

DANet: spatial gene expression prediction from H&E histology images through dynamic alignment

Yulong Wu^{1,‡}, Jin Xie^{1,‡}, Jing Nie^{2,3,*}, Jiale Cao⁴, Yuansong Zeng^{1,5,*}, Zheng Wang^{5,*}

¹School of Big Data and Software Engineering, Chongqing University, Chongqing 400044, China

²School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China

³Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China

⁴School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

⁵Jinfeng Laboratory, Chongqing 401329, China

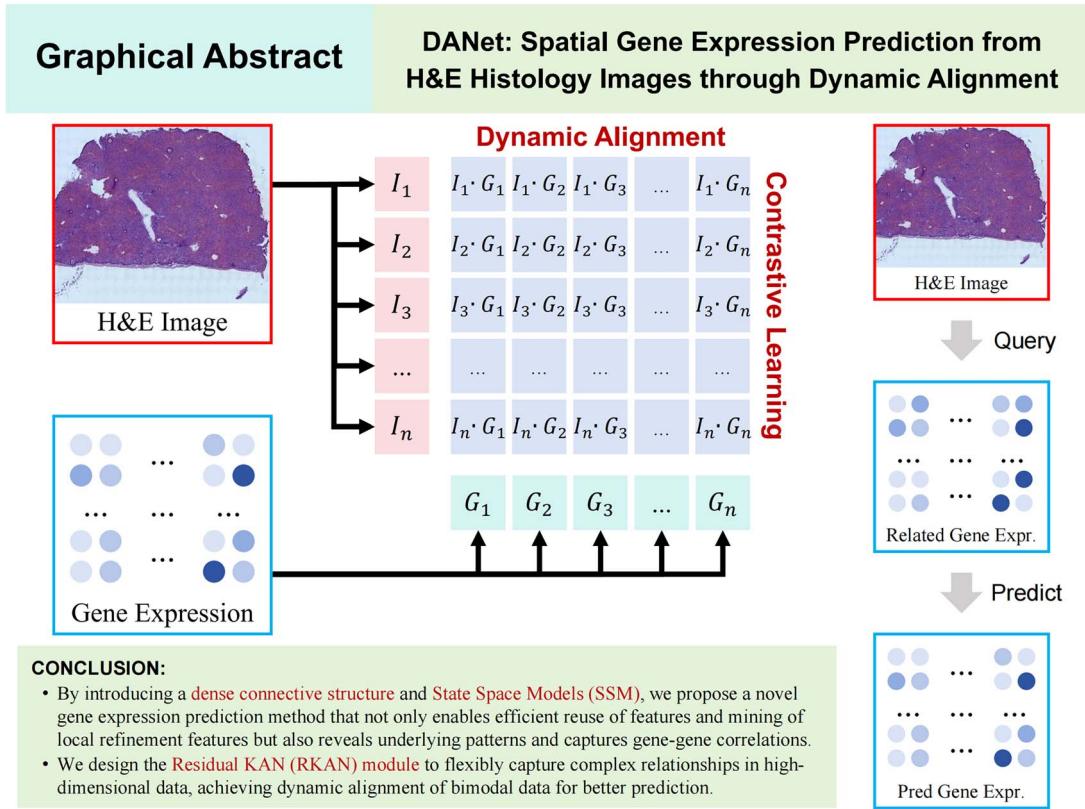
*Corresponding authors. Jing Nie, E-mail: jingnie@cqu.edu.cn; Yuansong Zeng, E-mail: zengys@cqu.edu.cn; Zheng Wang, E-mail: biowz@mail.ustc.edu.cn.

‡Yulong Wu and Jin Xie contributed equally to this work.

Abstract

Predicting spatial gene expression from Hematoxylin and Eosin histology images offers a promising approach to significantly reduce the time and cost associated with gene expression sequencing, thereby facilitating a deeper understanding of tissue architecture and disease mechanisms. Achieving accurate gene expression prediction requires the extraction of highly refined features from pathological images; however, existing methods often struggle to effectively capture fine-grained local details and model gene-gene correlations. Moreover, in bimodal contrastive learning, dynamically and efficiently aligning heterogeneous modalities remains a critical challenge. To address these issues, we propose a novel method for predicting gene expression. First, we introduce a dense connective structure that enables efficient feature reuse, thereby enhancing the capturing and mining of local refinement features. Second, we leverage the state space models to uncover underlying patterns and capture dependencies within 1D gene expression data, enabling more accurate modeling of gene-gene correlations. Furthermore, we design the Residual Kolmogorov-Arnold Network (RKAN) that uses a learnable activation function to dynamically adjust bimodal mappings based on input characteristics. Through continuous parameter updates during contrastive training, RKAN progressively refines the alignment between modalities. Extensive experiments conducted on two publicly available datasets, GSE240429 and HER2+, demonstrate the effectiveness of our approach and its significant improvements over existing methods. Source codes are available at <https://github.com/202324131016T/DANet>.

Graphical Abstract



Keywords: gene expression prediction; dynamic alignment; state space models; neural network

Introduction

Predicting spatial gene expression by the pathological histological image is of great significance to clinical diagnosis and scientific research [1–3]. First, the method gets rid of the dependence on expensive gene sequencing, and can quickly obtain gene expression information, which greatly promotes the development of clinical diagnosis and scientific research. Second, the combination of morphological features in the pathological histological image and gene expression can help to understand the complex relationship between biological characteristics and genes of pathological tissues more comprehensively, and provide more abundant biological basis for accurate diagnosis and treatment. Additionally, analyzing the correlation between the pathological histological image and gene expression of spot can help identify novel biologic therapeutic targets and facilitate the development of personalized medicine. The task of predicting spatial gene expression through the pathological histological image not only improves the efficiency and accuracy of clinical diagnosis, but also provides new perspectives and tools for disease research and treatment [4–6].

In recent years, some scholars have begun to pay attention to related fields of research, and some existing methods have made good progress in the above tasks, such as ST-Net [7], HisToGene [8] and BLEEP [9]. However, they still face many challenges: First, in pathology images, it is difficult to efficiently extract refinement features of local areas due to the density and complexity of the cells. Second, traditional neural networks [e.g. convolutional

neural networks (CNNs) and multilayer perceptrons (MLPs)] primarily focus on local spatial patterns in data. However, gene-gene interactions are often driven by global biological mechanisms. Since traditional architectures lack the ability to directly model long-range dependencies, they struggle to effectively capture gene-gene correlations. Finally, existing methods seldom pay attention to the importance of bimodal data dynamic alignment for the task of spatial gene expression prediction. In this paper, we try to solve the above problems. Specifically, the novelty and characteristic are as follows:

- We propose a novel spatial gene expression prediction method and try to introduce the dense connective structure into this task, which realizes efficient reuse of features and is conducive to capturing and mining local refinement features.
- In order to better model the gene-gene correlations, we introduce the state space models (SSMs), which can reveal underlying patterns and capture dependencies within 1D data.
- Considering the importance of bimodal data dynamic alignment, we design the Residual Kolmogorov-Arnold Network (RKAN) module that uses a learnable activation function to dynamically adjust bimodal mappings based on input characteristics and flexibly capture complex relationships in high-dimensional data.
- We conducted comparison and ablation experiments on two public datasets, the GSE240429 dataset and the HER2+ dataset. Compared with the state-of-the-art method BLEEP [9], our method DANet performs better.

Related work

In order to efficiently and accurately obtain spatial gene expression through the histology image, various traditional spatial transcriptomics methods have emerged in the past period of time, including: MERFISH [10], Visium [11], smFISH [12], STARmap [13], seqFISH+ [14], and Targeted ExSeq [15]. However, these methods that are not based on deep learning are often low throughput or low content. They also tend to be time-consuming, expensive and require specialized equipment and extensive domain expertise to optimize.

