

基于改进 Faster R-CNN 的苹果采摘视觉定位与检测方法

李翠明 杨柯 申涛 尚拯宇

(兰州理工大学机电工程学院, 兰州 730050)

摘要: 针对采摘机器人对场景中目标分布密集、果实相互遮挡的检测及定位能力不理想问题,提出一种引入高效通道注意力机制(ECA)和多尺度融合特征金字塔(FPN)改进 Faster R-CNN 果实检测及定位方法。首先,利用表达能力较强的融合 FPN 的残差网络 ResNet50 替换原 VGG16 网络,消除了网络退化问题,进而提取更加抽象和丰富的语义信息,提升模型对多尺度和小目标的检测能力;其次,引入注意力机制 ECA 模块,使特征提取网络聚焦特征图像的局部高效信息,减少无效目标的干扰,提升模型检测精度;最后,采用一种枝叶插图数据增强方法改进苹果数据集,解决图像数据不足问题。基于构建的数据集,使用遗传算法优化 K-means++ 聚类生成自适应锚框,提高模型定位准确性。试验结果表明,改进模型对可抓取和不可直接抓取苹果的精度均值分别为 96.16% 和 86.95%,平均精度均值为 92.79%,较传统 Faster R-CNN 提升 15.68 个百分点;对可抓取和不可直接抓取的苹果定位精度分别为 97.14% 和 88.93%,较传统 Faster R-CNN 分别提高 12.53 个百分点和 40.49 个百分点;内存占用量减少 38.20%,每帧平均计算时间缩短 40.7%,改进后的模型参数量小且实时性好,能够更好地应用于果实采摘机器人视觉系统。

关键词: 苹果采摘机器人; 目标定位与检测; Faster R-CNN; 注意力机制; 特征金字塔

中图分类号: TP391.4; S24

文献标识码: A

文章编号: 1000-1298(2024)01-0047-08

OSID:



Vision Detection Method for Picking Robots Based on Improved Faster R-CNN

LI Cuiming YANG Ke SHEN Tao SHANG Zhengyu

(School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: To address the issue of poor detection and positioning capabilities of fruit picking robots in scenes with densely distributed targets and fruits occluding each other, a method to improve the fruit detection and positioning of Faster R-CNN was proposed by introducing an efficient channel attention mechanism(ECA) and a multiscale feature fusion pyramid(FPN). Firstly, the commonly used VGG16 network was replaced with a ResNet50 residual network with strong expression capability and eliminate network degradation problem, thus extracting more abstract and rich semantic information to enhance the model's detection ability for multiscale and small targets. Secondly, the ECA module was introduced to enable the feature extraction network to focus on local and efficient information in the feature map, reduce the interference of invalid targets, and improve the model's detection accuracy. Finally, a branch and leaf grafting data augmentation method was used to improve the apple dataset and solve the problem of insufficient image data. Based on the constructed dataset, genetic algorithms were used to optimize K-means++ clustering and generate adaptive anchor boxes. Experimental results showed that the improved model had average precision of 96.16% for graspable apples and 86.95% for non-graspable apples, and the mean average precision was 92.79%, which was 15.68 percentages higher than that of the traditional Faster R-CNN. The positioning accuracy for graspable and non-directly graspable apples were 97.14% and 88.93%, respectively, which were 12.53 percentages and 40.49 percentages higher than that of traditional Faster R-CNN. The weight was reduced by 38.20%. The computation time was reduced by 40.7%. The improved model was more suitable for application in fruit-picking robot visual systems.

Key words: apple picking robot; target localization and detection; Faster R-CNN; attention mechanism; feature pyramid

收稿日期: 2023-06-26 修回日期: 2023-07-13

基金项目: 国家自然科学基金项目(52265065, 51765031)

作者简介: 李翠明(1976—),女,副教授,博士,主要从事移动机器人场景理解和导航研究, E-mail: li_goddess@163.com

0 引言

采摘机器人要完全替代人力完成复杂自然环境中的采摘作业,首先要解决的问题是如何在复杂的自然环境中快速准确地对采摘对象进行定位和检测,特别是自然环境果园中果实目标分布密集、相互遮挡较为常见,要求在复杂环境中具有更强目标识别能力的检测及定位算法。因此,研究针对自然环境中目标分布密集、相互遮挡问题的目标定位与检测方法,提升模型对密集目标的检测与定位性能,对降低新鲜水果的采摘成本,提高采摘效率具有重要意义^[1]。

深度学习技术在采摘方面已经有了广泛的应用^[2-5],基于深度学习的目标检测分为两类^[6-7]:二阶段检测算法和单阶段检测算法^[8]。二阶段检测算法的代表是基于区域的卷积神经网络(Region-based convolution neural networks,R-CNN),即 R-CNN 系列算法。该类算法检测精度较高,但检测速度较慢。文献[9]使用低成本 Kinect V2 传感器开发室外机器视觉系统,利用深度特征过滤背景物体来提高苹果检测的准确性并采用两种基于 Faster R-CNN 的架构 ZFNet 和 VGG16 来对比检测效果。文献[10]在非结构化环境下的草莓检测中开发了特征金字塔增强的 Mask R-CNN。相比二阶段检测算法,单阶段检测算法检测速度更快,但检测精度较低。文献[11]使用改进的 13 层 YOLO v3 算法,证明了在不同光照方向、不同苹果生长阶段和不同采摘时间下使用深度学习算法是可行的。文献[12]使用改进的 YOLO v3-Tiny 网络达到了很高的检测速度。文献[13]使用改进的 EfficientNet-

