

## 基于深度学习的类球状水果采摘识别算法研究进展

李辉<sup>1</sup>, 张俊<sup>2\*</sup>, 俞烁辰<sup>2</sup>, 李志鑫<sup>2</sup>

(<sup>1</sup>湖州师范学院信息工程学院, 浙江湖州 313000; <sup>2</sup>浙江省农业科学院食品科学研究所重点实验室, 浙江杭州 310021)

**摘要:**【目的】中国在水果产量方面处于全球领先地位, 但面临着人力资源减少和老龄化问题, 传统的人工采摘方式已经无法满足快速高效的采摘需求, 研发集成计算机视觉的自动化水果采摘设备成为解决劳动力短缺难题的关键。【方法】水果大多呈类球状, 相关的识别算法研究居多。本文探讨了柑橘、蜜桃等类球状水果的识别算法。根据应用场景的不同, 本文分析了传统类球状水果识别算法与基于深度学习的类球状水果识别算法在网络结构方面的改进, 对水果采摘识别算法进行总结并提出算法的未来发展趋势。【结果】在复杂环境下, 本文通过对比发现传统算法在简单场景下表现有效, 但在复杂环境中往往会受到设计特征的限制, 基于深度学习的算法因其高效性和准确性更适合自动化水果采摘的需求。【结论】本文总结了类球状水果识别算法的研究进展, 深度学习算法在处理复杂环境下具有良好的有效性和适应性, 更适合部署在自动化采摘设备。提出了未来的研究方向, 包括优化算法性能、数据集构建及扩增, 以及结合多模态数据以提升算法的精度和适应性。

**关键词:** 目标检测算法;深度学习;卷积神经网络;计算机视觉;水果采摘

中图分类号: S66 文献标志码: A 文章编号: 1009-9980(2025)02-0001-08

## Research Progress of Globular Fruit Picking Recognition Algorithm Based on Deep Learning

LI Hui<sup>1</sup>, ZHANG Jun<sup>2</sup>, YU Shuo Chen<sup>2</sup>, LI Zhixin<sup>2</sup>

(<sup>1</sup>School of Information Engineering, Huzhou University, Huzhou 313000, China; <sup>2</sup>Key Laboratory of Food Science Research Institute, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China)

**Abstract:** 【Objective】China is a global leader in fruit production, and fruit picking mainly relies on manual labor, which helps to select fruits according to fruit size and quality to reduce loss in this way, different techniques and tools can be adopted according to the characteristics and picking needs of each fruit tree. However, the present picking field is faced with the problem of decreasing human resources and aging problem. Meanwhile, the traditional manual picking method has become unable to meet the demand for fast and efficient picking. To solve the problem of labor shortage, the research and development of automated fruit picking equipment with integrated computer vision has become the key to solve the problem of labor shortage. It can effectively improve the efficiency and quality of fruit picking. 【Methods】Automatic picking equipment combined with computer vision often uses object detection algorithms to identify objects, and object detection algorithms can be divided into traditional algorithms and deep learning-based object detection algorithms. Traditional algorithms identify the position and bounding box of a specific object in an image or video, usually by preprocessing the image (Scaling, grayscale or normalization), feature extraction (using traditional hand-designed features or automatic learning based

收稿日期: 2024-06-18 接受日期: 2024-12-06

基金项目: 国家柑橘产业技术体系 (CARS-26-29)

作者简介: 李辉, 男, 硕士, 研究方向为基于 3D 视觉的采摘机器人检测算法研究。E-mail: 3023763876@qq.com

\*通信作者 Author for correspondence. 张俊, 男, 博士, 副研究员, 硕士生导师, 主要研究农产品加工智能装备。E-mail: hunterzju@163.com

on machine learning), classification or regression (confirming object class and location), and non-maximum suppression to further optimize and filter detected objects. When traditional fruit detection algorithms process images in complex environments, their limited expression ability and robustness are easily affected by illumination, occlusion and other factors, resulting in a decline in recognition accuracy. Furthermore, with the increase of feature complexity and computation amount, the algorithm processing speed will be reduced. When changing scenes, adding fruit types, and updating features, the feature extractor needs to be redesigned and adjusted, and in special cases, the entire system needs to be retrained. Compared with traditional fruit detection algorithms, the fruit detection algorithm based on deep learning can extract and learn rich features from a large amount of data, and has higher accuracy and robustness when processing noisy data. When changing new environments and adding new categories, the fruit detection algorithm based on deep learning can improve the recognition ability and recognition accuracy of the model through transfer learning, data enhancement, multi-model combination, feature fusion and multi-modal data. **【Result】** Fruit detection algorithms based on deep learning can be divided into two categories: one-stage target detection algorithm and two-stage target detection algorithm. The one-stage object detection algorithm achieves end-to-end detection by using a single convolutional neural network to directly predict the target location and category. This method achieves fast detection while maintaining high accuracy, transforms the problem of target detection into a regression problem, and completes the location and classification of the target directly. In the training and deployment phase of the algorithm, the first-stage object detection algorithm uses pruning and quantization techniques to reduce the model size, which is suitable for running in mobile devices or embedded systems with limited resources. The two-stage target detection algorithm is called the target detection algorithm based on region of interest or region suggestion, which is usually divided into two stages: 1) generate a large number of candidate regions by selective search, regional suggestion network (RPN) and other methods; 2) through the network processing including classifiers and boundary box regressors, the candidate region is identified and accurately located. Traditional algorithms are effective in simple scenarios, but are often limited by design features in complex environments. Algorithms based on deep learning are more suitable for automated fruit picking due to their high efficiency and accuracy. This paper summarizes the improvement and application of traditional Object detection algorithm and deep learning-based Object detection algorithm. And, This paper summarizes the improvements and applications of traditional spherical fruit detection algorithms and deep learning-based spherical fruit detection algorithms, and analyzes the advantages and disadvantages of these algorithms in different use scenarios. **【Conclusion】** This paper summarizes the fruit picking recognition algorithm and puts forward the future development trend of the algorithm. With model optimization and lightweight as the starting point, efficient network architecture or model compression technology is adopted to reduce computational complexity and model size, improve model processing speed and adapt to mobile automatic picking equipment. Enhance data processing, improve model generalization by preprocessing and synthesizing data, and optimize model adaptability in changing environments. The accuracy and robustness of model recognition are improved by combining spectral, infrared, laser and other sensor data. The model adaptive adjustment algorithm was developed to adjust strategies and parameters according to real-time feedback and adapt to different fruit picking operations and different picking environmental conditions. In the fruit picking recognition algorithm based on deep learning, YOLO can directly predict the boundary box and category

probability of the target in a single forward propagation to achieve near real-time detection, which is very important for the fruit picking robot in the orchard that needs fast response. The end-to-end design of YOLO simplifies the training inspection process, reduces complexity, and enables faster deployment in picking robot systems. In the changeable environment of orchards and groves, YOLO can effectively distinguish between fruit and background, improving the accuracy of detection. With the continuous research of domestic scholars, YOLO algorithm is also continuously iteratively optimized, and its detection ability of objects of different sizes and shapes is significantly improved, which can adapt to the maturity, size and occlusion of fruits, and improve the detection performance in complex environments.

**Keywords:** Object detection algorithms; Deep learning; Convolutional neural networks; Computer vision; Fruit picking

我国水果产量位居全球第一<sup>[1]</sup>, 常见的类球状水果如柑橘、苹果和桃子, 其采摘工作主要依赖人工。人工采摘能根据果实成熟程度、大小和品质选择果实减少损耗, 并根据不同果树特点和采摘需求采用不同技术与工具。随着人口老龄化导致劳动力减少, 仅靠人工采摘无法高效完成采摘任务, 开发自动化水果采摘设备成为解决这一问题的重要途径。

结合计算机视觉的自动化水果采摘设备能利用采摘识别算法识别果园中的水果, 判断水果品质并精确定位, 为采摘设备提供准确的引导<sup>[2]</sup>。与传统的人力采摘相比, 结合计算机视觉的自动化采摘设备具有识别速度快、精度高、成本低等优点, 能够提升水果采摘的效率和质量<sup>[3]</sup>。在采摘类球状水果时, 类球状水果在二维图像中呈现接近圆形的轮廓, 稳定的几何形状能帮助算法减少形状复杂性带来的误检问题<sup>[4]</sup>。并且, 类球状水果规则的形状和一致的颜色分布, 容易进行有效的数据增强操作, 增加训练数据的多样性, 提升算法的泛化能力, 更好地适应不同果园环境, 进一步提高设备的通用性和稳定性。自动化采摘设备在自然环境下作业同样会遇到许多问题, 例如光照变化、遮挡问题、多样性、复杂性、背景干扰以及实时性要求, 这些问题都会影响识别的准确率。为克服这些挑战, 需要持续改进目标检测算法的鲁棒性以及适应性, 在有限的算力资源下降低算法参数量提高运行速度, 更好地应用于自动化采摘设备中<sup>[5]</sup>。

