Ensemble Graph Neural Spatial Clustering: A Robust Framework for Spatial Domain Discovery in Spatial Transcriptomics

Shuyan Guo*, Yiting Bu†
*Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, USA
Email: sg4600@cumc.columbia.edu
†Department of Neurobiology and Behavior, University of California, Irvine, USA
Email: yitinb1@uci.edu

***Abstract*—Spatial transcriptomics provides a powerful approach for investigating cellular organization and tissue heterogeneity by integrating gene expression profiles with spatial information. A fundamental step in spatial transcriptomics analysis is *spatial domain discovery* (clustering), which aims to delineate distinct tissue regions based on molecular and spatial features. Although various spatial domain discovery methods have been proposed, existing approaches often struggle to produce stable and reproducible results across different datasets. This instability primarily arises from high levels of noise, stochastic behavior in model training, and limited generalizability. To address these challenges, we propose a robust ensemble graph-neural framework that integrates multiple spatial graph neural networks to achieve consistent and biologically interpretable tissue segmentation. Specifically, our approach adopts a *feature bagging* and *graph perturbation* strategy to capture diverse spatial representations. An ensemble mechanism is then employed, incorporating *Hungarian label alignment* and *spectral co-association consensus clustering* to fuse the results and derive the final spatial domains. Experimental studies on human dorsolateral prefrontal cortex (DLPFC) sections and the Human Breast Cancer (HBC) dataset demonstrate that our ensemble framework achieves superior clustering accuracy and enhanced stability compared with several representative spatial domain discovery methods.**

***Index Terms*—spatial transcriptomics, ensemble learning, graph neural networks, domain segmentation.**

I. Introduction

The integration of spatial information into gene expression profiling has revolutionized our ability to investigate tissue organization and cellular communication [1]–[5]. Spatial transcriptomics technologies, such as the *10x Genomics Visium* platform, enable the quantification of mRNA abundance at defined spatial coordinates, thereby establishing a direct link between molecular profiles and histological context. These technologies have become indispensable tools for reconstructing cell–cell interactions and delineating the functional architecture of tissues. A fundamental step in spatial transcriptomics analysis is *spatial domain discovery* (clustering), which aims to identify distinct tissue regions based on both molecular and spatial features [6]–[10].

For the spatial domain discovery task, various approaches have been proposed [11]–[13]. Among them, graph neural network (GNN)-based models [14]–[16] have emerged as particularly effective, as they jointly model gene expression similarity and spatial adjacency to capture the underlying spatial structure of tissues [17]–[20]. For example, GraphST is a graph self-supervised contrastive learning method that integrates gene expression profiles with spatial information in spatial transcriptomics data to enable spatially informed clustering for identifying tissue domains, batch integration to handle multiple tissue sections, and cell-type deconvolution by projecting single-cell RNA-seq data onto spatial spots [30]. stMMR is a multimodal geometric deep learning method that integrates gene expression, spatial location, and histological information from spatially resolved transcriptomics data to accurately identify spatial domains, using graph convolutional networks, self-attention, and similarity contrastive learning to address challenges in multimodal analysis [27]. SpatialMGCN employs a multi-view graph convolutional network with an attention mechanism to identify spatial domains in spatial transcriptomics data, by constructing separate neighbor graphs from gene expression and spatial information, extracting unique and shared embeddings, and reconstructing the expression matrix using a zero-inflated negative binomial decoder while incorporating spatial regularization to preserve neighbor information [28]. STAIG integrates gene expression, spatial coordinates, and histological images from spatial transcriptomics data using graph contrastive learning with dynamic graph augmentation and neighbor-based contrastive loss to enable accurate spatial domain identification and alignment-free batch integration without manual alignment or preprocessing [29].

Despite their promise, these methods remain sensitive to random initialization, hyperparameter choices, and optimization instability inherent in non-convex training objectives. A single model instance may capture only a subset of spatial patterns, leading to unstable predictions and poor reproducibility. Furthermore, complex architectures with many learnable parameters can overfit to technical noise or local spatial artifacts, reducing generalization to tissues with distinct morphology or expression landscapes [21]–[26].

To overcome these limitations, we propose an **ensemble graph neural spatial clustering framework** for robust spatial

domain discovery. Specifically, our approach adopts *feature bagging* and *graph perturbation* strategies to capture diverse spatial representations, while the proposed ensemble framework integrates multiple independently trained graph-based models—each initialized with distinct random seeds, feature subsets, and graph perturbations—to extract complementary spatial information and enhance the robustness of spatial domain identification. In the ensemble fusion stage, the outputs of individual models are aligned using the *Hungarian matching* algorithm and subsequently fused through *spectral co-association clustering* to generate a consensus spatial segmentation. As result, our method performs structured graph subtraction to isolate the most informative spatial relationships while suppressing redundancy. Collectively, these strategies enable consistent and biologically interpretable tissue segmentation across diverse datasets.

The main contributions of this work can be summarized as:

- We propose an ensemble graph neural network approach that integrates multiple independently trained models to capture complementary spatial representations, enhancing clustering stability and reproducibility.
- We develop a simplified ensemble strategy that performs structured graph subtraction to isolate informative spatial relationships and suppress redundancy, improving interpretability and regularization.
- Experimental studies on human dorsolateral prefrontal cortex (DLPFC) sections and the Human Breast Cancer (HBC) dataset show superior clustering accuracy, robustness, and biological relevance compared with several representative spatial domain discovery methods.

