



超像素稀疏注意力引导的中药高光谱图像分割方法

蒋雄杰, 张辉, 刘立柱, 尹阿婷, 王耀南

引用本文:

蒋雄杰, 张辉, 刘立柱, 等. 超像素稀疏注意力引导的中药高光谱图像分割方法[J]. *智能系统学报*, 2026, 21(1): 272-283.

JIANG Xiongjie, ZHANG Hui, LIU Lizhu, et al. Superpixel sparse attention-guided hyperspectral image segmentation network for traditional Chinese medicine[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(1): 272-283.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202507028>

您可能感兴趣的其他文章

双向特征融合与注意力机制结合的目标检测

Target detection based on bidirectional feature fusion and an attention mechanism
智能系统学报. 2021, 16(6): 1098-1105 <https://dx.doi.org/10.11992/tis.202012029>

空洞卷积与注意力融合的对抗式图像阴影去除算法

An antagonistic image shadow removal algorithm based on dilated convolution and attention mechanism
智能系统学报. 2021, 16(6): 1081-1089 <https://dx.doi.org/10.11992/tis.202011022>

多视图主动学习的多样性样本选择方法研究

Diversity sample selection method of multiview active learning classification
智能系统学报. 2021, 16(6): 1007-1014 <https://dx.doi.org/10.11992/tis.202007037>

结合模糊特征检测的鲁棒核相关滤波跟踪法

Robust KCF tracking algorithm combined with fuzzy feature detection
智能系统学报. 2021, 16(2): 323-329 <https://dx.doi.org/10.11992/tis.201912010>

基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism
智能系统学报. 2020, 15(5): 956-963 <https://dx.doi.org/10.11992/tis.201903001>

基于Object Proposals并集的显著性检测模型

Saliency detection model based on the union of Object Proposals
智能系统学报. 2018, 13(6): 946-951 <https://dx.doi.org/10.11992/tis.201801009>

DOI: 10.11992/tis.202507028

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20251218.1341.006>

超像素稀疏注意力引导的中药高光谱图像分割方法

蒋雄杰¹, 张辉^{2,3}, 刘立柱³, 尹阿婷³, 王耀南³

(1. 长沙理工大学电气与信息工程学院, 湖南长沙 410004; 2. 湖南大学人工智能与机器人学院, 湖南长沙 410082; 3. 湖南大学机器人视觉感知与控制技术国家工程研究中心, 湖南长沙 410082)

摘要: 在中医药质量检测中, 针对传统红绿蓝三通道图像 (red green blue, RGB) 检测方法因缺乏光谱信息难以实现精准鉴别、高光谱方法在“类间差异小、类内差异大”场景下面临精度与效率不足的问题, 本文构建了线扫描式高光谱成像系统, 建立 5 个具有像素级标注的中药材高光谱数据集, 并提出了一种空间-光谱超像素稀疏注意力引导的中药高光谱图像分割网络。该网络通过三阶段架构实现中药质量无损高精度检测, 基于空间距离和光谱距离将高光谱图像的同质区域分割为超像素块, 增强语义一致性, 提升特征学习效率; 利用双流超像素稀疏注意力模块, 通过构建超像素块间关联矩阵过滤不相关区域, 实现强关联区域间的空间维度全局特征提取与光谱维度的波段显著性建模; 通过多尺度特征融合解码器实现像素级检测。实验结果表明, 在真伪鉴别、产地溯源和炮制品鉴别数据集上, 本文方法平均像素精度 (mean pixel accuracy, MPA) 和平均交互比 (mean intersection over union, MIoU) 分别为 0.968、0.933, 均优于现有方法。本文方法可为中药质量精准检测提供参考。

关键词: 中药质量检测; 高光谱图像; 注意力机制; 语义分割; 特征提取; 跨尺度融合; 超像素; 稀疏注意力

中图分类号: TP391.4; TH744 **文献标志码:** A **文章编号:** 1673-4785(2026)01-0272-12

中文引用格式: 蒋雄杰, 张辉, 刘立柱, 等. 超像素稀疏注意力引导的中药高光谱图像分割方法 [J]. 智能系统学报, 2026, 21(1): 272-283.

英文引用格式: JIANG Xiongjie, ZHANG Hui, LIU Lizhu, et al. Superpixel sparse attention-guided hyperspectral image segmentation network for traditional Chinese medicine[J]. CAAI transactions on intelligent systems, 2026, 21(1): 272-283.

Superpixel sparse attention-guided hyperspectral image segmentation network for traditional Chinese medicine

JIANG Xiongjie¹, ZHANG Hui^{2,3}, LIU Lizhu³, YIN Ating³, WANG Yaonan³

(1. School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410004, China; 2. School of Artificial Intelligence and Robotics, Hunan University, Changsha 410082, China; 3. National Engineering Laboratory for Robot Visual Perception and Control Technology, Hunan University, Changsha 410082, China)

Abstract: In the context of quality control in traditional Chinese medicine, to address the limitations of conventional RGB-based methods—which struggle to achieve accurate identification due to the lack of spectral information—and the challenges faced by existing hyperspectral approaches in scenarios characterized by small inter-class differences and large intra-class variations, a line-scan hyperspectral imaging system was developed and five pixel-level annotated hyperspectral datasets were constructed for TCM materials. Moreover, a spatial-spectral superpixel, sparse attention-guided hyperspectral image segmentation network was proposed for non-destructive, high-precision quality detection via a three-stage architecture. First, the network segments homogeneous regions of the hyperspectral image into superpixel blocks based on the spatial and spectral distances to enhance the semantic consistency and improve feature learning. Thereafter, a dual-stream superpixel sparse attention module constructs a superpixel correlation matrix, filters out irrelevant regions, and enables the extraction of global spatial features and the modeling of spectral band saliency in strongly correlated areas. Finally, a multi-scale feature fusion decoder achieves pixel-level segmentation. As demonstrated experimentally, the proposed method achieves an mean pixel accuracy(MPA) of 0.968 and an mean intersection over union(MIoU) of 0.933 across datasets for authenticity identification, origin tracing, and processing-type classification, outperforming existing approaches. These results indicate that the proposed framework offers a robust and effective solution for precise quality assessment of traditional Chinese medicinal materials.

Keywords: traditional Chinese medicine quality identification; hyperspectral images; attention mechanism; semantic segmentation; feature extraction; cross-scale fusion; superpixel; sparse attention

收稿日期: 2025-07-25. 网络出版日期: 2025-12-19.

基金项目: 国家自然科学基金项目 (62027810); 国家自然科学基金重大研究计划项目 (92148204); 国家自然科学基金重点项目 (62433010).

通信作者: 张辉. E-mail: zhanghui1983@hnu.edu.cn.

中医药作为我国独有的医疗、科技、经济、生态与文化资源, 在临床诊治中发挥了至关重要的作用。中药质量直接影响临床疗效, 严格的质量控制是保证中药疗效和安全性的关键^[1]。由于道

地药材品种变异、假冒伪劣药材和原料重金属污染等问题,导致市场上中药材质量良莠不齐。这些问题不仅影响中药的疗效和安全性,而且严重阻碍中医产业的高质量发展。目前中药质量检测主要采用液相色谱法^[2-3]和质谱法^[4]等理化分析方法,然而这些基于理化分析的方法具有时间成本高、破坏原材料和无法做到全数检测的痛点^[5]。因此研究高效的中药真伪、产地、等级和不同炮制品检测方法具有重要意义。

基于图像识别的检测作为一种高效无损的方法,被广泛应用于食品和病害检测。当前基于图像识别的中药质量检测方法利用深度神经网络分析中药红绿蓝三通道图像(red green blue, RGB)中的纹理、形状和颜色等外在特征来实现不同中药的区分^[6]。然而,在大部分质量检测场景中,中药外在特征高度相似,其差异主要体现在内在的成分和含量上,而RGB图像提供的空间信息无法详细反映内在的差异性信息,导致基于RGB图像的检测方法难以取得令人满意的结果。

