

基于改进 RT-DETR 的草莓叶枯病实时检测

林 晨¹ 钟雅露¹ 孔彦琪¹ 陈仁凡¹ 谢 知^{1,2}
LIN Chen ZHONG Yalu KONG Yanqi CHEN Renfan XIE Zhi

摘 要

为实现草莓叶枯病的快速实时诊断,文章以端到端的实时目标检测模型 RT-DETR 为基础架构,提出一种改进型 MNV2IF-DETR 轻量级草莓叶枯病识别方法。使用 MobileNetV2 替换原 RT-DETR 骨干网络的 ResNet,最优特征子集选择 MobileNetV2 网络架构的 $\{S_2, S_4, S_6\}$ 特征输出层组合,在丰富特征提取能力的同时,使模型轻量化。通过把原架构的普通卷积替换为 DW 深度卷积,并添加 SE 注意力机制,构建改进的 ImFuze 特征融合模块,大幅降低计算量,同时减少冗余特征,提高编码器对关键特征的响应能力与提取效率。试验结果表明,相较于原模型,改进的 MNV2IF-DETR 模型的 mAP 达 90.3%,准确率提高 0.5%,模型大小仅为 22.1 MB,减小 48.4%,另外,参数量和浮点计算量也分别下降 31.7% 和 67.6%,在 GPU 上的单张平均检测时间仅 14.2 ms,速度提高 28.3%。模型综合性能优于其他对比检测算法,可以为草莓叶枯病害防治领域的技术发展提供参考。

关键词

草莓叶枯病;病害诊断;RT-DETR;目标检测

doi: 10.3969/j.issn.1672-9528.2025.01.018

0 引言

草莓作为极具经济价值的水果品类,在农业生产中占据重要地位。然而近年来,叶枯病连年肆虐,致使草莓果实品质下滑,产量锐减^[1]。在草莓刚感染叶枯病早期阶段,通过实现精确诊断,进而施用适配的杀菌剂,对提高草莓产量起着至关重要的作用^[2]。

近年来,作物的实时目标检测取得显著成效,能够实时检测对象的类别和位置,大幅度提高了其生产和作业的效率。DETR 模型以端到端的方式直接输出预测的对象框集合,简化了目标检测的整体流程^[3]。诸多研究通过骨干修改,查询机制设计和注意力细化等方式,得到了 Deformable-DETR^[4]、Conditional DETR^[5]、DAB-DETR^[6]、RT-DETR 等改进 DETR 模型。在农业应用领域,Wu 等人^[7]提出的病害分割检测器 DS-DETR,高效分割番茄早疫病和晚疫病造成的病斑,疾病分级准确率达到 96.4%。Zhao 等人^[8]通过解耦多尺度特征的尺度内交互和跨尺度融合,构造了首个实时端到端目标检测 RT-DETR。

本文在 RT-DETR 基础上使用 MobileNetV2^[9] 替换原始骨干网络 ResNet,通过实验测试选择 MobileNetV2 的最佳特征

输出层;优化特征融合模块,提高编码器的效率和关键特征捕捉能力,增强模型的检测性能和实时处理能力。通过这些改进,实现模型的轻量化,并提升其检测效率。

1 材料与方法

1.1 草莓叶枯病数据集的构建

本文数据图片来自公开数据集 plantvillage^[10],使用软件 X-AnyLabeling 进行标注处理,数据集图片示例如图 1。数据集制作过程中采用数据增强技术,包括随机采用裁剪、翻转、扩展、颜色扭曲,以及高斯模糊等操作,在数据量有限或类别不平衡的情况下,可以提升模型泛化能力。经过数据增强环节,重新将草莓叶枯病病害数据集分为训练集、验证集、测试集,样本数分别为 1423、630、277 张。

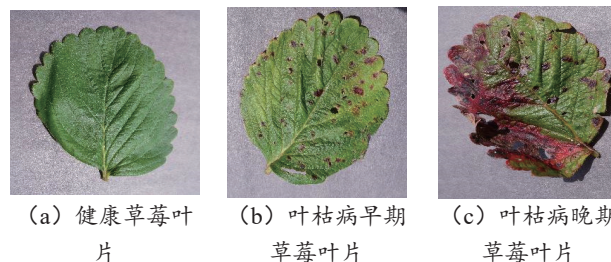


图 1 草莓叶片样本示例

1.2 模型的选取与改进

1.2.1 RT-DETR 目标检测模型

本文采用端到端实时目标检测模型 RT-DETR,其

1. 福建农林大学 福建福州 350002

2. 福建省农业信息感知技术重点实验室 福建福州 350002

[基金项目]福建省自然科学基金项目(2019J01403);福建农林大学科技创新专项基金项目(KFB23165A)

