

doi:10.6041/j.issn.1000-1298.2026.20.029

基于改进 YOLO 11 的梨园复杂环境下梨果检测方法

周建军¹ 樊林¹ 刘泉乐¹ 王兴民¹ 王彬¹ 邱权²
(1. 北京石油化工学院信息工程学院, 北京 102617; 2. 北京农学院智能科学与工程学院, 北京 102206)

摘要: 针对梨园环境下梨果检测易受光照变化、目标尺度差异、分布密集以及严重遮挡等复杂因素影响的问题, 本文提出一种基于改进 YOLO 11 的梨果检测方法。该方法在 YOLO 11 基础网络中引入自监督形状感知特征增强 (Self-supervised shape-aware feature enhancement, SS-SAFE) 模块, 通过自监督学习预测目标长宽比, 从而增强对梨果独特形状特征的感知与提取能力; 同时, 采用自适应核卷积 (Adaptive kernel convolution, AKConv) 替换原始 C3K2 (Cross stage partial network with kernel size 2) 模块中的标准卷积, 构建 C3K2-AKConv (Cross stage partial network with kernel size 2 and adaptive kernel convolution) 深层特征提取结构, 提升网络对不同尺度梨果的适应性; 此外, 将 C2PSA (Convolutional block with parallel spatial attention) 模块替换为 C2PSA-HSAN (Cross-stage partial with hybrid spectral attention network) 模块, 通过优化特征交互与融合机制, 可显著增强模型在复杂环境下的特征融合能力。试验结果表明, 改进模型在精确率、召回率、F1 值、mAP@0.5 和 mAP@0.5:0.95 分别达到 88.9%、81.7%、85.2%、88.6% 和 67.2%, 相较于 YOLO 11n 模型分别提升 3.8、1.9、2.7、3.2、2.9 个百分点。研究结果可为梨园智能化管理与自动化采摘提供技术支撑。

关键词: 梨果检测; YOLO 11; 自监督学习; 自适应核卷积; 特征融合

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2026)20-0328-11 OSID: 

Pear Detection Method in Complex Orchard Environments Based on Improved YOLO 11

Zhou Jianjun¹ Fan Lin¹ Liu Quanle¹ Wang Xingmin¹ Wang Bin¹ Qiu Quan²
(1. College of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617, China
2. College of Intelligent Science and Engineering, Beijing University of Agriculture, Beijing 102206, China)

Abstract: Aiming to address the challenges of detecting pear fruits in orchard environments, such as variable illumination, multi-scale targets, dense distributions, and severe occlusions, a pear detection method was proposed based on an improved YOLO 11 framework. The proposed approach integrated a self-supervised shape-aware feature enhancement (SS-SAFE) module into the base YOLO 11 network, which learned to predict the aspect ratio of targets through self-supervised learning, thereby strengthening the perception and extraction of distinctive pear-shape features. Additionally, an adaptive kernel convolution (AKConv) was used to replace the standard convolution in the original cross stage partial network with kernel size 2 (C3K2) module, forming a cross stage partial network with kernel size 2 and adaptive kernel convolution (C3K2-AKConv) deep feature extraction structure, improving the network's adaptability to pears of varying scales. Furthermore, the C2PSA (convolutional block with parallel spatial attention) module was replaced by the C2PSA-HSAN (cross-stage partial with hybrid spectral attention network) module, which optimizes the feature interaction and fusion mechanism, significantly enhancing the model's feature fusion ability in complex environments. Experimental results show that the improved model achieves a precision of 88.9%, recall of 81.7%, F1-score of 85.2%, mAP@0.5 of 88.6%, and mAP@0.5:0.95 of 67.2%, outperforming the baseline YOLO 11n by 3.8, 1.9, 2.7, 3.2, 2.9 percentage points, respectively. This study offers a technical reference for the intelligent management and automated harvesting in pear orchards.

Key words: pear detection; YOLO 11; self-supervised learning; adaptive kernel convolution; feature fusion

0 引言

梨果是我国重要的经济林果之一,其采摘作业长期依赖人工,存在劳动强度大、效率低、成本高等问题^[1-2]。随着智慧农业的发展,自动化采摘已逐渐成为一种趋势,基于机器视觉的果实检测是自动化采摘关键环节^[3]。同时,梨果检测易受光照变化、枝叶遮挡、果实密集重叠、梨果自身纹理相似以及现实中梨果尺度差异较大等复杂因素影响,导致目标边界模糊、背景干扰较强,给梨果检测带来挑战。为实现梨园智能化管理和自动化采摘,亟需一种较高精度的梨果检测方法。

早期梨果识别研究多依赖颜色、纹理等基础特征并结合传统分类器进行识别。然而这类方法对图像背景噪声较敏感、特征工程繁琐,并且其泛化能力与鲁棒性存在一定不足,难以满足实用场景对准确性与实时性的要求^[4-6]。近年来,基于深度学习的目标检测算法(如 Faster R-CNN、YOLO 与 DETR 系列)凭借端到端特征学习与多尺度建模能力显著提升了检测效果^[7-10],并已逐步应用于多类果实的检测与定位,在密集、遮挡与小目标等复杂农业场景中展现出较好性能^[11-13]。此外,研究者围绕注意力机制、结构裁剪、可分离卷积与多模态融合等方向的持续改进,也为农业目标检测应用提供了有力技术支撑。

尽管相关研究已在果实检测精度与特征提取速度上取得一定进展,但面向梨果这一特定目标所处的实际场景的研究仍存在不足。首先,现有模型对梨果特有的几何先验(如近椭圆轮廓)利用不足,难以充分利用梨果的特定纵横比约束来提升模型在复杂背景下的判别能力。同时,由于果实生长分布的差异以及采集视角的变化,梨果表现出较大尺度波动,而现有模型对非均匀尺度分布的感知能力有限,导致其泛化与鲁棒性仍显不足^[14-16];其次,高精度模型往往伴随较高的计算与存储开销,难以在资源受限的边缘采摘设备上高效部署,精度与效率之间仍需要权衡^[17-18]。在增强特征表达与上下文建模方面,形状特征感知增强、动态自适应核卷积与增加注意力机制的多尺度特征融合机制已表现出良好潜力。然而如何将梨果特定的形状先验知识与尺度变化特征深度融合,仍有待深入研究^[19-21]。

本文提出一种基于改进 YOLO 11 的梨果检测方法,提升复杂梨园环境下梨果检测性能。设计自监督形状感知特征增强模块,利用梨果形状先验知识提升特征表达能力;引入自适应核卷积,以增强网络对不同尺度目标的适应性检测精度;构建优化

的多尺度特征融合机制,以提高复杂场景下的检测性能。在自建梨果数据集上对改进模型进行消融试验,将改进模型与主流 YOLO 系列模型性能进行测试分析,以验证改进模型的有效性。

1 材料与方法

1.1 数据集构建

1.1.1 数据采集

2024 年 10 月梨果成熟期在北京市大兴区赵村梨园进行数据采集。品种包括酥梨、京丰梨、鸭梨、红肖梨、子母梨等。采集数据覆盖近距离、中距离、远距离等多种拍摄距离,以及正视、侧视和俯视等多种拍摄角度,并分别在晴天、阴天等典型天气下完成采集。

