Briefings in Bioinformatics, 2026, 27, bbag070
https://doi.org/10.1093/bib/bbag070
Published: 19 February 2026
Problem Solving Protocol

OXFORD

An unsupervised method for spatial transcriptomics analysis based on adversarial autoencoder

Wei Lan[1], Guohang He[1], Lingzhi Zhu[2,*], Ruiqing Zheng[3], Min Li[3], Yi Pan[4]

[1]Guangxi Key Laboratory of Multimedia Communications and Networks Technology, School of Computer, Electronic and Information, Guangxi University, No. 100 Daxue Road, Nanning, Guangxi 530004, China
[2]School of Computer and Information Science, Hunan Institute of Technology, No. 18 Henghua Road, Hengyang, Hunan 421002, China
[3]Hunan Provincial Key laboratory on Bioinformatics, School of Computer Science and Engineering, Central South University, No. 932 Lushan South Road, Changsha, Hunan 410083, China
[4]Faculty of Computer Science and Control Engineering, Shenzhen Institute of Advanced Technology, No. 1 Gongchang Road, Shenzhen, Guangdong 518055, China

*Corresponding author. E-mail: zlz@hnit.edu.cn

Abstract

Spatial transcriptomics (ST) offers unprecedented opportunities to decode the spatial organization of gene expression, yet the inherent noise and complexity of ST data pose substantial challenges for accurate analysis. Here, we present DACN, a unified framework that integrates an improved adversarial autoencoder (AAE) with a graph convolutional network (GCN) to robustly analyze ST data across varying resolutions and throughputs. DACN employs a hybrid encoder that couples multi-head attention with residual connections to capture fine-grained local expression patterns while retaining critical global information. The hybrid encoder and generator jointly construct the AAE module, which denoises expression profiles and learns stable latent representations. The GCN component further exploits spatial neighborhood relationships to refine these embeddings. Across multiple ST datasets with varying resolutions, DACN consistently outperforms existing methods in accuracy and robustness. All code and datasets are publicly available at https://github.com/lanbiolab/DACN.

Keywords spatial transcriptomics, adversarial autoencoder, deep learning, spatial domain

Introduction

The rapid advancement of single-cell omics has offered a novel perspective for bioinformatics research. Single-cell sequencing technologies provide RNA information at the single-cell level with varying sequencing depths, thereby providing critical technical support for biological analysis [1, 2]. For instance, single-cell RNA sequencing (scRNA-seq) captures gene expression profiles at the resolution of individual cells, allowing researchers to investigate the regulatory mechanisms across different cell types or tissues [3, 4]. However, the spatial location of cells within tissues often influences their functional intensity and interactions with other cells, which makes spatial information equally essential for understanding biological processes. The emergence of spatial transcriptomics (ST) has introduced a new approach to addressing this challenge [5].

In recent years, a range of ST technologies have been developed, including 10× Visium [6], Slide-seq [7], Slide-seqV2 [8], MERFISH [9], SeqFISH+ [10], and STARmap [11]. 10× Visium represents an early approach characterized by limited spatial resolution. This limitation has been addressed by technologies like Slide-seq and Slide-seqV2 that provide substantial improvements. Furthermore, methods such as MERFISH, SeqFISH+, and STARmap achieve spatial resolution at the subcellular scale. Spatial resolution which is defined by the physical size of each spot is a critical parameter of ST technologies. Higher-resolution technologies yield smaller spots that encompass fewer cells. As a highly regarded research area, ST not only provides gene expression information within the spatial context of tissues but also reveals functional relationships from a spatial perspective [12].

In the early stages, machine learning is commonly applied to the analysis of ST data. In instance, Giotto [13] utilizes a hidden Markov model (HMM) to estimate spatial similarity between nodes and performs spatial clustering to identify spatial domain. BayesSpace [14] adopts a Bayesian framework to cluster spatial spots with similar expression profiles, thereby improving the resolution of spatial domain. With the success of deep learning in various tasks, researchers have begun to apply deep learning to the analysis of ST data. SpaGCN [15] is a method to apply graph neural network

Received: September 8, 2025. **Revised:** December 24, 2025. **Accepted:** January 12, 2026

(GNN) for spatial domain identification. It uses graph convolutional network (GCN) to capture features between spatially adjacent nodes and performs clustering to identify spatial domains. SEDR [16] integrates an autoencoder with GCN to jointly model gene expression and spatial relationships. By jointly modeling the gene expression matrix and the spatial adjacency matrix, the method learns a latent embedding that captures both transcriptomic and spatial structural information. In addition, it introduces a clustering loss to achieve smoother spatial domain identification. STAGATE [17] is also a GNN-based method that employs an adaptive attention mechanism to dynamically assign weights to neighboring nodes during training, enhancing the consistency at spatial domain boundaries. SpaceFlow [18] introduces contrastive learning into ST data analysis. It constructs negative samples by randomly perturbing the structure of spatial graph and training the model to distinguish them from real graphs, thereby improving the performance of spatial domain identification. GraphST [19] further combines self-supervised contrastive learning with GCN to better exploit contextual information among spatial nodes. It allows GraphST to adaptively refine feature representations during training and enhance spatial domain identification. SiGra [20] is a method that performs single-cell spatial analysis through image-enhanced graph transformers, leveraging imaging features to identify spatial domains and enhance sparse and noisy transcriptomic data. Although these methods have achieved notable successes, several limitations also exist in these methods. The variability in sequencing depth and spatial resolution generates complex noise and leads to inconsistent gene detection across datasets, thereby compromising model stability and hindering robust spatial transcriptomic analysis. Consequently, these methods face challenges in accurately identifying distinct spatial domains, particularly in regions with sparse coverage or heterogeneous expression patterns. In addition, most approaches primarily rely on local adjacency information, limiting their ability to capture long-range spatial information.

To address these issues, we propose an unsupervised learning method (DACN) that integrates an improved adversarial autoencoder (AAE) with a GCN. The AAE component mitigates noise and dropout effects inherent in ST data, while a multi-head attention mechanism is incorporated to capture long-range spatial dependencies and simultaneously extract local feature representations for each spatial spot. This design enables DACN to effectively distinguish key regions within complex tissue domains by leveraging subtle local expression differences. A residual connection is further introduced to preserve essential spatial expression feature and prevent overfitting to local patterns. The GCN module constructs a spatial graph to represent neighborhood relationships among spots, and a spatial regularization term constrains domain boundaries, resulting in smoother and more coherent spatial clustering [21]. The learning embedding is further utilized for downstream tasks such as spatial domain identification and batch effect correction. To demonstrate the robustness and generalizability of DACN, we evaluate its spatial clustering performance across multiple ST datasets with varying spatial resolutions. The experimental results show that DACN achieves the most accurate and smooth spatial clustering results, outperforming other methods across multiple clustering metrics.

