

基于改进型 YOLO 的复杂环境下番茄果实快速识别方法

刘芳¹ 刘玉坤^{1,2} 林森² 郭文忠² 徐凡² 张白¹

(1. 北方民族大学电气信息工程学院, 银川 750021; 2. 北京农业智能装备技术研究中心, 北京 100097)

摘要: 为实现温室环境下农业采摘机器人对番茄果实的快速、精确识别, 提出了一种改进型多尺度 YOLO 算法 (IMS-YOLO)。对 YOLO 网络模型进行筛选和改进, 设计了一种含有残差模块的 darknet-20 主干网络, 同时融合多尺度检测模块, 构建了一种复杂环境下番茄果实快速识别网络模型。该网络模型层数较少, 能够提取更多特征信息, 且采用多尺度检测结构, 同时返回番茄果实的类别和预测框, 以此提升番茄果实检测速度和精度。采用自制的番茄数据集对 IMS-YOLO 模型进行测试, 并分别对改进前后网络的检测性能以及主干网络层数对特征提取能力的影响进行了对比试验。试验结果表明, IMS-YOLO 模型对番茄图像的检测精度为 97.13%, 准确率为 96.36%, 召回率为 96.03%, 交并比为 83.32%, 检测时间为 7.719 ms; 对比 YOLO v2 和 YOLO v3 等网络模型, IMS-YOLO 模型可以同时满足番茄果实检测的精度和速度要求。最后, 通过番茄温室大棚采摘试验验证了本文模型的可行性和准确性。

关键词: 番茄; 采摘机器人; 识别; 深度学习; 改进型 YOLO; darknet-20

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2020)06-0229-09

OSID:



Fast Recognition Method for Tomatoes under Complex Environments Based on Improved YOLO

LIU Fang¹ LIU Yukun^{1,2} LIN Sen² GUO Wenzhong² XU Fan² ZHANG Bai¹

(1. College of Electrical Information Engineering, North Minzu University, Yinchuan 750021, China
2. Beijing Agricultural Intelligent Equipment Technology Research Center, Beijing 100097, China)

Abstract: In order to implement the fast and accurate recognition of tomatoes for agricultural harvesting robots under greenhouse environments, an improved multi-scale YOLO detection algorithm named IMS-YOLO was presented. A new backbone network structure, which was named darknet-20, with one residual block was designed based on a series of the previous YOLO algorithms, and a multi-scale detection structure was utilized simultaneously for the detection algorithm. Therefore, a new kind of neural network model was formed for the fast recognition of tomatoes under complex environments. Due to some features of the method such as the fewer layers required, the larger amount of information extracted, and by using the multi-scale structure to return both the detection categories and the bounding boxes, the detection speed and accuracy were improved. IMS-YOLO model was tested on our own tomato dataset, and the detection performance of the network before and after the improvement as well as the influence of the variation of the backbone network layers on the feature extraction capacity were analyzed respectively. The test results showed that the proposed method had ideal features with a precision of tomato image detection of 97.13%, an accuracy of 96.36%, a recall rate of 96.03%, an intersection over union (I_{ou}) of 83.32% and a detection time of 7.719 ms. Furthermore, compared with YOLO v2, YOLO v3 and some other neural networks mentioned, IMS-YOLO can meet the requirements of both detection accuracy and speed. At last, the feasibility of the proposed algorithm applying to the robots was verified by the harvesting tests of the ripe tomatoes under the greenhouse environments.

Key words: tomato; picking robot; recognition; deep learning; improved YOLO; darknet-20

收稿日期: 2019-11-12 修回日期: 2020-02-06

基金项目: 国家自然科学基金项目(31601794)、宁夏回族自治区重点研发计划项目(2018BBF02024)和宁夏回族自治区重点研发计划重大科技项目(2017BY067)

作者简介: 刘芳(1970—),女,教授,博士生导师,主要从事智能控制研究,E-mail: fangliu214@163.com

0 引言

我国是世界最大的番茄生产和消费国家,番茄生产是农民增收致富和出口创汇的重要途径之一^[1]。番茄采摘机器人对减少生产成本、降低劳动强度、提高作业效率具有重要意义^[2]。采摘机器人的设计难点是番茄的识别和定位,其准确性关系到采摘机器人的工作效率。番茄果实的生长姿态各异,果实之间重叠,叶片、枝干、果柄等对果实遮挡严重,光照环境复杂,这些因素给采摘机器人识别带来了一定的困难。因此,对温室复杂环境下的番茄果实快速、精确识别是研发番茄采摘机器人亟待解决的关键问题^[3]。

温室环境下番茄果实的传统识别方法是基于颜色和形状特征进行信息提取分析^[4-6]。传统目标识别流程包括基于穷举策略的区域选择、基于尺度不变特征变换(SIFT)和方向梯度直方图(HOG)等方法的特征提取、基于支持向量机(SVM)和自适应提升(Adaboost)等方法的分类器分类等。郑小东等^[7]利用颜色特征差异提取红色成熟番茄果实和绿色枝叶背景信息,通过阈值分割的方法识别红色成熟番茄果实,该方法对图像质量要求很高,对噪声影响较为敏感。王玉翰^[8]提取颜色、形状、纹理等5个特征,利用SVM分类器对番茄果实进行识别,该方法有较好的分类识别能力,但分类耗时过长,不能满足实时性需求。梁喜凤等^[9]提出一种番茄果实串采摘点识别方法,将整个番茄果实串视为一个整体,提取其形状特征,该方法对垂直向下的番茄果实串采摘点识别效果较好,但现实中番茄果实生长形态各异,该方法不能对其他姿态的番茄果实进行识别。赵源深等^[10]提出一种基于非颜色编码的番茄果实识别算法,通过基于Haar-like特征的阈值判断获得若干弱分类器,利用AdaBoost算法将多个弱分类器通过学习训练得到一个强分类器,用于识别红色成熟番茄果实。在光照条件较差和果实遮挡较为严重的情况下,该方法对成熟番茄果实的识别率为93.3%,识别精度偏低。冯青春等^[11]在2R-G-B色差模型下提取红色成熟番茄果实颜色特征,利用动态阈值分割的方法识别红色成熟番茄果实,该方法识别耗时较长,且未考虑叶片遮挡等复杂环境下番茄果实的识别。综上所述,采用传统方法进行番茄果实识别无法达到较好的精度和实时性要求。此外,上述研究大多没有考虑温室复杂环境下的影响因素,对多样的特征变化鲁棒性不足,因此难以满足实际要求。

