

分类号: TP181

密 级: 公开

学校代码: 10363

学 号: 2210120158



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于深度学习的草莓检测
与定位方法研究

论文作者 彭勇

校内导师 乔印虎

校外导师 贡军

专业学位类别 机械

研究方向(领域) 智能农业装备

论文提交日期: 2024 年 06 月 02 日

分类号：TP181
密 级：公开

单位代码：10363
学 号：2210120158

基于深度学习的草莓检测与定位方法研究

Research on Strawberry Detection and Localization Method Based on Deep Learning

论文作者：	彭勇
指导教师：	乔印虎（教授）
校外教师：	贡军（高级工程师）
学科（专业）：	机械
研究方向：	智能农业装备

论文答辩日期：2024 年 06 月 02 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：彭勇

日期：2024 年 06 月 02 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 _____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：



指导教师签名：



日期： 2024 年 06 月 02 日

日期： 2024 年 06 月 02 日

基于深度学习的草莓检测与定位方法研究

摘要

草莓具有较高的营养价值、医疗价值和生态价值。草莓浆果芳香多汁，营养丰富，素有“水果皇后”的美称，又是果树中上市最早的鲜果，素有“早春第一果”的美称。中国是草莓最大的消费地区之一，其产量在持续增加，于是对草莓的品质要求和采摘效率也在不断提高，而采摘时对目标进行准确地检测和定位是保证草莓品质和效率的最重要的一个环节。针对人工采摘劳动强度大、成本高、采摘效率低，现有的草莓采摘设备产品成本较高、检测识别精度偏低、检测模型在边缘设备中识别效率较低等问题，本文提出了基于深度学习的草莓检测与定位方法研究，主要内容有：

(1) 草莓数据集的收集与制作。草莓图片来自 StrawDI 收集的不同生长周期的草莓，共 3100 张。将收集到的草莓图片进行标注，按照农业行业标准，将其分为草莓花、未成熟草莓和成熟草莓三类，并划分出训练集、验证集和测试集，得到草莓目标检测数据集。

(2) 基于深度学习的草莓检测算法研究。本文在深度学习目标检测算法 YOLOv5s 的基础上，针对算法对草莓的检测精度有待进一步提高、参数量大、模型复杂度较高、复杂环境下的草莓检测效果不佳这些问题，首先通过 Ghost 结构来减少模型参数量，降低复杂度。之后使用 SIOU 作为定位损失函数，不仅考虑了边界框回归时的重叠面积、中心点距离和纵横比，而且增加了边界框回归之间的向量角度约束，提升了模型在复杂环境下对草莓的检测效果。然后通过添加高效通道注意力机制 ECA，由高权重来聚焦草莓特征，用低

权重忽略背景信息，达到对草莓特征信息的关注，并使用 SimSPPF 特征融合模块提升模型对不同尺度特征的融合能力，从而提高模型对草莓的检测精度。通过测试集进行测试，实验结果表明：改进后的 YOLOv5s 模型对草莓的检测精度为 94.9%，模型大小仅为 7.76MB，与原始 YOLOv5s 相比，具有更高的检测精度和更低的模型规模，对复杂环境下的草莓，例如遮挡、重叠等，检测效果也更为优越。

(3) 草莓定位方法的研究。首先通过 Halcon 软件，通过图像处理一系列方法实现了对成熟草莓的二维定位。构建了草莓的三维定位模型，推导了草莓像素坐标与世界坐标的转换关系，使用张正友标定法，对 RGB-D 相机和被动双目相机进行标定，获取相机参数，并对被动双目相机进行了立体矫正与立体匹配，对主动双目相机进行了图像对齐，并通过 python 程序，将改进 YOLOv5s 检测方法与相机定位方法结合，检测类别的同时获得了目标的三维坐标，实现了三维定位。实验结果表明：相比于被动双目相机，RGB-D 相机计算出的深度与实际深度的平均误差仅为 $\pm 3.75\text{mm}$ ，具有更为优秀的定位性能。

(4) 草莓检测系统设计。基于 Pycharm 集成开发平台，采用本文改进的 YOLOv5s 模型，设计了可视化用户界面，搭建了草莓检测系统，并对系统界面和各个功能模块进行了相应的分析，结果表明：该系统操作简单，使用便捷，稳定性较高，具有较高的实际应用价值。

(5) 草莓检测模型的边缘部署。为了实现草莓检测模型的实际应用，通过 TensorRT 加速技术，实现了对检测模型的加速推理，使其部署到低算力的边缘设备中具有更快的检测速度。并以 Jetson

nano 为实验平台，进行草莓检测模型的边缘部署，实验结果表明：使用 TensorRT 加速推理后的模型，相较于未使用的模型，速度提升大约 51.1%，较大幅度提升了模型部署到边缘设备的检测速度。

关键字：草莓，深度学习，目标检测，定位，边缘部署

RESEARCH ON STRAWBERRY DETECTION AND LOCALIZATION METHOD BASED ON DEEP LEARNING

ABSTRACT

Strawberry has high nutritional value, medical value and ecological value. Strawberry berries are aromatic, juicy and nutritious, known as the "Queen of Fruits", and the earliest fresh fruit on the market in the fruit tree, known as the "first fruit in early spring". China is one of the largest consumers of strawberries, and its production continues to increase, so the quality of strawberries and picking efficiency is also increasing, and the accurate detection and positioning of the target during picking is the most important part of ensuring the quality and efficiency of strawberries. Aiming at the problems of high labor intensity, high cost and low picking efficiency of manual picking, high cost of existing strawberry picking equipment products, low detection and recognition accuracy, and low recognition efficiency of the detection model in the edge equipment, this paper proposes the research of strawberry detection and localization method based on deep learning, and the main contents of the research are:

(1) Strawberry dataset collection and production. Strawberry pictures are from strawberries with different growth cycles collected by StrawDI, totaling 3100 pictures. The collected strawberry pictures were labeled and classified into three categories of strawberry flowers, immature strawberries and ripe strawberries according to the agricultural industry standard, and the training set, validation set and test set were divided to obtain the strawberry object detection dataset.

(2) Research on strawberry detection algorithm based on deep

learning. In this paper, based on the deep learning object detection algorithm YOLOv5s, for the problems that the algorithm's detection accuracy of strawberry needs to be further improved, the number of parameters is large, the model complexity is high, and the effect of strawberry detection in complex environments is not good, we firstly reduce the number of parameters of the model through the Ghost structure to reduce the complexity. After that, SIOU is used as the localization loss function, which not only takes into account the overlap area, centroid distance and aspect ratio in the bounding box regression, but also increases the vector angle constraints between the bounding box regressions, which improves the model's detection effect of strawberries in complex environments. Then by adding the efficient channel attention mechanism ECA, the strawberry features are focused by high weights and the background information is ignored with low weights to achieve the attention to the strawberry feature information, and the SimSPPF feature fusion module is used to enhance the model's fusion ability for different scale features, so as to improve the model's detection accuracy for strawberries. Tests are conducted through the test set, and the experimental results show that the improved YOLOv5s model has a detection accuracy of 94.9% for strawberries, and the model size is only 7.76MB, which has higher detection accuracy and lower model size compared with the original YOLOv5s, and also has more superior detection effect for strawberries in complex environments, such as occlusion and overlapping.

(3) Research on strawberry localization methods. Firstly, the two-dimensional localization of ripe strawberries was realized by a series of methods of image processing through Halcon software. The three-dimensional localization model of strawberry was constructed, and the

conversion relationship between the pixel coordinates of strawberry and the world coordinates was deduced. Using the Zhang Zhengyou calibration method, the RGB-D camera and the passive binocular camera were calibrated to obtain the camera parameters, and stereo rectification and stereo matching were carried out on the passive binocular camera, and the active binocular camera was image-aligned, and through the python program, the improved YOLOv5s detection method was combined with the camera localization method, and the 3D coordinates of the object were obtained while detecting the categories to achieve 3D localization. The experimental results show that compared with the passive binocular camera, the average error between the calculated depth and the actual depth of the RGB-D camera is only $\pm 3.75\text{mm}$, which has more excellent localization performance.

(4) Strawberry detection system design. Based on Pycharm integrated development platform, using the improved YOLOv5s model in this paper, a visual user interface was designed to build a strawberry detection system, and the system interface and each functional module were analyzed accordingly.

(5) Edge deployment of strawberry detection model. In order to realize the practical application of the strawberry detection model, accelerated reasoning of the detection model is realized by TensorRT acceleration technology, so that it can be deployed to edge devices with low computing power with faster detection speed. And Jetson Nano is used as the experimental platform for the edge deployment of the strawberry detection model, and the experimental results show that the speed of the model after using TensorRT accelerated inference is improved by about 51.1% compared with that of the unused model, which improves the detection speed of the model deployed to the edge device

more substantially.

KEY WORDS: Strawberry, deep learning, object detection, localization, edge deployment

目录

摘要	I
ABSTRACT	IV
第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 研究现状	3
1.2.1 果实目标检测国内外研究现状	3
1.2.2 果实定位国内外研究现状	5
1.3 目前存在的问题	6
1.4 论文主要的研究内容	7
1.5 论文技术路线和章节安排	8
1.5.1 技术路线	8
1.5.2 章节安排	9
1.6 本章小结	10
第 2 章 基于深度学习的草莓检测相关理论	11
2.1 卷积神经网络	11
2.1.1 卷积层	11
2.1.2 池化层	12
2.1.3 全连接层	13
2.2 经典目标检测算法	13
2.2.1 双阶段目标检测算法	13
2.2.2 单阶段目标检测算法	15
2.3 本章小结	17
第 3 章 改进 YOLOv5 草莓检测模型研究	18
3.1 草莓目标检测数据集采集与制作	18
3.1.1 数据集采集与制作	18
3.1.2 草莓目标检测数据集展示	19
3.2 YOLOv5 算法	21
3.2.1 YOLOv5 网络结构	21
3.2.2 YOLOv5 损失函数	22
3.3 相关模块及改进方法	22
3.3.1 ECA 注意力机制	22
3.3.2 Ghost 模型	23
3.3.3 SIoU 损失函数	24
3.3.4 SimSPPF 特征融合模块	27
3.3.5 改进的 YOLOv5 模型	27
3.4 实验结果与分析	28
3.4.1 训练策略和实验平台	28
3.4.2 评价指标	29
3.4.3 改进算法性能测试	30

3.5 本章小结	36
第 4 章 草莓定位方法研究	37
4.1 图像处理二维定位研究	37
4.1.1 图像预处理	37
4.1.2 颜色空间	37
4.1.3 图像分割	39
4.1.4 成熟草莓的二维定位	40
4.2 双目视觉原理	42
4.2.1 相机成像原理	42
4.2.2 双目视觉定位原理	46
4.3 被动双目相机三维定位研究	47
4.3.1 实验设备	47
4.3.2 标定实验	48
4.3.3 双目立体矫正	52
4.3.4 双目立体匹配	52
4.4 RGB-D 相机三维定位研究	54
4.5 本章小结	55
第 5 章 草莓检测模型系统搭建与边缘部署	56
5.1 系统需求分析	56
5.2 搭建环境介绍	56
5.3 系统功能介绍与测试	57
5.3.1 登录界面	57
5.3.2 主界面	57
5.3.3 草莓检测界面	58
5.4 边缘 AI 介绍	59
5.5 设备及环境搭建	60
5.5.1 基于 TensorRT 技术推理	60
5.5.2 模型的边缘部署	61
5.6 边缘设备识别实验	62
5.7 本章小结	64
第 6 章 草莓检测与定位实验	65
6.1 草莓检测实验	65
6.1.1 草莓检测实验内容	65
6.1.2 草莓检测结果分析	66
6.2 草莓定位实验	67
6.2.1 草莓定位实验内容	67
6.2.2 草莓定位结果分析	69
6.3 本章小结	70
第 7 章 总结与展望	71
7.1 论文工作总结	71
7.2 未来展望	72
参考文献	74

攻读硕士学位期间发表的学术论文及专利.....	80
致谢.....	80

第 1 章 绪论

1.1 研究背景与意义

草莓，亦称为红莓，也被称为洋莓或地莓等，其正式分类为蔷薇科草莓属植物，属于多年生草本植物，其高度介于 10 至 40 厘米之间。草莓的果实为聚合瘦果，外观呈鲜红色尖卵形，含有浓郁的水果芳香。草莓营养价值丰富，含有丰富的维生素、胡萝卜素、纤维素、铁、钙等营养物质，而且含量均远高于市场大部分水果。草莓中的维生素 A 和胡萝卜素，可以养肝和保护视力，有效缓解夜盲症，并且维生素 C 含量比葡萄与苹果高 7 至 10 倍。此外，草莓中还含有膳食纤维，能够促进胃肠道内的食物消化，改善便秘，预防肠癌和痤疮。草莓外观近似心形，鲜美红嫩，香气浓郁，果肉酸甜多汁，色、香、味俱佳且营养价值丰富，是名副其实的“水果皇后”^[1]。

随着我国经济的发展，人民生活水平的提高，对草莓的需求也越来越大，市场需求也吸引越来越多的人参与草莓产业，越来越多的农户种植草莓^[2]。如图 1-1 所示，通过国家数据统计，至 2022 年，我国已拥有草莓播种面积 147.45 千公顷，是全球最大的草莓生产与消费国，预计后面几年依然呈递增趋势。

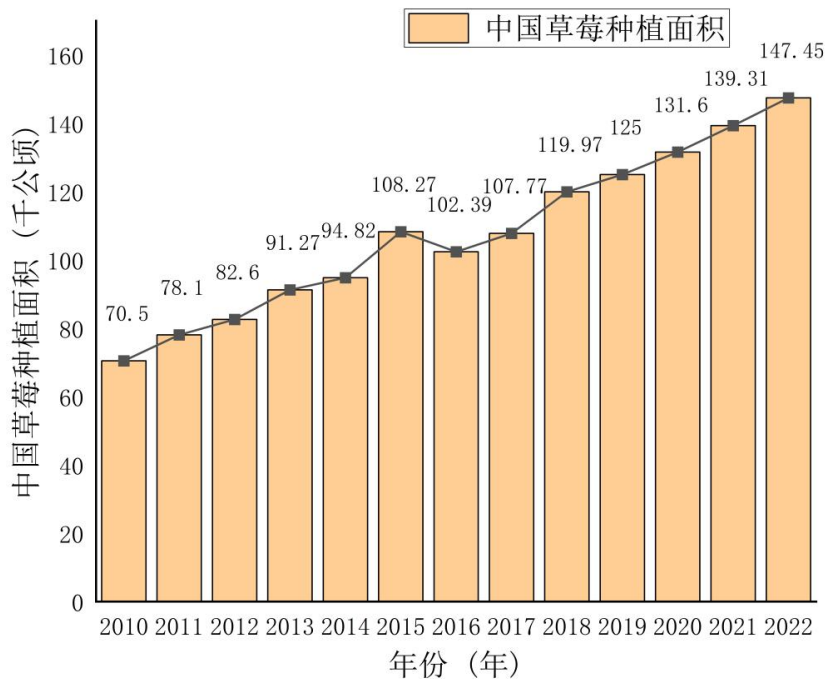


图 1-1 2010—2021 年中国草莓播种面积变化趋势

目前，人工采摘是我国草莓的主要采摘方式，人工采摘灵活性强、精度高，保证采摘时草莓不被碰伤等。但是草莓的采摘季节短暂且集中，其过程需要大量的人力资源。然而，受限于采摘期的短暂性，人手不足的问题常常难以避免。近年来，随着城市化步伐的加快和人口老龄化趋势的加剧，农村地区的劳动力呈现出不断减少的趋势。这一现象尤其体现在年轻一代对回归农村的意愿降低，这无疑进一步加剧了农业采摘工作中劳动力短缺的问题，对农业农村的持续发展产生了深远的影响。此外，随着劳动力成本不断提升，致使草莓人工采摘成本不断增加^[3]。

随着计算机视觉技术的飞速发展，自动化智能化设备的应用逐渐在各个领域铺开^[4]，机器视觉算法也在农业领域展开实验研究，各类农业自动化智能化设备开始发展，草莓采摘的自动化和智能化成为国内外研究的热点。与机械筛选相比，机器视觉算法因其具备更好的鲁棒性、更高的效率从而更适合大规模的检测和生产^[5-6]。近年来，深度学习逐渐兴起^[7-8]，神经网络可以通过训练对目标的特征进行学习，训练好的神经网络可以对信号、图像等进行识别。将农业智能机器人与深度学习相结合，可以使农业机器人具有学习能力，相较于单纯的机械型设备，具有很大的优势^[9-10]。在多台智能机器人并行运作时，其所学习到的特征信息可实现相互间的分享与交流，进而产生协同效应，共同提升机器人执行任务的精确性。

用机器人结合机器视觉采摘草莓替代人工是目前实现草莓自动化采摘的主要方式^[11-13]，并在部分方向成功实现了投产^[14-15]。基于机器视觉方式采摘草莓的工业机器人采摘速度略低于人工但可以实现 24 小时作业、能识别草莓的品种和成熟度，研究基于深度学习的草莓检测与定位方法有助于提升草莓自动采摘单位时间作业效率、扩展其作业范围、实现草莓成熟度精细化分选，还能提高当地居民的收入，带动农村当地企业和教育的发展，从根本上实现产业科技进步、生态可持续发展，对农业高质量发展具有深远的影响^[16]。将深度学习技术应用于草莓采摘机器人，极大地提高了草莓的采摘效率和采摘机器人的精度，对实现农业智能化具有重要意义^[17-18]。

1.2 研究现状

国内草莓采摘设备产品和样机主体上还处于试验和完善阶段，还未实现量产和应用。自动化采摘草莓的主要技术可以归结为草莓的检测识别^[19-21]、草莓定位以及采摘^[22-24]，其中，草莓的检测识别和定位是最关键的环节。近年来，随着科技的不断进步，国内外学者纷纷投身于农业智能化领域的研究，旨在通过技术手段提升农业生产的效率与便捷性。智能农业机器人的应用逐渐扩展，涵盖了土地施肥^[25-27]、果实采摘^[28-30]以及果实分拣^[31-33]等多个方面，为现代农业发展注入了新的活力。

其中，国外研究农业机器人的时间较早，像日本、美国、荷兰等智能机械较为发达的国家已经拥有了比较成熟的农业机器人技术^[34]。与这些国家相比，我国研究农业机器人的时间还很短，前期发展速度也比较缓慢，但是，经过近几年来国家对智能化农业生产的扶持以及科研人员的不断努力，农业机器人也有了很好的发展^[35]。

在机器人对果实进行采摘时，对目标进行准确地检测和定位是采摘机器人执行任务过程中最重要的一个环节^[36]。近年来，国内外研究者在果实检测与采摘点定位技术上取得了显著进展，这些成果为草莓采摘机器人的实际应用提供了有力的技术支持。以下，我们将详细论述这两个领域的研究现状。

1.2.1 果实目标检测国内外研究现状

在对成熟果实进行目标检测的前期，主要是运用基于颜色、区域、形状、边缘等特征的传统目标检测方法^[37-38]，例如 RGB、HSV、SIFT 等。朱坚民^[39]等人通过提取水果图像关注区域，提出了一种基于图像处理的水果颜色和形状特征参数的提取方法、基于灰色关联分析和模糊隶属度匹配的球形水果自动识别方法；Li 等^[40]通过对草莓图像进行线条绘制，得到草莓特征，然后使用 K-means 聚类方法对草莓图像进行分类，通过利用机器视觉技术实现了对草莓的分级。Kurtulmus F 等^[41]检测自然环境中的桃果实使用了统计分类器和神经网络提取了桃子的纹理特征，并利用支持向量机作为分类器，该方法检测精度较高，但检测速度较慢，不能满足采摘机器人的实时性要求。卢军等^[42]对拍摄的树上柑橘图像，基于水果表面的光照分布应用最大稳定极值区域算法提取图像中的

感兴趣区域，建立感兴趣区域周围的分层轮廓图，并利用霍夫变换拟合每一级轮廓获得分层圆形目标，最后进行拟合圆嵌套分析得到绿色柑橘水果目标，单幅图像平均处理时间为 3.70 s。郭峰等^[43]基于 OHTA 颜色空间的图像分割方法从背景中分割草莓，通过计算二值化草莓 blob 的惯性主轴来判断草莓的姿态，根据草莓成熟度实现了选择性采摘。以上方法虽然都取得了一定的效果，但是传统目标检测方法对复杂背景非常敏感，只适用于具有较为显著颜色、特征的目标，依赖于手工提取特征；当果实存在遮挡、重叠、光照等复杂因素时，识别准确率会显著降低，甚至无法识别目标，鲁棒性较差，这些问题导致传统目标检测方法在采摘机器人的应用具有一定的局限性。

近年来，基于深度学习的目标检测技术在农业果实识别中得到了很多的研究。刘小刚^[44]等人基于 YOLOv3，对静态草莓图片和动态视频进行了检测识别，得到了较好的检测效果。Chu^[45]等开发了一种用于苹果检测的新型抑制 Mask R-CNN，图片通过主干网络特征学习之后获得苹果候选的边界框，然后将边界框中的图像块送入特征抑制，去除错误标记的苹果候选框，在苹果数据集测试后，F1 分数为 90.5%，每帧检测时间为 0.25 秒；Yu 等^[46]使用 Mask R-CNN 用于草莓的识别与检测，主干网络使用 ResNet50，并与 FPN 网络结构结合，该方法在一定环境下，对重叠和遮挡以及在不同光照条件下的果实，具有较好的通用性和鲁棒性，但是网络结构复杂，实时性较差。赵德安等^[47]为提高苹果采摘机器人的工作效率和环境适应性，使用 YOLOv3 实现了对复杂环境下苹果的检测，得到了较好的识别准确率和速度。Zhou 等人^[48]通过对 YOLOv4 网络采用分组卷积，集成卷积层和批量归一化层，提出一种改进的轻量级网络，用于对草莓花进行检测，但是仅仅针对了单一的草莓花目标，并且模型依然具有轻量化的空间。Hu 等人^[49]使用 Faster R-CNN 模型实现了对垄上成熟番茄的快速准确检测。首先采用 Faster RCNN 对番茄进行定位，然后利用 HSV 色彩空间中 H 和 S 的高斯密度函数对番茄区域进行分割，去除背景像素，最后实现了不同环境下的番茄检测。孙俊等^[50]基于 YOLOv4 网络，通过采用 Ghost 卷积、添加注意力机制和修改损失函数，使网络检测精度得到了提升，模型规模得到了降低，但是训练集数量较少，模型鲁棒性较差。Kang 等^[51]采用 SSD 算法快速检测荔枝品质。

Ni^[52]等人提出了一种增强的 AlexNet，对草莓病害进行识别，从而用于草莓品质的评价。Hu 等^[53]通过使用 YOLOv3 检测网络对未成熟和成熟的草莓进行检测，得到较好的检测效果。Wang 等人^[54]通过添加扩展卷积、添加 ECA 注意力机制和多层感知器模块，提出了 BerryNet-Lite，对草莓病害得到了较好的检测效果。Zhou 等^[55]通过对 YOLOv4 的骨干网络和颈部网络进行改进，提出了一种轻量级参数网络，实现了对草莓花、番茄等目标的识别。Wang 等人^[56]将 ResNext 网络和 SVM 结合起来，用于识别草莓的外观质量。

基于深度学习的目标检测算法，模型通过数据集训练后，就可以检测图片中的相应目标，用边界框将目标表示出来，并显示对应的类别和置信度，使人们清晰地观看到目标的位置信息。它可以利用丰富的数据集，对复杂环境具有较强的适应能力，并且具有较高的检测精度，基于端到端的训练，相比于传统目标检测方法，节省了大量手工提取特征的时间，能够更好地适应复杂环境下果实的检测任务。

1.2.2 果实定位国内外研究现状

果实的目标检测是后期采摘定位的基础，采摘机器人首先要检测到需要采摘的目标，然后根据图像计算出在现实世界中的位置，从而通过末端执行器实施采摘。目前常用的三维定位方法主要有立体视觉匹配、深度相机、光基 3D 相机飞行时间、单目彩色相机和激光测距仪等。近年来，国内外学者采用了不同的方法来实现果实的采摘定位，其中：毕松等^[57]通过颜色信息和饱和度信息，并结合 YOLOv3 模型，进行成熟度和遮挡判断，完成精度区域点云分割，实现草莓的位姿估计和采摘点估计。张凯良等^[58]使用两个 CMOS 摄像头和一个激光发生器，首先采用镜像匹配法计算出草莓果轴在 X-Y 平面内的位置信息，然后利用扇形激光束和几何光学，用来辅助判断采摘位置和末端执行器深度方向的距离，实现对草莓的定位采摘。王粮局等^[59]首先进行图像预处理去除图像噪声，通过外接矩阵的方法得到了草莓域轮廓，然后根据运动学关系，获得摄像机光心和其理想位置的实际偏差，通过 PID 控制策略对定位机构的 X 轴速度进行调节，最后采用运动恢复结构的方法计算出精确的深度信息，根据针孔摄像机的原理，计算出了末端执行器需升高的距离和夹指宽度等采摘参数。陈燕等^[60]首

