

基于多模态遥感图像融合的檀香树幼苗精准检测

张宇¹ 黄泽众¹ 董金让¹ 韦文正¹ 赵懿琨¹ 涂淑琴²

(1. 华南农业大学电子工程学院(人工智能学院), 广州 510642;

2. 华南农业大学数学与信息学院, 广州 510642)

摘要: 檀香树幼苗在生长过程中依赖伴生植物,且常与其他作物混种,导致其生长环境复杂,难以实现精准检测,进而影响檀香树生产力的有效评估。为解决这一问题,本研究提出一种基于多光谱图像多模态融合与改进 YOLO v8n 的轻量级检测算法(YOLO v8n improved for sandalwood plant seedlings detection, YOLO-SPS)。利用 SwinFusion 模型融合檀香树幼苗的近红外与可见光遥感图像,增强其纹理与颜色特征;基于 YOLO v8n 模型在骨干网络中引入结合空间重构与通道重构的模块(Cross stage partial network fusion_spatial and channel reconstruction convolution, C2f_SCConv),提升特征提取能力,在颈部网络中加入部分自注意力机制(Partial self-attention, PSA),优化全局感知能力并降低计算成本,采用 Shape-IoU 作为边界框损失函数,进一步提高检测精度。试验结果表明, YOLO-SPS 模型精确率、召回率和平均精度均值分别达到 92.8%、93.2% 和 95.9%,较原 YOLO v8n 模型性能显著提升,且优于 YOLO v5s、YOLO v6n、YOLO v7-tiny、YOLO v9-t、YOLO v10n 等 5 个主流检测模型。研究结果为檀香树幼苗期精准监测提供了有效的技术支持。

关键词: 檀香树幼苗; 多模态融合; 遥感图像; 目标检测; YOLO v8n

中图分类号: TP753

文献标识码: A

文章编号: 1000-1298(2026)10-0164-09

OSID:



Accurate Detection of Sandalwood Seedlings Based on Multi-modal Remote Sensing Image Fusion

ZHANG Yu¹ HUANG Zezhong¹ DONG Jinrang¹ WEI Wenzheng¹ ZHAO Yikun¹ TU Shuqin²

(1. College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China

2. College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: Sandalwood seedlings rely on companion plants during their growth, which are often interplanted with other crops, resulting in a complex growth environment that makes accurate detection challenging, thereby affecting the effective assessment of sandalwood productivity. To address this issue, a lightweight detection algorithm, (YOLO v8n improved for sandalwood plant seedlings detection, YOLO-SPS), was proposed based on multispectral image fusion and an improved YOLO v8n. The SwinFusion model was utilized to fuse near-infrared and visible light remote sensing images of sandalwood seedlings, enhancing their texture and color features. The cross stage partial network fusion_spatial and channel reconstruction convolution module, which combined spatial and channel reconstruction, was introduced into the backbone network of the YOLO v8n model to improve feature extraction capabilities. The partial self-attention (PSA) mechanism was incorporated into the neck network to optimize global perception and reduce computational costs. Finally, Shape-IoU was adopted as the bounding box loss function to further enhance detection accuracy. Experimental results demonstrated that the YOLO-SPS model achieved precision, recall, and mean average precision of 92.8%, 93.2%, and 95.9%, respectively, significantly outperforming the original YOLO v8n model and surpassing five mainstream detection models, including YOLO v5s, YOLO v6n, YOLO v7-tiny, YOLO v9-t, and YOLO v10n. The research result can provide effective technical support for the precise monitoring of sandalwood seedlings during their early growth stages.

Key words: sandalwood seedlings; multimodal fusion; remote sensing images; object detection; YOLO v8n

收稿日期: 2025-02-25 修回日期: 2025-04-02

基金项目: 2025 年度农村科技特派员专题(SL2024D04T00033)

作者简介: 张宇(1976—),女,副教授,主要从事计算机视觉研究, E-mail: zhangyu@scau.edu.cn

通信作者: 涂淑琴(1978—),女,讲师,主要从事计算机视觉研究, E-mail: tushuqin@163.com

0 引言

檀香树是一种半寄生常绿植物,具有很高的经济价值,生长周期一般为 10 ~ 15 年^[1-2]。在檀香树前 5 年生长期内,其生长状况直接影响后期生产力。然而,檀香树幼苗在生长初期植株特征不明显,且需要伴生植物及其他作物的穿插种植,导致其生长环境复杂,传统的人工监测方法难以实现精准、高效的生长状态评估。随着智能化算法和遥感技术的快速发展,基于无人机的多光谱遥感技术为檀香树幼苗的精准监测提供了新的解决方案。然而,现有方法主要依赖单一可见光检测,难以应对复杂背景下檀香树幼苗的精准检测需求,亟需一种高效、轻量化的检测方法。

近年来,基于深度学习的无人机遥感影像目标检测技术取得了显著进展。现有方法主要分为两阶段目标检测算法(如 R-CNN^[3-5] 系列)和单阶段目标检测算法(如 YOLO 系列)。两阶段算法通过生成候选区域并进行精细化定位,虽检测精度较高,但模型复杂、计算成本高,难以满足实时检测需求^[6-7]。相比之下,单阶段算法通过端到端的检测框架,在保证较高精度的同时显著提升了检测效率。YOLO 系列算法在农业领域的应用表现出色,如玉米雄穗检测^[8]和落叶松毛虫虫害识别^[9]等。然而,这些方法主要针对单一光谱数据,难以应对檀香树幼苗检测因复杂背景导致的检测精度下降问题。

针对上述问题,本研究提出一种改进 YOLO v8n 的檀香树幼苗检测模型 YOLO-SPS。通过 SwinFusion^[10] 模型融合可见光与近红外图像,构建 RGB-NIR 融合数据集,增强檀香树幼苗的纹理与颜色特征;在 YOLO v8n 骨干网络中引入 C2f-SCConv^[11] 模块,结合空间重构与通道重构机制,提升特征提取能力;在颈部网络中添加 PSA^[12] 注意力机制,优化全局感知能力并降低计算开销;采用 Shape-IoU^[13] 损失函数替代 CIoU^[14],综合考虑边界框的形状与尺度因素,进一步提升检测精度。以期在复杂背景下提升檀香树幼苗检测能力,为檀香树幼苗精准监测提供一种高效、轻量化的技术解决方案。

