

Supplementary Materials: Heterogeneous Graph Guided Contrastive Learning for Spatially Resolved Transcriptomics Data

Anonymous Authors

1 COMPARED METHODS

To demonstrate the superior performance of the proposed stGCL, we have chosen six benchmark methods for comparative analysis, including DeepST [6], GraphST [3], SCANPY [5], SCGDL [2], SpaGCN [1], and Spatial-MGCN [4].

- We propose a stGCL for spatial transcriptomics data that end-to-end combines the heterogeneity of genetic and spatial a priori distributions to learn the intrinsic local organization of cells, providing a novel perspective on the mechanism of cellular interactions to address the coordination between different views in tissue.
- **DeepST** [6] employs a denoising autoencoder network and a variational graph autoencoder to derive latent embeddings for individual spots.
- **GraphST** [3] combines GNNs with self-supervised contrastive learning to capture information and distinctive representations of spots, revealing different cell types.
- **SCANPY** [5] is an extensible toolkit designed specifically for analyzing single-cell gene expression data. It offers methods for preprocessing, visualization, and clustering. This toolkit offers researchers a comprehensive solution for managing and scrutinizing single-cell RNA sequencing data.
- **SCGDL** [2] combines deep graph infomax (DGI) with residual gated graph convolutional neural network for spatial domain identification.
- **SpaGCN** [1] effectively leverages GCN to combine gene expression, spatial coordinates, and histological images for spatial domain identification.
- **Spatial-MGCN** [4] employs an attention-driven multi-view GCN to derive spatial and gene expression information for spatial clustering.

2 DATA PREPARATION

To reduce technical noise in spatial transcriptomics data, we first eliminate spots located outside the primary tissue regions. Subsequently, we employ the SCANPY [5] package to filter out genes with low expression or variance and select the top 3000 genes with high variability [18]. Finally, these genes are normalized utilizing a scale factor [14]. The normalization function is defined as:

$$X_{ij} = \frac{\text{count}_{ij}}{\sum_j \text{count}_{ij}} \times 10000, \quad (1)$$

where X_{ij} represents the raw expression value of the j -th gene in the i -th spot.

3 ANALYZE OF SPATIAL TRANSCRIPTOMICS

We analyze the traditional fusion paradigm of spatial transcriptomics from an information theory perspective.

Theorem 1. In the fusion of multiple views into a compact fusion feature, the mapping to global spatial deflation strategy is better expressed and more effective than direct fusion.

Proof. According to the fundamental principles of information theory, let us consider two non-orthogonal modes X and Y with sizes x and y , respectively, and an aggregate mode F of size f . When $f \geq x + y$, a mode Z can be learned that contains all information of X and Y , that is,

$$\begin{aligned} I(F; X, Y) &= H(X, Y) - H(X, Y|F), \\ &= - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log p(x_i, y_j), \end{aligned} \quad (2)$$

where I denotes mutual information and H denotes entropy. This implies that Z has the same entropy as the joint distributions of X and Y . The entropy of F is expressed as:

$$H(F) = H(X) + H(Y) + H(A) > H(X) + H(Y) > I(F; X, Y) = H(X, Y), \quad (3)$$

where $H(A)$ represents the entropy of the association information between X and Y .

Given the sparsity of gene data, the aggregation process should satisfy $f < x + y$, aiming for a compact model that captures the association and original representations of X and Y within a smaller distribution. Typically, we observe that $f = x = y$ in spatial transcriptomics data. The mutual information learned, $I(F; X) + I(F; Y)$, represents the sum of information about X and Y that can be recovered in F . When $f < x + y$, the maximization tends to focus on the shared information of X and Y , denoted as:

$$\max\{I(F; X) + I(F; Y)\}, \quad (4)$$

while neglecting the orthogonal (independent, non-overlapping) information components, leading to a biased representation of F towards either X or Y . This results in an incoherent expression of the aggregated modality.

Considering spatial and genetic data heterogeneity in spatial transcriptomics, the joint entropy $H(X, Y)$ is minimal and challenging to aggregate, complicating the development of a compact, efficient, and effective representation. This complexity is particularly evident in multiple downstream tasks, where a task may depend on the independent information components of X or Y , potentially leading to a collapse of the learned modal aggregation information. Consequently, the learned Z may need more capacity to adequately express X and Y .

