doi:10.6041/j.issn.1000-1298.2026.05.015

基于改进 YOLO 11n 的密集遮挡环境百香果识别方法

汤　晨[1]　刘振青[1]　邵　阳[1]　乐　凯[1]　高　玲[2]　顾　晔[3]　宋　鹏[1,4]

(1. 华中农业大学植物科学技术学院, 武汉 430070; 2. 中国热带农业科学院热带作物品种资源研究所, 海口 571737;
3. 海南丰歌实业开发有限公司, 乐东 572500; 4. 作物遗传改良全国重点实验室, 武汉 430070)

摘要: 针对果园密集遮挡环境下百香果自动化采摘任务中目标检测精度与模型轻量化难以兼顾的问题,本文提出一种基于改进 YOLO 11n 的百香果视觉检测模型。在特征提取阶段及检测头中引入部分卷积(Partial convolution, PConv)替换标准卷积块进行轻量化改进;设计融合特征图切片机制的改进 SimAMs 3D 注意力模块(Sliced 3D spatial and channel attention module, SimAMs),强化跨通道-空间域特征融合表达。实验结果显示,改进模型精确率和平均精度均值(mAP@0.5)分别为 93.32%、93.08%,较原始模型分别提升 1.28、0.26 个百分点,其参数量、计算量和内存占用量分别降低 21.2%、23.8% 和 20.0%,且检测帧率(GPU 和 CPU)分别为 EfficientNetV2、FasterNet 等对比实验模型平均值的 1.36、1.68 倍。实验表明,在深度距离 400 ~ 500 mm 下单果采摘时间用时 13 s,成功率达 91.7%,较近、远端深度距离高 11.1、2.8 个百分点。本研究为果园密集遮挡环境下百香果高精度实时检测与机械臂采摘作业提供了有效技术支撑。

关键词: 百香果; YOLO 11n; 轻量化; 3D 注意力模块; 机械臂采摘; 密集遮挡环境

中图分类号: S126　　**文献标识码**: A　　**文章编号**: 1000-1298(2026)05-0167-10　　**OSID**:

Passion Fruit Recognition Method in Densely Occluded Environments Based on Improved YOLO 11n

TANG Chen[1]　LIU Zhenqing[1]　SHAO Yang[1]　YUE Kai[1]　GAO Ling[2]　GU Ye[3]　SONG Peng[1,4]

(1. *College of Plant Science and Technology, Huazhong Agricultural University, Wuhan* 430070, *China*
2. *Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou* 571737, *China*
3. *Hainan Fengge Industrial Development Co., Ltd., Ledong* 572500, *China*
4. *National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan* 430070, *China*)

Abstract: Aiming to address the challenge of balancing detection accuracy and model lightweightness in automated passion fruit harvesting under densely occluded orchard environments, an improved YOLO 11n-based visual detection model was proposed. Partial convolution (PConv) was introduced in both the feature extraction stage and detection head to replace standard convolution blocks for lightweight optimization. A sliced 3D spatial and channel attention module (SimAMs), integrating a feature map slicing mechanism, was designed to enhance cross-channel and spatial feature fusion. Experimental results showed that the improved model achieved a precision of 93.32% and mAP@0.5 of 93.08%, with increases of 1.28 percentage points and 0.26 percentage points over the original model. The parameter count, computation, and memory usage were reduced by 21.2%, 23.8%, and 20.0%, respectively. Detection speeds on GPU and CPU were approximately 1.36 times and 1.68 times faster than the average speed of the compared models, including EfficientNetV2 and FasterNet. Harvesting tests showed a single-fruit picking time of 13 s and a success rate of 91.7% at a depth range of 400 ~ 500 mm, 11.1 percentage points and 2.8 percentage points higher than that at near and far distances. The research result can provide effective technical support for high-precision, real-time detection and robotic harvesting of passion fruit in complex orchard environments with dense occlusion.

Key words: passion fruit; YOLO 11n; lightweight; 3D attention module; robotic arm harvesting; densely occluded environments

收稿日期: 2025-04-29　修回日期: 2025-06-12

基金项目: 湖北省重点研发计划项目(2024BBB053)

作者简介: 汤晨(2003—),男,硕士生,主要从事农业机器人技术研究,E-mail: tcyoung@webmail.hzau.edu.cn

通信作者: 宋鹏(1986—),男,副教授,博士,主要从事作物表型机器人技术研究,E-mail: songp@mail.hzau.edu.cn

0 引言

百香果作为热带特色经济作物,广泛种植于湿润、光照充足地区,其果实多生长在藤蔓缠绕、枝叶交错的冠层区域[1],易被叶片和茎秆遮挡,增加了采收作业的复杂性。随着农业自动化技术快速发展,果园智能巡检与采摘机器人等装备逐步取代传统人工劳作[2-3],采摘机械臂通常需要在复杂藤蔓网络中快速识别成熟果实并规划采摘路径,而密集遮挡、果实堆叠以及果皮与叶片颜色相近等因素使传统检测方法易出现目标丢失或定位偏差[4-5],制约着自动化设备的作业效率与可靠性。

传统机器学习、数字图像处理等方法主要通过特征工程和分类算法来实现目标检测[6]。例如,LINKER 等[7]将区域生长算法生成的种子区域结合轮廓分割提取特征弧段后识别苹果,检测率达到 85%,但因光照、色彩等变化出现大量误检。LU 等[8]提出了一种基于颜色信息和轮廓片段的算法,通过色差分析和归一化 RGB 模型进行初始分割,结合轮廓片段特征选择和椭圆拟合技术,实现在自然光照条件下检测柑橘类水果并恢复被遮挡的果实,相对误差为 5.27%。这类方法的特征工程算法相对复杂,在面对光照变化、枝叶遮挡等场景时会出现特征提取适应性差、鲁棒性和效率低等问题。而基于深度学习的目标检测方法采用卷积神经网络经过监督学习和损失函数优化,实现自适应提取深度特征[6,9],以 Faster R-CNN[10]、SSD[11]、YOLO[12]为代表的深度学习目标检测模型显著提升了复杂场景下的检测性能。例如,SA 等[13]提出了一种 DeepFruits 的改进 Faster R-CNN 模型,融合 RGB 和近红外 NIR 多模态数据来提升苹果、芒果等多类水果的检测。TU 等[14]结合 RGB-D 图像和浅层特征提取出 MS-FRCNN 网络,以多尺度特征提取器获取浅层卷积特征图,增强了复杂环境下百香果的检测能力。TIAN 等[15]提出一种结合 DenseNet 的改进 YOLO v3 模型,提高了低分辨率特征层复用能力,用于复杂环境中不同生长阶段苹果的检测。在上述研究中改进模型虽在检测精度上有一定提升,但带来了参数量增多、网络结构复杂化等问题,难以在硬件资源有限的情况下部署和应用。为提升检测性能、轻量化目标检测模型,现有研究[16-17]主要通过模块轻量化设计、通道剪枝、知识蒸馏等方法来降低模型计算开销,但也降低了模型在复杂场景下对目标细节特征的捕捉能力而导致检测精度下降。

