GATCL: An Adaptive Contrastive Learning Framework Based on MHGAT for Spatial Domain Identification in Spatial Transcriptomics

Shilin Zhang[1], Weiliang Huo[2], Qingchen Zhang[3*], Xiulong Liu[1*]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China
[2]School of Information and Communication Engineering, Hainan University, Hainan, China
[3]School of Computer Science and Technology, Hainan University, Hainan, China
zhang_shilin_sd@163.com,{wlhuo, zhangqingchen}@hainanu.edu.cn,xiulong_liu@tju.edu.cn

Abstract

Recent advances in spatial transcriptomics have enabled the simultaneous measurement of gene expression profiles and spatial location information, offering a more comprehensive and in-depth view for studying the tissue microenvironment. Spatial domain identification is a crucial step in analyzing spatial transcriptomics. However, current methods have poor accuracy and visualization because they lack self-adaptability to different tissue data, and moreover, they cannot effectively extract spatial location information. To address these issues, we propose an adaptive graph contrastive learning framework based on multi-head graph attention networks (GATCL) for spatial domain identification. Specifically, we design a data augmentation module to mask and shuffle the pre-processed gene expression data to generate more differentiated negative samples. In addition, we construct the multi-head graph attention networks (MHGAT) to encode gene expression profiles and spatial location information. More importantly, we design an adaptive graph contrastive learning model that works both with positive and negative samples from spatial transcriptomics. We introduce the attention pooling mechanism to dynamically and adaptively aggregate the spots' neighborhood information, and to improve the model's generalization ability for different spatial transcriptomics data. Furthermore, we design a discriminator that adds spectral normalization to bilinear functions. Experimental results on DLPFC, breast cancer, and mouse somatosensory cortex datasets demonstrate that the average Adjusted Rand Index (ARI) scores are 0.5746, 0.6182, and 0.6496, respectively, significantly outperforming baseline methods. More importantly, GATCL provides a more detailed visualization of different spatial transcriptomics data.

Introduction

The tissue microenvironment is essential for studying the mechanisms of disease onset and progression. For example, analyzing the distribution of tumor cells and immune cells in the tumor microenvironment and the spatial heterogeneity in gene expression can facilitate accurate tumor therapy. In recent years, unlike single cell RNA sequencing(Zhu et al. 2024)(Liu et al. 2019)(Ren et al. 2024), spatial transcriptomics has been able to obtain both gene expression profiles and spatial location information (Asp, Bergenstråhle, and Lundeberg 2020)(Rao et al. 2021), allowing for a more comprehensive characterization of the tissue microenvironment, which can help elucidate disease mechanisms and improve medical diagnosis (Moncada et al. 2020).

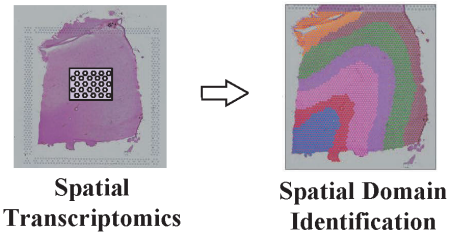

Figure 1: An illustration of spatial domain identification.

Spatial transcriptomics analysis includes three basic tasks: spatial domain identification, deconvolution (Li et al. 2022a), and intercellular interactions (Shao et al. 2022), of which spatial domain identification is an important task because it can describe spatial heterogeneity in gene expression (See Figure 1 as an example). The current spatial domain identification methods can be divided into two main categories according to whether they make use of spatial location information. The first type of methods like Louvain (Blondel et al. 2008) and K-means utilize only single gene expression profile data. Without using any spatial location information, methods of this type cannot effectively identify the spatial domain. Additionally, it is difficult for them to obtain the continuous smooth domain. The second type of method combines spatial location information and gene expression profile data and sometimes incorporates histological images. Li *et al.* proposed an unsupervised cell clustering method (CCST) based on graph convolutional networks (GCN), which combines spots' spatial information and gene expression information to achieve better clustering results (Li et al. 2022b). Hu *et al.* proposed a clustering method (SpaGCN) based on GCN, which used GCN to learn a joint representation of information in different dimensions to achieve unified modeling of gene expression, spatial location, and histological information (Hu et al. 2021). However, the above methods have two limitations. One is their poor generalization ability since they lack self-adaptability to different tissue data. The other is that they cannot extract

*Corresponding authors: Qingchen Zhang and Xiulong Liu.

