



**Representation Learning for Cellular
Development in Spatiotemporal Transcriptomics**

Hung-Jen Wang

A thesis submitted in partial fulfillment of the
requirements for Honors in the Department of Computer
Science at Brown University

April 2026

Acknowledgment

I would like to sincerely thank Professor Ritambhara Singh for her guidance, support, and mentorship. She has been instrumental in shaping my academic journey, and I am deeply grateful to have had her as my research advisor.

I would also like to thank Jiaqi Zhang and Whitney Sloneker for their guidance as doctoral students. They mentored me thoroughly across their personal projects, and I am grateful for the opportunity to be a part of their research journey.

The Singh Lab as a whole also welcomed me with open arms. I learned a lot from the weekly journal clubs and undergraduate meetings, and it has been an honor to be a part of this community.

Last but not least, I would like to thank my parents, brother, and friends. You have made my undergraduate journey all the more meaningful, and I am truly grateful for the love and support you have given me.

Abstract

Recent advancements in spatiotemporal transcriptomics have facilitated the study of cellular dynamics across multiple timepoints. However, because cells are destructively sampled at each stage, reconstructing continuous developmental trajectories remain a fundamental challenge, and capturing meaningful cell-fate trajectories require models that can learn biological, spatial, and dynamic structures. We propose a representation learning framework for modeling cellular dynamics in spatiotemporal transcriptomics. Specifically, we integrate graph convolutional networks to encode spatial relationships, optimal transport to establish cell correspondence across temporal slices, neural ordinary differential equations to model temporal progressions, and contrastive learning to improve representation qualities. Together, these components build a coherent latent space that captures both local cell microenvironment and global developmental dynamics of spatiotemporal transcriptomics.

Contents

1	Introduction	5
1.1	Background	5
1.2	Motivation	6
2	Related Works	6
3	Methodology	6
4	Experiment Settings	8
4.1	Experiment 1: Structure Preservation	8
4.2	Experiment 2: Temporal Alignment	9
5	Results	9
5.1	Successful Structure Preservation of All Properties	9
5.2	Limitations of Existing Models in Multi-slice Alignment	10
6	Discussion and Conclusion	12

1 Introduction

Spatial Transcriptomics (ST) datasets have attracted considerable attention in recent years, due to their capacity to retain spatial context while capturing transcriptomic information at single-cell resolution. Often formatted as an expression matrix paired with spatial coordinates within a tissue section, ST datasets help to facilitate the study of spatially resolved cell trajectory analysis [8, 19] and microenvironment cellular interactions [23].

Building upon the growing interest in ST analyses, recent techniques have advanced on temporally-resolved ST datasets, which capture gene expression profiles across multiple spatially-resolved tissue slices at successive time points. Unlike conventional ST which provides a static snapshot of tissue organizations, temporal ST enables the investigation of dynamic cellular processes, allowing researchers to trace the progression of cell states and developmental trajectories over time. Such analyses can reveal fundamental insights into processes such as embryogenesis [11] and stem cell differentiation [3], and they can support predictive modelings of cell fate decisions. For instance, a study leveraged temporal ST data of embryo development to identify the spatial origins of anterior second heart field progenitors [6], highlighting the capacity of temporal ST to uncover both spatial and temporal aspects of cellular organization. Therefore, analyses of such temporal ST data are critical for understanding how cells evolve across time within biological systems.

1.1 Background

There have been contributions on the analyses of ST datasets at single time points or single tissue slices, focusing on cell clusterings and spatial cell niche identifications [1, 4, 7, 13, 16, 17]. These approaches leverage Graph Neural Networks (GNNs) to model the spatial relationships among cells, as they provide a flexible framework to represent cells as nodes and their spatial interactions as edges. This allows for further integration of both gene expression and neighborhood information in downstream analyses. For example, a study leverages a GNN to encode each spot by pooling information from its spatial neighbors, compelling a locally coherent latent representation [13]. Although these methods are highly effective for characterizing tissue architecture and cellular heterogeneity within a single snapshot, they are inherently constrained to individual ST slice and cannot capture temporal dynamics or the progression of cellular states. Consequently, they are inadequate for investigating developmental processes that unfold across multiple stages and time points.

