

# 基于改进型 YOLOv4-LITE 轻量级神经网络的密集圣女果识别

张 伏<sup>1,2,3</sup>, 陈自均<sup>1</sup>, 鲍若飞<sup>1</sup>, 张朝臣<sup>1</sup>, 王治豪<sup>1</sup>

(1. 河南科技大学农业装备工程学院, 洛阳 471003; 2. 江苏大学现代农业装备与技术教育部重点实验室, 镇江 212013;  
3. 机械装备先进制造河南省协同创新中心, 洛阳 471003)

**摘 要:** 对密集圣女果遮挡、粘连等情况下的果实进行快速识别定位, 是提高设施农业环境下圣女果采摘机器人工作效率和产量预测的关键技术之一, 该研究提出了一种基于改进 YOLOv4-LITE 轻量级神经网络的圣女果识别定位方法。为便于迁移到移动终端, 该方法使用 MobileNet-v3 作为模型的特征提取网络构建 YOLOv4-LITE 网络, 以提高圣女果果实目标检测速度; 为避免替换骨干网络降低检测精度, 通过修改特征金字塔网络 (Feature Pyramid Networks, FPN) + 路径聚合网络 (Path Aggregation Network, PANet) 的结构, 引入有利于小目标检测的  $104 \times 104$  尺度特征层, 实现细粒度检测, 在 PANet 结构中使用深度可分离卷积代替普通卷积降低模型运算量, 使网络更加轻量化; 并通过载入预训练权重和冻结部分层训练方式提高模型的泛化能力。通过与 YOLOv4 在相同遮挡或粘连程度的测试集上的识别效果进行对比, 用调和均值、平均精度、准确率评价模型之间的差异。试验结果表明: 在重叠度为 0.50 时所提出的密集圣女果识别模型在全部测试集上调和均值、平均精度和准确率分别为 0.99、99.74% 和 99.15%, 同比 YOLOv4 分别提升了 0.15、8.29、6.55 个百分点, 权重大小为 45.3 MB, 约为 YOLOv4 的 1/5, 对单幅  $416 \times 416$  (像素) 图像的检测, 在图形处理器 (Graphics Processing Unit, GPU) 上速度可达 3.01 ms/张。因此, 该研究提出的密集圣女果识别模型具有识别速度快、识别准确率高、轻量化等特点, 可为设施农业环境下圣女果采摘机器人高效工作以及圣女果产量预测提供有力的保障。

**关键词:** 机器视觉; 模型; YOLO; 深度学习; 图像识别; 目标检测

doi: 10.11975/j.issn.1002-6819.2021.16.033

中图分类号: TP391, TP81

文献标志码: A

文章编号: 1002-6819(2021)-16-0270-09

张伏, 陈自均, 鲍若飞, 等. 基于改进型 YOLOv4-LITE 轻量级神经网络的密集圣女果识别[J]. 农业工程学报, 2021, 37(16): 270-278. doi: 10.11975/j.issn.1002-6819.2021.16.033 http://www.tcsae.org

Zhang Fu, Chen Zijun, Bao Ruofei, et al. Recognition of dense cherry tomatoes based on improved YOLOv4-LITE lightweight neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2021, 37(16): 270-278. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2021.16.033 http://www.tcsae.org

## 0 引 言

视觉识别系统作为果蔬采摘机器人系统的组成部分, 在果蔬目标识别定位、自动采摘和果蔬估产等方面<sup>[1-2]</sup>具有至关重要作用。圣女果 (Cherry tomatoes) 因其营养价值高、风味独特而被广泛种植。由于圣女果果实高度密集 (粘连)、果实小、叶片遮挡严重, 且其生长周期长、果蔬高度不一, 采收环节成为最费时耗力的工作<sup>[3-4]</sup>, 因此研究圣女果快速精确识别, 对实现采摘机器人自动采摘、提高作业效率具有较大应用价值。

近年来, 国内外基于传统机器视觉和图像处理技术, 对自然环境下果蔬目标识别提出了解决方法。马翠花等<sup>[5]</sup>提出了使用基于密集和稀疏重构的显著性检测方法与改进随机 Hough 变换识别方法, 对番茄果簇中的绿色单果番茄进行识别, 识别准确率为 77.6%。李寒等<sup>[6]</sup>通过

RGB-D 图像和 SOM-K-means 算法获取果实的位置信息与轮廓形状, 该方法对多个番茄重叠粘连识别正确率达 87.2%。Payne 等<sup>[7]</sup>提出一种基于机器视觉的芒果产量估测方法, 利用 RGB 和 YCbCr 颜色空间分割和基于相邻像素变异性的纹理分割方法, 从背景像素中有效分割出水果, 该方法可计算芒果果实的数量及产量估测。李昕等<sup>[8]</sup>将边缘预检测、快速定位圆心点等模块添加在 Hough 算法中, 对遮挡油茶果识别率为 90.70%。Gong 等<sup>[9]</sup>提出一种改进的 8 连通性的编码识别算法, 进行簇状柑橘的产量预测。Gong 等<sup>[9]</sup>提出一种改进的 8 连通性的编码识别算法, 进行簇状柑橘的产量预测。Xu 等<sup>[10]</sup>采用基于面向梯度直方图 (Histogram of Oriented Gradient, HOG) 描述符与支持向量机 (Support Vector Machine, SVM) 分类器相结合的方法, 分类器检测率为 87%。此外, 还有学者对粘连、重叠的果实进行识别研究<sup>[11-15]</sup>, 检测准确率均在 90% 左右。上述方法利用果实的形状特征、纹理特征与背景颜色差异特征等单特征或多特征组合实现果实的识别, 一旦遇到光照改变、果实粘连重叠、枝叶遮挡、相似度高的背景等情况识别准确率会降低。另外, 运用传统的机器视觉技术识别果实, 受制于自身分类器算法的限制, 无法满足当前复杂环境下果实目标识别任务<sup>[16-17]</sup>。

收稿日期: 2021-06-02 修订日期: 2021-08-14

基金项目: 国家自然科学基金资助项目 (52075149); 河南省科技攻关计划项目 (212102110029); 现代农业装备与技术教育部重点实验室和江苏省农业装备与智能化高技术重点实验室开放基金课题 (JNZ201901); 河南省高等教育教学改革研究与实践项目 (研究生教育) 成果 (2019SJJGLX063Y)  
作者简介: 张伏, 教授, 研究方向为农业信息化与农业装备仿生技术。  
Email: zhangfu30@126.com