Inspired by archery, our method leverages the prior distributions $P(T|(X, Y))$ of X and Y through an intermediate distribution T of size t , where $t > x + y > f$. As illustrated in Equation 1,

$$H(T) = H(X) + H(Y) + H(A), \quad (5)$$

indicating that T captures information from both X and Y . The information within T is subsequently subjected to low-pass filtering, and T is compressed through a low-pass filter to obtain unbiased

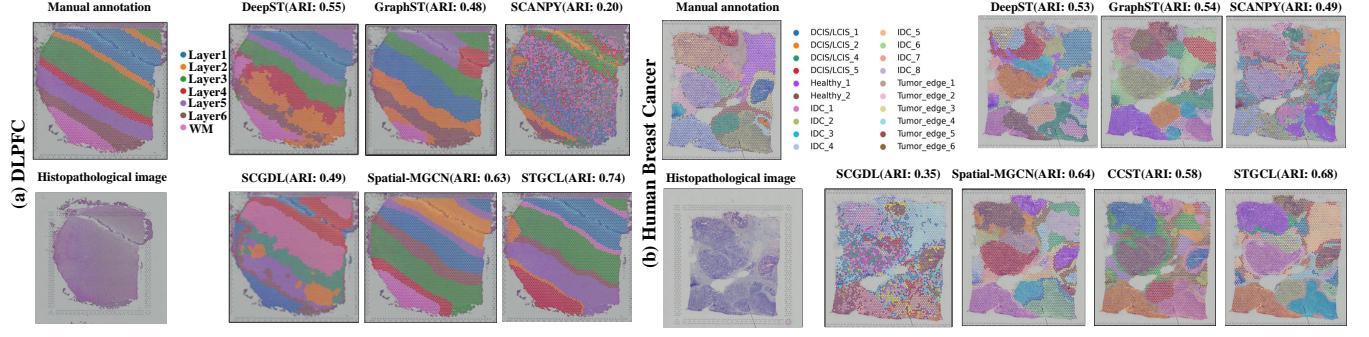


Figure 1: Identify spatial domains for comparison experiments between DLPFC and Human Breast Cancer datasets. The manual annotation of the slice # 151507 in DLPFC.

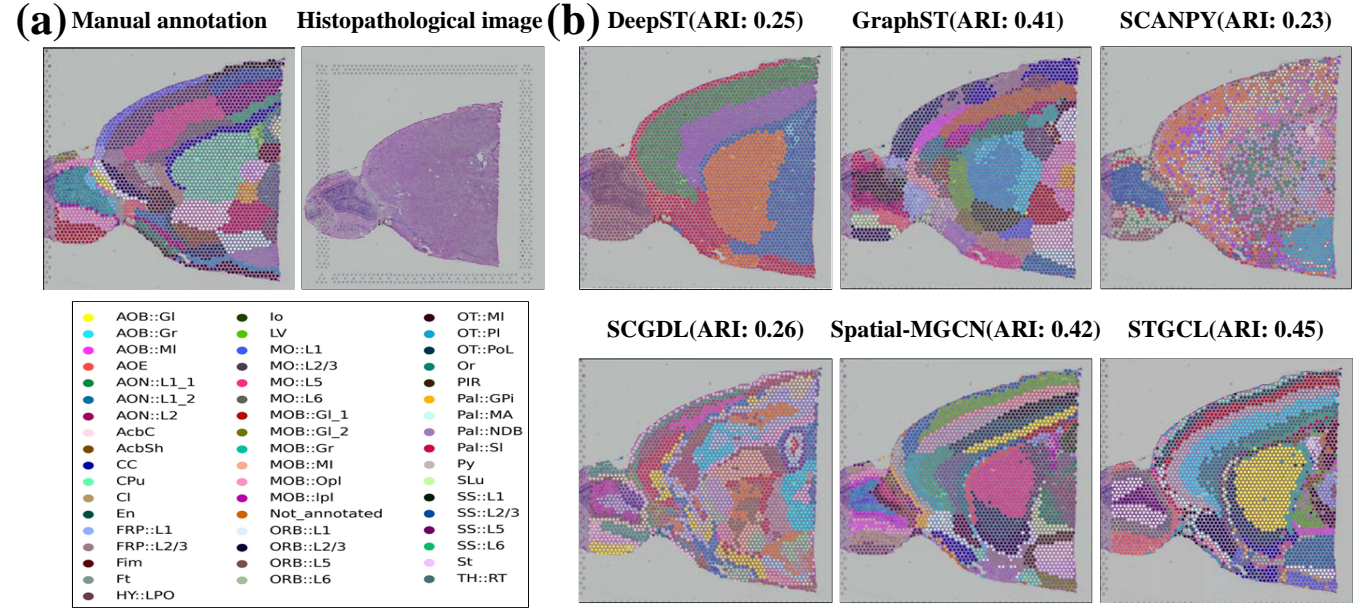


Figure 2: Identify spatial domains on Mouse Brain Anterior Tissue dataset. (a) Manual annotation layer structure and the histopathological image for human breast cancer dataset. (b) Spatial domains are detected with stGCL and five methods.

