

# 基于改进 YOLO v7 - tiny 的甜椒畸形果识别算法

王昱<sup>1</sup> 姚兴智<sup>1</sup> 李斌<sup>2,3</sup> 徐赛<sup>2,3</sup> 易振峰<sup>1</sup> 赵俊宏<sup>2,3</sup>

(1. 华南农业大学工程学院, 广州 510642; 2. 岭南现代农业科学与技术广东省实验室, 广州 510600;

3. 广东省农业科学院设施农业研究所, 广州 510640)

**摘要:** 甜椒在生长发育过程中容易产生畸形果, 机器代替人工对甜椒畸形果识别和摘除一方面可提高甜椒品质和产量, 另一方面可解决当前人工成本过高、效率低下等问题。为实现机器人对甜椒果实的识别, 提出了一种基于改进 YOLO v7 - tiny 目标检测模型, 用于区分正常生长和畸形生长的甜椒果实。将无参数注意力机制 (Parameter-free attention module, SimAM) 融合到骨干特征提取网络中, 增强模型的特征提取和特征整合能力; 用 Focal - EIOU (Focal and efficient intersection over union) 损失替换原损失函数 CIOU (Complete intersection over union), 加快模型收敛并降低损失值; 使用 SiLU 激活函数代替原网络中的 Leaky ReLU, 增强模型的非线性特征提取能力。试验结果表明, 改进后的模型整体识别精确度、召回率、平均精度均值 (Mean average precision, mAP) mAP<sub>0.5</sub>、mAP<sub>0.5-0.95</sub> 分别为 99.1%、97.8%、98.9%、94.5%, 与改进前相比, 分别提升 5.4、4.7、2.4、10.7 个百分点, 模型内存占用量为 10.6 MB, 单幅图像检测时间为 4.2 ms。与 YOLO v7、Scaled - YOLO v4、YOLOR - CSP 等目标检测模型相比, 模型在 F1 值上与 YOLO v7 相同, 相比 Scaled - YOLO v4、YOLOR - CSP 分别提升 0.7、0.2 个百分点, 在 mAP<sub>0.5-0.95</sub> 上分别提升 0.6、1.2、0.2 个百分点, 而内存占用量仅为上述模型的 14.2%、10.0%、10.0%。本文所提出的模型实现了小体量而高精度, 便于在移动端进行部署, 为后续机械化采摘和品质分级提供技术支持。

**关键词:** 甜椒畸形果; YOLO v7 - tiny; 目标检测; 机器视觉

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2023)11-0236-11

OSID:



## Malformed Sweet Pepper Fruit Identification Algorithm Based on Improved YOLO v7 - tiny

WANG Yu<sup>1</sup> YAO Xingzhi<sup>1</sup> LI Bin<sup>2,3</sup> XU Sai<sup>2,3</sup> YI Zhenfeng<sup>1</sup> ZHAO Junhong<sup>2,3</sup>

(1. College of Engineering, South China Agricultural University, Guangzhou 510642, China

2. Guangdong Laboratory of Lingnan Modern Agricultural Science and Technology, Guangzhou 510600, China

3. Institute of Facility Agriculture, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China)

**Abstract:** Sweet peppers are prone to malformed fruits during the growth and development process. Machine replace manual identification and removal of deformed sweet peppers, on the one hand, it can improve the quality and yield of sweet peppers; on the other hand, it can solve the current problems of high labor costs and low efficiency. In order to realize the identification of sweet pepper fruits by robots, an improved YOLO v7 - tiny target detection model was proposed to distinguish between normal and abnormal growth of sweet pepper fruits. The parameter-free attention module (SimAM) was integrated into the backbone feature extraction network to enhance the feature extraction and feature integration capabilities of the model; the original loss function CIOU was replaced with Focal - EIOU loss, Focal - EIOU can speed up model convergence and reduce loss value; the SiLU activation function was used to replace the Leaky ReLU in the original network to enhance the nonlinear feature extraction ability of the model. The test results showed that the overall recognition precision, recall rate, mAP<sub>0.5</sub> and mAP<sub>0.5-0.95</sub> of the improved model were 99.1%, 97.8%, 98.9% and 94.5%, compared with that

收稿日期: 2023-07-31 修回日期: 2023-08-20

**基金项目:** 岭南现代农业科学与技术广东省实验室科研项目 (NT2021009)、国家自然科学基金面上项目 (32372002)、广东省农业科学院科技人才引进专项资金项目 (R2019YJ-YB3003)、广东省农业科学院协同创新中心项目 (XT202201) 和广东省重点领域研发计划项目 (2023B0202090001)

**作者简介:** 王昱 (1987—), 女, 副教授, 博士, 主要从事结构优化设计方法与智能农机装备研究, E-mail: yu-wang@scau.edu.cn

**通信作者:** 赵俊宏 (1989—), 男, 助理研究员, 博士, 主要从事农业信息化方法与智能农业装备研究, E-mail: junhongzhao@126.com

before improvement, it was increased by 5.4 percentage points, 4.7 percentage points, 2.4 percentage points, and 10.7 percentage points, respectively, the model weight size was 10.6 MB, and the single image detection time was 4.2 ms. Compared with YOLO v7, scaled - YOLO v4, YOLOR - CSP target detection models, the model had the same F1 score as YOLO v7. Compared with scaled - YOLO v4, YOLOR - CSP was increased by 0.7 and 0.2 percentage points, respectively, mAP0.5 - 0.95 was increased by 0.6 percentage points, 1.2 percentage points and 0.2 percentage points, respectively, and the weight size was only 14.2%, 10.0%, 10.0% of the above model. The model proposed achieved small size and high precision, and it was easy to deploy on the mobile terminal, providing technical support for subsequent mechanized picking and quality grading.

**Key words:** malformed sweet pepper; YOLO v7 - tiny; target detection; machine vision

## 0 引言

甜椒具有很高的营养价值和经济价值,在我国广泛种植<sup>[1]</sup>。但在生长发育阶段,甜椒非常容易出现畸形果,甜椒畸形是果实生长异常的一种生理疾病,主要是因为花芽分化期间遇到极端天气从而影响甜椒花芽分化,导致花粉受精不良,花器发育不完全,主要表现为果实生长不正常,具体表现为果实呈扁圆形、表面凹陷、顶部增生、底部尖化及其他不规则形状。一方面,外观美感缺失会导致经济价值的降低,另一方面,果实之间养分的争夺影响正常甜椒果实的生长从而影响甜椒的产量。因此在甜椒的生长发育阶段需要对果实适时检查,并对畸形果进行及时修剪去除,使用人工进行生长监测会耗费大量的人力和时间,且当前人工成本较高,无法全天候掌握果实生长发育状况,对果实的生长发育状态也无法及时地汇总并处理。为更好地解决上述问题,巡检机器人成为解决问题的有效方案,而要更好地实现巡检机器人监测甜椒畸形果,需要对甜椒畸形果进行准确地识别。近年来,随着计算机视觉技术的发展,世界各地的研究人员针对水果和蔬菜的识别主要可以分为基于传统图像处理和基于深度学习的目标检测<sup>[2]</sup>。

