

基于 CenterNet 的密集场景下多苹果目标快速识别方法

杨福增^{1,2} 雷小燕^{1,2} 刘志杰^{1,2} 樊攀^{1,2} 闫彬^{1,3}

(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)

摘要: 为提高苹果采摘机器人的识别效率和环境适应性,使其能在密集场景下对多苹果目标进行快速、精确识别,提出了一种密集场景下多苹果目标的快速识别方法。该方法借鉴“点即是目标”的思路,通过预测苹果的中心点及该苹果的宽、高尺寸,实现苹果目标的快速识别;通过改进 CenterNet 网络,设计了 Tiny Hourglass-24 轻量级骨干网络,同时优化残差模块提高了目标识别速度。试验结果表明,该方法在非密集场景下(即近距离场景)测试集的识别平均精度(Average precision, AP)为 98.90%, F1 值为 96.39%;在密集场景下(即远距离场景)测试集的识别平均精度为 93.63%, F1 值为 92.91%,单幅图像平均识别时间为 0.069 s。通过与 YOLO v3、CornerNet-Lite 网络在两类测试集下的识别效果进行对比,该方法在密集场景测试集上比 YOLO v3 和 CornerNet-Lite 网络的平均精度分别提高了 4.13、29.03 个百分点;单幅图像平均识别时间比 YOLO v3 减少 0.04 s、比 CornerNet-Lite 减少 0.646 s。该方法无需使用锚框(Anchor box)和非极大值抑制后处理,可为苹果采摘机器人在密集场景下快速准确识别多苹果目标提供技术支撑。

关键词: 采摘机器人; 密集场景; 多苹果识别; Tiny Hourglass-24; CenterNet

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2022)02-0265-09

OSID:



Fast Recognition Method for Multiple Apple Targets in Dense Scenes Based on CenterNet

YANG Fuzeng^{1,2} LEI Xiaoyan^{1,2} LIU Zhijie^{1,2} FAN Pan^{1,2} YAN Bin^{1,3}

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
2. Apple Full Mechanized Scientific Research Base, 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)

Abstract: In order to improve the recognition efficiency and environmental adaptability of the apple picking robot, so that it can quickly and accurately recognize multiple apple targets in dense scenes, a rapid recognition method for multiple apple targets in dense scenes was proposed. The method drew on the idea of “point is the target”, and realized the rapid identification of apple targets by predicting the center point of apple and the width and height of apple. By improving the CenterNet network, the Tiny Hourglass-24 lightweight backbone network was designed, and the residual module was optimized to improve the target recognition speed. The test results showed that the average recognition accuracy of this method on the test set in non-dense scenes (images taken in close-range scenes) was 98.90%, and F1 was 96.39%. In the dense scene (images taken in the remote scene), the recognition average precision (AP) of the test set was 93.63%, the F1 was 92.91%, and the average recognition time of a single image was 0.069 s. By comparing with the recognition effect of YOLO v3 and CornerNet-Lite network under the two types of test sets, the AP of this method was increased by 4.13 percentage points and 29.03 percentage points respectively on the dense scene test set. The average image recognition time was 0.04 s faster than that of YOLO v3 and 0.646 s faster than that of CornerNet-Lite. This method did not need to use anchor box and non-maximum suppression post-processing, and can provide technical support for the apple picking robot to quickly and accurately identify multiple apple targets in dense scenes.

Key words: picking robot; dense scene; multiple apple recognition; Tiny Hourglass-24; CenterNet

收稿日期: 2021-01-27 修回日期: 2021-02-21

基金项目: 陕西省科技重大专项(2020zdxx03-04-01)和国家重点研发计划项目(2016YFD0700503)

作者简介: 杨福增(1966—),男,教授,博士生导师,主要从事基于无人机械的信息与集成技术研究,E-mail: yfz0701@163.com

0 引言

我国是世界上最大的苹果生产国和消费国,但苹果的采摘仍以传统的人工作业为主,人工采收不仅劳动强度大,且随着农村劳动力的减少,出现了用工荒的现象,亟需研发苹果采摘机器人以降低对人工的过度依赖^[1-2]。苹果目标的准确快速识别是机器人实现自主采摘的重要前提^[3-5]。在果园环境中,密集场景(即远距离场景)指远距离拍摄的单幅图像,该图像包含大量的果实,且图像中普遍存在果实重叠、遮挡的现象^[6-7]。实现密集场景下多苹果目标的准确快速识别,对提高采摘机器人的识别效率、进而实现苹果的智能化采摘至关重要。

传统的果实识别方法主要基于果实的颜色、纹理和形状特征等^[8-11],这类方法对单个果实或相邻果实识别效果较好,但对果园中互相重叠或受枝条、树叶遮挡的果实识别准确率明显降低^[12]。近年来卷积神经网络(Convolutional neural networks, CNN)在目标检测方面性能卓越,已经被广泛应用于果实识别^[13-15]。果实识别从原理上可划分为两类:一类需要将果实的识别问题分为两步,先通过网络产生可能包含果实的候选区域,再对候选区域进行分类,这类网络的识别时间普遍较长^[16-17]。另一类方法直接通过 CNN 给出果实的置信度和位置坐标,其特点是能提高果实的识别速度,但同样使用锚框(Anchor box)导致遮挡类果实的识别精度降低^[18-19]。上述两类算法均是基于锚框的识别算法,针对不同的数据集,需要进行不同大小、多个数量的锚框设定,由于锚框尺寸固定,导致网络对不同尺寸目标识别的通用性降低,对重叠、遮挡的小目标果实可能无法匹配到锚框,从而导致漏检。在预测阶段需要通过非极大值抑制(Non-maximum suppression, NMS)删除重复的候选框,使得目标识别时间更长。

