

单统华, 刘立群. 基于 YOLOv8 的轻量化苹果异源目标检测模型[J]. 智能计算机与应用, 2025, 15(11): 62-67. DOI: 10.20169/j. issn. 2095-2163. 251110

# 基于 YOLOv8 的轻量化苹果异源目标检测模型

单统华, 刘立群

(甘肃农业大学 信息科学技术学院, 兰州 730070)

**摘要:** 果园环境具有多样性且存在复杂的外界因素干扰, 如光照条件、枝干遮挡、树叶重叠等, 异源果实目标检测是一项极具挑战性的任务。针对这些问题, 本文提出了一种基于 YOLOv8 的轻量化苹果异源目标检测模型 (YoloMN)。在该模型的目标检测部分, 引入了一个数据预处理模块, 能够同时处理异源彩色图像和深度图像, 通过图像特征的融合来增强特征表达能力; 采用 MobileNetV3 对主干网络进行了改进, 并在其后的头部输入部分加入了 NAM attention 模块, 显著提升了模型的特征提取和表达能力, 进一步强化了模型在复杂环境中的异源目标检测能力。本文提出的模型在保持原有目标检测精度提高的同时参数量减少 60% 和模型大小减小 50%。在复杂环境下, 本文模型展现出了较好的目标检测能力, 在减少参数量的情况下, 本文模型的目标检测能力依旧出色。

**关键词:** 异源目标; YOLOv8; MobileNetV3; 目标识别; NAM Attention

中图分类号: TP391 文献标志码: A 文章编号: 2095-2163(2025)11-0062-06

## Lightweight apple heterogeneous object detection model based on YOLOv8

SHAN Tonghua, LIU Liqun

(College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China)

**Abstract:** Due to the diversity of the orchard environment and the presence of complex external interference factors such as lighting conditions, branch obstruction, and overlapping leaves, detecting heterogeneous fruit targets is a highly challenging task. To address these issues, this paper proposes a lightweight apple heterogeneous target detection model based on YOLOv8, named YoloMN model. In the target detection part of this model, we introduced a data preprocessing module that can simultaneously process heterogeneous RGB images and depth images, enhancing feature expression capabilities through the fusion of image features. To further optimize the model, we improved the backbone network using the MobileNetV3 module and added the NAM Attention module to the head input part of the model, significantly enhancing the model's feature extraction and expression capabilities, thereby strengthening its heterogeneous target detection ability in complex environments. While maintaining the original target detection accuracy, our model achieves significant optimizations with a 60% reduction in parameter count and a 50% reduction in model size. In complex environments, our model demonstrates excellent target detection capabilities, and even with a reduced parameter count, its target detection performance remains outstanding.

**Key words:** heterogeneous target; YOLOv8; MobileNetV3; target detection; NAM Attention.

## 0 引言

随着农业科技的不断进步和智能化水平的日益提高, 果实识别技术已成为现代农业领域中的一项关键技术<sup>[1]</sup>。果实异源目标识别是指在复杂的果园环境中, 利用目标识别传感器以及图像处理技术,

对当前采集画面中的不同种类、不同生长状态的果实进行识别。

目标检测的应用不仅有助于提高果园的生产效率, 减少人工劳动强度, 对于果实品质的监测和控制也具有重要意义, 特别是在精准农业、自动化采摘以及智能农业管理等方面, 果实识别发挥着越来越重

基金项目: 甘肃省高校教师创新基金 (2023A-051); 甘肃农业大学青年导师基金 (GAU-QDFC-2020-08); 甘肃省科技计划资助 (20JR5RA032)。

作者简介: 单统华(1997—), 男, 硕士, 主要研究方向: 深度学习, 图像处理。

通信作者: 刘立群(1982—), 女, 硕士, 教授, 研究生导师, 主要研究方向: 智能计算, 深度学习。Email: llqhjy@126.com。

收稿日期: 2024-02-21

要的作用。

果实异源目标识别的研究面临诸多挑战。果园环境中光照、阴影、遮挡等因素的影响,果实图像的获取和处理难度较大;不同种类果实之间的形态差异较小,分类算法的准确性和稳定性面临挑战;果实生长过程中的颜色、大小、形状等特征也会随着生长阶段变化,增加了识别的难度。因此,研究和发展果实异源目标识别技术,对于提升农业智能化水平、促进农业生产方式的转型升级具有重要意义。近年来,随着计算机视觉和人工智能技术的快速发展,果实异源目标识别技术取得了显著的进步。利用双目相机采集器可以快速获取果园中果实的图像信息<sup>[2]</sup>;利用图像处理算法可以自动识别并分类不同种类的果实<sup>[3]</sup>。目标检测在果园的应用不仅可以提高采摘效率,降低劳动成本,而且还可以为果实品质的监测和控制提供有力的支持。目标检测模型的计算量增大,对于实验设备的成本需求也在增加,所以在保证识别精度的同时,对模型进行轻量化处理同样是一项亟待解决的问题。

