

doi: 10.6041/j.issn.1000-1298.2022.09.003

# 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别

闫彬<sup>1,2</sup> 樊攀<sup>1,2</sup> 王美茸<sup>1,2</sup> 史帅旗<sup>1,3</sup> 雷小燕<sup>1,4</sup> 杨福增<sup>1,2</sup>

- (1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100;  
2. 农业农村部苹果全程机械化科研基地, 陕西杨凌 712100;  
3. 农业农村部北方农业装备科学观测实验站, 陕西杨凌 712100;  
4. 黄土高原土壤侵蚀与旱地农业国家重点实验室, 陕西杨凌 712100)

**摘要:** 为准确识别果树上的不同苹果目标,并区分不同枝干遮挡情形下的果实,从而为机械手主动调整位姿以避免枝干对苹果的遮挡进行果实采摘提供视觉引导,提出了一种基于改进 YOLOv5m 面向采摘机器人的苹果采摘方式实时识别方法。首先,改进设计了 BottleneckCSP-B 特征提取模块并替换原 YOLOv5m 骨干网络中的 BottleneckCSP 模块,实现了原模块对图像深层特征提取能力的增强与骨干网络的轻量化改进;然后,将 SE 模块嵌入到所改进设计的骨干网络中,以更好地提取不同苹果目标的特征;进而改进了原 YOLOv5m 架构中输入中等尺寸目标检测层的特征图的跨接融合方式,提升了果实的识别精度;最后,改进了网络的初始锚框尺寸,避免了对图像里较远种植行苹果的识别。结果表明,所提出的改进模型可实现对图像中可直接采摘、迂回采摘(苹果上、下、左、右侧采摘)和不可采摘果实的识别,识别召回率、准确率、mAP 和 F1 值分别为 85.9%、81.0%、80.7% 和 83.4%。单幅图像的平均识别时间为 0.025 s。对比了所提出的改进算法与原 YOLOv5m、YOLOv3 和 EfficientDet-D0 算法在测试集上对 6 类苹果采摘方式的识别效果,结果表明,所提出的算法比其他 3 种算法识别的 mAP 分别高出了 5.4、22、20.6 个百分点。改进模型的体积为原始 YOLOv5m 模型体积的 89.59%。该方法可为机器人的采摘手主动避开枝干对果实的遮挡,以不同位姿采摘苹果提供技术支撑,可降低苹果的采摘损失。

**关键词:** 苹果; 采摘机器人; YOLOv5m; 采摘方式识别; 视觉引导; 深度学习

**中图分类号:** TP391.4; S24 **文献标识码:** A **文章编号:** 1000-1298(2022)09-0028-11

**OSID:**



## Real-time Apple Picking Pattern Recognition for Picking Robot Based on Improved YOLOv5m

YAN Bin<sup>1,2</sup> FAN Pan<sup>1,2</sup> WANG Meirong<sup>1,2</sup> SHI Shuaiqi<sup>1,3</sup> LEI Xiaoyan<sup>1,4</sup> YANG Fuzeng<sup>1,2</sup>

- (1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China  
2. Apple Full Mechanized Scientific Research Base of Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China  
3. Scientific Observing and Experimental Station of Agricultural Equipment for the Northern China, Ministry of Agriculture and Rural Affairs, Yangling, Shaanxi 712100, China  
4. State Key Laboratory of Soil Erosion and Dryland Farming on Loess Plateau, Yangling, Shaanxi 712100, China)

**Abstract:** In order to accurately identify the different fruit targets on apple trees, and automatically distinguish the fruit occluded by different branches, providing visual guidance for the mechanical picking end-effector to actively adjust the pose of apple picking to avoid the shelter of the branches, a real-time recognition method of apple picking pattern based on improved YOLOv5m for picking robot was proposed. Firstly, BottleneckCSP module was designed and improved to BottleneckCSP-B module which was used to replace the BottleneckCSP module in backbone architecture of original YOLOv5m network. The ability of image deep feature extraction of the original BottleneckCSP module was enhanced, and the original YOLOv5m backbone network was lightweight designed and improved. Secondly, SE module was inserted to the proposed improved backbone network, to better extract the features of different apple targets. Thirdly, the bonding fusion mode of feature maps, which were input to the target detection layer of

收稿日期: 2022-04-01 修回日期: 2022-06-15

基金项目: 陕西省科技重大专项(2020zdzz03-04-01)

作者简介: 闫彬(1994—),男,博士生,主要从事苹果采摘机器人智能化感知与控制技术研究,E-mail: yanbin@nwfau.edu.cn

通信作者: 杨福增(1966—),男,教授,博士生导师,主要从事智能农业装备及其相关技术研究,E-mail: yangfzkm@nwfau.edu.cn

medium size in the original YOLOv5m network, were improved, and the recognition accuracy of apple was improved. Finally, the initial anchor box sizes of the original network were improved, avoiding the misrecognition of apples in farther plant row. The experimental results indicated that the graspable, circuitous-graspable (up-graspable, down-graspable, left-graspable, right-graspable) and ungraspable apples could be identified effectively by using the proposed improved model in the study. The recognition recall, precision, mAP and F1 were 85.9%, 81.0%, 80.7% and 83.4%, respectively. The average recognition time was 0.025 s per image. Contrasted with original YOLOv5m, YOLOv3 and EfficientDet-D0 model, the mAP of the proposed improved YOLOv5m model was increased by 5.4 percentage points, 22 percentage points and 20.6 percentage points, respectively on test set. The size of the improved model was 89.59% of original YOLOv5m model. The proposed method can provide technical support for the picking end-effector of robot to pick apples in different poses avoiding the shelter of branches, to reduce the loss of apple picking.

**Key words:** apple; picking robot; YOLOv5m; picking pattern recognition; visual guidance; deep learning

## 0 引言

目前苹果采摘机器人<sup>[1-5]</sup>可以实现对其视觉范围内未被遮挡或仅被树叶遮挡苹果的直接采摘,而针对被枝干遮挡的苹果,若未经精确识别而直接对其进行采摘可能会造成果实损伤,或机械手、机械臂损坏<sup>[6]</sup>。

深度学习技术因具有能够对所采集信息数据的特征进行较好地挖掘与提取的优点,近年来,已经被广泛应用于目标识别领域中<sup>[7-10]</sup>。截至目前,在基于深度学习的苹果目标识别方面,已经有许多深度学习网络架构,如YOLOv2<sup>[11]</sup>、YOLOv3<sup>[12-14]</sup>、YOLOv4<sup>[15]</sup>、YOLOv5s<sup>[6]</sup>、Faster R-CNN<sup>[16-19]</sup>、DaSNet-v2<sup>[20]</sup>、R-FCN<sup>[21]</sup>、LedNet<sup>[22]</sup>、FCOS<sup>[23]</sup>、DaSNet<sup>[24]</sup>和Mask R-CNN<sup>[25]</sup>等,都被成功应用于检测苹果树上的果实目标。

然而,现有的算法大多将复杂果园环境(苹果被树叶遮挡、枝干遮挡、果实遮挡和混合遮挡等)下的不同苹果划分为同一类目标进行识别,而现实中极有可能会造成果实的损伤,或采摘手、机械臂的损坏。本研究团队前期针对苹果被树枝干遮挡的情形进行了识别<sup>[6]</sup>,以引导机器人避开对这些果实的采摘,但会导致一部分苹果漏采。另一方面,若根据苹果被枝干遮挡的不同情形而相应地改变采摘手的位姿,则可以使机器人实现对该类苹果的迂回采摘,以降低不必要的采摘损失。然而,目前尚未见可以区分枝干单侧遮挡(即枝干遮挡苹果的上、下、左、右侧)、多侧遮挡情形下苹果识别算法的研究报道。

为解决苹果采摘机器人采摘识别时所面临的上述问题,本文基于人工智能算法,提出一种采摘机器人的苹果采摘方式识别方法,拟实现对苹果树上不同枝干遮挡情形下果实的实时识别,从而为机械手

