

YOLO-Apple: 一种用于苹果幼果检测的改进型目标检测方法

韩昕辰^{1,2}, 安一文^{1,2*}, 武红欣^{1,2}

¹塔里木大学信息工程学院, 新疆 阿拉尔

²塔里木绿洲农业教育部重点实验室, 新疆 阿拉尔

收稿日期: 2026年4月23日; 录用日期: 2026年5月22日; 发布日期: 2026年5月29日

摘要

针对自然果园场景中, 苹果幼果因通体青绿色、与叶片色调相近, 且本身尺度偏小、分布密集又易被枝叶遮挡, 检测时易出现漏检与误检的问题, 文章提出了一种面向苹果幼果检测的改进型目标检测模型 YOLO-Apple。该方法以 YOLOv11n 为基线, 在主干 C2PSA 模块引入轻量特征增强单元 Mona (C2PSA_Mona) 以强化复杂背景下的特征映射能力, 改造 C3k2 模块并在其增强块嵌入 EMA 注意力 (C3k2_EMA), 提升特征表达与选择能力。实验采用来自新疆阿克苏市与石河子市等自然场景采集的 5558 张苹果幼果图像数据集, 按 7:2:1 划分为训练、验证与测试集。结果表明, YOLO-Apple 在测试集上实现 $P = 91.5\%$ 、 $R = 75.9\%$ 、 $F1 = 83.0$ 、 $mAP@0.5 = 84.4\%$ 的检测性能, 在 6.5 GFLOPs 计算量与 2.6 M 参数量的轻量级开销下优于多款主流对比模型; 消融实验验证了 C2PSA_Mona 与 C3k2_EMA 模块的改进有效性, 为自然果园苹果幼果检测提供了新方法。

关键词

苹果幼果, 目标检测, YOLO, 注意力机制

YOLO-Apple: A Modified Object Detection Method for Apple Young-Fruit Detection

Xinchen Han^{1,2}, Yiwen An^{1,2*}, Hongxin Wu^{1,2}

¹College of Information Engineering, Tarim University, Alaer Xinjiang

²Key Laboratory of Tarim Oasis Agriculture, Ministry of Education, Alaer Xinjiang

Received: April 23, 2026; accepted: May 22, 2026; published: May 29, 2026

*通讯作者。

文章引用: 韩昕辰, 安一文, 武红欣. YOLO-Apple: 一种用于苹果幼果检测的改进型目标检测方法[J]. 计算机科学与应用, 2026, 16(5): 525-533. DOI: 10.12677/csa.2026.165203

Abstract

To address the missed and false detection of young apple fruits in natural orchard scenes caused by their turquoise color, highly similar to leaf tones, along with small target scale, dense spatial distribution, and easy occlusion by branches and leaves, this paper proposes an improved object detection model, YOLO-Apple, for young apple fruit detection. Based on the YOLOv11n baseline, this method introduces the lightweight feature enhancement unit Mona into the backbone C2PSA module (denoted as C2PSA_Mona) to strengthen the feature mapping capability in complex backgrounds, and modifies the C3k2 module by embedding the EMA attention mechanism into its enhancement block (denoted as C3k2_EMA) to improve the feature representation and selection capabilities. Experiments were performed on a dataset of 5558 young apple fruit images collected from natural scenes in Aksu City and Shihezi City of Xinjiang, which was partitioned into training, validation and test sets at a ratio of 7:2:1. Experimental results show that YOLO-Apple achieves the detection performance of $P = 91.5\%$, $R = 75.9\%$, $F1 = 83.0$ and $mAP@0.5 = 84.4\%$ on the test set, and outperforms multiple mainstream comparison models with a lightweight overhead of 6.5 GFLOPs and approximately 2.6 M parameters. Ablation experiments further verify the effectiveness of the C2PSA_Mona and C3k2_EMA modules, thus providing a novel method for young apple fruit detection in natural orchard scenes.

