doi:10.6041/j.issn.1000-1298.2023.06.023

基于改进 SSD 卷积神经网络的苹果定位与分级方法

张立杰　周舒骅　李　娜　张延强　陈广毅　高　笑
(河北农业大学机电工程学院, 保定 071001)

摘要: 为实现苹果果径与果形快速准确自动化分级,提出了基于改进型 SSD 卷积神经网络的苹果定位与分级算法。深度图像与两通道图像融合提高苹果分级效率,即对从顶部获取的苹果 RGB 图像进行通道分离,并提取分离通道中影响苹果识别精度最大的两个通道与基于 ZED 双目立体相机从苹果顶部获取的苹果部分深度图像进行融合,在融合图像中计算苹果的纵径相关信息,实现了基于顶部融合图像的多个苹果果形分级和信息输出;使用深度可分离卷积模块替换原 SSD 网络主干特征提取网络中部分标准卷积,实现了网络的轻量化。经过训练的算法在验证集下的识别召回率、精确率、mAP 和 F1 值分别为 93.68%、94.89%、98.37% 和 94.25%。通过对比分析了 4 种输入层识别精确率的差异,实验结果表明输入层的图像通道组合为 DGB 时对苹果的识别与分级 mAP 最高。在使用相同输入层的情况下,比较原 SSD、Faster R-CNN 与 YOLO v5 算法在不同果实数目下对苹果的实际识别定位与分级效果,并以 mAP 为评估值,实验结果表明改进型 SSD 在密集苹果的 mAP 与原 SSD 相当,比 Faster R-CNN 高 1.33 个百分点,比 YOLO v5 高 14.23 个百分点。并且在不同硬件条件下验证了该算法定位分级效率的优势,单幅图像在 GPU 下的检测时间为 5.71 ms,在 CPU 下的检测时间为 15.96 ms,检测视频的帧率达到 175.17 f/s 和 62.64 f/s。该研究可为自动化分级设备在高速环境下精准定位并分级苹果提供理论基础。

关键词: 苹果分级; 信息融合; 改进型 SSD; 卷积神经网络; 目标检测
中图分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2023)06-0223-10　　**OSID**:

Apple Location and Classification Based on Improved SSD Convolutional Neural Network

ZHANG Lijie　ZHOU Shuhua　LI Na　ZHANG Yanqiang　CHEN Guangyi　GAO Xiao
(*School of Mechanical and Electrical Engineering*, *Hebei Agricultural University*, *Baoding* 071001, *China*)

Abstract: An apple localization and grading algorithm was proposed based on an improved SSD convolutional neural network to achieve fast and accurate automatic grading of apple fruit diameter and shape. The efficiency of apple grading was improved by improving the input layer of the original SSD network. Channel separation was performed on the color apple image obtained from the top, and the two channels in the separation channel that had the most significant impact on the apple recognition accuracy were extracted. A fused image was composed of the two channels and the apple depth image from the top based on the binocular camera. The longitudinal diameter-related information of the apple was calculated in the fused image. Moreover, multiple apple shape grading and information output based on the fused image were realized through this method. The depthwise-separable convolution module was used to replace part of the standard convolution in the original SSD network backbone feature extraction network, which achieved the light weighting of the network. The recognition recall, accuracy, mAP and F1 values of the trained model under the verification set were 93.68%, 94.89%, 98.37% and 94.25%, respectively. By comparing and analyzing the differences in recognition accuracy among the four input layers, the experimental results showed that the highest recognition and grading mAP for apples was achieved when the image channel combination of the input layer was DGB. The actual recognition

收稿日期: 2022-10-28　修回日期: 2023-02-14
基金项目: 河北省重点研发计划项目(21321902D)
作者简介: 张立杰(1969—),男,教授,博士,主要从事机器人技术与自主系统研究,E-mail: ljzhang@ysu.edu.cn
通信作者: 李娜(1981—),女,副教授,博士,主要从事机器人智能化与控制研究,E-mail: res_lina@126.com

localization and grading effects of the original SSD, Faster R - CNN and YOLO v5 algorithms for apples with different numbers of fruits were compared by using the same input layer and evaluated in terms of mAP. The experimental results showed that the improved SSD had a comparable mAP to the original SSD for dense apples, which was higher than that of Faster R - CNN by 1.33 percentage points and higher than YOLO v5 by 14.23 percentage points. The advantages of the algorithm localization and grading efficiency were verified under different hardware conditions. The detection time of an image was 5.71 ms under GPU and 15.96 ms under CPU, and the actual frame rate of the detected video reached 175.17 f/s and 62.64 f/s. The research result can provide a theoretical basis for automated grading equipment to accurately locate and grade apples in a high-speed environment.

Key words: apple classification; information fusion; improved SSD; convolutional neural network; object detection

0 引言

随着科技的发展,水果自动化分级设备日益完善,视觉识别系统作为水果自动化分级设备的一部分,其目标检测的速度与分级准确率对水果自动化分级的工作效率和工作时长有较大影响。稳定快速的水果识别分级视觉系统可以让水果自动化分级设备长时间稳定而快速的工作,压缩劳动成本,提高生产效率。高精度且高效率的苹果定位分级算法对提高农业生产力有较大的应用价值和现实意义。

目前水果视觉定位与分级方面的研究已经取得了一定的进展[1]。李先锋等[2]通过 D-S 证据理论进行决策级融合,对苹果形状和颜色进行分级,分级正确率达 93.75%。夏卿等[3]提出基于粒子群优化(Particle swarm optimization, PSO)改进算法的最小二乘向量机(LS-SVM)的苹果分级方法,分级准确率达到 96%。相比于常规方法,深度卷积神经网络在苹果识别领域显现出巨大的优越性[4-8],卷积神经网络由于其对目标高维特征的提取,使得对各等级苹果的快速在线识别变成可能。何进荣等[9]提出了基于多卷积神经网络融合 DXNet 模型的苹果外部品质分级方法,分级准确率达到 97.84%。目前的目标检测与分类卷积神经网络主要分为两种,一种是基于区域预测的目标检测算法,典型的算法有 RCNN[10]、Fast R-CNN[11] 和 Faster R-CNN[12-13],称为两步法(Two-stage)目标检测,核心思想是先获得建议区域然后在当前区域内进行分类。另一种方法是无区域预测的目标检测算法,典型算法有 YOLO[14-15]、SSD[16],称为一步法(One-stage)目标检测,核心思想是用单一的卷积神经网络直接使用整幅图像来预测目标的位置及其种类。

