

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2220910111



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于双目视觉的果园苹果识别
与定位方法的研究与应用

论 文 作 者 王彦辉

校 内 导 师 汪 军 (教授)

校 外 导 师 袁纪坤 (经理)

专业学位类别 计算机技术

研究方向 (领域) 智能系统及应用

论文提交日期: 2025 年 5 月 9 日

分类号：TP391

密 级：公开

单位代码：10363

学 号：2220910111

题 目 基于双目视觉的果园苹果识别与定位方法的研究与应用

题 目 RESEARCH AND APPLICATION OF ORCHARD APPLE RECOGNITION
AND LOCALISATION METHOD BASED ON BINOCULAR VISION

学 生 姓 名：王彦辉

校 内 教 师：汪 军（教授）

校 外 教 师：袁纪坤（经理）

专业 学位 类别：计算机技术

研究方向（领域）：智能系统及应用

论文答辩日期：2025 年 5 月 10 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：王彦辉

日期：2025年5月9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。

本学位论文属于 保密 ☐，在 _____ 年解密后适用本版权书。
不保密 ☒。

学位论文作者签名：王彦辉

日期：2015 年 5 月 9 日

指导教师签名：王军

日期：2015 年 5 月 9 日

基于双目视觉的果园苹果识别与定位方法的研究与应用

摘 要

中国是全球最大的苹果种植国之一,但传统人工采摘方式存在季节性强、劳动强度大、成本高等问题,难以满足现代果园对高效采摘的需求。图像识别作为智能苹果采摘机器人的核心技术,其精度与速度直接影响采摘效率。本文以卷积神经网络为基础,结合双目立体视觉技术,深入探究苹果目标的识别和定位技术。从分析果园的实际环境出发,提出能够提高苹果识别精度和鲁棒性的方法,从而准确识别到苹果,并获取三维空间坐标,为后续机器人提供更为精确的采摘位置信息,实现高效、精准的苹果采摘。本文的研究结果如下:

(1) 针对果园环境中苹果在遮挡严重、尺度不一等复杂条件下检测精度不足的问题,提出基于 YOLOv5 的检测改进方法。在模型改进方面,YOLOv5 引入了软非极大值抑制替代传统非极大值抑制算法,提高重叠目标的检测精度;采用最优传输分配优化标签分配策略,增强模型在多目标环境下的匹配准确性;同时将主干网络替换为 FasterNet,以提高特征提取效率和速度。

(2) 针对复杂果园场景中苹果检测在细节检测方面仍存在的困难,进一步提出基于 YOLOv7 的检测优化方法,以提升模型对颜色、纹理等细节特征的检测能力。具体改进包括引入最小点距离交并比作为新的回归损失函数,提高边界框回归的准确性;加入高效通道注意力机制,增强模型对颜色、纹理等细节特征的敏感性;并将主干网络中的标准 3×3 卷积替换为深度可分离卷积,在降低计算复杂度的同时保持检测性能。

(3) 针对果园环境中苹果仍存在多尺度分布和空间位置变化下检测精度下降的问题,再进一步提出基于 YOLOv11 的融合增强方法,以提升模型在复杂环境中的检测鲁棒性与精度。首先在颈部网络中引入空间增强注意力模块,提升模型对目标空间分布的关注能力;其次使用全维度动态卷积替代原有卷积结构,使模型能动态适应不同特征尺度,提高复杂场景下的表现力;最后加入 C3k2_ScConv 模块,结

合 C3k 模块与空间和通道重建卷积进行结构重构，进一步增强模型的空间与通道特征融合能力。

(4) 针对苹果目标精准检测与空间定位后的机械化自动采摘的综合需求，提出基于 YOLOv11 与双目视觉融合的检测定位方法。本文结合 Intel RealSense D415 双目相机进行三维距离测量，首先完成相机标定以获取内外参矩阵，并通过立体匹配算法生成视差图转换为深度信息，随后采用排序算法与双边滤波对深度数据进行去噪处理，最后基于三角测量原理计算苹果的空间坐标，实现高精度定位。并在国产高性能计算平台 RK3588 对 YOLOv11 模型进行了转换与优化，以适应边缘设备的算力限制并提升推理效率。实验结果表明，该系统平均深度误差为 0.0148m，检测定位效果较稳定，基本满足果园自动采摘的实际应用需求。

关键词：苹果检测，双目定位，深度学习，目标检测，模型部署

RESEARCH AND APPLICATION OF ORCHARD APPLE RECOGNITION AND LOCALISATION METHOD BASED ON BINOCULAR VISION

ABSTRACT

China is one of the world's largest apple-producing countries, but traditional manual picking methods face issues such as strong seasonality, high labour intensity, and high costs, making it difficult to meet the modern orchard's demand for efficient harvesting. Image recognition, as the core technology of intelligent apple-picking robots, directly influences picking efficiency through its accuracy and speed. This paper is based on convolutional neural networks, combined with binocular stereo vision technology, to explore the recognition and positioning technology of apple targets in depth. Starting from the analysis of the actual environment of the orchard, methods are proposed to enhance the accuracy and robustness of apple recognition, thereby accurately identifying apples and obtaining three-dimensional spatial coordinates, providing more precise picking location information for subsequent robots, achieving efficient and precise apple harvesting. The research results of this paper are as follows:

(1) In response to the issue of insufficient detection accuracy of apples in orchard environments under severe occlusion and varying scales, an improved detection method based on YOLOv5 is proposed. In terms of model improvement, YOLOv5 introduces soft non-maximum suppression to replace the traditional non-maximum suppression algorithm, enhancing the detection accuracy of overlapping targets; it adopts optimal transport allocation to optimise label assignment strategies, improving the model's matching accuracy in multi-target environments; at the same time, the backbone network is replaced with FasterNet to improve feature extraction efficiency and speed.

(2) In response to the difficulties in detail detection for apple detection in complex orchard scenarios, an optimisation method based on YOLOv7 is further proposed to enhance the model's ability to detect details such as colour and texture. The specific improvements include the introduction of the minimum point distance intersection over union as a new regression loss function to improve the accuracy of bounding box regression; the incorporation of an efficient channel attention mechanism to enhance the model's sensitivity to detailed features such as colour and texture; and the replacement of standard 3×3 convolutions in the backbone network with depthwise separable convolutions to reduce computational complexity while maintaining detection performance.

(3) To address the issue of decreased detection accuracy of apples under multi-scale distribution and spatial position variations in orchard environments, a fusion enhancement method based on YOLOv11 is further proposed to improve the model's detection robustness and accuracy in complex environments. First, a spatial enhancement attention module is introduced in the neck network to enhance the model's focus on the spatial distribution of targets; secondly, full-dimensional dynamic convolution replaces the original convolution structure, enabling the model to dynamically adapt to different feature scales and improve performance in complex scenes; finally, the C3k2_ScConv module is added, combining the C3k module with spatial and channel reconstruction convolution for structural reconstruction, further enhancing the model's ability to fuse spatial and channel features.

(4) To meet the comprehensive needs of precise detection and spatial positioning for mechanised automatic apple picking, a detection and positioning method based on YOLOv11 and binocular vision fusion is proposed. This paper combines the Intel RealSense D415 binocular camera for three-dimensional distance measurement, first completing camera calibration to obtain the intrinsic and extrinsic parameter matrices, and generating disparity maps converted into depth information

through a stereo matching algorithm. Subsequently, denoising processing of the depth data is carried out using sorting algorithms and bilateral filtering, and finally, the spatial coordinates of the apples are calculated based on the principle of triangulation to achieve high-precision positioning. The YOLOv11 model was converted and optimised on the domestic high-performance computing platform RK3588 to accommodate the computing power constraints of edge devices and improve inference efficiency. Experimental results show that the system has an average depth error of 0.0148m, the detection and positioning effect is relatively stable, and it basically meets the practical application needs of automatic harvesting in orchards.

KEY WORDS: apple detection, binocular positioning, deep learning, object detection, model deployment

目录

摘 要.....	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 苹果检测研究现状.....	3
1.2.2 苹果定位技术研究现状	4
1.2.3 边缘端研究现状	4
1.3 论文主要研究内容	5
1.4 论文组织框架	6
第 2 章 相关技术理论基础	8
2.1 卷积神经网络的组成.....	8
2.1.1 输入层	9
2.1.2 卷积层	9
2.1.3 池化层	10
2.1.4 激活函数层.....	11
2.1.5 全连接层	12
2.2 双目定位技术	12
2.2.1 定位原理	12
2.2.2 定位步骤	13
2.3 边缘计算	14
2.4 模型转换与模型部署	15
2.5 本章小结	16
第 3 章 基于深度学习的果园苹果检测算法	17
3.1 构建苹果数据集	17
3.1.1 数据增强	18
3.1.2 数据标注	19

3.2 改进 YOLOv5	19
3.2.1 YOLOv5 网络结构.....	19
3.2.2 Soft_NMS	21
3.2.3 最优传输分配	22
3.2.4 FasterNet	24
3.3 改进 YOLOv7	25
3.3.1 YOLOv7 网络结构.....	25
3.3.2 MPDIU	26
3.3.3 ECA	28
3.3.4 DSConv	29
3.4 改进 YOLOv11	30
3.4.1 YOLOv11 网络结构.....	31
3.4.2 SEAM	32
3.4.3 ODConv	34
3.4.4 C3k2_ScConv.....	34
3.5 模型训练与实验	37
3.5.1 实验环境	37
3.5.2 评价指标	37
3.5.3 对比实验结果与分析	38
3.5.4 消融实验结果与分析	39
3.5.5 检测结果分析	41
3.6 本章小结	42
第 4 章 苹果三维定位.....	43
4.1 深度相机	43
4.2 坐标原理	44
4.2.1 仿射变换	44
4.2.2 透视投影	46
4.2.3 刚体变换	46
4.3 相机标定	48
4.4 深度信息获取	50
4.5 苹果的三维坐标	51

4.6 实验结果分析	51
4.7 本章小结	52
第 5 章 果园苹果检测定位系统的实现.....	54
5.1 系统功能	54
5.2 系统设计	55
5.2.1 RK3588.....	55
5.2.2 功能设计	56
5.3 模型转化与部署	59
5.3.1 RKNPU	59
5.3.2 模型转化	59
5.4 系统测试	60
5.5 本章小结	61
第 6 章 总结与展望	63
6.1 研究工作总结	63
6.2 研究成果意义	63
6.3 局限性	64
6.4 展望	64
参考文献	65
攻读学位期间发表的学术论文目录	70
致 谢	71

第 1 章 绪论

1.1 研究背景及意义

随着我国农业现代化进程的不断推进，提升果蔬生产效率与品质已成为推动农业高质量发展的重要方向。苹果作为我国种植最广泛的水果之一，不仅在国内市场占据重要地位，更在全球水果产业中扮演着关键角色。据统计，我国苹果种植面积和产量均占全球总量的 50% 以上^[1]，2021 年全国苹果产量高达 4597.34 万吨，同比增长 4.33%^[2]。如此庞大的产量不仅关系到国民经济发展，也深刻影响着农民收入和地方经济结构。因此，围绕苹果产业链开展相关技术研究具有重要的战略意义。

在整个苹果产业链中，采摘环节是影响苹果产量和品质的关键因素，直接决定了苹果的商品价值和后续供应链的运作效率。当前，我国苹果采摘普遍依赖人工操作，这种方式不仅劳动强度大、效率低，还存在苹果损伤率高、成本上升等问题。特别是在城镇化进程加快、农村人口持续流失与老龄化加剧的背景下，苹果采摘面临劳动力短缺的严峻挑战，制约了果业的可持续发展。

进一步来看，我国苹果主产区多位于高原、丘陵、山地等地形复杂、交通不便的区域，且以家庭分散种植为主，种植规模小、果树密植，使得传统大型采摘机械难以施展。这一现实客观上要求我们研发能够适应复杂场景的灵活、高效的智能采摘装备，以提高采摘效率，降低人工成本，保障苹果品质。

在此背景下，随着人工智能、计算机视觉与智能传感器等技术的发展，自动化采摘机器人逐渐成为农业科技创新的前沿方向^[3]。通过机器替代人工完成高强度、重复性的采摘任务，不仅是解放农村劳动力、应对人口老龄化的有效手段，更是推动农业高质量发展、提高农业生产智能化水平的必然选择^[4]。然而，当前采摘机器人仍面临检测精度不足、环境适应性差、成本高等瓶颈问题，其中最基础也是最关键的技术环节便是苹果的精准检测与定位。

苹果检测与定位技术作为采摘机器人实现高效作业的前提，是实现自动采摘、提升操作效率与稳定性的核心环节^[5]。在复杂自然环境下，苹果可能因遮挡、光照变化、颜色相近等因素而难以准确检测，这对图像处理、目标检测与空间定位算法提出了更高要求^[6]。因此，深入研究苹果检测与定位技术，不仅具有重要的理论研究价值——推

动计算机视觉与农业场景融合发展，还具有重要的应用价值——为智能采摘装备提供技术支撑，加快果园智能化转型，提升果农收益，助力乡村振兴战略落地。

1.2 国内外研究现状

早期的苹果检测方法多为传统图像处理方法，如颜色分割、边缘检测和形态学处理等，用于提取苹果的特征^[7]。尽管这些方法计算速度快，但鲁棒性较差，容易受到各种复杂因素的影响。随着深度学习的图像检测技术快步发展提升，基于深度学习的目标检测算法(如 Faster R-CNN^[8]、YOLO^[9-17]、SSD^[18]等)广泛应用到苹果检测任务中。这些算法能够自动提取学习苹果的多层次特征，不仅具有较高的检测精度，还展现出更强的鲁棒性。其中 YOLO 系列算法相比 Faster R-CNN 等两阶段算法能够更好的兼具精度和实时性，同时也具有良好的部署生态，更适合用于实际应用中。YOLOv5 相较于早期 YOLO 版本，在推理速度、模型大小和精度之间达到了很好的平衡。尤其是其轻量级变种非常适合边缘设备或资源受限场景，便于快速部署和优化。YOLOv7 在多个公开检测数据集上刷新了精度记录，在速度和准确率之间实现了新的最优点(State-of-the-Art)。它引入了 E-ELAN 结构、任务自适应赋权(Task-Aware Learning)等机制，使得模型在小目标检测、复杂场景下表现更加出色。它适合对检测准确率要求较高、硬件资源相对充足的场景，能够提升整体检测效果。YOLOv11 作为 YOLO 系列的最新版本，融合了 Transformer、改进的注意力机制、动态卷积等新技术，相比前代在检测精度、泛化能力和复杂场景适应性上进一步提升。它还针对高分辨率图像检测和超大规模数据集做了优化，适合在更高要求(如复杂果园环境、多类别目标、小目标密集)的检测任务中使用，能进一步提升系统的整体鲁棒性与精度。因此本文选用这三种模型作为基准模型进行研究。

在定位方面，研究者们也将深度学习与传统测距技术相结合，通过单目、双目和深度相机等方法实现苹果的三维定位^[19]。单目相机结构简单、成本较低，适合快速部署，但由于缺乏真实深度信息，其测距依赖苹果尺寸等先验知识，易受苹果大小、姿态变化和环境光照等因素干扰，定位精度有限。深度相机则可直接获取目标点云信息，并结合深度神经网络进行三维重建，适用于结构化或半结构化场景中对精度要求极高的任务，但其对环境光线敏感、成本较高，且在强光或户外复杂环境中稳定性较差。相比之下，双目相机在成本、精度和环境适应性之间取得了良好平衡。通过提取左右图像间的视差信息，双目系统能够实现无接触、非侵入式的高精度距离估计，定位误

差可稳定控制在厘米级，足以满足果园采摘作业的精度需求。同时，双目相机结构灵活，适配性强，可在复杂光照和非结构化场景下稳定运行，适合部署在移动平台或机械臂等自动采摘设备上。更重要的是，双目视觉系统在开源算法、软硬件协同开发等方面具备良好生态，便于系统集成与后期优化。因此本文选用双目相机进行研究。

1.2.1 苹果检测研究现状

在蔬果目标检测领域 Font 等^[20]针对夜间和遮挡条件下的红葡萄检测问题，提出了一种基于球形表面镜面反射峰的算法，当遮挡率超过 75% 时，误差仅为 14%。何斌等^[21]通过改进 YOLOv5 模型并结合自适应锚框，成功检测夜间环境下的红色西红柿，检测精度达 97.6%。王金鹏等^[22]提出轻量级 YOLOv4-LITE 检测火龙果的检测算法，在检测准确率和检测速度上有着不错的表现，但由于训练数据集较少且火龙果背景较为单一，该模型的通用性受到了限制。Yue 等^[23]开发了一款适用于移动设备的草莓成熟度检测应用，能够快速且准确地检测草莓。张同勋等^[24]提出一种适用于自然光和日光灯条件下的葡萄摘取点检测方法，精度达到 90%。张瑞合等^[25]通过直方图曲线分割算法，准确区分西红柿与叶子，实现精确检测。易诗等^[26]提出的 YOLOv4 春见柑橘检测模型有着不错的检测精度，平均精度为 94.6%，但未充分考虑遮挡物的影响。Arfi 等^[27]提出的西红柿检测模型，基于西红柿成熟与未成熟之间的颜色差异，精度达到 96%。Huang 等^[28]首先检测图像中的草莓，然后根据其质量和最低点位置计算摘除点，准确检测率为 84%。李国进等^[29]提出的 ISD-YOLOv3 芒果检测模型，在测试集上精度达到 94.91%，检测速度达到 85 帧每秒。

具体到苹果检测领域，Feng 等^[30]提出一种基于链码信息的相邻苹果分离算法。该方法通过图像预处理提取水果区域，跟踪轮廓获取链码信息，并利用链码差变化寻找凹点，最终匹配凹点对实现苹果分离。实验表明其平均分离成功率超过 91.2%，具备实际应用价值。但该方法依赖明显的边界凹凸，难以应对链码信息利用不足及果实被遮挡等问题。江梅等^[31]提出的 K-means 凸壳目标识别算法，利用苹果的形状特征提取其位置信息，识别效果较好。但在苹果遮挡严重的情况下，识别性能陡然下降。荆伟斌等^[32]提出的 Faster-RCNN 苹果检测网络虽识别精度达到了 91%，但模型部署受限且检测时间长实时性差。龙燕等^[33]设计的疏果期苹果检测网络平均精度为 95.2%，检测准确率为 92.7%，召回率为 91.0%，但存在内存占用较大、实时性较差的问题。

尽管当前在大的蔬果领域和具体的苹果领域目标检测技术已取得一定的进展，但

在精度、速度和部署性能方面仍存在不足，尤其是在背景复杂或对实时性要求较高的应用场景中，仍需探索更具精度与实时性的模型，以更好地满足实际应用需求。

1.2.2 苹果定位技术研究现状

为了实现苹果采摘的自动化，首先需要通过图像检测系统获取苹果目标，然后利用相机计算出其位置信息。

在单目相机方面，Mehta 等^[34]通过单目相机成功测出取柑橘的三维坐标，且误差在 15mm 以内。然而，由于苹果尺寸变化会影响检测效果，该方法更适用于尺寸固定的苹果检测。王壮飞等^[35]利用 OpenCV 的级联分类器与单目视觉测距技术，实现了采摘机器人对目标水果的检测与定位。李振雨等^[36]单目相机针对不同风力条件下苹果的采摘技术，利用角点提取技术计算曲率峰值点，获得苹果握柄点的位置。

在双目相机方面，Si 等^[37]的双目苹果检测与定位系统，定位误差在 40cm 至 150cm 距离范围内小于 2cm。牟其松^[38]利用 SGBM 算法获得定位关键的视差信息，成功计算出苹果的三维坐标。

在深度相机方面，赵海东^[39]使用 Kinect 相机并结合点云处理技术，获取到九个苹果目标的二维点云坐标平均值，计算出苹果的三维空间坐标。Jordi Gene-Mola 等^[40]则采用两个 Kinect 相机外加深度传感器来构建了一个 RGB-D 相机系统，实现对苹果高精度的检测与定位。

1.2.3 边缘端研究现状

随着物联网和人工智能的快速发展，边缘计算成为解决云计算中心化问题的重要手段。作为边缘计算的核心组件，边缘设备的研究现状主要集中在硬件性能提升、能耗优化、兼容性设计以及应用场景适配等方面。目前，边缘设备已大量应用在智能工业、智慧农业、无人驾驶和医疗监测等领域。由于边缘设备通常运行在能源受限的环境中(如电池供电设备)，能耗优化成为研究的核心问题之一。特别是在智慧农业和果园采摘等应用中，系统通常运行在室外复杂环境中，依赖无线通信，且对响应速度、能源效率和设备稳定性要求极高。相比传统依赖云端计算的方式，边缘设备可以在本地快速完成目标检测与决策处理，大幅降低通信延迟和对网络连接的依赖，提升系统的实时性和可靠性。

因此，本研究选用边缘端设备作为苹果检测与定位系统的计算平台，不仅能够提升系统的实时性、稳定性与能效比，还能更好地适配野外环境的复杂部署需求，为实

现自主、智能、高效的果园作业提供有力支撑。

张冬妍等^[41]提出一种轻量化 YOLOv8s 果园李子检测算法，并成功部署在 Jetson Nano 4GB 边缘设备，检测帧率达到 48.3FPS。郑镇辉等^[42]对 YOLOv8l 模型进行改进，部署英伟达边缘设备后，对竹笋的检测和分割分别达到 mAP@0.5 为 0.896 和 0.891，单张图片推理时间 0.0914 秒，且仅需 12.2GB 显存。

1.3 论文主要研究内容

本文研究以果园内的苹果作为研究对象，采用目标检测算法和三维定位技术实现对苹果的检测与定位。本文主要有以下几点工作内容：

(1) 针对果园环境中苹果在遮挡严重、尺度不一等复杂条件下检测精度不足的问题，提出基于 YOLOv5 的检测改进方法。YOLOv5 在轻量化与实用性之间达到了较好平衡，适合作为初始探索阶段的基准模型。在数据处理方面，采用了对比度调整、亮度调整、随机遮挡、反色图像等多种数据增强策略，结合人工标注构建了鲁棒性强的数据集。在模型改进方面，YOLOv5 引入了软非极大值抑制(Soft-NMS)替代传统 NMS，提高重叠目标的检测精度；采用最优传输分配(OTA)优化标签分配策略，增强模型在多目标环境下的匹配准确性；同时将主干网络中的部分卷积模块替换为 FasterNet 块，以提高特征提取效率和速度。

