DUSTED: Dual-Attention Enhanced Spatial Transcriptomics Denoiser

Jun Zhu[1,2,3*†], Yifu Li[4,5*], Zhenchao Tang[6*], Cheng Chang[2,3†]

[1]School of Life Sciences, Tsinghua University, Beijing, 100084, China
[2]National Center for Protein Sciences (Beijing), Beijing, 102206, China
[3]Beijing Institute of Lifeomics, Beijing, 102206, China
[4]National Superior College for Engineers,Beihang University, Beijing, 100191, China
[5]School of Biological Science and Medical Engineering, Beihang University, Beijing, 100191, China
[6]School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen, 518107, China
zhuj21@mails.tsinghua.edu.cn,ferryliyifu@buaa.edu.cn, tangzhch7@mail2.sysu.edu.cn, changcheng@ncpsb.org.cn

Abstract

Spatially Resolved Transcriptomics (SRT) has become an indispensable tool in various fields, including tumor microenvironment identification, neurobiology, and the study of complex tissue architecture. However, the accuracy of these insights is often compromised by noise in spatial transcriptomics data due to technical limitations. While recent advancements in denoising methods have shown some promise, they frequently fall short by neglecting spatial features, overlooking the variability in noise levels among genes, and relying heavily on external histological images for supplementary information. In our study, we propose DUSTED, a Dual-Attention Enhanced Spatial Transcriptomics Denoiser, designed to address these challenges. Built on a graph autoencoder framework, DUSTED utilizes gene channel attention and graph attention mechanisms to simultaneously consider spatial features and noise variability in gene expression data. Additionally, it integrates the negative binomial distribution with or without zero-inflation, ensuring a more accurate fit for gene expression distributions. Benchmark tests using simulated datasets demonstrate that DUSTED outperforms existing methods. Furthermore, in real-world applications with the HOCWTA and DLPFC datasets, DUSTED excels in enhancing the correlation between gene and protein expression, recovering spatial gene expression patterns, and improving clustering results. These improvements underscore its potential impact on advancing our understanding of tumor microenvironments, neural tissue organization, and other biologically significant areas.

Code — https://github.com/Lifeomics/DUSTED

Introduction

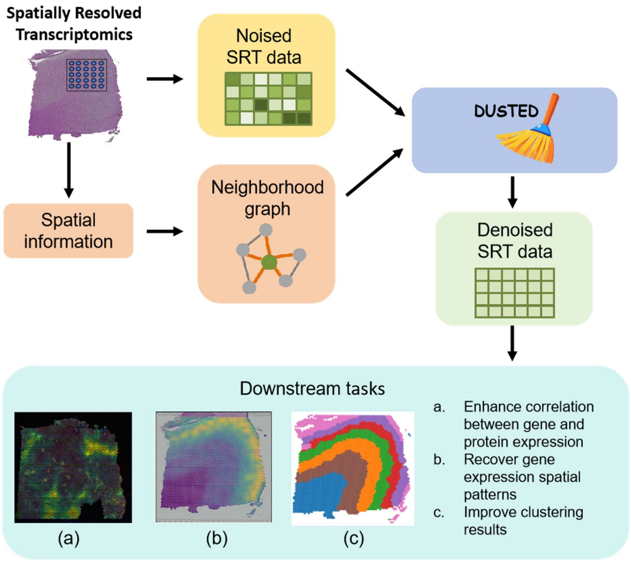
Spatially Resolved Transcriptomics
Noised SRT data
DUSTED
Spatial information
Neighborhood graph
Denoised SRT data
Downstream tasks
(a)
(b)
(c)
a. Enhance correlation between gene and protein expression
b. Recover gene expression spatial patterns
c. Improve clustering results

Figure 1: Overview of DUSTED denoising procedure. DUSTED utilizes the gene expression matrix and neighborhood graph constructed using spatial information to estimate clean gene expression, which has been proven to enhance performance in downstream tasks.

Spatially Resolved Transcriptomics (SRT) is a powerful technique for analyzing gene expression within tissue samples, capturing the spatial distribution of transcriptional activity (Moses and Pachter 2022). Unlike single-cell RNA sequencing (scRNA-seq), SRT retains spatial context, providing deeper insights into the roles, distribution, and interactions of various cell types in both healthy and diseased tissues (Tian, Chen, and Macosko 2023).

However, SRT data often suffer from technical noise introduced during sample preparation and sequencing (Du et al. 2023). This noise leads to sparse and imprecise gene expression matrices and blurred spatial patterns, which can compromise downstream analyses and yield inaccurate results (Wang et al. 2022). As SRT data becomes more abundant, deep learning-based denoising techniques have been developed to address these challenges (Heumos et al. 2023).

Initially, denoising methods from single-cell transcriptomics were directly applied to SRT data. Techniques such as MAGIC (Van Dijk et al. 2018) and kNN-smoothing (Wagner, Yan, and Yanai 2017) smooth gene expression

[*]These authors contributed equally.
[†]Corresponding author.

