

文章编号 1004-924X(2025)09-1456-15

分层特征细化与频域增强学习结合的空间组学细胞核分割

李修齐^{1,2}, 李金泽^{1,2}, 杨 弃², 李莹雪², 赵才荣³, 周连群^{1,2*}, 姚 佳^{1,2*}

(1. 中国科学技术大学 生物医学工程学院(苏州) 生命科学与医学部, 安徽 合肥 230026;

2. 中国科学院 苏州生物医学工程技术研究所
生物医学检验技术重点实验室, 江苏 苏州 215163;
3. 同济大学 电子与信息工程学院, 上海 201804)

摘要:精确地将生物分子信息映射到组织切片的空间坐标是空间多组学分析的关键,细胞核分割精度直接决定靶标分子定位准确性。然而,细胞核形态多样性、组织结构复杂性以及细胞核聚集等因素造成细胞核分割困难,现有分割方法难以实现细胞核精准分割,进而影响空间多组学分析结果。针对上述问题,结合分层特征细化和频域增强学习提出细胞核分割网络 FFVM-UKAN,深度整合用于特征提取的浅层视觉状态空间模块及用于特征细化的深层令牌化 Kolmogorov-Arnold 网络,并提出并行频域学习模块实现细胞核分割所需精细特征捕捉,增强网络跳跃连接效果。该方法在公开数据集 MoNuSeg 上实现了细胞核分割, mIoU 和 Dice 分别为 69.09% 和 81.72%, 在私有数据集上达到 85.95% 和 92.45%。此外,在 10X Genomics 的人类肝脏数据集上验证基因、细胞核映射效果,结果显示基因映射准确率达 90.63%。上述结果表明,本文所述方法在细胞核分割精度和模型泛化能力方面具有较好效果,实现了空间基因与细胞核的高精度映射,展现了该方法在空间多组学分析中的应用潜力。

关键词:细胞核分割;分层特征细化;频域增强学习;Kolmogorov-Arnold 网络

中图分类号: TP391.4 文献标识码: A

doi: 10.37188/OPE.20253309.1456 CSTR: 32169.14.OPE.20253309.1456

Hierarchical feature refinement with frequency-enhanced learning method for nuclei segmentation in spatial omics

LI Xiuqi^{1,2}, LI Jinze^{1,2}, YANG Qi², LI Yingxue², ZHAO Cairong³, ZHOU Lianqun^{1,2*}, YAO Jia^{1,2*}

(1. Division of Life Sciences and Medicine, School of Biomedical Engineering (Suzhou),
University of Science and Technology of China, Hefei 230026, China;

2. Key Laboratory of Biomedical Detection Technology, Suzhou Institute of Biomedical Engineering and
Technology, Chinese Academy of Sciences, Suzhou 215163, China;

3. College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

* Corresponding author, E-mail: zhoulq@sibet.ac.cn; yaojia@sibet.ac.cn

Abstract: Accurate mapping of biomolecular information onto tissue section image coordinates is a funda-

收稿日期: 2025-01-09; 修订日期: 2025-03-18.

基金项目: 国家重大科研仪器研制项目 (No. 82327802); 江苏省高端医疗器械技术创新中心项目 (No. BM2022012)

mental requirement in spatial multi-omics analysis, where the precision of nuclei segmentation critically determines the accuracy of target molecule localization. However, existing segmentation methods frequently produce suboptimal outcomes due to challenges such as highly heterogeneous nuclear morphology, complex tissue architecture, and densely packed cellular regions, all of which compromise the reliability of downstream spatial genomics analyses. To overcome these challenges, FFVM-UKAN, a novel encoder-decoder architecture, is proposed. This architecture integrates shallow Visual State Space modules for feature extraction with a deep tokenized Kolmogorov-Arnold Network for feature refinement. Additionally, a Parallel Frequency Domain Learnable Module is incorporated to significantly enhance skip connections by effectively capturing fine-grained frequency-level features essential for high-precision nuclei segmentation. The proposed method achieved a mean Intersection over Union (mIoU) of 69.09% and a Dice coefficient of 81.72% on the MoNuSeg dataset, and 85.95% and 92.45% respectively on an in-house dataset. Further validation using the 10X Genomics human liver dataset for gene-to-nuclei mapping yielded an accuracy of 90.63%. Experimental results demonstrate that the proposed approach delivers superior nuclei segmentation accuracy and robust model generalization. This enables highly precise gene-to-nuclei mapping, underscoring the significant potential of this method to advance spatial multi-omics research and its applications.

Key words: nuclei segmentation; hierarchical feature refinement; frequency-enhanced learning; Kolmogorov-Arnold network

1 引言

空间多组学通过整合基因组学、转录组学、蛋白质组学和代谢组学等数据,为生物样本提供全面的分析,这对于揭示复杂的生物系统至关重要^[1-2]。准确地将靶标分子映射至样本中单个细胞核是空间多组学尤其是空间基因组分析的重要基础^[3]。然而,在H&E染色图像中进行细胞核精准分割面临着诸多挑战,例如细胞核形态的多样性^[4]、组织结构的复杂性^[5]、染色强度和均匀性的变化^[6]以及细胞核之间的重叠或黏连^[7]。这些因素会导致细胞核分割精度下降,进而影响基因定位准确性以及后续生信分析的可靠性。目前,空间多组学中的细胞核分割多依赖传统算法^[8]或机器学习算法^[9-10](ilastik^[11])。然而,传统算法例如分水岭法,往往高度依赖预定义参数设置,在处理复杂多变的组织图像时,难以达到足够的准确性和稳定性^[12]。机器学习算法例如随机森林,用于细胞核分割时的性能可能受特征选择和数据噪声的影响。尽管其训练和推理过程相对简单,但模型的表达能力有限,特别是在处理复杂或大规模数据集时,会遇到精度和效率的瓶颈。

深度学习方法在推动细胞核分割技术的进步方面发挥了巨大作用^[13-15],U形架构已成为图像分割任务中的基本框架^[16-17]。基于卷积神经网络(Convolutional Neural Network, CNN)的U形架构广泛应用于细胞核分割任务,U-Net^[18]以其卓越的特征提取能力和图像分割效果而著称,并催生了一系列改进和衍生模型。CENet^[19]将密集空洞卷积块和残差多核池化模块集成到U-Net结构中,创建了用于细胞核分割的新网络。HoverNet^[20]利用细胞核像素的水平和垂直距离来区分聚集的细胞核。尽管CNN在捕捉局部图像特征方面取得了显著成就,但它们的局部感受野也限制了网络捕捉上下文信息的能力。基于Transformer的U型架构(如CellViT^[21])及基于Swin Transformer的Swin-UNet^[22],通过利用视觉Transformer架构改善了对长程依赖的处理能力,但面临着二次计算复杂度的挑战。近年来,基于状态空间模型^[23](State Space Model, SSMs)的Mamba^[24]因在处理序列数据时具有的线性复杂度,而受到广泛关注。该模型可以高效处理长程依赖,很快被引入计算机视觉领域。基于Mamba的U型结构,例如VM-UNet^[25],它以视觉状态空间(Visual State Space, VSS)模块作

为特征提取和重建的核心,兼顾长程依赖的捕获和线性计算复杂度,在细胞核分割任务中的表现优于 CNN 和 Transformer 框架。然而,在细胞核聚集或边界模糊等复杂场景下,为了确保基因映射的可靠性,基于 Mamba 的 U 型架构仍需进一步增强,以满足空间多组学分析对细胞核分割精度的要求。

本文提出了一种结合分层特征细化和频域增强学习的细胞核分割网络(FFVM-UKAN),编解码器整合了用于特征提取的浅层 VSS 模块和用于特征细化的深层令牌化的 Kolmogorov-Arnold 网络(Tokenized Kolmogorov-Arnold Network, Tok-KAN^[26]),并提出了并行频域学习模块(Parallel Frequency Domain Learnable Module, PFDLM)用于跳跃连接,通过在 U 形架构的每一层整合空间和频域信息,使网络能在多个层面上学习和捕获更精细的图像特征。FFVM-UKAN 显著提升了细胞核边缘细节的处理能力,从而提高了细胞核分割精度。

2 原理

如图 1 所示,FFVM-UKAN 包括补丁嵌入层、编码器、解码器、最终投影层和用于跳跃连接的 PFDLM,编码器和解码器由四层 VSS 模块和两层 Tok-KAN 模块组成。浅层 VSS 模块负责从输入图像中提取初步特征;深层 Tok-KAN 模块则进一步细化这些特征,增强模型表达能力。VSS 模块提取的特征通过跳跃连接位置的 PFDLM 输入解码器,在此过程中自适应地进行频域学习以提高分割质量。解码器结合这些特征逐步重建分割图,实现组织切片中细胞核的精确分割。

