

DOI:10.12171/j.1000-1522.20240145

# SSFYOLO: 一种面向复杂森林场景的树干检测算法

杨 灿 范习健 张九于

(南京林业大学信息科学技术学院, 人工智能学院, 江苏 南京 210037)

**摘要:**【目的】因为在复杂环境下树干目标尺寸差距大且易受遮挡, 所以树干检测容易出现漏检、错检等问题。为有效解决这个问题, 提出一种基于单阶段目标检测框架的树干检测算法 SSFYOLO。【方法】首先, 设计了空间感知网络模块 SAM。SAM 模块能够高效处理多尺度和多分辨率的特征信息, 在保证计算效能的同时, 实现对各类特征的充分整合与精确提取, 提高目标检测的准确性和效率。其次, 设计多尺度特征增强自适应网 FastScaleNet, 用于替代 YOLO 模型中的 C2f 结构。FastScaleNet 通过更为精细的多尺度特征融合与优化, 并且利用跳跃连接和特征融合策略, 有效保留不同层次的特征信息, 增强模型对不同尺度目标的表达能力, 提升模型的稳健性和广泛适用性。最后, 引入加权 IoU(WIoU)机制, 实现对小目标损失权重的动态优化。WIoU 机制根据目标尺寸的不同, 动态调整损失权重, 使模型在面对小尺寸目标时, 能够灵活调整参数, 从而灵活适应不同尺寸目标的检测需求, 进一步提高小目标检测的准确性和鲁棒性。【结果】对复杂场景下树干数据集进行检测实验, 与主流检测算法 YOLOv8 相比, SSFYOLO 算法在缩小参数数量的同时, 具有更好的检测精度, 其参数量减少了 20%, 平均精度均值(mAP)和召回率分别提升了 1.6 和 0.7 个百分点。【结论】本研究设计了面向复杂森林场景的树干检测算法 SSFYOLO。SSFYOLO 算法在复杂环境树干检测中表现出色, 具有广阔的应用前景。

**关键词:** 目标检测; 深度学习; 深度卷积神经网络; 特征提取; 计算机视觉; 树干检测; 复杂场景

**中图分类号:** TP751.1 **文献标志码:** A **文章编号:** 1000-1522(2025)02-0132-11

**引文格式:** 杨灿, 范习健, 张九于. SSFYOLO: 一种面向复杂森林场景的树干检测算法 [J]. 北京林业大学学报, 2025, 47(2): 132-142. Yang Can, Fan Xijian, Zhang Jiuyu. SSFYOLO: a trunk detection algorithm for complex forest scenarios[J]. Journal of Beijing Forestry University, 2025, 47(2): 132-142.

## SSFYOLO: a trunk detection algorithm for complex forest scenarios

Yang Can Fan Xijian Zhang Jiuyu

(College of Information Science and Technology & Artificial Intelligence,  
Nanjing Forestry University, Nanjing 210037, Jiangsu, China)

**Abstract:** [Objective] Due to the large size variation and frequent occlusion of trunks in complex environments, trunk detection is susceptible to issues such as missed detections and false positives. To effectively solve these problems, a trunk detection network named SSFYOLO, based on a single-stage target detection framework, was proposed. [Method] Firstly, a spatial awareness module (SAM) was designed. The SAM module can efficiently process multi-scale and multi-resolution feature information, ensuring computational efficiency while fully integrating and accurately extracting various features, thereby improving the accuracy and efficiency of target detection. Secondly, a multi-scale feature enhancement adaptive network (FastScaleNet) was designed to replace the C2f structure in the YOLO model. By refining multi-scale feature fusion and optimization, and utilizing skip connections and feature fusion strategies, FastScaleNet effectively retained feature information at different levels, enhanced the model's ability to express targets of different scales, and improved the model's robustness and versatility. Finally, the WIoU

收稿日期: 2024-05-11 修回日期: 2024-12-15

基金项目: 自然资源部重点实验室开放课题(NRMSSHR2023Y-07)。

第一作者: 杨灿。主要研究方向: 计算机视觉。Email: yangcan20000401@163.com 地址: 210037 江苏省南京市龙蟠路 159 号南京林业大学信息科学技术学院、人工智能学院。

责任作者: 范习健, 博士, 副教授。主要研究方向: 计算机视觉、智慧林业。Email: xijian.fan@njfu.edu.cn 地址: 同上。

本刊网址: <http://j.bjfu.edu.cn>; <http://journal.bjfu.edu.cn>

mechanism was introduced to dynamically optimize the loss weights for small targets. Depending on the target size, the WIoU mechanism dynamically adjusted the loss weights, allowing the model to flexibly adjust parameters when facing small targets. This enabled the model to adapt to the detection requirements of targets of different sizes, further improving the accuracy and robustness of small object detection.

**[Result]** Detection experiments were conducted on a trunk dataset in complex scenarios. Compared with the mainstream detection algorithm YOLOv8, the proposed SSFYOLO algorithm achieved better detection accuracy with reduced parameter volume. Specifically, the parameter volume was reduced by 20%, mAP and the recall rate represented improvements of 1.6 and 0.7 percentage points, respectively.

**[Conclusion]** This study has designed a tree trunk detection algorithm named SSFYOLO, which is specifically tailored for complex forest scenarios. The proposed algorithm demonstrates excellent performance in trunk detection in complex environments, and has broad application prospects.

**Key words:** object detection; deep learning; deep convolutional neural networks (CNN); feature extraction; computer vision; trunk detection; complex scenarios

森林树干检测技术对林业机器人在复杂森林环境中的精确感知和定位树木主干位置至关重要。尤其在复杂的森林场景中,树干的识别变得极具挑战,不同种类的树木、密集的植被和不规则的地形都增加了检测的难度。这项技术不仅能提升林业机器人的导航和避障能力,还有助于管理者更深入地了解森林结构和组成,从而显著提高森林资源监测的精准性<sup>[1]</sup>。因此,针对复杂森林场景的树干快速自动化检测成为当前研究的热点之一。

传统的树干检测方法通常利用视觉信息来区分复杂背景中的树干区域,例如:刘慧等<sup>[2]</sup>融合彩色和纹理特征,并利用简单线性迭代聚类算法,对树干进行定位识别;Chen 等<sup>[3]</sup>将颜色直方图与训练分类器相结合,以提高树干检测性能。但是,各树种树干颜色和纹理的差异,以及树干与背景色的相似性,限制了利用色彩空间进行树干检测的性能表现。

近年来,以深度卷积神经网络(convolutional neural network, CNN)为代表的深度学习方法,在计算机视觉领域的目标检测任务中受到广泛关注和应用<sup>[4]</sup>。基于 CNN 的目标检测方法具有良好的特征提取能力,有助于弥补传统方法对图像颜色和纹理等信息依赖较强的不足,从而提高树干检测的准确性和鲁棒性。当前,基于 CNN 的目标检测算法主要分为双阶段检测算法和单阶段检测算法。双阶段目标检测算法如 R-CNN<sup>[5]</sup>、Fast R-CNN<sup>[6]</sup>、Cascade R-CNN<sup>[7]</sup>等。双阶段检测算法首先需定位图像中潜在目标区域,随后对这些区域内的目标进行分类和位置标注。然而,在实际场景中,常伴随着模型规模庞大和检测速度较慢等挑战。单阶段检测算法,例如 SSD(single shot multi-box detector)<sup>[8]</sup>和 YOLO(you only look once)<sup>[9]</sup>系列算法,融合了分类与回归思想,可直接处理图像,一次性完成目标的分类与位置信

