

南京农业大学学报
Journal of Nanjing Agricultural University
ISSN 1000-2030, CN 32-1148/S

《南京农业大学学报》网络首发论文

题目: 基于改进 YOLOv5s 的不同成熟度苹果目标检测方法
作者: 王勇, 陶兆胜, 石鑫宇, 伍毅, 吴浩
收稿日期: 2023-05-26
网络首发日期: 2023-09-27
引用格式: 王勇, 陶兆胜, 石鑫宇, 伍毅, 吴浩. 基于改进 YOLOv5s 的不同成熟度苹果目标检测方法[J/OL]. 南京农业大学学报.
<https://link.cnki.net/urlid/32.1148.S.20230926.1201.002>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于改进 YOLOv5s 的不同成熟度苹果目标检测方法

王勇，陶兆胜*，石鑫宇，伍毅，吴浩

(安徽工业大学机械工程学院，安徽 马鞍山 243032)

摘要：[目的]在自然环境下进行机械自动化采摘苹果时，对不同成熟度的果实做到精确检测尤为重要。针对因苹果之间遮挡和同级成熟度苹果的纹理特征分布差异所导致的在自然环境下不同成熟度苹果目标检测精度较低的问题，提出了一种不同成熟度苹果检测模型 SODSTR-YOLOv5s (YOLOv5s with Small Detection Layer and Omni-Dimensional Dynamic Convolution and Swin Transformer Block)。**[方法]**首先改进 YOLOv5s 的多尺度目标检测层，在 Prediction 中构建检测 160×160 特征图的检测头，提高小尺寸的不同成熟度苹果的检测精度；其次在 Backbone 结构中融合 Swin Transformer Block，加强同级成熟度的苹果纹理特征融合，弱化纹理特征分布差异带来的消极影响，提高模型泛化能力；然后将 Neck 结构的 Conv 模块替换为动态卷积模块 ODConv，细化局部特征映射，实现局部苹果细粒度特征的充分提取，最后基于不同成熟度苹果数据集进行试验，验证改进模型的性能。**[结果]**试验结果显示，改进模型的精确率、召回率、平均精度均值分别为 89.1%、95.5%、93.6%，高、中、低成熟度苹果精度均值分别为 94.1%、93.1%、93.7%，平均检测时间为 16 ms，参数量为 7.34 M，相比于 YOLOv5s 模型，精确率、召回率、平均精度均值分别提高了 3.8%、5.0%、2.9%。**[结论]**虽然改进模型的参数量与平均检测时间分别增加了 0.32 M 和 5 ms，但提升了在自然环境下对不同成熟度苹果的检测能力，较好地满足实际采摘苹果的检测要求，为基于深度学习的不同成熟度苹果检测方面的研究提供思路。

关键词：苹果；成熟度；目标检测；YOLOv5s；深度学习；自然环境

中图分类号：TP391.41 **文献标志码：**A

Apple target detection method with different ripeness based on improved YOLOv5s

WANG Yong TAO Zhaosheng* SHI Xinyu WU Yi WU Hao

(College of Mechanical Engineering, Anhui University of Technology, Ma'anshan 243032, China)

Abstract: [Objectives] When picking apples mechanically and automatically in natural environment, it is very important to accurately detect fruits of different ripeness. Aiming at the problem of low detection accuracy of apples with different ripeness under natural environment due to the difference in texture feature distribution between apples and apples with the same ripeness, a detection model of apples with different ripeness SODSTR-YOLOv5s (YOLOv5s with Small Detection Layer and Omni-Dimensional Dynamic Convolution and Swin Transformer Block) was proposed. [Methods] Firstly, the multi-scale target detection layer of YOLOv5s was improved, and the detection head of 160×160 feature map was constructed in Prediction to improve the detection accuracy of small-size apples with different ripeness. Secondly, Swin Transformer Block was integrated into the Backbone structure to strengthen the texture feature fusion of apples at the same level of ripeness, weaken the negative impact of the difference in texture feature distribution, and improve the generalization ability of the model. Then, the Conv module of Neck structure was replaced by the dynamic convolution module ODConv, and the local feature mapping was refined to realize the full extraction of local apple fine-grained features. Finally, experiments were conducted based on different ripeness apple data sets to verify the performance of the improved model. [Results] The test results showed that the accuracy rate, recall rate and average accuracy of the improved model were 89.1%, 95.5% and 93.6%,

收稿日期：2023-05-26

基金项目：安徽省自然科学基金面上项目(2108085ME166)、安徽高校自然科学基金项目重点项目(KJ2021A0408)

作者简介：王勇，硕士研究生。*通信作者：陶兆胜，博士，副教授，主要从事机器视觉、机电一体化研究，E-mail: taozs@ahut.edu.cn

respectively. The average accuracy of high, medium and low ripeness apples was 94.1%, 93.1% and 93.7%, respectively. The average detection time was 16 ms and the number of parameters was 7.34 M. Compared with the YOLOv5s model, the accuracy rate, recall rate and average accuracy are increased by 3.8%, 5.0% and 2.9%, respectively. **[Conclusions]** Although the number of parameters and the average detection time of the improved model increased by 0.32 M and 5 ms respectively, it improved the detection ability of apples with different ripeness under natural environment, better met the detection requirements of actual apples picked, and provided ideas for the research of apples with different ripeness based on deep learning.

Key words: apple; ripeness; target detection; YOLOv5s; deep learning; natural environment

当前苹果采摘方式主要有人工采摘与机械自动化采摘两种方式^[1-2]。针对不同成熟度苹果进行人工采摘时因其采摘时间较长易导致错过最佳采摘期,智能机械自动化采摘在实际应用下采摘不同成熟度苹果时,主要以苹果表皮颜色判定其成熟度,但因无法准确判定不同成熟度苹果易产生误采和漏采等现象。因此如何构建高准确率的不同成熟度苹果目标检测系统成为众多学者研究的关键技术之一。

随着机器视觉的快速发展,各种机器视觉检测算法在不同成熟度水果检测方面得到了广泛的应用。基于机器视觉的不同成熟度水果检测技术从图像中提取特定信息(如水果成熟的生理生化特性)进行分析和分析,判断其与成熟度之间的相关性^[3]。Mazen 等^[4]构建了一种自动识别不同成熟度香蕉的机器视觉系统,并基于人工神经网络(Artificial neural network, ANN)算法,利用颜色、褐斑生长和 Tamura 纹理等特征对不同成熟度的香蕉分类。Yu 等^[5]提出了 RGB-D 相机结合机器学习以判断荔枝成熟度。但由于传统机器学习在自然环境下鲁棒性和实时性较差,故难于满足水果智能机械化采摘的需求^[6]。