本文提出用改进型 SSD 算法来实现苹果自动化高速分级与定位,在不影响精度的情况下压缩网络规模来加快检测速度[17-21],在不同的输入层条件下分别进行网络的识别实验,以验证不同输入层对检测精度的影响,在相同输入层条件下进行网络的检测实验,以检测网络的效果,并实验验证实时性。

1 目标识别分级算法

1.1 识别分级对象分析

实验所用苹果均为产自陕西省洛川市的红富士苹果,图像采集使用苹果自动化分级设备中的视觉采集装置,其具有 LED 补光灯,通过 ZED 双目立体相机采集苹果顶部的 RGB 图像与深度图像。根据表 1 中的各项指标将苹果分为特等果、一等果、二等果、等外果 4 个等级,分别采集人工分级后苹果顶部的 RGB 图像与其对应的深度图像。

表 1 苹果等级分级标准

Tab.1 Apple classification standard

指标	特等果	一等果	二等果	等外果
果径/mm	≥75	[70,75)	[65,70)	<65
果形系数	≥0.85	[0.80,0.85)	[0.75,0.80)	<0.75

4 种等级苹果顶部的 RGB 图像与深度图像如图 1 所示,苹果均顶部朝上,并随机地分布在传送带上,RGB 图像中苹果的形状特征清晰,并与绿色的传送带形成鲜明对比,深度图像中苹果表面的深度信息完整,与传送带形成明显的距离差。

如果仅使用苹果顶部的 RGB 图像对苹果进行分级,只能通过苹果的果径特征对苹果进行分类。为了对苹果的果形特征进行分级,一般需要同时采集苹果的侧面 RGB 图像获取苹果纵径来计算果形系数,对于这种随机分布在传送带上的苹果,无法对其侧面图像进行获取,因此无法对其果径与果形进行综合分级操作。

因此本文考虑用 RGB 图像与深度图像进行特征融合后的图像,通过卷积神经网络来提取苹果特征,采用端到端的整体训练让神经网络自适应地学习不同等级苹果的特征,实现对随机分布在传送带上苹果的识别和分级。

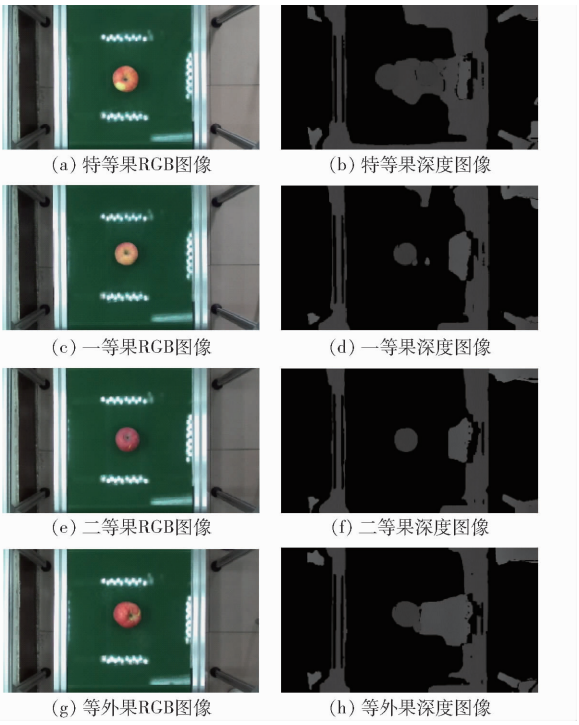

图 1　4 种等级苹果的 RGB 图像与深度图像

Fig. 1　Four levels of apple's RGB image and depth image

1.2　SSD 算法

SSD 算法是目前为止最先进的目标检测方案之一，它能够在一幅图像中同时检测和分类对象。

1.2.1　基于 SSD 算法的苹果识别分级

本文提出的基于改进型 SSD 算法的苹果定位与分级方法能够定位并分级视频中的苹果并返回其坐标与对应等级。不同于区域预测的目标检测算法（Faster R－CNN），SSD 算法通过单个卷积神经网络检测整个图像回归目标的类别和位置，SSD 算法采用 CNN 直接来进行检测，而不是像 YOLO 那样在全连接层之后进行检测。为让卷积神经网络遍历整幅图像后能够得到固定格式的预测尺寸，首先将图像调整为固定尺寸 300 像素 ×300 像素，再分成 6 种不同尺寸的网格；然后在这些网格上面的每一个点构造 6 个不同尺度的默认框（Default boxes），分别进行检测和分类，生成多个初步符合条件的默认框；最后，将不同尺寸网格获得的默认框结合起来，经过非极大抑制（Non-maximum suppression，NMS）方法过滤掉一部分重叠或者不正确的默认框，生成最终检测结果，具体包含 5 个预测参数：代表目标坐标 x、y，目标外包矩形的宽度 w 和高度 h，用来通过阈值对预测结果进行取舍的置信度。

1.2.2　神经网络改进设计

原 SSD 算法卷积层主要由 VGG－Base、Extra－Layers、Pred－Layers 3 部分组成。VGG－Base 作为基础框架用来提取图像的特征，Extra－Layers 对 VGG 的 feature 做进一步处理，增加算法对图像的感受野，使得 Extra－Layers 得到的特征图承载更多抽象信息。待预测的特征图由 6 种特征图组成，最终通过 Pred－Layer 得到 6 种特征图预测框的坐标、置信度、类别信息。为了提高算法的检测效率，运用一种深度可分离卷积块替代 VGG－Base 中的标准卷积以减小算法数据量和算法检测所用时间。同时本文对苹果的分级条件是苹果的果径和果形这两个因素，并且仅通过苹果的顶部图像进行分级，需要引入与苹果纵径相关的图像信息，所以对输入层进行修改，同时输入 RGB 图像与深度图像。深度图像与 RGB 图像通道融合，将 RGB 图像的 3 个通道进行分离并去除对目标检测影响最小的通道，使用深度图代替去除的通道将融合后的三通道图像作为输入层输入网络进行检测[22]，通道融合过程如图 2 所示。融合后的图像包含 2 个颜色通道和 1 个深度通道，其中深度通道用于引入深度信息实现对纵径的预测，由于 ZED 双目立体相机与传送带之间的距离是固定的，所以深度图中可以通过苹果顶部深度特征来预测苹果纵径。而结合保留的 GB 两个通道所具备的颜色与果径特征可以通过卷积神经网络对苹果的位置与等级进行综合预测。

