

## 改进YOLOv5轻量级网络的柑橘检测方法

高新阳<sup>1</sup>, 魏 晟<sup>2</sup>, 温志庆<sup>2</sup>, 于天彪<sup>1</sup>

1. 东北大学 机械工程与自动化学院, 沈阳 110000

2. 智能机器人工程研究中心 季华实验室, 广东 佛山 528000

**摘 要:**针对现有的柑橘检测算法准确率低、模型参数量大、检测实时性差、不适用移动采摘设备等问题,提出一种基于改进轻量模型YOLO-DoC的柑橘检测方法。引入Bottleneck结构的ShuffleNetV2网络作为YOLOv5骨干网络模型,构造轻量化网络。同时加入无参型SimAM注意力机制提高复杂环境下对目标的识别精度。为了提高检测网络对于目标果实的边界框定位精度,通过引入Alpha-IoU边界框回归损失函数的方法来获取目标的边界框。实验显示,YOLO-DoC模型的P(precision)值和mAP(mean average precision)值分别为98.8%和99.1%,参数量缩减为YOLOv5网络的1/7,模型的大小为2.8 MB。改进后的模型相比于原网络模型具有识别速度快、定位准度高以及占用内存少的优势,在满足精准采摘工作要求的前提下可以提高采摘效率。

**关键词:**神经网络;注意力机制;YOLOv5;柑橘检测;损失函数;ShuffleNetV2;Alpha-IoU

**文献标志码:**A **中图分类号:**TP391.4 **doi:**10.3778/j.issn.1002-8331.2212-0023

## Citrus Detection Method Based on Improved YOLOv5 Lightweight Network

GAO Xinyang<sup>1</sup>, WEI Sheng<sup>2</sup>, WEN Zhiqing<sup>2</sup>, YU Tianbiao<sup>1</sup>

1. School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110000, China

2. Jihua Laboratory, Intelligent Robot Engineering Research Center, Foshan, Guangdong 528000, China

**Abstract:** Aiming at the problems of the existing citrus detection algorithms, such as low accuracy, large amount of model parameters, poor real-time detection, and unsuitability for mobile picking equipment, a citrus detection method based on the improved lightweight model YOLO-DoC is proposed. This paper introduces the ShuffleNetV2 network of the Bottleneck structure as the YOLOv5 backbone network model to construct a lightweight network. At the same time, the non-parametric SimAM attention mechanism is added to improve the recognition accuracy of targets in complex environments. In order to improve the positioning accuracy of the bounding box of the target fruit by the detection network, the bounding box of the target is obtained by introducing the method of Alpha-IoU bounding box regression loss function. Experiments show that the P(precision) value and mAP(mean average precision) value of the YOLO-DoC model are 98.8% and 99.1%, respectively, and the number of parameters is reduced to 1/7 that of the YOLOv5 network, and the size of the model is 2.8 MB. Compared with the original network model, the improved model has the advantages of fast recognition speed, high positioning accuracy and less memory usage. It can improve the picking efficiency under the premise of meeting the requirements of precise picking work.

**Key words:** neural network; attention mechanism; YOLOv5; citrus detection; loss function; ShuffleNetV2; Alpha-IoU

我国土地资源中丘陵山区是以柑橘为代表的水果主要产地,其面积高达69%的占比<sup>[1]</sup>。丘陵山区的水果种植产业普遍存在种植密度大、规模小、分散性大、地形起伏多变和陡弯多等特点<sup>[2-5]</sup>。而柑橘采收基本以人工为主,基于丘陵地区的复杂地势和种植管理规范

相应缺乏等原因,导致采摘作业成为既耗时费力,又对于人身存在不小安全隐患的一个环节,约占整个生产过程作业量的40%左右<sup>[6]</sup>。因此,为实现基于视觉系统的机械臂柑橘果实采摘项目,对于柑橘果实的识别与定位研究尤为重要,如何在不同因素的影响下提高柑橘果实

**基金项目:**广东省基础与应用基础研究基金(2020B1515120070)。

**作者简介:**高新阳(1996—),男,硕士研究生,研究方向为机器视觉、图像处理,E-mail:gaoxinyang006@163.com;魏晟(1990—),男,博士研究生,研究方向为机器视觉、机器人操作系统。

**收稿日期:**2022-12-02 **修回日期:**2023-03-13 **文章编号:**1002-8331(2023)11-0212-10

的视觉定位精准度也是重要的研究方向。

在果蔬采摘领域已有很多国内外学者基于传统机械视觉提出了多种方案。Bac等<sup>[7]</sup>融合自适应阈值分割、支撑电缆显著性检测和视觉三维特征表达方法,在变化的光照条件下实现了大棚作物采摘点的定位。Huang等<sup>[8]</sup>在OHTA色彩空间中结合色彩直方图和三元阈值图像分割法提取了草莓果实的承重枝条,通过匹配草莓果实区域质心和枝条的梯度信息定位采摘点。Sengupta等<sup>[9]</sup>结合形状分析、Canny检测器、Hough变换、纹理支持向量机和SIFT算子,实现了野外环境中柑橘果实的检测。He等<sup>[10]</sup>提出了一种基于线性判别分析的未成熟荔枝卷积特征提取方法,并结合Hough圆变换实现荔枝果实定位,但由于未能较好融入果实与绿叶干扰物之间的色调差异特性,该方法仅获得了约76.4%的果实识别与定位准确率。

由于传统机械视觉受限于对环境依赖性强、复杂环境下鲁棒性差等劣势,在实际工况中未能满足高效智能的采摘需求。因此大量学者将人工智能和神经网络在视觉检测领域的创新应用于农业果实采摘项目。黄彤铨等<sup>[11]</sup>提出一种改进版的YOLOv5柑橘识别模型,为模型引入CBAM<sup>[12]</sup>注意力机制并使用 $\alpha$ -IoU损失函数,提高了遮挡目标的检测精度和边界框的定位精度。吕石磊等<sup>[13]</sup>将MobileNetV2<sup>[14]</sup>作为YOLOv3的骨干网络提出了YOLOv3-LITE轻量级网络,提高了柑橘果实的检测速度。冯硕等<sup>[15]</sup>以ResNet<sup>[16]</sup>为检测网络,利用双目视觉原理进行果实定位并设计机械臂控制系统,从而建立起一套奇异果识别定位采摘流程。陈锋军等<sup>[17]</sup>提出一种基于EfficientDet<sup>[18]</sup>网络的深度模型进行油橄榄果实成熟度检测的方法,检测准确率可以达到92.8%。张皓婷等<sup>[19]</sup>利用FasterRCNN<sup>[20]</sup>方法进行番茄果实的视觉检测研究,对于番茄的平均检测结果的准确率可达90.1%且执行效果较好。