YOLO v4 数据集和叶子遮挡数据增强方法取得了较好效果。文献[14]在传统 Mask R-CNN 中增加了一个抑制分支来抑制非苹果特征,检测速度表现良好。

上述方法将单一的果实作为识别对象,或者侧重于背景的分类,然而当采摘对象被树枝或果实遮挡,采摘机器人^[15]就无法直接采摘,若直接采摘果实,可能会造成机器人的末端执行器和机械采摘臂被树枝损坏,导致采摘操作失败。考虑复杂果园实际环境,本文将果实分为可抓取类(未被叶子遮挡或仅被叶子遮挡)和不可直接抓取类(被树枝或果实遮挡)。针对复杂果园环境,以果实为检测目标,提出一种引入注意力机制的改进 Faster R-CNN 果实采摘机器人的视觉定位和检测方法。采用表达能力较强的融合特征金字塔 (Feature pyramid networks,FPN) 的残差网络 ResNet50 替换原模型主干网络,并嵌入注意力模块强化模型学习果实特征;针对遮挡问题,利用插图数据增强方法构建全新数据集,使用基于遗传算法的 K-means++ 聚类生成自适应锚框,优化边界框尺寸,以提高前景中果实的检测精度。

1 改进的苹果采摘目标检测模型

Faster R-CNN^[16] 有两个网络:生成区域提案的区域提案网络 (Region proposal network,RPN) 和提案检测对象网络。RPN 对区域框排序,并建议最有可能包含对象的框。改进的 Faster R-CNN 检测模型架构如图 1 所示,该模型包含 4 个主要组成部分:特征提取网络、区域建议生成、感兴趣区域 (Region of interest,ROI) 池化、分类和边界框回归。模型利用 ResNet50 网络进行特征提取,使用高效通

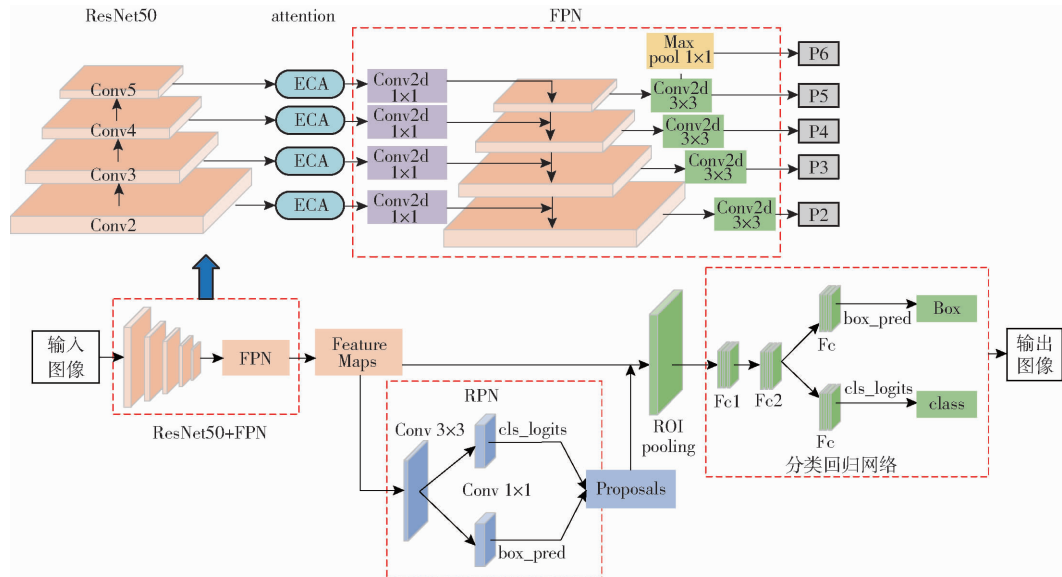


图 1 改进的 Faster R-CNN 检测模型结构图

Fig. 1 Improved Faster R-CNN detection model structure

道注意力 (Efficient channel attention, ECA) 模块增强卷积神经网络 (CNN) 的特征提取能力, 采用 FPN 将网络的不同尺度特征图融合在一起, 提高检测小目标的准确性, 通过优化区域建议网络中的锚框, 提高模型对果实的定位准确度。

1.1 改进特征提取网络结构

在自然环境中不同拍摄视角采集的果实图像, 存在枝叶对果实的遮挡、大量无关和复杂背景等制约因素, 传统检测模型的特征提取网络 VGG16 对上述制约因素下的果实特征刻画丰富度不足, 难以获得满意的检测效果。为了提高网络获取果实特征的丰富度, 提取更为抽象的特征, 需要加深网络的层数, 由此将导致网络退化。本文采用残差网络 ResNet50^[17] 作为特征提取网络, 以解决网络深度增

