

文章编号:1007-2780(2026)02-0303-15

基于 ViT 与联合优化策略的医学图像分割

马 龙¹, 白 昆¹, 冯星泰¹, 檀帅兵², 王璐瑶², 薛 尧², 卢弈斐³, 王晓田^{3*}

(1. 西安现代控制技术研究, 陕西 西安 710065;

2. 西安交通大学 信息与通信工程学院, 陕西 西安 710049;

3. 西北工业大学 无人系统技术研究院, 陕西 西安 710072)

摘要:细胞分割是生物医学影像分析的基本任务,在病理诊断和细胞生物学中具有重要应用价值。精确的细胞核分割对疾病筛查、肿瘤微环境分析和治疗方案制定至关重要。然而,组织图像中细胞形态高度异质、重叠频繁,加之噪声与照明不均,显著增加了自动分割的难度。尽管传统与深度学习方法已获进展,但在细胞重叠、边界模糊等复杂条件下仍面临明显局限。为此,本文提出一种基于视觉变换器(VisionTransformer, ViT)的NucleiViT模型,通过结合特征联合优化、注意力机制优化与区域边界感知损失函数的设计,实现细胞精准检测及分割。该模型的核心创新在于结合滑动窗口膨胀注意力机制,在共享模块中实现不同任务分支的特征交互,最大化多任务学习优势;同时引入联合边界区域感知损失函数,增强对细胞边界的识别灵敏度。实验结果表明,NucleiViT在细胞分割精度和计算效率上均优于现有主流方法,尤其在细胞密集重叠场景中表现出更优的鲁棒性,有效服务于临床病理图像分析。

关键词:细胞分割;多任务联合优化;边界感知损失;神经网络

中图分类号:R318;TP391.4 文献标识码:A doi:10.37188/CJLCD.2025-0222 CSTR:32172.14.CJLCD.2025-0222

Medical image segmentation based on ViT and joint optimization strategy

MA Long¹, BAI Kun¹, FENG Xingtai¹, TAN Shuaibing², WANG Luyao²,
XUE Yao², LU Yifei³, WANG Xiaotian^{3*}

(1. Xi'an Modern Control Technology Research Institute, Xi'an 710065, China;

2. School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

3. Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Cell segmentation is a fundamental task in biomedical image analysis, holding significant application value in pathological diagnosis and cell biology. Accurate nuclear segmentation is crucial for disease screening, tumor microenvironment analysis, and treatment planning. However, the high morphological heterogeneity and frequent overlapping of cells in tissue images, coupled with noise and

收稿日期:2025-11-01;修订日期:2025-12-06.

基金项目:国家自然科学基金(No.62106193);上海航天八院自主研发基金(No.D5203240888);中央高校基本科研业务费(No.D5000240033)

Supported by National Natural Science Foundation of China (No.62106193); Shanghai Academy of Spaceflight Technology Self-Research and Development Fund (No.D5203240888); Fundamental Research Funds for Central Universities (No.D5000240033)

*通信联系人, E-mail:xtwang0111@nwpu.edu.cn

uneven illumination, significantly increase the difficulty of automated segmentation. Despite advances in traditional and deep learning methods, they still face notable limitations under complex conditions such as cell overlap and blurred boundaries. To address these challenges, this study proposes NucleiViT, a Vision Transformer(ViT)-based model designed for efficient and accurate cell detection and segmentation. The core innovation of the proposed model lies in the integration of a sliding-window dilated attention mechanism, which enables feature interaction among different task branches within a shared module, thereby maximizing the advantages of multi-task learning. Additionally, a joint boundary-aware loss function is introduced to enhance sensitivity to cell boundary delineation. Experimental results demonstrate that NucleiViT outperforms existing state-of-the-art methods in terms of segmentation accuracy and computational efficiency, exhibiting superior robustness in densely packed and overlapping cellular regions. This approach provides an efficient and reliable solution for clinical pathological image analysis.

Key words: cell segmentation; multi-task joint optimization; boundary-aware loss; neural network

1 引言

细胞核分割是从组织病理图像中精准分离并识别单个细胞,为后续的细胞计数、形态分析及类型分类提供支撑。在临床实践中,病理学家通过观察细胞核的大小、形态、染色强度及分布模式判断病变程度,例如肿瘤细胞的核异型性是乳腺癌、结直肠癌等恶性肿瘤诊断的关键指标。传统分割方法受病理学家经验限制,分割效率和一致性差。随着人工智能的发展,自动化分割技术为细胞分割提供了新途径,能够很好地解决分割效率和一致性问题。

早期细胞分割方法主要依赖传统图像处理技术,如基于阈值的分割^[1]、分水岭算法^[2]及形态学滤波法^[3],需通过手工设计特征(如灰度值、纹理、边缘梯度)实现分割,在面对细胞密集重叠、背景噪声干扰或染色不均时,鲁棒性显著下降。如,分水岭算法虽能处理简单细胞簇的分割,但易因局部灰度极值导致过分割。

随着深度学习技术的兴起,细胞分割也进入了新的阶段,以 U-Net^[4]及其变体(例如 SwinUNETR^[5])为代表的卷积神经网络(CNN)成为主流框架。U-Net 能够很好地保留图像细节信息,通过编码-解码结构在分割过程中展现出优异性能^[4];HoVer-Net 提出基于“细胞核像素到质心的水平/垂直距离(HoVer 距离)”的深度学习方法,能够有效解决细胞核重叠问题^[6];MaskR-CNN 则在 FasterR-CNN 基础上新增掩码预测分

支,通过“检测-分割”两阶段流程,实现细胞核实例分割与分类^[7]。然而,CNN 依赖局部感受野提取特征,使其难以学习长距离空间依赖去捕捉图像全局上下文信息,在高分辨率病理图像(如 1 000 像素×1 000 像素)中易忽略细胞间的空间关联^[8-9]。此外,现有方法多采用单任务学习模式,分割、分类等任务独立优化,未能充分利用任务间的互补信息,导致在细胞边界模糊或形态复杂场景中精度不足。

视觉变换器(Vision Transformer, ViT)是一种基于 Transformer 架构的图像处理模型,其将输入图像分成若干个图块,然后将每个图块的像素展平成向量,并添加位置嵌入(Position Embeddings)来表示每个图块在图像中的位置信息^[10]。ViT 将图像划分为补丁序列,通过捕捉全局依赖关系来准确识别目标,在大规模数据预训练后,其在高分辨率图像处理中展现出优于传统 CNN 的全局信息整合能力(尤其在语义关联建模场景中)^[11]。基于 ViT 的医学图像分割模型(如 UNETR^[12]、SwinUNETR)已在器官分割、肿瘤区域检测等任务中验证了有效性,但在细胞级精细分割中仍存在挑战:(1)ViT 原始架构计算复杂度高(自注意力计算量随序列长度的平方增长)^[11],难以直接应用于大尺寸病理图像;(2)现有 ViT 分割模型多聚焦单一任务,未充分考虑细胞分割中“定位-分割-分类”的多任务协同需求;(3)对细胞边界的建模不足,在密集重叠场景中易出现边界混淆。

针对上述挑战,本文提出 NucleiViT 模型,

旨在提供更加高效和精准的细胞检测及分割解决方案。NucleiViT 在 ViT 中引入了一种基于共享滑动窗口膨胀注意力机制的信息交互与融合框架,应用于细胞核分割(Segmentation, Seg)、水平垂直距离(Distance, Dist)和细胞类型分类(Classification, Cls),实现多任务学习,通过多任务的协调优化,提升模型分割性能。本模型还提出了一种联合边界和区域损失函数,该损失函数不仅强调分割的整体性能,还特别关注细胞的边界精度,能够使网络对细胞的边界进行更加细致的学习,同时在确保关注边界的同时,不会忽略内部区域的分割精度,从而实现更精准的细胞分割。

2 相关工作

细胞分割技术经历了传统方法与深度学习方法的演变过程。传统方法中,基于形态学、分水岭变换及边缘检测的方法是主流方向。Malpica 等人提出基于核形态与分水岭算法的分割方案,通过标记控制分水岭变换解决传统分水岭算法对聚集细胞的过分割问题。这种方法精准解决了聚集细胞分割的难题,对规则形态的细胞核表现出较强的适配性,但由于其标记参数(如 h -穹顶的 h 值、距离变换阈值)需人工适配样本类型,导致泛化性受限(如无法直接迁移至不规则形态肿瘤细胞核)^[13]。传统方法的核心局限在于手工特征的泛化性差,难以适应不同组织类型或染色条件的病理图像。

