

自然环境下基于改进的YOLOv8s的脐橙检测方法

罗宇航, 陈骁扬, 汪成江, 罗坤, 黄帅永

江西理工大学信息工程学院, 江西 赣州

收稿日期: 2024年5月19日; 录用日期: 2024年6月15日; 发布日期: 2024年6月25日

摘要

近年来, 基于深度学习的目标检测算法发展极为迅速。YOLOv8s能够兼顾检测精度和速度, 有利于脐橙智能化采摘的实现。本文分别从数据集扩充和YOLOv8s模型改进两方面提高脐橙的检测精度。数据集扩充方面, 对晴天采集的脐橙图像进行加雨和加雾扩充, 提高模型在复杂环境下的检测能力。模型改进方面, 先将YOLOv8s颈部的特征融合改进为权重化的融合, 以突出重要的特征; 然后将YOLOv8s原浅层处的检测头更换至更浅层, 以检测出因遮挡严重而表现为小的目标; 最后, 将YOLOv8s主干网络的第二个卷积改进为全维度动态卷积, 以在卷积的空间、输入通道、输出通道和整个卷积核四个维度上关注重要特征。实验结果表明, 与未对数据集扩充且采用原始YOLOv8s模型的方法相比, 本文方法获得的精确率、召回率和平均精确率均值都得到大幅度提升。

关键词

YOLOv8s, 特征融合, 检测头, 全维度动态卷积, 脐橙检测

Detection of Navel Orange Based on the Improved YOLOv8s under Natural Environment

Yuhang Luo, Xiaoyang Chen, Chengjiang Wang, Kun Luo, Shuaiyong Huang

School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi

Received: May 19th, 2024; accepted: Jun. 15th, 2024; published: Jun. 25th, 2024

Abstract

In recent years, deep learning-based object detection algorithms have developed rapidly. YOLOv8s

文章引用: 罗宇航, 陈骁扬, 汪成江, 罗坤, 黄帅永. 自然环境下基于改进的YOLOv8s的脐橙检测方法[J]. 计算机科学与应用, 2024, 14(6): 41-49. DOI: 10.12677/csa.2024.146140

can balance detection accuracy and speed, which is conducive to the implementation of intelligent navel orange picking. In this paper, we improve the detection performance from two aspects: dataset expansion and model improvement. In terms of dataset expansion, rain and fog are added into the navel orange images collected on sunny days to improve the detection ability of the model in complex environments. In terms of model improvement, the weighed concatenation operation is adopted in feature fusion to highlight the important features; the detection head at the shallow layer in the original model is transferred to a shallower layer to detect targets that appear small due to the severe occlusion; the second convolution operation in the backbone is replaced by a omni-dimensional convolution to focus on important features in four dimensions: spatial position, input channel, output channel and the entire convolution kernel. The experimental results show that when compared with the method using the original YOLOv8s without data augmentation, the precision, the recall and the mean average precision obtained by our method have been significantly improved.

Keywords

YOLOv8s, Feature Fusion, Detection Head, Omni-Dimensional Dynamic Convolution, Navel Orange Detection

Copyright © 2024 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

脐橙不仅营养成分高, 而且具有预防多种疾病的功效, 其种植地已遍布全球, 但脐橙的智能化采摘仍面临着各种难题。其中, 脐橙的高精度实时检测问题亟待解决。近年来, 基于深度学习的目标检测技术得到了深入发展, 主要分为双阶段和单阶段两大类[1] [2] [3] [4]。其中, 双阶段指的是先提取目标候选区域后分类, 典型的模型有 R-CNN、Fast R-CNN 和 Faster R-CNN 等; 单阶段指的是一次性完成分类和回归, 典型的模型有 SSD 和 YOLO 系列等。在单阶段检测技术中, YOLO 系列模型因其检测速度优势显著, 在水果检测中发挥着重要的作用。薛等人先对数据集做了增强和扩充处理, 然后将密集模块引入到 YOLOv2 中, 有效地提高了未成熟芒果的检测性能[4]。唐等人通过去除 YOLOv3 中大尺度目标检测模块, 总体上提高了百香果的检测效率; 通过在 YOLOv3 中增加 DenseNet 模块, 提高了中等尺寸百香果的检测精度[5]。王等人对 YOLOv4 中的特征提取模块轻量化, 并引入注意力机制到特征融合模块, 实现了自然环境下苹果的高精度实时检测[6]。张等人首先对数据进行离线与在线增强, 然后将 Transform 思想引入 YOLOv5 的主干网络, 并将双向特征金字塔网络(Bi-directional Feature Pyramid Network, BiFPN)引入到颈网络, 有效地提高了樱桃的检测精度[7]。杜等人首先对数据集进行增强处理, 然后将全局注意力机制引入到 YOLOv8 中, 实现了复杂环境下苹果的高精度实时检测[8]。