近两年,基于无锚框(Anchor free)的目标识别思路开始兴起,无锚框是基于预测目标关键点的识别算法,如 CenterNet^[20]、CornerNet - Lite^[21]、ExtremeNet^[22]等在人体姿态检测上得到成功应用。CenterNet 通过预测每个目标的中心点,并预测其宽、高尺寸实现目标识别。CornerNet - Lite 网络通过预测目标的左上角点和右下角点,两个关键点进行组合确定目标。ExtremeNet 网络通过预测一幅图像每个目标的上、下、左、右 4 个极点和中心点,通过几何关系对预测的关键点进行分组实现目标识别。与 CornerNet 等关键点预测模型相比,CenterNet 只预测目标的中心点,没有预测点的分组问题,不需要 NMS 抑制冗余框,是无锚框领域识别精度和速度较优的算法。

在密集场景下,图像中的苹果尺寸较小,且苹果间互相重叠遮挡、苹果被枝条和树叶遮挡等现象严重,这些因素给苹果目标的准确识别带来困难。为实现密集场景中多苹果目标的准确识别,本文通过改进基于无锚框的 CenterNet 模型,设计 Tiny Hourglass - 24 轻量级骨干网络,同时优化残差模块进而提高苹果的识别速度,实现密集场景下多苹果目标的准确快速识别。

1 试验数据

1.1 图像采集

图像采集地点为陕西省乾县农业科技试验示范基地的标准化现代果园,采集设备为索尼 FDR - AX45 型相机。相机与树冠距离为 0.5 ~ 2 m,选择多个角度和距离进行拍摄,采集时间为白天,包括树叶遮挡、枝条遮挡、果实间重叠遮挡、顺光和逆光等环境下的图像,最终采集苹果图像 1 852 幅。部分采集图像如图 1 所示,可以看到在密集场景中(图 1b ~ 1d),果园苹果呈多簇密集生长的情况,存在大量重叠、遮挡的现象,导致密集场景下的苹果目标识别存在较大难度。



图 1 果园环境下的苹果图像

Fig. 1 Images of apples in an orchard environment

1.2 数据集制作

利用 Labellmg 标注软件绘制苹果矩形框,标注

后生成相对应的 xml 文件,其中标注信息包含每个苹果目标真实的左上角点和右下角点像素坐标,数

据集保存为 PASCAL VOC 格式。为了验证在密集场景下算法的识别性能,将标注后随机选择的 852 幅测试集分为 2 类,其中测试集 A 为近距离场景下的苹果,共 397 幅图像,包含 3 542 个苹果,平均一幅图像包含 8.92 个苹果目标。测试集 B 为摄像机远距离拍摄的密集场景下的图像,共 455 幅图像,包含 11 578 个苹果,平均每幅图像有 25.45 个苹果目标,其果实在图像中的尺寸较小,遮挡现象较为普遍,两类测试集上均包含不同光照情况下的图像。

由于深度学习需要大量数据集提升网络的识别性能,对剩余的 1 000 幅图像进行亮度、锐度、色度的增强和减弱,加入 0.01 的高斯噪声、进行图像翻转等数据增强方式来丰富数据集,最终共生成图像 28 020 幅用于网络训练。

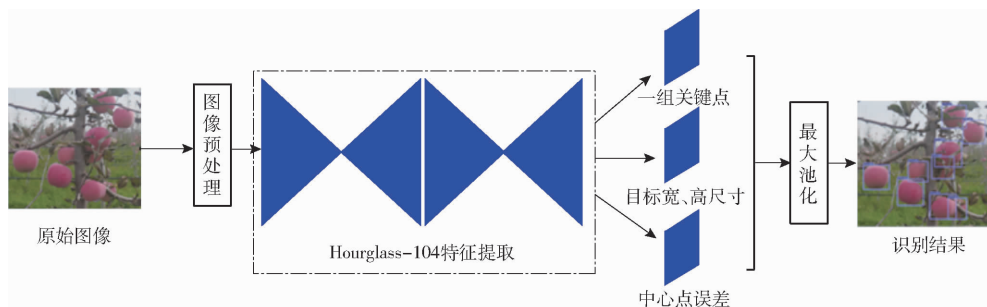


图2 CenterNet 网络识别流程

Fig. 2 Progress of CenterNet recognition

2.2 苹果中心点预测

CenterNet 预测苹果目标中心点的过程分为 3 步:首先根据标注保存的 xml 文件内的像素坐标信息,求取每个苹果真实的中心点坐标,将真实中心点位置的得分置为 1;其次,考虑到苹果真实中心点附近的区域也近似代表苹果目标,通过高斯核函数 Y_{xyc} 在真实中心点附近生成一个二维的正态分布,使真实的苹果中心点附近的区域值为 0~1 之间的数;最后,通过网络训练预测出一组苹果可能的中心点,预测的中心点得分 $\hat{Y}_{xyc}$ 在 0~1 之间,预测的中心点和真实中心点之间的最小误差为 L_k ,误差函数采用改进的 Focal loss^[23],即

$$L_k =$$

$$\begin{cases} -\frac{1}{N} \sum_{xyc} (1 - \hat{Y}_{xyc})^\alpha \lg \hat{Y}_{xyc} & (Y_{xyc} = 1) \\ -\frac{1}{N} \sum_{xyc} (1 - Y_{xyc})^\beta \hat{Y}_{xyc} \lg (1 - \hat{Y}_{xyc}) & (Y_{xyc} \neq 1) \end{cases} \quad (1)$$

$$\text{其中 } Y_{xyc} = \exp \left(-\frac{(x - \tilde{p}_x)^2 + (y - \tilde{p}_y)^2}{2\sigma_p^2} \right) \quad (2)$$