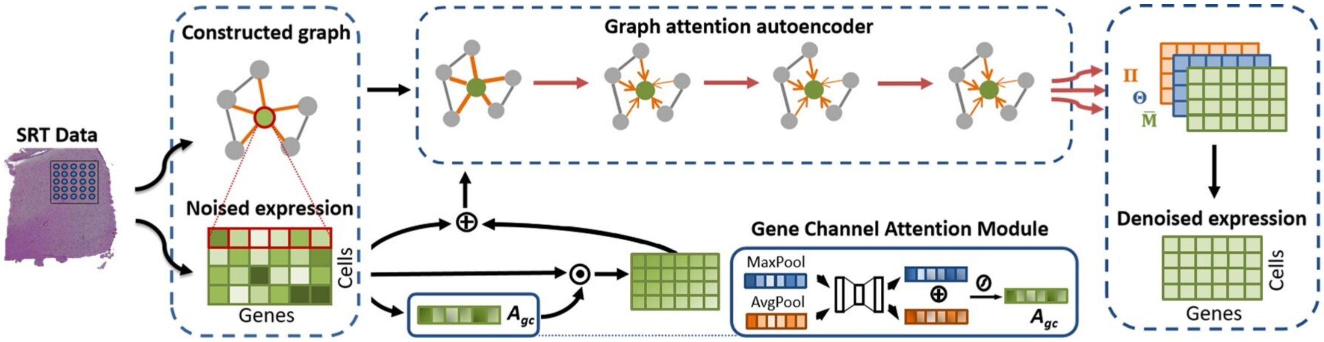

Figure 2: The framework of DUSTED. The input consists of a noised gene expression matrix and an distance-based neighborhood graph. The outputs are three matrices $\overline{M}, \Theta, \Pi$, which stand for mean value, dispersion and drop-out rate of estimated gene expression respectively, fitted by NB or ZINB distribution. We take $\overline{\mathrm{M}}$ as the denoised expression. A_{gc} denotes the gene channel attention weights between genes, reflecting the differences in noise levels between genes.

based on cell similarity. Matrix-based methods like ScRMD (Chen et al. 2020) and ZIFA (Pierson and Yau 2015) recover denoised gene profiles by decomposing sparse expression matrices into low-rank components. Deep learning approaches, including AutoImpute (Talwar et al. 2018), scVI (Lopez et al. 2018), and DCA (Eraslan et al. 2019), utilize autoencoders for denoising single-cell data. However, these methods often ignore the spatial information inherent in SRT data, limiting their effectiveness.

Recent efforts in SRT denoising have recognized the importance of integrating spatial information with gene expression data. Sprod (Wang et al. 2022) builds a similarity graph from SRT and image data, refining gene expression with a regularized least squares method based on local context. Dong and Zhang (2022) proposed graph attention autoencoders that combine spatial location with gene expression data. Su et al.(2023) developed a framework that integrates spatial and histological features to establish prior spatial dependencies. Pham et al. (2023) proposed stSME, a spatial graph-based neural network using imaging, spot distances, and gene expression similarity. SiGra (Tang et al. 2023) employs image-augmented graph transformers for SRT denoising and spatial recognition. These methods have conducted extensive explorations on how to utilize spatial information. However, the reliance on ancillary data, such as pathological images, in many of these methods may constrain their applicability. Furthermore, the omission of gene expression distributions, interactions, and the heterogeneity of noise levels across various genes can be a significant oversight.

To overcome these challenges, we propose DUSTED, a Dual-Attention Enhanced Spatial Transcriptomics Denoiser built on a graph autoencoder framework. DUSTED employs a dual-attention mechanism that simultaneously focuses on spatial features and noise variations in transcriptional expressions. In this mechanism, the gene expression at any location can be interpolated based on the gene expression of neighboring spots. In addition, to make the generated SRT counts conform more realistically, we refined the output to align with the negative binomial distribution with or without zero-inflation. By leveraging prior biological distribution, our method more accurately fits the true gene expression profiles, achieving superior performance in self-supervised SRT data denoising without the need for external auxiliary information.

Since SRT denoising is an emerging task, we design a benchmark simulation data generation method for evaluating SRT data denoising performance, providing a quantitative benchmark for the evaluation and comparison of denoising methods. For real-world evaluation, we use two real datasets, HOCWTA and DLPFC. Experimental results show that our method can achieve realistic and robust denoising performance.

Related Works

Graph Autoencoder

Autoencoders (AEs) are renowned for feature extraction and reconstruction in unsupervised learning. (Kipf and Welling 2016)(2016) pioneered graph-based AEs with Variational Graph Auto-Encoders (VGAE) and Graph Auto-Encoders (GAE) for citation network link prediction. Over time, the structure of GAEs has been optimized and extended, leading to their widespread use in graph-structured tasks such as social network analysis (Yang et al. 2020), recommendation systems (Zhang et al. 2023), and chemical molecular structure prediction (Oliveira, Da Silva, and Quiles 2022). Dong and Zhang(2022) used graph attention (GAT) layers to construct encoders and decoders, adaptively learning the similarity of adjacent points, and proposed a unified framework for spatial domain recognition, data denoising, and extraction of 3D spatial domains for SRT data.

Channel Attention

The channel attention mechanism captures interdependencies between feature channels, enabling adaptive rescaling of each channel's features. By assigning varying weights, it recalibrates the importance of input channels. Channel attention is widely used in computer vision tasks. Hu et al.(2018) first applied adaptive average pooling (squeeze) to the spatial dimension, followed by two fully connected layers to learn channel attention (excitation). Most subsequent channel attention works are based on the Squeeze-Excitation concept. For instance, CBAM (Woo et al. 2018) uses both adaptive average pooling and adaptive max pooling for the squeeze step, then combines channel attention with spatial attention. Fu et al.(2019) extracted channel attention and position attention through self-attention operations and performed ele-ment-wise summation.

In the context of SRT technology, where the PCR process may preferentially amplify certain genes and the expression levels of different genes vary (Kanagawa 2003), the resulting noise levels can differ between genes. The application of the channel attention mechanism to address the noise differences provides biological interpretability for SRT denoising, making it a valuable tool in this domain.

Spatial Transcriptomics Denoiser

Recent research has underscored the importance of spatial information in denoising Spatially Resolved Transcriptomics (SRT) data, leading to several innovative approaches. Sprod (Wang et al. 2022) uses a similarity graph that integrates SRT data with image features, adjusting gene expression through regularized least squares. Dong and Zhang (2022) use graph attention autoencoders to combine spatial location with gene expression data, creating low-dimensional embeddings for denoising. Su et al. (2023) developed a framework that integrates spatial and histological features, using a hidden Markov random field for interpolation denoising. Pham et al. (2023) developed a spatial graph-based imputation method with neural network, which which leverages imaging data, the proximity of spots, and the similarity of gene expression. SiGra (Tang et al. 2023) integrates single-cell multimodal data using three graph transformer-based autoencoders (imaging, transcriptomics, and hybrid), thereby recovering the true gene expression.

