



基于机器视觉和机器学习技术的浙贝母外观品质等级区分

董成烨, 李东方, 冯槐区, 龙思放, 奚特, 周琴安, 王俊*

(浙江大学生物系统工程与食品科学学院, 浙江 杭州 310058)

摘要 为区分浙贝母外观品质等级,本研究利用数字电子眼系统及图像标注工具构建浙贝母数据集,选择若干统计学习算法和目标检测算法在该数据集上进行训练与测试。结果表明:目标检测算法YOLO(you only look once)系列YOLO-X所得模型的效果最佳。为优化YOLO-X,根据浙贝母数据集的特点,针对性地向YOLO-X的主干特征提取网络末端嵌入空洞卷积结构,以加强模型对尺度特征的敏感度。改进后模型(空洞率为4)的平均精确率均值为99.01%,对于特级、一级、二级、虫蛀、霉变、破碎浙贝母的平均精确率分别为99.97%、98.33%、98.47%、98.71%、99.73%、98.85%,精确率和召回率的加权调和平均数(F_1)分别为0.99、0.92、0.94、0.97、0.99、0.97。本研究在不增加参数量、计算量或者对算法进行大规模改动的情况下,改善了模型的检测效果,为后续浙贝母检测平台的搭建提供了科学依据。

关键词 浙贝母; 统计学习; 深度学习; 目标检测; 目标检测算法YOLO-X; 空洞卷积

中图分类号 TP391.4; S238; S567.231

文献标志码 A

引用格式 董成烨,李东方,冯槐区,等.基于机器视觉和机器学习技术的浙贝母外观品质等级区分[J].浙江大学学报(农业与生命科学版),2023,49(6):881-892. DOI:10.3785/j.issn.1008-9209.2022.10.181

DONG Chengye, LI Dongfang, FENG Huaiqu, et al. Classification of *Fritillaria thunbergii* appearance quality based on machine vision and machine learning technology[J]. *Journal of Zhejiang University (Agriculture & Life Sciences)*, 2023, 49(6): 881-892.

Classification of *Fritillaria thunbergii* appearance quality based on machine vision and machine learning technology

DONG Chengye, LI Dongfang, FENG Huaiqu, LONG Sifang, XI Te, ZHOU Qin'an, WANG Jun* (*College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, Zhejiang, China*)

Abstract In order to classify the appearance quality level of *Fritillaria thunbergii*, the *F. thunbergii* dataset was constructed with the DigiEye system followed by an image annotation tool. Several statistical learning and object detection algorithms were selected to train and test the *F. thunbergii* dataset. The results showed that the model trained by the YOLO-X of YOLO (you only look once) series had relatively better performance. In addition, to optimize YOLO-X, according to the unique features of *F. thunbergii* dataset, a dilated convolution structure was embedded into the end of the backbone feature extraction network of YOLO-X as it could improve the model sensitivity to the dimension feature. The mean average precision (mAP) of the improved model was raised to 99.01%; the average precision (AP) for superfine, level one, level two, moth-eaten, mildewed, and broken *F. thunbergii* were raised to 99.97%, 98.33%, 98.47%, 98.71%, 99.73%, and 98.85%, respectively; and the weighted harmonic mean of precision and recall (F_1) were raised to 0.99, 0.92, 0.94, 0.97, 0.99, and 0.97, respectively. The

基金项目 浙江省重点研发计划项目(2021C02016)。

*通信作者(Corresponding author): 王俊(<https://orcid.org/0000-0001-5767-6149>), E-mail: jwang@zju.edu.cn

第一作者(First author): 董成烨(<https://orcid.org/0000-0003-2716-3339>), E-mail: DongChengye@zju.edu.cn

收稿日期(Received): 2022-10-18; 接受日期(Accepted): 2023-04-25

tune-up in this study enhanced the detection performance of the model without increasing the number of parameters, computational complexity, or major changes to the original model. This study provides a scientific basis for the subsequent construction of *F. thunbergii* detection platform.

Key words *Fritillaria thunbergii*; statistical learning; deep learning; object detection; object detection algorithm YOLO-X; dilated convolution

浙贝母(*Fritillaria thunbergii*)是一种百合科(Liliaceae)贝母属(*Fritillaria*)草本植物,是传统中药材之一,主要药用成分为生物碱类、皂苷类,具有镇咳、祛痰、抗炎、抑菌、抗肿瘤等功效。据统计,2017—2018年中国浙贝母总产量已超4 000 t,种植产值约为60亿元^[1]。根据T/CACM 1021.24—2018《中药材商品规格等级 浙贝母》,浙贝母在满足无霉变、无虫蛀、无破碎的情况下,按直径分为特级、一级、二级^[2],而霉变、虫蛀、破碎的浙贝母不能入药,需剔除。然而,目前对浙贝母的分级通常利用筛网、圆柱辊等实现^[3-6],由于浙贝母形状不规则,分级效果往往不理想,此外,筛分过程易对浙贝母造成机械损伤,也无法筛分出霉变、虫蛀、破碎的浙贝母。因此,开发精准检测浙贝母外观品质及直径的算法,并研制一种快速、无损的分级生产线十分必要。

目前,针对农产品的无损检测主要依靠声学、光学、电磁、电子鼻和机器视觉等技术^[7]。机器视觉是利用摄像机和电脑代替人眼对目标进行识别、跟踪和测量的技术,具备准确、快速、经济等特点,因此,浙贝母的外观品质及直径的检测与分级可通过机器视觉技术实现。机器学习是处理机器视觉任务的重要工具^[8],包括统计学习与深度学习。统计学习是利用研究人员的先验知识对数据进行硬编码,以更简单的方式来表示数据,从而分析、挖掘数据内部潜在的规律,但是其特征需要手动提取且数量有限,因此很难快速、高效地利用原有数据,并且操作过程复杂、耗时^[9]。

深度学习是一种以人工神经网络为架构,对数据进行高层抽象的算法。深度学习与统计学习的关键区别在于前者可以从数据中自动学习更为深层的特征,所以深度学习更易建模,且能高效利用数据。目前,基于深度学习的目标检测算法分为一阶段(one stage)型和两阶段(two stage)型,前者以YOLO(you only look once)系列算法(YOLO、YOLO-V2、YOLO-V3、YOLO-V4、YOLO-V5、YOLO-X)为代表^[10-14],后者以R-CNN(region-based convolutional

neural network)系列算法(R-CNN、Fast R-CNN、Faster R-CNN)为代表^[15-17]。上述目标检测算法已应用于农产品检测领域,在田间或生产线上苹果^[18-19]、冬青果^[20]、茶芽^[21]、番茄^[22-23]、小麦^[24-25]、青椒^[26]等的检测任务中均取得了较好的效果。然而,不同于其他物料对象,浙贝母等级较多,且特级、一级和二级的外观特征相近,因此实现对浙贝母的精准检测具有一定难度。目前,还鲜有研究将目标检测算法应用于浙贝母的外观品质及直径的检测与分级中。

本研究将机器学习应用于浙贝母的无损检测中,创建浙贝母数据集,并在该数据集上训练、测试若干统计学习算法和基于深度学习的目标检测算法。为改善检测效果,本研究根据浙贝母数据集的特点,在YOLO-X主干特征提取网络的末端添加空洞卷积(dilated convolution)结构,并通过可视化中间激活的方式,从直观的角度对比、分析改进前后模型学习特征的差异,以实现浙贝母的无损检测,也为后续在线检测平台的搭建提供科学依据。

