

基于改进YOLOv5的辣椒目标检测算法研究

王启锬, 王旭*

贵州大学大数据与信息工程学院, 贵州 贵阳

收稿日期: 2023年8月9日; 录用日期: 2023年9月12日; 发布日期: 2023年9月20日

摘要

在自然环境下, 辣椒果实生长比较密集, 目标检测算法很难区分被枝叶遮挡或被其他辣椒遮挡的情况, 这对自动检测辣椒果实增加了困难, 在本文中, 我们采用现有的辣椒数据集, 基于YOLOv5算法模型进行了辣椒图像识别。为了更好的对辣椒进行检测, 减小因为枝叶遮挡和被其他辣椒遮挡而造成的检测难度, 我们在原来的YOLOv5模型中加入了CBAM注意力机制, 构建了CM-YOLO模型, 我们的模型mAP可以达到93.1%, 比原来的YOLOv5高0.4个百分点。我们所构建的CM-YOLO模型提高了对辣椒的检测能力, 同样该模型也可以应用于农产品识别和分类等其他领域。

关键词

辣椒图像识别, YOLOv5, CBAM注意力机制

Research on Chili Target Detection Algorithm Based on Improved YOLOv5

Qikun Wang, Xu Wang*

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: Aug. 9th, 2023; accepted: Sep. 12th, 2023; published: Sep. 20th, 2023

Abstract

In natural environments, the growth of chili peppers is relatively dense, making it difficult for object detection algorithms to distinguish between situations obstructed by branches and leaves or other chili peppers. This increases the difficulty of automatically detecting chili peppers. In this article, we used existing chili pepper datasets and conducted chili image recognition based on the YOLOv5 algorithm model. In order to better detect chili peppers and reduce the detection difficulty caused by branch and leaf occlusion and being obstructed by other chili peppers, we added the

*通讯作者。

CBAM attention mechanism to the original YOLOv5 model and constructed a CM-YOLO model. Our model's mAP can reach 93.1%, which is 0.4 percentage points higher than the original YOLOv5 model. The CM-YOLO model we have constructed has improved the detection ability of chili peppers, and can also be applied to other fields such as agricultural product recognition and classification.

Keywords

Chili Pepper Image Recognition, YOLOv5, CBAM Attentional Mechanisms

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

现如今,人工智能技术已经渗透到我们的日常生活中,其在农业采摘领域的应用也越来越广泛[1] [2]。在采摘辣椒的行业中,传统的人工采摘效率低且成本高,因此,实现辣椒的自动采摘、保证辣椒果实的及时采摘、并降低采摘成本,已经成为该行业发展的必然趋势。对于实现自动采摘辣椒果实,辣椒信息的智能感知和定位技术是关键所在[3] [4]。因此,为了提高辣椒自动采摘效率,必须实现辣椒果实目标的快速、准确的识别和定位。

自然生长环境下辣椒通常会存在重叠遮挡现象,此类果实的识别是自动采摘辣椒的重要问题[5] [6]。重叠遮挡生长状态下的辣椒果实识别方法不同于无遮挡形态果实以及其他生长状态下的果实,因此,能够自动识别有无重叠遮挡状态下的辣椒果实成为关键。

近年来,目标检测和定位技术发展迅速,在各个领域都有这些技术的参与,其中在农业领域也取得了显著的成果,这为辣椒果实地识别与定位提供了研究基础。目标检测算法可分为传统的检测算法和基于深度学习的算法。然而,传统目标检测算法在处理复杂多变的背景时,其检测效果容易受到较大的影响。因此,研究者们一直在不断改进这一方法。Whittaker等[7]为了检测番茄,使用 Sobel 算子检测果实的边缘,并利用 Hough 变换计算番茄的中心坐标。该方法能够很好地识别重叠的番茄,但是处理时间较长,识别效率较低。Gongal等[8]为了实现对苹果的自动化检测,结合了 HSI 和 RGB 颜色空间,对苹果图像进行了处理。利用 Otsu 阈值分割和 Hough 变换对苹果进行了分割识别,平均识别精度为 78.9。该方法不适用于数量较多、环境较复杂的辣椒。随着深度学习卷积神经网络技术的不断发展,目标识别检测相较于传统方法展现出更高的优越性[9],目前的卷积神经网络主要分为两大类,一种是端到端的方法[10],主要有 YOLO、SSD、RetinaNet,另一种是基于区域建议的方法[11] [12] [13],主要有 Fast RCNN、Faster RCNN、Mask RCNN。

