

基于 CBAM-YOLOv8 的温室番茄果实识别

廖新芝¹, 孔国希¹, 林桂潮¹, 曹惠茹², 李小敏¹

(1. 仲恺农业工程学院机电工程学院 广州 510225; 2. 广州工程技术职业学院信息工程学院 广州 510075)

摘要: 在农业自动化和精准农业的背景下, 果实识别检测是智能收获系统中的重要环节。快速且准确地识别番茄果实大小、数量、位置等信息, 对番茄生长状况实时监测、产量检测和提高采摘效率具有重要意义。然而, 现有的基于深度学习的目标检测算法在番茄果实检测领域仍面临挑战, 由于现实应用的需要, 期待目标检测模型更加精确化和快速化, 然而这两者难以同时兼顾, 所以在实际应用中, 根据应用需求来权衡这两者的关系。笔者提出一种基于 YOLOv8 的番茄检测改进模型, 将 CBAM 注意力机制加入到 YOLOv8 模型的特征提取环节。CBAM 体积小, 可方便集成在任一 CNN 架构中, 有效抑制噪声和无关信息, 动态调整特征图的权重, 提升模型检测的精确率。经过试验, 该模型的精确率、召回率、平均精度分别达到了 91%、78% 和 90%, 与 SSD 相比, 分别提升 21、14 和 15 个百分点; 与 Faster RCNN 相比, 分别提升了 13、11 和 15 个百分点; 与原始的 YOLOv8 相比, 分别提升了 5、3 和 7 个百分点。这说明了 CBAM-YOLOv8 模型降低了误检率和漏检率。在预测时间方面, YOLOv8 的综合耗时最短。相比之下, CBAM-YOLOv8 模型所需的预测时间有所增加, 推断速度更慢, 从而增加了计算成本。因此, 在实际应用中需要在精度与速度之间进行权衡。综上所述, CBAM-YOLOv8 模型为番茄果实精准、高效识别提供了一种有效的解决方案, 将来可应用于番茄实时监测、计数和采摘中。

关键词: 番茄识别; 目标检测; 注意力机制; CBAM-YOLOv8

中图分类号: S641.2

文献标志码: A

文章编号: 1673-2871(2025)09-048-09

Greenhouse tomato fruit recognition based on CBAM-YOLOv8

LIAO Xinzhi¹, KONG Guoxi¹, LIN Guichao¹, CAO Huiru², LI Xiaomin¹

(1. College of Mechanical and Electrical Engineering, Zhongkai University of Agricultural Engineering, Guangzhou 510225, Guangdong, China; 2. College of Information Engineering, Guangzhou Institute of Technology, Guangzhou 510075, Guangdong, China)

Abstract: Efficient and accurate detection of tomato fruits is the key to achieving intelligent harvesting. To address the challenge of balancing accuracy and speed in existing deep learning object detection models, this study proposed an improved YOLOv8 model. This model incorporated the CBAM attention module during the YOLOv8 feature extraction stage, which dynamically adjusted feature weights to effectively suppress noise and irrelevant information, and improved the accuracy of model detection. The experiments showed that CBAM-YOLOv8 performed well in tomato detection, with accuracy, recall, and average precision reaching 91%, 78%, and 90%, respectively. Compared with SSD, Faster RCNN, and original YOLOv8, the performance had significantly improved. This model effectively reduced the rates of false positives and false negatives. In terms of prediction time, YOLOv8 has the shortest comprehensive time consumption. In contrast, the prediction time required by CBAM-YOLOv8 model has increased, and the inference speed is slower, thereby increasing the computational cost. Therefore, in practical applications, a balance needs to be made between accuracy and speed. In conclusion, CBAM-YOLOv8 provides an effective solution for real-time monitoring, yield estimation, and efficient harvesting of tomato fruits.

Key words: Tomato recognition; Target detection; Attention mechanism; CBAM-YOLOv8

在众多广泛种植的蔬菜作物中, 番茄具有不可忽视的重要营养和经济价值。然而, 在番茄种植和收获过程中, 果实的成熟程度、大小、位置等信息对于采摘和后续处理至关重要^[1-4]。传统的番茄果实

目标检测方法主要依赖于人工肉眼观察, 存在番茄果实识别速度慢、精确率低等问题, 因此, 开发一种基于机器学习的番茄果实检测方法具有重要的实际应用价值^[5]。

收稿日期: 2024-08-20; 修回日期: 2025-03-08

基金项目: 国家自然科学基金面上项目(32472015)

作者简介: 廖新芝, 女, 在读硕士研究生, 研究方向为农业物联网等。E-mail: 15089916924@163.com

通信作者: 李小敏, 男, 副教授, 研究方向为农业大数据。E-mail: xiaominli_09@163.com

目标检测是计算机视觉领域的核心问题,在过去 20 多年里,目标检测技术迅速发展^[6],经历了从早期的传统方法到现代的深度学习技术。传统目标检测方法主要依赖于手工设计的特征和机器学习算法。如 Viola 等^[7]提出的 Viola-Jones 检测器,该方法利用了 Haar-like 特征和 AdaBoost 分类器,并通过级联结构来逐步精细化检测过程,在人脸检测领域取得了巨大成功。Dalal 等^[8]提出 HOG,广泛应用于行人检测。Felzenszwalb 等^[9]提出 DPM,主要用于解决目标的尺度变化、遮挡和姿态变化,能够适应复杂场景的检测需求。自 2014 年深度学习技术兴起以来,目标检测算法的研究取得了重大进展。根据检测流程的差异,主流算法可分为两类,即一阶段目标检测算法(如 YOLO^[10]、SSD^[11]和 CenterNet^[12])与二阶段目标检测算法(如 RCNN^[13]、SPPNet^[14]和 Pyramid Networks^[15])。与二阶段目标检测算法不同,一阶段目标检测算法是在一个步骤中直接完成目标物体的分类和位置预测,无需预先生成候选区域。因此,一阶段检测算法检测速度更快,适用于实时检测的场景。