现有轻量化检测方法在果园中密集遮挡、光照变化等复杂场景下,难以兼顾模型检测鲁棒性和计算效率。本文基于改进 YOLO 11n 模型,提升密集遮挡场景下百香果检测性能;并搭建六自由度机械臂抓取平台,开展百香果采摘实验,以期为果园自动化作业提供高效可靠的视觉解决方案。

1 复杂果园环境下百香果数据集构建

百香果图像拍摄于 2024 年 12 月,地点为中国热带农业科学院品种资源研究所百香果标准化种植示范基地(海南省乐东黎族自治县丰歌百香果基地);采集时百香果处于果实膨大期至成熟前期的过渡阶段,果皮色泽与功能叶片呈现色度特征趋同现象。如图 1 所示,果园采用标准化藤架栽培模式,由若干种植区组成,单区包含两组平行排列的栽培行,每组栽培行由间距 0.5 m 的独立植株组成,各植株主蔓沿钢丝网格呈“T”形定向牵引,形成高度约 2.2 m 的三维栽培结构。

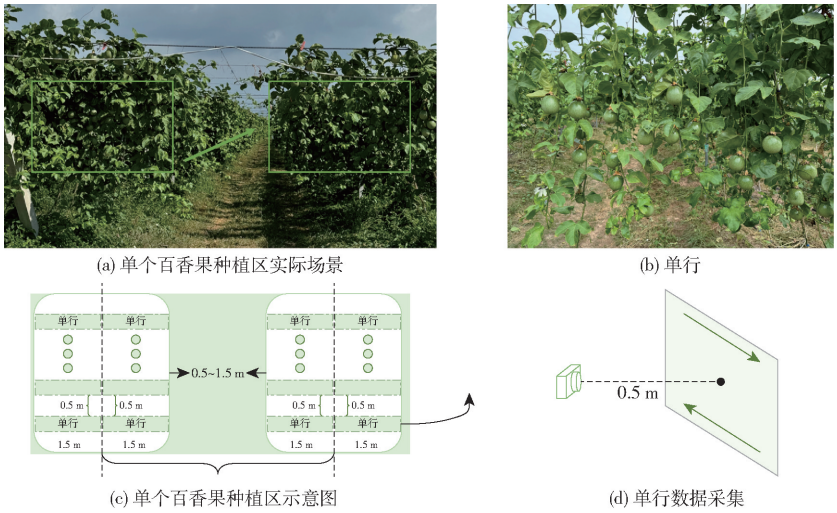


图 1 果园种植环境与数据采集

Fig. 1 Orchard planting environment and data collection

使用 Intel RealSense D435i 型深度相机采集图像，分辨率为 1 280 像素 ×720 像素，保存为 PNG 格式。相机固定在距离栽培行 0.5 m 的位置沿植株行方向前后移动拍摄，模拟实际采摘视角变化。为增强数据多样性，于 2024 年 12 月 17—21 日不同时段下采集图像，覆盖 200 多个单行植株区，在数据筛选与质量评估后共获取 422 幅图像。图 2 展示了侧光、顺光、逆光、炫光等环境下的百香果图像，可见百香果主蔓枝叶茂密、果实遮挡严重。

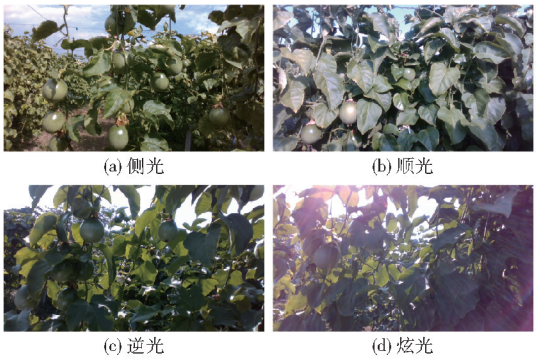


图 2　不同环境下百香果图像

Fig. 2　Passion fruit images under different environments

实验使用 LabelImg 工具对 422 幅图像进行标注，矩形框类别均为“passion_fruit”，共计 3 158 个目标对象。数据集以比例 8∶1∶1 划分为训练集、验证集和测试集。为增强模型对光照变化、尺度差异及遮挡等情况的鲁棒性，在训练阶段引入 HSV 调整、几何变换、Mosaic 拼接与随机擦除等多维数据增强策略，如图 3 所示。

2　改进 YOLO 11n 百香果检测模型

YOLO（You only look once）系列模型[12]作为目标检测领域代表性算法，将定位与分类整合为端到端回归任务，实现高效检测。YOLO 11 系列[18]在“Backbone - Neck - Head”架构基础上，引入 C3k2 模块，利用跨阶段部分连接结构将传统大卷积核分解为 2 ×2 小核级联，结合残差跳跃连接强化梯度传播，优化特征提取效率，有效降低浮点运算量；新增并行空间注意力模块（C2PSA）融合通道分组与空间注意力机制，采用双分支结构分别执行特征重标定和区域聚焦、自适应权重融合强化关键区域感知能力，提升复杂场景下的小目标检测精度；并在 Head 层关键路径部署双 DWConv 结构，结合动态核参数调整策略，减少冗余计算，提升推理速度。相较 YOLO v8，YOLO 11 在参数量更少的前提下，提高了 mAP，加快了推理速度。