(2) 针对复杂果园场景中苹果检测在细节检测方面仍存在的困难，进一步提出基于 YOLOv7 的检测优化方法，以提升模型对颜色、纹理等细节特征的检测能力。YOLOv7 在结构设计和性能表现方面优于 YOLOv5，特别适用于对细节要求较高的果园场景。具体改进包括引入最小点距离交并比(MPDIOU)作为新的回归损失函数，提高边界框回归的准确性；加入高效通道注意力机制(ECA)，增强模型对颜色、纹理等细节特征的敏感性；并将主干网络中的标准 3×3 卷积替换为深度可分离卷积(DSConv)，在降低计算复杂度的同时保持检测性能。

(3) 针对果园环境中苹果仍存在多尺度分布和空间位置变化下检测精度下降的问题，进一步提出基于 YOLOv11 的融合增强方法，以提升模型在复杂环境中的检测鲁棒性与精度。YOLOv11 融合了多种更先进的模块和机制，具备更强的表达能力和适应能力。首先在颈部网络中引入空间增强注意力模块(SEAM)，提升模型对目标空间分布的关注能力；其次使用全维度动态卷积(ODConv)替代原有卷积结构，使模型能动态适应不同特征尺度，提高复杂场景下的表现力；最后加入 C3k2_ScConv 模块，结合

C3k 模块与 ScConv 进行结构重构, 进一步增强模型的空间与通道特征融合能力。

(4) 针对苹果目标精准检测与空间定位后的机械化自动采摘的综合需求, 提出基于 YOLOv11 与双目视觉融合的检测定位方法。根据前面三个模型优化的实验结果, 系统最终选用精度与鲁棒性兼备的 YOLOv11 模型作为目标检测核心模块, 以提升在复杂果园环境下的识别能力与应用稳定性。同时, 结合 Intel RealSense D415 双目相机进行三维距离测量, 首先完成相机标定以获取内外参矩阵, 并通过立体匹配算法生成视差图转换为深度信息, 随后采用排序算法与双边滤波对深度数据进行去噪处理, 最后基于三角测量原理计算苹果的空间坐标, 实现高精度定位。为实现系统的边缘端部署, 本文选用国产高性能计算平台 RK3588, 并对 YOLOv11 模型进行了转换与优化, 以适应边缘设备的算力限制并提升推理效率。实验结果表明, 该系统平均深度误差为 0.0148m, 检测定位效果较稳定, 基本满足果园自动采摘的实际应用需求。

1.4 论文组织框架

本文由六章构成, 具体结构如下:

第一章: 绪论。本章首先是对果园苹果检测的研究意义和研究背景作相关的介绍, 重点突出苹果检测和定位技术对苹果采摘自动化的重要经济意义。其次是就国内外苹果的检测技术和苹果的定位技术的研究现状展开讨论, 概述该领域的当前研究进展, 接着详细阐述本文的研究目标、方法及预期成果。最后, 介绍本文的结构框架。

第二章: 相关技术理论基础。本章的主要内容是关于苹果检测和定位技术的相关理论知识进行详细的介绍和说明。从深度学习神经网络的具体理论知识, 再到双目定位技术的原理以及定位的方法步骤, 最后到应用层面, 包括边缘计算、模型转换以及模型部署等进行详细的说明。

第三章: 基于深度学习的果园苹果检测算法。本章先是通过数据增强、数据标注等步骤构建出一个高质量的果园苹果数据集。然后再详细说明 YOLOv5、YOLOv7、YOLOv11 三种果园苹果检测算法模型的网络结构及其模块的改进方法。最终, 通过评估指标、对比与消融实验以及实验结果分析等手段, 验证了改进算法的有效性。

第四章: 苹果三维定位。本章主要是对双目定位技术的具体说明。首先是介绍并分析各种相机的优缺点, 其次是对坐标原理和坐标转换的具体方法进行说明, 接着是通过标定板对相机进行标定, 获取相机的内外参数等, 最后说明深度信息的获取方法, 并通过实验数据说明有效性和实际应用价值。

第五章：果园苹果检测定位系统的实现。本章主要是对果园检测系统的实现作一个详细的介绍。主要从系统的功能分析出发到硬件的选型和介绍，再到系统的设计，最终搭建出苹果检测定位系统，并对系统进行测试。

第六章：总结与展望。本章主要是对全文内容的一个总结，并对未来的研究工作做出说明。

第 2 章 相关技术理论基础

深度学习是在机器学习和人工神经网络基础上的一种深层神经网络，它结合大数据与高性能计算的发展而兴起。其核心技术特点在于具备自动提取特征的能力，这些特征被称为深度特征或深度表示。相比人工设计的特征，深度特征具有更强的表示能力和鲁棒性，使得深度学习的本质可归结为特征表示学习。

深层神经网络是深度学习实现特征自动提取的关键模型，其核心在于通过一系列非线性变换的嵌套逐步提取数据中的复杂特征。深度学习作为一个跨学科技术领域，融合了数据科学、统计学、工程学、人工智能和神经生物学的研究成果，是机器学习的重要分支。图 2-1 展示以增量方式学习的五层深度网络。

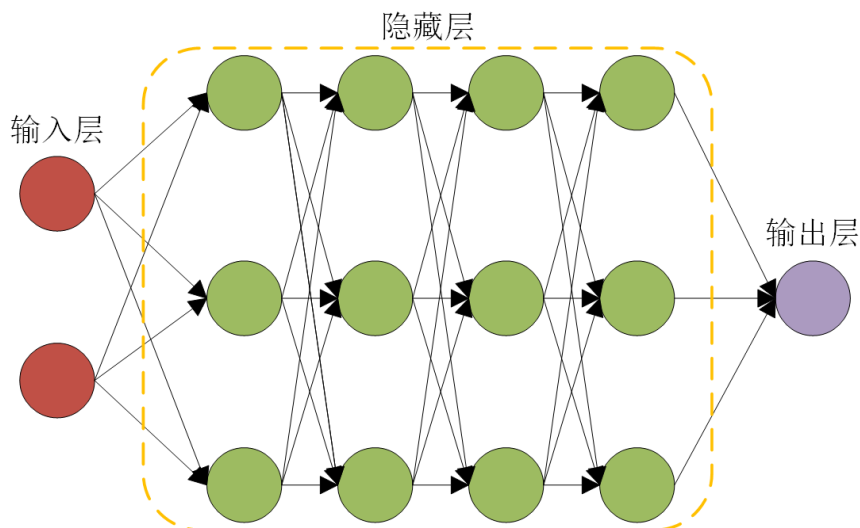


图 2-1 五层深度学习模型

Figure 2-1 Five-layer deep learning model

2.1 卷积神经网络的组成

卷积神经网络(Convolutional Neural Network, CNN)^[43]是一种深度学习模型，其设计灵感源于生物视觉系统，通过模拟人类视觉的处理方式，可以高效地检测和处理图像数据。CNN 的核心特点包括局部连接、权值共享和层次化结构。CNN 包含输入层、卷积层、池化层、激活函数层和全连接层。CNN 通过多个卷积层和池化层的组合完成特征提取任务，并通过一个或多个全连接层将提取的特征映射到最终的输出结果，其结构如图 2-2 所示。

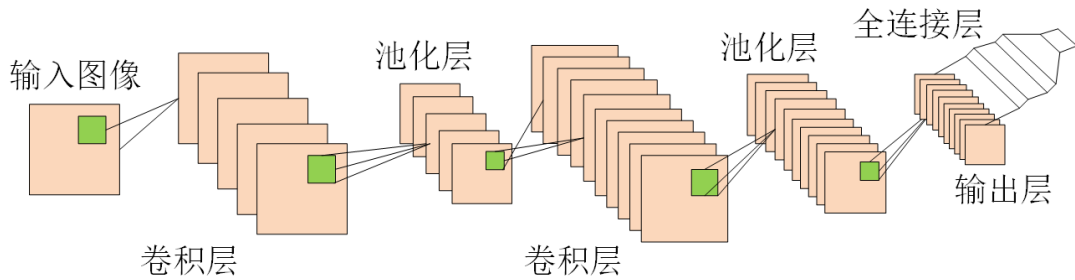


图 2-2 卷积神经网络架构示意构图

Figure 2-2 Convolutional neural network architecture schematic composition

2.1.1 输入层

输入层负责接收网络的原始数据或经过预处理的数据。根据应用领域的不同，输入数据可以有多种形式。图像检测任务中，输入层接收的是原始的 RGB 三维彩色图。

卷积神经网络会将输入层的数据传递到后续的卷积层、池化层等，进行特征提取和转化，最终通过全连接层对提取到的特征进行汇总，并输出分类结果。

在处理过程中，根据计算能力、存储容量和网络结构的不同，卷积神经网络通常采用批量处理方式。每次输入层可以同时处理多个图像，这个批量大小是由硬件配置和具体任务决定的。

2.1.2 卷积层

卷积层主要功能是特征提取，通过卷积核(也称滤波器或特征矩阵)对输入数据进行处理，捕捉原始数据中潜在的局部关联性和特征。卷积运算实质上是对输入数据矩阵与卷积核执行逐元素相乘并求和的数学操作，最终生成一个特征映射，用来表示图像中的特定局部特征。

在卷积操作中，输入数据矩阵与卷积核滑动相乘的过程中涉及不同的填充策略，主要包括零填充和有效卷积。零填充是将零值添加到输入矩阵的边缘，以保持输出特征图的尺寸不变或略有减少，从而保留更多的边缘信息。图 2-3 的上半部分展示了这一过程。有效卷积则是在不进行任何填充的情况下，直接对输入矩阵进行卷积操作。这种策略会舍弃无法完整运算的边缘部分，导致输出特征图的尺寸减小。图 2-3 的下半部分展示了有效卷积的过程。

通过卷积核的滑动窗口机制，卷积层能够在输入数据的不同局部区域进行特征提取，并通过多层卷积叠加逐步提取从低级特征(如边缘、纹理)到高级特征(如物体形状、结构)的信息。这种逐层特征抽象和表达的能力，是卷积神经网络实现高效图像处理

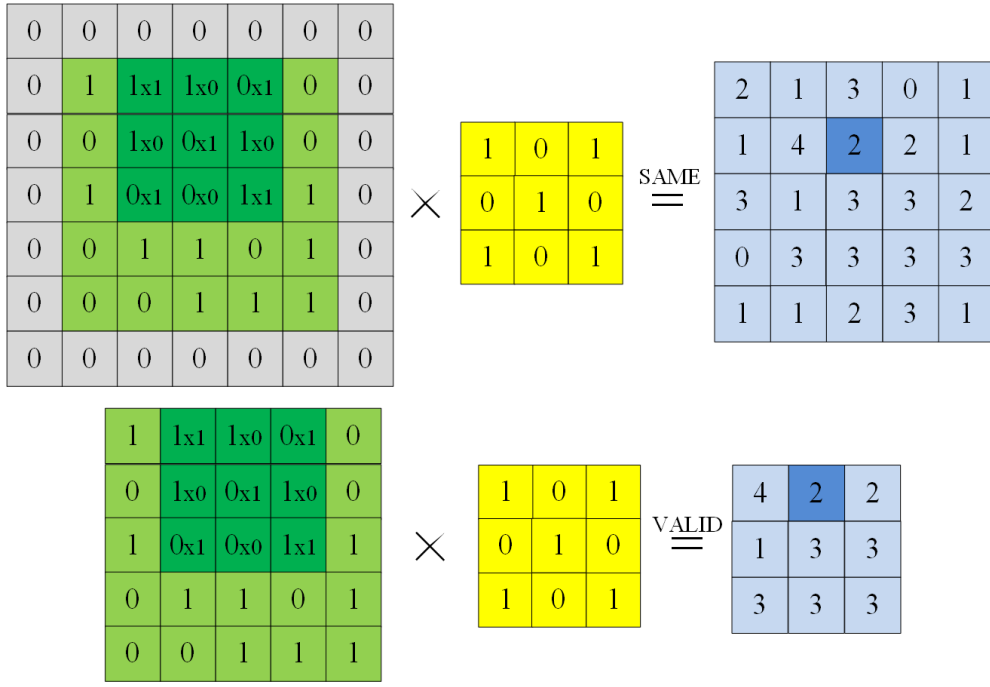


图 2-3 卷积填充策略

Figure 2-3 Convolution-filled strategy

关键。卷积神经网络的计算公式为：

$$N = (W - F + 2P) / S + 1 \quad (2-1)$$

式 2-1 中 S 为步长大小， P 为填充值的大小， F 为卷积核大小， W 表示输入大小， N 为输出大小。

2.1.3 池化层

池化层，亦被称作下采样层，其核心功能在于对感受野内的特征进行筛选与降维，提炼出区域内的关键特征。该层通过缩减输入特征图的空间维度，降低输出特征的规模，进而减少了模型的参数总量及计算负担。池化层不仅提高了模型的计算效率，还增强了模型的平移不变性，使其对输入图像的微小平移或变形不那么敏感。

池化层根据池化操作的不同，通常分为最大池化、平均池化和求和池化三种类型。最大池化是指提取感受野内的最大值，通常能保留更为明显的特征信息，常用于大多数任务中。平均池化是指计算感受野内所有数值的平均值，常用于一些特定的应用中，能够较好地保留整体特征，但对于细节信息的保留不如最大池化。求和池化是指计算感受野内所有数值的和，虽然不如最大池化常见，但在某些任务中也能提供有效的信息。在实际应用中，最大池化和平均池化最为常用，池化操作如图 2-4 所示。

池化操作本身不包含可训练的参数，只有两个重要的超参数：步幅(Stride)和卷积

核大小(Kernel_size)。池化过程类似于卷积操作，通常是通过滑动窗口的方式对输入图像进行处理。通常情况下，步幅和卷积核大小相等，以保持特征图的空间结构。

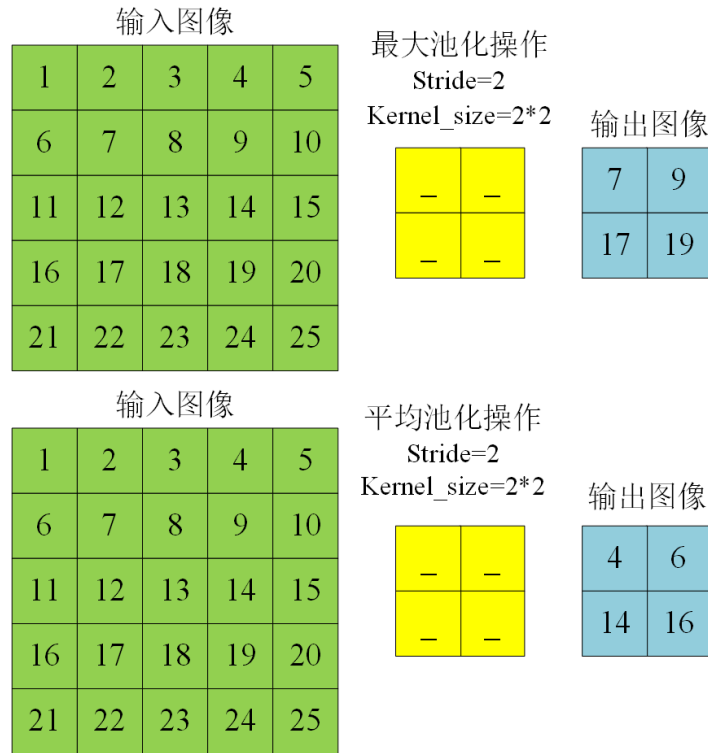


图 2-4 最大池化操作和平均池化操作

Figure 2-4 Maximum and average pooling operations

2.1.4 激活函数层

激活层的主要功能是对卷积层所提取的特征实施非线性转换。鉴于卷积运算实质上是输入矩阵与卷积核矩阵的线性结合，激活层通过引入非线性映射，有效地突破这种线性限制，使得神经网络能够学习和表示更复杂的特征，并拟合非线性的数据模式。

激活层主要由激活函数组成，激活函数是一个非线性函数，它作用于卷积层的输出结果，将线性变化转化为非线性变化。如此一来，激活函数可增强网络的表达能力，从而更好地检测输入数据中的复杂模式。在卷积神经网络中，ReLU(Rectified Linear Unit)常被用作激活函数。ReLU 的函数形式如式 2-2 所示：

$$\text{ReLU}(x) = \max(0, x) \quad (2-2)$$

ReLU 函数的特点是：当输入值小于 0 时，输出为 0；而当输入值大于或等于 0 时，输出值保持不变。通过这种方式，ReLU 能够有效地引入非线性，同时避免了传统激活函数可能出现的梯度消失问题。

2.1.5 全连接层

全连接层在 CNN 中用于汇总前面各层提取的特征，并将高维特征映射到最终输出空间。全连接层通常位于网络的末端，特别是在分类任务中，它将卷积层和池化层提取的高维特征转化为所需的输出。

全连接层中，每个神经元与上一层的所有神经元相连接，从而能够整合来自不同特征图的信息，形成对整个输入数据的综合表示。全连接层的输出通常是二维的，一维代表样本批次(Batch)，另一维对应任务的目标输出(例如分类任务中的类别数)。全连接层拥有大量参数，通常位于网络的后几层，负责将深层特征抽象为更高层次的语义信息，并最终输出特定任务的预测结果。

2.2 双目定位技术

双目定位(Stereo Localization, SL)，又称双目视觉定位或双目立体视觉定位技术，是一种通过两台相机(或摄像头)获取图像数据来进行三维重建和空间定位的方法。其基本原理是利用立体视觉原理，通过两个从不同视角拍摄同一场景的相机，分析图像差异，计算物体在三维空间中的位置。

2.2.1 定位原理

双目定位技术的核心是立体视觉。立体视觉通过两个相机(通过模拟人眼的左右视角)捕捉同一场景的不同视图，利用视差——即同一物体在左、右相机图像中位置的差异——来计算物体的深度信息。具体而言，视差是指在两个相机成像平面上，某个物体在左、右相机的成像位置之间的水平差异。通过计算视差，可以推算物体到相机的距离，得出其深度信息。

相机成像原理中，每个相机拍摄到的图像是二维的代表场景中物体的投影。物体的三维位置可以通过相机的内参(如焦距、主点坐标等)和外参(如相机位置、姿态等)反向推算得到。给定两个相机拍摄的图像，假设某个物体在左相机图像中的像素位置为 (X_L, Y_L) ，在右相机图像中的像素位置为 (X_R, Y_R) ，则视差 d 可以通过 $d = X_L - X_R$ 来定义。

立体匹配是双目视觉系统中的关键步骤，它指的是在左、右图像中找出对应的点。这个过程通常需要通过图像处理技术(如特征提取、匹配算法等)来实现。常见的匹配方法包括：基于区域的方法，通过计算图像块之间的相似度来寻找匹配点；基于特征的方法，提取图像中的关键点或特征(如 SIFT^[44]、SURF^[45]、ORB^[46]等)，再进行匹配；以及全局优化方法，如图割等优化算法，可以通过全局信息来提升匹配精度。

在立体视觉系统中,几何模型用于描述相机如何将三维点投影到二维图像平面上。常用的几何模型是针孔相机模型,数学表达式如式 2-3 所示,其数学表达式通过考虑相机的内外参数,准确描述三维空间到二维图像的投影关系。

$$z = \frac{fX}{x} \text{ and } z = \frac{fY}{y} \quad (2-3)$$

式 2-3 中 f 是相机的焦距, X,Y,Z 是物体在三维空间中的坐标, x,y 是物体在图像平面上的坐标。

2.2.2 定位步骤

双目定位基于立体匹配和三角测量的原理,通常分为以下几个主要步骤:

(1) 相机标定。相机标定的目的是消除畸变、获取相机的内参和外参以及确定相机与世界坐标系之间的关系。标定过程包括内、外参标定。内参标定用于获取焦距、主点位置、径向畸变系数等信息;外参标定则确定两个相机的相对位置和方向,通常通过棋盘格标定法或立体标定法完成。

(2) 图像采集。双目相机系统由两个摄像头或两个单目相机组成,通常这两个摄像头或两个单目相机以平行方式放置,并且具有固定的基线(即相机之间的距离)。两个相机从不同的视角拍摄同一场景,生成左右图像。为了确保获取准确的深度信息,两个相机必须尽可能精确对齐,并保持适当的视差(即相机间的距离和角度)。

(3) 图像预处理。图像预处理包括去噪、对齐、灰度化等操作,这些步骤有助于提高后续匹配和深度计算的准确性。

(4) 特征提取和匹配。由于两个摄像头视角不同,同一物体在左图和右图中的投影位置会有偏移。通过匹配左右图中的对应特征点,可以计算出每个特征点的视差(Disparity),即物体在两个图像中的水平位置差异。常用的特征提取方法包括角点检测(如 Harris 角点)、SIFT/SURF 特征点(尺度不变特征转换)和 ORB 特征点(基于 BRIEF 的快速算法)。

(5) 视差计算。根据两幅图像相互匹配的特征点,计算图像中每对对应点之间的视差。视差图是一个二维图像,显示了每个像素对应的视差值。此外视差与物体离相机的距离是呈反比关系的。

(6) 深度计算。通过视差图和相机的几何模型,可以将视差转换为深度信息。根据视差值以及相机的参数(如基线长度、焦距等),利用三角测量原理计算出每个特征点的

深度，从而得到物体的距离坐标。具体计算公式为：

$$Z = \frac{fB}{d} \quad (2-4)$$

式 2-4 中 Z 是物体到相机的深度(距离)， f 是相机的焦距， B 是两个相机间的基线长度(相机间的水平距离)， d 是视差，即两个图像中对应点的水平距离差。

(7) 定位与应用。通过相机计算场景中物体的距离信息，并结合物体检测获得的二维坐标，可以确定物体的三维位置，从而实现精准定位。

最后与其他定位技术相比双目定位具有多个优势。首先，经过精确的深度计算，双目视觉能够实现高精度的三维定位，尤其在近距离内表现突出。与单目视觉相比，双目系统无需依赖额外的传感器，只需利用标准相机即可实现精准的空间定位。此外，双目视觉系统的硬件成本相对较低，相较于激光雷达等传感器，成本更具优势。双目定位还具有较强的适应性，能够在不同光照条件下工作，适用于多种室内外场景。