Despite these advances, challenges remain, including dependency on external data, inadequate handling of gene interactions, and limitations in accurately modeling gene expression distributions. DUSTED overcomes these issues by eliminating the need for auxiliary data, effectively capturing gene interactions and noise variations, and providing a more accurate fit for gene expression distributions.

Methodology

The overview of the proposed framework is shown in Figure 2. The input of the model is an undirected neighborhood graph $G(V, E)$, constructed using Euclidean distance as the radius. In this graph, V represents each spot or cell within the tissue, while X denotes the features of the nodes, with X_v representing the normalized gene expression of node v. The edges E capture the neighborhood connections between these spots. Our model is designed to overcome the limitations of existing SRT denoising methods by integrating prior biological knowledge and spatial location information, eliminating the need for auxiliary image data. The dual-attention mechanism in our model includes Gene Channel Attention (GCA) and Graph Attention, which together address the challenges of small sample sizes and complex noise in SRT data. The GCA module captures gene interactions and accounts for noise variations between different genes, enhancing model performance by leveraging the inherent biological relationships. Meanwhile, the Graph Attention module focuses on spatial features of gene expression, ensuring that the denoising process is both highly effective and interpretable, without relying on external imaging inputs.

Gene Channel Attention Module

As illustrated in Figure 2, the gene channel attention module begins by processing the input gene expression features. It aggregates information through average pooling and max pooling operations along the gene dimension, resulting in average pooled features and max pooled features.

These pooled features are then passed through a shared network with a Multi-Layer Perceptron (MLP) structure comprising two layers. The activation size of the hidden layer is set to $\mathrm{R}^{\mathrm{N} \times 1 \times 1}$ to reduce parameter overhead. The output of this shared network generates vector representations for the two channel attention maps. These vectors are ele-ment-wise summed and then passed through a Sigmoid function to form the final gene attention, which is given by the following expression:

$$A_{gc} = \sigma\left(\mathrm{MLP}\left(A_{1 \times G}\right) + \mathrm{MLP}\left(M_{1 \times G}\right)\right) \tag{2.1}$$

The gene channel attention is subsequently multiplied with the original features along the gene dimension. Finally, the adjusted gene expression profiles are connected through a residual connection, incorporating the weights of the residual connection, as illustrated in the following calculation process:

$$X'_{\mathrm{N} \times \mathrm{G}} = X_{\mathrm{N} \times \mathrm{G}} + \alpha \cdot \left(\mathrm{A}_{\mathrm{gc}} \odot X_{\mathrm{N} \times \mathrm{G}}\right) \tag{2.2}$$

where α represents the weights of the residual connection.

Unsupervised Learning Setup

Encoder: Let X be the gene expression of the nodes, and L be the number of layers in the encoder. By treating the expression profiles as the initial node embeddings $\left(\mathrm{h}_{\mathrm{i}}^{(0)} = x_i, \forall i \in \{1, 2, \ldots, N\}\right)$, with N being the number of sequencing nodes, the k-th layer ($k \in \{1, 2, \ldots, L\}$) of the encoder generates the embeddings of node i in layer k as follows:

$$\mathrm{h}_{\mathrm{i}}^{(\mathrm{k})} = \mathrm{ELU}\left(\sum_{\mathrm{i} \in N_{\mathrm{i}}} \alpha_{\mathrm{ij}}^{(\mathrm{k})}\left(\mathrm{W}_{\mathrm{k}}\, \mathrm{h}_{\mathrm{j}}^{(\mathrm{k}-1)}\right)\right) \tag{2.3}$$

where W_{k} is a trainable weight matrix, ELU is the Exponential Linear Unit activation function, N_{i} is the set of neighbors of node i in the spatial neighborhood graph (including i itself), and $\alpha_{\mathrm{ij}}^{(\mathrm{k})}$ is the edge weight between nodes i and j as output by the k-th graph attention layer. The final output of the encoder is the latent node embeddings at layer L.

Decoder: The output of the encoder serves as the input to the decoder, which reconstructs the latent embeddings to produce the denoised gene expression profiles. The k-th layer

$(\mathrm{k} \in \{2, \ldots, \mathrm{L}-1, \mathrm{L}\}$) of the decoder reconstructs the embeddings of node i as shown in (2.4):

$$\widehat{\mathrm{h}}_{\mathrm{i}}^{(\mathrm{k}-1)} = \mathrm{ELU}\left(\sum_{\mathrm{j}\in\ \mathrm{S}_{\mathrm{i}}} \widehat{\mathrm{gat}}_{\mathrm{ij}}^{(\mathrm{k}-1)} \left(\widehat{\mathrm{W}}_{\mathrm{k}}\ \hat{\mathrm{h}}_{\mathrm{j}}^{(\mathrm{k})}\right)\right) \tag{2.4}$$

It's important to note that the final layer of the decoder comprises three output layers instead of one, representing three parameters for each gene. These parameters are used to form the gene-specific loss function, enabling the model to fit the true expression distribution of each gene. The formula for the three parameters of node i in the final output layers as follows:

$$\overline{\mathrm{M}}_{\mathrm{i}} = \exp\left(\hat{\mathrm{h}}_{\mathrm{i}}^{(\mathrm{L})}\right) \tag{2.5}$$

$$\Theta_{\mathrm{i}} = \log(1 + \exp\left(\hat{\mathrm{h}}_{\mathrm{i}}^{(\mathrm{L})}\right)) \tag{2.6}$$

$$\Pi_{\mathrm{i}} = \mathrm{sigmoid}\left(\hat{\mathrm{h}}_{\mathrm{i}}^{(\mathrm{L})}\right) \tag{2.7}$$