加导致的网络退化问题, 在提高模型精度的同时降低模型参数量。

改进的特征提取网络结构如图 2 所示, ResNet50 模型结构如图 2c 所示。第 1 阶段, 使用 7×7 卷积层和 3×3 最大池化进行采样以减小输入图像的大小; 阶段 2 中使用 Conv2、Conv3、Conv4 和 Conv5 残差结构提取高维特征; 最后, 将提取的高维特征送至第 3 阶段的全连接层进行分类。残差块有附加比例的卷积残差块和特征尺度未改变的恒等残差块两种类型的结构, 如图 2a、2b 所示。卷积残差块图通过跳跃连接 (skip connection) 方式将输入的特征矩阵维度缩小一半, 以更好地应对模型退化; 恒等残差块图的输入和输出特征矩阵维度相同, 允许串联操作以加深网络深度。

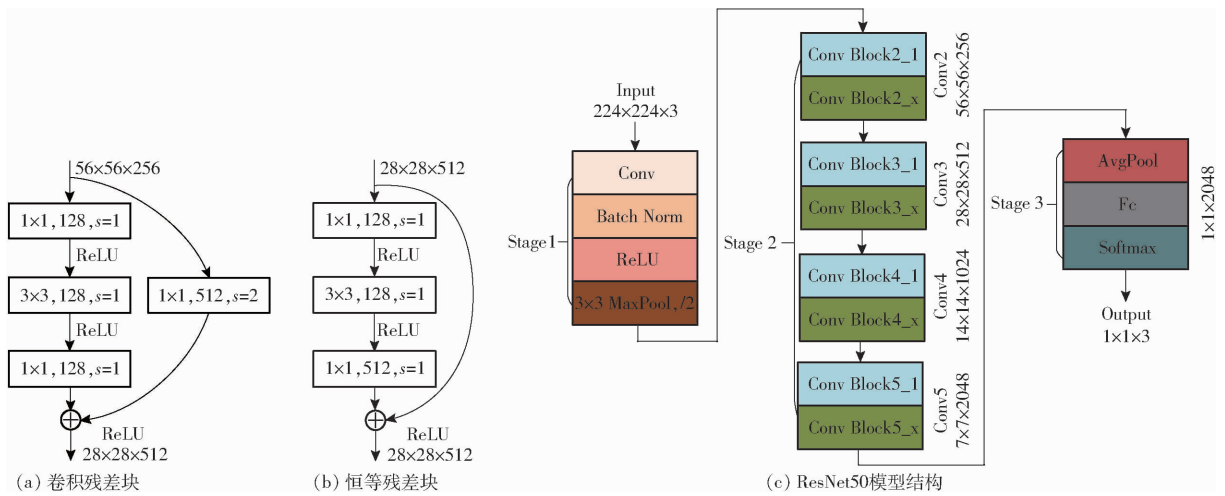


图 2 改进的特征提取网络结构图

Fig. 2 Improved feature extraction network structure

1.2 引入注意力机制的 ECA 模块

注意力机制 ECA 模块是 SE (Squeeze excitation) 模块^[18] 的改进版本, 其结构如图 3 所示。ECA 模块^[19] 提出了一种不降维的局部跨渠道交互策略, 该策略考虑通道全局平均池化 (Global average pooling, GAP) 后的每个通道及其 k 个近邻来捕获本地跨通道交互信息。ECA 模块首先对输入特征图进行空间特征压缩, 得到 $1 \times 1 \times C$ 特征图; 其次, 进行通道特征学习, 通过动态卷积核的一维卷积捕获跨通道交互信息, 由卷积核自适应函数公式自适应选择确定局部跨通道交互作用的覆盖范围; 最后, 将激活函数计算生成的输出特征通道权重向量与原始输入特征图进行逐通道相乘, 输出具有通道注意力的特征图, 从而提高有效特征的提取。ECA 模块避免了通道维度减缩, 允许模型学习更有效的通道注意力, 模块具有少量参数, 仅由其卷积核大小 k 决定, 其公式为

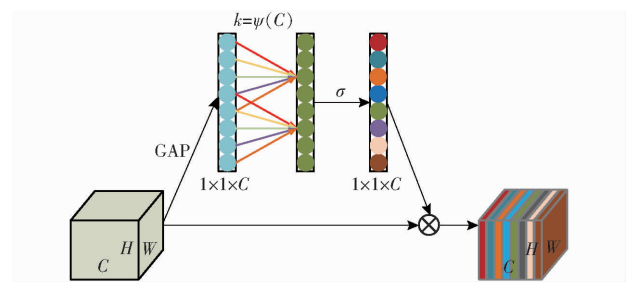


图 3 ECA 模块

Fig. 3 ECA module

$$k = \psi(C) = \left\lceil \frac{\lg(C)}{\gamma} + \frac{b}{\gamma} \right\rceil_{\text{odd}} \quad (1)$$