随着深度学习的发展,许多学者将其应用于农产品的视觉检测领域特别是不同成熟度水果检测领域^[7]。Wan 等^[8]提出一种将颜色特征值与反向传播神经网络(Back propagation neural network, BPNN)分类技术相结合的算法,对绿色、橙色和红色 3 种成熟度等级的番茄进行分类。龙洁花等^[9]提出将改进的 Mask R-CNN 模型应用于番茄采摘机器人,并在温室内开展不同成熟度番茄识别。Parvathi 等^[10]采用 ResNet-50 网络的 Faster R-CNN 算法实现复杂背景下不同成熟度椰子的实时检测。Altaheri 等^[11]构建了一种分类模型,以区分不同成熟度椰果。Fan 等^[12]针对夜间草莓采摘精度低、易误采的问题,提出了结合暗通道增强的改进 YOLOv5 算法。陈锋军等^[13]针对油橄榄果实之间相互遮挡的问题,提出了引入卷积注意力模块和改进特征融合网络的改进 EfficientDet 算法。张志远等^[14]为解决樱桃果实之间相互遮挡以及图像的小目标樱桃检测困难等问题,提出一种基于改进 YOLO v5 的自然环境下樱桃果实识别方法。耿磊等^[15]提出一种基于融合注意力机制的自动识别和分类模型 EBm-Net,通过提取苹果表面的颜色纹理特征对苹果品种分类。

虽然 YOLO 算法^[16]所具有的一阶段目标检测特性能快速准确地检测不同目标,但自然环境下苹果状态较为复杂,如苹果大小不一、枝叶和果实遮挡、同级成熟度苹果的特征分布存在较大差异,影响检测模型对不同成熟度苹果的检测准确率^[17-18]。为了解决因苹果纹理特征分布差异以及枝叶和果实遮挡所导致的检测精度较低的问题,提高在非结构环境下不同成熟度苹果的检测精度,本文提出了一种基于改进 YOLOv5s 的不同成熟度苹果目标检测模型,意在进一步提高检测模型在自然环境下对不同成熟度苹果的检测效果。

1 材料与方法

1.1 试验图像采集

本文图像采集地点为安徽省宿州市砀山县官庄镇前王庄,以红富士与嘎啦两种品种的苹果为研究对象,于 2022 年 7—8 月和 9—10 月使用苹果 13pro 手机采集自然环境下的晴天、雨后、白天和傍晚的不同树的不同成熟度的苹果图像,共采集 1 545 幅苹果图像作为苹果成熟度检测原始数据集,图像统一调整为 960×640 像素。

1.2 苹果成熟度划分

常规苹果成熟度划分方式根据果皮颜色以及果肉色泽分为完全成熟、可食用成熟、可采摘成熟与未成熟^[19-20]。由于在自然环境下检测苹果的成熟度时,无法观察果肉颜色变化,因此本文根据苹果果皮颜色以及市场需求对成熟度进行划分^[21],分为 3 级成熟度,即高成熟度(果皮基本全红)、中成熟度(果皮红绿相

间)和低成熟度(果皮基本全绿)3种类型进行标注,如图1所示。



图1 苹果成熟度等级

Fig.1 Apple ripeness level

A. 高成熟度 High ripeness; B. 中成熟度 Medium ripeness; C. 低成熟度 Low ripeness.

1.3 试验数据集构建

为了增强模型对苹果个体尺寸差异和光照变化的鲁棒性以及模型的泛化能力,避免过拟合,本文运用旋转,高斯噪声等数据增强方法扩充数据集,扩充后的数据集共5158张,采用LabelImg软件对图像中不同成熟度的苹果进行标注(共标注苹果成熟度标签16842个,其中高成熟度苹果标签5597个,中成熟度苹果标签5604个,低成熟度苹果标签5641个),训练集、验证集和测试集以8:1:1的比例随机划分。数据增强示例见图2。

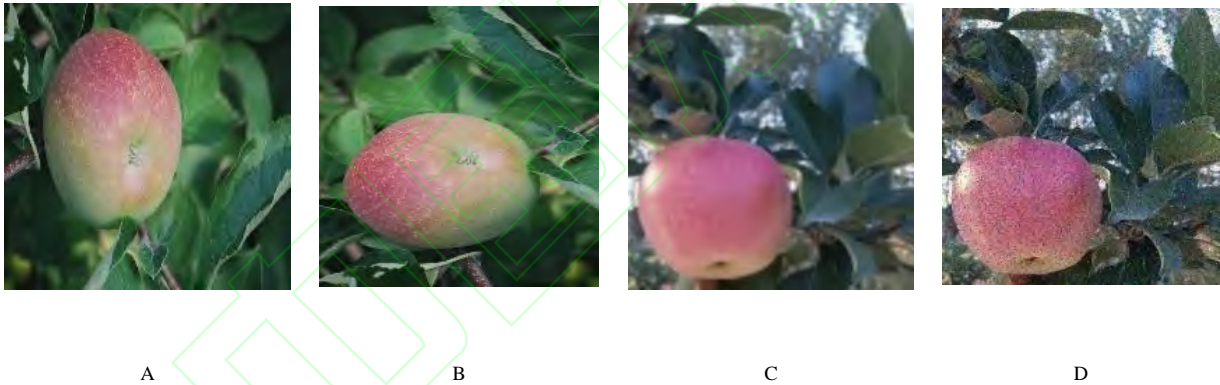


图2 数据增强示例

Fig.2 Example of data enhancement

A. 原始图片 Original image; B. 旋转 90° Rotation 90° ; C. 原始图片 Original image; D. 高斯噪声 Gaussian noise

1.4 YOLOv5s 检测模型

YOLOv5s 运用端到端机制将图像归一化后,输入卷积神经网络以便回归预测目标物体的类别和位置信息。YOLOv5s 主要包括输入端、骨干网络(Backbone)、颈部网络(Neck)和预测层(Prediction)4 部分。输入端建立图像处理策略和锚框生成机制,自适应计算训练集的最佳锚框值; Backbone 运用卷积运算提取不同层次的特征图,采用跨阶段连接方式融合特征,减少参数冗余,提高模型精度; Neck 使用特征金字塔网络结构(Feature Pyramid Networks, FPN^[22])自上而下增强语义特征,构建路径聚合网络结构(Path Aggregation Network, PANet^[23])自底向上增强特征信息; Prediction 生成预测目标的类别概率和位置信息,应用 3 个检测头(Head)分别在 3 个不同尺度特征图上预测图像的大中小目标,网络结构图如图 3 所示。

