doi:10.6041/j.issn.1000-1298.2026.05.011

基于改进 YOLO v8n - ByteTrack 的温室番茄果实巡检计数方法

冯青春[1,2]　李豪博[1,2]　陈诗琪[1,2]　刘　静[2]　李亚军[2]

(1. 江苏大学农业工程学院, 镇江 212013; 2. 北京市农林科学院智能装备技术研究中心, 北京 100097)

摘要: 番茄果实数量的精确统计是产量评估和智能化管理的重要基础。为实现温室环境下丛生枝叶间果实动态检测与计数,提出了以改进 YOLO v8n 检测模型和 ByteTrack 跟踪算法组合为主要框架的番茄果实动态计数方法。在 YOLO v8n 检测模型中引入小波下采样模块和 P2 检测头,并设计 MLLA 注意力机制,提高模型在复杂背景下的检测性能,并基于 ByteTrack 引入自适应低分匹配重试策略。最后,基于改进后的 ALRM - Track 提出一种基于序列区域匹配机制计数。试验结果表明,改进后的 YOLO v8n - MAFP 模型在番茄果实数据集上的检测平均精度均值达到 96.9%,较原始模型提升 2.4 个百分点;结合 YOLO v8n - MAFP 改进的 ALRM - Track 算法,多目标跟踪准确率提升至 88.2%,ID 跳变次数减少至 2 次;采用序列区域匹配机制计数,在 5 组温室视频试验中的平均绝对误差仅为 1.4,平均计数精度达到 96.6%,显著优于传统划线计数和区域计数方法。温室环境试验表明,基于改进 YOLO v8n - ByteTrack 模型适用于温室番茄的估产统计需求。

关键词: 温室; 番茄; 果实计数; 生产巡检; 多目标追踪; YOLO v8n; ByteTrack

中图分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)05-0127-11　　**OSID**:

Greenhouse Tomato Fruit Inspection and Counting Method Based on Improved YOLO v8n - ByteTrack

FENG Qingchun[1,2]　LI Haobo[1,2]　CHEN Shiqi[1,2]　LIU Jing[2]　LI Yajun[2]

(1. *College of Agricultural Engineering*, *Jiangsu University*, *Zhenjiang* 212013, *China*

2. *Intelligent Equipment Technology Research Center*, *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*)

Abstract: Accurate counting of tomato fruits is crucial for yield assessment and intelligent management. To achieve dynamic detection and counting of fruits among clustered branches in greenhouse environments, a dynamic tomato fruit counting method was proposed, primarily based on an improved YOLO v8n detection model and the ByteTrack tracking algorithm. A wavelet downsampling module and a P2 detection head were introduced into the YOLO v8n detection model, and an MLLA attention mechanism was designed to improve the model's detection performance in complex backgrounds. An adaptive low-score matching retry strategy was introduced based on ByteTrack. Finally, a counting method based on sequence region matching was proposed by using the improved ALRM - Track. Experimental results showed that the improved YOLO v8n - MAFP model achieved an average detection accuracy of 96.9% on the tomato fruit dataset, a 2.4 percentage points improvement over the original model. Combined with the improved ALRM - Track algorithm of YOLO v8n - MAFP, the multi-target tracking accuracy was increased to 88.2%, and the number of ID switching operations was reduced to 2. Using a sequence region matching mechanism for counting, the average absolute error in five greenhouse video experiments was only 1.4, with a counting accuracy of 96.6%, significantly outperforming traditional line-based counting and region-based counting methods. Greenhouse environment experiments demonstrated that the improved YOLO v8n - ByteTrack model was suitable for the statistical needs of greenhouse tomato yield estimation.

Key words: greenhouse; tomato; fruit counting; production inspection; multi-target tracking; YOLO v8n; ByteTrack

收稿日期: 2025-12-22　修回日期: 2025-12-29

基金项目: 北京市农林科学院创新能力建设专项(KJCX20260901)

作者简介: 冯青春(1987—),男,研究员,博士,主要从事农业机器人关键技术和产业应用研究,E-mail: fengqc@nercita.org.cn

通信作者: 李亚军(1995—),男,助理研究员,博士,主要从事人工智能与农业机器人技术研究,E-mail: liyj@nercita.org.cn

0 引言

番茄精准计数[1]是设施农业智能化管理中的一项重要基础性工作[2-3],准确的番茄数量是温室实现番茄产量评估、精准水肥管理和合理调配资源的前提。在温室环境下,番茄数量统计主要依赖巡检机器人[4]在种植行间的动态巡检。然而,温室内番茄分布密集且容易出现枝叶遮挡与果实堆叠等情况,传统的人工计数方法不仅效率低下、劳动成本高[5],而且容易产生漏数或重复计数等问题。此外,人工干预频繁进入温室会破坏种植环境,影响植株正常生长。因此,设施农业迫切需要研究一种基于智能感知的自动化、高精度番茄计数方法,以提升巡检效率和计数准确性[6-9]。

番茄计数方法主要分为基于图像的静态计数与基于视频的动态计数。其中基于图像的静态计数主要是基于 Fast R-CNN[10-11]、YOLO(You only look once)等[12-13]目标检测算法进行检测计数,从而获取到视场内果实的产量。文献[14]提出一种基于改进 Fast R-CNN 模型识别刺梨果实的识别方法,平均识别精度为 92.01%,但单幅图像识别时间为 200 ms,检测时间过长。文献[15]通过提出一种改进 YOLO v8n 的方法实现油茶果的识别计数,最终检测精度为 81.9%。文献[16]提出了一种基于改进 YOLO v5n 的猪只盘点算法,实现了猪只盘点的功能,单幅图像识别速度满足实时要求。但静态计数受视场大小的局限性、路径带来的视角变化等影响,难以通过拼接图像获得大区域的精准产量估计,这种静态计数难以满足生产过程中实时测产的要求。

基于视频的动态计数主要是将目标检测算法与 MOT、DeepSORT 等多目标跟踪[17-19]算法相结合实现番茄计数。这类方法一般是先对每一帧的果实进行识别定位,然后再通过跟踪算法进行目标匹配,实现在连续帧跟踪同一目标,避免了目标的重复计数。文献[20]利用 YOLO v4-tiny 和改进匈牙利算法对苹果视频进行识别与计数,计数精度达到 69.14%。文献[21]通过 YOLO v7 模型结合 DeepSORT 算法对猕猴桃跟踪,基于匈牙利算法匹配得到唯一 ID 来估算猕猴桃产量,准确率为 89.87%,跟踪速度为 20.19 f/s。然而,这些研究仅通过跟踪算法匹配得到的 ID 进行计数,但是受枝叶遮挡与果实重叠的影响,导致对目标进行匹配失败,使得 ID 发生变化,大大增加了漏检和重复计数的概率。

综上,通过单一目标检测算法[22]实现番茄计数时,其计数精度易受温室枝叶遮挡和果实聚集等因素影响。基于目标检测[23]与多目标跟踪结合的方法虽然在一定程度上提升了计数精度[24],但往往增加了计算开销,并在果实数量快速变化时存在不稳定性。因此,现有方法多适用于试验环境下的番茄计数,难以完全满足温室复杂动态环境[25]下的实时巡检计数精准需求。尽管已有研究尝试在巡检机器人平台上实现动态计数[26],但算法仍然存在精度不高和实时性不强等问题,难以直接应用于实际生产场景。