传统图像处理的目标检测方法,主要是对数字图像中被测物体的形状、颜色、轮廓等特征,通过分割<sup>[3]</sup>、聚类、增强、形态学等操作进行合理地特征提取,进而实现目标检测<sup>[4-7]</sup>。对于一些水果或蔬菜,因其具有特定的颜色特征,且在不同的颜色空间中有不同的特征,可通过对不同颜色空间中的特征进行聚类,实现对目标的检测与分割<sup>[8]</sup>。颜色的产生是由于果蔬表面对不同波长的吸收与反射,因此可以对不同波长进行阈值分割,分离出对应波长的颜色,再使用分类器进行分类,最终实现对目标的识别与检测<sup>[9-10]</sup>。然而,传统图像处理只分析了物体的表面特征(形状、颜色等),无法对深层特征进行很好地提取,难以满足实际复杂环境中水果和蔬菜的

识别。

与传统图像处理方法相比,深度学习方法使用深度神经网络,可以自主学习图像中的各种特征,并进行特征融合,被广泛用于农业生产环境下的果蔬识别,如苹果、芒果、葡萄、番茄等。深度神经网络主要可以分为两种类型,一种是注重精度的两阶段目标检测网络,如 Faster R - CNN<sup>[11]</sup>, Mask R - CNN<sup>[12-13]</sup>,两阶段网络需要在图像上生成许多的候选框,并根据非极大值抑制过滤多余的候选框,两阶段目标检测网络虽然在检测精度上可以获得很好的效果,但是由于需要更多的浮点运算,导致其模型规模较大和检测速度较慢。另一种是注重检测速度的单阶段目标检测网络,如 YOLO v4<sup>[14-15]</sup>、YOLO v5<sup>[16-17]</sup>、YOLO v7<sup>[18]</sup>、SSD<sup>[19]</sup>等,与两阶段目标检测网络不同,单阶段目标检测网络可直接在图像生成预测框,大幅提升了检测速度。在农业场景中,由于自然环境下的果蔬识别往往拥有复杂的背景,容易对识别造成干扰,因此,使用注意力机制(如通道注意力(Channel attention, CA)<sup>[14,17]</sup>、空间注意力(Spatial attention, SA)<sup>[20]</sup>、卷积注意力(Convolutional block attention module, CBAM)<sup>[21]</sup>)来增强网络的特征提取能力成为了许多识别模型的解决方案。在甜椒果实识别方面,文献[12]采用 Mask R - CNN 对绿色甜椒的果实和果蒂进行检测和分割。文献[13]将 Swin - Transformer 融合到 Mask R - CNN 主干网络中,并使用 UNet3 + 来改善掩模头部和掩模的分割质量,用于对甜椒的实例分割。文献[22]提出了一种识别甜椒并规划采摘序列的算法 AYDY,并将卷积块注意力模块嵌入到 YOLO v4 网络中用于识别和定位甜椒。文献[11]通过融合 RGB 和近红外(Near - infrared, NIR)的多模态信息,并使用 Faster R - CNN 对甜椒进行识别。

综上,当前在甜椒识别领域,针对甜椒果实的识别已经取得了较好的结果,但仍存在以下待解决的问题:当前未有对正常果与畸形果的判断研究;畸形果与正常果实的特征相近,极易造成误检;当前针对

甜椒识别的模型体量较大,对硬件要求高,很难部署到移动机器人上。因此,本文以甜椒畸形果实和正常果实为识别对象建立甜椒果实数据集,基于参数量较少的 YOLO v7-tiny 目标检测模型进行改进,通过替换不同的激活函数和损失函数,融合 SimAM 注意力模块,在不增加参数量情况下提高模型的特征提取能力,实现快速且准确地识别甜椒畸形果与正常果,为甜椒品质分级及采摘机器人识别与定位提供参考。

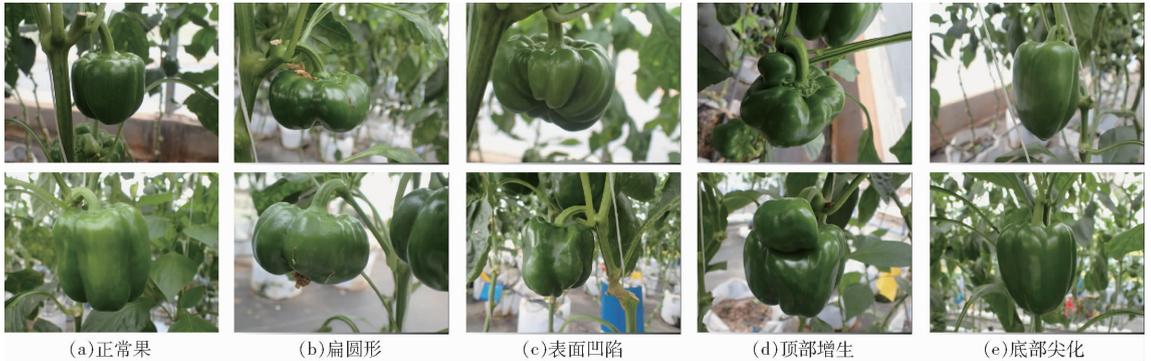


图1 正常果及畸形果

Fig. 1 Normal fruit and malformed fruit

## 1.2 数据采集

本研究所使用的甜椒品种为鲁斯卡,分别于2022年4月2日和2022年5月9日在广东省农业科学院设施农业研究所实验室环境下拍摄被采摘下来的畸形果图像240幅(每幅图像内仅包含1个果实),采集设备为 FLIR 工业摄像头(分辨率为  $2\,448$  像素  $\times$   $2\,048$  像素,像素尺寸  $2.74\ \mu\text{m}$ ,最高支持 12 bit 色深),和广东省农业科学院白云示范基地甜椒种植大棚内采集设施环境下的甜椒图像1 005幅(包含畸形果实和正常果实),采集设备为便于携带的奥林巴斯 TG-4 型数码相机(分辨率为  $3\,200$  像素  $\times$   $2\,400$  像素,焦距  $4.5 \sim 18\ \text{mm}$ ,快门速度  $1/2 \sim 1/2\,000$ ,感光度  $\text{ISO}100 \sim 6400$ )。为了使所拍摄的图像更好地表征果实形状轮廓特点以更好地区分正常果和畸形果,模拟机器人在实际场景中的工作视角,拍摄角度设为果实侧面俯仰  $30^\circ \sim 45^\circ$ ,拍摄距离为  $30 \sim 50\ \text{cm}$ ,一方面,在这个范围内所拍摄的甜椒果实图像可以将甜椒整体形状轮廓表征下来,另一方面,在数据标注时可以减少人工误判,机器人在大棚中工作视角及相机拍摄示意图如



图3 数据增强方法及效果

Fig. 3 Data augmentation method and effect

## 1 试验数据获取

### 1.1 畸形果判断

在甜椒生长发育阶段,容易受环境影响导致畸形果的产生。正常果外观如图 1a 所示,其表面光滑平整,底部收缩,整体饱满,宽高比例约为  $1:1 \sim 1:2$ ;与正常果相比,畸形果形状主要可以归类为扁圆形(图 1b)、表面凹陷(图 1c)、顶部增生(图 1d)及底部尖化(图 1e)。