息提取,具备检测速度快速的优势。因为林业机器人对树干检测具有高实时性的要求,所以研究人员更加关注并使用单阶段算法,如:Zhang 等<sup>[10]</sup>通过将 YOLOv5 模型中的 C3 模块改为 GhostNet V2,将 CIoU 损失函数改为 SIoU,并添加协调注意机制,减少了网络参数,提高了检测框的回归准确性,并减少了背景干扰;Lü等<sup>[11]</sup>提出了两个模块,Embedded-A 通过增加特征图深度和残差结构提高了模型轻量化,Embedded-B 通过深度卷积和池化融合提高了特征图采样精度,从而增强了检测性能。

但是上述模型在处理复杂森林背景,应对多尺度特征和小目标检测方面仍显不足。复杂的背景干扰会导致误检和漏检,多尺度特征处理和小目标检测精度欠佳。此外,模型的实时性和精度平衡仍需优化,模型轻量化在保持检测能力上存在局限。针对以上问题,同时兼顾检测精度与检测实时性,本研究提出一种基于改进 YOLOv8 的树干检测模型。首先,为有效降低复杂背景干扰,设计一种空间关注度模块(spatial awareness module, SAM),以促进特征之间的信息流动,增强特征的抽象表达能力,从而有助于模型更准确地捕获树干目标的语义信息。其次,为解决树干尺寸差距大的问题,设计一种快速尺度适应的架构(FastScaleNet),专注于处理树干多尺度和多分辨率特征,旨在增强模型的稳定性和泛化能力。此外,引入一种加权交并比损失函数,动态调整小目标的损失权重,从而使模型更好地适应不同尺寸树干目标的检测需求。

## 1 复杂森林场景的树干检测算法框架

### 1.1 基础算法框架 YOLOv8

本研究采用 YOLOv8 模型作为基础算法框架,该模型是 Ultralytics 公司在 2023 年发布的一款新算

法。相较于 YOLO 系列之前的优秀模型(如 YOLOv5 和 YOLOv7), YOLOv8 是一款先进且前沿的模型, 它提供了更高的检测精度和速度。YOLOv8 网络结构主要由 3 个部分组成: 主干网络(backbone)、颈部网络(neck)以及头部网络(head)。

YOLOv8 采用改进后的 CSPDarknet53 作为主干网络, 输入特征经过 5 次下采样得到 5 个不同尺度的特征, 依次记为 B1 ~ B5。原主干网络中的 CSP (cross stage partial)模块被替换为 C2f 模块, C2f 模块的结构如图 1 所示。C2f 模块采用了梯度分流连接方式, 在保持轻量化的同时丰富了特征提取网络的信息流。CBS 模块对接收到的输入信息进行卷积

操作后, 紧接着进行批量归一化处理, 并最终利用 SiLU 激活函数对信息流进行激活以获取输出结果, 这一过程如图 1 所示。

主干网络最后运用 SPPF 模块(空间金字塔池化快速模块)对输入的特征图进行池化操作, 将其转换为固定尺寸的特征图, 以便适应不同尺寸的输出。相比于传统的 SPP(空间金字塔池化)结构, SPPF 模块通过顺序连接 3 层最大池化层的方式降低了计算量, 减少了延迟。

受 PANet 启发, YOLOv8 在颈部设计了 PAN-FPN 结构。相较于 YOLOv5 和 YOLOv7 模型的颈部结构, YOLOv8 在 PAN<sup>[12]</sup>结构中去除了上采样之

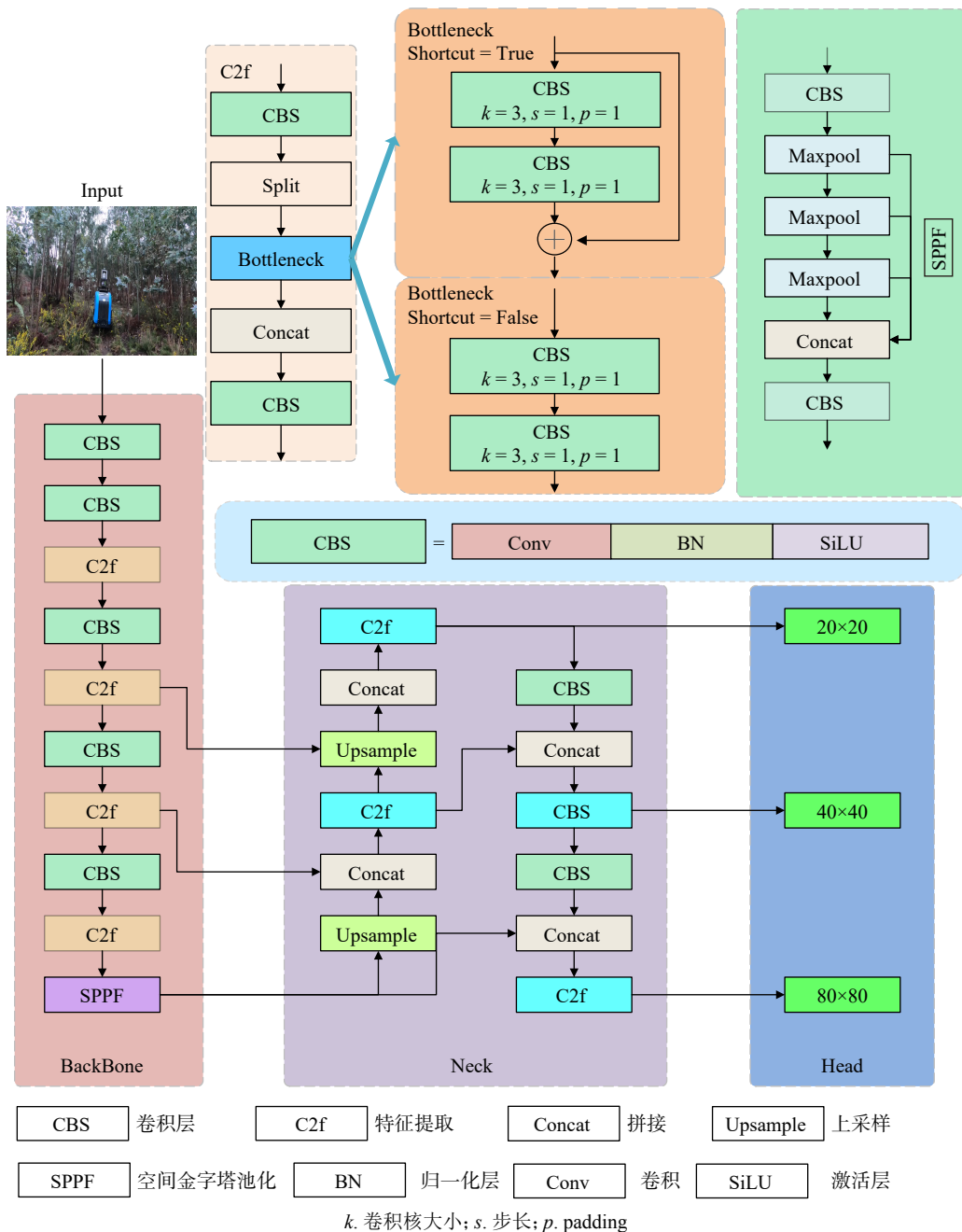


图1 YOLOv8 结构图

Fig. 1 YOLOv8 structural diagram

后的卷积操作,从而在保持原有性能的基础上实现了模型轻量化。分别用 P4-P5 和 N4-N5 表示 YOLOv8 模型中 PAN 结构和 FPN 结构中的两种不同尺度特征。传统的 FPN<sup>[13]</sup>采用自顶向下的方式传递深层语义信息,通过融合 B4-P4 和 B3-P3 增强特征的语义信息,但会损失一定的目标定位信息。为缓解这个问题, PAN-FPN 在 FPN 基础上添加了 PAN 模块。PAN 通过融合 P4-N4 和 P5-N5 增强了位置信息的学习,以自顶向下的形式实现路径增强。PAN-FPN 构建了一种自顶向下与自底向上的网络结构,通过特征融合实现了浅层位置信息与深层语义信息的互补性,进而产生多样性和完整性的特征,提升了特征表达的质量。

