2023年4月 第42卷第4期

DOI:10. 19652/j. cnki. femt. 2304668

# 基于 YOLOv5 的草莓轻量化网络检测模型\*

杨世忠 王瑞彬 高 升 邵明伟 (青岛理工大学信息与控制工程学院 青岛 266520)

**摘 要:**随着现代农业技术的发展,草莓生产和采摘的自动化是一个必然的趋势,而草莓目标检测是实现采摘自动化的关键 环节。基于 YOLOv5 目标检测算法,采用 ShuffleNet 轻量级网络结构替代原模型的特征提取网络,并在骨干网络提取的特征 图后加入 SE 通道方向的注意力机制,结合 EloU 和 Alpha-IoU 损失函数,设计了一个 α-EloU 损失函数,给定参数 α 的值为 3, 统一指数化 IoU 损失函数,据此获得更准确的边界框回归和目标检测。改进的模型在草莓小目标数据集上平均检测精度均 值达到了 97.6%,其中成熟草莓的准确率为 99.4%,与 YOLOv3、YOLOv4 和 YOLOv5 相比,平均精度均值(mAP)分别提高 了 5.4%、2.9%和1.1%,该模型识别图像传输帧率为 125 fps,比原 YOLOv5 模型提升了 38 fps,该实验模型更适应于移动端 部署,为草莓采摘识别的自动化提供了一些理论基础。

关键词:目标检测;YOLOv5;轻量化网络模型;草莓识别;损失函数

**中图分类号:** TP183;S24 文献标识码:A 国家标准学科分类代码: 520.60

### YOLOv5-based lightweight network model for strawberry detection

Yang Shizhong Wang Ruibin Gao Sheng Shao Mingwei

(1School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, China)

Abstract: With the development of modern agricultural technology, the automation of strawberry production and picking is an inevitable trend, and strawberry object detection is a key link to achieve picking automation. In this paper, based on the YOLOv5 target detection algorithm, the ShuffleNet lightweight network structure is used to replace the feature extraction network of the original model, and an attention mechanism in the direction of the SE channel is added after the feature map extracted by the backbone network. Combining the EIoU and Alpha-IoU loss functions, an  $\alpha$ -EIoU loss function was designed, given a value of 3 for the parameter  $\alpha$ , to unify the exponentiated IoU loss function, whereby a more accurate bounding box regression and object detection was obtained. The improved model in this paper achieved a mean detection accuracy of 97.6% on the average strawberry small object dataset, with 99.4% accuracy for ripe strawberries, and improved mAP by 5.4%, 2.9% and 1.1% compared to YOLOv3, YOLOv4 and YOLOv5 respectively. The model recognises images transmitted at 125 fps, an improvement of 38 fps over the original YOLOv5 model. The experimental model is more adaptable to mobile deployment and provides some theoretical basis for the automation of strawberry picking recognition.

Keywords: object detection; YOLOv5; lightweight network model; strawberry detection; loss function

### 0 引 言

草莓是世界上具有较高经济价值的主要作物之一,在 人类的日常生产生活中具有越来越重要的研究意义<sup>[1]</sup>。 根据联合国粮食及农业组织(FAO)的数据,到 2026 年全 球草莓市场预计将达到 224.5 亿美元,高于 2020 年的 183.7 亿美元,其中 2008~2018 年,全球草莓产量增长了 39.4%,预计 2021~2026 年的复合年增长率为 3.4%<sup>[2]</sup>。 然而草莓很难储存,而且它们的香气或味道会随着时间的 推移而明显减少,草莓在采摘过程中会导致生产成本的增

收稿日期:2023-01-28

<sup>\*</sup>基金项目:山东省自然科学基金(ZR2020QF101)项目资助

加,目标识别与定位等新技术的应用在自动采摘过程中越 来越普及,采摘机器人已经成为当前农业领域的研究方向 之一<sup>[3]</sup>。在自然环境下,采摘过程中的光照条件,作物生 长状态,枝叶遮挡等是草莓小目标识别过程中的一大难 题<sup>[4]</sup>,草莓的目标检测功能受到复杂的农业环境的影响, 因此优化目标检测技术尤为重要。