2.3 边缘计算

边缘计算是一种在数据源附近做计算、存储和网络功能的分布式计算架构。通过将数据处理任务分配到网络边缘的设备(如路由器、传感器、智能终端等)，此技术能够降低数据传输延迟，减轻中心服务器的负担，并增强实时处理能力。

在传统的云计算模型中，数据通常会从终端设备传输到远程数据中心进行处理，然后将结果返回。这种做法会导致高延迟，特别是在要求实时性较高的应用场景中。边缘计算通过在本地进行数据处理，大大降低了响应时间，数据可以更快地得到处理和反馈，适用于需要低延迟、高带宽的场景。

边缘计算的关键特点是低延迟、低带宽、可靠性与容错性，以及分布式存储和计算。由于数据处理发生在接近数据源的地方，边缘计算大大减少了数据在网络中的传输时间，降低了延迟，这对自动驾驶和智能制造等需要实时数据处理的应用极为重要。同时，边缘计算在边缘设备上进行初步数据处理，只将精简后的重要数据发送至数据中心，减轻网络带宽的压力并优化带宽的使用。即便部分边缘设备或连接出现故障，边缘计算架构仍能保持本地计算和处理，提升系统的可靠性和容错性。此外，边缘计算将计算和存储资源分布在网络的边缘节点上，这些节点具备一定的计算和存储能力，能够在本地高效处理大量数据，并做出快速响应。

边缘计算的应用领域涵盖自动驾驶、智能制造、物联网和视频监控等多个行业。

在自动驾驶中，车辆需要实时处理来自激光雷达、摄像头等传感器的大量数据，边缘计算使得车辆能够在本地快速分析这些数据并作出决策，减少对远程服务器的依赖，保证系统的响应速度和安全性。在智能制造领域，边缘计算能够实时监控生产设备，优化工艺流程并预警潜在故障，在本地分析机器生成的数据，能够更迅速地采取措施，提升生产效率并降低成本。在物联网应用中，边缘计算可以在本地对大量设备生成的数据进行预处理，筛选出有价值的数据后再发送至云端，减少数据传输量、降低带宽消耗，实现实时响应。视频监控系统中，边缘计算可以在摄像头附近进行视频分析，实时检测和处理图像中的事件或异常行为，减少视频数据的传输量，提高响应速度。

虽然边缘计算在许多领域有着优势，但也面临一些难题。如边缘计算涉及大量设备和节点，如何高效地管理和维护这些设备成为一大难题。随着设备数量的增加，设备配置、升级和故障修复需要更加灵活和高效的管理机制。另外由于数据在边缘设备上进行处理和存储，必须要保证数据在传输、存储以及处理过程中的隐私性和安全性。

2.4 模型转换与模型部署

模型转换是指将一个机器学习或深度学习模型从一个开发环境或框架迁移到另一个框架或平台的过程。这通常是因为不同的深度学习框架或硬件平台对模型的支持有所不同，需要进行适配。其步骤如下：

(1) 首先根据边缘部署设备的计算能力、内存限制以及对实时性的要求，选择合适的目标检测模型。模型的选择应充分考虑设备的硬件特点和应用场景的需求。YOLO系列模型以其较高的检测精度和推理速度而闻名，适用于需要实时性较高的应用场景。YOLO模型的优势在于其能够在保证精度的同时，有较快的推理速度，这使得它成为资源受限设备上的理想选择。

(2) 其次为了减轻边缘设备的存储和计算负担，并尽可能保持目标检测模型的精度，必须对模型进行压缩和优化。

(3) 然后对压缩和优化后的目标检测模型进行转换，以使其适应不同的硬件和平台。常见的模型转换格式包括 ONNX(Open Neural Network Exchange)、TensorRT、MNN(Mobile Neural Network)、NCNN(Neural Network Inference Framework) 和 RKNN(Rockchip Neural Network)。这些格式可以提高模型的兼容性和可移植性，确保在多种边缘设备上顺畅运行。例如，ONNX 作为一个开源的跨平台深度学习模型交换格式，支持多种主流深度学习框架(如 TensorFlow、PyTorch、Caffe2 等)之间的模型转

换，为开发者提供了极大的灵活性。其他格式如 TensorRT 专为 NVIDIA GPU 加速优化，MNN 和 NCNN 则更适合在移动设备或嵌入式设备上部署。通过将优化后的模型转换为适应不同硬件加速的格式，可以最大程度地提高模型推理速度，降低硬件资源的消耗，满足边缘设备对实时性和计算效率的需求。

(4) 最后选择合适的推理引擎和框架，根据不同设备类型和操作系统进行部署。不同设备类型和操作系统需要根据其特点选择最适合的推理引擎。例如，TensorFlow Lite 和 Paddle Lite 是专为移动设备和嵌入式设备优化的推理引擎，能够提升推理效率并减少计算资源消耗。而 RKNPU 则专为 RKNN 格式的模型进行优化，能够加速推理过程并充分发挥 Rockchip 硬件的计算能力。这些推理引擎和框架不仅支持硬件加速，还能够实现异构计算，提升模型的推理速度和整体性能。

(5) 此外，为了确保模型在实际运行中达到预期效果，还需要对边缘设备进行适当的配置和调试。这包括优化硬件资源的利用，调整推理引擎的参数设置，进行性能测试和精度验证等，以确保最终的推理结果既符合精度要求，又能在实时性和效率上满足边缘计算的需求。

2.5 本章小结

本章主要介绍了与目标检测相关的基础技术与理论，涵盖卷积神经网络架构、双目定位技术、边缘计算，以及模型转换与部署等方面的内容。

首先，在卷积神经网络架构部分，系统阐述了从输入层到激活函数层的各个关键组成部分，解释了其在特征提取中的作用与机制。

随后，介绍了双目定位技术，通过对定位原理和定位流程的解析，说明了如何利用两台摄像头获取的视差信息实现空间目标的三维定位。

接着，阐述了边缘计算的基本概念，并分析了其在提升数据处理效率、降低系统延迟及节省带宽资源方面的优势。边缘计算在大规模物联网环境中尤为重要，能够满足实时响应的处理需求，尤其适用于对时效性要求较高的应用场景。

最后，模型转换与部署部分探讨了在实际应用中如何选择合适的模型结构和硬件平台，并将训练完成的模型有效地转换为适配目标设备的格式，从而确保其在边缘设备上高效稳定运行。

本章内容为后续章节中关于目标检测与模型优化的深入研究提供了必要的技术背景和理论支撑。

第3章 基于深度学习的果园苹果检测算法

本章将以 YOLO 系列中的 YOLOv5、YOLOv7 和 YOLOv11 三种目标检测算法为基准，针对果园苹果检测任务的特定需求，进行相应的优化与改进。

在目标检测算法中 YOLO 系列能够很好的兼具精度和实时性，同时也具有良好的部署生态，更适合用于实际应用中。其中 YOLOv5 相较于早期 YOLO 版本，在推理速度、模型大小和精度之间达到了很好的平衡。尤其是其轻量级变种非常适合边缘设备或资源受限场景，便于快速部署和优化。YOLOv7 在多个公开检测数据集上刷新了精度记录，在速度和准确率之间实现了新的最优点(State-of-the-Art)。它引入了 E-ELAN 结构、任务自适应赋权(Task-Aware Learning)等机制，使得模型在小目标检测、复杂场景下表现更加出色。它适合对检测准确率要求较高、硬件资源相对充足的场景，能够提升整体检测效果。YOLOv11 作为 YOLO 系列的最新版本，融合了 Transformer、改进的注意力机制、动态卷积等新技术，相比前代在检测精度、泛化能力和复杂场景适应性上进一步提升。它还针对高分辨率图像检测和超大规模数据集做了优化，适合在更高要求(如复杂果园环境、多类别目标、小目标密集)的检测任务中使用，能进一步提升系统的整体鲁棒性与精度。

通过对这三种算法的优化与测试，可以全面评估它们在果园苹果检测任务中的性能表现。基于实验结果，最终选取性能最优的模型，为果园苹果检测技术的实际应用提供可靠支持。

3.1 构建苹果数据集

数据集的质量对苹果的检测具有重要的作用，它直接决定了检测的准确率，并影响后续定位任务的精度。因此，构建一个高质量、适合应用场景的数据集是关键。尽管网络上已有大量关于苹果的数据集，但其中大多数仅用于分类任务，这些数据集通常只包含少量苹果目标，且背景相对简单单一。这类数据集训练出的模型通常鲁棒性较差，难以适应果园环境中复杂多变的实际情况，无法满足本文对检测任务的需求。

果园环境具有高度复杂性，苹果在自然场景中可能受到多种遮挡，如树枝和树叶遮挡苹果、苹果与苹果之间重叠遮挡等。当苹果被遮挡时，遮挡物会干扰苹果之间的特征表达，模型难以有效检测和推理，从而导致模型在实际应用场景中的性能下降。

因此，为提升模型的泛化能力，本文从实际果园场景出发，收集自然环境下的果园苹果图像样本。这些样本涵盖了多种苹果遮挡情况，包括枝叶遮挡、苹果重叠等情形，共计 1357 张分辨率为 960×1280 的图像。部分果园苹果图像样本如图 3-1 所示。



图 3-1 部分果园苹果图像样本

Figure 3-1 Image sample of some orchard apples

3.1.1 数据增强

数据增强是一种提高模型泛化和减少过拟合的有效手段。通过多样化处理训练数据，数据增强不仅能有效扩充数据集规模，还能提升模型在不同环境下的鲁棒性。针对果园场景中苹果检测的实际需求，本文设计并采用了一系列数据增强策略，以增强模型在复杂自然场景中的表现。

本文主要采用了对比度调整、亮度调整、随机遮挡和反色图像生成等数据增强方法，最终得到共 2500 张数据图片。在颜色变换方面，通过调整对比度和亮度，模拟光照条件变化以及苹果表面反射光强度差异对检测任务的影响，提升模型的适应能力。对于复杂的果园场景，随机遮挡方法被用于模拟枝叶或其他苹果对目标的遮挡情况。此外，生成反色图像用于模拟特殊光照或成像条件下可能出现的颜色反转现象，进一步提高模型的鲁棒性。部分果园苹果数据增强图像样本如图 3-2 所示。



图 3-2 部分果园苹果数据增强图像样本

Figure 3-2 A sample of data-enhanced images of some orchard apples

3.1.2 数据标注

数据标注是训练深度学习模型的基础。针对果园苹果检测任务的需求，本文采用半自动标注，首先使用检测模型对图像进行初步标注，然后由人工进行校验和调整。此方法不仅提高了标注效率，同时保证了标注的一致性和精确性，减少了标注工作量。

为应对果园复杂环境中的各种情况，本文在标注过程中考虑了树枝和树叶的遮挡、苹果间相互重叠遮挡以及背景干扰等因素。对于被枝叶或其他苹果遮挡的苹果，仅标注其可见部分的边界框，以帮助模型学习遮挡物的特征。针对苹果重叠的情况，每个重叠目标都进行了独立标注，确保每个目标的特征得以准确描述。同时，对小尺寸的目标进行了精细标注，并严格区分苹果目标与背景干扰(如叶片和阴影)，以提高标注数据的准确性和一致性。

3.2 改进 YOLOv5

针对果园环境中苹果在遮挡严重、尺度不一等复杂条件下检测精度不足的问题，本文提出一种基于 YOLOv5 的目标检测改进方法。YOLOv5 相较于早期 YOLO 版本，在推理速度、模型大小和检测精度之间实现了良好平衡，尤其其轻量级变体非常适合部署在边缘设备或资源受限的场景中，具备快速部署与优化的优势。此外，YOLOv5 引入了多尺度检测策略，通过在不同尺度上对图像进行特征提取与目标检测，有效提升了对不同大小目标的识别能力。同时，其自适应训练机制能够根据目标的检测难度动态调整训练过程，从而增强模型在复杂果园环境中的鲁棒性与泛化能力。

3.2.1 YOLOv5 网络结构

YOLOv5 主要由主干网络(Backbone)的 CSPDarknet53(Cross Stage Partial)、颈部网络(Neck)的 PANet(Path Aggregation Network)、头部网络(Head)的检测头三部分组成，其具体结构如图 3-3 所示。

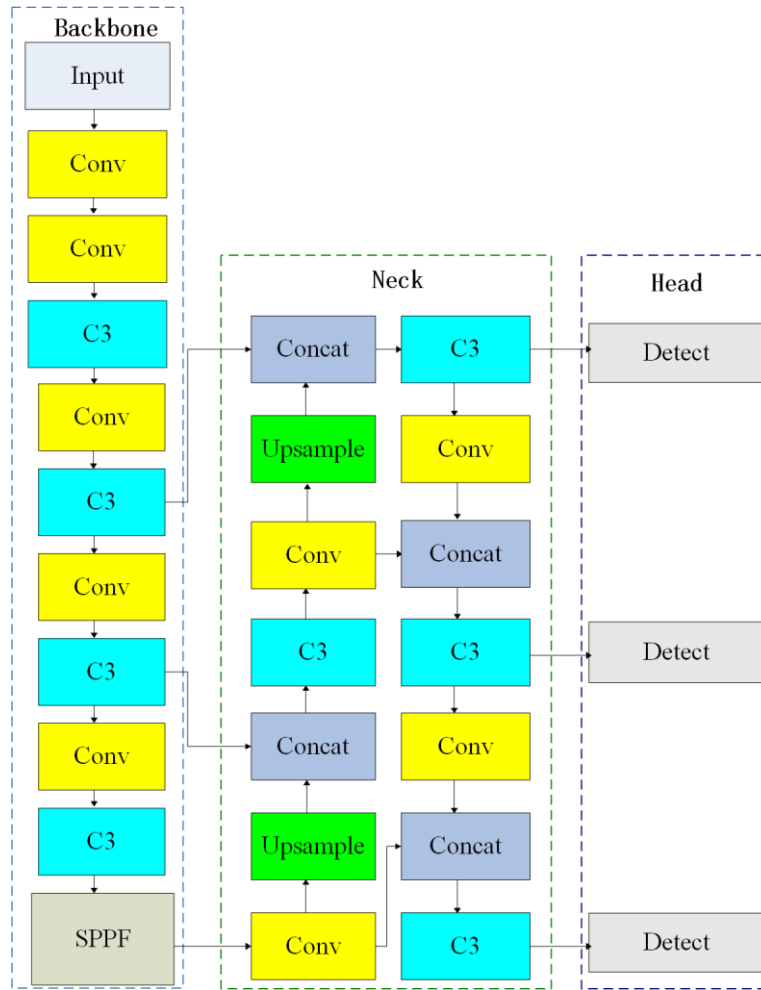


图 3-3 YOLOv5 网络结构

Figure 3-3 YOLOv5 network structure

主干网络：YOLOv5 使用 CSPDarknet53 作为主干网络。CSPDarknet53 由多个卷积层和残差块组成，能够在不同尺度下提取和组合特征。通过逐步提取图像的特征，从低级特征逐渐转化为高级特征，如边缘、纹理、语义和抽象概念，为后续的目标检测任务提供特征信息。

颈部网络：采用 PANet 作为连接主干网络和头部网络的中间层，用于跨层级特征融合。PANet 通过上下文感知和信息融合来提高目标检测的精度，它能够处理来自不同层级的特征图。

头部网络：YOLOv5 的头部结构由卷积层和全连接层组成。头部结构负责生成目标检测的边界框和类别预测。YOLOv5 采用了一种简化的头部结构，使用 3 个不同尺度的输出层来预测不同大小的目标。而检测是模型的最后一部分，它利用一些后处理技术将网络输出的特征图转换为目标检测结果，包含边界框的位置坐标和类别概率。

YOLOv5 引入软非极大值抑制替代传统非极大值抑制算法，提高重叠目标的检测

精度；采用最优传输分配优化标签分配策略，增强模型在多目标环境下的匹配准确性；同时将主干网络替换为 FasterNet，以提高特征提取效率和速度。如图 3-4 是模型修改后的网络结构图。

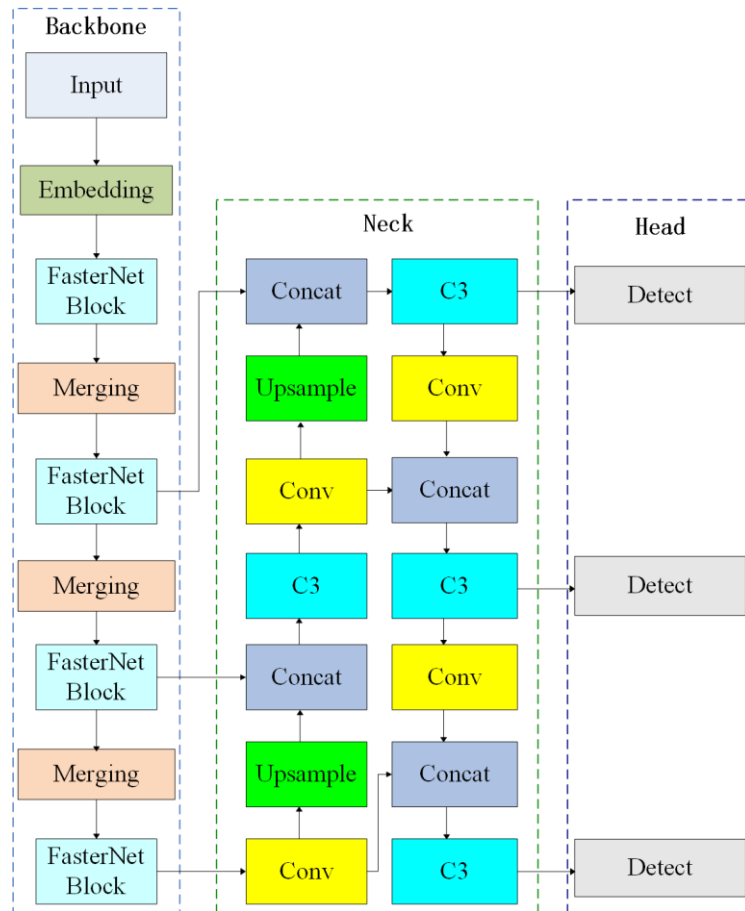


图 3-4 YOLOv5 改进后网络结构

Figure 3-4 Improved network structure of YOLOv5

3.2.2 Soft_NMS

果园环境中的苹果在生长的过程中往往会密集地分布在树上不同的区域，苹果间会相互重叠遮挡，原模型中的非极大值抑制(Non-Maximum Suppression, NMS)算法会将与最高置信度框重叠的其他框抑制掉，导致遮挡重叠的苹果被错误地抑制掉，从而出现错检、漏检的情况使得检测精度下降。

在 YOLOv5 原模型中的 NMS 步骤为：

首先，对每个类别的预测框按照置信度高低进行排序，将具有较高置信度的框放置在前面。

从置信度最高的预测框开始，计算其与其他预测框之间的重叠度。而交并比 (Intersection Over Union, IOU) 是用来衡量边界框与边界框间(或区域与区域间)重叠程

度的指标。它通过计算两个框的交集面积与并集面积的比值来度量它们的重叠程度。

如果某个预测框与置信度最高的框之间的重叠程度超过预先设定的阈值，那么会将该预测框抑制掉，因为它很可能对应着同一个目标物体。然后不断的对余下的预测框进行判断和抑制，直到所有的预测框都被处理完毕。NMS 可以描述为式 3-1 所示：

$$S_i = \begin{cases} S_i, iou(M, b_i) < N_t \\ 0, iou(M, b_i) \geq N_t \end{cases} \quad (3-1)$$

最后，经过 NMS 处理后，每个类别都只保留了置信度最高的框，并移除与其重叠程度较高的冗余框。

软非极大值抑制(Soft Non-Maximum Suppression, Soft_NMS)^[47]在原来的 NMS 的基础上引入一个衰减函数，加入惩罚机制，以概率的形式来降低与最高置信度框重叠的其他框的得分，而不再是直接将得分大于阈值的全置为 0。加入线性衰减函数的 Soft_NMS 公式为：

$$S_i = \begin{cases} S_i, iou(M, b_i) < N_t \\ S_i(1 - iou(M, b_i)), iou(M, b_i) \geq N_t \end{cases} \quad (3-2)$$

式 3-2 中 S_i 为第 i 个预测框的得分， N_t 为阈值， b_i 表示第 i 个预测框， M 表示得分最高的预测框。

引入高斯衰减函数的公式为：

$$S_i = S_i e^{-\frac{iou(M, b_i)^2}{\sigma}}, \forall b_i \notin D \quad (3-3)$$

式 3-3 中，当 σ 取值为 0.5 时衰减幅度最大，预测框筛选速度最快。

引入衰减函数后会降低检测框的得分，而不再直接设置为 0，这就可以保留某些得分高的预测框，从而在后续处理中也有被重新视为正确预测框的可能性。通过这种软性的抑制方式，能够有效缓解模型错检漏检的问题。

3.2.3 最优传输分配

最优传输分配(Optimal Transport Assignment, OTA)^[48]是基于最优传输理论的一种方法，它将标签分配视为最优传输问题。它通过计算网络生成的预测框与真实的人工标注框之间的距离矩阵，并利用线性规划方法找到最优分配方案，从而有效地匹配预测框与真实标注框之间的对应关系，使总距离最小化。原 YOLOv5 网络主要关注于单个对象的最优分配策略，而很少从全局的角度关注理解上下文信息。对于本实验中密

集的苹果目标来说，遮挡重叠的模糊框就难以完整利用上下文信息，从而出现对目标错检漏检的问题。在目标检测中，遮挡的本质原因在于一些预测框无法与真实标注框完全匹配的情况，以及一个真实标注框对应多个预测框的情况。因此通过使用 OTA 来优化标签分配策略，可以很好地解决这种不完整匹配和多对一匹配的情况，并结合全局信息更准确地度量预测框与真实标注框之间的差异从而有效解决苹果的密集遮挡带来检测精度低的问题。

OTA 的最优传输理论方法应用在苹果的目标检测中，将真实框(Ground Truth, GT) 视为供应商，预测框视为需求者，然后在 GT 和所有预测框之间计算运输成本，并通过寻找最低运输成本的合适映射关系来进行标签分配。假定 GT 的数量为 m ，预测框的数量为 n ，将标签从 GT 运输到预测框的成本 c_{ij}^{fg} 为：

$$c_{ij}^{fg} = L_{cls}(P_j^{cls}(\theta), G_i^{cls}) + \alpha L_{reg}(P_j^{box}(\theta), G_i^{box}) \quad (3-4)$$