式中 C ——通道数

用 γ 、 b 改变 C 和 k 之间的比例, 一般取 $\gamma = 2$, $b = 1$ 。

本文将低复杂度和高效率的 ECA 模块嵌入骨干网络 ResNet50, 提高模型的表达能力, 最终达到提高模型识别及定位性能的目的。

1.3 采用特征金字塔的多尺度特征融合

图像中可能有多个不同大小的果实, 需要不同

的尺度特征加以区分。然而,随着 CNN 提取的特征更为抽象,其特征图尺度逐渐缩小,导致详细信息丢失。本文选取各尺度上都具有强语义特征的 FPN^[20] 进行融合,以充分利用 ResNet50 网络提取的各阶段特征。构建的多尺度特征融合 FPN 结构如图 4 所示,特征提取网络 ResNet50 包含 4 个残差体: C2、C3、C4 和 C5。选择每个残差体的最后一个残差结构输出特征图,共获得 4 类特征图。将这 4 类特征图输入信道 ECA 模块进行特征校正和滤波,对网络提取的有效特征信息赋予更高的权重,

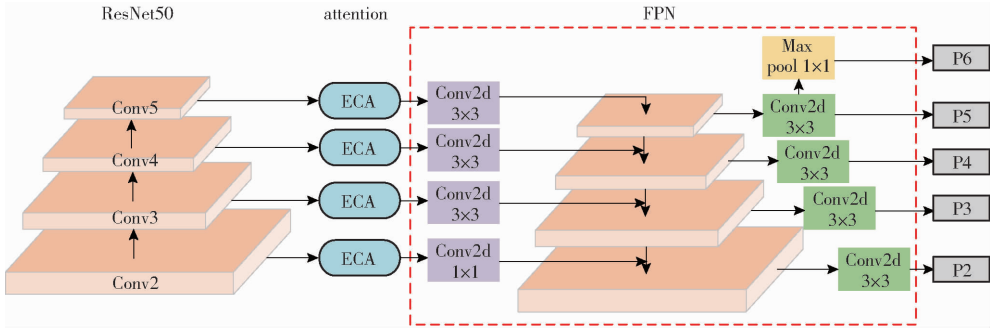


图 4 多尺度特征融合 FPN 结构图

Fig. 4 Feature pyramid networks with multiscale feature fusion

1.4 基于改进的 K-means + + 算法的自适应锚框

传统 Faster R - CNN 模型的初始检测锚框为 3 种尺寸 (128 × 128, 256 × 256, 512 × 512), 每个特征图设置 3 种比例 (1:1、1:2、2:1)。由于改进的检测模型中特征提取网络有 5 个预测特征层, 故本文增加 2 种尺寸 (32 × 32、64 × 264)。图像中可能有多个不同大小的果实检测目标, 模型训练中设置的 3 种比例 (1:1、1:2、2:1) 与原始定位点的相应尺寸并不匹配, 将会导致候选框尺寸与图中目标无法适应和精确匹配, 从而影响模型检测的准确性, 因此, 采用一种生成锚框 (anchors) 的自适应方法对所用数据集进行分析, 设计自适应采摘对象的尺寸和比例的锚框。生成 anchors 的自适应方法有 3 个步骤: 首先, 标注所有采摘目标的相应坐标并设置聚类中心数 n ; 其次, 运用 K-means + + 聚类得到 n 个 anchors, 对数据集中每个样本到各个聚类中心点的距离, 采用基于 1 - IOU (bboxes, anchors) 距离的方法进行 K-means 聚类; 最后, 利用遗传算法对 K-means + + 聚类的结果进行优化。

2 模型训练与试验

2.1 数据集构建与插图数据增强

本文以果园中需要采摘的苹果作为目标检测与定位对象, 运用 Python 语言开发图像爬虫在百度和谷歌上基于关键字“苹果”、“苹果树”和“苹果园”等批量下载苹果图像。为了确保图像质量, 将抓取

完成信道权重信息的标定, 过滤无用信息后输入 FPN 结构中。由于拍摄距离变化的影响, 捕获的果实图像有大小不同的目标, 不同目标具有不同的特征, 故需要将得到的不同尺度的特征图添加水平连接, 进行不同尺度的特征融合, 进而构建 FPN 结构。图 4 中 FPN 结构输出的特征图 P2 ~ P5 与 ResNet50 中的 C2 ~ C5 相对应, P6 是在 P5 的基础上通过最大池化得到的更高层次的抽象特征。将 P2 ~ P6 不同比例的特征图作为 RPN 的输入, 查找 ROI。

图像设置为至少大于 500 像素, 通过手动筛选重复、模糊和不一致的图像后共得到 928 幅高质量图像, 其中仅有单个苹果图像 130 幅, 含有苹果被树叶遮挡、被树枝遮挡、混合遮挡以及苹果相互重叠的图像 798 幅, 不同条件下苹果图像如图 5 所示。

