

基于改进 YOLOv5 和 BiFPN 的苹果检测方法^{*}

温彬彬

(河北工业职业技术大学, 河北 石家庄 054000)

摘要:【目的】解决果蔬采摘机器人在背景复杂且苹果尺寸不一的情况下检测准确率低的问题, 提供更加高效、便捷的农业管理手段。【方法】提出一种基于改进 YOLOv5 和 BiFPN 的苹果检测方法。首先, 使用 YOLOv5 作为基础目标检测模型, 以实现高效的目标检测。其次, 引入 BiFPN, 通过权重系数和双向特征传播机制来增强特征融合能力, 以更好地捕捉多尺度、多层次的特征, 提高模型对不同尺度目标的检测能力和鲁棒性。【结果】将改进的模型在自建数据集上进行实验, 改进后的 BiFPN-YOLOv5 模型同原始 YOLOv5 模型对比, 在准确率上提高了 2.2 个百分点, 召回率提高了 2.4 个百分点, 平均精度提高了 3.2 个百分点, 但是训练时间增加了 0.6 ms/幅。【结论】训练时间的增加是由 BiFPN 的结构引起的, 其他指标均有一定程度的提升, 证明了改进算法的有效性, 能够提高苹果小目标的检测精度。

关键词: 目标检测; YOLOv5; BiFPN; 特征融合; 特征金字塔

中图分类号: TP391.4

文献标志码: A

DOI: 10.3969/j.issn.1672-3872.2024.09.002

苹果是全球最受欢迎、最重要的经济作物之一, 具有低投入、高产出、高收益等特点, 是促进农民脱贫致富和推动乡村振兴的重要抓手^[1]。数字化、智能化的农业正在高速发展, 苹果目标检测对于智能果园的生产管理、病虫害预测、果实采摘以及果实质量评估等方面具有重要作用, 快速且准确的苹果检测技术为果农提供了更加高效、便捷的农业管理手段, 加速推动了农业领域的智能化发展。

1 目标检测发展现状

目标检测是计算机视觉领域的一项重要任务, 它的目标是在图像或视频中准确地定位和识别出感兴趣的目标对象。早期的目标检测方法主要基于手工设计的特征, 如 Haar 特征、HOG 特征等^[2], 主要通过寻找特定的局部特征模式, 然后使用分类器进行目标检测。但是由于手工特征的设计通常依赖于领域知识和经验, 且难以适应复杂的场景和目标变化, 因此在性能和泛化能力上存在一定的局限性。为了解决目标尺寸和位置的变化, 研究者提出了基于滑动窗口的方法。这类方法通过在图像上移动不同尺寸和位置的窗口, 并使用分类器对每个窗口进行分类^[3], 以判断是否包含目标。然而, 这种方法需要对大量的窗口进行分类, 计算量较大, 且容易产生大量的误检测。随着深度学习的兴起, 目标检测领域也

迎来了革命性的突破。基于深度学习的目标检测方法通常采用卷积神经网络(CNN)来提取图像特征^[4], 并通过分类器和回归器来预测目标的类别和位置。一系列经典的方法如 R-CNN、Fast R-CNN、Faster R-CNN、YOLO 和 SSD 等被相继提出, 并在各自的任务中取得了较好的性能。这些方法通过端到端的训练方式, 可以自动学习图像特征和目标检测模型, 大大提升了检测性能和效率。但 CNN 是两阶段的检测器, 模型具有大量的参数和复杂的计算结构, 计算和推理速度较慢, 而且大量研究表明, CNN 对小目标的检测性能较弱。为了进一步提升检测速度和准确率, 研究者们提出了一系列单阶段检测器, 如 YOLOv2、YOLOv5、RetinaNet 和 EfficientDet 等^[5]。这些方法通过设计新的网络结构、特征融合机制和损失函数等, 实现了在速度和准确率上的平衡。