先使用双目相机采集图像，通过改进的 YOLOv3 网络实现荔枝串的检测，通过同行顺序一致性约束实现荔枝串的配对，最后基于双目立体视觉的三角测量原理计算出荔枝串的空间坐标。张勤等^[61]提出了一种基于 RGB-D 信息融合和目标检测的草莓串采摘点识别定位方法。首先通过 YOLOv4 网络，检测出可采摘果梗的感兴趣区域，然后，融合图像中的深度信息和颜色特征识别采摘点，并综合目标果梗深度信息均值填充算法得到采摘点精确深度值，实现采摘点精确定位。Lin 等^[62]利用 SVM 与双目视觉结合的方法，对田间柑橘进行检测，该方法在野外条件下对柑桔类水果的检测具有较强的鲁棒性，然而，其执行时间较长，难以达到实时。M.Benavides 等人^[63]开发出一个计算机视觉系统，该系统能够利用图像增强、边缘检测、图像分割等图像处理工具检测出成熟西红柿，后续利用三角学几何知识获取了单个西红柿的采摘点，从而实现西红柿的定位。Yoshida 等^[64]获取了番茄束状作物的三维图像，检测了束状花序梗的位置，研究只使用了 6 张样本图像，并且每张图像的计算时间没有指定。Ge 等^[65]通过调查不同相机模式和相机类型的影响，利用每幅图像中被检测目标的像素、相应的深度值以及从检测结果中提取的二维边界框的位置，从而提供目标的三维位置。

1.3 目前存在的问题

根据国内外果实目标检测和采摘定位研究现状，目前的果实目标检测和采摘定位方法都存在着一一定的局限性，主要有以下问题：

(1) 传统目标检测方法基于滑动窗口的区域遍历图片选择方法没有针对性，产生的冗余窗口较多，致使模型训练用时较长；手工设计特征需要花费大量的时间，导致模型训练后的鲁棒性差、泛化能力弱、检测速度和精度都比较低下，可能会产生多个正确识别的结果，即一个物体类别预测出了多个类别，无法实现端到端方案，使很多任务不能一次性解决，需要多个步骤，较为繁琐。

(2) 目前深度学习草莓检测并没有公开使用的数据集，采集图像比较耗时耗力，相当部分的研究仅仅针对小部分数据集展开研究，草莓目标检测数据集的缺乏给深度学习技术在草莓检测中的应用带来挑战。

(3) 深度学习目标检测算法对草莓的检测精度具有提升的空间，并且当检

测环境较为复杂，例如果实遮挡、重叠和数量较多时，存在着较高的漏检率。

(4) 目前深度学习草莓检测模型都比较复杂，模型参数量较多、模型占用内存较大，训练和边缘部署时对硬件要求较高。

(5) 关于草莓的定位，国内外多采用传统方法对目标进行图像分割、形态学处理、边缘处理等多种图像处理技术对成熟草莓果实进行定位，过程比较复杂繁琐；使用激光雷达比较昂贵，成本较高，同时草莓主要是靠表面着色度判断成熟程度，激光雷达无法识别颜色信息。所以，需要简化采摘定位过程，提升采摘机械的快速响应能力，从而提高草莓检测和采摘的效率，同时降低硬件成本，节省开支。

1.4 论文主要的研究内容

(1) 草莓数据集的构建。对收集的草莓图片进行标注、分类，分为训练集、验证集和测试集，用于后面深度学习目标检测算法的训练及测试，并使用多种数据增强算法进行数据增强，提升模型训练后的泛化性能。

(2) 基于深度学习的草莓检测方法研究。对现有的 YOLOv5 单阶段目标检测算法的网络结构进行改进，通过使用 Ghost 卷积技术、添加高效通道注意力机制 ECA，将原先的特征融合模块由 SPPF 替换为高效的 SimSPPF 模块，损失函数由 CIoU 替换为 SIoU，从而降低了模型复杂度，提升了模型的检测精度和对遮挡、密集草莓的定位性能。

(3) 基于图像处理和双目相机的草莓定位方法研究。通过图像预处理、颜色空间转换和图像分割实现对成熟草莓果实的二维定位；分别对被动双目相机与 RGB-D 相机进行标定，得到相机的外参矩阵和内参矩阵，然后对被动双目相机进行立体矫正和立体匹配，通过相机成像原理和双目相机定位原理，实现对草莓目标的深度距离测量。

(4) 通过图形程序框架 PyQt5，设计了草莓检测系统，对改进模型的预测模块进行可视化用户界面设计，并能够对模型、输入文件类型、交并比阈值和置信度阈值等相关信息进行解析和调整，使其更为直观，方便用户进行查找和使用。

(5) 将所得到的草莓检测模型进行 TensorRT 加速，使其更适合在边缘设备上部署，并对其进行边缘设备部署实验分析。

1.5 论文技术路线和章节安排

1.5.1 技术路线

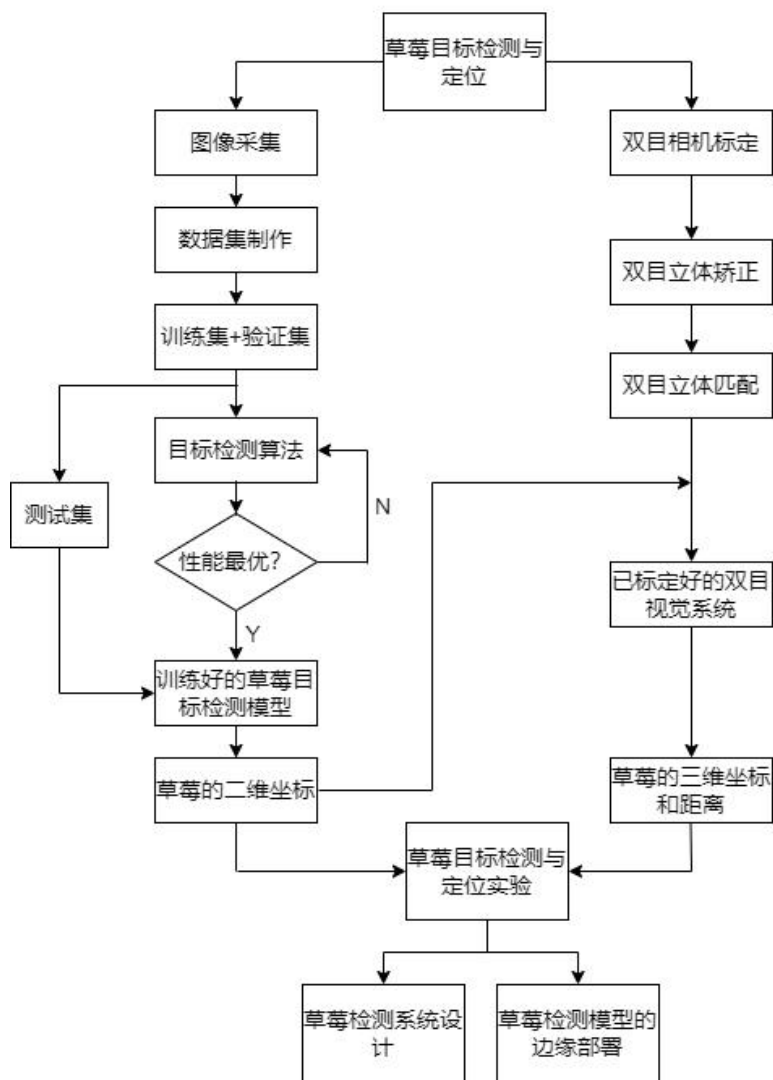


图 1-2 技术路线

基于深度学习的草莓检测与定位方法研究，主要对收集到的草莓图片进行标注，生成草莓目标检测数据集。对深度学习目标检测模型进行改进优化，将其中的训练集输入到模型中进行训练，得到相应的权重参数，从而实现对草莓的目标检测，在此基础上进行检测系统的开发和边缘部署工作。通过图像处理

实现了对成熟草莓的二维定位，并使用双目相机，实现了对草莓的三维定位。主要的技术路线如图 1-2 所示。

1.5.2 章节安排

本文主要使用深度学习技术对草莓果实的检测识别和使用相机对草莓的定位展开研究，各章节主要内容如下：

第一章：绪论。主要介绍了草莓采摘中目标检测与定位的研究背景和意义，通过草莓目标检测与定位的国内外研究现状了解现在行业研究的热点以及目前研究下存在的一些问题，提出了本课题总体的研究内容和技术路线。

第二章：基于卷积神经网络草莓检测相关理论介绍。主要介绍了卷积神经网络中卷积层、池化层和全连接层的结构和定义，同时介绍了双阶段目标检测算法 R-CNN、Fast R-CNN 和 Faster R-CNN 的网络构成和特点，以及单阶段目标检测算法 SSD 和 RetinaNet 的网络构成和特点，为深入研究草莓目标检测模型奠定了一定的理论基础。

第三章：草莓目标检测方法研究。首先，介绍了草莓目标检测数据集的收集与制作，为训练深度学习草莓目标检测模型提供了基础条件；然后，介绍了 YOLOv5 模型的网络结构和损失函数，针对原始 YOLOv5 对草莓检测精度较低、在复杂条件下，如遮挡，重叠条件下检测效果较差、训练时模型复杂度较高、模型规模较大等问题，对模型进行了改进优化：先通过使用 Ghost 模块降低了模型复杂度，减小了模型训练后的体积，接着采用高效通道注意力机制 ECA 和 SimSPPF，提升了模型对草莓的检测精度，然后使用 SIOU 定位损失函数提升模型对遮挡和重叠目标的检测效果，并使用测试集验证改进方法的有效性。

第四章：草莓定位方法研究。首先，通过图像预处理、颜色空间转换和图像分割方法，实现对成熟草莓目标在图片中的二维定位；然后，介绍了相机成像原理和双目相机定位原理，通过研究被动双目相机，对其进行标定，得到了相机的外参矩阵、内参矩阵和畸变系数，并通过立体矫正和立体匹配；通过对 RGB-D 相机进行标定，得到了相机的内参矩阵和畸变系数，为实现三维定位奠定了基础。

第五章：草莓检测系统搭建和边缘部署。首先，对草莓检测系统进行需求分析，并介绍其搭建环境；然后，使用程序创建草莓检测系统，并进行登录界面和检测主界面设计；最后，通过使用 Jetson Nano 设备，采用 TensorRT 加速，实现了改进模型在边缘设备的部署工作。

第六章：草莓检测与定位实验。通过草莓检测实验结果验证了本文提出的改进方法的有效性；并通过草莓的三维定位的实验结果，验证了使用 RGB-D 相机相较于被动双目相机，具有更为优越的定位性能。

第七章：总结与展望。对本课题基于深度学习的草莓检测与定位方法研究进行综合分析总结，根据对草莓检测和草莓定位存在的问题进行对应的分析，给出未来改进方向及规划。

1.6 本章小结

本章主要阐述了草莓目标检测和采摘定位的研究背景与意义、研究现状和目前存在的一些问题，最后对本文的主要研究内容做出了总结并给出了整体的技术路线。

第2章 基于深度学习的草莓检测相关理论

深度学习，主要是通过学习样本数据的内在规律和表示层次，让机器能够具有类似于人的分析学习能力，最终目标是使机器能够识别和解释各种数据。目标检测，是深度学习计算机视觉中非常重要并且研究难度较高的领域之一。模型通过数据集训练后，就可以检测图片中的相应目标，用边界框将目标表示出来，并显示对应的类别和置信度，使人们清晰地观看到目标的位置信息，与传统方法相比，它可用于在高度复杂的场景中实现更可靠、更高效的检测。近年来，已经运用到生活的各个领域，如自动驾驶、实时监控、智能机器人等。虽然应用得越来越广泛，但是目标检测模型在对遮挡、光照、形状、背景环境等因素较为敏感，对于草莓等果实高品质采摘的检测能力仍然具有较大的提升空间。

2.1 卷积神经网络

卷积神经网络在深度学习领域，尤其是在图片分类、目标检测和目标跟踪等方面，拥有广泛而重要的应用^[66-68]。卷积神经网络架构由多个非线性卷积层组成，这些卷积层可以学习比浅层网络更复杂的特征，从而实现高度复杂的特征抽象。本质是一个多层感知机，采用了局部连接和权值共享的方式，从而大大降低了参数数量，它主要由卷积层、池化层、激活层、全连接层和损失函数组成。其中，卷积层的作用是提取输入图片中的信息；池化层的作用主要是降低数据的空间维度和计算复杂度，保持不变性；激活层的作用是引入激活函数，加入非线性因素，提高网络的表达能力；全连接层的作用负责将卷积输出的二维特征图转化成一维的一个向量，由此实现了端到端的学习过程；损失函数的作用是为了衡量预测值与实际数据的差距程度。

2.1.1 卷积层

卷积层在卷积神经网络中是为了执行卷积操作，进行特征提取，通过矩阵点乘和求和运算得到结果，它的基本公式是 $y = wx + b$ ，其中， x 表示输入特征

图， w 表示卷积核， b 表示偏置，如图 2-1 所示，输入特征图宽和高为 5，通道数为 1，卷积核大小为 3×3 ，偏置为 1，输入特征图像素与对应位置卷积核参数相乘后累加得到输出特征图的一个位置的像素值。



图 2-1 卷积层的计算过程

输出特征图宽和高的大小可由计算公式 (2-1) 和 (2-2) 得出：

$$H_{out} = \frac{H_{in} + 2p - H_k}{s} + 1 \quad (2-1)$$

$$W_{out} = \frac{W_{in} + 2p - W_k}{s} + 1 \quad (2-2)$$

其中， H_{in} ， H_{out} 和 W_{in} ， W_{out} 分别表示输入特征图和输出特征图的高和宽； s 表示步长， p 表示填充，这里取步长为 1，填充为 0。

2.1.2 池化层

池化层，一般放在卷积层之后，主要是为了减少特征图的宽高维度，进行降采样，减少参数，降低模型复杂度，同时保持维度变化后的特征不变性。通过池化层，线性，减可以提高模型的计算速度和鲁小网络过合模拟率。池化操作通常使用最大值池化和平均值池化，这两种池化操作过程如图 2-2 所示。

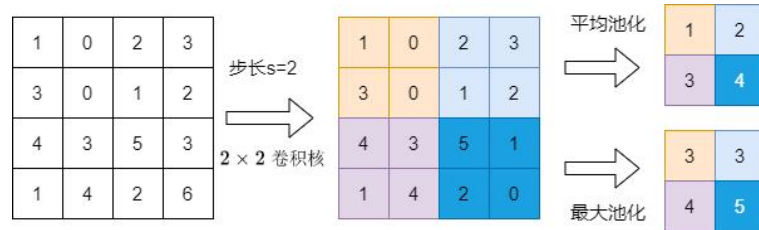


图 2-2 池化操作

由图 2-2 可以看到，池化层不需要学习新的参数，只需要指定池化操作的类型，假设输入特征图的高和宽为 h 、 w ，卷积核大小为 $k \times k$ ，步长 s 为 1，则输出特征图的宽 w' 和高 h' 可由公式 (2-3)、(2-4) 计算得出：

$$w' = \frac{w - k + s}{s} \quad (2-3)$$

$$h' = \frac{h - k + s}{s} \quad (2-4)$$

2.1.3 全连接层

全连接层一般连接到卷积神经网络最终得到的特征图之后，负责将经过卷积操作后输出的二维特征图转换为一个一维特征向量，由此实现了端到端的学习过程，实现分类。它与前后层的所有节点连接，这就使全连接层具有较多的参数，占用的空间较大，主要包括三个部分，输入层、隐藏层和输出层，其中，输出层神经元的个数等于预测分类类别的个数，如图 2-3 所示。

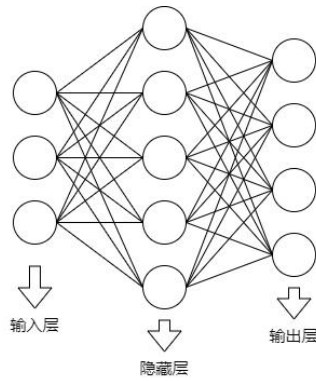


图 2-3 全连接层

2.2 经典目标检测算法

目标检测是计算机视觉中非常重要并且研究难度较高的领域之一，它是分类和定位任务的混合，确保图像中目标的精确检测及其与背景的明确区分，运用多对象像素掩码与边界框技术，针对每个对象实例进行细致处理。同时，结合置信度分数与对象标签，实现目标的精确定位。主要分为双阶段目标检测算法和端到端的单阶段目标检测算法，双阶段网络的检测精确率高于单阶段网络，检测速度却相反。

2.2.1 双阶段目标检测算法

针对传统目标检测的问题，Ross B.Girshick 等人^[69]创建了 R-CNN，如图 2-4 所示，它的候选区域是预先找出图中目标可能出现的位置，利用 selective search 算法，根据颜色、边缘、纹理等在图像中提取 2000 个左右的候选区域，

扭曲成固定大小，然后用卷积神经网络提取每个候选区域特征，送入 SVM 分类器进行分类，代替了传统的滑动窗口选出候选区域和手工设计特征，使得目标检测技术取得巨大突破，并开启了基于深度学习目标检测的热潮，但是存在训练速度过慢、占用磁盘空间大、训练步骤繁琐等问题。

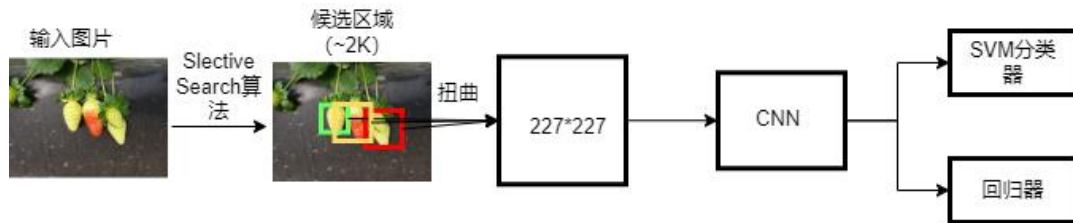


图 2-4 R-CNN 流程图

Ross Girshick 等人^[70]结合 R-CNN 和 SPPNet 提出 Fast R-CNN，如图 2-5 所示，Fast R-CNN 使用 ROI（Region of Interest，感兴趣区域）对所有候选区域进行相同尺度的特征变换，避免了裁剪和扭曲（crop/warp）操作，只需要对图像进行一次特征提取操作，从而节约了较多的训练时间；引入了多任务损失函数，同时将边界框回归直接加入卷积神经网络中进行训练，用 softmax 函数代替了 SVM 分类器，提升了多分类任务的准确性。但是，使用 selective search 进行候选区域的提取依然需要花费较多的训练时间。

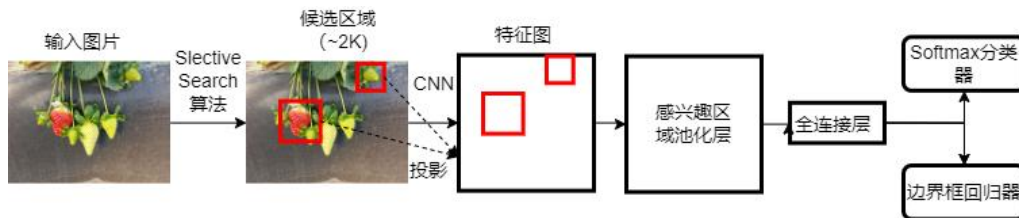


图 2-5 Fast R-CNN 流程图

Ren 等人^[71]提出了 Faster R-CNN，如图 2-6 所示。通过 RPN（Region Proposal Networks）网络，进行候选区域的选择，使用卷积神经网络直接产生候选区域，从而舍弃了 R-CNN 和 Fast R-CNN 一直采用的非常耗时的 selective search 算法，极大地提升了模型的检测速度和检测精度。

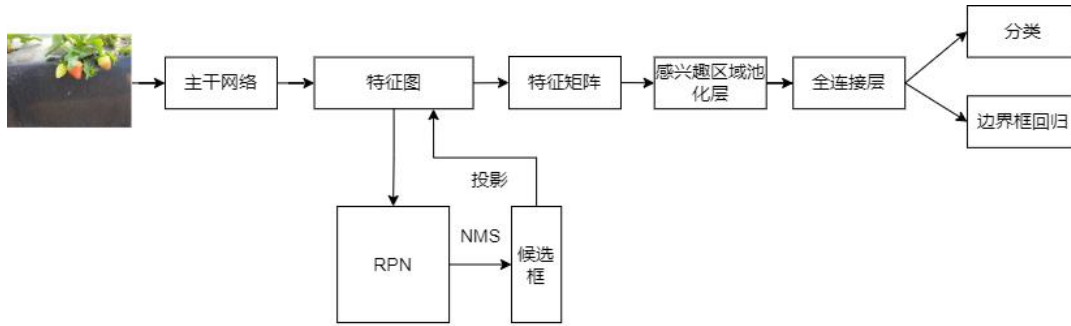


图 2-6 Faster R-CNN 流程图

2.2.2 单阶段目标检测算法

Faster R-CNN 的这种双阶段方法虽然对目标检测精度较高，但是检测速度较慢，不能满足实时的要求。Redmon 等人^[72]提出 YOLOv1，单阶段的方法慢慢显现出其重要性。这种方法采纳了回归的理念，即在接收到输入图像后，直接在其多个位置回归出目标的边界框及其所属类别。这种方法显著提升了检测速度。同时，每个网络在预测目标窗口时，充分利用了整幅图像的信息，拥有充分的上下文信息，使得将负样本预测为正样本的比例大幅降低。但是 YOLOv1 也存在问题：没有了候选区域机制，只使用固定大小的网格回归使得其目标不能精准定位，致使 YOLOv1 的检测精度不是很高，特别对小物体检测效果较差，因此，由其出现了 SSD、RetinaNet 等一系列单阶段目标检测算法，以提升其目标检测精度和速度。

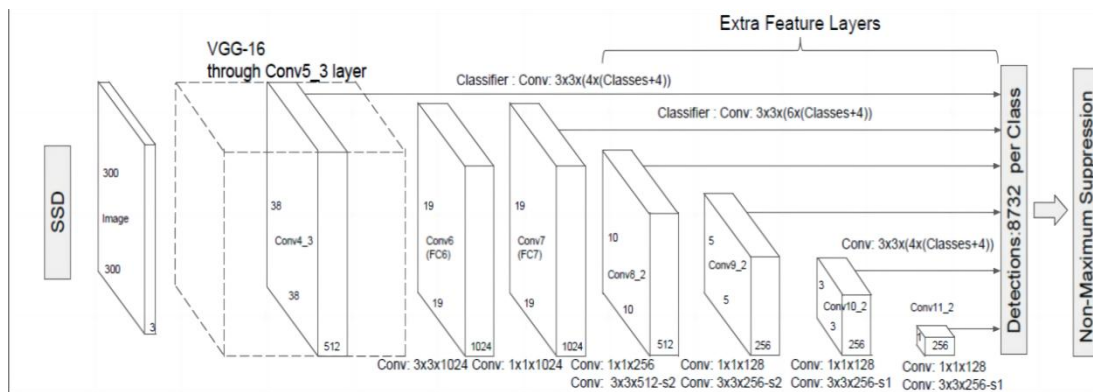


图 2-7 SSD 流程图^[70]

SSD (Single Shot MultiBox Detector) 算法是一种基于深度学习的目标检测方法，它由 Wei Liu 等人^[73]于 2016 年提出，该算法将回归思想和锚框机制结合，消除了双阶段算法中候选区域生成和随后的像素或特征重采样阶段，并将所有计算封装在一个网络中，使其易于训练，速度较快。SSD 网络可以分成特征提

取和检测框生成两部分，特征提取采用的基础网络是从分类网络 VGG16^[74]借鉴而来的，SSD 在裁剪的基础网络之后添加了卷积层，这些层的特征图大小是逐步减小的，从而实现在多尺度下进行预测，具体结构如图 2-7 所示。

RetinaNet 是由 Tsung-Yi Lin 等人^[75]提出的一种单阶段目标检测算法，由 ResNet^[76]、FPN 和 SubNet 三部分组成，它的损失函数为 Focal loss，其最大的贡献就是解决了前景背景种类不均衡问题，使单阶段目标检测算法的精度达到了双阶段目标检测算法的水平。具体结构如图 2-8 所示。其中，ResNet 作为它的主干网络，负责对输入的图片进行特征提取；FPN（特征金字塔网络）负责将不同尺度的特征融合在一起，从而提高信息的丰富性；SubNet 负责对网络输出的目标进行分类，并对目标所需的边界框进行回归。

