

基于改进 YOLOv5 的番茄识别检测算法

丁仁集, 陈丙三

(福建理工大学 机械与汽车工程学院, 福建 福州 350118)

摘要: 为提升果园自动采摘机器人的工作效率, 提高对番茄果实识别的准确率, 提出了一种基于改进 YOLOv5 的番茄识别检测算法。该识别检测算法以 YOLOv5 算法为基础, 改进了原算法中颈部网络的 BottleneckCSP 模块, 通过增加批归一化层与 SiLU 激活函数提升了网络对目标深层语义信息的提取能力; 采用轻量级通用上采样算子 CARAFE 扩大感受野, 减少漏检并保持轻量化; 结合轻量化注意力机制 SimAM 为网络提供三维的注意力权重, 过滤掉冗余信息, 提高模型的准确性和鲁棒性; 用 SIoU 替换 CIoU 损失函数, 有效减少冗余框的同时加快了预测框的收敛和回归。研究表明, 改进后算法对番茄目标检测的平均精度均值达到 96.5%, 比原始算法提高 3.4%, 对小番茄及番茄被遮挡时的漏检率也有效降低, 且满足实时要求。

关键词: 番茄检测; 改进 YOLOv5; SimAM; CARAFE

中图分类号: TP391.4

文献标志码: A

文章编号: 1672-4348(2023)06-0585-07

Tomato recognition and detection algorithm based on improved YOLOv5

DING Renji, CHEN Bingsan

(School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, China)

Abstract: In order to improve the efficiency of automatic picking robots in orchards and the accuracy of tomato fruit recognition, a tomato recognition detection algorithm based on improved YOLOv5 is proposed. This algorithm is based on the YOLOv5 algorithm, and improves the BottleneckCSP module in the neck network, which enhances the network's ability to extract the deep semantic information of the target by adding a batch normalization layer and the SiLU activation function. It adopts the lightweight universal upsampling operator CARAFE to expand the sensory field, reduce leakage detection and maintain lightweight. Combined with the lightweight attention mechanism SimAM, it provides a three-dimensional attention weight for the network, filters out redundant information, and improves the accuracy and robustness of the model. Replacing the CIoU loss function with SIoU effectively reduces the redundant frames while accelerating the convergence and regression of the prediction frames. Experimental results show that the improved tomato identification and detection algorithm achieves 96.5% mean average precision for tomato target detection, which is 3.4% higher than that of the original algorithm, and the omission rate for small tomatoes and tomatoes occluded is also effectively reduced, and the real-time requirements are met.

Keywords: tomato detection; improved YOLOv5; SimAM; CARAFE

随着智慧农业的发展, 机器人自动采摘在番茄产业中得到了推广和应用。比起人工采摘, 机器人自动采摘虽然省时高效, 但对番茄果实的识

别准确性还需要进一步提高。研究目标识别检测算法有助于提高番茄机器人自动采摘效率, 促进番茄产业的智能化发展^[1-2]。

收稿日期: 2023-05-30

基金项目: 福建省高校创新团队发展计划(闽教科[2020]12号); 福建省自然科学基金(2023J01342)

第一作者简介: 丁仁集(1996—), 男, 湖北黄冈人, 硕士研究生, 研究方向: 机器视觉。

通信作者: 陈丙三(1980—), 男, 安徽宿州人, 教授, 博士, 研究方向: 智能材料及工程应用、机器视觉。

基于卷积神经网络的目标识别算法已在农场番茄识别检测中得到广泛应用。张文静^[3]等提出用改进的 Faster-RCNN^[4]模型识别自然环境下的番茄,但其检测效率低下且在复杂环境下不能保证较高的检测精度。成伟^[5]等提出改进的 YOLOv3^[6]模型,用于识别温室中的番茄,平均精度为 95.7%,但对于小番茄仍存在误检和漏检。近年来, YOLO 系列目标检测算法被不断更新, YOLOv5 的设计在精度、召回率和平均精度等方面都达到了最佳性能,是番茄检测的理想模型,但在番茄果实存在遮挡或果实与背景颜色相似情况下,检测效果仍不理想。

