

优化 FCOS 网络复杂果园环境下绿色苹果检测模型

张中华¹, 贾伟宽^{1, 2*}, 邵文静¹, 侯素娟¹, Ji Ze³, 郑元杰¹

1. 山东师范大学信息科学与工程学院, 山东 济南 250358

2. 机械工业设施农业测控技术与装备重点实验室, 江苏 镇江 212013

3. School of Engineering, Cardiff University, Cardiff CF24 3AA, United Kingdom

摘要 目标果实的精准识别是实现果园测产和机器自动采摘的基本保障。然而受复杂的非结构化果园环境、绿色苹果与枝叶背景颜色接近等因素的影响, 制约着可见光谱范围内目标果实的检测精度, 给机器视觉识别带来极大挑战。针对复杂果园环境下的不同光照环境和果实姿态, 提出一种优化的一阶全卷积(FCOS)神经网络绿色苹果识别模型。首先, 新模型在 FCOS 的基础上融合了卷积神经网络(CNN)的特征提取能力, 消除了对锚框的依赖, 以单阶段、全卷积、无锚框的方式预测果实置信度与边框偏移, 在保证检测精度的前提下提升了模型的识别速度; 其次, 增加了自底向上的特征融合架构, 为模型提供了更加准确的定位信息, 进一步优化绿色苹果的检测效果; 最后根据 FCOS 末端三个输出分支设计整体损失函数, 完成模型训练。为尽可能模拟真实果园环境, 分别采集不同光照环境、光照角度、遮挡类型、摄像距离的绿色苹果图像, 制作数据集并用以模型训练。挑选最优训练模型在包含不同场景的验证集上进行评估, 结果为: 在检测效果方面, 平均精度为 85.6%, 与目前最先进的检测模型 Faster R-CNN, SSD, RetinaNet, FSAF 相比, 分别高出 0.9, 10.5, 2.5, 1.9 个百分点; 在模型设计方面, FCOS 的模型参数量与整个检测流程所需的计算量分别为 32.0 M 和 47.5 GFLOPs (10 亿次浮点运算), 与 Faster R-CNN 相比, 分别降低了 9.5 M 和 12.5 GFLOPs。对比表明, 在可见光谱范围内, 对复杂果园环境中绿色苹果, 提出的新模型具有更高的检测精度和识别效率, 为苹果果园测产和自动化采摘提供理论和技术支撑; 也可为其他果蔬的球形绿色目标果实识别提供借鉴。

关键词 FCOS 网络; 绿色果实; 目标检测

中图分类号: TP317.4 **文献标识码**: A

DOI: 10.3964/j.issn.1000-0593(2022)02-0647-07

引言

在实际苹果园环境下, 目标果实的精准识别是实现果园测产和自动采摘的基础, 果园的自动化和智能化作业是苹果产业的发展以及缓解农业劳动力缺乏的重要保障。受非结构化的复杂环境制约, 当前机器视觉系统难以应对各种干扰, 如果实受枝叶遮挡、果实间相互重叠、光照角度和强度、天气等因素; 特别是绿色苹果由于与实景颜色极为接近, 进一步加大了识别难度, 容易出现漏检或果实与叶片之间的混检。绿色苹果的高效识别和精准定位是实现果园智能作业的关键。

在可见光谱范围, 利用传统的机器学习在水果检测等方面积累了一定的研究基础^[1-3], 然而这些方法大多基于目标果实的颜色、纹理、形状等其他表征的组合进行目标果实的检测, 当果实由于光照强度、角度或自然环境变化等因素出现纹理特征不明显, 枝干、树叶遮挡或果实间相互重叠等因素出现形状缺失, 目标果实与同色系叶片背景相似出现颜色干扰等问题时, 该类方法的识别精度明显下降, 难以满足果实采摘机器人部署到实际应用时对精度和速度的要求。近年来, 随着深度学习技术的发展, 在目标识别过程中, 能够自行深度提取图像特征, 实现端到端的检测, 大幅提升了模型识别的精度与鲁棒性等多个方面的优势。在此基础上, 衍生出的众多目标检测和图像分割算法开始被广泛应用到目标果

收稿日期: 2021-01-02, 修订日期: 2021-02-01

基金项目: 国家自然科学基金项目(62072289, 61973141, 81871508), 山东省重点研发计划项目(2019GNC106115), 山东省自然科学基金项目(ZR2020MF076, ZR2019ZD04)资助

