

基于改进 YOLO11 的森林倒木图像实例分割^①



赵云鹏, 宋文龙, 莫冲, 王广来, 黄建平

(东北林业大学 计算机与控制工程学院, 哈尔滨 150040)

通信作者: 黄建平, E-mail: jphuang@nefu.edu.cn

摘要: 倒木是森林生态系统的重要组成部分, 其准确识别与分割是开展森林资源清查与动态监测的重要基础. 针对森林倒木图像中倒木目标呈现出形态多变、尺度差异显著及边界模糊等复杂特性, 提出一种基于改进 YOLO11 的倒木图像实例分割算法 DSM-YOLO. 针对倒木形状不规则问题, 引入了双动态任务对齐分割头 (dual dynamic task align head, DDTAH) 代替原分割头, 通过多个卷积层联合提取特征以提升任务之间的互补性, 有效增强网络的特征提取和特征表达能力, 从而提升分割精度. 针对倒木尺度变化较大的问题, 在骨干网络中使用共享膨胀卷积金字塔 (shared dilated convolution pyramid, SDCP) 模块, 通过膨胀卷积多尺度特征, 使模型具备更强的多尺度适应能力, 确保不同尺度倒木的分割精度. 针对倒木边缘与背景过渡模糊、边界定位不精确的问题, 在骨干网络中使用多尺度边缘注意力 (multi-scale edge attention, MSEA) 模块, 通过增强边缘特征, 使模型对倒木的边缘细节更敏感, 从而减少背景干扰, 提升倒木目标的边界分割效果. 实验结果表明, 本文提出模型的 $mAP50$ 比原始 YOLO11 模型提高了 4.3%. 与其他分割模型相比, 本文所提模型精度更高, 证明了 DSM-YOLO 的有效性, 为森林资源的精细调查提供技术支持.

关键词: YOLO11; 倒木; 实例分割; 森林调查

引用格式: 赵云鹏, 宋文龙, 莫冲, 王广来, 黄建平. 基于改进 YOLO11 的森林倒木图像实例分割. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10114.html>

Forest Fallen Tree Instance Segmentation Based on Improved YOLO11

ZHAO Yun-Peng, SONG Wen-Long, MO Chong, WANG Guang-Lai, HUANG Jian-Ping

(College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China)

Abstract: Fallen trees represent a vital component of forest ecosystems, and their accurate identification and segmentation is a crucial basis for forest resource inventory and dynamic monitoring. Due to the complex characteristics of fallen trees in images, such as variable morphology, significant scale differences and blurred boundaries, an instance segmentation algorithm named DSM-YOLO is proposed based on an improved YOLO11 framework. To address the irregular shapes of fallen trees, a dual dynamic task align head (DDTAH) is introduced to replace the original segmentation head. It jointly leverages multiple convolutional layers to extract features, improving inter-task complementarity and enhancing the network's capacity for feature extraction and expression, thereby improving segmentation accuracy. To handle large-scale variations of fallen trees, a shared dilated convolution pyramid (SDCP) module is incorporated into the backbone network. It utilizes dilated convolution to capture multi-scale features, enabling the model to better adapt to objects of different sizes and ensuring segmentation accuracy across scales. To mitigate fuzzy edge-background transitions and inaccurate boundary localization, a multi-scale edge attention (MSEA) module is integrated into the backbone. It enhances edge-related features, making the model more sensitive to detailed contours of fallen trees, which reduces background

① 收稿时间: 2025-09-23; 修改时间: 2025-10-14; 采用时间: 2025-10-29; csa 在线出版时间: 2026-03-09

interference and improves boundary segmentation. The experimental results indicate that the proposed model achieves a 4.3% improvement in $mAP50$ over the original YOLO11. It also outperforms other segmentation models in accuracy, demonstrating the effectiveness of DSM-YOLO and providing technical support for detailed forest resource surveys.

Key words: YOLO11; fallen tree; instance segmentation; forest survey

森林是地球上重要的自然资源,在木材生产、水源调节、碳固存和维持生物多样性等方面发挥重要作用,是实现生态系统可持续发展的重要基础^[1],倒木是指森林中倾斜度(偏离垂直方向)超过 45° ,粗头直径 $\geq 10\text{ cm}$,长度通常大于 1 m 的粗木质残体^[2].倒木是森林生态系统的重要组成部分,具有重要的生态功能和价值,如在温带森林中,多达 $1/3$ 的物种在其生命周期中依赖于倒木^[3].此外,倒木还可以分解产生无机物和有机物,供植物和土壤吸收,并为微生物群落提供了重要的栖息地和养分,这些过程不仅促进养分循环,还会影响碳的释放和储存^[4].

因此,高效准确地获取森林中倒木空间分布信息,为森林资源的精细调查、森林参数的高精度提取以及生物量的准确估算等提供可靠的数据支持,对生物多样性保护以及维持森林生态系统的健康具有重要意义.

传统人工森林调查方式在大范围调查中存在成本高、效率低、受森林地形条件限制等问题,难以满足现代森林调查的发展要求^[5].随着激光硬件技术的高速发展,激光雷达在森林调查中的应用日益广泛.目前,基于激光雷达进行倒木数据分析的研究,多以航空激光雷达(ALS)数据为基础展开.Lindberg等^[6]提出了一种从机载激光扫描点云数据中提取倒木的直线模板匹配算法,该算法生成的栅格数据直观地显示了线状元素,但由于定位误差和复杂的现场条件,结果很难与实地测量结果进行比较.Nyström等^[7]也应用了直线模板匹配,开发了一种利用两个海拔模型之间的差异检测倒木的方法,该方法无法检测不同长度和形状复杂的树木.Duan等^[8]开发了一种由粗到精的倒木精细提取技术,利用直线检测算法实现了树木个体的精细提取,完整性为 75.7% ,正确率为 92.5% ,但是提取范围较小且遮挡较少.针对上述方法的缺点,Polewski等^[9]提出了一种使用归一化切割算法从ALS点云中检测掉落茎的综合方法,该方法在复杂的森林场景中检测到 90% 的倒木,但会受地上植被覆盖率影响.而基于TLS识别倒木的研究,马振宇等^[10]通过TLS形态特征实现了倒

木的分割与特征提取,样本地区中倒木识别率为 100% ,但在不同倒木分布情境下仍需调整策略以确保分割准确性.

尽管激光雷达可以提供精确的倒木三维几何信息进行倒木点云分类和参数拟合,但其成本高,能耗高,数据获取复杂,且在复杂野外环境中存在安装与维护困难,限制了在大规模应用中的普及性.相较于激光雷达,RGB图像能够提供丰富的纹理与色彩信息,具有成本低、获取便捷、适应性强等优势.随着计算机视觉和深度学习的快速发展,RGB图像借助深度学习模型,可有效支持倒木识别与分割等任务.Redder等^[11]根据无人机正射影像提出了一种基于U-Net模型架构用于风吹倒木的语义分割和定位,达到了优于未预训练的模型,但在泛化能力方面仍存在提升空间.Polewski等^[12]提出了一种基于活动轮廓线的图像分割方法,在分割细长物体(如倒伏的茎)时,精确度和召回率均提高了 6% ,但图像中的阴影或遮挡部分会导致茎分裂成不相交的部分.Jiang等^[13]提出了一种优化的FCN-DenseNet算法,用于在复杂的温带森林环境中检测枯木(即枯立木和倒木),准确度可达 99% ,但对数据集要求很高,数据集准备时间很长.Nasimi等^[14]采用YOLOv8x-seg,结合目标检测和实例分割方法,进行被龙卷风影响地区倒下树木的分割,倒木识别成功率超过 73% ,但模型是为特定地区定制,泛化能力有限.