近年来,随着计算机视觉与深度学习技术在农业领域的广泛应用,深度卷积神经网络(Deep Convolutional Neural Networks, DCNN)在果蔬目标检测中凸显出巨大的优越性,其主要分为两种,一种是以 RCNN<sup>[18]</sup>、Fast RCNN<sup>[19]</sup>、Faster RCNN<sup>[20]</sup>为代表的两阶段(two stage)目标检测方法,其算法思想是先获得目标区域建议框再在区域建议框中进行分类。闫建伟等<sup>[21]</sup>提出改进的 Faster RCNN 自然环境中不同成熟度和遮挡情况下刺梨果实的识别方法,识别召回率达 81.4%,准确率达 95.53%,F1 值达 94.99%,识别速度为 0.2 s/张,其模型精度高、泛化能力强,但两阶段(two stage)目标检测方法生成候选区域步骤需占用大量资源,检测时间较长。另一种方法是以 SSD(Single Shot MultiBox Detector)<sup>[22]</sup>、YOLO(You Only Look Once)<sup>[23]</sup>等为代表的单阶段(one stage)目标检测方法,其算法思想是目标建议框与分类标签在同一网络下完成。成伟等<sup>[24]</sup>基于改进的 YOLOv3 网络识别温室番茄,识别模型估产平均精度(mean Average Precision, mAP)为 95.7%,单幅图像处理耗时 15 ms,并且对密集和受遮挡的果实具有更好的识别效果。吕石磊等<sup>[25]</sup>通过引入 GIoU 边框回归损失函数和,使用 MobileNet-v2 作为模型的骨干网络,构建了基于改进型 YOLOv3-LITE 轻量级神经网络,对自然环境下密集、叶片遮挡严重的柑橘进行识别定位,在全部测试集上调和均值为 93.69%,平均精度值为 91.13%,对单张 416×416(像素)的图片识别速度为 16.9 ms,因此使用 YOLO 算法对果实进行识别,在保证较高的识别准确率情况下,又具备检测速度快、内存占用少等特点<sup>[26-29]</sup>。

为解决设施农业环境下密集度高、果实小、遮挡粘连严重的圣女果精确、快速识别问题,提出改进型 YOLOv4-LITE 轻量级神经网络,在特征融合结构中引入实现对小目标检测的特征层,增加模型细粒度检测,并进行剪枝操作和引入深度可分离卷积结构压缩网络量,加快识别速度,同时通过载入预训练权重并冻结部分层的训练方式提高模型精度,以期对圣女果快速精准识别提供参考。

## 1 材料与方法

### 1.1 试验数据获取

圣女果图像的采集地点为河南省洛阳市孟津区某采摘园基地温室内,拍摄时间为 2021 年 3 月 19 日上午 10:00-12:00。在白天自然光条件下采集,采集时,使用高清数码单反相机距圣女果果实 20~80 cm 的距离拍摄,为模拟采摘机器人识别系统,拍摄角度为左侧拍、右侧拍、俯拍、仰拍和正面拍共 5 个方向角度<sup>[30]</sup>,共采集成熟圣女果原始图像 249 张,挑选出 214 张。所采集圣女果图像大小为 6 000×4 000(像素),种类涵盖顺光、逆光、重叠、遮挡、粘连等各种实际温室中圣女果生长环境的图像。温室环境下的部分圣女果如图 1 所示,可以发现圣女果果实出现重度粘连、遮挡现象。

### 1.2 数据扩增

在深度学习模型训练阶段,数据采集的越充分、越

全面,模型识别效果越显著,因此通过数据扩增方法(Data augmentation)扩充样本数量。为更能模拟复杂场景下拍摄状态和应用到深度神经网络中,在 Keras 框架下结合 Opencv,先对采集数据裁剪和压缩至 416×416(像素),再对图像进行平移、随机旋转、镜像翻转、水平翻转、垂直翻转、颜色增强、亮度改变、添加高斯噪音等不同程度与组合的物理变换对图像进行数据增强<sup>[18,31]</sup>。增强后共得到 10 710 张图像作为数据集。

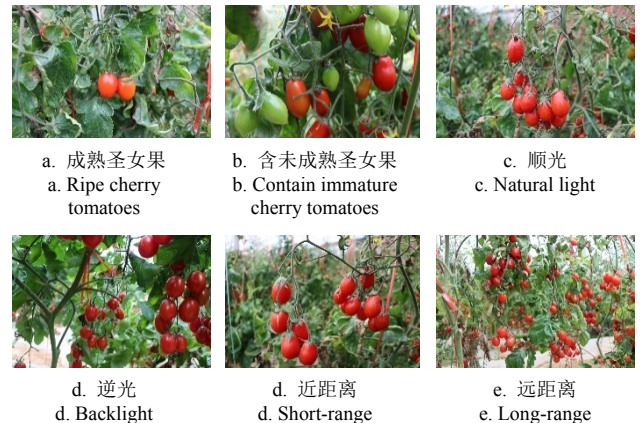


图 1 温室环境下的部分圣女果图像  
Fig.1 Part of cherry tomato images in greenhouse

### 1.3 数据集准备

使用 LabelImg 工具对图像中圣女果果实进行手工标记矩形框。标注时,对完全裸露的圣女果切于矩形框内侧进行标注,对遮挡或粘连的圣女果果实裸露于图像的部分进行矩形框标注,对图像边界出现部分和遮挡的圣女果程度小于 10%时进行无标注处理。标注完成后得到包含 ground truth 的.xml 文件,之后将数据集按照 9:1 比例划分,其中 90%部分再按照 9:1 分为训练集和验证集,剩余 10%作为测试集,最终得到训练集、验证集和测试集的样本数量分别为 8 674、965、1 071 张。

## 2 密集圣女果识别网络

### 2.1 YOLOv4 网络模型

相对于 R-CNN、Fast R-CNN、Faster R-CNN 等两阶段(two-stage)的目标检测方法,YOLOv4<sup>[32]</sup>作为经典的单阶段(one-stage)深度识别网络模型,直接在网络中生成预测物体分类(Classification)和边界框(Bounding box),极大提高了目标检测速度,其结构简图如图 2 所示。YOLOv4 网络模型由 3 大部分组成:主干网络(Backbone network)、颈部网络(Neck network)、头部网络(Head network)。主干网络为 CSPDarknet53 特征提取网络,由 CSP1~CSP5 共 5 个模块构成,每个模块由 CSPX 模块和 CBM 或 CBL 模块相互交替堆叠而成,YOLOv4 在主干网络中使用了 CSPnet 结构,如图 2c 所示,在残差块 X Res unit 堆叠的另一边经 CMB 处理后形成一个大的残差边,增强了 CNN 的学习能力。在 CSPnet 结构中又引入 Mish 激活函数替代 Leaky ReLU 激活函数,Mish 函数具有无上界、有下界、非单调、无穷阶连续性、