针对上述问题,本研究提出一种基于改进 YOLOv5 的番茄识别检测算法,优化了 YOLOv5 原始算法的主要模块 BottleneckCSP,通过增加批归一化层与 SILU 激活函数提升网络对目标深层语义信息的提取能力;使用轻量级通用上采样 CARAFE 算子增加感受野;引入无参注意力机制 SimAM 增强有效特征,抑制干扰特征;使用 SIoU Loss 作为模型的损失函数优化回归过程,提升模型检测性能。

1 YOLOv5 算法

YOLOv5 是一种单阶段目标检测算法,它汲取了许多优秀网络结构的优点,确保高检测精度的同时还能维持较高的检测速度,能够做到实时检测目标。YOLOv5 由输入端、主干网络(Backbone)、颈部网络(Neck)和输出检测层(Head)4 部分组成,其结构如图 1 所示。

2 改进 YOLOv5 算法

2.1 优化 BottleneckCSP 模块

BottleneckCSP 模块是 YOLOv5 的重要组成部分,用于构建骨干网络。其中 Bottleneck 模块(如图 2 所示)是 BottleneckCSP 的主要构成,该模块采用的是残差结构,即将卷积核尺寸为 1×1 的卷积层与 3×3 的卷积层相连,并为每一个卷积层配置了一个 BN 层和一个激活函数。再将该部分的输出与输入相加,形成残差结构,作为 Bottleneck 模块的输出。

