

谢家兴,廖飞,王卫星,等.基于改进Faster R-CNN的荔枝病虫害检测[J].华中农业大学学报,2025,44(1):62~73.  
DOI:10.13300/j.cnki.hnlkxb.2025.01.007

## 基于改进 Faster R-CNN 的荔枝病虫害检测

谢家兴<sup>1,2</sup>,廖飞<sup>1</sup>,王卫星<sup>3</sup>,高鹏<sup>1,2</sup>,胡凯<sup>1</sup>,吴佩文<sup>1</sup>,邓钲奇<sup>1</sup>,刘洪山<sup>1,2</sup>

1. 华南农业大学电子工程学院(人工智能学院),广州 510642;

2. 广东省农情信息监测工程技术中心,广州 510642; 3. 华南农业大学珠江学院,广州 510900

**摘要** 针对荔枝园复杂背景下荔枝小目标病虫害检测困难的问题,提出一种基于改进Faster R-CNN的荔枝病虫害检测方法。以Faster R-CNN为基础,使用特征提取能力更优秀的Swin Transformer代替原有主干网络VGG16;通过特征金字塔网络(feature pyramid network, FPN)提升Faster R-CNN模型的多尺度特征融合能力,均衡提高每一类荔枝病虫害的识别精确率;引入感兴趣区域对齐(region of interest align, ROI Align)策略提升模型的候选框定位精度,进一步提升模型的整体检测效果。结果显示,改进后的模型平均精度均值达到92.76%,相较原始Faster R-CNN检测器提升了30.08个百分点,在5类荔枝病虫害图像(藻斑病、炭疽病、煤烟病、毛毡病、叶瘿蚊)中的检测精度分别为93.05%、94.81%、96.57%、87.03%和92.34%,平均精度均值比SSD512、RetinaNet、EfficientDet-d0和YOLOv5s模型分别提高了20.50、5.70、13.08和3.26个百分点。结果表明,改进后的Faster R-CNN模型能准确检测复杂背景下的荔枝病虫害目标,具有较高的应用价值,能为农作物病虫害快速、准确识别研究提供参考。

**关键词** 荔枝; 病虫害检测; Faster R-CNN; Swin Transformer; 多尺度特征融合

**中图分类号** TP391.4; S436.629    **文献标识码** A    **文章编号** 1000-2421(2025)01-0062-12

荔枝口感美味,含有丰富的微量元素以及蛋白质,营养价值高,深受人们的青睐<sup>[1-2]</sup>,市场对荔枝的需求量也快速增加<sup>[3]</sup>。荔枝产地主要分布于两广、海南、福建等南方省份,多栽培于气候湿润、光热资源丰富、降水充沛地区,湿热的气候为荔枝病虫害的泛滥提供了绝佳的条件,荔枝病虫害成为制约荔枝产量提升的重要因素<sup>[4-5]</sup>。荔枝病虫害种类繁多、防治困难,传统病虫害检测主要依赖于人工依据经验进行现场观察与判定,存在判定效率低、成本高以及受主观因素影响大等问题,无法满足现代农业大规模、实时检测的需求。近年来,深度学习逐渐被引入到农作物的病虫害检测领域,该领域迅速发展并涌现出大量的研究成果,为我国农业病虫害精准防控提供了坚实的理论基础<sup>[6]</sup>。深度学习方法主要包含基于卷积神经网络(convolutional neural network, CNN)的图像分类方法和目标检测方法。CNN能够

自动从图像中提取特征,进而实现对农作物病虫害的高效分类识别。Waheed等<sup>[7]</sup>提出一种基于改进DenseNet架构的分类模型,对3类玉米病害的分类准确率达到98.06%。Chen等<sup>[8]</sup>提出LeafNet茶树病害检测模型,通过卷积神经网络提取茶叶病害特征,并使用支持向量机(support vector machine, SVM)和多层次感知机(multi layer perceptron, MLP)进行分类,7类茶叶病害分类准确率达到90.16%。谢家兴等<sup>[9]</sup>基于ShuffleNet V2模型,嵌入轻量型通道注意力模块ECA,增进特征图之间相互依赖关系,对荔枝病虫害的分类准确率达到了99.04%。以上研究采用深度学习对农作物病虫害进行分类识别,分类网络主要用于确定背景简单的输入图像属于哪一类预定义类别,无法对复杂图像中的目标物体进行定位。

自然环境下的检测任务中,荔枝病虫害图像背景往往是复杂的、非结构的,并且图像中往往存在多

收稿日期:2024-09-11

基金项目:教育部产学合作协同育人项目(230702595161723);国家现代农业产业技术体系专项(CARS-32-09);华南农业大学新农村发展研究院农业科技合作共建项目(2021XNYNYKJHZGJ032);广东省重点领域研发计划项目(2023B0202090001);广东省现代农业产业技术创新团队建设专项(2023KJ108);广东省科技创新战略专项资金项目(pdjh2023a0074);大学生创新创业训练计划项目(S202410564127)

谢家兴,E-mail:xjx1998@scau.edu.cn

通信作者:刘洪山,E-mail:hugouliu@scau.edu.cn

种病虫害目标。随着计算机视觉技术的发展,对输入图像中病虫害目标同时进行分类和定位的目标检测方法,能够对农作物病虫害进行更加高效且精确的诊断防控。徐会杰等<sup>[10]</sup>基于YOLOv3算法提出了YOLOv3-corn改进模型,相较于原始YOLOv3精度均值提升了4.03个百分点。王卫星等<sup>[11]</sup>采用轻量化卷积模块Ghost Module对荔枝病虫害目标进行检测,检测精度达到89.76%,相比原始YOLOv4提升4.13个百分点。何颖等<sup>[12]</sup>使用加权双向特征融合技术对YOLOv5算法进行改进,并对20种经济林木虫害进行检测,最终平均精度值达到了92.3%。Xiao等<sup>[13]</sup>提出一种改进的轻量级网络YOLOv7-MGPC,使用轻量级的主干网络GhostNet v1,添加CBAM注意力机制和通道剪枝策略,荔枝毛毡病、叶瘿蚊、炭疽病和煤烟病的平均检测精度均值达到88.6%。以上病虫害检测研究是基于单阶段目标检测算法YOLO进行的,与之相比,双阶段目标检测算法Faster R-CNN在第一阶段通过生成目标候选区域,有效减少了大量低质量背景目标的干扰,从而降低了少样本图像的误检率。Ghoury等<sup>[14]</sup>在Faster R-CNN算法的基础上使用Inception v2作为骨干网络,加强了模型的特征提取能力,并得到了95.37%的分类准确率。姜晟等<sup>[15]</sup>针对茶园复杂背景下茶叶叶部病害识别较为困难的问题,基于Faster R-CNN模型,通过优化区域建议框的尺寸、更换骨干网络和采用RS损失函数,茶叶病虫害检测的平均精度均值达到了88.06%。Faster R-CNN是一种卷积神经网络,依赖于不同尺寸的卷积核感知图像中局部像素间的关系。随着卷积层的增加,网络能够捕捉到更广泛的像素特征,但这种方法会减少网络对图像特征长期依赖性的捕捉。