高光谱图像(hyperspectral images, HSI)是一种同时包含丰富的空间信息和光谱信息的影像数据。相比于RGB图像,高光谱图像丰富的光谱信息反映了被测物的成分与含量信息,使其能够进行更精细的分析。因此,本文搭建了近红外(near-infrared, NIR)高光谱成像系统,用以构建真伪鉴别、产地溯源和炮制品鉴别中药高光谱数据集。

高光谱图像的分类主要包括基于光谱信息的方法和融合空谱特征的方法。前者依赖数学统计模型^[7]或神经网络^[8-9]结合光谱信息进行分类。但由于数据特征复杂,单一光谱信息难以实现高精度分类。因此研究者们将空间信息引入分类方法,He等^[10]提出多尺度3D深度卷积神经网络(deep convolutional neural network, DCNN),以端到端方式学习二维多尺度空间特征与一维光谱特征;Liao等^[11]设计的空间-光谱Transformer网络,能有效捕获卷积神经网络(convolutional neural network, CNN)难以表征的跨空间与波段的长距离相关性;王瑞婷等^[12]构建混合卷积与三重注意力模型,以增强多尺度纹理和结构特征提取能力并提升精度与效率;涂潮等^[13]提出结合空洞卷积与密集网络的方法,以改善有限样本下的分类表现;Roy等^[14]提出的HybridSN通过混合卷积结构实现空谱联合建模;Hong等^[15]提出的SpectralFormer利用谱域Transformer结构强化光谱维度的依赖建模,为高光谱分类提供了新思路。

在遥感和病理检测领域,由于样品在外在或内在差异较为显著,现有方法能够取得较好的检

测效果,例如基于编码-解码结构的UNet^[16]、基于融合多尺度特征的Hyper-net^[17]及基于光谱依赖性建模的Spectr^[18]在病理高光谱图像分割中均表现优异。然而在中药质量检测场景中,检测样品由于类间差异小、类内差异大,导致存在差异性特征微弱和冗余特征严重的特点,因此现有高光谱检测方法在中药检测场景面临3个问题:1)同质区域影响计算效率与检测精度,高光谱中药图像含有大量低信息熵的同质区域,现有检测方法无差别处理特征区域与同质区域导致增加计算量,并影响特征差异微弱区域的特征提取;2)规则窗口采样导致窗口内部语义不一致,同时在不相关区域间计算远程依赖性,增加无效计算量,影响差异性特征表达,降低检测精度;3)空间-光谱特征信息耦合,影响微弱差异信息提取。空间和光谱特征的信息互相干扰,导致微弱差异性特征丢失,降低模型对细微光谱变化的敏感性,影响分类精度。

针对上述问题,本文提出一种空间-光谱超像素稀疏注意力引导的中药高光谱图像分割网络,将高光谱图像结合深度学习进行中药质量检测,以实现中药质量的精准高效检测。首先,基于空间距离和光谱距离将图像划分为若干个具有相似特征的超像素块,避免了同质区域被多次重复采样而造成额外的计算开销,同时增强语义一致性,提高差异性特征的表达;其次,超像素稀疏注意力通过距离相似性度量超像素块间的相关性,引导模型仅在相似度高的超像素块间计算细粒度令牌的注意,以实现在较少的计算开销下,捕获特征之间的远距离依赖性;此外,引入基于双流超像素稀疏注意力的特征提取模块,分别沿空间和光谱维度进行特征提取,以解决空间-光谱特征提取过程中信息耦合导致的特征相互影响问题;最后,通过双线性插值逐步将特征大小恢复到图像大小,得到高精度的端到端检测分割结果。本文的主要贡献:1)构建了用于医药检测的近红外线扫高光谱成像系统,并建立了5个具有像素级标注的中药质量检测数据集;2)提出了一个双流超像素稀疏注意力模块用于学习空间和光谱信息的全局差异性依赖关系;3)设计了用于中药质量检测的空间-光谱超像素稀疏注意力分割网络。利用超像素稀疏注意力模块,提高在同质区域和不相关区域的特征提取效率。

1 高光谱图像检测相关工作

1.1 高光谱成像系统与检测算法

由于不同材料表面对光的反射、吸收及透射

特性不同,这些差异在特定波长区间会形成具有鉴别性的光谱指纹。高光谱成像技术通过捕捉物体在连续波段的光谱响应特征,结合空间维度信息,实现物质组成与结构特征的精准解析^[19]。基于此技术,学者们搭建了多种不同的高光谱成像系统,并结合相应检测算法,以实现物质的无损识别与精准检测。Delwiche 等^[20]搭建了一台由 NIR 相机、光谱仪、可移动载物台、照明组件和控制软件组成的 NIR 高光谱成像系统对 1 000~1 700 nm 波长范围内的小麦进行成像,并结合线性判别分析 (linear discriminant analysis, LDA) 检测算法,以检测小麦是否感染镰刀菌枯萎病;Zhang 等^[21]搭建了一台成像波段范围为 430~995 nm 的线扫高光谱成像平台,并结合偏最小二乘判别分

析 (partial least squares discrimination analysis, PLS-DA) 检测算法,实现了玉米种子种的黄曲霉素浓度检测;史卓林等^[22]设计并搭建了一套推扫式双相机高光谱成像系统,并结合偏最小二乘回归 (partial least squares regression, PLSR) 检测算法以实现玉米叶片全氮含量测定。

为了对中药质量进行无损精准检测,如图 1 所示,本文搭建了用于医药检测的高光谱成像系统。高光谱成像部分主要由 3 个部件组成:1) 光源,其作用为照亮样品;2) 光谱分离装置 (光栅),它的作用是将入射光的光谱按波段分散到探测器阵列中;3) 图像传感器 (complementary metal oxide semiconductor, CMOS),它将接受到的光能转换为电能,电能大小取决于有效区域光量。

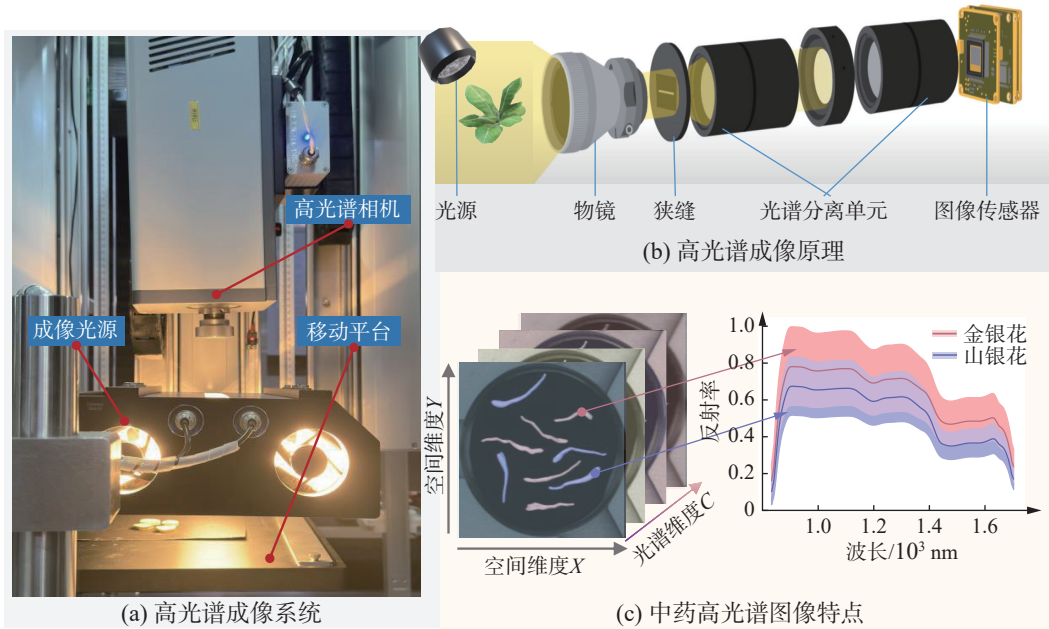


图 1 高光谱成像系统原理及高光谱数据特点

Fig. 1 Principle of hyperspectral imaging system and characteristics of hyperspectral data