深度学习方法中,Ronneberger 等人设计的 U-Net 奠定了生物医学图像分割基础^[4]。Ali 等人聚焦于细胞核边界的精准定位,提出“多步骤图像增强(如强度归一化、随机翻转/缩放)+形态学操作(如残缺块卷积、反卷积)”流程,通过强化局部边缘梯度提升边界完整性,实现细胞核边界提取。该方法实现了高精度的边界定位,且无需依赖于细胞核形态,适配了部分不规则的细胞,但其在低对比度图像中边缘定位误差大,易出现边界断裂的情况,同时对于重叠细胞的边界区分能力较弱^[12]。在此基础上,研究者提出一系列 U-Net 变体:为解决原始 U-Net 对重叠细胞(如密集肿瘤细胞、重叠白细胞)分割效果差的问题,HoVer-Net^[6]新增水平-垂直距离图分支,通过预测每个像素到细胞核质心的距离,实现重叠细胞

核的分离,同时支持细胞核分割与分类,在 H&E 染色、荧光染色等多组织病理图像中均表现优异;Micro-Net^[14]简化 U-Net 结构,减少参数数量,在保证精度的同时提升推理速度;为解决原始 U-Net 及多数变体对不同组织类型(如肝、肺、乳腺组织)的细胞分割泛化性差的问题,TSFD-Net^[15]引入特征金字塔与组织分类分支,增强对不同组织类型的适应性。然而,这些基于 CNN 的模型仍受限于局部感受野,无法捕捉细胞密集区域的全局关联,在细胞密集区域易因全局信息缺失而容易导致分割错误^[8-9]。

近年来,ViT 架构为细胞分割提供了新范式。CellViT 首次将 ViT 用于多类别细胞核分割,通过 ViT 编码器提取全局特征,结合 U-Net 解码器恢复细节,在 PanNuke 数据集上实现了 0.55 的 PQ 指标,首次验证了 ViT 在多类别细胞核分割中的有效性,为后续基于 ViT 的分割模型提供“全局编码器+局部解码器”的混合架构范式^[16]。UNETR 优化了 ViT 与解码器的连接方式,将标准 ViT 与 U-Net 解码器结合,通过多尺度跳跃连接提升分割精度,ViT 在全局信息捕捉方面具有一定的优势^[12]。当前 ViT 框架实现了全局特征捕捉能力的显著提升,增强了多模态与多组织的泛化性,为复杂细胞分割任务奠定了基础。但现有 ViT 分割模型仍存在两点不足:(1)计算复杂度高。ViT 的自注意力计算量随图像尺寸呈平方级增长(复杂度为 $O(N^2)$, N 为图像块(patch)数量),例如 ViT-B/16 在 $1\,024 \times 1\,024$ 图像上的自注意力计算量是同尺寸下 U-Net(复杂度 $O(N)$)的 3 倍;(2)多任务协同不足。当前 ViT 分割模型多聚焦“单一切割任务”,未充分利用分割与分类、计数等任务间的互补信息。

3 NucleiViT 模型架构

精确分割细胞核是数字病理诊断的基础。针对细胞重叠、边界模糊及背景噪声带来的分割误差,本文提出 NucleiViT:一种轻量级 ViT,通过“共享滑动窗口膨胀注意力”同时完成细胞核分割、核-质心距离回归与细胞类型分类 3 项任务。设计“联合边界-区域损失”强化边缘定位,利用任务间互补特征提升重叠区鲁棒性,无需额外降低 ViT 计算复杂度即可保持实时推理(16 FPS)。

在 PanNuke、CoNSEP 与 MoNuSeg 上的全景质量 (Panoptic Quality, PQ) 分别达 0.554、0.612 与 0.765, 优于 CellViT 等最新方法, 验证了多任务协同优势。

本节将详细介绍 NucleiViT 模型的架构设计, 重点讨论 ViT 编码器在细胞图像特征提取中的作用、后处理步骤的应用、以及解码器与跳跃连接在恢复空间分辨率和细节恢复方面的优势。

3.1 模型架构

图 1 展示了 NucleiViT 的整体架构。该架构采用 ViT 作为图像编码器的主体结构, 并通过跳跃连接将其与上采样的解码器网络相连接。这样的设计能够充分利用视觉变换器进行实例分割, 同时保留细粒度的信息。预训练模型考虑 ViT₂₅₆ 和 ViT_{SAM} 的影响, 网络选择了一个不修改原始结构的 ViT 网络架构。

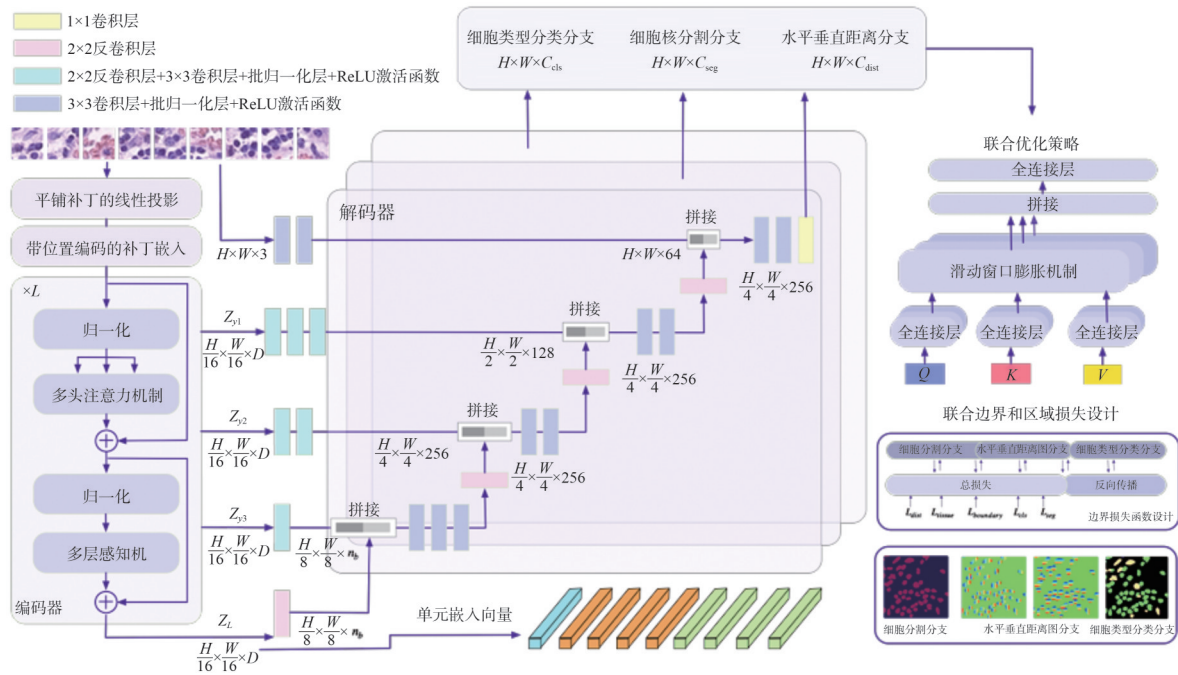


图 1 NucleiViT 整体架构图

Fig. 1 Overall architecture diagram of NucleiViT

3.1.1 ViT 输入层的处理

与自然语言处理(NLP)中的方法^[17]类似, 视觉变换器将输入图像划分为一系列一维标记序列^[8-17]。对于给定的输入图像 $x_p \in \mathbb{R}^{H \times W \times C}$ (其中 H 是图像高度, W 为宽度, C 为通道数), 首先将其划分为若干大小为 $P \times P$ 的图像块, 生成一个扁平的标记序列 $x_p \in \mathbb{R}^{N \times P^2 \times C}$, 其中每个标记对应一个图像块。通过公式 $N = \frac{HW}{P^2}$ 可计算出标记的数量 N , 这是输入序列的有效长度^[12]。接下来使用一个线性投影层 $E \in \mathbb{R}^{N \times D}$, 将每个平展后的标记 x_p 映射到 D 维的潜在空间。