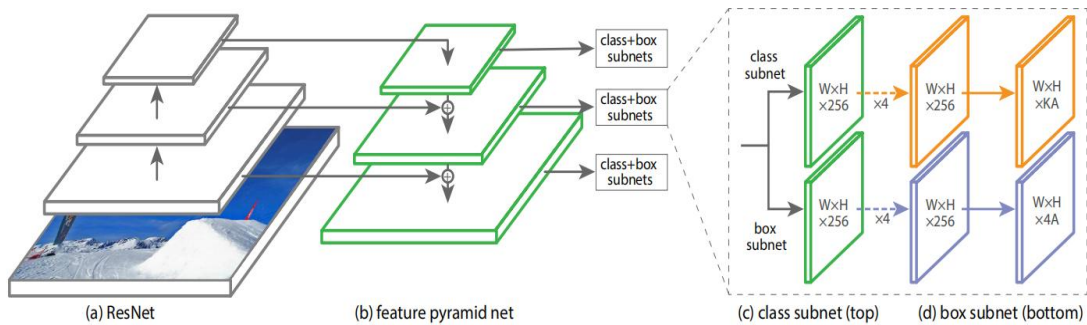


图 2-8 RetinaNet 流程图^[72]

综上，虽然双阶段目标检测算法相对于单阶段目标检测算法具有更高的检测精度，但是需要首先通过候选区域进行筛选，然后进行分类与回归，消耗较多的时间成本，并且所要求的硬件设施非常昂贵。在实际的草莓采摘中，由于成熟草莓数量非常多，丰产期采摘间隔很短，若不及时采摘，草莓过熟便会腐烂变质，危害别的未成熟的果实品质，对草莓采摘机器具有很高的实时性要求，故本文使用速度更快的单阶段目标检测算法，并进行改进，提升算法对草莓的检测精度、复杂环境下的检测效果、并进一步降低模型复杂度，使其更为容易地部署到嵌入式设备中。

2.3 本章小结

本章首先详细地介绍了卷积神经网络中卷积层、池化层和全连接层的原理和作用。然后对目前比较经典的双阶段目标检测算法和单阶段目标检测算法进行了一定结构和原理上的阐述。

第3章 改进 YOLOv5 草莓检测模型研究

随着计算机视觉的不断发展，深度学习技术逐渐运用到工业和农业的各个方面，尤其是卷积神经网络，在图像分类、目标检测、语义分割等领域得到了许多具有创新性的应用。在图像方面，卷积神经网络可以提取特征，对不同大小的特征进行融合，通过学习到的权重和偏置参数，进而对图片中的目标进行分类、定位和分割。对不同生长状态的露地种植的草莓，本章通过在 YOLOv5 目标检测网络下对其进行网络结构和损失函数等方面进行改进优化，并与其它经典目标检测网络进行对比分析，通过消融实验，证明所得模型的优越性能。

3.1 草莓目标检测数据集采集与制作

3.1.1 数据集采集与制作

深度学习的发展需要三大因素：数据集、计算力和算法，它们是相辅相成、相互提升，三者缺一不可，要想实现草莓的检测，制作草莓目标检测数据集作为基础步骤是必不可少的。首先，尽可能多地收集了农业规模化种植的草莓图片，在训练目标检测模型之前，必须人为对图像中的目标进行分类，并用合适大小的边界框标注位置信息，生成记录信息的文件标签，然后输入到目标检测模型中进行训练，从而得到可以对不同生长状态的草莓进行检测分类和定位的识别检测模型。



图 3-1 草莓示例图片

本文草莓图像为 StrawDI 团队收集的不同生长周期的草莓^[77]，部分图片如

图 3-1 所示，收集草莓图片共 3100 张，图片大小为像素 1008×756，按照 8: 1: 1 的比例分为训练集（2480 幅）、验证集（310 幅）和测试集（310 幅）。

草莓成熟期的划分与判断，最重要的指标是果面着色程度，生产上可以以此作为确定采收成熟度的标准。根据中华人民共和国农业行业标准中草莓采收与贮存技术规范规定^[78]：1）当地市场鲜销的草莓，宜在果面着色达到 90%以上时采收。2）需要贮藏或者物流运输销售的草莓，果实表面着色达到 80%以上且保持较高的硬度时采收，避免过度成熟。本文使用 Make Sense 工具进行标注，标注类别分为三类，分别为（1）flowering：草莓花，标签个数为 1100；（2）fruit：未成熟的草莓，即表面着色度小于等于 80%的草莓，标签个数为 10218；（3）ripe：已经成熟可采摘的草莓，即表面着色度大于 80%的草莓，标签个数为 4718。其中，遮挡草莓目标占比约为 10.30%，重叠草莓目标占比约为 1.28%，标注时部分示例图如图 3-2 所示，标注结束后生成 YOLO 数据格式的.txt 文件和 xml 文件。

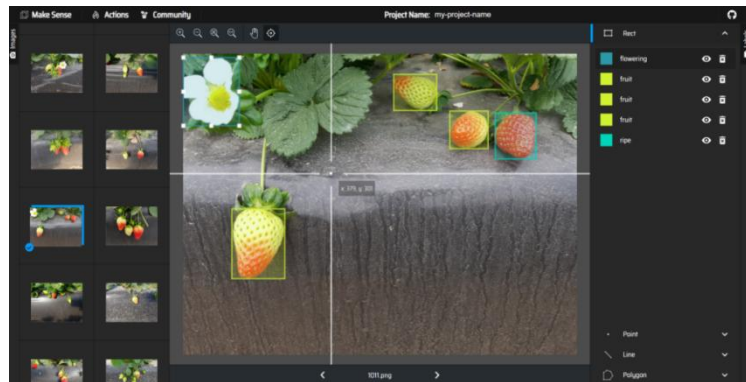


图 3-2 Make Sense 标注界面

3.1.2 草莓目标检测数据集展示

本文模仿 POSCALL VOC2012 数据集格式^[79]进行制作，文件构成如图 3-3 所示。

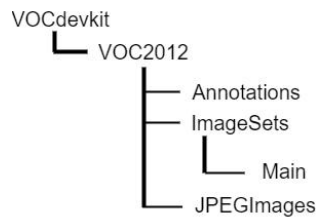


图 3-3 草莓目标检测数据集文件结构

其中，Annotations 文件夹主要储存由程序 Make Sense 标注后产生的 xml 文件，xml 文件主要记录了标注图片的信息，如图 3-4 所示，filename 表示图片名称；size 表示图片尺寸，由宽（width）、高（height）和通道数（depth）三部分组成；object 表示目标，主要部分是目标类别（name）；bndbox 表示目标位置信息，标注框为矩形，故 xmin、xmax、表示左横坐标和右横坐标，ymin、ymax 表示下纵坐标和上纵坐标。Main 文件夹主要是将数据集分为训练集、验证集和测试集的文本文件，JPEGImages 文件夹里为本次论文实验中收集的 3100 张露地种植的草莓图片。

```
<?xml version="1.0" encoding="UTF-8" ?>
<annotation>
  <folder>WOC2012</folder>
  <filename>02001.jpg</filename>
  <source>
    <database>Unknown</database>
  </source>
  <size>
    <width>1008</width>
    <height>756</height>
    <depth>3</depth>
  </size>
  <object>
    <name>fruit</name>
    <pose>Unspecified</pose>
    <truncated>0</truncated>
    <difficult>0</difficult>
    <bndbox>
      <xmin>347</xmin>
      <ymin>72</ymin>
      <xmax>393</xmax>
      <ymax>146</ymax>
    </bndbox>
  </object>
</annotation>
```

图 3-4 生成的 XML 文件

表 3-1 数据增强参数

数据增强方法	概率
HSV_H	0.015
HSV_S	0.7
HSV_V	0.4
图像旋转	0.1
图像平移	0.1
尺寸变换	0.5
图像裁剪	0.5
图像透视	0.001
水平翻转	0.5
上下翻转	0.5
Mosaic	1.0
Mixup	0.1

为了使训练出来的模型效果最佳，增强模型的泛化能力，提升鲁棒性，使用了 Mosaic、Mixup、随机缩放、随机调整 HSV、随机翻转、图像旋转、图像平移、尺寸变换、图像裁剪、图像透视 10 种方式进行数据增强。其中，Mosaic 将四张图片组合成一张图片；Mixup 将两张图片按照一定的透明度融合在一起；随机缩放将原始图片缩放到一个标准尺寸；随机调整 HSV 随机调整图片的色调、饱和度和明度；图像透视将使图像里的目标更加透明；组合这些方法，进行数据增强，具体设置参数如表 3-1 所示。

3.2 YOLOv5 算法

3.2.1 YOLOv5 网络结构

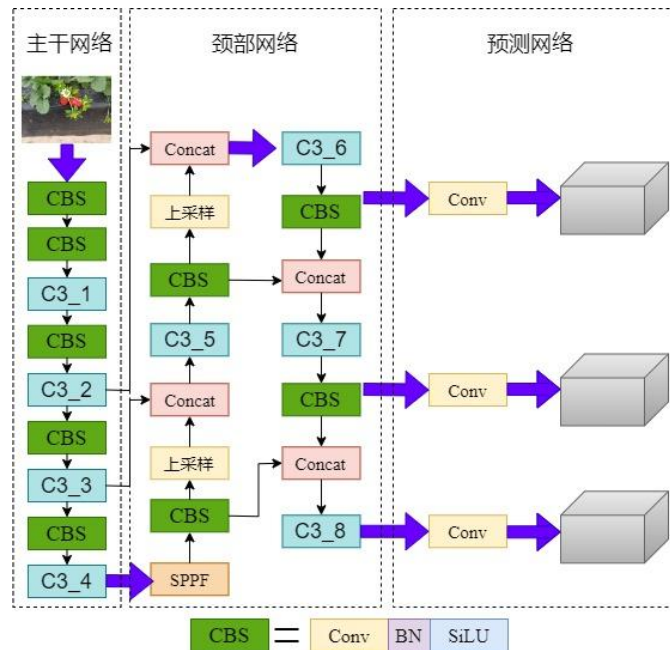


图 3-5 YOLOv5 基础模型

YOLOv5 是基于 YOLOv4^[80]出现的一种新的单阶段目标检测算法，相较于前面提到的 SSD、RetinaNet 模型，具有训练和推理速度快、使用灵活性高等特点，主要由主干网络（Backbone）、颈部网络（Neck）和预测网络（Head）这三部分组成。主干网络负责特征提取，首先输入大小为 640×640 的图片，通过卷积核大小为 6×6 ，步长和填充为 2 的卷积层，得到其特征图，然后通过跨阶段网络 CSPDarkNet53 结构，完成对特征图通道维度的拆分和拼接。颈部网络的核心构成是路径聚合网络（Path Aggregation Network, PAN），其关键任务在

于整合不同尺寸的特征图，实现特征的全面融合。此举旨在提升模型对小目标和大目标的综合检测能力，确保在各种尺度下都能实现精准识别。最后通过预测网络实现对目标类别、置信度和边界框的预测回归，YOLOv5 整体模型如图 3-5 所示。

3.2.2 YOLOv5 损失函数

YOLOv5 的损失由三部分组成，分别为分类损失、置信度损失和定位损失，分类和置信度损失使用的是适合多标签分类任务的 BCE 损失函数，如公式（3-1）所示。

$$Loss(o,t) = -\frac{1}{n} \sum_i w[i] (t[i] \times \log(o[i]) + (1-t[i]) \times \log(1-o[i])) \quad (3-1)$$

其中， n 表示类别数； $w[i]$ 表示权重系数； $t[i]$ 表示目标类别，0 或者为 1； $o[i]$ 表示目标类别预测分数；

YOLOv5 的定位损失使用的是 CIoU 损失函数，如公式（3-2）所示。

$$L_{CIoU} = 1 - IoU + \left(\frac{\rho^2(b, b^{gt})}{c^2} \right) + \alpha v \quad (3-2)$$

其中， IoU 表示交并比； ρ^2 表示两个中心点的欧式距离； b, b^{gt} 分别表示预测框和真实框的中心点； c 表示能包含标注框和真实框的包围区域的对角线距离； α 为权重参数； v 为用来度量长宽比的相似性。

3.3 相关模块及改进方法

3.3.1 ECA 注意力机制

由于草莓存在着遮挡、体积较小多变等问题，导致模型检测精度不佳，故本研究在 YOLOv5 模型中添加高效通道注意力机制 ECA（Efficient Channel Attention）^[81]，它是一种轻量级的注意力机制，通过考虑通道信息、避免维度缩减、适当的跨通道交互来降低模型复杂度，由高权重来聚焦草莓特征，用低权重忽略背景信息，达到对草莓特征信息的关注，提升对草莓的检测性能，其模型结构如图 3-6 所示。

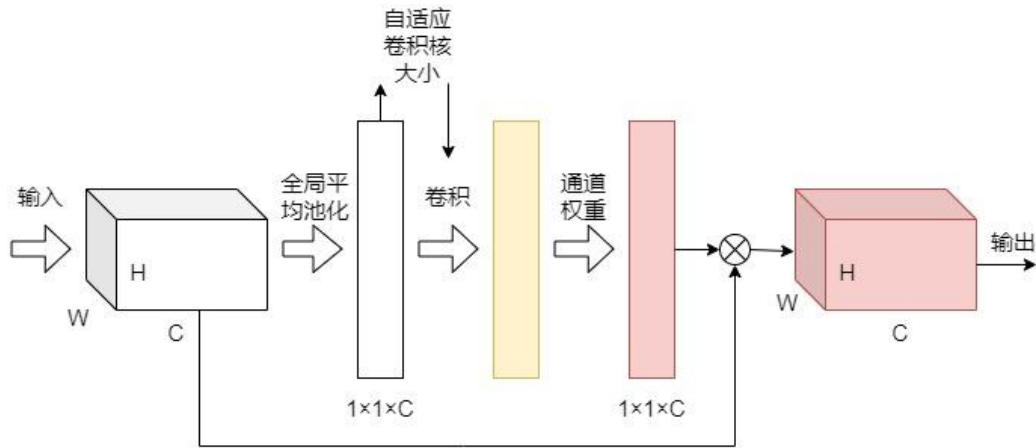
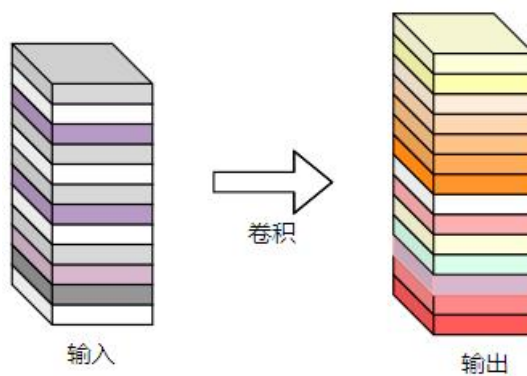


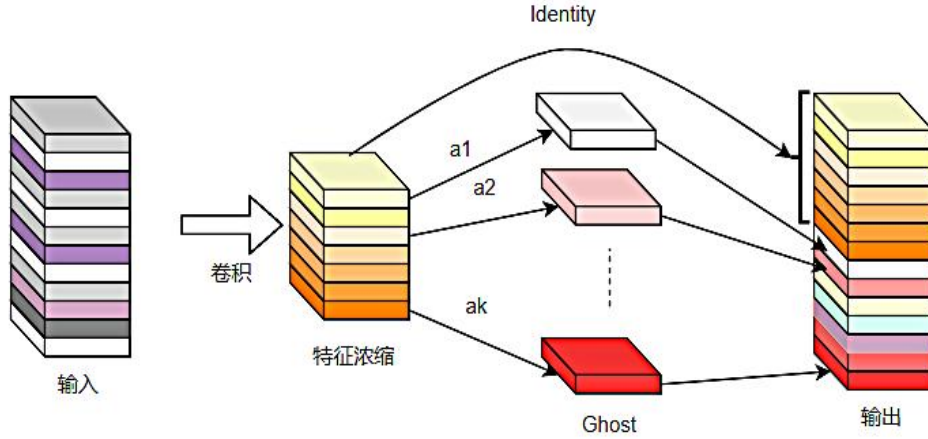
图 3-6 ECA 注意力模块

3.3.2 Ghost 模型

GhostNet^[82]是由华为提出来的一个轻量化模型。当图片输入到卷积神经网络后，会生成特别多的特征图，它们之间存在着两两相似度特别高的特征图，这时候就会产生冗余，作者将相似的特征图认为是彼此的 Ghost，这个网络就被称为 GhostNet。通过使用 Ghost，可以在保证良好检测效果的前提下，减少模型参数量，降低模型复杂度。如图 3-7 所示，Ghost 模组分为两个部分，首先通过一个 1×1 卷积，负责特征整合，生成特征浓缩，然后进行深度可分离卷积，即 Cheap Operations，通过特征浓缩生成 Ghost 特征图，其中 Identity 代表恒等变换， $a_1, a_2 \dots a_k$ 代表 Cheap Operations。



(a) 普通卷积



(b) Ghost 模块

图 3-7 普通卷积和 Ghost 模块示意图

3.3.3 SIoU 损失函数

在卷积神经网络中，损失函数作为一种优化模型的惩罚措施，需要在训练时最小化，它对模型检测的识别准确度、定位精度起到了非常大的作用。由公式（3-2）可知，CIoU 虽然考虑了边界框回归时的中心点距离、重叠面积和纵横比，但是却忽略了真实框与预测框方向不匹配的问题，对识别定位准确度造成了一定损失，而 SIoU^[83]通过考虑边界框回归之间的向量角度，对边界框回归方向定义了惩罚指标，提高了模型的定位性能。SIoU 由四部分组成，分别为角度损失、距离损失、形状损失和交并比损失，具体解释为：

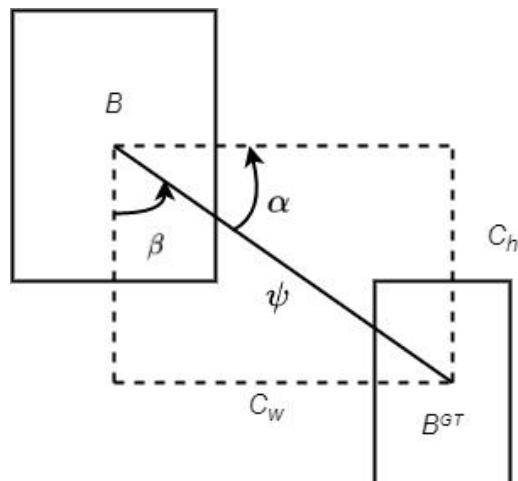


图 3-8 角度损失示意图

角度损失，如式（3-3）和图 3-8 所示，描述了中心点连线与 x-y 轴之间的

最小角度，如果 $\alpha \leq \frac{\pi}{4}$ 首先尝试最小化 α ，否则最小化 $\beta = \frac{\pi}{2} - \alpha$ ，主要是为了减少距离有关的变量数量，同时，首次在 x 轴和 y 轴实现预测，并继续在相关轴之间进行处理，角度损失函数用以下公式定义：当中心点在 x 轴或 y 轴对齐时， $A=0$ ；当中心点连线与 x 轴成 45 度时， $A=1$ 。

$$A = 1 - 2 \times \sin^2 \left(\arcsin(x) - \frac{\pi}{4} \right) \quad (3-3)$$

其中： $x = \frac{C_h}{\sigma} = \sin(\alpha)$ ， $\psi = \sqrt{(b_{Cx}^{gt} - b_{Cx})^2 + (b_{Cy}^{gt} - b_{Cy})^2}$ ， $C_h = \max(b_{Cy}^{gt}, b_{Cy}) - \min(b_{Cy}^{gt}, b_{Cy})$ ； b_{Cx}^{gt} ， b_{Cx} 表示真实框和预测框中心点的横坐标； b_{Cy}^{gt} ， b_{Cy} 表示真实框和预测框中心点的纵坐标； ψ 表示真实框与预测框中心点连线距离； α ， β 表示角度值； C_w ， C_h 表示真实框与预测框中心点的横纵距离。通过对角度实施惩罚，可以引导锚框（anchor box）最接近目标框的轴上，从而减少了预测框的总自由度数。

距离损失主要描述了中心点之间的距离，如式（3-4）所示，它的惩罚代价与角度损失呈正相关，其通过角度损失，重新定义了一个距离损失函数，如图 3-9 所示。

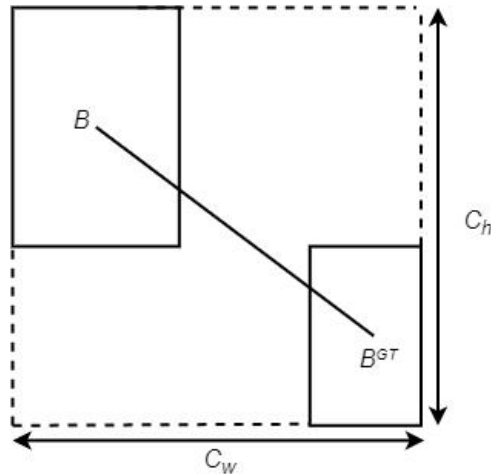


图 3-9 距离损失示意图

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho t}) \quad (3-4)$$

其中： $\rho_x = \left(\frac{b_{C_x}^{gt} - b_{C_x}}{c_w} \right)^2$, $\rho_y = \left(\frac{b_{C_y}^{gt} - b_{C_y}}{c_h} \right)^2$, $\gamma = 2 - \Lambda$; Λ 表示角度损失； Δ

表示距离损失； c_w , c_h 表示真实框与预测框最小外接矩阵的宽和高。当预测框与真实框近乎平行时， α 趋近于 0，此时角度损失接近 0， γ 接近 2，此时距离损失对整体损失的贡献迅速下降；当 α 趋近 $\frac{\pi}{4}$ 时，此时角度损失接近 1， γ 接近 1，此时对整体损失的贡献越大，此时两框之间的距离损失需要重视。

形状损失计算如式（3-5）所示，通过计算预测框与真实框宽或者长的差值与最大宽或长的比值，即考虑两框之间的长宽比来定义的。

$$\Omega = \sum_{t=w,h} \left(1 - e^{-\omega_t} \right)^\theta \quad (3-5)$$

其中 $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$, $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$; θ 表示控制对形状损失的关注程度，为了避免过于关注形状损失而降低对预测框的移动，设置范围为[2,6]； w , w^{gt} 表示预测框和真实框的宽； h, h^{gt} 表示预测框和真实框的高； Ω 表示形状损失。

交并比（IoU）损失，如公式（3-6）所示，主要是为了衡量预测框与真实框之间的重叠程度，形状损失和交并比损失示意图如图 3-10 所示。

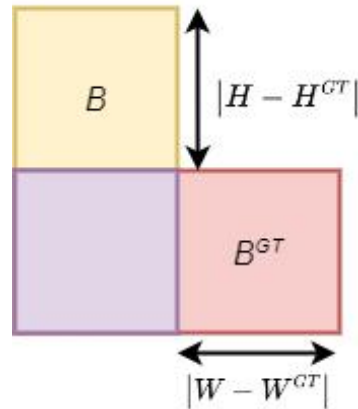


图 3-10 形状损失和交并比损失示意图

$$IoU = \frac{|B \cap B^{GT}|}{|B \cup B^{GT}|} \quad (3-6)$$

所以，SIoU 损失函数最终定义如式（3-7）所示。

$$SIoU_Loss = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (3-7)$$

3.3.4 SimSPPF 特征融合模块

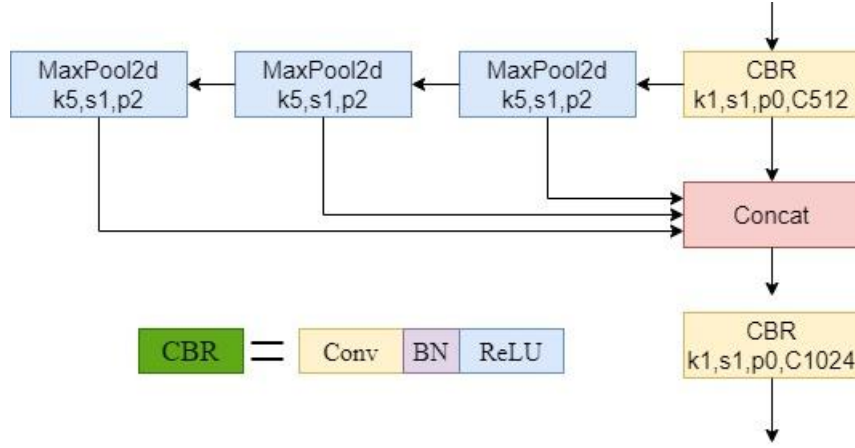


图 3-11 SimSPPF 模块

对于卷积神经网络，低层的卷积层感受野较小，小目标的语义信息非常丰富，得到的特征图适合检测小尺寸目标，相反，在高层中，感受野较大，大目标的语义信息比较丰富，得到的特征图适合检测大尺寸目标。在 YOLOv5 中，SPPF（Fast spatial pyramid pooling module）特征融合模块的作用是将不同尺度的特征进行融合，即将低级特征图和高级特征图进行融合，以提取更丰富的特征表示。由于 SPPF 中使用 SiLU 作为激活函数，需要消耗较多的计算资源，可能会导致梯度消失或者梯度爆炸，而 YOLOv6^[84]中的 SimSPPF 中使用 ReLU 作为激活函数比较稳定，训练也更加高效。故使用 SimSPPF 特征融合模块，具体结构如图 3-11 所示。