Adjacently, recent advances in dynamic modeling for single-cell transcriptomics, such as work on neural Ordinary Differential Equations (ODEs) [22], have driven research on the reconstruction of continuous cellular trajectories from time-resolved expression data. However, most of these approaches are primarily designed for non-spatial gene expression data, and they ignore important spatial relations when directly applied upon ST frameworks. Neighborhood local interactions and contexts, which are critical for understanding microenvironment cellular behaviors, would undesirably be neglected for downstream analyses.

To overcome these limitations, recent research have focused on studying multiple ST slices to facilitate temporal analyses. However, in single-cell spatial transcriptomic technologies, cells are typically dissociated or destructively sampled at each time point, resulting in different sets of cells being measured across time points or slices. Consequently, direct comparisons between time points are confounded by both tissue heterogeneity and cell sampling variability, hindering the reconstruction of continuous developmental trajectories from observations. Therefore, in order to trace cellular trajectories and capture dynamic changes over time, cell alignment between temporal ST slices is required.

State-of-the-art methods [5, 9, 20, 21] employ Optimal Transport (OT) [24] frameworks to establish cell correspondences between ST slices. For example, a study solves the Fused Gromov–Wasserstein (FGW) OT problem to match transcriptomic features between two slices [21]. While these methods facilitate alignment and provide insights into changes between individual time points, they are generally limited to pairwise ST

slice alignment and do not capture continuous cellular dynamics across the full temporal sequence. This restricts their ability to reconstruct long-range developmental trajectories.

1.2 Motivation

To address these challenges, we propose a novel framework that jointly analyzes multiple ST slices across time points, explicitly incorporating the long-range temporal progression of cellular developments and providing a comprehensive view of the dynamic system. Our approach first integrates GNNs to capture spatial relationships and local neighborhood information among cells. We then employ an optimal transport (OT) framework based on the Wasserstein distance to establish latent correspondences across timepoints. A neural ODE, modeled by a neural network framework, is then used to represent cellular velocities across the latent space. Finally, we leverage contrastive learning to improve representation qualities and prevent spurious correlations. By integrating these complementary techniques, our framework aims to reconstruct complex developmental trajectories, map cell fate transitions, and predict future cellular states while accounting for the spatial microenvironment.

2 Related Works

There are several published works that are related to our studies. Considering biological, spatial, and dynamic structure preservations of temporal ST processes, scNODE [22] and GraphST [13] are two relevant models. scNODE [22] is a method that operates on cell expressions across time points and constructs a meaningful latent representation from this information. It then mathematically connects these time stages via a neural ODE, building a cell-trajectory inference pipeline for both interpolation and extrapolation tasks. GraphST [13] is another method that focuses on compelling locally coherent latent representations in individual ST slices, leveraging graph-based contrastive learning to preserve neighborhood structure. Compared to these methods, our novelty lies in the ability to capture all three relevant structures, modeling temporal dynamics with both biological and spatial signal consideration.

Adjacently, we also consider the issue of temporal ST slice alignment. Three contemporary models that focus on ST alignment are Probabilistic Alignment of ST Experiments (PASTE) [21], Developmental Spatiotemporal Optimal Transport (DeST_OT) [5], and Multi-omics Single-cell Optimal Transport (moscot) [9]. Published in 2022, PASTE is a foundational work in ST alignment. Its pairwise alignment focuses on aligning across tissue thickness, but most existing alignment models still benchmark against PASTE. On the other hand, DeST_OT and moscot have alignment algorithms that function appropriately across the temporal axis. Published in 2025, they are considered two of the most modern models of spatio-temporal transcriptomics alignment. Compared to these methods, our novelty lies in the ability to model spatiotemporal transcriptomics as a coherent latent dynamic process, rather than directly mapping discrete slice-to-slice alignments.

3 Methodology

We propose to adopt techniques from scNODE [22] to build a dynamic-aware latent representation across time points. Specifically, with $\mathbf{X}^{(t)}$ representing the gene expression of observed cells at timepoint t and $\mathbf{S}^{(t)}$ representing the spatial coordinates matrix at timepoint t , there exists a cell graph $\mathcal{G}^{(t)} = (\mathbf{X}^{(t)}, \mathbf{A}^{(t)})$ at timepoint t , where nodes correspond to cells with their gene expressions, and edges are represented by $\mathbf{A}^{(t)}$, the adjacency matrix based on spatial coordinates of $\mathbf{S}^{(t)}$. To incorporate these spatial features, a Graph Convolutional Network (GCN) architecture, under the GNN family, replaces the original encoder in the scNODE [22] encoder-decoder network.