平滑性等特点, 有助于模型实现正则化、稳定网络梯度流。Mish 函数表达式为

$$\text{Mish} = x \cdot \tanh(\ln(1 + e)) \quad (1)$$

式中  $x$  为输入值,  $\tanh$  为双曲正切函数,  $\ln$  为以常数  $e$  为底数的对数函数。

经过 CSPDarknet53 主干网络得到  $52 \times 52 \times 256$  (特征层 P3)、 $26 \times 26 \times 512$  (特征层 P4)、 $13 \times 13 \times 1024$  这 3 个特征层, 其中对  $13 \times 13 \times 1024$  特征层在空间金字塔池化 (Spatial Pyramid Pooling, SPP) 中分别进行  $13 \times 13$ 、 $9 \times 9$ 、 $5 \times 5$ 、 $1 \times 1$  四个不同尺度的最大池化处理, 然后通过 Concatenate 运算整合得到特征层 P5, SPP 结构极大地增大感受野, 获得了更多上下文特征。P3、P4、P5 三个特征层先经过特征金字塔网络 (Feature Pyramid Networks, FPN) 进行自下向上的上采样 (Upsample) 融合, 但 FPN 融合低、高层特征方法路径长, 细节信息传递困难, 因此 YOLOv4 设计路径聚合网络 (Path Aggregation Network, PANet) 通过下采样 (Downsample) 对 3 个特征层完成自上向下的路径增强, 极大缩短了信息传播路径, 同时利用了低层特征的精准定位信息。YOLOv4 使用了 Mosaic 数据增强、Label Smoothing 平滑、CIOU 回归损失函数和学习率余弦退火衰减法等训练技巧。

YOLOv4 的损失函数 Loss 包括回归损失函数 Loss

(coord)、置信度损失函数 Loss (conf) 和分类损失函数 Loss (cls)。损失函数公式如下:

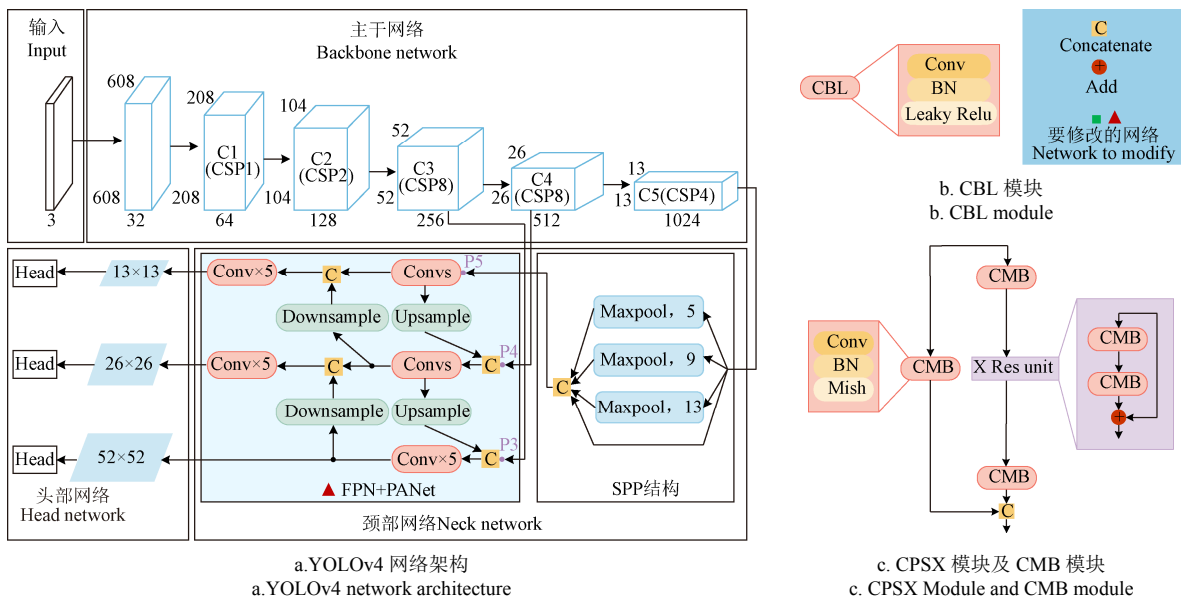
$$\text{Loss} = \text{Loss}(\text{coord}) + \text{Loss}(\text{conf}) + \text{Loss}(\text{cls}) \quad (2)$$

$$\text{Loss}(\text{coord}) = \lambda_{\text{coord}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} (2 - w_i \times h_i) [L_{\text{CIOU}}] \quad (3)$$

$$\begin{aligned} \text{Loss}(\text{conf}) &= - \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{obj}} [\hat{C}_i \lg C_i + (1 - \hat{C}_i) \lg(1 - C_i)] \\ &\quad - \lambda_{\text{noobj}} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{\text{noobj}} [\hat{C}_i \lg C_i + (1 - \hat{C}_i) \lg(1 - C_i)] \end{aligned} \quad (4)$$

$$\begin{aligned} \text{Loss}(\text{cls}) &= - \sum_{i=0}^{K \times K} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} [\hat{p}_i(c) \lg p_i(c) \\ &\quad + (1 - \hat{p}_i(c)) \lg(1 - p_i(c))] \end{aligned} \quad (5)$$

式中  $K$  为网格大小,  $I$  表示 feature map 的第  $i$  个方格,  $j$  表示该方格预测出的第  $j$  个预测框,  $w$  和  $h$  分别代表 ground truth 的宽度和高度, 角标 obj 和 noobj 分别表示第  $i$  个方格中有、无对象,  $C_i$ 、 $\hat{C}_i$  分别表示预测和真实框的类别,  $p_i(c)$  为预测目标的置信度,  $\hat{p}_i(c)$  为实际目标的置信度,  $\lambda_{\text{coord}}$  和  $\lambda_{\text{noobj}}$  为惩罚系数,  $L_{\text{CIOU}}$  为 bounding box 的回归损失函数。



注: CSP 表示跨阶段部分 (Cross stage partial) 结构, Conv 表示卷积 (convolutional), BN 表示批量正则化 (Batch Norm), CBL 表示 Conv+批量正则化 (Batch Norm, BN)+Leaky relu 激活函数合成模块, CBM 表示 Conv+批量正则化 (Batch Norm, BN)+Mish 激活函数合成模块, Maxpool 表示池化操作。  
Note: CSP stands for Cross stage partial structure, Conv stands for convolutional, BN stands for Batch Norm, and CBL stands for Conv+Batch Norm (BN)+Leaky relu activation function synthesis module, CBM means Conv+Batch Norm (BN)+Mish activation function synthesis module, Maxpool represents the pooling operation.