estimates F of X and Y , where $I(F; T)$. The refined entropies are given by

$$H(F) = H(X_{\text{refine}}) + H(Y_{\text{refine}}) + H(A_{\text{refine}}), \quad (6)$$

where $X_{\text{refine}} \subseteq X$, $Y_{\text{refine}} \subseteq Y$, and $Z_{\text{refine}} \subseteq Z$. This strategy F learns multi-view orthogonality and shared essential features by aggregating X and Y through T rather than directly through F , thus enabling the alignment mapping of X and Y to be found in F .

4 EXPERIMENTAL RESULTS

4.1 Experimental settings

We assess the clustering performance by utilizing the Adjusted Rand Index (ARI) as a metric, which evaluates the similarity between the predicted cluster labels and the ground truth cluster labels. A higher ARI value signifies superior clustering performance. The

calculation is performed as follows:

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}, \quad (7)$$

where $E[RI]$ indicates the RI that would be achieved by random labeling.

4.2 More Visualization

Due to page limitations, we have zoomed in on key visualization images from the main text to demonstrate the superiority of our stGCL.

REFERENCES

- [1] Jian Hu, Xiangjie Li, Kyle Coleman, Amelia Schroeder, Nan Ma, David J Irwin, Edward B Lee, Russell T Shinohara, and Mingyao Li. 2021. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and

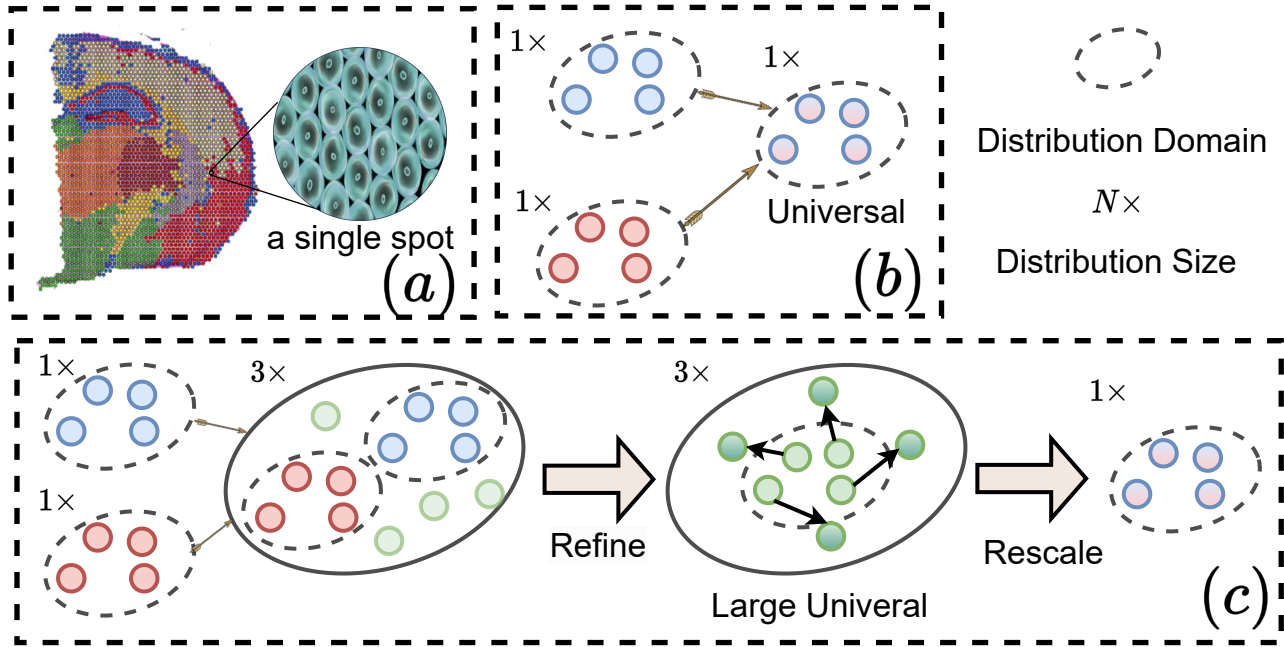


Figure 3: (a) A spot in spatial transcriptomics indicates multiple cellular genes. (b) Typical methods aggregate multi-modal data into a compact universal distribution. (c) Our proposed mapping-to-expansion paradigm maximizes the preservation of the original modal molecular information.