然而,上述研究尚未充分考虑实际森林中倒木目标的多样性与复杂性.在实际森林环境中,倒木具有明显的尺寸跨度大,姿态各异,也可能存在断裂、重叠、弯曲等形态差异,同时,倒木在空间分布上常与立木、植被等其他森林元素交错分布,且采集的图像中倒木易受枝叶、地形起伏或光照变化所造成的阴影、遮挡影响,还常包含大量颜色、纹理与倒木相似的干扰信息.这些问题在实际环境中是同时存在且相互交织的,使得倒木实例分割任务面临极高的挑战.现有方法虽在理想数据条件下取得了良好的分割效果,但针对上述多源干扰因素的问题仍未得到系统性

解决。

为此,本文以 YOLO11 框架为基础,提出面向森林复杂场景中的倒木分割改进方法,重点解决倒木尺寸多样,形状各异和边界模糊,且会被冗余信息干扰等挑战。具体改进主要包含以下 3 个方面。

(1) 针对倒木存在形状不规则影响分割精度的问题,引入 DDTAH 分割头代替原分割头,强化在多个卷积层中提取交互特征,结合了不同层次的特征信息,从而识别到更多的局部细节和全局信息,实现更精确的边界检测和分割。

(2) 针对倒木呈现出多尺度的问题,引入 SDCP 模块替换 SPPF 模块,SDCP 模块通过膨胀卷积来进行多尺度特征提取,低膨胀率卷积能够捕捉倒木的局部细节,高膨胀率卷积能够扩大感受野,捕捉到全局信息,从而有效提升了对于多尺度目标的实例分割性能。

(3) 针对倒木边缘与背景过渡模糊、边界定位不精确,受冗余信息干扰影响的问题,采用 MSEA 模块替

换原模型骨干网络中的 C3k2 模块。MSEA 模块通过增强边缘特征,使模型对倒木的边缘更敏感,从而更好地区分倒木与背景。降低复杂背景对倒木分割的影响,提升倒木实例分割的性能。

1 方法

1.1 YOLO11 算法

YOLO11 是 Ultralytics 公司于 2024 年推出的 YOLO 系列模型,支持多种计算机视觉任务,并在结构上相比 YOLOv8 进行了进一步优化升级,整合了 C3k2 与改进型 C2PSA 模块,并延续了 Backbone-Neck-Head 的 3 阶段架构,在提升模型检测精度的同时显著降低了计算成本,因而为包括森林倒木识别在内的多类任务提供了强有力的技术支持。

本文在 YOLO11 的基础上提出了一种高效的实例分割模型 DSM-YOLO。DSM-YOLO 模型的整体结构如图 1 所示。

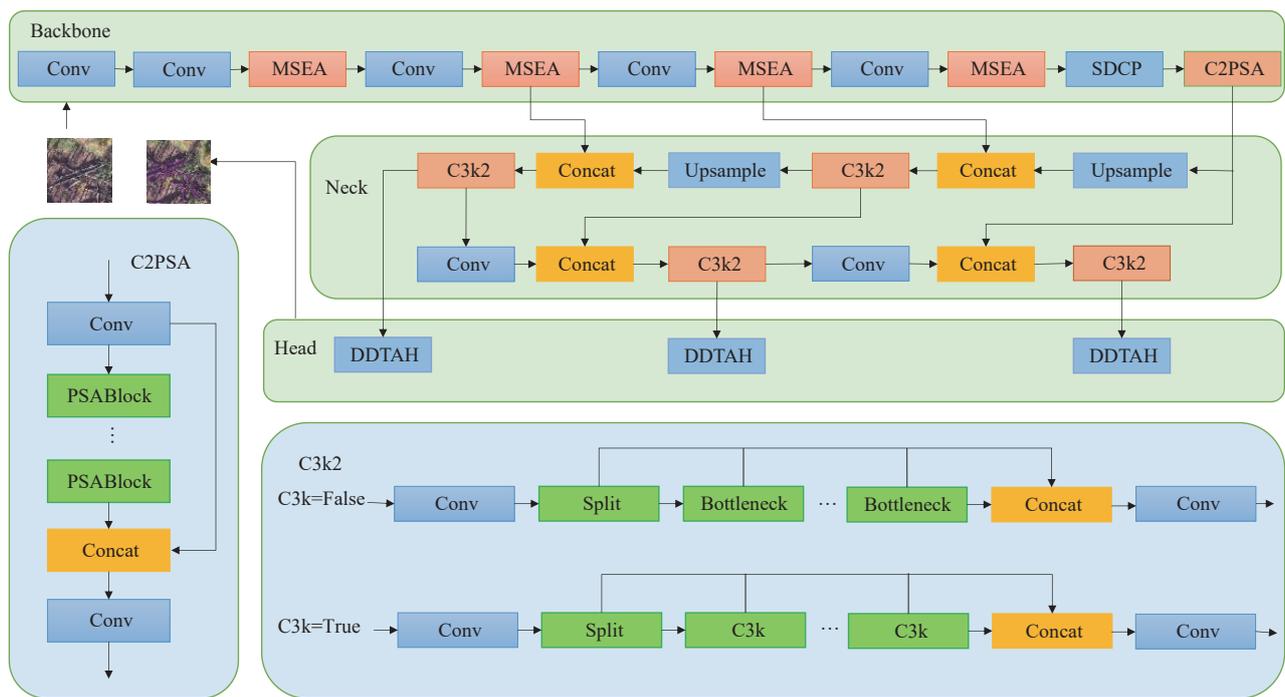


图 1 DSM-YOLO 网络结构图

1.2 DDTAH 模块

在倒木图像实例分割任务中,由于倒木形状不规则、姿态方向多变等问题,导致传统的分割头在分类和定位两个子任务之间缺乏有效信息协同,难以同时兼顾类别判别能力和空间定位精度,从而影响整体分割效果。为解决上述问题,本文引入双动态任务对齐分

割头 (dual dynamic task align head, DDTAH) 代替原分割头,DDTAH 结构如图 2 所示。首先采用共享卷积层 (shared convolution) 提取输入的多尺度特征,同时采用 GroupNorm (GN)^[15]进行归一化,提升分类和回归的稳定性。相比于传统分割头使用独立卷积的方式,DDTAH 通过共享基础卷积层,使分类、回归与方向预测等任

务在共享特征的基础上展开,有效减少参数量,使模型更轻量化.然后再把经过共享卷积处理后的特征进行拼接,融合后的特征能够同时捕获大目标和小目标

的信息,提高模型对不同尺度目标的泛化能力并且为任务解耦(task decoupling)提供更丰富的输入特征信息.

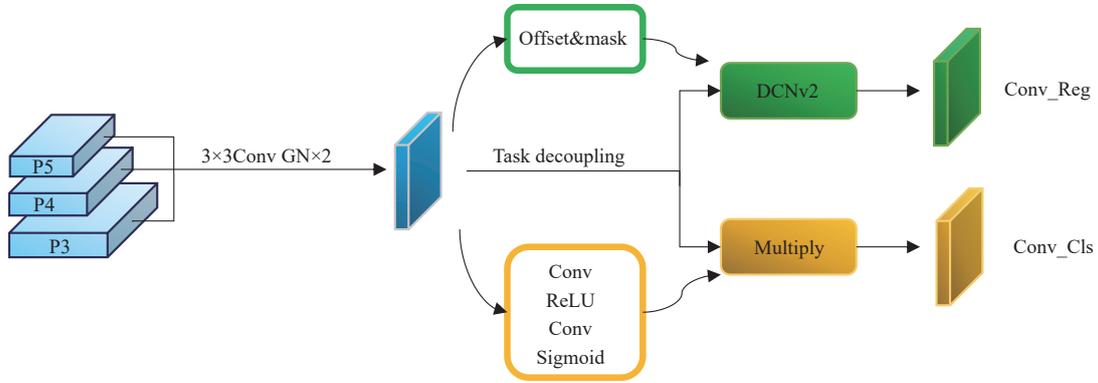


图2 DDTAH 结构图

特征通过任务解耦被分离为不同任务的特征流:定位任务特征和分类任务特征.传统的目标检测器头部通常使用独立的分类和定位分支,这会导致两个任务之间缺乏交互,信息无法共享.而 DDTAH 采用任务对齐策略^[16],在保留任务独立优化优势的同时,通过多个卷积层学习到联合特征,构建任务间的信息桥梁,使得分类与定位任务能共享部分信息,提升任务间的协同表达能力.同时采用了任务解耦,初步将任务特征进行分离,并进一步细化任务,使得分类和定位任务更加专注于各自的目标,同时保持一定的交互性.