针对上述问题,提出基于 YOLO v8n-ByteTrack 的番茄精准计数方法。基于目标检测算法 YOLO v8n 对主干网络进行改进,并引入注意力机制,以提高番茄在复杂环境下的识别精度;基于多目标跟踪算法 ByteTrack 引入自适应低分匹配重试策略;基于 ALRM-Track 跟踪算法建立序列区域匹配机制计数,在真实环境中对番茄进行动态计数,以提高番茄计数精度,满足温室番茄估产统计需求。

1 图像数据集获取与处理

1.1 温室番茄植株图像采集

训练与测试数据集由安装在巡检机器人升降杆上的海康威视萤石相机连续拍摄采集,图像分辨率设定为 1 280 像素×720 像素,采集帧率为 30 f/s。试验环境及设备部署如图 1 所示。采集过程中,巡检机器人以 0.5 m/s 的恒定速度沿图中红色箭头指示的路径行进,相机距地面高度 0.6 m,并与邻近植株保持 0.4 m 的横向间距,以确保其视场能够完整覆盖有效监测区域(图中淡蓝色区域),即植株 80 cm 高度范围内的熟果分布区。

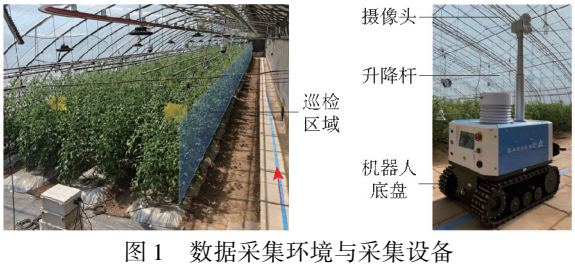


图 1 数据采集环境与采集设备

Fig. 1 Data collection environment and devices

1.2 数据集构建

采用 LabelImg 工具对采集的原始图像数据进行人工标注。标注过程中,仅对能够清晰识别的番茄目标及被遮挡比例小于 50% 的果实进行标记,对位于垄后或掉落在地面的番茄不予标注。标注方式采用目标的最小外接矩形框作为真实标注框,以保证目标边界的一致性。

为提升检测模型的泛化能力和鲁棒性,在数据集划分后,分别对训练集和验证集进行了相应的数

据增强与预处理。首先，将原始 900 幅图像按照比例 7∶3随机划分为训练集和验证集，确保训练与评估数据彼此独立。随后，对训练集实施了多样化的数据增强操作，包括随机水平翻转、添加高斯噪声、运动模糊、随机裁剪与填充、亮度和对比度调整，将训练集扩充至 2 000 幅，显著提升了样本多样性。其中，训练集用于模型参数学习，验证集用于模型性能评估和超参数调优。

2　番茄识别与跟踪算法

2.1　基于改进 YOLO v8n 模型的番茄检测

在 YOLO v8n 模型框架基础上，针对温室环境下丛生枝叶间果实识别的问题，对模型进行改进并提出有针对性的 YOLO v8n－MAFP 网络优化策略。

首先，在主干网络中引入 Down_wt 小波下采样模块，并构建 P2 目标检测头与小波特征融合机制，在降低特征图分辨率的同时最大程度保留高频细节信息，扩充模型的多尺度检测范围，增强对遮挡目标的感知能力。其次，在主干网络末端嵌入 MLLA（Mamba－Inspired linear attention）注意力模块，利用其线性注意力机制与遗忘门设计，使网络能够自适应聚焦于与任务相关的关键特征。改进后的 YOLO v8n－MAFP 网络结构如图 2 所示。

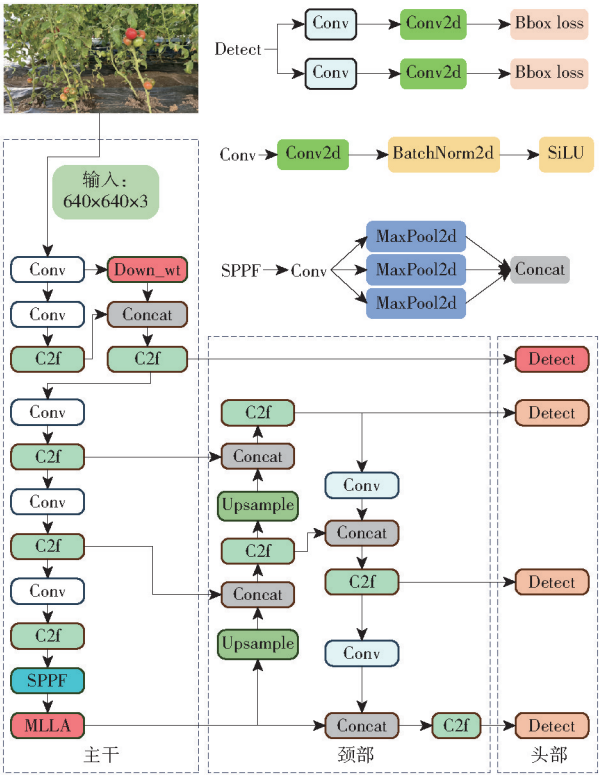


图 2　YOLO v8n－MAFP 网络模型结构图

Fig. 2　Network structure of YOLO v8n－MAFP

2.1.1　小波下采样模块

针对遮挡造成的果实边界模糊，传统模型仅依赖单一尺度的特征融合，往往难以兼顾全局轮廓与局部纹理，极易造成果实漏检。为此，采用在主干网络中引入 Down_wt 小波下采样模块[28]，并结合新增的小目标检测头进行多尺度特征融合。利用小波变换的无损频域特性与 P2 层丰富的位置信息，使模型能精准提取微小目标的边缘细节，从而提升模型对遮挡番茄的检测精度。

小波特征下采样模块（Wavelet-based down sampling module，WDM）结构如图 3 所示，主要由无损特征编码（Lossless feature encoding）和特征学习（Feature learning）两个子阶段组成，分别对特征的频率分量和语义表达进行处理，旨在解决传统池化操作造成的空间信息丢失问题，提升神经网络对纹理和边缘的感知能力。该模块能够在保持较低计算成本的同时实现特征分辨率的下采样，并可嵌入至现有的卷积神经网络架构中。

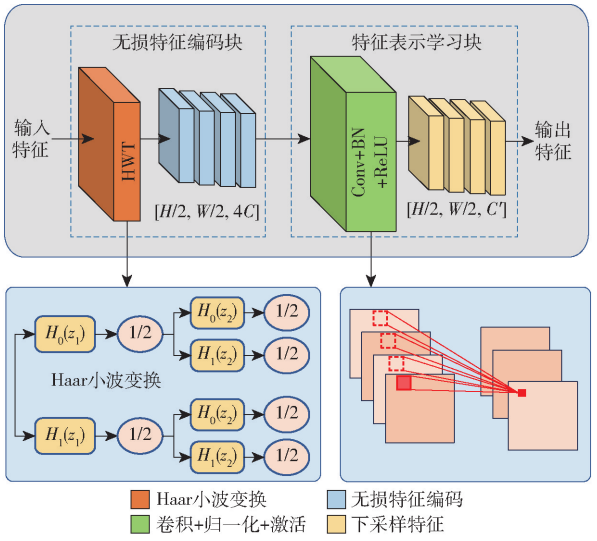


图 3　Dwon_wt 小波下采样模块结构图

Fig. 3　Dwon_wt down sampling module structure

为有效解决传统池化操作导致的信息丢失问题，提出小波特征下采样模块（WDM），采用“频域变换–特征融合”的串联架构对输入特征进行优化处理。无损特征编码块首先利用 Haar 小波变换将维度为 $H \times W \times C$ 的输入特征在空间维度进行离散分解，无损地将空间信息转换为 4 个频率子带并拼接为 $[H/2, W/2, 4C]$ 的张量，最大程度保留了全局结构与局部纹理细节；随后进入特征表示学习块，利用 Conv＋BN＋ReLU 复合单元对上述高维频率特征进行非线性映射与降维处理，旨在深度融合不同频率分量的信息并将通道压缩至 C'，最终生成兼具丰富语义与精细特征的输出特征。