近年来兴起的基于深度学习的深度卷积神经网络

(Deep convolutional neural network, DCNN)为番茄果实识别提供了新的思路。DCNN在图像目标检测上具有较大的优势,其检测方法可归纳为两类:一是基于区域生成的检测方法,先由算法生成一系列候选框,再对候选框中的目标进行分类,R-CNN^[12]、Fast R-CNN^[13]、Faster R-CNN^[14]等网络模型都属于此类检测方法。这类方法识别错误率低、漏识别率也较低,但速度较慢,难以满足实时检测场景需求。二是基于回归的方法,该类方法在实现目标定位的同时预测目标分类,YOLO^[15-17]系列网络属于此检测方法。该类方法识别速度快,可以达到实时性要求,而且准确率也基本能达到第1类方法的水平。目前,DCNN在农业设施智能化研究方面备受关注^[18-22]。研究表明^[23],第1类目标检测算法检测精度较高,但占用大量计算资源,导致检测时间较长,影响采摘机器人工作效率;第2类目标检测算法兼顾了检测精度和速度,但其网络结构设计过于精简,导致识别精度低于第1类算法。

实际温室环境下番茄果实重叠、遮挡严重,稀疏程度不同、大小不一,这些成为快速、精确识别番茄果实的难点。本文以DCNN中检测速度较快的YOLO网络模型为基础,根据温室环境下番茄果实的特征改进主干网络得到一种新型主干网络darknet-20,以提升检测速度;采用YOLO v3-tiny快速精确的检测结构,融合多尺度信息,以提升检测精度,构建一种温室复杂环境下番茄果实识别的网络模型。

1 数据样本采集与预处理

番茄果实图像数据采集于北京市农林科学院连栋温室,采集设备为FM810-IX-A 3D型相机(图漾科技),其二维RGB图像分辨率为640像素×480像素,像素级精度为2 mm。为降低由训练样本多样性不足导致网络模型过拟合的概率,采集时分远景和近景,远景、近景分别距离果实500~600 mm和200~300 mm,远景、近景情况下各采集左、中、右3个角度的图像,每个角度采集3幅图像,共1 800幅图像。为了增加样本的多样性,图像样本中包含成熟红色番茄果实和未成熟绿色番茄果实,包含果实数量、稠密程度、遮挡程度不同的情况,还包含顺光、逆光等光照情况。为获取准确的数据参数,还需要对图像进行人工标注,标注时将番茄的最小外接矩形框作为真实框(Ground truth),以减少真实框内的背景像素。表1为各类番茄图像数量,图1为一组典型复杂环境下的番茄果实图像。

表 1 各类番茄图像数量

Tab.1 Number of tomato images under varies conditions

参数	顺光				逆光			
	单果		多果		单果		多果	
	遮挡	未遮挡	遮挡	未遮挡	遮挡	未遮挡	遮挡	未遮挡
图像数量/幅	153	72	840	319	43	23	251	99



图 1 复杂环境下的番茄果实图像

Fig. 1 Tomato images under complex environments in greenhouse

温室环境下光照情况复杂,在光照很强或很弱时采集到的番茄果实图像颜色差异很大。此外,番茄果实的生长姿态各异、重叠遮挡严重,导致果实形状特征难以完整提取。使用传统方法难以有效提取温室番茄果实的特征。因此,在番茄果实识别时需要对采集样本进行预处理^[24-25]。本文采用自适应直方图均衡化方法减小光照强度对图像质量的影响;通过缩放、水平/垂直翻转或旋转正负 90°等方法,解决番茄果实生长姿态不同的问题,改善样本集数据不均衡。通过以上方法,最终扩增样本集图像至 15 120 幅,其中训练集 70%、验证集 10%、测试集 20%。

2 复杂环境下的番茄果实快速识别方法

YOLO (You only look once) 网络模型采用 darknet 深度学习框架实现输入图像端到端的训练,其突出特点是快速性。与 Faster R-CNN 使用候选区域(Region proposal)特征提取方式不同,YOLO 对输入图像的全局区域进行训练,可加快训练速度且能更好地区分目标和背景。利用主干网络完成番茄

果实特征提取之后,采用预测框(Bounding box) 预测的方式,同时预测出目标类别和预测框。

2.1 YOLO 系列网络模型

YOLO 系列网络模型中,YOLO v1 存在网络模型检测精度不高、目标定位不准确等问题^[15];YOLO v2 中加入了锚点框、批量归一化、高分辨率分类器、更改网络模型结构等操作来提升网络模型性能,尤其是检测速度突出,但不适用于检测目标重叠的情况^[16];YOLO v3 中引入了多尺度融合训练、残差结构、改变网络模型结构、锚点框选择机制、分类方法等操作,使得网络模型性能得到了极大提升,但其主干网络深度达 53 层且采用多尺度融合,导致检测速度不高,番茄果实检测的实时性无法得到保证^[17]。因此,本文在剖析 YOLO v2、YOLO v3 网络模型特性的基础上对其进行重组,构建一种新型网络模型结构,解决复杂环境下番茄果实检测识别问题。