Transformer模型依托于自注意力机制,能够精准捕捉图像中的长距离依赖关系,在目标检测任务中展现出卓越的建模能力,因此ViT(Vision Transformer)和DETR(Detection Transformer)先后被提出并分别应用于图像分类和目标检测任务。蒲攀等<sup>[16]</sup>利用Transformer在自然语言处理领域的特性,提高苹果生产领域实体识别的准确性, $F_1$ 值达到了92.66%。杨信廷等<sup>[17]</sup>构建基于Swin Transformer与GRU的番茄成熟度时序预测模型,对番茄成熟度正识别率达到了95.78%。王杨等<sup>[18]</sup>在标准ViT模型基础上加入增强分块序列化和掩码多头注意力机制,在PlantVillage数据集的番茄子集中得到99.63%

的分类准确率。以上基于Transformer的农作物病虫害识别研究大多基于实验室背景,对复杂背景下(如杂草、树枝等干扰物)的病虫害目标检测研究相对较少,并且纯Transformer架构(如DETR)目标检测模型对小目标的检测性能较差<sup>[19]</sup>。

荔枝病虫害图像背景复杂,病斑面积较大但数量较少的样本与病斑面积较小但数量较多的样本并存,部分类别的病斑与雨水或光斑相似,极易造成病虫害斑点的误检与漏检。针对以上问题,本研究结合双阶段目标检测算法Faster R-CNN与Transformer的特性,提出基于改进Faster R-CNN的荔枝病虫害检测方法,以期为林间精准监测荔枝病虫害提供参考。

## 1 材料与方法

### 1.1 荔枝病虫害数据采集

本研究的荔枝病虫害图像采集地点为广东省农业技术推广总站,采集时间为2021年8月中旬和2022年4月下旬,图像采集设备为Nikon D7200相机和高清智能手机,相机传感器尺寸约22.3 mm×14.9 mm,像素为1 800万像素,设置相机分辨率为4 032像素×3 016像素和3 000像素×3 000像素。为真实还原荔枝园复杂环境,分别在晴天、雨后和阴天3种不同天气采集数据,在自然光下对荔枝病虫害叶片进行东、南、西、北4个方向拍摄,拍摄距离为0.3~0.5 m。在荔枝病虫害专家的指导下完成数据集分类,数据集共包含5类荔枝病虫害图像,病害种类包括藻斑病、煤烟病和炭疽病,虫害包括毛毡病和叶瘿蚊,数据集图像总数为4 093。数据集样例如图1所示。

### 1.2 荔枝病虫害数据集构建

采用可视图像标注工具LabelImg手动标注图像中的荔枝病虫害目标,标注格式为COCO格式,标注过程中不会标注接近图像边缘的病虫害目标。为增强模型的鲁棒性,本研究在每次迭代训练期间使用随机缩放、随机旋转和高斯模糊3种方法对输入图像进行增强。训练集、验证集和测试集按照8:1:1的比例划分。数据集的5类样本划分情况见表1。

### 1.3 Faster R-CNN模型

如何更准确地提取小目标类别病斑特征和回归边界框,进而准确识别已知病虫害类别与发病位置是本文的研究重点。随着R-CNN<sup>[20]</sup>和Fast R-CNN<sup>[21]</sup>的发展,Faster R-CNN模型将特征提取网

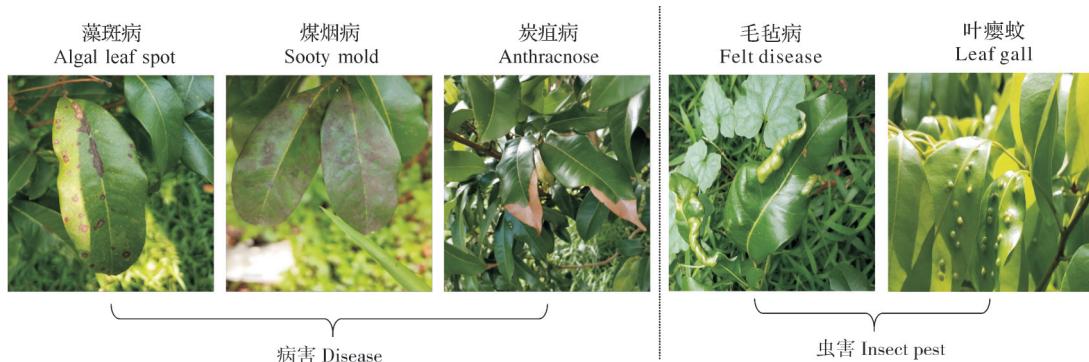


图1 病虫害实例

Fig. 1 Case of diseases and insect pests

表1 荔枝病虫害数据集

Table 1 Data set of litchi diseases and pests

类别 Category	训练集 Train set	验证集 Validation set	测试集 Test set	合计 Total set
毛毡病 Felt disease	686	86	86	858
煤烟病 Sooty mold	633	79	79	791
炭疽病 Anthracnose	623	78	78	779
藻斑病 Algal leaf spot	638	80	80	798
叶瘿蚊 Leaf gall	694	86	87	867
合计 Number of images	3 274	409	410	4 093

络、区域候选网络(region proposal network, RPN)、感兴趣区域池化(region of interest pooling, ROI Pooling)和区域卷积神经网络(regions with CNN features, RCNN)都整合在了一个网络中,综合性能有较大提升。Faster R-CNN算法的主要流程为:将原图输入特征提取网络VGG16<sup>[22]</sup>得到一系列特征图;RPN网络在原图上生成多个不同宽高比例的锚框,将锚框映射到特征图并选取特定数量的锚框,得到区域建议框;RCNN网络将区域建议框映射到特征图层,得到感兴趣区域,下采样层将不同大小的ROI全部调整为统一尺寸(7像素×7像素),最后送入回归器与分类器,分别得到预测框的位置坐标值与类别置信度。Faster R-CNN模型的整体结构如图2所示。

#### 1.4 改进Faster R-CNN模型

1) 主干网络。2017年,由Google研究团队提出的Transformer模型<sup>[23]</sup>采用通用的编码器-解码器网络架构。伴随Transformer在NLP领域中的研究愈发深入,Vision Transformer<sup>[24]</sup>开创了Transformer在图像处理领域有效工作的先河。Swin Transform-

er<sup>[25]</sup>在ViT的基础上,通过分层结构提取多尺度信息,降低了计算复杂度,在多目标检测方面优于ViT。

如图3所示,Swin Transformer通过分块和滑动窗口机制处理特征图,实现特征图尺寸减半和深度翻倍,逐层扩大感受野以获取全局信息。各阶段输出特征图尺寸分别为 $\frac{H}{4} \times \frac{W}{4} \times C$ , $\frac{H}{8} \times \frac{W}{8} \times 2C$ , $\frac{H}{16} \times \frac{W}{16} \times 4C$ , $\frac{H}{32} \times \frac{W}{32} \times 8C$ 。其中H与W表示输入图片的高度和宽度,C表示通道数(本研究设为96)。核心模块Swin Transformer Block将Vision Transformer中的多头自注意力模块(multi-head self-attention, MSA)改进为基于窗口的多头自注意力模块(window-based multi-head self-attention, W-MSA)和基于移位窗口的多头自注意力模块(shifted window multi-head self-attention, SW-MSA),输入特征图依次通过归一化层(layer normalization, LN)、W-MSA或者SW-MSA后,继续通过归一化层和多层次感知机(multi layer perceptron, MLP),实现跨窗口的交互,获得全局建模的能力,能够为后续处理提供多尺度特征。

