

基于改进 YOLO v5 的自然环境下樱桃果实识别方法

张志远^{1,2} 罗铭毅^{1,2} 郭树欣^{1,2} 刘刚^{1,2} 李淑平^{1,2} 张瑶^{1,2}

(1. 中国农业大学信息与电气工程学院, 北京 100083;

2. 中国农业大学现代精细农业系统集成研究教育部重点实验室, 北京 100083)

摘要: 为提高对樱桃果实识别的准确率,提升果园自动采摘机器人的工作效率,使用采集到的樱桃原始图像以及其搭配不同数据增强方式得到的数据图像共 1 816 幅建立数据集,按照 8:2 将数据集划分成训练集与测试集。基于深度学习网络,利用 YOLO v5 模型分别对不同数据增强方式以及组合增强方式扩增后的樱桃数据集进行识别检测,结果表明离线增强与在线增强均对模型精度提升有一定的正向促进作用,其中采用离线数据增强策略能够显著且稳定的增加检测精度,在线数据增强策略能够小幅度提高检测精度,同时使用离线增强以及在线增强能够最大幅度的提升平均检测精度。针对樱桃果实之间相互遮挡以及图像中的小目标樱桃检测难等导致自然环境下樱桃果实检测精度低的问题,本文将 YOLO v5 的骨干网络进行改动,增添具有注意力机制的 Transformer 模块,Neck 结构由原来的 PAFPN 改成可以进行双向加权融合的 BiFPN,Head 结构增加了浅层下采样的 P2 模块,提出一种基于改进 YOLO v5 的自然环境下樱桃果实的识别网络。实验结果表明:相比于其他已有模型以及单一结构改进后的 YOLO v5 模型,本文提出的综合改进模型具有更高的检测精度,使平均精度均值 2 提高了 2.9 个百分点。结果表明该方法有效的增强了识别过程中特征融合的效率 and 精度,显著地提高了樱桃果实的检测效果。同时,本文将训练好的网络模型部署到安卓(Android)平台上。该系统使用简洁,用户设备环境要求不高,具有一定的实用性,可在大田环境下对樱桃果实进行准确检测,能够很好地满足实时检测樱桃果实的需求,也为自动采摘等实际应用奠定了基础。

关键词: 樱桃; 目标检测; YOLO v5; 数据增强; 注意力机制

中图分类号: S126 **文献标识码:** A **文章编号:** 1000-1298(2022)S1-0232-09

Cherry Fruit Detection Method in Natural Scene Based on Improved YOLO v5

ZHANG Zhiyuan^{1,2} LUO Mingyi^{1,2} GUO Shuxin^{1,2} LIU Gang^{1,2} LI Shuping^{1,2} ZHANG Yao^{1,2}

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agricultural University, Beijing 100083, China)

Abstract: In order to improve the accuracy of cherry fruit recognition and the working efficiency of orchard automatic picking robot, totally 1 816 sets of cherry original images collected in Yantai Academy of Agricultural Sciences and data images obtained with different data enhancement methods were used to establish the data set, the data set was divided into training set and test set according to rate of 8:2, and YOLO v5 model was used to identify and detect cherry data sets enhanced by different data enhancement methods and combined enhancement methods based on the in-depth learning network. The results showed that offline enhancement and online enhancement had a certain positive effect on the improvement of model accuracy. The offline data enhancement strategy could significantly and stably increase the detection accuracy, and the online data enhancement strategy could slightly improve the detection accuracy. Using the combination of offline enhancement and online enhancement at the same time could greatly improve the average detection accuracy. In view of the mutual occlusion between cherry fruits and

收稿日期: 2022-06-14 修回日期: 2022-08-25

基金项目: 山东省自然科学基金项目(ZR2020MC084)

作者简介: 张志远(1998—),男,硕士生,主要从事深度学习在农业智能检测方面的应用研究,E-mail: zhangzhiyuan@cau.edu.cn

通信作者: 张瑶(1988—),女,副教授,主要从事农业定量遥感应用研究,E-mail: zhangyao@cau.edu.cn

the difficulty in detecting small cherry targets in the picture, the detection accuracy of cherry fruits in the natural environment was low, the backbone network of YOLO v5 was changed, the transformer module with attention mechanism was added, and the neck structure was changed from the original pafpn to bifpn which could carry out two-way weighted fusion. The P2 module of shallow down sampling was added to the head structure. The experimental results showed that compared with other existing models and the improved YOLO v5 model with a single structure, the comprehensive improved model proposed had the highest detection accuracy, and the mAP@ 0.5 : 0.95 was increased by 2.9 percentage points. The results showed that this method effectively enhanced the efficiency and accuracy of feature fusion in the recognition process, and significantly improved the detection effect of cherry fruit. At the same time, the trained network model was deployed on the Android platform. The system was simple and clear to use, and the requirements of user equipment environment were not high. Therefore, the system had certain practicability. It could detect cherry fruit in real time and accurately in the field environment, which laid a foundation for practical applications such as automatic service picking in the future.

Key words: cherry; target detection; YOLO v5; data enhancement; attention mechanism

0 引言

当前樱桃采摘通常采用人工方式,由于樱桃成熟期短^[1],农村劳动人口大大减少,采摘劳动强度大,樱桃的种植规模被约束。自动化采摘机器人可以提高采摘效率,降低对果实的损伤,又能够减少人工成本^[2]。但采摘机器人的运行环境复杂,不确定性因素多,造成采摘难度大。高效快捷的果蔬采摘需要精准的目标识别和三维定位的支撑。采摘机器人的视觉系统运行经历 4 个阶段,分别是目标检测、目标识别^[3]、三维重建和三维定位。目标识别和定位的精度将直接决定采摘机器人的采摘效率,农作物是否会受到破坏,采摘机器人本体是否碰撞损坏等。

传统数字图像处理技术需要精确的目标果实特征信息,但在光照条件变化、果实存在遮挡^[4]以及果实与背景颜色相似情况下,该方法不能保持较高的检测精度。随后,基于数据的机器学习被用于目标检测和识别,但机器学习需要提前设定参数,且参数对目标分类结果影响较大。近年来,鉴于深度学习对目标高维特征^[5]具有较强提取能力,故被广泛运用于采摘机器人目标检测识别中。而尽管基于深度学习的目标识别方法提高了精度和实时检测能力^[6],但是面对图像中小目标的樱桃果实以及被遮挡的樱桃果实等复杂情景下的识别能力还难以满足实际的需求。