主动调整位姿以避开枝干的遮挡进行果实采摘提供视觉引导,以期降低苹果的采摘损失。

## 1 苹果图像数据获取与预处理

### 1.1 图像采集方法

以现代标准果园中纺锤形栽培模式下的红富士苹果果实为研究对象,该栽培模式下的苹果树行距约4 m,株距约1.2 m,树高约3.5 m,适合苹果采摘机器人入园进行自动化采摘作业。采摘机器人所面临的实际采摘情况如图1所示。



图1 机器人所面临的苹果采摘情况示意图

Fig. 1 Schematic of apple picking situation that robot confronted

对位于陕西省乾县农业科技试验示范基地的现代标准苹果园与西北农林科技大学白水苹果试验站中的苹果树进行图像采集。分别采集晴天与阴天条件下的苹果树图像,拍摄时段包括上午、中午与下午,图像采集设备为佳能 Canon PowerShot G16 型相机,分别在不同的拍摄距离下(0.5~1.5 m)选择多种角度进行图像采集,共采集苹果图像1 400幅。具体包含以下情况:果实被树叶遮挡、果实被枝干遮挡(枝干遮挡苹果的上、下、左、右侧或多侧遮挡)、混合遮挡、果实之间重叠、顺光角度、逆光角度和侧光角度等(图2、3)。所采集图像分辨率为4 000像素×3 000像素,格式为JPEG。



图2 不同情况下的苹果图像

Fig.2 Apple images under different conditions

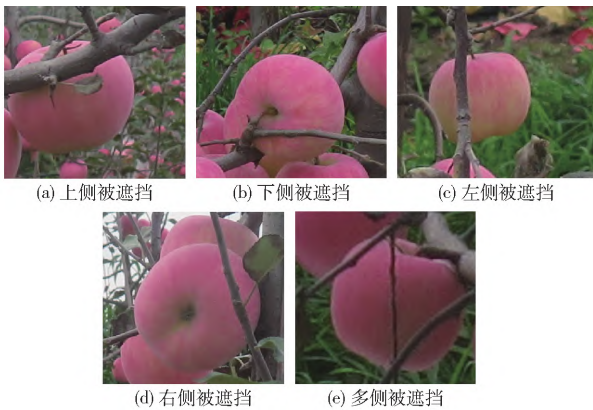


图3 果实被枝干遮挡的不同情形

Fig.3 Different situations of apples occluded by branches

## 1.2 图像数据预处理

从所采集的图像中随机选取拍摄质量稳定的344幅图像(晴天与阴天各172幅)作为测试集,另随机选取1014幅图像用于模型训练。测试集图像中含不可采摘果实1202个,可直接采摘果实1952个,上侧采摘果实658个,下侧采摘果实909个,左侧采摘果实950个和右侧采摘果实865个。

使用LabelImg图像数据标注软件在压缩后的果树图像中绘制苹果的外接矩形框以实现果实的人工标注。图像标注时,需要基于每个苹果的最小外接矩形进行标注,以保证矩形框内尽可能少地包含背景。其中,根据机器人在苹果园采摘作业时面临的实际情况,将图像中需要识别的苹果目标分别划归入6个类别进行数据标注,具体的类别划分规则为:将图像中未被遮挡或仅被树叶遮挡的苹果标注为“可直接采摘”类,将果实上侧被枝干遮挡的苹果标注为“下侧采摘”类,下侧被枝干遮挡的苹果标注为“上侧采摘”类,左侧被枝干遮挡的苹果标注为“右侧采摘”类,右侧被枝干遮挡的苹果标注为“左侧采摘”类,其他情况下的苹果标注为“不可采摘”类。

另一方面,由于较远处种植行(非机器人所处的当前果树种植行)中的苹果与采摘机器人间的距离过大,故机器人无法对其进行采摘作业。而机器人视觉系统所获取的苹果树图像中会不可避免地拍摄到较远种植行中的苹果,因此并不能将其作为有效的待识别/采摘目标。因而模型需要避免对较远处他行苹果的识别。故在苹果图像数据集标注时,对于较远处他行的苹果均不进行标签标注。最后,保存标注后所生成的XML格式文件。

为了丰富训练集的图像数据,以更好地提取不同标注类别苹果的特征,对训练集图像进行数据扩增。由于存在光照与天气等不确定因素,导致机器人在果园进行识别采摘作业时的视觉感知环境十分复杂,为了提高苹果采摘方式识别模型的泛化能力,对训练集的图像分别进行亮度增强与减弱的数据增强方式:首先,将原始图像转换至HSV彩色空间;然后,将图像的V分量(亮度分量)乘以不同的权重系数;最后,将新合成的HSV彩色空间图像转换至RGB彩色空间,实现图像亮度的增强和减弱。在本研究中,利用亮度增强产生2种亮度的图像: $H + S + 1.2V$ 和 $H + S + 1.6V$ ,其中H为色调分量,S为饱和度分量;利用亮度减弱产生2种亮度的图像: $H + S + 0.6V$ 和 $H + S + 0.8V$ 。

另外,考虑到图像采集装置在拍摄过程中可能产生的噪声以及由于设备晃动或树枝干抖动而使所获取的图像存在模糊的情况,分别对图像添加了方差为0.02的高斯噪声,进行了运动模糊处理。在运动模糊处理中,将运动滤波器的参数( $LEN, \theta$ )( $LEN$ 代表长度,表示摄像机线性运动的像素; $\theta$ 代表其逆时针方向旋转的角度)设置为(7,45)。然后利用Matlab中的imfilter函数对图像进行模糊处理。将经过数据扩增处理后所得到的共7098幅图像作为训练集数据用于后续苹果采摘方式识别模型的训练。

## 2 改进的YOLOv5m网络架构

### 2.1 YOLOv5m网络架构

YOLOv5网络架构具有检测精度高、运行速度快的优势,最高检测速度可达140帧/s<sup>[26]</sup>。另一方面,该网络模型的权重文件较小,与YOLOv4模型相比缩小了近90%,这使得YOLOv5模型适合部署到嵌入式设备上以实现目标的实时检测。由于模型的检测精度、实时性与是否轻量化直接关系到机器人识别果实采摘方式的准确率与效率,因此本研究以YOLOv5架构为基础,改进设计苹果采摘机器人的果实采摘方式识别网络。

YOLOv5 网络<sup>[6,27-28]</sup> 具体包含了 YOLOv5s、YOLOv5m、YOLOv5l 与 YOLOv5x 4 种架构,其主要区别为在网络的特定位置处所包含的特征提取模块数量和卷积核数量不同,4 种模型的参数数量和体积依次增大,指标参数如表 1 所示。

表 1 YOLOv5 4 种架构模型指标参数

Tab.1 Model parameters of four YOLOv5 architectures

模型架构	深度	宽度	层数	参数数量
YOLOv5s	0.33	0.50	191	$7.26 \times 10^6$
YOLOv5m	0.67	0.75	263	$2.15 \times 10^7$
YOLOv5l	1.00	1.00	335	$4.78 \times 10^7$
YOLOv5x	1.33	1.25	407	$8.90 \times 10^7$

由于本研究共需要识别 6 类目标,且对模型的识别实时性与轻量化要求较高,综合考虑模型识别的准确率、效率及模型的体积,确定以 YOLOv5m 架构(图 4)为基础,改进设计苹果采摘方式识别网络。

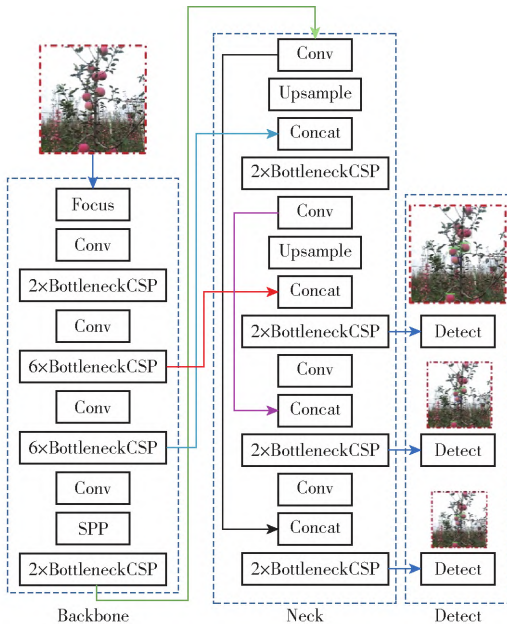


图 4 原始的 YOLOv5m 网络结构

Fig. 4 Architecture of original YOLOv5m network

YOLOv5m 架构主要由 Backbone 网络、Neck 网络和 Detect 网络组成。Backbone 网络即特征提取网络,是在不同的图像细粒度上聚合并形成图像特征的卷积神经网络。Backbone 网络的第 1 层为 Focus 模块(图 5),该模块的设计主要是为了减少模型的计算量以加快训练速度,其主要功能是:首先通过切片操作将输入的 3 通道图像(YOLOv5m 架构默认的图像输入尺寸为  $3 \times 640 \times 640$ )切分成 4 份  $3 \times 320 \times 320$  的切片,然后使用 Concat 操作从深度上连接这 4 个切片,输出的特征图尺寸为  $12 \times 320 \times 320$ ,进而再通过由 48 个卷积核组成的卷积层,生成  $48 \times 320 \times 320$  的输出,最后再经过 BN 层(Batch normalization)和 Hardswish 激活函数后将结果输入