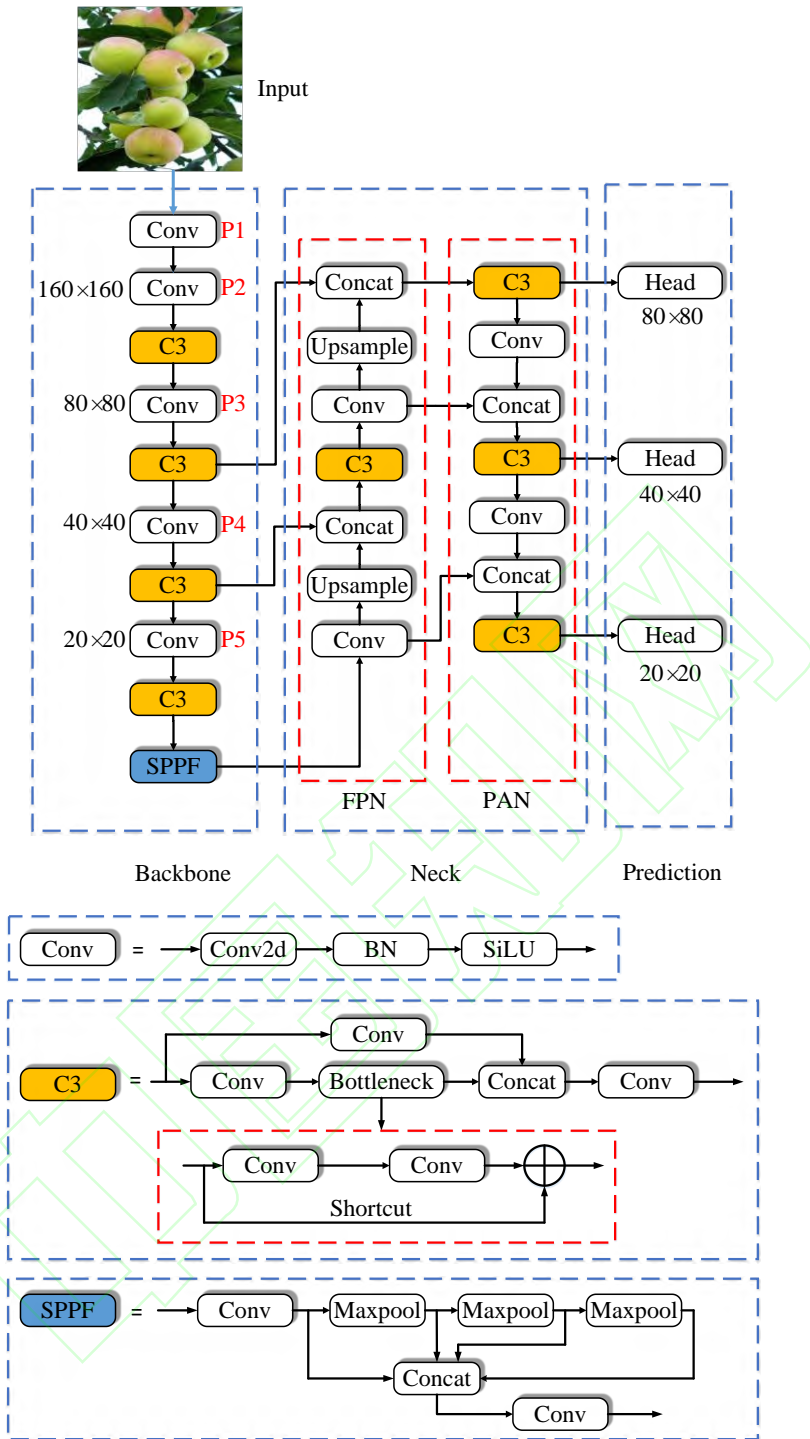


图 3 YOLOv5s 网络结构图

Fig.3 Network structure of YOLOv5s

1.5 改进 YOLOv5s 苹果成熟度检测模型构建

1.5.1 改进多尺度检测层 由于本文构建的自制数据集图片存在部分苹果尺寸较小的情况，原 YOLOv5s 所运用的 80×80 尺寸检测头无法精准检测小尺寸的不同成熟度苹果，但 YOLOv5s 在 P2 层所生成的 160×160 特征图则包含了较多小尺寸苹果的浅层语义特征，故本文运用上采样(Upsample)和拼接(Concat)算法融合 Backbone 结构的 P2 层生成的苹果浅层语义特征，以便后续生成的苹果深层语义特征包含较为丰富的小尺寸苹果轮廓纹理信息，并在 Prediction 中构建 160×160 特征图的检测头，提高小尺寸的不同成熟度苹果的检测精度，如图 4 所示。

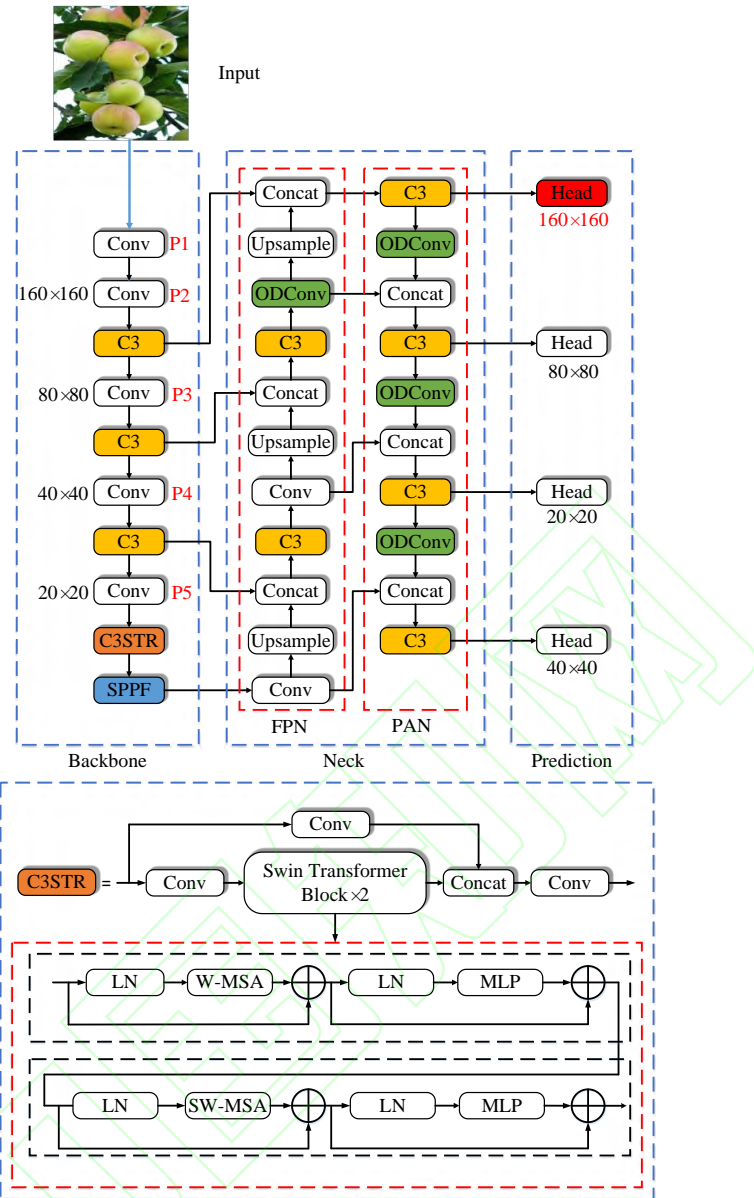


图 4 改进 YOLOv5s 网络结构图

Fig.4 Network structure of improved YOLOv5s

1.5.2 骨干网络融合 Swin Transformer Block 为弱化同级成熟度苹果纹理特征差异，提高模型的泛化能力，进而提高同级成熟度苹果的检测精度，引入 Swin Transformer^[24]模型的 Swin Transformer Block。该模块首先对输入特征图进行独立窗口划分，并在窗口内独立计算以减少模型计算复杂度，然后利用滑动划分窗口实现跨窗口的特征信息交互，实现同级成熟度苹果的纹理特征交互融合。