对深度学习领域提出的目标检测算法按照结构上可以大体划分为双阶段(Two-stage)检测算法和单阶段(One-stage)检测算法^[7]。传统的目标检测算法,是将整个目标检测分为两个阶段:候选区域选择以及候选区域目标分类,典型的算法为 R - CNN、Fast R - CNN、Faster R - CNN 以及 Mask R - CNN 等系列算法。双阶段目标检测方法中存在区域生成这

一计算过程,实现了检测从粗到细的一个过程,但导致整体计算量增加,增加了训练难度,同时难以满足低硬件环境的实时推理需求。为了解决算法效率问题,研究人员提出了端到端的单阶段检测算法,最为经典的是 YOLO(You only look once)系列算法,其中 YOLO v5 目标检测算法是目前较为先进的目标检测算法,其平均检测速度相较于 YOLO v4 有大幅的提升,平衡了总体的精度与速度,在灵活性与速度上远优于 YOLO v4,因此在模型的快速部署上具有非常大的优势,能够适用于自然环境下樱桃果实的快速检测。

本研究提出一种基于改进 YOLO v5 的目标检测方法,优化樱桃果实实时检测流程,克服复杂多变的环境背景因素造成的误差和影响,提升果实检测性能。

1 樱桃图像数据采集与图像预处理

1.1 数据采集

山东省是全国闻名的樱桃产量大省,而烟台市也是山东省产出樱桃的重要支柱^[9]。本研究以山东大樱桃^[10]为研究对象,在烟台市农业科学研究院以及烟台福山樱桃种植基地进行数据采集。在 2021 年 5—7 月,在樱桃树种植基地选择多棵长势较好、果实成熟且完整的樱桃树,并使用 iphone12 的超广角摄像头进行拍摄。数据中包含多方向、多品种、不同距离、不同果实密集度、不同时间段的果树图像,以保证识别的稳定性和准确度。最终采集到 1 726 幅高清晰度成熟樱桃图像,图像格式为 JPG,图像的分辨率为 4 000 像素 × 3 000 像素。与其他果实数据集不同的是,自然状况下,樱桃果实存在大量的果实密集重叠、叶片遮挡。如图 1 所示。

1.2 数据筛选以及图像标注

本研究主要目的是提高对自然环境下的樱桃果



图 1 成熟樱桃图像
Fig.1 Ripe cherry diagram

实识别的精度,因此在筛选过程中首先删除采集到的樱桃图像中部分文件受到损坏的,以及拍摄设备没有完全聚焦导致图像模糊等人为采集时无法避免的意外情况对检测算法识别造成干扰的图像。同时,为了提升整体模型的鲁棒性,在选择樱桃图像时,着重选择一部分包括树枝或树叶遮挡樱桃果实以及樱桃果实之间相互遮挡的自然情况下的图像。最终保留的是从不同角度(从多个角度拍摄的樱桃图像能够让 YOLO 目标检测算法全方位的学习樱桃果实各个角度的表层特征)、不同光照、不同距离拍摄的清晰樱桃图像,便于 YOLO 目标检测算法更好地去学习标注好的樱桃果实目标的细节特征,提高整体模型识别精度。

通过数据筛选,得到高质量图像共计 454 幅。使用 Labellmg 软件对得到的 454 幅图像进行樱桃果实标注,得到包含樱桃果实的中心坐标 (x_c,y_c) 、宽、高信息的 xml 文件。由于 YOLO 需要的标注文件类型为 txt 文件,因此使用 python 编程将 xml 文件转化为 txt 格式的标注文件,得到了可以在 YOLO 模型中运行的数据集。YOLO 数据格式如图 2 所示。

class	$\bar{x}_c$	$\bar{y}_c$	$\bar{w}$	$\bar{h}$
-------	-------------	-------------	-----------	-----------

图 2 YOLO 数据格式
Fig.2 Data format of YOLO

其中 class 表示数值化的类别标签,标签值从 0 开始索引,就本研究而言,class 值只有 0; $\bar{x}_c$ 、 $\bar{y}_c$ 、 $\bar{w}$ 、 $\bar{h}$ 的计算式为

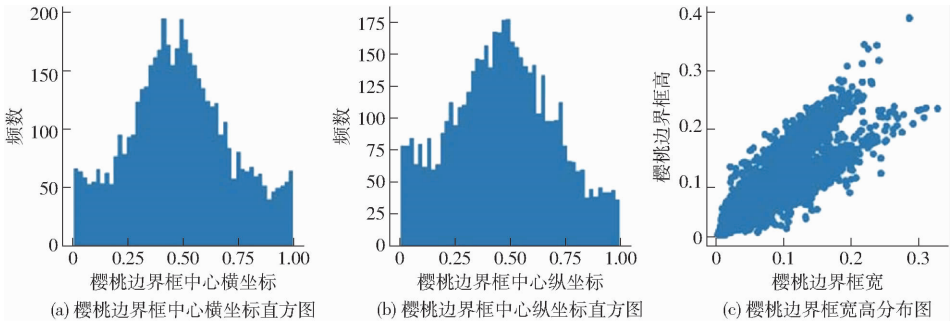


图 4 樱桃边界框统计
Fig.4 Statistics for cherry bounding box

$$\begin{cases} \bar{x}_c = \frac{x_c}{W} \\ \bar{y}_c = \frac{y_c}{H} \end{cases} \tag{1}$$

$$\begin{cases} \bar{w} = \frac{w}{W} \\ \bar{h} = \frac{h}{H} \end{cases} \tag{2}$$

式中 W ——图像宽度 H ——图像高度
 $\bar{w}$ ——归一化后目标物体宽度
 $\bar{h}$ ——归一化后目标物体高度
 $\bar{x}_c$ ——归一化后目标中心 x 坐标
 $\bar{y}_c$ ——归一化后目标中心 y 坐标