本文提出了一种基于 YOLOv8 的轻量化苹果异源目标检测模型(YOLOMN),采用图像特征融合预处理模块,将彩色图像与深度图像特征融合后输出特征图像作为本文网络的输入;使用 MobileNetV3 模块改进 YOLOv8 模型主干网络部分;对 YOLOv8 模型目标检测头进行改进,加入 NAM attention 模块,增强识别精度。

## 1 相关工作

目标检测是计算机视觉领域中的关键任务之一,其核心目标是在输入的图像或视频序列中精确识别并定位出特定的目标对象<sup>[4]</sup>。随着人工智能技术的持续演进,目标检测算法的研究也展现出了显著的进步与突破。早期阶段,目标检测主要依赖于手工提取特征的方法,如尺度不变特征转换算法(Scale-Invariant Feature Transform, SIFT)、方向梯度直方图算法(Histogram of Oriented Gradient, HOG)等,这些方法需要大量的人工干预和调整,对于不同的目标类型和场景适应性较差。随着深度学习技术的兴起,目标检测开始采用卷积神经网络(CNN)提取特征,这种方法自动从数据中学习特征,大大提高了目标检测的准确性和效率<sup>[5-6]</sup>。在人工智能的发展过程中,目标检测领域出现了许多重要的模型,如 Faster R-CNN (Faster Region Convolutional Neural Network)、YOLO (You Only Look Once, YOLO)、SSD

(Single Shot MultiBox Detector)等,这些模型在准确性和实时性方面都取得很好的平衡<sup>[7-9]</sup>。YOLO 系列算法将目标检测和分类任务结合起来,具有较高的检测准确度和检测速度;Faster R-CNN 则通过引入基于区域的卷积神经网络(Region Proposal Networks, RPN),提高了目标检测的精度和速度<sup>[10]</sup>。传统的 CNN 网络模型在保持较高准确性的同时,具有更小的模型大小和更快的运行速度,适用于移动设备和嵌入式设备等资源受限的场景,近年来还有一些研究工作关注于轻量级模型的目标检测,如 MobileNetV3、ShuffleNet 等<sup>[11-12]</sup>。目标检测技术的发展经历了从手工特征提取到深度学习特征提取的转变,同时也经历了从单一模型到多模型、从大型模型到轻量级模型的演变。目前,目标检测技术已经在许多领域得到了广泛应用,如人脸识别、物体跟踪、自动驾驶等<sup>[13]</sup>。

## 2 基础理论知识

### 2.1 YOLOv8 模型

YOLOv8 模型是 ultralytics 公司在 2023 年初发布的模型,如图 1 所示,支持图像分类、物体检测和实例分割任务。YOLOv8 模型是一个领先的 SOTA (State Of The Art) 模型,吸收了 YOLO 系列之前版本模型的精华,并在模型中加入了新的功能,从而实现了性能和灵活性的双重提升<sup>[14]</sup>。YOLOv8 模型整体架构涵盖了创新的骨干网络、革新的无锚点检测头以及优化的损失函数,主要用于目标检测任务,能够准确地检测出图像或视频中的目标物体,并给出其位置和类别信息。

### 2.2 MobileNetV3 模型

MobileNetV3 模型如图 2 所示,是由 Google 团队在 2019 年提出的用于移动设备以及嵌入式设备上运行的轻量化网络模型,在移动端图像分类、目标检测等任务上的表现均很出色<sup>[15-17]</sup>。MobileNetV3 模型采用了针对通道注意力的 SE (Squeeze-and-Excitation) 模块、参数搜索方法 NAS (Neural Architecture Search) 等,有利于进一步提升网络的性能;采用深度可分离卷积 DSC (Depthwise Separable Convolution) 和逆残差结构等,使模型在减少计算量的同时,拥有良好的分类性能<sup>[18]</sup>;将卷积的运算过程分为 2 个步骤,即逐点卷积和点卷积,减少了计算量和模型大小;采用了互补搜索组合,在整体网络结构中,前移平均池化层并移除最后一个卷积层,并使用 H-Swish 激活函数进行激活。

