

基于改进 YOLO 的自然环境下桃子成熟度快速检测模型

曾俊, 陈仁凡, 邹腾跃

(福建农林大学机电工程学院, 福建 福州 350002)

摘要:【目的】解决自然环境下不同成熟度桃子快速准确检测的问题, 课题组提出一种基于改进YOLOv5s的目标检测算法YOLO-Faster。【方法】使用YOLOv5s网络模型作为基础网络, 将主干特征提取网络替换为FasterNet, 使模型轻量化, 并在主干和颈部之间增加串联的CBAM卷积注意力模块和常规卷积块, 增强对图像重要特征的捕捉与表达, 同时引入SIOU损失函数缓解预测框与真实框之间方向的不匹配。【结果】改进后模型的 mAP 为88.6%, 与YOLOv5s相比提升1个百分点, 模型权重缩减39.4%, 浮点运算量降低44.3%, 在GPU、CPU上的单张图像平均检测时间分别减少12.6%和24%。此外, 本研究将训练好的模型部署到嵌入式设备Jetson Nano上, 模型在Jetson Nano上的检测时间比YOLOv5s减少30.4%。【结论】改进后的轻量级模型能够快速准确地检测自然环境下不同成熟度的桃子, 可以为桃子采摘机器人的视觉识别系统提供技术支持。

关键词: 目标检测; 桃子成熟度; YOLOv5s; FasterNet; 注意力机制; 快速识别

中图分类号: TP391.4

文献标志码: A

DOI: 10.3969/j.issn.1672-3872.2023.24.006

0 引言

桃子是一种生活中常见的水果, 具有较高的营养价值和经济价值。据统计, 2019年我国桃种植面积和产量分别为89.0万 hm^2 和1 599.3万t, 均居世界第一位^[1]。但当前桃子采摘主要依赖人工来完成, 效率低, 成本高, 引入智能化的农业机械装备用于果实采摘十分必要, 果实采摘机器人成为农业发展的重要方向^[2]。因此, 作为果实采摘机器人的关键技术, 设计一种高准确率的自然环境下桃子果实成熟度快速检测系统对实现桃子自动采摘具有重大意义。

传统的果实目标识别方法主要基于颜色、几何形状等特征进行检测和分类^[3]。Liu等^[4]使用椭圆边界模型的机器视觉算法对柚子果实成熟度进行检测研究, 将图像从RGB空间转换至Y'CbCr空间, 达到了93.5%的准确率, 但该研究对于未成熟的柚子果实的检测效果较差。Lin等^[5]提出了一种Hough变换方法, 对轻度遮挡的目标能够识别, 但对于遮挡严重的目标仍然无法准确识别。

上述的传统机器视觉方法依赖人工提取特征, 鲁棒性和泛化能力差。近年来, 深度学习发展迅速, 在果实识别定位中发挥着越来越重要的作用^[6]。崔潇等^[7]提出一种以MTCNN为主干的桃子检测模型, 该模型对成熟桃子具有89.3%的检测准确率。Liu等^[8]在YOLOv3中引入一种密集架构并将矩形边界框替换为圆形边界框, 识别遮挡等条件下的番茄, 精度达

到96.4%。王立舒等^[9]基于YOLOv4-Tiny结合卷积注意力模块构建蓝莓成熟度目标检测模型。

总之, 深度学习在果实目标检测中具有较大优势。但基于深度学习的网络模型通常比较大, 检测速度不够快, 不能满足果实采摘机器人的实际作业需求。基于以上问题, 课题组提出一种基于YOLOv5s的自然环境下快速检测桃子成熟度的轻量级网络, 以期桃子采摘机器人的视觉识别系统提供技术支持。

1 材料与方法

1.1 数据及制作

本研究所用的图像采集于湖北省黄冈市一处桃子种植园, 桃子品种为“胭脂桃”。共采集自然条件下桃子果实原始图像1 860幅。使用LabelImg软件对原始图像进行标注, 将桃子果实标注为未成熟、半成熟以及成熟, 并按照比例8:1:1划分训练集、验证集和测试集。