2) 特征金字塔网络。本研究数据中的荔枝藻斑病与叶瘿蚊的病斑面积小,直径通常不超过5 mm,而荔枝炭疽病、煤烟病和毛毡病病斑相对较大,针对小目标和多尺度检测难点问题,本研究引入Lin等<sup>[26]</sup>提出的FPN改进Faster R-CNN模型。

如图4所示,FPN通过在网络中添加自下而上、自上而下和中间横向连接的3条线路,将特征提取网络不同层级提取的特征图进行融合,构建了金字塔型结构的特征表示。自下而上线路为骨干网络自选定的特定深度层级产生的一系列不同尺度与分辨率的特征图,底层特征图分辨率较大、语义信息较少,随着特征提取网络层级加深,高层次特征图语义信

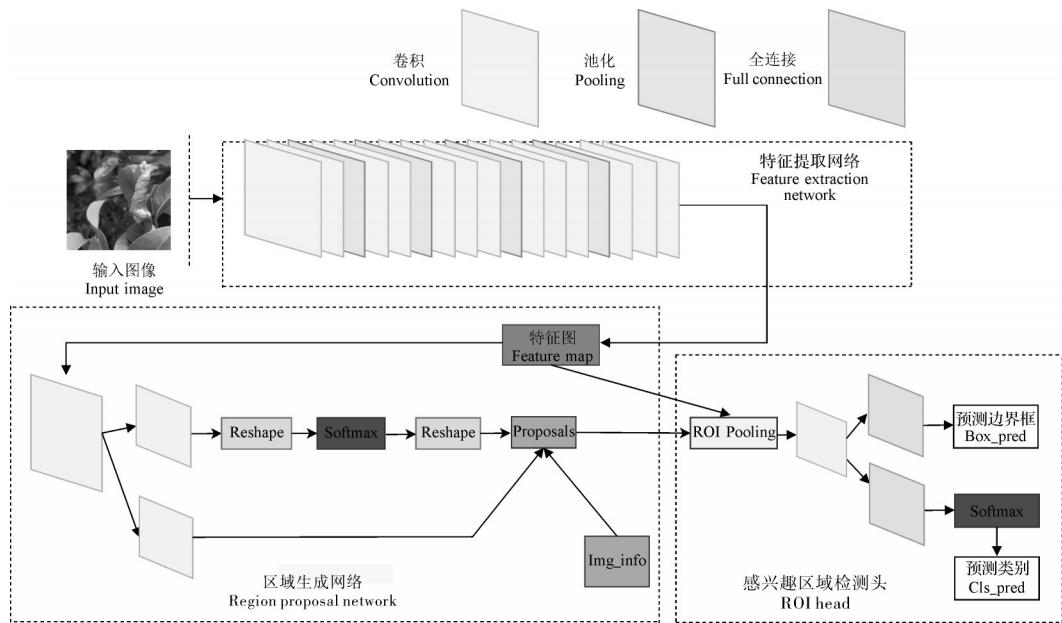
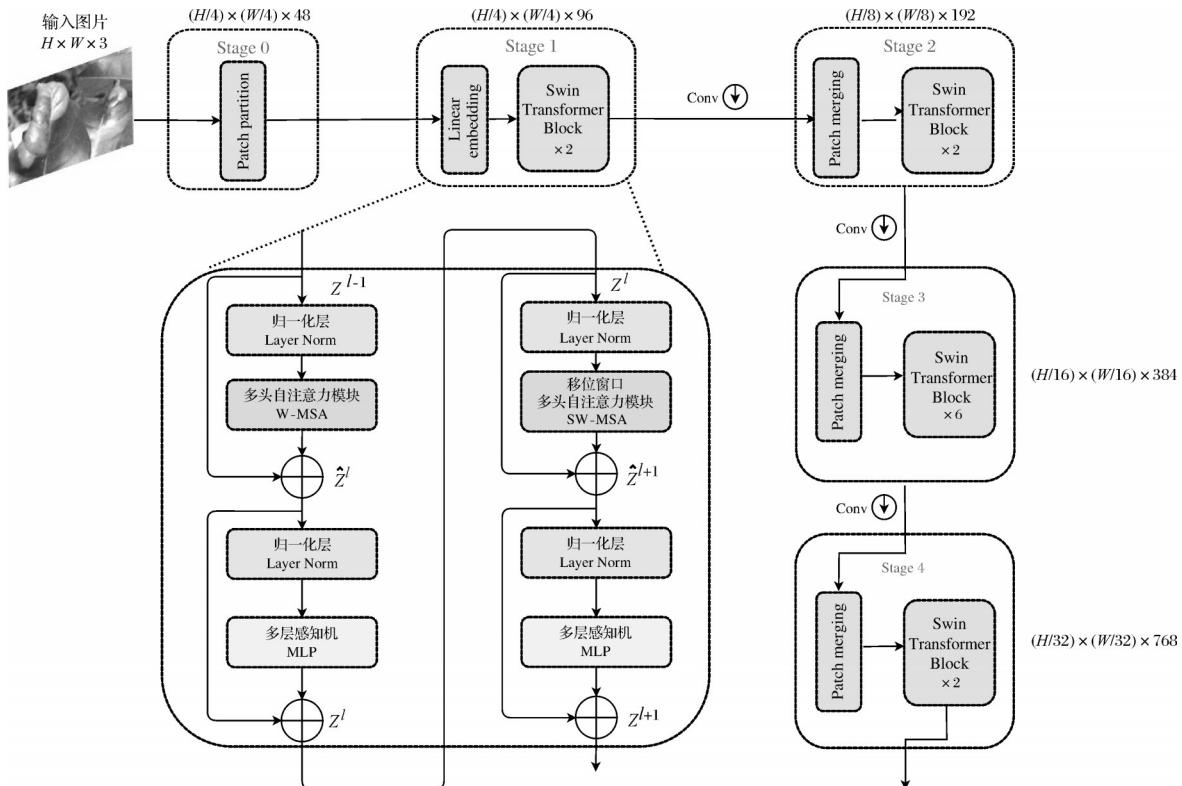


图2 Faster R-CNN模型结构

Fig. 2 Model structure of Faster R-CNN



$H$  和  $W$  分别表示高度和宽度  $H$  and  $W$  represent height and width respectively; Patch partition 表示块分割操作 Patch partition represents block segmentation operation; Linear embedding 表示线性嵌入操作 Linear embedding represents linear embedding operation;  $\oplus$  表示矩阵逐元素相加操作  $\oplus$  represents matrix element addition operation.

