

基于改进 YOLOX 的自然环境下辣椒果实检测方法

李 旭^{1,2}, 刘 青¹, 匡敏球¹, 潘建东³, 刘大为^{1,2*},
向 阳^{1,2}, 吴艳华⁴, 谢方平^{1,2}

(1. 湖南农业大学机电工程学院, 长沙 410128; 2. 智能农机装备湖南省重点实验室, 长沙 410128; 3. 浙江大学生物系统工程与食品科学学院, 杭州 310000; 4. 长沙县多莱农业科技有限公司, 长沙 410100)

摘 要: 针对不同光照、枝叶遮挡和果实遮挡条件下模型适应能力差和检测精度较低的问题, 该研究提出了一种基于 YOLOX 的改进辣椒果实检测模型 YOLOX_Pepper。首先, 在 YOLOX 特征融合网络中添加融合高效通道 CA (coordinate attention) 注意力机制, 提升不同光照条件下模型捕捉辣椒果实关键特征的能力; 其次, 将主干网络特征聚合模块中的卷积模块替换为可变形卷积 DCNv2 (deformable convnets v2), 提升了模型对不同遮挡情况下辣椒多样几何特征的感知能力。试验结果表明, 改进的 YOLOX_Pepper 模型平均检测精度为 93.30%, 与 Faster R-CNN、YOLOv5、YOLOv7 以及 YOLOX 相比, 分别提高了 3.99、1.58、3.19 和 2.84 个百分点, F1 分数为 96%, 单张图片检测平均用时 0.026 s。改进的 YOLOX_Pepper 模型对自然环境不同光照和遮挡条件的辣椒果实均能进行准确快速的检测。该方法可为辣椒智能化生产提供技术基础。

关键词: 农业; 机器视觉; YOLOX; 辣椒; 自然环境; 果实检测

doi: 10.11975/j.issn.1002-6819.202405175

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2024)-21-0119-08

李旭, 刘青, 匡敏球, 等. 基于改进 YOLOX 的自然环境下辣椒果实检测方法[J]. 农业工程学报, 2024, 40(21): 119-126. doi: 10.11975/j.issn.1002-6819.202405175 <http://www.tcsae.org>

LI Xu, LIU Qing, KUANG Minqiu, et al. Detecting chili pepper fruits in a natural environment using improved YOLOX[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(21): 119-126. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202405175 <http://www.tcsae.org>

0 引 言

辣椒在中国种植广泛, 西起西藏, 东至上海, 南起海南岛, 北至黑龙江, 均能种植, 是中国种植范围最广的蔬菜之一^[1]。当前, 鲜食辣椒的生产面临着劳动强度大、效率低下的问题^[2-6], 为应对这些问题, 辣椒生产正朝着机械化和智能化生产的方向转变^[7-12]。在实现智能化生产的过程中, 对辣椒果实的精准检测成为首要任务^[13-15]。

目前, 国内外研究人员应用深度学习针对自然环境下的果实检测开展了一系列研究。围绕二阶段目标检测算法, HABARAGAMUWA 等^[16]使用深度卷积神经网络 DCNN (deep convolutional neural networks) 对温室中的草莓进行分类识别, 将其分为成熟和未成熟两类, 研究结果显示, DCNN 在成熟和未成熟草莓的分类准确率分别为 88.03% 和 77.21%, 单张图像平均检测时间为 0.22 s, 该研究未对草莓遮挡情况进行研究分析。LIU Y 等^[17]提出了一种改进的 Faster R-CNN 方法, 用于多簇绿柿子的

识别, 该方法利用 DetNet 作为主干特征提取网络, 并引入了加权 ECA 通道注意力机制, 试验结果显示, 改进后的 Faster R-CNN 模型在多簇绿柿子识别任务中 mAP (mean average precision) 为 98.4%, 比传统模型提高了 11.8 个百分点, 单张图像平均检测时间为 0.54 s, 该研究未对自然环境下多簇绿柿子的光照和遮挡进行详细研究, 且检测速度较慢。傅隆生等^[18]提出了一种基于 LeNet 卷积神经网络的深度学习模型, 用于多簇猕猴桃果实图像的识别, 该研究针对遮挡果实的识别提供了思路和方法, 试验结果显示, 该模型的整体准确率为 89.29%, 单张图像平均检测时间为 0.27 s, 在遮挡果、重叠果、相邻果和分离果的识别方面, mAP 分别为 78.97%、83.11%、91.01% 和 94.78%。

围绕一阶段目标检测算法, LI X 等^[19]提出了一种基于 YOLOv4-tiny 的自适应空间特征金字塔方法, 结合 CBAM 注意力机制和多尺度预测, 提升了遮挡和小目标青椒的检测效果。研究显示该方法的 mAP 为 95.11%, 准确率 96.91%, 召回率 93.85%, 单张图像检测时间为 11.24 ms。苗荣慧等^[20]改进了 YOLOv7 模型, 将 MobileNetV3 作为骨干网络并加入全局注意力机制用于樱桃成熟度检测, 改进后模型的准确率、召回率和 mAP 分别为 98.6%、98.1% 和 98.2%, 平均检测时间为 82 ms。熊俊涛等^[21]基于 YOLOv5-Lite 模型改进了 Focus 层, 引入 ShuffleNetv2, 并将中间的 1×1 组卷积改为普通卷积, 以减少深层神经

收稿日期: 2024-05-24 修订日期: 2024-09-28

基金项目: 国家重点研发计划项目 (2021YFD1600300-4/06; 2022YFD2002003-3)

作者简介: 李旭, 博士, 副教授, 研究方向为农业装备智能控制技术。

Email: leexu@hunau.edu.cn

*通信作者: 刘大为, 博士, 副教授, 研究方向为蔬菜生产全程机械化技术与装备。Email: liudawei8361@163.com

网络块中 C3 层的数量,并在 FPN 和 PAN 中设置了相同的通道数,该模型木瓜检测 mAP 为 92.4%,检测时间为 7 ms,内存占用 11.3MB。

综合分析,二阶目标检测算法如 DCNN、Faster RCNN、Mask RCNN 等在精度方面表现较好,适用于物体精细识别;一阶目标检测算法如 YOLOv4-tiny、YOLOv5-Lite、YOLOv7 等则具有较快的检测速度,适用实时应用场景。上述研究在果实检测的精度和速度上取得了一定成果,但大多数试验研究未考虑自然环境下不同光照和遮挡条件对模型检测精度的影响。因此,本文以东山光皮辣椒为检测对象,对 YOLOX 模型进行了改进,引入融合高效通道 CA 注意力机制和可变形卷积 DCNv2,前者有助于提高不同光照条件下模型捕捉辣椒果实关键特征的能力,后者提高因遮挡导致辣椒长度、宽度和长宽比几何特征呈现多样性情况下模型的感知能力,进而增强了模型的鲁棒性和适用性,提升了模型对辣椒果实的特征提取效率和检测精度,为辣椒的智能化生产提供技术支撑。

