

李 颀, 陈哲豪. 基于改进单次多目标检测器的果面缺陷冬枣实时检测[J]. 江苏农业学报, 2022, 38( 1 ): 119-128.

doi: 10.3969/j.issn.1000-4440.2022.01.014

## 基于改进单次多目标检测器的果面缺陷冬枣实时检测

李 颀<sup>1</sup>, 陈哲豪<sup>2</sup>

(1. 陕西科技大学电子信息与人工智能学院, 陕西 西安 710021; 2. 陕西科技大学电气与控制工程学院, 陕西 西安 710021)

**摘要:** 为实现果面缺陷冬枣实时检测, 并解决缺陷的尺寸与位置不同影响检测精度的问题, 提出一种基于改进单次多目标检测器(Single shot multibox detector, SSD)的果面缺陷冬枣实时检测方法。以陕西大荔冬枣中的虫蛀、轮纹和木质化3种缺陷果和正常果为研究对象, 在数据采集设备下采集实际分拣图像, 然后通过数据增强由400张扩充至2 000张。改进SSD, 建立MobileNetV3-SSD模型, 为实时检测奠定基础; 引入改进感受野块(RFB)可实现模型多尺寸提取冬枣缺陷特征的能力; 用空间注意力模块(SAM)代替挤压和激励通道注意力模块(SE)增强模型定位冬枣缺陷特征的能力。试验结果表明, 本研究模型在果面缺陷冬枣数据集上的表现均优于目前先进目标检测网络模型(RetinaNet和EfficientDet-D0), 该模型对4类冬枣的整体检测精准性(*mAP*)达到91.89%, 检测速度达到1 s 40.85帧。因此本研究模型较好地平衡了实时性和精准性, 可应用于果面缺陷冬枣分拣流水线。

**关键词:** 冬枣; 果面缺陷; 实时检测; 单次多目标检测器; 多尺寸; 空间注意力模型

**中图分类号:** TP391.4

**文献标识码:** A

**文章编号:** 1000-4440(2022)01-0119-10

## Real-time surface defect detection of winter jujube based on improved single shot multibox detector

LI Qi<sup>1</sup>, CHEN Zhe-hao<sup>2</sup>

(1. School of Electronic Information and Artificial Intelligence, Shaanxi University of Science & Technology, Xi'an 710021, China; 2. School of Electrical and Control Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China)

**Abstract:** In order to realize the real-time surface defect detection of winter jujube and solve the problems that different sizes and positions affected the detection accuracy, a real-time surface defect detection method of winter jujube based on improved single shot multibox detector (SSD) was proposed. Three kinds of defective winter jujubes (worm, wheel-pattern and lignification) and normal winter jujubes from Dali (Shaanxi province) were taken as the research objects. The actual sorting images were collected by data acquisition equipment, and then expanded from 400 to 2 000 by data enhancement. The SSD was improved, and MobileNetV3-SSD model was established to lay the foundation for real-time detection. The introduction of improved receptive field block (RFB) could realize the ability of model to extract the defect feature of winter jujube at multiple scales. Spatial attention module (SAM) was used to replace squeeze-and-excitation (SE) block, so the ability of the model to locate the defect feature of winter jujube was enhanced. The test results showed that the performance of the proposed model on the dataset of defective winter jujube was better than the current advanced target detection network models (RetinaNet and EfficientDet-D0). The overall detection accuracy of the model for four

types of winter jujube was 91.89%, and the detection speed was 40.85 frames per second. Therefore, the model established in this study can balance the real-time performance and accuracy, and can be applied to sorting pipeline of winter jujube with surface defect.

**Key words:** winter jujube; surface defect; real-time detection; single shot multibox detector; multi-scale; spatial attention module

收稿日期: 2021-09-08

基金项目: 陕西省农业科技创新工程项目[201806117YF05NC13(1)]; 陕西省科技厅农业科技攻关项目(2015NY028); 陕西科技大学博士科研启动基金项目(BJ13-15)

作者简介: 李 颀(1973-), 女, 陕西西安人, 博士, 教授, 主要研究方向为农业智能化、信息化深度学习。(E-mail) liqidq@sust.edu.cn

通讯作者: 陈哲豪, (E-mail) 773905920@qq.com

2020 年陕西大荔冬枣产量超  $5 \times 10^5 \text{ t}$ <sup>[1]</sup>。庞大的产量带来了巨大的果面缺陷冬枣分拣量,而人工应对巨大分拣量时精度骤降且效率低下<sup>[2-3]</sup>。此外采后冬枣常温下易失水的缺点会导致其品质受损,所以需要尽快完成分拣以冷藏保鲜,可见实现果面缺陷冬枣实时检测至关重要<sup>[4-5]</sup>。

常规的果面缺陷检测方法包括传统机器学习和深度学习两类。Lu 等<sup>[6]</sup>使用主成分分析获得苹果第一主成分图像,基于类内方差法分割图像,根据非零像素判断苹果损伤,准确率为 84.20%。孙世鹏等<sup>[7]</sup>利用降维算法确定冬枣损伤处一致的特征波长,经处理获得一致波长灰度图,并用支持向量机分类,检测准确率为 81.80%。

显然传统机器学习的精度与所提取的果面缺陷特征息息相关。但是冬枣果面缺陷尺寸不一且位置随机,传统机器学习的精度也会因特征提取难度的升高而下降<sup>[8]</sup>。然而深度学习能够挖掘到人类难以发现的有效特征,从而在果面缺陷检测任务中脱颖而出。

薛勇等<sup>[9]</sup>使用 GoogLeNet 自动提取苹果果面缺陷特征,对测试样本检测准确率达 91.91%。费琦琪等<sup>[10]</sup>利用 CNN 自主提取冰糖橙缺陷特征,对测试样本检测准确率达 96.67%。李小占等<sup>[11]</sup>通过改进 VGG 网络训练哈密瓜表面缺陷检测模型,检测准确率为 93.50%,静态下处理单幅图像需 0.7 s。虽然深度学习在上述任务中表现出色,但是卷积神经网络的巨大计算量导致模型实时性受限。因此果面缺陷目标检测模型的实时性和精准性的平衡问题亟待解决。