式中 N ——单幅图像中目标的数量

x, y ——4 倍下采样后目标真实的中心点坐标

c ——识别目标的类别, c 取 1

2 目标快速识别方法

2.1 CenterNet 网络

CenterNet 网络比基于锚框的识别算法更简单,更快且更准确^[20]。CenterNet 采用“点即是目标”的思路,通过寻找待识别物体的中心点以及宽、高来确定目标。CenterNet 网络首先对输入图像预处理,4 倍下采样降低图像分辨率;然后通过骨干网络 Hourglass-104 进行特征提取,产生一组关键点、目标宽高尺寸以及中心点误差;最后用最大池化(Max pooling)筛选保留的关键点,在关键点附近产生预测框。CenterNet 网络识别流程如图 2 所示。网络总的输出通道数为 $C + 4$, C 为识别目标的类别数,4 为通道数,表示目标的宽、高尺寸以及目标中心点的横向和纵向坐标误差。

α, β ——每个点贡献的超参数, α 取 2, β 取 4

$\tilde{p}_x, \tilde{p}_y$ ——4 倍下采样后预测的中心点坐标

σ_p ——自适应苹果目标的标准差

在进入骨干网络训练之前,首先对图像进行 4 倍下采样预处理,真实的中心点坐标在缩小 4 倍后会产生量化误差,因此需要对每一个中心点预测中心点偏移量 $\hat{O}_p$,偏移损失 L_{off} 计算公式为

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \hat{O}_p - \left(\frac{p}{4} - \left\lfloor \frac{p}{4} \right\rfloor \right) \right| \quad (3)$$

其中

$$\hat{O}_p = (\delta \hat{x}_{i_p}, \delta \hat{y}_{i_p}) \quad (4)$$

$$p = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right) \quad (5)$$

式中 x_1, y_1 ——苹果目标左上角点像素坐标

x_2, y_2 ——苹果目标右下角点像素坐标

p ——目标真实的中心点

$\left\lfloor \frac{p}{4} \right\rfloor$ ——4 倍下采样后预测的中心点

$\delta \hat{x}_{i_p}, \delta \hat{y}_{i_p}$ —— x, y 坐标的位置偏移量

在预测阶段,对输出的特征图用 3×3 最大池化提取特征图上的中心点,选出特征图上大于或等于周围 8 个相邻值的点,作为初步预测的中心点,对预测的点按照置信度由高到低进行排序,选取前 50 个

点作为预测的中心点,保留置信度大于 0.3 的点作为最终预测的中心点。

2.3 苹果宽、高尺寸预测

网络除了要预测苹果的中心点,也要预测每个中心点对应的宽和高。CenterNet 计算出图像中苹果的真实宽度和高度 s ,通过网络训练预测出苹果目标的宽、高尺寸 $\hat{S}$ 。尺寸误差 L_{size} 的计算公式为

$$L_{size} = \frac{1}{N} \sum_{k=1}^N |\hat{S} - s| \tag{6}$$

其中 $s = (x_2 - x_1, y_2 - y_1)$ (7)

训练的损失函数由 3 部分构成,分别是预测的关键点损失 L_k 、预测框的尺寸损失 L_{size} 以及中心点的偏移损失 L_{off} ,总体目标损失函数 L_{det} 的公式为

$$L_{det} = L_k + \alpha_{off} L_{off} + \beta_{size} L_{size} \tag{8}$$

式中 α_{off} ——偏移损失权重,取为 1

β_{size} ——尺寸损失权重,取为 0.1

每个中心点附近对应的预测框为 $(\hat{x}_{i_p} + \delta\hat{x}_{i_p} - \hat{w}_{i_p}/2, \hat{y}_{i_p} + \delta\hat{y}_{i_p} - \hat{h}_{i_p}/2, \hat{x}_{i_p} + \delta\hat{x}_{i_p} + \hat{w}_{i_p}/2, \hat{y}_{i_p} + \delta\hat{y}_{i_p} + \hat{h}_{i_p}/2)$ 。其中, $\hat{x}_{i_p}$ 、 $\hat{y}_{i_p}$ 为网络预测的中心点坐标, $\hat{w}_{i_p}$ 、 $\hat{h}_{i_p}$ 为网络预测的苹果的宽和高。

2.4 CenterNet 网络的改进设计

骨干网络是一个卷积神经网络的重要组成部分,其主要对目标的特征进行提取。CenterNet 原型使用了 Hourglass - 104 骨干网络进行特征提取,Hourglass 最先提出用于复杂的人体姿态估计^[23],该网络由一个或多个沙漏模块组成,每个沙漏模块通过下采样和上采样的方式对输入图像进行特征提取,上、下采样过程均采用多个残差(Residual)模块,网络深度为 104 层,其结构复杂庞大,且受巨大参数量制约,需要多块高性能显卡带动训练,网络运行速度很慢。