在本文中,为解决辣椒果实生长密集、在识别过程中可能会遇到漏检或错检的问题,构建了 CM-YOLO 模型。

2. CM-YOLO 模型结构

2.1. YOLOv5 模型

YOLO [14] [15]系列网络模型是最为经典的 one-stage 算法,在目标检测的网络里面,它是在工业领域使用最多的网络模型。YOLOv5 网络模型在继承了原有 YOLO 网络模型优点的基础上,具有更优的检

测精度和更快的推理速度。模型的整体结构如图 1 所示。YOLOv5 网络结构由输入端(Input)、Backbone、Neck、Head 组件组成。各部分完成的主要功能, 所述如下:

(1) Input: YOLOv5 采用了 YOLOv4 中的 Mosaic 数据增强方法来丰富数据集, 降低硬件需求, 从而减少了 GPU 的使用。此外, 自适应锚框计算功能被嵌入到整个训练代码中, 可以根据需要自行调节开关, 同时也实现了自适应图片缩放, 这有助于提高目标检测的推理速度。

(2) Backbone: Focus 模块将输入的图片数据分成四份, 每一份数据相当于进行了 2 倍下采样, 然后将这四份数据进行拼接, 得到一个尺寸缩小一半的新特征图。最后, 通过卷积操作将信息进行融合, 并改变特征图的通道数。这种方法的优点是能够最大程度地减少信息损失和计算量, 同时增加了难样本和数据的多样性。

(3) Neck: 在 YOLOv5 网络模型中, 颈部网络主要负责对主干网络提取的特征进行增强处理, 以提高后续预测的精度。原始的 FPN 结构采用自顶向下的特征融合方式来处理目标检测领域中的多尺度变化问题, 已经在许多模型中得到了广泛应用。但是, 如果仅使用 FPN 结构来融合上下文信息, 则无法实现上层信息与底层信息的交流。因此, 在 FPN 的基础上, YOLOv5 网络还加入了 PAN 结构, 引入了一条自下而上的信息流, 充分实现了网络上下信息流的融合, 从而提高了网络的检测能力。

