

基于轻量级 YOLO v8 改进模型的苹果入库高密度计数方法

郭成波¹ 姜文文¹ 郭艳玲² 孙术发¹

(1. 东北林业大学土木与交通学院, 哈尔滨 150040; 2. 东北林业大学机电工程学院, 哈尔滨 150040)

摘要: 在规模化苹果自动入库场景中, 基于计算机视觉的计数模型需兼顾轻量化与检测精度, 传统模型参数量大、计算成本高, 难以实时运行; 同时在苹果密集、遮挡严重的复杂环境下, 存在边界模糊、误检率高等问题。为此, 本文提出一种改进的 CGW-YOLO v8 模型。首先, 通过将主干网络中的 C2f 模块替换为 GhostNet 轻量化模块, 结合特征通道重加权机制, 显著减少了模型的参数量。其次, 采用 CSPHet 模块, 通过异构多分支卷积与双路径特征融合策略, 在降低参数量的同时增强密集苹果目标的边界区分能力。最后, 采用基于 Wasserstein distance loss 的损失函数替代传统的 IoU 度量, 有效降低了密集堆叠场景中的误检率。实验表明, 本模型平均精度均值 mAP@0.5 提升至 95.8%, 较原模型提升 1 个百分点, 精确率和召回率较原模型分别提升 1.1、1.3 个百分点, 参数量与运算量较基准模型分别减少 24.4% 和 23.2%。针对入库生产环节中对计数实时性与准确性双重要求, 本文集成 DeepSORT 追踪算法实现苹果在视频帧间的持续跟踪与准确计数。设计了基于轨迹管理的计数策略, 通过虚拟计数线仅在目标首次越过时进行计数, 有效避免了重复统计与漏计问题。实验结果表明, 所提出的改进方法在复杂背景下, 尤其是苹果密集排列与部分遮挡的场景中, 展现了较强的鲁棒性和较高的计数准确性。

关键词: 苹果入库计数; 目标检测; GhostNet; CSPHet; Wasserstein 距离; DeepSORT

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-1298(2026)05-0342-11

OSID:



High-density Apple Counting Method Based on Lightweight Improved YOLO v8

GUO Chengbo¹ JIANG Wenwen¹ GUO Yanling² SUN Shufa¹

(1. Faculty of Civil Engineering and Transportation, Northeast Forestry University, Harbin 150040, China

2. Faculty of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China)

Abstract: In the scenario of large-scale automatic apple warehousing, the counting model based on computer vision must balance lightweight design with detection accuracy. Traditional models involve large parameters and high computational costs, making real-time operation challenging; moreover, in complex environments characterized by dense apples and severe occlusion, issues such as boundary blurriness and high false detection rates arise. To address these challenges, an improved CGW-YOLO v8 model was proposed. Firstly, by replacing the C2f module in the backbone network with the lightweight GhostNet module and incorporating a feature channel reweighting mechanism, the model's parameter count was significantly reduced. Secondly, the CSPHet module was employed, which utilized heterogeneous multi-branch convolution and a dual-path feature fusion strategy to enhance the boundary distinguishing capability of densely packed apples while decreasing the number of parameters. Lastly, a loss function based on Wasserstein distance loss was adopted to replace the traditional IoU metric, effectively reducing the false detection rate in densely stacked scenarios. Experimental results indicated that the model's mean average precision (mAP@0.5) was improved to 95.8%, representing a 1 percentage points increase compared with that of the original model, with precision and recall increased by 1.1 percentage points and 1.3 percentage points, respectively, compared with that of the original model. Both parameter count and computational volume were decreased by 24.4% and 23.2% relative to that of the baseline model. To meet the dual requirements of real-time counting and accuracy in the inventory production process, the DeepSORT tracking algorithm was integrated to achieve continuous tracking and accurate counting of

收稿日期: 2025-07-14 修回日期: 2025-09-19

基金项目: 国家自然科学基金项目(52075090)和黑龙江省自然科学基金项目(PL2024C003)

作者简介: 郭成波(1987—), 男, 副教授, 博士, 主要从事智能物流装备研究, E-mail: chengbo.guo@nefu.edu.cn

apples across video frames. A counting strategy based on trajectory management was designed, which counted only when the target first crossed a virtual counting line, effectively avoiding issues of duplicate statistics and missed counts. Experimental results demonstrated that the proposed improved method exhibited strong robustness and high counting accuracy, particularly in complex backgrounds, such as densely arranged apples and partially obscured scenes.

Key words: apple warehousing count; object detection; GhostNet; CSPHet; Wasserstein distance; DeepSORT

0 引言

苹果计数盘点是入库管理中的关键环节,其准确性对库存管理效率和物流优化具有直接影响。然而,考虑到苹果数量庞大且在入库过程中常常发生堆叠与遮挡,传统的人工计数方法不仅增加了运营成本,且容易导致盘点精度降低,甚至可能造成苹果的挤压或损坏。此外,在数字化与智能化趋势下,引入基于计算机视觉的检测模型成为提升苹果计数盘点效率与精度的有效途径。但仓库环境复杂,硬件资源有限,部署的检测模型需要具备轻量化特性。因此,开展基于轻量化目标检测模型的自动计数方法研究至关重要。有助于提升库存管理效率,满足实时盘点的需求,进一步推动苹果产业的智能化转型和可持续发展。

目标检测领域主要存在2种主流路线:一阶段和二阶段检测算法,其中二阶段算法通常包括特征提取、生成预选框,以及分类和定位回归,精度较高但计算速度较慢,常见的算法有R-CNN(Region-based convolutional neural network)^[1]、SPP-Net(Spatial pyramid pooling network)^[2]、Fast R-CNN^[3]、Faster R-CNN^[4-5]和R-FCN(Region-based fully convolutional networks)^[6]等;而一阶段算法则在特征提取后直接进行分类和定位回归,速度较快,代表算法有OverFeat^[7]、YOLO(You only look once)系列^[8-13]、SSD(Single shot multiBox detector)^[14]和RetinaNet^[15]等。总体而言,二阶段方法侧重精度,适合高精度需求的场景,而一阶段方法则强调速度,适合实时检测任务。

近年来,轻量化目标检测算法在农业检测领域^[16-19]和农业计数方法研究领域^[20-21]取得了重要研究进展。张传栋等^[22]提出了融合随机注意力机制与Wise-IoU损失的SAW-YOLO v8n模型,用于复杂环境下的葡萄簇幼果检测,在模型内存占用量仅为6.20 MB的同时,实现了96.10%的mAP,显著优于多种轻量化基准模型。冀汶莉等^[23]针对农田杂草识别,构建了基于YOLO v5的轻量化模型,通过引入PP-LCNet、Ghost卷积与NAM注意力模块,将模型内存占用量压缩至6.23 MB,并在自建数据