鉴于本研究的识别为单一的苹果类别,浅层网络也可以实现苹果目标的特征提取^[19],因此本文改进设计了轻量化的 CenterNet 骨干网络,以提高苹果的识别速度。考虑到密集场景下苹果多为小目标,特征图分辨率过小可能丢失小目标苹果的信息,因此在骨干网络中,本文对图像仅进行 3 倍下采样降低特征图分辨率,再通过 3 倍上采样恢复特征图的分辨率。为了降低网络参数量,提升苹果目标识别的速度,对 Hourglass - 104 的单个 Residual 模块进行了优化,将原有 Residual 模块第 1 层的 3×3 卷积的通道数压缩 2 倍,将第 2 层的 3×3 卷积替换为 1×1 的分组卷积^[24],改进后的 G - Residual 模块改进设计如图 3 所示。图 3 中 n 为 Residual 模块的通道数量。

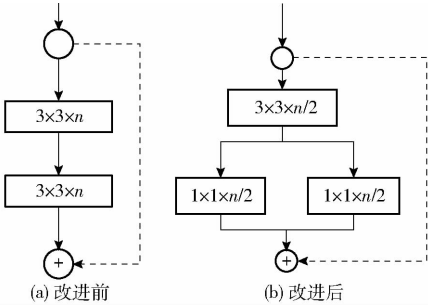


图 3 Residual 模块改进结果

Fig. 3 Residual module improvement result

重新设计的沙漏模块命名为 Tiny Hourglass,由多个 G - Residual 组成,单个 Tiny Hourglass 模块由下采样层、上采样层、跳跃连接层构成,每层均使用 1 个 G - Residual 模块。下采样层对特征图进行 3 倍下采样提取不同尺度下的特征,3 倍下采样后特征图的分辨率降为 16 像素 $\times$ 16 像素,再通过上采样层 3 倍上采样进行特征提取,恢复图像的分辨率。输出特征图的分辨率为 128 像素 $\times$ 128 像素,大分辨率的特征图有利于目标的准确识别。在每个下采样层和上采样层之间,添加跳跃连接层弥补下采样丢失的目标信息。改进的骨干网络由 2 个相同的 Tiny Hourglass 模块堆叠而成,整个骨干网络为全卷积网络,网络深度为 24 层,2 个 Tiny Hourglass 模块采用中间监督^[23] (Intermediate supervision) 方式,将第 1 个 Tiny Hourglass 模块输出的特征图和输入特征图作为第 2 个 Tiny Hourglass 模块的输入,基于 Tiny Hourglass - 24 骨干网络的 CenterNet 网络结构如图 4 所示,图 4 中 A、B、C 为跳跃连接层。

3 网络训练

3.1 试验运行平台

本试验所运行的硬件平台为 T7920 塔式工作站,处理器为 Intel(R) Xeon(R) 5118 CPU, 2.30、2.29 GHz (双处理器), 32 GB 运行内存,显卡为 Nvidia RTX2080TI - 11G。在 Windows 10 操作系统上运行,借助 Pytorch1.2 框架、CUDA 10.0、CUDNN 7.3 软件工具搭建运行环境。

3.2 网络训练

网络训练的批处理量为 68,初始学习率为 0.000 25,网络训练共迭代 100 轮,学习率在 60 轮和 90 轮时分别下降 10 倍。训练前对图像进行标准化和归一化处理,使用整个网络最后一层的特征图来进行预测。网络训练的损失值如图 5 所示。由图 5 可知,网络在前 20 轮训练时,损失值快速下降,80 轮之后,损失值基本趋于稳定。