1 材料与方法

1.1 试验样品

本研究所用浙贝母产自浙江省金华市磐安县(28°57'05"—29°01'58" N, 120°28'05"—120°33'40" E; 海拔319 m)。浙贝母表面呈类白色或淡黄色,形状呈扁圆形,其稳定的摆放姿态有2种,分别为正面(鳞叶开合一端)朝上和背面(鳞叶带梗一端)朝上(图1A₁~A₆)。特级、一级、二级浙贝母在形状、颜色、纹理等特征方面差异较小,主要以直径划分,直径越大,等级越高(图1A₁~A₃);霉变、虫蛀的浙贝母形状与正常浙贝母类似,但前者因微生物在其内部或表面大量繁殖,导致其外观呈深灰褐色(图1A₄),后者表面有多个虫孔且常伴随絮状物,其外观颜色偏淡黄色(图1A₅);破碎的浙贝母形状不规则、特征多变,很难利用固定的模型或阈值来进行表征(图1A₆)。

1.2 图像采集

为模拟浙贝母实时在线检测场景,本研究采用 DigiEye System-700mm Cube 数字电子眼系统(英国 VeriVide 公司)作为图像采集设备,其硬件由 D7000 单反相机(日本 Nikon 公司)、带有反光涂层的黑箱、条形光源和盛放样本的底板组成(图 1B)。为使图像最大程度地还原浙贝母真实的颜色、形状以及纹理等特征,减少物料表面由光源所造成的高

光区域,经反复试验,本试验光源采用漫反射光照模式,拍摄距离为 45 cm。基于上述条件,对 2 种摆放姿态下的 6 类浙贝母各拍摄 200 张图像,共计拍摄 2 400 张单目标图像,部分图像如图 1A₁~A₆所示。为丰富数据集种类,提高目标检测模型的泛化能力,本研究还拍摄了 400 张多目标图像。为平衡不同级别浙贝母的数量,多目标图像中各类不同摆放姿态的浙贝母数量相同,且呈单列化分布。

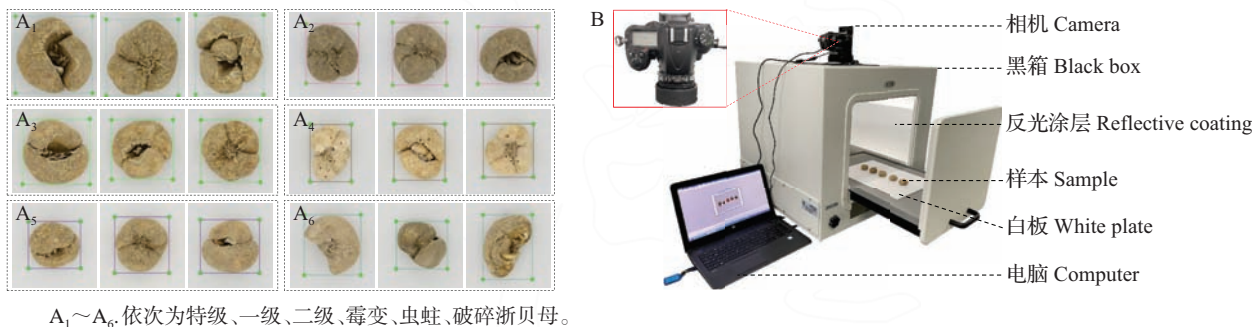


图 1 数据集部分图像(A₁~A₆)及图像采集设备(B)

Fig. 1 Partial images of data sets (A₁~A₆) and image acquisition equipments (B)

1.3 研究方法

本研究总体思路如图 2 所示。首先准备浙贝母数据集,用于训练和测试不同的统计学习算法和目

标检测算法;然后在检测效果最好的预选算法的基础上,针对浙贝母数据集的特点对算法进行相应的改进与优化,以改善模型的检测效果。

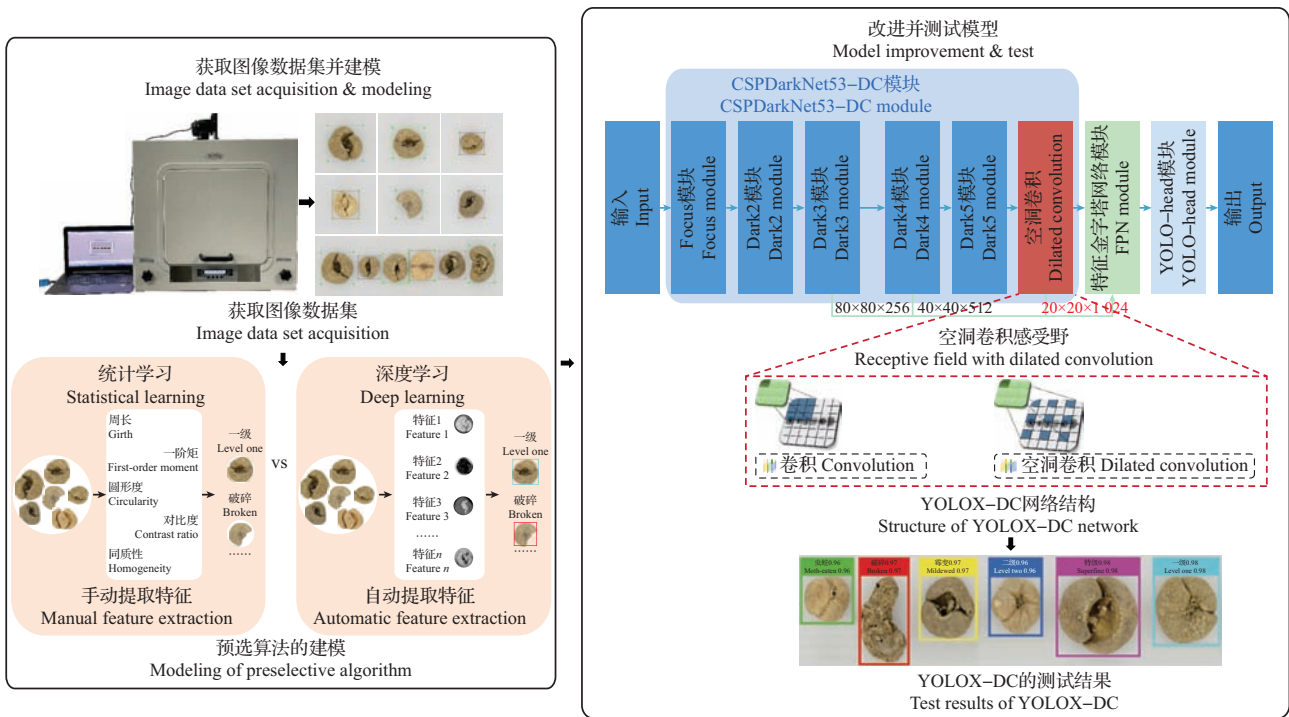


图 2 研究方法示意图

Fig. 2 Schematic diagram of the research method

1.3.1 数据集制作

目标检测算法有固定的输入尺寸,一般为32的倍数,当输入图片不满足要求时,会调整其大小,以满足尺寸要求。为适应神经网络的输入,将图像从4 928×3 264剪裁成3 200×3 200,以便目标检测算法对其进行等比放缩,避免对图像中的感兴趣区(region of interest, ROI)造成拉伸、扭曲,从而使图像保留原来的形状、尺寸等特征。利用Labelimg软件对图像进行标注,并将带注释的标签文件以扩展名为xml的格式保存。图像数据集共包含“superfine”“level one”“level two”“moth-eaten”“mildewed”和“broken”6类标签,分别对应特级、一级、二级、虫蛀、霉变、破碎浙贝母,每类浙贝母均有800个标签,因此浙贝母图像数据集共有4 800个标签。该数据集的ROI占比较大,但图像数量较少(共2 800个),因此训练集、验证集、测试集按6:2:2的比例划分,即训练集1 680个,验证集560个,测试集560个。