上述复杂自然环境下基于 YOLO 系列的水果检测面临的共同问题有: 1) 数据采集过程中存在各种光线不佳的问题; 2) 图像中存在果实被树叶遮挡和果实之间相互遮挡问题; 3) 采集的数据集有限问题。解决这些问题的主要方法是数据集的增强和扩充, 以及 YOLO 模型的改进。这些问题和解决思路同样适用于脐橙目标。熊等人为了去除光照等噪声, 采用导向滤波处理; 为了提高脐橙检测精度, 对 YOLOv3 中的尺度融合模块和损失函数分别改进[9]。章等人先对采集的脐橙图像进行了随机畸变、混合和随机裁剪等增强处理, 然后对 PP-YOLO 模型进行特征提取和融合网络的改进, 同时结合 K-means 算法获取适合脐橙尺寸的检测框, 有效地提高了各种光线下的检测精度和速度[10]。本文采用能够兼顾精度和速度的

YOLOv8s 开展脐橙检测研究, 为了提高模型在不同光照和各种噪声干扰下的检测能力, 对晴天采集的数据集进行加雨和加雾扩充; 为了提高脐橙被树叶和其他果实遮挡情况下的检测能力, 对 YOLOv8s 的主干网络、颈网络和检测头进行改进, 实验结果验证了本文方法的有效性。

2. 数据集的扩充

本文采用的脐橙数据集为天气晴朗时拍摄, 图像采集过程中不可避免地存在侧光和逆光等情况。某些脐橙图像虽然为顺光拍摄, 但一些被树叶遮挡的区域也表现为暗区域。为了提高各种光照下脐橙图像中所有区域内的脐橙检测性能, 采用加雨和加雾对数据集进行扩充。

2.1. 加雨扩充

图像加雨的过程其实是加噪, 不同密度的随机噪声对应于不同的雨量。对脐橙图像加噪后, 一定程度上增加了脐橙目标检测的难度, 因此, 加雨扩充有助于提高模型的检测能力。

将雨的生成当成随机噪声的生成, 则脐橙图像加雨扩充的主要步骤如下:

- 1) 对待加雨脐橙图像 I_0 , 获取其分辨率, 并生成该分辨率下的二维随机噪声图像 N_0 ;
- 2) 对 N_0 中的随机噪声拉长、旋转, 以获取不同大小和方向的雨滴噪声, 并形成二维噪声图像 N_1 ;
- 3) 将生成的雨滴噪声图像 N_1 和原始脐橙图像 I_0 进行叠加, 得到包含雨滴的脐橙图像 I_1 。

给定任意一幅脐橙图像, 如图 1(a)所示。按照该图像的分辨率生成二维随机噪声图像, 如图 1(b)所示。将图 1(b)中的噪声拉长和旋转处理后, 得到近似的雨滴图像, 如图 1(c)所示。最后将雨滴图像和原始脐橙图像相叠加, 得到的加雨脐橙图像如图 1(d)所示。

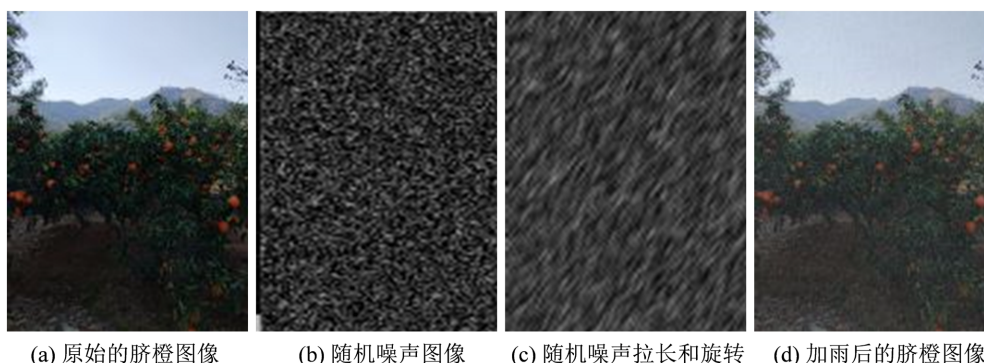


Figure 1. Process of adding rain to an image
图 1. 图像加雨过程

2.2. 加雾扩充

采用标准的光学模型[11]对图像加雾, 如图 2 所示。由于相机获取有雾图像时的光线一部分来自于穿透雾的目标反射光, 另一部分来自于被雾反射的大气光, 因此, 有雾图像为两部分之和:

$$I_2(x, y) = I_0(x, y)t(x, y) + A(1 - t(x, y)) \quad (1)$$

其中, (x, y) 为图像中各像素的空间坐标位置, t 表示图像中各像素对应的光透射率, I_0 表示无雾图像, I_2 表示有雾图像, A 为大气光。

采用中心点合成雾的思想对图像加雾。假设 α 表示散射系数, D 表示图像中心像素到其他像素的距离, 则光透射率 t 的计算表达式为:

$$t(x, y) = \exp(-\alpha D(x, y)) \quad (2)$$

显然，离雾中心距离越远的像素，雾合效果越弱。图像加雾的主要步骤如下：

- 1) 将图像的中心像素作为雾化中心，计算图像中心到其他像素之间的距离 D ；
- 2) 根据公式(2)计算有雾情况下的光透射率；
- 3) 选取合理的 A ，根据公式(1)生成加雾图像。

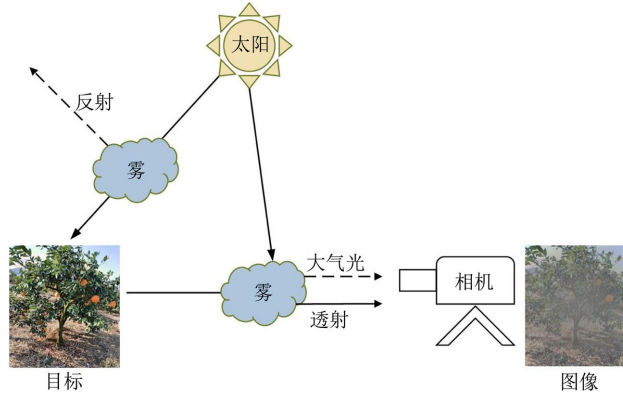


Figure 2. Atmospheric scattering model
图 2. 大气散射模型

3. YOLOv8s 的改进

YOLOv8 按照模型的深度和宽度又可分为 n/s/m/l/x 五种。本文综合考虑脐橙检测精度和速度，选用其中的 YOLOv8s 并对其改进，即在原始的 YOLOv8s 基础上，进行三个方面的改进。首先，对颈网络(Neck)改进，将其路径聚合网络(Path Aggregation Network, PAN)和特征金字塔网络(Feature Pyramid Network, FPN)中采用的拼接(Concat)方式改为加权的拼接(Weighted Concat, WConcat)。然后，将浅层检测头(Detect)更换至更浅层；最后，将主干网络(Backbone)中的第二个卷积(Conv)改为全维度动态卷积(Omni-Dimensional Dynamic Convolution, ODConv)。改进后的 YOLOv8s 主体结构和各模块如图 3 所示，其中红色虚线框表示相应的三个改进点。

