

基于SAM的中药粉末显微图像分阶段框监督实例分割方法

马景鑫, 赵 晶, 江 宁, 谢怡宁

东北林业大学控制与信息工程学院, 黑龙江 哈尔滨

收稿日期: 2026年9月23日; 录用日期: 2026年10月20日; 发布日期: 2026年10月27日

摘 要

针对中药粉末显微图像中实例分割面临的目标密集分布、背景杂质干扰与矩形框标注缺乏像素级边界信息等问题, 本文提出一种基于Segment Anything Model (SAM)的分阶段框监督实例分割框架。该框架依次包含三个阶段: 首先, 利用网格提示SAM自动生成高召回候选掩码, 并通过面积约束去除明显碎片, 解决目标密集场景下的实例召回问题; 其次, 构建形状与外观双分支候选拒识网络, 外观分支以候选区域RGB图像与掩码为四通道输入, 利用轻量卷积网络学习纹理判别特征; 形状分支提取17维几何特征, 包括紧致度、凸实度、Hu不变矩等, 利用多层感知机学习轮廓判别特征; 两分支在logit空间进行加权融合, 有效抑制背景杂质和过分割碎片造成的假阳性; 最后, 利用内缩框构造伪掩码微调SAM的掩码解码器, 以选择性补全不完整掩码, 并引入精修前后掩码的自一致性交并比self-IoU门控机制防止精修漂移。在包含3191幅中药粉末显微图像、约44.7万个标注实例的数据集上的实验结果表明, 本文方法在仅使用矩形框标注的条件下, F1分数达到0.863, 实例级平均交并比mIoU达到0.809, 显著优于多个主流方法。此外, 消融实验也验证了各组件的独立贡献与有效性。

关键词

SAM, 图像分割, 框监督学习

A Staged Box-Supervised Instance Segmentation Method Based on SAM for TCM Powder Microscopy Images

Jingxin Ma, Jing Zhao, Ning Jiang, Yining Xie

School of Control and Information Engineering, Northeast Forestry University, Harbin Heilongjiang

Received: September 23, 2026; accepted: October 20, 2026; published: October 27, 2026

文章引用: 马景鑫, 赵晶, 江宁, 谢怡宁. 基于SAM的中药粉末显微图像分阶段框监督实例分割方法[J]. 计算机科学与应用, 2026, 16(10): 156-170. DOI: 10.12677/csa.2026.1610319

Abstract

To address the challenges of dense target distribution, background impurity interference, and the lack of pixel level boundary information from bounding box annotations in instance segmentation of traditional Chinese medicine powder microscopy images, this paper proposes a staged, box supervised instance segmentation framework built upon the Segment Anything Model. The framework comprises three stages. First, a grid prompted SAM automatically generates high recall candidate masks, and an area constraint removes obvious debris to address the instance recall problem in dense target scenarios. Second, a shape appearance dual branch candidate rejection network is constructed: the appearance branch takes a four channel input combining RGB image patches of candidate regions with their masks, and employs a lightweight convolutional network to learn texture discriminative features; the shape branch extracts a 17 dimensional geometric feature vector including compactness, solidity, Hu invariant moments, and so forth, and employs a multi layer perceptron to learn contour discriminative features. The two branches are fused in logit space via weighted summation, effectively suppressing false positives caused by background impurities and over segmentation fragments. Finally, shrunken box pseudo masks are leveraged to fine tune the SAM mask decoder for selective completion of incomplete masks, and a self IoU gating mechanism comparing masks before and after refinement is introduced to prevent refinement drift. Experimental results on a dataset containing 3191 TCM powder microscopy images with approximately 447,000 annotated instances demonstrate that, using only bounding box annotations, the proposed method achieves an F1 score of 0.863 and an instance level mean IoU of 0.809, significantly outperforming multiple mainstream methods. In addition, ablation studies validate the individual contributions and effectiveness of each component.

Keywords

SAM, Image Segmentation, Box-Supervised Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中药材品种的准确鉴定是保障用药质量与安全的重要环节。药材经过切制、破碎或粉末化后,部分宏观性状会减弱,而组织碎片、细胞、结晶、纤维和导管等微观结构仍可为品种鉴别提供依据[1]-[4]。因此,显微鉴定在中药粉末及混合样品的质量评价中具有重要作用。然而,传统方法需要专业人员在大量视野中搜索和辨认特征结构,工作量大,鉴定结果还容易受到人员经验、样本制备和成像质量的影响。利用图像分析方法提高显微鉴定的客观性与重复性,已成为中药材质量评价数字化的重要研究方向。

随着数字显微成像和深度学习的发展,卷积神经网络已被用于中药材显微特征的分类与检测,并通过注意力机制、数据增强和多尺度特征学习提高识别性能[5][6]。这类方法通常以预先裁切的区域或矩形框作为识别对象,但中药粉末显微图像中往往同时存在大量目标实例和背景结构,分类前仍需获得较为完整、背景干扰较少的实例区域。若同一目标被分割为多个碎片,或候选区域包含较多杂质,便会影响后续裁剪、分类和计数。FCN、U-Net 和 Mask R-CNN 等方法推动了图像分割的发展[7]-[9],但其训练通常依赖像素级标注。对于实例密集且边界模糊的中药粉末显微图像,逐实例勾画轮廓成本较高,实际数据更容易获得矩形框标注。虽然框监督方法能够降低标注负担[10][11],但矩形框同时包含目标与背景,

难以直接提供准确的边界监督。

Segment Anything Model (SAM)能够根据点、框或掩码提示生成目标区域,为弱标注条件下的图像分割提供了新的思路[12]。相关研究已将 SAM 应用于医学图像和显微图像,并通过领域适配改善其在专业场景中的表现[13]-[15]。然而,直接将 SAM 用于中药粉末显微图像仍面临明显困难。基于网格点提示的自动分割需要布置密集提示点以搜索潜在目标,同一目标内部的多个提示可能产生重叠或破碎的候选掩码,背景污渍和纤维团块也可能被识别为目标。简单提高面积阈值或 SAM 置信度阈值虽然能够删除部分假阳性,却可能同时丢失边界模糊、对比度较低的真实目标。此外,矩形框标注只能限定目标的大致位置,无法直接约束掩码边界。因此,该任务的关键并非单纯提高模型容量,而是在有限监督条件下兼顾目标召回、候选可靠性和掩码完整性。

针对上述问题,本文提出一种面向中药粉末显微图像的分阶段实例分割框架,整体流程如图 1 所示。该框架利用网格点提示的 SAM 生成候选掩码,并通过面积约束去除明显的小碎片。随后,利用形状与外观信息判断候选质量,筛除背景杂质和过分割结果。对于保留的候选,利用矩形框构造伪掩码并微调 SAM 的掩码解码器,以实施选择性补全。为避免精修掩码偏离原目标,本文根据精修前后掩码的自一致性交并比(self-IoU)决定是否接受精修结果。该框架分别处理候选生成、候选质量判断和掩码补全,使各阶段对应不同的分割失效原因。