Graph Attention Layer: We utilize GAT (Graph Attention Networks) to adaptively learn the similarity between neighboring nodes. GAT is a single-layer feedforward neural network with shared parameters, parameterized by a weight vector. In the k-th encoder layer, the edge weight from node i to its neighboring node j is calculated as shown in (2.8):

$$\mathrm{e}_{\mathrm{ij}}^{(\mathrm{k})} = \sigma\left(\mathrm{v}_{\mathrm{s}}^{(\mathrm{k})^{\mathrm{T}}}\left(\ \mathrm{W}_{\mathrm{k}}\ \mathrm{h}_{\mathrm{i}}^{(\mathrm{k}-1)}\right) + \mathrm{v}_{\mathrm{r}}^{(\mathrm{k})^{\mathrm{T}}}\left(\ \mathrm{W}_{\mathrm{k}}\ \mathrm{h}_{\mathrm{j}}^{(\mathrm{k}-1)}\right)\right) \tag{2.8}$$

where $V_s^{(k)}$ and $V_r^{(k)}$ are learnable weight vectors. To make the spatial similarity weights comparable, we normalize them using the Softmax function, as shown in (2.9):

$$\mathrm{gat}_{\mathrm{ij}}^{(\mathrm{k})} = \frac{\exp\left(\mathrm{e}_{\mathrm{ij}}^{(\mathrm{k})}\right)}{\sum_{\mathrm{i}\in N_{\mathrm{i}}} \exp\left(\mathrm{e}_{\mathrm{ij}}^{(\mathrm{k})}\right)} \tag{2.9}$$

These learned attention weights are further used to update the latent embeddings in both the encoder and decoder.

To avoid overfitting and reduce parameter overhead, we set the attention weight matrix of the k-th decoder layer to be equal to that of the k-th encoder layer, i.e., $\widehat{\mathrm{gat}}^{(\mathrm{k})} = \mathrm{gat}$ $^{(\mathrm{k})}$ (Dong and Zhang 2022).

Loss Function

The Negative Binomial (NB) model and the Zero-Inflated Negative Binomial (ZINB) model are well-suited for fitting the distribution of gene expression in spatial transcriptomics.

The expression for the Negative Binomial distribution as follows:

$$\mathrm{NB}(\mathrm{x}; \mu, \theta) = \frac{\Gamma(\mathrm{x}+\theta)}{\Gamma(\theta)}\left(\frac{\theta}{\theta+\mu}\right)^{\theta}\left(\frac{\mu}{\theta+\mu}\right)^{\mathrm{x}} \tag{2.10}$$

where μ is the mean of the Negative Binomial distribution, and θ is the inverse dispersion parameter, which controls the variance

For transcriptomics data based on Unique Molecular Identifiers (UMIs), which have a lower dropout rate (Chen et al. 2018), the simpler Negative Binomial distribution is appropriate. However, for data without UMI technology, where the dropout rate is higher, and results in a sparser expression matrix. In this case, using the Zero-Inflated Negative Binomial distribution can better fit the data distribution (Eraslan et al. 2019). The expression for the Zero-Inflated Negative Binomial distribution is:

$$\mathrm{ZINB}(x; \pi, \mu, \theta) = \pi\delta_0(x) + (1-\pi)\,\mathrm{NB}(x; \mu, \theta) \tag{2.11}$$

where $\delta_0(\mathrm{x})$ is the Dirac delta function representing dropout-induced zeros, and π is the mixture coefficient representing the probability of drop-out.

Given the similarities in amplification and sequencing procedures between SRT and single-cell transcriptomics, we similarly assume that SRT data follow either NB or ZINB distribution. Consequently, we provide both types of loss functions during model training. The Negative LogLikelihood (NLL) functions for NB and ZINB distributions are denoted as $\mathcal{L}_{\mathrm{NB}}$ and $\mathcal{L}_{\mathrm{ZINB}}$ respectively, as shown in (2.12) and (2.13):

$$\mathcal{L}_{\mathrm{NB}} = \mathrm{argmin}_{\Pi,\mathrm{M},\Theta}(-\,\mathrm{logNB}(\mathrm{X}; \mathrm{M}, \Theta)) \tag{2.12}$$

$$\mathcal{L}_{\mathrm{ZINB}} = \mathrm{argmin}_{\Pi,\mathrm{M},\Theta}(-\,\mathrm{logZINB}(\mathrm{X}; \Pi, \mathrm{M}, \Theta)) \tag{2.13}$$

By minimizing the NLL, the model can effectively recover gene expressions that fit the underlying distribution characteristics.

Experiments

Experimental Setup

Datasets. We evaluate our algorithm using both simulated and real datasets to ensure a comprehensive assessment.

Simulated datasets provide SRT data along with the corresponding true gene expressions. By comparing the difference between the SRT data before and after denoising with the true gene expressions, we can quantitatively assess the denoising performance of our model. Our method for generating simulated SRT data consists of three steps: (i)Generate scRNA-seq data and corresponding true gene expression value by Symsim (Zhang, Xu, and Yosef 2019). (ii)Assign singlecell gene expression value to randomly generated spatial patterns by scCube (Qian et al. 2024). (iii)Apply dropout events (a technical noise arising from low RNA content or undetected gene expression, leading to missing data (Patruno et al. 2021) of specific spatial patterns by spatially varying the drop-out rate. Figure 3 shows six different simulated spatial patterns of SRT data we generated in steps (i) and (ii), each with varied cell types and gene expression features, as well as technical noise introduced in the simulated PRC procedure. By setting different drop-out rate within the edge range compared to the overall drop-out rate, we introduced more drop-out noise with spatial patterns in step (iii). The application details are shown in Table 1. By applying 4