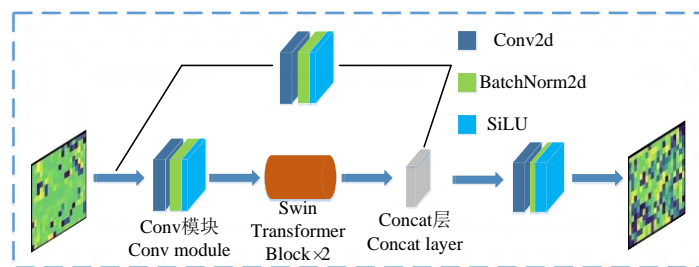


图 5 C3STR 模块结构图

Fig.5 Structure of C3STR module

虽然 C3 模块能较好地融合残差特征，但其 Bottleneck 模块无法实现对同级成熟度苹果的纹理特征交

互，故本文使用 Swin Transformer Block 替换 C3 模块的 Bottleneck 模块构建 C3STR 模块，在保留 C3 模块融合残差特征的基础上，增强了苹果的纹理特征交互能力，如图 5 所示。在改进模型中应用 C3STR 模块（如图 4 所示），利用其移位窗口和分层结构提高输入特征图的分辨率，逐层扩大苹果纹理特征的感受野，实现像素邻域间的苹果纹理特征交互。

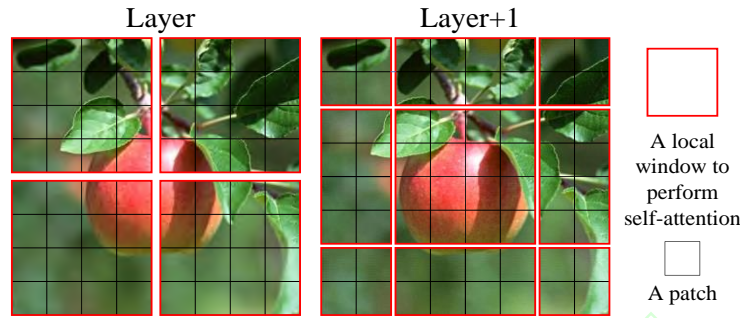


图 6 Swin Transformer Block 窗口划分流程图

Fig.6 S Flow chart for dividing the Swin Transformer Block window

图 6 为 Swin Transformer Block 窗口划分流程图。图 6(Layer)所示的 Swin Transformer Block 采用基于窗口的多头自注意力机制(Window-Multi-head Self Attention, W-MSA)的常规窗口划分策略，将 8×8 patch 的输入特征图均匀划分为 2×2 个 4×4 patch 的窗口，并实现在每个窗口内进行独立自注意力计算，但该计算缺少跨窗口的连接，即窗口之间未建立联系从而无法实现 Swin Transformer Block 对同级成熟度苹果的纹理特征交互的能力。故本文在图 6(Layer+1)所示的 Swin Transformer Block 中应用基于滑动窗口多头自注意力机制(Shifted Window-Multi-head Self Attention, SW-MSA)的移位划分窗口机制，将常规划分的窗口移位 $\left[\frac{M}{2}, \frac{M}{2}\right] (M=4)^{[24]}$ 像素，实现不同窗口间的同级成熟度苹果的纹理特征交互。

1.5.3 引入 ODConv 动态卷积模块 由于自然环境的苹果间、苹果与树叶间的遮挡，从而导致原 YOLOv5s 无法准确定位不同成熟度苹果的特征区域，进而产生苹果成熟度特征信息丢失的情况，为此本文在 YOLOv5s 的 Neck 结构中引入动态卷积模块 ODConv^[25]，构建并行策略的四维注意力机制，精确定位不同成熟度苹果特征所在的目标区域，细化苹果的局部细粒度特征，进而提取不同成熟度苹果的显著表型特征，实现准确检测不同成熟度苹果。

动态卷积模块 ODConv 运用一维卷积、空间、输入通道、输出通道等维度的动态性，应用并行策略制定多维注意力机制，针对核空间的四个维度（空间核尺寸、输入通道数、输出通道数、卷积核数量）学习卷积核的注意力，使本文所提出的改进模型学习经 SPPF(Spatial Pyramid Pooling-Fast)模块融合后的苹果细粒度特征。