到下一层。

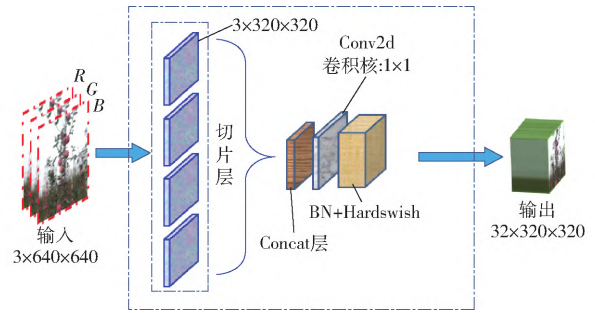


图 5 Focus 模块

Fig. 5 Focus module

Backbone 网络的第 3 层为 BottleneckCSP 模块,该模块的主要作用是为了更好地提取图像的深层特征。其中,BottleneckCSP 模块主要由 Bottleneck 模块(图 6)构成,该模块是一种残差结构的网络,即将卷积核尺寸为  $1 \times 1$  的卷积层(Conv2d 层 + BN 层 + Hardswish 激活函数)与  $3 \times 3$  的卷积层相连,再将该部分的输出通过残差结构与输入相加作为最终 Bottleneck 模块的输出。

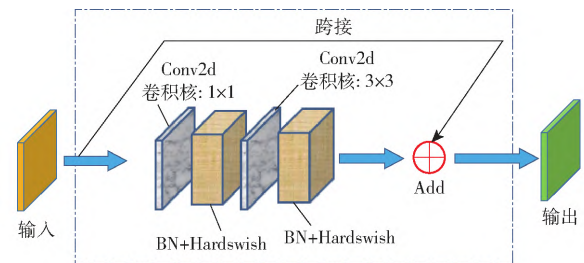


图 6 Bottleneck 模块

Fig. 6 Bottleneck module

而 BottleneckCSP 模块(图 7)的主要功能是:将初始的输入分别送入 2 个分支,分别对这 2 个分支进行卷积操作使其特征图的通道数减半,其中,分支 2 再与 Bottleneck 模块相连,进而在经过 Conv2d 层后,使用 Concat 操作从深度上连接分支 1 与分支 2 的输出特征图。最后再依次经过 BN 层与 Conv2d 层后得到 BottleneckCSP 模块的输出特征图,该特征图的尺寸与输入 BottleneckCSP 模块的特征图尺寸相同。

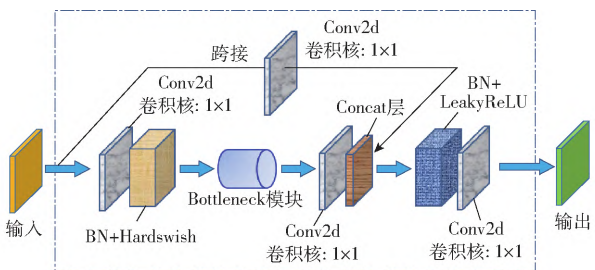


图 7 BottleneckCSP 模块

Fig. 7 BottleneckCSP module

Backbone 网络的第 9 层为空间金字塔池化模

块(Spatial pyramid pooling, SPP) (图8), 该模块的主要功能是为了使任意尺寸的特征图都能够被转换成固定尺寸的特征向量, 以提高网络的感受野。其中, YOLOv5m 中 SPP 模块的输入特征图尺寸为  $768 \times 20 \times 20$ , 首先经过  $1 \times 1$  的卷积层后输出尺寸为  $384 \times 20 \times 20$  的特征图, 然后将该特征图与其分别经过 3 个并行的 Maxpooling 层(最大池化层) 进行下采样后的输出特征图从深度上进行连接, 输出尺寸为  $1536 \times 20 \times 20$  的特征图, 最后再经过卷积核数量为 768 的卷积层后输出尺寸为  $768 \times 20 \times 20$  的特征图。

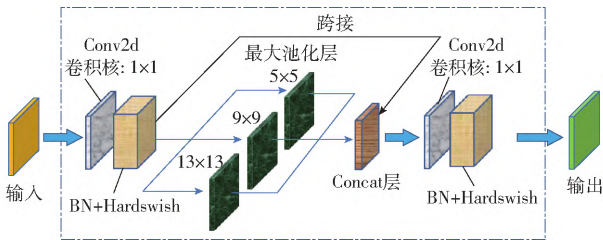


图8 SPP 模块  
Fig. 8 SPP module

Neck 网络是一系列组合图像特征的特征融合网络, 采用自顶向下与自底向上联合的特征融合方式, 更好地融合了多尺度特征, 因此改善了低层特征的传播, 增强了模型对于不同缩放尺度对象的检测, 从而能够更好地识别不同尺寸和尺度的同一目标对象。

Detect 网络主要用于模型最终的推理与检测, 该网络在上一层 Neck 网络输出的特征图上应用锚框(Anchor box), 最后输出包含目标对象的类别概率、对象得分和包围框位置的向量。YOLOv5m 架构的预测网络由 3 个 Detect 层组成, 其输入尺寸分别为  $80 \times 80$ 、 $40 \times 40$  与  $20 \times 20$  的特征图, 用于检测图像中不同尺寸的目标物, 每个 Detect 层最终输出一个 33 通道的向量(3 个锚定框, 每个锚定框包括 6 个类别、1 个类别概率、4 个包围框位置坐标, 即  $(6+1+4) \times 3$ ), 进而在原始图像中生成并标记出所预测目标的位置边界框和类别, 实现对图像中目标对象的检测。

## 2.2 BottleneckCSP 模块和骨干网络改进

对于设计面向采摘机器人的苹果采摘方式识别算法, 既要使其能够准确区分复杂果园环境中的多种果实采摘方式, 还需要尽可能压缩算法模型的体积以有利于后期将其部署在硬件设备中。因此, 在 YOLOv5m 架构的基础上, 对其骨干网络进行优化改进, 在能有效提取图像的深层特征以保证识别精度的前提下, 降低网络的权重参数量, 缩小其体积, 以实现苹果采摘方式识别网络的轻量化改进设计。

由于所设计的识别算法需要区分 6 种不同的果实采摘方式, 因此识别网络需要对苹果图像的深层特征进行充分提取, 以精确辨识不同苹果采摘方式间的差异。为此, 本研究对 BottleneckCSP 模块进行了改进设计, 以增强其对图像深层特征提取的能力。

使用 Bottleneck 模块替换原 BottleneckCSP 模块中 Concat 特征图拼接层之前的 Conv2d 层。由图 6 可知, Bottleneck 模块中既含有跨接融合层, 又包含多个 Conv2d 卷积层, 因而该模块既可实现对图像低层特征与高层特征的融合(可弥补低分辨率的高层特征图中空间信息的损失), 又可实现对图像更深层特征的提取。故该模块相较于单一的 Conv2d 卷积层具有更强的图像特征提取能力。进而, 在上述改进的基础上, 再去掉原 BottleneckCSP 模块跨接分支上的卷积层, 则实现了将分辨率更高、包含更多目标物位置与细节信息的低层特征图与经过多个特征提取模块后包含更强语义信息的高层特征图进行融合, 使得改进后 BottleneckCSP 模块的输出特征图含有更强的语义信息与更多的细节信息, 以提升该模块的特征提取能力。