包含骨干网络 ResNet、高效混合编码器 (efficient hybrid encoder)、配有辅助预测头的 Transformer 解码器 (RTDETRDecoder)，总体模型架构如图 2 所示。骨干网络由一系列卷积层 (convolutional layers) 和残差块 (residual blocks) 组成；高效混合编码器由基于注意力的尺度内特征交互模块 (attention-based intrascale feature interaction, AIFI) 与跨尺度特征融合模块 (cross-scale feature-fusion module, CCFM) 组成，其中 CCFM 通过侧连接卷积、上采样、下采样及特征融合等步骤，融合高层与低层特征，再完成最终特征的组合，从而将相邻的特征融合成新的特征，提升模型对不同尺度目标的检测能力；Transformer 解码器通过变形注意力模块 (deformable attention modules) 使模型获得更强的目标定位能力与抗背景干扰能力，且支持灵活调用不同解码器层以调整推理速度，有助于检测器的实时应用。

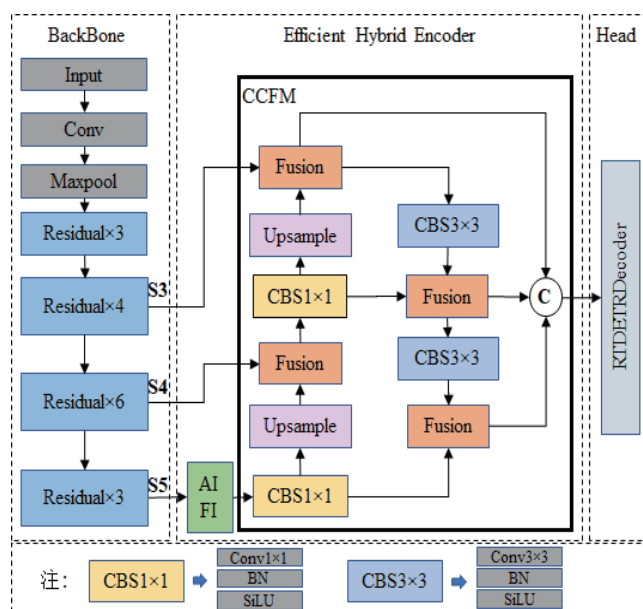


图 2 RT-DETR 模型架构

1.2.2 RT-DETR 网络架构的改进

为实现网络轻量化，本文选用 MobileNetV2 替换 RT-DETR 的原骨干网络 ResNet，MobileNetV2 网络结构使用了特殊的深度可分离卷积 (depthwise separable convolution, DSC)，包括深度卷积 (depthwise convolution, DW) 与逐点卷积 (pointwise convolution, PW) 两个步骤。

DSC 计算量约为普通卷积的 1/9，DSC 与普通卷积理论计算公式分别为：

$$\text{DSC 卷积: } D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \quad (1)$$

$$\text{普通卷积: } D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \quad (2)$$

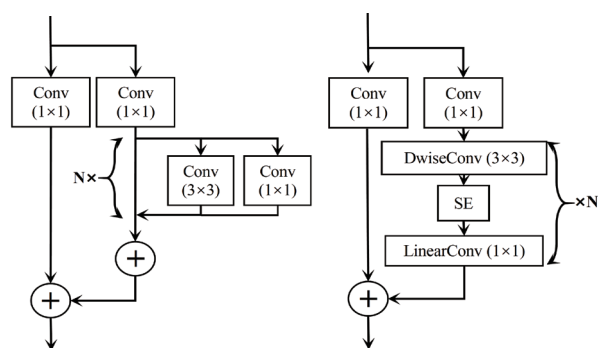
式中： D_K 为卷积核的大小； M 为输入特征矩阵的 channel； N 为输出特征矩阵的 channel； D_F 为输入特征矩阵的宽高。