1 试验数据获取

1.1 数据来源

自建的檀香树数据集 2023 年 3 月采自广东省江门市台山市某檀香树种植园,檀香树种植园所在经纬度为 112.84°E、22.14°N。多光谱无人机为大疆 P4 Multispectral 型,如图 1a 所示。在数据采集过程中,设置无人机云台拍摄角度为垂直地面,航向重

叠率 80%,旁向重叠率 70%。拍摄高度分别为 5、10、20 m。经筛选后共获得 836 幅包含檀香树的有效图像,分辨率为 1 600 像素×1 300 像素。这些图像中,每幅可见光图像都对应有 1 幅近红外图像,如图 1c、1d 所示。不同高度的图像数量比例约为 7:2:1。

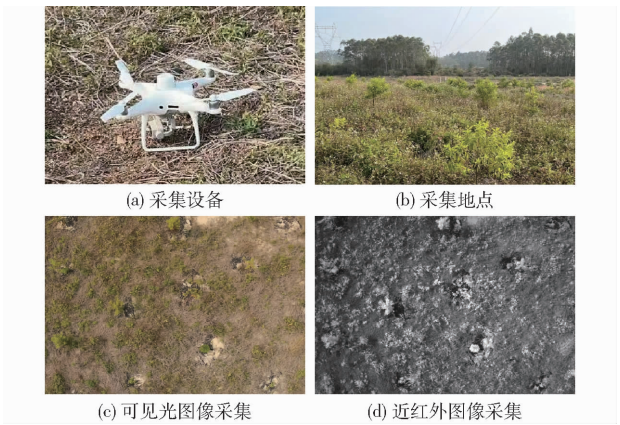


图 1 数据采集

Fig. 1 Data acquisition

1.2 多模态图像融合

针对无人机单可见光下采集的檀香树图像分辨率低,植株特征不明显以及与背景相似度高导致检测精度低的问题,将檀香树可见光与近红外 2 种不同模态的图像进行融合。可见光图像提供了丰富的颜色信息,但在檀香树植株与杂草颜色相似时,单凭可见光图像难以准确识别檀香树幼苗。而近红外图像因不同植物对叶绿素反射率敏感度不同,能有效区分檀香树与其他植被或背景,但由于缺少颜色信息,检测精度较低。为充分利用不同模态图像的特点,使用 SwinFusion 模型融合檀香树可见光模态与近红外模态图像。SwinFusion 模型网络结构如图 2 所示。

SwinFusion 模型通过卷积神经网络(Convolutional neural network, CNN)提取可见光图像 I_1 和近红外图像 I_2 的特征。通过多头自注意力机制(Multi-head self-attention, MSA)进行处理,将特征图转换为查询(Query, Q)、键(Key, K)和值(Value, V)。其中查询用于捕捉特征之间的关系,键用于与查询进行匹配,而值则提供与键相关的信息。通过这种机制,融合模型能够关注到特征图的重要区域,从而增强局部信息的表达。MSA 通过多个头并行计算,使得模型能够在不同子空间中学习特征之间的关系,提高特征表示的丰富性和多样性。经 MSA 处理后,特征图将通过层归一化(Layer normalization, LN)进行标准化,提高训练的稳定性 and 收敛速度。经过 MSA 和 LN 处理后的特征被送入前馈神经网络(Feed forward network, FFN),FFN 由

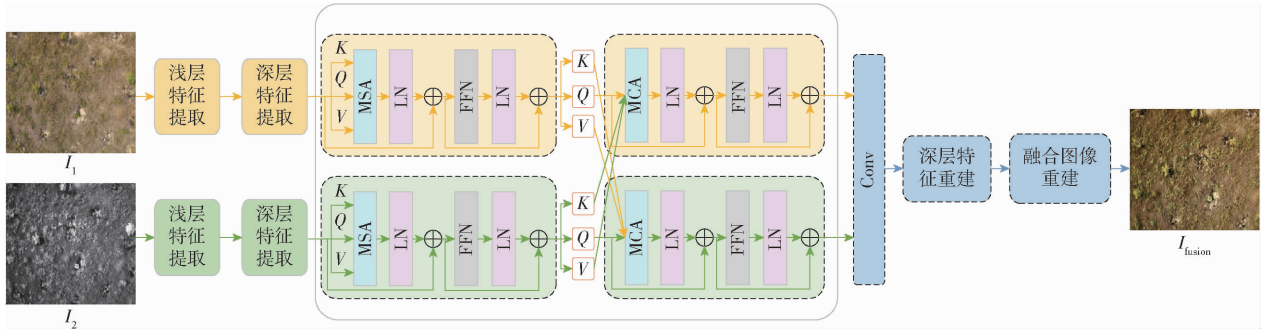


图 2 用于多模态图像融合的 SwinFusion 框架
Fig. 2 Framework of SwinFusion for multi-modal image fusion

2 个线性变换和一个高斯误差线性单元 (Gaussian error linerar units, GELU) 作为激活函数组成, 进一步增强特征表达能力。随后, 引入多头交叉注意力机制 (Multi-head cross-attention, MCA) 以实现可见光特征与近红外特征之间的融合。在这一过程中, 来自可见光图像的查询与近红外图像的键进行匹配, 从而提取互补信息。MCA 通过将一个域的特征作为查询, 另一个域的特征作为键和值, 使得模型能够有效整合 2 种图像的特征, 充分发挥各自的优势。最后, 融合后的特征通过重建模块被转换回图像空间, 生成最终的融合图像 I_{fusion} 。

1.3 数据集构建

对筛选后的原始图像实施数据增强策略, 包括图像翻转、镜像和随机平移等操作, 提升数据样本的多样性, 有效预防模型过拟合的风险, 最终获得檀香树可见光与近红外图像各 2 608 幅。

如图 3 所示, 使用 SwinFusion 模型融合可见光图像和近红外图像, 得到融合图像数据集。本研究构建 3 个独立的数据集, 檀香树可见光数据集, 檀香树近红外数据集以及檀香树融合图像数据集。为确保数据的一致性和可比性, 数据集中的每 1 幅可见光图像和近红外图像都是无人机在同一地点拍摄的。使用 LabelImg 工具对可见光数据集进行人工标注, 并按照比例 8: 1: 1 随机划分为训练集、验证集、测试集。近红外与融合图像数据集共用可见光数据集的标签文件, 并按照对应的可见光图像划分训练集、验证集与测试集。这种严格的对应关系保证了数据集之间的一致性, 可以精确地评估不同数