为了确保数据的全面性和代表性,分别在晴天和阴天对稀疏分布和密集分布 2 种典型场景进行了图像采集。共采集原始图像 1 104 幅,图像分辨率为 1 920 像素×1 280 像素。其中包括晴天场景下梨果稀疏分布图像 265 幅,晴天场景下梨果密集分布 277 幅,阴天场景下梨果稀疏分布图像 273 幅,阴天场景下梨果密集分布图像 289 幅。图 1 为不同天气场景下梨果图像。晴天光照充足,梨果表面易出现光斑;阴天光照较弱,梨果与背景对比度较低。梨果稀疏分布场景中,梨果数量较少,相互遮挡较轻;而在密集分布场景中,梨果重叠较多、遮挡现象明显。



图 1 不同环境下梨果图像

Fig. 1 Pear fruit images under different environments

1.1.2 数据标注与增强

采用 LabelImg 工具对梨果进行标注,使用矩形

框标注梨果位置,并将标注类别统一设定为“pear”。标注过程中严格遵循标注规范,确保边界框紧贴目标轮廓,最大限度减少背景区域纳入,以提升标注一致性并有利于后续模型的训练效果。

为了丰富训练数据,对原始数据采用增强技术进行扩充。首先,按照比例 7:2:1 随机划分训练集、验证集和测试集,分别为 773、221、110 幅图像,再对训练集样本进行增强。采用多种差异化增强策略:随机旋转与缩放、水平翻转、对比度调整、遮挡模拟、随机光斑等。通过循环使用不同增强策略,有效提升了数据特征的多样性。

增强后部分图像如图 2 所示,经过数据增强,训练集从原始 773 幅图像扩充至 1 546 幅图像。

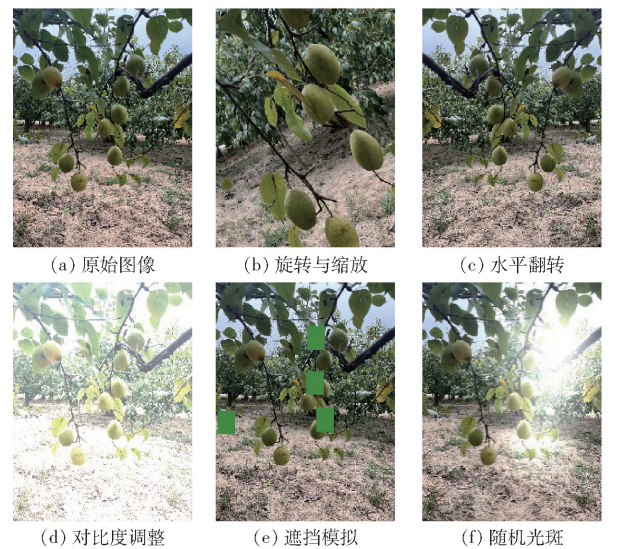


图 2 数据增强示例

Fig. 2 Examples of data augmentation

1.2 YOLO 11 目标检测网络结构

YOLO 系列算法作为单阶段目标检测代表,凭借其端到端的高效架构在实时检测任务中表现优异,与 R-CNN 系列等两阶段检测方法相比,YOLO 在网络结构上更为紧凑高效,能够在保持较高检测精度的同时保持优越的实时检测性能。尽管 YOLO 当前已有较新版本(如 YOLO 13、YOLO 26 等),但考虑到模型稳定性以及软硬件协同方面的成熟度,YOLO 11 更具优势。此外,最新版本虽可能带来理论上的性能提升,但其网络结构更加复杂可能会增加额外算力开销,因此极易引入过拟合风险与不必要的计算冗余。

YOLO 11 在网络结构上保持了紧凑高效的特点,主要由骨干特征提取网络(Backbone)、特征融合颈部网络(Neck)以及预测检测头(Head)3 部分组成。其提供了多个版本,其中 YOLO 11n 在模型参数量与检测能力之间取得了更好的平衡,在稳定性与软硬件兼容方面具有更优的综合表现。考虑到果

园环境中普遍存在的光照变化、枝叶遮挡、果实密集重叠等复杂干扰因素,选择 YOLO 11n 为基线模型。

1.3 改进 YOLO 11 模型

针对果园环境中存在的复杂干扰因素,以 YOLO 11 为基础框架,通过引入新的模块及调整原始结构中的模块,提升梨果的识别精度。改进 YOLO 11 模型结构如图 3 所示。设计并引入 SS-SAFE 模块,使模型能够学习并利用目标的形状先验知识,引导模型关注梨果的全局轮廓与几何特征,从而有效提升模型在复杂环境下的覆盖能力与检测精度。使用 AKConv 替换原始 C3K2 模块中的部分标准卷积,构建了 C3K2-AKConv 结构,并将其应用于主干网络中,增强网络自适应调整空间采样坐标的能力,提高网络对不同尺度梨果的适应性,实现更高效特征提取的同时增强模型的多尺度泛化能力。最后,使用 C2PSA-HSAN 替换原结构中的 C2PSA 模块,该机制通过优化跨阶段特征交互路径,在特征融合过程中,精准加强小目标及被遮挡果实的微弱特征响应,从而有效提升了远处小目标梨果以及遮挡梨果的特征融合能力。

1.3.1 自监督形状感知特征增强模块

在真实果园场景,由于枝叶遮挡,果实自身形态差异等因素,通用目标检测模型对梨果识别仍面临较高漏检率,尤其在对严重遮挡或非典型形状目标的检测中表现不佳。现有方法大多未充分挖掘类别特有的形状先验知识^[22],而通过标注每个实例的形状特征成本过高,制约了检测精度的进一步提升。

为克服这一局限,根据梨果具有的典型椭圆形状特征,设计了 SS-SAFE 模块,通过对目标的长宽比进行预测,基于预测的长宽比,采用以梨果典型形状为中心的高斯分布生成基础权重,并对高长宽比目标引入补偿机制^[23]。该模块与整个模型联合训练,不引入额外的损失函数,而是通过模型的分类和回归损失反向传播,在长宽比预测过程中学习到有利于检测的形状信息。在训练中,预测输出宽高影响注意力权重分布,从而增强模型特征感知能力,网络在最小化检测损失的过程中自发调整预测过程中的参数,使其关注与目标几何属性相关的区域,从而提升对梨果目标的检测性能^[24]。该机制可以在不引入人工标注的情况下,使模型能够自主感知并强化目标的形状特征,提升困难样本识别能力^[25-26]。SS-SAFE 模块整体架构如图 4 所示。

