

基于改进 YOLO v8 的轻量化棉铃识别模型与产量预测方法研究

刘祥^{1,2} 项若雪^{1,2} 班成龙^{1,2} 田敏^{1,2} 谭明天^{1,2} 黄凯文^{1,2}

(1. 石河子大学机械电气工程学院, 石河子 832003; 2. 农业农村部西北农业装备重点实验室, 石河子 832003)

摘要: 单株总铃数是棉花重要的表型性状之一,也是种植者估算棉花产量的重要参考因素。因此,从真实复杂的棉田图像中高效准确地识别棉花,对于确保棉花产业生产的经济效益和增强农业管理至关重要。然而,许多现有的卷积神经网络在棉花识别方面优先考虑准确性,缺乏了对识别效率的关注。因此,以脱叶期新疆密植棉花为对象,提出了一种改进的轻量化 YOLO(IML-YOLO)棉铃快速识别模型。IML-YOLO 模型结合了轻量化卷积特征提取和 YOLO 模型实时快速识别的优势,构建了一种全新的 RepGhostCSPELAN 轻量化模块,同时为了降低由轻量化带来的模型识别精度下降的问题,结合 CAHSFPN 特征融合提高对不同尺度棉铃的识别精度,还提出了一种 Focaler-MPDIoU 损失函数,有效提高了模型的识别精度。通过消融试验和可解释性分析证实了这些设计的有效性和显著性。与基准 YOLO v8n 模型相比,IML-YOLO 模型在浮点运算次数、模型内存占用量和参数量方面分别显著降低了 32.1%、47.5% 和 50%,同时平均精确度提升了 10.1 个百分点。将 IML-YOLO 模型应用于棉花产量预测,平均相对误差为 7.22%。该模型为棉铃检测算法与产量预测提供了新途径,为棉花智能化管理提供了技术支持。

关键词: 棉铃检测; 产量预测; YOLO v8; 特征融合; 无人机遥感

中图分类号: S126; TP79

文献标识码: A

文章编号: 1000-1298(2025)05-0130-11

OSID:



Lightweight Cotton Boll Detection Model and Yield Prediction Method Based on Improved YOLO v8

LIU Xiang^{1,2} XIANG Ruoxue^{1,2} BAN Chenglong^{1,2} TIAN Min^{1,2} TAN Mingtian^{1,2} HUANG Kaiwen^{1,2}

(1. College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832003, China

2. Key Laboratory of Northwest Agricultural Equipment, Ministry of Agriculture and Rural Affairs, Shihezi 832003, China)

Abstract: Cotton boll count is a critical phenotypic trait for estimating cotton yield and plays a vital role in precision agricultural management. However, accurately detecting cotton bolls in densely planted fields remained challenging due to complex backgrounds, occlusion, and varying illumination conditions. High-resolution UAV imagery was employed to capture cotton field scenes in a densely planted area of Xinjiang. A comprehensive dataset was developed through image segmentation and augmentation techniques, ensuring diverse representations of field conditions. To address the trade-off between detection accuracy and computational efficiency, an improved lightweight detection model IML-YOLO was proposed. The model integrated a novel GRGCE module that combined efficient ghost convolution with a RepGhostCSPELAN structure for feature extraction, a CAHSFPN feature fusion mechanism to enhance multi-scale representation, and a Focaler-MPDIoU loss function to refine localization accuracy. Extensive experiments demonstrated that IML-YOLO reduced computational complexity by 32.1%, decreased model size by 47.5%, and lowered parameter count by 50% compared with that of the baseline YOLO v8n, while boosting mean average precision by 10.1 percentage points. Furthermore, when applied to cotton yield prediction, the model achieved an average relative error of only 7.22%. These findings indicated that the proposed IML-YOLO model and yield prediction methodology can offer an effective solution for real-time cotton boll detection and significantly contribute to the advancement of intelligent cotton management.

Key words: cotton boll detection; yield prediction; YOLO v8; feature fusion; UAV remote sensing

收稿日期: 2025-01-27 修回日期: 2025-03-04

基金项目: 新一代人工智能国家科技重大专项(2022ZD0115803)、国家重点研发计划项目(2022YFD2002400)和兵团科技攻关计划项目(2023AB014)

作者简介: 刘祥(1998—),男,博士生,主要从事农业计算机视觉技术研究,E-mail: lx@stu.shzu.edu.cn

通信作者: 田敏(1968—),男,教授,博士生导师,主要从事农业信息化研究,E-mail: tm_inf@shzu.edu.cn

0 引言

棉花产量预测精度直接关系到供应链的稳定和农业经济效益^[1-2]。在棉花产量预测中,棉铃数量作为关键指标,不仅是产量估算的核心参数,更是优化种植管理、实现精准农业的重要依据。然而,传统的产量预测方法主要依赖人工采样和经验观察,存在效率低下、主观性强、覆盖面有限等固有缺陷,难以满足现代农业对精准化、智能化的需求。随着人工智能技术的快速发展,机器学习方法在棉花产量预测领域的应用显著提升了预测精度,为农业现代化提供了新的技术支撑^[3]。尽管如此,现有技术 在棉铃检测中仍面临诸多挑战,尤其是在复杂田间环境和密植种植条件下,检测精度和效率难以兼顾,严重制约了产量预测的可靠性和实用性。

在棉铃检测技术发展过程中,颜色、纹理和空间特征等传统图像处理方法发挥了重要作用^[4-6]。然而,这些传统方法在复杂田间环境中面临着光照变化、阴影干扰、背景混杂等严峻挑战,检测精度和鲁棒性难以满足实际应用需求。特别是在新疆等高密度种植区域,棉铃分布密集且相互遮挡,现有算法难以实现精准定位和计数,严重限制了产量预测的精度和效率。

近年来,无人机高分辨率图像采集技术与深度学习算法的结合为棉铃检测开辟了新途径。文献[7]提出的CD-SegNet模型能够有效分割棉花像素,但在棉铃计数方面存在局限;文献[8]通过结合SAM和YOLO算法实现了精确定位,显著提升了检测准确性;文献[9]对YOLO模型进行改进,引入时空上下文信息以增强模型鲁棒性,但在处理密植棉田中的遮挡和干扰问题时仍面临挑战。此外,现有深度学习模型往往在提高精度的同时牺牲了效率,难以满足实时检测需求,且在不同环境条件下的泛化能力不足,限制了其实际应用价值。

针对上述挑战,本文提出一种轻量级棉铃检测方法——IML-YOLO,旨在解决复杂环境下棉铃检测的精度、效率和泛化能力问题。本文提出RGCE(RepGhostCSPoELAN)特征提取模块,通过结构优化和特征复用,提升模型在复杂环境下的特征提取能力,同时降低参数量和计算成本;并设计CAHSFPN特征融合机制,通过多层次特征融合和上下文信息增强,以提升模型对密集目标的区分能力,并提高检测精度;引入FMPDI损失函数,通过优化目标检测的损失计算方式,增强模型在不同环境下的适应性和鲁棒性。

