Spatially Resolved Transcriptomics Data Clustering with Tailored Spatial-scale Modulation

Yuang Xiao[1], **Yanran Zhu**[1], **Chang Tang**[2 *], **Xiao Zheng**[3], **Yuanyuan Liu**[1], **Kun Sun**[1] and **Xinwang Liu**[4]
[1]School of Computer Science, China University of Geosciences, Wuhan, China
[2]School of Software Engineering, Huazhong University of Science and Technology, China
[3] School of Computer Science, Hubei University of Technology, China
[4] College of Computer Science and Technology, National University of Defense Technology, China
{leon_shawn, zhuyanran, tangchang, liuyy}@cug.edu.cn, {zhengxiao, xinwangliu}@nudt.edu.cn, sunkun.hust@gmail.com

Abstract

Spatial transcriptomics, comprising spatial location and high-throughput gene expression information, provides revolutionary insights into disease discovery and cellular evolution. Spatial transcriptomic clustering, which pinpoints distinct spatial domains within tissues, reveals cellular interactions and enhances our understanding of the intricate architecture of tissues. Existing methods typically construct spatial graphs using a static radius based on spatial coordinates, which hinders the accurate identification of spatial domains and complicates the precise partitioning of boundary nodes within clusters. To address this issue, we introduce a novel spatially resolved transcriptomics data clustering network (TSstc). Specifically, we employ a tailored spatial-scale modulation approach, constructing different spatial graphs incrementally as the radius of the spatial domain expands, and a Spatiality-Aware Sampling (SAS) strategy is proposed to aggregate node representations by considering the spatial dependencies between spots. We then use GCN encoders to learn gene embedding with gene graphs and multiple spatial embeddings with spatial graphs. During training, we incorporate cross-view correlation-based tailored spatial regularization constraints to preserve high-quality neighbor relationships across spatial embeddings at different scales. Finally, a zero-inflated negative binomial model is utilized to capture the global probability distribution of gene expression profiles. Extensive experimental results demonstrate that our approach surpasses existing state-of-the-art methods in clustering tasks and related downstream applications.

1 Introduction

Spatial transcriptomics technology [Song and Su, 2021; Zhai *et al.*, 2023; Rao *et al.*, 2021] provides a novel approach to studying the functions and interactions of cells within tissues. Offering quantitative gene expression data alongside the spatial location of cells overcomes the limitations of single-cell sequencing [Mrabah *et al.*, 2023], particularly the loss of spatial information. In recent years, spatial transcriptomics has been widely applied across various fields, including embryonic development [Rossant and Tam, 2022], disease research [Moore *et al.*, 2024], and plant science [van Dijk *et al.*, 2021].

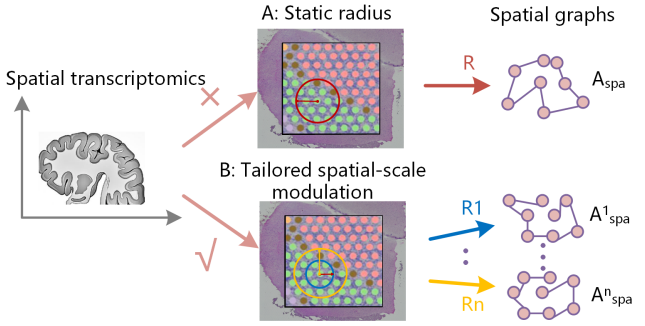

Figure 1: It is a histopathological image composed of numerous spots with different colors representing distinct clusters. The way A displays a standard static radius-based approach for constructing spatial graphs. The way B illustrates our proposed tailored spatial-scale modulation method for constructing spatial graphs.

Accurate identification of spatial domains within transcriptomic data enables a deeper understanding of the distribution patterns of cell populations in tissues and their biological functions at specific spatial locations. Early methods for spatial domain identification primarily relied on gene expression data for clustering analysis, such as K-means [Ahmed *et al.*, 2020], Louvain [Zhang *et al.*, 2021], and Seurat [Pereira *et al.*, 2021] algorithms. However, these methods neglected spatial relationships between cells, often leading to discontinuous domain identification, where adjacent cells were mistakenly assigned to separate domains.

Consequently, researchers have increasingly explored deep clustering methods integrating gene expression data with spatial location information. For example, stGCL [He *et al.*,

*Corresponding author.

