

# SPHENIC: Topology-Aware Multi-View Clustering for Spatial Transcriptomics

Chenkai Guo<sup>2,5</sup>, Yikai Zhu<sup>1</sup>, Renxiang Guan<sup>3</sup>, Jinli Ma<sup>1</sup>, Siwei Wang<sup>4</sup>,  
Ke Liang<sup>3</sup>, Guandong Peng<sup>2,5</sup> and Dayu Hu<sup>1,†</sup>

<sup>1</sup>College of Medicine and Biological Information Engineering, Northeastern University

<sup>2</sup>Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences

<sup>3</sup>College of Computer Science and Technology, National University of Defense Technology

<sup>4</sup>Intelligent Game and Decision Lab, Academy of Military Sciences

<sup>5</sup>University of Chinese Academy of Sciences

<sup>†</sup> Corresponding Author

shaunak.guo@gmail.com, hudy@bmie.neu.edu.cn

## Abstract

Spatial transcriptomics clustering is pivotal for identifying cell subpopulations by leveraging spatial location information. While recent graph-based methods modeling cell–cell interactions have improved clustering accuracy, they remain limited in two key aspects: (i) reliance on local aggregation in static graphs often fails to capture robust global topological structures (e.g., loops and voids) and is vulnerable to noisy edges; and (ii) dimensionality reduction techniques frequently neglect spatial coherence, causing physically adjacent spots to be erroneously separated in the latent space. To overcome these challenges, we propose SPHENIC, a **Spatial Persistent Homology–Enhanced Neighborhood Integrative Clustering** method. Specifically, it explicitly incorporates topology-invariant features into the clustering network to ensure robust representation learning against noise. Furthermore, we design a dual-regularized optimization module that imposes spatial constraints alongside distributional optimization, ensuring that the embedding space preserves the physical proximity of cells. Extensive experiments on 11 benchmark datasets demonstrate that SPHENIC outperforms state-of-the-art methods by 4.19%–9.14%, validating its superiority in characterizing complex tissue architectures.

## 1 Introduction

Spatial transcriptomics has transformed the study of biological systems by providing spatially resolved gene-expression profiles in intact tissues [Rao *et al.*, 2021; Moses and Pachter, 2022; Tian *et al.*, 2023; Williams *et al.*, 2022; Jain and Eadon, 2024]. A critical step in this approach is the accurate spatial clustering required to delineate spatially coherent cellular domains in which cells exhibit both molecular similarity and physical proximity. Accurate identification of these domains may have far-reaching implications, including guiding therapeutic target discovery and advancing our understanding of

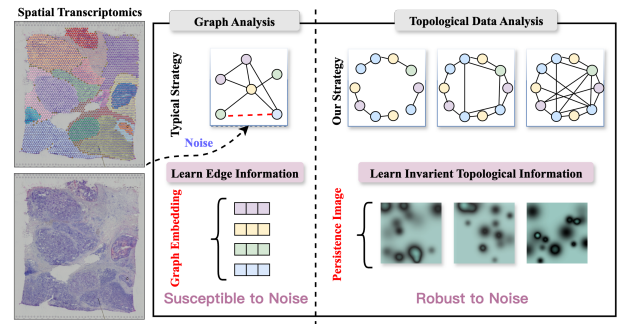


Figure 1: Comparison of static graph learning and invariant topological feature learning.

complex biological systems and cellular behaviors [Gulati *et al.*, 2025; Wang *et al.*, 2024].

Current spatial transcriptomics clustering methods can be broadly classified into three categories: (i) traditional community-detection approaches (e.g., Seurat and Scanpy), which primarily emphasize gene expression while giving limited consideration to spatial organization [Wolf *et al.*, 2018; Satija *et al.*, 2015]; (ii) deep-learning-based methods, including GraphST and STAIG, which explicitly model spatial–transcriptional relationships through graph structures to identify biologically meaningful spatial domains [Long *et al.*, 2023; Yang *et al.*, 2025]; and (iii) supervised methods, such as PASSAGE, which incorporates phenotypic guidance into the clustering process [Guo *et al.*, 2025]. The latter two categories have attracted substantial attention and are widely regarded as significant improvements over approaches that rely solely on the gene expression matrix [Li *et al.*, 2025]. Among these approaches, graph-based analysis plays a pivotal role, as accurate learning of the underlying topological structure is essential for identifying cell–cell associations and, consequently, achieving high clustering accuracy.

Despite recent progress, existing graph-based spatial clustering approaches still face two major challenges: (i) **Susceptibility to Noise**: Most methods rely on graph representations to capture topological patterns (Fig. 1). However, such representations are often sensitive to noise, making them unreli-

able for preserving global geometric structures, such as loops and voids. Although some approaches incorporate persistent homology, they often fail to capture complex multidimensional structures, which hampers the learning of robust and topology-invariant representations. (ii) **Inadequate Spatial Coherence:** In addition, most existing methods model spatial relationships using simple adjacency graphs. This strategy fails to capture sufficient spatial context, leading to embeddings that distort neighborhood relationships and incorrectly separate physically adjacent spots in the latent space.

To address these limitations, we propose SPHENIC, a spatial clustering framework tailored for spatial transcriptomics data. Specifically, SPHENIC extends persistent homology to effectively learn topology-invariant features. It then constructs modality-specific views and fuses them using a multi-view Graph Convolutional Network (GCN) to learn comprehensive consensus representations. In addition, we develop a dual-regularized optimization module, which incorporates spatial constraints and distribution optimization to further refine cell distributions and embedding quality by increasing the similarity between a cell’s embedding and those of its spatial neighbors while decreasing similarity to non-neighboring cells. Extensive experiments demonstrate the superiority of our proposed SPHENIC. Our main contributions are summarized as follows:

- *Framework:* We pioneer the integration of extended persistent homology into spatial clustering frameworks. To the best of our knowledge, this study represents one of the earliest attempts to incorporate topology-invariant features for improving spatial clustering.
- *Algorithm:* We develop a novel dual-regularized optimization component that integrates spatial constraints with distributional optimization to further enhance cell distribution accuracy and embedding quality. Moreover, this component ensures that physically adjacent spots are assigned similar embeddings.
- *Evaluation:* Extensive experiments have been conducted across three distinct spatial transcriptomics datasets, encompassing a total of eleven tissue slices, and eight current state-of-the-art baselines, demonstrating the consistent superiority of SPHENIC over existing state-of-the-art methods.

## 2 Related Work

### 2.1 Clustering in Spatial Transcriptomics

Clustering strategies for spatial transcriptomics data can be broadly classified into three paradigms [Liu *et al.*, 2024; Zahedi *et al.*, 2024; Hu *et al.*, 2024; Yuan *et al.*, 2024]. (i) Traditional community-detection methods, implemented in tools such as Seurat and Scanpy, employ the Louvain or Leiden algorithm to cluster cells on the basis of gene expression similarity [Wolf *et al.*, 2018; Satija *et al.*, 2015; Blondel *et al.*, 2008; Traag *et al.*, 2019]. Although computationally efficient, these methods typically treat spatial information as secondary and may therefore overlook the critical aspects of tissue organization. (ii) Deep-learning-based techniques have recently emerged to more effectively integrate

spatial context with gene expression. For example, SpaGCN employs GCN to model spatial neighbourhoods, whereas stLearn jointly incorporates the expression profiles, spatial coordinates, and histological image features [Hu *et al.*, 2021; Pham *et al.*, 2023; Kipf, 2016]. Spatial-MGCN further advances this paradigm by introducing multi-view graph learning framework [Wang *et al.*, 2023]. (iii) Supervised methods leverage phenotypic guidance to delineate biologically meaningful spatial domains and molecular signatures. PAS-SAGE exemplifies this strategy by learning the phenotype-associated signatures through a two-step contrastive graph-learning framework, thereby identifying domains that correspond to the known anatomical or pathological phenotypes [Guo *et al.*, 2025]. However, none of the existing approaches integrates invariant topological information into the spatial clustering process.