Method

Model architecture

DACN is composed of three key components: AAE module, GCN, and decoder. The architecture of DACN is shown in Fig. 1. The AAE module consists of hybrid encoder, discriminator layer, and generator. The AAE and GCN are designed to extract gene expression features and spatial features, respectively. The decoder is responsible for reconstructing the latent embeddings. Next, we provide a detailed description of each component.

Adversarial autoencoder

AAE are commonly used for denoising and reconstructing sparse images by enhancing feature reconstruction capabilities through adversarial training [22]. However, conventional AAE often fail to capture the contextual information of cells when applied to gene expression profiles [23]. To solve this issue, we design an improved AAE to mitigate the loss of highly expressed gene information. The AAE is primarily composed of a hybrid encoder and GAN. The primary function of the hybrid encoder is to map ST data into a low-dimensional latent space. The hybrid encoder consists of multi-head attention and fully connected layers. Similar designs have demonstrated strong capability in feature integration and hierarchical representation. For instance, DFT_ANPD employs a dual-feature bidirectional attention mechanism to effectively integrate heterogeneous information [24]. Moreover, methods that incorporate part–whole hierarchies into fully convolutional networks for scene parsing introduce hierarchical representations to capture structural relationships within complex objects, thereby enhancing semantic understanding [25]. The multi-head attention layer is employed to characterize complex tissue regions that often exhibit highly heterogeneous and disordered expression patterns. Each attention head captures distinct local gene expression relationships among spatial spots, enabling more accurate delineation of these intricate regions. In addition, a residual connection is integrated within the multi-head attention layer to preserve essential information from the original spatial gene expression while preventing the model from overfitting or developing an undue bias toward local feature representations. The hybrid encoder is defined as follows:

$$x_i' = \left(\mathbf{softmax}\left[\frac{(W^Q x_i)\cdot(W^K x_i)}{\sqrt{d_k}}\right]\right)\cdot(W^V x_i),$$
$$i \in \{1,2,3,\ldots,n\} \tag{1}$$

$$x' = \mathbf{Concat}\left\{x_1', x_2', x_3', \ldots, x_n'\right\} \tag{2}$$

$$z_h = FC((1-\alpha)x' + \alpha x) \tag{3}$$

where x represents the feature matrix of spatial spots, with a dimensionality of 200 obtained from the preprocessed gene expression matrix. n denotes the number of attention heads in the multi-head attention layer. The features matrix of each spatial spot is divided into n segments, each processed by a corresponding attention head for feature extraction. x_i denotes the feature subset processed by the ith attention head, with a dimensionality of 200/n. The three projection matrices, W^Q, W^K, and W^V are used to transform the input spatial representations into query, key, and value vectors, respectively. The d_k is used to scale the dot product between $W^Q x_i$ and $W^K x_i$, which helps to avoid excessively large result and mitigates the risk of gradient explosion. x_i' denotes the low-dimensional embedding of the spatial spots extracted by ith attention head. Concat refers to the concatenation function that aggregates the low-dimensional embedding of the spatial spots from all attention heads. The parameter α represents a weighting factor to control the fusion between the input and output

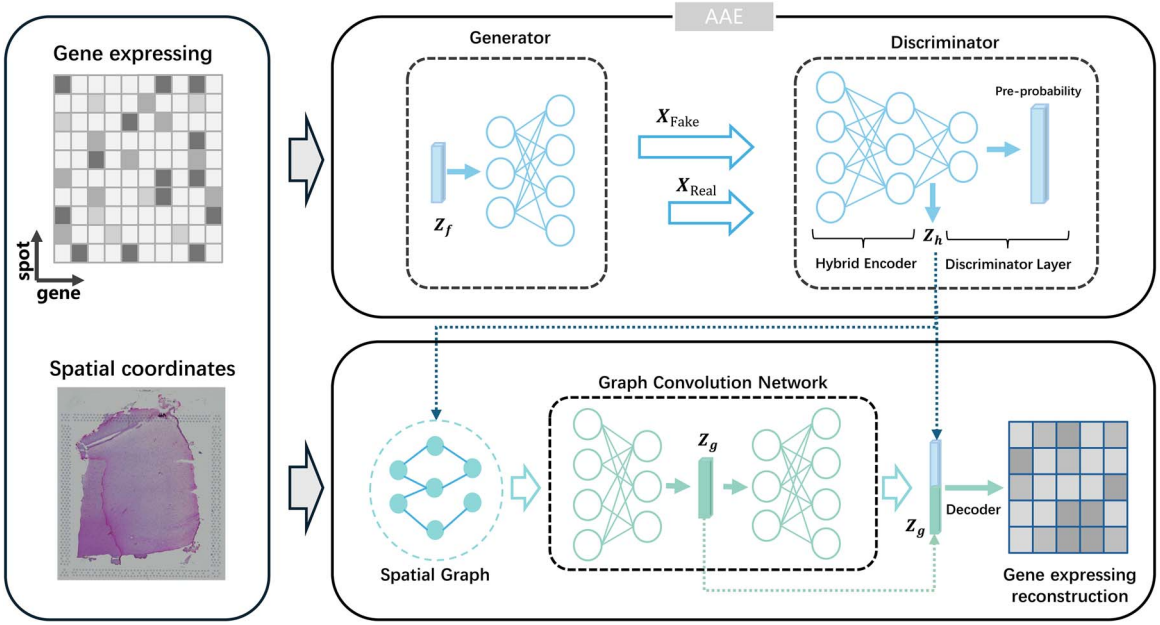

Figure 1 The workflow of DACN. DACN consists of three main components: an AAE, a GCN, and a decoder. The AAE enhances feature extraction from gene expression profiles through adversarial training to obtain their latent embeddings. The GCN captures spatial latent embeddings from the spatial graph. These two embeddings are then integrated via a fusion module and reconstructed by the decoder.

features in the residual connection. FC denotes a fully connected layer to implement linear transformation. z_h represents the latent embedding of the spatial spots.