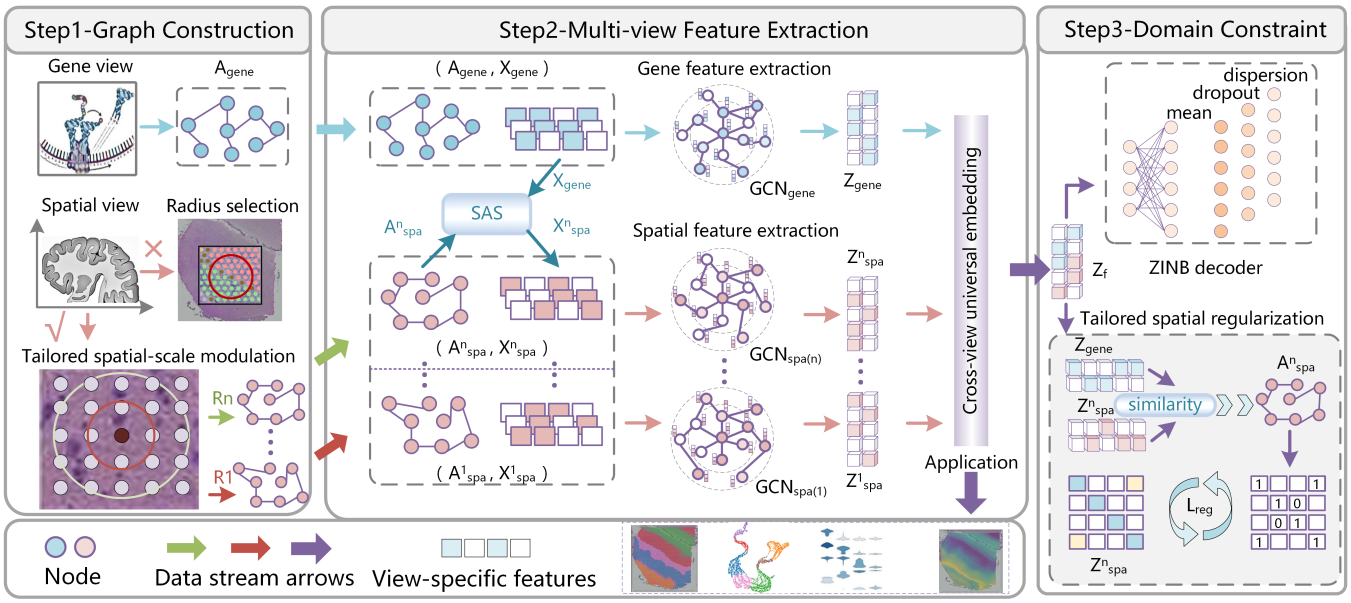

Figure 2: The architecture of TSstc. (A) Graph Construction module: we employ a tailored spatial-scale modulation approach to construct different spatial graphs $\mathbf{A}^v_{spa}$ incrementally as the radius of the spatial domain expands. (B) Multi-view Feature Extraction module: we aggregate node representations to obtain different spatial feature matrixes $\mathbf{X}^v_{spa}$ by the Spatiality-Aware Sampling (SAS) strategy. Then, we use GCN encoders to learn gene embedding $\mathbf{Z}_{gene}$ and multiple spatial embeddings $\mathbf{Z}^v_{spa}$. Concurrently, the Cross-view Fusion module is proposed to integrate them for $\mathbf{Z}_f$. (C) Spatial Domain Constraints module: we incorporate cross-view correlation-based tailored spatial regularization constraints to preserve high-quality neighbor relationships across spatial embeddings $\mathbf{Z}^v_{spa}$ at different scales and utilize a ZINB decoder model to reconstruct the original expression matrix.

2024] constructs a heterogeneous graph coordinating module that maintains view independence during the fusion process. stMMR [Zhang *et al.*, 2024] introduces a multimodal geometric method to effectively fuse histological, gene expression, and spatial location information. SCGDL [Liu *et al.*, 2023a] applies a Bayesian Gaussian mixture model to generate spatial domains. Recently, many methods based on Graph Neural Networks (GNNs) have also been proposed, such as MAFN [Zhu *et al.*, 2024], SpaGCN [Hu *et al.*, 2021], and Spatial-MGCN [Wang *et al.*, 2023]. However, these methods construct spatial graphs using a conventional approach with a static radius based on spatial coordinates, as shown in Fig. 1 A. For the central node, considering all nodes within the circle as neighbors, including those from other clusters, hinders the accurate identification of spatial domains and complicates the precise partitioning of boundary nodes in clusters. The radius used to construct the graph is crucial: if it is too small, fails to capture adequate spatial neighbors; and too large, it introduces noise into the spatial domain.

To address this issue, we propose a novel spatially resolved transcriptomics data clustering network (TSstc) with tailored spatial-scale modulation, as illustrated in Fig. 2. Specifically, we employ a tailored spatial-scale modulation approach that incrementally constructs different spatial graphs $\mathbf{A}^v_{spa}$ as the spatial domain radius expands. Additionally, we introduce a Spatiality-Aware Sampling (SAS) strategy to aggregate node representations by accounting for the spatial dependencies between spots. Next, we use graph convolutional network (GCN) encoders to learn gene embeddings $\mathbf{Z}_{gene}$ from gene graph and multiple spatial embeddings $\mathbf{Z}^v_{spa}$ from spatial graphs. The Cross-view Fusion module is then applied to integrate the gene embedding and spatial embedding features for $\mathbf{Z}_f$. During training, we incorporate cross-view correlation-based tailored spatial regularization constraints across spatial embeddings $\mathbf{Z}^v_{spa}$ at different scales to remove noisy neighbors while preserving high-quality neighbor relationships. Finally, we utilize a zero-inflated negative binomial (ZINB) model to reconstruct the original expression matrix, capturing the global probability distribution of gene expression profiles. Extensive experimental results demonstrate that our approach outperforms existing state-of-the-art methods in various spatial transcriptomic tasks and related downstream applications.

The main contributions of this work are summarized as follows:

- We propose a novel network for clustering spatially resolved transcriptomics data, termed TSstc. To our knowledge, we are the first to explore spatial graph construction scales and address the challenge of identifying boundary nodes within clusters.
- We design a tailored spatial-scale modulation approach that incrementally constructs spatial graphs and employs the Spatiality-Aware Sampling (SAS) strategy to aggregate node representations. Additionally, we implement cross-view correlation-based tailored spatial regularization constraints to preserve high-quality neighbor relationships across spatial embeddings at different scales.
- We conducted experiments on three benchmark datasets

to validate the effectiveness of our proposed method. The results demonstrate the superiority of TSstc compared to state-of-the-art methods.

2 Related Work

2.1 Spatial Transcriptomics Techniques

Spatial transcriptomics (ST) techniques [Khan *et al.*, 2024; Yang *et al.*, 2024c], widely used in fields such as embryonic development [Rossant and Tam, 2022], disease research [Moore *et al.*, 2024], and plant science [van Dijk *et al.*, 2021], provide insights that traditional transcriptomics cannot offer due to its lack of spatial resolution. ST enables the capture of quantitative gene expression data while preserving the spatial locations of cells, thereby overcoming the spatial information loss inherent in single-cell sequencing. As a result, significant efforts have been made to develop deep learning networks that utilize spatial domains. Hu et al. [Hu *et al.*, 2021] proposed a graph convolution-based network that aggregates gene expression of each spot and its neighboring spots to identify coherent spatial domains. Wang et al.[Wang *et al.*, 2023] introduced an attention mechanism to combine gene expression and spatial embeddings, followed by spatial regularization to constrain the integrated features. However, these methods typically map aggregated features into a unified representation, overlooking gene heterogeneity. Xiao et al. [He *et al.*, 2024] constructed a heterogeneous graph coordinating module that maintains view independence during the fusion process. Zhu et al.[Zhu *et al.*, 2024] designed a Cross-view Fusion Module to dynamically integrate multiple features and employed a CCR strategy to reduce irrelevant information. Recently, Graph Neural Networks (GNNs) have shown remarkable effectiveness in handling non-Euclidean structural data, leading to their growing application in spatial transcriptomic clustering. Notable examples include SEDR [Xu *et al.*, 2024], DeepST [Xu *et al.*, 2022], CCST [Li *et al.*, 2022], GraphST [Long *et al.*, 2023], and others. These approaches leverage graph deep learning to accurately model the spatial distribution and interactions of gene expression within cells, thereby improving the precision of spatial domain delineation.