上述方法虽然在检测模型的精度和速度上都有所改进,且针对水果采摘流程也给出了完整可行方案,但无法兼顾提升检测速度和精度的要求。本文使用轻量化网络ShuffleNetV2<sup>[21]</sup>作为骨干网络,并在ShuffleNetV2中引入Bottleneck瓶颈结构同时加入SimAM注意力机制<sup>[22]</sup>,提出一种轻量级的满足柑橘采摘流程需要的检测模型YOLO-DoC。并使用Alpha-IoU<sup>[23]</sup>边界框回归损失函数作为评价指标,提高柑橘果实的定位精度,进一步满足了的抓取环节工作要求。同时在不同光照、果实遮挡情况的数据集中进行对比训练,以证明改进模型的优越性。经实验验证,该套流程所涉及的算法和理论满足了柑橘果实采摘的需求,对遮挡果实也能有效地检测与定位,降低了工况环境和设备的成本部署,同时也能满足低算力平台的运行能力要求。

## 1 材料与方法

### 1.1 数据采集

本文所使用的柑橘图像数据在广东省佛山市季华实验室柑橘种植区采集,使用Intel RealSense相机在距离柑橘目标0.3至0.5米处对果树进行拍摄。为适应自然光照下的多种采摘环境,拍摄角度环视各个方向,充分采集到晴天、阴天、顺光和逆光等多种不同光线下300张照片。为提高对不同种类柑橘果实的识别能力,通过网络获取200张不同柑橘种植区的柑橘照片。随机选取原图像数据分别进行水平、垂直翻转操作。再对部分原数据通过调整饱和度和色差进行光照适应性增强,从而将数据集增加到1500张。光照适应性增强效果如图1所示。通过人工筛选去除图像质量较差的样本之后,最终数据集为1435张。



图1 光照适应性增强

Fig.1 Enhanced light adaptability

### 1.2 数据集准备

对上述处理后的数据集用LabelImg数据标注工具进行人工标定,并按照7:2:1的比例将数据集划分为训练集、测试集和验证集,图片的数量分别为1005张,287张和143张。然后,将标注好的标签以xml的格式进行存储。考虑到影响柑橘检测效果最直接的因素就是枝叶枝干以及果实之间的遮挡问题,因此对数据集中不同遮挡情况进行分类。将测试集中含有被遮挡样本数低于30%的图片划分为轻度遮挡数据集A,共207张图片。含有被遮挡样本数高于30%的图片划分为高度遮挡数据集B,共80张图片。其中数据集A包含的样本数量分别为1522个,数据集B中包含的样本数量为604个。

## 2 YOLO-DoC改进算法

### 2.1 通道混洗结构

通道混洗(channel shuffle)结构是ShuffleNet网络



的核心内容,它在分组卷积(group convolution)的基础上进行一定的改进。分组卷积将输入层的不同特征图进行分组,然后采用不同的卷积核对每个组分别进行卷积,从而降低卷积的计算量,如图2(a)所示。常规的卷积操作是在所有输入的特征图上进行卷积,是一种通道紧密连接方式,即全通道卷积,而大量的全通道卷积无疑会需要繁重的计算量,从而影响网络结构的高效性。而分组卷积由于只进行组内卷积操作,阻挡了不同组间的特征图通信功能,从而会降低网络的特征提取能力。而通道混洗在不增长计算量的前提下将不同组内的特征图进行重组,因此可以改善分组卷积的不足,保证不同组特征图间的信息交流,如图2(b)所示。在程序实现上将输入层分为  $G$  组,每组通道数设为  $N$ ,将所有通道整理为  $(G \times N)$  的矩阵形式,然后将该矩阵进行转置得到  $(N \times G)$ ,最后将矩阵降维至一维即可完成通道混洗。

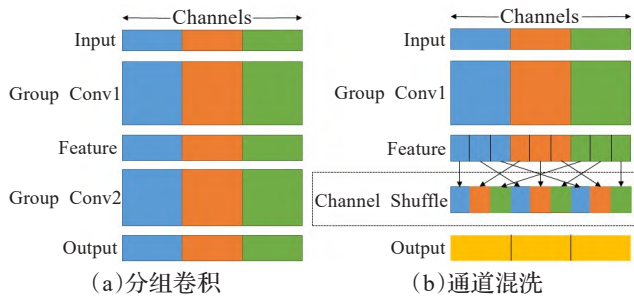


图2 通道混洗原理图

Fig.2 Channel shuffle schematic

## 2.2 SimAM 注意力模块

通过在 ShuffleNetV2 中增加 SimAM 注意力机制模块增强检测模型对复杂环境下柑橘果实的检测精度,提高对目标的关注度。

SimAM 注意力模块是一种基于人脑注意力机制特性提出的无参化 3D 注意力模块。该模块通过能量函数的解析计算提高了注意力权值的计算速度,相比于现有的通道/空域注意力模块具有轻量性和灵活性的优势。

SimAM 注意力模块通过度量神经元之间的线性可分性的方法来评估每个神经元的重要性。因此提出能量函数,经一系列简化,最小能量函数表达式如式(1)所示:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (1)$$

其中,  $t$  表示神经元,  $\lambda$  为超参数,  $\hat{\mu}$  和  $\hat{\sigma}^2$  的表达式如式(2)和式(3)所示:

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i \quad (2)$$

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2 \quad (3)$$

其中,  $x$  为神经元  $t$  的相邻神经元。由式(1)可知,能量越低,神经元  $t$  与周围神经元的区别越大,重要性则

越高,因此,神经元的重要性可由  $1/e_t^*$  获得。最后通过式(4)所示的方式对特征进行增强处理。

$$\tilde{X} = \text{sigmoid}(\frac{1}{E}) \odot X \quad (4)$$

与 SE 单通道注意力和 CBAM 空间注意力和通道注意力串行组合不同, SimAM 直接对独立的神经元进行重要性评估,能够在不提高模型的参数量的同时对重要的神经元赋予更高的权重。

