

基于 YOLOv8-MT 与 RGB-D 相机的番茄检测方法

江 健¹
JIANG Jian

摘 要

针对番茄采摘机器人视觉系统在作业中面临的果实遮挡、光照变化以及在嵌入式平台上部署等具体挑战,文章开发了一种集成增强型 YOLOv8-MT 模型与 RGB-D 相机的轻量化视觉检测系统。YOLOv8-MT 模型通过引入双路径架构瓶颈注意力模块,以提升遮挡环境下的小目标番茄检测能力。采用共享卷积与自适应缩放层的轻量化检测头,在减少参数的同时增强多尺度特征表征能力。通过将 2D 检测中心与深度数据对齐,无需计算密集的点云即可实现番茄的实时三维定位。实验表明,该系统在嵌入式平台上实现 51 帧/s 推理速度的同时,检测精度达到 93.9% (提升 2.5%), 参数量减少 33.3%。最大三维定位误差分别控制在 11 ms、15 ms、14 ms 以内,满足了视觉检测系统对轻量化、实时性与准确性的综合要求,为自动化番茄采摘提供了技术支持。

关键词

番茄检测; 实时检测; 深度相机融合; 自动化番茄采摘

doi: 10.3969/j.issn.1672-9528.2026.02.022

0 引言

番茄作为全球广泛栽培的蔬菜作物^[1],在传统采收方式下面临劳动力短缺与效率低下的双重挑战。在此背景下,番茄采摘机器人应运而生,而视觉系统在其中起着关键作用。通过精准分析番茄的形态特征与三维空间信息,视觉系统为机械臂作业提供重要的决策依据^[2]。针对番茄果实生长密集、相互遮挡、质地脆弱等特点,视觉系统需具备较高的检测精度,以实现精确的成熟度判别与空间定位^[3],最终支撑不同成熟度番茄的选择性采摘与分级管理。

视觉识别系统的设计主要基于计算机视觉技术。传统方

法通常依赖轮廓、颜色和纹理等手工特征进行果实识别。Liu 等^[4]开发了一种适用于常规彩色图像的番茄自动识别算法,实现了 94% 的检测准确率。Li 等^[5]提出通过果实颜色和果实目标轮廓重建的三维模型检测方法进行果实识别。在自然环境中,光照角度的变化易导致颜色特征提取失效,检测目标的复杂背景和果实形态的多样性则使轮廓提取、几何拟合与实际目标的对应性较弱。此外,这些方法的设计过程通常较为繁琐,难以同时检测植物不同发育阶段生长特征差异显著的性状。

随着计算机视觉的快速发展,深度学习技术已成为果实识别的主流方法^[6]。相较于传统算法,深度学习方法能够从数据中自主学习和提取更多多样化的特征,展现出更优异的鲁

1. 武汉轻工大学 湖北武汉 430023

参考文献:

- [1] 邵华,李军.基于移动边缘计算的无人机应急场景中绿色通信方案研究[J].中国军转民,2025(10):53-54.
- [2] 中立军,韦智琬.基于边缘计算存储的无线电监测数据“存管管用”标准化技术研究[J].中国无线电,2025(5):75-78.
- [3] 肖琛,牟云飞.5G 网络下边缘计算任务卸载优化策略研究[J].软件,2025,46(5):175-177.
- [4] 黎志生,覃健诚,王磊.基于边缘计算和云融合构建智慧农用集群系统[J].自动化与仪表,2025,40(5):1-6.
- [5] 蔡洁聪,国旭涛,陈思艺,等.基于边缘计算和机器学习

的光伏面板热斑检测系统设计[J].工业控制计算机,2025,38(5): 65-66.

