

RESEARCH ARTICLE

SpaLSTF: Diffusion-based generative model with BiLSTM and XCA-Transformer for spatial transcriptomics imputation

Lin Yuan^{1,2,3}, Yufeng Jiang^{1,2,3}, Boyuan Meng^{1,2,3}, Qingxiang Wang^{1,2,3}, Cuihong Wang^{4*}, De-Shuang Huang^{5,6*}

1 Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, **2** Shandong Engineering Research Center of Big Data Applied Technology, Faculty of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China, **3** Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan, China, **4** The Second Qilu Hospital of Shandong University, Shandong University, Jinan, China, **5** Institute for Regenerative Medicine, Medical Innovation Center and State Key Laboratory of Cardiology, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China, **6** Ningbo Key Laboratory of Multi-Omics & Multimodal Biomedical Data Mining and Computing, Eastern Institute of Technology, Ningbo, China

* xqawc1120@163.com (CW), dshuang@eitech.edu.cn (D-SH)



OPEN ACCESS

Citation: Yuan L, Jiang Y, Meng B, Wang Q, Wang C, Huang D-S (2026) SpaLSTF: Diffusion-based generative model with BiLSTM and XCA-Transformer for spatial transcriptomics imputation. PLoS Comput Biol 22(2): e1013954. <https://doi.org/10.1371/journal.pcbi.1013954>

Editor: Guang-Zhong Wang, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, CHINA

Received: October 18, 2025

Accepted: January 28, 2026

Published: February 10, 2026

Copyright: © 2026 Yuan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The source code for SpaLSTF is available at <https://github.com/nathanyl/SpaLSTF>.

Funding: This work is supported by the National Natural Science Foundation of China

Abstract

Spatial transcriptomics (ST) technologies provide powerful tools for analyzing spatial distribution patterns of gene expression in tissue samples. However, they are limited by sparse gene detection and incomplete expression coverage. Several computational approaches based on reference scRNA-seq have been proposed to impute ST data and have achieved impressive results. However, these methods fail to fully explore latent temporal dependencies among cells and cannot accurately capture hidden gene-level regulatory mechanisms. To overcome those limitations, we propose SpaLSTF, a novel method for enhancing ST gene expression using a conditional diffusion model guided by scRNA-seq data. SpaLSTF captures gene expression relationships through a dual Markov process: one progressively perturbs scRNA-seq data with noise, while the other denoises it to reconstruct the original distribution. To effectively model contextual dependencies among cell states, we adopt a bidirectional long short-term memory (BiLSTM) network. Furthermore, we design a cross-covariance attention mechanism within a Transformer (XCA-Transformer) to efficiently compute attention coefficients between gene expression and accurately predict the noise added at each step. In addition, we introduce a variational lower bound (VLB) objective and introduce Kullback-Leibler (KL) divergence as a regularization term, along with mean squared error loss, to ensure that the generated noise follows the target distribution. We compared the performance of SpaLSTF with seven state-of-the-art methods on twelve cross-platform datasets covering a variety of tissues and organs using nine evaluation metrics. Experimental results demonstrated

(62472239 to LY; 62333018, 62372255 to DSH), the Shandong Provincial Natural Science Foundation (ZR2024MF011 to LY), the Youth Innovation Team of Colleges and Universities in Shandong Province (2023KJ329 to LY), the Joint Project of National Natural Science Foundation of China and Russian Science Foundation (W2412087 to DSH), the Natural Science Foundation of Guizhou Province (ZK2024ZD035 to DSH), the Natural Science Foundation of Ningbo City (2023J199 to DSH), the Key Research and Development (Digital Twin) Program of Ningbo City (2023Z219 and 2023Z226 to DSH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

that SpaLSTF outperforms competing methods in gene expression imputation, cell population identification, and spatial structure preservation.

Author summary

Computational approaches based on reference scRNA-seq have been proposed to impute ST data and have some limitations. We propose SpaLSTF, a novel method for enhancing ST gene expression using a conditional diffusion model guided by scRNA-seq data. SpaLSTF captures gene expression relationships through a dual Markov process. We adopt a bidirectional long short-term memory (BiLSTM) network to effectively model contextual dependencies among cell states. Furthermore, we design a cross-covariance attention mechanism within a Transformer (XCA-Transformer) to efficiently compute attention coefficients between gene expression and accurately predict the noise added at each step. In addition, we adopt a variational lower bound (VLB) objective and introduce Kullback-Leibler (KL) divergence as a regularization term, along with mean squared error loss, to ensure that the generated noise follows the target distribution. Experimental results demonstrated that SpaLSTF outperforms competing methods in gene expression imputation, cell population identification, and spatial structure preservation.

Introduction

Spatial transcriptomics (ST) technologies have greatly promoted the exploration of spatial cellular organization [1,2]. However, ST-derived gene expression data are highly sparse with limited detected genes, hindering accurate gene quantification and spatial analysis. Gene imputation in ST data faces huge challenges [3,4]. Single-cell RNA sequencing (scRNA-seq) provides single-cell resolution high-throughput transcriptome data that substantially improve the quality and interpretability of spatial gene expression data [5,6], facilitating more accurate ST studies [7–9]. In recent years, computational approaches based on reference scRNA-seq data have become indispensable for gene imputation and comprehensively decoding the spatial transcriptome landscape [10–13].

Numerous computational approaches (Tangram [14], gimVI [15], stPlus [16], SpaGE [17], uniPort [18], and SpatialScope [19]) have emerged to recover missing gene expression leveraging reference scRNA-seq data. These approaches typically presume that gene expression patterns in scRNA-seq and ST data are comparable. By analyzing expression patterns of shared genes, they compute the similarity between scRNA-seq cells and ST spots, and utilize scRNA-seq profiles to impute undetected genes in ST data. Consequently, accurate cross-modal cell alignment is crucial for high-fidelity imputation. However, both ST and scRNA-seq data suffer from sparsity, and the number of genes shared across modalities is often limited,

which complicates the matching process [20]. In addition, batch effects between datasets further distort similarity measures, and neither decoder nor K-nearest neighbor (KNN)-based methods employed in the aforementioned computational approaches can effectively address batch effects. As a result, the predicted gene expression and original ST measurements may reside in distinct batch spaces, reducing the accuracy of imputation and complicating downstream analyses [21–23].

To address these challenges, a diffusion-based model [24,25] stDiff [26] was developed. Unlike cell-similarity-based methods, stDiff extracts latent gene regulatory signals from scRNA-seq data to guide the imputation of ST data. Based on the assumption that ST and scRNA-seq data from the same tissue region share analogous gene expression profiles, stDiff learns the inter-gene dependencies present in scRNA-seq and leverages this information to predict undetected genes in ST, thereby enabling reconstruction of a more complete ST data [27]. However, stDiff does not fully explore latent temporal dependencies among cells, which may lead to imputation results that fail to accurately reflect true gene expression patterns within tissue microenvironments. In addition, traditional self-attention mechanisms operate at the cellular level and is difficult to accurately capture the hidden gene-level regulatory mechanisms, which leads to noise prediction bias and reduces imputation accuracy.

To overcome these limitations, we propose SpaLSTF, as illustrated in Fig 1, a conditional diffusion model incorporating BiLSTM [28], XCA-Transformer, and KL divergence [29] regularization to improve imputation accuracy while preserving the structural integrity of ST data and enhancing generalization capability. Specifically,

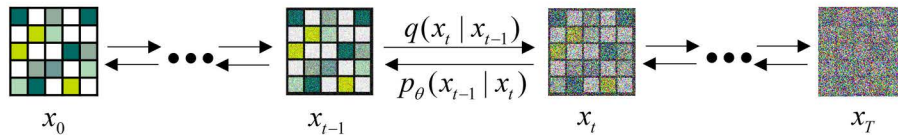
SpaLSTF incorporates BiLSTM network into conditional data and utilize its sequential modeling capabilities to capture complex interrelationships among cell states, thereby preserving the spatiotemporal structure of ST data. Although ST data are not temporal sequences, spatially adjacent spots often exhibit strong biological continuity and correlated gene expression profiles. By capturing bidirectional contextual information, BiLSTM effectively models this spatial continuity, thereby improving the accuracy of imputing unmeasured genes. We introduce a Transformer module with cross-covariance attention mechanism [30], called XCA-Transformer, which directly computes attention relationships between gene expression abundances. The XCA (cross-covariance attention) module is designed to explicitly model dependencies among genes along the gene dimension. Compared with conventional self-attention mechanisms that primarily focus on spatial or cell-level interactions, XCA is more suitable for capturing gene–gene co-expression and regulatory relationships, which are critical for high-dimensional gene expression imputation, enhancing the model’s ability to capture gene interaction patterns and improving the accuracy of noise prediction. To enhance the generative capability of the model, we adopt the concept of variational lower bound [31] (VLB) and use Kullback-Leibler (KL) divergence as a regularization term combined with mean squared error (MSE) loss [32]. This combination guides generated gene expression toward the target distribution while alleviating batch bias. To the best of our knowledge, this is the first time that a conditional diffusion model combining BiLSTM, XCA-Transformer and KL divergence regularization has been applied to the task of ST data imputation. SpaLSTF can fully capture the latent temporal dependencies among cells, accurately capture the hidden gene-level regulatory mechanisms, and improve the accuracy of imputation. We compared the performance of SpaLSTF with seven methods (one shallow learning algorithm and six state-of-the-art deep learning (DL)-based methods) on twelve datasets (see Table 1) from multiple platforms and organizations. Experimental results demonstrated that SpaLSTF not only improves gene imputation accuracy but also better preserves the topological structure among ST cell populations, offering a novel and effective solution for high-precision ST data enhancement.