据集对检测檀香树的影响, 并探索更有效的检测技术。

2 檀香树检测模型

2.1 YOLO v8 模型

YOLO v8 是一个单阶段的实时目标检测模型, 包含 YOLO v8n、YOLO v8s、YOLO v8m、YOLO v8l 和 YOLO v8x 等版本。其中, YOLO v8n 采用轻量级设计, 降低了算法和存储需求, 使无人机能够更好地处理图像数据。YOLO v8n 模型网络结构主要由骨干网络 (Backbone)、颈部网络 (Neck) 和头部网络 (Head) 3 部分组成。

(1) 骨干网络是模型基础, 从原始输入数据中提取特征。在 YOLO v8 中, 骨干网络主要由 Conv、C2f 和 SPPF 模块构成。Conv 模块对输入的图像进行卷积、批归一化 (Batch normalization, BN) 和激活函数 SiLU 操作。C2f 模块借鉴了 YOLO v7 中的 ELAN 思想, 引入更多并行梯度流分支, 丰富梯度流信息, 同时减少计算量。相比 YOLO v5 中 C3 模块, C2f 去掉上采样之前的 1×1 卷积, 将 Backbone 不同阶段输出的特征直接送入上采样操作。SPPF 模块基于空间金字塔池化 (SPP) 改进, 将并行结构改为串行结构, 减少计算量并保持功能不变。

(2) 颈部网络位于骨干网络与头部网络之间, 主要进行特征融合和增强。在 YOLO v8 中, 颈部网络采用路径聚合网络 (Path aggregation network, PANet) 和特征金字塔网络 (Feature pyramid network, FPN) 的结合。PANet 通过自顶向下路径和

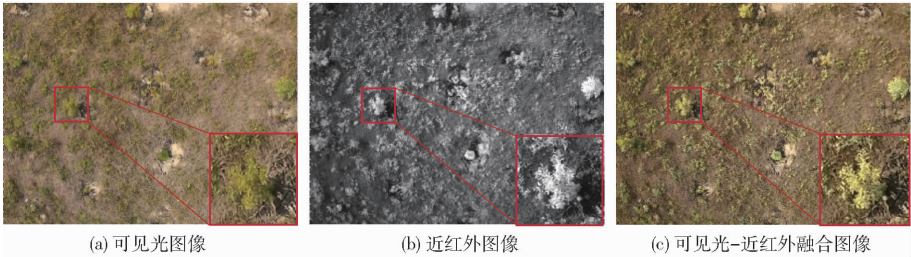


图 3 不同模态数据集示例
Fig. 3 Example of different modal datasets

自底向上路径的结合,增强了特征金字塔的语义信息传递能力,从而在不同尺度上更好地捕捉目标信息。FPN 通过构建多尺度特征金字塔,进一步提高了模型对不同尺寸目标的检测能力。

(3)头部网络是目标检测模型的决策部分,负责产生最终的检测结果。在 YOLO v8 中,头部网络采用 YOLO 系列的一贯设计,通过多个预测头在不同尺度上进行目标分类和边界框回归。每个预测头包括卷积层、分类层和回归层,分别用于特征提取、目标分类和边界框回归。这种设计使 YOLO v8 能

够在不同分辨率上检测出大小不同的目标,提高检测的准确性和鲁棒性。

2.2 YOLO v8n 模型改进

本研究采用 YOLO v8n 模型,为进一步提升模型对檀香树幼苗检测任务中的性能,采取 3 项改进措施。引入 C2f_SCConv 模块,增强模型骨干网络特征提取能力,并添加 PSA 注意力机制以及 Shape-IoU 损失函数,综合设计用于檀香树幼苗检测的 YOLO-SPS 模型,网络结构如图 4 所示。

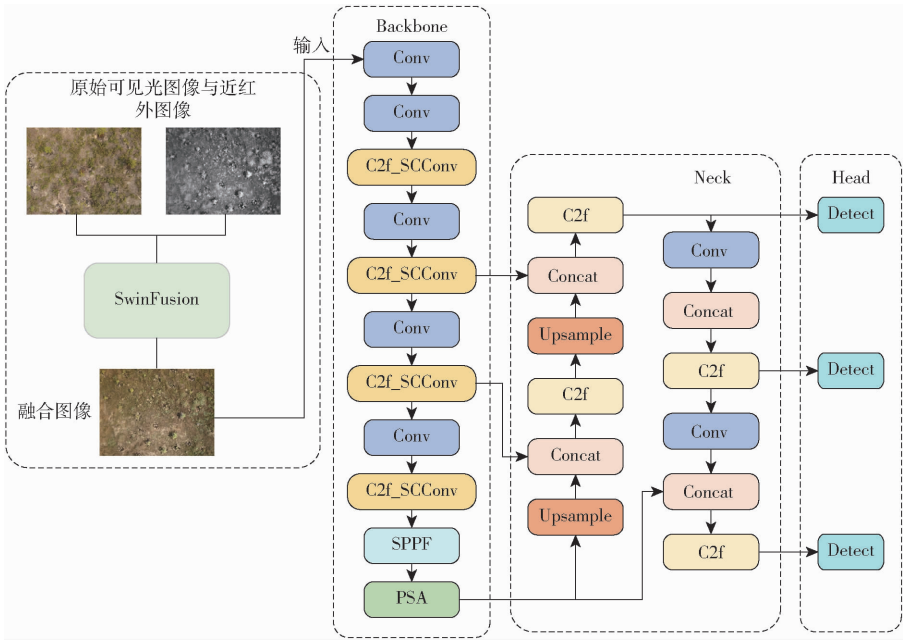


图 4 YOLO-SPS 模型结构
Fig.4 Model structure of YOLO-SPS

2.2.1 C2f_SCConv 模块

遥感图像檀香树植株遥感图像特征不明显,易受杂草等其他植物干扰,YOLO v8 中固有的 C2f 结构主要关注通道维度的特征融合,无法充分捕捉檀香树在空间维度上的细微纹理和形态特征,且檀香树遥感图像数据集由不同高度的数据混合构建,C2f 模块的固定通道分割方式缺乏足够的灵活性。因此在原有 YOLO v8n 模型骨干网络中,结合 C2f 模块及 SCConv 模块,引入 C2f_SCConv 替换 C2f 模块,结构如图 5 所示。