2.2 特征快速提取网络结构

在使用 DCNN 提取特征时,增加网络深度意味着增加检测时间。本文只需区分红、绿两类番茄果实,其特征明显、易于提取。由图 2a 可以看出,YOLO v2 主干网络 darknet-19 由 19 层卷积层和 5 层最大池化层交替组成。其结构简单、占用计算资源较少,但特征信息在传递时逐层丢失,导致检测精度下降。本文以 darknet-19 为基础,借鉴 YOLO v3 网络结构,在 darknet-19 主干网络中引入残差结构,提出 darknet-20 主干网络如图 2b 所示,实现前后层特征复用和融合,同时满足番茄果实特征提取的快速性和精确性。该网络的构建思想如下:去掉 darknet-19 中的 5 层最大池化层,用第 2、3、6、9、14 层卷积层进行 2 倍下采样代替原最大池化层的下采样,从而缩小特征图尺度以提取更多特征信息传递给后续卷积层。去掉 darknet-19 最后一层卷积层以便与检测层相连,保留其余各卷积层;在 darknet-19 的第 2 层卷积层后添加两层卷积层以构建一个残差模块,由此构成由 20 层卷积层顺序连接的主干网络。借鉴残差思想构建残差模块。通过恒等映射、跳跃连接其前层卷积输出特征信息,将其直接传递给后面某层输出,如图 2 中蓝线箭头所指,以进行信息融合。需要指出的是,在第 4、5 次下采样单元后各有 2 个残差模块,其作用是在相应尺度下融合更丰富的特征信息输出给检测层。

综上,darknet-20 网络由 20 层卷积层构成,其中包含了首层卷积、5 个下采样卷积层和 7 个残差块,每个残差块由 2 层卷积层组成。相较于 darknet-19 主干网络的 19 层卷积层,darknet-20 只多了一层卷积层,网络深度变化不大,对检测速度的影响不大;使

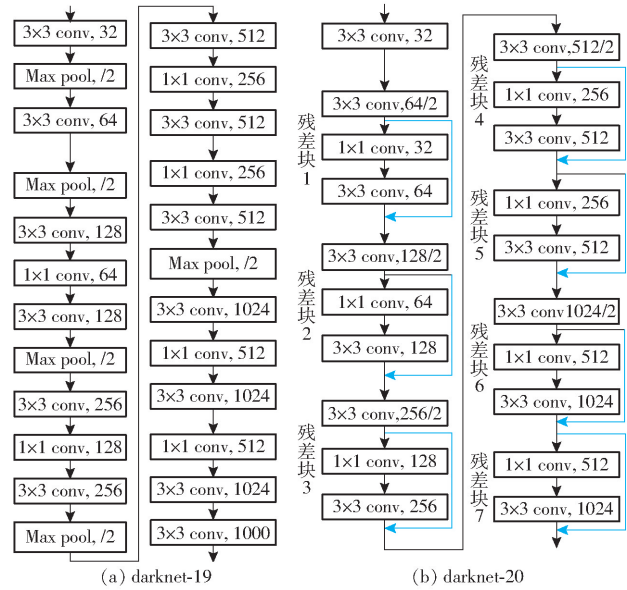


图2 darknet-19 及 darknet-20 主干网络结构
Fig.2 Backbone structure diagrams of darknet-19 and darknet-20

用卷积代替最大池化层进行下采样,能够保留更多的特征信息;在特征提取过程中,引入残差结构可以更好地保留番茄特征信息、促进网络前后层的特征信息融合,在不增加冗余参数和计算量的前提下可以更好地提取番茄果实特征,进一步提高检测精度。

2.3 多尺度检测模型结构

为了得到精准的番茄果实位置和类别信息,本文采用多尺度检测模型结构同时进行预测,其结构如图3所示。输入图像尺寸为416像素×416像素,经由darknet-20主干网络进行特征提取时,主干网络依次完成5次下采样,称为 2^m ($m=1,2,\cdots,5$)倍下采样。每经历一次下采样,特征图将缩小为原输入图像尺寸的1/2,最终在32倍下采样后生成尺寸为13像素×13像素的特征图。对该图像进行一系列的卷积操作之后,再进行一次2倍上采样,生成26像素×26像素的特征图,将其与主干网络下的16倍下采样生成的26像素×26像素的特征图进行张量拼接(Concat),融合两个图像的特征信息,生成一个同等尺寸且含有更多特征信息的特征图,以此提升小目标的检测精度。针对上述过程中32倍下采样生成的13像素×13像素特征图及特征融合后的26像素×26像素特征图,分别进行独立检测,生成尺寸不变且含有番茄果实预测框中心坐标(x,y)、宽 w 、高 h 等4个位置信息,1个置信度和番茄果实颜色类别 c 的特征图,从而得到番茄果实的预测框和颜色类别。采用多尺度检测结构,能够更好地提取不同光照条件及果实重叠、遮挡严重等温室复杂环境下的番茄果实特征,从而可以极大地提升检测精度,为采摘机器人的高效采

摘奠定基础。

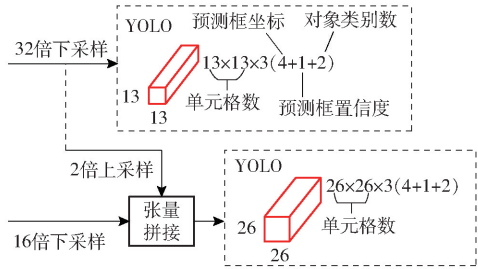


图3 番茄果实多尺度检测结构
Fig.3 Multi-scale detection structure for tomato

本文借鉴YOLO v3中的检测方式生成番茄果实预测框,图4为番茄果实预测框生成流程图。由darknet-20主干网络提取的特征图包含真实框的信息,如图4a绿色框所示;特征图通过回归预测得到若干预测框,如图4b黄色框所示;通过非极大值抑制(Non-maximum suppression, NMS)算法筛选出番茄果实预测框,如图4c紫色框所示。在使用NMS方法时,根据上述多尺度检测模型获得2个类别的置信度,取置信度较高的类别作为预测框中对象所属的类别,并将置信度低于置信度阈值 $a=0.5$ 的所有预测框删除;然后将剩余的所有预测框按照置信度从高到低排序,将置信度最高的预测框作为要保留类别的第1个预测框;再按照由大到小的顺序计算其他预测框与第1个预测框的交并比 I'_{ou} ,删除 I'_{ou} 大于其阈值 $b=0.5$ 的预测框,即完成第1次迭代;接着从剩下的预测框中取置信度最高的预测框作为保留的第2个预测框,进行第2次迭代。通过不断迭代,输出最后的番茄果实预测框。通过以上方法得到番茄果实预测框,即可得到番茄果实在图像中的准确位置。