2 YOLO 算法简介

YOLO 系列算法(You Only Look Once)由于实时性和高准确性而备受关注。YOLOv1 是 YOLO 系列的第一版^[6], 是 Joseph Redmon 等在 2016 年提出的一种端到端的物体识别与检测方法, 采用了多个网格划分与多尺度预测的策略, 相较于同期方法如 Faster R-CNN、SSD 等有着更快的推理速度, 但是 YOLOv1 在检测小目标和困难目标上效果不佳。为了解决这些问题, Joseph Redmon 等在 2017 年提出了 YOLOv2, 使用 Anchor Boxes 来预测目标的位置和尺

基金项目: 石家庄市科技计划项目(231130351)

作者简介: 温彬彬(1989—), 女, 河北邢台人, 硕士, 讲师, 研究方向为图像处理及模式识别。

寸,使用Darknet-19网络结构来提取图像特征,使用批量归一化技术来加速模型的训练等,在精度和检测数目上有明显的提升。YOLOv3进一步提高了检测精度,引入了特征金字塔(Feature Pyramid Network, FPN)技术和多尺度特征图预测的策略,更适合小目标的检测,然而YOLOv3在速度方面的表现并不理想。为了达到速度和检测性能之间的平衡,2020年4月Alexey等推出了YOLOv4,提出了一系列的优化措施来平衡速度与精度;同年6月,YOLOv5在YOLOv4的基础上进一步优化并公开代码,提出在特征提取阶段使用更强的CSPDarkNet网络模型,在目标边框损失计算上使用CIoU损失(Complete Intersection over Union Loss)来更精确地描述两个目标框之间的位置偏差^[7],大量实验表明,YOLOv5相较于前代算法YOLOv4,在速度和精度方面都有着不小的提升。

YOLO系列算法在农业领域的应用非常广泛,特别是在农作物的检测、分类和病害诊断方面。杨佳昊等^[8]利用LCNN替换YOLOv5s主干对模型主体进行缩减,对番茄黄化曲叶病检测模型开展轻量化试验;刘忠意等^[9]提出一种改进的YOLOv5方法,在主干网络部分使用RepVGG(re-param VGG)模块替换原始C3模块,加强网络对特征信息的提取能力。还有一些研究人员使用深度学习来进行苹果识别。张立杰等^[10]提出了一种基于改进SSD卷积神经网络的苹果定位与分级算法,所提方法具有良好的分割效果。曹志鹏等^[11]采用轻量化的主干网络Ghostnet替代YOLOv4中的CSPdarknet53主干网络,再采用深度可分离卷积替换YOLOv4中的颈部网络,该算法加快了检测速度,减少了权重文件的大小。周桂红等^[5]提出了一种基于改进YOLOv4和基于阈值的边界框匹配合并算法的全景图像苹果识别方法,该方法在精确率和召回率上有明显的提升。张境锋等^[12]提出一种Des-YOLOv4算法与苹果检测方法,该方法具有检测速度快、检测精度高的优点。

虽然YOLOv5在农业领域应用较为广泛,但是自然环境下成熟的苹果存在果实大小不一、果实之间遮挡严重等问题。为了解决目标检测中的这些问题,本文提出一种基于改进YOLOv5网络的果实检测方法,在中间层的特征增强模块中改变原特征金字塔简单双向融合特征图,采用BiFPN特征融合网络^[13],将

不同尺度的特征融合,可以在保持高分辨率特征的同时,提升检测模型对小目标的感知能力。

3 改进的YOLOv5模型

3.1 YOLOv5算法介绍

YOLOv5是YOLO系列的第五代算法,相较于初代算法,其在模型架构、锚框匹配、损失函数等方面进行了大量改进,并最终在速度和精度之间取得了更好的平衡。其结构包括以下几个组成部分:

1) Backbone网络。由多个卷积层和池化层组成,将原始图像转化为一系列特征图,这些特征图可以捕捉到不同尺度和语义层次的信息,有助于提高目标检测算法在不同尺度上的检测能力。YOLOv5使用CSPDarknet作为其主干网络(backbone network),它是一种轻量级的卷积神经网络。CSPDarknet通过使用CSP(Cross Stage Partial Network)结构来提高特征的表达能力和计算效率。

2) Neck网络。将特征图进行融合,将低层次的细节特征和高层次的语义特征结合起来,从而提高模型对目标的检测能力,为了实现更好的目标检测性能,YOLOv5还引入了一种名为PANet(Path Aggregation Network)的Neck网络结构。PANet可通过多尺度特征融合的方式,将不同层次的特征进行聚合。