近年来, 针对类球状水果采摘中的目标检测问题提出了不同的算法和技术, 主要分为两大类: 传统目标检测算法和基于深度学习的目标检测算法。基于深度学习的目标检测算法进一步细分为单阶段目标检测算法<sup>[6]</sup>和两阶段目标检测算法<sup>[7]</sup>。本文介绍应用于类球状水果采摘的目标检测算法, 并对每类算法的研究成果进行归纳分析。

## 1 传统目标检测算法

传统目标检测算法<sup>[8]</sup>在识别图像或视频中特定对象的位置和边界框, 通常步骤包括对图片进行预处理(如: 缩放、灰度化或归一化)、特征提取(利用传统手工设计特征或是基于机器学习的自动学习特征)、分类或回归(确认对象类别和位置)以及利用非极大值抑制等方法对检测到的对象进一步优化和过滤。Liu等<sup>[9]</sup>提出使用简单线性迭代聚类算法对果园图像进行超像素块分割, 从块中提取颜色特征确定候选区域, 采用方向梯度直方图描述果实形状, 用于检测定位果实。夏康利等<sup>[10]</sup>提出基于 HSV 颜色空间统计特征的水果识别技术, 将 RGB 水果图像转换为 HSV 颜色空间, 将色调分布近似为拉普斯分布作为果实的特征描述, 使用 Meanshift 算法<sup>[11]</sup>进行图像分割, 通过计算输入水果图像的色调数据的马氏距离与预设的特征马

氏距离进行比较,判断输入的水果类别。邹伟<sup>[12]</sup>采用工业相机获取柑橘 RGB 图像,将柑橘图像的 RGB 颜色空间转换为 HSV 颜色空间,按 H、S、V 三通道计算颜色直方图,利用 H 通道峰值以及对应色调值对柑橘成熟度进行判断,实验证明对柑橘成熟度的检测准确率达到 90%以上。陈雪鑫等<sup>[13]</sup>提出一种基于多颜色特征和纹理特征的水果识别算法,采用颜色矩算法和非均匀量化算法对图像 RGB、HSV 颜色空间提取颜色特征,使用局部二值化提取纹理特征,优化组合颜色和纹理特征向量,利用 BP 神经网络作为分类器对样本训练分类,最终实验表明多特征的结合让分类准确率达 90%以上,高于单一特征算法准确率。徐惠荣等<sup>[14]</sup>设计了基于彩色信息的树上柑橘识别算法,对各种天气、光照场景采集图像,对图像进行颜色提取,利用柑橘、枝叶在 R-B 颜色指标的差异建立柑橘识别颜色空间,利用动态阈值法,将柑橘从背景分割,实现对树上单个或多个柑橘的识别。上述传统目标检测算法主要依赖于水果的形状、颜色等单一或组合特征来进行识别。通过对背景进行建模和特征融合,此类算法能提取果实信息并对其进行分割,从而在自然环境下实现对水果的有效检测。

传统目标检测算法在特定场景下表现良好,但依赖手工设计的特征,难以适应复杂场景和识别变化。在自然环境中,传统检测算法的表达能力和鲁棒性有限,易受到光照变化、枝叶遮挡、果实重叠等因素的影响,导致识别准确率下降。当场景变更、添加水果种类和更新特征时,需要重新设计和调整特征提取器,特殊情况甚至需要重新训练整个系统。相比之下,基于深度学习的目标检测算法能从大量数据提取学习到丰富的特征,具备更高的准确性和鲁棒性。场景变更和添加水果种类时,基于深度学习的目标检测算法可以通过迁移学习、数据增强、多模型组合、特征融合以及多模态数据提高模型的识别能力与鲁棒性。

## 2 基于深度学习的目标检测算法

基于深度学习的目标检测算法可分为两大类:单阶段目标检测算法和两阶段目标检测算法。单阶段目标检测算法通过使用单个卷积神经网络(Convolutional Neural Network, CNN)直接预测目标位置和类别,实现端到端的检测。这种方法保持高准确率的同时实现快速检测,将目标检测问题转化为回归问题,直接对目标完成位置定位与分类。在算法的训练和部署阶段,单阶段目标检测算法采用剪枝和量化等技术减小模型尺寸,适合在资源有限的移动设备或嵌入式系统中运行;而两阶段目标检测算法被称为基于感兴趣区域或基于区域建议的目标检测算法,这类算法运行时通常分为两个阶段:1)利用选择性搜索、区域建议网络(Region Proposal Network, RPN)<sup>[15]</sup>等方法生成大量候选区域;2)通过包含分类器和边界框回归器的网络处理,对候选区域进行识别目标与精确定位。

### 2.1 单阶段目标检测算法

单阶段目标检测算法省略生成候选区域步骤,直接在特征图中生成类概率和位置坐标,再进行分类回归。常见的单阶段目标检测算法有 YOLO、SSD<sup>[16]</sup>、MobileNet<sup>[17]</sup>、ShuffleNet<sup>[18]</sup>、Swin-Transformer<sup>[19]</sup>,其中 YOLO 系列模型应用最多。Redmon 等<sup>[20]</sup>为解决两阶段目标检测算法检测速度慢、提取特征区域重复等问题提出了 YOLOv1 算法,将目标检测转化为回归问题,使用全局特征预测边界框。YOLOv1 通过图像均匀分割避免重复计算,提高了检测速度,适用于实时性要求高的自动化采摘设备,但其精度较低。Redmon 等<sup>[21]</sup>提出的 YOLOv2 算法,通过引入锚框和联合训练提升了精度,但在复杂场景和小目标检测中仍存在误检。为了进一步提升多尺度特征提取能力,Redmon 等<sup>[22]</sup>提出了 YOLOv3 算法,采用更深的 DarkNet-53 骨

干网络，并结合特征金字塔网络（Feature Pyramid Networks, FPN）<sup>[23]</sup>进行特征融合，显著增强了不同尺度目标的检测能力。Bochkovskiy 等<sup>[24]</sup>提出的 YOLOv4 算法，引入 Mosaic 数据增强、CSPDarkNet-53 骨干网络和 SPP 模块，提升了复杂背景和遮挡情况下的精度，同时保持实时性。在 YOLOv4 取得成功的基础上，研究者进一步推出了 YOLOv5 算法<sup>[25]</sup>，YOLOv5 通过自适应锚框、Focus 模块和轻量化设计，适合资源受限的设备。Li 等<sup>[26]</sup>提出的 YOLOv6 算法取消了锚框，采用 EfficentRep 和 Rep-PAN 模块，虽然提高了检测精度，但其复杂的结构不适合部署在资源受限的移动采摘设备上。Alexey 等<sup>[27]</sup>提出了 YOLOv7 算法。通过 BConv、E-ELAN 和 MPCConv 层优化特征提取，不依赖锚框，提升了硬件适应性。Ultralytics 公司提出 YOLOv8 算法<sup>[28]</sup>，YOLOv8 采用 C2F 模块和解耦头，进一步提升了检测速度和精度，适应多场景需求。YOLO 系列算法经过长期演变，其核心优化始终围绕速度与精度的平衡展开。YOLO 以实时性为目标，在保持高检测精度的同时，能够满足各类应用场景对快速响应的需求（如表 1 罗列的各类 YOLO 算法改进点及目标），而这正是自动化水果采摘领域的关键所在。因此，YOLO 成为果实识别任务中应用最广泛的目标检测算法。随着技术的发展，许多基于 YOLO 的优化模型不断涌现，为提高识别精度和实时性能提供了有效的解决方案。