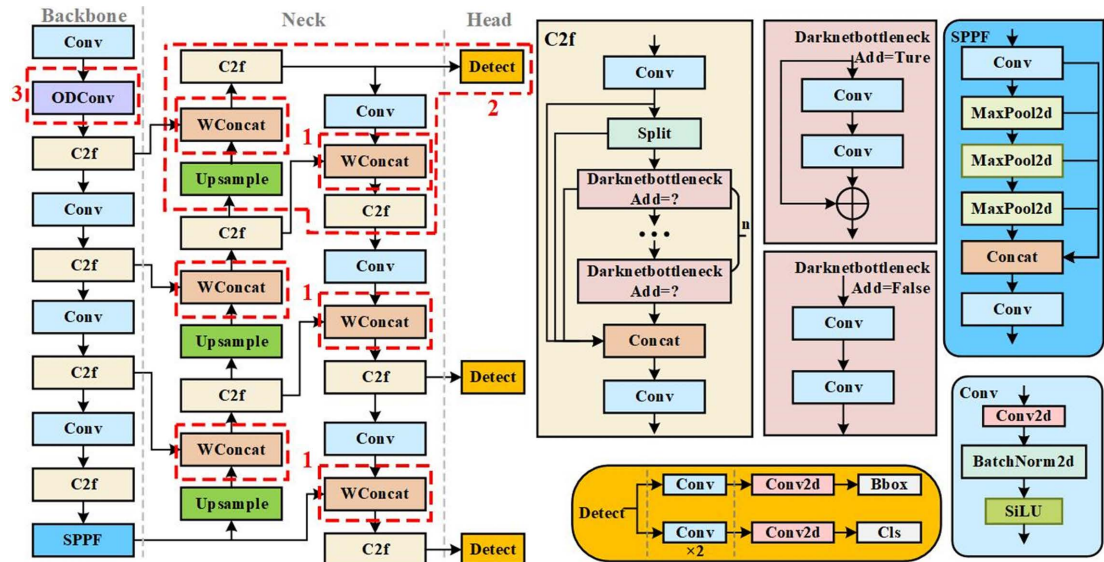


Figure 3. Structure of the improved YOLOv8s
图 3. 改进后的 YOLOv8s 结构

3.1. WConcat

受文献[12]中 BiFPN 的启发, 将 PAN-FPN 中的 Concat 用 WConcat 替换, 如图 3 中 6 个标记为 1 的红色虚线框。事实上, 不同分支输入特征对输出结果的贡献不一样, 采用 WConcat 对不同分支输入特征进行不同的加权, 突出重要特征的作用。进一步地, 为防止模型不稳定, 对权值进行归一化。

若特征融合前, 来自两条不同分支的特征分别记为 \mathbf{F}_1 和 \mathbf{F}_2 , 则融合后的特征 \mathbf{O} 表示为:

$$\mathbf{O} = \left[\frac{w_1}{\sum_{j=1}^2 w_j + \varepsilon} \cdot \mathbf{F}_1, \frac{w_2}{\sum_{j=1}^2 w_j + \varepsilon} \cdot \mathbf{F}_2 \right] \quad (3)$$

其中, $w_i \geq 0$ ($i = 1, 2$)且 ε 为任意小的常数。

3.2. Detect

将原浅层 Detect 更换至更浅层, 如图 3 中标记为 2 的红色虚线框, 原因是更浅层包含了丰富的目标形状和位置等信息, 有利于模型检测出被树叶或其他橙遮挡严重的小目标。为了使得 Detect 的输入为更浅层特征, 将 Backbone 的第一个跨阶段局部网络融合模块(Cross Stage Partial Network Fusion, C2f)提取出的特征作为 Neck 的一部分输入。此外, 在 Neck 的 FPN 部分增加一个上采样(Upsample)、WConcat 和 C2f 层; PAN 部分增加一个 Conv、WConcat 和 C2f 层。

3.3. ODConv

将 Backbone 中的第二个 Conv 改为 ODConv, 如图 3 中标记为 3 的红色虚线框。数学上, 动态卷积可以定义为:

$$y = (\alpha_{w1}W_1 + \dots + \alpha_{wn}W_n) * x \quad (4)$$

其中, $x \in \mathcal{R}^{h \times w \times c_{in}}$ 表示输入特征, $y \in \mathcal{R}^{h \times w \times c_{out}}$ 表示输出特征, 输入和输出特征图的高和宽分别为 h 和 w , 输入特征通道数为 c_{in} , 输出特征通道数为 c_{out} ; W_i 表示第 i 个卷积核; $\alpha_{wi} \in \mathcal{R}$ 是用于加权 W_i 的注意力标量; $*$ 代表卷积运算。

进一步地, ODConv [13]可以定义为:

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) * x \quad (5)$$

其中, α_{si} 、 α_{ci} 、 α_{fi} 和 α_{wi} 分别表示沿着卷积核的空间位置、输入通道、输出通道和整个卷积核四个维度的注意力权重。 \odot 代表相乘运算。

在动态卷积中, 注意力标量 $\alpha_{wi} \in \mathcal{R}$ 是在输入特征上计算注意力函数得到, 如图 4(a)所示。在 ODConv 中, $\alpha_{si} \in \mathcal{R}^{k \times k}$ ($k \times k$ 表示空间核尺寸)、 $\alpha_{ci} \in \mathcal{R}^{c_{in}}$ 、 $\alpha_{fi} \in \mathcal{R}^{c_{out}}$ 和 α_{wi} 同样是在输入特征上计算多注意力函数, 如图 4(b)所示。

在图 4(b)中, 输入特征 x 被挤压成一个长度为 c_{in} 的特征矢量, 然后通过逐通道的全局平均池化(Global Average Pooling, GAP)、全连接层(Fully Connected, FC)、非线性 ReLU 激活和四个头分支计算得到四个维度的注意力权重。其中, 四个头分支各包含一个 FC 层和 Sigmoid(或 Softmax)函数; 各分支 FC 层的输出尺寸分别为 $k \times k$ 、 $c_{in} \times 1$ 、 $c_{out} \times 1$ 和 $n \times 1$; 各分支 Sigmoid(或 Softmax)函数分别用于生成归一化的注意力系数 α_{si} 、 α_{ci} 、 α_{fi} 和 α_{wi} 。