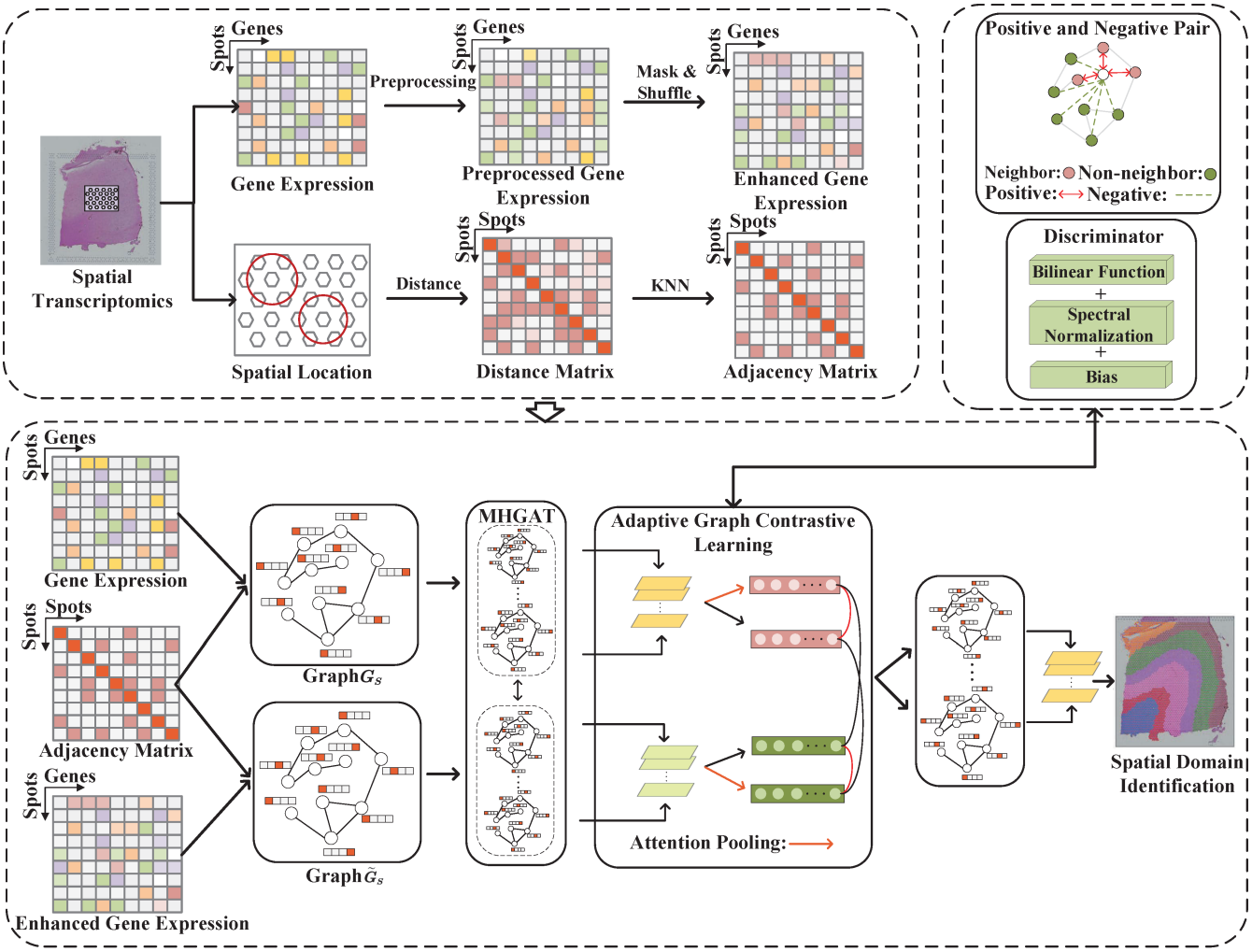

Figure 2: The overall framework of GATCL.

spatial location information effectively.

To address the above limitations, we propose an adaptive graph contrastive learning framework based on multi-head graph attention networks (GATCL) for spatial domain identification. The spatial location adjacency matrix and the preprocessed gene expression profiles are input into MHGAT for feature encoding. The adaptive graph contrastive learning module is constructed to further learn potential features with spatial location information and a higher representational ability for effective spatial domain identification. Among them, the spots' gene expression matrix is masked and shuffled to obtain negative samples, and attention pooling adaptively adjusts the weights to improve the generalization ability for different tissue datasets. We conduct experiments on datasets from different tissues, including the human dorsolateral prefrontal cortex (DLPFC), breast cancer, and the mouse somatosensory cortex, and compare the results with existing methods. The experimental results show that the average Adjusted Rand Index (ARI) scores on the three datasets are 0.5746, 0.6182, and 0.6496, respectively, significantly outperforming other methods. The visual results of our method have clear continuous hierarchical boundaries and are closest to the manually labeled seven-layer structure in the DLPFC dataset. In addition, experiments on breast cancer and mouse somatosensory cortex datasets also validate that GATCL has better model generalization ability for different tissue datasets.

The main contributions of the paper are as follows:

- We propose an adaptive graph contrastive learning framework based on multi-head graph attention networks (GATCL) for spatial domain identification, which can effectively extract spatial location information and gene expression profile information from spatial transcriptomics data for joint characterization.
- We design a data augmentation module to mask and shuffle the pre-processed gene expression data to generate more differentiated negative samples.
- We design an adaptive graph contrastive learning module based on MHGAT to encode features and introduce atten-

tion pooling to adaptively adjust the weights, thereby enhancing the characterization ability for different datasets. Additionally, we design a discriminator that incorporates spectral normalization into bilinear functions to improve GATCL's generalization ability.

- We validate the performance of GATCL on different tissue datasets: DLPFC, human breast cancer, and mouse somatosensory cortex datasets. GATCL has the best performance and can identify clear and continuous hierarchical boundaries.

Related Work

Graph Neural Networks

Graph Neural Networks (GNNs), as an application of deep learning to non-Euclidean data, have garnered extensive attention in recent years (Zhang et al. 2019). Huang *et al.* proposed ClusterDrop to solve the Battle Over-Smoothing problem for GNNs (Huang, Lu, and Yang 2024). Zhao *et al.* proposed a contrastive self-supervised GNNs (CSGNN) to predict molecular interactions (Zhao et al. 2021). In this work, we construct the multi-head graph attention networks (MHGAT) to encode gene expression profiles and spatial location information.