表 1 各类 YOLO 算法改进点及目标

Table 1 Improvement and objectives of various Yolo algorithms

模型 Model	骨干网络 Backbone network	改进点及目标 Improvements and objectives
YOLOv1	GoogLeNet	改进点：将目标检测任务视为回归问题，使用全局图像特征预测边界框。 目标：实现快速检测，适用于实时性高的场景，但精度有限。 Improvement: Treats object detection as a regression problem, using global image features to predict bounding boxes. Objective: Achieves fast detection, suitable for real-time scenarios, but with limited accuracy.
YOLOv2	Darknet-19	改进点：引入批量归一化、多尺度训练和锚框。 目标：提升模型精度，保持较高响应速度。 Improvement: Introduced batch normalization, multi-scale training, and anchor boxes. Objective: Enhance model accuracy while maintaining a high response speed.
YOLOv3	Darknet-53	改进点：多尺度检测和 FPN 特征融合。 目标：增强对不同尺度目标的检测能力，适应复杂背景，同时保持快速响应。 Improvement: Multi-scale detection and feature fusion using FPN. Objective: Enhance the detection capability for objects of different scales, adapt to complex backgrounds, while maintaining fast response.
YOLOv4	CSPDarknet53	改进点：引入 CSP 结构减少计算量，Mish 激活函数，集成多种数据增强技术提高鲁棒性。 目标：在提高精度的同时，保持实时性，尤其适用于复杂背景下的应用。 Improvement: Introduced CSP structure to reduce computational load, Mish activation function, and integrated various data augmentation techniques to enhance robustness. Objective: Maintain real-time performance while improving accuracy, particularly suitable for applications in complex backgrounds.
YOLOv5	CSPDarknet-53	改进点：Focus 结构、自适应锚框和图像缩放，优化网络架构，模型轻量化。

		<p>目标：增加模型灵活性和效率，保证精度的同时，进一步提升轻量化，适合嵌入式和移动设备部署。</p> <p>Improvement: Focus structure, adaptive anchor boxes, and image scaling, optimizing the network architecture for model light weighting.</p> <p>Objective: Increase model flexibility and efficiency, ensuring accuracy while further enhancing lightweight design, suitable for deployment on embedded and mobile devices.</p> <p>改进点：通过 PAN 拓扑结构整合多尺度特征。</p>
YOLOv6	EfficientRep	<p>目标：针对多尺寸目标进行优化，但复杂性增加。</p> <p>Improvement: Integrated multi-scale features through PAN topology.</p> <p>Objective: Optimize for targets of various sizes, but increased complexity.</p> <p>改进点：引入新的模型缩放策略，确保模型能在不同硬件平台有效缩放，通过改进网络架构的层聚合方法，提高特征提取的效率和效果。</p>
YOLOv7	E-ELAN	<p>目标：提高模型在不同硬件平台上的性能，同时保持精度。</p> <p>Improvement: Introduced new model scaling strategies to ensure effective scaling across different hardware platforms, and improved feature extraction efficiency and effectiveness through enhanced layer aggregation methods in the network architecture.</p> <p>Objective: Enhance model performance on various hardware platforms while maintaining accuracy.</p> <p>改进点：骨干网络采用 C2F 模块，解耦头分离分类与定位，删除 object 分支，提升检测速度与准确性。</p>
YOLOv8	CSPDarkNet-53	<p>目标：进一步提升速度与精度，适应多种应用场景。</p> <p>Improvement: The backbone network employs the C2F module, decoupled heads separate classification from localization, eliminating the object branch to enhance detection speed and accuracy.</p> <p>Objective: Further improve speed and precision, suitable for various application scenarios.</p>

复杂的自然环境中，早期柑橘果实与背景的枝叶颜色相近，传统算法很难精确识别果实，常出现把绿色枝叶背景错误识别为果实以及漏检情况，为解决类似问题。宋中山等<sup>[29]</sup>提出 D-YOLOv3 算法，采用密集连接卷积网络（Densely Connected Convolutional Networks, DenseNet）<sup>[30]</sup>，加强特征传播，实现特征的复用。在构建数据集时，Song 等采集不同天气状况的柑橘图片，对图片进行高斯模糊、色彩平衡等处理，增加了数据集的多样性，有效提高了模型的泛化能力与鲁棒性，实验表明 D-YOLOv3 对早期柑橘果实的识别精确率达 83.01%。吕强等<sup>[31]</sup>基于 YOLOv5s 优化改进提出了早期柑橘果实检测算法 YOLO-GC。针对模型精度低、模型大的问题，Lv 等将骨干网络换为轻量级 GhostNet，嵌入全局注意力机制（Global Attention Mechanism, GAM）<sup>[32]</sup>提升对果实特征的提取能力。为改善枝叶遮挡、重叠造成的漏检问题，YOLO-GC 采用 GIoU 损失函数结合非极大值抑制（Soft-Non-Maximum Suppression, Soft-NMS）<sup>[33]</sup>算法优化边界框回归机制，最终实验表明，YOLO-GC 与 YOLOv5s 相比，权重文件大小为 6.69 MB，平均精度达到了 97.8%，在边缘安卓移动端 APP 对绿色柑橘果实检测，实验推理仅用时 108 ms。帖军等<sup>[34]</sup>提出一种基于混合注意力机制并改进 YOLOv5 模型的柑橘识别方法 YOLOv5-SC，Tie 等在骨干网络嵌入 SE<sup>[35]</sup>注意力与 CA<sup>[36]</sup>注

意力，使网络不仅能捕获方向和位置信息，也能捕获通道信息，让模型更好提取、定位柑橘图像特征。YOLOv5-SC 引入 Varifocal Loss<sup>[37]</sup>作为损失函数，能够更加平衡正负样本的损失。实验表明，YOLOv5-SC 平均精度达到了 95.09%，改善了将绿色背景误检成绿色柑橘果实的问题。

自然环境下，类球状水果往往以各种姿态分布在果树上，对果园这类远距离大视场的场景识别时，树叶遮挡、果实目标较小或果实密集分布等因素会导致目标检测算法在识别过程中出现漏检或误检的情况。为了解决此类问题，马帅等<sup>[38]</sup>提出基于 YOLOv4 改进的梨果实识别算法，将 SPP 模块中的最大池化法改为平均池化法，更多地保留目标信息，解决了漏检和误检的问题。其次，该算法 SPP 模块前后的卷积、PANet 中的部分卷积以及输出部分的卷积被替换为深度可分离卷积，在保证卷积效果不变的效果下减少了模型所占空间。使用训练后的改进 YOLOv4 模型对新获取的图像样本进行测试，改进后的模型权重文件大小为 136 MB，平均精度达到 90.18%。刘忠意等<sup>[39]</sup>对 YOLOv5 进行改进，提出了一种橙子果实识别算法。Liu 等将骨干网络部分 C3 模块替换为 RepVGG 模块，加强特征提取能力，将颈部网络中的普通卷积替换成鬼影混洗卷积（Ghost Shuffle Convolution, GConv）<sup>[40]</sup>，保证精度的同时降低模型参数量。为提高定位目标信息的准确率，该算法在预测头前加入了高通道注意力机制（Efficient Channel Attention, ECA）<sup>[41]</sup>，最后经过实验证明，改进后算法对橙子检测平均精度达到 90.1%，误检漏检的问题被有效解决，检测效果如图 1 所示，算法在无遮挡、复杂光照、枝叶遮挡及密集小目标场景下均展现了良好的检测效果，表明其具备较强的鲁棒性和泛化能力。

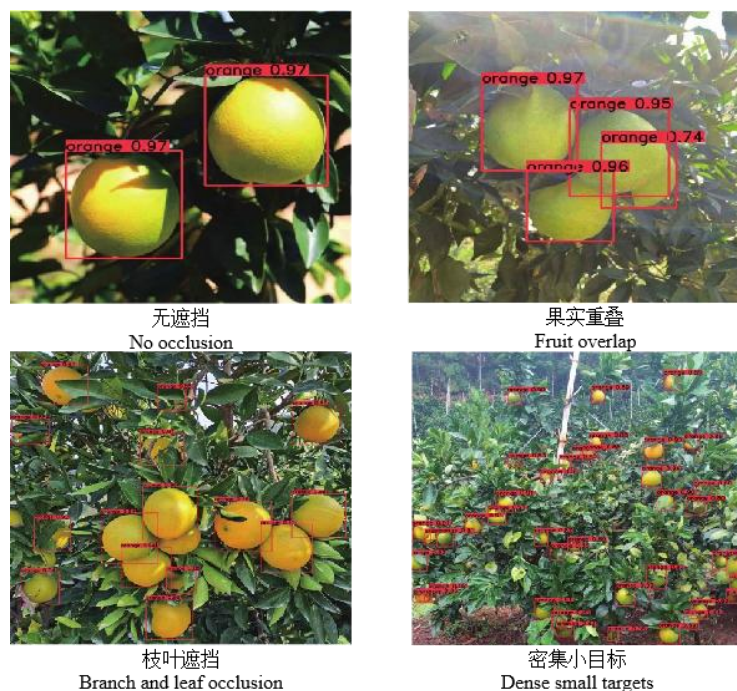


图 1 基于 YOLOv5 改进的算法检测效果图

Fig 1. Detection effect of improved algorithm based on Yolov5

贺英豪等<sup>[42]</sup>设计了一种基于 YOLOv5s 的改进算法，有效提升了对李识别的准确率，该算法骨干网络中的下采样卷积被替换成 FM 模块，保证模型下采样时不丢失严重遮挡目标和小目标的特征信息，使用 focal loss 和交叉熵函数的加权损失作为分类损失，提升密集目标的识别能力，最后测试模型性能发现平均精度提高 2.84%达到了 97.56%，小目标识别的平均精度达到 92.01%。为实现对柑橘果实的精确识别，黄彤颖