图 2 所示,经过数据筛选,最终图像总数为 982 幅。

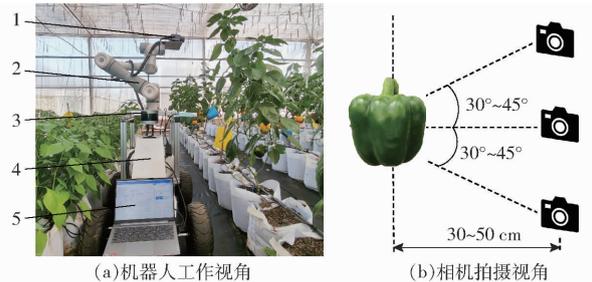


图2 机器人视角及相机拍摄视角和距离

Fig. 2 Robot perspective together with camera shooting perspective and distance

1. 多功能相机 2. 六轴机械臂 3. 激光雷达 4. 移动底盘 5. 计算机

### 1.3 数据增强

为保证模型的训练准确,提高模型的鲁棒性,加强模型对不同环境下甜椒果实的识别能力,避免因数据量不足导致模型过拟合,对上述图像数据分别进行亮度提高和降低、添加高斯噪声以及水平翻转等操作模拟不同光照环境和噪声环境,增强效果如图 3 所示,增强后数据量为 4 910 幅。

## 1.4 图像标注

将采集的图像数据导入到百度飞桨 EasyDL 中, 使用 EasyDL 手动标注一部分图像后使用智能标注功能将剩余图像进行自动标注, 自动标注完成后使用图像标注软件 Labellmg 对标注后的甜椒图像进行检查, 对自动标注有误的结果进行修改, 生成对应 YOLO 数据格式的 xml 文件, 文件中包括图像名称、图像中标签数量、标签名及各标签对应真实值 (Ground truth) 的  $x$ 、 $y$  坐标, 其中标签名分为两类, 即正常果 (Normal) 和畸形果 (Malformed)。并将标注好的标签文件按照 7:2:1 的比例划分为训练集 (3 437 幅)、验证集 (982 幅) 和测试集 (491 幅), 根据 PASCAL VOC 数据集格式完成本研究数据集的构建。

## 2 甜椒畸形果识别方法

### 2.1 YOLO v7-tiny 甜椒畸形果识别

YOLO v7<sup>[23]</sup> 是 YOLO v4<sup>[24]</sup> 作者于 2022 年发布

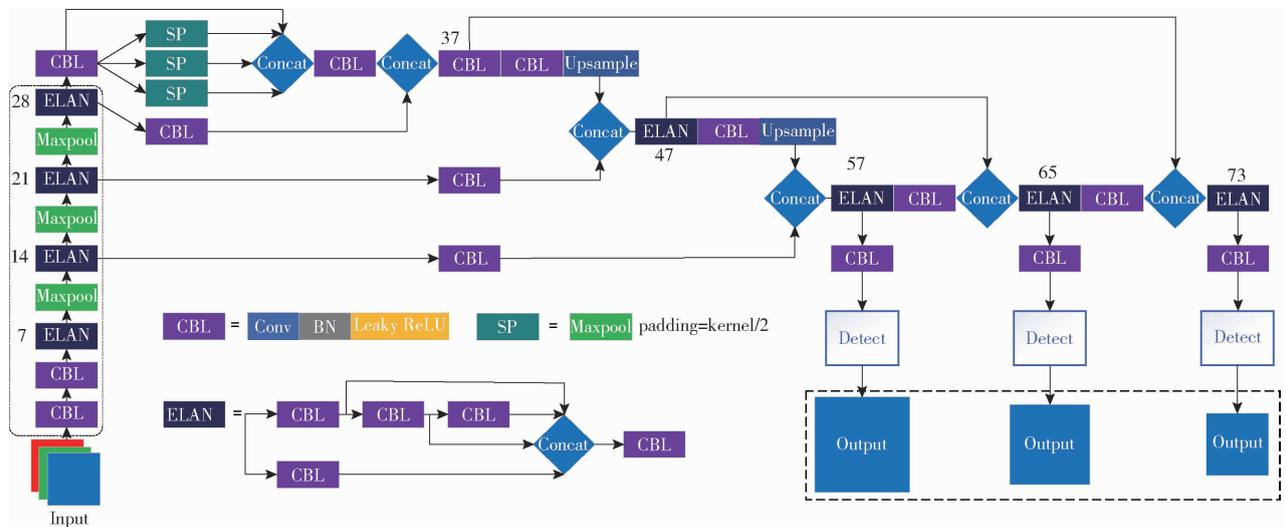


图 4 YOLO v7-tiny 网络结构

Fig. 4 YOLO v7-tiny network structure

### 2.2 训练平台

本研究的训练和测试均在同一台计算机上完成, 计算机硬件及软件信息为: Intel Core i7-11700K 3.6 GHz 的 16 核台式计算机, 机身内存为 64 GB, GPU 为 NVIDIA GeForce RTX 3090, 拥有 24 GB 显存, 操作系统为 64 位的 Ubuntu 18.04, Pytorch 深度学习环境, Python 版本为 3.8.16, CUDA 版本为 11.1, OpenCV 版本为 4.7.0.68, Python IDE 为 PyCharm Community Edition 2022。

### 2.3 模型训练

训练过程采用随机梯度下降 (Stochastic gradient descent, SGD) 优化算法, 模型训练批次 (Batch size) 设置为 32, 初始学习率 (Learning

的新一代 YOLO 目标检测网络。该网络采用新型骨干网络, 引入高效聚合网络 (Efficient layer aggregation networks, ELAN) 结构, 在保证模型性能的同时降低了模型计算复杂度; 改进了特征金字塔, 引入一种新的特征金字塔融合机制 SPPCSPC<sup>[23]</sup>, 可以将不同层级的特征融合地更加紧密; 使用卷积重参数化 (Re-parameterization convolution, RepConv)<sup>[25]</sup> 优化出了更有效的目标检测头部等, 此外 YOLO v7 还使用深度监督技术, 在网络中间层添加额外的辅助头 (Aux head), 可以更好地结合浅层特征, 提高训练精度。YOLO v7-tiny 是 YOLO v7 系列针对边缘计算设备的小模型, 具体特征为模型参数量小, 其整体网络结构与 YOLO v7 基本相同, 主要区别在于主干网络中的 ELAN 结构相较于 YOLO v7 减少了一个连接分支; 使用多个最大值池化代替特征金字塔, 且在网络检测头部未使用重参卷积以及使用 Leaky ReLU 激活函数和 CIoU<sup>[26]</sup> 损失函数, 其网络结构如图 4 所示。

rate) 为 0.01, 动量因子 (Momentum) 为 0.937, 权重衰减率 (Weight decay) 为 0.0005, 训练过程中使用 mosaic 增强, 色调 (Hue)、饱和度 (Saturation) 和亮度 (Value) 增强系数分别设置为 0.015、0.7 和 0.4, 训练轮数 (Epochs) 设定为 300, 输入图像尺寸设置为 640 像素 × 640 像素。完成 300 轮训练耗费时间 10.45 h, 训练完成后, 保存得到的识别模型的权重文件, 并输出模型在验证集上的精确度 (Precision,  $P$ )、召回率 (Recall,  $R$ )、IOU (交并比) 为 0.5 时的平均准确度 (mAP0.5)、IOU 在不同阈值 (从 0.5 到 0.95, 步长 0.05) 上的平均精度均值 (mAP0.5-0.95) 和 F1 值。