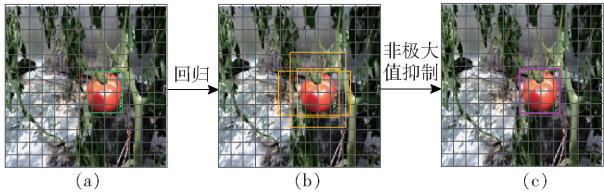


图4 番茄果实预测框生成流程图
Fig.4 Flow chart of generating bounding box for prediction tomatoes

2.4 番茄果实识别模型架构

本文构建的基于darknet-20主干网络的复杂环境下番茄果实快速识别模型架构如图5所示。该架构主要包含2个模块:数据预处理模块和番茄果实快速检测网络模块。前者是在数据样本采集后进行的;后者以darknet-20为主干网络,通过融合多尺度检测模块构建,称为IMS-YOLO (Improved multi-scale YOLO)检测网络。该网络从训练集和验证集

中快速提取复杂环境下的番茄果实相应特征,并融合多尺度特征信息,同时得到番茄果实预测框和类别,从而快速精确地识别温室复杂环境下番茄果实。其中训练集用于拟合检测网络,验证集用于调整检测网络的超参数以及对网络性能进行初步评估。对检测网络进行样本训练的目的在于利用从高分辨率的

番茄果实原始图像中学习到的特征来识别低分辨率的实时采集图像。番茄果实识别架构的工作流程如下:首先将采集的原始图像进行预处理并将预处理后的图像输入检测网络进行训练,然后根据预测框、置信度和类别的损失函数不断调整训练参数,最终得到识别模型。

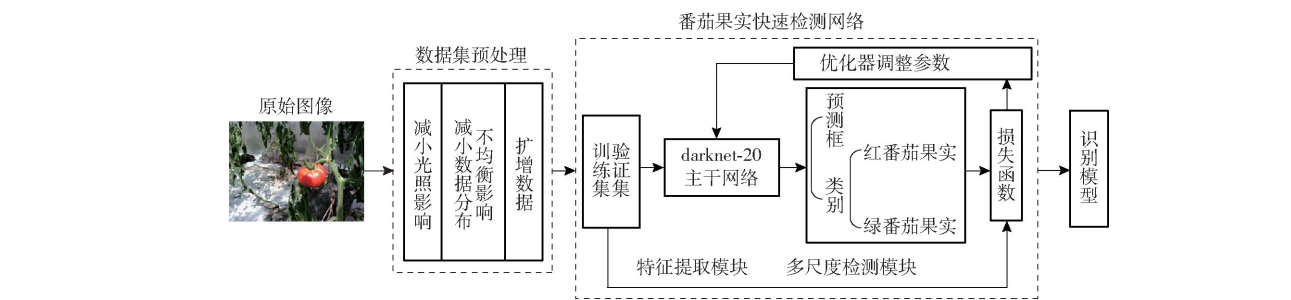


图5 复杂环境下番茄果实快速识别模型的架构

Fig.5 Framework of fast recognition model for tomatoes under complex environment

3 检测网络训练与结果分析

3.1 番茄果实检测网络训练

为了得到复杂环境下番茄果实快速精确的检测网络,需要优化损失函数,使预测框、置信度、类别的训练误差达到平衡。YOLO v3 的损失函数^[17]为

$$L = k_{xy} \delta(x,y) + k_{wh} \delta(w,h) + k_{conf} \delta(c_{conf}) + k_c \delta(c) \tag{1}$$

式中 $\delta(x,y)$ ——预测边界框中心坐标 (x,y) 的误差函数

$\delta(w,h)$ ——预测边界框宽 w 、高 h 的误差函数

$\delta(c_{conf})$ ——预测置信度 c_{conf} 的误差函数

$\delta(c)$ ——类别的误差函数

$k_{xy}、k_{wh}、k_{conf}、k_c$ ——误差权重系数

训练误差曲线可以监测损失函数的变化趋势。对训练集和验证集进行约2 500次迭代训练后的误差变化曲线如图6所示。可以看出,前500次迭代中网络快速拟合;1 500次迭代后损失函数基本稳定,此后检测网络收敛。

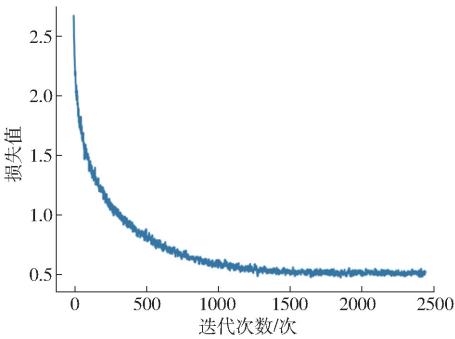


图6 训练误差变化曲线

Fig.6 Training error changing curve

本文选用的性能评价指标主要包括均值平均精

度 M_{AP} (Mean average precision)、红色和绿色番茄果实的检测精度 A_{PR} 和 A_{PG} 、检测时间 t_f 、准确率 P 、召回率 R 、交并比 I_{OU} 。本文目的是快速精确识别番茄果实,因此把 $M_{AP}、t_f$ 作为主要评价指标。交并比 I_{OU} 是预测框和真实框的重合程度,表示番茄果实的检测定位精度。 $A_{PR}、A_{PG}$ 是针对单一类别进行分析计算的, M_{AP} 是 $A_{PR}、A_{PG}$ 的平均值。计算公式为