在 DDTAH 定位分支中,引入了可变形卷积(deformable convolution v2, DCNv2),并通过交互特征生成 DCNv2 所需的偏移量(offset)与掩码(mask),以增强模型对目标边界的回归能力.相较于传统卷积在处理目标形变、姿态变化或复杂背景时定位精度下降的问题,DCNv2 能够根据目标的形变、尺度变化和复杂背景,动态调整卷积采样点的位置和权重,提高特征提取的灵活性与适应性,从而显著提升边界框回归的精度.

标准卷积的特征矩阵如式(1):

$$y(p) = \sum_{k=1}^K W_k \cdot x(p + p_k) \quad (1)$$

其中, W_k 表示卷积核权重, p_k 为第 k 个采样点的位置, K 为卷积核的采样点总数.

DCNv2 在每个采样位置引入了归一化后的权重因子以抑制无关区域的响应,DCNv2 特征表达式如式(2):

$$y(p) = \sum_{k=1}^K W_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (2)$$

其中, p_k 为对应的二维偏移矩阵, Δm_k 表示归一化后的注意力调节因子,相比传统卷积,DCNv2 在目标检测与实例分割任务中展现出更优的鲁棒性和精度表现.

分类分支在 DDTAH 中采用交互特征进行动态特征选择,通过 ReLU (rectified linear unit) 操作引入非线性变换,使得分类预测在不同层级特征下都能保持稳定,使模型能学习复杂的特征表示.激活函数(Sigmoid)则生成动态权重,使分类和定位特征的交互更平稳、可控.特征交互(Multiply)将分类和定位任务的特征进行融合,使得分类分支能够感知定位信息,增强任务对齐能力.

1.3 SDCP 模块

1.3.1 SPPF 模块

SPPF 是 YOLOv5 中对传统 SPP 结构的优化版本,其结构如图 3(a) 所示,与传统 SPP 使用多个不同尺寸的池化核不同, SPPF 采用多层小尺度(如 3x3)的最大池化操作,依次堆叠形成等效的大感受野,从而代替单一的大尺寸池化核.这种设计不仅减少了计算资源消耗,还保留了多尺度特征融合的能力.相较传统的 SPP 结构, SPPF 在不影响检测精度的情况下,实现了更快的推理速度.

然而,由于最大池化核的大小是固定的, SPPF 在应对不同尺度目标时缺乏自适应调节能力,限制了其在处理尺度变化较大的目标时的表现.

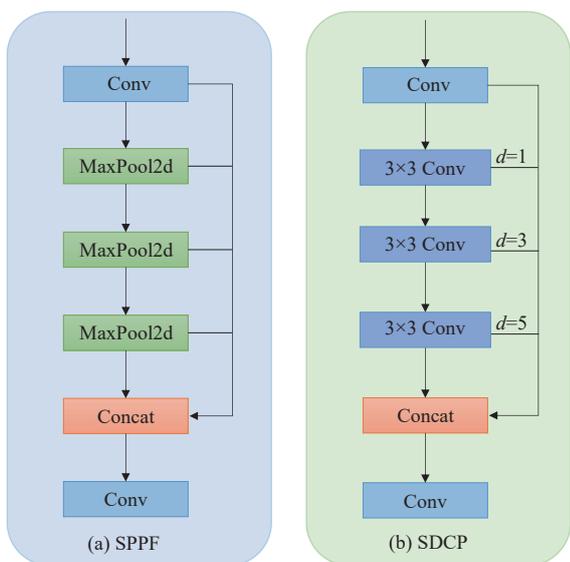


图3 SPPF与SDCP模块结构图

1.3.2 SDCP 模块

针对实际环境中的倒木图像存在不同尺度的倒木目标, 本文引入共享膨胀卷积金字塔 (shared dilated convolution pyramid, SDCP) 模块代替 SPPF 模块. SDCP 模块以更灵活的方式提取多尺度信息, 使得模型在特征提取阶段具有更好的多尺度感知能力. SDCP 模块在 SPPF 的基础上做出了重大改进, 主要是采用了膨胀卷积 (dilated convolution) 代替最大池化, 以更灵活的方式提取多尺度信息, 其结构如图 3(b) 所示.

SDCP 首先进行一次普通的 1×1 卷积 (Conv), 进行通道调整和初步特征提取. 然后使用 3 层 3×3 的膨胀卷积, 不同膨胀率 ($d=1, d=3, d=5$) 捕捉不同尺度的

特征.

$d=1$: 普通 3×3 卷积, 捕捉最基础的局部细节信息.

$d=3$: 适中的膨胀率, 能够在保持局部信息的同时, 引入更大的感受野.

$d=5$: 较大的膨胀率, 提供更宽的感受野, 捕捉全局信息.

最后将这些不同尺度的特征进行拼接, 得到最终的输出.

SDCP 通过不同膨胀率的卷积 ($d=1, d=3, d=5$) 提取不同尺度的特征, 低膨胀率捕捉局部细节, 高膨胀率捕捉全局信息, 使得局部特征和全局特征能够更好地结合, 这对捕捉图像中不同大小和不同特征非常有利. 使用卷积操作进行特征提取还能够捕捉更加细粒度的特征. 相比之下, SPPF 的池化操作可能会丢失一些细节信息. 卷积操作在特征提取时具有更高的灵活性和表达能力, 可以更好地捕捉图像中的细节和复杂模式. 相比于每个膨胀率使用独立的卷积层, 共享卷积层能够减少冗余, 提升模型效率, 减少了模型的存储和计算开销, 提升了计算效率.

1.4 MSEA 模块

在实际森林环境中, 倒木边缘通常与地面背景过度模糊, 并且可能与地面、植被等背景混杂在一起, 且地面、叶子、阴影等复杂环境因素会对模型的决策造成影响. 为了解决这个问题, 本文使用多尺度边缘注意 (multi-scale edge attention, MSEA) 模块代替 YOLO11 骨干网络中 C3k2 模块, 更精准地筛选任务相关的边缘特征, 其结构如图 4 所示.