## 2.4 模型验证与测试

### 2.4.1 模型评价指标

采用精确度、召回率、mAP0.5、mAP0.5-0.95和F1值等指标对训练好的正常果和畸形果识别模型进行性能评估。

### 2.4.2 验证结果与评估

使用YOLO v7-tiny在验证集上对识别结果作进一步分析,在982幅图像的验证集上共有1159个识别目标,其中正常果实(Normal)737个,畸形果

实(Malformed)422个,具体识别结果如表1所示。结果表明,对于正常果实,模型精确度、召回率、mAP0.5、mAP0.5-0.95和F1值分别为93.0%、95.5%、96.9%、84.0%和94.2%,对于畸形果实,模型精确度、召回率、mAP、mAP0.5-0.95和F1值分别为94.3%、90.6%、96.1%、83.7%和92.4%,总体模型精确度、召回率、mAP、mAP0.5-0.95和F1值分别为93.7%、93.1%、96.5%、83.8%和93.4%,训练产生的模型内存占用量为12.3 MB。

表1 YOLO v7-tiny 训练结果

Tab.1 YOLO v7-tiny training results

类别	图像数量/幅	标签/个	P/%	R/%	mAP0.5/%	mAP0.5-0.95/%	F1值/%
整体	982	1159	93.7	93.1	96.5	83.8	93.4
正常果	982	737	93.0	95.5	96.9	84.0	94.2
畸形果	982	422	94.3	90.6	96.1	83.7	92.4

## 2.5 测试与评估

使用训练好的模型权重在测试集上评估识别结果,输入图像尺寸设置为640像素×640像素,置信度阈值设置为0.4,IoU阈值设置为0.5,使用GPU对测试集上的491幅图像进行检测,单幅图像平均检测时间为4.0 ms,部分识别结果如图5所示。识别结果表明,在未经训练过的测试集上进行识别,存在同一目标被重复识别以及叶片被识别成果实等问题。

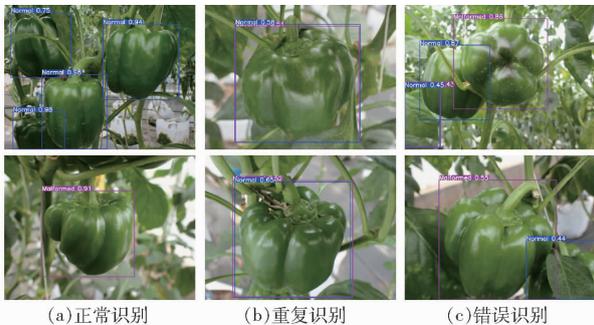


图5 YOLO v7-tiny 识别结果

Fig.5 YOLO v7-tiny recognition results

## 3 改进YOLO v7-tiny模型

根据2.5节测试结果可知,由于正常果和畸形果部分特征相接近,原始YOLO v7-tiny模型容易将同一目标识别出多个标签以及将部分叶子识别成果实,无法很好地满足甜椒正常果和畸形果识别,造成上述结果的原因可能是模型特征提取能力不足,从而导致模型拟合程度不好。为解决上述问题,需要对YOLO v7-tiny进行改进,具体改进为:将激活函数Leaky ReLU改为SiLU;将损失函数CIOU改为Focal-EIOU;在主干网络中添加SimAM。

### 3.1 Leaky ReLU 改为 SiLU

激活函数是确定神经网络输出的数学方程式,可以给神经元引入非线性因素,使得神经网络可以应用到众多非线性模型中。Leaky ReLU作为神经网络的激活函数之一,是在ReLU激活函数基础上,解决神经元接近零或为负值的时函数梯度为0,从而导致无法反向传播,使得神经元在负值时无法进行学习的问题;Leaky ReLU在负区域具有一个很小的斜率,使得神经元可以进行反向传播,其数学表达式为

$$L_{eakyReLU}(x) = \begin{cases} x & (x \geq 0) \\ \alpha x & (x < 0) \end{cases} \quad (1)$$

$\alpha$ 为控制负斜率的角度,默认为0.01。由于Leaky ReLU在不同区间上有不一样的函数,导致其非线性程度不高,无法为正负输入值提供一致的关系预测。

SiLU<sup>[27]</sup>激活函数结合了ReLU和Sigmoid函数,具有无上界有下界、平滑、非单调的特性,其数学表达式为

$$S_{iLU}(x) = \frac{x}{1 + e^{-x}} \quad (2)$$

在深度神经网络中,既有ReLU激活函数的优点,又因为SiLU是平滑函数,在整个定义域内都可以求导,更有利于优化。

### 3.2 CIOU 改成 Focal-EIOU

CIOU是在DIOU(Distance intersection over union)<sup>[26]</sup>损失函数的基础上考虑了预测框(Bounding box)的长宽比,进一步提高了模型精度,CIOU和DIOU均来自同一文献,数学表达式为

$$L_{oss_{CIOU}} = 1 - I_{OU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{(gt)})}{C^2} + \alpha'v \quad (3)$$

其中

$$\alpha' = \frac{v}{1 - I_{OU} + v}$$

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

式中  $\mathbf{b}$ 、 $\mathbf{b}^{(gt)}$ ——预测框、真实框中心点  
 $\rho$ ——计算两个中心点间的欧氏距离  
 $C$ ——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  
 $\alpha'$ ——适用于做权衡的参数  
 $v$ ——用来衡量长宽比一致性的参数  
 $I_{OU}$ ——交并比  
 $w^{(gt)}$ 、 $w$ ——真实框、预测框宽度  
 $h^{(gt)}$ 、 $h$ ——真实框、预测框高度

CIoU 解决了预测框和真实框是包含关系的情况下或者处于水平和垂直方向上 GIoU (Generalized intersection over union)<sup>[28]</sup> 损失几乎退化成为 IOU 损失的问题,且可以直接优化两个目标框的距离,收敛更快,其结构示意图如图 6 所示;但是由于是通过其公式中的  $v$  反映纵横比的差异,而不是从宽、高层面反映对置信度的影响,所以有时会阻碍有效的优化相似性。

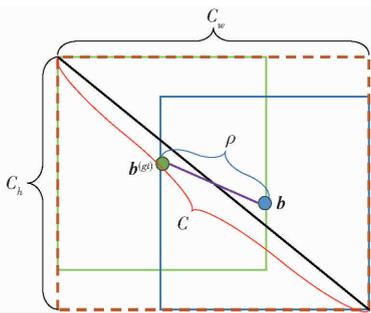


图 6 CIOU 计算结构示意图

Fig. 6 CIOU calculation structure diagram

因此,针对这一问题,Focal-EIOU<sup>[29]</sup> 在 CIOU 的基础上将纵横比拆开,并加入 Focal 聚类优质的锚框。EIOU 损失函数主要包含重叠损失、中心距离损失和宽高损失,前两部分与 CIOU 一致,不同的是宽高损失是直接使预测框和真实框的宽度和高度之差最小,具有更快的收敛速度。EIOU 损失函数和 Focal-EIOU 损失函数表达式为

$$\begin{cases} L_{oss_{EIOU}} = 1 - I_{OU} + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{(gt)})}{C^2} + \frac{\rho^2(w, w^{(gt)})}{C_w^2} + \\ \frac{\rho^2(h, h^{(gt)})}{C_h^2} \\ L_{oss_{Focal-EIOU}} = \gamma L_{oss_{EIOU}} \end{cases} \quad (4)$$

式中  $\gamma$ ——控制异常值抑制程度的参数,取 3  
 $C_w$ 、 $C_h$ ——最小外接矩形框宽度、高度

Focal-EIOU 引入 Focal Loss 优化了边界框回归任务中的样本不平衡问题,即减少与目标框重叠较少的大量锚框 (Anchor box) 对 Bounding Box 回归的优化贡献,使回归过程专注于高质量锚框。