At each layer, the GCN model leverages the adjacency matrix to update a single cell’s representation by aggregating information from its neighboring cells. This supports the encoding of spatial relationships between nearby cells, and after multiple layers of aggregations and nonlinear transformations, the GCN model

outputs a latent representation that effectively captures both the expression profile and surrounding environment. Specifically, it is defined as the following, with $\tilde{\mathbf{A}}$ representing the normalized adjacency matrix, $\mathbf{W}$ and $\mathbf{b}$ representing the trainable weight matrix and bias vector, l representing each GCN layer, and $\sigma(\cdot)$ representing a nonlinear activation function.

$$\begin{aligned} \text{GCN}(\mathbf{X}, \mathbf{A}) &= \mathbf{Z}^L, \text{ where} \\ \mathbf{Z}^l &= \sigma(\tilde{\mathbf{A}}\mathbf{Z}^{l-1}\mathbf{W}^{l-1} + \mathbf{b}^{l-1}) \text{ for } l = 1, 2, \dots, L \text{ and } \mathbf{Z}^0 = \mathbf{X} \end{aligned} \quad (1)$$

That said, using GCNs as the encoder, the system then pre-trains on Mean Squared Error (MSE) to learn stable and expressive latent representations within each timepoint, with the loss defined as the following.

$$\begin{aligned} \mathbf{Z}_{enc}^{(t)} &= \text{GCN}(\mathbf{X}^{(t)}, \mathbf{A}^{(t)}) \\ \hat{\mathbf{X}}^{(t)} &= \text{Dec}(\mathbf{Z}_{enc}^{(t)}) \\ \mathcal{L}_{pre} &= \sum_t \text{MSE}(\mathbf{X}^{(t)}, \hat{\mathbf{X}}^{(t)}) \end{aligned} \quad (2)$$

After pre-training, neural ODEs are then leveraged to capture cellular dynamics across this latent space. The neural ODE describes dynamics of latent representations as the following, where *Drift* is a nonlinear neural network representing cell velocities in the latent space.

$$\begin{aligned} \mathbf{Z}_{dyn}^{(0)} &= \text{GCN}(\mathbf{X}^{(0)}, \mathbf{A}^{(0)}) \\ \mathbf{Z}_{dyn}^{(t+\Delta t)} &= \mathbf{Z}_{dyn}^{(t)} + \text{Drift}(\mathbf{Z}_{dyn}^{(t)}, t)\Delta t \end{aligned} \quad (3)$$

To train the neural ODE, we first minimize the distribution differences between the input $\mathbf{X}$ and reconstructed gene expression $\hat{\mathbf{X}}$ using the Wasserstein metric. We then follow the dynamic regularization strategy in scNODE [22], where we enforce consistency between encoder latents $\text{GCN}(\mathbf{X}, \mathbf{A})$ and dynamically evolved latents $\mathbf{Z}_{dyn}$. Specifically, we minimize the Wasserstein distance between these two latent representations, effectively constraining the ODE trajectory to pass through regions of latent space that is supported by the observed data points. That said, the dynamic loss can be written as the following.

$$\begin{aligned} \mathbf{Z}_{enc}^{(t)} &= \text{GCN}(\mathbf{X}^{(t)}, \mathbf{A}^{(t)}) \\ \mathcal{L}_{dyn} &= \sum_t \text{Wass}(\mathbf{X}^{(t)}, \hat{\mathbf{X}}^{(t)}) + \beta \sum_t \text{Wass}(\mathbf{Z}_{enc}^{(t)}, \mathbf{Z}_{dyn}^{(t)}) \end{aligned} \quad (4)$$

To further improve representation quality, we incorporate a self-supervised contrastive learning (SCL) objective, inspired by GraphST's [13] methodologies. The key idea is to encourage the model to learn meaningful relationships between the gene expression and spatial context, while avoiding potentially spurious correlations. To achieve this, in addition to original examples where we have correct gene expressions and neighborhood information ($\mathbf{X}$), we also construct corrupted examples by randomly shuffling gene expression values while keeping the spatial structure unchanged ($\tilde{\mathbf{X}}$). This constructs corrupted versions of the cell graph and latent encoder representations, defined as the following.