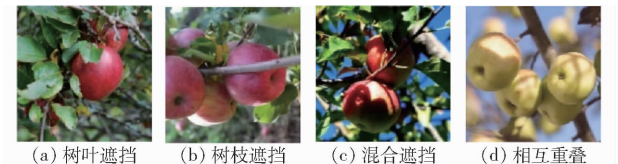


图 5 不同条件下的苹果图像

Fig. 5 Different types of apple images under different conditions

为满足检测模型需要的大量训练图像数据, 本文在传统数据增强如镜像、裁剪、亮度、模糊、丢失、旋转、缩放和平移操作等方法的基础上, 提出了一种枝叶插图数据增强方法。该方法首先利用“LabelImg”图像数据标注软件, 将图像中未遮挡或仅被叶子遮挡苹果标记为“可抓取”类, 被树枝或者苹果遮挡的苹果及其他条件下的苹果标记为“不可直接抓取”类, 然后使用一些叶子或树枝图像随机插入原始图像中对数据集进行数据增强处理。由此提取属于不同标记类别的苹果的特征, 避免训练得到的模型过度拟合。插图如图 6、7 所示。插图增强前后效果对比如图 8、9 所示。

通过枝叶插图数据增强方法将 928 幅图像扩展到 5 050 幅图像。随机划分 4 545 幅图像作为训练

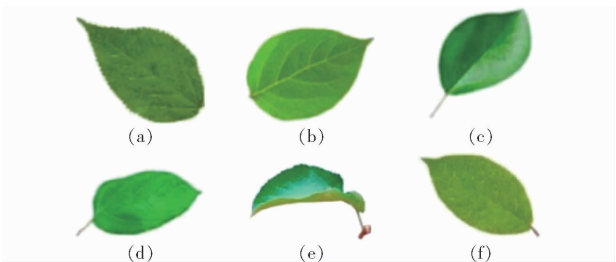


图 6 6 种苹果叶插图
Fig. 6 Six apple leaf illustrations

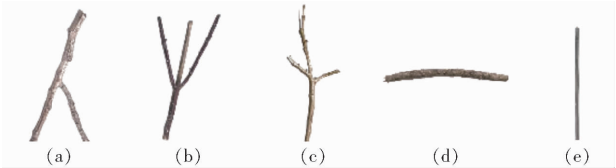


图 7 5 种树枝插图
Fig. 7 Five illustrations of tree branches



图 8 不同叶子插图增强前后效果
Fig. 8 Comparison of different leaf illustration data enhancement effects before and after enhancement

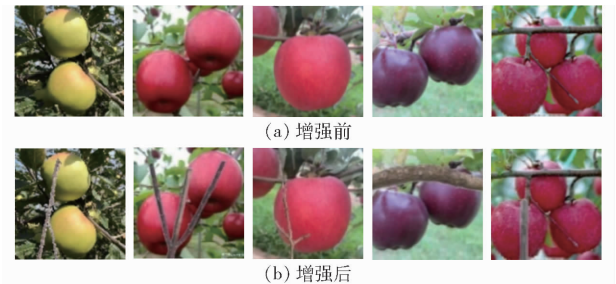


图 9 不同树枝插图增强前后效果
Fig. 9 Comparison of different branch illustration data enhancement effects before and after enhancement

集来训练检测模型,505 幅图像作为测试集来验证检测性能。

2.2 试验环境配置及参数设置

试验所用硬件平台为 Intel(R) Core(TM) i9 - 9900K CPU @ 3.60 GHz 3.60 GHz 处理器,内存为 128 GB,显卡为英特尔®超核芯显卡 630,显存为 64 GB。在 Windows 10 64 位系统环境下运行,采用 Pytorch 1.12 版本的深度学习开源框架配以 CUDA 11.7 进行环境配置。模型训练时,批量大小 (batchsize) 设置为 8,初始学习率为 0.01,动量参数设置为 0.9。

3 结果与分析

3.1 模型训练

使用平均精度 (Average precision, AP) 和平均精度均值 (Mean average precision, mAP) 作为苹果检测性能有效性的评价指标。定位性能采用平均交并比 (Mean intersection over union, MIoU) 定量评估,MIoU 综合考虑了模型提取目标和非目标交并比 (Intersection over union, IoU),值越高表示模型提取效果越好。

使用改进的 K-means ++ 算法对数据集所有标注框的宽度和高度进行聚类,设置 5 个聚类中心,聚类结果如图 10 所示。取这 5 个聚类中心宽度和高度平均值作为 anchor 的尺寸,将聚类中心的宽高比分为 3 组,最终改进后的 anchor 共有 5 种尺度 (54^2 、 65^2 、 116^2 、 282^2 、 415^2),3 种纵横比 (0.73、1.0、1.2)。

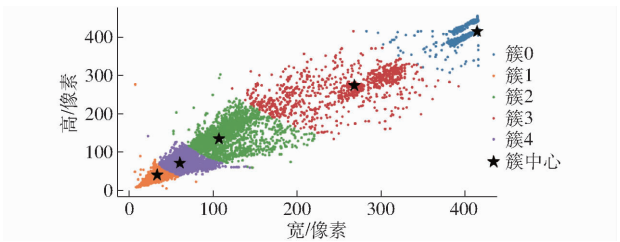


图 10 改进的 K-means ++ 聚类结果
Fig. 10 Improved K-means ++ clustering results

为验证改进 Faster R - CNN 模型对比原始模型的优势,分别对改进前后模型用同样的训练参数和数据集进行训练,最终模型训练 65 次。损失值变化曲线如图 11a 所示,改进后的模型在前 10 次中迅速拟合,曲线垂直下降,在 15 次之后渐渐平稳,只有些许的振荡,由此可知改进后模型收敛性更强,后期波动小。mAP 随迭代次数的变化如图 11b 所示,当迭代次数达到 35 时,mAP 已经稳定,不再发生变化,其中 mAP 最大值为 90.99%。