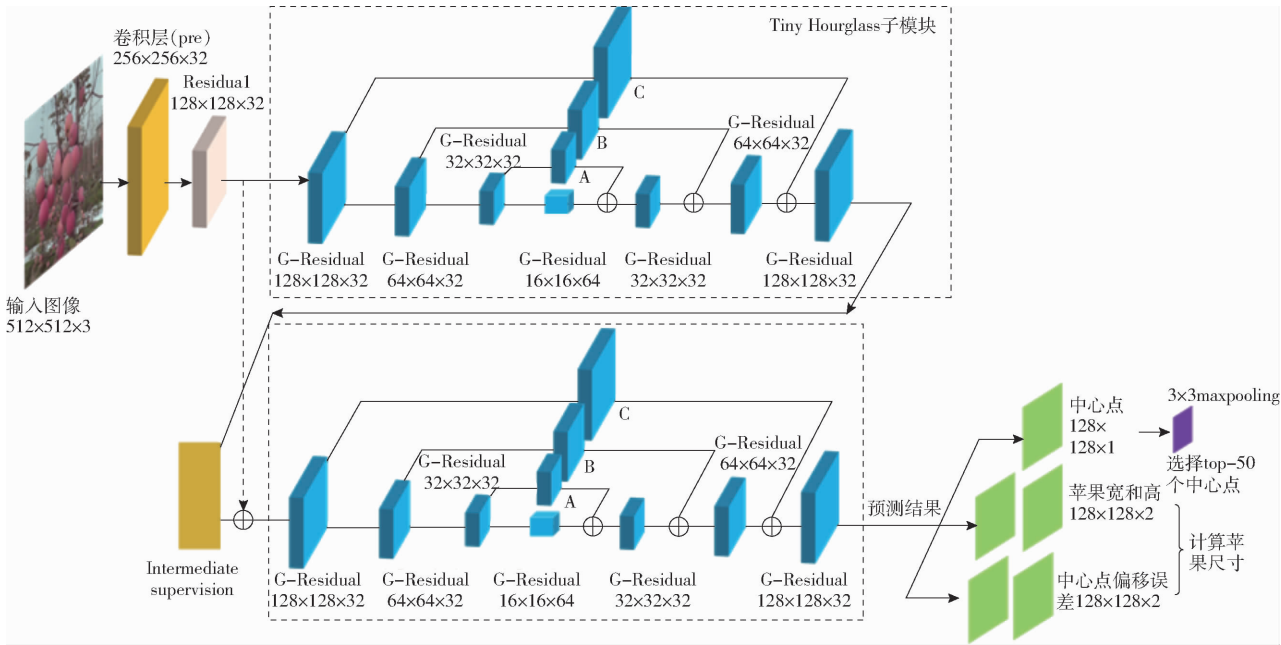


图 4 基于 Tiny Hourglass - 24 的 CenterNet 网络结构

Fig.4 CenterNet network structure based on Tiny Hourglass - 24

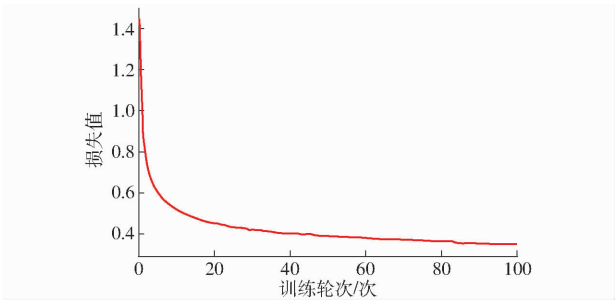


图 5 网络训练损失曲线

Fig.5 Network training loss curve

3.3 评价指标

为验证本文所提网络的识别性能,利用 F1 值和平均精度来对网络的识别性能进行评价。

4 结果与分析

4.1 不同深度 Tiny Hourglass 网络识别结果对比

网络深度对识别性能有很大影响,浅层网络会导致其特征提取能力不强,加深网络层数会提取更加复杂的深层特征,但会导致识别速度降低。为了在不降低识别精度的同时提升网络的识别速度,在确定骨干网络深度之前,本试验设计了 3 种不同深度的 Tiny Hourglass 网络进行性能对比,设置深度分别为 32、24、12 层,网络在测试集 A、B 下的识别结果如表 2 所示。

由表 2 可知,随着网络深度的降低,网络识别的 F1 值下降,且网络在测试集 B 上的平均精度和 F1 值降低较为严重。当骨干网络深度为 24 层时,在测试集 A 上识别的平均精度和 F1 值分别为 98.90% 和 96.39%;在测试集 B 上识别的平均精度和 F1 值

表 2 不同深度骨干网络识别结果				
Tab.2 Backbone network identification results of different depths				
测试集	骨干网络	平均精度/%	F1 值/%	单幅图像平均识别时间/s
A	Tiny Hourglass - 32	98.48	96.86	0.078
	Tiny Hourglass - 24	98.90	96.39	0.068
	Tiny Hourglass - 12	93.49	94.80	0.065
B	Tiny Hourglass - 32	94.10	94.00	0.101
	Tiny Hourglass - 24	93.63	92.91	0.069
	Tiny Hourglass - 12	74.18	83.78	0.072

为 93.63% 和 92.91%,单幅图像的平均识别时间为 0.069 s,在两类测试集上的识别精度和速度均较好。网络深度在 32 层时,在测试集 B 识别的平均精度和 F1 值与 24 层时基本相当,但 Tiny Hourglass - 24 网络识别一幅图像的时间比 Tiny Hourglass - 32 减少 0.032 s。当网络降为 12 层时,在测试集 B 上,识别的平均精度和 F1 值相较于 24 层分别下降了 19.45 个百分点和 9.13 个百分点,可见,12 层的浅层网络对密集场景果实识别效果较差。由于在测试集 B 上,单幅图像苹果数目增多,果实 在 图 像 中 的 尺 寸 变 小,苹 果 间 互 相 重 叠 遮 挡、苹 果 被 枝 条 遮 挡 严 重,12 层骨干网络提取特征能力有限,对复杂场景下多目标的识别效果较差。综上,Tiny Hourglass - 24 在识别速度和识别精度上均可以取得较好的结果,最终选择深度为 24 层的 Tiny Hourglass 网络作为骨干网络。

4.2 不同骨干网络识别结果对比

为了验证 Tiny Hourglass - 24 网络对苹果目标

的识别性能,本文将 Tiny Hourglass - 24 网络与另外两类骨干网络 DLA - 34、ResNet - 18 进行比较,在两类测试集下的识别结果见表 3。