2.1.2　MLLA 注意力模块

在真实温室场景中，番茄植株生长茂密，成熟果实常被厚重的枝叶遮挡或与邻近果实相互堆叠，导

致目标特征提取困难。针对这一问题,在 YOLO v8n 主干网络末端引入 MLLA 注意力模块[29]。该模块通过线性注意力机制从全局维度捕捉上下文信息,显著增强模型对被遮挡番茄关键区域的关注能力,提升特征提取的鲁棒性与表达效果。MLLA 模块整体结构如图 4 所示。

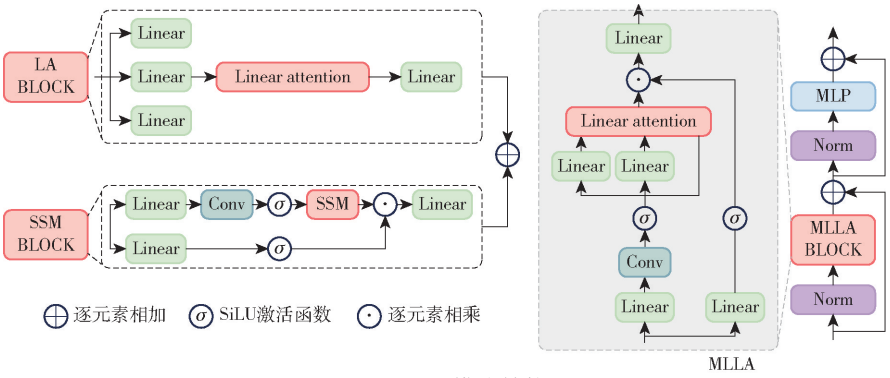


图 4 MLLA 模块结构图

Fig. 4 MLLA module structure

MLLA 模块旨在以线性的计算复杂度捕获全局长距离依赖信息。MLLA 模块通过重构注意力机制的运算方式,摒弃了标准 Transformer 中计算密集的 Softmax 操作,在大幅降低参数量和推理延迟的条件下,显著扩展了模型的全局感受野。该机制允许模型在保持高分辨率特征的同时,高效地聚合全局上下文信息,从而提升特征图的语义表达完整性。

MLLA 模块(图 4)主要由双流并行分支组成。首先,第 1 个分支为局部特征提取,采用线性映射与 3×3 深度卷积,以保留输入特征的局部与空间细节。第 2 个分支为全局线性注意力(对应图中的 LA BLOCK,内部集成 SSM BLOCK 以实现状态空间模型机制),该分支通过改进的线性注意力机制与遗忘门策略,有助于在低计算成本下获取长距离的全局上下文信息并抑制遮挡噪声。其次,利用逐元素相乘操作对两个分支的特征进行门控融合,使局部特征受到全局语义的动态加权。最后,通过 1×1 卷积进行特征投影,并结合残差连接恢复至原始维度,融合局部与全局信息,实现对复杂遮挡场景下番茄特征的精准表达。

2.2 基于改进 ByteTrack 的番茄果实跟踪

2.2.1 ByteTrack 算法原理

ByteTrack[30] 是一种基于 TBD(Tracking by detection)范式的高效多目标跟踪算法,其核心思想在于充分利用低置信度检测框参与数据关联,突破了传统多目标跟踪算法仅依赖高置信度检测框进行匹配的限制。在每一帧中,ByteTrack 采用二级级联匹配策略完成轨迹关联:首先,将高置信度检测框与已有轨迹进行匹配,优先保证高分目标的连续性;随后,将未匹配成功的轨迹与低置信度检测框进行第 2 次关联。

在 ByteTrack 的第 2 阶段关联中,算法首先计算低置信度检测框与未匹配轨迹预测框之间的 IoU 矩阵,随后利用匈牙利算法求解全局最优匹配。最终,算法引入一个固定的 IoU 阈值对匹配结果进行验证:仅当匹配对的 IoU 高于该阈值时,才确认为关联成功。然而,在温室番茄巡检场景中,这种固定阈值策略显露出极大的局限性。当番茄受到枝叶严重遮挡或相机快速移动时,检测框会存在形变,卡尔曼滤波的预测位置可能出现偏差。这导致同一目标的检测框与预测框 IoU 值可能短暂下降至固定阈值以下。尽管匈牙利算法正确地将它们关联在了一起,但由于未达到固定阈值,导致低分检测框被丢弃。因此,设计一种能根据目标状态动态调整的自适应阈值机制显得尤为重要。

2.2.2 自适应 IoU 阈值

为了解决原生 ByteTrack 算法中使用固定 IoU 阈值导致在果实目标遮挡或运动突变场景下匹配失效的问题,设计了一种基于统计特性的自适应阈值策略(Adaptive threshold strategy),该策略的核心思想是利用当前帧的 IoU 分布统计量来量化匹配的难度,并据此动态调整关联阈值。具体流程如下:

首先,算法基于当前帧的低分检测框集合与卡尔曼滤波预测的轨迹集合构建 IoU 代价矩阵。随后,计算每个检测框与所有活跃轨迹的最大重叠程度,以评估该目标的潜在匹配质量。对于第 i 个检测框 d_i,其最大 IoU 值 u_i 定义为

$$u_i = \max_{j=1}^{N} \mathrm{IoU}(d_i, t_j) \tag{1}$$

式中 t_j——第 j 条活跃轨迹

N——当前帧活跃轨迹的总数量

u_i 反映了第 i 个检测框在当前时刻找到对应轨迹的最大可能性。

然后,为了获得具有鲁棒性的全局匹配难度估

计，选取所有检测框 u_i 值的中位数作为基准。相较于均值，中位数能够有效抵抗个别极端离群值（如完全错误的检测框）的干扰。自适应基准值 μ 的计算公式为

$$\mu = \text{median}(\{u_1, u_2, \cdots, u_M\}) \tag{2}$$

式中　M——当前帧低分检测框的数量

当 μ 较高时，表明大部分检测框与轨迹重叠良好，场景简单；当 μ 较低时，表明普遍存在遮挡或预测误差，最后，基于前期试验中对检测框 IoU 分布特性的统计分析，为了防止阈值过低导致误匹配或过高导致漏匹配，将最终的关联阈值 τ_{adapt} 限制在合理的经验区间（0.2，0.5）内，即

$$\tau_{adapt} = \max(0.2, \min(0.5, \mu)) \tag{3}$$

2.2.3　基于自适应阈值的低分匹配重试策略

针对温室复杂环境下因枝叶、果实遮挡导致的目标检测置信度剧烈波动，进而引发轨迹中断与 ID 频繁切换（ID switch）问题，在 ByteTrack 算法的基础上，提出了一种自适应低分重试匹配策略（Adaptive low-score re-matching，ALRM）。该策略构建了一个三级级联关联架构，通过动态调整匹配阈值与多轮次重试机制，显著提升了弱观测条件下目标的召回能力。ALRM－Track 的算法流程如图 5 所示，具体实施步骤如下：