Keywords

Apple Young Fruit, Object Detection, YOLO, Attention Mechanism

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

苹果因其独特的口感和较好的营养价值，在新疆阿克苏地区被广泛种植[1][2]。在苹果幼果阶段，统计其数量对种植者至关重要，因为其有助于资源的高效利用和配置[3]。传统人工计数的方法不仅劳动强度大，且容易受到客观因素的影响导致数据错误[4][5]。因此，种植者迫切需要一种高通量且廉价的苹果幼果检测方法。

随着计算机视觉技术的发展，一些研究者将计算机视觉技术应用到苹果识别中。Gao 等人[6]基于 YOLOv4-tiny 模型，在果园中实现自动化视频提取苹果果实，为果园中的苹果数量统计提供了新的方法。郁家丞等人[7]通过改进 YOLOv7，在主干网络中添加 CBAM 注意力机制、更换损失函数和更换卷积的方法，以缓解果园复杂环境下苹果目标难以检测、正负样本不均衡和检测速度慢的问题，在提升检测精度的同时，满足了实时检测的需求。黄志豪等人[8]基于 YOLOv11n 模型，结合动态卷积和幽灵卷积，改进 C3k2 模块，以增强模型特征提取能力的同时，降低模型复杂度，并且对网络颈部结果进行重构，以增强模型局部和全局建模能力。通过这些改进，在有效提升模型检测能力的同时，还降低了模型复杂度，使其更便于在田间部署，为苹果果园的信息化发展提供了坚实的数据支撑。然而，现有研究大多聚焦于苹果近成熟期或成熟期的检测实验——此时苹果呈粉红色或红色，与果园绿色背景对比鲜明，检测难度较低；针对苹果幼果的检测研究则相对匮乏。苹果幼果为青绿色，与叶片颜色极为相近，这无疑为检测任务带来了不小的挑战。

基于自然环境下苹果幼果检测的应用需求，本文提出一种面向苹果幼果的自动化检测模型 YOLO-Apple。该模型以 YOLOv11n 为基线模型，融合改进的 C2PSA 模块，旨在强化对复杂背景下目标的特征映

射能力，从而提升检测精度；同时对 C3k2 模块进行改造，增强其特征表达性能。最后，通过消融实验验证了各模块的有效性，并与其他主流模型进行对比，证明模型的优越性，以为自动化苹果幼果检测提供新的解决方案。

2. 材料与方法

2.1. 数据集构建

在本研究中，数据集图像来自 AI Studio 的苹果幼果目标识别数据集[9]。其图片于 5 月和 6 月在新疆阿克苏市和石河子市的农田、果园和街道拍摄，拍摄周期为两周一次，涵盖一天中的不同时段，苹果幼果此时处于青绿色状态，如图 1 所示。

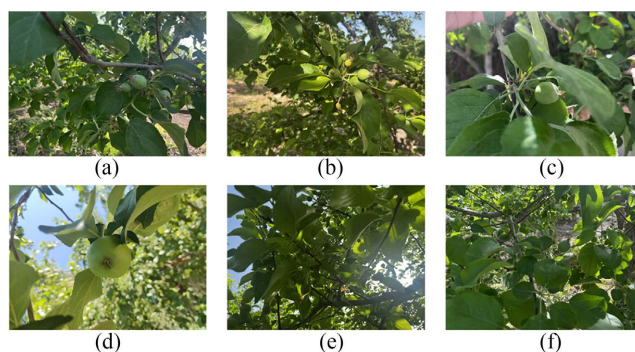


Figure 1. Display of young apples

图 1. 苹果幼果展示

该数据集共包含 5558 张图像。本研究对该数据集的图像和标注进行混合，按照 7:2:1 的比例重新划分为训练集、验证集和测试集，具体数量为：训练集 3890 张、验证集 1111 张和测试集 557 张。在训练过程中，全程使用 YOLO 系列默认的数据增强方法，包括马赛克数据增强等。