$$\begin{aligned} \tilde{\mathcal{G}} &= (\tilde{\mathbf{X}}, \mathbf{A}) \\ \tilde{\mathbf{Z}}_{enc} &= \text{GCN}(\tilde{\mathbf{X}}, \mathbf{A}) \end{aligned} \quad (5)$$

The model then learns to distinguish between true ($\mathbf{Z}_{enc}$) and corrupted representations ($\tilde{\mathbf{Z}}_{enc}$) by pairwise comparing each cell's embedding with an aggregated representation of its respective neighbors (g_i), defined as the following, with $N(i)$ representing the set of neighbors for some cell i .

$$g_i = \frac{1}{|N(i)|} \sum_{j \in N(i)} z_j \quad (6)$$

Positive pairs consist of consistent cell-neighborhood relationships, while negative pairs consist of mismatched or corrupted signals. We then leverage conventional Binary Cross-Entropy (BCE) and a probabilistic discriminator ($\Phi(\cdot)$) of positive pairs to construct a contrastive loss term ($\mathcal{L}_{con}$). Specifically, it is defined as the following.

$$\begin{aligned} \text{SCL}(\mathbf{X}, \mathcal{G}, \tilde{\mathcal{G}}) &= \sum_{i=1}^n \mathbb{E}_{\mathcal{G}}[\log \Phi(z_i, g_i)] + \mathbb{E}_{\tilde{\mathcal{G}}}[\log(1 - \Phi(\tilde{z}_i, g_i))] \\ \mathcal{L}_{con} &= \sum_t \text{SCL}(\mathbf{X}^{(t)}, \mathcal{G}^{(t)}, \tilde{\mathcal{G}}^{(t)}) \end{aligned} \quad (7)$$

This encourages the model to produce embeddings on a tighter learning trajectory, building more biologically and spatially meaningful representations. Combining with the previous dynamic loss term, our overall loss would be defined as the following.

$$\mathcal{L}_{all} = \mathcal{L}_{dyn} + \gamma \mathcal{L}_{con} \quad (8)$$

4 Experiment Settings

We evaluate our model with two perspectives. First, we assess biological, spatial, and dynamic structure preservation to determine whether the learned representation retains meaningful patterns in the data. Second, we evaluate temporal ST alignment performances to examine cell-fate trajectories and slice-wise correspondences. Together, these experiments provide us with a comprehensive assessment of both representation quality and alignment/development accuracy.

4.1 Experiment 1: Structure Preservation

Specifically, to evaluate on structure preservation, we benchmark with three different temporal ST datasets: Zebrafish Embryogenesis [12], Human Heart [14], and Human Pancreas [15], described in the following table.

	Dataset 1	Dataset 2	Dataset 3
Name	Zebrafish Embryogenesis	Human Heart	Human Pancreas
Number of Slices	6	7	4
Number of Cells	13,166	25,309	5,531

Summary of Spatiotemporal Transcriptomics Datasets

We then calculate the NMI, Moran’s I, and Time Correlation scores on these respective datasets to assess biological, spatial, and dynamic structure preservation for our model, GraphST [13], and scNODE [22].

$$\begin{aligned} \text{NMI} &= \frac{2 I(Y; C)}{H(Y) + H(C)} \\ \text{Moran's I} &= \frac{N}{W} \cdot \frac{\sum_{i,j} w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2} \\ \text{Time Corr.} &= \text{dCor}(\mathbf{Z}, \mathbf{t}) \end{aligned} \quad (9)$$

NMI compares ground-truth cell types (Y) with latent space cluster assignments (C) to assess the consistency between biological labels and learned representations. $I(Y; C)$ represents the mutual information between Y and C , and $H(\cdot)$ represents the entropy. Moran’s I evaluates the spatial autocorrelation, capturing the extent to which nearby cells in the latent space have similar representations. Specifically, with N representing the total number of cells and W representing the sum of the weighting/adjacency matrix, it compares the values

of cell representations (z_i, z_j) across spatially adjacent pairs defined by a weighting matrix (w_{ij}), relative to the global mean of cell representations ($\bar{z}$). Finally, Time Correlation leverages distance correlations [18] between cell representations ($\mathbf{Z}$) and timepoint labels ($\mathbf{t}$) to evaluate the inferred dynamic trajectories. For all three metrics, a higher value indicates better preservation of the respective category.