SCConv 模块主要由空间重构单元 (Spatial reconstruction unit, SRU) 和通道重构单元 (Channel

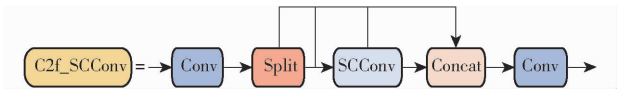


图 5 C2f_SCConv 结构
Fig.5 Structure of C2f_SCConv

reconstruction unit, CRU) 组成,具体结构如图 6 所示。通过 SRU 与 CRU 的组合,SCConv 模块能同时处理空间和通道维度的冗余,提高模型对檀香树的检测效率和准确性,同时降低计算复杂度,使模型更有效地提取檀香树的细微纹理特征和形态特征。

SRU 通过分离和重构操作减少特征图的空间冗余,如图 7 所示,是一个给定中间特征图,其中 N 为批量轴, C 为通道轴, H 和 W 为输入特征图的高

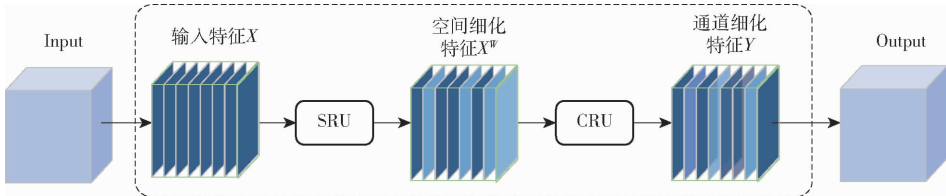


图 6 SCConv 结构
Fig.6 SCConv structure

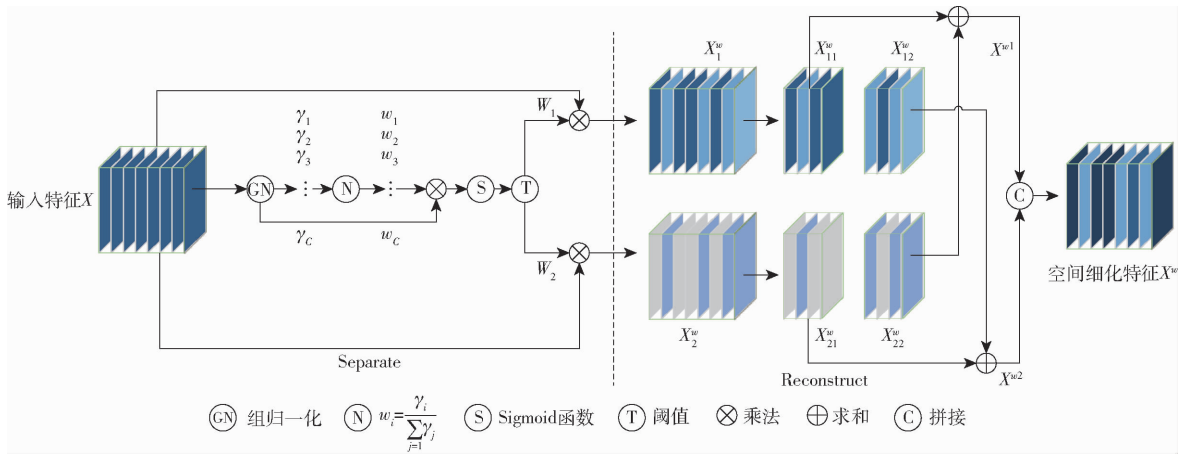


图 7 SRU 结构

Fig. 7 SRU structure

度和宽度,首先通过组归一化(GN)处理得到 X_{out} 为

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{1}$$

式中 μ, σ —— X 均值和标准差
GN (X)——对输入特征 X 执行 Group Normalization 操作
 ε ——确保除法稳定性小的正常数
 β ——仿射变换训练参数
归一化相关权重计算式为

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^c \gamma_j} \quad (i, j = 1, 2, \dots, C) \tag{2}$$

式中 γ_i ——第 i 个通道对应的 GN 层缩放因子
 C ——输入特征图通道数
 $\sum_{j=1}^c \gamma_j$ ——所有通道缩放因子总和

然后,通过 Sigmoid 函数将特征图的权重映射到 0 ~ 1 范围,并通过阈值进行门控。权重高于阈值的设为 1,否则设为 0,得到非信息权重。计算式为

$$W = \text{Gate}(\text{Sigmoid}(W_\gamma(\text{GN}(X)))) \tag{3}$$

式中 Gate($\cdot$)——设定阈值为 0.5 的二值化操作
最后,将输入特征 X 分别乘以 W_1 和 W_2 ,得到信

息丰富的 X_1^w 和信息较少的 X_2^w ,为减少空间冗余,通过交叉重构操作,具有丰富信息的特征与较少信息的特征充分结合,生成更具代表性和表现力的特征,同时抑制空间维度上的冗余特征。

获得空间优化的特征图计算式为

$$\begin{cases} X_1^w = W_1 \otimes X \\ X_2^w = W_2 \otimes X \\ X_{11}^w \oplus X_{22}^w = X^{w1} \\ X_{21}^w \oplus X_{12}^w = X^{w2} \\ X^{w1} \cup X^{w2} = X^w \end{cases} \tag{4}$$

式中 $\otimes$ ——元素级乘法
 $\oplus$ ——元素级求和
 $\cup$ ——特征连接操作

如图 8 所示,CRU 通过分割-变换-融合策略来减少特征的通道冗余。将输入特征 X 在通道维度上分割为上部分 X_{up} 和下部分 X_{low} 。 X_{up} 通过组卷积(Group-wise convolution, GWC)和逐点卷积(Point-wise convolution, PWC)进行变换,提取出具有丰富信息和代表性的特征 Y_1 ,计算式为