基于浙贝母图像,本研究提取了浙贝母形态(M)、颜色(C)和纹理(T)特征,建立了浙贝母特征数据集 $F=[M, C, T]=[M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18}, T_1, T_2, T_3, T_4, T_5, T_6, T_7]$,其中: $M_1 \sim M_9$ 代表形态特征,包括周长、面积、长轴、短轴、直径、矩形体、圆形体、紧凑度和细长度; $C_1 \sim C_{18}$ 代表颜色特征,包括图像在RGB(红、绿、蓝)、HSV(色调、饱和度、亮度)颜色空间下各通道的一、二、三阶矩; $T_1 \sim T_7$ 代表纹理特征,包括基于灰度共生矩阵提取的能量、对比度、相关度、熵、同质性、均值和标准差。因此,特征数据集 F 的属性空间为34维,样本量为2 400个。由于样本较为稀疏,本研究使用Origin 9.1软件对浙贝母特征数据集进行降维处理。降维后的属性空间为9维,信息保有量为原来的96.56%。

1.3.2 预选算法

本研究利用决策树(decision tree, DT)和支持向量机(support vector machine, SVM)对降维后的浙贝母特征数据集进行建模。DT是一种监督学习方式的分类和回归模型,利用数据特征对数据进行一系列条件判定(if or else)与划分,以提高划分后数据子集的信息纯度。DT通过训练集确定划分属性及其在决策树中的位置,从而产生一棵处理未见示例能力强的决策树。SVM是一种监督学习方式的分类模型,在样本空间中寻找一个超平面 $w^T x + b = 0$,

使其最近的正负样本(支持向量)到该超平面的距离(间隔)最大,从而区分不同类别的样本。其中: x 为样本特征向量, $w=(w_1, w_2, \dots, w_d)$ 为超平面法向量, T 为矩阵转置, b 为位移项。

本研究利用浙贝母图像数据集训练和测试了YOLO系列(YOLO-V3、YOLO-V4、YOLO-V5和YOLO-X)和Faster R-CNN目标检测算法。其中,Faster R-CNN会在图像中生成大量可能包含待检测物体的建议区域(region proposal, RP),然后通过分类器判别每个RP内是否包含待检测物体,并对物体及其位置进行分类和回归,从而得到待检测物体的边界框(bounding box),再过滤掉置信度不高或重叠的建议区域,进而得到检测结果。YOLO系列为实现更快的检测速度,采用预定义建议区域的方法,将生成候选区和检测目标合二为一,直接在输出层回归建议区域的位置和所属类别,并通过将建议区域划分成不同尺寸网格的方式,检测不同尺度的目标。YOLO系列含有主干特征提取网络(backbone)、特征金字塔网络(feature pyramid network, FPN)、预测结构(YOLO-head)3个模块。表1为本试验所选YOLO系列主干特征提取网络的特点总结。利用主干特征提取网络提取图片特征,并在特定的位置将其输入到FPN中,以进行进一步的特征提取与融合,其输出张量的通道数(D)按照公式(1)计算:

$$D = s \times n \times (p + x_{\text{offset}} + y_{\text{offset}} + h + w + 1). \quad (1)$$

式中: s 为建议区域被划分成网格的数量; n 为每个网格中存在的边界框数量; p 为属于某一类标签的置信度; x_{offset} 、 y_{offset} 为网格中心点调整参数; h 、 w 分别为网格高和宽的调整参数。将FPN的输出结果输入YOLO-head模块,以整合特征并调整通道数,最终获得预测结果。

1.3.3 训练设备及策略

试验采用相同配置的计算机来训练上述预选算法。计算机的操作系统为Windows 2019;中央处理器(central processing unit, CPU)型号为英特尔9900k;图形处理器(graphics processing unit, GPU)型号为NVIDIA Tesla V100-SXM2(32 GB DDR4 RAM);通用统一计算设备架构(compute unified device architecture, CUDA)的版本为11.2,CUDA深层神经网络库(CUDA deep neural network library, cuDNN)的版本是8.1.1;编程语言采用Python 3.7;深度学习框架采用PyTorch 1.7。

表1 YOLO系列的主干特征提取网络特点

Table 1 Characteristics of backbone feature extraction network of YOLO series

预选算法 Preselective algorithm	主干特征提取网络 Backbone feature extraction network	主干特征提取网络特点 Characteristics of backbone feature extraction network
YOLO-V3	DarkNet-53	应用DarkNet结构和LeakyReLU激活函数
YOLO-V4	CSPDarkNet-53	在DarkNet结构中添加CSPNet结构,变为CSPDarkNet结构,应用Mish激活函数
YOLO-V5	CSPDarkNet-53	在CSPDarkNet结构中添加Focus和空间金字塔池化(SPP)结构,应用SiLU激活函数
YOLO-X	CSPDarkNet-53	在CSPDarkNet结构中添加Focus和SPP结构,应用SiLU激活函数

DT主要有信息增益、信息增益率和基尼指数3种选择最优划分属性的方法,其代表算法分别为迭代二叉树3代(iterative dichotomiser 3, ID3)、C4.5和分类回归树(classification and regression tree, CART), CART可处理分类和回归任务。由于浙贝母特征数据集的属性取值为连续型数值,因此本研究采用CART决策树方法进行试验,并使用基尼指数来选择划分属性,使其生成叶子节点较小的树。为避免过拟合,本研究对决策树进行剪枝。经前期试验发现,当决策树的深度增加时,过拟合现象就会更严重,其深度为5时拟合状态最佳。利用SPSS PRO软件进行模型训练。核函数的选择是SVM性能的关键,本研究基于浙贝母特征数据集,训练并测试了不同核函数以及不同多分类融合策略的模型。

深度学习采取微调训练策略,即将视觉目标分类(visual object class, VOC)数据集或上下文中常见对象(common objects in context, COCO)数据集中预训练好的模型权重作为预选算法的初始权重,并在浙贝母训练集上进行迁移训练。理想的模型是其训练损失曲线不再有明显波动或刚好介于欠拟合与过拟合之间,为找到该界限,经多次试验发现,当训练总轮次达到200时,模型的训练损失曲线不再有明显波动,趋向平缓,说明此时模型接近收敛或已收敛。为优化训练策略,经多次试验发现,将冻结训练轮次设为50,解冻训练轮次设为150,学习率设为0.001,冻结批大小设为8,非冻结批大小设为4时,训练损失曲线的下降速度最快。为找出训练所得的最优模型,本试验保存了每轮次训练所得模型的权重文件,待训练轮次满200后,从中挑选出训练损失函数值最小的模型作为对应算法的最终训练所得模型。

