

# 基于改进 YOLO v5 的苹果采摘机器人目标检测方法

胡仕林<sup>1</sup>, 陈伟<sup>1</sup>, 张境锋<sup>1</sup>, 魏庆宇<sup>1</sup>, 金学广<sup>2</sup>

(1.江苏科技大学 电子信息学院, 江苏 镇江 212003; 2.常州信息职业技术学院, 江苏 常州 213164)

**摘要:** 为实现复杂环境中苹果采摘机器人目标快速检测,克服传统 YOLO v5 网络结构复杂、计算性能弱的缺点,提出了一种基于深度可分离卷积 YOLO v5 的采摘机器人目标检测方法。在采集苹果样本图像并制作实验数据集后,进行模型训练和测试,引入深度可分离卷积 YOLO v5 网络对苹果图像进行特征提取,解决了网络中参数冗余问题,提高了采摘机器人的识别速度;采用 CIoU-Loss 损失函数和 DIoU-NMS 非极大值抑制方法,对损失函数进行优化,提升了机器人视觉系统对苹果的定位精度。机器人采摘试验结果表明:算法检测精度达 95.8%,检测速度达 53 帧/s,机器人单次采摘时间为 4.7s,采摘成功率达 93.9%。检测方法在减少模型参量的同时可保证检测精度和效率,具有较强的工程实用性。

**关键词:** 苹果采摘机器人; 目标检测; YOLOv5; 深度可分离卷积

**中图分类号:** S225; TP242

**文献标识码:** A

**文章编号:** 1003-188X(2024)06-0048-08

**DOI:**10.13427/j.cnki.njy.2024.06.009

## 0 引言

苹果采摘机器人可以提高农业自动化水平,降低劳动成本。目标果实的快速精准检测是苹果采摘机器人的关键技术之一,对提升农业生产力和采摘机器人的实用化程度具有重要的现实意义。

目前,国内外在机器人视觉检测和图像处理等方面展开了大量的探索。Kelman 等<sup>[1]</sup>通过分析果实图像的多个强度分布实现了重叠苹果的定位,准确率达到 93%,但计算过程耗时较长,实时性较差。Kapach 等<sup>[2]</sup>研究了苹果颜色检测方法,但对于未成熟苹果或有枝叶遮挡、苹果与背景相似等情况,算法检测效果并不理想。刘晓洋等<sup>[3]</sup>提出了一种基于超像素特征的果实分割方法,解决了苹果颜色分布不均的问题,但在光照强度变化时会影响目标区域的颜色和纹理信息特征,导致颜色信息失真,影响算法检测的准确性。吕继东等<sup>[4]</sup>对苹果果实进行 OTSU 动态阈值分割,但图像对于光照过于敏感,在跟踪目标苹果时会发生跟踪漂移而造成误差,无法保证苹果的完整和准确检测。

近年来,深度卷积神经网络在目标检测领域呈现出巨大的优势<sup>[5-7]</sup>。由于深度神经网络能够对图像高维特征实现深度提取,使得其在复杂情况下的苹果准确检测变成可能<sup>[8-11]</sup>。概括起来主要分为两种方法:一种是基于区域建议的方法,包括 R-CNN 方法<sup>[12]</sup>、Fast R-CNN 方法<sup>[13]</sup>和 Faster R-CNN 方法<sup>[14]</sup>等,核心思想是两阶段目标检测,即图像分类处理<sup>[15]</sup>。Faster R-CNN 模型虽然精度很高,泛化能力强,但是由于区域步骤消耗大量的计算资源,检测时间较长,因此满足不了实时性要求。另一种方法是无区域建议的方法,包括 SSD 算法<sup>[16]</sup>、YOLO 算法<sup>[17]</sup>,核心思想是基于整幅图像采用单卷积网络直接来预测目标的位置及其属性。

综上所述,现有方法对单一环境中的苹果图像检测具有较好的实时性和鲁棒性,但由于模型参数量大,导致苹果采摘机器人系统计算量大。因此,为了在真实复杂的果园环境下实现快速、精确的机器识别,基于深度可分离卷积 YOLO v5 网络结构,在减少模型参数量的同时进一步实现网络多层特征的复用和融合,对损失函数进行优化,使得在复杂果园环境下苹果检测速度和准确性得到了改善,从而进一步推动了国产苹果采摘机器人的实用化进程。