集上取得了97.8%的mAP。邓铭辉等^[24]为提升群养猪只的检测效率,提出了Ghost-YOLOX-BiFPN模型,利用Ghost卷积与结合CBAM的BiFPN结构,在参数量降低63%的情况下将检测精度提升至95.80%。王金鹏等^[25]为应对油茶果的严重遮挡与小目标挑战,改进了COF-YOLO v8n模型,通过采用MPDIoU损失函数,增加小目标检测层与SCConv轻量化模块,将mAP提升至99%,并实现了动、静态条件下的果实计数。这些研究通过结构轻量化、引入注意力机制和优化损失函数等策略,有效平衡了检测精度与计算效率,推动了轻量化目标检测算法在农业智能化中的实际应用。

针对苹果在入库时受传送带运行、苹果滚动及设备振动等动态干扰,要求算法兼具高精度单帧检测与实时跟踪能力,以解决目标运动模糊造成的重复计数或跟踪丢失问题。现场仓库环境要求系统必须部署于资源受限、低功耗的边缘设备。基于YOLO v8-DeepSORT的轻量化密集目标计数方法中,YOLO v8能够在复杂背景下精准识别并实现计数效果,同时支持快速响应,适用于需要低延迟的应用场景。在现有的跟踪算法中,基于深度学习的DeepSORT通过融合卷积神经网络特征提取与排序匹配机制,有效解决了复杂场景下的目标关联问题,展现了优异的稳定性和实时性。研究表明,作为多目标跟踪系统的前置环节,目标检测技术的精度直接影响后续跟踪算法的性能。然而,在密集场景中使用时YOLO v8进行计数时,存在精度下降问题,尤其是当物体密集度极高时,模型容易混淆相似的物体,降低识别准确性。而且在对密集计数任务中,需要对更多物体进行识别和定位,这会增加计算负担,影响实时性。

针对以上问题,本文提出一种基于YOLO v8的轻量化改进的密集计数方法。通过GhostNet模块重构主干网络降低模型复杂度;使用CSPHet模块增强密集目标边界区分能力,降低参数量;引入归一化Wasserstein距离损失函数优化定位精度。以期显著提升检测精度与实时性,为解决密集场景下的目标检测与计数问题提供有效方法。

1 轻量化 CGW - YOLO v8 - DeepSORT 网络模型

本文构建的面向苹果密集入库场景的轻量化目标检测与追踪方法整体流程如图 1 所示。首先,进行数据采集和标注,构建适用于检测任务的数据集。随后,将数据输入改进后的 CGW - YOLO v8 检测模型以完成目标识别训练,获取稳定的模型参数。在

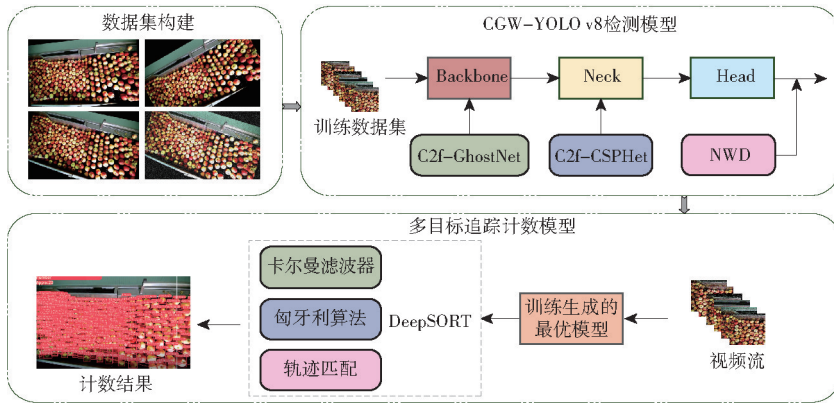


图 1 轻量化目标检测与追踪方法

Fig. 1 Lightweight object detection and tracking methods

层级实现了创新性优化。其 Backbone 网络通过将 C3 模块升级为 C2f 模块来降低模型复杂度,并结合 SPPF 模块的多尺度特征融合能力增强小目标检测性能;Neck 部分在保留 PAN 结构的基础上,通过移除 1×1 降采样层减少特征信息损失,同时针对不同尺度模型动态调整通道维度以平衡计算效率与特征分辨率。Head 部分创新性地引入解耦头结构,采用 Anchor-Free 机制替代传统锚框依赖策略,通过分类与回归任务的物理分离有效提升检测精度。在损失函数设计上,该模型构建了分类任务专用的 VFL loss (Varifocal loss) 与回归任务联合优化的 DFL loss (Distribution focal loss) + CIOU loss (Complete intersection over union) 组合,VFL loss 通过对难易样本赋予不同权重,优化类别分布不均带来的训练偏差,从而提升分类鲁棒性,后者结合概率分布建模与几何约束优化实现精准边界框回归。

1.2 改进后的 CGW - YOLO v8 网络模型

CGW - YOLO v8 模型针对苹果入库计数场景下模型轻量化需求及复杂背景下误判率高的问题进行了相关优化。CGW - YOLO v8 模型结构见图 2,具体改进方案如下:

(1) 重构轻量化主干网络,将主干网络中的 C2f 模块替换为 GhostNet^[26] 轻量化模块架构,通过特征通道重加权机制减少参数量。

(2) 优化特征融合结构,在 Neck 部分采用新型 CSPHet 模块替代原 C2f 结构,利用异构卷积核组合

此基础上,采用 DeepSORT 算法对检测结果进行帧间关联,实现对苹果目标在视频序列中的连续追踪与数量统计。该方法在兼顾模型推理效率的同时,具备一定的目标区分能力,适用于密集场景下的目标检测与计数需求。

1.1 YOLO v8 网络模型

YOLO v8 通过系统整合 YOLO v5、YOLO v6 及 YOLOX 等模型的架构优势,在目标检测框架的多

增强特征表达能力,减少参数量,提升密集排列下苹果识别的鲁棒性。

(3) 改进损失函数,引入基于最优运输理论的 Wasserstein distance loss 替代传统 IoU 度量,通过概率分布匹配策略有效降低密集堆叠场景的误检。

1.3 GhostNet 模块

为提高苹果密集计数时的准确率以及减少模型参数提高实时性,采用 GhostNet 模块对骨干网络中的原生 C2f 结构进行优化重构。GhostNet 通过解耦式双路径特征生成机制实现高效特征表达:主路径采用普通卷积进行高密度语义信息的提炼与抽象,生成紧凑的特征图;辅路径则通过轻量化 Ghost 模块对特征进行空间维度重组与线性变换,生成补偿性细节特征。通过主辅路径的协同融合,GhostNet 能够在减少冗余计算的同时,保留多尺度目标的细粒度表征能力。GhostNet 的引入显著降低了模型的参数量,直接转换为更低的计算资源消耗。优化后的模型内存更小、计算需求更低,使其极适合部署在资源受限的边缘设备上,例如广泛用于农业自动化场景的 NVIDIA Jetson TX2 或类似嵌入式平台。这种轻量化特性有效保障了模型在小型果品分拣线或仓库入口传送带等场景下的实时处理能力,解决了传统大型模型部署成本高、实时性差的问题。

