

# 基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法

吕石磊<sup>1,2,3</sup>, 卢思华<sup>1</sup>, 李震<sup>1,2,3\*</sup>, 洪添胜<sup>1,2,4</sup>, 薛月菊<sup>1,2,3</sup>, 吴奔雷<sup>1</sup>

(1. 华南农业大学电子工程学院, 广州 510642; 2. 国家柑橘产业技术体系机械化研究室, 广州 510642;  
3. 广东省农情信息监测工程技术研究中心, 广州 510642; 4. 华南农业大学工程学院, 广州 510642)

**摘要:** 柑橘识别是实现柑橘园果实自动采摘、果树精细化管理以及实现果园产量预测的关键技术环节。为实现自然环境下柑橘果实的快速精准识别, 该文提出一种基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法。在采摘机器人领域, 果实识别回归框的准确率直接决定了机器手的采摘成功率, 该方法通过引入 GIoU 边框回归损失函数来提高果实识别回归框准确率; 为便于迁移到移动终端, 提出一种 YOLOv3-LITE 轻量级网络模型, 使用 MobileNet-v2 作为模型的骨干网络; 使用混合训练与迁移学习结合的预训练方式来提高模型的泛化能力。通过与 Faster-RCNN 以及 SSD 模型对比在不同遮挡程度的测试样本下模型的识别效果, 用 F1 值与 AP 值评估各模型的差异, 试验结果表明: 该文提出的模型识别效果提升显著, 对于果实轻度遮挡的数据集, 该文提出的柑橘识别模型的 F1 值和 AP 值分别为 95.27% 和 92.75%, Average IoU 为 88.65%; 在全部测试集上, F1 值和 AP 值分别为 93.69% 和 91.13%, Average IoU 为 87.32%, 在 GPU 上对柑橘目标检测速度可达 246 帧/s, 对单张 416×416 的图片推断速度为 16.9 ms, 在 CPU 上检测速度可达 22 帧/s, 推断速度为 80.9 ms, 模型占用内存为 28 MB。因此, 该文提出的柑橘识别方法具有模型占用内存低、识别准确率高及识别速度快等优点, 可为柑橘采摘机器人以及柑橘产业产量预测提出新的解决方案, 为柑橘产业智能化提供新的思路。

**关键词:** 神经网络; 果树; 算法; 柑橘; YOLOv3-LITE; 混合训练; 迁移学习; GIoU 边框回归损失函数

doi: 10.11975/j.issn.1002-6819.2019.17.025

中图分类号: TP301.6; S274.2

文献标志码: A

文章编号: 1002-6819(2019)-17-0205-10

吕石磊, 卢思华, 李震, 洪添胜, 薛月菊, 吴奔雷. 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. doi: 10.11975/j.issn.1002-6819.2019.17.025 <http://www.tcsae.org>  
Lü Shilei, Lu Sihua, Li Zhen, Hong Tiansheng, Xue Yueju, Wu Benlei. 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) doi: 10.11975/j.issn.1002-6819.2019.17.025 <http://www.tcsae.org>

## 0 引言

随着农业机械化与信息化的兴起, 计算机视觉以及深度学习技术的不断突破, 果实识别、机器人自动采摘、果实产量预测是近年来的研究热点<sup>[1]</sup>。在智慧农业方面<sup>[2]</sup>, 人们通过对果实的识别来实现果树的精细化管理。柑橘产业是我国南方水果产业的重要支柱<sup>[3]</sup>, 由于我国柑橘果园环境复杂, 果实密集度高、叶片遮挡严重, 且柑橘果树树冠较低, 果实阴影程度较高, 为实现柑橘果树的自动采摘, 柑橘果实的精准识别研究至关重要。

近年来, 针对于自然环境下柑橘目标识别问题, 国内外相关文献基于传统机器视觉技术提出了多种解决方法。Hussin 等<sup>[4]</sup>和谢忠红等<sup>[5]</sup>分别使用圆形 Hough 变换进行柑橘目标检测, 检测精度较低, 且在柑橘密集、重叠的情况下, 容易导致错误检测。熊俊涛等<sup>[6]</sup>提出利用

K-means 聚类分割法结合优化 Hough 圆拟合方法实现柑橘果实分割, 利用直线检测确定柑橘采摘点, 但误检率会随目标密集度变化而增加。卢军等<sup>[7-9]</sup>提出遮挡轮廓恢复的方法进行柑橘识别, 利用 LBP 特征的识别准确率可达 82.3%, 之后利用水果表面光照分布的分层轮廓特征进行树上绿色柑橘的检测, 但在图像采集过程中需要人工采集光源且场景适应性不高。Zhao 等<sup>[10]</sup>提出一种 SATD 匹配方法对潜在的水果像素点进行检测, 使用支持向量机 (support vector machine, SVM) 进行分类, 识别准确率达到 83% 以上, 但该方法随着目标数的增加, 检测的复杂度和难度随之增加, 从而影响检测精度。Dorj 等<sup>[11]</sup>提出利用颜色特征来预测柑橘产量, 首先将 RGB 图像转换为 HSV 图像, 再对图像进行橙色检测, 利用分水岭分割法对柑橘进行计数, 但该方法的检测准确率受距离和背景变化的影响较大。Liu 等<sup>[12]</sup>提出一种基于成熟柑橘区域特征的识别方法, 利用特征映射表来降低特征向量的维数, 但该方法在阴影情况下的识别效果较差。

另外, 在传统机器视觉领域, 针对苹果、桃子、番茄等常见水果的检测方法也相继被提出。马翠花等<sup>[13]</sup>提出基于密集和稀疏重构<sup>[14]</sup> (dense and sparse reconstruction, DSR) 的未成熟番茄识别方法, 识别准确率为 77.6%。张春龙等<sup>[15]</sup>使用基于 SVM 分类器和阈值分

收稿日期: 2019-06-24 修订日期: 2019-08-27

基金项目: 国家自然科学基金项目(61601189, 31971797); 现代农业产业技术体系建设专项资金(CARS-26); 广东省科技计划项目(2016A020210088); 广州市科技计划项目(201803020037)

作者简介: 吕石磊, 博士, 副教授, 主要从事农业信息化研究。

Email: lvshilei@scau.edu.cn

\*通信作者: 李震, 博士, 教授, 博士生导师, 主要从事机电一体化技术应用研究。Email: lizhen@scau.edu.cn

类器的混合分类器来实现近色背景中绿色苹果的识别与计数,平均识别准确率为 89.3%,但该方法对单张图片检测时间过长,对重叠果实识别精度较低。Peng 等<sup>[16]</sup>提出利用形状不变矩等方法综合果实的颜色和形状特征,利用 SVM 分类器对提取的特征向量进行分类识别,但该方法在复杂环境下适用性较低。Wajid 等<sup>[17]</sup>提出基于决策树分类的成熟柑橘、未成熟柑橘识别方法,但受决策树特性影响,该方法的识别准确率稳定性有待进一步提高。以上对不同水果的识别方法主要是基于颜色特征和纹理特征,都受限于自身算法的局限性,无法找到通用的特征提取模型,无法同时识别所有类型的水果,而且受光照以及环境的影响较大,泛化性较差,很难满足实时检测的要求,算法没有得到很好的推广。