### 3.3 在主干网络中加入 SimAM

SimAM<sup>[30]</sup> 是一种无参数的注意力机制。不同于传统的通道注意力生成一维权重和空间注意力生成二维权重,SimAM 是具有全三维权重的注意力机制 (Full 3-D weights for attention),全三维权重可以更好的细化特征,其结构示意图如图 7c 所示。在视觉任务中,为了更好地实现注意力,表现出明显空间抑制效果的神经元应该被给予更高的优先级,通过定义一个能量函数度量神经元之间的线性可分性,以此寻找重要神经元且该能量函数具有解析解,可以加快权重计算,能量函数公式为

$$e_t(w_t, b_t, \varphi, x_i) = (\varphi_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (\varphi_o - \hat{x}_i)^2 \quad (5)$$

其中  $\hat{t} = w_t t + b_t \quad (6)$

$\hat{x}_i = w_i x_i + b_i \quad (7)$

式中  $t$ ——目标神经元  
 $x_i$ ——其他神经元  
 $w_t$ ——权重  
 $b_t$ ——偏差  
 $M$ ——神经元数量

最小化上述公式等价于训练同一通道内神经元  $t$  与其他神经元之间的线性可分性,采用二值标签 (即  $\varphi_t$ 、 $\varphi_o$  为 1 和 -1) 简化计算,并添加正则项,最终的能量函数定义为

$$e_t(w_t, b_t, y, x_i) = (1 - t)^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - \hat{x}_i)^2 + \lambda w_t^2 \quad (8)$$

式中  $\lambda$ ——正则化平衡参数

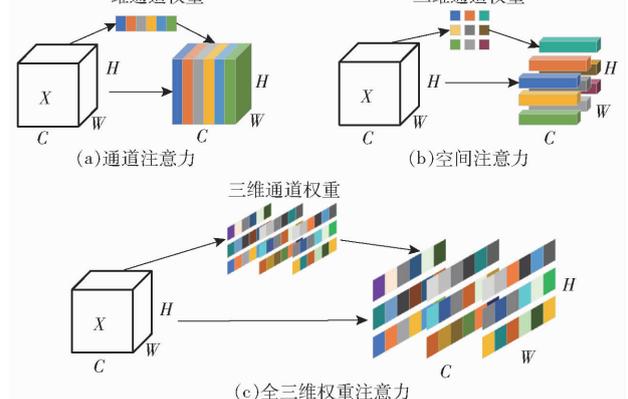


图 7 SimAM 结构示意图

Fig. 7 Schematic of SimAM structure

且具有解析解,公式为

$$\begin{cases} w_i = -\frac{2(t - u_i)}{(t - u_i)^2 + 2\sigma_i^2 + 2\lambda} \\ b_i = -\frac{1}{2}(t + u_i)w_i \end{cases} \quad (9)$$

其中  $u_i$  和  $\sigma_i^2$  是该通道除  $t$  以外的所有神经元的均值和方差,因此,最小能量公式为

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (10)$$

因此,公式(10)表明能量越低,神经元  $t$  与周围

神经元区别越大,重要性越高。最后通过 Sigmoid 函数抑制过大的值,从而满足注意力机制的定义,公式为

$$X = \text{Sigmoid}\left(\frac{1}{e_i^*}\right) \odot X \quad (11)$$

综上,对 YOLO v7-tiny 网络进行改进,改进后的网络结构如图 8 所示,具体改进为在骨干网络中 ELAN 模块后面加入 SimAM 模块,整合 ELAN 模块获得的特征,增强特征提取能力,同时将 Leaky ReLU 替换为 SiLU。

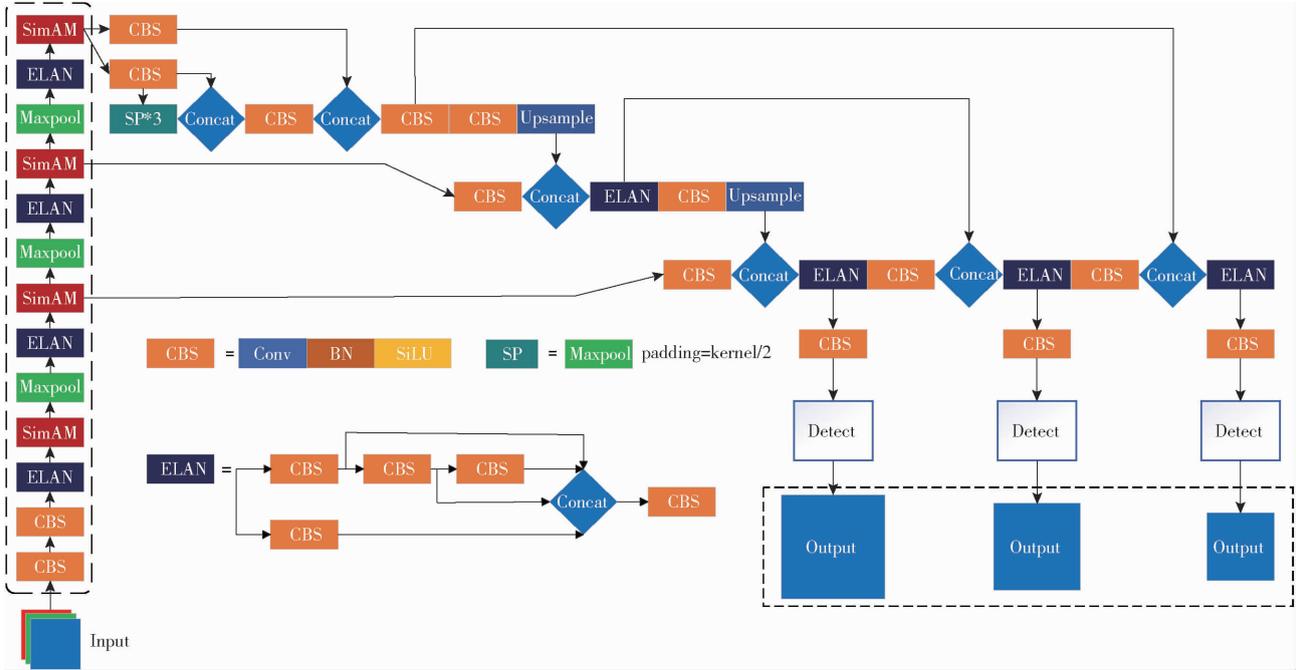


图 8 改进后的 YOLO v7-tiny 网络结构图

Fig. 8 Improved YOLO v7-tiny network structure diagram

## 4 结果与分析

### 4.1 改进模型试验结果

#### 4.1.1 不同激活函数试验结果

本试验在 YOLO v7-tiny 基础上,将激活函数 Leaky ReLU 更改为 SiLU 训练目标检测模型,将训练的结果进行分析,验证改进方案的可行性。由表 2 可知,当使用 SiLU 为模型的激活函数时,模型整体精确度相较于原始模型提升 4.2 个百分点,召回率提升 4.5 个百分点,mAP0.5 和 mAP0.5-0.95 分别提升 2.5、5.9 个百分点,内存占用量为 12.3 MB 保持不变。

#### 4.1.2 不同 IOU 损失函数试验结果

原始 YOLO v7-tiny 模型采用的是 CIOU 作为目标框和真实框之间的损失函数,但由于 CIOU 在训练过程中没有分别考虑宽高比对检测结果的影响,导致收敛速度慢。因此本试验在使用 SiLU 作为