四种注意力系数 α_{si} 、 α_{ci} 、 α_{fi} 和 α_{wi} 乘以卷积核的过程分别如图 5(a)~(d)所示。图 5(a)中, α_{si} 赋予不同的注意力到 $k \times k$ 卷积空间; 图 5(b)中, α_{ci} 赋予不同的注意力到 c_{in} 输入通道; 图 5(c)中, α_{fi} 赋予不

同的注意力到 c_{out} 输出通道；图 5(d)中， α_{wi} 赋予给整个卷积核，不同的卷积核具有不同的 α_{wi} 。四种类型的注意力互补，有利于捕获丰富的上下文信息。

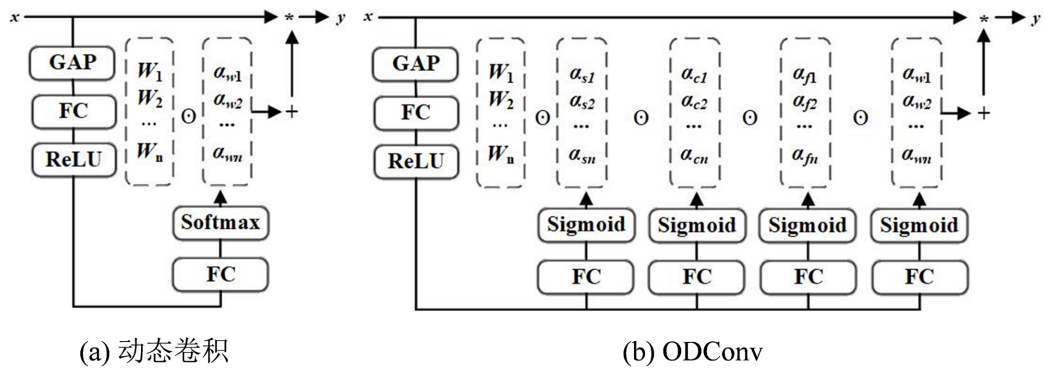


Figure 4. Principles of dynamic convolution and omnidirectional dynamic convolution
图 4. 动态卷积和 ODConv 原理

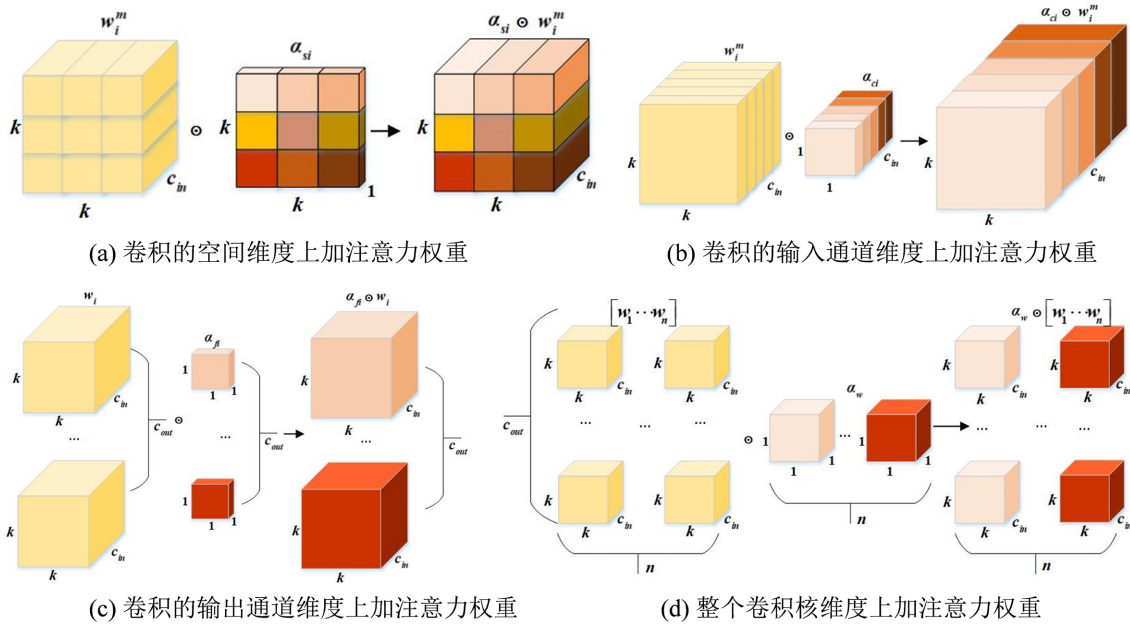


Figure 5. Adding attention weights to four dimensions in omnidirectional dynamic convolution
图 5. ODConv 中四个维度上加注意力权重