图3 Swin Transformer模型结构

Fig. 3 Model structure of Swin Transformer

息越来越丰富,分辨率越来越低,含小目标的信息也会降低。自上而下部分从顶层特征图开始,通过上采样操作使其与相邻一层特征图尺寸相匹配,随后进行逐元素相加,这一部分操作能将细节特征与上下文特征相融合,新得到的特征图兼顾小目标信息与高语义信息,使之后的预测网络检测效果大大提升。

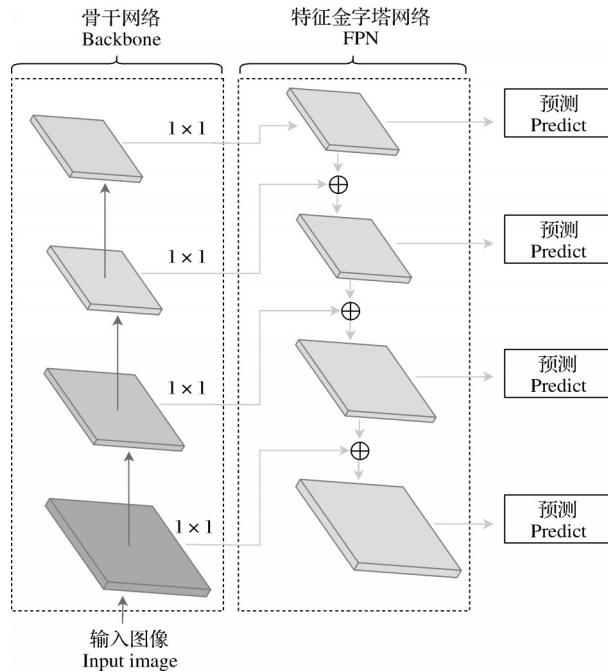


图4 特征金字塔网络结构

Fig. 4 Feature pyramid network architecture

3) 感兴趣区域对齐。ROI Pooling 涉及 2 次量化:第1次将 ROI 坐标从浮点数量化为整数,第2次在 ROI 划分子区域时向下取整。这导致 Faster R-CNN 得到的候选框与初始位置产生了一定的偏差,这对小目标检测效果影响很大。本研究借鉴 Mask R-CNN,用 ROI Align 改进网络以提高小目标检测效果。

ROI Align 机制的运作流程可概括为 3 个关键步骤,如图 5 所示:首先将预测的候选区域映射至相应的特征层,并对此进行详尽的遍历;每个候选区域被均匀地划分成大小为  $k \times k$  的多个小网格,且在此过程中边界点的坐标保持其精确性,避免了量化过程中可能出现的误差;每个小网格内,通过双线性插值法精确计算出 4 个关键采样点的数值,这些数值随后被用于执行最大池化操作,以确定该网格的综合特征表示。通过整合 ROI Align 策略,模型在处理候选区域时不仅能够适应统一的输入尺寸,还能实现更为精细的定位精度,增强模型对小病虫害目标的识别能力。

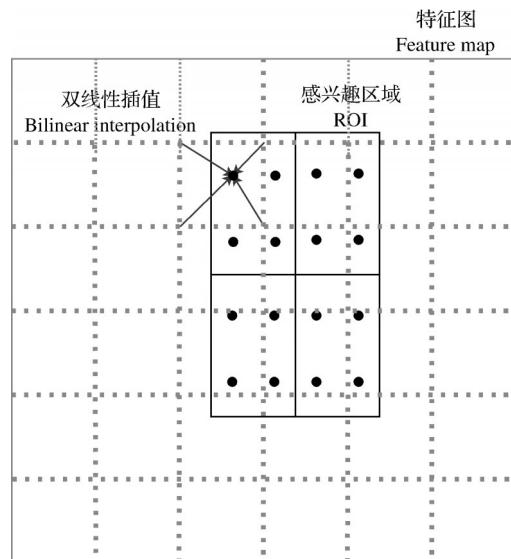


图5 感兴趣区域对齐原理

Fig. 5 Principle of ROI Align

为了提高 Faster R-CNN 对复杂背景下的荔枝小目标病虫害检测精度,本研究使用 Swin Transformer 替代原始 Faster R-CNN 所使用的 VGG16 特征提取骨干网络;通过特征金字塔网络融合高级特征与低级特征,丰富特征图信息;引入 ROI Align 替代原始 Faster R-CNN 模型的 ROI Pooling 策略,提升模型的候选框定位精度。改进后的 Faster R-CNN 模型结构如图 6 所示。

### 1.5 实验设置与评价指标

1) 实验平台与训练超参数设置。本研究平台基于操作系统为 Ubuntu18.04(x86-64bit)的服务器,该服务器 CPU 为 Inter(R) Xeon(R) Gold 6240 CPU@2.59 GHz, GPU 为 2 块 NVIDIA GeForce GTX 3090,显存为 24 GB,运行内存 192 GB。编程语言选择 Python 3.8.10,深度学习框架选择 Pytorch 1.9.0, CUDA 驱动选择 11.1,Cudnn 版本选择 8.5.0。

为了提升模型训练的效率和稳定性,本研究采用了迁移学习方法,使用预训练的 Swin Transformer 骨干网络权重。训练前对所有待处理图像尺寸标准化至 640 像素  $\times$  640 像素,并采取随机缩放、旋转和添加高斯噪声 3 种数据增强措施,以提高模型鲁棒性。在训练阶段,设定初始学习率 0.002,批量大小设置为 64,配合 Adam 优化器执行模型的反向传播。此外,模型在每完成 250 次迭代后,使用 0.5 的交并比(intersection over union, IoU)阈值对验证集进行性能评估,并通过非极大值抑制(non-maximum suppression, NMS)预测结果,其中 NMS 的 IoU 阈值设定