式 3-4 中 θ 是模型的参数， L_{cls} 表示分类损失 Focal loss， L_{reg} 为回归损失 IOU loss， G_i 为真实的标签， α 是可调节的超参数， P_j^{cls} 与 P_j^{box} 表示的是 a_j 的预测类别和检测框。

在训练过程中，除了正标签分配之外，许多锚框被视为负样本。而一个负样本的运输成本为：

$$c_{ij}^{fg} = L_{cls}(P_j^{cls}(\theta), \emptyset) \quad (3-5)$$

式 3-5 中， $\emptyset$ 表示的背景类。

通过上述方法计算每个 GT 和每个预测框的成本，以及计算每个 GT 需要的预测框数量 k ，就可以通过使用 Sinkhorn-Knopp 算法求得最小成本的传输 π^* 。具体的算法公式如 3-6 所示：

$$\begin{aligned} \pi^* &= \pi_{ij} : \\ \min_{\pi} & \sum_{i=1}^m \sum_{j=1}^n c_{ij} \pi_{ij} \\ \sum_{i=1}^m \pi_{ij} &= d_j, \sum_{j=1}^n \pi_{ij} = s_i, \sum_{i=1}^m s_i = \sum_{j=1}^n d_j \\ \pi_{ij} &\geq 0, i=1, 2, \dots, m, j=1, 2, \dots, n \end{aligned} \quad (3-6)$$

式 3-6 中 s_i 表示 GT 拥有的正标签数， d_j 表示预测框对应的标签。

最终计算得到最小传输成本 π^* 后，根据每个预测框分配最多标签的 GT 来实现标签的最优分配。

3.2.4 FasterNet

FasterNet^[49]是一种高效的轻量化特征提取网络，旨在在保持模型精度的同时显著减少计算量与参数规模。该网络结构主要基于部分卷积（Partial Convolution, PConv）与逐点卷积（Pointwise Convolution, PWConv），并采用了模块化的设计策略以提升结构灵活性和适应性。FasterNet 由四个阶段构成，每个阶段包含一个嵌入层（embedding layer）或合并层（merge layer），以及若干个重复堆叠的 FasterNet 块（FasterNet Block）。嵌入层用于对输入特征图进行空间下采样并扩展通道数，以便于后续网络提取更高层次的语义信息；而合并层则用于融合来自不同尺度的特征，以增强网络对多尺度目标的感知能力。

FasterNet 块是整个网络结构的核心组成部分，其采用了一种倒置残差（Inverted Residual）结构，结构紧凑且易于部署。具体而言，FasterNet 块由一个 PConv 层和两个 PWConv 层构成。PConv 层在保留输入图像空间结构的同时，只对部分通道进行卷积操作，这种不对称的处理方式在保证特征表达能力的前提下大幅减少了计算量；而两个 PWConv 层分别用于通道压缩与通道恢复，第一层 PWConv 旨在减少特征图的通道数，从而降低后续计算复杂度，第二层 PWConv 则负责恢复通道数，确保输出特征具有足够的表达能力。在块的尾部，通过残差连接（shortcut connection）将输入与输出特征图相加，进一步缓解梯度消失问题，并提升网络的训练稳定性。

FasterNet 块的结构及其所构建的改进后主干网络如图 3-5 所示。其整体流程为：输入特征图首先通过 PConv 层提取关键的空间信息，随后依次通过两个 PWConv 层完成通道压缩与通道还原操作，最后与输入特征进行逐元素加和，输出最终结果。这种设计不仅提升了模型在特征提取阶段的效率，还增强了对局部和全局信息的联合建模能力。

在本文的研究中，针对 YOLOv5 主干网络结构参数冗余、计算开销较大的问题，引入 FasterNet 作为其替代主干。通过用 FasterNet 替换 YOLOv5 原有的 CSPDarknet 主干网络，可以有效减少模型参数数量与计算量，达到轻量化部署的目标，同时仍然保持较高的目标检测精度。这种替换方式不仅适用于边缘设备部署的需求，还为实际场景中的实时目标检测系统提供了更加可行的解决方案。

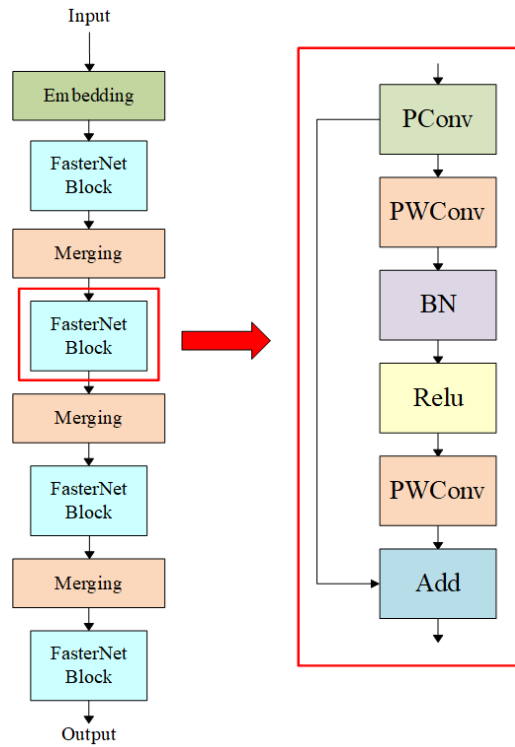


图 3-5 FasterNet 块

Figure 3-5 FasterNet block

3.3 改进 YOLOv7

针对复杂果园场景中苹果在颜色、纹理等细节特征上的检测难题，本文提出基于 YOLOv7 的检测优化方法，以进一步提升模型在细节感知方面的能力。YOLOv7 是 Ultralytics 公司于 2022 年 4 月推出的最新一代目标检测模型，在多个公开数据集上刷新了精度纪录，在检测速度与准确率之间实现了新的最优平衡（State-of-the-Art）。该模型引入了增强型扩展特征融合结构（E-ELAN）和任务自适应赋权机制（Task-Aware Learning），显著提升了在小目标检测和复杂场景下的表现，特别适用于对检测精度要求较高且硬件资源较为充足的应用环境。YOLOv7 的整体架构由输入层、主干网络和统一的头部网络组成，其中头部网络整合了传统 YOLO 中的 Neck 与 Head 功能。这种结构设计有助于提升特征提取效率和检测精度，为果园环境中高精度苹果检测提供更可靠的技术支持。

3.3.1 YOLOv7 网络结构

YOLOv7 模型的网络结构图如图 3-6 所示。

YOLOv7 引入最小点距离交并比作为新的回归损失函数，提高边界框回归的准确性；加入高效通道注意力机制，增强模型对颜色、纹理等细节特征的敏感性；并将主

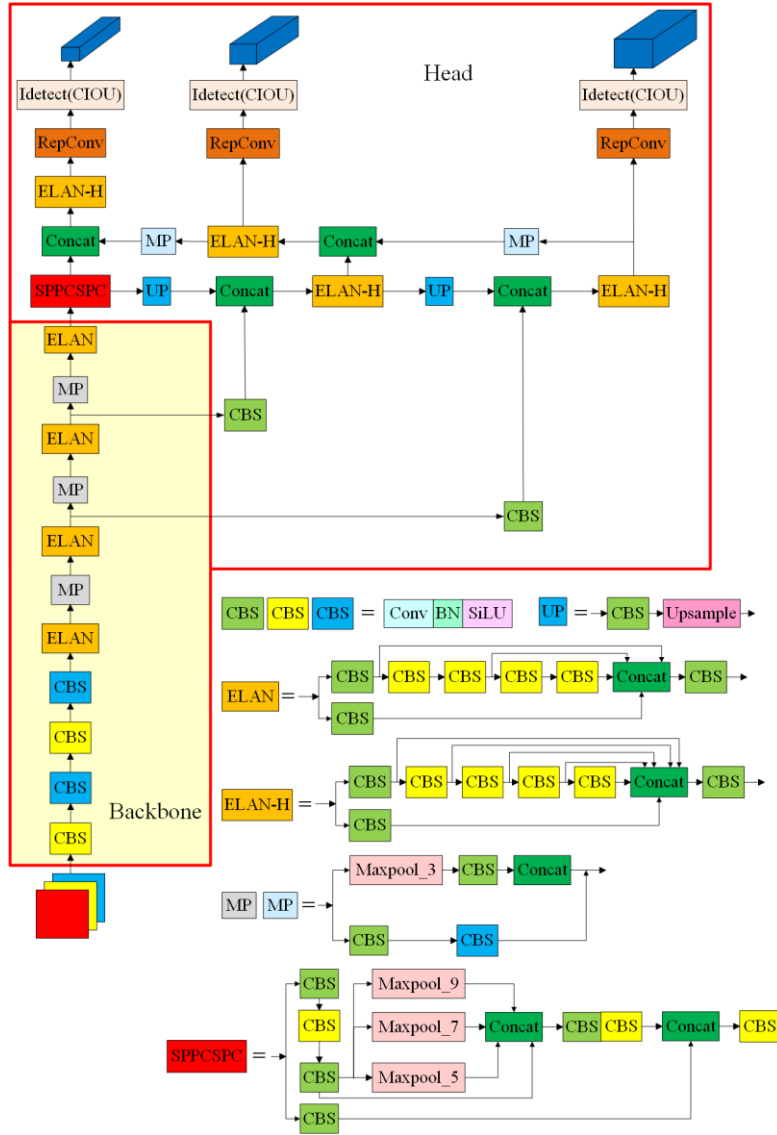


图 3-6 YOLOv7 网络结构

Figure 3-6 YOLOv7 network structure

干网络中的标准 3×3 卷积替换为深度可分离卷积，在降低计算复杂度的同时保持检测性能。如图 3-7 是模型修改后的网络结构图。

3.3.2 MPDIOU

原始的 YOLOv7 模型使用 CIOU(Complete IOU)损失函数进行边界框回归。然而，CIOU 具有一定的局限性，难以处理预测框与真实框具有相同宽高比但宽高值完全不同的情况。在这种情况下，CIOU 难以有效地捕捉预测框与真实框之间的差异，妨碍了回归损失的优化。为了解决这些问题，本文采用了一种基于最小点距离的边界框相似度比较度量法最小点距离交并比(Minimum Point Distance Intersection Over Union, MPDIOU)^[50]损失函数，以增强边界框回归。MPDIOU 直接计算预测边界框与真实边界

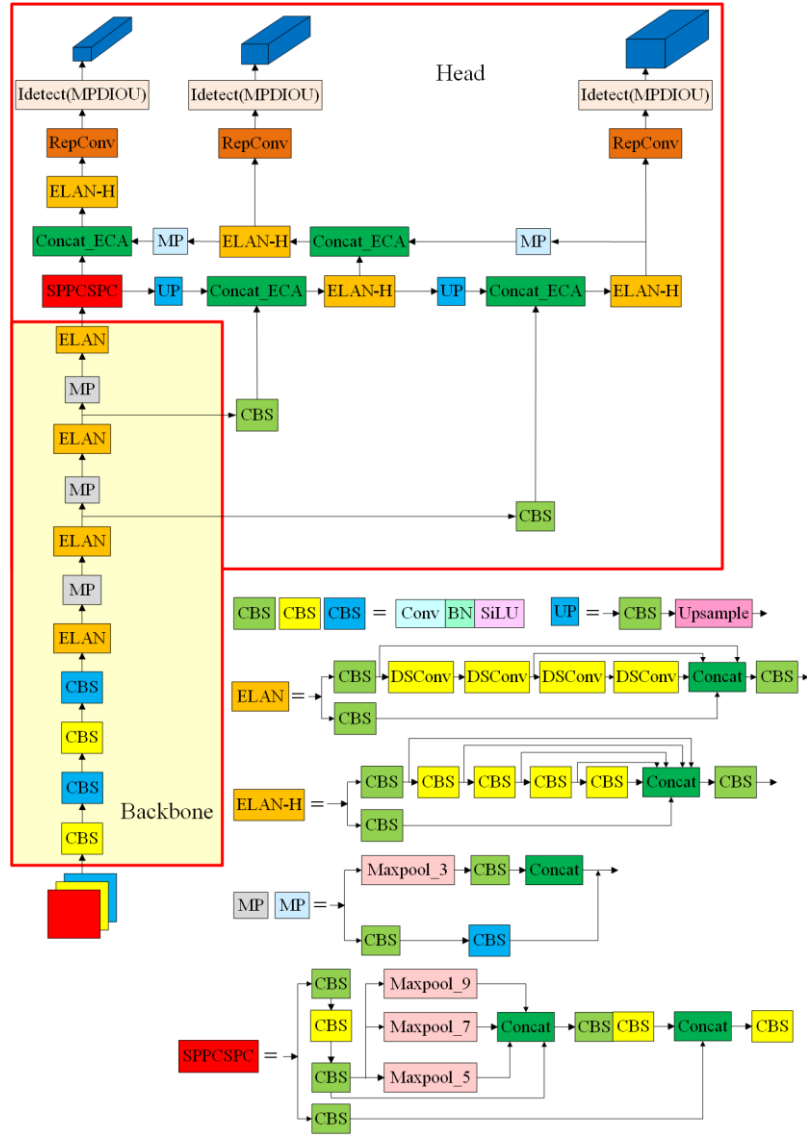


图 3-7 YOLOv7 改进后网络结构

Figure 3-7 Improved network structure of YOLOv7

框的左上角和右下角点之间的最小距离，即使尺寸相差很大也能有效地捕捉预测框与真实框之间的空间关系。

MPDIOU 的计算分为如下三个步骤，具体如图 3-8 所示：

距离计算:计算预测边界框与地面真实框的左上角和右下角点之间的距离 d ，其中 x 和 y 表示左上角点和右下角点的坐标，A 表示真实框，B 表示预测框，公式如 3-7 和 3-8 所示：

$$d_1^2 = (x_1^A - x_1^B)^2 + (y_1^A - y_1^B)^2 \quad (3-7)$$

$$d_2^2 = (x_2^A - x_2^B)^2 + (y_2^A - y_2^B)^2 \quad (3-8)$$

面积计算:计算预测框与真实框之间的重叠面积(IOU)，公式如 3-9 所示：

$$IOU = \frac{A \cap B}{A \cup B} \quad (3-9)$$

MPDIOU 计算:使用以下公式将距离和 IOU 度量组合成一个 MPDIOU 分数。其中 H 表示图像的高度, W 表示图像的宽度, 公式如 3-10 所示:

$$MPDIOU = IOU - \frac{d_1^2}{W^2 + H^2} - \frac{d_2^2}{W^2 + H^2} \quad (3-10)$$

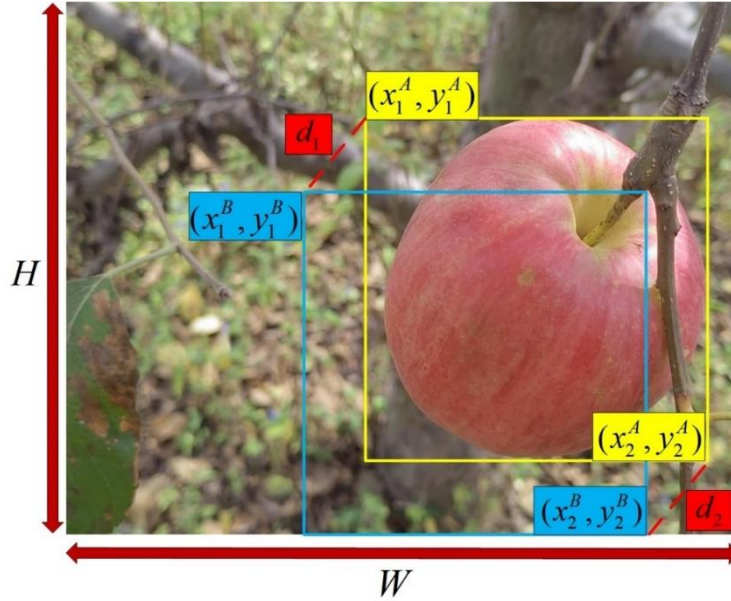


图 3-8 MPDIOU

Figure 3-8 MPDIOU

通过在果园苹果检测中引入 MPDIOU,能够增强对密集苹果的检测能力,MPDIOU 能够准确地衡量预测框与真实框之间的距离和方向差异,即使是紧密相邻的苹果也能有效区分,能够提高对不同形状苹果检测的鲁棒性。

3.3.3 ECA

高效通道注意力机制(Efficient Channel Attention, ECA)^[51]注意力机制是一种用于深度卷积神经网络的轻量级通道注意力机制。它的核心思想是在通道维度上进行注意力计算,以强化有用特征并减弱无用特征。与其他注意力机制相比,ECA 注意力机制具有计算效率高、模型轻量化、有效性好等优点。这是因为它只需要对每个通道进行一个全局平均池化,然后经过一个简单的线性变换即可得到注意力权重,并且本身的参数量很少。ECA 注意力机制如图 3-9 所示。

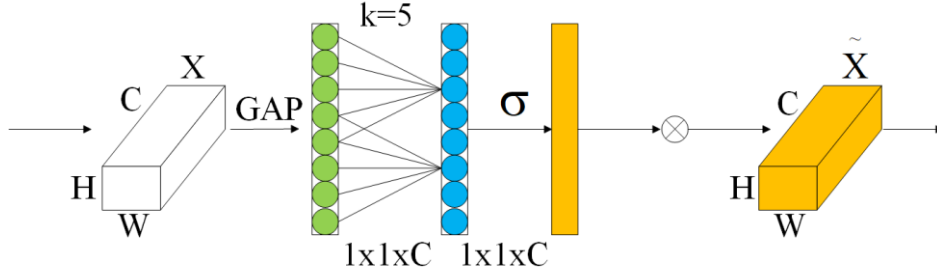


图 3-9 ECA

Figure 3-9 ECA

ECA 注意力机制由全局平均池化、一维卷积和 Sigmoid 函数组成。在不同尺度的特征图进入融合网络之前，输入为大小为 $H \times W \times C$ 的图像。通过全局平均池化层，将数据转换为尺度为 $1 \times 1 \times C$ 的特征数据。然后，使用一维卷积核为 k 的卷积操作，在不降低维度的情况下实现局部跨通道交互，以减小模型参数。最后，通过应用 Sigmoid 函数使网络更加非线性化。

本文算法中通过在 YOLOv7 的第 55, 67, 80, 93 层的 Concat 中嵌入 ECA 注意力机制，如图 3-10 所示，以增强模型对苹果特征的提取和融合能力。这可以使模型在不同层次的特征融合过程中更加关注有用的通道信息，从而提高对苹果目标的检测性能。

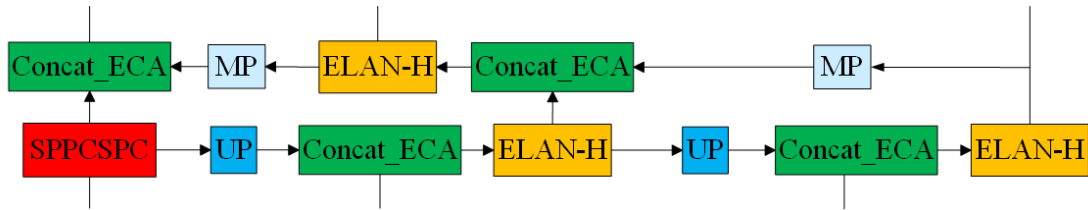


图 3-10 嵌入 ECA 注意力机制

Figure 3-10 Embed ECA attention mechanism

3.3.4 DSConv

深度可分离卷积(Depthwise Separable Convolution, DSConv)^[52]，是一种轻量级的卷积操作，通过分解标准卷积操作，可以减少参数量和计算量，同时又能保持与标准卷积相当的性能。在本文算法中将 DSConv 模块引入到 YOLOv7 的主干网络中，替换主干网络中原有的 3×3 标准卷积，以达到降低模型的计算量的目的，具体如图 3-11 所示。

标准卷积使用一个三维滤波器对输入特征图进行卷积操作。该滤波器由多个通道组成，每个通道对该滤波器包含多个通道，每个通道对应一个输出特征图。标准卷积

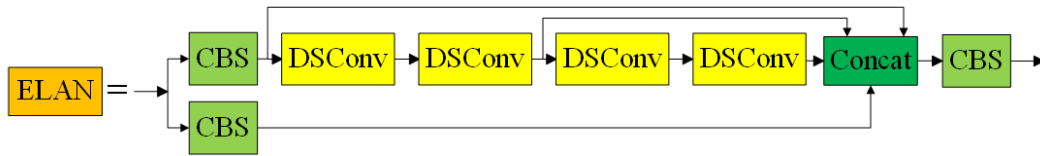


图 3-11 Backbone 嵌入 DSConv

Figure 3-11 Backbone is embedded in DSConv

的计算量和参数量与滤波器的尺寸和通道数成正比，因此标准卷积的计算成本较高。

DSConv 将标准卷积操作分解为逐深度卷积(Depthwise Convolution, DWC)和逐点卷积(Pointwise Convolution, PWC)两个步骤。

DWC 使用一个逐深度滤波器对每个输入通道单独进行卷积。由于该滤波器仅包含一个通道，因此只提取空间特征，而不会改变通道数，输入特征图通道数等于卷积核个数等于输出特征图个数。逐深度卷积的计算量远低于标准卷积，这是因为它只使用了单个通道的滤波器。

PWC 使用一个逐点滤波器对每个输出特征图单独进行卷积，它实质上是用 1×1 的卷积核进行升维。逐点卷积是通过包含多个通道的滤波器来实现的，它可以融合来自不同输入通道的信息，并生成最终的输出特征图。与标准卷积相比，逐点卷积的计算量仍然较少，因为它与输入特征图的通道数成正比

通过引入深度可分离卷积模块，可以在 YOLOv7 的主干网络中减少计算量，而不会明显损失性能。这可以提高模型的效率和推理速度，同时保持对苹果特征的准确提取和融合能力。

3.4 改进 YOLOv11

针对果园环境中苹果在多尺度分布和空间位置变化条件下检测精度下降的问题，本文提出基于 YOLOv11 的融合增强方法，以提升模型在复杂环境下的检测鲁棒性与精度。作为 YOLO 系列的最新版本，YOLOv11 在检测架构与训练机制方面进行了全面升级，融合了 Transformer 结构、改进的注意力机制以及动态卷积等先进技术，显著提升了模型在目标检测中的精度、泛化能力和复杂场景适应性。