图 3 为樱桃边界框标注示意图。

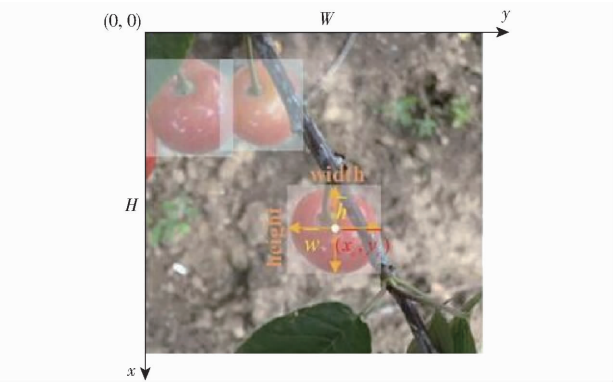


图 3 樱桃数据标注示意图
Fig.3 Cherry data annotation diagram

对标注的樱桃边界框数据进行统计,得到图 4 所示的统计结果,图中边界框坐标中心 (x,y) 和边界框的宽高已经经过归一化处理。从统计结果可以看出,因为 1 幅图像中存在多个樱桃果实,因此 454 幅图像经过标注后得到了 5 004 个樱桃图像实例。

从图 4a、4b 可以看出,大部分樱桃边界框的中心 (x,y) 分布在整个图像中心,且呈现正态分布规律;可以看出其纵横比接近 1,没有出现太大的形状差异,这也和樱桃数据本身特点有关,因为樱桃形状呈椭圆形,使用矩形框标注的宽高基本接近。这种较为规则的标注框也降低了检测模型在位置预测过程中的过拟合情况。另外从图 4c 还可以发现樱桃

数据集的一个特点,边界框的面积占比大都在 9% 以内(宽高之积与归一化总面积之比),因此总体呈现出小目标的特点,这也是在后续算法改进中需要解决的一个重点问题。

1.3 数据增强

采用不同数据增强方式,记录果园樱桃果实识别准确性的波动。选择 YOLO v5 模型作为基准模型,对不同数据集进行训练和测试。

将原始数据集中 20% 数据作为测试集,共 91 幅图像,979 个标注实例。训练集、验证集和测试集的划分比例为 7:1:2,对扩增后的数据集,按照 8:2 划分训练集和验证集,测试集采用原始数据集分割得到的测试集,数据划分结果如表 1 所示。训练和验证时以离线扩增和 Mixup 扩增后的数据集的数量为标准数量,其他在数量上与该数据集有差异的数据集在训练前通过复制一定数量以保证各个训练集和验证集数量基本相等。例如原始数据集在训练前将训练集与验证集按照 10 倍进行复制,Mixup 扩增

数据集按照 2 倍进行复制。

表 1 数据集图像划分

Tab. 1 Dataset partitioning				幅
数据集	训练集	验证集	测试集	总数
原始数据集	318	45	91	454
Mixup 扩增数据集	1 543	363		1 816
Mosaic 扩增数据集	1 543	363		1 816
离线扩增数据集	1 543	363		1 816
Mixup 扩增与离线扩增	3 086	726		3 632
Mosaic 扩增与离线扩增	3 086	726		3 632

首先分析不同扩增方式下每一轮训练后在验证集上的效果,如图 5 所示。从结果可以看出,在只有原始数据集的基础上进行 Mixup 或者 Mosaic 在线扩增都提高了平均精度均值 (Mean average precision, mAP) 和召回率,精度大都比较接近。没有任何在线扩增方式下,只采用离线扩增的方式在验证集的性能评价指标上也优于原始数据集训练得到的模型,说明数据扩增能使模型性能得到提升。

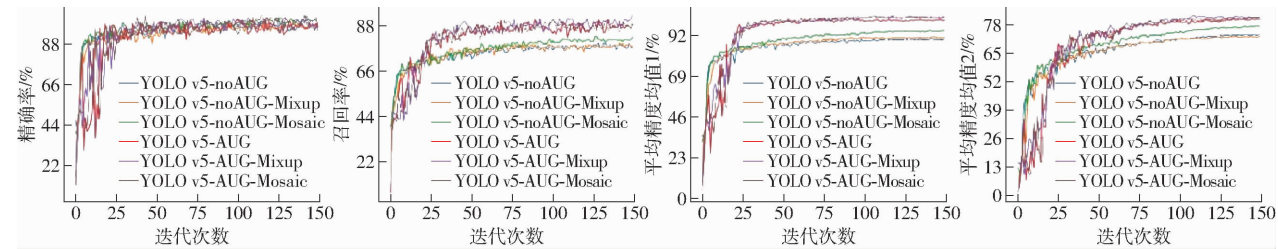


图 5 不同数据集训练过程验证集性能评估结果

Fig. 5 Performance evaluation results of validation sets during training of different data sets

从图 5 中还可以得知(图中平均精度均值 1 指 mAP@0.5,平均精度均值 2 指 mAP@0.50~0.95),采用离线扩增方式后模型在前 25 次训练过程中较使用原始数据集具有更多震荡,其原因是离线扩增加入了不同的图像变换,样本比较丰富,模型在前 25 轮训练中还未能学习到经离线扩增后的完整特征。

图 6 为 150 轮训练过程中在训练集和验证集的锚框损失值以及目标损失值,锚框损失描述预测框与锚框之间的匹配程度,目标损失描述预测物体类别与真实类别的差异程度。从图中可以看出,使用离线扩增数据集其锚框损失越小,目标损失也相对较低,说明经过扩增后模型收敛更快。