With the rapid progress of deep learning-related technologies, some neural network-based methods to predict the spatial gene expression have been proposed. Several existing approaches have shown promising results in predicting expression from histology images including HE2RNA [16] and hist2rna [17]. The above two methods focus on clinical applications and gene expression prediction for specific cancer types, primarily for cancer diagnosis and treatment decisions. In addition to this, there have been promising advances in some approaches applicable to broader tissue research and clinical applications, such as: ST-Net [7], HisToGene [8] and BLEEP [9]. The ST-Net [7] aims to predict gene expression from histological images of tissue sections. It demonstrates good performance on breast cancer datasets, capturing the heterogeneity within tumors and showing potential in predicting the spatial co-localization of tumor and immune-related genes. Besides, the HisToGene [8] is also designed to accomplish the same task. The model employs the Vision Transformer architecture, capable of capturing the spatial dependencies of gene expression, which helps reveal the molecular characteristics of tissues. However, the performance of the aforementioned methods relies on high-quality training data. Their prediction accuracy is significantly affected by experimental noise and sequencing depth, resulting in poor performance with low-quality data. Additionally, obtaining large-scale, high-quality training data is challenging due to the complexity and cost of spatial transcriptomics technology. Furthermore, these methods have limitations when dealing with different tissue types or disease states, exhibiting poor generalization capabilities. Recently, Ronald Xie et al. have proposed a bimodal embedding framework called BLEEP [9]. This framework uses contrastive learning to construct a low-dimensional joint embedding space for aligning image features and gene expression features. BLEEP demonstrates high prediction accuracy and is currently the state-of-the-art method on the GSE240429 dataset. It not only preserves the biological heterogeneity in the original data but also shows robustness against experimental artifacts and batch effects. However, its effectiveness has not yet been verified on related datasets.

Methods

In this paper, we focus on the task of predicting spatial gene expression through the Hematoxylin and Eosin (H&E) histology image. Meanwhile, we aim to address existing challenges and achieve accurate gene expression prediction. The contrastive learning-based method we employ has significant advantages over methods that directly predict gene expression. First, gene expression data is characterized by high dimensionality and high noise, making direct regression prediction susceptible to sparsity and outliers. In contrast, contrastive learning constructs similarity relationships between image-gene pairs, transforming the problem into implicit association modeling, which significantly enhances the model's robustness. Secondly, there are often

complex nonlinear mappings between pathological features and gene expression, which traditional methods struggle to capture in terms of the synergistic regulatory relationships of multi-scale features. Contrastive learning, however, aligns cross-modal representations to autonomously uncover potential association patterns between morphological features and gene modules [18, 19]. Lastly, the contrastive learning-based approach decouples feature learning from prediction tasks, allowing the model to focus on biologically meaningful cross-modal associations. This achieves dynamic alignment of bimodal data and enables more efficient utilization of limited data. The overall structure of our method, DANet, is shown in the Fig. 1.

Encoder

The DenseNet [20–23] is a type of deep CNN that connects each layer directly to all previous layers. This design effectively alleviates the vanishing gradient problem and promotes feature reuse, enhancing feature representation capabilities. Due to feature reuse, DenseNet requires fewer parameters to achieve the same performance, offering high parameter efficiency, and reduces overfitting by increasing path diversity. In the task of predicting spatial gene expression through the pathological histological image, it is difficult to efficiently extract refinement features of local areas due to the density and complexity of the cells in the pathology image. Therefore, we try to introduce the DenseNet and rely on its efficient feature reuse to capture and mine locally refinement features. As shown in Fig. 1 (left), E_I represents the vector obtained by embedding the histology image, and F_I represents the features extracted from E_I using the DenseNet (the Image Encoder).

In addition, existing methods are challenging to reveal underlying patterns, capture dependencies within 1D data, and model the gene-gene correlations. Traditional neural networks (such as CNN and MLP) primarily focus on local spatial patterns in data, but interactions between genes are often driven by global biological mechanisms. When modeling gene expression data, the latent relationships between genes often involve interactions across the entire genome and between different spots. Moreover, traditional networks lack the ability to directly model long-range dependencies between genes. The SSM, through its selective state transition mechanism (state transformation equation), enables the model to focus specifically on key gene associations while reducing the weight of information transmission for irrelevant or weakly associated genes. This mechanism adaptively captures long-range dependencies between genes, efficiently modeling gene-gene correlations. In order to solve the above problem, we introduce the Mamba based on the SSMs. Mamba [24] is a selective structural SSM that alleviates the modeling constraints of CNNs through global receptive field and dynamic weighting. Moreover, the SSMs [25–27] can effectively capture and maintain dependency relationships [28, 29] without significantly increasing computational complexity. The processing details are given as follows:

$$F_G = fc(SSM(\sigma(conv(fc(E_G)))) \otimes \sigma(fc(E_G))), \quad (1)$$

where the E_G is a vector obtained from the spatial gene expression after data preprocessing and embedding. In addition, \otimes means the element-wise multiplication, $fc()$ means the fully connected layer, $conv()$ denotes the 1-D convolution operation, $SSM()$ means the SSMs and the σ is the Sigmoid Linear Unit (SiLU) activation

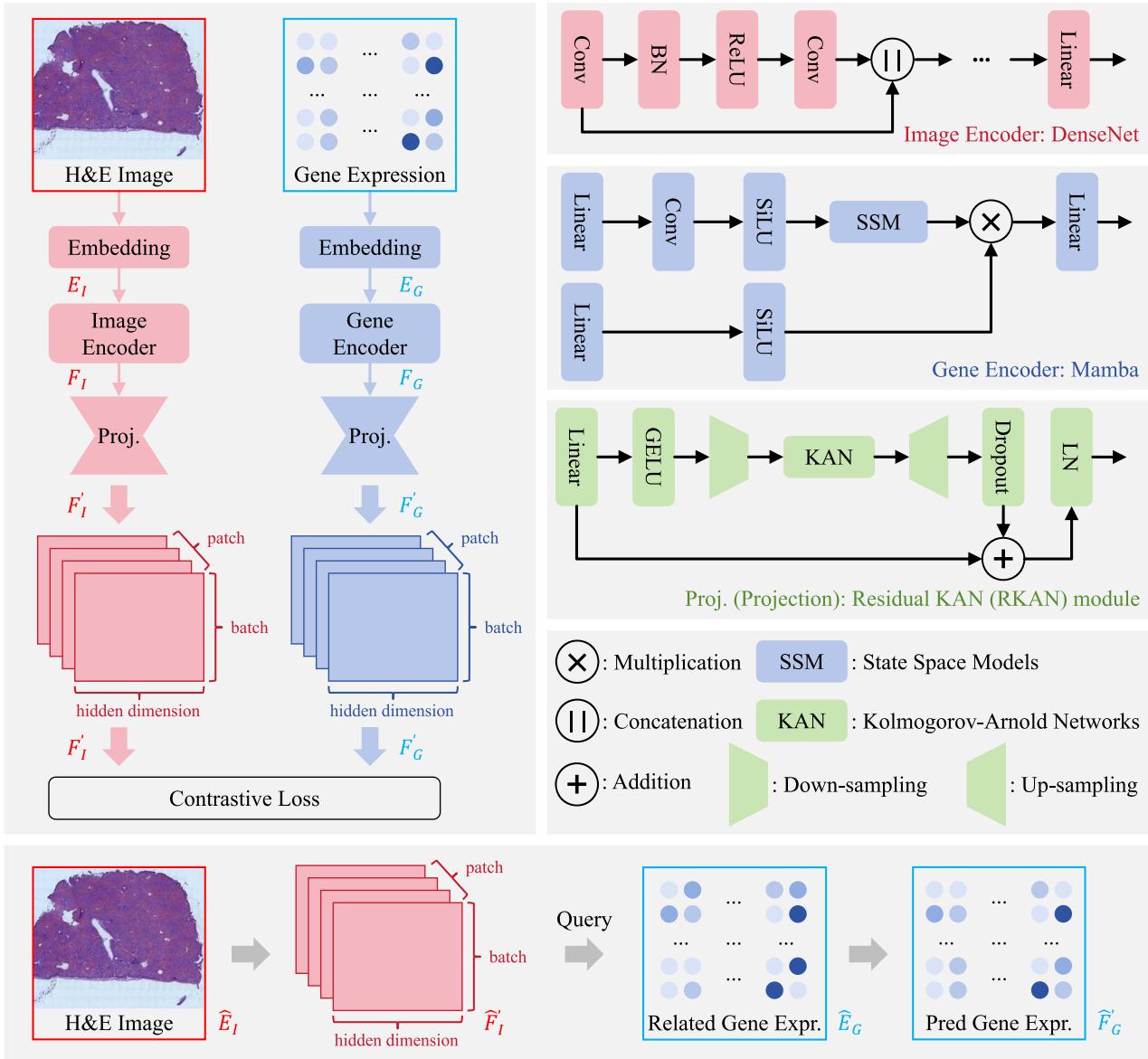


Figure 1. The overall structure of our DANet consists of a training process, where H&E histology images and real spatial gene expressions are embedded, encoded separately by the Image Encoder and Gene Encoder, and then projected into a latent space by the proposed RKAN module for contrastive learning, and a prediction process, where features from H&E histology images are projected into the latent space to query related spatial gene expressions and generate the predicted expression.

function. Finally, as shown in Fig. 1 (left), the F_G is the feature acquired by the Mamba module (the Gene Encoder).