1 材料与方法

1.1 数据集建立

东山光皮辣椒起源于湖南长沙县黄兴镇、黎托乡等沿河两岸,具有皮滑肉厚、色泽艳丽有光泽、微辣带甜和耐贮运等特点,是优质的地方辣椒品种。试验用辣椒果实图像分别于 2023 年 7-9 月在湖南省长沙市长沙县黄兴镇多莱农业科技有限公司东山光皮辣椒种植基地采集,采集设备为佳能 M100 相机,图像分辨率为 4 288×2 848 像素。由于自然环境光照的复杂性和辣椒果实的高遮挡,为提升模型泛化能力,采集过程中选择了不同的时间、光照条件、拍摄角度等,增加了目标数据的多样性,共拍摄照片 1 069 张,图 1 为自然条件下采集的部分东山光皮辣椒果实图像。



图 1 辣椒果实图像

Fig.1 Image of pepper fruit

用 Labellmg 软件对图像中的青椒和红椒分别进行标记,青椒标记为 green pepper,红椒标记为 pepper。为了防止数据扩展后的数据泄露,原始数据集以 8:2 的比例分为数据扩展集和测试集,测试集将用于后续模型效果测试。

针对东山光皮辣椒果实呈子弹形、分布密集、枝叶遮挡严重等特点,对数据扩展集的辣椒图片进行了数据增强处理。其中 30% 的照片引入高斯噪声,以提高模型的鲁棒性和泛化能力;20% 的照片调整了 HSV 通道,调高模型对不同色调、饱和度、亮度的适应能力;10% 的照片进行了模糊处理,以消除噪声、平滑图像或减少细节;40% 的照片进行了大小、剪切和旋转处理,增加图像的多样性。通过图像增强,数据扩展集照片增加为 3 189 张,将图像按照 8:2 的比例分为训练集和验证集,图 2 为不同数据增强方式处理后的图像实例。

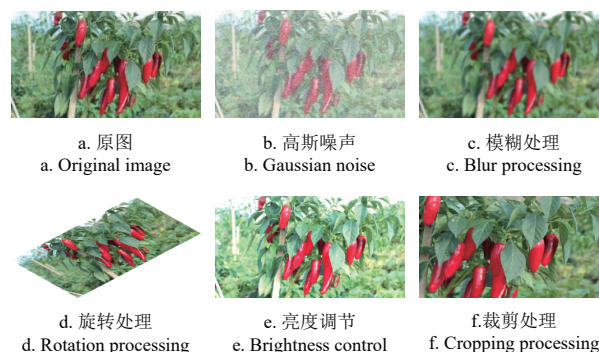


图 2 辣椒果实数据增强

Fig.2 Enhancement of fruit data of peppers

1.2 辣椒果实目标检测方法

1.2.1 YOLOX 目标检测模型

YOLOX 模型框架主要由输入端、主干网络、特征融合和预测头 4 个模块组成^[22]。YOLOX 基于 CSPDarknet 主干网络,结合 Focus 模块、SiLU 激活函数和 CSPLayer,提升了特征提取与融合能力,从而增强模型性能。通过特征金字塔网络(FPN),YOLOX 融合多尺度特征,适应不同大小目标的检测能力^[23]。网格点操作和解码机制提升了模型对目标位置和尺度的精确度。解耦检测头分别预测置信度和回归框,进一步提升了分类和回归的处理能力以及检测精度。因此,该研究选择对 YOLOX 进行模型的设计与改进。

1.2.2 模型的改进

为了提升对辣椒果实的特征提取效率和检测精度,在 YOLOX 模型中特征提取网络的 3 个输出结果处添加融合高效通道 CA 注意力机制,并将主干网络特征聚合模块中的卷积模块替换为可变形卷积 DCNv2,图 3 为改进后的 YOLOX_Pepper 模型网络结构。