训练神经网络需要的数据量较大, 较大的数据集训练得到的模型泛化能力和鲁棒性也相对较好, 因此对采集到的原始图像进行了数据增强。本研究采用图像翻转、修改对比度、图像缩放、添加噪声的方法对数据集进行了增强, 最终得到3 638幅图像用于训练模型。

1.2 YOLOv5s网络模型

YOLOv5s目标检测算法是YOLO系列的最新版本之一, 整体结构主要由Backbone、Neck、Head三

作者简介: 曾俊(1998—), 男, 湖北黄冈人, 硕士研究生, 研究方向为计算机视觉及目标检测。

通信作者: 邹腾跃(1984—), 男, 福建福州人, 博士, 副教授, 研究方向为人工智能、模式识别与智能系统。

个部分组成,其主干网络使用了CSPNet结构,通过CSPNet的交叉连接和信息整合使得网络能够更好地传递和融合不同阶段的特征,提高了模型的特征表示能力。颈部使用了路径聚合网络结构PANet,生成具有多尺度信息的特征图。头部使用了CIOU_Loss和非极大值抑制(NMS),完成目标检测结果的输出。

1.3 改进的桃子成熟度检测模型

1.3.1 FasterNet

本研究使用FasterNet^[10]重新搭建了YOLOv5s的主干特征提取网络。FasterNet提出一种新颖的PConv(Partial Conv),可以同时减少计算冗余和内存访问,提升模型的检测速度。它利用了特征图中的冗余,并系统性地仅在一部分输入通道上应用常规卷积(Conv),而不影响其余通道。对于连续或者常规的内存访问,PConv将第一个或者最后一个连续的通道视为整个特征图的代表进行计算,在不失一般性的情况下认为输入和输出特征图具有相同的数量。

FasterNet基本结构BasicStage的结构如图1所示,它由PConv和两个逐点卷积(PWConv)构成,并使用残差连接缓解梯度消失问题。由于使用了PConv和PWConv,BasicStage模块的参数量和浮点运算量大幅降低。

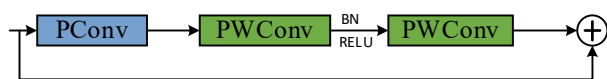


图1 BasicStage结构

1.3.2 CBAM 卷积注意力机制模块

在深度学习模型中增加注意力机制可以增强模型对于输入数据中有意义部分的关注,优化模型的表现。Woo等^[11]于2018年提出CBAM卷积注意力模块(Convolutional Block Attention Module),其结构如图2所示,该模块包含两个主要组成部分:通道注意力模块(CAM)和空间注意力模块(SAM)。在CAM中,输入特征图F经过池化并通过全连接层进行通道数压缩以及激活函数(ReLU)处理,生成两个激活后的特征图并经过加和以及sigmoid操作得到输入特征的每个通道的权重MC。将这些权重与输入特征的对元素相乘得到SAM输入特征图F'。在SAM中,F'经过池化操作得到2个二维向量,然后将其进行拼接和卷积操作并经过sigmoid得到输入特征的每一通道的权重MS,最后将权重MS与输入特征对应元素相乘得到最终的特征图。

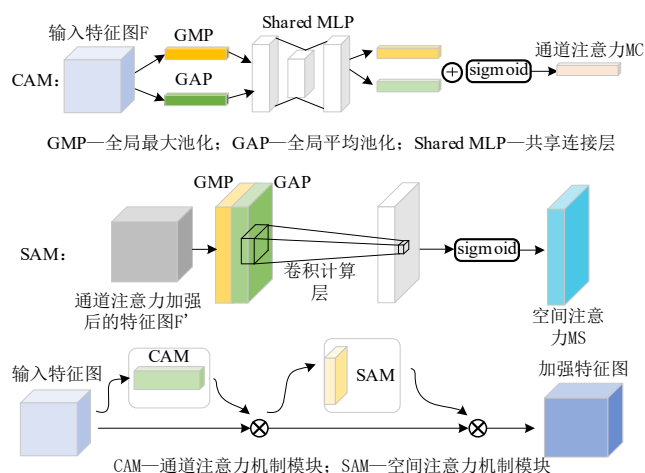


图2 CBAM结构

本研究在主干和颈部之间增加了串联的CBAM注意力模块和常规卷积块,增强了模型对图像中重要特征的捕捉和表达,同时增强了主干和颈部之间的有效信息交流和特征融合。其结构如图3所示。