YOLO 作为一种一阶段目标检测算法,自 2015 年提出以来便获得广泛关注。在此基础上,许多研究人员也致力于对 YOLO 模型的改进,专注于模型优化。如 Huang 等^[16]提出了密集连接(增强特征利用)和空间金字塔池化(捕捉多尺度信息)的 YOLO 算法,即 DC-SPP-YOLO,提高了 YOLOv2 的目标检测精度。刘芳等^[17]提出了一种 IMS-YOLO 算法,实现了温室环境下农业采摘机器人对番茄果实快速、精准检测。Zhang 等^[18]针对 YOLOv4 进行了两处改进:首先,将 GhostNet 特征提取网络与坐标注意力模块结合;其次,通过深度可分离卷积重构 neck

部分和 head 部分。试验表明,改进后模型的检测速度提升了 5.7 fps。李丽等^[19]改进了 YOLOv5 算法,在颈部网络插入创新的多通道信息融合模块,修改颈部网络的拼接方法,在识别部分嵌入聚类算法模块,实现更精确的检测。这些改进的模型比原始模型参数量更少且计算复杂度更低,有效提高目标检测速度,更适应在资源有限的农业环境中工作。

在农业环境中,目标果实的检测主要应用于农业机器人采摘任务,需应对复杂场景及不确定性因素,基于深度学习的目标识别方法在识别和定位方面仍要进一步研究。通过在原始 YOLOv8 模型中融入卷积块注意力模块(Convolutional Block Attention Module, CBAM),动态增强果实特征的代表能力,以期为农业机器人的实时采摘任务提供可靠的技术支持。

1 温室番茄果实图像数据的构建

1.1 数据采集

图像数据收集于广东省湛江科学研究所的番茄种植基地,拍摄时间为 2024 年 3 月,拍摄设备选用 HONOR Magic5 Pro,图像以 JPG 格式保存。番茄果实多为球形,不同成熟度颜色有所差异,呈现为绿色、橙色或红色,而绿色果实与背景色相近,可能给识别过程带来一定的干扰。还有一些其他干扰,如果实被枝叶等严重遮挡、光照异常(如过曝或欠曝)导致特征不明显等。为构建具有多样性的数据集,在以下 3 种条件下采集了 5013 张图像:(1)多位置(植株顶部、中部和底部);(2)多时段(6:00—18:00 每小时采集);(3)多光照(自然光、补光灯和阴影组合),以此提升模型泛化能力与鲁棒性。采集的部分番茄图像如图 1 所示。



图 1 采集的番茄图像

Fig. 1 Tomato images captured

1.2 数据集标注与扩充

为了显著提升模型的泛化能力和鲁棒性,使其能够学习更多样化的特征,从原始图像中挑选了部分质量稳定的图像进行一系列数据扩充操作,这些操作包括随机裁剪、缩放、翻转和旋转等,旨在模拟不同拍摄角度和条件,从而丰富模型的训练样本^[20]。经过扩充和筛选之后,最终形成一个含有7200张图像的数据集。

将数据集划分为2个子集,80%的数据用作训练集,20%的数据作为测试集。按此分配比例,将5760幅图像分配为训练集,主要目的是用于模型的主体训练过程,确保模型能够充分学习到番茄果实识别的关键特征;测试集则由1440幅图像构成,用于最终测试模型的实际应用效果。

为了对预处理后的数据集进行精确标注,选用labellmg图像标注软件对图像进行标注。该软件对番茄果实进行了详细的标注,具体方法是围绕每个目标番茄果实绘制最小的外接矩形,并为这些矩形分配了单一类别标签“tomato”,并采用标准txt格式保存标注信息。

2 基于CBAM-YOLOv8温室番茄果实识别方法

2.1 CBAM-YOLOv8总体框架

YOLOv8^[21]是一种优秀的一阶段目标检测算法,它的网络结构分为主干网络backbone层、连接

Neck层和输出Head层。主干网络backbone层的作用是提取番茄数据集中的特征;连接Neck层的作用是进行特征融合;输出层基于Neck Head层是融合的多尺度特征,通过回归和分类预测目标边界框及类别。在YOLOv8中,conv为卷积操作,学习番茄图片中的局部特征和空间关系;C2f是特征融合操作,将番茄图片中的不同分支的特征图进行融合;UpSample为特征图上采样操作,在保持特征图宽高比的同时,增加特征图的通道数;Concat为特征图拼接。