Subsequently, we add a discriminator layer to the hybrid encoder, allowing it to act as the discriminator in the GAN framework. During the training process, the generator progressively generates simulated data to optimize the gene expression profiles, effectively compensating for missing high-expression signals. The discriminator is trained to distinguish between real and simulated data, thereby enhancing the ability of DACN to recognize critical gene features, particularly for key genes. Through adversarial training with the generator, the discriminator gradually enhances its ability to differentiate real data from fake data. In the early stages of training, the discriminator may lack the ability to properly identify real and fake data, causing it to converge too rapidly. This can hinder effective gradient updates for the generator and lead to unstable training. To balance the training dynamics between the generator and the discriminator, we introduce a hyperparameter N that controls the number of discriminator updates per iteration. This strategy helps improve the overall training stability of DACN. The AAE is defined as follows:

$$\mathcal{L}_D = \frac{1}{2}\left[\mathrm{Bce}(D(z_t), 1) + \mathrm{Bce}(D(z_f), 0)\right] \tag{4}$$

$$\mathcal{L}_G = \mathrm{Bce}(D(z_f), 1) \tag{5}$$

where z_t and z_f denotes the latent embedding of real data and simulated data, respectively. D denotes the discriminator that is trained to assign predictive labels to both real and synthetic latent embeddings. The loss of discriminator and generator loss ($\mathcal{L}_D$ and $\mathcal{L}_G$) are calculated using the binary cross-entropy (Bce) function between the true and predicted labels.

Graph convolutional network

After obtaining the low-dimensional embeddings from the AAE, we construct a graph based on the spatial coordinates of each spot. A three-layer GNN is then designed to extract spatial features from the graph. Specifically, the first layer of GCN maps the low-dimensional embeddings (z_h) to match the dimensionality of the feature matrix of spatial x. Subsequently, the second and third layers of GCN are used to estimate the mean and variance, respectively. To further enhance the performance of DACN, we employ the reparameterization trick that allows graph embeddings to be resampled from the latent space based on the learned mean and variance. The GCN is defined as follows:

$$H^{(1)} = \mathbf{ReLU}(A \cdot z_h W_1) \tag{6}$$

$$\mu = \mathbf{ReLU}\left(A \cdot H^{(1)} W_2\right) \tag{7}$$

$$\log \sigma^2 = \mathbf{ReLU}\left(A \cdot H^{(1)} W_3\right) \tag{8}$$

$$z_g = \mu + \epsilon \cdot \sigma \tag{9}$$

where $H^{(1)}$ denotes the graph embeddings obtained from the first layer of GCN. W_1, W_2, and W_3 represent three independent learnable weight matrices. A is an adjacency matrix which is constructed based on the spatial coordinates of each spot. μ and σ correspond to the mean vector and the normalized log-variance of the latent space, respectively. ϵ is a random noise vector sampled from a standard normal distribution. z_g is the latent embedding which is generated by using the reparameterization trick.

Decoder

The decoder is employed to reconstruct features from the low-dimensional embeddings extracted by the AAE and the GCN module, enabling the computation of the reconstruction loss. The decoder is

defined as follows:

$$z = \mathbf{concat}(z_h, z_g) \tag{10}$$

$$\tilde{x} = \mathbf{ReLU}(A \cdot \mathbf{Dropout}(z)W) \tag{11}$$

where z denotes the final latent embedding obtained by concatenating z_h and z_g. Dropout is a technique that applies a dropout operation to reduce overfitting. A and W represent the adjacency matrix and weight matrix, respectively. $\tilde{x}$ denotes the reconstructed features.

Results

Dataset

To comprehensively evaluate the performance of DACN in various ST datasets, we conduct experiments on multiple publicly datasets with different spatial resolutions. The human dorsolateral prefrontal cortex (DLPFC) dataset is generated by using the 10× Genomics Visium and consists of 12 sections derived from human dorsolateral prefrontal cortex tissues. Each section contains ∼4000 spatial spots and manual annotations of cortical layers. Due to its clearly defined spatial structure and widespread use in benchmarking studies, this dataset is used to evaluate the performance of spatial domain identification as a reliable dataset. The BaristaSeq dataset is derived from mouse primary cortex tissue. This dataset is generated using BaristaSeq technology with high resolution. It comprises three tissue sections, with 1525, 2042, and 1690 cells, respectively. The osmFISH dataset is obtained by using the osmFISH (osmotic-based fluorescent *in situ* hybridization) technique and offers spatial transcriptomic data from the mouse somatosensory cortex at single-cell resolution. It includes a total of 4839 cells.

Evaluation of spatial domain identification on the DLPFC Dataset

We evaluate the clustering performance of DACN on DLPFC dataset. To ensure a comprehensive comparison, we select some representative methods: SEDR [16], SpaGCN [15], GraphST [19], SpaceFlow [18], DeepST [26], sigra [20], and STAGATE [17]. For 12 spatial tissue sections, we compute adjusted rand index (ARI) and normalized mutual information (NMI) which are two unsupervised clustering metrics to assess the agreement between predicted clusters and ground truth. ARI measures the similarity between the predicted clusters and the true labels by quantifying their pairwise agreements. NMI evaluates the mutual dependence between the predicted and true labels based on information theory. The clustering results of all methods are visualized by using boxplots, which is shown in Fig. 2a. Overall, DACN demonstrates more stable and superior clustering performance across all 12 samples, highlighting its robustness and effectiveness in spatial domain identification. In addition, we perform t-tests to assess the statistical significance of differences in ARI and NMI between DACN and other methods. Asterisks indicate that DACN achieved significantly higher ARI and NMI scores than the other methods (P <.05).

To further validate the reliability of the low-dimensional embedding extracted by DACN, we select slice sample 151 674 from the DLPFC dataset for clustering analysis. The spatial domain identification results for all methods are presented in Fig. 2b. From the results, both SpaGCN and GraphST exhibit suboptimal performance, with evident inconsistencies between the predicted spatial layers and the ground truth. Their domain boundaries are often ambiguous. Although STAGATE and SpaceFlow achieve relatively low ARI scores, they still produce clearer and more structured spatial partitions. On the other hand, SEDR and DeepST generate a spatial configuration more consistent with the true layer distribution. However, a misalignment appears in Layers 2 and 3, with their predicted positions nearly reversed. In contrast, DACN effectively captures spatial structures that closely align with the true anatomical organization. Its segmentation boundaries are smoother and more consistent, with no noticeable misplacement between adjacent layers. It demonstrates its superior performance in spatial domain identification. It is worth noting that a gray region appears in the lower right corner of the ground truth spatial annotation, corresponding to unannotated spots in the dataset. It can be observed that this region is adjacent to Layers 2 and 3, suggesting a strong biological relevance to these two layers. However, except for DACN and STAGATE, all other methods failed to correctly identify this relationship, instead assigning these spots to layers spatially distant from their actual locations. Although STAGATE successfully categorized the region as part of Layer 2, it still misclassified some spots as belonging to Layer 1. In contrast, DACN produces more accurate spatial domain identification, demonstrating its advantage in modeling spatial domain relationships.