若卷积核的大小取 3，计算量比较所得式为：

$$\frac{1}{N} + \frac{1}{D_K \cdot D_K} = \frac{1}{N} + \frac{1}{9} \quad (3)$$

相较于原骨干网络使用的残差架构，MobileNetV2 引入倒置残差结构。首先利用扩展层来扩展输入矩阵维度，因为在特征数据提取时低维矩阵的利用率远低于高维矩阵，所以通过扩展层来保证拥有足够多的特征信息；然后利用 DW 深度卷积提取特征，延缓维度扩展引起的计算量陡增；最后通过压缩层 PW 逐点卷积来压缩数据，避免网络结构无节制地扩展。

为优化 RT-DETR 的混合编码器，本文构建改进的特征融合模块 ImFuze，以提高编码器的性能和效率，与原特征融合模块 Fusion 的对比如图 3 所示。



(a) Fusion 特征融合

(b) ImFuze 改进特征融合

图 3 改进特征融合模块与原特征融合模块

首先将原架构的普通 3×3 卷积替换为 DW 深度卷积，大幅减少计算量；然后添加 SE (squeeze-and-excitation) 注意力机制，减少冗余特征，提升编码器对关键特征的响应能力与提取效率；最后通过 PW 逐点卷积降低特征维度，在保持特征图空间尺寸不变的情况下，省略叠加步骤。

既往研究表明 RT-DETR 可通过减少低层的特征融合，减少融合消耗时间，提高准确率^[7]。本文根据实验数据的测试分析 (见 2.3 节)，选取 $\{S_2, S_4, S_6\}$ 层，再通过多尺度解耦并进行交互处理多尺度特征获得最佳性能，具体公式定义为：

$$Q = K = V = \text{Flatten}(S_6) \quad (4)$$

$$F_6 = \text{Reshape}(\text{Attn}(Q, K, V)) \quad (5)$$

$$\text{Output} = \text{CCFM}(\{S_2, S_4, S_6\}) \quad (6)$$

结合 MobileNetV2 和 ImFuze 改进得到的 MNV2IF-DETR 模型架构见图 4。利用骨干网络中 S_6 的输出特征作为 AIFI 的输入，高效混合编码器通过 CCFM 模块将 AIFI 的处理结果与 S_2 、 S_4 输出特征交互，然后转换为图像特征序列。最后，采用感知查询选择从编码器输出序列中选取固定数量的图像特征作为解码器的初始对象查询。

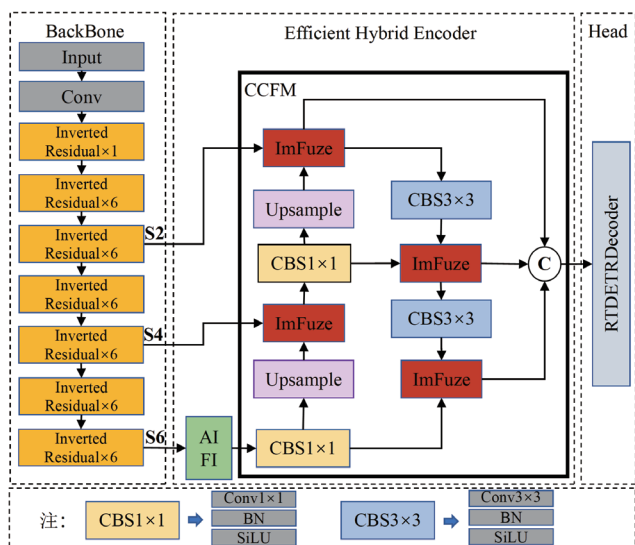


图4 MNV2IF-DETR 网络架构

2 结果与分析

2.1 评估指标

本文采用准确率（Accuracy）、精确度（Precision）、召回率（Recall）、平均精度（average precision, AP）评估模型性能。准确率、精确度、召回率、平均精度的定义公式分别为：

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (10)$$

式中：TP 和 TN 分别表示标记为正确识别的正负样本数量；FP 表示标记为错误识别的正样本数量；FN 表示标记为错误识别的负样本数量；平均精度 AP 表示 PR（precision-recall）曲线下的面积；PR 曲线表示精度与召回率的关系。本文采用 7 种精确回归曲线参数对模型的表现进行细分评估，具体如图 5（d）所示，其中 C75 和 C50 分别表示在 IOU 阈值为 75% 和 50% 时的性能表现；Loc 代表定位错误（localization error）；Sim 表示相似类别错误（similar class error）；Oth 表示其他类别错误（other class error）；BG 表示背景错误（background error）；FN 则表示漏检错误（false negative error）。计算所有类别得到的平均精度均值 mAP（mean average precision）。