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{2}$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \tag{3}$$

$$M_{AP} = \frac{1}{2} \sum_{k=1}^2 J(P,R)_k \tag{4}$$

式中 T_p ——实际为正样本且被检测为正样本的数量

F_p ——实际为负样本且被检测为正样本的数量

F_N ——实际为负样本且被检测为负样本的数量

k ——类别编号

$J(P,R)_k$ ——平均精度函数,即类别编号为 k 时准确率 P 与召回率 R 所构成 $P-R$ 曲线的面积

经过NMS方法处理后得到的预测框和对应类别并非全部正确,因此将置信度大于阈值 $a=0.5$ 的预测框定义为正样本,反之则为负样本;将正样本中与真实框的交并比大于阈值 $d=0.6$ 的样本视为 T_p ,反之视为 F_p 。若负样本中存在实际为正样本的,则视为 F_N 。

检测网络IMS-YOLO在darknet深度学习框架中进行训练。硬件环境为Intel Xeon CPU,E5-2680

v3@2. 50 GHz ×48 处理器,64 GB 运行内存,2 TB 硬盘,两个显存为 12 GB 的 GeForce GTX 1080ti GPU。软件环境为 Ubuntu16. 04 移动操作系统。输入图像尺寸 416 像素 ×416 像素,权值衰减速率为 0. 000 5,初始学习率为 0. 001,动量设置为 0. 9。

3. 2 番茄果实识别效果

由于 IMS-YOLO 模型是基于 YOLO v2 主干网络,同时融合了多尺度检测结构进行改进的,因此有

表 2 4 种网络模型的检测结果
Tab. 2 Detection results of four network models

模型	结构特点	红色番茄	绿色番茄	检测精度 $M_{AP}/\%$	检测时间 t_r/ms	准确率 $P/\%$	召回率 $R/\%$	交并比 $I_{OU}/\%$
		检测精度	检测精度					
		$A_{PR}/\%$	$A_{PG}/\%$					
IMS-YOLO	darknet-20 + 多尺度	97. 90	96. 35	97. 13	7. 719	96. 36	96. 03	83. 32
MS-YOLO v2	darknet-19 + 多尺度	97. 20	95. 35	96. 28	7. 626	96. 16	95. 24	81. 94
ISS-YOLO	darknet-20 + 单尺度	94. 00	88. 66	91. 33	6. 749	95. 35	90. 10	75. 30
YOLO v2	darknet-19 + 单尺度	93. 32	86. 60	89. 96	6. 580	93. 75	88. 34	74. 38

由表 2 可知,IMS-YOLO 比 MS-YOLO v2 的 M_{AP} 、 I_{OU} 分别提升 0. 85 个百分点和 1. 38 个百分点,而 t_r 仅增加了 0. 093 ms。IMS-YOLO 比 ISS-YOLO 的 M_{AP} 、 I_{OU} 分别提升了 5. 8 个百分点和 8. 02 个百分点, t_r 增加了 0. 97 ms。可以看出,分别更换了主干网络和检测结构,除了 t_r 略有增加,检测模型的性能均有不同程度提升,说明本文网络结构的改进方法对模型性能具有促进作用。而与 YOLO v2 相比,IMS-YOLO 的多项性能指标均有显著提升,尤其是 M_{AP} 和 I_{OU} 分别提高了 7. 17 个百分点和 8. 94 个百分点,极大地提高了番茄果实的检测精度和检测定位精度。另外,IMS-YOLO 比 YOLO v2 的 t_r 增加了 1. 139 ms,两者 t_r 相差不多,均能达到较好的实时性。综上,IMS-YOLO 的检测精度和检测定位精度最高。其原因为 IMS-YOLO 使用多尺度检测结构,可以较为精确地预测番茄果实预测框类别和位置,从而大幅度提升了检测精度和定位精度;IMS-YOLO 的主干网络增加了残差结构,可以促进前后卷积层的信息融合,有效提取番茄果实类别和真实框的信息,使训练误差更小,可以进一步提升检测精度。综合考虑各项检测指标,本文方法对番茄果实的检测性能更佳。

IMS-YOLO 的 A_{PR} 、 A_{PG} 分别是 97. 90% 和 96. 35%,较 MS-YOLO v2 分别提升了 0. 70 个百分点和 1. 00 个百分点;而相较于 ISS-YOLO 和 YOLO v2,IMS-YOLO 的 A_{PR} 分别提升了 3. 9 个百分点和 4. 58 个百分点, A_{PG} 则分别提升了 7. 69 个百分点和 9. 75 个百分点。可以看出,IMS-YOLO 的 A_{PR} 、 A_{PG} 提升最为显著,且 A_{PG} 提升更为明显;此外,多尺度检

必要对改进前后网络的番茄果实检测性能进行对比分析。为此,对 YOLO v2 保留 darknet-19 主干网络不变,采用本文所述多尺度检测模块与其一起构成 MS-YOLO v2 检测模型;基于 darknet-20 主干网络的改进型单一尺度 (Single-scale) 检测模型称为 ISS-YOLO。为了对比改进效果,同时与 YOLO v2 进行对比。上述 4 种 YOLO 系列检测模型的特点及其检测结果如表 2 所示。

测模型的检测效果远比单尺度的效果好。这是因为在 IMS-YOLO 中主干网络加入了残差模块且采用了多尺度检测结构,其提取番茄果实特征、区分背景和目标信息的能力更为突出。