（1）第 1 阶段：高置信度主要关联

首先，利用卡尔曼滤波预测当前帧轨迹的位置。将检测框按置信度划分为高分集合 D_{high} 与低分集合 D_{low}。在本阶段，采用严格的固定 IoU 阈值，优先将高分检测框 D_{high} 与活跃轨迹进行匈牙利匹配，以确保清晰目标的跟踪精度与纯度。

（2）第 2 阶段：低置信度的自适应关联

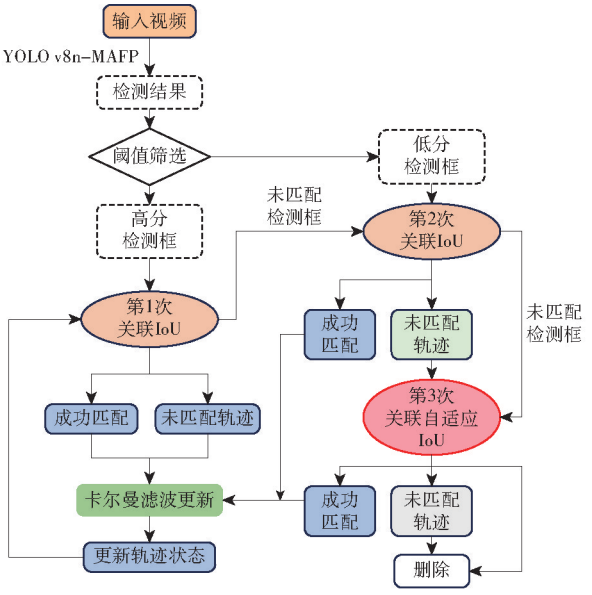


图 5　ALRM－Track 算流程图

Fig. 5　ALRM－Track algorithm flowchart

针对步骤（1）未匹配的剩余轨迹与低分检测框 D_{low}，传统的固定阈值往往难以适应检测框质量的动态变化。为此，本阶段引入自适应阈值机制。系统实时计算当前帧 D_{low} 中所有检测框的平均置信度，并据此动态调整 IoU 匹配阈值 τ_{adapt}。当平均置信度下降时，适当降低 τ_{adapt} 以增强对模糊目标的包容度，从而在新的自适应阈值下完成二次匹配。

（3）第 3 阶段：基于自适应阈值的重试匹配

为了进一步挖掘潜在的有效目标，防止因局部遮挡导致的轨迹意外中断，设计了第 3 级“重试”机制。对于前两阶段处理后仍未匹配的孤立轨迹，系统采用极低的 IoU 阈值进行最后一次宽容匹配尝试。该阶段旨在捕捉那些检测评分极低、且位置偏差较大，但实际上属于同一目标的“极限样本”。

2.3　基于序列区域匹配的果实计数

为了避免动态巡检计数过程中目标的重复计数，在 ALRM－Track 对特定番茄果实跟踪后，采用序列区域匹配机制对番茄果实精准计数，如图 6 所示。

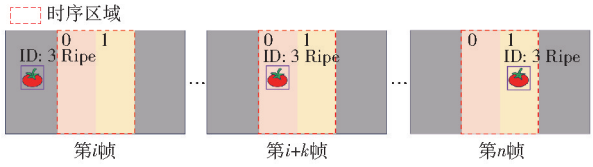


图 6　序列区域匹配机制计数

Fig. 6　Sequence region matching mechanism counting

算法如下：

输入：每帧 t 的跟踪番茄集合（包含中心点 p、唯一 ID、类别标签）；缓存列表 List0（区域 0 的 ID）、List1（区域 1 的 ID）

输出：成熟番茄的计数结果 O

```
Initialize List0 = { }, List1 = { }, O = 0
for t = 1→N do
    for c_i in t do        // c_i：当前帧中的每一个跟踪到的番茄
        p ← center_point(c_i) //p：番茄的中心点坐标
        id ← ID(c_i)       // id：番茄的唯一跟踪 ID
        if p ∈ Region 0 then
            if id ∉ List0 then
                Add id ← List0
            end if
        else if p ∈ Region 1 then
            if id ∉ List1 then
                Add id → List1
            end if
            if id ∈ List0 then
                O = O + 1
```

```
            Remove id from List0
            Remove id from List1
        end if
    end if
end for
end for
return O        // O 是当前帧计数结果
```

该机制通过构建基于中轴线时序区域匹配实现精准计数。该方法在图像中轴线两侧分别构建宽度为 150 像素的前序(Region 0)与后序(Region 1)匹配缓冲区,以有效兼容帧间像素间隔、目标中心波动及短时遮挡干扰。在此基础上,算法遵循严格的时序跨越逻辑,当跟踪目标中心点依次经过前序注册区与后序检索区,且后序捕获 ID 在前序列表中匹配成功时,判定为一次有效跨越并触发计数器累加,随即清除双侧列表中的对应记录以实现状态闭环,从而确保果实计数的唯一性与稳定性。

3 试验与结果分析

3.1 试验环境与模型训练

试验中的番茄检测、追踪及计数算法的开发与测试均在高性能计算工作站上完成。硬件平台搭载 Intel Core i7 - 13700K CPU,配备 64 GB 内存,图像加速处理采用 NVIDIA GeForce RTX 4090 (24 GB)显卡。操作系统为 Windows 11 专业版。软件环境基于 Python 3.8 编程语言,深度学习框架采用 PyTorch 2.4.0 及其配套的 TorchVision 0.19.0,图像处理基础库为 OpenCV 4.10.0,可视化部分则结合 Pillow 库与自定义绘图函数实现。

在检测模型训练阶段,为了加速模型收敛并提升特征提取能力,采用基于预训练权重的迁移学习策略。训练过程中,优化器选用随机梯度下降(SGD),动量系数设置为 0.937,权重衰减系数为 0.000 5。初始学习率设定为 0.000 1,批量大小(Batch size)设置为 16。在此参数设置下,通过观察训练期间的损失函数变化趋势可知,模型收敛状态稳定,且未出现过拟合现象。此外,通过随机裁剪、旋转变换等数据增强策略,扩充训练样本的多样性,从而显著提升模型在复杂场景下的泛化性能。

3.2 性能评价指标

为全面评估所提方法的性能,检测部分采用精确率、召回率、mAP@0.5、mAP@0.5:0.95 及模型参数量作为评价指标,mAP@0.5 表示 IoU 为 0.5 的平均精度,mAP@0.5:0.95 表示 IoU 阈值从 0.5 到 0.95 的平均精度。跟踪部分使用 MOTA、HOTA、IDF1 和 IDSW 衡量多目标跟踪效果。其中,MOTA 综合反映漏检、误检及 ID 切换情况;HOTA 同时考察检测、关联和定位能力;IDF1 用于衡量多目标识别与跟踪的准确性;IDSW 表示同一目标的 ID 变更次数,值越小越稳定。计数任务采用绝对误差、平均绝对误差(MAE)和平均计数精度(ACP)对计数结果进行量化评价。