## 1 数据采集与预处理

### 1.1 实验数据采集

采集果园环境中的图像素材,以此进行训练和测试,从而遴选适用于苹果采摘机器人的算法参数。苹

**收稿日期:** 2022-06-20

**基金项目:** 江苏省现代农业项目(BE2020406);常州市科技支撑计划(农业)项目(CE20212025);镇江市国际科技合作项目(GJ202009)

**作者简介:** 胡仕林(1997-),男,安徽淮南人,硕士研究生,(E-mail) h1072226109@126.com。

**通讯作者:** 陈伟(1981-),男,江苏镇江人,副教授,硕士生导师,(E-mail) cwchenwei@aliyun.com。

果图像的采集地为江苏省扬中市雷公岛苹果示范基地(北纬 $32^{\circ}17'$ ,东经 $119^{\circ}43'$ ),使用数码相机等设备拍摄自然光照下的苹果,共计2700幅图像。

为保证苹果图像数据的复杂性,素材中包含不同果实个数、遮挡情况、重叠程度和光照条件的图像。图1为一组典型复杂环境下的苹果图像。除拍摄的图像素材外,还通过网络爬虫技术获取1300张苹果照片,总共2000张素材,包含4950个目标。



图1 复杂环境下的苹果图像

Fig.1 Apple image in complex environment

## 1.2 数据预处理

在对YOLO神经网络训练之前,需要对苹果图像进行预处理。为了更加全面、准确地采集目标场景中的苹果图片,需要对数据集进行扩充、衍生,让训练的模型可以更好地适应各种复杂环境中的苹果图像。首先,将苹果图像水平、垂直或随机旋转、放缩,改变图像的纵横比,组合更多数据来丰富扩展训练集,通过调整饱和度和色调、直方图均衡、中值滤波等图像处理技术来增强苹果图像数据;随机选择4组图片,通过Mosaic方法生成一组新的苹果图片,使算法在不损失检测准确率的同时提高其泛化能力。

最后,使用labelImg工具对目标果实的位置坐标、类别等信息进行框选标注,并将制作成标准的Pascal Voc数据集形式。利用MatLab,将共计10100幅苹果的数据集随机划分成训练集、验证集和测试集。为保证数据集的均匀分布,苹果图像按照70%、10%、20%的比例随机分为训练集、验证集和测试集。训练集样本7070个,验证集样本1010个,测试集样本2020个。

## 2 复杂环境下的苹果快速检测方法

苹果图像检测是采摘机器人进行采摘作业的信息来源,也是影响采摘成功率的重要因素。本文通过改进YOLO v5算法,实现采摘机器人在复杂环境下快速检测苹果,具有较高的检测速度与准确率。

### 2.1 基于YOLO v5的苹果检测

YOLO v5模型主要由Input、Neck、Backbone、Prediction 4部分组成。苹果Input图像主要由mosaic数据增强、自适应锚框计算和自适应图片缩放等对图像进行预处理。YOLO v5的主干特征提取网络是Backbone,形成对苹果图像特征的卷积神经网络,主要结构包含Focus(焦点层)、CSP(跨阶段局部网络层)和SPP(空间金字塔池化层)3部分。其中,Focus结构对图片进行切片操作,通过Concat层将其进行通道数的合并,增加苹果图像的特征数,可得到双倍信息的下苹果采样特征图<sup>[18]</sup>,再经过CBL层提取输入的不同特征,有助于找到特定的局部苹果图像特征;其次,通过Batch norm(批量规范层),实现结果归一化,使各个批量的差异不会过于明显;最后,用Leaky relu激活函数产生结果到下一层卷积。CSP进行卷积处理将图像梯度的变化集成到特征图中,这样可以减少模型尺寸,使通道数下降为原来的1/2,再经过Concat通道数的合并把需要引入的参数数量大幅降低,提高了模型的学习能力,有利于提升模型的准确度和训练速度。SPP提取苹果的特征信息,通过对特征层3次卷积后进行最大池化,扩大感受野,增强网络的非线性表达能力。Neck部分是对FPN(特征金字塔)结构的改进,解决了多尺度问题,加强了网络的特征融合和网络中推理信息的传输。YOLO v5整体网络结构如图2所示。