YOLOv11 在高分辨率图像处理和超大规模数据集训练方面进行了特别优化，适用于多类别、小目标密集、背景复杂等高要求场景，尤其适合果园等复杂自然环境中的目标检测任务。此外，YOLOv11 还引入了更加精炼的网络架构与优化的训练流程，在提升检测精度的同时实现了更快的处理速度，在精度与性能之间达到了新的平衡。

3.4.1 YOLOv11 网络结构

如图 3-12 所示, YOLOv11 的网络结构主要包括骨干网络、特征融合网络和检测头部网络等几个部分。

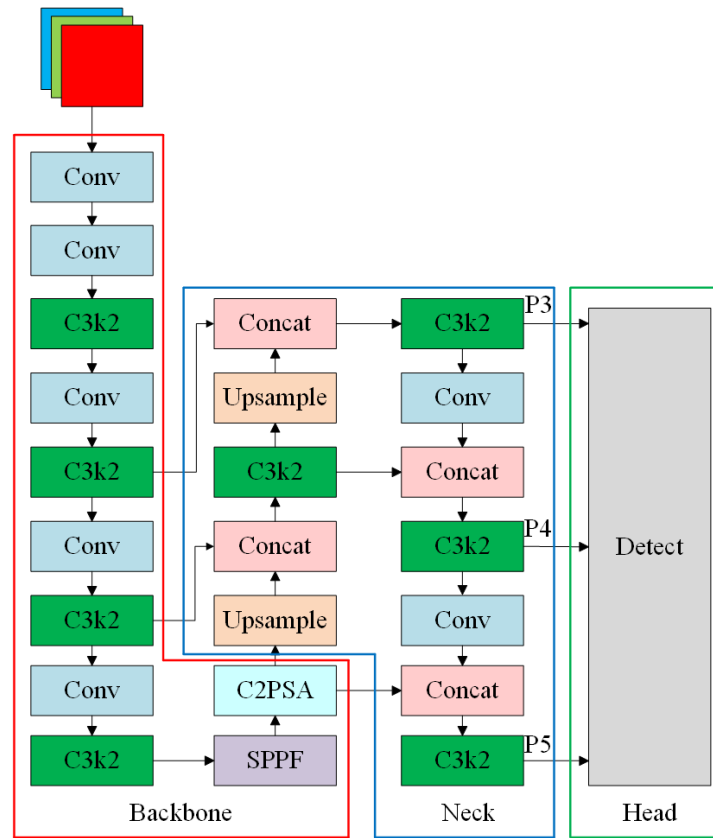


图 3-12 YOLOv11 网络结构

Figure 3-12 YOLOv11 network structure

Backbone 是目标检测模型的核心, 主要负责提取图像的深层特征。骨干网络除以往系列的卷积模块、快速空间金字塔池化(Spatial Pyramid Pooling Fast, SPPF)还包含 C3k2 模块和 C2PSA(Pointwise Spatial Attention, PSA)等模块构成。C3K2 模块源自 YOLOv8 的 C2F 模块, 在提取更丰富特征的同时降低了参数量和计算量。当参数 C3k 设置为 FALSE 时, C3K2 模块为 C2F 模块, 此时其 Bottleneck 为普通 Bottleneck; 反之, 当 C3k 为 TRUE 时, Bottleneck 模块则替换为 C3 模块。SPPF 是用来增强网络提取不同尺度目标的特征。C2PSA 是对 C2F 模块的扩展, 结合了点式空间注意力块, 以增强特征提取和注意力机制, 提高模型对重要特征的捕捉能力。

Neck 部分连接 Backbone 和 Head, 负责对 Backbone 提取的特征进行融合和增强, 为检测头提供多尺度特征。由特征金字塔网络和路径聚合网络组成的特征融合网络, 可以深层次融合高层语义特征与低层细节特征, 生成更具表达能力的特征图。

Head 是模型的输出部分，负责生成最终的检测结果，它包含多个检测头，每个检测头负责预测不同尺度目标的边界框、类别和置信度。

YOLOv11 首先在颈部网络中引入空间增强注意力模块，提升模型对目标空间分布的关注能力；其次使用全维度动态卷积替代原有卷积结构，使模型能动态适应不同特征尺度，提高复杂场景下的表现力；最后加入 C3k2_ScConv 模块，结合 C3k 模块与空间和通道重建卷积进行结构重构，进一步增强模型的空间与通道特征融合能力。如图 3-13 所示为改进后 YOLOv11 网络结构图。

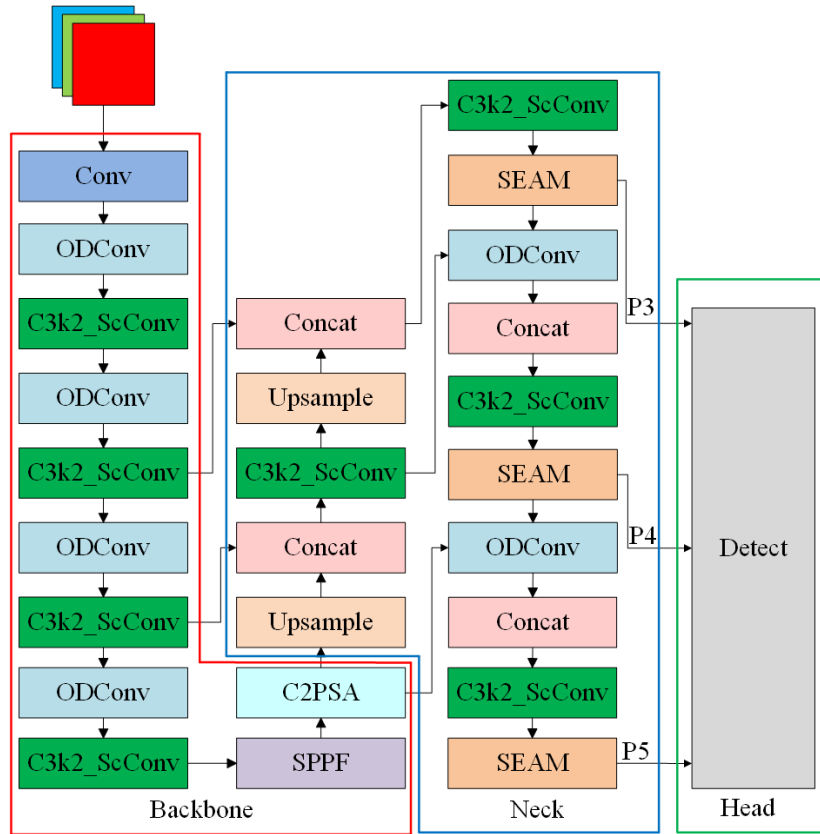


图 3-13 改进的 YOLOv11 神经网络设计

Figure 3-13 Improved YOLOv11 neural network design

3.4.2 SEAM

面对复杂的果园环境，原模型在遮挡、光照变化及背景干扰等场景下容易出现错检和漏检现象，导致检测精度下降。主要原因在于，真实果园场景中苹果常常被枝叶、苹果重叠等因素遮挡，使模型难以准确提取目标特征。因此引入空间增强注意力模块 (Separated and Enhancement Attention Module, SEAM)^[53]来提升模型对遮挡目标的检测精度。SEAM 的核心作用是增强未遮挡区域的特征表达，以弥补遮挡区域的特征损失。通过学习遮挡区域与未遮挡区域之间的关联性，动态调整特征图中的关注点，使模型

能够有效捕捉遮挡区域的关键信息，从而提升模型对果园苹果的检测能力。SEAM 的具体工作流程为：

(1) 特征提取：输入苹果图像或特征图经过一系列卷积操作，提取出包含丰富信息的特征表示。

(2) 生成空间注意力图：SEAM 模块对提取到的特征进行处理，生成空间注意力图。注意力图中的每个像素值代表对应位置特征的重要性,值越大表示该位置更可能包含目标信息。

(3) 将空间注意力图作为软掩码：通过与原始特征图逐像素点乘的方式进行加权，抑制背景噪声并增强苹果区域的特征表示。

(4) 输出：经过加权后的特征图被送入后续网络层，进一步完成检测任务。

图 3-14 展示了 SEAM 的架构以及 CSMM(Channel and Spatial Mixing Module, CSMM)的结构。

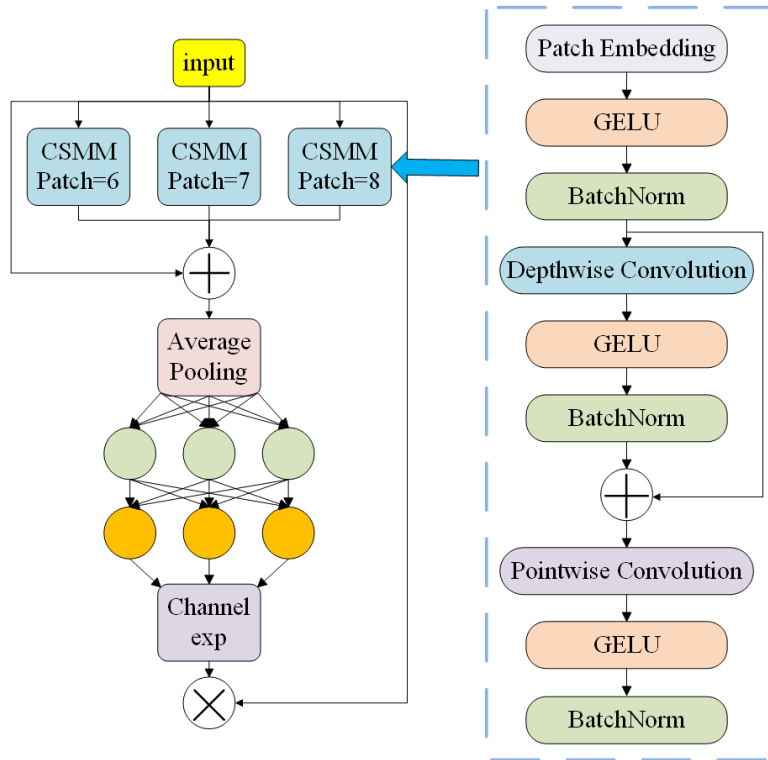


图 3-14 SEAM 架构与 CSMM 结构

Figure 3-14 SEAM architecture and CSMM structure

SEAM 的整体架构在左侧展示，包含三种不同尺寸的 CSMM 模块。各模块通过对不同尺度特征的提取，能够在局部区域和全局区域之间建立关联，增强特征表示多样性。多尺度特征在果园环境中尤为关键，能够有效弥补苹果被枝叶遮挡时的特征损失。

CSMM 模块的详细结构在右侧展示。该模块通过空间和通道信息的交互建模，进一步优化特征表达，增强对关键区域的关注。具体结构包括：

- (1) Patch Embedding: 对输入图像划分的 Patch 进行嵌入，提取局部特征。
- (2) GELU(Gaussian Error Linear Unit): 非线性激活函数，用于增强特征表达能力。
- (3) BatchNorm: 批量归一化，稳定训练过程并加速收敛。
- (4) Depthwise Convolution: 逐通道卷积，减少参数量的同时保留空间特征。
- (5) Pointwise Convolution: 逐点卷积，通过 1×1 卷积核对通道特征进行融合，增强特征图的全局表示。

3.4.3 ODConv

全维度动态卷积(Omni-Dimensional Dynamic Convolution, ODConv)^[54]能够增强模型的特征提取能力。果园场景中苹果的尺度、姿态和遮挡情况各异，传统卷积操作难以同时兼顾这些多样化特征，导致检测性能下降。通过引入 ODConv，模型在空间维度、输入通道和输出通道等多个维度上具有动态性，从而使得卷积核能够自适应地捕捉数据中的关键特征。根据动态卷积的定义，ODConv 可以描述成如下形式：

$$Y = (a_{w1} \odot a_{f1} \odot a_{c1} \odot a_{s1} \odot W_1 + \dots + a_{wn} \odot a_{fn} \odot a_{cn} \odot a_{sn} \odot W_n) * X \quad (3-11)$$

式 3-11 中，卷积核 W_i 的注意力标量由 a_{wi} 表示，空间、输入通道以及输出通道三个维度分别对应 $a_{si} \in R^{k \times k}$ ， $a_{ci} \in R^{C_{in}}$ ， $a_{fi} \in R^{C_{out}}$ 三个注意力值。它们的值通过多头注意力模块计算得出。

在苹果检测任务中，利用 ODConv 通过自适应学习多尺度特征并动态调整卷积核权重的特点，有效提升对遮挡苹果、密集苹果和小目标的检测能力。通过多维度动态加权，模型可以更加精准地提取苹果区域的特征，抑制背景干扰，最终提高检测精度。

ODConv 通过对卷积核的不同维度赋予不同的注意力值来实现自适应性，具体包括：空间维度(图 3-15a)、输入通道维度(图 3-15b)、输出通道维度(图 3-15c)以及整个卷积核维度(图 3-15d)。

3.4.4 C3k2_ScConv

在果园环境中，苹果常因枝叶遮挡而导致特征提取困难，进而影响模型检测精度。为增强模型对遮挡苹果的检测能力，本文采用空间和通道重建卷积(Spatial and Channel reconstruction Convolution, ScConv)模块^[55]替代 C3k2 中的 C3k 和 Bottleneck 结构，并

命名为 C3k2_ScConv 模块(C3k2_ScConv 模块替换结构如图 3-16 所示), ScConv 模块通过对空间特征和通道特征的联合重构,有效增强了模型对关键特征的提取能力,减少遮挡背景的干扰。ScConv 模块由空间重校准单元(Spatial Recalibration Unit, SRU)和通道重校准单元(Channel Recalibration Unit, CRU)两部分组成,其结构如图 3-17 所示。

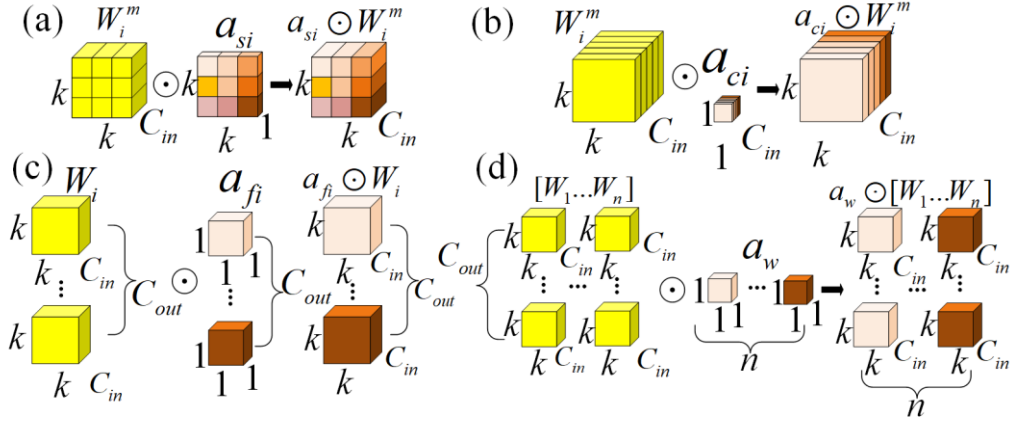


图 3-15 ODConv 四维注意力机制对卷积核加权示意图

Figure 3-15 Schematic diagram of ODConv four-dimensional attention mechanism weighting to convolutional kernel

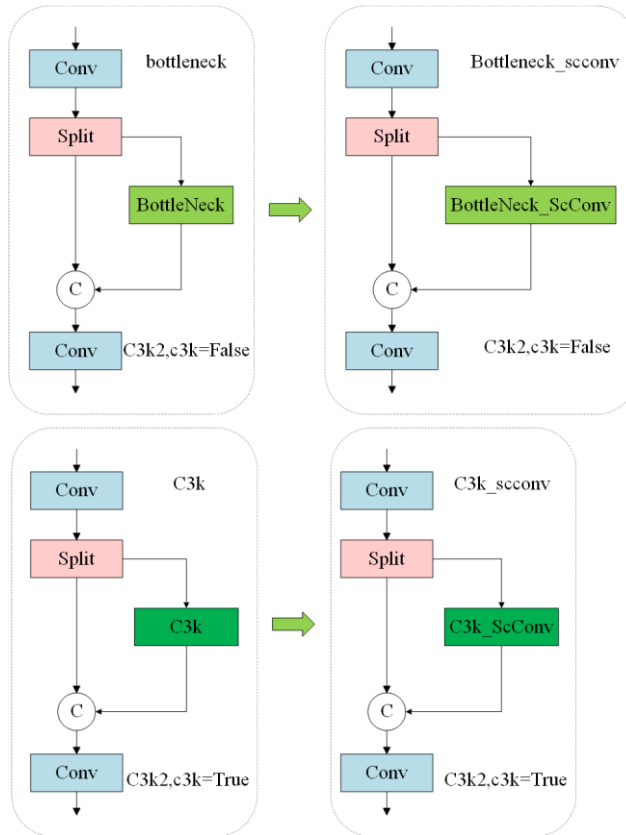


图 3-16 C3k2_ScConv 模块替换图

Figure 3-16 C3k2_ScConv Module Replacement Diagram

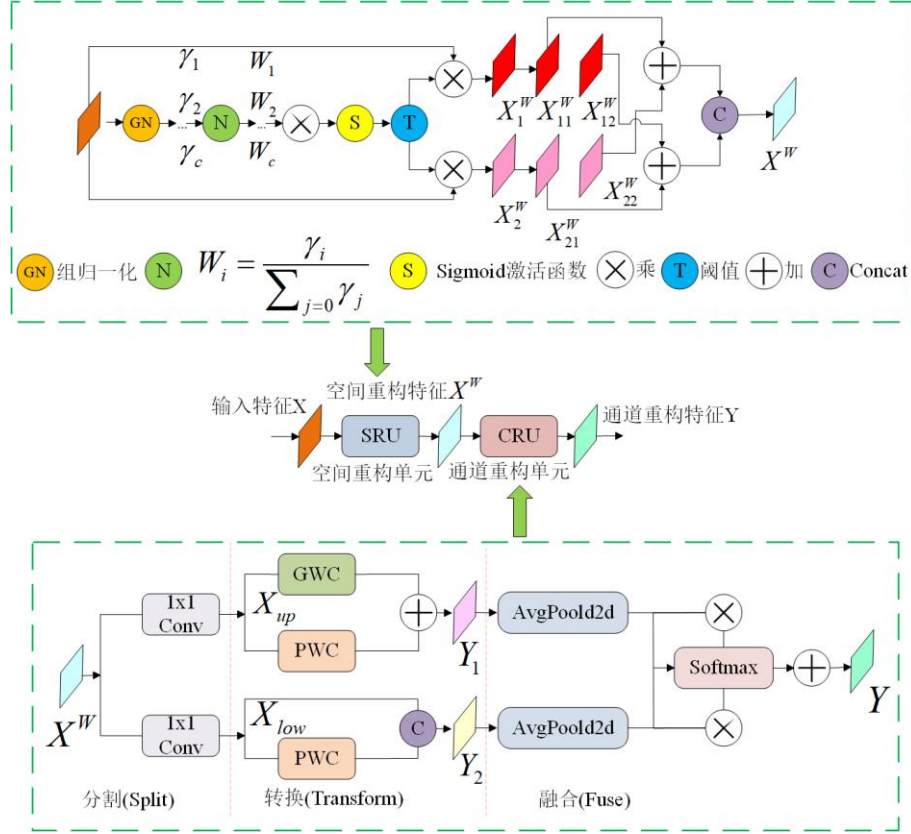


图 3-17 ScConv 模块结构图

Figure 3-17 Structure diagram of the ScConv module

特征分离操作通过组归一化缩放因子评估特征重要性，具体计算如下：

$$X_{out} = GN(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (3-12)$$

$$W_\gamma = \frac{\gamma_i}{\sum_{j=1}^C \gamma_j}, i, j = 1, 2, \dots, C \quad (3-13)$$

$$W = Gate(Sigmoid(W_\gamma(GN(X)))) \quad (3-14)$$

式 3-12、3-13 和 3-14 中，参数 μ ， σ 分别表示均值和标准差， γ ， β 为可学习参数， ε 为一极小常量，用于保证模型的稳定性。参数 γ 的大小直接反映了像素间变化程度，即空间信息的丰富程度。

重构操作将筛选出的特征图加权融合，生成信息更丰富的特征表示。通过交叉重构，模型能够有效整合关键信息，同时抑制背景干扰。具体融合方式如式 3-15 所示，其中 $\otimes$ 是元素乘法， $\oplus$ 是元素加法， $\cup$ 是求并集。

$$\begin{cases} X_1^W = W_1 \otimes X, \\ X_2^W = W_2 \otimes X, \\ X_{11}^W \oplus X_{22}^W = X^{W1}, \\ X_{21}^W \oplus X_{12}^W = X^{W2}, \\ X^{W1} \cup X^{W2} = X^W. \end{cases} \quad (3-15)$$

尽管 SRU 能有效去除空间冗余,但无法消除通道维度上的冗余。CRU 通过分割、转换和融合三个操作进一步优化通道特征。输入特征图 X^W 被划分为两部分,分别经过独立的卷积处理,得到 X_{up} 和 X_{low} 。

转换操作包括分组卷积(Group-Wise Convolution, GWC)和 PWC,分别提取局部特征和全局特征。融合操作通过全局平均池化和 Softmax 归一化生成权重向量 β_1 和 β_2 ,最终得到通道加权的提炼特征 $Y = \beta_1 Y_1 + \beta_2 Y_2$ 。

改进后的 C3k2_ScConv 模块通过 ScConv 对空间特征和通道特征的联合重构,增强模型的特征提取能力,能够更精准地聚焦于目标区域,抑制背景干扰,提高苹果检测任务的准确率和鲁棒性。

3.5 模型训练与实验

3.5.1 实验环境

实验主机的处理器为 Intel® Core™ i7-5820K CPU @ 3.30GHz(12 核),运行内存为 32GB,显卡为 GeForce GTX TITAN X,显存大小为 12GB,操作系统为 Ubuntu 20.04.6 LTS。深度学习框架使用的是 PyTorch 1.12.0(Python 3.8.13),CUDA 版本为 12.2。