3) Head网络。YOLOv5的头部网络(head network)由一系列卷积层和全连接层组成。它负责对骨干网络提取的特征进行进一步处理,以输出目标的检测框、类别标签和置信度得分。

4) Anchor Boxes。YOLOv5使用了Anchor Boxes来预测目标的位置和尺寸。Anchor Boxes是一组预定义的矩形框,每个框都与一个特定的尺度和纵横比相关联。YOLOv5通过预测Anchor Boxes的偏移量和尺度缩放因子来生成最终的检测框。

5) Loss函数。YOLOv5使用了一种组合的损失函数来训练模型,包括目标检测损失(object detection loss)和分类损失(classification loss)。目标检测损失主要用于衡量检测框的准确性,分类损失主要用于衡量类别的预测准确性。

综上所述,YOLOv5算法通过Backbone网络提取图像特征,通过Neck网络进行多尺度特征融合,然后通过Head网络生成检测结果,并使用Loss函数进行模型训练和优化。这种结构设计使得YOLOv5在目标检测任务中具有较高的准确性和较快的检测速度。

YOLOv5模型的准确率很大程度上是由中间层特征增强模块决定的,中间层特征增强模块的关键是特征金字塔,传统的特征金字塔网络PANet使用简单双向融合网络在不同层级的特征图之间进行上下文信息的传递,如图1(a)所示。简单双向融合网络可能会造成特征的分辨率降低和信息丢失,在进行特征融合时,会导致不同层级的特征在数量和质量上存在不平衡的问题,影响目标检测的准确性。

3.2 BiFPN网络结构

为了解决信息传递不够充分而导致模型准确率下降的问题,在YOLOv5模型中使用一种改进的特征提取模块BiFPN(Bi-directional Feature Pyramid Network),它是一种用于目标检测任务的特征金字塔网络。

BiFPN通过引入双向的信息传递机制,更好地利用了不同层级特征的上下文信息,如图1(b)所示。BiFPN模块包括两个阶段:自上而下和自下而上。在自上而下的阶段,高层特征图通过上采样进行细化,并与低层特征图进行融合。这样可以使得高层特征图获得更丰富的上下文信息。在自下而上的阶段,低层特征图通过下采样进行粗化,并与高层特征图进行融合。这样可以使得低层特征图获得更多的细节信息。BiFPN结构中的每个层都与上下两个相邻的层进行连接,这种双向连接的设计可以有效地将不同层级的特征信息进行融合和交互,提高了特征的表达能力。BiFPN结构通过上下两个分支分别进行不同尺度的特征融合,可以在保持高分辨率特征的同时,将更高层级的语义信息引入,提升检测模型对小目标的感知能力。BiFPN结构的特征融合过程中,采用了一种动态权重分配的方式,根据特征的质量和重要性,自适应地调整各个特征通道的权重,以保证高质量特征的有效利用。BiFPN结构可以根据需要进行多次重复堆叠,每次重复都通过进一步的特征融合和优化,提高了网络的深度可扩展性,可以适应不同复杂度和精度要求的目标检测任务。

BiFPN网络通过加权特征融合机制、卷积、下采样的方式实现融合,以 P_6out 为例, P_6out 是将第6层输入的 P_6td 、 P_6in 和第5层的 P_5out 加权融合得到的, P_6td 是自上而下第6层的中间特征, P_6td 和 P_6out 的表达式分别如式(1)、式(2)所示。

$$P_6td = \text{Conv} \left(\frac{\omega_1 \cdot P_6in + \omega_2 \cdot \text{Resize}(P_7in)}{\omega_1 + \omega_2 + \varepsilon} \right) \quad (1)$$

$$P_6out = \text{Conv} \left(\frac{\omega'_1 \cdot P_6in + \omega'_2 \cdot P_6td + \omega'_3 \cdot \text{Resize}(P_5out)}{\omega'_1 + \omega'_2 + \omega'_3 + \varepsilon} \right) \quad (2)$$