补丁嵌入层将输入图像 $x \in \mathbf{R}^{H \times W \times 3}$ 分割成不重叠的 4×4 的块,并将图像维度映射到 C (默认为 96) 维。这个过程得到嵌入图像 $x' \in \mathbf{R}^{\frac{H}{4} \times \frac{W}{4} \times C}$ 。随后,应用层归一化,并将补丁嵌入层的输出传递给编码器进行特征提取。编码器由六个阶段组成,其中浅层的 VSS 模块位于编码器的前四阶段。每个阶段使用两个 VSS 块,通道数分别为 $[C, 2C, 4C, 8C]$ 。前三个阶段在 VSS 块后采用补丁合并操作,以减少

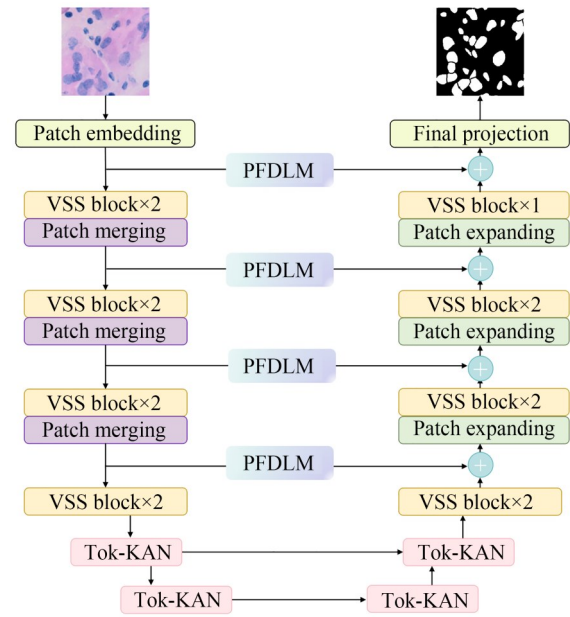


图 1 FFVM-UKAN 的网络结构

Fig. 1 Network architecture of FFVM-UKAN

输入特征的高度和宽度,同时增加通道数。在使用多个 VSS 块进行特征提取后,输出的特征图被令牌化,并由堆叠的 KAN 层细化特征。其中,两个 Tok-KAN 模块的通道数分别为 $10C$ 和 $16C$ 。解码器也分为六个阶段,解码器的前两个阶段由 Tok-KAN 模块组成,通道数分别为 $16C$ 和 $10C$,后四个阶段,每个阶段使用 $[2, 2, 2, 1]$ 个 VSS 块,通道数分别为 $[8C, 4C, 2C, C]$ 。最后三个阶段在 VSS 块之前采用补丁扩展操作,以减少特征通道数,同时增加特征图的高度和宽度。解码器后的最终投影层用于恢复特征尺寸并与分割目标匹配。具体来说,通过补丁扩展层恢复特征的高度和宽度,通过投影层重建通道数。

2.1 编码器和解码器

编码器-解码器架构融合了 Mamba 和 Kolmogorov-Arnold 网络(Kolmogorov-Arnold Network, KAN)。该架构中,浅层的 VSS 模块作为特征提取的核心,深层的 Tok-KAN 模块则进一步细化特征。

2.1.1 浅层特征提取模块

浅层 VSS 模块如图 2(a) 所示,在经过层归一化之后,输入数据分到两个分支处理。在第一个分支中,输入数据依次通过线性层和激活函数。在第二个分支,输入数据先通过线性层、深

度可分离卷积和激活函数处理,然后输入 2D 选择性扫描(2D Selective Scan, SS2D)模块进行特征提取。接着,对特征应用层归一化,并与第一分支的输出进行逐元素相乘,来合并两个分支。最后,使用线性层混合特征,并将结果与残差连接相结合,形成 VSS 模块的输出。其中,使用 SiLU^[27]作为激活函数。

SS2D 模块主要由扫描扩展操作,S6 块和扫描合并操作 3 个部分组成。如图 2(b)所示,扫描扩展操作沿着 4 个不同的方向将输入图像展开成

序列。然后,通过如图 2(c)所示的 S6 块对序列进行特征提取,确保各个方向的信息被彻底扫描,从而捕获不同的特征。S6 模块源自 Mamba,在 S4^[28]的基础上引入基于输入调整参数的选择性机制,使模型可以过滤掉不相关信息,保留相关信息。如图 2(d)所示,扫描合并操作将来自 4 个方向的序列求和并合并,从而将输出图像恢复到与输入图像相同的尺寸。全过程使 SS2D 模块能够有效地提取和整合多向特征,增强了模型处理复杂图像数据的能力。

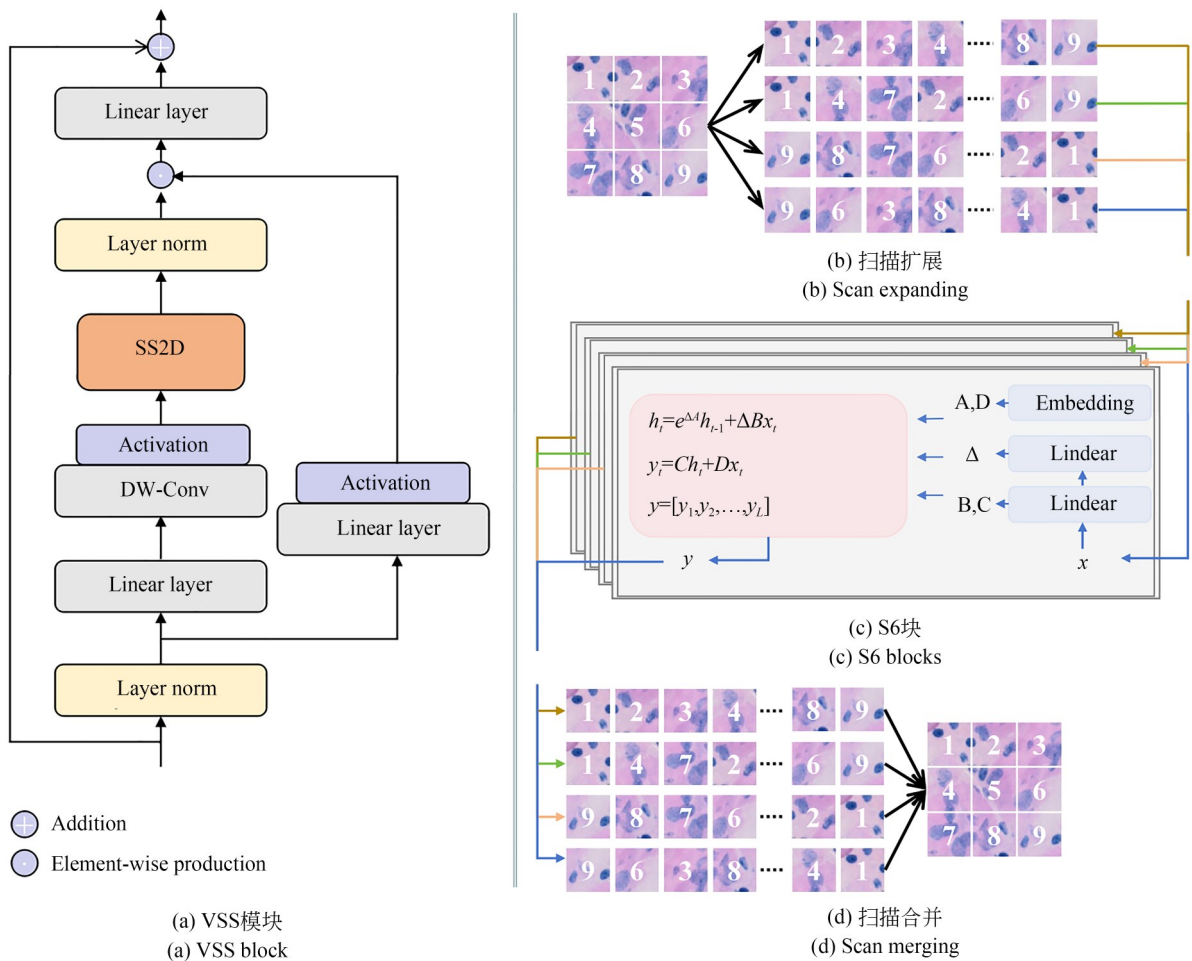


图 2 浅层特征提取模块的网络架构及操作细节

Fig. 2 Network architecture and operational details of shallow feature extraction module

2.1.2 深层特征细化模块

在高维数据中,高效且精确地近似复杂函数十分重要。然而,多层感知机(Multilayer Perceptron, MLP)通过固定的激活函数来引入非线性,

这种固有结构使得它在处理高维数据时存在局限性。具体而言,MLP的计算过程可以表示为:

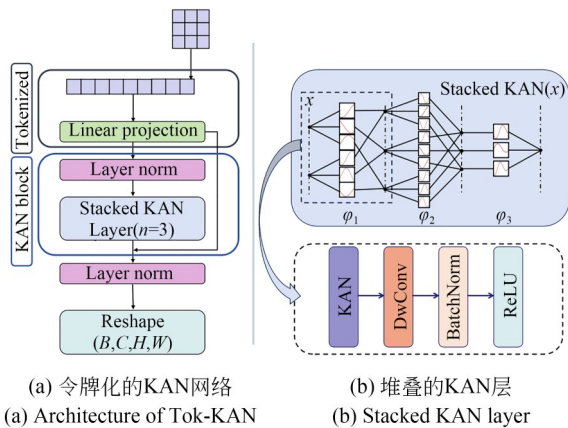
$$MLP(x) = (W_{L-1} \circ \sigma \circ W_{L-2} \circ \sigma \circ \dots \circ W_1 \circ \sigma \circ W_0)x, \quad (1)$$

其中: W 表示线性变换, σ 表示固定的非线性激活函数。这种结构在面对高维数据时, 往往难以充分挖掘复杂的模式和关系。为了解决这一问题, KAN 应运而生, 它通过使用可学习的激活函数替代固定激活函数, 从根本上消除了对线性权重矩阵的依赖, 每个权重参数都可以被一个单变量函数替代, 这些函数通常以样条函数的形式进行参数化。KAN 可以表示为:

$$\text{KAN}(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \dots \circ \Phi_2 \circ \Phi_1)x, \quad (2)$$

其中 Φ 表示可学习的样条激活函数。与传统的固定激活函数相比, 样条函数的灵活性允许网络通过调整其形状来适应建模数据中的复杂关系, 从而最小化近似误差。这种特性显著增强了网络从高维数据中学习细微模式的能力, 对于特征的进一步细化具有重要意义, 使得 KAN 在特征提取和表示上更加精准和高效。

Tok-KAN 模块用于特征细化, 主要包括令牌化嵌入和 KAN 网络两个部分。如图 3(a) 所示, 最后一个 VSS 块的输出特征 M_L 被重塑成一系列大小为 $P \times P$ 的扁平二维块, 特征块的数量为 $N = H_L \times W_L / P^2$ 。这些块随后通过一个可训练的线性投影 $E \in \mathbb{R}^{(N \times C_L) \times D}$ 映射到潜在的嵌入空间 $Z_0 = [M_L^1 E; M_L^2 E; \dots; M_L^N E]$ 中。嵌入的张量经过层归一化后被传递到堆叠的 KAN 层 ($N=3$) 中, 这一过程增强了网络的非线性建模能力。在每个 KAN 层之后, 如图 3(b) 所示, 特征会依次通过深度卷积层 (Depthwise Convolution, DwConv)、批量归一化层 (Batch Normalization,



(a) 令牌化的KAN网络

(a) Architecture of Tok-KAN

(b) 堆叠的KAN层

(b) Stacked KAN layer

图 3 深层特征细化模块的网络架构及操作细节

Fig. 3 Network architecture and operational details of deep feature refinement module

BatchNorm) 和 ReLU 激活函数。在堆叠的 KAN 层之后, 嵌入的令牌通过残差连接加进来。最后, 输出特征在传递到下一个块之前会经过层归一化处理。深层的特征细化模块由两个 Tok-KAN 块组成。编码器中, 第 K 个 Tok-KAN 表示如下:

$$Z_K = \text{LN}(\Theta(Z_{K-1}) + \text{KAN}_{\text{Stacked}}(\text{LN}(\Theta(Z_{K-1}))))), \quad (3)$$

其中: $\Theta(\cdot)$ 代表“令牌化”操作, $\text{KAN}_{\text{Stacked}}$ 表示使用堆叠的 KAN 层进行特征处理和细化。

在解码器的 Tok-KAN 结构中, 特征首先通过一个复合卷积块, 这个块由两个卷积层组成, 每个卷积层后面都紧跟着一个批量归一化层和一个 ReLU 激活函数。之后, 通过双线性插值进行上采样。接着, 编码器中相应层的输出通过跳跃连接添加进来, 然后这些特征进入堆叠的 KAN 块。给定跳跃连接的特征 Z_k 和最后一个上采样块的特征 Z'_{k+1} , 第 k 个上采样块的特征 Z'_k 如下:

$$Z'_k = \text{Cat}(Z'_{k+1}, Z_k), \quad (4)$$

其中 $\text{Cat}(\cdot)$ 表示特征连接操作。

2.2 跳跃连接位置的并行频域学习模块

精准的细胞核分割在很大程度上依赖网络对细微边缘特征的捕捉能力, 这些关键的边缘特征通常与图像频域中的高频分量相对应, 空间域特征提取方法往往难以有效捕捉这些细节。以往的研究成果^[29-30]表明频域方法能够显著促进分割效果。在 FFVM-UKAN 中, VSS 模块提取的特征通过跳跃连接位置的 PFDLM 传输到解码器中, 该模块通过自适应地学习频域特征来增强网络的分割能力, 如图 4 所示。PFDLM 模块采

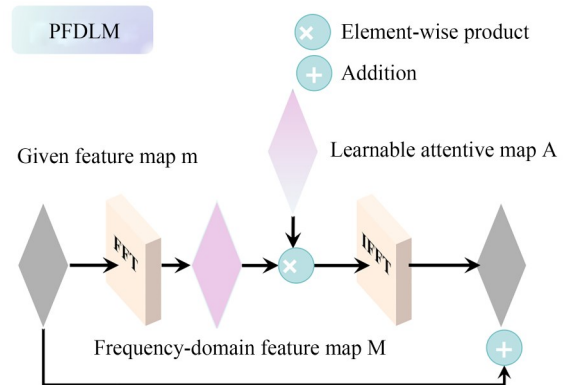


图 4 并行频域学习模块的网络架构

Fig. 4 Network architecture of PFDLM

用双分支策略:第一分支,即频域学习分支,利用参数化的频域注意力图在频域中自适应地学习特征,频域注意力图 A 通过可参数化复值权重矩阵实现,作为频域特征选择器,这些权重参数通过反向传播算法与损失函数梯度引导进行优化,逐步收敛至能够捕获最佳频域特征表示的状态。在前向传播过程中,该复值权重与输入特征的频域表示进行乘积操作,实现对特定频率成分的自适应选择与增强,从而对频域特征图进行全局调制;第二分支则通过残差连接来增强特征。这种设计既考虑了经过一系列变换后的特征,也保留了原始输入特征,因此网络能够综合不同层次的图像特征,捕捉到更精细、更稳健的特征表示,有效地提升了细胞核分割精度。

PFDLM 模块的具体实现如下:首先,沿空间维度对特征 $m \in \mathbf{R}^{H \times W \times C}$ 进行二维快速傅里叶变换(Fast Fourier Transform, FFT),由于 FFT 对实数输入具有 Hermitian 对称性,只保留宽度方向上的前 $W//2+1$ 列,避免输出冗余,即:

$$M = F[m] \in \mathbf{C}^{H \times (W//2+1) \times C}, \quad (5)$$

其中: $F[\cdot]$ 表示二维 FFT, M 为输入特征的频域表示。然后,将 M 与参数化可学习的注意力图 $A \in \mathbf{C}^{H \times (W//2+1) \times C}$ 相乘来调制频谱。可学习的注意力图定义为形状为 $(C, H, W//2+1, 2)$ 的复值参数张量,其中 C 表示通道数, H 和 W 分别代表输入 PFDLM 模块前特征图的高度与宽度,最后一个维度 2 对应复数的实部与虚部。这些参数通过梯度下降优化算法进行端到端训练,使网络能够自适应地学习最优频域表示,并对频域特征图 M 进行全局调制,得到:

$$M' = A \odot M, \quad (6)$$

其中: \odot 表示逐元素乘积, M' 表示调制后的频域特征。理论上,该设计构建了一种频域选择性滤波器,允许模型对不同频率分量进行差异化增强或抑制,从而捕获对任务性能至关重要的频域特征。之后,通过逆快速傅里叶变换(Inverse Fast Fourier Transform, IFFT)将 M' 恢复到空间域,即:

$$m' = F^{-1}[M'] \in \mathbf{R}^{H \times W \times C}, \quad (7)$$

其中 $F^{-1}[\cdot]$ 表示 IFFT。最后,在第二分支中,将原始特征 m 与经过第一分支处理的特征图 m' 相加,利用残差连接实现特征的融合与增

强,即:

$$m_{\text{out}} = m + m'. \quad (8)$$

PFDLM 可视为一种可学习的频域滤波器,区别于局部调整特征的空间注意力机制,它能够对图像中的特定频率成分进行全局性的调整,从而捕捉到对细胞核分割至关重要的边缘和纹理等特征。通过整合空间域和频域信息,PFDLM 在细胞核密集或边界不清晰的区域中,更有效地识别和保留这些关键特征,有效提高了分割的准确性。