II. Methods

In this section, we provide detailed descriptions of the proposed model.

A. Data Preprocessing

In our model, spatial transcriptomics data preprocessing involved several key steps. First, quality control was performed to remove low-quality spots and lowly expressed genes. Next, total-count normalization followed by logarithmic transformation was applied to standardize expression levels across spots. The top ∼2,000 highly variable genes (HVGs) were then selected for downstream modeling to capture the most informative transcriptional signals. Thus,

For each tissue section, four adjacency graphs were constructed: (1) `fadj`, a functional adjacency graph based on gene expression similarity; (2) `sadj`, a spatial adjacency graph defined by physical proximity of spots; (3) `graph_pos`, a positive k-nearest neighbor graph capturing local connectivity; and (4) `graph_neg`, a randomized negative control graph used for contrastive analysis and robustness evaluation.

B. Base Graph Neural Clustering Model

In the proposed ensemble spatial clustering framework, a key component lies in the design of the base model, which is responsible for jointly modeling gene expression profiles and spatial neighborhood information. To achieve this, we employ a multi-graph neural network that integrates both molecular features and spatial topology to learn biologically meaningful tissue domains across cortical layers. This design enables the model to capture complex spatial dependencies while preserving transcriptional heterogeneity within local microenvironments.

In detail, given the constructed adjacency graphs `sadj` and `fadj`, the base learner in our ensemble framework is designed to jointly model functional and spatial relationships through a multi-graph neural network. Specifically, for each spot i, we denote its feature vector as $\mathbf{x}_i \in \mathbb{R}^d$, representing the expression. The model learns two latent embeddings: a functional embedding $\mathbf{H}_f$ from `fadj`, and a spatial embedding $\mathbf{H}_s$ from `sadj`. Each embedding is obtained via graph convolutional propagation as follows:

$$\mathbf{H}_f = \sigma\left(\tilde{\mathbf{A}}_f \mathbf{X} \mathbf{W}_f\right),$$

and

$$\mathbf{H}_s = \sigma\left(\tilde{\mathbf{A}}_s \mathbf{X} \mathbf{W}_s\right).$$

where $\tilde{\mathbf{A}}_f$ and $\tilde{\mathbf{A}}_s$ denote the normalized adjacency matrices for `fadj` and `sadj`, respectively; $\mathbf{W}_f$ and $\mathbf{W}_s$ are trainable weight matrices; and $\sigma(\cdot)$ represents a nonlinear activation function (e.g., ReLU).

The resulting embeddings $\mathbf{H}_f$ and $\mathbf{H}_s$ capture complementary information—functional similarity and spatial context—which are then fused to obtain a unified representation for spatial domain discovery:

$$\mathbf{H} = \alpha\, \mathbf{H}_f + (1 - \alpha)\, \mathbf{H}_s,$$

where $\alpha \in [0, 1]$ controls the relative contribution of functional and spatial information.

In addition, spatially neighboring spots should exhibit similar representations, while distant or non-neighboring spots should be dissimilar. To enforce this spatial smoothness and discriminative structure, we introduce a spatial regularization constraint loss that leverages both the positive and negative graphs: `graph_pos` and `graph_neg`. Formally, given the learned embedding matrix $\mathbf{H}$, the spatial constraint is defined as:

$$\mathcal{L}_{\text{reg}} = -\sum_{i=1}^{N_{\text{spot}}} \left(\sum_{j \in \texttt{graph\_pos}_i} \log(\sigma(S_{ij})) + \sum_{k \notin \texttt{graph\_neg}_i} \log(1 - \sigma(S_{ik})) \right). \tag{1}$$

where S denotes the cosine similarity matrix based on the learned latent representations $\mathbf{H}$.

Furthermore, to more accurately model the distribution of spatial gene expression data, we adopt a **Zero-Inflated Negative Binomial (ZINB)** loss function. This loss explicitly accounts for both the over-dispersion of RNA-seq counts and the excess of zero-valued entries caused by technical dropout