2.2. YOLO-Apple

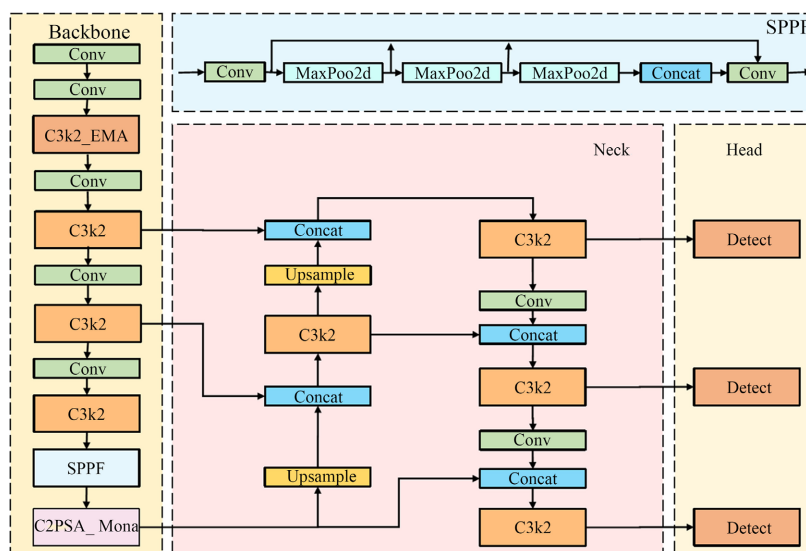


Figure 2. Structure diagram of YOLO-Apple

图 2. YOLO-Apple 结构图

在本研究中，使用 YOLOv11n 作为基线模型，该模型为 YOLO 系列的较新模型。在确定基线模型前，本文选取了多种主流模型作为对比实验。结果显示，YOLOv11n 在苹果幼果检测任务中不仅检测精度表现突出，且模型复杂度更低，最终被选定为本研究的基线模型。以此为基础，本文通过对 C3k2 模块和 C2PSA 模块进行针对性改进，构建得到 YOLO-Apple 检测模型，其结构图如图 2 所示。

2.3. C2PSA_Mona

在果园自然场景中，苹果幼果通常具有尺度小、密集分布、易遮挡等特点，且常与叶片高光、枝条纹理在局部外观上相似，容易引发误检与漏检。YOLOv11n 主干中的 C2PSA 通过“分支保留 + 注意力增强”的方式提升表示能力，但其注意力与 FFN 更偏向全局依赖与通道映射，对于幼果这种细粒度局部纹理与多尺度边界信息仍可能不够敏感。因此，本文将 Mona [10]作为轻量级的特征增强单元插入 C2PSA 的 PSABlock 中：在 Attention 残差与 FFN 残差之后分别引入 Mona，对关键中间特征进行分布校正和多尺度局部建模，以强化幼果的边缘、纹理响应并抑制复杂背景干扰。C2PSA_Mona 的结构图如图 3 所示。

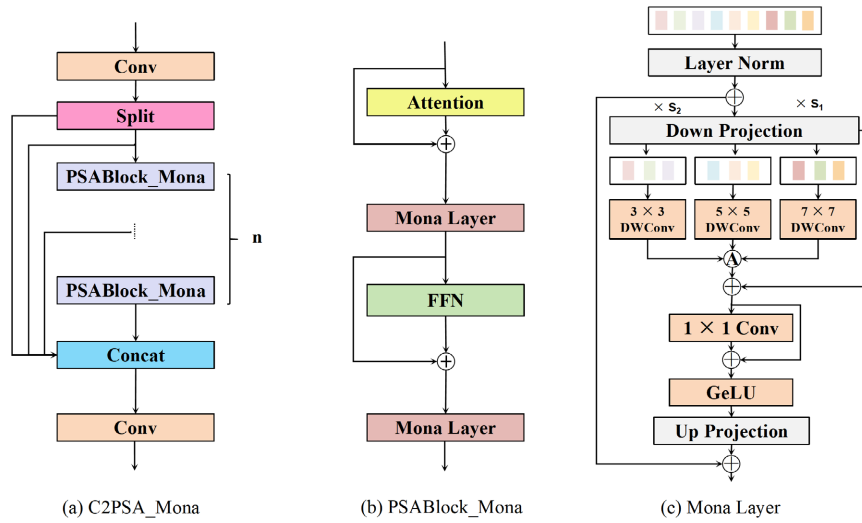


Figure 3. Structure diagram of C2PSA_Mona
图 3. C2PSA_Mona 结构图

设 PSABlock_Mona 的输入特征为 $x \in \mathbb{R}^{B \times C \times H \times W}$ ，则 PSABlock_Mona 可用数学表达为：