图 2 YOLOv4 网络整体框架

Fig.2 YOLOv4 network holistic framework

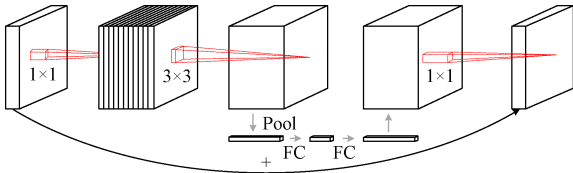
## 2.2 改进的 YOLOv4-LITE 网络模型设计

YOLOv4 网络含有 CSPnet 结构的 CSPDarknet53 网络, 虽降低了参数量, 获得更快的速度, 但模型计算仍复杂, 需消耗较多的内存空间。本研究提出一种基于目标检测的轻量级网络模型, 在传统的 YOLOv4 网络基础上, 将 MobileNet-V3 网络作为特征提取主干网络, 构建

YOLOv4-LITE 轻量级网络模型。MobileNet 是一种基于移动端和嵌入式设备的高效 CNN 模型, 具有更轻量级、更快速的特点。MobileNet-V3 既继承了 MobileNet-V1 的深度可分离卷积 (depthwise separable convolutions), 又融合了 MobileNet-V2 的具有线性瓶颈的逆残差结构 (the inverted residual with linear bottleneck) 的结构特点,



MobileNet-V3 由多个 block 堆叠而成, block 的 bneck 结构如图 3 所示, 其依次经过  $1 \times 1$ 、 $3 \times 3$ 、 $1 \times 1$  的卷积提升通道数量、进行深度卷积和降低维度。MobileNet-V3 在 bottleneck 结构中引入 SE (Squeeze and Excitation) 结构的轻量级注意力机制, 并基于 MobileNet-V2 的结构: 将头部的卷积核通道数量由 32 降低至 16; 在尾部通过 avg pooling 将特征图大小由  $7 \times 7$  降到了  $1 \times 1$ , 并且舍弃了纺锤形的  $3 \times 3$  和  $1 \times 1$  的卷积, 在减少消耗时间的同时又提高了精度。



注: FC 为全连接网络; Pool 为池化。

Note: FC is a fully connected network; Pool is a pooling.

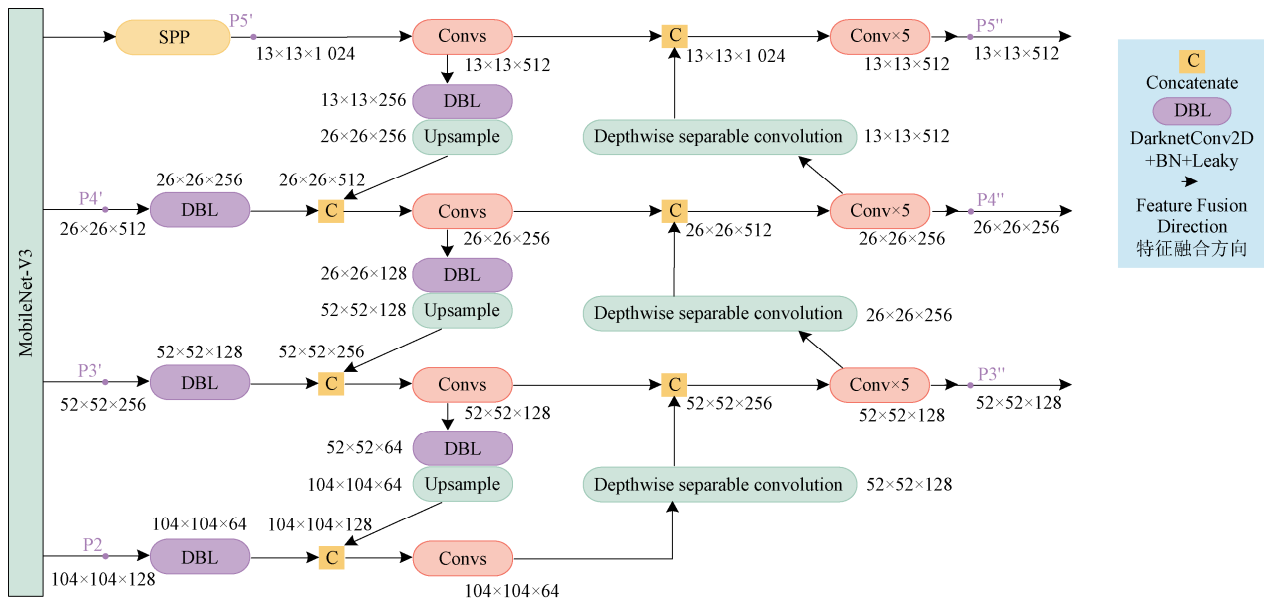
图 3 MobileNet-V3 的 bneck 结构图

Fig.3 The bneck structure diagram of MobileNet-V3

圣女果作为识别目标, 其尺度各有差异、不尽相同, 因此为避免使用 MobileNet-V3 骨干网络降低对小目标的检测精度, 通过修改 FPN 网络, 如图 4 所示, 从 YOLOv4-LITE 轻量级神经网络的主干网络 MobileNet-V3 中, 输出  $13 \times 13$  (特征层 P5')、 $26 \times 26$  (特征层 P4')、 $52 \times 52$  (特征层 P3') 和  $104 \times 104$  (特征层 P2) 共 4 个尺

度, P5'感受野较大适合大尺寸目标检测, P4'适合中等目标检测, 在 P3'基础上进行上采样, 融合 P2 特征层后获得丰富的浅层信息, 使之对小目标检测更加敏感, 实现细粒度的检测。在特征传播过程中  $13 \times 13$  尺度特征层仍然经过 SPP 结构得到特征层 P5', 本研究将特征层 P5'、P4'、P3'和 P2 在 FPN 结构中通过上采样进行不同金字塔级别的 Feature map 结合, 每一个特征层经过 DBL 和 Upsample 操作变换, 获得与上一个特征层相同的尺度和通道数量, 然后经过 Concatence 与上一个特征层融合获得信息更加丰富的 Feature map。为防止网络过于冗余, 对 FPN 特征融合后输出的 4 个特征层进行剪枝操作, 也即将 FPN 输出的  $104 \times 104$  尺度特征层不再进行 YOLO Head 的预测输出, 直接在 PANet 结构上采样, 因此改进算法仍保留了 YOLO Head 的  $13 \times 13$ 、 $26 \times 26$ 、 $52 \times 52$  三个尺度的特征层 (P5''、P4''、P3'') 预测输出。

但在改进 FPN 网络, 增加  $104 \times 104$  特征层负责检测小目标的同时, 网络数量必然会相应增加, 为降低网络运算数量, 在 PANet 结构中引入深度可分离卷积 (Depthwise separable convolution) 代替原网络 Downsample 中普通卷积进行下采样, 实现自上而下的特征信息交互, 有效减少网络计算量和参数; 同时深度可分离卷积可由自身的  $1 \times 1$  Filter 数量任意决定输出通道数量, 解除了普通卷积核个数和步长的限制, 代替了池化操作的作用, 在节省内存消耗的同时也提高了模型精度<sup>[33]</sup>。



注: P5'、P4'、P3'和 P2 为改进 FPN+PANet 结构的输入特征层; P5''、P4''、P3''为改进 FPN+PANet 结构的输出。

Note: P5', P4', P3', and P2 are the input feature layers of the improved FPN+PANet structure, and P5'', P4'', P3'' are the output of the improved FPN+PANet structure.