长宽比预测分支采用自监督学习策略,通过分析特征图的全局统计信息预测目标宽度和高度,无需额外形状标注。长宽比计算公式为

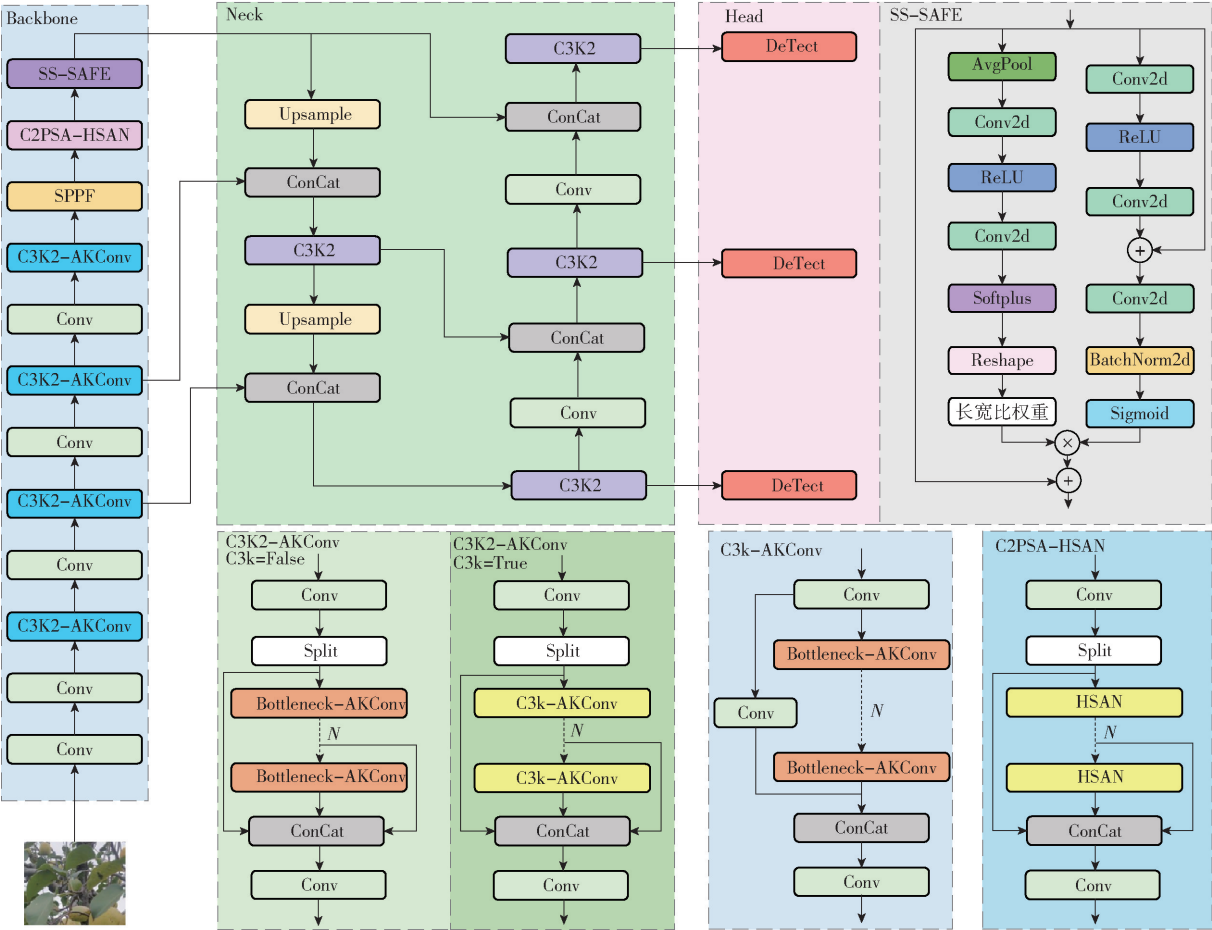


图 3 改进 YOLO 11 网络结构

Fig. 3 Improved YOLO 11 network structure

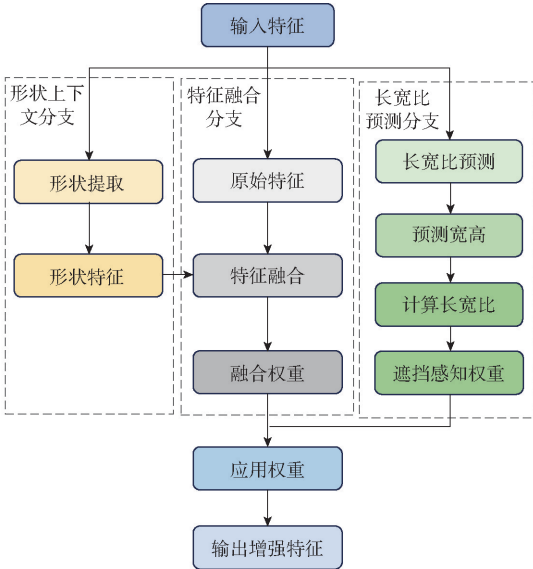


图 4 SS-SAFE 模块架构

Fig. 4 Architecture of SS-SAFE module

$$r = \frac{\max(b, h)}{\min(b, h) + \varepsilon} \tag{1}$$

式中 r ——长宽比
 b ——预测目标宽度
 h ——预测目标高度
 ε ——平滑常数,用于数值稳定性

为进一步将长宽比转换为权重信号,设计了一种基于高斯分布的基础权重生成策略,基础权重计算公式为

$$w_b = \exp \left(-\frac{1}{2} \left(\frac{r - \mu}{\sigma} \right)^2 \right) \tag{2}$$

式中 w_b ——基础权重 σ ——标准差
 μ ——典型梨果形状比中心值

根据典型梨果形状设定中心值 $\mu = 1.3$,标准差 $\sigma = 0.5$ 。为加强对疑似遮挡样本的感知能力,引入补偿权重机制,补偿权重计算公式为

$$w_c = \begin{cases} 1.0 + 0.3 \frac{r - \theta}{\theta} & (r > \theta) \\ 1.0 & (r \leq \theta) \end{cases} \tag{3}$$

式中 w_c ——补偿权重 θ ——长宽比阈值

长宽比阈值 $\theta = 1.5$,若 r 大于 θ ,则按式(3)计算 w_c ,否则取值为 1。遮挡感知权重为基础权重与补偿权重的乘积,计算公式为

$$w_d = \text{clip}(w_b w_c, 0.5, 2.0) \tag{4}$$

式中 w_d ——遮挡感知权重
 $\text{clip}()$ ——裁剪函数

为保证训练稳定性,将计算得到的遮挡感知权重限制在 0.5 ~ 2.0 之间。形状上下文分支对主干

网络传来的特征进行进一步处理,使用分组卷积高效地提取与目标轮廓和形状强相关的上下文信息。特征融合分支将原始特征与增强的形状上下文特征结合,生成空间自适应的注意力权重。该分支确保在检测过程中有效地利用形状先验,从而提升特征的表达能力,有效地减少被遮挡实例的漏检,增强模型的目标覆盖能力,并将原始特征与各项权重进行融合计算,特征融合计算公式为