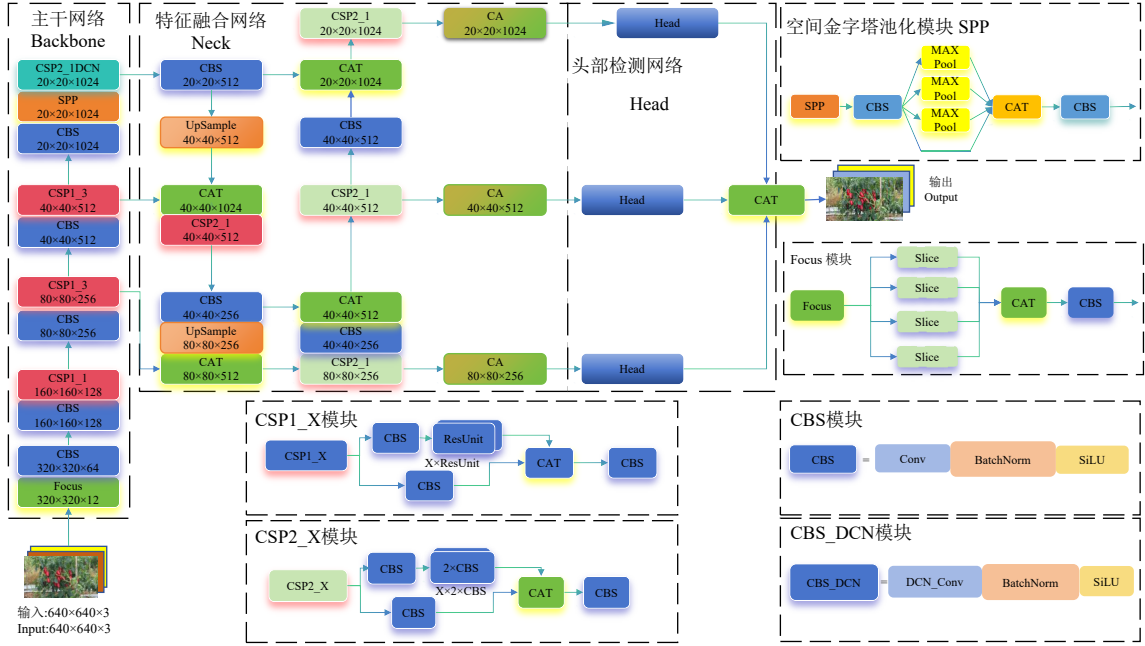
1) 注意力机制

东山光皮辣椒果实簇生,受生长环境和成熟度等因素的影响,果实外形和颜色复杂多变。准确获取辣椒果实的空间信息对于提高检测能力至关重要。本研究在 YOLOX 模型中添加了 CA 注意力机制,CA 注意力机制能够提高神经网络对空间信息的感知能力,根据辣椒果实空间位置,动态调整特征图中不同位置的重要性,使神经网络能够更专注于关键区域。通过在特征融合层末端添加 CA 注意力机制模块,提升模型对辣椒果实关键特征的捕捉能力,提高模型在不同光照条件

下的检测精度。

CA 注意力机制结构如图 4 所示^[24], CA 注意力机制的基本思想是根据输入特征图中每个位置的坐标信息,

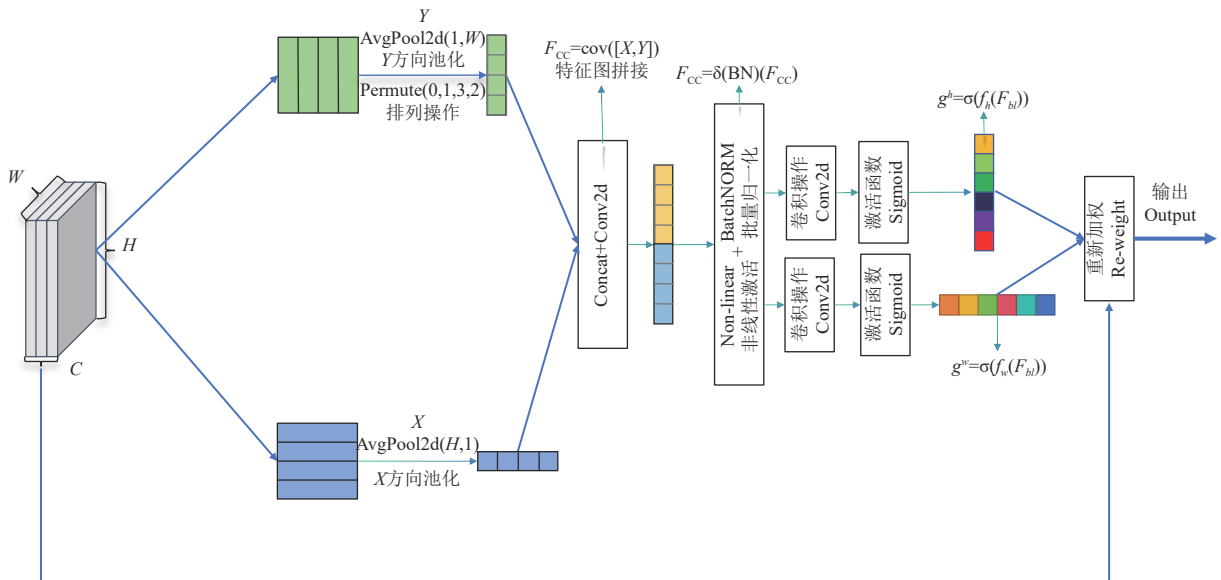
动态地调整特征图中不同位置的权重, 从而使网络能够更好地利用空间位置信息^[25], CA 注意力机制的数学表达式如式 (1) 所示:



注: Maxpool 为最大池化操作, Focus 模块为加速数据处理, CBS 模块为 Conv+BN+SiLU 激活函数组成模块, CBS_DCN 模块为 DCNv2Conv+BN+SiLU 激活函数组成模块, UpSample 模块为上采样操作, SPP 为空间金字塔池化模块, CSP1_X 和 CSP2_X 模块为跨阶段局部网络的两种结构, CA 为融合高效通道注意力机制。

Note: Maxpool refers to the max pooling operation; the Focus module is designed to accelerate data processing. The CBS module consists of Conv + BN + SiLU activation functions, while the CBS_DCN module includes DCNv2Conv + BN + SiLU activation functions. The UpSample module is responsible for upsampling operations, and SPP stands for the spatial pyramid pooling module. CSP1_X and CSP2_X refer to two structures of the cross stage partial network, and CA integrates the efficient channel attention mechanism.

图 3 YOLOX_Pepper 模型网络结构
Fig.3 YOLOX_Pepper model network structure



注: C 是通道数, H 是特征图的高度, W 是宽度, $\text{AvgPool2d}(1, W)$ 为 Y 方向池化, $\text{AvgPool2d}(H, 1)$ 为 X 方向池化, Permute 为排列操作, $\text{Concat} + \text{Conv2d}$ 为特征图拼接, $\text{BatchNorm} + \text{Non-linear}$ 为批量归一化和非线性激活, Conv2d 为卷积操作, Sigmoid 为激活函数, Re-weighting 为重新加权。 $g^h = \sigma(f_h(F_{h0}))$ 为生成特征图在 H 方向上的注意力权重, $g^w = \sigma(f_w(F_{w0}))$ 为生成特征图在 W 方向上的注意力权重。

Note: C is the number of channels, H is the height of the feature map, W is the width, $\text{AvgPool2d}(1, W)$ is Y -direction pooling, $\text{AvgPool2d}(H, 1)$ is X -direction pooling, Permute is the permutation operation, $\text{Concat} + \text{Conv2d}$ is the feature map stitching, $\text{BatchNorm} + \text{Non-linear}$ is the BatchNormal and Non-linear activation, Conv2d is convolution operation, Sigmoid is the activation function, Re-weighting is re-weighting, $g^h = \sigma(f_h(F_{h0}))$ is the attention weight of the generated feature map in the H direction, $g^w = \sigma(f_w(F_{w0}))$ is the attention weight of the generated feature map in the W direction.