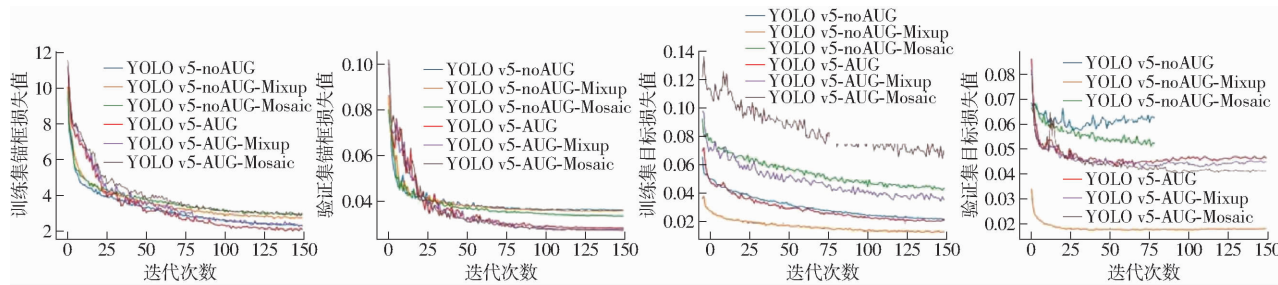


图 6 不同数据集训练过程损失值变化曲线

Fig. 6 Loss in training process of different data sets

通过不同的数据集训练得到最优模型后,在测试集上检测模型性能,其结果如表 2 所示。

从表 2 可以看出,在原始数据集基础上使用 Mixup 和 Mosaic 在线扩增策略使平均精度均值 2 分别提升 0.9、1.3 个百分点,而使用离线扩增方式其

平均精度均值 2 提升 4.2 个百分点,在使用离线扩增数据集的基础上采用 Mixup 和 Mosaic 在线扩增策略使平均精度均值 2 提高 4.9、5.2 个百分点,在有些指标上反而下降了。使用离线增强与 Mosaic 增强能够最大限度的提高数据容量,丰富数据集的

表 2 不同训练数据所得模型测试集性能评估结果

数据集类型	精确率	召回率	平均精度	
			均值 1	均值 2
原始数据集	93.4	89.9	94.5	72.6
原始数据集 + Mixup	93.1	89.3	93.8	73.5
原始数据集 + Mosaic	94.0	90.1	94.0	73.9
离线扩增数据集	97.8	90.4	95.5	76.8
离线扩增数据集 + Mixup	96.1	90.4	94.8	77.5
离线扩增数据集 + Mosaic	95.7	90.5	95.0	77.8

多样性,以小容量样本获得大容量数据集,并且由于数据集的多样性获得了提升,对于目标检测的精度也能够获得一定程度的提升。AUG - Mosaic 模型的性能如图 7 所示。

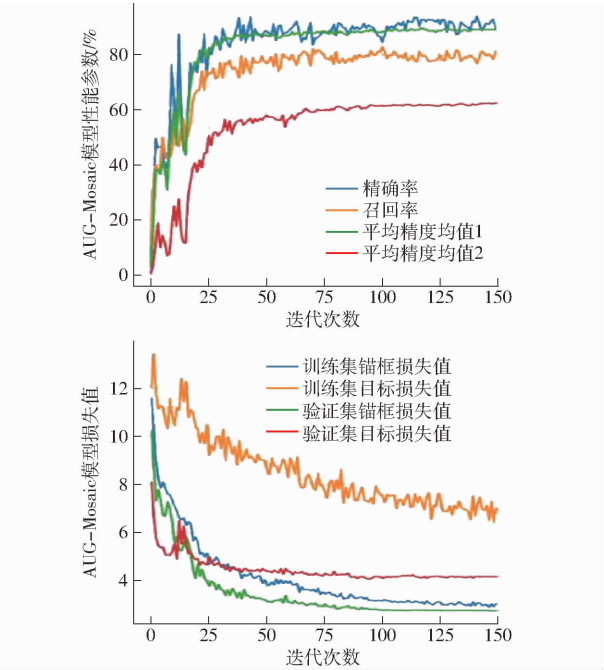


图 7 AUG - Mosaic 模型的性能参数与损失值变化曲线

Fig. 7 Performance and loss of AUG - Mosaic model

采用离线数据增强策略能够显著且稳定的增加检测精度,在线数据增强能够小幅度提高检测精度,同时使用离线增强以及在线增强能够大幅度地提升平均检测精度,但是稳定性将会略微下降。

2 基于改进 YOLO v5 的樱桃识别模型

2.1 YOLO v5 算法结构优化

目前主流的目标检测算法大都在公共数据集 COCO^[11]上进行训练和性能评估^[12],该数据集在目标检测方面有超过 1.2×10^5 幅图像,标注对象超 8.8×10^5 个,共 80 个类别^[13],涵盖多种场景。采用该数据集可以很大程度上抑制模型的过拟合,也是检验目标检测模型鲁棒性最简单有效的方法。但这也对模型参数训练带来弊端,检测类别太多会抑制某些特征的表达^[14],本研究只涉及到樱桃这一类别,为增强特征表达,提高樱桃检测性能,对 YOLO v5 算法中骨干网络、Neck 网络以及 Head 网络等结构进行优化。

2.1.1 骨干网络优化

首先对 YOLO v5 模型骨干网络进行改进,卷积模块最大的弊端是需要手动设置卷积参数,例如卷积核数量(即输出通道数)、卷积核尺寸、步长、组别数等。

本研究借鉴 Transform 思想,对 YOLO v5 的 CSPDarknet53 骨干网络进行了优化改进。图 8 为 YOLO v5 模型的原始骨干网络以及改进后的骨干网络结构示意图。从图 8 中可以看出,改进后的网络将用于感受野特征融合的池化金字塔网络前 CSP Bottleneck 模块替换成了 Transformer 模块,其输出结果与 P3、P4 和 P5 特征输出共同输入至下一级 Neck 网络。