图 4 改进的 FPN+PANet 结构图

Fig.4 Improve FPN+PANet structure diagram

## 2.3 模型的训练与测试

### 2.3.1 试验平台

本研究使用 Tensorflow 和 Keras 框架来改进 YOLOv4 网络。试验环境如表 1 所示。

### 2.3.2 密集圣女果识别网络训练

密集圣女果目标检测网络流程如图 5 所示。通过对比改进模型的不同修改之处进行对比试验, 并在相同验

证集上验证模型的效果。首先对采集的数据进行预处理, 然后在 Labellmg 工具上进行目标圣女果手工标注, 标注时对图像边界出现部分和遮挡的圣女果程度小于 10% 时进行无标注处理, 对标注的图像数据与对应的.xml 文件同时进行扩增, 并以 PASCALVOV 数据格式保存, 分别采用使用预训练权重、不使用预训练权重的训练方法进行网络模型的训练, 使用预权重训练时先冻结预训练权

重网络层,将更多的资源放在后面网络的参数训练上,后解冻这部分网络层的训练方式,通过先冻结部分层后解冻的训练方式可有效保证权值。模型训练时,采用 Keras 训练中的早停法(early stopping)技巧,当模型在训练集上的损失值不再下降时,停止训练,有效防止过拟合。

表 1 试验环境  
Table 1 Experimental environment

配置 Configuration	参数 Parameter
CPU	Intel(R) Xeon(R) Silver 4210R
GPU	NVIDIA Quadro P2200 16 G
操作系统	Windows10
加速环境	cuda10.0、CUDNN7.4.1.5
开发平台	PyCharm
其他	Numpy1.17.4、Opencv3.4.1

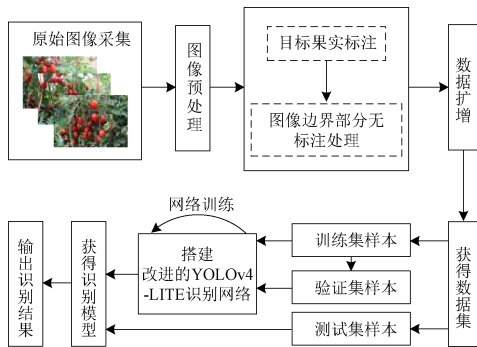


图 5 密集圣女果检测网络流程图

Fig.5 Flowchart for dense cherry tomatoes detection network

网络训练参数设置。不使用预训练权重训练时,模型超参数设置批样本(batch size)数量为 2,动量因子(Momentum)为 0.9,权值初始学习率(learning rate)为 0.001,衰减系数为 0.000 5;使用预训练权重时,模型冻结层超参数设置代训练(epoch)为 50,批样本数量为 8,动量因子为 0.9,权值初始学习率为 0.000 1,衰减系数为 0.000 5,解除冻结后超参数设置代训练为 50,批样本数量为 2,动量因子为 0.9,权值初始学习率为 0.001,衰减系数为 0.000 5,总代训练为 100。2 种训练方式均采用 BN(batch normalization)正则化进行网络层权值更新,在训练集上每经过一代训练(epoch)保存一次权重文件,并生成日志文件输出训练集和验证集的损失值。训练集损失和验证集损失曲线如图 6 所示。

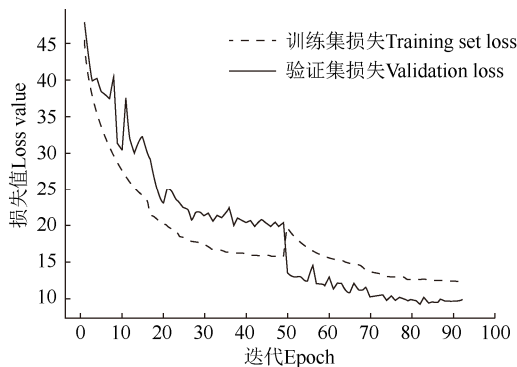


图 6 损失值随迭代次数变化曲线

Fig.6 Loss value change curve with epoch times

在前 50 代训练时验证集损失虽在前期出现震荡但总体趋于减小趋势,训练集损失持续下降;在第 50 代训练时由于解冻训练使训练集损失升高,验证集损失减少;在第 50 代训练之后由于验证集损失的减小,训练集损失不断减小,当验证集损失不再发生大幅变化时,训练集损失也不再变化,模型收敛。

### 2.3.3 模型测试

为客观衡量模型对密集圣女果的目标检测效果,使用调和均值 F1 值(F1-score)、召回率(Recall)、准确率(Precision)、检测速度、网络参数量、权重大小来评价训练后的模型,其中 F1、Recall、Precision 计算公式如公式(6)~(8)所示。

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

式中 TP 为真实的正样本数量,FP 为虚假的正样本数量,FN 为虚假的负样本数量。平均精度(Average Precision, AP)计算公式如式(9)所示。

$$AP = \int_0^1 \text{Precision} \times \text{Recall} dr \quad (9)$$

式中  $r$  为积分变量,是对召回率与精确度乘积的积分。AP 为 PR(Precision-Recall)曲线与坐标轴围成的面积,取值在 0~1 之间。 $AP_{50}$  为重叠度(Intersection Over Union, IOU)=0.5 时不同查全率下的精度平均值; $AP_{75}$  为 IOU=0.75 时不同查全率下的精度平均值。

## 3 结果与分析

为验证密集圣女果目标识别算法的性能,分别对是否载入预训练权重并冻结部分层方法和不同骨干网络在相同密集圣女果数据集和参数设置上进行训练,之后对提出的模型的部分改进之处进行对比试验分析。

### 3.1 载入预训练权重并冻结部分层的检测结果

以 YOLOv4-LITE 网络为对象网络,分别对其进行使用预训练权重并冻结部分层的训练方式和不使用预训练权重的训练方式的 2 种方式进行训练。不使用预训练权重即在训练集上对所有参数进行初始训练;使用预训练权重并冻结部分层的训练方式为载入 COCO 数据集的 MobileNet-V3 权重,初始训练冻结 YOLOv4-LITE 中的 MobileNet-V3 网络层,先训练后半部分网络,在训练 50 代后开始解冻冻结部分网络层,进行全部网络的训练。表 2 比较了 2 种训练策略的结果。

试验结果表明,相比于不使用预训练权重,使用预训练权重并冻结部分层的训练方式在 IOU=0.50 和 IOU=0.75 时平均精度值提升约 3 个百分点,同时调和均值和准确率也有不同程度的提升。

### 3.2 不同骨干特征提取网络的比较

为满足设施农业圣女果采摘机器人与相关嵌入式移

动端检测设备的实际生产需求，以 MobileNet-V3 作为骨干网络设计 YOLOv4-LITE 轻量级神经网络。为验证设计的 YOLOv4-LITE 的合理性，在 1 071 张相同测试集上进行对比试验，试验结果如表 3 所示。从表 3 可以看出，相较于原 YOLOv4 以 CSPDarkNet-53 作为骨干网络，YOLOv4-LITE 以 MobileNet 系列轻量级网络作为骨干网络，权重大小、检测速度和网络参数量都有大幅改善，但 MobileNet-V1 和 MobileNet-V2 作为骨干网络，在 IOU=0.50 和 IOU=0.75 时，大部分 F1 值、AP、Recall 和 Precision 均低于原 YOLOv4 网络。采用 MobileNet-V3 作为骨干网络，模型在 F1 值、AP 和 Precision 指标上均有不同程度的提升，在 IOU=0.50 时，且相对于原 YOLOv4 方法在 IOU=0.50 时，F1 值提升了 0.01，AP 值提升了 1.71 个百分点，准确率提升了 3.69 个百分点，对单张 416×416（像素）图像的检测速度高达 2.78 ms/张，权重大小减小了 142 MB，网络参数量也减少了 37.82%，因此 YOLOv4-LITE 轻量级神经网络在移动设备或嵌入