常规卷积操作结构图如图 3 所示,计算公式为

$$Y = f * X$$

$$(X \in \mathbf{R}^{c \times h \times w}, f \in \mathbf{R}^{c \times k \times k \times n}, Y \in \mathbf{R}^{n \times h' \times w'}) \quad (1)$$

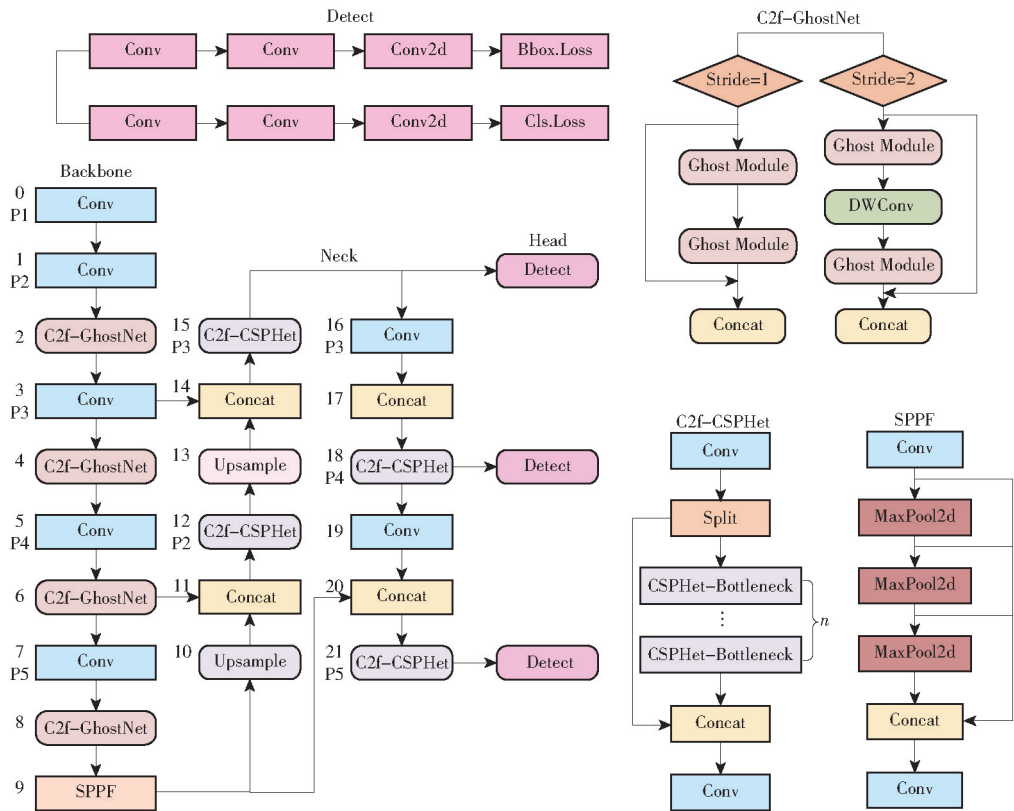


图 2 CGW-YOLO v8 网络结构图

Fig. 2 Structure of CGW-YOLO v8 network

式中 X ——输入特征 f ——卷积核
 $*$ ——卷积运算
 Y ——输出特征
 h, w ——输入图像的高度和宽度
 h', w' ——输出图像的高度和宽度
 c ——通道数
 n ——卷积核数量
 $k \times k$ ——卷积核的内核大小

这个操作的计算量为 $nh'w'cck$ 。

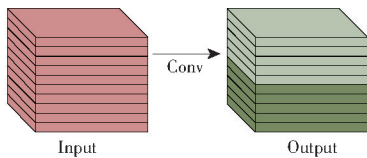


图 3 普通卷积示意图

Fig. 3 Schematic of ordinary convolution

在 GhostNet 操作中,为了简化参数量,减少冗余的特征图,整体特征图生成过程被划分为 2 条路径。首先,主路径对输入特征图进行 1 次标准卷积操作,生成 m 个原始特征图,构成主路径特征表示

$$Y' = f' * X \quad (2)$$

式中, $Y' \in \mathbf{R}^{m \times h' \times w'}$, $f' \in \mathbf{R}^{c \times k \times k \times m}$ 。此时操作计算量为 $mh'w'cck$ 。随后,辅路径对上述 m 个主路径特征图分别进行线性变换操作

$$y_{ij} = \Phi_{i,j}(y'_i) \quad (\forall i=1,2,\dots,m; j=1,2,\dots,s-1) \quad (3)$$

式中, y'_i 为第 i 个原始特征图,经过线性变化 $\Phi_{i,j}$,得到第 j 个幻影特征图 y_{ij} 。共得到 $s-1$ 个幻影特征图, s 为 GhostNet 生成比例,表示每幅主特征图最终对应 s 幅输出特征图。将主路径中的特征图进行恒等映射和廉价操作后的特征图进行拼接,可以得到最后的输出结果 $n = m + (s-1)m$ 。线性卷积核大小为 $d \times d$,操作计算量为 $mhw(s-1)dd$ 。具体结构图见图 4。

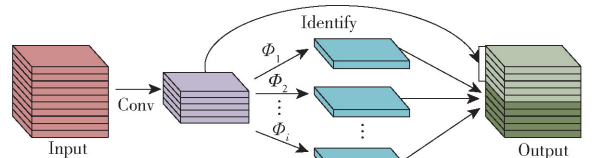


图 4 GhostNet 卷积示意图

Fig. 4 Schematic of GhostNet convolution

GhostNet 卷积和普通卷积的计算量比为 R_s , 计算式为

$$R_s = \frac{nh'w'cck}{mh'w'cck + mhw(s-1)dd} \approx s \quad (4)$$

根据公式(4)可知,普通卷积在生成相同数量输出通道时,其计算量约为 Ghost 模块的 s 倍。为了减少网络结构中的计算量,本文基于 Ghost 卷积构建了 Ghost Bottleneck 结构。这一结构通过堆叠多

个 Ghost 模块,以实现更深层次的特征表示。Ghost Bottleneck 模块包含 2 个阶段:扩展阶段和压缩阶段。在扩展阶段,输入特征图经过一个 Ghost Module 生成较多的通道,以增加网络的特征表达能力;在压缩阶段,另一个 Ghost Module 将通道数恢复到原始水平,以便和分支路径的特征图匹配。具体结构如图 5 所示。