相比于 CSP Bottleneck 模块,Transformer 模块具

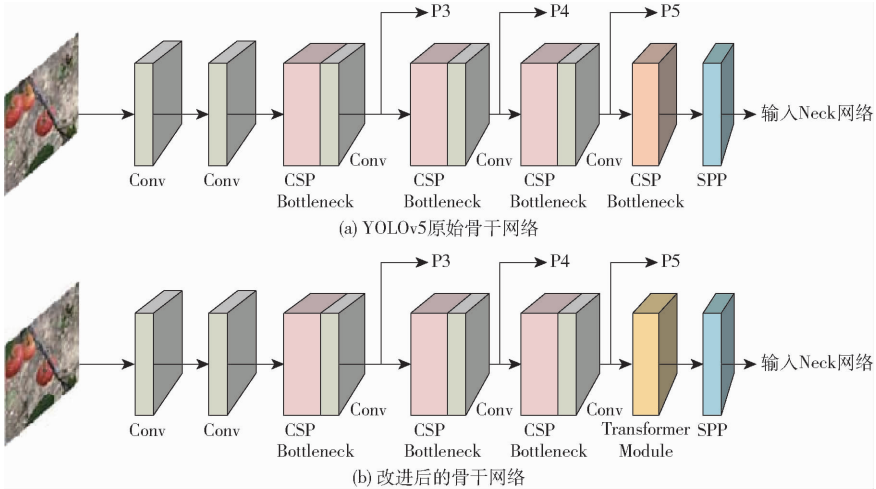


图 8 改进后的骨干网络与原始网络对比

Fig. 8 Comparison between improved backbone network and original network

有注意力机制,可以对全局特征进行重点提取。这一点可以帮助检测系统有意识的提取出属于樱桃的特征点,从而实现更好的检测效果以及更高的检测精度。

2.1.2 Neck 网络优化

PANet 的提出证明了双向融合的有效性^[15],而 PANet 的双向融合较为简单,本文在此基础上引入了 BiFPN 结构对其进行改进,如图 9 所示。从图 9 中可以看出,特征金字塔结构采用简单的自上而下特征融合方式,这种结构在 Faster R - CNN、RetinaNet 和 YOLO v3^[16]中较常见,而且实验结果也表明这种特征融合方式取得了很好的检测效果。在 YOLO v4 和 YOLO v5 模型中特征融合方式采用 PANET 结构,可以看出,该结构在 FPN 基础^[17]上增加了自下而上的特征融合分支。

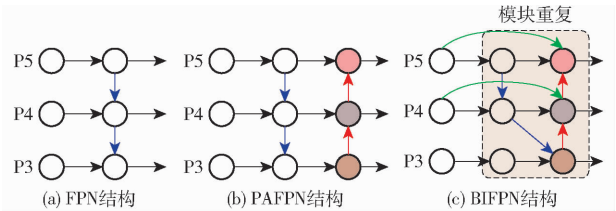


图 9 不同 Neck 网络结构对比

Fig. 9 Comparison diagram of different neck network structures

以上两种结构在特征融合过程中信息流都是基于临近特征结果,本研究使用 BiFPN 结构作为特征融合方法,这是一种高效的双向跨尺度和加权特征融合方式。

针对目前樱桃识别的研究,采用 BiFPN 结构作为 Neck 网络进行特征融合主要解决了以下问题:

(1)樱桃小目标检测问题:BiFPN 作为模块重复的特征网络层,相比 PAFPN 可以获得更高级的特征融合方式,增加各尺度特征的耦合,特别是有助于小目标检测的浅层特征。

(2)樱桃目标重合问题:由于 BiFPN 采用了跨尺度连接方式,可以对不同检测特征按照跨尺度权重进行抑制或者增强特征表达,从而缓解检测目标重合导致的识别不准确问题。

2.1.3 Head 网络优化

在 YOLO v5 算法中,Head 网络^[18]的输入特征分别选取下采样为 32、16 和 8 的卷积输出结果,即分别对应图中的 P5、P4 和 P3。下采样越大,卷积越深,其特征的语义表达越强,一般而言对分类效果更好,但失去了更多的位置信息;而下采样较小的浅层特征包含更多位置信息,有利于小目标物体的位置检测^[19]。根据神经网络的这一特点,本文对 Head 输出网络进行了如下优化:将浅层特征输出 P2 (4

倍下采样率)作为 Neck 网络的一个输入特征,共同融合 P3、P4 和 P5 特征,Neck 网络输出 4 个不同尺度^[20]的特征作为 Head 分支的输入,从而提高小目标位置检测精度。其优化结果如图 10 所示。

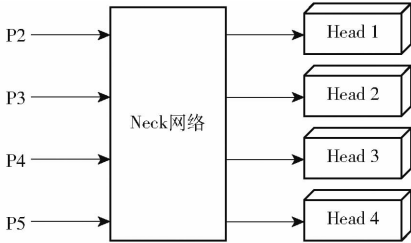


图 10 Head 网络优化

Fig. 10 Head network optimization

2.2 算法优化实验对比

训练数据集统一采用离线扩增后的数据,不采用任何在线增强策略,测试集采用原始数据集 20% 的数据,表 3 为不同模型与其对应说明。

表 3 模型名称及其说明

Tab.3 Model name and description

模型名称	说明
YOLO v5	YOLO v5 模型
YOLO v5 - Transformer	骨干网络增加 Transformer 模块
YOLO v5 - BiFPN	Neck 网络增加 BiFPN 模块
YOLO v5 - Transformer - BiFPN	Transformer 模块与 BiFPN 模块
YOLO v5 - P2	Head 网络利用 P2 特征
YOLO v5 - BiFPN - P2	Neck 网络增加 BiFPN 模块, Head 网络利用 P2 特征
YOLO v5 - Transformer - P2	骨干网络增加 Transformer 模块, Head 网络利用 P2 特征
Faster R - CNN	双阶段检测算法
Mask R - CNN	双阶段检测算法
YOLO v3	单阶段检测算法
SSD 512	单阶段检测算法
RetinaNet	单阶段检测算法

不同 YOLO v5 改进算法在 150 轮训练过程中的性能指标如图 11 所示。

从图 11 可得,单独使用 Transformer 模块或者单独使用 BiFPN 模块并未对模型性能带来提升,而同时结合 Transformer 模块和 BiFPN 模块后模型性能得到了大幅提升,尽管在前 25 轮计算中比较振荡,但模型训练到后期各性能指标都超越了基准模型和单模块优化的模型,且训练过程中在验证集的锚框损失和目标损失都较其他模型快速收敛,且损失降低至一个较低的值,说明具有注意力机制 Transformer 模块和高效特征融合 BiFPN 模块对目标分类和检测具有一定效益,如图 12 所示。