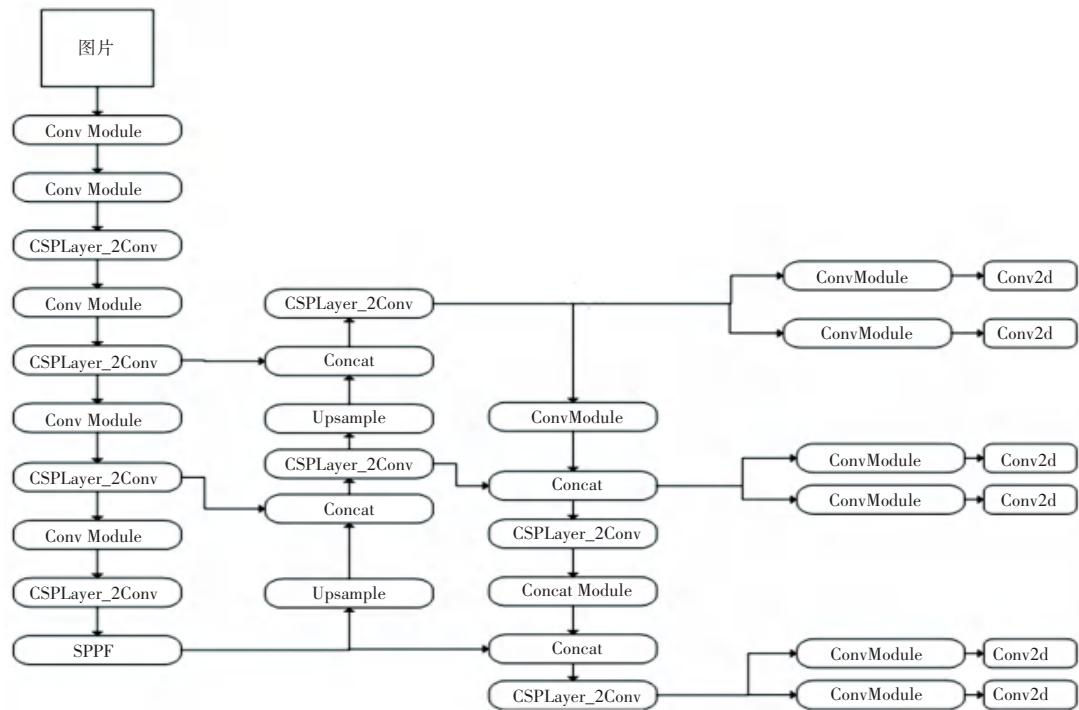


图 1 YOLOv8 模型

Fig. 1 YOLOv8 model

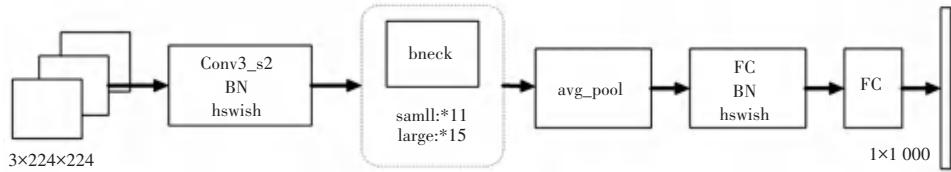


图 2 MobileNetV3 模型

Fig. 2 MobileNetV3 model

### 2.3 NAM Attention 模块

NAM attention 模块如图 3 所示, 是一种基于归一化的注意力模块, 在保证无额外计算过程增加的

基础上, 降低处理不太显著的特征的权重, 提高显著特征权重。NAM attention 的核心思想是将输入特征中的每个位置都表示为一个注意力向量。

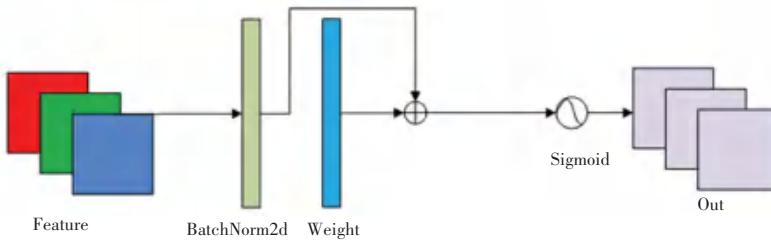


图 3 NAM attention 模块

Fig. 3 NAM attention module

### 3 YOLOMN 模型

YOLOMN 模型如图 4 所示, 首先在输入端采用跨平台计算机视觉库 (Open Computer Vision, OpenCV) 对数据进行特征融合预处理, 将异源 RGB

