肖永帅,陈若冰,时雷,等.基于改进RT-DETR的苹果病害检测算法[J].华中农业大学学报,2026,45(3):87-97.
DOI:10.13300/j.cnki.hnlkxb.2026.03.008

基于改进RT-DETR的苹果病害检测算法

肖永帅,陈若冰,时雷,郑光,尹飞

河南农业大学信息与管理科学学院,郑州 450046

摘要 针对复杂果园环境下苹果病害检测中存在的背景干扰强、多尺度病斑识别困难以及模型轻量化部署需求,提出一种基于改进RT-DETR的轻量化苹果病害检测模型EGA-DETR。模型从3个方面进行了改进:设计RGLAN轻量化特征聚合模块,通过特征分流与重参数化卷积减少冗余计算并增强特征表达;构建BiFPN-GLSA多尺度特征融合模块,通过双向特征传递和全局-局部自注意力机制提升不同尺度病斑的表征能力;引入Inner-Shape-IoU损失函数,以增强模型对不规则病斑目标的定位能力。试验结果显示,EGA-DETR在苹果病害数据集上取得了91.8%的精确率、88.9%的召回率和92.1%的mAP_{50},较基线RT-DETR-18分别提升3.5、3.5、1.5百分点。同时,模型参数量降至11.8×10^6,较基线模型减少40.4%,推理速度达到120帧/s。综上所述,EGA-DETR模型在检测精度与计算效率之间实现了较优平衡,可为苹果病害精准、实时检测提供技术支撑。

关键词 苹果病害检测;RT-DETR;轻量化特征聚合;双向特征金字塔网络;全局-局部自注意力;Inner-Shape-IoU

中图分类号 TP391.41 **文献标识码** A **文章编号** 1000-2421(2026)03-0087-11

苹果是全球种植最广泛的水果之一,苹果产业的健康发展直接关系到农业经济与民生保障。然而,各类病害频繁暴发会显著降低果实品质与产量,威胁产业利润和可持续发展。传统病害检测主要依赖人工诊断[1],该方法不仅耗时耗力,而且容易因主观判断差异导致误诊或漏诊,进而延误病害防治并造成经济损失[2]。随着计算机视觉技术的快速发展,深度学习特别是卷积神经网络的出现开启了自动特征学习的新纪元,卷积神经网络凭借优异的特征提取能力和实时处理性能,迅速成为作物病害诊断的核心技术。为兼顾检测精度与推理速度,以YOLO[3]为代表的单阶段检测器将目标检测建模为端到端回归任务,在病害识别领域得到了广泛应用。研究者围绕模型轻量化与检测鲁棒性展开了大量探索:赵晓等[4]提出改进YOLOv10n,对复杂环境苹果叶片病害检测的mAP_{50}达到90.8%;曾林涛等[5]基于改进YOLOv8提出高效病害检测模型,在苹果叶部病害数据集上mAP_{50}达到97.8%。然而,这些基于CNN[6]的方法存在共同的结构性局限:卷积层依赖局部感受野进行特征提取,在显式建模长距离空间依赖关系方面存在不足,导致模型在果实遮挡、光照变化等复杂场景下难以充分整合全局语义信息,从而在一定程度上影响检测性能的稳定性和鲁棒性。

近年来,基于Transformer[7]架构的检测方法通过自注意力机制有效捕获图像的长程依赖关系,为解决CNN的局部性限制提供了新思路。DETR[8]作为典型代表模型已在农业智能化领域得到应用,Ji等[9]将多维特征提取网络与Transformer模块结合,对与背景近色的绿色苹果实现AP_{50}达97.12%的检测精度。Zhang等[10]融合Transformer和Diffusion机制,对沙漠环境枣树病害检测的mAP_{50}达91%。尽管基于Transformer的DETR模型具有良好的识别性能,但其庞大的参数规模导致计算开销过高,难以满足农业现场对实时性和资源效率的双重要求。RT-DETR[11]作为首个实现端到端实时推理的Transformer检测架构,通过融合CNN的高效局部特征提取与Transformer的全局建模优势,在保持较高检测精度的同时兼顾了推理速度。

然而,苹果病害检测面临诸多技术挑战:病害症状尺度跨度大,微小病斑呈高密度分布;果实与周围

收稿日期:2025-10-10
基金项目:河南省科技攻关项目(242102521027);河南省科技研发计划联合基金项目(222301420113)
肖永帅,E-mail:17657318695@163.com
通信作者:尹飞,E-mail:yin.fei@henau.edu.cn

植株组织在色彩特征上存在显著相似性，背景干扰严重；现有方法参数规模较大，难以满足边缘设备实时部署需求。针对上述问题，本研究基于RT-DETR提出轻量化苹果病害检测模型EGA-DETR。该模型通过设计RGLAN轻量化特征聚合模块、构建BiFPN-GLSA多尺度特征融合模块以及引入Inner-Shape-IoU损失函数，旨在实现模型轻量化、增强多尺度特征表征能力，以提升复杂果园环境下苹果病害目标的检测精度与推理效率。

1 材料与方法

1.1 数据集的构建

本研究采用Roboflow[12]的“Apple Disease Detection Computer Vision Model”数据集，该数据集涵盖了褐斑病和炭疽病2种常见病害类型。褐斑病：典型表现为不规则圆形病斑，具有明显的同心环纹，颜色呈褐色至深褐色。炭疽病：表现为凹陷的深色病斑，形状不规则，常呈现多个小斑点融合状态，这些形态学差异为基于深度学习的检测提供了可区分的视觉特征基础。数据集图像来源多样，具有场景多样性和光照变化特性，为模型提供了全面的训练素材。为确保数据质量，本研究对原始数据集进行了系统性审查，剔除了模糊不清、标注错误及重复样本，最终筛选出2 100张有效图像。使用Labelimg标注软件对数据集样本进行标注。按照7∶2∶1的比例将数据集划分为训练集、验证集和测试集，其中训练集1 470张、验证集420张、测试集210张，所有图像均调整为640像素×640像素。为防止因样本数量不足导致的过拟合和泛化性能下降问题，仅对训练集实施数据增强操作，验证集与测试集保持原始状态，以确保模型评估结果的客观性与可信度。对训练集图像分别独立实施水平翻转、垂直翻转、随机角度旋转和高斯模糊4种增强操作，连同原始图像共扩充至7 350张。图1展示了数据增强效果。



A：原图；B：水平翻转；C：垂直翻转；D：随机角度变化；E：模糊。A：Original image；B：Horizontal flip；C：Vertical flip；D：Random angle variation；E：Blur.