## 2.3 ShuffleNetV2 改进结构

ShuffleNetV2 网络是由旷视科技在 ECCV 中提出的一种轻量型网络结构。论文中作者提出了四条评价高效适用于轻量化网络的准则,用于衡量网络计算的复杂度。输入输出通道数相同,使用少量的分组并且减少小卷积的使用可以对轻量网络进一步优化。基于该准则在 ShuffleNetV1 网络的基础上对 block 进行新的设计。在 Shuffle Block1 模块中对输入的特征图在通道维度进行了分支(channel split),等比例划分为两组通道。将其中一组进行连续三组卷积操作并且输入和输出通道数相同,其中两组为  $1 \times 1$  的普通卷积,中间一组为  $3 \times 3$  的深度可分离卷积(DWConv)。完成卷积操作后再与另一组通道进行通道数相加(Concat)操作,紧接着对两分支通道相加后的结果进行 Channel Shuffle 从而保证两个分支的信息交流,通道融合。Shuffle Block2 将特征图数量等分进行分支,两个分支均采用了  $3 \times 3$  步长为 2 的 DWConv 卷积进行降维处理,再将两分支各自处理后的结果进行通道数相加操作,使其通道数为输出通道数的两倍,最后同样使用通道混洗操作实现不同组间通道的信息交流。

为使 ShuffleNetV2 网络满足柑橘检测的工作需求,对该网络结构进行如下改进:

(1)针对原结构中 ReLU 激活函数不平滑特性可能带来的部分神经元失效的情况,将 ReLU 激活函数用 SiLU 激活函数进行替换。ReLU 和 SiLU 函数图像对比如图 3 所示。SiLU 函数的平滑性能够有效地提高神经元的使用效率,从而提高模型的整体性能。

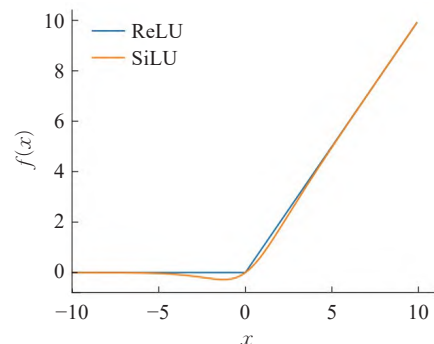


图3 ReLU 和 SiLU 激活函数对比

Fig.3 Comparison of ReLU and SiLU activation functions

(2)将 Shuffle Block 模块中右分支 DWConv 卷积后的  $1 \times 1$  普通卷积替换为 Bottleneck 瓶颈结构。Bottleneck

结构在ResNet网络中提出,其作用是用多个小尺寸卷积代替大尺寸卷积,从而减少模型的参数量。本文使用的Bottleneck结构将输入特征层一分为二,其中一部分经过kernel size为 $1\times 1$ ,stride为1的卷积、BN和SiLU层,之后经过kernel size为 $3\times 3$ ,stride为1,padding为1的卷积、BN和SiLU层,通道数经过卷积后保持不变。最后与另一部分特征层进行相加操作,具体结构如图4所示。而用Bottleneck结构替代 $1\times 1$ 卷积一定程度会增加模型整体的参数量,但同时会提高Shuffle Block模块的目标检测能力。

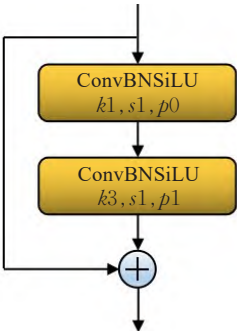


图4 Bottleneck 结构图

Fig.4 Bottleneck structure diagram

改进后的Shuffle Block模块如图5所示,为方便对比较,命名为S2 Block。

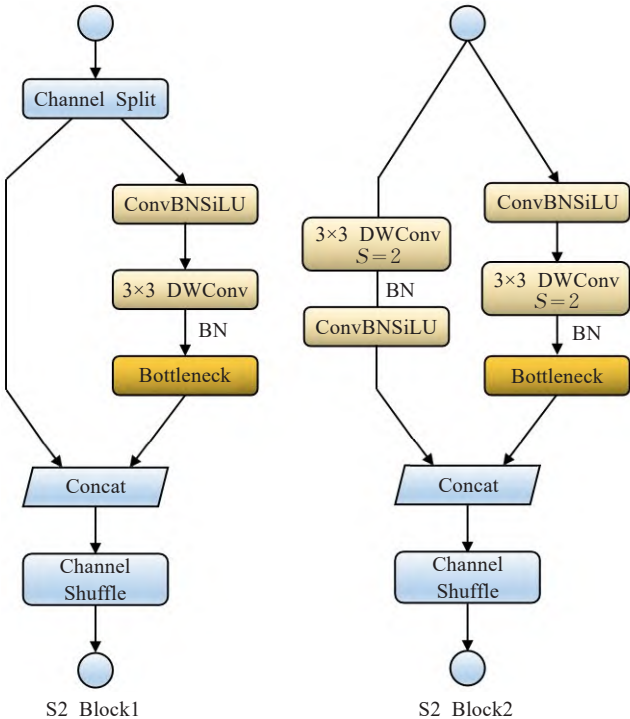


图5 S2 Block 单元模块结构

Fig.5 S2 Block unit module structure

(3)由于增加Bottleneck模块导致模型参数量相应增加,因此在不明显降低模型检测能力的情况下对ShuffleNetV2结构中的stage2、stage3、stage4堆叠结构进行缩减,并将原网络中的MaxPool层用 $3\times 3$ Conv2卷积