### 2.2 Topological Data Analysis

While traditional geometric deep learning outperforms in capturing local dependencies, it often struggles to quantify global structural properties invariant to deformation. Topological data analysis (TDA) leverages the rigorous framework of algebraic topology to extract multi-scale structural features, like connected components, loops, and voids, which offer a robust complement to standard graph representations [Pham *et al.*, 2025; Wasserman, 2018; Su *et al.*, 2025]. Recent progress can be summarized along three key directions: (i) Topologically-Augmented Architectures, standard Message Passing Neural Networks (MPNNs) are fundamentally limited by the Weisfeiler-Lehman hierarchy, often fail to distinguish non-isomorphic graphs with identical local neighborhoods (e.g., cycles vs. separated paths) [Gilmer *et al.*, 2020]. To bridge this gap, architectures like TopoGCL [Chen *et al.*, 2024] augment graph neural networks by injecting persistent homology features into the message-passing schema, thereby enabling more robust modeling of higher-order interactions in molecular and social graphs; (ii) Topological Regularization Techniques, exemplified by TopoReg [Chen *et al.*, 2019], introduce the filtration-based losses that preserve critical connectivity patterns within latent representations, which significantly improve performance of graph-classification task; and (iii) Topology-Guided Generation, in the generative domain, ensuring the global validity of synthesized structures remains challenging for purely probabilistic models. Therefore, methods like TopoDiff [Zhang *et al.*, 2025] employ topological signatures to steer the synthesis of graphs possessing desired structural properties, a capability that is particularly valuable for drug-discovery applications. Collectively, these advances demonstrate topology’s transformative potential for graph representation learning by capturing multi-scale relationships, enabling a wide range of biomedical applications [Hernández-Lemus *et al.*, 2024; Skaf and Laubenbacher, 2022].

## 3 Methods

### 3.1 Preliminaries

**Notations.** Formally, given a spatial transcriptomics dataset consisting of  $N$  spots and  $G$  genes, we denote the gene ex-

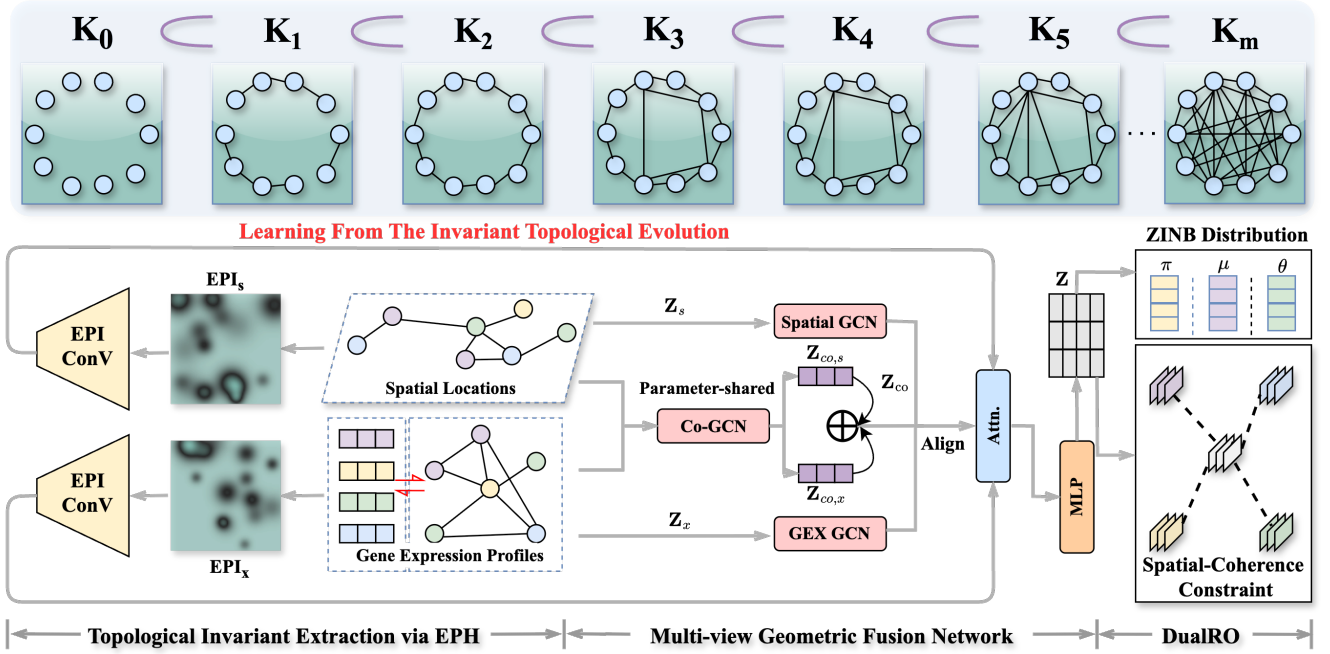


Figure 2: The proposed SPHENIC framework. Invariant topological information is incorporated into the multi-view graph fusion network to enhance the representation of different cells by EPH feature extraction. Then, the Dual-Regularized Optimization module is designed to ensure the spatial distribution and consistency of spatial clustering results.

pression matrix as  $\mathbf{X} \in \mathbb{R}^{N \times G}$ , where each row  $\mathbf{x}_i$  represents the gene expression profile of the  $i$ -th spot. Associated with these profiles is the spatial coordinate matrix  $\mathbf{S} \in \mathbb{R}^{N \times 2}$ , where  $\mathbf{s}_i$  denotes the two-dimensional physical coordinates of the  $i$ -th spot. Additionally, we formulate the learning process over the set of modalities  $\mathcal{M} = \{s, x\}$ , where  $s$  denotes the spatial domain and  $x$  denotes the gene expression profile (GEX). To capture both transcriptomic similarity and physical proximity, we model the data as a dual-view graph system over the vertex set  $\mathcal{V} = \{v_1, \dots, v_N\}$ . We define the following graphs:

- **The Spatial Graph**  $\mathcal{G}_s = (\mathcal{V}, \mathcal{E}_s, \mathbf{A}_s)$ , in which the edge set  $\mathcal{E}_s$  and the weighted adjacency matrix  $\mathbf{A}_s$  are constructed based on pairwise Euclidean distances in  $\mathbf{S}$ , thereby encoding the underlying physical tissue architecture
- **The Expression Graph**  $\mathcal{G}_x = (\mathcal{V}, \mathcal{E}_x, \mathbf{A}_x)$ , where edges connect spots with similar transcriptomic profiles derived from  $\mathbf{X}$ , capturing biological functional relationships among spots.

**Problem Formulation.** In our problem setting, the goal is to learn a mapping function  $f : (\mathbf{X}, \mathbf{S}, \mathcal{G}_s, \mathcal{G}_x) \rightarrow \mathbf{Z}$ , where  $\mathbf{Z} \in \mathbb{R}^{N \times d}$  represents a low-dimensional latent embedding with  $d \ll G$ . Ideally, the learned embedding  $\mathbf{Z}$  should satisfy three key properties: (i) *Topological Invariance*, preserving global geometric structures (e.g., loops and voids) in a noise-robust manner; (ii) *Semantic Consistency*, aiming to learn a consensus representation that effectively integrates heterogeneous gene expression and spatial profiles; and (iii) *Spatial Coherence*, ensuring that physically adjacent spots share similar embeddings, as dimensionality reduction may other-

wise erroneously separate neighboring locations in the latent space. Finally, the learned embeddings  $\mathbf{Z}$  are clustered to identify spatial domains.