1 数据集建立与模型构建

1.1 数据采集

1.1.1 无人机遥感影像获取与预处理

本文以新疆密植棉花为研究对象,在新疆塔城地区沙湾县121团(44°92'N,85°40'E)进行实地数据采集。该区域地势平坦,土壤肥沃,以高密度种植棉花为主,具备典型的密植棉田特征。数据采集于2023年10月1—7日进行,期间晴天3d,多云2d,阴天2d。考虑到天气状况可能影响图像的光照条件,数据采集在14:00—16:00进行,以减少光照角度变化带来的影响。

本研究采用DJI Matrice 350 RTK型无人机搭载Zenmuse H20T型多传感器相机进行数据采集。无人机飞行高度设定为10m,航线规划采用预设航点模式,影像侧重于垂直俯拍,以确保数据的一致性和高分辨率。相机参数设置如下:分辨率4056像素×3040像素。飞行过程中,相邻影像的横向与纵向重叠率分别为70%和80%,确保图像拼接和特征识别的精度。在棉田内部,采用五点采样法随机选取5个样方点,每个样方尺寸为1m×1m,并在不同气象条件下重复采集。最终,共计获得500幅高分辨率棉田图像,涵盖不同光照和阴影条件,以增强数据的泛化能力。

1.1.2 数据集制作

为了便于模型训练同时充分利用高分辨率图像所包含的信息,本文将分辨率为4056像素×3040像素的图像连续分割为多个尺寸为640像素×640像素的子图像,每个子图像拥有10%的重叠区域。同时为增强检测模型的学习能力,对原始图像分别进行亮度增强、噪声模糊、翻转和旋转。通过上述数据增强方法,共得到5000幅样本数据。构建棉花数据集的主要任务是从图像中标记棉花,本研究利用Labeling软件标记图像中棉花的位置。参考数据集划分基本原则^[10],最终将标记好的棉花数据集按照比例7:2:1划分为训练集、验证集和测试集。

1.2 密植棉铃识别模型

1.2.1 IML-YOLO检测模型

由于棉花种植密度较高的特殊性,在复杂的棉田中存在许多半遮挡棉花以及小目标棉花^[11],普通的检测模型很难精准识别。同时由于新疆气候多变性大,YOLO系列目标检测模型对环境变化鲁棒性较差,泛化性能不佳,不同环境下性能落差明显。同时YOLO检测模型中非极大值抑制造成推理速度降低并且额外超参数的引入导致速度与准确度的不稳

定,此外非极大值抑制阈值的选择在不同场景下也需慎重考虑^[12]。为了应对这些挑战,本研究通过网络结构优化、改进特征融合等方式提出了 IML -

YOLO 检测模型,该模型作为一种轻量级检测模型,专门用于实时且精确检测棉田中成熟棉铃数量。图 1 展示了 IML - YOLO 的网络结构。

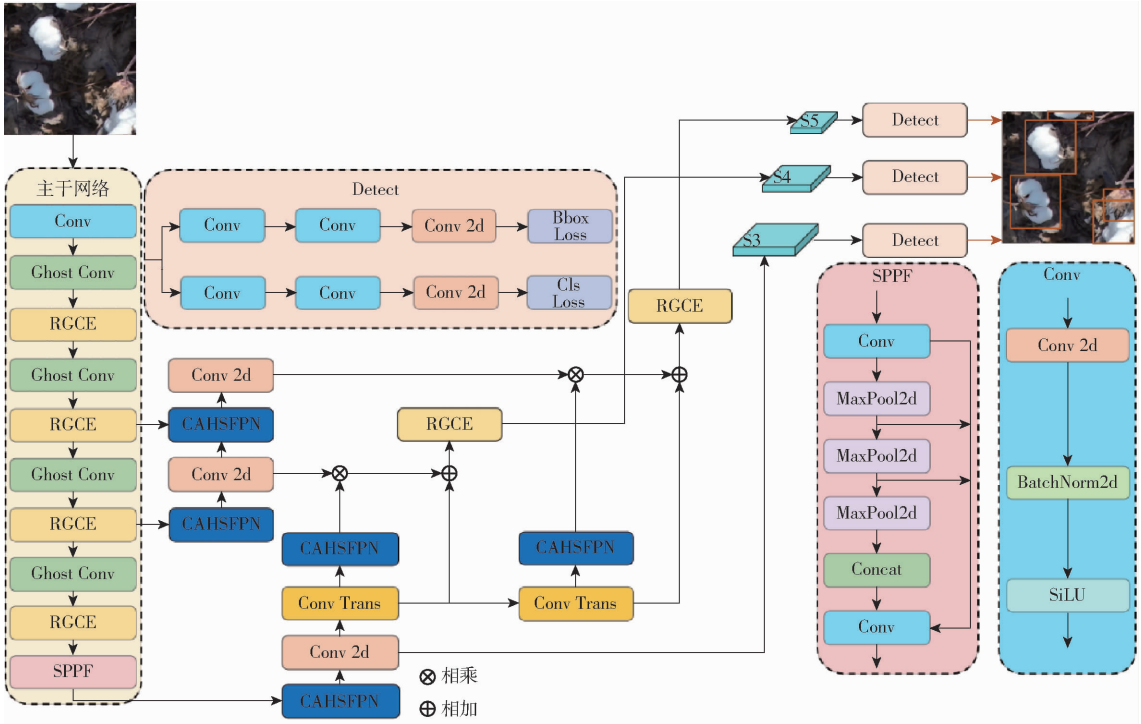


图 1 IML - YOLO 结构

Fig. 1 IML - YOLO structure

1. 2. 2 轻量化特征提取网络模块:GRGCE

在目标检测任务中,轻量化模型能够满足资源受限设备(如边缘计算和移动设备)的部署需求,但轻量化设计通常会带来特征提取能力的损失,导致检测精度下降。为了在降低计算量和参数量的同时维持高检测精度,本研究提出了一种全新的轻量化特征提取模块——GRGCE (Ghost Conv and RepGhostCSPELAN),该模块由幻影卷积 (Ghost Conv) 和 RepGhostCSPELAN (RGCE) 组成,专为实现高密度目标检测场景优化。常规卷积通常是直接使用卷积生成特征图,计算量与特征图的数量呈强相关线性增长关系。GhostNet^[13] 作为一种轻量级的卷积神经网络架构,与常规卷积相比,GhostNet 的优势在于通过一系列廉价运算能够显著地减少模型运算量。GhostNet 结构如图 2 所示。

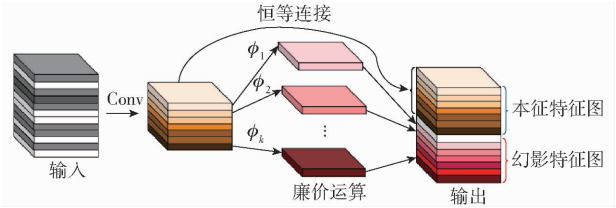


图 2 GhostNet 结构

Fig. 2 GhostNet structure

本文使用每秒浮点运算次数衡量模型的复杂

度,常规卷积运算复杂度和 Ghost Conv 运算复杂度计算公式为

$$F_N = nw'h'ck^2 \tag{1}$$

$$F_c = \frac{n}{s}w'h'[ck^2 + (s-1)d^2] \tag{2}$$

常规卷积与幻影卷积计算量和参数量之比计算公式为

$$r_s = \frac{ck^2}{\frac{1}{s}[ck^2 + (s-1)d^2]} \approx \frac{sc}{s+c-1} \approx s \tag{3}$$

$$r_c = \frac{nck^2}{\frac{n}{s}[ck^2 + (s-1)d^2]} \approx \frac{sc}{s+c-1} \approx s \tag{4}$$