4.2 Experiment 2: Temporal Alignment

To evaluate on temporal ST alignment, we only benchmark on the Zebrafish Embryogenesis [12] dataset with our model, PASTE [21], DeST_OT [5], and moscot [9]. Specifically, for each method, we obtain correspondences between cells across timepoints and construct transition matrices of cell types to capture piece-to-piece ST developments. We construct sankey diagrams and heatmaps for visualizations, as they are common tools to show flow across proportional distributions. This will provide us with an interpretable view for discussion on cell-fate evolutions over time, and we assess downstream whether these trajectories are biologically plausible/feasible.

5 Results

We conclude the two experiments with disparate yet impactful results. First, our model performs exceedingly well in structure preservation of all three properties, and this indicates a strong representation learning framework for cellular development in temporal ST. Second, we observe pitfalls in our model with temporal ST alignment and slice-by-slice transition matrices. We verify our performance in a biological context and reference similar issues with other state-of-the-art methods.

5.1 Successful Structure Preservation of All Properties

Structure Preservation of Zebrafish Embryogenesis Dataset [12]

Method	Biological (NMI)	Spatial (Moran’s I)	Dynamic (Time Corr.)
Our Model	0.380	0.511	0.512
GraphST [13]	0.252	0.497	0.369
scNODE [22]	0.298	0.321	0.374

Structure Preservation of Human Heart Dataset [14]

Method	Biological (NMI)	Spatial (Moran’s I)	Dynamic (Time Corr.)
Our Model	0.260	0.443	0.433
GraphST [13]	0.132	0.244	0.240
scNODE [22]	0.364	0.374	0.541

Method	Biological (NMI)	Spatial (Moran's I)	Dynamic (Time Corr.)
Our Model	0.078	0.478	0.340
GraphST [13]	0.039	0.306	0.292
scNODE [22]	0.061	0.143	0.376

The two baseline models here represent different preservation strategies: GraphST [13] is known for its spatial preservation capabilities, while scNODE [22] is known for its ability to capture biological and dynamic patterns. That said, our model still consistently ranks among the top two scorers across all three evaluation criteria on each dataset. This shows that our method effectively preserves biological, spatial, and dynamic structures on a caliber comparable to state-of-the-art methods, indicating that it successfully identifies a reliable and expressive latent representation of temporal ST slices.

5.2 Limitations of Existing Models in Multi-slice Alignment

For all of the following figures regarding the zebrafish embryogenesis dataset [12], the cell-type labels are ordered for each timepoint in the sankey diagram, and the rows and columns of the transition matrices correspond to these cell-type orderings at each stage. On the heatmaps, darker shades indicate higher alignment probabilities, while lighter shades indicate lower probabilities, and they are row-normalized to sum up to one. To focus the discussion, we also only include the Sankey diagram of our model for conciseness.