综上所述,本研究以产自陕西大荔的虫蛀、轮纹和木质化 3 种果面缺陷冬枣和正常冬枣为试验材料,提出基于改进单次多目标检测器 (Single shot multibox detector, SSD) 的果面缺陷冬枣实时检测模型,对缺陷果和正常果进行实时检测,使检测模型兼具实时性和精准性。

## 1 果面缺陷冬枣实时检测模型

### 1.1 SSD 目标检测网络

单位时间内检测果面缺陷冬枣个数是决定模型实时性的主要因素之一,这就要求选择的网络能应用于图中存在多个冬枣目标的检测任务,因此果面缺陷冬枣实时检测模型应选择目标检测网络。目前

目标检测网络分为单阶段检测网络和双阶段检测网络。双阶段检测网络原理是先获得若干个建议框,最后在建议框中筛选出预测框。而单阶段检测网络则跳过获得建议框的阶段,直接获得预测框。因此单阶段网络相较于双阶段网络的实时性有一定提升。Liu 等<sup>[12]</sup>学习 YOLO 的回归理念,提出先验框预测机制,使 SSD 具有出色的实时性和精准性。

### 1.2 MobileNetV3 主干网络

VGG16 的计算量高达 14.2 G FLOPs,因此 VGG16 作为主干网络与 SSD 搭配存在计算量过大的缺点,这会减缓果面缺陷冬枣实时检测模型的检测速度<sup>[13]</sup>。但是检测果面缺陷冬枣的速度是决定模型实时性的另一个主要因素,为了进一步加强模型的实时性,本研究用 MobileNetV3 代替 VGG16 作为主干网络与 SSD 搭配。

一方面 MobileNetV3 通过使用深度可分离卷积代替大量标准卷积,有效压缩了果面缺陷冬枣实时检测模型的计算量,因而模型的实时性得到加强<sup>[14]</sup>。

深度可分离卷积与标准卷积的区别是,前者将后者的卷积操作分为 2 步来执行:

(1) 对输入特征图各通道应用一个滤波器进行深度卷积但不合成特征图。

(2) 利用逐点卷积计算输入特征图的线性组合构建输出特征图。

深度可分离卷积由深度卷积和逐点卷积两部分组成<sup>[15]</sup>。深度卷积和逐点卷积的计算量分别如公式(1)和公式(2)所示。

$$N_{\text{Depthwise}} = D_K \times D_K \times M \times D_F \times D_F \quad (1)$$

式中  $N_{\text{Depthwise}}$  是深度卷积计算量,  $D_K$  是输入特征图大小,  $M$  是输入特征图通道数,  $D_F$  是输出特征图大小。

$$N_{\text{Pointwise}} = M \times N \times D_F \times D_F \quad (2)$$

式中  $N_{\text{Pointwise}}$  是逐点卷积计算量,  $M$  是输入特征图通道数,  $N$  是输出特征图通道数,  $D_F$  是输出特征图大小。

标准卷积计算量如公式(3)所示。

$$N_{\text{STD}} = D_K \times D_K \times M \times N \times D_F \times D_F \quad (3)$$

式中  $N_{\text{STD}}$  是标准卷积计算量,  $D_K$  是输入特征图大小,  $M$  是输入特征图通道数,  $N$  是输出特征图通道数,  $D_F$  是输出特征图大小。

根据推导,深度可分离卷积与标准卷积的计算

量之比如公式(4)所示。

$$\frac{N_{\text{Depthwise}} + N_{\text{Pointwise}}}{N_{\text{STD}}} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} \quad (4)$$

MobileNetV3 中,  $D_K = 3$  或  $D_K = 5$ , 当  $D_K = 3$  时, 公式(4)比例关系近 1:9, 可见使用深度可分离卷积极大地减少了模型的计算量。

另一方面, MobileNetV3 通过使用倒残差模块, 提高模型学习冬枣果面特征的效率, 减少模型冗余计算。

倒残差模块由扩张层、深度可分离卷积层和投

影层组成。其设计思路是:为了深度可分离卷积层在高维度特征图中学习到更多冬枣果面缺陷特征, 通过使扩张层中的扩张系数  $factor = 6$  升高特征图维度, 为了有效降低模型向下传递特征图时的计算量, 又通过投影层压缩特征图维度。

本研究将 MobileNetV3 的池化层及其之后的卷积层删除, 然后与 SSD 拼接组成 MobileNetV3-SSD。MobileNetV3-SSD 将第 13 个倒残差模块和池化层之前的卷积层的输出作为预测特征层, 最后通过预测特征层和 SSD 附加层得到预测结果(图 1)。