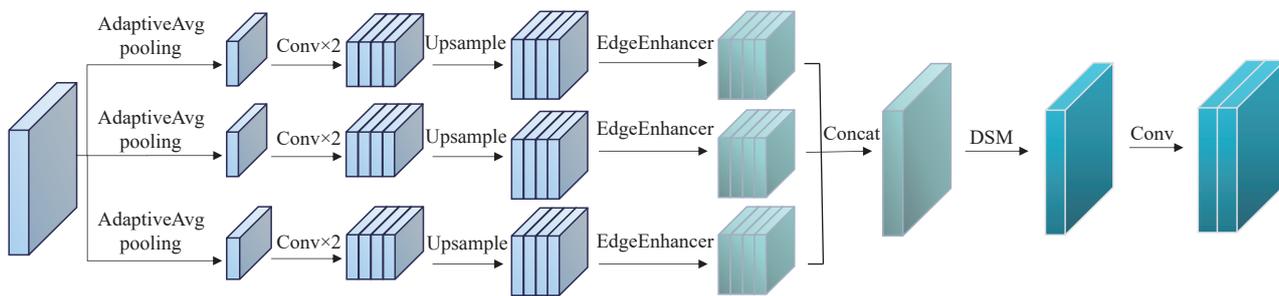


图4 MSEA模块结构图

MSEA 模块的整体目标是融合多尺度信息并突出边缘特征, 以提升模型在复杂场景下的表征能力. 主要包括多尺度特征提取、边缘增强 (EdgeEnhancer)、特征拼接 (Concat) 以及双域选择机制 (DSM) 这 4 个核心部分. 首先, 输入特征通过自适应平均池化层 (AdaptiveAvg

pooling) 进行不同尺度的池化操作. 该设计使得网络能够同时捕捉全局与局部的上下文信息, 确保不同感受野下的特征被提取和利用. 随后, 通过卷积层 (Conv) 对池化后的特征进行处理, 用于提取局部判别特征并保持通道数一致. 接着采用上采样 (Upsample) 操作对

特征进行对齐,使来自不同尺度的特征能够在空间尺寸上保持一致,为后续融合奠定基础.在此基础上,引入了 EdgeEnhancer 模块,其结构如图 5 所示.传统的多尺度融合往往容易忽视边缘与高频细节信息,而 EdgeEnhancer 的设计目标正是通过专门的特征提取与增强操作,将目标边缘区域的信息显式建模.首先,对输入特征图进行平均池化以获得平滑特征,并通过差分操作提取高频边缘信息.随后,利用卷积层对边缘特征进行进一步建模.最后,将增强的边缘特征与原始特征进行残差融合,从而在保持整体语义信息的同时显著强化边缘细节.它不仅能够缓解在下采样过程中边缘细节丢失的问题,还能够突出目标边界的语义特征,从而提升网络在分割任务中对边界的敏感度.具体而言,EdgeEnhancer 更强调梯度突变区域,使得模型在复杂背景或目标边缘模糊的情况下依然能够保持良好的判别能力.

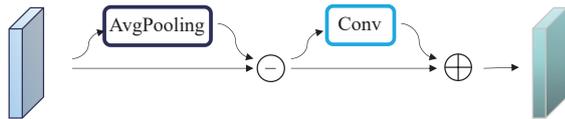


图 5 EdgeEnhancer 结构图

经过多个尺度的 EdgeEnhancer 模块处理后,特征在 Concat 层中被拼接融合,形成一个包含多尺度、边缘增强信息的综合特征表示.为了进一步筛选其中对任务最有价值的部分,引入了双域选择机制 (dual domain selection mechanism, DSM)^[17],其结构如图 6 所示. DSM 的核心思想是在空间域与频率域两个角度同时建模特征相关性.

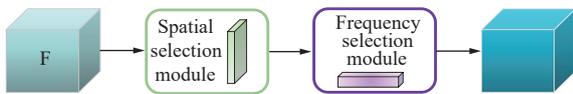


图 6 DSM 结构图

在空间域,DSM 能够捕捉图像中的边缘区域与目标结构信息,突出与任务高度相关的区域.

在频率域,DSM 更关注高频信号与细节特征,能够有效抑制背景噪声和低相关性信息.

通过这两方面的选择性筛选,DSM 实现了对多尺度特征的高效整合,使模型能够自适应地聚焦于与任务目标最相关的特征子集,从而在减少冗余信息的同时,显著提升整体分割性能.最后,通过卷积层进一步整合特征,生成最终输出,为后续的任务分支提供高质

量输入.

2 实验结果与分析

2.1 数据集

本文实验使用两个公开倒木数据集组成的混合数据集,数据集 1^[11]是通过无人机采集的遭受风暴严重影响的森林,数据集原始包含 560 棵倒伏山毛榉树的完整正射影像记录,经数据筛选(剔除模糊、遮挡严重及重复影像)后,最终保留 571 张有效图像;数据集 2^[14]是通过无人机采集的遭受龙卷风后森林的航空图像,树木类型包括落叶树和针叶树,包含 5 020 张图像.混合数据集最后包含 5 591 张倒木图像,部分图像如图 7 所示,其中,训练集有 3 473 张图片,验证集有 1 058 张图片,测试集有 1 060 张图片.



图 7 数据集部分倒木图像

本研究采用多种数据增强策略以提升模型的泛化能力.同时引入 YOLOX 中的 HSVRandomAug 方法^[18]对图像颜色进行扰动,以增强对光照和色彩变化的鲁棒性.所有增强过程均在训练时在线完成^[19],无需额外存储,有助于提升模型对真实环境的适应能力.

2.2 实验细节和评价指标

本文中所有的实验均在表 1 所示的硬件环境和软件环境中完成.

表 1 实验环境

名称	相关配置
操作系统	Windows 11系统
处理器	Intel(R) Core(TM) i5-13400F
显卡	NVIDIA RTX 3060
Python	3.8.20
深度学习框架	PyTorch 1.10.1

训练使用随机梯度下降法 (SGD),初始学习率为 0.01,动量因子为 0.937,正则化系数为 0.000 5,输入图像为 640×640 像素,批量 (batchsize) 为 32,训练轮次

(epochs) 为 300 次.

我们通过精确率 (precision, P)、召回率 (recall, R) 和平均精度均值 (mean average precision, mAP) 评估模型性能, 同时以参数量 (Params) 和计算量 (GFLOPS) 衡量模型复杂度. 参数量指的是构成深度学习模型的参数总数, 计算量指的是每秒 10 亿次的浮点运算次数. 精确率 P 、召回率 R 和平均精度 (mAP) 计算公式为:

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \int_0^1 P(R) dR \quad (5)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP \quad (6)$$

其中, TP 表示正确预测的数量; FP 指错误预测的数量; FN 为正确样本被错误预测的数量. $mAP50$ 表示在 50% 的 IoU 阈值下的平均精度均值, $mAP50-95$ 表示在 50%–95% 的 IoU 阈值下的平均精度均值.

2.3 消融实验

为了评估 DDTAH 模块、SDCP 模块以及 MSEA 模块对算法性能的贡献, 本文设计了消融实验, 以逐步添加不同模块的方式评估各自对模型性能的影响.

从表 2 可以看出, YOLO11 模型在逐步引入各改进模块后, 其倒木图像实例分割性能呈现明显上升趋势. 在第 1–8 组的对比中, 随着 DDTAH 模块、SDCP 模块以及 MSEA 模块的逐个加入, $mAP50$ 指标分别提高了 0.9、1.2 和 2.2 个百分点, 累计提升达到 4.3 个百分点, 充分体现出每项改进在提升模型分割精度中的关键作用.

表 2 消融实验表

实验编号	DDTAH	SDCP	MSEA	P	R	$mAP50$	$mAP50-95$	Params (M)	GFLOPS
1	×	×	×	0.682	0.597	0.652	0.286	2.83	10.2
2	√	×	×	0.683	0.599	0.661	0.296	2.37	11.9
3	×	√	×	0.684	0.603	0.660	0.294	2.98	11.2
4	×	×	√	0.685	0.607	0.663	0.298	2.93	11.4
5	√	√	×	0.687	0.608	0.673	0.305	2.52	11.9
6	√	×	√	0.685	0.608	0.668	0.299	2.36	12.1
7	×	√	√	0.684	0.607	0.667	0.297	2.86	12.6
8	√	√	√	0.692	0.659	0.695	0.322	2.51	12.1

注: √表示使用该模块, ×表示不使用该模块; 加粗表示最优结果

第 2 组实验中只引入 DDTAH 模块后, $mAP50$ 提升 0.9 个百分点、 $mAP50-95$ 提升了 1 个百分点、精确率提升了 0.1 个百分点、召回率提升了 0.2 个百分点, 并将参数量减少约 16.3%, GFLOPS 仅为 11.9, 这说明 DDTAH 模块能够有效增强不同层次特征之间的信息交互与融合能力, 从而提升模型对倒木图像的细节与整体结构的感知能力.