表 3 不同骨干网络的识别结果

Tab. 3 Detection results of different backbone networks				
测试集	骨干网络	平均 精度/%	F1 值/ %	单幅图像平均 识别时间/s
A	Tiny Hourglass - 24	98. 90	96. 39	0. 068
	DLA - 34	96. 44	95. 90	0. 103
	ResNet - 18	94. 95	94. 63	0. 065
B	Tiny Hourglass - 24	93. 63	92. 91	0. 069
	DLA - 34	91. 16	93. 41	0. 103
	ResNet - 18	81. 30	88. 28	0. 064

由表 3 可知,Tiny Hourglass - 24 在识别精度和识别速度上性能最优。DLA - 34 骨干网络在测试集 B 上平均精度为 91. 16%,比 Tiny Hourglass - 24 低 2. 47 个百分点,此外,DLA - 34 网络模型较大、运行时间更长,在识别速度上,Tiny Hourglass - 24 在测试集 B 上对单幅图像的识别时间比 DLA - 34 减少 0. 034 s。ResNet - 18 骨干网络相较于 Tiny Hourglass - 24 和 DLA - 34 更加轻量化,在识别速度

上优于前两种骨干网络,但由于网络较浅,特征提取能力有限,在密集场景的测试集 B 上识别性能明显降低,其 F1 值和平均精度相较于 Tiny Hourglass - 24 分别下降了 4. 63 个百分点和 12. 33 个百分点。Tiny Hourglass - 24 骨干网络通过 3 倍下采样和上采样的方式使特征图的分辨率较高,有利于小目标的特征提取;结合跳跃连接层弥补下采样丢失的信息,增加了网络特征提取能力,更适用于密集场景下多目标的识别。

图 6 为基于 Tiny Hourglass - 24 的 CenterNet 识别结果,其中紫色框为网络识别到的苹果,红色框为漏识别苹果。由图 6 可以看出,网络在测试集 A 上对阴天和晴天逆光、顺光条件下的识别效果较好,对遮挡面积严重的苹果也可以准确识别。在测试集 B 中,网络在阴天环境下识别效果较好,且对重叠和遮挡的苹果识别效果较好。在逆光情况下,苹果表面颜色较暗,网络在苹果被枝干重度遮挡的情况下存在少量漏识别;在顺光情况下,苹果表面亮度增强,部分表面颜色特征变为白色,导致对遮挡苹果存在个别漏检。



图 6 基于 Tiny Hourglass - 24 的 CenterNet 识别结果

Fig. 6 Detection results of CenterNet based on Tiny Hourglass - 24

4. 3 不同算法识别结果对比

为了进一步验证基于无锚框的 CenterNet 算法对苹果目标的识别性能,本文将改进的 CenterNet 与 YOLO v3、CornerNet - Lite 算法在两类测试集上进行

了对比,以平均精度作为评价指标。YOLO v3 网络属于经典的基于锚框的识别思想,CornerNet - Lite 网络是另外一种基于无锚框的识别方法,它通过预测物体的左上角点和右下角点,将预测的关键

点进行匹配实现目标识别,网络识别结果如表 4 所示。

表 4 不同网络识别结果

Tab.4 Different network detection results

测试集	识别网络	平均	单幅图像
		精度/%	平均识别时间/s
A	改进 CenterNet	98.90	0.068
	YOLO v3	91.70	0.071
	CornerNet-Lite	90.40	0.610
B	改进 CenterNet	93.63	0.069
	YOLO v3	89.50	0.109
	CornerNet-Lite	64.60	0.715

由表 4 可知,本文提出的基于 Tiny Hourglass-24 的 CenterNet 网络在测试集 A 和测试集 B 上的平均精度均最高。在测试集 A 上,改进 CenterNet 网络的平均精度比 YOLO v3 和 CornerNet-Lite 分别高 7.20、8.50 个百分点。在测试集 B 上,改进 CenterNet 网络的平均精度高于 YOLO v3 网络 4.13 个百分点,高于 CornerNet-Lite 29.03 个百分点。图 7 为基于 Tiny Hourglass-24 的 CenterNet 网络在近距离场景和密集场景下中心点识别结果,可以看出,在树枝遮挡、果实互相重叠遮挡的情况下,网络均可以准确识别到苹果的中心点。



图 7 改进 CenterNet 网络预测的中心点结果
Fig.7 Results of improved CenterNet network prediction center

3 种网络的识别结果如图 8 所示,其中左图为在测试集 A 上的识别结果,右图为在测试集 B 上的识别结果。由图 8a 可知,改进 CenterNet 在近距离场景的识别效果较好,在果实被树干遮挡和果实之间互相重叠遮挡的情况,均能取得良好的识别结果。在密集场景的测试集上,网络在树叶遮挡、果实之间重度遮挡的情况下,均可以准确识别到苹果。由图 8b 可知,当果实被树干遮挡以及果实被树叶遮挡时,YOLO v3 存在漏识别的现象,这是由于 YOLO v3 是基于锚框的多尺度预测模型,由于锚框参数需要设置,无法灵活适应苹果目标的尺寸,在密集场景下遮挡严重,容易丢失苹果的特征,使得 YOLO v3 对小尺度遮挡类的苹果存在漏检。