3.3 基于改进 YOLO v8n 番茄检测结果与分析

3.3.1 消融试验

基于 YOLO v8n 模型的改进模型,在骨架网络中引入 Down_wt 小波下采样模块和检测头,并在主干网络末端嵌入 MLLA 注意力模块。为验证改进后 YOLO v8n - MAFP 模型的合理性和有效性,在保持其他训练参数一致的情况下,根据性能评价指标设置了消融试验,试验结果如表 1 所示。

表 1 消融试验结果

Tab.1 Ablation test results

模型	Down_wt	MLLA	精确率/%	召回率/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量
YOLO v8n	×	×	87.7	88.0	94.5	72.1	3.01×10^6
改进	√	×	91.3	84.6	95.2	73.7	3.13×10^6
	×	√	87.9	92.3	95.7	75.1	3.05×10^6
	√	√	92.5	90.6	96.9	75.7	3.08×10^6

注:“√”表示采用此改进方法,“×”表示未采用,下同。

引入 Down_wt 模块后,模型在多个关键指标上均表现优越,尤其在检测精度方面提升显著。模型的精确率和平均精度(mAP@0.5)提升了 3.6 个百分点和 0.7 个百分点,模型的精确率达到 91.3%,平均精度提升至 95.2%。表明 Down_wt 模块的检测头与小波特征融合机制有效增强了模型对目标的识别能力。在主干网络中引入 MLLA 模块后,虽然精确率仅提升 0.2 个百分点,但召回率显著提升了 4.3 个百分点,mAP@0.5:0.95 提升了 3.0 个百分点。这表明 MLLA 模块通过聚合全局上下文信息,更适应于增强复杂遮挡场景下的目标检出能力。当结合 Down_wt 和 MLLA 后,模型的精确率和召回率分别达到 92.5% 和 90.6%,且参数量仅增加 7×10^4,几乎不影响推理效率。模型所有指标均达到最优性能,证明了模块组合的有效性。

3.3.2　经典检测模型对比试验

为了全面评估所提出的 YOLO v8n－MAFP 模型在番茄目标检测任务中的性能表现，选取多种具有代表性的轻量化检测模型进行对比，包括 YOLO v8n、YOLO v5s、YOLO v7－tiny、YOLO v9n 以及最新的轻量化模型 YOLO 11n。所有模型均使用相同的训练集和测试集，以确保对比结果在数据条件上的一致性和可比性。各模型在精确率、召回率、mAP@0.5 以及模型参数量方面的性能对比如表 2 所示。

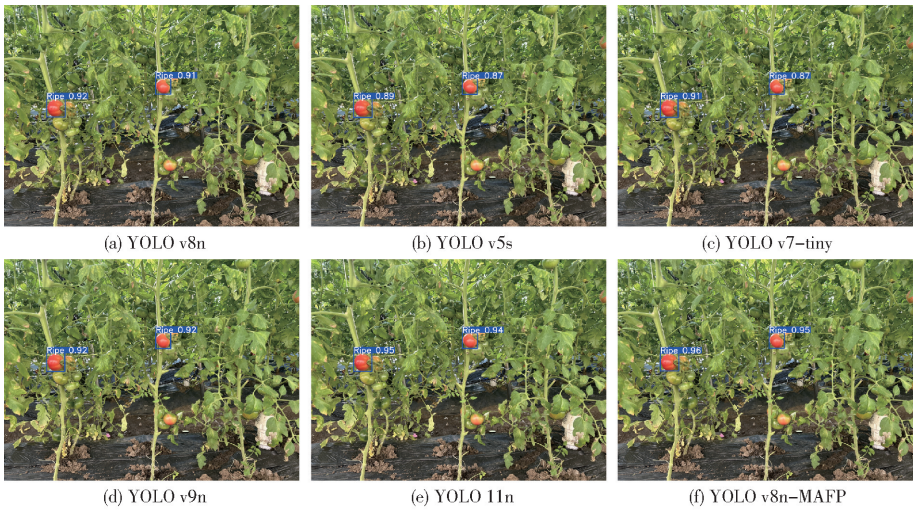
表 2　经典检测模型性能对比

Tab.2　Performance comparison of classic detection models

模型	精确率/%	召回率/%	mAP@0.5/%	参数量
YOLO v8n	87.7	88.0	94.5	3.01×10^6
YOLO v5s	86.1	86.5	93.4	7.20×10^6
YOLO v7－tiny	88.4	87.5	94.2	5.90×10^6
YOLO v9n	87.2	88.9	95.1	2.90×10^6
YOLO 11n	89.5	89.2	96.6	2.70×10^6
YOLO v8n－MAFP	92.5	90.6	96.9	3.08×10^6

通过表 2 可知，在检测精度方面，改进后模型的精确率和召回率、平均精度均值均优于 YOLO v8n、YOLO v5s、YOLO v7－tiny，与 YOLO v9n 相比，YOLO v8n－MAFP 模型的精确率、召回率和 mAP@0.5 分别提升 5.3、1.7、1.8 个百分点。与 YOLO 11n 相比，改进后的模型在精确率、召回率和 mAP@0.5 上分别提高 3.0、1.4、0.3 个百分点。该模型在参数量仅为 3.08×10^6 的轻量化结构下，实现了检测精度的显著提升。在复杂度方面，其参数规模大幅低于 YOLO v5s（7.20×10^6）和 YOLO v7－tiny（5.90×10^6），与 YOLO v8n 基本持平；尽管略高于 YOLO v9n（2.90×10^6）和 YOLO 11n（2.70×10^6），但换取了更优的识别表现。综合评估模型精度与复杂度，该方法展现出更佳的性能平衡，更胜任复杂环境下的番茄识别任务。

为验证改进 YOLO v8n－MAFP 模型在实际场景中的检测性能，从番茄数据集中选取光照充足场景进行评估，旨在全面评估改进模型在实际巡检中的检测精度和鲁棒性表现。各模型检测效果如图 7 所示，通过观察图 7a～7c 发现，在背景光线充足的情况下，YOLO v5s 和 YOLO v7－tiny 模型置信度均低于 YOLO v8n 模型；通过对比图 7a、7d、7e、7f 发现，虽然 YOLO v9n 和 YOLO 11n 在检测效果上均优于 YOLO v8n 模型，但是识别出的置信度低于改进后的模型。综合评估检测精度与实际应用表现，该模型在真实场景下的番茄识别任务中展现出更优异的综合性能。