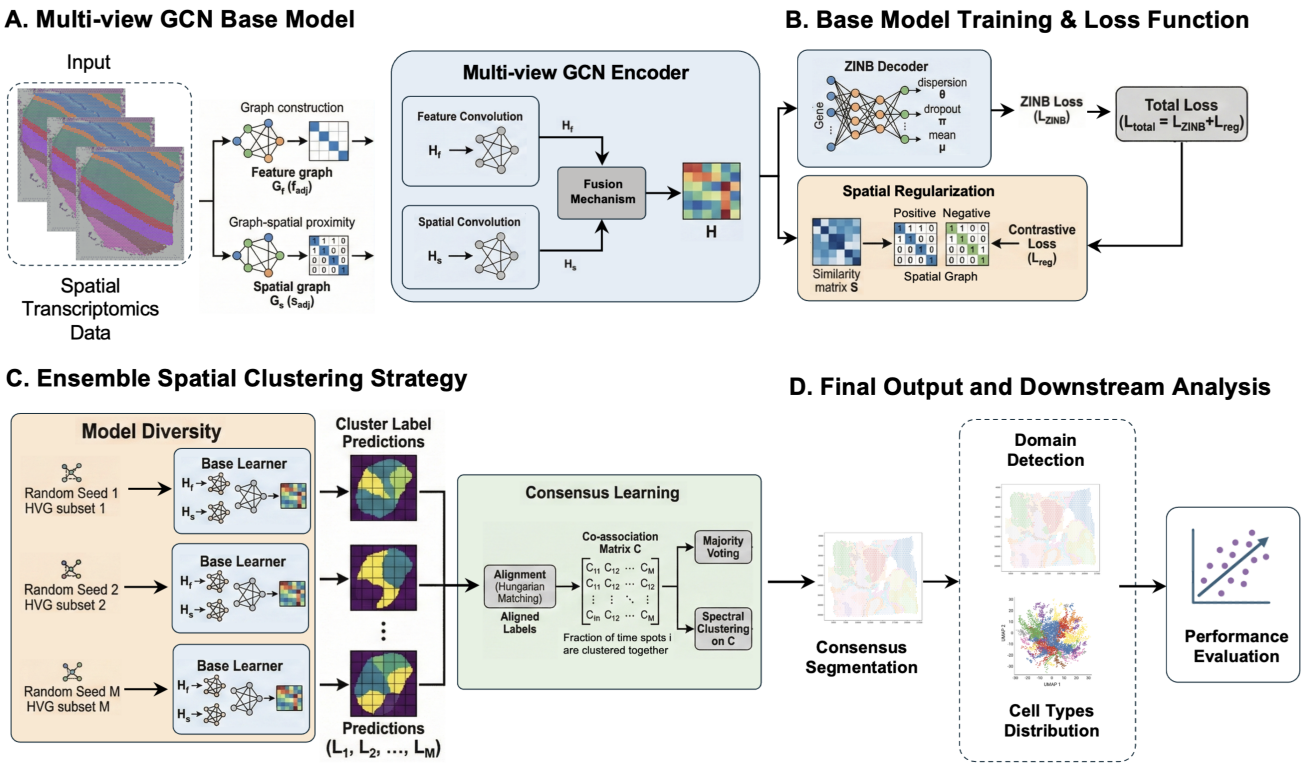

Fig. 1. Overview of the proposed ensemble graph neural network framework for spatial transcriptomics analysis. **A.** Base graph neural model: input data, construction of feature and spatial graphs, and a graph neural encoder that produces latent representations $\mathbf{H}$ via feature and spatial convolutions followed by a fusion module. **B.** Base-model training and loss functions: a ZINB decoder reconstructs the observed counts (loss $\mathcal{L}_{\text{ZINB}}$), and a contrastive spatial regularization term ($\mathcal{L}_{\text{cont}}$), derived from the similarity matrix and spatial graph, together define the total loss $\mathcal{L}_{\text{total}}$. **C.** Ensemble spatial clustering strategy: multiple base learners, initialized with different random seeds and highly variable gene (HVG) subsets, generate diverse cluster label predictions. Consensus learning performs alignment matching, constructs the co-association matrix $\mathbf{C}$, and applies majority voting and spectral clustering to obtain a robust consensus segmentation. **D.** Final output and downstream analyses: the consensus segmentation is used for spatial domain detection, visualization of cell-type distributions, and performance evaluation.

events. More specifically, for a given observed count $\mathbf{x}_{ij}$ (gene j in spot i), the likelihood is modeled as:

$$p(\mathbf{x}_{ij} \mid \mu_{ij}, \theta_j, \pi_{ij}) = \pi_{ij}\, \delta_0(\mathbf{x}_{ij}) + (1-\pi_{ij})\, \mathrm{NB}(\mathbf{x}_{ij};\, \mu_{ij}, \theta_j), \tag{2}$$

where μ_{ij} is the mean expression, θ_j is the gene-specific dispersion parameter, π_{ij} is the dropout probability, and $\delta_0(\cdot)$ denotes the point mass at zero. The corresponding **negative log-likelihood loss** used for training is:

$$\mathcal{L}_{\text{ZINB}} = -\sum_{i,j} \log\, p(x_{ij} \mid \mu_{ij}, \theta_j, \pi_{ij}). \tag{3}$$

Therefore, the total loss for the base graph neural network can be defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Reg}} + \mathcal{L}_{\text{ZINB}} \tag{4}$$

This joint formulation enables the network to capture both spatial dependencies and statistical characteristics of expression profiles.

C. Ensemble Spatial Clustering

The proposed ensemble spatial clustering framework comprises two key components: model diversity and consensus learning.

Model Diversity: Each model takes as input the normalized gene expression matrix and four adjacency graphs `fadj`, `sadj`, `graph_pos` and `graph_neg`, which jointly encode both functional similarity and spatial neighborhood structure among tissue spots. To enhance robustness and mitigate bias from individual model initialization, we introduce a sampling-based strategy to generate diverse base learners. Specifically, multiple base models are trained using different random seeds (`3, 5, 7, 9, 11`) to capture distinct local minima of the non-convex optimization landscape. To further improve generalization capability, a feature bagging strategy is employed, in which approximately 70% of highly variable genes (HVGs) are randomly sampled for each run. This approach prevents over-reliance on specific gene subsets and promotes complementary feature representations across base models. Moreover, we introduce a graph perturbation technique, termed neighbor jitter, in which adjacency matrices are randomly smoothed or

sharpened during training. This perturbation enables the model to perceive spatial boundaries at multiple resolutions, thereby encouraging the extraction of spatially coherent domains under varying local connectivity patterns.