$$x_1 = x + \text{Attn}(x) \tag{1}$$

$$x_2 = \text{Mona}(x_1) \tag{2}$$

$$x_3 = x_2 + \text{FFN}(x_2) \tag{3}$$

$$y = \text{Mona}(x_3) \tag{4}$$

其中，Attn 为注意力机制，FFN 为前馈网络，Mona 为 Mona Layer。对于 Mona 而言，其由两个部分构成：输入优化与多尺度多认知卷积滤波。

对于输入优化，Mona 使用 LayerNorm 和两个可学习的缩放因子调节输入分布与比例：

$$x_{norm} = s_1 \cdot \text{LN}(x_0) + s_2 \cdot x_0 \tag{5}$$

其中， x_0 为 Mona 的输入， s_1 、 s_2 为可学习标量，LN(·) 为 LayerNorm。

对于多尺度多认知视觉滤波，为模拟苹果幼果不同尺度视觉认知并融合的机制，使用三路不同大小

的卷积(卷积核大小分别为 7×7 、 5×5 和 3×3)，再取平均，并用 1×1 的卷积进行融合，同时在两处均加入残差进行增强：

$$f_{dw} = x + \text{avg} \left(\sum_{i=1}^3 \omega_{dw}^{(i)} \otimes_{dw} x \right) \tag{6}$$

$$f_{pw} = x + \omega_{pw} \otimes_{pw} x \tag{7}$$

其中， $\text{avg}(\cdot)$ 为多尺度分支结果求平均融合， \otimes_{dw} 表示深度卷积， \otimes_{pw} 表示 1×1 点卷积。

在经过多尺度多认知滤波后，Mona 的整体映射，可表示为：

$$x = x_0 + U_l \left(\sigma \left(f_{pw} \left(f_{dw} \left(D_l(x_{norm}) \right) \right) \right) \right) \tag{8}$$

其中， $D_l(\cdot)$ 和 $U_l(\cdot)$ 分别为第 l 个 Mona 的下投影和上投影， $\sigma(\cdot)$ 为 GeLU。

由于 Mona 使用深度卷积代替标准卷积，控制了额外参数量，因此在 C2PSA 中插入 Mona 能在较小额外开销下提升局部与多尺度表征。

2.4. C3k2_EMA

为增强特征表达并在较低计算开销下引入注意力建模，本文在 C3k2 框架中嵌入 EMA 注意力[11]，构建 C3k2_EMA。整体结构与 C3k2 结构基本一致，即输入特征经 1×1 卷积后分裂为两支，其中一路保持直连，另一路依次通过若干个增强块(Bottleneck_EMA 或 C3k_EMA)，最终将两支与各级输出拼接并用 1×1 卷积融合得到输出。EMA 的结构如图 4 所示。