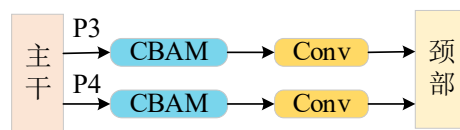


图3 颈部增强模块结构

1.3.3 SIoU

目标检测的有效性在很大程度上取决于损失函数的定义,传统的目标检测损失函数是交并比损失(Intersection over Union, IoU),通过比较预测框(A)和真实框(B)的交集面积和并集面积来度量它们之间的重叠程度。

YOLO使用的IoU损失函数为CIOU,忽略了预测框和真实框之间不匹配的方向,从而导致模型训练效率不高。SIoU^[12](SCYLLA-IoU)进一步考虑两框之间的向量角度,重新定义了相关损失,分别为角度损失、距离损失、形状损失和IoU损失,并重新设计了损失函数的惩罚项,考虑了预测框与真实框之间不匹配的方向,提高了模型的训练效率。

综上所述,本研究提出的自然环境下桃子成熟度检测模型YOLO-Faster结构如图4所示。在YOLOv5s的基础上,将主干特征提取网络替换为FasterNet使网络轻量化,同时在主干和颈部之间增加串联的CBAM卷积注意力模块和常规卷积块,增强模型对重要特征的提取以及主干和颈部之间的有效信息交流和特征融合,最后引入SIoU作为模型的损失函数缓解预测框与真实框之间方向不匹配的问题。

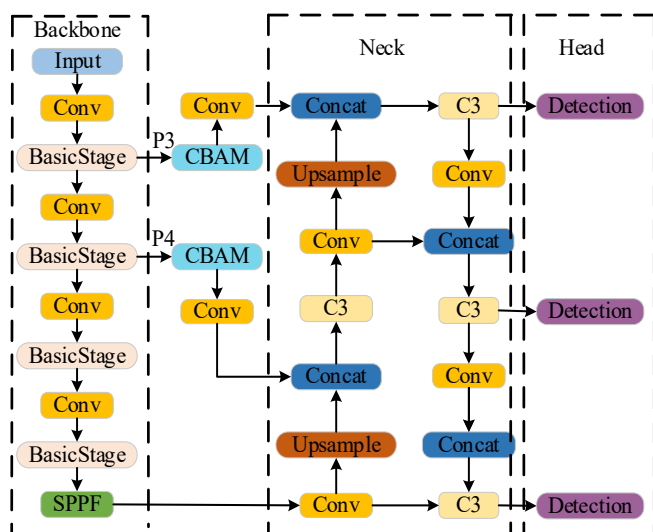


图4 YOLO-Faster 网络结构

1.4 评价指标

模型所用的评价指标为精确率 (Precision, P)、召回率 (Recall, R)、平均精度均值 (mean Average Precision, mAP)、模型权重大小、参数量、浮点运算量 (FLOPs) 以及推理速度。有关计算公式如下:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

$$AP = \int_0^1 P(R) dR \times 100\% \quad (3)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \times 100\% \quad (4)$$

其中, TP 为正确将目标预测为正样本的数量, FP 为错误将非目标预测为正样本的数量, FN 为漏检实际目标的数量, n 为检测对象类别, AP 为每一种类别的平均精度值, mAP 为所有类别的平均精度值。

表1 不同网络的消融试验结果

基线网络	试验	FasterNet	CBAM	Siou	P	R	mAP	权重大小 / Mb	FLOPs / G
YOLOv5s	1				80.6%	79.0%	87.6%	13.70	15.8
	2	✓			80.5%	80.5%	87.1%	8.28	8.7
	3	✓	✓		83.4%	78.2%	88.0%	8.32	8.8
	4	✓	✓	✓	84.3%	78.2%	88.6%	8.32	8.8