图 4 CA 注意力机制结构
Fig.4 Structure of the coordinate attention (CA) mechanism

$$y_i = \sum_{j=1}^{H \times W} \frac{e^{f(x_j)}}{\sum_{k=1}^{H \times W} e^{f(x_k)}} x_j \quad (1)$$

式中 y_i 表示输出特征图中的第 i 个位置的值, x_j 表示输入特征图中的第 j 个位置的值, x_k 表示输入特征图中的第 k 个位置的值, $f(x_k)$ 为映射函数, 用于计算每个位置的注意力分数。该映射函数通常是一个全连接网络或者是一个卷积神经网络, 用于学习每个位置的权重。 $H \times W$ 表示输入特征图的高度和宽度, j 是特征图中的位置索引。

2) 可变形卷积

由于枝叶遮挡和果实遮挡, 辣椒果实的长度、宽度和长宽比呈现多样性, 原始 YOLOX 使用普通卷积操作, 卷积层在特征提取时不能适应辣椒果实多样几何特征导致的不同位置区域感受野尺寸的变化, 削弱了网络模型对辣椒果实特征的提取能力。针对该问题, 将 YOLOX 主干网络特征聚合模块中的卷积模块替换为可变形卷积 DCNv2, 其结构如图 5 所示。DCNv2 的基本思想是在卷积层中引入可变形卷积操作, 以捕获输入特征图中物体的变形信息^[26-29]。相比较传统的固定形状卷积核, 可变形卷积核具有可学习的偏移量, 能够根据输入特征图中物体的形变情况自适应地调整卷积核的形状和大小, 能更好地适应因遮挡导致的不同长度、宽度和长宽比的辣椒果实, 提高模型对几何特征多样的辣椒果实特征的提取能力。传统卷积和可变形卷积调整卷积核效果对比如图 6 所示, 可变形卷积引入了可学习的偏移量, 使得卷积核能够根据输入数据的不同形状和空间位置进行自适应变形。保持相同的感受野大小和卷积计算的有效性。可变形卷积的参数比传统卷积更加紧凑, 并且卷积核的形状可以根据输入特征自适应调整, 因此更适合处理不同形状和位置的特征。

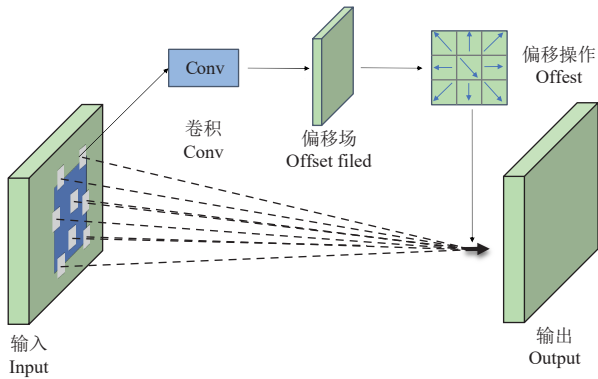


图 5 可变形卷积原理结构

Fig.5 Principle structure of deformable convolution

DCNv2 的数学表达式如式 (2) 所示:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \times x(p_0 + p_n + \Delta p_n) \quad (2)$$

式中 $y(p_0)$ 表示输出特征图中的像素位置 p_0 的值, x 表示输入特征图, w 表示采样位置的权重, R 表示卷积核的区域, p_n 表示在 R 中相对于 p_0 的相对位置, Δp_n 表示偏移量。

DCNv2 的关键之处在于 Δp_n 偏移量的引入, 它允许卷积核在空间上进行动态调整, 从而捕获目标对象的空

间变化信息。通常情况下, 偏移量 Δp_n 是一个小数, 这导致了输入特征图上像素位置 x 的值无法直接获得, 需要通过双线性插值估计。双线性插值如式 (3) 所示:

$$G(q, p) = \prod_{i=0}^1 (1 - |q_i - p_i|) \quad (3)$$

式中 $G(q, p)$ 是双线性插值核函数, q 表示特征图映射中的所有积分空间位置, p 表示需要估计的像素位置, q_i 表示输入特征图的第 i 个维度中的整数网格位置, p_i 表示第 i 维度中需要估计的像素位置。通过双线性插值可变形卷积能够根据偏移量动态调整卷积核的采样位置, 从而更准确地捕获辣椒果实的空间变化信息, 针对辣椒的几何特征进行特征优化, 提高了模型对目标的感知能力和检测精度。

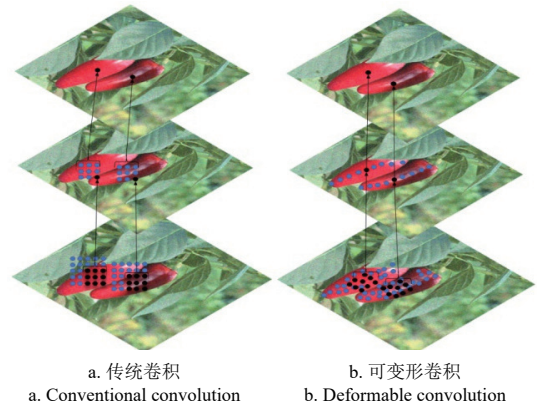


图 6 两种卷积方法效果对比

Fig.6 Comparison of the effect of the two convolution methods

1.3 试验方法

1.3.1 试验平台与参数设置

本文试验平台操作系统为 64 位 Windows 10, 内存为 16G, 显卡驱动为 NVIDIA GeForce RTX 3080TI, 显存为 12G, 搭载 AMD Ryzen 7 5800X 处理器, 主频为 3.8 GHz, 采用 Pytorch 1.7 版本的深度学习框架, CUDA 版本为 11.6, Python 版本为 3.8, 编译器为 Pycharm 2023.1.5。模型输入尺寸为 640×640 像素, 训练批次设置为 16, 初始学习率设为 0.01, 学习退火参数为 0.1, 模型学习率从 0.01 呈余弦变化衰退至 0.001, 采用随机梯度下降法优化模型参数, 迭代次数为 300, 对东山光皮辣椒果实检测模型的训练和测试均在同一环境下进行。

1.3.2 模型评价指标

本试验采用目标检测中常用的准确率 P (precision, %)、召回率 R (recall, %)、F1 分数 (F1 score)、平均精确度 AP (average precision, %)、平均精度均值 mAP (mean average precision, %)、参数量 (params, 10^6 M) 作为模型的评价指标。