与 UNETR 网络架构不同, 本章所提出的方法还引入了一个可学习的类别标记 X_{class} ^[11], 它可用于分类任务并被附加到标记序列中。这意味

着需要将输入数据的类别信息融入到模型中来帮助分类任务。类别标记是一个科学系的嵌入, 将在模型训练时被优化。

为了使模型能够感知二维图像中的位置空间信息, 引入了一个可学习的一维位置嵌入 $E_{\text{pos}} \in \mathbb{R}^{(N+1) \times C}$, 将其加入到投影后的标记中以保留上下文信息^[12]。最终 Transformer 编码器的最终输入序列 z_0 为:

$$z_0 = [x_{\text{class}}, x_p^1 E, x_p^2 E, \dots, x_p^N E] + E_{\text{pos}}. \quad (1)$$

3.1.2 ViT 编码器与解码器

Transformer 块由交替的多头自注意力层 (MHA) 和多层感知器 (MLP) 全连接网络构成, 前者用于捕捉输入序列中不同部分之间的关系, 后者处理编码器中的特征, 多个 Transformer 块

堆叠组成 Transformer 编码器。经编码器处理后,每一层提取的特征表示如下:

$$\begin{cases} z'_i = \text{MHA}(\text{Norm}(z_{i-1})) + z_{i-1}, i = 1 \cdots L \\ z_i = \text{MLP}(\text{Norm}(z_{i-1})) + z_{i-1}, i = 1 \cdots L \end{cases} \quad (2)$$

将输入 x 通过两个大小为 3×3 的卷积层,使用批量归一化与 ReLU 函数作为激活函数,实现第一个跳跃连接。对于剩下的 4 个跳跃连接,输入为从中间层和瓶颈层提取的潜在的标记(tokens) $z_j, j \in \left\{ \frac{L}{4}, \frac{2L}{4}, \frac{3L}{4}, L \right\}$, 这些标记经提取后重塑为二维张量 $Z_j \in \mathbb{R}^{H' \times W' \times D}$ 。该操作只在 Transformer 块的数量 L 是 4 的倍数时有效。每个特征图 z_j 通过反卷积层进行处理。反卷积层将分辨率在两个方向上增加 2 倍,并通过卷积调整潜在的维度。每个阶段转换后的特征图会再与对应的跳跃连接融合,然后依次通过解码器进行处理。

3.1.3 多任务分支

通过解码器输出 3 个分支:分割分支(Seg)、距离图分支(Dist)、分类分支(Cls),3 个分支共享

解码器底层结构,并使用相同的跳跃连接及其转换,仅在顶层设置独立输出层。在多任务学习中,细胞图像通过多个任务分支进行处理,各个分支的特征图经过拼接,形成一个包含所有任务信息的大特征矩阵,而后通过共享滑动膨胀注意力机制进行特征融合与交互。

3.2 任务间联合优化在分割任务中的使用

如图 2 所示,本研究提出了一种基于共享滑动窗口膨胀注意力机制的信息交互与融合框架,通过细胞核分割(Segmentation, Seg)、水平垂直距离(Distance, Dist)和细胞类型分类(Classification, Cls)3 个分支的并行处理,实现多任务学习。 C_{dist} 预测每个细胞核像素到其质心的水平和垂直距离,并将其归一化到 $-1 \sim 1$ 之间; C_{seg} 预测二值化的细胞核图; C_{cls} 预测细胞核类型。每个任务分支在输入图像的基础上提取不同的特征图用于不同的学习任务。当输入图像的尺寸为 $H \times W \times C_{\text{in}}$ (C_{in} 为输入图像的通道数),每个任务分支的输出特征图维度为 $H' \times W' \times C_{\text{out}}$ (C_{out} 为每个任务所生成的特征图的通道数)。

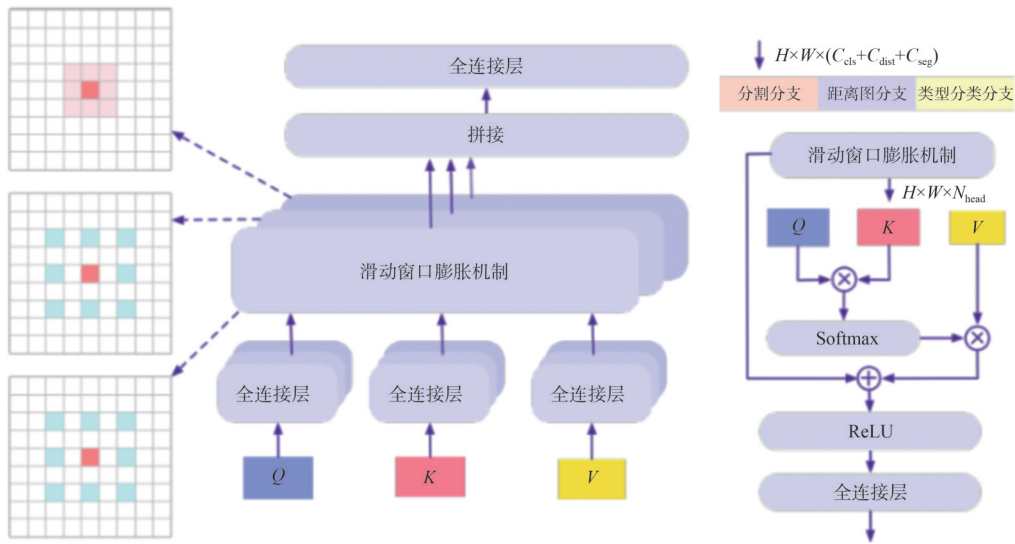


图2 联合优化模块架构

Fig. 2 Structure of the joint optimization module

图2的右侧详细展示了滑动窗口膨胀操作在空间层面对特征的处理,左侧则展示了这一操作如何与后续的计算步骤结合作为模型结构的一部分。Seg分支表示细胞核分割,该分支的任务是分割细胞核,输入的图像特征通过卷积层提

取后生成了一个具有细胞核边界信息的特征图 $H \times W \times C_{\text{seg}}$, C_{seg} 是细胞核分割特征的通道数。Dist分支为水平-垂直距离图分支,该分支用于计算每个细胞核的水平和垂直距离图,输出特征图形状为 $H \times W \times C_{\text{dist}}$, 其中 C_{dist} 是距离图的特征

通道数。Cls 分支为细胞类型分类分支,该分支主要输出每个像素对应的细胞类型,输出特征图的形状为 $H \times W \times C_{\text{cls}}$,其中 C_{cls} 是细胞类型分类的特征通道数。为了让这些任务分支之间的信息互相传递和共享,本实验将每个任务分支的特征进行拼接,如果每个任务分支的输出特征图为 $[H \times W \times C_{\text{seg}}]$ 、 $[H \times W \times C_{\text{dist}}]$ 和 $[H \times W \times C_{\text{cls}}]$,那么把所有任务的特征图拼成一个大的特征矩阵。拼接后的特征矩阵形状为:

$$\mathbf{X} = [\mathbf{X}_{\text{seg}} \oplus \mathbf{X}_{\text{dist}} \oplus \mathbf{X}_{\text{cls}}], \mathbf{X} \in \mathbb{R}^{H' \times W' \times (C_{\text{seg}} + C_{\text{dist}} + C_{\text{cls}})}. \quad (3)$$

该矩阵包含了所有的任务信息。接下来用滑动窗口膨胀注意力机制对其进行处理,实现任务间的信息交互和融合。

为确保不同任务的特征可以有效融合,首先需要每个任务分支的特征进行统一处理。任务分支不同,相应通道数也不同,可以通过卷积操作与线性变换对特征进行映射处理,使得维度相同。在滑动窗口膨胀注意力机制中,拼接后的特征矩阵 $\mathbf{X} \in \mathbb{R}^{H' \times W' \times C_{\text{total}}}$ 会被切分为若干局部窗口,通过在特征图上应用滑动窗口来限制计算范围,在每个窗口内计算注意力以减少计算量。膨胀系数设为 $\delta=1, \delta=2$,选择较小的窗口大小 3×3 与 5×5 ,以确保仅计算局部区域的注意力,这会使得窗口逐步扩大,从而适应不同尺度的细胞图像信息。

对于分割任务,尤其是在细胞密集的图像中,较小的窗口有助于细节的捕捉,同时可以避免引入过多的背景信息。在细胞边缘和结构较小的区域, 3×3 窗口能够保证高分辨率处理。设置较小的膨胀系数($\delta=1$)提供了较小的感受野,使得模型能够精准捕捉细节,处理细胞边缘和小区域等局部信息。