通过训练得到各模型优化参数,使其在统一的测试集上进行测试,结果如表 4 所示。

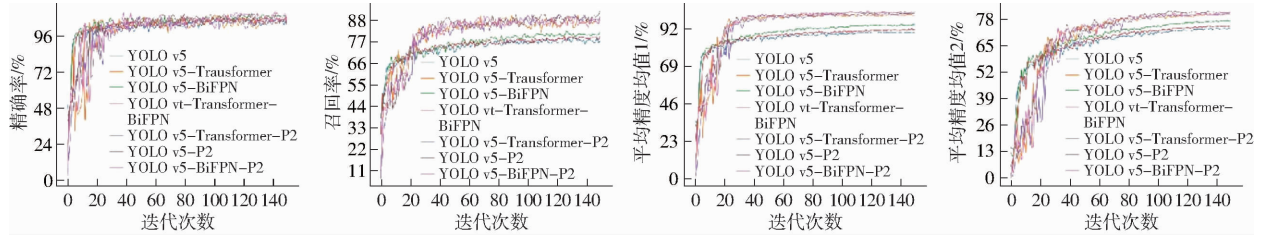


图 11 不同 YOLO 改进算法在训练过程中的性能参数变化曲线

Fig. 11 Performance results of different YOLO improved algorithms in training process

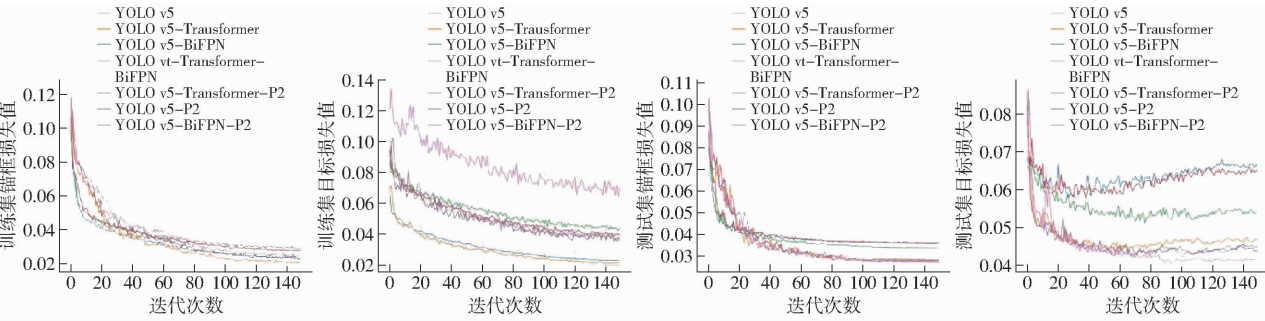


图 12 不同 YOLO 改进算法在训练过程中的损失

Fig. 12 Loss of different YOLO improved algorithms in training process

表 4 算法优化实验测试集性能评估结果

Tab.4 Performance evaluation results of algorithm optimization experiment test set

%

模型	精确率	召回率	平均精度	
			均值 1	均值 2
YOLO v5	97.8	90.4	95.5	76.8
YOLO v5 - Transformer	97.6	89.9	95.2	77.1
YOLO v5 - BiFPN	97.9	91.0	95.9	77.9
YOLO v5 - Transformer - BiFPN	98.4	91.6	96.3	79.7
YOLO v5 - Transformer - P2	98.2	92.3	97.1	78.6
YOLO v5 - P2	97.2	89.6	94.9	76.1
YOLO v5 - BiFPN - P2	98.0	91.2	99.0	78.2

从表 4 可得,相比于 YOLO v5 模型,单独采用 Transformer 模块、单独采用 BiFPN、单独采用 P2 特征的模型在测试集上的平均精度均值 2 分别提升 0.3、1.1、-0.7 个百分点,该结果与在训练过程中验证集的性能结果相近,说明单独应用一种优化算法对樱桃数据检测性能提升有限。

当优化方法进行两两组合时,其检测效果得到提升,从表 4 可以得出,Transformer 模块与 BiFPN 共同优化使平均精度均值 2 提升 2.9 个百分点,Transformer 模块与 P2 特征共同优化该消融使平均精度均值 2 提升 1.8 个百分点;BiFPN 网络与 P2 特征共同优化使平均精度均值 2 提升 1.4 个百分点。

通过消融实验,证实了对 YOLO v5 算法进行结构优化可以有效提升樱桃检测性能,这种提升需要对骨干网络、Neck 网络和 Head 网络共同优化改进,其中骨干网络优化结合 Neck 网络优化得到的

YOLO v5 - Transformer - BiFPN 模型能够获得最优检测效果。Transformer 与 BiFPN 共同优化模型的性能与损失值如图 13 所示。