图1 苹果病害数据集增强案例

Fig. 1 Augmentation examples of the apple disease dataset

1.2 EGA-DETR苹果病害检测模型

RT-DETR是首个端到端实时目标检测的Transformer架构，由CNN骨干网络、混合编码器、Transformer解码器和辅助预测头组成。CNN骨干提取低、中、高级多尺度特征；混合编码器中，AIFI（attention-based intra-scale feature interaction，基于注意力的尺度内特征交互）模块增强高级特征的层次性，CCFM（CNN-based cross-scale feature-fusion module，基于CNN的跨尺度特征融合）模块促进跨尺度特征融合；处理后的特征经Transformer解码器生成目标预测，其中IoU感知查询用于优化解码器初始化。本研究在RT-DETR-18基础上提出轻量化检测模型EGA-DETR，模型结构如图2所示。EGA-DETR由3个部分组成：轻量化主干网络、高效混合编码器和解码预测模块。输入图像经标准化预处理调整至统一尺寸后送入主干网络；主干网络中的BasicBlock由RGLAN模块替换，从S3、S4、S5共3个语义层级输出多尺度特征；高效混合编码器以BiFPN-GLSA模块替换原有Neck结构，对多尺度特征进行跨尺度融合，并结合AIFI模块完成特征序列化；解码预测模块基于IoU感知查询选择策略筛选高质量特征序列作为初始目标查询，由解码器迭代细化后输出边界框与置信度分数；边界框回归过程采用Inner-Shape-IoU损失函数。

1.3 设计RGLAN模块

针对RT-DETR主干网络中的特征冗余与计算开销问题，本研究设计了RGLAN轻量化特征聚合模块，结构如图3所示。RGLAN模块采用双分支结构。输入特征首先经1×1卷积调整通道数，随后由Split操作按通道维度等比例划分为2个分支：第一分支以零参数恒等映射保留原始特征，第二分支依次经过RepConv（reparameterized convolution，重参数

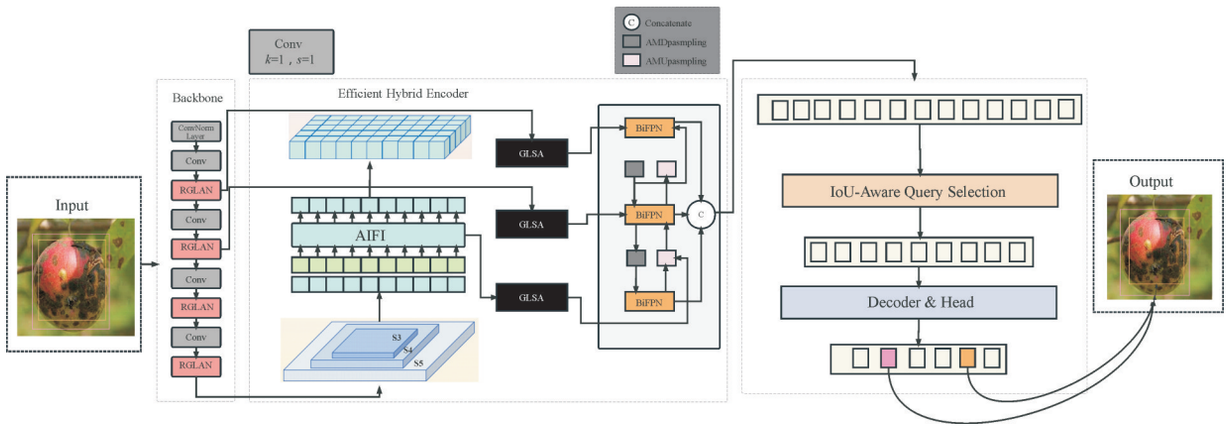


图 2　EGA-DETR 模型网络架构

Fig. 2　EGA-DETR model network architecture

化卷积)、2次3×3卷积和1个1×1卷积完成特征变换;2个分支的输出经Concat操作拼接后,再由1×1卷积层完成通道整合,得到模块输出特征。该结构仅对50%的输入通道执行密集计算,其余通道直接传递,由此降低浮点运算速度与显存占用。RepConv在训练阶段以多分支结构提升特征表达,推理阶段通过参数融合等价转换为单分支卷积,因此在不增加推理开销的前提下增强了特征学习能力。

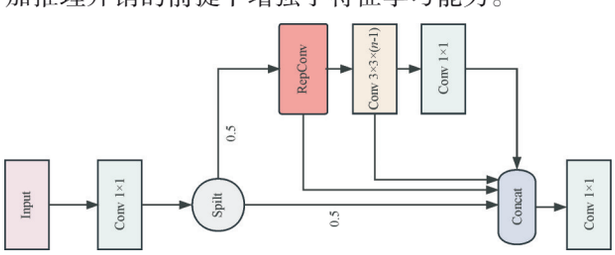


图 3　RGLAN模块结构图

Fig. 3　Structure diagram of the RGLAN module

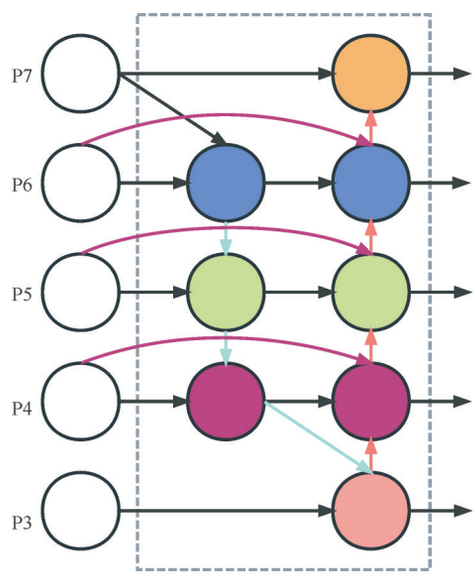


图 4　BiFPN模块结构图

Fig. 4　Structure of the BiFPN module

1.4　构建 BiFPN-GLSA 模块

针对多尺度目标检测中特征表征不充分的问题,本研究引入BiFPN[13]构建高效多尺度特征融合框架,BiFPN结构如图4所示。该架构通过重构传统特征金字塔的信息传递拓扑,实现跨分辨率语义信息的耦合传递。BiFPN移除了仅有单向输入的冗余节点,同时引入跨越多个分辨率层级的直接传播路径:自顶向下路径承载高阶语义信息的下沉,自底向上路径承载空间精度信息的上传。在特征融合层面,BiFPN的多个融合节点采用可学习加权系数的自适应聚合策略,为不同输入分支分配独立权重参数,使网络能够根据各尺度特征的贡献度动态调节融合结果,缓解异构特征直接叠加产生的语义冲突,并在全局病变模式与局部纹理细节之间实现自适应平衡。