2.2 Deep Learning-based Multi-view Clustering

In recent years, deep learning-based multi-view clustering methods have attracted significant attention. These methods can be broadly categorized into three groups: generative adversarial network-based (GAN) [Yang *et al.*, 2024b; Yang *et al.*, 2022], variational autoencoder-based (VAE) [Li *et al.*, 2023; Yuan *et al.*, 2023], and graph convolutional network-based (GCN) [Xiao *et al.*, 2025]. Specifically, GANs generate realistic view data samples to improve clustering performance. Li et al. [Li *et al.*, 2019] used an adversarial module to capture the multi-view data distribution and disentangle the latent space. Meanwhile, VAEs map input data into a probabilistic distribution in the latent space, from which latent variables are sampled. Xu et al. [Xu *et al.*, 2021] have defined view-common and view-specific variables within a generative model, extracting a common clustering factor through an approximation of the discrete Gumbel-Softmax distribution. Additionally, recent literature also benefits from the strong representation power of graph convolutional networks, showing a strong tendency to utilize GCNs to further improve clustering performance. Zhou et al. [Zhou *et al.*, 2024] developed a self-encoding framework that incorporates view-specific GCN networks to learn high-level representations of each view's data features while considering prognosis prediction results. In our proposed method, TSstc, we aim to explore how to integrate gene information and spatial relationships to improve clustering performance.

3 Methodology

In this section, we first provide an overview of the TSstc model. We then describe the individual modules of the proposed model in detail, followed by an explanation of the training loss function.

3.1 Overview of TSstc

Our proposed TSstc model consists of four main components: the Graph Construction module, the Spatiality-Aware Sampling (SAS) module, the Feature Extraction module, and the Spatial Domain Constraints module.

The TSstc model is trained on spatial transcriptomic datasets, which include the gene feature matrix $\mathbf{X}_{gene} \in \mathbb{R}^{N\times D}$ and the 2D spatial coordinates $\mathbf{S_c} \in \mathbb{R}^{N\times 2}$, where N denotes the number of cells and D denotes the dimension of each cell. We construct the gene graph $\mathbf{A}_{gene} \in \mathbb{R}^{N\times N}$ and different spatial graphs $\{\mathbf{A}^v_{spa}\}^n_{v=1} \in \mathbb{R}^{N\times N}$ incrementally as the radius of the spatial domain expands. Then, we aggregate node representations to obtain different spatial feature matrixes $\{\mathbf{X}^v_{spa}\}^n_{v=1} \in \mathbb{R}^{N\times D}$ by the SAS module and we utilize GCN to learn the latent representations $\mathbf{Z}_{gene} \in \mathbb{R}^{N\times d}$ and $\mathbf{Z}^v_{spa} \in \mathbb{R}^{N\times d}$. To obtain more comprehensive cross-view features $\mathbf{Z}_f$, TSstc uses an attention fusion module to integrate the gene and spatial features. Furthermore, we use the Spatial Domain Constraints module to preserve high-quality spatial neighbor relationships and capture the global probability distribution of gene expression profiles.

3.2 Graph Construction Module

Compared to directly constructing spatial graphs with a fixed radius based on spatial coordinates $\mathbf{S_c}$, we employ a tailored spatial-scale modulation approach to construct multiple spatial graphs. For each dataset, we analyze the spatial distribution of spots and incrementally construct spatial graphs $\mathbf{A}^v_{spa}$ as the Euclidean distance-based radius R^v increases, thereby evaluating the connectivity between two specified spots as follows:

$$\mathbf{A}^v_{spa_{ij}} = \begin{cases} 1, & \text{if } L^v_{ij} \le R^v \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

where L^v_{ij} represents the Euclidean distance of spot i and j, calculated based on the spatial coordinates $\mathbf{S_c}$.

Then we measure the cosine similarity between two points i and j by calculating the similarity of gene expression vectors $\mathbf{X}_{gene_i}$ and $\mathbf{X}_{gene_j}$. We then construct the K-NN [Guo *et al.*, 2003] neighbor graph, denoted as $\mathbf{A}_{gene}$, using a binary

classification system, where:

$$\mathbf{A}_{gene_{ij}} = \begin{cases} 1, & \text{if } j \text{ is a neighbor of } i \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

We utilize the tailored spatial-scale modulation approach to construct various spatial graphs $\mathbf{A}_{spa}^{v}$ that identify spatial domains of spots. Additionally, the gene graph $\mathbf{A}_{gene}$ is constructed to capture the similarity between spots.

3.3 Spatiality-aware Sampling (SAS) Module

Understanding the spatial domain of the target spot and sampling its spatial neighborhood is crucial for spatially resolved transcriptomics data. We propose the Spatiality-Aware Sampling (SAS) module and combine the various spatial graphs $\mathbf{A}_{spa}^{v}$ to perform an aggregation operation on the features of the target spot and its spatial neighborhood, yielding the spatial feature representation $\mathbf{X}_{spa}^{v}$ before training, where:

$$\boldsymbol{X}_{spa_i}^{v} = \text{ AGGREGATION } \left(\left\{\boldsymbol{X}_{gene_j} : v_j \in \mathcal{N}_i\right\}\right), \tag{3}$$

where $\mathcal{N}_i$ represents the spatial neighborhood of spot i, and the aggregation function can take various forms, such as mean, max, sum, attention, and ensemble, which learn the spatial feature representations. In this case, we choose the mean operation.

3.4 Feature Extraction Module

In general, spatially resolved transcriptomics data are prone to noise and redundant information. To filter out this noise and retain more discriminative features in the latent space, we use the $\mathbf{GCN}_{gene}$ encoder and $\mathbf{GCN}_{spa(v)}$ encoders [Yang *et al.*, 2024a; Qin *et al.*, 2024] to learn specific features of the gene graph and spatial graph, respectively. The propagation rule is defined as follows:

$$\mathbf{Z}_{gene}^{(l+1)} = \sigma\left(\widetilde{\mathbf{A}}_{gene}\mathbf{Z}_{gene}^{(l)}\mathbf{W}_1^{(l)}\right), \tag{4}$$

$$\mathbf{Z}_{spa}^{v^{(l+1)}} = \sigma\left(\widetilde{\mathbf{A}}_{spa}^{v}\mathbf{Z}_{spa}^{v^{(l)}}\mathbf{W}_2^{v^{(l)}}\right). \tag{5}$$

We simplify the gene and spatial normalized adjacency matrix as $\widetilde{\mathbf{A}}$, which is given by $\tilde{\mathbf{A}} = (\mathbf{D})^{-\frac{1}{2}}(\mathbf{I}+\mathbf{A})(\mathbf{D})^{-\frac{1}{2}}$, where $\mathbf{D}_{ii} = \sum_j (\mathbf{A}_{ij} + \mathbf{I}_{ij})$, and the identity matrix $\mathbf{I} \in \mathbb{R}^{N\times N}$ ensures self-loops for each spot. The matrices $\mathbf{W}_1^{(l)}$ and $\mathbf{W}_2^{v^{(l)}}$ are the weights for the l-th layer and $\sigma(\cdot)$ is the Relu activation function.