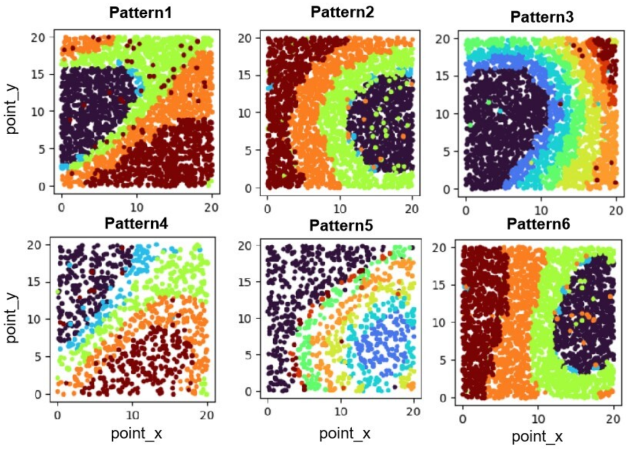

Figure 3: Six different spatial patterns of simulated SRT data. Different colored dots represent different types of cells. By adding four settings of drop-out noise respectively, we generated 24 simulated SRT datasets.

Setting	Overall Drop-out Rate	Increased Drop-out Rate	Affected Edge Distance (units)
1	0	0	None
2	0.1	0.2	3
3	0.1	0.3	5
4	0.2	0.5	6

Table 1: Application details of four drop-out settings. By adjusting drop-out rate at the edges, we introduced drop-out noise with spatial patterns to simulated SRT data.

settings of drop-out noise to 6 patterns of SRT data respectively, we obtain 24 datasets of simulated SRT data varying in gene numbers, cell numbers, spatial distribution, and noise levels.

Real datasets are used to validate that the denoised data retains biologically relevant information, demonstrating the practical value of our methods. The Human Ovarian Cancer: Whole Transcriptome Analysis (HOCWTA) dataset (10x Genomics 2020) includes 10x Visium SRT data for ovarian endometrioid adenocarcinoma tissue, along with corresponding four-channel fluorescence imaging (DAPI, FITC, TRITC, and Cy5). We use this dataset to test whether DUSTED can improve the correlation between gene and protein expressions. The Human Dorsolateral Prefrontal Cortex (DLPFC) dataset (Maynard et al. 2021) consists of SRT data for 12 dorsolateral prefrontal cortex tissue sections, including images and manual annotations for six layers and white matter. Additionally, it provides 126 enriched genes specific to different layers, which serve as marker genes to reveal the positions and spatial structures of the layers. This dataset is used to evaluate whether our denoising algorithm can improve the spatial patterns of gene expression and spots clustering.

Hyperparameters. The encoder and decoder of DUSTED are both set to two layers, with feature dimensions from input to output being [2000, 512, 30, 512, 2000]. The parameter updates are optimized using the Adam optimizer, and the model is trained for 500 epochs. For the simulated dataset, the learning rate is set to 10^{-4}, α (the weights of the residual connection) is set to 0.3 , and the loss function is selected as $\mathcal{L}_{\mathrm{NB}}$. For the HOCWTA dataset, the learning rate is 2×10^{-4}, α is set to 1 , and the loss function is selected as $\mathcal{L}_{\mathrm{NB}}$. For the DLPFC dataset, the learning rate is 10^{-4}, α is set to 1.5 , and the loss function is selected as $\mathcal{L}_{\mathrm{ZINB}}$. For constructing the neighborhood graph, the radius r is adjusted for different datasets to ensure each node has 5 6 neighboring nodes. The model is built and trained using the PyTorch deep learning framework, and all experiments are conducted on a single NVIDIA Quadro GV100 GPU with 32GB of memory.

Evaluation Metrics. For experiments using simulated datasets, we employ Mean Square Error (MSE) as the primary evaluation metric. MSE quantitatively measures the denoising effectiveness by comparing the difference between the SRT data and the true gene expression across the gene dimension before and after denoising. For the HOCWTA dataset, we assess the denoising performance by calculating Spearman and Pearson correlation coefficients (Schober, Boer, and Schwarte 2018) to measure the relationship between the expression levels of the PTPRC gene and the CD45 protein before and after denoising. In the DLPFC dataset, we use the Global Moran's I index (Moran 1950) to evaluate spatial autocorrelation, which measures the clustering and continuity of layer marker genes in space. Additionally, we employ the Adjusted Rand Index (ARI) (Vinh, Epps, and Bailey 2009), Normalized Mutual Information (NMI) (Pfitzner, Leibbrandt, and Powers 2009), and Homogeneity Score (HS) (Rosenberg and Hirschberg 2007) to measure the similarity between the clustering results and manual annotations.

Baselines. To evaluate the performance of the DUSTED algorithm, we compare it with state-of-the-art transcriptomics denoising algorithms, including methods designed for both single-cell transcriptomics and spatial transcriptomics:

- DCA (Eraslan et al. 2019): An autoencoder-based method for single-cell transcriptomics that models each gene count as a negative binomial (NB) or zero-inflated negative binomial (ZINB) distribution to constrain reconstruction errors.
- MAGIC (Van Dijk et al. 2018): A method that uses data diffusion to characterize cell-to-cell similarity via a power Markov matrix. The original data is then multiplied by this matrix to obtain denoised data.
- STAGATE (Dong and Zhang 2022): This approach uses a graph attention autoencoder to integrate spatial location and gene expression data, learning latent embeddings that are decoded to produce denoised spatial gene expression data.
- Smoother (Su et al. 2023): This method constructs a weighted spatial graph from physical, histological, and other features, then employs a stochastic process to create a covariance matrix indicating spatial dependencies. For denoising, it applies a Smoother method leveraging