YOLO 11 系列在通用目标检测任务中表现优异，但在百香果园等密集遮挡场景中仍存局限：现有注意力机制无法准确捕捉果实重叠及枝叶遮挡环境下的目标特征，易造成特征混淆而出现目标粘连和误检问题；且多尺度特征融合机制在应对光照变化与枝叶晃动等干扰时难以平衡检测精度与实时性。因此，本研究以 YOLO 11n 作为基准模型，优化网络结构并引入改进 3D 注意力机制，提升模型在果园密集遮挡场景下的检测精度与稳定性。

2.1　PConv 模块

部分卷积[19]（Partial convolution，PConv）选取输入通道前 1/4 进行常规卷积操作，其余通道保持不变传递，浮点运算量显著降低至标准卷积的1/16，内存访问量减少为 1/4，显著降低内存 I/O 负担。在图 4 所示的各类轻量化卷积模块设计中，深度可分离卷积（Depthwise separable convolution，DWConv）将空间卷积与通道卷积解耦来降低计算量，但非连续的内存访问模式易引发 GPU 缓存命中率下降和跨通道信息交互不足的缺陷；分组卷积（Grouped convolution，GConv）对输入通道分组处理压缩参数规模，但各组独立计算导致跨组语义关联断裂，缺乏跨通道特征融合的完整性；Ghost 卷积（Ghost convolution，GhostConv）通过廉价线性变换生成冗余特征图减少计算量，其线性操作固有局限性带来的低质量伪影干扰会降低特征判别力；而部分卷积仅对输入通道的部分子集执行高密度空间卷积，保留了完整通道信息流拓扑结构，优化连续内存

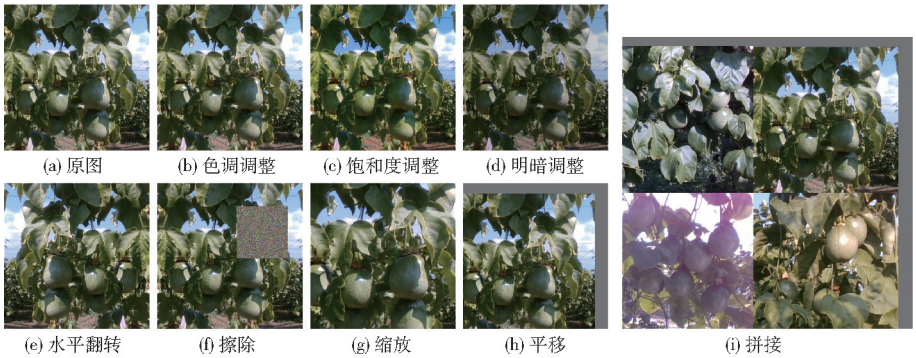


图 3　数据增强示例

Fig. 3　Data augmentation examples

访问和规则计算模式,更加适配现代 GPU 并行计算架构,在实际硬件部署中具备更高的计算效率。

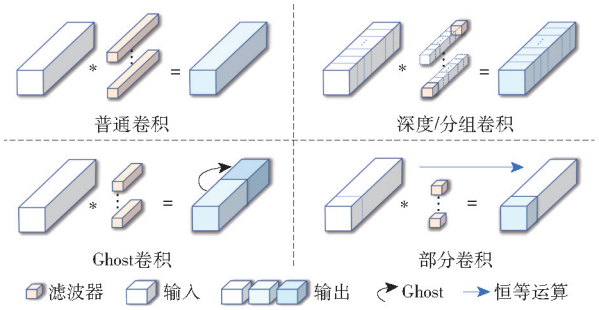


图 4 PConv 卷积原理图

Fig. 4 Principle diagram of partial convolution

2.2 改进 Sliced SimAM 3D 注意力机制

在传统注意力机制中,通道注意力(Channel-wise attention,例如 SE[20]、ECA[21]、CBAM[22])采用全局池化,即以全连接或卷积操作生成各通道的权重因子,组成 1D 权重向量对通道维度进行动态标定,强化特征通道之间的语义关联和信息传递;空间注意力(Spatial-wise attention,例如 SA[23]、CBAM)使用 2D 权重矩阵对空间位置选择性增强,突出重要区域并抑制背景干扰,实现局部结构信息提取[24];但 2 种注意力方式局限于单一维度的重标定,不能兼顾局部特征响应和跨通道语义关联,难以实现通道-空间联合特征优化。而图 5 所示的 SimAM[25] 3D 注意力机制(Simple, parameter-free attention module, SimAM)创新性地提出基于神经科学"空间抑制"理论的能量函数,直接评估并量化每个神经元与其周围神经元之间的线性可分性,无需任何学习参数的条件下生成 3D 注意力权重强化关键特征,同步抑制背景噪声,有助于增强复杂场景下目标区域辨识度;SimAM 核心闭式解析特征统计量有效规避了传统注意力机制中复杂的权重扩展与参数学习过程,具有无参数和轻量化的特性。

在果园密集遮挡且果皮色度与叶片高度趋同的复杂场景下,SimAM 对特征图的全局能量计算会因高比例相似背景(叶片)而均质化、平滑果实关键区域注意力分布,难以捕捉果柄、斑纹等细微纹理特征,影响被遮挡百香果和边缘信息响应。因此,实验提出具有对特征图切片能力的 SimAMs 3D 注意力机制(Sliced 3D spatial and channel attention module, SimAMs),如图 6 所示,SimAMs 对输入特征图沿宽高(W、H)维度进行切分,形成宽为 $W/2$,高为 $H/2$ 的 4 个空间子区域,每个切片后的子区域独立计算区域内神经元均值与方差,通过闭式解生成局部敏感的 3D 注意力权重后拼接重构为完整特征图。相较于 SimAM 注意力机制,SimAMs 分块统计有效规避叶片色度相似、大面积遮挡对果实关键区域均质化的干扰,增强局部聚焦、提升细粒度纹理感知能力,保持无参计算的轻量化特性,适应实时检测需求。