YOLOv8 的检测部分采用了分离式头结构(图 1)。该分离式头结构使用两个独立的分支分别处理物体分类任务和预测边框回归任务,并针对这两种类型的任务采用了不同的损失函数。对于分类任务,采用二元交叉熵损失(BCE Loss)<sup>[14]</sup>。而对于预测边界框回归任务,则采用分布焦点损失(distribution focal loss, DFL)和 CIoU 损失函数。这种检测结构能够提

高检测准确性,并加速模型收敛。YOLOv8 是一个无锚点检测模型,简洁地定义了正负样本。同时,它还采用了任务对齐分配器动态分配样本,这进一步提高了模型的检测准确性和鲁棒性。

### 1.2 SSFYOLO 模型

本研究根据基础模型 YOLOv8 提出 SSFYOLO 模型,整体结构如图 2 所示。树干尺寸差异大,易受遮挡,导致复杂场景下漏检和误检。为解决这个问题,设计空间感知 SAM 模块,以处理多尺度和多分辨率的特征信息,在确保计算效率的同时,对各类特征实现深度整合与精确提取;设计 FastScaleNet 自适应网络结构,其替代传统特征金字塔中的 C2f 模块,通过优化多尺度特征表达,提升模型在不同环境下的稳定性和适用范围,减轻计算负担,提高检测的精确度;利用加权损失函数(weighted intersection over union, WIoU)方法<sup>[15]</sup>实现小目标损失权重的动态调整,使模型能够灵活应对各种尺寸树干的检测需求,从而显著增强整体的检测性能与效果。

#### 1.2.1 空间感知 SAM

针对目前树干检测面临的精度不足和运算效率

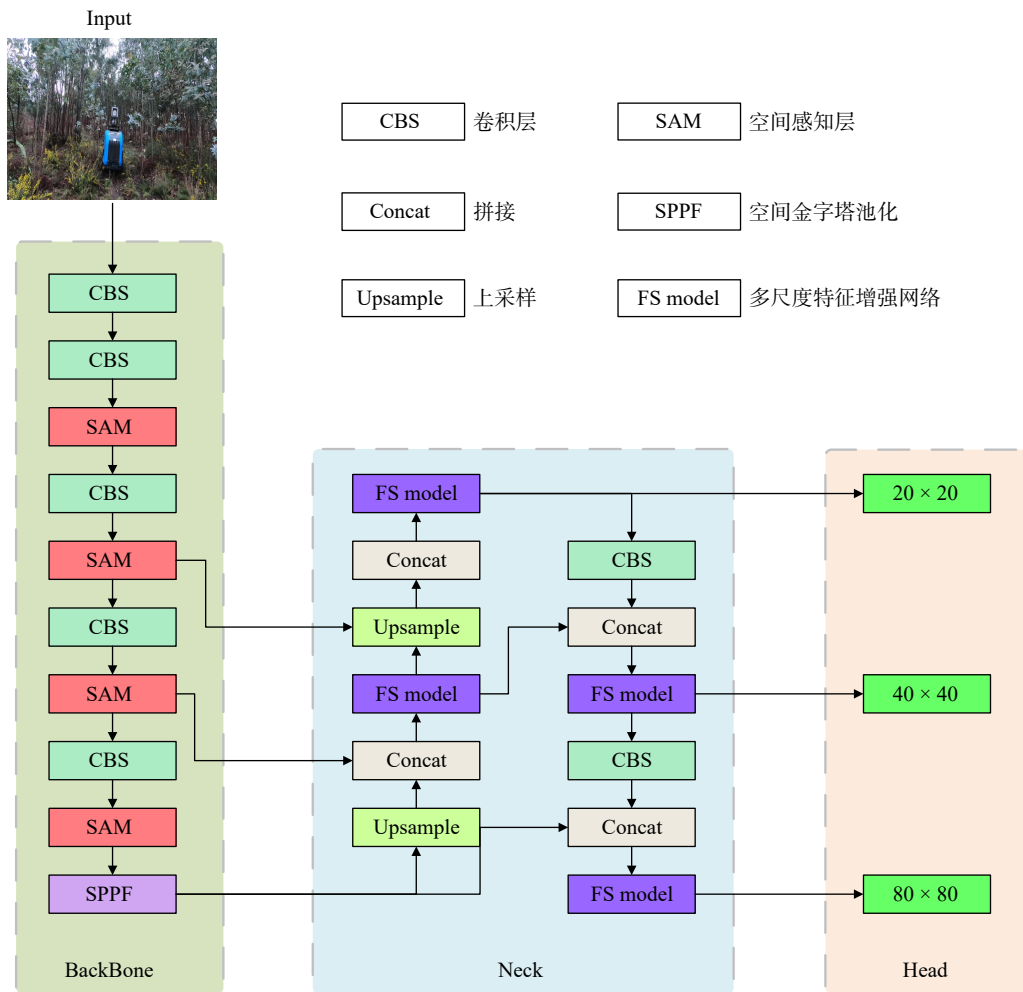


图 2 SSFYOLO 模型图

Fig. 2 SSFYOLO model diagram

低下等挑战,设计一种空间感知模块 SAM 的全新架构设计。相较于传统的 Transformer 模型<sup>[16]</sup>, SAM 架构在检测性能、计算效率、参数利用率以及模型适应性等方面均显示出显著的优势。

在本研究中,将 SAM 架构融入到 YOLOv8 目标检测模型的主干网络之中,旨在更好地挖掘和利用特征图中的空间位置关联与通道间的信息交互,进而提升对树干这类小目标的检测精确度。SAM 模块首先采用分组卷积,让模型并行处理不同的特征子空间,每个通道组内的信息可以被单独学习和提取(图 3)。将每个输入通道分成多个组,每个组内的卷积操作独立地捕捉特定的特征信息。然后,逐点卷积允许在不同的特征子空间之间进行信息交互和整合。通过这种组合, SAM 模块能够捕捉到不同层次特征之间的内在联系;分组卷积和逐点卷积操作的共同作用,使每个通道组内的特征能够在更高层次上进行整合和提取,从而显著增强模型对微小物体(如树干边缘和纹理特征)的识别能力。