作者简介: 张中华, 1997 年生, 山东师范大学信息科学与工程学院硕士研究生 e-mail: zzh9714@163.com

* 通讯作者 e-mail: wkjia@sdu.edu.cn

实的识别^[4-5]。Bargoti 首先利用多尺度多层感知器和卷积神经网络将苹果图像分割, 提取出图像中的苹果目标, 然后用水分岭分割和圆形 Hough 变换法对苹果目标进行识别和计数^[6]。Jia 通过改进实例分割模型 Mask R-CNN 以适应苹果目标的检测, 结合 ResNet 与 DenseNet 作为原模型的特征提取网络, 大幅提高了重叠及枝叶遮挡环境下苹果目标的检测精度^[7]。熊俊涛借助 ResNet 与 DenseNet, 设计了 Des-YOLOv3 算法实现了网络多层特征的复用与融合, 加强了小目标和重叠遮挡果实的鲁棒性, 为夜间成熟柑橘的识别与检测提供了技术支持^[8]。以上基于卷积神经网络(convolutional neural networks, CNN)的果实识别算法在多个维度的表现均优于传统机器视觉方法。

当前以 Faster R-CNN, Mask R-CNN 为代表的双阶段检测分割算法, 以及以 SSD, YOLO 为代表的单阶段检测算法, 都沿用了先在原图上设置预定义锚框, 再通过网络进行调整得到边界框的方式进行预测。该类方法对计算与存储资源的需求较高, 识别速度慢, 实时性差, 正常作业时的稳定与功耗得不到保证, 尚不适合将其迁移部署至移动采摘设备。针对上述问题, 本研究提出一种基于无锚框检测器的一阶全卷积(fully convolutional one-stage, FCOS)高效绿色苹果检测优化模型, 首先借助 CNN 的特征提取能力, 消除对锚框的依赖, 在保证检测精度的前提下提升了模型的识别速度及识别不同外形果实的适应能力, 兼顾准确率与效率。在多尺度 FCOS 的基础上添加了自底向上的特征融合架构, 为模型提供更加准确的定位信息, 提升模型的检测效果。

1 实验部分

1.1 图像采集

图像采集地点为山东省烟台市福山区, 如图 1 所示, 挑选不同光照强度(清晨、中午、夜间)、不同光照角度(顺光、逆光、测光)、不同遮挡类型(枝叶遮挡、叶片遮挡、果实间相互重叠)的拍摄场景, 并按照不同的拍摄距离采集图像, 尽可能模拟苹果采摘机器人真实作业环境, 最终共采集绿色苹果图像 256 张。



图 1 果园环境下的绿色苹果图像

Fig. 1 Green apple images in orchard environment

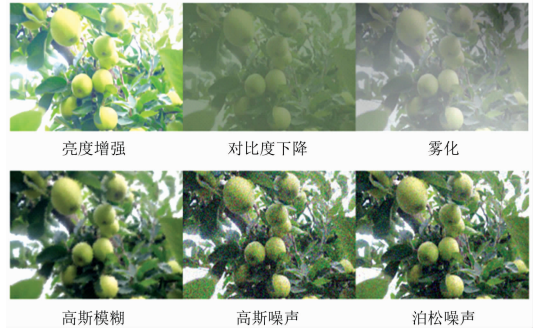


图 2 不同类型的图像增强示例

Fig. 2 Examples of different types of image enhancement

1.2 数据集生成

采用检测任务标注软件 labelme, 挑选 256 张图像中的 230 张进行标注, 统一将分辨率缩小至 600×400 像素, 图像中每个果实的最小外接矩阵作为真实框并生成对应的 json 文件。标注完成后对 230 幅图像进行数据增强, 增强类型包括亮度增强、对比度下降、雾化、高斯噪声、脉冲噪声、泊松噪声等, 如图 2 所示, 每种增强类型又分为不同增强程度, 最终共生成图像 5 290 张。由原始图像生成的增强图像共享同一 json 文件中标注信息, 在每一种干扰类型的每一种干扰程度中以 7 : 3 的比例划分训练集与验证集, 最终得到训练集共 3703 张图片, 验证集共 1587 张图片, 并分别生成 MS COCO 数据集^[9]格式的标注文件。

2 基于优化 FCOS 的绿色苹果检测网络

FCOS 不同于以往的单阶段或双检测检测框架^[10], FCOS 摆脱了模型对锚框的依赖, 在保证检测精度的前提下缩减模型容量与算法复杂度, 减少采摘机器人正常作业时的功耗并提升稳定性。FCOS 在整体架构上以全卷积、单阶段的方式进行预测, 模型结构与 RetinaNet 基本类似, 但正负样本的判定方式、预测目标的生成及模型的训练却大相径庭。如图 3 所示, 基础 FCOS 主要分为以下几个部分: 图像输入、ResNet50、特征金字塔网络(feature pyramid networks, FPN)和图像输出(类别、中心度、边界框)。