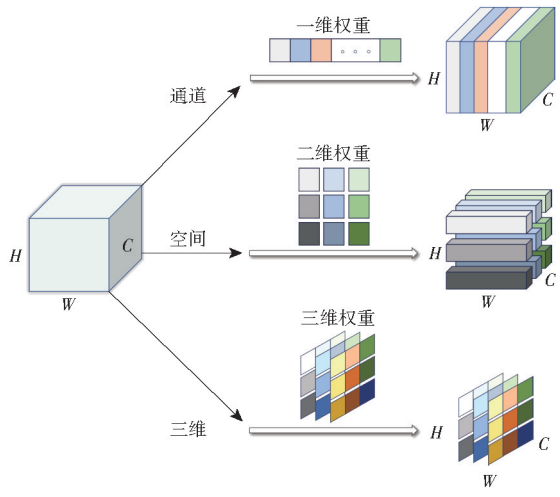


图 5 SimAM 3D 注意力模块原理图

Fig. 5 Principle diagram of SimAM 3D attention module

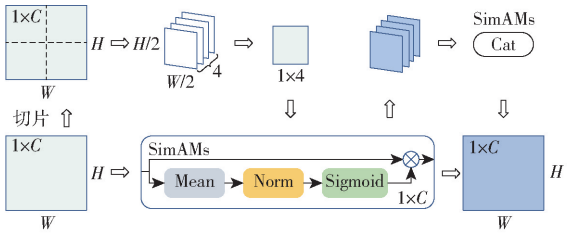


图 6 SimAMs 3D 注意力模块原理图

Fig. 6 Principle diagram of SimAMs 3D attention module

2.3 改进 YOLO 11n 轻量化网络

果园中百香果在藤蔓纵横、枝叶交织的环境中生长,密集遮挡及背景色彩的高度相似性易导致漏检、误检等问题;而智能采摘、自动巡检等果园作业对检测模型的实时性、精确度提出更高要求。因此,本研究基于 YOLO 11n 融合轻量化改进并优化特征提取、注意力建模与检测头回归计算以提升模型性能,如图 7 所示。

针对骨干(Backbone)、颈部(Neck)网络中的特征提取模块 C3k2,将用于串联多个卷积层实现特征复用与感受野扩展的 Bottleneck 模块轻量化,即使用 PConv 替换标准卷积得到 Bottleneck - PConv,实现 C3k - PConv、C3k2 - PConv 模块改进;优化模块以选择性处理高激活区域方法减少冗余特征提取过程,保留果柄、斑纹等细粒度重要特征,提升计算效率并进一步降低计算量,如图 8 所示。同时,将检测头 Detect 中用于 CIoU 边界框回归计算的标准卷积替换为部分卷积得到 Detect - PConv,加快预测推理速度。在 C2PSA 模块的 PSABlock 结构中,将原始多头注意力机制替换为改进 SimAMs 3D 注意力机

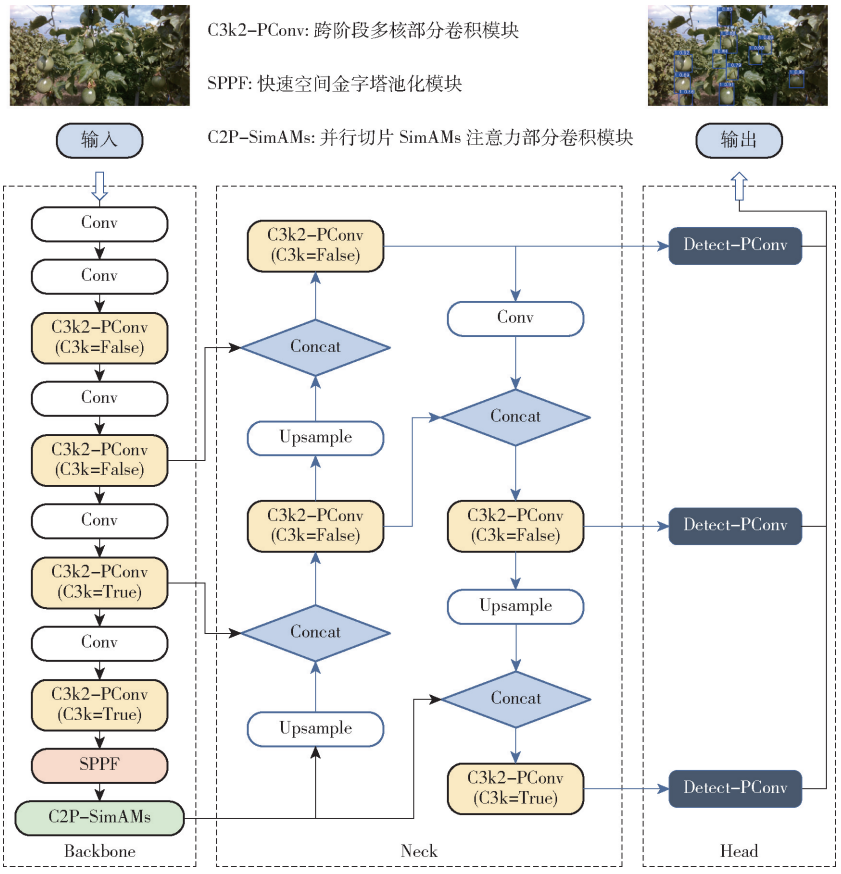


图 7　改进 YOLO 11n 结构图

Fig. 7　Architecture diagram of improved YOLO 11n

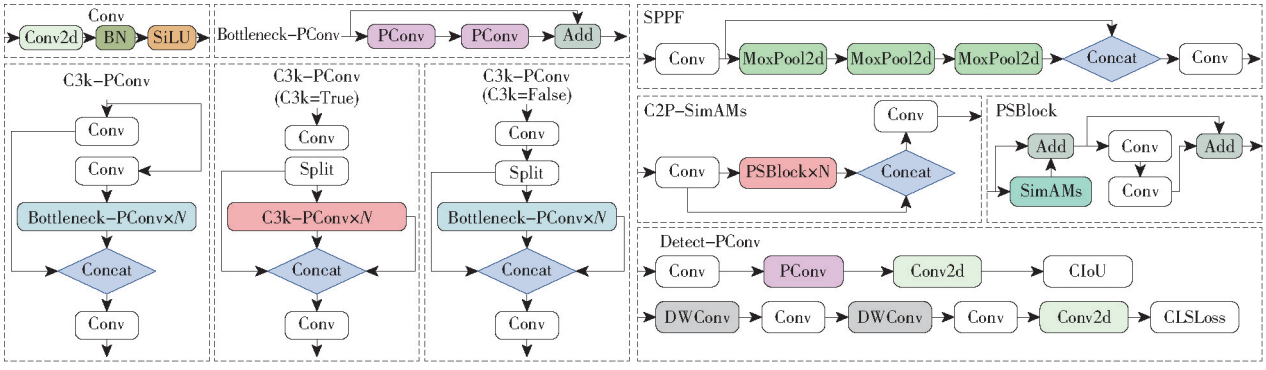


图 8　改进 YOLO 11n 中优化模块结构图

Fig. 8　Structure diagram of optimized modules in improved YOLO 11n

制，优化得到 PSBlock 和 C2P－SimAMs 模块。

3　改进模型消融与对比实验

3.1　实验环境与参数配置

实验配置环境如下：Windows 10；CPU 为 13th Gen Intel(R) Core(TM) i7－13700KF @ 3.40 GHz；GPU 为 NVIDIA GeForce RTX4090，显存为 24 GB；运行内存为 32 GB；Python 3.10；PyTorch 2.0.0；适配对应的 CUDA 11.8 和 CUDNN v8.7.0。