每一步卷积操作之后, SAM 模块都配置了激活

函数 GELU 进行非线性变换(式 1),紧接着运用 BatchNorm 进行归一化处理(式 2)。这种连续的操作序列有助于稳定训练过程,并进一步优化特征表达。

$$z_l = \text{BN}(\sigma\{\text{ConvDepthwise}(z_{l-1})\} + z_{l-1}) \quad (1)$$

$$z_{l+1} = \text{BN}(\sigma\{\text{ConvDepthwise}(z_l)\}) \quad (2)$$

式中:BN表示批标准化层,σ代表激活函数,ConvDepthwise 代表深度可分离卷积操作,  $z_{l-1}$  表示上一层的输出,  $z_l$  表示当前层的输出,  $z_{l+1}$  表示下一层的输出。

最后,使用 MLP 结构整合不同位置和不同通道的特征信息,并在整个特征映射上进行全局的上下文建模。通过这种方式,MLP 可以帮助模型更好地理解特征之间的关联性,从而提高模型对于复杂数据的抽象能力和泛化能力。

与传统的卷积神经网络相比, SAM 在抽取特征的过程中能有效地保留输入原始结构的完整性,这一特性使得 SAM 在作为目标检测模型的主干网络组件时展现卓越的表现力。同时,相比于 Transformer

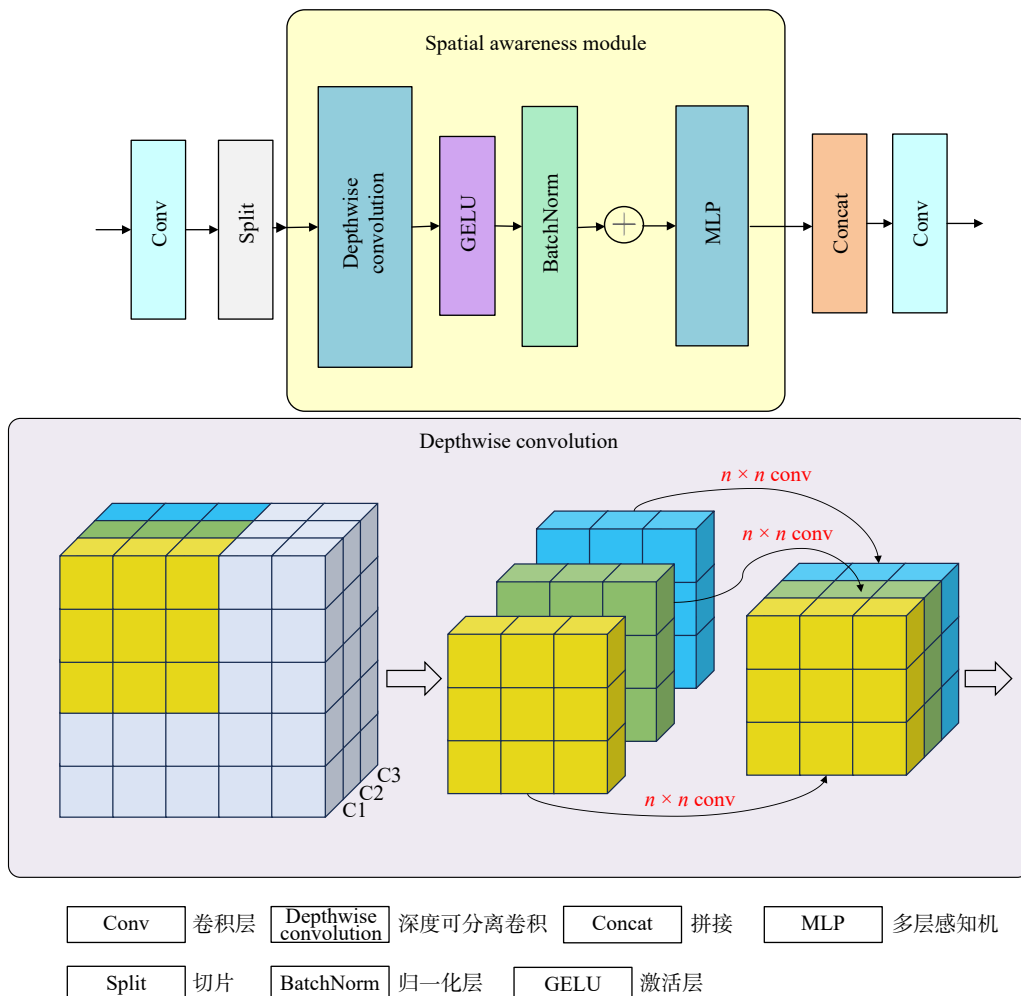


图3 SAM 结构图

Fig. 3 Structure diagram of SAM

架构, SAM 在设计上更为简洁明了, 既便于理解, 也易于实施和优化, 从而为提升树干检测任务的整体性能提供有力支持。

1.2.2 多尺度特征增强网络 FastScaleNet

针对树木生长过程中树干尺寸多变且密集分布的难题, 构建了一种快速尺度适应网络 FastScaleNet。该结构设计的目的在于灵活适应并精准定位不同尺度下的树干目标, 以确保检测精度的同时, 大幅提升检测效率。具体结构如图 4 所示。首先, 为有效应对目标物体尺寸跨度较大的难题, 采用特征 split 拆分策略, 让网络在多个层级上都能够捕获到更为多样且细致的特征表达, 从而拓宽网络的感受野, 并强化其对局部细微特征的敏锐感知能力。但是特征拆分会导致一定程度上的特征冗余, 这些冗余特征无疑会加重网络的计算负荷。为此, 引入了高效特征选择卷积(PCConv), 在精简去除冗余特征的同时, 最大限度地保留关键的特征信息, 从而实现特征的精密调控与优化。面对树干尺寸变异大且密度不均匀的现实挑战, 进一步嵌入增强型移动平均注意力(enhanced moving average attention, EMA)机制。该结构采纳了期望最大化(expectation-maximization, EM)算法框架, 通过迭代优化一组紧凑的基础向量集, 以此为基础进行注意力运算。具体来说, 在 EM 算法流程中, E 步骤负责依据当前估计更新注意力分布图, M 步骤则专注于更新这一组基础向量集合; 这两步迭代循环进行, 直至达到收敛状态。

在 E 步骤中, 使用当前参数  $\theta^{old}$  求出由  $p(Z|X, \theta^{old})$  给出的 Z 的后验分布。然后, 使用后验分布求出完全数据似然  $Q(\theta, \theta^{old})$  的期望, 计算公式为

$$Q(\theta, \theta^{old}) = \sum_Z p(Z|X, \theta^{old}) \ln p(X, Z|\theta) \quad (3)$$