In addition, we perform differential gene expression analysis for each cluster identified by DACN, with the results displayed in Fig. 2c. Among these clusters, MBP and MALAT1 are identified as the most significantly expressed genes in the white matter (WM) and Layer 1 regions, respectively. Their spatial expression patterns are visualized in Fig. 2c. It is shown that the distributions of MBP and MALAT1 are consistent with the ground truth, with MBP predominantly localized in the WM region and MALAT1 mainly enriched in Layer 1. MBP is well known for its essential role in the maintenance and function of white matter in the brain [27]. We further visualized the clustering results of all methods by using UMAP, assigning the predicted spatial domain identities at their respective centroids. As shown in Fig. 2d, the clustering results of the DACN exhibit more compact intra-cluster structures and minimal inter-cluster overlap. These findings suggest that the low-dimensional latent representation learned by DACN effectively captures the intrinsic structural organization of spatial transcriptomic data, thereby achieving superior clustering performance compared with existing methods.

Demonstrating the effectiveness of spatial domain identification and batch effect correction on the BaristaSeq dataset

To further evaluate the applicability of DACN to ST datasets with varying spatial resolutions, we conduct experiments on the BaristaSeq dataset. BaristaSeq is a high-resolution ST technology with a spot diameter ranging from 5 to 10 μm. The dataset consists of three tissue sections. For each section, we calculate the ARI and the NMI to quantitatively assess clustering performance across different methods. As shown in Fig. 3a, DACN consistently outperformed other methods on all three sections, demonstrating its superior ability to model high-resolution ST data. In addition, we visualize the results of spatial domain identification on slice 2, as shown in Fig. 3b. The spatial domains identified by SpaGCN, DeepST and GraphST appear disorganized, lacking a clear laminar structure. It indicates limited spatial recognition capabilities on high-resolution data. In contrast, DACN, SEDR, SpaceFlow, and STAGATE all successfully identified the

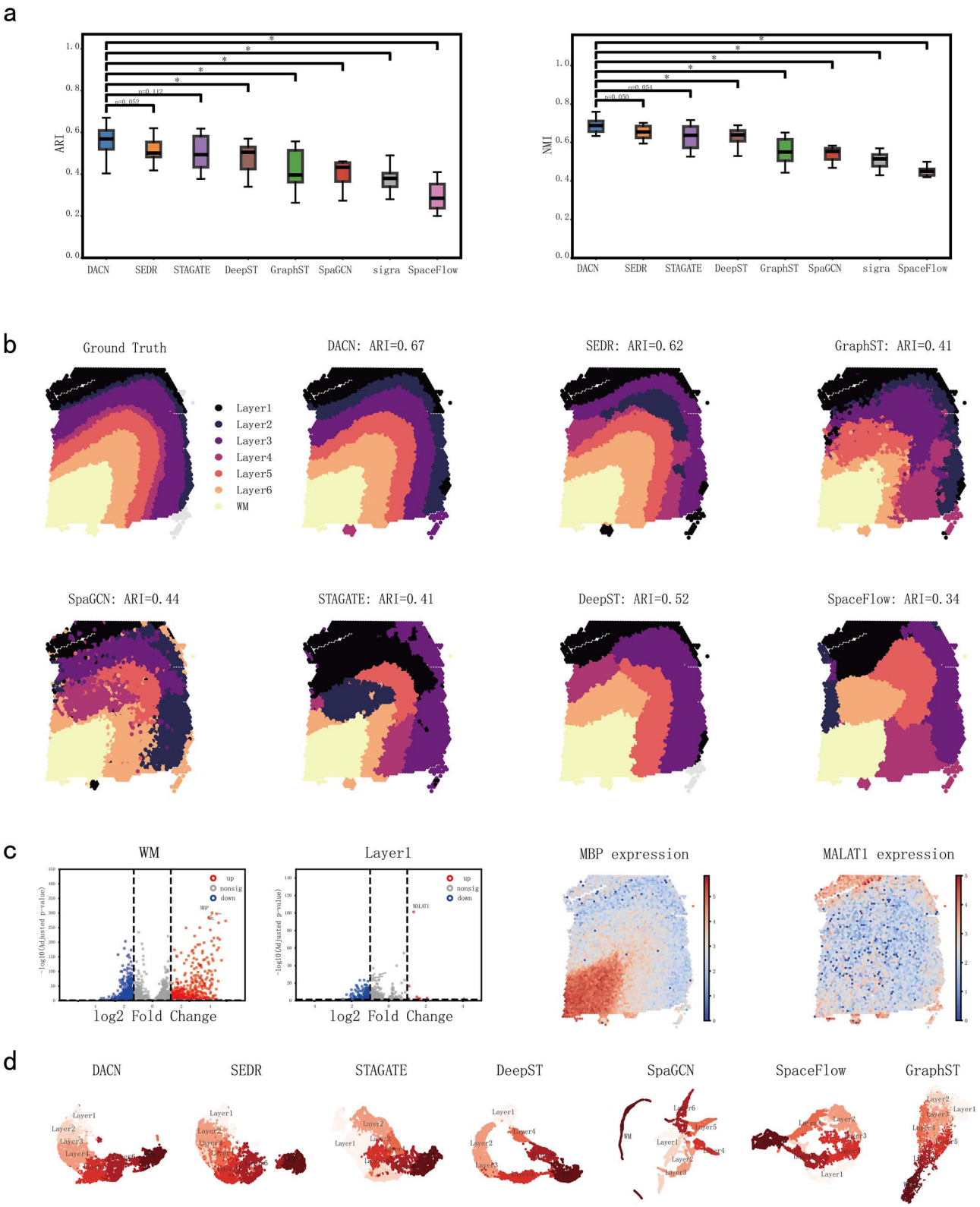

Figure 2 Evaluation of DACN in identifying spatial domains. (a) ARI and NMI box plots comparing different methods across 12 DLPFC slices. (b) Spatial domain identification on slice 151 674 of the DLPFC dataset. (c) Spatial clustering results validated by the spatial expression patterns of marker genes consistent with known anatomical domains. (d) UMAP representations of clustering results on slice 151 674 for each method.

spatial architecture, showing strong spatial pattern recognition abilities. However, some misclassifications were still observed in specific regions across SEDR, SpaceFlow, and STAGATE. DACN exhibits fewer misclassified areas and yields more coherent spatial domains with more accurate boundaries. Compared with STAGATE and SpaceFlow, DACN demonstrates a slight advantage in overall