(4) Head: YOLOv5 采用损失函数 CIoU-Loss, 可以将一些遮挡重叠的目标准确识别出来。

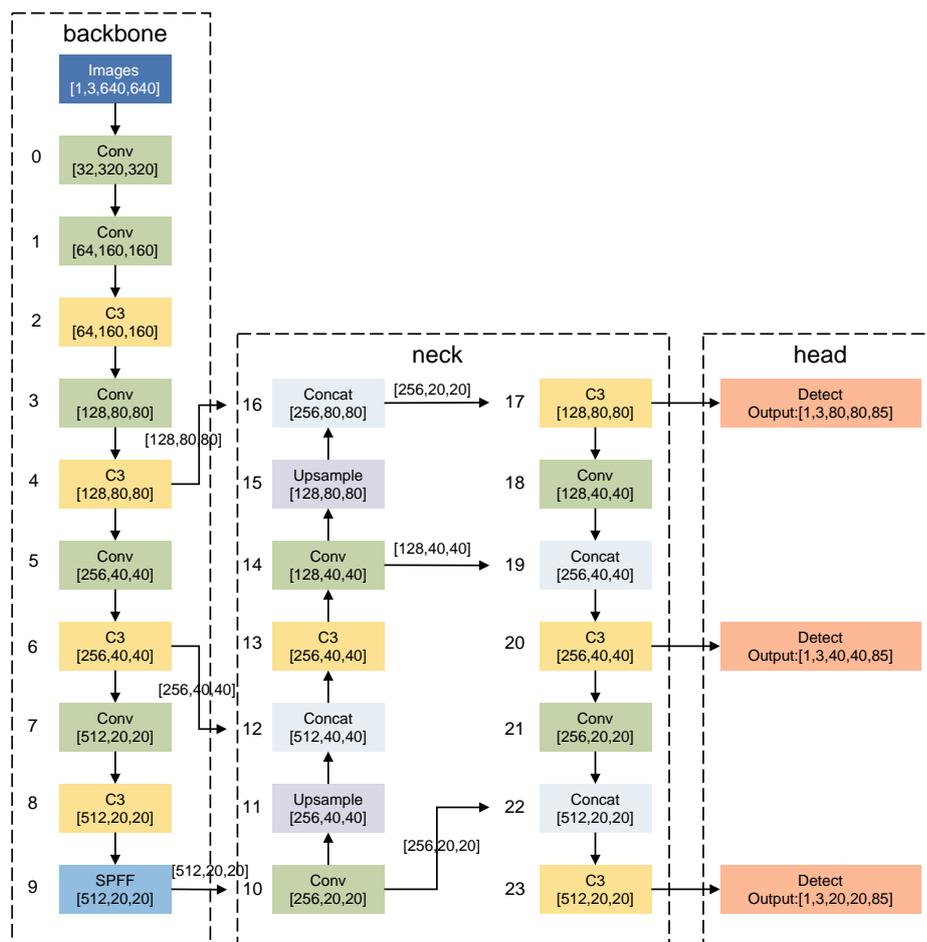


Figure 1. Original YOLOv5 model

图 1. 原始 YOLOv5 模型

2.2. CM-YOLO 模型架构

为了实现微小目标的有效检测,我们对 YOLOv5 模型进行了改进,增加了小目标检测层的 YOLOv5 模型如图 2 所示。我们利用 P2 层提取更丰富的底层特征信息,并与主干网络的同尺度特征融合,得到了三个输入特征的融合结果。此外,我们还增加了一个底层特征检测头[16],它能够更加敏感地检测微小目标。虽然该检测头增加了模型的计算量和内存开销,但有效提高了我们对密集生长的辣椒果实的检测能力,并缓解了尺度方差的负面影响。该检测头分辨率为 160×160 像素,相当于在主干网络中只进行了 2 次下采样。另外,我们的 P2 层检测头与原始的三个检测头一起使用,可以进一步提高检测效果。这些检测头针对不同的特征图生成,可以有效地覆盖不同尺度的目标。通过将它们结合起来,我们能够在不影响速度的情况下提高检测的精度和鲁棒性。

总之,我们的改进方案在不增加过多计算量和内存开销的情况下,有效地提高了模型对微小目标的检测能力。这对于像辣椒这样密集生长的农作物的自动化检测非常重要,也为后续的研究提供了有力的基础。