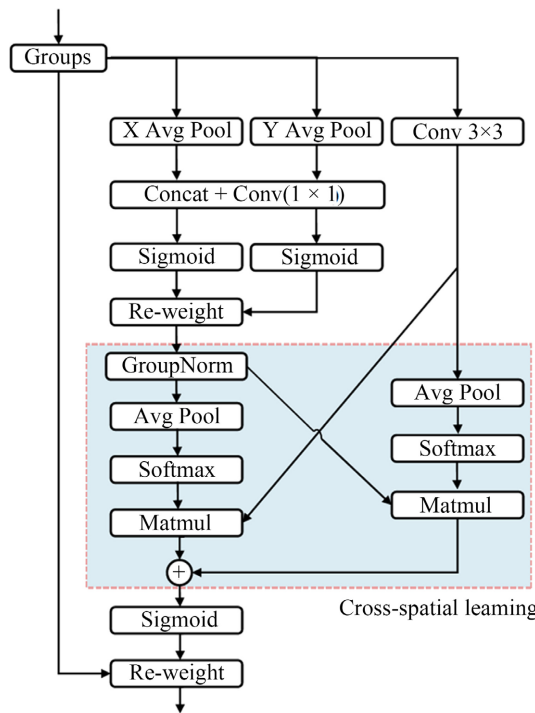


Figure 4. EMA structure diagram

图 4. EMA 结构图

设 EMA 输入特征为 $X \in \mathbb{R}^{B \times C_1 \times H \times W}$ 。首先将通道划分为 G 组，并重排为组内特征：

$$X_g = \text{Reshape}(X) \in \mathbb{R}^{(BG) \times (C/G) \times H \times W} \tag{9}$$

随后, EMA 分别沿高度和宽度方向进行自适应平均池化以提取方向上下文:

$$X_h = \text{Pool}_h(X_g) \in \mathbb{R}^{(BG) \times (C/G) \times H \times 1}, \quad X_w = \text{Pool}_w(X_g) \in \mathbb{R}^{(BG) \times (C/G) \times 1 \times W} \quad (10)$$

将两者在空间维度拼接并经过 1×1 卷积融合后, 再拆分得到沿 H/W 的门控项实现坐标调制; 该调制结果经 GroupNorm 得到一支特征 X_1 , 同时对 X_g 施加 3×3 卷积得到局部上下文分支 X_2 :

$$X_1 = \text{GN}(X_g \odot \sigma(\phi_h) \odot \sigma(\phi_w)), \quad X_2 = \text{Conv}_{3 \times 3}(X_g) \quad (11)$$

在此基础上, EMA 使用“GAP + Softmax”生成跨分支的全局查询向量, 并与另一分支的展平特征进行矩阵乘, 以形成空间权重图:

$$W = \text{MatMul}(\text{Softmax}(\text{GAP}(X_1)), \text{vec}(X_2)) + \text{MatMul}(\text{Softmax}(\text{GAP}(X_2)), \text{vec}(X_1)) \quad (12)$$

最终将权重经 Sigmoid 激活后对组内特征进行重标定并恢复原形状:

$$Y = \text{Reshape}(X_g \odot \sigma(W)) \in \mathbb{R}^{B \times C \times H \times W} \quad (13)$$

由此, EMA 同时引入方向敏感的全局上下文与局部结构信息, 并通过轻量的跨分支全局交互生成空间注意权重, 从而在较低开销下实现对关键区域的自适应强调与冗余抑制, 增强了复杂背景下苹果幼果的特征表达和选择能力, 提升检测精度。

2.5. 评估指标

为选择出最佳模型, 本研究使用多种评估指标用于评估模型的性能和计算效率, 包括精确率(Precision, P)、召回率(Recall, R)、mAP@0.5、F1 分数(F1-score)、计算量(GFLOPs)和参数量(Parameters)等。其计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (14)$$

$$R = \frac{TP}{TP + FN} \quad (15)$$

$$F1 = \frac{2PR}{P + R} \quad (16)$$

$$\text{mAP} @ 0.5 = \frac{1}{C} \sum_{c=1}^C AP_c |_{IoU \geq 0.5} \quad (17)$$

其中, TP (True Positive)表示被模型正确检测并识别为苹果幼果的目标数量; FP (False Positive)表示被模型误检为苹果幼果的目标数量; FN (False Negative)表示实际为苹果幼果但未被检测到(被漏检或被误判为背景、其他目标)的目标数量。此外, F1 分数(F1-score)为精确率与召回率的调和平均, 能够在两者之间取得平衡, 适用于同时关注误检与漏检的评价场景。平均精度(Average Precision, AP)定义为精确率 - 召回率(Precision-Recall, PR)曲线下的面积, 用于综合衡量模型在不同置信度阈值下的检测性能。需要注意的是, 除非特别说明, 本文所报告的所有评估指标均基于测试集的结果。