由于传统机器视觉方法在复杂场景下鲁棒性差,很难满足采摘机器人在复杂场景下的工作需求。近年来,卷积神经网络<sup>[18]</sup>在目标检测领域不断完善,体现出巨大的优越性,其主要分为两类,一类是先由算法生成一系列作为样本的候选框,再通过卷积神经网络进行样本分类,代表有 RCNN<sup>[19]</sup>、Fast RCNN<sup>[20]</sup>和 Faster RCNN<sup>[21]</sup>;一类直接将目标边框定位的问题转化为回归问题处理,不需要产生候选框,标志性算法包括 SSD<sup>[22]</sup>、YOLO<sup>[23]</sup>等。傅隆生等<sup>[24]</sup>提出基于卷积神经网络的田间多簇猕猴桃图像识别方法,该方法对相邻果实和独立果实的识别率分别为 91.01%和 94.78%,但对遮挡、重叠果实识别准确率较低。Sa 等<sup>[25]</sup>和熊俊涛等<sup>[26]</sup>分别采用 Faster R-CNN 深度卷积神经网络来识别甜椒和柑橘,但模型训练速度较慢,检测时间较长。彭红星等<sup>[27]</sup>提出基于 SSD 深度学习模型对 4 种不同水果进行识别,平均检测精度可达 89.53%,有较好的泛化性和鲁棒性。薛月菊等<sup>[28]</sup>提出基于 YOLOv2<sup>[29]</sup>的未成熟芒果识别方法,识别速度和识别精度可显著提升。以上研究的开展为卷积神经网络应用于果实识别提供了参考和可行性依据,且能避免传统机器视觉方法中特征提取过程的不足。

该文结合 SSD 网络以及 YOLO 系列网络的优点,通过改进 YOLOv3 网络模型<sup>[30]</sup>,使用一种新的 IoU 边框回归损失函数,并提出一种混合训练<sup>[31]</sup>与迁移学习<sup>[32]</sup>结合的方式将模型在 COCO 数据集中学习到的知识迁移到柑橘图像识别过程中,提出一种基于 GIoU<sup>[33]</sup>的 YOLOv3-LITE 轻量级神经网络的柑橘识别方法。在不损失精度的情况下,该方法较传统 YOLOv3 网络的检测速度提升了近 4 倍,模型占用内存缩小近 8 倍。

## 1 材料与方法

### 1.1 试验数据的采集

柑橘图像的采集地为广东省梅州市柑橘果园,使用数码相机、高清手机等多种设备拍摄距离 1 m 左右的自然光照下柑橘树冠图像,拍摄角度朝东南西北 4 个方向,并采集树冠下方含阴影的果实,共采集原始图片 500 张,图像包括阴天、晴天、雨天,涵盖顺光、逆光等所有光照情况。为保证柑橘图像数据的多样性,通过网络爬虫获取 200 张柑橘照片,挑选出 120 张,所得数据集为 620 张。

### 1.2 数据增强

使用 Matlab 工具对原始数据集进行数据扩增,对原始图像进行旋转,旋转角度随机取 $-30^\circ$ 、 $-15^\circ$ 、 $15^\circ$ 、 $30^\circ$ ;对原始图像随机进行镜像翻转、水平翻转、垂直翻转;通过裁剪以及缩放等方式进行扩展数据集;通过调整饱和度和色调、直方图均衡化、中值滤波等图像处理技术对数据进行增强。考虑到数据增强会导致图片中图像形状变化以及质量变化较为严重,对每张图片随机采用以上一种方式进行扩增,得到 1 240 张,筛选出符合试验的数据作为最终的数据集,最终数据集为 1 130 张。

### 1.3 数据集准备

将上述数据集使用 labelling 工具对检测目标进行标记。考虑标签和数据的对应关系以及确保数据集分布统一,使用 Matlab 工具将数据集按照 70%、10%、20%的比例随机拆分为训练集、验证集、测试集,其中训练集含边框标注样本为 7 148 个,验证集含边框标注样本为 1 006 个,测试集含边框标注样本为 2 243 个。将最终数据集按照 PASCAL VOC 数据集的格式存储,再将测试集分为 2 部分:目标平均遮挡程度小于 30%的数据集(轻度遮挡,用 A 表示)、目标平均遮挡程度大于 30%且较密集的数据集(重度遮挡,用 B 表示)。其中,测试集 A 包含标注样本 682 个,测试集 B 包含标注样本 1 561 个,最终数据集如表 1 所示。

表 1 数据集及其数量

Table 1 Data sets and size

数据集 Dataset	训练集 Training set	验证集 Validation set	测试集 Test set		总数量 Total number
			A	B	
图片个数 Number of images	791	113	131	95	1 130
轻度遮挡样本个数 Number of lightly obscured samples	2 253	317	682	0	3 252
重度遮挡样本个数 Number of severely obscured samples	4 895	689	0	1 561	7 145

## 2 改进的 YOLOv3-LITE 柑橘识别网络

### 2.1 YOLOv3 网络模型

YOLO 网络模型是一种 one-stage 的方法,YOLO 将输入图像分成  $S \times S$  个格子,如果一个物体的中心落在某个网格内,则对应的网格负责检测该物体。其中,YOLOv1 将输入图像的尺寸改为  $448 \times 448$ ,然后送入 CNN 网络训练,采用非极大抑制算法进行预测,YOLOv1 对小目标检测效果不好,定位也不够精确。Redmon 等<sup>[29]</sup>提出的 YOLOv2 吸收了 Faster R-CNN 中 RPN 的思想,去掉 YOLOv1 中的全连接层,加入 Anchor Boxes,提高了召回率,且提出 k-means 聚类计算 Anchor Boxes, Average IoU 也提升了,但对小目标检测效果一般。YOLOv3 在 YOLOv2 的基础上进行了一些改进,加深了网络,提出了多标签分类预测,使用逻辑回归(logistic regression)对方框置信度进行回归,同时提出了跨尺度预测,使用类似 FPN(feature pyramid networks)的融合做法,在多

个尺度的特征图上进行位置和类别预测，对小目标的检测效果提升明显。在类别预测中，YOLOv3 不使用 Softmax 函数对每个框分类，而是对每个类别独立地使用逻辑回归，在训练过程中，使用二元交叉熵损失（binary cross-entropy loss）来进行类别预测，可以更好地处理多标签任务。YOLOv3 目标检测网络损失函数如式（1）所示。

$$\begin{aligned} \text{Loss Function} = & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{obj}} \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ & + \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{obj}} \left[ (w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] \\ & + \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{obj}} \left( C_i \log \hat{C}_i + (1 - C_i) \log (1 - \hat{C}_i) \right) \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{s^2} \ell_i^{\text{obj}} \sum_{C \in \text{Class}} (p_i(C) - \hat{p}_i(C))^2 \quad (1) \end{aligned}$$