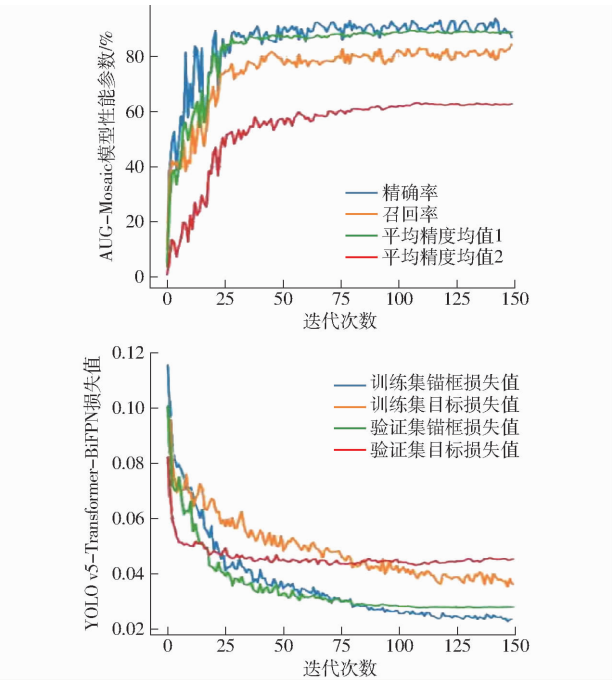


图 13 YOLO v5 - Transformer - BiFPN 的性能参数与损失变化曲线

Fig. 13 Performance and loss performance of YOLO v5 - Transformer - BiFPN

2.3 算法优化后的樱桃实例检测

使用上述优化算法,针对樱桃图像实例进行了测试,如图 14 所示。在樱桃实例检测中,检测效果通过误检情况、漏检情况与正例的置信度区间 3 个指标评价。

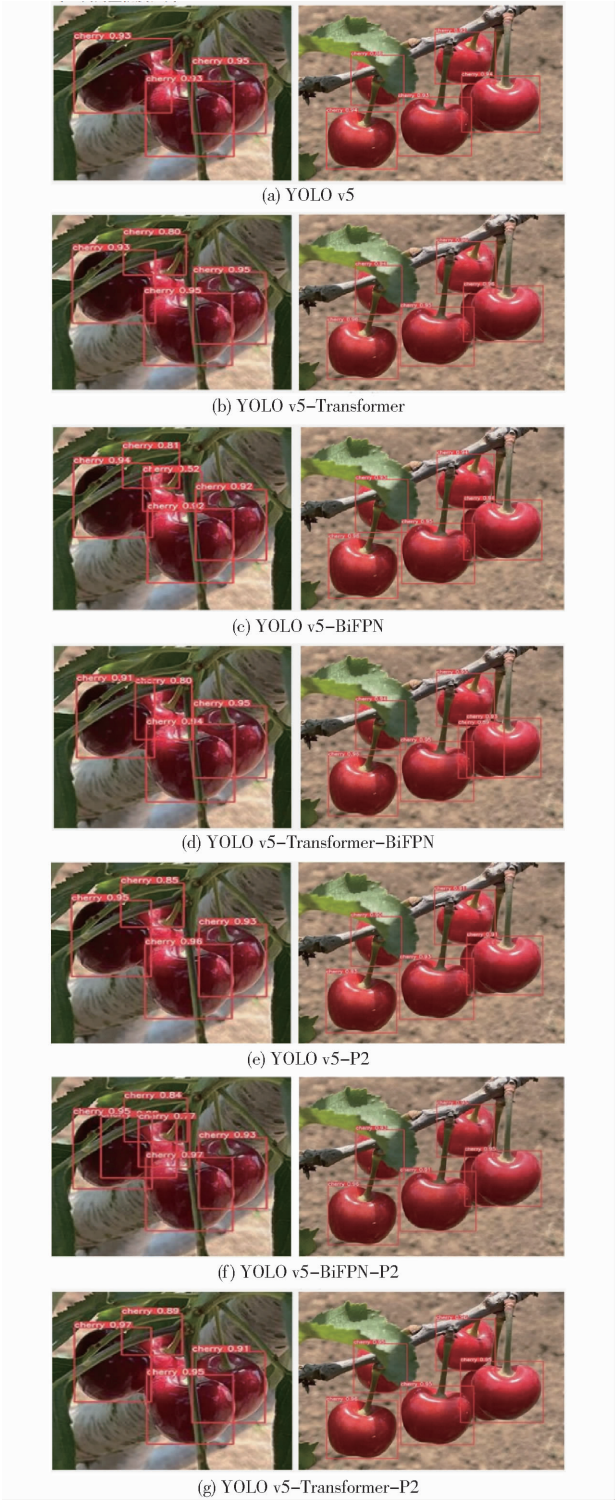


图 14 算法优化后樱桃实例检测对比

Fig. 14 Comparison of cherry instance detection after algorithm optimization

本文选择了不同光线条件、不同环境背景、不同遮挡情况下的两个场景进行樱桃实例检测对比分析。从图 14 可以看出,对于左图像场景,YOLO v5 模型存在漏检重叠樱桃的情况,而其余改进后的模型均能够识别出全部樱桃果实。其中 YOLO v5 - BiFPN - P2 与 YOLO v5 - BiFPN 模型则出现了过多的检测结果,存在误检情况;其余模型恰好正确识别

出樱桃数量,但仅有 YOLO v5 - Transformer - BiFPN 能够正确地检测出重叠樱桃的真实位置。

右图的背景简单、光线条件良好,在果实识别中可以判断为几乎没有外界因素干扰的图像,着重关注点为最右侧两个重叠樱桃中后一个樱桃的检测。在所有模型中仅有 YOLO v5 - Transformer - BiFPN 能够识别出这一重叠度极高的樱桃,而其余模型则无法识别出这一樱桃。

综上所述,本文进行算法优化后得到的 YOLO v5 - Transformer - BiFPN 模型在樱桃实例检测中能够达到最佳的检测效果,在保证识别准确度的前提下,能够有效地防止误检和漏检情况发生。

3 模型安卓端部署

本文将训练好的 YOLO v5 - Transformer - BiFPN 模型部署到移动设备的应用系统上,建立基于 Android 平台的樱桃果实实时检测系统,本系统导出安卓 apk 安装包,Android SDK 版本为 29,该系统基本满足了樱桃果实检测所需的功能,图 15 为功能模块流程图。该系统使用简单,用户设备环境要求不高,因此该系统具有一定的实用性,可在大田环境下实时准确地进行樱桃果实检测。

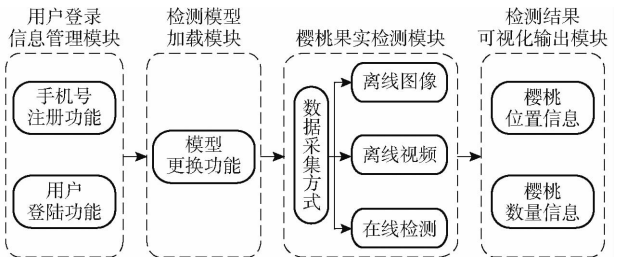


图 15 功能模块流程图

Fig. 15 Flow chart of functional models

本系统开发环境如下:Android Studio 4. 1. 2 on Windows 10,使用该 Android 集成开发工具用于开发和调试 Android 应用。PyCharm Professional,使用 python 集成开发工具用于开发 Android 的 HTTP 请求服务端,提供单幅图像检测时的云服务器端响应服务。使用 Flask 框架开发 Linux 服务器上的服务端系统。使用 Pytorch 框架对樱桃检测模型进行训练、验证并部署到服务器端。