式中 c ——输入特征图通道数
 n ——输出特征图通道数
 w' ——宽度 h' ——高度
 k ——常规卷积核尺寸
 d ——线性变换卷积核尺寸
 s ——产生特征图个数

通过对比,可以看出 GhostNet 可以显著地减少模型的计算量和参数量。为了降低计算量和参数量,同时提升特征提取能力,RGCE 采用廉价操作生成部分特征图,同时利用 CSP (Cross stage partial network) 的跨阶段部分连接策略,将特征图分割并

跨阶段传递, 结合 ELAN (Extended efficient layer aggregation networks) 的多尺度特征融合, 将不同阶段的特征图进行多尺度融合, 提升特征表达能力, 进一步提升网络的特征提取能力。同时 RGCE 摒弃了 YOLO v5 与 YOLO v8 中常用的 BottleNeck 结构。为弥补舍弃残差块所带来的性能损失, 在梯度流通分支上引入了 RepConv (Re-parameterizable convolution), 将训练阶段的复杂卷积结构转换为推理阶段的简单卷积结构, 显著增强了特征提取和梯度流通的能力。此外, RGCE 可以通过缩放因子调节尺寸, 从而兼顾小模型和大模型的需求。模块通过可调节的通道缩放因子适配不同规模模型的需求, 在边缘设备和高性能服务器等场景中均能灵活应用。RGCE 的结构如图 3 所示。RGCE 是一个高效且灵活的卷积神经网络模块, 结合了 RepConv 和多层卷积特征处理技术, 能够在推理阶段显著提高速度, 同时在训练阶段保持较高的特征提取能力。

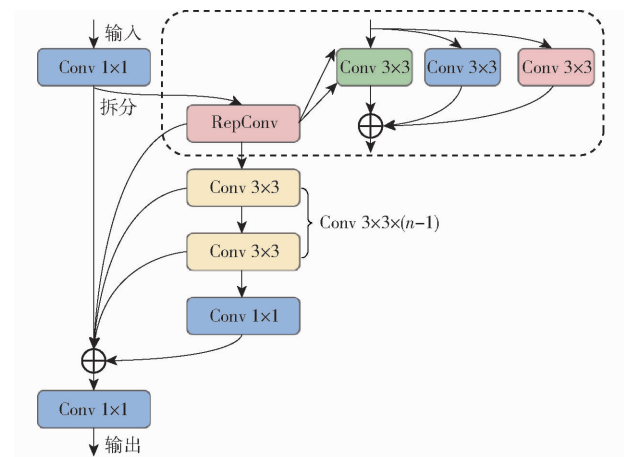


图 3 RGCE 结构
Fig. 3 RGCE structure

1.2.3 CAHSFPN 特征提取模块

坐标高级筛选特征融合金字塔网络 (Coordinate attention and high-level screening feature fusion pyramid networks, CAHSFPN) 主要用于解决多尺度问题。它的基本原理包括两个关键部分: 特征选择模块和特征融合模块。CAHSFPN 通过这两个模块协同工作, 显著地提高了检测的准确性和鲁棒性。CAHSFPN 特征融合模块结构如图 4 所示。

(1) 特征选择模块: CA (Coordinate attention) [14] 注意力机制能够将横向和纵向的位置信息编码到通道注意力中, 在不增加计算量的基础上使得网络能够关注大范围的位置信息。该模块利用 CA 注意力机制和维度匹配机制对不同尺度的特征图进行筛选。通过池化操作和权重计算, 该模块有效地提取了每个通道中的重要信息。

(2) 特征融合模块: 该模块通过选择性特征融

合 (Selective feature fusion, SFF) [15] 机制将筛选后的低级特征和高级特征结合起来。高级特征扩展后, 通过双线性插值或转置卷积进行尺度调整, 然后与低级特征融合, 从而增强模型对图像特征的表达能力。

SFF 是 CAHSFPN 中的一个关键组件, 它的主要作用是融合不同尺度的特征图。SFF 通过使用高级特征作为权重来过滤低尺度特征中的重要信息。然后, 利用高级特征作为注意力权重, 筛选出低尺度特征中有用的信息。这种融合方法能够有效地结合高级特征的语义信息和低尺度特征的细节信息, 从而提高模型在处理多尺度问题时的性能。

1.2.4 Focaler - MPDIOU 损失函数

目标检测的定位精度很大程度上依赖于基于 Intersection Over Union (IoU) 提出的边界框回归损失函数 (Boundary box regression, BBR), 然而, 大多数现有的边界框回归损失函数在预测框与实际标注框具有相同的宽高比, 但宽度和高度完全不同的情况下, 无法进行有效的优化。

Focaler - IoU [16] 通过对 IoU 进行线性区间映射, 使其更关注于特定范围内的值, 从而更好地处理正负样本不平衡问题。然而, Focaler - IoU 只关注边界框的重叠部分, 没有考虑边界框之间的空间位置关系, 导致目标定位不够准确。同时, 对于小目标检测, Focaler - IoU 无法提供足够的信息来区分细微的重叠差异。MPDIOU (Modified pan-diagonal intersection over union) [17] 结合了边界框中心点和边缘距离, 能够更好地反映边界框之间的实际重叠程度和定位误差。通过考虑最小垂直距离, MPDIOU 在边界框的精确定位上表现更好, 特别是对于形状不规则或相互靠近的目标。MPDIOU 示意图如图 5 所示。

因此, 本文提出了一种 Focaler - MPDIOU (FMPDI) 的边界框损失函数。它有效地考虑了边界框的重叠程度和空间位置关系, 提供了更全面的度量。通过结合 MPDIOU 的最小垂直距离信息, 可以在小目标棉花检测中获得更好的表现。同时, 借助 Focaler - IoU 的线性区间映射, 有助于缓解正负样本不平衡的问题, 提高了在大量背景样本情况下处理样本不平衡的能力。FMPDI 计算公式为

$$d_1^2 = (x_1^{gt} - x_1^{prd})^2 + (y_1^{gt} - y_1^{prd})^2 \quad (5)$$

$$d_2^2 = (x_2^{gt} - x_2^{prd})^2 + (y_2^{gt} - y_2^{prd})^2 \quad (6)$$

$$I_F = \begin{cases} 0 & (IoU < d) \\ \frac{IoU - d}{u - d} & (d \leq IoU \leq u) \\ 1 & (IoU > u) \end{cases} \quad (7)$$