由表 4 和图 8c 可知,基于 CornerNet-Lite 网络

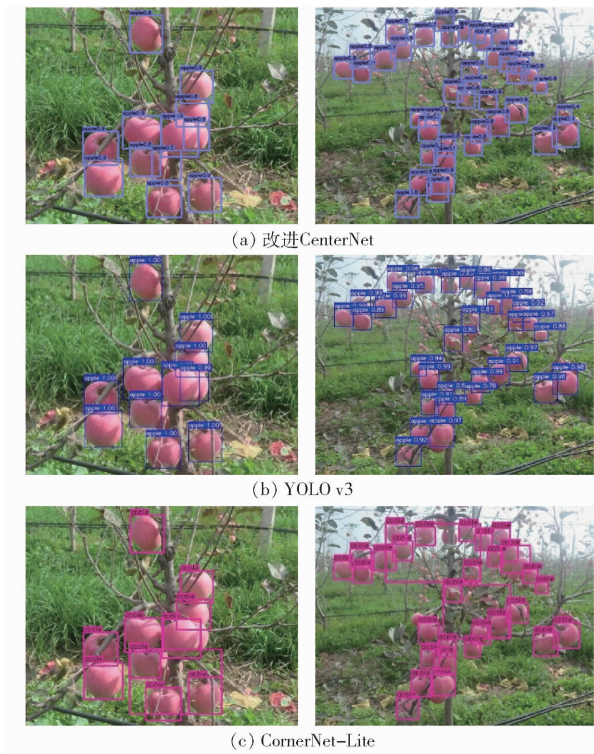


图 8 3 种模型在两类测试集中的识别结果
Fig.8 Detection results of three models under two test sets

的识别结果在近距离场景测试集上的性能远优于其在密集场景测试集。在近距离场景和密集场景中,网络识别均存在较大的误检框,这是由于 CornerNet-Lite 网络对每个目标预测出两个关键点,在预测出苹果的左上角和右下角两个关键点后,在关键点匹配阶段,存在部分不同苹果目标的关键点被错误组合为一个苹果,导致了网络的错检。在密集场景中,由于图像中苹果目标较小,果实与果实之间距离更加紧密,存在较大的误检框,降低了在密集场景下的识别精度。因此,基于多个关键点分组的思路并不适用于密集场景下的小目标识别,而基于 Tiny Hourglass-24 骨干网络的 CenterNet 网络,只预测一个目标中心的关键点,没有预测点的分组问题,也不需要 NMS 删除重复框,更适合密集环境下的小目标苹果识别。

4.4 性能验证

为了验证本文方法的识别性能,分别与 Faster R-CNN^[25]、R-FCN^[26]、改进 YOLO v3^[19] 的研究结果进行了比较,结果见表 5。

由表 5 可以看出,Faster R-CNN、R-FCN、改进 YOLO v3 算法在单幅图像苹果数约为 13 个的情况下,F1 值分别为 89.8%、90.2%、75.15%,而本文方法在单幅图像平均 25.45 个苹果的测试集上识别的 F1 值为 92.91%,比 Faster R-CNN、R-FCN、改进 YOLO v3 分别高 3.11、2.71、17.76 个百分点。

表 5 不同方法比较结果

Tab.5 Comparison of results of different methods					
方法	测试集图	总果实	平均每幅图	平均精	F1
	像数量/幅	数/个	像果实数/个	度/%	值/%
Faster R – CNN	193		12. 84	94. 80	89. 80
R – FCN	332	4234	12. 75		90. 20
改进 YOLO v3	58	759	13. 08		75. 15
本文方法	455	11 578	25. 45	93. 63	92. 91

这 3 类算法由于锚框的限制,导致对遮挡类目标识别精度不高,锚框的设计不仅增加了网络参数量,且容易丢失小目标苹果的信息,不利于密集场景下多苹果目标的识别。本文提出的密集场景下多苹果目标的快速识别方法,采用无锚框的设计使其能更好地适用遮挡类的多目标苹果识别,不需要使用 NMS 操作,节省了大量时间。本文方法在密集场景测试集上识别的平均精度为 93. 63% ,F1 值为 92. 91% ,可以实现密集场景下多苹果目标的准确识别。

5 结论

(1)提出了一种密集场景下多苹果目标的快速

识别方法。借鉴“点即是目标”的思路,直接预测苹果的中心点和宽、高尺寸来识别苹果。设计了基于 Tiny Hourglass – 24 骨干网络的 CenterNet 模型,并进一步优化 Hourglass 网络的残差结构来提高识别速度。该模型在近距离场景测试集下识别的平均精度为 98. 90% ,F1 值为 96. 39% ;在密集场景测试集下识别的平均精度和 F1 值分别为 93. 63% 和 92. 91% ,单幅图像平均识别时间为 0. 069 s。

(2)通过将改进 CenterNet 与 YOLO v3、CornerNet – Lite 网络在两类测试集上进行对比验证,试验结果表明,改进 CenterNet 网络在测试集 B 上的平均精度比 YOLO v3 提高 4. 13 个百分点,比 CornerNet – Lite 高 29. 03 个百分点。基于 Tiny Hourglass – 24 的 CenterNet 通过预测目标中心点的方法,比基于锚框的识别方法、基于多个关键点进行匹配的识别方法更具优势,整个识别过程不使用锚框和 NMS 后处理,减少了网络参数,更适用于密集场景下多苹果目标的识别。

参 考 文 献

[1] JIA W,ZHANG Y,LIAN J,et al. Apple harvesting robot under information technology: a review[J]. International Journal of Advanced Robotic Systems,2020,17(3):1–16.