In this subsection, we introduce the Cross-view Attention-induced Feature Fusion module, which allocates variable weights to effectively integrate gene expression and spatial embedding features. Specifically, we integrate the latent representations $\mathbf{Z}_{gene} \in \mathbb{R}^{N\times d}$ and $\{\mathbf{Z}_{spa}^{v}\}_{v=1}^{n} \in \mathbb{R}^{N\times d}$ to obtain the global cross-view feature $\mathbf{Z}_f \in \mathbb{R}^{N\times d}$.

To incorporate the $(v+1)$-th view information, we concatenate the d-dimensional feature vectors of $\mathbf{Z}_{spa}^{v}$ and $\mathbf{Z}_{gene}$ along the feature dimension to form $\mathbf{Z}^c \in \mathbb{R}^{N\times(d\times(v+1))}$. An attention weight matrix $\mathbf{H} \in \mathbb{R}^{N\times(v+1)}$ is then generated by the network as follows:

$$\mathbf{H} = \phi(\mathbf{W}_2 \cdot \delta(\mathbf{W}_1 \cdot \mathbf{Z}^c + \mathbf{b}_1) + \mathbf{b}_2), \tag{6}$$

where $\phi(\cdot)$ denotes the Softmax function, $\delta(\cdot)$ is the Tanh activation function, and $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are parameters of the fusion network.

Next, a broadcasting mechanism is applied to ensure dimensional compatibility, followed by the application of attention weights to each feature vector. Finally, the weighted feature vectors are summed to yield the fused feature vector $\mathbf{Z}_f \in \mathbb{R}^{N\times d}$ as follows:

$$\mathbf{Z}_f = \sum_v \mathbf{H}^v \cdot \mathbf{Z}_{spa}^{v} + \mathbf{H}_{last} \cdot \mathbf{Z}_{gene}, \tag{7}$$

where $\mathbf{H}^v$ denotes the v-th column of the attention weight matrix $\mathbf{H}$ and $\mathbf{H}_{last}$ is its last column.

3.5 Spatial Domain Constraints Module

Raw spatial transcriptomics data include spots with spatial relationships and location information, making it crucial to preserve these relationships in the latent space. Spatial-MGCN [Wang *et al.*, 2023] and MAFN [Zhu *et al.*, 2024] use spatial graph-based regularization loss to enforce embedding constraints. However, if the spatial graph contains noisy nodes or connections between nodes from different clusters, the regularization loss hinders the learning of high-quality embedding features. To address this, we propose a cross-view correlation-based tailored spatial regularization constraint loss to preserve high-quality neighbor relationships across spatial embeddings at different scales. We begin by evaluating the sample correlation $\mathbf{S}^v$ between the gene features $\mathbf{Z}_{gene}$ and each spatial feature $\mathbf{Z}_{spa}^{v}$, as follows:

$$\mathbf{S}_{ij}^{v} = \frac{\left(\mathbf{Z}_{spa_i}^{v}\right)\left(\mathbf{Z}_{gene_j}\right)^{\mathrm{T}}}{\left\|\mathbf{Z}_{spa_i}^{v}\right\|\left\|\mathbf{Z}_{gene_j}\right\|}, \forall i,j \in [1,N], \tag{8}$$

where $\mathbf{S}_{ij}^{v}$ represents the cosine similarity between the i-th node embedding in the v-th spatial view and j-th node embedding in the gene view. For each spot in the spatial embedding feature $\mathbf{Z}_{spa}^{v}$, we select its neighboring nodes with high expression values based on the correlation matrix $\mathbf{S}^v$ to refine the spatial graph. This process removes some connections that may originate from different clusters or contain noise, with low expression values. We then apply regularization loss to the spatial embedding features, as shown below:

$$\sum_v \mathcal{L}_R^v = -\sum_{i=1}^{N}\left(\sum_{j\in\mathbf{R}_i^v}\log\left(\sigma\left(\mathbf{Z}_{spa_{ij}}^{v}\right)\right) + \sum_{k\notin\mathbf{R}_i^v}\log\left(1-\sigma\left(\mathbf{Z}_{spa_{ij}}^{v}\right)\right)\right), \tag{9}$$

where $\mathbf{R}_i^v$ denotes the set of spatial neighbors for spot i in the v-th spatial embedding. It is generally accepted that data spots in close physical proximity should also be close in latent space, while non-adjacent spots should be distant.

As done in previous methods [He *et al.*, 2024; Zhu *et al.*, 2024], we use the ZINB decoder module [Wang *et al.*, 2023] to reconstruct the gene expression matrix, assuming that the gene expression matrix $\mathbf{X}_{gene}$ and the spatial expression matrix $\mathbf{X}_{spa}$ follow a ZINB distribution. For simplicity, we denote them as $\mathbf{X}$ in the following equations:

$$\begin{aligned} p_{\text{zinb}}\left(\mathbf{X}_{ij} \mid b_i\right) &= \text{ZINB}\left(\mathbf{X}_{ij} \mid \pi_{ij}, v_{ij}, \theta_{ij}, b_i\right) \\ &= 0 + (1-\pi_{ij})\, p_{nb}\left(\mathbf{X}_{ij} \mid b_i\right), \end{aligned} \tag{10}$$