3.3.5 改进的 YOLOv5 模型

本研究使用 YOLOv5s 作为草莓目标检测模型所使用的基础网络，首先将主干网络和颈部网络中的卷积层使用 Ghost 卷积，同时将 C3 模块中的瓶颈结构（BottleNeck）使用 Ghost BottleNecks 进行替换，通过使用 Ghost 卷积和 Ghost BottleNecks，减少网络由于特征提取而造成的参数量过多，降低模型复杂度；将 YOLOv5 的定位损失由 CIoU 替换为 SIoU，考虑边界框回归的向量角度，增加约束；然后在 SPPF 模块前加入高效通道注意力机制 ECA，提升模型对草莓

的关注能力；最后，将快速空间金字塔模块 SPPF 替换为 SimSPPF 模块，进一步降低训练所需的资源，减少模型复杂度，提高检测精度。改进的草莓检测模型整体结构如图 3-12 所示。

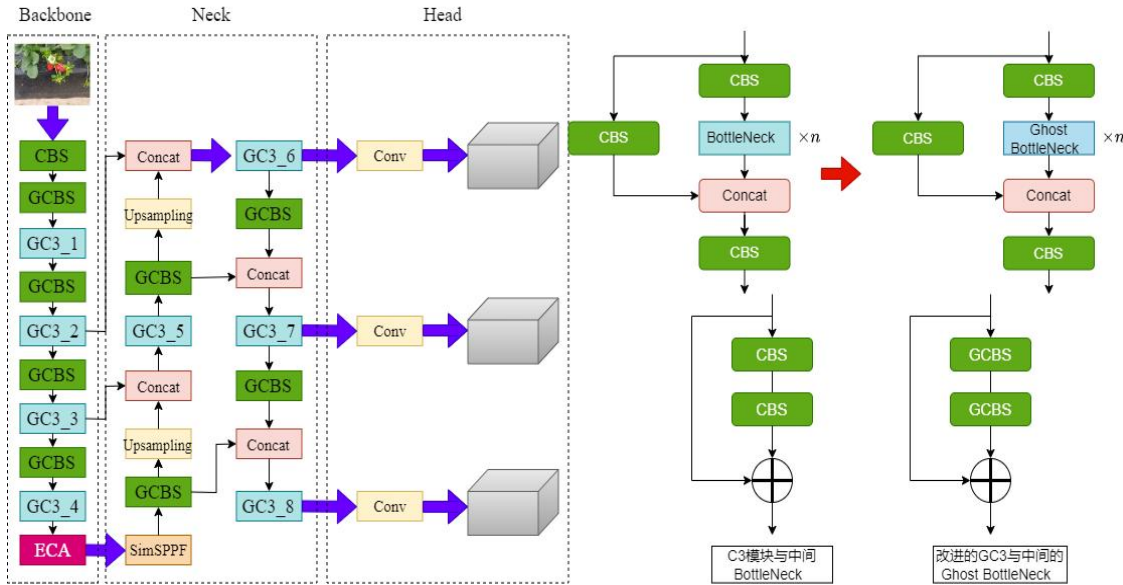


图 3-12 改进的草莓目标检测模型

3.4 实验结果与分析

3.4.1 训练策略和实验平台

表 3-2 硬件和超参数配置

名称	配置
CPU 配置	Intel(R) Xeon(R) W-2255 CPU @ 3.70GHz
显卡型号	NVIDIA RTX A4000
显存	16GB
迭代次数	300
批大小	36
训练图片大小	640×640
优化器	SGD
学习率	0.01
权重衰减	0.0005
动量	0.937

本次实验基于 Python3.8.13、CUDA11.7 和 PyTorch1.13.0 搭建深度学习框架，将制作好的草莓数据集输入到改进的 YOLOv5 模型中进行训练，具体硬件配置和超参数配置如表 3-2 所示。

3.4.2 评价指标

本文主要采用精确率（Precision，P），召回率（Recall，R）和平均精度均值（mean Average Precision，mAP）等作为评估指标来评价目标检测模型的性能，其中，精确率指模型预测正确的正样本个数除以预测为正样本的个数；召回率指模型预测正确的正样本个数除以实际为正样本的个数。精确率、召回率和平均精度均值计算如式（3-8）、（3-9）和（3-10）所示：

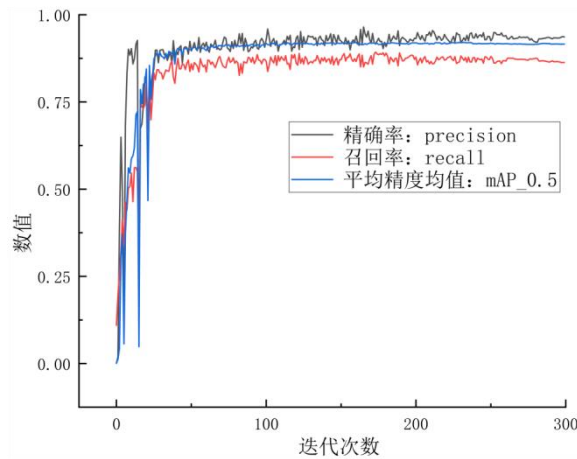
$$Precision = \frac{TP}{TP + FP} \quad (3-8)$$

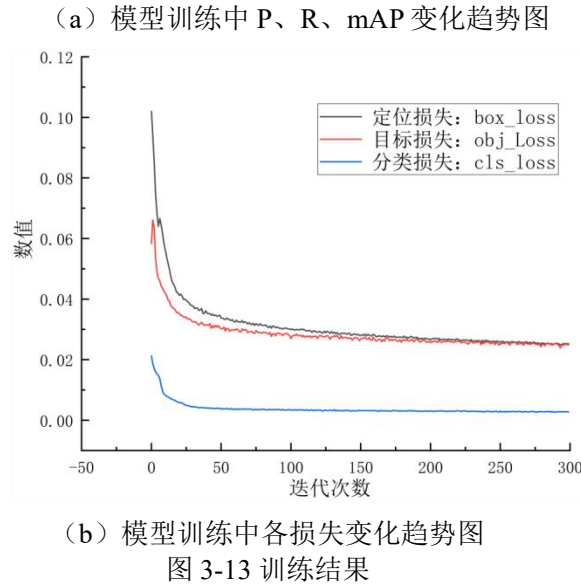
$$Recall = \frac{TP}{TP + FN} \quad (3-9)$$

$$mAP = \frac{\sum_{i=1}^C \int_0^1 P_i dR_i}{C} \quad (3-10)$$

其中， $\int_0^1 P_i dR_i$ 表示 P-R 曲线的面积，即平均精度（Average Precision，AP）；C 表示检测的类别总数；TP 表示模型将正样本预测为正样本的个数；FP 表示模型将负样本预测为正样本的个数；FN 表示模型将正样本预测为负样本的个数。

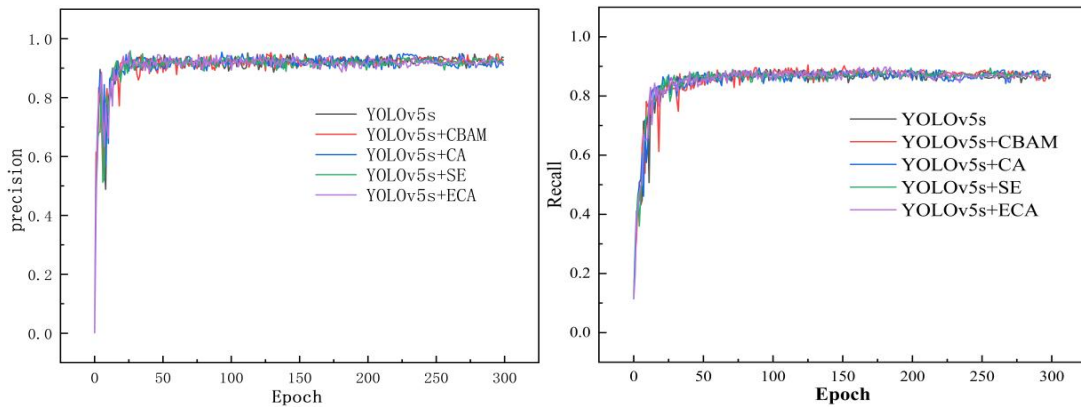
改进模型共训练了 300 轮次，训练初期，模型学习效率较高，精确率、召回率和阈值为 0.5 时的平均精度均值（mAP_0.5）增长较快，模型的定位损失（box_loss）、目标损失（obj_loss）和分类损失（cls_loss）值逐渐减小，大约迭代到 250 次，趋于稳定，模型收敛。300 次迭代后，模型的精确率为 93.64%，召回率为 86.33%，平均精度均值为 91.60%。图 3-13 为模型训练各值的变化趋势图。





3.4.3 改进算法性能测试

考虑到模型检测精度和模型复杂度，本文引用了高效通道注意力模块 ECA，并与当前比较先进的注意力机制进行对比，实验结果如图 3-14 和表 3-3 所示，可知，添加注意力机制后 YOLOv5s 的检测精度得到了一定的提升，其中，添加 SE 注意力机制和 ECA 注意力机制提升的最多，均为 93.5%，相较于 YOLOv5s，mAP_{0.5} 提升了 1.0%。使用 ECA 注意力机制时的模型参数量和浮点运算数较 SE 注意力机制分别少了 32765 和 0.1，添加后的检测模型规模最小，故本文使用高效通道注意力机制 ECA 作为研究模块。



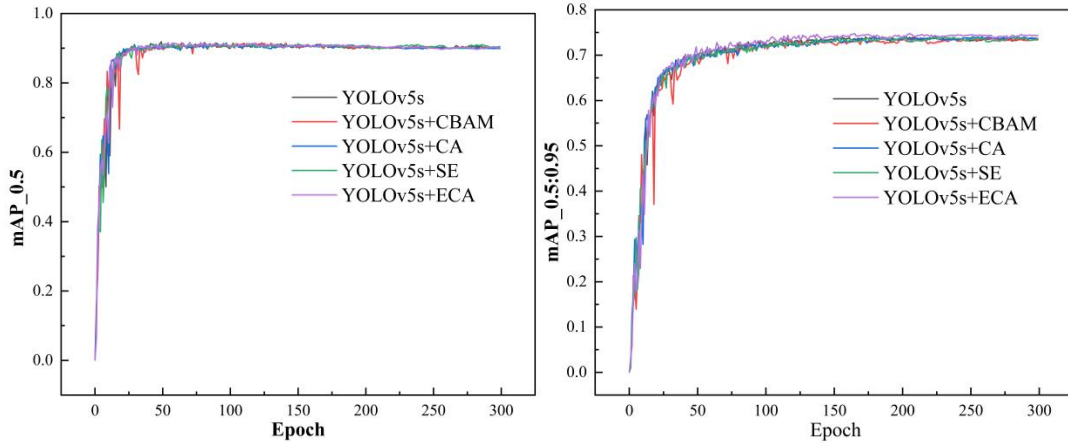


图 3-14 不同注意力模块训练对比图

表 3-3 不同注意力模块对比实验

模型	P/%	R/%	mAP@0.5	模型参数量	浮点运算数
YOLOv5s	91.4	89.6	92.5	7027720	15.9
YOLOv5s+CBAM	90.5	90.9	93.1	7236714	16.3
YOLOv5s+CA	91.7	91.1	93.3	7229496	16.3
YOLOv5s+SE	91.2	91.7	93.5	7236616	16.3
YOLOv5s+ECA	92.9	90.5	93.5	7203851	16.2

为了验证本文改进算法的性能，分别对 YOLOv5 的主干网络和颈部网络使用 Ghost 卷积、SIoU 定位损失函数替换、添加高效通道注意力机制 ECA 和替换特征融合模块为 SimSPPF 的消融实验，同样使用自制的草莓数据集集中的测试集进行测试，实验结果如表 3-4 所示。

试验使用了参数量和浮点运算数来衡量模型的复杂度，两者数值越小，表示网络复杂度越低，如表 3-4 所示，采用 Ghost 模块后，相较于 YOLOv5，模型参数量降低了大约 47.5%，浮点运算数降低了大约 49.1%，说明 Ghost 模块具有降低模型复杂度，轻量化网络，较为显著提升网络性能的作用。

由表 3-3 和表 3-4 数据可以看出，使用高效通道注意力机制 ECA 相比于 YOLOv5 网络 P 值上升了 1.5%，R 值上升了 0.9%，mAP 值提升了 1.0%，虽然参数量和浮点运算数有所增加，但是检测精度得到了提升。说明 ECA 模块能够提升模型对草莓的关注，抑制无用特征，有效提高网络性能。当使用 SIoU 作为定位损失函数时，P 值虽降低 0.2%，但 R 值提升了 1.1%，mAP 值提升了 0.1%，

说明 SIOU 损失函数能够通过考虑边界框回归角度，从而增强定位性能，可小幅提升模型的检测精度。当将 YOLOv5 中的 SPPF 替换为 SimSPPF 后，mAP 提升了 0.2%。综合四种方法，改进模型的 P 值、R 值和 mAP 优于 YOLOv5 基础网络，提升了 3.1%、0.9%和 2.4%，分别为 94.5%、90.5%和 94.9%，计算时的模型参数量减少了 3161653，浮点运算数降低了 7.4 GFLOPs，较为显著地提升了 YOLOv5 对草莓的检测精度，降低了模型大小和复杂度。

表 3-4 消融实验

主干网络	Ghost	SimSPPF	SIOU	P/%	R/%	mAP@0.5/ %	模型参 数量	浮点运算数 /GFLOPs
YOLOv5	√			91.2	88.8	92.2	3689936	8.1
		√		90.7	91.9	92.7	7027720	15.9
			√	91.2	90.7	92.6	7027720	15.9
	√			92.5	90.7	93.6	3866067	8.5
YOLOv5 +ECA		√		90.5	91.7	93.8	7203851	16.2
			√	92.1	90.2	93.5	7203851	16.2
	√	√		93.8	88.0	93.0	3866067	8.5
	√		√	93.4	89.6	94.1	3866067	8.5
		√	√	92.2	88.3	93.1	7203851	16.2
	√	√	√	94.5	90.5	94.9	3866067	8.5

为了评价本文改进模型的综合性能，改进 YOLOv5 模型检测结果如图 3-15 所示，直角坐标系中横坐标是 R 值，纵坐标是 P 值。曲线下方围成的面积是每个类别的平均精度 AP，草莓花、未成熟的草莓和成熟的草莓的 AP 值分别为 89.3%、96.7%和 98.7%，mAP@0.5 为 94.9%，对草莓花检测精度较低主要是由于一些未成熟草莓中带有一两片未脱落的花瓣，造成干扰，后续会尝试增加草莓花图片数量，提升模型对草莓花特征的学习能力。

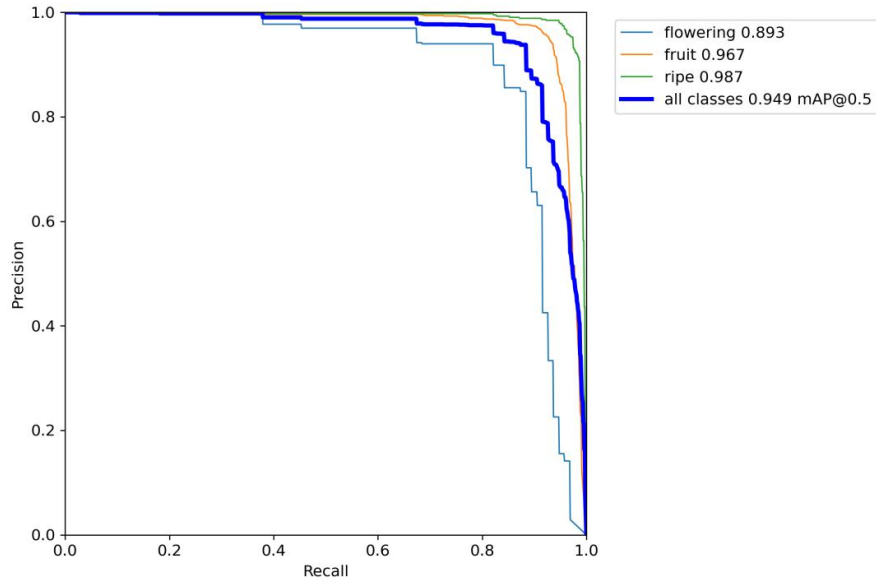


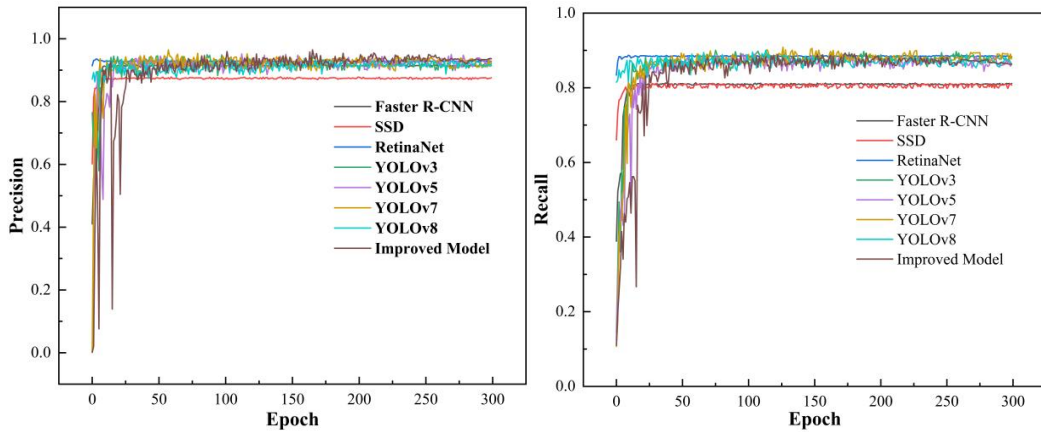
图 3-15 改进模型 P-R 曲线

表 3-5 综合方法对比

模型	P/%	R/%	mAP@0.5/%	模型大小/MB
Faster R-CNN	93.1	76.9	93.1	629.0
SSD	83.7	65.7	83.7	102.0
RetinaNet	92.7	77.6	92.7	245.0
YOLOv3	91.4	90.9	92.8	117.0
YOLOv5	91.4	89.6	92.5	13.7
YOLOv6	89.3	79.3	89.3	38.7
Yolov7	93.9	87.3	87.6	71.3
Yolov8	92.2	86.8	91.7	21.4
DETR	91.7	78.7	91.7	473
改进模型	94.5	90.5	94.9	7.76

本文还与其它经典目标检测模型进行了对比实验，如表 3-5 和图 3-16 所示。可以看出改进模型具有最高的平均精度均值，为 94.9%，比 Faster R-CNN、SSD、RetinaNet、YOLOv3^[85]和 YOLOv5 分别多出 1.8%、11.2%、2.2%、2.1%、2.4%，具有最高的检测精度；在模型大小方面，改进 YOLOv5 模型比 Faster R-CNN、SSD、RetinaNet、YOLOv3 和 YOLOv5 分别减少了 621.24MB、94.24MB、237.24MB、109.24MB 和 5.94MB。本文还与最新的检测模型 YOLOv6、YOLOv7、YOLOv8 和 DETR（Detection Transformer）进行对比，由表 3-5 可知，改进模型的 P 值分别多出 5.2%、0.6%、2.3%、2.8%，为 94.5%。R 值分别多出 11.2%、3.2%、3.7%、11.8%，为 90.5%。mAP 值分别多出 5.6%、7.3%、3.2%、3.2%，模型大小仅为 7.76MB，拥有最高的检测精度和最低的模型规模，对移

动式和嵌入式的硬件要求最低。



3-16 不同检测算法对比

不同方法检测结果如图 3-17 所示，蓝色箭头指向算法漏检的目标，黄色箭头指向算法错误检测或者边界框拟合较差的目标，从检测效果可以看出，RetinaNet、SSD、YOLOv5 和 YOLOv8 网络对遮挡目标和小目标都出现了一定程度的漏检现象；Faster R-CNN、SSD、YOLOv5 和 YOLOv8^[86]在对遮挡的成熟草莓进行识别时，检测框的面积无法较为准确地拟合遮挡的草莓外表；同时，RetinaNet 会将上下遮挡同状态的两个草莓目标错误地检测为三个目标，SSD 和 YOLOv5 会错误地检测为一个目标，而改进后的 YOLOv5 模型不仅减少了漏检现象，而且正确地将上下遮挡的同状态的草莓目标进行了检测框的预测，预测后的检测框比较符合遮挡目标，得到了较为准确的检测框拟合结果。由前面的对比实验和图 3-17 的检测结果对比，证明了本文改进模型对草莓拥有较为优越的检测性能。



(a) Faster R-CNN



(b) SSD



(c) RetinaNet



(d) YOLOv5



(e) YOLOv8



(f) 本文改进模型

图 3-17 模型检测效果

3.5 本章小结

针对草莓目标，目前大部分检测模型存在着检测精度较低、模型复杂度高和检测效果差等问题，本文提出一种基于 YOLOv5 的改进方法，实现了对草莓的高效检测。为了提高模型检测速度，降低复杂度，通过将 YOLOv5 主干网络和颈部网络中的卷积层替换为 Ghost 卷积，同时将 C3 模块中的瓶颈结构使用 Ghost 瓶颈结构进行替换；为了提高模型对草莓特征的关注，使用高效通道注意力模块 ECA 对网络模型进行了优化，将定位损失函数由 CIoU 替换为 SIoU，并将特征融合模块由 SPPF 改为 SimSPPF，提升了模型的检测能力。实验结果表明：相较于目前一些经典的目标检测模型，本文所提的改进 YOLOv5 算法平均精度均值可以达到 94.9%，模型大小仅为 7.76MB，具有更高的检测精度、更低的模型规模和更好的检测效果，验证了本文改进算法用于草莓检测的有效性和优越性。

第 4 章 草莓定位方法研究

前面一章已经实现了对不同生长状态草莓的目标检测，要想实现农业机器人对草莓的自动化采摘，必须计算出草莓的位置，本章将研究使用传统的图像处理，实现草莓的二维定位，使用被动双目相机和主动深度相机，实现对草莓目标的三维定位。

4.1 图像处理二维定位研究

4.1.1 图像预处理

图像预处理，作为图像处理流程中的核心环节，旨在滤除图像中无关紧要的信息，还原并凸显出关键的真实信息，增强这些信息的可辨识度，并尽量简化数据构成。通过这一系列的处理，图像的特征提取、分割、匹配以及识别等后续操作的准确性得以显著提升。通常，图像预处理的过程包含以下几个步骤：

(1) 灰度化：灰度化是指将 RGB 彩色图像转换为灰度图像的过程。当 R（红色）、G（绿色）和 B（蓝色）三个分量的值相等时，该颜色即为灰度颜色。此时， $R=G=B$ 的值被称为灰度值，其取值范围在 0 至 255 之间。为了实现彩色图像的灰度化，常用的方法包括分量法、最大值法、平均值法和加权平均法。这些方法均能有效地将彩色图像转换为灰度图像。

(2) 几何变换：可以纠正图像的几何失真，如扭曲、拉伸、旋转等。

(3) 图像增强：通过增强图像的对比度、清晰度等，使图像更加清晰、易于处理和理解。

在图像预处理中，还有一些常用的技术，如滤波、平滑、锐化等，这些技术可以进一步提高图像的质量和清晰度，减少噪声和干扰，以便更好地提取特征和信息。

4.1.2 颜色空间

同 RGB 颜色空间相比，Lab 颜色空间并不常见。Lab 颜色空间是一个全面覆盖人眼可见所有颜色的色彩空间，因此其色域极为宽广。在此色彩空间中，每一组色值均唯一对应一种特定的、与设备无关的色彩。这种色彩空间通过数