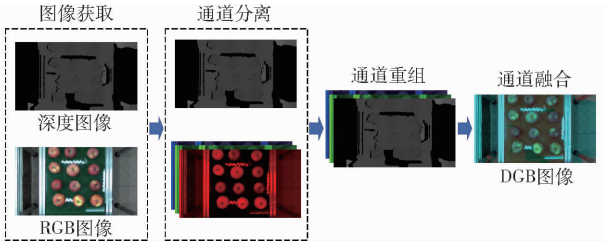

图 2　输入层信息融合过程示意图

Fig. 2　Schematic of input layer information fusion process

本文设计的神经网络使用 1 层卷积核为 3 ×3、步长为 2 的标准卷积对输入层的图像进行初步特征提取，输出为 150 ×150 ×64，继续用 13 层深度可分离卷积块对图像特征进行提取，最后输出为 19 ×19 ×1 024，深度可分离卷积块的结构如图 3 所示。首先，逐通道卷积（Depthwise convolution）分别对每

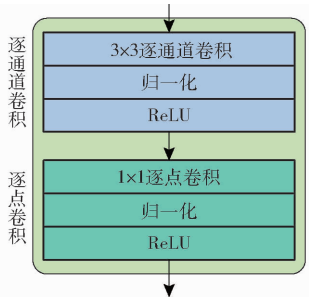

图 3　深度可分离卷积块

Fig. 3　Depthwise separable convolution

个输入通道进行 3×3 卷积,并对其进行归一化,使用 ReLU 函数进行激活。然后,逐点卷积(Pointwise convolution)进行 1×1 卷积,以调整输出通道数,使深度可分离卷积块与标准卷积具有相同的输出维度,同样进行归一化并使用 ReLU 函数进行激活[23-24]。接着使用 4 层标准卷积继续进行特征提取,最后输出张量为 1×1×16。分别将输出张量为 38×38×256、19×19×1 024、10×10×512、5×5×256、3×3×256、1×1×256 的 6 个特征层上的每个像素映射到输入图像上产生若干个默认框,然后对每个默认框进行分类和回归,该分类和回归操作是通过 3×3×n×(classes+4)的卷积操作完成的。其中 n 代表 6 个特征层上每个像素生成的默认框数量,在大小为 38×38、3×3、1×1 的特征层上数量为 4,其余 3 个特征层上数量为 6;classes 为类别数量,本文设置为 4。对每个默认框逐级分类和回归后进行非极大抑制操作,即得出目标检测结果。具体结构如图 4 所示。

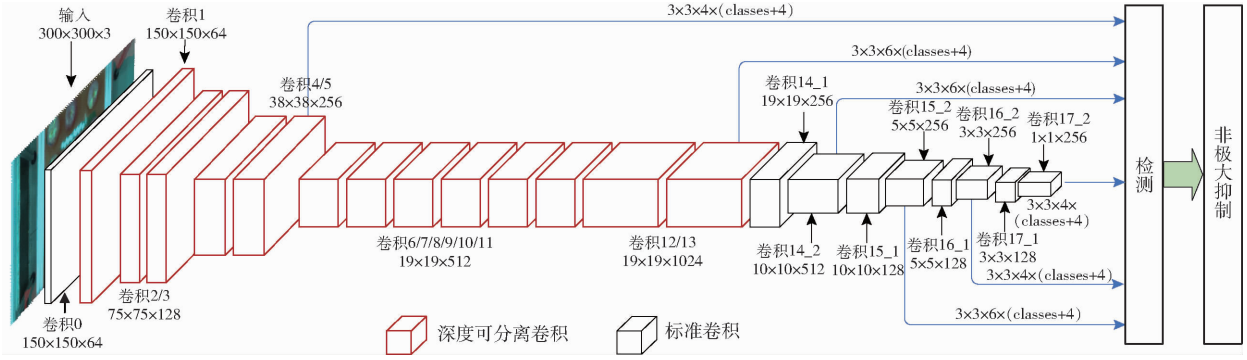

图 4 算法框架图

Fig. 4 Algorithm frame diagram

1.2.3 SSD 算法损失函数

SSD 算法的损失函数包含两部分,分别为对目标位置进行预测的位置误差(Localization loss, loc)与对目标分类进行预测的置信度误差(Confidence loss, conf)的加权和,分别采用 Smooth loss 和 Softmax loss 进行计算,总的系统损失计算式为

$$L(r,c,l,g)=\frac{1}{N}(L_{\text{conf}}(r,c)+\alpha L_{\text{loc}}(r,l,g)) \quad (1)$$

式中 r——预测框与真实框的匹配结果,如果匹配则设为 1,不匹配设为 0

c——Softmax 函数对每一个类别的置信度

l——预测标签框

g——实际标签框

N——默认框数量

L_{conf}、L_{loc}——类别置信损失函数和位置损失函数

α——加权系数,设置分类损失和位置损失的比例

损失函数会遍历所有默认框,如果默认框与真实框的标签匹配,就对损失函数进行计算,否则损失函数值为 0。

1.2.4 预测过程

预测过程是对于每个预测框,首先根据类别置信度确定其类别(置信度最大者)与置信度,并滤除属于背景的预测框。然后根据置信度阈值过滤掉阈值较低的预测框。对留下的预测框进行解码,根据默认框得到其真实的位置参数。最后进行 NMS,过滤掉那些重叠度较大的预测框。最后剩余的预测框就是检测结果。

1.3 苹果分级设备

分级设备如图 5 所示,苹果是顶部朝上并随机排列的,通过传送带送入分级设备然后进行分级。在实际应用中设备运行是按照以下过程:将苹果顶部朝上随机放在传送带上→输送苹果进入分级框→ZED 双目立体相机图像采集→改进型 SSD 网络识别与分级→Delta 机器人分拣。