近年来,随着计算机视觉技术的发展,出现了许多基 于深度学习的目标检测算法模型,与传统的机器学习的目 标检测相比,网络模型对目标特征的提取能力和物体检测 精度都有所提高,模型结构也具有很强的鲁棒性和泛化能 力。目前主流的物体检测算法主要有以 YOLO(you only look once)<sup>[5]</sup>和 SSD(single shot multi-box detector)<sup>[6]</sup>系 列为代表的单阶段目标检测算法和以 Fast R-CNN 系 列<sup>[7-8]</sup>为代表的双阶段目标检测算法。水果的目标检测已 经成为现代农业的重点研究方向之一,近年来取得了重大 进展。李林升等<sup>[9]</sup>在原有目标检测算法框架 Faster-RC-NN 的基础上加入两种不同卷积核大小的滑动窗口,提高 了模型特征提取能力,对苹果识别的准确度达到了 97.6%。刘云等<sup>[10]</sup>基于卷积神经网络结合 RGB 彩色背 景分割算法,对正常和有缺陷的苹果识别准确率达到了 97.3%,检测速度也达到了每秒5个样本。郭鑫等[11]基于 改进方向梯度直方图和支持向量机对苹果小目标同时进 行浅层和深层特征提取并加入特征金字塔获取特征提取 图特征融合,检测准确率平均值达到 90.46%,相比较传 统机器学习算法提高了检测精度。以 Fast R-CNN 系列 为代表的双阶段目标检测算法在特征提取过程中产生了 大量的候选区域,并使用神经网络来识别和定位候选区 域,耗时较长,不适合草莓目标的实时检测,单阶段目标检 测算法将图像直接输入特征提取网络,在输出层获得目标 的位置和类别信息,检测速度大幅提高,在实时目标检测 中更为普遍。蒋晨等<sup>[12]</sup>将 SSD 目标检测算法的特征提取 网络替换为可变形卷积的 ResNet50 网络结构,融合不同 网络层的特征语义信息,相比于传统 SSD 算法对小目标 的识别准确率达到了 86.8%。徐印赟等[13]提出了一种 YOLOv4 目标检测和非极大抑制相结合的水果检测算 法,改进原网络中重新分布池化核,增强感受野范围,对水 果目标检测的平均精度均值达到了 96.65%。徐晓光 等[14] 通过多尺度特征融合的方法改进 YOLO 算法,加入 跳跃卷积层,融合不同尺寸的特征图所包含的位置信息, 提高了模型的定位精度,在数据集上平均精度均值提高了 1.53%,但检测速度有轻微的降低。综上所述,在目标检 测过程中,检测速度和识别精度一直是评价模型的标准; 在深化网络的过程中,精度的提高会带来检测速度降低的 问题,采用轻量级模型结构识别精度又大幅度下降,对嵌 入式设备也是一个考验,应用到实际生产生活中便需要实 现目标检测速度和精度的双重考虑。

本文在原有 YOLOv5 目标检测模型的基础上<sup>[15]</sup>,用 轻量级网络模型代替骨干特征提取网络,加快模型训练速

## ■研究与开发

度,在特征提取图后加入挤压和激励(squeeze and excitation, SE)通道注意力机制,增强了对特征信息的提取,提 高检测精度,并通过改进损失函数加速模型的收敛,实现 了目标检测精度和识别速度的提高,完善了对草莓小目标 检测的识别模型。

由于现阶段没有单独的草莓公共数据集,本文建立了 一个草莓的小目标数据集,该数据集可以用来解决草莓研 究中的一些问题;用一个轻量级的网络模型替换深度网 络,通过在特征提取图后加入 SE 通道关注机制来增强特 征信息的提取,实现了对识别精度和检测速度的双重考 虑,为部署到嵌入式设备提供了模型基础;本文损失函数 是在原有 EIoU 的基础上,结合 Alpha-IoU 的思想,统一 了指数化 IoU 的计算方法,并赋值 α为3设计了损失函数 α-EIoU,更加灵活地实现了不同层次先验框的回归精度, 加速模型收敛,具有更强的鲁棒性。

### 1 材料与方法

#### 1.1 YOLOv5 目标检测模型

目前就目标检测速度而言,单阶段物体检测网络更适合于实时检测网络。YOLOv5 算法是在 YOLOv3<sup>[16]</sup>和 YOLOv4<sup>[17]</sup>基础上改进的目标检测算法模型,加入深度倍数(depth-multiple)和宽度倍数(width-multiple)参数分别 控制模型的深度和卷积核的数量,以获得不同的网络结构 模型,网络结构示意图如图 1 所示。

YOLOv5 主要有 5 种不同深度的网络的物体检测模 型,分别为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv51 和 YOLOv5x。本文选择 YOLOv5s 网络模型结构来检测 草莓图像的小目标数据集,并将识别精度和检测速度作为 识别标准。作为一个新的单阶段目标检测的网络模型, YOLOv5 与以前的网络模型相比,在输入侧的模型训练 阶段,使用了 Mosaic 数据增强等自适应方法来随机缝合 数据集。主干网络采用 FOCUS 结构对图像进行切片,通 过对原始输入的 RGB 三通道图像的宽度和高度信息进行 打分并汇总成通道,提高了网络速度。主干网络中的标准 卷积层采用 Conv+BN+SiLU 激活函数的组合,网络中 的 C3 结构主要是基于区域建议网络的思想,将输入的特 征图分成两部分,一部分进行卷积,另一部分与卷积运算 的输出融合,有效地提高了网络模型的学习能力, YOLOv5 模型的 Neck 层使用 FPN 结构与金字塔注意网 络(PAN)<sup>[18]</sup>相结合,以纳入全局信息。通过改变网络的 连接性来实现多尺度检测,有效地提高了网络的特征融合 能力,从而改善了小目标检测的性能。