等<sup>[43]</sup>提出一种基于 YOLOv5 改进模型的识别方法。该算法通过引入卷积注意力模块 (Convolutional Block Attention Module, CBAM)<sup>[44]</sup>提高网络的特征提取能力,改善遮挡目标与小目标的漏检问题,采用 Alpha-IoU<sup>[45]</sup>损失函数代替 GIoU 损失函数作为边界框回归损失函数,提高边界框定位精度。最后结果显示模型平均精度达到 91.3%,对单张柑橘果实图像的检测时间为 16.7 ms。苑迎春等<sup>[46]</sup>提出基于改进 YOLOv4-Tiny 的果园环境下桃子实时识别算法, YOLOv4-Tiny-Peach 在骨干网络引入 CBAM,颈部网络添加大尺度浅层特征层,提高小目标识别精度,采用双向特征金字塔网络 (Bidirectional Feature Pyramid Network, BiFPN)<sup>[47]</sup>对不同尺度特征进行融合。通过训练, YOLOv4-Tiny-Peach 平均精度达 87.88%,与 YOLOv4-Tiny 相比,在大视场和早期桃子识别场景下模型提升明显。为提升全天候自动化采摘设备在夜间环境中的视觉检测能力,熊俊涛等<sup>[48]</sup>提出 Des-YOLOv3 算法,借鉴 ResNet<sup>[49]</sup>与 DenseNet,实现对多层特征的复用、融合,加强了夜间环境下对算法小目标、重叠遮挡果实识别的鲁棒性,检测效果如图 2 所示,实验表明, Des-YOLOv3 平均精度达 97.67%。此后,熊俊涛等<sup>[50]</sup>再次针对夜间采摘作业,提出基于 YOLOv5s 改进和主动光源结合的柑橘识别算法 BI-YOLOv5s,利用 BiFPN 进行多尺度交叉连接和加权特征融合,引入 CA 注意力加强定位信息提取,采用 C3TR 模块减少计算量并提取全局信息,实验后发现,在光源色环境下,模型对夜间柑橘识别准确率达 95.3%,实现了全天候自动化采摘作业。余圣新等<sup>[51]</sup>利用全维动态卷积替换 YOLOv8 系列模型中的部分普通卷积,以提高 YOLOv8 系列的鲁棒性,并将损失函数替换为 MPDIoU<sup>[52]</sup>,解决了原本 CIoU 损失函数退化的问题,通过实验验证,改进后的 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x 模型的平均精度都提高至 88.25%、89.32%、89.57%、89.90%、90.12%。岳有军等<sup>[53]</sup>基于 YOLOv8 设计了一个新的特征融合网络 Rep-YOLOv8 实现高层语义和低层空间特征融合。通过集成 EMA 注意力模块到 YOLOv8 中,使模型更关注果实区域,抑制背景和枝叶遮挡等一般特征信息。最后,将 C2f 模块替换为三支路 DWR 模块,通过多尺度特征融合提高小目标检测能力,使用 Inner-SIoU<sup>[54]</sup>损失函数提高模型精度。在果园环境中,以苹果作为检测对象,进行不同果实数量、不同成熟度的实验对比。实验结果表明,该算法平均精度达到 94%,在成熟果实大视场的识别场景下,改进后的算法各项指标均有显著提升,为果实识别任务提供有效支持。





图 2 Des-YOLOv3 算法检测效果图

Fig. 2 Detection effect of Des-yolov3 algorithm

目标检测算法通常包含庞大参数量和复杂的网络结构, 将这些模型部署到嵌入式平台时, 有限的计算资源会严重限制模型的实时响应速度。为解决这个问题, 吕石磊等<sup>[55]</sup>提出基于 YOLOv3 改进的轻量化柑橘识别方法 YOLO-LITE, 使用 MOBILENet-v2 作为骨干网络, 便于部署到移动终端, 引入 GIoU<sup>[56]</sup>边框回归损失函数。最终实验表明, YOLO-LITE 对柑橘目标检测速度可以达到 246 帧·s<sup>-1</sup>, 权重文件大小为 28 MB。王卓等<sup>[57]</sup>以 YOLOv4 算法为基础提出轻量级苹果实时检测算法 YOLOv4-CA, 使用轻量级网络 MobileNet-v3 作为特征提取网络, 并将 SE 注意力模块集成其中作为颈部基本块, 提高网络对特征通道的敏感程度, 增强特征提取能力。为有效压缩模型参数量和计算量, Wang 等将特征融合网络的普通卷积全部换为深度可分离卷积。最终实验表明算法平均检测精度达到 92.23%, 在嵌入式平台检测速度为 15.11 帧/秒, 内存占用量 54.1 MB, 在保证精度的同时, 满足采摘机器人实时性的需求。曾俊等<sup>[58]</sup>提出了 YOLO-Faster 算法对桃子完成实时快速检测, 在 YOLOv5s 基础上将骨干网络替换为 FasterNet<sup>[59]</sup>, 引入部分卷积 (Partial Convolution, PConv)<sup>[60]</sup>有效减少计算冗余和内存访问, 模型检测速度提升变得更加轻量化。骨干网络和颈部网络之间, 增加串联的卷积注意力模块和常规卷积模块, 强化骨干网络和颈部网络之间的特征融合和特征提取能力, 提高检测的准确性。算法采用 SIOU<sup>[61]</sup>作为损失函数, 解决预测框与真实框之间不匹配的问题, 更好地衡量预测框和真实框之间的匹配程度, 提高检测结果的质量。经过在自建数据集的训练和嵌入式设备 Jetson Nano 上的部署, 该算法平均精度达到了 88.6%, 权重文件大小为 8.3 MB, 相较于 YOLOv5s, 平均精度提升了 1%。赵辉等<sup>[62]</sup>提出基于改进 YOLOv3 的苹果识别算法, 将 DarkNet-53 网络残差模块与 CSP 模块<sup>[63]</sup>结合降低网络计算量, 加入 SPP 模块将全局、局部特征融合, 提高小目标召回率。采用 SoftNMS<sup>[64]</sup>算法增强重叠遮挡果实识别能力。改进后的算法平均精度达到 96.3%, 相较于 YOLOv3 提高了 3.8%, 满足了苹果自动采摘识别的准确性和实时性要求。然而, 当光线不足果实表面纹理特征不明显时, 算法的准确率可能会受到影响。Yan 等<sup>[65]</sup>对 YOLOv5S 算法进行了优化改进, 提高模型表达能力和空间信息损失处理能力, 并使其更适合部署在嵌入式设备上。Yan 等将模型骨干网络的 BottleneckCSP 模块桥分支上的卷积层移除, 把 BottleneckCSP 模块输入特征映射与另一个分支的输出特征映射直接进行深度连接, 减少模块中的参数数量。其次, 将 SE 注意力嵌入到网络模型中, 通过学习自动获得一种新的特征重新校准策略, 有效提高了模型的表达能力。最后, 将下层感知视野较大的特征提取层输出与位于中等大小目标检测层之前的特征提取层输出进行融合, 以弥补高层特征分辨率低造成的空间信息损失, 检测效果如图 3 所示。王乙涵<sup>[66]</sup>致力于实现精确且高效的柑橘识别采摘任务, 为此构建了适用于采摘机器人的轻量化目标检测模型 LT-YOLOv7, 以解决 YOLOv7 模型存储空间需求高、不适合移动终端等问题。Wang 采用 RepVGG<sup>[67]</sup>作为骨干网络, 将其得到的多尺度特征图与 YOLOv7 的颈部网络进行多尺度特征拼接, 以保留全局特征并降低整体网络的计算量。颈部网络引入深度可分离卷积, 以减少参数量、节省内存并提高模型精度。此外, 通过引入 ECA, 增强特征表示对目标的判别能力, 降低叶片、枝干等因素对目标识别的干扰。在预测阶段, 模型采用 soft DIoU\_NMS 算法进行目标预测框的筛选, 以优化对重叠物体的识别能力, 优化后的 LT-YOLOv7 模型对重叠遮挡柑橘果实平均精度达到了 96.98%, 如图 4 所示, 即使在果实被遮挡的情况下, 算法仍然能够实现良好的检测效果。