学理论进行量化，使得不同设备之间的色彩能够进行相互比较、模拟和匹配。在 Lab 颜色空间中，颜色由三个参数表示，分别是 L（亮度）、a 颜色和 b 颜色。L 通道没有颜色，a 通道和 b 通道只有颜色。Lab 颜色空间中的 L 分量，其数值域跨度为[0,100]，这代表着从深邃的黑色渐变至纯净的白色。而 a 分量，其数值域为[127,-128]，则涵盖了从热烈的红色过渡到静谧的绿色的色域。同时，b 分量，其数值域同样为[127,-128]，则代表着从鲜明的黄色渐变至深邃的蓝色的色域。这三个分量共同构成了 Lab 颜色空间，正为暖色，负为冷色，它具有如下特点：（1）基于人对颜色的感觉，更符合人类视觉；（2）亮度 L 和颜色 a、b 分离，方便调节亮度、色彩平衡等；（3）色域宽，适合所有颜色；（4）与设备无关，可以相互比较、模拟和匹配不同设备的色彩。

在本文使用中，需要借助 XYZ 颜色空间将 RGB 颜色空间的图片转为 Lab 颜色空间，首先在 RGB 颜色空间中，对图片中各像素点进行归一化，转换到 XYZ 颜色空间，之后在 XYZ 颜色空间中，通过白色参考点归一化的结果，转换到 Lab 颜色空间。RGB 与 XYZ 颜色空间转换关系如式（4-1），其中 R、G、B 为对应像素值除以 255 归一化得到的值，范围为[0,1]。

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4-1)$$

在 XYZ 颜色空间中，对其进行归一化，如公式（4-2）所示：其中， $X_n = 0.95047$ ， $Y_n = 1.0$ ， $Z_n = 1.08883$ 。

$$\begin{cases} x = \frac{X}{X_n} \\ y = \frac{Y}{Y_n} \\ z = \frac{Z}{Z_n} \end{cases} \quad (4-2)$$

XYZ 与 Lab 颜色空间非线性转换关系如式（4-3）所示，其中，t 代表 x，y，z。

$$t = \begin{cases} t^{1/3}, & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \left(\frac{1}{3}\right)\left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & \text{else} \end{cases} \quad (4-3)$$

XYZ 与 Lab 颜色空间线性转换关系如式（4-4）所示：

$$\begin{cases} L = 116y - 16 \\ a = 500(x - y) \\ b = 200(y - z) \end{cases} \quad (4-4)$$

4.1.3 图像分割

图像分割指的是通过一系列技术和过程，将一幅图像细致地划分为若干特定区域，每个区域均具有其独特的性质。此过程不仅涉及图像的基本处理，更关键的是为后续的图像分析提供了重要的基础。在图像处理领域，图像分割被视为由处理迈向分析的桥梁和纽带。在图像分割中，人们通常只对部分目标感兴趣，这些目标通常占据一定的区域，并且在某些特性上和临近的图像有差别，这些特性差别可能非常明显，也可能很细微，以至肉眼察觉不出来。因此，图像分割的目的是将这些感兴趣的目标从图像中分离出来，以便更好地提取特征、进行目标跟踪、分类等后续任务。现有的图像分割方法主要包括以下几类：基于阈值的分割、基于区域的分割、基于边缘的分割以及基于特定理论的分割。这些方法各自具有独特的优势和适用场景，在图像处理领域发挥着重要作用。

（1）基于阈值的分割方法：这种方法主要是通过设定一个阈值，将图像分成两部分，一部分的像素值大于阈值，另一部分的像素值小于或等于阈值。这种方法简单易行，但对于复杂的图像或噪声较多的图像，效果可能不太理想。

（2）基于区域的分割方法：这种方法主要是通过将图像分成若干个区域，并对每个区域进行特征提取和分类。这种方法对于复杂的图像或噪声较多的图像，效果可能更好。

（3）基于边缘的分割方法：这种方法主要是通过检测图像中的边缘来分割图像，边缘是指图像中不同区域之间的交界线，通过检测这些交界线，可以将图像分割成不同的区域。

（4）基于特定理论的分割方法：这种方法主要是通过应用特定的理论或算

法来进行图像分割，例如，基于图割的方法、基于小波变换的方法等。

在图像分割的过程中，还可以使用诸如颜色、纹理、形状等特征来构建区域。对于像素的邻近区域，可以利用图割方法将图像分割成两个或多个区域，图割的基本思想是相似且彼此相近的像素应该划分到同一个区域。

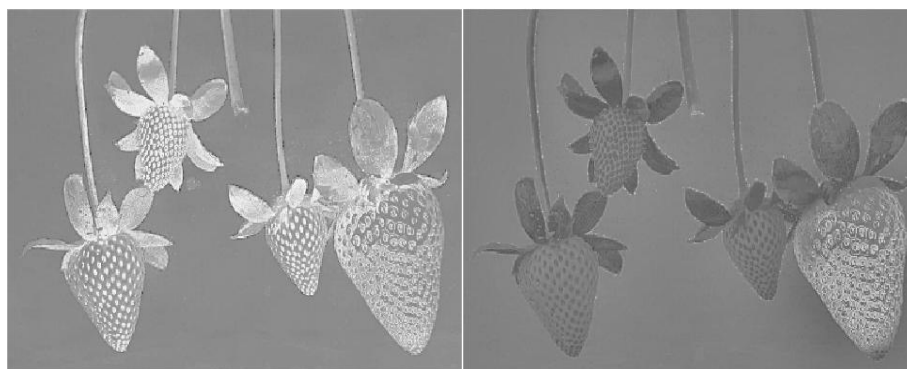
4.1.4 成熟草莓的二维定位



图 4-1 二维定位流程

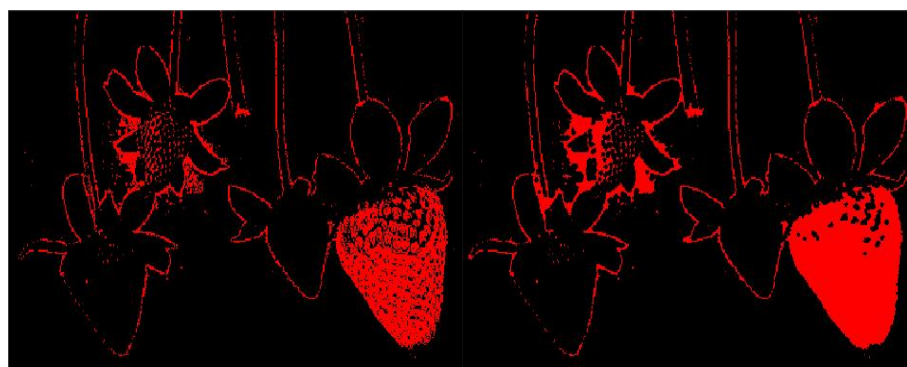
为了通过图像的像素处理实现草莓的二维定位，本文通过图像预处理改变图片颜色空间从而进行图像分割等步骤实现了对草莓的二维定位，使用的工具为机器视觉软件 Halcon。实现步骤如图 4-1 所示。

首先通过 `emphasize()` 函数对原图片进行图像增强，增加图片对比度；然后通过 `decompose3()` 和 `trans_from_rgb()` 函数将 RGB 颜色空间转为 Lab 颜色空间，从而增加图片的色域宽，提升草莓图片处理的效果，如图 4-2a 所示；再对输出的图片，使用 `illuminate()` 增强图像对比度和图像的高频区域，即边缘和拐角，使图像看起来更清晰，如图 4-2b 所示；使用 `threshold()` 进行阈值分割，阈值范围设置为[140,255]，分割出成熟草莓的大致轮廓，如图 4-2c 所示；虽然通过阈值分割得到了成熟草莓的大致轮廓，内部却充满了孔洞，会对后续的图像处理造成干扰，这时需要使用 `closing_circle()` 函数进行形态学闭运算，消除孔洞，如图 4-2d 所示；由于分割之后的区域为一个整体，这时我们需要使用 `connection()` 函数进行区域打散，如图 4-2e 所示；对于不同区域，需要选择出成熟草莓所在的区域，使用 `select_shape()` 进行面积的阈值分割并进行圆相似度选择，分割出单个成熟草莓目标，如图 4-2f 所示；使用 `area_center()` 和 `gen_cross_contour_xld()` 指定目标中心并使用蓝色十字准星进行显示，如图 4-2g 所示，从而完成了对成熟草莓目标的二维定位。



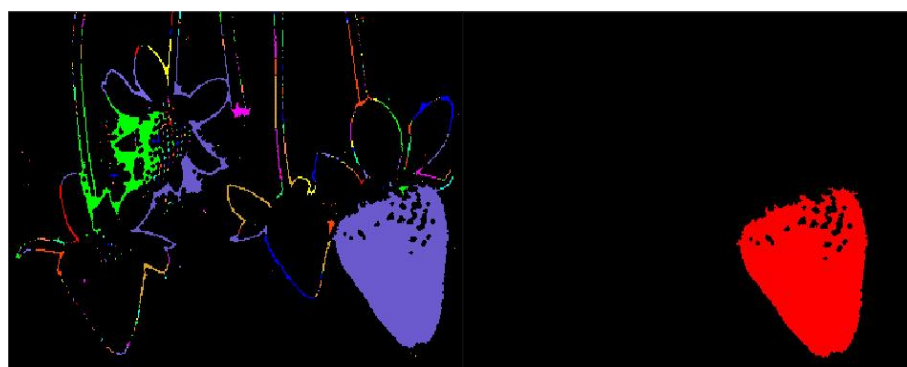
(a) Lab 颜色空间

(b) 图像增强



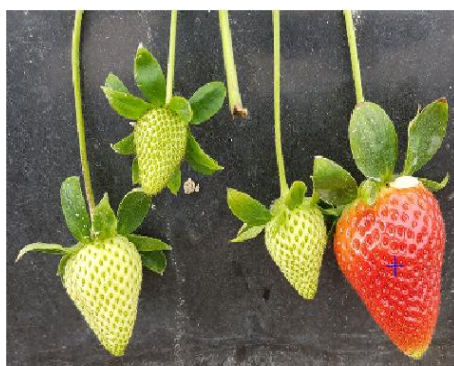
(c) 阈值分割

(d) 形态学闭运算



(e) 区域打散

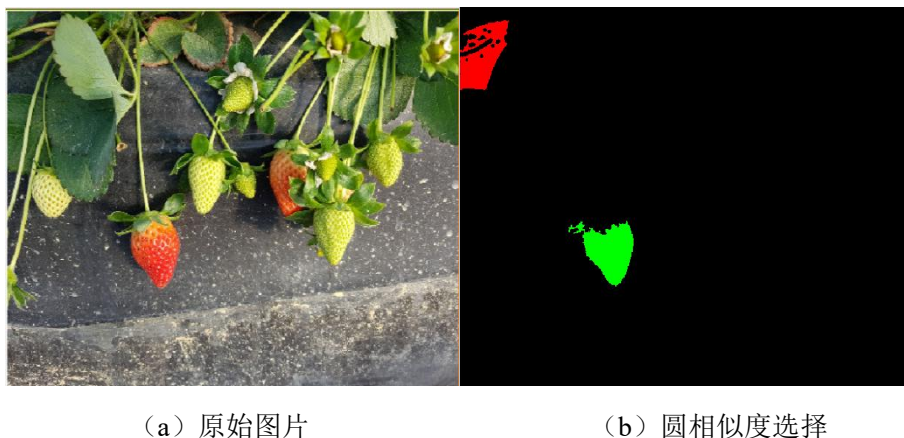
(f) 圆相似度选择



(g) 十字准星定位

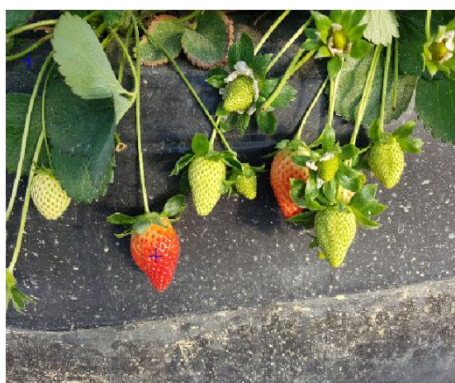
图 4-2 简单环境下草莓图片图像处理效果

基于 Lab 颜色空间的图像处理方法虽然对未遮挡的成熟草莓目标具有较好的二维定位的效果，但是当处于复杂条件下，例如成熟草莓被遮挡、草莓数量较多等时，就会出现误定位的情况，不具有泛用性，鲁棒性较差，如图 4-3a，4-3b，4-3c 所示。故本文使用 YOLOv5 首先进行草莓图片的检测识别，回归出对应类别的边界框，并对类别和置信度进行显示，然后以边界框的中心点为主要参考点，对图片中草莓进行二维定位。



(a) 原始图片

(b) 圆相似度选择



(c) 十字准心定位

图 4-3 复杂条件下草莓图片图像处理效果

4.2 双目视觉原理

4.2.1 相机成像原理

相机的镜头由一组透镜组合而成。当光线与主光轴保持平行状态并穿越这些透镜时，它们会精确地汇聚于一点，这个点被定义为焦点。焦距 f ，作为一个关键参数，精确地描述了焦点与透镜中心之间的空间距离。相机的镜头相当于

凸透镜，感光元件处在凸透镜的焦点附近，当焦距 f 近似为凸透镜并且到感光元件的距离为 s 时，便是相机成像采用最多的小孔成像模型，如图 4-4 所示，在此模型下，物体的三维空间坐标与二维图像坐标之间满足线性关系。

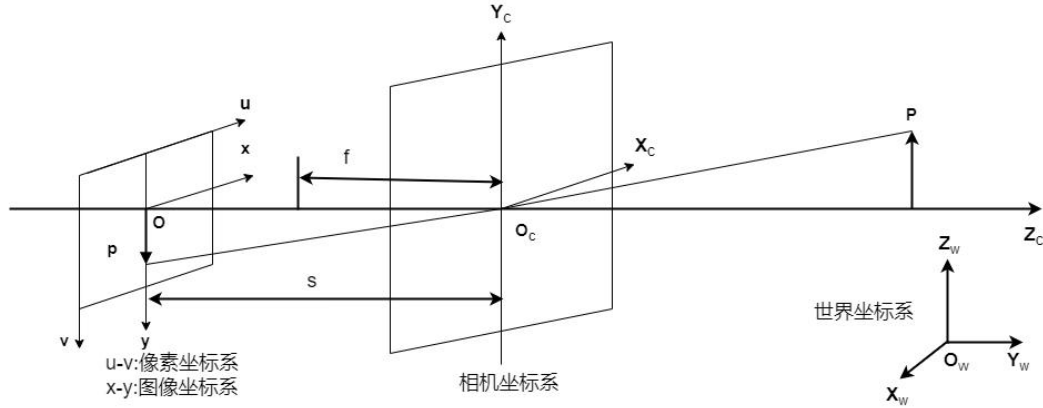


图 4-4 相机成像原理

相机的成像过程涉及一系列精确的坐标转换步骤，包括刚体变换、透视投影、畸变矫正以及二次转换。这些步骤涉及世界坐标系、相机坐标系、图像坐标系和像素坐标系之间的复杂关系，确保了从现实世界到图像像素的精确映射。转换关系如图 4-5 所示。



图 4-5 坐标系转换关系

(1) 刚体变换，即世界坐标系到相机坐标系的变换，此时物体不会发生形变，只需要进行旋转和平移，即通过旋转矩阵 R 与平移向量 T ，就能将世界坐标系中的 $P_w(X_w, Y_w, Z_w)$ 转换为相机坐标系中的 $P_c(X_c, Y_c, Z_c)$ ，转换关系如式 (4-5) 所示。

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = R \begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} + T \quad (4-5)$$

为了方便矩阵运算，通常将公式（4-5）转换为齐次坐标形式，如公式（4-6）所示。

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (4-6)$$

其中，旋转矩阵 R 为 3×3 矩阵，平移向量 T 为 3×1 矩阵，矩阵 $\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$ 被称为外参矩阵，此时单位为 mm。

（2）透视投影，即相机坐标系到图像坐标系的转换，如图 4-6 所示，此时，真实世界中的某点会投影在相机的成像平面上，利用针孔成像原理，在相机坐标系中，任意一点 P_c 与图像坐标系中的点 p 之间存在一种特定的关系。具体来说，当从点 P_c 引出一条连线至相机光心 O_c ，形成线段 O_cP_c ，该线段与像面的交点即为点 p 。点 p 即为空间点 P_c 在图像平面上的投影。

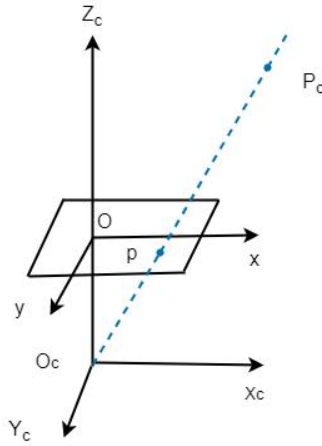


图 4-6 透视投影示意图

该过程为透视投影，利用相似三角形原理可得公式（4-7）。

$$\frac{x}{X_C} = \frac{y}{Y_C} = \frac{z}{Z_C} = \frac{f}{s} \quad (4-7)$$

其中， f 为焦距， s 为 Z 轴方向上 P_c 点图像平面的距离，展开可得式（4-

8)。

$$s \cdot x = Xc \cdot f, s \cdot y = Yc \cdot f, s \cdot z = Zc \cdot f \quad (4-8)$$

将公式 (4-8) 转换为矩阵形式, 则可得投影关系 (4-9)。

$$s \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Xc \\ Yc \\ Zc \\ 1 \end{bmatrix} \quad (4-9)$$

其中, $[Xc, Yc, Zc, 1]^T$ 为空间点 Pc 在相机坐标系 $OcXcYcZc$ 中的齐次坐标;

$(x, y, 1)^T$ 为像点 p 在图像坐标系 oxy 中的齐次坐标, 此时单位为 mm。

(3) 畸变矫正, 在摄影或成像过程中, 由于光学元件的形状、位置或材料等因素的影响, 图像中的像素点可能会偏离理想的规律分布, 导致图像失真, 这种现象被称为畸变。畸变主要分为径向畸变和切向畸变两类。为了提高图像质量和准确性, 必须对畸变进行矫正处理。

(4) 二次转换, 即图像坐标系与像素坐标系之间的转换, 如图 4-7 所示。两者均位于成像平面上, 但它们的原点和度量单位存在差异。图像坐标系的原点定义为相机光轴与成像平面的交点, 通常位于成像平面的中心位置。而像素坐标系的原点则通常位于成像平面的左上角, 其度量单位为像素 (pixel)。两者之间的转换关系如公式 (4-10) 所示。

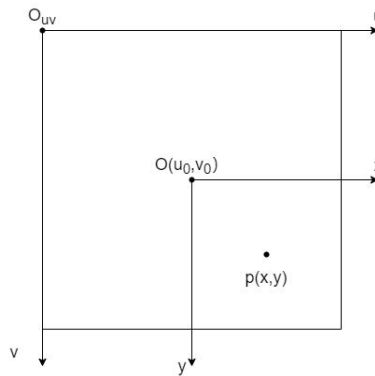


图 4-7 像素坐标系与图像坐标系

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \quad (4-10)$$

将公式（4-10）转换为矩阵表达式，得到公式（4-11）。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4-11)$$

其中，dx 与 dy 表示每行或者每列的长度，即 1 pixel=dx mm。

通过公式（4-6）（4-9）（4-11）可得世界坐标系转换为像素坐标系的总体公式，如式（4-12）所示。

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4-12)$$

在公式（4-12）中， $\begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ 为相机的内参矩阵，外参矩阵和内参矩阵具

体数值可通过张正友标定法获得。

4.2.2 双目视觉定位原理

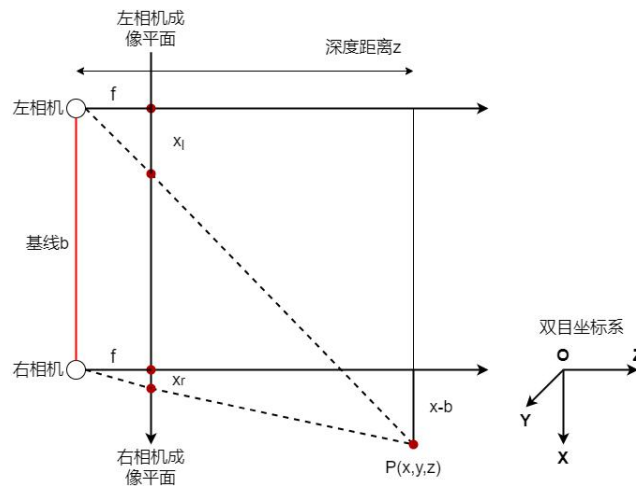


图 4-8 双目相机成像原理

双目立体视觉深度相机与人的双眼极为相似，其工作原理与基于 TOF、结构光原理的深度相机截然不同。它无需主动投射光源，而是依赖于拍摄的两幅

彩色或灰度图像来精确计算深度信息，因此亦被称作被动双目深度相机。双目相机通过对同一位置用两个不同视角的相机进行成像，两图像的深度信息不同，从而可以进行定位和距离测量，其具体原理如图 4-8 所示。

根据图 4-8，由相似三角形原理可得公式（4-13）。

$$\frac{z}{f} = \frac{x}{x_l} = \frac{y}{y_r} = \frac{x-b}{x_r} \quad (4-13)$$

由公式（4-13），求解方程可得公式（4-14）。

$$x = \frac{x_l z}{f}, y = \frac{y_r z}{f}, z = \frac{fb}{x_l - x_r} = \frac{fb}{d} \quad (4-14)$$

其中，d 表示左相机与右相机的视差；相机的焦距 f 和基线可通过相机的出厂制作参数得出，也可以通过张正友标定方法得到，由公式（4-14）就可以得出点 P 的三维坐标。

4.3 被动双目相机三维定位研究

4.3.1 实验设备

在实际的草莓采摘定位中，需要考虑多方面的因素，例如种植环境、光照、图片清晰度、相机拍摄定位速率和实验条件等。本文使用的是汇博视捷公司生产的 HBV-1780-2 S2.0 型号的双目相机如图 4-9 所示，具体各项参数如表 4-1 所示，它可以通过 USB 接口直接与笔记本电脑相连。



图 4-9 双目相机实物图

表 4-1 双目相机参数

参数名称	参数值
产品体积	80mm × 22mm × 24.6mm ± 0.5mm
传感器类型	CMOS
最大分辨率	2560 × 720
传感器像素	100 万
视场角/焦距	100 度 无畸变/2.1mm
帧速	30FPS
感光芯片	OV9732
支持系统	WinXP/7/8/10/11 Linux Android
工作温度	-30℃ ~ 70℃
接口	USB2.0
工作电压	DC 5V ± 5%

4.3.2 标定实验

为了实现草莓的三维定位，首先需要得到所用被动双目相机的外参矩阵、内参矩阵和畸变系数，才可以实现世界坐标系到像素坐标系的转换关系。本文将通过 Matlab 里的工具箱 Stereo Camera Calibrator，使用张正友标定法对双目相机进行标定，主要分为三个过程：采集标定板图像并转为灰度图片、识别标定图像和计算标定结果。