[2] 赵春江. 智慧农业发展现状及战略目标研究[J]. 智慧农业,2019,1(1):1–7.
ZHAO Chunjiang. State-of-the-art and recommended developmental strategic objectives of smart agriculture[J]. Smart Agriculture,2019,1(1):1–7. (in Chinese)

[3] BECHAR A,VIGNEAULT C. Agricultural robots for field operations: concepts and components[J]. Biosystems Engineering, 2016,149:94–111.

[4] WEI X,JIA K,LAN J,et al. Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot[J]. Optik-International Journal for Light Electron Optics,2014,125(19):5684–5689.

[5] GONGAL A,AMATYA S,KARKEE M,et al. Sensors and systems for fruit detection and localization: a review[J]. Computers and Electronics in Agriculture,2015,116:8–19.

[6] ZHANG Q,LIU Y,GONG C,et al. Applications of deep learning for dense scenes analysis in agriculture: a review[J]. Sensors,2020,20(5):1520–1553.

[7] INKYN S,ZONGYUAN G,FERAS D,et al. DeepFruits: a fruit detection system using deep neural networks[J]. Sensors,2016,16(8):1222–1245.

[8] RONG Z,DAMEROW L,SUN Y,et al. Using colour features of cv. ‘Gala’ apple fruits in an orchard in image processing to predict yield[J]. Precision Agriculture,2012,13(5):568–580.

[9] 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.

[10] 顾宝兴,姬长英,王海青,等. 智能移动水果采摘机器人设计与试验[J]. 农业机械学报,2012,43(6):153–160.
GU Baoxing,JI Changying,WANG Haiqing,et al. Design and experiment of intelligent mobile fruit picking robot[J]. Transactions of the Chinese Society for Agricultural Machinery,2012,43(6):153–160. (in Chinese)

[11] SILWAL A,DAVIDSON J R,KARKEE M,et al. Design, integration, and field evaluation of a robotic apple harvester[J]. Journal of Field Robotics,2017,34(6):1140–1159.

[12] 王丹丹,宋怀波,何东健. 苹果采摘机器人视觉系统研究进展[J]. 农业工程学报,2017,33(10):59–69.
WANG Dandan,SONG Huaibo,HE Dongjian. Research advance on vision system of apple picking robot[J]. Transactions of the CSAE,2017,33(10):59–69. (in Chinese)

[13] KAMILARIS A,PRENAFETA-BOLDÚ F X. Deep learning in agriculture: a survey[J]. Computers and Electronics in Agriculture,2018,147(1):70–90.

[14] KOIRALA A,WALSH K B,WANG Z,et al. Deep learning for real-time fruit detection and orchard fruit load estimation;

- benchmarking of ‘MangoYOLO’[J]. Precision Agriculture,2019,20(6):1107–1135.
- [15] 熊俊涛,郑镇辉,梁嘉恩,等. 基于改进 YOLO v3 网络的夜间环境柑橘识别方法[J]. 农业机械学报,2020,51(4):199–206.
XIONG Juntao,ZHENG Zhenhui,LIANG Jiaen,et al. Citrus detection method in night environment based on improved YOLO v3 network[J]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(4):199–206. (in Chinese)
- [16] 熊俊涛,刘振,汤林越,等. 自然环境下绿色柑橘视觉检测技术研究[J]. 农业机械学报,2018,49(4):45–52.
XIONG Juntao,LIU Zhen,TANG Linyue,et al. Visual detection technology of green citrus under natural environment[J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(4):45–52. (in Chinese)
- [17] 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]. 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]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(10):24–34. (in Chinese)
- [18] KANG H,CHEN C. Fast implementation of real-time fruit detection in apple orchards using deep learning[J]. Computers and Electronics in Agriculture,2020,168:105108.
- [19] 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 Rendu,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)
- [20] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points [C] // IEEE Conference on Computer Vision and Pattern Recognition,2019:1904.07850.
- [21] LAW H, TENG Y, RUSSAKOVSKY O, et al. CornerNet – Lite: efficient keypoint based object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition,2019:1904.08900.
- [22] ZHOU X, ZHUO J, KRÄHENBÜHL P. Bottom-up object detection by grouping extreme and center points [C] // IEEE Conference on Computer Vision and Pattern Recognition,2019:850–859.
- [23] LAW H, DENG J. CornerNet: detecting objects as paired keypoints [J]. International Journal of Computer Vision, 2020, 128(3):642–656.
- [24] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6):84–90.
- [25] 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.
- [26] 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)
- ~~~~~

(上接第 264 页)

- [20] HUI T, TANG X, LOY C C. Liteflownet: a lightweight convolutional neural network for optical flow estimation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:8981–8989.
- [21] DOSOVITSKIY A, FISCHER P, ILG E, et al. Flownet: learning optical flow with convolutional networks [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015:2758–2766.
- [22] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C] // 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017:3645–3649.
- [23] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking [C] // 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016:3464–3468.
- [24] ZHANG Y, WANG C, WANG X, et al. A simple baseline for multi-object tracking [J]. arXiv preprint arXiv,2020:2004.01888.
- [25] 宋怀波,阴旭强,吴岷华,等. 基于自适应无参核密度估计算法的运动奶牛目标检测 [J]. 农业机械学报,2019,50(5):196–204.
SONG Huaibo, YIN Xuqiang, WU Dihua, et al. Detection of moving cows based on adaptive kernel density estimation algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(5):196–204. (in Chinese)