3. 3 不同方法识别效果对比试验

基于 IMS-YOLO 的温室复杂环境下的番茄果实识别结果如图 7 所示。对照图 1,可以看到番茄果实数量不同、稠密程度不同、光照不同、叶片枝干遮挡、果实重叠等环境下均具有很高的识别准确率,且番茄果实在图像中定位准确。因此,本文方法具有较强的鲁棒性,可以适应温室下各种复杂情况。

为进一步验证本文网络模型对温室番茄果实的识别性能,主干网络分别取 ResNet34 和 VGGNet,且 VGGNet 选用前 13 层网络,去掉最后 3 层全连接层以降低参数量;检测模块统一采用 YOLOv3-tiny 的检测结构,构成 MS-ResNet34、IMS-VGGNet。将其与 IMS-YOLO 网络进行对比试验,测试网络层数较少的主干网络对番茄果实特征的提取能力。为验证 IMS-YOLO 的实时性,选取 YOLO v3 网络模型与其进行对比试验。番茄果实识别效果对比试验结果如图 7 和表 3 所示。图 7 中由左至右依次为单果、多果、逆光、顺光、重叠和遮挡。

由表 3 可以看出,IMS-YOLO 模型的 I_{OU} 最大,检测定位精度最高。IMS-YOLO 比 IMS-VGGNet 的 M_{AP} 高 0. 41 个百分点。比 MS-ResNet34 模型的 M_{AP} 高 3. 26 个百分点,且 t_r 较小。原因是 MS-ResNet34 在主干网络开始训练时先进行 8 倍下采样,特征信息丢失严重,因此检测精度降低;同时,其卷积层较多,计算量大,因此检测速度较慢。相对于 YOLO v3,

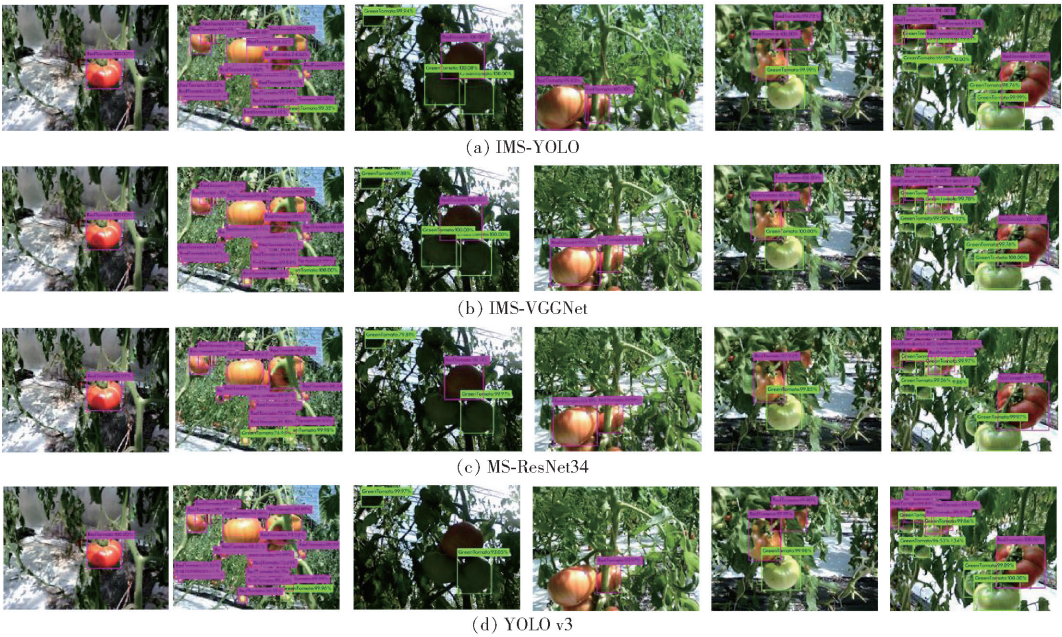


图7 复杂环境下4种模型对番茄果实的识别结果

Fig. 7 Recognition results of tomatoes under complex environment of four models

表3 4种模型的试验结果

Tab.3 Test results of four models

模型	红色番茄检测	绿色番茄检测	检测精度	检测时间	准确率	召回率	交并比
	精度 $A_{PR}/\%$	精度 $A_{PG}/\%$	$M_{AP}/\%$	t_t/ms	$P/\%$	$R/\%$	$I_{OU}/\%$
IMS-YOLO	97.90	96.35	97.13	7.719	96.36	96.03	83.32
IMS-VGGNet	97.59	95.85	96.72	6.420	96.25	95.31	81.85
MS-ResNet34	95.57	91.80	93.87	7.762	97.39	86.89	73.73
YOLO v3	99.23	98.07	98.65	20.126	98.24	85.13	77.85

虽然IMS-YOLO检测精度降低了1.52个百分点,但检测时间减少了12.407 ms,实时性较好。可以看出,本文方法整体性能表现最佳,可以同时满足识别精度和速度的需求。

另外,IMS-YOLO模型的 A_{PR} 、 A_{PG} 稍低于YOLO v3而高于其他两种模型,原因是IMS-YOLO的网络结构比YOLO v3简单得多,对番茄果实的特征提取能力偏低,因而检测精度偏低;而与IMS-VGGNet相比,两者的网络深度相近,但IMS-YOLO采用了残差结构,可提取更丰富的特征信息,因而检测精度较高;MS-ResNet34网络模型整体精度最低,这是因为其先进行8倍下采样,导致特征信息丢失严重,继而影响网络的拟合能力,对绿色番茄果实与背景的颜色区分能力下降,导致绿色番茄果实检测精度降低。4种网络模型对红色番茄果实的检测精度均高于绿色番茄果实,这是因为红色果实目标与背景颜色差异大,红色颜色特征易提取。