此外,为进一步提升模型对苹果病害特征的识别精度,引入GLSA[14](global-local self-attention,全局-局部自注意力)模块。GLSA模块融合了全局空间注意力(global spatial attention,GSA)单元和局部空间注意力(local spatial attention,LSA)单元的计算结果,双流设计同时保持全局语义建模和局部细节感知能力,并通过通道分离策略平衡检测精度与计算资源消耗。GLSA模块结构如图5所示。

具体而言,特征图$\{F_i|i\in(2,3,4)\}$的64个通道被均匀划分为2个特征图组F_i^1和F_i^2,分别输入到全局空间注意力(GSA)模块和局部空间注意力(LSA)模块中。2个注意力单元的输出最终通过拼接操作和1×1卷积层融合。整个过程可表示为:

$$F_i^1,F_i^2=\mathrm{Split}(F_i) \tag{1}$$

$$F_i'=\mathrm{C}_{1\times1}\left(\mathrm{Concat}\left(F_{\mathrm{G}}(F_i^1),F_{\mathrm{L}}(F_i^2)\right)\right) \tag{2}$$

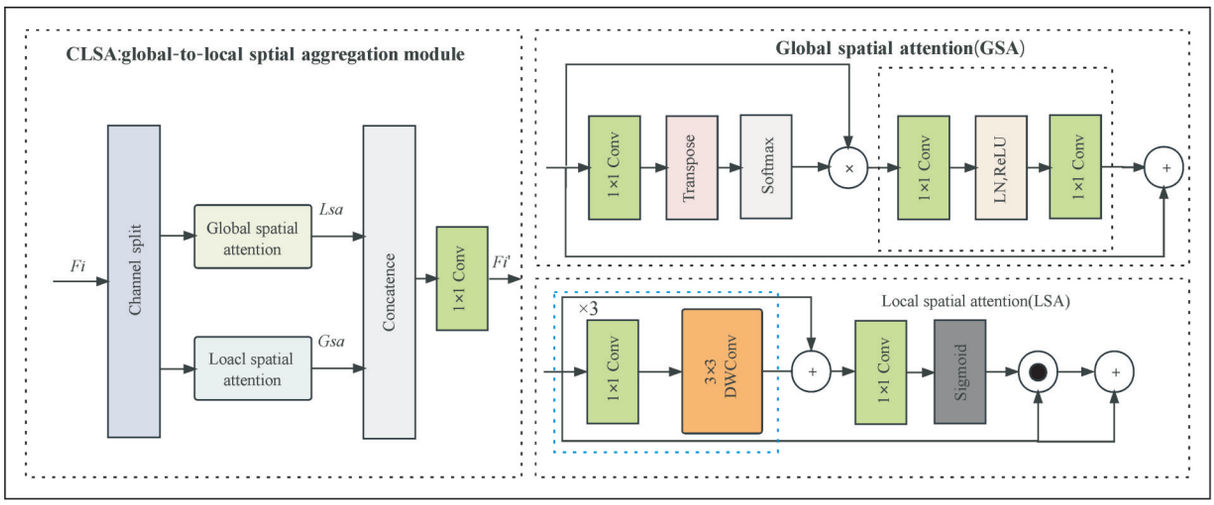


图5 GLSA模块结构图

Fig. 5 Structure diagram of the GLSA module

式(1)~(2)中，$F_G(\cdot)$表示全局空间注意力，$F_L(\cdot)$表示局部空间注意力，F_i'为各尺度对应的输出特征图，C1×1表示1×1卷积操作，Split表示通道分离操作，Concat表示拼接操作。

全局空间注意力(GSA)模块：F_G强调空间维度中每个像素的长距离关联关系，可作为局部空间注意力的有效补充。受长距离交互提取方法的启发，通过简单的方式生成全局空间注意力图($F_G \in \mathrm{R}^{(H/8\times W/8\times 32)}$)，其计算过程如下：

$$A_G(F_i^1)=\mathrm{Softmax}\left(\mathrm{Transpose}\left(\mathrm{Conv}_{1\times1}(F_i^1)\right)\right) \quad (3)$$

$$F_G(F_i^1)=P_{ml}(A_G(F_i^1)\otimes F_i^1)+F_i^1 \quad (4)$$

式(3)~(4)中，$A_G(\cdot)$为注意力操作，$\mathrm{Conv}_{1\times1}$表示1×1卷积，$\otimes$表示矩阵乘法；$P_{ml}(\cdot)$由2个全连接层、ReLU非线性激活和归一化层组成，第一层将输入变换到高维空间(扩张比为2)，第二层恢复到与输入相同的维度；H、W分别为原始输入图像的高度和宽度。

局部空间注意力(LSA)模块：F_L模块在给定特征图的空间维度上有效提取感兴趣区域的局部特征，特别适合捕获苹果病害的微小目标特征。F_L模块的局部空间注意力响应($F_L \in \mathrm{R}^{(H/8\times W/8\times 32)}$)的计算过程为：

$$A_L(F_i^2)=\sigma\left(\mathrm{C}_{1\times1}(f_c(F_i^2))+F_i^2\right) \quad (5)$$

$$F_L(F_i^2)=A_L(F_i^2)\odot F_i^2+F_i^2 \quad (6)$$

式(5)~(6)中，$f_c(\cdot)$表示3个1×1卷积层和1个3×3深度可分离卷积层的级联结构，通道数在f_c中调整为32。$A_L(\cdot)$为局部注意力操作，$\sigma(\cdot)$为sigmoid函数，$\odot$表示逐点乘法。该结构能够高效聚合局部空间信息并使用更少的参数，使网络既能定位大面积苹果病变区域，又能识别微小病斑的细节特征，适用于果园环境中形态多样、尺度差异显著的病害目标。

1.5 Inner-Shape-IoU损失函数

RT-DETR架构默认集成GIoU[15]作为边界框回归的优化准则。然而，针对苹果病害检测这一特定应用领域，GIoU对不规则形状病害区域的几何建模能力不足，导致检测效果仍有提升空间。为此，本研究融合了Inner-IoU[16]和Shape-IoU[17]，构建复合损失函数Inner-Shape-IoU，该方法在区域重叠度量的基础上引入边界框内部几何形状结构化表征，建立更精细的相似性评估准则。