$$Y_1 = M^C X_{\text{up}} + M^{P1} X_{\text{up}} \tag{5}$$

式中 M^C, M^{P1} ——可学习的权重矩阵

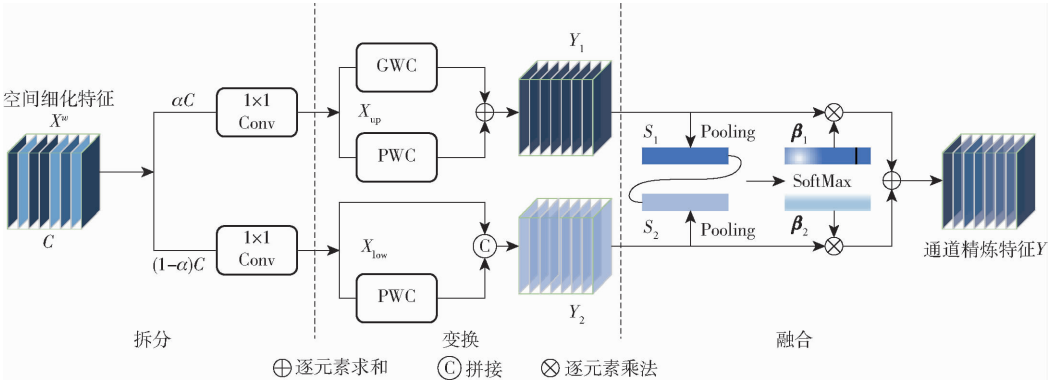


图 8 CRU 结构

Fig. 8 CRU structure

下部分 X_{low} 通过廉价的 1×1 逐点卷积生成补充特征 Y_2 , 计算式为

$$Y_2 = M^{P_2} X_{\text{low}} \cup X_{\text{low}} \quad (6)$$

式中 M^{P_2} ——下支路 1×1 逐点卷积的卷积核权重

通过简化的 SKNet 方法自适应地融合 Y_1 和 Y_2 , 得到通道细化特征 Y , 计算式为

$$Y = \beta_1 Y_1 + \beta_2 Y_2 \quad (7)$$

式中 β_1, β_2 ——特征重要性向量

2.2.2 PSA 注意力机制

PSA 注意力机制是一种优化的自注意力框架, 结构如图 9 所示。输入特征图被划分为 2 部分, 一部分被送入多头自注意力 (MSA) 模块。在 MSA 模块内, 通过计算查询、键和值, 并利用缩放点积得出注意力分数, 再经过 SoftMax 归一化处理, 突出显示特征间的相互依赖性。归一化的注意力分数与值结合后, 进行特征融合, 以获得更丰富的特征表示。这些特征被送入前馈网络 (FFN), FFN 由 2 个线性变换和 1 个 ReLU 激活函数组成, 进一步提炼特征并提升模型的非线性建模能力。FFN 输出与未经 MSA 处理的特征图另一部分进行拼接、融合, 最终通过 1 个 1×1 卷积层整合信息, 输出最终的特征图, 这些特征图包含了更丰富的上下文信息和增强的表示能力, 为檀香树幼苗检测任务提供了更准确的特征支持。

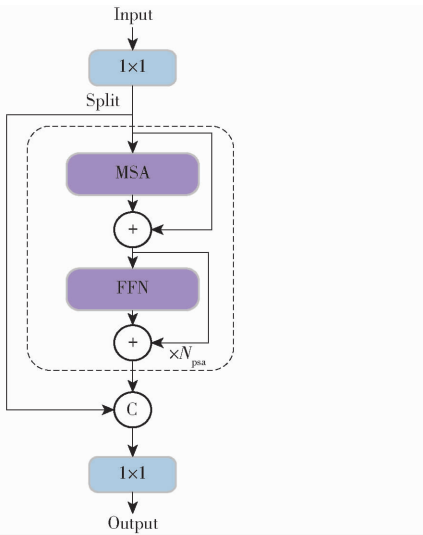


图 9 PSA 注意力模块结构

Fig. 9 PSA attention module architecture

2.2.3 Shape-IoU 改进

传统的 YOLO v8n 模型使用 CIoU 作为损失函数, 计算式为

$$CIoU = \frac{IoU - \rho^2(b, b^{gt})}{c^2} - \alpha \nu \quad (8)$$

$$\alpha = \frac{\nu}{1 - IoU + \nu} \quad (9)$$

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2 \quad (10)$$

式中 $\rho^2(b, b^{gt})$ ——预测框和真实框的中心点 b 和 b^{gt} 之间的欧氏距离

c ——预测框与真实框最小外接矩形对角线长度

ν ——修正因子

CIoU 方法主要考虑边界框之间的几何关系, 通过计算边界框的相对位置和形状来计算损失。然而, CIoU 忽略了边界框本身的形状和尺度因素对回归结果的影响。实际上, 不同的目标可能具有不同的形状和尺度, 忽略这些因素可能导致回归结果不准确。

Shape-IoU 可以计算边界框形状和尺度之间差异, 并将这个差异作为损失函数的一部分, Shape-IoU 方法能够更准确地回归目标的位置和形状。

Shape-IoU 损失函数计算式为

$$L_{\text{Shape-IoU}} = 1 - IoU + distance^{\text{shape}} + 0.5 \Omega^{\text{shape}} \quad (11)$$

$$IoU = \frac{B \cap B^{gt}}{B \cup B^{gt}} \quad (12)$$

$$distance^{\text{shape}} = h_h \frac{(x_c - x_c^{gt})^2}{c^2} + w_w \frac{(y_c - y_c^{gt})^2}{c^2} \quad (13)$$

$$\Omega^{\text{shape}} = \sum_{t=w, h} (1 - e^{-w_t})^\theta \quad (\theta=4) \quad (14)$$

$$\begin{cases} \omega_w = h_h \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = w_w \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (15)$$

式中 B ——预测框 B^{gt} ——真实框

x_c, y_c ——预测框中心坐标

x_c^{gt}, y_c^{gt} ——真实框中心坐标

$distance^{\text{shape}}$ ——考虑边界框形状和尺度中心点距离的距离项

h_h, w_w ——垂直和水平方向权重系数, 与真实框形状相关

Ω^{shape} ——考虑边界框形状和尺度的角度差异项

θ ——超参数, 用于控制形状差异对损失函数的敏感度

ω_w, ω_h ——宽度和高度的形状损失因子

w, h ——预测框宽度和高度

w^{gt}, h^{gt} ——真实框宽度和高度

2.3 模型训练与评价指标

2.3.1 试验环境与训练参数

用于训练的操作系统是 Windows Service, CPU 型号为 Intel(R) Core(TM) i9-12900K, GPU 型号

为 NVIDIA GeForce RTX 3090 24 G。训练阶段不采用预训练权重,输入图像分辨率为 640 像素 × 640 像素,总训练轮数(epoch)设置为 300,初始学习率为 0.01,采用随机梯度下降(SDG)进行优化,批处理(batchsize)大小为 16,其他参数采用默认值。