本文主要贡献包括:

- 1) 针对矩形框标注缺少像素级边界信息的问题,提出一种面向中药粉末显微图像的分阶段实例分割框架。该框架依次生成候选、剔除错误候选并补全不完整掩码,以兼顾目标召回、候选可靠性和掩码完整性。
- 2) 提出一种形状与外观信息相结合的双分支候选拒识方法,利用图像表观特征和候选几何特征判断掩码质量,以抑制背景杂质和过分割碎片造成的假阳性。
- 3) 提出一种框监督选择性掩码补全方法,通过微调 SAM 的掩码解码器补全候选,并利用 self-IoU 回退机制控制精修漂移,在无需像素级标注的条件下提高掩码精修的稳定性。

2. 相关工作

2.1. 中药材显微图像智能分析

数字显微成像与深度学习的发展,为中药材显微特征的自动识别提供了新的技术路径。已有研究主要围绕显微结构分类与目标检测展开。Ma 等[5]利用卷积神经网络学习中药材显微图像的判别特征,验证了深度表征用于显微鉴别的可行性。Zhu 等[6]进一步结合分割组合式数据增强与双重注意机制,以增强复杂显微背景下的特征表达。这些方法提高了特征结构的识别能力,但研究重点多为类别判定或位置检测,通常未对密集目标的实例轮廓进行精细建模。

在更广泛的显微图像分析领域,StarDist 以星凸多边形表示细胞或细胞核实例[16],Cellpose 则通过通用化的深度学习模型提高了不同成像条件下的细胞分割适用性[17]。近期研究开始将基础分割模型用于自动化显微实例分割。CellSAM 通过检测器生成矩形框提示,并利用 SAM 输出细胞实例掩码[18]。SAMCell 面向无标记细胞显微图像进行领域训练,以提高跨细胞类型和跨显微设备的适用性[19]。mSAMUNet 进一步将 SAM 与多尺度 U-Net 编码结构结合,用于处理密集细胞、模糊边界和尺度变化[20]。这些方法说明显微图像分割正由特定形状建模转向自动提示和领域适配,但其训练仍主要依赖像素级实例标注或大规模细胞数据。中药粉末中的组织碎片、导管、纤维和结晶等结构在尺度、形状及纹理上差异较大,现有细胞分割范式难以直接兼顾标注成本与实例提取质量。

2.2. 矩形框监督的弱监督分割

矩形框标注成本低于像素级掩码，因此，利用框标注学习目标轮廓已成为弱监督分割的重要方向。BoxSup 通过交替更新候选区域和分割网络，将检测框转化为可用于网络训练的区域监督[10]。BoxInst 利用掩码投影约束和基于颜色相似性的像素对约束，在不改变实例分割网络主体的情况下实现框监督训练[11]。为减小伪掩码噪声的影响，BoxTeacher 引入教师模型生成伪掩码，并依据掩码质量调整训练权重与损失[21]。上述研究表明，矩形框能够为实例掩码学习提供有效约束，但框内同时包含目标和背景，监督信息具有天然歧义。当标注框不够紧致时，分割性能还可能明显下降[22]。

在更广义的低标注分割研究中，SAM-Match 利用自动生成的提示驱动微调后的 SAM 修正中间伪标签，并将精修结果反馈给半监督模型[23]。这一思路体现了基础模型辅助伪标签优化的潜力，但其目标仍是训练端到端语义分割模型。中药粉末显微图像中的目标密集分布、相互接触且类别间形态差异显著，自动生成的候选还可能同时包含背景杂质、不完整碎片和重复实例。在这种情况下，仅约束预测掩码与矩形框的一致性，难以判断候选是否对应完整、可靠的目标。因此，除利用框标注补充边界监督外，还需要在候选层面显式抑制错误实例，降低低质量候选对后续掩码学习与图像分类的影响。

2.3. SAM 在医学与显微图像分割中的应用

SAM 通过点、框或粗掩码提示生成目标区域，为低标注成本的通用分割提供了基础模型[12]。多数数据集评测表明，SAM 在不同成像模态和目标类型上的表现差异较大，分割质量会受到图像对比度、目标尺度、边界复杂度和提示方式的共同影响[13][24]。为提高专业场景中的适用性，MedSAM 利用大规模医学图像数据进行领域适配[14]。Medical SAM Adapter 通过轻量适配模块向 SAM 注入医学领域知识，减少全模型微调需求[25]。Segment Anything for Microscopy 则面向光学与电子显微图像扩展了交互式 and 自动实例分割能力[15]。这些进展表明，SAM 在专业图像上的改进主要沿领域表征适配、提示生成和解码器调整展开。

在与本文较为接近的密集微粒场景中，CryoSegNet 将注意力门控 U-Net 与 SAM 自动掩码生成器结合，用于低信噪比冷冻电镜图像中的蛋白质颗粒提取，并通过阈值和后处理减少错误候选[26]。Gaillochet 等将矩形框约束、基础分割模型生成的伪标签与提示学习结合，用于弱监督医学图像分割[27]。这些工作表明，SAM 与专用网络或框约束结合能够改善领域图像分割，但密集提示产生的冗余候选、背景结构造成的假阳性以及精修结果可能发生的目标漂移，仍需针对具体任务进行控制。基于此，本文将 SAM 作为高召回候选生成器，在此基础上利用形状与外观信息进行候选质量判断，再通过框监督的掩码解码器实施选择性补全。该设计从候选生成、质量判断和掩码精修三个环节处理不同误差，与主要依赖领域微调、提示优化或固定阈值后处理的方法形成区别。

3. 方法

3.1. 问题定义与框架概览

本文研究矩形框监督条件下的中药粉末显微图像类别无关实例分割问题。设训练集为：

$$\mathcal{D} = \left\{ (I_n, \mathcal{B}_n) \right\}_{n=1}^N \quad (1)$$

其中， $I_n \in \mathbb{R}^{H \times W \times 3}$ 表示第 n 幅显微图像， $\mathcal{B}_n = \{b_{n,j}\}_{j=1}^{J_n}$ 表示其矩形框标注集合。训练数据不提供像素级实例轮廓。给定测试图像 I ，模型需要输出类别无关的实例掩码集合 $\mathcal{M}^* = \{m_i^*\}_{i=1}^K$ 。分割阶段不判断药材类别，而是利用实例掩码定位和裁剪显微颗粒，为后续分类提供目标区域。

该任务面临三类困难。单幅显微图像中的颗粒数量与位置未知，依赖单一人工提示难以覆盖全部目标。粉末图像中普遍存在碎屑、纤维、气泡及染色不均等干扰，通用分割模型生成的候选中同时包含背景结构、重复片段和覆盖不足的目标。此外，矩形框只能描述目标的大致空间范围，无法直接提供真实边界，用其训练掩码预测器容易引入框内背景。针对这些问题，本文构建由候选生成、双分支拒识和选择性补全组成的分阶段框监督实例分割框架，如图 1 所示。