2 结果与分析

2.1 消融试验

为了验证模型改进的有效性, 结合测试集开展消融试验^[11], 在 YOLOX 模型的基础上逐步改进, 以验证每一步改进的有效性, 试验结果如表 1 所示。从表 1 可以看出添加 CA 注意力机制, 相比 YOLOX 模型, mAP 和召回率分别提高了 1.46 和 1.51 个百分点, 说明 CA 注意力

机制增强了模型对空间信息的感知能力，使得神经网络更加专注于关键区域，提升模型对辣椒果实关键特征的捕捉能力，进而提升模型的检测精度；引入可变形卷积 DCNv2，相比 YOLOX 模型，mAP、召回率和 F1 分数分别提升 1.2、1.54 和 1 百分点，说明可变形卷积 DCNv2 通过自适应调整卷积核的形状和大小，使模型更好地适应因遮挡导致的不同长度、宽度和长宽比的辣椒果实，

提高了模型对辣椒果实多样的几何特征感知能力和检测精度；在添加 CA 注意力机制的同时引入可变形卷积 DCNv2 的 YOLOX_Pepper 模型在综合性能上表现最优，mAP 为 93.30%，召回率为 93.31%，相较于 YOLOX 分别提高了 2.84、2.77 个百分点，对红椒和绿椒的检测精度也分别提升了 1.47 和 4.21 个百分点。综合消融试验，证明了模型改进的有效性。

表 1 消融试验
Table 1 Ablation test

CA (coordinate attention) 注意力机制	可变形卷积 DCNv2	平均精度均值 Mean average precision mAP/%	召回率 Recall R/%	准确率 Precision P/%	参数量 Params(P_s)/ $\times 10^6$ M	F1 分数 F1 score/%	检测精度	
							Average precision AP/%	
							红椒 Red pepper	绿椒 Green pepper
—	—	90.46	90.54	99.48	8.94	95	93.48	87.44
√	—	91.92	92.05	98.75	8.97	95	94.39	89.45
—	√	91.66	92.08	95.55	9.07	94	94.02	89.30
√	√	93.30	93.31	98.51	9.11	96	94.95	91.65

注：“—”表示不使用此模块；“√”表示使用此模块。
Note: “—” indicates that this module is not used; “√” indicates the use of this module

2.2 不同检测模型性能对比

为验证模型对东山光皮辣椒果实的检测能力，本研究选择了主流目标检测模型 Faster R-CNN、YOLOv5、YOLOv7、YOLOX 和 YOLOX_Pepper 进行对比试验，结果如表 2 所示。YOLOX_Pepper 在各项指标上表现最佳，mAP 为 93.30%，相较于 Faster R-CNN、YOLOv5、YOLOv7 和 YOLOX 分别提升了 3.99、1.58、3.19 和 2.84 个百分点。

表 2 不同检测模型性能对比

Table 2 Performance comparison of different detection models

模型 Model	mAP/%	R/%	P/%	P_s / $\times 10^6$ M	F1 /%	AP/%	
						红椒 Red pepper	绿椒 Green pepper
Faster R-CNN	89.31	92.73	79.68	136.71	86	95.93	82.70
YOLOv5	91.72	91.97	98.60	46.56	95	92.71	90.73
YOLOv7	90.11	88.28	97.60	37.62	92	91.97	88.05
YOLOX	90.46	90.54	99.48	8.94	95	93.48	87.44
YOLOX_Pepper	93.30	93.31	98.51	9.11	96	94.95	91.65

YOLOX_Pepper 的 F1 分数达到 96%，高于其他模型。YOLOX_Pepper 的模型参数量仅为 9.11×10^6 M，比 YOLOv5、YOLOv7 和 Faster R-CNN 模型参数量大幅度减少。在红椒和绿椒的检测精度上，YOLOX_Pepper 分别达到 94.95% 和 91.65%。综合不同检测模型性能对比，YOLOX_Pepper 模型具有更高的精度和效率，Faster R-CNN 模型参数量远超其他模型，不适合实时检测移动端部署。

2.3 不同场景下检测结果分析

为了验证不同场景下 YOLOv5、YOLOv7、YOLOX、YOLOX_Pepper 模型对辣椒果实的检测性能，在采集的原始图像中选取 130 张不同光照条件和 116 张不同遮挡条件的辣椒果实图像进行测试。

不同光照条件具体为 41 张强光图像，41 张阴影图像和 48 张光照不足的图像，三种光照条件的图像中绿椒的数量分别为 120、154 和 111，红椒的数量分别为 339、282 和 412。不同遮挡条件具体为 58 张果实遮挡图像，58 张枝叶遮挡图像，两种遮挡条件的图像中绿椒的数量

分别为 83 和 90，红椒数量分别为 682 和 429。

2.3.1 不同光照条件下检测结果分析

辣椒果实不同光照条件下的检测结果如表 3 所示，在强光条件下，YOLOX_Pepper 模型绿椒和红椒平均检测精度分别为 69.16% 和 89.67%，绿椒和红椒正确检测数分别为 83 和 304 个；在阴影条件下，YOLOX_Pepper 模型绿椒和红椒平均检测精度分别为 77.21% 和 90.42%，绿椒和红椒正确检测数分别为 119 和 255 个；在光照不足条件下，YOLOX_Pepper 模型绿椒和红椒平均检测精度分别为 77.38% 和 75.47%，绿椒和红椒正确检测数分别为 86 和 311 个。

表 3 辣椒果实不同光照条件下的检测结果

Table 3 Detection results of pepper fruits under different light condition

场景 Scenes	模型 Model	种类 Type	mAP/%	总数 Sum/个	检测数目 Number of detection/ 个	正确检测 Number of correct detection/ 个	误检数目 Number of misdeteci on /个
强光 Strong light	YOLOv5	绿椒	68.33	120	82	82	0
		红椒	87.50	339	302	298	4
	YOLOv7	绿椒	60.75	120	74	73	1
		红椒	85.82	339	293	291	2
	YOLOX	绿椒	67.50	120	81	81	0
		红椒	88.79	339	302	301	1
阴影 Shadow	YOLOX_Pepper	绿椒	69.16	120	84	83	1
		红椒	89.67	339	306	304	2
	YOLOv5	绿椒	77.40	154	123	120	3
		红椒	88.90	282	253	251	2
	YOLOv7	绿椒	76.19	154	121	118	3
		红椒	85.35	282	243	241	2
光照不 足 Lack of light	YOLOX	绿椒	77.35	154	122	119	3
		红椒	90.36	282	259	255	4
	YOLOX_Pepper	绿椒	77.21	154	124	119	5
		红椒	90.42	282	261	255	6
	YOLOv5	绿椒	77.40	111	87	86	1
		红椒	72.65	412	303	300	3
	YOLOv7	绿椒	75.31	111	87	84	3
		红椒	72.48	412	301	299	2
	YOLOX	绿椒	76.57	111	87	85	2
		红椒	75.44	412	313	310	3
	YOLOX_Pepper	绿椒	77.38	111	90	86	4
		红椒	75.47	412	314	311	3