在YOLOv8算法的基础上,将CBAM注意力机制模块嵌入到Head层的C2f模块与检测层之间,这样在特征提取的过程中,CBAM模块可以自动地识别出哪些通道和区域是对番茄识别更为关键的,并给它们分配更高的权重,由于YOLOv8算法本身具有较快的检测速度,结合CBAM注意力机制后,模型可以更快地定位到关键区域,从而提高识别效率。CBAM-YOLOv8网络模型如图2所示。

2.2 YOLOv8模型

2.2.1 Backbone部分 从图2中YOLOv8的Backbone层部分可以看出,在Backbone层中图像进行多次运用C2f特征融合操作和conv图像卷积操作。与YOLOv5^[22]相比,它做了一些改进,其中最为显著的改变是C3模块的替换。YOLOv5中的C3模块通过跨阶段的部分连接来增强网络的特征提取能力,而在v8中被替换成了C2f模块。C2f模

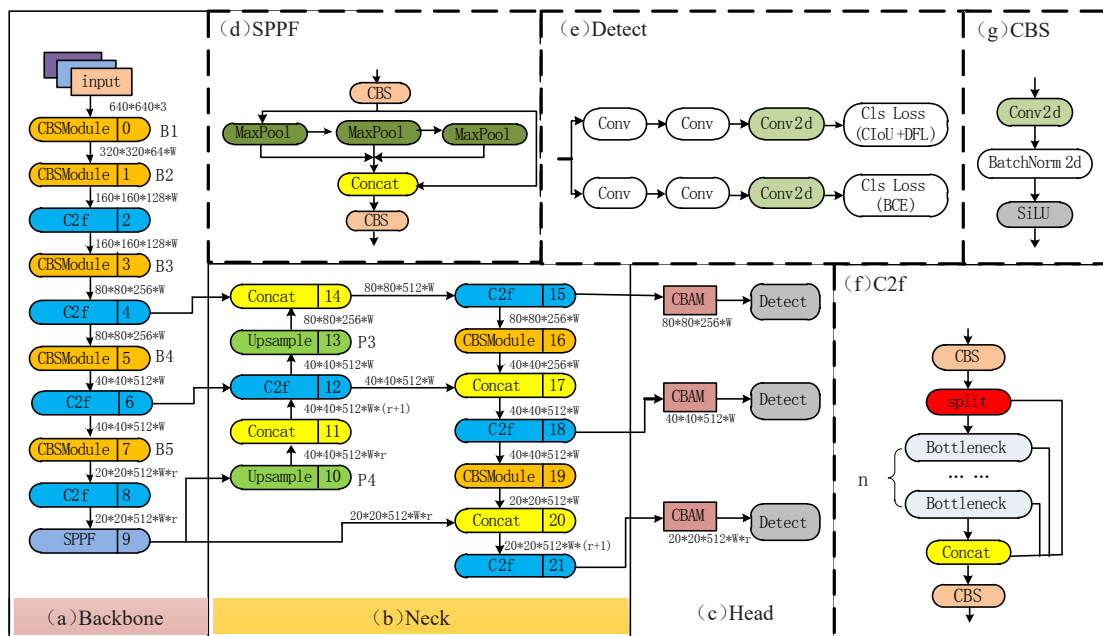


图2 CBAM-YOLOv8模型
Fig. 2 CBAM-YOLOv8 model

块是 YOLOv8 Backbone 中的核心模块,它借鉴了 YOLOv7 中的 ELAN (Efficient Layer Aggregation Network) 设计思想,在保持轻量化模型的同时,获得多样化的梯度流信息。C2f 模块在保证模型轻量化的同时,具有更强的特征提取能力和梯度传播能力,从而提高了模型识别的精确率和鲁棒性。

YOLOv8 在 YOLOv5 的基础上,对 SPP 模块改进并引入更为高效的 SPPF 模块,SPPF 模块既保留了空间金字塔设计中同时考虑局部特征和全局特征的优势,又进一步提升了模型训练过程中的计算效率。经过两者的融合后,进一步提升了特征图的表达,对于需要检测的图像里对象大小差距过大的情况以及对多目标这种复杂情况下的检测精度有很大的提升。SPPF 模块摒弃掉原来的 5×5 、 9×9 、 13×13 的最大池化,改用 3 个 5×5 的最大池化,可以在保留原有功能的同时,更有效融合不同的特征图,这种改进在提升模型的训练计算速度的同时实现了性能和效率的双重提升。

2.2.2 Neck 部分 从图 2 网络结构中的 Neck 层可以看出,图像在 Neck 中进行一系列上采样(Upsample)、C2f 特征融合、conv 卷积操作和 concat 特征图拼接,作用是接收来自 Backbone 的多个尺度的特征图,并将它们进行融合。这种融合可以包括上采样、下采样、卷积等操作,旨在提取不同尺度的特征信息并有效地结合起来,以提高模型对目标的检测能力。通过 Neck 层的处理,原始特征图可以得到增强,使得模型能够更好地捕捉目标的细节和上下文信息。这有助于提升目标检测的准确性和鲁棒性。