Method	Avg_MSE			
	setting 1	setting 2	setting 3	setting 4
Raw	1.0831 ± 0.0975	1.0406 ± 0.1479	1.2575 ± 0.1299	1.6977 ± 0.1124
Smoother (Su et al. 2023)	0.8004 ± 0.1855	1.1291 ± 0.1514	1.3489 ± 0.1281	1.8076 ± 0.0779
STAGATE (Dong and Zhang 2022)	0.4886 ± 0.0847	0.5523 ± 0.0874	0.6275 ± 0.0855	0.8408 ± 0.0975
MAGIC (Van et al. 2018)	0.4004 ± 0.0456	0.4097 ± 0.0426	0.4179 ± 0.0428	0.4390 ± 0.0521
DCA (Eraslan et al. 2019)	0.3925 ± 0.0477	0.3916 ± 0.0435	0.3929 ± 0.0459	0.4054 ± 0.0567
DUSTED	**0.3839 ± 0.0533**	**0.3840 ± 0.0485**	**0.3842 ± 0.0478**	**0.3929 ± 0.0463**

Table 2: The average MSE between simulated SRT dataset and ground truth before (RAW) and after denoising. DUSTED gets the lowest average MSE.

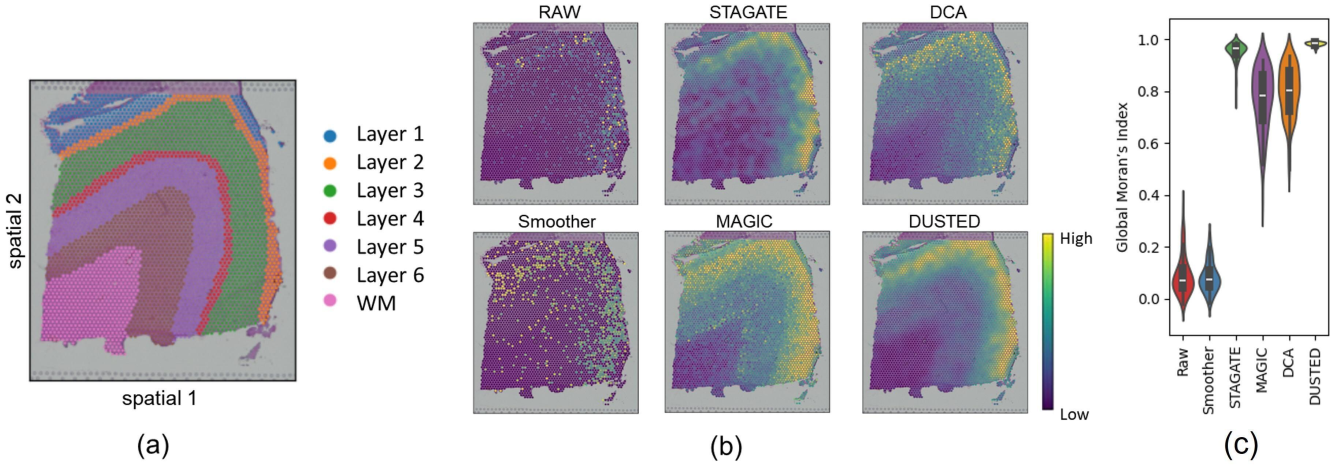

Figure 4: (a)Ground-truth segmentation of cortical layers and white matter (WM) in the DLPFC section 151673. (b)Spatial expression intensity of the marker gene CARTPT in the DLPFC section 151673. It is intuitively observed that DUSTED results have a clear expression profile consistent with the Layer 3 spatial pattern. (c) Global Moran's index of marker genes in the DLPFC section 151673.

hidden Markov random fields based on these spatial priors, with the denoised gene expression as the field's mean at each location.

Experimental Result

Performance of DUSTED on Simulated Dataset. We computed the MSE between denoised data and ground truth for 24datasets of simulated data. We adjusted the random seed and repeated the experiment three times per dataset, yielding the average MSE for the 24 datasets as shown in Table 2. The results indicate that DUSTED reduced the average MSE by 69.6% compared to the pre-denoising state, outperforming the comparative methods.

Performance of DUSTED on HOCWTA. CD45 is a transmembrane glycoprotein that plays a crucial role in regulating activation mediated by T and B cell antigen receptors. It is encoded by the PTPRC gene, and there is typically a strong correlation between CD45 protein expression and PTPRC gene expression. Using Squidpy (Palla et al. 2022), we extracted the intensity of CD45 protein expression at each sequencing point and computed its correlation with the corresponding PTPRC gene expression intensity, as shown in Figure 5. The results clearly demonstrate that, after denoising with the DUSTED model, the Pearson and Spear man correlation coefficients between PTPRC gene expression and CD45 protein expression are significantly higher compared to the other four comparative methods. Compared to MAGIC, which achieved the second best performance, DUSTED improved the Pearson correlation by 0.0264 (8.45%) and the Spearman correlation by 0.0426 (11.31%). These findings underscore the DUSTED model's exceptional ability to recover true gene expression and enhance the correlation between gene and protein expression.

Performance of DUSTED on DLPFC. Firstly, we assessed the model's ability to recover spatial patterns of gene expression across 12 datasets of the DLPFC. We selected the top

Task	Simulated Datasets	HOCTWA		DLPFC		
Evaluation	Avg_MSE ↓	Pearson ↑	Spearman ↑	Avg_ARI ↑	Avg_NMI ↑	Avg_HS ↑
without GCA	0.3904 ± 0.0087	0.30096	0.40337	0.4475 ± 0.0965	0.6158 ± 0.0665	0.6158 ± 0.0665
with GCA	$\mathbf{0.3865 \pm 0.0070}$	**0.33852**	**0.42193**	$\mathbf{0.5125 \pm 0.0859}$	$\mathbf{0.6542 \pm 0.0692}$	$\mathbf{0.6542 \pm 0.0692}$

Table 3: Results of Ablation Study on Simulated datasets and two real datasets. The bold font indicates the best performance.