（1）采集标定板图像：

本文选用了 12×9 的棋盘格标定板，每个方格的尺寸为 15mm×15mm，如图 4-10 所示。

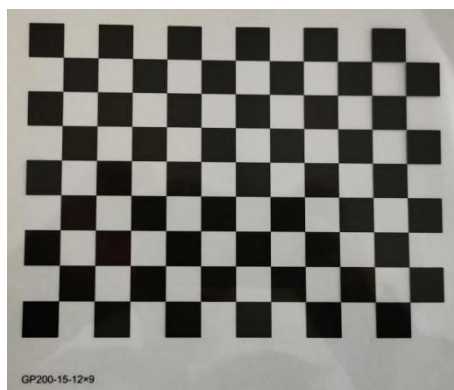


图 4-10 标定板图像

将标定板平铺在水平桌面上，使用双目相机通过改变不同位置和姿态进行

拍摄，共拍摄图片左右各 25 张并转为灰度图片，保留在不同的文件夹中，图片如图 4-11 所示。

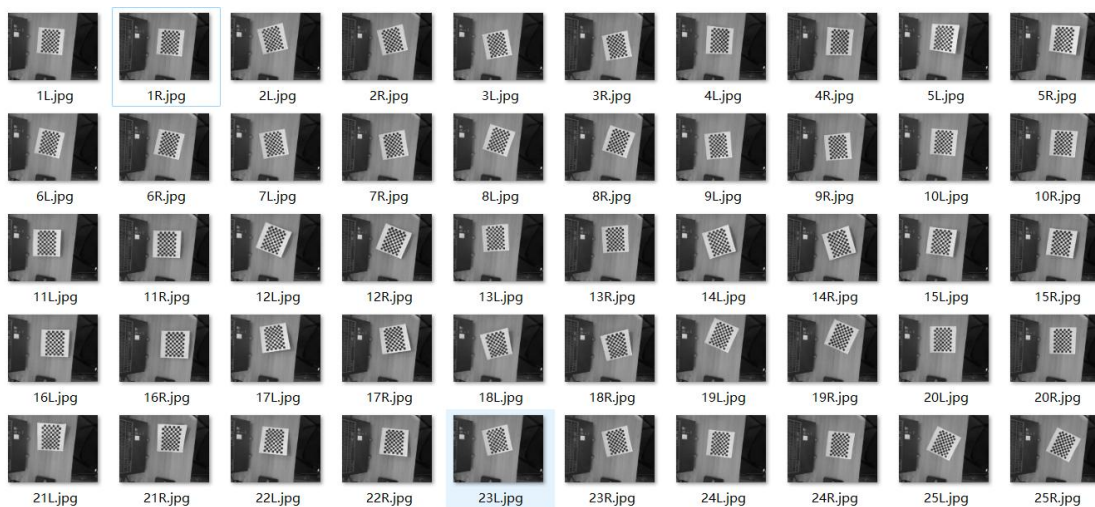


图 4-11 采集的标定板图片

(2) 识别标定图像：

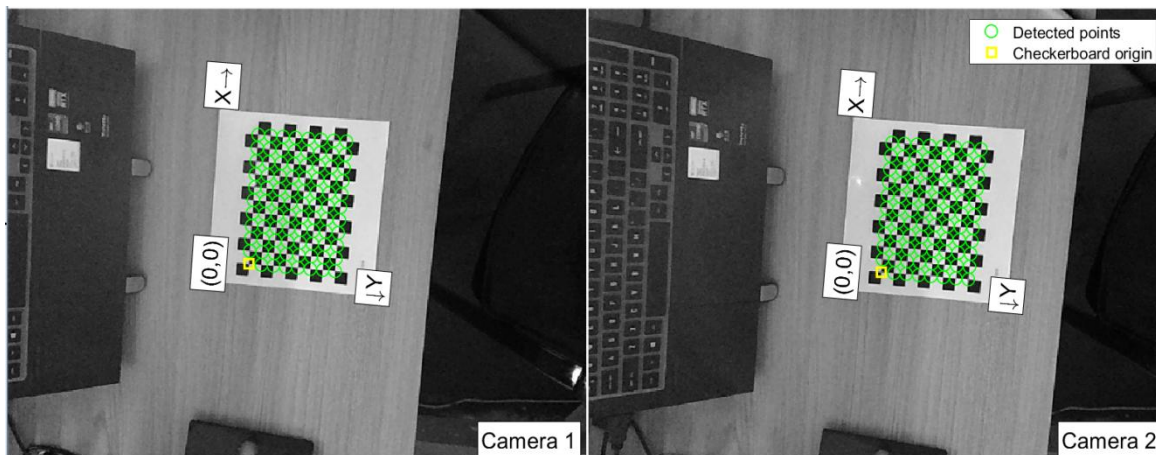


图 4-12 角点检测结果

打开 Matlab，使用 APP 中图像处理与计算机视觉中的 stereo Camera Calibrator 工具箱，勾选 3 Coefficients 和 Tangential Distortion 选项，导入左右相机采集的标定图片，设置棋盘格的尺寸为 15mm，载入后，界面会自动检测出标定板图片的角点，如图 4-12 所示。

(3) 计算标定结果：

点击 Calibrate 按钮，开始计算标定结果，标定结果如图 4-13 所示。

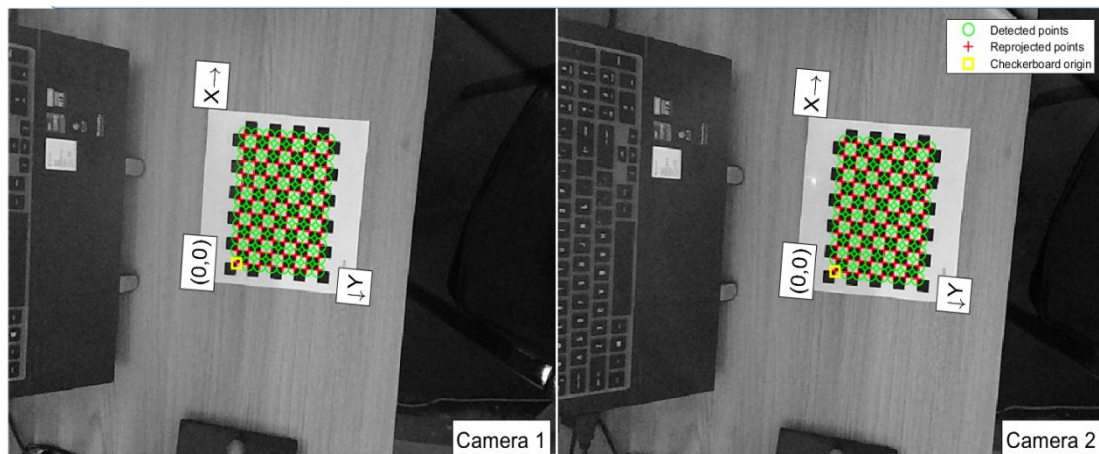


图 4-13 标定结果

标定完成后，会生成每组图片产生的重投影误差图，如图 4-14 所示，标定板位置空间图如图 4-15 所示。一般认为平均重投影误差低于 0.5 个像素代表标定效果较好^[87]，通过计算可得试验误差平均为 0.15 个像素，标定结果满足本课题要求。

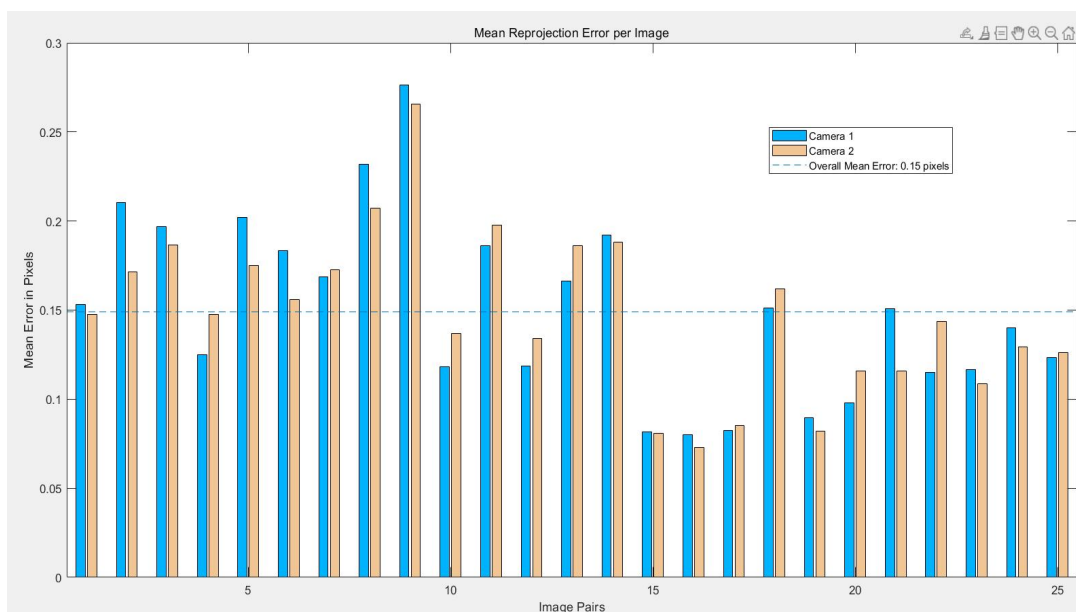


图 4-14 重投影误差

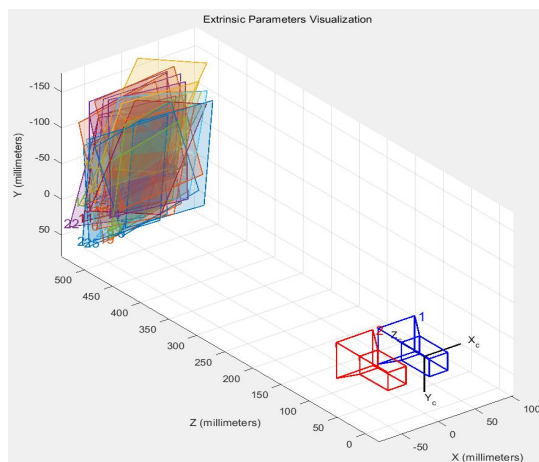


图 4-15 双目相机与标定板之间的位置关系图

点击 **Export Camera Parameters**，导出标定结果，即左相机与右相机的旋转矩阵、平移向量、内参矩阵和畸变系数，结果如表 4-2 所示。

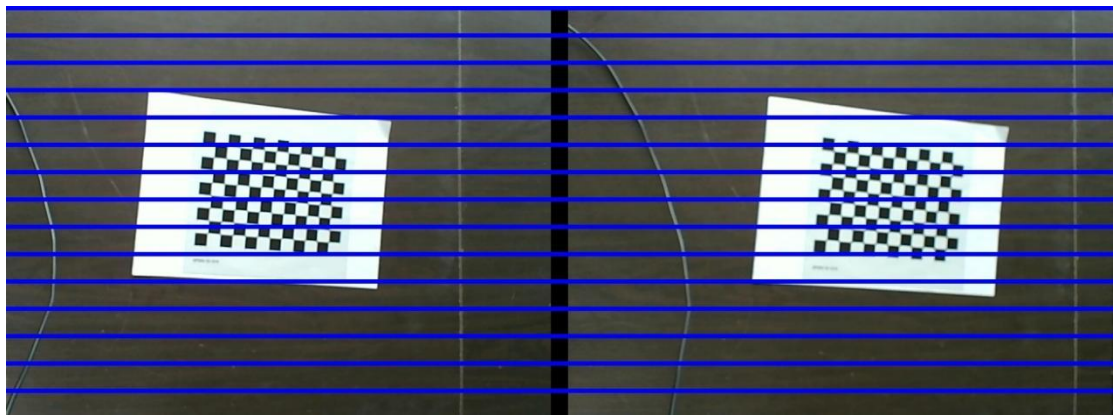
表 4-2 标定结果参数表

相机参数	左相机
	右相机
旋转矩阵	$\begin{bmatrix} 1.0000 & 2.2516 \times 10^{-4} & -0.0073 \\ -2.0550 \times 10^{-4} & 1.0000 & 0.0027 \\ 0.0073 & -0.0027 \times 10^{-2} & 1.0000 \end{bmatrix}$
平移向量	$[56.2199 \quad 0.0193 \quad -1.2566]$
内参矩阵	$\begin{bmatrix} 489.0515 & 0 & 0 \\ 0.2215 & 487.8746 & 0 \\ 301.3951 & 233.6376 & 1 \end{bmatrix}$
径向畸变	$[-0.0063 \quad 0.1734 \quad 1.2897]$
切向畸变	$[-0.0550 \quad 0.6159 \quad -1.1463]$ $[-0.0051 \quad 0.0080]$ $[-0.0030 \quad 0.0068]$

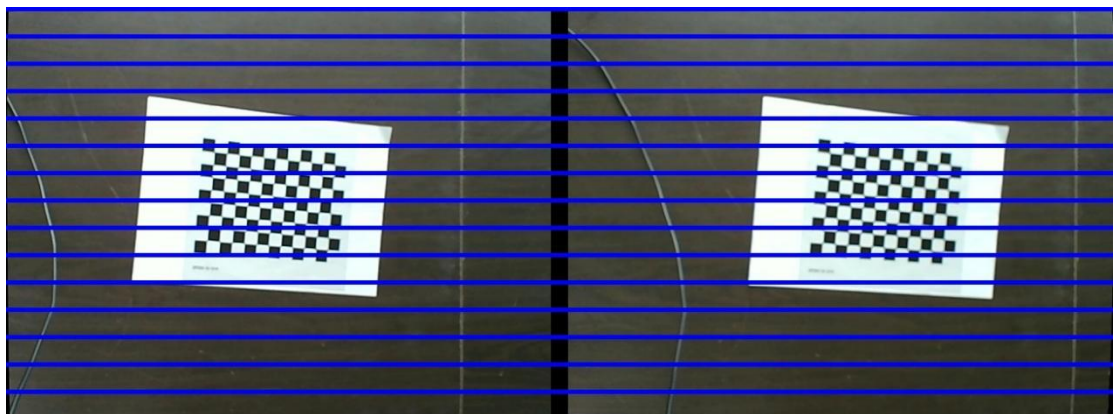
4.3.3 双目立体矫正

在理想的双目立体视觉系统中，可以通过公式（4-14）求解出视差，进而求出物体的三维坐标。但是，在实际中使用的双目立体视觉系统中，相机的制造和安装都会存在一定的误差，从而导致左相机与右相机的成像平面处于非平行状态，即产生了图像畸变。这时，左右相机拍摄的图像中的物体对应点不处于同一直线，会极大地增加立体匹配时的计算量，这时我们就需要对双目相机进行立体矫正，使其满足极线约束条件，然后进行立体匹配，将左右图像相应的点进行对应，进而获取视差值。

本文立体矫正使用了 `stereoRectify()`、`initUndistortRectifyMap()` 和 `remap()` 函数进行立体矫正，矫正图像如图 4-16 所示。



(a) 未立体矫正



(b) 已立体矫正

图 4-16 立体矫正结果

4.3.4 双目立体匹配

在对左右相机进行立体矫正后，两图像的 v 轴坐标相同，此时只需要考虑

u 轴方向对应像素点的匹配，通过立体匹配，计算出需要的视差，从而实现图中物体的三维定位。

在双目立体匹配中，使用的方法主要有三种，分别是 BM、SGBM 和 GC，在检测速度上，BM 最快，GC 最慢；在检测效果上，GC 最好，BM 最差。综合考虑检测速度和效果，SGBM 都处于中间水平，故本文将使用 SGBM 立体匹配算法，它是一种用于计算双目视觉中视差的半全局匹配算法，主要分为四个过程，分别为预处理、代价计算、动态规划和后处理，即首先通过水平 Sobel 算子对图像进行处理，如公式（4-15）所示，然后通过一个映射函数，如公式（4-16）所示，将前面处理后的图像上的每个像素点映射成一个新的图像，从而得到图像的梯度信息，其中 P 表示像素值， P_{NEW} 表示新图像上的像素值， $preFilterCap$ 为一个常数参数。

$$Sobel(x, y) = 2[P(x+1, y) - P(x-1, y)] + P(x+1, y-1) - P(x-1, y-1) + P(x+1, y+1) - P(x-1, y+1) \quad (4-15)$$

$$P_{NEW} = \begin{cases} 0 & , P < preFilterCap \\ P + preFilterCap & , -preFilterCap \leq P \leq preFilterCap \\ 2 * preFilterCap & , P \geq preFilterCap \end{cases} \quad (4-16)$$

代价计算有两部分组成：（1）将预处理阶段得到的图像的梯度信息经过采样的方法得到梯度代价。（2）将原图像经过采样的方法得到 SAD 代价，匹配算法 SAD 的基本思想是对经行对准后的左右视图图像的对应像素块的对应像素差的绝对值进行求和，如公式（4-17）所示，其中 d 表示视差；

$$C(x, y, d) = \sum_{i=-n}^n \sum_{j=-n}^n |L(x+i, y+j) - R(x+d+i, y+j)| \quad (4-17)$$

动态规划算法存在着拖尾效应，在视差突变处易产生错误的匹配，SGBM 利用动态规划在每个方向进行一维能量积累，会将错误的视差信息传播给后面的路径上，利用多个方向上的信息，试图消除错误信息的干扰，能明显减弱动态规划算法产生的拖尾效应，最后将各个方向上的匹配代价相加得到总的匹配代价，如公式（4-18）中， L 为当前路径累积的代价函数， $P1$ 、 $P2$ 为像素点与相邻点视差存在较小和较大差异情况下的平滑惩罚， $P1 < P2$ ，第三项无意义，

仅仅是为了消除各个方向路径长度不同造成的影响，将所有 r 方向的匹配代价相加得到总的匹配代价，如公式（4-19）所示。

$$L_r(p, d) = c(p, d) + \min \left\{ \begin{array}{l} L_r(p-r, d) \\ L_r(p-r, d \pm 1) + p_1 \\ \min_{i=d_{\min}, \dots, d_{\max}} L_r(p-r, i) + p_2 \end{array} \right\} - \min_{i=d_{\min}, \dots, d_{\max}} L_r(p-r, i) \quad (4-18)$$

$$s(p, d) = \sum_r L_r(p, d) \quad (4-19)$$

在动态规划完成后，SGBM 再进行唯一性检测、亚像素插值和左右一致性检测，完成双目的立体匹配，用 SGBM 立体匹配算法获得视差 d 后，就可以通过公式（4-14），获得目标在空间中的三维信息。

4.4 RGB-D 相机三维定位研究

RealSense D435i 是英特尔公司推出的一款消费级深度相机，它主要包含一个 RGB 相机、两个红外相机以及一个红外发射器，此外还有一个 IMU（惯性测量单元），可用于检测六个自由度的运动和旋转，能够改进其在摄像头移动情况下的深度感知，它的深度成像原理是主动立体红外成像，不是传统意义上理解的双目 RGB 相机成像，拥有自标定功能，RealSense D435i 相机如图 4-17 所示，它的主要参数如表 4-3 所示。



图 4-17 英特尔 RealSense D435i 相机

英特尔 RealSense D435i 的定位原理基于红外投射和结构光技术。红外投射技术通过发射红外光源并捕捉反射光来获取物体的深度信息，结构光技术则通

过发射结构化的光纹并捕捉光纹的形状和位置信息来实现三维重建和定位，这两种技术的结合使得 D435i 相机具备了高精度和高稳定性的成像能力。除此之外，D435i 亦运用三角测量原理以测定深度。其左右两侧的红外相机负责深度测量，而中央的红外点阵投射器则起到补光灯的作用，即便在不开启的情况下，仍可进行深度测量，但效果会有所降低。位于最右侧的 RGB 相机则负责采集彩色图像，进而实现彩色视频流与深度流的精确对齐。

表 4-3 英特尔 RealSense D435i 相机主要参数列表

参数名称	参数值
RGB 帧分辨率	1920×1080
深度输出分辨率	高达 1280 x 720
RGB 额定视场角 (H×V)	69° × 42°
深度额定视场角 (H×V)	87° × 58°
RGB 帧率	30 FPS
深度帧率	90 FPS

4.5 本章小结

本章详细地介绍了使用图像处理对草莓图片的二维定位、相机成像原理和双目相机定位原理。通过对汇博视捷 HBV-1780-2 S2.0 型号的被动双目相机进行标定，得到了左右相机的旋转矩阵、平移向量、内参矩阵和畸变矩阵，然后对其进行了立体矫正，使其满足极线约束条件，即左右图像处于相同水平线，减少定位过程中的计算量。然后，使用 SGBM 算法对左右相机产生的图像进行立体匹配，根据立体匹配获得的视差进行三维重建；对英特尔 RealSense 435i RGB-D 相机的定位原理和关键参数进行了介绍，从而为使用被动双目相机和 RGB-D 相机实现草莓的三维定位奠定了基础。

第 5 章 草莓检测模型系统搭建与边缘部署

为了对研究的草莓检测系统进行试验以及后期推广使用，本章在基于第三章提出的改进 YOLOv5 草莓目标检测算法上，设计了一个初步具备草莓检测功能的图像识别系统，便于产品的使用。首先对草莓检测系统的实际需求进行分析，然后对系统的整体框架和主要功能进行介绍，最后通过运行加载的算法进行系统性能测试，将本课题研究的草莓检测内容集成到软件模型中。最后，通过 Jetson nano 设备，对训练好的草莓检测模型进行边缘部署，并通过 Tensor RT 加速技术，进一步提高检测速度，可以更好地在实际采摘机器中进行应用。

5.1 系统需求分析

为了提升农业的智能化水平，使深度学习目标检测方法更好地为农业服务，可视化系统是必不可少的。本文设计了一个基于 YOLOv5 的草莓视觉检测系统，具有以下优点：（1）草莓检测系统通过采用图形方式显示用户操作界面，可以使用户更容易地查找、理解和使用；（2）草莓检测系统允许用户通过鼠标的方式来控制设备，从而取代基于文本的复杂命令，降低用户的学习难度；（3）草莓检测系统设计通过结合计算机科学、美学、心理学、行为学，及各商业领域需求分析的人机系统工程，符合市场需求；（4）当今市场中，草莓检测系统的 GUI 设计已成为软件应用程序编程中以用户为中心的设计标准，缺乏 GUI 的应用程序可能会显得过时且难以使用。

5.2 搭建环境介绍

本文的检测系统的开发、运行以及测试是在笔记本电脑上搭建的深度学习环境中完成的，具体硬件配置为 Intel(R) Core(TM) i7-11800H @ 2.30GHz；8GB 运行内存；6G 独显；内置英伟达 GeForce RTX3060 显卡进行加速图形化处理，操作系统为 Window11；采用的开发工具主要为 PyCharm，其中，PyCharm 是一种 Python 集成开发环境（IDE），由 JetBrains 开发，它专为 Python 编程语言而

设计，提供了一系列工具和功能，旨在提高 Python 开发人员的效率，具有代码编写、编译、分析和调试功能；软件环境主要包括开发语言 Python、Pytorch、Opencv 库和图形程序框架 PyQt5 并结合可视化开发工具 Qt Designer 设计和开发草莓检测系统。

5.3 系统功能介绍与测试

5.3.1 登录界面

对于本文的草莓检测系统设计了一个提供用户和管理员的登录界面，如图 5-1 所示。登陆时，可选择对应的身份、用户或者管理员，账号和密码都是“12345678”。选择用户登录时可以使用此系统对草莓的检测功能，实现对草莓花、未成熟草莓和成熟草莓的监控作用；当选择管理员身份登陆时，可以打开此检测系统的代码开发界面，对系统进行调试和修改。输入好账号和密码，选择身份后点击登录即可，点击退出则关闭登录界面。



图 5-1 登录界面

5.3.2 主界面

用户经过登录界面登陆后就可以进入到系统主界面，如图 5-2 所示，主界面主要包括模型选择模块、输入选择模块、交并比阈值调节模块、置信度阈值调节模块、延迟模块、存储模块、统计结果模块和详情显示模块。其中，模型

选择模块中包含的权重文件是已经用草莓数据集训练好的 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 和本文改进的 YOLOv5s 权重，可点击右边箭头下拉进行选择；输入模块主要选择输入模型文件设备类型，可以选择图片读取模块、视频读取模块和摄像头实时采集模块进行检测，通过点击相应的按钮进行选择，如果选择了文件按钮或者视频按钮，会弹出对话框，用户可选择图片或者视频资源进行添加；如果选择了摄像头按钮，系统会自动检测连接的摄像头个数，并给出索引，用户选择想要使用的摄像头索引即可打开对应的摄像头，进行实时检测。交并比调节模块和置信度调节模块可通过上下箭头或者黄色圆点进行拖动进行调节，进而适应不同的草莓生长环境，实现不同的检测效果和精度。存储模块中有一个自动存储点按钮，用户可点击进行选择是否对检测的图片、视频结果进行保存。统计结果模块主要是检测的草莓花、未成熟的草莓和成熟的草莓数量，并用数字进行统计。延迟模块主要是设置输入视频资源后帧的延迟时间，单位为毫秒，详情显示模块主要是显示原图像或者视频和它们的检测结果。



图 5-2 草莓检测系统主界面

5.3.3 草莓检测界面

首先输入图像、视频或者实时摄像到检测系统中，然后选择本文之前训练好的检测模型参数，并调整好用户所需要的交并比阈值、置信度阈值和延迟，点击运行按钮，系统就会自动调用用户选择的模型参数，对输入的文件进行推

理预测，将结果显示在原图像的右边，并在统计结果那里显示对应类别的数量，方便用户进行统计，如图 5-3 是系统对输入图片进行检测的运行示例图。



图 5-3 草莓检测界面

5.4 边缘 AI 介绍

边缘 AI 是指在边缘计算环境中实现的人工智能技术，它允许在设备附近进行计算，而不是在集中式云计算设施或异地数据中心进行计算，边缘 AI 结合了边缘计算和人工智能两种新兴技术，旨在支持在端点设备上处理数据和算法。通过在设备上执行人工智能计算，边缘 AI 可以加快设备的决策速度，并减少对互联网连接或云计算的依赖，这种本地化处理方式使得设备能够在几毫秒内做出决策，同时保证数据的隐私和安全性。总之，边缘 AI 技术的发展潜力巨大，在未来有望成为人工智能领域的重要发展方向之一。