3.5.2 评价指标

本文采用多个指标对模型进行客观量化的评估,这些指标包括精确率(Precision, P)、召回率(Recall, R)、均值平均精度(mean Average Precision, mAP)、计算量以及检测帧率。均值平均精度使用 IOU(Intersection Over Union, IOU)阈值为 0.5 进行计算,即用 mAP@0.5 来衡量目标模型的准确性。以下是各类指标的详细介绍:

(1) 精确率。精确率是一个常用的分类模型评价指标,表示正确预测为正类的样本占有所有预测为正类样本的比例。式 3-16 中, TP(True Positives, TP)是模型预测正确的正例数, FP(False Positives, FP)则是模型将负例误判为正例的次数。

$$P = \frac{TP}{TP + FP} \quad (3-16)$$

(2) 召回率(查全率)。它是衡量模型找到所有正样本的能力,表示模型能找到多少

真正的正样本，反映了模型对所有希望被预测的样本的覆盖能力。式 3-17 中，FN(False Negatives, FN)表示模型错判为负类的正样本数。

$$R = \frac{TP}{TP + FN} \quad (3-17)$$

(3) 均值平均精度是目标检测任务中比较常用的性能评价指标。mAP 通过衡量模型在各类别上的检测精度来反映整体检测性能。mAP 是所有类别 AP(Average Precision, AP)值的平均，能够比较全面地反映模型的整体检测性能。mAP 的取值范围在 0 到 1 之间，数值越高表示模型的检测性能越优，式 3-18 中 d 表示种类数， AP_i 表示种类 i 的平均检测精度。

$$mAP = \frac{1}{d} \sum_{i=1}^d AP_i \quad (3-18)$$

(4) 计算量。计算量即浮点运算的次数，用来衡量算法的时间复杂度。通常用 GFLOPs(Giga Floating Point Operations per Second, GFLOPs)作为单位进行衡量。模型的计算量与其参数量和数据量是息息相关的。计算量的多少可以反映模型的计算复杂度、计算资源需求以及模型的训练和推理速度。因此在实际的应用中需要综合考虑计算量和计算资源之间的平衡，以满足任务需求。

(5) 检测帧数。FPS(Frames Per Second, FPS)是指每秒钟图像帧的数量，反映了视频图像的流畅度，也是衡量计算机视觉任务中模型推理计算速度的重要指标。FPS 越高，表示模型推理速度越快，视频图像越流畅。计算公式如 3-19 所示，其中 t 表示模型对单张图片预处理、推理以及后处理所需的时间的总和。

$$FPS = \frac{1}{t} \quad (3-19)$$

3.5.3 对比实验结果与分析

为了验证本文所提出的改进型 YOLOv5、YOLOv7 和改进型 YOLOv11 算法具有更好的效果，使用控制变量法，在保证实验训练集、验证集、测试集完全相同的情况下，使用 YOLOv3、Faster-RCNN、YOLOv6、YOLOv8、RTDETR^[56]、YOLOv9、YOLOv10 目前较为流行的多种目标检测模型进行对比实验。本研究以 mAP、精确率和召回率等作为主要评价指标。实验结果如表 3-1 所示。

表 3-1 模型对比实验表

Table 3-1 Model comparison experiment table

模型	P	R	mAP@0.5	GFLOPs	FPS
YOLOv3 ^[11]	0.792	0.683	0.762	155.3	32
YOLOv5L	0.826	0.698	0.795	108.3	34
改进型 YOLOv5L	0.831	0.724	0.817	71.3	36
Faster-RCNN ^[8]	0.833	0.711	0.820	323.5	21
YOLOv6L ^[12]	0.779	0.678	0.744	391.9	19
YOLOv7 ^[13]	0.819	0.727	0.803	105.1	37
改进型 YOLOv7	0.855	0.738	0.831	59.2	49
YOLOv8L ^[14]	0.836	0.675	0.797	165.4	31
RTDETR-L ^[56]	0.816	0.692	0.778	-	20
YOLOv9 ^[15]	0.806	0.703	0.774	84.1	41
YOLOv10L ^[16]	0.827	0.673	0.782	120	32
YOLOv11L ^[17]	0.829	0.714	0.810	87.3	31
改进型 YOLOv11L	0.864	0.736	0.848	58.3	38

根据表 3-1 的实验结果可以得出结论，本文提出的改进型 YOLOv11 算法能够高效地实现果园苹果的检测与检测。该算法在各类目标的检测中取得了 84.8% 的 mAP，优于 YOLOv3、Faster-RCNN、YOLOv6、YOLOv8、RTDETR、YOLOv9、YOLOv10 等模型 8.6%，2.8%，10.4%，5.1%，7.0%，7.4%，6.6%，同时也超过了原始 YOLOv5、YOLOv7 以及其改进版本 5.3%，3.1%，4.5%，1.7%。此外，YOLOv11 在精确度、召回率和计算量等方面均表现优异。尽管其推理帧率(FPS)比改进型 YOLOv7 低 11 和比 YOLOv9 低 3，但仍能够满足实时检测的需求。而且，较低的计算量也会降低对边缘设备的硬件要求，同时在检测效果上保持良好的表现。通过实验数据对比表明，相比较其他的检测模型以及在改进模型中，改进的 YOLOv11 表现最好，综合性能最优，能够满足实际果园场景下苹果检测的需求。

3.5.4 消融实验结果与分析

本文通过设置三组消融对比实验来验证本文提出的三个改进模型的各个改进模块的优化作用。

改进型 YOLOv5 实验结果如表 3-2 所示, 以 YOLOv5L 为基础模型, 先使用 Soft_NMS 来替换模型原有的 NMS, 改变原模型对预测框的处理方式, mAP 提升 2.1%。然后再使用 OTA 来优化原模型的标签分配方式, 将标签分配看作最优传输问题, mAP 再提升 0.5%。最后使用 FasterNet 块替换主干网络中的卷积模块作为新的特征提取网络, mAP 降低 0.4%, GFLOPs 减少 34.2%, FPS 增加 2。

表 3-2 改进型 YOLOv5 模型消融实验对照表

Table 3-2 Comparison table of ablation experiments of the improved YOLOv5 model

模型	P	R	mAP@0.5	GFLOPs	FPS
YOLOv5	0.826	0.698	0.795	108.3	34
YOLOv5+Soft_NMS	0.812	0.745	0.816	108.3	30
YOLOv5+Soft_NMS+OTA	0.825	0.752	0.821	108.3	30
YOLOv5+Soft_NMS+OTA+FasterNet	0.831	0.724	0.817	71.3	36

改进型 YOLOv7 实验结果如表 3-3 所示, 以 YOLOv7 为基础模型, 先引入 MPDIU 损失函数来替换模型原有的 IOU 损失函数, mAP 提升 1.2%。再增加 ECA 注意力机制, mAP 再提升 2.5%。最后使用 DSConv 替换主干网络中 3*3 的卷积, 在保持较高检测精度的同时有效降低模型的计算复杂度和保证检测速度, mAP 降低 0.9%, GFLOPs 减少 43.7%, FPS 增加 2。

表 3-3 改进型 YOLOv7 模型消融实验对照表

Table 3-3 Comparison table of ablation experiments of the improved YOLOv7 model

模型	P	R	mAP@0.5	GFLOPs	FPS
YOLOv7	0.819	0.727	0.803	105.1	37
YOLOv7+MPDiou	0.840	0.728	0.815	105.1	36
YOLOv7+MPDiou+ECA	0.836	0.774	0.840	105.1	38
YOLOv7+MPDiou+ECA+DSConv	0.855	0.738	0.831	59.2	49

改进型 YOLOv11 实验结果如表 3-2 所示, 以 YOLOv11L 为基础模型, 先在颈部网络中引入了 SEAM 注意力机制, 以增强模型的空间注意力能力 mAP 提升 1.1%。再用 ODCConv 替换网络中的卷积模块, 使模型能够自适应学习不同尺度和特征的卷积核, mAP 再提升 2.5%。最后使用 C3k2_ScConv 模块, 增强模型的特征提取能力, 能够更

精准地聚焦于目标区域，mAP 提升 1.5%，GFLOPs 减少 33.2%，FPS 增加 10。

表 3-4 改进型 YOLOv11 模型消融实验对照表

Table 3-4 Comparison table of ablation experiments of the improved YOLOv11 model

模型	P	R	mAP@0.5	GFLOPs	FPS
YOLOv11	0.829	0.714	0.810	87.3	31
YOLOv11+SEAM	0.848	0.698	0.821	89.3	29
YOLOv11+SEAM+ODConv	0.859	0.703	0.833	64.9	36
YOLOv11+SEAM+ODConv+C3k2_ScConv	0.864	0.736	0.848	58.3	41

3.5.5 检测结果分析

在果园苹果检测任务中，需要综合考虑检测精度、计算量以及模型复杂度等因素，以确保检测算法在实际应用中的可行性与稳定性。针对果园中苹果存在遮挡严重、尺度变化显著等复杂环境条件，本文从实际问题出发，依次提出了三种改进策略：首先，为提升模型在基础场景下的检测效果，提出一种基于 YOLOv5 的改进型检测方法，有效缓解了遮挡和尺度变化带来的识别困难；其次，针对复杂果园环境中苹果在颜色、纹理等细节特征识别上的难点，进一步提出基于 YOLOv7 的优化策略，以增强模型对细节信息的感知能力；最后，为解决苹果在多尺度分布和空间位置变化条件下检测鲁棒性不足的问题，设计了基于 YOLOv11 的融合增强方法，以提升模型在复杂环境中的稳定性与检测精度。

通过消融实验可知，三个改进方法在精度与帧率方面均有提升，且计算量有所下降，验证了各改进方案的有效性。在进一步的对比实验中，改进型 YOLOv11 在多个主流检测模型中表现最优，在准确率、召回率及计算量方面均领先，表现出其在满足实际需求方面的综合优势。考虑到后续三维定位等环节对检测精度的更高要求，改进型 YOLOv11 更适合作为果园自动采摘系统中的检测核心模型。

为进一步验证改进型 YOLOv11 在实际果园环境中的应用效果，图 3-18 展示了该模型的热力图输出结果。热力图通过颜色变化（红色表示高响应区域，蓝色表示低响应区域）直观反映了模型对目标区域的关注能力。在无遮挡情况下，改进前后的模型均能准确识别苹果位置；而在遮挡场景中，改进前的模型存在明显的漏检和误检，部分被遮挡的苹果未能准确识别为高响应区域。相较之下，改进后的 YOLOv11 模型在遮挡区域仍能准确识别所有苹果目标，表现出更强的特征提取能力与目标判别性能。

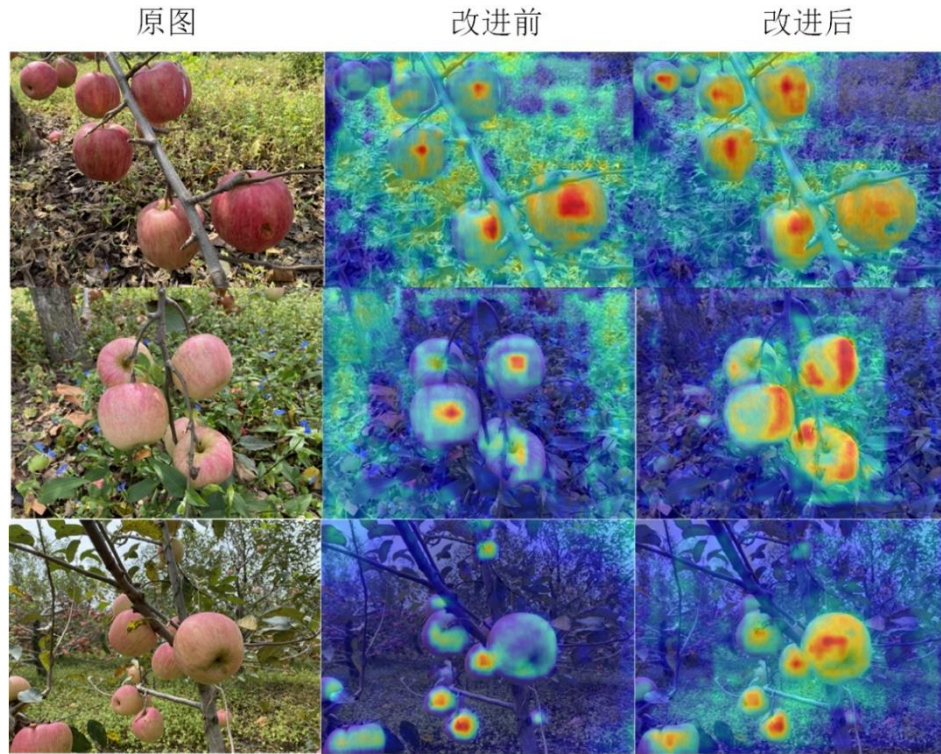


图 3-18 热力图

Figure 3-18 Test results of experimental scenarios

3.6 本章小结

本章针对果园苹果检测任务的特点，从数据集构建、模型改进和实验验证三个方面进行研究。首先结合果园场景的复杂性，构建了苹果检测数据集。随后针对目标检测算法，分别对 YOLOv5、YOLOv7 和 YOLOv11 三个一阶段检测模型进行了改进。

对于 YOLOv5 模型，添加了 Soft_NMS 算法以优化后处理阶段，采用 OTA 分配策略改进目标匹配过程，并引入 FasterNet 模块以提升特征提取能力。在 YOLOv7 的改进中，集成了 MPDIU 损失函数、ECA 注意力机制和 DSCConv 模块，提升了模型对小目标的检测能力与特征表达能力。在 YOLOv11 的研究中，优化了网络结构，结合 SEAM、ODConv 和 C3k2_ScConv 等模块，强化模型的表达能力和定位精度。

通过实验部分，本文搭建了完善的实验环境，采用标准评价指标对改进模型的性能进行了全面评估，并通过消融实验验证了各个改进策略的具体效果。对比实验进一步表明，本章提出的改进方法的有效性。

通过本章的研究工作，成功构建了适用于果园场景的苹果检测数据集，并提出了针对一阶段目标检测算法的多种改进策略。这些为果园苹果检测检测提供了技术支持，同时也为后续研究和应用奠定了坚实的基础。

第4章 苹果三维定位

在第3章中,本文对YOLOv5、YOLOv7和YOLOv11模型进行了深入研究,并对其进行了优化和改进,使得检测模型能够实现对苹果的准确、实时检测与定位。然而,此阶段的模型仅能在二维平面上完成苹果的定位,获取的仅是苹果的二维坐标信息。若要获取苹果的三维坐标信息,则需要继续借助相机进行测距。单目相机结构简单、成本较低,适合快速部署,但由于缺乏真实深度信息,其测距依赖苹果尺寸等先验知识,易受苹果大小、姿态变化和环境光照等因素干扰,定位精度有限。深度相机则可直接获取目标点云信息,并结合深度神经网络进行三维重建,适用于结构化或半结构化场景中对精度要求极高的任务,但其对环境光线敏感、成本较高,且在强光或户外复杂环境中稳定性较差。相比之下,双目相机在成本、精度和环境适应性之间取得了良好平衡。通过提取左右图像间的视差信息,双目系统能够实现无接触、非侵入式的高精度距离估计,定位误差可稳定控制在厘米级,足以满足果园采摘作业的精度需求。同时,双目相机结构灵活,适配性强,可在复杂光照和非结构化场景下稳定运行,适合部署在移动平台或机械臂等自动采摘设备上。更重要的是,双目视觉系统在开源算法、软硬件协同开发等方面具备良好生态,便于系统集成与后期优化。因此,本章将进一步研究如何利用双目相机获取苹果的三维信息,为实际应用场景中的精确定位及后续操作提供更全面的数据支持。

4.1 深度相机

深度相机是一种能够捕捉物体或场景的三维深度信息的传感器设备,与传统二维相机不同,它除了获取物体的二维图像信息外,还可以记录物体与相机之间的距离信息。深度相机广泛应用于机器人、无人驾驶以及手势检测等领域。以下是深度相机的几种类型,包含每种类型优点、缺点及代表设备。

结构光的原理是通过向目标物体投射特定的光图案(如条纹或点阵),并根据图案的变形来计算物体的三维深度信息。其主要优点是适合短距离和高精度的场景,但缺点在于对环境光较为敏感,因此在室外条件下表现不佳。代表设备包括 Kinect V1 和 Intel RealSense D400 系列。

飞行时间技术通过测量发射光信号到达物体并返回的时间差来计算深度信息。这

种方法的优点是实时性强，适用于大范围场景，并且具有较强的抗环境光干扰能力。然而，其精度相对较低，分辨率也存在一定的限制。代表设备包括 Kinect V2 和 iPhone 的 LiDAR 扫描仪。

双目视觉使用两个摄像头模拟人眼，通过视差原理计算深度信息。这种方法的优点是不需要主动光源，适合室外使用；但其缺点是算法复杂度较高，并且对纹理较少的表面表现不佳。代表设备有 ZED Camera 和部分 Intel RealSense D400 系列产品。

主动立体视觉是结构光与双目视觉的结合，通过利用主动投影的光源增强双目深度测量的可靠性。这种方法兼顾了室内外环境适应性和较高精度的需求。代表设备包括 Intel RealSense L500 系列。

因此，综合考虑果园室外环境的复杂性、目标检测对精确性的高要求，以及硬件设备的价格等多方面因素，本文最终选择了 Intel RealSense D415 相机作为三维定位的核心设备。这款相机能够在多变的室外光照条件下稳定工作，成本相对较低，在开源算法、软硬件协同开发等方面具备良好生态，便于系统集成与后期优化，精度也可满足果园采摘作业的需求。

4.2 坐标原理

获取图像的深度信息需经过四个坐标系的转换，依次为像素坐标系、图像坐标系、相机坐标系和世界坐标系。具体的转换流程如图 4-1 所示。通过这些坐标系的逐步转换，可以将从深度相机获取的二维图像数据转化为三维空间中的实际位置，从而实现精确的目标定位和深度感知。

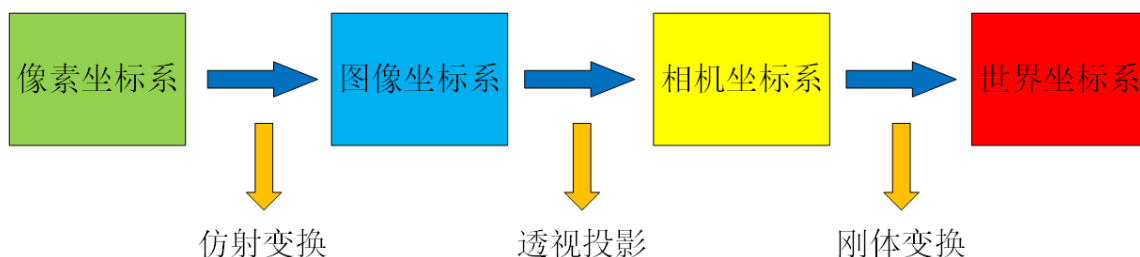


图 4-1 坐标系的转换流程图

Figure 4-1 Diagram of the transformation flow of a coordinate system

4.2.1 仿射变换

仿射变换用于将像素坐标系转换为图像坐标系，是图像处理中从离散像素位置到相机归一化坐标的过渡步骤。图像坐标系以图像光学中心为原点，通常采用毫米为单位，表示像素在成像平面上的实际物理位置。

如图 4-2 所示, $O_{uv}uv$ 代表像素坐标系, O_{uv} 为坐标原点, $O(u_0, v_0)xy$ 代表图像坐标系, 其中 $O(u_0, v_0)$ 为坐标原点。

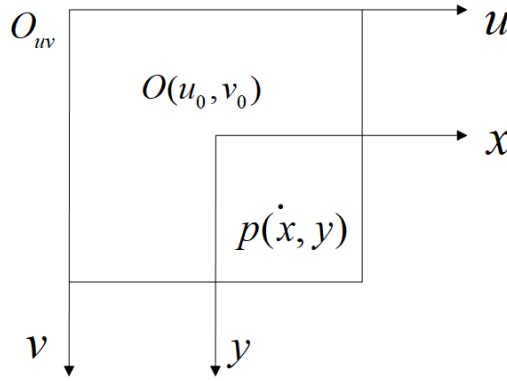


图 4-2 仿射变换

Figure 4-2 Affine transformation

图像上的任意点 p 在像素坐标系与图像坐标系之间的转换关系可表示为:

$$\begin{cases} u = \frac{x}{dx} + u_0 \\ v = \frac{y}{dy} + v_0 \end{cases} \quad (4-1)$$

写成齐次坐标矩阵为:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4-2)$$

在图像坐标系建立完成之后, 目标检测模块输出的二维边界框 (即目标在图像平面上的位置) 为后续三维坐标的计算提供了关键基础。目标检测模型通过对输入的 RGB 图像进行分析, 准确地定位出苹果在图像中的位置, 并以边界框的形式表示目标区域。每个边界框通常包含中心点坐标及其宽度和高度等信息, 其中, 中心点或其他具有代表性的关键点的二维图像坐标 (u, v) 是进行空间重建的核心数据。

这些二维坐标将与深度图像或深度传感器提供的对应像素点的深度信息 d 相结合, 通过相机的内参矩阵完成像素坐标到三维相机坐标系下空间点的转换, 从而得到苹果在三维空间中的位置 (X, Y, Z) 。因此, 准确的目标检测不仅关系到图像中目标的识别效果, 更直接影响到后续三维重建与定位精度, 是整个系统实现果实空间感知与后续操作的基础环节。

4.2.2 透视投影

透视投影是将图像平面上的物理坐标转换为三维空间中相机坐标的过程。此过程中利用深度信息将二维坐标扩展为三维坐标，并以相机光学中心为原点，定义三维空间中的位置。

如图 4-3 所示， oxy 为图像坐标系，其原点为 o ； $O_cX_cY_c$ 为图像坐标系，原点为 O_c 。点 $p(x,y)$ 是点 $P(X_c,Y_c,Z_c)$ 在图像坐标系上的投影， f 为相机焦距。