在 YOLOv8 中,Neck 层通常采用 PAN (Path Aggregation Network) 路径聚合网络结构或其改进版本。路径聚合网络结构由两部分组成:自顶向下的特征金字塔和自底向上的特征金字塔。自顶向下的特征金字塔部分结构将高层次的特征图通过双线性插值等上采样操作和横向连接的方式将高层次特征图与较低层次的特征图进行融合。这种融合可以使得低层次的特征图获得高层次更丰富的语义信息,从而提升对较小目标的检测精度;自底向上的特征金字塔则与自顶向下的结构相反,将低层次的特征图通过最大池化或卷积等下采样操作和横向连接的方式与高层次的特征图进行融合。这种融合可以使得高层次的特征图获得更精细的空间信息,有助于检测较大的目标。

2.2.3 Head 部分 从图 2 网络结构中看出,Head

部分相比之前的版本(如 YOLOv5)有了较大的变化。YOLOv8 在 Head 部分进行了优化,将原先的耦合式头部设计改进为更加高效的解耦头结构。这种结构使得模型在检测时能够更加灵活,可以根据不同的任务需求进行调整。同时,从 Anchor-based 转变为 Anchor-free,减少了计算量和内存占用空间,提高了模型的实时性。同时,YOLOv8 在 Loss 计算方面采用了 Task-Aligned Assigner 正样本分配策略,并引入了 Distribution Focal Loss。这种策略能够更准确地分配正样本,提高模型的训练效率和检测精度。Head 部分的具体结构会根据不同的任务而有所不同。在 YOLOv8 中,Head 部分会接收来自 Neck 层的特征图,并进行一系列的卷积、池化等操作,以进一步提取和融合特征,最终输出检测结果。检测结果包括边界框坐标、目标类别和目标类别置信度等信息,用于后续的目标检测任务。

2.3 CBAM 模块

CBAM 是一种用于增强卷积神经网络(CNN)性能的注意力机制模块^[23],在 2018 年,由 Woo 等^[24]提出。其小型、通用,可以方便集成到任何 CNN 架构中,有效抑制噪声和无关信息,动态地调整特征图的权重,从而极大地提升模型在处理图像数据时的准确性和效率。CBAM 模块可分为通道注意力模块(CAM)和空间注意力模块(SAM),如图 3 是 CBAM 的结构图。具体流程是:将 YOLO 模型处理的中间特征图作为 CBAM 的输入,先后经过通道注意力和空间注意力模块的处理,分别创建出一维的通道注意图 $M_c R^{C \times 1 \times 1}$ 和二维的空间注意图 $M_s R^{1 \times H \times W}$,最终得到 CBAM 模型的输出,再次回到 YOLO 模型进行后续处理。CBAM 整体处理过程可以概括为:

$$F' = M_c(F) \otimes F; \quad (1)$$

$$F'' = M_s(F') \otimes F'. \quad (2)$$

其中, F 表示一个中间特征图, $F R^{C \times H \times W}$,此为 CBAM 的输入; F' 为经过 CAM 处理后的输出特征图; F'' 为经过 SAM 处理后的输出特征图; $M_c(F)$ 为 CAM; $M_s(F')$ 为 SAM; \otimes 表示乘法运算。CBAM 模型如图 3 所示。

通道注意力结构图如图 4 所示,它用来强化每个通道的特征表示。输入的中间特征图进行平行的最大池化和平均池化,分别捕捉了通道内的最大特征值和平均特征值,从而生成 2 个反映通道全局信息的 $C \times 1 \times 1$ 向量;随后,将提取到的中间特征值

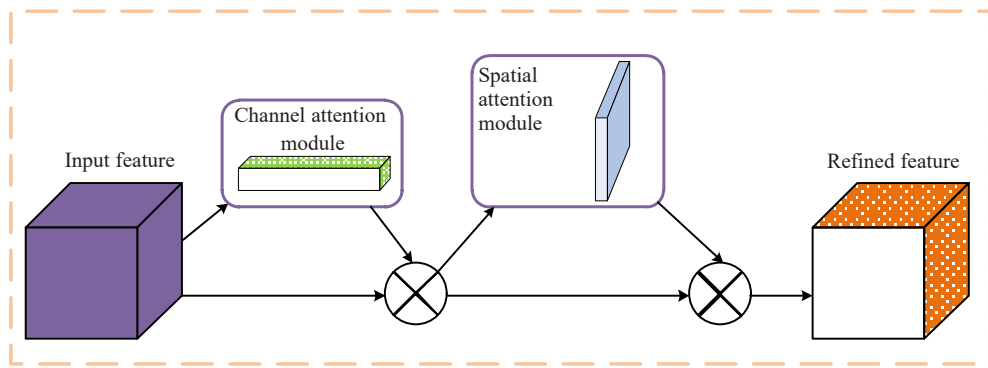


图3 CBAM 模块
Fig.3 CBAM module

输入到多层感知器中,在 Share MLP 中将得到的特征值转化为每个通道的注意力权重,使得模型在训练过程中能自适应判断哪些通道信息对目标检测任务更加重要。最终,将 MLP 输出的特征通过对位相乘并加权求和到原始特征中,再通过 Sigmoid 激活函数处理后,得到通道注意力 $M_c(F)$ 。计算公式如下:

$$M_c(F) = \sigma[MLP(AvgPool(F)) + MLP(MaxPool(F))] \\ = \sigma[W_1(W_0(F_{avg}^c)) + W_2(W_0(F_{max}^c))] \quad (3)$$