3.4 温室环境中番茄果实识别验证

为验证实际应用效果,将本文模型集成在采摘机器人操作系统中进行采摘试验。如图8所示,采摘机器人核心部件由遨博6轴协作机械臂、FM810-

IX-A型3D相机、智能移动底盘(国兴)以及一台工控机组成。采摘试验过程如下:将智能移动底盘移动到合适位置,由工控机调用IMS-YOLO模型识别番茄果实并控制3D相机对番茄果实进行检测;若3D相机检测到红色成熟番茄果实,则选择置信度概率最大的番茄果实目标为第1采摘目标并进行定位,输出番茄果实类别和三维坐标至工控机,再由工控机控制机械臂实现采摘。针对顺光、逆光两种光照情况共设计了8组试验,对单果/多果果实在有遮挡、无遮挡环境条件下进行番茄果实采摘试验,本次试验对枝叶遮挡程度大于60%以上的番茄果实不进行采摘。由于机器人只采摘红色成熟番茄果实,且每次只采摘一个,所以只统计红色番茄的识别率。若识别到番茄果实,系统会返回番茄果实三维坐标值,否则,返回至“未识别到番茄”。系统返回番茄坐标值即识别成功,因番茄未识别成功而导致采摘不成功的情况为识别失败,因机械臂路径规划失误等导致采摘不成功的情况不计入识别失败。

各种情况下番茄果实识别准确率如表4所示。由表4可以看出,逆光比顺光识别准确率低,原因是逆光时果实与背景界限不明显,果实边缘特征



图8 温室环境中机器人采摘试验

Fig.8 Test of harvesting robot under greenhouse environments

表4 温室环境下番茄果实采摘试验结果

Tab.4 Results of tomato harvesting tests under greenhouse environments

参数	顺光				逆光			
	单果		多果		单果		多果	
	遮挡	未遮挡	遮挡	未遮挡	遮挡	未遮挡	遮挡	未遮挡
番茄数量/ 个	50	50	100	100	50	50	100	100
识别准 确率/%	100	100	97	98	98	98	93	95

提取难度增加,导致识别难度增加;多果比单果识别准确率低,原因是密集情况下果实之间有重叠,果实形状特征提取困难;遮挡比未遮挡识别准确率偏低,原因是有枝叶遮挡时增加了果实形状特征提取的难度。

为节约成本,采摘机器人使用了只有 CPU 的工控机。工控机为单 CPU Intel i7-7500u 2.7 GHz 处理器、16 GB 内存。由于工控机的计算能力下降,识别试验中番茄果实的识别时间比图像识别时间偏长。

4 结论

(1)提出了改进型多尺度检测模型 IMS-YOLO,采用含有残差模块的 darknet-20 主干网络实现特征提取和检测速度的提升,借鉴 YOLO v3-tiny 检测结构,融合多尺度信息,提升检测精度。该模型对复杂环境下番茄果实的检测精度为 97.13%,检测时间为 7.719 ms,准确率为 96.36%,召回率和交并比分别为 96.03% 和 83.32%,红、绿番茄果实检测精度分别为 97.90% 和 96.35%。对比其他网络模型,该模型兼顾了识别精度与速度的要求,且检测定位精度最高,综合性能最佳。

(2)采用多尺度检测结构的 IMS-YOLO、MS-YOLO v2 的检测精度、召回率、交并比比单尺度的 ISS-YOLO、YOLO v2 更高;与采用 darknet-19 的 MS-YOLO v2、YOLO v2 相比,采用 darknet-20 主干网络的 IMS-YOLO、ISS-YOLO 的检测时间增加不多,对检测实时性影响不大;与 YOLO v2 相比,IMS-YOLO 的多项性能指标均有显著提升,极大地提高了番茄果实的检测精度和检测定位精度。

(3)对比 MS-ResNet34 与 IMS-VGGNet 模型,IMS-YOLO 模型精度最高;IMS-YOLO 比 YOLO v3 检测时间减少了 12.407 ms;对比其他 3 种网络模型,IMS-YOLO 的交并比最高,即检测定位精度最高,可以为采摘机器人提供精确的定位指导。

(4)在番茄温室大棚进行了机器人实际采摘试验,进一步验证了 IMS-YOLO 模型的可行性和准确性。该模型检测精度高、速度快,可兼顾精度和实时性的要求,在复杂环境下鲁棒性强、检测定位精度高,可以满足温室复杂环境下采摘机器人识别番茄果实的需求。

参 考 文 献

[1] 霍建勇. 中国番茄产业现状及安全防范[J]. 蔬菜, 2016(6): 1-4.
HUO Jianyong. Status and safety prevention of tomato industry in China [J]. Vegetable, 2016(6): 1-4. (in Chinese)

[2] 徐铭辰,牛媛媛,余永昌. 果蔬采摘机器人研究综述[J]. 安徽农业科学, 2014, 42(31): 11024-11027, 11057.
XU Mingchen, NIU Yuanyuan, YU Yongchang. Review of research on picking robots of fruits and vegetables [J]. Journal of Anhui Agri. Sci., 2014, 42(31): 11024-11027, 11057. (in Chinese)

[3] 刘继展. 温室采摘机器人技术研究进展分析[J/OL]. 农业机械学报, 2017, 48(12): 1-18.
LIU Jizhan. Research progress analysis of robotic harvesting technologies in greenhouse [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12): 1-18. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20171201&flag=1. DOI:10.6041/j.issn.1000-1298.2017.12.001. (in Chinese)

[4] 陶华伟,赵力,奚吉,等. 基于颜色及纹理特征的果蔬种类识别方法[J]. 农业工程学报, 2014, 30(16): 305-311.
TAO Huawei, ZHAO Li, XI Ji, et al. Fruits and vegetables recognition based on color and texture features [J]. Transactions of the CSAE, 2014, 30(16): 305-311. (in Chinese)