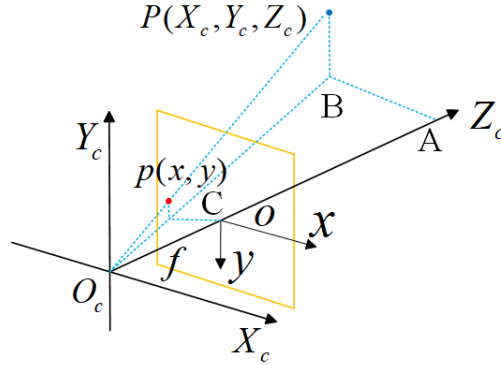


图 4-3 透视投影

Figure 4-3 Perspective projection

根据三角形相似原理可以得到：

$$\frac{AB}{oC} = \frac{AO_c}{oO_c} = \frac{PB}{pC} = \frac{X_c}{x} = \frac{Z_c}{f} = \frac{Y_c}{y} \quad (4-3)$$

$$x = f \frac{X_c}{Z_c}, y = f \frac{Y_c}{Z_c} \quad (4-4)$$

写成齐次坐标矩阵为：

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = [K][0] \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (4-5)$$

式 4-5 中矩阵 K 就是相机的内参矩阵。

4.2.3 刚体变换

刚体变换指的是从相机坐标系到世界坐标系的转换过程，即将相机坐标系中的三维点映射到全局世界坐标系中的位置。该过程通常涉及相机的外参，包括旋转矩阵(用于表征相机坐标系与世界坐标系两者之间的旋转关系)和平移向量(用于描述两者原点间的相对位移)。

刚体变换不会改变物体的形状，仅对物体的空间位置(平移)和朝向(旋转)产生影响。该变换可由两个变量描述：正交单位旋转矩阵 R 和三维平移向量 t 。同样，世界坐标系到相机坐标系的转换也遵循该关系。

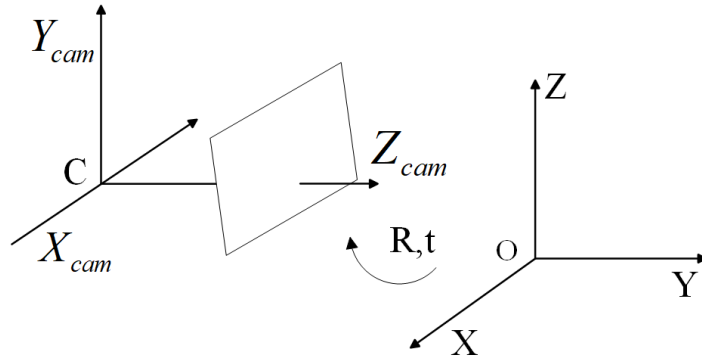


图 4-4 世界与相机坐标系之间的欧式变换

Figure 4-4 Euclidean transformation between the world and the camera's coordinate system

平移是指将世界坐标系的原点转换至相机坐标系的原点，而旋转则涉及三个自由度，即围绕 X 轴、 Y 轴和 Z 轴的旋转，根据不同轴向的旋转角度，可分别得到对应的旋转矩阵 R_x 、 R_y 、 R_z ，最终的整体旋转矩阵由这三个矩阵相乘得到，即旋转矩阵 $R=R_x \times R_y \times R_z$ 。

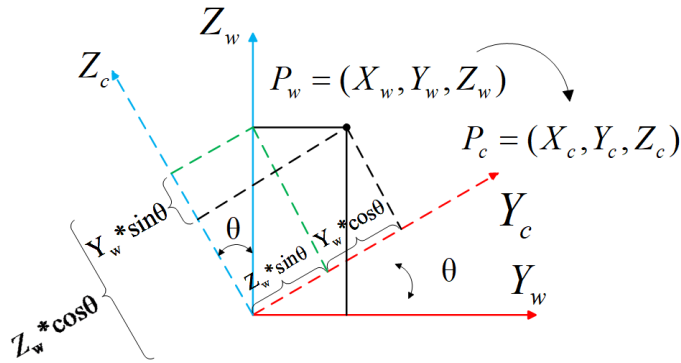


图 4-5 坐标系旋转变换图

Figure 4-5 Coordinate system rotation transformation diagram

以绕 X 轴旋转为例，求解 R_x 。如图 4-5 所示，坐标轴 $Z_w Y_w$ 绕原点旋转 θ 角后得到 $Z_c Y_c$ 。则可得出：

$$\begin{cases} X_c = X_w \\ Y_c = Y_w * \cos \theta + Z_w * \sin \theta \\ Z_c = -Y_w * \sin \theta + Z_w * \cos \theta \end{cases} \quad (4-6)$$

写成矩阵形式为：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = R_x \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \quad (4-7)$$

同理可得出绕 Y 轴和绕 Z 轴的关系：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = R_y \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \quad (4-8)$$

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = R_z \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \quad (4-9)$$

因此可以得到旋转矩阵 $R=R_xR_yR_z$ ，T 为平移矩阵：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T \quad (4-10)$$

写成齐次坐标系的形式为：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4-11)$$

式 4-11 中 M 就是相机的外参矩阵。

4.3 相机标定

计算相机的内参和外参矩阵是相机标定是主要目标。内参矩阵包含了相机的焦距、主点坐标(光心)以及图像的畸变系数，外参矩阵用于描述相机坐标系与世界坐标系之间的转换关系。

本文实验中采用了尺寸为 12×9 的标定板，其中每个黑白小方格的边长为 20 毫米。为了确保计算结果的准确性，实验使用固定相机，并通过移动和旋转标定板的方法，在不同角度和距离下拍摄标定板的照片。标定板的多角度和多距离拍摄能够提高标定过程的精度和鲁棒性，如图 4-6 所示。



图 4-6 相机标定标定板图片

Figure 4-6 Picture of the camera calibration board

拍摄照片后，采用张正友标定法^[57]对相机进行校准，并计算相机的内参及畸变系数 k_1 、 k_2 、 p_1 、 p_2 、 k_3 。其中， k_1 、 k_2 、 k_3 为径向畸变系数， p_1 、 p_2 为切向畸变系数，具体如式 4-12、式 4-13、式 4-14 所示：

$$K = \begin{bmatrix} 332.65042114 & 0 & 266.67412324 \\ 0 & 340.41986084 & 141.08411644 \\ 0 & 0 & 1 \end{bmatrix} \quad (4-12)$$

$$k_1, k_2, k_3 = [1.57496469\text{e-}01 \quad -4.28461445\text{e-}01 \quad 2.52681728\text{e-}01] \quad (4-13)$$

$$p_1, p_2 = [-8.80127089\text{e-}05 \quad 1.72501204\text{e-}03] \quad (4-14)$$

在获取相机的系数矩阵后，通过建立棋盘格物体点与图像点的对应关系，采用 solvePnPRansac 方法估算棋盘格在相机坐标系中的位姿，即获取其旋转向量和平移向量。基于这些位姿信息，进一步计算出此时相对标定板的相机外参矩阵(式 4-15 为旋转矩阵，式 4-16 为平移矩阵)，用来描述世界坐标系与相机坐标系之间的变换关系。

$$R = \begin{bmatrix} 0.01802874 & 0.99843473 & 0.05294383 \\ -0.99418607 & 0.01227957 & 0.10697324 \\ 0.10615567 & -0.05456461 & 0.99285129 \end{bmatrix} \quad (4-15)$$

$$T = \begin{bmatrix} -0.44466793 & -0.76249192 & -0.469975 \\ 0.40936621 & -0.63970312 & 0.65053765 \\ -0.79667417 & 0.09688135 & 0.59659389 \end{bmatrix} \quad (4-16)$$

在此基础上，将一个虚拟立方体和三维坐标轴的顶点投影到图像平面，并绘制在标定图像上。绘制结果直观地展现了相机坐标系与世界坐标系的相对位置和方向，如图 4-7 所示，通过这种可视化方式来验证标定结果的准确性。

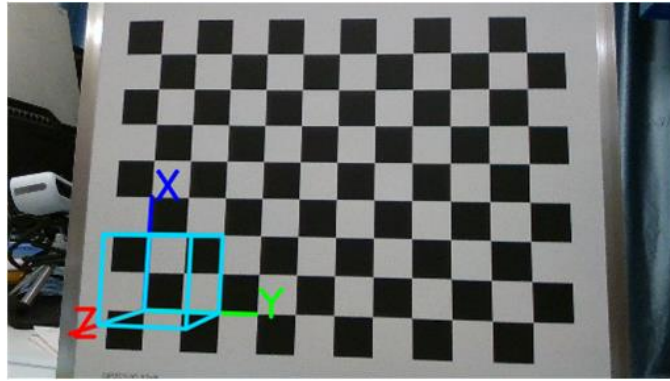


图 4-7 相机标定可视化结果图

Figure 4-7 Camera calibration visualisation plot

4.4 深度信息获取

三角测量法(Triangulation)是计算机视觉中基于几何原理的深度信息获取方法，广泛应用于立体视觉和多视图几何等领域。其核心思想是通过分析同一场景点在不同相机视角下的投影位置，推算该点在三维空间中的深度和位置。

三角测量法的关键概念是视差(Disparity)，即同一物体在不同视角下的投影位置差异。视差的大小与物体的深度成反比：物体距离相机越近，视差越大；距离越远，视差越小。因此，通过视差计算，可以获取场景点的深度信息。

立体匹配是计算视差的核心步骤，其通过分析双目相机采集的两幅图像，建立像素点的对应关系，即在左图中选定一个像素点，并在右图中找到其对应位置。这两个位置的横坐标之差即为视差。所有像素点的视差值构成视差图，该视差图可进一步转换为深度图。

本文采用 Intel RealSense D415 双目相机采集左右视图，并使用视差计算生成初始视差图。考虑到实际场景中由于光照变化、遮挡、纹理缺失等因素会引入噪声，导致深度图不连续或不准确，故在生成深度图后，需对原始深度数据进行后处理。本文采用快速排序对视差值进行局部排序平滑，同时结合双边滤波保留边缘信息，去除高频噪声，从而生成更为准确和平滑的深度图。

具体计算方法如图 4-8 所示，其中 O_L 、 O_R 为两个相机的光心， d 为基线距离， f 为相机焦距， P_L 、 P_R 为点 P 在两个相机感光器上的成像点， X_L 与 X_R 的差值即为视差。

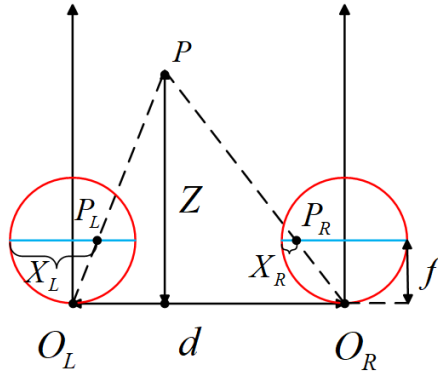


图 4-8 三角测量法

Figure 4-8 Triangulation

因此目标点 P 的距离为：

$$Z = \frac{fd}{|X_L - X_R|} \quad (4-17)$$

4.5 苹果的三维坐标

在果园智能采摘系统中，准确获取苹果在三维空间中的位置是实现机械臂精确抓取的关键。本文基于 YOLOv11 目标检测模型与双目视觉系统相结合的方式，提出了一种高精度的苹果三维坐标获取方法。该方法综合利用 YOLOv11 检测出的二维图像坐标与相机计算出的深度信息，最终恢复出苹果在世界坐标系下的三维空间位置。

YOLOv11 作为前端目标检测模块，能够在 RGB 图像中快速准确地识别出苹果的位置，并输出每个检测框的二维图像中心坐标(u,v)。这一坐标对应于图像平面中目标的像素位置，表示目标在二维图像中的水平和垂直位置。

随后，双目相机通过视差计算获取每个像素点的深度信息，进而为对应的图像坐标提供距离信息 Z，即苹果与相机之间的距离。

在获取了二维图像坐标(u,v)及其对应的深度值 Z 后，利用相机的内参矩阵 K，可依据针孔相机模型完成图像坐标的反投影，从而将其还原为相机坐标系下的三维空间坐标(X,Y,Z)。反投影公式如 4-18 所示。

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = Z \cdot K^{-1} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (4-18)$$

4.6 实验结果分析

在经过相机标定、视差图计算以及三角测量计算后，就可以计算出苹果质量中心

在图像中的中心坐标信息。从实际的果园采摘范围出发，将三维定位的实验距离区间设定在 0.3m-1m 范围内，定位实验的顺序按距离减小的规律来排列，通过计算人工测量的苹果实际距离和相机测量距离的误差来分析定位的准确性。从降低实验的复杂性方面的考虑，此次实验仅在距离上做实验分析，不考虑图像上的坐标，部分实验数据如表 4-1 所示。

经过分析表中的数据可知，双目相机到苹果之间的测量距离以及实际距离的平均绝对误差值是 0.0148m，平均相对误差为 2.29%，因此说明本实验中的方法可以用于苹果的定位。从数据上能够反映出距离越远误差越大，距离越近误差越小。造成误差的可能原因有在标定过程中产生的误差，立体匹配算法计算出的视差的误差，相机镜头的畸变，基线长度过短等。

表 4-1 苹果定位实验结果表

Table 4-1 Apple positioning experiment results table

序号	测量距离(m)	实际距离(m)	相对误差	绝对误差
1	0.968	1.000	0.032	0.032
2	0.874	0.900	0.029	0.026
3	0.778	0.800	0.028	0.022
4	0.686	0.700	0.020	0.014
5	0.585	0.600	0.025	0.015
6	0.492	0.500	0.016	0.008
7	0.439	0.450	0.024	0.011
8	0.390	0.400	0.025	0.010
9	0.342	0.350	0.023	0.008
10	0.298	0.300	0.007	0.002

4.7 本章小结

本章深入探讨了苹果三维定位的关键技术，内容涵盖深度相机、坐标变换原理、相机标定以及深度信息获取等多个方面。首先，介绍了深度相机的主要几种类型，包括结构光、飞行时间、双目视觉以及主动立体视觉，并介绍了其优缺点及其在三维定位中的作用。接着，深入分析了三维坐标系中的转换方法，为后续的定位计算奠定了理论基础。随后，通过详细的相机标定步骤，确保了三维定位系统的精确性和稳定性。

之后，探讨了如何利用深度相机有效获取深度信息，并结合检测模型获取到了二维坐标信息从而获取苹果的三维坐标，实现对苹果的三维定位。最后，分析实验结果，表明了本章提出方法的有效性和实际应用价值。本章的研究为苹果的三维定位提供了坐标信息的支持，同时为后续章节的应用打下了基础。

第 5 章 果园苹果检测定位系统的实现

5.1 系统功能

目前, 针对果园苹果检测任务, 基于深度学习的目标检测模型在高性能服务器端通常能够取得较为理想的检测精度与实时性表现。然而, 这类服务器通常价格昂贵, 搭载了高算力 GPU (如 NVIDIA RTX 或 A100 系列), 不仅初期采购成本高, 而且在运行维护中还需消耗大量电力资源。此外, 服务器端部署通常需要专业的算法工程师与运维人员进行持续技术支持, 这对于大多数普通果农或小型农业合作社来说, 不仅存在较高的经济门槛, 也面临着技术人员短缺与操作复杂度高等现实问题, 极大地限制了该类方案在农业一线的落地推广与应用普及。

因此, 为了实现果园自动化检测方案的低成本部署与便捷操作, 本文选择以边缘设备作为计算平台, 结合高性价比的传感器与轻量化目标检测模型, 打造一套实用、高效、便捷的果园苹果检测系统。该系统旨在减少对高性能计算资源的依赖, 提升终端智能化水平, 从而使普通用户也能够轻松部署与使用。在硬件方面, 本文选用具备较强环境适应能力且价格相对低廉的 Intel RealSense D415 深度相机进行图像数据采集。该相机具备同时采集 RGB 图像与深度信息的能力, 能够为苹果的空间定位提供重要的三维感知支持, 适用于果园等非结构化环境中的复杂检测需求。

基于上述设计目标, 本文提出并实现了一套融合相机、边缘计算设备与 YOLO 目标检测模型的果园苹果检测系统。系统架构中, D415 相机采集的图像数据作为输入, 经由边缘设备内部加载的检测模型进行推理分析, 实现对苹果目标的实时检测与精准定位。检测系统主要包括三个核心功能模块: 图像采集、图像推理、结果分析与处理, 具体流程如下。

(1) 收集图像。

系统使用 Intel RealSense D415 深度相机对果园现场进行图像采集。该相机集成了红外投射器、红外摄像头和 RGB 摄像头, 能够在不同光照条件下提供稳定的深度感知能力。系统通过 RealSense SDK 与相机进行通信, 实现对 RGB 与深度图像的同步采集。为提升检测模型的输入质量, 采集到的图像在送入检测模型之前将经过图像增强、去噪、尺寸归一化等预处理操作, 以提升模型对不同环境条件下苹果目标的鲁棒性和

检测精度。

(2) 推理图像。

在图像采集模块获取图像数据后，这些图像被送入 YOLO 系列的目标检测模型中进行目标识别。YOLO 模型以其高效、轻量和端侧友好性被广泛应用于边缘计算场景中。本文部署模型于搭载了 RK3588 芯片的边缘计算设备上，并借助其内置的 NPU（RKNPU）对模型推理过程进行加速，从而有效提升系统的实时响应能力和图像处理吞吐量。推理过程输出苹果目标的二维边界框位置、置信度分值及类别信息，为后续的空间定位打下基础。

(3) 分析与处理结果。

YOLO 模型完成二维检测后，系统将检测到的边界框信息与 D415 获取的深度图进行融合，利用深度图中对应区域的像素深度值对每个苹果进行三维重建，获取其在果园环境中的实际空间坐标。通过该空间信息，系统能够输出每个苹果在相机坐标系下的三维位置，为机器人执行后续的自动采摘任务或路径规划提供精确目标点。同时，系统支持将分析结果进行记录、可视化展示或生成定制化的检测报告，以便农业管理者进行数据追踪与决策优化。

5.2 系统设计

5.2.1 RK3588

RK3588 是 Rockchip(瑞芯微)推出的一款高性能 AI 处理器(SoC, 系统级芯片)。该处理器采用 8 核 CPU 设计，基于 Arm Cortex 架构，由 4 个主频最高可达 2.4GHz 的 Cortex-A76 高性能核心与 4 个主频最高可达 1.8GHz 的 Cortex-A55 高能效核心组成。其支持 big.LITTLE 异构架构，在保证高性能的同时兼顾能耗优化，实现高效的计算处理能力。其 GPU 集成了 Arm Mali-G610 MP4，支持 OpenGL ES、Vulkan 和 OpenCL，可高效处理复杂图形任务及高分辨率视频播放。此外，RK3588 内置 6 TOPS 的神经网络处理单元(NPU)，兼容 TensorFlow、PyTorch 和 Caffe 等主流深度学习框架，能够高效执行目标检测、人脸检测及自然语言处理等计算密集型 AI 推理任务。

在存储和接口方面，RK3588 支持 LPDDR4 和 LPDDR5 内存，最大容量可达 32GB，并提供高速存储接口，包括 eMMC 5.1、UFS 3.1 和 SD 卡。其接口配置丰富，拥有多达 8 个 USB 接口(支持 USB 3.1 和 USB 2.0)，4 个独立 HDMI 2.1/DP 接口用于多屏显示，以及 MIPI-DSI 和 eDP 接口适配嵌入式显示设备。此外，还支持 PCIe 3.0 和 SATA

3.0 接口，可满足外接存储或扩展卡的需求。

在多媒体处理方面，RK3588 支持 8K@60fps 的视频解码(包括 H.265、H.264、VP9 等格式)，并兼容最新的 AV1 编解码标准。同时，其视频编码能力可达 8K@30fps，支持 H.265/H.264 编码，内置高性能图像信号处理器(ISP)，可处理多路摄像头输入。

软件方面，RK3588 兼容多种操作系统，包括 Android、Linux(如 Debian 和 Ubuntu)以及 RTOS，为开发者提供了灵活的选择。此外，其 AI 软件支持完善，提供 SDK 和工具链，支持模型转换、优化和部署，并兼容 TensorFlow Lite、ONNX 和 RKNN 等推理框架，使 AI 应用的开发和集成更加高效。

RK3588 的应用场景涵盖边缘计算、智能视觉、多媒体处理和工业自动化。它的高性能 CPU、GPU 和 NPU 能够高效执行本地推理任务，减少对云端的依赖；其强大的视频编解码能力适用于家庭影院和智能电视盒子；而丰富的传感器输入与控制接口使其成为工业机器人和自动化设备的理想选择。

RK3588 基于先进的 8nm 制程工艺，兼具高性能与低功耗。其 big.LITTLE 架构与电源管理技术进一步提升了在能耗敏感场景下的适应性。综合来看，RK3588 是一款性能强劲、接口丰富且生态完善的 SoC，更加适合 AI 推理、高分辨率多媒体处理和多任务并行计算等高要求场景。

5.2.2 功能设计

为了实现果园环境中苹果的自动检测与三维定位，本文设计并实现了一套基于边缘计算平台的果园苹果检测系统。该系统集成了图像采集、目标检测、空间定位、数据分析与系统管理等多项功能，能够在非结构化的果园环境中高效、稳定地运行。

本系统主要由以下功能模块构成：图像采集模块、苹果检测与定位模块、数据处理与报告模块，以及系统管理模块。系统硬件设备如图 5-1 所示，为本研究中使用的 Intel RealSense D415 深度相机与搭载瑞芯微 RK3588 芯片的边缘计算平台的实物图。图 5-2 展示了系统的功能流程图，包括相机初始化、模型加载、图像输入、图像处理、苹果检测、深度计算、坐标提取与结果保存等关键步骤。各模块协同运行，构成了完整的数据采集与处理流程。以下将对各模块进行详细说明。



图 5-1 双目相机和 RK3588

Figure 5-1 Binocular camera and RK3588

图像采集模块作为系统的前端数据输入单元，负责采集果园现场的图像数据。系统采用 Intel RealSense D415 深度相机，该相机支持同步采集 RGB 图像与深度图像，可在强光、阴影等复杂户外条件下保持良好的感知稳定性。通过 USB 3.0 接口，图像数据被实时传输至 RK3588 边缘平台，为后续图像处理与目标检测提供基础数据支持。

在图像采集过程中，系统利用 RealSense SDK 与 D415 相机进行通信，支持对相机参数进行灵活配置，包括图像分辨率、帧率、曝光补偿等。为了提升检测精度，采集图像在送入模型前会进行图像增强、去噪、颜色校正等预处理操作，从而提升目标检测模型对不同环境变化的鲁棒性。