替换,在完成下采样的同时提取样本特征,生成复杂度不同的网络。

改进后ShuffleNetV2的总体网络结构如表1所示,对每个模块中的通道数量进行缩放,生成复杂度不同的网络。

表1 改进后ShuffleNetV2网络结构

Table1 Improved ShuffleNetV2 network structure

Layer	Output size	KSize	Stride	Repeat	Output channels
Image	640×640				1×
ConvBNSiLU	320×320	3×3	2	1	32
ConvBNSiLU	160×160	3×3	2	1	
Stage2	80×80		2	1	64
	80×80		1	1	
SimAM	80×80			1	64
Stage3	40×40		2	1	128
	40×40		1	3	
SimAM	40×40			1	128
Stage4	20×20		2	1	256
	20×20		1	1	
SimAM	20×20			1	256
Parameters					528 704
# of Weights					2.8 MB

2.4 YOLO-DoC 结构

YOLOv5的网络结构主要由Backbone和Head两部分组成,提取图像特征的功能主要在Backbone中实现,用上文改进的ShuffleNetV2网络替换YOLOv5原有的CSPDarkNet53主干网络,减少检测网络的运算参数和占用显存,提高检测速度,从而满足采摘工况实时性的需求。更改后的网络结构YOLO-DoC如图6所示。YOLO-DoC在数据输入部分进行了自适应图像填充,随机仿射变换,Mosaic图像拼接等数据增强方式,以增强检测目标的辨识度和准确度。Backbone部分为上文改进后的ShuffleNetV2网络,主要包括S2 Block模块和SimAM注意力机制模块。S2 Block模块主要用于图像的特征提取,包含目标的位置信息和细节信息,能够满足图像中心区域特征的提取需求。引入S2 Block模块的创新之处在于将ShuffleNetV2的Channel Shuffle思想引入YOLOv5网络,相对于原Backbone减少参数量的同时提升了特征提取能力。SimAM注意力机制能对不同stage模块提取的特征进行优化,以端到端的方式降低背景噪声对目标检测的干扰,从而有效地获取不同的特征信息。在Head部分,YOLO-DoC沿用YOLOv5的特征金字塔结构(feature pyramid network,FPN)+路径聚合网络结构(path aggregation network,PAN)。FPN通过对高底层特征的融合提升网络对小目标样本的检测能力。PAN通过增加自下而上的路径将底层定位信息传递到上层网络,从而实现对不同检测层进行特征聚合。该结构在不同特征层充分进行特征融合,可以

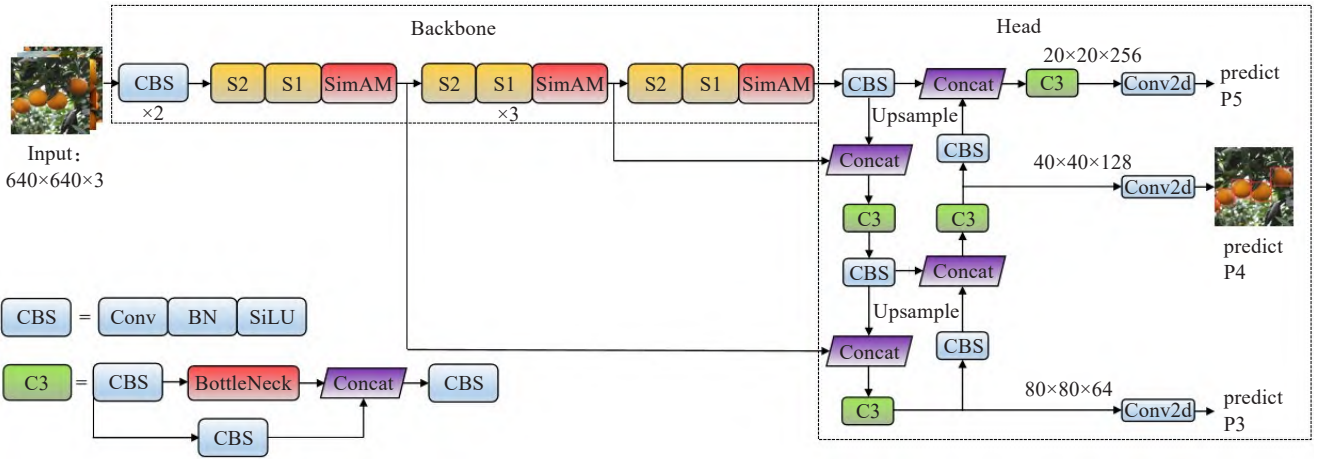


图6 YOLO-DoC 网络结构图

Fig.6 YOLO-DoC network structure diagram

增强多尺度上的定位能力。

## 2.5 alpha-IoU

在YOLOv5原网络中,边界框回归损失函数使用的是GIoU Loss,GIoU将预测框和真实标注框的外接矩形面积考虑在损失计算之中,改善了IoU只关注预测框和真实标注框的重叠区域,而不能反映出二者相交的情况。损失函数定义如式(5)和(6)所示:

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \quad (5)$$

$$Loss_{GIoU} = 1 - GIoU \quad (6)$$

其中,  $A$  和  $B$  分别表示预测框和真实标注框,  $C$  表示  $A$  和  $B$  的外接矩形。当  $A$  与  $B$  不相交时,  $A \cup B$  的值不变,  $IoU$  的值为0,只有最小化  $C$  才能满足最大化  $GIoU$ ,即最小化  $GIoU$  Loss,而  $C$  在不断变小的过程就是  $A$  和  $B$  两个框不断接近的过程。尽管  $GIoU$  引入了最小外接框最为损失函数的惩罚项,但从  $A$  与  $B$  不相交开始到相交的过程需要。此外,若出现目标框包裹预测框的情况,如图7所示,此时  $GIoU$  的值等同于  $IoU$ ,因此仅通过  $GIoU$  无法反映出预测框和真实标注框的具体位置情况。

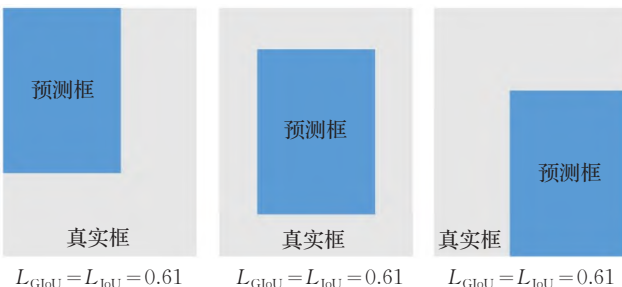


图7 预测框位于真实框内部

Fig.7 Prediction box inside ground truth box

针对  $GIoU$  使用当中所存在的问题,本文使用具有更多惩罚项的  $\alpha$ -IoU,  $\alpha$ -IoU 将  $IoU$  现有的损失统一化并以幂指数形式表示。通过引入可调节的超参