YOLOX_Pepper 模型在各种光照条件下均表现出一一定优势, 尤其是在检测数目和检测精度上面优于 YOLOV5、YOLOV7 和 YOLOX 模型。不同光照条件下各模型对辣椒果实实际检测效果如图 7 所示。

2.3.2 不同遮挡条件下检测结果分析

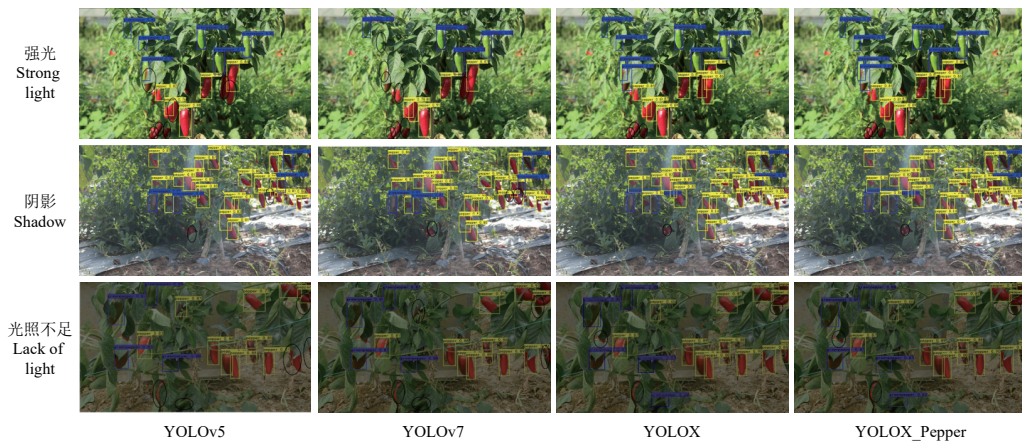
辣椒果实不同遮挡条件下的检测结果如表 4 所示, 在果实遮挡条件下, YOLOX_Pepper 的绿椒和红椒平均检测精度分别为 71.15% 和 94.87%, 绿椒和红椒正确检测数分别为 79 和 650 个; 在枝叶遮挡条件下, YOLOX_Pepper 的绿椒和红椒平均检测精度分别为 83.98% 和 87.10%, 绿椒和红椒正确检测数分别为 88 和 394 个。

在不同遮挡条件下, 改进后的 YOLOX_Pepper 模型在辣椒果实检测方面表现出色, 相比 YOLOv5、YOLOv7 和 YOLOX 对辣椒果实检测效果更好, 在复杂环境中, YOLOX_Pepper 模型展现出优秀的检测性能, 证明了改进模块的有效性, 为辣椒的智能化生产提供了可靠的技术支持。不同遮挡条件下各模型对辣椒果实实际检测效果如图 8 所示。

表 4 辣椒果实不同遮挡条件下的检测结果

Table 4 Detection of pepper fruits under different occlusion conditions

场景 Scenes	模型 Model	种类 Type	mAP/%	总数 Sum/个	检测数目 Number of detection/ 个	正确检测 Number of correct detection/ 个	误检数目 Number of misdeteci on/个
果实遮挡 Fruit occlusion	YOLOv5	绿椒	70.15	83	105	75	30
		红椒	91.07	682	639	624	15
	YOLOv7	绿椒	65.98	83	101	74	27
		红椒	90.06	682	631	617	14
	YOLOX	绿椒	67.12	83	104	76	28
		红椒	92.78	682	650	635	15
	YOLOX_P epper	绿椒	71.15	83	115	79	36
		红椒	94.87	682	672	650	22
枝叶遮挡 Branch and foliage occlusion	YOLOv5	绿椒	70.38	90	129	75	54
		红椒	81.78	429	439	369	70
	YOLOv7	绿椒	72.52	90	126	76	50
		红椒	80.61	429	431	372	59
	YOLOX	绿椒	73.28	90	126	75	51
		红椒	82.90	429	447	375	72
	YOLOX_P epper	绿椒	83.98	90	128	88	40
		红椒	87.10	429	441	394	47



注: 黄色矩形框表示检测出的红椒, 蓝色矩形框表示检测出的绿椒, 黑色椭圆框表示未检测出的辣椒, 下同。

Note: Yellow rectangular boxes indicate the detected red peppers, blue rectangular boxes indicate the detected green peppers, and black oval boxes indicate the non-detected peppers, the same as below.

图 7 不同光照条件下各模型对辣椒果实实际检测效果

Fig.7 Actual detection performance of each model for pepper fruits under different light conditions

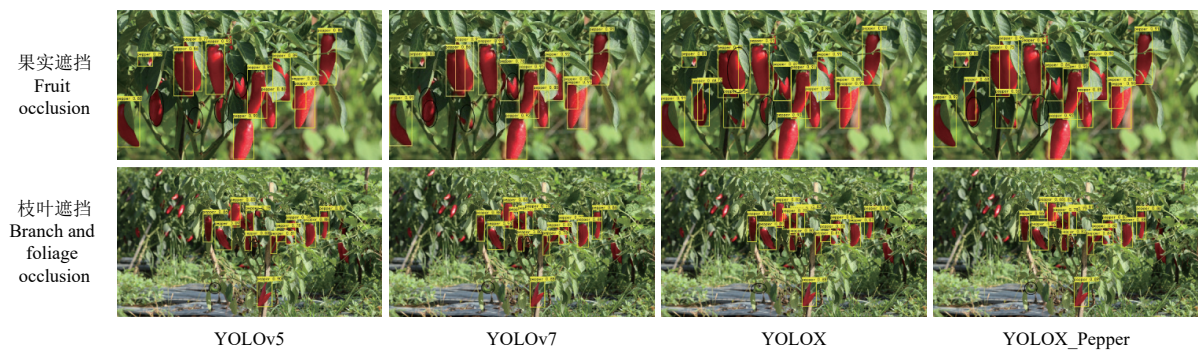


图 8 不同遮挡条件下各模型对辣椒果实实际检测效果

Fig.8 Actual detection performance of each model for pepper fruits under different occlusion conditions