Contrastive Learning

Traditional supervised learning methods heavily rely on labeled training data, whereas contrastive learning, as a form of self-supervised learning, excels in achieving high performance without labeled data (Jin et al. 2023)(Lu et al. 2023)(Chu et al. 2021)(Chen et al. 2024). Guan *et al.* designed a molecule graph contrastive learning approach combined with Transformer (Guan and Zhang 2023). Zheng *et al.* proposed a contrastive learning-based model (ConGRI) for identifying gene regulatory interactions (Zheng et al. 2022). Liu *et al.* achieved good performance by combining contrastive learning for depression detection (Liu et al. 2024). While the aforementioned approaches have made significant strides in contrastive learning, they often struggle to adapt to specific downstream tasks involving spatial transcriptomic data.

Spatial Domain Identification

Spatial domain identification is an important task of spatial transcriptomics analysis. Dong *et al.* proposed a spatial domain identification method (STAGATE) based on a graph attention autoencoder (Dong and Zhang 2022). Fu *et al.* proposed a spatial domain identification method (SEDR) based on a depth autoencoder and a variational autoencoder, which obtained potential representations through the joint optimization of the two autoencoders (Fu et al. 2021). In this work, we construct an adaptive graph contrastive learning framework based on MHGAT for spatial domain identification, which can achieve better identification performance. Moreover, it can achieve finer visualization to different spatial transcriptomics.

Method

Overview of GATCL

GATCL makes full use of the gene expression profile information and spatial location information in spatial transcriptomics data to identify spatial domains. As shown in Figure 2, GATCL consists of four main modules: spatial transcriptomics data preprocessing module, data augmentation module, MHGAT module, and adaptive graph contrastive learning module. GATCL is described in Algorithm 1.

Data Preprocessing

GATCL takes as input the spatial transcriptomics data, which contains the gene expression matrix and spatial location information. For the gene expression matrix, we first use the scanpy package (V 1.7.1) (Wolf, Angerer, and Theis 2018) to extract the top 3000 highly variable genes. Then, normalization and logarithmic conversion are carried out on it, and finally, scaling to unit variance is performed to obtain the preprocessed gene expression matrix. The preprocessing strategy provided in (Long et al. 2023) is mainly followed here.

Graph Construction

Spots in the same spatial domain are usually physically close together in tissues. To effectively utilize the spatial location information in the spatial transcriptomics data, we use the KNN algorithm to calculate the Euclidean distance between all spots, and select the top 3 spots with the closest distance as the neighbors of the current spot. Thus, the spatial graph adjacency matrix is constructed. Specifically, the characterization of spatial transcriptomics data is defined as a graph $G_s = (V, E)$. Each spot forms the set of nodes V, and the connections between spots form the set of edges E. In graph G_s, the adjacency matrix $M \in \mathbb{R}^{N_s \times N_s}$ and the gene expression matrix $F = \{f_1, f_2, \cdots, f_{N_s}\}$ for each spot is included. Let m_{ij} denote the nearest neighbor relationship between spot i and spot j. m_{ij} is defined as follows:

$$m_{ij} = \begin{cases} 1, & \text{If } i \text{ and } j \text{ are close neighbors} \\ 0, & \text{Otherwise} \end{cases} \tag{1}$$

where $i, j \in V$.

Data Augmentation

To generate negative samples for adaptive contrastive learning, we retain the structure of the original adjacency matrix M and only take a 60% probability mask and random shuffling of the feature matrix F corresponding to the spots to obtain $\tilde{F}$. The negative sample $\tilde{G}_s$ is formed by the adjacency matrix M and the processed feature matrix $\tilde{F}$. $\tilde{F}$ is calculated as follows:

$$\tilde{F} = \varphi(F \cdot Z) \tag{2}$$

where $\varphi(\cdot)$ is random disruption, Z follows a Bernoulli distribution.

Multi-head Graph Attention Network for Feature Encoding

The spatial transcriptomics data of each tissue have features that are distinct from those of other tissues. To capture the distinct features adaptively, we construct MHGAT to extract effective feature representations and improve the generalization of the model (Velickovic et al. 2017). Given the set of neighbors N_i for the i-th spot, the attention coefficient of spot i and spot j is calculated as follows:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^T\left(\left[\boldsymbol{W}\boldsymbol{f}_i \| \boldsymbol{W}\boldsymbol{f}_j\right]\right)\right)\right)}{\sum_{\boldsymbol{k}\in N_i}\exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^T\left(\left[\boldsymbol{W}\boldsymbol{f}_i \| \boldsymbol{W}\boldsymbol{f}_k\right]\right)\right)\right)} \tag{3}$$

where W is the linear transformation weight matrix, a is the weight vector and $LeakyReLU$ is the activation function.

The features of neighbor spots are aggregated according to the attention coefficient α_{ij} as follow:

$$e_i = \sigma\left(\sum_{j\in N_i}\alpha_{ij}\boldsymbol{W}\boldsymbol{f}_j\right) \tag{4}$$

where $\sigma(\cdot)$ is the nonlinear activation function.