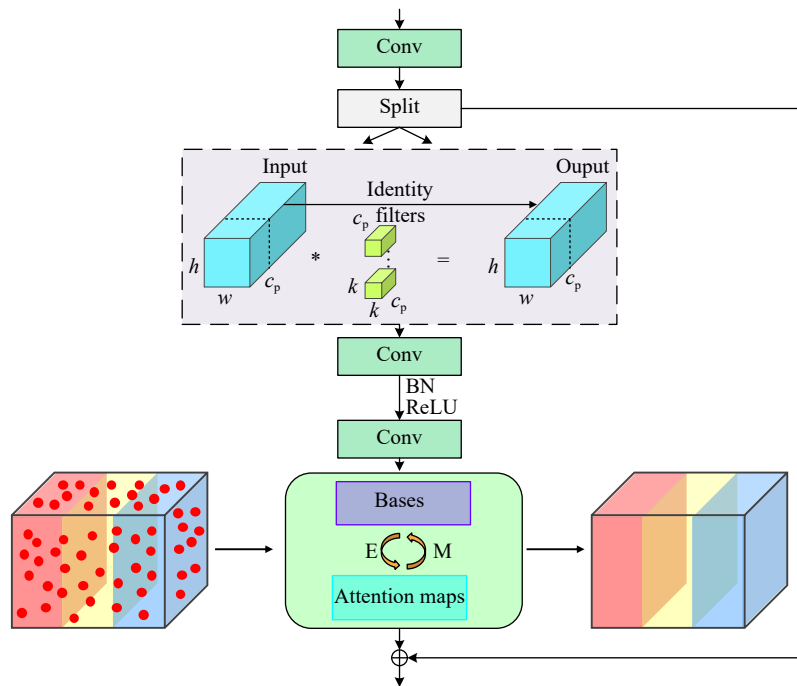
然后, 在 M 步骤中, 通过最大化函数确定修订的参数  $\theta^{new}$ 。

$$\theta^{new} = \operatorname{argmax} Q(\theta, \theta^{old}) \quad (4)$$

最后, 利用这一优化后的注意力结果重构特征图, 从而使网络具备更强大的自适应性和灵活性, 进一步优化了检测过程, 使其具有更快的响应速度和更高的效率。这种整合的方法不仅在树干目标检测的效率方面取得了巨大成功, 而且在处理尺度差异大, 数量密集的情况下表现出色, 从而有效应对了树木生长监测中的尺度变化和数量密集的挑战, 为树木生长过程中的监测和管理提供了可靠的技术支持。

1.2.3 加权损失函数 WIoU

在树干目标检测任务中, 距离导致远处的树干目标较小, 因此需要优化损失函数以提升模型对较小树干的检测性能。YOLOv8 利用 DFL(distance-IoU loss)和 CIoU(complete IoU)计算边界框的回归损失, 但 CIoU 存在以下不足之处。首先, CIoU 并未考虑到难易样本之间的平衡问题。这意味着在训练过程中, 对于简单和困难的样本, CIoU 损失函数可能会给予相同的重视程度, 而不利于模型对难以定位的目标进行学习。其次, CIoU 损失函数中使用了



Conv. 卷积层; Split. 切片; h. 高度; w. 宽度; c<sub>p</sub>. 通道数; filters. 卷积核; k. 卷积核数量; \*. 卷积; BN. 归一化层; ReLU. 激活函数; Bases. 基础特征; E. 更新注意力图; M. 更新这组基; Attention maps. 注意力映射

图 4 FastScaleNet 结构图

Fig. 4 Structure diagram of FastScaleNet

宽高比作为惩罚因子之一,如果真实边界框和预测边界框的宽高比相同,但是宽度和高度的具体数值不同,此时损失函数的惩罚项并不能准确反映出这两个框之间的真实差异。第三,CIoU的计算涉及到反三角函数,这会增加模型运算时的计算负担。CIoU损失函数的计算为

$$L_{CIoU} = 1 - I_{oU} + \frac{\rho^2(b, b^{gt})}{(c_w)^2 + (c_h)^2} + \frac{4}{\pi^2} \left( \tan^{-1} \frac{w^{gt}}{h^{gt}} - \tan^{-1} \frac{w}{h} \right) \quad (5)$$

式中:  $I_{oU}$  表示预测框和真实框的交集面积与并集面积之比,  $\rho^2(b, b^{gt})$  表示真实框与预测框中心点之间的欧氏距离,  $h$  和  $w$  分别代表预测框的高度和宽度,  $h^{gt}$  和  $w^{gt}$  则分别代表真实框的高度和宽度,  $c_h$  和  $c_w$  表示由预测框和真实框共同构成的最小外接矩形的高度和宽度。

针对树干目标检测任务中 CIoU 存在的局限性,引入 WIoU v3 作为损失函数。WIoU 不仅考虑了宽高比、中心点距离和重叠区域,还引入了动态非单调聚焦机制,并且采取了一种合理的梯度增益分配策略评估锚框的质量。Tong 等<sup>[15]</sup>提出了 WIoU 的 3 个版本,其中 WIoU v1 设计了基于注意力的预测框损失,而 WIoU v2 和 WIoU v3 在此基础上增加了聚焦系数。WIoU v1 引入距离作为一种注意力度量指标。当目标物体框与预测框在一定范围内重叠时,减少几何度量的惩罚措施使模型能够获得更好的泛化能力。WIoU v1 的计算公式为

$$L_{IoU} = 1 - I_{oU} \quad (6)$$

$$L_{WIoUv1} = R_{WIoU} L_{IoU} \quad (7)$$

$$R_{WIoU} = \exp \left( \frac{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2}{c_w^2 + c_h^2} \right), R_{WIoU} \in (1, e) \quad (8)$$

式中:  $L_{IoU} \in [0, 1]$ ,  $c_h$  和  $c_w$  表示由预测框和真实框共同构成的最小外接矩形的高度和宽度,  $R_{WIoU}$  表示为预测框和真实框中心点之间的归一化距离。

WIoU v2 在 WIoU v1 的基础上构建了单调聚焦系数  $L^*_{IoU}$ , 有效地降低了损失值中简单样本的权重。然而,在模型训练过程中,随着  $L^*_{IoU}$  的减小,  $L_{IoU}$  随之减小,这可能导致模型收敛速度变慢。因此,引入了  $L_{IoU}$  平均值规范化  $L^*_{IoU}$ 。WIoU v2 的计算公式为

$$L_{WIoUv2} = \left( \frac{L^*_{IoU}}{L_{IoU}} \right)^\gamma L_{WIoUv1}, \gamma > 0 \quad (9)$$

式中:  $\gamma$  是非单调焦点因子,  $L^*_{IoU}$  为单调焦点系数。

WIoU v3 定义了一个异常值  $\beta$  来衡量锚框的质量,并基于  $\beta$  构建了一个非单调聚焦因子  $\gamma$ 。将  $\gamma$  应用

于 WIoU v1。当  $\beta$  较小时,锚框质量较高,赋予较小的  $\gamma$  值,从而降低高质量锚框在较大损失函数中的权重;当  $\beta$  较大时,锚框质量较低,赋予较小的梯度增益,这样可以减少由低质量锚框产生的有害梯度。WIoU v3 通过采用一种合理的梯度增益分配策略,在损失函数中动态优化高质量和低质量锚框的权重,从而使模型更多关注平均质量样本,进而提升模型的整体性能。WIoU v3 的计算公式为

$$L_{WIoUv3} = \gamma L_{WIoUv1} \quad (10)$$

$$\gamma = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (11)$$

式中:  $\delta$  和  $\alpha$  是可调的超参数,可以根据不同模型进行调整,  $\beta$  是离群度。

在树干目标检测任务中,小目标的检测一直是一个挑战。而 WIoU v3 通过动态优化小目标损失权重,能够更好地适应不同尺寸目标的检测需求。WIoU v3 引入了加权机制,根据目标尺寸动态调整损失权重,使得小目标在训练过程中的占比更大,从而提高了检测精度。此外, WIoU v3 还优化了损失函数,使模型在复杂背景下具有更强的鲁棒性和稳定性。引入 WIoU v3 后,模型在真实世界的树干场景下表现出色,不仅能显著提升目标检测性能,还能有效减少误检和漏检,从而满足实际应用的需求。