(Red-Green-Blue, RGB) 与深度图像 (Depth) 进行图像特征融合成为特征图像 (Feature), 作为本文目标识别的数据输入; 其次, 将原 YOLOv8 模型主干网络部分使用 MobileNetV3 模型结构进行改进, 把主干网络中原 Conv 网络模块替换为批归一化-HSwish

组合网络模块, 以及 C2f 模块整体替换为 MobileNetV3 网络模块; 最后, 为了增强识别能力与图片中关键特征表达, 在头部分层次中加入了注意

力机制, 将获取到的特征表达进行提纯, 突出显著的特征权重, 提高每次训练时的精确度。

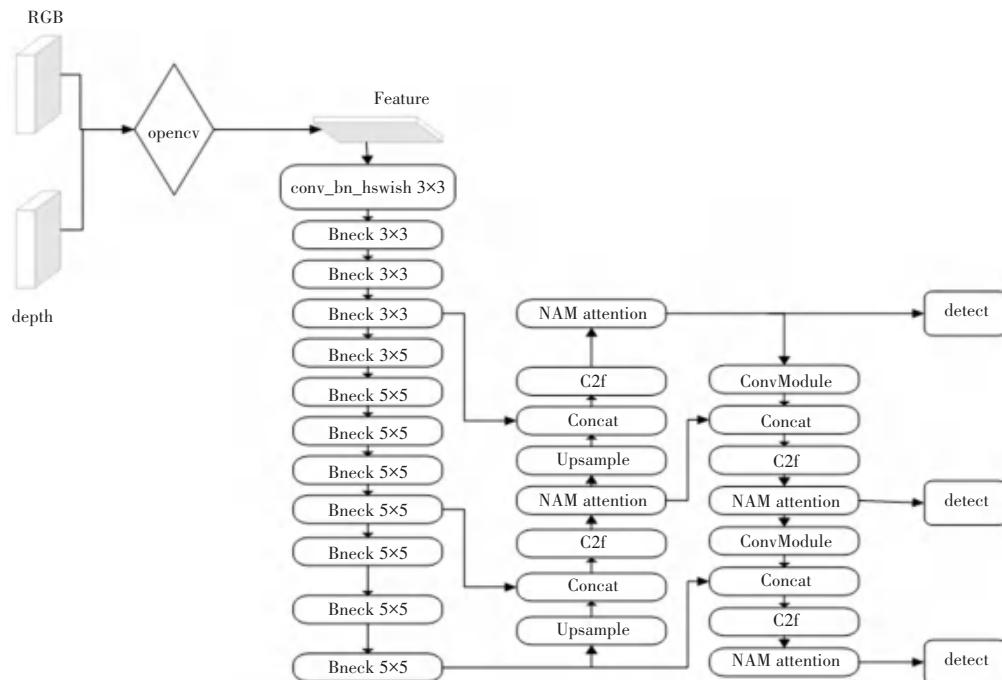


图 4 YOLOMN 模型

Fig. 4 YOLOMN model

## 4 实验

### 4.1 数据获取

为了验证本文目标检测模型, 使用两个目标检测数据集: RD-fuji 公开数据集和通过 RGB-TOF 双目采集系统采取的果树图像自建数据集, 每个数据集的训练集、验证集和测试集的比例为 8 : 1 : 1。

### 4.2 实验环境

本文实验环境: CPU 型号为 Intel (R) Core (TM) I9-9900KF @ 3.6 GHz, GPU 型号为 NVIDIA GeForce RTX 2060s, CUDA 版本为 11.6, 在 Windows

10 系统环境下进行训练, 使用 Pytorch1.12.0 与 python 3.8 作为深度学习基础环境。

### 4.3 对比实验设计

将本文提出的 YOLOMN 模型与 YOLOv3-tiny, YOLOv4-tiny, YOLOv5, YOLOv7-tiny, YOLOv8 共计 5 种目标检测模型在 RD-fuji 公开数据集与自建数据集上进行对比实验, 采用准确率、召回率、MAP50 (Mean Average Precision 0.5) 和 MAP50-95 作为实验评价指标。在 RD-fuji 公开数据集上的对比实验结果见表 1, 在自建数据集上的对比实验结果见表 2。