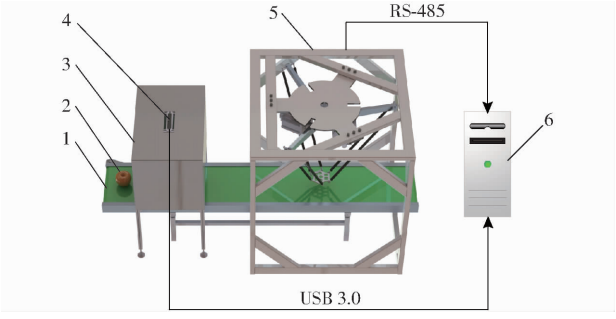

图 5 系统的硬件结构

Fig. 5 Hardware structure of system

1. 传送带 2. 苹果 3. 视觉采集箱 4. ZED 双目立体相机 5. Delta 机器人 6. 工作站

2 算法训练与调整

采集得到的苹果图像为 2 624 幅,其中特等果图像 480 幅,一等果图像 1 024 幅,二等果图像 558

幅，等外果图像 562 幅。其中 2 362 幅图像用来训练算法，262 幅图像用来验证算法，比例约为 9∶1。

2.1　实验运行平台

本文中所有的训练和测试均在同一台工作站进行，工作站的主要配置为 Intel Core（TM）i7－11700k CPU@ 3.6 GHz、8 GB 的 GPU GeForce RTX 3060 Ti 和 16 GB 的运行内存。所有程序均是由 Python 语言编写并调用 CUDA、Cudnn、OpenCV、Tensorflow 1.15、Keras 2.3.1 库并在 Ubuntu 20.04 系统下运行，训练和测试算法时分别单独使用 CPU 和 GPU。

训练时采用 16 个样本作为一个处理单元，每次更新权值时用 BN（Batch normalization）进行正则化，并且在每层中加入丢弃层（Dropout），以保持每层和提取特征的相互独立，其丢弃比例为 0.5，动量（Momentum）设置为 0.937，权值衰减（Decay）设置为 5×10^{-4}，优化器（Optimizer）选用 Adam，最大学习率（Max learning rate）设置为 0.002，最小学习率（Min learning rate）设置为 2×10^{-5}，学习率衰减遵循余弦退火（Cosine annealing），实际计算式为

$$\eta_t=\eta_{\min}^i+\frac{1}{2}(\eta_{\max}^i-\eta_{\min}^i)\left(1+\cos\left(\frac{T_{\text{cur}}}{T_i}\pi\right)\right) \tag{2}$$

式中　η_t——学习率

i——训练运行次数

$\eta_{\max}^i$、$\eta_{\min}^i$——学习率最大值、最小值

T_{cur}——当前迭代周期（Epoch）数

T_i——迭代周期总数

2.2　数据集预处理

选取经人工分级的不同等级苹果图像计算每个像素点对应的深度值获取深度图，分辨率与 RGB 图像相同。将 RGB 图像的 GB 通道提取出来并将深度图像与其重新组合得到 DGB 图像作为训练图像。为了更精准地对 DGB 图像中的苹果进行标注，首先使用 LabelImg 对 RGB 图像中的苹果位置进行准确标注，标注信息采用 PASCAL VOC 数据集的格式进行保存，其中包含目标苹果的类别信息和外包边框。然后将 RGB 图像替换为 DGB 图像，其中 DGB 图像的苹果位置和等级与 RGB 图像中的完全相同。接着进行归一化处理，将目标的实际数据除以图像的宽度和高度，得到的数据都在 0～1 的范围之内，使得训练时能够更快地读取数据，并且能够训练不同尺寸的图像，其具体格式为 5 个参数作一组数据，分别包含 index（类别的序列），x（目标中心的 x 坐标），y（目标中心的 y 坐标），w（目标的宽），h（目标的高）。各参数的计算公式为

$$\begin{cases} x=\dfrac{x_{\max}+x_{\min}}{2w_i} \\ y=\dfrac{y_{\max}+y_{\min}}{2h_i} \\ w=\dfrac{x_{\max}-x_{\min}}{w_i} \\ h=\dfrac{y_{\max}-y_{\min}}{h_i} \end{cases} \tag{3}$$

式中　$x_{\max}$、$y_{\max}$——边框右下角坐标值

$x_{\min}$、$y_{\min}$——边框左上角坐标值

w_i、h_i——图像宽和高

深度学习需要大量的数据进行训练，2 600 幅图像数据量较少。训练前对数据进行数据增强来增加数据，其中旋转角度的变化范围从 45°到 315°，并对其进行镜像，最终生成了 11 200 幅图像以供训练使用。

2.3　算法测试与评估

最终算法一共训练 700 个迭代周期（Epoch），每个迭代周期进行 467 次训练（Iteration），训练共使用了7 840 000 幅图像（在 11 200 幅图像中随机抽取并重复使用），其训练的损失值（Loss）变化图如图 6 所示，可以看出算法在前 50 个迭代周期中迅速拟合，Loss 值快速变小，在 200 个迭代周期后逐渐稳定，只有稍许振荡。

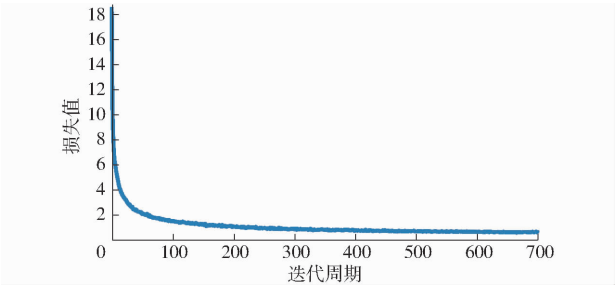

图 6　网络训练损失曲线

Fig. 6　Network training loss curve

在训练中，平均每隔 10 个迭代周期输出一次权值（Weights），算法性能并不是迭代的次数越多越好，过多的训练可能导致过拟合，所以需要对训练出来的权值进行测试和评估。

本文采用客观评价标准来评估苹果识别与分级系统，使用精确率（Precision）、召回率（Recall）和平均精度均值（Mean average precision，mAP）来评估算法。并且为算法找出合适的阈值（Threshold），通过算法预测的置信度来选择合适的目标。

首先找到 mAP 最高的权值文件，以筛选出整体性能足够高的权值。然后在单独找出的权值中不断地调整阈值，衡量精确率、召回率和平均精度，使之能按照自身需求去检测当前环境下的苹果。