### 3.2 Topological Invariant Extraction via EPH

To capture high-order structural invariants that are typically overlooked by graph neural networks, we employ Extended Persistent Homology (EPH). This approach transforms discrete graph data into rigorous topological descriptors through a unified analytical pipeline. The foundation of this pipeline is a filtration process, which constructs a sequence of nested subgraphs by varying a threshold parameter  $\epsilon$  over each modality graph. While standard filtrations often rely solely on feature similarity, we introduce an additional geometric regularizer in the expression view to suppress long-range noise.

Let  $r$  denote a fixed spatial radius, and let  $\epsilon$  represent a dynamic threshold for expression dissimilarity. At scale  $\epsilon$ , we define the edge set  $\mathcal{E}_x^\epsilon = \{(u, v) \mid \omega_x(u, v) \leq \epsilon \text{ and } \omega_s(u, v) \leq r\}$ . Based on this definition, we construct a filtration sequence  $\mathbb{F}_m$  for each modality  $m \in \mathcal{M}$ , consisting of nested subgraphs that evolve as  $\epsilon$  increases.

$$\mathbb{F}_m = \{\mathcal{G}_m^\epsilon = (\mathcal{V}, \mathcal{E}_m^\epsilon) \mid \mathcal{G}_m^{\epsilon_1} \subseteq \mathcal{G}_m^{\epsilon_2} \dots \text{ for } \epsilon_1 \leq \epsilon_2\}. \quad (1)$$

Upon constructing the filtration sequence, we track the evolution of topological features to identify significant structural invariants. By applying homology inference to  $\mathbb{F}_x$ , we track the birth and death moment of connected components and cycles, thereby obtaining the Extended Persistence Diagram (EPD). The resulting diagram is represented as a multi-set  $\mathcal{D}_m$  in  $\mathbb{R}^2$ , defined as follows:

$$\mathcal{D}_m = \text{EPH}(\mathbb{F}_m) = \{\mu_m^k = (b_m^k, d_m^k) \mid k = 1 \dots K\}, \quad (2)$$

where each point  $\mu_m^k \in \mathbb{R}^2$  represents a topological feature that is born at threshold  $b_m^k$  and persists until  $d_m^k$  for modality  $m \in \mathcal{M}$ . Although the EPD fully characterizes the topological evolution induced by the filtration sequence in the form of a strict birth–death point set, it remains an unordered and variable-length multiset that is sensitive to minor perturbations caused by noise. Also, A multiset could not be directly amenable to numerical computation. These properties hinder the direct integration of EPD into end-to-end deep learning models. Therefore, we further parameterize the EPD into a fixed-dimensional Extended Persistence Image (EPI). This process begins by assigning a weight to each feature  $\mu_k$  to emphasize robust structures with long lifespans over transient noise, using a weighting function  $w(\mu_k)$  defined as  $w(\mu_m^k) = |d_m^k - b_m^k|^\theta$ , where  $\theta$  denotes a scaling factor.

Then, we map the weighted points onto a discretized grid  $\Omega$  using Gaussian kernels. The value of the EPI at pixel  $p$ , centered at  $\mathbf{z}_p$ , is computed by summing the contributions of all features in the discrete persistence diagram  $\mathcal{D}_m$  for modality  $m \in \mathcal{M}$ , and integrating this surface over each pixel grid  $\Omega$ , which is formulated as follows:

$$\text{EPI}_m(p) = \iint_{\Omega} \sum_{\mu_m^k \in \mathcal{D}_m} \frac{w(\mu_m^k)}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{z}_p - \mu_m^k\|_2^2}{2\sigma^2}\right) d\mathbf{z}. \quad (3)$$

The resulting EPI effectively serves as the topological embedding for the following fusion module.

### 3.3 Multi-view Geometric Fusion Network

To effectively learn from both spatial and GEX data, while capturing cross-modality biological interdependencies, we construct separate views for each data modality. The resulting views are fused using a multi-view GCN network, which systematically integrates spatial organization, GEX profiles, and topological information.

**Intra-Modality Graph Encoding.** To incorporate both spatial information and GEX patterns into the proposed model, we apply graph convolutions in parallel to the symmetrically normalized Laplacian matrix  $\mathcal{T}$  of the expression-similarity adjacency matrix and the spatial adjacency matrix; we formalize the operation as follows:

$$\mathbf{Z}_m^{(l+1)} = \text{ReLU}\left(\mathcal{T}_m \mathbf{Z}_m^{(l)} \Theta_m^{(l)}\right), \quad \forall m \in \mathcal{M}, \quad (4)$$

where  $\mathbf{Z}_m^{(l)}$  denotes the output embedding of the  $l$ -th GCN layer applied to the GEX graph or spatial graph. We define  $\tilde{\mathbf{A}}_m = \mathbf{A}_m + \mathbf{I}$  as the expression or spatial similarity adjacency matrix with self-loops, while  $\tilde{\mathbf{D}}_m$  denotes its diagonal degree matrix, thus, the symmetrically normalized Laplacian matrix  $\mathcal{T}_m = \tilde{\mathbf{D}}_m^{-\frac{1}{2}} \tilde{\mathbf{A}}_m \tilde{\mathbf{D}}_m^{-\frac{1}{2}}$ .  $\Theta_m$  denotes the learnable weight matrix of the convolution layer for modality  $m$ .

**Cross-Modality Alignment & Joint Learning.** Motivated by the inherent correlation between gene-expression patterns and tissue architecture, we model both modality-specific characteristics and their shared latent representations. To this end, we employ a GCN with shared parameters to learn joint embeddings for the two modalities. We formalize this with

the following propagation rule:

$$\mathbf{Z}_{co,m}^{(l+1)} = \text{ReLU}\left(\mathcal{T}_m \mathbf{Z}_{co,m}^{(l)} \Theta_{co}^{(l)}\right), \quad \forall m \in \mathcal{M}, \quad (5)$$

where  $\mathbf{Z}_{co,m}^{(l)}$  denote the output co-spatial or co-expression embeddings of the  $l$ -th co-convolution layer, respectively.  $\Theta_{co}$  denotes the shared learnable weight matrix of the co-convolution layer. With these two co-convolution embeddings for spatial locations and GEX profiles, we compute the final consensus co-embedding  $\mathbf{Z}_{co}$  using an aggregation function  $\mathcal{F}_{agg}$  as follows:

$$\mathbf{Z}_{co}^{(l)} = \mathcal{F}_{agg}\left(\{\mathbf{Z}_{co,m}^{(l)}\}_{m \in \mathcal{M}}\right) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \mathbf{Z}_{co,m}^{(l)}. \quad (6)$$

To ensure that the co-embedding captures biological patterns shared across both modalities and that the latent representations progressively align during optimization, we introduce a consistency loss that minimizes the divergence between different modality, formalized as follows:

$$\mathcal{L}_{con} = \sum_{m,n \in \mathcal{M}, m \neq n} \|\tilde{\mathbf{Z}}_{co,m} \tilde{\mathbf{Z}}_{co,n}^T - \tilde{\mathbf{Z}}_{co,n} \tilde{\mathbf{Z}}_{co,m}^T\|_2^2, \quad (7)$$

where  $\tilde{\mathbf{Z}}_{co,m}$  and  $\tilde{\mathbf{Z}}_{co,n}$  denote the matrices obtained by normalizing  $\mathbf{Z}_{co,m}$  and  $\mathbf{Z}_{co,n}$ , respectively.