候选生成模块利用分割基础模型的密集网格提示覆盖潜在颗粒，并通过面积约束去除明显的小碎片。双分支拒识模块分别从局部外观与几何形状两个角度判断候选质量，以抑制与真实标注区域不一致的候选。选择性补全模块利用矩形框构造范围型伪监督，仅微调掩码解码器，再依据精修前后掩码的一致性接受新掩码或回退至原候选。三个模块分别处理潜在实例召回、低质量候选抑制和不完整掩码修正问题。第 3.2 至 3.4 节依次介绍各模块，完整训练与推理过程见第 3.5 节。

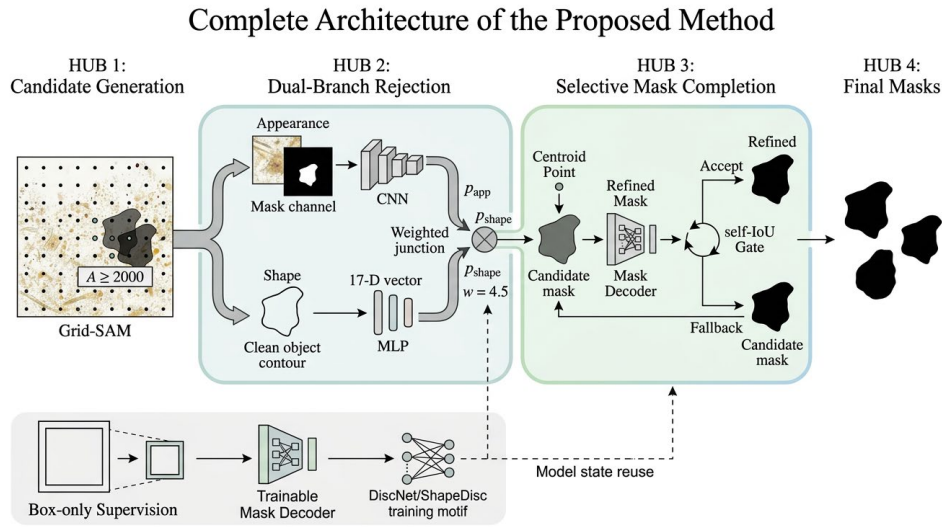


Figure 1. Architecture of the model

图 1. 模型结构图

3.2. 基于网格提示的高召回候选生成

测试图像不提供目标数量、位置或人工提示，因此，本研究首先利用 SAM [12] 自动掩码生成器在整幅图像上布置均匀点网格。设网格提示集合为 $\mathcal{P} = \{p_i\}_{i=1}^L$ ，SAM 对不同位置的提示点进行编码，并结合预测交并比、稳定性分数和非极大值抑制产生初始候选集合：

$$\mathcal{C}_0 = \text{SAM}_{\text{auto}}(I, \mathcal{P}) \quad (2)$$

本文采用 ViT-H 版本 SAM，每个方向布置 32 个提示点，预测交并比阈值与稳定性阈值分别设为 0.88 和 0.95，并在自动生成阶段删除面积小于 30 像素的连通区域。密集提示有利于发现尺寸、形态与对比度差异较大的颗粒，但也会同时产生大量细小纹理和过分割碎片。为降低后续拒识模块的计算负担，进一步采用面积阈值 A_0 进行初筛：

$$\mathcal{C} = \{m_i \in \mathcal{C}_0 \mid |m_i| \geq A_0\} \quad (3)$$

其中， $|m_i|$ 表示候选掩码 m_i 的前景像素数，本文设置 $A_0 = 2000$ 。面积初筛只删除尺度明显不符合目标分布的区域，不承担完整的真假候选判别，以避免低对比度或边界模糊的真实颗粒在早期被误删。

3.3. 形状与外观双分支候选拒识

面积相近并不意味着候选具有相同语义。背景污渍和纤维团块可能具有与真实颗粒接近的尺度，同一颗粒也可能被分割为多个局部片段。仅依据 SAM 自身置信度难以同时识别这些错误。颜色与纹理能够反映候选的局部表现，但易受照明和染色变化影响。轮廓形状具有较强的可解释性，却不能充分描述候选内部信息。为利用二者的互补性，本文分别建立外观分支 f_a 和形状分支 f_s ，并在 logit 空间融合其预测结果，如图 2 所示。

3.3.1. 候选样本与监督标签构建

由于训练集没有像素级真值掩码，本文将矩形框 b_j 填充为二值区域 g_j ，以候选与所有标注区域的最大交并比构造候选级监督。对于面积初筛后的候选 m_i ，其标签定义为：

$$y_i = \mathbb{I} \left[\max_j \text{IoU}(m_i, g_j) \geq \tau_m \right] \quad (4)$$

其中， $\mathbb{I}[\cdot]$ 为指示函数， $\tau_m = 0.5$ 。满足条件的候选被视为可保留候选，其余候选作为拒识样本。需要说明的是，负样本不仅包括背景杂质，还包括严重偏移、覆盖不足和过分割形成的局部片段。因此，该模块学习的是候选与标注目标范围的一致性，即候选质量判别，而不是简单的前景材质分类。

3.3.2. 基于 RGB 与掩码联合表示的外观分支

对于候选 m_i ，根据其外接矩形 r_i 分别裁取原始 RGB 区域和对应的二值掩码。二者缩放至 64×64 后沿通道维拼接，得到四通道输入：

$$x_i^a = \text{Resize}_{64} \left(\left[I|_{r_i}, m_i|_{r_i} \right] \right) \in \mathbb{R}^{64 \times 64 \times 4} \quad (5)$$

其中，RGB 通道描述候选及其邻域的颜色与纹理，掩码通道显式指出网络需要判断的区域，避免仅凭裁剪图像推断候选位置。外观分支采用轻量卷积网络 DiscNet，由 4 个卷积块组成，通道数依次为 32、64、128 和 256。每个卷积块包含 3×3 卷积、批归一化、ReLU 激活和最大池化。卷积特征经全局平均池化、Dropout 和全连接层后输出二分类概率：

$$p_i^a = \left[\text{Softmax} \left(f_a(x_i^a) \right) \right]_1 \quad (6)$$

训练时采用带类别权重的交叉熵损失，并使用 90° 旋转、水平翻转和垂直翻转增强，以减小显微颗粒方向变化对判断结果的影响。

3.3.3. 基于显式几何描述的形状分支

为提供不依赖局部颜色的判别依据，本文从每个候选的最大外轮廓中提取 17 维形状向量 x_i^s 。其中，10 个轮廓统计量包括面积、周长、紧致度、凸实度、外接矩形占比、长宽比、离心率、拟合椭圆轴比、圆度和凸包周长比。其余 7 维为经对数变换的 Hu 不变矩。以候选面积 A_i 、周长 P_i 、凸包面积 A_i^h 及外接矩形宽高 w_i, h_i 为例，部分特征可表示为：

$$c_i = \left(\frac{4\pi A_i}{P_i^2} \right), s_i^{\text{sol}} = \left(\frac{A_i}{A_i^h} \right), e_i^{\text{ext}} = \left(\frac{A_i}{w_i h_i} \right) \quad (7)$$