第 3 组实验中只引入 SDCP 模块. 实验结果表明, 模型在 $mAP50$ 和 $mAP50-95$ 上分别提升 0.8 和 0.8 个百分点, 尽管增加了结构复杂度, 但参数量仅增加 5.3%, GFLOPS 为 11.2, 保持了良好的精度与效率的平衡.

第 4 组实验只引入 MSEA 模块改进了 C3k2 结构. 在仅对骨干网络结构进行调整的前提下, 模型在 $mAP50$ 和 $mAP50-95$ 上分别提升 0.4 和 1.2 个百分点, 参数量仅提高 3.5%, GFLOPS 为 11.4, 验证了 MSEA 模块在低计算成本下的有效性.

第 5 组在第 2 组的基础上进一步加入 SDCP 模块. 相较于第 2 组, 精确率提升 0.4 个百分点, 召回率

提升 0.9 个百分点, $mAP50$ 与 $mAP50-95$ 分别提升 1.2 和 0.9 个百分点, 参数量降低 11.3%, GFLOPS 仅为 11.9. 表明 SDCP 通过多膨胀率卷积增强了模型对多尺度特征的提取能力, 既能捕捉倒木的局部细节, 又能感知全局语义信息, 从而有效缓解了传统卷积在多尺度表达上的不足.

为了进一步探索模块之间的互补性与协同效应, 本文还增加了两种“两两组合”实验 (第 6、7 组). 从结果看, 在第 6 组实验中, 虽然未使用 SDCP 模块, 但仍获得了 $mAP50$ 为 0.668、 $mAP50-95$ 为 0.299 的表现, 并将参数量减少约 16.6%, GFLOPS 仅增至 12.1, 表明 MSEA 有效缓解了由于尺度变化或纹理模糊所带来的边缘定位不准确问题, 这说明 MSEA 不仅在细节提取方面表现良好, 也具备一定的多尺度适应能力, 第 7 组实验中, 虽然未包含 DDTAH 模块, 但模型的 $mAP50$ 达到了 0.667, $mAP50-95$ 为 0.297, 相比基础模型提升明显, 参数量仅增加约 1%, GFLOPS 为 12.6, 表明 MSEA 与 SDCP 的结合能够同时加强多尺度感知和边

缘识别能力.

综上所述,在YOLO11模型中引入所有3种改进方法而构建的DSM-YOLO在 $mAP50$ 和 $mAP50-95$ 指标上都得到了最高值,模型精确率提升了1个百分点,召回率提升了6.2个百分点, $mAP50$ 提升4.3个百分点, $mAP50-95$ 提升3.6个百分点,参数量减少11.3%,

GFLOPS仅为12.1,证明了本文改进方法能够有效提升模型的实例分割性能.

为直观验证本文模型各部分改进的有效性,各模块改进前后的倒木分割模型效果对比图如图8所示.图中“A”为倒木形状不规则情况,“B”为倒木尺度变化情况,“C”为倒木边缘模糊情况.

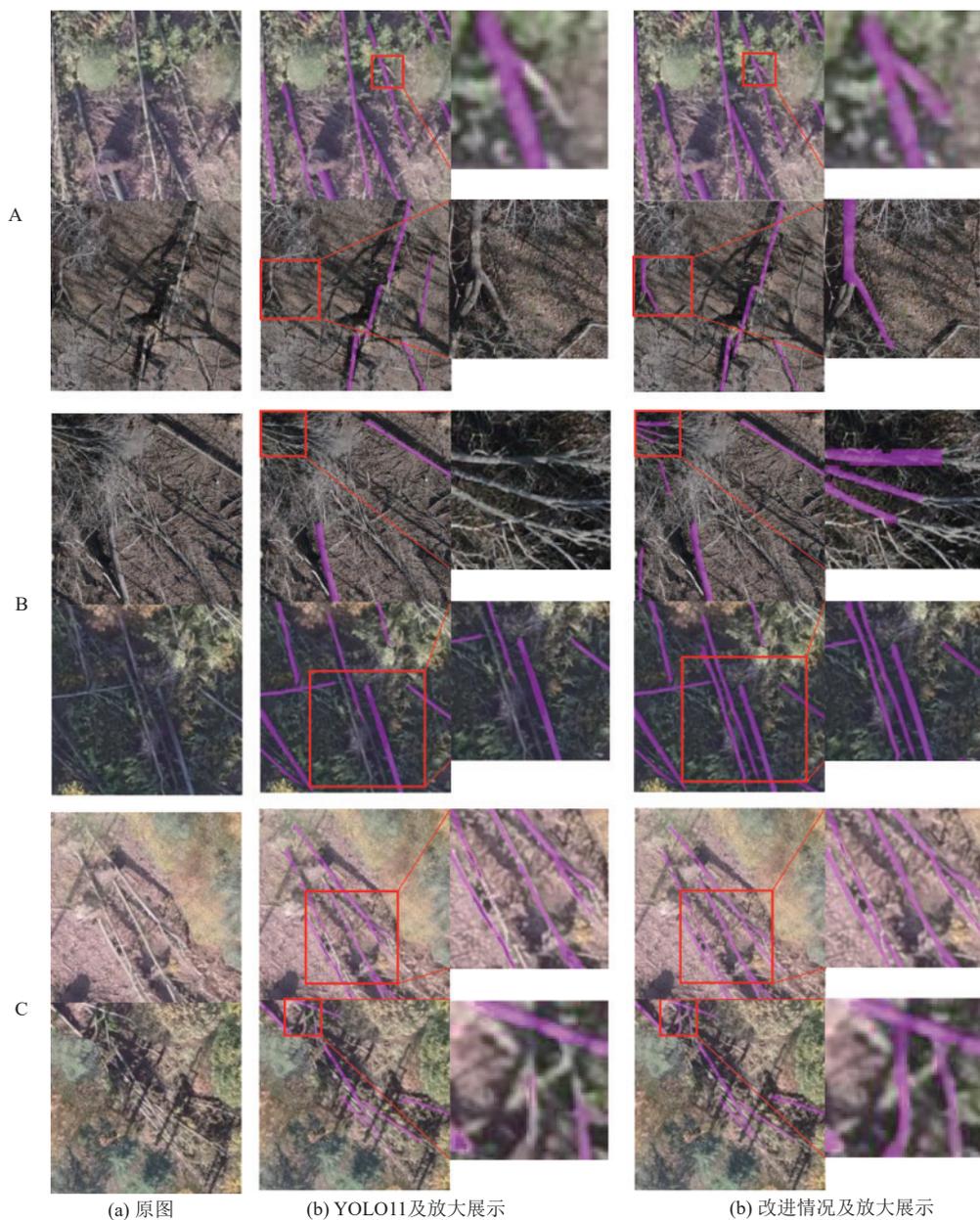


图8 改进前后实例分割可视化对比

图8中A行展示了倒木形状不规则的情况.在该场景中,原始YOLO11模型在面对细长弯曲、结构复杂的倒木形态时存在边界断裂、轮廓缺失的问题,尤

其是在目标末端或弯折处分割不完整.而改进后的DSM-YOLO能够更好地融合局部和全局信息,实现了更为完整和连续的分割. B行展示了倒木尺度变化大

的情况. YOLO11 对于目标尺度变化较大的倒木目标表现出显著的分割不一致性, 容易出现漏分的现象. 而 DSM-YOLO 在不同尺度的倒木目标分割上均表现出更强的适应性和稳定性. C 行针对的是倒木边缘与背景过渡模糊的复杂背景情况. 在这种场景下, YOLO11 往往难以准确定位倒木边缘, 与周围植被、地面纹理等元素发生混淆, 导致边界识别不准确. 而 DSM-YOLO 对倒木边缘更敏感, 显著减少了背景干扰, 有效提升了边界分割精度.