BottleneckCSP 模块(如图 3 所示)的主要作用是将输入特征图分别送入两个分支,并对这

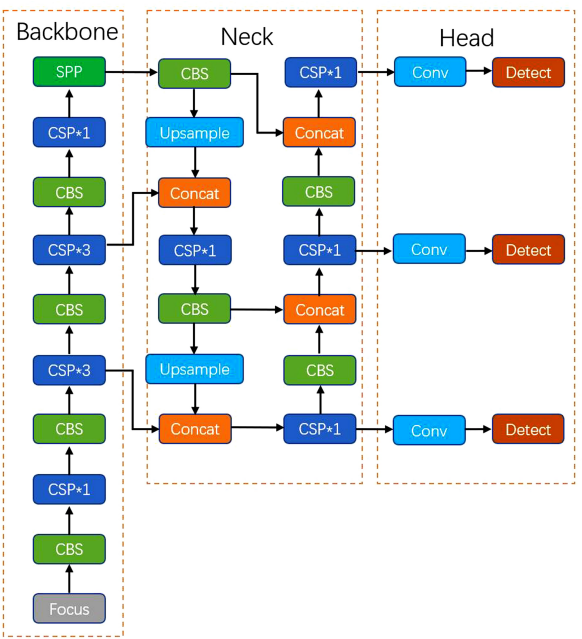


图 1 YOLOv5 结构

Fig.1 Structure of YOLOv5

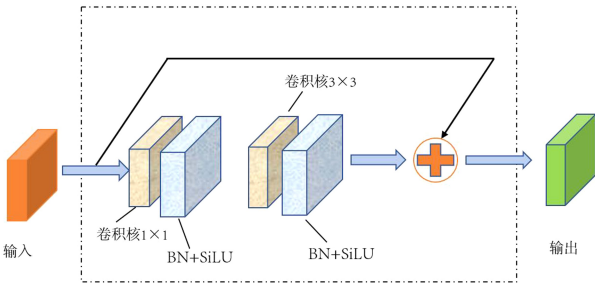


图 2 Bottleneck 结构

Fig.2 Structure of Bottleneck

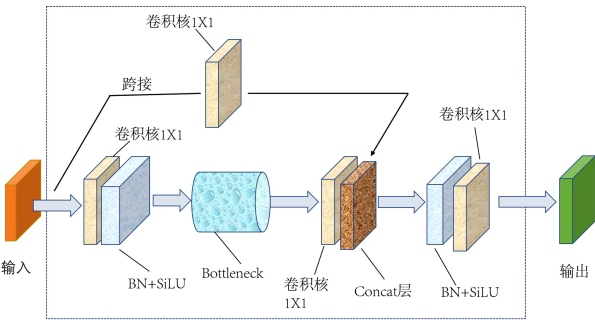


图 3 BottleneckCSP 结构

Fig.3 Structure of BottleneckCSP

两个分支进行卷积操作,使其特征图的通道数减半。其中一个分支先与 Bottleneck 模块相连进行数据降维操作,再通过常规卷积进行升维;然后经过 Conv 层,使用 Concat 操作将两个分支的输出特征图深度连接;最后通过 BN 层和 Conv

层依次处理,得到 BottleneckCSP 模块的输出特征图,其尺寸与输入 BottleneckCSP 模块特征图尺寸相同。

本研究在 BottleneckCSP 模块的基础上进行改进,在其上层的卷积层后面添加了一个批归一化层和一个 SiLU 激活函数,形成了卷积模块 CBS。通过加入批归一化层,既提升了算法的特征提取能力,又确保了算法的稳定性。引入 SiLU 激活函数可以增加算法的非线性因素,将含有更多目标位置与细节信息的低层特征图、含有更强语义信息的高层特征图进行融合,使得改进后的 BottleneckCSP 模块的输出特征图可以包含更多细节信息和更强的语义信息,从而提升了 BottleneckCSP 模块的特征提取能力。改进后的 BottleneckCSP 模块如图 4 所示,命名为 BottleneckCSP_D 模块。

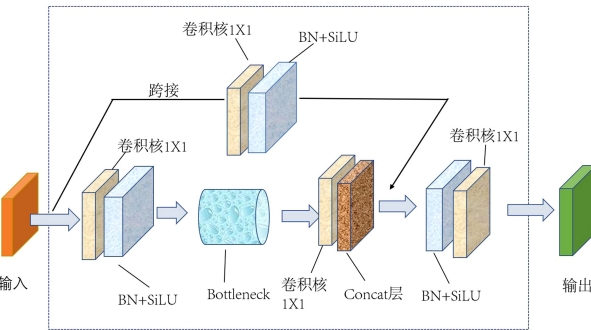


图 4 BottleneckCSP_D 结构图

Fig.4 Structure of BottleneckCSP_D

2.2 CARAFE 轻量级通用上采样算子

YOLOv5 原始算法默认使用最近邻插值,仅根据像素点的空间位置来确定上采样核,并未充分利用特征图的语义信息,可以将其看作是一种“均匀”的上采样方式,但感知域通常较小。为了解决该问题,本研究的番茄识别检测算法引入了轻量高效的上采样算子 CARAFE^[7],通过少量参数和计算保持轻量级功能并扩大感受野。CARAFE 由上采样预测模块和特征重组两个模块构成,如图 5 所示。

2.3 SimAM 注意力机制

当待检测目标和背景环境相似时,在神经网络中添加注意力机制可以有效地剔除不相关信息的干扰。改进的番茄识别检测算法使用了三维的无参数 SimAM^[8]注意力机制改进特征融合网络。

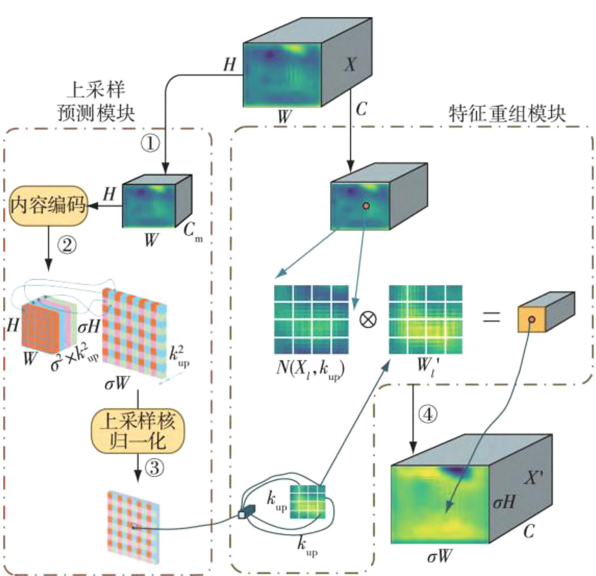


图 5 CARAFE 上采样结构

Fig.5 Up-sampling structure of CARAFE

与一维、二维的注意力机制相比,三维注意力机制在没有引入参数的情况下就能够更加全面高效地平衡特征的重要性,从而增强目标区域的特征权重并降低背景特征的权重。SimAM 注意力机制模块如图 6 所示。