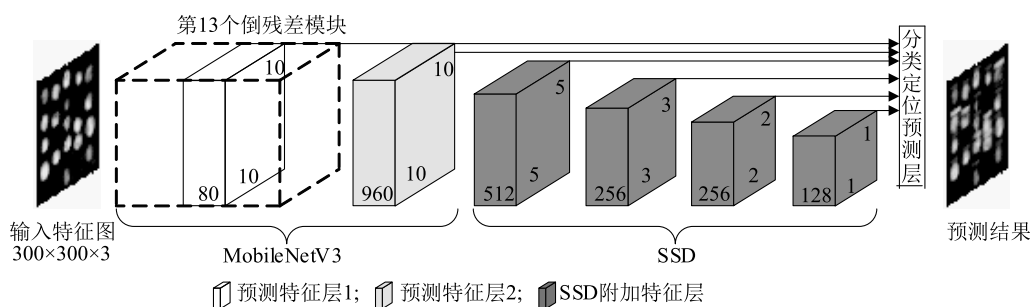


图 1 MobileNetV3-SSD 框架图

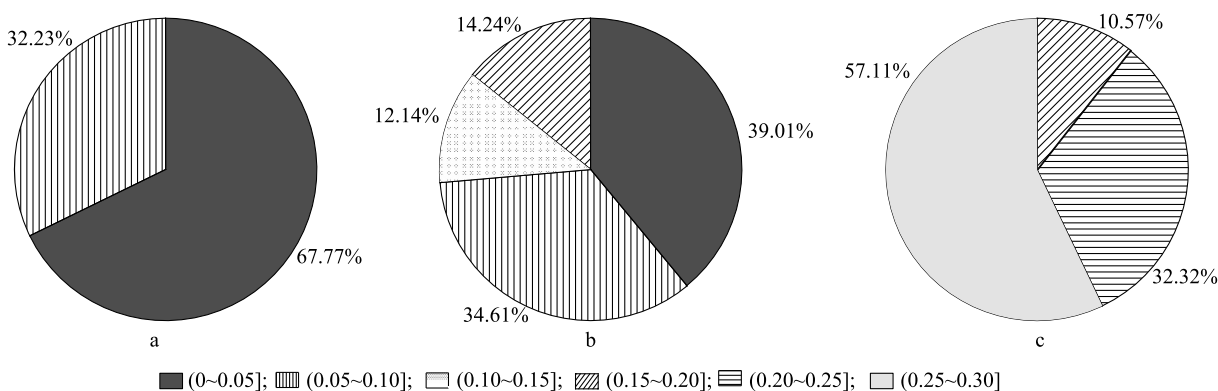
Fig.1 Framework diagram of MobileNetV3-SSD

### 1.3 改进感受野块

本研究借助 OpenCV 库的 ContourArea 函数计算冬枣果面缺陷面积与果面总面积的比例来分析冬枣果面缺陷尺寸。由图 2 可以看出, 虫蛀、轮纹和木质化的尺寸位于 6 个区间内, 分别是  $(0, 0.05]$ ,  $(0.05, 0.10]$ ,  $(0.10, 0.15]$ ,  $(0.15, 0.20]$ ,  $(0.20,$

$0.25]$  和  $(0.25, 0.30]$ , 然而单尺寸感受野无法胜任提取不同尺寸冬枣果面缺陷特征的任务<sup>[16]</sup>。

针对上述问题, 本研究将改进的感受野块 (Receptive field block, RFB) 引入果面缺陷冬枣实时检测模型, 让模型从 3 个尺寸提取冬枣果面缺陷特征, 旨在提高模型分类果面缺陷冬枣的精度<sup>[17]</sup>。



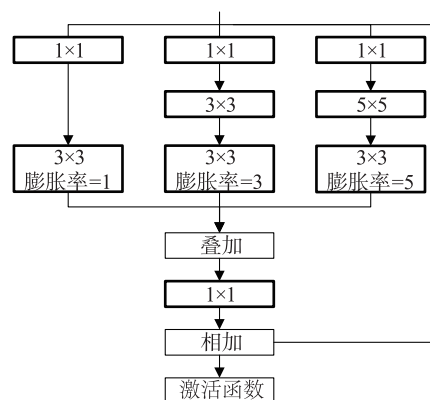
a: 果面虫蛀面积与果面总面积比例; b: 果面轮纹面积与果面总面积比例; c: 果面木质化面积与果面总面积比例。

图 2 冬枣果面缺陷尺寸分析图

Fig.2 Scale analysis diagram of surface defect in winter jujube

RFB 是一种模仿人眼群体感受野 (Population receptive field, pRF) 的仿生卷积结构, 该结构通过标

准卷积和空洞卷积模拟 pRF 尺寸与偏心率的正相关实现特征图感受野<sup>[18-19]</sup>。为了获得 3 种尺寸的冬枣果面缺陷特征,引入模型的 RFB 设置为 3 个分支卷积结构,其中每个分支由卷积核大小不同的标准卷积和膨胀率不同的空洞卷积组成,前者用于模拟不同尺寸的 pRF,后者用于模拟不同偏心率的 pRF。由图 3 可以看出,空洞卷积核的膨胀率随着标准卷积核的尺寸增大而增大,两者的变化关系则复现了 pRF 的尺寸与偏心率的关系。最后 RFB 各分支输出的特征图都将叠加到一起,以实现多尺寸果面缺陷特征融合。



粗线框为标准卷积。

图 3 感受野块 (RFB) 架构图

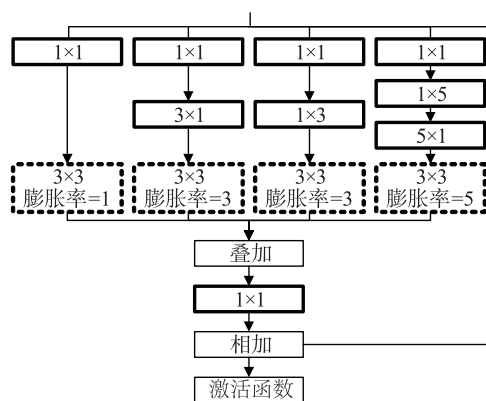
Fig.3 Architecture of receptive field block (RFB)

为了减轻引入 RFB 对模型实时性的影响,本研究对 RFB 进行改进:将 3x3 标准卷积拆分成并联的 1x3 标准卷积和 3x1 标准卷积;将 5x5 标准卷积拆分成串联的 1x5 标准卷积和 5x1 标准卷积;同时用深度可分离卷积代替标准卷积实现空洞卷积,最后得到改进的 RFB 模块(图 4)。由于在浅层特征图中更容易提取到尺寸普遍偏小的冬枣果面缺陷特征,本研究将改进 RFB 模块嵌入 MobilenetV3-SSD 的第 13 个倒残差模块和分类定位预测层之间。

#### 1.4 空间注意力模块

通过观察图 5 中 3 类果面缺陷出现的位置,发现缺陷可能处于果面边缘,也可能处于果面中央,甚至遍布果面。针对缺陷在冬枣果面上任意分布的问题,本研究引入空间注意力机制帮助果面缺陷冬枣实时检测模型确定果面缺陷的位置,继而提高模型定位冬枣果面缺陷位置的精度。