1.3.4 基于YOLO-X的改进算法

前期预试验发现,目标检测模型对特级、一级、

二级浙贝母的检测能力较弱,推测图像集中的ROI较大,导致目标检测模型对浙贝母的尺度特征不敏感。

目标检测模型对尺度特征的敏感度与感受野的大小有关,通常可通过调整网络深度来控制。一般而言,网络层数越深,感受野越大,目标检测模型的特征提取能力越强,检测效果就越好。但网络层数达到一定限度后,随着网络层数的增加,模型的错误率也会增加,信息丢失会更加严重;同时,网络层数加深可能会导致梯度消失和色散问题,从而影响模型的检测精度。此外,在训练过程中,当模型对特征图进行下采样时,也会丢失一些小目标特征信息,导致检测精度下降。空洞卷积是针对图像语义分割问题中因下采样导致图像分辨率降低、信息丢失而提出的一种卷积方法,能够在不增加卷积核数量或大小的条件下,提供更大的感受野,使之输出更大范围的信息(long-ranged information),也可避免因进行池化(pooling)而导致小目标信息丢失的情况。感受野大小按照公式(2)~(3)计算,带填充的空洞卷积输出图片的高和宽按照公式(4)~(5)计算。

$$F_1 = (r - 1) \times (k - 1) + k; \quad (2)$$

$$F_2 = 2 \times (r - 1) \times (k - 1) + k; \quad (3)$$

$$H_{out} = \frac{H_{in} + 2 \times p - r \times (k - 1) - 1}{s} + 1; \quad (4)$$

$$W_{out} = \frac{W_{in} + 2 \times p - r \times (k - 1) - 1}{s} + 1. \quad (5)$$

式中: F_1 为正常空洞卷积的感受野大小; F_2 为带填充的空洞卷积的感受野大小; r 为空洞率,其值越大,表示模型的感受野越大; k 为卷积核大小; p 为边缘填充值; H_{in} (H_{out})和 W_{in} (W_{out})分别为输入(输出)图片的高和宽; s 为步距。

预选算法中YOLO-X的检测效果较好,因此本研究选择在YOLO-X主干特征提取网络的末端(主

干特征提取网络输出层 Dark5 处)添加一层空洞卷积结构(图3A)。为使所添加的空洞卷积结构与原模型相应位置的输入、输出张量尺寸相匹配,本研究选择带填充的空洞卷积作为嵌入结构。如图3B所示,原 Dark5 模块处输出张量尺寸为 $20 \times 20 \times 1\,024$,在经过空洞卷积处理后,根据公式(1)、(3)~(5)计算,其张量尺寸仍为 $20 \times 20 \times 1\,024$,与 FPN 对应接收端的张量尺寸保持一致,以便后续特征融合与重用。为探究最佳空洞率,本研究测试了空洞率为2、3、4、5、6时的空洞卷积结构。改进后的主干特征提取网络命名为 CSPDarkNet53-DC,对应的目标检测算法命名为 YOLOX-DC(图3)。

1.3.5 测试结果评价指标

对于统计学习模型,采用精确率(precision, P)、召回率(recall, R)、 F_1 和准确率(accuracy, A)作为评价指标;对于目标检测模型,采用平均精确率(average precision, AP)、平均精确率均值(mean average precision, mAP)、 F_1 和每秒传输帧数(frames per second, FPS)作为评价指标^[27]。试验根据预测结果对正负例(正负样本)划分的正确性将所有

样本划分为4种类型,分别为真正例(true positive, TP)、真负例(true negative, TN)、假正例(false positive, FP)、假负例(false negative, FN)。其中,TP代表被正确分类的正样本,即某类浙贝母被正确分类的数量;TN代表被正确分类的负样本,即其余类别浙贝母被正确分类的数量;FN代表被错误分类的正样本,即某类浙贝母被错误分类的数量;FP代表被错误分类的负样本,即其余类别浙贝母被错误分类的数量。

利用上述4类数据对 P 、 R 、 A 进行定义。 P 的具体含义为在预测结果是正样本的所有样本中,模型预测正确的比例。 R 的具体含义为在真实结果是正样本的所有样本中,模型预测正确的比例。 A 的具体含义为在总样本中,模型预测正确的比例。 P 、 R 、 A 分别按照公式(6)~(8)计算。

$$P = \frac{TP}{FP + TP} \times 100\%; \quad (6)$$

$$R = \frac{TP}{FN + TP} \times 100\%; \quad (7)$$

$$A = \frac{TP + TN}{FP + TP + FN + TN} \times 100\%. \quad (8)$$

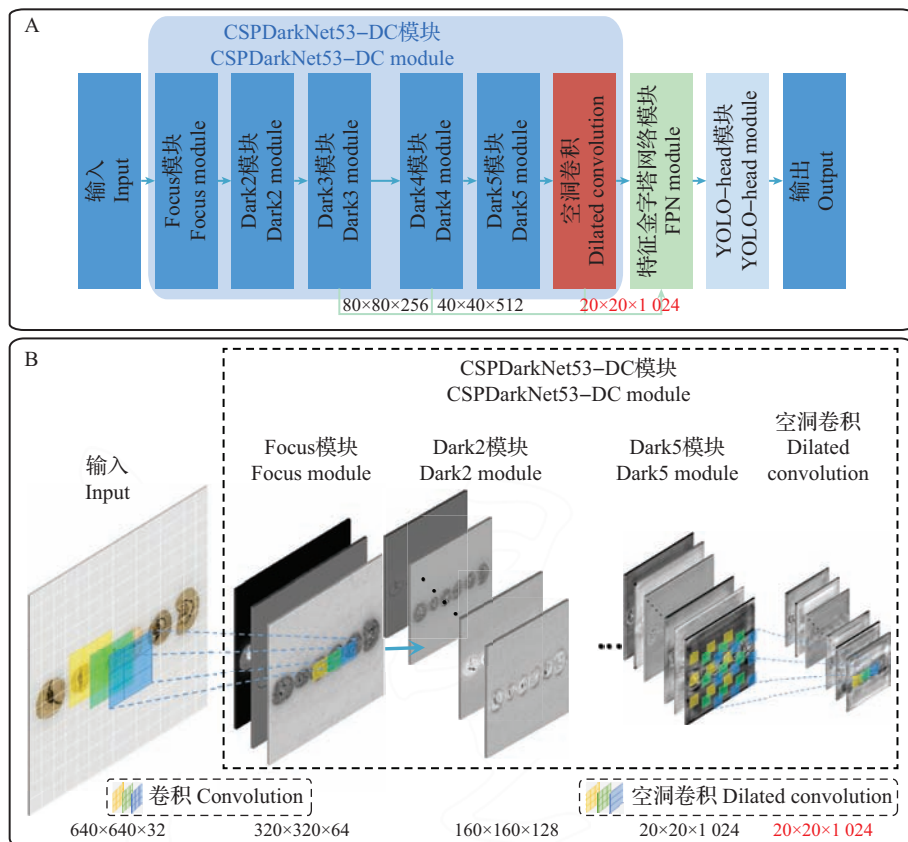


图3 YOLOX-DC结构(A)及其主干特征提取网络推断流程(B)示意图

Fig. 3 Schematic diagrams of YOLOX-DC structure (A) and its backbone feature extraction network inference process (B)