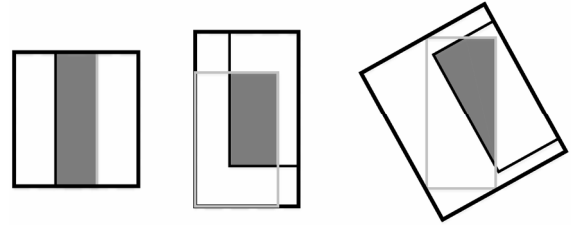
式中  $i$  表示第  $i$  个方格， $j$  表示该方格预测出的第  $j$  个方框； $\text{obj}$  表示含有目标， $\text{noobj}$  即没有目标， $C_i$  为预测目标的类别， $\hat{C}_i$  为真实目标的类别， $\lambda_{\text{coord}}$ 、 $\lambda_{\text{noobj}}$  是惩罚系数。损失函数包括 5 个部分，第 1 项包含预测和实际的  $x$ 、 $y$  坐标，该部分是计算预测的中心坐标的损失；第 2 项包含预测和实际的宽和高，即  $w$ 、 $h$ ，该部分是计算预测边界框的宽高的损失；第 3 项是方框置信度损失，回归目标是预测方框与实际方框的 IoU 值；前 3 项都只在预测方框和实际方框有对应关系时才计算损失；第 4 项是没有对应物体的方框置信度损失，3、4 项是计算预测的置信度损失，第 5 项是分类损失，当方格中有物体时，才计算分类损失。

## 2.2 改进的 YOLOv3-LITE 轻量级神经网络模型设计

### 2.2.1 基于 GIoU 的边框回归损失函数

IoU 为预测框与原来图片中标记框的重合程度，目标检测领域常使用边框回归 IoU 值作为评价指标。但是，大部分检测框架没有结合该值优化损失函数，IoU 可以被反向传播，它可以直接作为目标函数去优化。考虑到优化度量本身与使用替代的损失函数之间的选择，最佳选择是优化度量本身。作为损失函数，传统 IoU 有两个缺点：如果 2 个对象不重叠，IoU 值将为零，则其梯度将为零，无法优化；2 个物体在多个不同方向上重叠，且交叉点水平相同，其 IoU 将完全相等，IoU 无法精确的反映两者的重合度大小，如图 1 所示，3 个不同的方法重叠两个矩形具有完全相同的 IoU 值，但他们的重合度是不一样的，最左边的图回归的效果最好，最右边的回归效果最差，其中最右边的图预测边框为旋转候选边框<sup>[34]</sup>。因此，IoU 函数的值并不能反映两个对象之间如何发生重叠。在柑橘采摘机器人的果实识别中，回归框位置的精确度直接决定了机器手采摘的成功率。因此，该文提出通过引入 GIoU 来解决 IoU 的缺点。IoU 取值 $[0,1]$ ，而 GIoU 有对称区间，取值范围 $[-1,1]$ ，在两者重合的时候取最大值

1，在两者无交集且无限远的时候取最小值-1。因此，GIoU 是一个非常好的距离度量指标，与 IoU 只关注重叠区域不同，GIoU 不仅关注重叠区域，还关注其他的非重合区域，能更好的反映两者的重合度，GIoU Loss 可以替换掉大多数目标检测算法中边框回归的损失函数，如式（2）-（5）所示。



注：黑色矩形代表预测的边界框，灰色矩形代表原始标记的边界框。  
Note: The black rectangle represents the predicted bounding box; the gray rectangle represents the original marker's bounding box.

图 1 两个矩形重叠示意

Fig.1 Diagram of two overlapping rectangles

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

$$\text{GIoU} = \text{IoU} - \frac{|C(A \cap B)|}{|C|} \quad (3)$$

$$L_{\text{GIoU}} = 1 - \text{GIoU} \quad (4)$$

$$\begin{aligned} \text{GIoULoss} = & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{obj}} (1 - \text{GIoU}) \\ & + \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{obj}} \left( C_i \log \hat{C}_i + (1 - C_i) \log (1 - \hat{C}_i) \right) \\ & + \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^B \ell_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\ & + \sum_{i=0}^{s^2} \ell_i^{\text{obj}} \sum_{C \in \text{Class}} (p_i(C) - \hat{p}_i(C))^2 \quad (5) \end{aligned}$$

式中  $A$  和  $B$  为任意两个矩形框， $C$  为包围  $A$ 、 $B$  矩形框的最小外接矩形， $S$  为  $A$ 、 $B$  所在空间， $A, B \subseteq S \in \mathbb{R}^n$ 。

### 2.2.2 YOLOv3-LITE 网络模型设计

传统 YOLOv3 采用自定义骨干网络 Darknet-53，模型计算复杂，对存储空间要求较高，在 GPU 上一张 416×416 的图片推断速度为 30 ms，在 CPU 上推断速度为 255.8 ms。该文提出一种实时目标检测的轻量级神经网络模型，在传统 YOLOv3 网络的基础上，设计了 YOLOv3-LITE 网络，在 GPU 上的推断速度可达 16.9 ms，在 CPU 上推断速度可达 80.9 ms。MobileNet<sup>[35]</sup>是一种基于移动端的轻量级神经网络，该文使用 MobileNet-v2<sup>[36]</sup>轻量级神经网络作为 YOLOv3-LITE 的骨干网络，MobileNet-v2 网络模型采用反残差模块与深度可分离卷积结合，首先通过 1×1 卷积提升通道数，后通过 3×3 卷积进行 Depthwise convolution（深度卷积），再用 1×1 卷积降低维度。MobileNet-v1 使用 ReLU6 替换 ReLU 激活函数，控制线性增长部分。MobileNet-v2 将非线性激活函数 ReLU6 去掉，即不使用激活函数，直接线性输出，减

少了信息丢失。其深度可分离卷积层具体结构如图 2 所示, 图中采用输入为 RGB 三通道图片, MobileNet-v2 在 Depthwise convolution 之前添加一层升通道卷积层(逐点卷积), 添加了这一层升通道卷积之后, Depthwise convolution 的 Filter 数量取决于升通道卷积之后的通道数, 而这个通道数是可以任意指定的, 因此解除了  $3 \times 3$  卷积核个数的限制。将普通卷积用深度可分离卷积代替, 使得计算量大大降低, 同时可以通过增加通道数来提升模型的精度, 对速度和精度有较好的提升, 且便于迁移到嵌入式以及移动设备等较小系统上。YOLOv3 最突出的特点是它可以在 3 种不同的尺度上进行检测, 从 75 到 105 层为 YOLO 网络的特征交互层, 分为 3 个尺度, 每个尺度内, 通过卷积核的方式实现局部的特征交互, 作用类似于全连接层。为避免采用 MobileNet-v2 网络会对小目标检测精度降低, 该文将特征图融合改为在 19、34 层做深度连接, 对于输入为  $416 \times 416$  的图像, 卷积网络在 53 层后, 经过卷积得到  $13 \times 13$  的特征图, 这里的特征图感受野较大, 适合检测尺寸比较大的对象, 即第 1 次预测输出; 为实现细粒度的检测, 卷积层 53 层的特征图

往右开始上采样, 得到与 34 层相同分辨率的特征图, 经过残差模块然后与 34 层特征图融合, 故 65 层经卷积后得到  $26 \times 26$  的特征图, 具有中等尺寸的感受野, 适合检测中等尺寸的对象; 最后 65 层特征图再次上采样, 得到与 19 层相同分辨率的特征图, 经过残差模块然后与 19 层特征图融合, 最后得到相对输入图像 8 倍下采样得到  $52 \times 52$  的特征图, 此时感受野较小, 适合检测小尺寸的对象, 具体结构如图 3, 网络参数如表 2 所示。