### 1.2 迁移学习

在目标检测和识别领域,由于收集和注释数据集的难度,这限制了数据集的构建。数据集往往是训练过程中最 重要的部分。迁移学习是对基础网络的应用,通过在一个 任务中的训练,学习到另一个依赖性任务,新任务不需要 从头开始训练目标模型,大大减少了目标训练数据的要求





和训练时间<sup>[19]</sup>,迁移学习在很大程度上解决了训练数据 不足的问题。

迁移学习分为4类,基于实例的深度迁移学习、基于 映射的深度迁移学习、基于网络的深度迁移学习和基于对 抗的深度迁移学习。由于从0开始的自建数据集的训练 权重过于随机,特征提取效果不佳。深度迁移学习是将在 源域中预先训练好的网络的一部分,在包括其网络结构和 连接参数,并重新利用它,将其转化为目标数据集中深度 神经网络的一部分。

如图 2 所示,它将网络分为两部分,第一部分是特征转换,最后一层是分类器,首先,网络在源域使用大规模训 练数据集进行训练;其次,为源域预训练的网络的一部分 被转化为为目标域设计的新网络的一部分;最后,用微调 策略更新被转移的子网络。



#### 1.3 轻量化网络模型

随着人工智能的快速发展,深度神经网络模型被广泛 应用于图像分类、物体检测、物体跟踪和其他计算机视觉 任务。深度神经网络在日常生活中的应用性能逐渐引起 人们的关注。在边缘侧平台上部署高性能的神经网络模 型,使目标检测技术的应用能够在实际场景中实时运行, 如移动/嵌入式设备。由于移动平台的内存资源、存储空 间和算术资源较少,处理器性能较低,功耗受到限制,深度 网络模型无法在这些平台上部署并实现实时运行<sup>[20]</sup>。轻 量级网络模型在保持模型精度的同时,减少了模型参数的 数量和复杂性,与深度神经网络模型相比,更容易部署到 移动嵌入式平台。

轻量级网络模型旨在减少模型参数的数量和复杂性, 同时保持目标模型的准确性。主要分为轻量级网络结构 设计和模型压缩两类。轻量级网络结构一般采用卷积结 构方法,如深度可分离卷积、分组卷积,或使用全局池化操 作来代替完全连接的层。这些方法实现了通道降维,减少 了卷积过程中模型参数的数量,从而加快了网络模型的收 敛速度。常见的轻量级网络工作结构包括 GhostNet<sup>[21]</sup>、 MobileNet 系列<sup>[22-24]</sup>和 ShuffleNet<sup>[25]</sup>等模型。轻量化网络 结构模型解决了卷积神经网络的效率问题,减少了网络模 型参数,能够更广泛地应用于移动端应用。

#### 2 改进的 YOLOv5 目标检测模型

在使用迁移学习思想训练草莓数据集的过程中,使用 了 YOLOv5 网络模型结构,本文 YOLOv5 的目标特征提 取网络被替换为轻量级的网络模型,省去原网络模型的 FOCUS+CSPDarkNet53 结构,简化网络模型,并在主干 网络中加入图 3 所示的 CBRM 结构,区别于原模型的 FOCUS 结构,接着将轻量化网络 ShuffleNet 的子模块加 入迭代网络结构中,优化网络结构,加快训练速度,减少计 算参数,降低实际运行时间。该模型通过 SE 注意力机制

模块处理从骨干网络中提取的特征图,用不同权重处理的 特征图更注重目标的有效特征,抑制无关的特征信息。该 模型将损失函数分为3部分:面积损失、面积框架距离损 失和宽度高度损失,损失函数被赋予了一个参数值,以设 计α-EIOU损失函数,修改后的损失函数加速了模型训 练,并自适应地提高了回归精度。改进后的网络模型的结 构如图3所示。



图 3 改进的轻量化网络结构模型

### 2.1 ShuffleNet 轻量化网络结构

为了便于将该模型嵌入到移动平台中, ShuffleNet 网络结构是以点群卷积和通道清洗为主要结构进行去设计的,该网络减少了模型的计算量,同时保持了识别的准确性;点群卷积通过使用通道洗牌进行1×1卷积,增强了特征通道的信息交换,消除了卷积层堆叠的冗余性,由于减少了点群卷积的计算复杂度,模型的速度更快。

如图 4 所示, ShuffleNet 网络结构被设计为从使用逐 点组卷积和信道中断操作的残差块开始。该模型将残余 分支中的第一个卷积操作替换为逐点组卷积, 然后通过一



图 4 ShuffleNet 网络结构

## 研究与开发

个带有 ReLU 函数的 BN 层,接着进行信道混合操作,并 使用计算成本较低的 3×3 DW 卷积代替普通卷。在 DW 卷积过程之后,需要进行点群卷积来恢复通道的数量,以 便对捷径分支的输入通道数量进行求和,该网络结构利用 串联来实现通道方向的拼接,既使用了残差结构,又扩大 了通道的维度,计算量很小。

