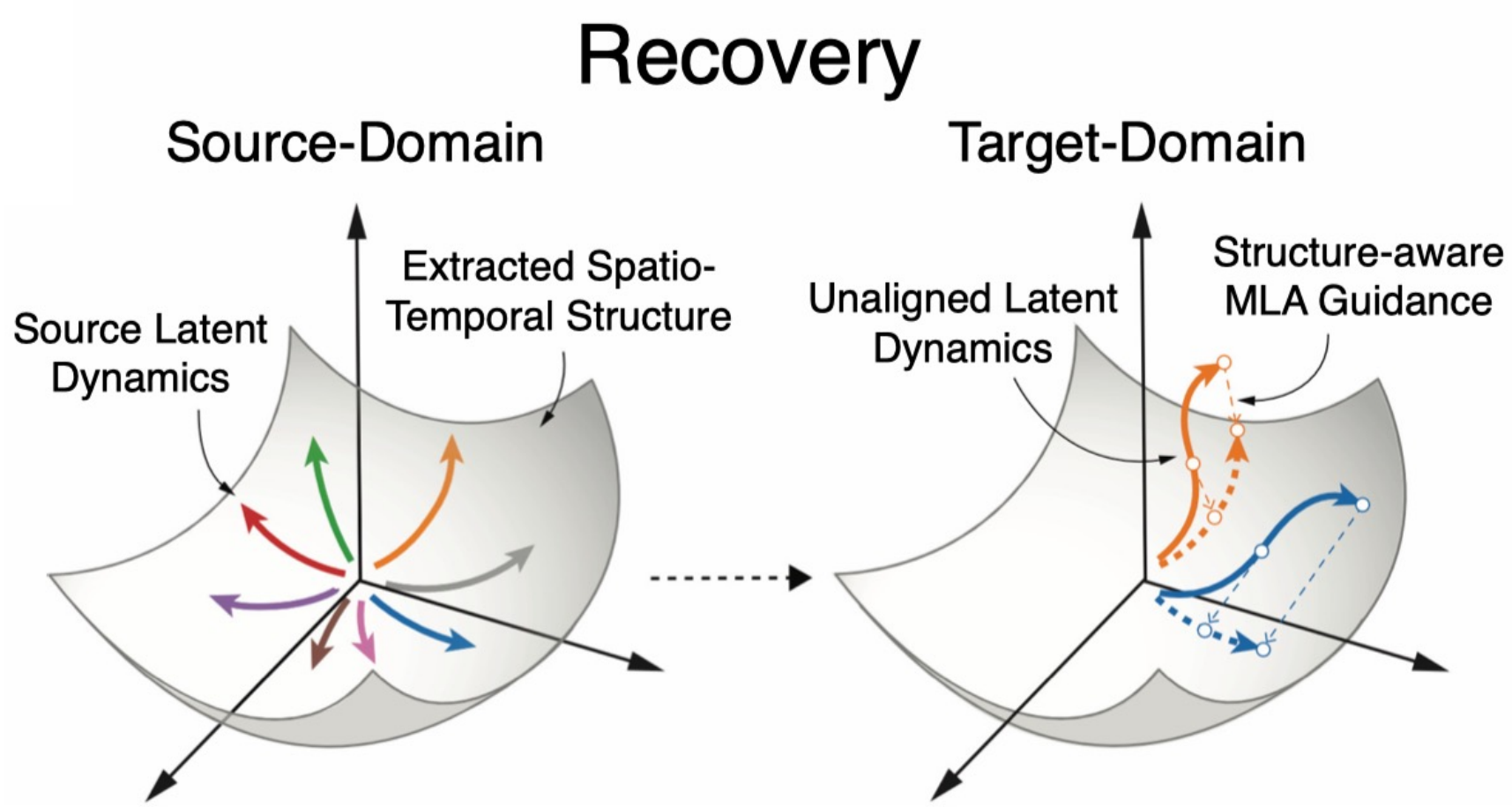
- The divergence objective can be approximated using the Hutchinson–Skilling trace estimator [1], making the whole optimization objective computationally tractable.

- We note that the recovery of spatio-temporal structure is primarily conducted by the denoising score matching part.