### 2.2 网络结构改进

对于苹果图像的多样化,适度优化网络结构可以提高采摘机器人识别效率,但网络结构参数也会不断增加,将使计算量庞大,占用多余资源。引入深度可分离卷积,在保证模型表达能力良好的前提下极大地降低了参数量。

为了使网络具有较高检测精度,并降低注意力机制对网络检测速度的影响,引入深度可分离卷积代替主干网络中的原有卷积操作,使神经网络执行空间卷

积,同时保持通道分离,再进行深度卷积,其结构如图3所示。

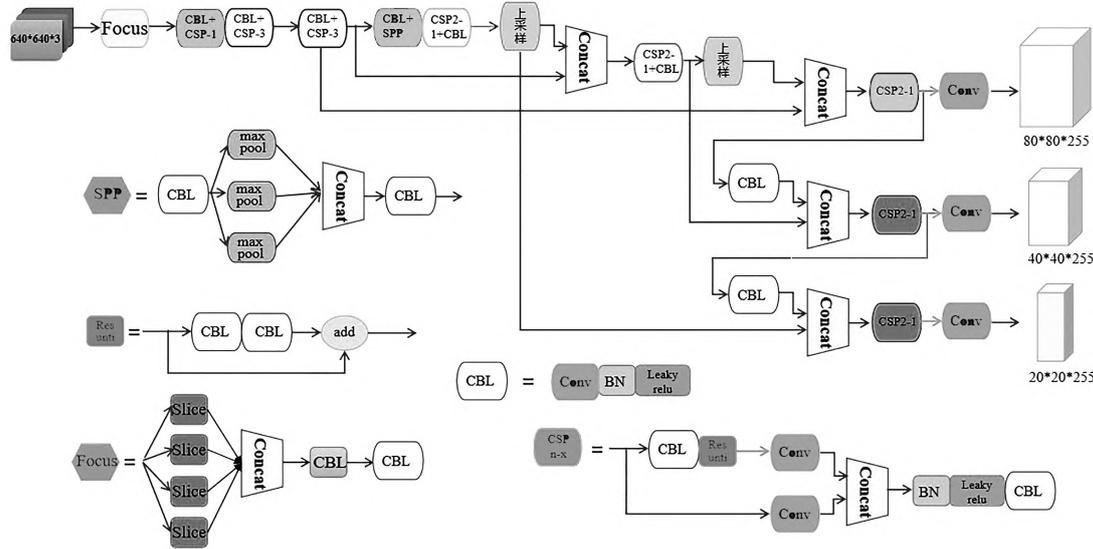


图2 YOLO v5 网络结构图及目标检测模型

Fig.2 YOLO v5 network structure diagram and target detection mod

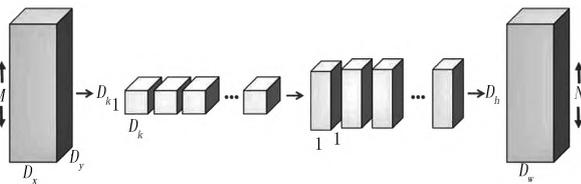


图3 深度可分离卷积结构

Fig.3 Depthwise Separable Convolutional Architecture

图3中, $D_x$ 为数据输入长度; $D_y$ 为数据输入宽度; $D_w$ 为数据输出长度; $D_h$ 为数据输出宽度; $D_k$ 为卷积核大小; $M$ 为数据输入通道数; $N$ 为数据输出通道数。网络中原有卷积主要是对通道特征图进行卷积操作,其计算量  $U_1$ 为

$$U_1 = D_k^2 \cdot M \cdot N \cdot D_w \cdot D_h \quad (3)$$

深度可分离卷积将卷积操作分解为1个  $3 \times 3$  的深度卷积和1个  $1 \times 1$  的逐点卷积。假设输入的特征图  $F$  为  $D_x \times D_y \times M$ ,经过深度卷积操作后得到  $D_h \times D_y \times M$  的特征图  $G$ ,其计算量  $U_2$ 为

$$U_2 = D_k^2 \cdot M \cdot D_w \cdot D_h + M \cdot N \cdot D_w \cdot D_h \quad (4)$$