Results

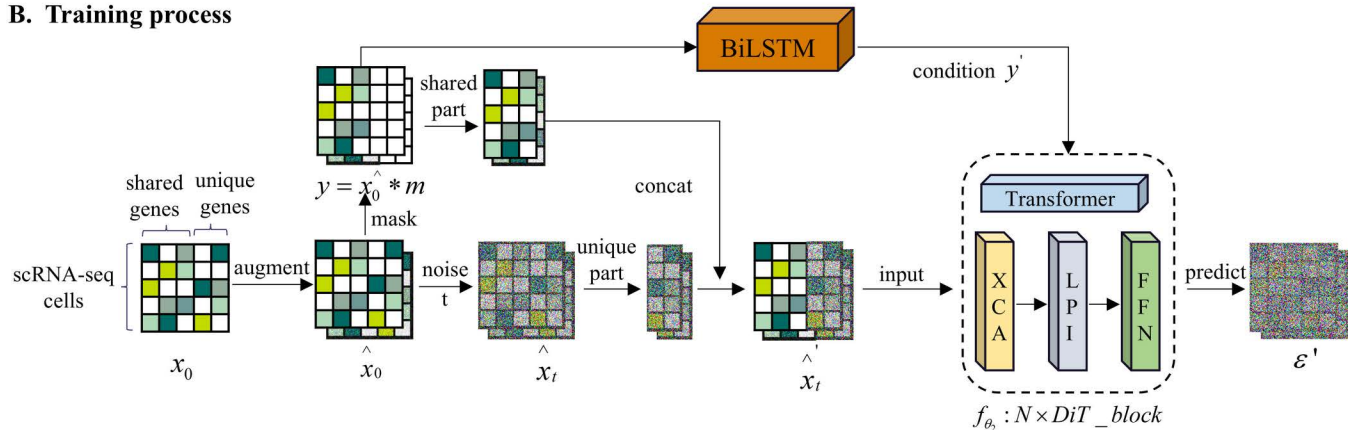
Ablation experiments

To evaluate the contribution of each key component in SpaLSTF, we conducted a comprehensive ablation study including both removal-based and retention-based variants. Specifically, we constructed two conventional ablation models: w/o BiLSTM and w/o XCA. The w/o BiLSTM variant removes the BiLSTM module from SpaLSTF, while the w/o XCA variant

A. Framework of DDPM



B. Training process



C. Inference process

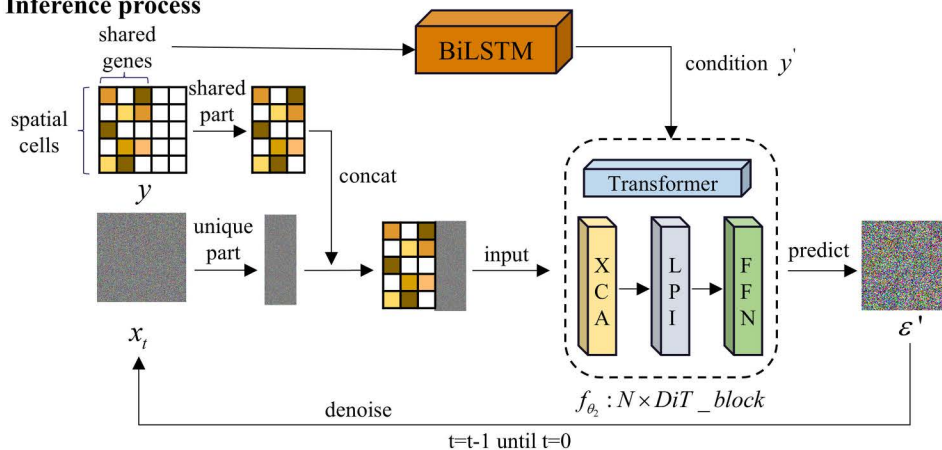


Fig 1. Overview of the SpaLSTF Framework. (A) Simplified structure of DDPM. (B) Training process of SpaLSTF on scRNA-seq data. (C) Inference process of SpaLSTF on ST data.

<https://doi.org/10.1371/journal.pcbi.1013954.g001>

replaces the XCA mechanism with a standard self-attention module. In addition, we introduced two complementary control variants that retain only the diffusion backbone with a single auxiliary component, namely Diffusion + BiLSTM only and Diffusion + XCA only.

We evaluated the performance of all methods on twelve ST datasets using two representative metrics: SPCC for gene expression similarity and ARI for clustering accuracy. As shown in Fig 2, removing the BiLSTM module leads to a clear performance decline in both metrics, highlighting its importance in modeling cell-level contextual dependencies. A more pronounced degradation is observed when the XCA module is removed, underscoring its critical role in capturing gene-level dependency structures and mitigating noise within the denoising network. Notably, the two “only” variants further exhibit lower SPCC and ARI values compared with their corresponding w/o counterparts, indicating that the absence of KL regularization negatively affects training stability and overall performance.

Table 1. Summary of the twelve validation dataset pairs.

Data pair	Tissue	Spatial data			scRNA-seq data		
		Cell num	Gene num	Reference	Cell num	Gene num	Reference
Dataset1_oscFISH	Somatosensory cortex	3405	33	[30]	14 249	34 041	[31]
Dataset2_seqFISH+	Cortex	524	10 000	[32]	14 249	34 041	[31]
Dataset3_MERFISH	Primary visual cortex	2399	268	[33]	14 249	34 041	[31]
Dataset4_ISS	Primary visual cortex	6000	119	[33]	14 249	34 041	[31]
Dataset5_FISH	Embryo	3039	84	[34]	1297	8924	[34]
Dataset6_BARISTAseg	Primary visual cortex	11 426	80	[35]	14 249	34 041	[31]
Dataset7_MERFISH	Mop	5551	247	[36]	14 249	34 041	[31]
Dataset8_seqFISH	Embryonic	175	45	[37]	9991	16 477	[38]
Dataset9_STARmap	Visual cortex	1549	1020	[39]	14 249	34 041	[31]
Dataset10_STARmap	Prefrontal cortex	1380	166	[39]	7737	14 837	[40]
Dataset11_ExSeq	Primary visual cortex	1154	42	[41]	14 249	34 041	[31]
Dataset12_MERFISH	Osteosarcoma	645	12 903	[42]	9234	19 098	[43]

<https://doi.org/10.1371/journal.pcbi.1013954.t001>

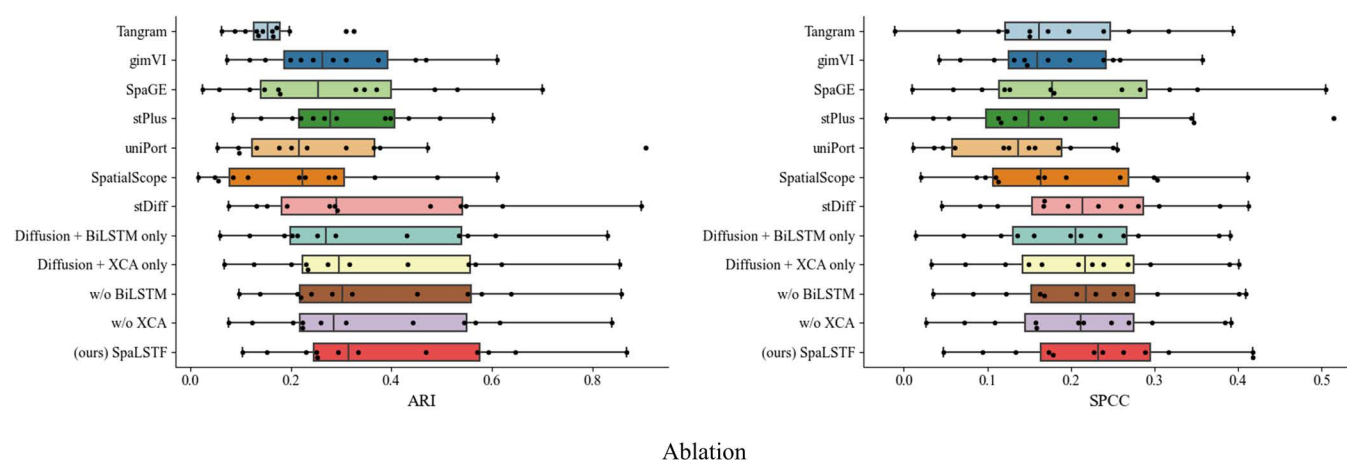


Fig 2. Boxplots showing the ablation experiments on all twelve datasets.

<https://doi.org/10.1371/journal.pcbi.1013954.g002>

Although the ablation variants still outperform most competing methods, the complete SpaLSTF model consistently achieves the best performance across all metrics, demonstrating that optimal performance is achieved only when BiLSTM, XCA, and KL regularization operate jointly. For all ablation experiments, we used the same training configuration as the full model, including batch size, learning rate, number of epochs, and optimizer settings, and only modified the corresponding target components.

Performance comparison of methods for preserving cellular topological structure

To evaluate the method's ability to preserve structural relationships among cells, we performed five-fold cross-validation on ST datasets without utilizing cell type labels. We then performed Leiden clustering [33] on the imputed data, used the clustering results of the original ST data as the reference standard, and evaluated the clustering performance using ARI, AMI, Homo, and NMI [33–35].

We compared the performance of SpaLSTF with Tangram [14], gimVI [15], SpaGE [17], stPlus [16], uniPort [18], SpatialScope [19], and stDiff [26] on twelve datasets (see Table 1) from multiple platforms and organizations. Comprehensive clustering evaluation of all methods across twelve datasets was presented in S1 Table. We selected six representative datasets (Dataset1_osmFISH, Dataset2_seqFISH+, Dataset3_MERFISH, Dataset4_ISS, Dataset5_FISH, Dataset6_BARISTAseg) from various sequencing platforms for detailed analysis, and performance evaluation was shown in Fig 3.

As shown in Fig 3, SpaLSTF outperforms all competing methods on four evaluation metrics. For instance, in Dataset1_osmFISH, SpaLSTF achieved scores of 0.251 (ARI) and 0.328 (AMI), exceeding the second-best method, stPlus, by 24.3% and 2.5%, respectively. In Dataset2_seqFISH+, SpaLSTF achieved Homo and NMI scores of 0.789 and 0.743, representing improvements of 4.6% and 4.8% over stDiff (the second-best method). These results indicate that the imputed data generated by SpaLSTF better preserve cellular neighborhood relationships, facilitating the identification of cellular populations. In contrast, Tangram performed worst across all datasets, with average scores of 0.235 and 0.135 for ARI and AMI, significantly lower than those of SpaLSTF. Additionally, gimVI, SpaGE, stPlus, uniPort, and SpatialScope exhibited inconsistent performance depending on the dataset. While some of these methods ranked second or third in certain datasets, their overall performance lacked stability. Notably, SpaLSTF consistently outperformed stDiff (another diffusion model-based method) across almost all datasets (see S1 Table). These results highlight SpaLSTF's superior ability in capturing spatial transcriptomic patterns while effectively preserving intercellular similarity relationships.