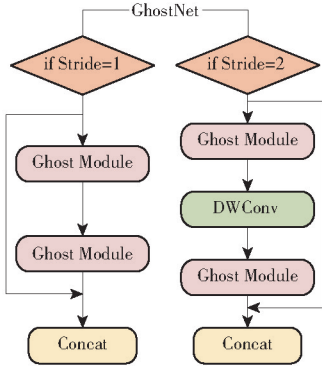


图 5 Ghost Bottleneck 模块示意图

Fig. 5 Schematic of Ghost Bottleneck module

1.4 CSPHet 模块

在 YOLO v8 架构中,传统 Neck 模块在高密度、小间距目标下对于跨实例特征干扰的抑制能力不足,常导致边缘信息弱化与实例可分性下降,从而削弱检测精度。并且该部分计算负载较重,在处理高密度目标场景时可能导致实时性下降。CSPHet 参考了 CSPNet^[27] 的分支设计理念,将输入特征分为两路进行差异化处理,最终通过特征融合实现轻量化与高效推理。主路径通过异构多分支卷积代替普通卷积,其中 3×3 深度卷积侧重于捕获局部边缘细节, 1×1 点卷积侧重于跨通道语义整合,二者协同作用可在目标高度聚集区域显著增强邻近实例边界的梯度区分度。分路径直接传递部分特征以减少重复计算。将两路径进行通道拼接,通过 1×1 卷积调整通道维度。具体结构图可见图 6。CSPHet Bottleneck 由 HetConv^[28] 堆叠组成,HetConv 结合了不同大小的卷积核,部分核为 3×3 ,部分核为 1×1 。由于 1×1 卷积核比 3×3 卷积核需要更少的参数,故相比标准操作能够减少模型的参数数量。

在 HetConv 中,卷积核被分为几部分,参数 P 代表了这些部分的数量。当 $P=2$ 时,意味着有一半的卷积核是 3×3 ,另一半是 1×1 。图 7 展示了不同 P 值之间的流程区别。

在一个普通卷积层中,输入特征图尺寸为 $H \times H \times C$, H 为输入特征图的宽度和高度, C 为输入的通道数。输出特征图为 $W \times W \times C_0$, W 为输出特征图的宽度和高度, C_0 为输出的通道数。使用 N 个 $Q \times Q \times M$ 的卷积核,其中 Q 是卷积核的宽度和高

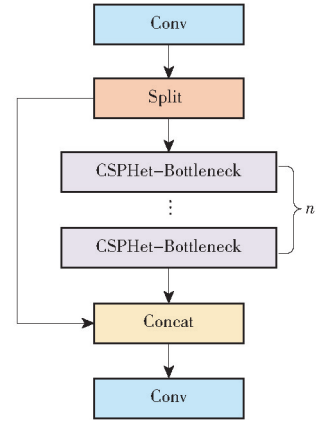


图 6 CSPHet 模块示意图

Fig. 6 Schematic of CSPHet module

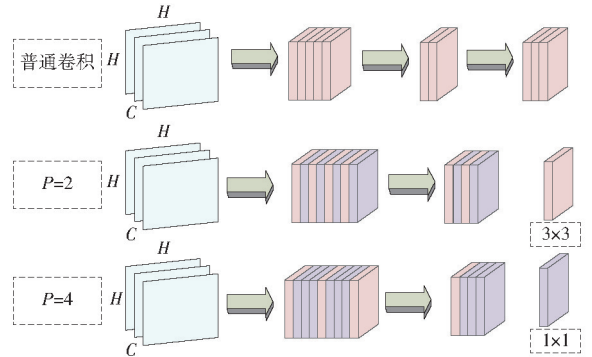


图 7 不同 P 值卷积示意图

Fig. 7 Schematic of convolutions with different P values

度, M 为每个卷积核需要处理的通道数,即输入特征图的通道数。由此普通卷积层总计算量为

$$F_c = WWMNQQ \quad (5)$$

在 HetConv 中,一般采用 1×1 或者 3×3 的卷积核, 3×3 的卷积核部分的计算量 F_0 为

$$F_0 = 3 \times 3 \times WW \frac{M}{P} N \quad (6)$$

1×1 的卷积核部分的计算量 F_1 为

$$F_1 = 1 \times 1 \times WW \frac{M(P-1)}{P} N \quad (7)$$

总计算量 F_H 为

$$F_H = F_0 + F_1 \quad (8)$$

普通卷积和 HetConv 卷积计算量的比值 R 为

$$R = \frac{F_c}{F_H} = \frac{9P}{P+8} \quad (9)$$

由此可以得出,随着 P 的增大,比值 R 会相应增大。在本文中,为了减少参数量,输入通道按 $P=4$ 分组处理, $1/4$ 的卷积核为 3×3 ,剩余的为 1×1 。

1.5 NWD 损失函数

在密集苹果场景中,苹果的分布通常呈现紧密排列或部分重叠,原本的损失函数 CIoU 因仅关注局部差异,易导致相邻苹果的预测框重叠时,难以区分

个体边界。密集堆叠中的错位会被放大为显著误差。Wasserstein distance loss^[29]通过最小化将预测分布“搬运”到真实分布所需的最小成本,实现全局分布对齐。通过最优传输规划容忍苹果可能因机械振动或堆叠压力发生微小位移的几何变化。基于最优传输理论,NWD 采用 Wasserstein 距离来计算 2 个分布之间的距离。对于 2 个二维高斯分布 $\mu_1 = N(m_1, \Sigma_1), \mu_2 = N(m_2, \Sigma_2)$ 之间的二阶 Wasserstein 距离定义为

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \|\Sigma_1^{1/2} - \Sigma_2^{1/2}\|_F^2 \quad (10)$$

式中 $\|\cdot\|_2$ ——向量欧几里得范数

$\|\cdot\|_F$ ——矩阵元素平方和的平方根,简称 Frobenius 范数

假设边界框 $A = (x_a, y_a, w_a, h_a), B = (x_b, y_b, w_b, h_b), (x_a, y_a), (x_b, y_b)$ 代表边界框中心坐标, w_a, w_b 为边界框宽度, h_a, h_b 为边界框高度。则 N_a 和 N_b 为边界框对应建模的高斯分布,公式可进一步简化为

$$W_2^2(N_a, N_b) = \left\| \left(x_a, y_a, \frac{w_a}{2}, \frac{h_a}{2} \right)^T, \left(x_b, y_b, \frac{w_b}{2}, \frac{h_b}{2} \right)^T \right\|_2^2 \quad (11)$$

式中, $W_2^2(N_a, N_b)$ 是距离单位,不能直接用作相似度度量,因此,使用它的指数形式归一化并获得称为归一化 Wasserstein 距离^[30](NWD)的新度量

$$\text{NWD}(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{Z}\right) \quad (12)$$