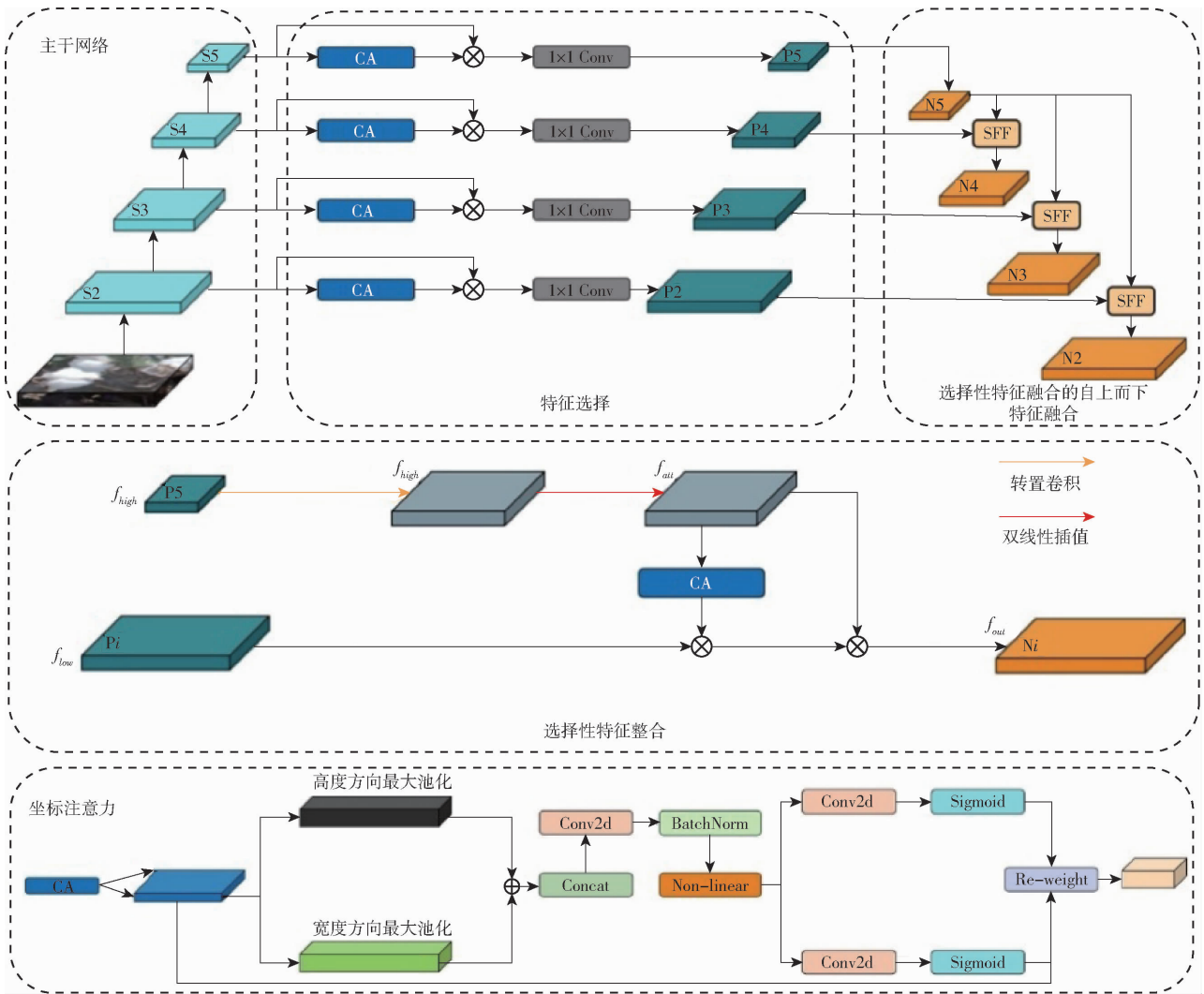


图 4 CAHSFPN 特征融合模块

Fig. 4 CAHSFPN feature fusion block

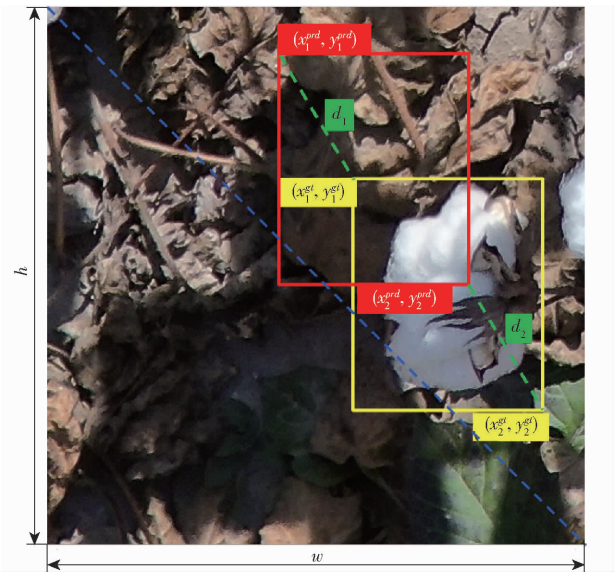


图 5 MPDIoU 示意图

Fig. 5 Schematic of MPDIoU

$$L_F = 1 - I_F + \frac{d_1^2 + d_2^2}{w^2 + h^2}$$

(8)

其中, (x_1^{gt}, y_1^{gt}) 为真实框左上角坐标; (x_2^{gt}, y_2^{gt}) 为真实框右下角坐标; (x_1^{prd}, y_1^{prd}) 为预测框左上角坐标; (x_2^{prd}, y_2^{prd}) 为预测框右下角坐标; $[d, u] \in [0, 1]$, 通过调整 d 和 u 的值使 FMPDI 关注不同的回归样本; w 和 h 分别为图像的宽度和高度; d_1 为真实框和预测框左上角顶点之间距离; d_2 为真实框和预测框右下角顶点之间距离; I_F 表示 Focaler - IoU; L_F 表示名为 Focaler - MPDIoU 的损失函数; IoU 表示交并比。

1.3 识别模型评价指标与试验环境

1.3.1 评价指标

为了进一步分析模型的优劣性,本研究使用了 3 个关键指标:精确度、召回率和平均精确度,这些指标用于定量分析验证本研究中模型识别棉花的性能。识别速度单位为 f/s。

1.3.2 试验环境

本研究模型训练和验证均在同一台设备上完成,设备采用 Ubuntu 20.04 操作系统,中央处理器为 AMD EPYC 7T83 64 - Core,显卡型号为 NVIDIA

GeForce RTX 4090 (24GB);编程语言版本为 Python 3.9.16;模型训练架构为 Torch 2.2.0,加速平台为 CUDA 11.7,CUDNN 11.7.64。模型训练初始学习率设置为 10^{-4} ,最小学习率设置为 10^{-6} ,学习率下降方式为余弦下降。训练轮次设置为 100。

1.4 棉花产量预测模型

本文将每个样方通过 IML-YOLO 棉铃检测模型所识别的棉铃数作为预测值,人工采集得到的棉铃数作为实际值。通过比较每个样方中棉铃实际数量和预测值之间的相对误差评价本文棉铃检测模型的可靠性。将人工采集棉铃在电子天平 (BSA224S 型,量程 220 g,精度 0.1 mg) 上进行单铃称量,每个棉铃重复 5 次,取 5 次平均值作为该棉铃的单铃质量。随机取出 30 朵棉铃求得平均单铃质量作为该样方的单铃质量。将单个样方中所有棉铃质量进行求和作为该样方的实际产量。