Comparison of imputation performance of methods at the gene-level

In this section, we evaluated the imputation performance of the method at the gene-level using five-fold cross-validation combined with four quantitative metrics (SPCC, SSIM, RMSE, and JS). We compared the performance of SpaLSTF with Tangram [14], gimVI [15], SpaGE [17], stPlus [16], uniPort [18], SpatialScope [19], and stDiff [26] on twelve datasets (see Table 1). We selected four representative datasets (Dataset7_MERFISH, Dataset8_seqFISH, Dataset9_STARmap, and

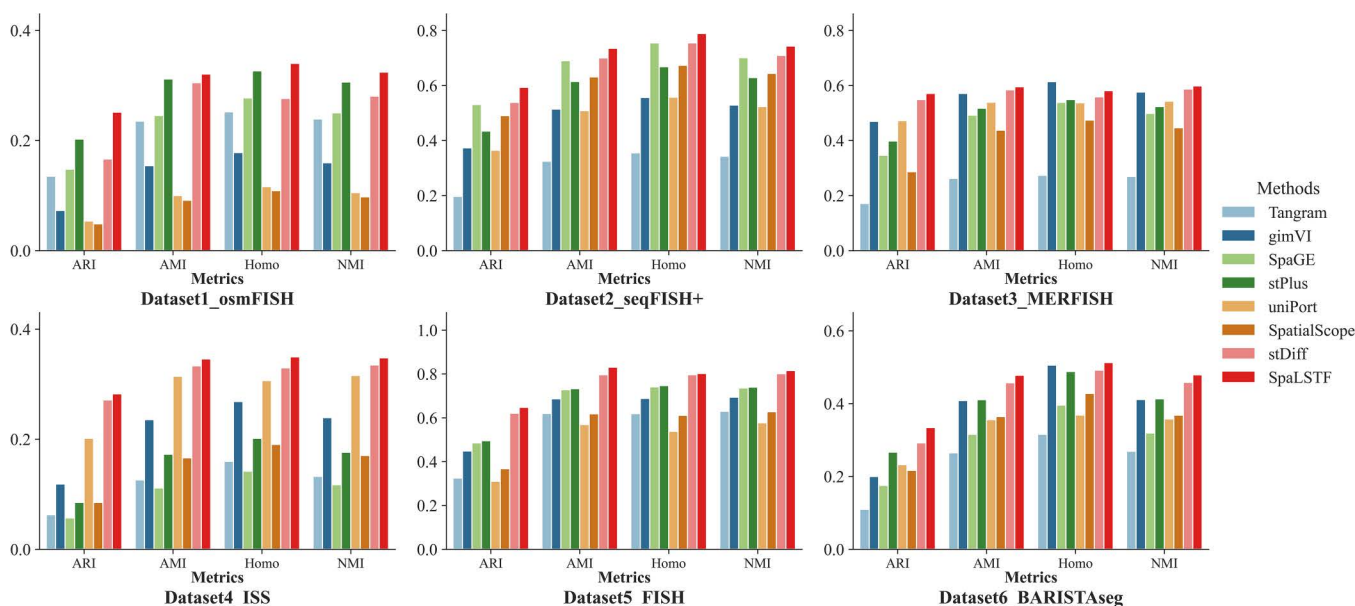


Fig 3. Clustering metrics (ARI, AMI, Homo, NMI) were used to assess the topological similarity between real ST data and imputed ST data across various ST platforms.

<https://doi.org/10.1371/journal.pcbi.1013954.g003>

Dataset10_STARmap) for detailed analysis (Fig 4), and the comparison results of the methods on the remaining eight datasets were shown in S1 Fig.

As shown in Fig 4, SpaLSTF achieves the highest similarity to ground truth across the four datasets. For example, in Dataset8_seqFISH, SpaLSTF achieved a median SPCC score of 0.134, surpassing the second-best method, stDiff, by 24%. Similarly, the median SSIM score of SpaLSTF is 0.122, representing a 15.1% improvement over stDiff. Furthermore, SpaLSTF outperformed most methods in terms of RMSE and JS, further demonstrating that its imputed gene expression data closely approximated real ST data.

It is noteworthy that among all methods, SpaLSTF not only achieves the best performance on gene-level metrics but also maintains the highest consistency in cellular topological structure (see S1 Table). This suggests that SpaLSTF is more effective in preserving the overall spatial topology while reconstructing the true gene expression pattern. However, despite their superior performance, all methods still exhibit limitations in gene-level imputation accuracy, indicating that there is still room for further improvement in aligning imputed data with real data. Moreover, in some datasets, even SpaLSTF shows relatively high RMSE values, highlighting the challenge of noise reduction in imputing gene expression data [36–38].

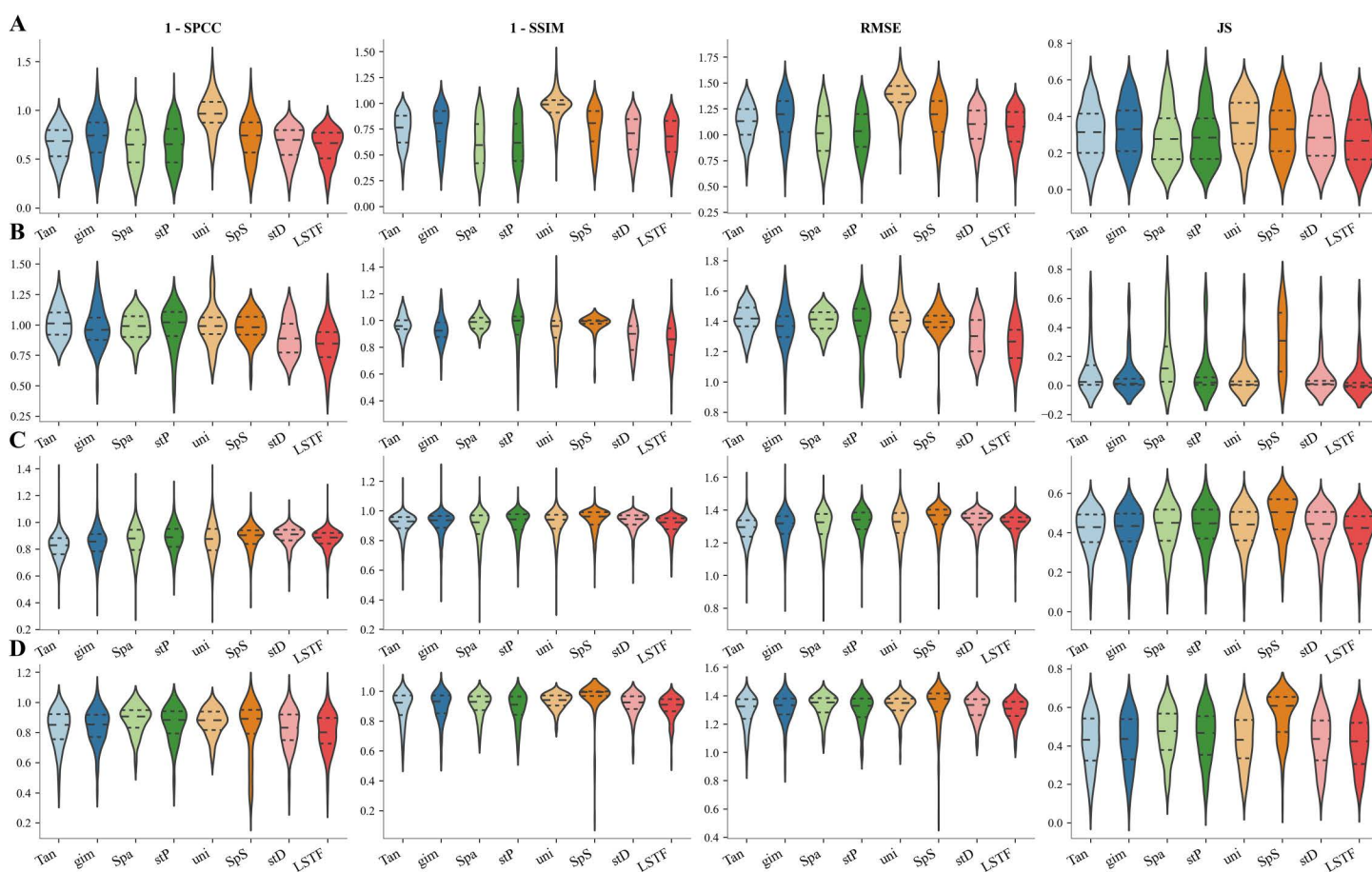


Fig 4. Evaluation metrics (1-SPCC, 1-SSIM, RMSE, JS) were used to quantify gene expression similarity between real ST data and predictions from Tangram (Tan), gimVI (gim), SpaGE (Spa), stPlus (stP), uniPort (uni), SpatialScope (SpS), stDiff (stD), and SpaLSTF (LSTF) across multiple ST platforms. (A)–(D) correspond to Dataset7_MERFISH, Dataset8_seqFISH, Dataset9_STARmap, and Dataset10_STARmap from Table 1.

<https://doi.org/10.1371/journal.pcbi.1013954.g004>

Comparison of alignment performance between imputed data and real ST data

To observe the alignment between imputed data and real ST data in a visualized form, we performed five-fold cross-validation and applied UMAP for dimensionality reduction to project scRNA-seq, real ST, and imputed ST data into a shared low-dimensional space. In the results of the imputation method, the distributions of imputed ST and real ST data should be close to each other while both are far from the scRNA-seq data. We compared the performance of SpaLSTF with Tangram, gimVI, SpaGE, stPlus, uniPort, SpatialScope, and stDiff on Dataset11_ExSeq and Dataset12_MERFISH.