4. 实验

实验共采集了 1881 幅脐橙图像，按照 4:1 随机分成训练集和验证集。根据加雨和加雾步骤对训练样本扩充，得到的总训练样本数为 2631。采用精确率(Precision, P)、召回率(Recall, R)和所有类别平均精确率的均值(Mean Average Precision, mAP)作为性能评价指标。mAP50 是指预测框和真实框的交并比(Intersection over Union, IoU)为 50%时对应的 mAP 值；mAP50-95 是指在 mAP50 和 mAP95 之间以一定间隔取值后计算所有值的平均。

4.1. 消融实验结果

消融实验时，基准方法是指采用原始的 YOLOv8s 对原脐橙数据检测。在此基础上，先进行加雨和加

雾扩充训练样本数, 后将 PAN-FPN 特征融合中的 Concat 换成 WConcat, 再更换浅层 Detect 位置, 最后将主干网络中的第二个 Conv 换成 ODConv, 各阶段实验结果如表 1 所示。在加雨和加雾扩充训练样本数后, P、R、mAP50 和 mAP50-95 分别提升了 0.7%、1%、0.7% 和 0.7%, 即四项指标均得到大幅度提升, 因此, 增加数据集的辨别难度有利于提高模型的检测能力。将 Concat 更改为 WConcat 后, 四项指标依次提升了 0.7%、0.2%、0.1% 和 0.3%, 此时 P 的提升幅度大; 将浅层的 Detect 更换至更浅层后, 四项指标依次提升了 0.2%、0.5%、0.2% 和 0.4%, 此时 R 的提升幅度大; 再将卷积换成 ODConv 后, 四项指标依次提升了 0.6%、0.1%、0.2% 和 0.9%, 此时 P 和 mAP50-95 提升幅度大。显然, 模型的每一次改进, 均能有效提高脐橙的检测性能。此外, 不同的改进方式所带来的指标提升情况有所不同, 原因是 WConcat 和 ODConv 均用于获取重要的目标特征, 而更换 Detect 位置用于检测出被遮挡的小目标。总体上, 本文提出的数据集扩充和模型改进方法在四项指标上比基准方法依次提升了 2.2%、1.8%、1.2% 和 2.3%, 具有更好的检测性能。

Table 1. Results of ablation experiment

表 1. 消融实验结果

YOLOv8s	扩充数据	WConcat	Detect	ODConv	P	R	mAP50	mAP50-95
√					0.918	0.896	0.944	0.724
√	√				0.925	0.906	0.951	0.731
√	√	√			0.932	0.908	0.952	0.734
√	√	√	√		0.934	0.913	0.954	0.738
√	√	√	√	√	0.94	0.914	0.956	0.747

4.2. 实验结果对比分析

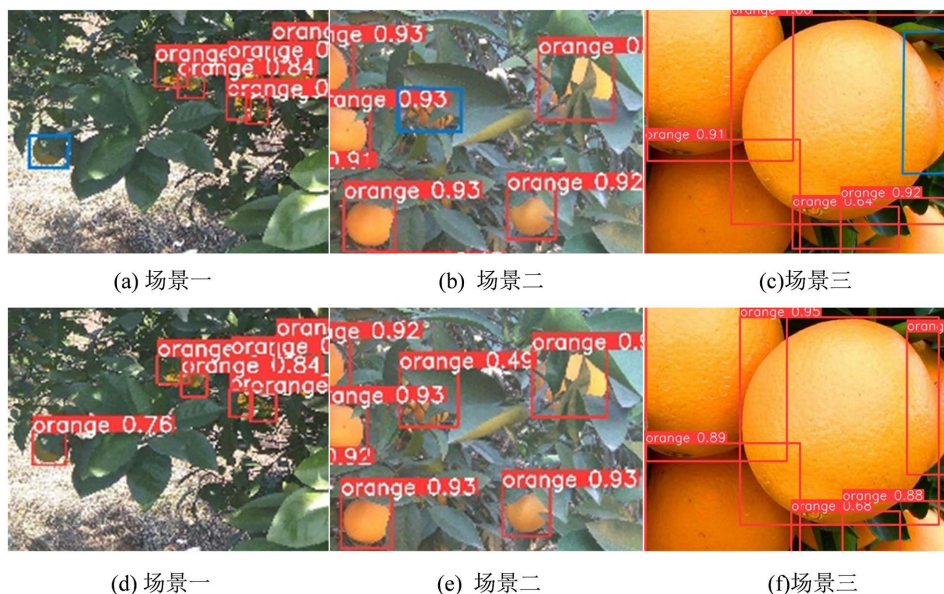


Figure 6. Comparison of detection results between the baseline and our method in three scenarios