模型训练中，图像尺寸为 1 280 像素 ×1 280 像素，训练次数为 500，批量大小为 32，启用自动优化器选择及数据增强。为减少因数据集加载顺序、随机初始化等因素对验证指标的影响，在每个模型训练时设置随机种子(0、1、2、3、4)共完成 5 次训练，最终评估结果以均值 ± 标准差的形式统计，更客观地反映各个模型实际性能和稳定性。

3.2　评价指标

采用精确率(Precision)、召回率(Recall)、平均精度均值(mAP@0.5、mAP@0.5:0.95)等指标来评估模型检测性能，其中 mAP@0.5 和 mAP@0.5:0.95 分别表示模型在 IoU 阈值为 0.5 及多个 IoU 阈值区间内的平均精度均值；衡量模型计算开销与实时推理能力采用参数量、浮点运算量、模型内存占用

量和帧率等指标。

3.3 消融实验

在 YOLO 11n 基准模型上逐步引入改进模块开展消融实验,验证轻量化设计中基于部分卷积的特征提取模块(C3k2 - PConv)和高效检测头(Detect - PConv)对推理速度的提升效果,并评估改进无参数 SimAMs 3D 注意力模块(C2P - SimAMs)对遮挡目标的检测准确性,结果如表 1 所示。

表 1 改进模型消融实验结果

Tab. 1 Ablation experiment results of improved model

PConv	SimAMs	精确率/%	召回率/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量	浮点运算量	内存占用量/MB	GPU 帧率/(f·s^{-1})	CPU 帧率/(f·s^{-1})
×	×	92.04 ± 2.49	85.78 ± 2.12	92.82 ± 0.95	77.38 ± 0.81	2.58×10^6	6.3×10^9	5.30	155.7	4.8
√	×	93.68 ± 1.41	85.22 ± 0.98	92.92 ± 0.51	77.24 ± 0.63	2.08×10^6	4.9×10^9	4.33	160.7	5.3
×	√	91.52 ± 0.54	87.50 ± 0.72	93.34 ± 0.42	77.52 ± 0.53	2.53×10^6	6.3×10^9	5.20	156.7	5.1
√	√	93.32 ± 3.02	86.04 ± 2.12	93.08 ± 1.16	77.78 ± 0.64	2.03×10^6	4.8×10^9	4.24	158.5	5.8

注:√ 表示采用此模块,× 表示未采用此模块。

由表 1 可知,引入 PConv 卷积块后,模型参数量、浮点运算量和内存占用量分别下降 19.3%、22.2% 和 0.97 MB,在 GPU 和 CPU 上对分辨率为 1 280 像素 × 1 280 像素图像检测帧率分别提升 5、0.5 f/s;召回率和 mAP 基本保持稳定,精确率提升 1.64 个百分点,证明该模块有效减少冗余卷积计算并保留更多特征,提升推理速度并维持检测精度。将 C2PSA 模块中原始注意力替换为改进 SimAMs 3D 模块后,参数量小幅降低 1.9%、模型内存占用量减少 0.1 MB,推理速度略微提升,说明 SimAMs 的特征图分块处理方式不会显著增加计算开销,且无参数设计保证了模型轻量化;召回率提升 1.72 个百分点,验证其在遮挡和边缘目标特征的聚焦能力更强,降低了密集遮挡下的漏检率,但放大了部分低置信度背景特征,导致精确率下降 0.52 个百分点,但 mAP@0.5 和 mAP@0.5:0.95 分别提升 0.52、0.14 个百分点,且多次评估结果标准差大幅减小表明其鲁棒性更强。

PConv 卷积块和 SimAMs 3D 注意力模块组合优化模型后,PConv 部分卷积策略中和了 SimAMs 对输入特征图切分与权重融合局部关联性计算的增量开销,参数量和浮点运算量显著下降 21.2% 和 23.8%,模型内存占用量减少 20.0%,在 CPU 上推理速度提升 1 f/s;且在密集遮挡场景下,PConv 有效抑制背景区域噪声干扰并构建高信噪比特征基底,提高 SimAMs 在特征图分块中遮挡区域特征响应,增强了模型对三维特征图全局上下文与局部遮挡区域的协同表征能力。

3.4 对比实验

为衡量改进模型在检测任务中的精度、浮点运算量和推理速度等,实验选取主流轻量级网络(EfficientNetV2 - S[26]、FasterNet_T0[19]、GhostNetV2 × 0.5[27]、MobileNetV3 - Small × 0.5[28] 和 ShuffleNetV2 × 1.0[29])对应的同量级计算复杂度版本作为 YOLO 11n 主干网络,在实验环境与参数配置不变的情况下进行对比,结果如表 2 所示。

表 2 不同轻量化模型对比实验结果

Tab. 2 Comparative experiment results of different lightweight models

模型	精确率/%	召回率/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量	浮点运算量	内存占用量/MB	GPU 帧率/(f·s^{-1})	CPU 帧率/(f·s^{-1})
EfficientNetV2 - S	94.44 ± 1.78	84.10 ± 1.95	92.78 ± 0.30	77.64 ± 0.59	2.09×10^6	5.2×10^9	4.47	112.1	3.3
FasterNet_T0	92.28 ± 1.65	83.86 ± 2.28	92.00 ± 0.76	75.96 ± 0.84	3.90×10^6	9.2×10^9	7.82	140.4	3.6
GhostNetV2 × 0.5	93.26 ± 2.40	83.88 ± 1.28	92.84 ± 1.45	75.10 ± 1.81	2.78×10^6	4.6×10^9	5.95	78.5	3.0
MobileNetV3 - Small × 0.5	92.76 ± 1.96	83.40 ± 1.78	91.72 ± 1.49	75.52 ± 1.49	2.35×10^6	4.1×10^9	4.88	124.3	4.2
ShuffleNetV2 × 1.0	93.30 ± 1.98	84.88 ± 1.50	92.38 ± 0.67	76.54 ± 0.77	2.52×10^6	5.9×10^9	5.25	127.6	3.2
本文	93.32 ± 3.02	86.04 ± 2.12	93.08 ± 1.16	77.78 ± 0.64	2.03×10^6	4.8×10^9	4.24	158.5	5.8