Inner- IoU引入辅助边界框机制，根据回归样本的IoU状态动态调整辅助边界框尺度：对于高IoU样本，采用较小尺度辅助边界框以产生更大的梯度绝对值，加速收敛；对于低IoU样本，采用较大尺度辅助边界框以扩展有效回归范围，增强对远距离目标的定位能力。如图6所示，真实框(ground truth，GT)和预测框分别记为B^{gt}和B，GT框中心坐标记为(x_c^{gt}, y_c^{gt})，预测框中心坐标记为(x_c, y_c)；内GT框和内预测框分别与原GT框、原预测框共中心，按比例因子r缩放得到。GT框宽度和高度分别为w^{gt}和h^{gt}，预测框宽度和高度分别为w和h；r为辅助边界框比例因子，取值范围为[0.5，1.5]；b_l^{gt}、b_r^{gt}、b_t^{gt}、b_b^{gt}分别为真实边界框的左、右、上、下边界；b_l、b_r、b_t、b_b分别表示为预测边界框的左、右、上、下边

界(下标l、r、t、b分别为left、right、top、bottom的首字母)；其中式(7)中的“±”对应左右边界；式(8)中的“±”对应上下边界；S_{inter}为辅助边界框的交集面积，S_{union}为辅助边界框的并集面积；上标gt表示真实框(ground truth)。

$$b_{l,r}^{\text{gt}}=x_{\text{c}}^{\text{gt}}\pm\frac{w^{\text{gt}}\cdot r}{2} \tag{7}$$

$$b_{t,b}^{\text{gt}}=y_{\text{c}}^{\text{gt}}\pm\frac{h^{\text{gt}}\cdot r}{2} \tag{8}$$

$$S_{\text{inter}}=\left[\min(b_{\text{r}}^{\text{gt}},b_{\text{r}})-\max(b_{\text{l}}^{\text{gt}},b_{\text{l}})\right]\cdot\left[\min(b_{\text{b}}^{\text{gt}},b_{\text{b}})-\max(b_{\text{t}}^{\text{gt}},b_{\text{t}})\right] \tag{9}$$

$$S_{\text{union}}=w^{\text{gt}}\cdot h^{\text{gt}}\cdot r^2+w\cdot h\cdot r^2-S_{\text{inter}} \tag{10}$$

$$\text{IoU}_{\text{inner}}=\frac{S_{\text{inter}}}{S_{\text{union}}} \tag{11}$$

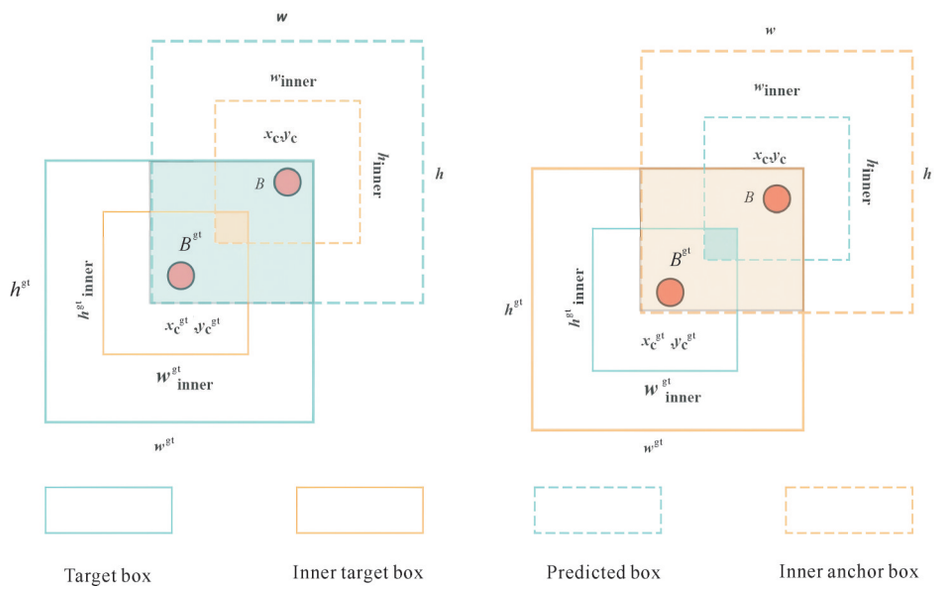


图6　Inner-IoU计算原理

Fig. 6　Inner-IoU calculation principle

Shape-IoU综合考虑边界框的角度关系、距离特征和形状相似性，对不规则形状目标提供几何约束。其角度损失通过计算预测框与真实框中心连线相对于坐标轴的最小角度，引导网络优先沿坐标轴方向对齐，适用于条纹状或椭圆形病斑的定位；距离损失结合角度成本动态调整中心点距离权重，适用于不规则形状或聚集分布病斑的中心定位。Shape-IoU计算原理如图7所示，计算过程如下：

$$W_w=\frac{2\cdot(w^{\text{gt}})^s}{(w^{\text{gt}})^s+(h^{\text{gt}})^s} \tag{12}$$

$$W_h=\frac{2\cdot(h^{\text{gt}})^s}{(h^{\text{gt}})^s+(w^{\text{gt}})^s} \tag{13}$$

$$\omega_{\text{w}}=W_h\cdot\frac{|w-w^{\text{gt}}|}{\max(w,w^{\text{gt}})} \tag{14}$$

$$\omega_{\text{h}}=W_w\cdot\frac{|h-h^{\text{gt}}|}{\max(h,h^{\text{gt}})} \tag{15}$$

$$\Omega_s=\sum_{t=w,h}\left(1-\text{e}^{-\omega_t}\right)^{\theta} \tag{16}$$

$$d_s=W_{\text{w}}\cdot\frac{(x_{\text{c}}-x_{\text{c}}^{\text{gt}})^2}{c^2}+W_{\text{h}}\cdot\frac{(y_{\text{c}}-y_{\text{c}}^{\text{gt}})^2}{c^2} \tag{17}$$

$$L_{\text{Shape-IoU}}=1-\text{IoU}+d_s+0.5\cdot\Omega_s \tag{18}$$

式(12)～(18)中，s为病害数据集的几何适应因子；W_{w}和W_{h}分别为水平方向和垂直方向的几何权重系数；ω_{w}和ω_{h}分别为宽度方向和高度方向的形状偏差；c为预测框与真实框最小外接矩形的对角线长度；θ为形状偏差敏感性调节参数；Ω_s为形状损失项；d_s为形状距离项；IoU为预测框与真实框交并比；$L_{\text{Shape-IoU}}$为Shape-IoU损失函数值：