光源激发目标药物后,不同药物因其成分差异呈现特异性的光谱反射特性。反射光经物镜聚焦和狭缝整形后形成线状光束,进入分光单元后,光栅将线光束按波长分散为连续窄波段,最后传感器同步记录各像素点信息。通过整合每个像素的空间维度 (X, Y) 信息与光谱维度 (C) 强度信息,形成包含完整目标物空间特征与光谱特征的三维数据立方体。同时,对于中药这类样品,其“类间差异小、类内差异大”的特性在光谱曲线上表现为不同类别样品的光谱曲线整体形态高度相似,区间分布存在显著重叠,导致类别边界模糊;而同一类别内部则因产地、加工方式及成分含量差异,曲线分布范围较广,呈现出显著的类内异质性。

1.2 超像素与注意力机制

超像素分割通过将相邻且特征相似的像素聚

合为紧凑区域,在保留关键边界信息的同时,显著减少图像特征表达的数据量。用少量的超像素表达图像特征,能有效降低后续图像处理的复杂度。

基于数理统计的超像素分割方法因其简单且高可靠性,被广泛应用于图像分割和目标识别等任务。分水岭 (Watershed)^[23]方法基于拓扑理论和数学形态学原理。Mean Shift^[24]通过迭代模态搜索生成规则形状的超像素。Quick-Shift^[25]依据像素间的相似性和空间距离实现分割。Turbopixels^[26]采用几何流优化,确保超像素大小均匀、连通性好且形状紧凑。SLIC (simple linear iterative clustering)^[27]基于颜色与空间距离相似性进行分割。

注意力机制旨在学习像素之间的长距离关系,并关注重点信息。早期的注意力机制主要由卷积神经网络实现,如通道注意力^[28]和空间-通

道混合注意力^[29]。然而由于卷积缺乏长距离特征提取能力, 2017 年, 谷歌提出了自注意力机制并成功应用于自然语言处理^[30]。Vision Transformer (ViT) 将图像划分为若干小块 (patches), 并通过自注意力机制在局部区域间建立全局依赖关系: 每个图像块通过线性变换生成查询 (query)、键 (key) 和值 (value), 再通过点积计算相似度并得到注意力权重, 从而更新特征表示。凭借在图像分类任务中的强大建模能力, ViT 被逐步引入到自然图像分割领域。SETR^[31] 将 Transformer 应用于语义分割, 实现了基于全局上下文的像素级预测; SegFormer^[32] 提出轻量化分层编码器结构, 有效兼顾了全局依赖与多尺度特征建模; 而 Mask2Former^[33] 引入统一的查询-掩码机制, 并通过改进的注意力结构, 在语义分割和实例分割等多任务场景下展现出更优的表现。

由于自注意计算所有图像块之间的相关性, 导致其需要大量的训练资源。针对此问题, 研究人员引入了稀疏注意力, 如 Loacl Attention^[34]、Dilated Attention^[35] 和 Deformable Attention^[36]。这些稀疏注意力通过使用不同的策略来选择进行采样的窗口, 以减少计算量。上述稀疏注意力机制虽能降低全局注意力的计算冗余, 但其规则的采

样窗口导致在面对类间差异小、类内差异大的中药图像检测时, 难以实现令人满意的检测结果。类内存在显著但对分类无效的差异特征, 需将类内区域聚合为少量语义一致的特征区域, 以滤除无用信息, 但现有稀疏注意力的规则采样窗口易引入语义信息不一致的无关区域干扰; 另一方面, 类间光谱差异小 (如不同产地同种药材, 仅在某些成分含量上有区别), 需局部注意力以重点捕捉细微光谱差异特征。

为此, 本文提出了超像素稀疏注意力机制, 通过超像素分割构建具有语义一致性的采样区域, 稀疏注意力仅在差异性信息的超像素之间捕获跨区域长程依赖, 实现低计算复杂度下的精准检测。

2 超像素稀疏注意力引导的分割网络

本文提出用于中药高光谱图像分割的空间-光谱超像素稀疏注意力方法主要包含超像素分割模块、空间-光谱双流注意力模块和特征解码模块 3 个模块。方法整体如图 2(a) 所示。首先对输入特征应用超像素分割模块得到语义信息一致的超像素块, 然后基于超像素块进行空间-光谱并行远程依赖性建模, 最后利用特征解码模块对深层特征进行空间分辨率恢复, 得到逐像素预测结果。

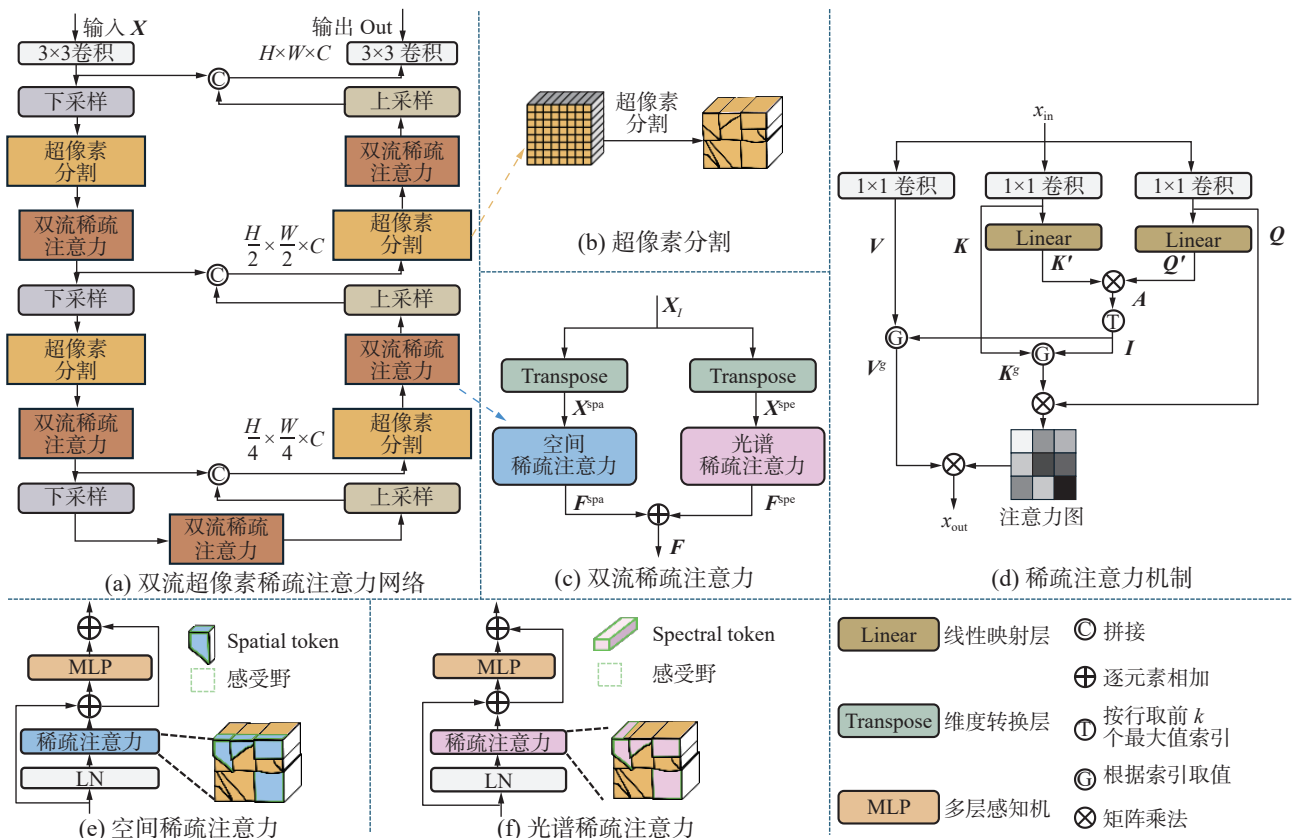


图 2 超像素稀疏注意力引导的高光谱图像分割网络

Fig. 2 Superpixel sparse attention-guided hyperspectral image segmentation network

