Cluster-guided contrastive learning with masked autoencoder for spatial domain identification based on spatial transcriptomics

Juan Wang[1,2], Qi Gao[1], Shasha Yuan[1], and Junliang Shang[1]

1. School of Computer Science, Qufu Normal University, Rizhao 276826, China

2. RizhaoQufu Normal University Joint Technology Transfer Center, Qufu Normal University, Rizhao 276826, China

A. Details of dataset

(1) Human dorsolateral prefrontal cortex (DLPFC): dataset includes 12 slices manually annotated by Maynard et al. [1] based on gene markers. Each slice contains 4–6 cortical layers and one WM layer.

(2) Human breast cancer (HBC): This dataset [2] consists of four primary regions, including ductal carcinoma in situ/lobular carcinoma in situ (DCIS/LCIS), healthy tissue, invasive ductal carcinoma (IDC), and low malignancy tumor-adjacent regions (Tumor_edge), encompassing 20 regions.

(3) Mouse brain anterior (MBA): The MBA dataset is manually annotated into 52 identifiable regions based on the Allen Mouse Brain Reference Atlas.

(4) Mouse visual cortex (MVC): The MVC dataset [3] is organized into 7 structural layers.

(5) Mouse embryo (ME9.5): The ME9.5 dataset represents the E9.5 developmental stage and consists of 12 tissue regions manually annotated by Chen et al. [4].

(6) Mouse olfactory bulb (MOB): The MOB dataset obtained from the Stereo-seq platform, features subcellular resolution (14 um) and is broadly divided into 7 structural layers: olfactory nerve layer (ONL), glomerular layer (GL), external plexiform layer (EPL), mitral cell layer (MCL), internal plexiform layer (IPL), granule cell layer (GCL), and rostral migratory stream (RMS).

(7) Mouse olfactory bulb (MOBV2): The MOBV2 dataset from the Slide-seqV2 platform (10 um), provides a higher-resolution description of spatial expression. Similar to the MOB dataset, it is divided into 7 layers.

B. Hyperparameters selection and analysis

Effect of KNN (k)

We investigate the impact of the k value on different datasets (Fig. S1A). The results indicate that k too high or too low impairs the model's ability to capture complex relationships between spots, and the model performance is better when the k value is between 6 and 10.

Effect of ς

The ς controls the number of high-confidence samples. Therefore, ς plays a crucial role in the performance. As shown in Fig. S1B, an appropriately selected value of ς leads to higher-quality positives and negatives, which further improves the discriminability of the learned embeddings.

Effect of learning rate (lr)

In this section, we vary learning rate (lr) to investigate its impact. Fig. S1C reveals that the MVC dataset is more sensitive about lr and exhibits noticeable fluctuations. Moreover, lower lr values (0.005 ~ 0.00005) are more conducive to stable training, with the best performance achieved when lr is set to 0.0001.

Effect of mask ratio and remask ratio

The masking and re-masking probability affects the construction of views and influences the reconstruction of the expression matrix, respectively. To illustrate this, we visualize the effects of these parameters on the HBC dataset in Fig. S1D. The results show that a higher re-mask rate and mask rate (e.g., 0.8) enable the model to learn informative embeddings.

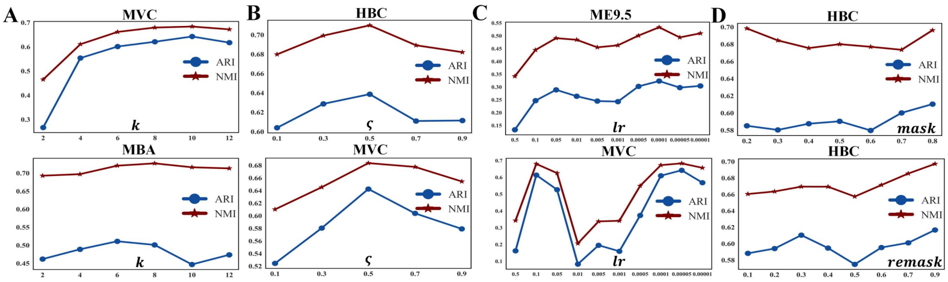

Fig. S1: Impact of hyperparameter settings.