数  $\alpha$  满足不同水平的边界框回归精度,并且经过实验表明  $\alpha$ -IoU 的损失函数效果优于目前所有的  $IoU$ ,具有更高的回归精度以及较好的鲁棒性。

$\alpha$ -IoU 损失函数的定义如式(7)所示,通过对超参数  $\alpha$  进行调节可以推导出现有损失中的大多数  $IoU$ 。

$$\alpha - IoU = \frac{1 - IoU^\alpha}{\alpha}, \alpha > 0 \quad (7)$$

再通过增加惩罚项将上述  $\alpha$ -IoU 损失函数扩展到一般形式,如式(8)所示,则可以根据  $\alpha$  值对现有的基于  $IoU$  的损失进行概括。

$$Loss_{\alpha - IoU} = 1 - IoU^{\alpha_1} + \rho^{\alpha_2}(B, B^{gt}) \quad (8)$$

依据上述  $\alpha$ -IoU 的一般形式可以归纳出常用的  $IoU$  损失,本文选择  $\alpha$ -CIoU,具体如式(9)所示:

$$Loss_{\alpha - CIoU} = 1 - IoU^\alpha + \frac{\rho^{2\alpha}(B, B^{gt})}{c^{2\alpha}} + (\beta\nu)^\alpha \quad (9)$$

式中,  $\rho$  为预测框和真实标注框的欧氏距离,  $c$  代表的是能够同时包含预测框和真实框的最小闭包区域的对角线距离,  $\nu$  用来衡量长宽比的一致性,  $\beta$  为用于平衡  $\nu$  的权衡参数。具体表达式如式(10)和(11)所示:

$$\beta = \frac{\nu}{(1 - IoU) + \nu} \quad (10)$$

$$\nu = \frac{4}{\pi^2} \left( \arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)^2 \quad (11)$$

式中,  $\frac{\omega^{gt}}{h^{gt}}$  和  $\frac{\omega}{h}$  分别为真实标注框和预测框的宽高比。

## 3 实验结果分析

### 3.1 实验环境与实验数据处理

本实验基于 Ubuntu20.04 操作系统, CPU 为 Intel® Core i7-9700 8核, GPU 为 Nivida A10 图形显卡,深度学习框架为 pytorch1.12 版本, CUDA 版本为 11.4。

本文实验流程如图8所示。首先将正常光源下采集的有效数据集通过对比度增强及亮度增强方式模拟



不同自然界光照情况,增加样本的多样性。之后通过数据增强对样本数量进行扩增,筛选出合适样本进行数据标注,并以xml格式保存标注文件。再对样本进行划分,分别放入三种模型进行训练。训练的过程采用迁移学习的方法,即使用MS COCO训练集的预训练模型对模型进行部分层参数的训练,目的是减少到达训练最优解的迭代次数,更快地达成模型收敛。

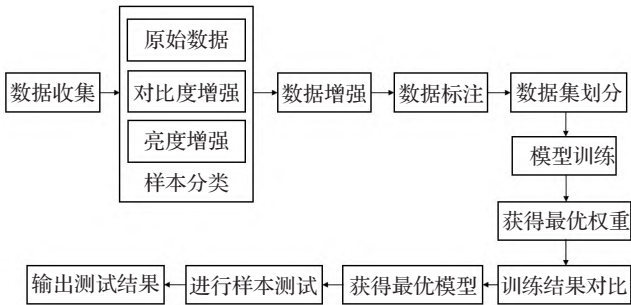


图8 柑橘检测模型实验流程图

模型训练时设定单批次的样本数量为32,初始学习率为0.001,动量因子为0.937,输入图像的尺寸为640×480,共训练100轮,将每轮训练后的结果进行保存,用于绘制模型收敛趋势对比图。

3.2 模型训练

本研究用每张图片检测时间作为模型的推理速度,用准确率P(precision),平均精度AP(average precision),召回率R(recall)作为模型的检测效果评价指标,同时用预测框回归损失函数(Box\_Loss)作为IoU预测效果的评价指标。具体的计算公式如式(12)(13)(14)所示:

Precision = T\_P / (T\_P + F\_P) (12)

Recall = T\_P / (T\_P + F\_N) (13)

AP = ∫\_0^1 P(R) dR (14)

式中,  $T_P$  表示模型预测正样本为真的数量,  $F_P$  表示模型预测负样本为真的数量,  $F_N$  表示模型预测正样本为假的数量。

将改进模型YOLO-DoC中的alphaIoU损失函数分别使用如式(6)所示GIoU损失函数,未加入alpha因子的CIoU损失函数替换,与改进的损失函数进行对比训练。并以训练结果中的检测框损失值box\_loss为纵坐标绘制不同IoU对模型效果损失的对比曲线,如图9所示。图中YOLO-DoC-CIoU、YOLO-DoC-GIoU和YOLO-DoC为YOLO-DoC网络分别以CIoU、GIoU以及alphaIoU为检测框损失评价指标绘制的损失图。完成100轮迭代之后三个模型都基本趋于稳定,其中alphaIoU与CIoU的收敛趋势较为相似,在迭代60轮左右时趋近平稳。GIoU的边界框回归损失值稍显动荡,大概迭代70轮时趋于平稳。由此可见,对于YOLO-DoC模型,采

用alphaIoU损失函数会有效降低边界框损失值。对检测模型的定位能力起到一定的优化作用。

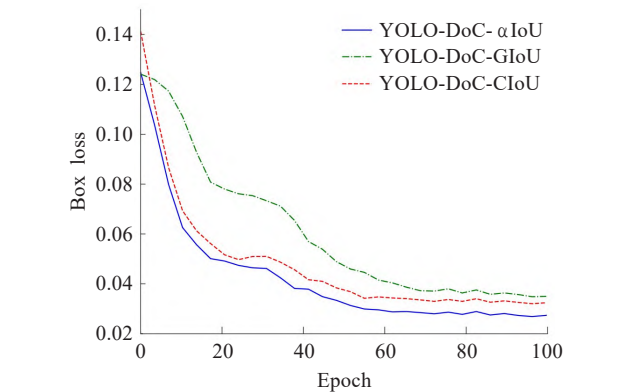


图9 不同IoU的预测框回归损失对比

Fig.9 Comparison of prediction box loss for different IoU

3.3 结果与分析

为验证YOLO-DoC模型的有效性,以下对各个改进模块进行对比分析。主要包括SimAM与其他空间和通道注意力机制的对比实验,针对YOLO-DoC模型的改进做消融实验和改进前后的对比实验,使用基于YOLO模型的不同骨干网络的轻量化模型对比,同时考虑到自然环境下的光照条件的多样性,比较模型对不同亮度与对比度的数据集的检测效果。