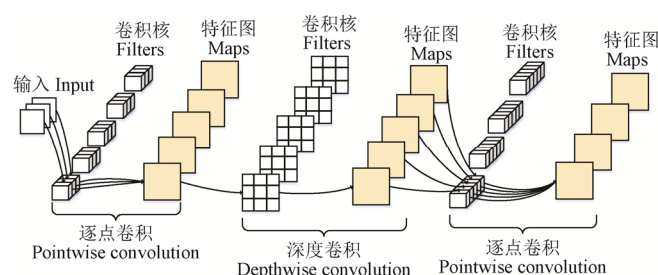
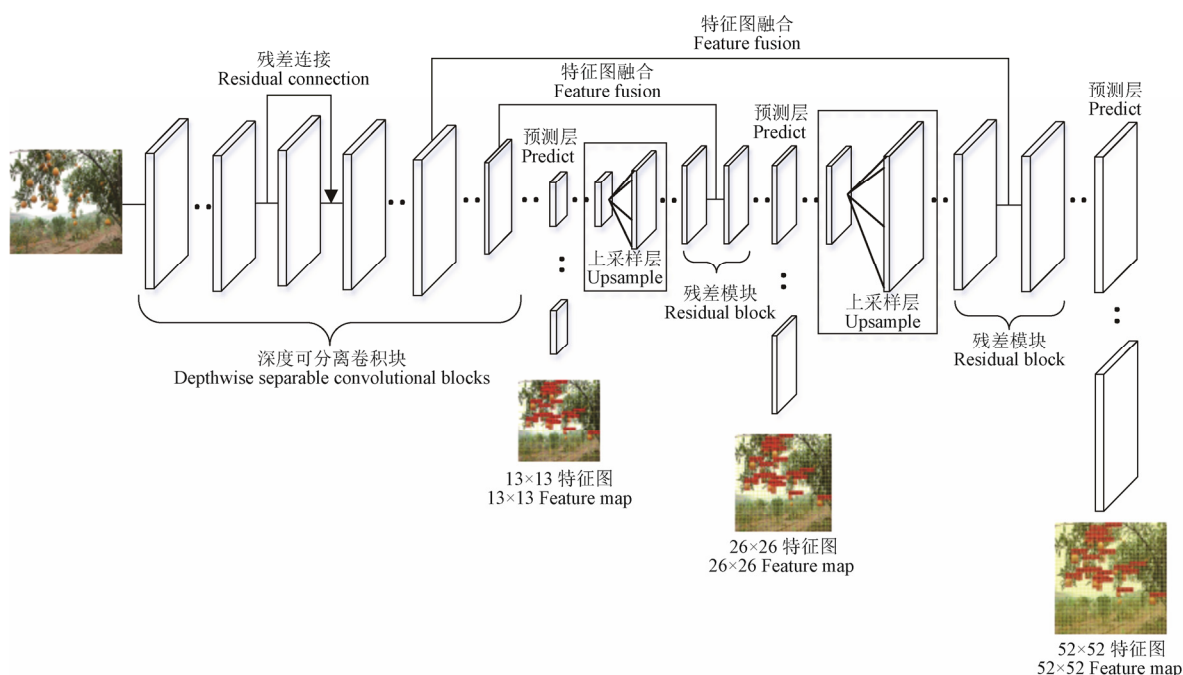


图 2 深度可分离卷积层结构示意图

Fig.2 Structure diagram of depth wise separable convolutional layer



注: 图 3 中深度可分离卷积块的单层结构如图 2 所示。

Note: The single-layer structure of depthwise separable convolutional blocks used in this figure is shown in Fig.2.

图 3 YOLOv3-LITE 结构

Fig.3 Structure diagram of YOLOv3-LITE

### 2.2.3 一种混合训练与迁移学习结合的预训练方式

在自然环境下, 柑橘目标遮挡情况严重, 且由于样本集数据涵盖的场景有限, 只识别单一类别的柑橘将使得模型泛化能力受限。该文使用一种为训练目标检测网络而设计的视觉相干图像混合方法 (visually coherent image mixup), 该方法可以有效提升模型的泛化能力, 减少过拟合问题。Mixup 是指将 2 张输入图像按照一定权重合并成一张图像, 基于这种合成图像进行训练的模型更加鲁棒, 可以达到目标遮挡的效果, 能够有效降低图像之间差异性的影响, 如图 4 所示。迁移学习 (transfer

learning) 是把已训练好的模型学习到的知识迁移到新的模型来帮助新模型训练。Yosinski 等<sup>[32]</sup>通过迁移学习试验并可视化证明底层的卷积神经网络能够学习到物体的通用特征, 例如几何变化、边缘、色彩变化等, 而高层网络则负责提取特定的特征细节。小数据集通过迁移学习也能够达到较好的训练效果。该文采用混合训练的方式对 COCO 数据集进行预训练, 通过迁移学习将模型从 COCO 数据集学习到的知识迁移到柑橘图像识别中, 通过冻结部分卷积层, 使得在反向传播修正模型参数时只对部分卷积层进行修正模型参数, 使用迁移学习与混合



训练结合的方式，降低了模型训练的时间和节省了内存消耗，柑橘目标识别效果提升明显。

表 2 YOLOv3-LITE 网络参数

Table 2 Parameters of YOLOv3-LITE networks

网络层 Layers	输入尺寸 Input size	步长 Step	重复次数 Numbers	通道数 Filters
Conv 1×1	416×416×3	2	1	32
Bottleneck/1	208×208×32	1	1	16
Bottleneck/6	208×208×16	2	2	24
Bottleneck/6	104×104×24	2	3	32
Bottleneck/6	52×52×32	1	4	64
Bottleneck/6	52×52×64	2	3	96
Bottleneck/6	26×26×96	2	3	160
Bottleneck/6	13×13×160	1	1	320
Conv 1×1	13×13×320	1	1	1 280
Conv 3×3	13×13×1 280	1	2	1 280
Conv Set	13×13×1 280	—	1	—
Conv 1×1	13×13×1 024	1	1	255
Detection	13×13×255	—	1	—
Conv 1×1	13×13×512	1	1	256
Upsample 2x	13×13×256	—	1	—
Route	—	—	1	—
Conv Set	26×26×768	—	1	—
Conv 1×1	26×26×512	1	1	255
Detection	26×26×255	—	1	—
Conv 1×1	26×26×256	1	1	128
Upsample 2x	26×26×128	—	1	—
Route	—	—	1	—
Conv Set	52×52×384	—	1	—
Conv 1×1	52×52×256	1	1	255
Detection	—	—	1	—

注：Bottleneck 是指深度可分离卷积块与残差网络的集合，Bottleneck/1，Bottleneck/6 代表中间通道数相对于输入通道的倍增系数为 1 和 6；Conv Set 代表 YOLOv3 的基本模块，包含 1×1 卷积，3×3 卷积以及 ReLu 层、BN 层；Up sample 2x 代表 2 倍上采样；Route 层为特征图融合层。

Note: Bottleneck refers to the set of deep separable convolution blocks and residual networks, Bottleneck/1 and Bottleneck/6 represent the expansion factor of the number of intermediate channels relative to the input channels of 1 and 6; Conv Set represents the basic module of YOLOv3, including 1×1 convolution, 3×3 convolution, ReLu layer and BN layer; Upsample 2x means double Upsample, Route is the feature map fusion layer.