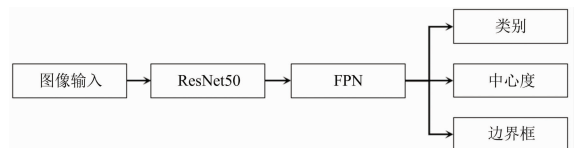


图 3 FCOS 网络结构简图

Fig. 3 FCOS network structure diagram

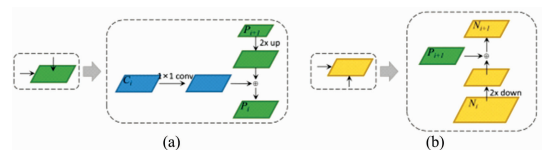


图 4 (a) 自顶向下融合方式; (b) 自底向上融合方式

Fig. 4 (a) Top-down pathway; (b) bottom-up pathway

2.1 FCOS 模型构建

2.1.1 ResNet 骨干框架

ResNet 是 FCOS 网络中的骨干部分，主要负责提取图像特征。其通过残差学习的方式，在网络中加入残差模块来帮助网络实现恒等映射，解决了 CNN 由于层数过深而导致的梯度爆炸/消散、退化问题。其基本思想为在网络前向过程中，当浅层的输出已经足够成熟，让深层网络后面的层能够实现恒等映射的作用；在反向过程中帮助传导梯度，让更深的模型能够成功训练。在 ResNet 中输出相同尺度的网络部分称为一级，每经过一级会通过卷积等降采样操作逐渐丰富深层网络的语义信息和表达能力，得到尺寸为原图尺寸 $1/2^l$ 的特征图 (feature map)， l 为层级索引，每一级最后一个残差块的输出分别记为 $\{C_1, C_2, C_3, C_4, C_5\}$ ，如图 5 中左侧 Backbone 部分所示。

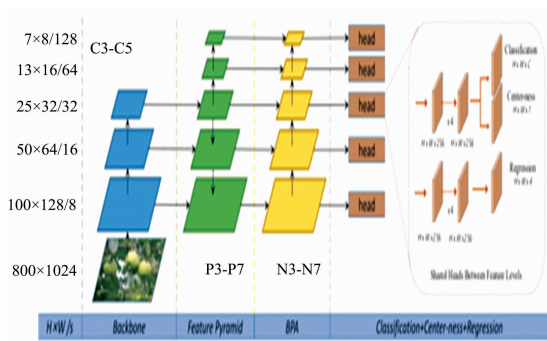


图 5 优化后的 FCOS 网络

Fig. 5 Optimized FCOS network

2.1.2 FPN 特征提取网络

ResNet 网络最后输出的特征图 C_5 虽然包含了丰富的语义信息，但经过不断的降采样之后，其分辨率很低，边界等细节信息基本丢失，因此只适合预测大尺度目标。FPN 主要解决了目标检测中的多尺度预测问题，其通过融合深层特征图的语义信息和浅层特征图的细节信息来构建特征金字塔，将不同尺度的待检测目标分发给金字塔中不同层级的特征图负责预测。该网络能有效提升模型对于不同尺度，尤其是小尺度目标果实的检测效果，从而能够更好的指导苹果采摘机器人的识别采摘与路径规划。在 FCOS 中，采用了 ResNet 中后三个层级输出的特征图 C_3, C_4, C_5 来构建特征金字塔， C_3, C_4, C_5 通过 1×1 卷积的横向连接与 2 倍上采样的自顶向下连接进行融合，得到 P_3, P_4, P_5 ，如图 4(a) 所示，再由 P_5 经过两次下采样得到 P_6 和 P_7 ，得到构建好的特征金字塔 $\{P_3, P_4, P_5, P_6, P_7\}$ ，如图 5 中 Feature Pyramid 部分所示。

2.1.3 结果输出

基础 FCOS 采用特征金字塔 $\{P_3, P_4, P_5, P_6, P_7\}$ 中的每一层特征图分别预测。如图 5 中最右侧所示，每一层特征图分别接以分类、中心度、回归三个分支，分别负责预测特征图上每一个空间位置属于苹果的置信度、距离苹果中心的偏移程度及对应在原图坐标到真实框四条边的距离。其中分类与回归分支分别使用两条不同的全卷积通道解耦预测，使