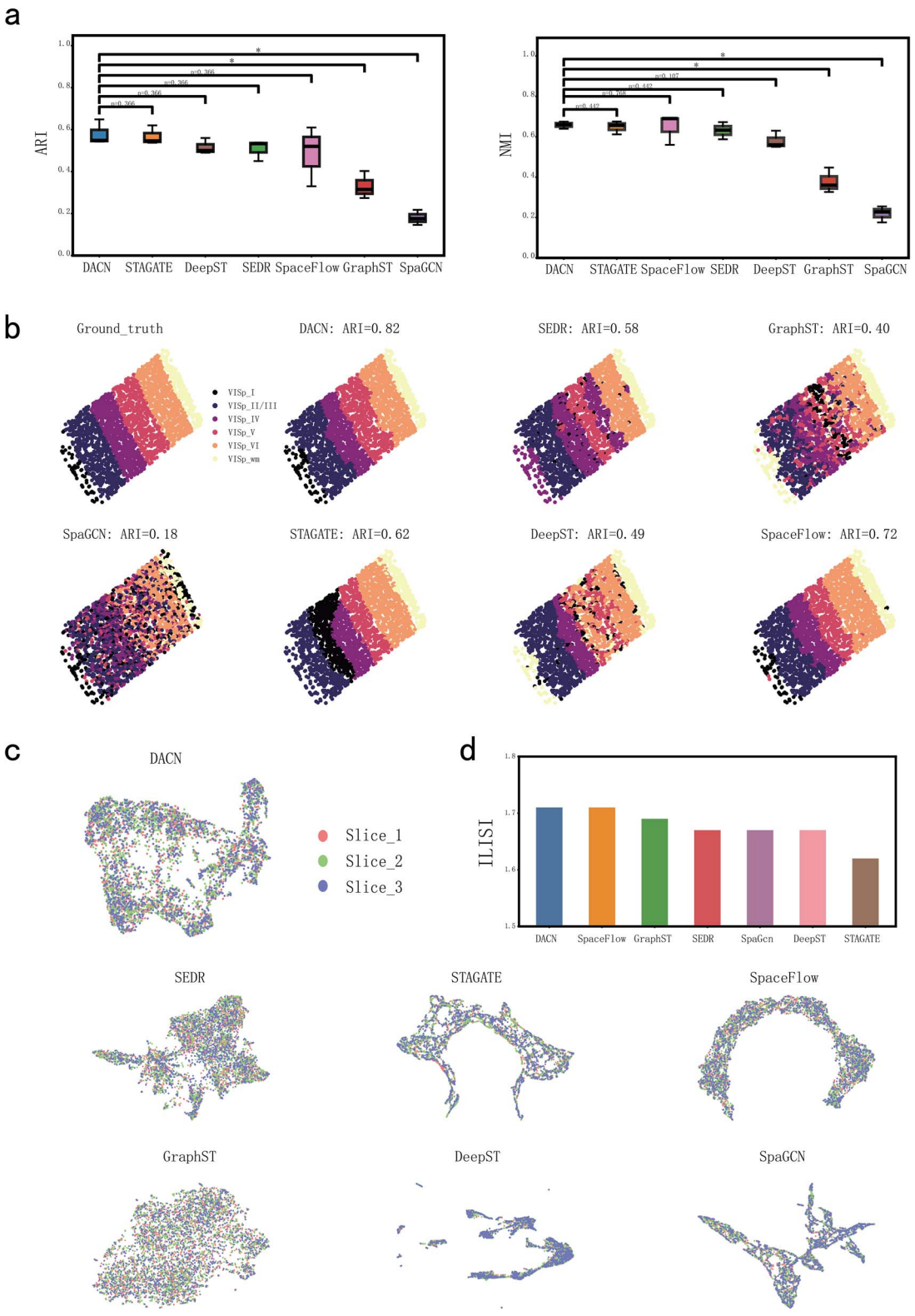

Figure 3 Performance comparison of DACN on the high-resolution dataset (BaristaSeq). (a) Box plots of ARI and NMI scores for different methods on three slices from the BaristaSeq dataset. (b) Visualization of spatial domain identification results on slice 2 of the BaristaSeq dataset. (c) Comparison of batch effect correction performance across different methods. (d) ILISI scores of different methods for evaluating batch effect correction effectiveness.