表 2 不同激活函数识别结果对比

Tab. 2 Comparison of recognition results of different activation functions

激活函数	P/%	R/%	mAP0.5/ %	mAP0.5-0.95/ %	模型内存 占用量/MB
Leaky ReLU	93.7	93.1	96.5	83.8	12.3
SiLU	97.9	97.6	98.5	89.7	12.3

激活函数的模型基础上,通过对比不同的 IOU 损失函数,分析各损失函数对模型性能的影响,试验结果如表 3 所示,不同损失函数在训练集和验证集上的损失下降曲线分别如图 9a 和图 9b 所示。结合图表可知,EIOU 对模型的各项指标虽然均有所提升,但由于其在验证集上的损失曲线表现出先下降再上升的趋势,表明模型出现严重的过拟合现象,所得结果无法满足收敛要求;SIOU (Soft intersection over union) 在各项指标上均有所下降,而 Focal-EIOU 相较于 CIOU,整体波动较小,其损失下降得更快且

拥有更小的收敛损失,在精确度和 mAP0.5 上提升 0.1 个百分点,在召回率和 mAP0.5 - 0.95 分别下降 0.6、0.3 个百分点。

表 3 不同 IOU 损失函数识别结果对比

Tab.3 Comparison of recognition results of different IOU loss functions

损失函数	P/%	R/%	mAP0.5/ %	mAP0.5 - 0.95/ %	模型内存 占用量/MB
CIOU	97.9	97.6	98.5	89.7	12.3
EIOU	98.9	97.6	99.0	95.9	12.3
SIOU	97.0	95.4	97.5	86.7	12.3
Focal - EIOU	98.0	97.0	98.6	89.4	12.3

### 4.1.3 YOLO v7 - tiny 改进前后试验结果

在上述使用 SiLU 激活函数和 Focal - EIOU 损失函数的试验基础上,对 YOLO v7 - tiny 网络进行改进。通过在骨干网络中加入 SimAM,增强模型的特征提取能力,进一步验证改进模型的性能,结果如表 4 所示。由表可知,SimAM 结合 Focal - EIOU 和 SiLU,可以大幅度的提升模型整体性能,其中在精准度上提升 5.4 个百分点,在召回率上提升 4.7 个百分点,在 mAP0.5 上提升 2.4 个百分点,在 mAP0.5 - 0.95 上提升 10.7 个百分点;模型内存占用量

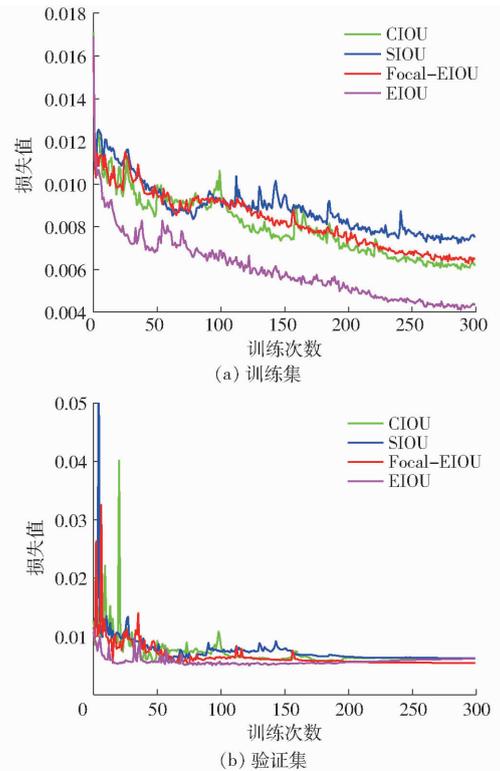


图 9 不同 IOU 损失函数曲线

Fig.9 Different IOU loss function curves

10.6 MB,降低 13.8%。

表 4 改进前后识别结果对比

Tab.4 Comparison of recognition results before and after improvement

模型	P/%	R/%	mAP0.5/%	mAP0.5 - 0.95/%	模型内存占用量/MB	单幅图像检测时间/ms
YOLO v7 - tiny	93.7	93.1	96.5	83.8	12.3	4.0
改进 YOLO v7 - tiny	99.1	97.8	98.9	94.5	10.6	4.2

此外,加入注意力机制可以让模型更加关注于甜椒果实本身,为了更加直观地展现添加注意力机制前后的区别,可以使用 Grad - CAM (Gradient-weighted class activation mapping)<sup>[31]</sup> 分别获取模型最后 3 层的热力图,分别对应模型最后的检测层。Grad - CAM 是神经网络可视化工具,可以根据不同部位的权重生成热力图,详细地描述模型效果,如图 10 所示,图中红色越深表示特征权重越大,可以明显看出,未使用 SimAM 的模型对一些与甜椒表面接近的叶子的特征也给予了较大的权重,容易使得

识别结果出错,造成误判,而使用 SimAM 的模型更加注重甜椒果实区域的特征提取,对于其他部分则较少关注,进而证明 SimAM 模块的有效性。

### 4.1.4 消融试验

改进 YOLO v7 - tiny 模型在 YOLO v7 - tiny 基础上融合了 SimAM 注意力模块,使用了 SiLU 激活函数和 Focal - EIOU 损失函数。为了能更加直观地分析改进 YOLO v7 - tiny 模型相较于原始模型的提升,分别进行 4 组试验并对比试验的 mAP0.5 - 0.95,试验结果如表 5 所示。

表 5 消融试验结果

Tab.5 Ablation test results

试验 编号	Leaky ReLU	CIOU	SiLU	Focal - EIOU	SimAM	mAP0.5 - 0.95/%
1	√	√				83.8
2		√	√			89.7
3			√	√		89.4
4			√	√	√	94.5

注:√表示使用该方法。

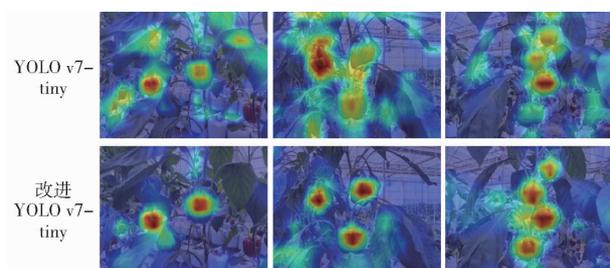


图 10 加入 SimAM 前后模型热力图

Fig.10 Model heat maps before and after adding SimAM

试验 1 为原始 YOLO v7 - tiny 模型, 在使用 Leaky ReLU 激活函数和 CIUO 损失函数的条件下 mAP0.5 - 0.95 为 83.8%, 与试验 1 不同, 试验 2 是将 Leaky ReLU 替换为 SiLU 激活函数, mAP0.5 - 0.95 提升 5.9 个百分点, 试验 3 是在试验 2 的基础上, 将 CIUO 替换为 Focal - EIOU, mAP0.5 - 0.95 下降 0.3 个百分点, 试验 4 是在试验 3 基础上加入 SimAM 注意力模块, mAP0.5 - 0.95 提升 5.1 个百分点。

#### 4.2 不同目标检测算法对比试验

通过上述方法对 YOLO v7 - tiny 进行一系列改进取得了精度上的提升, 为了进一步验证改进 YOLO v7 - tiny 的识别性能, 在相同数据集上, 与 YOLO v7、Scaled - YOLO v4<sup>[32]</sup>、YOLOR - CSP<sup>[33]</sup> 等目标检测模型进行试验对比, 各试验参数保持一致, 比较不同模型在验证集上的整体 F1 值、mAP0.5、mAP0.5 - 0.95、模型内存占用量以及单幅图像检测时间。由表 6 可知, 改进 YOLO v7 - tiny 在 F1 值上基本达到 YOLO v7 的水平, 比 Scaled - YOLO v4、YOLOR - CSP 分别高 0.7、0.2 个百分点, mAP0.5 比 YOLO v7 和 YOLOR - CSP 分别降低 0.2、0.4 个