其中, σ 为 Sigmoid 激活函数; MLP 为多层感知器; $AvgPool$ 为平均池化; $MaxPool$ 为最大池化; $W_0 \in R^{C \times c}$, $W_1 \in R^{C \times Cr}$ 。

空间注意力结构如图 5 所示,其将 CAM 模块生成的特征图经过最大池化和平均池化处理后,得到 2 个 $H \times W \times 1$ 的特征图。再使用 7×7 的过滤器将 2 个特征进行连接和卷积,合成 1 个 $H \times W \times 1$ 的特征图。最后通过 Sigmoid 激活函数输出空间注意力权重 $M_s(F')$ 。计算公式如下:

$$M_s(F') = \sigma[f^{7 \times 7}(AvgPool(F'); MaxPool(F'))] \quad (4)$$

其中, $M_s(F')$ 为空间注意力模块; $f^{7 \times 7}$ 表示过滤

器大小为 7×7 的卷积运算。

3 结果与分析

3.1 模型训练与评价

3.1.1 试验平台与参数设置 对所提算法进行试验设计,主要采用 PyCharm 集成开发环境与 Python 为基础的开发框架。试验所用硬件设备中,处理器型号为 12th Gen Intel (R) Core (TM) i5- 12400F 2.50 GHz;显卡为 NVIDIA GeForce RTX 4060,内存和硬盘容量分别为 16G 和 1TB。软件开发环境中,使用了 Win 11 操作系统,3.8.3 版本的 Python、2.1.2 版本的 Pytorch 和 12.2 版本的 CUDA。训练基本参数设置如下:训练周期为 100 个 epoch,默认批次大小为 16。

3.1.2 评价标准 使用精确率 P (precision)、召回率 R (recall)和平均精度均值 mAP (mean of average precision)来综合评价模型的性能。具体的计算公式如(5)~(8)所示:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

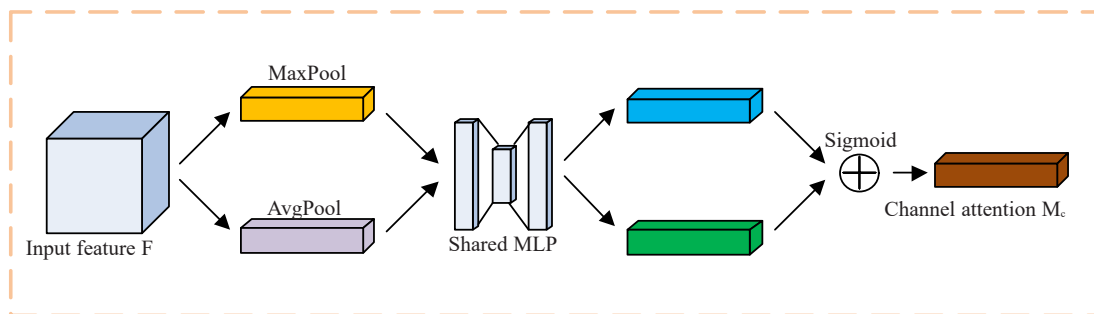


图4 通道注意力结构图
Fig.4 Channel attention structure diagram

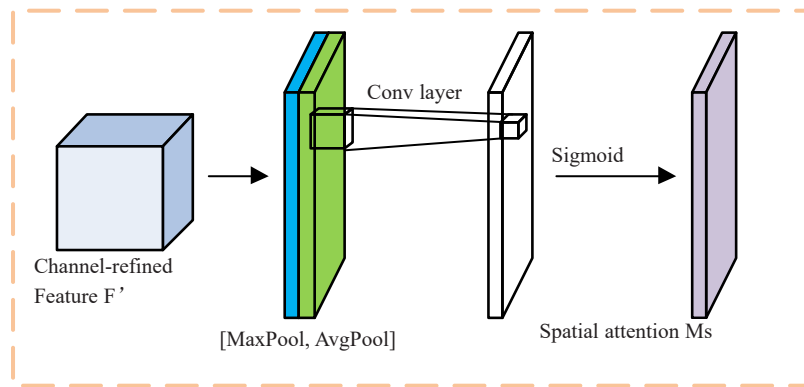


图5 空间注意力结构图

Fig. 5 Spatial attention structure diagram

$$AP = \int_0^1 P(R) dR ; \tag{7}$$

$$mAP = \frac{1}{N} \sum_{k=1}^N AP_k . \tag{8}$$

其中, TP (真正例)表示实际为正确类别“tomato”且被模型正确检测为“tomato”的实例数量; FP (假正例)则是指实际为错误类别或者背景类别“background”但被模型错误地检测为正确类别“tomato”的实例; FN (假负例)代表实际为正确类别“tomato”但被模型错误地检测为错误类别或背景类别“background”的实例; TN (真负例)则是实际为错误类别“background”且被模型正确检测为错误类别“background”的实例数量。

精确率 P 衡量的是模型检测为正样本的实例中真正为正样本的比例,也就是检测为正样本的准确率,精确率越高,说明模型在正确类别中检测的准确性越高;召回率 R 即模型正确识别出的正确类别的数量占实际正确类别总数的比例。召回率越高,表示模型在正确类别的识别中越准确,漏检的情况越少; AP 是 P-R 曲线下的面积; mAP 为所有类别 AP 值的均值, N 表示目标识别的类别个数, k 指待评估的某个类别,只有一个类别,即“tomato”, $N=1$ 。此外,实时性对目标检测也非常重要,因此需要考虑模型的检测速度。