由式(3)、式(4)可以推出深度可分离卷积与标准卷积之间的计算量比值为

$$\frac{U_2}{U_1} = \frac{1}{N} + \frac{1}{D_k^2} \quad (5)$$

通过引入深度可分离卷积可以将原网络的计算量降低和参数减少,从而能显著提高苹果图像检测速度。采用  $3 \times 3$  的卷积核,输入通道数为3,输出通道数

为256,网络总计算量是原标准卷积计算量的  $1/8$ 。

常规的YOLO v5网络参数量庞大,对重要程度不同的苹果特征采用相同的加权方式,并不利于对苹果特征的提取。因此,将YOLO v5特征提取网络中的标准卷积更换为深度可分离卷积,并且在深度可分离卷积中加入一个池化注意力模块(Pooling Block Attention Module, PBAM)。在池化注意力模块(PBAM)中,采用将苹果图像先压缩后扩张的方式,利用多尺度卷积核对输入图像模块进行处理。增强学习浅层特征采样的关键点,对不同池化得到的苹果特征进一步添加注意力,获取到苹果更全面、更重要的特征信息。为了增强特征的复用和提升深层网络中学习效率,在连续的卷积中引入了残差单元的结构;输出苹果特征图与输入苹果特征图分辨率保持一致,保证了网络在较深的情况下不会出梯度爆炸的问题,实现了更有效的融合高分辨率等特性。任意网络结构均可被池化注意力模块嵌入,具有结构优化、识别效率高优势。Concat通过建立候选区和特征层次之间的信息路径,达到苹果图像自适应识别特征的目的。深度可分离卷积YOLO v5网络选择不同的加权方式,来增强本尺度特征图中对苹果检测有用的信息,削减无效苹果特征,能够减少对网络结构的重复传递和运算。图4为改进后的YOLO v5网络结构图。其中,改进后的YOLO v5算法不仅可以缓解深度可分离卷积,减轻注意

力机制模块带来的计算压力,而且改进后的网络结构可以更快地计算将浅层特征图抽取转化为更多的深

层苹果特征图,苹果关键图像快速批量化处理,使网络的整体检测效果可以得到大幅提升。

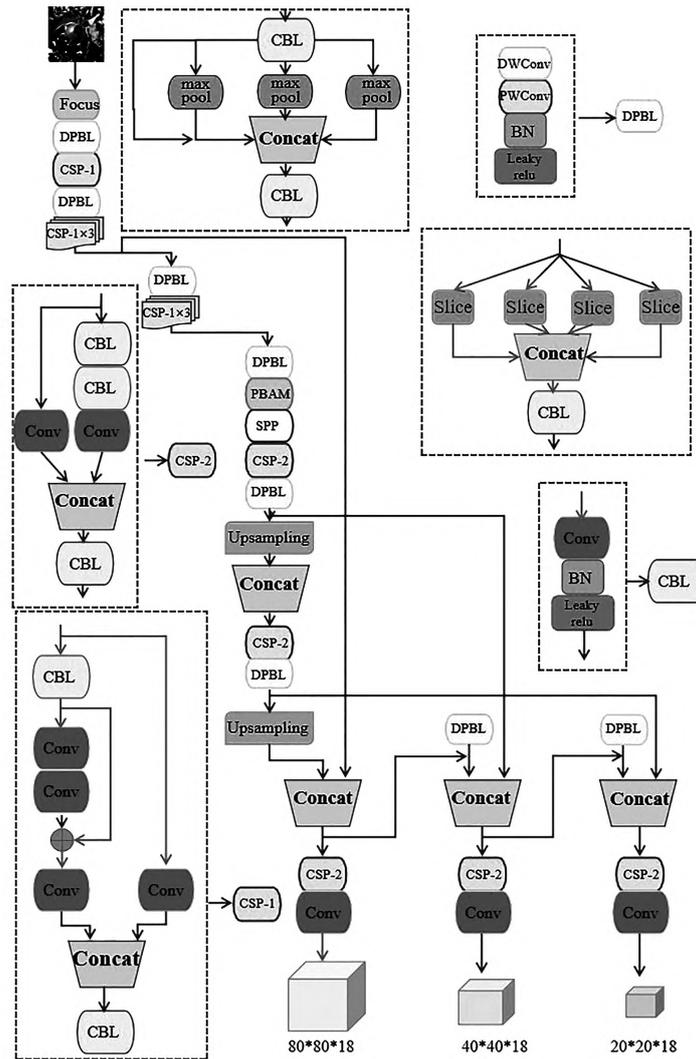


图4 深度可分离卷积YOLO v5的网络结构及目标检测模型

Fig.4 Network Structure and Target Detection Model of Depthwise Separable Convolution YOLO v5