**Topological Structure Embedding.** The constructed EPI topological embedding is presented as an image and further processed using a two-dimensional convolutional layer. This layer extracts high-dimensional feature maps from the input EPI by applying a ReLU activation and a following max-pooling operation, thereby downsampling the features while preserving salient topological information. In brief, we propose the following extended persistence embedding (EPE) formulation:

$$\text{EPE}_{i,j}^{(m)} = \text{ReLU}\left(\sum_{p=0}^{k-1} \sum_{q=0}^{k-1} W_{p,q}^{(m)} \text{EPI}_{i\ell+p, j\ell+q}^{(m)} + b^{(m)}\right), \quad (8)$$

where  $\text{EPI}^{(m)}$  denote the EPI for each modality  $m \in \mathcal{M}$ .  $\mathbf{W}^{(m)}$  and  $\mathbf{b}^{(m)}$  denote the learnable weight matrices and bias vectors. While  $k$  denotes the convolution kernel size, and  $\ell$  denotes the stride length.

**Adaptive Attention Aggregation.** To allocate the contributions of different modalities, we construct an attention-fusion layer that adaptively generates attention scores for each modality. Specifically, the attention score for each modality's EPE is computed as follows:

$$\mathbf{a}_{Em} = \text{softmax}\left(\mathbf{W}_2 \sigma(\mathbf{W}_1 \text{EPE}^{(m)} + \mathbf{b}_1)\right), \quad (9)$$

where  $\mathbf{W}_1$  and  $\mathbf{W}_2$  denote learnable feature transformation matrices, respectively, and  $\mathbf{b}_1$  is the bias vector. Accordingly, we assign attention weights  $\mathbf{a}_{Em}$  to the EPE for modality  $m \in \mathcal{M}$ , attention weights  $\mathbf{a}_n$  to the GCN embedding for modality  $n \in \mathcal{M}$ , and  $\mathbf{a}_{co}$  to the spatial-expression co-embedding. We then aggregate the above embeddings using the computed attention weights to produce the final representations as follows:

$$\mathbf{Z} = \text{MLP}\left(\sum_{m \in \mathcal{M}} \mathbf{a}_{Em} \cdot \text{EPE}^{(m)} + \sum_{n \in \mathcal{M}} \mathbf{a}_n \mathbf{Z}_n + \mathbf{a}_{co} \mathbf{Z}_{co}\right). \quad (10)$$

### 3.4 Dual-Regularized Optimization

Spatial transcriptomics data naturally contain spatial location information and exhibit a sparse, zero-inflated distribution. To leverage these two critical properties, we developed double design-optimization modules.

**Spatial-Coherence Constraint.** To enhance the model’s ability to learn from spatial locations, we propose a spatial-constraint module that explicitly preserves local neighborhood relationships in the embedding space. In brief, it operates via a two-part mechanism: (i) it increases the similarity between the embedding of cell  $i$  and those of its spatial neighbors ( $j \in \mathcal{N}_i$ ), and (ii) it decreases the similarity with non-neighboring cells for both the co-embedding and the view-specific embeddings. This strategy is formalized as follows:

$$\mathcal{L}_{sco} = - \sum_{i=1}^N \sum_{j \in \mathcal{N}_i} \left[ y_{ij} \log(\sigma(S_{ij}) \cdot \Phi_{ii}^+) + (1 - y_{ij}) \log((1 - \sigma(S_{ij})) \cdot \Phi_{ij}^-) \right], \quad (11)$$

where  $y_{ij} = 1$  if  $j \in \mathcal{N}_i$ . And  $S_{ij} = C(\mathbf{Z}_i, \mathbf{Z}_j)$  is the cosine similarity between two vectors, while  $\sigma$  denotes the hyperbolic tanh activation function. View-specific embedding  $\Phi_{ii}^+ = \sum_{v=1}^V \exp(C(\mathbf{Z}_i, \mathbf{Z}_i^v))$  and  $\Phi_{ij}^- = \sum_{v=1}^V \exp((1 - S_{ij}) \cdot C(\mathbf{Z}_i, \mathbf{Z}_j^v))$ .  $N$  denotes the number of cells in the spatial slice, and  $V$  denotes the number of views.

**ZINB Probabilistic Reconstruction.** Spatial transcriptomics data exhibit characteristic statistical properties: discrete counts with overdispersion and high sparsity. These properties can be modeled by a zero-inflated negative binomial (ZINB) distribution, whose mathematical form is given below:

$$\text{NB}(X = x | \mu, \theta) = \frac{\Gamma(x + \theta)}{x! \Gamma(\theta)} \left( \frac{\theta}{\theta + \mu} \right)^\theta \left( \frac{\mu}{\theta + \mu} \right)^x, \quad (12)$$

$$\text{ZINB}(X = x | \pi, \mu, \theta) = \pi I(x) + (1 - \pi) \text{NB}(x), \quad (13)$$

where  $\mu$  and  $\theta$  denote the mean and dispersion of gene expression, while  $\pi$  denotes the probability of zero inflation,  $I(\cdot)$  denote the indicator function. To incorporate these characteristics into spatial domain identification, we employ a three-layer MLP  $f(\cdot)$  to estimate the zero-inflation probability matrix  $\Pi$ , the mean matrix  $\mathbf{M}$ , and the dispersion matrix  $\Theta$ , as follows:

$$\begin{aligned} \Pi &= \text{sigmoid}(\mathbf{W}_\pi f(\mathbf{Z})); \\ \mathbf{M} &= \exp(\mathbf{W}_\mu f(\mathbf{Z})); \\ \Theta &= \exp(\mathbf{W}_\theta f(\mathbf{Z})), \end{aligned} \quad (14)$$

where  $\mathbf{W}_\pi$ ,  $\mathbf{W}_\mu$ , and  $\mathbf{W}_\theta$  denote the learnable weight matrices for  $\Pi$ ,  $\mathbf{M}$ , and  $\Theta$ , respectively. To learn the model parameters, we minimize the negative log-likelihood of the ZINB model with respect to the observed  $X$ , which can be defined as follows:

$$\mathcal{L}_{\text{rec}} = - \sum \log(\text{ZINB}(X | \pi, \mu, \theta)). \quad (15)$$

**Total Objective.** Ultimately, the fused embedding is jointly optimized using the consistency loss, the ZINB loss, and the spatial-constraint loss. We minimize the following total loss:

$$\mathcal{L}_f = \mathcal{L}_{\text{rec}} + \lambda_1 \mathcal{L}_{\text{con}} + \lambda_2 \mathcal{L}_{\text{sco}}, \quad (16)$$

where  $\lambda_1$  and  $\lambda_2$  denote the trade-off hyperparameters for respective loss function terms.

## 4 Experiments

### 4.1 Experiment Settings

**Benchmark Datasets and Evaluation.** The proposed SPHENIC model was evaluated on three publicly available spatial transcriptomics datasets, DLPFC (comprising 9 distinct tissue sections), 10x Visium human breast cancer dataset (HBC), and mouse brain anterior dataset (MBA) [Maynard *et al.*, 2021; Buache *et al.*, 2011]. Clustering performance was assessed using the adjusted Rand index (ARI) and normalized mutual information (NMI).

**Compared Baseline Methods.** To enable a fair comparison, we benchmark our model against eight current representative baseline methods: SCANPY [Wolf *et al.*, 2018], SpaGCN [Hu *et al.*, 2021], DeepST [Xu *et al.*, 2022], stLearn [Pham *et al.*, 2023], SCGDL [Liu *et al.*, 2023], GraphST [Long *et al.*, 2023], Spatial-MGCN [Wang *et al.*, 2023] and STAIG [Yang *et al.*, 2025].

**Implementation Details.** All experiments were conducted on a single NVIDIA virtual GPU (vGPU) rented from AutoDL, providing up to 48 GB of memory and a 25-core Intel Xeon Platinum 8481C processor. The model was implemented in PyTorch 2.1.0 and executed with CUDA 12.1. For each dataset, training was terminated after 100 epochs. All baseline methods were run with the default hyperparameters recommended in their original publications.