Integrating the results of multiple attention mechanisms yields feature encoding as follow:

$$e_i = \|_{k=1}^{K}\sigma\left(\sum_{j\in N_i}\alpha_{ij}^k\boldsymbol{W}^k\boldsymbol{f}_j\right) \tag{5}$$

where K is the number of heads in the multi-head attention module, and $\sigma(\cdot)$ is the nonlinear activation function.

After MHGAT encoding, the potential feature E_s is obtained: $E_{MH}(F, M) = E_s = \{e_1, e_2, \cdots, e_{N_s}\}$, and then E_s is reconstructed to obtain the reconstructed gene expression R_s. Training is conducted by minimizing the following loss function:

$$L_r = \sum_{i=1}^{N_s}\|f_i - r_i\|_F^2 \tag{6}$$

where f_i is the original gene expression and r_i is the reconstructed gene expression.

Adaptive Graph Contrastive Learning for Representation Enhancement

To further enhance the characterization capability of the MHGAT encoder for spatial transcriptomics data, the adaptive graph contrastive learning (AGCL) module is constructed to extract the global features of the data and improve the accuracy of spatial domain identification.

First, the graph G_s formed by the original adjacency matrix M and the original gene expression matrix F, and the graph $\tilde{G}_s$ formed by the original adjacency matrix M and the gene expression matrix $\tilde{F}$ after mask and shuffle processing are input into the MHGAT encoder to obtain the encoded feature representation matrix: $E_s \in \mathbb{R}^{N_s\times D}$ and $\tilde{E}_s \in \mathbb{R}^{N_s\times D}$.

Different from DGI (Velickovic et al. 2019) and GraphST (Long et al. 2023), we adopt the attention pooling to adaptively aggregate neighbor spots to learn the gene expression characteristics of spots in the graph structure and determine the weight dynamically and adaptively according to the relationship between different spots and their neighbor spots, thus improving the adaptive capability and performance of the model. Specifically, the features encoded by MHGAT are taken as input to obtain the global feature representation G_g by the attention pooling AP. The formula is defined as follows:

$$G_g = AP\left(E_{MH}(F, M)\right) \tag{7}$$

$$AP : \mathbb{R}^{N_s\times D} \to \mathbb{R}^D \tag{8}$$

$$s = Linear_2\left(ReLU\left(Linear_1\left(E_s\right)\right)\right) \tag{9}$$

$$w_i = \frac{\exp(s_i)}{\sum_{j=1}^{N_s}\exp(s_j)} \tag{10}$$

$$g = \sum_{i=1}^{N_s}(E_s * w_i) \tag{11}$$

$$G_g = \frac{g}{\|g\|_2} \tag{12}$$

Algorithm 1: GATCL

Initialize: Data G_s, $E_{MH}(\cdot)$

1: **while** not reached maximum epochs **do**
2: $\tilde{G}_s = G_{augm}(G_s)$ # Data augmentation
3: $E_s = E_{MH}\left(\tilde{G}_s\right)$ # Multi-head graph attention network for feature encoding
4: $G_g = AP(E_s)$ # Attention pooling for global feature
5: $D_s : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ # Discriminator
6: **define** $L_r = \sum_{i=1}^{N_s}\|f_i - r_i\|_F^2$
7: **define**

$$L_1 = -\frac{1}{2N_s}\left(\sum_{i=1}^{N_s}\left(E_{MH}(F, M)\left[\log D_s\ (e_i, G_g)\right]\right.\right.$$
$$\left.\left.+E_{MH}\left(\tilde{F}, M\right)\left[\log D_s\ (1 - D_s\ (\tilde{e}_i, G_g))\right]\right)\right)$$

8: **define**

$$L_2 = -\frac{1}{2N_s}\left(\sum_{i=1}^{N_s}\left(E_{MH}\left(\tilde{F}, M\right)\left[\log D_s\ \left(\tilde{e}_i, \tilde{G}_g\right)\right]\right.\right.$$
$$\left.\left.+E_{MH}(F, M)\left[\log D_s\ \left(1 - D_s\ \left(e_i, \tilde{G}_g\right)\right)\right]\right)\right)$$

9: $L = \lambda_1 L_r + \lambda_2(L_1 + L_2)$
10: Update encoder $E_{MH}(\cdot)$ to minimize L
11: **end while**
12: **return** Encoder $E_{MH}(\cdot)$

The local feature encoding e_i and the global feature representation G_g form a positive sample pair, and the same

$\tilde{e}_i$ and G_g form a negative sample pair. The goal of AGCL is to minimize the distance between positive sample pairs and maximize the distance between negative sample pairs to learn feature representations so that spots in the same spatial domain are closer together in the feature space and spots in different spatial domains are more dispersed in the feature space.

The discriminator D_s: $\mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ is introduced to obtain the probability scores of positive and negative sample pairs, respectively. The discriminator adds a spectral normalization to the bilinear function to alleviate the gradient disappearance problem and adds bias to improve the model's robustness.