【作者简介】

王善程(1993—),男,河南信阳人,本科,助理工程师,研究方向:计算机。

陈玉林(1993—),女,河南信阳人,本科,助教,研究方向:计算机。

(收稿日期: 2025-07-11 修回日期: 2026-01-27)

棒性与准确性。目前基于深度学习的番茄检测方法正呈现多元化发展趋势,卷积神经网络^[7]在视觉系统中得到广泛应用。Hu等^[8]通过将IFS与Faster R-CNN结合,在单个成熟番茄检测中达到80%的准确率,但该方法在处理密集簇生或严重遮挡的番茄时,其性能和计算效率会受限。Liu等^[9]使用YOLO-Tomato模型检测复杂环境下的番茄果实,具有较好的检测性能,但其网络结构较为复杂。Chen等^[10]提出的YOLOv3-DPN樱桃番茄检测模型将平均检测精度提升至96.8%,但其实时性有待进一步提高。

除二维识别外,果实采收作业还需具备精确的三维空间定位能力。近年来,研究者已探索多种三维感知方案。Luo等^[11]采用双目立体视觉实现葡萄检测与采摘点定位,但地形复杂性和图像噪声会影响精度。Liu等^[12]通过多光谱视觉与三角测距法获取苹果的空间位置与成熟度信息。Tao等^[13]基于RGB-D相机采集的点云数据,开发了苹果自动识别与定位方法,但该方法高度依赖高分辨率点云数据处理,导致计算开销较大,对实现自动化采收的实时性构成显著挑战。

为开发适用于番茄采摘机器人的视觉检测系统,本研究提出了YOLOv8-MT模型,实现对不同生长阶段番茄的实时、高精度检测与定位,以支撑选择性采摘与精准管理。本文主要贡献如下:

(1) 引入双路径架构瓶颈注意力模块(BAM),通过并行的通道与空间注意力机制,在光照不足条件下实现双维度特征选择与增强,提升复杂光照下的鲁棒性。

(2) 引入共享卷积与自适应缩放层的轻量化检测头(SLDH),在保持精度的同时显著减少参数量,增强多尺度特征提取能力。

(3) 建立高效三维定位机制,通过将二维检测框中心与深度图对齐获取三维坐标,巧妙绕过了处理完整点云的计算开销,更好地满足实时性要求。

1 方法

1.1 YOLOv8-MT 网络结构

提出的YOLOv8-MT番茄检测方法整体框架如图1所示。该方法在主干网络的两个C2f模块之后及颈网络的C2f模块之前,分别嵌入BAM注意力机制。该机制通过通道与空间维度的并行处理,对模糊特征进行选择增强,精准聚焦关键特征,同时有效抑制无关信息。其次,将优化后的特征输入检测头中的SLDH模块。该模块通过共享卷积减少参数量,并利用自适应缩放层实现多尺度特征适配,有效缓解了不同检测头间的目标尺度不一致问题。

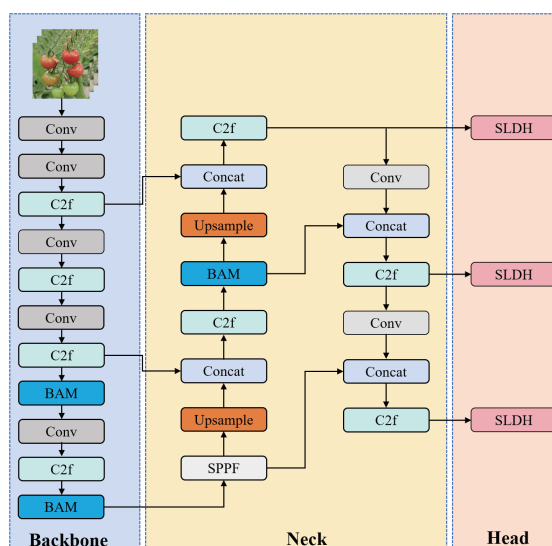


图1 YOLOv8-MT 网络结构图

1.2 BAM 注意力机制

由于小目标在特征图中占据的像素有限,网络中的逐层下采样操作会导致细节特征逐渐丢失,致使检测率明显下降。当未成熟番茄与绿叶背景具有相似颜色和纹理时,这一问题尤为突出。为提升小目标检测能力,本研究引入BAM注意力模块。其结构如图2所示,该模块通过并行的通道与空间注意力路径实现特征重加权:通道注意力机制通过分析不同通道的特征响应,增强与番茄特征相关通道的权重;空间注意力机制则通过关注特征图中的空间位置信息,强化包含小番茄的关键区域。这种双注意力机制有效提升了模型对小目标的特征提取能力,使其能够在复杂背景中准确识别未成熟的小尺寸番茄。

