文章编号:1000-1638(2025)05-0532-13 DOI:10.13484/j.nmgdxxbzk.20250508

STConVM:基于多视图多模态的对比学习方法识别空间域*

司佳宝[1],赵翔宇[1],刘和鑫[1],代冰杰[2],冯振兴[1]
(1. 内蒙古工业大学理学院,呼和浩特 010051; 2. 内蒙古大学生命科学学院,呼和浩特 010021)

摘要:在空间转录组学的空间域识别研究中,当前多数算法在空间转录组多模态数据整合方面普遍面临多模态特征融合能力不足、识别精度有限和计算复杂度高等挑战。因此,本研究提出了一种基于多视图多模态对比学习框架的STConVM方法,旨在实现基因表达谱、空间坐标信息与组织学图像的深度融合,进而提升空间域识别的精度与鲁棒性。多个数据集的对比实验结果表明,STConVM能够更精准地识别组织功能区域,为解析生物复杂组织微环境提供了有效的工具。

关键词:空间转录组学; 空间域识别; 多视图; 多模态; 对比学习
中图分类号:Q811.4; TP391; O29 **文献标志码**:A

空间转录组学(Spatial transcriptomics,ST)技术的发展实现了基因表达谱与空间定位的同步记录[1],这类技术通过整合基因表达的空间异质性数据,为解析哺乳动物发育、疾病发生发展和组织微环境提供了前所未有的多维视角[2]。当前ST技术可分为两大分支:基于成像的原位分析技术(如MERFISH[3]、seqFISH[4]和osmFISH[5])与基于测序的空间捕获技术(如10x Visium[6]、Slide-seqV2[7]和Stereo-seq[8])。前者凭借亚细胞级分辨率优势实现荧光标记分子的精准定位,后者通过高通量测序获得深度转录组覆盖,两类技术共同构建了多模态、多尺度、多分辨率的空间组学数据体系[9-11]。

解析ST数据的前提是实现组织区域的生物学功能划分,即空间聚类任务[12-13]。该任务旨在通过整合基因表达、空间位置等多模态数据,将组织切片划分为具有功能异质性的细胞亚群,进而揭示细胞簇的生物学功能、组织重构规律及细胞间相互作用机制。现有空间域识别方法主要有两类:非空间聚类和空间聚类。经典的非空间聚类方法K-means[14]、Louvain[15]和Seurat[16]主要依赖基因表达矩阵进行无监督聚类。尽管此类方法对数据模态具有普适兼容性,但其聚类结果缺乏对空间信息和组织学图像的考虑,难以捕捉空间位置隐含的生物学关联。为突破非空间方法的局限性,研究者通过引入空间位置信息提升聚类精度。例如Giotto[17]构建隐马尔可夫随机场(HMRF)模型,联合基因表达计数矩阵与物理坐标实现空间邻域依赖建模。SEDR利用双网络架构,通过深度自编码器学习基因表达特征,结合变分图自编码器(VGAE)捕获空间位置的图结构信息,生成低维潜在嵌入[18]。STAGATE融合基因表达与空间坐标,借助图注意力机制强化局部邻域特征交互,实现空间域的识别[19]。SpaNCMG通过构建邻域互补混合视图的图卷积网络,实现空间域的精准识别[20]。Space-

* **收稿日期**:2025-05-08; **修回日期**:2025-06-23
基金项目:内蒙古自然科学基金面上项目(2024MS06027);自治区直属高校基本科研业务费项目(JY20230067);自治区级大学生创新创业训练计划项目(S202410128012)
作者简介:司佳宝(2000—),男,内蒙古赤峰人,2023级硕士研究生。E-mail: 2822883813@qq.com
通信作者:冯振兴(1988—),男,河南遂平人,副教授,博士。主要从事生物数学、生物统计学算法开发等方面研究。E-mail: zxfeng@imut.edu.cn

Flow利用两层图卷积神经网络,结合对比学习和正则化策略优化特征表示,提升了模型聚类鲁棒性[21]。GraphST采用自监督对比学习与图卷积神经网络联合建模,通过挖掘图结构的依赖关系,增强特征学习,实现了更精准的空间域划分[22]。

在ST数据分析中,基因表达、空间坐标与组织学图像等多模态信息共同构成了生物特征解析的核心维度。近年来,随着多模态整合技术的发展,一系列的方法通过不同策略实现了图像数据与基因表达、空间位置的融合,推动了ST空间域识别算法的进步。stLearn提出SME标准化方法,整合了基因表达矩阵、空间坐标及组织学图像的RGB特征,构建了包含三种模态信息的修正表达矩阵,为后续聚类分析提供了增强的多模态输入数据[23]。SpaGCN构建无向加权图模型,将图像模态的RGB值映射为三维特征坐标,通过图卷积网络实现了多模态特征的邻域传播,结合无监督迭代聚类完成了空间域划分[24]。DeepST采用预训练的卷积神经网络(CNN)提取图像模态的特征,与基因表达、空间位置数据融合后构建空间增强的基因表达矩阵,进一步通过图神经网络自编码器与去噪自编码器联合学习数据的潜在表征[25]。EfNST利用EfficientNet[26]的复合尺度网络学习图像的多尺度特征来准确识别组织结构域,从而促进空间转录组学数据的多模态整合[27]。尽管上述方法在多模态数据整合技术上取得了一定的进步,但面对复杂的ST多模态数据的噪声和稀疏性,全面且高效地提取并整合多模态信息仍然是一个挑战。为此,本文提出了一个多视图多模态的对比学习框架STConVM,从多个角度更全面地提取了多模态特征并结合了VGAE和注意力机制,通过对比学习方法优化特征的提取。最后,本文在10x Visium平台的3个公开数据集,即:人类背外侧前额叶皮质(DLPFC)数据集、小鼠大脑前部数据集、人类乳腺癌数据集,进行了基准测试,与主流算法对比,STConVM在复杂组织的空间域识别、基因功能分析和数据去噪方面具有较高的准确性,为解析复杂组织微环境的分子机制提供了分析工具。

1　材料与方法

1.1　数据预处理

本研究主要选择10x Visium平台作为空间转录组的数据来源。该平台商业成熟度高、测序成本低、应用广泛,为后续的空间解析、细胞类型定位和机制研究提供了可靠的数据支撑。数据预处理能够为模型提供高质量的输入数据。STConVM模型的输入数据包括组织学图像、基因表达矩阵和空间位置信息。为减少高表达基因和低表达基因间的差异并降低噪声干扰,首先对原始基因表达矩阵进行对数变换,并标准化为统一文库大小,然后将标准化后的基因表达矩阵进一步缩放至零均值和单位方差。为更聚焦对细胞变化影响显著的基因,本研究选取了前2000个高变异基因(HVG)作为输入,提高分析效率。预处理后的基因表达矩阵记为$\boldsymbol{X}=\{x_1,x_2,\cdots,x_N\}\subseteq R^{N\times F}$,其中$N$表示斑点(spot)总数,$F$表示每个斑点的特征维度。对于组织学图像,首先以每个斑点为中心,截取一个正方形区域作为其局部图像特征代表并进行图像缩放,然后使用torchvision. transforms[28]对图像进行数据增强操作,包括归一化、旋转与清晰度调整等,然后从预训练的卷积神经网络(默认采用ResNet18[29])中提取每个图像块的特征表示。为更有效表征斑点的形态特征,采用主成分分析(PCA)提取前200个主成分作为潜在特征。