#### 2.2 通道注意机制

注意机制模块在提高特征图的特征学习方面有很大 改进。本文采用的挤压和激励注意机制模块<sup>[26]</sup>是通过挤 压和激励机制对特征图的通道信息进行处理。通道信息 被赋予不同的权重到特征图的不同位置,以获得更多的特 征学习信息,增强特征提取图的信息表示,SE通道注意力 机制模块的结构如图 5 所示。





SE 注意力机制首先通过全局池化操作来整合输入特征图的全局空间信息,将特征图的全局空间信息压缩到通 道描述符中,如式(1)所示表示对特征图 U ∈ R<sup>H×w×c</sup>上的 通道元素进行挤压机制处理。

$$Z_{c} = F_{sq}(U_{c}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_{c}(i,j)$$

$$\tag{1}$$

经过挤压机制处理汇总的信息特征再经过激励机制的处理,输入到全连接结构层和 ReLU 激活函数,SE 注意力机制模块的激励机制如式(2)所示。通道权重值 S 是通过两个全连接层  $W_1W_2$  对上一步得到的向量 Z 进行处理而得到的,对于重新分配的通道权重信息,S 通过式(3)所示的  $F_{scale}$  函数重新对特征图 U 进行权重赋值,得出通道信息处理后的U,尺寸大小不变,其大小是恒定的,SE 注意力机制只处理通道方向的信息特征,然后调整改变后的特征图,将特征图权重信息与输入的特征进行乘法运算。

 $\mathbf{S} = F_{ex}(\mathbf{Z}, \mathbf{W}) = \sigma(g(\mathbf{Z}, \mathbf{W})) = \sigma(W_2\delta(W_1\mathbf{Z}))$ 

(2)

$$U_{c} = \boldsymbol{F}_{scale}(U_{c}, \boldsymbol{S}_{c}) = \boldsymbol{S}_{c}U_{c}$$
(3)

### 2.3 交并比(IoU)损失函数

在物体检测中,IoU被用来衡量预测结果的位置信息的准确性<sup>[27]</sup>,IoU通过计算模型预测的目标位置与目标实际位置的偏差大小来处理预测的图像<sup>[28]</sup>。如式(4)所示,目标的实际面积与预测面积重叠越大,其值就越大,即 $0 \ll IoU \ll 1$ 。IoU的值越接近1,效果越好,IoU的值越大,预测区域的位置就越准确。在下面的公式中, $A \cap B$  代表目标的实际面积和预测面积之间的重叠面积, $A \cup B$  代表实际面积和预测面积作为一个整体所占的面积,重叠面积

损失的计算如式(5)所示。

 $IoU = \frac{|A \cap B|}{|A \cup B|}$ (4)  $L_{IoU} = 1 - IoU$ (5)

在 IoU 损失的计算中,预测区域框的位置修正是基于 IoU 损失的反向传播迭代产生的,并结合重叠面积对坐标 比例进行归一化处理,但它不能准确反映预测区域框与真 实区域框的相交或包含过程。利用最小外框最大化重叠 面积来优化两框之间的交叉过程,并充分考虑预测区域框 和真实区域框在不同情况下带来的垂直和水平方向,通过 计算两个区域框的中心点之间的归一化距离,重新确定最 小外框最大化重叠面积,并结合先验框加入惩罚项,加速 损失向回归精度收敛。CIoU<sup>[29]</sup>的计算方法如式(6)所示, 其中 b 表示预测区域框架的中心点,b<sup>st</sup>表示真实区域框 架的中心点,ρ 表示两个区域疯狂中心点之间的欧氏距 离,c 表示包含预测区域框架和真实区域框架的最小封闭 区域的对角线距离,然后加入影响因子 av 以反映区域框 架长宽比的差异,其中 a 的计算方法如式(7)所示,v 的计 算方法如式(8)所示,用来衡量长宽比的一致性。

$$L_{CloU} = L_{loU} + \frac{\rho^2(b, b^{gt})}{c^2} + av$$
(6)

$$a = \frac{v}{(1 - IoU) + v} \tag{7}$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2$$
(8)

为了提高模型的最佳相似度,EIoU<sup>[30]</sup>使用区域框的 宽度和高度的真实差异及其置信度,并分离出长宽比的影 响因素 av,分别计算区域框和锚定框的 w 和 h。EIoU 的 计算分为重叠区域的损失、两个区域框之间中心点距离的 损失以及宽度和高度的损失,如式(9)所示,宽高损失是在 CIoU 的基础上利用区域框和先验框之间的宽高差计算 的,其中  $C_w$  和  $C_h$  表示同时覆盖预测区域框和真实区域 框的最小外框的宽度和高度,同时加入基于 IoU 的现有损 失的统一幂化的 Alpha-IoU 损失函数<sup>[31]</sup>,具有单个参数  $\alpha$ ,更加适用于精确的先验框回归和目标检测,为小数据集 提供了更强的鲁棒性,随着区域框回归精度的提高,自适 应地改善损失和梯度的权重。修改后的  $\alpha$ -EIoU 损失函数 如式(10)所示。

$$L_{EloU} = L_{IoU} + \frac{\rho^{2}(b, b^{gt})}{c^{2}} + \frac{\rho^{2}(w, w^{gt})}{C_{w}^{2}} + \frac{\rho^{2}(h, h^{gt})}{C_{h}^{2}}$$
(9)
$$L_{a \cdot EloU} = 1 - IOU^{a} + \frac{\rho^{2a}(b, b^{gt})}{c^{2a}} + \frac{\rho^{2a}(w, w^{gt})}{C_{w}^{2a}} + \frac{\rho^{2a}(w, w^{gt})}{C_{w}^{2a}}$$
(10)