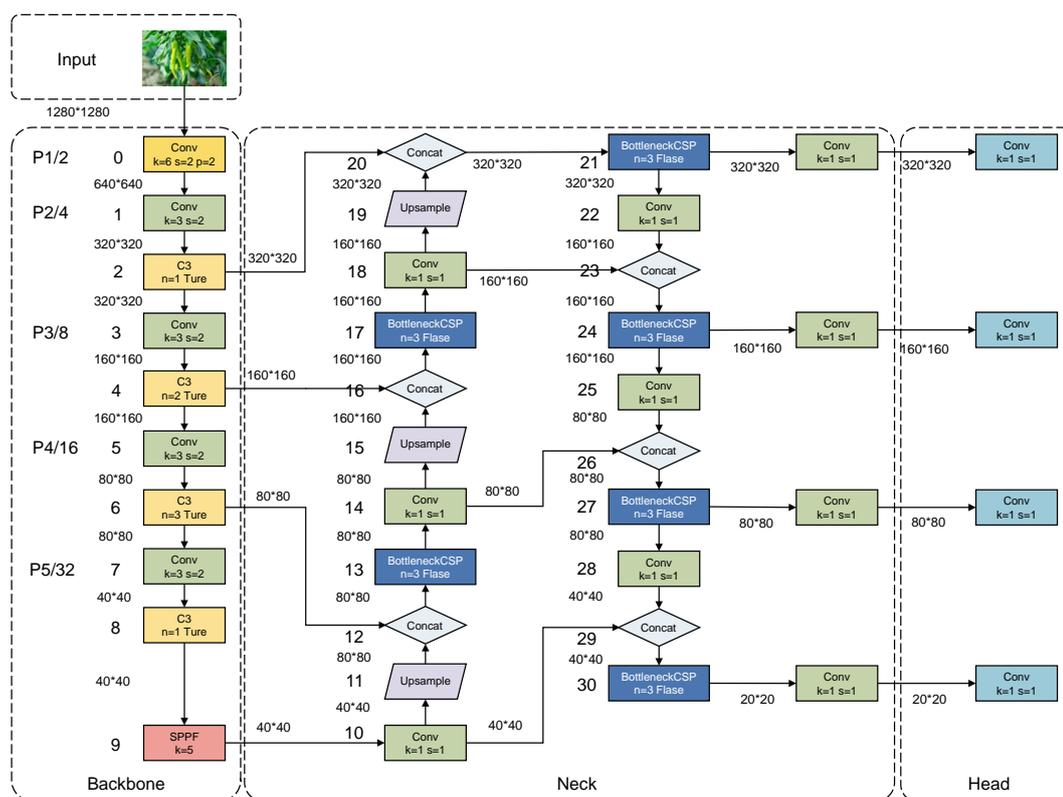


Figure 2. YOLOv5 model with added small target detection layer

图 2. 增加小目标检测层的 YOLOv5 模型

2.3. 添加 CBAM 注意力机制模块

为了改善辣椒果实的识别效果,我们选择在特征融合网络中应用融合通道注意力和空间注意力的 CBAM 模块[17] [18]。这种卷积注意力模块能够帮助网络模型更好地关注辣椒果实的关键信息,特别是在自然场景下的辣椒生长环境比较复杂,且辣椒图像的前景和背景较为相似的情况下。我们将 CBAM 模块融合到原始 YOLOv5 的 C3 模块后,如图 3 所示。这样做可以增强特征融合网络的性能,并提高对辣椒果实的识别能力。