2.3.2 模型评价指标

为验证模型性能,选择精确率、召回率、平均精度均值(mAP)、模型参数量(Parameters)和浮点运算量(GFLOPs)等作为评价指标。

3 结果与分析

3.1 试验结果

3.1.1 不同模型在不同模态数据集下性能试验

为评估不同模态数据集对檀香树幼苗检测性能的影响,以 YOLO v8n 为基线模型,分别对 3 种模型在不同模态的数据集分别训练与测试,试验结果见图 10 和表 1。由表 1 可知,融合图像数据集训练的模型在檀香树幼苗检测任务上取得了最佳效果,精确率为 88.8%,召回率为 87.9%,平均精度均值为 93.1%;相较于可见光数据集,精确率、召回率、平均精度均值分别提高 4.7、2.1、2.9 个百分点。相较于近红外数据集,精确率、召回率、平均精度均值分别提高 15、12.8、10.4 个百分点。综上,融合图像数据集能够综合利用可见光和近红外图像的优势,增强了模型对幼苗特征的识别能力。因此,后续试验均使用融合图像数据集进行训练。

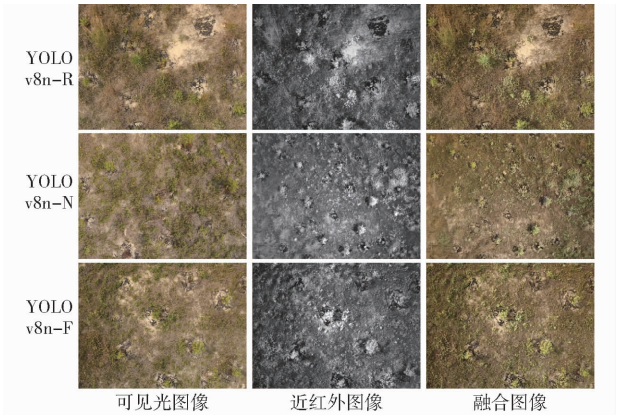


图 10 不同模态数据集

Fig. 10 Different modal datasets

表 1 不同模型在不同模态数据集下性能试验结果

Tab.1 Experimental results of different models on multimodal datasets				
模型	数据集	精确率	召回率	平均精度均值
YOLO v8n-R	可见光数据集	84.1	85.8	90.2
YOLO v8n-N	近红外数据集	73.8	75.1	84.3
YOLO v8n-F	融合图像数据集	88.8	87.9	93.1

3.1.2 不同主干网络、注意力机制与损失函数性能试验

为检验主干网络引入 C2f_SCCConv 的有效性,在主干网络中分别引入 SPDConv^[15]、AKConv^[16] 以及 RFACConv^[17] 进行试验,试验结果如表 2 所示。由表 2 可知,C2f_SCCConv 模块在保持低参数量和计算量的同时,精确率、召回率及 mAP 均有提升。因此,C2f_SCCConv 能够有效提升模型整体性能。

表 2 主干网络改进模块性能试验结果

Tab.2 Performance evaluation results of backbone network improvement modules

模型	精确率/%	召回率/%	平均精度均值/%	模型参数量	浮点运算量
YOLO v8n-F	88.8	87.9	93.1	3.2 × 10 ⁶	8.3 × 10 ⁹
YOLO v8n-SPDConv	89.0	86.8	93.7	4.4 × 10 ⁶	1.0 × 10 ¹⁰
YOLO v8n-AKConv	87.7	88.5	94.0	2.7 × 10 ⁶	7.5 × 10 ⁹
YOLO v8n-RFACConv	88.6	87.4	94.2	3.2 × 10 ⁶	8.1 × 10 ⁹
YOLO v8n-C2f_SCCConv	89.1	90.2	94.5	2.7 × 10 ⁶	7.1 × 10 ⁹

为探究 PSA 注意力机制在檀香树幼苗检测任务中性能,将几种常见注意力 SE^[18]、CBAM^[19]、SimAM^[20] 分别添加到同一位置与 PSA 注意力机制进行试验,试验结果如表 3 所示。由表 3 可知,4 种注意力的添加均能使模型性能提升,其中,添加 PSA 注意力的模型精确率、召回率、平均精度均值处于领先地位,分别达到 89.6%、89.8%、94.1%,证明了其在檀香树幼苗检测任务中的有效性。

表 3 不同注意力机制性能试验结果

Tab.3 Performance evaluation results of different attention mechanisms

模型	精确率/%	召回率/%	平均精度均值/%	模型参数量	浮点运算量
YOLO v8n-SE	89.4	88.6	93.4	3.2 × 10 ⁶	7.9 × 10 ⁹
YOLO v8n-CBAM	88.3	87.6	93.8	3.6 × 10 ⁶	8.2 × 10 ⁹
YOLO v8n-SimAM	88.4	88.5	93.7	3.0 × 10 ⁶	8.1 × 10 ⁹
YOLO v8n-PSA	89.6	89.8	94.1	3.0 × 10 ⁶	7.8 × 10 ⁹

为探究不同损失函数对模型性能的影响,分别采用 CIoU、SIoU^[21]、WIoU^[22]、MPDIoU^[23] 和 Shape-IoU 作为损失函数进行对比试验,试验结果如表 4 所示。由表 4 可知,在其他参数设置相同时,SIoU、Wise-IoU、MPDIoU、Shape-IoU 分别使平均精度均值相较原模型提升 1.2、0.3、0.7、1.6 个百分点。由此可看出,将 Shape-IoU 作为模型的损失函数,能够更准确地捕捉檀香树幼苗的形状变化和尺度差

表 4 不同损失函数性能试验结果

Tab.4 Performance evaluation results of different loss functions			%
损失函数	精确率	召回率	平均精度均值
CIoU	88.7	87.9	93.1
SIoU	86.3	87.1	94.3
WIoU	89.3	89.8	93.4
MPDIoU	91.1	92.3	93.8
Shape-IoU	92.8	93.2	94.7