图 3 基于 YOLOv5s 改进算法的晴天和阴天检测效果

Fig. 3 Detection effect of the improved YOLOv5s algorithm on sunny and cloudy days



图 4 LT-YOLOv7 算法检测效果图

Fig. 4 Detection effect of Lt-yolov7 algorithm

Yang 等<sup>[68]</sup>针对苹果果实密度高、重叠、网络模型参数化问题，提出了 MobileOne-YOLOv7 算法。MobileOne-YOLOv7 采用多尺度特征提取方法，构建特征金字塔输入模型中。多尺度训练提高模型的鲁棒性，避免多尺度特征提取中的过多计算。Yang 等将骨干网络的最后一个 ELAN 模块替换 MobileOne 模块，增加模型的非线性和表示能力。同时，Yang 等还将 SPPCSPC 模块更改为 SPPFCSPC 模块，将串行通道变为并行通道，在保证感受野不变的情况下加快特征融合速度。此外，在颈部网络增加了一个预测头，提高了对不同尺度物体的检测精度。通过引入可重参数化的分支，训练时增加模型容量，推理时简化结构，降低内存访问成本。Zhang 等<sup>[69]</sup>提出基于 YOLOv7 改进的轻量化苹果识别算法，将多分支堆叠模块里的部分普通卷积换成 PConv，以降低模型的参数量和计算量。同时，该算法加入 ECA 改善遮挡目标错检漏检问题，保证了模型的精度平衡。在模型训练过程中，该算法采用了基于麻雀搜索算法 (Sparrow Search Algorithm, SSA)<sup>[70]</sup>的学习率优化策略，显著提高了模型的检测精度，实验表明模型平均精度达到了 97%，模型参数量和计算量分别降低了 22.93%、27.41%，适合部署在嵌入式设备中。

## 2.2 两阶段目标检测算法

Girshick 等<sup>[71]</sup>提出了 R-CNN 算法，R-CNN 利用区域建议网络提取大约 2000 个自上而下独立于类别的区域建议。通过大型 CNN 计算这些区域的固定长度特征，使用线性支持向量机 (Support Vector Machines, SVM)<sup>[72]</sup>对这些特征进行分类，确定每个区域是否包含特定的目标类别。He 等<sup>[73]</sup>提出的 SPP-Net 算法改进了 R-CNN，使其能够处理任意比例的图像。通过金字塔池化，SPP-Net 提取不同尺度的特征并整合，生

成固定长度的输出。与 R-CNN 相比, SPP-Net 无需处理所有候选区域, 只需输入整张图像即可获得特征图, 直接从中提取感兴趣区域的特征, 减少冗余计算并提高速度。Girshick 等<sup>[74]</sup>在 SPP-Net 的基础上提出了 Fast R-CNN 算法。Fast R-CNN 将整个对象建议与整张图像作为输入, 通过多个卷积和最大池化层生成特征图。通过一次卷积操作解决了多次卷积产生的冗余问题。Fast R-CNN 利用感兴趣区域 (Region Of Interest, ROI) 池化层从特征图中获取固定长度的特征向量, 然后通过全连接层进行处理, 最终分为分类和回归两个输出层。

Ren 等<sup>[75]</sup>提出的 Faster R-CNN 算法摒弃传统的选择性搜索算法, 引入了 RPN。RPN 通过滑动窗口生成不同尺寸的锚框, 并根据设定的阈值对其进行正负判断, 输出候选边界框及概率数据。这些候选区域经过 ROI 池化操作后, 被映射为固定大小的特征图, 然后通过全连接层进行物体类别判断和位置精确定位。Dai 等<sup>[76]</sup>提出了基于区域的 R-FCN 算法, 由共享的全卷积结构组成。R-FCN 生成位置敏感分数图作为输出, 编码了相对空间位置信息。其 ROI 池化层从分数图中提取信息。另一方面, Cai 等<sup>[77]</sup>提出了 Cascade R-CNN 算法, 包括提议子网络和 ROI 检测子网络。Cascade R-CNN 利用级联边界框回归将回归任务分解, 每一步骤都使用专门回归器。通过级联回归作为重采样机制, 解决初始假设分布严重偏向低质量的问题。其中, Faster R-CNN 是第一个实现端到端的基于深度学习的目标检测算法。

在两阶段目标检测算法的基础上, 研究者优化检测模型提出高效准确的类球状水果目标检测算法应用于自动化采摘作业当中, 任会等<sup>[78]</sup>利用果园内采集的柑橘果实图像, 通过实验比较传统检测算法和 Faster R-CNN 对柑橘果实的识别效果, 实验发现传统检测算法在增强预处理且果实无遮挡的情况下, 识别效果要优于 Faster R-CNN, 但当果实重叠或遮挡, 则 Faster R-CNN 更优。Wan 等<sup>[79]</sup>提出了一种基于 Faster R-CNN 改进的多类水果检测框架。骨干网络为 VGG-16 包含 13 卷积层、13 个 ReLu 层和 4 个池化层。为避免样本较少训练出现过拟合问题, 平衡模型复杂度和数据量, 该算法使用正则化对高位参数进行权值衰减, 增加两个损失函数优化卷积层和池化层参数根据拍摄角度自动调整保证每个卷积层大小以及核参数的合理, 提高检测精度。Liu 等<sup>[80]</sup>提出基于 R-FCN 改进的水果识别定位算法, 由 RPN 和 FCN 组成。RPN 用于生成候选区域框, FCN 进行像素级特征提取, 通过反卷积可视化检测结果。黄磊磊等<sup>[81]</sup>为解决算法识别遮挡重叠柑橘果实精度低的问题提出基于深度学习的重叠柑橘分割与形态复原算法, 引入 Pointend 分支的 Mask R-CNN 实现对柑橘识别及边缘细化的实例分割, 采用编解码器结构的 U-Net 作为主体网络提出形态粗复原模型, 设计局部惩罚损失函数及交并比形状损失函数, 通过机器视觉方法根据粗复原结果提取 ROI, 最后利用基于 PConv 的形态精复原模型完成果实的形态复原。采用该方法对柑橘果实识别的平均精度达 93.66%, 分割精确度达 96.30%。荆伟斌等<sup>[82]</sup>针对苹果园果实产量预估提出了一种基于不同特征网络的苹果树侧面果实识别方法。研究人员通过采集果园内苹果树的侧视图, 测试不同特征提取网络与 Faster R-CNN 模型结合的识别效果。实验中, 分别选用了 VGG-16 和 ResNet-50 作为特征提取网络, 对两个 Faster R-CNN 模型进行训练。结果显示, 虽然两者使用了相同的学习参数, VGG-16 作为特征网络的 Faster R-CNN 模型在各项指标上优于 ResNet-50, 识别精度达 91%, 单幅图像的推理时间为 1.4 秒。贾艳平等<sup>[83]</sup>利用相机采集自然环境中不同水果的 RGB 图像, 在 Faster R-CNN 添加似然函数和正则化函数保证卷积层的大小和核参数在合理范围内, 对不同水果进行识别测试, 整体识别准确率达 99.69%, 对橙子的识别准确率为 77.3%。Lu<sup>[84]</sup>

利用相机采集果园内绿色柑橘果实图像，采用深浅层特征融合策略增加 Mask R-CNN 骨干网络每易阶段提取的特征信息，通过引入骨干网络之间的组合连接块减少通道数并提高模型精度，改进后的 Mask R-CNN 在识别绿色柑橘果实的平均精度达 95.36%，比原模型提高了 1.42%。Min 等<sup>[85]</sup>为了聚合 CNN 不同层次的注意力特征，设计了多尺度注意力网络（Multi Scale Attention Network, MSANet）。MSANet 引入混合注意力机制，能将空间通道注意力和不同层的多个注意力特征聚合到最终的统一表示，使最终表示更加鲁棒、全面。

### 3 总结及未来发展趋势

#### 3.1 总结

本文对水果采摘领域中表现优异的检测算法进行了综述，重点分析了传统目标检测算法和基于深度学习的目标检测算法。针对类球状水果识别任务的传统目标检测算法依赖手工设计特征，通过明确的规则提取，使得算法的各个步骤具备高度的可解释性。传统算法对数据需求较少，仅需少量标记数据就能实现模型的调试和优化，且无需复杂的深度神经网络运算，计算复杂度相对较低，对计算资源的要求不高。然而，在自然环境中，传统水果检测算法在处理果实重叠、光照变化和枝条遮挡等复杂场景时，往往难以准确地提取有效信息。在更换识别水果种类时，可能需要人工更改算法，缺乏良好的泛化能力。