式终端上应用具有明显优势，但在 IOU=0.75 时的 AP 为 43.94%低于原 YOLOv4 网络的 50.75%，可能是由于在目标识别过程中，对于小目标的特征检测不够充分。

表 2 不同方式训练结果比较

Table 2 Comparison of training results in different ways			
训练方式及重叠度值 Training method and Intersection over Union value	F1 值 F1 score	平均精度 Average precision/%	准确率 Precision/%
使用预训练权重且重叠度为 0.50 The pre-training weight is used and IOU=0.50	0.85	93.16	96.00
不使用预训练权重且重叠度为 0.50 The pre-training weight is not used and IOU=0.50	0.82	90.05	93.73
使用预训练权重且重叠度为 0.75 The pre-training weight is used and IOU=0.75	0.50	43.94	47.81
不使用预训练权重且重叠度为 0.75 The pre-training weight is not used and IOU=0.75	0.47	40.89	45.84

表 3 不同骨干网络检测结果对比

Table 3 Comparison of detection results of different backbone networks

网络模型 Network models	骨干网络 Backbone networks	IOU 得分 IOU score	F1 值 F1 score	平均精度 Average precision/%	召回率 Recall/%	准确率 Precision/%	权重大小 Weight size/MB	检测速度 Detection speed (ms·张 <sup>-1</sup> )	网络参数量 Network parameter
YOLOv4	CSPDarkNet-53	0.50	0.84	91.45	77.26	92.61	244.0	4.62	64 407 389
		0.75	0.46	50.75	50.75	42.86			
YOLOv4-LITE	MobileNet-V1	0.50	0.77	80.21	63.72	85.37	105.0	3.39	41 005 757
		0.75	0.43	22.45	45.37	30.76			
YOLOv4-LITE	MobileNet-V2	0.50	0.83	90.28	75.23	92.08	98.4	3.31	39 124 541
		0.75	0.58	57.96	71.66	48.20			
YOLOv4-LITE	MobileNet-V3	0.50	0.85	93.16	75.72	96.30	102.0	2.78	40 043 389
		0.75	0.50	43.94	52.52	47.81			

3.3 FPN+PANet 结构的改进结果比较

在原始的 YOLOv4 算法检测圣女果目标时，其尺度各有差异、不尽相同，往往由于识别目标密集度高、重叠和目标小等问题，出现漏检、误检的情况。因此提出改进型 YOLOv4-LITE 轻量级密集圣女果识别神经网络模型，具体改进方式如下：①骨干网络替换。以 MobileNet-V3 作为骨干网络。②改进 FPN 结构。在 FPN

结构中增加 104×104 特征层。兼顾大、中目标检测的同时，增加了有利于小目标检测的 Feature map。③改进 PANet。在 PANet 引入深度可分离卷积，为减少因增加小目标检测 Feature map 后网络参数量的增大。为验证改进型 YOLOv4-LITE 中 FPN+PANet 结构改进的优越性，在 1 071 张相同测试集上进行对比试验，试验结果如表 4 所示。

表 4 不同改进结构检测结果对比

Table 4 Comparison of test results of different improved structures

算法 Algorithms	IOU 得分 IOU score	F1 值 F1 score	平均精度 Average precision/%	召回率 Recall/%	准确率 Precision/%	权重大小 Weight size/MB	检测速度 Detection speed (ms·张 <sup>-1</sup> )	网络参数量 Network parameter
YOLOv4+②	0.50	0.98	99.74	97.10	99.16	248.0	4.89	65 364 061
	0.75	0.70	65.76	77.66	63.69			
YOLOv4+①+②	0.50	0.98	99.70	97.46	98.87	103.0	3.60	41 000 061
	0.75	0.63	65.72	75.08	52.04			
YOLOv4+①+②+③ (本文改进网络)	0.50	0.99	99.74	98.00	99.15	45.3	3.01	12 026 685
	0.75	0.79	67.19	78.08	78.99			

注：①骨干网络替换；②改进 FPN 结构；③改进 PANet。

Note: ① Replace backbone network; ② Improve FPN structure; ③ Improve PANet.

在 YOLOv4 基础上改进 FPN 结构比原 YOLOv4 的 AP<sub>50</sub> 提高了 8.29 个百分点，AP<sub>75</sub> 提高了 15.01 个百分点，F1 值在相应的 IOU 阈值下相应提升了 0.14 和 0.24，但其权重大小增加了 4 MB，检测速度增加了 0.27 ms/张，网络参数量增加了 14.85%。与只替换骨干网络(YOLOv4-LITE+

MobileNet-V3)相比，骨干网络替换+改进 FPN 结构网络，AP<sub>50</sub> 提高了 6.58 个百分点，AP<sub>75</sub> 提高了 21.82 个百分点，F1 值在相应的 IOU 阈值下相应提升了 0.13 和 0.20，但其权重大小增加了 146 MB，检测速度增加了 2.11 ms/张，网络参数量增加了 63.23%。说明 YOLOv4 和 YOLOv4+ MobileNet-V3



通过增加小目标的 Feature map 提高了模型细粒度的检测, 但模型权重大小、检测速度和参数量会相应增大。

骨干网络替换+改进 FPN 结构+改进 PANet 网络, 在保证 F1、AP、Recall 和 Precision 较高的同时, 模型权重大小为 45.3 MB, 检测速度为 3.01 ms/张, 网络参数量为 12 026 685。相比于 YOLOv4 网络, 其模型权重大小减小了 198.7 MB, 检测速度减少了 1.61 ms/张, 模型参数量减少了 81.33%; 相比于 YOLOv4+MobileNet-V3, 其模型权重大小减小了 56.7 MB, 检测速度增加了 0.23 ms/张, 模型参数量减少了 69.97%。说明改进 PANet 的策略在不影响精度等情况下, 可有效减小内存消耗、降低模型参数量、加快模型识别速度。

总体来看, 本研究提出的改进型 YOLOv4-LITE 轻量

级网络, 与 YOLOv4 网络相比, 在 IOU=0.50 时, F1 值提升了 0.15, AP 提高了 8.29 个百分点, 准确率提高了 6.55 个百分点, 权重大小约为 YOLOv4 的 1/5, 检测速度降低了 34.85%。

图 7 对比了 YOLOv4 网络与改进型 YOLOv4-LITE 轻量级网络在设施环境下对密集圣女果的识别结果, 图中深色框标注为算法对圣女果的识别结果, 浅色框为对识别结果图像的人工标注, 表明有个别果实未能识别出来, 即算法间对果实识别的差异。从图中可以看出, YOLOv4 在识别高密度圣女果时, 对高度粘连和遮挡的圣女果果实出现漏检的情况, 对较小目标未能成功检测识别, 相比之下, 本文所提出的算法对高度粘连、严重遮挡和小目标果实有较高识别率和良好的泛化性能。



注: 深色框标注为算法对圣女果的识别结果; 浅色框为对识别结果图像的人工标注。

Note: The dark box is labeled as the algorithm's recognition result of the cherry tomatoes, and the yellow box and the light box are the artificial labels of the recognition result image.