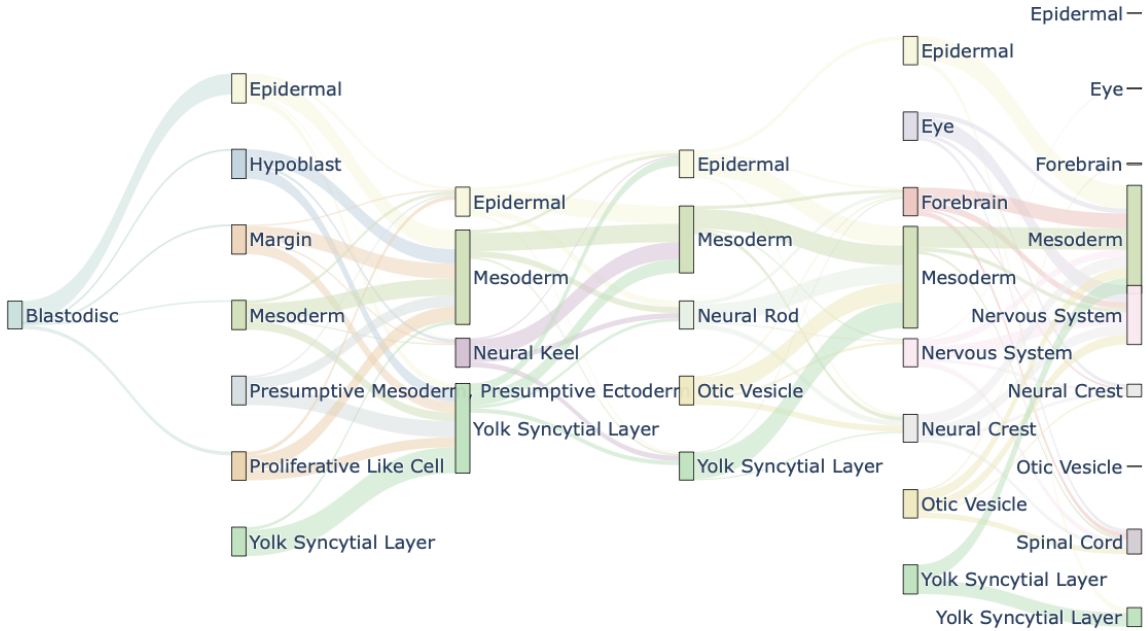


Figure 1: Sankey Diagram of Our Model

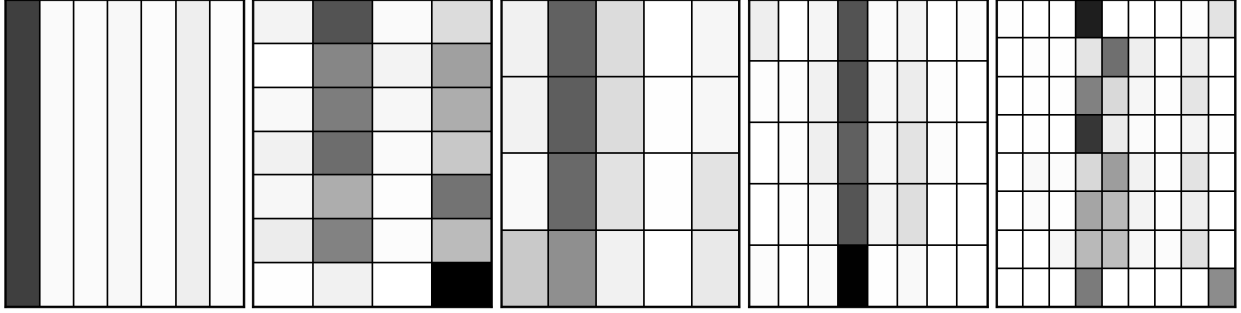


Figure 2: Temporal Alignment of Our Model

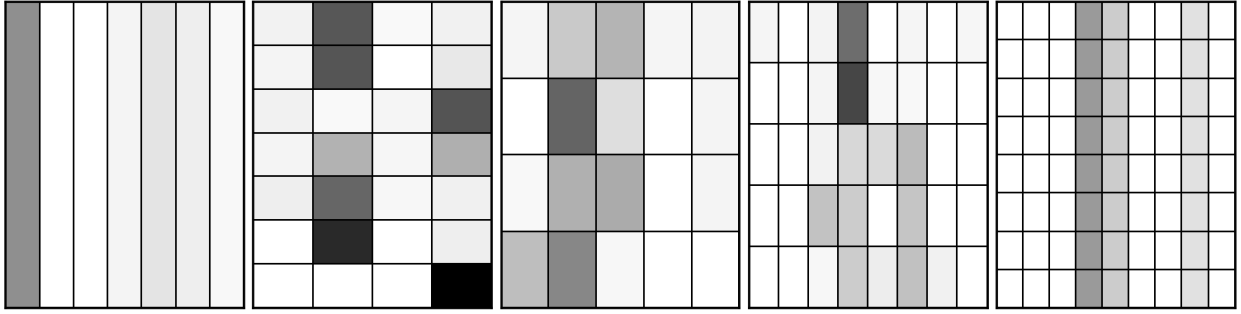


Figure 3: Temporal Alignment of PASTE [21]