如果阈值选择不合适下,会出现如图 7 所示结果,当阈值过低时会出现误识别与重复识别,当阈值过高时会漏识别。

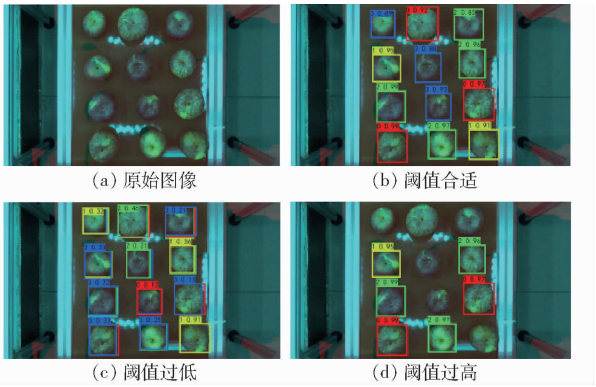

图 7 阈值对检测的影响

Fig. 7 Impact of threshold on detection

3 实验结果与分析

算法训练好后需要对其性能进行评估以找到最优权值,用 mAP 评估算法的整体性能,然后在 mAP 最高的权值中通过对比不同阈值下精确率、召回率和 mAP 的变化选择最优阈值。之后对不同输入层下的苹果识别与分级效果进行实验,并与其他算法进行比较。

3.1 最优权值寻找

在本文的实验中,对 4 种不同的输入层组合分别进行训练,每种输入层训练迭代 700 个迭代周期,每 10 个迭代周期输出一次权值,所以每种输入层得到 70 个权值,需要在这 70 个权值中找出一个 mAP 最高的输入层组合。计算出来的实验结果如图 8 所示。

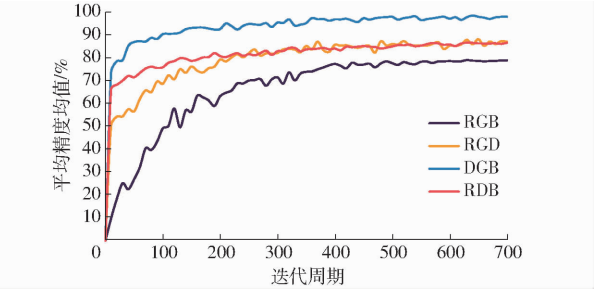

图 8 4 种输入层的 mAP 曲线

Fig. 8 The mAP curves of four input layers

从图 8 中可以看出,当迭代周期数分别达到 400、500、400、300 时输入层分别为 RGB、RGD、DGB、RDB 的算法 mAP 已经稳定,不再发生变化,其中输入层组合为 DGB 时 mAP 总体大于其它输入层组合,mAP 最大为 98.37%,此为本文选用的权值。

3.2 类别阈值选择

算法预测出目标的置信度后需要使用预设的阈值对其进行筛选。通过调整置信度阈值,比较衡量算法在不同阈值下对测试集中 4 个等级苹果识别与分级的精确率、召回率、F1 值和 mAP 的变化,结合自动化分级设备对苹果进行分级的实际需求,确定算法的最佳阈值。经实验测试,不同置信度阈值下算法的精确率、召回率、F1 值、mAP 曲线如图 9 所示。

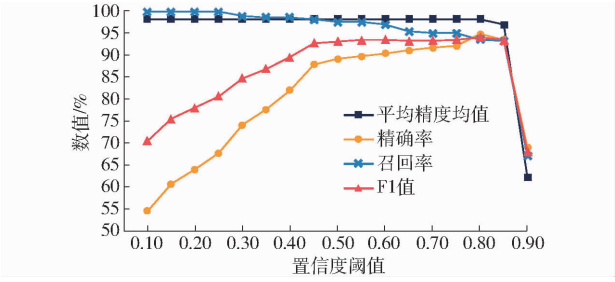

图 9 不同阈值下参数的变化

Fig. 9 Changes in various parameters at different thresholds

对于面向自动化分级设备的苹果等级,需要对传送带识别区内的苹果进行实时识别,同时对识别到的苹果进行分级,因而在算法识别的精确率和召回率之间需要优先考虑精确率。另一方面,在选择阈值时需要辅助参考用于评估算法综合性能的指标 F1 值和 mAP,其能同时兼顾精确率和召回率。

由图 9 可知,当置信度小于 0.8 时,识别精确率较低,不足 90%;当置信度阈值高于 0.8 时,识别算法的 mAP 降低至 90% 以下;因而综合考虑算法的识别精确率与 mAP,当置信度为 0.8 时,算法的性能表现最好,此时算法的精确率、召回率、F1 值、mAP 分别为 94.89%、93.68%、94.28% 和 98.37%。

3.3 实际效果检验

3.3.1 苹果等级识别结果与分析

为了验证所设计的苹果等级识别算法的性能,对该算法在测试集图像上的识别结果进行进一步分析。测试集 262 幅图像中共有 1 348 个苹果目标,其中特等果、一等果、二等果、等外果的数量分别为 243、503、284、318 个。

算法的具体识别结果如表 2 所示,可以看出,针对特等、一等、二等、等外果,本文所提出的算法对其识别的精确率及 F1 值分别为 95.00%、99.69%、99.51%、99.28% 及 92.00%、97.00%、92.00%、96.00%,总体的识别精确率、召回率、mAP(4 个等级苹果的平均精度)及 F1 值均在 90% 以上,分别为 94.89%、93.68%、98.37%、94.25%,满足苹果识别分级的精度要求。

3.3.2 不同目标检测算法识别结果对比

为了进一步分析所提出的苹果定位与分级算法的性能,将改进型 SSD 算法与原 SSD、Faster R-CNN、YOLO v5 算法在测试集图像上进行了识别

表 2　基于改进型 SSD 的苹果定位分级结果

Tab. 2　Apple positioning and grading results based on improved SSD

苹果等级	数量/个	精确率/%	召回率/%	平均精度均值/%	F1 值/%
特等	243	92.02	91.13	95.00	92.00
一等	503	96.55	98.25	99.69	97.00
二等	284	92.89	90.70	99.51	92.00
等外	318	98.15	94.64	99.28	96.00
均值		94.89	93.68	98.37	94.25

结果对比。以 mAP、平均识别耗时等作为评价指标，各算法的识别结果、算法体积（存储空间占用量）及参数数量如表 3 所示。

表 3　不同目标检测网络识别性能对比

Tab. 3　Performance comparison of various target detection networks

检测算法	mAP/%	单幅图像平均识别时间/ms	参数数量	算法体积/MB
改进型 SSD	98.37	5.60	6 683 952	27.51
SSD	98.68	11.14	24 013 232	96.72
Faster R－CNN	96.61	66.67	136 708 989	113.93
YOLO v5	87.64	11.83	46 709 073	28.84