### 2.3 损失函数优化

损失函数能够很好地反映模型和实际数据间的差异。GIoU-Loss 作为边界框回归损失函数,用以评判预测边界框和真实边界框的距离,即

$$GIoU = IoU - \frac{A^c - U}{A^c} \quad (6)$$

$$L_{GIoU} = 1 - GIoU \quad (7)$$

其中,  $IoU$  为预测边界框和真实边界框交集与并集的比值;  $A^c$  为将预测边界框和真实边界框同时包含的最小的矩形框的面积;  $U$  为预测边界框和真实边界框的并集;  $L_{GIoU}$  为  $GIoU$  损失。当有若干组相交的苹果目标框和苹果预测框、预测框在目标框内部且两框面积相近时,  $IoU$ -Loss 和  $GIoU$ -Loss 得到的结果是相同的,则此时损失函数会退化为  $IoU$ -Loss,那么苹果图像检测的准确性是无法保证的。

$CIoU$ -Loss 可充分考虑重叠面积、中心点距离、长宽比例的惩罚等因素。因此,更改  $CIoU$ -Loss 作为 YOLO v5 的损失函数,解决了  $GIoU$ -Loss 损失函数的不足,即

$$CIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \quad (8)$$

$$L_{CIoU} = 1 - CIoU \quad (9)$$

其中,  $b$  和  $b^{gt}$  为预测边界框和真实边界框的中心点;  $\rho^2$  为  $b$  和  $b^{gt}$  两个中心点之间的欧式距离;  $c$  为能同时覆盖预测边界框和真实边界框的最小矩形的对角线;  $\alpha$  为权衡长宽比例造成的损失的平衡因子;  $v$  为预测边界框和真实边界框的长度与宽度比例差值的归一化。  $\alpha$  和  $v$  的定义为

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (10)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (11)$$

其中,  $\omega^{gt}$ 、 $h^{gt}$  和  $\omega$ 、 $h$  分别为真实边界框和预测边界框的宽度和高度。CIoU-Loss 通过更多维度考虑真实边界框和预测边界框的差异,效果更好。

常规 NMS 在预测阶段,若目标周围有很多高置信度的预测框,则目标检测算法会输出多个预测框,常常会出现对苹果重叠以及遮挡情况的错误抑制。采用加权 NMS 移除重复的苹果图像,从而获取最优目标框,某个检测框与预测得分最高的检测框的交并比 IoU 作为评判指标。果园环境较为密集复杂,由于苹果与枝叶之间存在相互遮挡,且果实在气候条件不稳定或风力扰动作用时目标果实检测框会出现重合等不利状况,相邻苹果果实的中心点间距非常接近,重叠面积较多,则另外一个苹果的预测框很有可能会被滤除,导致苹果目标检测精度下降。而 DIoU-NMS 可以在两个检测框的 IoU 值接近或相同时会对预测框和真实框的中心点距离进行更深的计算筛查,能充分判定重叠率较高的两个边界框是否属于同一个果实,从而高效地对冗余的边界框进行抑制作用,可有效降低苹果的漏检率。DIoU-NMS 在 IoU 的基础上更加全面地考虑了两个边界框中心点的距离,则

$$DIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} \quad (12)$$

其中,  $b^{gt}$ 、 $b$ 、 $\rho^2$ 、 $c$  定义与(8)中一致。

DIoU-NMS 可以有效去除假阳性样本,非常有效地解决了果实和枝叶或果实和果实间的遮挡所导致的漏检问题,即本文使用 CIoU-Loss 和 DIoU-NMS 进行的优化能更有效地检测重叠苹果目标和不显著苹果目标,提高采摘机器人苹果检测精度。