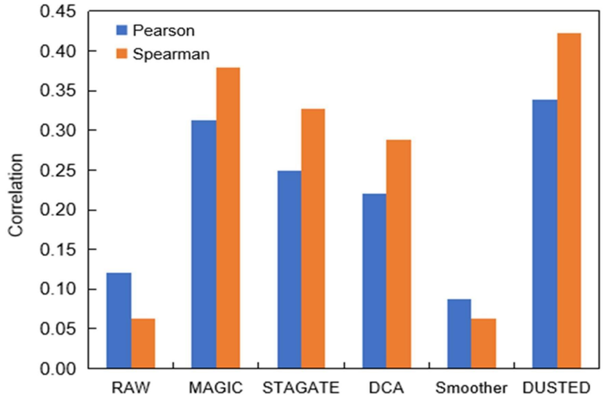

Figure 5: Correlation between PTPRC and CD45 expression intensity before and after denoising on HOCWTA Dataset. DUSTED achieves the best results in both Pearson and Spearman correlation.

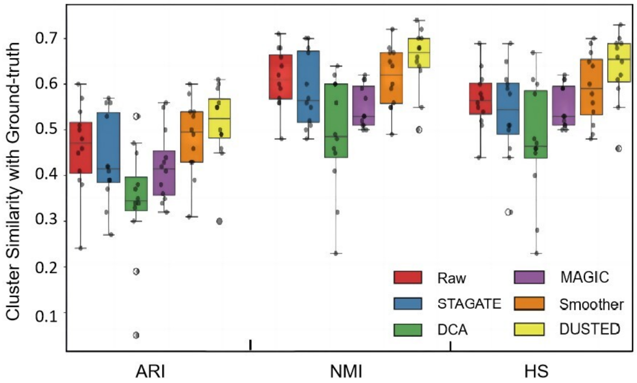

Figure 6: Comparison of clustering results with annotated ground truth across 12 sections in the DLPFC dataset.

3000 highly variable genes and intersected them with layerenriched genes provided by the dataset as marker genes for evaluation. As shown in Figure 4(c), DUSTED consistently achieved Moran's I values closest to 1 with the smallest standard deviation, indicating best and most robust spatial continuity improvement. This trend was replicated across all twelve sections. Figure 4(b) depicts the spatial expression intensity of the marker gene CARTPT in section 151673 from Layer 3. It is evident that the gene expression intensity processed by DUSTED achieved the clearest spatial pattern of CARPET expression intensity, and is most consistent with the manually annotated layer 3 in Figure 4(a). Similar results are also observed in marker genes of other layers, demonstrating the effectiveness of DUSTED in restoring gene expression spatial patterns.

Next, we evaluated the model's ability to enhance clustering results. We first reduced the dimensionality of gene expression profiles using STAGATE, then applied the mclust (Scrucca et al. 2016) clustering method on the low-dimensional embeddings. Clustering performance was assessed using ARI, NMI, and HS metrics against the annotated ground truth, as shown in Figure 6. The box plots indicate that DUSTED denoised data achieved higher median values of ARI (0.515), NMI (0.685), and HS (0.665) compared to raw data and other methods. Additionally, the small interquartile ranges and narrow bounds demonstrate low variability and stable performance.

Ablation Study. We evaluated the impact of the Gene Channel Attention (GCA) mechanism in DUSTED through experiments on both simulated and real datasets. As shown in Table 3, removing GCA significantly reduced performance on the HOCWTA dataset, with Pearson and Spearman correlations decreasing by 11.1% and 4.4%, respectively. Additionally, incorporating GCA led to substantial improvements in average ARI, NMI, and HS across 12 sections, increasing by 14.53%, 6.24%, and 9.47%. These results highlight the critical importance of the GCA mechanism.

Conclusion

We introduced DUSTED, a denoising model for Spatially Resolved Transcriptomics (SRT) data that employs a dual-attention mechanism within a graph autoencoder framework. DUSTED excels in self-supervised denoising without the need for external image. The gene channel attention effectively captures spatial features and addresses noise variations across genes. Furthermore, modeling gene distributions with NB/ZINB ensures that denoised expressions closely reflect true values. Comprehensive experiments on both simulated and real datasets demonstrate DUSTED's effectiveness and practical utility in enhancing SRT data quality.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2021YFA1301603).

References

10x Genomics. 2020. Human Ovarian Cancer: Whole Transcriptome Analysis. https://www.10xgenomics.com/datasets/human-ovarian-cancer-whole-transcriptome-analysis-stains-dapi-anti-pan-ck-anti-cd-45-1-standard-1-2-0. Stains: DAPI, Anti-PanCK, Anti-CD45.

Chen, C.; Wu, C.; Wu, L.; Wang, X.; Deng, M.; and Xi, R. 2020. scRMD: imputation for single cell RNA-seq data via robust matrix decomposition. *Bioinformatics*, 36(10): 3156–3161.

Chen, W.; Li, Y.; Easton, J.; Finkelstein, D.; Wu, G.; and Chen, X. 2018. UMI-count modeling and differential expression analysis for single-cell RNA sequencing. *Genome biology*, 19: 1–17.

Dong, K.; and Zhang, S. 2022. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications*, 13(1): 1739.

Du, J.; Yang, Y.-C.; An, Z.-J.; Zhang, M.-H.; Fu, X.-H.; Huang, Z.-F.; Yuan, Y.; and Hou, J. 2023. Advances in spatial transcriptomics and related data analysis strategies. *Journal of Translational Medicine*, 21(1): 330.

Eraslan, G.; Simon, L. M.; Mircea, M.; Mueller, N. S.; and Theis, F. J. 2019. Single-cell RNA-seq denoising using a deep count autoencoder. *Nature communications*, 10(1): 390.

Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; and Lu, H. 2019. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3146–3154.

Heumos, L.; Schaar, A. C.; Lance, C.; Litinetskaya, A.; Drost, F.; Zappia, L.; Lücken, M. D.; Strobl, D. C.; Henao, J.; Curion, F.; et al. 2023. Best practices for single-cell analysis across modalities. *Nature Reviews Genetics*, 24(8): 550–572.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Kanagawa, T. 2003. Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of bioscience and bioengineering*, 96(4): 317–323.

Kipf, T. N.; and Welling, M. 2016. Variational graph autoencoders. *arXiv preprint arXiv:1611.07308*.

Lopez, R.; Regier, J.; Cole, M. B.; Jordan, M. I.; and Yosef, N. 2018. Deep generative modeling for single-cell transcriptomics. *Nature methods*, 15(12): 1053–1058.

Maynard, K. R.; Collado-Torres, L.; Weber, L. M.; Uytingco, C.; Barry, B. K.; Williams, S. R.; Catallini, J. L.; Tran, M. N.; Besich, Z.; Tippani, M.; et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience*, 24(3): 425–436.

Moran, P. A. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2): 17–23.

Moses, L.; and Pachter, L. 2022. Museum of spatial transcriptomics. *Nature methods*, 19(5): 534–546.

Oliveira, A. F.; Da Silva, J. L.; and Quiles, M. G. 2022. Molecular property prediction and molecular design using a supervised grammar variational autoencoder. *Journal of Chemical Information and Modeling*, 62(4): 817–828.

Palla, G.; Spitzer, H.; Klein, M.; Fischer, D.; Schaar, A. C.; Kuemmerle, L. B.; Rybakov, S.; Ibarra, I. L.; Holmberg, O.; Virshup, I.; et al. 2022. Squidpy: a scalable framework for spatial omics analysis. *Nature methods*, 19(2): 171–178.

Patruno, L.; Maspero, D.; Craighero, F.; Angaroni, F.; Antoniotti, M.; and Graudenzi, A. 2021. A review of computational strategies for denoising and imputation of single-cell transcriptomic data. *Briefings in bioinformatics*, 22(4): bbaa222.

Pfitzner, D.; Leibbrandt, R.; and Powers, D. 2009. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19: 361–394.

Pham, D.; Tan, X.; Balderson, B.; Xu, J.; Grice, L. F.; Yoon, S.; Willis, E. F.; Tran, M.; Lam, P. Y.; Raghubar, A.; et al. 2023. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature communications*, 14(1): 7739.

Pierson, E.; and Yau, C. 2015. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome biology*, 16: 1–10.

Qian, J.; Bao, H.; Shao, X.; Fang, Y.; Liao, J.; Chen, Z.; Li, C.; Guo, W.; Hu, Y.; Li, A.; et al. 2024. Simulating multiple variability in spatially resolved transcriptomics with scCube. *Nature Communications*, 15(1): 5021.

Rosenberg, A.; and Hirschberg, J. 2007. V-measure: A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 410–420.

Schober, P.; Boer, C.; and Schwarte, L. A. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, 126(5): 1763–1768.

Scrucca, L.; Fop, M.; Murphy, T. B.; and Raftery, A. E. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R journal*, 8(1): 289.

Su, J.; Reynier, J.-B.; Fu, X.; Zhong, G.; Jiang, J.; Escalante, R. S.; Wang, Y.; Aparicio, L.; Izar, B.; Knowles, D. A.; et al. 2023. Smoother: a unified and modular framework for incorporating structural dependency in spatial omics data. *Genome Biology*, 24(1): 291.

Talwar, D.; Mongia, A.; Sengupta, D.; and Majumdar, A. 2018. AutoImpute: Autoencoder based imputation of single-cell RNA-seq data. *Scientific reports*, 8(1): 16329.

Tang, Z.; Li, Z.; Hou, T.; Zhang, T.; Yang, B.; Su, J.; and Song, Q. 2023. SiGra: single-cell spatial elucidation through an image-augmented graph transformer. *Nature Communications*, 14(1): 5618.

Tian, L.; Chen, F.; and Macosko, E. Z. 2023. The expanding vistas of spatial transcriptomics. *Nature Biotechnology*, 41(6): 773–782.

Van Dijk, D.; Sharma, R.; Nainys, J.; Yim, K.; Kathail, P.; Carr, A. J.; Burdziak, C.; Moon, K. R.; Chaffer, C. L.; Pattabiraman, D.; et al. 2018. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3): 716–729.

Vinh, N. X.; Epps, J.; and Bailey, J. 2009. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th annual international conference on machine learning*, 1073–1080.

Wagner, F.; Yan, Y.; and Yanai, I. 2017. K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. *BioRxiv*, 217737.

Wang, Y.; Song, B.; Wang, S.; Chen, M.; Xie, Y.; Xiao, G.; Wang, L.; and Wang, T. 2022. Sprod for de-noising spatially resolved transcriptomics data based on position and image information. *Nature methods*, 19(8): 950–958.

Woo, S.; Park, J.; Lee, J.-Y.; and Kweon, I. S. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, 3–19.

Yang, C.; Zhang, J.; Wang, H.; Li, S.; Kim, M.; Walker, M.; Xiao, Y.; and Han, J. 2020. Relation learning on social networks with multi-modal graph edge variational autoencoders. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 699–707.

Zhang, X.; Xu, C.; and Yosef, N. 2019. Simulating multiple faceted variability in single cell RNA sequencing. *Nature communications*, 10(1): 2611.

Zhang, Y.; Zhang, Y.; Yan, D.; Deng, S.; and Yang, Y. 2023. Revisiting graph-based recommender systems from the perspective of variational auto-encoder. *ACM Transactions on Information Systems*, 41(3): 1–28.