$$f = x \odot f_w \odot w_d \tag{5}$$

式中 f ——融合后的特征
 x ——原始特征
 $\odot$ ——逐元素相乘
 f_w ——特征融合权重

其中, f_w 为形状上下文分支与特征融合分支生成的特征融合权重,将原始特征与特征融合权重进行运算后得到融合后的特征。

1.3.2 自适应核卷积

传统卷积受限于固定的局部窗口与规则采样网格,难以充分适应多尺度与异形目标的特征提取需求,多尺度目标要求模型能够同时捕捉不同大小的视觉结构,而异形目标则需卷积核具备灵活适应不规则轮廓与空间分布的能力。为克服这一局限,引入 AKConv,核心思想是通过动态调整卷积核的采样位置,增强模型对复杂视觉模式的建模能力。AKConv 允许卷积核具备任意数量的参数,并支持不规则采样形状,通过初始采样坐标生成、可学习偏移、特征重采样的机制,实现对采样位置的自适应调整,从而显著提升特征提取的灵活性与准确性^[27-28],AKConv 的执行流程如图 5 所示。初始采样坐标生成阶段,通过网格生成函数 Grid() 构建出初始采样坐标集从而得到初始采样偏移,形成覆盖目标区域的采样模板,初始采样坐标集计算公式为

$$p_n = \text{Grid}(K) = \{(x_t, y_t)\}_{t=1}^K \tag{6}$$

式中 p_n ——初始采样偏移
 K ——采样点数量
 x_t, y_t ——第 t 个采样点相对横坐标、相对纵坐标

其中, (x_t, y_t) 为第 t 个采样点在局部坐标系中的相对位置,该模板可随着任务进行改变为规则形式或不规则形式,以支持多样化感受野结构。

可学习偏移优化阶段,在网络训练过程,通过一个偏移预测模块学习每个采样点的空间偏移量 Δp 。对于输出特征图中位置 (i, j) 处的卷积核,最终采样位置计算公式为

$$\begin{cases} p_0 = (is, js) \\ p = p_0 + p_n + \Delta p \end{cases} \tag{7}$$

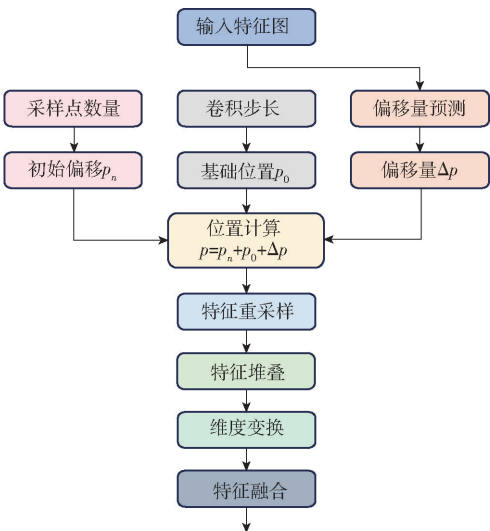


图 5 AKConv 执行流程
Fig. 5 Execution process of AKConv

式中 p_0 ——基础采样位置
 i ——特征图空间索引相对横坐标
 j ——特征图空间索引相对纵坐标
 s ——卷积步长
 Δp ——网络学习到的自适应偏移量
 p ——最终采样位置

其中, p_0 由输出位置与卷积步长共同决定, p_n 定义卷积核基本几何结构, Δp 实现对采样位置细粒度空间校正。

特征重采样与融合阶段,基于调整后的采样点坐标,采用双线性插值算法从输入特征图中提取对应位置的特征值。对每个目标位置 $p = (p_x, p_y)$,首先通过向下取整的计算方法确定其 4 个最近邻整数坐标点,具体计算公式为

$$\begin{cases} q_{ll} = (\lfloor p_x \rfloor, \lfloor p_y \rfloor) \\ q_{rt} = (\lfloor p_x \rfloor + 1, \lfloor p_y \rfloor) \\ q_{lb} = (\lfloor p_x \rfloor, \lfloor p_y \rfloor + 1) \\ q_{rb} = (\lfloor p_x \rfloor + 1, \lfloor p_y \rfloor + 1) \end{cases} \tag{8}$$

式中 p_x, p_y ——目标位置相对横、纵坐标
 q_{ll}, q_{rt} ——左上、右上坐标邻点
 q_{lb}, q_{rb} ——左下、右下坐标邻点
 $\lfloor \cdot \rfloor$ ——向下取整

再根据各整数邻点与目标位置的相对距离计算插值权重,计算公式为

$$\begin{cases} \Delta x = p_x - \lfloor p_x \rfloor \\ \Delta y = p_y - \lfloor p_y \rfloor \\ g_{ll} = (1 - \Delta x)(1 - \Delta y) \\ g_{rt} = \Delta x(1 - \Delta y) \\ g_{lb} = (1 - \Delta x)\Delta y \\ g_{rb} = \Delta x\Delta y \end{cases} \tag{9}$$

式中 Δx ——采样点与左上邻点水平距离
 Δy ——采样点与左上邻点垂直距离
 g_{lt} ——左上点权重
 g_{rt} ——右上点权重
 g_{lb} ——左下点权重
 g_{rb} ——右下点权重
最终的特征值通过加权求和得到,计算公式为

$$x_{\text{offset}} = \sum_{k \in \{lt, rt, lb, rb\}} g_k x_{bk} \tag{10}$$

式中 x_{offset} ——最终特征值
 g_k ——每个点双线性插值权重
 x_{bk} ——整数坐标位置原始特征值
所有采样点的重采样结果沿通道维度叠加,经维度变换通过点卷积实现跨通道特征融合和输出映射。该机制使 AKConv 能够根据输入内容动态调整感受野形状与分布,显著增强模型对不同尺度目标的特征提取能力。

1.3.3 优化的多尺度特征融合机制

为提高小目标梨果以及遮挡场景下特征融合能力,设计了 C2PSA - HSAN 模块以优化特征融合,该模块主要由混合谱注意力网络 (Hybrid spectral attention network, HSAN) 组成,结构如图 6 所示。

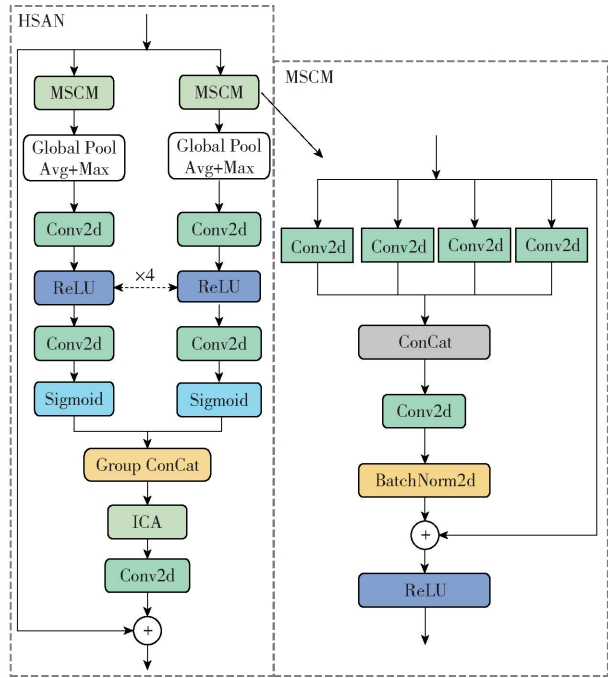


图 6 HSAN 与 MSCM 结构

Fig. 6 Structure of HSAN and MSCM

HSAN 主要通过结合多尺度卷积 (Multi-scale convolution module, MSCM) 模块与改进的坐标注意力 (Improved coordinate attention, ICA) 机制,以增强多尺度目标的特征融合能力^[29]。MSCM 以 1×1 卷积聚焦点特征和通道关系, 3×3 卷积捕捉局部纹理和边缘, 5×5 与 7×7 卷积则负责捕获更大范围的