The loss function L_1 and L_2 for the overall AGCL is defined as follows (Long et al. 2023):

$$L_1 = -\frac{1}{2N_s}\left(\sum_{i=1}^{N_s}\left(E_{MH}\left(F, M\right)\left[\log D_s\ \left(e_i, G_g\right)\right]\right.\right. \\ \left.\left. + E_{MH}\left(\tilde{F}, M\right)\left[\log\left(1 - D_s\ \left(\tilde{e}_i, G_g\right)\right)\right]\right)\right) \tag{13}$$

$$L_2 = -\frac{1}{2N_s}\left(\sum_{i=1}^{N_s}\left(E_{MH}\left(\tilde{F}, M\right)\left[\log D_s\ \left(\tilde{e}_i, \tilde{G}_g\right)\right]\right.\right. \\ \left.\left. + E_{MH}\left(F, M\right)\left[\log\left(1 - D_s\ \left(e_i, \tilde{G}_g\right)\right)\right]\right)\right) \tag{14}$$

The three loss functions are trained jointly, that is, the L_r, L_1, and L_2 loss functions are optimized simultaneously to improve the performance and generalization ability of the model.

$$L = \lambda_1 L_r + \lambda_2\left(L_1 + L_2\right) \tag{15}$$

where λ_1 and λ_2 are the weighting factors.

Finally, the learned latent features are processed by PCA and are input into the mclust clustering algorithm (Fraley et al. 2014) to identify the tissue's spatial domains.

Experiments

Datasets

In order to comprehensively evaluate the performance of GATCL, we conduct experiments on three real datasets containing different resolutions, different platforms, human and mouse, healthy tissue, and cancer tissue (Table 1). The descriptions of the datasets are as follows:

Human dorsolateral prefrontal cortex (DLPFC) dataset: DLPFC is a publicly available ST dataset for the 10x Genomics Visium platform (Maynard et al. 2021). It contains 12 tissue samples, each with a minimum of 5 layers and a maximum of 7 layers of structure.

Breast cancer dataset: The dataset is 10X Visium spatial transcriptomics data and has 20 manually annotated regions.

Mouse somatosensory cortex dataset: The dataset is a publicly available single-cell spatial transcription dataset collected using the osmFISH technique and has 11 manually annotated regions (Codeluppi et al. 2018).

The Compared Methods and Evaluation Metric

To evaluate the effectiveness of GATCL in spatial domain identification, we compare it experimentally with four baseline methods (CCST, SpaGCN, STAGATE, and GraphST) used for spatial transcriptomics data. The default parameters of the method are used in the comparison experiments. The Adjusted Rand Index (ARI) is used as the evaluation metric to assess the performance of the GATCL and comparison methods.

Implementation

Our method and the compared methods are trained on the Linux platform and run on a workstation that is equipped with two NVIDIA A100 80 GB GPUs. The model employs unsupervised training with a learning rate of 0.001 and 5000 epochs. λ_1 and λ_2 are set to 10 and 1, respectively.

Spatial Domain Identification on the DLPFC Dataset

We first quantitatively compare the spatial domain identification effects of GATCL with the four other baseline methods (CCST, SpaGCN, STAGATE, and GraphST) on the DLPFC dataset. As shown in Figure 3, the manual annotations provided by Maynard *et al.* as ground truth (Maynard et al. 2021) are used to calculate the evaluation metrics for clustering performance. The ARI results of different unsupervised methods are shown in Table 2. Our method achieves the best spatial domain identification effect on the DLPFC dataset, and the average ARI is 0.5746. GATCL produces ARIs that are 0.0595, 0.1950, 0.1574, and 0.0651 higher than those of STAGATE, SpaGCN, CCST, and GraphST, respectively. This result indicates that GATCL can extract features with stronger characterization ability from spatial transcriptomics data.

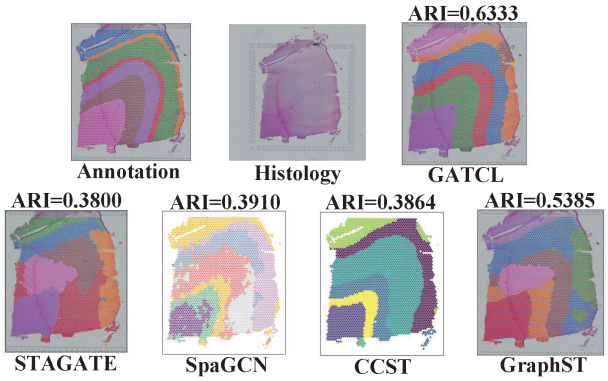

Figure 3: The visualization results on the DLPFC 151674.

The visualization results of each method on sample number 151674 are shown in Figure 3. GATCL can effectively distinguish the hierarchical structure and obtain the best clustering effect, and the ARI is 0.6333. In addition, the results of GATCL correspond to the hierarchical structure of the manual annotations, with clear and continuous hierarchical boundaries. The spatial domain identified by STAGATE is fragmentary, the hierarchical boundary identified

Datasets	Spots/cells	Genes	Tissue	Domains	Technology	Size/radius	Tissue type
DLPFC	3460-4789	33538	12	5-7	10x Visium	$55\mu m$	Human, Health
Breast cancer	3798	36601	1	20	10x Visium	$55\mu m$	Human, Cancer
Mouse somatosensory cortex	4839	33	1	11	osmFISH	$\leq$1 cell	Mouse, Health

Table 1: The statistical information of the datasets.

by SpaGCN is discontinuous, and the hierarchical results of CCST and GraphST are inconsistent.