3.2 模型的消融试验

已有相关试验表明 CBAM 模块加在 YOLO 模型的 Neck 层和 Head 层之间其综合评价指标最佳^[25],本消融试验旨在进一步研究加入 CBAM 模块后以及每增加一层 CBAM 模块对整体模型的影响,本试验通过 3 个关键指标来清晰衡量模型的改进效果,分别是参数量、精确率 P 和 mAP 。研究结果如表 1 所示。

表 1 消融试验结果

Table 1 Ablation experiment results

YOLOv8	第 1 层 CBAM First layer CBAM	第 2 层 CBAM Second layer CBAM	第 3 层 CBAM Third layer CBAM	参数量 Parameters/ M	精确率 Precision ratio/%	mAP / %
✓				31.6	86.0	83.0
✓	✓			32.3	86.3	83.5
✓	✓	✓		33.1	88.0	87.0
✓	✓	✓	✓	33.7	91.0	90.0

注: ✓表示加入模块。

Note: ✓ indicate add module.

从表 1 可以看出,在原始 YOLOv8 模型基础上每加入 1 层 CBAM 模块,其参数量平均增加 0.7 M,加入 3 层后,参数量为 33.7 M,增加了 2.1 M;而精确率明显提升,当加入 1 层 CBAM 模块时,提升 0.3 个百分点,加入 2 层和 3 层后迅速提升 1.7 和 3 个百分点,精确率达到了 91.0%;另外, mAP 值也是逐步增加,从原始的 83.0%至 90.0%,最后提升了 7 个百分点。因此, CBAM 模块体积小、可灵活集成到任何 CNN 架构中,在添加了 3 层 CBAM 模块至 YOLOv8 模型后,其参数量只增加了 2.1 M,但对整体模型在精确率和 mAP 的提高非常明显,主要是由于 CBAM 模块可以有效抑制噪声和无关信息,动态调整特征图的权重,例如在小目标番茄的识别中,因这些番茄占据像素数较少,其特征信息极易在深层网络中丢失,而加入 CBAM 模块后,通过增大小目标番茄在整张特征图的权重,促使小目标信息更容易被模型学习,从而增加整体模型的识别精确率。

3.3 不同检测方法的对比试验

为了验证 CBAM-YOLOv8 模型的优越性,将

其与经典机器学习算法(SSD、Faster RCNN)和原始YOLOv8模型进行检测效果对比,如图6展现了部分检测结果。在本研究中,以YOLOv8作为基础模型,首先对其进行了训练以评估其基准性能。接着,在YOLOv8的基础上添加了CBAM注意力机制,并再次在番茄数据集上进行训练。模型训练完

成后,对测试集中的图像进行目标检测性能评估。在本次模型检测结果对比研究中,输入图像大小统一为640*640像素。图6中对检测结果图像进行了裁剪,选择了具有代表性和差异性的部分进行分析。

不同模型在对完整番茄的识别精度上存在明

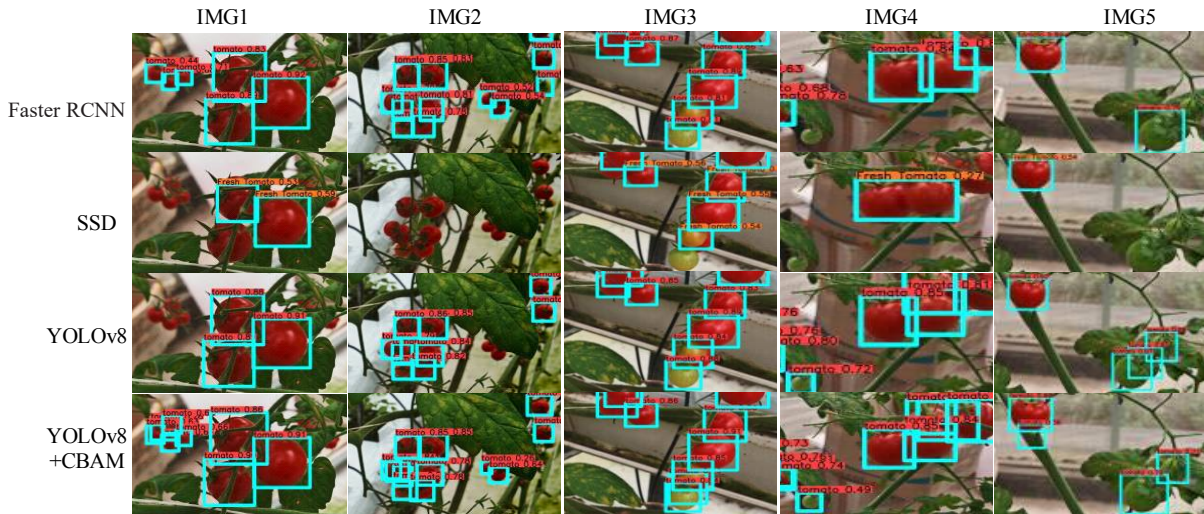


图6 不同算法的番茄识别模型检测结果

Fig. 6 Detection results of tomato recognition model under different algorithms

显的差异。其中,从检测结果图上看,SSD模型表现最为逊色,很多清晰完整的正例没有识别出来,特别是在绿色番茄的识别中,且该模型检测完整番茄的置信度范围较低,仅在0.64~0.73。相比SSD模型,Faster RCNN模型优势凸显,检测的置信度大致稳定在0.8左右。然而,当转向YOLOv8模型时,其检测置信度有了明显提升,达到了0.9左右,显示了较高的识别可靠性。然而,在引入CBAM注意力机制对YOLOv8模型进行改进后,虽然在一定程度上优化了模型的某些特性,但在完整番茄的识别中,其置信度却有所下降,稳定在0.85左右。