F_1 为 P 和 R 的加权调和平均数,综合考虑 R 和 P , F_1 按照公式(9)计算。

$$F_1 = \frac{2 \times P \times R}{P + R}. \quad (9)$$

将所有样本按其分类的置信度进行排序,并计算每个样本所对应的置信度作为正、负样本划分阈值中的 P 和 R ,对不同置信度下的 P 和 R 作图,得到精确率-召回率(precision-recall, P - R)曲线, P - R 曲线与坐标轴所围的面积为AP,且AP按照公式(10)计算。

$$AP = \int_0^1 P(R) dR \times 100\%. \quad (10)$$

mAP为所有类别AP的平均值,且mAP按照公式(11)计算。

$$mAP = \frac{\sum_{c=1}^C AP(c)}{C} \times 100\%, \quad c \in [1, C]. \quad (11)$$

式中, c 、 C 为样本类别数量。

FPS用来反映模型处理图像的速度,且FPS按照公式(12)计算。

$$FPS = \frac{N}{T_n}. \quad (12)$$

式中, N 和 T_n 分别表示模型在一段时间内处理图像的数量以及处理该帧数图像所用的时间。

2 结果与讨论

2.1 预选算法测试结果

试验拍摄了2 800张浙贝母图像作为原始数据集,并基于该数据集训练并测试DT、SVM、YOLO-V3、YOLO-V4、YOLO-V5、YOLO-X和Faster R-CNN算法。以 P 、 R 、 F_1 、 A 、AP、mAP和FPS作为模型的评价指标。

DT和SVM对破碎浙贝母的分类能力普遍较弱,但对其余类别浙贝母的分类更为准确。推测是由于破碎的浙贝母形状不规则、特征多变,因此很难利用固定的特征或阈值对其进行表征。此结果也验证了传统机器学习的局限性,即当原有数据的规律不明显以及研究人员无法充分提取有效特征时,数据便无法被有效利用,最终导致模型分类效果不佳。

对于目标检测算法,训练过程中特级、一级和二级浙贝母的训练收敛速度慢于虫蛀、霉变、破碎浙贝母,且所得模型对仅直径不同的特级、一级、二

级浙贝母的检测能力普遍较弱,其AP和 F_1 均低于虫蛀、破碎、霉变浙贝母的对应该指标(表2)。其中,YOLO-X对特级、一级、二级浙贝母的整体检测效果优于其他算法,其AP分别为98.39%、72.22%、96.59%, F_1 分别为0.90、0.63、0.84,但仍不及该模型对虫蛀、破碎、霉变浙贝母的检测效果。推测是因为虫蛀、破碎、霉变浙贝母颜色、形状和纹理等特征的差异可被目标检测算法有效识别,从而进行区分,而特级、一级、二级浙贝母在上述特征方面相似,所以只能以直径尺度为特征进行区分。此外,本研究的图像数据集中的ROI占比较大,模型的感受野相对较小,导致其对直径尺度特征不敏感,并且一级浙贝母的直径刚好介于特级浙贝母和二级浙贝母之间,所以模型容易将一级浙贝母误判为特级浙贝母或二级浙贝母,最终导致对一级浙贝母的检测效果不佳。这也说明在农产品检测领域,此类算法可能更适合检测以形状、颜色、纹理等为差异特征的产品。在检测速度方面,各目标检测算法的FPS均能满足实际检测要求。

2.2 YOLOX-DC 测试结果

根据预选算法的测试结果推测,如果要改善统计学习模型对破碎浙贝母的检测效果,就需要人为提取更多形态、颜色和纹理方面的特征;如果要改善目标检测模型对特级、一级、二级浙贝母的检测效果,则需要加强算法对尺度特征的敏感度。由于预选算法中YOLO-X所得模型的效果最佳,因此后续研究以YOLO-X为基础进行改进(改进后的算法命名为YOLOX-DC)。

目标检测模型对尺度特征的敏感度与感受野的大小有关,而空洞卷积可在不增加卷积核数量或大小的情况下提供更大的感受野,使之输出更大范围的信息,因此,本研究在YOLO-X主干特征提取网络的末端分别添加了一层空洞率为2、3、4、5、6的带有填充的空洞卷积结构。结果(表3)表明:空洞率为4时,模型的检测效果最佳,其mAP达到98.95%,特级、一级、二级浙贝母的AP分别为99.97%、98.33%、98.47%, F_1 分别为0.99、0.92、0.94。与原模型相比,空洞率为4的改进后模型(以下简称“改进后模型”)的mAP提高了4.95个百分点,一级浙贝母的AP提高了26.11个百分点, F_1 提高了0.29。原模型与改进后模型的检测效果如图4~5所示:前者对图像的误检或漏检多发生在特级、一级、二级浙贝母上;后者

表 2 预选算法训练所得模型在浙贝母测试集上的测试结果
Table 2 Test results of models trained by preselective algorithms on the test set of *F. thunbergii*

模型 Model	指标 Index	浙贝母品质 Quality of <i>F. thunbergii</i>					
		特级 Superfine	一级 Level one	二级 Level two	虫蛀 Moth-eaten	霉变 Mildewed	破碎 Broken
DT	$P/\%$	95.00	66.14	70.83	88.06	77.12	68.80
	$R/\%$	64.41	67.20	80.19	94.40	77.78	75.19
	F_1	0.77	0.67	0.75	0.91	0.78	0.72
	$A/\%$	76.53					
SVM	$P/\%$	95.76	84.40	85.04	99.23	90.83	86.32
	$R/\%$	87.60	95.97	95.58	98.47	86.84	75.23
	F_1	0.92	0.90	0.90	0.99	0.89	0.80
	$A/\%$	90.28					
YOLO-V3	$AP/\%$	61.14	40.00	37.90	97.67	93.04	95.48
	F_1	0.24	0.05	0.00	0.93	0.85	0.94
	mAP/%	70.87					
	FPS	29.10					
YOLO-V4	$AP/\%$	92.93	68.86	88.37	99.25	94.85	96.87
	F_1	0.85	0.76	0.85	0.92	0.92	0.95
	mAP/%	90.19					
	FPS	42.33					
YOLO-V5	$AP/\%$	64.32	84.22	96.18	98.95	99.89	96.20
	F_1	0.70	0.71	0.87	0.97	0.98	0.97
	mAP/%	89.96					
	FPS	30.95					
YOLO-X	$AP/\%$	98.39	72.22	96.59	98.84	99.53	98.81
	F_1	0.90	0.63	0.84	0.95	0.98	0.96
	mAP/%	94.06					
	FPS	30.65					
Faster R-CNN	$AP/\%$	88.91	69.75	92.52	98.91	99.48	98.61
	F_1	0.79	0.66	0.83	0.96	0.97	0.95
	mAP/%	91.36					
	FPS	36.28					