[1] J. Skilling, “The eigenvalues of mega-dimensional matrices,”

Georgia Tech

ERDiff Spatio-Temporal Structure Recovery



Georgia Tech

Experiment

Experiment

Synthetic Datasets – Results

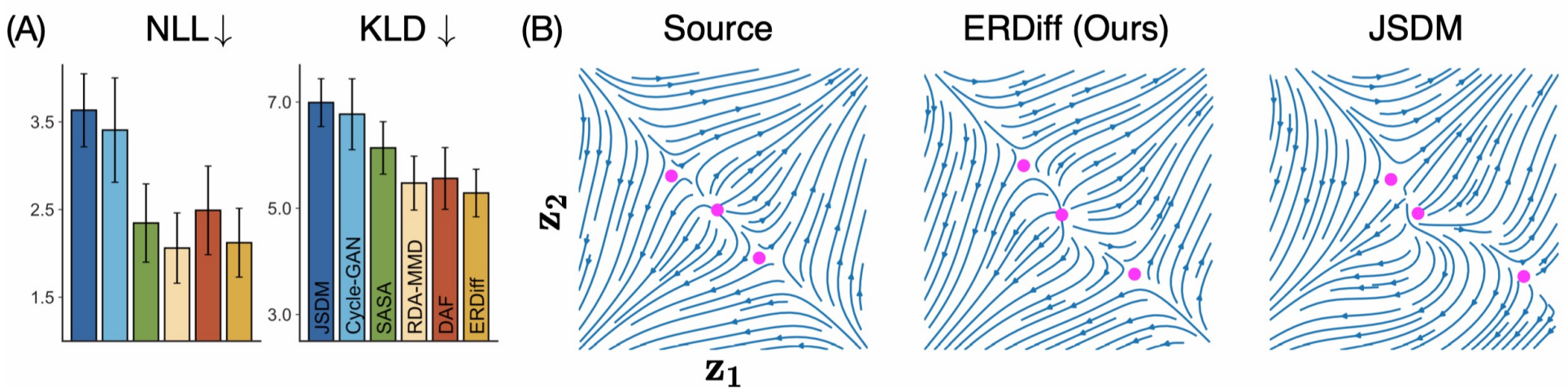


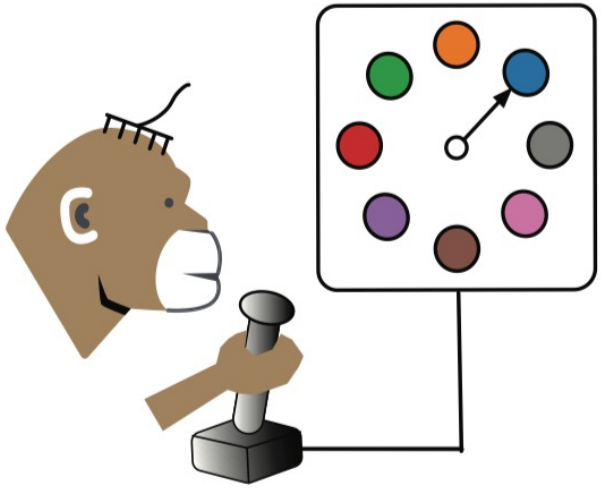
Figure 3: **Experimental Results on the synthetic dataset**. **(A)** Performance comparison on trial-average negative log-likelihood (NLL) and KL Divergence (KLD). ↓ means the lower the better. ERDiff achieves the second-lowest NLL and the lowest KLD. **(B)** True continuous Bernoulli dynamics in the source domain compared to the latent dynamics aligned by ERDiff and JSDM in the target domain (blue dots denote the fixed points). ERDiff preserves the spatio-temporal structure of latent dynamics much better.

Georgia Tech

Experiment

Neural Datasets – Setup

- We conduct experiments with datasets collected from the primary motor cortex (M1) of non-human primates
- The primates have been trained to reach one of eight targets at different angles
- We perform cross-day (recordings of the same primate performing the task on different days) and inter-subject (recordings of different primates) experiments.