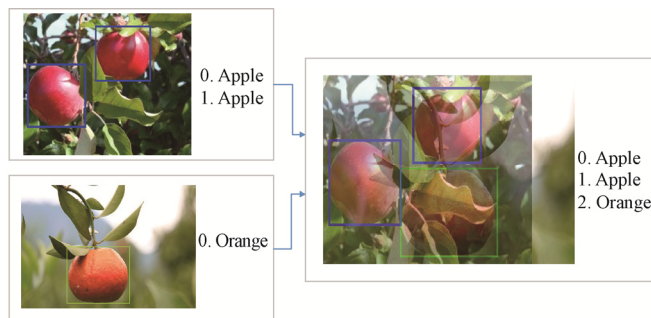


图 4 视觉相干图像混合示意

Fig.4 Diagram of visually coherent image mixup

## 2.3 模型的训练与测试

### 2.3.1 试验平台

该文训练采用的操作系统为 Ubuntu 18.04，测试的框架为 Tensorflow、Darknet，处理器为 E5-2620 V4@2.10GHz，八核，16GB RAM，显卡为 Nvidia GeForce RTX 2080TI，使用 CUDA 10.0 版本并行计算框架配合 CUDNN 7.3 版本的深度神经网络加速库。

### 2.3.2 柑橘识别网络训练

柑橘目标检测网络训练的流程如图 5 所示。该文采

用对比实验的方式，使用网络模型 Faster-RCNN、SSD、YOLOv3、以及 GIoU+YOLOv3-LITE（即改进的 YOLOv3-LITE）进行对比试验，并在不同的数据集上验证模型效果。首先，将采集到的数据分为重度遮挡（目标平均遮挡程度大于 30%）与轻度遮挡（目标平均遮挡程度小于 30%）的 2 部分数据集，通过数据增强对图像进行扩充，对样本进行标注，以 PASCAL VOC 的格式进行存储。分别采用混合训练与迁移学习结合的预训练方式、单独使用迁移学习的预训练方式、不采用预训练方式等 3 种方法来训练网络模型，并分别结合反向传播算法修正模型参数，使损失函数不断减小，当平均损失小于 0.01，且多次迭代损失函数不再减少时，停止训练。

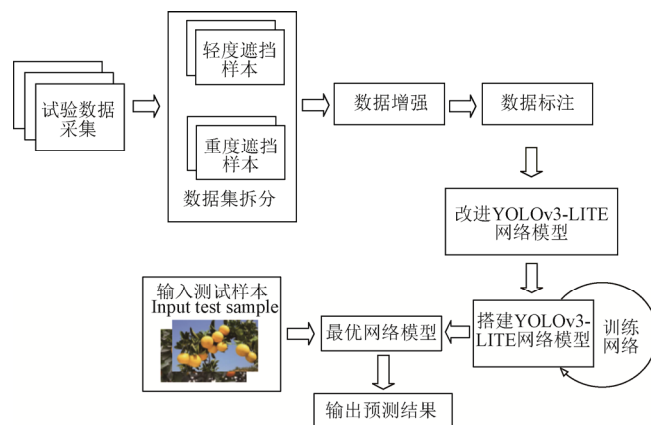


图 5 柑橘目标检测网络的流程

Fig.5 Flowchart of orange detection network

模型超参数设置为每批量样本数为 32，动量因子为 0.9，初始学习率为 0.001，每过 5 000 次迭代训练，将学习率降低 10 倍，模型每训练 100 次保存一次权重。

### 2.3.3 模型测试

该文使用  $F1$  值 ( $F1$  score)、 $AP$  值 (average precision) 来评价损失函数训练出来的模型，其中  $F1$  值计算公式如式 (8) 所示， $AP$  值计算公式如式 (9) 所示。表示如下

$$P = \frac{TP}{TP+FP} \quad (6)$$

$$R = \frac{TP}{TP+FN} \quad (7)$$

$$F1 = \frac{2PR}{P+R} \quad (8)$$

$$AP = \int_0^1 P(R) dR \quad (9)$$

式中  $P$  为准确率， $R$  为召回率， $TP$  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。

分别使用如式 (5) 所示的 GIoU 损失函数与如式 (1) 所示的原始 YOLO 损失函数训练 YOLOv3-LITE 网络模型，训练时间分别为 10.6 与 12.8 h，采用训练的 Loss 曲线图如图 6a 所示，图中损失值即损失函数的值，损失函数包含分类损失、置信度损失、边框回归损失 3 部分。采用 YOLO Loss 的模型在训练集上迭代的 loss 曲线如

图 6a 中曲线 YOLO loss-train 所示, 在验证集上面迭代的 loss 曲线如图 6a 中曲线 YOLO loss-val 所示。采用 GIoU 损失函数的模型在训练集上迭代的 loss 曲线如图 6a 中曲线 GIoU loss-train 所示, 在验证集上面迭代的 loss 曲线如图 6a 中曲线 GIoU loss-val 所示。训练的 Average IoU 曲线图如 6b 所示。

根据图 6 所示, 采用 GIoU Loss 训练的模型在验证集上面的拟合程度要优于传统 YOLO Loss, 并且 Average IoU 值要明显高于 YOLO Loss, 采用 YOLO Loss 的模型在 3 000 次之后渐渐稳定, 而采用 GIoU Loss 的模型在 9 000 次之后逐渐稳定。该文采用验证集去验证模型的优劣, 并且通过对比验证集与训练集的 loss 曲线去调节模型的超参数, 图 6a 为不断调参过程中选择出来最佳的超参数训练所迭代的 loss 曲线图, 其中最优超参数同 2.3.2

节模型所设置的超参数。每迭代 100 次保存一次权重, 对训练出来的模型进行测试和评估。该文通过每 100 次训练保存的权重, 使用客观的评价标准 (F1 值、AP 值以及 Average IoU 值) 进行评估模型的优劣。

3 结果与分析

为了验证该文的柑橘目标检测方法的性能, 对提出的策略进行单个验证, 对比使用混合训练与迁移学习以及不同骨干网络在柑橘目标检测中的精度与速度, 之后对比采用新型损失函数对模型检测精度的影响。采用网络模型 SSD、YOLOv2、YOLOv3、Faster-RCNN 以及该文采用的改进 YOLOv3-LITE 进行对比实验, 数据集如表 1 所示, 比较模型在不同的数据集的优劣程度, 重点分析模型对遮挡果实识别的准确率。