距离任务需要考虑空间关系和局部的上下文。 5×5 窗口能够适当捕捉这些信息,同时不会因窗口过大而引入过多不相关的背景信息。合适的膨胀系数($\delta=2$)保证了适度的感受野扩展,有助于模型捕捉细胞之间的相对距离关系,同时避免过大的窗口带来的计算冗余。

分类任务需要更多全局的特征信息,但细胞密集的图像通常包含较多的局部信息。 5×5 窗口大小能够适当地覆盖较大的上下文区域,同时

避免计算过于冗杂。为了控制计算量,将膨胀系数设置为 2,这能够平衡全局信息的提取与计算效率,尤其在细胞密集的情况下,可在确保不损失关键信息的同时避免过度计算,具体的过程如下:

首先对拼接后的特征矩阵进行卷积操作,得到每个窗口的 Query、Key 和 Value 矩阵:

$$\mathbf{Q}_i = \mathbf{W}_Q^i \mathbf{X}_{\text{local}}, \mathbf{K}_i = \mathbf{W}_K^i \mathbf{X}_{\text{local}}, \mathbf{V}_i = \mathbf{W}_V^i \mathbf{X}_{\text{local}}. \quad (4)$$

计算每个 Query 与所有 Key 之间的点积,并通过缩放点积注意力获得注意力权重 α :

$$\alpha_{ij} = \frac{\exp\left(\frac{\mathbf{Q}_i \mathbf{K}_j^T}{\sqrt{d_i}}\right)}{\sum_j \exp\left(\frac{\mathbf{Q}_i \mathbf{K}_j^T}{\sqrt{d_j}}\right)}. \quad (5)$$

然后,通过计算加权和得到每个局部窗口内特征的输出:

$$\text{Attention Output} = \sum_j \alpha_{ij} \mathbf{V}_{ij}. \quad (6)$$

最后,将每个局部窗口的输出拼接起来,得到全局特征表示,再通过一个线性层映射到最终的输出空间:

$$\mathbf{O} = \mathbf{W}_o [\text{Window}_1 \oplus \text{Window}_2 \oplus \dots \oplus \text{Window}_n]. \quad (7)$$

虽然在拼接特征矩阵后,对所有任务的特征进行了融合和注意力计算,但在信息传递和融合过程中,只是使用了共享的滑动窗口膨胀注意力机制来增强任务之间的协同效应,每个任务的输出并没有合并成单一的结果,在矩阵拼接前后,它们依然保持每个分支独立的输出。

在共享注意力模块输出处,分割、距离、分类 3 支特征图的互信息由 0.31 升至 0.57,余弦相似度由 0.62 升至 0.79,对应 PQ 指标提升 0.008,表明联合优化后特征分布趋同且边缘响应增强。

3 项任务植根于同一细胞核形态空间,天然互补:分割掩膜给出目标区域,为距离图提供有效计算范围;距离图通过质心约束将重叠核强制分离,其几何信息反向修正掩膜边界;分类分支则利用细胞类别先验,引导网络在易混淆区域关注判别性纹理,从而细化边缘。共享编码器使底层边缘与纹理特征同时受到 3 个损失监督,等价于增加了标注样本量,减轻过拟合;共享解码器则让不同任务的梯度在同一套参数上累积,形成正则化效果,提升泛化能力。

然而任务目标并不总是一致:分割任务关注像素级二值决策,距离回归对目标尺度与位置敏感;分类任务强调语义抽象。若简单加权,分类梯度易在边界处引入过度平滑,距离损失则可能因尺度差异放大而主导训练。为此,在总损失中采用动态加权策略:训练初期赋予分割损失最大权重,确保区域提取可靠;随着轮次(epoch)增加,按指数衰减降低距离与分类权重,避免后者干扰边缘细节;同时保留边界感知损失以独立优化边界像素,使各任务在共享表征的前提下各司其职。通过共享特征提取与动态权重解耦,模型既利用任务互补性提升精度,又缓解目标冲突带来的负面影响。

3.3 联合边界和区域损失设计

传统的分割损失函数(如交叉熵损失、Dice损失等)更侧重于图像的整体分割精度,忽略了对边界区域的重点关注,而仅依赖边界损失则可能会忽略整个区域尤其是细胞内部区域的精准

度。因此本研究提出了一种联合边界和区域损失函数,该损失函数不仅强调分割的整体性能,还特别关注细胞的边界精度。该损失函数将边界损失与区域损失等进行联合优化,平衡了边界和区域之间的权重,从而提高整体预测质量。

图3展示了细胞分割与分类任务中的联合边界和区域损失方法。图3上图展示了细胞分割、水平垂直距离分割以及细胞分类的过程,其中细胞分割将细胞区域从图像中提取出来,水平垂直距离通过计算细胞间的距离来优化分割效果,细胞类型分类则对分割后的细胞进行进一步的分类。图3下图展示了包含多个损失项的总损失函数,包括细胞分割损失、距离损失、分类损失、组织损失和边界损失,这些损失项共同作用,优化细胞分割和分类的精度。图3右侧展示了一个实际应用示例,其中通过不同颜色标出细胞的分割、分类及边界,体现了该方法在处理组织图像时的效果和精度。

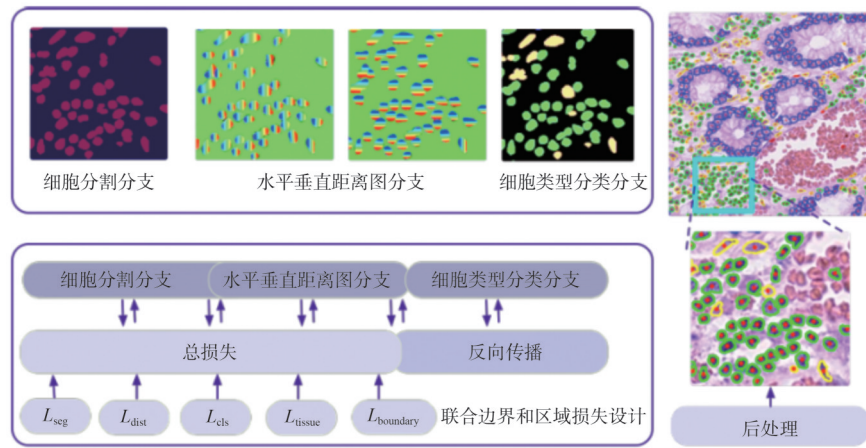


图3 联合边界和区域损失

Fig. 3 Joint boundary and region loss

总损失函数如下:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \mathcal{L}_{\text{dist}} + \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{tissue}} + \mathcal{L}_{\text{boundary}}. \quad (8)$$

3.3.1 分割损失

$\mathcal{L}_{\text{seg}}$ 是优化后的细胞核分割损失,这个损失包含了Dice损失和FocalTversky损失,如式(9)所示:

$$\mathcal{L}_{\text{seg}} = \lambda_{\text{segFT}} \cdot \mathcal{L}_{\text{FT}} + \lambda_{\text{segDICE}} \cdot \mathcal{L}_{\text{DICE}}, \quad (9)$$

其中, λ_{segFT} 和 λ_{segDICE} 是调节参数,分别控制Focal-Tversky损失和Dice损失的相对重要性。

$\mathcal{L}_{\text{FT}}$ 损失旨在通过计算预测与实际目标之

间的相似度来优化分割任务,解决类别不平衡问题:

$$\mathcal{L}_{\text{FT}} = \sum_{c=1}^C \left(1 - \frac{\sum_{i=1}^{N_{\text{px}}} y_{ic} \hat{y}_{ic} + \epsilon}{\sum_{i=1}^{N_{\text{px}}} y_{ic} + \alpha_{\text{FT}} \sum_{i=1}^{N_{\text{px}}} y_{ic} \hat{y}_{ic} + \beta_{\text{FT}} \sum_{i=1}^{N_{\text{px}}} y_{ic} \hat{y}_{ic}} \right)^{\frac{1}{\gamma_{\text{FT}}}}, \quad (10)$$

公式(10)的核心思想是通过 α_{FT} 和 β_{FT} 参数进行调整,强化错误类别的惩罚,它通过加权调整来

减小类别不平衡对模型训练的影响。 y_{ic} 和 $\hat{y}_{ic}$ 分别是样本 i 在类别 c 下的真实标签和预测标签。

Dice 损失常用于衡量预测掩码与真实掩码的重叠度。它的值在 0 和 1 之间,值越大,代表预测越准确。该损失主要用于计算分割模型的准确性,如式(11)所示:

$$\mathcal{L}_{\text{DICE}} = 1 - \frac{2 \times \sum_{i=1}^{N_{\text{px}}} y_{ic} \hat{y}_{ic} + \epsilon}{\sum_{i=1}^{N_{\text{px}}} y_{ic} + \sum_{i=1}^{N_{\text{px}}} \hat{y}_{ic} + \epsilon}, \quad (11)$$