2.4 对比实验

2.4.1 DDTAH 模块有效性分析

本文通过实验评估 DDTAH 模块在处理更多和更细节的倒木图像信息方面能力的有效性. 表 3 为不同分割头的对比实验, 以 LSCD 模块和 EfficientHead 模块作为对比.

表 3 DDTAH 模块有效性分析

模型	P	R	$mAP50$	$mAP50-95$	Params (M)
LSCD	0.639	0.587	0.635	0.285	2.56
EfficientHead	0.663	0.585	0.653	0.29	2.56
DDTAH	0.683	0.599	0.661	0.296	2.37

从实验结果可知, 采用 DDTAH 模块后, 模型在多个评价指标上均展现出显著的性能提升. 在精确率方面 DDTAH 达到 0.683, 相比 EfficientHead 提高了 2 个百分点, 相比于 LSCD 提高 4.4 个百分点, 说明 DDTAH 更侧重于优化模型在精确性和全面性之间的平衡, 这得益于其任务对齐的分割头设计, 增强了分类和定位任务之间的交互, 使得更多目标能够被正确分割出来. 在召回率上, DDTAH 达到了 0.599, 明显高于 LSCD 的 0.587 和 EfficientHead 的 0.585, 分别提升 1.2 和 1.4 个百分点. 召回率的提升表明 DDTAH 在确保捕获所有正类目标方面更强, 减少了漏检情况. DDTAH 的 $mAP50$ 达到 0.661, 相比于 LSCD 的 0.635 和 EfficientHead 的 0.653 分别提高了 2.6 个百分点和 0.8 个百分点. $mAP50$ 衡量的是高置信度目标的检测能力, DDTAH 在此项指标上占据明显优势, 表明它在目标边界精确度上表现更优, 能够更准确地界定目标区域. 在 $mAP50-95$ 上, DDTAH 达到了 0.296, 相比于 LSCD 的 0.285 和 EfficientHead 的 0.29 分别提高了 1.1 和 0.6 个百分点. 这表明 DDTAH 在不同尺度目标的检测能力更强, 具备更好的泛化能力. 综上所述, DDTAH 模块显著增强了模型在倒木图像实例分割任务中的性能.

2.4.2 SDCP 模块有效性分析

本文通过对比两种常见的模块 LSKA 和 AIFI, 以验证 SDCP 模块在提升模型性能上的有效性. 如表 4 所示, 引入 SDCP 模块后, 原始模型 YOLO11 在多个评价指标上均实现了显著的性能提升. 在精确率方面, SDCP 模块的应用使得模型精度提升了 0.2 个百分点, 召回率也提升了 0.6 个百分点, 而在 $mAP50$ 指标上, 提升幅度达到了 0.8 个百分点, 同时 $mAP50-95$ 值也实现了 0.8 个百分点的增长. 这表明 SDCP 模块在提高模型对倒木图像实例分割的精度上具有显著优势, 尤其是在处理多尺度目标时, 模块利用多种膨胀率的卷积层来捕捉不同尺度的特征. 这对捕捉图像中不同大小和不同上下文的信息非常有利, 从而显著提升整体的分割精度.

表 4 SDCP 模块有效性分析

模型	P	R	$mAP50$	$mAP50-95$	Params (M)
LSKA	0.676	0.601	0.652	0.291	3.10
AIFI	0.679	0.6	0.653	0.288	3.46
SDCP	0.684	0.603	0.660	0.294	2.98

LSKA 和 AIFI 模块的 $mAP50$ 值分别为 0.652 和 0.653, 相较于 YOLO11 原始模型的 0.652 提升较小, 在 $mAP50-95$ 指标上较 YOLO11 的 0.286 分别提升 0.5 个百分点和 0.2 个百分点, 整体提升幅度均不及 SDCP 模块, 且 SDCP 模块的参数量相比 LSKA 模块和 AIFI 模块更低, 进一步证明了该模块在计算效率上的优化.

综上所述, SDCP 模块在提升倒木图像实例分割任务中的性能方面展现出了更为突出的优势, 其通过多尺度特征提取和高效的通道变换, 显著增强了模型的分割能力.

2.4.3 MSEA 模块有效性分析

为了获得 MSEA 模块最好的改进效果, 分别将其替换到主干网络、颈部网络以及整个网络中的 C3k2 模块, 结果如表 5 所示.

表 5 MSEA 实验结果

模型	$mAP50$	$mAP50-95$	Params (M)	GFLOPS
YOLO11	0.652	0.286	2.83	10.2
MSEA (主干)	0.663	0.298	2.93	11.4
MSEA (颈部)	0.651	0.288	2.97	11.4
MSEA (全部)	0.661	0.292	2.96	11.6

由表 5 可知, 相较于仅在颈部网络中单独应用 MSEA 模块, 或同时应用于主干网络和颈部网络中, 将

MSEA 模块应用于主干网络中,能够显著提升模型在实例分割任务上的性能表现,从而实现更高的分割精度.与 YOLO11 原始模型相比,MSEA 模块在主干网络应用时, $mAP50$ 提升了 1.1 个百分点,从 0.652 提升至 0.663, $mAP50-95$ 提升了 1.2 个百分点.这表明,MSEA 模块在处理倒木与背景边界的分割时,能够有效减少复杂背景(如树木、岩石等)的干扰.在面对复杂背景和遮挡的挑战时,MSEA 模块通过增强边缘特征和多尺度信息聚合,提升了模型的鲁棒性和精度.相比之下,MSEA 仅应用于颈部网络时, $mAP50-95$ 下降了 1 个百分点,是由于颈部特征聚合阶段对局部边缘信息利用效率不足所致.而当 MSEA 同时作用于主干和颈部网络时,参数量和计算量略有增加,分割性能反而下降,这归因于边缘特征过度提取所引起的信息冗余和干扰,降低了模型的判别能力.

虽然 MSEA 模块在主干网络中单独引入时提升幅度有限,但其在完整 DSM-YOLO 模型中与 DDTAH、SDCP 模块联合使用时,整体分割性能实现了显著提升.这表明 MSEA 模块在边缘特征增强方面的作用具有一定的结构依赖性,在模型完整架构中的边缘增强能力与其他模块形成互补,在细化倒木边界、抑制背景干扰方面起到关键作用,其优势需通过特征融合与解码过程进一步发挥,才能有效提升模型对复杂背景中倒木目标的识别精度.

2.4.4 不同模型对比实验

为了更好地证明改进算法的实例分割效果,在同样的数据集上将改进算法与当前主流的实例分割算法 SOLOv2^[20]、Mask R-CNN^[21]、Mask Scoring R-CNN^[22]、PointRend^[23]、QueryInst^[24]、YOLO12^[25]、YOLO13^[26] 以及与文献[15]中使用的 YOLOv8x-seg 模型进行比较.实验结果如表 6 所示,加粗表示最优结果.