由式(10)可知,由于如果预测区域框的宽度和高度满 足特定值,CloU中惩罚项的作用就会下降,所以 w 和 h 值不能一直同步增加和减少。EloU 损失将 CloU 惩罚项

### 2023年4月 第42卷第4期

中的宽度和高度的相对比例替换为区域盒的宽度和高度 的实际值。该损失函数分3部分计算,直接使用边缘长度 作为惩罚项,解决了宽度和高度值不能同时放大或缩小的 问题,在惩罚项的基础上,结合给定参数值  $\alpha$ 设计了  $\alpha$ -EloU损失函数,以加速收敛,自适应提高回归精度。

#### 3 模型训练和分析

### 3.1 实验环境

本文使用 YOLOv5 来改进基本的网络检测模型,使 用轻量化网络模型 MobileNet、GhostNet 和 ShuffleNet 来 比较和实验最佳的轻量化目标检测模型。为了加快网络 模型的训练,实验环境使用了 Pytorch1.8 深度学习框架、 Ubuntu 18.04 操作系统和 NVIDIA RTX 3090(24 GB) GPU,配置 CUDA 11.0 版本及其对应的 cudnn 8.0 版本 加速模型训练,并在训练过程中使用随机梯度下降(SGD) 来优化更新网络模型。

原 YOLO 目标检测模型采用迁移学习的方法直接应 用到本文草莓数据集,冻结主干特征提取部分,加载官方 模型预训练权重,加速收敛,其余轻量化网络模型实验为 得到更精确的实验结果采用不加载预训练模型的方法重 新对数据集进行训练,以获得更准确的实验结果。

本文使用的数据集是自行采集的田间自然环境中的 草莓图像,主要是在山东省青岛市红石崖草莓种植中心用 华为手机拍摄的,黑白像素 2 400×10<sup>4</sup> pixels,彩色像素 1 600×10<sup>4</sup> pixels采集的白天温室栽培的草莓图像,从最 初采集的图像中选择了近 2 000 幅草莓图像进行目标检 测,经过数据增强操作得到近 8 000 张草莓图像,训练集、 测试集和验证集的比例被设定为 8:1:1进行实验,改进后 的模型使用表 1 所示的超参数进行训练,不加载源模型的 预训练权重,实验中使用的具体网络超参数如表 1 所示。

表1 网络超参数

训练超参数名称	参数值	
图像输入尺寸	$640 \times 640$	_
训练批次	300	
批处理大小	32	
优化器	SGD	
学习率下降方式	cos	
初始学习率	0.01	

### 3.2 评价指标

在本文中,目标检测模型的经典性能评价指标有召回率(recall)、精确度(precision)、每个类别的平均精度(average precision, AP)和平均精度均值(mean average precision, mAP)<sup>[32]</sup>,以及模型处理图片的帧率来对比各个模型间的性能,其中mAP主要代表每个类别的P-R曲线所围成的面积,帧率值越高,证明模型结构运行更迅速流畅。本文草莓数据集划分为两个类别,其中成熟的草莓

## 研究与开发

用标签 ripe 表示,未成熟草莓用标签 unripe 表示。精确度 和召回率的计算方法分别如式(11)和(12)所示。

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$Recall = \frac{TP}{TP + FN}$$
(12)

式(11)中精确率是衡量分类器中能正确识别样本的 能力,即在某种类别中被识别出正样本的目标中能正确预 测的样本所占的总比例,式(12)为召回率主要代表某种类 别的预测正确的样本数量占实际样本总数量的比值。*TP* 是指某种类别里正样本数据被正确识别出的目标数量,在 本文草莓数据集中,即是被正确识别出的成熟草莓数量或 未成熟草莓数量;FN 为某种类别里正样本数据没有被正 确识别出的目标样本数量,在本文草莓数据集中,即是未 被正确识别出的成熟草莓数量或未成熟的草莓数量;FP 是指某种类别里非正样本数据识别结果成正样本目标的 数量,如在成熟草莓类别识别中不是成熟草莓的被模型识 别为了成熟草莓。

$$AP = \int_{0}^{1} P(R) dR$$
(13)

$$mAP = \frac{1}{N} \sum_{i=1}^{N} \int_{0}^{1} P(R) dR$$
(14)