表 1 不同主干网络对模型的影响
Tab.1 Impact of different backbone networks on model

模型	精确度/%	召回率/%	平均精确度/%	浮点运算次数	内存占用量/MB	参数量
YOLO v8n	83.2	80.4	86.7	8.10×10^9	5.96	3.2×10^6
GhostNet-YOLO	82.2	78.0	84.8	5.00×10^9	3.58	1.7×10^6
MobileNetv4-YOLO	86.3	81.3	90.7	2.25×10^{10}	11.10	5.7×10^6
GRGCE-YOLO	91.7	88.8	93.1	5.50×10^9	3.97	1.9×10^6

与原始 YOLO v8n 模型相比,GRGCE-YOLO 模型的浮点运算次数、参数量和模型内存占用量分别减少了约 32.1%、36.7% 和 33.4%,平均精确度提升了 6.4 个百分点。在一阶段和二阶段目标检测算法上,GhostNet 能降低大部分计算量已被其他研究人员证实。与 GhostNet-YOLO 模型相比,GRGCE-YOLO 模型的平均精确度提升了 8.3 个百分点,但是模型浮点运算次数、参数量和内存占用量略有增加。MobileNetv4^[18] 作为在移动设备中被广泛使用的轻量型检测主干网络,它的检测速度和检测精度都被高度认可。与之相比,本文提出 GRGCE 主干特征提取网络,不但在检测精度方面显著地超越了 MobileNetv4,提升了 2.4 个百分点,而且模型的浮点运算次数、内存占用量和参数量还分别减少了约 75.6%、66.1% 和 66.7%,由该结果可见 GRGCE 对比其他轻量化模型具有显著的优势,主要得益于 RGCE 通过采用廉价操作生成部分特征图有效地减少了模型中的参数量,从而减少模型内存占用量。同时为弥补舍弃模型轻量化所带来的性能损失,在模型中引入跨阶段传递、多尺度融合和 RepConv 等策略用于保证模型的检测准确度。

2 结果与分析

2.1 改进模块性能比较

2.1.1 不同主干网络的比较试验

通过对比分析,本研究发现现阶段的目标检测特征提取主干网络主要存在以下两类问题:部分模型虽然检测精度很高,但是模型参数量较大,导致模型整体尺寸较大,同时严重降低了实时检测速度;模型通过轻量化降低了模型尺寸,但是模型的检测精度显著下降。

因此,为了平衡模型的整体尺寸和检测精度,本研究将 Ghost Conv 与自研的 RGCE 结合作为 YOLO v8 的特征提取主干网络。通过试验对比分析,GRGCE 在显著地减小了模型整体尺寸的同时提升了模型的检测精度,与其他轻量化特征提取主干网络比较结果如表 1 所示。

2.1.2 不同特征融合模块的比较试验

在 10 m 飞行高度下,无人机视角中棉株底部的棉铃和顶部棉铃大小存在显著性的差异。在棉铃检测任务中,模型能够同时检测不同尺寸棉铃对象的图像区域十分具有挑战性。尺寸差异往往会导致检测精度降低。YOLO v8n 的主干网络中通过逐层卷积,提取多个层次的特征。其中,高层次网络感受野大,语义信息表征能力强,但特征图分辨率比较低,空间几何特征缺乏;低层次网络虽然感受野小,语义信息表达能力也较弱,但特征图分辨率大,空间几何细节更加充分。为了提升模型在多尺度棉铃检测任务中的性能,本研究将 YOLO v8n 中原始的 Neck 模块替换为 CAHSFPN 模块,CAHSFPN 通过特征选择模块和特征融合模块协同工作,将低级特征和高级特征相融合显著地增强了模型提取特征的能力。

如表 2 所示,将 6 种特征融合模块引入 YOLO v8n,其检测精度均得到提升,同时浮点运算次数、内存占用量和参数量也有所下降。具体而言,CAHSFPN 相比 AFPN、BiFPN、CGRFPN、ASSF 和 CGAF 特征融合模块,检测精度分别提升了 2.2、3.5、1.8、0.9、0.5 个百分点,同时参数量也得到减

少,模型浮点运算次数显著降低。通过试验分析,AFPN 通过逐步融合低层和高层特征,避免了传统方法中信息丢失或退化的问题,与 CAHSFPN 采用了相似的特征融合策略,但是 AFPN 每一层的特征更新仅依赖于相邻的上一层和下一层。这种逐步的特征融合方式会导致信息在传播过程中逐渐丢失,尤其是对于深度神经网络来说,传递到更高层的特征已经被严重弱化。BiFPN 通过多层次双向融合来实现特征融合,主要依赖于简单的加权和操作来合并不同分辨率的特征图。它没有内置机制用于筛选出最有用的特征,因此在面对复杂场景的特征时,无法有效过滤冗余信息。CGRFPN 主要依赖于上下文引导的特征重建机制,将全局上下文信息和局部细节信息进行融合,以实现目标边界的精细识别,全局的语义理解较弱。ASSF 强调通过注意力机制来动态选择重要特征,在多尺度融合的过程中有很强的灵活性和精细化处理,尤其适合小物体和复杂边界

的检测,但计算复杂度较高,实时性较差。CGAF 利用内容引导的注意力机制来选择和融合多尺度特征,增强目标区域的表示能力,同时抑制背景干扰,提高检测精度;但是局部信息关注过多,在极端复杂的场景中(如多物体遮挡、强背景干扰或大规模噪声),该方法过于依赖局部特征的注意力。CAHSFPN 引入了高层筛选机制和注意力机制,通过筛选高层次特征来选择对目标检测和分类有用的信息。这种方法能够过滤冗余的特征,减轻噪声干扰,从而提高特征融合的质量和效率。此外,本研究采用 Grad-CAM^[24]来显示模型在复杂现场场景中对棉铃的注意程度和定位精度。Grad-CAM 技术使用明亮的颜色突出显示与检测结果相关的区域,较亮的颜色表示与棉铃的相关性较强。这种可视化为模型的检测结果提供了试验证据,如图 6 所示。热力图结果表明,CAHSFPN 在复杂背景中更加聚焦于棉铃。

表 2 不同特征融合模块对模型的影响
Tab. 2 Impact of different feature fusion modules on model

模型	精确度/%	召回率/%	平均精确度/%	浮点运算次数	内存占用量/MB	参数量
YOLO v8n	83. 2	80. 4	86. 7	8.1×10^9	5. 96	3.2×10^6
AFPN ^[19] -YOLO	86. 9	85. 3	92. 0	8.4×10^9	5. 60	2.6×10^6
BiFPN ^[20] -YOLO	83. 1	81. 9	90. 7	7.1×10^9	4. 04	2.0×10^6
CGRFPN ^[21] -YOLO	86. 8	88. 2	92. 4	8.2×10^9	6. 83	3.4×10^6
ASSF ^[22] -YOLO	88. 7	85. 6	93. 3	8.5×10^9	6. 06	3.1×10^6
CGAF ^[23] -YOLO	89. 8	86. 2	93. 7	8.5×10^9	6. 60	3.2×10^6
CAHSFPN-YOLO	90. 3	89. 1	94. 2	6.9×10^9	3. 90	1.9×10^6