Projection

In CLIP [30], the projection helps map features of different scales to the same dimension, enabling subsequent contrastive learning to effectively measure the similarity between them and drive model training. With the deepening of the research on artificial neural networks, more and more scholars begin to pay attention to how to flexibly capture and represent complex relationships in high-dimensional data. Among these methods, self-learning non-linear mapping structures [Kolmogorov-Arnold Networks, KAN [31]] are more flexible than traditional MLP structures that rely on expert knowledge. In our DANet, we propose the RKAN module as the projection. Based on the Kolmogorov-Arnold theorem, KAN efficiently decomposes complex functions. Unlike traditional

networks, KAN places learnable activations on edges (weights) instead of nodes (neurons), enabling flexible weight adaptation for complex data relationships. Moreover, some methods [32–35] have made a series of improvements on the basis of KAN, achieving significant progress.

In our work, we attempt to achieve semantic association and matching between two biologically related but distinct modalities (histology images and spatial gene expression) through dynamic alignment in the model's representation space. This alignment allows the model to understand their relationships, facilitating gene expression prediction from histology images. Notably, the correlation patterns between tissue morphology and gene expression vary significantly across different regions, making predefined fixed mapping rules insufficient. Dynamic adjustment of mapping rules is therefore essential. To address this, inspired by existing successful methods [36–39], we design the RKAN module as the projection for both modalities (as shown in Fig. 1). The RKAN

module employs learnable activation functions that dynamically adjust mapping parameters based on input characteristics, flexibly capturing complex data relationships and enabling dynamic alignment of bimodal features. The detailed calculation process is as follows:

$$\begin{aligned} F'_I &= \text{LN}(\text{Drop}(U(\text{KAN}(D(\mu(\text{fc}(F_I)))))) + F_I), \\ F'_G &= \text{LN}(\text{Drop}(U(\text{KAN}(D(\mu(\text{fc}(F_G)))))) + F_G), \end{aligned} \quad (2)$$

where the μ is the Gaussian Error Linear Units (GELU) activation function, the $D()$ means the Down-sampling operation, the $U()$ means the Up-sampling operation and the KAN means the Kolmogorov-Arnold Networks. In addition, the $\text{Drop}()$ and $\text{LN}()$ represent the Dropout (0.1) and the Layer Normalization, respectively. Finally, as shown in Fig. 1, the features of both modes are mapped to 256-dimensional vectors (F'_I and F'_G) for subsequent contrastive learning.

Contrastive loss

The Contrastive Language-Image Pre-training (CLIP) is a multi-modal pre-training model, which maps images and texts to the same feature space through comparative learning, so as to realize cross-modal understanding and matching. The model can align medical images with text information and improve the interpretability and accuracy of the model. Moreover, it performs well in tasks such as medical image classification and anomaly detection [30, 40, 41]. In addition, its versatility and transfer learning ability make it have wide application potential in cross-modal tasks of medical imaging. In the BLEEP [9], the model structure and the Contrastive Loss [30] are first introduced to the task of predicting spatial gene expression through the pathological histological image. In our method, we use the same loss function, which is calculated as follows:

$$\begin{aligned} \mathbf{S} &= \frac{\text{sim}(F'_I, (F'_G)^T)}{\tau}, \\ \mathbf{T} &= \text{softmax}\left(\frac{\text{sim}(F'_I, (F'_I)^T) + \text{sim}(F'_G, (F'_G)^T)}{2\tau}\right), \\ \mathcal{L} &= \frac{1}{2} (\mathcal{L}_{CE}(\mathbf{S}, \mathbf{T}) + \mathcal{L}_{CE}(\mathbf{S}^T, \mathbf{T}^T)), \end{aligned} \quad (3)$$

where, the $\text{sim}()$ means the matrix similarity calculated by matrix multiplication, the τ represents the scaling factor, and the notation $(\cdot)^T$ represents the transpose of a matrix. Besides, the \mathbf{S} represents the similarity scores between the histology image and the spatial gene expression. The \mathbf{T} represents the target probability distribution of the matching histology image and spatial gene expression. The model's predicted distribution should be as close as possible to this target probability distribution. Finally, the Contrastive Loss \mathcal{L} is calculated by the Cross-Entropy loss functions \mathcal{L}_{CE} .

Spatial gene expression prediction

The detailed process of spatial gene expression prediction is shown at the bottom of Fig. 1. In this step, we follow the processing of the BLEEP [9]: First, extract the features from the H&E histology image (\hat{F}_I) in the test set and map them to a 256-dimensional vector (\hat{F}'_I). Next, use the obtained vector \hat{F}'_I as an index to query the ground truth (GT) (\hat{E}_G) of the related spatial gene expression in the training set. Finally, obtain the predicted spatial gene expression \hat{F}'_G .

Experimental results

Datasets

To fully compare the performance of various approaches, we conduct extensive experiments on two popular public datasets, the GSE240429 dataset [42, 43] and the HER2+ dataset [44].

For the GSE240429 dataset, its data consists of four consecutive 16 micrometer thick slices of human liver tissue from neurologically deceased donor livers suitable for transplantation that were OCT embedded, frozen, sliced with a cryostat and imaged using the 10 \times Genomics Visium platform [43]. After quality control, the slices contain 2378, 2349, 2277, and 2265 spots respectively. We adopt the processing method of the BLEEP [9], the image patches of the H&E histology image are 224 \times 224, and there are a total of 3467 genes used for training and prediction. The HER2+ dataset consists of histology images, gene expressions at spatial spots, and their coordinates. Specifically, there are 36 tissue sections from eight patients, the image patch of the histology image is 112 \times 112, and there are a total of 785 genes.

In this task, spatial gene expression data represents the expression levels of multiple genes in localized regions of pathological tissue, where each continuous data corresponds to the expression level of one gene. For example, in the GSE240429 dataset, a tissue section is divided into 2277 spots, and the expression levels of 3467 genes are measured for each spot. Besides, the H&E staining method is widely used in pathology to clearly visualize the morphological structures of tissue cells under a microscope, facilitating pathological observation and diagnosis. All pathological images in both datasets were prepared using standard H&E staining protocols.

Evaluation metrics

In the task of predicting spatial gene expression through the pathological histological image, we measure the performance of various methods by two common metrics, the Pearson correlation coefficient (PCC) [45] and the adjusted rand index (ARI) [46]. The accuracy of predicted spatial gene expression is measured by the PCC. It measures the linear relationship between two variables, ranging from -1 to 1. A value of 1 indicates perfect positive correlation, -1 indicates perfect negative correlation, and 0 indicates no correlation. In addition, the ARI can measure the quality of clustering results. It is used to assess the similarity between clustering results and the GT. Its value ranges from -1 to 1. The closer the value is to 1, the more consistent the clustering result is with the GT; the closer the value is to -1, the more inconsistent the clustering result is with the GT, and the distributions are opposite.

Implementation details

To ensure a fair comparison, we ensure that the experimental settings are reasonable and consistent. First, in the GSE240429 dataset, the division of the training set and the test set is exactly the same as the method BLEEP [9]. In addition, we adjust some hyperparameters to meet the needs of our model and equipment. Specifically, the batch size is set to 256, the learning rate is set to 0.001, and our model is able to converge within fewer epochs (25 epochs). Secondly, in the HER2+ dataset, to fully utilize all the data, unlike the data partitioning from previous methods [8], we do not remove any data and have adopted a more reasonable partitioning method. Specifically, among the 36 tissue sections, 8 tissue sections with manually annotated GT are used as the test set, while the remaining data are used as the training set.

Table 1. The average PCC of predicted spatial gene expression for 8 MGs, top 50 most HEGs and top 50 most HVGs on the GSE240429 dataset

Method	MG	HEG	HVG
HisToGene	0.097±0.015	0.072±0.018	0.071±0.011
ST-Net	0.099±0.020	0.126±0.005	0.091±0.007
BLEEP	0.217±0.002	0.175±0.016	0.173±0.011
DANet (Ours)	0.265±0.003	0.230±0.023	0.200±0.013

The correlation values show the average of three replicate experiments. Accompanied uncertainty values denote the maximum difference from the mean of the three replicate experiments.