对小目标、堆叠等复杂情况的番茄识别中,加入CBAM模块后明显提高识别精度。如图6的IMG1中位于左上角和IMG2中偏右下的模糊、虚化的小目标番茄,采用SSD和YOLOv8均未识别出这些正例,其他2种模型,尤其加入注意力机制后的YOLOv8模型,其置信度更高;再如在IMG3中,由于目标清晰,除SSD外3种模型检测效果都很好,但只有CBAM-YOLOv8模型检测出了堆叠的番茄,这都显示了它们在处理模糊背景和复杂环境时的局限性。当使用YOLOv8模型进行预测时,

其置信度达到了0.71,在YOLOv8的基础上添加了CBAM注意力机制后,其置信度相比YOLOv8提升了约6个百分点。尽管YOLOv8在清晰样本上表现出色,但在处理模糊或细节不足的图像时,其性能会受到影响。

综合来看,SSD模型的识别效果最为逊色,Faster RCNN模型不能很好地识别出绿色番茄。直接使用YOLOv8在预测模糊和复杂背景下的番茄时,会出现较多的错误预测,导致预测到的番茄数量明显多于实际数量。而添加了CBAM注意力机制后,增强模型的特征提取和注意力分配能力,会对特征图权重调整,虽然导致完整番茄置信度降低,但模糊的小目标番茄置信度增高。因此,这种更关注小目标样本的模型,导致预测分布更分散,但漏检减少对实际应用更有价值。

为了进一步验证CBAM-YOLOv8模型的优越性,将其与经典目标检测模型(SSD、Faster RCNN)、YOLO其他系列算法(YOLOv5、YOLOv7)和原始YOLOv8模型在精确率、召回率和*mAP*这3个关键指标上进行比较。由表2可知,SSD模型在3个指标上的得分最低。Faster RCNN模型相较于SSD

模型在3个指标上分别提高了8个百分点、3个百分点和4个百分点。YOLO系列算法在3个指标上均优于二阶段检测算法Faster RCNN,这展现了YOLO这个一阶段目标识别算法的优越性。而添加了CBAM注意力机制的YOLOv8模型在这3个关键指标上的得分最高,精确率、召回率和*mAP*值分别为91%、78%和90%,较原始YOLOv8模型分别提高了5个百分点、3个百分点和7个百分点,这充分证明了CBAM注意力机制在提升模型识别精度和召回率方面的有效性。

表2 不同模型训练结果对比
Table 2 Comparison of training results of different models

模型 Models	精确率 Precision ratio/%	召回率 Recall rate/%	<i>mAP</i> /%
SSD	70	64	75
Faster RCNN	78	67	79
YOLOv5	80	76	78
YOLOv7	86	74	80
YOLOv8	86	75	83
YOLOv8+CBAM	91	78	90

图7是对不同模型的检测速度进行对比分析的结果。对测试集中的45%的图像进行识别检测,逐一记录不同模型对每幅图像的预测时间,并求取平均值。由图7可知,从模型预测时间结果来看,YOLO系列模型的预测时间最短,然后是Faster RCNN和YOLOv8+CBAM,最后是SSD。YOLO通过单阶段端到端的预测框架,直接输出目标位置

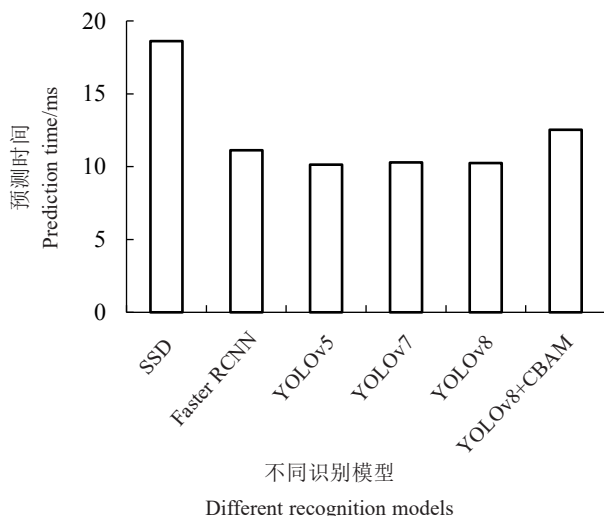


图7 不同模型的预测时间差异
Fig. 7 Differences in prediction time among different models

与类别,避免了区域提议网络产生的冗余计算,从而提升检测速度。这种全局推理的方式使得YOLO在处理复杂场景时能够更快速、更准确地检测目标。并且YOLO的结构简单,在番茄果实识别试验中,由于YOLO采用了单阶段检测的方式,避免了传统两阶段检测算法中的候选区域生成和筛选过程,因此其处理速度更快。两阶段检测算法的推理速度显著低于一阶段YOLO系列模型,且检测精度也不及YOLO算法,因此,在本次试验中,选择将经典目标检测算法作为对照组,以便更全面地评估YOLO算法在番茄果实识别任务中的性能。