As shown in Fig 5, the imputation results of Tangram, gimVI, stPlus, SpaGE, uniPort, and SpatialScope show significant differences from the real ST data, while the imputation results (in orange) of the diffusion-based methods SpaLSTF and stDiff closely approximate the real ST data (in green). For example, in Dataset11_ExSeq (Fig 5A), the distributions of imputation results from Tangram, SpaGE, stPlus, and uniPort were more consistent with the scRNA-seq data than with the real ST data, suggesting that these methods retain batch effects from scRNA-seq in their imputed ST data. In Dataset12_MERFISH (Fig 5B), compared to 9,234 scRNA-seq cells, the number of ST cells is only 645. Despite this, SpaLSTF and stDiff still provided stable and accurate imputation of ST data, while other methods exhibited substantial deviation from real ST data. Overall, SpaLSTF outperformed the other methods in aligning imputed data with the real ST data.

The imputation strategy of SpaLSTF is fundamentally different from that of other methods. SpaLSTF utilizes a diffusion process combined with BiLSTM and XCA-Transformer modules to model gene expression dependencies in scRNA-seq data. As a result, the imputed ST data is more consistent with the real ST data. In contrast, other methods typically rely on cellular similarity between scRNA-seq and ST data. These strategies include averaging gene expression from the k -nearest scRNA-seq neighbors, reconstructing ST data using decoders trained on scRNA-seq data, or drawing samples from the scRNA-seq distribution guided by ST data with batch effects eliminated. These methods essentially embed ST data into the batch space of scRNA-seq data, enhancing expression patterns within that space. However, due to the inherent batch differences between ST and scRNA-seq data, such mapping often hinders accurate recovery of the real ST data, resulting in noticeable discrepancies.

SpaLSTF enables accurate reconstruction of gene expression while preserving well-defined spatial patterns

In addition to quantitatively assessing the similarity between real and imputed ST gene expression, we also evaluated the consistency of spatial patterns by visualization. We selected four genes that exhibited well-defined patterns from the Dataset5_FISH embryonic tissue dataset. We compared the performance of SpaLSTF with Tangram, gimVI, SpaGE, stPlus, uniPort, SpatialScope, and stDiff.

As shown in Fig 6, in the ground truth, the lower half of the region of the *sna* gene shows a horizontal spatial pattern, which is accurately captured by SpaLSTF, SpaGE, stPlus, uniPort, and stDiff. These methods effectively delineated the expression boundary in the lower region, with SpaLSTF and stDiff show superior precision in both the left and right regions. In contrast, Tangram and gimVI exhibited a more disordered expression distribution between the high and low expression areas, while SpatialScope excessively constrained the high expression region.

For the *trn* gene following the vertical spatial pattern, SpaLSTF successfully reconstructed the vertical boundaries and captured the contours with high precision. In contrast, other methods produced blurrier boundaries. stPlus, SpaGE, and uniPort consistently overestimated expression levels across the entire spatial domain, while Tangram, gimVI, and SpatialScope generally underestimated overall gene expression levels. The results of stDiff showed some blurring and discontinuities on some vertical contour lines.

For the *tkv* gene exhibiting a complex spatial pattern, SpaLSTF's predictions closely matched the intricate expression pattern, particularly in the high-expression region in the upper-left corner. In contrast, the results from other methods showed significant discrepancies from the true spatial pattern.

For the *Antp* gene, which exhibits a sharply localized expression in a narrow central region, SpaLSTF successfully reproduced the true spatial pattern. In contrast, all other methods, except SpatialScope, significantly overestimated the

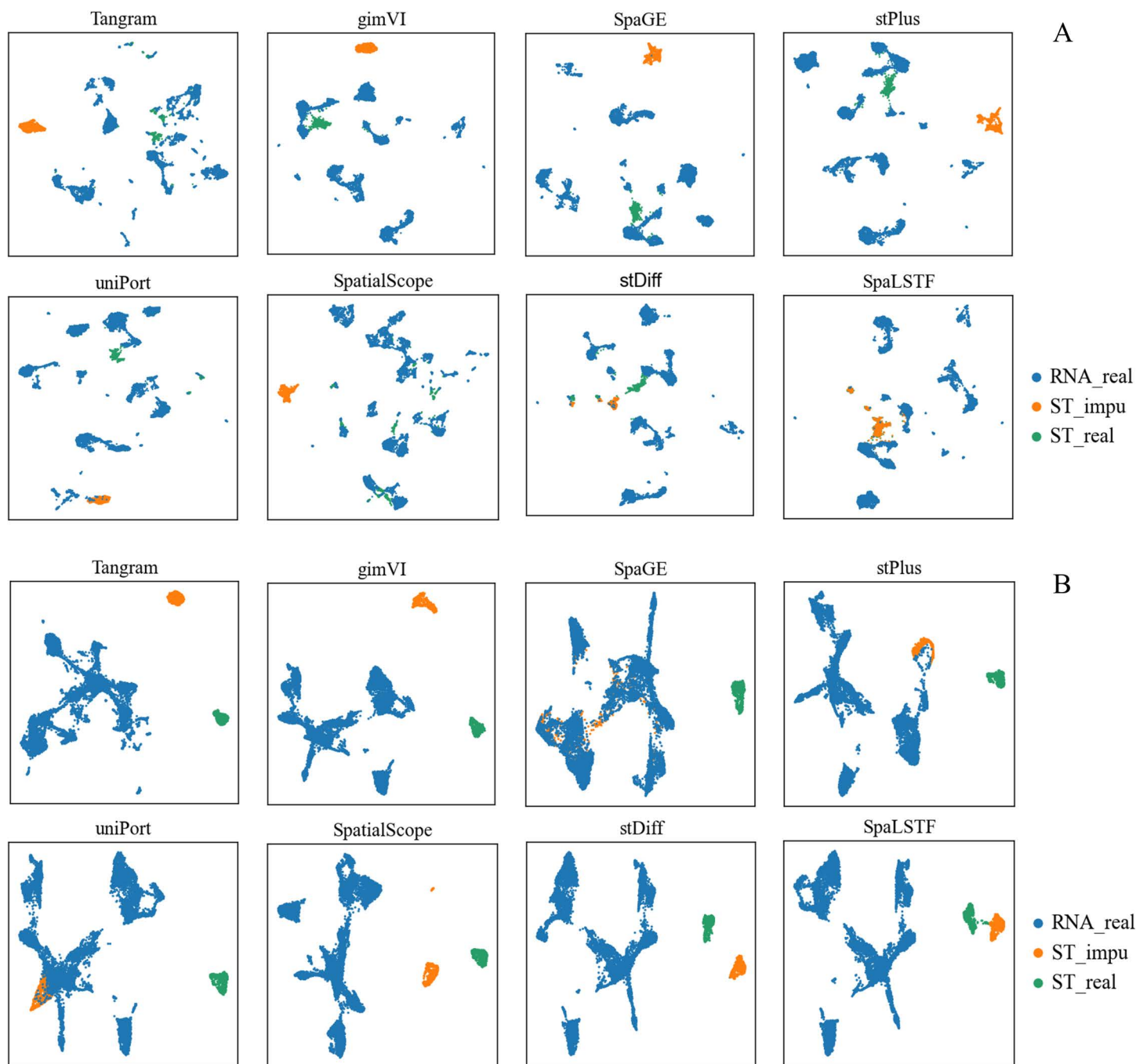


Fig 5. UMAP visualizations of scRNA-seq, true ST data, and imputed ST data produced by SpaLSTF and other competing methods. (A) and (B) correspond to Dataset11_ExSeq and Dataset12_MERFISH from Table 1, respectively.

<https://doi.org/10.1371/journal.pcbi.1013954.g005>

spatial distribution of the *Antp* gene. Although the expression boundaries generated by SpatialScope are close to the true spatial pattern, its predictions lack the clarity observed in the real data. Furthermore, SpatialScope consistently underestimated the expression levels of marker genes across all four genes (*sna*, *trn*, *tkv*, *Antp*).

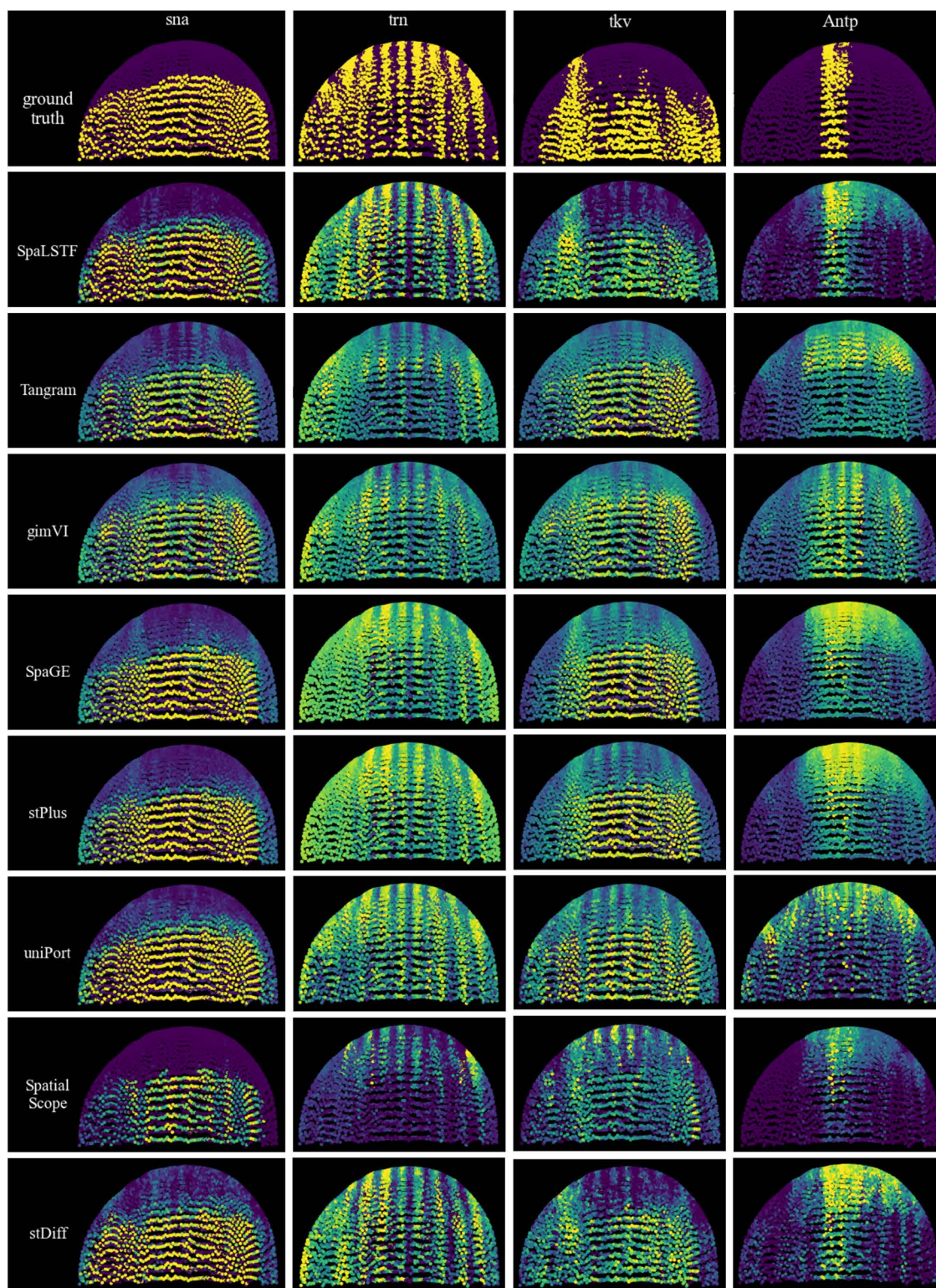


Fig 6. Spatial expression predictions of genes with known patterns in Dataset5_FISH. Each column represents a gene with distinct spatial features. The top row shows the ground truth, while the following rows display predictions from SpaLSTF and competing methods.