Besides, the learning rate of our model is set to 0.0002 in this dataset, and our model will converge in 25 epochs. More important experimental settings can be found in [Supplementary Table S1](#). Moreover, the inference times of BLEEP and DANet were measured using an NVIDIA L20 GPU (48G). On the GSE240429 dataset, BLEEP required ~3.1 minutes for inference, while DANet took around 3.5 minutes for inference. On the HER2+ dataset, BLEEP completed inference in 3.4 minutes, whereas DANet required 4.2 minutes for inference. Overall, both methods exhibited comparable computational costs, falling within a practical and acceptable range.

Results on GSE240429 dataset

State-of-the-art comparison

[Table 1](#) shows the comparison results of different methods on GSE240429 dataset. The results of ST-Net [7], HisToGene [8], and BLEEP [9] are taken from Ronald Xie et al. [9]. At the same time, the results are consistent with the BLEEP [9] report. The correlation values show the average of three replicate experiments. Accompanied uncertainty values denote the maximum difference from the mean of the three replicate experiments. In addition, we also report the average PCC of predicted spatial gene expression for 8 marker genes (MGs) derived from Andrews et al. [42], top 50 most highly expressed genes (HEGs) and top 50 most highly variable genes (HVGs) on the GSE240429 dataset. From the last two rows of [Table 1](#), it can be seen that compared to the current state-of-the-art method BLEEP, our proposed DANet achieves the best performance in all three average PCCs. Specifically, it improves by 0.048 (22%) on MG, 0.055 (31%) on HEG, and 0.027 (15%) on HVG. This demonstrates the superiority of our method and also proves that our proposed DANet can efficiently extract more discriminative features for spatial gene expression prediction from the pathological histology images.

[Table 2](#) shows the PCC of predicted spatial gene expression for top-5 genes and their average values on the GSE240429 dataset. We follow the BLEEP and report the representative results of one replicate experiment. Besides, the results of the first three methods are reported by Ronald Xie et al. [9]. Meanwhile, the data presented in black bold is the most effective among all methods. It can be observed that our proposed DANet outperforms other methods. In detail, our DANet achieves the best or second-best performance across all metrics, while also attaining the highest average value (Avg.) for the top-5 PCC.

To more intuitively compare the performance of our proposed method, [Fig. 2](#) presents visualization results comparing the predictions of BLEEP and DANet (Ours) with the GT for the expression levels of four genes (CYP3A4, ORM1, CYP2E1, and CYP1A2) on the GSE240429 dataset. The gene expression results shown in [Fig. 2](#) are all derived from the top-5 in [Table 2](#). Overall, both our

DANet and the current state-of-the-art method BLEEP demonstrate promising performance, with predicted gene expressions closely approximating the GT. In certain details, DANet outperforms BLEEP. For example, comparing the last three subfigures in [Fig. 2](#), the “bright spots” in GT ([Fig. 2j](#)) and DANet ([Fig. 2l](#)) appear denser, while the results from BLEEP ([Fig. 2k](#)) are slightly sparser. We provide additional comparative visualization results, as shown in [Supplementary Fig. S1](#). By comparing the three visualizations in each row, we can observe that DANet’s detail processing is closer to GT compared to BLEEP.

To further compare the performance of our method with the state-of-the-art method BLEEP, we report the clustering results (ARI) for both methods, as shown in [Fig. 3](#). It can be observed that our method achieves better performance (0.175 at the [Fig. 3c](#)). The [Fig. 3](#) presents clustering results based on gene expression data, categorizing the 2277 spots into 5 distinct biologically relevant regions or cell types. The clustering results were overlaid onto H&E histology images using color-coding, revealing the spatial relationship between gene expression data and tissue morphology. In the upper-left corner (the red regions) of the GT ([Fig. 3a](#)), there are experimental artifacts. It can be observed that the predictions of our DANet and BLEEP for this region are relatively similar, which also demonstrates DANet’s capability to handle low-quality data. In our model, the SSM and KAN can model the global dependencies of the data while flexibly capturing complex relationships in high-dimensional data, providing a foundation for effectively handling low-quality data.

[Figure 4](#) shows a comparison between predicted gene expression profiles and reference gene expression profiles normalized by gene count means or gene count variances for BLEEP and DANet. [Figure 4a and b](#) are plotted after normalizing by gene expression means and sorting according to the true gene expression values of the reference. The closer the red scatter points (Predicted Values) are to the blue line (GT), the more accurately the method captures the average gene expression levels, highlighting the relative expression levels. From the enlarged local region in the upper left corner, it can be observed that the scatter points (Predicted Values) of our DANet are closer to the blue line (GT), indicating that it can better balance the expression levels of all genes. [Figure 4c and d](#) are plotted after normalizing by gene expression variance and sorting according to the true gene expression values of the reference. Here, the closer the red scatter points (Predicted Values) are to the blue line (GT), the better the method preserves the spatial heterogeneity of gene expression, emphasizing the degree of spatial variation. Similarly, as seen in the enlarged local region in the upper left corner, the scatter points (Predicted Values) of our DANet are closer to the blue line (GT), indicating its superior ability to retain biologically relevant local expression differences. In addition, the [Supplementary Fig. S2](#) displays gene-gene correlation heatmaps calculated using the predicted expressions for GT, BLEEP, and DANet (Ours) on the GSE240429 dataset. The color gradients in the heatmaps visually represent the correlations in spatial expression patterns between different genes. By comparing the heatmaps of GT, BLEEP, and DANet, we can verify whether the predicted results preserve true biological relationships.

Ablation study

The results of ablation experiments on GSE240429 dataset are shown in [Table 3](#). Variables tested include: smoothed objective versus the original CLIP objective (Smoothed Obj.), choice of top K (Top-K), the method of aggregation (Aggregation) and the modules we proposed (Modules). From the first half of the table, it can be

Table 2. The PCC of predicted spatial gene expression for top-5 genes and their average values (Avg.) on the GSE240429 dataset

HisToGene		ST-Net		BLEEP		DANet	
Gene Name	PCC	Gene Name	PCC	Gene Name	PCC	Gene Name	PCC
CYP3A4	0.549	CYP3A4	0.549	CYP3A4	0.741	CYP3A4	0.739
CYP1A2	0.542	CYP1A2	0.532	CYP1A2	0.681	ORM1	0.681
GLUL	0.488	CYP2E1	0.530	CYP2E1	0.675	CYP2E1	0.673
CYP2E1	0.330	GLUL	0.463	GLUL	0.656	CYP1A2	0.661
FABP1	0.328	SLCO1B3	0.375	FABP1	0.503	GLUL	0.624
Avg.	0.447	Avg.	0.490	Avg.	0.651	Avg.	0.676

We follow the BLEEP and report the representative results of one replicate experiment. Besides, the data presented in black bold is the most effective among all methods.