3 实验及结果分析

3.1 实验数据集

实验使用两个数据集(MoNuSeg^[31]和一个私有数据集^[32])进行分割评估,并使用 10X Genomics 的人类肝脏数据集^[33]进行进一步的基因映射验证。MoNuSeg 数据集包含 30 张训练图像和 14 张测试图像,每张图像的尺寸为 $1\,000 \times 1\,000$ 像素,来源不同器官的全切片。为了构建数据集,从每张图像的 4 个角落提取 4 个 512×512 像素的图像块,并将每个图像块按 1:1 的宽高比均匀划分为 4 个 256×256 像素的小块。最终,数据集由 480 张训练图像和 224 张测试图像构成,每张图像均为 256×256 像素。

私有数据集由小鼠肾脏组织的 H&E 染色图像组成。为了促进后续的实例分割,在贴近细胞核的边缘位置引入一像素宽的细线作为分隔。经过裁剪和数据清洗后,共获得 367 张 256×256 像素的图像。最终,私有数据集由 276 张训练图像、59 张验证图像和 32 张测试图像构成。

来自 10X Genomics 的人类肝脏数据集包含基因表达数据,涵盖 377 个靶向基因。肝脏全切片被分割为 256×256 像素的图像块,并使用在 MoNuSeg 数据集上训练的模型进行细胞核分割测试。由于缺乏金标准,从中随机选择 100 张图像进行数据清洗和手动细胞核标注,并通过与模型输出结果比较来评估其泛化能力。

模型训练过程采用了一系列的数据增强手段:按照 50% 的概率随机水平翻转图像,以 50% 的概率随机垂直翻转图像,以 50% 的概率将图像随机旋转 $0^\circ \sim 360^\circ$ 之间的一个角度。

3.2 实验环境和参数设置

实验使用的 PyTorch 版本为 1.13.0+cu117, 在 Nvidia A100 上进行。在 MoNuSeg 数据集和私有数据集上进行对比实验, 按照文章中^[18,22,25,34]的参数设置执行。消融实验过程中, 所有模型的初始学习率设置为 0.000 5, 使用余弦退火的学习率衰减策略, 批量大小为 32。使用 AdamW 优化器, 损失函数为 BceDiceLoss, 训练轮次设为 50。这里将编码器和解码器浅层特征提取模块的权重初始化为在 ImageNet-1k 上预训练的 VMamba-T 模型的权重。

3.3 评价指标

为了对实验结果进行定量分析, 本文采用以下指标: 平均交并比 (mean Intersection over Union, mIoU)、Dice 系数 (Dice Coefficient, Dice)、准确率 (Accuracy, Acc)、灵敏度 (Sensitivity, Sen) 和特异性 (Specificity, Spe), 从多个角度对细胞核分割的性能进行评估。其计算公式如下:

$$mIOU = \frac{TP}{TP + FP + FN}, \quad (9)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN}, \quad (10)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (11)$$

$$Sen = \frac{TP}{TP + FN}, \quad (12)$$

$$Spe = \frac{TN}{TN + FP}, \quad (13)$$

其中: 真阳性 (True Positive, TP) 指正确识别为属于阳性类别的像素数量; 假阳性 (False Positive, FP) 指错误地将实际属于阴性类别的像素识别为阳性类别的数量; 真阴性 (True Negative, TN) 指正确识别为属于阴性类别的像素数量; 假阴性 (False Negative, FN) 指错误地将实际属于

阳性类别的像素识别为阴性类别的数量。

本文将上述基于图像的评估矩阵迁移到空间多组学领域来评估基因映射的准确性。与传统的计算像素分类情况的混淆矩阵不同, 本文构建了专用于评估基因是否正确映射到细胞核内外的混淆矩阵。具体而言, 真阳性 (TP_{gene}) 指在真实值和预测结果中均位于细胞核内的基因数量; 真阴性 (TN_{gene}) 指在真实值和预测结果中均位于细胞核外的基因数量; 假阳性 (FP_{gene}) 指模型错误地将实际位于细胞核外的基因预测为细胞核内的基因数量; 假阴性 (FN_{gene}) 指模型错误地将实际位于细胞核内的基因预测到细胞核外的基因数量。本研究还基于该混淆矩阵, 计算了基因的 Dice 相似系数 $Dice_{gene}$ 、准确率 (Acc_{gene})、灵敏度 (Sen_{gene})、特异性 (Spe_{gene}) 以及平均交并比 ($mIoU_{gene}$) 来定量评估基因映射。

$$mIOU_{gene} = \frac{TP_{gene}}{TP_{gene} + FP_{gene} + FN_{gene}}, \quad (14)$$

$$Dice_{gene} = \frac{2 \times TP_{gene}}{2 \times TP_{gene} + FP_{gene} + FN_{gene}}, \quad (15)$$

$$Acc_{gene} = \frac{TP_{gene} + TN_{gene}}{TP_{gene} + TN_{gene} + FP_{gene} + FN_{gene}}, \quad (16)$$

$$Sen_{gene} = \frac{TP_{gene}}{TP_{gene} + FN_{gene}}, \quad (17)$$

$$Spe_{gene} = \frac{TN_{gene}}{TN_{gene} + FP_{gene}}. \quad (18)$$

3.4 消融实验

通过消融实验来证明 PFDLM 和 Tok-KAN 模块的有效性, 结果如表 1 和图 5 所示。其中, FFVM-UNet 表示用 PFDLM 取代 VM-UNet 原本跳跃连接的新结构, VM-UKAN 表示用分层特征细化编解码器取代 VM-UNet 原始编码器的新结构。

表 1 私有数据集和 MoNuSeg 数据集上的消融实验

Tab. 1 Ablation experiments on in-house and public MoNuSeg datasets

Model	PFDLM	Tok-KAN	In-house dataset					MoNuSeg dataset				
			mIoU	Dice	Acc	Spe	Sen	mIoU	Dice	Acc	Spe	Sen
VM-UNet			0.845 4	0.916 2	0.961 1	0.973 0	0.921 3	0.665 4	0.799 1	0.916 8	0.936 7	0.836 1
FFVM-UNet	✓		<u>0.855 3</u>	<u>0.922 0</u>	<u>0.964 1</u>	0.977 5	<u>0.919 2</u>	<u>0.685 2</u>	<u>0.813 2</u>	<u>0.922 8</u>	<u>0.941 2</u>	0.848 4
VM-UKAN		✓	0.850 6	0.919 2	0.963 4	<u>0.981 6</u>	0.902 6	0.678 7	0.808 6	0.920 8	0.939 5	0.845 1
FFVM-UKAN	✓	✓	0.859 5	0.924 5	0.965 6	0.981 9	0.911 3	0.690 9	0.817 2	0.925 0	0.944 2	<u>0.847 3</u>

Note: Bolded text indicates the best results in each column, and underlined text indicates the second-best results in each column.