本文模型浮点运算量略高于 MobileNetV3 和 GhostNetV2,但参数量、浮点运算量和模型内存占用量相较于 FasterNet 分别减少 47.9%、47.8% 和 45.8%,实现了计算资源高效节省;且模型推理速度 GPU 较 FasterNet 和 GhostNetV2 分别提升 12.9% 和 101.9%,CPU 较 MobileNetV3 和 GhostNetV2 分别提升 38.1% 和 93.3%,适用于实时检测场景。改进模型在保持参数量、模型内存占用量和推理速度最优的同时,精度仅次于 EfficientNetV2,但召回率和 mAP@0.5 较 MobileNetV3 高 2.64、1.36 个百分点,

mAP@0.5:0.95 相对 GhostNetV2 高 2.68 个百分点，均显著优于所有对比模型，验证了改进模型在应对光照变化、枝叶遮挡和果实堆叠等复杂干扰因素时的鲁棒性与适应性。对比实验结果显示改进模型“高效-精准”的协同设计在硬件效率与检测性能均衡性上更具优势，为无人机巡检、自动化果园监测、采摘机器人等边缘计算设备部署提供了更高效稳定的实时检测解决方案。

3.5　密集遮挡场景下检测效果对比

为突出改进模型在目标密集遮挡场景下的特征聚焦能力，选取检测优势明显的 YOLO 11n 基准模型和改进模型 YOLO 11n - PConv - SimAMs 分析注意力热力图与检测结果。

使用一种用于可视化深度卷积网络目标关注区域的 Grad - CAM + +[30] 热力图工具对 2 种模型综合 P3、P4、P5 特征提取层分析。图 9 为改进前后模型热力图结果，其中圆圈表示漏检，方框表示误检；改进模型对不同遮挡程度及果实重叠条件下的百香果轮廓区域红色高响应覆盖范围均相对更广。在枝叶遮挡场景图中，当百香果处于非线性直射光条件下，且果皮颜色与叶片高度趋同（RGB 颜色空间平均差值分布在 2 ~ 15 之间）、枝叶遮挡率达 70% 时，基准模型因颜色混淆、特征丢失无法有效区分目标与背景；而在果实重叠场景图中，重叠率达 50% 、无枝叶遮挡干扰时，模型均能实现准确识别，在果实重叠与枝叶遮挡同时存在的情况下，改进模型仍能有效分辨目标。热力图结果进一步表明，改进模型在目标检测框内及全局热力图的激活强度与聚焦性显著优于基准模型，验证了其在密集遮挡环境中具备更高的目标感知敏感度和鲁棒性。

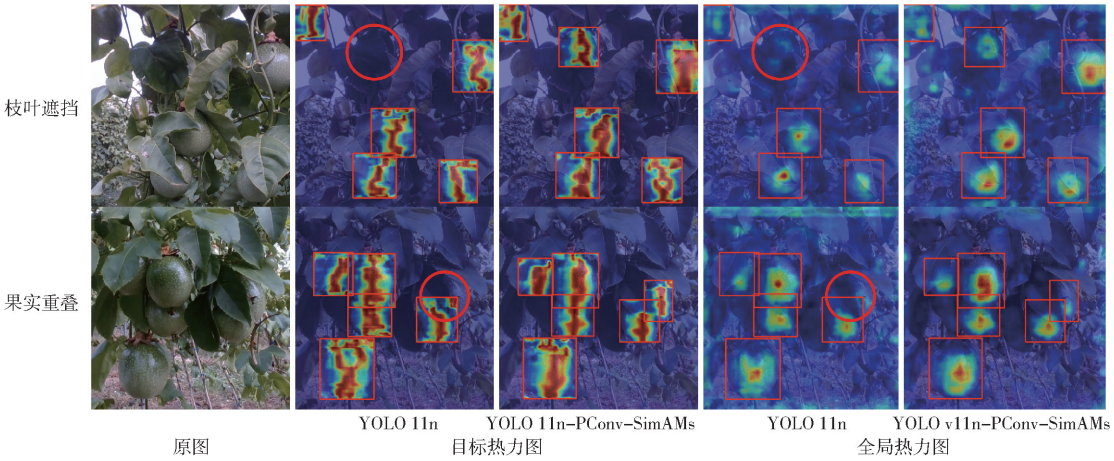


图 9　改进前后模型注意力热力图对比

Fig. 9　Comparison of attention heatmaps between original and improved models

图 10 展示了 2 种模型在不同光线下密集遮挡场景的检测效果，其中圆圈表示漏检，方框表示误检。基准模型在背光和遮挡干扰下出现更多漏检、误检现象，部分检测框与目标边界定位模糊；而改进模型相对识别出更多被遮挡的目标物体，在光照不足或遮挡严重区域保持较高检测准确率，且被遮挡