### 3 实验结果与分析

#### 3.1 苹果检测算法实验

本次实验中,训练机核心处理器为 AMD 3900×3.8GHz CPU,显卡为 NVIDIA RTX3060Ti。采用 C++ 程序,调用 OpenCV、PyTorch 等操作库,来训练一种深度可分离卷积 YOLO v5 网络。模型训练时,将动量因子值设为 0.94,以避免模型陷入局部最优解或跳过最优解。将网络训练的前 300 轮学习率设置为 0.01,后 200 轮训练的学习率变为原来的 0.1 倍,将权重衰减正则项设置为 0.0005,以防止网络在训练过程中出现过拟合。为了验证算法优势,将提出算法与 Faster R-CNN 和 YOLO v5 进行对比,最后经过 500 轮的模型迭代训练,得到最优模型权重,结果如图 5 和图 6 所示。

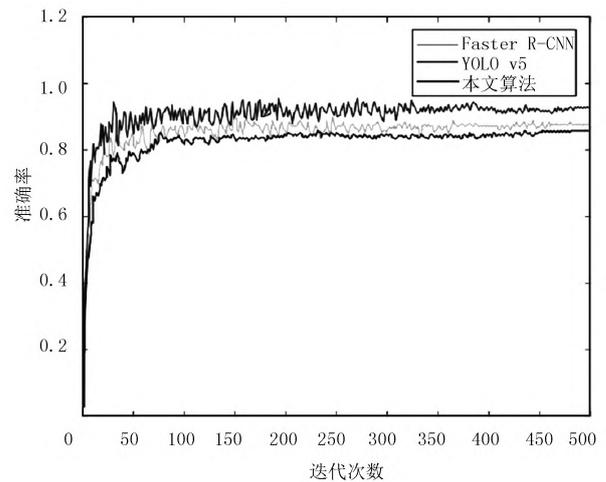


图 5 准确率变化曲线

Fig.5 Accuracy change curve

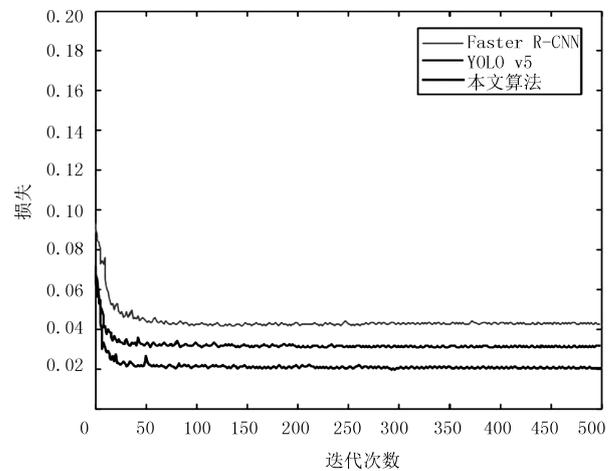


图 6 损失曲线

Fig.6 Loss curve

由图 5 可以看出:Faster R-CNN 和 YOLO v5 算法每轮迭代 10 次,迭代到 50 轮左右准确率才逐渐上升到 0.8 左右,并最终稳定在 0.85 和 0.82 左右;而本文算法迭代了 50 轮后准确率大致在 0.9 左右,并最终稳定在 0.91 左右。

由图 6 可以看出:Faster R-CNN 损失在迭代到 50 轮才逐渐降低到 0.05,最终稳定在 0.042 左右;YOLO v5 网络的初始损失为 0.07 左右,最终稳定在 0.03 左右;优化损失函数后,改进后网络损失值降低且收敛速度加快,初始损失值在 0.06 左右,最终稳定在 0.021。通过在 YOLO v5 算法结构中引入深度可分离卷积,利用该模块的通道注意力,使网络自动获取每个通道的重要程度,并将每个通道的信息与权重共享的多层感知器进行叠加,从而赋予每个通道不同的权重,来强化目标苹果特征,能准确识别不同大小和尺寸的同一种苹果。由仿真曲线可以看出,改进后的检测模型达到了非常理想的训练效果。

为了进一步验证深度可分离卷积 YOLO v5 算法的有效性,将其与多种检测算法进行对比实验。本文主要评估了 Faster R-CNN、YOLOv5 和深度可分离卷