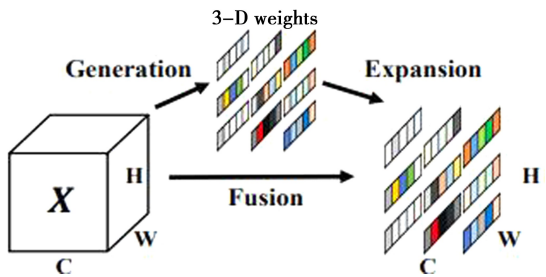


图 6 SimAM 注意力机制结构图

Fig.6 Structure of SimAM attention mechanism

SimAM 注意力机制的计算过程如式(1)所示。

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{e_i^*}\right) \odot X \tag{1}$$

式中, X 为输入特征, $\tilde{X}$ 为输出特征, e_i^* 则为每个通道上的能量函数,其计算过程如式(2)所示。

$$e_i^* = \frac{4(\hat{\sigma}^2 + \hat{\lambda})}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\hat{\lambda}} \tag{2}$$

式中, $\hat{\lambda}$ 为超参数; t 为输入特征值; $\hat{\mu}$ 和 $\hat{\sigma}^2$ 分别代表 X 中每个通道上的平均值和方差,其计算公式如式(3)和式(4)所示。

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M X_i \quad (3)$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (X_i - \hat{\mu})^2 \quad (4)$$

式中, X 为输入特征, M 为输入特征图的尺寸。通过对神经元进行操作, 赋予带有关键信息的神经元更高的权重, 提高网络的识别和定位精度, 而不会增加网络的额外参数。本研究将注意力机制添加在检测头前, 如图 7。

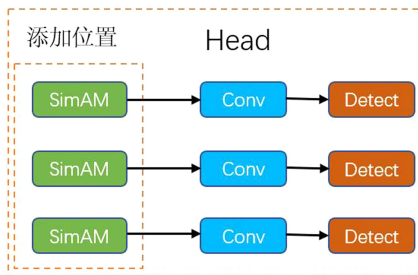


图 7 注意力机制添加区域

Fig.7 Addition area of attention mechanism

2.4 SIoU 损失函数

SIoU^[9] 损失函数在 CIoU 损失函数的基础上, 定义了预测框向真实框靠近的方向, 避免了预测框的游荡现象, 从而可以进一步提高模型的回归精度。因此, 改进的番茄识别检测算法选择 SIoU 损失函数来计算回归预测损失。

SIoU_Loss 损失函数由 4 个 Cost 函数组成: Angle cost、Distance cost、Shape cost 和 IoU cost, 包含的数学定义为式 (5) ~ 式 (11)。Angle cost 添加了 LF 角度感知组件, 模型首先将预测框带到 X 或 Y 轴, 然后沿着相对应的轴继续接近。为实现这一点, 引入和定义 LF 组件如下:

$$\Lambda = 1 - 2\sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \quad (5)$$