其中, ϵ 是一个小的常数,用于避免除零错误。该损失通过计算真实标签和预测标签的交集与并集来衡量它们的相似性。

3.3.2 距离图损失

$\mathcal{L}_{\text{dist}}$ 用于预测细胞核边界的水平和垂直距离图,包含均方误差损失(MSE)和梯度损失(GradientLoss)。 λ_{distMSE} 和 $\lambda_{\text{distGrad}}$ 是这两部分损失的权重系数,调节这两部分对总损失函数的贡献:

$$\mathcal{L}_{\text{dist}} = \lambda_{\text{distMSE}} \cdot \mathcal{L}_{\text{MSE}} + \lambda_{\text{distGrad}} \cdot \mathcal{L}_{\text{grad}}, \quad (12)$$

其中, $\mathcal{L}_{\text{MSE}}$ 是所有样本的预测值与真实值之间差异的平方和的平均值。它是用于回归任务中常见的损失函数,衡量了预测值与实际值之间的差异,差异越大,损失越大:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_i (\hat{d}_i - d_i)^2. \quad (13)$$

$\mathcal{L}_{\text{grad}}$ 是一个衡量图像或信号的梯度(变化率)差异的损失函数,通常用于图像处理任务,以确保预测结果在边缘或细节部分的变化符合真实的图像梯度:

$$\mathcal{L}_{\text{grad}} = \sum_i \left| \nabla \hat{d}_i - \nabla d_i \right|^2. \quad (14)$$

3.3.3 细胞核分类损失

$\mathcal{L}_{\text{cls}}$ 是用于细胞核分类任务的损失,结合了二元交叉熵损失和加权 Dice 损失,用于提升对细胞核类型的精确分类:

$$\mathcal{L}_{\text{cls}} = \lambda_{\text{clsFT}} \cdot \mathcal{L}_{\text{FT}} + \lambda_{\text{clsDICE}} \cdot \mathcal{L}_{\text{DICE}} + \lambda_{\text{clsBCE}} \cdot \mathcal{L}_{\text{BCE}}. \quad (15)$$

二元交叉熵损失 $\mathcal{L}_{\text{BCE}}$:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{n} \sum_{i=1}^{N_{\text{px}}} \sum_{c=1}^C y_{ic} \log \hat{y}_{ic}. \quad (16)$$

3.3.4 组织分类损失

$\mathcal{L}_{\text{tissue}}$ 是组织分类任务的损失,采用标准的交叉熵损失:

$$\begin{cases} \mathcal{L}_{\text{tissue}} = \lambda_{\text{ce}} \cdot \mathcal{L}_{\text{ce}} \\ \mathcal{L}_{\text{ce}} = -\sum_{c_T=1}^{C_T} y_{c_T} \log \hat{y}_{c_T} \end{cases}. \quad (17)$$

3.3.5 边界感知损失

$\mathcal{L}_{\text{boundary}}$ 主要用于增强细胞核分割任务的边界精确度,重点优化在边缘像素的表现。通过最小化这个损失函数,可以让模型在边界处的预测更加准确,从而提高整个模型的精度。

$$\mathcal{L}_{\text{boundary}} = \sum_{i \in \text{boundary}} |y_i - \hat{y}_i|. \quad (18)$$

以 Canny 算子+形态学细化在 GT 掩膜上提取单像素宽细胞边界,得到二值边界图 B_{gt} (边缘为 1,其余为 0)。

模型预测掩膜经相同后处理得到 B_{pred} ,联合边界损失直接计算 B_{pred} 与 B_{gt} 的 BCE 与 Dice,使网络在边缘像素处获得额外梯度。

3.4 训练细节与参数设置

通过多项实验以及不同算法对比分析验证本研究优势。首先,实验考察了不同 ViT 配置下补丁大小与重叠策略对模型性能的影响,特别分析了 512 像素和 1 024 像素补丁在细胞核检测任务中的表现。在 CoNSeP 数据集上,分别使用 512 像素和 1 024 像素的补丁大小,进一步比较了这些设置对分割精度和计算效率的影响。其次,将模型与现有方法进行了对比,比较了本研究提出的方法与多个基准模型(如 Mask-RCNN^[7]、HoVer-Net^[6]、TSFD-Net^[15]、CPP-Net^[18]、CellViT256^[16]、CellViTSAM^[16])的性能,特别关注了补丁大小和重叠策略的变化对模型结果的影响,并使用 MoNuSeg 数据集进行评估。此外,还进行了消融实验,使用 1 024 像素的补丁和 6.25% 与 12.5% 的重叠设置,通过比较添加与删除正则化方法(如联合优化、边界损失)的效果,评估了不同策略对检测精度和分割性能的影响。利用 CoNSeP 数据集进一步分析细胞核的嵌入。在此实验中,使用预训练的 PanNuke 数据集的模型对 CoNSeP 图像进行推理,并提取每个细胞核的标记嵌入。通过均匀流形近似和投影(UMAP)方法,将细胞嵌入向量降维并可视化在二维散点图中。此外,还训练了一个线性分类器,将检测到的细胞分类到 CoNSeP 细胞核类别中,并在测试图像上进行了验证。训练超参数如表 1 所示。

表1 所有训练运行的超参数

Tab. 1 Hyperparameters for all training runs

参数	数值
损失	$\lambda_{\text{segFT}} = 1, \lambda_{\text{segDICE}} = 1, \lambda_{\text{distMSE}} = 1.5,$ $\lambda_{\text{segGrad}} = 5, \lambda_{\text{clsFT}} = 0.5, \lambda_{\text{clsDICE}} = 0.25,$ $\lambda_{\text{dist}} = 0.4, \lambda_{\text{ce}} = 0.1$
采样	$\gamma_s = 0.9$
优化器	AdamW
训练	epochs=150, batch-size=16, lr-scheduling=0.9

4 实验结果与分析

本节对 NucleiViT 进行了系统的实验分析与结果评估,旨在全面验证其显著的鲁棒性与高精度性。首先,介绍了所使用的数据集和评价指标,为后续实验奠定了基础框架。随后,通过不

同 ViT 配置下补丁大小与重叠策略的影响对比,展示了 NucleiViT 与多种先进算法的对比,验证了其在细胞分割任务中的优势。为了进一步深入理解模型的性能,还进行了细胞实例计数准确性分析多任务分支独立评估。为直观说明分割有效性,进行了可视化分析,展示最终细胞分割与分类结果,所有图中细胞均已圈出。

4.1 数据集与评价指标

4.1.1 实验数据集介绍

本研究选用 PanNuke、MoNuSeg、CoNSeP 3大核心数据集,从多组织类型、多任务需求、多评估维度覆盖细胞形态分析,为细胞分割与分类模型提供全面而丰富的样本支撑。此外,为了展示分割效果,还使用了少量高分辨率的癌症基因组图谱(TCGA)进行结果的可视化。图4展示了实验所用部分数据集。

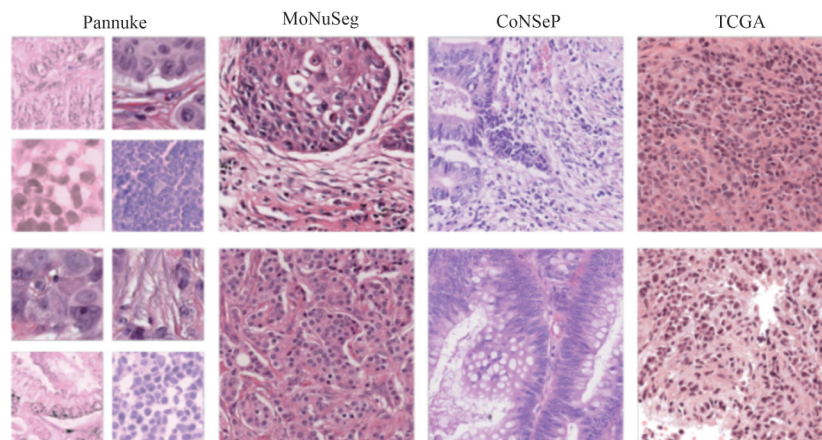


图4 部分数据集展示

Fig. 4 Demonstration of partial datasets

PanNuke 是细胞核实例分割与分类数据集,包含 481 个视野,原始数据来自 312 个多源视野的 20 000+ 张不同放大倍率全切片图像(WSI);按任务划分为 5 000 张训练图像、2 000 张验证图像、2 000 张测试图像,共 189 744 个标注细胞核,覆盖 19 种组织类型(如肾上腺、乳腺、肺、结肠等);按折叠划分为 3 个 fold,每个 fold 均包含 19 类细胞图。该数据集包含完整的“分割+分类”双维度标注,可同时评估模型的分割精度(如 Dice、mIoU)与分类准确率。