3.2 数据增强方法的影响

传统的数据增强方法与本文所提出的增强方法对改进的 Faster R - CNN 模型的影响如表 1 所示。

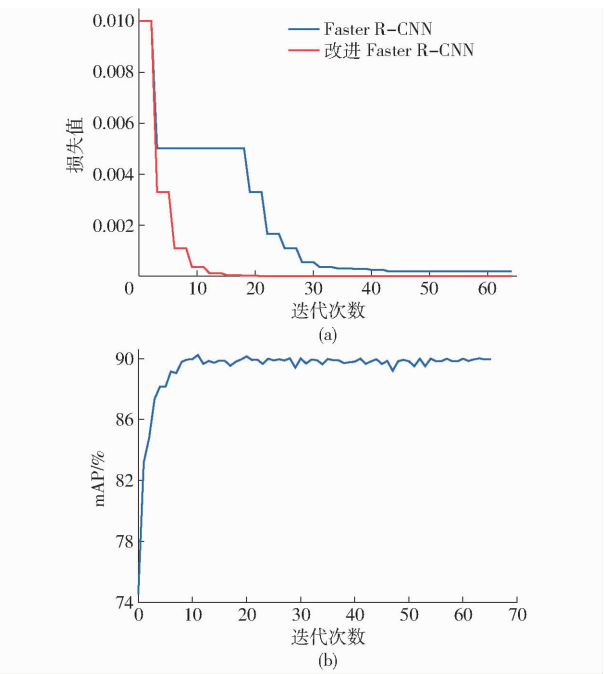


图 11 损失值变化曲线与 mAP 曲线

Fig. 11 Loss value curves and mAP value curves

表 1 插图增强前后性能

Tab.1 Enhancing illustrations improved before and after performance				%
数据增强方法	AP		mAP	
	可抓取类	不可直接抓取类		
传统增强	91.64	79.43	85.51	
插图增强	96.16	86.95	92.79	

由表 1 可知,在有插图增强时改进模型对可抓取苹果和不可直接抓取苹果 AP 分别提高约 5 个百分点和 7 个百分点,说明对不可直接抓取苹果的检测精度提高得更为明显,插图数据增强的方法对丰富训练集的多样性有一定贡献。这是因为与常见的数据增强方法相比,插图数据增强为图像生成了新的背景和纹理,对增强模型检测的鲁棒性有很大帮助,特别是在茂密的树叶干扰下。将常见的数据增强与提出的图像数据增强方法结合生成苹果图像,一方面弥补了训练图像的不足,大大减少标注工作量,另一方面增强方法丰富了训练图像中的叶遮挡场景,为模型学习提供了更丰富的特征,从而有助于提高模型学习能力和检测结果。

3.3 消融试验

为了评估对 Faster R - CNN 模型的各项改进是否有效及其相互作用,使用消融试验分析每个改进组件的因果关系。将试验分成 5 组即 M0 ~ M4 进行训练:M0 为传统 Faster R - CNN 模型;M1 将骨干网络替换为 ResNet50,特征融合使用 FPN 方法;M2、M3 是在 M1 的基础上分别将注意力机制 CBAM 和 ECA 嵌入骨干网络 ResNet50;M4 则

是基于 M3 试验,增加了 RPN 中初始锚框的自适应。在构建的苹果数据集上不同模型性能测试结果如表 2 所示,其中“√”表示包含该结构,“-”表示不含该结构。

表 2 消融试验结果

Tab.2 Results of ablation experiments						
模型	组别	ResNet50 + 注意力机制			初始锚框	mAP/%
		FPN	CBAM	ECA		
Faster R - CNN	M0	-	-	-	-	77.11
	M1	√	-	-	-	90.45
改进 Faster R - CNN	M2	√	√	-	-	90.32
	M3	√	-	√	-	92.58
	M4	√	-	√	√	92.79

观察表 2 发现,M1 模型与基础 Faster R - CNN 模型比较,平均精度明显提升 13.34 个百分点,由此可知特征提取网络的替换以及特征融合使得模型提取特征的能力提升,且有效提高了模型的精度。不同注意力模块下不同模块的检测能力:当 M2 模型采用 CBAM 双通道注意力机制时,模型性能较差,精度相比于 M1 下降 0.13 个百分点;M3 相比于 M1 模型有提升,说明新的特征提取网络以及特征融合与 CBAM 模块存在功能冲突,导致一定的副作用。通过将 CBAM 注意力模块替换为 ECA 注意力模块,M3 模型精度为 92.58%,明显高于 M2。该结果表明,高效通道注意力 ECA 模块比 CBAM 能更大程度地提高模型的精度,而 M4 基于 M3,增加了初始锚框的自适应,提高对建议框的定位准确性,使精度提升 0.21 个百分点。因此研究中采用 M4 模型作为改进的 Faster R - CNN 模型的最终版本。