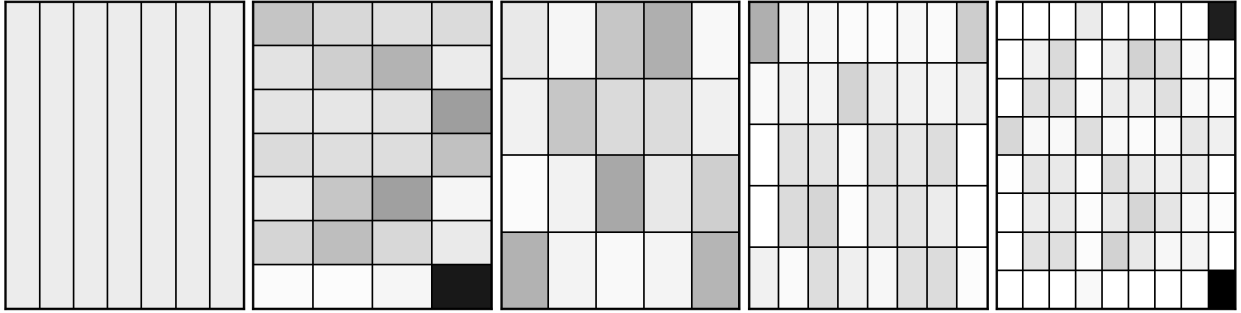


Figure 4: Temporal Alignment of DeST_OT [5]

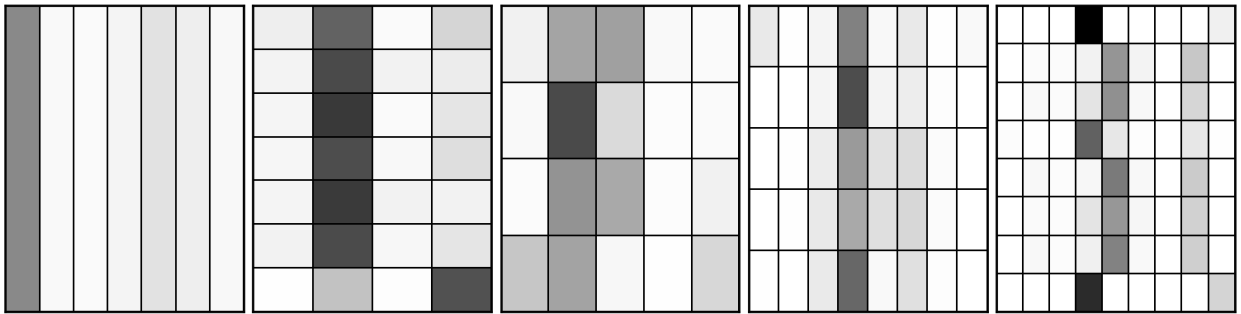


Figure 5: Temporal Alignment of moscot [9]

We observe a few issues with the temporal ST alignment and transition matrices of our model. First, while strong Mesoderm and Yolk Syncytial Layer continuity is biologically accurate, there exhibits excessive mixing into Mesoderm from other cell types. This indicates that the Mesoderm cell type seems to be a convenient sink for our model, and there may be bias and/or over-smoothing in the latent space. Furthermore, the model appears to have difficulty in resolving more fine-grained structures, particularly the Epidermal, Eye, and Forebrain cell types. The lack of distinct branches suggests an insufficient resolution in representation, and the OT matching may again be overly smoothed. Interestingly, we also observe that while PASTE [21] and moscot [9] models yield similar alignments as our model’s, DeST_OT [5] is presented as a much more overly distributed option. This vast difference is likely due to a tradeoff between signal confidence and regularization, and we should aim to position our model at the middle ground.

6 Discussion and Conclusion

In summary, our results showed that the proposed model is effective in learning representations that preserve biological, spatial, and dynamic structures in ST data. Across three temporal ST datasets, our model consistently outperforms or performs competitively against other state-of-the-art approaches, indicating an output of a robust and meaningful latent representation. In terms of temporal slice-wise alignment, the inferred cell-fate trajectories reveal key challenges between alignment resolution and regularization, hinting that our model is currently only recovering coarser developmental structures. Similar issues resonate with the other contemporary models, suggesting that biologically faithful trajectories remain an open problem for researchers to tackle upon.