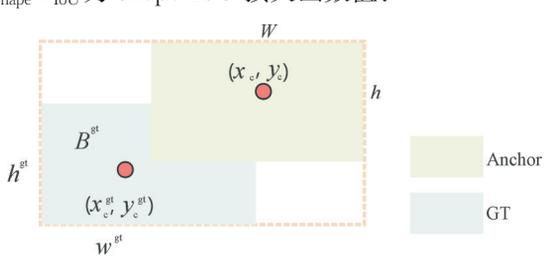


图7　Shape-IoU计算原理

Fig. 7　Shape-IoU computation principle

1.6　试验环境、参数配置与训练策略

实验硬件配置采用Intel Core i5-12600KF处理器、NVIDIA RTX 4060 Ti 8 GB显卡和32 GB内存；软件环境基于Python 3.8.0和PyTorch 1.10.0构

建。模型训练软硬件配置如表1所示。

表1 模型训练环境

Table 1 Experimental environment for model training

组件 Component	规格 Specification
CPU	Intel i5-12600KF
GPU	NVIDIA RTX 4060 Ti 8 GB
内存 Memory	32 GB
深度学习框架 Framework	PyTorch 1.10.0
编程语言 Programming language	Python 3.8.0

模型超参数配置如下：优化器采用自动模式(auto)，由训练框架根据数据属性自行选择算法；初始学习率设为0.01，动量参数设为0.937，权重衰减系数设为0.000 5；训练轮次根据预实验结果设定为200。

1.7 评价指标

表2给出了本研究的模型评价指标体系，从检测准确性、运行效率和模型复杂度3个方面进行评估。检测准确性包括精确率(precision，P)、召回率(recall，R)和平均精度均值(mean average precision，mAP；公式中以P_{mA}表示)3个指标；运行效率采用每秒帧数(frames per second，FPS；公式中以F_{PS}表示)表征；模型复杂度包括参数量(公式以N_P表示)、模型权重大小和浮点运算速度(floating point operations per second，FLOPs；公式中以N_F表示)3个指标。各指标的计算公式如下：

$$P=\frac{T_p}{T_p+F_p}\times 100\% \tag{19}$$

$$R=\frac{T_p}{T_p+F_N}\times 100\% \tag{20}$$

$$P_{mA}=\frac{1}{n}\sum_{i=1}^{n}P_{A,i} \tag{21}$$

$$P_{mA\ 50:95}=\frac{1}{10}\sum_{i=0}^{9}P_{mA\ 0.50+0.05i} \tag{22}$$

$$F_{PS}=\frac{1}{T_{inf}} \tag{23}$$

$$N_P=k^2\times C_{in}\times C_{out} \tag{24}$$

$$N_F=k^2\times C_{in}\times C_{out}\times H_{out}\times W_{out} \tag{25}$$

式(19)～(25)中，T_p为真正例数量(true positive，实际为正样本且被预测为正样本的数量)；F_p为假正例数量(false positive，实际为负样本但被预测为正样本的数量)；F_N为假负例数量(false negative，实际为正样本但被预测为负样本的数量)；AP为单类别平均精度，公式中以P_A表示；n为类别总数；T_{inf}为单帧推理时间，s；k为卷积核大小；C_{in}和C_{out}分别为卷积层的输入通道数和输出通道数；H_{out}和W_{out}分别为输出特征图的高度和宽度。

表2 模型评价指标

Table 2 Evaluation metrics of the model

指标 Metric	描述 Description
精确率(P) Precision	模型正确预测的正样本占预测正样本总数的比例
召回率(R) Recall	模型正确预测的正样本占实际正样本总数的比例
平均精度均值 Mean average precision	在IoU阈值为0.50时各类别平均精度的均值
检测速度(FPS) Frames per second	模型每秒可处理的图像帧数
参数量 Number of parameters	模型中可训练参数的总数
模型权重大小 Weight size	模型文件的存储大小
浮点运算速度(FLOPs) Floating point operations per second	模型推理过程的计算复杂度

2 结果与分析

2.1 特征融合方法对比分析

基于RT-DETR-18基准模型，本研究对比了3种特征融合方案：BiFPN、BiFPN-Weight和BiFPN-GLSA[18]。BiFPN采用标准的双向特征金字塔网络；BiFPN-Weight在BiFPN基础上引入可学习的自适应权重，动态调节不同尺度特征的贡献度；BiFPN-GLSA将全局-局部空间注意力模块嵌入BiFPN架构，实现特征的空间维度增强。试验结果如表3所示。引入BiFPN后，精确率由基线的88.3%降至86.6%，下降1.7百分点；mAP_{50}由90.6%降至89.7%，下降0.9百分点；检测速度由117帧/s降至103帧/s，且FLOPs由56.9×10^9增至64.6×10^9，BiFPN在RT-DETR架构下未能直接带来性能提升。BiFPN-Weight引入自适应加权后，精确率回升至87.8%、mAP_{50}回升至90.1%，但仍低于基线水平，且参数量由19.8×10^6增加至26.8×10^6，增加35.4%。BiFPN-GLSA在所有改进方案中表现最优：精确率达到89.2%，较基线提升0.9百分点；参数量降至17.7×10^6，较基线减少10.6%；FLOPs降至49.1×10^9，较基线减少13.7%；检测速度为115帧/s。结果表明，GLSA的双分支设计(全局分支捕获长距离依赖、局部分支保留空间细节)结合通道分离策略，能够在精度与计算量之间实现较优平衡。

2.2 损失函数对比分析

为进一步验证损失函数改进的有效性，以RT-DETR-18作为基线网络，对Inner-Shape-IoU与原网络使用的GIoU、CIoU和DIoU[19]进行对比分析，不同损失函数对比试验结果如图8所示。在精确率指标上，4种损失函数依次为：88.3%(GIoU)、89.1%

表 3　不同特征融合方法对比

Table 3　Comparison of different feature fusion methods

模型 Model	精确率/% Precision	召回率/% Recall	mAP_{50}/%	检测速度 FPS	参数量(×10^6) Parameters	浮点运算速度/(×10^9 s^{-1})FLOPs
RT-DETR-18	88.3	85.4	90.6	117	19.8	56.9
BiFPN	86.6	85.8	89.7	103	20.4	64.6
BiFPN-Weight	87.8	86.9	90.1	110	26.8	59.4
BiFPN-GLSA	89.2	86.6	90.4	115	17.7	49.1