## 2 研究方法

采用前面提出的 SSFYOLO 算法对复杂场景下树干数据集进行检测实验。本实验采用的操作系统环境是 Ubuntu 20.04, 配备的 GPU 型号为 NVIDIA RTX 3080, 其显存容量大小为 10 GB。选用 Pytorch 1.8.0 的深度学习框架,并结合 CUDA 11.3 版本的并行计算能力以及 CUDNN 8.3.4 版本的神经网络加速库,以提升运算效率。所有实验模型初始化时,学习率设定为 0.01, 梯度动量参数设定为 0.937, 权重衰减系数精确至 0.000 5。为了更好地收敛和稳定训练过程,采用 3 个 epoch 的 warm-up 策略预热学习率,其中动量参数设定为 0.8。在预热阶段结束后,进一步应用余弦退火算法动态调整学习率。实验中的训练 batch size 设定为 16, 训练总轮次为 200 个 epoch, 以确保模型得到充分优化与训练。

### 2.1 数据集

本研究采用的 Forest 数据集是文献 [17] 展示的数据集的最新扩充版本,新增超过 2 000 张带有详尽标注的新图像资料。这些增补图像源于文献<sup>[18]</sup>中的 Robot Operating System(ROS)bag 数据集资源库。采集地点横跨葡萄牙三大地理区域: 41°11'02"N、8°29'06"W 的 Lobão 地区, 41°21'01"N、8°44'03"W

的 Vila do Conde 区域, 以及位于  $40^{\circ}59'01''\text{N}$ 、 $8^{\circ}29'02''\text{W}$  的 Valongo 地区。这 3 个林区主要由桉树 (*Eucalyptus* spp.)、松树 (*Pinus* spp.) 两个树种组成。不同采集地区的环境条件和地理特征存在差异, 树干图像样本呈现一定的异质性, 这使得实验能够更充分地验证所提算法检测的有效性与适应性。实验中使用数据集共包含 2 690 张树干图像, 其中将数据集按 8:2 的比例划分为训练集和测试集, 具体而言, 训练集中包含 2 152 张图像, 而测试集则包含 538 张图像。

## 2.2 算法的评价

### 2.2.1 改进模块的有效性评价

为对比验证各个改进模块的优化作用, 探究本算法的有效性, 在 YOLOv8 基础上进行消融实验, 每组实验均使用相同的数据集、超参数设置和训练策略。

### 2.2.2 算法的对比评价

为进一步验证 SSFYOLO 算法改进的合理性和有效性, 在实验环境和数据集不变的情况下, 与当前主流的各种目标检测算法 (Faster-RCNN、YOLOv5 不同模型、YOLOv6 和 YOLOv7-tiny 算法) 进行对比实验。

## 2.3 评价指标

为衡量改进模型的检测效果, 采用精确度 (precision,  $P$ )、召回率 (recall,  $R$ )、平均精度 (mean average precision, mAP)、参数量和计算量作为评价指标。精确度表示在所有被分类器判定是正类的样本中, 实际为正类的样本; 召回率表示在所有实际为正类的样本中, 被正确预测为正类的比例。参数量表示模型训练时所有网络层的参数之和, 计算量表示模型的浮点运算数, 二者数值可作为算法复杂程度的衡量标准。通过混淆矩阵对精确度、召回率和平均精度进行计算, 其中混淆矩阵包含 4 个指标:  $T_P$  指模型正确预测的正样本数量;  $T_N$  指模型正确预测的负样本数量;  $F_P$  指错误预测的正样本数量;  $F_N$  指错误预测的负样本数量。具体计算公式为

$$P = \frac{T_P}{T_P + F_P} \times 100\% \quad (12)$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \quad (13)$$

$$P_A = \int_0^1 P(R) dR \quad (14)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N P_{Ai} \quad (15)$$

式中:  $N$  为数据集类别数目;  $P_A$  为  $P$ 、 $R$  曲线所围成的面积。

## 3 结果与分析

### 3.1 改进模块的有效性评价

为验证方法有效性, 本研究选择 IoU 阈值为 0.5 时的平均精度作为评价指标。表 1 显示: 在单个模块的消融实验中, 通过改进 YOLOv8 的主干网络, 使用 SAM 模块平均精度提高了 1.2 个百分点, 使模型能够更加准确地捕捉和利用数据中的关键特征, 从而提高了其在目标检测或分类任务中的整体性能。在使用 SAM 模块的基础上, 进一步将设计的 FastScaleNet 模块应用于模型中后, 相比于单独使用 SAM 模块的模型, 其精确度、召回率和平均精度分别提高了 0.8、0.2 和 0.1 个百分点。这是因为 FastScaleNet 模块使模型更好地学习数据的关键特征, 从而提高了模型的表征能力。在不应用前两个改进, 单独将损失函数由原来的 CIoU 替换为 WIoU v3 后, 平均精度提高了 1 个百分点, 这是因为 WIoU 可以让模型更好地学习目标的位置和形状信息, 从而提升目标检测的性能和准确度。在同时应用上述 3 种改进后, 本模型相较于原始的 YOLOv8, 召回率和平均精度分别提升了 0.7 和 1.6 个百分点, 参数量降低了 20%。由此可见模型在显著降低参数量的同时, 依然能够保证较为优异的检测性能, 这也表明模型具备在实际应用场景中的部署能力。

SSFYOLO 在一定召回率区间内的精确率略高于 YOLOv8, 尤其是在召回率较高的情况下,

表 1 消融实验结果

Tab. 1 Results of ablation experiment

原始模型	+SAM	+FastScaleNet	+WIoU v3	参数量	精确率( $P$ )/%	召回率( $R$ )/%	平均精度(mAP)/%
				$3.0 \times 10^6$	89.2	89.2	92.8
	√			$2.8 \times 10^6$	88.5	87.9	94.0
		√		$3.0 \times 10^6$	88.7	87.3	93.5
YOLOv8			√	$2.2 \times 10^6$	89.6	86.8	93.8
	√	√		$2.8 \times 10^6$	89.3	88.1	94.1
	√	√	√	$2.4 \times 10^6$	88.9	89.9	94.4

注: “√”表示使用改进方法。

SSFYOLO 的精确率下降得相对缓慢(图 5), 这意味着其在处理复杂森林场景时表现更好。YOLOv8 在高置信度区间内的召回率下降更为明显(图 6), 表明它的高置信度阈值下可能会错过一些正类样本。而 SSFYOLO 在高置信度区间内的召回率下降得相对缓慢(图 6), 这意味着其在极端情况时表现更好。

### 3.2 对比实验

为进一步验证 SSFYOLO 算法改进的合理性和有效性, 在实验环境和数据集不变的情况下, 与当前主流的各种目标检测算法 Faster-RCNN、YOLOv5s、YOLOv5m、YOLOv6 和 YOLOv7-tiny 算法进行了对比(表 2)。