3 结论

1) 提出了基于改进 YOLOX 的自然环境下辣椒果实检测模型 YOLOX_Pepper, 通过添加 CA 注意力机制和

可变形卷积 DCNv2, 提高了模型对辣椒果实特征的捕捉能力。改进的模型 mAP 和召回率分别为 93.30% 和 93.31%, 相较于 YOLOX 分别提高了 2.84 和 2.77 个百分点, 对红

椒和青椒的检测精度也分别提升了 1.47 和 4.21 个百分点。

2) 与其他主流目标检测模型进行试验对比, YOLOX_Pepper 平均检测精度为 93.30%, 相较于 Faster-RCNN、YOLOv5、YOLOv7 和 YOLOX, 分别提高了 3.99、1.58、3.19 和 2.84 个百分点, 改进的模型 F1 分数为 96%, 参数量为 9.11×10^6 M, 单张图片检测平均用时为 0.026 s, 模型具有较好的检测能力。

3) 不同光照条件下, YOLOX_Pepper 模型对青椒的平均检测精度分别为 69.16%、77.21% 和 77.38%, 对红椒的平均检测精度分别为 89.67%、90.42% 和 75.47%; 不同遮挡条件下, YOLOX_Pepper 模型对青椒的平均检测精度分别为 71.15% 和 83.98%, 对红椒的平均检测精度分别为 94.87% 和 87.10%。相较于 YOLOv5、YOLOv7 和 YOLOX 均具有一定优势, 模型的鲁棒性更好, 对辣椒果实检测的准确性更高。

[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): 1715-1726.
ZOU Xuexiao, MA Yanqing, DAI Xiongze, et al. Spread and industry development of pepper in China. Acta Horticulturae Sinica, 2020, 47(9): 1715-1726. (in Chinese with English abstract)
- [2] KASHYAP P K, KUMAR S, JAISWAL A, et al. Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network[J]. IEEE Sensors Journal, 2021, 21(16): 17479-17491.
- [3] 胡仕林, 陈伟, 张境锋, 等. 基于改进 YOLO v5 的苹果采摘机器人目标检测方法[J]. 农机化研究, 2024, 46(6): 48-55.
HU Shilin, CHEN Wei, ZHANG Jingfeng, et al. Target detection method of apple harvesting robot based on improved YOLO v5[J]. Journal of Agricultural Mechanization Research, 2024, 46(6): 48-55. (in Chinese with English abstract)
- [4] MIRHAJI H, SOLEYMANI M, ASAKEREH A, et al. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions[J]. Computers and Electronics in Agriculture, 2021, 191: 106533.
- [5] TIAN H, WANG T, LIU Y, et al. Computer vision technology in agricultural automation: A review[J]. Information Processing in Agriculture, 2020, 7(1): 1-19.
- [6] 李华, 曹卫彬, 李树峰, 等. 辣椒穴盘苗自动取苗机构运动学分析与试验[J]. 农业工程学报, 2015, 31(23): 20-27.
LI Hua, CAO Weibin, LI Shufeng, et al. Kinematic analysis and test on automatic pick-up mechanism for chili plug seedling[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(23): 20-27. (in Chinese with English abstract)
- [7] ARAD B, BALENDONCK J, BARTH R, et al. Development of a sweet pepper harvesting robot[J]. Journal of Field Robotics, 2020, 37(6): 1027-1039.
- [8] CHEN Y, XU H, ZHANG X, et al. An object detection method for bayberry trees based on an improved YOLO algorithm[J]. International Journal of Digital Earth, 2023, 16(1): 781-805.
- [9] POLIC M, TABAK J, ORSAG M. Pepper to fall: A perception method for sweet pepper robotic harvesting[J]. Intelligent Service Robotics, 2022, 15(2): 193-201.
- [10] 孙俊, 陈义德, 周鑫, 等. 快速精准识别棚内草莓的改进 YOLOv4-Tiny 模型[J]. 农业工程学报, 2022, 38(18): 195-203.
SUN Jun, CHEN Yide, ZHOU Xin, et al. Fast and accurate recognition of the strawberries in greenhouse based on improved YOLOv4-Tiny model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(18): 195-203. (in Chinese with English abstract)
- [11] GAI R, CHEN N, YUAN H. A detection algorithm for cherry fruits based on the improved YOLO-v4 model[J]. Neural Computing and Applications, 2023, 35(19): 13895-13906.
- [12] MASHORI S, AZMI M A A, SAHARI N, et al. Development of pesticide sprayer robot prototype for chilli farm agricultural application[J]. International Journal of Integrated Engineering, 2023, 15(3): 217-226.
- [13] 王昱, 姚兴智, 李斌, 等. 基于改进 YOLO v7-tiny 的甜椒畸形果识别算法[J]. 农业机械学报, 2023, 54(11): 236-246.
WANG Yu, YAO Xingzhi, LI Bin, et al. Malformed sweet pepper fruit identification algorithm based on improved YOLO v7 tiny[J]. Transactions of the Chinese Society of Agricultural Machinery (Transactions of the CSAE), 2023, 39(4): 124-134. (in Chinese with English abstract)
- [14] HUYNH Q K, NGUYEN V C, NGUYEN C N, et al. A 3D location estimating model for harvesting the fresh chili fruit using Yolov5[C]//International Conference on Advances in Mechanical Engineering and Material Science. Singapore: Springer Nature Singapore, 2023: 187-195.
- [15] LUO Y, LIU Y, WANG H, et al. YOLO-CFruit: A robust object detection method for Camellia oleifera fruit in complex environments[J]. Frontiers in Plant Science, 2024, 15: 1389961.
- [16] HABARAGAMUWA H, OGAWA Y, SUZUKI T, et al. Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network[J]. Engineering in Agriculture, Environment and Food, 2018, 11(3): 127-138.
- [17] LIU Y, REN H, ZHANG Z, et al. Research on multi-cluster green persimmon detection method based on improved Faster RCNN[J]. Frontiers in Plant Science, 2023, 14: 1177114.
- [18] 傅隆生, 冯亚利, Elkamil Tola, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205-211.
FU Longsheng, FENG Yali, ELKAMIL Tola, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(2): 205-211. (in Chinese with English abstract)
- [19] LI X, PAN J, XIE F, et al. Fast and accurate green pepper detection in complex backgrounds via an improved Yolov4-tiny model[J]. Computers and Electronics in Agriculture, 2021, 191: 106503.
- [20] 苗荣慧, 李志伟, 武锦龙. 基于改进 YOLO v7 的轻量化樱桃番茄成熟度检测方法[J]. 农业机械学报, 2023, 54(10): 225-233.
MIAO Ronghui, LIZhiwei, WUJinlong, et al. Lightweight Maturity Detection of Cherry Tomato Based on Improved YOLO v7[J]. Transactions of the Chinese Society of Agricultural Machinery (Transactions of the CSAE), 2023, 54(10): 225-233. (in Chinese with English abstract)
- [21] 熊俊涛, 韩咏林, 王潇, 等. 基于 YOLO v5-Lite 的自然环境木瓜成熟度检测方法[J]. 农业机械学报, 2023, 54(6): 243-252.
XIONG Juntao, HAN Yonglin, WANG Xiao, et al. Papaya Maturity Detection Method in Natural Environment Based on YOLO v5-Lite[J]. Transactions of the Chinese Society for