其更专注于各自的训练任务，中心度的预测相对简单，因此和分类分支共享同一部分全卷积网络，最后分别通过 1 个 1×1 卷积核和 C 个 1×1 卷积核得到中心度与分类输出。通过比较 RetinaNet 与 FCOS 在网络设计上的差别，不同之处在于 FCOS 添加了一个中心度分支。FCOS 虽然消除了模型对锚框的依赖，但其与基于锚框的检测器之间仍然存在着一定的距离，原因在于距离目标中心较远的位置产生了很多低质量的预测边框。因此，FCOS 提出了一种简单而有效的策略来抑制这些低质量的预测边框，而且不涉及任何超参数。这种策略即引入中心度分支，预测特征点与目标中心的偏移程度。测试时，将预测的中心度与相应的分类分数相乘，得到边框的最终得分，因此，中心度可以降低远离对象中心的边框的权重，从而通过非极大值抑制等后处理过程将其过滤掉。

2.2 FCOS 的优化

本工作主要针对 FCOS 中的特征融合方式进行改进。对目标检测而言，深层的语义信息与浅层的细节信息都很重要，特征融合通过 FPN 实现，其虽然通过横向连接与自顶向下的架构方式逐渐丰富了浅层特征图的语义容量，但深层特征图的细节信息仍是由 ResNet 通过几十甚至一百多个网络层传递得到，即浅层到深层的传递路径过长，导致浅层信息丢失严重，增加了定位信息流动的难度。因此，借鉴 PANet 中的自底向上的路径增强 (bottom-up path augmentation, BPA) 模块并将其融合至 FCOS 网络中。如图 5 中 BPA 部分所示，对通过 FPN 得到的特征金字塔 $\{P_3, P_4, P_5, P_6, P_7\}$ 再连接一个自底向上的特征融合路径，生成一个新的特征金字塔 $\{N_3, N_4, N_5, N_6, N_7\}$ ，更好的保留浅层特征信息，具体融合方式如图 4(b) 所示。

2.2.1 正负样本判定及训练目标生成

设特征图 N_l 上的某一空间位置为 $A: (x, y)$ ，按照当前层级的下采样数 s_l 可得其在输入图片上对应感受野区域的中心坐标为 $A': (\lfloor s_l/2 \rfloor + xs_l, \lfloor s_l/2 \rfloor + ys_l)$ ，如图 6 所示。如果 A' 落在了真实标注框 $G: (x_1, y_1, x_2, y_2)$ 内，则将空间位置 A 视为训练阶段的正样本点。反之，则全部视为负样本点。

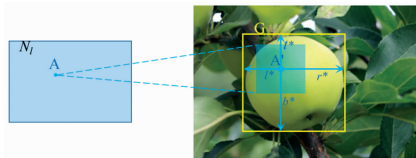


图 6 坐标映射图

Fig. 6 Coordinate mapping

正样本在分类分支的训练目标为苹果标签，因为本工作的目的是检测图像中的绿色苹果，是一个二分类问题，因此，分别设正负样本的训练目标 P^* 取值为 1, 0。正样本在回归分支的训练目标为一个四维向量 $d^* = (l^*, t^*, r^*, b^*)$ ，如图 6 所示， l^*, t^*, r^*, b^* 分别为映射坐标到与其相关真实框的四条边的距离。最后，对于中心度分支的训练目标，其计算方式如式(1)所示

$$\text{centerness}^* = \sqrt{\frac{\min(l^*, r^*)}{\max(l^*, r^*)} \times \frac{\min(t^*, b^*)}{\max(t^*, b^*)}} \quad (1)$$

式(1)中, centerness^* 的取值范围为 $[0, 1]$, 当映射坐标位于真实框正中心时, centerness^* 取值为 1, 随着映射坐标朝着真实框四周偏移而逐渐衰减, 如图 7 所示, 红色、蓝色和其他颜色分别表示 1, 0 和 $[0, 1]$ 之间的渐变值。模型测试阶段, 将中心度的预测值与分类分支对应位置的预测值相乘作为最后预测框的置信度, 从而有效抑制远离苹果果实中心的正样本点预测的边界框。

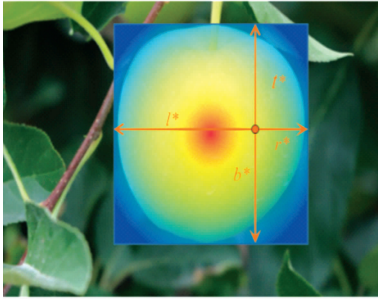


图 7 centerness^* 取值变化

Fig. 7 Centerness^* value change

2.2.2 损失函数

模型实现检测绿色果实能力的方式, 是通过迭代训练计算预测边框与真实框之间的损失, 并按照损失值反向传播梯度来更新模型参数, 直至损失减少并收敛至一个取值区间。对应 FCOS 网络末端的三个输出分支, 其损失函数由 L_{cls} , L_{reg} , L_{center} 三部分组成, 分别采用 Focal Loss, IoU (intersection of union) Loss 和 BCE Loss 计算, 模型整体损失函数如式(2)所示。

$$L(\{p_{x,y}\}, \{d_{x,y}\}, \{\text{center}_{x,y}\}) = \frac{1}{N_{\text{pos}}}_{x,y} \sum L_{\text{cls}}(p_{x,y}, p_{x,y}^*) + \frac{\lambda}{N_{\text{pos}}}_{x,y} \sum p_{x,y}^* L_{\text{reg}}(d_{x,y}, d_{x,y}^*) + \frac{\beta}{N_{\text{pos}}}_{x,y} \sum p_{x,y}^* L_{\text{center}}(\text{center}_{x,y}, \text{center}_{x,y}^*) \quad (2)$$