轮廓和形状信息,该结构能够更有效地捕捉被遮挡果实以及较小果实的特征信息。通过该机制,模型能够以不同感受野处理各组特征,同步捕捉远距离小目标梨果与被遮挡梨果的微弱特征,以及近距离梨果的显著特征。

ICA 模块结构如图 7 所示,其分别沿着高度和宽度 2 个方向进行全局池化,减少因拼接和共享卷积造成的信息干扰,从而提取出 2 个相互独立的方向感知特征。在此基础上,模块进一步学习每个空间方向上的重要程度,并生成相应的注意力权重图。最终,2 个方向的权重图同时作用于特征图,帮助模型更准确地锁定目标在空间中的位置。在获取多尺度特征后,ICA 机制突出关键特征,同时抑制背景噪声和干扰,并融合各尺度结果。

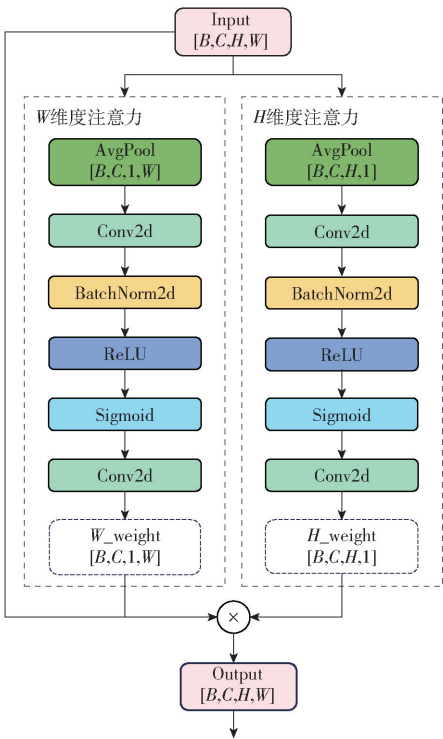


图 7 改进的坐标注意力结构

Fig. 7 Improved coordinate attention structure

用 C2PSA - HSAN 替换模型原有的 C2PSA 模块,替换后的网络能够更加准确高效地捕获小目标及遮挡目标的特征,能够更加有效地聚合多尺度上下文信息并建立长距离空间依赖关系,为后续检测头提供更有效的特征表示。

1.4 试验环境与参数配置

试验环境配置: Windows 11 (64 位); CPU 为 13th Gen Intel (R) Core (TM) i9 - 13900HX @ 2.20 GHz; GPU 为 NVIDIA GeForce RTX 4070 Laptop (显存为 8 GB); 运行内存为 32 GB; Python 3.8.20; PyTorch 2.4.1; 适配对应的 CUDA 12.1 和 cuDNN 9.0.1。

模型训练过程,所有图像尺寸均统一调整为 640 像素×640 像素,训练周期设置为 200 轮,批量大小设置为 16。为提升训练效率,启用数据缓存机制并设置 4 个工作进程进行数据加载。损失函数权重经过调优,边界框回归损失权重设为 7.5,分类损失权重设置为 0.9,分步焦点损失权重设置为 1.5。训练采用渐进式学习率策略,初始学习率为 0.000 8,最终学习比率为 0.01,并包含 5 个周期的学习率预热阶段,确保训练过程稳定收敛。试验设定了随机种子值,保证试验结果可复现。

1.5 评价指标

为评估改进后模型的检测性能,采用精确率(P)、召回率(R)、F1 值、平均精度均值(mAP)进行

评估。其中,mAP@ 0.5 使用交并比(IoU)阈值为 0.5 进行计算,对所有类别的 AP 值取平均。mAP@0.5;0.95 为 IoU 阈值从 0.5 到 0.95,步长 0.05,共 10 个,并对这 10 个值取均值。

2 试验结果与分析

2.1 消融试验

为验证各改进模块的有效性,以 YOLO 11n 为基线模型,设计消融试验,逐步引入改进模块,验证 SS-SAFE 通过形状知识先验是否有效降低漏检,以及由 AKConv 构建的 C3K2-AKConv 通过学习偏移能否提升检测精确度,并评估 C2PSA-HSAN 模块的融合效果,结果如表 1 所示。

表 1 消融试验结果
Tab.1 Ablation experiments results

模型	SS-SAFE	C3K2-AKConv	C2PSA-HSAN	精确率/ %	召回率/ %	F1 值/ %	mAP@0.5/ %	mAP@0.5;0.95/ %
YOLO 11n	×	×	×	85.1	79.8	82.4	85.4	64.3
SS-SAFE	✓	×	×	86.9	81.5	84.1	87.3	65.5
C3K2-AKConv	×	✓	×	87.1	79.9	83.3	86.3	66.3
C2PSA-HSAN	×	×	✓	87.3	80.1	83.5	87.1	66.1
SS-SAFE + C2PSA-HSAN	✓	×	✓	88.3	81.2	84.6	88.1	66.9
SS-SAFE + C3K2-AKConv	✓	✓	×	87.2	80.5	83.7	87.0	65.9
C3K2-AKConv + C2PSA-HSAN	×	✓	✓	87.9	80.8	84.2	87.2	66.8
改进模型	✓	✓	✓	88.9	81.7	85.1	88.6	67.2

注:✓表示使用该模块,×表示未使用该模块。

加入 SS-SAFE 模块后,召回率显著提升,较原模型提升 1.7 个百分点。表明添加了该模块后,改进模型有效地减少了被遮挡目标的漏检问题,增强了模型的覆盖能力,能够检测到更多的实例,这也验证了形状感知特征增强在梨果检测任务中的关键作用。

将原始 C3K2 模块替换为 C3K2-AKConv 模块后,改进模型在与基线模型召回率保持持平的基础上,精确率与平均精度均值(mAP@0.5;0.95)均提升约 2 个百分点。表明使用该模块替换后能够使改进模型更加准确地提取复杂场景中的梨果的特征信息,验证了 C3K2-AKConv 特征提取结构对多尺度目标以及异形目标检测的有效性。

使用 C2PSA-HSAN 模块显著的提升了精确率与平均精度均值,其中,精确率提升 2.2 个百分点,平均精度均值(mAP@0.5;0.95)提升 1.8 个百分点。验证了该模块能够更加有效地融合不同尺度的特征信息并建立长距离空间依赖关系,能够为检测头提供更加有效且具有鲁棒性的特征信息。