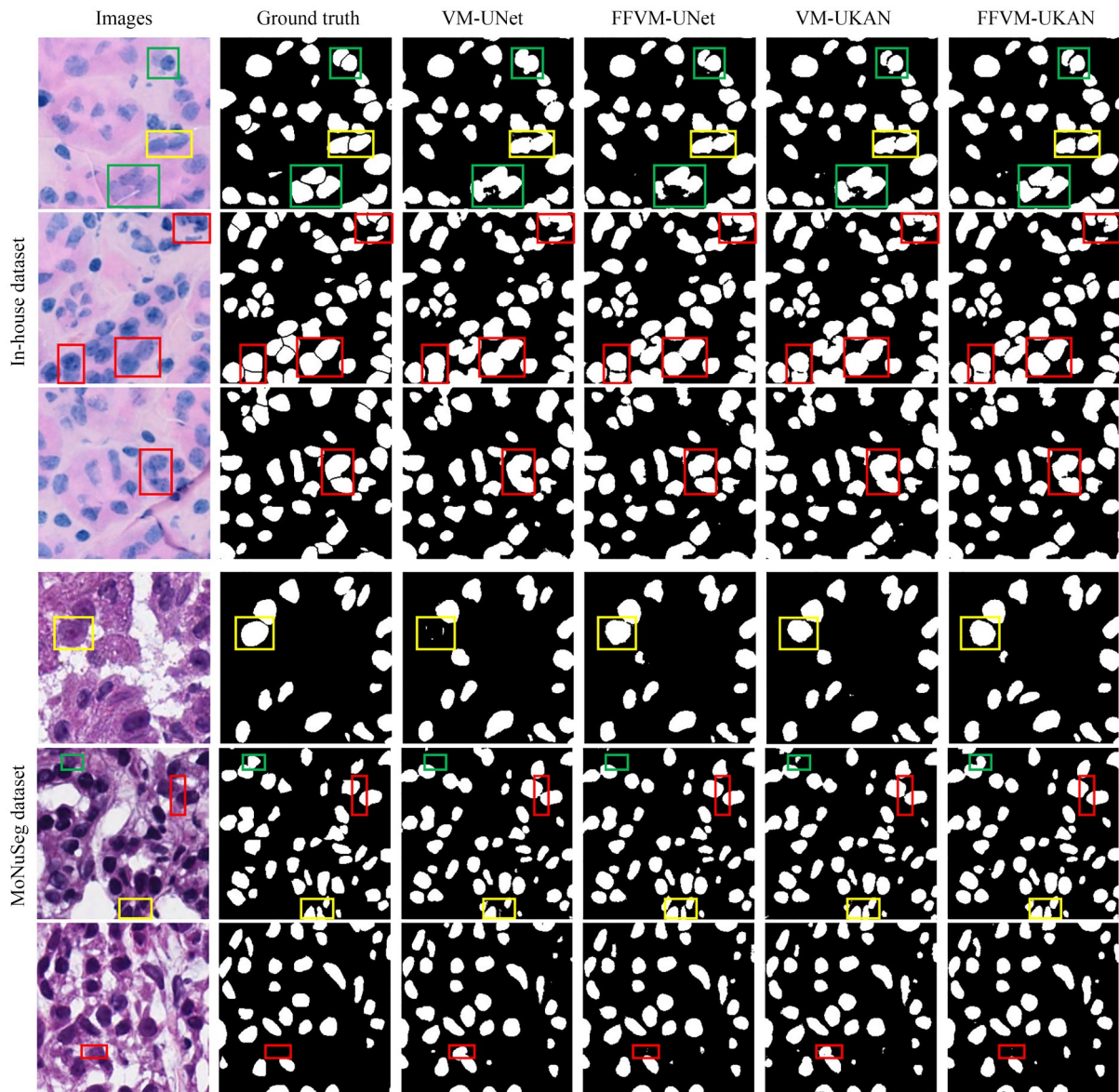


图5 模块消融后不同方法的分割结果比较

Fig. 5 Comparison of different segmentation results after ablation of each module

在私有数据集上,FFVM-UKAN在mIoU, Dice, Acc, Sen和Spe指标上分别达到了0.8595,0.9245,0.9656,0.9819和0.9113;在MoNuSeg数据集上,相应的值分别为0.6909,0.8172,0.9250,0.9442和0.8473。两个模块的加入显著提升了mIoU, Dice和Acc等指标。与VM-UNet相比,FFVM-UNet和VM-UKAN在私有数据集上的mIoU分别提高了1.2%和0.6%,在MoNuSeg数据集上分别提高了3.0%和2.0%。如图5所示(彩图见期刊电子版),PFDLM模块显著增强了红色标记区域内的分割

效果,通过频域学习更精确地分割细胞核贴近或边界模糊区域。绿色标记区域展示了对细胞核边界的细化,这得益于Tok-KAN卓越的非线性建模能力,黄色标记区域的分割改善是两者共同作用的结果。

为了进一步分析这两个模块的具体作用,实验采用梯度加权类激活映射(Gradient-weighted Class Activation Mapping, Grad-CAM^[35])方法生成类激活热力图。图6是VM-UNet,FFVM-UNet,VM-UKAN和FFVM-UKAN的类激活热力图,通过对比不同列的热力响应强度解释模型

决策。

通过对比 VM-UNet 列和 FFVM-UNet 列解释模型决策可知:PFDLM 模块通过频域变换建立了跨区域的长程依赖关系,细胞核区域热力响应范围显著扩展;将 PFDLM 模块集成到 U 型架构的不同层次,通过融合多尺度信息,不同尺寸的细胞核均呈现更均匀的热力分布,印证了表 1

中 FFVM-UNet 相较于 VM-UNet 在 mIoU, Dice, Acc, Spe 均有所提升的结果,验证了 PFDLM 模块在促进细胞核分割方面的作用。

通过对比 FFVM-UNet 列和 FFVM-UKAN 列解释模型决策可知:Tok-KAN 模块的加入使细胞核边缘或染色较浅的区域热力响应强度有所提升,说明 Tok-KAN 模块通过可学习的非线性

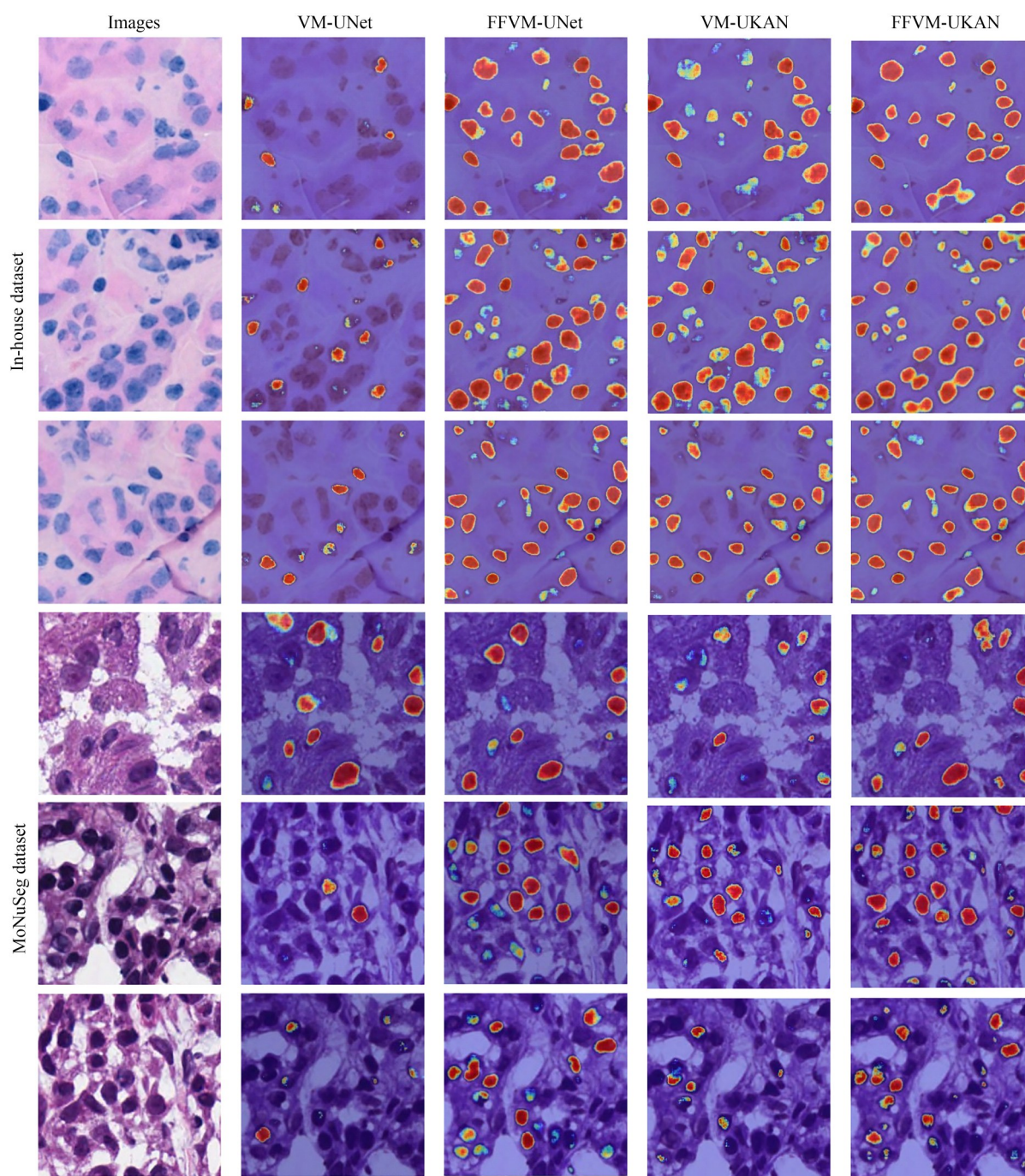


图 6 消融实验中不同方法的类激活热力图

Fig. 6 Class activation heatmaps with different methods in ablation experiments

性样条函数增强了模型的特征表达能力,使得模型更加关注易忽略的细微结构,尤其是切片中不明显的细胞核区域或背景结构复杂的区域,印证了表 1 中 FFVM-UKAN 相较于 FFVM-UNet 在 mIoU, Dice, Acc, Spe 均有所提升的结果,表明 Tok-KAN 模块对细胞核分割具有促进作用。该模块显著提升了模型在处理不同特征的组织区域的分割性能,并能够精确捕捉细微且易混淆的特征和边缘信息,从而为细胞核分割提供更为丰富和准确的支持。两个模块协同作用显著提升了整体的分割性能,PFDLM 增强全局信息建模能力,而 Tok-KAN 聚焦局部特征校准,有利于在空间多组学中的应用。