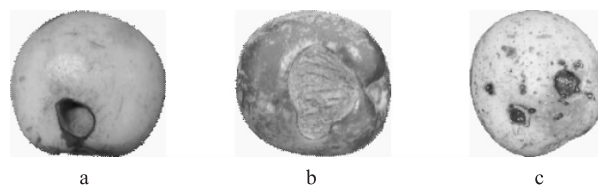
倒残差模块中采用了挤压和激励模块



粗线框为标准卷积,虚线框为深度可分离卷积。

图 4 改进的 RFB 架构图

Fig.4 Architecture of improved RFB



a: 虫蛀冬枣; b: 木质化冬枣; c: 轮纹冬枣

图 5 3 类果面缺陷冬枣

Fig.5 Three kinds of winter jujubes with surface defects

(Squeeze-and-Excitation, SE), SE 属于通道注意力机制,该机制利用各通道特征图之间的关系来提取冬枣果面缺陷特征,最后生成基于果面缺陷特征的通道注意力特征图。可见通道注意力机制有利于提高模型提取果面缺陷特征的能力<sup>[20]</sup>。但是冬枣果面缺陷位置不确定影响了模型定位果面缺陷冬枣的精度。解决这个问题需要增强模型定位冬枣果面缺陷的能力,而空间注意力机制恰好能改善模型的定位能力,因此本研究用卷积层注意力模块(CBAM)中的空间注意力模块(SAM)取代 SE<sup>[21-23]</sup>。

如图 6 所示,本研究将倒残差模块中深度可分离卷积层的输出作为 SAM 的输入特征图, SAM 首先通过  $F$  推导出空间注意力特征图  $M_s(F)$ , 然后  $M_s(F)$  与  $F$  进行哈达玛积运算,求得输出特征图  $F'$ ,最后 SAM 的输出作为投影层的输入。

SAM 计算公式如公式(5)所示。

$$F' = M_s(F) \otimes F \quad (5)$$

式中  $F'$  是输出特征图,  $M_s(F)$  是空间注意力特征图,  $F$  是输入特征图。

执行哈达玛积运算的目的是为  $F$  中出现冬枣果面缺陷特征的位置分配相应的关注度。



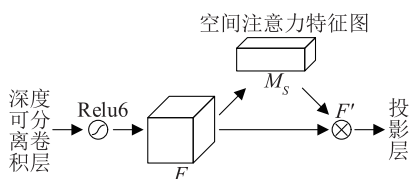


图6 空间注意力模块与倒残差模块组合结构图

Fig.6 Structure diagram of combination of spatial attention module and inverted residual module

如图7所示,为了推导  $M_s(F)$ ,首先需要通过最大池化和平均池化整合  $F$  的通道信息,生成平均池化特征图( $F_{avg}^s$ )和最大池化特征图( $F_{max}^s$ ),2种池化操作的目的是凸显冬枣果面缺陷部位特征<sup>[24]</sup>;然后叠加  $F_{avg}^s$  和  $F_{max}^s$ ,生成叠加特征图  $F_{concat}^s$ ;最后  $F_{concat}^s$  依次经过  $7 \times 7$  标准卷积和 sigmoid 激活函数处理生成空间注意力特征图 [ $M_s(F)$ ]。

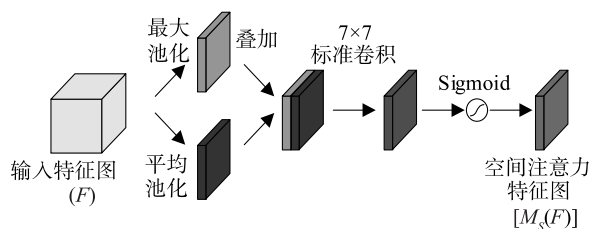


图7 空间注意力模块结构图

Fig.7 Structure diagram of spatial attention module

$M_s(F)$  计算公式如公式(6)所示。

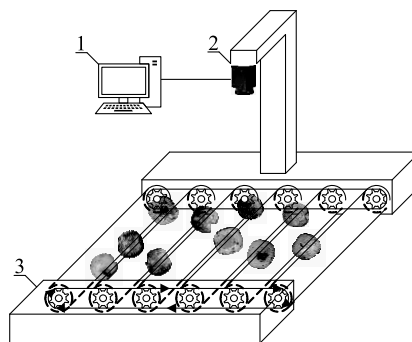
$$M_s(F) = \sigma \{ f^{7 \times 7} [ \text{AvgPool}(F); \text{Maxpool}(F) ] \} = \sigma [ f^{7 \times 7} (F_{avg}^s; F_{max}^s) ] = \sigma [ f^{7 \times 7} (F_{concat}^s) ] \quad (6)$$

式中  $M_s(F)$  是空间注意力特征图,  $F$  是输入特征图,  $F_{avg}^s$  是平均池化特征图,  $F_{max}^s$  是最大池化特征图。

## 2 果面缺陷冬枣实时检测模型训练

### 2.1 果面缺陷冬枣数据集建立

2.1.1 数据采集 采集时间为2020年7-11月,采集地点为陕西中农厚朴农业科技有限公司。果面缺陷冬枣数据集以陕西大荔冬枣为数据采集对象。数据集由3类果面缺陷冬枣和正常冬枣组成。如图8所示,数据集通过数据采集设备完成采集工作,其中电荷耦合器件(Charge coupled device, CCD)工业相机使用杭州海康威视数字技术股份有限公司生产的MV-CE013-50UC相机,最大分辨率为  $1280 \times 960$ ,即  $1.30 \times 10^{10}$  像素。采集时冬枣随机摆放,同时CCD工业相机在自然光照下以俯视视角采集图像。



1:PC机;2:CCD工业相机;3:滚筒机。

图8 数据采集设备

Fig.8 Data collection equipment

2.1.2 数据集预处理及生成 原始数据集共400张图像,本研究使用Labellmg标注数据,标注文件格式参照Pascal VOC。

为了提高果面缺陷冬枣实时检测模型的检测精度和泛化能力,本研究使用Python库函数Imgaug对数据集进行数据增强<sup>[25]</sup>。考虑到图像内的冬枣果面特征不会因观察方向的变化而改变,并且为了增强模型在不同光照下的鲁棒性,本研究使用上下翻转、左右翻转、镜像翻转和亮度增减4种方法实现数据增强。在Imgaug的帮助下,标注框会随着数据增强而变换,避免再次绘制。