2.2 与主流算法对比

将改进模型与当前主流的目标检测算法进行对

比,选取当前具有代表性的 YOLO 系列轻量型版本(包括 YOLO v5n、YOLO v8n、YOLO v10n、YOLO 11n、YOLO 13n、YOLO 26n)以及 RT-DETR v2-s 进行试验,对比结果如表 2 所示。

表 2 不同模型性能
Tab.2 Performance of different models %

模型	精确率	召回率	F1 值	mAP@0.5	mAP@0.5;0.95
RT-DETR v2-s	80.3	77.1	78.7	81.3	62.9
YOLO v5n	84.4	79.5	81.9	84.3	61.7
YOLO v8n	83.8	79.9	81.8	84.6	63.4
YOLO v10n	84.8	79.1	81.9	83.9	62.3
YOLO 11n	85.1	79.8	82.4	85.4	64.3
YOLO 13n	83.2	79.0	81.1	83.8	62.4
YOLO 26n	86.9	78.6	82.5	85.0	64.0
改进模型	88.9	81.7	85.2	88.6	67.2

由表 2 可知,改进模型各指标均优于其他模型。具体来看,在精确率方面,改进模型达到 88.9%,较 YOLO 11n 提升 3.8 个百分点,有效降低了误检率。在召回率方面,改进模型为 81.7%,相较于 YOLO v8n 提升 1.8 个百分点,表明该方法在保持较高精

度的同时,具备更强的目标覆盖能力。在目标检测的核心指标 $mAP@0.5$ 和 $mAP@0.5:0.95$ 上,改进模型分别达到 88.6% 和 67.2%,相较 YOLO 11n 分别提升 3.2、2.9 个百分点。 $mAP@0.5:0.95$ 显著提升表明改进方法在不同 IoU 阈值下都能保持较高的检测精度。此外,改进模型的 F1 值(85.2%)相较于 YOLO 11n 提升 2.8 个百分点,表明改进模型在精确率与召回率之间实现了更加有效的平衡。

2.3 可视化分析

为更加直观地展示改进模型在各类环境下区分梨果与其他复杂背景方面的性能,对梨果图像使用 Grad CAM 热力图进行可视化,红色区域颜色越深,表示该区域对模型最终预测的贡献越大。如图 8 所示,从数据中筛选出晴天、阴天、稀疏与密集 4 类具有代表性的场景,分别使用原图、YOLO 11n、改进模型还原梨果检测效果。

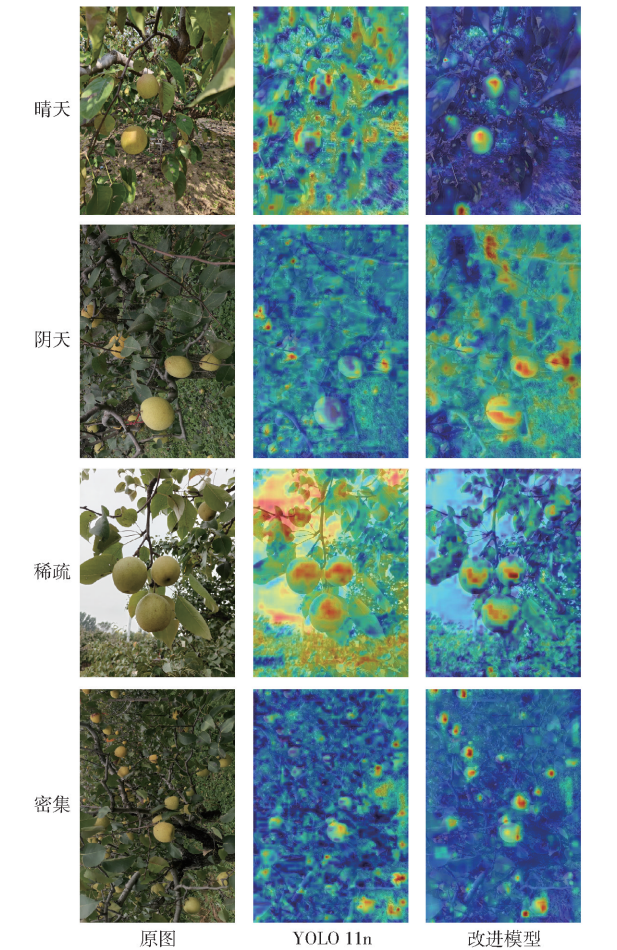


图 8 热力图可视化效果

Fig. 8 Effect of heatmap visualization

由图 8 可知,与 YOLO 11n 相比,改进模型能够更加精准聚焦与梨果目标的纹理细节,还能有效抑制梨叶、树干等复杂背景的干扰响应。在原始图像中,由于梨果与梨叶具有相似的颜色特征,YOLO 11n 的热力图显示其对梨叶部分存在误激活现象,

同时在真实的梨果目标区域未能产生有效激活,而改进模型则能够更有效的区分梨果目标与背景。其中,SS-SAFE 模块有效强化了梨果的形态特征表达,同时,C2PSA-HSAN 模块在通道维度上突出关键特征,在空间维度上抑制背景噪声和干扰,从而实现更精准的目标与背景之间的特征解耦。

2.4 不同场景下检测效果

在晴天、阴天、稀疏、密集场景下测试模型性能,结果如表 3 所示。

表 3 不同场景下检测效果

Tab. 3 Detection effects in different scenarios %

场景	精确率	召回率	F1 值	$mAP@0.5$	$mAP@0.5:0.95$
晴天	89.9	82.3	85.9	89.2	67.7
阴天	87.7	80.2	83.8	87.1	66.0
稀疏	91.2	86.6	88.8	93.8	78.0
密集	85.9	80.1	82.9	86.5	65.1

由表 3 可知,晴天、阴天 2 种场景下性能接近,精确率分别为 89.9% 与 87.7%,F1 值分别为 85.9% 与 83.8%,表明模型对天气变化具有良好鲁棒性与适应性;在目标稀疏分布下模型表现最佳, $mAP@0.5$ 为 93.8%;在目标密集场景中, $mAP@0.5:0.95$ 下降至 65.1%,但整体仍维持较高检测能力。

2.5 不同场景下不同模型检测效果

2.5.1 晴天场景下检测效果

图 9 为在晴天不同密集程度场景下,基线模型与改进模型检测效果,图中红色方框表示模型产生的漏检或误检。

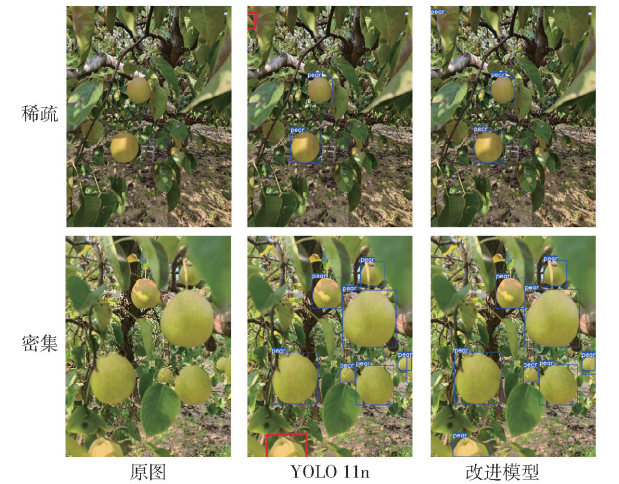


图 9 晴天场景下检测效果

Fig. 9 Detection performance in sunny scenarios

在晴天目标稀疏场景下,由于梨果表面易形成光斑、梨果与梨树叶颜色接近等问题,部分异形目标果实极易产生漏检,改进模型能够更加精准地捕捉到部分异形与目标,有效地降低了漏检实例目