改进后的 BottleneckCSP 模块如图 9 所示, 将其命名为 BottleneckCSP-B 模块。另一方面, 为了缩小识别模型的体积, 降低骨干网络的整体参数量, 以实现模型的轻量化, 本研究将原骨干网络中的 4 处共使用到 16 个 BottleneckCSP 模块的地方都分别替换为 4 个相连的 BottleneckCSP-B 模块。

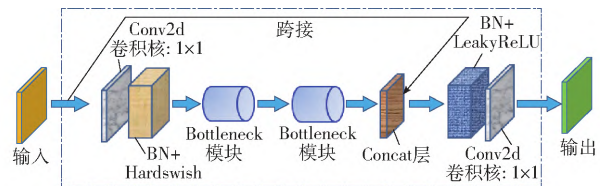


图9 改进 BottleneckCSP (BottleneckCSP-B) 模块  
Fig. 9 Improved BottleneckCSP (BottleneckCSP-B) module

## 2.3 骨干网络中 SE 模块嵌入设计

由于苹果外形和颜色与树枝干、树叶及图像中的背景物相比具有一定的差异, 因此为了提升果实采摘方式的识别精度, 将机器视觉中的注意力机制<sup>[24]</sup>(Attention mechanism) 引入苹果采摘方式识别网络的架构中, 以更好地提取不同果实图像的特征。SE 模块<sup>[20, 29]</sup>(Squeeze and excitation networks, SENet) 是视觉注意力机制网络的一种, 其采用了一种全新的特征重标定策略, 即通过学习的方式自动获取到每个特征通道的重要程度, 然后据此提升有用的特征并抑制不重要的特征。该模块主要包含 3 种操作: 挤压(Squeeze)操作、激励(Excitation)操作、缩放(Scale)操作。由于该模块的计算量不大, 且能

有效提升模型的表达能力,优化模型所学习的内容,因此将其嵌入所改进设计的 YOLOv5m 架构的骨干网络中,以提升模型的识别精度。SE 模块结构示意图如图 10 所示,该模块在改进 YOLOv5m 网络(图 11)中的具体嵌入布局如下:分别在改进后骨干网络的第 5、7、11、14 层嵌入 SE 模块。第 7 层的嵌入,改进 YOLOv5m 架构的第 1 个检测层(Detect 层),可实现对图像中尺寸相对较大目标物的识别。在改进后的网络结构中,第 7 层输出的特征图与第 20 层输出的高层特征图进行拼接后输入至第 1 个 Detect 层,因而在前端第 7 层位置处嵌入 SE 模块,使第 20 层输出的含有高维特征信息的特征图与第 7 层经过 SE 模块凸显了尺寸较大目标物信息的特征图进行融合,有利于第 1 检测层对于图像中相对较大目标物的准确识别。

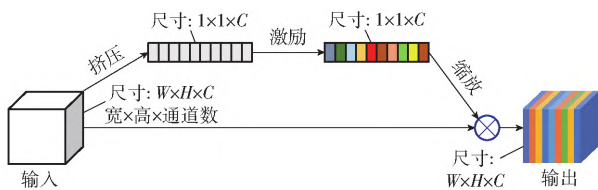


图 10 SE 模块结构示意图  
Fig. 10 SE module

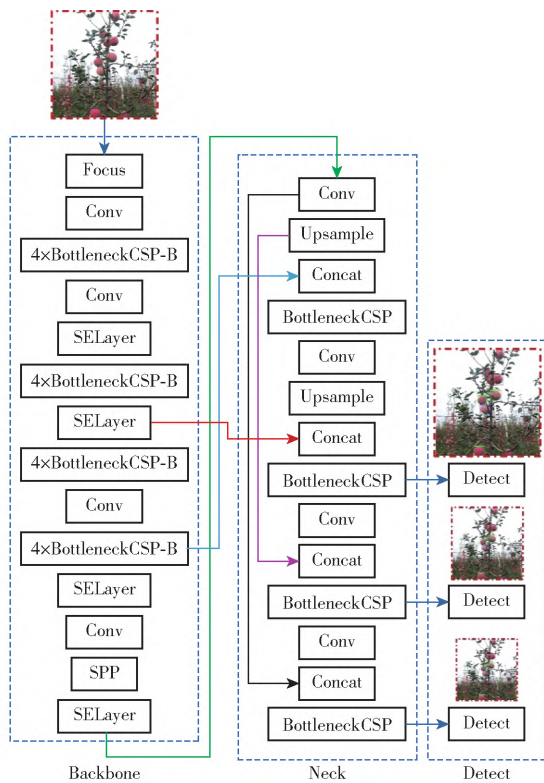


图 11 改进的 YOLOv5m 网络结构  
Fig. 11 Architecture of improved YOLOv5m network

第 14 层的嵌入。改进网络第 13 层的 SPP 模块实现了局部特征和全局特征的融合,丰富了特征图的表达能力,有利于对图像中目标大小差异较大的

情况进行识别,在其后嵌入 SE 模块,以进一步提升 SPP 模块输出特征图中明显的特征,并抑制不重要的特征。

改进后的 YOLOv5m 骨干网络中存在多个含卷积操作的特征提取模块,如 BottleneckCSP-B 模块与 Conv 模块,都具有对图像特征进行提取的功能。因此,在这些特征提取模块后接入 SE 模块可以在得到深层特征图的基础上,挖掘特征图中各通道之间的相互关系,提取细节信息,进一步优化经过卷积层所提取的特征。故在上述改进的基础上,在改进后骨干网络的第 1~7 层、7~14 层之间嵌入 SE 模块。

第 5 层的嵌入。由于骨干网络对图像特征的提取程度是随层数递进的,当骨干网络前端对特征提取还不充分时,在过于靠前的位置嵌入 SE 模块以进行特征优化的效果并不明显。另一方面,2 个 SE 模块相连的架构又会造成资源冗余,故将 SE 模块嵌入改进后骨干网络的第 5 层。

第 11 层的嵌入。由于第 7~14 层之间各模块的输出特征图维度包含了 192、384、768,且维度为 192 与 768 的特征图均经过 SE 模块进行特征优化,故在 7~14 层间输出特征图维度为 384 的模块后接入 SE 模块。另一方面,为了将低层特征与包含强语义信息的更高层特征图进行跨接融合,以增强网络的特征提取能力,确定在 7~14 层间将 SE 模块嵌入至引出了低层跨接线(图 11 中浅蓝色跨接线)的第 10 层 BottleneckCSP-B 模块之后的第 11 层。

SE 模块的输入与输出特征图的维度是相同的。在将该模块嵌入 YOLOv5m 骨干网络时,需要确保设置其输入特征图维度值与上一层模块的输出特征图维度值相同。由于本研究是在改进 YOLOv5m 骨干网络的第 5、7、11 和 14 层分别嵌入 SE 模块,根据嵌入位置前一模块的输出特征图维度尺寸,将各嵌入 SE 模块的输入特征图维度值分别设置为 192、192、384、768。

### 2.4 跨接融合特征层改进

融合不同尺度的特征是提高目标检测网络识别性能的一个重要手段。特征融合是将从不同图像中提取到的特征合成更具判别能力的特征图。低层特征图的分辨率较高,包含较多的目标物位置与细节信息,但由于卷积层对其特征进行提取较少,因而特征图的语义性较低,所含噪声较多。而高层特征图具有较强的语义信息,但其分辨率较低,对图像中细节的感知能力较差。因此,对高层与低层特征进行有效融合,是改善模型检测性能的关键。

基于 2.2 节和 2.3 节中对 YOLOv5m 架构骨干网络的改进设计,结合改进后网络各层输出特征图

的尺寸,将原 YOLOv5m 架构的第 5 层与第 17 层(图 4 中红色跨接线)、第 7 层与第 13 层(图 4 中浅蓝色跨接线)、第 11 层与第 23 层(图 4 中黑色跨接线)融合,分别更改为本研究所设计网络的第 7 层与第 21 层(图 11 中红色跨接线)、第 10 层与第 17 层(图 11 中浅蓝色跨接线)、第 15 层与第 27 层(图 11 中黑色跨接线)融合。