(a) YOLO v8n　(b) YOLO v5s　(c) YOLO v7-tiny

(d) YOLO v9n　(e) YOLO 11n　(f) YOLO v8n-MAFP

图 7　经典检测模型检测效果对比

Fig.7　Comparison of detection performance of classic detection models

3.4　基于改进 ByteTrack 番茄多目标跟踪结果与分析

3.4.1　消融试验

为验证所提出改进策略的有效性，基于 ByteTrack 算法分别引入自适应 IoU 阈值与低分重试匹配策略，进行了消融试验。试验结果如表 3 所示：在仅引入自适应 IoU 阈值时，MOTA、HOTA 和 IDF1 分别提升至 86.7%、74.8% 和 85.6%，ID 跳变次数由 16 次下降至 13 次，说明该方法在一定程度上改善了跟踪的稳定性。当仅引入低分重试匹配策略时，MOTA、HOTA 和 IDF1 分别达到 85.1%、72.5% 和 83.2%，ID 跳变次数显著降低至 8 次，整体性能提升幅度更为明显。进一步将两种策略结合使用时，MOTA、HOTA 和 IDF1 分别提升至 88.2%、75.3% 和 87.9%，ID 跳变次数进一步降低至 2 次，相比原始 ByteTrack 算法在稳定性与准确性上均有

大幅提升。

表 3 改进 ByteTrack 消融试验结果

Tab. 3 Ablation test results of improved ByteTrack

模型	自适应 IoU 阈值	低分重试匹配策略	MOTA/%	HOTA/%	IDF1/%	ID 跳变次数/次
ByteTrack	×	×	84.9	71.8	82.5	16
改进	√	×	86.7	74.8	85.6	13
	×	√	85.1	72.5	83.2	8
	√	√	88.2	75.3	87.9	2

为了更加直观地说明本算法的有效性，原 ByteTrack 和改进的 ALRM－Track 在遮挡场景下的轨迹连续性对比如图 8 所示。采用 ByteTrack 对目标进行跟踪时，图 8a 展示了果实堆叠场景下的数据样本下算法未能保持有效跟踪，导致部分被遮挡的番茄目标丢失，即算法未能成功跟踪到被遮挡的番茄目标。而在图 8b 中对同一位置的目标跟踪时，算法成功跟踪到被遮挡目标。此外如图 8c 光线明亮的场景中，改进的算法能够稳定跟踪到视场中被遮挡的目标。表明了 ALRM－Track 算法在温室复杂场景中对目标跟踪的稳定性与优越性。



(a) 改进前 (b) 改进后 (c) 光线明亮

图 8 改进前后跟踪效果对比

Fig. 8 Comparison of tracking effects before and after improvement

3.4.2 经典目标跟踪算法对比试验

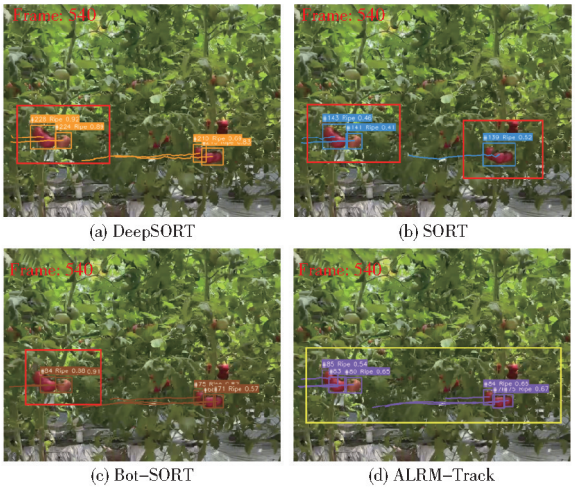
为验证改进算法的有效性，将其与 DeepSORT、SORT 及 Bot－SORT 等主流多目标跟踪算法进行了对比试验，详细数据如表 4 所示。结果显示，该方法在各项关键指标上均表现出显著优势。具体而言，ID 跳变次数降低至 2 次，反映出算法在维持目标身份一致性方面具有极高的稳定性；MOTA 提升至 88.2%，相较于 DeepSORT、Bot－SORT 和 SORT 分别增长了 6.5、4.6、13.3 个百分点；HOTA 与 IDF1 指标也分别达到 75.3% 和 87.9%，显著优于对比组算法。此外，该算法在保证高精度的同时兼顾了良好的实时性，平均帧率达 23.8 f/s，实现了检测精度与跟踪速度的有效平衡，能够满足温室巡检场景下的实时跟踪与计数需求。

表 4 经典跟踪算法对比

Tab. 4 Comparison of classic tracking algorithms

模型	ID 跳变次数/次	MOTA/%	HOTA/%	IDF1/%	平均帧率/(f·s^{-1})
DeepSORT	21	81.7	71.9	78.8	17.3
SORT	28	74.9	67.0	76.7	24.6
Bot－SORT	19	83.6	70.6	82.0	18.2
本文	2	88.2	75.3	87.9	23.8

改进后的 ALRM－Track 算法与 DeepSORT、SORT、Bot－SORT 的跟踪可视化对比结果如图 9 所示。观察发现，各算法在番茄密集遮挡场景下的表现差异显著：DeepSORT 与 SORT 算法（图 9a、9b）在红色标记处出现明显的跟踪偏差，倾向于将多个相邻番茄错误合并为单一目标；Bot－SORT 算法（图 9c）则存在漏匹配问题，未能建立有效的目标轨迹。相比之下，ALRM－Track 在同等条件下展现出更优的鲁棒性，实现了对所有番茄目标的稳定跟踪且无漏检，确立了高度遮挡下的轨迹连续性。这种性能提升主要得益于引入的自适应阈值与低分匹配重试机制，通过有效关联低置信度检测框，成功召回严重遮挡的目标，显著增强了算法对复杂环境的适应能力。



(a) DeepSORT (b) SORT

(c) Bot-SORT (d) ALRM-Track

图 9 改进后的 ALRM－Track 与 DeepSORT、SORT、Bot－SORT 算法跟踪结果对比

Fig. 9 Comparison of tracking results of ALRM－Track with DeepSORT, SORT, and Bot－SORT algorithms

3.5 经典计数方法试验结果与分析

基于改进的 YOLO v8n－MAFP 检测器与 ALRM－Track跟踪算法，选取 5 个现场采集的番茄视频序列，所有视频均采集于同一时段的光照环境下，但涵盖了果实与枝叶不同遮挡程度的场景，从而

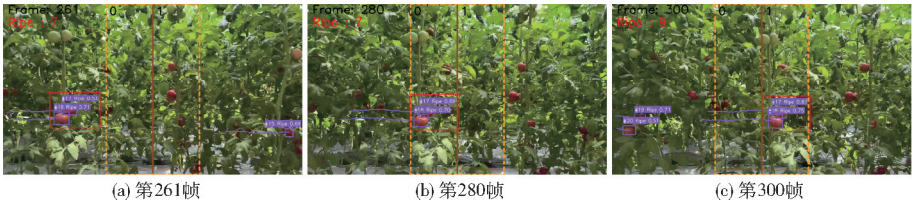
构建了差异化的视频序列。对划线计数、区域计数及序列区域匹配机制 3 种计数策略进行了对比试验(结果见表 5)。试验结果表明,序列区域匹配机制的累计绝对误差仅为 7 个,显著低于其余两种方法。此外,该机制在视频 03 和 05 中实现了零误差或近零误差,在其余场景中亦将偏差控制在 3 个以内,展现出极高的稳定性。相比之下,划线计数对边界遮挡较为敏感,易导致漏计,例如在视频 01 和 02 中分别出现 6 个和 8 个的绝对误差;区域计数则易受目标丢失与 ID 频繁切换的干扰,引发严重的过计现象,如视频 02 和 03 的误差分别高达 14 个和 12 个。综合指标显示,序列区域匹配机制的平均计数精度(ACP)高达 96.6%,大幅领先于划线计数(88.0%)与区域计数(75.1%),且平均绝对误差(MAE)低至 1.4 个,充分验证了其在高精度计数与稳定性方面的显著优势。