2.2 不同注意力机制对比试验

为了探究不同注意力机制对该模型的影响, 本研究在消融实验4的基础上使用四种注意力机制替换颈部增强模块中的CBAM注意力机制, 结果如表2所示。可以看出, 颈部增强模块使用CBAM注意力机制时, 模型拥有最好的综合性能。

2.3 不同检测模型对比试验

将改进的模型与YOLOv4、YOLOv5s、YOLOv7-tiny、YOLOv8s进行对比, 结果如表3所

1.5 实验环境及参数设置

本研究所进行的实验的软硬件环境如下: CPU为Intel®core™ i5-12490KF, GPU为NVIDIA GeForce RTX 3060, 操作系统为Windows 10, 使用的软件平台为PyCharm, 学习框架为Pytorch。采用AdamW优化器进行优化, Batchsize设置为16, 初始学习率为0.01, 使用以上参数训练100轮次。

本研究的模型部署于Jetson Nano平台, Jetson Nano是一款由NVIDIA推出的嵌入式人工智能计算平台, 采用高性能的GPU架构, 具备出色的并行计算能力。

2 结果与分析

2.1 消融试验

用改进后的YOLO-Faster和原始网络在桃子数据集上分别进行4组试验, 以便更好地分析改进后模型的效果, 消融试验结果如表1所示。

试验1是原始的YOLOv5s算法。试验2采用FasterNet为主干特征提取网络, 浮点运算量相比基线网络下降了7.1 G, 权重文件减少了5.42 Mb, 模型实现轻量化, 但同时 mAP 也稍有下降; 试验3在试验2的基础上, 在主干和颈部之间串联了卷积注意力模块CBAM, 模型的准确率上升了2.8个百分点, mAP 上升了0.4个百分点; 试验4在试验3的基础上使用Siou损失函数, 进一步提升了模型的准确率和 mAP , 分别达到了84.3%和88.6%。结果表明, 改进后的模型不仅分别将准确率和 mAP 提升3.7和1个百分点, 还大幅度地减少了浮点运算量和权重文件大小。

改进后的模型的 mAP 为88.6%, 高于其他模型, 同时浮点运算量、参数量和权重大小明显低于其他模型。YOLO-Faster在GPU和CPU上的检测时间分别为7.6 ms和96.5 ms, 在所有模型中检测时间均最短, 相比YOLOv5s分别快了1.1 ms和30.5 ms。此外, 本研究还将表中所列5种模型部署到嵌入式设备Jetson Nano上, 对比它们的检测速度, YOLO-Faster的检测时间为145.8 ms, 表现明显优于其他模型。

表2 不同注意力机制对比结果

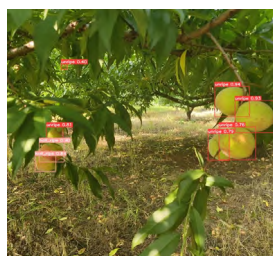
注意力机制	P	R	mAP	检测时间/ms	
				GPU	CPU
SE	81.0%	79.7%	87.1%	8.2	95.5
ECA	80.2%	80.5%	88.2%	8.1	95.3
CA	82.2%	80.2%	87.8%	8.5	96.9
TA	79.7%	80.9%	87.4%	8.6	99.5
CBAM	84.3%	78.2%	88.6%	7.6	96.5

表3 不同模型的检测性能对比

模型	P	R	mAP	检测时间/ms			FLOPs/G	参数量	权重大小/Mb
				GPU	CPU	Jetson Nano			
YOLOv4	80.7%	68.0%	78.0%	39.2	511.0	639.0	60.5	64.4×10^6	244.0
YOLOv5s	80.6%	79.0%	87.6%	8.7	127.0	209.5	15.8	7.0×10^6	13.7
YOLOv7-tiny	77.4%	83.8%	88.2%	8.6	116.9	184.7	13.0	6.0×10^6	11.7
YOLOv8s	78.4%	80.1%	87.9%	9.5	117.0	235.3	28.7	11.1×10^6	21.4
Ours	84.3%	78.2%	88.6%	7.6	96.5	145.8	8.8	4.1×10^6	8.3