最后,使用余弦距离计算第i个斑点S_i和其相邻点S_j之间的形态相似性权重MS_{ij}:

$$MS_{ij}=1-\frac{S_i\cdot S_j}{\|S_i\|_2\|S_j\|_2}。$$

1.2　空间邻域网络构建

K近邻算法(KNN)[30]擅长捕捉数据的局部结构特征[31]。r半径方法[32-33]不受KNN中邻居数量的限制,能够适应不同密度和分布的数据,有助于揭示数据的全局结构[34]。为充分利用相邻点之间的相似性,本研究同时采用KNN和r半径方法,构建两个无向邻域图$G^{(n)}=\left(V,E^{(n)}\right),n=1,2,V$表

示点的集合,$E^{(n)}$表示第n个图中边的集合,其中n表示图的数量。定义$\boldsymbol{A}^{(n)}\in\mathbb{R}^{N\times N}$为图$G^{(n)}$的邻接矩阵,其中$N$为节点总数。

1.3 注意力机制

不同的邻接矩阵构建方法为复杂的组织样本提供了多视图表示,因此需要对这些邻接矩阵进行综合分析,以实现更全面的数据理解。由于各邻接矩阵可能包含互补甚至矛盾的信息,应为其生成的潜在嵌入分配不同的重要性,从而实现有效的数据整合。本研究采用注意力聚合层,来自适应地整合多个邻接矩阵,注意力聚合层通过为关键潜在嵌入分配更高权重,从而突出了更重要的邻接矩阵。算法首先通过计算注意力系数g_i^m以学习各邻接矩阵的重要性,其公式如下:

$$g_i^m=\boldsymbol{v}^{\mathrm{T}}\cdot\tanh\left(\boldsymbol{W}_i\,y_i^m+b_i\right)。$$

式中:y_i^m表示点i在视图m的潜在嵌入y^m中所对应的向量;g_i^m表示视图m对点i表示的重要性关注系数;$\boldsymbol{W}_i$和$\boldsymbol{v}$为可学习的权重矩阵;b_i为偏置项。然后使用softmax函数对注意力系数进行归一化,公式如下:

$$\beta_i^m=\frac{\exp(g_i^m)}{\sum_{m=1}^{M}\exp(g_i^m)}。$$

式中:β_i^m为归一化后的注意力得分,表示视图m对点i表示的贡献;M是视图总数。先令$\beta^m=\left[\beta_i{}^m\right]$,再使$\beta^m=\mathrm{diag}\left(\beta^m\right)$得到具体的注意力得分$\beta^m$。

最后,算法通过注意力得分β^m聚合各视图特定表示y^m,得到了最终的表示矩阵Z_g,公式如下:

$$Z_g=\sum_{m=1}^{M}\beta^m\cdot y^m。$$

1.4 编码器

基因表达的潜在表示由深度自动编码器学习,编码器由两层全连接层堆叠组成并从掩蔽后的基因表达矩阵$\boldsymbol{X}'\in\mathbb{R}^{n\times m}$生成低维表示$Z_f\in\mathbb{R}^{n\times d_f}$。同时,单层全连接解码器从潜在表示$Z\in\mathbb{R}^{n\times d}$中重构基因表达矩阵$\hat{\boldsymbol{X}}\in\mathbb{R}^{n\times m}$,该潜在表示$Z\in\mathbb{R}^{n\times d}$由低维表示$Z_f$与空间嵌入$Z_g\in\mathbb{R}^{n\times d_g}$拼接而成。这里$d_f$、$d_g$和$d$分别是编码器学习的低维表示、VGAE[35]学习并通过注意力机制融合的空间嵌入和ST-ConVM的最终潜在表示的维度,其中$d=d_f+d_g$。STConVM的解码器包含两种模式:聚类模式与基因去噪模式。在聚类模式下,解码器采用图卷积神经网络,可更有效地捕获空间信息;而在基因去噪模式中,使用线性解码器可以避免图卷积可能引起的特征过度平滑问题。深度自编码器的目标函数是最大化输入基因表达与重构结果之间的相似性,该相似性通过均方误差(MSE)损失函数$\sum\left(X-\hat{X}\right)^2$计算。

利用邻接矩阵$\boldsymbol{A}^{(n)}$及其对应的度矩阵$\boldsymbol{D}^{(n)}$,通过VGAE学习嵌入$Z_g{}^{(n)}$,其形式如下:$g:(A^{(n)})\rightarrow Z_g{}^{(n)}$。VGAE的推理过程由两层图卷积神经网络进行参数化:

$$g\left(Z_g{}^{(n)}|\boldsymbol{A}^{(n)}\right)=\prod g(z_i|\boldsymbol{A}^{(n)}),\ g\left(z_i|\boldsymbol{A}^{(n)}\right)=\mathcal{N}(z_i|\boldsymbol{\mu}_i{}^{(n)},\mathrm{diag}(\boldsymbol{\sigma}_i^{(n)2}))。$$

式中:$\boldsymbol{\mu}^{(n)}$为均值矩阵;$\boldsymbol{\sigma}^{(n)}$为标准差矩阵。$\boldsymbol{\mu}^{(n)}$和$\ln(\boldsymbol{\sigma}^{(n)})$的图卷积神经网络定义如下:

$$\mathrm{GCN}_{\boldsymbol{\mu}^{(n)}}\left(\boldsymbol{A}^{(n)}\right)=\tilde{\boldsymbol{A}}^{(n)}\mathrm{ReLU}\left(\tilde{\boldsymbol{A}}^{(n)}\boldsymbol{W}_0\right)\boldsymbol{W}_{\mu^{(n)}},$$

$$\mathrm{GCN}_{\ln(\boldsymbol{\sigma}^{(n)})}\left(\boldsymbol{A}^{(n)}\right)=\tilde{\boldsymbol{A}}^{(n)}\mathrm{ReLU}\left(\tilde{\boldsymbol{A}}^{(n)}\boldsymbol{W}_0\right)\boldsymbol{W}_{\sigma^{(n)}}。$$

式中:$\boldsymbol{W}_0$、$\boldsymbol{W}_{\mu^{(n)}}$、$\boldsymbol{W}_{\sigma^{(n)}}$为权重矩阵;$\tilde{\boldsymbol{A}}^{(n)}=\boldsymbol{D}^{(n)-\frac{1}{2}}\boldsymbol{A}^{(n)}\boldsymbol{D}^{(n)-\frac{1}{2}}$为对称归一化邻接矩阵。空间嵌入$Z_g{}^{(n)}$的分布难以直接建模,因此通过重参数化进行采样,公式如下:

$$Z_g{}^{(n)}=\boldsymbol{\mu}^{(n)}+\boldsymbol{\sigma}^{(n)}\odot\epsilon。$$

式中:$\epsilon\sim\mathrm{Normal}(0,1)$。在获得空间嵌入$Z_g{}^{(n)}$之后,通过注意力机制生成融合后的潜在表示$Z_g$,与$Z_f$进行拼接获得合并的潜在表示$Z$,并进一步重构邻接矩阵$\hat{\boldsymbol{A}}^{(n)}$:

$$\hat{A}^{(n)}=\sigma(Z\cdot Z^{\mathrm{T}})。$$

VGAE的优化目标是同时最小化原始邻接矩阵$A^{(n)}$与重构邻接矩阵$\hat{A}^{(n)}$之间的交叉熵损失，以及空间嵌入分布$g\left(Z_g^{(n)}|A^{(n)}\right)$与标准高斯先验$p(Z_g^{(n)})=\prod_i \mathcal{N}(z_i|0,I)$之间的KL散度。