Datasets	DeepST	GraphST	SCANPY	SCGDL	SpaGCN	Spatial-MGCN	stLearn	stCluster	Ours
151507	54.76	39.09	20.43	49.83	39.21	63.05	49.21	64.21	**68.53**
151508	42.36	49.36	15.34	34.75	33.78	46.72	51.33	53.75	**54.73**
151509	43.47	52.13	19.04	32.75	35.14	54.22	45.68	47.24	**63.24**
151510	51.14	49.57	14.99	31.47	37.45	51.42	44.24	46.34	**58.74**
151670	33.75	46.24	10.24	26.57	33.47	35.25	23.75	50.49	**53.95**
151671	52.67	61.27	12.24	31.75	42.57	59.37	68.44	70.14	**72.11**
151672	48.62	63.24	12.96	34.27	52.77	**77.14**	34.58	68.24	75.36
151674	55.39	43.25	22.75	27.63	31.36	60.43	38.99	55.42	**63.24**
HBC	53.13	53.78	49.15	35.32	55.65	63.43	55.26	57.88	**63.83**
MBA	25.63	40.59	23.43	25.95	31.65	42.46	38.42	36.88	**44.82**

Table 1: The comparative performance of our method against others across ten datasets is evaluated using metrics ARI. The best performance is denoted by bold font, while an underline indicates the second-best.

where $p_{nb}\left(\mathbf{X}_{ij} \mid b_i\right)$ is given by:

$$\begin{aligned} & p_{nb}\left(\mathbf{X}_{ij} \mid b_i\right) = NB\left(\mathbf{X}_{ij} \mid v_{ij}, \theta_{ij}, b_i\right) \\ & = \frac{\Gamma(\mathbf{X}_{ij}+\theta_{ij})}{\Gamma(\mathbf{X}_{ij}+1)\Gamma(\theta_{ij})}\left(\frac{\theta_{ij}}{\theta_{ij}+v_{ij}}\right)^{\theta_i}\left(\frac{v_{ij}}{v_{ij}+\theta_{ij}}\right)^{\mathbf{X}_{ij}}. \end{aligned} \tag{11}$$

Here, π_{ij} and v_{ij} are the zero-inflation and mean parameters, respectively, while θ_{ij} and b_i represent the discretization of the decoder output and the bias vector. The latent representation $\mathbf{Z}_{gene}$ and $\mathbf{Z}^{v}_{spa}$ serve as input to the ZINB decoder, yielding estimates for these parameters. The negative log-likelihood of the ZINB distribution is used as the reconstruction loss:

$$\mathcal{L}_C = -\frac{1}{m\left(n_s+n_t\right)} \sum_{i=1}^{n_s+n_t} \sum_{j=1}^{m} \ln p_{\text{zinb}}\left(\mathbf{X}_{ij} \mid b_i\right). \tag{12}$$

3.6 The Total Loss Function

Our total loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}_R + \alpha \mathcal{L}_C. \tag{13}$$

The parameter α represents the weighting factor used to balance the influences of $\mathcal{L}_R$ and $\mathcal{L}_C$.

4 Experiments

This section assesses the performance of our approach on three benchmark datasets and compares its clustering effectiveness with that of eight state-of-the-art methods for spatial transcriptomics.

4.1 Datasets

We evaluate our method on three widely used spatial transcriptomics datasets. The first dataset is the LIBD human dorsolateral prefrontal cortex DLPFC [Maynard *et al.*, 2021], comprising 8 slices each containing 33,538 genes and 5–7 regions. The second dataset is the 10x Visium dataset of human breast cancer [Maynard *et al.*, 2021] which contains 20 regions and 36,601 genes. The third dataset is mouse brain anterior tissue [Buache *et al.*, 2011], annotated with 52 regions and containing 32,285 genes. Preprocessing for these datasets follows a method similar to that described in MAFN [Zhu *et al.*, 2024]. We exclude spots outside the primary tissue areas and use the SCANPY [Wolf *et al.*, 2018] toolkit to filter out genes with minimal expression or variance, selecting the top 3,000 genes with the highest variability [Zeng *et al.*, 2023]. Finally, gene normalization is performed by applying a scaling factor.

4.2 Compared Methods

To validate the effectiveness of our method, we compare it against eight approaches: SCANPY [Wolf *et al.*, 2018], SpaGCN [Hu *et al.*, 2021], DeepST [Xu *et al.*, 2022], GraphST [Long *et al.*, 2023], SCGDL [Liu *et al.*, 2023b], Spatial-MGCN [Wang *et al.*, 2023], stLearn [Pham *et al.*, 2023], stCluster [Wang *et al.*, 2024]. All comparative experiments are conducted using default parameters.

4.3 Training Details and Metrics

We conduct experiments to evaluate the proposed TSstc on the PyTorch platform using a single NVIDIA GeForce RTX 3090. Identifying the optimal spatial domain is crucial for different datasets. First, we analyze the spatial distribution of spots and incrementally construct spatial graphs as the Euclidean distance-based radius R increases. Next, we input these spatial graphs into the GCN and train TSstc on all benchmark datasets for at least 150 iterations until convergence. We apply a weight decay of 0.0005 and use the Adam optimizer with a learning rate of 1e-3 to optimize our model. The parameter α is set to 10. To validate the effectiveness of clustering, we use the Adjusted Rand Index (ARI) [Meilă, 2007] and Normalized Mutual Information (NMI) [Knops *et al.*, 2006] as evaluation metrics.

4.4 Clustering Results and Analysis

Table 1 summarizes the performance of TSstc in terms of ARI compared to other methods across ten datasets, with NMI provided in the supplementary materials. From this table, we observe the following:

(1) In the DLPFC dataset, TSstc outperforms nearly all methods across 8 slices, achieving 4.32% higher ARI scores in 151507 slice compared to stCluster [Wang *et al.*, 2024], which uses graph contrastive learning. Several methods show significant drops below 40% on specific slices, reducing their reliability in scientific and clinical settings. In contrast, TSstc