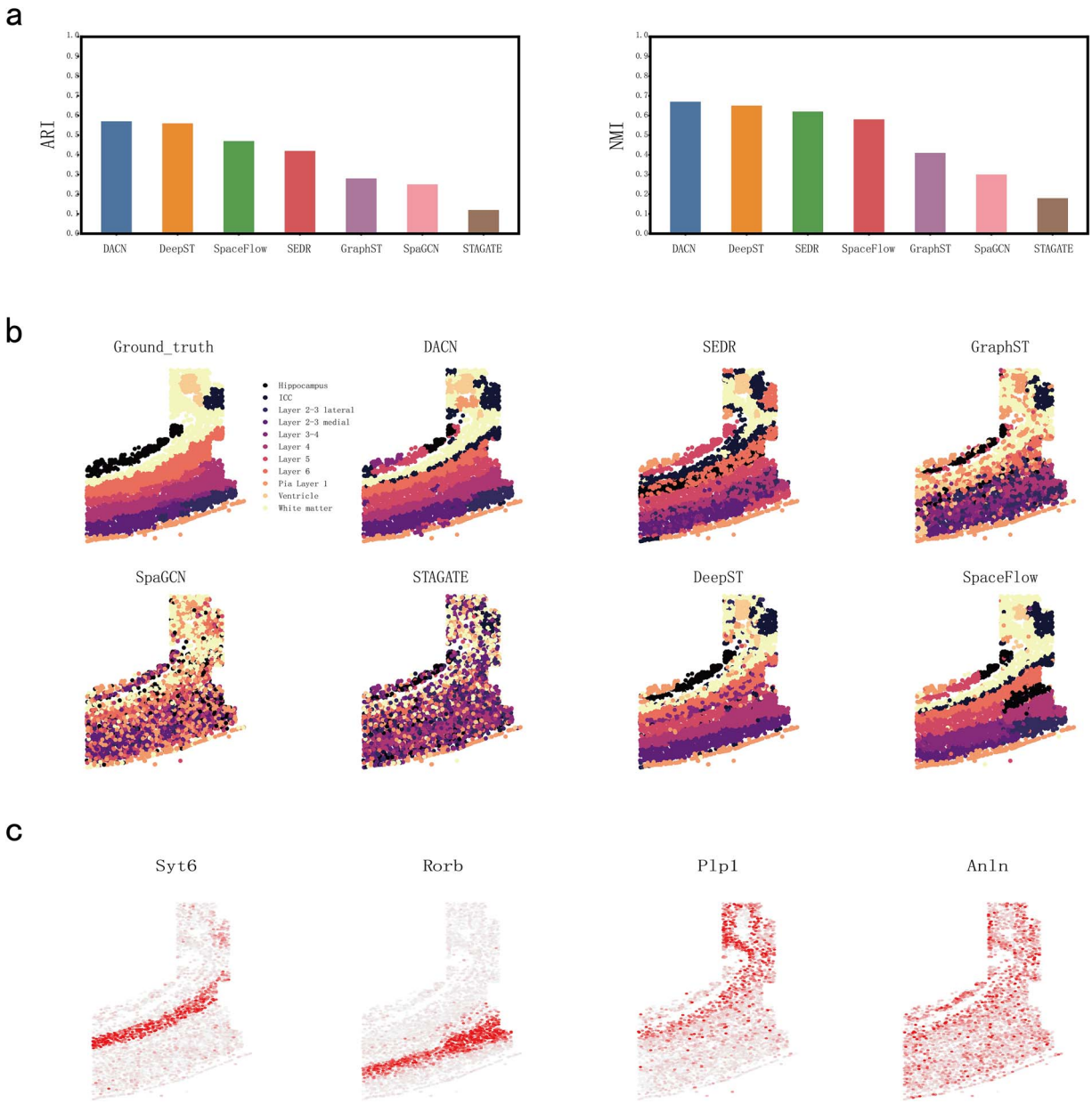

Figure 4 Application of DACN on subcellular-resolution spatial transcriptomics dataset (osmFISH). (a) Comparison of ARI and NMI scores for different methods on the osmFISH dataset. (b) Spatial domain identification results on the osmFISH dataset. (c) Spatial expression patterns of marker genes for glutamatergic neurons and oligodendrocytes.

clustering accuracy. When compared with SEDR, DACN shows a better ability to capture the inter-layer boundaries, leading to higher spatial identification precision.

In addition, we evaluate the performance of DACN in addressing batch effects. Specifically, we utilize Harmony to compare the integration capabilities of DACN with other approaches on BaristaSeq dataset. Figure 3d shows the visualization of the latent space after batch correction. As observed, the spot from different batches in results of DACN are more evenly distributed, with no obvious separation among tissue sections. This demonstrates that DACN is effective in mitigating batch effects. To quantitatively assess the integration performance across batches, we employ the Integration Local Inverse Simpson's Index (iLISI) to evaluate the degree of batch mixing in the latent space. An iLISI value closer to the true number of batches indicates better batch effect correction. As shown in Fig. 3c, DACN achieves iLISI score which is closest to the actual number of batches and outperforms other methods. It further validates its effectiveness in batch effect correction.

Improved identification of spatial domain on the subcellular-resolution osmFISH dataset

To assess the performance of DACN in spatial domain identification on ST data with subcellular spatial resolution, we conduct clustering evaluation on the osmFISH dataset. As shown in Fig. 4a, DACN achieves the highest performance in both ARI and NMI, significantly outperforming other methods. To further compare spatial domain identification capabilities, we visualize the clustering results from different methods on the osmFISH dataset, which is shown in Fig. 4b. The results show

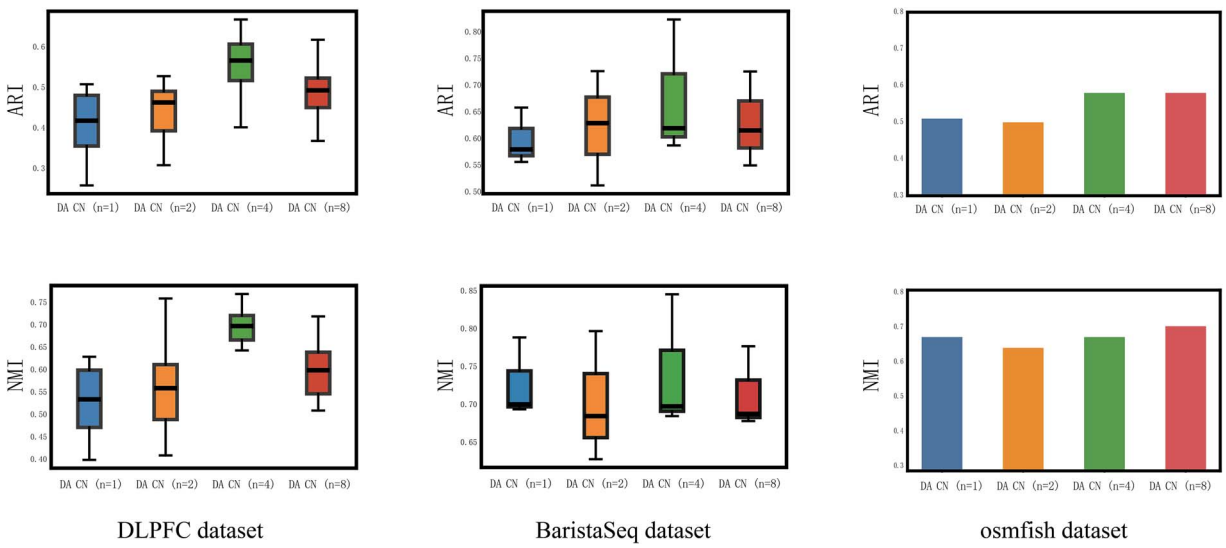

Figure 5 Parameter sensitivity analyses. ARI and NMI performance across varying numbers of attention heads (n).

that SpaGCN and STAGATE fail to accurately delineate the spatial structures, exhibiting poor identification performance. GraphST and SEDR can accurately identify part of the spatial domain. In contrast, the spatial domain identified by DACN presents smoother and more clear boundaries. These results demonstrate that DACN effectively captures key spatial features in ST datasets with subcellular resolution and low throughput.

Furthermore, we visualize the spatial expression patterns of several marker genes for specific neuronal cell types in the osmFISH dataset (Fig. 4c). Specifically, Syt6 and Rorb are established markers of glutamatergic neurons, whereas Plp1 and Anln serve as markers of oligodendrocytes. Glutamatergic neurons are predominantly distributed across cortical layers L2–L6, while oligodendrocytes are mainly localized within the white matter [28]. Consistently, Syt6 and Rorb show strong expression in layers L4–L6, whereas Plp1 and Anln are enriched in the white matter domain. As shown in Fig. 4b, the spatial distributions of these genes closely coincide with the regions identified by DACN, further supporting the biological relevance of the spatial clustering results.