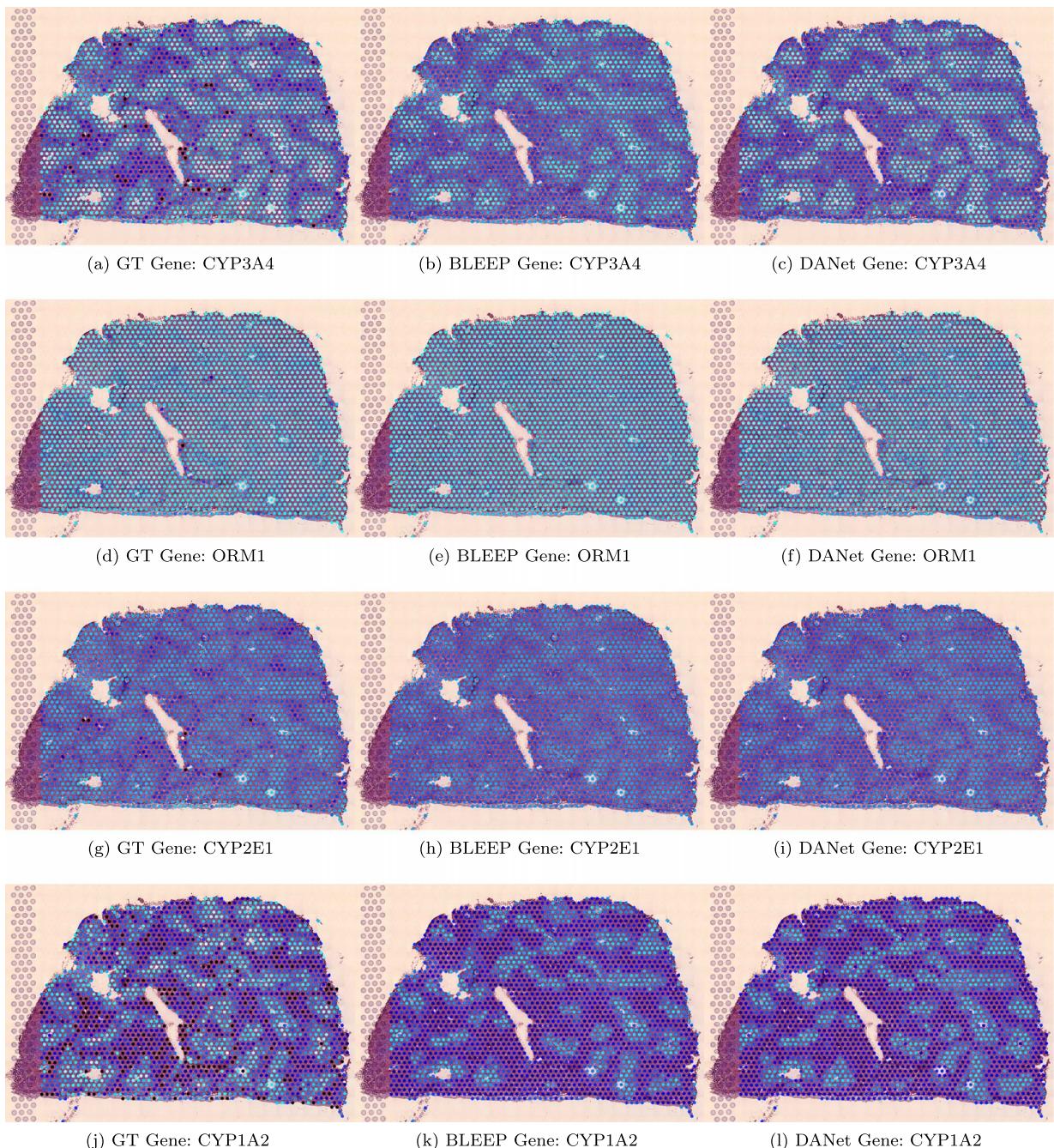


Figure 2. Visualization results comparing the predictions of BLEEP and DANet (Ours) with the GT for the expression levels of four genes CYP3A4, ORM1, CYP2E1, and CYP1A2 on the GSE240429 dataset.

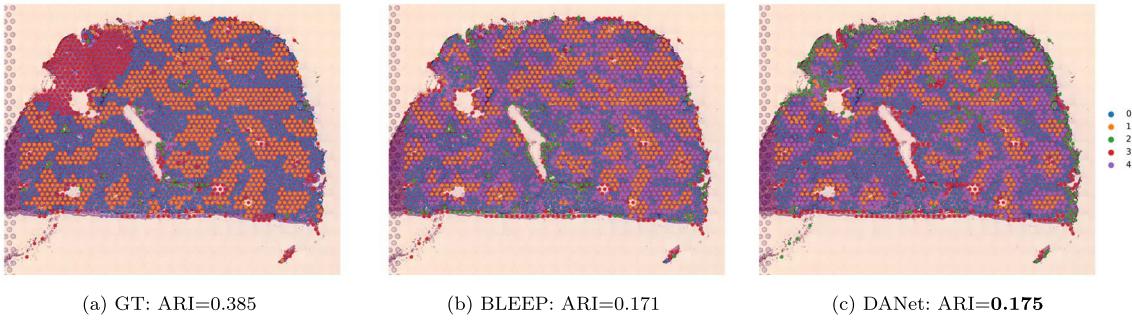


Figure 3. The comparative clustering results of predicted spatial gene expression for GT, BLEEP, and ours on the GSE240429 dataset.

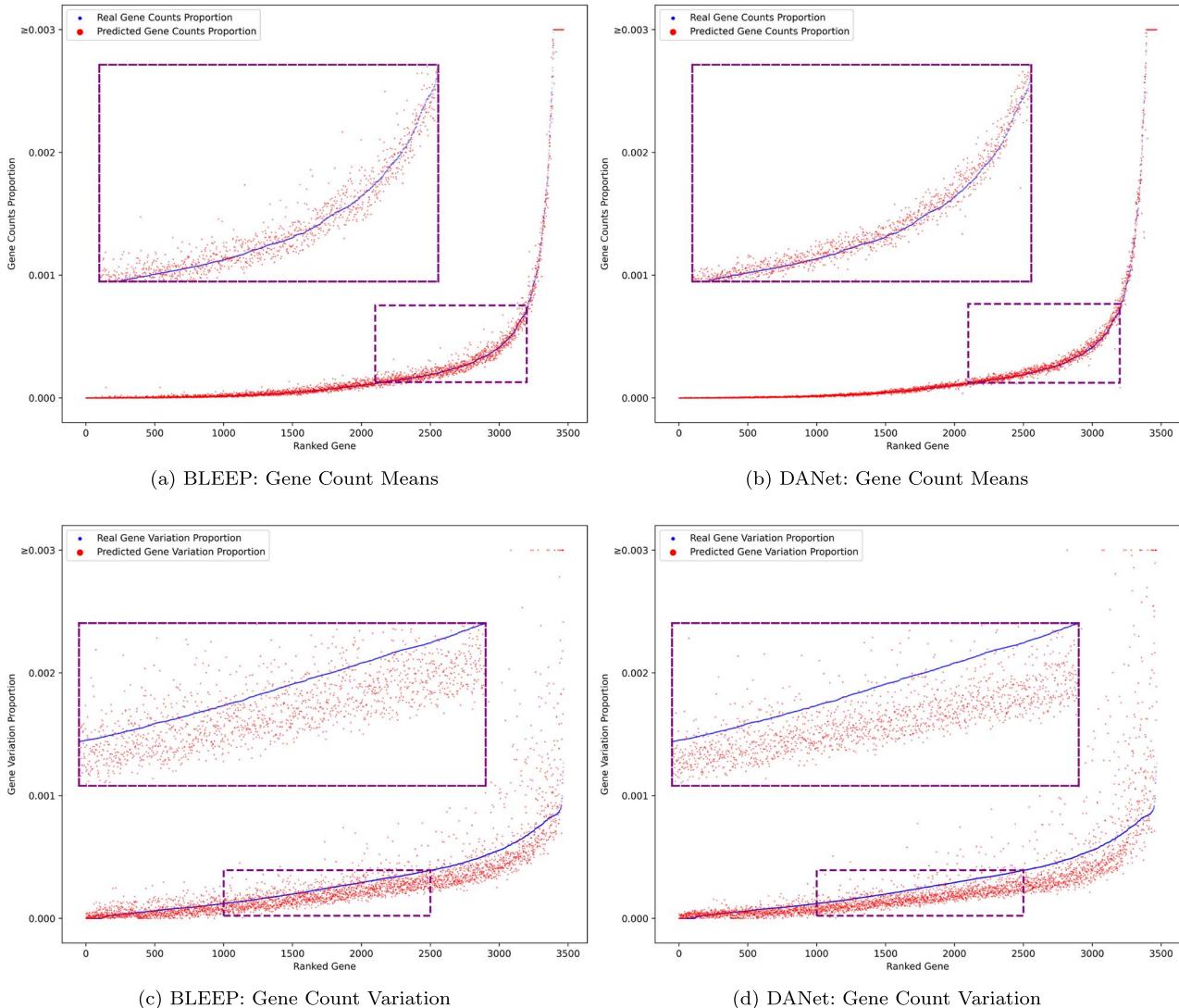


Figure 4. Comparison of predicted and real spatial gene expressions on the GSE240429 dataset, normalized by gene count means (upper panels) or gene count variance (lower panels), with the upper left corner of each subplot showing a zoomed-in view of the corresponding local region.

seen that our method achieves the best result in the fourth row, i.e. by using the smoothed objective with Top-K set to 100, and the method of aggregation is set to average. The last three lines show the overall impact of the different modules we propose on the method. Specifically, when specific modules are removed, the performance of the method decreases to varying degrees. This phenomenon is most evident when the SSM module is removed,

followed by our designed RKAN module. The results indicate that our proposed improvements can effectively address the current challenges. Moreover, mining local refinement features, identifying gene-gene correlations, and achieving dynamic alignment of bimodal data collectively contribute to improving the task of predicting spatial gene expression from pathological histological images.