积 YOLO v5 在相同条件下的检测效果。实验用采摘机器人对选取不同数目和大小的多目标苹果图像进行识别检测对比实验,结果如图 7 和表 1 所示。



图 7 不同检测算法对比

Fig.7 Comparison of Different Detection Algorithm

表 1 检测算法对比

Table 1 Comparison of Detection Algorithms

算法	平均精度值/%	检测速度/ $f \cdot s^{-1}$
Faster R-CNN	88.1	9
YOLOv5	87.9	51
基于深度可分离卷积 YOLOv5	95.8	53

由表 1 可以看出,深度可分离卷积 YOLO v5 算法在检测苹果方面的性能优于其他算法。当苹果较少时,几种算法检测结果相似,但基于深度可分离卷积 YOLO v5 的检测速度更快,平均精度值相对较高;当苹果较多时,虽然 YOLO v5 算法可以检测到的苹果数量更多,但在实际应用中距离太远的苹果是无效的采摘目标。综上所述,采用深度可分离卷积 YOLO v5 算法具有更快的检测速度,也具有较高的准确率,有利于提高苹果采摘机器人的作业效果和效率。

### 3.2 机器人采摘苹果实验

在实验室环境下,进行了苹果采摘机器人的目标检测和苹果采摘实验。实验的苹果采摘机器人如图 8 所示。



图 8 苹果采摘机器人样机图

Fig.8 Apple Picking Robot Prototype

采摘实验中,机器人首先对采集的苹果图像进行视觉处理,通过深度可分离卷积 YOLO v5 算法识别出图像中的苹果目标;然后,计算出目标在机械臂坐标系中的位置,通过视觉伺服控制算法控制抓手向苹果目标方向移动,从而完成机器人采摘任务。

图9为机器人进行采摘作业的完整流程。本实验使用果树模型道具模拟采摘作业环境,总共进行了250余次采摘实验,平均单次采摘时间为4.7s,综合采摘成功率为93.9%,满足采摘机器人的实时性采摘要求。

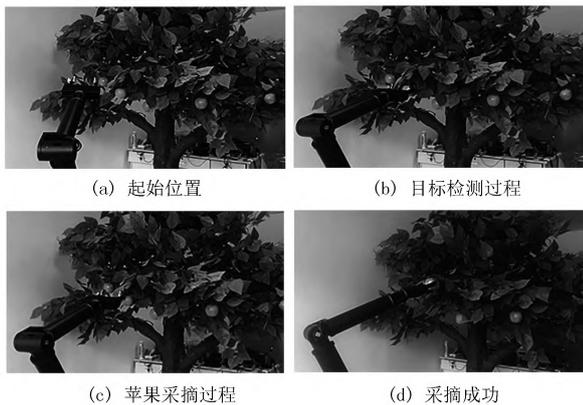


图9 机器人采摘实验

Fig.9 Harvesting Experiment of Robot

## 4 结论

1) 为提高采摘机器人的检测速度,对 YOLO v5 网络结构进行了改进。通过与深度可分离卷积结合,提出了一种深度可分离卷积 YOLO v5 网络结构,有效减少了原网络的参数,并提高了网络提取苹果图像特征的能力,使改进后的网络具有更快的检测速度。

2) 针对采集数据中存在网络检测准确度不高问题,CIoU-Loss 和 DIoU-NMS 进行的优化能更有效地检测重叠苹果目标和不显著苹果目标,可以减少苹果误检的概率,使得整个苹果识别网络检测精度得到很好的改善。

3) 通过自制的苹果数据集对所提算法进行了测试。深度可分离卷积 YOLO v5 算法对苹果检测的平均精度值为 95.8%,检测速度为每秒 53 帧;对比 Faster R-CNN 等算法,深度可分离卷积 YOLOv5 可以同时满足苹果检测的精度和速度要求。