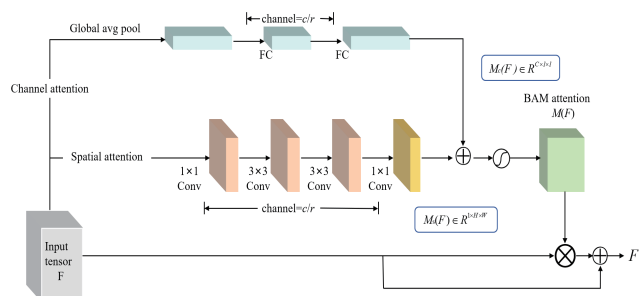


图2 BAM 模块结构

BAM 模块的计算流程用公式表示为:

$$F' = F \otimes (1 + M) \quad (1)$$

$$M(F) = M_c(F) \otimes M_s(F) \quad (2)$$

$$M_c(F) = \sigma(\text{MLP}(\text{GAP}(F))) \quad (3)$$

$$M_s(F) = \sigma(f_s(F)) \quad (4)$$

对于输入特征 F , BAM模块通过空间与通道注意力分支分别生成空间权重向量 $M_s(F)$ 和通道权重向量 $M_c(F)$ 。两个

注意力向量经整合后与原始特征相乘,得到BAM增强特征,再通过残差连接与原始输入融合,形成最终输出。在通道注意力分支中,首先采用全局平均池化使各通道获得全局感受野,接着通过多层感知器(MLP)校准通道重要性,最后进行归一化处理。在空间注意力分支中, f_s 表示一系列卷积操作:先采用 1×1 卷积降维,再通过两个 3×3 卷积层捕获空间上下文信息,最后使用 1×1 卷积恢复原始维度。

1.3 SLDH

在番茄采摘机器人视觉系统的实际应用中,嵌入式设备的计算资源限制成为关键瓶颈。YOLOv8n原有的解耦式检测头虽然通过任务分离提升了检测精度,但其双分支结构中的卷积模块会产生大量参数与计算开销,难以在硬件平台上实现高效推理。为解决此问题,本文设计了如图3所示的轻量化检测头SLDH。该模块首先对三个输入特征层分别进行复合卷积处理,其中每个复合卷积依次包含标准卷积、组归一化和激活函数。随后采用两个 3×3 复合卷积构建共享卷积层,在多个检测层之间复用。与传统的独立参数分配方式相比,这种设计显著降低了模型复杂度和过拟合风险。更重要的是,共享卷积机制能够引导模型学习具有通用性的特征模式,更好地捕捉跨尺度的共性特征,从而提升泛化性能。

为解决各检测头间的尺度不一致的问题,本研究在共享卷积基础上引入了自适应缩放层(ASL)。该模块能够依据各检测层的检测状态动态生成缩放参数,优化边界框回归结果。通过自适应调整番茄目标特征缩放比例,该层有效适应了不同检测层的特性需求,显著提升了模型对数据分布的适应能力,最终改善了整体检测性能。

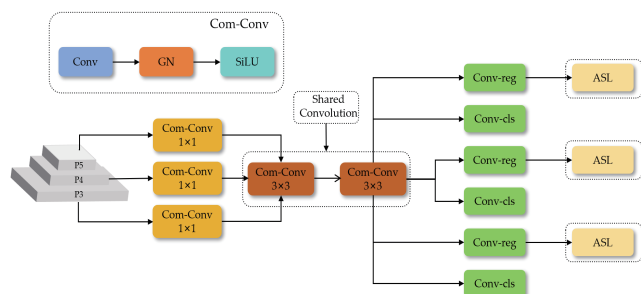


图3 SLDH 模块结构

2 结果和讨论

在这项工作中,主要实验环境为Windows11操作系统和NVIDIA RTX 4060TI显卡。各项参数设置如下:

初始学习率为0.01,动量系数为0.937,权重衰减系数为0.0005,输入图像尺寸为 $640\text{ px}\times 640\text{ px}$,批处理大小为8,训练轮次设置为200。使用的深度学习框架为PyTorch1.12.2,编程语言为Python3.10。