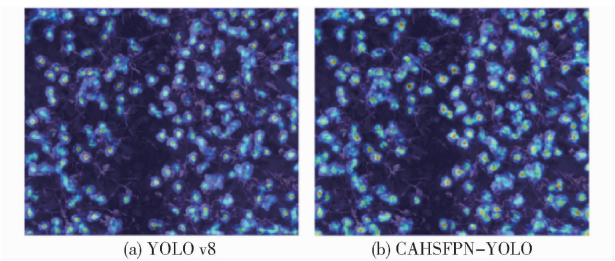


图 6 热力图比较结果

Fig. 6 Heat map comparison results

2. 1. 3 不同损失函数的比较试验

IoU 作为目标检测任务中一个关键指标,用于精确评估物体检测中预测边界框与真实边界框之间

的匹配程度。密植棉田中背景相对复杂且背景对象类型多,为了有效缓解复杂背景带来的准确度损失,同时提高模型中边界框的定位精度。本研究提出将 Focaler-IoU 和 MPDIoU 融合,试验结果表明在密植棉田的棉铃检测任务中 FMPDI 表现出了优异的检测性能。

根据表 3 可知,改进损失函数并没有对模型的浮点运算次数、模型内存占用量以及推理速度产生显著性的影响。但是在平均精确度方面,无论是 Focaler-IoU 损失函数还是 MPDIoU 损失函数,都提升了 YOLO v8n 检测模型的检测精度,但是提升幅

表 3 不同损失函数对模型的影响
Tab. 3 Impact of different loss functions on model

模型	精确度/%	召回率/%	平均精确度/%	浮点运算次数	内存占用量/MB	参数量
YOLO v8n	83. 2	80. 4	86. 7	8.1×10^9	5. 96	3.2×10^6
Focaler-IoU-YOLO	84. 7	83. 1	87. 9	8.1×10^9	5. 96	3.2×10^6
MPDIoU-YOLO	86. 0	84. 4	87. 6	8.1×10^9	5. 96	3.2×10^6
FMPDI-YOLO	90. 7	87. 3	89. 3	8.1×10^9	5. 96	3.2×10^6

度有限。FMPDI – YOLO 模型的平均精确度相较于 YOLO v8n 基准模型提升了 2.6 个百分点。该结果表明改进损失函数是一种有效提高检测精度的方法。对于需要高精度的密植棉田棉铃检测任务来说,改进是非常有价值的,因为它可以在不牺牲实时性能的前提下提高检测质量。图 7 展示了 3 种损失函数曲线,观察曲线可以看出,FMPDI 收敛速度最快,在迭代次数为 60 后波动趋于平稳,说明它能够在较少的迭代次数内达到一个较低的损失值,对于节省计算资源和加速模型训练是非常有利的。

2.2 IML – YOLO 模型消融试验

本研究基于 YOLO 模型进行了消融试验,以进一步研究所提出的增强技术在执行棉铃检测任务中

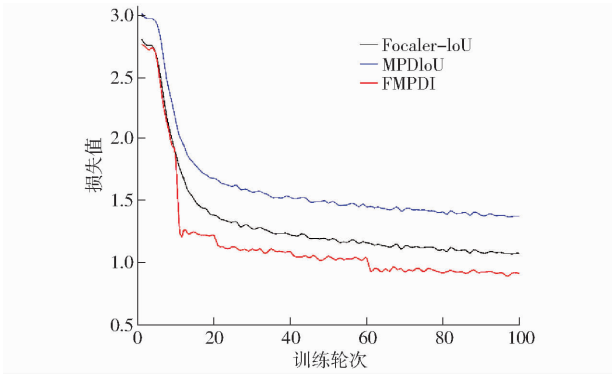


图 7 损失函数曲线

Fig.7 Loss function curves

表 4 消融试验结果

Tab.4 Results of ablation experiments

YOLO	GRGCE	HSFPN	FMPDI	精确度/%	召回率/%	平均精确度/%	浮点运算次数	内存占用量/MB	参数量
√	×	×	×	83.2	80.4	86.7	8.1×10^9	5.96	3.2×10^6
√	√	×	×	91.7	88.8	93.1	5.5×10^9	3.97	1.9×10^6
√	×	√	×	90.3	89.1	94.2	6.9×10^9	3.90	1.9×10^6
√	×	×	√	90.7	87.3	89.3	8.1×10^9	5.96	3.2×10^6
√	√	√	×	90.8	86.6	95.7	5.5×10^9	3.13	1.5×10^6
√	√	×	√	93.2	86.6	95.4	5.5×10^9	3.97	1.9×10^6
√	×	√	√	93.1	85.4	95.7	6.9×10^9	3.90	1.9×10^6
√	√	√	√	93.2	91.2	96.8	5.5×10^9	3.13	1.5×10^6

注:√表示使用该模块,×表示未使用该模块。

与基准 YOLO v8n 模型相比,本研究提出的 IML – YOLO 将模型的浮点运算次数、内存占用量和参数量分别降低了约 32.1%、47.5% 和 50%,平均精确度提升了 10.1 个百分点。该模型有效地平衡了模型检测精度和模型尺寸。此外,IML – YOLO 具有最佳的准确率和召回率,高准确率表明模型的预测更准确,高召回率表明模型能够以更高的准确性捕获阳性类别样本。

2.3 不同识别模型的对比试验

本文使用原始无人机图像进行模型性能测试,其中图像尺寸为 4 056 像素 × 3 040 像素。将 IML – YOLO 模型与经典目标检测模型进行比较,以证明其在棉铃实时检测任务中的优越性,比较结果如表 5 所示。该结果证明了 IML – YOLO 在检测棉铃的有效性。改进的 IML – YOLO 模型为 3.13 MB,具有 96.8% 的平均精确度,参数量仅有 1.5×10^6 。不

表 5 不同模型识别结果

Tab.5 Results of different detection models

模型	精确度/%	召回率/%	平均精确度/%	浮点运算次数	内存占用量/MB	参数量	推理速度/(f·s ⁻¹)
Faster R – CNN ^[25]	60.1	58.4	63.7	4.70×10^{11}	108	2.83×10^7	2.9
SSD ^[26]	64.2	62.8	66.1	3.36×10^{10}	45.2	1.43×10^7	14.8
YOLO v9 ^[27]	85.1	84.4	92.7	1.07×10^{10}	5.81	2.60×10^6	49.6
YOLO v10 ^[28]	85.5	82.0	90.6	8.20×10^9	5.50	2.70×10^6	50.7
RT – DETR ^[29]	88.1	80.5	91.0	5.69×10^{10}	38.5	1.99×10^7	27.1
EFA – YOLO ^[30]	89.2	83.9	92.2	4.60×10^9	3.30	1.40×10^6	52.4
Le – YOLO ^[31]	89.7	85.6	93.7	4.51×10^9	3.96	1.90×10^6	53.5
IML – YOLO	93.2	91.2	96.8	5.50×10^9	3.13	1.50×10^6	67.9