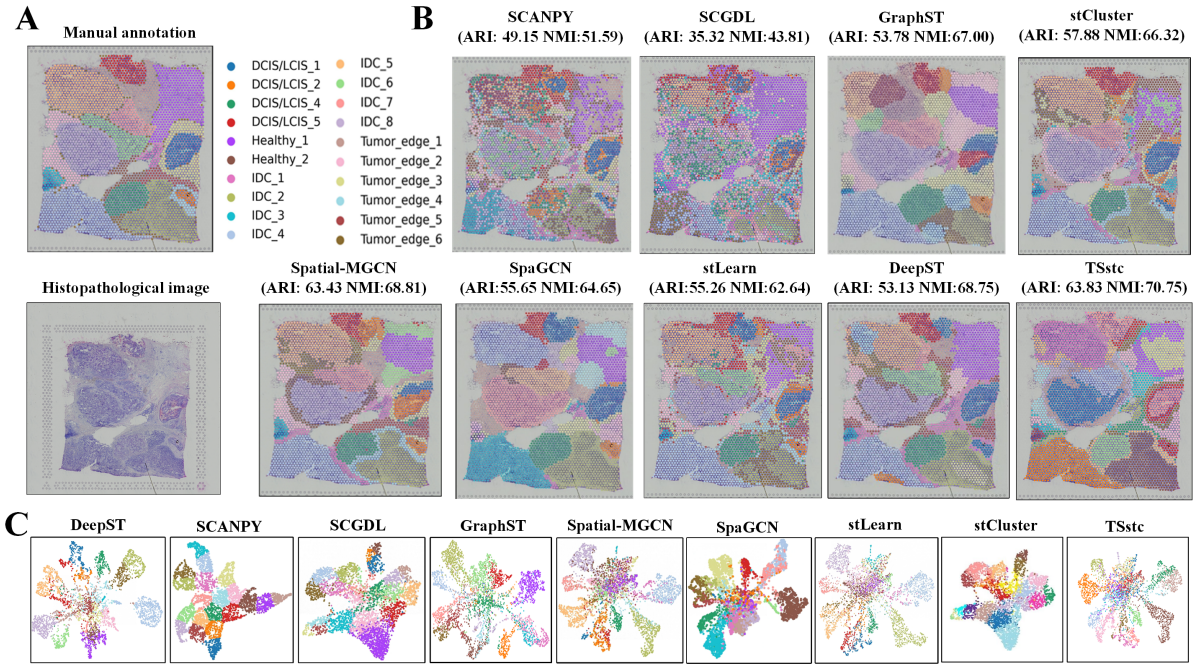

Figure 3: A: The manually annotated layer structure and the corresponding histopathological image of the Human Breast Cancer dataset. B: Spatial domain detection results for various methods. C: UMAP visualization of the experimental results.

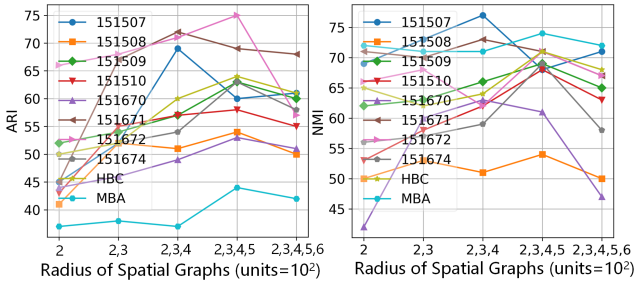

Figure 4: A comparison of the ARI and NMI metrics across different radius combinations for all datasets.

outperforms the average of competing methods by 23%, demonstrating the effectiveness of the tailored spatial-scale modulation approach.

(2) In the Human Breast Cancer dataset, our tailored spatial-domain modulation improves ARI by 0.4% and NMI by 1.94% compared to the best-performing model, Spatial-MGCN [Wang *et al.*, 2023]. We also detect the spatial domain, and TSstc clusters similar spots effectively, leading to well-defined boundaries, as shown in Fig.3. In contrast, DeepST[Xu *et al.*, 2022] and GraphST [Long *et al.*, 2023] fail to preserve spatial coherence, despite their strong local representation capabilities. By using the Spatiality-Aware Sampling (SAS) mechanism, TSstc integrates spatial information at multiple scales, ensuring a biologically meaningful connection between neighboring spots based on expression patterns. In the Mouse Brain Anterior dataset, Tab. 1 compares TSstc with other methods, and the supplementary materials show the spatial domain detection results. TSstc achieves the highest ARI of 44.82%, outperforming other methods, while stLearn [Pham *et al.*, 2023], stCluster [Wang *et al.*, 2024], and SpaGCN [Hu *et al.*, 2021] report lower ARI scores of 38.42%, 36.88%, and 31.65%, respectively.

4.5 Ablation Studies

This section presents an ablation study to examine the impact of each component in our TSstc model. The ablation experiments are categorized into four sections, and Tab. 2 shows ARI performance across ten real datasets, with NMI results provided in the supplementary materials.

Effect of $\mathcal{L}_R$ As shown in Tab. 2, w/o $\mathcal{L}_R$ refers to the removal of the spatial domain constraints module. In the 151672 slice, this results in a 27% decrease in performance compared to TSstc, emphasizing the importance of spatial domain constraints in preserving spatial relationships and location details and integrating $\mathcal{L}_R$ into the spatial embedding features ensures the retention of high-quality spatial neighborhood information.

Effect of $\mathcal{L}_C$ w/o $\mathcal{L}_C$ represents the removal of the ZINB decoder module. Analysis of the third row in Tab. 2 indicates that incorporating $\mathcal{L}_C$ improves the average performance by over 9.9%, further validating the effectiveness of the ZINB decoder. This module reconstructs the gene expression matrix, mitigating noise issues caused by the discrete and sparse nature of spatial transcriptomics data.

Effect of SAS module w/o SAS refers to the omission of the Spatiality-Aware Sampling (SAS) module. In the Human Breast Cancer dataset, TSstc achieves approximately 7% higher ARI scores compared to w/o SAS, demonstrat-

Datasets	151507	151508	151509	151510	151670	151671	151672	151674	HBC	MBA
w/o $\mathcal{L}_R$	55.24	43.75	51.96	47.34	41.57	64.37	48.75	49.27	47.23	32.66
w/o $\mathcal{L}_C$	56.37	46.37	55.37	49.37	43.11	58.76	62.34	48.37	58.77	42.79
w/o SAS	64.37	49.37	58.77	51.34	47.99	62.37	67.77	60.71	57.24	38.97
w/o TSM	66.37	51.37	61.73	52.79	52.44	68.37	73.88	51.22	58.74	41.96
TSstc	**68.53**	**54.73**	**63.24**	**58.74**	**53.95**	**72.11**	**75.36**	**63.24**	**64.83**	**44.82**

Table 2: This table presents the ablation study in term of ARI for each component, where w/o $\mathcal{L}_R$, w/o $\mathcal{L}_C$, w/o SAS, and w/o TSM, respectively.