经过增强,数据集共2000张图像,按8:1:1的比例划分为训练集、验证集和测试集。

### 2.2 试验环境

试验硬件环境选择Inter(R) Xeon(R) CPU E5-2637 @ 3.50 Hz 和 NVIDIA GeForce RTX2070 SUPER 8 G;操作系统为Windows Server 2008 R2 Enterprise 64bit;软件环境选择Python3.6.7和Keras 2.1.5。

### 2.3 模型训练

本研究模型训练运用迁移学习策略,需要在Pascal VOC07+12数据集上训练得到预训练模型。鉴于丰富的Pascal VOC07+12数据集中必然有数据与本研究数据存在关联,因此迁移学习能把学到的模型参数应用于本研究数据集训练,达到加快并优化训练的目的,避免人工初始化参数直接在数据集上训练,导致模型表现不佳<sup>[26]</sup>。

另外模型训练还运用了冻结训练策略,即主干网络MobileNetV3的模型参数不会随着模型训练而更新。冻结训练可以有效避免模型过拟合。本研究将前50次训练设为冻结训练阶段,输入图像大小统

一修改为  $300 \times 300$ , 批数据量为 16, 学习率为 0.000 5。第 50 次以后训练进入解冻训练阶段, 批数据量减半, 学习率降为 0.000 1。其中学习率下降以验证损失为标准, 当验证损失经过 3 次训练不下降, 学习率减半。训练使用自适应矩估计优化器,  $\beta_1 = 0.900, \beta_2 = 0.999$ 。Adam 优化器可预防学习步长过小导致模型收敛缓慢的问题。

## 2.4 模型性能评价指标

由于果面缺陷目标检测模型在实际应用中需要平衡实时性和精准性, 因此试验中模型性能的优劣以这 2 个性能为衡量标准。每秒帧数 (FPS) 可表示模型对果面缺陷冬枣检测的实时性, 平均精度 (AP) 可表示模型对各类冬枣检测的精准性,  $mAP$  可表示模型对 4 类冬枣的整体检测精准性。因此本研究将 FPS、AP 和  $mAP$  作为评价指标。

计算 AP 需求得精度 ( $p$ ) 和召回率 ( $r$ ),  $p$  和  $r$  计算公式如公式 (7) 和公式 (8) 所示。

$$p = \frac{TP}{TP + FP} \quad (7)$$

$$r = \frac{TP}{TP + FN} \quad (8)$$

式中  $TP$  表示分类正确的正样本数,  $FP$  表示分类错误的正样本数,  $FN$  表示分类错误的负样本数。

AP 和  $mAP$  的计算公式如公式 (9) 和公式 (10) 所示。

$$AP = \int_0^1 p(r) dr \quad (9)$$

$$mAP = \frac{\sum_{i=1}^n AP(i)}{n} \quad (10)$$

式中  $n$  表示冬枣果面类别数, 故  $n = 4$ ;  $AP(i)$  表示第  $i$  类冬枣果面的平均精度。

表 2 不同尺寸感受野对模型性能的影响

Table 2 Effects of receptive fields with different sizes on model performance

感受野 尺寸	平均精度 (%)				$mAP$ (%)	FPS	模型计算量 (MFLOPs)
	正常冬枣	虫蛀冬枣	轮纹冬枣	木质化冬枣			
单尺寸	94.34	88.94	92.19	85.28	90.19	22.79	48.28
多尺寸	94.73	91.17	94.25	86.60	91.69	22.06	49.61

$mAP$ 、FPS 见表 1 注。

此外多尺寸感受野的应用对于正常冬枣的分类精度提高不明显, 原因是相对于果面缺陷冬枣, 正常冬枣的果面不存在多尺寸果面特征, 因此多尺寸感受野更适用于提取不同尺寸的冬枣果面缺陷特征。

## 3 结果与分析

### 3.1 不同主干网络对模型性能的影响

为了分析 MobileNetV3 对模型检测冬枣果面缺陷速度的影响, 本研究对 SSD 分别搭配 VGG16 和 MobileNetV3 的性能进行比较试验。由表 1 可知, 当主干网络由 VGG16 更换为 MobileNetV3 后, 模型计算量减少了 43.16 MFLOPs。庞大计算量的减少优化了模型性能,  $mAP$  仅损失 1.98%, FPS 获得 81.22% 的提升, 达到 41.30。因此将 MobileNetV3 作为主干网络可为果面缺陷冬枣实时检测模型的实际应用奠定基础。

表 1 不同主干网络试验结果对比

Table 1 Comparison of results of different backbone networks

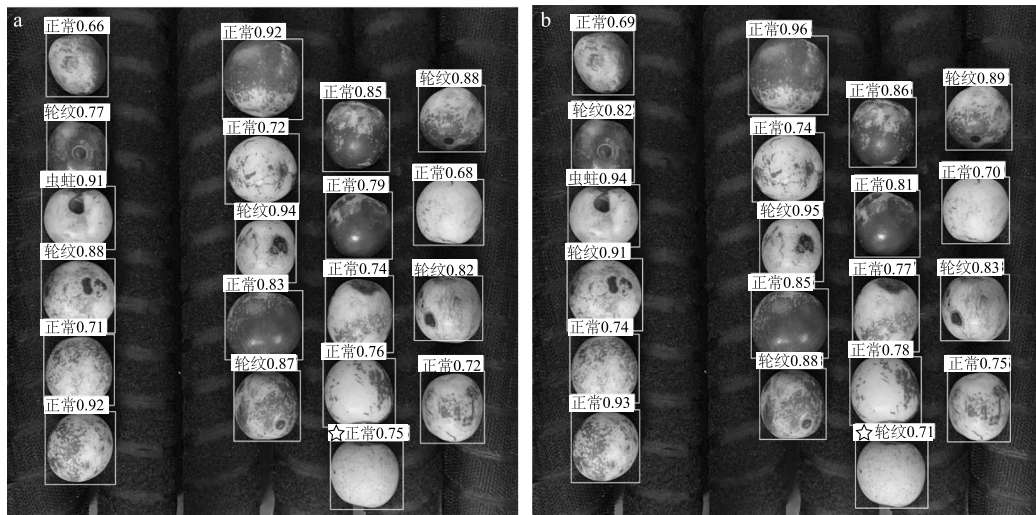
主干网络	FPS	模型计算量 (MFLOPs)	$mAP$ (%)
VGG16	22.79	48.28	90.19
MobileNetV3	41.30	5.12	88.21