相比之下，基于深度学习的目标检测算法采用多层级的神经网络架构，研究人员可以通过对网络内模块的调整，增强特征表达能力、减少模型参数量和提升图像推理速度等。神经网络架构的优势在于，能通过大规模数据的学习，自主提取复杂且抽象的多层次特征，并通过层次化特征学习逐步捕捉从低级到高级的语义信息，提高模型在面对复杂目标和多变环境时的检测精度和鲁棒性。

在类球状水果的识别任务中，研究人员通过引入轻量级特征提取网络（如 MobileNet、GhostNet）和不同注意力机制模块（如 SE、CA 和 CBAM），降低计算成本和内存占用，使得这些算法更适合在资源有限的嵌入式设备上运行。同时，这些改进增强了模型对关键特征的敏感性，使网络能够更精准地捕捉到目标对象的关键特征，抑制一般特征信息，从而提高检测精度，提升算法在复杂环境中的表现，解决由于背景复杂性导致的误检和漏检问题。此外，研究人员还运用多种先进的损失函数（如 SIOU、Alpha-IoU 损失等），平衡了正负样本的影响，提高边界框的回归精度。

#### 3.2 发展趋势

近年来，目标检测算法在类球状水果识别任务方面有广泛应用，在遮挡重叠果实、产量预测、水果分类分级和表面缺陷等复杂任务展现出优越的性能。但是，由于果园环境条件复杂多变，现有的类球状水果目标检测算法识别能力的普适性仍有待提高。根据类球状水果目标检测算法的发展趋势分析，未来的研究可以重点集中在以下几个方向：

- （1） 模型优化：基于深度学习的目标检测算法需要根据果实识别需求不断改进，可以通过引入注意力机制、改变特征提取网络结构、优化损失函数和调整网络深度宽度等方法，提高果实目标识别的准确率、加快识别速度以及降低漏检误检率。参考其他领域大模型，研究领域内表现优异的模型是否可以经过调整用于果实目标识别，进一步提高识别准确率和效率。
- （2） 数据集构建与扩增：根据不同采摘任务需求，收集各个生长阶段和品种的一类球状水果图像，构建

一个包含不同天气、光照条件（顺光、逆光）、果实重叠程度以及遮挡情况的数据集。结合图像处理方法（如图像旋转、翻转、裁剪、缩放、加噪声、色彩变换等）或生成对抗网络（Generative Adversarial Network, GAN）<sup>[86]</sup>合成图像生成技术进行数据扩增。在多样化的数据集上进行训练，可以增强模型的泛化能力和鲁棒性。

- (3) 多模态数据结合: 为了进一步提升类球状水果识别的精度与适应性, 未来研究可以结合激光雷达、深度相机获取的三维信息, 更全面地获取果实形态和位置信息, 特别是在果实被严重遮挡或光照条件极差的情况下, 多模态数据有助于增强模型的鲁棒性。

### 参考文献 References:

- [1] 刘袁, 黄彪, 陈昌银, 杨文达, 张华东, 杨涛. 水果采摘机器人采摘装置机研究现状[J]. 农业科学, 2021, 11(2): 129-132. LIU Yuan, HUANG Biao, CHEN Changyin, YANG Wenda, ZHANG Huadong, YANG Tao. Research status of fruit picking robot picking device[J]. Journal of Agricultural Sciences, 2021, 11(2):129-132.
- [2] 戴军. 机器视觉技术在瓜菜检测应用中的研究进展[J]. 中国瓜菜, 2023, 36(11): 1-9. DAI Jun. Research progress of machine vision technology in the detection of cucurbits and vegetables[J]. China Cucurbits and Vegetables, 2023, 36(11): 1-9.
- [3] 吴剑桥, 范圣哲, 贡亮, 苑进, 周强, 刘成良. 果蔬采摘机器人系统设计与控制技术研究现状和发展趋势[J]. 智慧农业 (中英文), 2020, 2(4): 17-40. WU Jianqiao, FAN Shengzhe, GONG Liang, YUAN Jin, ZHOU Qiang, LIU Chengliang. Research status and development direction of design and control technology of fruit and vegetable picking robot system[J]. Smart Agriculture, 2020, 2(4): 17-40.
- [4] 初广丽, 张伟, 王延杰, 丁南南, 刘艳滢. 基于机器视觉的水果采摘机器人目标识别方法[J]. 中国农机化学报, 2018, 39(2): 83-88. CHU Guangli, ZHANG Wei, WANG Yanjie, DING Nannan, LIU Yanying. A method of fruit picking robot target identification based on machine vision[J]. Journal of Chinese Agricultural Mechanization, 2018, 39(2): 83-88.
- [5] 杨健, 杨啸治, 熊串, 刘力. 基于改进 YOLOv5 的番茄果实识别估产方法[J]. 中国瓜菜, 2024, 37(6): 61-68. YANG Jian, YANG Xiaozhi, XIONG Chuan, LIU Li. An improved YOLOv5-based method for tomato fruit identification and yield estimation[J]. China Cucurbits and Vegetables, 2024, 37(6): 61-68.
- [6] DENG J, XUAN X J, WANG W F, LI Z, YAO H W, WANG Z Q. A review of research on object detection based on deep learning[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1684(1): 012028.
- [7] DU L X, ZHANG R Y, WANG X T. Overview of two-stage object detection algorithms[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1544(1): 012033.
- [8] 蒋焕煜, 彭永石, 申川, 应义斌. 基于双目立体视觉技术的成熟番茄识别与定位[J]. 农业工程学报, 2008, 24(8): 279-283. JIANG Huanyu, PENG Yongshi, SHEN Chuan, YING Yibin. Recognizing and locating ripe tomatoes based on binocular stereovision technology[J]. Transactions of the Chinese Society of Agricultural Engineering, 2008, 24(8): 279-283.
- [9] LIU X Y, ZHAO D A, JIA W K, JI W, SUN Y P. A detection method for apple fruits based on color and shape features[J]. IEEE Access, 2019, 7: 67923-67933.
- [10] 夏康利, 何强. 基于颜色统计的水果采摘机器人水果识别的研究[J]. 南方农机, 2022, 53(24): 11-16. XIA Kangli, HE Qiang. Research on fruit recognition for fruit-picking robots based on color statistics[J]. China Southern Agricultural Machinery, 2022, 53(24): 11-16.
- [11] COMANICIU D, MEER P. Mean shift: A robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.

- [12] 邹伟. 基于机器视觉技术的柑橘果实成熟度分选研究[J]. 农业与技术, 2023, 43(17): 41-44.  
ZOU Wei. Research on citrus fruit maturity sorting based on machine vision technology[J]. Agriculture and Technology, 2023, 43(17): 41-44.
- [13] 陈雪鑫, 卜庆凯. 基于多颜色和局部纹理的水果识别算法研究[J]. 青岛大学学报(工程技术版), 2019, 34(3): 52-58.  
CHEN Xuexin, BU Qingkai. Research on fruit recognition algorithm based on multi-color and local texture[J]. Journal of Qingdao University (Engineering & Technology Edition), 2019, 34(3): 52-58.
- [14] 徐惠荣, 叶尊忠, 应义斌. 基于彩色信息的树上柑橘识别研究[J]. 农业工程学报, 2005, 21(5): 98-101.  
XU Huirong, YE Zunzhong, YING Yibin. Identification of citrus fruit in a tree canopy using color information[J]. Transactions of the Chinese Society of Agricultural Engineering, 2005, 21(5): 98-101.
- [15] FAN Q, ZHUO W, TANG C K, TAI Y W. Few-shot object detection with attention-RPN and multi-relation detector[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 4012-4021.
- [16] LIU W, ANGUELOV D, ERHAN D, SZEGEDY C, REED S, FU C Y, BERG A C. SSD: Single shot MultiBox detector[C]//Computer Vision—ECCV 2016: 14th European Conference. Amsterdam, The Netherlands: Springer International Publishing, 2016: 21-37.
- [17] HOWARD A G, ZHU M L, CHEN B, KALENICHENKO D, WANG W J, WEYAND T, ANDREETTO M, ADAM H, HEATON J. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. 2017: 1704.04861. <https://arxiv.org/abs/1704.04861v1>.
- [18] ZHANG X Y, ZHOU X Y, LIN M X, SUN J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6848-6856.
- [19] LIU Z, LIN Y T, CAO Y, HU H, WEI Y X, ZHANG Z, LIN S, GUO B N. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2021: 9992-10002.
- [20] REDMON J, DIVVALA S, GIRSHICK R, FARHADI A. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [21] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6517-6525.
- [22] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. 2018: 1804.02767. <https://arxiv.org/abs/1804.02767v1>
- [23] LIN T Y, DOLLÁR P, GIRSHICK R, HE K M, HARIHARAN B, BELONGIE S. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944.
- [24] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. 2020: 2004.10934. <https://arxiv.org/abs/2004.10934v1>
- [25] JOCHER G. YOLOv5byUltralytics (Version7.0) Computersoftware[CP]. <https://doi.org/10.5281/zenodo.3908559>, 2020.
- [26] LI C Y, LI L L, JIANG H L, WENG K H, GENG Y F, LI L, KE Z D, LI Q Y, CHENG M, NIE W Q, LI Y D, ZHANG B, LIANG Y F, ZHOU L Y, XU X M, CHU X X, WEI X M, WEI X L. YOLOv6: A single-stage object detection framework for industrial applications[EB/OL]. arxiv preprint arxiv, 2022: 2209.02976.
- [27] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 7464-7475.
- [28] VARGHESE R, M S. YOLOv8: A novel object detection algorithm with enhanced performance and robustness[C]//2024