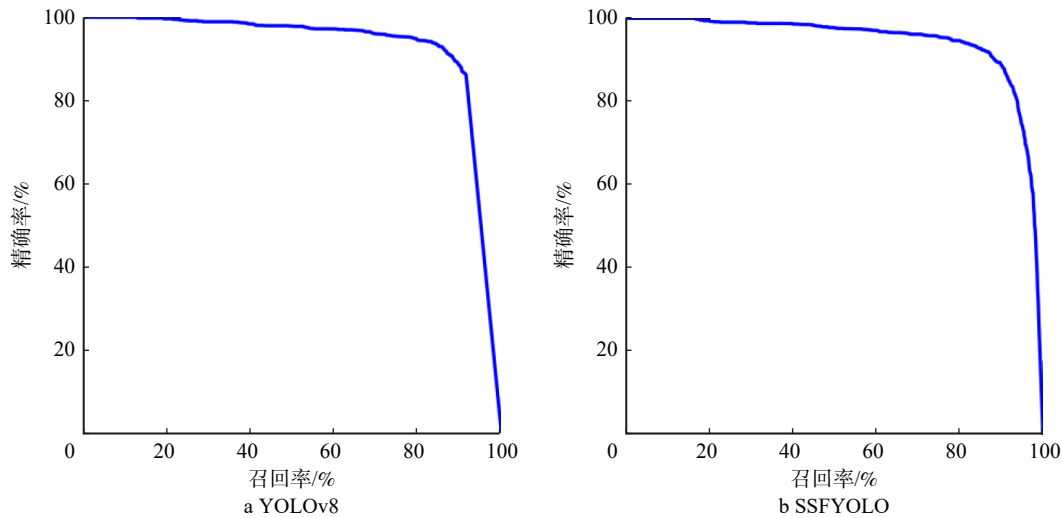


图 5 精确率-召回率曲线对比图

Fig. 5 Precision-recall curve comparison chart

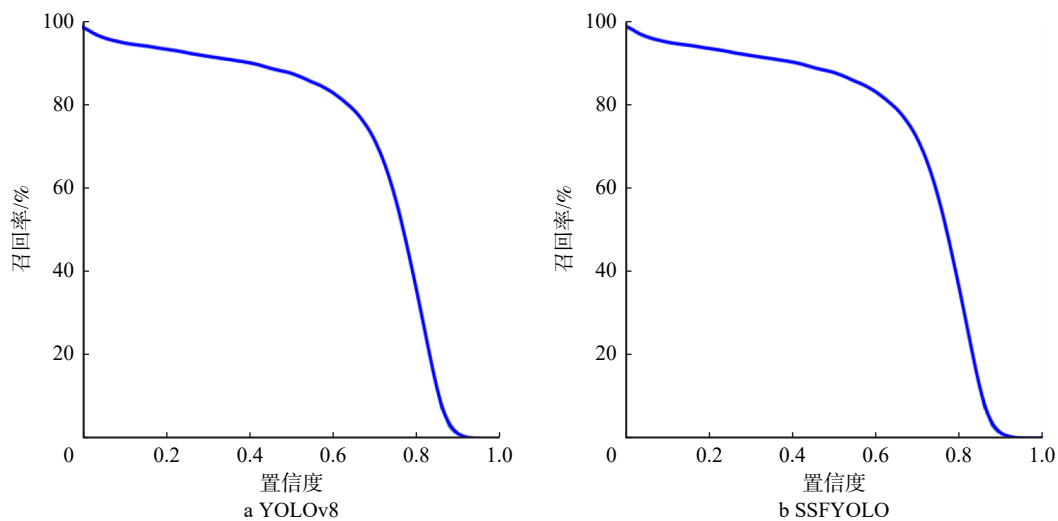


图 6 召回率对比图

Fig. 6 Comparison chart of recall

表 2 对比实验结果

Tab. 2 Comparative experimental results

模型	参数量	FLOPs	P/%	R/%	mAP/%
Faster-RCNN	$41.3 \times 10^6$	$251.4 \times 10^9$	83.5	84.7	87.9
YOLOv5s	$7.0 \times 10^6$	$15.8 \times 10^9$	88.7	89.4	93.8
YOLOv5m	$20.8 \times 10^6$	$47.9 \times 10^9$	90.7	89.5	93.7
YOLOv6	$9.7 \times 10^6$	$24.8 \times 10^9$	88.6	87.3	93.5
YOLOv7-tiny	$6.9 \times 10^6$	$13.8 \times 10^9$	80.0	72.0	81.7
SSFYOLO	$2.4 \times 10^6$	$7.1 \times 10^9$	88.9	89.9	94.4

注: 计算量采用每秒浮点运算次数(floating-point operations per second, FLOPs)表示。

首先,从模型参数量角度看,SSFYOLO 模型将参数量进一步减小至  $2.4 \times 10^6$ ,体现了较高的模型压缩效率。YOLOv7-tiny 参数量与 YOLOv5s 相当,但其在精度和召回率上的表现相对较弱。其次,在计算效率上,SSFYOLO 以 FLOPs 为  $7.1 \times 10^9$  的低计算量展现了优秀的实时性能,远低于 FasterRCNN 的  $251.4 \times 10^9$ ,且在保证了高效性的同时,精度并未显著降低。在精确度和召回率方面,YOLOv6 和 YOLOv5 系列整体表现优秀,尤其是 YOLOv5m,其精确度高达 90.7%,召回率为 89.5%。然而,

SSFYOLO 在保持相对较小的参数量和较低计算复杂度的同时,精确度和召回率依然分别达到了 88.9% 和 89.9%,在平均精度这一关键指标上达到 94.4%,超越了所有对比模型,展现出了强大的目标检测能力和良好的平衡性。综上所述,通过本次对比实验可以看出,相较于已有的经典目标检测模型,SSFYOLO 在保持较高精度和召回率的前提下,成功实现了模型的小型化和计算效率的提升,具有显著的竞争优势和实际应用价值。