对各类别浙贝母的判别均正确,且置信度较高,预测框与浙贝母紧密贴合,呈现外接矩形状态,说明模型对其位置的回归也相对更为准确。

2.3 可视化中间激活

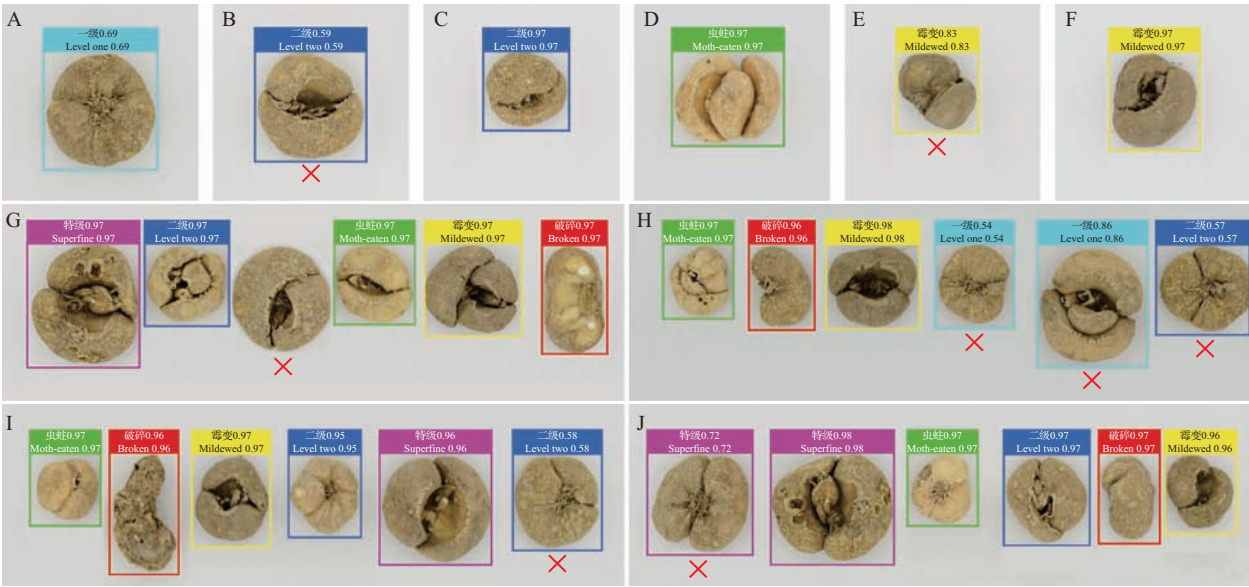
相较于手动提取特征的统计学习,深度学习可以自动提取更为抽象的特征。为分析原模型与改进后模型所学特征的差异,本研究通过可视化中间激活的方式加以探究。所谓可视化中间激活,即将网络中各卷积层和池化层等输出的特征图进行可视化展示。由于特征图数量庞大,本试验基于浙贝母单目标和多目标图像,将YOLO-X与YOLOX-DC模型每层输出的特征图间隔5张进行可视化,并选择2个模型的有效主干特征提取网络(Dark3、Dark4、

Dark5)对应位置处的中间激活进行对比。

如图6~7所示:随着神经网络层数的加深以及对输入图片的重复下采样,其中间激活的像素值越来越小,图像愈发模糊、抽象,但YOLOX-DC模型的中间激活较YOLO-X更加清晰,尤其在深度为Dark5时,YOLO-X输出的中间激活基本失真,而YOLOX-DC输出的中间激活仍保有其外形轮廓。此外,对于单目标图像,YOLOX-DC中间激活的ROI普遍比YOLO-X大(图6);对于多目标图像,YOLOX-DC中间激活的ROI外层普遍有一圈清晰可见的包络线(图7),说明改进后模型在对图像进行前向推断的过程中,与浙贝母之间保持着较强的联系,即改进后模型将多目标看作单一整体,避免

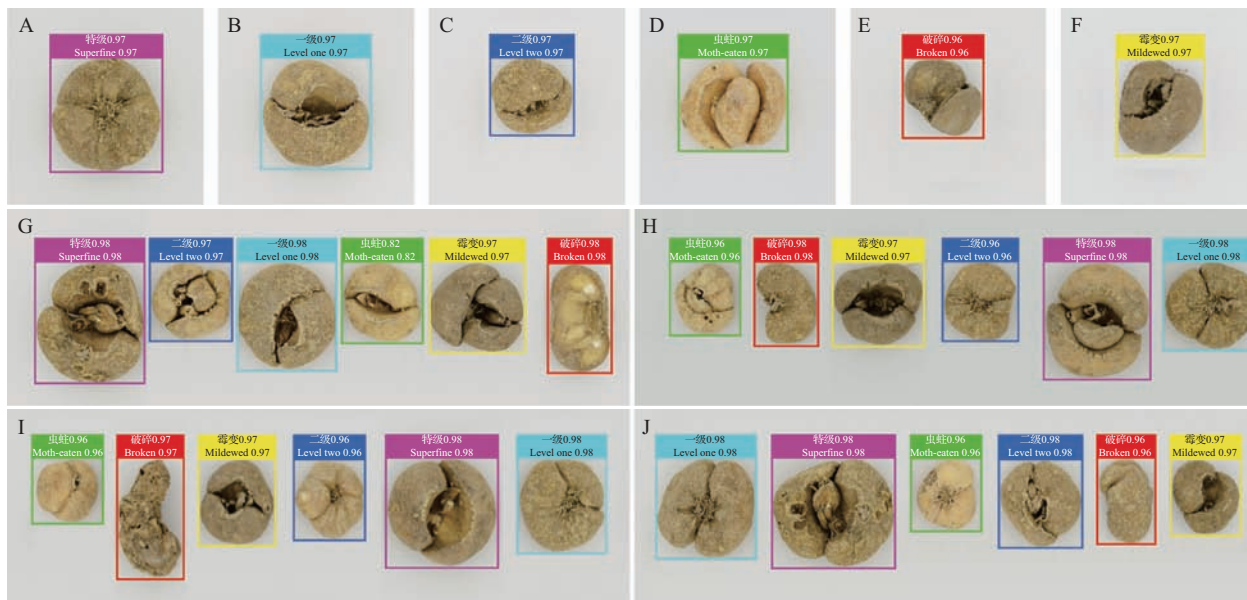
表 3 不同空洞率的YOLOX-DC训练所得模型在浙贝母测试集上的测试结果
Table 3 Test results of models trained by YOLOX-DC with different dilated rates on the test set

空洞率 Dilated rate	指标 Index	浙贝母品质 Quality of <i>F. thunbergii</i>					
		特级 Superfine	一级 Level one	二级 Level two	虫蛀 Moth-eaten	霉变 Mildewed	破碎 Broken
2	AP/%	99.51	93.53	96.97	97.85	99.77	97.48
	F_1	0.95	0.84	0.81	0.96	0.98	0.97
	mAP/%	97.52					
	FPS	28.97					
3	AP/%	99.37	98.52	97.62	97.75	99.93	98.37
	F_1	0.97	0.94	0.93	0.97	0.99	0.95
	mAP/%	98.59					
	FPS	29.18					
4	AP/%	99.97	98.33	98.47	98.71	99.73	98.85
	F_1	0.99	0.92	0.94	0.97	0.99	0.97
	mAP/%	99.01					
	FPS	29.13					
5	AP/%	98.63	95.83	94.89	98.34	99.86	98.99
	F_1	0.94	0.81	0.89	0.96	0.98	0.97
	mAP/%	97.76					
	FPS	29.02					
6	AP/%	98.16	94.31	98.77	98.65	99.55	97.11
	F_1	0.92	0.86	0.75	0.97	0.92	0.96
	mAP/%	97.76					
	FPS	29.05					



A~F. 单目标图像检测结果;G~J. 多目标图像检测结果。图片下方红色叉代表模型检测时发生误检或漏检。标签中数字为模型检测浙贝母类别时的置信度,图 5 同。
A~F. Single object image detection results; G~J. Multi-object image detection results. The red crosses at the bottom of the image represent false detections or missed detections during the model detection. The number in the label is the confidence of the model when detecting the class of *F. thunbergii*, and the same as in Fig. 5.