Consensus Learning: To achieve stable and reproducible spatial domain segmentation, we develop a consensus learning strategy that integrates all independently trained base models into a unified ensemble. This approach captures complementary spatial features across multiple runs, mitigates stochastic variance arising from random initialization, and establishes a consistent framework for downstream spatial domain discovery. Specifically, cluster label predictions from individual models must first be aligned to resolve arbitrary label permutations. To this end, we employ the *Hungarian matching algorithm*, which performs pairwise label alignment by maximizing overlap between clusters, thereby ensuring consistent mapping across different seeds and model instances. Given two sets of cluster assignments, the algorithm finds an optimal one-to-one mapping between clusters that maximizes total overlap. Formally, let L_{ref} and L_{target} denote cluster assignments of n spots from a reference and a target model, respectively. The optimal permutation matrix $\mathbf{P}^*$ is obtained by solving the following:

$$\arg\max_{\mathbf{P}\in\Pi} \sum_{k=1}^{K}\sum_{l=1}^{K} P_{kl}\left|\{i \mid L_{\text{ref}}(i)=k\}\cap\{i \mid L_{\text{target}}(i)=l\}\right|, \tag{5}$$

where Π denotes all possible permutation matrices and $|\cdot|$ is set cardinality. The target labels are then relabeled according to $\mathbf{P}^*$ to ensure consistent cluster mapping across models.

After alignment, we computed a co-association matrix $C \in \mathbb{R}^{N\times N}$ that captures pairwise agreement across the ensemble. Each entry C_{ij} quantifies how often spots i and j were assigned to the same cluster among the M aligned predictions:

$$C_{ij} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}[L_m(i) = L_m(j)], \tag{6}$$

where N is the number of spatial spots and $\mathbb{I}$ is the indicator function. This matrix serves as a robust, model-averaged estimate of spatial co-membership, smoothing over stochastic variation across individual runs.

To obtain the final ensemble segmentation, we adopted a two-step consensus procedure. First, we applied simple majority voting to generate an initial per-spot consensus label based on the aligned predictions. Second, we performed spectral clustering directly on the co-association matrix C (with the number of clusters fixed to that of the reference model) to refine the global spatial structure and produce an alternative consensus partition. We then select consensus solutions as the final ensemble output.

The overview of the proposed model can be found in Figure 1 and the related optimization processing can be found in Algorithm 1.

Algorithm 1 The optimization of the ensemble graph neural spatial clustering framework

Input: Normalized spatial transcriptomics expression matrix $\mathbf{X} \in \mathbb{R}^{N\times G}$, spatial coordinates $\{\mathbf{s}_i\}_{i=1}^{N}$, number of clusters K, number of base learners M.

Output: Consensus spatial domain labels $\hat{\mathbf{y}} \in \{1,\ldots,K\}^N$.

1: **Preprocessing:** Normalize and log-transform $\mathbf{X}$; select highly variable genes (HVGs).

2: **Graph construction:** Using $\mathbf{X}$ and spatial coordinates, construct
- functional adjacency graph fad_j (gene expression similarity),
- spatial adjacency graph sad_j (physical proximity),
- positive graph $graph\_pos$ (local k-NN connectivity),
- negative graph $graph\_neg$ (randomized negative control).

3: **Base model learning:**
for $m = 1$ to M
- Sample a subset of HVGs $G_m \subseteq G_{\text{HVG}}$ (e.g., 70% of HVGs).
- Initialize base graph neural network with random seed.
- Optionally perturb adjacency graphs via neighbor jitter to obtain perturbed graphs.
- Compute feature embedding $\mathbf{H}_f$ from $(\mathbf{X}, \mathbf{A}_f)$ and Compute spatial embedding $\mathbf{H}_s$ from $(\mathbf{X}, \mathbf{A}_s)$.
- Fuse embeddings to obtain $\mathbf{H}$.

4: **Consensus learning:**
- **Label alignment:** Choose one run r as reference. For each $m \neq r$, align $L^{(m)}$ to $L^{(r)}$ using the Hungarian algorithm to resolve label permutations, yielding aligned labels $L^{(m)}$.
- **Co-association matrix construction:** Build $C \in [0,1]^{N\times N}$ where C_{ij} is the fraction of base learners assigning spots i and j to the same cluster across $\{\tilde{\mathbf{L}}^{(m)}\}_{m=1}^{M}$.
- **Consensus segmentation:** (a) Obtain per-spot majority-vote labels $\hat{\mathbf{y}}^{\text{mv}}$ from $\{L^{(m)}\}_{m=1}^{M}$. (b) Apply spectral clustering with K clusters to C to obtain refined labels $\hat{\mathbf{y}}^{\text{sc}}$.
- **Clustering:** Select the final consensus labels $\hat{\mathbf{y}}$ as the ensemble spatial domain segmentation.