1.5　DEC聚类

空间关系仅存在于每个独立的空间转录组中，不同转录组之间的斑点之间不存在直接的空间联系。令A^k和Z_f^k分别表示第k个空间转录组的邻接矩阵和深度基因表示。本研究构建了一个块对角邻接矩阵A^k，并将所有斑点的基因表示在特征维度上进行拼接，公式如下：

$$A=\begin{bmatrix}A^1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & A^K\end{bmatrix},Z_f=\begin{bmatrix}Z_f^1\\ \vdots\\ Z_f^K\end{bmatrix}。$$

式中，k表示空间转录组的数量。基于上述构建方式，多个空间转录组被统一转换为一个块对角邻接矩阵形式的图结构，作为STConVM的输入。

为消除批次效应并提升潜在表示的紧凑性，STConVM引入无监督深度嵌入聚类（DEC）方法[36]，以迭代方式对斑点进行聚类。本研究使用scikit-learn中的KMeans函数对潜在表示进行聚类初始化，聚类簇的数量作为超参数预先设定。在上述初始化基础上，DEC通过两步无监督迭代优化聚类结果。在第一步中，采用学生t分布计算第i个潜在点z_i（$i=1,2,\cdots,N$）到第j个聚类中心μ_j（$j=1,2,\cdots,K$）的软分配概率q_{ij}，公式如下：

$$q_{ij}=\frac{\left(1+\left|\left|z_i-\mu_j\right|\right|^2\right)^{-1}}{\sum_{j'}^{K}\left(1+\left|\left|z_i-\mu_{j'}\right|\right|^2\right)^{-1}}。$$

在第二步中，构建基于q_{ij}的辅助目标分布p_{ij}，通过其高置信度样本进一步细化聚类结果：

$$p_{ij}=\frac{q_{ij}^2/\sum_i^N q_{ij}}{\sum_{j'}^{K}(q_{ij'}^2/\sum_i^N q_{ij'})}。$$

基于软分配q_{ij}与目标分布p_{ij}，通过KL散度定义聚类目标函数，公式如下：

$$\mathrm{KL}(P\|Q)=\sum_i^N\sum_j^K p_{ij}\ln\frac{p_{ij}}{q_{ij}}。$$

最后，采用带动量的随机梯度下降法（SGD）同时优化STConVM的模型参数与聚类中心。

1.6　自监督对比学习

为了从基因表达和组织学图像两个模态中同时提取联合特征嵌入，本研究在共享特征空间中采用SimCLR框架下的NT-Xent损失函数进行对比学习，旨在将来自同一个斑点的基因表达特征与图像特征（正对）拉近，同时将不匹配的特征对（负对）分离开来。NT-Xent损失（归一化温度缩放交叉熵损失）用于引导两个模态的特征共同学习统一的嵌入表示。每个损失值通过M个斑点的基因表达特征及其对应裁剪图像块的图像特征计算得到。在每个训练批次中，定义一组正对后，其余$2(M-1)$个特征对被视为负样本，相同斑点的基因特征与图像特征之间的对比损失函数定义如下：

$$L_{z_i,z_j'}=-\ln\frac{\exp\left(\frac{\mathrm{sim}\left(Z_i,Z_j'\right)}{\tau}\right)}{\sum_{k=1}^{2N}1_{[k\neq i]}\exp\left(\frac{\mathrm{sim}\left(Z_i,Z_k'\right)}{\tau}\right)}。$$

式中:Z_i是点i的基因表达特征与空间位置特征进行拼接获得的潜在表示;Z_j',Z_k'分别是点i、k通过卷积神经网络提取的图像特征;$1_{[k\neq i]}\in\{0,1\}$为指示函数,当$k\neq i$时取值为1;τ为温度参数,用于控制负样本分离程度。$\mathrm{sim}\left(Z_i,Z_j'\right)$表示$l_2$归一化后的基因表达矩阵与图像特征之间的点积,定义如下:

$$\mathrm{sim}\left(Z_i,Z_j'\right)=\frac{(Z_i)^{\mathrm{T}}Z_j'}{\|Z_i\|\|Z_j'\|}。$$

最终对比损失通过所有具有正对关系的特征对(如Z_i、Z_j'和Z_j、Z_i')共同计算得到。

2 结果与讨论

2.1 STConVM框架

在ST数据中,为充分挖掘基因表达谱、空间坐标信息与组织学图像三种模态数据在空间域识别中的协同潜力,本研究提出了一种多模态多视图对比学习框架STConVM,其整体架构如图1所示。该框架包含5个核心模块:空间位置特征提取、基因表达特征提取、组织学图像特征提取、对比学习模块以及解码器。对于空间位置信息模态,采用KNN和r半径方法分别构建邻域图,作为VGAE的输入以生成潜在表示,并通过注意力机制融合两个嵌入向量(图1-A)。对于基因表达信息模态,通过过滤程序去除原始表达矩阵中的低变异基因,并进行数据掩码,随后输入至深度自编码器以生成潜在表示(图1-B)。对于图像信息模态,经过预处理后,组织学图像被输入至卷积神经网络(CNN)中,以提取潜在特征表示(图1-C)。在对比学习中,将空间位置和基因表达的潜在表示拼接后,与图像特征进行对比学习。通过拉近同一斑点的正对特征、远离不同斑点负对特征,生成了优化的多模态嵌入表示(图1-D)。然后对优化后的嵌入向量进行解码,恢复数据结构(图1-E)。最终,采用Mclust方法对优化后的潜在嵌入执行DEC聚类,实现了空间域识别与基因表达去噪(图1-F)。本研究在多个ST数据集上进行了测试与比较,包括人类背外侧前额叶皮质(DLPFC)数据集、小鼠大脑前部数据集及人类乳腺癌数据集。

2.2 STConVM有效解析了DLPFC数据集的层级结构

DLPFC数据集包含12个组织切片,每个切片有4或6个皮质层以及1个白质层[37]。由于该数据集具有清晰的层次结构,常被用于评估空间域识别算法的性能。图2-A展示了151674切片的组织学图像及其人工标注结果。

为评估STConVM在空间域识别中的表现,本研究将其与7种主流方法进行对比,包括EfNST、SpaNCMG、SpaGCN、SEDR、STAGATE、DeepST和GraphST。图2-B为8种算法在12个切片上的调整兰德指数(ARI)值的箱线图。结果表明STConVM的结果显著优于对比方法,在151672切片上取得了最高ARI值(0.72),较次优方法GraphST(ARI=0.63)提升了14%。

在151674切片上,STConVM与其他方法的可视化对比如图2-C—D所示,SpaGCN识别效果较差,聚类结果混乱,边界模糊,难以区分层级结构;而DeepST、SpaNCMG、GraphST、EfNST、SEDR与STAGATE虽实现了对7个层的基本分离,但未能精确还原人工标注,边界仍不清晰,且存在噪声干扰。STConVM展现出更优的分层效果,ARI值达0.67,较次优方法STAGATE提升6%。其结果与人工标注高度一致,仅在第4层存在差异。由于第4层在图像中占比较小,对整体ARI值影响有限。上述结果表明,STConVM在聚类性能上相较于其他算法展现出了显著的优势。

原始ST数据常受到高噪声和随机缺失的干扰,这严重影响基因表达分析的准确性与可靠性。因此,本研究进一步验证了STConVM的去噪能力。以151676切片为例,本研究对该切片中6个层特异性标记基因(*SLC1A2*、*CAMK2N1*、*NEFH*、*PCP4*、*SCGB2A2*、*MBP*)在去噪前后的表达情况进行了对比(图3-A)。结果显示,在原始表达状态下,这些基因分布杂乱,难以提取有效的生物学信息,而在使用STConVM增强处理后,这些层标记基因呈现出更为清晰的空间富集特征。去噪后的嵌入数据能够准确描绘皮质层之间的边界,且切片清晰展现了各标记基因的空间富集特征。例如,