另一方面,经过对所获取的苹果树图像进行分析,相对于整幅图像,所需识别的苹果目标大部分属于中等尺寸。由于所改进设计网络架构的第 25 层输出特征图是作为中等尺寸目标检测层的输入,因此,为提升对当前种植行苹果目标的检测精度,弥补高层特征因其自身的低分辨率所导致的空间信息的损失,对原 YOLOv5m 架构中输入中等尺寸目标检测层的特征图的跨接融合(第 15 层与第 20 层融合,见图 4 中粉红色跨接线)进行改进,将低层感受野更大的特征提取层的输出与中等尺寸目标检测层之前的特征提取层输出进行融合,即将改进网络的第 16 层与第 23 层的输出特征图融合(图 11 中粉红色跨接线)。改进设计的苹果采摘方式识别网络架构如图 11 所示。

## 2.5 初始锚框尺寸改进

YOLOv5m 架构针对输入多尺度检测层用于识别小、中、大目标物的 3 种尺寸特征图(80×80、40×40、20×20)各设置了 3 种初始检测锚框尺寸,分别为:10×13、16×30、33×23;30×61、62×45、59×119;116×90、156×198、373×326。

对于机器人视觉系统所获取的苹果树图像,由于处在图像里较远处种植行中的苹果与采摘机器人间的距离过大,因此并不能将其作为有效的待识别/采摘目标。为了避免对图像中较远种植行小苹果的识别并提高对机器人所处的当前果树种植行苹果目标的识别准确率,在对所获取图像中当前种植行里的小、中、大苹果的尺寸以及图像中较远种植行小苹果的尺寸及图像尺寸进行综合分析的基础上,对原 YOLOv5m 网络中的小、中尺度目标检测层的初始锚框尺寸进行了改进设计,分别修改为 80×70、75×75、85×100;95×110、130×110、115×125。以期实现图像里当前种植行果实目标的准确识别。

## 2.6 网络损失函数

YOLOv5 网络的损失函数  $L$  主要由回归框预测损失  $L_{loc}$ 、置信度损失  $L_{conf}$  与目标分类损失  $L_{class}$  组成,计算公式为

$$L = L_{conf} + L_{class} + L_{loc} \quad (1)$$

其中,置信度损失和目标分类损失采用二进制交叉熵损失函数(BCELoss),计算公式为

$$BCELoss = \begin{cases} -\log p' & (y=1) \\ -\log(1-p') & (y=0) \end{cases} \quad (2)$$

式中, $BCELoss$  表示 BCE 损失函数, $p'$  表示样本的预测值, $y$  表示样本的真实类别, $y=1$  表示属于该类目标, $y=0$  则表示不属于该类目标。

回归框预测损失采用的是 GIoU Loss 函数。GIoU loss 不仅考虑到了真实框与预测框的重叠区域,还关注了非重叠区域。这样能较好地反映两框之间的距离,因而目标框的回归会更加稳定,避免了使用 IOU 进行模型训练时出现的目标框回归发散问题。

## 3 模型训练与评价

### 3.1 网络训练

#### 3.1.1 训练平台

基于联想 Legion Y7000P 型计算机(Intel(R) Core(TM) i7-9750H CPU, 2.6 GHz, 16 GB 内存; NVIDIA GeForce RTX 2060 GPU, 6 GB 显存),在 Windows 10 操作系统下搭建了 Pytorch 深度学习框架,使用 Python 语言编写程序并调用 CUDA、Cudnn 和 OpenCV 等所需的库,实现了对面向采摘机器人的果实采摘方式识别模型的训练和测试。

采用随机梯度下降法(SGD)以端到端的联合方式训练所改进设计的 YOLOv5m 网络。模型训练时采用 4 个样本作为一个批处理单元(Batch size),每次更新权值时使用 BN 层进行正则化,动量因子(Momentum)设置为 0.937,权值衰减率(Decay)设置为 0.0005,初始学习率设置为 0.01,IOU 阈值设置为 0.01,色调( $H$ )、饱和度( $S$ )与明度( $V$ )的增强系数分别设置为 0.015、0.7 和 0.4,共训练 300 轮(Epochs)。模型训练结束后保存所得到的权重文件,在测试集上对识别模型的性能进行评估。在经过非极大值抑制(Non-maximum suppression, NMS)等后处理操作消除了大量冗余的预测框后,网络的最终输出为置信度得分最高的苹果采摘方式预测类别,并返回果实位置预测框的坐标。

#### 3.1.2 训练结果

网络训练的损失(Loss)曲线见图 12,由图 12 可知,网络在前 50 轮训练时,损失值快速下降,训练 250 轮之后,损失值基本趋于稳定。因此本研究将训练 300 轮结束后输出的模型确定为面向采摘机器人的苹果采摘方式识别模型。

### 3.2 模型测试与评价

#### 3.2.1 苹果采摘方式识别性能评价指标

采用准确率( $P$ )、召回率( $R$ )、平均精度(Average precision, AP)、平均精度均值(mAP)和 F1

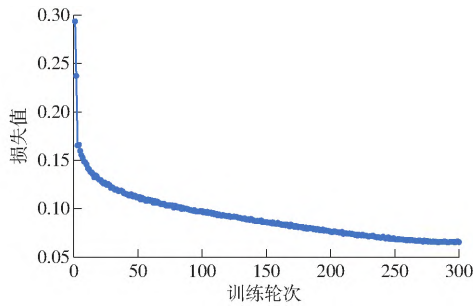


图 12 网络训练损失曲线  
Fig. 12 Network training loss curve

值(F1 Score) 客观评判标准来评估所训练出的模型针对苹果采摘方式的识别性能。

### 3.2.2 最优模型和预测类别阈值确定

识别模型在预测出目标所属类别的置信度之后需要依据预设的阈值对预测框进行筛选。基于同一个识别模型使用不同的置信度阈值进行预测,其识别结果的准确率和召回率是不同的。

若识别模型的置信度阈值选择不合适,会出现如图 13 所示的预测结果: 置信度阈值设置过低时会误将图像中较远种植行的小苹果识别出来(图 13a 中的黄色椭圆标识), 阈值设置过高时可能会漏掉当前种植行中的苹果目标(图 13b 中的黄色椭圆标识)。因此,需要结合具体的识别任务为模型确定合适的置信度阈值,以准确地筛选出需要识别的苹果目标。

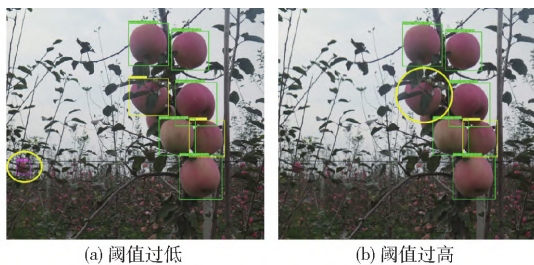


图 13 预测类别阈值对识别结果的影响

Fig. 13 Impact of confidence threshold on recognition result

基于训练所得的苹果采摘方式识别模型,通过调整置信度阈值,比较衡量模型在不同阈值下对测试集共 344 幅苹果树图像中的 6 类果实目标识别的准确率、召回率和 mAP 的变化,结合采摘机器人果实采摘方式识别任务的实际需求,确定模型的最佳预测类别阈值。经试验测试,不同置信度阈值下模型的识别准确率、召回率、mAP 曲线如图 14 所示。

对于面向采摘机器人的苹果采摘方式识别,需要对当前种植行(即采摘机械手可抓取范围内)中的果实进行识别,同时排除较远处他行苹果树上果实的干扰,因而在模型识别的准确率和召回率之间需要优先考虑准确率。另一方面,在选择阈值时需要辅助参考用于评估模型综合性能的指标 mAP,因

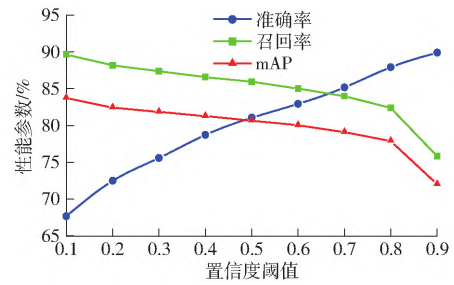


图 14 不同置信度阈值下模型性能变化曲线  
Fig. 14 Changes of performance of model with different confidence thresholds