百分点, 相比 Scaled - YOLO v4 提升 0.1 个百分点, mAP0.5 - 0.95 上相比 YOLO v7、Scaled - YOLO v4、YOLOR - CSP 分别提升 0.6、1.2、0.2 个百分点, 模型内存占用量上分别降低 85.8%、89.9%、89.9%, 仅为 YOLO v7、Scaled - YOLO v4、YOLOR - CSP 的 14.2%、10.1%、10.1%。

表 6 不同目标检测模型识别结果对比

Tab. 6 Comparison of recognition results of different target detection models

模型	F1 值/ %	mAP0.5/ %	mAP0.5 - 0.95/%	模型内 存占用 量/MB	单幅图像 检测时 间/ms
YOLO v7	98.5	99.1	93.9	74.8	4.2
Scaled - YOLO v4	97.7	98.8	93.3	105.5	4.5
YOLOR - CSP	98.2	99.3	94.3	105.5	4.5
改进 YOLO v7 - tiny	98.4	98.9	94.5	10.6	4.2

为了进一步反映不同模型的识别性能, 将不同模型分别在同一测试集上进行目标检测, 其中输入图像尺寸为 640 像素 × 640 像素, 置信度阈值和 IOU 阈值均与 2.5 节保持一致, 将各识别结果与真实值进行对比, 如图 11 (图中红色箭头表示将叶子或其

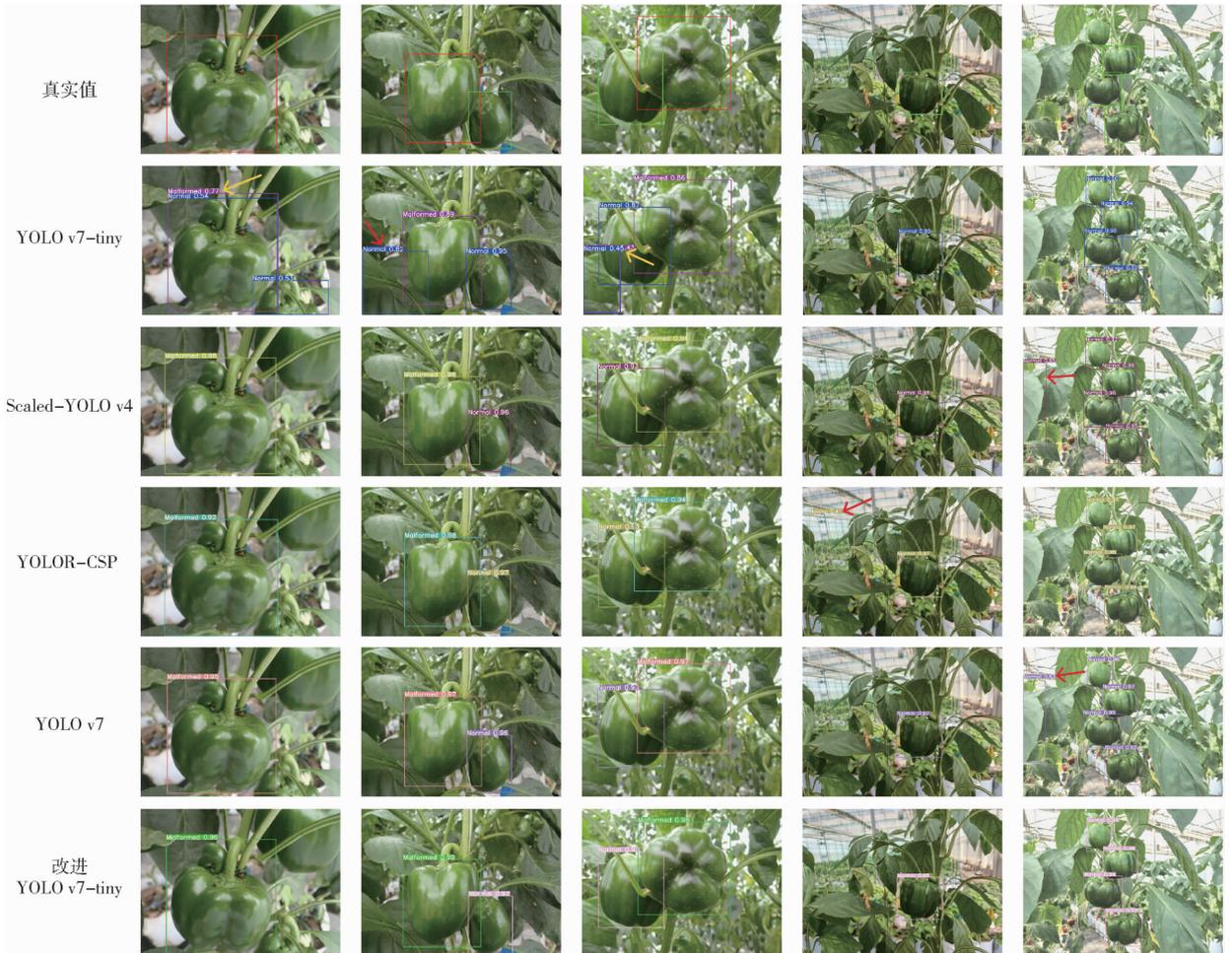


图 11 不同模型识别结果

Fig. 11 Different model recognition results

他物体识别成果实,黄色箭头表示对同一果实识别出两种结果)所示,从识别结果可以看出,改进后的模型具有很强的识别能力,相较于原始的 YOLO v7-tiny 网络,没有出现对同一目标识别出两种结果的情况,且由于注意力机制的加入,让模型能够更好地关注甜椒的表面特征以及区分不同特征,避免将甜椒叶子识别成目标,而在其他模型的识别结果中均出现将叶子或其他物体识别成甜椒果实的情况。

## 5 结束语

针对设施场景中甜椒容易出现畸形果这一问题,构建了正常果实和畸形果实不同环境下的数据集,提出了基于改进 YOLO v7-tiny 的甜椒畸形果实识别模型。为解决 YOLO v7-tiny 识别中出现的识别错误和重复识别问题,增强模型的特征提取能力,提高模型的检测精度,在原始模型的骨干网络中融入了 SimAM 注意力模块,该模块具有无参数的

特征,可以降低参数量而又提升精度;除此之外,将原始模型中的激活函数 Leaky ReLU 用 SiLU 激活函数代替,增强模型的非线性特征提取能力;损失函数由 CIOU 损失替换为 Focal-EIOU 损失,加快模型收敛并降低损失值。试验结果表明:改进后的 YOLO v7-tiny 的整体精确度、召回率、mAP0.5、mAP0.5-0.95 分别为 99.1%、97.8%、98.9%、94.5%,相较于 YOLO v7-tiny 分别提升 5.4、4.7、2.4、10.7 个百分点,模型内存占用量为 10.6 MB,单幅图像的检测时间为 4.2 ms。与 YOLO v7、Scaled-YOLO v4、YOLOR-CSP 等目标检测模型相比,改进 YOLO v7-tiny 的 mAP0.5-0.95 分别提升 0.6、1.2、0.2 个百分点,而模型内存占用量为仅为上述模型的 14.2%、10.0%、10.0%。得到了小体量而高精度的目标检测模型,能够在快速识别目标的同时保证较高精度,进一步验证了改进方法对甜椒果实识别的有效性,为后续机械化采摘和品质分级提供技术支持。