去噪后,*SLC1A2*基因在第1层表现出明显的差异表达,结果与文献[37]的研究结果一致,而其原始表达则较为混乱。STConVM揭示的层状富集特征得到了艾伦脑图谱中公开的原位杂交(ISH)数据[38]的验证。本研究对原始与去噪后的表达进行了量化比较(图3-B),结果表明这些层标记基因的空间表达模式显著增强。STConVM不仅有效减少了噪声干扰,还突出了表达的关键特征,从而更准确地解读了基因的空间分布模式。此外,通过对比原始与去噪后的表达,验证了STConVM在增强层特异性基因空间表达方面的有效性。

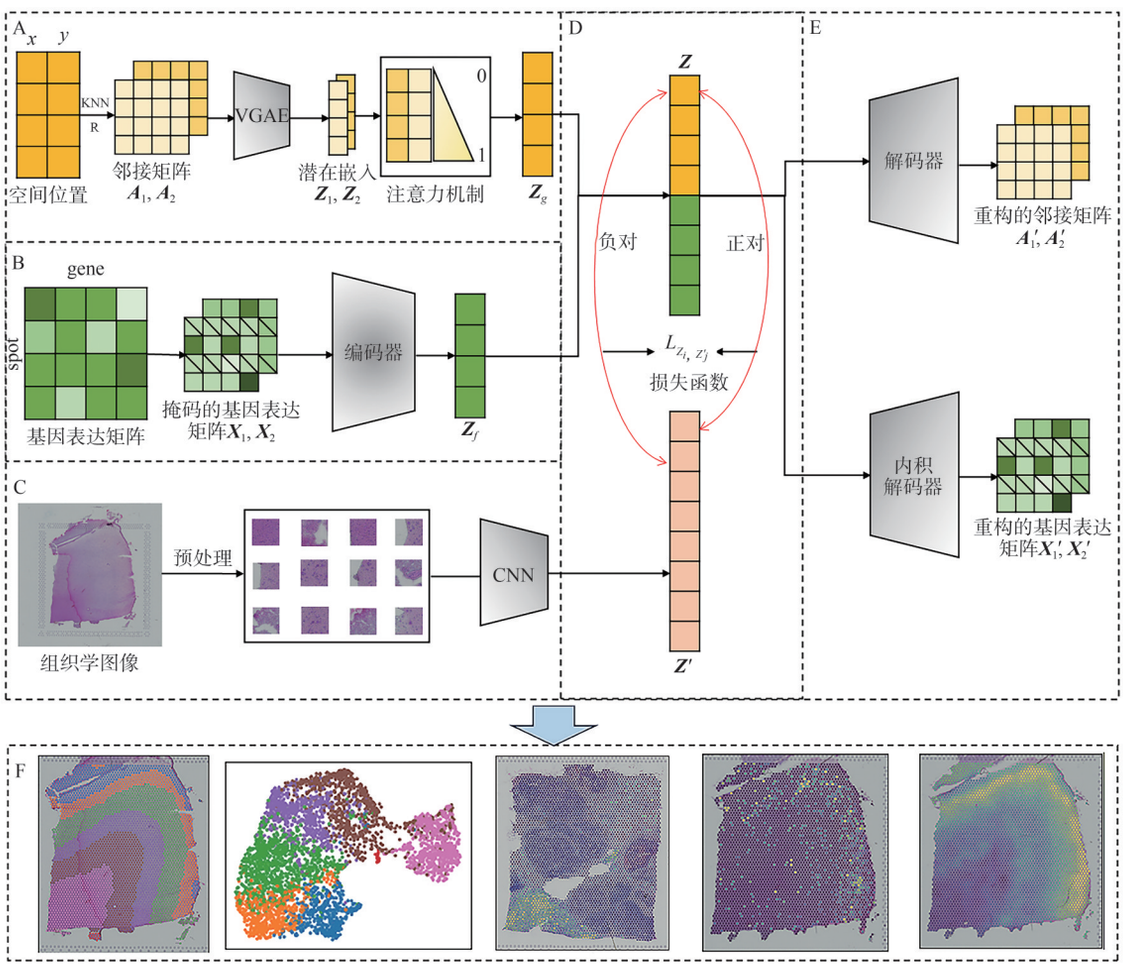

图1　STConVM的流程图

Fig. 1　Workflow of the STConVM

2.3　STConVM更精确地描述了小鼠脑前部组织中微小区域

小鼠脑前部数据集提供了特定脑区、亚区域及细胞类型,有助于验证算法在精确识别复杂组织结构方面的能力。如图4-A所示,共手动注释标记了52个区域。图4-B展示了STConVM与其他5种方法在ARI指标上的直方图比较。结果显示,STConVM取得了最高的ARI得分。图4-C展示了6种方法的聚类结果可视化。可以明显看出,SpaNCMG与STAGATE聚类性能较差,不同区域混杂,边界模糊不清;DeepST与GraphST虽识别出相对清晰的边界,但与手动注释相比仍存在较多不一致;SEDR的聚类结果更接近手动注释,但部分区域仍存在混合,边界相对不清晰;相比之下,STConVM能够清晰识别小鼠大脑前部数据集中全部52个簇,边界明确。此外,本研究进一步鉴定了小鼠大脑主要区域的标记基因。结果显示,在第2区(纹状体)中检测到表达最强的标记基因为*Gpr88*(图4-D),与文献[39]的结果一致。研究表明,*Gpr88*参与调控多种神经传递过程,

包括多巴胺传递、精神活动及运动控制[40],因此,该基因可能成为治疗帕金森病、精神分裂症等精神疾病的潜在靶点。同样,在巨噬细胞(MOB)第23区的肾小球旁颗粒细胞(GR)中,本研究检测到潜在标记基因 *Cdhr1*,并通过艾伦脑图谱数据库(https://portal.brain-map.org/,实验ID:71662897)验证了该结果。相关研究表明,*Cdhr1* 在视网膜发育中发挥关键作用,为视网膜疾病的研究与治疗提供新思路[41]。