III. Experiment

In this section, we evaluate the effectiveness of the proposed framework on multiple spatial transcriptomic datasets. All experiments were conducted following a standardized protocol to ensure reproducibility across models and datasets. We first describe our experimental setup from three perspectives: datasets, evaluation metrics, and competing methods. We then conduct a comprehensive evaluation of the proposed framework by addressing the following key questions:

- **Q1.Superiority**: Does the proposed model outperform

state-of-the-art methods in spatial transcriptomics clustering?

- **Q2. Effectiveness**: Are the proposed graph enhancement strategies effective in improving clustering performance?
- **Q3. Spatial domain identification**: Can the proposed framework accurately and robustly identify distinct spatial domains?

A. Experiment Setting

1) Datasets: We evaluate the proposed framework on two spatial transcriptomics datasets:

- **DLPFC**: Eleven human dorsolateral prefrontal cortex (DLPFC) Visium sections (IDs 151508–151676) with known cortical layer annotations (L1–L6 and white matter). Including an additional reference section, this results in a total of 12 datasets for analysis [31].
- **HBC**: A Human Breast Cancer (HBC) dataset containing both tumor and stromal regions [30].

Detailed statistics of the datasets used in this study are summarized in Table I.

TABLE I
STATISTICS OF THE DATASETS.

Tissues	Slices	# Spots	# Genes
Dorsolateral pre-frontal cortex (DLPFC)	151507	4226	20494
	151508	4384	20083
	151509	4789	20732
	151510	4634	20475
	151669	3661	20583
	151670	3498	20338
	151671	4110	21037
	151672	4015	10725
	151673	3639	21267
	151674	3673	21897
	151675	3592	20783
	151675	3460	20806
Breast cancer	Section_1	3798	36601

2) Comparison Methods: We compared the proposed framework against several representative spatial clustering models: 1. SCANPY [35], 2. Spatial-MGCN [28], 3. SpaGCN [32], 4 GraphST [30], 5. StMMR [27], 6. DeepST [33] and 7. stLearn [34].

All models were trained for 100–150 epochs using the Adam optimizer (learning rate 0.0005, batch size 128) on GPU clusters to ensure reproducibility. All experiments were repeated five times with independent random seeds (`3, 5, 7, 9, 11`). Metric means and standard deviations were reported in the Results section.

3) Metrics: Performance was evaluated using two widely adopted metrics: Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI). These metrics collectively provide a comprehensive assessment of both clustering accuracy and structural consistency. Higher values of NMI and ARI indicate stronger agreement between the predicted clustering results and the ground-truth annotations, reflecting more accurate and reliable spatial domain identification.

B. Results and Analysis (*Q1.Superiority*)

Tables II and III summarize the clustering performance of our method in comparison with several state-of-the-art spatial transcriptomics clustering baselines on two benchmark datasets: the 12-section DLPFC dataset and the HBC Visium dataset. For the DLPFC dataset, GraphST exhibits the weakest performance, with consistently low NMI and ARI values across all sections. SpaGCN performs moderately but shows substantial variability, particularly in sections with more complex spatial architectures. Spatial-MGCN provides a strong baseline; however, our model surpasses it in seven out of the nine DLPFC sections, with the largest gains appearing in regions of high structural heterogeneity. Overall, our method consistently achieves the highest NMI and ARI across the majority of tissue sections. In particular, it outperforms GraphST and SpaGCN by substantial margins, frequently exceeding them by more than $+0.20$ in ARI. Relative to Spatial-MGCN, our model demonstrates notable improvements in structurally complex tissues such as sections 151508, 151669, and 151671, with ARI gains of $+0.26$, $+0.30$, and $+0.18$, respectively. We further evaluate our method on the Human Breast Cancer (HBC) Visium dataset, which contains heterogeneous tumor and stromal regions. Under the same preprocessing pipeline and identical hyperparameter settings, our model again achieves the best overall performance, surpassing GraphST, SpaGCN, and Spatial-MGCN on all evaluation metrics.

Overall, these results highlight the robustness and superior generalization capability of our approach, particularly in complex and spatially heterogeneous transcriptomics scenarios.

C. Ablation Study (*Q2. Effectiveness*)

To evaluate the efficacy of our proposed graph enhancement strategies in improving clustering performance, we compared our ensemble model against the baseline graph neural network clustering framework on the 12-section DLPFC dataset and the HBC dataset. The results are presented in Tables IV and V.

From these results, we observe that incorporating the graph enhancement components and the ensemble strategy generally leads to improved or at least comparable performance across most sections and evaluation metrics, indicating that the proposed design effectively strengthens spatial domain detection while maintaining robustness. For example, improvements of $+0.11$ (151508), $+0.09$ (151509), and $+0.10$ (151670) demonstrate that the enhanced graph mechanism strengthens the model's ability to capture more discriminative spatial structures.

Overall, the ensemble model consistently yields more accurate and robust clustering outcomes, validating the effectiveness of the proposed ensemble learning mechanisms.