<https://doi.org/10.1371/journal.pcbi.1013954.g006>

Comparison of overall performance of methods on twelve datasets

In section ‘Performance comparison of methods for preserving cellular topological structure’ and section ‘Comparison of imputation performance of methods at the gene-level’, we benchmarked the performance of eight imputation methods across twelve datasets using four clustering (ARI, AMI, NMI, and Homo) and four similarity metrics (SPCC, SSIM, RMSE, and JS). In this section, we introduced an aggregated AS [39] composite index for a comprehensive comparison.

We used AS to combine the results of the four clustering metrics to evaluate the consistency of the cellular topological structure between the imputed data and the real data. As shown in Fig 7A, SpaLSTF achieves the best overall clustering performance, with its median AS index exceeding the upper quartile of all competing methods. Following closely behind are gimVI, stPlus, and stDiff, with stDiff showing relatively stable results. However, Tangram performed poorly, ranking last in most datasets.

We used AS to combine the results of the four similarity metrics to evaluate gene-level similarity between imputed and real data. As shown in Fig 7B, SpaLSTF outperforms competing methods. Fig 7C presents a summary of the integrated evaluation results of both cell clustering and gene-level similarity, and SpaLSTF performs the best and is the most robust, ranking first in both metrics. Although Tangram performed relatively well in gene-level similarity, its clustering accuracy was substantially lower, suggesting a potential disruption of cellular structural information and limited use in identifying cell populations. gimVI, SpaGE, and stPlus showed comparable overall performance, with stPlus favoring clustering accuracy, while gimVI and SpaGE prioritized gene-level similarity but showed reduced consistency in cell clustering. stDiff ranked second overall, offering a balanced and competitive performance in both metrics. SpatialScope performed poorly on both metrics, resulting in the lowest composite index among all methods.

Biological pathway consistency analysis across cluster pairs

To further investigate the biological interpretability of SpaLSTF, we conducted a pathway-level consistency analysis on Dataset2_seqFISH+ to evaluate whether the imputed gene expression profiles preserve biologically meaningful functional signals beyond gene-wise accuracy.

Specifically, we performed differential expression analysis between multiple pairs of cell clusters identified by Leiden clustering in the ground-truth spatial transcriptomics data. For each cluster pair, the top differentially expressed genes were subjected to Gene Ontology (GO) biological process enrichment analysis. The resulting enriched pathways derived from the imputed data were then compared with those obtained from the ground truth.

Two complementary metrics were used to quantify biological consistency: (1) the overlap ratio of the top enriched pathways between the imputed data and the ground truth, and (2) the Spearman correlation of enrichment scores across

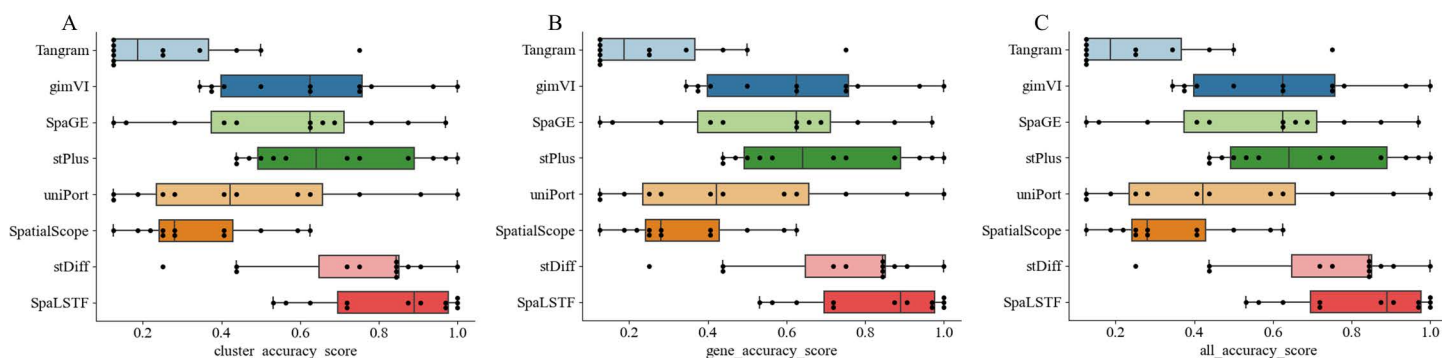


Fig 7. AS scores across twelve datasets for all methods. (A) Clustering metrics. (B) Gene similarity metrics. (C) Overall AS combining all metrics.

<https://doi.org/10.1371/journal.pcbi.1013954.g007>

commonly enriched pathways. These metrics respectively reflect the agreement in functional categories and the consistency of pathway importance ranking. As shown in Fig 8, SpaLSTF consistently achieved the highest pathway overlap ratio and enrichment score correlation among all compared methods when averaged across multiple cluster pairs (mean \pm standard deviation). In contrast, baseline methods exhibited lower overlap and more variable enrichment correlations, indicating a reduced ability to preserve coherent biological programs across cellular states.

These results demonstrate that SpaLSTF not only improves gene expression imputation accuracy, but also better maintains biologically interpretable functional structures at the pathway level. This suggests that the integration of diffusion-based denoising with BiLSTM and XCA modules enables SpaLSTF to capture gene regulatory patterns that are more consistent with underlying biological processes.

Discussion

In this study, we proposed SpaLSTF, a novel conditional diffusion-based model to impute the missing gene expression in ST data. By incorporating BiLSTM network, XCA-Transformer, and KL divergence regularization, SpaLSTF can fully capture the latent temporal dependencies among cells, accurately capture the hidden gene-level regulatory mechanisms, and improve the accuracy of imputation. We compared the performance of SpaLSTF with Tangram, gimVI, SpaGE, stPlus, uniPort, SpatialScope, and stDiff on twelve cross-platform datasets covering different tissues and organs using nine evaluation metrics (SPCC, SSIM, RMSE, JS, ARI, AMI, NMI, Homo, and AS). Experimental results demonstrated that SpaLSTF consistently outperforms competing methods in terms of gene expression accuracy and cell topological structure preservation. This indicates that SpaLSTF is a powerful solution for ST data imputation, capable of accurately enhancing ST data, particularly in cases of sparse data and limited gene coverage.

One limitation of SpaLSTF is that attention-based components such as XCA require larger attention matrices as the number of genes increases, which may lead to higher memory consumption; however, this overhead remains manageable under the commonly used GPU resources in our experimental settings.

Histological images and protein expression information can provide richer biological context information for the model, thereby enhancing the model's ability to distinguish cellular identities and improving the accuracy and robustness of gene expression prediction. By developing a multimodal integration framework, SpaLSTF is expected to enable more efficient ST imputation in challenging scenarios with low gene coverage or high technical noise. From a modeling perspective, the diffusion process mitigates batch effects by gradually perturbing input data and learning to reconstruct signals from noise, which reduces sensitivity to platform-specific technical variations and emphasizes shared distributional characteristics across datasets.

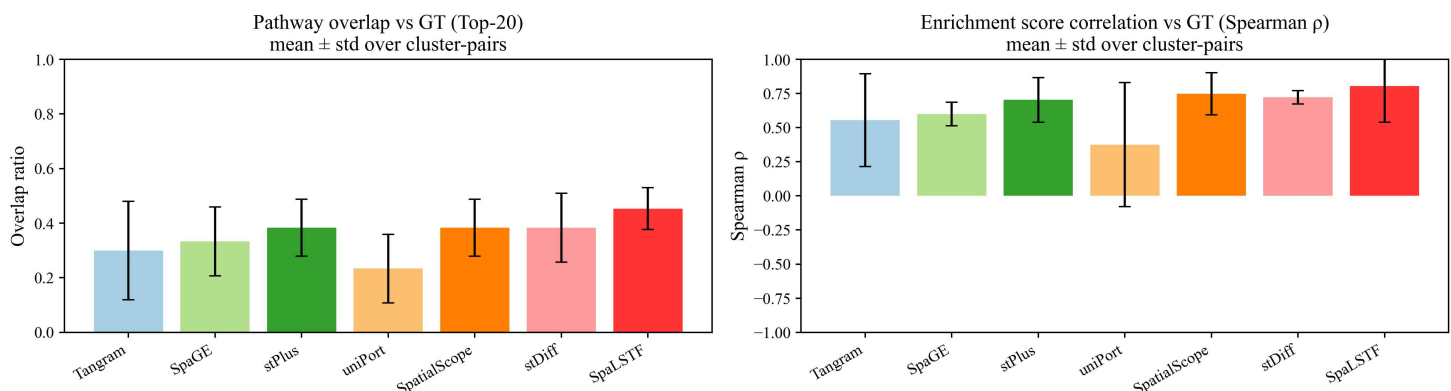


Fig 8. Pathway-level biological consistency analysis across cluster pairs.