Neural Datasets – Neural Manifold Analysis

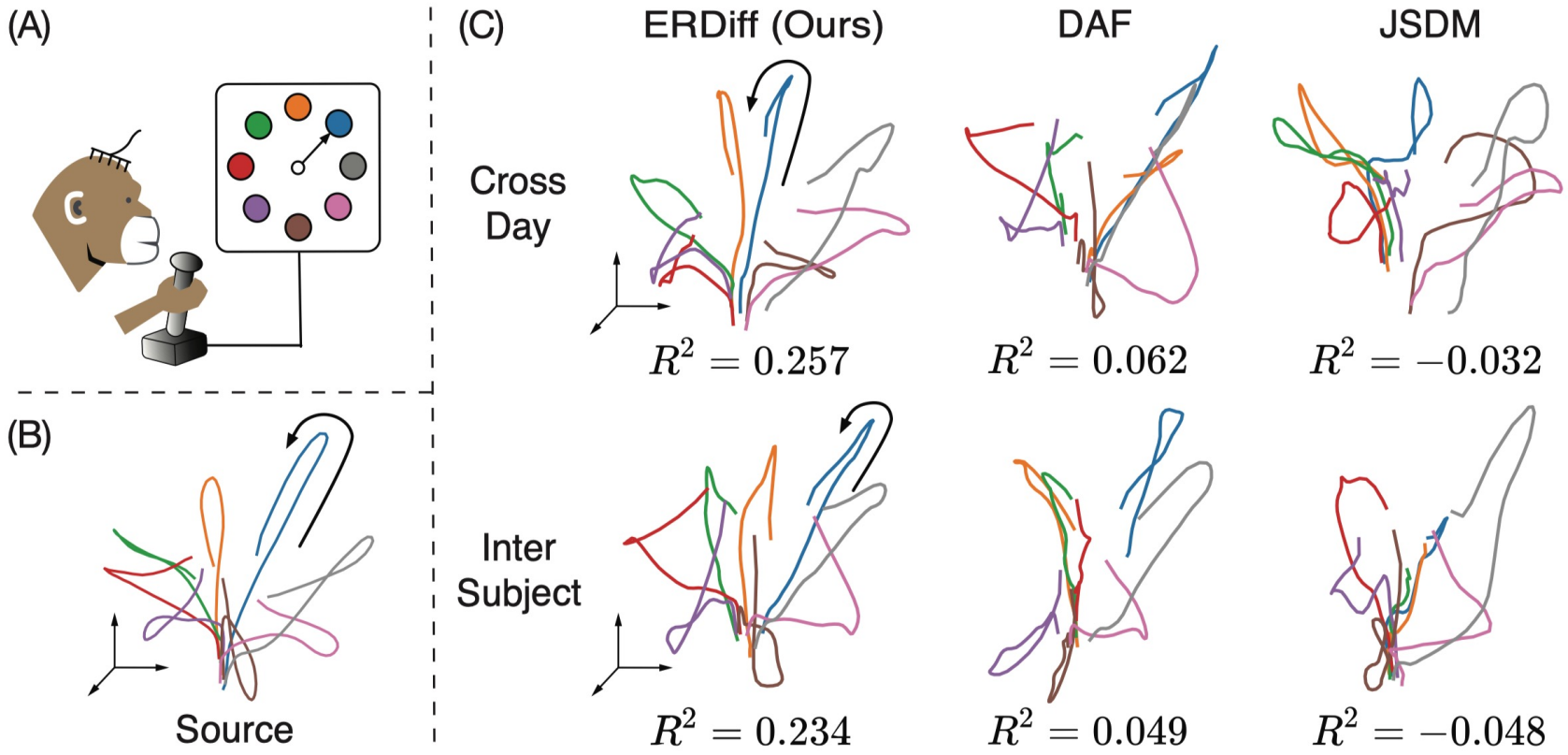


Figure 4: **Motor cortex dataset and Experimental Results**. **(A)** Illustration of the center-out reaching task of non-human primates. **(B)** The 3D Visualization of trial-averaged latent dynamics corresponding to each reaching direction in the source domain. **(C)** The 3D Visualization of trial-averaged latent dynamics corresponding to each reaching direction aligned by ERDiff, DAF, and JSDM given the target distribution from cross-day and inter-subject settings. We observe that ERDiff preserves the spatio-temporal structure of latent dynamics well.

Georgia Tech

Neural Datasets – Behavior Decoding Performance

Table 1: The R-squared values (R^2, in %) and RMSE of the methods on the motor cortex dataset. ERDiff w/o S is short for a variant of our proposed method that removes the spatial transformer layer in the DM. ERDiff w/o T is short for a variant of our proposed method that removes the temporal transformer layer in the DM. The boldface denotes the highest score. Each experiment condition is repeated with 5 runs, and their mean and standard deviation are listed.