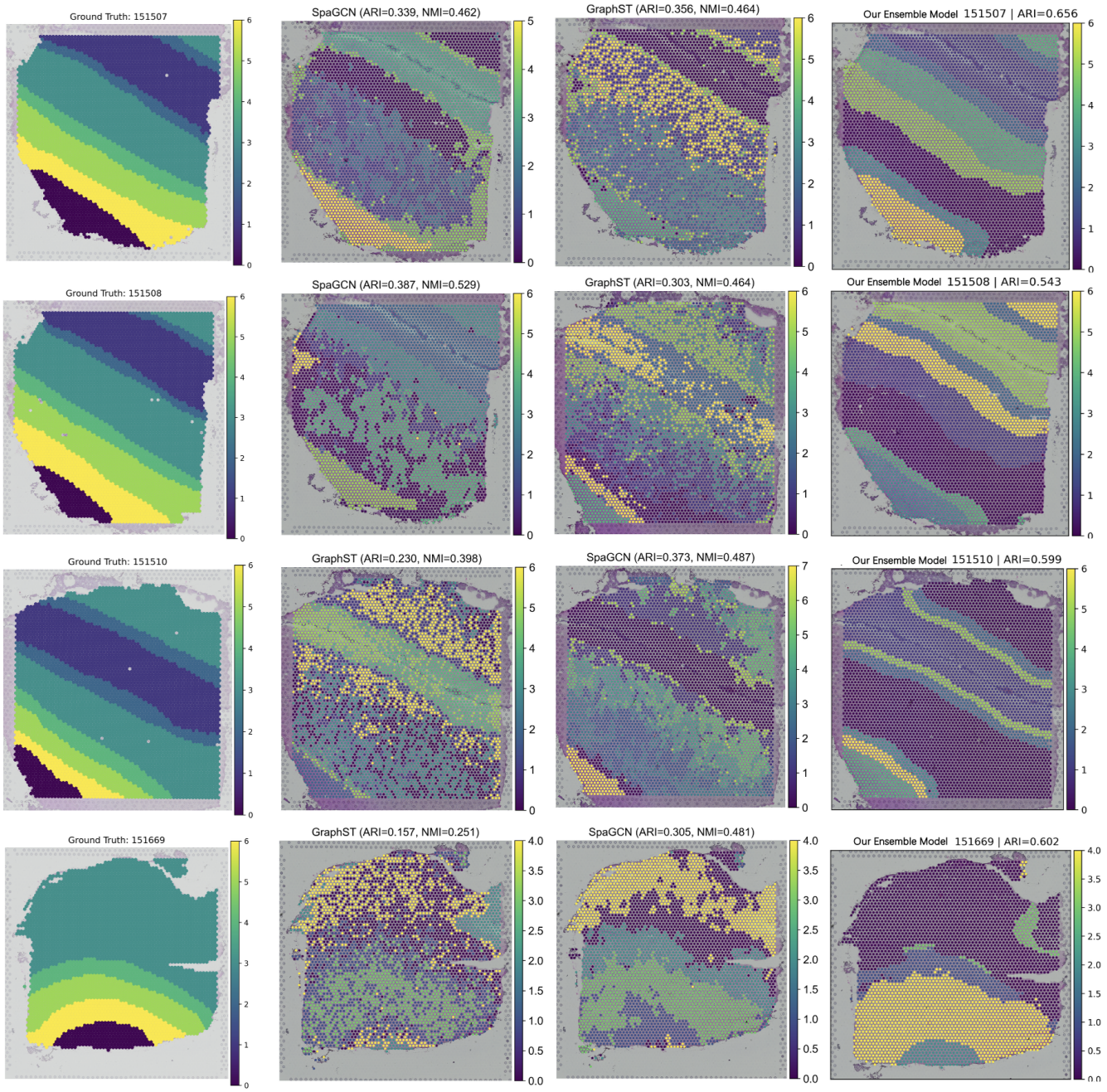

Fig. 2. **Comparative spatial domain segmentation for a DLPFC sample.** Shown are the ground-truth cortical layer annotations (left) alongside the results of four representative methods: GraphST, SpaGCN, Spatial-MGCN, and our model. Among all approaches, our method produces the most coherent laminar organization with minimal boundary noise, achieving the closest correspondence to the biological ground truth.

TABLE II
CLUSTERING PERFORMANCE OF COMPETING SPATIAL TRANSCRIPTOMICS MODELS IN TERMS OF ARI. BOLD ENTRIES INDICATE THE BEST RESULTS.

Adjusted Rand Index (ARI)									
Method	151507	151508	151509	151510	151669	151670	151671	151672	HBC
SCANPY	0.20	0.15	0.19	0.14	0.10	0.09	0.12	0.12	0.49
Spatial-MGCN	0.63	0.46	0.54	0.51	0.39	0.35	0.60	0.77	0.64
SpaGCN	0.43	0.33	0.41	0.37	0.23	0.35	0.51	0.53	0.56
GraphST	0.48	0.39	0.52	0.50	0.48	0.46	0.61	0.63	0.54
stMMR	0.59	0.51	0.58	0.69	0.49	0.48	0.68	0.63	0.62
DeepST	0.55	0.42	0.43	0.50	0.44	0.33	0.52	0.48	0.53
stLearn	0.49	0.31	0.45	0.44	0.32	0.23	0.39	0.34	0.55
Our Model	**0.65**	**0.55**	**0.62**	**0.67**	**0.52**	**0.40**	**0.77**	**0.76**	**0.61**

TABLE III
CLUSTERING PERFORMANCE OF ALL SPATIAL TRANSCRIPTOMIC CLUSTERING METHODS IN TERMS OF NMI. BOLD ENTRIES INDICATE THE BEST RESULTS.