式中, Z 是与数据集密切相关的常数。引入 IoU - Loss 是为了消除训练和测试之间的性能差距。然而, IoU - Loss 在以下 2 种情况下无法为优化网络提供梯度: 预测的边界框 K 和真实框 G 之间没有重叠或框 K 完全包含框 G , 反之亦然。尽管 CIoU 和 DIoU 可以处理上述 2 种情况,但它们对小物体的位置偏差很敏感,它们都是基于 IoU 的。为了处理上述问题,将 NWD 度量设计为损失函数

$$L_{\text{NWD}} = 1 - \text{NWD}(N_k, N_g) \quad (13)$$

式中 N_k ——预测框 K 的高斯分布模型

N_g ——真实标注框 G 的高斯分布模型

1.6 DeepSORT 算法

DeepSORT^[31]是一种基于检测结果的多目标追踪算法,在传统 SORT 算法的基础上引入深度外观特征嵌入,以提升目标在遮挡、重识别等复杂场景下的关联准确性。该算法主要由 2 部分组成: 卡尔曼滤波器用于状态预测,匈牙利算法用于帧间数据关联。在目标匹配过程中 DeepSORT 同时考虑了运动信息(位置、速度)和外观特征的余弦相似度,从而

有效缓解了因目标外观相似或短时遮挡导致的目标身份切换问题。凭借较高的实时性与鲁棒性, DeepSORT 广泛应用于视频监控、交通分析及工业检测等多目标追踪任务中。

为实现苹果在入库过程中的自动计数,本文首先利用改进后的轻量化 CGW - YOLO v8 模型对视频序列中的每帧图像进行目标检测,提取苹果目标的边界框及其深度特征信息。随后,采用 DeepSORT 追踪算法对检测结果进行帧间关联;通过卡尔曼滤波器对目标的运动状态(包括位置、速度等)进行预测,并使用匈牙利算法将当前帧的检测结果与上一帧的预测轨迹进行最优匹配,从而实现目标的连续追踪。此外,本文设计了基于轨迹管理的计数策略,为每个目标分配唯一标识符(ID),并使用轨迹列表持续记录其运动过程。在目标因遮挡或移出画面而短暂消失时,系统将暂存其轨迹,并在后续帧中尝试重新匹配关联,以确保轨迹连续性与计数准确性。通过引入虚拟计数线,仅在目标首次越过计数线时进行计数,有效避免重复统计与漏计问题。该方法在应对苹果密集排列、部分遮挡等复杂入库场景方面表现出较强的鲁棒性,能够为果品入库管理提供稳定的数量监测支持。

2 实验设置

2.1 训练环境

使用的训练环境包括基于 NVIDIA RTX 4090 显卡和 12 个 vCPU 的 Intel Xeon Platinum 8352V 处理器的高性能计算平台,能够高效支持大规模神经网络训练。软件则采用 YOLO v8 作为目标检测框架,基于 PyTorch 2.0.0 和 Python 3.8 进行开发,利用 CUDA 11.8 版本进行 GPU 加速,确保了训练过程的高效性。

2.2 数据采集与预处理

在某苹果自动化仓库运输带上方放置摄像头(HMA201-S1型)来收集构建密集苹果计数数据集。在不同时间段采集视频数据,包括清晨、正午、黄昏及夜间人工照明情况,收集了 1 000 幅高质量苹果图像,全面捕捉苹果在不同光照下的外观特征,如色泽、纹理。为增强数据集多样性和规模,采用旋转、缩放、翻转等数据增强技术,将图像扩展至 3 000 幅,具体情况见图 8、9。此数据集不仅增加了训练样本量,还提升了模型对环境变化的鲁棒性,有助于学习到更加泛化的特征表示,减少过拟合风险。

2.3 参数设置

本研究涉及的所有目标检测网络均使用统一数

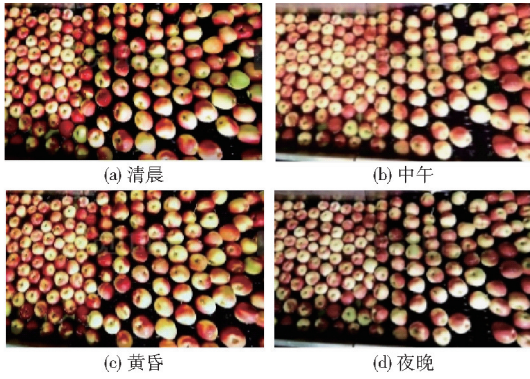


图8 不同光照下的数据图例

Fig. 8 Data under different lighting

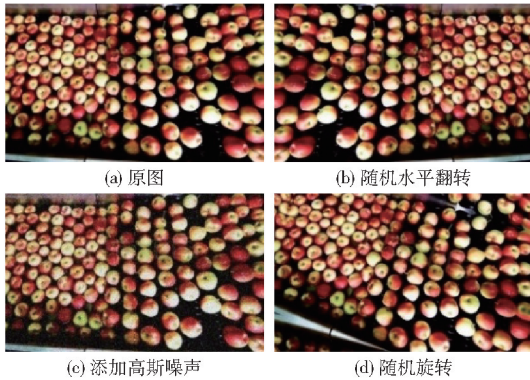


图9 数据增强图例

Fig. 9 Data augmentation

数据集,数据集划分为训练集、验证集、测试集,比例为7:2:1。训练过程中,训练轮数为300轮,初始学习率设置为0.01,并通过学习率衰减因子将其调整至0.0001,SGD的动量设定为0.937,以加速收敛并提高训练稳定性。此外,为防止过拟合,使用了0.0005的权重衰减。在训练初期,设置了3个epoch的warmup阶段,其间初始动量为0.8,偏置学习率为0.1,以帮助模型逐步适应学习率的调整。该训练环境的配置保证了模型在高效的计算资源下稳定收敛,完成大规模图像识别任务。

表1 不同检测模型性能对比

Tab. 1 Comparison of different detection models

模型	精确率/%	召回率/%	mAP@0.5/%	参数量	内存占用量/MB	浮点数运算量
YOLO v5n	93.8	91.3	94.4	2.5×10^6	5.0	7.2×10^9
YOLO v6n	95.5	93.3	96.0	4.2×10^6	8.3	1.19×10^{10}
YOLO v7	94.5	93.9	93.2	3.71×10^7	71.3	1.051×10^{11}
YOLO v7-tiny	94.2	93.8	93.7	6.0×10^6	46.3	1.32×10^{10}
YOLO v8n	94.8	93.5	94.8	3.0×10^6	5.9	8.2×10^9
CGW-YOLO v8-DeepSORT(本文模型)	95.9	94.8	95.8	2.2×10^6	4.6	6.3×10^9