Sample	STAGATE	SpaGCN	CCST	GraphST	GATCL
151507	0.5468	0.4678	0.4581	0.4008	0.4363
151508	0.5432	0.3883	0.3391	0.4910	0.5459
151509	0.5070	0.4552	0.3561	0.4276	0.5065
151510	0.5325	0.4431	0.3317	0.4998	0.4766
151669	0.5118	0.2479	0.3033	0.6094	0.6566
151670	0.2521	0.1951	0.3619	0.5203	0.4456
151671	0.6020	0.3395	0.4609	0.6117	0.6400
151672	0.5183	0.4131	0.5243	0.6162	0.7692
151673	0.5956	0.4947	0.3826	0.4008	0.6294
151674	0.3800	0.3910	0.3864	0.5385	0.6333
151675	0.5960	0.4011	0.5182	0.5688	0.5567
151676	0.5960	0.3186	0.5836	0.4290	0.5992
Average	0.5151	0.3796	0.4172	0.5095	**0.5746**

Table 2: The ARI of different methods on the DLPFC.

Spatial Domain Identification on the Breast Cancer Dataset

To verify the spatial domain identification ability and generalization of GATCL for cancer tissues, we conduct an experimental comparison with the four baseline methods (CCST, SpaGCN, STAGATE, and GraphST) on the breast cancer public dataset. As shown in Figure 4, the dataset has 20 manually annotated regions. The same clustering number setting as for the annotation is used, and the ARI clustering metric is used to quantitatively compare the performance of the different methods. As shown in Figure 5, GATCL achieves the highest ARI (ARI=0.6182). This result indicates that GATCL has good generalization for cancer tissues, and the adaptive strategy is effective in terms of obtaining feature representations for spatial transcriptomics data from different tissues.

As shown in Figure 4, we have visualized the results for different methods. The regions identified by GATCL have the highest agreement with the annotations. In addition, there are continuous boundaries between different regions.

Spatial Domain Identification on the Mouse Somatosensory Cortex Dataset

To verify the spatial domain identification ability of GATCL for single-cell spatial transcriptomics, we conduct an experimental comparison with the four baseline methods (CCST, SpaGCN, STAGATE, and GraphST) on the mouse somatosensory cortex dataset. We adopt the image-free mode on the spaGCN method because the mouse somatosensory cortex dataset does not contain histological images. As

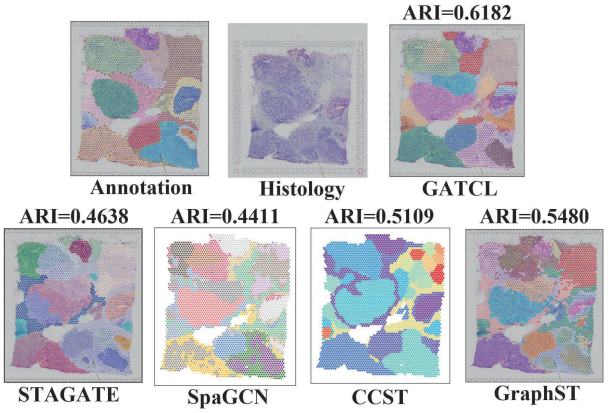

Figure 4: The visualization results on the breast cancer dataset.

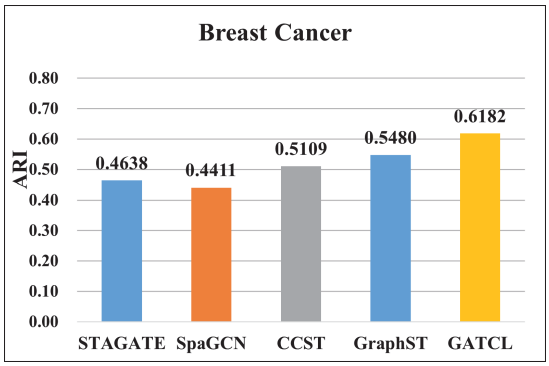

Figure 5: The ARI of different methods on the breast cancer dataset.

shown in Figure 6, the dataset has 11 manually annotated regions. The clustering number setting is consistent with the annotation, and the ARI clustering metric is employed to evaluate and compare the performance of the various methods. As shown in Figure 7, GATCL achieved the highest ARI (ARI=0.6496). This result indicates that GATCL also has a good effect on spatial domain identification of spatial transcriptomics at single-cell resolution.

The visualization results of each method on the mouse somatosensory cortex dataset are shown in Figure 6. GATCL can accurately identify Layer 4, Layer 6, and Pia Layer 1. Besides Layer 2-3 lateral, other layers also exhibit high overlap with the ground truth.

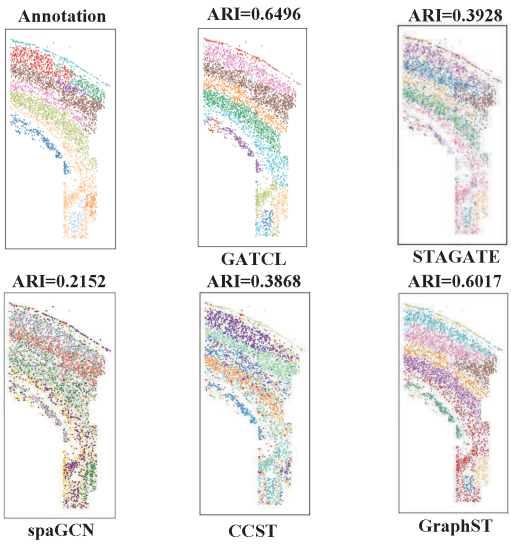

Figure 6: The visualization results on the mouse somatosensory cortex.