3.3.1 不同注意力机制实验对比

为验证SimAM注意力机制在YOLO-DoC柑橘采摘模型中的高效性和轻量性,本小节将SimAM注意力机制与SE、ECA、CBAM注意力机制进行对比实验。实验模型为YOLO-DoC网络结构,将主干网络中stage2、stage3、stage4模块后添加注意力机制模块进行测试,实验结果如表2所示。由表2可知添加SimAM注意力机制的网络模型在各项精度和参数量的综合衡量下优于其他注意力机制,且由于其无参性的特点并不会增加模型的参数量。因此在模型中添加SimAM注意力机制。

表2 不同注意力机制对比测试

Fig.2 Comparison test of different attention mechanisms				
Attention Mechanism	Precision/%	mAP@0.5/%	mAP@0.5:0.95/%	Parameters
SE	97.2	98.5	88.2	10 752
ECA	97.5	98.7	87.7	9
CBAM	98.1	99.1	88.9	11 046
SimAM	98.8	99.4	88.6	0

3.3.2 YOLO-DoC模型消融实验

为验证本文所进行的改进实验对检测效果有一定的提升作用,以基于ShuffleNetV2为骨干网络的YOLOv5网络作为对比实验模型,分别将SimAM、Bottleneck和alphaIoU作为验证模块进行消融实验。训练结果如表3所示。可以看出,分别对模型添加SimAM注意力机制, Bottleneck模块和alphaIoU都会提高模型检测效

表3 YOLO-DoC 消融实验的检测结果  
Table 3 Detection results of YOLO-DoC ablation experiment

Model	Modules			Precision/%	mAP@0.5(±)	Box Loss
	SimAM	Bottleneck	Alpha-IoU			
YOLOv5-ShuffleNetV2	—	—	—	96.7	0	0.043
	✓	—	—	97.5	+0.4	0.041
	—	✓	—	97.9	+1.3	0.039
	—	—	✓	96.8	0	0.025
	✓	✓	—	98.2	+0.9	0.033
	—	✓	✓	98.3	+0.7	0.028
	✓	—	✓	97.3	+1.2	0.027
YOLO-DoC	✓	✓	✓	98.8	+1.9	0.019

果。其中,引入SimAM注意力机制和Bottleneck结构会一定程度提高模型的检测精度,而引入Alpha-IoU会明显降低预测框损失。不同的组合都会对模型的检测效果有一定的提升,但YOLO-DoC模型的训练结果中准确率为98.8%,平均精度为相比于改进前YOLOv5-ShuffleNetV2模型提高1.9个百分点,Box Loss为0.019,均为最优结果。可以看出YOLO-DoC模型对于柑橘的检测定位更加准确,提高了采摘过程的高效性。

3.3.3 使用不同骨干网络训练

为证明YOLO-DoC网络在不失准度且快速检测的突出优势,本小节将基于YOLOv5网络的不同骨干网络模型进行对比,测试模型包括以DarkNet53为骨干网络的YOLOv5s原模型,以MobileNetV3为骨干网络的YOLOv5模型,以ShuffleNetV2为骨干网络的YOLOv5模型,即为改进前的模型,以及改进后的YOLO-DoC模型。训练过程如图10和图11的曲线图所示。当模型训练趋于稳定时,YOLOv5s和YOLO-DoC模型的mAP@0.5和mAP@0.5:0.95平均精度最高,其次是YOLOv5-ShuffleNetV2,相对最低的为YOLOv5-MobileNetV3。

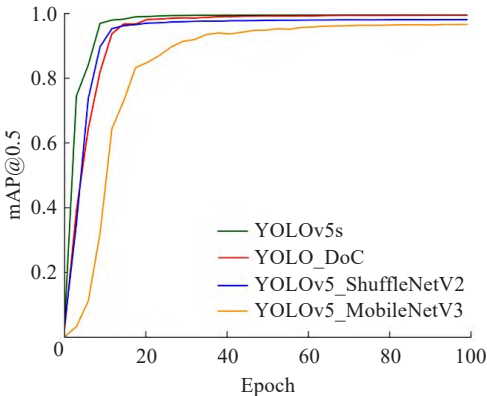


图10 不同主干网络模型mAP@0.5精度对比

Fig.10 Comparison of mAP@0.5 of different backbone network models

最终的训练结论如表4所示。YOLOv5s相较于其他骨干网络模型最复杂,训练权重最大,相应的检测速率也最慢,但总体的训练效果更优异。使用ShuffleNetV2作为骨干网络的YOLOv5模型在检测效

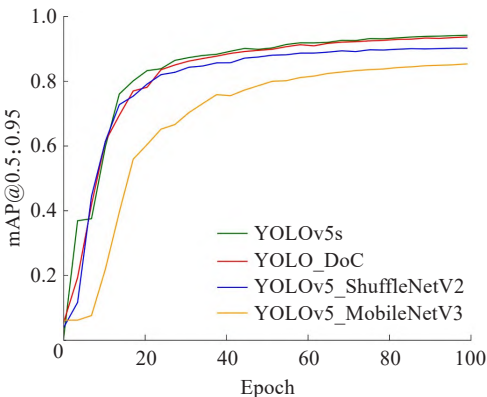


图11 不同主干网络模型mAP@0.5:0.95精度对比

Fig.11 Comparison of mAP@0.5:0.95 of different backbone network models

果上略低于YOLOv5原模型,准确率相差2.3个百分点,平均精度相差1.6个百分点,训练速度提升39 FPS。MobileNetV3作为骨干网络的YOLOv5模型在准确率测试结果略低于ShuffleNetV2模型。而YOLO-DoC模型与YOLOv5原骨干网络模型的检测效果接近,权重大小为所有模型中最小,在GPU上的检测速度可以达到187 FPS,在CPU上的检测速度为32 FPS。YOLO-DoC模型在检测精度和准确率满足采摘作业需求的前提下减小了模型大小,大幅提升了模型检测速度,并且在移动终端上也可以满足实时检测的需求。