所有形状特征使用训练集统计量进行标准化，得到 $\tilde{x}_i^s$ 。ShapeDisc 采用 $17 \rightarrow 64 \rightarrow 32 \rightarrow 1$ 的多层感知机结构，中间层使用 ReLU 和 Dropout，输出形状分支的正类概率：

$$p_i^s = \sigma \left(f_s(\tilde{x}_i^s) \right) \quad (8)$$

其中, $\sigma(\cdot)$ 为 Sigmoid 函数。该分支使用带正类权重的二元交叉熵损失进行优化。显式形状量能够直接刻画候选的轮廓紧凑性、填充程度与拉伸程度, 为外观分支提供可解释的几何补充。

3.3.4. 双分支 logit 融合

直接对两个概率取均值会压缩高置信度预测的差异。为此, 本文先将分支输出映射到 logit 空间, 再进行加权融合:

$$z_i = \text{logit}(p_i^a) + \lambda \text{logit}(p_i^s), q_i = \sigma(z_i) \quad (9)$$

其中, $\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$, λ 控制形状分支对最终决策的影响, 本文取 $\lambda = 4.5$ 。保留的候选集合为:

$$C_r = \{m_i \in C \mid q_i \geq \tau_q\} \quad (10)$$

其中, $\tau_q = 0.5$ 。外观分支关注候选内部的颜色、纹理及其邻域关系, 形状分支关注轮廓规则性与完整程度。logit 级融合将两类证据统一到可加的决策空间, 使最终拒识不依赖单一特征来源。

Shape-Appearance Dual-Branch Rejection

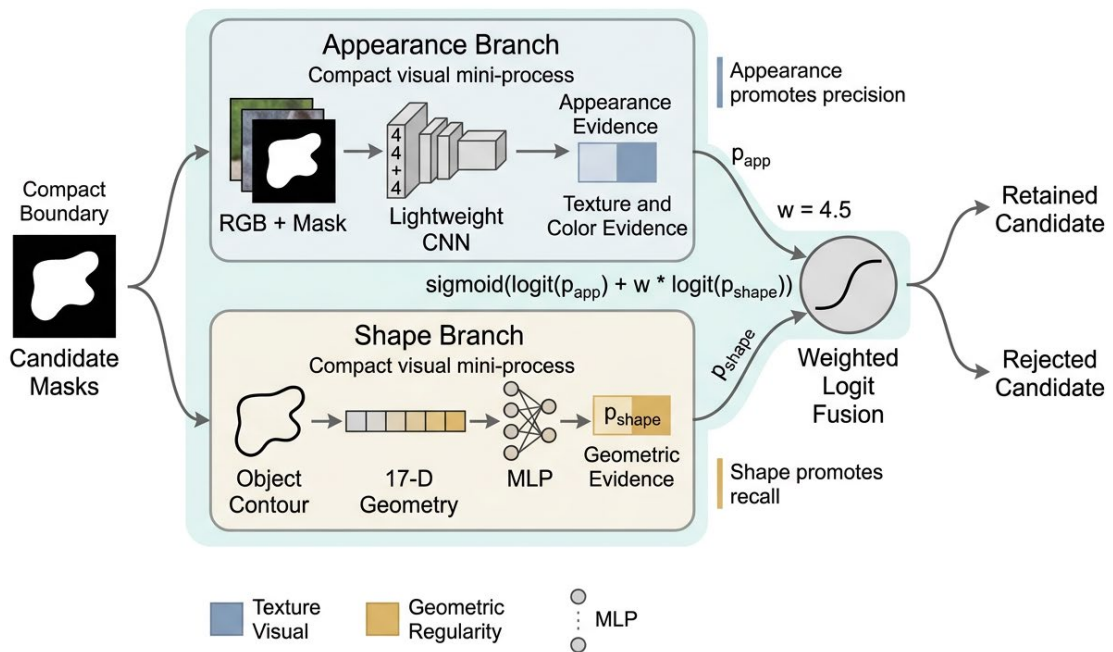


Figure 2. Schematic of the dual-branch candidate rejection method for shape and appearance

图 2. 形状与外观双分支候选拒识方法示意图

3.4. 框监督选择性掩码补全

双分支拒识能够删除低质量候选, 却不会主动扩大被截断或覆盖不足的真实候选。直接用填充矩形框训练分割模型, 会将框内背景误作目标。无条件采用微调模型的输出, 也可能使原本正确的候选扩张至相邻颗粒。为此, 本文将弱监督掩码学习与候选一致性控制结合。训练阶段利用内缩矩形构造范围型伪掩码, 仅更新 SAM 掩码解码器。推理阶段以已有候选为起点生成精修结果, 并通过 self-IoU 决定是否接受精修结果, 如图 3 所示。

3.4.1. 内缩框伪监督与掩码解码器微调

设标注框为 $b_j = (x_0, y_0, x_1, y_1)$ ，其中心、宽度和高度分别为 (c_x, c_y) 、 w 和 h 。为减小宽松标注框中背景区域的影响，以缩放系数 ρ 构造同中心内缩框：

$$\tilde{b}_j = \left(c_x - \left(\frac{\rho w}{2} \right), c_y - \left(\frac{\rho h}{2} \right), c_x + \left(\frac{\rho w}{2} \right), c_y + \left(\frac{\rho h}{2} \right) \right) \quad (11)$$

并将其内部填充为伪掩码 $\tilde{g}_j$ 。本文设置 $\rho = 0.8$ 。训练时，每轮从一幅图像的标注框集合中随机抽取一个实例，并在其内缩框中随机采样一个正提示点。图像经冻结的 SAM 图像编码器得到图像嵌入，提示点经冻结的提示编码器得到提示嵌入，训练过程中仅更新掩码解码器。设解码器输出为 $\hat{g}_j$ ，其训练目标为：

$$\mathcal{L}_{\text{mask}} = \text{BCE}(\hat{g}_j, \tilde{g}_j) + \text{DiceLoss}(\hat{g}_j, \tilde{g}_j) \quad (12)$$

冻结图像编码器和提示编码器，可在保留预训练图像表征与提示响应能力的同时，将弱监督更新限制在掩码生成端。内缩框伪掩码并非真实轮廓，其作用是向解码器提供相对保守的目标范围约束。具体边界仍由 SAM 的预训练形状先验和测试时的候选提示共同决定。

3.4.2. 候选引导的单点精修

推理阶段只对双分支保留的候选进行精修。对于 $m_i \in \mathcal{C}_r$ ，设其外接矩形为 $(x_i^L, y_i^T, x_i^R, y_i^B)$ ，其中上标分别表示左、上、右、下边界。外接矩形中心为：

$$u_i = \left(\left(\frac{x_i^L + x_i^R}{2} \right), \left(\frac{y_i^T + y_i^B}{2} \right) \right) \quad (13)$$