式(2)中, $p_{x,y}$, $d_{x,y}$, $\text{center}_{x,y}$ 分别为分类分支, 回归分支, 中心度分支在空间位置 (x, y) 处的预测值, $p_{x,y}^*$, $d_{x,y}^*$, $\text{center}_{x,y}^*$ 对应空间位置 (x, y) 处的训练目标, 三个损失项中, L_{reg} , L_{center} 只针对正样本而言, N_{pos} 表示正样本数量, λ , β 为各损失项的平衡系数。

关于式(2)中的分类损失 L_{cls} , 如式(3)和式(4)所示

$$L_{\text{cls}}(p_{x,y}, p_{x,y}^*) = -\alpha_t (1 - p'_{x,y})^\gamma \log(p'_{x,y}) \quad (3)$$

$$p'_{x,y} = \begin{cases} p_{x,y} & \text{if } p_{x,y}^* = 1 \\ 1 - p_{x,y} & \text{otherwise} \end{cases}, \quad \alpha_t = \begin{cases} \alpha & \text{if } p_{x,y}^* = 1 \\ 1 - \alpha & \text{otherwise} \end{cases} \quad (4)$$

式(3)和式(4)中, α 负责平衡正负样本之间的重要性, γ 负责调节简单样本权重降低的速率。

关于式(2)中的回归损失 L_{reg} , 如式(5)所示

$$L_{\text{reg}}(d_{x,y}, d_{x,y}^*) = -\ln \frac{\text{Intersection}(d_{x,y}, d_{x,y}^*)}{\text{Union}(d_{x,y}, d_{x,y}^*)} \quad (5)$$

式(5)中, $\text{Intersection}(d_{x,y}, d_{x,y}^*)$ 与 $\text{Union}(d_{x,y}, d_{x,y}^*)$ 分别为预测框与真实框之间的相交面积与合并面积。关于式(2)中的中心度损失 L_{center} , 如式(6)所示

$$L_{\text{center}}(\text{center}_{x,y}, \text{center}_{x,y}^*) = -(\text{center}_{x,y} \log(\text{center}_{x,y}^*) + (1 - \text{center}_{x,y}) \log(1 - \text{center}_{x,y}^*)) \quad (6)$$

3 结果与讨论

为验证优化 FCOS 模型对绿色苹果检测的有效性, 进行以下试验。首先介绍试验运行平台及模型在训练与测试阶段的相关实施细节; 然后挑选最优训练模型, 在验证集上进行评估并分析试验数据; 接着分别对基础 FCOS 与优化 FCOS 进行训练, 分析 BPA 模块的绿色苹果检测的有效性; 最后, 选取目前最先进的各类型目标检测, 测试并对比模型在检测绿色苹果时在精度与速度两方面性能上的差异。

3.1 试验运行平台

服务器的主要配置环境为 Ubuntu 16.04 操作系统、32 GB 的 Tesla V100 显卡和 10.0 的 CUDA 环境。所有模型均使用 Python 语言及 Pytorch 1.4 深度学习库, 并借助 MMDetection 框架相关模块进行搭建。

3.2 试验实施

训练阶段: 正式训练之前, 抽取 COCO 数据集中包含苹果图像的 1 586 张图像并使用这些图像进行预训练, 将预训练完成之后参数迁移至 FCOS 网络作为初始化参数以更好的提升模型精度和鲁棒性。正式训练时, 采用 mini-batch 的方式迭代训练 12 个 epoch, 每次迭代使用 2 个样本作为一个 batch, 共迭代 31 740 次, 训练过程中的三个分支产生的损失变化如图 8(a)所示, 横轴为迭代次数, 纵轴为损失值。每训练完一个 epoch 后在验证集上进行评估, 得到 AP 变化曲线图如图 8(b)所示。每次更新权值时使用 BN 进行正则化, 使用 SGD(stochastic gradient descent)更新模型参数, 学习率、权重衰减、动量分别设为 0.002 5, 0.000 1 和 0.9。网络训练之前统一尺寸至(1 200, 800), 并依次经过随机翻转、正则化、填充等预处理操作。