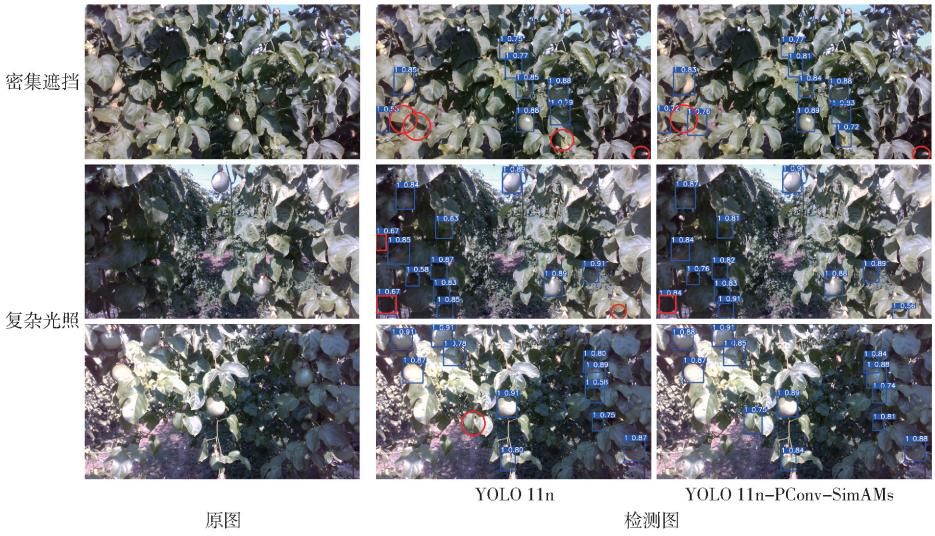


图 10　改进前后模型检测效果对比

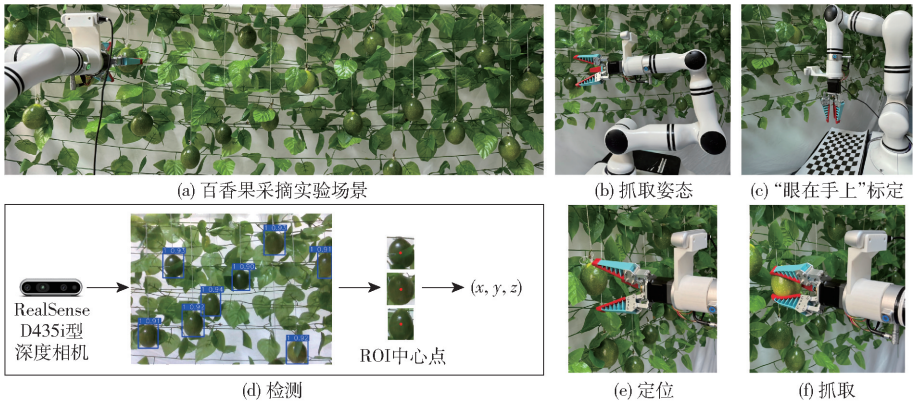
Fig. 10　Comparison of detection performance between original and improved models

目标检测框置信度明显高于基准模型,表明了改进 SimAMs 3D 注意力机制在多目标重叠、枝叶遮挡严重场景中对关键目标区域响应更为显著,提升了目标边缘轮廓判别精度与整体检测稳定性,充分验证了所提出改进策略在密集遮挡场景中的有效性与实用性。

4 百香果采摘实验

为模拟真实果园场景下百香果检测识别、深度定位与机械臂协同采摘,选取同品种、同生长时期的百香果在室内搭建长 1.75 m、高 0.6 m 单行种植区,还原叶片遮挡、藤蔓缠绕等场景特征。实验平台采用具备主动红外立体视觉和 IMU 功能的 Intel RealSense D435i 型深度相机实时获取 RGB - D 图像,结合改进模型实现百香果检测与高精度三维目标定位;运动执行单元选用睿尔曼 RM65 - B 型六自由度协作机械臂,其工作范围覆盖半径 610 mm 球体,有效负载可达 5 kg,具备 ±0.05 mm 重复定位精度,满足采摘作业需求;末端执行器选用 WHEELTEC 型电动三指柔性机械夹爪,尺寸为 85 mm × 98 mm × 155.5 mm,适用于范围 10 ~ 120 mm 和 0 ~ 1 kg 物体抓取任务。

如图 11a 所示,在实验场景中,机械臂挂载在轮式小车上沿平行于百香果单行种植区方向匀速步进,并以固定姿态按“抓取距离末端执行器最近目标”的策略逐一完成相机检测画面内百香果采摘,每完成一轮单行种植区采摘,人工调整果实位置与遮挡布局后小车往返采摘,两轮采摘构成一次实验。在距离百香果单行种植区 300 ~ 400 mm、400 ~ 500 mm 和 500 ~ 600 mm 共 3 种不同深度下设置 3 次采摘实验,对比分析不同作业距离下的单果平均采摘时间与采摘成功率,结果如表 3 所示。



(a) 百香果采摘实验场景 (b) 抓取姿态 (c) “眼在手上”标定 (d) 检测 (e) 定位 (f) 抓取

图 11 百香果采摘实验过程

Fig. 11 Passion fruit harvesting experiment procedure

表 3 不同深度下采摘性能

Tab. 3 Harvesting performance at different depth distances

深度/mm	单果平均采摘时间/s	采摘成功率/%
300 ~ 400	11	80.6
400 ~ 500	13	91.7
500 ~ 600	16	88.9

如图 11d 所示,改进模型在相对简化的遮挡环境中能够高效准确地识别出百香果并提取 ROI 中心点,融合 RGB - D 信息进行三维坐标定位用于机械臂抓取路径规划,其中相机标定及末端执行器多次定位累计误差在水平、深度和垂直方向上分别为(3.56 ± 3.47) mm、(1.88 ± 1.78) mm 和(2.45 ± 1.74) mm;相比基准模型,改进模型在遮挡干扰下提升了 ROI 区域提取稳定性与空间定位一致性,降低因识别偏差导致的抓取失败率。由表 3 可知,随着深度减小,单果采摘时间逐渐缩短;但位于检测画面上下边缘的百香果处于机械臂固定姿态与末端执行器长度共同作用下的可达临界位置,因此,在深度 300 ~ 400 mm 内,机械臂因姿态调整空间受限易发生更多抓取限位问题,整体采摘成功率仅为 80.6%;而随着深度增加,目标可达性逐渐提升,在距离 400 ~ 500 mm 下整体采摘成功率达到最高(91.7%);但进一步增大深度时,部分果实超出机械臂工作范围,导致成功率下降至 88.9%。

5 结论

(1)提出一种面向果园密集遮挡场景下的轻量化百香果检测模型。在 YOLO 11n 特征提取模块与检测头 CIoU 边界框回归计算中引入 PConv 替换标准卷积块,显著降低模型参数量与计算复杂度,减小内存占用量的同时加快推理速度;融合改进 SimAMs 3D 注意力机制,有效增强遮挡目标特征空间感知与关键区域响应能力,改善枝叶遮挡和果实堆叠下的漏检与误检问题,实现模型结构轻量化和对遮挡果实特征表达。