3.5 对比实验

为了验证模型的优越性,将本文提出的 FFVM-UKAN 与 UNet^[18], UNet++^[34], Swin-

Unet^[22] 和 VM-UNet^[25] 进行了对比,实验结果如表 2 所示,FFVM-UKAN 在 mIoU, Dice, Acc 等指标上均优于其他模型。在私有数据集上,FFVM-UKAN 相比 UNet, UNet++, Swin-UNet 和 VM-UNet 的 mIoU 分别高出 5.0%, 4.4%, 5.9% 和 1.7%。在 MoNuSeg 数据集上,FFVM-UKAN 的 mIoU 分别比 UNet, UNet++, Swin-UNet 和 VM-UNet 分别高出 14.3%, 15.0%, 10.2% 和 3.8%。

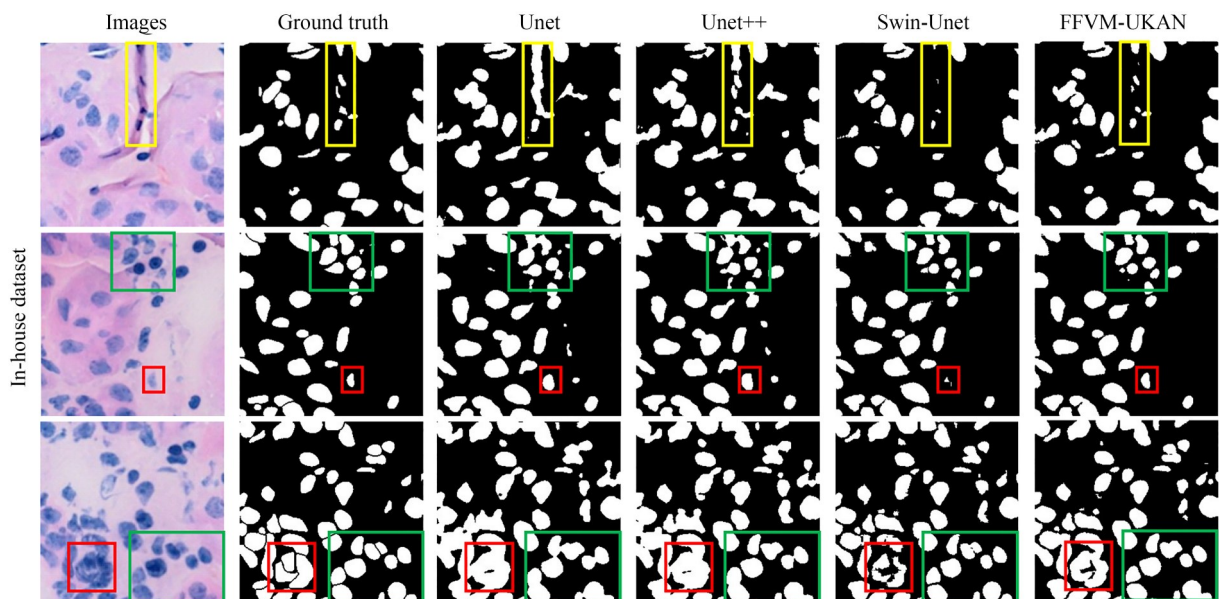
UNet 和 UNet++ 在细胞核边界位置常常出现分割错误,尤其在细胞核聚集区域,容易将背景误分类为细胞核。尽管 Swin-UNet 在特异性上表现最好,但它极易将原本属于细胞核的像素误分类为背景。如图 7 所示(彩图见期刊电子版),在红色标记区域,FFVM-UKAN 相较于 Swin-UNet 展现出更准确的分割效果;而在绿色

表 2 私有数据集和 MoNuSeg 数据集上不同方法的对比

Tab. 2 Comparison of different methods on in-house and public MoNuSeg datasets

Model	In-house dataset					MoNuSeg dataset				
	mIoU	Dice	Acc	Spe	Sen	mIoU	Dice	Acc	Spe	Sen
UNet ^[18]	0.818 4	0.900 2	0.953 7	0.968 6	0.903 8	0.604 2	0.746 5	0.887 0	0.898 7	0.850 8
UNet++ ^[34]	0.823 0	0.902 9	0.954 8	0.968 1	0.910 2	0.600 9	0.745 3	0.892 2	0.912 4	0.818 9
Swin-UNet ^[22]	0.811 3	0.895 8	0.950 6	0.991 1	0.832 2	0.627 2	0.770 8	0.914 6	0.961 3	0.725 5
VM-UNet ^[25]	<u>0.845 4</u>	<u>0.916 2</u>	<u>0.961 1</u>	0.973 0	0.921 3	<u>0.665 4</u>	<u>0.799 1</u>	<u>0.916 8</u>	0.936 7	0.836 1
FFVM-UKAN	0.859 5	0.924 5	0.965 6	<u>0.981 9</u>	<u>0.911 3</u>	0.690 9	0.817 2	0.925 0	<u>0.944 2</u>	<u>0.847 3</u>

Note: Bolded text indicates the best results in each column, and underlined text indicates the second-best results in each column.