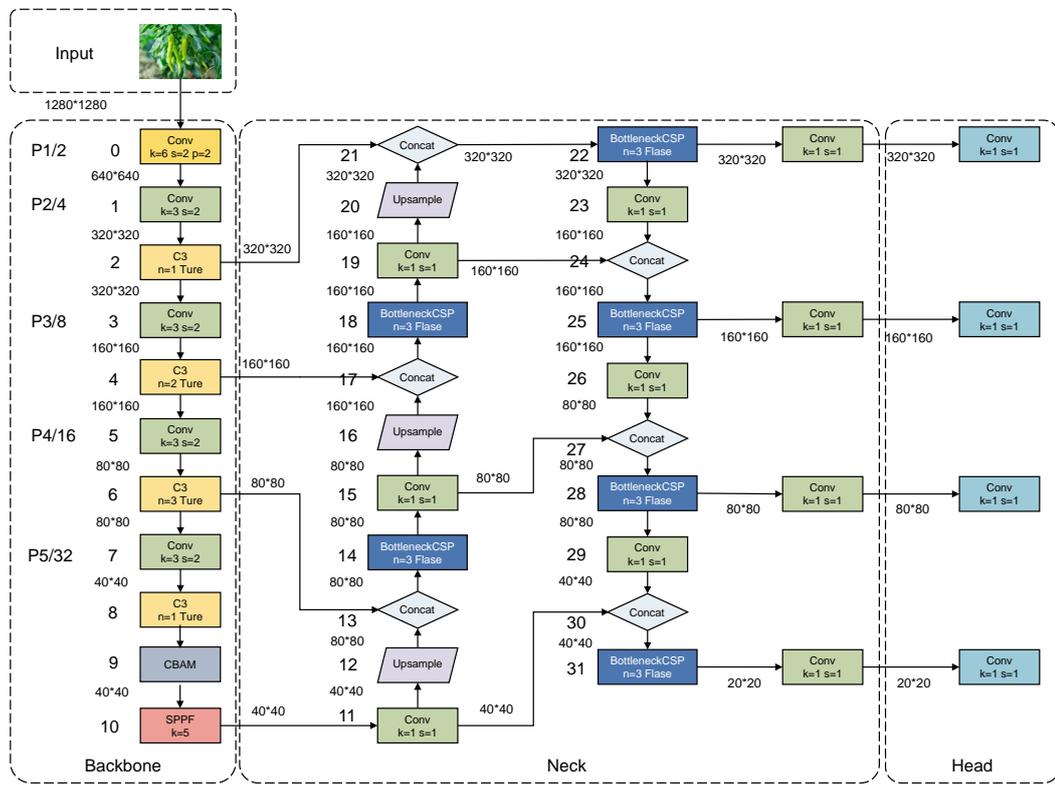


Figure 3. YOLOv5 model fused with CBAM
图 3. 融合 CBAM 的 YOLOv5 模型

CBAM 是一种简单而有效的轻量级注意力模块，可以轻松地集成到 CNN 架构中，并进行端到端的训练。通过在给定特征映射上逐步生成通道和空间维度的注意力映射，并将这些映射与输入特征映射相乘，CBAM 注意力模块可以自适应地细化特征。CBAM 注意力模块的结构如图 4 所示。当遇到目标尺寸较小、特征不显著且数量较少的情况时，将 CBAM 注意力模块嵌入主干网络中可以增强网络对目标特征的提取能力，同时改善颈部网络中特征融合部分的性能。在目标检测任务中，CBAM 注意力模块可以有效地协助模型提取关注区域，从而提高检测性能。

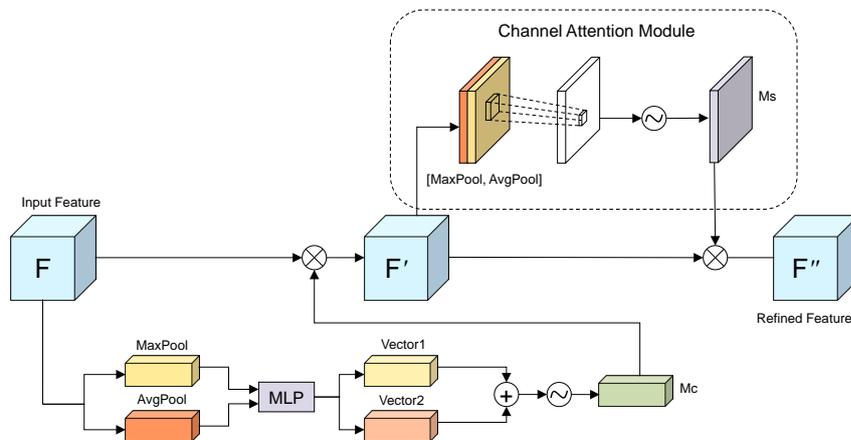


Figure 4. CBAM module structure diagram
图 4. CBAM 模块结构图

3. 实验评估

3.1. 数据集

实验中数据集一部分来自 Zhang [19]等构建的辣椒数据集, 一部分为我们采集的辣椒图像, 以露地栽培模式下的二荆条辣椒为研究对象, 露地栽培模式下, 辣椒树行距约 10 cm, 株距约 40 cm。辣椒处于绿熟成长期, 为了正确反映辣椒生长姿态及环境的复杂性, 采集了晴天和阴天的辣椒图像。拍摄阶段包括上午、中午和下午。收集了 2000 张图像, 这些图像在不同光照和角度条件下采集, 其中包括: 无遮挡和无重叠的辣椒图像、果实部分重叠的辣椒图像、叶片遮挡的辣椒图像、枝干遮挡的辣椒图像以及叶片和枝干混合遮挡的辣椒图像。采集设备为手机自带摄像头, 对比度、饱和度和锐度设置为标准模式。包括以下条件: 被叶子、树枝遮挡的辣椒, 混合遮挡的辣椒与重叠的辣椒、自然光角度、背光角度和侧光角度等。此外, 在图像采集中还考虑了云、阴影、高光、弱光、反射等环境因素。俯视图 785 张(正光 462 张, 背光 323 张), 平视图 1215 张(正光 871 张, 背光 344 张), 数据集部分原始图像如图 5 所示。