式中, ω_1 、 ω_2 、 ω'_1 、 ω'_2 和 ω'_3 是权重参数; Resize 是下采样操作; ε 是一个参数,用来避免由于权重参数过小而导致的数值不稳定。

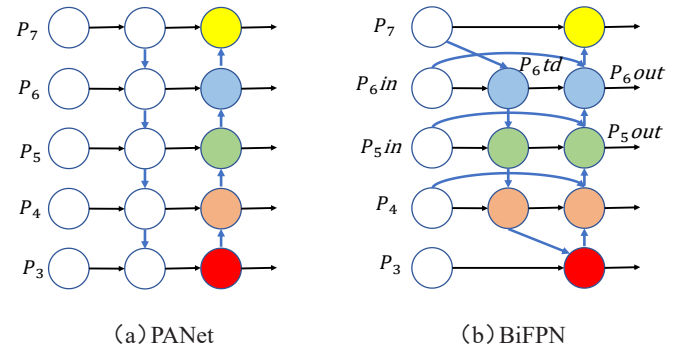


图1 PANet和BiFPN结构

BiFPN模块的引入使得YOLOv5模型在苹果识别任务中具有更好的性能和更高的准确性。它能够提取多尺度的特征,并且能够更好地处理不同尺度的目标。BiFPN结构基于双向连接、多尺度特征融合、动态权重分配和深度可扩展性等特点,有效地提升了目标检测模型的性能和表达能力。

4 实验与分析

4.1 实验环境

本实验配置环境为:台式计算机,CPU采用Intel-core i9-9900k处理器,显卡采用NVIDIA GeForce GTX 2080ti,显存为12 GB,内存为32 GB,采用pytorch框架,编程语言使用python 3.6。

4.2 实验数据集处理

本实验采用自建数据集,数据集共包含图片1 800张,包括不同遮挡情况、不同果实数量等的数据图片,共计3 000个目标。同时,为了避免过拟合,保证网络模型训练效果,在模型训练之前对数据集进行图像增强,例如缩放、旋转、调整对比度和饱和度等。

4.3 实验结果分析

为了保持一致,使用YOLOv5原网络进行相同的测试。选择准确率、召回率、平均精度、 F_1 值和检测速度作为模型性能评价指标,实验结果如表1所示。

表1 两种算法的检测效率对比

模型	准确率	召回率	平均精度	F_1 值	检测速度/(ms/幅)
YOLOv5	85.5%	86.4%	91.6%	87.1%	33.0
BiFPN-YOLOv5	87.7%	88.8%	94.8%	88.5%	33.6

通过实验数据可以看出,改进后 (下转第26页)

- 于绿色发展视角[J].科技管理研究, 2020, 40(9): 60-67.
- [2] 陈耀, 赵芝俊, 高芸. 中国区域农业科技创新能力排名与评价[J]. 技术经济, 2018, 37(12): 53-60.
- [3] 王丹, 杜旭, 郭翔宇. 中国省域农业科技创新能力评价与分析[J]. 科技管理研究, 2021, 41(1): 1-8.
- [4] 曹琼, 李成标. 基于熵权TOPSIS法的农业科技创新能力评价: 以湖北省为例[J]. 南方农业学报, 2013, 44(10): 1751-1756.
- [5] 孙长东, 陈酉宜. 新时期我国农业科技创新能力评价与提升对策研究[J]. 中国农业资源与区划, 2020, 41(6): 91-99.
- [6] 张小彦, 曹方, 叶得明, 等. 基于灰色综合评价法的甘肃农业科技创新能力分析[J]. 广东农业科学, 2014, 41(22): 191-195.
- [7] 吴敬学, 毛世平, 王志丹. 提升我国农业科技原始创新能力的思路与对策[J]. 农业经济问题, 2012, 33(8): 56-59+111.
- [8] 王丹, 赵新力, 郭翔宇, 等. 国家农业科技创新理论框架与创新能力评价: 基于二十国集团的实证分析[J]. 中国软科学, 2018(3): 18-35.
- [9] 张鸿, 郑林用. 省级农科院进一步提升科技创新与转化能力的战略思考[J]. 科技管理研究, 2009, 29(2): 180-182+188.
- [10] 杨传喜, 王亚萌. 中国区域农业科技资源配置能力及比较优势研究: 基于省级农业科学院视角[J]. 中国农业资源与区划, 2018, 39(12): 80-87.
- [11] 池敏青, 许正春, 刘健宏. 省级农业科研机构科技投入产出相关性实证分析[J]. 科技管理研究, 2017, 37(2): 111-116.
- [12] 杨传喜, 徐硕强, 孔令孜, 等. 农业科学院科技资源配置效率研究: 基于30个省级农业科学院的面板数据分析[J]. 南方农业学报, 2015, 46(1): 170-174.
- [13] 胡慧英, 申红芳, 廖西元, 等. 农业科研机构科技创新能力的影响因素分析[J]. 科研管理, 2010, 31(3): 78-88.