2.2 最优特征子集设定

最优特征子集是影响检测模型计算资源和时间消耗的重要因素。高效的特征提取方法既可以消除低层因为卷积过少

产生的大量噪声，又能防止高层因为分辨率过低导致的细节处理不佳，同时实现减少模型训练和推理的时间。本文设置了不同的特征输出层组合，得到的结果如表 1。

表 1 不同特征输出层组合的性能表现

输出层	AP/%	AP ₅₀ /%	AP ₇₅ /%
2,3,6	89.7	92.6	90.9
2,4,6	90.3	92.9	91.4
2,5,6	89.5	92.5	90.8
3,4,6	89.9	92.7	91.4
3,5,6	90.1	92.2	91.9
4,5,6	89.0	91.8	90.3

测试数据表明，输出层组合 {S₂, S₄, S₆} 的平均精度 AP 最高，达 90.3%，且在 AP₅₀ 和 AP₇₅ 上也表现优异，分别为 92.9% 和 91.4%。其整体检测性能最为优越，因此改进模型中选择 {S₂, S₄, S₆} 作为特征层组合。

2.3 模型训练分析

训练过程中评估结果如图 5 所示，包含分类损失、回归损失、mAP 变化及精确回归曲线：（a）分类损失从 1.6 迅速降至 0.6，随后缓慢下降并在 0.4 趋于稳定，表明模型完成了快速学习和后期微调；（b）回归损失从 6 迅速降至接近 1，之后缓慢下降并趋稳，反映目标定位性能优化；（c）mAP 在训练初期迅速从 40% 增至 80%，后期逐渐平缓并稳定在 89% 左右，表明模型快速学习有效特征后逐步优化性能；（d）精确回归曲线显示不同召回率下精度表现，随召回率提升，C75（0.869）、C50（0.903）、Loc（0.956）、Sim（0.977）、Oth（0.977）、BG（1.000）和 FN（1.000）。随着召回率的提高，模型在捕捉更多目标时精度逐渐下降。识别更多的目标，但也因此可能引入误检。

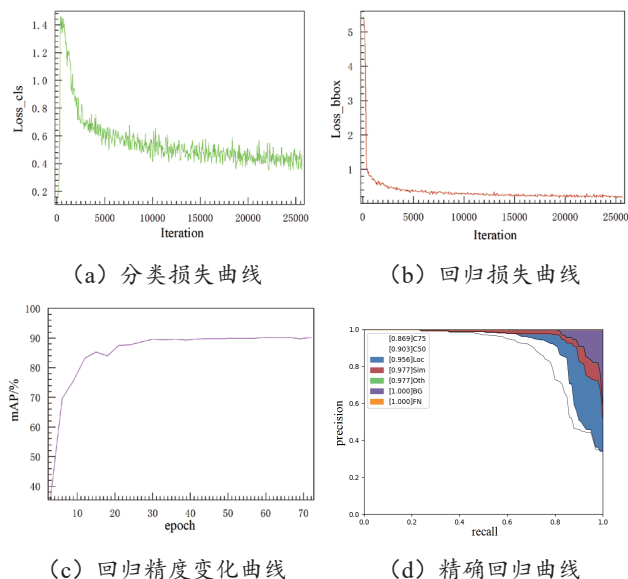


图5 评估参数曲线

如图 6 可见, MNV2IF-DETR 测试结果的混淆矩阵显示, MNV2TF-DETR 的预测值密集分布在对角线上, 检测效果良好。

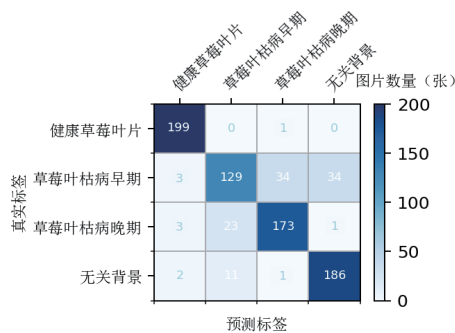


图 6 MNV2IF-DETR 模型混淆矩阵

2.4 不同检测模型的测试对比

将 MNV2IF-DETR 与目前较成熟的目标检测模型 Faster R-CNN、SSD、YOLOv8s、RT-DETR(ResNet18) 进行对比测试, 试验结果如表 2 所示。精确率和平均精度较其他网络高出 0.2~5.3 个百分点, 且浮点运算量大大降低, 减少了 68%~438%, 参数量上, 相较于原模型减少 31.7%, 更少的参数量代表着更低的部署成本, 平衡模型内存占用量和检测速度, 有利于病害检测系统在边缘设备上部署。