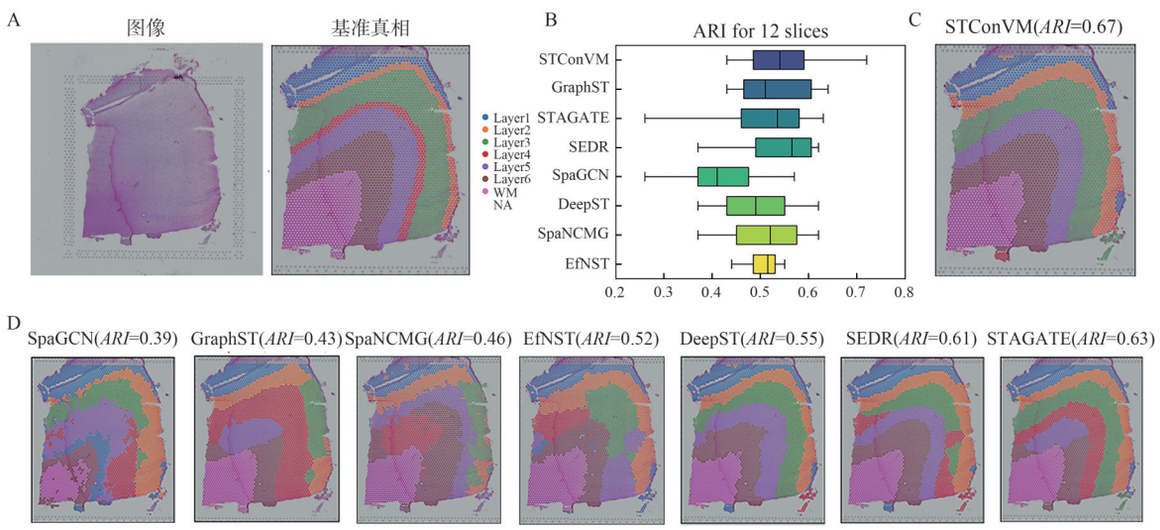

图2 STConVM在DLPFC数据集的聚类结果与比较

Fig. 2 Clustering results and comparison of STConVM on the DLPFC dataset

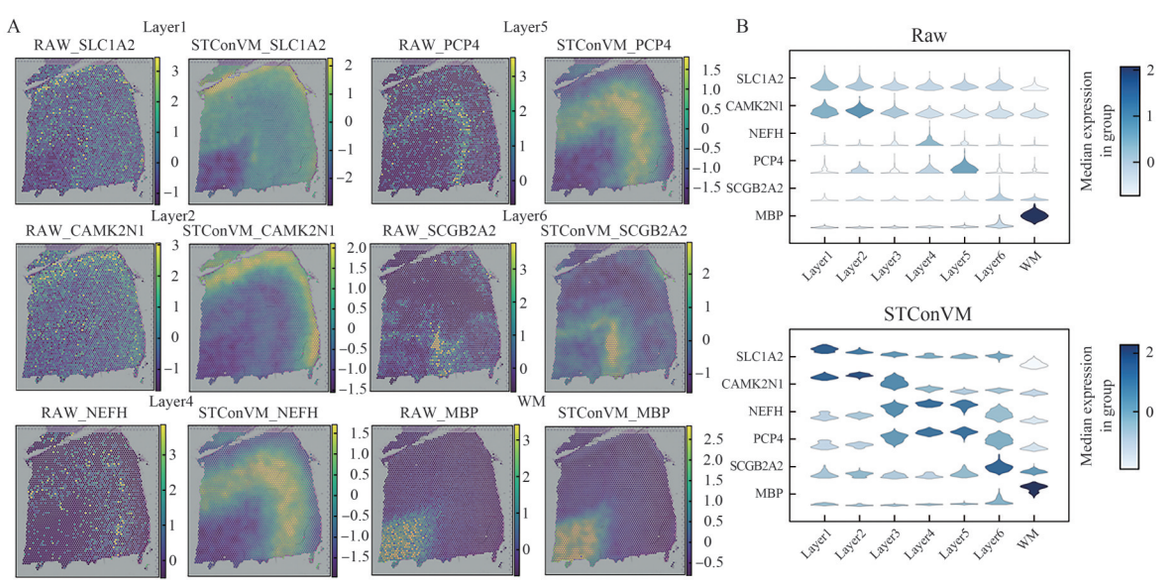

图3 STConVM增强了DLPFC数据集中层标记基因的空间模式

Fig. 3 STConVM enhances the spatial patterns of layer-marker genes on the DLPFC dataset

2.4 STConVM揭示了人类乳腺癌症数据集的精细组织

乳腺癌组织具有复杂的结构和异质性,通过对组织切片进行聚类分析,可以更好地理解不同细胞类型之间的空间关系,揭示肿瘤内部结构和组织异质性。本文在10x Visium平台上的人类乳腺癌数据中进行了测试。该数据集共注释为20个空间域,归类为四种形态类型:DCIS/LCIS、健康组织、IDC和肿瘤边缘(图5-A)。在空间域识别任务中,STConVM依然取得了最高的ARI得分(图5-B)。聚类结果如图5-C所示,本文对其进行了可视化展示。SpaGCN、SpaNCMG和STAGATE得到的聚

类结果存在明显的碎片化现象,边界不规则且存在严重的噪声点,影响了聚类结果的准确性。尽管GraphST、DeepST、EfNST 和 SEDR 的聚类结果具有更清晰的边界,但与实际的人类乳腺癌组织结构仍相去甚远。这些方法的聚类结果大多混淆了非肿瘤区域和部分肿瘤区域,这不利于癌症患者肿瘤组织的研究。与其他 7 种方法相比,STConVM 的聚类结果中区域之间边界清晰且分割平滑,具有更强的区域连续性和更少的噪声干扰,与人类乳腺癌组织结构分区更加一致。这些结果表明了 STConVM 在识别人类肿瘤组织空间域方面的优越性。

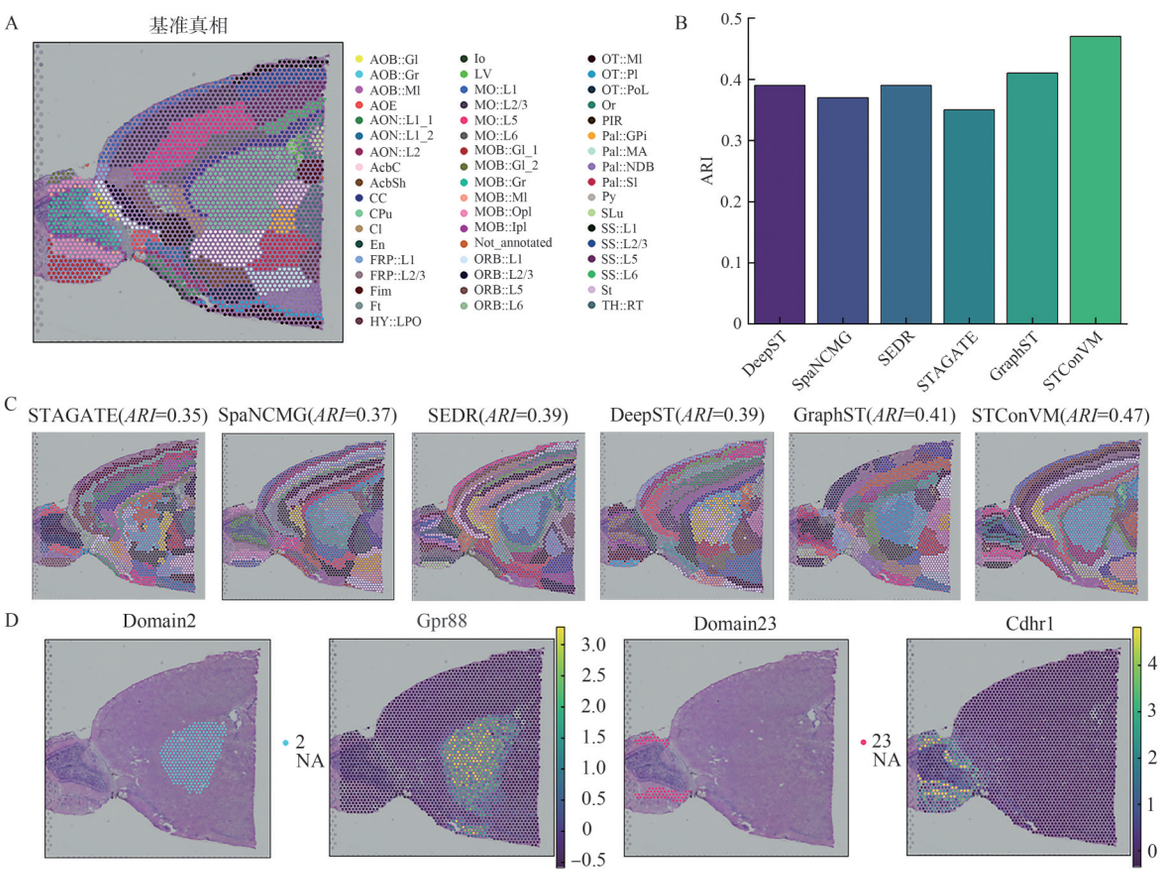

图 4　STConVM 优化了小鼠脑前部已知组织结构的识别

Fig. 4　STConVM optimizes the identification of known tissue structures in the mouse brain anterior