由表1可知,相较于基线模型YOLO v8n,本模型在保持最低参数量 2.2×10^6 、最小模型内存占用量4.6 MB及最低浮点数运算量 6.3×10^9 的前提下,实现了检测精度的显著提升:精确率达95.9%,

2.4 评价指标

在目标检测模型的性能评估中,本文采用多维评价体系以全面衡量模型效率与检测精度。效率指标方面,使用通过参数量(Params)、浮点数运算量(GFLOPs)和模型内存占用量(Model size)量化模型的复杂性与计算需求。精度指标方面,使用精确率(Precision)与召回率(Recall)分别评价模型检测可靠性及目标覆盖能力,平均精度均值(mAP@0.5)作为核心综合指标,以交并比(IoU)阈值为0.5时的检测结果为基础,融合分类置信度与定位精度,反映模型在密集场景下的鲁棒性。针对苹果静态识别任务,引入平均绝对误差(MAE)与均方根误差(RMSE)作为数量级误差指标,直接衡量预测数量与真实值的偏差。针对动态苹果计数任务,本文选取了多项指标对模型性能进行评估,包括多目标跟踪准确度(Multiple object tracking accuracy, MOTA)、多目标跟踪精度(Multiple object tracking precision, MOTP)、高级目标检测精度(Higher order tracking accuracy, HOTA)、定位准确率(Localization accuracy, LocA)以及帧率(Frames per second, FPS)。其中,MOTA综合反映了漏检、误检与身份保持的整体表现,MOTP和LocA用于衡量目标空间定位的精度,HOTA则在检测与关联之间提供更加平衡的评估。而FPS反映了模型的实时处理能力,满足农业自动化场景中对目标统计精度的需求。

3 结果分析

3.1 模型对比实验

为了验证本文改进算法对密集苹果检测的有效性,与其他主流算法进行对比,包括YOLO v5n、YOLO v6n、YOLO v7、YOLO v7-tiny、YOLO v8n。训练过程中,所有算法采用相同的实验设备、数据集和参数设置,结果如表1所示。

召回率提升至94.8%,mAP@0.5达到95.8%。YOLO v6n的mAP@0.5虽达96.0%,其浮点数运算量为 1.19×10^{10} ,是本模型的1.89倍,其参数量也比本模型高 2×10^6 。难以满足边缘端实时性需求。

以上表明,本模型通过结构优化实现了精度与效率的帕累托最优,为资源受限场景下的高精度目标检测提供了可行性方案。

3.2 消融实验

本研究采用 YOLO v8 模型作为基础框架,在控

制实验环境变量一致的前提下,利用自建数据集执行了一系列消融实验,旨在系统地分析 GhostNet、CSPHet 及 NWD 损失函数对模型性能的优化作用及其有效性。结果如表 2 所示。

基准模型在密集检测任务中 mAP@0.5 为

表 2 消融实验结果

Tab.2 Dissolution experiment results

GhostNet	CSPHet	NWD	精确率/%	召回率/%	mAP@0.5/%	参数量	内存占用量/MB	浮点数运算量
×	×	×	94.8	93.5	94.8	3.0×10^6	5.9	8.2×10^9
√	×	×	94.4	92.9	94.3	2.5×10^6	5.1	7.0×10^9
×	√	×	94.6	92.9	94.8	2.7×10^6	5.4	7.6×10^9
×	×	√	94.4	93.6	94.4	3.0×10^6	5.9	8.2×10^9
√	√	×	94.7	93.0	95.4	2.2×10^6	4.6	6.3×10^9
√	×	√	95.9	94.2	96.1	2.5×10^6	5.1	7.0×10^9
×	√	√	94.5	93.8	94.9	2.7×10^6	5.4	7.6×10^9
√	√	√	95.9	94.8	95.8	2.2×10^6	4.6	6.3×10^9

注:√表示使用该模块,×表示未使用该模块。

94.8%,但需 3.0×10^6 的参数量与 5.9 MB 模型内存占用量。单独引入 GhostNet 模块后,通过特征解耦重构策略将参数量降低 16.7%,浮点数运算量下降 14.6%,模型内存占用量下降 13.5%,虽导致 mAP@0.5 下降 0.5 个百分点,但其生成的轻量化特征为后续模块协同优化奠定基础。当独立采用 CSPHet 模块时,跨阶段特征压缩机制在保持 mAP@0.5 基准值的同时参数量减少 10%,浮点数运算量下降 4.8%,模型内存占用量下降 8.4%,验证了其高效特征融合能力。GhostNet 与 CSPHet 联合使用使 mAP@0.5 提升至 95.4%,参数量进一步压缩了 26.6%,模型内存占用量和浮点数运算量分别下降 22% 和 23.2%。证明轻量化特征生成与跨阶段压缩的互补性。而 NWD 模块的引入使召回率显著提升 1.8 个百分点,精确率提升 1.2 个百分点。表明其基于归一化 Wasserstein 距离的匹配策略有效缓解了密集目标的漏检问题。GhostNet 与 NWD

组合的精确率和 mAP@0.5 均实现了最佳单指标性能,表明特征解耦可增强分布匹配的鲁棒性;三者全集成时模型达到最优均衡状态,在参数量最低的情况下,mAP@0.5 对比基准模型提升 1.0 个百分点,达到 95.8%;精确率提升 1.1 个百分点,达到 95.9%;召回率提升 1.3 个百分点,达到 94.8%。该结果表明,GhostNet 的冗余特征消除、CSPHet 的跨阶段特征融合及 NWD 的分布对齐机制形成三重优化路径,共同推动模型在精度-效率权衡边界上扩展,为密集目标检测任务提供了高精度与低资源消耗并重的解决方案。

3.3 静态检测实验

为了检验不同 YOLO 模型在不同苹果数量 (203、230、260) 下的检测性能,进行了静态检测对比实验。并聚焦检测结果、漏检与误检的核心数据,系统分析模型结构优化对检测精度的影响。具体数据如表 3 所示。

表 3 静态检测实验结果

Tab.3 Static detection experiment results

模型	苹果数量 203			苹果数量 230			苹果数量 260			平均绝对误差	均方根误差
	检出数	漏检数	误检数	检出数	漏检数	误检数	检出数	漏检数	误检数		
YOLO v5n	213	20	30	227	16	13	251	26	17	20.33	21.17
YOLO v6n	180	23	24	223	13	10	247	28	15	18.83	19.93
YOLO v7	210	16	23	234	18	22	249	23	12	19.00	19.43
YOLO v7-tiny	209	21	27	231	25	26	244	30	14	23.83	24.38
YOLO v8n	208	20	25	230	16	16	252	25	17	19.83	20.21
YOLO v8n + GhostNet	233	48	36	231	17	18	239	30	9	26.33	29.43
YOLO v8n + CSPHet	220	23	40	243	20	33	255	26	21	27.17	28.09
YOLO v8n + NWD	210	23	30	233	14	17	251	25	16	20.83	21.58
YOLO v8n + GhostNet + CSPHet	223	41	21	241	15	26	252	27	19	24.83	26.18
YOLO v8n + GhostNet + NWD	204	22	23	227	11	8	245	25	10	16.50	17.90
YOLO v8n + CSPHet + NWD	216	20	33	238	22	30	258	29	27	26.83	27.21
YOLO v8n + GhostNet + CSPHet + NWD	208	19	24	231	9	10	243	23	6	15.17	16.75