经过第三章对草莓检测模型的优化后，需将其部署至边缘设备以适用于实际采摘场景。由于边缘设备与训练设备算力差异，直接部署可能导致实时性能差和检测耗时。本研究基于 Jetson nano，使用 TensorRT 加速技术优化草莓检测模型，以此提升模型在边缘设备的实时检测速度，为草莓采摘机器人的检测系统打下基础，进行模型的边缘部署实验，并对检测速度结果进行对比分析。

5.5 设备及环境搭建

5.5.1 基于 TensorRT 技术推理

TensorRT 是一个深度学习推理加速库，它可以为深度学习应用提供低延迟、高吞吐率的部署推理，是英伟达推出的一款高性能神经网络推理优化器，该优化器通过无需对模型结构进行额外更改的方式，实现了推理加速。该优化器可显著提升图像处理器的推理速度，并对应用于边缘设备的识别模型进行了有效优化。TensorRT 的优化策略涵盖调整数据精度、模型重构优化、内核调整、动态张量内存分配以及多输入流执行等多个方面。

在 TensorRT 中，主要的优化方法为层间融合或张量融合。这种优化方法主要针对计算图中制约计算速度的 CUDA 核心计算张量，通过减少卷积加载和内存读写，有效缓解内存带宽瓶颈和降低 GPU 资源浪费，具体而言，采用了层间横向与纵向合并的策略，显著减少了网络中的层数。这一改进不仅优化了计算流程，还提升了整体系统的稳定性和可靠性。通过横向合并，卷积层、偏置层和激活层可以被整合为一个单一的 CBR 结构（Convolution-Bias-ReLU），从而优化计算效率，仅需一个 CUDA 核心即可实现。而纵向合并则可将具有相同结构但权值不同的层合并为一个更宽的层，同样仅需一个 CUDA 核心，具体实现过程如图 5-4 所示。这种合并策略有助于提升计算效率，减少资源占用，优化深度学习模型的性能。

同时，TensorRT 还可以将高精度的存储精度降低，目前主流的深度学习框架，例如 PyTorch 和 TensorFlow，采用的是 FP32 或者更高的 FP64，经过深度卷积神经网络训练的目标检测模型，其模型占用相对较大。若采用高精度的存储方式，对硬件设备的要求会显著提高。在训练过程中，高精度的存储方式会在反向传播时对训练结果产生影响。然而，当模型训练完成后，实际调用模型时无需进行反向传播。因此，为了降低对硬件资源的需求，将模型转换为 FP32 甚至更低精度的 FP16 是一个合理的选择，这样可以显著减少所需的运行内存，此时对硬件要求更低，同时，调整数据结构还可以增加自动校准，从而减少性能损失。

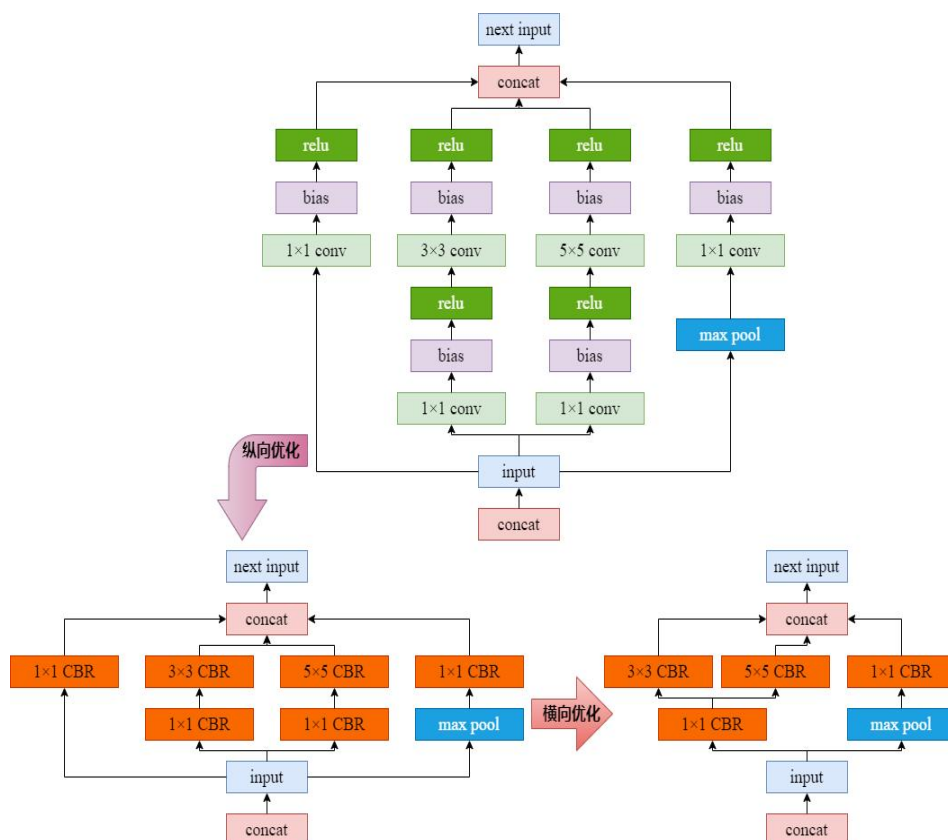


图 5-4 Inception 模块合并优化

5.5.2 模型的边缘部署

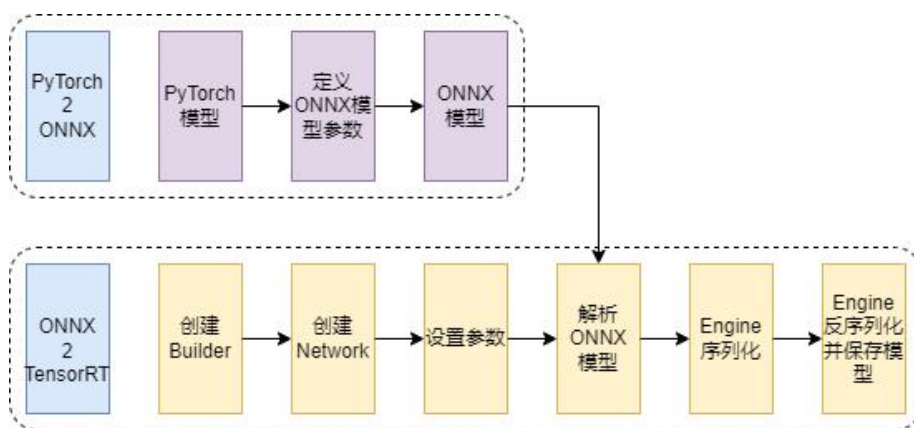


图 5-5 模型转换流程图

在进行 TensorRT 转换之前，首要步骤是将模型转换为 ONNX 格式。ONNX 模型作为一种中间表示形式，能够在不同的深度学习框架之间实现数据的无缝交互。随后，我们采用了 TensorRT 内置的工具 `trtexec` 来进行 ONNX-TensorRT 模型的转换工作。此工具通过集成的 Builder 模块，根据网络定义构建了一个高效的推理引擎（Engine）文件。这一过程正是前述所提的优化技术方法的实现。

优化后的引擎创建了执行上下文，将文件反序列化并本地保存，通过执行此引擎文件，就可以实现推理过程的加速。图 5-5 展示了模型转换的完整流程图。

5.6 边缘设备识别实验

一般在深度学习中，卷积神经网络的训练都是在普通电脑上进行的，边缘设备由于硬件条件受限，直接在上面进行训练并不适宜。为确保模型的训练效果与效率，我们先在普通电脑上完成深度学习目标检测模型的训练工作，随后再将其部署至边缘设备中。其中，Jetson nano 作为一款由英伟达公司精心研发的 AI 边缘计算设备，凭借其卓越的性能和紧凑的体积，在业界备受瞩目，能支持多个神经网络并行运行，实现图像分类、人脸识别、语音处理、目标检测及物体识别追踪等功能，还可以用于创建各种各样的项目，从小型智能 IoT（物联网）应用程序到更复杂的机器人，人工视觉系统和对象识别，以及通过评估一系列传感器参数做出智能反应的设备，小型自动驾驶汽车等，设备如图 5-6 所示。表 5-1 为训练草莓检测模型的计算机与 Jetson nano 参数对比表。对上文提出的改进模型，使用 TensorRT 加速技术，部署到 Jetson nano 开发板上。在本次实验中，随机从制作的草莓数据集中的测试集挑选出 10 张图片进行测试，表 5-2 为模型平均对每张图片的推理速度对比。



图 5-6 Jetson Nano 设备图

表 5-1 硬件设备参数对比

产品参数	训练计算机	Jetson Nano BO1
中央处理器	Intel(R) Xeon(R) W-2255 CPU @ 3.70GHz	ARM A57 @ 1.43GHz
图形处理器核心	6144	128
专用 GPU 内存	16GB	4GB
系统	Windows	Linux for Ubuntu

表 5-2 推理速度对比

设备	模型	推理速度/s
Jetson Nano	改进 YOLOv5s	0.45
	改进 YOLOv5s+TensorRT 加速	0.22

由表 5-1 和 5-2 中可以看出，由于硬件的性能差距较大，改进 YOLOv5 模型不使用 TensorRT 技术加速，在 Jetson Nano 中进行推理，平均对每张图片的推理速度为 0.45s。使用 TensorRT 技术进行加速之后，可以看到改进的 YOLOv5s 模型平均对每张图片的推理速度为 0.22s，提升了大约 51.1%，大幅度提升了对每张图片的检测速度。通过本次实验可以证明 TensorRT 加速技术可以帮助图像处理器加速推理时间，对于应用到边缘设备的识别模型有很好的优化效果。图 5-7 为使用 TensorRT 加速技术后模型在 Jetson nano 中对草莓果实识别的部分结果图，其中，0 代表草莓花；1 代表未成熟草莓；2 代表已成熟草莓。由图 5-7 可以看到，模型边缘部署之后，对于遮挡目标具有较好的检测效果，进行了正确的分类识别。

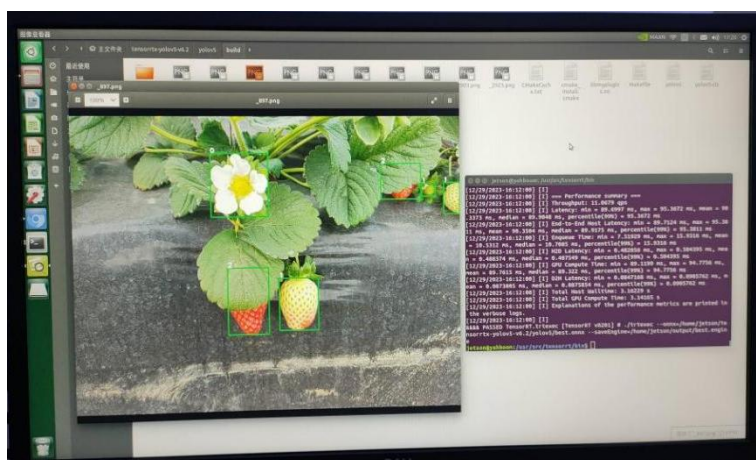


图 5-7 识别结果

5.7 本章小结

本章首先进行了草莓检测系统的需求分析，使用 PyQt5 图形用户界面库，实现了对草莓检测系统的搭建和界面的设计工作，并对系统搭建的环境、登录界面、主界面和里面的功能模块进行了详细的阐述。经过 TensorRT 优化后，将草莓检测模型成功移植至边缘设备 Jetson nano 中。在此平台上，我们对本文的草莓检测模型，以及经过 TensorRT 加速的草莓检测模型进行了识别时间的对比分析。实验结果显示，经过 TensorRT 加速的模型，其平均识别速度相较于未使用的提升了约 51%，这一显著的性能提升验证了 TensorRT 加速在提升边缘设备实时识别能力方面的有效性，为深度学习目标检测模型在草莓采摘机器人上的实际应用提供了一定的参考。

第 6 章 草莓检测与定位实验

为了验证本文改进草莓检测算法和定位方法的有效性，分别在真实场景中进行草莓检测与定位实验，其中检测实验主要为了测试本文改进模型对实际种植园草莓的检测效果，而定位实验主要是为了检验双目相机与 RGB-D 相机对草莓三维定位的精确性，对两者得出的实验数据对比并进行分析。

6.1 草莓检测实验

6.1.1 草莓检测实验内容

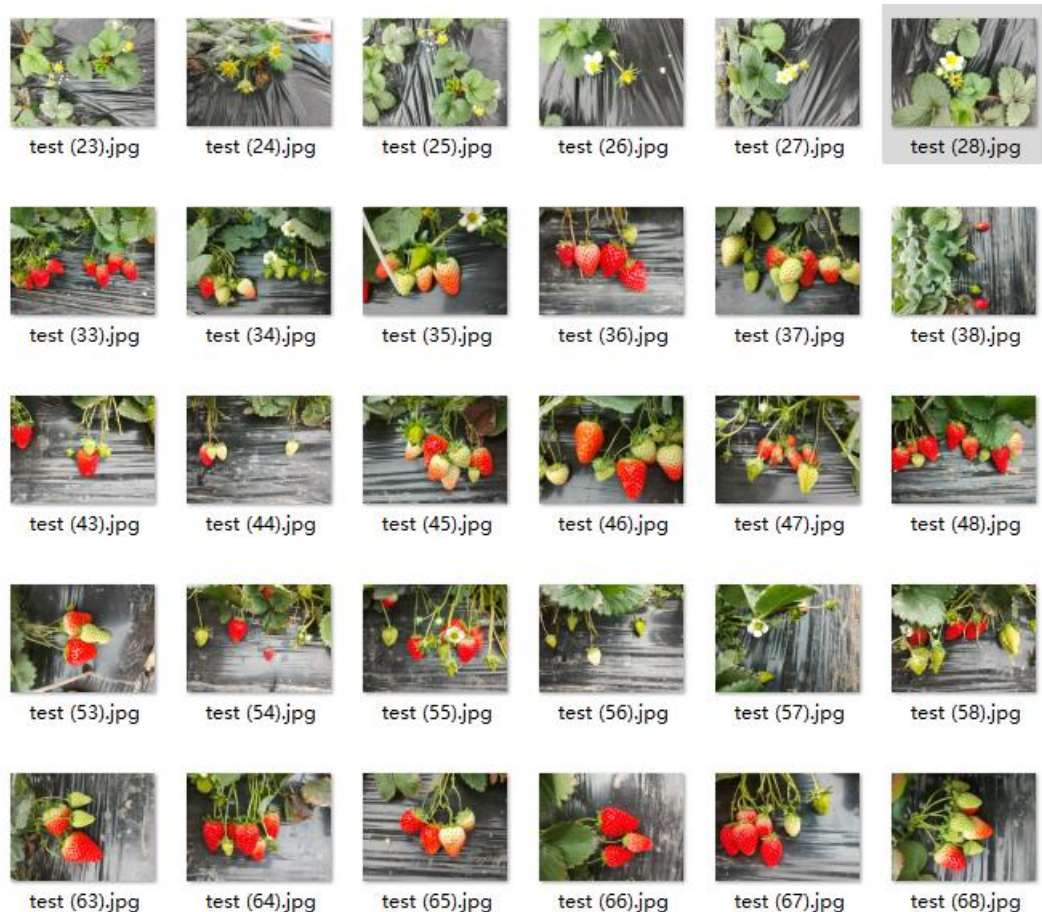


图 6-1 拍摄图片

为了验证本文改进 YOLOv5 算法对实际种植草莓的检测效果，在凤阳县梅山草莓种植基地，使用 oppo ace2 手机随机拍摄草莓图片 100 张，包含了本文标注的三个类别状态，分别是 flowering（草莓花）、fruit（未成熟草莓）和 ripe（成熟草莓），拍摄图片部分示例如图 6-1 所示，统计出拍摄的图片中。草莓

花目标数量为 94，未成熟草莓数量为 461，已成熟草莓数量为 195，然后将拍摄好的草莓实际种植图片输入到已经训练好的草莓检测模型中进行检测，配置环境与第三章相同，通过实验，得到实验数据。

6.1.2 草莓检测结果分析

调用前面使用自制草莓目标检测数据集训练完成的改进模型的权重参数文件，将采集的某草莓种植园的图片输入到网络模型中进行推理检测，通过实验得到改进模型对图片检测后出现的漏检和错检的数据统计，实验结果如表 6-1 所示，检测部分效果如图 6-2 所示。

表 6-1 实验结果统计

类别	真实类别 数量/个	正确检测 数量/个	漏检数 量/个	漏检率/%	错检数 量/个	错检率 /%	识别准确 率/%
flowering	94	78	16	17.02	0	0	82.98
fruit	461	291	154	33.41	16	3.47	63.12
ripe	195	182	13	6.67	0	0	93.33

由表 6-1 实验数据统计可以看到，在 94 个真实草莓花目标中，改进模型正确检测数量为 78 个，错检率为 0，识别准确率为 82.98%，对草莓花的漏检数量为 16 个，漏检率约为 17.02%，通过观察改进模型检测后的图片，发现当两朵草莓花距离很近，相互接触时，模型有时会将其检测为一个目标，从而对另一个草莓花目标出现一定程度的漏检现象；在 461 个真实未成熟草莓目标中，改进模型正确检测数量为 291 个，错检个数为 16，错检率为 3.47%，对未成熟草莓的漏检数量为 154，漏检率为 33.41%，识别准确率较低，为 63.12%。通过观察分析改进模型检测后的图片，主要可能的原因为：1) 较多未成熟草莓在果实表面还没有任何着色，处于绿熟期阶段，与草莓的叶片和草莓植株的茎秆颜色非常接近，遮挡重叠数量较多，模型会错误地将其识别为背景，故而导致漏检现象较多。2) 根据国家标准^[75]，表面着色度达到 80%以上时才被认为是可采收的成熟草莓，处于转色期的未成熟草莓表面着色度还没有达到要求，但是有时模型会错误地检测为成熟草莓，从而出现了一定程度的错检现象；在 195 个真实成熟草莓目标中，改进模型正确检测数量为 182 个，错检率为 0，拥有较高的识别准确率，达到了 93.33%，有 13 个成熟草莓漏检，漏检率约为 6.67%，

通过观察改进模型检测后的图片，可能的原因主要有两个：1) 当成熟草莓数量比较密集时，重叠区域过多时，改进模型会出现一定程度的漏检现象，将多目标错误地识别为单目标。2) 当成熟草莓体积被草莓叶片等遮挡过多时，也会出现漏检现象。

由本次实验可以看到，虽然改进模型依然存在着一定的错检漏检现象，但是对于成熟草莓目标，改进模型具有较高的识别准确率，达到了 93.33%，漏检率只有 6.67%，对于应用到草莓智能采摘机器人中具有较大的应用前景。



图 6-2 检测效果图

6.2 草莓定位实验

6.2.1 草莓定位实验内容

为了验证本文提出的草莓三维定位方法的有效性，本文以仿真的成熟草莓果实模型为研究对象，分别使用被动双目相机 HBV-1780-2 S2.0 和 RGB-D 相机 Intel RealSense D435i 进行定位实验，图 6-3 为验证实验的场景图，被动双目相

机的三维定位操作流程如图 6-4（a）所示，RGB-D 相机的三维定位操作流程如图 6-4（b）所示。



图 6-3 实验场景图

对于已经标定好的被动双目相机，首先通过左右相机获取彩色 RGB 图像，然后将 RGB 图像输入到本文的草莓检测模型中，模型会自动识别出图像中草莓花、未成熟草莓和成熟草莓区域并用矩形框进行标识，计算出中心点的二维像素坐标；同时，对左右采集的彩色图像进行畸变矫正和立体匹配，得到左右相机视差；接下来通过分析深度信息图像中对应像素坐标的数据，获取草莓的深度距离；最后根据相机成像原理和双目视觉定位原理，计算出草莓中心点的三维空间位置。

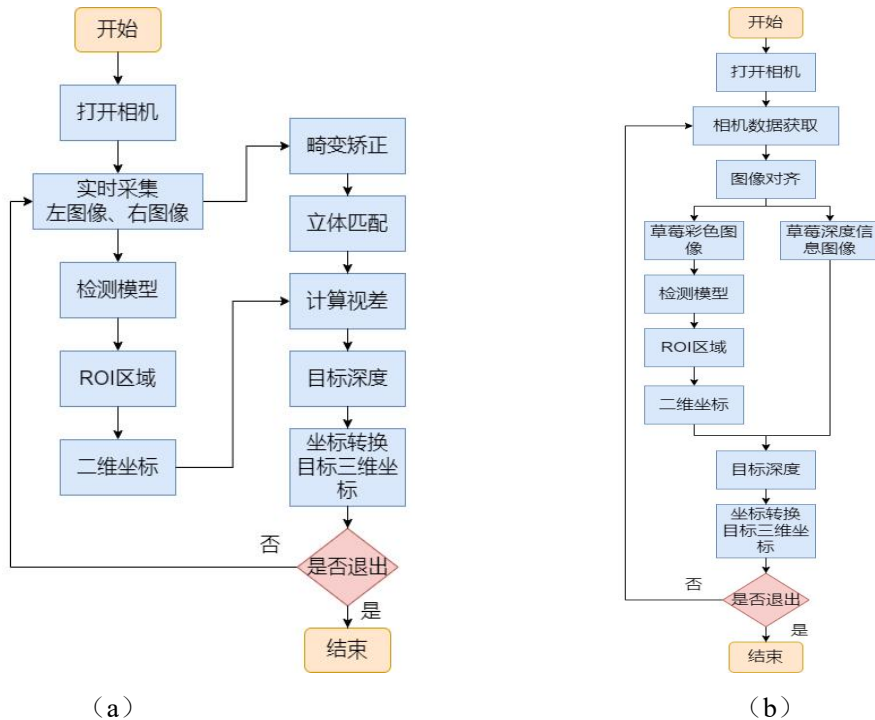


图 6-4 相机三维定位流程

对于已经标定好的 RGB-D 相机，首先通过 RGB 相机获取草莓彩色 RGB 图片，将彩色图片输入到本文已经训练好的草莓检测模型中，模型会自动识别出图像中草莓花、未成熟草莓和成熟草莓区域并用矩形框进行标识，计算出中心点的二维像素坐标；同时，通过左红外相机和右红外相机获取深度图像；然后，将深度图像与彩色 RGB 图像进行对齐，得到对齐后的草莓深度信息图像；接下来通过分析深度信息图像中对应像素坐标的数据，获取草莓在相机坐标系下的深度距离；最后，根据相机成像原理和双目视觉定位原理，计算出草莓中心点的三维空间位置。

6.2.2 草莓定位结果分析

图 6-5 为 Intel Realsense D435i 深度相机获取的 RGB 彩色图像，从图 6-5 中可以看出，成功检测了 4 个成熟草莓目标，把这些目标的像素坐标代入到对应的深度信息中可得到目标的深度数据，再根据双目视觉定位原理计算出草莓目标三维坐标。表 6-2 和表 6-3 分别列出了 HBV-1780-2 S2.0 相机和 Intel Realsense D435i 深度相机获取的深度数据和三维坐标。

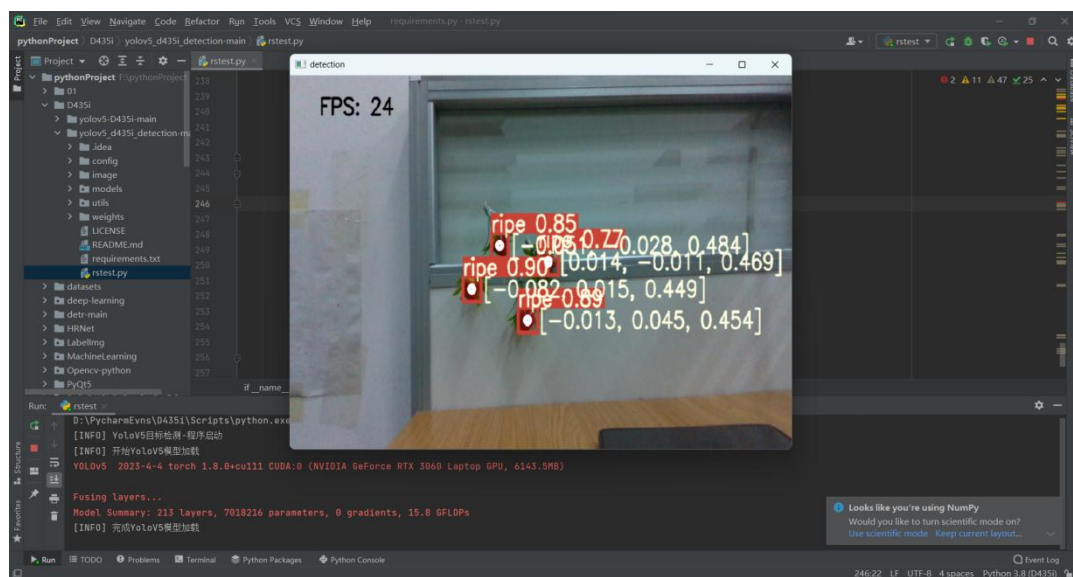


图 6-5 Intel Realsense D435i 深度相机三维定位测试图

实验结果表明：确定草莓中心点的二维像素坐标和深度数据后，被动双目相机能够通过双目视觉定位原理计算出草莓中心点基于世界坐标系的三维空间坐标，RGB-D 也能够通过深度相机计算得出。忽略人工测量的误差，通过将人工测量的深度作为实际深度，与相机测量的深度值进行对比，可以得到使用被动双目相机 HBV-1780-2 S2.0 计算出的深度与目标果实中心点的实际深度的平