本研究实施了多种数据增强方法对 PanNuke 数据集进行增强,增加数据多样性,图像像素值

被归一化到 $[0, 1]$ 区间,确保所有图像在训练过程中具有一致的数值分布。对于较大的图像,先进行 patch 切割,以保持细胞信息的完整性,同时满足模型输入的尺寸要求。该数据集被分成多个子集,其中 80% 用于训练,另外 20% 用于测试。

由 Graham 等人策划的 CoNSeP 数据集,包含 41 张 H&E 染色的结直肠腺癌整张切片图像,分辨率 $0.25 \mu\text{m}/\text{px}$ 。原始数据从 16 个患者的结直肠腺癌全切片图像(WSI)中提取,使用 OmnyxVL120 扫描仪采集,确保临床数据真实性。该数据集涵盖多种组织区域(基质、上皮、肌肉、胶原、脂肪、肿瘤区域)及多类型细胞核,有助

于进行全面的表型分析。

MoNuSeg 数据集常作为医学影像分割任务的补充数据集,弥补多类别数据集在“纯分割性能评估”上的不足。在本研究中仅使用 MoNuSeg 的测试集来评估模型,该测试集包含 14 幅图像,是以 40 倍放大率和 0.25 $\mu\text{m}/\text{px}$ 分辨率采集的高分辨率图像。测试集共标注了超过 7 000 个细胞核,涉及 7 种器官(肾脏、肺、结肠、乳腺、膀胱、前列腺和大脑)以及多种疾病状态(包括良性肿瘤和不同阶段的肿瘤)。由于该数据集不包含细胞核的分类标签,因此不能用于评估分类性能,只能用于分割任务的性能评估。

4.1.2 评价指标

图像分割是将图像划分为多个特定区域的过程,这些区域通常具有独特的性质或对应于感兴趣的目标。从集合论的角度观察,图像分割是将图像划分成互不相交的集合^[19]。

细胞核实例分割评估通常采用 Dice 系数或 Jaccard 指数作为医学图像分割的评估标准。但这两种指标在评估细胞核检测的质量时并不理想。为满足从背景中分离出细胞核,识别独立的细胞核并处理重叠的细胞核,对每个实例进行评估,本实验采用 PQ 指标专门针对实例分割进行评估。PQ 的定义为:

$$\text{PQ} = \frac{|\text{TP}|}{|\text{TP}| + \frac{1}{2}\text{FP} + \frac{1}{2}\text{FN}} \times \frac{\sum_{(y,y') \in \text{TP}} \text{IoU}(y,y')}{|\text{TP}|}, \quad (19)$$

公式(19)结合了检测质量 DQ 和分割质量 SQ 的评价。其中:TP 为真阳性实例,FP 为假阳性实例,FN 为假阴性实例,IoU 为交并比。

为了确保公平比较,使用二元 PQ(bPQ)假设所有核都归类为一个类别(核与背景),它忽略细胞类型,仅区分细胞核与背景,可以评估二值分割性能;同时采用更具挑战性的多类 PQ(mPQ),以评估类别级分割性能。

核分类评估的目的是评估模型在检测任务中的质量,本文采用常见的指标。与核实例分割中使用的 PQ 分数的方法类似,真实(GT)和预测实例被分为真阳性(TP)、假阳性(FP)和假阴性(FN)。评估中使用了精度 P_d 、召回率 R_d 及其调和平均数 $F_{1,d}$,索引 d 指代所有类别 c 的整体二

元核检测指标。因此,二元检测分数可被定义如下:

$$\begin{cases} F_{1,d} = \frac{2\text{TP}_d}{2\text{TP}_d + \text{FP}_d + \text{FN}_d} \\ P_d = \frac{\text{TP}_d}{\text{TP}_d + \text{FP}_d} \\ R_d = \frac{\text{TP}_d}{\text{TP}_d + \text{FN}_d} \end{cases}. \quad (20)$$

TP_d 进一步细分为每个类别 $c(\text{TP}_c)$ 、 $c(\text{FP}_c)$ 和 $c(\text{FN}_c)$,以获得针对细胞类型的特定分数。基于此,针对每个核类别, $F_{1,c}$ -score 分数、精度 P_c 和召回率 R_d 均可被定义:

$$\begin{cases} F_{1,c} = \frac{2(\text{TP}_c + \text{TN}_c)}{2(\text{TP}_c + \text{TN}_c) + 2\text{FP}_c + 2\text{FN}_c + \text{F}_d + \text{FN}_d} \\ P_c = \frac{\text{TP}_c}{\text{TP}_c + \text{TN}_c + 2\text{FP}_c + \text{F}_d} \\ R_c = \frac{\text{TP}_c + \text{TN}_c + 2\text{FN}_c}{\text{TP}_c + \text{TN}_c + 2\text{FN}_c + \text{FN}_d} \end{cases}. \quad (21)$$

为对不同核类型的分类进行优先排序,依据官方 PanNuke 评估指标的建议引入了额外的加权因子。由于无法使用 $\text{IoU}(y,\hat{y}) > 0.5$ 标准进行 GT 实例与预测之间的匹配,采用 Sirinukunwatana 等人^[20]提出的方法来定义匹配 $(y,\hat{y})$,条件为两个质心在分别为 6 px(0.50 $\mu\text{m}/\text{px}$) 和 12 px(0.25 $\mu\text{m}/\text{px}$) 的半径内。

4.2 实验结果对比

本节设置了细胞分割数据集上表格中列出的不同模型在多种配置下的表现实验,评估指标包括 PQ、bPQ、mPQ、 P_d 、 R_d 和 $F_{1,d}$ 。此外,还展示了模型在不同 patch 与重叠策略下的表现以及在模型中加入的 SAM 和 BAL 模块的效果。

表 2 展示了不同模型的性能指标,主要对比了多个深度学习模型在不同 patch 大小下的表现,其中 SAM 和 BAL 模块为本文提出的联合优化和边界感损失模块。通过观察表 2 数据, NucleiViT_{SAM} 模型在 512 和 1 024 补丁大小下相较于 NucleiViT₂₅₆ 模型在多个指标上表现出明显的性能提升,尤其在 PQ 和 bPQ 指标上,表明 SAM 模块的引入有效地改善了模型在细胞实例分割任务中的表现。例如, NucleiViT_{SAM} 在 512 补丁下的 PQ 值达到 0.554 1,相较于 CellViT₂₅₆^[16] 的 0.553 9 有所提升。此外,在 1 024 补丁下,

表 2 NucleiViT_{SAM}与现有方法的对比结果

Tab. 2 Results of NucleiViT_{SAM} compared with existing methods

Models	Patch	PQ	bPQ	mPQ	P_d	R_d	$F_{1,d}$
MaskRCNN	1 024	0.470 8	0.538 9	0.583 7	0.740	0.760	0.745
HoVerNet	1 024	0.464 5	0.448 9	0.478 2	0.760	0.750	0.720
TSFDNet	1 024	0.498 2	0.517 3	0.602 4	0.760	0.750	0.755
CppNet	1 024	0.532 4	0.526 1	0.607 5	0.770	0.750	0.745
CellViT ₂₅₆	1 024	0.553 9	0.543 0	0.590 2	0.765	0.690	0.760
CellViT _{SAM}	1 024	0.552 1	0.510 9	0.601 2	0.775	0.770	0.765
NucleiViT ₂₅₆	512	0.503 4	0.527 6	0.608 4	0.745	0.720	0.760
	1 024	0.541 8	0.535 2	0.612 5	0.750	0.755	0.765
NucleiViT _{SAM}	512	0.554 1	0.532 4	0.609 3	0.780	0.775	0.750
	1 024	0.532 0	0.542 1	0.602 3	0.760	0.755	0.740