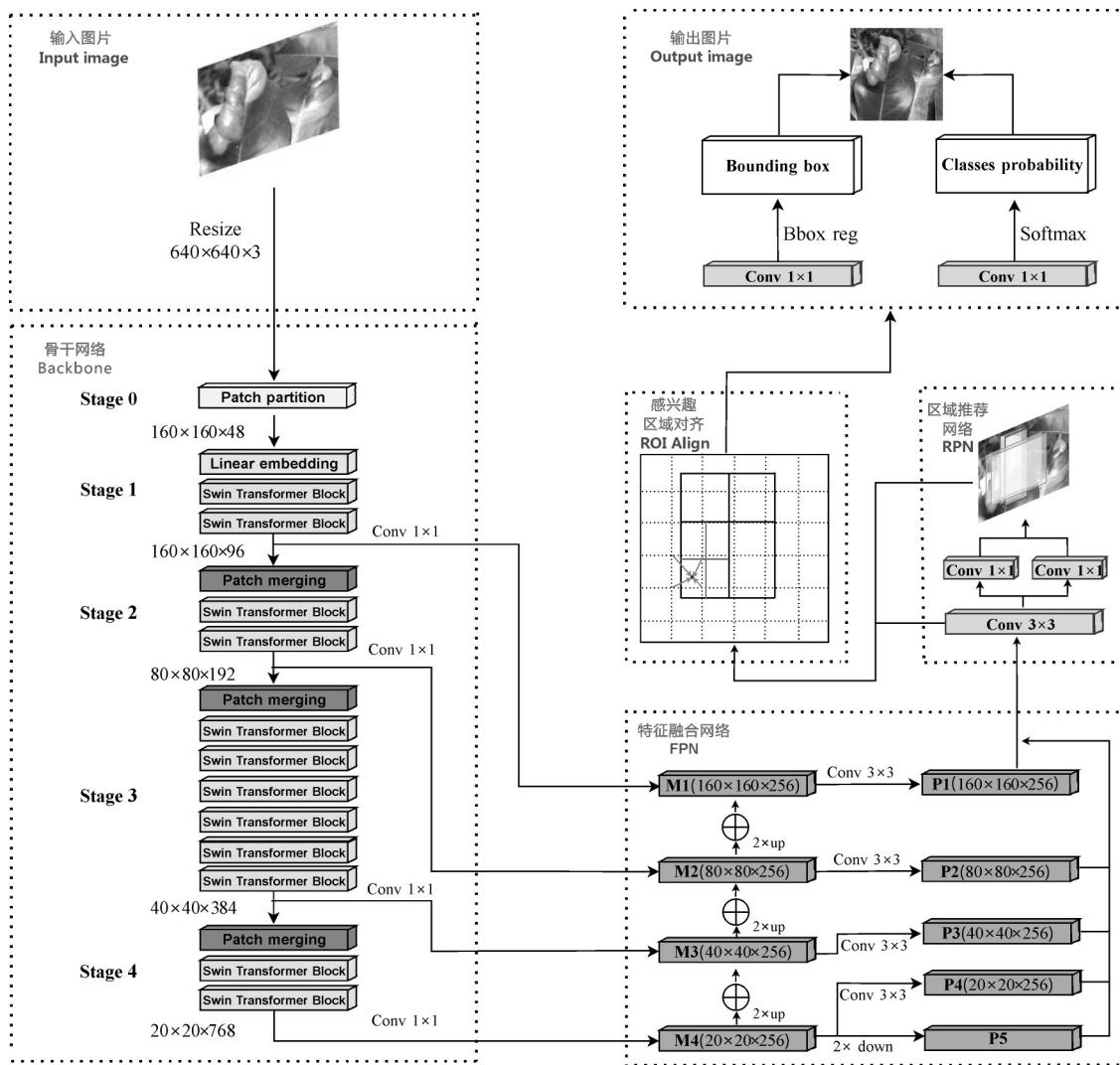


图6 改进后 Faster R-CNN 完整结构

Fig. 6 Improved complete structure of Faster R-CNN

为0.5。

2) 评价指标。本研究使用平均精度(average precision,  $P_A$ )作为模型在各类别荔枝病虫害数据集上的评价指标。平均精度与模型的精确率(precision,  $P$ )和召回率(recall,  $R$ )有关, 在给定一个特定检测类别以及IoU阈值下, 平均精度表示为精确率与召回率的曲线面积之和, 进一步可以得到各类别 $P_A$ 的平均精度均值(mean average precision,  $P_{mA}$ )。其中, 精确率、召回率和平均精度的定义如式(1)~(3)所示。

$$P_A = \int_0^1 P(R) dR \times 100\% \quad (1)$$

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (2)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (3)$$

式(1)~(3)中,  $T_p$ 表示预测为正例的正样本数量;  $F_p$ 表示预测为正例的负样本数量;  $F_N$ 表示预测为负例的正样本数量。

## 2 结果与分析

### 2.1 改进前后 Faster R-CNN 模型检测结果

为验证提出的改进方法对5类荔枝病虫害的实际检测效果, 本研究将改进Faster R-CNN模型与原始Faster R-CNN模型基于相同测试集进行测试。由表2可知, 原始Faster R-CNN模型对藻斑病和叶瘿蚊2类小目标检测效果较差, 这是由于原模型的骨干网络-VGG16的特征提取能力较差, 容易造成小目标区域漏检。改进Faster R-CNN模型将骨干网络更换为Swin Transformer并加入特征金字塔网络后, 小目标类别(藻斑病和叶瘿蚊)的平均检测精度分别

提升了60.12、71.24百分点,病斑面积较大类别(炭疽病、煤烟病和毛毡病)平均检测精度分别提升了3.39、9.11和6.52个百分点,整体精度提升30.08个百分点,以上结果表明改进Faster R-CNN模型提升了小目标病虫害检测精度。

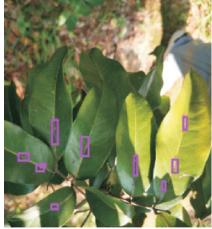
为直观感受改进后的模型对荔枝病虫害图像的检测效果,分别使用改进前后的模型对比检测了晴天阳光直射、晴天阳光非直射、阴天、雨天4种环境下的荔枝病虫害图像。由图7可知,改进前后的Faster

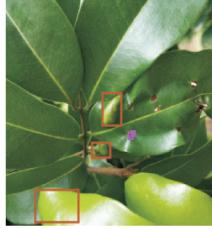
R-CNN模型均能很好地检测出大尺度病斑所属类别(煤烟病和炭疽病),但在检测大尺度病斑目标时仍存在预测框回归效果差以及预测框位置与病斑实际位置存在较大偏差的问题;针对面积小、数量多的病斑(藻斑病和叶瘿蚊病)检测效果,改进Faster R-CNN模型未出现漏检和错误识别的问题。以上结果表明,改进Faster R-CNN模型具有更优的抗干扰能力、小目标检测能力和预测框回归能力,能更准确地识别复杂环境下荔枝病虫害目标。

表2 改进前后模型的性能对比

Table 2 Performance comparison of the model before and after improvement

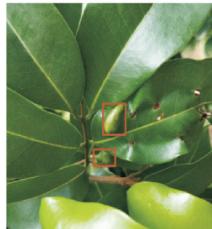
模型 Model	病害平均精度 Disease average precision			虫害平均精度 Insect average precision		$P_{mA}$
	藻斑病 Algal leaf spot	炭疽病 Anthracnose	煤烟病 Sooty mold	毛毡病 Felt disease	叶瘿蚊 Leaf gall	
基线模型 Baseline model	32.93	91.42	87.46	80.51	21.10	62.68
改进模型 Improved model	93.05	94.81	96.57	87.03	92.34	92.76




原始Faster R-CNN检测效果  
Original Faster R-CNN detection effect




改进Faster R-CNN检测效果  
Improved Faster R-CNN detection effect

藻斑病  
Algal leaf spot
煤烟病  
Sooty mold
炭疽病  
Anthracnose
毛毡病  
Felt disease
叶瘿蚊  
Leaf gall

图7 改进前后模型对每类荔枝病虫害的检测效果

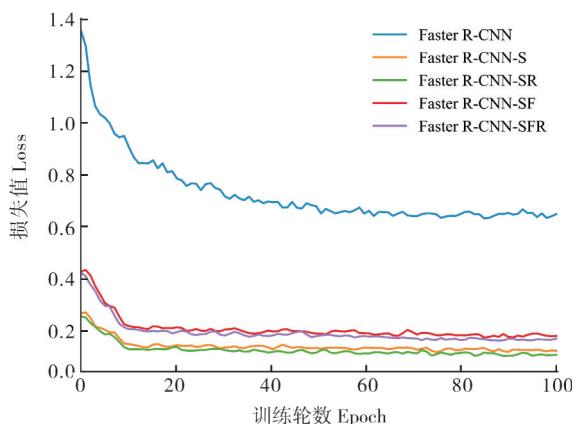
Fig. 7 Detection effect of improved models on each kind of litchi diseases and insect pests

## 2.2 消融实验

为验证Swin Transformer骨干网络、特征金字塔网络和感兴趣区域对齐策略对原始Faster R-CNN模型带来的性能提升效果,本研究在统一测试集和损失函数的基准下,设置了一系列消融实验。由图8可知,Swin Transformer骨干网络能显著降低模型损失值,且能更快收敛,在第10个训练轮次左右,各消融实验模型损失曲线基本收敛,而原始Faster R-

CNN模型则在第50个训练轮次左右,损失值才趋向稳定,且在后续的训练过程中依然出现较大的波动,模型拟合效果较差。原始Faster R-CNN模型添加了FPN和ROI Align后,训练过程中收敛曲线更加稳定,降低了模型损失绝对值,整体提升了精度。

由表3可知,将特征提取网络由原来的VGG16替换成Swin Transformer后,精确率和平均精度均值分别提升30.02、11.82个百分点,精确率提升效果显著,



S, F, R 分别代表原始模型添加 Swin Transformer、FPN 和 ROI Align。S, F and R respectively represent that the original model is added with Swin Transformer, FPN and ROI Align.