(2)消融实验结果表明,引入 PConv 卷积块后,模型参数量和浮点运算量显著下降 21.2% 和 23.8%,模型内存占用量减少 20.0%,保持检测性能的同时实现了模型轻量化;改进 SimAMs 3D 注意力模块强化了遮挡目标特征激活,使模型在枝叶遮挡、果实堆叠和光照不均等复杂场景下的精确率、召回率、mAP @ 0.5 分别达 93.32%、86.04%、93.08%,分别较原始模型提升 1.28、0.26、0.26 个百分点。

(3)与 EfficientNetV2、FasterNet、GhostNetV2、MobileNetV3 和 ShuffleNetV2 等主流轻量化模型对比实验中,本文模型内存占用量和参数量最低,推理速度最快,在 GPU 和 CPU 端的单幅图像平均检测帧率分别为 158.5 f/s 和 5.8 f/s,且召回率与平均精度均值优于对比模型,验证了其在轻量化部署和高精度检测方面的综合性能优势。

(4)在百香果单行种植区采摘实验中,改进模型在简化遮挡场景下能够快速、精准识别百香果,并对 ROI 中心实现高精定位;3 次不同深度距离采摘结果表明,在深度 400 ~ 500 mm 下百香果采摘成功率达 91.7%,而随着深度距离减少或增加会带来更多限位和超出可达范围的问题。

参 考 文 献

[1] FONSECA A M A, GERALDI M V, JUNIOR M R M, et al. Purple passion fruit (*Passiflora edulis* f. edulis): a comprehensive review on the nutritional value, phytochemical profile and associated health effects[J]. Food Research International, 2022, 160: 111665.

[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] 郑太雄, 江明哲, 冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述[J]. 仪器仪表学报, 2021, 42(9): 28 - 51.
ZHENG Taixiong, JIANG Mingzhe, FENG Mingchi. Vision based target recognition and location for picking robot: a review[J]. Chinese Journal of Scientific Instrument, 2021, 42(9): 28 - 51. (in Chinese)

[5] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报, 2019, 35(3): 164 - 173.
ZHAO Dean, WU Rendi, LIU Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the CSAE, 2019, 35(3): 164 - 173. (in Chinese)

[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] 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.

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

[9] 宋怀波, 尚钰莹, 何东健. 果实目标深度学习识别技术研究进展[J]. 农业机械学报, 2023, 54(1): 1 - 19.
SONG Huaibo, SHANG Yuying, HE Dongjian. Review on deep learning technology for fruit target recognition[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(1): 1 - 19. (in Chinese)

[10] 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.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// Computer Vision - ECCV 2016: 14th European Conference, 2016: 21 - 37.

[12] 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.

[13] SA I, GE Z, DAYOUB F, et al. Deepfruits: a fruit detection system using deep neural networks[J]. Sensors, 2016, 16(8): 1222.

[14] TU S, PANG J, LIU H, et al. Passion fruit detection and counting based on multiple scale faster R - CNN using RGB - D images[J]. Precision Agriculture, 2020, 21: 1072 - 1091.

[15] 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.

[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] SHEN L, SU J, HE R, et al. Real-time tracking and counting of grape clusters in the field based on channel pruning with YOLO v5s[J]. Computers and Electronics in Agriculture, 2023, 206: 107662.
[18] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements[J]. arXiv Preprint, arXiv:2410.17725, 2024.
[19] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021 - 12031.
[20] 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.
[21] WANG Q, WU B, ZHU P, et al. ECA - Net: efficient channel attention for deep convolutional neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534 - 11542.
[22] 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.
[23] ZHANG Q L, YANG Y B. SA - Net: shuffle attention for deep convolutional neural networks[C]// ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021: 2235 - 2239.
[24] 张宸嘉, 朱磊, 俞璐. 卷积神经网络中的注意力机制综述[J]. 计算机工程与应用, 2021, 57(20):64 - 72.
ZHANG Chenjia, ZHU Lei, YU Lu. Review of attention mechanism in convolutional neural networks[J]. Computer Engineering and Applications, 2021, 57(20):64 - 72. (in Chinese)
[25] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]// International Conference on Machine Learning. PMLR, 2021: 11863 - 11874.
[26] TAN M, LE Q. EfficientNetv2: smaller models and faster training[C]// International Conference on Machine Learning. PMLR, 2021: 10096 - 10106.
[27] 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.
[28] 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.
[29] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet v2: practical guidelines for efficient CNN architecture design[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 116 - 131.
[30] CHATTOPADHAY A, SARKAR A, HOWLADER P, et al. Grad-CAM + +: generalized gradient-based visual explanations for deep convolutional networks[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 839 - 847.

(上接第 **166** 页)

[15] GU Z, HE D, HUANG J, et al. Simultaneous detection of fruits and fruiting stems in mango using improved YOLO v8 model deployed by edge device[J]. Computers and Electronics in Agriculture, 2024, 227: 109512.
[16] ZHENG C, CHEN P, PANG J, et al. A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard[J]. Biosystems Engineering, 2021, 206: 32 - 54.
[17] LI H, CHEN J, GU Z, et al. Optimizing edge-enabled system for detecting green passion fruits in complex natural orchards using lightweight deep learning model[J]. Computers and Electronics in Agriculture, 2025, 234: 110269.
[18] Ultralytics. YOLO v8[CP/OL]. (2023 - 01 - 10)[2025 - 10 - 08]. https://github.com/ultralytics/ultralytics.
[19] WANG C Y, YEH I H, LIAO H Y M. YOLO v9: learning what you want to learn using programmable gradient information[C]// European Conference on Computer Vision. Springer Nature Switzerland, 2024: 1 - 21.
[20] WANG A, CHEN H, LIU L, et al. YOLO v10: real-time end-to-end object detection[C]// arXiv Preprint, arXiv:2405.14458, 2024.
[21] KHANAM R, HUSSAIN M. YOLO v11: an overview of the key architectural enhancements[C]// arXiv Preprint, arXiv:2410.17725, 2024.
[22] TIAN Y, YE Q, DOERMANN D. YOLO v12: attention-centric real-time object detectors[C]// arXiv Preprint, arXiv:2502.12524, 2025.
[23] LEI M, LI S, WU Y, et al. YOLO v13: real-time object detection with hypergraph-enhanced adaptive visual perception[C]// arXiv Preprint, arXiv:2506.17733, 2025.