$$x = \frac{c_h}{\sigma} = \sin(\alpha) \quad (6)$$

Distance cost 重新定义为

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho^t}) \quad (7)$$

Shape cost 的定义为

$$\Omega = \sum_{t=w,h} (1 - e^{-\eta t})^\theta \quad (8)$$

其中, θ 的值定义为每个数据集的 Shape cost, 并且其值是唯一的, 它控制着对 Shape cost 的关注程度。

IoU cost 的定义为

$$I = \frac{|B \cap B^T|}{|B \cup B^T|} \quad (9)$$

其中, B 表示真实框, B^T 表示预测框。回归损失函数 SIoU 为

$$L_{\text{box}} = 1 - I + \frac{\Delta + \Omega}{2} \quad (10)$$

总损失函数为

$$L = W_{\text{box}} L_{\text{box}} + W_{\text{cls}} L_{\text{cls}} \quad (11)$$

其中, L_{cls} 是 focal loss, W_{box} 和 W_{cls} 分别是框和分类损失权重。

3 试验及结果分析

3.1 数据集

研究采用的数据集为 Kaggle 公开的数据集, 由 895 张带有标注的番茄图像组成, 同时实地采集 400 张番茄图像标注之后与公共数据集混合。为扩大数据集数量, 分别对图像进行随机翻转、改变亮度、Mosaic 增强和 Mixup 增强等操作, 加上原始数据后一共有 3 980 张图片。对增强后的数据集按照 8:1:1 的比例划分为训练集、验证集、测试集, 数据集图像, 如图 8 所示。



图 8 番茄数据集

Fig.8 Tomato dataset

3.2 实验环境与参数设定

实验运行环境如表 1 所示。训练时, 使用余弦退火算法调整学习率, 采用 SGD 优化器进行参数更新, 参数设置如表 2 所示。

表 1 实验环境配置

Tab.1 Experimental environment configuration	
名称	参数
GPU	RTX3060Ti,显存 8G
CPU	7 核 Intel Xeon CPU E5-2680 v4@ 2.4GHz
操作系统	Windows10
编程语言	Python3.8.1
深度学习框架	PyTorch1.9.0+CUDA 1.11.0
编译软件	PyCharm

表 2 参数设置

Tab.2 Parameter settings	
参数	参数值
学习率	0.01
动量	0.937
权重衰减系数	0.0005
批量大小	16
训练轮数	100

3.3 评价指标

采用的算法评价指标有精确率 P 、召回率 R 、平均精度 AP 和平均精度均值 mAP 等指标衡量算法性能,计算公式如下:

$$P=\frac{TP}{TP+FN} \tag{12}$$

$$R=\frac{TP}{TP+FP} \tag{13}$$

$$AP=\int_0^1 p(r) \mathrm{d} r \tag{14}$$

$$\mathrm{mAP}=\frac{\mathrm{AP}_1+\mathrm{AP}_2+\mathrm{AP}_3+\cdots+\mathrm{AP}_N}{N} \tag{15}$$

其中,TP 表示正确检测出的番茄目标,FP 表示错误检测的番茄目标,FN 表示实际存在但未被检测的番茄目标, N 表示标签类别数。

3.4 训练结果

在相同的配置环境下,将原始算法和改进后的算法分别进行了 100 轮训练。平均精度均值曲线的对比如图 9 所示,其中虚线代表原始的 YOLOv5 算法,实线代表改进后的 YOLOv5 算法,图 9 中横坐标表示训练次数,总共为 100 轮,纵坐标表示平均精度均值,从图 9 可以看出,在前 20 轮训练时,算法的收敛速度非常快,经过 30 轮训练迭代后,两

个算法曲线都趋于水平,改进后算法平均精度均值明显提高,表明算法改进方案是有效的。

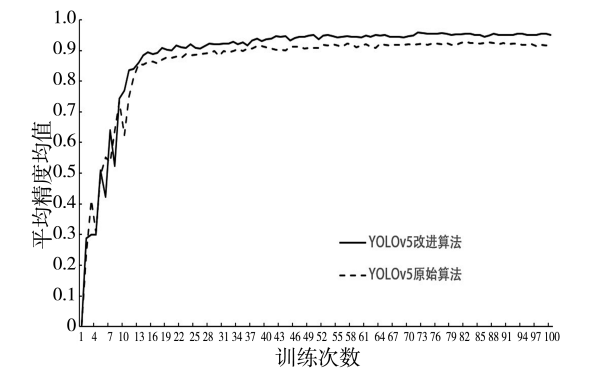


图 9 YOLOv5 改进前后对比曲线
Fig.9 Comparison curve of YOLOv5 before and after improvement