图 7 不同算法对密集圣女果的识别结果

Fig.7 Recognition results of dense cherry tomatoes by different algorithms

## 4 结 论

1) 提出了一种改进型的 YOLOv4-LITE 轻量化神经网络检测高度密集、严重粘连的圣女果目标识别算法。利用 MobileNet-V3 作为骨干网络构建 YOLOv4-LITE 轻量化网络, 在特征融合层 (Feature Pyramid Networks, FPN) 结构中引入小目标检测的特征层以增加图像检测细粒度, 对并在 PANet 网络中引入深度可分离卷积网络代替传统网络使改进网络更轻量化, 更利于嵌入式设备和移动端的部署, 为实现采摘机器人自动采摘提供理论依据。并通过采用载入预训练权重并冻结部分层的训练方式比不使用预训练权重平均精度提高约 3%。

2) 在 1 071 张相同测试集上以调和平均值、平均精度和准确率为判断依据, 通过对比试验验证提出方法的

可行性与优越性。与 YOLOv4 网络相比, 通过替换骨干网络、改进特征金字塔网络和改进路径聚合网络, 在 IOU=0.50 时, 本文提出的方法调和均值提升了 0.15, 平均精度提高了 8.29 个百分点, 准确率提高了 6.55 个百分点, 权重大小约为 YOLOv4 的 1/5, 检测速度为 3.01 ms/张。通过对比试验, 验证了该方法具有显著优势。

## [参 考 文 献]

- [1] Tang Y C, Chen M Y, Wang C L, et al. Recognition and localization methods for vision-based fruit picking robots: A review[J]. Frontiers in Plant Science, 2020, 11: 1-17.
- [2] 赵献立, 王志明. 机器学习算法在农业机器视觉系统中的应用[J]. 江苏农业科学, 2020, 48(12): 226-231.

Zhao Xianli, Wang Zhiming. Application of machine learning algorithm in agricultural machine vision system[J]. Jiangsu Agricultural Sciences, 2020, 48(12):

- 226-231. (in Chinese with English abstract)
- [3] 林伟明, 胡云堂. 基于 YUV 颜色模型的番茄收获机器人图像分割方法[J]. 农业机械学报, 2012, 43(12): 176-180. Lin Weiming, Hu Yuntang. Image segmentation method based on YUV color space for tomato harvesting robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(12): 176-180. (in Chinese with English abstract)
- [4] Ochida C O, Itodo A U, Nwanganga P A. A review on postharvest storage, processing and preservation of tomatoes (lycopersicon esculentum mill)[J]. Asian Food Science Journal, 2018, 6(2): 1-10.
- [5] 马翠花, 张学平, 李育涛, 等. 基于显著性检测与改进 Hough 变换方法识别未成熟番茄[J]. 农业工程学报, 2016, 32(14): 219-226. Ma Cuihua, Zhang Xueping, Li Yutao, et al. Identification of immature tomatoes base on salient region detection and improved Hough transform method[J]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(14): 219-226.
- [6] 李寒, 陶涵斌, 崔立昊, 等. 基于 SOM—K-means 算法的番茄果实识别与定位方法[J]. 农业机械学报, 2021, 52(1): 23-29. Li Han, Tao Hanxiao, Cui Lihao, et al. Recognition and localization method of tomato based on SOM—K-means algorithm. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1): 23-29.
- [7] Payne A B, Walsh K B, Subedi P P, et al. Estimation of mango crop yield using image analysis-segmentation method[J]. Computers and Electronics in Agriculture, 2013, 91: 57-64.
- [8] 李昕, 李立君, 高自成, 等. 改进类圆随机 Hough 变换及其在油茶果实遮挡识别中的应用[J]. 农业工程学报, 2013, 29(1): 164-170. Li Xin, Li Lijun, Gao Zicheng, et al. Revised quasi-circular randomized Hough transform and its application in camellia-fruit recognition[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(1): 164-170. (in Chinese with English abstract)
- [9] Gong A, Yu J, He Y, et al. Citrus yield estimation based on images processed by an Android mobile phone[J]. Biosystems Engineering, 2013, 115(2): 162-170.
- [10] Xu Y, Imou K, Kaizu Y, et al. Two-stage approach for detecting slightly overlapping strawberries using HOG descriptor[J]. Biosystems Engineering, 2013, 115(2): 144-153.
- [11] 谢忠红, 姬长英, 郭小清, 等. 基于改进 Hough 变换的类圆果实目标检测[J]. 农业工程学报, 2010, 26(7): 157-162. Xie Zhonghong, Ji Changying, Guo Xiaoqing, et al. Target detection of fruit-like fruit based on improved Hough transform[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(7): 157-162. (in Chinese with English abstract)
- [12] Zhao C, Lee W S, He D. Immature green citrus detection based on colour feature and sum of absolute transformed difference (SATD) using colour images in the citrus grove[J]. Computers and Electronics in Agriculture, 2016, 124: 243-253.
- [13] Liu S, Yang C, Hu Y, et al. A method for segmentation and recognition of mature citrus and branches-leaves based on regional features[C]//Chinese Conference on Image and Graphics Technologies. Singapore: Springer, 2018: 292-301.
- [14] 卢军, 桑农. 变化光照下树上柑橘目标检测与遮挡轮廓恢复技术[J]. 农业机械学报, 2014, 45(4): 76-81. Lu Jun, Sang Nong. Detection of citrus fruits within tree canopy and recovery for occlusion contour in variable illumination[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(4): 76-81. (in Chinese with English abstract)
- [15] Stein M, Bargoti S, Underwood J. Image based mango fruit detection, localisation and yield estimation using multiple view geometry[J]. Sensors, 2016, 16(11): 1915.
- [16] Xu Y, Imou K, Kaizu Y, et al. Two-stage approach for detecting slightly overlapping strawberries using HOG descriptor[J]. Biosystems engineering, 2013, 115(2): 144-153.
- [17] Wachs J P, Stern H I, Burks T, et al. Low and high-level visual feature-based apple detection from multi-modal images[J]. Precision Agriculture, 2010, 11(6): 717-735.
- [18] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014: 580-587.
- [19] Girshick R. Fast R-CNN[C]//IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440-1448.
- [20] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Annual Conference on Neural Information Processing Systems, Montreal, Canada, 2015: 91-99.
- [21] 闫建伟, 赵源, 张乐伟, 等. 改进 Faster-RCNN 自然环境下识别刺梨果实[J]. 农业工程学报, 2019, 35(18): 143-150. Yan Jianwei, Zhao Yuan, Zhang Lewei, et al. Recognition of Rosa roxbunghii in natural environment based on improved Faster RCNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(18): 143-150. (in Chinese with English abstract)
- [22] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European Conference on Computer Vision Springer, Cham, 2016: 21-37.
- [23] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 779-788.
- [24] 成伟, 张文爱, 冯青春, 等. 基于改进 YOLOv3 的温室番茄果实识别估产方法[J]. 中国农机化学报, 2021, 42(4): 176-182. Cheng Wei, Zhang Wenai, Feng Qingchun, et al. Method of greenhouse tomato fruit identification and yield estimation based on improved YOLOv3[J]. Journal of Chinese Agricultural Mechanization. 2021 42(4): 176-182. (in Chinese with English abstract)
- [25] 吕石磊, 卢思华, 李震, 等. 基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法[J]. 农业工程学报, 2019, 35(17): 205-214. Lü Shilei, Lu Sihua, Li Zhen, et al. Orange recognition method using improved YOLOv3-LITE lightweight neural network[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(17): 205-214. (in Chinese with English abstract)
- [26] 薛月菊, 黄宁, 涂淑琴, 等. 未成熟芒果的改进 YOLOv2 识别方法[J]. 农业工程学报, 2018, 34(7): 173-179. Xue Yueju, Huang Ning, Tu Shuqin, et al. Immature mango detection based on improved YOLOv2[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(7): 173-179. (in Chinese with English abstract)
- [27] 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 Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 164-173. (in Chinese with English abstract)
- [28] 张健. 基于改进 YOLOv3 的果园行人检测方法研究[D]. 镇江: 江苏大学, 2020.
- Zhang Jian. Research on Orchard Pedestrian Detection Method based on Improved YOLOv3[D]. Zhenjiang: Jiangsu University, 2020. (in Chinese with English abstract)
- [29] 蔡逢煌, 张岳鑫, 黄捷. 基于 YOLOv3 与注意力机制的桥梁表面裂痕检测算法[J]. 模式识别与人工智能, 2020, 33(10): 926-933.
- Cai Fenghuang, Zhang Yuexin, Huang Jie. Bridge surface crack detection algorithm based on YOLOv3 and attention mechanism[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(10): 926-933. (in Chinese with English abstract)
- [30] Hu X L, Liu Y, Zhao Z X, et al. Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network[J]. Computers and Electronics in Agriculture, 2021, 185: 106135.
- [31] 李就好, 林乐坚, 田凯, 等. 改进 Faster R-CNN 的田间苦瓜叶部病害检测[J]. 农业工程学报, 2020, 36(12): 179-185.
- Li Jiu hao, Lin Lejian, Tian Kai, et al. Detection of leaf diseases of balsam pear in the field based on improved Faster R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(12): 179-185. (in Chinese with English abstract)
- [32] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004. 10934.
- [33] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]. // IEEE/CVF International Conference on Computer Vision, Seoul, SOUTH KOREA, 2019: 1314-1324.