Effect of loss term

To explore the impact of each loss on model performance and their relative importance, we evaluate how variations of individual loss terms affect cluster metrics on the HBC dataset. In Fig. S2, we observe that the L_{cos} has minimal impact on the stability of NMI and ARI, but it has a more pronounced effect on F1 and ACC. Additionally, L_{multi} (sum of L_{d} and L_{kl}) shows that increasing the weight of L_{d} leads to a noticeable decline in F1 and also degrades other cluster metrics. In contrast, L_{kl} exhibits minor fluctuations in its impact. Considering the combined effects of L_{d} and L_{kl}, we set their weights to 1. Moreover, L_{latent} achieves the best performance when its weight is 1, and its values being too high or too low may result in values degradation. Lastly, ACC and ARI appear to be more sensitive to L_{rec}, suggesting that the reconstruction loss helps improve overall accuracy.

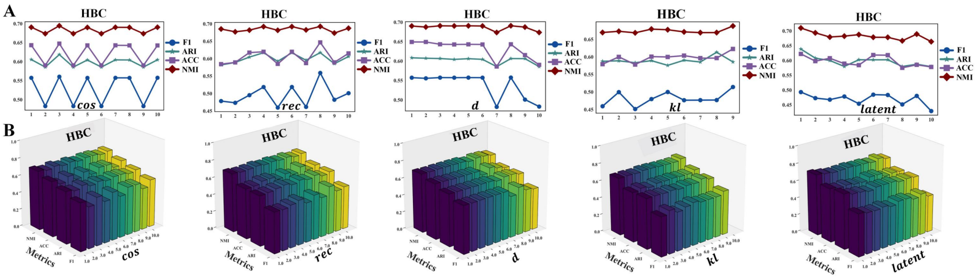

Fig. S2: Sensitivity analysis of loss parameters.

C. Impact of different GNN convolutional blocks

We evaluate the impact of GNN convolutional blocks in STMCCL, specifically comparing the use of GIN and GCN for the encoder (E_1), decoder (D), and masked encoder (E_2). Table III presents the results of four datasets. On the HBC and MVC datasets, the configuration with a GIN-based E_1 and GCN-based D and E_2 achieves the highest accuracy. Conversely, for the MBA dataset, the combination of a GIN-based E_1 and E_2 pair with a GCN-based D produces the most accurate cluster outcomes. Overall, with a GIN-based E_1 and GCN-based D and E_2, it proves to be more effective cluster results.

GNNs			DLPFC		HBC		MBA		MVC	
E1	D	E2	ARI	NMI	ARI	NMI	ARI	NMI	ARI	NMI
GCN	GCN	GCN	**0.6625**	0.6970	0.5606	0.6644	0.4576	0.7188	0.5684	0.6569

GCN	GCN	GIN	0.6442	0.6803	0.5534	0.6635	0.4480	0.7156	0.5390	0.6623
GIN	GCN	GCN	0.6522	**0.6978**	**0.6387**	**0.7098**	0.5112	0.7206	**0.6425**	**0.6836**
GIN	GCN	GIN	0.6295	0.6687	0.6232	0.6843	**0.5559**	**0.7386**	0.6138	0.6805
GCN	GIN	GCN	0.6463	0.6740	0.5587	0.6644	0.4538	0.7198	0.6101	0.6712
GCN	GIN	GIN	0.6435	0.6868	0.5468	0.6724	0.4389	0.7126	0.5268	0.6254
GIN	GIN	GCN	0.6424	0.6776	0.6143	0.6699	0.4146	0.7089	0.5067	0.5977
GIN	GIN	GIN	0.6061	0.6661	0.6225	0.6879	0.4996	0.7207	0.5017	0.6041

TABLE SI: The impact of different graph convolution blocks.

D. Clustering evaluation metrics

For different datasets, we used various clustering metrics to assess the expressiveness of the embeddings extracted by STMCCL. The ARI [5] is formulated as:

$$ARI = \frac{RI\text{-}E[RI]}{max(RI)\text{-}E[RI]},$$

where the unadjusted rand index (RI) is defined as $RI=(a+b)/C_n^2$, and a denotes the number of pairs correctly labeled as coming from the same set. Additionally, b is the number of pairs correctly labeled as not in the same set, C_n^2 defines the total number of possible pairs, and $E[RI]$ represents the expected RI of random labeling.