4) 在实验室环境下进行了采摘苹果实验,结果表明:处理单帧图像的时间为 0.4s,单次采摘时间为 4.7s,综合采摘成功率为 93.9%。

### 参考文献:

- [1] KELMAN E E, LINKER R. Vision-based localisation of mature apples in tree images using convexity[J]. *Biosystems engineering*, 2018, 11(8): 174-185.
- [2] KAPACH K, BARNEA E, MAIRON R, et al. Computer vision for fruit harvesting robots—state of the art and challenges ahead[J]. *International journal of computational vision and robotics*, 2018, 3(2): 4-34.
- [3] 刘晓洋, 赵德安, 贾伟宽, 等. 基于超像素特征的苹果采摘机器人果实分割方法[J]. *农业机械学报*, 2019, 50(3): 15-23.
- [4] 吕继东, 赵德安. 苹果采摘机器人目标果实快速跟踪识别方法[J]. *农业机械学报*, 2017, 45(1): 65-72.
- [5] 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]. *Journal of technology and science*, 2020, 14(11): 214-223.
- [6] YING B Y, XU Y C, ZHANG S, et al. Weed detection in images of carrot fields based on improved YOLO v4[J]. *Traitement du signal*, 2021, 38(2): 384-418.
- [7] RAKUN J, STAJNKO D, ZAZULA D. Detecting fruits in natural scenes by using spatial-frequency based texture analysis and multiview geometry[J]. *Computers and electronics in agriculture*, 2022, 76(1): 80-88.
- [8] AGELOPOLOU A D, BOCHTIS D, FOUNTAS, et al. Yield prediction in apple orchards based on image processing[J]. *Precision agriculture*, 2021, 12(3): 448-456.
- [9] D Wu, S Lv, M Jiang. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments[J]. *Computers and electronics in agriculture*, 2021, 24(6): 737-739.
- [10] WEI JI, XIANGLI MENG, YUN TAO, et al. Fast segmentation of colour apple image under all-weather conditions for vision recognition of picking robots[J]. *International journal of advanced robotic systems*, 2019, 13(3): 86-87.
- [11] E K NYARKO, I VIDOVIC, K Cupec. A nearest neighbor approach for fruit recognition in RGB-D images based on detection of convex surfaces[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii: AAAI Press, 2020: 893-907.
- [12] MEHMET TURAN, YASIN ALMALIOGLU, HELDER ARAUJO, et al. Deep EndoVO: a recurrent convolutional neural network based visual odometry approach for endoscopic capsule robots[J]. *Neurocomputing*, 2019, 15(4): 273-281.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2019: 580-587.
- [14] GAO F F, FU L S, ZHANG X, et al. Multi-class fruit

- plant detection for apple in SNAP system using faster R-CNN[J]. Computers and electronics in agriculture, 2021, 17(6): 152-161.
- [15] S. KHAN, H. RAHMANI, AND S. A. shah. A guide to convolutional neural networks for computer vision[J]. Synthesis lectures on computer vision, 2022, 17(8): 112-124.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision. Amsterdam: Springer, Cham, 2021: 21-37.
- [17] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the 2019 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2019: 779-788.
- [18] L FU. Fast and accurate detection of kiwifruit in orchard using improved YOLOv4-tiny model[J]. Precision agriculture, 2020, 38(3): 68-69.

## Target Detection Method of Apple Harvesting Robot Based on Improved YOLO v5

Hu Shilin<sup>1</sup>, Chen Wei<sup>1</sup>, Zhang Jingfeng<sup>1</sup>, Wei Qingyu<sup>1</sup>, Jin Xueguang<sup>2</sup>

(1. School of Electronic and Information, Jiangsu University of Science and Technology, Zhenjiang 212003, China; 2. Changzhou Information Vocational and Technical College, Changzhou 213164, China)

**Abstract:** In order to achieve rapid detection of apple picking robot targets in complex environments, and overcome the shortcomings of traditional YOLO v5 network structure and weak computing performance, a target detection method based on depthwise separable convolution YOLO v5 is proposed. After collecting apple sample images and making experimental data sets, model training and testing are performed, and a depthwise separable convolutional YOLO v5 network is introduced to extract features from apple images, which solves the problem of parameter redundancy in the network and improves the recognition speed of picking robots; The CIoU-Loss loss function and the DIoU-NMS non-maximum suppression method are used to optimize the loss function and improve the positioning accuracy of the robot vision system for apples. The results of the robot picking experiment show that the detection accuracy of the algorithm is 95.8%, the detection speed is 53 frames per second, the single picking time of the robot is 4.7s, and the picking success rate is 93.9%. The detection method can ensure detection accuracy and efficiency while reducing model parameters, and has strong engineering practicability.

**Key words:** apple picking robot; object detection; YOLO v5; depthwise separable convolution