- Agricultural Machinery, 2023, 54(6): 243-252. (in Chinese with English abstract)
- [22] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. 2021[2024-05-24]. <https://doi.org/10.48550/arXiv.2107.08430>.
- [23] ZHANG Y, ZHANG W, YU J, et al. Complete and accurate holly fruits counting using YOLOX object detection[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107062.
- [24] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Montreal, QC, Canada. 2021: 13713-13722.
- [25] JIA Y, FU K, LAN H, et al. Maize tassel detection with CA-YOLO for UAV images in complex field environments[J]. *Computers and Electronics in Agriculture*, 2024, 217: 108562.
- [26] WANG R, SHIVANNA R, CHENG D, et al. Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems[C]//Proceedings of the web conference 2021. New York, United States. 2021: 1785-1797.
- [27] ZHANG Z, HU Y, LI X, et al. Autonomous navigation method of jujube catch-and-shake harvesting robot based on convolutional neural networks[J]. *Computers and Electronics in Agriculture*, 2023, 215: 108469.
- [28] 刘安稳, 谢方平, 向阳, 等. 基于机器视觉的菠萝自动去眼方法与试验[J]. *农业工程学报*, 2024, 40(1): 80-89. LIU Anwen, XIE Fangping, XIANG Yang, et al. Automatic removal of pineapple eye using machine vision[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2024, 40(1): 80-89. (in Chinese with English abstract)
- [29] BADGUJAR C M, POULOSE A, GAN H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review[J]. *Computers and Electronics in Agriculture*, 2024, 223: 109090. 2024, 55(2): 278-286. (in Chinese with English abstract)

Detecting chili pepper fruits in a natural environment using improved YOLOX

LI Xu^{1,2}, LIU Qing¹, KUANG Minqiu¹, PAN Jiandong³, LIU Dawei^{1,2*},
XIANG Yang^{1,2}, WU Yanhua⁴, XIE Fangping^{1,2}

(1. College of Mechanical and Electrical Engineering, Hunan Agricultural University, Changsha 410128, China; 2. Hunan Key Laboratory of intelligent Agricultural Machinery Equipment, Changsha 410128, China; 3. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310000, China; 4. Changsha DuoLai Agricultural Science and Technology Company Limited, Changsha 410100, China)

Abstract: Chili pepper is one of the most widely planted vegetables in China. The current production of fresh chili peppers, such as field management and harvesting, faces the challenges of high labor intensity and low efficiency. The chili pepper industry is ever transitioning towards mechanization and intelligent production. The rapid and accurate detection of chili fruits in the natural environment is of great significance for the automatic picking of chili peppers. However, it is still lacking in the adaptive ability and detection accuracy of the model under different light and occlusion conditions. In this study, an improved detection model, called YOLOX_Pepper, was proposed for chili fruit using YOLOX. Firstly, a fusion-efficient channel CA (coordinate attention) attention mechanism was added to the YOLOX feature fusion network, in order to capture the key features of chili fruits. Secondly, the convolution module in the feature fusion module of the backbone network was replaced with Deformable Convolutional DCNv2 (Deformable ConvNets v2), in order to improve the perceptual ability of the model in the case of the complex geometric features of chili pepper length, width, and aspect ratio, due to branch and fruit occlusion. The experimental results showed that the improved YOLOX_Pepper model had mAP (mean average precision) of 93.30%, which was 3.99, 1.58, 3.19, and 2.84 percentage points higher than that of Faster R-CNN, YOLOv5, YOLOv7, and YOLOX, respectively, with an F1 score of 96%, and an average time for the single-image detection of 0.026s. Under strong light conditions, the mAP of green and red chili fruits of the YOLOX_Pepper model was 69.16% and 89.67%, respectively, and the number of correctly detected green and red peppers was 83 and 304, respectively. Under shadow conditions, the mAP of green and red peppers of the YOLOX_Pepper model was 77.21% and 90.42%, respectively, and the number of green and red peppers was 119 and 255 correctly detected. Under the lack of light conditions, the mAP of the YOLOX_Pepper model for green peppers and red peppers were 77.38% and 75.47%, respectively, and the number of correctly detected green and red peppers were 86 and 311, respectively. The YOLOX_Pepper model performed better in various light conditions, especially in the number and accuracy of detections, compared with the YOLOv5, YOLOv7, and YOLOX models. Under fruit occlusion conditions, the mAP of YOLOX_Pepper was 71.15% and 94.87% for green and red peppers, respectively, and the number of correct detections was 79 and 650 for green and red peppers, respectively. Under branch and foliage occlusion conditions, the mAP of YOLOX_Pepper was 83.98% and 87.10% for green and red peppers, respectively, and the number of correctly detected green and red peppers was 88 and 394, respectively. The improved YOLOX_Pepper model performed better in the chili fruit detection under different occlusions, compared with the YOLOv5, YOLOv7, and YOLOX models. The YOLOX_Pepper model showed excellent performance of detection in complex environments. The effectiveness of the improved module can also provide the intelligent production of chili peppers with reliable technical support.

Keywords: agriculture; machine vision; YOLOX; pepper; natural environment; fruit detection