图 6. 三种场景下基准方法和本文方法的检测结果对比

挑选自然环境下三个复杂场景, 将本文方法和基准方法的检测结果进行可视化对比。基准方法的结

果分别如图 6(a)~(c)所示, 本文方法的结果分别如图 6(d)~(f)所示。第一个场景光线不好, 且脐橙被树叶遮挡, 如图 6(a)和图 6(d)所示; 第二个场景整体光线较佳, 但存在脐橙被树叶遮挡非常严重的情况, 如图 6(b)和图 6(e)所示; 第三个场景不仅光线好, 而且脐橙没有被树叶遮挡, 但脐橙之间相互遮挡比较严重, 如图 6(c)和图 6(f)所示。显然, 基准方法难以检测出的脐橙(图 6(a)~(c)中蓝色方框区域内)能被本文方法正确检测出。

将基准方法和本文方法在模型训练过程中的四项指标变化情况进行对比, 分别如图 7(a)~(d)所示, 蓝色表示基准方法, 橙色表示本文方法。结果表明, 当训练轮数小于 100 时, 基准方法和本文方法的变化曲线基本上重合; 当训练轮数大于 100 时, 本文方法获得的四项指标明显比基准方法高; 当训练轮数大于 500 轮时, 两种方法获得的变化曲线均能收敛, 且本文方法获得的四项指标均比基准方法更高, 这与表 1 的结论一致。

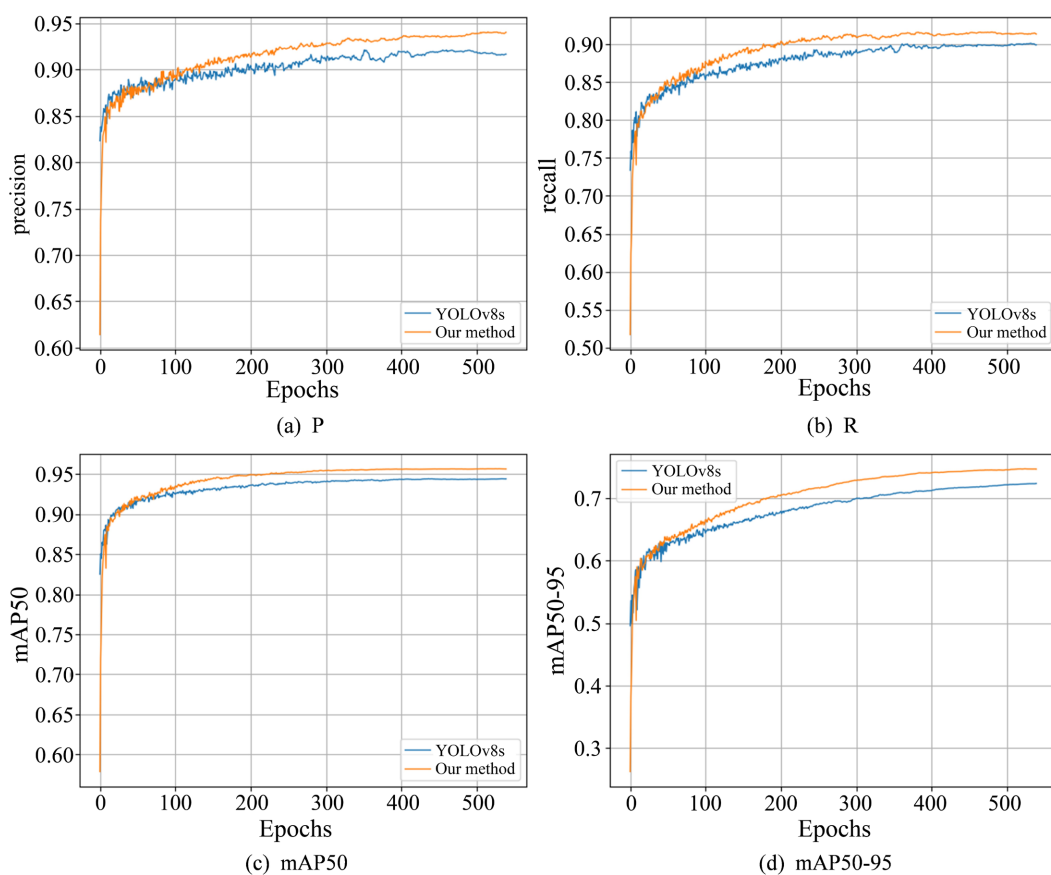


Figure 7. Four metrics curves obtained by the baseline and our method during the model training