表4 不同骨干网络的检测结果

Table 4 Detection results of different backbone networks

Models	Precision/%	Average Precision/%	Weight Size/MB	Detect Speed/FPS
YOLOv5	98.9	99.3	15.8	96
YOLOv5-MobileNetV3	94.8	94.5	10.3	116
YOLOv5-ShuffleNetV2	96.6	97.7	8.6	135
YOLO-DoC	98.8	99.4	2.8	187

3.3.4 不同先进检测算法对比实验

为证明YOLO-DoC检测网络对于柑橘果实检测的优越性,本小节将YOLO-DoC网络与其他常用检测网络以及近年先进检测网络在自建柑橘图像数据集上进

行对比测试。测试结果如表5所示。为了直观展示先进检测网络的特点和创新,将原论文中添加和改进的模块加在模型名后。其中,M2和M3分别表示MobileNetV2和MoblieNetV3网络结构。

由表中结论可知,常用检测网络FasterRCNN-ResNet50由于是以ResNet50为Backbone的两阶段网络,整体模型权重大,检测速度慢,且检测精度相较于YOLO-DoC也稍有逊色。而单阶段网络SSD-300和YOLOv4网络对于柑橘果实的检测率偏低,模型权重偏大,实验结果均劣于其余检测模型。例如与YOLOv4网络相比,模型缩减254.4 MB且检测精度提升了9.1个百分点。

对于先进网络,以YOLOv3为框架,MobileNetV2为主干网络且引入GIoU回归损失函数的YOLOv3-M2-GIoU网络模型虽大幅缩减模型大小并提升检测速度,但在柑橘数据集上的平均精度仍低于YOLO-DoC平均精度7.2个百分点;在YOLOv5s网络中引入CBAM注意力机制和AlphaIoU的YOLOv5s-CBAM-AlphaIoU网络模型虽然在检测精度接近YOLO-DoC,但模型大小和检测速度均有一定劣势;进一步改进的两个网络均以轻量化网络代替YOLOv5s的主干网络。分别为以MobileNetV3为Backbone且引入GIoU和CBAM的YOLOv5s-M3-GIoU-CBAM网络以及以EfficientNet为Backbone的YOLOv5s-EfficientNet网络。由实验结论可以看出,本文算法与该两者网络相比,在模型大小缩减到1/8到1/9的同时,检测精度仍有0.4到1.2个百分点的提升。因此可以验证YOLO-DoC网络在柑橘检测方面效果最优越。

3.3.5 不同遮挡条件下的模型检测结果

考虑到不同程度的样本遮挡情况对检测结果的影响,本节将对遮挡程度划分好的数据集A、数据集B和

数据集A+B分别进行准确率和平均精度的测试,测试样张如图12所示。其中图(a)和(b)为改进前后模型对重度遮挡样本检测的对比示例图,部分样本受到大量枝叶枝干遮挡或果实重叠情况,可以看到,YOLO-DoC模型对于此类型样本的检测精度为98.9%,目标损失值为0.037。图(c)和(d)为改进前后模型对轻度遮挡样本检测的对比示例图,部分样本或多或少收到部分枝叶遮挡,但遮挡情况不严重,YOLO-DoC模型对于此类型样本的检测精度为97.5%,目标损失值为0.056。综合测试结果如表6所示。可以看到,YOLO-DoC对不同遮挡程度的数据集,相比于改进前的网络模型检测效果均有一定程度的提升。而重度遮挡样本由于被遮挡样本特征不完整、距离远像素模糊从而人工标注时对不明显样本的忽略,以及神经网络在进行卷积操作时对小目标和边缘信息造成丢失等原因,造成准确率和精度稍有下降,目标损失值稍有增大。但考虑实际采摘时,机械臂会首先对非遮挡以及目标识别率高的样本进行抓取,因此上述损失并不会对实际采摘工作的效率造成影响。

4 结语

本文针对柑橘检测项目,设计了一套YOLO-DoC轻型检测网络用于柑橘检测环节。将引入Bottleneck结构和SimAM注意力机制的ShuffleNetV2作为YOLOv5的骨干网络,模型的权重大小降至2.8 MB,模型的检测速度提升,在GPU上可以达到187 FPS,在CPU上为32 FPS。在模型做到轻量化的同时准确率为98.7%,mAP@0.5为99.1%,检测效果较好,因此该模型适用于嵌入式移动端设备。用Alpha-IoU代替YOLOv5的GIoU降低了损失函数的边界框回归损失,提高了对于样本的边界定位,有利于抓取工作的进行。为了验证该模型的优越性,将模型进行消融实验和与其他先进检测

表5 不同先进检测网络对比  
Table 5 Comparison of different advanced detection networks

Model	Precision/%	Average Precision/%	Weight Size/MB	Detect Speed/FPS
FasterRCNN-ResNet50	96.6	97.2	196.5	5
SSD-300	93.7	94.1	212.8	19
YOLOv4	89.6	90.3	257.2	11
YOLOv5s-CBAM-AlphaIoU <sup>[11]</sup>	98.5	99.1	17.8	96
YOLOv3-M2-GIoU <sup>[13]</sup>	91.4	92.2	21.0	91
YOLOv5s-M3-GIoU-CBAM <sup>[24]</sup>	97.5	98.2	12.6	112
YOLOv5s-EfficientNet <sup>[25]</sup>	98.2	99.0	22.3	108
YOLO-DoC	98.8	99.4	2.8	187

表6 不同遮挡程度数据检测结果  
Table 6 Detection results of different occlusion degrees

Model	Precision/%			Object Loss		
	A	B	A+B	A	B	A+B
改进前	98.1	95.7	97.3	0.053	0.072	0.059
YOLO-DoC	98.9	97.5	98.3	0.037	0.056	0.041