注:粗体数字表示最优值

NucleiViT_{SAM}的 P_d 的0.780优于其他模型。这些结果表明,SAM模块的引入提升了模型的泛化能力和细粒度分割精度,为细胞实例分割任务提供了更为强大的支持。

4.3 可视化分析

图5、图6、图7展示了Pannuke数据集的不同

组织类型在影像分割任务中的预测效果与真实标签之间的对比。

每一列代表一种不同的组织类型,包括乳腺、结肠、肺、前列腺、胃和卵巢等19种不同组织。每个组织类型的图像有两行,第一行为真实标签,第二行为预测结果。

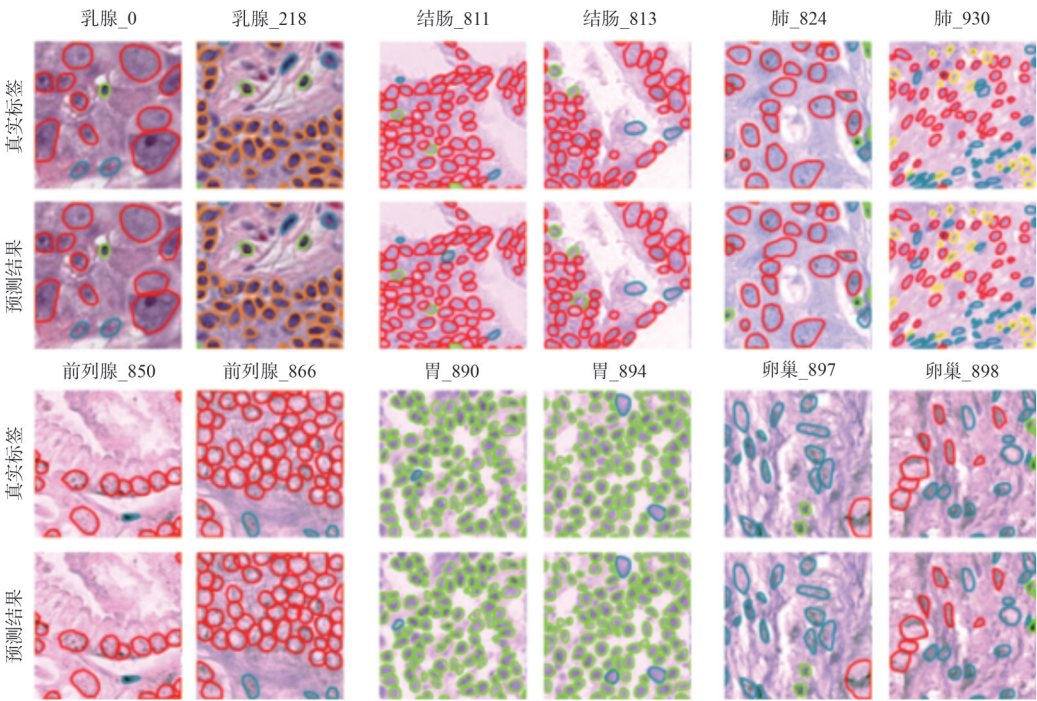


图5 Pannuke数据集的不同组织类型的实验结果1

Fig. 5 Experimental results of different tissues in Pannuke dataset 1

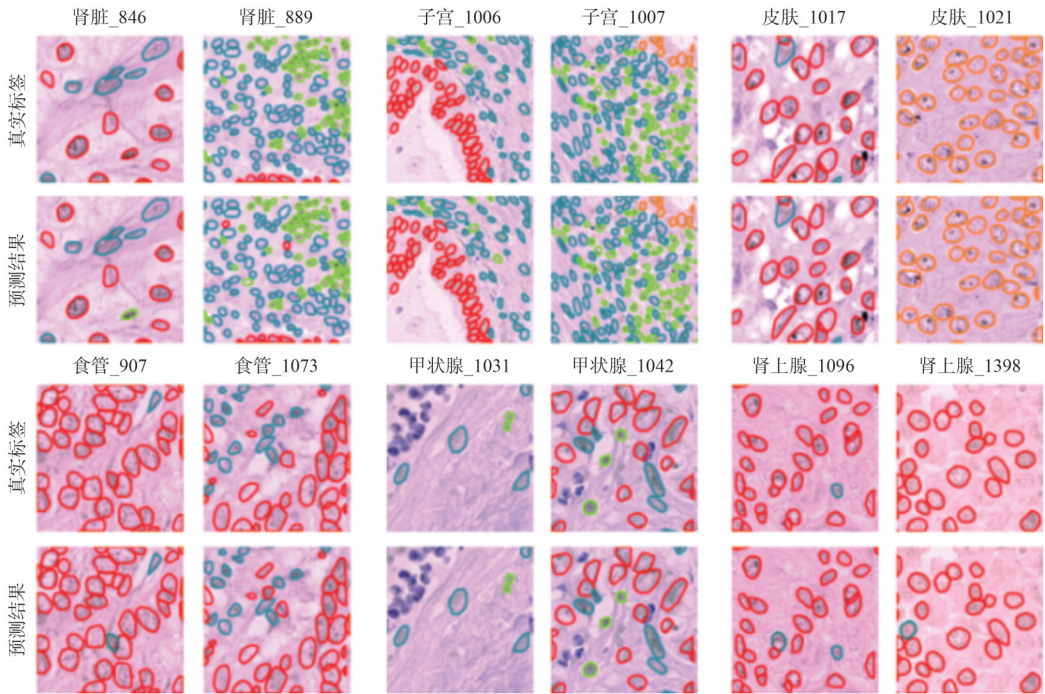


图 6 Pannuke 数据集的不同组织类型的实验结果 2
Fig. 6 Experimental results of different tissues in Pannuke dataset 2

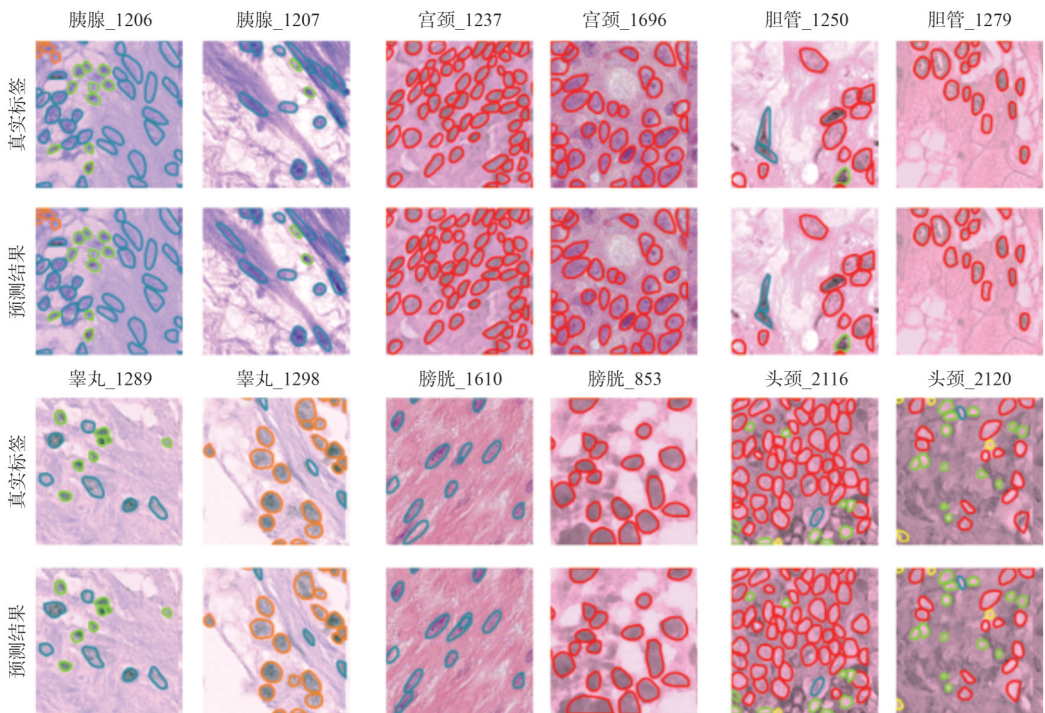


图 7 Pannuke 数据集的不同组织类型的实验结果 3
Fig. 7 Experimental results of different tissues in Pannuke dataset 3