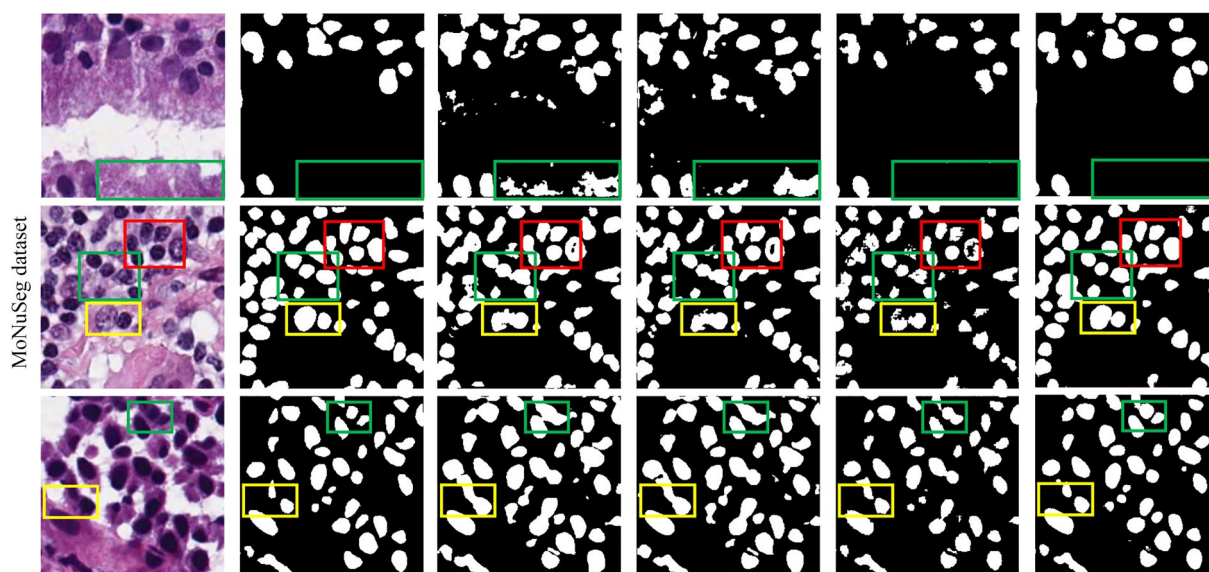


图7 私有数据集和MoNuSeg数据集上不同方法的比较

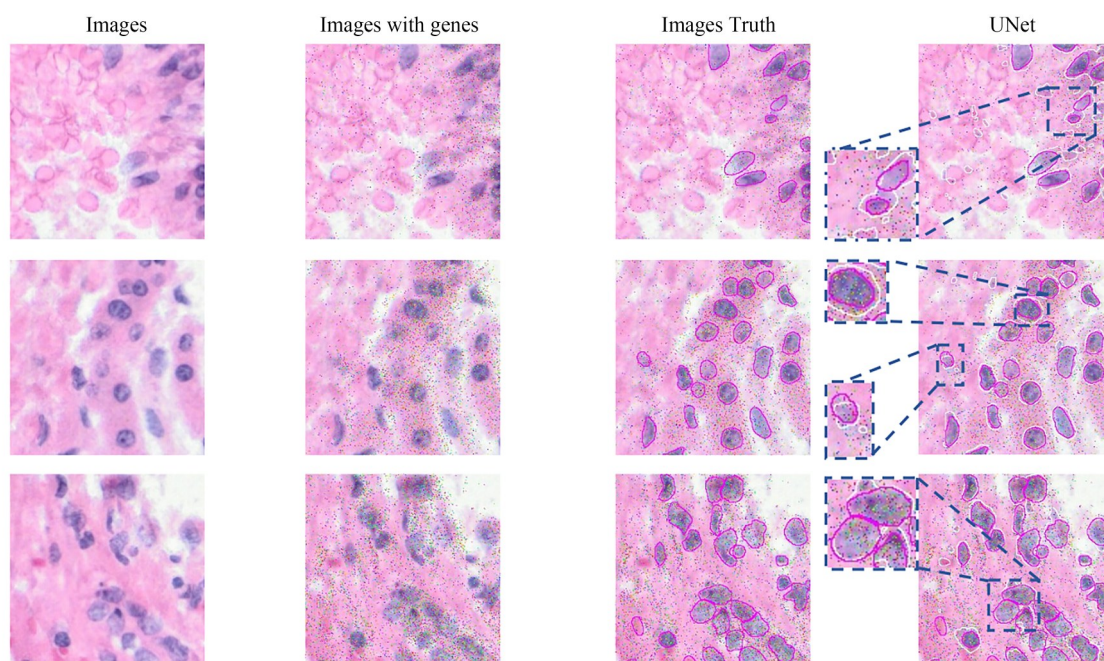
Fig. 7 Comparison of different methods on in-house and public MoNuSeg datasets

标记区域,FFVM-UKAN在细胞核聚集区域的分割效果明显优于UNet和UNet++。综合来看,在处理复杂的组织结构时,FFVM-UKAN能够提供更加精细和准确的分割结果。

3.6 FFVM-UKAN在空间多组学中的应用

为验证本文模型在空间基因组的分析效果,采用来自10X Genomics的人类肝脏数据集进行测试,分割结果如图8所示。单像素的彩色点代表不同的基因。可见UNet和UNet++相较于

FFVM-UKAN过分割,导致部分位于细胞核外部的基因被错误分配到细胞核内;Swin-UNet相较于FFVM-UKAN欠分割,导致部分位于细胞核内部的基因被错误分配到细胞核外;FFVM-UKAN相较于VM-UNet,有更多细胞核区域的正确识别以及边缘位置更精细的处理,在细胞核边缘位置上基因到细胞核的映射准确率更高。分割性能及基因到细胞核的映射指标详见表3,在基因映射方面,FFVM-UKAN的 $mIoU_{gene}$,



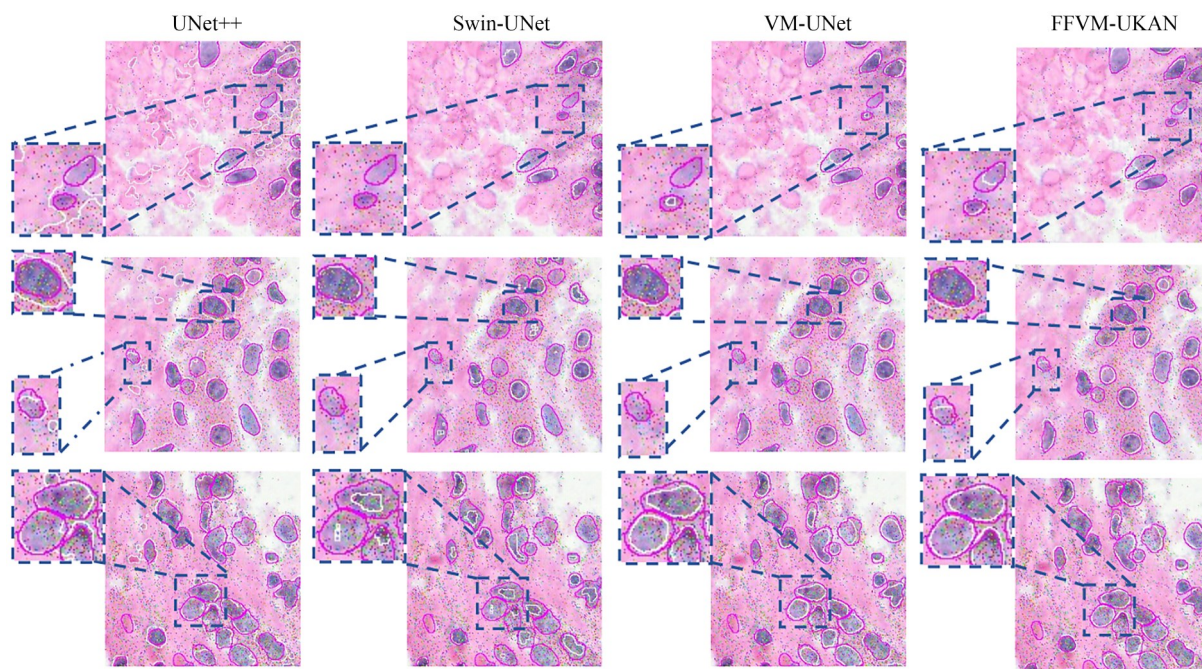


图 8 10X Genomics 人类肝脏样本上不同方法的比较

Fig. 8 Comparison of different methods on human liver sample from 10X Genomics

表 3 10X Genomics 数据集上基于细胞核分割和基因映射的对比实验

Tab. 3 Comparison experiments on human liver sample from 10X Genomics with nuclei segmentation and gene mapping

Model	mIoU	Dice	Acc	Spe	Sen	Average time/s	mIoU _{gene}	Dice _{gene}	Acc _{gene}	Spe _{gene}	Sen _{gene}
UNet	<u>0.687 8</u>	<u>0.815 0</u>	0.931 7	0.947 9	0.855 9	0.048 7	0.690 7	0.817 0	0.870 4	0.844 0	0.928 8
UNet++	0.677 3	0.807 6	0.931 4	0.955 6	<u>0.819 0</u>	0.076 7	<u>0.696 8</u>	0.821 3	0.879 0	0.872 9	<u>0.892 6</u>
Swin-UNet	0.320 9	0.469 6	0.877 4	0.998 0	0.325 1	<u>0.020 3</u>	0.331 1	0.497 5	0.789 7	0.995 9	0.334 1
VM-UNet	0.681 5	0.809 1	<u>0.939 6</u>	<u>0.986 2</u>	0.733 7	0.018 9	0.699 1	<u>0.822 9</u>	<u>0.900 6</u>	<u>0.972 7</u>	0.741 2
FFVM-UKAN	0.700 5	0.822 4	0.942 9	0.980 0	0.779 9	0.048 1	0.723 1	0.839 3	0.906 3	0.961 1	0.785 3

Note: Bolded text indicates the best results in each column, and underlined text indicates the second-best results in each column.

Dice_{gene}, Acc_{gene}, Spe_{gene}, Sen_{gene} 分别为 0.723 1, 0.839 3, 0.906 3, 0.961 1 和 0.785 3, 在 mIoU_{gene}, Dice_{gene}, Acc_{gene} 3 个指标上均明显优于其他对比方法。在模型推理效率方面, UNet, UNet++, Swin-UNet, VM-UNet 和 FFVM-UKAN 对每张图像的平均推理时间分别约为 0.048 7, 0.076 7, 0.020 3, 0.018 9 和 0.048 1 s。尽管 FFVM-UKAN 的推理速度相较于 VM-UNet 有所增加, 但在基因映射准确率上的显著提升弥补了这一差距, 并且其在推理时间上与 UNet 接近, 优于 UNet++。空间多组学研究中细胞核分割的实时性要求相对较低, 因此, FFVM-UKAN 以略

微增加的推理时间换取更高的基因映射准确率是值得的。

4 结 论

本文提出了面向空间多组学细胞核分割的网络架构 FFVM-UKAN。它采用创新的编码器-解码器结构, 并在跳跃连接处加入了 PFDLM 模块, 显著提高了细胞核分割的准确性。与传统使用固定激活函数引入非线性的神经网络不同, 该架构在编解码器的较深层次引入可学习的样条激活函数, 增强网络的非线性建模能力, 深度

整合了用于特征提取的浅层 VSS 模块及用于特征细化的深层 Tok-KAN; 针对细胞核聚集区域易出现的欠分割和过分割问题, 提出一种并行频域学习模块 PFDLM, 通过频域学习增强了网络的细胞核分割能力; FFVM-UKAN 在公开数据集 MoNuSeg^[27-28] 和私有数据集的细胞核分割任务中均具有优异表现。在私有数据集上, FFVM-UKAN 的 mIoU, Dice, Acc, Spe 和 Sen 上分别为 0.859 5, 0.924 5, 0.965 6, 0.981 9 和 0.911 3; 在 MoNuSeg 数据集上分别为 0.690 9, 0.817 2, 0.925 0, 0.944 2 和 0.847 3, 展现出稳健的性能。为了验证其在空间多组学领域的应用价值, 本文将在 MoNuSeg 数据集上训练的模型