2.1 数据集采集和构建

数据集全部通过手动采集温室环境中的番茄植株图像构建。采集时相机与目标保持30~120 cm的距离,该范围与实际采摘机械臂的典型工作距离相匹配。为提升数据多样性,采集过程中主动调整了光照条件与拍摄角度。经筛选去除模糊及无番茄目标的图像后,共获得1850张有效原始图像,每张图像至少包含一个番茄目标。通过数据增强技术,数据集规模扩展至3800张图像。增强方法包括:缩放、平移、旋转、剪切,以及添加高斯噪声、传感器噪声、太阳耀斑模拟和雾雨效果模拟。标注的目标类别包含成熟番茄与未成熟番茄两类,对应标签颜色分别为红色与绿色。番茄果实的成熟度判定标准为:当果实红色部分占比超过70%时定义为成熟,低于70%则定义为未成熟,如图4所示。数据集中2600张图像作为训练集,其余1200张作为测试集。

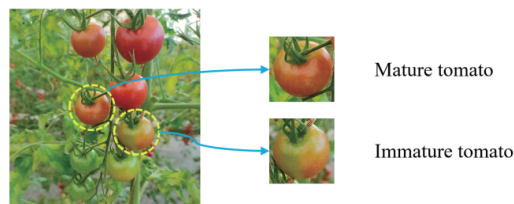


图4 成熟番茄和未成熟番茄

2.2 评估指标

本研究采用均值平均精度(mAP)、精确率(P)、召回率(R)以及参数量(Params)作为模型性能的评估指标。具体计算方法为:

$$P = \frac{TP}{TP+FP} \quad (5)$$

$$R = \frac{TP}{TP+FN} \quad (6)$$

$$mAP = \frac{\sum AP}{N} \quad (7)$$

式中:TP表示真阳性样本数;FP表示假阳性样本数;FN表示假阴性样本数;P表示精确率;R表示召回率;AP表示P-R曲线下面积;mAP表示所有类别AP的平均值;N表示类别总数。

2.3 消融实验

为验证YOLOv8-MT中所设计模块对模型性能的贡献,本研究对各个模块进行消融实验。结果如表1所示,表中“√”代表采用对应模块。由表1可知,采用共享卷积机制的SLDH轻量检测头使模型参数量降低43.3%,mAP仅下降0.1%,显著优化了模型复杂度与计算成本。引入BAM注意力模块后mAP提升了2.7%,其通道注意力增强了对番茄颜色纹理的区分力,空间注意力则有效抑制了枝叶等背景干扰。

所有改进模块集成后，模型在番茄检测任务中达到最优精度，验证了多模块的协同增益效应。

表 1 YOLOv8-MT 中各模块的消融实验结果

Base	SLDH	BAM	mAP/%	Params/ 10^6
√			91.4	3.0
√	√		91.3	1.7
√		√	94.1	3.1
√	√	√	93.9	2.0

2.4 网络对比实验

为验证 YOLOv8-MT 模型的先进性，本研究在相同数据集和 Jetson Orin NX 嵌入式平台上进行了对比实验。对比实验中选取了具有代表性的主流目标检测模型进行对比。完整的实验结果汇总于表 2 中，其中最优性能指标已用加粗字体标出。

表 2 YOLOv8-MT 与其他网络模型的性能对比实验结果

Model	mAP /%	Params / 10^6	帧 /s	
			RTX 4060Ti	Jetson Orin NX
Faster R-CNN	75.3	116.5	25	23
RT-DETR	74.8	32.8	40	36
YOLOv5	91.3	1.7	53	47
YOLOv6n	91.4	4.6	43	39
YOLOv7	93.5	6.0	37	34
YOLOv8n	91.7	3.0	51	46
YOLOv9-tiny	91.3	2.0	41	37
YOLOv10n	91.0	2.3	45	42
YOLOv11n	93.2	2.6	48	44
YOLOv8-MT(Ours)	93.9	2.0	56	51

YOLOv8-MT 以 93.9% 的 mAP 达到最优检测精度，同时在 Jetson Orin NX 上实现了 51 帧/s 的推理速度。相比 YOLOv7，在显著降低参数量的同时获得更高精度，更适应嵌入式平台的部署约束。表明 YOLOv8-MT 在精度、效率与轻量化方面取得最佳平衡，其目标检测能力与优化的模型结构，有效满足采摘机器人在复杂环境下对视觉感知系统高实时性、高精度的作业要求。