在引入注意力机制到YOLO算法后,尽管模型可能在某些方面获得了更高的检测精度,但其检测速度却相应地有所减缓。采用改进CBAM注意力机制的YOLOv8模型在检测速度方面虽略微不及原始模型,但显著降低了小目标的漏检率。

4 结论

在番茄果实识别方面,笔者提出了一种名为CBAM-YOLOv8的模型,该模型在检测精度和推理速度之间取得了良好的平衡,满足农业环境中番茄果实的实时检测需求。通过实际试验得到以下结论。

(1)在YOLOv8模型的Head层加入了CBAM注意力机制,改进后的模型对模糊的、背景复杂的小目标番茄有更好的检测效果。与原始YOLOv8相比,虽耗费一定的计算资源,检测速度略有下降,但提升了检测的精确率。

(2)对改进模型进行消融试验,验证了每添加一层CBAM模块,在精确率和*mAP*值上的有效提升。

(3)CBAM-YOLOv8模型相对于其他经典目标检测模型,其精确率、召回率和*mAP*值分别可达到91%、78%和90%;较原始YOLOv8模型分别提高了5个百分点、3个百分点和7个百分点。

综上所述,采用改进CBAM注意力机制的YOLOv8模型不仅保证了识别过程的精确性,还提高了实时计数的准确率和效率。因此,尽管存在细微的速度差异,但该模型无疑是实现精准、快速番茄识别和实时目标计数、采摘的有力工具,将来可以为其他类似水果目标检测提供宝贵的研究经验。

参考文献

- [1] 刘维孟,卢伟源,尹秉奎,等.番茄采摘机器人数字孪生仿真系统的开发研究[J].南方农机,2023,54(22):5-8.
- [2] 李赟,刘思雨,朱川,等.基于深度学习的水果识别系统设计[J].农机化研究,2023,45(10):187-191.
- [3] 邹伟.基于机器视觉技术的柑橘果实大小分类研究[J].农业与技术,2023,43(15):48-51.
- [4] 丛志文.基于深度学习的番茄识别技术及采摘路径规划研究[D].山东淄博:山东理工大学,2021.
- [5] 杨健,杨啸治,熊串,等.基于改进YOLOv5的番茄果实识别估产方法[J].中国瓜菜,2024,37(6):61-68.
- [6] ZOU Z X, CHEN K Y, SHI Z W, et al. Object detection in 20 years: A Survey[J]. Proceedings of the Ieee, 2023, 111(3): 257-276.
- [7] VIOLA P, JONES M J. Rapid object detection using a boosted cascade of simple features[J]. Computer Vision and Pattern Recognition, 2021(1):1.
- [8] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[J]. Iee Computer Society Conference on Computer Vision, 2005:1063-1069.
- [9] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model[C]//IEEE. 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008.
- [10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//IEEE. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [11] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[J]. Computer Vision, 2016:21-37.
- [12] DUAN K W, BAI S, XIE L X, et al. Centernet: Keypoint triplets for object detection[J]. Computer Vision and Pattern Recognition, 2019:6568-6577.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. Iee Conference on Computer Vision and Pattern Recognition, 2014:580-587.
- [14] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. Computer Vision and Pattern Recognition, 2014, 37: 1904-1916.
- [15] LIN T Y, DOLLAR P, GIRSHICK R B, et al. Feature pyramid networks for object detection[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017:936-944.
- [16] HUANG Z C, WANG J L, FU X S, et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection[J]. Information Sciences, 2020, 522:241-258.
- [17] 刘芳,刘玉坤,林森,等.基于改进型YOLO的复杂环境下番茄果实快速识别方法[J].农业机械学报,2020,51(6):229-237.
- [18] ZHANG C X, KANG F, WANG Y X. An improved apple object detection method based on lightweight YOLOv4 in complex backgrounds[J]. Remote Sensing, 2022, 14(17):4150.
- [19] 李丽,梁继元,张云峰,等.基于改进YOLO v5的复杂环境下柑橘目标的精准检测与定位方法[J].农业机械学报,2024, 55(8):208-290.
- [20] 王杨,李迎春,许佳炜,等.基于改进Vision Transformer网络的农作物病害识别方法[J].小型微型计算机系统,2024, 45(4):887-893.
- [21] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023, 5(4):1680-1716.
- [22] JOCHER G, CHAURASIA A, STOKEN A, et al. Ultralytics/YOLOv5: v6.2-YOLOv5 classification models, apple ml, reproducibility, clearml and deci. ai integrations[J]. Zenodo, 2022.
- [23] 王欢欢,李舒晴.基于注意力机制的三维模型特征提取[J].现代计算机,2024,30(4):48-52.
- [24] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Springer, Cham. Proceedings of the European Conference on Computer Vision (ECCV), 2018:3-19.
- [25] 杨巧梅,崔婷婷,袁永榜,等.轻量化YOLO模型在农作物微小病虫害检测中的应用研究[J].中国农机化学报,2024, 45(9):265-270.