式中:AP 是指以 Recall 值为横轴, Precision 值为纵轴所 围成的 P-R 曲线面积,以积分的形式计算 P-R 曲线围成面 积;mAP 是取数据集中所有类别的准确率的平均值,也是 由目标类别的召回率和精确率决定;N 表示数据集中的 目标类别个数。本文实验用平均准确率和平均精度均值 衡量目标检测网络模型的性能。

#### 3.3 实验分析

模型训练过程中有 3 种不同类型的损失,分别为先验 框损失(box-loss)、目标损失(obj-loss)和分类损失(clsloss)。先验框损失表示算法能够定位物体中心的程度,以 及预测的区域框覆盖目标的程度;目标损失是对目标区域 存在的概率的计算,以确定该区域是否包含目标;分类损 失表示模型将目标预测为正确类别的能力。mAP@0.5 表示该模型在 IoU 阈值设置为 0.5 时成熟草莓 ripe 和未 成熟草莓 unripe 两个类别上识别准确率的平均值即平均 精度均值,mAP@0.5:0.95 表示在不同 IoU 阈值(0.5~ 0.95,步长为 0.05)时目标各类别的平均精度均值,IoU 阈 值设置的数值越大,精度越低。

本文改进的轻量级模型的训练时间约为 3 h,模型训 练过程中在训练集和验证集上的不同表现如图 6 所示,从 图 6 可以看出,训练阶段和验证阶段的损失值在前 100 个 epoch 中下降最快,而后损失值逐渐趋于平稳,同时随着损 失的快速下降,改进模型的精确度、召回率在 100 次迭代 轮数快速上升,各个阶段不同阈值的各类别平均精度均值 在大约 200 个 epoch 后趋于平稳。



图 6 本文改进模型实验曲线

本文改进模型的 P-R 曲线如图 7 所示,由图 7 可知本 文改进模型对成熟草莓 ripe 的识别精确度达到了 99.4%,未成熟草莓 unripe 的识别精度为 95.7%,平均精 度均值达到了 97.6%,识别效果得到了显著提升。

### 3.4 实验结果

YOLO 是第1个提出的单阶段物体检测算法,取消了 双阶段检测的区域建议框筛选过程。现阶段 YOLO 算法 中常见的物体检测模型有 YOLOv3、YOLOv4 和 YOLOv5,其中 YOLOv3采用残差网络模型结构来深化 网络结构,骨干网络采用 DarkNet-53 结构,并采用多尺 度预测;YOLOv4 使用 CSPDarket53 骨干特征提取网络 的马赛克数据增强方法,并在骨干特征提取网络后引入 SPP 结构,结合 Down Sampling 操作和特征金字塔来设计 目标检测模型;YOLOv5 采用 Mosaic 数据增强方法,在骨 干网络中加入了 FOCUS 结构,并将 SPP 结构改为 SPPF 网络结构,在网络结构的隐含层中使用 ReLU 激活函数, 在检测层中使用 Sigmoid 激活函数,使用二元交叉熵和 Logits 损失函数计算类概率和目标得分的损失。如表 2 所示,上述目标检测模型 1~3 的结构采用迁移学习方法, 加载预训练权重,并应用于本文的草莓数据集作为对比测 试模型。

由于卷积层等结构的差异性,不同的神经网络在训练



图 7 改进模型 7 的 P-R 曲线

模型	方法	骨干网络	网络参数		
1	YOLOv3	DarkNet53	61 529 119		
2	YOLOv4	CSPDarkNet53	63 943 071		
3	YOLOv5	CSPDarkNet53	7 025 023		
4	YOLOv5	MobileNetv2	4 859 583		
5	YOLOv5	GhostNet	5 486 979		
6	YOLOv5	ShuffleNet	845 823		
7	YOLOv5	$ShuffleNet\!+\!SE$	3 797 439		

表 2 模型结构参数

### 2023年4月 第42卷第4期

过程中会有不同的参数量,一般来说,层数越大,参数数越 多,参数数对模型的训练有不同的影响。原有模型 YOLOv5s物体检测算法采用了 CSPDarkNet53 网络结构,这是结合 CSPNet 结构和 DarkNet 网络的思想提出的 特征提取网络,具有残差结构。为了减少卷积过程中模型 参数量,本文采用了轻量级网络模型结构作为特征提取网 络,模型参数的数量大幅减少。本文的轻量级网络物体检 测模型是替代 YOLOv5s 的骨干网络结构,模型参数量的 减少加快了网络模型的速度。相比之下,ShuffleNet 轻量 级网络模型的参数数量最少,更适合于在嵌入式移动平台 上部署。

本文基于 YOLOv5s 的目标检测模型用轻量化网络 取代了原网络模型的骨干网络,并替换了特征提取网络, 在保持精度的同时实现了模型参数数量的压缩,比较实验 中使用的轻量级网络模型是 MobileNetv2、GhostNet 和 ShuffleNet。具有最佳精度的 ShuffleNet 轻量级网络模型 加入到 SE 通道注意力机制,以进一步提高小目标检测精 度。从表 2 可以看出,轻量级网络模型的参数数量呈指数 级减少,对目标的检测速度得到了极大的提高。