表 1 RD-fuji 数据集对比实验结果数据表

Table 1 Comparison of experimental results data of RD-fuji dataset

模型	准确率	召回率	MAP50	MAP50-95	模型推理耗时/ms	模型大小/MB
YOLOv3-tiny	0.912	0.892	0.947	0.665	3.8	23.20
YOLOv4-tiny	0.895	0.898	0.950	0.665	3.3	14.80
YOLOv5	0.909	0.883	0.945	0.661	2.7	5.02
YOLOv7-tiny	0.901	0.900	0.949	0.663	4.6	15.70
YOLOv8	0.892	0.891	0.945	0.668	4.2	5.95
本文模型	0.923	0.857	0.940	0.649	3.6	2.53

表 2 自建数据集对比实验结果数据表

Table 2 Comparison of experimental results data of self our dataset

模型	准确率	召回率	MAP50	MAP50-95	模型推理耗时/ms	模型大小/MB
YOLOv3-tiny	0.952	0.863	0.944	0.678	3.6	23.20
YOLOv4-tiny	0.926	0.861	0.940	0.659	3.5	14.80
YOLOv5	0.926	0.848	0.933	0.658	3.1	5.02
YOLOv7-tiny	0.928	0.900	0.963	0.705	4.5	15.70
YOLOv8	0.935	0.896	0.959	0.690	3.7	5.95
本文模型	0.945	0.913	0.962	0.699	4.2	2.55

从表 1 和表 2 可以看出,本文提出的模型在实现轻量化设计的同时,展现出优异的综合性能。本文模型大小显著小于其他 5 种对比模型,轻量化优势明显;在关键指标 MAP50 和 MAP50-95 上,虽未达到最优水平,但与最优结果的差距在 0.01 以内,保持了较高的检测精度;而在准确率和召回率两项核心指标上,本文模型表现突出,说明本文模型在检

测性能与模型效率间的良好平衡与优势。

#### 4.4 目标检测实验结果与分析

为验证本文所提模型在目标识别任务中的准确性与优越性,分别在 RD-fuji 公开数据集与本文构建的自建数据集上进行目标检测对比实验, RD-fuji 公开数据集上目标检测对比结果见表 3,在自建数据集上的目标检测对比结果见表 4。

表 3 KFuji 公开数据集对比实验目标检测表

Table 3 Comparison of experimental results for object detection of KFuji dataset

模型	YOLOv3-tiny	YOLOv4-tiny	YOLOv5	YOLOv7-tiny	YOLOv8	本文
验证集标签						
验证集识别结果						

表 4 自建数据集对比实验目标检测表

Table 4 Comparison of experimental results for object detection of our dataset

模型	YOLOv3-tiny	YOLOv4-tiny	YOLOv5	YOLOv7-tiny	YOLOv8	本文
验证集标签						
验证集识别结果						

如表 3 与表 4 中对比实验结果所示,可以发现本文提出的 YOLOMN 模型在两个数据集中对苹果目标检测识别数量多,漏标与误标情况很少,说明本文模型能捕捉到果园复杂环境中,苹果目标的重要特征,将其准确地与背景进行区分,从而实现了高精

确率的目标检测。而其他对比模型均存在目标检测中漏标与识别错误过量标识的情况,如 YOLOv8 与 YOLOv7-tiny 模型在两个数据集中有部分漏标, YOLOv3-tiny, YOLOv4-tiny, YOLOv7-tiny 模型将图片中某些复杂的叶片识别为苹果。

本文模型在准确目标检测与定位方面的卓越能力,更凸显了其在复杂果园环境下所展现出的鲁棒性和稳定性;在各种光照条件、枝干遮挡、树叶重叠等复杂因素的影响下,依然能够准确地识别并定位异源果实目标,充分证明了本文模型在实际应用中的潜力和价值。