2.5 实验结果与对比分析

为直观展示 YOLOv8-MT 模型在番茄检测任务中的性能优势，图 5 为原始模型 YOLOv8n 与改进模型 YOLOv8-MT 在同一温室环境图像中可视化检测结果的对比，箭头标注了关键差异区域。对比结果清晰显示，当原始模型因严重遮挡而漏检小尺寸番茄时（如箭头所示），YOLOv8-MT 仍能保持准确检测能力，成功检测出这些被忽略的目标。

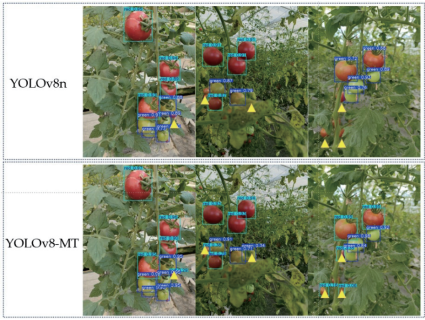


图 5 成熟与未成熟番茄检测结果对比图

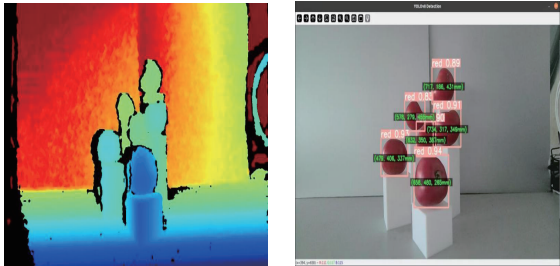
2.6 基于 RGB-D 视觉的番茄三维定位

为实现番茄抓取精准化，视觉系统需同步完成目标检测与三维定位。本研究基于改进 YOLOv8-MT 模型获取番茄二维边界框后，采用 Orbbec Gemini 335 深度相机（如图 6 所示）进行三维空间定位：以检测框中心像素坐标 (X, Y) 为基准，从对齐深度图中提取对应深度值 Z ，并通过 5×5 区域中值滤波抑制噪声，最终融合生成相机坐标系下的三维坐标 (X, Y, Z) 。该坐标可直接用于机械臂抓取路径规划，整套方案在确保定位精度的同时满足实际采收作业的实时性要求。



图 6 Orbbec Gemini 335 深度相机

为系统评估番茄三维坐标的获取精度，本研究基于 Orbbec Gemini 335 深度相机，结合番茄固定装置与激光测距仪，构建了坐标测量实验平台。为排除复杂田间环境的干扰因素，实验在受控室内环境中进行，以保障测量数据的稳定性。图 7 展示了三维定位全流程的关键可视化结果。为进一步量化定位性能，选取单颗番茄作为测试对象，以激光测距仪获取的空间距离作为基准真值，与视觉检测系统输出的三维坐标进行系统性比对与误差分析。详细的误差统计结果汇总于表 3 中。



(a) 番茄配准深度图 (b) 番茄三维定位结果可视化

图 7 番茄三维定位过程示意图

表3 番茄三维坐标检测结果

Group	Ground Truth (X,Y,Z)/mm	Estimated Value (X,Y,Z)/mm	Error/mm
1	(511, 449, 327)	(504, 444, 319)	(7, 5, 8)
2	(601, 328, 450)	(597, 321, 437)	(4, 7, 13)
3	(428, 317, 523)	(425, 307, 516)	(3, 10, 7)
4	(735, 484, 638)	(744, 471, 634)	(9, 13, 4)
5	(456, 274, 753)	(461, 286, 760)	(5, 12, 7)
6	(947, 746, 857)	(955, 757, 866)	(8, 11, 9)
7	(749, 486, 848)	(736, 481, 840)	(13, 5, 8)
8	(943, 730, 872)	(954, 721, 886)	(11, 9, 14)
9	(830, 245, 941)	(823, 230, 928)	(7, 15, 13)
10	(694, 287, 1064)	(705, 295, 1068)	(11, 8, 14)
11	(596, 847, 1183)	(606, 854, 1195)	(10, 7, 12)
12	(693, 751, 1346)	(684, 760, 1332)	(8, 9, 14)