(DIoU)、89.9%(CIoU)、90.8%(Inner-Shape-IoU)；Inner-Shape-IoU 较 GIoU 提升 2.5 百分点，较 CIoU 提升 0.9 百分点。在召回率指标上，Inner-Shape-IoU 达到 87.0%，分别较 GIoU、DIoU、CIoU 提升 1.6、1.0、0.7 百分点。在 mAP_{50} 指标上，Inner-Shape-IoU 达到 91.2%，较 GIoU、CIoU 分别提升 0.6、0.4 百分点；DIoU 的 mAP_{50} 为 89.4%，反而较 GIoU 下降 1.2 百分点，表明仅引入中心点距离约束未能改善综合检测性能。上述结果表明，CIoU 在 GIoU 基础上引入长宽比一致性约束，3 项指标均有提升，但仅依靠长宽比约束难以充分刻画苹果病斑的不规则轮廓特征，性能上限受限；DIoU 虽在精确率和召回率上较 GIoU 略有提升，但 mAP_{50} 出现回落，说明距离约束对苹果病害这类形状不规则、尺度差异大的目标适应性不足。Inner-Shape-IoU 在长宽比和距离约束之外，进一步通过辅助边界框尺度调节回归过程，并将形状偏差因子纳入损失计算，对边缘模糊、形态多样的病斑目标更具适应性，因而在 3 项指标上均取得最优结果。

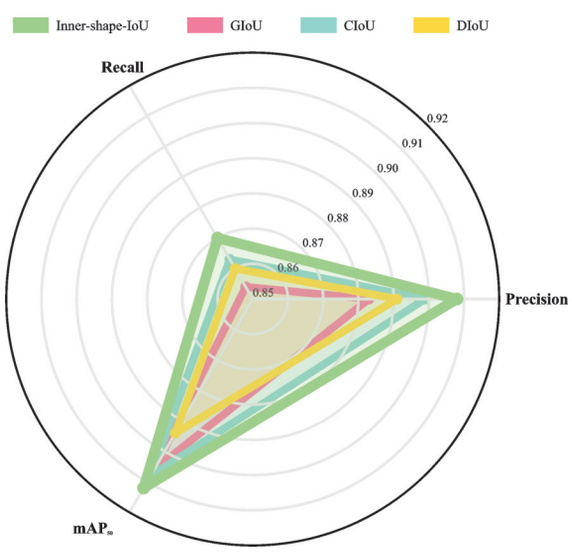


图 8　不同损失函数对比

Fig. 8　Comparison of different loss functions

2.3　对比试验

在相同试验条件下，将 EGA-DETR 与 RT-DETR 系列(RT-DETR-34、RT-DETR-X、RT-DETR-50)和 YOLO 系列(YOLOv7、YOLOv8L、YOLOv11X)共 6 种主流目标检测模型进行对比试验，结果如表 4 所示。

在检测精度方面，EGA-DETR 精确率为 91.8%，召回率为 88.9%，mAP_{50} 为 92.1%，$mAP_{50:95}$ 为 61.8%，4 项指标均高于其他对比模型。其中，mAP_{50} 较 RT-DETR-X、YOLOv7、YOLOv8L 分别提升 1.3、4.8、1.1 百分点；$mAP_{50:95}$ 较 RT-DETR-34、RT-DETR-X、YOLOv7、YOLOv8L 分别提升 3.3、3.7、6.3、2.9 百分点。在计算复杂度方面，EGA-DETR 参数量为 11.8×10^6，FLOPs 为 36.1×10^9，均为对比模型中最低值；与参数量次低的 RT-DETR-34 相比，参数量减少 62.4%，FLOPs 降低 60.2%。在推理速度方面，EGA-DETR 达到 120 帧/s，高于其他所有对比模型。综合上述指标，EGA-DETR 在精度、轻量化、推理速度 3 个维度上同时取得最优结果。

不同模型对苹果病害的检测结果如图 9 所示。YOLO 系列在测试图像中存在不同程度的漏检：YOLOv7、YOLOv8L、YOLOv11X 分别漏检 4、2、2 个目标。漏检集中出现在低对比度病斑场景，反映出 YOLO 系列特征提取网络在与背景颜色接近的病害区域上敏感度不足。RT-DETR 系列虽对小目标特征更敏感，但在与背景视觉特征相近的果实区域出现错检：RT-DETR-34、RT-DETR-X、RT-DETR-50 分别错检 2、3、3 个目标。EGA-DETR 在测试图像中未出现漏检与错检，主干网络的 RGLAN 模块降低冗余计算的同时保留细粒度特征，BiFPN-GLSA 模块的全局-局部双分支结构兼顾远距离依赖和空间细节，配合 Inner-Shape-IoU 对不规则病斑边界的拟合，整体检测稳定性较 RT-DETR 与 YOLO 系列更高。

表4 EGA-DETR与6种模型的试验对比

Table 4 Comparison of EGA-DETR with six models

评价指标 Evaluation metrics	EGA-DETR	RT-DETR-34	RT-DETR-X	RT-DETR-50	YOLOv7	YOLOv8L	YOLOv11X
精确率/% Precision	**91.8**	90.7	89.6	89.4	80.6	84.9	89.2
召回率/% Recall	**88.9**	85.9	82.6	84.7	79.6	79.9	78.8
mAP_{50}/%	**92.1**	90.6	90.8	89.3	87.3	91.0	88.7
检测速度/(帧/s) FPS	**120**	110	105	101	91	112	113
$mAP_{50:95}$/%	**61.8**	58.5	58.1	57.1	55.5	58.9	56.3
参数量($\times10^6$) Parameters	**11.8**	31.4	67.4	36.6	36.5	43.7	56.2
浮点运算速度/(10^9 s^{-1}) FLOPs	**36.1**	90.6	108	134	105	160	190



绿色方框代表正确检测；红色方框代表漏检；蓝色方框代表错检。 The green box represents correct detection; the red box represents missed detection; and the blue box represents false detection.