均误差为 $\pm 18\text{mm}$ ，使用 RGB-D 相机 Intel Realsense D435i 计算出的深度与目标果实中心点的实际深度的平均误差仅为 $\pm 3.75\text{mm}$ 。通过对比可得，使用 RGB-D 相机，比使用被动双目相机误差减少了 14.25mm ，更加适合草莓采摘机器人定位的精度要求。

表 6-2 HBV-1780-2 S2.0 相机定位数据

目标序号	实际深度值/mm	三维坐标/mm	深度误差/m
1	0.480	[-0.43, -0.036, 0.497]	$\Delta 0.017$
2	0.474	[0.029, -0.025, 0.452]	$\Delta 0.022$
3	0.453	[-0.101, 0.026, 0.434]	$\Delta 0.019$
4	0.456	[-0.019, 0.052, 0.441]	$\Delta 0.015$

表 6-3 Intel Realsense D435i 深度相机定位数据

目标序号	实际深度/m	三维坐标/m	深度误差/m
1	0.480	[-0.051, -0.028, 0.484]	$\Delta 0.004$
2	0.474	[0.014, -0.011, 0.469]	$\Delta 0.005$
3	0.453	[-0.082, 0.015, 0.449]	$\Delta 0.004$
4	0.456	[-0.013, 0.045, 0.454]	$\Delta 0.002$

6.3 本章小结

本章主要进行了草莓目标检测和三维定位实验来验证本文方法的有效性，目标检测实验表明：本文改进的草莓检测算法对成熟草莓拥有较好的检测效果，通过对收集的某草莓种植基地的图片进行检测实验，改进草莓检测模型对成熟草莓的识别准确率达到了 93.33% ，漏检率只有 6.67% ，可以有效地对实际种植的草莓进行检测。三维定位实验表明：使用 RGB-D 相机进行草莓三维定位的精度优于被动双目相机，相较于被动双目相机，测量时的平均深度误差减少了 14.25mm ，仅为 $\pm 3.75\text{mm}$ ，更加适合草莓采摘机器人定位的精度要求。

第 7 章 总结与展望

本文“基于深度学习的草莓检测与定位方法研究”通过综合果实检测与定位研究领域的进展并发现其不足点，通过分析国内外的果实检测与定位方式，基于深度学习目标检测算法和双目视觉分别研究了草莓的检测和定位方法。首先，通过改进深度学习目标检测模型，从而提升了模型对草莓的检测精度，在复杂环境下对草莓的检测效果，减少了模型对草莓的漏检和错检现象；同时，对模型进行了轻量化设计，降低了模型复杂度；其次，通过图像处理方法，实现了对草莓的二维定位；再次，通过使用被动双目相机和 RGB-D 相机，实现了对草莓的三维定位；同时对草莓检测系统进行了设计工作；然后运用 TensorRT 加速技术，对改进的草莓检测模型进行了边缘部署工作；最后，本文结合研究的草莓检测与定位方法，进行了实际环境下的检测与定位实验，证明了研究内容的有效性。

7.1 论文工作总结

(1) 本论文综述了近年来果实检测与定位的国内外研究，详细阐述了深度学习中的目标检测模型及其结构特性和应用，强调了其在采摘领域的重要性。此外，分析了双目视觉在果实采摘定位中的应用，特别指出了在草莓采摘中的良好前景。

(2) 基于 StrawDI 收集的 3100 张图片，按照农业标准分类为草莓花、未熟草莓和成熟草莓，按照 8:1:1 比例，将数据集随机分为训练集、验证集和测试集，并按 Pascal VOC2007 格式标注，构建了草莓目标检测数据集。

(3) 针对检测精度、模型复杂度及复杂环境下性能不足的问题，采用 Ghost 结构轻量化模型，引入 SIoU 损失函数提升复杂环境下的检测效果，添加 ECA 注意力机制和 SimSPPF 特征融合模块提高检测精度。测试结果显示，改进模型 P 值 94.5%，R 值 90.5%，mAP 值 94.9%，模型参数量减少约 50%，规模减小约 43.4%。与原始 YOLOv5s 相比，性能显著提升，并在复杂环境下表现出良好的检测效果。

(4) 研究了基于图像处理的成熟草莓二维定位和双目视觉的三维定位方法。在 Halcon 软件中, 通过预处理、颜色空间转换、图像分割等步骤实现了二维定位。推导了草莓像素坐标与世界坐标的转换关系, 并使用张正友标定法标定 RGB-D 和被动双目相机, 结合改进的 YOLOv5s 检测方法和双目相机定位, 实现了草莓的三维定位。实验显示, RGB-D 相机定位性能更佳, 深度平均误差为 $\pm 3.75\text{mm}$ 。

(5) 设计了基于改进模型的草莓检测系统及其图形用户界面。利用 Pycharm 和 PyQt5 中的 Qt Design 工具完成了系统搭建, 包括登录、主界面和草莓检测界面。通过 Pytorch、OpenCV 和 Numpy 等库实现系统功能, 该系统便利用户, 具有实际应用价值。

(6) 改进草莓检测模型通过 TensorRT 加速技术成功部署于边缘设备, 实现了加速推理。在 Jetson nano 平台上测试, 使用 TensorRT 后推理速度提升至每张图片 0.22s, 速度提升约 51.1%。实验证明了 TensorRT 技术能显著提高边缘设备上的检测速度, 实现了深度学习模型在低算力边缘设备的应用。

7.2 未来展望

本文实现了基于深度学习的草莓检测与定位方法研究, 为自动化草莓采摘机器的视觉检测和定位方面的研究提供了一些方法, 但限于时间等因素仍然存在着以下不足之处需要进行改进及后续进一步研究:

(1) 本文使用 Ghost 模块对 YOLOv5s 进行了模型轻量化工作达到了一定的效果, 但由于 Ghost 卷积内部使用的是 Depthwise Conv (深度可分离卷积), CUDNN (是基于 CUDA 的深度学习 GPU 加速库) 对 Depthwise Conv 并没有针对性的优化, 使得整个操作碎片化, 另外, 因为 3×3 Conv 用得最多, 而 CUDNN 专门优化过, 故拉大了对比, 导致改进的 YOLOv5 模型检测速度有待进一步提升, 后续可以考虑优化 CUDNN 底层代码、使用 MobileNet 作为主干网络、模型剪枝等其它方法, 对模型进一步压缩, 减少程序的检测时间。

(2) 深度学习目标检测算法目前正在不断地更新迭代。由于研究的局限性, 今后将在其它更为先进的深度学习目标检测网络模型的基础上进行改进, 以期

获得对草莓目标更好的检测性能。

(3) 本文研究的草莓检测数据集数量仅采用 StrawDI 团队收集的不同生长周期的草莓图片，对草莓检测定位系统的未来推广到更多类型或环境使用有一定的限制，今后将考虑对自制的草莓检测数据集进行扩充，建立更大的草莓目标检测数据集来完善模型。

(4) 本文只对草莓果实进行了检测与定位研究，设计了草莓检测系统，并将草莓检测模型部署到嵌入式设备中。由于时间与设备的问题，如何搭建草莓采摘的定位系统，设计嵌入式设备与草莓采摘机械臂之间的控制系统，实现草莓采摘点的位置转换，后面有待进一步研究。

参考文献

- [1] 张智,李曼宁,杨志等.基于多指标协同的草莓水肥耦合综合调控[J].农业机械学报,2020,51(02):267-276.
- [2] 张更,颜志明,王全智,等.我国设施草莓无土栽培技术的研究进展与发展建议[J].江苏农业科学,2019,47(18):58-61.
- [3] 韩振兴,刘东辉,常向阳.乡村振兴背景下采摘型休闲农业消费者选择偏好研究——基于选择实验法的实证分析[J].农林经济管理学报,2023,22(03):322-329.DOI:10.16195/j.cnki.cn36-1328/f.2023.03.34
- [4] 刘继展,吴硕.草莓全程生产机械化技术与装备研究进展[J].农业机械学报,2021,52(05):1-16.
- [5] 刘成良,林洪振,李彦明等.农业装备智能控制技术研究现状与发展趋势分析[J].农业机械学报,2020,51(01):1-18.
- [6] 赵博,张巍朋,苑严伟等.农业装备运维与作业服务管理信息化技术研究进展[J].农业机械学报,2023,54(12):1-26.
- [7] 孙钰,韩京治,陈志泊等.基于深度学习的大棚及地膜农田无人机航拍监测方法[J].农业机械学报,2018,49(02):133-140.
- [8] 贾少鹏,高红菊,杭潇.基于深度学习的农作物病虫害图像识别技术研究进展[J].农业机械学报,2019,50(S1):313-317.
- [9] 孙红,李松,李民赞等.农业信息成像感知与深度学习应用研究进展[J].农业机械学报,2020,51(05):1-17.
- [10] 翟肇裕,曹益飞,徐焕良等.农作物病虫害识别关键技术研究综述[J].农业机械学报,2021,52(07):1-18.
- [11] Le Louëdec J, Cielniak G. 3D shape sensing and deep learning-based segmentation of strawberries[J]. Computers and Electronics in Agriculture, 2021, 190: 106374.
- [12] He Z, Karkee M, Zhang Q. Detecting and localizing strawberry centers for robotic harvesting in field environment[J]. IFAC-PapersOnLine, 2022, 55(32): 30-35.
- [13] Ge Y, Xiong Y, Tenorio G L, et al. Fruit localization and environment perception for strawberry harvesting robots[J]. IEEE Access, 2019, 7: 147642-147652.
- [14] 齐飞,李恺,李邵等.世界设施园艺智能化装备发展对中国的启示研究[J].农业工程学报,2019,35(02):183-195.
- [15] 刘宏新,王登宇,郭丽峰等.先进设计技术在农业装备研究中的应用分析[J].农业机械学报,2019,50(07):1-18.
- [16] 王晓兵.推动农业机械化智能化,强化粮食安全装备支撑[J].农业经济与管

- 理,2023,(01):21-23.
- [17] 王鹏新,田惠仁,张悦等.基于深度学习的作物长势监测和产量估测研究进展[J].农业机械学报,2022,53(02):1-14.
- [18] 张晓东,张朝昆,赵继军.边缘智能研究进展[J].计算机研究与发展,1-22[2023-12-13]<http://kns.cnki.net/kcms/detail/11.1777.tp.20230310.1731.006.html>.
- [19] Yu Y, Zhang K, Yang L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. Computers and Electronics in Agriculture, 2019, 163: 104846.
- [20] 刘小刚,范诚,李加念等.基于卷积神经网络的草莓识别方法[J].农业机械学报,2020,51(02):237-244.
- [21] 杨珍,龚惟新,李凯等.高架草莓的果实识别与果梗分割[J].农业工程学报,2023,39(17):172-181.
- [22] Xiong Y, Ge Y, Grimstad L, et al. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation[J]. Journal of Field Robotics, 2020, 37(2): 202-224.
- [23] 毕松,隗朋峻,刘仁学.温室高架栽培草莓空间姿态识别与采摘点定位方法[J].农业机械学报,2023,54(09):53-64+84.
- [24] Xiong Y, Ge Y, Grimstad L, et al. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation[J]. Journal of Field Robotics, 2020, 37(2): 202-224.
- [25] 朱新华,李洪春,李旭东等.果园有机肥机械化环沟施肥方法及其实现[J].农业工程学报,2023,39(17):60-70.
- [26] Escalante H J, Rodríguez-Sánchez S, Jiménez-Lizárraga M, et al. Barley yield and fertilization analysis from UAV imagery: a deep learning approach[J]. International journal of remote sensing, 2019, 40(7): 2493-2516.
- [27] Agrahari R K, Kobayashi Y, Tanaka T S T, et al. Smart fertilizer management: the progress of imaging technologies and possible implementation of plant biomarkers in agriculture[J]. Soil Science and Plant Nutrition, 2021, 67(3): 248-258.
- [28] 陈青,殷程凯,郭自良等.苹果采摘机器人关键技术研究现状与发展趋势[J].农业工程学报,2023,39(04):1-15.
- [29] 张晴晖,孔德肖,李俊萩等.基于逆运动学降维求解与 YOLO v4 的果实采摘系统研究[J].农业机械学报,2021,52(09):15-23.
- [30] Zhong Z, Xiong J, Zheng Z, et al. A method for litchi picking points calculation in natural environment based on main fruit bearing branch detection[J]. Computers and Electronics in Agriculture, 2021, 189: 106398.

-
- [31] 乔艳军.基于深度学习的采摘机器人水果识别定位与采后自动分级技术研究[D].天津理工大学,2022.DOI:10.27360/d.cnki.gtlgy.2022.000219
 - [32] Chithra P L, Henila M. Apple fruit sorting using novel thresholding and area calculation algorithms[J]. *Soft Computing*, 2021, 25(1): 431-445.
 - [33] Bellon V, Rabatel G, Guizard C. Automatic sorting of fruit: sensors for the future[J]. *Food Control*, 1992, 3(1): 49-54.
 - [34] Li Z, Miao F, Yang Z, et al. Factors affecting human hand grasp type in tomato fruit-picking: A statistical investigation for ergonomic development of harvesting robot[J]. *Computers and electronics in agriculture*, 2019, 157: 90-97.
 - [35] Hua X, Li H, Zeng J, et al. A Review of Target Recognition Technology for Fruit Picking Robots: From Digital Image Processing to Deep Learning[J]. *Applied Sciences*, 2023, 13(7): 4160.
 - [36] Tao Y, Zhou J. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking[J]. *Computers and electronics in agriculture*, 2017, 142: 388-396.
 - [37] Hassanein, M.; Lari, Z.; El-Sheimy, N. A new vegetation segmentation approach for cropped fields based on threshold detection from hue histograms[J]. *Sensors*. 2018, 18(4): 1253.
 - [38] Nalepa, J. Recent advances in multi-and hyperspectral image analysis[J]. *Sensors*. 2021, 21(18): 6002.
 - [39] 朱坚民, 雷静桃, 翟东婷, 等. 基于灰关联分析和模糊隶属度匹配的球形水果自动识别方法[J]. *仪器仪表学报*, 2012, 33 (08): 1826-1836.
 - [40] Liming, X.; Yanchao, Z. Automated strawberry grading system based on image processing[J]. *Comput. Electron. Agric.* 2020, 71 (Suppl. 1), S32–S39.
 - [41] Kurtulmus F, Lee W S, Vardar A. Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network[J]. *Precision agriculture*, 2014, 15: 57-79.
 - [42] 卢军,胡秀文.弱光复杂背景下基于 MSER 和 HCA 的树上绿色柑橘检测[J].*农业工程学报*,2017,33(19):196-201.
 - [43] 郭峰, 曹其新, 崔永杰, 等. 用于草莓收获机器人的果实定位和果柄检测方法 [J]. *农业工程学报*,2008(10):89-94.
 - [44] 刘小刚, 范诚, 李加念, 等. 基于卷积神经网络的草莓识别方法 [J]. *农业机械学报*,2020,51(02):237-244.
 - [45] Chu P, Li Z, Lammers K, et al. Deep learning-based apple detection using a suppression mask R-CNN[J]. *Pattern Recognition Letters*, 2021, 147: 206-211.
 - [46] Yu Y, Zhang K, Yang L, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. *Computers and Electronics in Agriculture*, 2019, 163: 104846.

-
- [47] 赵德安,吴任迪,刘晓洋,等.基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J].农业工程学报,2019,35(03):164-173.
 - [48] Zhou, B.; Lin, X.; Zhou, J.; et al. An improved Lightweight Parameters Network for Strawberry Flowers Detection[J]. IEEE Access. 2023: Volume 11, Pages 63761-63772.
 - [49] Hu, C.; Liu, X.; Pan, Z.; et al. Automatic Detection of Single Ripe Tomato on Plant Combining Faster R-CNN and Intuitionistic Fuzzy Set[J]. IEEE Access. 2019,7, 154683–154696.
 - [50] 孙俊,陈义德,周鑫,等.快速精准识别棚内草莓的改进 YOLOv4-Tiny 模型[J].农业工程学报,2022,38(18):195-203.
 - [51] Kang, H.W.; Zhou, H.Y.; Wang, X.; et al. Real-Time Fruit Recognition and Grasping Estimation for Robotic Apple Harvesting[J]. Sensors. 2020, 20 (19), 5670.
 - [52] Ni, J.; Gao, J.; Li, J.; et al. E-AlexNet: quality evaluation of strawberry based on machine learning. Journal of Food Measurement and Characterization[J]. 2021, 15(5): 4530-4541.
 - [53] Hu, H.; Kaizu, Y.; Zhang, H.; et al. Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN. International Journal of Agricultural and Biological Engineering[J]. 2022, 15(6): 175-179.
 - [54] Wang J, Li Z, Gao G, et al. BerryNet-Lite: A Lightweight Convolutional Neural Network for Strawberry Disease Identification[J]. Agriculture, 2024, 14(5): 665.
 - [55] Zhou, B.; Lin, X.; Zhou J.; et al. An improved Lightweight Parameters Network for Strawberry Flowers Detection[J]. IEEE Access. 2023,63761-63772.
 - [56] Wang, G.; Zheng, H.; Li, X. ResNeXt-SVM: a novel strawberry appearance quality identification method based on ResNeXt network and support vector machine. Journal of Food Measurement and Characterization[J]. 2023: 1-12.
 - [57] 毕松,隗朋峻,刘仁学.温室高架栽培草莓空间姿态识别与采摘点定位方法[J].农业机械学报,2023,54(09):53-64+84.
 - [58] 张凯良,杨丽,张铁中.草莓采摘位置机器视觉与激光辅助定位方法[J].农业机械学报,2010,41(04):151-156.
 - [59] 王粮局,张立博,段运红等.基于视觉伺服的草莓采摘机器人果实定位方法[J].农业工程学报,2015,31(22):25-31.
 - [60] 陈燕,王佳盛,曾泽钦等.大视场下荔枝采摘机器人的视觉预定位方法[J].农业工程学报,2019,35(23):48-54.
 - [61] 张勤,陈建敏,李彬等.基于 RGB-D 信息融合和目标检测的番茄串采摘点识别定位方法[J].农业工程学报,2021,37(18):143-152.
 - [62] Lin G, Tang Y, Zou X, et al. In-field citrus detection and localisation based on RGB-D image analysis[J]. Biosystems Engineering, 2019, 186: 34-44.
 - [63] Benavides M, Cantón-Garbín M, Sánchez-Molina J A, et al. Automatic tomato and peduncle

- location system based on computer vision for use in robotized harvesting[J]. *Applied Sciences*, 2020, 10(17): 5887.
- [64] Yoshida, T.; Fukao, T.; Hasegawa, T. A Tomato Recognition Method for Harvesting with Robots using PointClouds[J]. In *Proceedings of the 2019 IEEE/SICE International Symposium on System Integration (SII)*, Paris, France, 14–16 January 2019; pp. 456–461.
- [65] Ge Y, Xiong Y, From P J. Three-dimensional location methods for the vision system of strawberry-harvesting robots: development and comparison[J]. *Precision Agriculture*, 2023, 24(2): 764-782.
- [66] 李国厚,李泽旭,金松林,等.混合卷积神经网络用于高光谱小麦品种鉴别[J].*光谱学与光谱分析*,2024,44(03):807-813.
- [67] Li Z, Liu F, Yang W, et al. A survey of convolutional neural networks: analysis, applications, and prospects[J]. *IEEE transactions on neural networks and learning systems*, 2021, 33(12): 6999-7019.
- [68] Kamilaris A, Prenafeta-Boldú F X. A review of the use of convolutional neural networks in agriculture[J]. *The Journal of Agricultural Science*, 2018, 156(3): 312-322.
- [69] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014: 580-587.
- [70] Girshick R. Fast r-cnn[C]//*Proceedings of the IEEE international conference on computer vision*. 2015: 1440-1448.
- [71] 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.
- [72] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 779-788.
- [73] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//*European conference on computer vision*. Springer, Cham, 2016: 21-37.
- [74] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv:1409.1556*, 2014.
- [75] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the IEEE international conference on computer vision*. 2017: 2980-2988.
- [76] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770-778.
- [77] Perez-Borrero I, Marin-Santos D, Gegundez-Arias M E, et al. A fast and accurate deep learning method for strawberry instance segmentation[J]. *Computers and Electronics in Agriculture*, 2020, 178: 105736.
- [78] 中华人民共和国农业农村部. 草莓采收与贮运技术规范: NY/T 2787-2015[S]. 2015-05-21.
- [79] Everingham M , Gool L V , Williams C K I ,et al.The Pascal Visual Object Classes (VOC)

-
- Challenge[J].International Journal of Computer Vision, 2010, 88(2):303-338.DOI:10.1007/s11263-009-0275-4.
- [80] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
- [81] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11534-11542.
- [82] Han K, Wang Y, Tian Q, et al. Ghostnet: More features from cheap operations[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 1580-1589.
- [83] Gevorgyan Z. SIOU Loss: More Powerful Learning for Bounding Box Regression[J]. arXiv preprint arXiv:2205.12740, 2022.
- [84] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.
- [85] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [86] Wang G, Chen Y, An P, et al. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.
- [87] 姜 锐 . 柑橘采摘机器人目标识别、定位与采摘研究 [D]. 江苏大学,2022.DOI:10.27170/d.cnki.gjsuu.2022.001423.

攻读硕士学位期间发表的学术论文及专利

彭勇,乔印虎,姚杰.基于改进的 ResNet34 网络模型的苹果叶病害识别[J].安徽工程大学学报,2022,37(06):51-57.

彭勇, 乔印虎, 姚杰, 等。改进 YOLOv5 模型的草莓检测方法研究[J].中国农机化学报,2023.10（已录用）

Chunyan Zhang, **Yong Peng** , Yinhu Qiao, et al. Research on strawberry detection method based on improved YOLOv5 model[J]. Engineering Letters（正在复审）

姚杰,张春雨,彭勇,等.基于计算机视觉的小麦收获边缘导航线提取方法[J].中国农机化学报,2024.02（已见刊）

一种联合收获机割台自动对行装置：姚杰，张春雨，周爽，彭勇（实用新型专利,2023.07（已授权）

致谢

时光荏苒，白驹过隙，我的硕士生涯已接近尾声。在这段难忘的时光里，我收获了丰富的知识和珍贵的友情，这一切都离不开众多师长、亲人和朋友的关心与支持。在此，我谨以最诚挚的心，向他们表达最深的感激之情。

首先，我要感谢我的导师乔教授。乔老师严谨的治学态度、深厚的学术造诣和悉心的指导，使我在学术研究的道路上受益匪浅。特别是在论文撰写过程中，乔老师不仅提供了宝贵的建议，更在我遇到困难时给予了耐心的鼓励与支持。在此，我向乔老师表示最崇高的敬意和衷心的感谢。

同时，我要感谢学院的张老师，她的专业知识和丰富的教学经验，使我收获颇丰，在我书写论文过程中给予了宝贵的意见和建议，为我的论文顺利完成提供了有力保障。

其次，我要感谢我的同窗好友们。在论文写作过程中，我们共同探讨、互相学习、互相鼓励，共同度过了一段难忘的时光。感谢你们在我遇到困难时给予的帮助和支持，你们的陪伴让我的硕士生涯更加充实和美好。

此外，我要感谢参与本次评审的各位专家，你们的参与为本研究提供了重要的见解。你们的专业知识评析使我能够更深入地理解和分析自己的研究内容。向你们说一声辛苦了

最后，我还要感谢我的家人。父母的无私奉献和关爱，使我在求学路上无后顾之忧。他们的理解和支持是我不断前进的动力源泉。在此，我要向他们表达我最真挚的感激之情。

在即将告别硕士生涯之际，我深知这段经历是我人生中宝贵的财富。在未来的日子里，我将铭记师长、亲人和朋友的关爱与支持，继续努力、不断进步。同时，我也将怀着一颗感恩的心，将这份感激化作行动，回馈社会、回馈每一个关心和帮助过我的人。

最后，再次感谢所有关心和帮助过我的师长、亲人和朋友们。愿我们在人生的道路上继续携手前行，共同成长与进步。

尚德敏学 唯实惟新