其能同时兼顾准确率和召回率。

由图 14 可知,当置信度阈值小于 0.5 时,模型的识别准确率较低,不足 80%; 当置信度阈值高于 0.5 时,识别模型的 mAP 降低至 80% 以下; 因而综合考虑模型的识别准确率与 mAP, 当置信度阈值为 0.5 时,模型的性能表现最好,此时模型识别的准确率、召回率、mAP 分别为 81.0%、85.9% 和 80.7%。

## 4 结果与讨论

### 4.1 苹果采摘方式识别结果与分析

为了验证所设计的苹果采摘方式识别模型的性能,对该模型在测试集图像上的识别结果进行进一步分析。344 幅测试集图像中共有 6 536 个苹果目标,其中可直接采摘果实目标数量为 1 952 个,不可采摘果实目标数量为 1 202 个,上、下、左、右侧采摘果实的数量分别为 658、909、950、865 个。

模型的具体识别结果如表 2 所示,可以看出,针对可直接采摘果实、不可采摘果实和上、下、左、右侧采摘果实,本文所提出的模型对其识别的 AP 及 F1 值分别为 91.6%、90.0%、68.9%、73.5%、73.2%、82.5%、82.4%、83.3%、84.5%、84.3%、83.4%、84.3%。总体的识别准确率、召回率、mAP(6 类苹果采摘方式的平均精度)及 F1 值均在 80% 以上,分别为 81.0%、85.9%、80.7% 和 83.4%,基本满足苹果采摘方式识别精度要求。

表 2 基于改进 YOLOv5m 网络的苹果采摘方式识别结果

Tab. 2 Recognition results of apple picking pattern using improved YOLOv5m network

苹果采摘方式类别	数量/个	准确率/%	召回率/%	平均精度/%	F1 值/%
不可采摘	1 202	70.8	76.5	68.9	73.5
可直接采摘	1 952	84.9	95.7	91.6	90.0
上侧采摘	658	94.7	73.1	73.2	82.5
下侧采摘	909	80.1	86.7	82.4	83.3
左侧采摘	950	78.1	91.6	84.5	84.3
右侧采摘	865	77.8	91.9	83.4	84.3
总体	6 536	81.0	85.9	80.7	83.4

本文所提出的模型对不同天气与光照条件下的6种苹果采摘方式的识别结果示例如图15所示。其中,所识别出的不可采摘、可直接采摘和上、下、左、右侧采摘果实分别使用黄色、绿色和蓝色、粉色、深蓝色、红色框标识。由图15可以看出,所提出的识别模型不仅适用于阴天采集的光照均匀的图像,也适用于晴天光照条件下采集的图像,且对于顺光、侧光及逆光条件下的图像均能较好地识别出不同苹果采摘方式。

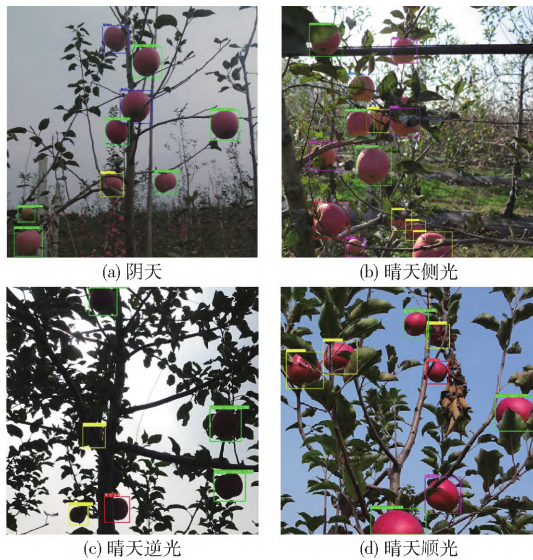


图15 基于改进YOLOv5m的苹果采摘方式识别示例  
Fig. 15 Recognition examples of apple picking pattern using improved YOLOv5m

#### 4.2 不同目标检测算法识别结果对比

为了进一步分析所提出的苹果采摘方式识别算法的性能,将改进的YOLOv5m网络与原YOLOv5m、YOLOv3、EfficientDet-D0网络在测试集图像上进行了识别结果对比。以mAP、平均识别速度等作为评价指标,各网络模型的识别结果、模型体积(占用存储空间量)及参数数量如表3所示。

表3 不同目标检测网络识别性能对比  
Tab. 3 Performance comparison of various target detection networks

目标检测网络	mAP/ %	单幅图像平均 识别时间/s	参数数量	模型体积/ MB
YOLOv5m	75.3	0.020	$2.15 \times 10^7$	41.3
YOLOv3	58.7	0.053	$6.15 \times 10^7$	235.0
EfficientDet-D0	60.1	0.040	$3.83 \times 10^6$	15.0
本文模型	80.7	0.025	$1.92 \times 10^7$	37.0

由表3可以看出,本文提出的改进YOLOv5m识别模型的mAP最高,与原YOLOv5m网络相比提高了5.4个百分点,与YOLOv3、EfficientDet-D0网络相比分别高出了22、20.6个百分点。针对模型的识别速度,本文提出的改进YOLOv5m识别模型在测试

集上单幅图像平均识别时间可达0.025 s(40 f/s),能够满足苹果采摘方式实时识别的要求,分别为EfficientDet-D0与YOLOv3网络识别时间的62.5%与47.17%。另一方面,由表3可以看出,本文所提出识别模型的体积为37 MB,为原始YOLOv5m模型的89.59%。说明所提出的网络在保证识别精度的同时,有效地实现了网络的轻量化。

综上,本文提出的模型在4种网络模型中具有最高的mAP。所提出模型的识别速度明显优于EfficientDet-D0与YOLOv3网络,虽然略低于原始的YOLOv5m网络,但识别平均帧率可达40 f/s,能够满足苹果采摘方式实时识别的需求。

4种模型分别针对阴天与晴天条件下的苹果采摘方式识别结果如图16所示,可以看出,本文提出的改进YOLOv5m网络的识别结果较准确,无误识与漏识现象。

针对阴天条件下的苹果采摘方式识别:EfficientDet-D0网络的识别结果较准确,无漏识与误识现象;YOLOv3网络的识别结果中出现了漏识别(图16d阴天图像中白色椭圆标识)和将右侧采摘果实识别为不可采摘果实的误识别现象(图16d阴天图像中黑色椭圆标识);YOLOv5m网络对当前种植行中苹果采摘方式的识别结果虽然较准确,但却过度识别了较多较远种植行中的苹果(如图16b阴天图像)。

针对晴天条件下的苹果采摘方式识别:YOLOv5m网络出现了将右侧、下侧采摘果实误识别为不可采摘果实现象(图16b晴天图像中黑色椭圆标识);EfficientDet-D0网络和YOLOv3网络均出现了漏识别现象(图16c晴天图像、图16d晴天图像中白色椭圆标识)和将不可采摘果实误识别为可采摘果实现象(图16c晴天图像、图16d晴天图像中黑色椭圆标识)。其中,EfficientDet-D0网络还出现了将右侧、上侧采摘果实误识别为不可采摘果实现象,YOLOv3网络还出现将不可采摘、右侧采摘果实误识别为左侧采摘果实现象(图16c晴天图像、图16d晴天图像中黑色椭圆标识)。

#### 4.3 与其他多分类苹果目标识别算法的对比

现有的苹果识别算法大多将苹果树上各种情况下的果实作为一类目标进行识别,对苹果目标多分类识别的研究较少。GAO等<sup>[17]</sup>将Faster R-CNN网络用于识别苹果树上不同条件下的共4类苹果目标,包括无遮挡、树叶遮挡、树枝/线遮挡和果实遮挡的苹果。YAN等<sup>[6]</sup>将YOLOv5s网络用于苹果识别中,将不同条件下的果实划分为可采摘(未被遮挡或仅被树叶遮挡的果实)与不可采摘