2.6. 实验环境

所有实验均在同一台设备上, 该设备配备 AMD5500 处理器, NVIDIA GeForce RTX 5080 显卡, 该显卡具有 16 GB 显存, 所使用系统为 Ubuntu 22.04, 内存为 48 GB。在实验环境上, 使用 Python 3.11.5、CUDA12.8 和 PyTorch 2.9.1 作为框架。在训练设置上, 图片大小为 640×640 , 训练轮次为 200 轮, 使用 SGD 作为优化器, 批量大小设置为 8。

3. 结果

3.1. 消融实验

为验证所提出模块对苹果幼果检测性能的贡献, 本文以基线模型(YOLOv11n)为对照, 逐步引入各模块进行消融实验, 结果如表 1 所示。

Table 1. Results of ablation experiments

表 1. 消融实验结果

Model	C2PSA_Mona	C3k2_EMA	P (%)	R (%)	F1-Score	mAP@0.5 (%)	GFLOPs	Parameters (M)
1	×	×	92.3	73.6	81.9	82.3	6.3	2.6
2	√	×	91.3	74.8	82.2	82.9	6.4	2.6
3	√	√	91.5	75.9	83.0	84.4	6.5	2.6

与基线模型相比, 仅加入 C2PSA_Mona (Model 2)后, Recall 从 73.6%提升至 74.8%, F1-score 从 81.9 提升至 82.2, mAP@0.5 从 82.3%提升至 82.9%, 说明该模块能有效减少漏检并提升整体检测质量; 同时计算量与参数量几乎不变。

在此基础上进一步引入 C3k2_EMA (Model 3)后, 性能提升更为明显: Recall 提升至 75.9%, F1-score 达到 83.0, mAP@0.5 提升至 84.4%。相较基线, mAP@0.5 提高 2.1 个百分点, F1-score 提高 1.1。与此同时, 计算量仅小幅增加, 参数量保持不变, 表明所提改进在几乎不增加模型规模的前提下显著提升了苹果幼果检测效果。

3.2. 对比实验

为验证 YOLO-Apple 模型在苹果幼果检测中的性能优势, 本研究选取了多种主流的目标检测算法作为对比实验, 包括 YOLOv8n、YOLOv9-tiny [11]、YOLOv10n [12]、YOLOv12n [13]和 YOLOv13n [14]。实验结果如表 2 所示。

Table 2. Comparative experimental results

表 2. 对比实验结果

Model	P (%)	R (%)	F1-Score	mAP@0.5 (%)	GFLOPs	Parameters (M)
YOLOv8n	90.6	74.4	81.7	82.4	8.1	3.0
YOLOv9n	91.6	73.3	81.4	82.0	7.7	2.0
YOLOv10n	92.4	72.8	81.5	82.2	6.7	2.3
YOLOv12n	91.8	73.4	81.6	82.1	5.8	2.5
YOLOv13n	91.2	73.5	81.4	81.8	6.1	2.4
YOLO-Apple	91.5	75.9	83.0	84.4	6.5	2.6

从检测精度综合指标看, YOLO-Apple 的 mAP@0.5 达到 84.4%, 显著高于所有对比模型 (81.8%~82.4%), 相对最优基线 YOLOv8n (82.4%)提升 2.0 个百分点, 说明模型对幼果目标的定位与分类更准确。同时, YOLO-Apple 的 F1-score 为 83.0, 也高于各对比模型(81.4~81.7), 表明其在 Precision 与 Recall 之间取得了更好的平衡。

从召回率看, YOLO-Apple 的 Recall 达到 75.9%, 为所有模型中最高, 相比 YOLOv8n (74.4%)提升

1.5 个百分点, 相比 YOLOv10n (72.8%)提升 3.1 个百分点, 说明其能够有效减少幼果漏检, 尤其适用于果实密集、遮挡或小目标比例较高的场景。