但模型最小,而且具有最佳的检测效果,性能优势显著。

表 5 展示了 8 种目标检测模型的综合性能对

比。IML – YOLO 在检测精度、计算效率与实时性方面均展现出显著优势。其精确度、召回率与平均精确度分别达到 93.2%、91.2% 和 96.8%,与次优模

型 Le - YOLO 相比分别提升 3.5、5.6、3.1 个百分点,表明其在减少误检与漏检上的卓越能力。在计算资源占用方面,IML - YOLO 的浮点运算量、参数量及模型内存占用量仅有传统模型 Faster R - CNN 的 1.2%、5.3% 和 2.9%,同时较轻量化改进模型 EFA - YOLO 和 Le - YOLO 进一步压缩存储空间,显示出更强的嵌入式部署适应性。实时性上,IML - YOLO 的推理速度高达 67.9 f/s,较 YOLO v10 提升 34%,较实时检测 Transformer 模型 RT - DETR 提升 1.5 倍,且远超两阶段检测器 Faster R - CNN 的 2.9 f/s,凸显其在动态场景下的高效处理能力。与同类改进模型相比,IML - YOLO 实现了精度与效率的协同优化:相较于 YOLO v9,其平均精确度提升 4.1 个百分点的同时模型内存占用量缩减 46.1%;面对 RT - DETR,浮点运算次数降低 90.3% 且推理速度翻倍;相比 EFA - YOLO,其精度指标提升 4 个百分点,推理速度提升 29.4%。尽管其浮点运算量略高于部分轻量化模型,但通过结构优化有效平衡了计算成本与性能增益,综合表现验证了其在复杂场景中高精度、低延迟与轻量化的协同优势,为棉铃实时检测任务提供了高效解决方案。

图 8 展示了 8 种模型在棉铃检测任务中的实际表现,其中蓝色检测框表示模型识别结果,红色圈部分表示漏检,绿色圈部分表示重叠漏检,橙色圈部分

表示冗余检测。从对比结果可以看出,各模型在识别能力和误检控制方面存在明显差异。Faster R - CNN 模型能够成功检测到部分顶部棉铃,但由于该模型对小目标和密集目标的识别能力有限,导致出现较多漏检。SSD 模型虽然检测速度相较 Faster R - CNN 更快,但在棉铃被部分遮挡时,其特征提取能力较弱,漏检率较高,尤其在复杂背景下容易忽略目标。YOLO v9 和 YOLO v10 模型在漏检问题上有所改善,能够较好地识别大部分棉铃目标。然而,当棉铃间距较近时,这两种模型容易将多个相邻棉铃误识别为一个整体,影响检测精度。RT - DETR 模型在整体识别效果上表现较好,能够较全面地检测棉铃目标,但在多个棉铃紧密聚集的情况时,容易产生大量冗余检测,导致不必要的重复识别。EFA - YOLO 和 Le - YOLO 模型在目标检测的准确性方面进一步优化,相较前述模型,其漏检率有所降低,同时提高了对遮挡棉铃的识别能力。然而,仍然存在部分区域的冗余检测和少量漏检。IML - YOLO 模型的检测性能最为优秀,其在棉铃被遮挡的情况下仍能准确识别目标,同时实现最低的漏检率。相比其他模型,IML - YOLO 有效减少了冗余检测,增强了对密集目标的区分能力,使得检测结果更加精准和稳定。整体来看,该模型在复杂环境下的棉铃检测任务中具有最优的鲁棒性和识别效果。

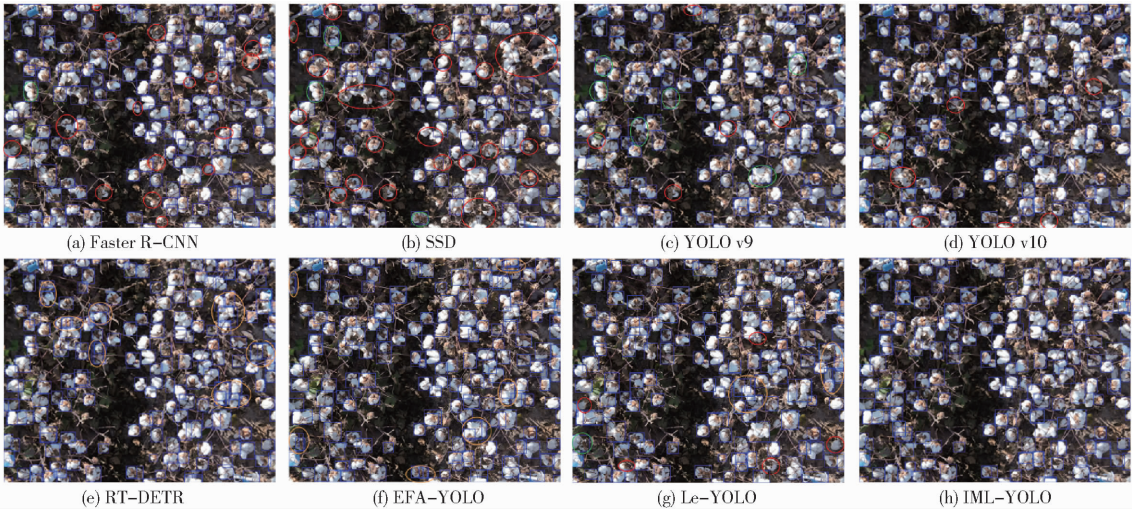


图 8 不同模型的检测结果
Fig. 8 Detection results of different models

2.4 棉花产量预测模型结果分析

为了验证本文所提出的 IML - YOLO 棉铃检测模型在实际应用中的有效性,本文设计并实施了基于 5 个样方棉铃的实际检测精度评估与样方产量预测试验。利用 IML - YOLO 模型对密植种植模式下的棉铃目标进行检测和计数,检测结果如图 9 所示。

从结果分析可以看出,模型整体表现优异,对不同样方的棉铃检测误差均较小,平均相对检测误差仅为 4.86%。其中,样方 1 的检测误差最高,为 7.69%。这是由于该样方的棉铃密度为所有样方中最高,背景干扰较为复杂,增加了检测的难度。而在样方 2 中,检测误差最低,仅为 2.68%,可能得益于该样方棉铃分布较为均匀且背景简单。综合分析表