Table 3. The results of ablation experiments on the GSE240429 dataset

Smoothed Obj.	Top-K	Aggregation	Modules			PCC		
			DenseNet	Mamba	RKAN	MG	HEG	HVG
Yes	-	Simple	✓	✓	✓	0.120±0.009	0.107±0.033	0.086±0.011
Yes	50	Average	✓	✓	✓	0.255±0.013	0.223±0.025	0.190±0.012
Yes	50	Weighted	✓	✓	✓	0.223±0.014	0.189±0.038	0.164±0.022
Yes	100	Average	✓	✓	✓	0.265±0.003	0.230±0.023	0.200±0.013
Yes	100	Weighted	✓	✓	✓	0.229±0.014	0.195±0.039	0.171±0.022
No	100	Average	✓	✓	✓	0.025±0.045	0.037±0.031	0.026±0.025
Yes	100	Average		✓	✓	0.252±0.005	0.204±0.013	0.187±0.006
Yes	100	Average	✓		✓	0.232±0.008	0.185±0.021	0.177±0.017
Yes	100	Average	✓	✓		0.250±0.017	0.193±0.018	0.184±0.014

We report the average of three replicate experiments, and the accompanied uncertainty values denote the maximum difference from the mean of the three replicate experiments.

Results on HER2+ dataset

State-of-the-art comparison

To fully assess the generalization capabilities of various methods (ST-Net [7], HisToGene [8], BLEEP [9] and ours), we conducted extensive experiments on a publicly available relevant dataset (the HER2+ dataset), and the results are shown in Tables 4 and 5. As we mentioned in (Section Implementation details), we adopt different ways of dividing the training and test sets on the HER2+ dataset, i.e. 8 tissue slices with manually annotated GT were used as the test set (A1, B1, C1, D1, E1, F1, G2, and H1), while the remaining data served as the training set. Additionally, we conducted three replicate experiments for the four methods on the HER2+ dataset and reported the mean and standard deviation (Mean±SD). Replicate experiments can effectively reduce the impact of random errors on the results, further verifying the stability and reliability of different methods. The combined use of the mean and standard deviation not only reflects the central tendency of the experimental results but also clearly presents the degree of dispersion of the experimental results, enabling a more comprehensive evaluation of the accuracy and robustness. Moreover, in the Tables 4 and 5, the * indicates the results that are reproduced under the new dataset partitioning. Besides, the data presented in black bold is the most effective among all methods.

Table 4 presents the mean and standard deviation (Mean±SD) of the PCC and ARI for predicted spatial gene expression across all genes in the HER2+ dataset, calculated from three replicate experiments. Here, the Avg. denotes the mean value of PCCs and ARIs from eight test sets in each column. Besides, it can be observed that all methods exhibit some degree of data fluctuation, but these fluctuations generally fall within an acceptable range. Overall, our method achieved the best or second-best results on all eight test sets of the HER2+ dataset, demonstrating the robustness and superiority of our approach.

Table 5 presents the PCC of predicted spatial gene expression for top-5 genes and their average values on the HER2+ dataset. Here, the Avg. represents the average of the top-5 PCCs in each row. Besides, to maintain consistency with Table 2 and due to the inability to display excessive data in a single table, we have only presented the results from one experiment in Table 5. The complete results from the three replicate experiments are shown in Supplementary Tables S2–S4. In the Table 5, it can be observed that our method achieved the best results in 5 out of 8 averages and the second-best results in 2 out of 8 averages. Specifically, our method demonstrates significant advantages on the F1 and D1 test sets and achieves strong performance on the B1, E1, and

H1 test sets. In contrast, it performs the worst on the G2 test set, but the gap compared to the best result is marginal (only 0.038 lower). Moreover, we presented the PCCs of predicted spatial gene expression for the top-5 genes and their average values from another two replicate experiments on the HER2+ dataset, as shown in Supplementary Tables S2 and S3. Additionally, in Table S4 of the supplementary material, we reported the mean and standard deviation (Mean±SD) of PCCs for the predicted spatial gene expression of the top-5 genes on the HER2+ dataset from three replicate experiments. The extensive experimental data in the above tables demonstrate that our method performs superiorly and stably in most cases.

Ablation study

Similarly, to analyze the effect of the proposed modules, we conduct ablation experiments on the HER2+ dataset for the modules we proposed, and the results are shown in the Table 6. The average value in the last column can more intuitively reflect the performance of the model. By comparing the first three rows with the last row, it can be seen that the performance of our model significantly declines after removing any one of the modules, whether in terms of PCC or ARI. Specifically, after removing the DenseNet and Mamba, the PCC and ARI significantly decreased on almost all 8 test sets. In addition, the performance of the model was significantly improved on the three test sets E1, F1 and H1 after adding the RKAN module. More importantly, in terms of the average values (Avg.) of the two metrics across the 8 test sets, the three modules we introduced and designed to address the numerous challenges of the spatial gene expression prediction task can complement each other effectively, yielding remarkable results. This also demonstrates the effectiveness of the modules and methods we proposed.

Discussion and conclusion

In this article, we have proposed a novel spatial gene expression prediction method, termed DANet. To capture and mine local refinement features, we attempt to introduce the dense connective structure to achieve efficient reuse of feature. Meanwhile, relying on the ability of the SSM to reveal underlying patterns and capture dependencies within 1D data, the Mamba is employed to model the correlations of gene-gene. In addition, we recognize the importance of bimodal data dynamic alignment for the task of spatial gene expression prediction, and design the RKAN module that uses a learnable activation function to dynamically adjust bimodal mappings based on input characteristics

Table 4. The mean and standard deviation (Mean \pm SD) of the PCC and ARI for predicted spatial gene expression across all genes in the HER2+ dataset, calculated from three replicate experiments

Metric	Test Set	HisToGene*	ST-Net*	BLEEP	DANet (Ours)
PCC	A1	0.025 \pm 0.002	0.005 \pm 0.001	0.042\pm0.003	0.031 \pm 0.010
	B1	0.193 \pm 0.002	0.066 \pm 0.010	0.233 \pm 0.022	0.245\pm0.002
	C1	0.107 \pm 0.002	0.029 \pm 0.007	0.154\pm0.030	0.131 \pm 0.019
	D1	0.085 \pm 0.003	0.020 \pm 0.001	0.108 \pm 0.001	0.118\pm0.015
	E1	0.002\pm0.000	-0.003 \pm 0.001	-0.011 \pm 0.006	-0.007 \pm 0.005
	F1	-0.001 \pm 0.000	0.001 \pm 0.002	0.003\pm0.003	0.002 \pm 0.011
	G2	0.113 \pm 0.001	0.025 \pm 0.005	0.120\pm0.006	0.110 \pm 0.010
	H1	0.110 \pm 0.001	0.031 \pm 0.001	0.112 \pm 0.004	0.122\pm0.003
	Avg.	0.079 \pm 0.001	0.022 \pm 0.003	0.095\pm0.009	0.094 \pm 0.009
ARI	A1	0.066\pm0.005	0.036 \pm 0.020	0.061 \pm 0.060	0.055 \pm 0.008
	B1	0.270 \pm 0.007	0.214 \pm 0.008	0.276 \pm 0.050	0.287\pm0.027
	C1	0.056\pm0.002	0.056\pm0.026	0.056\pm0.052	0.025 \pm 0.037
	D1	0.295 \pm 0.010	0.257 \pm 0.014	0.217 \pm 0.027	0.302\pm0.025
	E1	0.039 \pm 0.002	-0.015 \pm 0.012	0.042 \pm 0.030	0.049\pm0.009
	F1	0.129\pm0.004	0.023 \pm 0.013	0.048 \pm 0.024	0.092 \pm 0.052
	G2	0.147 \pm 0.010	0.144 \pm 0.007	0.099 \pm 0.015	0.165\pm0.013
	H1	0.302 \pm 0.007	0.273 \pm 0.048	0.269 \pm 0.021	0.328\pm0.016
	Avg.	0.163\pm0.006	0.123 \pm 0.018	0.133 \pm 0.035	0.163\pm0.023

The data presented in black bold is the most effective among all methods. Besides, the *indicates the results that are reproduced under the new dataset partitioning.