图8 消融实验训练损失值

Fig. 8 Training loss results of ablation experiments

说明 Swin Transformer 骨干网络能有效提高模型的特征提取能力,丰富特征图信息;融合 FPN 后模型的精确率、召回率和平均精度均值分别提升了 38.45、17.52 和 26.52 百分点,模型多尺度和小目标检测能力获得提升;综合图 8 与表 3 可得出,融合 FPN 前后,

将 Faster R-CNN 模型的池化策略更改为 ROI Align,模型损失函数绝对值虽无太大区别,但召回率分别提升了 3.71、3.41 百分点,说明 ROI Align 能增强模型整体预测框回归效果,进而使预测框更精准地覆盖荔枝病斑。添加改进策略后的 Faster R-CNN 相较于原始模型,  $P_{mA}$  提高了 30.08 百分点,  $F_1$  分数提高了 37.71 百分点,说明这些改进策略显著提升了 Faster R-CNN 模型的检测性能。

### 2.3 与其他模型检测效果对比

为综合对比改进 Faster R-CNN 模型与其他不同模型对荔枝病虫害的检测效果,在相同训练环境和超参数设置下,将当前主流目标检测模型 SSD512、YOLOv5s、EfficientDet-d0 和 RetinaNet 与本研究提出的改进 Faster R-CNN 模型进行测试对比。

由表 4 可知,改进 Faster R-CNN 模型获得了最高的召回率、 $F_1$  分数和平均精度均值,其中平均精度均值比 SSD512、YOLOv5s、EfficientDet-d0 和 RetinaNet 分别高出 20.50、3.26、13.08 和 5.70 百分点,召回率分别高出 19.52、8.65、10.67 和 7.32 百分点,证明了改进的 Faster R-CNN 模型能很好地提取有效特征,准确预测了病斑位置坐标与类别信息。

表3 消融实验结果

Table 3 Ablation results

骨干网络 Backbone	特征金字塔网络 FPN	感兴趣区域对齐 ROI Align	精确率/% $P$	召回率/% $R$	$F_1$ 分数 $F_1$ score	平均精度均值/% $P_{mA}$
VGG16	×	×	38.44	72.64	50.01	62.68
Swin Transformer	×	×	68.46	78.37	73.08	74.50
Swin Transformer	×	√	75.30	82.08	78.54	79.00
Swin Transformer	√	×	76.89	90.16	83.20	89.20
Swin Transformer	√	√	82.57	93.57	87.72	92.76

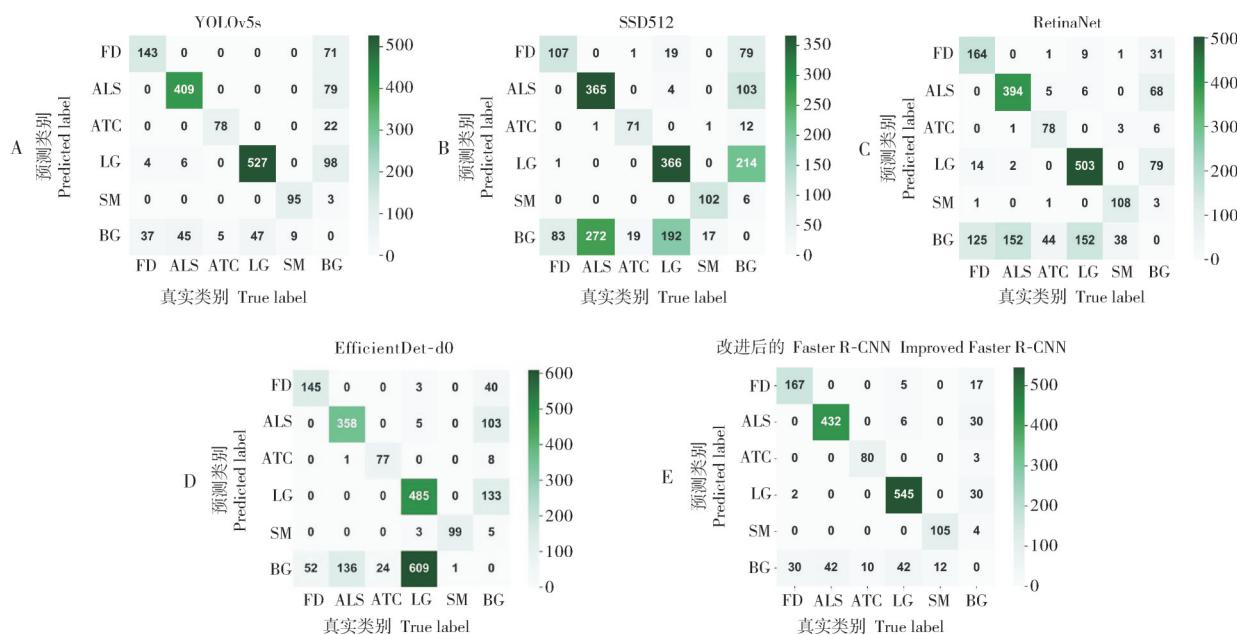
表4 不同模型的性能对比

Table 4 Performance comparison of different models

模型 Model	骨干网络 Backbone	精确率/% $P$	召回率/% $R$	$F_1$ 分数 $F_1$ score	平均精度均值/% $P_{mA}$
SSD512	VGG16	67.85	74.05	70.81	72.26
YOLOv5s	CSPDarknet53	87.33	84.92	86.05	89.50
EfficientDet-d0	EfficientNet	72.61	82.90	77.41	79.68
RetinaNet	ResNet18	66.65	86.25	75.19	87.06
改进 Faster R-CNN	Swin Transformer	82.57	93.57	87.72	92.76
Improved Faster R-CNN					

由图 9 可知,改进 Faster R-CNN 模型在 5 类荔枝病虫害类别上,均获得最多的正确预测框。在 5 类荔枝病虫害中,藻斑病和叶瘿蚊病斑数目多且面积小,SSD512、RetinaNet 和 EfficientDet-d0 三种模型对藻斑病和叶瘿蚊检测效果较差,容易将病斑漏检为

背景类别,或者误检为其他类别;YOLOv5s 模型和改进 Faster R-CNN 模型对 5 类病虫害检测效果较好,但改进 Faster R-CNN 模型拥有更低的漏检率,而 YOLOv5s 模型相对更容易将背景物体检测为病斑,从而造成误检。



FD, ALS, ATC, LG, SM, BG 分别代表荔枝毛毡病、藻斑病、炭疽病、叶瘿蚊、煤烟病、背景类别。FD, ALS, ATC, LG, SM, BG represents felt disease, algal leaf spot, anthracnose, leaf gall, sooty mold and background category, respectively.