作为单个正点提示，并将 u_i 与测试图像共同输入微调后的 SAM，得到精修掩码 $\hat{m}_i$ 。该设计复用候选所提供的目标位置，使补全阶段与前一阶段保留的实例一一对应，避免重新执行密集网格提示而产生额外的重复候选。

3.4.3. 基于 Self-IoU 的风险控制

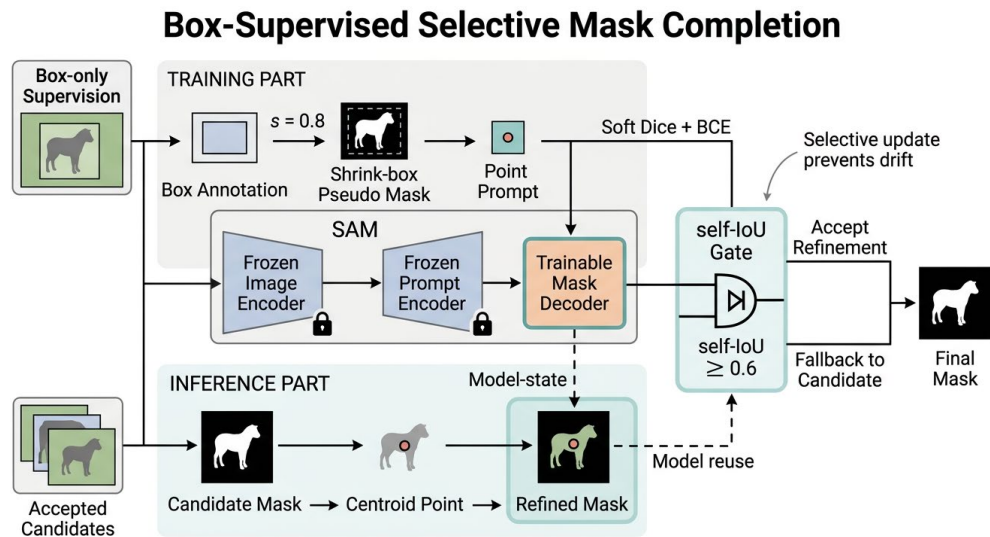


Figure 3. Schematic of the frame-supervised selective mask completion method

图 3. 框监督选择性掩码补全方法示意图

单点提示可能使掩码扩张到邻近结构，且框监督微调无法保证每次精修均优于原候选。因此，本文以精修前后掩码的交并比衡量更新幅度：

$$r_i = \text{IoU}(m_i, \hat{m}_i) \quad (14)$$

并将最终掩码定义为：

$$m_i^* = \begin{cases} \hat{m}_i, & r_i \geq \tau_s \\ m_i, & r_i < \tau_s \end{cases} \quad (15)$$

其中， $\tau_s = 0.6$ 。当精修结果与原候选保持足够重叠时，接受精修掩码。当二者差异过大时，回退至原候选。self-IoU 不用于判断掩码与未知真实轮廓的绝对精度，而是衡量精修结果相对原候选的变化幅度，以防止弱监督解码器产生过大的目标偏移。

4. 实验

4.1. 实验设置

4.1.1. 数据集

本文使用的数据集包含 3191 幅中药粉末显微图像，图像分辨率为 2048×2048 像素，均为 RGB 彩色图。数据涵盖 12 个中药品种，共包含约 44.7 万个显微目标实例。所有标注均采用 LabelMe 工具以四点轴对齐矩形框(axis-aligned bounding box)形式给出，不提供像素级实例掩码。数据集按照药材批次进行分层划分(stratified split)，确保同一批次的图像仅出现在训练集、验证集或测试集中的某一个子集内，以验证模型在未见批次上的泛化能力。数据集按照约 7:1.5:1.5 的比例划分为训练集、验证集和测试集，分别包含约 2230 幅、480 幅和 481 幅图像。

4.1.2. 评估指标

本文采用类别无关的实例级评估方式，即仅评价预测掩码与标注框的空间对应关系，不区分药材类别。具体使用以下四项指标：准确率(Precision, P)：以交并比(Intersection over Union, IoU)大于等于 0.5 作为正匹配阈值，计算预测掩码中被正确匹配的比例。召回率(Recall, R)：标注目标中被正确匹配的比例。F1 分数(F1-score)： P 与 R 的调和平均数，综合衡量分割结果的整体质量。例级平均交并比(instance-level mean IoU, mIoU)：所有成功匹配的实例对之间 IoU 的平均值，反映分割掩码相对于标注区域的边界精度。

4.1.3. 实现细节

候选生成阶段采用 ViT-H 版本的 SAM [12]作为分割基础模型。网格提示点设置为 32×32 的均匀分布，共 1024 个提示点。自动掩码生成器的预测 IoU 阈值设为 0.88，稳定性分数阈值设为 0.95，并删除面积小于 30 像素的连通区域。面积初筛阈值设为 2000 像素，用于滤除明显小于目标分布尺度的碎片。

外观判别网络以四通道输入(RGB 三通道 + 候选掩码二值图)，将候选区域缩放至 64×64 像素后，经四个卷积块逐级提取特征。卷积块通道数依次为 32、64、128 和 256，每个卷积块包含 3×3 卷积、批归一化、ReLU 激活和 2×2 最大池化。卷积特征经全局平均池化、Dropout 和全连接层输出二分类概率。网络总参数量约 0.5 M，采用 Adam 优化器，初始学习率为 1×10^{-3} 的负 3 次方，训练 30 个 epoch。数据增强包括 90 度随机旋转、水平翻转和垂直翻转。

形状判别器从每个候选掩码的最大外轮廓中提取 17 维几何特征向量，包括：面积、周长、紧致度、凸实度、外接矩形占比、长宽比、离心率、拟合椭圆轴比、圆度、凸包周长比以及 7 个经对数变换的 Hu 不变矩。所有形状特征使用训练集统计量进行 Z-score 标准化。ShapeDisc 采用三层全连接网络结构，隐藏层维度为 64 和 32，中间层使用 ReLU 激活和 Dropout，输出层经 Sigmoid 函数得到正类概率。优化器

采用 AdamW，初始学习率为 1×10 的负 3 次方，训练 100 个 epoch。

用于训练拒识模块的候选样本自动从训练集图像中提取。对于每幅训练图像，先运行网格 SAM 生成初始候选，经面积初筛后，将每个候选与所有填充标注矩形框计算 IoU。若候选与任意标注框的最大 IoU 大于等于 0.5，则标记为正样本；否则标记为负样本。最终获得约 2000 个训练样本，其中正样本约 1709 个，负样本约 293 个。训练时对正样本施加类别权重以缓解类别不平衡。

掩码补全阶段的内缩系数设为 0.8，即标注框的长和宽各缩减至原来的 0.8 倍后填充为伪掩码。训练时冻结 SAM 的图像编码器和提示编码器，仅更新掩码解码器。损失函数为 Dice 损失与二元交叉熵损失的等权组合。采用 AdamW 优化器，初始学习率为 1×10 的负 4 次方，训练 30 个 epoch，等效批量大小为 4。推理阶段，self-IoU 门控阈值设为 0.6，当精修掩码与原候选掩码的 IoU 大于等于此阈值时接受精修结果，否则回退至原始候选掩码。