## 5 结束语

本文提出了一种基于 YOLOv8 的轻量化苹果异源目标检测模型 (YOLOMN),比改进前原模型目标检测精度上有一定提高,并实现了显著的轻量化优化。训练后文件大小仅为改进前文件大小的一半,参数量也大幅减少至原参数量的 30%;在保持精准度的同时进行的轻量化改进,使得模型在实际应用中更加高效和灵活。虽然 YOLOMN 在目标检测方面已经取得了一定成果,但精准度仍有一定的提升空间。在模型的输出部分仍可以继续进行轻量化处理,以进一步提升模型检测的效率,减少参数量。今后工作将致力于进一步优化本文模型,也将关注减少目标检测在训练部分的时间损耗,提高目标检测的反应速率,从而进一步提升模型在实际应用中的效率和效果。

## 参考文献

- [1] HAN Y, LI S, YANG Y, et al. Vision based fruit recognition and positioning technology for harvesting robots [J]. Computers and Electronics in Agriculture, 2023, 213: 108258–108261.
- [2] RACHMAWATI E, SUPRIANA I, KHODRA M L, et al. Integrating semantic features in fruit recognition based on perceptual color and semantic template [J]. Agricultural Information Processing, 2022, 9(2): 316–334.
- [3] ALRESHEEDI K M, ALADHADH S, KHAN R U, et al. Dates fruit recognition: From classical fusion to deep learning [J]. International Journal of Computer Systems Science & Engineering, 2022, 40(1): 151–166.
- [4] SHI Jichao, WANG Ziheng, ZHAO Xianchao, et al. Target detection algorithm based on human judge mechanism [J]. Journal of Shanghai Jiaotong University, 2022, 27 (5): 660–670.
- [5] LIN M, TENG S, CHEN G, et al. Optimal CNN-based semantic segmentation model of cutting slope images [J]. Frontiers of Structural and Civil Engineering, 2022, 16(4): 414–433.
- [6] QIU Q, LAU D. Real-time detection of cracks in tiled sidewalks using YOLO – based method applied to unmanned aerial vehicle (UAV) images [J]. Automation in Construction, 2023, 147: 104745–104759.
- [7] MEKHALFI M L, NICOLO C, BAZI Y, et al. Contrasting YOLOv5, transformer, and efficientdet detectors for crop circle detection in desert [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1–5.
- [8] ZHU W, ZHANG H, EASTWOOD J, et al. Concrete crack detection using lightweight attention feature fusion single shot multibox detector [J]. Knowledge – Based Systems, 2023, 261: 110216–110228.
- [9] JIANG H, LEARNED-MILLER E. Face detection with the faster R – CNN [C]//Proceedings of 2017 12<sup>th</sup> IEEE International Conference on Automatic Face & Gesture Recognition. Piscataway, NJ: IEEE, 2017: 650–657.
- [10] ZHANG Fucheng, YANG Han, HAN Zongwang. Weld defect detection based on improved attention RPN [J]. Modeling and Simulation, 2023, 12 (1): 582–589.
- [11] TONG Z, WU Y, LIU Y. Single – stage multi – scale receptive field improvement lightweight object detection network based on MobileNetV3 [C]// Proceedings of 2022 21<sup>st</sup> International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES). Piscataway, NJ: IEEE, 2022: 103–106.
- [12] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 6848–6856.
- [13] MAHMOOD B A, KURNAZ S. An investigational FW–MPM–LSTM approach for face recognition using defective data [J]. Image and Vision Computing, 2023, 132: 104644.
- [14] MEKHALFI M L, NICOLO C, BAZI Y, et al. Contrasting YOLOv5, Transformer, and EfficientDet detectors for crop circle detection in desert [J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1–5.
- [15] ZHANG Y, CHEN X. Lightweight semantic segmentation algorithm based on MobileNetV3 network [C]// Proceedings of 2020 International Conference on Intelligent Computing, Automation and Systems (ICICAS). Piscataway, NJ: IEEE, 2020: 429–433.
- [16] JI J, ZHU X, MA H, et al. Apple fruit recognition based on a deep learning algorithm using an improved lightweight network [J]. Applied Engineering in Agriculture, 2021, 37 (1): 123–134.
- [17] LIANG X, LIANG J, YIN T, et al. A lightweight method for facial expression recognition based on improved MobileNetV3 [J]. IET Image Process, 2023, 17(8): 2375–2384.
- [18] PRASETYO E, PURBANINGTYAS R, ADITYO R D, et al. Combining MobileNetV1 and depthwise separable convolution bottleneck with expansion for classifying the freshness of fish eyes [J]. Agricultural Information Processing, 2022, 9(4): 485–496.