Ablation Study

To further investigate the contribution of each module to our method GATCL, we performed an ablation study on the DLPFC dataset. We ablated the MHGAT, the attention pooling mechanism, and the adaptive graph contrastive learning modules, respectively, to evaluate their contributions to the performance of GATCL.

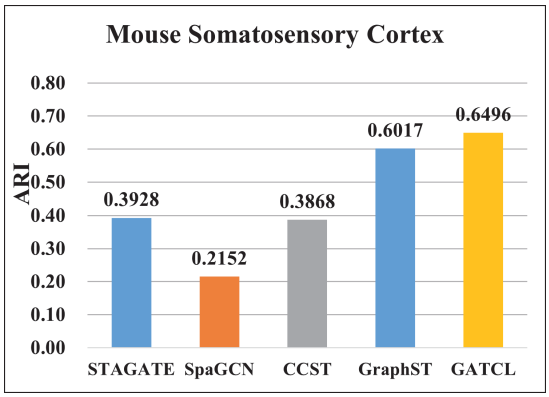

Figure 7: The ARI of different methods on the mouse somatosensory cortex.

GATCL-w/o-MHGAT: Remove MHGAT and replace it with graph convolutional neural Network (GCN).

GATCL-w/o-AP: Remove attention pooling and replace it with average pooling.

GATCL-w/o-AGCL: Remove the adaptive graph contrastive learning modules.

As shown in Table 4, we first tested whether the performance of GATCL benefits from the MHGAT module. We found that the performance decreases by 5.78% if MHGAT is not used. Secondly, we can observe that the performance of GATCL-w/o-AGCL is reduced by 10.84% compared to GATCL, highlighting its importance in feature encoding. Further, GATCL-w/o-AP showed higher results than GATCL-w/o-AGCL and lower results than GATCL. This suggests that attention pooling is able to adaptively aggregate neighbor spots to learn the gene expression characteristics of spots. Finally, it was further verified on two datasets of breast cancer and mouse somatosensory cortex dataset. As shown in Table 4, we found similar trends as on the DLPFC dataset.

Sample	w/o-MHGAT	w/o-AP	w/o-AGCL	GATCL
151507	0.4206	0.3670	0.4316	0.4363
151508	0.5341	0.5145	0.3900	0.5459
151509	0.4873	0.4300	0.4154	0.5065
151510	0.4701	0.4991	0.5376	0.4766
151669	0.5022	0.6561	0.5123	0.6566
151670	0.5505	0.6459	0.4565	0.4456
151671	0.6465	0.6504	0.6488	0.6400
151672	0.6581	0.5218	0.6527	0.7692
151673	0.6089	0.6315	0.6338	0.6294
151674	0.5627	0.5805	0.6348	0.6333
151675	0.5858	0.5348	0.4509	0.5567
151676	0.4705	0.6257	0.3835	0.5992
Average	0.5414	0.5548	0.5123	**0.5746**

Table 3: Ablation study on the DLPFC.

Method	Breast cancer	MSC
GATCL-w/o-MHGAT	0.5314	0.5440
GATCL-w/o-AP	0.5791	0.5896
GATCL-w/o-AGCL	0.5201	0.5136
GATCL	**0.6182**	**0.6496**

Table 4: Ablation study on the breast cancer and mouse somatosensory cortex dataset (MSC).

Conclusion

In this paper, we propose an adaptive graph contrastive learning framework based on multi-head graph attention networks (GATCL) for spatial domain identification. One property of the proposed method is that it combines MHGAT with adaptive graph contrastive learning to encode gene expression profiles and spatial location information. In addition, to improve the embedding representation of the graph, the attention pooling mechanism dynamically and adaptively aggregates the neighborhood information of the spots and improves the model's generalization ability for different spatial transcriptomics data. The experimental results demonstrate that GATCL achieves the best performance on the DLPFC and human breast cancer and mouse somatosensory cortex datasets, and has clear continuous hierarchical boundaries. Therefore, the proposed method provides new tools and ideas for understanding tissue complexity and gaining spatial biological insights. In the future, it is considered to combine with histological images to further improve the effect of spatial domain identification, and to explore the mechanism of occurrence and development in cancer.

Acknowledgments

This study is supported by the grant with No. 62572157, No. KYQD(ZR)-21079 and Hainan Provincial Natural Science Foundation of China with No. 825CXTD608.

References

Asp, M.; Bergenstråhle, J.; and Lundeberg, J. 2020. Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays*, 42(10): 1900221.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10): P10008.

Chen, W.; Wan, H.; Wu, Y.; Zhao, S.; Cheng, J.; Li, Y.; and Lin, Y. 2024. Local-global history-aware contrastive learning for temporal knowledge graph reasoning. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 733–746. IEEE.

Chu, G.; Wang, X.; Shi, C.; and Jiang, X. 2021. CuCo: Graph Representation with Curriculum Contrastive Learning. In *IJCAI*, 2300–2306.

Codeluppi, S.; Borm, L. E.; Zeisel, A.; La Manno, G.; van Lunteren, J. A.; Svensson, C. I.; and Linnarsson, S. 2018. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods*, 15(11): 932–935.