所有实验在单张 NVIDIA RTX 3080 Ti (12 GB 显存)上完成。

4.2. 对比实验

为全面评估本文方法的有效性，选取了以下七种具有代表性的基线方法进行对比。Otsu + Watershed：经典图像处理方法，以 Otsu 自适应阈值进行二值化后使用分水岭算法分割连通区域，代表传统无监督分割范式。U-Net [8]：生物医学图像分割的经典全卷积网络，以填充标注框作为伪掩码监督进行训练。DeepLabV3+：基于空洞卷积的语义分割模型，采用 ResNet-101 骨干网络，训练监督方式同 U-Net。SegFormer：基于层次化 Transformer 的语义分割模型，采用 MiT-B2 编码器，训练监督方式同上。Mask R-CNN [9]：代表性实例分割模型，以 ResNet-50 + FPN 为骨干网络，利用矩形框标注训练检测分支，并以填充框训练掩码头。YOLOv8-seg：基于 YOLOv8 的实例分割变体，训练方式与 Mask R-CNN 类似。Cellpose [17]：面向细胞显微图像的通用分割模型，使用其预训练权重直接推理，不加微调。对比实验结果如表 1 所示。

Table 1. Comparison with existing methods on the test set

表 1. 测试集上不同方法的对比实验结果

方法	Precision	Recall	F1	mIoU
Otsu + Watershed	0.002	0.001	0.001	0.035
U-Net	0.337	0.357	0.342	0.696
DeepLabV3+	0.325	0.236	0.271	0.650
SegFormer	0.210	0.263	0.225	0.648
Mask R-CNN	0.649	0.554	0.594	0.788
YOLOv8-seg	0.749	0.324	0.442	0.799
Cellpose	0.236	0.106	0.140	0.627
Ours	0.871	0.855	0.863	0.809

从表 1 可以看出，Otsu + Watershed 方法在复杂纹理背景的中药粉末图像上几乎完全失效，F1 分数仅为 0.001。其根本原因在于显微图像中存在大量染色不均、纤维结构和气泡干扰，单一的全局阈值无法区分目标颗粒与背景纹理，分水岭算法产生了严重的过分割。

三类语义分割模型(U-Net、DeepLabV3+和 SegFormer)的 F1 分数在 0.225 至 0.342 之间，显著优于传统方法但仍远低于实例分割模型。这些模型以填充标注矩形框为监督信号进行训练，由于矩形框内包含

大量背景像素,网络难以学习精确的目标边界。U-Net 得益于其跳跃连接结构对多尺度特征的保持能力,在三者中取得了最高的 F1。SegFormer 作为基于 Transformer 的模型,在该数据集上表现最弱,可能因为 Transformer 对局部细节的保持能力弱于 CNN,而中药粉末显微图像的纹理细节对于区分目标与背景至关重要。

Mask R-CNN [9]利用检测分支进行目标定位, F1 达到 0.594, 召回率为 0.554, 表明其能够召回超过一半的真实目标。然而,其掩码头以填充矩形框为监督,生成的掩码精度受限, mIoU 为 0.788。YOLOv8-seg 呈现出典型的高准确率低召回率模式,说明其检测器倾向于仅输出高置信度目标,漏检了大量中等或低对比度的显微颗粒。

Cellpose [17]作为专门面向细胞图像设计的分割模型,在中药粉末图像上表现不佳。其预训练模型对细胞形态(近圆形、均匀纹理)有较强的归纳偏置,而中药粉末颗粒形状多样(从近圆形淀粉粒到细长纤维和导管碎片),且背景复杂度远超典型细胞显微图像。这一结果也印证了领域专用适配对于专业性显微图像分割的必要性。

本文方法在 F1 分数上达到 0.863,显著优于所有基线方法。与最佳的 Mask R-CNN 相比, F1 提升了 0.269。具体而言,准确率提升 34.2%,召回率由 0.554 提升至 0.855。这一大幅提升表明,本文采用的分阶段策略是先利用 SAM 获取高召回候选,再通过双分支拒识和选择性补全提升质量,比端到端地从矩形框中学习分割更为有效。在 mIoU 方面,本文方法达到 0.809,同样优于 Mask R-CNN 和 YOLOv8-seg。这一结果验证了选择性掩码补全策略的有效性:掩码补全策略通过内缩框伪监督与掩码解码器微调,使 SAM 输出的掩码从粗粒度的候选区域逐步收敛至颗粒的真实轮廓,在框 IoU 指标上也实现了对填充框式监督方法的超越。综合 F1 和 mIoU 两项指标,本文方法在实例级匹配质量和掩码边界精度上均达到了当前最优水平。

4.3. 消融实验

为系统验证各模块的贡献,本节以 ViT-H 版本 SAM 为基础,在测试集上逐步叠加各个组件,消融实验结果如表 2 所示,效果如图 4 所示。

Table 2. Ablation study of each component on the test set

表 2. 各组件消融实验结果

配置	Precision	Recall	F1	mIoU
Grid-SAM	0.317	0.866	0.464	0.689
+ 面积初筛	0.803	0.866	0.833	0.692
+ DiscNet	0.875	0.825	0.849	0.696
+ ShapeDisc	0.872	0.850	0.861	0.696
+ 选择性掩码补全	0.871	0.855	0.863	0.809

Grid-SAM 原始输出呈现出极高的召回率与极低的准确率, F1 仅为 0.464。32 × 32 密集网格提示能够覆盖绝大部分真实显微颗粒,但同时也对每个颗粒内部、纹理区域和背景杂质生成了大量冗余和碎片化候选。加入面积初筛后,准确率从 0.317 跃升至 0.803,召回率维持 0.866 不变, F1 提升至 0.833。仅通过删除面积小于 2000 像素的候选,就使 F1 提升了 0.369。这一显著改善说明 Grid-SAM 产生的大量假阳性候选主要来自小碎片和纹理噪点,面积阈值作为一个简单且无参数的筛选条件,能够以几乎零成本过滤大部分噪声。同时,召回率未受任何影响,验证了阈值 2000 的设置低于真实目标尺寸分布的下界,

不会误删真实颗粒。

在外观 DiscNet 介入后, 准确率进一步从 0.803 提升至 0.875, 但召回率从 0.866 下降至 0.825, F1 提升至 0.849。外观分支能够有效识别并拒识面积较大但与真实颗粒纹理不一致的背景区域(如染色斑块、纤维团块), 因此准确率继续上升。召回率的下降则说明部分低对比度或边缘模糊的真实颗粒在外观特征上与负样本接近, 被外观分支误判为杂质。对于仅依赖局部 RGB 纹理的外观判别器而言, 这种误判难以完全避免。