根据表3的测量结果可知,番茄三维坐标在X、Y、Z轴方向的最大定位误差分别为11.0 mm、15.0 mm和14.0 mm,平均误差分别为8.0 mm、9.3 mm和10.3 mm。所有误差值均在系统允许容差范围内,满足实际采摘作业的精度要求。

3 结论

本研究针对番茄采摘面临的目标遮挡、光照变化及果簇干扰等问题,提出了一个基于YOLOv8-MT与RGB-D相结合的番茄检测视觉系统,并构建了一个番茄数据集。在YOLOv8-MT模型中,通过引入BAM模块并行增强通道与空间特征,强化对番茄色彩特性的响应。引入SLDH模块,通过共享卷积与自适应缩放,在保持多尺度检测能力的同时降低参数量与计算复杂度。同时,系统搭载了Orbbec Gemini 335深度相机,实现了番茄实时检测与毫米级三维定位。实验结果表明,该方法在精度与效率间取得了良好平衡,为采摘机器人提供了可靠的视觉感知支持。

参考文献:

- [1]GAO F C, LI H J, MU X G, et al. Effects of organic fertilizer application on tomato yield and quality: a meta-analysis[J]. Applied sciences, 2023, 13(4):2184.
- [2]JI W T, HUANG X H, HE X K, et al. A comprehensive review of the research of the "Eye-Brain-Hand" harvesting system in smart agriculture[J]. Agronomy, 2023, 13(9):2237.
- [3]CHENG S S, YU Z W, LI Z, et al. Design and analysis of a lightweight redundant-degree-of-freedom fruit-picking robot arm[J]. Journal of field robotics, 2025, 42(6):2815-2825.
- [4]LIU G X, MAO S Y, KIM J H. A mature-tomato detection algorithm using machine learning and color analysis[J]. Sensors, 2019, 19(9):2023.

- [5]LI L, OU-YANG Z K, TIAN W Q, et al. Location and model reconstruction algorithm for overlapped and sheltered spherical fruits based on geometry[J/OL]. International journal of advanced robotic systems,2022[2025-12-23].https://www.researchgate.net/publication/357904480_Location_and_model_reconstruction_algorithm_for_overlapped_and_sheltered_spherical_fruits_based_on_geometry.
- [6]MELNYCHENKO O, SCISLO L, SAVENKO O, et al. Intelligent integrated system for fruit detection using Multi-UAV imaging and deep learning[J]. Sensors, 2024, 24(6):1913.
- [7]QIU D K, GUO T H, YU S Q, et al. Classification of apple color and deformity using machine vision combined with CNN[J]. Agriculture, 2024, 14(7):978.
- [8]HU C H, LIU X, PAN Z, et al. Automatic detection of single ripe tomato on plant combining Faster R-CNN and intuitionistic fuzzy set[J]. IEEE access,2019, 7:154683-154696.
- [9]LIU G X, NOUAZE J C, MBOUEMBE P L T, et al. YOLO-Tomato: a robust algorithm for tomato detection based on YOLOv3[J]. Sensors, 2020, 20(7):2145.
- [10]CHEN J Q, WANG Z K, WU J H, et al. An improved YOLOv3 based on dual path network for cherry tomatoes detection[J]. Journal of food process engineering, 2021, 44(10):e13803.
- [11]LUO L F, TANG Y C, ZOU X J, et al. Vision-based extraction of spatial information in grape clusters for harvesting robots[J]. Biosystems engineering, 2016, 151:90-104.
- [12]LIU Z X, LIU G. Apple maturity discrimination and positioning system in an apple harvesting robot[J]. New zealand journal of agricultural research,2007, 50(5):1103-1113.
- [13]TAO Y T, ZHOU J. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking[J]. Computers and electronics in agriculture, 2017, 142:388-396.

【作者简介】

江健(2000—),男,湖北大冶人,硕士研究生,研究方向:深度学习、图像处理。

(收稿日期:2025-12-23 修回日期:2026-02-06)