从表 3 可以看出,本文模型 6 选择轻量化网络结构模型 ShuffleNet 作为骨干特征提取网络,与模型 4 采用 MobilenetV2 作为骨干网络的模型 5 采用 GhostNet 作为骨 干网络相比,更适应本文的草莓数据集,其中成熟草莓的 识别准确率达到 99.4%,平均准确率平均值也达到

		成熟草莓		未成熟草莓			m A D@ 0 E		
模型 方法 F	Precision / %	Recall/%	$\mathrm{AP}/\%$	Precision / %	Recall/%	$\mathrm{AP}/\frac{9}{10}$	/ %	帧率/fps	
1	YOLOv3	95.4	94.7	93.7	94.1	91.4	90.6	92.2	73
2	YOLOv4	99.4	93.5	96.6	91.8	89.1	92.7	94.7	69
3	YOLOv5	99.7	97.0	98.4	95.8	91.1	94.6	96.5	87
4	MobileNetv2	99.4	92.5	96.1	93.1	80.7	88.6	92.4	115
5	GhostNet	99.6	91.7	95.7	94.6	87.1	91.9	93.8	109
6	ShuffleNet	99.6	93.7	99.4	97.8	88.1	93.7	95.2	134
7	Shufflenet + SE	99.7	97.7	99.4	88.7	92.7	95.7	97.6	125

表 3 各模型实验结果

95.2%,分别比轻量级模型 MobileNetv2 和 GhostNet 提高 2.8% 和 1.4%。模型 6 比原模型 3 的帧率提高了 47 fps,但准确率下降了 1.3%。

从图 7 和表 3 的实验结果可以看出,与模型 6 相比, 本文改进后的模型 7 将 mAP 的平均精度提高了 2.4%, 未成熟草莓的识别精度提高了 2.0%。与模型 3 相比,平 均准确率平均值 mAP 提高了 1.1%,尤其是将成熟和未 成熟草莓的识别准确率提高了 1.0%和 1.1%,每秒处理 的图像数也提高了 38 fps。改进后的模型 7 在识别成熟 草莓方面达到了 99.5%的准确率,完成了识别准确率和 检测速度都得到了提高。 各实验模型检测效果如图 8 所示,本文改进模型检测 效果如图 9 所示,其中图 9(a)表示草莓检测过程的热力 图,图 9(b)表示草莓检测效果的框选图。由图 8 可以看 出,模型 1~6 对目标草莓的识别框选情况,红色框表示成 熟的草莓,蓝色框表示未成熟的草莓,目标框右上角的数 值表示草莓目标识别过程中位置信息的置信度,各个模型 对草莓的识别情况各个不同,其中效果最好的是模型 3 的 YOLOv5 原模型,本文将主干网络换成轻量化网络 Shufflenet 后,模型 6 识别率轻微下降,识别框选效果也略微下 降,本文模型 7 在此基础上加入 SE 通道注意力机制,并改 进损失函数,其识别效果如图 9 所示,对小型目标的识别



(d) 模型4

(e)模型5

(f) 模型6





(a) 热力图

(b) 检测图

图 9 本文改进模型 7 检测效果

效果最好,框选准确度最高,从图 9(a)可以看出,热力图 是一个学习识别目标中心的空间和语义信息的训练过程, 草莓检测过程中,通过去掉检测框对物体特征信息的分 析,图片中的颜色信息越深,学习到的物体特征信息越多, 目标检测框的置信度越高,热力图的分布主要从物体的中 心点开始,与原图比较,图片中的草莓被提取了位置特征 信息,越靠近草莓物体的颜色越深,红色位置表示草莓物 体中心点的位置。如图 9(b)所示,识别框已经准确地框 定了草莓目标,其中草莓的分类显示正确,置信度代表了 图中目标的原始位置与算法所能识别的目标位置之间的 相似度,其中对小型草莓目标也实现了框选,综合对比,识 别效果最好。

从以上实验结果可以看出,本文的轻量级草莓目标识 别模型在小型草莓物体的位置选择和识别方面取得了良 好的效果,模型的轻量级结构也为嵌入式移动终端的部署 提供了理论依据,可以更好地应用于草莓采摘机器人。

### 4 结 论

为了将小目标检测技术应用于草莓自动采摘的日常

生产生活中,本文基于 YOLOv5 物体检测算法设计了一 个轻量级网络模型。该模型采用轻量级网络的 ShuffleNet 网络结构替代原模型的特征提取网络,并在后骨网 络提取的特征图后加入 SE 通道方向的注意机制。同时 结合 EIoU 和 Alpha-IoU 损失函数,给定一个参数 α,设计 一个 α-EIoU 损失函数,以满足不同层次的先验框回归精 度,同时对小数据集提供一定的鲁棒性。