<https://doi.org/10.1371/journal.pcbi.1013954.g008>

Materials and methods

Data preparation

We evaluated the performance of SpaLSTF on twelve paired ST and scRNA-seq datasets [39–52]. These ST datasets come from a wide variety of experimental protocols, cover different tissues and organs, and contain different numbers of genes and cells. The details and sources of these datasets are listed in Table 1. Notably, the first dataset contains pre-defined cell type labels, while the remaining ST datasets have no such annotations.

Experiment settings

For the diffusion process, we set the training epoch to 1200 and validation step to 1500 for all datasets. When the number of genes exceeds 500, we set the batch size to 512 and the hidden size to 1024 to provide sufficient representational capacity for modeling high-dimensional gene expression patterns; otherwise, we use a batch size of 2048 and a hidden size of 512 for improved computational efficiency. For the BiLSTM module, we set the hidden_size to 128 and number of layers to 2. For the XCA-Transformer module, the number of multi-head cross-covariance attention layers is set to 6 with 16 attention heads. The model is optimized with the AdamW [52] optimizer with a learning rate of 1e-4. All results were obtained on a NVIDIA RTX 4090 GPU with 128GB of memory. We executed the competing methods using their default parameters.

The SpaLSTF framework

SpaLSTF is formulated as a denoising diffusion probabilistic model (DDPM) [24], including two coupled Markov chains: a forward diffusion process and a reverse denoising process. The overall framework is illustrated in Fig 1.

The DDPM framework

As shown in Fig 1A, the forward process incrementally adds Gaussian noise to the input gene expression matrix x_0 according to the given distribution $q(x_t|x_{t-1})$, which is a predefined forward diffusion distribution and is fixed during training. The process is carried out over time steps t , gradually transforming the data into a standard Gaussian distribution. In contrast, the reverse diffusion process gradually reconstructs the original data using the learned denoising conditional distribution $p_\theta(x_{t-1}|x_t)$. Starting from pure noise $x_T \sim N(0, I)$, the model performs iterative denoising to reconstruct a final gene expression matrix \hat{x}_0 in $c \times g$, c and g denote the number of cells and genes, respectively.

Training procedure

Fig 1B outlines the training process of SpaLSTF, where the model learns the complex interactions among gene expression levels from reference scRNA-seq data through both noise addition and denoising. First, we perturb the reference scRNA-seq data x_0 by injecting stochastic noise to address the batch effect discrepancies between ST and scRNA-seq modalities, while adaptively aligning their distributions during training to enhance SpaLSTF's robustness. This strategy not only diversifies the training dataset but also maintains intrinsic gene-gene relationships, ensuring that the model emphasizes the gene-gene interactions rather than the absolute expression values. The resulting perturbed dataset, denoted as \hat{x}_0 , is then used for training.

A time index $t \sim U(1, T)$ is sampled in each training iteration, and Gaussian noise ε is added to \hat{x}_0 at step t , resulting in the corrupted input \hat{x}_t , as described in (1).

$$\hat{x}_t = \sqrt{\gamma_t} \hat{x}_0 + \sqrt{1 - \gamma_t} \varepsilon, \quad \varepsilon \sim N(0, I), \quad (1)$$

where γ_t is computed as $\prod_{i=1}^t \alpha_i$. The hyperparameters $\alpha_{i:T}$ are defined using the cosine function, ensuring that $0 < \alpha_t < 1$ for all t . At each iteration, these parameters control the noise's mean and variance.

Subsequently, the unique gene component from \hat{x}_t and the shared gene component from \hat{x}_0 are specifically isolated and merged to construct $\hat{x}_t = \hat{x}_0 * m + \hat{x}_t * (1 - m)$. 'unique gene component' represents gene expressions observed only in the scRNA-seq dataset, whereas 'shared gene component' represent gene expressions observed in both modalities. $m \in \{0, 1\}^{c \times g}$ is a binary matrix where entries corresponding to shared genes are assigned value 1, while entries corresponding to unique genes are assigned the value 0. Operator $*$ represents element-wise multiplication. A BiLSTM module is then used to capture the relationships in the masked conditional cell state data $y' = \text{BiLSTM}(\hat{x}_0 * m)$. Finally, \hat{x}_t and y' are fed to the XCA-Transformer denoising module to estimate the added noise $\varepsilon' = \text{XCA_Transformer}(\hat{x}_t, y', t)$ at time-step t .

During training, the loss is calculated based on the noise associated with the masked unique gene components:

$$\text{loss} = \text{MSE}(\varepsilon * (1 - m), \varepsilon' * (1 - m)) + \alpha \cdot D_{\text{KL}}(q(\varepsilon' * (1 - m)) \parallel p(\varepsilon * (1 - m))) \quad (2)$$

where $q(\cdot)$ and $p(\cdot)$ represent the predicted noise distribution and the real noise distribution, respectively. The weighting coefficient α is set according to the dimensionality of the predicted gene space: $\alpha = 2.8$ for low-dimensional gene prediction tasks and $\alpha = 0.02$ for high-dimensional settings. KL divergence is incorporated as a regularization term in combination with the MSE loss to ensure both the accuracy of predicted noise values and the consistency between generated noise and the target distribution.

Inference procedure

In the inference phase (Fig 1C), SpaLSTF leverages the functional mapping learned during training to reconstruct whole-transcriptome ST data. Initially, ST data is augmented to form a conditional vector y , wherein the unique gene component is zero-filled. At time $t = T$, stochastic noise vector x_t is generated. The shared gene component from y provides conditional signal that guide the reverse diffusion process, which also passes through the BiLSTM module and is then concatenated with the unique gene components of x_t to construct the complete input $x'_t = y * m + x_t * (1 - m)$. This composite input x'_T , along with y' generated by BiLSTM module, is then processed by the pre-trained transformer denoising network to estimate the noise at time T . The estimated noise is subsequently utilized to generate the state x'_{T-1} at time $t = T - 1$ according to equation (3). The reverse diffusion process is iteratively performed over T steps, culminating in the final prediction at $t = 0$:

$$x_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} \text{transformer}(x'_t, y', t) \right) + \sqrt{1 - \alpha_t} \varepsilon_t \quad (3)$$

where $\varepsilon_t \sim N(0, I)$.

BiLSTM module

BiLSTM simultaneously processes information in both forward and backward directions, enabling the model to capture global context from both ends of the cell sequence data, it captures bidirectional contextual dependencies among cellular states in the conditional feature space, where transcriptionally similar cells tend to exhibit coherent functional and latent structural relationships. SpaLSTF leverages the memory mechanism of this module to effectively capture long-distance dependencies among cells in the conditional data y , thereby more accurately reflecting the complex interactions between cells. As a result, richer and more accurate conditional information y' is generated to guide the process. LSTM units control the flow of information through three gates (input, forget and output), and their calculation formulas are described in equation (4):

$$\begin{cases} i_t = \sigma(W_{xi} y_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \\ f_t = \sigma(W_{xf} y_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc} y_t + W_{hc} h_{t-1} + b_c) \\ o_t = \sigma(W_{xo} y_t + W_{ho} h_{t-1} + W_{co} c_t + b_o) \\ h_t = o_t \odot \tanh(c_t) \end{cases} \quad (4)$$

where y_t is the current input vector, h_{t-1} and c_{t-1} are the hidden state and cell state at the previous time step, respectively. σ denotes sigmoid activation function, \tanh denotes hyperbolic tangent activation function, \odot denotes element-wise multiplication, while W and b are the corresponding weight matrices and bias vectors.

Specifically, BiLSTM processes the cell sequence data from both ends to obtain the forward hidden states and backward hidden states, and then concatenates them to produce a representation that integrates bidirectional information:

$$\begin{cases} \vec{h}_t = LSTM(y_t, \vec{h}_{t-1}), t = 1, 2, \dots, n \\ \overleftarrow{h}_t = LSTM(y_t, \overleftarrow{h}_{t+1}), t = n, n-1, \dots, 1 \\ h_t = [\vec{h}_t; \overleftarrow{h}_t] \end{cases} \quad (5)$$

Finally, the entire conditional sequence is encoded by BiLSTM to form the new conditional feature matrix $y' = \{h_1, h_2, \dots, h_n\}$.

XCA-Transformer module

The new conditional data y' and the encoded time t are fused to serve as the key and are input into the attention layer along with x_t . As shown in [equation \(6\)](#), we introduce a multi-head cross-covariance attention (XCA) mechanism into the diffusion Transformer model to replace the traditional self-attention mechanism, thereby focusing on the calculation of attention coefficients among gene expression abundances rather than among cells. This modification plays an important role in capturing the relationships among gene expression abundances and more accurately predicting the noise data ε , in x_t at time t .

$$\begin{cases} H(\varepsilon') = \text{concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_k) \times W^O \\ \text{XC_Attention}(Q, K, V) = V \cdot \text{softmax}\left(\frac{K^T Q}{\sqrt{d_k}}\right) \\ \text{head}_i = \text{XC_Attention}\left(\hat{x}_t^i W_i^Q, (y' + t) W_i^K, (y' + t) W_i^V\right) \end{cases} \quad (6)$$

where k denotes the number of heads and the weight matrix W^O aggregates attention heads, Q, K, V represent 'Query', 'Key' and 'Value', respectively. Additionally, W_i^Q, W_i^K , and W_i^V are weight matrices. The Attention Map, denoted as $\left(\frac{K^T Q}{\sqrt{d_k}}\right)$, has a shape of $g \times g$, storing attention weights contributed by Query and Key representing each gene. V represents the value component in the cross-covariance mechanism. Finally, the attention matrix of each attention head can be attained by multiplying weight coefficients calculated by the values of V and $\text{softmax}(\cdot)$.

The local patch interaction (LPI) block is introduced after each XCA block to enable explicit communication between patches. The feed-forward network (FFN) enables global feature interaction across all features.