Normalized Mutual Information (NMI)									
Method	151507	151508	151509	151510	151669	151670	151671	151672	HBC
SCANPY	0.21	0.21	0.27	0.22	0.16	0.16	0.24	0.23	0.52
Spatial-MGCN	0.74	0.60	0.68	0.67	0.58	0.56	0.72	0.75	0.69
SpaGCN	0.54	0.42	0.55	0.50	0.42	0.45	0.60	0.61	0.56
GraphST	0.64	0.54	0.64	0.64	0.59	0.68	0.70	0.61	0.67
stMMR	0.72	0.65	0.71	0.71	0.56	0.56	0.72	0.72	0.65
DeepST	0.62	0.57	0.62	0.62	0.57	0.51	0.59	0.60	0.68
stLearn	0.64	0.53	0.62	0.59	0.49	0.41	0.54	0.47	0.63
Our Model	**0.76**	**0.66**	**0.71**	**0.68**	**0.64**	**0.57**	**0.73**	**0.73**	**0.65**

TABLE IV
CLUSTERING PERFORMANCE OF THE PROPOSED ENSEMBLE MODEL AND BASE MODEL IN TERMS OF ARI. BOLD ENTRIES INDICATE THE BEST RESULTS.

Adjusted Rand Index (ARI)									
Method	151507	151508	151509	151510	151669	151670	151671	151672	HBC
Base	0.64	0.41	0.53	0.47	0.39	0.31	0.61	0.77	0.59
Our Model	**0.65**	**0.55**	**0.62**	**0.67**	**0.52**	**0.40**	**0.77**	**0.76**	**0.61**

TABLE V
CLUSTERING PERFORMANCE OF THE PROPOSED ENSEMBLE MODEL AND BASE MODEL IN TERMS OF NMI. BOLD ENTRIES INDICATE THE BEST RESULTS.

Normalized Mutual Information (NMI)									
Method	151507	151508	151509	151510	151669	151670	151671	151672	HBC
Base	0.74	0.59	0.68	0.64	0.50	0.44	0.72	0.66	0.67
Our Model	**0.76**	**0.66**	**0.71**	**0.68**	**0.64**	**0.57**	**0.73**	**0.73**	**0.65**

*D. Visualization of spatial domain detection (**Q3. Spatial domain identification**)*

To evaluate the effectiveness of the proposed framework in accurately and robustly identifying distinct spatial domains, we visualize the spatial clustering results of our method alongside two competing approaches (SpaGCN and GraphST) in Figure 2. Spatial overlay maps were generated to examine the correspondence between predicted clusters and histological structures, with each domain color-coded on the Visium tissue grid to assess boundary sharpness, layer continuity, and overall structural coherence within both cortical and tumor-associated tissues.

From the visual comparison, it is evident that our model more accurately recovers cortical layer boundaries than the competing methods. SpaGCN and GraphST often produce fragmented or noisy segmentations, particularly in regions with subtle transitions or heterogeneous expression patterns. In contrast, our ensemble model yields smoother and more coherent laminar structures, with clearly delineated interfaces between adjacent layers and substantially reduced boundary noise. This improvement is especially pronounced in samples with complex spatial organization (e.g., 151508 and 151510), where competing methods fail to preserve layer continuity. These qualitative observations further reinforce the quantita-

tive gains reported in Tables II and III, demonstrating the enhanced ability of our framework to capture biologically meaningful spatial patterns.

IV. Conclusion

In this work, we proposed an ensemble graph neural network framework for spatial transcriptomics analysis that jointly exploits gene expression similarity and spatial neighborhood information. By constructing complementary feature and spatial graphs, encoding spots with a graph neural encoder, and decoding gene expression with a ZINB-based likelihood, our method effectively captures both the statistical properties of count data and the spatial organization of tissue. A contrastive spatial regularization term further encourages locally coherent representations, while an ensemble clustering strategy with diverse base learners and consensus learning yields robust spatial domain segmentation. Extensive experiments on the 12-section DLPFC dataset and the HBC dataset demonstrate that our framework consistently achieves superior performance compared with existing graph-based and deep clustering baselines.

References

[1] Williams, Cameron G., et al. "An introduction to spatial transcriptomics for biomedical research." Genome medicine 14.1 (2022): 68.

[2] Moses, Lambda, and Lior Pachter. "Museum of spatial transcriptomics." Nature methods 19.5 (2022): 534-546.

[3] Rao, Anjali, et al. "Exploring tissue architecture using spatial transcriptomics." Nature 596.7871 (2021): 211-220.

[4] Anderson, Ana C., et al. "Spatial transcriptomics." Cancer cell 40.9 (2022): 895-900.

[5] Du, Jun, et al. "Advances in spatial transcriptomics and related data analysis strategies." Journal of translational medicine 21.1 (2023): 330.

[6] Sun, Yidi, et al. "A comprehensive survey of dimensionality reduction and clustering methods for single-cell and spatial transcriptomics data." Briefings in Functional Genomics 23.6 (2024): 733-744.

[7] Teng, Haotian, Ye Yuan, and Ziv Bar-Joseph. "Clustering spatial transcriptomics data." Bioinformatics 38.4 (2022): 997-1004.