### 4.2 Clustering Performance

The comparative performance of nine methods is summarized in Table 1. Fig. 3 provides a visual comparison of the clustering outputs produced by each method. The above results reveal that:

(i) SPHENIC delivers state-of-the-art clustering accuracy in architecturally complex tissues. With respect to the ARI term, SPHENIC exceeds the next-best method (Spatial-MGCN) by 4.19% on the HBC dataset and 9.14% on the mouse-anterior dataset compared to the second-best method (SCANPY), underscoring its strength in identifying domains within challenging tissue architectures.

(ii) In spatially simpler structures such as the DLPFC slices, SPHENIC surpasses the next-best method (Spatial-MGCN for slice 151507, 151508, and 151569) by 2.65%, 2.88%, and 3.90% ARI, respectively. These findings confirm SPHENIC’s generalizability and its ability to extract fundamental biological signals irrespective of spatial complexity.

(iii) SPHENIC accurately delineates tumor boundaries in the HBC dataset, as illustrated in the Fig. 3. Our SPHENIC shows strong agreement with manual annotations in identifying key tumor-microenvironment components (e.g., ‘IDC\_4’, ‘healthy\_1’). Unlike baselines that yield fragmented clusters and tumor-immune misclassifications, SPHENIC preserves biologically valid structures, achieving expert-level spatial recognition in carcinomas.

Additionally, we performed scalability analysis of all the baselines in the HBC dataset, the 151507 dataset, and the MBA dataset, which are demonstrated in the supplementary materials.

| Datasets | ARI               |                  |                         |                  |                  |                  |                         |                         |                         |
|----------|-------------------|------------------|-------------------------|------------------|------------------|------------------|-------------------------|-------------------------|-------------------------|
|          | SCANPY'18         | SpaGCN'21        | DeepST'22               | stLearn'23       | SCGDL'23         | GraphST'23       | Spatial-MGCN'23         | STAIG'25                | SPHENIC                 |
| HBC      | 48.52 $\pm$ 0.14  | 56.79 $\pm$ 2.81 | 56.30 $\pm$ 0.89        | 55.02 $\pm$ 3.36 | 32.88 $\pm$ 2.51 | 51.83 $\pm$ 0.57 | <u>64.04</u> $\pm$ 3.24 | 50.69 $\pm$ 2.65        | <b>68.23</b> $\pm$ 2.36 |
| MBA      | 39.65 $\pm$ 0.95  | 34.42 $\pm$ 0.89 | 25.91 $\pm$ 3.43        | 34.13 $\pm$ 0.19 | 25.81 $\pm$ 1.23 | 29.33 $\pm$ 0.13 | 39.58 $\pm$ 0.02        | 37.01 $\pm$ 0.71        | <b>48.79</b> $\pm$ 1.68 |
| 151507   | 25.53 $\pm$ 1.23  | 38.26 $\pm$ 3.05 | 52.24 $\pm$ 2.76        | 42.95 $\pm$ 1.04 | 46.26 $\pm$ 2.41 | 44.79 $\pm$ 0.30 | <u>62.68</u> $\pm$ 4.60 | 54.22 $\pm$ 1.68        | <b>65.33</b> $\pm$ 1.74 |
| 151508   | 14.25 $\pm$ 0.70  | 39.48 $\pm$ 2.15 | 42.89 $\pm$ 5.77        | 40.03 $\pm$ 0.68 | 35.54 $\pm$ 1.82 | 35.18 $\pm$ 1.01 | <u>50.66</u> $\pm$ 1.04 | 47.69 $\pm$ 3.98        | <b>53.54</b> $\pm$ 4.89 |
| 151509   | 14.31 $\pm$ 1.17  | 34.08 $\pm$ 2.39 | 47.60 $\pm$ 6.47        | 29.52 $\pm$ 0.75 | 37.37 $\pm$ 3.40 | 36.71 $\pm$ 5.81 | <u>58.10</u> $\pm$ 3.87 | 55.02 $\pm$ 0.46        | <b>60.51</b> $\pm$ 6.14 |
| 151510   | 17.04 $\pm$ 0.52  | 38.36 $\pm$ 3.26 | 46.57 $\pm$ 2.11        | 34.09 $\pm$ 4.33 | 30.57 $\pm$ 0.34 | 39.73 $\pm$ 4.05 | <b>57.06</b> $\pm$ 3.64 | 51.26 $\pm$ 0.21        | <u>53.97</u> $\pm$ 3.41 |
| 151669   | 25.50 $\pm$ 2.86  | 26.70 $\pm$ 4.79 | 27.73 $\pm$ 1.81        | 25.23 $\pm$ 0.19 | 25.15 $\pm$ 0.84 | 29.98 $\pm$ 5.75 | <u>45.14</u> $\pm$ 5.91 | 33.40 $\pm$ 2.65        | <b>49.04</b> $\pm$ 5.35 |
| 151670   | 24.34 $\pm$ 1.61  | 23.80 $\pm$ 2.88 | 28.56 $\pm$ 2.76        | 23.46 $\pm$ 0.60 | 20.94 $\pm$ 1.59 | 25.97 $\pm$ 1.45 | <b>42.67</b> $\pm$ 5.79 | 24.62 $\pm$ 2.96        | <u>42.60</u> $\pm$ 5.61 |
| 151671   | 21.97 $\pm$ 1.58  | 40.76 $\pm$ 1.08 | 37.24 $\pm$ 3.75        | 33.22 $\pm$ 4.78 | 23.43 $\pm$ 0.26 | 38.83 $\pm$ 0.95 | <b>65.75</b> $\pm$ 0.61 | 48.87 $\pm$ 5.95        | <u>65.14</u> $\pm$ 6.40 |
| 151673   | 17.46 $\pm$ 1.90  | 44.73 $\pm$ 1.04 | 52.82 $\pm$ 0.54        | 35.34 $\pm$ 2.69 | 32.48 $\pm$ 1.03 | 48.00 $\pm$ 2.28 | <u>56.92</u> $\pm$ 4.93 | 51.88 $\pm$ 5.73        | <b>58.17</b> $\pm$ 2.14 |
| 151674   | 23.95 $\pm$ 1.97  | 33.26 $\pm$ 2.97 | 54.51 $\pm$ 3.19        | 30.60 $\pm$ 2.76 | 23.41 $\pm$ 0.78 | 44.03 $\pm$ 3.65 | <u>56.10</u> $\pm$ 1.71 | 56.03 $\pm$ 3.84        | <b>56.41</b> $\pm$ 2.86 |
| Avg      | 24.78 $\pm$ 10.62 | 37.33 $\pm$ 8.85 | 42.94 $\pm$ 11.37       | 34.87 $\pm$ 8.79 | 30.35 $\pm$ 7.57 | 38.58 $\pm$ 8.17 | <u>54.43</u> $\pm$ 8.82 | 46.43 $\pm$ 10.20       | <b>56.52</b> $\pm$ 7.95 |
|          |                   |                  |                         |                  |                  |                  |                         |                         |                         |
| Datasets | NMI               |                  |                         |                  |                  |                  |                         |                         |                         |
|          | SCANPY'18         | SpaGCN'21        | DeepST'22               | stLearn'23       | SCGDL'23         | GraphST'23       | Spatial-MGCN'23         | STAIG'25                | SPHENIC                 |
| HBC      | 60.60 $\pm$ 0.21  | 65.71 $\pm$ 0.59 | 69.33 $\pm$ 0.58        | 66.47 $\pm$ 1.84 | 48.28 $\pm$ 5.68 | 65.86 $\pm$ 0.39 | <u>69.41</u> $\pm$ 1.88 | 68.08 $\pm$ 1.54        | <b>69.54</b> $\pm$ 1.88 |
| MBA      | 63.89 $\pm$ 0.19  | 64.88 $\pm$ 1.51 | 55.23 $\pm$ 2.75        | 64.55 $\pm$ 0.04 | 62.04 $\pm$ 0.36 | 67.69 $\pm$ 0.15 | 63.65 $\pm$ 1.99        | <u>67.75</u> $\pm$ 0.14 | <b>69.09</b> $\pm$ 1.65 |
| 151507   | 34.36 $\pm$ 0.67  | 49.61 $\pm$ 3.40 | 64.54 $\pm$ 3.99        | 54.35 $\pm$ 0.55 | 52.96 $\pm$ 1.26 | 57.40 $\pm$ 0.69 | <u>70.92</u> $\pm$ 1.03 | 68.49 $\pm$ 2.00        | <b>74.08</b> $\pm$ 2.13 |
| 151508   | 19.54 $\pm$ 1.03  | 51.04 $\pm$ 1.35 | 56.81 $\pm$ 4.56        | 51.57 $\pm$ 0.76 | 45.61 $\pm$ 0.16 | 52.57 $\pm$ 0.60 | 60.42 $\pm$ 2.38        | <u>61.93</u> $\pm$ 2.18 | <b>63.12</b> $\pm$ 1.16 |
| 151509   | 26.00 $\pm$ 0.51  | 51.35 $\pm$ 3.52 | 62.92 $\pm$ 1.87        | 49.01 $\pm$ 0.96 | 51.34 $\pm$ 1.18 | 53.74 $\pm$ 1.58 | 67.08 $\pm$ 2.28        | <u>67.83</u> $\pm$ 0.21 | <b>69.61</b> $\pm$ 0.27 |
| 151510   | 26.29 $\pm$ 1.93  | 49.88 $\pm$ 3.84 | 61.26 $\pm$ 1.08        | 47.77 $\pm$ 2.03 | 47.34 $\pm$ 0.42 | 54.70 $\pm$ 0.70 | 65.39 $\pm$ 2.72        | <u>65.44</u> $\pm$ 1.25 | <b>65.63</b> $\pm$ 0.88 |
| 151669   | 26.79 $\pm$ 1.08  | 45.41 $\pm$ 6.29 | 52.16 $\pm$ 0.93        | 45.18 $\pm$ 0.67 | 41.26 $\pm$ 0.61 | 51.02 $\pm$ 2.71 | <u>59.48</u> $\pm$ 6.34 | 49.48 $\pm$ 1.88        | <b>61.84</b> $\pm$ 0.48 |
| 151670   | 24.73 $\pm$ 4.05  | 42.24 $\pm$ 1.90 | 51.65 $\pm$ 2.08        | 43.79 $\pm$ 1.22 | 36.75 $\pm$ 1.30 | 45.76 $\pm$ 1.64 | <u>55.13</u> $\pm$ 2.60 | 48.86 $\pm$ 1.29        | <b>60.23</b> $\pm$ 1.52 |
| 151671   | 23.34 $\pm$ 2.53  | 55.50 $\pm$ 1.48 | 58.84 $\pm$ 3.32        | 50.76 $\pm$ 1.53 | 40.04 $\pm$ 0.45 | 57.27 $\pm$ 0.77 | 63.41 $\pm$ 0.42        | <u>66.78</u> $\pm$ 2.72 | <b>73.50</b> $\pm$ 2.19 |
| 151673   | 29.44 $\pm$ 1.73  | 59.08 $\pm$ 1.20 | <u>67.71</u> $\pm$ 0.66 | 51.92 $\pm$ 0.76 | 43.04 $\pm$ 0.46 | 62.54 $\pm$ 1.46 | 67.08 $\pm$ 2.14        | 67.36 $\pm$ 2.64        | <b>68.63</b> $\pm$ 0.94 |
| 151674   | 30.66 $\pm$ 1.39  | 45.86 $\pm$ 1.48 | 64.31 $\pm$ 2.88        | 49.52 $\pm$ 2.22 | 36.33 $\pm$ 0.93 | 55.32 $\pm$ 2.48 | 64.94 $\pm$ 1.93        | <u>66.97</u> $\pm$ 1.86 | <b>65.56</b> $\pm$ 1.21 |
| Avg      | 33.24 $\pm$ 14.87 | 52.78 $\pm$ 7.73 | 60.43 $\pm$ 5.99        | 52.26 $\pm$ 7.21 | 45.91 $\pm$ 7.66 | 56.72 $\pm$ 6.51 | <u>64.26</u> $\pm$ 4.59 | 63.54 $\pm$ 7.33        | <b>67.30</b> $\pm$ 4.48 |