实验数据显示,引入 GhostNet、CSPHet 与 NWD 三重模块组合的模型在综合性能上显著优于其他模型。如平均绝对误差和均方根误差均为最低。本模型在苹果数量 260 时的漏检数与误检数分别为 23 和 6,均为所有模型中最低,体现出良好的边界区分能力与遮挡适应性。且平均漏检和误检数量较基准模型 YOLO v8n 分别降低 23.5% 与 17.1%,验证了其通过模块协同优化实现的高鲁棒性。对比其他模型,单一模块的性能均存在局限性:例如, YOLO v5n、YOLO v6n,在 3 个不同苹果数量下的漏检和误检数量均高于改进后模型。YOLO v7n 虽单个指标存在优势,但在整体上仍低于改进后的模型。而单独使用 GhostNet 的 YOLO v8n + GhostNet 在苹果数量为 203 时漏检数高达 48 个,而 YOLO v8n + CSPHet 在苹果数量为 230 时误检数达 33 个,表明单一模块易过度轻量化而丢失准确性;相比之下,三重模块组合通过 GhostNet 的冗余特征消除、CSPHet 的跨阶段融合及 NWD 的分布对齐机制,显著抑制了复杂场景下的漏检与误检。相较之下,轻量化模型如 YOLO v7-tiny 在苹果数量增加时性能显著下降,而基准模型 YOLO v8n 虽检测出的苹果数量结果较高,但其误检数与漏检数却高于三重模块组合,进一步证明多模块协同设计对提升密集目标检测稳定性的必要性。综上, YOLO v8n + GhostNet + CSPHet + NWD 通过结构优化与模块互补性,在保持较高检测结果的同时,实现了漏检与误检的双重抑制,为高精度轻量化检测任务提供了更具竞争力的解决方案。

为直观展示静态检测效果并验证所提方法的有效性,选取经数据增强处理的 3 种不同数量的苹果图像,分别评估漏检与误检情况。其中,绿色框表示

检测正确的目标,红色框则表示漏检或误检的目标。由图 10 可以看出,在 3 种不同数量条件下,所提出的 CGW-YOLO v8 模型相较于原始模型均表现出更优的检测性能。即使在模型体积大幅压缩的情况下,CGW-YOLO v8 仍能有效缓解特征混淆带来的漏检与误检问题,并显著减少对边缘模糊目标的识别错误,表现出良好的鲁棒性与实用性。

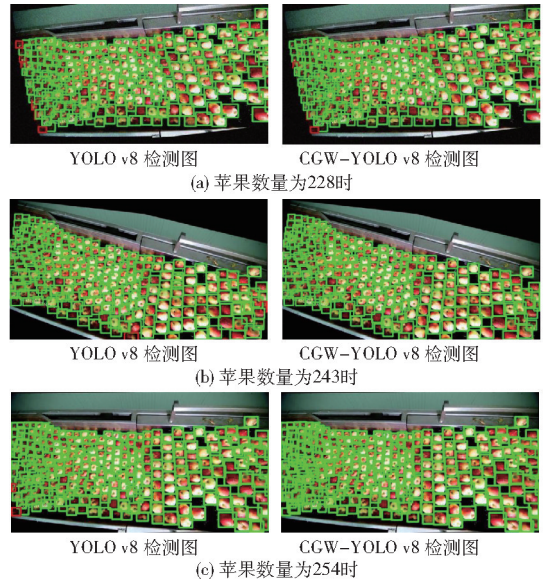


图 10 部分检测结果对比图

Fig. 10 Comparison charts of some test results

3.4 计数实验

为对比不同模型在多目标追踪计数任务中的性能,本文选取标注完成的 6 段视频样本,分别在 YOLO v5n - DeepSORT、YOLO v6n - DeepSORT、YOLO v8n - DeepSORT 以及本文提出的 CGW-YOLO v8n - DeepSORT 模型上进行测试。各模型在相同条件下完成目标检测与追踪任务,并依据统一评估指标进行性能对比。实验结果如表 4 所示。

表 4 多目标追踪评价指标对比实验结果

Tab. 4 Comparative experiments of evaluation metrics for multi-target tracking

模型	HOTA/%	LocA/%	MOTA/%	MOTP/%	帧率/(f·s ⁻¹)
YOLO v5n - DeepSORT	92.7	79.7	89.2	80.1	10
YOLO v6n - DeepSORT	93.0	78.8	89.6	79.3	7
YOLO v8n - DeepSORT	91.9	79.7	88.2	79.9	9
CGW-YOLO v8 - DeepSORT	91.9	79.1	88.3	79.6	13

从精度指标分析可见,本文模型 HOTA 和 MOTA 与 YOLO v8n 模型表现持平,分别达到 91.9% 和 88.3%,表明其在检测与目标关联方面具备良好的稳定性与鲁棒性。同时,MOTP 相较 YOLO v6n 略有提升,说明模型在空间定位精度方面具有一定优势。尽管 YOLO v5n 与 YOLO v6n 的 HOTA 和 MOTA 略有领先,但本模型的 LocA 优于 YOLO v6n,反映出更强的目标定位一致性。

在实时性方面,所提出模型在 Intel Core i7-12700H 处理器(CPU 环境)下推理速度为 13 f/s,显著优于 YOLO v6n(7 f/s)和 YOLO v5n(10 f/s),表现出更高的推理效率和应用可行性。该优势使得本文模型在农业自动化等对运算资源与处理时延敏感的实际场景中具有更强的部署潜力。图 11 展示了模型在不同密度条件下进行苹果入库计数的实际效果。从图 11 中可以看出,当摄像头聚焦范围所涵盖