在图像中,红色、蓝色、橙色和绿色的轮廓代表了不同类型的细胞或组织区域。红色代表肿瘤细胞,蓝色代表结缔组织细胞,橙色代表上皮

细胞,绿色代表炎症细胞,而黄色代表死亡细胞。这些轮廓通常围绕不同类型的细胞或组织,以展示模型的预测结果与标注结果之间的对比。

UMAP 是一种常用的降维技术,它可以将高维数据压缩到低维空间,以便更直观地展示数据的结构。它适用于大规模数据集,帮助理解高维数据中的模式和聚类结构。每个细胞通过模型提取出一个高维的“嵌入向量”,该向量包含了细胞的各种特征。每个细胞的嵌入向量都是一个高维的数据点,这些高维的细胞嵌入向量被输入到 UMAP 中进行降维操作。UMAP 将这些高维向量映射到 2D 空间,使得细胞的特征能在二维平面上展示出来。细胞之间的相似性可通过其在 UMAP 降维后二维图中的位置体现。相似的细胞会被投影到离得较近的地方,而不同类型的细胞可能会分散在图的不同区域。根据细胞的不同类型,使用不同的颜色标识每个细胞,这就能帮助观察者直观地看到不同细胞类型在图中的分布,并且查看它们之间的聚类结构。

图 8 展示了使用 NucleiViT_{SAM} 模型对 TCGA 数据集进行的二维 UMAP 嵌入可视化。这意味着模型将高维的细胞数据映射到二维空间,帮助更直观地展示细胞数据的聚类结构。不同的细胞类型通过不同的颜色编码,便于观察细胞类型之间的分布与关联。每个细胞都通过模型提取出一个细胞标记,即每个细胞会有一个对应的嵌入向量,这个向量包含了该细胞的所有特征信息。这使得细胞类型的识别更加精确,并且可以清晰地标注每个细胞的类别。在此基础上在细胞嵌入(从训练数据集中提取的嵌入向量)上训练了一个线性分类器,将检测到的细胞分类到 TCGA 数据集中的不同细胞类型。通过这种方法,可以在测试图像上验证分类器的准确性,并进行模型性能的评估。右侧展示了使用 NucleiViT_{SAM} 模型得到的聚类结果,这些聚类与相应的组织图像一起展示。图像中不同颜色表示细胞的不同类别,即 TCGA 数据集中每种细胞类型的标注。通过对比聚类结果和组织图像,可以直观地看到模型在不同细胞类型分类上的表现。这种基于深度学

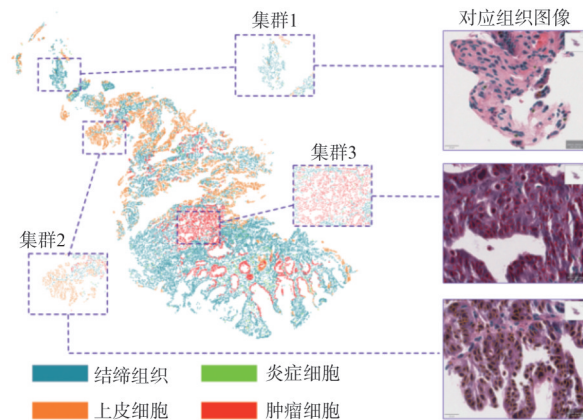


图 8 细胞类型的聚类结果及组织图像

Fig. 8 Clustering results of cell types and tissue images

习的细胞类型识别与数据嵌入技术为医疗人员提供了辅助工具。通过自动化的细胞分类与结构可视化,医生能够更加高效、准确地分析组织样本,为疾病的诊断和治疗提供更可靠的依据。特别是在复杂或大量样本的检查中,该技术能够显著提高工作效率,并减少人为错误,从而提高医疗决策的质量。

5 结 论

本文提出一种基于视觉变换器的 NucleiViT 模型,通过任务间特征共享和联合优化边界区域感知损失来解决密集场景和多组织类型的问题;通过滑动窗口膨胀注意力机制在提升细胞分割精度和效率的同时,有效减少了任务间的特征冗余。在处理细胞密集重叠、边界模糊等复杂场景时,表现出显著的鲁棒性、良好的精度与效率。在 3 个公开医学数据集上的实验结果表明,NucleiViT 在分割精度(PQ 最高 0.585)与计算效率(16 FPS)上均优于现有主流方法。在未来工作中,我们还将探索引入细胞微环境的相关信息并通过图神经网络等方法进一步分析细胞间的关系的可行性,从而实现更为全面的细胞分割与分析。

参考文献:

- [1] MINAEE S, BOYKOV Y, PORIKLI F, *et al.* Image segmentation using deep learning: a survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7): 3523-3542.
- [2] WU Q, MERCHANT F A, CASTLEMAN K R. *Microscope Image Processing* [M]. Boston: Academic Press, 2008.

- [3] MALPICA N, DE SOLÓRZANO C O, VAQUERO J J, *et al.* Applying watershed algorithms to the segmentation of clustered nuclei [J]. *Cytometry*, 1997, 28(4): 289-297.
- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241.
- [5] HATAMIZADEH A, NATH V, TANG Y C, *et al.* Swin UNETR: Swin Transformers for semantic segmentation of brain tumors in MRI images [C]. 7th International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Online: Springer, 2022: 272-284.
- [6] 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.
- [7] HE K M, GKIOXARI G, DOLLÁR P, *et al.* Mask R-CNN [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 2980-2988.
- [8] HU H, ZHANG Z, XIE Z D, *et al.* Local relation networks for image recognition [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 3463-3472.
- [9] 孔韦韦, 李泽江, 何磊磊, 等. 基于 Restormer 与双重注意力机制的非对称医学图像融合模型[J]. 液晶与显示, 2025, 40(8): 1189-1201.
KONG W W, LI Z J, HE L L, *et al.* Asymmetric medical image fusion model based on Restormer and dual attention mechanism [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(8): 1189-1201. (in Chinese)
- [10] RÜDIGER S, MADESTA F, NIELSEN M, *et al.* Multi-scale fully convolutional neural networks for histopathology image segmentation: From nuclear aberrations to the global tissue architecture [J]. *Medical Image Analysis*, 2021, 70(12): 101996.
- [11] HATAMIZADEH A, TANG Y C, NATH V, *et al.* UNETR: Transformers for 3D medical image segmentation [C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE, 2022: 1748-1758.
- [12] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16×16 words: Transformers for image recognition at scale [C]. 9th International Conference on Learning Representations. Online: OpenReview. net, 2021.
- [13] MALPICA N, DE SOLÓRZANO C O, VAQUERO J J, *et al.* Applying watershed algorithms to the segmentation of clustered nuclei [J]. *Cytometry*, 1997, 28(4): 289-297.
- [14] RAZA S E A, CHEUNG L, SHABAN M, *et al.* Micro-Net: a unified model for segmentation of various objects in microscopy images [J]. *Medical Image Analysis*, 2019, 52: 160-173.
- [15] ILYAS T, MANNAN Z I, KHAN A, *et al.* TSFD-Net: tissue specific feature distillation network for nuclei segmentation and classification [J]. *Neural Networks*, 2022, 151: 1-15.
- [16] HÖRST F, REMPE M, HEINE L, *et al.* CellViT: vision Transformers for precise cell segmentation and classification [J]. *Medical Image Analysis*, 2024, 94: 103143.
- [17] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [18] CHEN S C, DING C X, LIU M F, *et al.* CPP-Net: context-aware polygon proposal network for nucleus segmentation [J]. *IEEE Transactions on Image Processing*, 2023, 32: 980-994.
- [19] 李更生, 刘国军. 融合统计范数度量的局部区域图像分割 [J]. 光学精密工程, 2024, 32(16): 2550-2563.
LI G S, LIU G J. Local region image segmentation by fusion of statistical norm metrics [J]. *Optics and Precision Engineering*, 2024, 32(16): 2550-2563. (in Chinese)

- [20] SIRINUKUNWATTANA K, RAZA S E A, TSANG Y W, *et al.* Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images [J]. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1196-1206.

作者简介:



马 龙,男,硕士,工程师,2020 年于南京理工大学获得硕士学位,主要从事图像目标跟踪、目标检测以及图像处理方面的研究。E-mail:mltx_203@163.com



王晓田,男,博士,副教授,2020 年于西北工业大学获得博士学位,主要从事目标检测与跟踪、计算机视觉和遥感图像处理方面的研究。E-mail:18710993786@163.com