Table 5. The PCC of predicted spatial gene expression for top-5 genes and their average values on the HER2+ dataset

	Method	Gene	PCC	Gene	PCC	Gene	PCC	Gene	PCC	Avg.	
A1	HisToGene*	FASN	0.378	GNAS	0.275	SCD	0.271	DMAP1	0.188	AKR1A1	0.180
	ST-Net*	GATC	0.192	EZH2	0.176	FLNA	0.163	CCT4	0.162	PHF11	0.160
	BLEEP	Gene	0.422	SCD	0.366	MUC11	0.342	CXCL10	0.320	IGKC	0.287
	DANet	FASN	0.355	SCD	0.332	C3	0.278	GNAS	0.270	MUC11	0.218
B1	HisToGene*	GNAS	0.532	FN1	0.498	RHOB	0.496	GBP3	0.471	TMEM123	0.469
	ST-Net*	TMEM14B	0.484	FN1	0.440	GNAS	0.411	SCD	0.393	CPNE1	0.386
	BLEEP	TMEM14B	0.560	FN1	0.553	FAM117A	0.532	SDHB	0.512	COL4A1	0.507
	DANet	FN1	0.584	TMEM14B	0.576	FAM117A	0.563	STMN1	0.537	ZFP36	0.537
C1	HisToGene*	GNAS	0.430	CLDN4	0.391	NDUFB9	0.376	FDPS	0.346	PXDN	0.345
	ST-Net*	SAT1	0.356	CPSF1	0.269	TMEM14B	0.256	GNAS	0.255	TMEM123	0.238
	BLEEP	UFD1L	0.471	FDPS	0.470	GNAS	0.458	ZDHHC12	0.453	CLDN4	0.453
	DANet	ITGB6	0.441	GNAS	0.435	UFD1L	0.402	ZDHHC12	0.383	MRPL51	0.377
D1	HisToGene*	FN1	0.517	SCD	0.453	ITGB6	0.436	C3	0.392	SRSF1	0.389
	ST-Net*	TMEM123	0.242	SCD	0.228	FN1	0.223	BTG1	0.207	PPP1R37	0.201
	BLEEP	FN1	0.466	SCD	0.432	ITGB6	0.398	C3	0.397	SRSF1	0.397
	DANet	C3	0.523	FN1	0.518	SCD	0.495	ITGB6	0.469	TXND17	0.444
E1	HisToGene*	IGHG3	0.132	IGHM	0.129	IFI27	0.125	DNAJB2	0.114	PLXND1	0.109
	ST-Net*	SLC35A2	0.161	JCHAIN	0.137	PLVAP	0.110	CGGBP1	0.109	IGLC7	0.108
	BLEEP	IGKC	0.141	IGHG4	0.126	CD14	0.119	IGLC3	0.116	IFI27	0.114
	DANet	IGKC	0.160	IGHC2	0.153	S100A9	0.150	TMEM86A	0.122	ANKLE2	0.122
F1	HisToGene*	FASN	0.190	SCD	0.183	LUC7L3	0.178	MYL9	0.154	GNAS	0.154
	ST-Net*	GNAS	0.147	LIMCH1	0.140	JAK3	0.110	IGLC7	0.108	PLP2	0.107
	BLEEP	LUC7L3	0.173	SCD	0.149	GNAS	0.141	HCST	0.141	IGKC	0.131
	DANet	IGKC	0.341	IGLC2	0.289	FASN	0.259	IGHG3	0.233	IGLC3	0.228
G2	HisToGene*	TMEM123	0.404	GNAS	0.383	NDRG1	0.371	CLDN4	0.369	FN1	0.360
	ST-Net*	CLDN4	0.302	TMEM123	0.281	BTG1	0.234	SAT1	0.225	MIF	0.219
	BLEEP	TMEM123	0.399	GNAS	0.391	ITGB6	0.390	MYL12B	0.382	ZNF217	0.372
	DANet	GNAS	0.379	TMEM123	0.350	CX3CL1	0.348	FASN	0.337	ITGB6	0.330
	HisToGene*	GNAS	0.426	HLA-DRA	0.425	ITGB6	0.376	FASN	0.351	IGHA1	0.340
H1	ST-Net*	HLA-DRA	0.314	CPNE1	0.270	BTG1	0.252	SRSF5	0.247	GATA3	0.232
	BLEEP	IGKC	0.485	GNAS	0.439	IGHG4	0.417	HLA-DRA	0.406	MGP	0.391
	DANet	IGKC	0.481	GNAS	0.453	IGHG4	0.439	HLA-DRA	0.429	ITGB6	0.421
										0.445	

We report the representative results of one replicate experiment. The data presented in black bold is the most effective among all methods. Besides, the *indicates the results that are reproduced under the new dataset partitioning.

Table 6. The results of ablation experiments on the HER2+ dataset

Metric	Modules		A1	B1	C1	D1	E1	F1	G2	H1	Avg.	
	DenseNet	Mamba	RKAN									
PCC		✓	✓	0.020	0.200	0.131	0.071	-0.016	0.000	0.079	0.088	0.072
	✓		✓	0.034	0.224	0.147	0.119	-0.003	0.009	0.105	0.119	0.094
	✓	✓		0.024	0.233	0.142	0.107	-0.012	-0.007	0.116	0.125	0.091
	✓	✓	✓	0.042	0.246	0.150	0.134	-0.001	0.014	0.102	0.119	0.101
	✓		✓	0.024	0.241	0.064	0.140	0.012	0.014	0.079	0.181	0.094
	✓	✓		0.041	0.240	0.044	0.247	0.026	0.025	0.157	0.268	0.131
ARI	✓	✓		0.062	0.303	-0.005	0.332	0.032	0.061	0.130	0.267	0.148
	✓	✓	✓	0.051	0.257	0.011	0.281	0.044	0.145	0.180	0.332	0.163

The Avg. represents the average in each row.

and flexibly capture complex relationships in high-dimensional data. Extensive experiments conducted on two popular datasets, the GSE240429 dataset and the HER2+ dataset, demonstrate the superiority and generalization capabilities of our DANet.

In the task of predicting spatial gene expression through the pathological histological image, the core challenge lies in establishing semantic correspondence between histology images and spatial gene expression data (two biologically related yet fundamentally distinct modalities). While existing methods often employ static projection rules, we argue that predefined mappings are insufficient for this task due to significant regional variations in tissue morphology-gene expression correlations. Our RKAN module dynamically adjusts projection parameters through its trainable activation functions, allowing real-time adaptation to input characteristics. This capability is particularly critical for dynamically aligning features across modalities, as it enables the model to adaptively capture latent correlations between the two data modalities, thereby mitigating inter-modal heterogeneity bias while enhancing the generalization capability of joint representations.

Traditional spatial gene expression measurement techniques face challenges such as high costs and lengthy experimental cycles, which limit their large-scale application in patient cohorts. Deep learning-based methods can predict molecular-level spatial gene expression using easily accessible histological images, thereby overcoming the limitations of spatial transcriptomics data. These approaches enable rapid processing and analysis of pathological histology images, improving data analysis efficiency while reducing time and costs. Additionally, by learning complex relationships between different modalities to achieve dynamic alignment of bimodal data, this approach enables more precise spatial gene expression prediction. Such advancement facilitates the elucidation of disease molecular mechanisms and assists pathologists in accurately determining disease types, stages, and prognostic outcomes. Furthermore, understanding patients' spatial gene expression profiles supports clinicians in developing personalized treatment plans.

Our proposed DANet demonstrates significant potential in related tasks, yet several challenges remain to be addressed. First, while the dense connective structure enhances feature reuse efficiency, the computational overhead may increase substantially with network depth, necessitating a balance between model capacity and computational efficiency. Second, the advantages of SSM in gene relationship modeling still require validation through comparative experiments, particularly against graph neural network-based approaches. Additionally, although the KAN module possesses the potential for flexibly modeling high-dimensional nonlinear relationships, the effectiveness of

cross-modal data alignment also depends on the design of the loss function, calling for further exploration of more robust and efficient solutions.

Key Points

- We propose DANet, a deep learning-based model for spatial gene expression prediction, utilizing dense connections to realize efficient feature reuse and mining local refinement features.
- For improved modeling of gene-gene correlations, we employ the state space models, which are capable of identifying underlying patterns and capturing dependencies in 1D data.
- Leveraging Kolmogorov-Arnold Network's ability to dynamically adjust bimodal mappings based on input characteristics and flexibly capture complex relationships in high-dimensional data, we design the Residual Kolmogorov-Arnold Network module to achieve dynamic alignment of bimodal data.
- Experiments conducted on the GSE240429 and HER2+ datasets demonstrate the superiority and robustness of our proposed DANet.

Acknowledgments

We sincerely thank the anonymous reviewers for their valuable and constructive suggestions.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

This research was supported by the Science and Technology Innovation Key R&D Program of Chongqing (grant No. CSTB2023TIAD-STX001), the New Chongqing Youth Innovative Talents Project (grant No. CSTB2024NSCQ-QCXMLX0077), the National Key Research and Development Program of China (grant No. 2022ZD0160400), the Science and Technology Innovation Key R&D Program of Chongqing (grant No. CSTB2024TIAD-STX0032), and the National Natural Science Foundation of China (grant No. 62206031, 62301092, and 62402071).