在计算开销方面, YOLO-Apple 的计算量为 6.5 GFLOPs, 低于 YOLOv8n, 与 YOLOv9-tiny、YOLOv10n、YOLOv12n 和 YOLOv13n 处于同一轻量级范围; 参数量约为 2.6 M, 同样保持在较小规模。综合来看, YOLO-Apple 在基本不显著增加模型规模的前提下取得更高的检测精度与召回性能, 体现了其在苹果幼果检测任务中的有效性与应用潜力。

为直观展现 YOLO-Apple 的检测优势, 本研究随机从测试集选取两张图像进行推理, 结果如图 5 所示。从检测结果可以看出, 其他模型均有错检或漏检现象, YOLO-Apple 检测效果较好。

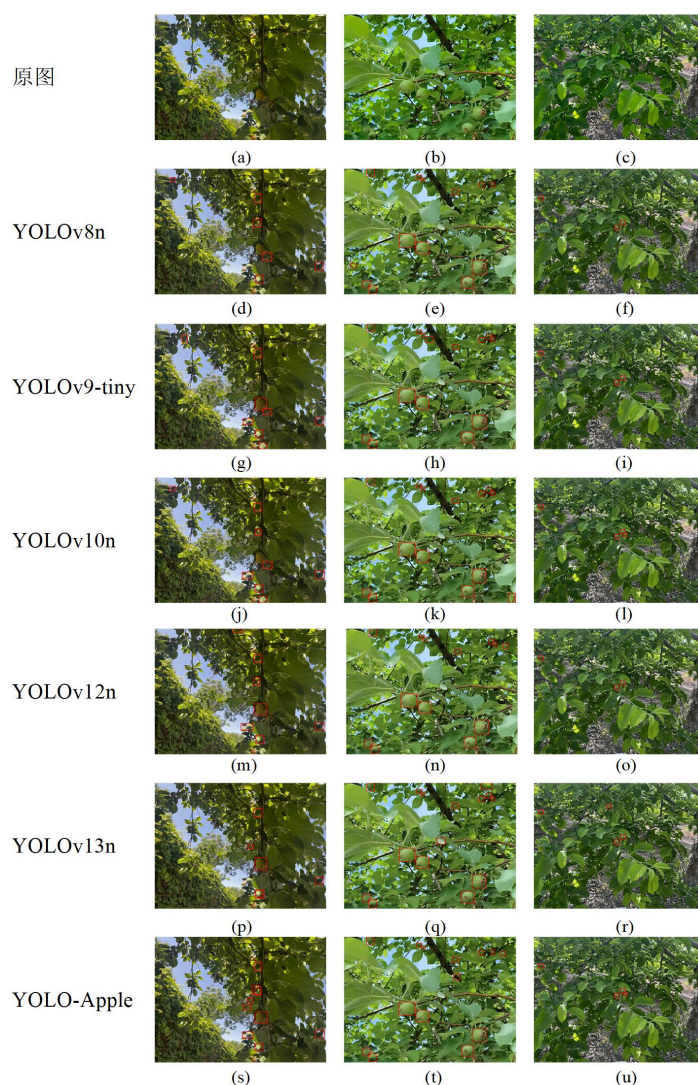


Figure 5. Display of test results
图 5. 检测结果展示

4. 结论与讨论

本研究针对苹果幼果小目标、遮挡与密集分布等检测难点, 提出检测模型 YOLO-Apple。在实验数据集上, YOLO-Apple 在保持较低复杂度(6.5 GFLOPs, 2.6 M 参数)的前提下取得最佳综合性能: $P=91.5\%$ 、

$R = 75.9\%$ 、 $F1 = 83.0$ 、 $mAP@0.5 = 84.4\%$ 。与多种主流轻量模型对比，YOLO-Apple 的 $mAP@0.5$ 相比最优基线提升约 2.0%，召回率也达到最高，说明其对幼果漏检抑制更有效。消融实验进一步表明，逐步引入 C2PSA_Mona 与 C3k2_EMA 可稳定提升检测效果，最终在几乎不增加参数量的情况下实现精度与召回的同步改善。