Table 1: Comparison with state-of-the-art methods on the eleven distinct spatial transcriptomic slices. The optimal and second-best results are denoted by **boldface** and underlining, respectively.  $\uparrow$  means improved performance compared with the second-best baseline.

| Dataset | Module   |        | Metric                  |                         |
|---------|----------|--------|-------------------------|-------------------------|
|         | Topo     | DualRO | ARI                     | NMI                     |
| HBC     | baseline |        | 52.89 $\pm$ 2.38        | 62.61 $\pm$ 2.14        |
|         | ✓        |        | 57.99 $\pm$ 2.73        | 63.93 $\pm$ 2.54        |
|         |          | ✓      | 63.39 $\pm$ 4.23        | 64.94 $\pm$ 2.71        |
|         | ✓        | ✓      | <b>68.23</b> $\pm$ 2.36 | <b>69.54</b> $\pm$ 1.88 |
| 151507  | baseline |        | 53.22 $\pm$ 1.37        | 63.79 $\pm$ 1.33        |
|         | ✓        |        | 59.92 $\pm$ 3.08        | 66.95 $\pm$ 2.06        |
|         |          | ✓      | 62.03 $\pm$ 2.65        | 71.78 $\pm$ 1.70        |
|         | ✓        | ✓      | <b>65.33</b> $\pm$ 1.74 | <b>74.08</b> $\pm$ 2.13 |

Table 2: Ablative study of key modules in our SPHENIC on HBC and 151507 datasets. This table demonstrates the effects of the EPH module and Dual-Regularized Optimization over the baselines, which remove these modules respectively or simultaneously. 'Topo' means EPH module.

### 4.3 Ablation Study

To demonstrate the effectiveness of the proposed modules, we construct several variants of SPHENIC: (i) Omits the invariant topological feature-extraction module; (ii) Excludes the dual-regularized optimization module; and (iii) Removes both modules. As shown in the Table 2, in the HBC dataset, omitting the topological module ('w/o Topo') reduces ARI by 4.84 % and 3.30 % relative to the full model, thereby validating the contribution of the topological learning component. Similarly, removing the dual-regularized optimization module ('w/o DualRO') decreases ARI by 10.24 % and 5.61 %, confirming the importance of explicitly modeling the spatial

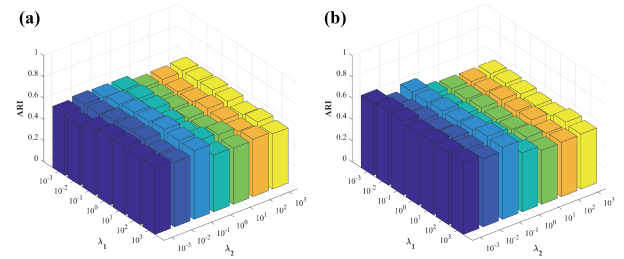


Figure 4: Hyperparameter sensitivity analysis on HBC dataset and DLPFC-151507, respectively, in ARI term.

neighbourhoods and non-neighbourhoods. We further evaluate a variant lacking both modules. The complete SPHENIC model improves ARI by 15.34 % and 6.93 % over this variant, demonstrating the synergistic benefit of combining topological and spatial-constraint learning. Ablation experiment on the MBA dataset is demonstrated in the supplementary materials.