由表 3 可以看出，本文提出的改进型 SSD 算法的 mAP 仅比原 SSD 算法降低 0.31 个百分点，与 Faster R－CNN、YOLO v5[25-28] 算法相比分别高 1.76、10.73 个百分点。针对算法的识别耗时，本文提出的改进型 SSD 算法在测试集上单幅图像平均识别时间可达 5.60 ms（177.5 f/s），能够满足实时高速识别的要求，比原 SSD 算法检测耗时减少 49.73%，与 Faster R－CNN、YOLO v5 算法相比分别减少 91.60%、52.66%。另一方面还可以看出，本文所提出算法的存储空间占用量为 27.51 MB，为原 SSD 算法的 43.64%，说明所提出的网络在保证精度的同时有效地实现了网络的轻量化。

3.3.3　不同果实数目下各等级苹果的分级精确率对比实验

因为在实际分级过程中苹果数量在不断变化，为验证改进前后在不同果实数目条件下不同等级苹果的分级情况，因此设置改进前后网络对不同果实数目条件下各等级苹果的分级精确率进行对比实验，分别对 4 个等级的苹果设置单个苹果、多个苹果和密集苹果，对比改进前后的识别精确率。

实验中测试集共有 300 幅图像，其中根据图像所包含苹果数量为单个、2～10 个以及 10 个以上 3 种，每种数量的图像均包含 4 种等级的苹果图像各 25 幅，每幅图像中只包含一个等级的苹果，将不同数目的苹果分别使用改进前后的 SSD 算法进行识别，对比不同果实数目下各等级苹果的分级精确率，其最终结果如表 4 所示。

表 4　不同果实数目条件下各等级苹果分级精确率

Tab. 4　Classification accuracy of apples of different grades under different fruit numbers

苹果数量/个	苹果等级	检测算法	精确率/%
1	特等	改进型 SSD	97.62
		SSD	98.85
	一等	改进型 SSD	98.77
		SSD	98.88
	二等	改进型 SSD	97.91
		SSD	97.96
	等外	改进型 SSD	98.26
		SSD	98.69
[2,10]	特等	改进型 SSD	95.25
		SSD	95.98
	一等	改进型 SSD	96.73
		SSD	98.34
	二等	改进型 SSD	96.38
		SSD	97.66
	等外	改进型 SSD	98.02
		SSD	98.67
>10	特等	改进型 SSD	95.56
		SSD	97.69
	一等	改进型 SSD	98.19
		SSD	98.88
	二等	改进型 SSD	97.98
		SSD	98.13
	等外	改进型 SSD	98.14
		SSD	98.61

从表 4 可以看出，不同数目果实条件下，各等级苹果的分级精确率均高于 95%，且改进后的算法与原 SSD 算法相比，精确率仅有略微降低，最大相差仅 2.13 个百分点，完全满足苹果分级要求。

3.3.4　随机分布条件下不同果实数目的对比实验

在自动化分级设备运行过程中，随着苹果的随机上料，进入视觉采集区域的苹果数量与等级会随机变化，数目较多时识别难度将会增大。因此设置不同数目下随机分布苹果检测的对比实验，分别分单个苹果、多个苹果和密集苹果，对比 4 种算法在不同数目苹果下随机分布苹果的检测性能，检测效果如图 10 所示。

实验测试集共 320 幅图像，包含 2 030 个苹果，按照果实数量分为 3 类，其中单个苹果 80 幅图像包含 80 个苹果，多个苹果 186 幅图像，包含 1 302 个苹果，密集苹果 54 幅图像包含 648 个苹果。每类随机抽取 30 幅图像作为实验测试集，用 4 种不同的算法进行检测，计算出 mAP。重复以上步骤 3 次，取平

均值,最后将 3 类结果取平均得到综合效果,其最终结果如表 5 所示。

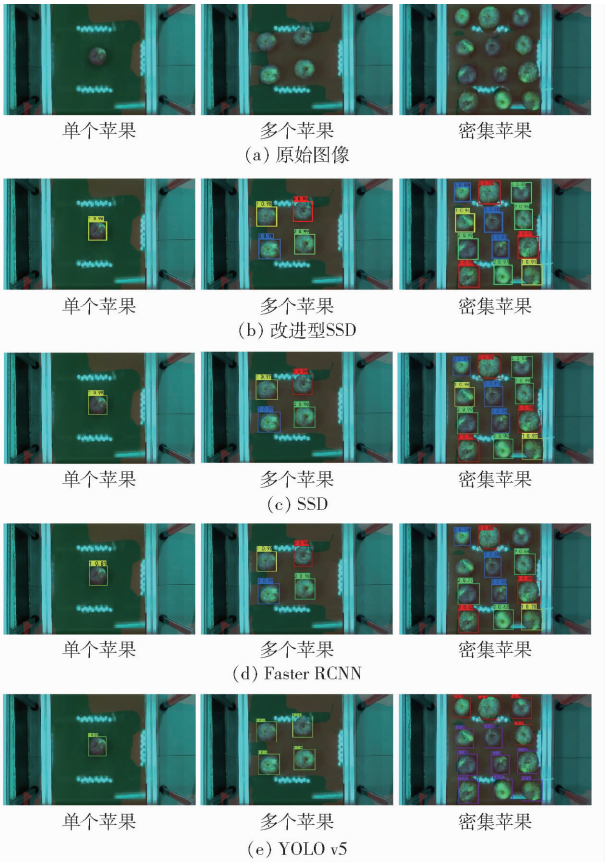

图 10 4 种算法对不同数量苹果的检测效果

Fig. 10 Detection effect of four algorithms on different numbers of apples

表 5 4 种算法对不同数量苹果图像的实验结果

Tab. 5 Experimental results of four algorithms for images with different numbers of apples

苹果数量/个	检测算法	mAP/%			
		重复 1	重复 2	重复 3	平均值
1	改进型 SSD	97.81	98.86	99.79	98.82
	SSD	99.21	99.15	99.56	99.31
	Faster R-CNN	98.56	97.89	97.94	98.13
	YOLO v5	92.82	91.95	92.68	92.48
[2,10]	改进型 SSD	98.37	98.51	98.12	98.33
	SSD	98.16	98.42	98.35	98.31
	Faster R-CNN	97.82	97.21	96.11	97.05
	YOLO v5	88.53	88.05	87.51	88.03
>10	改进型 SSD	90.83	91.55	90.53	90.97
	SSD	91.23	93.12	91.08	91.81
	Faster R-CNN	89.53	88.48	90.92	89.64
	YOLO v5	75.41	79.53	75.29	76.74
平均值	改进型 SSD	95.67	96.31	96.15	96.04
	SSD	96.23	96.90	96.33	96.49
	Faster R-CNN	95.30	94.53	94.99	94.94
	YOLO v5	85.59	86.51	85.16	85.75