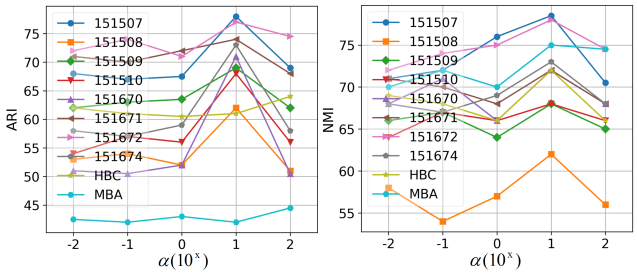

Figure 5: This figure presents a comparison of the ARI and NMI metrics for various values of α across ten datasets.

ing the module's effectiveness. By leveraging the SAS mechanism, TSstc integrates spatial information across multiple scales, ensuring a biologically meaningful association between neighboring spots based on expression patterns.

Effect of TSM module To evaluate the effectiveness of our tailored spatial-scale modulation (TSM) module, we replace it with a static radius, following the approach used in Spatial-MGCN [Wang *et al.*, 2023]. The results show that w/o TSM leads to a 4% lower ARI in the Mouse Brain Anterior dataset, with consistently lower ARI scores across other datasets. Furthermore, we evaluate the effect of different spatial graph combinations across all datasets, as shown in Fig. 4. For each dataset, spatial graphs are incrementally constructed as the Euclidean distance-based radius R increases, facilitating the identification of the optimal spatial graph combination. For the 151672 slice, the radius combination {2, 3, 4, 5} yields spatial graphs that achieve approximately 4% higher ARI scores and 3% higher NMI scores compared to the second-best combination. For the 151507 slice, the radius combination {2, 3, 4} achieved the best scores. Overall, these findings highlight the importance of the TSM module in identifying the optimal spatial domain, effectively capturing spatial relationships, and enhancing the understanding of cellular survival patterns.

4.6 Parameter Sensitivity Analysis

This section investigates the impact of the parameter α on clustering performance by varying its values within the set {0.01, 0.1, 1, 10, 100} across the DLPFC 8 slices, Human Breast Cancer, and Mouse Brain Anterior datasets. α serves as a weighting factor that balances the influences of $\mathcal{L}_R$ and $\mathcal{L}_C$. As shown in Fig. 5, in the 151507 slice, the ARI increases gradually with α, while in the Mouse Brain Anterior datasets, it decreases or remains constant. The Human Breast Cancer dataset shows minimal performance changes, indicating that its clustering performance is less sensitive to variations in α. In our experiments, the parameter α is fixed at 10.

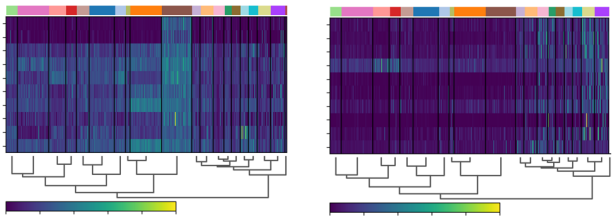

Figure 6: Heatmap of the top 10 differentially expressed genes (DEGs) between Healthy1 and DCIS/LCIS5 for the Human Breast Cancer dataset.

4.7 Downstream Applications

We explore tumor tissue heterogeneity by examining the top 10 differentially expressed genes (DEGs) in the Healthy1 and DCIS/LCIS5 clusters, as shown in Fig. 6. The identified gene expression variations highlight regional molecular differences, underscoring TSstc's ability to capture the complex heterogeneity of Human Breast Cancer tissues. We also conduct several downstream applications, including UMAP visualization and gene imputation. Detailed results are provided in the supplementary materials.

5 Conclusion

In this paper, we propose a novel spatially resolved transcriptomics data clustering network (TSstc) to accurately identify spatial domains and precisely partition the boundary nodes within clusters. Specifically, we introduce a tailored spatial-scale modulation approach that incrementally constructs spatial graphs as the spatial domain radius expands, and a Spatiality-Aware Sampling (SAS) strategy is proposed to aggregate node representations. We incorporate cross-view correlation-based tailored spatial regularization constraints and employ a zero-inflated negative binomial model to capture the global probability distribution. Extensive experiments demonstrate that our method outperforms state-of-the-art clustering approaches and benefits related to downstream applications. For future work, we plan to explore the spatial distribution patterns of spots to refine clustering performance.

Acknowledgments

The work was supported in part by the National Natural Science Foundation of China under grants 62476258 and 62325604, and in part by the Natural Science Foundation of Hubei Province under grant 2025AFA113, and in part by the Fundamental Research Funds for National Universities, China University of Geosciences (Wuhan) under grant 2024XLB6.

References

[Ahmed *et al.*, 2020] Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8):1295, 2020.

[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.

[Guo *et al.*, 2003] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, and Kieran Greer. Knn model-based approach in classification. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, pages 986–996. Springer, 2003.

[He *et al.*, 2024] Xiao He, Chang Tang, Xinwang Liu, Chuankun Li, Shan An, and Zhenglai Li. Heterogeneous graph guided contrastive learning for spatially resolved transcriptomics data. In *Proceedings of the ACM International Conference on Multimedia*, pages 8287–8295, 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.

[Khan *et al.*, 2024] Shawez Khan, Jong Joo Kim, et al. Spatial transcriptomics data and analytical methods: an updated perspective. *Drug Discovery Today*, page 103889, 2024.

[Knops *et al.*, 2006] Zeger F Knops, JB Antoine Maintz, Max A Viergever, and Josien PW Pluim. Normalized mutual information based registration using k-means clustering and shading correction. *Medical Image Analysis*, 10(3):432–439, 2006.

[Li *et al.*, 2019] Zhaoyang Li, Qianqian Wang, Zhiqiang Tao, Quanxue Gao, Zhaohua Yang, et al. Deep adversarial multi-view clustering network. In *International Joint Conference on Artificial Intelligence*, volume 2, page 4, 2019.

[Li *et al.*, 2022] Jiachen Li, Siheng Chen, Xiaoyong Pan, Ye Yuan, and Hong-Bin Shen. Cell clustering for spatial transcriptomics data with graph neural networks. *Nature Computational Science*, 2(6):399–408, 2022.

[Li *et al.*, 2023] Jingci Li, Guangquan Lu, Zhengtian Wu, and Fuqing Ling. Multi-view representation model based on graph autoencoder. *Information Sciences*, 632:439–453, 2023.

[Liu *et al.*, 2023a] 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), 04 2023.

[Liu *et al.*, 2023b] 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), 2023.

[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.