Parameter analysis

To comprehensively evaluate the effect of the number of attention heads (n), we compare the performance of DACN across different values of n (1, 2, 4, and 8). To ensure a robust and representative comparison, all experiments are performed on three ST datasets previously used, each with distinct spatial resolutions. The results are summarized in Fig. 5. It can be observed that DACN ($n = 4$) exhibits the most stable performance across all three datasets. It demonstrates a clear performance advantage in the DLPFC dataset, whereas the performance gains are less pronounced in BaristaSeq dataset and osmFISH dataset. The limited improvement may stem from the relatively low gene throughput of datasets BaristaSeq and osmFISH, which constrains the effectiveness of the multi-head attention mechanism under low-throughput conditions.

Ablation study

To assess the contribution of the core AAE module, we ablate the AAE component (DACN w/o AAE) and evaluate its effect on model performance. For a fair and robust comparison, all experiments are performed on three previously used ST datasets with distinct spatial resolutions. As summarized in Fig. 6, DACN consistently outperforms its ablated counterpart across all datasets. On the low-resolution DLPFC dataset, DACN achieves moderate improvements in both ARI and NMI. Notably, on the high-resolution BaristaSeq and osmFISH datasets, the inclusion of the AAE led to markedly greater performance gains. These results demonstrate that integrating the AAE substantially enhances the capacity of DACN to capture complex spatial transcriptomic patterns.

Implementation

The DACN method is implemented by using Python 3.9 and R 4.3. We use the PyTorch framework together with the Scanpy and Anndata packages for ST data analysis. The rpy2 package enables interaction between Python and R, and the Mclust package in R is used for clustering. The model training is performed on a workstation equipped with an NVIDIA RTX 1060 GPU (8 GB memory). Consistent training parameters are applied across all datasets, including a learning rate of 0.01 and 550 training epochs. An early stopping mechanism monitors the loss and saves the model weights corresponding to the minimum loss value. The complete training procedure is available in our GitHub repository (https://github.com/lanbiolab/DACN).

Conclusion

With the advent of ST technologies, single-cell omics studies have gained access to crucial spatial information [29, 30]. However, ST data often suffer from expression noise and dropout effects, which frequently lead to inaccurate or unreliable spatial transcriptomic analyses. To address these challenges, we propose DACN, which uses the improved AAE to address the impact of noise on training and adapts to ST datasets with varying resolution. To evaluate the performance of DACN, we conduct experiments on multiple ST datasets and compared it against other state-of-the-art methods. In addition, we select ST datasets with varying spatial resolutions to demonstrate the robustness of the model, including the low-resolution DLPFC dataset, the high-resolution BaristaSeq dataset, and the subcellular-resolution

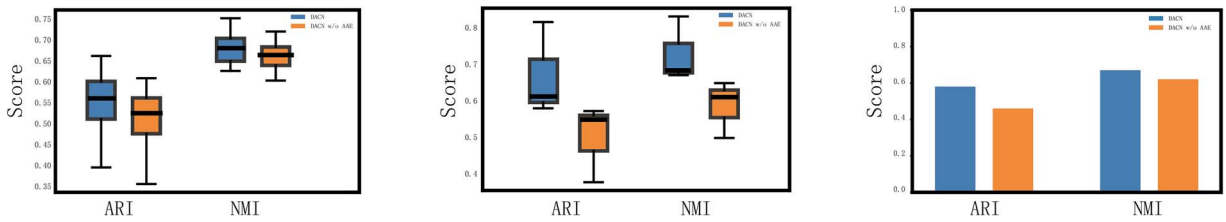

Figure 6 Ablation study analyses. Comparative performance of DACN and DACN without the AAE module (DACN w/o AAE) on the DLPFC, BaristaSeq, and osmFISH datasets.

osmFISH dataset. Experimental results demonstrate that DACN consistently outperforms existing methods in spatial clustering tasks on all datasets, exhibiting superior performance in batch effect correction and denoising capabilities.

Overall, DACN has demonstrated remarkable performance on ST datasets with varying spatial resolutions. Nevertheless, DACN relies on the spatial adjacency matrix that may fail to capture higher-order relationships among spatial spots. Moreover, the computational cost increases with the number of spatial spots, posing challenges for the analysis of large-scale datasets [31, 32]. In future work, we plan to integrate scRNA-seq data to bridge gene expression patterns across individual cells [33]. Furthermore, incorporating protein–protein interaction networks could enable the modeling of higher-order biological relationships, providing a more comprehensive understanding of spatial organization at the molecular level. To address the computational overhead, we will explore more efficient neural architectures [34, 35] that offers lower computational complexity.

Key Points

- At present, resolution-dependent variability in gene throughput remains a key barrier to accurate spatial domain identification in spatial transcriptomics.
- We present DACN, a robust framework for spatial transcriptomics (ST) analysis across datasets of varying resolutions. To accommodate differences in gene throughput, DACN integrates residual connections and multi-head attention into an enhanced AAE to improve feature representation and denoise ST data, while GCNs are employed to couple spatial context with gene expression.
- Across diverse ST datasets, DACN consistently achieves superior accuracy in spatial domain identification and effective batch-effect correction.

Conflicts of interest

None declared.

Funding

This work was supported by the Natural Science Foundation of Guangxi Zhuang Autonomous Region of China (grant no. 2024GXNSFFA010006), the National Natural Science Foundation of China (grant nos U24A20256 and 62472108), the Guangxi Bagui Youth Talent Program, the Project of Guangxi Key Laboratory of Eye Health (grant no. GXYJK-202407), the 2025 Guangxi Education Department Graduate Innovation Project (grant no. ZX01030030725006), Natural Science Foundation of Hunan Province (grant no. 2023JJ50117), the project of Hunan Intelligent Rehabilitation Robot and Auxiliary Equipment Engineering Technology Research Center (grant no. 2025AQ104) and the Project of Guangxi Health Commission Eye and Related Diseases Artificial Intelligence Screen Technology Key Laboratory (grant no. GXYAI-202402).

Data availability

The codes and datasets are publicly available online at our GitHub repository (https://github.com/lanbiolab/DACN).