- spatially variable genes by graph convolutional network. *Nature methods* 18, 11 (2021), 1342–1351.
- [2] Teng Liu, Zhao-Yu Fang, Xin Li, Li-Ning Zhang, Dong-Sheng Cao, and Ming-Zhu Yin. 2023. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *Briefings in Bioinformatics* 24, 3 (2023), bbad146.
- [3] 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. 2023. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nature Communications* 14, 1 (2023), 1155.
- [4] Bo Wang, Jiawei Luo, Ying Liu, Wanwan Shi, Zehao Xiong, Cong Shen, and Yahui Long. 2023. Spatial-MGCN: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. *Briefings in Bioinformatics* 24, 5 (2023), bbad262.
- [5] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome biology* 19 (2018), 1–5.
- [6] Chang Xu, Xiyun Jin, Songren Wei, Pingping Wang, Meng Luo, Zhaochun Xu, Wenyi Yang, Yideng Cai, Lixing Xiao, Xiaoyu Lin, et al. 2022. DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Research* 50, 22 (2022), e131–e131.

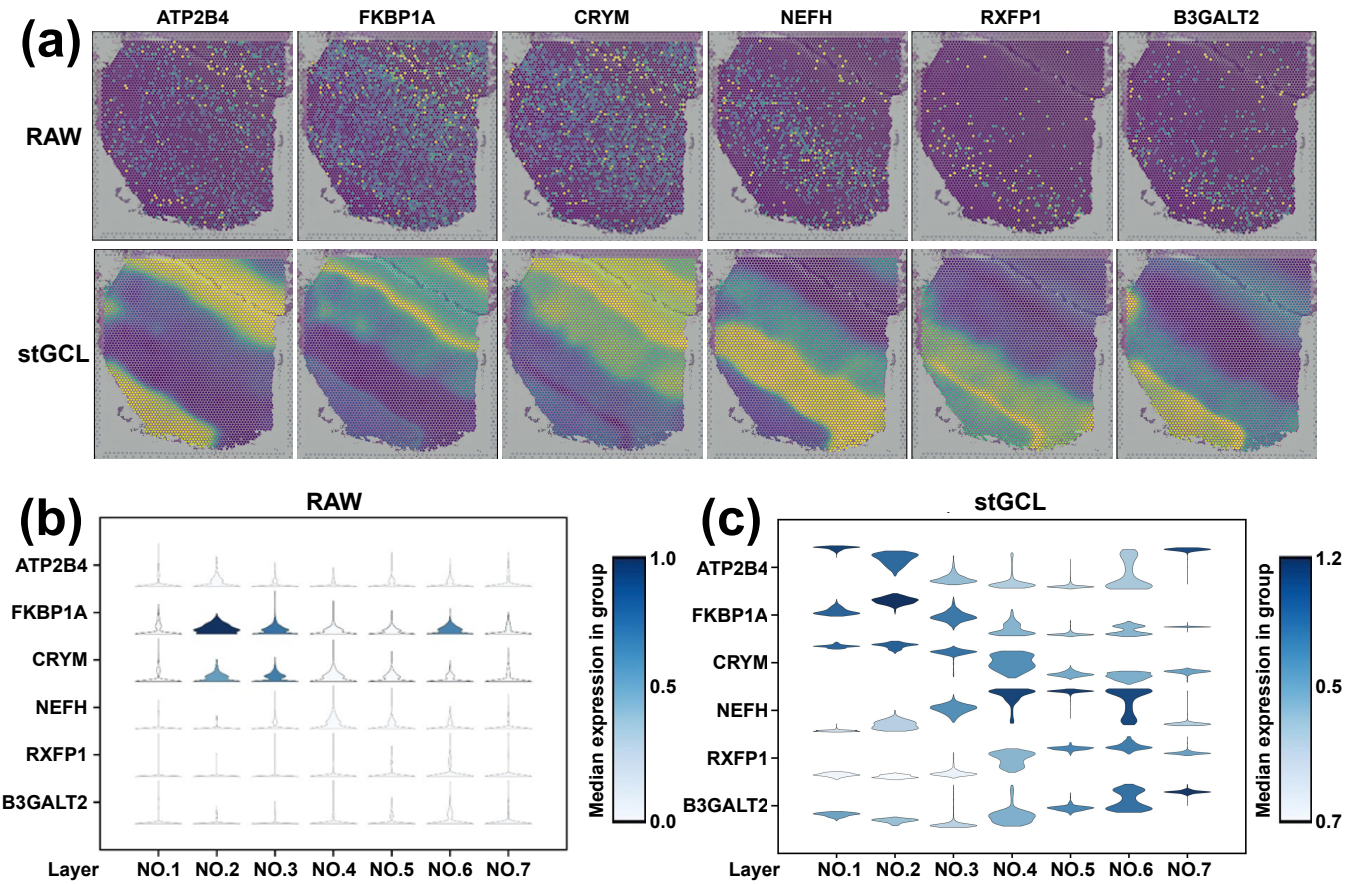


Figure 4: (a) The visualization of raw expression of layer marker genes and expression after stGCL imputation. (b) The violin plots of raw cortical marker gene expression and stGCL cortical marker gene expression. (c) The violin plot of cortical marker gene expression imputed by stGCL.

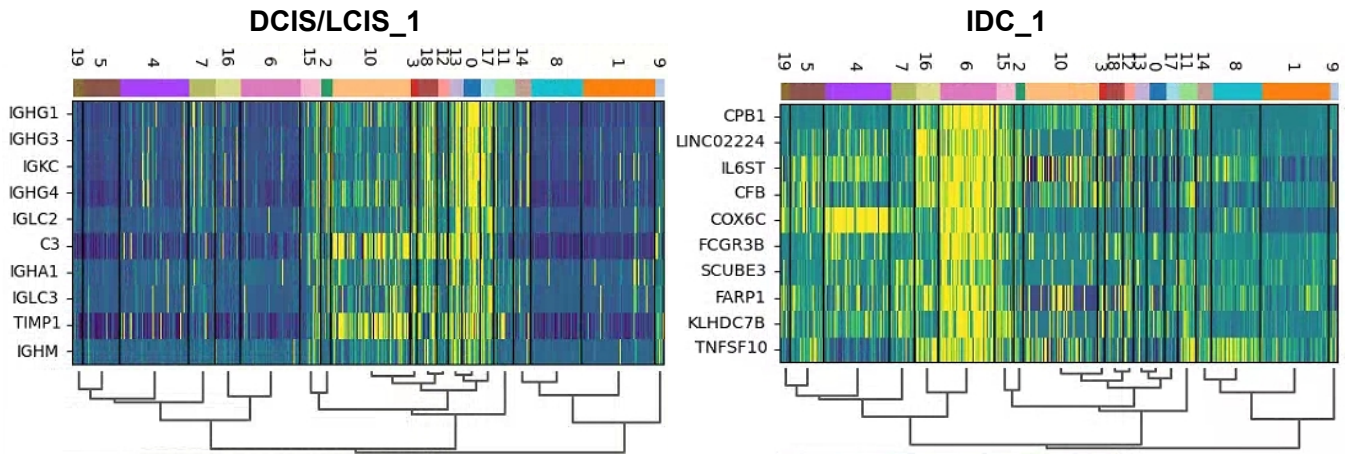


Figure 5: The Heatmap of the expression of the structural domains on the top 10 DEGs between Healthy 1 and DCIS/LCIS.