表 6 对比实验表

Method	$mAP50$	$mAP50-95$	Params (M)	GFLOPS
SOLOv2	0.566	0.214	65.585	227.0
Mask R-CNN	0.429	0.138	62.963	233.0
Mask Scoring R-CNN	0.453	0.157	118.000	370.0
PointRend	0.616	0.261	59.810	134.0
QueryInst	0.608	0.258	216.000	495.0
YOLOv8x-seg	0.69	0.315	717.216	343.7
YOLO12	0.643	0.282	2.761	9.7
YOLO13	0.645	0.282	2.701	10.0
YOLO11	0.652	0.286	2.835	10.2
DSM-YOLO	0.695	0.322	2.510	12.1

根据表 6 的数据,本文改进的算法显著提高了 YOLO11 模型的实例分割精度.本文改进算法的计算量小于除 YOLO11、YOLO12 和 YOLO13 外的所有实例分割算法,与 9 种实例分割算法相比, $mAP50$ 分别提高了 12.9、26.6、24.2、7.9、8.7、0.5、5.2、5.0、4.3 个百分点. $mAP50-95$ 也达到了最优结果.说明本文算法不仅在大目标分割上表现突出,同时在小目标和复杂纹理场景下也具备更强的鲁棒性.相较于基线模型,本文方法在 $mAP50$ 上提升 4.3 个百分点,在 $mAP50-95$ 上提升 3.6 个百分点,性能的提升主要得益于 DDTAH、SDCP 和 MSEA 模块的引入,它们在特征共享、边缘增强和多尺度特征融合方面有效提升了特征表达能力,使模型能更准确地分割细长、纹理复杂的倒木区域.而 Mask R-CNN 与 Cascade Mask R-CNN 这两类算法依赖候选框生成与多阶段推理,虽然在通用分割任务中具有较高精度,但在倒木数据集中,由于目标形态细长、背景复杂,其区域提议阶段难以准确捕捉边界,导致 $mAP50$ 分别降低 26.6 和 24.2 个百分点.同时,它们的 GFLOPS 远高于本文算法,计算代价大,难以在边缘设备上部署.SOLOv2 在通用场景表现优秀,但由于未针对倒木细节特征设计,其在本数据集上 $mAP50$ 比本文低 12.9 个百分点.PointRend 和 QueryInst 同样未能充分利用倒木的线性与方向性信息,分割边缘存在断裂问题, $mAP50$ 分别降低 7.9 和 8.7 个百分点.YOLOv8x-seg 在分割精度方面获得了除本文所提算法外的最佳效果,但其计算量和参数量远远高于本文模型,且由于 YOLOv8x 模型未从倒木分割任务特性角度进行改进,因此在处理倒木分割任务时,实际效果不及本文所提算法.

此外,本文方法计算量仅略高于 YOLO11,远低于其他模型,计算复杂度较低,较好地平衡了模型精度和模型复杂度,更适合部署于资源受限的边缘设备,如移动终端或无人机平台,具备较高的实际应用价值.然而,尽管模型整体性能表现优异,其结构设计仍存在一定局限性.例如,DSM-YOLO 的精度提升主要依赖于多模块的叠加增强,这在训练阶段可能导致特征冗余、优化难度增加及微弱改进效果不易独立量化.在消融实验中也可观察到,当 MSEA 模块单独加入模型时,其性能增益较小,表明该模块更适合与其他结构联合使用以发挥协同优势.此外,尽管引入多模块增强特征表达能力,但模型的结构仍基于 YOLO11 的主干设计,并未

在更深层次(如 Transformer 编码器结构)实现系统优化. 因此, 可能在面对极端复杂场景中结构模糊或特征弱化的倒木目标时表达能力不足. 这需要在未来的工作中继续研究和优化, 以适应更广泛的实际应用需求.

2.5 结果可视化分析

为直观展示改进效果, 本文展示不同算法和本文改进算法 DSM-YOLO 的实例分割结果, 如图 9 所示.

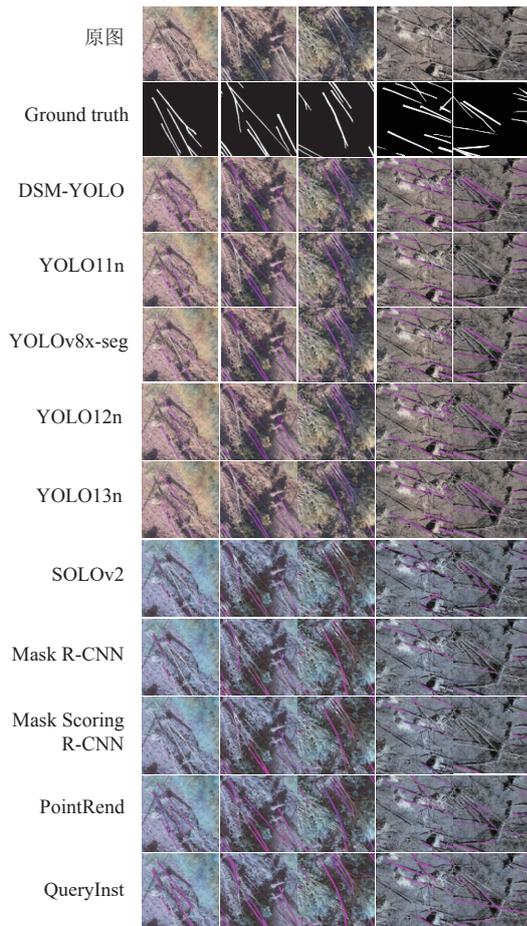


图 9 不同算法分割效果对比

从图 9 中可以明显看出, DSM-YOLO 在倒木边界的识别上最清晰准确, 能够较完整地覆盖目标区域并保持良好的边界连续性, 尤其是在倒木较细、被植被遮挡、背景复杂的场景中仍能保持稳定的分割性能. 相比之下, YOLO11 虽能较好地检测到倒木目标, 但会出现漏检, 识别不完整的问题. Mask R-CNN 和 Mask Scoring R-CNN 模型在处理森林高复杂度背景时的能力明显不足, 普遍存在漏检、误检现象, 有目标轮廓破碎、完整性差、边界不连续的问题. PointRend、QueryInst、YOLOv8x-seg、YOLO12n 和 YOLO13n 在边缘细节上

有所提升, 但对于形状、背景复杂和细长结构的倒木目标, 仍未能实现精确的分割.

综上所述, DSM-YOLO 展现出了较强的目标感知能力和边界处理能力, 尤其在处理森林场景中遮挡严重、目标密集、边缘模糊等难点问题上效果显著优于其他算法, 验证了本文提出的结构改进在提升倒木实例分割性能方面的有效性和鲁棒性.

为进一步验证本文的改进在不同场景的适用性, 本文选用不同于本文所用的两种数据集场景的倒木图片来进行检验, 如图 10 所示. 可以看出, 改进后的模型相比基线模型效果有明显提升, 图中用圆圈标记出原始模型和改进模型分割倒木图像的差异. 在基础模型出现漏检或误检情况时, 改进后的模型都能准确地分割出目标, 能够较好地复杂环境下完成分割任务.



图 10 对比结果

3 结论与展望

针对现有实例分割算法在森林复杂场景中对倒木目标分割精度不足的问题, 本文基于 YOLO11 算法提出了一种高效的倒木图像实例分割方法——DSM-YOLO.