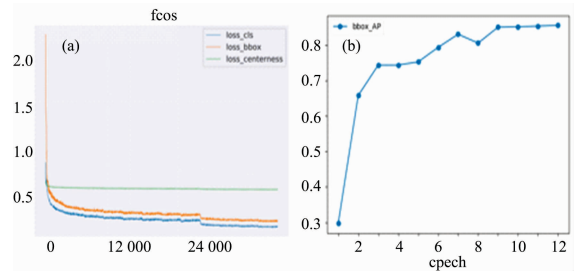


图 8 损失变化曲线(a)与 AP 变化曲线图(b)

Fig. 8 Loss value curve (a) and AP value curve (b)

测试阶段: 图像输入网络之前同样经过剪裁、随机翻转、正则化、填充等预处理操作; 网络推断结束后, 首先剔除置信度小于 0.05 的低质量预测框, 再使用非极大值抑制

(non maximum suppression, NMS) 筛选重叠度过高的预测框, 筛选时使用 IoU 等于 0.5 作为阈值, 筛选完成后按照置信度排序, 每张图片至多保留前 100 置信度的预测框。

3.3 试验评估指标

由于需要对模型的检测结果作出综合评价, 所以评价指标需要同时考虑准确率和召回率, 因此本工作主要选择 AP 作为评估指标, 首先, 在特定 IoU 阈值下准确率 P 与召回率 R 定义如式(7)所示

$$P = TP_s / (TP_s + FP_s) \quad R = TP_s / (TP_s + FN_s) \quad (7)$$

式(7)中, TP_s 为实际为苹果且被模型预测为苹果的个数, 即真实的正样本数量; FP_s 为实际为背景却被模型预测为苹果的个数, 即虚假的正样本数量; FN_s 为实际为苹果但没有被网络识别为苹果的个数, 即虚假的负样本个数。进一步可得特定 IoU 阈值下平均精度 $AP^{IoU=i}$ 定义如式(8)所示

$$AP^{IoU=i} = 1/101 \sum_{r \in R} p_{interp}(r) = 1/101 \sum_{r \in R} \max_{r': r' \geq r} \hat{p}(r') \quad (8)$$

如式(8)所示, 特定 IoU 阈值下的 AP 值是通过取 101 个召回率 $R: [0, 0.01, \dots, 1]$ 下对应的准确率并取平均得到的。求得特定 IoU 阈值下的, 最终的评估指标 AP 定义如式(9)所示

$$AP = \frac{1}{10} \sum_{i \in I} AP^{IoU=i} \quad (9)$$

其中, I 为 IoU 阈值数组: $[0.5, 0.55, \dots, 0.95]$, 即 AP 值为分别计算 IoU 阈值从 0.5 开始, 每间隔 0.05~0.95 结束的 10 个 $AP^{IoU=i}$ 值并取平均。本工作还使用以下评估指标, 具体定义细节请参照 MS COCO。

AP^{small} : 评估模型对小尺度目标果实的检测效果(真实框面积小于 32^2); AP^{medium} : 评估模型对中等尺度目标果实的检测效果(真实框面积介于 $32^2 \sim 96^2$); AP^{large} : 评估模型对大尺度目标果实的检测效果(真实框面积大于 96^2); AR: 每个图像中检测到固定数量 100 时的最大召回, 同样平均 10 个 IoU 阈值下的结果; AR^{small} , AR^{medium} , AR^{large} : 模型对不同尺度的平均召回; Params: 模型包含的参数数量, 用来衡量模型的简洁程度; FLOPs: 模型推断时所需的浮点运算次数, 用来衡量模型的计算复杂度。

3.4 模型检测效果

如图 8(b) 所示, 模型在训练完最后一个 epoch 后 AP 值最高, 为 85.6%, 因此挑选该模型作为最优模型来评估网络整体性能, 模型在验证集上的详细评估结果如表 1 所示。

表 1 FCOS 网络评估结果

Table 1 Evaluation results of FCOS network

Metric	AP	$AP^{IoU=0.5}$	AP^{small}	AP^{medium}	AP^{large}	AR	AR^{small}	AR^{medium}	AR^{large}	Params	FLOPs
Value	85.6%	90.0%	30.7%	73.6%	87.6%	87.6%	46.3%	81.2%	89.4%	32.0 M	47.5 G