### 4.4 Model Analysis

**Hyperparameter Sensitivity Analysis.** We evaluate the effect of different hyperparameters on the performance of our proposed SPHENIC in the HBC dataset, which includes the trade-off coefficients of the total loss function ( $\lambda_1$  and  $\lambda_2$ ). Fig. 4 demonstrates the ARI term of our SPHENIC when  $\lambda_1$  and  $\lambda_2$  are varied from  $10^{-3}$  to  $10^3$ . From this figure, the proposed SPHENIC is insensitive to both  $\lambda_1$  and  $\lambda_2$  in

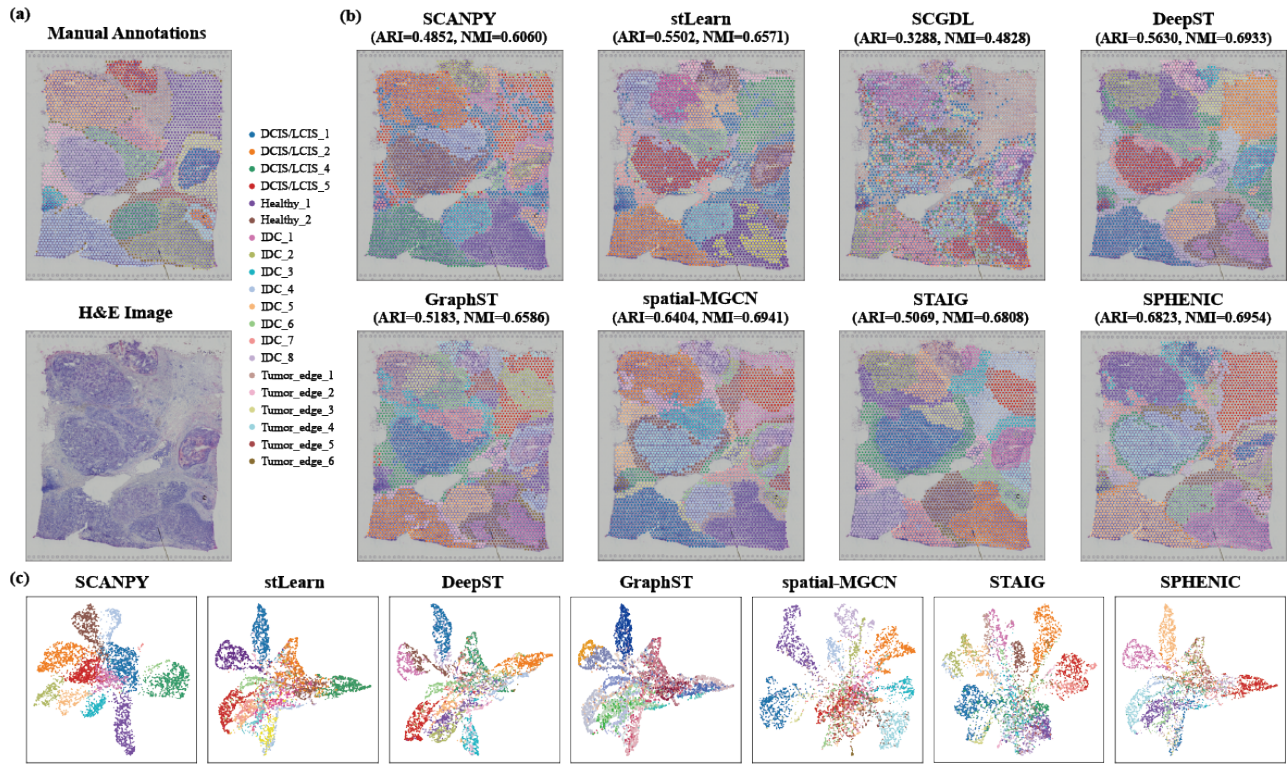


Figure 3: Comparison experiments on HBC dataset between SPHENIC and other baselines. (a) Manual ground-truth annotations and histological image of the HBC dataset. (b) Spatial clustering result visualization by SCANPY, stLearn, SCGDL, DeepST, GraphST, spatial-MGCN, STAIG, and SPHENIC. (c) UMAP visualization of detected clusters by the above baselines.

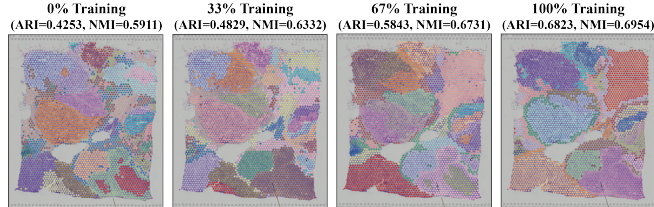


Figure 5: Visualization of spatial clustering by SPHENIC model in different training stages

the range  $10^{-3}$  to  $10^{-1}$ , further illustrating the robustness of our proposed SPHENIC. Sensitivity analysis of the EPI hyperparameters is further demonstrated in the supplementary materials.

**Visualization Analysis.** In addition, to further verify the effectiveness of our SPHENIC, we visualize the spatial clustering results of the proposed SPHENIC in different training stages (0%, 33%, 67% and 100% of training epochs) by the UMAP visualization methods in Fig. 5 [McInnes *et al.*, 2018]. We can clearly observe that the spatial clustering structure of our comprehensive multi-view fused embedding  $\mathbf{Z}$  progressively refines across the training epochs, along with consistent improvement in both ARI and NMI terms.

**Gene Enhancement Analysis.** To validate the proposed SPHENIC’s dual-module optimization strategy in preserving spatial expression patterns while reducing the noise artifacts, we evaluated its performance on the HBC dataset by exam-

ining the spatial distributions of characteristic marker genes (*DSP*, *IFT122*, *CCL21*, *CTTN*, *GNG5*, *FARP1*). Fig. S1 in the supplementary materials shows that SPHENIC successfully reconstructs biologically plausible expression patterns while significantly reducing technical noise compared to raw data. The system maintains consistent spatial distributions of these markers that align with established biological knowledge, confirming its ability to (i) filter out non-informative noise, (ii) recover true biological signals from degraded data, and (iii) faithfully reconstruct spatial expression profiles.

## 5 Conclusions

In this paper, we present SPHENIC, a topology-aware clustering framework for spatial transcriptomics data. By leveraging extended persistent homology features, the proposed model learns more robust topological information and is less sensitive to intrinsic noise. Meanwhile, we leverage distribution constraints and spatial coordinates to supervise the embedding space through a novel DualRO module, forcing proximity between a cell’s embedding and those of its spatial neighbors while separating it from non-neighboring cells, thereby producing high-quality spatial embeddings for applications. Extensive validation experiments demonstrate the superiority of the proposed SPHENIC. Nevertheless, the principal limitation of our approach is the computational overhead associated with extended persistent homology, which may impede scalability in application scenarios.