3.5 注意力机制对比试验

为了验证不同的注意力机制对 YOLOv5 算法的性能的影响,对比 SE、CBAM、CA 和 SimAM4 种注意力机制,结果如表 3 所示。从表 3 可见,除了 CA 之外,其它 3 种算法的平均精度均值相比原始 YOLOv5 算法都有不同程度的提升。其中,SE 的精确度提升较大,但召回率提升不够明显;CBAM 和 SimAM 的精确率、召回率和平均精度均值 3 个指标都有一定提升,但 SimAM 的提升更为显著,分别比 YOLOv5 原始算法提高了 1.9%、2.1%和 1.7%,综合检测性能更好。因此,本研究选择 SimAM 模块进行识别检测。

表 3 不同注意力机制的对比

Tab.3 Comparison of different attention mechanisms			
注意力机制	$P/\%$	$R/\%$	mAP/%
YOLOv5	90.7	85.3	93.1
YOLOv5+SE	92.9	85.5	94.1
YOLOv5+CBAM	91.4	86.5	93.9
YOLOv5+CA	89.5	84.3	92.4
YOLOv5+SimAM	92.6	87.4	94.8

3.6 消融试验

为了验证 4 种改进方案的检测效果,本研究在番茄数据集上进行了消融试验和对比分析。为了确保试验的严谨性,在试验过程中使用了相同的参数配置,结果如表 4 所示。相较于原始 YOLOv5 算法,4 种改进算法的性能都有不同程度

的提升。

表 4 消融试验结果
Tab.4 Ablation test results

编号	CSP_D	CAR FE	SimAM	SIoU	mAP/%
0	×	×	×	×	93.1
1	√	×	×	×	94.3
2	×	√	×	×	95.1
3	×	×	√	×	94.8
4	×	×	×	√	95.4
5	√	√	√	√	96.5

注：“X”表示使用该模块；“√”表示使用该模块。

3.7 主流目标检测模型对比

为了进一步评估改进算法对番茄的检测性能,将改进后的算法与 SSD^[10]、Faster-RCNN、YOLOv4-tiny 以及 YOLOv5 目标检测算法进行了比较。从表 5 可以看出,改进后的 YOLOv5 算法的平均精度均值和召回率都优于其他算法,相较于原始 YOLOv5,平均精度均值提高了 3.4%,达到 96.5%;召回率提高了 3.8%,达到 89.1%。由于召回率的提高,改进后的算法会检测到更多的番茄目标,帧率有所下降,但仍快于 SSD、Faster-RCNN

和 YOLOv4-tiny,满足现实番茄检测的实时性要求。

表 5 主流算法性能对比

Tab.5 Performance comparison of mainstream algorithms

模型	mAP/%	R/%	帧/f · s ⁻¹
SSD	75.7	72.6	18
Faster-RCNN	81.6	87.5	9
YOLOv4-tiny	88.4	84.1	52
YOLOv5	93.1	85.3	90
改进后 YOLOv5	96.5	89.1	82

3.8 应用检测评价

为了进一步证明改进 5 算法对番茄检测的有效性,对 YOLOv5 原始算法和改进算法在实际种植场景进行可视化比较。选择了番茄果园中的枝叶遮挡、果实遮挡、背景相似和远距离的 4 种复杂场景。如图 10 所示,原始 YOLOv5 算法在 4 种复杂场景下存在不同程度的误检和漏检问题,且检测置信度偏低。改进后的算法有效降低了原始 YOLOv5 检测过程中存在的误检、漏检问题,并且提高了置信度,如图 11 所示。综上,改进的算法在实际番茄种植场景中具有更好的检测性能。