Dong, K.; and Zhang, S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature Communications*, 13(1): 1739.

Fraley, C.; Raftery, A.; Scrucca, L.; Murphy, T.; and Fop, M. 2014. mclust: Normal mixture modeling for model-based clustering, classification, and density estimation. *R Package Version*, 4(7).

Fu, H.; Xu, H.; Chong, K.; Li, M.; Ang, K. S.; Lee, H. K.; Ling, J.; Chen, A.; Shao, L.; Liu, L.; et al. 2021. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Biorxiv*. https://doi.org/10.1101/2021.06.15.448542.

Guan, X.; and Zhang, D. 2023. T-MGCL: Molecule Graph Contrastive Learning Based on Transformer for Molecular Property Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Hu, J.; Li, X.; Coleman, K.; Schroeder, A.; Ma, N.; Irwin, D. J.; Lee, E. B.; Shinohara, R. T.; and Li, M. 2021. 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.

Huang, Y.; Lu, W.; and Yang, Y. 2024. An efficient prototype-based clustering approach for edge pruning in graph neural networks to battle over-smoothing. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 4201–4209.

Jin, S.; Chen, Z.; Yu, S.; Altaf, M.; and Ma, Z. 2023. Self-Augmentation Graph Contrastive Learning for Multi-view Attribute Graph Clustering. In *Proceedings of the 2023 Workshop on Advanced Multimedia Computing for Smart Manufacturing and Engineering*, 51–56.

Li, B.; Zhang, W.; Guo, C.; Xu, H.; Li, L.; Fang, M.; Hu, Y.; Zhang, X.; Yao, X.; Tang, M.; et al. 2022a. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nature Methods*, 19(6): 662–670.

Li, J.; Chen, S.; Pan, X.; Yuan, Y.; and Shen, H.-B. 2022b. Cell clustering for spatial transcriptomics data with graph neural networks. *Nature Computational Science*, 2(6): 399–408.

Liu, H.; Li, C.; Zhang, X.; Zhang, F.; Wang, W.; Ma, F.; Chen, H.; Yu, H.; and Zhang, X. 2024. Depression Detection via Capsule Networks with Contrastive Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 22231–22239.

Liu, Z.; Liu, F.; Hong, C.; Gao, M.; Chen, Y.-P. P.; Liu, S.; and Zhang, W. 2019. Detection of cell types from single-cell RNA-seq data using similarity via kernel preserving learning embedding. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 451–457. IEEE.

Long, Y.; Ang, K. S.; Li, M.; Chong, K. L. K.; Sethi, R.; Zhong, C.; Xu, H.; Ong, Z.; Sachaphibulkij, K.; Chen, A.; et al. 2023. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications*, 14(1): 1155.

Lu, J.; Lin, H.; Zhang, X.; Li, Z.; Zhang, T.; Zong, L.; Ma, F.; and Xu, B. 2023. Hate speech detection via dual contrastive learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Maynard, K. R.; Collado-Torres, L.; Weber, L. M.; Uytingco, C.; Barry, B. K.; Williams, S. R.; Catallini, J. L.; Tran, M. N.; Besich, Z.; Tippani, M.; et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature Neuroscience*, 24(3): 425–436.

Moncada, R.; Barkley, D.; Wagner, F.; Chiodin, M.; Devlin, J. C.; Baron, M.; Hajdu, C. H.; Simeone, D. M.; and Yanai, I. 2020. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature Biotechnology*, 38(3): 333–342.

Rao, A.; Barkley, D.; França, G. S.; and Yanai, I. 2021. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871): 211–220.

Ren, L.; Wang, J.; Li, W.; Guo, M.; and Yu, G. 2024. Single-cell RNA-seq data clustering by deep information fusion. *Briefings in Functional Genomics*, 23(2): 128–137.

Shao, X.; Li, C.; Yang, H.; Lu, X.; Liao, J.; Qian, J.; Wang, K.; Cheng, J.; Yang, P.; Chen, H.; et al. 2022. Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk. *Nature Communications*, 13(1): 4429.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *Stat*, 1050(20): 10–48550.

Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep graph infomax. *ICLR (Poster)*, 2(3): 4.

Wolf, F. A.; Angerer, P.; and Theis, F. J. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19: 1–5.

Zhang, X.; Liu, H.; Li, Q.; and Wu, X. M. 2019. Attributed graph clustering via adaptive graph convolution. In *28th International Joint Conference on Artificial Intelligence, IJCAI 2019*, 4327–4333. International Joint Conferences on Artificial Intelligence.

Zhao, C.; Liu, S.; Huang, F.; Liu, S.; and Zhang, W. 2021. CSGNN: Contrastive Self-Supervised Graph Neural Network for Molecular Interaction Prediction. In *IJCAI*, 3756–3763.

Zheng, L.; Liu, Z.; Yang, Y.; and Shen, H.-B. 2022. Accurate inference of gene regulatory interactions from spatial gene expression with deep contrastive learning. *Bioinformatics*, 38(3): 746–753.

Zhu, Q.; Li, A.; Zhang, Z.; Zheng, C.; Zhao, J.; Liu, J.-X.; Zhang, D.; and Shao, W. 2024. Discriminative Domain Adaption Network for Simultaneously Removing Batch Effects and Annotating Cell Types in Single-Cell RNA-Seq. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.