[Meilă, 2007] Marina Meilă. Comparing clusterings—an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, 2007.

[Moore *et al.*, 2024] Kristine A Moore, Angela J Mehr, Julia T Ostrowsky, Angela K Ulrich, Nicolina M Moua, Petra C Fay, Peter J Hart, Josephine P Golding, Virginia Benassi, Marie-Pierre Preziosi, et al. Measures to prevent and treat nipah virus disease: research priorities for 2024–29. *The Lancet Infectious Diseases*, 24(11):e707–e717, 2024.

[Mrabah *et al.*, 2023] Nairouz Mrabah, Mohamed Mahmoud Amar, Mohamed Bouguessa, and Abdoulaye Banire Diallo. Toward convex manifolds: A geometric perspective for deep graph clustering of single-cell rna-seq data. In *International Joint Conference on Artificial Intelligence*, pages 4855–4863, 2023.

[Pereira *et al.*, 2021] Wendell J Pereira, Felipe Marques Almeida, D Conde, KM Balmant, PM Triozzi, HW Schmidt, C Dervinis, GJ Pappas, and M Kirst. Asc-seurat: analytical single-cell seurat-based web application. *BMC bioinformatics*, 22:1–14, 2021.

[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.

[Qin *et al.*, 2024] Guangming Qin, Jianpeng Qi, Bin Wang, Guiyuan Jiang, Yanwei Yu, and Junyu Dong. Multi-relational graph attention network for social relationship

inference from human mobility data. In *International Joint Conference on Artificial Intelligence*, pages 2315–2323, 2024.

[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.

[Rossant and Tam, 2022] Janet Rossant and Patrick PL Tam. Early human embryonic development: Blastocyst formation to gastrulation. *Developmental cell*, 57(2):152–165, 2022.

[Song and Su, 2021] Qianqian Song and Jing Su. Dstg: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. *Briefings in bioinformatics*, 22(5), 2021.

[van Dijk *et al.*, 2021] Aalt Dirk Jan van Dijk, Gert Kootstra, Willem Kruijer, and Dick de Ridder. Machine learning in plant science and plant breeding. *Iscience*, 24(1), 2021.

[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), 2023.

[Wang *et al.*, 2024] Tao Wang, Han Shu, Jialu Hu, Yongtian Wang, Jing Chen, Jiajie Peng, and Xuequn Shang. Accurately deciphering spatial domains for spatially resolved transcriptomics with stcluster. *Briefings in Bioinformatics*, 25(4), 2024.

[Wolf *et al.*, 2018] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19:1–5, 2018.

[Xiao *et al.*, 2025] Yuang Xiao, Dong Yang, Jiaxin Li, Xin Zou, Hua Zhou, and Chang Tang. Dual alignment feature embedding network for multi-omics data clustering. *Knowledge-Based Systems*, 309:112774, 2025.

[Xu *et al.*, 2021] Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9234–9243, 2021.

[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.

[Xu *et al.*, 2024] Hang Xu, Huazhu Fu, Yahui Long, Kok Siong Ang, Raman Sethi, Kelvin Chong, Mengwei Li, Rom Uddamvathanak, Hong Kai Lee, Jingjing Ling, et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Medicine*, 16(1):12, 2024.

[Yang *et al.*, 2022] Ren Yang, Radu Timofte, and Luc Van Gool. Perceptual learned video compression with recurrent conditional gan. In *International Joint Conference on Artificial Intelligence*, pages 1537–1544, 2022.

[Yang *et al.*, 2024a] Niya Yang, Ye Wang, Zhizhi Yu, Dongxiao He, Xin Huang, and Di Jin. Joint domain adaptive graph convolutional network. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 2496–2504, 2024.

[Yang *et al.*, 2024b] Wenqi Yang, Minhui Wang, Chang Tang, Xiao Zheng, Xinwang Liu, and Kunlun He. Trustworthy multi-view clustering via alternating generative adversarial representation learning and fusion. *Information Fusion*, 107:102323, 2024.

[Yang *et al.*, 2024c] Wenyi Yang, Pingping Wang, Shouping Xu, Tao Wang, Meng Luo, Yideng Cai, Chang Xu, Guangfu Xue, Jinhao Que, Qian Ding, et al. Deciphering cell–cell communication at single-cell resolution for spatial transcriptomics with subgraph-based graph attention network. *Nature Communications*, 15(1):7101, 2024.

[Yuan *et al.*, 2023] Wenxuan Yuan, Shanchuan He, and Jianwen Dou. Mstcn-vae: An unsupervised learning method for micro-gesture recognition based on skeleton modality. In *International Joint Conference on Artificial Intelligence*, 2023.

[Zeng *et al.*, 2023] Yuansong Zeng, Rui Yin, Mai Luo, Jianing Chen, Zixiang Pan, Yutong Lu, Weijiang Yu, and Yuedong Yang. Identifying spatial domain by adapting transcriptomics with histology through contrastive learning. *Briefings in Bioinformatics*, 24(2), 2023.

[Zhai *et al.*, 2023] Yuyao Zhai, Liang Chen, and Minghua Deng. Realistic cell type annotation and discovery for single-cell rna-seq data. In *International Joint Conference on Artificial Intelligence*, pages 4967–4974, 2023.

[Zhang *et al.*, 2021] Jicun Zhang, Jiyou Fei, Xueping Song, and Jiawei Feng. An improved louvain algorithm for community detection. *Mathematical Problems in Engineering*, 2021(1):1485592, 2021.

[Zhang *et al.*, 2024] Daoliang Zhang, Na Yu, Zhiyuan Yuan, Wenrui Li, Xue Sun, Qi Zou, Xiangyu Li, Zhiping Liu, Wei Zhang, and Rui Gao. stmmr: accurate and robust spatial domain identification from spatially resolved transcriptomics with multimodal feature representation. *GigaScience*, 13:giae089, 2024.

[Zhou *et al.*, 2024] Lin Zhou, Ning Wang, Zhengzhi Zhu, Hongbo Gao, Yi Zhou, and Mingxing Fang. Identification of subtypes in digestive system tumors based on multi-omics data and graph convolutional network. *International Journal of Machine Learning and Cybernetics*, pages 1–11, 2024.

[Zhu *et al.*, 2024] Yanran Zhu, Xiao He, Chang Tang, Xinwang Liu, Yuanyuan Liu, and Kunlun He. Multi-view adaptive fusion network for spatially resolved transcriptomics data clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2024.