FPS: 每秒帧数;  $mAP$ : 模型对 4 类冬枣的整体检测精准性。

### 3.2 不同尺寸感受野对模型性能的影响

为了分析应用多尺寸感受野对模型分类冬枣果面缺陷精度的影响, 并避免 MobileNetV3 作为主干网络对试验的干扰, 本研究对 VGG16-SSD 的引入是否改进 RFB 的性能进行试验, 即试验 SSD 分别应用多尺寸感受野和单尺寸感受野对模型性能的影响。试验结果如表 2 所示, 改进的 RFB 后, 模型对正常冬枣、虫蛀冬枣、轮纹冬枣和木质化冬枣的 AP 分别提升了 0.39、2.23、2.06、1.32 个百分点,  $mAP$  提升了 1.50 个百分点。同时改进的 RFB 为轻量级卷积块, 将其引入模型后, 模型计算量仅增加 2.75%, FPS 仅下降 3.20%, 说明模型的实时性基本不受影响。

如图 9 所示, 应用单尺寸感受野的模型忽略了果面上存在小尺寸轮纹的冬枣, 导致分类错误。而应用多尺寸感受野的模型可以从多个尺寸提取冬枣果面缺陷特征, 因此对果面上存在不同尺寸缺陷的冬枣分类无误。



a: 单尺寸感受野; b: 多尺寸感受野。星号表示小尺寸轮纹的冬枣。

图 9 不同尺寸感受野检测结果比较

Fig.9 Comparison of detection results of receptive fields with different sizes

### 3.3 不同注意力模块对模型性能的影响

为了分析应用空间注意力机制对模型定位冬枣果面缺陷精度的影响,本研究对 MobileNetV3-SSD 使用不同注意力模块进行比较试验。试验结果(表 3)表明,当注意力模块由 SE 更换为 SAM 后,模型对正常冬枣、虫蛀冬枣、轮纹冬枣和木质化冬枣的  $AP$  分别提升了 2.63、2.27、3.34、3.18 个百分点,

$mAP$  提升了 2.85 个百分点。同时由于删减掉 CBAM 中属于通道注意力机制的 CAM,仅将 SAM 引入模型,模型计算量降低了 42.58%, $FPS$  提升了 3.70%。

模型对正常冬枣的  $AP$  提升的原因是正常冬枣的果面特征单一且分布规律,易被模型定位,因此 SAM 可提高模型定位正常冬枣的精度。

表 3 不同注意力机制对模型性能的影响

Table 3 Effects of different attention mechanisms on model performance

注意力模块	$mAP$	$AP$ (%)				$FPS$	模型计算量 (MFLOPs)
		正常冬枣	虫蛀冬枣	轮纹冬枣	木质化冬枣		
通道注意力模块(SE)	88.21	92.03	86.29	90.37	84.14	41.30	5.12
空间注意力模块(SAM)	91.06	94.66	88.56	93.71	87.32	42.83	2.94

$mAP$ : 模型对 4 类冬枣的整体检测精准性;  $AP$ : 平均精度;  $FPS$ : 每秒帧数。

如图 10 所示,应用通道注意力模块的模型无法定位轮纹特征,导致轮纹冬枣定位不准确。而应用空间注意力模块的模型具备定位轮纹特征的能力,因此对轮纹冬枣定位无误。

### 3.4 不同改进组合对模型性能的影响

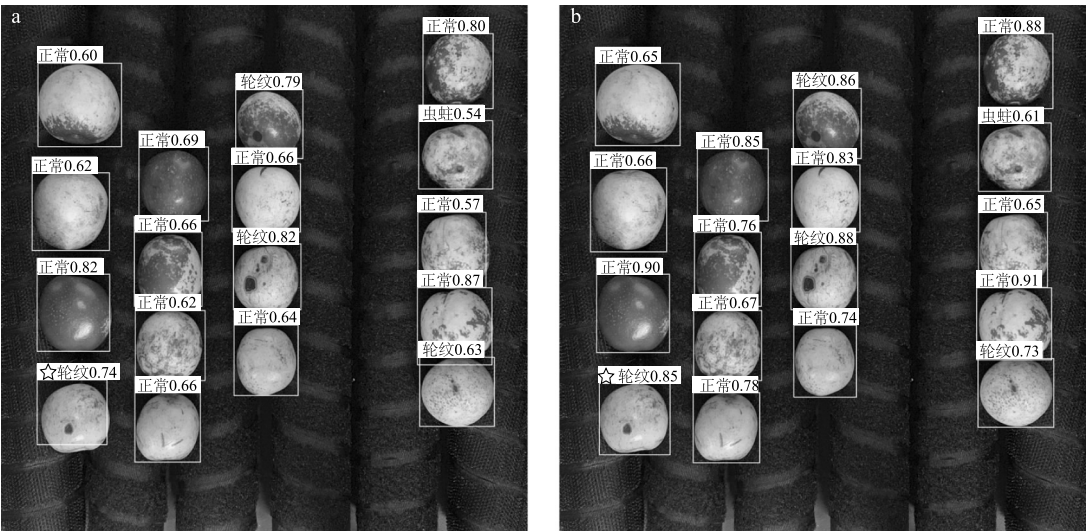
本研究分别从主干网络、感受野尺寸以及注意力模块 3 个方面对 SSD 进行改进。SSD 为表 4 中的模型 1。在上述试验中不同改进单独与 SSD 的组合分别记作表 4 中的模型 2 至模型 4。如表 5 所示,模型 1~模型 4 4 个模型的性能较好,为了分析所有改进与 SSD 组合对模型的影响,本研究将所有改进与 SSD 组成的模型记