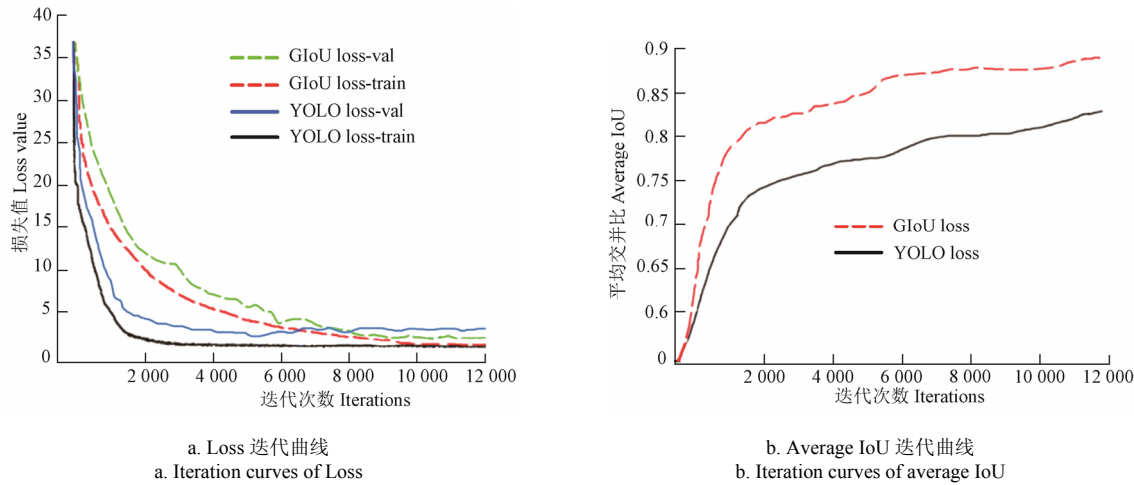


图 6 迭代曲线图  
Fig.6 Iteration curves

3.1 使用混合训练与迁移学习的检测结果

以 YOLOv3-LITE 网络为基础网络, 传统方法为不使用预训练模型, 用训练集从头开始对所有参数进行训练, 迁移学习的方法是使用 COCO 数据集的预训练模型对模型进行部分层参数的训练, 该文提出使用混合训练与迁移学习结合的预训练模型进行模型的微调, 3 种方法检测的结果在测试集 A+B (如表 1 所示) 的对比结果如表 3 所示。相对于传统方法, 使用迁移学习对模型提升效果明显, 使用混合训练与迁移学习结合, 模型的 F1 值上升 3.79%, AP 值上升 2.75%, Average IoU 值上升 2.74%。

表 3 不同训练方式检测结果的比较

Table 3 Comparison of detection results using different training methods

训练方式 Training methods	F1 score/%	平均精度 Average precision/%	IoU Average IoU/%
传统方法 Original method	88.97	87.63	80.47
迁移学习方法 Transfer learning method	91.61	89.49	82.78
混合训练+迁移学习方法 Mixup+Transfer learning method	92.76	90.38	83.21

3.2 不同骨干网络检测结果对比

为了证明模型在 YOLO 网络框架下改进的优越性, 该文通过对比不同的网络框架以及不同骨干网络进行对比, 在测试集 A+B (如表 1 所示) 的检测结果如表 4 所示。其中, 采用 DarkNet-53 为骨干网络的传统 YOLOv3 模型相对于 YOLOv2 模型由于网络模型更复杂, 使得检测速度略低于 YOLOv2, 但模型 F1 值提升了 5.45%, AP 值提升了 6.36%, 识别准确率提升明显。从表 4 中可以看到 YOLOv3-Thiny 减少了模型的层数, 检测速度与模型大小都取得了很好的提升, 但检测精度下降明显。采用 MobileNet-v1 作为 YOLOv3 的骨干网络, 其速度达到 270 帧/s, 模型大小为 23 MB, 速度和模型占用内存取得了最优的结果, 但检测精度较传统 YOLOv3 模型下降明显。而骨干网络采用 MobileNet-v2 的 YOLOv3 (即 YOLOv3-LITE), 其 F1 值与 AP 值在所有模型中取得了最优的效果, 且相对于传统 YOLOv3 方法, 模型 AP 值提升 1.68%, F1 值提升了 0.84%, 且模型权重所占内存减小 208 MB, 在 GPU 上的检测速度对视频检测的帧率高达 246 帧/S, 在 CPU 上的检测速度可以达到 22 帧/s。由于在果园采摘机器人中绝大部分采用的是嵌入式终端或者移动设备, YOLOv3-LITE 网络具有明显的优势。



表 4 不同骨干网络检测结果对比

Table 4 Comparison of detection results using different backbone networks

网络模型 Network models	骨干网络 Backbone networks	F1 score/%	平均精度 Average precision /%	权重大小 Weight size	检测速度 Detection speed
YOLOv2	DarkNet-19	85.75	82.34	195 MB	70
YOLOv3	DarkNet-53	91.92	88.70	236 MB	62
YOLOv3-Thiny	Thiny	78.67	77.21	34 MB	220
YOLOv3	MobileNet-v1	88.37	86.49	23 MB	270
YOLOv3	MobileNet-v2	92.76	90.38	28 MB	246

3.3 使用 GIoU 损失函数的检测结果

使用 YOLOv3-LITE 为基础网络，将 GIoU 回归损失函数替代传统 YOLOv3 的回归框损失函数，对模型进行训练，在测试集 A+B（如表 1 所示）的检测结果对比如表 5 所示。采用 GIoU 替代模型的边框回归损失函数对模型的  $F1$  值提升了 0.93%，AP 值提升 0.75%，而 Average IoU 值提升了 4.11%，且相对于传统 YOLOv3 提升显著，可以看出将 GIoU 作为损失函数，对边框回归的准确率影响较大，这使得柑橘果实的定位更加精准，能为柑橘采摘机器人提供精度较高的定位信息。

表 5 使用 GIoU 损失函数的检测结果

Table 5 Detection results using GIoU loss function

网络模型 Network models	F1 score/%	平均精度 Average precision/%	IoU Average IoU/%
YOLOv3	91.92	88.70	82.60
YOLOv3-LITE	92.76	90.38	83.21
GIoU+YOLOv3-LITE	93.69	91.13	87.32

3.4 不同遮挡程度对比试验

由于叶片对样本的遮挡以及样本之间互相重叠，都会对模型的检测精度带来较大的影响。该文将遮挡程度作为控制变量，使用网络模型为改进的 YOLOv3-LITE 模型，分别取测试集 A 和 B 以及 A+B（见表 1）验证检测结果，见表 6 和图 7。对轻度遮挡果实的识别，模型的  $F1$  值能达到 95.27%，AP 能达到 92.75%，Average IoU 高达 88.65%。且在全部测试集中， $F1$  值达到 93.69%，AP 值为 91.13%，Average IoU 为 87.32%。由表 6 可知，在果实遮挡以及重叠的情况下，模型的识别精度会有所降低，一方面是因为卷积神经网络在对小目标以及目标密集的图像进行卷积计算时，在深层网络会丢失若干信息；另一个原因是考虑到实际应用场景，对单个柑橘遮挡面积超过 70%，和距离较远、目标极小的柑橘未做数据标记，这对识别效果也会有一定的影响。但从图表数据可以看出，在严重遮挡、且目标密集的环境下，模型也能达到 91.43% 的  $F1$  值和 89.10% 的 AP 值。模型对测试集 A+B 检测的 P-R 曲线如图 8 所示。