## 参 考 文 献

- [1] 邹学校,马艳青,戴雄泽,等. 辣椒在中国的传播与产业发展[J]. 园艺学报, 2020, 47(9): 1715-1716.  
ZOU Xuexiao, MA Yanqing, DAI Xiongze, et al. Spread and industry development of pepper in China[J]. Acta Horticulturae Sinica, 2020, 47(9): 1715-1716. (in Chinese)
- [2] TANG Y, CHEN M, WANG C, et al. Recognition and localization methods for vision-based fruit picking robots: a review[J]. Front Plant Sci., 2020, 11: 510.
- [3] 程演演,祁力钧,程一帆,等. 基于 M-LP 特征加权聚类的果树冠层图像分割方法[J]. 农业机械学报, 2020, 51(4): 191-198, 260.  
CHENG Zhenzhen, QI Lijun, CHENG Yifan, et al. Fruit tree canopy image segmentation method based on M-LP features weighted clustering[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 191-198, 260. (in Chinese)
- [4] 刘慧,朱晟辉,沈跃,等. 基于多特征融合的树干快速分割算法[J]. 农业机械学报, 2020, 51(1): 221-229.  
LIU Hui, ZHU Shenghui, SHEN Yue, et al. Fast segmentation algorithm of tree trunks based on multi-feature fusion[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(1): 221-229. (in Chinese)
- [5] CHEN M, TANG Y, ZOU X, et al. Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology[J]. Computers and Electronics in Agriculture, 2020, 174: 105508.
- [6] SYAZWANI R W N, ASRAF H M, AMIN M A M S, et al. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning[J]. Alexandria Engineering Journal, 2022, 61(2): 1265-1276.
- [7] 罗陆锋,邹湘军,熊俊涛,等. 自然环境下葡萄采摘机器人采摘点的自动定位[J]. 农业工程学报, 2015, 31(2): 14-21.  
LUO Lufeng, ZOU Xiangjun, XIONG Juntao, et al. Automatic positioning of grape picking robot picking points under natural environment[J]. Transactions of the CSAE, 2015, 31(2): 14-21. (in Chinese)
- [8] 王青,朱德利,李炜. 增强的 OHTA 颜色空间融合 FPFH 识别点云中的甜椒果实[J]. 重庆师范大学学报(自然科学版), 2022, 39(2): 111-119.  
WANG Qing, ZHU Deli, LI Wei. Enhanced OHTA color space fusion FPFH to identify sweet pepper fruits in point cloud[J]. Journal of Chongqing Normal University (Natural Science Edition), 2022, 39(2): 111-119. (in Chinese)
- [9] BAC C W, HEMMING J, VAN HENTEN E J. Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper[J]. Computers and Electronics in Agriculture, 2013, 96: 148-162.
- [10] VITZRABIN E, EDAN Y. Adaptive thresholding with fusion using a RGBD sensor for red sweet-pepper detection[J]. Biosystems Engineering, 2016, 146: 45-56.
- [11] SA I, GE Z, DAYOUB F, et al. DeepFruits: a fruit detection system using deep neural networks[J]. Sensors, 2016, 16(8): 1222.
- [12] LÓPEZ-BARRIOS J D, ESCOBEDO CABELLO J A, GÓMEZ-ESPINOSA A, et al. Green sweet pepper fruit and peduncle detection using Mask R-CNN in greenhouses[J]. Applied Sciences, 2023, 13(10): 6296.

- [13] CONG P, LI S, ZHOU J, et al. Research on instance segmentation algorithm of greenhouse sweet pepper detection based on improved Mask RCNN[J]. *Agronomy*, 2023, 13(1): 196.
- [14] QIU C, TIAN G, ZHAO J, et al. Grape maturity detection and visual pre-positioning based on improved YOLO v4[J]. *Electronics*, 2022, 11(17): 2677.
- [15] 宋怀波, 马宝玲, 尚钰莹, 等. 基于 YOLO v7-ECA 模型的苹果幼果检测[J]. *农业机械学报*, 2023, 54(6): 233-242.  
SONG Huaibo, MA Baoling, SHANG Yuying, et al. Detection of young apple fruits based on YOLO v7-ECA model[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(6): 233-242. (in Chinese)
- [16] 张志远, 罗铭毅, 郭树欣, 等. 基于改进 YOLO v5 的自然环境下樱桃果实识别方法[J]. *农业机械学报*, 2022, 53(增刊1): 232-240.  
ZHANG Zhiyuan, LUO Mingyi, GUO Shuxin, et al. Cherry fruit detection method in natural scene based on improved YOLO v5[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2022, 53(Supp.1): 232-240. (in Chinese)
- [17] CHEN S, ZOU X, ZHOU X, et al. Study on fusion clustering and improved YOLO v5 algorithm based on multiple occlusion of *Camellia oleifera* fruit[J]. *Computers and Electronics in Agriculture*, 2023, 206: 107706.
- [18] ZHAO K, ZHAO L, ZHAO Y, et al. Study on lightweight model of maize seedling object detection based on YOLO v7[J]. *Applied Sciences*, 2023, 13(13): 7731.
- [19] 张立杰, 周舒骅, 李娜, 等. 基于改进 SSD 卷积神经网络的苹果定位与分级方法[J]. *农业机械学报*, 2023, 54(6): 223-232.  
ZHANG Lijie, ZHOU Shuhua, LI Na, et al. Apple location and classification based on improved SSD convolutional neural network[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(6): 223-232. (in Chinese)
- [20] LI K, WANG J, JALIL H, et al. A fast and lightweight detection algorithm for passion fruit pests based on improved YOLO v5[J]. *Computers and Electronics in Agriculture*, 2023, 204: 107534.
- [21] 张万枝, 曾祥, 刘树峰, 等. 基于改进 YOLO v5s 的马铃薯种薯芽眼检测方法[J]. *农业机械学报*, 2023, 54(9): 260-269.  
ZHANG Wanzhi, ZENG Xiang, LIU Shufeng, et al. Detection method of potato seed bud eye based on improved YOLO v5s[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2023, 54(9): 260-269. (in Chinese)
- [22] NING Z, LUO L, DING X M, et al. Recognition of sweet peppers and planning the robotic picking sequence in high-density orchards[J]. *Computers and Electronics in Agriculture*, 2022, 196: 106878.
- [23] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464-7475.
- [24] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection[J]. *arXiv preprint arXiv: 2004. 10934*, 2020.
- [25] DING X, ZHANG X, MA N, et al. Repvgg: making vgg-style convnets great again[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13733-13742.
- [26] 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.
- [27] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. *Neural Networks*, 2018, 107: 3-11.
- [28] 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.
- [29] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, 506: 146-157.
- [30] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]//*International Conference on Machine Learning*. PMLR, 2021: 11863-11874.
- [31] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 618-626.
- [32] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-Yolov4: scaling cross stage partial network[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13029-13038.
- [33] WANG C Y, YE H I, LIAO H Y M. You only learn one representation: unified network for multiple tasks[J]. *arXiv Preprint, arXiv: 2105.04206*, 2021.