融合 ShapeDisc 后, 召回率提升了 0.025, 准确率从 0.875 微降至 0.872, F1 提升至 0.861。形状分支的加入使得部分被外观分支误拒的低对比度真实颗粒得以恢复, 这些颗粒虽然纹理不显著, 但其轮廓具有典型的紧致性和规则性, 几何特征层面与杂质有明显的可分离性。准确率的微降则说明极少数背景杂质具有与真实颗粒相似的轮廓形状, 被形状分支错误地保留。两个分支在 logit 空间的加权融合($w = 4.5$)实现了信息互补, 最终 F1 达到 0.861。

选择性掩码补全的加入使 mIoU 提升了 0.113, F1 微升至 0.863。这一阶段不与拒识模块争夺候选的真假判断, 而是专注于改善已保留候选的掩码完整性。mIoU 的大幅提升表明, 内缩框伪监督微调后的掩码解码器能够有效感知颗粒轮廓, 使掩码从“覆盖大致的候选区域”向“贴合颗粒边界”方向改善。F1 的微升则来自两方面贡献: 部分原先 IoU 略低于 0.5 阈值的候选经精修后跨过阈值成为真阳性, 但也有少量候选因精修漂移而退化为假阴性, 后者的存在限制了 F1 的进一步提升空间。为进一步理解双分支拒识的互补机制, 表 3 给出了形状分支与外观分支的独立及联合性能。

Table 3. Performance comparison of individual branches and fusion

表 3. 形状分支与外观分支的独立及融合性能对比

判别方式	Precision	Recall	F1	mIoU
仅外观 DiscNet	0.875	0.825	0.849	0.696
仅 ShapeDisc	0.842	0.858	0.850	0.696
双分支融合	0.872	0.850	0.861	0.696

从表 3 可以看出, 外观分支偏向高准确率, 形状分支偏向高召回率, 二者呈现典型的“准确率 - 召回率互补”模式。融合后 F1 高于任一单独分支, 验证了双分支互补设计的有效性。

进一步分析两个分支 logit 输出之间的 Pearson 相关系数, 其值为 $r = 0.42$, 属于中等偏低的相关性水平。这表明外观和形状两个模态包含部分重叠的判别信息, 但同时也存在相当程度的独立判别能力。融合过程中, 外观分支预测概率处于不确定区间中的 23% 的判断错误被形状分支纠正, 这一比例与相关系数 $r = 0.42$ 所暗示的非共享信息量在数量级上一致。为验证 self-IoU 门控机制的必要性, 表 4 给出了有无门控的性能对比。

Table 4. Effect of self-IoU gating mechanism

表 4. Self-IoU 门控机制的消融实验

配置	Precision	Recall	F1	mIoU
无门控	0.861	0.845	0.853	0.751
Self-IoU 门控	0.871	0.855	0.863	0.809

若移除 self-IoU 门控, 对全部保留候选无条件应用精修, mIoU 从 0.809 骤降至 0.751 (下降 0.058), F1 从 0.863 降至 0.853。详细分析发现, 无条件精修导致 14 个原本正确的候选($\text{IoU} \geq 0.5$)因掩码漂移至相

邻颗粒或背景区域而退化为假阴性。门控机制使 mIoU 获得 0.058 的显著提升, 同时成功拦截了漂移案例, 在 F1 上也保持优势(+0.010)。

对 self-IoU 阈值在 0.5 至 0.7 范围内的参数扫描实验表明, 阈值为 0.6 时 F1 最优。阈值过低(如 0.5)时门控过于宽松, 无法有效拦截漂移案例; 阈值过高(如 0.7)时门控过于严格, 拒绝了大量有益的边界精修。0.6 在两者之间取得了最佳平衡, 既保留了大部分有效的精修, 又成功拦截了绝大多数漂移案例。

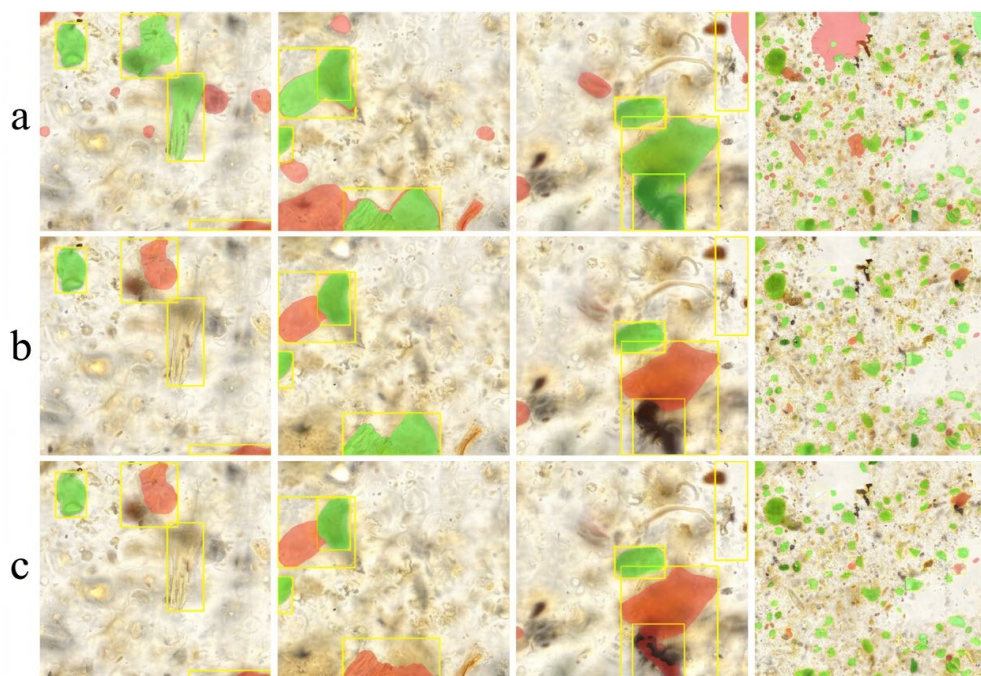


Figure 4. Ablation experiment visualization (a, b, and c show the effects after adding each module sequentially)
图 4. 消融实验可视化(a, b, c 分别为逐个模块加入后的效果)

5. 结论

本文针对中药粉末显微图像中实例分割面临的三个核心困难: 目标数量未知且分布密集、背景杂质与过分割碎片产生大量假阳性、矩形框标注缺乏像素级边界信息, 提出了一种分阶段的框监督实例分割框架。该框架依次通过高召回候选生成、形状-外观双分支拒识和选择性掩码补全三个阶段, 并且经过实验验证, 在仅使用矩形框标注的条件下实现了高质量的类别无关实例分割。在未来的研究中, 我们将探索的方向包括: 探索针对低对比度区域的专门增强策略, 如在 SAM 编码器前端引入对比度自适应模块; 将当前类别无关的分割框架与下游药材分类任务端到端联合优化; 利用多批次、多成像条件下的数据扩充, 进一步提高模型对不同制备工艺和显微设备的泛化能力; 探索主动学习或交互式标注策略, 以较低的人工成本获取部分像素级精标注, 突破当前框监督的性能上限。