4 结论

(1)提出了在线增强与离线数据增强相结合的数据增强策略。与原始数据集相比,离线增强对于精度的提高较大,单独使用离线增强可以使平均精度均值 2 提高 4. 2 个百分点,使用离线增强与在线增强结合的方式可以使模型训练精度达到最高,其

中离线增强与 Mosaic 增强相结合的方式能使平均精度均值 2 提高 5.2 个百分点。

(2)本研究以小目标、易遮挡重叠物体的检测任务为目标,分别从 3 个方面对传统 YOLO v5 模型进行改进,将 YOLO v5 的骨干网络增添具有注意力的 Transformer,这一模块可以形成与卷积的互补,并且注意力机制能够帮助模型更好的检测樱桃这类小目标物;neck 结构由 PaFPN 改成高效的双向跨尺度

和加权特征融合方式 BiFPN,能够有效缓解小目标与目标重叠问题;在 head 结构中增加下采样较小的浅层特征 P2 进行特征输出,浅层特征包含更多的位置信息,因此对小目标物体的位置检测有一定帮助。实验结果表明,单独使用一种改进方法对模型精度提升不太显著,将改进优化方法两两结合能够显著提升精度,其中 YOLO v5 - Transformer - BiFPN 模型的精度最高,较原始模型提高 2.9 个百分点。

参 考 文 献

- [1] 何斌,张亦博,龚健林,等. 基于改进 YOLO v5 的夜间温室番茄果实快速识别[J]. 农业机械学报, 2022, 53(5):201 - 208.
HE Bin,ZHANG Yibo,GONG Jianlin,et al. Fast recognition of tomato fruit in greenhouse at night based on improved YOLO v5 [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):201 - 208. (in Chinese)
- [2] JI W, ZHAO D, CHENG F, et al. Automatic recognition vision system guided for apple harvesting robot[J]. Computers & Electrical Engineering, 2012, 38(5):1186 - 1195.
- [3] KURTULMUS F, LEE W S, VARDAR A. Green citrus detection using 'eigenfruit', color and circular Gabor texture features under natural outdoor conditions[J]. Computers and Electronics in Agriculture, 2011, 78(2):140 - 149.
- [4] RAKUN J, STAJNKO D, ZAZULA D. Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry[J]. Computers & Electronics in Agriculture, 2011, 76(1):80 - 88.
- [5] JUNOS M H, MOHD K A S, THANNIRMALAI S, et al. An optimized YOLO-based object detection model for crop harvesting system[J]. IET Image Processing, 2021, 15(9): 2112 - 2125.
- [6] PAYNE A B, WALSH K B, SUBEDI P P, et al. Estimation of mango crop yield using image analysis-Segmentation method[J]. Computers & Electronics in Agriculture, 2013, 91:57 - 64.
- [7] MOALLEM P, SERAJODDIN A, POURGHASSEM H. Computer vision-based apple grading for golden delicious apples based on surface features[J]. Information Processing in Agriculture, 2017, 4(1): 33 - 40.
- [8] 周云成,许童羽,郑伟,等. 基于深度卷积神经网络的番茄主要器官分类识别方法[J]. 农业工程学报,2017,33(15):219 - 226.
ZHOU Yuncheng, XU Tongyu, ZHENG Wei, et al. Classification and recognition approaches of tomato main organs based on DCNN[J]. Transactions of the CSAE,2017,33(15):219 - 226. (in Chinese)
- [9] WANG C, TANG Y, ZOU X, et al. A robust fruit image segmentation algorithm against varying illumination for vision system of fruit harvesting robot[J]. Optik-International Journal for Light and Electron Optics, 2016, 131:626 - 631.
- [10] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO - V3 model[J]. Computers and Electronics in Agriculture, 2019, 157:417 - 426.
- [11] SENG W C, MIRISAE S H. A new method for fruits recognition system [C] // International Conference on Electrical Engineering and Informatics. IEEE, 2009.
- [12] LINKER R, COHEN O, NAOR A. Determination of the number of green apples in RGB images recorded in orchards[J]. Computers & Electronics in Agriculture, 2012, 81(1):45 - 57.
- [13] KURTULMUS F, LEE W S, VARDAR A. Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network[J]. Precision Agriculture, 2014, 15(1):57 - 79.
- [14] WANG D, HE D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning[J]. Biosystems Engineering, 2021, 210(6):271 - 281.
- [15] ZHAO Y, GONG L, ZHOU B, et al. Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis[J]. Biosystems Engineering, 2016, 148:127 - 137.
- [16] JI W, ZHAO D, CHENG F, et al. Automatic recognition vision system guided for apple harvesting robot[J]. Computers & Electrical Engineering, 2012, 38(5):1186 - 1195.
- [17] 李寒,张漫,高宇,等. 温室绿熟番茄机器视觉检测方法[J]. 农业工程学报,2017,33(增刊):328 - 334.
LI Han, ZHANG Man, GAO Yu, et al. Green ripe tomato detection method based on machine vision in greenhouse [J]. Transactions of the CSAE,2017,33(Supp.):328 - 334. (in Chinese)
- [18] 王卓,王健,王泉雄,等. 基于改进 YOLO v4 的自然环境苹果轻量级检测方法[J]. 农业机械学报,2022,53(8):294 - 302.
WANG Zhuo,WANG Jian,WANG Xiaoxiong,et al. Lightweight real-time apple detection method based on improved YOLO v4 [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(8):294 - 302. (in Chinese)
- [19] 刘小刚,范诚,李加念,等. 基于卷积神经网络的草莓识别方法[J]. 农业机械学报,2020,51(2):237 - 244.
LIU Xiaogang,FAN Cheng,LI Jianian,et al. Identification method of strawberry based on convolutional neural network [J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(2):237 - 244. (in Chinese)
- [20] 张晴晖,孔德肖,李俊毅,等. 基于逆运动学降维求解与 YOLO v4 的果实采摘系统研究[J]. 农业机械学报,2021,52(9):15 - 23.
ZHANG Qinghui,KONG Dexiao,LI Junqiu,et al. Design of fruit picking system based on inverse kinematics dimension reduction and YOLO v4 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(9): 15 - 23. (in Chinese)