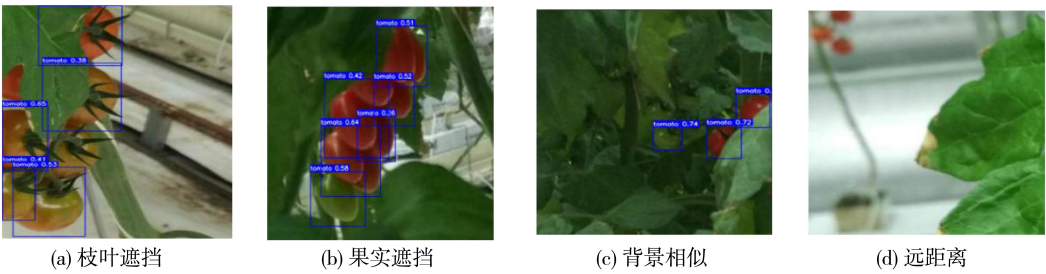


图 10 YOLOv5 改进前的效果

Fig.10 Effects before improvement of YOLOv5

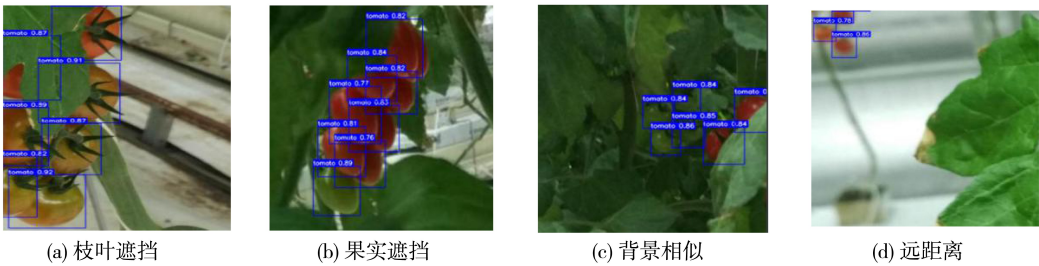


图 11 YOLOv5 改进后的效果

Fig.11 Effects after improvement of YOLOv5

4 结论

本研究针对当前番茄机器人检测精度较差以及对遮挡目标存在漏检的问题,提出一种基于改进YOLOv5算法的番茄目标识别检测算法。算法通过设计BottleneckCSP_D模块替换BottleneckCSP模块,扩大感受野,增强模型的特征提取能力;引入三维无参数SimAM注意力机制,抑制了背景中无关信息的干扰,保持了检测速度的同时

提高了检测精度;更换上采样为CARAFE算子,增强特征融合准确度,充分融合上下文语义信息;改进损失函数,优化边界框回归过程,提升预测锚框准确率。结果表明,相比原始算法,改进后的YOLOv5的mAP提高了3.4%,置信度更高,对遮挡环境和背景相似环境中番茄目标的检测性能也得到了提高,其检测精度能够满足农场番茄实际采摘的需求。该算法可为番茄自动采摘提供了技术支持。

参考文献:

- [1] 黄彤斌,黄河清,李震,等. 基于YOLOv5改进模型的柑橘果实识别方法[J]. 华中农业大学学报,2022,41(4):170-177.
- [2] 王海楠,弋景刚,张秀花. 番茄采摘机器人识别与定位技术研究进展[J]. 中国农机化学报,2020,41(5):188-196.
- [3] 张文静,赵性祥,丁睿柔,等. 基于Faster R-CNN算法的番茄识别检测方法[J]. 山东农业大学学报(自然科学版),2021,52(4):624-630.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] 成伟,张文爱,冯青春,等. 基于改进YOLOv3的温室番茄果实识别估产方法[J]. 中国农机化学报,2021,42(4):176-182. DOI: 10.13733/j.jcam.issn.2095-5553.2021.04.25.
- [6] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2018-04-08]. <https://arxiv.org/abs/1804.02767>.
- [7] WANG J Q, CHEN K, XU R, et al. CARAFE: content-aware ReAssembly of FEatures[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 3007-3016.
- [8] YANG L, ZHANG R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International Conference on Machine Learning, 2021: 11863-11874.
- [9] GEORGIAN Z. SIoU loss: More powerful learning for bounding box regression[J]. Springer International Publishing, 2016: 21-37.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

(责任编辑:方素华)