标数量。在晴天目标密集场景下,由于光照较强,明暗对比过大,梨果数量较多,且梨果表皮光滑,易形成镜面反射区域,这些区域极易丢失目标特征信息,从而导致模型产生漏检或误检。改进模型能够有效地降低因光照过强而产生的漏检,验证了改进模型在光照较强环境下的鲁棒性。

2.5.2 小目标检测效果

图 10 为小目标梨果较多检测场景下,基线模型与改进模型检测效果。

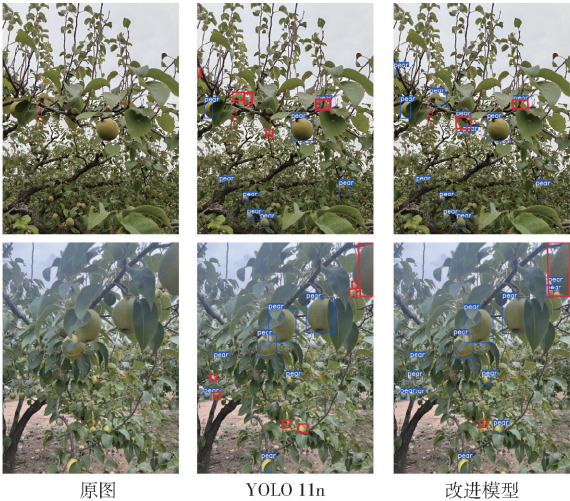


图 10 小目标检测效果

Fig. 10 Detection performance on small objects

在该场景下,由于目标梨果像素占比较小、且目标梨果颜色与梨树树叶颜色接近,目标特征在模型下采样过程中易丢失。这导致深层特征图难以提供足够的特征判别信息,从而引发大范围的漏检与误检现象。改进模型通过优化特征提取结构并加强自上而下的深层语义信息与浅层高分辨率特征的融合,显著提升了网络对小尺度目标的表征能力,有效降低了小目标漏检率。该结果表明,改进模型在应对小目标果实检测任务中具有良好的鲁棒性与有效性。

2.5.3 阴天场景下检测效果

图 11 为阴天不同密集程度下,基线模型与改进模型检测效果。

在阴天梨果稀疏场景下,存在光照不足、对比度较低等问题,导致模型提取特征困难,极易将颜色相近叶片误检为目标果实,使模型产生更多误检,改进模型能够有效地降低误检目标数量。在阴天梨果密集场景下,由于光线柔和,缺乏阴影,使相邻梨果之间的边界模糊不清,以及果实亮度不均等问题,导致模型提取目标的有效真实特征较难,在实际识别过程中,极易产生因边缘模糊而导致的重复检测或漏检。改进模型降低了漏检目标数量,同时能够更加精确的识别果实与果实轮廓,验证了改进模型在光

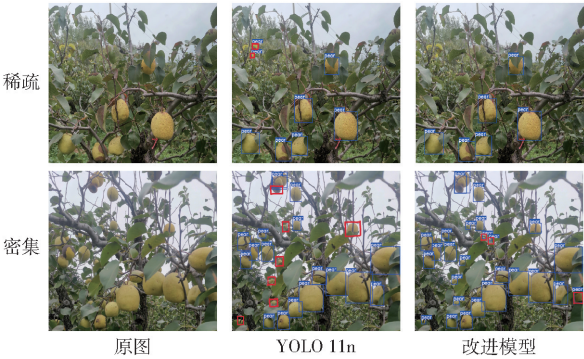


图 11 阴天场景下检测效果

Fig. 11 Detection performance in cloudy scenarios

照不足环境下模型的鲁棒性。

2.5.4 遮挡目标检测效果

图 12 为遮挡目标较为严重的场景,基线模型与改进模型检测效果。

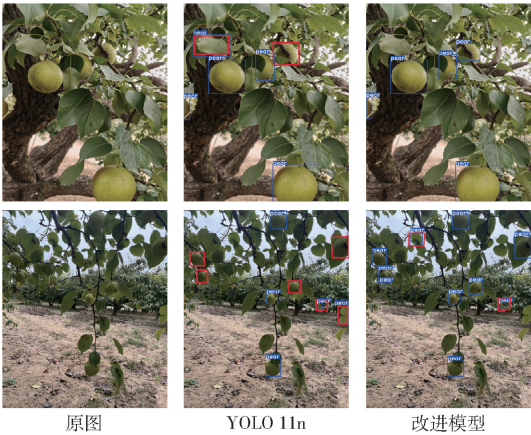


图 12 遮挡目标检测效果

Fig. 12 Detection performance for occluded objects

在遮挡情况普遍存在的检测场景下,目标常因被枝叶或相邻果实部分遮盖而出现外形不完整或形态异常,导致其表观信息残缺,显著增加了模型识别与定位的难度。改进模型通过引入自监督形状感知特征增强模块并结合高效的注意力机制,使网络能够聚焦于目标的可见区域同时增强该区域的特征提取,并利用上下文信息推断被遮挡部分的可能位置。试验表明,该方法有效提升了对遮挡目标以及异形果实的召回率,验证了改进模型在复杂遮挡环境下良好的泛化能力。

2.5.5 重叠目标检测效果

图 13 为目标重叠较多检测场景下,基线模型与改进模型检测效果。

在目标间重叠度较高的检测场景下,相邻果实边界模糊,导致模型难以对单个实例的边界进行有效区分,易产生重复检测或漏检。改进模型通过增强网络的自适应采样能力,并利用更有效的多尺度特征交互与融合策略,提升了模型对重叠目标的轮

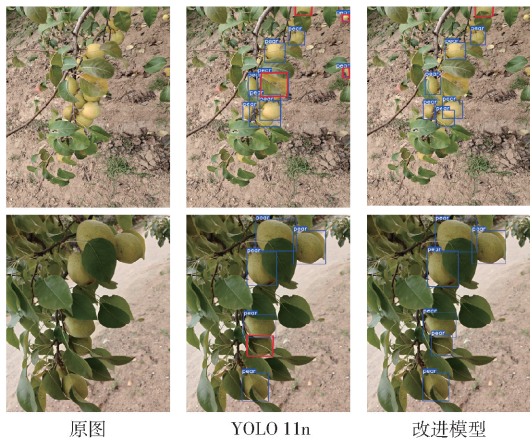


图 13 重叠目标检测效果

Fig. 13 Detection performance for overlapping objects

廓辨识能力与实例区分能力,从而显著减少了重复框与合并检测的现象,增强了在果实高度密集区域与相互重叠场景下的检测准确性。

3 结论

(1)设计了 SS-SAFE 模块,能够有效利用梨果

的形状先验知识,通过预测长宽比增强特征表达能力,有效地降低了漏检率,提升了模型检测覆盖能力,召回率提升 1.7 个百分点。

(2)采用的 C3K2-AKConv 模块可根据目标尺度自适应调整感受野与采样位置,提高了模型对不同尺度梨果的特征提取能力与检测能力,有效地解决了部分异形梨果与远处小梨果难识别的问题。