3.4 模型检测性能

考虑改进模型是一种用于抓取的实时检测方法,若对苹果目标进行更多的类别划分将导致识别模型参数量增加,牺牲一定的检测时间,因此将苹果分为可抓取类(未被遮挡或仅被叶子遮挡的苹果)和不可直接抓取类(被树枝遮挡或者苹果重叠及其他条件下)2 类,能够满足实际采摘工况。使用构建的苹果数据集,将本文改进的 Faster R - CNN 模型与传统模型进行性能比较,结果如表 3 所示。

表 3 模型检测性能对比

Tab. 3 Comparison of model detection performance					
模型	AP/%		mAP/ %	检测	内存占
	可抓	不可直接		速率/	用量/
	取类	抓取类		(ms·f ⁻¹)	MB
Faster R - CNN	83.46	70.77	77.11	226	521
改进 Faster R - CNN	96.16	86.95	92.79	134	322

从表 3 可以看出,改进模型与传统模型相比对可抓取和不可直接抓取苹果的 AP 分别提升 12. 7、16. 18 个百分点,整体 mAP 也大幅提升 15. 68 个百分点;而每帧平均计算时间缩短 40. 7%,满足采摘机器人对实时苹果识别的要求;内存占用量减少 38. 20%,有利于以后在硬件设备中部署模型。

对比改进前后两种模型在测试集上对苹果的检测结果,部分测试结果如图 12 所示,其中绿色和蓝色的矩形分别指可抓取苹果和不可直接抓取苹果,而手动添加的对应颜色圆圈表示每个类中遗漏的水果,红色圆圈表示在错误类别中检测到的水果。对于同一图像,改进模型检测到的苹果果实多于传统模型,正确率也更高。尽管使用同一数据集训练两个模型,但同一水果也会被错误检测类别。

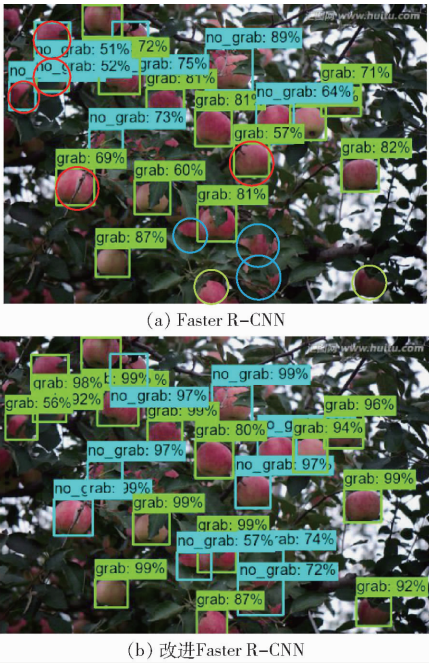


图 12 苹果检测结果

Fig. 12 Apple detection results

图 12a 中红色圆圈所示部分,在图 12b 中则都被正确识别出。图 12a 中遗漏了 5 个苹果,错误识别了 5 个苹果,而图 12b 的识别结果准确,没有错误识别或遗漏。

检测结果表明,当苹果果实较大且无遮挡时,包含更多的特征信息,也更容易被模型精准检测,改进前后模型均能取得较好的检测结果,但当果实较小、果实重叠或被枝叶遮挡时,由于果实暴露面积小,包含的有效特征信息少,检测难度比较大,此时传统模型出现了漏检的情况,而改进模型则不受影响,正确检测出了所有果实目标。

3.5 定位精度

将提出的模型与传统 Faster R - CNN 模型的定位精度对比,同时为体现每一步改进对模型的定位性能影响,进一步对比了有注意力机制和改进初始锚框的定位精度。表 4 显示了对可抓取类和不可直接抓取类苹果的定位精度。

表 4 定位精度 MIoU 对比

Tab. 4 Comparison of positioning accuracy %

模型	可抓取类	不可直接 抓取类
Faster R - CNN	84. 61	48. 44
ResNet50 + FPN	94. 71	87. 89
改进 Faster R - CNN 注意力机制	96. 69	88. 92
初始锚框	97. 14	88. 93

从表 4 结果可以看出,对可抓取类和不可直接抓取类苹果的定位精度分别提高了 12. 53 个百分点和 40. 49 个百分点,改进的模型比传统 Faster R - CNN 的定位精度更优,尤其是对遮挡类的苹果定位精度的提升更大。对于自然光下的苹果果实,该算法能够有效分离重叠果实,在枝叶遮挡、光照条件差等情况下能够更准确地检测目标果实,实现快速准确地定位,进行目标标注。如果图像中的苹果数量不止一个,则确定图像中面积最大的苹果果实,即靠近镜头的苹果果实作为目标,这也满足了苹果采摘机器人采摘作业时的实际需求。