图9 不同模型对苹果病害检测结果的比较

Fig.9 Comparison of detection results of different models

2.4 消融实验

为验证各改进模块的贡献，对RT-DETR-18基线进行3项替换：(A) RGLAN替换BasicBlock；(B) BiFPN-GLSA替换原始Neck；(C) Inner-Shape-IoU替换GIoU。各模块单独引入及组合引入的试验结果如表5所示。

单模块引入结果显示，RGLAN(+A)侧重轻量化：精确率较基线提升1.3百分点，参数量由19.8×10^6降至13.0×10^6，FLOPs由56.9×10^9降至44.5×10^9，推理速度由117帧/s提升至122帧/s。BiFPN-GLSA(+B)侧重特征融合：精确率提升0.9百分点，召回率提升1.2百分点，参数量降至17.7×10^6，FLOPs降至49.1×10^9，但推理速度小幅下降至115帧/s(受双向特征传播带来的额外计算路径影响)。Inner-Shape-IoU(+C)不改变网络结构，仅替换损失函数，因此参数量、FLOPs、模型大小均与基线一致；3项准确性指标全部提升，精确率90.8%、召回率87.0%、mAP_{50} 91.2%。

双模块组合结果显示，A+B组合在轻量化方向叠加：参数量降至11.8×10^6，较基线降低40.4%、FLOPs降至36.1×10^9，较基线降低36.6%、模型大小压缩至22.9 MB，较基线降低40.7%；A+C组合在精度方向叠加：精确率91.0%、召回率87.4%，较基线分别提升2.7和2.0百分点；B+C组合精度提升幅度有限，且推理速度回落至111帧/s。

三模块替换后的模型A+B+C在7项指标上均达到最优组合：精确率91.8%、召回率88.9%、mAP_{50} 92.1%，同时维持11.8×10^6参数和22.9 MB模型大小以及120帧/s的推理速度。比较A+B与A+B+C的差异可见，Inner-Shape-IoU在不增加任何计算

表 5　不同方法消融实验对比

Table 5　Ablation study comparison of different methods

方法 Method	精确率/% Precision	召回率/% Recall	mAP_{50}/%	检测速度/(帧/s) FPS	参数量($\times10^6$) Params	浮点运算速度/(10^9 s^{-1}) FLOPs	权重大小/MB Weight size
RT-DETR-18	88.3	85.4	90.6	117	19.8	56.9	38.6
RT-DETR-18+A	89.6	86.2	90.9	122	13.0	44.5	27.3
RT-DETR-18+B	89.2	86.6	90.4	115	17.7	49.1	34.3
RT-DETR-18+C	90.8	87.0	91.2	116	19.8	56.9	38.6
RT-DETR-18+A+B	90.7	87.6	91.1	120	11.8	36.1	22.9
RT-DETR-18+A+C	91.0	87.4	91.3	121	13.0	44.5	27.3
RT-DETR-18+B+C	89.7	86.9	90.8	111	17.7	49.1	34.3
RT-DETR-18+A+B+C	**91.8**	**88.9**	**92.1**	120	**11.8**	**36.1**	**22.9**

注：A 表示用 RGLAN 替换 BasicBlock；B 表示用 BiFPN-GLSA 替换原始 Neck；C 表示用 Inner-Shape-IoU 替换 GIoU。最优指标加粗。

Note: A indicates replacing BasicBlock with RGLAN; B indicates replacing the original Neck with BiFPN-GLSA; C indicates replacing GIoU with Inner-Shape-IoU. The best metrics are in bold.

开销的前提下进一步提升了边界框定位精度，与 RGLAN 的特征聚合及 BiFPN-GLSA 的多尺度融合形成功能互补。为深入探究各改进模块对特征提取的影响机制，本研究采用梯度加权类激活映射技术（gradient-weighted class activation mapping++，Grad-CAM++）[20]对模型注意力分布进行可视化分析，结果如图 10 所示。基线 RT-DETR-18 的激活区域较为分散，对病斑区域的响应强度不足；引入 RGLAN 后，激活区域向目标病斑收敛；进一步引入 BiFPN-GLSA 后，病斑边缘和内部区域同时产生强响应，体现双向特征传递与全局-局部注意力的作用；完整模型 A+B+C 的高激活区域与病斑实际轮廓重合度最高，对不规则形状病斑的定位最准确，与表 5 中 mAP_{50} 从 90.6% 升至 92.1% 的定量结果一致。

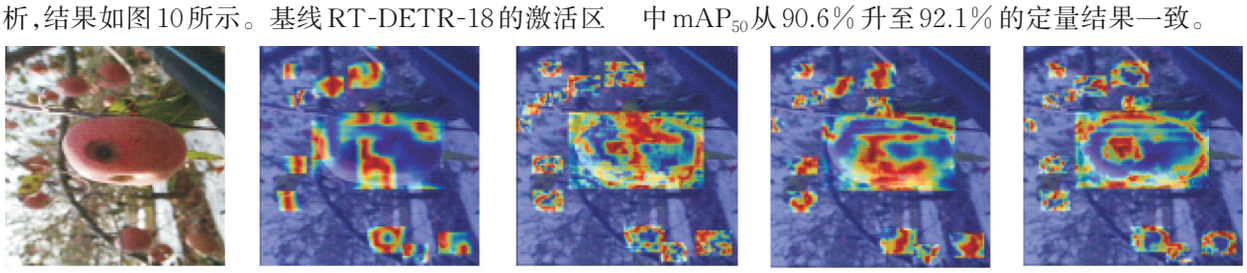


图 10　Grad-CAM++热力图

Fig. 10　Grad-CAM++ heatmap

3　讨　论

针对苹果病害检测中存在的背景干扰强、多尺度病斑识别困难以及模型轻量化部署需求等问题，本研究对基线模型 RT-DETR 进行多项改进，构建轻量化检测模型 EGA-DETR。试验结果表明，EGA-DETR 在苹果病害数据集上取得了 91.8% 的精确率、88.9% 的召回率和 92.1% 的 mAP_{50}，较基线 RT-DETR-18 分别提高 3.5、3.5 和 1.5 百分点；参数量降至 11.8×10^6，FLOPs 降至 36.1×10^9，推理速度达到 120 帧/s。表明所提方法在提升检测性能的同时，能够有效降低模型复杂度，具有较好的实时检测能力。本研究工作主要体现在以下 3 个方面。(1)RGLAN 轻量化特征聚合模块通过特征分流与重参数化卷积相结合，减少主干网络中的冗余计算，在降低模型复杂度的同时增强了特征表达能力；(2)BiFPN-GLSA 多尺度特征融合模块通过双向特征传递与全局-局部自注意力机制相结合，提升复杂背景下不同尺度病斑目标的特征表征效果；(3)Inner-Shape-IoU 损失函数增强模型对不规则病斑目标的几何建模能力，从而改善了边界框回归精度。消融实验和对比试验结果表明，上述改进具有较好的协同增益效果，能够有效提升模型综合检测性能。

本研究所采用的数据集主要包含褐斑病和炭疽病 2 类苹果病害，数据规模和场景类型相对有限，因此模型在更多病害种类、不同生长阶段及跨场景条件下的泛化能力仍有待进一步验证。后续研究可从

扩充数据集规模与类型、增强自然环境下模型鲁棒性以及推进移动端和嵌入式平台部署等方面展开。总体而言，本研究提出的EGA-DETR模型为苹果病害实时检测提供了一种兼顾检测精度与计算效率的实现思路，也可为农业场景下轻量化目标检测模型的设计与优化提供参考。

参考文献 References

[1] 郑超杰，李少波，蒲睿强，等.基于轻量化卷积神经网络的番茄叶片病害识别[J].江苏农业科学，2024，52(11)：225-231. ZHENG C J，LI S B，PU R Q，et al.Tomato leaf disease recognition based on lightweight convolutional neural network[J]. Jiangsu agricultural sciences，2024，52(11)：225-231 (in Chinese with English abstract).