Figure 5. Original images of some datasets

图 5. 部分数据集原始图像

3.2. 结果分析

本实验训练过程在 Ubuntu18.0、CUDA11.0 环境下进行, GPU 配置: NVIDIA GeForce RTX 3090, 24GB 显存, 调用 GPU 进行训练。所有实验训练参数设置: 输入图片大小为 640×640 , 优化器采用带动量的 SGD 优化器, 初始学习率设置为 0.001、批大小为 16, 共训练 200 轮。检测结果如图 6 所示:

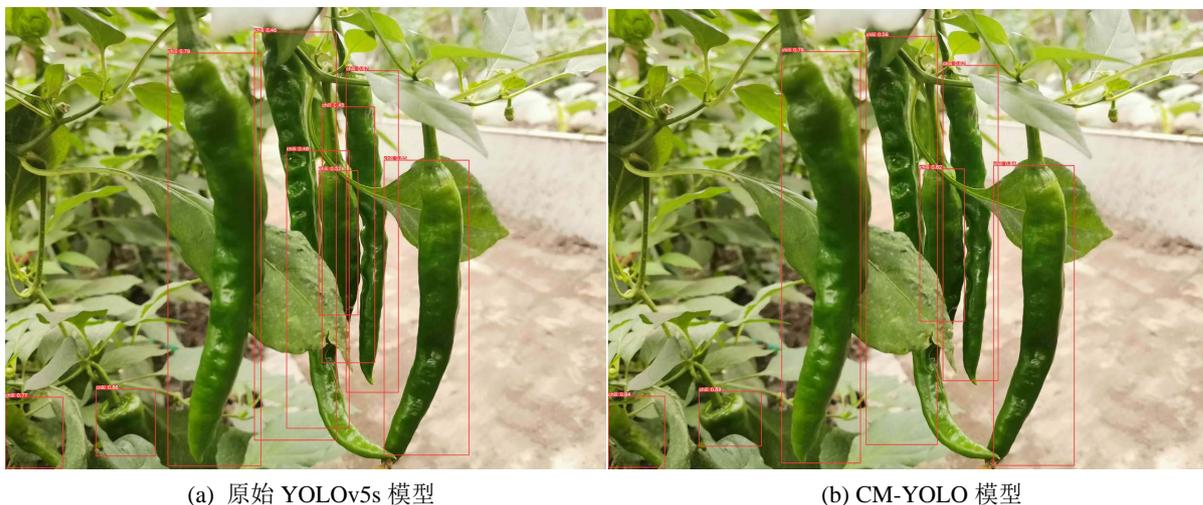


Figure 6. Comparison of detection results between original YOLOv5s and CM-YOLO
图 6. 原始 YOLOv5s 与 CM-YOLO 的检测结果比较

由测试结果可知, 经过改进的网络识别准确度获得较大提高, 从图(a)和图(b)两张图片中可以看出, 改进后的 CM-YOLO 网络的 Map 比原始 YOLOv5s 网络有所提升, 如在有叶子遮挡的情况下, 原始 YOLOv5s 网络识别目标框的 Map 为 0.52, 如图 6(a)所示, CM-YOLO 模型的 Map 为 0.82, 如图 6(b)所示。因此 CM-YOLO 网络目标识别框的位置更加精准。

通过实验结果比较原始 YOLOv5s 模型与 CM-YOLO 模型对于辣椒果实目标检测识别的效果, 从准确率 Precision, 平均精度(mAP), 召回率 Recall 对算法进行比较, 比较结果如表 1 所示, 具体地。

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

TP 表示实际为正样本且被分类器正确预测为正样本的样本数量, FP 表示实际为负样本但被分类器错误地预测为正样本的样本数量, FN 表示实际为正样本但被分类器错误地预测为负样本的样本数量。Precision 表示分类器正确预测为正样本的样本数量占分类器预测为正样本的所有样本数量的比例, Recall 表示分类器正确预测为正样本的样本数量占所有实际为正样本的样本数量的比例。