Evaluation methods

- 1) *Gene Expression Prediction*: At the gene level, we use cross-validation combined with four evaluation metrics to evaluate the consistency between the predicted and ground-truth ST data. The four evaluation metrics are Spearman's

rank correlation coefficient [53] (SPCC), structural similarity index measure (SSIM), root mean square error (RMSE), and Jensen-Shannon divergence (JS). Higher SPCC and SSIM scores, or lower RMSE and JS values, reflect improved prediction performance.

$$SPCC(i) = 1 - \frac{6 \sum_{j=1}^n d_{ij}^2}{n(n^2 - 1)} \quad (7)$$

where i denotes the i -th gene, j denotes the j -th cell, and n denotes the cell count. d_{ij} denotes the difference between the predicted and true expression values of the i -th gene in the j -th cell.

$$SSIM(i) = \frac{(2\mu(T_i)\mu(P_i) + C_1^2)(2cov(T_i, P_i) + C_2^2)}{(\mu^2(T_i) + \mu^2(P_i) + C_1^2)(\sigma^2(T_i) + \sigma^2(P_i) + C_2^2)} \quad (8)$$

where P_i and T_i denote the vectors of the i -th gene in predicted values and ground truth, respectively. $\mu()$ denotes the average value and $\sigma()$ denotes the standard deviation calculation process.

$$RMSE(i) = \sqrt{\frac{1}{n} \sum_{j=1}^n (z'_{ij} - z_{ij})^2} \quad (9)$$

where z and z' denote the scores from the ground truth and predicted expressions, respectively.

$$JS = \frac{1}{2} \left[KL\left(\varphi_i(T) \middle| \frac{\varphi_i(T) + \varphi_i(P)}{2}\right) + KL\left(\varphi_i(P) \middle| \frac{\varphi_i(T) + \varphi_i(P)}{2}\right) \right]$$

$$\varphi_i(x) = \frac{x_{ij}}{\sum_{j=1}^n x_{ij}}, \quad KL(T_i | P_i) = \sum_{j=1}^n \left(T_{ij} \times \log \frac{T_{ij}}{P_{ij}} \right) \quad (10)$$

where $\varphi_i(x)$ denotes the distribution probability of gene i , and $KL()$ computes the KL divergence.

2) *Clustering Performance*: To assess spatial clustering performance [54–56], we use the adjusted rand index (ARI), adjusted mutual information (AMI), normalized mutual information (NMI) and homogeneity (Homo) to measure the consistency between imputation results and ground truth.

$$ARI = \frac{\sum_{ij} \binom{n_{ij}}{2} - \frac{\left[\sum_i \binom{\alpha_i}{2} \sum_j \binom{\beta_j}{2} \right]}{\binom{n}{2}}}{\frac{1}{2} \left[\sum_i \binom{\alpha_i}{2} + \sum_j \binom{\beta_j}{2} \right] - \frac{\sum_i \binom{\alpha_i}{2} \sum_j \binom{\beta_j}{2}}{\binom{n}{2}}} \quad (11)$$

where α_i and β_i denote the count of spots in the i -th cluster of predicted cluster P and the j -th cluster of true cluster T , respectively, n_{ij} denotes overlap between the i -th cluster of P and the j -th cluster of T .

$$NMI = \frac{MI(P_c, T_c)}{\sqrt{H(P_c) \cdot H(T_c)}} \quad (12)$$

$$AMI = \frac{MI(P_c, T_c) - E(MI(P_c, T_c))}{avg(H(P_c), H(T_c)) - E(MI(P_c, T_c))} \quad (13)$$

$$Homo = 1 - \frac{H(T_c|P_c)}{H(T_c)} \quad (14)$$

where $MI(P_c, T_c)$ denotes the mutual information between P_c and T_c , $H(P_c)$ and $H(T_c)$ denote the entropy of P_c and T_c , and $E(MI(P_c, T_c))$ denotes the expected mutual information under the stochastic model.

- 3) *Overall Accuracy Score*: We assess different imputation methods using various evaluation metrics from both gene and cell perspectives. In this section, we introduce the accuracy score (AS) [39] to calculate the overall accuracy score of each method. For each dataset and evaluation metric, we sort the methods in ascending order and assign corresponding ranks. AS refers to the average ranking across all datasets and metrics, with higher scores indicating better overall effectiveness.
- 4) *Comparison Methods*: We compared SpaLSTF with Tangram [14], gimVI [15], SpaGE [17], stPlus [16], uniPort [18], SpatialScope [19], and stDiff [26] to test the performance of SpaLSTF. These methods include one shallow learning algorithm (SpaGE) and six state-of-the-art DL-based methods (Tangram, gimVI, stPlus, uniPort, SpatialScope, and stDiff). Tangram (2021) used a deep learning framework to map multimodal single data on spatial support to predict spatial patterns. gimVI (2019) is a deep generative model for ST imputation. SpaGE (2020) is a ST imputation method that combines principal components analysis (PCA), singular value decomposition (SVD) and KNN. stPlus (2021) imputed ST data via autoencoders and weighted KNN. uniPort (2022) utilized a coupled variational autoencoder (coupled-VAE) and minibatch unbalanced optimal transport (Minibatch-UOT) to impute ST data. SpatialScope (2023) is a ST imputation method using deep generative models. stDiff (2024) is a diffusion-based imputation method.

Supporting information

S1 Table. Cluster results of all methods in all datasets.
(DOCX)

S1 Fig. Evaluation metrics (1-SPCC, 1-SSIM, RMSE, JS) were used to quantify gene expression similarity between real ST data and predictions from Tangram (Tan), gimVI (gim), SpaGE (Spa), stPlus (stP), uniPort (uni), SpatialScope (SpS), stDiff (stD), and SpaLSTF (LSTF) across multiple ST platforms. (E)–(L) correspond to Dataset1_osmFISH, Dataset2_seqFISH+, Dataset3_MERFISH, Dataset4_ISS, Dataset5_FISH, Dataset6_BARISTAseg, Dataset11_ExSeq and Dataset12_MERFISH from Table 1.
(TIFF)

Author contributions

Conceptualization: Lin Yuan, Yufeng Jiang, Boyuan Meng, Qingxiang Wang, Cuihong Wang, De-Shuang Huang.

Data curation: Lin Yuan, Yufeng Jiang, Cuihong Wang, De-Shuang Huang.

Funding acquisition: Lin Yuan, De-Shuang Huang.

Investigation: Lin Yuan, Yufeng Jiang, Cuihong Wang, De-Shuang Huang.

Methodology: Lin Yuan, Yufeng Jiang, Boyuan Meng, Qingxiang Wang, Cuihong Wang, De-Shuang Huang.

Project administration: Lin Yuan, Cuihong Wang, De-Shuang Huang.

Software: Lin Yuan, Yufeng Jiang, Boyuan Meng, Cuihong Wang.

Supervision: De-Shuang Huang.

Visualization: Lin Yuan, Yufeng Jiang, Boyuan Meng.

Writing – original draft: Lin Yuan, Yufeng Jiang, Boyuan Meng, Qingxiang Wang, Cuihong Wang, De-Shuang Huang.

Writing – review & editing: Lin Yuan, Yufeng Jiang, Boyuan Meng, Qingxiang Wang, Cuihong Wang, De-Shuang Huang.

References

- Wang J, Gao Q, Yuan S, Shang J. Cluster-Guided Contrastive Learning With Masked Autoencoder for Spatial Domain Identification Based on Spatial Transcriptomics. *IEEE J Biomed Health Inform.* 2026;30(2):1809–20. <https://doi.org/10.1109/JBHI.2025.3586483> PMID: [40622835](#)
- Jiang J, Liu Y, Qin J, Chen J, Wu J, Pizzi MP, et al. METI: deep profiling of tumor ecosystems by integrating cell morphology and spatial transcriptomics. *Nat Commun.* 2024;15(1):7312. <https://doi.org/10.1038/s41467-024-51708-9> PMID: [39181865](#)
- Li X, Zhu F, Min W. SpaDiT: diffusion transformer for spatial gene expression prediction using scRNA-seq. *Brief Bioinform.* 2024;25(6):bbae571. <https://doi.org/10.1093/bib/bbae571> PMID: [39508444](#)
- Zhang Z, Cui F, Su W, Dou L, Xu A, Cao C, et al. webSCST: an interactive web application for single-cell RNA-sequencing data and spatial transcriptomic data integration. *Bioinformatics.* 2022;38(13):3488–9. <https://doi.org/10.1093/bioinformatics/btac350> PMID: [35604082](#)
- Gong J, Xu K, Ma Z, Lu ZJ, Zhang QC. A deep learning method for recovering missing signals in transcriptome-wide RNA structure profiles from probing experiments. *Nat Mach Intell.* 2021;3(11):995–1006. <https://doi.org/10.1038/s42256-021-00412-0>
- Li X, Meng X, Chen H, Fu X, Wang P, Chen X, et al. Integration of single sample and population analysis for understanding immune evasion mechanisms of lung cancer. *NPJ Syst Biol Appl.* 2023;9(1):4. <https://doi.org/10.1038/s41540-023-00267-8> PMID: [36765073](#)
- Fletcher M. Improved tool for scRNA-seq analysis. *Nat Genet.* 2024;56(12):2589. <https://doi.org/10.1038/s41588-024-02040-x> PMID: [39653731](#)
- Hu D, Liang K, Dong Z, Wang J, Zhao Y, He K. Effective multi-modal clustering method via skip aggregation network for parallel scRNA-seq and scATAC-seq data. *Brief Bioinform.* 2024;25(2):bbae102. <https://doi.org/10.1093/bib/bbae102> PMID: [38493338](#)
- Katzenelenbogen Y, Sheban F, Yalin A, Yofe I, Svetlichnyy D, Jaitin DA, et al. Coupled scRNA-Seq and Intracellular Protein Activity Reveal an Immunosuppressive Role of TREM2 in Cancer. *Cell.* 2020;182(4):872–885.e19. <https://doi.org/10.1016/j.cell.2020.06.032> PMID: [32783915](#)
- Li B, Tang Z, Budhkar A, Liu X, Zhang T, Yang B, et al. SpaIM: Single-cell spatial transcriptomics imputation via style transfer. *bioRxiv.* 2025.
- Zhao Y, Wang K, Hu G. DIST: spatial transcriptomics enhancement using deep learning. *Brief Bioinform.* 2023;24(2):bbad013. <https://doi.org/10.1093/bib/bbad013> PMID: [36653906](#)
- Zou G, Shen Q, Li L, Zhang S. stAI: a deep learning-based model for missing gene imputation and cell-type annotation of spatial transcriptomics. *Nucleic Acids Res.* 2025;53(5):gkaf158. <https://doi.org/10.1093/nar/gkaf158> PMID: [40057378](#)
- Wang Z, Geng A, Duan H, Cui F, Zou Q, Zhang Z. A comprehensive review of approaches for spatial domain recognition of spatial transcriptomes. *Brief Funct Genomics.* 2024;23(6):702–12. <https://doi.org/10.1093/bfpg/elae040> PMID: [39426802](#)
- Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods.* 2021;18(11):1352–62. <https://doi.org/10.1038/s41592-021-01264-7> PMID: [34711971](#)
- Lopez R, Nazaret A, Langevin M, Samaran J, Regier J, Jordan MI. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv preprint.* 2019. <https://doi.org/10.48550/arXiv.1905.02269>
- Shengquan C, Boheng Z, Xiaoyang C, Xuegong Z, Rui J. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics.* 2021;37(Suppl_1):i299–307. <https://doi.org/10.1093/bioinformatics/btab298> PMID: [34252941](#)
- Abdelaal T, Mourragui S, Mahfouz A, Reinders MJT. SpaGE: Spatial Gene Enhancement using scRNA-seq. *Nucleic Acids Res.* 2020;48(18):e107. <https://doi.org/10.1093/nar/gkaa740> PMID: [32955565](#)
- Cao K, Gong Q, Hong Y, Wan L. A unified computational framework for single-cell data integration with optimal transport. *Nat Commun.* 2022;13(1):7419. <https://doi.org/10.1038/s41467-022-35094-8> PMID: [36456571](#)
- Wan X, Xiao J, Tam SST, Cai M, Sugimura R, Wang Y, et al. Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat Commun.* 2023;14(1):7848. <https://doi.org/10.1038/s41467-023-43629-w> PMID: [38030617](#)
- Yan C, Zhu Y, Chen M, Yang K, Cui F, Zou Q, et al. Integration tools for scRNA-seq data and spatial transcriptomics sequencing data. *Brief Funct Genomics.* 2024;23(4):295–302. <https://doi.org/10.1093/bfpg/elae002> PMID: [38267084](#)
- Hu Y, Zhao Y, Schunk CT, Ma Y, Derr T, Zhou XM. ADEPT: Autoencoder with differentially expressed genes and imputation for robust spatial transcriptomics clustering. *iScience.* 2023;26(6):106792. <https://doi.org/10.1016/j.isci.2023.106792> PMID: [37235055](#)
- Yang Y, Hu L, Li G, Li D, Hu P, Luo X. Link-Based Attributed Graph Clustering via Approximate Generative Bayesian Learning. *IEEE Trans Syst Man Cybern, Syst.* 2025;55(8):5730–43. <https://doi.org/10.1109/tsmc.2025.3572738>
- Su X, Hu P, Li D, Zhao B, Niu Z, Herget T, et al. Interpretable identification of cancer genes across biological networks via transformer-powered graph representation learning. *Nat Biomed Eng.* 2025;9(3):371–89. <https://doi.org/10.1038/s41551-024-01312-5> PMID: [39789329](#)

24. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*. 2020;33:6840–51.
25. Dhariwal P, Nichol A. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*. 2021;34:8780–94.
26. Li K, Li J, Tao Y, Wang F. stDiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Brief Bioinform*. 2024;25(3):bbae171. <https://doi.org/10.1093/bib/bbae171> PMID: 38628114
27. Sun Y, Kong L, Huang J, Deng H, Bian X, Li X, et al. A comprehensive survey of dimensionality reduction and clustering methods for single-cell and spatial transcriptomics data. *Brief Funct Genomics*. 2024;23(6):733–44. <https://doi.org/10.1093/bfpg/ela023> PMID: 38860675
28. Siarni-Namini S, Tavakoli N, Namin AS. The performance of LSTM and BiLSTM in forecasting time series. In *IEEE International conference on big data (Big Data)*; 2019: IEEE.
29. Cui J, Qi X, Tian Z, Yu B, Zhang H, Zhong Z. Decoupled Kullback-Leibler Divergence Loss. In: *Advances in Neural Information Processing Systems 37*, 2024. 74461–86. <https://doi.org/10.52202/079017-2370>
30. El-Nouby A, Touvron H, Caron M, Bojanowski P, Douze M, Joulin A. Xcit: Cross-covariance image transformers. *arXiv preprint*. 2021. <https://doi.org/2106.09681>
31. Ma Z, Xie J, Lai Y, Taghia J, Xue J-H, Guo J. Insights Into Multiple/Single Lower Bound Approximation for Extended Variational Inference in Non-Gaussian Structured Data Modeling. *IEEE Trans Neural Netw Learn Syst*. 2020;31(7):2240–54. <https://doi.org/10.1109/TNNLS.2019.2899613> PMID: 30908264
32. Duan H, Zhang Q, Cui F, Zou Q, Zhang Z. MVST: Identifying spatial domains of spatial transcriptomes from multiple views using multi-view graph convolutional networks. *PLoS Comput Biol*. 2024;20(9):e1012409. <https://doi.org/10.1371/journal.pcbi.1012409> PMID: 39235988
33. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9(1):5233. <https://doi.org/10.1038/s41598-019-41695-z> PMID: 30914743
34. Ye J, Yu Y, Lu L, Wang H, Zheng Y, Liu Y, et al. DEP-Former: Multimodal Depression Recognition Based on Facial Expressions and Audio Features via Emotional Changes. *IEEE Trans Circuits Syst Video Technol*. 2025;35(3):2087–100. <https://doi.org/10.1109/tcsvt.2024.3491098>
35. Meng Y, Wang Y, Guo C, Tang X, Zhang Z, Cui F, et al. ST-GCP: a graph convolutional network model with contrastive consistency and permutation for spatial transcriptomics. *Brief Bioinform*. 2025;26(6):bbaf643. <https://doi.org/10.1093/bib/bbaf643> PMID: 41348600
36. Zha Y, Feng S, Gao P, Zou Q, Ma X. Enhancing and accelerating cell type deconvolution of large-scale spatial transcriptomics slices with dual network model. *Bioinformatics*. 2025;41(8):btaf419. <https://doi.org/10.1093/bioinformatics/btaf419> PMID: 40704686
37. Li D, Yang Y, Cui Z, Yin H, Hu P, Hu L. LLM-DDI: Leveraging Large Language Models for Drug-Drug Interaction Prediction on Biomedical Knowledge Graph. *IEEE J Biomed Health Inform*. 2026;30(1):773–81. <https://doi.org/10.1109/JBHI.2025.3585290> PMID: 40601466
38. Yang Y, Hu L, Li G, Li D, Hu P, Luo X. FMvPCI: A Multiview Fusion Neural Network for Identifying Protein Complex via Fuzzy Clustering. *IEEE Trans Syst Man Cybern, Syst*. 2025;55(9):6189–202. <https://doi.org/10.1109/tsmc.2025.3578348>
39. Li B, Zhang W, Guo C, Xu H, Li L, Fang M, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat Methods*. 2022;19(6):662–70. <https://doi.org/10.1038/s41592-022-01480-9> PMID: 35577954
40. Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunteren JA, Svensson CI, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods*. 2018;15(11):932–5. <https://doi.org/10.1038/s41592-018-0175-z> PMID: 30377364
41. Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*. 2018;563(7729):72–8. <https://doi.org/10.1038/s41586-018-0654-5> PMID: 30382198
42. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulana N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature*. 2019;568(7751):235–9. <https://doi.org/10.1038/s41586-019-1049-y> PMID: 30911168
43. Karaïskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, et al. The Drosophila embryo at single-cell transcriptome resolution. *Science*. 2017;358(6360):194–9. <https://doi.org/10.1126/science.aan3235> PMID: 28860209
44. Chen X, Sun Y-C, Church GM, Lee JH, Zador AM. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res*. 2018;46(4):e22. <https://doi.org/10.1093/nar/gkx1206> PMID: 29190363
45. Zhang M, Eichhorn SW, Zingg B, Yao Z, Cotter K, Zeng H, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature*. 2021;598(7879):137–43. <https://doi.org/10.1038/s41586-021-03705-x> PMID: 34616063
46. Takei Y, Yun J, Zheng S, Ollikainen N, Pierson N, White J, et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature*. 2021;590(7845):344–50. <https://doi.org/10.1038/s41586-020-03126-2> PMID: 33505024
47. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*. 2018;172(5):1091–1107.e17. <https://doi.org/10.1016/j.cell.2018.02.001> PMID: 29474909
48. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*. 2018;361(6400):eaat5691. <https://doi.org/10.1126/science.aat5691> PMID: 29930089
49. Joglekar A, Prijibelski A, Mahfouz A, Collier P, Lin S, Schlusche AK, et al. A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat Commun*. 2021;12(1):463. <https://doi.org/10.1038/s41467-020-20343-5> PMID: 33469025
50. Alon S, Goodwin DR, Sinha A, Wassie AT, Chen F, Daugharthy ER, et al. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science*. 2021;371(6528):eaax2656. <https://doi.org/10.1126/science.aax2656> PMID: 33509999

51. Zhou Y, Yang D, Yang Q, Lv X, Huang W, Zhou Z, et al. Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat Commun.* 2020;11(1):6322. <https://doi.org/10.1038/s41467-020-20059-6> PMID: [33303760](#)
52. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint. 2017. <https://doi.org/10.48550/arXiv.1711.05101>
53. Li N, Qi W, Jiao J, Li A, Li L, Xu W. SPCC: A superpixel and color clustering based camouflage assessment. *Multimed Tools Appl.* 2023;83(9):26255–79. <https://doi.org/10.1007/s11042-023-16425-3>
54. Romano S, Vinh NX, Bailey J, Verspoor K. Adjusting for chance clustering comparison measures. *Journal of Machine Learning Research.* 2016;17(134):1–32.
55. Yuan L, Xu Z, Meng B, Ye L. scAMZI: attention-based deep autoencoder with zero-inflated layer for clustering scRNA-seq data. *BMC Genomics.* 2025;26(1):350. <https://doi.org/10.1186/s12864-025-11511-2> PMID: [40197174](#)
56. Li X, Lin Y, Xie C, Li Z, Chen M, Wang P, et al. A Clustering Method Unifying Cell-Type Recognition and Subtype Identification for Tumor Heterogeneity Analysis. *IEEE/ACM Trans Comput Biol Bioinform.* 2023;20(2):822–32. <https://doi.org/10.1109/TCBB.2022.3203185> PMID: [36044493](#)