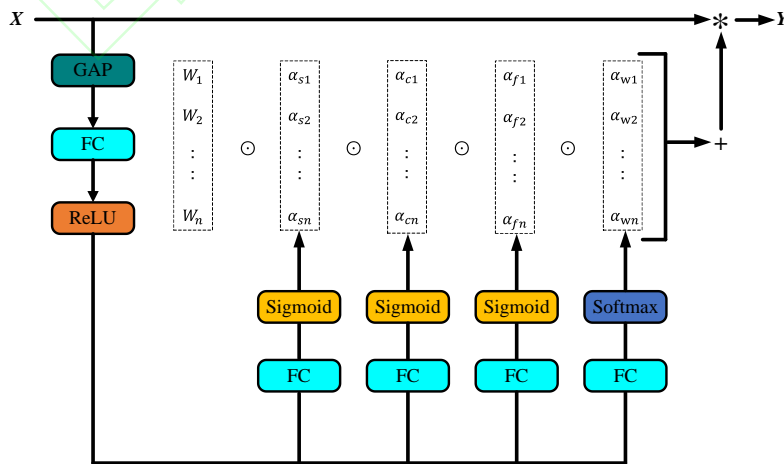


图 7 ODConv 卷积运算流程图

Fig.7 Flow chart of ODConv convolution operation

由图 7 可知, ODConv 模块首先应用全局平均池化层 GAP(global average pooling)将输入特征 X 压缩成长度为 c_{in} 的特征向量, 其次采用一个全连接层 FC 层(fully connected layers)将特征向量映射至缩减比 r 为 1/16 的低维空间, 降低动态卷积模块的复杂度, 接着运用 ReLU 函数去除特征向量的负值, 最后建立 4 个头部分支, 构建 ODConv 卷积模块的 4 类注意力标量。第一个分支经 FC 层预测每个卷积核不同的空间位置权重, 输出 $k \times k$ 的特征向量, 以便模块在卷积运算时精确定位苹果所在位置区域; 第二个分支经 FC 层预测每个卷积核不同的输入通道权重, 输出 $c_{in} \times 1$ 的特征向量, 从而使模块在卷积运算时有效分析通道内不同成熟度苹果的结构特征, 实现精确提取不同成熟度苹果的颜色特征; 第三个分支经 FC 层预测不同输出通道的权重, 输出 $c_{out} \times 1$ 的特征向量, 以避免模块在卷积运算时出现通道间苹果成熟度的颜色特征丢失情况; 第四个分支经 FC 层预测不同卷积核的权重, 输出 $n \times 1$ 的特征向量, 以便模块在卷积运算时细化不同成熟度苹果局部特征(如颜色特征)。第一、二、三个分支经 Sigmoid 激活函数分别将 $k \times k$ 、 $c_{in} \times 1$ 、 $c_{out} \times 1$ 的特征向量的变量值映射到区间(0,1)之内, 生成归一化注意力标量 α_{si} 、 α_{ci} 、 α_{fi} ; 第四个分支经 Softmax 激活函数将 $n \times 1$ 的特征向量的变量值映射到区间(0,1)之内, 并使变量值总和为 1, 实现按不同成熟度苹果的颜色特征分类, 生成归一化注意力标量 α_{wi} 。

本文模型在基于输入特征对卷积核进行调制的基础上进行卷积计算, 保证模型充分提取不同成熟度苹果的特征, 提高检测精度。

$$Y = \left(\sum_{i=1}^n \alpha_{oi} \odot \alpha_{ci} \odot \alpha_{fi} \odot \alpha_{si} \odot W_i \right) * X \quad (1)$$

式中, $\alpha_{wi} \in \mathbb{R}$ 表示整个卷积核 W_i 的注意力标量, $\alpha_{si} \in \mathbb{R}^{k \times k}$ 表示沿着卷积核 W_i 的空间维度引入的注意力标量, $\alpha_{ci} \in \mathbb{R}^{c_{in}}$ 表示沿着卷积核 W_i 的输入通道维度引入的注意力标量, $\alpha_{fi} \in \mathbb{R}^{c_{out}}$ 表示沿着卷积核 W_i 的输出通道维度引入的注意力标量, \odot 表示沿内核空间不同维度的乘法操作, $*$ 表示卷积运算, X 表示输入特征, Y 表示输出特征。

1.6 试验环境配与网络参数设置

本文所做的试验在 windows10 系统下进行模型训练与测试, 计算机使用的 CPU 为 Intel Core i7-11700K, GPU 为 NVIDIA RTX3070, 显存 8GB, 运行内存 32GB, 为提高网络训练速度, 使用 GPU 进行加速, cuda 版本为 11.3, cudnn 版本为 8.2.1, 软件环境为 Python3.8, 所有对比试验均在相同环境下运行。

试验迭代 epoch 数为 200, 批量处理大小 batchsize 为 16, 初始学习率为 0.001, 梯度优化器选择 Adam^[26], 其动量设置为 0.937, 使用余弦退火学习率衰减算法自动调整学习率。

1.7 试验评价指标

本文使用精确率(Precision, P)、召回率(Recall, R)、平均精度(Average Precision, AP)、平均精度均值(mean Average Precision, mAP)、平均检测时间(Mean Detection Time, MDT)和参数量(Parameters)作为评价指标以测试模型性能。参数量(Parameters)衡量模型对内存资源的消耗量, 单位为兆(M), 平均检测时间(Mean Detection Time, MDT)为模型检测单张测试集图像所用时间的均值, 单位为毫秒(ms)。在自然环境下进行机器人采摘时, 为了防止损伤苹果, 采摘机器人的采摘动作一般较慢, 故检测速度无较高的要求, 因此本文将评价指标的平均精度均值 mAP 作为第一评价指标, 对所改进的检测模型进行评价, 评价指标计算公式分别为:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$AP = \int_0^1 PRdR \quad (4)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (5)$$

式中， TP 表示模型预测为正样本且是正样的数量， FP 表示模型预测为正样本但是负样本的数量， FN 表示模型预测为负样本但是正样本的数量， n 表示数据的类别数量， AP_i 表示第 i 类检测目标平均精度， mAP 表示其所有类别 AP_i 均值。

2 结果与分析

2.1 C3STR模块

为了验证 C3STR 对模型的影响，相较于 YOLOv5s，本文从平均精度均值 mAP 、参数量、平均检测时间 MDT 三方面进行验证，如表 1 所示。

表 1 C3STR 模块替换位置对比结果

Tab.1 Experimental results of C3STR module replacement position

| 模型 Model | mAP /% | 参数量 Parameters/M | 平均检测时间/ms Mean Detection Time |
|-------------|--------|---------------------|----------------------------------|
| YOLOv5s | 90.7 | 7.02 | 11 |
| C3STR1 | 92.3 | 7.04 | 13 |
| C3STR2 | 92.6 | 7.07 | 14 |
| C3STR3 | 92.6 | 7.38 | 15 |
| C3STR4 | 92.8 | 7.15 | 14 |
| C3STR5 | 92.1 | 7.53 | 18 |
| C3STR6 | 91.9 | 7.57 | 19 |

在表1中，C3STR1表示在YOLOv5s模型的Backbone结构的第2层应用此模块，C3STR2表示在第4层应用此模块，C3STR3表示在第6层应用此模块，C3STR4表示在第8层应用此模块，C3STR5表示在第6、8层应用此模块，C3STR6表示在第4、6、8层应用此模块。由表1可知，在检测模型的第8层应用C3STR模块，能够最小程度地增加模型参数量和平均检测时间，且较为明显地提升模型的检测精度。

2.2 引用ODConv动态卷积模块

针对不同位置的ODConv动态卷积模块对改进模型的影响大小，相对于YOLOv5s，本文从平均精度均值 mAP 、参数量、平均检测时间MDT进行对比，结果如表2所示。

表 2 ODConv 卷积模块替换位置试验结果

Tab.2 Experimental results of ODConv module replacement position

| 模型 Model | mAP/% | 参数量/M Parameters | 平均检测时间/ms Mean Detection Time |
|-----------------|-------|---------------------|----------------------------------|
| YOLOv5s | 90.7 | 7.02 | 11 |
| S-YOLOv5s | 91.5 | 7.17 | 12 |
| ODConv_backbone | 92.5 | 7.20 | 18 |
| ODConv_neck1 | 92.7 | 7.24 | 15 |
| ODConv_neck2 | 92.9 | 7.19 | 14 |
| ODConv_neck3 | 92.8 | 7.18 | 14 |
| ODConv_all | 92.7 | 7.27 | 19 |

在表2中，S-YOLOv5s为改进多尺度检测层的检测模型，ODConv_backbone表示在改进模型的第1、3、5、7层应用此模块，ODConv_neck1表示在第14、18、22、25、28层应用此模块，ODConv_neck2表示在第18、22、25、28层应用此模块，ODConv_neck3表示在第22、25、28层采用此模块，ODConv_all表示在上述层中应用此模块。通过分析表2对比试验结果，本文确定在改进模型的第18、22、25、28层应用ODConv动态卷积模块，如图4所示，以便更为有效地细化苹果成熟度的细粒度特征。