异,有效提高模型检测精度。

3.1.3 消融试验

为进一步验证本文模型 YOLO-SPS 各改进模块的有效性,对各模块进行消融试验。其中 YOLO

v8n-F 为基线模型,模型 1、2、3 分别为在骨干网络中添加 C2f_SCConv、PSA 注意力机制、Shape-IoU 损失函数。模型 4 为在骨干网络中引入 C2f_SCConv 模块的同时添加 PSA 注意力机制,试验结果如表 5 所示。由表 5 可知,在 YOLO v8n 模型骨干网络中引入 C2f_SCConv 模块后,模型精确率、召回率、平均精度均值相对于基线模型分别提高 0.3、2.3、1.4 个百分点,再添加 PSA 注意力机制,模型的精确率、召回率、平均精度均值相较于基线模型分别提高 2.3、2.6、2.0 个百分点,相对于基线模型,YOLO-SPS 在保持轻量化的同时精确率、召回率、平均精度均值分别提高 4.0、5.3、2.8 个百分点,证明了各改进方法的有效性。

表 5 消融试验结果

Tab.5 Results of ablation test

模型	C2f_SCConv	PSA	Shape-IoU	精确率/%	召回率/%	平均精度均值/%	模型参数量	浮点运算量
YOLO v8n-F	×	×	×	88.8	87.9	93.1	3.2×10^6	8.3×10^9
1	√	×	×	89.1	90.2	94.5	2.7×10^6	7.1×10^9
2	×	√	×	89.6	89.8	94.1	3.0×10^6	7.8×10^9
3	×	×	√	89.0	89.5	94.7	3.2×10^6	8.3×10^9
4	√	√	×	91.1	90.5	95.1	2.8×10^6	7.0×10^9
YOLO-SPS	√	√	√	92.8	93.2	95.9	2.8×10^6	7.0×10^9

注:√表示采用此模块(方法),×表示不采用此模块(方法)。

3.1.4 不同检测模型性能试验

为验证本文模型 YOLO-SPS 在檀香树幼苗检测的有效性,与 YOLO 系列中当前流行的轻量级目标检测算法(YOLO v5s、YOLO v6n^[24]、YOLO v7-tiny^[25]、YOLO v9-t^[26]、YOLO v10n)进行对比试验,试验结果如表 6 所示。其中 YOLO v8n-R 为在檀香树可见光数据集上进行训练的模型,其他模型均在檀香树幼苗融合图像数据集上进行训练。

表 6 不同检测模型性能试验结果

Tab.6 Performance evaluation results of different detection models					
模型	精确率/%	召回率/%	平均精度均值/%	模型参数量	浮点运算量
YOLO v8n-R	84.1	85.8	90.2	3.2×10^6	8.3×10^9
YOLO v8n-F	88.8	87.9	93.1	3.2×10^6	8.3×10^9
YOLO v5s	86.3	84.3	92.4	7.0×10^6	7.1×10^9
YOLO v6n	84.2	84.0	91.4	4.2×10^6	1.2×10^{10}
YOLO v7-tiny	84.4	83.8	91.7	6.0×10^6	1.3×10^{10}
YOLO v9-t	90.8	91.6	94.0	2.0×10^6	7.7×10^9
YOLO v10n	91.7	92.3	95.3	2.7×10^6	8.2×10^9
YOLO-SPS	92.8	93.2	95.9	2.8×10^6	7.0×10^9

由表 6 可知,YOLO v8n-F 与 YOLO v8n-R 相比,精确率、召回率、平均精度均值分别提升 4.7、2.1、2.9 个百分点,原因为使用 SwinFusion 模型融

合檀香树幼苗可见光与近红外图像后,融合图像有效保留并整合了不同模态图像中的互补信息和细节特征,从而提升檀香树幼苗辨识度和清晰度,使得模型能够更准确地识别檀香树幼苗。YOLO-SPS 在保证模型轻量化的同时,精确率、召回率、平均精度均值均取得了最好表现。与 YOLO v5s、YOLO v6n、YOLO v7-tiny、YOLO v8n、YOLO v9-t 相比,YOLO-SPS 的平均精度均值分别提升 3.5、4.5、4.2、2.8、1.9 个百分点。与 YOLO v10n 相比,YOLO-SPS 在参数量相近的情况下,浮点运算量仅为 7.0×10^9 ,同时平均精度均值提升 0.6 个百分点。

3.1.5 检测热力图分析

热力图能直观展示不同检测模型对目标区域的关注程度,通过颜色强度分布反映模型特征提取能力和定位精度。由图 11 可知,在 YOLO v8n 基线模型中,基于单可见光检测模型 YOLO v8n-R 和基于融合图像检测模型 YOLO v8n-F 均存在漏检现象,但 YOLO v8n-F 模型表现优于 YOLO v8n-R。而 YOLO-SPS 模型目标检测热力图与真实檀香树幼苗区域的吻合度显著提高。结果表明 YOLO-SPS 模型检测效果优于原始模型,验证了本文模型在檀香树幼苗检测任务中具有更优越的性能。

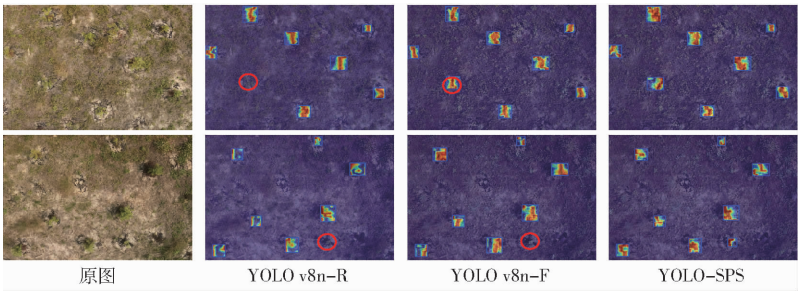


图 11 检测热力图

Fig. 11 Detection heatmap

4 结论

(1)针对檀香树幼苗在遥感图像中精准检测问题,提出了一种基于多模态遥感图像融合的高精度检测模型 YOLO-SPS,通过融合近红外与可见光图像,显著增强了檀香树幼苗的纹理与颜色特征,提高了复杂背景下幼苗的可识别性。

(2)引入 C2f_SCConv 模块、PSA 注意力机制以及 Shape-IoU 损失函数,显著提升模型特征提取能力、全局感知能力和检测精度,同时降低计算成本。

(3)YOLO-SPS 模型在檀香树融合图像验证集上表现优异,精确率、召回率和检测精度均优于其他主流轻量化模型,展现了更高的检测性能。

参 考 文 献

[1] ARUNKUMAR A, JOSHI G, RAM H M. Sandalwood: history, uses, present status and the future[J]. Current Science, 2012,103(12): 1408-1416.