Table 1. Comparison of test results
表 1. 检测结果比较

	mAP	Precision	Recall
原始 YOLOv5s	0.927	0.915	0.932
CM-YOLO	0.931	0.925	0.938

由表 1 可以看出, 基于 CM-YOLO 的辣椒检测平均精度和召回率更高, 能够较好地完成辣椒检测任务。图 7 是原始 YOLOv5 模型与 CM-YOLO 模型特征可视化对比, 由图 7 可以看出, CM-YOLO 模型提取的辣椒线条更加明显。

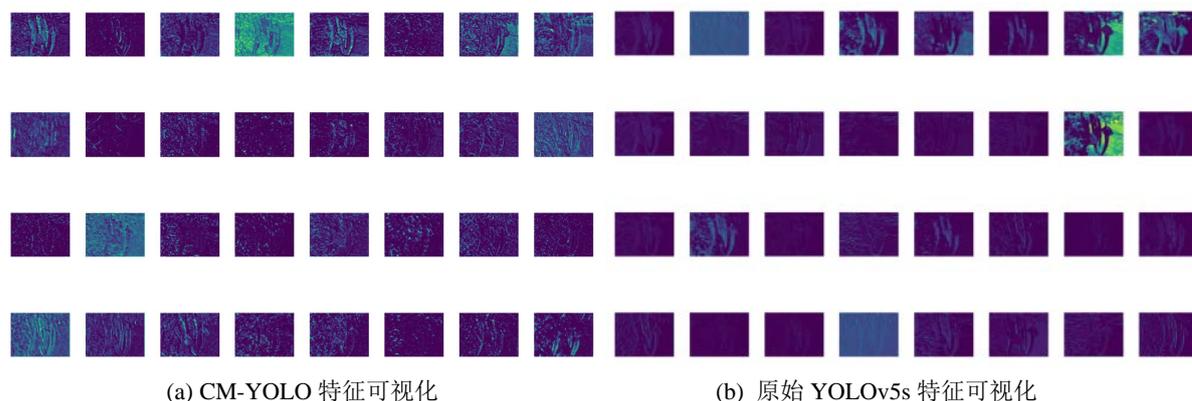


Figure 7. Comparison of feature visualization between the original YOLOv5 model and the CM-YOLO model
图 7. 原始 YOLOv5 模型与 CM-YOLO 模型特征可视化对比

4. 结论

针对自然环境下, 辣椒果实生长比较密集, 原始 YOLOv5 算法对实现有遮挡及密集生长的辣椒果实检测性能较差的问题, 进行了两个方面的改进: 首先, 利用 YOLOv5 的 P2 层提取更丰富的底层特征信息, 并与主干网络的同尺度特征融合, 得到了三个输入特征的融合结果, 并且增加了一个底层特征检测头, 更加敏感地检测辣椒果实目标。其次, 将 CBAM 模块融合到原始 YOLOv5 的 C3 模块后, 增强特征融合网络的性能, 并提高对辣椒果实的识别能力。通过实验验证, 该模型对自然场景下的辣椒果实检测性能良好, 为实现辣椒果实的目标检测提供了良好的技术支持。