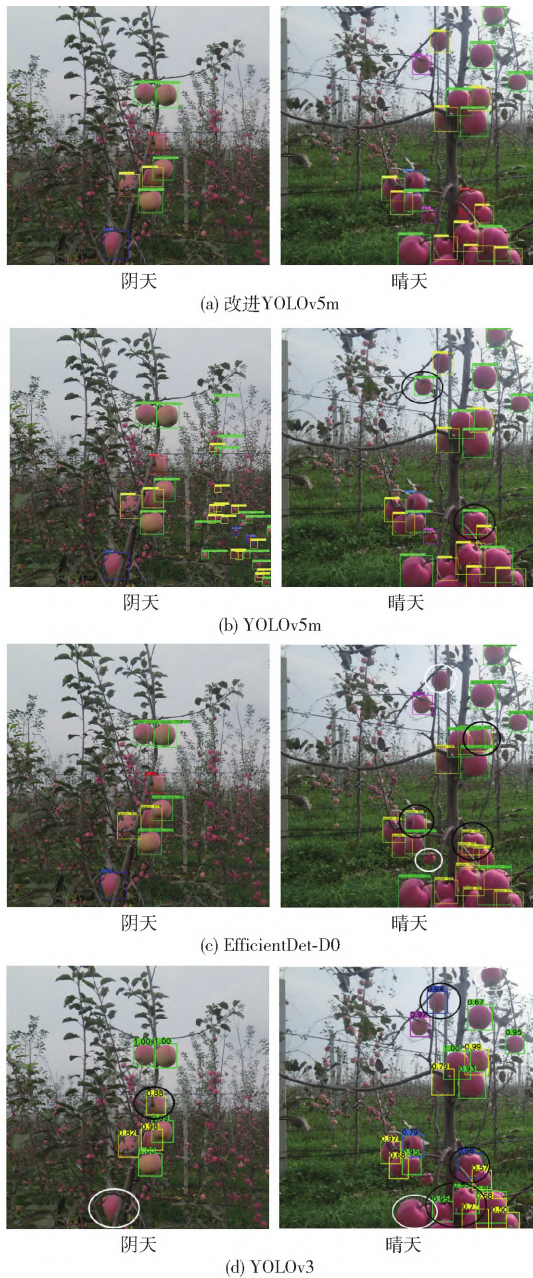


图 16 4 种网络模型的苹果采摘方式识别结果  
Fig. 16 Apple picking pattern recognition results of four network models

(被树枝干或被其他苹果遮挡的果实) 两类。为了验证所提出算法的识别性能,将本文算法与上述苹果检测算法的识别结果进行了比较,对比结果见表 4。

由表 4 可以看出,YAN 等<sup>[6]</sup>所提出的算法虽然在总体识别 mAP 与识别速度上优于本研究提出的算法,可以识别出被树枝干遮挡的苹果目标,以引导机器人避开对这些果实的采摘,但该算法却无法从视觉上引导机器人采摘被树枝干遮挡的苹果,从而导致一部分果实的漏采。而针对苹果采摘机器人,若根据苹果被枝干遮挡的不同情形而相应地改变机械手的采摘位姿,则可以实现对被枝干遮挡苹果的采摘。而本文所提出的识别方法实现了对不同苹果采摘方式的识别,从而可为机械手主动调整位姿以避免枝干对苹果的遮挡进行果实采摘提供视觉引导。

另一方面,GAO 等<sup>[17]</sup>所提出的算法也无法从视觉上引导机器人采摘被树枝干遮挡的苹果。该算法识别 4 类苹果的总 mAP 为 87.9%,比本文所提出识别方法的 mAP(80.7%)高 7.2 个百分点。但该文识别模型的体积较庞大,权重文件达到了 512 MB,而本文所提出的轻量化识别模型的体积为 37 MB,仅为该文模型的 7.23%,体积的轻量化有利于后期将模型部署在硬件设备中。另外,模型的识别速度对于机器人的采摘效率具有重要影响,该文模型的单帧图像识别时间为 0.241 s(帧率为 4.15 f/s),而本文模型的单帧图像识别时间仅为 0.025 s(帧率为 40 f/s),为该文模型识别速度的 9.64 倍,能够满足果实目标实时识别的需求。

## 5 结论

(1) 针对现有识别算法无法区分枝干单侧遮挡

表 4 与其他多分类苹果目标识别算法对比

Tab.4 Performance comparison with multi-class recognition method for apple

识别算法	总体 mAP/ %	平均精度均值/%						单幅图像 识别时 间/s	模型 体积/ MB
		可采摘		不可采摘					
		未被遮挡	树叶遮挡	枝干遮挡		果实重叠遮挡			
				可直接采摘	上侧 采摘	下侧 采摘	左侧 采摘		
改进 YOLOv5s	86.75	89.23		84.87				0.015	12.7
Faster R-CNN	87.90	90.90	89.90	85.8				0.241	512.0
本文模型	80.70	91.60	73.2	82.4	84.5	83.4	68.9	0.025	37.0

(即枝干遮挡苹果的上、下、左、右侧)、多侧遮挡情形下苹果目标的问题,提出了一种基于改进 YOLOv5m 面向采摘机器人的苹果采摘方式实时识

别方法,实现了对苹果树上不同果实采摘方式的识别,从而可为机械手主动调整位姿以避免树枝干的遮挡进行果实采摘提供视觉引导,以降低苹果的采

摘损失。

(2) 改进设计了 BottleneckCSP-B 特征提取模块并替换原 YOLOv5m 骨干网络中的 BottleneckCSP 模块,实现了原模块对图像深层特征提取能力的增强与骨干网络的轻量化改进;将 SE 模块嵌入到所改进设计的骨干网络中,更好地提取了不同苹果目标的特征;改进了原 YOLOv5m 架构中输入中等尺寸目标检测层的特征图跨接融合方式与网络的初始锚框尺寸,提升了当前种植行苹果目标的识别精度,避免了对图像里较远种植行苹果的识别。

(3) 所提出的改进网络模型可有效实现对图像

中可直接采摘、迂回采摘(苹果的上、下、左、右侧采摘)和不可采摘果实的识别,测试集试验结果表明,识别召回率为 85.9%,准确率为 81.0%,mAP 为 80.7%,F1 值为 83.4%,单幅(帧)图像平均识别时间为 0.025 s。

(4) 对比了所提出的改进 YOLOv5m 算法与原始 YOLOv5m、YOLOv3 和 EfficientDet-D0 算法在测试集图像上对 6 类苹果目标的识别效果,结果表明,所提出的改进算法与其他 3 种算法相比,识别的 mAP 分别高出了 5.4、22、20.6 个百分点。改进模型的体积压缩至原 YOLOv5m 模型的 89.6%。