为了验证 STConVM 在空间聚类中的准确性,本研究绘制了邻域富集热图用于评估不同空间区域之间的邻近关系。图 6 展示了基准真值(图 6-A)与 STConVM 聚类结果(图 6-B)的邻域富集热图。热图中横轴和纵轴为 20 个对应的区域标签。色标条上紫色到黄色的数值越来越大,表示两个区域之间的邻域富集程度越来越高。其中:黄色表示区域间在组织切片中呈正向富集,表明聚类内部区域在空间上高度相似,具有一致性;紫色表示两个区域在空间中呈负向富集,表明这些区域在空间上倾向于分散,聚类区分度高。热图中每个单元格的颜色对应两个空间域之间的富集得分,分数越高,表示富集程度越高。例如,图 6-A 中的两个黑圈分别显示 Tumor_Edge_3 靠近 DCIS/LCIS_1,Tumor_Edge_2 靠近 IDC_3;图 6-B 中的两个红圈显示了域 5 与域 8、域 7 与域 18 之间的空间接近关系。这些观察结果与图 5-C 中的注释相符,进一步验证了 STConVM 在识别人类肿瘤组织空间域方面的优越性。

2.5　消融实验

为验证 STConVM 中各模块的必要性与有效性,本文在 DLPFC 数据集的 151674 切片上进行了消融实验(图 7-A)。当仅使用单一的邻接矩阵构建方法(non-K,non-R)时,ARI 值分别为 0.6352 和

0.5410,这表明两种邻接矩阵构建方法提供的多视图信息提供了更全面的数据,在输入端促进了特征学习。当使用两种邻接矩阵构建方法但不使用注意力机制融合多视图信息(non-att)时,ARI值为0.6292,这表明注意力机制模块对两种视图实现有效的数据整合,增强了特征学习。在STConVM的基础上去掉对比学习模块(non-CLimg),ARI值只有0.6364,这表明对比学习模块有效地优化了基因表达的潜在嵌入,更好地融合了基因表达和图像的特征,STConVM算法克服了图像模态数据的噪声,增强了特征学习。

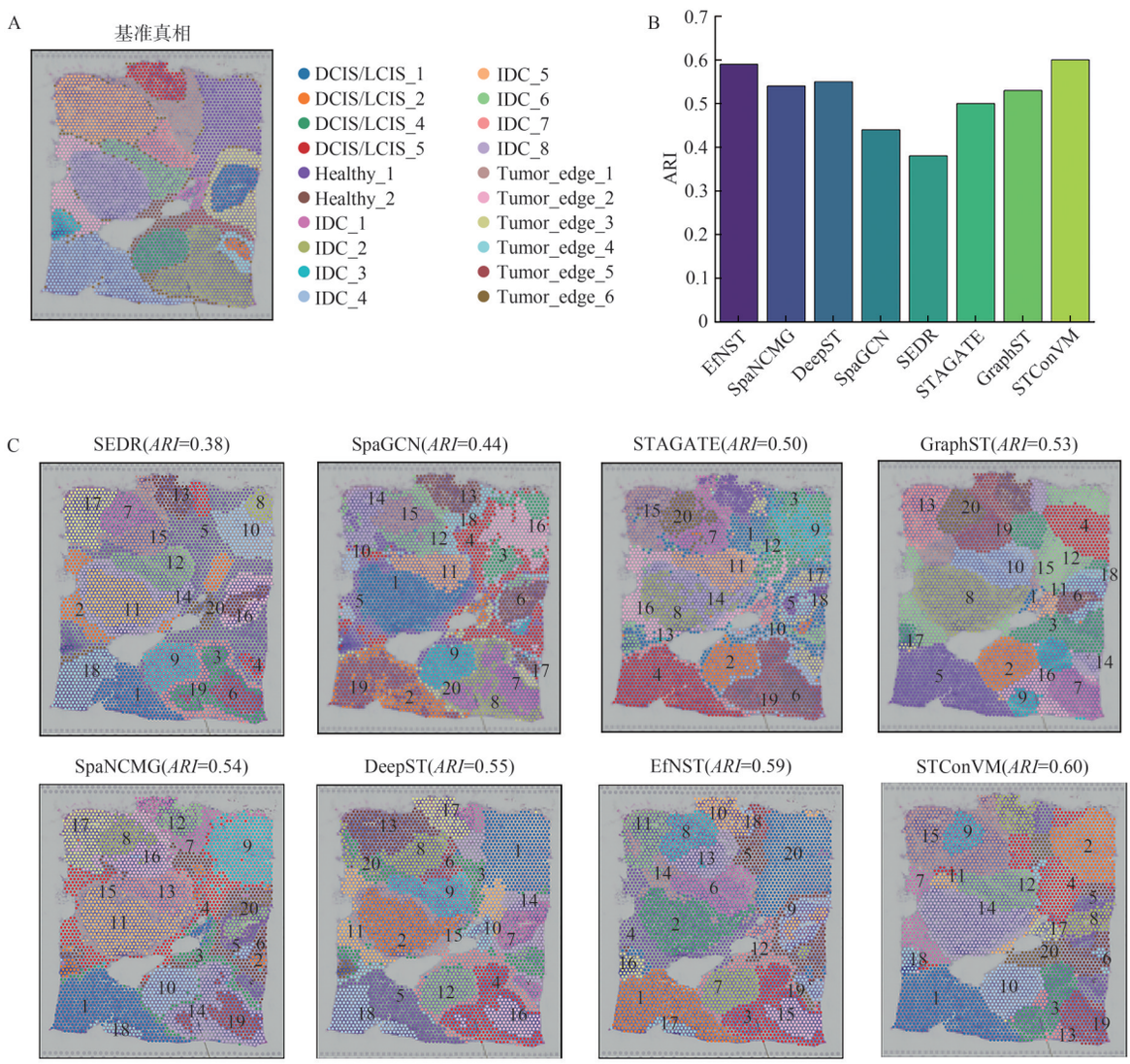

图5 STConVM解析人类乳腺癌组织的精细结构

Fig. 5 STConVM resolves the fine structure of human breast cancer tissue

本文在10x Visium平台上的人类乳腺癌数据集上亦进行了消融实验(图7-B),结果同样清晰展现了上述趋势。综合上述结果,进一步验证了3个模块在模型中各具独特且关键的作用,每一模块在特征学习与空间关系识别中均不可或缺。

此外,为了验证不同数量的主成分对实验结果的影响,本文在DLPFC数据集的151674切片上进行了对比实验(图8),结果表明当使用200个主成分时,算法性能达到最佳,证明了算法选择的主成分数量能高效压缩特征、保留主要生物变异、去除噪声,并为下游聚类、图构建和模型训练提供了稳健、低维、高信息量的表示。