图 4 YOLO-X 的预测效果
Fig. 4 Prediction results by YOLO-X



A~F. 单目标图像检测结果;G~J. 多目标图像检测结果。

A~F. Single object image detection results; G~J. Multi-object image detection results.

图5 YOLOX-DC(空洞率为4)的预测效果

Fig. 5 Prediction results by YOLOX-DC (dilated rate=4)

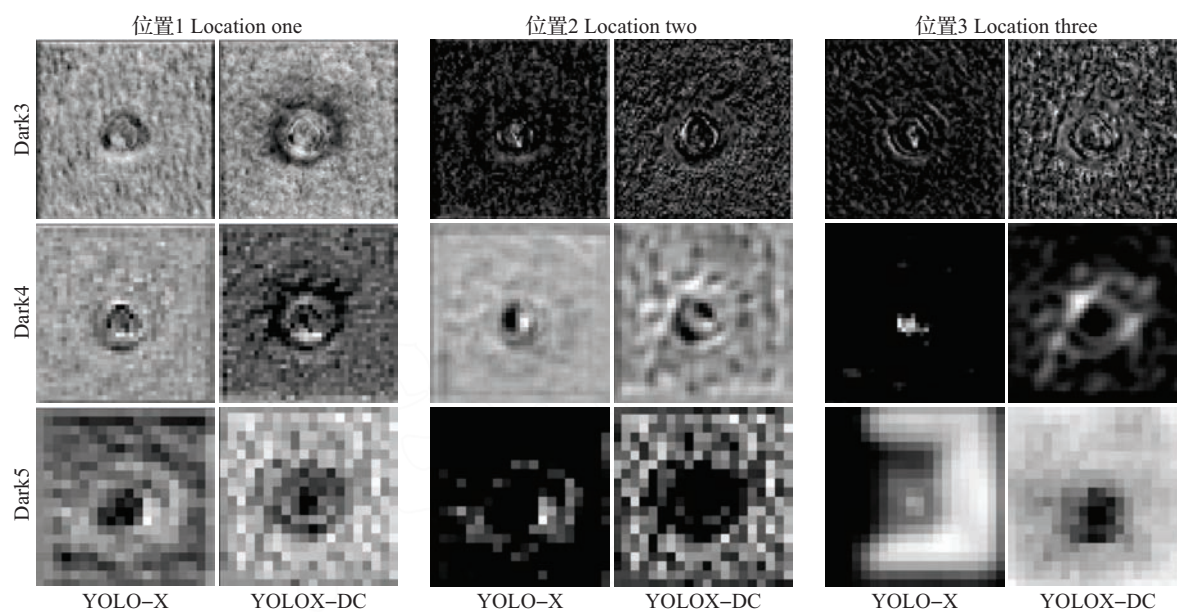


图6 YOLO-X和YOLOX-DC基于浙贝母单目标图像对应位置的中间激活

Fig. 6 Intermediate activation based on the corresponding positions of single object images of *F. thunbergii* by YOLO-X and YOLOX-DC

了“盲人摸象”的情况,这也是感受野扩大的直观体现。以上可视化中间激活的结果直观地证明,向模型的主干特征提取网络嵌入空洞卷积的方式可以提高其对尺度特征的敏感度。受此启发,对于其他数据集而言,YOLOX-DC也可根据其ROI的不同设置不同的空洞率,以调整感受野大小,从而适应

该数据集的尺度特征,进而更好地实现对主要以尺度为特征产品的检测。

3 结论

本研究将机器视觉和机器学习技术应用于浙

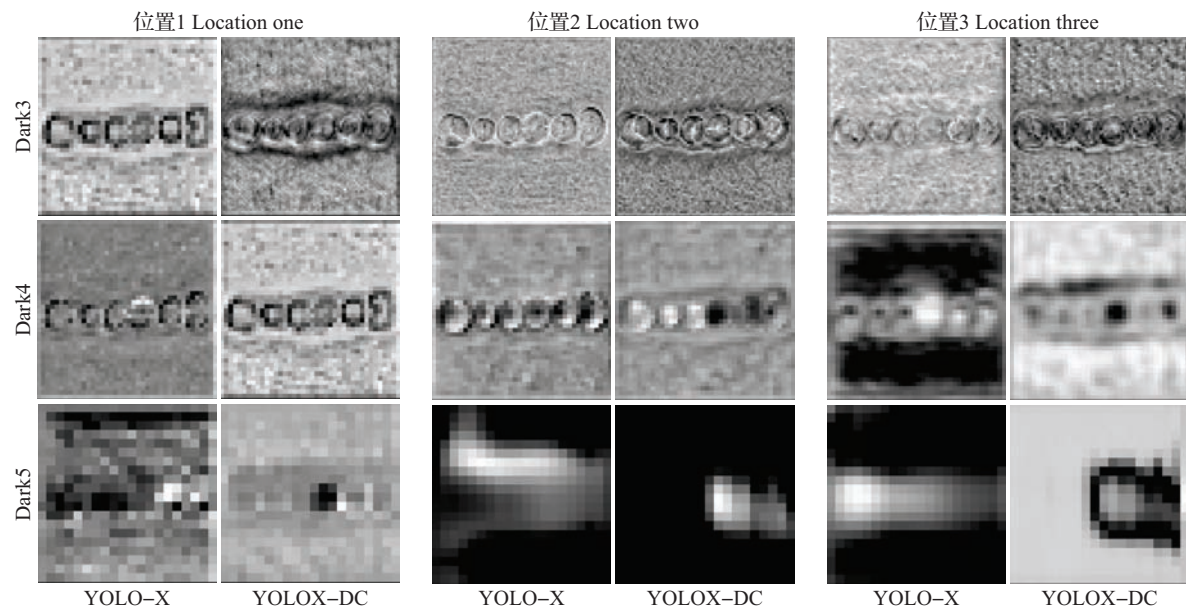


图7 YOLO-X和YOLOX-DC基于浙贝母多目标图像对应位置的中间激活

Fig. 7 Intermediate activation based on the corresponding positions of multi-object images of *F. thunbergii* by YOLO-X and YOLOX-DC

贝母外观品质及直径的检测与区分中。首先对比了统计学习和深度学习算法对浙贝母分级的效果,结果显示YOLO-X所得模型的检测效果较好;其次针对浙贝母的特点,向YOLO-X主干特征提取网络的末端嵌入一层空洞率为4的带填充的空洞卷积结构,在不增加参数量、计算量或对原模型进行大规模改动的情况下,促进了改进位点处的特征融合与重用,扩大了模型的感受野,提高了模型对尺度特征的敏感度,从而更好地适应试验创建的浙贝母数据集。与其他目标检测算法相比,改进后算法所得模型(空洞率为4)针对浙贝母图像数据集的检测效果更佳,其特级、一级、二级、虫蛀、霉变、破碎浙贝母的AP分别为99.97%、98.33%、98.47%、98.71%、99.73%、98.85%, F_1 分别为0.99、0.92、0.94、0.97、0.99、0.97,mAP为99.01%,FPS为29.13。本研究还通过可视化中间激活的方式,直观地对比了改进前后模型所学特征的差异,从而对所作改进加以解释与分析。本研究为浙贝母无损检测平台的搭建提供了科学依据,有望推动浙贝母市场的发展。同时,本研究提出的方法还可为其他同类产品的无损检测提供思路。