2.1 超像素分割模块

由于中药图像中同质区域在空间和光谱维度的高度相似性,因此利用空间与光谱相似性约束的可以将同质区域分割为超像素块,如图 2(b)所示。对于输入特征 \mathbf{X} ,首先设定 K 个超像素块(每个超像素块大小 N ,相邻两个中心的距离为 S),并在宽为 W 和高为 H 的图像上随机初始化 K 个超像素中心(根据深层特征分布特性设定多级超像素划分策略,在 3 个超像素分割模块中分别初始化 50、30、20 个超像素中心),并对每个中心点进行 3×3 邻域内像素梯度分析,动态调整中心至邻域内梯度极小值位置,从而避免初始中心落入轮廓边界干扰后续计算。其中超像素块大小 N 、距离 S 由下式定义:

$$N = \frac{HW}{K}$$

$$S = \sqrt{\frac{N}{K}}$$

在像素分配阶段,为每个中心点构建 $2S \times 2S$ 动态搜索区域,采用距离度量 D 综合评估像素点与其中心的空间距离 d_{xy} 光谱距离 d_s ,实现特征相似像素的精确聚类。其中 d_{xy} 、 d_s 和 D 分别由下式定义:

$$d_{xy} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}$$

$$d_s = \sqrt{(a_m - a_n)^2 + (b_m - b_n)^2 + \dots + (i_m - i_n)^2}$$

$$D = \sqrt{\left(\frac{d_{xy}}{N}\right)^2 + \left(\frac{d_s}{10}\right)^2}$$

式中 m 和 n 分别为两个不同位置的像素点。迭代过程中,通过重新计算各超像素簇的质心,并选择与簇内距离最小的特征点作为新中心,最终形成边界清晰、类内语义一致性强的超像素分割结果。该方法通过梯度优化初始化和动态距离权重调节,有效克服了传统算法对中药图像复杂纹理敏感性,显著提升了特征分割的鲁棒性。

2.2 双流超像素稀疏注意力模块

由于中药差异性光谱微弱,传统自注意力机制中固定窗口划分存在局部语义碎片化与计算冗余问题,本文设计了如图 2(c) 所示基于超像素的稀疏注意力机制,旨在提高远程依赖性关系提取能力和提取效率。同时,本文在超像素稀疏注意力机制的基础上提出了空间-光谱双流注意力模块,以解决现有方法中空间与光谱特征相互干扰导致的差异性特征退化问题。

2.2.1 超像素稀疏注意力机制

图 2(d) 所示超像素稀疏注意力机制 (super-pixel spares attention, SSA) 的关键思想是将语义信息一致的超像素块作为 Token,然后筛选出强相

关的区域,并仅在强相关的区域间应用 Token 到 Token 的自注意力。

具体来说,对于输入特征 $\mathbf{X} \in \mathbf{R}^{(H \times W \times C)}$,经过超像素分割模块后得到由 K 个超像素块组成的超像素块集合 $\mathbf{X}_i = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_i, \dots, \mathbf{X}_K\}$,其中每个超像素块 $\mathbf{X}_i \in \mathbf{R}^{M_i \times C_i}$, M_i 和 C_i 分别表示各超像素块包含的像素个数和其光谱维度。由于每个块的 M_i 值不同,为了利用矩阵乘法加速运算,本文将每个超像素块中的 M 统一用“0”填充将维度拓展到 M_i 中的最大值 M_p 。在稀疏注意力操作中输入特征可以表示为 $\mathbf{X}_i \in \mathbf{R}^{K_a \times M_a \times C_a}$ 。然后通过矩阵乘法对输入特征进行映射,得到查询 \mathbf{Q} 、键 \mathbf{K} 和值 \mathbf{V} 。

$$\mathbf{Q} = \mathbf{X}_i \mathbf{W}_q$$

$$\mathbf{K} = \mathbf{X}_i \mathbf{W}_k$$

$$\mathbf{V} = \mathbf{X}_i \mathbf{W}_v$$

由于 \mathbf{Q} 和 \mathbf{K} 包含了输入特征的信息和特征分布关系,因此可以沿注意力维度利用矩阵内积构建关联矩阵来计算每个区域之间的相关性。具体来说,本文首先通过对 \mathbf{Q} 和 \mathbf{K} 进行线性映射得到的映射矩阵 $\mathbf{Q}' \in \mathbf{R}^{K_a \times M_a \times C_a}$ 和 $\mathbf{K}' \in \mathbf{R}^{K_a \times M_a \times C_a}$,然后进行矩阵内积,得到关联矩阵 $\mathbf{A} \in \mathbf{R}^{K \times K}$,计算过程如下:

$$\mathbf{A} = \mathbf{Q}' (\mathbf{K}')^T$$

式中 T 为矩阵转置。

关联矩阵 \mathbf{A} 的每一行表示某个超像素块与其他超像素块间的相关性,其数值表示了超像素块之间的相关程度,数值小的低关联性区域不利于差异性特征的表达。因此可以根据关联矩阵 \mathbf{A} 每行中最大的 k 个数的索引构建索引矩阵 $\mathbf{I} \in \mathbf{R}^{K_a \times k}$,实现不相关区域的过滤。其计算为

$$\mathbf{I} = \text{TopkIndex}(\mathbf{A})$$

式中 TopkIndex 为按行取前 k 个最大数的索引。

最后通过索引矩阵,对高关联区域应用 Token 到 Token 的自注意力操作,进行特征远程依赖关系提取。具体来说,首先通过索引矩阵对 \mathbf{K} 和 \mathbf{V} 中的强相关区域进行提取得到 $\mathbf{K}^g \in \mathbf{R}^{k \times M_a \times C_a}$ 和 $\mathbf{V}^g \in \mathbf{R}^{k \times M_a \times C_a}$,然后对 \mathbf{Q} 、 \mathbf{K}^g 和 \mathbf{V}^g 沿 Token 维度分成 h 份 (h 为多头注意力中的头数),并应用 Token 到 Token 的多头自注意力进行特征提取。其计算过程为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}^g, \mathbf{V}^g) = \text{softmax} \left(\frac{\mathbf{Q} (\mathbf{K}^g)^T}{\sqrt{d_k}} \right) \mathbf{V}^g$$

$$h_i = \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i^g, \mathbf{V}_i^g)$$

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}^g, \mathbf{V}^g) = \text{Concat}(h_0, h_1, \dots, h_i)$$

式中: Attention 为自注意力操作, h_i 为第 i 个注意力头上的注意力, \mathbf{Q}_i 、 \mathbf{K}_i^g 、 \mathbf{V}_i^g 分别为第 i 个注意力头的查询、键和值, Concat 为沿 Token 维度拼接。

2.2.2 双流稀疏注意力模块

进一步地, 本文将以上超像素稀疏注意力机制应用于空间-光谱全局差异性特征提取, 并提出了双流稀疏注意力模块。如图 2(e) 和图 2(f) 所示, 该模块由并行的空间超像素稀疏注意力 (spatial superpixel spares attention, SASSA) 和光谱超像素稀疏注意力 (spectral superpixel spares attention, SESSA) 组成, 其分别沿空间和光谱维度对输入特征进行独立的特征提取, 以消除空间和光谱信息在提取过程中的相互干扰, 增强特征提取能力。