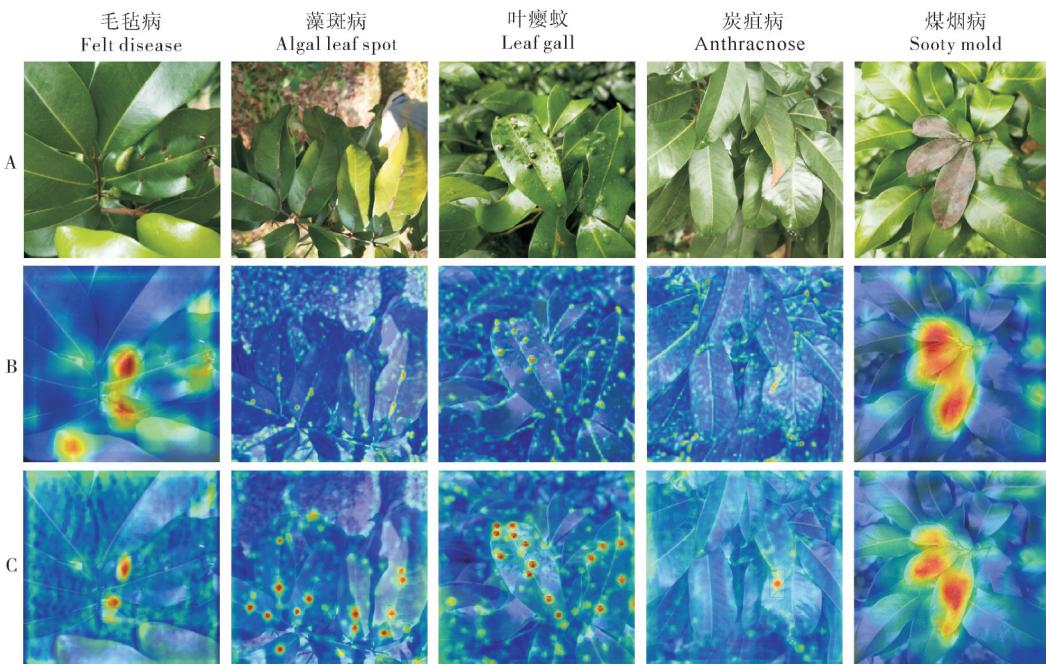
图9 不同模型的混淆矩阵

Fig. 9 Confusion matrices for different models

## 2.4 改进前后模型的热力图

本研究使用Grad-CAM (gradient-weighted class activation mapping)方法对改进 Faster R-CNN 模型进行热力图可视化, 颜色越红表示该区域对模型影响越大。

应越高。由图10可知, 本研究改进的 Faster R-CNN 模型的检测热力图更贴近真实病虫害目标, 能精准注意到阳光直射环境和雨天环境下的小目标藻斑病斑和叶瘿蚊虫斑特征。



A:荔枝病虫害图像 Images of litchi pests and diseases; B:原始 Faster R-CNN 模型检测热力图 Heatmaps of Faster R-CNN model detection; C:改进 Faster R-CNN 模型检测热力图 Heatmaps of improved Faster R-CNN model detection.

图10 改进前后模型检测热力图

Fig. 10 Heatmaps of model detection before and after improvement

### 3 讨 论

本研究针对复杂环境下的荔枝病虫害小目标检测难点,以5种不同荔枝病虫害图像作为研究对象,使用实地采集的4 093张图像,构建荔枝病虫害数据集,在原始Faster R-CNN模型的基础上,使用Swin Transformer作为模型的骨干网络,采用特征金字塔网络和感兴趣区域对齐策略改进模型。改进Faster R-CNN模型在训练集、验证集和测试集上的检测平均精度均值分别为97.70%、93.30%和92.76%,表明模型进行了很好的收敛且具有很好的泛化性;改进Faster R-CNN模型在毛毡病、炭疽病和煤烟病的检测精度分别为87.03%、94.81%和96.57%,在小目标类别(藻斑病和叶瘿蚊)的检测精度为93.05%和92.34%,对比改进前分别提升60.12和71.24个百分点,小目标检测能力获得大幅提升。

本研究提出的改进Faster R-CNN模型相比常见目标检测模型,很好地解决了荔枝病虫害斑的多尺度检测问题与小目标检测问题,具有较高的应用价值。但由于时间和资源限制,所使用数据集样本集中采集于广州地区,采集时间为4月下旬和8月中旬,存在地域局限性和时间局限性。因此在未来的研究中,需要进一步扩大样本数量和增加样本多样性,以提高研究结果的普适性和可靠性;在保持精度的情况下,设计更轻量化的模型嵌入移动端设备,为荔枝产业智能化管理提供有效参考。

### 参考文献 References

- [1] 陈剑文. 荔枝栽培管理及病虫害防治方法[J]. 吉林农业, 2018(19): 86. CHEN J W. Litchi cultivation management and pest control methods[J]. Agriculture of Jilin, 2018(19): 86 (in Chinese).
- [2] 钟谋. 荔枝和龙眼病虫害综合防治技术[J]. 乡村科技, 2018, 9(7): 101-102. ZHONG M. Integrated pest control techniques for litchi and longan[J]. Rural science and technology, 2018, 9(7): 101-102 (in Chinese).
- [3] 张敏,曾雯雯,陆耀. 荔枝种植中优质高效栽培技术研究[J]. 现代农机, 2022(5): 104-106. ZHANG M, ZENG W W, LU Y. Study on high-quality and efficient cultivation techniques in litchi planting[J]. Modern agricultural machinery, 2022(5): 104-106 (in Chinese).
- [4] 黄康泉. 荔枝主要病虫害的为害症状及防治方法[J]. 现代园艺, 2021, 44(22): 40-41. HUANG K Q. Damage symptoms and control methods of main diseases and insect pests in litchi [J]. Contemporary horticulture, 2021, 44 (22) : 40-41 (in Chinese).
- [5] 全林发,董易之,徐淑,等. 荔枝园荔枝蒂蛀虫常规爆发期的种群空间分布格局[J]. 环境昆虫学报, 2021, 43(4): 950-958. QUAN L F, DONG Y Z, XU S, et al. Spatial distribution patterns of *Conopomorpha sinensis* Bradley during its conventional population outbreak in litchi orchard [J]. Journal of environmental entomology, 2021, 43 (4) : 950-958 (in Chinese).
- [6] 邵明月,张建华,冯全,等. 深度学习在植物叶部病害检测与识别的研究进展[J]. 智慧农业, 2022, 4(1): 29-46. SHAO M Y, ZHANG J H, FENG Q, et al. Research progress of deep learning in detection and recognition of plant leaf diseases [J]. Smart agriculture, 2022, 4(1) : 29-46 (in Chinese with English abstract).
- [7] WAHEED A, GOYAL M, GUPTA D, et al. An optimized dense convolutional neural network model for disease recognition and classification in corn leaf [J/OL]. Computers and electronics in agriculture, 2020, 175: 105456 [2024-09-11]. <https://doi.org/10.1016/j.compag.2020.105456>.
- [8] CHEN J, LIU Q, GAO L W. Visual tea leaf disease recognition using a convolutional neural network model [J/OL]. Symmetry, 2019, 11(3): 343 [2024-09-11]. <https://doi.org/10.3390/sym11030343>.
- [9] 谢家兴,陈斌瀚,彭家骏,等. 基于改进ShuffleNet V2的荔枝叶片病虫害图像识别[J]. 果树学报, 2023, 40(5): 1024-1035. XIE J X, CHEN B H, PENG J J, et al. Image recognition of leaf pests and diseases based on improved ShuffleNet V2 in litchi [J]. Journal of fruit science, 2023, 40(5) : 1024-1035 (in Chinese with English abstract).
- [10] 徐会杰,黄仪龙,刘曼. 基于改进YOLOv3模型的玉米叶片病虫害检测与识别研究[J]. 南京农业大学学报, 2022, 45 (6): 1276-1285. XU H J, HUANG Y L, LIU M. Research on pest detection and identification of corn leaf based on improved YOLOv3 model[J]. Journal of Nanjing Agricultural University, 2022, 45(6) : 1276-1285 (in Chinese with English abstract).
- [11] 王卫星,刘泽乾,高鹏,等. 基于改进YOLOv4的荔枝病虫害检测模型[J]. 农业机械学报, 2023, 54(5) : 227-235. WANG W X, LIU Z Q, GAO P, et al. Detection of litchi diseases and insect pests based on improved YOLOv4 model[J]. Transactions of the CSAM, 2023, 54(5): 227-235 (in Chinese with English abstract).
- [12] 何颖,陈丁号,彭琳. 基于改进YOLOv5模型的经济林木虫害目标检测算法研究[J]. 中国农机化学报, 2022, 43(4): 106-115. HE Y, CHEN D H, PENG L. Research on object detection algorithm of economic forestry pests based on im-