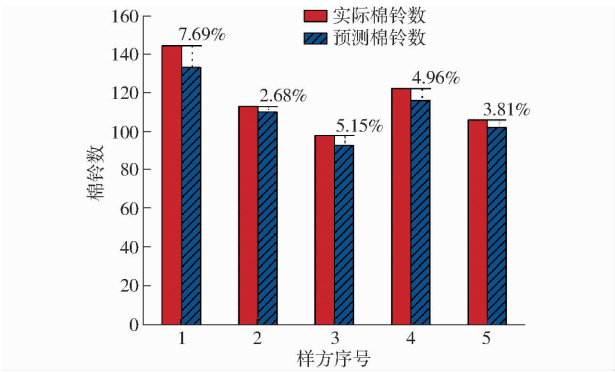


图9 试验样方棉铃检测结果

Fig. 9 Results of boll detection in experimental samples

明,IML – YOLO 模型能够在复杂场景下实现对棉铃目标的准确识别。

基于棉铃检测数量,进一步开展棉花产量预测相关研究,如表 6 所示,棉花产量相对误差趋势与棉铃数量相对误差呈现相同的趋势。同样,样方 1 的棉花产量预测相对误差最高,为 9.33%。样方 2 中相对误差最低,为 3.78%。5 个样方平均相对误差为 7.22%。该结果进一步说明了本文所提出 IML – YOLO 模型的有效性。

表 6 棉花实际产量与预测产量比较结果

Tab. 6 Results of comparison between actual and predicted cotton yields

样方序号	实际产量/g	预测产量/g	相对误差/%
1	902.57	818.4	9.33
2	702.34	654.0	3.78
3	625.77	570.4	8.85
4	771.23	713.0	7.55
5	670.45	626.2	6.60

3 结论

(1)IML – YOLO 模型的轻量级设计兼顾了棉铃的检测速度和精度。首先,设计了全新的 GRGCE 特征提取模块,显著减少了模型的参数。其次,引入了 CAHSFPN 特征融合模块,融合高层和低层语义信息提高了模型对多尺度棉铃的检测能力。最后,提出了 FMPDI 损失函数,有效地提升了棉铃检测精度。

(2)基于精准的棉铃检测结果,验证了目标检测方法在棉花产量预测应用中的稳定性与可靠性,为棉花产量预测提供了可落地的智能化决策支持。

参 考 文 献

[1] 张若薇,陈玉兰,赵达君. 新疆棉花种业发展现状及建议[J]. 合作经济与科技,2023(15):31 – 33.
ZHANG Ruowei, CHEN Yulan, ZHAO Dajun. Current situation and suggestions for the development of cotton seed industry in Xinjiang[J]. CO-Operative Economy & Science, 2023(15):31 – 33. (in Chinese)

[2] 胡灿,王兴旺,王旭峰,等. 基于 Android 手机的田间棉花产量预测系统设计[J]. 农业机械学报, 2023, 54(增刊 2): 252 – 259, 277.
HU Can, WANG Xingwang, WANG Xufeng, et al. Field cotton yield prediction system based on Android mobile phone[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(Supp. 2): 252 – 259, 277. (in Chinese)

[3] NIU H, PEDDAGUDREDDYGARI J R, BHANDARI M, et al. In-season cotton yield prediction with scale-aware convolutional neural network models and unmanned aerial vehicle RGB imagery[J]. Sensors, 2024, 24(8): 2432.

[4] SUN S, LI C, PATERSON A H, et al. Image processing algorithms for infield single cotton boll counting and yield prediction [J]. Computers and Electronics in Agriculture, 2019, 166: 104976.

[5] 刘金帅,赖惠成,贾振红. 基于 YCbCr 颜色空间和 Fisher 判别分析的棉花图像分割研究[J]. 作物学报, 2011, 37(7): 1274 – 1279.
LIU Jinshuai, LAI Huicheng, JIA Zhenhong. Image segmentation of cotton based on YCbCr color space and Fisher discrimination analysis[J]. Acta Agron. Sin., 2011, 37(7): 1274 – 1279. (in Chinese)

[6] MAJA J M J, CAMPBELL T, NETO J C, et al. Predicting cotton yield of small field plots in a cotton breeding program using UAV imagery data[C]// Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping. SPIE, 2016, 9866: 39 – 45.

[7] LI F, BAI J, ZHANG M, et al. Yield estimation of high-density cotton fields using low-altitude UAV imaging and deep learning [J]. Plant Methods, 2022, 18(1): 55.

[8] REDDY J, NIU H, SCOTT J L L, et al. Cotton yield prediction via UAV-based cotton boll image segmentation using YOLO model and segment anything model (SAM)[J]. Remote Sensing, 2024, 16(23): 4346.

[9] ZHANG M, CHEN W, GAO P, et al. YOLO SSPD: a small target cotton boll detection model during the boll-spitting period based on space-to-depth convolution[J]. Frontiers in Plant Science, 2024, 15: 1409194.

[10] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. Medical Image Analysis, 2017, 42: 60 – 88.

[11] LU Z, HAN B, DONG L, et al. COTTON – YOLO: enhancing cotton boll detection and counting in complex environmental conditions using an advanced YOLO model[J]. Applied Sciences, 2024, 14(15): 6650.

[12] LAVANYA G, PANDE S D. Enhancing real-time object detection with YOLO algorithm[J]. EAI Endorsed Transactions on

Internet of Things, 2023, 10: 12.

- [13] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580 – 1589.
- [14] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713 – 13722.
- [15] CHEN Y, ZHANG C, CHEN B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. Computers in Biology and Medicine, 2024, 170: 107917.
- [16] ZHANG H, ZHANG S. Focaler – IoU: more focused intersection over union loss[J]. arXiv preprint arXiv:2401.10525, 2024.
- [17] MA S, XU Y. Mpdious: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.
- [18] QIN D, LEICHNER C, DELAKIS M, et al. MobileNetV4: universal models for the mobile ecosystem[C] // European Conference on Computer Vision. Springer, Cham, 2025: 78 – 96.
- [19] YANG G, LEI J, ZHU Z, et al. AFPN: asymptotic feature pyramid network for object detection[C] // 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2023: 2184 – 2189.
- [20] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781 – 10790.
- [21] NI Z, CHEN X, ZHAI Y, et al. Context-guided spatial feature reconstruction for efficient semantic segmentation[C] // European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 239 – 255.
- [22] KANG M, TING C M, TING F F, et al. ASF – YOLO: a novel YOLO model with attentional scale sequence fusion for cell instance segmentation[J]. Image and Vision Computing, 2024, 147: 105057.
- [23] CHEN Z, HE Z, LU Z M. DEA – Net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002 – 1015.
- [24] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization[C] // Proceedings of the IEEE international Conference on Computer Vision, 2017: 618 – 626.
- [25] 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.
- [26] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: single shot multibox detector[C] // Computer Vision-ECCV 2016: 14th European Conference. Amsterdam: Springer International Publishing, 2016: 21 – 37.
- [27] WANG C Y, YE H I, MARK LIAO H Y. YOLO v9: learning what you want to learn using programmable gradient information[C] // European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 1 – 21.
- [28] WANG A, CHEN H, LIU L, et al. YOLOv10: real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2025, 37: 107984 – 108011.
- [29] ZHAO Y, LV W, XU S, et al. Detrs beat YOLOs on real-time object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965 – 16974.
- [30] PAN W, WANG X, HUAN W. EFA – YOLO: an efficient feature attention model for fire and flame detection[J]. arXiv preprint arXiv:2409.12635, 2024.
- [31] HOLLARD L, MOHIMONT L, GAVEAU N, et al. LeYOLO, new scalable and efficient CNN architecture for object detection[J]. arXiv preprint arXiv:2406.14239, 2024.