参考文献

- [1] 朱智惟, 单建华, 余贤海, 等. 基于 YOLOv5s 的番茄采摘机器人目标检测技术[J]. 传感器与微统, 2023, 42(6): 129-132. [https://doi.org/10.13873/J.1000-9787\(2023\)06-0129-04](https://doi.org/10.13873/J.1000-9787(2023)06-0129-04)
- [2] 赵敬, 王全有, 褚幼晖, 等. 农业采摘机器人发展分析及前景展望[J]. 农机使用与维修, 2023(6): 63-70. <https://doi.org/10.14031/j.cnki.njwx.2023.06.019>
- [3] 高帅, 刘永华, 高菊玲, 等. 基于 YOLOv3 算法与 3D 视觉的农业采摘机器人目标识别与定位研究[J]. 中国农机化学报, 2022, 43(12): 178-183. <https://doi.org/10.13733/j.jcam.issn.2095-5553.2022.12.026>
- [4] 李亚涛. 茶叶采摘机器人的视觉检测与定位技术研究[D]: [博士学位论文]. 杭州: 浙江理工大学, 2022. <https://doi.org/10.27786/d.cnki.gzjlg.2022.000006>
- [5] 魏天宇, 柳天虹, 张善文, 等. 基于改进 YOLOv5s 的辣椒采摘机器人识别定位方法[J]. 扬州大学学报(自然科学版), 2023, 26(1): 61-69. <https://doi.org/10.19411/j.1007-824x.2023.01.010>
- [6] 刘丽娟, 窦佩佩, 王慧. 自然环境下重叠与遮挡苹果图像识别方法研究[J]. 中国农机化学报, 2021, 42(6): 174-181. <https://doi.org/10.13733/j.jcam.issn.2095-5553.2021.06.27>
- [7] Whittaker, D.E., Miles, G.R., Mitchell, O.D. and Gaultney, L. (1987) Fruit Location in a Partially Occluded Image. *Transactions of the ASAE*, **30**, 591-596. <https://doi.org/10.13031/2013.30444>
- [8] Gongal, A., et al. (2016) Apple Crop-Load Estimation with Over-the-Row Machine Vision System. *Computers and Electronics in Agriculture*, **120**, 26-35. <https://doi.org/10.1016/j.compag.2015.10.022>
- [9] 李杰. 结合改进注意力机制的 YOLO 目标检测算法[J]. 计算机时代, 2023(7): 108-113. <https://doi.org/10.16644/j.cnki.cn33-1094/tp.2023.07.025>
- [10] Liu, W., et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV*, Springer International Publishing, Berlin, 21-37.
- [11] Mehta, S.S., Ton, C., Asundi, S. and Burks, T.F. (2017) Multiple Camera Fruit Localization Using a Particle Filter. *Computers and Electronics in Agriculture*, **142**, 139-154. <https://doi.org/10.1016/j.compag.2017.08.007>
- [12] Nyarko, E.K., Vidović, I., Radočaj, K. and Cupec, R. (2018) A Nearest Neighbor Approach for Fruit Recognition in

- RGB-D Images Based on Detection of Convex Surfaces. *Expert Systems with Applications*, **114**, 454-466. <https://doi.org/10.1016/j.eswa.2018.07.048>
- [13] 徐惠荣, 叶尊忠, 应义斌. 基于彩色信息的树上柑橘识别研究[J]. 农业工程学报, 2005, 21(5): 98-101.
- [14] Jiang, P., Ergu, D., Liu, F., Cai, Y. and Ma, B. (2022) A Review of Yolo Algorithm Developments. *Procedia Computer Science*, **199**, 1066-1073. <https://doi.org/10.1016/j.procs.2022.01.135>
- [15] 张睿敏, 杜叔强, 周秀媛. 多目标口罩规范佩戴检测算法及技术[J]. 兰州工业学院学报, 2023, 30(3): 43-47.
- [16] 李家俊, 杨杰. 基于改进 YOLOX 的微小目标快速检测方法[J]. 江西冶金, 2023, 43(2): 164-172. <https://doi.org/10.19864/j.cnki.jxye.2023.02.012>
- [17] 张鑫, 周顺勇. 改进 YOLOv5s 的摩托车头盔佩戴检测算法[J]. 四川轻化工大学学报(自然科版), 2023, 36(3): 50-58.
- [18] 孙想, 吴华瑞, 朱华吉, 等. 基于 LW-YOLOv3 模型的棉花主茎生长点检测与定位研究[J]. 河北农业大学学报, 2021, 44(6): 106-115.
- [19] Zhang, S. and Xie, M. (2023) Real-Time Recognition and Localization Based on Improved YOLOv5s for Robot's Picking Clustered Fruits of Chilies. *Sensors*, **23**, Article No. 3408. <https://doi.org/10.3390/s23073408>