实验结果表明,YOLOv3、YOLOv4和 YOLOv5模型 在不同程度上加深和拓宽了骨干特征提取网络,这使得参 与者的数量大大增加,增加了部署到移动端的难度,对比 测试表明,YOLOv5的物体检测算法在小型草莓目标数 据集上效果最好。实验采用轻量级网络模型 Mobile-NetV2、GhostNet和 ShuffleNet的轻量级网络结构来替 代YOLOv5模型的CSPDarkerNet53深度网络结构,模型 的大小被压缩,参数数量明显减少。在最有效的轻量级网 络模型 ShuffleNet 的基础上为提取的特征图增加了 SE 通道方向的关注机制和改进的 IoU 损失函数,基于 EIoU 和 α-IOU 损失函数设计了一种 α-EIoU 损失函数,加快了 网络模型的收敛速度,自适应优化了损失的加权值和梯

度,提高了先验框回归精度,提高了目标识别精度和检测 速度,本文的模型更适应于部署到嵌入式移动终端,为草 莓的自动识别应用提供了一定的实验基础。

#### 参考文献

- [1] KOLLE S R, SHANKARAPPA T H, RAHIMI B A, et al. Review of trends in strawberry research from 1960 to 2016 [J]. Journal of Agricultural & Food Information, 2019, 20(1): 25-38.
- [2] PARK S, KIM J W. Design and implementation of a hydroponic strawberry monitoring and harvesting timing information supporting system based on nano ai-cloud and iot-edge[J]. Electronics, 2021, 10(12): 1400.
- [3] 刘成良, 贡亮, 苑进, 等. 农业机器人关键技术研究现 状与发展趋势[J]. 农业机械学报, 2022, 53(7): 1-22, 55.
- [4] 郑太雄,江明哲,冯明驰.基于视觉的采摘机器人目标 识别与定位方法研究综述[J].仪器仪表学报,2021, 42(9):28-51.
- [5] 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, 2016: 779-788.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2016: 21-37.
- [7] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] 李林升,曾平平.改进深度学习框架 Faster-RCNN的 苹果目标检测[J]. 机械设计与研究,2019,35(5): 24-27.
- [10] 刘云,杨建滨,王传旭.基于卷积神经网络的苹果缺陷 检测算法[J].电子测量技术,2017,40(3):108-112.
- [11] 郭鑫,郝骞,杨风. 基于改进 HOG 和 SVM 的苹果多 目标检测方法[J]. 国外电子测量技术,2022,41(11): 154-159.
- [12] 蒋晨,钱永明,姚兴田,等. 基于可变形卷积改进 SSD 算法的目标检测方法[J]. 电子测量技术,2022, 45(16):116-122.
- [13] 徐印赟,江明,李云飞,等.基于改进 YOLO 及 NMS

### 2023年4月 第42卷第4期

的水果目标检测[J]. 电子测量与仪器学报,2022, 36(4):114-123.

- [15] 侯学良,单腾飞,薛靖国.深度学习的目标检测典型算 法及其应用现状分析[J].国外电子测量技术,2022, 41(6):165-174.
- [16] REDMON J, FARHADI A. YOLOv3: An incremental improvement [C]. Computer Vision and Pattern Recognition, 2018, DOI: 10. 48550/arXiv. 1804. 02767.
- [17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[C]. Computer Vision and Pattern Recognition, 2020.
- [18] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation [C]. Computer Vision and Pattern Recognition, 2018.
- [19] 刘鑫鹏,栾悉道,谢毓湘,等.迁移学习研究和算法综 述[J].长沙大学学报,2018,32(5):28-31,36.
- [20] 葛道辉,李洪升,张亮,等.轻量级神经网络架构综 述[J].软件学报,2020,31(9):2627-2653.
- [21] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [22] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[C]. Computer Vision and Pattern Recognition, 2017.
- [23] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [24] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [25] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [27] REZATOFIGHI H, TSOI N, GWAK J Y, et al.

Generalized intersection over union: A metric and a loss for bounding box regression[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.

- [28] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network [C]. Proceedings of the 24th ACM International Conference on Multimedia, 2016: 516-520.
- [29] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 52 (8): 8574-8586.
- [30] ZHANG Y F, REN W, ZHANG Z, et al. 2022. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing 2022, 506: 146-157.
- [31] HE J, ERFANI S, MA X, et al. α-IoU: A family of power intersection over union losses for bounding box regression[J]. Advances in Neural Information Pro-

cessing Systems, 2021, 34: 20230-20242.

[32] HENDERSON P, FERRARI V. End-to-end training of object class detectors for mean average precision [C]. Computer Vision-ACCV 2016: 13th Asian Conference on Computer Vision, Springer, 2017: 198-213.

### 作者简介

杨世忠,博士,副教授,主要研究方向为控制理论与应 用、建模与智能控制。

E-mail:yszxlj@126.com

王瑞彬(通信作者),硕士研究生,主要研究方向为计 算机视觉、智能检测。

E-mail:wangruibin009@163.com

高升,博士,主要研究方向为农产品无损智能检测、机 电一体化技术及装备。

E-mail:401116575@qq. com

邵明伟,博士,主要研究方向为计算机视觉。 E-mail:373851680@qq.com