## References

- [Blondel *et al.*, 2008] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- [Buache *et al.*, 2011] E Buache, N Etique, F Alpy, I Stoll, M Muckensturm, B Reina-San-Martin, MP Chenard, C Tomasetto, and MC Rio. Deficiency in trefoil factor 1 (tff1) increases tumorigenicity of human breast cancer cells and mammary tumor development in tff1-knockout mice. *Oncogene*, 30(29):3261–3273, 2011.
- [Chen *et al.*, 2019] Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A topological regularizer for classifiers via persistent homology. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 2573–2582. PMLR, 16–18 Apr 2019.
- [Chen *et al.*, 2024] Yuzhou Chen, Jose Frias, and Yulia R Gel. Topogcl: Topological graph contrastive learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 11453–11461, 2024.
- [Gilmer *et al.*, 2020] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Message passing neural networks. In *Machine learning meets quantum physics*, pages 199–214. Springer, 2020.
- [Gulati *et al.*, 2025] Gunsagar S Gulati, Jeremy Philip D’Silva, Yunhe Liu, Linghua Wang, and Aaron M Newman. Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nature Reviews Molecular Cell Biology*, 26(1):11–31, 2025.
- [Guo *et al.*, 2025] Chen-Kai Guo, Chen-Rui Xia, Guangdun Peng, Zhi-Jie Cao, and Ge Gao. Learning phenotype associated signature in spatial transcriptomics with passage. *Small Methods*, 9(5):2401451, 2025.
- [Hernández-Lemus *et al.*, 2024] Enrique Hernández-Lemus, Pedro Miramontes, and Mireya Martínez-García. Topological data analysis in cardiovascular signals: An overview. *Entropy*, 26(1):67, 2024.
- [Hu *et al.*, 2021] Jian Hu, Xiangjie Li, Kyle Coleman, Amelia Schroeder, Nan Ma, David J Irwin, Edward B Lee, Russell T Shinohara, and Mingyao Li. 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.
- [Hu *et al.*, 2024] Yunfei Hu, Manfei Xie, Yikang Li, Mingxing Rao, Wenjun Shen, Can Luo, Haoran Qin, Jihoon Baek, and Xin Maizie Zhou. Benchmarking clustering, alignment, and integration methods for spatial transcriptomics. *Genome Biology*, 25(1):212, 2024.
- [Jain and Eadon, 2024] Sanjay Jain and Michael T Eadon. Spatial transcriptomics in health and disease. *Nature Reviews Nephrology*, 20(10):659–671, 2024.
- [Kipf, 2016] TN Kipf. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- [Li *et al.*, 2025] Sijie Li, Heyang Hua, and Shengquan Chen. Graph neural networks for single-cell omics data: a review of approaches and applications. *Briefings in Bioinformatics*, 26(2), 2025.
- [Liu *et al.*, 2023] Teng Liu, Zhao-Yu Fang, Xin Li, Li-Ning Zhang, Dong-Sheng Cao, and Ming-Zhu Yin. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *Briefings in Bioinformatics*, 24(3):bbad146, 2023.
- [Liu *et al.*, 2024] Teng Liu, Zhao-Yu Fang, Zongbo Zhang, Yongxiang Yu, Min Li, and Ming-Zhu Yin. A comprehensive overview of graph neural network-based approaches to clustering for spatial transcriptomics. *Computational and Structural Biotechnology Journal*, 23:106–128, 2024.
- [Long *et al.*, 2023] Yahui Long, Kok Siong Ang, Mengwei Li, Kian Long Kelvin Chong, Raman Sethi, Chengwei Zhong, Hang Xu, Zhiwei Ong, Karishma Sachaphibulkij, Ao Chen, et al. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications*, 14(1):1155, 2023.
- [Maynard *et al.*, 2021] Kristen R Maynard, Leonardo Collado-Torres, Lukas M Weber, Cedric Uytingco, Brianna K Barry, Stephen R Williams, Joseph L Catallini, Matthew N Tran, Zachary Besich, Madhavi Tippani, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience*, 24(3):425–436, 2021.
- [McInnes *et al.*, 2018] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [Moses and Pachter, 2022] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5):534–546, 2022.
- [Pham *et al.*, 2023] Duy Pham, Xiao Tan, Brad Balderson, Jun Xu, Laura F Grice, Sohye Yoon, Emily F Willis, Minh Tran, Pui Yeng Lam, Arti Raghubar, et al. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature communications*, 14(1):7739, 2023.
- [Pham *et al.*, 2025] Phu Pham, Quang-Thinh Bui, Ngoc Thanh Nguyen, Robert Kozma, Philip S Yu, and Bay Vo. Topological data analysis in graph neural networks: Surveys and perspectives. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.
- [Rao *et al.*, 2021] Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.
- [Satija *et al.*, 2015] Rahul Satija, Jeffrey A Farrell, David Gennert, Alexander F Schier, and Aviv Regev. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*, 33(5):495–502, 2015.

- [Skaf and Laubenbacher, 2022] Yara Skaf and Reinhard Laubenbacher. Topological data analysis in biomedicine: A review. *Journal of Biomedical Informatics*, 130:104082, 2022.
- [Su *et al.*, 2025] Zhe Su, Xiang Liu, Layal Bou Hamdan, Vasileios Maroulas, Jie Wu, Gunnar Carlsson, and Guo-Wei Wei. Topological data analysis and topological deep learning beyond persistent homology-a review. *arXiv preprint arXiv:2507.19504*, 2025.
- [Tian *et al.*, 2023] Luyi Tian, Fei Chen, and Evan Z Mascosko. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6):773–782, 2023.
- [Traag *et al.*, 2019] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.
- [Wang *et al.*, 2023] Bo Wang, Jiawei Luo, Ying Liu, Wanwan Shi, Zehao Xiong, Cong Shen, and Yahui Long. Spatial-mgcn: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. *Briefings in Bioinformatics*, 24(5):bbad262, 2023.
- [Wang *et al.*, 2024] Ziyi Wang, Aoyun Geng, Hao Duan, Feifei Cui, Quan Zou, and Zilong Zhang. A comprehensive review of approaches for spatial domain recognition of spatial transcriptomes. *Briefings in Functional Genomics*, 23(6):702–712, 2024.
- [Wasserman, 2018] Larry Wasserman. Topological data analysis. *Annual review of statistics and its application*, 5(2018):501–532, 2018.
- [Williams *et al.*, 2022] Cameron G Williams, Hyun Jae Lee, Takahiro Asatsuma, Roser Vento-Tormo, and Ashraf Haque. An introduction to spatial transcriptomics for biomedical research. *Genome medicine*, 14(1):68, 2022.
- [Wolf *et al.*, 2018] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):15, 2018.
- [Xu *et al.*, 2022] Chang Xu, Xiyun Jin, Songren Wei, Pingping Wang, Meng Luo, Zhaochun Xu, Wenyi Yang, Yideng Cai, Lixing Xiao, Xiaoyu Lin, et al. Deepst: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research*, 50(22):e131–e131, 2022.
- [Yang *et al.*, 2025] Yitao Yang, Yang Cui, Xin Zeng, Yubo Zhang, Martin Loza, Sung-Joon Park, and Kenta Nakai. Staig: Spatial transcriptomics analysis via image-aided graph contrastive learning for domain exploration and alignment-free integration. *Nature Communications*, 16(1):1067, 2025.
- [Yuan *et al.*, 2024] Zhiyuan Yuan, Fangyuan Zhao, Senlin Lin, Yu Zhao, Jianhua Yao, Yan Cui, Xiao-Yong Zhang, and Yi Zhao. Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nature Methods*, 21(4):712–722, 2024.
- [Zahedi *et al.*, 2024] Roxana Zahedi, Reza Ghamsari, Ahmadreza Argha, Callum Macphillamy, Amin Beheshti, Roohallah Alizadehsani, Nigel H Lovell, Mohammad Lotfollahi, and Hamid Alinejad-Rokny. Deep learning in spatially resolved transcriptomics: a comprehensive technical view. *Briefings in Bioinformatics*, 25(2):bbae082, 2024.
- [Zhang *et al.*, 2025] Yuyang Zhang, Yuhang Liu, Zinnia Ma, Min Li, Chunfu Xu, and Haipeng Gong. Improving diffusion-based protein backbone generation with global-geometry-aware latent encoding. *Nature Machine Intelligence*, pages 1–15, 2025.