通过图 7 第 1 列对比可知,SSFYOLO 模型相

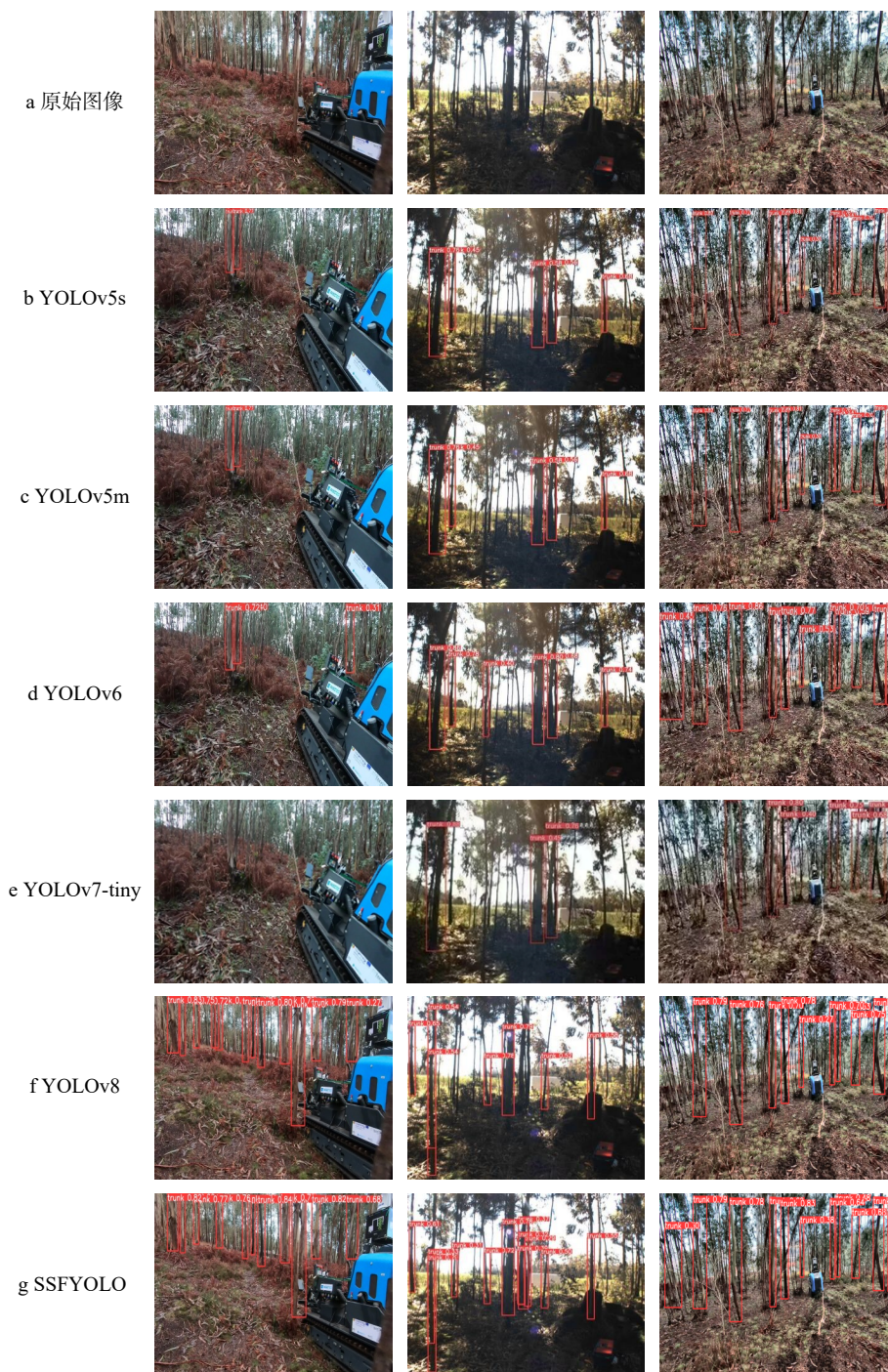


图 7 不同模型检测结果对比

Fig. 7 Comparison of detection results of different models

较于 YOLOv8 在目标置信度方面显著提升。通过图 7 第 2 列对比可知, SSFYOLO 模型成功降低了漏检和误检的发生率。图 7 第 3 列图像对比显示: SSFYOLO 模型同样表现出色, 不仅减少了原有的漏检与误检现象, 而且有效提高了整体的检测准确性。但由于林地场景通常具有丰富多样的背景, 如树木、树叶、灌木等, 这些背景元素可能与目标具有相似的纹理或颜色, 并且林地场景中的光照条件通常变化较大, 例如树木的阴影、阳光透过树叶的效果等, 这些变化可能导致目标外观发生变化, 导致检测仍然会出现一些误检或漏检。

## 4 结 论

为解决树干在复杂背景与密集种植环境下的检测难题, 本研究设计出了 SSFYOLO 算法。该算法专为实现高效且精确的树干识别任务而设计。其核心架构集成了空间感知模块 SAM, 通过优化特征层间的相互作用与整合过程, 显著增强了模型对目标深层次语义特征的理解与抓取能力。SSFYOLO 的独特之处还在于引入了一种创新的尺度自适应网络组件 FastScaleNet, 专门针对多尺度特征处理问题, 极大地提高了模型在各种条件下的稳健性和泛化能力。另外, 算法采用了动态调整的小目标损失权重策略, 确保模型在面对不同尺寸树干时, 能够灵活适应并精确检测。尽管 SSFYOLO 在树干检测领域已取得显著成就, 但仍存在进一步提升的空间, 特别是在提升检测精度与加快检测速度等方面。因此, 未来的研究工作将持续聚焦于网络结构优化、模型大小压缩、参数精简及检测速度提升等多个层面, 以期不断完善, 并增强 SSFYOLO 算法的整体效能, 使其在树干检测任务上发挥更大价值。

### 参 考 文 献

- [1] 罗海风, 谭月胜. 一种欠驱动双板分离式采摘器的研制[J]. 北京林业大学学报, 2018, 40(12): 110-116.  
Luo H F, Tan Y S. Research on an under actuated dual separation plate harvester[J]. Journal of Beijing Forestry University, 2018, 40(12): 110-116.
- [2] 刘慧, 朱晟辉, 沈跃, 等. 基于多特征融合的树干快速分割算法[J]. 农业机械学报, 2020, 51(1): 221-229.  
Liu H, Zhu S H, Shen Y, et al. Fast segmentation algorithm of tree trunks based on multi-feature fusion[J]. Transactions of the Chinese Society of Agricultural Machinery, 2020, 51(1): 221-229.
- [3] Chen X, Zhang B, Luo L. Multi-feature fusion tree trunk detection and orchard mobile robot localization using camera/ultrasonic sensors[J]. Computers and Electronics in Agriculture, 2018, 147: 91-108.
- [4] Zhao Z Q, Zheng P, Xu S, et al. Object detection with deep learning: a review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.
- [5] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// O'Conner L. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2014: 580-587.
- [6] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 1137-1149.
- [7] Cai Z W, Nuno V. Cascade R-CNN: delving into high quality object detection[C]// O'Conner L. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6154-6162.
- [8] Huo A Q, Zhang W L, Li Y. Traffic sign recognition based on improved SSD model [C]//Cristina C. Proceedings of International Conference on Computer Network, Electronic and Automation (ICCNEA). Xi'an: IEEE, 2020: 54-58.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// O'Conner L. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779-788.
- [10] Zhang J, Tian M, Yang Z, et al. An improved target detection method based on YOLOv5 in natural orchard environments[J]. Computers and Electronics in Agriculture, 2024, 219: 108780.
- [11] Lü F, Wang X, Li L, et al. Tree detection algorithm based on embedded YOLO lightweight network[J]. Journal of Shanghai Jiaotong University (Science), 2022, 29: 518-527.
- [12] Yang J F, Fu X Y, Hu Y W, et al. PanNet: a deep network architecture for pan-sharpening[C]//O'Conner L. Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 1753-1761.
- [13] Ghiasi G, Lin T Y. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]// O'Conner L. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 7029-7038.
- [14] Li Q F, Jia X, Zhou J C, et al. Rediscovering BCE loss for uniform classification[J/OL]. arXiv preprint arXiv, 2024 [2023-12-20]. <https://arxiv.org/abs/2403.07289>.
- [15] Tong Z J, Chen Y H, Xu Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[J/OL]. arXiv preprint arXiv, 2023 [2024-03-20]. <https://arxiv.org/abs/2301.10051>.
- [16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// Curran Associates, Inc. Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Red Hook, 2017: 6000-6010.
- [17] Silva D, Santos F, Sousa A, et al. Visible and thermal image-based trunk detection with deep learning for forestry mobile robotics[J]. Journal of Imaging, 2021, 7(9): 176-180.
- [18] Silva D, Santos F, Sousa A, et al. Unimodal and multimodal perception for forest management: review and dataset[J]. Computation, 2021, 9(12): 127-132.

(责任编辑 吴娟  
责任编辑 陈志泊)