Mutual information (MI) measures the similarity between ground truth and predicted clusters, which is defined as:

$$MI(U,V) = \Sigma_{i=1}^{|U|}\Sigma_{j=1}^{|V|}\frac{|U_i\cap V_j|}{N}log\frac{N|U_i\cap V_j|}{|U_i||V_j|}$$

where $|U_i|$ is the number of the samples in cluster U_i , and $|V_j|$ denotes the number of the samples in cluster V_j . Moreover, $|V_j|\cdot MI$ is generally higher for clustering results with larger number of clusters. To account this bias, the Normalized Mutual Information (NMI) [6] was calculated to remove the effect of cluster numbers:

$$NMI(U,V) = \frac{MI(U,V)}{F(H(U),H(V))}$$

where F can find functions of maximum, minimum, geometric mean and arithmetic mean.

$F1$ score is an indicator used in statistics to measure the accuracy of the model. It is denoted as:

$$F1 = 2\frac{P\cdot R}{P+R}$$

where P represents precision, which measures the proportion of correctly predicted positive samples among all samples classified as positive. Similarly, R denotes recall, also known as the retrieval rate, which quantifies the proportion of correctly identified positive samples among all actual positive samples.

Finally, the formulas for SC [7] and DB [8] are as follows:

$$SC: \ s(i) = \frac{a(i)\text{-}b(i)}{max(a(i),b(i))}; \quad DB: \ R_{pq} = 2\frac{S_p+S_q}{d_{pq}}$$

where $a(i)$ represents the average distance from point i to other points within the same cluster, and $b(i)$ defines the average distance from i to all points in the nearest cluster. Additionally, S_p and S_q are the average intra-cluster distances for cluster p and cluster q, respectively. Furthermore, d_{pq} is the distance between cluster p and cluster q. The DB index is the average of the maximum R_{pq} value across all cluster pairs.

E. Visualization

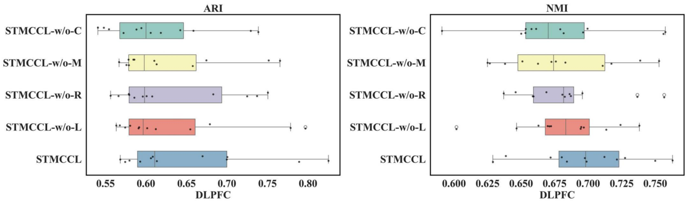

Fig. S3: The ARI and NMI boxplots of STMCCL and its variants on DLPFC.

F. References

[1] K. R. Maynard, L. Collado-Torres, L. M. Weber, C. Uytingco, B. K. Barry, S. R. Williams, J. L. Catallini, M. N. Tran, Z. Besich, M. Tippani et al., “Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex,” Nature neuroscience, vol. 24, no. 3, pp. 425–436, 2021.

[2] E. Buache, N. Etique, F. Alpy, I. Stoll, M. Muckensturm, B. Reina-San-Martin, M. Chenard, C. Tomasetto, and M. Rio, “Deficiency in trefoil factor 1 (tff1) increases tumorigenicity of human breast cancer cells and mammary tumor development in tff1-knockout mice,” Oncogene, vol. 30, no. 29, pp. 3261–3273, 2011.

[3] X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu et al., “Three-dimensional intact-tissue sequencing of single-cell transcriptional states,” Science, vol. 361, no. 6400, p. eaat5691, 2018.

[4] L. Richardson, S. Venkataraman, P. Stevenson, Y. Yang, J. Moss, L. Graham, N. Burton, B. Hill, J. Rao, R. A. Baldock et al., “Emage mouse embryo spatial gene expression database: 2014 update,” Nucleic acids research, vol. 42, no. D1, pp. D835–D844, 2014.

[5] W. M. Rand, “Objective criteria for the evaluation of clustering methods,” Journal of the American Statistical association, vol. 66, no. 336, pp. 846–850, 1971.

[6] A. Strehl and J. Ghosh, “Cluster ensembles—a knowledge reuse framework for combining multiple partitions,” Journal of machine learning research, vol. 3, no. Dec, pp. 583–617, 2002.

[7] A. M. Bagirov, R. M. Aliguliyev, and N. Sultanova, “Finding compact and well-separated clusters: Clustering using silhouette coefficients,” Pattern Recognition, vol. 135, p. 109144, 2023.

[8] S. Petrovic, “A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters,” in Proceedings of the 11th Nordic workshop of secure IT systems, vol. 2006. Citeseer, 2006, pp. 53–64.