从表 5 可以看出,改进型 SSD 算法 mAP 仅比原 SSD 算法降低 0.09 个百分点,且均高于其余 2 个算法,苹果数量越少,mAP 越高,数量由 1 增加到 10 以下时 mAP 下降不多。但是遇到密集苹果时,mAP 降低近 10 个百分点,因为密集苹果的图像中苹果尺寸不一,导致有些苹果没有被识别,同时因为卷积神经网络本身固有的特性,不断地卷积会在深层网络中丢失高分辨率时小目标的特征,而 Faster R-CNN 是基于区域推荐的算法,在 RPN 网络输出矩形候选区域时就已丢失了小目标区域,进行分类并输出预测结果时自然不会出现小苹果。但是改进型 SSD 在浅层网络中提取出特征用于定位与分类,并输出更小的栅格预测结果,降低了深层卷积对小目标的影响,因此在密集苹果上 mAP 高出 Faster R-CNN 1.33 个百分点。而 YOLO v5 虽定位准确度很高,但在遇到多个苹果和密集苹果时分类准确率很低,会错误识别苹果的等级。从综合结果来看,改进型 SSD 算法能够胜任不同数目苹果的检测。

3.4 实际检测时间测试

为了验证改进型 SSD 在检测速度上的优势,对 4 种算法分别在使用 GPU 与 CPU 的情况下的检测速度进行测试,统计结果如表 6 所示。

表 6 4 种算法对使用不同硬件的检测速度

Tab. 6 Four algorithms for detection speed by using different hardwares

硬件	检测算法	帧率/(f·s^{-1})
GPU	改进型 SSD	175.17
	SSD	106.03
	Faster R-CNN	15.69
	YOLO v5	95.03
CPU	改进型 SSD	62.64
	SSD	13.61
	Faster R-CNN	1.91
	YOLO v5	18.57

由于 4 种算法在做卷积运算前需要对图像进行缩放到固定尺寸,所以不同分辨率图像在本文算法下检测时间的差异只与缩放过程所用的时间有关,其余流程并没有区别。本文实验测试用的图像尺寸均为 1 920 像素 ×1 080 像素,考虑到苹果识别与分级系统是搭载在苹果分级设备的工作站中,有的工作站只有 CPU,为了减少硬件成本,需要在单 CPU 的情况下运行,所以分别在 GPU 和 CPU 下进行了测试,检测的精度没有变化,CPU 检测时长较 GPU 长。

在 GPU 下算法检测 200 幅图像用时 1.18 s,在 CPU 下检测 200 幅图像耗时 3.19 s,即为 175.17 f/s 和 62.64 f/s,每幅图像平均检测时间为 5.71 ms 和 15.96 ms,与原 SSD 算法相比检测用时分别减少

39.45% 和 78.28%。该检测速度能够满足自动化分级设备的实时且高速分拣需求。

4　结论

(1)针对苹果果径与果形自动化分级，提出了基于改进型 SSD 算法的苹果识别与分级方法。实验表明该算法的检测精度高、速度快，选用的网络在真实环境下进行检测，在 GPU 和 CPU 下每幅图像的平均检测时间分别达到 5.71 ms 和 15.96 ms，即检测视频的帧率分别达到 175.17 f/s 和 62.64 f/s，召回率、精确率、mAP 和 F1 值分别达到 93.68%、94.89%、98.37% 和 94.25%。

(2)针对自动化分级设备上高速识别环境，提出用改进型 SSD 网络定位并分级苹果，与原 SSD 算法相比 mAP 仅下降 0.31 个百分点，检测用时减少 49.73%，算法体积压缩至原 SSD 算法的 43.64%。

(3)对比了在不同输入层组合下的检测结果，当使用 DGB 图像组合作为输入层时，其 mAP 总体高于其他 3 种输入层组合，且仅使用带有颜色信息的图像作为输入层时，其 mAP 明显低于引入深度信息后的输入层组合。

(4)对比了改进型 SSD 网络与原 SSD 网络在不同果实数量下，对各等级苹果的分级准确率，其分级准确率与原 SSD 网络相差不大，准确率均在 95% 以上，分级准确率最多仅比原 SSD 算法下降 2.13 个百分点，满足苹果分级要求。

(5)对比了原 SSD、Faster R－CNN、YOLO v5 和改进型 SSD 在不同苹果数目下随机分布苹果的检测结果，其 mAP 总体比 Faster R－CNN 高 1.10 个百分点，比 YOLO v5 高 10.29 个百分点，并且在密集苹果的图像检测中，改进型 SSD 的 mAP 比 Faster R－CNN 高 1.33 个百分点，比 YOLO v5 高 14.23 个百分点，本文算法检测精度优势更为明显。

(6)由于深度学习需要大量的计算能力，在实际使用中还需要考虑功耗和稳定的问题，因此对比了原 SSD、Faster R－CNN、YOLO v5 算法和改进型 SSD 算法分别在使用 GPU 与 CPU 的硬件条件下的检测速度。与原 SSD 算法相比，本文所提出的改进算法检测用时在硬件为 GPU 和 CPU 时分别减少 39.45% 和 78.28%。与 Faster R－CNN、YOLO v5 算法相比，无论是 GPU 还是 CPU 下检测用时均大幅减少，且检测精度也高于这 2 种算法。

参　考　文　献

[1] 黄辰，费继友. 基于图像特征融合的苹果在线分级方法[J]. 农业工程学报，2017，33(1)：285－291.
HUANG Chen, FEI Jiyou. Online apple grading based on decision fusion of image features[J]. Transactions of the CSAE, 2017, 33(1): 285－291. (in Chinese)

[2] 李先锋，朱伟兴，花小朋，等. 基于 D－S 证据理论的决策级多特征融合苹果分级方法[J]. 农业机械学报，2011，42(6)：188－192.
LI Xianfeng, ZHU Weixing, HUA Xiaopeng, et al. Multi-feature decision fusion method based on D－S evidence theory for apple grading[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(6): 188－192. (in Chinese)

[3] 夏卿，李先锋. 基于改进 PSO 算法和 LS－SVM 的苹果分级检测[J]. 计算机与现代化，2012，1(5)：161－163，166.
XIA Qing, LI Xianfeng. Detection of apple grading based on improved PSO algorithm and LS－SVM[J]. Computer and Modernization, 2012, 1(5): 161－163, 166. (in Chinese)

[4] GUANJRYUI H, ENYU Z, JIANGUO Z, et al. Infield apple detection and grading based on multi-feature fusion[J]. Horticulturae, 2021, 7(9): 276.