表 6 不同遮挡程度的检测结果对比

Table6 Comparison of detection results in different obscured degrees

测试集 Test set	F1 score/%	平均精度 Average precision/%	IoU Average IoU/%
A	95.27	92.75	90.65
B	91.43	89.10	83.73
A+B	93.69	91.13	87.32



图 7 不同遮挡程度的结果对比

Fig.7 Comparison of results in different obscured degrees

3.5 不同检测方法对比试验

分别对 Faster-RCNN、SSD 和改进的 YOLOv3-LITE 网络进行训练，在不同测试集下对比试验。如表 7 和图 9

所示，在果实轻度遮挡的情况下，3 个模型的  $F1$  值都达到 91% 以上，AP 值在 89% 以上。改进的 YOLOv3-LITE 比 SSD、Faster-RCNN 在速度精度方面提升明显，且在重

度遮挡的情况下,  $F1$  值也能达到 90% 以上, 相对 SSD 与 Faster-RCNN 提升近 3 个百分点。在自然环境下, 其  $F1$  值和 AP 值比 SSD 分别高出 3.94% 和 3.49%, 而比 Faster-RCNN 高出 3.8% 和 3.26%。由图 9 可知, 改进的 YOLOv3-LITE 网络会漏检遮挡大于 70% 的柑橘, 这是由于遮挡大于 70% 的柑橘默认不做标记。其次, 在较暗的背景下也会影响柑橘目标的检测精度。另外, 从表中可以得出, SSD 与 Faster-RCNN 在检测精度方面比较接近。

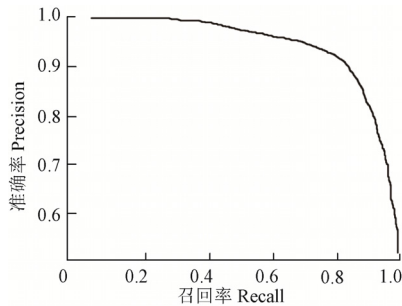


图 8 P-R 曲线  
Fig.8 P-R curve

在训练时间方面, 统一设置模型每次迭代训练的批次大小为 32, 其中 Faster-RCNN 模型训练时间为 16.7 h, SSD 模型训练时间为 12.2 h, 该文提出的 YOLOv3-LITE

模型训练的时间为 10.6 h, 由于 YOLOv3 与 SSD 去掉了全连接层, 所有在训练时间上相对 Faster-RCNN 会有显著提升, 并且该文采用混合训练与迁移学习的方式, 有效降低了训练时间。在检测速度方面, 在 GPU 上, YOLOv3-LITE 可以达到 246 帧/s 的检测速度, 比 SSD 检测速度提升近 4 倍, 比 Faster-RCNN 提升了近 20 倍。综合上述 5 组对比试验的结果分析表明, 该文提出的改进 YOLOv3-LITE 轻量级神经网络能够有效识别自然环境下的柑橘果实, 识别准确率和识别速度具有较显著的优势。

表 7 不同网络模型的检测结果对比  
Table 7 Comparison of detection results using different network model

测试集 Test set	网络模型 Network models	$F1$ score/%	平均精度 Average precision/%
A	GIoU+YOLOv3-LITE	95.27	92.75
	SSD-300	92.36	89.26
	Faster-RCNN	91.78	89.53
B	GIoU+YOLOv3-LITE	91.43	89.10
	SSD-300	88.72	86.89
	Faster-RCNN	89.13	87.08
A+B	GIoU+YOLOv3-LITE	93.69	91.13
	SSD-300	89.75	87.64
	Faster-RCNN	89.89	87.87



图 9 不同检测方法的结果对比  
Fig.9 Comparison of results using different detection methods

#### 4 结 论

1) 该文提出了一种改进的 YOLOv3-LITE 轻量级神经网络的柑橘识别方法。试验结果表明, 在自然环境下, 模型检测的  $F1$  和 AP 值分别达到 93.69% 和 91.13%。使用 GIoU 回归框损失函数替代传统损失函数边框回归的 MSE 均方误差部分, Average IoU 高达 87.32%, 为柑橘采摘机器人定位提供良好的技术支持。

2) 该文通过改进模型, 提出一种轻量级神经网络模

型, 模型占用内存为 28 MB, 且对单张  $416 \times 416$  像素的图片的推断速度, 在 GPU 上 YOLOv3-LITE 可以达到 16.9 ms, 在 CPU 上可以达到 80.9 ms 的检测速度, 可用于移植到嵌入式以及手机终端; 提出一种混合训练与迁移学习结合的方法, 将模型在 COCO 数据集学习到的知识迁移到柑橘识别模型中, 提高了模型的泛化能力, 大大降低了训练的时间和资源。

3) 为了验证提出方法的可行性和优越性, 该文以检测精度以及计算速度等为判别依据, 与不同模型进行对比



试验,对比了 Faster-RCNN 和 SSD 在不同场景下的识别准确率,在自然环境下,其  $F1$  值和 AP 值比 SSD 分别高出 3.94% 和 3.49%, 而比 Faster-RCNN 高出 3.8% 和 3.26%。在 GPU 上, YOLOv3-LITE 可以达到 246 帧/s 的检测速度,比 SSD 检测速度提升近 4 倍,而比 Faster-RCNN 提升了近 20 倍。通过对比,提出的方法具有显著优势。