图 7. 模型训练时基准方法和本文方法获得的四种性能指标变化曲线

5. 结论

本文选用 YOLOv8s 作为脐橙检测的基本模型, 并针对自然环境下脐橙检测面临的问题, 分别从数据集扩充和基本模型改进两方面提高脐橙检测精度。首先, 介绍了对晴天采集的脐橙训练样本进行加雨和加雾扩充的原理。然后, 详细介绍了 YOLOv8s 的改进方法, 将 Neck 的特征融合 Concat 改进为 WConcat; 将原浅层处的 Detect 更换至更浅层; 将 Backbone 的第二个 Conv 改进为 ODConv。最后, 开展了脐橙检测实验, 并对实验结果进行了分析。实验结果表明, 加雨和加雾扩充能够提高模型的检测能力, 且模型

的三个改进之处均有助于提升检测性能。与数据集不扩充且采用原始的 YOLOv8s 基准方法相比,本文方法获得的四项性能指标均得到大幅度提升。

基金项目

大学生创新创业训练计划项目: 202110407029, 以及江西省自然科学基金: 20224BAB202002。

参考文献

- [1] 朱旭, 马溟, 姬江涛, 金鑫, 赵凯旋, 张开. 基于 Faster R-CNN 的蓝莓冠层果实检测识别分析[J]. 南方农业学报, 2020, 51(6): 1493-1501.
- [2] 熊俊涛, 刘振, 汤林越, 林睿, 卜榕彬, 彭红星. 自然环境下绿色柑橘视觉检测技术研究[J]. 农业机械学报, 2018, 49(4): 45-52.
- [3] 彭红星, 黄博, 邵园园, 等. 自然环境下多类水果采摘目标识别的通用改进 SSD 模型[J]. 农业工程学报, 2018, 34(16): 155-162.
- [4] 薛月菊, 黄宁, 涂淑琴, 等. 未成熟芒果的改进 YOLOv2 识别方法[J]. 农业工程学报, 2018, 34(7): 173-179.
- [5] 唐熔钗, 伍锡如. 基于改进 YOLO-V3 网络的百香果实时检测[J]. 广西师范大学学报, 2020, 38(6): 32-39.
- [6] 王卓, 王健, 王泉雄, 时佳, 白晓平, 赵泳嘉. 基于改进 YOLOv4 的自然环境苹果轻量级检测方法[J]. 农业机械学报, 2022, 53(8): 294-302.
- [7] 张志远, 罗铭毅, 郭树欣, 刘刚, 李淑平, 张瑶. 基于改进 YOLOv5 的自然环境下樱桃果实识别方法[J]. 农业机械学报, 2022, 53(1): 232-240.
- [8] 杜宝侠, 唐友, 辛鹏, 杨牧. 基于改进 YOLOv8 的苹果检测方法[J]. 无线互联科技, 2023(13): 119-122.
- [9] 熊正午, 吴瑞梅, 黄俊仕, 李霸聪, 戴仕明, 艾施荣. 深度学习结合快速导向滤波识别自然环境下脐橙果实[J]. 江西农业大学学报, 2022, 44(3): 736-746.
- [10] 章倩丽, 李秋生, 胡俊勇, 谢湘慧. 基于 PP-YOLO 改进算法的脐橙果实实时检测[J]. 北京联合大学学报, 2022, 36(4): 58-66.
- [11] He, K., Sun, J. and Tang, X. (2011) Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **33**, 2341-2353. <https://doi.org/10.1109/TPAMI.2010.168>
- [12] Tan, M., Pang, R. and Le, Q. (2020) EfficientDet: Scalable and Efficient Object Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 10778-10787. <https://doi.org/10.1109/CVPR42600.2020.01079>
- [13] Li, C., Zhou, A. and Yao, A. (2022) Omni-Dimensional Dynamic Convolution. *International Conference on Learning Representations*, Virtual, 25-29 April 2022, 1-20.