作模型 5,并与其他 4 个模型进行性能比较试验。

由表 5 可知,模型 5 表现出众,该模型以 MobileNetV3 为主干网络,同时引入改进 RFB 并选择 SAM 取代 SE 完成组合。在实时性上,模型 5 的  $FPS$  达 40.85,仅与检测速度最快的模型 4 相差 1.98。在精准性上,模型 5 的  $mAP$  位列各模型之首,达到 91.89%。模型 5 有效平衡了模型的实时性与精准性之间的关系,足见所有改进共同作用的有效性,因此本研究将表现最优的模型 5 作为本研究提出的冬枣果面缺陷实时检测模型。

图 11 为本研究模型在测试集上的表现。





a:通道注意力模块;b:空间注意力模块。星号表示小尺寸轮纹的冬枣。

图 10 不同注意力模块检测结果比较

Fig.10 Comparison of detection results of different attention modules

表 4 不同改进组合对模型性能的影响

Table 4 Effects of different improved combinations on model perormance

模型编号	主干网络		感受野尺寸		注意力模块	
	VGG16	MobileNetV3	单尺寸	多尺寸	SE	SAM
1	✓		✓			
2	✓			✓		
3		✓	✓		✓	
4		✓	✓			✓
5		✓		✓		✓

✓代表该模型引入的改进。

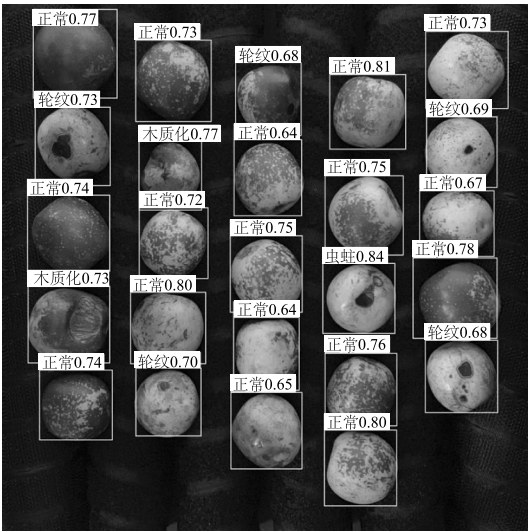


图 11 本研究模型检测结果

Fig.11 Detection results of the model established in this study

3.5 本研究模型与目前先进目标检测网络性能对比

为了分析本研究模型的性能,本研究将所提出的模型与目前先进的目标检测网络 RetinaNet 和 EfficientDet-D0 在果面缺陷冬枣数据集上进行比较试验。为保证比较试验的可靠性,2 种比较模型均按本研究模型训练方式进行训练。由表 6 可知,本研究模型的 *mAP* 和 *FPS* 均排在首位,其中比 EfficientDet-D0 的 *mAP* 和 *FPS* 分别高出 3.65 个百分点和 7.15。经分析,EfficientDet-D0 采用双向特征金字塔网络(BiFPN)结合特征加权策略,这有利于模型重视与冬枣果面缺陷尺寸相符的特征,忽略冗余特征,EfficientDet-D0 的性能略优于 RetinaNet<sup>[27]</sup>。

从图 12 可知,各模型在训练过程中收敛迅速,且验证损失均在第 70 次训练中便趋于稳定,这得益于迁移学习策略和 Adam 优化器的应用。同时由于



训练中使用冻结训练策略,各模型训练趋于稳定时验证损失与训练损失相近,第 100 次训练的验证损失均低于训练损失,未出现过拟合现象。

表 5 不同组合模型试验结果对比

Table 5 Comparison of test results of different combined models

模型编号	AP(%)				mAP(%)	FPS
	正常冬枣	虫蛀冬枣	轮纹冬枣	木质化冬枣		
1	94.34	88.94	92.19	85.28	90.19	22.79
2	94.73	91.17	94.25	86.60	91.69	22.06
3	92.03	86.29	90.37	84.14	88.21	41.30
4	94.66	88.56	93.71	87.32	91.06	42.83
5	94.85	89.68	94.77	88.25	91.89	40.85

模型 1~模型 5 见表 4, AP、mAP、FPS 见表 3 注。

表 6 本研究模型和目前先进目标检测网络试验结果对比

Table 6 Comparison of test results of proposed model and current advanced detection network models

模型	AP(%)				mAP(%)	FPS
	正常冬枣	虫蛀冬枣	轮纹冬枣	木质化冬枣		
RetinaNet	92.03	85.92	90.37	81.14	87.37	28.10
EfficientDet-D0	92.79	86.43	89.84	83.91	88.24	33.70
本研究模型	94.85	89.68	94.77	88.25	91.89	40.85