4 结束语

为了模拟果园苹果检测可能出现的复杂场景,提出一种枝叶插图数据增强方法来改进苹果数据集。鉴于 Faster R - CNN 模型的检测性能和计算复杂性,使用表达能力较强的多尺度特征融合 FPN 的残差网络 ResNet50 来取代原骨干网络 VGG16,并引入高效通道注意力机制 ECA,进一步优化不同残差的输出特征图。改进的 Faster R - CNN 模型的 mAP 为 92. 79%,每帧平均计算时间为 0. 134 s,定位可抓取和不可直接抓取的苹果目标 MIoU 分别为 97. 14% 和 88. 93%。将提出的 Faster R - CNN 模型与 Faster R - CNN 模型相比,mAP 也提高 15. 68 个百分点,内存占用量减少了 38. 20%,对不可直接抓取类的苹果定位精度提升了近 1 倍。本文将果实分为可抓取类(未被叶子遮挡或仅被叶子遮挡)和不可直接抓取类(被树枝或果实遮挡),在后续的工作中,对于不可直接抓取类则可以制定采摘策略(例如,果实采摘顺序、采摘机械臂轨迹规划和路径规划),避免树枝的潜在损坏,从而形成一套完整的苹果检测系统。

参考文献

- [1] 宋怀波,尚钰莹,何东健.果实目标深度学习识别技术研究进展[J].农业机械学报,2023,54(1):1-19.
SONG Huaibo,SHANG Yuying,HE Dongjian. Review on deep learning technology for fruit target recognition[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(1):1-19. (in Chinese)
- [2] GAO F,FU L,ZHANG X,et al. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN[J]. Computers and Electronics in Agriculture,2020,176:105634.
- [3] LIANG C,XIONG J,ZHENG Z,et al. A visual detection method for nighttime litchi fruits and fruiting stems[J]. Computers and Electronics in Agriculture,2020,169:105192.
- [4] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报,2020,48(6):1230-1239.
LUO Huilan,CHEN Hongkun. Survey of object detection based on deep learning[J]. Acta Electronica Sinica,2020,48(6):1230-1239. (in Chinese)
- [5] 王鹏新,田惠仁,张悦,等. 基于深度学习的作物长势监测和产量估测研究进展[J]. 农业机械学报,2022,53(2):1-14.
WANG Pengxin,TIAN Hui ren,ZHANG Yue,et al. Crop growth monitoring and yield estimation based on deep learning: state of the art and beyond[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(2):1-14. (in Chinese)
- [6] XU W,ZHAO L,LI J,et al. Detection and classification of tea buds based on deep learning[J]. Computers and Electronics in Agriculture,2022,192:106547.
- [7] ANAND K,KERRY B W,WANG Zhenglin,et al. Deep learning method overview and review of use for fruit detection and yield estimation[J]. Computers and Electronics in Agriculture,2019,162:219-234.
- [8] 孙颖,侯志强,杨晨,等. 基于双模态融合网络的目标检测算法[J]. 光子学报,2023,52(1):203-215
SUN Ying,HOU Zhiqiang,YANG Chen,et al. Object detection algorithm based on dual-modal fusion network[J]. Acta Photonica Sinica,2023,52(1):203-215. (in Chinese)
- [9] FU L,MAJEED Y,ZHANG X,et al. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting[J]. Biosystems Engineering,2020,197:245.
- [10] YU Y,ZHANG K,YANG L,et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. Computers and Electronics in Agriculture,2019,163:104846.
- [11] 赵德安,吴任迪,刘晓洋,等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报,2019,35(3):164-173.
ZHAO Dean,WU Rendi,LIU Xiaoyang,et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the CSAE,2019,35(3):164-173. (in Chinese)
- [12] MAZZIA V,KHALIQ A,SALVETTI F,et al. Real-time apple detection system using embedded systems with hardware accelerators[J]. An Edge AI Application. IEEE Access,2020,8:9102.
- [13] WU L,MA J,ZHAO Y,et al. Apple detection in complex scene using the improved YOLOv4 model[J]. Agronomy,2021,11(3):476.
- [14] CHU P,LI Z,LAMMERS K,et al. Deep learning-based apple detection using a suppression Mask R-CNN[J]. Pattern Recognition Letters,2021,147:206.
- [15] 孙俊,钱磊,朱伟栋,等. 基于改进 RetinaNet 的果园复杂环境下苹果检测[J]. 农业工程学报,2022,38(15):314-322.
SUN Jun,QIAN Lei,ZHU Weidong,et al. Apple detection in complex orchard environment based on improved RetinaNet[J]. Transactions of the CSAE,2022,38(15):314-322. (in Chinese)
- [16] 朱红春,李旭,孟杨,等. 基于 Faster R-CNN 网络的茶叶嫩芽检测[J]. 农业机械学报,2022,53(5):217-224.
ZHU Hongchun,LI Xu,MENG Yang,et al. Tea bud detection based on Faster R-CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):217-224. (in Chinese)
- [17] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016.
- [18] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018.
- [19] WANG Q,WU B,ZHU P,et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020.
- [20] LIN T Y,DOLLÁR P,GIRSHICK R,et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017.