参考文献

- [1] Zhao, Z., Liang, Z. and Ping, G. (2011) Macroscopic Identification of Chinese Medicinal Materials: Traditional Experiences and Modern Understanding. *Journal of Ethnopharmacology*, **134**, 556-564.
<https://doi.org/10.1016/j.jep.2011.01.018>
- [2] Miao, X.S., Cui, Q.Y., Wang, Z.Y., et al. (2017) Advance in Quality Assessment of Chinese Materia Medica Using Microscopic and Morphological Methods. *Chinese Journal of Natural Medicines*, **15**, 653-663.
[https://doi.org/10.1016/s1875-5364\(17\)30094-8](https://doi.org/10.1016/s1875-5364(17)30094-8)

- [3] Ichim, M.C., Häser, A. and Nick, P. (2020) Microscopic Authentication of Commercial Herbal Products in the Globalized Market: Potential and Limitations. *Frontiers in Pharmacology*, **11**, Article 876. <https://doi.org/10.3389/fphar.2020.00876>
- [4] Wang, H., Zou, H., Cheng, C., Wang, W., Wang, S., Wang, D., *et al.* (2023) Authentication of bistortae Rhizoma and Its Three Adulterants Based on Their Macroscopic Morphology and Microscopic Characteristics. *Microscopy Research and Technique*, **86**, 359-367. <https://doi.org/10.1002/jemt.24277>
- [5] Ma, Y., Zhong, Y., Su, Q., Xu, L., Song, H. and Wen, C. (2023) Study on Identification Algorithm of Traditional Chinese Medicinals Microscopic Image Based on Convolutional Neural Network. *Medicine*, **102**, e34085. <https://doi.org/10.1097/md.00000000000034085>
- [6] Zhu, X., Pang, G., He, X., Chen, Y. and Yu, Z. (2024) A Segmentation-Combination Data Augmentation Strategy and Dual Attention Mechanism for Accurate Chinese Herbal Medicine Microscopic Identification. *Frontiers in Plant Science*, **15**, Article 1442968. <https://doi.org/10.3389/fpls.2024.1442968>
- [7] Shelhamer, E., Long, J. and Darrell, T. (2017) Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 640-651. <https://doi.org/10.1109/tpami.2016.2572683>
- [8] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [9] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2020) Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 386-397. <https://doi.org/10.1109/tpami.2018.2844175>
- [10] Dai, J., He, K. and Sun, J. (2015) BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. 2015 *IEEE International Conference on Computer Vision (ICCV)*, Santiago, 7-13 December 2015, 1635-1643. <https://doi.org/10.1109/iccv.2015.191>
- [11] Tian, Z., Shen, C., Wang, X. and Chen, H. (2021) BoxInst: High-Performance Instance Segmentation with Box Annotations. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 5439-5448. <https://doi.org/10.1109/cvpr46437.2021.00540>
- [12] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
- [13] Mazurowski, M.A., Dong, H., Gu, H., Yang, J., Konz, N. and Zhang, Y. (2023) Segment Anything Model for Medical Image Analysis: An Experimental Study. *Medical Image Analysis*, **89**, Article 102918. <https://doi.org/10.1016/j.media.2023.102918>
- [14] Ma, J., He, Y., Li, F., Han, L., You, C. and Wang, B. (2024) Segment Anything in Medical Images. *Nature Communications*, **15**, Article No. 654. <https://doi.org/10.1038/s41467-024-44824-z>
- [15] Archit, A., Freckmann, L., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., *et al.* (2025) Segment Anything for Microscopy. *Nature Methods*, **22**, 579-591. <https://doi.org/10.1038/s41592-024-02580-4>
- [16] Schmidt, U., Weigert, M., Broaddus, C. and Myers, G. (2018) Cell Detection with Star-Convex Polygons. In: *Lecture Notes in Computer Science*, Springer, 265-273. https://doi.org/10.1007/978-3-030-00934-2_30
- [17] Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M. (2021) Cellpose: A Generalist Algorithm for Cellular Segmentation. *Nature Methods*, **18**, 100-106. <https://doi.org/10.1038/s41592-020-01018-x>
- [18] Marks, M., Israel, U., Dilip, R., Li, Q., Yu, C., Laubscher, E., *et al.* (2025) CellSAM: A Foundation Model for Cell Segmentation. *Nature Methods*, **22**, 2585-2593. <https://doi.org/10.1038/s41592-025-02879-w>
- [19] VandeLoo, A.D., Malta, N.J., Sangneriya, S., Aponte, E., van Zyl, C., Xu, D., *et al.* (2025) SAMCell: Generalized Label-Free Biological Cell Segmentation with Segment Anything. *PLOS ONE*, **20**, e0319532. <https://doi.org/10.1371/journal.pone.0319532>
- [20] Alam, M.S., Jackson, M., Lord, M. and Meijering, E. (2026) Cell Segmentation in Microscopy Images Using a Sam-Based U-Net Architecture and a Novel Dataset. *Computer Methods and Programs in Biomedicine*, **285**, Article 109470. <https://doi.org/10.1016/j.cmpb.2026.109470>
- [21] Cheng, T., Wang, X., Chen, S., Zhang, Q. and Liu, W. (2023) BoxTeacher: Exploring High-Quality Pseudo Labels for Weakly Supervised Instance Segmentation. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 3145-3154. <https://doi.org/10.1109/cvpr52729.2023.00307>
- [22] Wang, J. and Xia, B. (2024) Weakly Supervised Image Segmentation Beyond Tight Bounding Box Annotations. *Computers in Biology and Medicine*, **169**, Article 107913. <https://doi.org/10.1016/j.compbiomed.2023.107913>
- [23] Xu, G., Qian, X., Shao, H., Luo, J., Lu, W. and Zhang, Y. (2025) A Segment Anything Model-Guided and Match-Based Semi-Supervised Segmentation Framework for Medical Imaging. *Medical Physics*, **52**, 4513-4527. <https://doi.org/10.1002/mp.17785>

-
- [24] Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., *et al.* (2024) Segment Anything Model for Medical Images? *Medical Image Analysis*, **92**, Article 103061. <https://doi.org/10.1016/j.media.2023.103061>
 - [25] Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., *et al.* (2025) Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation. *Medical Image Analysis*, **102**, Article 103547. <https://doi.org/10.1016/j.media.2025.103547>
 - [26] Gyawali, R., Dhakal, A., Wang, L. and Cheng, J. (2024) CryoSegNet: Accurate Cryo-Em Protein Particle Picking by Integrating the Foundational AI Image Segmentation Model and Attention-Gated U-Net. *Briefings in Bioinformatics*, **25**, bbae282. <https://doi.org/10.1093/bib/bbae282>
 - [27] Gaillochet, M., Noori, M., Dastani, S., Desrosiers, C. and Lombaert, H. (2026) Prompt Learning with Bounding Box Constraints for Medical Image Segmentation. *IEEE Transactions on Biomedical Engineering*, **73**, 359-368. <https://doi.org/10.1109/tbme.2025.3582749>