[5] LI J, ZHU X, JIA R, et al. Apple-yolo: a novel mobile terminal detector based on YOLO v5 for early apple leaf diseases[C]//2022 IEEE 46th Annual Computers, Software, and Applications Conference, 2022.

[6] HOSSAIN M S, AL-HAMMADI M H, MUNAMMAD G. Automatic fruit classification using deep learning for industrial applications[J]. IEEE Trans. Industrial Informatics, 2019, 15(2): 1027－1034.

[7] SHI X, CHAI X, YAN C, et al. Vision-based apple quality grading with multi-view spatial network[J]. Computers and Electronics in Agriculture, 2022, 195: 168－169.

[8] 王立扬，张瑜，沈群，等. 基于改进型 LeNet－5 的苹果自动分级方法[J]. 中国农机化学报，2020，41(7)：105－110.
WANG Liyang, ZHANG Yu, SHEN Qun, et al. Automatic detecting and grading method of apples based on improved LeNet－5[J]. Journal of Chinese Agricultural Mechanization, 2020, 41(7): 105－110. (in Chinese)

[9] 何进荣，石延新，刘斌，等. 基于 DXNet 模型的富士苹果外部品质分级方法研究[J]. 农业机械学报，2021，52(7)：379－385.
HE Jinrong, SHI Yanxin, LIU Bin, et al. External quality grading method of Fuji apple based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(7): 379－385. (in Chinese)

[10] 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, 2014.

[11] GIRSHICK R. Fast R－CNN[C]//2015 IEEE International Conference on Computer Vision, 2015.

[12] REN S, HE K, GIRSHICK R, et al. Faster R－CNN: towards real-time object detection with region proposal networks[J].

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137 - 1149.
[13] 闫建伟,赵源,张乐伟,等. 改进 Faster - RCNN 自然环境下识别刺梨果实[J]. 农业工程学报,2019,35(18): 143 - 150.
YAN Jianwei, ZHAO Yuan, ZHANG Lewei, et al. Recognition of *Rosa roxbunghii* in natural environment based on improved Faster - RCNN[J]. Transactions of the CSAE, 2019, 35(18): 143 - 150. (in Chinese)
[14] REDMON J, DIVVALA S K, GIRSHICK R B, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[15] MARCO S, SILVIA C, ALESSIA C, et al. Automatic bunch detection in white grape varieties using Yolov3, Yolov4, and Yolov5 deep learning algorithms[J]. Agronomy, 2022, 12(2): 319.
[16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// Computer Vision-ECCV 2016, 2016.
[17] XIAO Dong, LI Hongzong, LIU Chenyi, et al. Large-truck safety warning system based on lightweight SSD model[J]. Computational Intelligence and Neuroscience, 2019, 2019(Supp. 1): 1 - 10.
[18] 郑少武,李巍华,胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报,2019,40(12):143 - 151.
ZHENG Shaowu, LI Weihua, HU Jianyao. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information[J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 143 - 151. (in Chinese)
[19] 高升,王巧华. 基于高光谱图像信息融合的红提糖度无损检测[J]. 发光学报,2019,40(12):1574 - 1584.
GAO Sheng, WANG Qiaohua. Comprehensive detection of internal quality of red globe grape extract based on near infrared spectroscopy[J]. Chinese Journal of Luminescence, 2019, 40(12): 1574 - 1584. (in Chinese)
[20] 吴天舒,张志佳,刘云鹏,等. 基于改进 SSD 的轻量化小目标检测算法[J]. 红外与激光工程,2018,47(7):47 - 53.
WU Tianshu, ZHANG Zhijia, LIU Yunpeng, et al. A lightweight small object detection algorithm based on improved SSD[J]. Infrared and Laser Engineering, 2018, 47(7): 47 - 53. (in Chinese)
[21] 王昱潭,薛君蕊. 改进 SSD 的灵武长枣图像轻量化目标检测方法[J]. 农业工程学报,2021,37(19):173 - 182.
WANG Yutan, XUE Junrui. Lightweight object detection method for Lingwu long jujube images based on improved SSD[J]. Transactions of the CSAE, 2021, 37(19): 173 - 182. (in Chinese)
[22] LEE E, MIN J. Explore influential color channels for quality classification of apples based on CNN[J]. Digital Contents Society, 2022, 23(8): 1477 - 1484.
[23] SHETTY S, S AV, MAHALE A. Ms-chexnet: an explainable and lightweight multi-scale dilated network with depthwise separable convolution for prediction of pulmonary abnormalities in chest radiographs[J]. Mathematics, 2022, 10(19): 3646.
[24] ZHANG L, WANG B, YUAN X, et al. Remaining useful life prediction via improved CNN, GRU and residual attention mechanism with soft thresholding[J]. IEEE Sensors Journal, 2022, 22(15): 15178 - 15190.
[25] 张志远,罗铭毅,郭树欣,等. 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(增刊 1):232 - 240.
ZHANG Zhiyuan, LUO Mingyi, GUO Shuxin, et al. Cherry fruit detection method in natural scene based on improved YOLOv5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(Supp. 1): 232 - 240. (in Chinese)
[26] 闫彬,樊攀,王美茸,等. 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别[J]. 农业机械学报,2022,53(9): 28 - 38,59.
YAN Bin, FAN Pan, WANG Meirong, et al. Real-time apple picking pattern recognition for picking robot based on improved YOLOv5m[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 28 - 38, 59. (in Chinese)
[27] 孙丰刚,王云露,兰鹏,等. 基于改进 YOLOv5s 和迁移学习的苹果果实病害识别方法[J]. 农业工程学报,2022, 38(11):171 - 179.
SUN Fenggang, WANG Yunlu, LAN Peng, et al. Identification of apple fruit diseases using improved YOLOv5s and transfer learning[J]. Transactions of the CSAE, 2022, 38(11): 171 - 179. (in Chinese)
[28] YAN B, FAN P, LEI X, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing, 2021, 13(9): 1619.