本文的主要改进体现在 3 个方面: 首先, 针对倒木存在形状不规则影响分割精度的问题, 引入 DDTAH 分割头代替原分割头, 识别到更多的局部细节和全局信息, 实现了更精确的边界检测和分割; 其次, 针对倒木目标中呈现出多尺度目标导致实例分割精度受限的问题, 引入 SDCP 模块替换 SPPF 模块, 通过膨胀卷积来进行多尺度特征提取, 从而有效提升了对于多尺度目标的实例分割性能; 最后, 针对倒木边缘与背景过渡模糊、边界定位不精确, 背景冗余信息干扰对分割精度的影响, 采用 MSEA 模块替换原模型的骨干网络中的 C3k2 模块, 使模型更敏感于倒木的边缘, 显著降低了复杂背景对倒木分割的负面影响, 进而提升了倒木实例分割的性能. 实验结果表明, DSM-YOLO 在真实森林倒木数据集上取得了显著优于原始 YOLO11 和主

流分割算法的性能,整体模型在保持较高推理速度的同时,实现了精度与效率的良好平衡,具备良好的应用潜力.本研究不仅提升了复杂森林环境下倒木识别与分割的精度,也为推动基于无人机的森林调查方法提供了一种可行且高效的技术路径.尽管如此,本文研究仍有进一步发展的空间.未来,将继续优化模型结构和训练策略,旨在使 DSM-YOLO 能够处理更加复杂的森林环境(遮挡,重叠等)、处理多种数据源,以实现更强的跨场景、跨数据集的泛化性能.

参考文献

- 1 Watson JEM, Evans T, Venter O, *et al.* The exceptional value of intact forest ecosystems. *Nature Ecology & Evolution*, 2018, 2(4): 599–610. [doi: [10.1038/s41559-018-0490-x](https://doi.org/10.1038/s41559-018-0490-x)]
- 2 Harmon ME, Sexton J. Guidelines for Measurements of Woody Detritus in Forest Ecosystems. Technical Report 4-1996. Seattle: University of Washington, 1996.
- 3 Müller J, Büttler R. A review of habitat thresholds for dead wood: A baseline for management recommendations in European forests. *European Journal of Forest Research*, 2010, 129(6): 981–992. [doi: [10.1007/s10342-010-0400-5](https://doi.org/10.1007/s10342-010-0400-5)]
- 4 Carmona MR, Armesto JJ, Aravena JC, *et al.* Coarse woody debris biomass in successional and primary temperate forests in Chiloé Island, Chile. *Forest Ecology and Management*, 2002, 164(1-3): 265–275. [doi: [10.1016/S0378-1127\(01\)00602-8](https://doi.org/10.1016/S0378-1127(01)00602-8)]
- 5 Koch B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010, 65(6): 581–590. [doi: [10.1016/j.isprsjprs.2010.09.001](https://doi.org/10.1016/j.isprsjprs.2010.09.001)]
- 6 Lindberg E, Hollaus M, Mücke W, *et al.* Detection of lying tree stems from airborne laser scanning data using a line template matching algorithm. *Proceedings of the 2013 ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Antalya, 2013. 169–174. [doi: [10.5194/isprsannals-II-5-W2-169-2013](https://doi.org/10.5194/isprsannals-II-5-W2-169-2013)]
- 7 Nyström M, Holmgren J, Fransson JES, *et al.* Detection of windthrown trees using airborne laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 30: 21–29. [doi: [10.1016/j.jag.2014.01.012](https://doi.org/10.1016/j.jag.2014.01.012)]
- 8 Duan FZ, Wan YC, Deng L. A novel approach for coarse-to-fine windthrown tree extraction based on unmanned aerial vehicle images. *Remote Sensing*, 2017, 9(4): 306. [doi: [10.3390/rs9040306](https://doi.org/10.3390/rs9040306)]
- 9 Polewski P, Yao W, Heurich M, *et al.* Detection of fallen trees in ALS point clouds using a Normalized Cut approach trained by simulation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 105: 252–271. [doi: [10.1016/j.isprsjprs.2015.01.010](https://doi.org/10.1016/j.isprsjprs.2015.01.010)]
- 10 马振宇, 庞勇, 李增元, 等. 地基激光雷达森林近地面点云精细分类与倒木提取. *遥感学报*, 2019, 23(4): 743–755. [doi: [10.11834/jrs.20197383](https://doi.org/10.11834/jrs.20197383)]
- 11 Reder S, Mund JP, Albert N, *et al.* Detection of windthrown tree stems on UAV-orthomosaics using U-Net convolutional networks. *Remote Sensing*, 2021, 14(1): 75. [doi: [10.3390/rs14010075](https://doi.org/10.3390/rs14010075)]
- 12 Polewski P, Shelton J, Yao W, *et al.* Instance segmentation of fallen trees in aerial color infrared imagery using active multi-contour evolution with fully convolutional network-based intensity priors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 178: 297–313. [doi: [10.1016/j.isprsjprs.2021.06.016](https://doi.org/10.1016/j.isprsjprs.2021.06.016)]
- 13 Jiang S, Yao W, Heurich M. Dead wood detection based on semantic segmentation of VHR aerial CIR imagery using optimized FCN-Densenet. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Munich: ISPRS, 2019. 127–133. [doi: [10.5194/isprs-archives-XLII-2-W16-127-2019](https://doi.org/10.5194/isprs-archives-XLII-2-W16-127-2019)]
- 14 Nasimi M, Wood RL. Using deep learning and advanced image processing for the automated estimation of tornado-induced treefall. *Remote Sensing*, 2024, 16(7): 1130. [doi: [10.3390/rs16071130](https://doi.org/10.3390/rs16071130)]
- 15 Tian Z, Shen CH, Chen H, *et al.* FCOS: Fully convolutional one-stage object detection. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 9626–9635. [doi: [10.1109/ICCV.2019.00972](https://doi.org/10.1109/ICCV.2019.00972)]
- 16 Feng CJ, Zhong YJ, Gao Y, *et al.* TOOD: Task-aligned one-stage object detection. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 3490–3499. [doi: [10.1109/ICCV48922.2021.00349](https://doi.org/10.1109/ICCV48922.2021.00349)]
- 17 Cui YN, Ren WQ, Cao XC, *et al.* Focal network for image restoration. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 12955–12965. [doi: [10.1109/ICCV51070.2023.01195](https://doi.org/10.1109/ICCV51070.2023.01195)]
- 18 Qiu ZR, Rong SY, Ye LK. YOLF-ShipNet: Improved RetinaNet with pyramid vision Transformer. *International Journal of Computational Intelligence Systems*, 2023, 16(1):

58. [doi: [10.1007/s44196-023-00235-4](https://doi.org/10.1007/s44196-023-00235-4)]
- 19 Athalye C, Arnaout R. Domain-guided data augmentation for deep learning on medical imaging. *PLoS One*, 2023, 18(3): e0282532. [doi: [10.1371/journal.pone.0282532](https://doi.org/10.1371/journal.pone.0282532)]
- 20 Wang XL, Zhang RF, Kong T, *et al.* SOLOv2: Dynamic and fast instance segmentation. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 1487.
- 21 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2980–2988. [doi: [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322)]
- 22 Huang ZJ, Huang LC, Gong YC, *et al.* Mask scoring R-CNN. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 6402–6411. [doi: [10.1109/CVPR.2019.00657](https://doi.org/10.1109/CVPR.2019.00657)]
- 23 Kirillov A, Wu YX, He KM, *et al.* PointRend: Image segmentation as rendering. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 9796–9805. [doi: [10.1109/CVPR42600.2020.00982](https://doi.org/10.1109/CVPR42600.2020.00982)]
- 24 Fang YX, Yang SS, Wang XG, *et al.* Instances as queries. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*. Montreal: IEEE, 2021. 6890–6899. [doi: [10.1109/ICCV48922.2021.00683](https://doi.org/10.1109/ICCV48922.2021.00683)]
- 25 Tian YJ, Ye QX, Doermann D. YOLOv12: Attention-centric real-time object detectors. *arXiv:2502.12524*, 2025.
- 26 Lei MQ, Li SQ, Wu YH, *et al.* YOLOv13: Real-time object detection with hypergraph-enhanced adaptive visual perception. *arXiv:2506.17733*, 2025.

(校对责编: 张重毅)