表 2 不同模型的检测性能对比

模型	精确率 /%	召回率 /%	平均精度 /%	浮点运算量 /($\times 10^9$)	存储容量 /MB	检测时间 /ms	参数量 /($\times 10^6$)
Faster R-CNN (ResNet50)	90.3	81.6	85.0	90.9	158	112.6	41.4
SSD (MobileNetV2)	89.3	83.7	87.1	32.6	27.6	37.3	14.8
YOLOv8s	91.9	86.5	90.1	28.4	21.4	14.0	11.2
RT-DETR (ResNet18)	90.1	93.9	89.8	52.1	42.8	19.8	19.9
MNV2IF-DETR	92.3	93.7	90.3	16.9	22.1	14.2	13.6

3 结论

本文提出了一种改进型 MNV2IF-DETR 目标检测模型。通过使用 Mobilenet V2 重构 RT-DETR 骨干网络, 选取并设置倒残差块中 $\{S_2, S_4, S_6\}$ 特征输出层组合, 并引入 DW 深度卷积和 SE 注意力机制改进特征融合模块。对比试验表明, 在同等条件下, 改进后模型比原模型 RT-DETR 检测平均精度均值提高 0.5%, 达 90.3%; 参数量减少 31.7%, 浮点计算量降低 67.6%; 检测速度提高 28.3%, 单张图片检测时间仅需 14.2 ms; 同时模型大小减少 48.4%, 仅为 22.1 MB, 实现了模型的轻量化。MNV2IF-DETR 轻量级检测模型具有高检测精度和高检测速度等优势, 能够为草莓叶枯病检测及相关植保装备的研发提供支持。

参考文献:

[1] 韩文心. 草莓枯萎病高效生防菌株筛选与田间应用研究[D]. 南京: 南京农业大学, 2022.

[2]DHAKATE M, INGOLE A B. Diagnosis of pomegranate plant diseases using neural network [C/OL] //2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG). Piscataway: IEEE, 2016[2024-04-13].<https://ieeexplore.ieee.org/document/7490056>.

[3]CARION N, MASSA F, SYNNAEVE G, et al. End-to-End object detection with transformers [C]//16th European Conference. Berlin: Springer, 2020:213-229.

[4]ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection [DB/OL]. (2021-03-18)[2024-04-11].<https://doi.org/10.48550/arXiv.2010.04159>.

[5]MENG D P, CHEN X K, FAN Z J, et al. Conditional DETR for fast training convergence [DB/OL].(2023-09-29)[2024-05-11].<https://doi.org/10.48550/arXiv.2108.06152>.

[6]LIU S L, LI F, ZHANG H, et al. DAB-DETR: dynamic anchor boxes are better queries for DETR [DB/OL]. (2022-03-30) [2024-05-10].<https://doi.org/10.48550/arXiv.2201.12329>.

[7]WU J S, WEI C J, CHEN H R, et al. DS-DETR: a model for tomato leaf disease segmentation and damage evaluation [J/OL].Special issue frontier studies in crop growth monitoring, diagnosis and precision operation,2023[2024-01-19].https://www.mdpi.com/2073-4395/12/9/2023/review_report.

[8]ZHAO Y A, LÜ W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [C/OL]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).Piscataway:IEEE,2024[2024-01-03].<https://ieeexplore.ieee.org/document/10657220>.

[9]SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [DB/OL].(2019-03-21)[2024-02-19].<https://doi.org/10.48550/arXiv.1801.04381>.

[10]GEETHARAMANI G, ARUN P J . Identification of plant leaf diseases using a nine-layer deep convolutional neural network [J]. Computers & electrical engineering, 2019, 76(3): 323-338.

【作者简介】

林晨(2000—), 男, 福建宁德人, 硕士研究生, 研究方向: 植物病害检测人工智能算法应用。

谢知(1982—), 通信作者(email: xz@fafu.edu.cn), 男, 博士, 副教授, 研究方向: 农业信息感知技术。

(收稿日期: 2024-10-14)