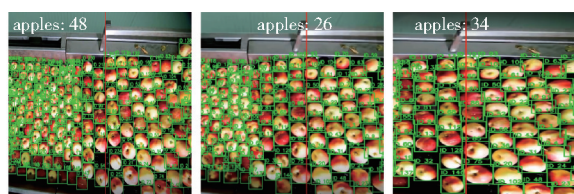


图 11 苹果计数可视化

Fig. 11 Visualization of apple counts

的部分越多,即目标密度升高时,误检现象有所加剧,表明模型在处理高密度目标时存在一定的识别误差。针对这一问题,未来将进一步优化模型,以提高其在高密度复杂场景中的检测精度和鲁棒性。总体而言,CGW - YOLO v8 在轻量化设计的基础上,兼顾了检测精度与效率,显著增强了在密集场景下的鲁棒性,为入库盘点计数的自动化管理提供了更加可靠的解决方案。

4 结论

(1) 针对苹果密集入库导致的检测精度与效率问题,提出一种基于 YOLO v8 的轻量化改进算法。实验表明,本模型平均精度均值 $mAP@0.5$ 提升至 95.8%,较原模型提升 1.0 个百分点,精确率和召回

率较原模型分别提升 1.1、1.3 个百分点。参数量与运算量较基准模型分别减少 24.4% 和 23.2%。

(2) 采用 GhostNet 模块替代 C2f 结构,通过特征解耦与通道重加权机制,在参数量减少 16.7% 的同时,保留密集目标的细粒度特征,显著降低模型计算量。

(3) 使用 CSPHet 模块,融合 3×3 深度卷积的局部纹理提取以及 1×1 点卷积的全局语义关联能力,增强密集目标的边界判别精度,在参数量降低 10% 的情况下, $mAP@0.5$ 达到 95.4%。

(4) 引入 NWD 损失函数,通过概率分布匹配策略解决密集苹果的边界模糊问题,精确率和召回率分别提升 1.2、1.8 个百分点。

(5) 将 CGW - YOLO v8 目标检测模型与 DeepSORT 目标跟踪算法相结合,实现了对苹果入库计数的实时检测与追踪功能。实验结果显示,该方法 HOTA 达到 91.9%,LocA 为 79.1%,MOTA 达到 88.3%,MOTP 为 79.6%,帧率为 13 f/s,验证了其在苹果入库计数任务中的高精度、良好的目标关联能力及实时性,也证明其在复杂场景中的有效性和应用潜力。

参 考 文 献

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 - 587.
- [2] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//Computer Vision-ECCV 2014. Cham: Springer, 2014: 346 - 361.
- [3] GIRSHICK R. Fast R - CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440 - 1448.
- [4] 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, 2017, 39(6): 1137 - 1149.
- [5] SHAHIN S, SADEGHIAN R, SAREH S. Faster R - CNN based decision making in a novel adaptive dual-mode robotic anchoring system[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an: IEEE, 2021: 11010 - 11016.
- [6] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627 - 1645.
- [7] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: integrated recognition, localization and detection using convolutional networks[J]. arXiv Preprint, arXiv: 1312.6229, 2013.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779 - 788.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517 - 6525.
- [10] FARHADI A, REDMON J. YOLOv3: an incremental improvement[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 1 - 6
- [11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 1 - 17.
- [12] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv Preprint, arXiv:2209.02976, 2022.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464 - 7475.
- [14] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer

Vision. Cham: Springer International Publishing, 2016: 21–37.

- [15] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2999–3007.
- [16] 奚小波,丁杰源,翁小祥,等. 基于轻量化 YOLO v5s – MCA 的番茄成熟度检测方法[J]. 农业机械学报,2025,56(3): 383–391,436.
XI Xiaobo, DING Jieyuan, WENG Xiaoxiang, et al. Tomato maturity detection method based on lightweight YOLO v5s – MCA [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 383–391, 436. (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 YOLOv5s[J]. Computers and Electronics in Agriculture, 2023, 206: 107662.
- [18] 刘子洋,徐慧英,朱信忠,等. Bi – YOLO:一种基于 YOLOv8n 改进的轻量化目标检测算法[J]. 计算机工程与科学,2024, 46(8):1444–1454.
LIU Ziyang, XU Huiying, ZHU Xinzong, et al. Bi – YOLO: a lightweight object detection algorithm based on YOLOv8n[J]. Computer Engineering & Science, 2024, 46(8): 1444–1454. (in Chinese)
- [19] 欧阳玉旋,张荣芬,刘宇红,等. 基于 Ghost 高效层聚合网络的多尺度融合 YOLOv7 – tiny 改进算法[J]. 计算机工程与科学,2025,47(8):1437–1448.
OUYANG Yuxuan, ZHANG Rongfen, LIU Yuhong, et al. An improved multi-scale fusion YOLOv7 – tiny algorithm based on Ghost efficient layer aggregation network[J]. Computer Engineering and Science, 2025, 47(8): 1437–1448. (in Chinese)
- [20] SHANG Y Y, GENG M Y, FANG H Y, et al. Using unmanned aerial vehicle acquired RGB images and density-cluster-count model for tree-level apple flower quantification[J]. Computers and Electronics in Agriculture, 2024, 226: 109389.
- [21] WU Z C, SUN X M, JIANG H H, et al. NDMFCS: an automatic fruit counting system in modern apple orchard using abatement of abnormal fruit detection[J]. Computers and Electronics in Agriculture, 2023, 211:108036.
- [22] 张传栋,高鹏,亓璐,等. 基于 SAW – YOLO v8n 的葡萄幼果轻量化检测方法[J]. 农业机械学报,2024,55(10):286–294.
ZHANG Chuandong, GAO Peng, QI Lu, et al. Lightweight detection method for grape cluster fruits based on SAW – YOLO v8n[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(10): 286–294. (in Chinese)
- [23] 冀汶莉,刘洲,邢海花. 基于 YOLO v5 的农田杂草识别轻量化方法研究[J]. 农业机械学报,2024,55(1):212–222,293.
JI Wenli, LIU Zhou, XING Haihua. Lightweight method for identifying farmland weeds based on YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 212–222, 293. (in Chinese)
- [24] 邓铭辉,龚俊杰,郑飘逸,等. 基于改进 YOLOX 的群养生猪轻量化目标检测方法[J]. 农业机械学报,2023,54(11): 277–285.
DENG Minghui, GONG Junjie, ZHENG Piaoyi, et al. Lightweight target detection method for group-raised pigs based on improved YOLOX[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(11): 277–285. (in Chinese)
- [25] 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)
- [26] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580–1589.
- [27] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 390–391.
- [28] SINGH P, VERMA V K, RAI P, et al. HetConv: beyond homogeneous convolution kernels for deep CNNs[J]. International Journal of Computer Vision, 2020, 128(8): 2068–2088.
- [29] SHEN J, QU Y, ZHANG W, et al. Wasserstein distance guided representation learning for domain adaptation[C]// Proceedings of the 2018 AAAI Conference on Artificial Intelligence. New York: AAAI, 2018: 4058–4065.
- [30] WANG J, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. arXiv Preprint, arXiv:2110.13389, 2021.
- [31] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3645–3649.