2.3 模型消融试验

为了验证改进模型中各模块的有效性,本文在自制数据集的基础上进行消融试验(如图8所示),OD-YOLOv5s 为引入 ODConv 动态卷积模块的检测模型, SOD-YOLOv5s 为改进多尺度检测层以及融入 ODConv 模块的检测模型, C3STR-YOLOv5s 为融入 C3STR 模块的检测模型, SODSTR-YOLOv5s 为改进多尺度检测层以及融入 ODConv 模块和 C3STR 模块的最终模型,即本文所提出的改进模型。

由图8(b)可知, YOLOv5s 模型对被遮挡的苹果特征提取能力较差,存在较为明显的漏检现象;由图8(c)可知, S-YOLOv5s 引入 160×160 特征图的检测头,提升了在密集情况下的小尺寸的不同成熟度苹果检测效果;由图8(d)可知, OD-YOLOv5s 的 ODConv 动态卷积模块细化苹果的细粒度特征,有效地提取不同成熟度苹果的颜色特征,在一定程度上解决了 YOLOv5s 模型因遮挡导致的漏检问题;由图8(e)可知, SOD-YOLOv5s 运用 ODConv 模块和 160×160 特征图的检测头提高了模型的检测效果,但依旧存在极少数漏检现象;由图8(f)可知, C3STR 模块的 Swin Transformer Block 将同级成熟度苹果的纹理特征进行融合以构建层级特征表,很大程度上解决了 YOLOv5s 存在的模型泛化能力较弱的问题;由图8(g)可知,本文所改进的模型 SODSTR-YOLOv5s 解决了其他对比模型所存在的漏检问题,并实现在同一张图像中准确检测中、低成熟度苹果,提高了检测的准确性。



图8 模型消融试验结果对比图

Fig.8 Comparison chart of model ablation test results

注: 图中数据表示置信度, □表示高成熟度, □表示中成熟度, □表示低成熟度。

Notes: The data in figures represent confidence, □ indicates high ripeness, □ indicates medium ripeness, □ indicates low ripeness.

由于人眼直观的观察试验结果无法全面客观地反映模型检测效果,为此本文采用平均精度 AP, 平均精度均值 mAP, 平均检测时间 MDT 以及参数量四项评价指标对不同模块进行有效性分析,结果如表3所示。

表3 SODSTR-YOLOv5s 模型消融试验结果

Tab.3 Results of SODSTR-YOLOv5s model ablation experiment

| 模型 Model | 高成熟度 High ripeness | AP/% 中成熟度 Medium ripeness | 低成熟度 Low ripeness | mAP/% | 平均检测时间/ms Mean Detection Time | 参数量/M Parameters |
|-------------|-----------------------|---------------------------------|----------------------|-------|-------------------------------------|---------------------|
| YOLOv5s | 91.0 | 90.5 | 90.6 | 90.7 | 11 | 7.02 |
| S-YOLOv5s | 91.9 | 91.1 | 91.5 | 91.5 | 12 | 7.17 |
| OD-YOLOv5s | 92.8 | 91.7 | 92.4 | 92.3 | 13 | 7.05 |
| SOD-YOLOv5s | 93.3 | 92.5 | 93.0 | 92.9 | 14 | 7.19 |

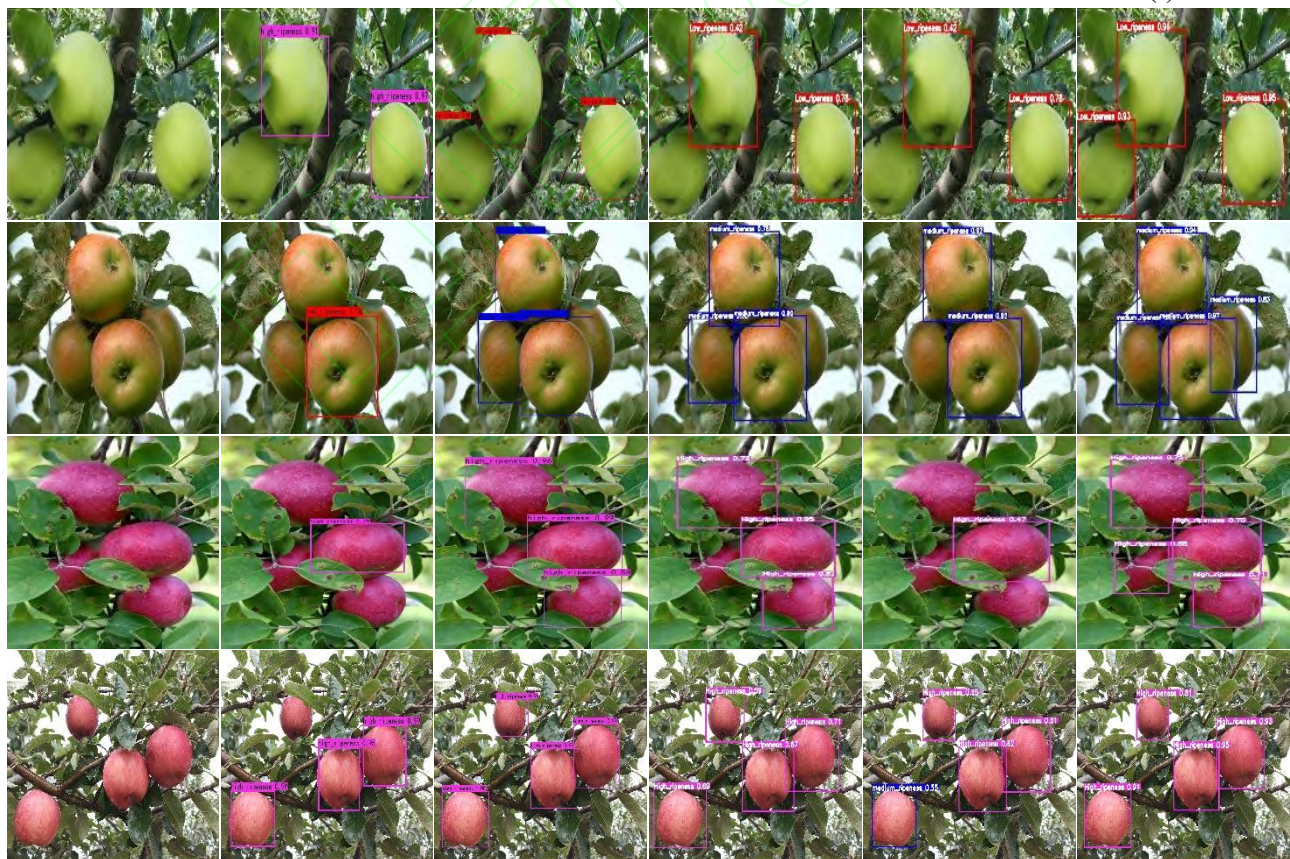
| | | | | | | |
|----------------|------|------|------|------|----|------|
| C3STR-YOLOv5s | 93.0 | 92.7 | 92.8 | 92.8 | 14 | 7.15 |
| SODSTR-YOLOv5s | 94.1 | 93.1 | 93.7 | 93.6 | 16 | 7.34 |