(a) 遮挡



(b) 重叠



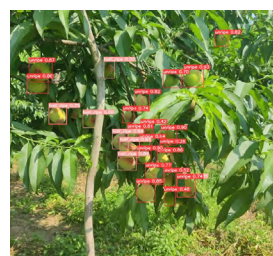
(c) 多目标



(d) 光照不足



(e) 逆光



(f) 重叠+遮挡+多目标

图5 YOLO-Faster模型在不同场景下的检测效果

3 结论

针对自然环境下桃子成熟度的快速检测问题,本研究在YOLOv5s算法的基础上进行改进,提出了YOLO-Faster桃子成熟度检测模型,并将其部署到嵌入式设备Jetson Nano上。YOLO-Faster在自建的数据集上训练后 mAP 为88.6%,浮点运算量为8.8 G,参数量为 4.1×10^6 ,权重大小为8.3 Mb,在GPU、CPU上的单幅图片检测时间分别为7.6 ms、96.5 ms,相较于YOLOv5s, mAP 提升了1个百分点,浮点运算量、参数量、权重大小和在GPU、CPU上的单幅图片检测时间分别降低了44.3%、41.4%、39.4%、12.6%和24%,在嵌入式设备上的检测时间降低了30.4%。结果表明,改进后的YOLO-Faster不仅 mAP 超越了YOLOv5s,还大幅度地减小了模型的参

为了验证YOLO-Faster在自然环境下对桃子成熟度的检测效果,分别选取了遮挡、重叠等自然环境下的图像进行检测。图5为模型在6种不同的自然条件下的检测效果图。从图5中可以看出,面对桃子生长所处的复杂的自然环境,模型仍能够对不同成熟度的桃子做出准确的识别。

数量、浮点运算量等,提升了检测速度,并且在嵌入式设备上也有着更好的综合性能,可以为桃子采摘机器人的视觉识别系统提供技术支持。

参考文献:

- [1] 王力荣.我国桃产业现状与发展建议[J].中国果树, 2021(10): 1-5.
- [2] ZHOU H, WANG X, AU W, et al. Intelligent robots for fruit harvesting: Recent developments and future challenges[J]. Precision Agriculture, 2022, 23(5): 1856-1907.
- [3] TIAN Y, YANG G, WANG Z, et al. Instance segmentation of apple flowers using the improved mask R-CNN model[J]. Biosystems engineering, 2020, 193: 264-278.
- [4] LIU T H, EHSANI R, TOUDESHEKI A, et al. Identifying immature and mature pomelo fruits in trees by elliptical model fitting in the Cr-Cb color space[J]. Precision Agriculture, 2019, 20: 138-156.

(下转第41页)

3 总结

1) 本设计的电机驱动电路均采用 MSP430G2553 微控制器和 AT8236 驱动芯片进行驱动, 采用 AMS1117-3.3 稳压模块对微控制器输入端稳压, 超声波测距模块采用了 HC-SRO4, 红外测距模块采用了 AM312, 压力感知模块采用了 KEYES, 采用 MG995 舵机和 Raspberry Pi Camera V2 摄像头, 在实现采摘机器人功能性的同时, 简化了电路结构, 使得技术实现方便且成本相对低廉。

2) 本设计的避障控制电路包括用于对农业全地形智能采摘机器人四周的障碍物进行检测的超声波测距模块、用于对农业全地形智能采摘机器人前进方向路面落差进行检测的前红外测距模块和后红外测距模块, 并借此实现对农业全地形智能采摘机器人四周的障碍物及路面严重坑洼均进行检测, 机器人避障功能强大。