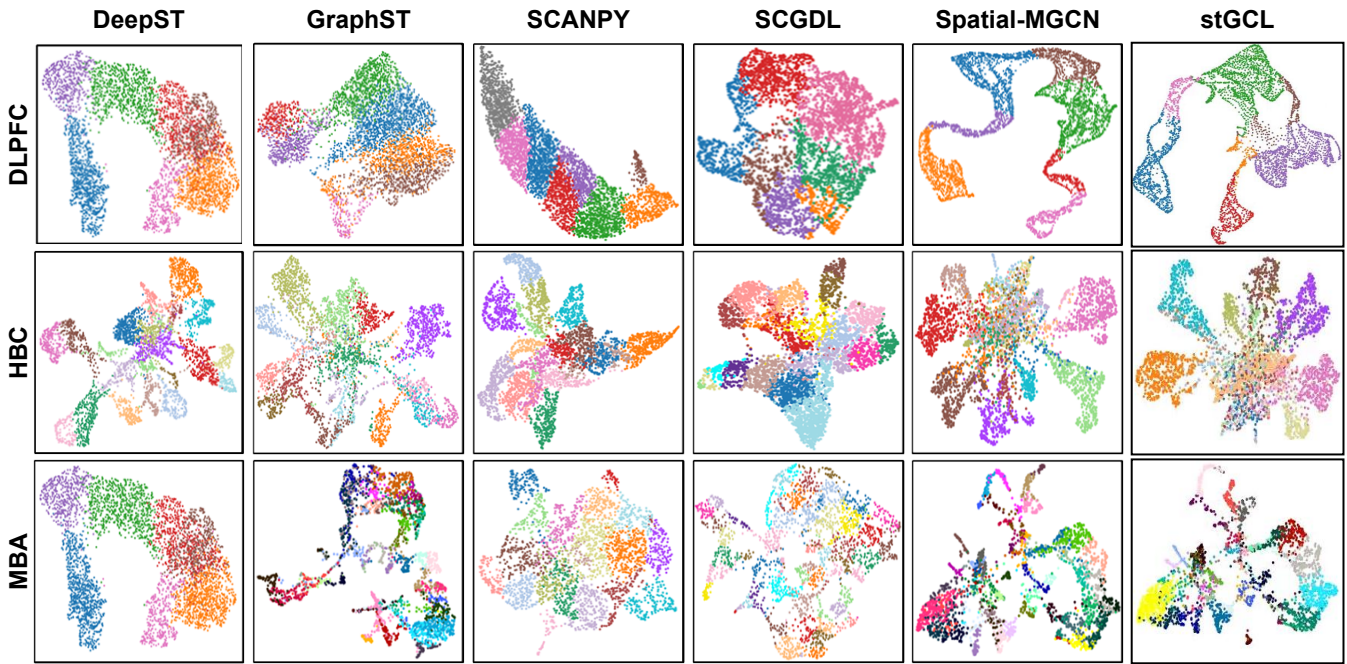


Figure 6: The overall architecture of the proposed stCHG framework.

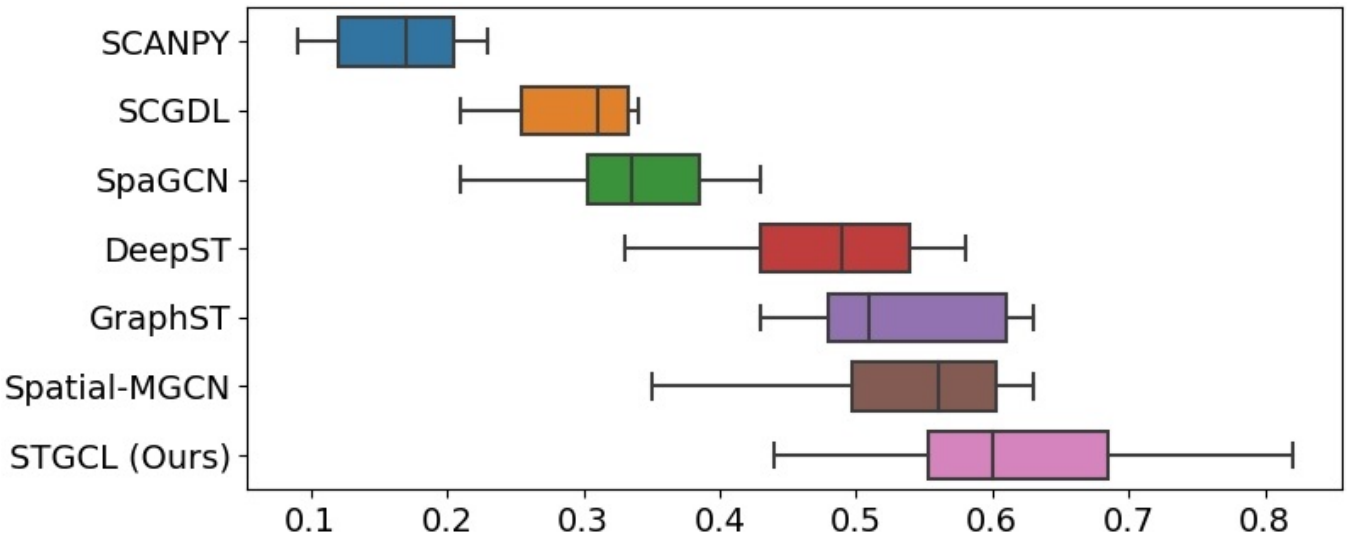


Figure 7: Boxplots of ARI values for seven methods across 12 slices of DLPFC.