首先, 对输入特征 $X_I \in \mathbf{R}^{K \times M_p \times C}$ 分别沿空间和光谱维度进行维度变换, 生成空间超像素稀疏注意力输入 $X^{\text{spa}} \in \mathbf{R}^{K \times M_p \times C}$ 和光谱超像素注意力输入 $X^{\text{spe}} \in \mathbf{R}^{K \times C \times M_p}$, 以进行空间和光谱信息能够独立处理, 为后续的特征提取提供更加精准的基础。其中 X^{spa} 和 X^{spe} 由下式定义:

$$X^{\text{spa}} = \text{transpose}(X_I)$$

$$X^{\text{spe}} = \text{transpose}(X_I)$$

式中 transpose 为维度变换。

随后, 采用超像素稀疏注意力机制分别对空间维度和光谱维度的特征进行处理。通过这一机制, 可以聚焦于局部区域内的关键特征, 提高特征提取的精度和效率, 同时减少冗余信息影响。在特征提取的过程中, 空间特征 $F^{\text{spa}} \in \mathbf{R}^{K \times M_p \times C}$ 和光谱特征 $F^{\text{spe}} \in \mathbf{R}^{K \times C \times M_p}$ 得到独立建模, 并保持各自的维度和结构特性。 F^{spa} 的计算过程为

$$F'_{\text{spa}} = \text{SASSA}(\text{LN}(X^{\text{spa}}) + X^{\text{spa}})$$

$$F^{\text{spa}} = \text{MLP}(\text{LN}(F'_{\text{spa}})) + F'_{\text{spa}}$$

式中: F'_{spa} 为中间空间特征, LN 为层归一化, MLP 为多层感知机。

F^{spe} 的计算过程如下式所示:

$$F'_{\text{spe}} = \text{SESSA}(\text{LN}(X^{\text{spe}}) + X^{\text{spe}})$$

$$F^{\text{spe}} = \text{MLP}(\text{LN}(F'_{\text{spe}})) + F'_{\text{spe}}$$

式中 F'_{spe} 为中间光谱特征。

最后, 采用动态加权融合策略对空间特征和光谱特征进行有针对性的融合。动态加权策略可以根据不同任务需求动态调整空间和光谱特征的重要性, 使得模型能够在不同场景下更加灵活地调整特征表示, 充分利用空间和光谱信息的互补性, 提升模型的整体性能和表达能力。计算过程为

$$F = \alpha F^{\text{spa}} + (1 - \alpha) \text{transpose}(F^{\text{spe}})$$

式中: $\alpha \in [0, 1]$ 为动态加权因子, $F \in \mathbf{R}^{K \times M_p \times C}$ 为融合后的深层特征。

2.3 特征解码模块

特征解码模块旨在将经过特征提取和处理后

的特征图恢复到与原始输入图像相同的空间维度, 以生成逐像素的精确预测结果。该模块主要由双线性插值和跳连接两部分组成。

首先, 双线性插值被用来将低分辨率的特征图逐步上采样至输入图像的大小, 同时确保解码后的特征图在空间结构上与原始图像一致。为了更好地保留高层次特征信息, 特征解码模块还结合了跳连接机制。跳连接通过将浅层的高分辨率特征与解码过程中的深层特征图进行拼接, 传递低级语义信息, 避免在解码过程中由于信息丢失导致细节丧失。定义如下式所示:

$$X = \text{Concat}(\text{Interp2d}(F), X_{\text{raw}})$$

式中: X_{raw} 为编码过程中的浅层特征, F 为解码过程中需要进行插值上采样的深层特征, Interp2d 为双线性插值, Concat 为矩阵拼接。

3 实验与结果分析

3.1 数据集介绍

3.1.1 近红外医药线扫高光谱成像平台

由于中药不同成分的光谱主要出现在可见光波段之后, 因此, 本研究搭建近红外医药线同时包含空间信息和光谱信息扫高光谱成像平台如图 1(a) 所示。该平台由近红外光谱成像相机、高精度电控平移台、卤素光源阵列以及计算机控制软件组成。近红外相机光谱分辨率为 1.65 nm, 能够覆盖 886~1735 nm 的光谱范围。在数据采集过程中, 样品置于配备消反光涂层的黑色移动平台上进行线性运动扫描, 结合双波段卤素光源的反射照明方案, 以进行高质量采集高光谱数据。成像系统通过推扫式扫描机制在近红外光谱区间实现连续谱段捕获, 每个空间位置共采集 512 个波段。

3.1.2 实验样品准备

本文中建立了 2 个真伪鉴别数据集、2 个产地鉴别数据集和 1 个炮制品鉴别数据集, 每个数据集均包含 200 张高光谱图像数据。部分样品的如图 3 所示。真伪鉴别数据集包括金银花数据集 (金银花 (真)、山银花 (伪))、人参数据集 (人参 (真)、西洋参 (伪)); 产地鉴别数据集为包括枸杞数据集 (宁夏枸杞、青海枸杞)、贝母数据集 (川贝母 (四川)、平贝母 (吉林)); 炮制品鉴别数据集包括半夏数据集 (清半夏和法半夏)。为了保证样品的真实有效性与样品丰富性, 所有样品均来自于农贸市场, 且在数据采集时, 样品不会重复使用, 以缓解同物异谱或者同谱异物现象。本文遵循深度学习检测方法中将训练集、测试集和验证集设定为 6:2:2 的比例设定。

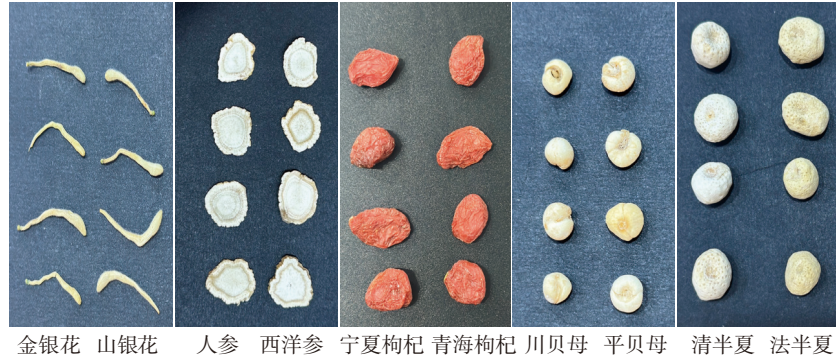


图 3 部分样品 RGB 图像示例

Fig. 3 examples of some samples RGB

3.1.3 数据预处理

在高光谱图像采集过程中,光照波动和电子器件噪声常常导致所获得的图像无法真实反映样品的特征。白板高光谱图像作为标准参考,能够有效反映数据采集过程中光照不均匀性和设备噪声等系统性误差。通过与白板图像的对比校正,可以有效去除光照波动和设备噪声带来的影响,从而恢复样品的真实光谱特征。因此本文采用逐波段白板校正方法对中药高光谱图像进行预处理。校正过程如下式所示:

$$I = \frac{R_{i,j}^n - D_{i,j}^n}{W_{i,j}^n - D_{i,j}^n}$$

式中: I 、 R 、 D 和 W 分别为校正后数据、原始数据、标准黑板数据和标准白板数据, n 为光谱波段, i 和 j 为空间位置。

3.2 实验设置

本文实验参数: Intel(R) Core(TM) i9 13900K, NVIDIA GeForce RTX 4090 24 GB, CUDA 版本为 12.0, Python 版本为 3.9, PyTorch 版本为 2.0。优化器为 SGD, 学习率为 0.01, 训练 Epoch 为 100, 批次大小为 2。

评价指标采用平均像素精度 (mean pixel accuracy, MPA) 和平均交互比 (mean intersection over union, MIoU) 进行模型性能评估。MPA 为正确像素总数与预测结果中像素总数之比, 表示该网络的像素分割精度; MIoU 是预测结果与真实值 (ground truth, GT) 之间的重叠比, 它代表网络的整体分割精度。定义如下式所示:

$$I_{\text{MPA}} = \frac{\sum_{i=0}^K \frac{P_{ij}}{\sum_{j=0}^K P_{ij} + P_{ji} + P_{ii} + P_{jj}}}{K+1} \times 100\%$$

$$I_{\text{MIoU}} = \frac{\sum_{i=0}^K \frac{P_{ij}}{\sum_{j=0}^K P_{ij} + \sum_{j=0}^K P_{ji} - P_{ii}}}{K+1} \times 100\%$$