## Recognition of dense cherry tomatoes based on improved YOLOv4-LITE lightweight neural network

Zhang Fu<sup>1,2,3</sup>, Chen Zijun<sup>1</sup>, Bao Ruofei<sup>1</sup>, Zhang Chaochen<sup>1</sup>, Wang Zhihao<sup>1</sup>

(1. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, China; 2. Key Laboratory of Modern Agricultural Equipment and Technology of Ministry of Education, Jiangsu University, Zhenjiang 212013, China; 3. Collaborative Innovation Center of Advanced Manufacturing of Machinery and Equipment of Henan Province, Luoyang 471003, China)

**Abstract:** Small and hidden conditions of dense cherry tomatoes have posed a great challenge to the rapid identification and positioning of fruits. New key technology with strong robustness is highly demanding to improve the efficiency and yield prediction of cherry tomatoes in the facility agriculture environment. In this study, a novel recognition method was proposed to locate the dense cherry tomatoes using an improved YOLOv4-LITE lightweight neural network. A mobile Net-v3 easy migration to mobile terminals was selected as the feature extraction network of the model to construct a YOLOv4-LITE for a higher detection speed of cherry tomatoes. A feature pyramid network was set as the modified (FPN) + Path Aggregation Network (PANet) structure, in order to avoid replacing the backbone network to reduce the detection accuracy. Specifically, a 104×104 Future map was introduced to achieve fine-grained detection for the small targets. More importantly, a deep separable convolution was used in the PANet structure to reduce the number of model calculations. The new network was more lightweight, where the generalization ability of the model was improved by loading pre-training weights and freezing partial layer training. A comparison was made on the recognition effect of YOLOv4, F1, and AP on the test set with the same degree of occlusion or adhesion, further to evaluate the difference between the models. The test results show that the improved FPN structure on the basis of YOLOv4 was higher than the AP<sub>50</sub> of the original YOLOv4 AP<sub>75</sub> increased by 15.00 percentage points, and the F1 increased by 0.14 and 0.24 under the corresponding IOU threshold. However, the weight increased by 4 MB, while the detection speed increased to 0.27 ms/sheet, and the number of network parameters increased by 14.85%. The improved FPN structure on the basis of YOLOv4+MobileNet-V3, AP<sub>50</sub> increased by 6.58 percentage points, AP<sub>75</sub> increased by 21.82 percentage points, F1 value increased by 0.13 and 0.20 under the corresponding IOU threshold, indicating that YOLOv4 and YOLOv4+MobileNet-V3 lacked small goals. Fortunately, the Future map of small targets was added to improve the fine-grained detection of the model, but the number of model parameters and weights increased accordingly. As such, the PANet structure was improved to introduce a deep separable convolutional network, while ensuring high F1, AP, Recall and Precision. Optimal performance was achieved, where the model weight was compressed to 45.3 MB, the detection speed was 3.01 ms/sheet, and the network parameters were 12 026 685. Specifically, the new network was reduced by 198.7MB, compared with the original YOLOv4. The data indicated that the improved PANet strategy presented similar accuracy under such circumstances, while effectively reduced memory consumption, and the number of model parameters, but accelerated the speed of model recognition. The F1, AP<sub>50</sub>, and recall of the proposed recognition model for the dense cherry tree on all test sets were 0.99, 99.74%, and 99.15%, respectively. The improved YOLOv4 increased by 0.15, 8.29, and 6.55 percentage points, respectively, and the weight size was 45.3MB, about 1/5 of YOLOv4. Additionally, the detection of a single 416×416 image reached a speed of 3.01ms/frame on the GPU. Therefore, the recognition model of dense cherry tomatoes behaved a higher speed of recognition, a higher accuracy, and lighter weight than before. The finding can provide strong support to the efficient production forecast of cherry tomatoes in the facility agriculture environment.

**Keywords:** computer vision; models; YOLO; deep learning; image recognition; target detection