[2] 张宇,徐浩然,牛家俊,等.基于 YOLO v4 和双重回归的复杂环境檀香树幼苗定位方法[J].农业机械学报,2022,53(11): 299-305,340.

ZHANG Yu, XU Haoran, NIU Jiajun, et al. Missing seedling localization method for sandalwood trees in complex environment based on YOLO v4 and double regression strategy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(11):299-305,340. (in Chinese)

[3] 朱红春,李旭,孟炀,等.基于 Faster R-CNN 网络的茶叶嫩芽检测[J].农业机械学报,2022,53(5):217-224.

ZHU Hongchun, LI Xu, MENG Yang, et al. Tea bud detection based on Faster R-CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5):217-224. (in Chinese)

[4] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.

[5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.

[6] LIU Y, CEN C, CHE Y, et al. Detection of maize tassels from UAV RGB imagery with Faster R-CNN[J]. Remote Sensing, 2020, 12(2):338.

[7] CUI J, ZHANG X, ZHANG J, et al. Weed identification in soybean seedling stage based on UAV images and Faster R-CNN[J]. Computers and Electronics in Agriculture, 2024, 227: 109533.

[8] 马中杰,罗晨,骆巍,等.基于改进 YOLO v7-tiny 的玉米种质资源雄穗检测[J].农业机械学报,2024,55(7):290-297.

MA Zhongjie, LUO Chen, LUO Wei, et al. Tassel detection method of maize germplasm resources based on improved YOLO v7-tiny[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7):290-297. (in Chinese)

[9] 林文树,张金生,何乃磊.基于改进 YOLO v4 的落叶松毛虫侵害树木实时检测方法[J].农业机械学报,2023,54(4): 304-312,393.

LIN Wenshu, ZHANG Jinsheng, HE Nailei. Real-time detection method of dendrolimus superans-infested Larix gmelinii trees based on improved YOLO v4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(4):304-312,393. (in Chinese)

[10] 李路,孙超奇,周玉凡,等.基于 FAS-YOLO v8n 的爬岸上草小龙虾多源图像融合识别方法[J].农业机械学报,2025, 56(8):526-534.

LI Lu, SUN Chaoqi, ZHOU Yufan, et al. Multi-source image fusion recognition of crawling bank grass crayfish based on FAS-YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(8):526-534. (in Chinese)

[11] LI J, WEN Y, HE L. SCConv: spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023:6153-6162.

[12] WANG A, CHEN H, LIU L, et al. YOLO v10: real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.

[13] ZHANG H, ZHANG S. Shape-iou: more accurate metric considering bounding box shape and scale[J]. arXiv preprint arXiv: 2312.17663, 2023.

Agronomy, 2022, 12(6): 1459.

[23] YANG G L, WANG J X, NIE Z L, et al. A Lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention[J]. Agronomy-basel, 2023, 13(7): 1824.

[24] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[25] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)CVPR, 2018.

[26] PIOTR D, RON A, SERGE B, et al. Fast feature pyramids for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1532 – 1545.

[27] CHEN J, JI C, ZHANG J, et al. A method for multi-target segmentation of bud-stage apple trees based on improved YOLOv8 [J]. Computers and Electronics in Agriculture, 2024, 220: 108876.

[28] TONG Z, CHEN Y, XU Z, et al. Wise – IoU: bounding box regression loss with dynamic focusing mechanism[J]. ArXiv, 2023, 1:10051.

[29] 罗志聪, 何陈涛, 陈登捷, 等. 基于轻量化 YOLO v8s – GD 的自然环境下百香果快速检测模型[J]. 农业机械学报, 2024, 55(8): 291 – 300.

LUO Zhicong, HE Chentao, CHEN Dengjie, et al. Passion fruit rapid detection model based on lightweight YOLO v8s – GD [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(8): 291 – 300. (in Chinese)

[30] 翟肇裕, 张梓涵, 徐焕良, 等. YOLO 算法在动植物表型研究中应用综述[J]. 农业机械学报, 2024, 55(11): 1 – 20.

ZHAI Zhaoyu, ZHANG Zihan, XU Huanliang, et al. Review of applying YOLO family algorithms to analyze animal and plant phenotype[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 1 – 20. (in Chinese)

[31] LIU Y, TANG L, LIAO C, et al. Optimized dropkey-based Grad – CAM: toward accurate image feature localization[J]. Sensors, 2023, 23(20): 8351.

(上接第 172 页)

[14] ZHENG Z, WANG P, LIU W, et al. Distance – IoU loss: faster and better learning for bounding box regression[J]. arXiv preprint arXiv:1911.08287, 2019.

[15] SUNKARA R, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2022: 443 – 459.

[16] ZHANG X, SONG Y, SONG T, et al. LDConv: linear deformable convolution for improving convolutional neural networks[J]. Image and Vision Computing, 2024(Sep.):149.

[17] 苑迎春, 耿俊, 许楠, 等. 基于 YOLO v8 – TRP 模型的小麦麦穗识别方法[J]. 农业机械学报, 2025, 56(11): 499 – 508.

YUAN Yingchun, GENG Jun, XU Nan, et al. Wheat ear recognition method using YOLO v8 – TRP model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(11): 499 – 508. (in Chinese)

[18] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 – 7141.

[19] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.

[20] 李尚平, 郑创锐, 文春明, 等. 基于无人机 RGB 图像与改进 YOLO v5s 的宿根蔗缺苗定位方法[J]. 农业机械学报, 2024, 55(12): 57 – 70.

LI Shangping, ZHENG Chuangrui, WEN Chunming, et al. Method for locating missing ratoon sugarcane seedlings based on RGB images from unmanned aerial vehicles and improve YOLO v5s[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(12): 57 – 70. (in Chinese)

[21] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[J]. Computer Vision and Pattern Recognition, 2022(5): 1 – 12.

[22] TONG Zanjia, CHEN Yuhang, XU Zewei, et al. Wise – IoU: bounding box regression loss with dynamic focusing mechanism [J]. arXiv preprint arXiv:2301.10051, 2023.

[23] MA S, XU Y. MPDIoU: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.

[24] LI C, LI L, JIANG H, et al. YOLO v6: a single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.

[25] WANG C Y, BOCHKOVSKJ Y A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464 – 7475.

[26] WANG C Y, YE H I, MARK L H Y. YOLO v9: learning what you want to learn using programmable gradient information [C]//European Conference on Computer Vision, 2024: 1 – 21.