- International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India: IEEE, 2024: 1-6.
- [29] 宋中山, 刘越, 郑禄, 帖军, 汪进. 基于改进 YOLOV3 的自然环境下绿色柑橘的识别算法[J]. 中国农机化学报, 2021, 42(11): 159-165.
- SONG Zhongshan, LIU Yue, ZHENG Lu, TIE Jun, WANG Jin. Identification of green citrus based on improved YOLOV3 in natural environment[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(11): 159-165.
- [30] HUANG G, LIU Z, VAN DER MAATEN L, WEINBERGER K Q. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2261-2269.
- [31] 吕强, 林刚, 蒋杰, 王明之, 张皓杨, 易时来. 基于改进 YOLOv5s 模型的自然场景中绿色柑橘果实检测[J]. 农业工程学报, 2024, 40(18): 147-154.
- LÜ Qiang, LIN Gang, JIANG Jie, WANG Mingzhi, ZHANG Haoyang, YI Shilai. Detecting green citrus fruit in natural scenes using improved YOLOv5s model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(18): 147-154.
- [32] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions[EB/OL]. 2021: 2112.05561. <https://arxiv.org/abs/2112.05561v1>
- [33] BODLA N, SINGH B, CHELLAPPA R, DAVIS L S. Soft-NMS—improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 5561-5569.
- [34] 帖军, 赵捷, 郑禄, 吴立锋, 洪博文. 改进 YOLOv5 模型在自然环境下柑橘识别的应用[J]. 中国农业科技导报, 2024, 26(7): 111-120.
- TIE Jun, ZHAO Jie, ZHENG Lu, WU Lifeng, HONG Bowen. Application of improved YOLOv5 model in citrus recognition in natural environment[J]. Journal of Agricultural Science and Technology, 2024, 26(7): 111-120.
- [35] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 7132-7141.
- [36] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 13708-13717.
- [37] ZHANG H Y, WANG Y, DAYOUB F, SÜNDERHAUF N. VarifocalNet: An IoU-aware dense object detector[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 8510-8519.
- [38] 马帅, 张艳, 周桂红, 刘博. 基于改进 YOLOv4 模型的自然环境下梨果实识别[J]. 河北农业大学学报, 2022, 45(3): 105-111.
- MA Shuai, ZHANG Yan, ZHOU Guihong, LIU Bo. Recognition of pear fruit under natural environment using an improved YOLOv4 model[J]. Journal of Hebei Agricultural University, 2022, 45(3): 105-111.
- [39] 刘忠意, 魏登峰, 李萌, 周绍发, 鲁力, 董雨雪. 基于改进 YOLO v5 的橙子果实识别方法[J]. 江苏农业科学, 2023, 51(19): 173-181.
- LIU Zhongyi, WEI Dengfen, LI Meng, ZHOU Shaofa, LU Li, DONG Yuxue. Orange fruit recognition method based on improved YOLO v5[J]. Jiangsu Agricultural Sciences, 2023, 51(19): 173-181.
- [40] HAN K, WANG Y H, TIAN Q, GUO J Y, XU C J, XU C. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1577-1586.
- [41] WANG Q L, WU B G, ZHU P F, LI P H, ZUO W M, HU Q H. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 11531-11539.
- [42] 贺英豪, 唐德钊, 倪铭, 蔡起起. 基于改进 YOLOv5 对果园环境中李的识别[J]. 华中农业大学学报, 2024, 43(5): 31-40.
- HE Yinghao, TANG Dezhao, NI Ming, CAI Qiqi. Recognizing plums in orchard environment based on improved YOLOv5[J]. Journal of Huazhong Agricultural University, 2024, 43(5): 31-40.

- [43] 黄彤斌, 黄河清, 李震, 吕石磊, 薛秀云, 代秋芳, 温威. 基于 YOLOv5 改进模型的柑橘果实识别方法[J]. 华中农业大学学报, 2022, 41(4): 170-177.
- HUANG Tongbin, HUANG Heqing, LI Zhen, LÜ Shilei, XUE Xiuyun, DAI Qiufang, WEN Wei. Citrus fruit recognition method based on the improved model of YOLOv5[J]. Journal of Huazhong Agricultural University, 2022, 41(4): 170-177.
- [44] WOO S, PARK J, LEE J Y, KWEON I S. CBAM: Convolutional block attention module[EB/OL]. 2018: 1807.06521. <https://arxiv.org/abs/1807.06521v2>
- [45] HE J B, ERFANI S, MA X J, BAILEY J, CHI Y, HUA X S. Alpha-IoU: A family of power intersection over union losses for bounding box regression[EB/OL]. 2021: 2110.13675. <https://arxiv.org/abs/2110.13675v2>
- [46] 苑迎春, 张傲, 何振学, 张若晨, 雷浩. 基于改进 YOLOv4-tiny 的果园复杂环境下桃果实实时识别[J]. 中国农机化学报, 2024, 45(8): 254-261.
- YUAN Yingchun, ZHANG Ao, HE Zhenxue, ZHANG Ruochen, LEI Hao. Peach fruit real-time recognition in complex orchard environment based on improved YOLOv4-tiny[J]. Journal of Chinese Agricultural Mechanization, 2024, 45(8): 254-261.
- [47] ZHU L, DENG Z J, HU X W, FU C W, XU X M, QIN J, HENG P A. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 122-136.
- [48] 熊俊涛, 郑镇辉, 梁嘉恩, 钟灼, 刘柏林, 孙宝霞. 基于改进 YOLO v3 网络的夜间环境柑橘识别方法[J]. 农业机械学报, 2020, 51(4): 199-206.
- XIONG Juntao, ZHENG Zhenhui, LIANG Jia'en, ZHONG Zhuo, LIU Bolin, SUN Baoxia. Citrus detection method in night environment based on improved YOLO v3 network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(4): 199-206.
- [49] HE K M, ZHANG X Y, REN S Q, SUN J. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [50] 熊俊涛, 霍钊威, 黄启寅, 陈浩然, 杨振刚, 黄煜华, 苏颖苗. 结合主动光源和改进 YOLOv5s 模型的夜间柑橘检测方法[J]. 华南农业大学学报, 2024, 45(1): 97-107.
- XIONG Juntao, HUO Zhaowei, HUANG Qiyin, CHEN Haoran, YANG Zhengang, HUANG Yuhua, SU Yingmiao. Detection method of citrus in nighttime environment combined with active light source and improved YOLOv5s model[J]. Journal of South China Agricultural University, 2024, 45(1): 97-107.
- [51] 余圣新, 韦莹莹, 方辉, 李敏, 柴秀娟, 曾志康, 覃泽林. 基于改进 YOLOv8 的自然环境下柑橘果实识别[J]. 湖北农业科学, 2024, 63(8): 23-27.
- YU Shengxin, WEI Yingying, FANG Hui, LI Min, CHAI Xiujian, ZENG Zhikang, QIN Zelin. Citrus fruit recognition in natural environment based on improved YOLOv8[J]. Hubei Agricultural Sciences, 2024, 63(8): 23-27.
- [52] MA S L, XU Y, MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[EB/OL]. 2023: 2307.07662. <https://arxiv.org/abs/2307.07662v1>
- [53] 岳有军, 漆潇, 赵辉, 王红君. 基于改进 YOLOv8 的果园复杂环境下苹果检测模型研究[J/OL]. 南京信息工程大学学报, 2024: 1-13(2024-07-15). <https://doi.org/10.13878/j.cnki.jnuist.20240410002>.
- YUE Youjun, QI Xiao, ZHAO Hui, WANG Hongjun. Research on apple detection model in complex orchard environments based on improved YOLOv8[J]. Journal of Nanjing University of Information Science & Technology, 2024: 1-13(2024-07-15). <https://doi.org/10.13878/j.cnki.jnuist.20240410002>.
- [54] ZHANG H, XU C, ZHANG S J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[EB/OL]. 2023: 2311.02877. <https://arxiv.org/abs/2311.02877v4>
- [55] 吕石磊, 卢思华, 李震, 洪添胜, 薛月菊, 吴奔雷. 基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法[J]. 农业工程学报, 2019, 35(17): 205-214.