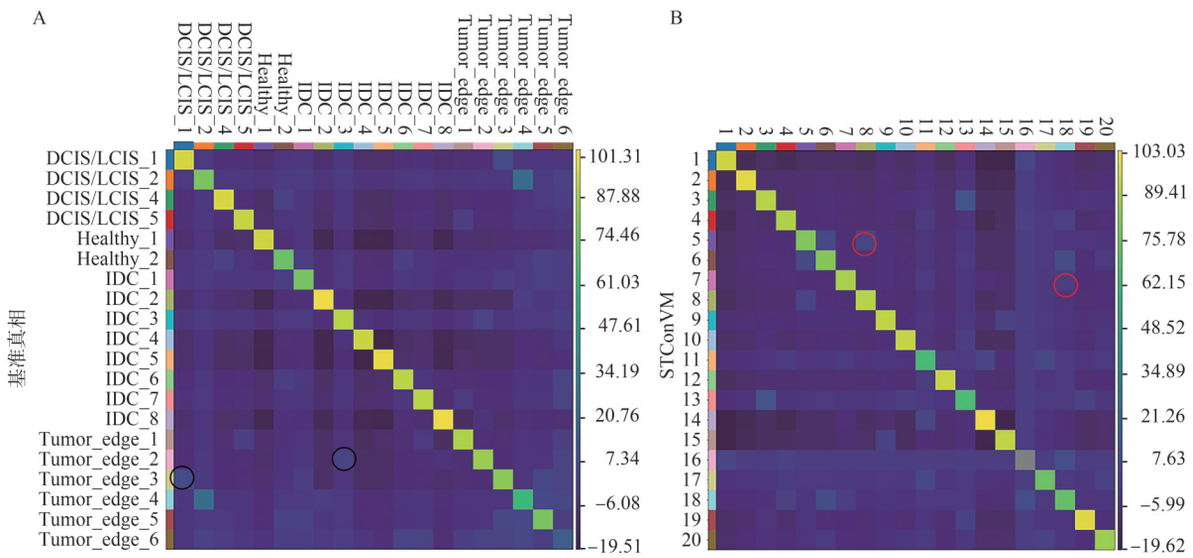

图 6　人类乳腺癌的邻域富集热图

Fig. 6　Neighborhood enrichment heatmap of the human breast cancer

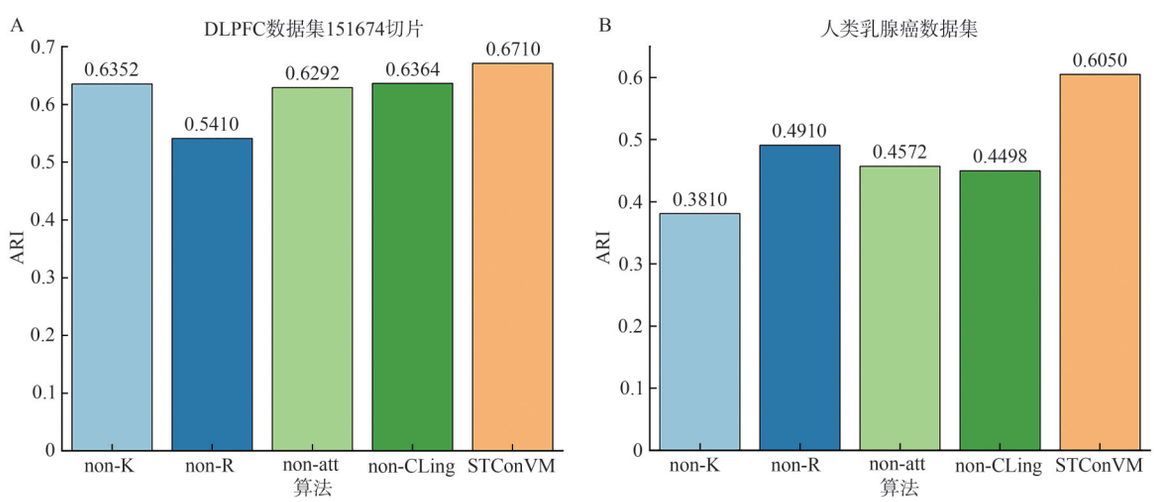

图 7　DLPFC 和人类乳腺癌数据集上的消融实验结果

Fig. 7　Results of ablation experiments on DLPFC dataset and human breast cancer dataset

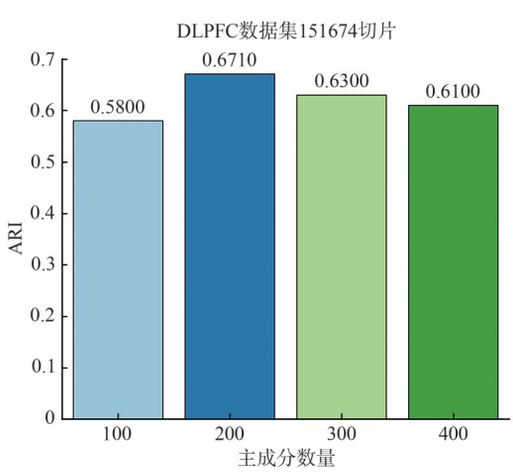

图 8　不同数量的主成分对结果的影响

Fig. 8　Effect of different number of principal components on the results

3 结论

准确解析ST数据中的空间域对揭示组织结构-功能关系至关重要。针对当前主流方法在复杂组织识别中存在的噪声敏感与模态融合不足问题,本研究提出了多视图多模态对比学习框架ST-ConVM。该算法在多视图空间建模方面,首先构建了KNN和r半径两个邻域图。经VGAE提取局部特征后,通过注意力机制进行融合。然后,将深度自编码器提取的基因表达特征与注意力机制生成的空间位置的潜在嵌入进行拼接。通过对比学习策略,该拼接后的特征与组织学图像特征进行跨模态对齐,强化多源信息的互补表征。最后,利用解码器重构基因表达分布,结合Mclust概率模型实现对噪声数据的鲁棒性聚类。

在多个公开数据集的验证与比较中,STConVM展现出显著的优越性。在DLPFC数据集12个组织切片中,STConVM的ARI值最高达0.72,较次优方法GraphST提升14%。在小鼠大脑前部数据集上,STConVM对已知组织结构的识别准确率较次优方法提升15%,证实其在识别精细结构方面的表现卓越。此外,在人类乳腺癌数据集中,STConVM在各项评估指标上的结果均优于其他方法,表明其可为研究肿瘤进化及其与微环境的相互作用提供有力支持。为了验证这些模块的有效性,本研究进行了消融实验,证实了STConVM模块间的协同有效性。本研究为ST数据分析提供了多模态融合技术工具,推动了空间组学分析方法的发展。本文代码已上传至https://github.com/sijiabao/STConVM.git。未来本研究将进一步针对更多ST技术(如Slide-seq、Stereo-seq)的分辨率差异与数据特性,开发模态自适应特征融合方法,并探索多组学数据的整合分析方法。

参考文献:

[1] RAO A, BARKLEY D, FRANÇA G S, et al. Exploring tissue architecture using spatial transcriptomics[J]. Nature, 2021, 596(7871): 211-220.

[2] 冯振兴,尚文婧,司佳宝,等. 基于空间转录组学数据的空间域识别算法综述[J]. 内蒙古大学学报(自然科学版), 2024, 55(6): 452-462.

[3] ZHANG M, EICHHORN S W, ZINGG B, et al. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH[J]. Nature, 2021, 598(7879): 137-143.

[4] SHAH S, TAKEI Y, ZHOU W, et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH[J]. Cell, 2018, 174(2): 363-376. e16.

[5] CODELUPPI S, BORM L E, ZEISEL A, et al. Spatial organization of the somatosensory cortex revealed by osmFISH[J]. Nature Methods, 2018, 15(11): 932-935.