[2] 杜英杰，宗哲英，王祯，等.农作物病害诊断方法现状和展望[J].江苏农业科学，2023，51(6)：16-23. DU Y J，ZONG Z Y，WANG Z，et al. Current situation and prospect of diagnostic methods for crop diseases[J]. Jiangsu agricultural sciences，2023，51(6)：16-23 (in Chinese with English abstract).

[3] BOCHKOVSKIY A，WANG C Y，LIAO H M.YOLOv4：optimal speed and accuracy of object detection[PP/OL].arXiv，2020：2004.10934 [2025-10-10]. https://arxiv. org/abs/2004.10934.

[4] 赵晓，杨梦婷，张懿丹.基于改进YOLO v10n的苹果叶片病害检测方法[J].农业机械学报，2025，56(9)：657-666，728. ZHAO X，YANG M T，ZHANG Y D.Apple leaf disease detection method based on improved YOLO v10n[J]. Transactions of the CSAM，2025，56(9)：657-666，728 (in Chinese with English abstract).

[5] 曾林涛，马嘉昕，丁羽，等.基于改进YOLO v8的苹果叶部病害检测方法[J].江苏农业科学，2025，53(5)：147-156. ZENG L T，MA J X，DING Y，et al.An apple leaf disease detection method based on improved YOLO v8[J].Jiangsu agricultural sciences，2025，53(5)：147-156 (in Chinese with English abstract).

[6] LIU L，OUYANG W L，WANG X G，et al. Deep learning for generic object detection：a survey[J]. International journal of computer vision，2020，128(2)：261-318.

[7] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach：Curran Associates，2017：6000-6010.

[8] CARION N，MASSA F，SYNNAEVE G，et al. End-to-end object detection with transformers[C]//Computer Vision - ECCV 2020.Cham：Springer，2020：213-229.

[9] JI W，ZHAI K L，XU B，et al. Green apple detection method based on multidimensional feature extraction network model and transformer module[J/OL]. Journal of food protection，2025，88 (1)：100397 [2025-10-10]. https://doi. org/10.1016/j.jfp.2024.100397.

[10] ZHANG L X，ZHAO Y，ZHOU C C，et al.JuDifformer：multimodal fusion model with transformer and diffusion for jujube disease detection[J/OL].Computers and electronics in agriculture，2025，232：110008 [2025-10-10]. https://doi. org/10.1016/j.compag.2025.110008.

[11] ZHAO Y A，LV W Y，XU S L，et al.DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2024：16965-16974.

[12] PFE. Apple disease detection computer vision model[DS/OL]. (2022)[2025-10-10]. https://universe.roboflow.com/pfe-ns5wl/apple-disease-detection-jspkx.

[13] TAN M X，PANG R M，LE Q V.EfficientDet：scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2020：10778-10787.

[14] TANG F L，XU Z X，HUANG Q M，et al.DuAT：dual-aggregation transformer network forMedical image segmentation[C]//Pattern Recognition and Computer Vision. Singapore：Springer，2024：343-356.

[15] REZATOFIGHI H，TSOI N，GWAK J，et al.Generalized intersection over union：a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach：IEEE，2020：658-666.

[16] ZHANG H，XU C，ZHANG S J.Inner-IoU：more effective intersection over union loss with auxiliary bounding box[PP/OL]. 2023：arXiv：2311.02877 [2025-10-10]. https://arxiv.org/abs/2311.02877.

[17] ZHANG H，ZHANG S J. Shape-IoU：more accurate metric considering bounding box shape and scale[PP/OL]. arXiv，2023：2312.17663 [2025-10-10]. https://arxiv. org/abs/2312.17663.

[18] XIAO J H，GUO H Q，ZHOU J，et al. Tiny object detection with context enhancement and feature purification[J/OL]. Expert systems with applications，2023，211：118665[2025-10-10]. https://doi.org/10.1016/j.eswa.2022.118665.

[19] ZHENG Z H，WANG P，LIU W，et al. Distance-IoU loss：faster and better learning for bounding box regression[J].Proceedings of the AAAI conference on artificial intelligence，2020，34(7)：12993-13000.

[20] SELVARAJU R R，COGSWELL M，DAS A，et al. Grad-CAM：visual explanations from deep networks via gradient-based localization[J].International journal of computer vision，2020，128(2)：336-359.

Apple disease detection algorithm based on improved RT-DETR

XIAO Yongshuai, CHEN Ruobing, SHI Lei, ZHENG Guang, YIN Fei

College of Information and Management Science, Henan Agricultural University, Zhengzhou 450046, China

Abstract To address the challenges of strong background interference, difficulty in recognizing multi-scale lesions, and the need for lightweight deployment in apple disease detection within complex orchard environments, an improved lightweight apple disease detection model named EGA-DETR was proposed based on RT-DETR. The model was enhanced in three key aspects. First, an RGLAN lightweight feature aggregation module was designed to reduce redundant computations and enhance feature representation through feature splitting and re-parameterized convolution. Second, a BiFPN-GLSA multi-scale feature fusion module was developed to improve the representation of lesions at various scales via bidirectional feature transmission and a global - local self-attention mechanism. Third, an Inner-Shape-IoU loss function was introduced to improve model's localization accuracy for irregularly shaped lesion targets. Experimental results showed that EGA-DETR achieved 91.8% precision, 88.9% recall, and 92.1% mAP_{50} on the apple disease dataset, with improvements of 3.5, 3.5, and 1.5 percentage points over the baseline RT-DETR-18, respectively. Meanwhile, the number of model parameters was reduced to 11.8×10^6, which is 40.4% fewer than that of the baseline model, and the inference speed reached 120 frames per second. In summary, the EGA-DETR model achieved a favorable balance between detection accuracy and computational efficiency, providing robust technical support for accurate and real-time apple disease detection.

Keywords apple disease detection; RT-DETR; lightweight feature aggregation; bidirectional feature pyramid network; global-local self-attention; Inner-Shape-IoU

（责任编辑：边书京）