Data availability

For the GSE240429 dataset, the data is publicly available for download at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE240429>. For the HER2+ dataset, the data is publicly available here: human HER2-positive breast tumor ST data <https://github.com/almaan/her2st>.

Code availability

Source codes are available at <https://github.com/202324131016T/DANet>.

References

1. Zeng Y, Wei Z, Weijiang Y. et al. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Brief Bioinform* 2022;**23**:bbac297.
2. Jia Y, Liu J, Chen L. et al. THIToGene: a deep learning method for predicting spatial transcriptomics from histological images. *Brief Bioinform* 2024;**25**:bbad464.
3. Min W, Shi Z, Zhang J. et al. Multimodal contrastive learning for spatial gene expression prediction using histology images. *Brief Bioinform* 2024;**25**:bbae551.
4. Van Darn S, Vosa U, van der Graaf A. et al. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief Bioinform* 2018;**19**:575–92.
5. Wu J, Zhao X, He Y. et al. IDMIR: identification of dysregulated miRNAs associated with disease based on a miRNA–miRNA interaction network constructed through gene expression data. *Brief Bioinform* 2024;**25**:bbae258.
6. Li X, Zhu F, Min W. SpaDiT: diffusion transformer for spatial gene expression prediction using scRNA-seq. *Brief Bioinform* 2024;**25**:bbae571.
7. He B, Bergensträhle L, Stenbeck L. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* 2020;**4**:827–34. <https://doi.org/10.1038/s41551-020-0578-x>
8. Pang M, Kenong S, Li M. Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors BioRxiv. 2021;2021:11.
9. Xie R, Pang K, Chung S. et al. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. In: *Advances in Neural Information Processing Systems* Curran Associates, Inc., New Orleans, Louisiana, USA, vol. **36**, pp. 70626–70637, 2023.
10. Chen KH, Boettiger AN, Moffitt JR. et al. Spatially resolved, highly multiplexed rna profiling in single cells. *Science* 2015; **348**:aaa6090.
11. Ståhl PL, Salmén F, Vickovic S. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;**353**:78–82. <https://doi.org/10.1126/science.aaf2403>
12. Codeluppi S, Borm LE, Zeisel A. et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods* 2018;**15**:932–5. <https://doi.org/10.1038/s41592-018-0175-z>
13. Wang X, Allen WE, Wright MA. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;**361**:eaat5691.
14. Eng C-HL, Lawson M, Zhu Q. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019;**568**: 235–9. <https://doi.org/10.1038/s41586-019-1049-y>
15. Alon S, Goodwin DR, Sinha A. et al. Expansion sequencing: spatially precise *in situ* transcriptomics in intact biological systems. *Science* 2021;**371**:eaax2656.
16. Schmauch B, Romagnoni A, Pronier E. et al. A deep learning model to predict RNA-seq expression of tumours from whole slide images. *Nat Commun* 2020;**11**:3877. <https://doi.org/10.1038/s41467-020-17678-4>
17. Mondol RK, Millar EKA, Graham PH. et al. hist2RNA: an efficient deep learning architecture to predict gene expression from breast cancer histopathology images. *Cancers* 2023;**15**:2569.
18. Zheng L, Liu Z, Yang Y. et al. Accurate inference of gene regulatory interactions from spatial gene expression with deep contrastive learning. *Bioinformatics* 2022;**38**:746–53.
19. Shen L-C, Liu Y, Liu Z. et al. Supervised contrastive learning enhances MHC-II peptide binding affinity prediction. *Expert Syst Appl* 2025;**269**:126463. <https://doi.org/10.1016/j.eswa.2025.126463>
20. Gao H, Liu Z, Van Der Maaten L. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Honolulu, Hawaii, USA, pp. 4700–8, 2017.
21. Kuang P, Ma T, Chen Z. et al. Image super-resolution with densely connected convolutional networks. *Appl Intell* 2019;**49**:125–36. <https://doi.org/10.1007/s10489-018-1234-y>
22. Li G, Zhang M, Li J. et al. Efficient densely connected convolutional neural networks. *Pattern Recognit* 2021;**109**:107610. <https://doi.org/10.1016/j.patcog.2020.107610>
23. Zhang J, Zhang Y, Jin Y. et al. MDU-Net: multi-scale densely connected U-Net for biomedical image segmentation. *Health Inf Sci Syst* 2023;**11**:13. <https://doi.org/10.1007/s13755-022-00204-9>
24. Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.
25. Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. In: *International Conference on Learning Representations*, OpenReview.net, Virtual, 2022.
26. Fu DY, Dao T, Saab KK. et al. Hungry hungry hippos: towards language modeling with state space models. In: *International Conference on Learning Representations*, OpenReview.net, Kigali, Rwanda, 2023.
27. Ali A, Zimmerman I, Wolf L. The hidden attention of mamba models. *arXiv preprint arXiv:2403.01590*. 2024.
28. Gu A, Dao T, Ermon S. et al. HiPPO: recurrent memory with optimal polynomial projections. In: *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Virtual, Vol. **33**, pp. 1474–87, 2020.
29. Gu A, Johnson I, Goel K. et al. Combining recurrent, convolutional, and continuous-time models with linear state space layers. In: *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Virtual, Vol. **34**, pp. 572–85, 2021.
30. Radford A, Kim JW, Hallacy C. et al. Learning transferable visual models from natural language supervision. In: *International Conference on Machine Learning*, PMLR, Virtual, pp. 8748–63, 2021.
31. Liu Z, Wang Y, Vaidya S. et al. KAN: Kolmogorov-Arnold networks. In: *International Conference on Learning Representations*, OpenReview.net, Singapore, 2025.
32. Drokin I. Kolmogorov-Arnold convolutions: design principles and empirical studies. *arXiv preprint arXiv:2407.01092*. 2024.
33. Hou Y, Zhang D. A comprehensive survey on Kolmogorov Arnold networks (KAN). *arXiv preprint arXiv:2407.11075*. 2024.
34. Brandt M, Chave J, Li S. et al. High-resolution sensors and deep learning models for tree resource monitoring. *Nat Rev Electr Eng* 2025;**2**:13–26.

35. Firsov N, Myasnikov E, Lobanov V. et al. HyperKAN: Kolmogorov-Arnold networks make hyperspectral image classifiers smarter. *Sensors (Basel, Switzerland)* 2024; **24**:7683.
36. He K, Zhang X, Ren S. et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Las Vegas, NV, USA, pp. 770–8, 2016.
37. Thekedath D, Sedamkar RR. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. *SN Comput Sci* 2020; **1**:79. <https://doi.org/10.1007/s42979-020-0114-9>
38. Wanni X, You-Lei F, Zhu D. ResNet and its application to medical image processing: research progress and challenges. *Comput Methods Programs Biomed* 2023; **240**: 107660.
39. Wang L, Xiangzheng F, Ye X. et al. PKAN: leveraging Kolmogorov-Arnold networks and multi-modal learning for peptide prediction with advanced language models. *IEEE J Biomed Health Inform* 2025.
40. Zhang Y, Jiang H, Miura Y. et al. Contrastive learning of medical visual representations from paired images and text. In: *Proceedings of the Machine Learning for Healthcare Conference*. PMLR, Durham, NC, USA, vol. **182**, pp. 2–25, 2022.
41. Zhao Z, Liu Y, Wu H. et al. CLIP in medical imaging: a comprehensive survey. *Medical Image Analysis* 2025; **102**:103551.
42. Andrews TS, Atif J, Liu JC. et al. Single-cell, single-nucleus, and spatial RNA sequencing of the human liver identifies cholangioocyte and mesenchymal heterogeneity. *Hepatol Commun* 2022; **6**: 821–40.
43. Andrews TS, Nakib D, Perciani C. et al. Single-cell and spatial transcriptomics reveals the human liver immunological landscape and myeloid dysfunction in psc. *bioRxiv*. 2023;2023–07.
44. Andersson A, Larsson L, Stenbeck L. et al Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat Commun*. 2021; **12**:6012.
45. Cohen I, Huang Y, Chen J. et al. Pearson correlation coefficient. *Noise Reduct Speech Process Springer Topics in Signal Processing*, 2009; **2**:1–4.
46. Rand WM. Objective criteria for the evaluation of clustering methods. *J Am Stat Assoc* 1971; **66**:846–50. <https://doi.org/10.1080/01621459.1971.10482356>