- proved YOLOv5[J]. Journal of Chinese agricultural mechanization, 2022, 43 (4) : 106-115 (in Chinese with English abstract).
- [13] XIAO J Y, KANG G B, WANG L H, et al. Real-time light-weight detection of lychee diseases with enhanced YOLOv7 and edge computing[J/OL]. Agronomy, 2023, 13(12):2866 [2024-09-11]. <https://doi.org/10.3390/agronomy13122866>.
- [14] GHOURY S, SUNGUR C, DURDU A. Real-time diseases detection of grape and grape leaves using Faster R-CNN and SSD mobilenet architectures [C]//International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES 2019), Apr 26-28, 2019, Alanya, Turkey. [S.l.:s.n.], 2019:39-44.
- [15] 姜晟,曹亚范,刘梓伊,等. 基于改进Faster RCNN的茶叶叶部病害识别[J]. 华中农业大学学报, 2024, 43(5):41-50. JIANG S, CAO Y P, LIU Z Y, et al. Recognition of tea leaf disease based on improved Faster RCNN[J]. Journal of Huazhong Agricultural University, 2024, 43(5):41-50 (in Chinese with English abstract).
- [16] 蒲攀,张越,刘勇,等. Transformer优化及其在苹果病虫命名实体识别中的应用[J]. 农业机械学报, 2023, 54(6):264-271. PU P, ZHANG Y, LIU Y, et al. Transformer optimization and application in named entity recognition of apple diseases and pests [J]. Transactions of the CSAM, 2023, 54(6) : 264-271 (in Chinese with English abstract).
- [17] 杨信廷,刘彤,韩佳伟,等. 基于Swin Transformer与GRU的低温贮藏番茄成熟度识别与时序预测研究[J]. 农业机械学报, 2024, 55(3):213-220. YANG X T, LIU T, HAN J W, et al. Low temperature storage tomato maturity recognition and time series prediction based on swin transformer-GRU [J]. Transactions of the CSAM, 2024, 55(3):213-220 (in Chinese with English abstract).
- [18] 王杨,李迎春,许佳炜,等. 基于改进Vision Transformer网络的农作物病害识别方法[J]. 小型微型计算机系统, 2024, 45(4):887-893. WANG Y, LI Y C, XU J W, et al. Crop disease recognition method based on improved Vision Transformer network [J]. Journal of Chinese computer systems, 2024, 45(4):887-893 (in Chinese with English abstract).
- [19] YAO Z Y, AI J B, LI B X, et al. Efficient DETR: improving end-to-end object detector with dense prior[DB/OL]. ArXiv, 2021: 2104.01318 [2024-09-11]. <https://doi.org/10.48550/arXiv.2104.01318>.
- [20] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, USA. Santiago:IEEE, 2014:580-587.
- [21] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV) , December 7-13, 2015, Santiago, Chile. Santiago:IEEE, 2015:1440-1448.
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [DB/OL]. ArXiv, 2014: 1409.1556 [2024-09-11]. <https://doi.org/10.48550/arXiv.1409.1556>.
- [23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, USA.Long Beach:Curran Associates Inc., 2017: 6000-6010.
- [24] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth  $16 \times 16$  words: transformers for image recognition at scale[DB/OL]. ArXiv, 2020: 2010.11929 [2024-09-11]. <https://doi.org/10.48550/arXiv.2010.11929>.
- [25] LIU Z, LIN Y T, CAO Y, et al. Swin transformer:hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV) , October 10-17, 2021, Montreal, Canada. Montreal: IEEE, 2021:9992-10002.
- [26] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , July 21-26, 2017, Honolulu, USA. Honolulu: IEEE, 2017: 936-944.
- [27] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//2017 IEEE International Conference On Computer Vision, October 22-29, 2017.Venice, Italy.Venice:IEEE, 2017: 2961-2969.

## Detecting diseases and insect pests in litchi based on improved Faster R-CNN

XIE Jiaxing<sup>1,2</sup>, LIAO Fei<sup>1</sup>, WANG Weixing<sup>3</sup>, GAO Peng<sup>1,2</sup>,  
HU Kai<sup>1</sup>, WU Peiwen<sup>1</sup>, DENG Zhengqi<sup>1</sup>, LIU Hongshan<sup>1,2</sup>

1. College of Electronic Engineering (Artificial Intelligence), South China Agricultural University,  
Guangzhou 510642, China;

2. Guangdong Engineering Research Center for Agricultural Information Monitoring,  
Guangzhou 510642, China;

3. Zhujiang College, South China Agricultural University, Guangzhou 510900, China

**Abstract** A method of detecting diseases and insect pests in litchi based on improved Faster R-CNN was proposed to solve the problems of detecting small targets of diseases and pests in complex backgrounds of litchi orchards. Swin Transformer with superior capabilities of extracting feature was used to replace the original backbone network VGG16 based on Faster R-CNN. The feature pyramid network (FPN) was used to enhance capability of the multi-scale feature fusion in the Faster R-CNN model, thereby improving the precision of identifying each type of diseases and insect pests in litchi in a balanced manner. The ROI Align strategy was introduced to refine the precision of the candidate box localization in the model, leading to the enhancement in the performance of overall detection in the model. The result showed that the average accuracy of the improved model was 92.76%, 30.08 percentage points higher than that of the original Faster R-CNN detector. The precision of detecting images of five types of diseases and insect pests including algal leaf spot, anthracnose, sooty mold, felt disease, and leaf gall in litchi was 93.05%, 94.81%, 96.57%, 87.03%, and 92.34%, respectively. The average precision was improved by 20.50, 5.70, 13.08, and 3.26 percentage points compared with that of SSD512, RetinaNet, EfficientDet-d0, and YOLOv5s model, respectively. It is indicated that the improved Faster R-CNN model can accurately detect diseases and insect pests in litchi with complex backgrounds, and has high value of application. It will provide a reference for studying the rapid and accurate identification of diseases and insect pests in crops.

**Keywords** litchi; detection of diseases and insect pests; Faster R-CNN; Swin Transformer; multi-scale feature fusion

(责任编辑:葛晓霞)