表 5　经典计数方法试验结果与分析

Tab. 5　Experimental results and analysis of classical counting methods

视频序列	真实数量(GT)/个	划线计数		区域计数		序列区域匹配机制计数	
		计数结果/个	绝对误差/个	计数结果/个	绝对误差/个	计数结果/个	绝对误差/个
01	46	40	6	53	7	45	1
02	51	43	8	65	14	48	3
03	37	35	2	49	12	37	0
04	34	29	5	43	9	33	1
05	41	37	4	51	10	41	2
平均绝对误差(MAE)/个		5		10.4		1.4	
平均计数精度(ACP)/%		88.0		75.1		96.6	

具体的计数试验流程如图 10 所示。在第 261 帧时,ID17 与 ID18 的番茄尚未进入前序区域 0,系统累计计数结果仍为 7。随后,在第 280 帧,ID17 与 ID18 的中心点进入区域 0,系统将其唯一标识符(ID)分别存入列表 0 中。随着视频播放至第 300 帧,ID17 与 ID18 的中心点进一步经过图像中轴线后序区域 1,此时系统将其 ID 同步写入列表 1,并在列表 0 中进行匹配检索。由于在列表 0 中成功检索到对应的 ID,系统判定该番茄完成了一次有效跨区域运动,计数器值由 7 增加至 9,并将该两条 ID 从列表 0 与列表 1 中同时移除,以避免重复计数。



(a) 第261帧　(b) 第280帧　(c) 第300帧

图 10　序列区域匹配机制计数试验结果

Fig. 10　Two-sided sequence matching mechanism counting experiment

4　结论

(1)针对温室番茄巡检中因枝叶、果实遮挡等复杂背景引起的检测与计数难题,构建了一种能够在复杂环境下实现稳定识别、连续跟踪与高精度计数的自动化方法。

(2)为解决小目标检测与遮挡干扰问题,模型在 YOLO v8n 基础上引入 Down_wt 小波下采样模块以无损保留频域细节,增设 P2 微小目标检测头以扩充多尺度检测范围,并融合 MLLA 注意力机制以强化对关键区域的特征聚焦。试验表明,改进后的 YOLO v8n－MAFP 的 mAP@0.5 达到 96.9%,较基线模型提升 2.4 个百分点,验证了其在复杂背景下对微小及被遮挡番茄具备更强的特征提取与辨识能力。

(3)针对巡检过程中因检测置信度波动导致的轨迹断裂现象,在 ByteTrack 框架中设计了自适应 IoU 阈值策略与低分重试机制。试验数据表明,改进算法的 MOTA、HOTA 与 IDF1 分别达到 88.2%、75.3% 和 87.9%,ID 跳变次数由 16 次大幅降至 2 次,同时平均帧率维持在 23.8 f/s。这表明该算法在显著提升跟踪连续性与稳定性的同时,有效兼顾了温室作业的实时性需求。

(4)为抑制动态巡检中因 ID 跳变引发的计数累计误差,基于 ALRM－Track 跟踪输出,构建了一种基于序列区域匹配的精准计数机制。该机制结合目标运动轨迹与中心点位置信息实施双重校验,有效过滤了重复计数与漏计现象。在 5 段实景视频测试中,该方法的平均计数精度(ACP)高达 96.6%,平均绝对误差(MAE)仅为 1.4 个,充分验证了该计数策略在动态巡检场景中具有极高的准确性与鲁棒性。

参 考 文 献

[1] MAHESWARI P, RAJA P, APOLO-APOLO O E, et al. Intelligent fruit yield estimation for orchards using deep learning based semantic segmentation techniques—a review[J]. Frontiers in Plant Science, 2021, 12: 684328.

[2] WANG S, JIANG H, YANG J, et al. Lightweight tomato ripeness detection algorithm based on the improved RT - DETR[J]. Frontiers in Plant Science, 2024, 15: 1415297.

[3] 赵九霄,张馨,史凯丽,等. 基于改进 Tiny - YOLO v5l 算法的串型番茄定位与计数[J]. 食品与机械,2022,38(12):79 - 86.
ZHAO Jiuxiao, ZHANG Xin, SHI Kaili, et al. Localization and counting of string tomatoes based on improved Tiny - YOLO v5l algorithm[J]. Food & Machinery,2022,38(12):79 - 86. (in Chinese)

[4] ZENG T, LI S, SONG Q, et al. Lightweight tomato real-time detection method based on improved YOLO and mobile deployment[J]. Computers and Electronics in Agriculture, 2023, 205: 107625.

[5] CÁMARA-ZAPATA J M, BROTONS-MARTÍNEZ J M, SIMÓN-GRAO S, et al. Cost-benefit analysis of tomato in soilless culture systems with saline water under greenhouse conditions[J]. Journal of the Science of Food and Agriculture, 2019, 99(13): 5842 - 5851.

[6] 王新, 唐灿, 朱建新, 等. 基于深度增强与特征抗噪的夜间串番茄成熟度识别方法[J]. 农业机械学报, 2025, 56(4): 363 - 374.
WANG Xin, TANG Can, ZHU Jianxin, et al. Tomato cluster ripeness recognition at night based on depth enhancement and feature noise reduction[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(4): 363 - 374. (in Chinese)

[7] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO - V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417 - 426.

[8] 赵春江. 智慧农业发展现状及战略目标研究[J]. 智慧农业,2019,1(1):1 - 7.
ZHAO Chunjiang. State-of-the-art and recommended developmental strategic objectivs of smart agriculture[J]. Smart Agriculture, 2019, 1(1): 1 - 7. (in Chinese)

[9] 蒋雪松, 肖伸平, 赵倩颖, 等. 基于 YOLO - DCL 的复杂环境油茶果遮挡检测与计数研究[J]. 农业机械学报, 2024, 55(10): 318 - 326,480.
JIANG Xuesong, XIAO Shenping, ZHAO Qianying, et al. *Camellia oleifera* fruits occlusion detection and counting in complex environments based on improved YOLO - DCL[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(10):318 - 326,480. (in Chinese)

[10] 周云成,许童羽,邓寒冰,等. 基于双卷积链 Fast R - CNN 的番茄关键器官识别方法[J]. 沈阳农业大学学报,2018,49(1):65 - 74.
ZHOU Yuncheng, XU Tongyu, DENG Hanbing, et al. Recognition method of tomato key organs based on dual convolution Fast R - CNN[J]. Journal of Shenyang Agricultural University,2018,49(1):65 - 74. (in Chinese)

[11] 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.

[12] XIE Z, LIU W, LI Y, et al. Enhanced litchi fruit detection and segmentation method integrating hyperspectral reconstruction and YOLO v8[J]. Computers and Electronics in Agriculture, 2025, 237: 110659.