参考文献(References):

- [1] 黄政晖,章勇杰,倪忠进,等.丘陵山区浙贝母机械化生产模式的效益分析[J].南方农机,2021,52(16):1-3. DOI:10.3969/

j.issn.1672-3872.2021.16.001

- HUANG Z H, ZHANG Y J, NI Z J, et al. Benefit analysis of mechanized production mode of *Fritillaria thunbergii* in hilly and mountainous areas[J]. *China Southern Agricultural Machinery*, 2021, 52(16): 1-3. (in Chinese with English abstract)
- [2] 中华中医药学会. 中药材商品规格等级 浙贝母: T/CACM 1021.24—2018[S]. 北京: 中国标准出版社, 2018.
- China Association of Chinese Medicine. *Commercial Grades for Chinese Medicinal Materials—Fritillariae thunbergii Bulbus: T/CACM 1021.24—2018*[S]. Beijing: Standards Press of China, 2018. (in Chinese)
- [3] 王金辉,田忠静,陈启永. 贝母筛分机的设计与研究[J]. 安徽农学通报, 2013, 19(3): 143-144. DOI:10.16377/j.cnki.issn1007-7731.2013.03.029
- WANG J H, TIAN Z J, CHEN Q Y. Design of the *Fritillaria* sieving machine[J]. *Anhui Agricultural Science Bulletin*, 2013, 19(3): 143-144. (in Chinese with English abstract)
- [4] 宋江,邱胜蓝. 基于UG的平贝母等级筛分机设计[J]. 机电工程技术, 2013, 42(8): 31-33. DOI:10.3969/j.issn.1009-9492.2013.08.007
- SONG J, QIU S L. Design of *Fritillaria ussuriensis* grade screen vessel based on UG[J]. *Mechanical & Electrical Engineering Technology*, 2013, 42(8): 31-33. (in Chinese with English abstract)
- [5] 宋江,邱胜蓝. 平贝母等级筛分机机架的有限元分析[J]. 机电工程技术, 2014, 43(8): 68-69. DOI:10.3969/j.issn.1009-9492.2014.08.019
- SONG J, QIU S L. The finite element analysis of *Fritillaria ussuriensis* grade screen vessel frame[J]. *Mechanical & Electrical Engineering Technology*, 2014, 43(8): 68-69. (in Chinese with English abstract)

- [6] 宋江,王明,刘丽华,等.振动式平贝母等级筛分机设计与试验[J].黑龙江八一农垦大学学报,2013,25(1):28-31. DOI:10.3969/j.issn.1002-2090.2013.01.006
SONG J, WANG M, LIU L H, et al. Design and test of vibration type *Fritillaria ussuriensis* grade screen vessel[J]. *Journal of Heilongjiang Bayi Agricultural University*, 2013, 25(1): 28-31. (in Chinese with English abstract)
- [7] 杨东,王纪华,陆安祥.肉品质量无损检测技术研究进展[J].食品安全质量检测学报,2015,6(10):4083-4090. DOI:10.19812/j.cnki.jfsq11-5956/ts.2015.10.054
YANG D, WANG J H, LU A X. Research development on non-destructive determination technology for meat product quality[J]. *Journal of Food Safety & Quality*, 2015, 6(10): 4083-4090. (in Chinese with English abstract)
- [8] 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. DOI: 10.1016/j.biosystemseng.2019.11.011
- [9] KAMILARIS A, PRENAFETA-BOLDÚ F X. Deep learning in agriculture: a survey[J]. *Computers and Electronics in Agriculture*, 2018, 147: 70-90. DOI: 10.1016/j.compag.2018.02.016
- [10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016: 779-788. DOI: 10.1109/CVPR.2016.91
- [11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017: 6517-6525. DOI: 10.1109/CVPR.2017.690
- [12] REDMON J, FARHADI A. YOLOv3: an incremental improvement[CP/OL]. *arXiv*, 2018: abs/1804.02767. <https://doi.org/10.48550/arXiv.1804.02767>
- [13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[CP/OL]. *arXiv*, 2020: abs/2004.10934. <https://doi.org/10.48550/arXiv.2004.10934>
- [14] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[CP/OL]. *arXiv*, 2021: abs/2107.08430. <https://doi.org/10.48550/arXiv.2107.08430>
- [15] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 580-587. DOI: 10.1109/CVPR.2014.81
- [16] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448. DOI: 10.1109/ICCV.2015.169
- [17] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137-1149. DOI: 10.1109/TPAMI.2016.2577031
- [18] TIAN Y N, YANG G D, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. *Computers and Electronics in Agriculture*, 2019, 157: 417-426. DOI: 10.1016/j.compag.2019.01.012
- [19] LÜ J D, XU H, HAN Y, et al. A visual identification method for the apple growth forms in the orchard[J]. *Computers and Electronics in Agriculture*, 2022, 197: 106954. DOI: 10.1016/j.compag.2022.106954
- [20] ZHANG Y C, ZHANG W B, YU J Y, et al. Complete and accurate holly fruits counting using YOLOX object detection [J]. *Computers and Electronics in Agriculture*, 2022, 198: 107062. DOI: 10.1016/j.compag.2022.107062
- [21] XU W K, ZHAO L G, LI J, et al. Detection and classification of tea buds based on deep learning[J]. *Computers and Electronics in Agriculture*, 2022, 192: 106547. DOI: 10.1016/j.compag.2021.106547
- [22] ZHENG T X, JIANG M Z, LI Y F, et al. Research on tomato detection in natural environment based on RC-YOLOv4[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107029. DOI: 10.1016/j.compag.2022.107029
- [23] QI J T, LIU X N, LIU K, et al. An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease[J]. *Computers and Electronics in Agriculture*, 2022, 194: 106780. DOI: 10.1016/j.compag.2022.106780
- [24] ZHANG D Y, LUO H S, WANG D Y, et al. Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YOLOv5 method[J]. *Computers and Electronics in Agriculture*, 2022, 198: 107086. DOI: 10.1016/j.compag.2022.107086
- [25] YAO Z S, LIU T, YANG T L, et al. Rapid detection of wheat ears in orthophotos from unmanned aerial vehicles in fields based on YOLOX[J]. *Frontiers in Plant Science*, 2022, 13: 851245. DOI: 10.3389/fpls.2022.851245
- [26] LI X, PAN J D, XIE F P, 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. DOI: 10.1016/j.compag.2021.106503
- [27] 谢鹏尧,富昊伟,唐政,等.基于RGB图像的冠层尺度水稻叶瘟病斑检测与抗性评估[J].浙江大学学报(农业与生命科学版),2021,47(4):415-428. DOI:10.3785/j.issn.1008-9209.2021.05.131
XIE P Y, FU H W, TANG Z, et al. RGB imaging-based detection of rice leaf blast spot and resistance evaluation at the canopy scale[J]. *Journal of Zhejiang University (Agriculture & Life Sciences)*, 2021, 47(4): 415-428. (in Chinese with English abstract)