式中: K 为数据集的类别数, P_{ij} 为像素数类别 i 被预测为类别 j , P_{ji} 为像素数类别 j 被预测为类别 i 。

3.3 实验结果

为验证本文所提出方法在的有效性, 本文选取了多个现有主流的分割检测方法进行对比分析。其中, 自然图像分割方法包括 SegFormer^[31]、SETR^[32] 和 Mask2Former^[33]; 遥感高光谱图像方法包括 HybridSN^[14] 和 SpectralFormer^[15]; 病理高光谱图像分割方法包括 UNet^[16]、Hyper-net^[17] 和 Spectr^[18]。为适配中药高光谱图像的光谱维度, 本文对 SegFormer、SETR 和 Mask2Former 的输入维度进行了修改, 其余训练设置均遵循原始文献的配置。表 1 给出了各方法在不同数据集上的分割性能指标。具体而言, 本文方法在所有数据集上的总体平均像素精度 MPA 达到了 0.968, 总体平均整体精度 MIoU 达到了 0.933, 显著优于对比的 8 种方法。相较于 SegFormer、SETR 和 Mask2Former, 本文方法在总体 MPA 和总体 MIoU 上分别提升了 8.8%、8.3%、21.4% 和 18.8%、18.1%、47.8%。该结果充分证明了本文方法在不同类型样本及复杂特征结构下均表现出良好的适应性与泛化能力。自然图像分割方法主要依赖空间纹理特征, 未能充分利用高光谱图像的光谱信息, 因此在空间差异较小的金银花、人参和枸杞数据集上, MIoU 均低于 0.80。重叠场景中, 由于空间信息丢失或改变, 严重依赖空间特征的方法易产生预测误差。相比之下, Mask2Former 的表现劣于 SegFormer 和 SETR, 主要原因在于其网络结构较深、注意力机制复杂, 对训练数据量依赖较大, 因而在样本有限的中药高光谱数据集上训练困难。与 HybridSN 和 SpectralFormer 相比, 本文方法在总体 MPA 上分别提升 12.0% 和 11.0%, 在总体 MIoU 上分别提升 25.7% 和 23.2%。虽然二者在光谱信息利用方面具有一定优势, 中心区域表现良好, 但因特征区分能力不足, 边缘与重叠区域误差较

高, 整体精度受限。SpectralFormer 通过注意力机制引入光谱维度的远程依赖信息, 性能优于 HybridSN, 但其单一维度建模方式在复杂样本结构下仍存在局限。病理高光谱图像分割方法能够同时建模空间纹理与光谱成分信息, 因此在 MMPa 和 MMIoU 上优于自然图像和遥感高光谱方法。但在金银花与人参数据集上, 因成分相似、形态相近, UNet 表现不及基于 Transformer 的 SegFormer 和 SETR, 说明 Transformer 架构在微弱光谱差异建模方面更具优势。尽管 Hyper-net 和 Spectr 在部分任务中取得较好结果, 但其未能有效建模空

间与光谱间的交互关系, 导致在微弱差异和局部谱特征处理中存在信息丢失, 尤其在形态细长且存在重叠的样本中表现明显。相比之下, 本文方法通过独立的空间与光谱特征提取模块, 有效避免了空间-光谱特征耦合带来的干扰, 显著提升了区分性特征的捕获能力。在总体 MPA 和 MIoU 上分别实现了 8.0%、7.0% 和 18.6%、16.0% 的提升, 充分说明其在整体精度上的优势。同时在样本重叠、细长形态及光谱差异微弱等复杂场景中依旧保持稳定性能, 避免了在边缘和重叠区域中常见的预测误差。

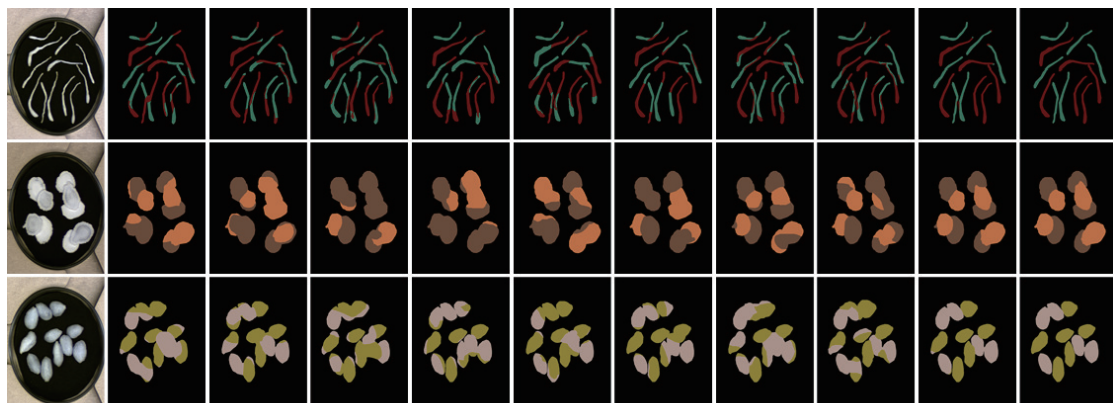
表 1 不同方法的分割结果比较
Table 1 Comparison of segmentation results by different methods

数据集	指标	SegFormer	SETR	Mask2Former	HybridSN	SpectralFormer	UNet	Hyper-net	Spectr	本文方法
金银花	MPA	0.830	0.832	0.712	0.792	0.809	0.809	0.830	0.845	0.961
	MIoU	0.686	0.688	0.502	0.621	0.656	0.657	0.679	0.703	0.901
人参	MPA	0.844	0.849	0.784	0.820	0.823	0.837	0.868	0.875	0.969
	MIoU	0.703	0.708	0.631	0.671	0.682	0.684	0.717	0.739	0.938
枸杞	MPA	0.876	0.881	0.774	0.832	0.839	0.842	0.878	0.892	0.952
	MIoU	0.757	0.763	0.602	0.691	0.703	0.706	0.759	0.783	0.927
贝母	MPA	0.934	0.939	0.852	0.924	0.929	0.925	0.939	0.941	0.978
	MIoU	0.859	0.865	0.704	0.841	0.849	0.843	0.869	0.871	0.942
半夏	MPA	0.962	0.965	0.862	0.955	0.961	0.953	0.968	0.977	0.981
	MIoU	0.921	0.927	0.715	0.887	0.899	0.883	0.901	0.934	0.957
平均值	MPA	0.889	0.893	0.797	0.864	0.872	0.873	0.896	0.904	0.968
	MIoU	0.785	0.790	0.631	0.742	0.757	0.755	0.784	0.804	0.933
总计	浮点数运算量/ 10^6	192.1	264.3	240.8	295.1	270.8	230.2	585.8	720.5	215.6

注: 加黑代表最优结果。

此外在计算量(浮点数运算量)上, 相对于其他基于 Transformer 的方法, 由于本文引入的稀疏注意力机制, 因此浮点数运算量仅有 215.6×10^6 , 均低于其他基于 Transformer 的方法和基于 3DCNN 的方法。这表明本文的方法具有较好的特征提取

效率和提取能力。这对检测方法的现实应用具有重要意义。本文还在图 4 中给出了定性分割结果, 以供比较。实验结果表明, 与其他方法相比, 本文方法能够得到更好的分割结果, 对不同任务场景具有更强的鲁棒性。



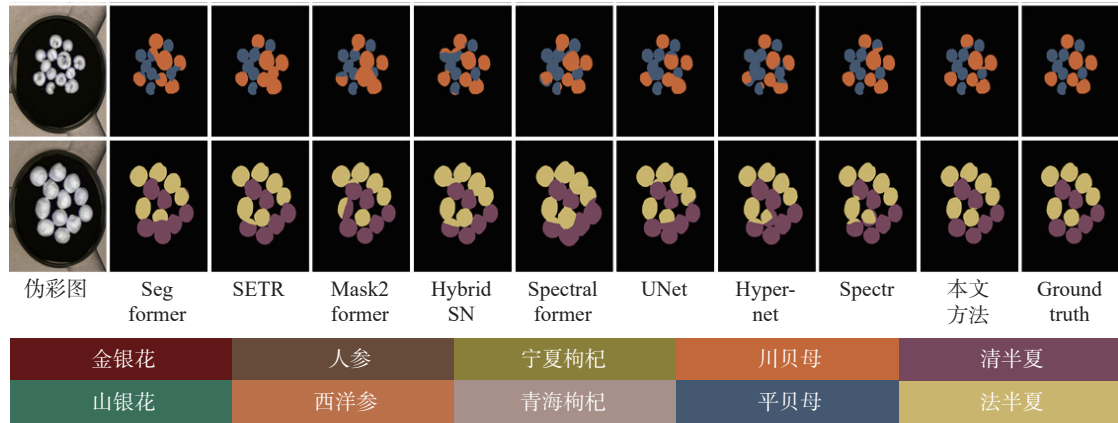


图 4 不同方法的可视化分割结果

Fig. 4 Segmentation results of different methods

3.4 消融实验

3.4.1 中药高光谱图像与 RGB 图像之间对比

为了验证 HSI 在中药材检测中的有效性, 本文使用来自 5 个自建中药材数据集的 RGB 图像 (波段为 450、550 和 650 nm) 进行了对比实验。图 5 给出了在 5 个数据集上, 使用 HSI 与使用 RGB 图像的 MPA 指标对比。由于 RGB 图像仅包含 3 个光谱波段, 其光谱信息有限, 因此 5 个数据集上的指标普遍不高。与 RGB 图像相比, 使用 HSI 检测的准确率能提升 15%~25%。尤其对