#### [参 考 文 献]

- [1] Ren D, Yang S X. Intelligent automation with applications to agriculture[J]. *Intelligent Automation & Soft Computing*, 2016, 22(2): 227—228.
- [2] 宋展, 胡宝贵, 任高艺, 等. 智慧农业研究与实践进展[J]. *农学学报*, 2018, 8(12): 95—100.  
Song Zhan, Hu Baogui, Ren Gaoyi, et al. Research progress of the smart agriculture extension system[J]. *Journal of Agriculture*, 2018, 8(12): 95—100. (in Chinese with English abstract)
- [3] 马驰, 李光林, 李晓东, 等. 丘陵山地柑橘果园多方位自动喷药装置研制[J]. *农业工程学报*, 2019, 35(3): 31—41.  
Ma Chi, Li Guanglin, Li Xiaodong, et al. Development of multi-orientation automatic spraying device for citrus orchards in hilly and mountainous areas[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2019, 35(3): 31—41. (in Chinese with English abstract)
- [4] Hussin R, Juhari M R, Kang N W, et al. Digital Image processing techniques for object detection from complex background image[J]. *Procedia Engineering*, 2012, 41: 340—344.
- [5] 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. An object detection method for quasi-circular fruits 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)
- [6] 熊俊涛, 邹湘军, 彭红星, 等. 扰动柑橘采摘的实时识别与采摘点确定技术[J]. *农业机械学报*, 2014, 45(8): 38—43.  
Xiong Juntao, Zou Xiangjun, Peng Hongxing, et al. Real-time identification and picking point localization of disturbance citrus picking[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(8): 38—43. (in Chinese with English abstract)
- [7] 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, 60.  
Lu Jun, Sang Nong. Detection of citrus fruits within tree canopy and recovery of occlusion contour in variable illumination[J]. *Transactions of the Chinese Society of Agricultural Machinery*, 2014, 45(4): 76—81, 60. (in Chinese with English abstract)
- [8] 卢军, 胡秀文. 弱光复杂背景下基 MSER 和 HCA 的树上绿色柑橘检测[J]. *农业工程学报*, 2017, 33(19): 196—201.  
Lu Jun, Hu Xiuwen. Detecting green citrus fruit on trees in low light and complex background based on MSER and HCA[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2017, 33(19): 196—201. (in Chinese with English abstract)
- [9] Lu J, Lee W S, Gan H, et al. Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis[J]. *Biosystems Engineering*, 2018, 171: 78—90.
- [10] 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.
- [11] Dorj U O, Lee M, Yun S. An yield estimation in citrus orchards via fruit detection and counting using image processing[J]. *Computers and Electronics in Agriculture*, 2017, 140: 103—112.
- [12] 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.
- [13] 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 (Transactions of the CSAE)*, 2016, 32(14): 219—226. (in Chinese with English abstract)
- [14] Li X, Lu H, Zhang L, et al. Saliency detection via dense and sparse reconstruction[C]//IEEE International Conference on Computer Vision (ICCV). Sydney: IEEE, 2013: 2976—2983.
- [15] 张春龙, 张楫, 张俊雄, 等. 近色背景中树上绿色苹果识别方法[J]. *农业机械学报*, 2014, 45(10): 277—281.  
Zhang Chunlong, Zhang Ji, Zhang Junxiong, et al. Recognition of green apple in similar background[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(10): 277—281. (in Chinese with English abstract)
- [16] Peng H, Shao Y, Chen K, et al. Research on multi-class fruits recognition based on machine vision and SVM[J]. *IFAC-PapersOnLine*, 2018, 51(17): 817—821.
- [17] Wajid A, Singh N K, Junjun P, et al. Recognition of ripe, unripe and scaled condition of orange citrus based on decision tree classification[C]//International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). Sukkur: IEEE, 2018: 1—4.
- [18] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems, 2012: 1097—1105.
- [19] 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: IEEE, 2014: 580—587.
- [20] Girshick R. Fast r-cnn[C]//IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440—1448
- [21] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91—99.
- [22] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Cham: Springer, 2016: 21—37.
- [23] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//IEEE conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779—788.
- [24] 傅隆生, 冯亚利, Elkamil Tola, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. *农业工程学报*, 2018, 34(2): 205—211.  
Fu Longsheng, Feng Yali, Elkamil Tola, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2018, 34(2): 205—211. (in Chinese with English abstract)
- [25] Sa I, Ge Z, Dayoub F, et al. DeepFruits: A fruit detection system using deep neural networks[J]. *Sensors*, 2016, 16(8): 1222.
- [26] 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 with English abstract)
- [27] 彭红星, 黄博, 邵园园, 等. 自然环境下多类水果采摘目标识别的通用改进 SSD 模型[J]. 农业工程学报, 2018, 34(16): 155–162.  
Peng Hongxing, Huang Bo, Shao Yuanyuan, et al. General improved SSD model for picking object recognition of multiple fruits in natural environment[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(16): 155–162. (in Chinese with English abstract)
- [28] 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)
- [29] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263–7271.
- [30] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. 2018, arXiv: 1804. 02767v1.
- [31] Zhang Z, He T, Zhang H, et al. Bag of freebies for training object detection neural networks[J]. 2019, arXiv: 1902. 04103v3.
- [32] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks?[C]//Advances in Neural Information Processing Systems, 2014: 3320–3328.
- [33] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 658–666.
- [34] Ma J, Shao W, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111–3122.
- [35] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for nobile vision applications[J]. 2017, arXiv: 1704. 04861v1.
- [36] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510–4520.

## Orange recognition method using improved YOLOv3-LITE lightweight neural network

Lü Shilei<sup>1,2,3</sup>, Lu Sihua<sup>1</sup>, Li Zhen<sup>1,2,3\*</sup>, Hong Tiansheng<sup>1,2,4</sup>, Xue Yueju<sup>1,2,3</sup>, Wu Benlei<sup>1</sup>

(1. College of Electronic Engineering, South China Agricultural University, Guangzhou 510642, China; 2. Division of Citrus Machinery, China Agriculture Research System, Guangzhou 510642, China; 3. Guangdong Engineering Research Center for Monitoring Agricultural Information, Guangzhou 510642, China; 4. College of Engineering, South China Agricultural University, Guangzhou 510642, China)

**Abstract:** Orange recognition is one of the key technologies for automatical fruits picking, delicacy management and orchard yield forecast etc. In order to realize the rapid and accurate identification of orange fruits in natural environment, this paper presents an orange recognition method using improved YOLOv3-LITE lightweight neural network. In the field of fruit-picking robots, the accuracy of bounding box regression for fruit recognition directly determines the success rate of robotic hand-picking. In this paper, the proposed method, which introduces GIoU bounding box regression loss function (GIoU reflects the relationship between the target frame and forecast box) to replace traditional loss function of MSE (mean square error) part of the bounding box regression, effectively improves the accuracy of bounding box regression. In order to facilitate the migration of the model to mobile terminals, the YOLOv3-LITE lightweight neural network model is proposed. Mobilenet-v2, a lightweight network proposed based on the mobile terminal, can effectively reduce the complexity of the model, and is thus used in this paper to replace the backbone network Darknet-53 in the original network model. The mixup-training method combines 2 pictures into 1 with a certain weight to achieve the effect of target occlusion, reduce the impact of differences between images, improve the generalization ability of the model, and reduce over fitting. In this paper, the mixup-training is combined with transfer learning for the purpose of pre-training. The deep learning framework built in the proposed model is TensorFlow and DarkNet. In order to verify the superiority and feasibility of the proposed model, its recognition effect is tested on test samples with different obscured degrees, comparing with the Faster-RCNN, SSD models and original YOLOv3 network model. The values of  $F1$  and AP are used to assess the differences among models. Test results show that, the model proposed in this paper can get significantly good recognition results. In the data sets of lightly obscured fruit, the  $F1$  value and AP value of the orange recognition model proposed in this paper can reach 95.27% and 92.75%, and the Average IoU is as high as 88.65%. In the data sets of severely obscured fruit, the  $F1$  value and AP value of the orange recognition model proposed in this paper are 91.43% and 89.10%, respectively, and the Average IoU is 83.73%. In all the test sets, the  $F1$  and AP values are 93.69% and 91.13%, respectively, and Average IoU is 87.32%. Compared with the original YOLOv3, the values of  $F1$  and AP increase by 1.77% and 2.43%, respectively, and Average IoU increases by 4.72%. The detection speed of orange target can reach 246 frames per second; the inferred speed of a single 416×416 picture on GPU is 16.9 ms. The detection speed on the CPU can reach 22 frames per second, and the inferred speed can reach 80.9ms. The memory occupied by the proposed model is 28 MB. Moreover, the pre-training method, combined with the transfer learning and mixup training, reduces the training time and memory consumption of the proposed model. Therefore, the orange recognition method proposed in this paper has the advantages of low model memory footprint, high recognition accuracy and fast recognition speed, etc. It can provide new solutions for orange picking robot and orchard yield forecast, and new ideas for intelligent orange industry as well.

**Keywords:** neural networks; fruits; algorithms; orange; YOLOv3-LITE; mixup-training; transfer learning; GIoU bounding box regression loss function