此外, 我们选取多幅包含了重叠、枝叶遮挡、夜间、远景及逆光等混合干扰情况下的图像进行检测, 并对检测效果进行研究分析, 如图 9 所示, 每对图片中的左列为人工标注图, 右列为 FCOS 模型检测的效果图。可以看出, 由于各种干扰及混合干扰导致苹果的形状、颜色和纹理发生缺失时, FCOS 均能取得很好的检测效果, 且在人工标注的目标果实检测图中, 存在一些果实由于面积遮挡严重或拍摄距离过远等原因而没有标注, FCOS 也能准确的将其检测出来, 说明模型不仅具有较好的检测效果与抗干扰能力之外, 而且具有较强的泛化能力。综上所述, 优化 FCOS 网络能够胜任不同复杂场景下的绿色苹果检测任务。



图 9 模型在不同干扰场景下的检测效果

Fig. 9 Detection effect of the model in different interference scenes

3.5 检测效果对比

为进一步说明优化 FCOS 网络在速度与精度上的高效

性, 选取目前最先进且同样基于深度学习的目标检测算法, 在验证集上进行评估并分析对比试验指标, 以说明模型在复杂自然环境下检测绿色苹果的表现。分别选择基于锚框的双阶段算法 Faster R-CNN^[11]、基于锚框的单阶段算法 SSD^[12] 和 RetinaNet^[13]、无锚框的单阶段算法 FSAF(feature selective anchor-free)^[14] 四种方法。试验运行在同一设备的相同环境下, 且使用相同的图像预处理流程, 模型检测完成后, 统一使用 NMS 和相同的 IoU 阈值筛选, 具体试验对比结果如表 2 所示。

表 2 各模型对绿色苹果的检测结果

Table 2 Detection results of different models

Methods	Metrics			
	AP /%	AR /%	Params /M	FLOPs /G
two-stage anchor-based				
Faster R-CNN	84.7	86.7	41.5	60.0
one-stage anchor-based				
SSD512	75.1	78.4	24.4	80.5
RetinaNet	83.1	84.7	36.1	48.4
one-stage anchor-free				
FSAF	83.7	84.8	36.0	48.0
ours				
FCOS	85.6	87.6	32.0	47.5

如 2 表所示, 优化 FCOS 在检测精度上, 其综合评估指

标 AP 与 AR 值最高, 分别为 85.6% 和 87.6%, 均高于其他四种不同类型的目标检测算法。其中 Faster R-CNN 与 FCOS 的检测精度相差较小, 仅相差 0.9 个百分点, 这得益于其先通过区域候选网络(region proposal network, RPN)得到候选框再进行校准的双阶段检测流程, 但代价也同样明显, 如表 2 所示, Faster R-CNN 的参数量和计算量分别为 41.5 M 和 60.0 GFLOPs, 远超 FCOS 12.0 M 和 12.5 GFLOPs。SSD512 模型的参数量仅为 24.4 M, 这是因为其提取特征时

使用的卷积神经网络为 VGG16, 相比 FCOS 模型使用的 ResNet50 更加简洁, 因此参数量更少, 但由于其是一种基于锚框的检测器, 因此所需的计算量同样很高。为更直观比较以上五种模型对绿色苹果的检测效果, 我们挑选了三幅具有代表性且检测难度较大的图像并可视化预测结果如图 10 所示。可以看出, 相较于其他四种算法, FCOS 几乎没有漏检、叶片与果实之间混检、多个相邻果实识别为一个的现象。



图 10 五种不同类型的检测器在不同复杂场景下的检测效果图

Fig. 10 Detection effect images of 5 detectors in different complex scenes

4 结 论

基于无锚框、单阶段的目标检测模型 FCOS, 并结合复杂自然场景下绿色苹果检测的实际要求进行优化。实验结果表明, 该方法的检测精度高, 在各种干扰条件下的鲁棒性强。相较于 Faster R-CNN 等不同类型的检测算法, 模型的计算及存储消耗更少, 架构设计更为简洁。另外, 在整个检测过程中, 我们借助了 FCOS 高效的检测能力, 其舍弃了以往深度视觉模型对锚框的依赖, 使得 FCOS 能直接被泛化应用到其他果实的识别中。在本文实验中采用的是绿色苹果果实, 形状为类球形, 若采用基于锚框的检测分割算法识别其

他果实时, 例如黄瓜、辣椒等, 又需要改变锚框大小等超参数以提升识别准确率, 同时还需要考虑锚框放置的密集程度以提升召回率。而 FCOS 不需要进行这些操作, 避免了超参数设置对果实识别效果的影响与众多锚框对计算资源的消耗。最后, 对最先进的基于深度学习的视觉识别算法, 选择并对比基于锚框的双阶段算法 Faster R-CNN、基于锚框的单阶段算法 SSD, RetinaNet、无锚框的单阶段算法 FSAF 在复杂自然条件下对绿色果实识别上的性能差异, 其各项评估指标更好, 优势显著。改进 FCOS 能够以更简洁的模型结构和更少的计算量达到更高的检测精度, 能够同时实现速度与精度上的高效性且适应复杂的果园环境。