于基于注意力机制的模型来说, 由于 HSI 提供了丰富的光谱特征, 从而实现更高精度的预测。因此, 采用 HSI 进行中药材质量检测是一种有效的方法。此外, 在利用 RGB 图像进行检测时, 与主流方法相比, 本文提出的方法取得了最优的效果, 这得益于本文所设计的空间与光谱并行注意力模块。即使在光谱信息有限的情况下, 空间注意力通道仍能有效学习图像的空间特征差异, 而光谱注意力分支也能从有限的光谱信息中挖掘有效特征, 从而实现更优的分割效果。

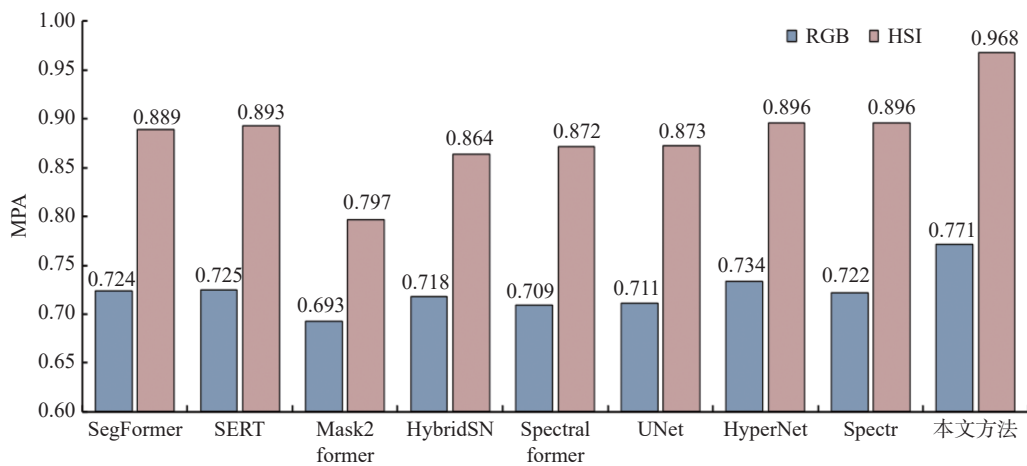


图 5 不同方法使用 HSI 与 RGB 图像的准确率对比

Fig. 5 Accuracy comparison between HSI and RGB image using state-of-the-art methods

3.4.2 模块消融对比

为了证明所提方法的有效性, 本文在 5 个数据集上进行了消融实验以评估超像素模块、稀疏注意力机制和双流超像素稀疏注意力的作用。共验证 11 种不同的组合, 其中 None 表示什么都不加入, 当仅加入空间注意力、光谱注意力、空间稀疏注意力、光谱稀疏注意力时, 注意力为标准的 ViT^[30] 网络。

表 2 给出了 11 种组合在 5 个数据集上的指标平均结果。相较于不加入任何模块, 当仅加入超像素模块时总体 MIoU 增加了 19.1%, 这证明了超像素模块的有效性; 相较于仅加入空间或光谱注意力, 当仅加入空间或光谱稀疏注意力时, 总体 MIoU 分别增加了 3.9% 和 2.9%, 这证明了超像素稀疏注意力的有效性。对比于单分支结构, 当使用双流网络结构时, MIoU 分别增加了 2.1% 和

1.4%, 这证明了双流网络的有效性。此外, 当使用本文的双流超像素注意力网络时, 总体 MIoU 分别增加了 2.5% 和 4.4%。因此, 设计的每个模块都在一定程度有助于提升中药高光谱图像的分割精度。

表 2 不同模块消融实验结果

Table 2 Ablation study results of different modules

方法	MPA	MIoU
None	0.754	0.621
超像素	0.884	0.812
空间注意力	0.901	0.839
光谱注意力	0.895	0.829
空间稀疏注意力	0.924	0.878

续表 2

方法	MPA	MIoU
光谱稀疏注意力	0.912	0.858
空间超像素稀疏注意力	0.941	0.908
光谱超像素稀疏注意力	0.930	0.889
双流注意力	0.918	0.856
双流稀疏注意力	0.939	0.891
双流超像素稀疏注意力	0.968	0.933

3.5 应用验证

为了验证本文提出方法的实际可用性, 如图 6、图 7 所示, 本文将检测算法部署到了 QT 上位机, 并结合高光谱成像控制系统, 构成检测软硬件系统。



图 6 检测软件系统

Fig. 6 Detection software system

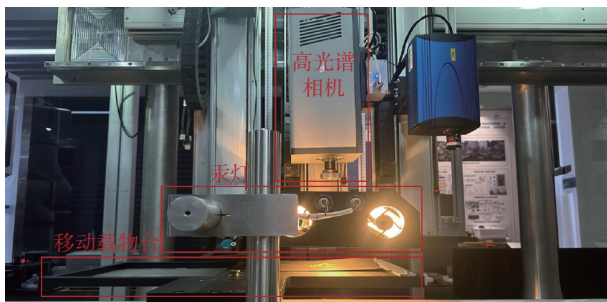


图 7 检测硬件系统

Fig. 7 Detection hardware system

同时, 本文使用另 10 份金银花-山银花、人参-西洋参、宁夏枸杞-青海枸杞、川贝母-平贝母和清半夏-法半夏进行测试。测试结果如表 3 所示。从表 3 中可以看出, 在实际场景中, 本文方法的总体平均像素精度和整体精度也能达到 0.966 和 0.927, 实现了中药质量的自动化高精度无损检测。

表 3 实际现场验证结果

Table 3 Results of real-world validation

数据集	MPA	MIoU
金银花-山银花	0.959	0.892
人参-西洋参	0.973	0.939
宁夏枸杞-青海枸杞	0.955	0.929
川贝母-平贝母	0.972	0.938
清半夏-法半夏	0.973	0.938
总计	0.966	0.927

4 结束语

为了提升中药质量无损检测的精度和效率, 提出了一种空间-光谱超像素稀疏注意力引导的中药高光谱图像分割网络, 并搭建了高光谱成像系统, 构建了具有像素级标注的中药质量检测高光谱数据集。本文结合超像素分割和空间-光谱稀疏注意力, 以解决中药检测场景中的同质区域

干扰和不相关区域干扰导致的微弱差异性特征提取能力弱效率低的问题。实验结果证明,本文提出的方法在中药高光谱数据集上总体平均像素精度 MPA 为 0.968, 总体平均总体精度 MIoU 为 0.933, 分类指标均高于主流检测方法。

在未来的工作中,将扩展数据集的规模和中药种类覆盖范围,并在保证检测精度的同时提高模型的检测速度。最终,模型将集成部署到中药质量检测产线上,确保市场流通中药的质量稳定可靠。同时,将引入机器人化操作,实现自动化的“检测-分离”,提高检测效率和智能化水平。

参考文献:

- [1] LI Yaolei, FAN Jing, CHENG Xianlong, et al. New revolution for quality control of TCM in industry 4.0: focus on artificial intelligence and bioinformatics[J]. *Trends in analytical chemistry*, 2024, 181: 118023.
- [2] YU Yang, YAO Changliang, GUO Dean. Insight into chemical basis of traditional Chinese medicine based on the state-of-the-art techniques of liquid chromatography-mass spectrometry[J]. *Acta pharmaceutica sinica B*, 2021, 11(6): 1469-1492.
- [3] LI Tao, ZHUANG Shuaixing, WANG Yiwei, et al. Flavonoid profiling of a traditional Chinese medicine formula of Huangqin Tang using high performance liquid chromatography[J]. *Acta pharmaceutica sinica B*, 2016, 6(2): 148-157.
- [4] PRIDEAUX B, LENAERTS A, DARTOIS V. Imaging and spatially resolved quantification of drug distribution in tissues by mass spectrometry[J]. *Current opinion in chemical biology*, 2018, 44: 93-100.
- [5] WANG Tianlong, LI Yicong, LIN Chunsheng, et al. Comprehensive analysis of natural polysaccharides from TCMs: a generic approach based on UPLC-MS/MS[J]. *Carbohydrate polymers*, 2022, 277: 118877.
- [6] ZHANG Fukai, JIN Xiaobo, LIN Gang, et al. Hybrid attention network for citrus disease identification[J]. *Computers and electronics in agriculture*, 2024, 220: 108907.
- [7] 黄鸿, 唐玉泉, 段宇乐. 面向高光谱分类的局部几何稀疏保持嵌入[J]. *自动化学报*, 2022, 48(10): 2496-2507.
HUANG Hong, TANG Yuxiao, DUANG Yule. Local geometry and sparsity preserving embedding for hyperspectral image classification[J]. *Acta automatica sinica*, 2022, 48(10): 2496-2507.
- [8] HU Wei, HUANG Yangyu, WEI Li, et al. Deep convolutional neural networks for hyperspectral image classification[J]. *Journal of sensors*, 2015, 2015: 258619.
- [9] SLAVKOVIKJ V, VERSTOCKT S, DE NEVE W, et al. Hyperspectral image classification with convolutional neural networks[C]//Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane: ACM, 2015.
- [10] HE Mingyi, LI Bo, CHEN Huahui. Multi-scale 3D deep convolutional neural network for hyperspectral image classification[C]//2017 IEEE International Conference on Image Processing. Beijing: IEEE, 2017.
- [11] LIAO Diling, SHI Cuiping, WANG Ligu. A spectral-spatial fusion transformer network for hyperspectral image classification[J]. *IEEE transactions on geoscience and remote sensing*, 2023, 61: 5515216.
- [12] 王瑞婷, 王海燕, 陈晓, 等. 基于混合卷积与三重注意力的高光谱图像分类网络[J]. *智能系统学报*, 2023, 18(2): 260-269.
WANG Ruiting, WANG Haiyan, CHEN Xiao, et al. Hyperspectral image classification based on hybrid convolutional neural network with triplet attention[J]. *CAAI transactions on intelligent systems*, 2023, 18(2): 260-269.
- [13] 涂潮, 刘万军, 赵琳琳, 等. 有限训练样本下的多尺度空洞密集网络高光谱影像分类[J]. *仪器仪表学报*, 2024, 45(4): 206-216.
TU Cao, LIU Wangjun, ZHAO Linlin, et al. Multiscale dilated dense network for hyperspectral image classification[J]. *Chinese journal of scientific instrument*, 2024, 45(4): 206-216.
- [14] ROY S K, KRISHNA G, DUBEY S R, et al. HybridSN: exploring 3-D-2-D CNN feature hierarchy for hyperspectral image classification[J]. *IEEE geoscience and remote sensing letters*, 2020, 17(2): 277-281.
- [15] HONG Danfeng, HAN Zhu, YAO Jing, et al. SpectralFormer: rethinking hyperspectral image classification with transformers[J]. *IEEE transactions on geoscience and remote sensing*, 2021, 60: 5518615.
- [16] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer Assisted Intervention. Munich: MICCAI Society, 2015.
- [17] WANG Qian, SUN Li, WANG Yan, et al. Identification of melanoma from hyperspectral pathology image using 3D convolutional networks[J]. *IEEE transactions on medical imaging*, 2021, 40(1): 218-227.
- [18] YUN Boxiang, LEI Baiying, CHEN Jieneng, et al. SpecTr: spectral transformer for microscopic hyperspectral pathology image segmentation[J]. *IEEE transactions on circuits and systems for video technology*, 2024, 34(6): 4610-4624.
- [19] SELCI S. The future of hyperspectral imaging[J]. *Journal of imaging*, 2019, 5(11): 84.
- [20] DELWICHE S R, KIM M S, DONG Yanhong. Fusarium

- damage assessment in wheat kernels by Vis/NIR hyperspectral imaging[J]. *Sensing and instrumentation for food quality and safety*, 2011, 5(2): 63–71.
- [21] ZHANG Jun, XU Binbo, WANG Zhiying, et al. Application of hyperspectral imaging in the detection of aflatoxin B1 on corn seed[J]. *Journal of food measurement and characterization*, 2022, 16(1): 448–460.
- [22] 史卓林, 杨增玲, 任朝霞, 等. 推扫式双相机高光谱成像系统设计及试验[J]. *农业机械学报*, 2024, 55(S1): 288–294, 305.
- SI Zhuoling, YANG Zengling, RENG Zhaoxia, et al. Design and test of push-broom dual-camera hyperspectral imaging system[J]. *Transactions of the Chinese society for agricultural machinery*, 2024, 55(S1): 288–294, 305.
- [23] VINCENT L, SOILLE P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations[J]. *IEEE transactions on pattern analysis and machine intelligence*, 1991, 13(6): 583–598.
- [24] COMANICIU D, MEER P. Mean shift: a robust approach toward feature space analysis[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2002, 24(5): 603–619.
- [25] VEDALDI A, SOATTO S. Quick shift and kernel methods for mode seeking[C]//10th European Conference on Computer Vision. Berlin: ECVA, 2008.
- [26] LEVINSHTEIN A, STEREA A, KUTULAKOS K N, et al. TurboPixels: fast superpixels using geometric flows[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2009, 31(12): 2290–2297.
- [27] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(11): 2274–2282.
- [28] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018.
- [29] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Munich: ECVA, 2018.
- [30] DOSOVITSKIY Alexey, BEYER Lucas, KOLESNIKOV Alexander, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020–10–20)[2025–06–16]. <https://arxiv.org/abs/2010.11929arXiv>.
- [31] XIE Enze, WANG Wenhai, YU Zhiding, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[J]. *Advances in neural information processing systems*, 2021, 34: 12077–12090.
- [32] ZHENG Sixiao, LU Jiachen, ZHAO Hengshuang, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021.
- [33] CHENG Bowen, MISRA I, SCHWING A G, et al. Masked-attention mask transformer for universal image segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022.
- [34] LUONG Minh-thang, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[EB/OL]. (2015–09–20)[2025–05–25]. <https://arxiv.org/abs/1508.04025>.
- [35] JIAO Jiayu, TANG Yuming, LIN Kunyu, et al. DilateFormer: multi-scale dilated transformer for visual recognition[J]. *IEEE transactions on multimedia*, 2023, 25: 8906–8919.
- [36] XIA Zhuofan, PAN Xuran, SONG Shiji, et al. Vision transformer with deformable attention[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022.

作者简介:



蒋雄杰, 硕士研究生, 主要研究方向为高光谱图像处理、计算机视觉。
E-mail: 1036046572@qq.com。



张辉, 教授, 博士生导师, 湖南大学人工智能与机器人学院院长, 入选国家高层次人才计划, 主要研究方向为机器视觉、高光谱、图像处理和机器人视觉。主持科技创新 2030—“新一代人工智能”重大项目课题、国家自然科学基金重点项目、国家重点研发计划子课题等 20 余项, 发表学术论文 60 余篇。E-mail: zhanghui1983@hnu.edu.cn。



王耀南, 教授, 博士生导师, 中国工程院院士。长期从事机器人感知与控制技术教学科研工作, 主持完成国家重大科技项目 20 余项, 成果获国家技术发明二等奖 1 项、国家科技进步二等奖 5 项、何梁何利基金科学与技术进步奖、国际 IEEE 机器人与自动化领域“工业应用最高奖”, 省部级一等奖 12 项。发表学术论文 500 余篇。E-mail: yaonan@hnu.edu.cn。