[8] Li, Jiachen, et al. "Cell clustering for spatial transcriptomics data with graph neural networks." Nature Computational Science 2.6 (2022): 399-408.

[9] Liu, Wei, et al. "Joint dimension reduction and clustering analysis of single-cell RNA-seq and spatial transcriptomics data." Nucleic acids research 50.12 (2022): e72-e72.

[10] Wei, Xindian, Tianyi Chen, Xibiao Wang, Wenjun Shen, Cheng Liu, Si Wu, and Hau-San Wong. "COME: contrastive mapping learning for spatial reconstruction of single-cell RNA sequencing data." Bioinformatics 41, no. 3 (2025): btaf083.

[11] Li, Jia, Yu Shyr, and Qi Liu. "aKNNO: single-cell and spatial transcriptomics clustering with an optimized adaptive k-nearest neighbor graph." Genome Biology 25.1 (2024): 203.

[12] Zhang, Hang, et al. "Kernel-bounded clustering for spatial transcriptomics enables scalable discovery of complex spatial domains." Genome Research 35.2 (2025): 355-367.

[13] Li, Zheng, and Xiang Zhou. "BASS: multi-scale and multi-sample analysis enables accurate cell type clustering and spatial domain detection in spatial transcriptomic studies." Genome biology 23.1 (2022): 168.

[14] Zhou, Jie, et al. "Graph neural networks: A review of methods and applications." AI open 1 (2020): 57-81.

[15] Corso, Gabriele, et al. "Graph neural networks." Nature Reviews Methods Primers 4.1 (2024): 17.

[16] Scarselli, Franco, et al. "The graph neural network model." IEEE transactions on neural networks 20.1 (2008): 61-80.

[17] Liu, Teng, et al. "Assembling spatial clustering framework for heterogeneous spatial transcriptomics data with GRAPHDeep." Bioinformatics 40.1 (2024): btae023.

[18] Zhang, Fangqin, et al. "SpaInGNN: Enhanced clustering and integration of spatial transcriptomics based on refined graph neural networks." Methods 233 (2025): 42-51.

[19] Lampert, Moritz, et al. "Cell-Type Prediction in Spatial Transcriptomics Data using Graph Neural Networks." ICLR 2024 Workshop on Machine Learning for Genomics Explorations.

[20] He, Xiao, et al. "Heterogeneous graph guided contrastive learning for spatially resolved transcriptomics data." Proceedings of the 32nd ACM International Conference on Multimedia. 2024.

[21] Hu, J., Schroeder, A., Coleman, K., Chen, C., Auerbach, B. J., and Li, M. (2021). Statistical and machine learning methods for spatially resolved transcriptomics with histology. Computational and structural biotechnology journal, 19, 3829-3841.

[22] Xu, H., Wang, S., Fang, M., Luo, S., Chen, C., Wan, S., ... and Qu, K. (2023). SPACEL: deep learning-based characterization of spatial transcriptome architectures. Nature Communications, 14(1), 7603.

[23] Xue, L., Liang, X., Wang, B., Liu, W., Zou, Z., Xiao, Q., ... and Luo, J. (2025, July). Identifying Spatial Domains by Fusing Spatial Transcriptomics and Histological Images Through Contrastive Learning. In International Conference on Intelligent Computing (pp. 3-14). Singapore: Springer Nature Singapore.

[24] Luo, J., Fu, J., Lu, Z., and Tu, J. (2025). Deep learning in integrating spatial transcriptomics with other modalities. Briefings in bioinformatics, 26(1), bbae719.

[25] 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.

[26] Liu, T., Fang, Z. Y., Li, X., Zhang, L. N., Cao, D. S., and Yin, M. Z. (2023). Graph deep learning enabled spatial domains identification for spatial transcriptomics. Briefings in Bioinformatics, 24(3), bbad146.

[27] Zhang, Daoliang, et al. "stMMR: accurate and robust spatial domain identification from spatially resolved transcriptomics with multimodal feature representation." GigaScience 13 (2024): giae089.

[28] Wang, Bo, et al. "Spatial-MGCN: a novel multi-view graph convolutional network for identifying spatial domains with attention mechanism." Briefings in Bioinformatics 24.5 (2023): bbad262.

[29] Yang, Yitao, et al. "STAIG: Spatial transcriptomics analysis via image-aided graph contrastive learning for domain exploration and alignment-free integration." Nature Communications 16.1 (2025): 1067.

[30] Long, Yahui, et al. "Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST." Nature Communications 14.1 (2023): 1155.

[31] Maynard, Kristen R., et al. "Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex." Nature neuroscience 24.3 (2021): 425-436.

[32] Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., ... 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.

[33] Xu, C., Jin, X., Wei, S., Wang, P., Luo, M., Xu, Z., ... Jiang, Q. (2022). DeepST: identifying spatial domains in spatial transcriptomics by deep learning. Nucleic Acids Research, 50(22), e131-e131.

[34] Pham, D., Tan, X., Xu, J., Grice, L. F., Lam, P. Y., Raghubar, A., ... Nguyen, Q. (2020). stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. biorxiv, 2020-05.

[35] Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis. "SCANPY: large-scale single-cell gene expression data analysis." Genome biology 19.1 (2018): 15.