[5] 项荣. 开放环境中番茄的双目立体视觉识别与定位[D]. 杭州:浙江大学, 2013.
XIANG Rong. Recognition and localization for tomatoes under open environments based on binocular stereo vision [D]. Hangzhou: Zhejiang University, 2013. (in Chinese)

[6] 蒋焕煜,彭永石,申川,等. 基于双目立体视觉技术的成熟番茄识别与定位[J]. 农业工程学报, 2008, 24(8): 279-283.
JIANG Huanyu, PENG Yongshi, SHEN Chuan, et al. Recognizing and locating ripe tomatoes based on binocular stereo vision technology [J]. Transactions of the CSAE, 2008, 24(8): 279-283. (in Chinese)

- [7] 郑小东,赵杰文,刘木华. 基于双目立体视觉的番茄识别与定位技术[J]. 计算机工程, 2004, 30(22): 155 – 156, 171. ZHENG Xiaodong, ZHAO Jiewen, LIU Muhua. Tomatoes recognition and location from nature background based on binocular stereo vision [J]. Computer Engineering, 2004, 30(22): 155 – 156, 171. (in Chinese)
- [8] 王玉翰. 基于双目立体视觉的番茄识别与定位技术研究[D]. 杭州:浙江大学, 2012. WANG Yuhan. Research on recognizing and locating tomato base on binocular stereo vision technology [D]. Hangzhou: Zhejiang University, 2012. (in Chinese)
- [9] 梁喜凤,章艳. 串番茄采摘点的识别方法[J]. 中国农机化学报, 2016, 37(11): 131 – 134, 149. LIANG Xifeng, ZHANG Yan. Recognition method of picking point for tomato cluster [J]. Journal of Chinese Agricultural Mechanization, 2016, 37(11): 131 – 134, 149. (in Chinese)
- [10] 赵源深,贡亮,周斌,等. 番茄采摘机器人非颜色编码化目标识别算法研究[J/OL]. 农业机械学报, 2016, 47(7): 1 – 7. ZHAO Yuanshen, GONG Liang, ZHOU Bin, et al. Object recognition algorithm of tomato harvesting robot using non-color coding approach[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(7): 1 – 7. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20160701&flag=1. DOI:10.6041/j.issn.1000-1298.2016.07.001. (in Chinese)
- [11] 冯青春,程伟,杨庆华,等. 基于线结构光视觉的番茄重叠果实识别定位方法研究[J]. 中国农业大学学报, 2015, 20(4): 100 – 106. FENG Qingchun, CHENG Wei, YANG Qinghua, et al. Identification and localization of overlapping tomatoes based on linear structured light vision system [J]. Journal of China Agricultural University, 2015, 20(4): 100 – 106. (in Chinese)
- [12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 – 587.
- [13] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440 – 1448.
- [14] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems, 2015: 91 – 99.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779 – 788.
- [16] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517 – 6525.
- [17] REDMON J, FARHADI A. YOLO v3: an incremental improvement [R]. arXiv:1804.02767v1, 2018.
- [18] 傅隆生,冯亚利,ELKAMIL T,等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205 – 211. FU Longsheng, FENG Yali, ELKAMIL T, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. Transactions of the CSAE, 2018, 34(2): 205 – 211. (in Chinese)
- [19] 周云成,许童羽,邓寒冰,等. 基于双卷积链 Fast R-CNN 的番茄关键器官识别方法[J]. 沈阳农业大学学报, 2018, 49(1): 65 – 74. ZHOU Yuncheng, XU Tongyu, DENG Hanbing, et al. Recognition method of tomato key organs based on dual convolution Fast R-CNN [J]. Journal of Shenyang Agricultural University, 2018, 49(1): 65 – 74. (in Chinese)
- [20] 杨洋,张亚兰,苗伟,等. 基于卷积神经网络的玉米根茎精确识别与定位研究[J/OL]. 农业机械学报, 2018, 49(10): 46 – 53. YANG Yang, ZHANG Yalan, MIAO Wei, et al. Accurate identification and location of corn rhizome based on Faster R-CNN [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(10): 46 – 53. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20181006&flag=1. DOI:10.6041/j.issn.1000-1298.2018.10.006. (in Chinese)
- [21] 毕松,高峰,陈俊文,等. 基于深度卷积神经网络的柑橘目标识别方法[J/OL]. 农业机械学报, 2019, 50(5): 181 – 186. BI Song, GAO Feng, CHEN Junwen, et al. Detection method of citrus based on deep convolution neural network [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 181 – 186. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20190521&flag=1. DOI:10.6041/j.issn.1000-1298.2019.05.021. (in Chinese)
- [22] 张领先,陈运强,李云霞,等. 基于卷积神经网络的冬小麦麦穗检测计数系统[J/OL]. 农业机械学报, 2019, 50(3): 144 – 150. ZHANG Lingxian, CHEN Yunqiang, LI Yunxia, et al. Detection and counting system for winter wheat ears based on convolutional neural network [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 144 – 150. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20190315&flag=1. DOI:10.6041/j.issn.1000-1298.2019.03.015. (in Chinese)
- [23] 穆龙涛,高宗斌,崔永杰,等. 基于改进 AlexNet 的广域复杂环境下遮挡猕猴桃目标识别[J/OL]. 农业机械学报, 2019, 50(10): 24 – 34. MU Longtao, GAO Zongbin, CUI Yongjie, et al. Kiwifruit detection of far-view and occluded fruit based on improved AlexNet [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(10): 24 – 34. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20191003&flag=1. DOI:10.6041/j.issn.1000-1298.2019.10.003. (in Chinese)
- [24] BARGOTI S, UNDERWOOD J. Deep fruit detection in orchards [C]//IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 3626 – 3633.
- [25] DING W, TAYLOR G. Automatic moth detection from trap images for pest management [J]. Computers and Electronics in Agriculture, 2016, 123: 17 – 28.