由表 3 可知，相比于 YOLOv5s 模型，S-YOLOv5s 的 mAP 提升了 0.8%，表明改进多尺度检测层能够提高小尺寸苹果的检测精度，OD-YOLOv5s 的 mAP 提升了 1.6%，说明 ODConv 能够精准定位苹果局部颜色特征，提高改进模型对被遮挡苹果的检测精度，C3STR-YOLOv5s 的 mAP 提升了 2.1%，证实 Swin Transformer Block 能够弱化同级成熟度苹果特征差异，有效地提高了改进模型对同级成熟度苹果的检测精度，改进模型 SODSTR-YOLOv5s 的 mAP 提升了 2.9%，参数量与平均检测时间 MDT 分别增加了 0.32M 和 5ms，虽然改进模型的参数量和平均检测时间略有增加，但对于不同成熟度苹果的检测精度有着较为明显的提升，较好地满足了自然环境下对不同成熟度苹果的检测要求。

2.4 不同目标检测模型性能对比

为了定性地评价本文所提出的改进模型，将改进模型与 4 种深度学习目标检测模型进行对比试验，试验结果如图 9 所示，图 9 中的图像从上到下依次为傍晚时低成熟度苹果图像、白天时中成熟度苹果图片、晴天时高成熟度苹果图像、雨天时高成熟度苹果图像。检测模型从左到右依次为 SSD^[27]、Faster R-CNN^[28]、YOLOv5-Transformer-BiFPN^[14]、YOLOv5s、改进模型。

由图 9(b)可知，SSD 模型存在错检漏检问题，降低了其检测精度；由图 9(c)可知，Faster R-CNN 模型的检测效果优于 YOLOv5s 模型和 SSD 模型，改善了漏检现象，但对处于光照程度较强的苹果仍然存在漏检情况；由图 9(d)可知，虽然 YOLOv5-Transformer-BiFPN 的检测效果优于 Faster R-CNN、SSD、YOLOv5s 模型，但对被遮挡的苹果检测效果较差，出现漏检问题；由图 9(e)可知，YOLOv5s 模型对苹果的高、中成熟度检测效果较差，存在较为严重的漏检问题。改进模型 SODSTR-YOLOv5s 不仅较好地解决了被遮挡苹果以及特征分布差异较大的苹果检测效果较差的问题，同时提高了对于小尺寸的苹果的检测效果，无错检漏检现象，从而明显地提升了检测模型对于不同情况下的不同成熟度苹果的检测效果，如图 9(f)所示。



(a)原始图片
Original images

(b)SSD

(c)Faster R-CNN

(d)YOLOv5-Transformer-BiFPN

(e)YOLOv5s

(f)SODSTR-YOLOv5s

图 9 不同目标检测模型的试验结果对比图

Fig.9 Comparison of experimental results of different target detection models

注：图中数据表示置信度；□表示高成熟度，□表示中成熟度，□表示低成熟度。

Notes: The data in figures represent confidence; □ indicates high ripeness, □ indicates medium ripeness, □ indicates low ripeness.

由于图 9 所示的试验结果对比图不能全面客观地反映改进模型的有效性，为此本文采用精确率 P，召回率 R，mAP 以及平均检测时间 MDT 和参数量五项评价指标评价，结果如表 4 所示。

表 4 不同目标检测模型性能对比

Tab.4 Performance comparison of different target detection models

| 模型 Models | 精确率/% Precision | 召回率/% Recall | mAP/% | 平均检测时间/ms Mean Detection Time | 参数量/M Parameters |
|--------------------------|--------------------|-----------------|-------|----------------------------------|---------------------|
| SSD | 79.6 | 88.6 | 90.9 | 56 | 24.79 |
| Faster R-CNN | 80.1 | 95.1 | 91.5 | 129 | 27.57 |
| YOLOv5-Transformer-BiFPN | 87.3 | 95.8 | 92.8 | 13 | 7.17 |
| YOLOv5s | 85.3 | 90.5 | 90.7 | 11 | 7.02 |
| SODSTR-YOLOv5s | 89.1 | 95.5 | 93.6 | 16 | 7.34 |

由表4可知，虽然SSD和Faster R-CNN模型的 mAP 优于YOLOv5s，但因参数量较大，从而导致检测速度较慢，故该两类模型无法满足自然环境下对不同成熟度苹果的实时检测要求。YOLOv5-Transformer-BiFPN模型在一定程度上解决了同级成熟度苹果特征分布差异的问题，提高模型的检测精度，且 mAP 高于 SSD、Faster R-CNN和YOLOv5s模型，但此模型精确率仍低于本文改进模型，表明该模型仍未能有效地解决检测被遮挡苹果时存在的漏检问题。SODSTR-YOLOv5s在召回率R上，相比于SSD、Faster R-CNN、YOLOv5s分别提高了6.9%、0.4%、5.0%；在 mAP 上，相比于四种对比检测模型，分别提高了2.7%、2.1%、0.8%、2.9%；在平均检测时间MDT上，相比于 SSD、Faster R-CNN分别减少了40ms、113ms，相比于 YOLOv5-Transformer-BiFPN、YOLOv5s增多了3ms、5ms；在参数量上，相比于 YOLOv5-Transformer-BiFPN、YOLOv5s增多了0.17M、0.32M，相比于 SSD、Faster R-CNN分别减少了 17.45M 、20.23M。本文改进模型虽然在参数量上略有瑕疵，但在精确率P，mAP上均优于其他检测模型，能够有效地解决因遮挡以及纹理特征分布差异所导致的检测精度较低的问题，从而有效地提升不同成熟度苹果的检测能力，实现在自然环境下对不同成熟度苹果较为精确的检测。

3 结论

本文以红富士、嘎啦苹果为研究对象，提出了一种不同成熟度苹果目标检测模型 SODSTR-YOLOv5s，以解决自然环境下因苹果遮挡以及特征分布差异所导致的检测精度较低问题。首先改进了 YOLOv5s 多尺度检测层，构建 160×160 检测头，增强小尺寸的不同成熟度苹果的检测精度，其次在骨干网络 Backbone 结构中融入 Swin Transformer Block 增强同级成熟度苹果的纹理特征融合，加强模型泛化性，最后在 Neck 结构中建立 ODConv 动态卷积模块细化不同成熟度苹果局部细粒度特征。

本文基于自制数据集进行改进模型试验，试验结果表明，本文所提出的改进模型 SODSTR-YOLOv5s 的精确率 P 为 89.1%，召回率 R 为 95.5%，平均精度均值 mAP 为 93.6%，参数量为 7.34 M，平均检测时间为 16 ms。本文改进模型相比于 YOLO v5s 虽然参数量和平均检测时间皆略有增加，但提高了不同成熟度苹果的检测精度，能有效地满足自然环境下实际应用的要求。

本文所做研究仍有一定的缺点，主要表现为以下两个方面：1) 改进模型检测效果虽提升较为明显，但对于检测模型的轻量化未进行改进；2) 本文所构建的数据集包含的苹果种类较少，数据集是以红富士、嘎啦苹果为基础进行构建，未涵盖一些特殊表皮颜色的苹果种类。后续将研究改进轻量化不同成熟度苹果目标检测模型以及多种类苹果数据集的构建，为苹果智能机械自动化采摘研究提供助力。

参考文献 References:

[1] 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(09):28-38,59.