从结果看，YOLO-Apple 的优势主要体现在召回率与整体检测质量(F1、mAP)的提升，这对幼果计数、疏果决策等应用尤为关键；同时计算量处于轻量级范围，具备端侧部署潜力。模块贡献方面，C2PSA_Mona 有助于增强特征表达、降低漏检，而 C3k2_EMA 进一步强化对关键区域的关注，使 $mAP@0.5$ 获得更明显增益，体现了注意力与特征融合改进对小目标场景的有效性。需要指出的是，当前工作仍可能受限于复杂光照、强遮挡、背景相似与跨果园/跨季节域偏移等因素；此外，极小目标或严重重叠情况下仍可能出现漏检与误检。后续可从三方面改进：一是引入更强的数据增强与域自适应以提升泛化；二是结合更精细的标注策略或多尺度监督提升极小目标识别；三是在部署端进行剪枝/量化与推理优化，实现更高帧率与更低功耗的实时检测。

基金项目

塔里木大学校长基金胡杨英才项目(TDZKSS202341)。

参考文献

- [1] 许瑛, 徐学里, 杨俊翔, 等. 阿克苏红富士苹果不同采收期微量元素及糖类物质的动态变化[J]. 果树资源学报, 2025, 6(5): 5-9.
- [2] 巫利梅, 张燕, 宋娟, 等. 新疆阿克苏苹果产业现状、存在问题与发展建议[J]. 中国果树, 2025(8): 129-133.
- [3] 龚圳玮, 彭伟, 田雅暄. 改进 YOLOv7 和 DeepSort 的视频苹果数量检测[J]. 计算技术与自动化, 2025, 44(4): 94-101.
- [4] 杨帆, 周杰, 吴昊荣, 等. 基于改进 YOLOv7 的苹果目标检测算法研究[J]. 农业装备与车辆工程, 2025, 63(9): 13-18.
- [5] 盛卉子, 王春宇, 许正荣, 等. 基于 RGB-D 图像的苹果检测[J]. 合肥大学学报, 2025, 42(2): 107-114.
- [6] Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., *et al.* (2022) A Novel Apple Fruit Detection and Counting Methodology Based on Deep Learning and Trunk Tracking in Modern Orchard. *Computers and Electronics in Agriculture*, **197**, Article 107000. <https://doi.org/10.1016/j.compag.2022.107000>
- [7] 郁家丞, 杨林初, 夏子林, 等. 基于改进 YOLOv7 的果园苹果目标识别方法研究[J]. 中国农机化学报, 2026, 47(1): 87-93.
- [8] 黄志豪, 卢承方, 崔艳荣, 等. YOLO-AP: 基于改进 YOLO11n 的轻量级苹果果实检测算法[J]. 中国农业科技导报(中英文), 2025, 27(10): 118-133.
- [9] 飞桨 AI Studio 星河社区. 苹果幼果目标识别数据集[EB/OL]. <https://aistudio.baidu.com/datasetdetail/351266>, 2026-01-08.
- [10] Yin, D., Hu, L., Li, B., Zhang, Y. and Yang, X. (2025) 5% > 100%: Breaking Performance Shackles of Full Fine-Tuning on Visual Recognition Tasks. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 10-17 June 2025, 20071-20081. <https://doi.org/10.1109/cvpr52734.2025.01869>
- [11] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., *et al.* (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. 2023 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 4-10 June 2023, 1-5. <https://doi.org/10.1109/icassp49357.2023.10096516>
- [12] Chen, H., Chen, K., Ding, G., Han, J., Lin, Z., Liu, L., *et al.* (2024) YOLOv10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems 37*, Vancouver, 10-15 December 2024, 107984-108011. <https://doi.org/10.52202/079017-3429>
- [13] Tian, Y., Ye, Q. and Doermann, D. (2025) YOLOv12: Attention-Centric Real-Time Object Detectors. <https://arxiv.org/abs/2502.12524>
- [14] Lei, M., Li, S., Wu, Y., Hu, H., Zhou, Y., Zheng, X., *et al.* (2025) YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception. <https://arxiv.org/abs/2506.17733>