References

1. Lan W, Chen J, Liu M *et al.* Deep imputation bi-stochastic graph regularized matrix factorization for clustering single-cell RNA-sequencing data. *IEEE Trans Comput Biol Bioinform* 2024. https://doi.org/10.1109/TCBB.2024.3387911
2. Lan W, Ling T, Chen Q *et al.* scMoMtF: an interpretable multitask learning framework for single-cell multi-omics data analysis. *PLoS Comput Biol* 2024;**20**:e1012679. https://doi.org/10.1371/journal.pcbi.1012679
3. Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 2015;**16**:133–45. https://doi.org/10.1038/nrg3833
4. Lan W, He G, Liu M *et al.* Transformer-based single-cell language model: a survey. *Big Data Mining and Analytics* 2024;**7**:1169–86.
5. Ståhl PL, Salmén F, Vickovic S *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**:78–82. https://doi.org/10.1126/science.aaf2403
6. Ji AL, Rubin AJ, Thrane K *et al.* Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell* 2020;**182**:497–514.e22. https://doi.org/10.1016/j.cell.2020.05.039
7. Rodriques SG, Stickels RR, Goeva A *et al.* Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;**363**:1463–7. https://doi.org/10.1126/science.aaw1219
8. Stickels RR, Murray E, Kumar P *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* 2021;**39**:313–9. https://doi.org/10.1038/s41587-020-0739-1
9. Zhang M, Eichhorn SW, Zingg B *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* 2021;**598**:137–43. https://doi.org/10.1038/s41586-021-03705-x
10. Eng C-HL, Lawson M, Zhu Q *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqfish+. *Nature* 2019;**568**:235–9. https://doi.org/10.1038/s41586-019-1049-y
11. Wang X, Allen WE, Wright MA *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;**361**:eaat5691. https://doi.org/10.1126/science.aat5691

12. Yang J, Zheng Z, Jiao Y *et al.* Spotiphy enables single-cell spatial whole transcriptomics across an entire section. *Nat Methods* 2025;**22**:724–736. https://doi.org/10.1038/s41592-025-02622-5
13. Dries R, Zhu Q, Dong R *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 2021;**22**:1–31.
14. Zhao E, Stone MR, Ren X *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* 2021;**39**:1375–84. https://doi.org/10.1038/s41587-021-00935-2
15. Jian H, Li X, Coleman K *et al.* SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 2021;**18**:1342–51.
16. Hang X, Huazhu F, Long Y *et al.* Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med* 2024;**16**:12.
17. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat Commun* 2022;**13**:1739. https://doi.org/10.1038/s41467-022-29439-6
18. Ren H, Walker BL, Cang Z *et al.* Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nat Commun* 2022;**13**:4076. https://doi.org/10.1038/s41467-022-31739-w
19. Long Y, Ang KS, Li M *et al.* Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nat Commun* 2023;**14**:1155.
20. Tang Z, Li Z, Hou T *et al.* SiGra: single-cell spatial elucidation through an image-augmented graph transformer. *Nat Commun* 2023;**14**:5618. https://doi.org/10.1038/s41467-023-41437-w
21. Lan W, Yang T, Chen Q *et al.* Multiview subspace clustering via low-rank symmetric affinity graph. *IEEE Trans Neural Netw Learn Syst* 2023;**35**:11382–95. https://doi.org/10.1109/TNNLS.2023.3260258
22. Wenju X, Keshmiri S, Wang G. Adversarially approximated autoencoder for image generation and manipulation. *IEEE Trans Multimed* 2019;**21**:2387–96. https://doi.org/10.1109/TMM.2019.2898777
23. Lan W, He GH, Zhou WH *et al.* scMCG: Analyzing a single-cell assay for transposase-accessible chromatin using sequencing data based on contrastive learning and generative adversarial network. *J Comput Sci Technol* 2025;**40**:1639–1649.
24. Norouzi R, Norouzi R, Abbasi K *et al.* DFT_ANPD: a dual-feature two-sided attention network for anticancer natural products detection. *Comput Biol Med* 2025;**194**:110442.
25. Abbasi K, Razzaghi P. Incorporating part-whole hierarchies into fully convolutional network for scene parsing. *Expert Systems with Applications* 2020;**160**:113662. https://doi.org/10.1016/j.eswa.2020.113662
26. Chang X, Jin X, Wei S *et al.* DeepST: identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res* 2022;**50**:e131–1.
27. Young VG, Halliday GM, Kril JJ. Neuropathologic correlates of white matter hyperintensities. *Neurology* 2008;**71**:804–11. https://doi.org/10.1212/01.wnl.0000319691.50117.54
28. Ma Y, Liu L, Zhao Y *et al.* HyperGCN: an effective deep representation learning framework for the integrative analysis of spatial transcriptomics data. *BMC Genomics* 2024;**25**:566. https://doi.org/10.1186/s12864-024-10469-x
29. Lan W, Liu M, Chen J *et al.* JLONMFSC: clustering scRNA-seq data based on joint learning of non-negative matrix factorization and subspace clustering. *Methods* 2024;**222**:1–9. https://doi.org/10.1016/j.ymeth.2023.11.019
30. Lan W, Tang Z, Liu M *et al.* The large language models on biomedical data analysis: a survey. *IEEE J Biomed Health Inform* 2025;**29**:4486–97. https://doi.org/10.1109/JBHI.2025.3530794
31. Lan W, Tang Z, Liao H *et al.* MULGONET: an interpretable neural network framework to integrate multi-omics data for cancer recurrence prediction and biomarker discovery. *Fundamental Research* 2025;**6**:99–110. https://doi.org/10.1016/j.fmre.2025.01.004
32. Lan W, Liao H, Chen Q *et al.* DeepKEGG: a multi-omics data integration framework with biological insights for cancer recurrence prediction and biomarker discovery. *Brief Bioinform* 2024;**25**:bbae185.
33. Yang W, Wang P, Shouping X *et al.* Deciphering cell–cell communication at single-cell resolution for spatial transcriptomics with subgraph-based graph attention network. *Nat Commun* 2024;**15**:7101. https://doi.org/10.1038/s41467-024-51329-2
34. Deng Z, Wang Z, Shan Y *et al.* COO-DuDo: computation overhead optimization methods for dual-domain sparse-view CT reconstruction. *Expert Systems with Applications* 2025;**286**:128109. https://doi.org/10.1016/j.eswa.2025.128109
35. Albert G, Dao T. *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*. In *First Conference on Language Modeling (COLM 2024)*. Philadelphia, PA, USA: OpenReview.net; 2024.

Briefings in Bioinformatics, 2026, **27**, bbag070
https://doi.org/10.1093/bib/bbag070
Problem Solving Protocol