### 参 考 文 献

- [1] FAN P, LANG G D, GUO P J, et al. Multi-feature patch-based segmentation technique in the gray-centered RGB color space for improved apple target recognition [J]. Agriculture, 2021, 11(3): 273.
- [2] FAN P, LANG G D, YAN B, et al. A method of segmenting apples based on gray-centered RGB color space [J]. Remote Sensing, 2021, 13(6): 1211.
- [3] FAN P, YAN B, WANG M R, et al. Three-finger grasp planning and experimental analysis of picking patterns for robotic apple harvesting [J]. Computers and Electronics in Agriculture, 2021, 188: 1-19.
- [4] FU L, GAO F, WU J, et al. Application of consumer RGB-D cameras for fruit detection and localization in field: a critical review [J]. Computers and Electronics in Agriculture, 2020, 177: 1-11.
- [5] ZHANG Z, IGATHINATHANE C, LI J, et al. Technology progress in mechanical harvest of fresh market apples [J]. Computers and Electronics in Agriculture, 2020, 175: 1-12.
- [6] YAN B, FAN P, LEI X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5 [J]. Remote Sensing, 2021, 13(9): 1619.
- [7] 刘莫尘, 高甜甜, 马宗旭, 等. 基于 MSRCR-YOLOv4-tiny 的田间环境玉米杂草检测模型 [J]. 农业机械学报, 2022, 53(2): 246-255, 335.  
LIU Mochen, GAO Tiantian, MA Zongxu, et al. Target detection model of corn weeds in field environment based on MSRCR algorithm and YOLOv4-tiny [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2): 246-255, 335. (in Chinese)
- [8] 王巧华, 顾伟, 蔡沛忠, 等. 基于改进 YOLO v4 的群体棉种双面破损检测方法 [J]. 农业机械学报, 2022, 53(1): 389-397.  
WANG Qiaohua, GU Wei, CAI Peizhong, et al. Detection method of double side breakage of population cotton seed based on improved YOLO v4 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(1): 389-397. (in Chinese)
- [9] 张宏鸣, 李永恒, 周利香, 等. 基于改进 YOLO v3 的肉牛多目标骨架提取 [J]. 农业机械学报, 2022, 53(3): 285-293.  
ZHANG Hongming, LI Yongheng, ZHOU Lixiang, et al. Multi-objective skeleton extraction method of beef cattle based on improved YOLO v3 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(3): 285-293. (in Chinese)
- [10] 张璐, 黄琳, 李备备, 等. 基于多尺度融合与无锚点 YOLO v3 的鱼群计数方法 [J]. 农业机械学报, 2021, 52(增刊): 237-244.  
ZHANG Lu, HUANG Lin, LI Beibei, et al. Fish school counting method based on multi-scale fusion and no anchor YOLO v3 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Supp.): 237-244. (in Chinese)
- [11] BRESILLA K, PERULLI G D, BOINI A, et al. Single-shot convolution neural networks for real-time fruit detection within the tree [J]. Frontiers in Plant Science, 2019, 10: 1-12.
- [12] 武星, 齐泽宇, 王龙军, 等. 基于轻量化 YOLO v3 卷积神经网络的苹果检测方法 [J]. 农业机械学报, 2020, 51(8): 17-25.  
WU Xing, QI Zeyu, WANG Longjun, et al. Apple detection method based on light-YOLO v3 convolutional neural network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 17-25. (in Chinese)
- [13] 赵德安, 吴任迪, 刘晓洋, 等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位 [J]. 农业工程学报, 2019, 35(3): 164-173.  
ZHAO Dean, WU Rendi, LIU Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background [J]. Transactions of the CSAE, 2019, 35(3): 164-173. (in Chinese)
- [14] 赵辉, 乔艳军, 王红君, 等. 基于改进 YOLO v3 的果园复杂环境下苹果果实识别 [J]. 农业工程学报, 2021, 37(16): 127-135.  
ZHAO Hui, QIAO Yanjun, WANG Hongjun, et al. Apple fruit recognition in complex orchard environment based on improved YOLO v3 [J]. Transactions of the CSAE, 2021, 37(16): 127-135. (in Chinese)
- [15] LU S, CHEN W, ZHANG X, et al. Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation [J]. Computers and Electronics in Agriculture, 2022, 193: 1-17.

- groove-tooth structure on high speed precision metering device of rapeseed [J]. Transactions of the CSAE, 2019, 35(1): 12 - 22. (in Chinese)
- [21] 崔涛, 刘佳, 杨丽, 等. 基于高速摄影的玉米种子滚动摩擦特性试验与仿真 [J]. 农业工程学报, 2013, 29(15): 34 - 41. CUI Tao, LIU Jia, YANG Li, et al. Experiment and simulation of rolling friction characteristic of corn seed based on high-speed photography [J]. Transactions of the CSAE, 2013, 29(15): 34 - 41. (in Chinese)
- [22] 王云霞, 梁志杰, 张东兴, 等. 基于离散元的玉米种子颗粒模型间接触参数标定 [J]. 农业工程学报, 2016, 32(22): 36 - 42. WANG Yunxia, LIANG Zhijie, ZHANG Dongxing, et al. Calibration method of contact characteristic parameters for corn seeds based on EDEM [J]. Transactions of the CSAE, 2016, 32(22): 36 - 42. (in Chinese)
- [23] CHEN Z, YU J, XUE D. An approach to and validation of maize-seed-assembly modelling based on the discrete element method [J]. Powder Technology, 2018, 328: 167 - 183.
- [24] 李婉宜, 曾攀, 雷丽萍, 等. 离散颗粒流动堆积行为离散元模拟及实验研究 [J]. 力学与实践, 2012, 34(1): 20 - 26. LI Wanyi, ZENG Pan, LEI Liping, et al. Simulation based on discrete element method and experiment on flow and packing behavior of particles [J]. Mechanics in Engineering, 2012, 34(1): 20 - 26. (in Chinese)
- [25] 史嵩, 张东兴, 杨丽, 等. 基于 EDEM 软件的气压组合孔式排种器充种性能模拟与验证 [J]. 农业工程学报, 2015, 31(3): 62 - 69. SHI Song, ZHANG Dongxing, YANG Li, et al. Simulation and verification of seed-filling performance of pneumatic-combined holes maize precision seed-metering device based on EDEM [J]. Transactions of the CSAE, 2015, 31(3): 62 - 69. (in Chinese)

~~~~~

(上接第 38 页)

- [16] FU L, MAJEED Y, ZHANG X, et al. Faster R - CNN-based apple detection in dense-foilage fruiting-wall trees using RGB and depth features for robotic harvesting [J]. Biosystems Engineering, 2020, 197: 245 - 256.
- [17] GAO F, FU L, ZHANG X, et al. Multi-class fruit-on-plant detection for apple in SNAP system using Faster R - CNN [J]. Computers and Electronics in Agriculture, 2020, 176: 1 - 10.
- [18] GENE-MOLA J, VILAPLANA V, ROSELL-POLO J R, et al. Multi-modal deep learning for Fuji apple detection using RGB - D cameras and their radiometric capabilities [J]. Computers and Electronics in Agriculture, 2019, 162: 689 - 698.
- [19] ZHANG J, KARKEE M, ZHANG Q, et al. Multi-class object detection using Faster R - CNN and estimation of shaking locations for automated shake-and-catch apple harvesting [J]. Computers and Electronics in Agriculture, 2020, 173: 1 - 10.
- [20] KANG H, CHEN C. Fruit detection, segmentation and 3D visualisation of environments in apple orchards [J]. Computers and Electronics in Agriculture, 2020, 171: 1 - 10.
- [21] 王丹丹, 何东健. 基于 R - FCN 深度卷积神经网络的机器人疏果前苹果目标的识别 [J]. 农业工程学报, 2019, 35(3): 156 - 163. WANG Dandan, HE Dongjian. Recognition of apple targets before fruits thinning by robot based on R - FCN deep convolution neural network [J]. Transactions of the CSAE, 2019, 35(3): 156 - 163. (in Chinese)
- [22] KANG H, CHEN C. Fast implementation of real-time fruit detection in apple orchards using deep learning [J]. Computers and Electronics in Agriculture, 2020, 168: 1 - 10.
- [23] 龙燕, 李南南, 高研, 等. 基于改进 FCOS 网络的自然环境下苹果检测 [J]. 农业工程学报, 2021, 37(12): 307 - 313. LONG Yan, LI Nannan, GAO Yan, et al. Apple fruit detection under natural condition using improved FCOS network [J]. Transactions of the CSAE, 2021, 37(12): 307 - 313. (in Chinese)
- [24] KANG H, CHEN C. Fruit detection and segmentation for apple harvesting using visual sensor in orchards [J]. Sensors, 2019, 19(20): 4599.
- [25] GENE-MOLA J, SANZ-CORTIELLA R, ROSELL-POLO J R, et al. Fruit detection and 3D location using instance segmentation neural networks and structure-from-motion photogrammetry [J]. Computers and Electronics in Agriculture, 2020, 169: 1 - 11.
- [26] ULTRALYTICS. YOLOv5 [R/OL]. 2020 - 05 - 18 [2022 - 03 - 01]. <https://github.com/ultralytics/yolov5>.
- [27] 宋怀波, 王亚男, 王云飞, 等. 基于 YOLOv5s 的自然场景油茶果识别方法 [J]. 农业机械学报, 2022, 53(7): 234 - 242. SONG Huaibo, WANG Ya'nan, WANG Yunfei, et al. *Camellia oleifera* fruit detection in natural scene based on YOLOv5s [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(7): 234 - 242. (in Chinese)
- [28] 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)
- [29] YAO J, QI J M, ZHANG J, et al. A real-time detection algorithm for kiwifruit defects based on YOLOv5 [J]. Electronics, 2021, 10(14): 1711.