应用于 10X Genomics 的人类肝脏样本, 达到了 90.63% 基因映射准确率。这一结果展示了该模型在空间多组学分析中的应用潜力, 为推动空间多组学研究提供了更为精准和高效的分析工具。

作者贡献声明:

李修齐: 方法提出, 论文构思和撰写;

李金泽: 论文审核与指导;

杨弃: 方法指导;

李莹雪: 数据集提供;

赵才荣、周连群: 论文指导;

姚佳: 论文指导与编辑。

参考文献:

- [1] VANDEREYKEN K, SIFRIM A, THIENPONT B, *et al.* Methods and applications for single-cell and spatial multi-omics [J]. *Nature Reviews Genetics*, 2023, 24(8): 494-515.
- [2] LIU X J, PENG T, XU M C, *et al.* Spatial multi-omics: deciphering technological landscape of integration of multi-omics and its applications [J]. *Journal of Hematology & Oncology*, 2024, 17(1): 72.
- [3] BRESSAN D, BATTISTONI G, HANNON G J. The dawn of spatial omics [J]. *Science*, 2023, 381(6657): eabq4964.
- [4] FISCHER E G. Nuclear morphology and the biology of cancer cells [J]. *Acta Cytologica*, 2020, 64(6): 511-519.
- [5] KOCHETOV B, BELL P D, GARCIA P S, *et al.* UNSEG: unsupervised segmentation of cells and their nuclei in complex tissue samples [J]. *Communications Biology*, 2024, 7(1): 1062.
- [6] CHLIPALA E, BENDZINSKI C M, CHU K, *et al.* Optical density-based image analysis method for the evaluation of hematoxylin and eosin staining precision [J]. *Journal of Histotechnology*, 2020, 43(1): 29-37.
- [7] 宋杰, 肖亮, 练智超. 级联稀疏卷积与决策树集成的病理图像细胞核分割方法 [J]. *自动化学报*, 2021, 47(2): 378-390.
SONG J, XIAO L, LIAN ZH CH. Cascade sparse convolution and decision tree ensemble model for nuclear segmentation in pathology images [J]. *Acta Au-*
- [8] tomatica Sinica, 2021, 47(2): 378-390. (in Chinese)
- [8] MOFFITT J R, BAMBAH-MUKKU D, EICHORN S W, *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region [J]. *Science*, 2018, 362(6416): eaau5324.
- [9] LOHOFF T, GHAZANFAR S, MISSAROVA A, *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis [J]. *Nature Biotechnology*, 2022, 40(1): 74-85.
- [10] JACKSON H W, FISCHER J R, ZANOTELLI V R T, *et al.* The single-cell pathology landscape of breast cancer [J]. *Nature*, 2020, 578(7796): 615-620.
- [11] BERG S, KUTRA D, KROEGER T, *et al.* Ilas-tik: interactive machine learning for (bio) image analysis [J]. *Nature Methods*, 2019, 16(12): 1226-1232.
- [12] MAHONY N O, CAMPBELL S, CARVALHO A, *et al.* Deep learning vs. traditional computer vision [EB/OL]. 2019: 1910.13796. <https://arxiv.org/abs/1910.13796v1>.
- [13] 张伟, 余浩, 袁波, 等. 基于高倍率细胞内镜系统的细胞核分割 [J]. *光学精密工程*, 2021, 29(11): 2574-2580.
ZHANG W, YU H, YUAN B, *et al.* Nuclear segmentation based on endocytoscopy system with high magnification [J]. *Opt. Precision Eng.*, 2021, 29(11): 2574-2580. (in Chinese)
- [14] ARSLAN S, ERSAHIN T, CETIN-ATALAY R, *et al.* Attributed relational graphs for cell nucle-

- us segmentation in fluorescence microscopy images [J]. *IEEE Transactions on Medical Imaging*, 2013, 32(6): 1121-1131.
- [15] KE J, LU Y Z, SHEN Y Q, *et al.* ClusterSeg: a crowd cluster pinpointed nucleus segmentation framework with cross-modality datasets[J]. *Medical Image Analysis*, 2023, 85: 102758.
- [16] 李大湘, 杨福杰, 刘颖, 等. 融入交叉注意力编码的皮肤病变分割网络[J]. *光学精密工程*, 2024, 32(4): 609-621.
- LI D X, YANG F J, LIU Y, *et al.* Skin lesion segmentation network with cross-attention coding [J]. *Opt. Precision Eng.*, 2024, 32(4): 609-621. (in Chinese)
- [17] 梁礼明, 周珑颂, 冯骏, 等. 基于高分辨率复合网络的皮肤病变分割[J]. *光学精密工程*, 2022, 30(16): 2021-2038.
- LIANG L M, ZHOU L S, FENG J, *et al.* Skin lesion segmentation based on high-resolution composite network[J]. *Opt. Precision Eng.*, 2022, 30(16): 2021-2038. (in Chinese)
- [18] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*. Cham: Springer International Publishing, 2015: 234-241.
- [19] GU Z W, CHENG J, FU H Z, *et al.* CE-net: context encoder network for 2D medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(10): 2281-2292.
- [20] GRAHAM S, VU Q D, RAZA S E A, *et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images [J]. *Medical Image Analysis*, 2019, 58: 101563.
- [21] HÖRST F, REMPE M, HEINE L, *et al.* CellViT: Vision Transformers for precise cell segmentation and classification[J]. *Medical Image Analysis*, 2024, 94: 103143.
- [22] CAO H, WANG Y Y, CHEN J, *et al.* Swin-unet: unet-like pure transformer for Medical image segmentation [C]. *Computer Vision-ECCV 2022 Workshops*. Cham: Springer Nature Switzerland, 2023: 205-218.
- [23] KALMAN R E. A new approach to linear filtering and prediction problems[J]. *Journal of Basic Engineering*, 1960, 82(1): 35-45.
- [24] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[J]. *ArXiv e-Prints*, 2023: arXiv: 2312. 00752.
- [25] Ruan J, Xiang S. Vm-unet: Vision mamba unet for medical image segmentation [J]. *arXiv preprint arXiv:240202491*, 2024.
- [26] LI C X, LIU X Y, LI W Y, *et al.* U-KAN makes strong backbone for medical image segmentation and generation[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, 39(5): 4652-4660.
- [27] KUMAR N, VERMA R, ANAND D, *et al.* A multi-organ nucleus segmentation challenge [J]. *IEEE Trans Med Imaging*, 2020, 39(5): 1380-1391.
- [28] KUMAR N, VERMA R, SHARMA S, *et al.* A dataset and a technique for generalized nuclear segmentation for computational pathology [J]. *IEEE Transactions on Medical Imaging*, 2017, 36(7): 1550-1560.
- [29] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning [J]. *Neural Networks*, 2018, 107: 3-11.
- [30] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces [J]. *arXiv preprint arXiv:211100396*, 2021.
- [31] WU J D, FU R, FANG H H, *et al.* MedSeg-Diff: medical image segmentation with diffusion probabilistic model[EB/OL]. 2022: 2211.00611. <https://arxiv.org/abs/2211.00611v5>.
- [32] WU J D, JI W, FU H Z, *et al.* MedSegDiff-V2: diffusion based medical image segmentation with transformer [EB/OL]. 2023: 2301.11798. <https://arxiv.org/abs/2301.11798v2>.
- [33] 10X Genomics, Human liver data with xenium human multi-tissue and cancer panel [D]. 2024, <https://www.10xgenomics.com/cn/datasets>
- [34] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, *et al.* UNet++: a nested U-Net architecture for medical image segmentation [J]. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, 11045: 3-11.
- [35] SELVARAJU R, COGSWELL M, DAS A, *et*

al. Grad-cam: Visual explanations from deep networks via gradient-based localization [C]. *Proceedings of the IEEE International Conference on Computer Vision*. 2017: 618-626.

ceedings of the IEEE International Conference on Computer Vision. 2017: 618-626.

作者简介:



李修齐(2000—),女,河北邢台人,硕士研究生,2022年于东北大学获得学士学位,主要从事图像分割及配准方面的研究。E-mail:kiki_li@mail.ustc.edu.cn

通讯作者:



周连群(1981—),男,山东金乡人,研究员,博士生导师,2010年分别于法国 Université de Franche-Comté 大学和中国科学院研究生院获得博士学位,主要从事微纳生物传感器及系统等方面的研究。E-mail:zhoulq@sibet.ac.cn

通讯作者:



姚佳(1986—),男,江苏苏州人,副研究员,2011年于南京航空航天大学获得硕士学位,2022年于苏州大学获得博士学位,主要从事生物传感芯片研发及应用等方面的研究。E-mail:yaojia@sibet.ac.cn