References

- [1] WANG Dan-dan, SONG Huai-bo, HE Dong-jian(王丹丹, 宋怀波, 何东健). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2017, 33(10): 59.
- [2] Jia W K, Zhang Y, Lian J, et al. International Journal of Advanced Robotic Systems, 2020, 17(5-6): 25310.
- [3] LI Da-hua, ZHAO Hui, YU Xiao(李大华, 赵 辉, 于 晓). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2019, 39(9): 2974.
- [4] Koirala A, Walsh K B, Wang Z, et al. Computers and Electronics in Agriculture, 2019, 162: 219.
- [5] Boogaard F P, Rongen K S A H, Kootstra G W. Biosystems Engineering, 2020, 192: 117.
- [6] Bargoti S, Underwood J P. Journal of Field Robotics, 2017, 34(6): 1039.
- [7] Jia W, Tian Y, Luo R, et al. Computers and Electronics in Agriculture, 2020, 172: 105380.
- [8] XIONG Jun-tao, ZHENG Zhen-hui, LIANG Jia-en, et al(熊俊涛, 郑镇辉, 梁嘉恩, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2020, 51(4): 199.
- [9] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context, European Conference on Computer Vision. Springer, Cham, 2014: 740.
- [10] Tian Z, Shen C, Chen H, et al. FCOS: Fully Convolutional One-Stage Object Detection, Proceedings of the IEEE International Conference on Computer Vision, 2019. 9627.
- [11] Ren S, He K, Girshick R, et al. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137.
- [12] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector, European Conference on Computer Vision. Springer, Cham, 2016. 21.

- [13] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection, Proceedings of the IEEE International Conference on Computer Vision, 2017. 2980.
- [14] Zhu C, He Y, Savvides M. Feature Selective Anchor-Free Module for Single-Shot Object Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

Green Apple Detection Based on Optimized FCOS in Orchards

ZHANG Zhong-hua¹, JIA Wei-kuan^{1,2*}, SHAO Wen-jing¹, HOU Su-juan¹, Ji Ze³, ZHENG Yuan-jie¹

1. School of Information Science and Engineering, Shandong Normal University, Ji'nan 250358, China

2. Key Laboratory of Facility Agriculture Measurement and Control Technology and Equipment of Machinery Industry, Zhenjiang 212013, China

3. School of Engineering, Cardiff University, Cardiff CF24 3AA, United Kingdom

Abstract In the visible spectrum range, the accurate recognition of target fruit is the fundamental guarantee for achieving orchard yield measurement and machine automatic picking. However, this task is susceptible to many interferences, such as the complex unstructured orchard environment, the close color between green apples and background leaves, etc., which significantly restrict the detection accuracy of target fruits and bring great challenges to recognition of machine vision. It targeted the different illumination environments and fruit postures under the complex orchard environment. An optimized convolution and one-stage (FCOS) fully neural network model for green apple recognition is proposed in this study. Firstly, the new model combines the feature extraction ability of convolutional neural network (CNN) based on FCOS, eliminates the dependence of previous detectors on anchor boxes, and switches to a novel manner of one-stage, full convolution and anchor-free for predicting the fruit confidence and boxes offsets, which greatly improves the recognition speed of the model while ensuring the detection accuracy simultaneously. Secondly, the bottom-up feature fusion architecture is embedded after the feature pyramid to provide more accurate positioning information for high-levels and thus further optimize the detection effect of green apple. Finally, the overall loss function is designed to complete the iterative training given three output branches of FCOS. To simulate the real orchard environment as possible, we collected green apple images in various environments with different lighting environments, illumination angle, occlusion type, camera distance for data sets generation and model training, and then evaluated the optimal model on validation set containing different scenes. The experimental results show that our proposed model's average precision (AP) is 85.6%, which is 0.9, 10.5, 2.5 and 1.9 percentage points higher than the state-of-the-art detection models Faster R-CNN, SSD, RetinaNet and FSAF, respectively. In the aspect of model design, the model parameters of FCOS and the calculation of the whole detection process are 32.0 M and 47.5 GFLOPs (billion floating-point operations), respectively, which are 9.5 M and 12.5 GFLOPs lower than those of Faster R-CNN. Comparisons of experimental results show that the new model has higher detection accuracy and recognition efficiency in the visible spectrum, which can provide theoretical and technical support for orchard yield measurement and automatic picking. In addition, the new model can also provide theoretical references for other kinds of fruits and vegetables.

Keywords FCOS network; Green fruits; Object detection

(Received Jan. 2, 2021; accepted Feb. 1, 2021)

* Corresponding author