(3)引入的 C2PSA-HSAN 模块能够高效地聚合多尺度上下文信息并建立长距离空间依赖关系,有效融合了不同尺度的特征信息,显著提升了模型在复杂场景下小目标与重叠目标的特征融合能力,改进后模型的平均精度均值在各场景下均有明显提升。

(4)改进模型在复杂自然场景下具备更强的梨果检测能力,在梨果检测数据集上,精确率、召回率、F1 值、mAP@ 0.5 和 mAP@ 0.5:0.95 分别达到 88.9%、81.7%、85.2%、88.6% 和 67.2%,各指标均优于其他模型。本文方法可为梨园智能化管理与自动化采摘提供技术支撑。

参 考 文 献

[1] 张绍铃, 谢智华. 我国梨产业发展现状、趋势、存在问题与对策建议[J]. 果树学报, 2019, 36(8): 1067-1072.
Zhang Shaoling, Xie Zhihua. Current status, trends, main problems and the suggestions on development of pear industry in China[J]. Journal of Fruit Science, 2019, 36(8): 1067-1072. (in Chinese)

[2] Zhou H, Wang X, Au W, et al. Intelligent robots for fruit harvesting: recent developments and future challenges[J]. Precision Agriculture, 2022, 23(5): 1856-1907.

[3] 宋鹏, 李正达, 杨蒙, 等. 作物表型机器人研究现状与展望[J]. 农业机械学报, 2025, 56(3): 1-17.
Song Peng, Li Zhengda, Yang Meng, et al. Current status and prospects of crop phenotyping robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 1-17. (in Chinese)

[4] Linker R, Cohen O, Naor A. Determination of the number of green apples in RGB images recorded in orchards[J]. Computers and Electronics in Agriculture, 2012, 81: 45-57.

[5] Lu J, Sang N. Detecting citrus fruits and occlusion recovery under natural illumination conditions[J]. Computers and Electronics in Agriculture, 2015, 110: 121-130.

[6] 李柯泉, 陈燕, 刘佳晨, 等. 基于深度学习的目标检测算法综述[J]. 计算机工程, 2022, 48(7): 1-12.
Li Kequan, Chen Yan, Liu Jiachen, et al. Survey of deep learning-based object detection algorithms[J]. Computer Engineering, 2022, 48(7): 1-12. (in Chinese)

[7] 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.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[PP/OL]. V1. arXiv(2020-04-23) [2026-01-15]. <https://doi.org/10.48550/arXiv.2004.10934>.

[10] 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.

[11] Fu L, Feng Y, Majeed Y, et al. Kiwifruit detection in field images using Faster R-CNN with ZFNet[J]. IFAC-PapersOnLine, 2018, 51(17): 45-50.

[12] 武星, 齐泽宇, 王龙军, 等. 基于轻量化 YOLOv3 卷积神经网络的苹果检测方法[J]. 农业机械学报, 2020, 51(8): 17-25.
Wu Xing, Qi Zeyu, Wang Longjun, et al. Apple detection method based on Light-YOLOv3 convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 17-25. (in Chinese)

[13] Koirala A, Walsh K B, Wang Z, et al. Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of ‘MangoYOLO’[J]. Precision Agriculture, 2019, 20(6): 1107-1135.

- [14] 朱立成, 王文贝, 赵博, 等. 基于 SDE - YOLO 的矮砧密植化果园苹果检测方法[J]. 农业机械学报, 2025, 56(9): 638 - 647.
Zhu Licheng, Wang Wenbei, Zhao Bo, et al. Detection method of apples in dwarf rootstock densification orchards based on SDE - YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(9): 638 - 647. (in Chinese)
- [15] 杨国亮, 孙刘衍, 洪鑫芳. 果园复杂环境下的酥梨检测算法研究[J]. 江苏农业科学, 2025, 53(11): 209 - 219.
Yang Guoliang, Sun Liuyan, Hong Xinfang. Study on pear detection algorithm in orchard complex environment[J]. Jiangsu Agricultural Sciences, 2025, 53(11): 209 - 219. (in Chinese)
- [16] 罗志聪, 李鹏博, 宋飞宇, 等. 嵌入式设备的轻量化百香果检测模型[J]. 农业机械学报, 2022, 53(11): 262 - 269, 322.
Luo Zhicong, Li Pengbo, Song Feiyu, et al. Lightweight passion fruit detection model based on embedded device[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(11): 262 - 269, 322. (in Chinese)
- [17] Tang Y, Han K, Guo J, et al. GhostNetv2: enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969 - 9982.
- [18] Howard A, Sandler M, Chu G, et al. Searching for mobilenetv3[C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314 - 1324.
- [19] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 - 7141.
- [20] 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.
- [21] Zhu X, Hu H, Lin S, et al. Deformable Convnets v2: more deformable, better results[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308 - 9316.
- [22] Liang J, Huang K, Lei H, et al. Occlusion-aware fruit segmentation in complex natural environments under shape prior[J]. Computers and Electronics in Agriculture, 2024, 217: 108620.
- [23] Zeng Y, Chen Y, Yang X, et al. ARS-DETR: aspect ratio-sensitive detection transformer for aerial oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1 - 15.
- [24] Dong Y, Wei M, Gao G, et al. SARFA - Net: shape-aware label assignment and refined feature alignment for arbitrary-oriented object detection in remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025, 18: 8865 - 8881.
- [25] Zhang J, Yang L, Mohammadabadi S M S, et al. A survey on self-supervised learning: recent advances and open problems[J]. Neurocomputing, 2025, 655: 131409.
- [26] 曹敏, 王尧. 基于稀疏形状先验与变分正则的典型红外目标分割[J]. 红外技术, 2025, 47(5): 611 - 618.
Cao Min, Wang Yao. Typical infrared object segmentation based on sparse shape prior and variational regularization[J]. Infrared Technology, 2025, 47(5): 611 - 618. (in Chinese)
- [27] Mi Z. YOLOv7-based concrete surface defect detection algorithm with enhanced AKConv[C] // 2025 5th International Symposium on Computer Technology and Information Science (ISCTIS). IEEE, 2025: 905 - 912.
- [28] Gupta S, Tripathi A K. Flora - NET: integrating dual coordinate attention with adaptive kernel based convolution network for medicinal flower identification[J]. Computers and Electronics in Agriculture, 2025, 230: 109834.
- [29] Li Y, Zhang M, Zhang C, et al. YOLO - CCS: vehicle detection algorithm based on coordinate attention mechanism[J]. Digital Signal Processing, 2024, 153: 104632.