图12 不同遮挡程度果实检测结果图

Fig.12 Fruit detection results with different degrees of occlusion

网络进行比较,准确率和检测速度均有一定优势。

本文设计的轻量化检测网络使采摘工作在满足高精度检测的同时对于目标检测速度有明显提升,节省了采摘时间提高了采摘工作的效率,果实的定位误差在不影响采摘工作的前提下达到可接受的范围内,采摘成功率高,能够满足采摘工作的要求。在今后研究中可以对环境复杂干扰因素多的不同柑橘果实进行检测算法的建立,提高模型对样本的检测能力。

参考文献:

[1] 柴光荣,王晨洋,等.丘陵山地果园运输机械的发展现状与展望[J].农业技术与装备,2019(3):30-31.  
CHAI Guangrong, WANG Chenyang, et al.Development status and prospects of orchard transportation machinery in hilly and mountainous areas[J].Agricultural Technology and Equipment,2019(3):30-31.

[2] 叶敏,邹湘军,罗陆锋,等.荔枝采摘机器人双目视觉的动态定位误差分析[J].农业工程学报,2016,32(5):50-56.  
YE Min,ZOU Xiangjun,LUO Lufeng,et al.Error analysis of dynamic localization tests based on binocular stereo vision on litchi harvesting manipulator[J].Transactions of the Chinese Society of Agricultural Engineering,2016, 32(5):50-56.

[3] 杨长辉,刘艳平,王毅,等.自然环境下柑橘采摘机器人识别定位系统研究[J].农业机械学报,2019,50(12):14-22.  
YANG Changhui,LIU Yanping,WANG Yi,et al.Research and experiment on recognition and location system for citrus picking robot in natural environment[J].Transactions of the Chinese Society for Agricultural Machinery, 2019,50(12):14-22.

[4] XIONG J,LIN R,LIU Z,et al.The recognition of litchi clusters and the calculation of picking point in a nocturnal natural environment[J].Biosystems Engineering,2018,166:

- 44-57.
- [5] ZHUANG J J, LUO S M, HOU C J, et al. Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications[J]. *Computers and Electronics in Agriculture*, 2018, 152: 64-73.
- [6] 冯娟, 刘刚, 司永胜, 等. 苹果采摘机器人激光视觉系统的构建[J]. *农业工程学报*, 2013, 29(25): 32-37.
- FENG Juan, LIU Gang, SI Yongsheng, et al. Construction of laser vision system for apple harvesting robot[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(25): 32-37.
- [7] BAC C W, HEMMING J, VAN HENTEN E J. Stem localization of sweet-pepper plants using the support wire as a visual cue[J]. *Computers and Electronics in Agriculture*, 2014, 105: 111-120.
- [8] HUANG Z, WANE S, PARSONS S. Towards automated strawberry harvesting: identifying the picking point[C]// *Annual Conference Towards Autonomous Robotic Systems*, 2017.
- [9] SENGUPTA S, LEE W S. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions[J]. *Bio-systems Engineering*, 2014, 117(1): 51-61.
- [10] HE Z L, XIONG J T, LIN R, et al. A method of green litchi recognition in natural environment based on improved LDA classifier[J]. *Computers & Electronics in Agriculture*, 2017, 140: 159-167.
- [11] 黄彤斌, 黄河清, 李震, 等. 基于YOLOv5改进模型的柑橘果实识别方法[J]. *华中农业大学学报*, 2022, 41(4): 170-177.
- HUANG Tongbin, HUANG Heqing, LI Zhen, et al. Citrus fruit recognition method based on improved model of YOLOv5[J]. *Journal of Huazhong Agricultural University*, 2022, 41(4): 170-177.
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// *Proceedings of the 2018 European Conference on Computer Vision: Lecture Notes in Computer Science*. Cham: Springer, 2018: 3-19.
- [13] 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.
- LV 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*, 2019, 35(17): 205-214.
- [14] 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.
- [15] 冯硕, 李旗. 奇异果采摘机器人机械臂控制系统研究——基于计算机视觉和卷积神经网络[J]. *农机化研究*, 2022, 44(10): 25-29.
- FENG Shuo, LI Qi. Research on the control system of the manipulator of kiwi fruit picking robot-based on computer vision and convolutional neural network[J]. *Journal of Agricultural Mechanization Research*, 2022, 44(10): 25-29.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. *arXiv:1512.03385*, 2015.
- [17] 陈锋军, 张新伟, 朱学岩, 等. 基于改进EfficientDet的油橄榄果实成熟度检测[J]. *农业工程学报*, 2022, 38(13): 158-166.
- CHEN Fengjun, ZHANG Xinwei, ZHU Xueyan, et al. Detection of the olive fruit maturity based on improved EfficientDet[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(13): 158-166.
- [18] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]// *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020.
- [19] 张皓婷, 李明, 李敏, 等. 基于Faster R-CNN的番茄果实检测算法应用[J]. *农业与技术*, 2022, 42(15): 12-16.
- ZHANG Haoting, LI Ming, LI Min, et al. Application of tomato fruit detection algorithm based on Faster R-CNN[J]. *Agriculture and Technology*, 2022, 42(15): 12-16.
- [20] 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.
- [21] MA N, ZHANG X, ZHANG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]// *European Conference on Computer Vision (ECCV)*. Cham: Springer, 2018.
- [22] YANG Lingxiao, ZHANG Ruyang, LI Lida, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]// *Proceedings of the 38th International Conference on Machine Learning*, 2021.
- [23] HE J B, ERFANI S, MA X J, et al. Alpha-IOU: a family of power intersection over union losses for bounding box regression[J]. *arXiv:2110.13675*, 2021.
- [24] 杨永波, 李栋. 改进YOLOv5的轻量级安全帽佩戴检测算法[J]. *计算机工程与应用*, 2022, 58(9): 201-207.
- YANG Yongbo, LI Dong. Lightweight helmet wearing detection algorithm of improved YOLOv5[J]. *Computer Engineering and Applications*, 2022, 58(9): 201-207.
- [25] LIU C, ZHU H J, GUO W, et al. EFDet: an efficient detection method for cucumber disease under natural complex environments[J]. *Computers and Electronics in Agriculture*, 2021, 189: 1063978.