AP、mAP、FPS 见表 3 注。

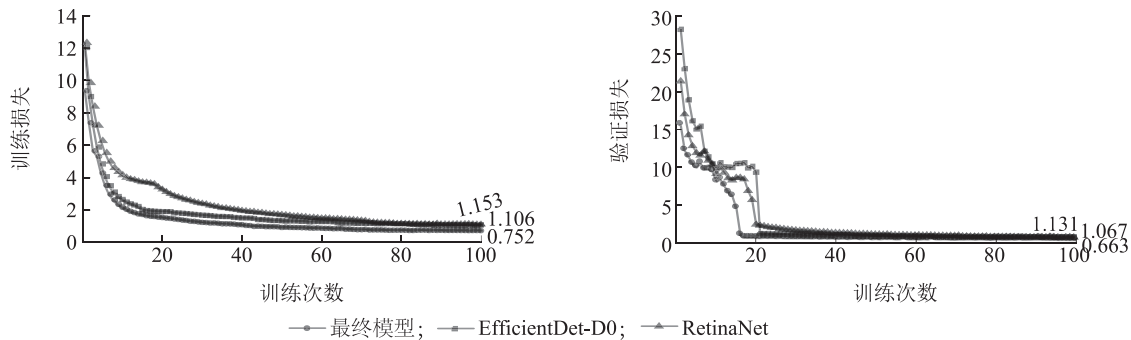


图 12 训练损失对比与验证损失对比

Fig.12 Comparison of training loss and comparison of validation loss

## 4 结 论

本研究针对实时检测在果面缺陷冬枣分拣中的重要性以及果面缺陷尺寸不一、分布随机影响检测精度的问题,对 SSD 进行改进:SSD 通过搭配 MobileNetV3,从检测速度方面提高本研究模型的实时性;SSD 通过引入改进 RFB 和 SAM 从分类和定位两方面提高本研究模型的精准性。通过对比试验,本研究模型的有效性得到验证,mAP 达到 91.89%,FPS 达到 40.85。在组合不同改进的模型中,本研

究模型平衡实时性和精准性的能力最强。在与目前先进目标检测网络模型的对比中,本研究模型比 EfficientDet-D0 的 mAP 和 FPS 分别高出 3.65 个百分点和 7.15,比 RetinaNet 的 mAP 和 FPS 分别高出 4.52 个百分点和 12.75。因此本研究模型可以为果面缺陷冬枣实时检测提供技术支持。经实测,本研究模型 1 h 分拣量可达 4 t。

### 参考文献:

[1] 王春焕,王 莹,张梦辉. 大荔县冬枣产业发展建议[J].陕西

- 林业科技, 2021, 49(2): 97-99.
- [2] 侯鹏帅, 刘玉乐, 宋欣, 等. 基于智能视觉识别的冬枣分选检测系统设计[J]. 中国农机化学报, 2020, 41(3): 109-114.
- [3] GENG L, XU W, ZHANG F, et al. Dried jujube classification based on a double branch deep fusion convolution neural network[J]. Japanese Society for Food Science and Technology, 2018, 24(6): 1007-1015.
- [4] NTURAMBIRWE J F I, OPARA U L. Machine learning applications to non-destructive defect detection in horticultural products[J]. Biosystems Engineering, 2020, 189: 60-83.
- [5] BHARGAVA A, BANSAL A. Fruits and vegetables quality evaluation using computer vision: a review[J]. Journal of King Saud University-Computer and Information Sciences, 2021, 33(3): 243-257.
- [6] LU Y, LU R. Development of a multispectral structured illumination reflectance imaging (SIRI) system and its application to bruise detection of apples[J]. Transactions of the ASABE, 2017, 60(4): 1379-1389.
- [7] 孙世鹏, 彭俊, 李瑞, 等. 基于近红外高光谱图像的冬枣损伤早期检测[J]. 食品科学, 2017, 38(2): 301-305.
- [8] JU J, ZHENG H, XU X, et al. Classification of jujube defects in small data sets based on transfer learning[DB/OL]. (2021-01-30) [2022-01-15]. <https://link.springer.com/article/10.1007/s00521-021-05715-2>.
- [9] 薛勇, 王立扬, 张瑜, 等. 基于 GoogLeNet 深度迁移学习的苹果缺陷检测方法[J]. 农业机械学报, 2020, 51(7): 30-35.
- [10] 费琦琪, 施杰, 夏敏, 等. 基于 CNN 的冰糖橙分级系统[J]. 江苏农业学报, 2020, 36(2): 513-519.
- [11] 李小占, 马本学, 喻国威, 等. 基于深度学习与图像处理的哈密瓜表面缺陷检测[J]. 农业工程学报, 2021, 37(1): 223-232.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[M]. Amsterdam, The Netherlands: Proceedings of the European Conference on Computer Vision, 2016.
- [13] 徐先峰, 赵万福, 邹浩泉, 等. 基于 MobileNet-SSD 的安全帽佩戴检测方法[J]. 计算机工程, 2021, 47(10): 298-305, 313.
- [14] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[M]. Seoul, Korea: International Conference on Computer Vision, 2019.
- [15] CHOLLET F. Xception: deep learning with depthwise separable convolutions[M]. Hawaii, United States: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [16] 黄林生, 罗耀武, 杨小冬, 等. 基于注意力机制和多尺度残差网络的农作物病害识别[J]. 农业机械学报, 2021, 52(10): 264-271.
- [17] 王春山, 周冀, 吴华瑞, 等. 改进 Multi-scale ResNet 的蔬菜叶部病害识别[J]. 农业工程学报, 2020, 36(20): 209-217.
- [18] LIU S, HUANG D, WANG Y. Receptive field block net for accurate and fast object detection[M]. Munich, Germany: Proceedings of the European Conference on Computer Vision, 2018.
- [19] WANDELL B, WINAWER J. Computational neuroimaging and population receptive fields[J]. Trends in Cognitive Sciences, 2015, 19(6): 349-357.
- [20] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[M]. Salt Lake City, United States: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [21] 高钦泉, 黄炳城, 刘文哲, 等. 基于改进 CenterNet 的竹条表面缺陷检测方法[J]. 计算机应用, 2021, 41(7): 1933-1938.
- [22] WOO S, PARK J, LEE J, et al. CBAM: convolutional block attention module[M]. Munich, Germany: Proceedings of the European Conference on Computer Vision, 2018.
- [23] CHEN L, ZHANG H, XIAO J, et al. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning[M]. Hawaii, United States: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [24] ZAGORUYKO S, KOMODAKIS N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer[M]. Paris, France: International Conference on Learning Representations, 2017.
- [25] TSUCHIYA H, FUKUI S, IWAHORI Y, et al. A method of data augmentation for classifying road damage considering influence on classification accuracy[J]. Procedia Computer Science, 2019, 159: 1449-1458.
- [26] 余小东, 杨孟辑, 张海清, 等. 基于迁移学习的农作物病虫害检测方法研究与应用[J]. 农业机械学报, 2020, 51(10): 252-258.
- [27] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[M]. Virtual: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020.

(责任编辑: 陈海霞)