[6] WANG X L, HE Y, ZHANG Q M, et al. Direct comparative analyses of 10x Genomics Chromium and Smart-seq2[J]. Genomics, Proteomics & Bioinformatics, 2021, 19(2): 253-266.

[7] STICKELS R R, MURRAY E, KUMAR P, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqV2[J]. Nature Biotechnology, 2021, 39(3): 313-319.

[8] CHEN A, LIAO S, CHENG M N, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays[J]. Cell, 2022, 185(10): 1777-1792. e21.

[9] PALLA G, FISCHER D S, REGEV A, et al. Spatial components of molecular tissue biology[J]. Nature Biotechnology, 2022, 40(3): 308-318.

[10] TEVES J M, WON K J. Mapping cellular coordinates through advances in spatial transcriptomics technology[J]. Molecules and Cells, 2020, 43(7): 591-599.

[11] WANG N, LI X, WANG R S, et al. Spatial transcriptomics and proteomics technologies for deconvoluting the tumor microenvironment[J]. Biotechnology Journal, 2021, 16(9): e2100041.

[12] JIANG R，LI Z，JIA Y H，et al. SINFONIA：Scalable identification of spatially variable genes for deciphering spatial domains[J]. Cells，2023，12(4)：604.

[13] TANG Z Y，ZHANG T L，YANG B J，et al. spaCI：Deciphering spatial cellular communications through adaptive graph model[J]. Briefings in Bioinformatics，2023，24(1)：bbac563.

[14] BU Z，LI H J，ZHANG C C，et al. Graph K-means based on leader identification，dynamic game，and opinion dynamics[J]. IEEE Transactions on Knowledge and Data Engineering，2020，32(7)：1348-1361.

[15] BLONDEL D V，GUILLAUME J，LAMBIOTTE R，et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics-Theory and Experiment，2008，2008(10)：10008.

[16] SATIJA R，FARRELL J A，GENNERT D，et al. Spatial reconstruction of single-cell gene expression data[J]. Nature Biotechnology，2015，33(5)：495-502.

[17] DRIES R，ZHU Q，DONG R，et al. Giotto：A toolbox for integrative analysis and visualization of spatial expression data[J]. Genome Biology，2021，22(1)：78.

[18] XU H，FU H Z，LONG Y H，et al. Unsupervised spatially embedded deep representation of spatial transcriptomics [J]. Genome Medicine，2024，16(1)：12.

[19] DONG K N，ZHANG S H. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder[J]. Nature Communications，2022，13(1)：1739.

[20] SI Z H，LI H S，SHANG W J，et al. SpaNCMG：Improving spatial domains identification of spatial transcriptomics using neighborhood-complementary mixed-view graph convolutional network[J]. Briefings in Bioinformatics，2024，25(4)：bbae259.

[21] REN H L，WALKER B L，CANG Z X，et al. Identifying multicellular spatiotemporal organization of cells with SpaceFlow[J]. Nature Communications，2022，13(1)：4076.

[22] LONG Y H，ANG K S，LI M W，et al. Spatially informed clustering，integration，and deconvolution of spatial transcriptomics with GraphST[J]. Nature Communications，2023，14(1)：1155.

[23] PHAM D，TAN X，BALDERSON B，et al. Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues[J]. Nature Communications，2023，14(1)：7739.

[24] HU J，LI X J，COLEMAN K，et al. SpaGCN：Integrating gene expression，spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network[J]. Nature Methods，2021，18(11)：1342-1351.

[25] XU C，JIN X Y，WEI S R，et al. DeepST：Identifying spatial domains in spatial transcriptomics by deep learning[J]. Nucleic Acids Research，2022，50(22)：e131.

[26] TAN M，LE Q. EfficientNet：Rethinking model scaling for convolutional neural networks[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach，California，USA：PMLR，2019：6105-6114.

[27] ZHAO Y N，LONG C S，SHANG W J，et al. A composite scaling network of EfficientNet for improving spatial domain identification performance[J]. Communications Biology，2024，7(1)：1567.

[28] PASZKE A，CROSS S，MASSA F，et al. PyTorch：An imperative style，high-performance deep learning library [C]//33rd Conference on Neural Information Processing Systems(NeurIPS 2019). Vancouver，Canada：NeurLPS，2019：01703.

[29] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas，NV，USA：IEEE，2015：03385.

[30] COVER T，HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory，1967，13(1)：21-27.

[31] HU Q，YU D，XIE Z. Neighborhood classifiers[J]. Expert Systems with Applications，2008，34(2)：866-876.

[32] TURAU V. Fixed-radius near neighbors search[J]. Information Processing Letters，1991，39(4)：201-203.

[33] BENTLEY J L，STANAT D F，WILLIAMS E H. The complexity of finding fixed-radius near neighbors[J]. Infor-

mation Processing Letters,1977,6(6):209-212.
[34] WANG Z,LI Y Q,LI D D,et al. Entropy and gravitation based dynamic radius nearest neighbor classification for imbalanced problem[J]. Knowledge-based Systems,2020,193:105474.
[35] KIPF N T,WELLING M. Variational graph auto-encoders[DB/OL]. arXiv,2016(2016-11-21)[2024-10-21]. https://arxiv.org/abs/1611.07308.
[36] XIE J,ROSS G,ALI F. Unsupervised deep embedding for clustering analysis[C]//33rd International Conference on Machine Learning. New York,NY,USA:PMLR,2016:478-487.
[37] MAYNARD K R,COLLADO-TORRES L,WEBER L M,et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex[J]. Nature Neuroscience,2021,24(3):425-436.
[38] SUNKIN S M,NG L,LAU C,et al. Allen brain Atlas:An integrated spatio-temporal portal for exploring the central nervous system[J]. Nucleic Acids Research,2013,41(Database issue):D996-D1008.
[39] VAN WAES V,TSENG K Y,STEINER H. GPR88:A putative signaling molecule predominantly expressed in the striatum:Cellular localization and developmental regulation[J]. Basal Ganglia,2011,1(2):83-89.
[40] LABOUTE T,GANDÍA J,PELLISSIER L P,et al. The orphan receptor GPR88 blunts the signaling of opioid receptors and multiple striatal GPCRs[J]. eLife,2020,9:e50519.
[41] HAER W L,VAN D R,VAN G M,et al. Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases[J]. Proceedings of the National Academy of Sciences of the United States of America,2020,117(5):2710-2716.

(责任编委 那顺布和)

STConVM: A Multi-View Multi-Modal Contrast Learning Approach to Identify Spatial Domains

SI Jiabao[1],ZHAO Xiangyu[1],LIU Hexin[1],DAI Bingjie[2],FENG Zhenxing[1]
(1. *College of Science*, *Inner Mongolia University of Technology*, *Hohhot* 010051, *China*;
2. *School of Life Sciences*, *Inner Mongolia University*, *Hohhot* 010021, *China*)

Abstract: In spatial domain recognition within spatial transcriptomics, existing algorithms often struggle with limited multimodal feature fusion capability, suboptimal recognition accuracy, and high computational complexity when integrating spatial transcriptomic modalities. To address these challenges, this study proposes a STConVM, a novel method based on a multi-view multimodal contrastive learning framework. STConVM enables the integration of gene expression profiles, spatial coordinates, and histological images, thereby enhancing the accuracy and robustness of spatial domain recognition. Extensive experiments on multiple public datasets demonstrate that STConVM more precisely identifies functional tissue regions and serves as an effective computational tool for decoding the complex microenvironment of biological tissue.

Key words: spatial transcriptomics; identify spatial domains; multi view; multi modal; contrast learning