[13] 曾志雄,黄迎辉,林锴,等. 基于 GSD - YOLO 的复杂场景仔猪检测和计数方法[J]. 农业机械学报,2025,56(6):247 - 257.
ZENG Zhixiong, HUANG Yinghui, LIN Kai, et al. Advanced piglet detection and counting in complex scenarios using GSD - YOLO architecture[J]. Transactions of the Chinese Society for Agricultural Machinery,2025,56(6):247 - 257. (in Chinese)

[14] 闫建伟,赵源,张乐伟,等. 改进 Faster - RCNN 自然环境下识别刺梨果实[J]. 农业工程学报,2019,35(18):143 - 150.
YAN Jianwei, ZHAO Yuan, ZHANG Lewei, et al. Recognition of *Rosa roxbunghii* in natural environment based on improved Faster - RCNN[J]. Transactions of the CSAE,2019,35(18):143 - 150. (in Chinese)

[15] 王金鹏, 何萌, 甄乾广, 等. 基于改进 COF - YOLO v8n 的油茶果静态与动态检测计数方法[J]. 农业机械学报, 2024, 55(4): 193 - 203.
WANG Jinpeng, HE Meng, ZHEN Qianguang, et al. *Camellia oleifera* fruit static and dynamic detection counting based on improved COF - YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery,2024,55(4):193 - 203. (in Chinese)

[16] 杨秋妹,陈淼彬,黄一桂,等. 基于改进 YOLO v5n 的猪只盘点算法[J]. 农业机械学报,2023,54(1):251 - 262.
YANG Qiumei, CHEN Miaobin, HUANG Yigui, et al. Pig counting algorithm based on improved YOLO v5n[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(1):251 - 262. (in Chinese)

[17] WANG Z, WALSH K, KOIRALA A. Mango fruit load estimation using a video based MangoYOLO - Kalman filter - Hungarian algorithm method[J]. Sensors, 2019, 19(12): 2742.

[18] 海兰,张丽娜,杨帆,等. 基于 YOLO v5 算法的舍饲羊只检测与跟踪方法研究[J]. 中国农机化学报,2025,46(12):70 - 76.
HAI Lan, ZHANG Li'na, YANG Fan, et al. Research on the detection and tracking method of housed sheep based on YOLO v5 algorithm[J]. Journal of Chinese Agricultural Mechanization,2025,46(12):70 - 76. (in Chinese)

[19] LUNEŽIČ A, VOJÍŘ T, ZAJC L Č, et al. Discriminative correlation filter tracner with channel and spatial reliability[J].

International Journal of Computer Vision, 2018, 126(7): 671－688.
[20] 高芳芳,武振超,索睿,等. 基于深度学习与目标跟踪的苹果检测与视频计数方法[J]. 农业工程学报,2021,37(21):217－224.
GAO Fangfang, WU Zhenchao, SUO Rui, et al. Apple detection and counting using real-time video based on deep learning and object tracking [J]. Transactions of the CSAE,2021,37(21):217－224. (in Chinese)
[21] 郭明月,刘雅晨,李伟夫,等. 基于视频跟踪算法的果园猕猴桃产量实时预估[J]. 农业机械学报,2023,54(6):178－185.
GUO Mingyue, LIU Yachen, LI Weifu, et al. Real-time production prediction of kiwifruit in orchard based on video tracking algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(6):178－185. (in Chinese)
[22] CHEN J, ZHANG M, LU B, et al. A lightweight citrus detection and counting method based on deep learning model[J]. Engineering Applications of Artificial Intelligence, 2025, 162: 112268.
[23] 傅隆生,孙世鹏,VÁZQUEZ－ARELLANO M,等. 基于果萼图像的猕猴桃果实夜间识别方法[J]. 农业工程学报,2017,33(2):199－204.
FU Longsheng, SUN Shipeng, VÁZQUEZ－ARELLANO M, et al. Kiwifruit recognition method at night based on fruit calyx image[J]. Transactions of the CSAE,2017,33(2):199－204. (in Chinese)
[24] LIN J, BAI D, XU R, et al. TSBA－YOLO: an improved tea diseases detection model based on attention mechanisms and feature fusion[J]. Forests, 2023, 14(3): 619.
[25] 李景顺, 刘美, 孟亚男, 等. 应用轻量化 FEB－YOLO 模型的荔枝果实动态识别计数方法[J]. 计算机测量与控制, 2025,33(2):229－237,261.
LI Jingshun, LIU Mei, MENG Ya'nan, et al. Dynamic recognition and counting method for litchi fruit using lightweight FEB－YOLO model[J]. Computer Measurement & Control, 2025,33(2):229－237,261. (in Chinese)
[26] 王少聪,杜肖鹏,丁小明,等. 温室巡检机器人关键技术研究进展与展望[J]. 江苏农业科学,2024,52(16):1－10.
WANG Shaocong, DU Xiaopeng, DING Xiaoming, et al. Research progress and prospect of key technologies of greenhouse inspection robot[J]. Jiangsu Agricultural Sciences,2024,52(16):1－10. (in Chinese)
[27] XU G, LIAO W, ZHANG X, et al. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation[J]. Pattern Recognition, 2023, 143: 109819.
[28] HAN D, WANG Z, XIA Z, et al. Demystify Mamba in vision: a linear attention perspective[J]. arXiv Preprint, arXiv:2405.16605,2024.
[29] ZHANG Y, SUN P, JIANG Y, et al. Bytetrack: multi-object tracking by associating every detection box[C]// European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 1－21.

(上接第 114 页)

[26] 张岳阳, 王宇珂, 吕宝轩,等. 单目视觉定位与 UWB 技术融合的育苗温室移动平台定位方法研究[J]. 西北农林科技大学学报(自然科学版), 2025, 53(5): 146－156.
ZHANG Yueyang, WANG Yuke, LÜ Baoxuan, et al. Research on the localization method of seedling greenhouse mobile platform based on fusion of visual beacons and UWB[J]. Journal of Northwest A&F University(Natural Science Edition), 2025, 53(5): 146－156. (in Chinese)
[27] LONG Z, XIANG Y, LEI X, et al. Integrated indoor positioning system of greenhouse robot based on UWB/IMU/ODOM/LiDAR[J]. Sensors,2022, 22(13): 4819.
[28] 张兆国, 朱时亮, 王法安,等. 基于改进自适应卡尔曼滤波算法的温室 UWB 定位技术[J]. 农业机械学报, 2025, 56(3): 494－502,522.
ZHANG Zhaoguo, ZHU Shiliang, WANG Faan, et al. UWB greenhouse positioning technology based on improved adaptive Kalman filter algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 494－502,522. (in Chinese)
[29] NIU Z, YANG H Z, ZHOU L, et al. Deep learning-based ranging error mitigation method for UWB localization system in greenhouse[J]. Computers and Electronics in Agriculture,2023, 205: 107573.
[30] LI M H, GAO H Y, ZHAO M X, et al. Development and experimentation of a real-time greenhouse positioning system based on IUKF－UWB[J]. Agriculture,2024, 14(9): 1479.
[31] 中华人民共和国住房和城乡建设部,中华人民共和国国家质检监督检验检疫总局. GB/T 51057—2015　种植塑料大棚工程技术规范[S]. 北京:中国计划出版社,2015.
[32] GUO H P, LI Y, WANG H, et al. Research on autonomous navigation system of greenhouse electric crawler tractor based on LiDAR[J]. Frontiers in Plant Science,2024, 15: 1377269.