There are also a few unsolved problems regarding our experimental setups. First, only the Zebrafish Embryogenesis dataset [12] is calibrated for batch effect with Harmony [10]. Regarding structure preservation, the Human Heart [14] and Human Pancreas [15] dataset were only ran and compared on the original versions, with no additional engineering to reduce batch effects. Second, regarding temporal ST alignments, only the Zebrafish Embryogenesis dataset [12] was used for benchmarking. The Human Heart [14] and Human Pancreas [15] dataset were not considered for this experiment. Third, only the structure preservation models had hyperparameter tuning from Optuna [2]. The temporal ST alignment models, other than our method, were based on default hyperparameters, and the underlying reason is largely due to time constraints.

Regarding future work, the first step would be to normalize the experimental setups and make sure each model is performing at the best of its respective capabilities. The second step would be to improve on alignment performances. Specifically, to address the sink-like behavior in dominant cell types such as Mesoderm, we should consider introducing additional regularization and constraints to prevent ambiguous mass from collapsing into convenient cell states. Furthermore, to improve on lineage-specific discrimination, we should sharpen the latent representation with higher resolution. Possible directions to conduct this could include OT lineage-aware penalties, cell-type contrastive losses, and fine-grained GCN encoders. This demonstrates that there are still actionable, concrete improvements that could be made to our model, and we plan on refining it till the method learns a biologically faithful alignment strategy.

References

- [1] Staligner enables the integration and alignment of multiple spatial transcriptomics datasets. *Nature Computational Science*, 3(831–832), 2023.
- [2] Akiba et al. Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2623–2631), 2019.
- [3] Crosse et al. Multi-layered spatial transcriptomics identify secretory factors promoting human hematopoietic stem cell development. *Cell Stem Cell*, 27(5)(822–839), 2020.
- [4] Dong et al. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13(1)(1739), 2022.
- [5] Halmos et al. Dest-ot: Alignment of spatiotemporal transcriptomics data. *Cell Systems*, 16(2)(101160), 2025.
- [6] He et al. Integration of single-cell and spatial transcriptomics by seu-tca reveals the spatial origin of early cardiac progenitors. *Genome Biology*, 26(1)(158), 2025.
- [7] Hu et al. Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods*, 18(11)(1342–1351), 2021.
- [8] Jin et al. Advances in spatial transcriptomics and its applications in cancer research. *Molecular Cancer*, 23(1)(129), 2024.
- [9] Klein et al. Mapping cells through time and space with moscot. *Nature*, 638(8052)(1065–1075), 2025.
- [10] Korsunsky et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nature Methods*, 16(12)(1289–1296), 2019.
- [11] Kumar et al. Spatiotemporal transcriptomic maps of whole mouse embryos at the onset of organogenesis. *Nature Genetics*, 55(7)(1176–1185), 2023.
- [12] Liu et al. Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Developmental Cell*, 57(10)(1284–1298), 2022.
- [13] Long et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications*, 14(1)(1155), 2023.
- [14] Lázár et al. Spatiotemporal gene expression and cellular dynamics of the developing human heart. *Nature Genetics*, 57(11)(2756–2771), 2025.
- [15] Olaniru et al. Single-cell transcriptomic and spatial landscapes of the developing human pancreas. *Cell Metabolism*, 35(1)(184–199), 2023.
- [16] Qian et al. Identification and characterization of cell niches in tissue from spatial omics data at single-cell resolution. *Nature Communications*, 16(1)(1693), 2025.
- [17] Ren et al. Identifying multicellular spatiotemporal organization of cells with spaceflow. *Nature Communications*, 13(1)(4076), 2022.
- [18] Székely et al. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35(6)(2769–2794), 2007.
- [19] Wang et al. Advances in spatial transcriptomics and its application in the musculoskeletal system. *Bone Research*, 13(1)(54), 2025.
- [20] Xia et al. Spatial-linked alignment tool (slat) for aligning heterogenous slices. *Nature Communications*, 14(1)(7236), 2023.

- [21] Zeira et al. Alignment and integration of spatial transcriptomics data. *Nature Methods*, 19(5)(567–575), 2022.
- [22] Zhang et al. scnode : Generative model for temporal single cell transcriptomic data prediction. *Bioinformatics*, 40(Suppl 2)(146-154), 2023.
- [23] Tjoonk. Five key applications of spatial transcriptomics. *Single Cell Discoveries*, 2023.
- [24] Williams. A short introduction to optimal transport and wasserstein distance. 2020.