(上接第6页)

的BiFPN-YOLOv5模型同原始YOLOv5模型对比, 在准确率上提高了2.2个百分点, 召回率提高了2.4个百分点, 平均精度提高了3.2个百分点, 但是训练时间增加了0.6 ms/幅, 这是由BiFPN的结构引起的。实验结果表明: 改进后的模型提升了准确率、召回率、平均精度和 F_1 值, 由此验证了改进后的BiFPN-YOLOv5模型具有一定的可行性。

5 结论

本文主要解决了果蔬采摘机器人在背景复杂且苹果尺寸不一的情况下检测准确率低的问题, 使用BiFPN特征网络替换传统YOLOv5模型中的PANet特征网络, 将浅层图像特征层引入到金字塔图像融合模块中, 改变原网络模型简单双向特征的融合方式, 提高了苹果小目标的检测精度。在自建数据集上进行实验验证, 改进后的模型提升了准确率、召回率、平均精度和 F_1 值, 证明了改进算法的有效性。

参考文献:

- [1] 王丽娟, 陈浩然, 李石军, 等. 机器视觉成熟度检测的苹果色选分选机设计[J]. 农业与技术, 2022, 42(12): 36-40.
- [2] 鄢开俊, 丁元. 宽型自注意力融合密集型残差网络的图像去雾[J]. 湖南大学学报(自然科学版), 2023, 50(8): 13-22.
- [3] 孙丰刚, 王云露, 兰鹏, 等. 基于改进YOLOv5s和迁移学习的苹果果实病害识别方法[J]. 农业工程学报, 2022, 38(11): 171-179.
- [4] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 07-13, 2015, Santiago, Chile. IEEE, 2015: 1440-1448.
- [5] 周桂红, 马帅, 梁芳芳. 基于改进YOLOv4模型的全景图像苹果识别[J]. 农业工程学报, 2022, 38(21): 159-168.
- [6] 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), June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 779-788.
- [7] CHEN X, LIAN Q, CHEN X, et al. Surface Crack Detection Method for Coal Rock Based on Improved YOLOv5[J]. Applied Sciences, 2022, 12(19): 9695.
- [8] 杨佳昊, 左昊轩, 黄祺成, 等. 基于YOLO v5s的作物叶片病害检测模型轻量化方法[J]. 农业机械学报, 2023, 54(S1): 222-229.
- [9] 刘忠意, 魏登峰, 李萌, 等. 基于改进YOLO v5的橙子果实识别方法[J]. 江苏农业科学, 2023, 51(19): 173-181.
- [10] 张立杰, 周舒骅, 李娜, 等. 基于改进SSD卷积神经网络的苹果定位与分级方法[J]. 农业机械学报, 2023, 54(6): 223-232.
- [11] 曹志鹏, 袁锐波, 杨肖, 等. 一种改进神经网络的苹果快速识别算法[J]. 重庆理工大学学报(自然科学), 2023, 37(5): 210-217.
- [12] 张境锋, 陈伟, 魏庆宇, 等. 基于Des-YOLO v4的复杂环境下苹果检测方法[J]. 农机化研究, 2023, 45(5): 20-25.
- [13] WANG A, CAO X, LU L, et al. Design of Efficient Human Head Statistics System in the Large-Angle Overlooking Scene[J]. Electronics, 2021, 10(15): 1851.