- LÜ Shilei, LU Sihua, LI Zhen, HONG Tiansheng, XUE Yueju, WU Benlei. Orange recognition method using improved YOLOv3-LITE lightweight neural network[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(17): 205-214.
- [56] REZATOFIGHI H, TSOI N, GWAK J, SADEGHIAN A, REID I, SAVARESE S. Generalized intersection over union: A metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 658-666.
- [57] 王卓, 王健, 王梟雄, 时佳, 白晓平, 赵泳嘉. 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 Zhuo, WANG Jian, WANG Xiaoxiong, SHI Jia, BAI Xiaoping, ZHAO Yongjia. Lightweight real-time apple detection method based on improved YOLO v4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8): 294-302.
- [58] 曾俊, 陈仁凡, 邹腾跃. 基于改进 YOLO 的自然环境下桃子成熟度快速检测模型[J]. 南方农机, 2023, 54(24): 24-27.
- ZENG Jun, CHEN Renfan, ZOU Tengyue. Rapid maturity detection model for peaches in natural environment based on improved YOLO[J]. China Southern Agricultural Machinery, 2023, 54(24): 24-27.
- [59] CHEN J R, KAO S H, HE H, ZHUO W P, WEN S, LEE C H, CHAN S H G. Run, Don't walk: Chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 12021-12031.
- [60] LIU G L, REDA F A, SHIH K J, WANG T C, TAO A, CATANZARO B. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European conference on computer vision (ECCV). Cham: Springer International Publishing, 2018: 85-100.
- [61] GEVORGYAN Z. Siou loss : More powerful learning for bounding box regression[EB/OL]. 2022 : 2205.12740. <https://arxiv.org/abs/2205.12740v1>
- [62] 赵辉, 乔艳军, 王红君, 岳有军. 基于改进 YOLOv3 的果园复杂环境下苹果果实识别[J]. 农业工程学报, 2021, 37(16): 127-135.
- ZHAO Hui, QIAO Yanjun, WANG Hongjun, YUE Youjun. Apple fruit recognition in complex orchard environment based on improved YOLOv3[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(16): 127-135.
- [63] WANG C Y, MARK LIAO H Y, WU Y H, CHEN P Y, HSIEH J W, YEH I H. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 1571-1580.
- [64] BODLA N, SINGH B, CHELLAPPA R, DAVIS L S. Soft-NMS—improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 5561-5569.
- [65] YAN B, FAN P, LEI X Y, LIU Z J, YANG F Z. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing, 2021, 13(9): 1619.
- [66] 王乙涵. 基于改进 YOLOv7 的自然环境下柑橘果实识别与定位方法研究[D]. 雅安: 四川农业大学, 2023.
- WANG Yihan. Research on detection and localization of citrus in natural environment based on improved YOLOv7[D]. Ya'an: Sichuan agricultural university, 2023.
- [67] DING X H, ZHANG X Y, MA N N, HAN J G, DING G G, SUN J. RepVGG: Making VGG-style ConvNets great again[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 13728-13737.
- [68] YANG H W, LIU Y Z, WANG S W, QU H X, LI N, WU J, YAN Y F, ZHANG H J, WANG J X, QIU J F. Improved apple fruit target recognition method based on YOLOv7 model[J]. Agriculture, 2023, 13(7): 1278.
- [69] 张震, 周俊, 江自真, 韩宏琪. 基于改进 YOLO v7 轻量化模型的自然果园环境下苹果识别方法[J]. 农业机械学报, 2024, 55(3): 231-242.
- ZHANG Zhen, ZHOU Jun, JIANG Zizhen, HAN Hongqi. Lightweight apple recognition method in natural orchard environment

- based on improved YOLO v7 model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 231-242.
- [70] XUE J K, SHEN B. A novel swarm intelligence optimization approach: Sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.
- [71] GIRSHICK R, DONAHUE J, DARRELL T, MALIK J. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [72] HEARST M A, DUMAIS S T, OSUNA E, PLATT J, SCHOLKOPF B. Support vector machines[J]. IEEE Intelligent Systems and Their Applications, 1998, 13(4): 18-28.
- [73] HE K M, ZHANG X Y, REN S Q, SUN J. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [74] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [75] REN S Q, HE K M, GIRSHICK R, SUN J. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2015, 28.
- [76] DAI J F, LI Y, HE K M, SUN J. R-FCN: Object detection via region-based fully convolutional networks[J]. Advances in Neural Information Processing Systems, 2016, 29.
- [77] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 6154-6162.
- [78] 任会, 朱洪前. 基于深度学习的目标橘子识别方法研究[J]. 计算机时代, 2021(1): 57-60.  
REN Hui, ZHU Hongqian. Research on the method of identifying target orange with deep learning[J]. Computer Era, 2021(1): 57-60.
- [79] WAN S H, GOUDOS S. Faster R-CNN for multi-class fruit detection using a robotic vision system[J]. Computer Networks, 2020, 168: 107036.
- [80] LIU J, ZHAO M R, GUO X F. A fruit detection algorithm based on R-FCN in natural scene[C]//2020 Chinese Control and Decision Conference (CCDC). Hefei: IEEE, 2020: 487-492.
- [81] 黄磊磊, 苗玉彬. 基于深度学习的重叠柑橘分割与形态复原[J]. 农机化研究, 2023, 45(10): 70-75.  
HUANG Leilei, MIAO Yubin. Overlapping citrus segmentation and morphological restoration based on deep learning[J]. Journal of Agricultural Mechanization Research, 2023, 45(10): 70-75.
- [82] 荆伟斌, 李存军, 竞霞, 赵叶, 程成. 基于深度学习的苹果树侧视图果实识别[J]. 中国农业信息, 2019, 31(5): 75-83.  
JING Weibin, LI Cunjun, JING Xia, ZHAO Ye, CHENG Cheng. Fruit identification with apple tree side view based on deep learning[J]. China Agricultural Informatics, 2019, 31(5): 75-83.
- [83] 贾艳平, 桑妍丽, 李月茹. 基于改进 Faster R-CNN 模型的水果分类识别[J]. 食品与机械, 2023, 39(8): 129-135.  
JIA Yanping, SANG Yanli, LI Yueru. Fruit identification using improved Faster R-CNN model[J]. Food & Machinery, 2023, 39(8): 129-135.
- [84] LU J Q, YANG R F, YU C R, LIN J H, CHEN W D, WU H W, CHEN X, LAN Y B, WANG W X. Citrus green fruit detection via improved feature network extraction[J]. Frontiers in Plant Science, 2022, 13: 946154.
- [85] MIN W Q, WANG Z L, YANG J H, LIU C L, JIANG S Q. Vision-based fruit recognition via multi-scale attention CNN[J]. Computers and Electronics in Agriculture, 2023, 210: 107911.
- [86] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, XU B, WARDE-FARLEY D, OZAIR S, COURVILLE A, BENGIO Y. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [87] 戈明辉, 张俊, 陆慧娟. 基于机器视觉的食品外包装缺陷检测算法研究进展[J]. 食品与机械, 2023, 39(9): 95-102.  
GE Minghui, ZHANG Jun, LU Huijuan. Research progress of food packaging defect detection based on machine vision[J]. Food & Machinery, 2023, 39(9): 95-102.

- [88] 任磊, 张俊, 陆胜民. 脱囊衣橘片自动分拣机器视觉算法研究[J]. 浙江农业学报, 2015, 27(12): 2212-2217.
- REN Lei, ZHANG Jun, LU Shengmin. Research on machine vision algorithm for automatic sorting of membrane-removed mandarin segments[J]. Acta Agriculturae Zhejiangensis, 2015, 27(12): 2212-2217.