3) 此电机驱动电路的设计多采用现有技术进行改造融合, 设计稳重合理, 使电路得到高稳定性和高可靠性, 能够实现对摄像头旋转电机、底座旋转电机、旋转臂驱动电机、摆臂驱动电机、手爪旋转电机稳定可靠的驱动。

4) 本设计核心控制模块采用 Raspberry Pi 4B 开发板, 技术稳定, 维修便捷, 且具备强大的功能扩展能力。

5) 本设计应用于能够实现各种功能(如抢险救灾、科考探测、智能家政、寻航喷淋)的履带机器人

上, 为机器人的行进驱动及避障识别提供了硬件电路设计参考方案, 具有强实用性, 便于推广与应用。

综上所述, 课题组所设计的电路结构简洁, 电路具备高稳定性、高可靠性和强实用性, 技术实现方便, 成本低廉、功能完备且功能扩展性强, 便于推广及应用, 可搭配收获机械、喷淋设备等, 也可投入抢险救灾、科考探测、物流运输等其他领域。

参考文献:

- [1] 杨立爽, 吴波, 潘相龙, 等. 全地形智能农业采摘机器人的设计[J]. 南方农机, 2023, 54(9): 4-6+20.
- [2] 黄钰, 李秀壁, 温德涌, 等. 电工电子技术(微课视频版)[M]. 北京: 清华大学出版社, 2023.
- [3] 张雪芹, 宋继荣. 电工电子实验教程[M]. 3版. 上海: 华东理工大学出版社, 2022.
- [4] 雅各布·弗雷登. 现代传感器手册: 原理、设计及应用[M]. 5版. 宋萍, 隋丽, 潘志强, 译. 北京: 机械工业出版社, 2020.
- [5] Stephen D. Umans. 电机学[M]. 7版. 刘新正, 苏少平, 高琳, 译. 北京: 电子工业出版社, 2021.
- [6] 何新强, 王宇俊, 谭兴军. 偏心圆腿六足机器人控制电路设计[J]. 河南教育学院学报(自然科学版), 2011, 20(1): 36-40.
- [7] 毕盛, 闵华清, 温正东, 等. 双足步行机器人控制电路设计与实现[J]. 科学技术与工程, 2010, 10(31): 7661-7664.
- [8] 皮骄阳, 李永新, 李鹏, 等. 基于 DSP 的足球机器人控制电路设计研究[J]. 电气自动化, 2006, 28(1): 27-29.
- [9] 吴善强, 孙立宁. 基于 DSP2407a 的爬壁机器人控制系统硬件电路设计[J]. 林业机械与木工设备, 2007(3): 35-37.
- [10] 孟祥荔. 飞滑式巡线机器人控制电路的设计研究[J]. 科技创新与应用, 2018(17): 91-93.

(上接第 27 页)

- [5] LIN G, TANG Y, ZOU X, et al. Fruit detection in natural environment using partial shape matching and probabilistic Hough transform[J]. Precision Agriculture, 2020, 21: 160-177.
- [6] 李宇豪, 孙宇朝, 杨玮. 基于深度学习的水果果实视觉检测技术研究进展[J]. 烟台果树, 2022(4): 6-8.
- [7] 崔潇, 马原东, 倪照凤. 基于深度学习的桃子检测研究[J]. 贵州科学, 2019, 37(2): 82-86.
- [8] LIU G, NOUAZE J C, TOUKO MBOUEMBE P L, et al. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3[J]. Sensors, 2020, 20(7): 2145.
- [9] 王立舒, 秦铭霞, 雷洁雅. 基于改进 YOLOv4-Tiny 的蓝莓成熟度识别方法[J]. 农业工程学报, 2021, 37(18): 170-178.

- [10] CHEN J, KAO S H, HE H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 12021-12031.
- [11] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module[Z/OL]. (2018-07-18). <https://arxiv.org/pdf/1807.06521.pdf>.
- [12] GEVORGYAN Z. SiLU Loss: More Powerful Learning for Bounding Box Regression[Z/OL]. (2022-05-25). <https://arxiv.org/ftp/arxiv/papers/2205/2205.12740.pdf>.