- Yan B, Fan P, Wang M R, 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(09):28-38,59 (in Chinese with English abstract) .
- [2]王丽娟,陈浩然,季石军,等. 机器视觉成熟度检测的苹果色选分选机设计[J]. 农业与技术,2022,42(12):36-40.
- Wang L J, Chen H R, Ji S J, et al. Machine Vision Ripeness Inspection for Apple Color Sorting Machine Design[J]. Agriculture and Technology, 2022, 42(12):36-40 (in Chinese) .
- [3]刘志刚,王丽娟,喜冠南,等. 水果成熟度检测技术的现状与发展[J]. 农业与技术,2020,40(08):17-21.
- Liu Z G, Wang L J, Xi G N, et al. Status and development of fruit ripeness detection technology[J]. Agriculture and Technology, 2020, 40(08):17-21 (in Chinese) .
- [4]Mazen F, Nashat A A. Ripeness classification of bananas using an artificial neural network[J]. Arabian Journal for Science and Engineering, 2019, 44(8): 6901-6910.
- [5]Yu L, Xiong J, Fang X, et al. A litchi fruit recognition method in a natural environment using RGB-D images[J]. Biosystems Engineering, 2021, 204: 50-63.
- [6]包晓敏,王思琪.基于深度学习的目标检测算法综述[J].传感器与微系统,2022,41(04):5-9.
- Bao X M, Wang S Q. Survey of object detection algorithm based on deep learning[J]. Transducer and Microsystem Technologies, 2022, 41(04):5-9 (in Chinese with English abstract) .
- [7]Ni X, Li C, Jiang H, et al. Deep learning image segmentation and extraction of blueberry fruit traits associated with harvestability and yield[J]. Horticulture research, 2020, 7.
- [8]Wan P, Toudeshki A, Tan H, et al. A methodology for fresh tomato ripeness detection using computer vision[J]. Computers and electronics in agriculture, 2018, 146: 43-50.
- [9]龙洁花,赵春江,林森,等. 改进 Mask R-CNN 的温室环境下不同成熟度番茄果实分割方法[J]. 农业工程学报,2021,37(18):100-108.
- Long J H, Zhao C J, Lin S, et al. Segmentation method of the tomato fruits with different maturities under greenhouse environment based on improved Mask R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(18):100-108(in Chinese with English abstract) .
- [10]Parvathi S, Selvi S T. Detection of maturity stages of coconuts in complex background using Faster R-CNN model[J]. biosystems engineering, 2021, 202: 119-132.
- [11]Altaheri H, Alsulaiman M, Muhammad G. Date fruit classification for robotic harvesting in a natural environment using deep learning[J]. IEEE Access, 2019, 7: 117115-117133.
- [12]Fan Y, Zhang S, Feng K, et al. Strawberry maturity recognition algorithm combining dark channel enhancement and YOLOv5[J]. Sensors, 2022, 22(2): 419.
- [13]陈锋军,张新伟,朱学岩,等. 基于改进 EfficientDet 的油橄榄果实成熟度检测[J]. 农业工程学报,2022,38(13):158-166.
- Chen F J, Zhang X W, Zhu X Y, et al. Detection of the olive fruit maturity based on improved EfficientDet[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022,38(13):158-166 (in Chinese with English abstract) .
- [14]张志远,罗铭毅,郭树欣,等. 基于改进 YOLO v5 的自然环境下樱桃果实识别方法[J]. 农业机械学报,2022,53(S1):232-240.
- Zhang Z Y, Luo M Y, Guo S X, et al. Cherry Fruit Detection Method in Natural Scene Based on Improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(S1):232-240 (in Chinese with English abstract).
- [15]耿磊,黄亚龙,郭永敏. 基于融合注意力机制的苹果品种分类方法[J].农业机械学报,2022,53(06):304-310+369.
- Geng L, Huang Y L, Guo Y M. Apple Variety Classification Method Based on Fusion Attention Mechanism[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022,53(06):304-310+369(in Chinese with English abstract) .
- [16]Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 779-788.
- [17]杨福增,雷小燕,刘志杰,等. 基于 CenterNet 的密集场景下多苹果目标快速识别方法[J]. 农业机械学报,2022,53(2):265-273.
- Yang F Z, Lei X Y, Liu Z J, et al. Fast Recognition Method for Multiple Apple Targets in Dense Scenes Based on CenterNet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(2):265-273 (in Chinese with English abstract).
- [18]王卓,王健,王泉雄,等. 基于改进 YOLO v4 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.
- Wang Z, Wang J, Wang X X, et al. Lightweight Real-time Apple Detection Method Based on Improved YOLOv4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8):294-302(in Chinese with English abstract).
- [19]表明新,于洪涛,江亚峰,等. 联合迁移学习和自适应学习率的苹果成熟度识别[J]. 中国农机化学报,2019,40(11):131-135.
- Yuan M X, Yu H T, Jiang Y F, et al. Apple ripeness recognition combining transfer learning and adaptive learning rate[J]. Journal of Chinese Agricultural Mechanization, 2019, 40(11):131-135(in Chinese with English abstract) .

[20]李豪. 基于多指标因子的苹果成熟度无损检测方法及设备研发[D]. 杨凌: 西北农林科技大学,2023.

LI H. Research on non-destructive testing method and equipment for apple ripeness based on multi-indicator factors[D]. Yangling: Northwest A & F University,2023 (in Chinese with English abstract).

[21]郁网庆,吕平,贾连文,等. 苹果成熟度确定方法[J]. 中国果菜,2012(11):37-39.

Yu W Q, Lü P, Jia L W, et al. Apple ripeness determination method[J]. Chinese Fruit and Vegetables, 2012(11):37-39(in Chinese).

[22]Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117-2125.

[23]Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 8759-8768.

[24]Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 10012-10022.

[25]Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[EB/OL]. [2022-09-16]. <https://arxiv.org/abs/2209.07947>

[26]Kingma D P, Ba J. Adam: A method for stochastic optimization[EB/OL]. [2014-12-22]. <https://arxiv.org/abs/1412.6980>

[27]Liu W, Anguelov D, Erhard D, et al. SSD: Single shot multibox detector[C]//Springer. European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[28]Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28:91-99.