Method	Cross-Day		Inter-Subject	
	$R^2(\%)\uparrow$	RMSE $\downarrow$	$R^2(\%)\uparrow$	RMSE $\downarrow$
Cycle-GAN	-24.83 ($\pm$3.91)	11.28 ($\pm$0.44)	-25.47 ($\pm$3.87)	12.23 ($\pm$0.46)
JSDM	-17.36 ($\pm$2.57)	9.01 ($\pm$0.38)	-19.59 ($\pm$2.77)	11.55 ($\pm$0.52)
SASA	-12.66 ($\pm$2.40)	8.36 ($\pm$0.32)	-14.33 ($\pm$3.05)	10.62 ($\pm$0.40)
DANN	-12.57 ($\pm$3.28)	8.28 ($\pm$0.32)	-18.37 ($\pm$3.24)	10.66 ($\pm$0.57)
RDA-MMD	-9.96 ($\pm$2.63)	8.51 ($\pm$0.31)	-6.31 ($\pm$2.19)	10.29 ($\pm$0.42)
DAF	-6.37 ($\pm$3.72)	8.17 ($\pm$0.48)	-11.26 ($\pm$3.64)	**9.57**($\pm$0.58)
ERDiff w/o S	-12.69 ($\pm$2.64)	8.57 ($\pm$0.50)	-14.60 ($\pm$2.88)	10.85 ($\pm$0.57)
ERDiff w/o T	-14.61 ($\pm$2.33)	8.93 ($\pm$0.50)	-17.10 ($\pm$3.23)	10.94 ($\pm$0.59)
ERDiff (Ours)	**18.81**($\pm$2.24)	**7.99**($\pm$0.43)	**10.29**($\pm$2.86)	9.78($\pm$0.50)

Georgia Tech

Computational Cost and Hyper-parameter Generalization

Table 2: Comparative analyses of computational cost between ERDiff and baseline methods during alignment. ERDiff has a comparable computational cost and maintains the stability of alignment.

Method	Cycle-GAN	JSDM	SASA	RDA-MMD	DAF	**ERDiff**
Add'l. Param	26K	0K	33K	65K	91K	28K
Add'l. Size	117KB	0KB	187KB	314KB	367KB	139KB
Align. Time	103ms	77ms	155ms	264ms	251ms	183ms
Stability	✗	✔	✔	✗	✗	✔

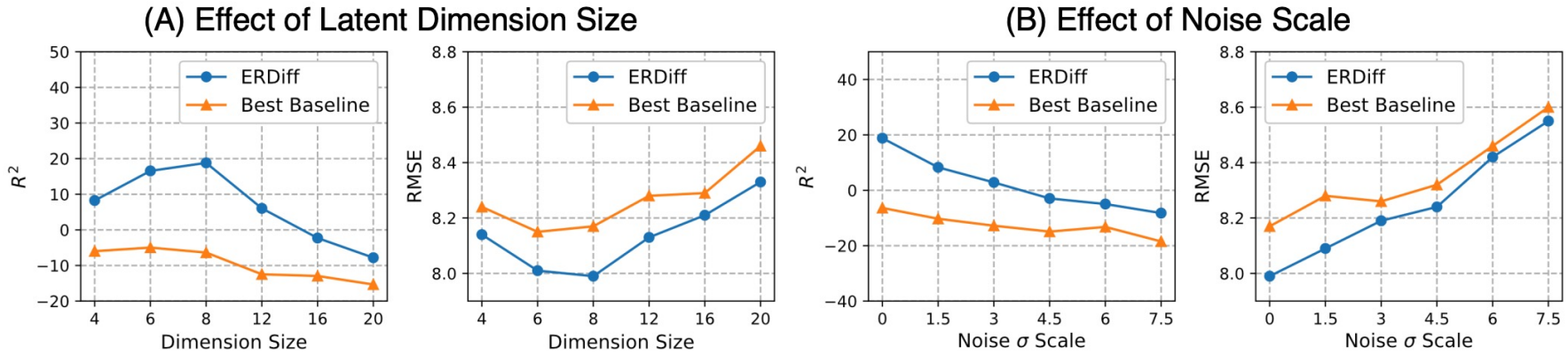


Figure 2: **Generalizability** and **Robustness** of ERDiff on alternative latent dimension sizes and noise scales (with 5 different random seeds). ERDiff consistently outperforms the best baseline method.

Georgia Tech

Thanks for listening!