苹果检测与定位模块是系统的核心功能模块，负责完成苹果的目标识别与三维空间定位。该模块基于 RK3588 平台强大的边缘计算能力，部署经过训练与优化的 YOLO 系列目标检测模型，对采集的 RGB 图像进行实时推理，输出苹果的二维边界框、类别标签及其置信度值。

检测完成后，系统将检测结果与 D415 获取的深度图像进行融合。根据相机内参模型，将二维图像坐标映射至三维空间，实现每个苹果目标的空间坐标估算。具体计算过程中，系统采用像素-深度对的方式进行点云重建，并结合相机内参矩阵与深度图信息，获取目标的真实三维坐标 (X, Y, Z)。该空间定位结果可为后续的自动采摘、导航规划等任务提供精确目标位置信息。

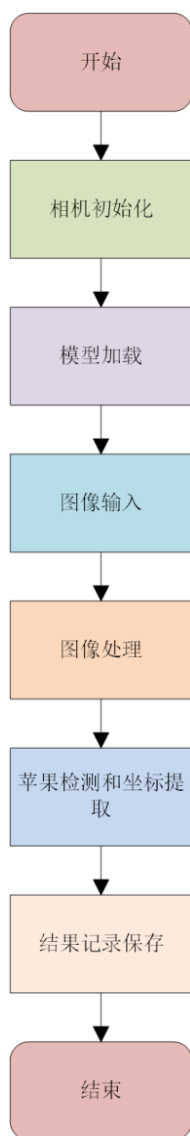


图 5-2 苹果检测系统功能流程图

Figure 5-2 Functional flow chart of Apple detection system

数据处理、存储与报告模块承担系统后端的数据管理任务，包括检测结果接收、信息处理、数据存储与报告生成等。其主要功能如下：

结果分析与处理。模块接收检测模块输出的二维检测结果及其对应的三维空间坐标，依次进行置信度阈值过滤、非极大值抑制和空间位置信息筛选等处理步骤，以提升目标检测结果的准确性与实用性。

数据存储与管理。系统将图像帧、检测结果与三维定位信息按时间顺序存储到数据库或本地存储设备中。支持基于时间戳、图像编号等关键字段进行索引与查询，便于数据溯源与历史分析。

报告与统计输出。系统能够根据检测数据自动生成检测报告，包括苹果目标数量、

空间分布图、检测置信度分布、区域统计等内容，支持可视化展示或导出为文本/图表文件，为果园管理与作业决策提供依据。

系统管理模块负责整体系统的运行控制与参数配置，确保系统的稳定性与可控性。主要功能包括：

参数配置与更新。用户可通过图形化界面或配置文件方式设置或更新系统参数，如模型路径、置信度阈值、采集频率等。

运行状态监控。模块实时监测系统运行状态，包括相机连接状态、边缘计算负载、帧率等关键指标，并将运行日志进行记录，便于系统调试与运维管理。

数据接口与权限控制。系统支持通过局域网或无线网络将检测结果同步上传至远程服务器或云平台，同时提供用户权限控制机制，保障数据安全性与系统稳定性

5.3 模型转化与部署

5.3.1 RKNPU

检测模型的训练旨在保证模型达到高精度，而模型的改进与优化则侧重于提升泛化能力，以训练出更高精度的模型。相比之下，模型推理的核心在于将输入数据(如图像)传递给已训练完成的模型进行预测，而无需大量的反向传播计算。因此，推理阶段的重点不在训练，而在于模型的部署效率、推理的实时性及准确性。

RKNPU(瑞芯微神经处理单元)经过多轮升级优化，能够在边缘计算设备中提供高效的推理性能，特别是在目标检测等任务的推理部署上表现良好，同时保持较低功耗，满足实时性与准确性的需求。并且 RKNPU 具有高吞吐量和低延迟，可以满足嵌入式系统对部署性能和资源效率的严格要求。

5.3.2 模型转化

训练好的模型需要将格式转换为 RKNPU 支持的格式——RKNN 格式，才能利用 RKNPU 加速推理。根据前面章节的实验和对比，YOLOv11 更适合用于本文的果园苹果检测任务，因此这里就是对 YOLOv11 模型进行转换。在此之前，必须先将模型转换为 ONNX(Open Neural Network Exchange)通用格式。ONNX 是一种开放的深度学习模型格式，作为 PyTorch 训练模型转换为 RKNN 格式的中间桥梁。通过标准化的数据类型和计算图结构，ONNX 使得不同深度学习框架之间的数据能够实现高效的交互和存储。对于 YOLOv11 模型，通常是先使用 PyTorch 进行训练，然后通过 `torch.onnx.export()` 方法将模型导出为 ONNX 格式。在导出时，需要确保输入张量的形状、动态批量处理

以及算子兼容性问题。

获得 ONNX 模型后, 需要使用 Rockchip 官方提供的 RKNN-Toolkit 工具进行转换。RKNN(Rockchip Neural Network)是专门为瑞芯微 NPU 设计的模型格式, 该工具提供了一整套转换和优化方案, 能够将 ONNX 及 TensorFlow 等框架训练的模型转换为 RKNN 格式。转换过程包括以下几个关键步骤:

- (1) 加载 ONNX 模型: 使用 `rknn.load_onnx()`方法读取 ONNX 文件。
- (2) 配置转换参数: 调用 `rknn.config()`进行模型优化配置, 包括输入通道顺序(NCHW 或 NHWC)、均值和标准差归一化、优化级别、量化方法等。
- (3) 编译和导出: 使用 `rknn.build()`进行模型编译, 并通过 `rknn.export_rknn()`生成最终的 RKNN 文件。

在转换过程中, 需特别注意使用官方提供的转换包和 NPU 兼容算子, 以确保模型能够在目标硬件上正确运行。此外, RKNN 模型本身并不包含后处理逻辑, 因此在推理阶段需要通过 RKNPU SDK 进行后处理。

5.4 系统测试

基于前文所述的系统设计与模块实现, 本文搭建了完整的果园苹果检测系统原型平台, 硬件由 Intel RealSense D415 深度相机和基于瑞芯微 RK3588 芯片的边缘计算平台构成。为了验证系统在实际果园场景下的功能实现情况和检测性能, 本文在室外自然光照条件下进行了模拟测试。

测试环境选取了一处具有果园特征的户外区域进行模拟, 现场具有一定复杂性, 包含多种遮挡、光照变化和背景干扰, 旨在模拟果园真实作业条件下的使用场景。测试过程中, D415 相机固定在与地面垂直或一定角度倾斜的安装架上, 以模拟农用机器人或采摘机械的安装方式。系统持续采集果园中不同角度和距离的苹果图像, 传输至 RK3588 平台进行实时检测与三维定位。

图 5-6 展示了部分检测过程中的效果图。从图中可以看出, 系统能够准确识别出图像中的苹果目标, 并在图像上绘制出对应的二维边界框。同时, 系统通过融合深度信息, 计算出每个目标在三维空间中的位置坐标, 并将该三维信息在图像界面中进行实时显示。

测试结果表明: 系统具备较高的实时性, 图像采集到结果输出过程延迟在 150~200ms 左右, 可满足移动平台或机械臂操作中的响应需求; 检测精度较高, 在遮

挡部分果实、光照变化明显的条件下依然能够保持较好的识别能力；三维坐标定位稳定可靠，坐标误差控制在 $\pm 2\text{ cm}$ 范围内，可用于大多数采摘引导与路径规划任务。

总体来看，本系统在室外非结构化环境下表现良好，能够稳定完成果园苹果的实时检测与三维定位任务，验证了本文提出的系统设计方案在实际农业应用中的可行性与有效性。该系统为智能果园视觉感知与自动作业提供了一定的技术支撑。

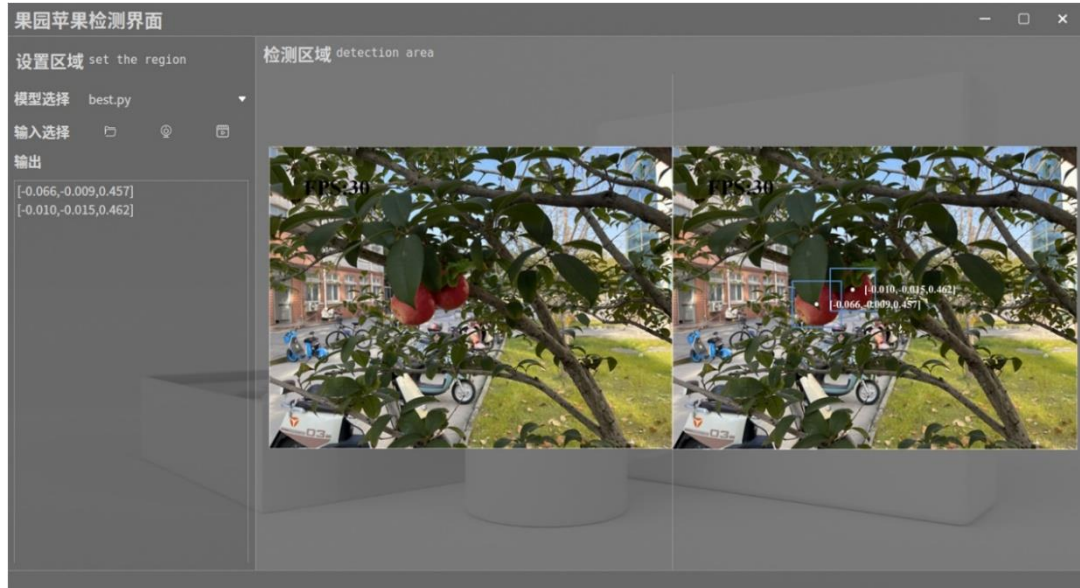


图 5-3 果园苹果检测效果图

Figure 5-3 Orchard apple detection renderings

5.5 本章小结

本章系统地阐述了果园苹果检测与定位系统的实现过程。首先，在系统功能设计部分，对系统的整体功能架构及其核心任务进行了概述，明确了目标检测与空间定位的基本需求，确立了系统设计的目标方向。

随后，在系统架构设计部分，重点介绍了所选硬件平台的组成与性能，特别是基于瑞芯微 RK3588 芯片的边缘计算平台。该平台具备强大的算力与丰富的接口资源，为系统实现实时图像处理和模型推理提供了有力保障。同时，结合系统功能需求，对各模块的结构与作用进行了详细分析，构建了完整的系统功能链路。

在模型转换与部署部分，详细讨论了 RKNN 平台中 RKNPU 的硬件加速特性，介绍了深度学习模型在部署前的转换、量化与优化流程。通过模型的高效适配，使目标检测模型能够在边缘设备上实现低延迟运行，确保了在复杂果园环境中的检测精度与响应速度。

最后,系统测试部分对系统在真实或近似真实果园环境中的运行效果进行了验证。测试结果表明,系统在稳定性、检测精度以及实时响应能力方面均表现良好,满足了果园自动检测与三维定位的实际应用需求。

综上所述,本章从硬件平台选择、系统功能设计到模型优化部署与实际测试,构建了完整的技术实现路径,充分验证了系统设计方案的可行性与有效性,也为后续相关研究与应用实践提供了一定的技术支撑与参考依据。

第6章 总结与展望

6.1 研究工作总结

在本研究中，针对果园苹果检测任务，系统地开展数据集构建、算法改进、双目定位及边缘部署等多个方面的工作。

首先，模型训练需要有较高质量的数据支持，因此自主构建了一个包含多种遮挡情况的果园苹果检测数据集，并对其进行数据增强和数据标注。然后根据果园苹果检测任务的特点，对 YOLOv5、YOLOv7 以及 YOLOv11 三种算法进行了改进。通过对网络结构、模块的改进，提高了算法模型的检测精度和速度，并且实验结果表明，改进后的算法在果园苹果检测任务上具有较好的表现。

其次，通过苹果的三维定位方法，利用张正友棋盘格标定法对 Intel RealSense D415 相机进行标定，成功得到了相机的内外参数并完成了坐标转换。经过立体匹配算法提取深度信息，从实验结果中可以看到，本文的方法基本满足后续在机械自动化采摘的应用需求。

最后，实现整个果园苹果检测定位系统。通过模型转化和优化等技术手段，成功地将模型部署在边缘设备上，并实现了果园苹果的实时检测和定位。在边缘端部署不仅可以减少数据传输的延迟，提高检测效率，还可以降低对高性能计算资源的需求，使得系统更加适合在实际场所中应用。

6.2 研究成果意义

本研究提出的果园苹果检测与定位系统具有重要的意义，该系统有利于推动农业智能化转型升级。它利用先进的计算机视觉和深度学习技术，实现了苹果的自动化、智能化检测和定位，有效替代了传统的人工检测方式，提高了果园管理的效率，并降低了人工成本，从而提升了农业生产的现代化水平和果园的生产效率以及经济效益，是推动农业智能化转型升级的关键一步。更重要的是，该系统为苹果采摘机器人等智能农机装备的应用奠定了坚实的技术基础，加速了农业生产的自动化进程，并有助于缓解农村劳动力短缺等问题。

此外本研究有助于促进相关技术的发展和應用。本研究在数据集构建、目标检测算法改进和边缘部署等方面进行了深入的研究和实践，为相关领域的技术发展提供了

有益的参考和借鉴。例如，针对果园环境特点改进的目标检测算法，不仅可以应用于苹果检测，还可以推广到其他农作物的检测任务中，具有广泛的应用前景，而边缘部署的经验则可以推广到其他农业场景的智能化应用中，促进农业智能化的全面发展。此外，研究中采用的三维定位方法，结合深度图匹配和相机标定技术，为机器人视觉导航和目标抓取提供了有效的解决方案，对机器人技术在农业领域的应用具有重要的推动作用，加速农业机器人的发展和应用进程。

6.3 局限性

尽管本研究取得了一定成果，但在实际应用中仍面临一些挑战。首先，数据集规模较为有限，尚难全面覆盖果园中多变的环境条件，如不同天气、果树品种等。此外，数据采集主要集中在特定果园和季节，设备型号单一，可能存在一定偏差，影响模型的泛化能力。

其次，基于双目视觉的三维定位精度仍受相机标定、深度匹配算法和环境变化等因素影响，复杂场景下可能出现误差，影响后续的抓取操作。

此外，尽管引入了模型转换与量化优化，模型在边缘设备上的推理速度仍受设备计算能力、存储与功耗等限制，实时性仍有提升空间。

6.4 展望

本研究所提出的果园苹果检测与定位系统为实现果园智能化管理和自动化采摘奠定了基础，但仍有很多不足的地方。未来的工作从以下方面着手工作：

(1) 扩大数据集规模，覆盖更多品种、生长阶段、成熟度的苹果，以及多样的果园环境，如不同天气、树形和树叶密度等。

(2) 开展实际应用测试，将系统集成到采摘机器人等设备中，在真实果园环境中验证其有效性，并优化系统性能。

(3) 拓展应用场景，将系统推广至其他果树或农作物的检测与定位，探索其在农业生产中如长势监测、杂草检测和精准施药等方面的潜力。

参考文献

- [1] 2021 年度中国苹果产业发展报告(总篇)精简版[J].中国果菜,2023,43(01):1-8.
- [2] 王壹.中国苹果产量稳居世界第一[N].农民日报,2023-11-18(007).
- [3] 陈青,吴玄博,殷程凯,等.苹果机械化采收技术与装备研究现状[J/OL].林业工程学报,1-12.
- [4] 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.
- [5] 鲍秀兰,包有刚,马萧杰,等.自然环境下柑橘采摘机器人避障规划研究[J].农业机械学报,2025,56(02):420-428.
- [6] 王丹丹,宋怀波,何东健.苹果采摘机器人视觉系统研究进展[J].农业工程学报,2017,33(10):59-69.
- [7] 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.
- [8] Shaoqing R,Kaiming H,Ross G, et al. Faster R-CNN: Towards real-time object detection with region proposal networks.[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137- 1149.
- [9] 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.
- [10] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7263-7271.
- [11] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [12] 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.
- [13] 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]//Proceedings of the IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.
- [14] Varghese R, Sambath M. 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). IEEE, 2024: 1-6.
- [15] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision. Springer, Cham, 2025: 1-21.
- [16] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.
- [17] Khanam R, Hussain M. YOLOv11: An Overview of the Key Architectural Enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.
- [18] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [19] 赵丽亚.基于深度学习的复杂背景下苹果检测与定位研究[D].吉林农业大学,2024.
- [20] Font D, Pallejà T, Tresanchez M, et al. Counting red grapes in vineyards by detecting specular spherical reflection peaks in RGB images obtained at night with artificial illumination[J]. Computers and Electronics in Agriculture, 2014, 108: 105-111.
- [21] 何斌,张亦博,龚健林,等.基于改进 YOLO v5 的夜间温室番茄果实快速识别[J].农业机械学报,2022,53(05):201-208.
- [22] 王金鹏,高凯,姜洪喆,等.基于改进的轻量化卷积神经网络火龙果检测方法[J].农业工程学报,2020,36(20):218-225.
- [23] Yue X Q, Shang Z Y, Yang J Y, et al. A smart data-driven rapid method to recognize the strawberry maturity[J]. Information Processing in Agriculture, 2020, 7(4): 575-584.
- [24] 张同勋,刘平.一种鲜食葡萄采摘点的快速高效识别方法[J].农机化研究,2020,42(02):189-193.
- [25] 张瑞合,姬长英,沈明霞,等.计算机视觉技术在番茄收获中的应用[J].农业机械学报,2001,(05):50-52+58.
- [26] 易诗,李俊杰,张鹏,等.基于特征递归融合 YOLOv4 网络模型的春见柑橘检测与计

- 数[J].农业工程学报,2021,37(18):161-169.
- [27] Arefi A, Motlagh A M, Mollazade K, et al. Recognition and localization of ripen tomato based on machine vision[J]. Australian Journal of Crop Science, 2011, 5(10): 1144-1149.
- [28] Huang Z, Wane S, Parsons S. Towards automated strawberry harvesting: Identifying the picking point[C]//Towards Autonomous Robotic Systems: 18th Annual Conference, TAROS 2017, Guildford, UK, July 19–21, 2017, Proceedings 18. Springer International Publishing, 2017: 222-236.
- [29] 李国进,黄晓洁,李修华.基于改进 YOLOv3 的树上成熟芒果检测方法[J].沈阳农业大学学报,2021,52(01):70-78.
- [30] Feng J, Wang S, Liu G, et al. A separating method of adjacent apples based on machine vision and chain code information[C]//Computer and Computing Technologies in Agriculture V: 5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011, Beijing, China, October 29-31, 2011, Proceedings, Part I 5. Springer Berlin Heidelberg, 2012: 258-267.
- [31] 江梅,孙飒爽,何东健,等.融合 K-means 聚类分割算法与凸壳原理的遮挡苹果目标识别与定位方法[J].智慧农业,2019,1(02):45-54.
- [32] 荆伟斌,李存军,竞霞,等.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(05):75-83.
- [33] 龙燕, 杨智优, 何梦菲.基于改进 YOLOv7 的疏果期苹果目标检测方法[J].农业工程学报, 2023, 39(14): 191-199.
- [34] Mehta S S, Burks T F. Vision-based control of robotic manipulator for citrus harvesting[J]. Computers and electronics in agriculture, 2014, 102: 146-158.
- [35] 王壮飞,卢海林,耿文涛,等.基于 OpenCV 的水果采摘机器人目标检测与定位系统研究[J].电子技术与软件工程,2022,(02):137-140.
- [36] 李振雨,王好臣.基于视觉识别定位的苹果采摘系统研究[J].图学报,2018,39(03):493-500.
- [37] Si Y, Liu G, Feng J. Location of apples in trees using stereoscopic vision[J]. Computers and Electronics in Agriculture, 2015, 112: 68-74.
- [38] 牟其松.基于深度学习的水果采摘通用检测模型研究[D].天津理工大学,2021.
- [39] 赵海东.基于 Kinect 的田间苹果检测与定位方法研究[D].西北农林科技大学,2019.

- [40] Vilaplana V, Rosell-Polo J R, et al. Multi-modal deep learning for Fuji apple detection using RGB-D cameras and their radiometric capabilities[J]. Computers and Electronics in Agriculture, 2019, 162: 689-698.
- [41] 张冬妍,陈诺,张淇,等.基于改进 YOLOv8s 的轻量级果园李子检测方法[J/OL].农业工程学报,1-7[2025-01-05].
- [42] 郑镇辉,张淡然,王槊,等.基于 YOLOv8 的竹笋分割方法研究[J/OL].中国农机化学报,1-6[2025-01-05].
- [43] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [44] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.
- [45] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[C]//Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer Berlin Heidelberg, 2006: 404-417.
- [46] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International conference on computer vision. Ieee, 2011: 2564-2571.
- [47] Bodla N, Singh B, Chellappa R, et al. Soft-NMS--improving object detection with one line of code[C]//Proceedings of the IEEE international conference on computer vision. 2017: 5561-5569.
- [48] Ge Z, Liu S, Li Z, et al. Ota: Optimal transport assignment for object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 303-312.
- [49] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.
- [50] Ma S, Xu Y. Mpdious: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.
- [51] 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.
- [52] Nascimento M G, Fawcett R, Prisacariu V A. Dsconv: Efficient convolution operator[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 5148-5157.
- [53] Wang Y, Zhang J, Kan M, et al. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 12275-12284.
- [54] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[J]. arXiv preprint arXiv:2209.07947, 2022.
- [55] Li J, Wen Y, He L. Scconv: Spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 6153-6162.
- [56] Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 16965-16974.
- [57] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations[C]//Proceedings of the seventh ieee international conference on computer vision. Ieee, 1999, 1: 666-673.

攻读学位期间发表的学术论文目录

- [1] 王彦辉,汪军.基于改进 YOLOv5s 的果园苹果检测算法[J/OL].重庆工商大学学报(自然科学版),1-9.
- [2] 王彦辉,汪军,黄翔翔.基于改进 YOLOv7 的果园苹果检测算法[J/OL].青岛理工大学学报, 已录用

攻读学位期间参与的科研项目

- [1] 安徽未来技术研究院企业合作项目, 面向工业安全的工业互联网个性化 AI 研究与应用(2023qyhz10).
- [2] 安徽高科电子股份有限公司企业合作项目, 基于机器视觉的焊缝坡口自动识别定位跟踪与自动焊接技术的研究与开发(HX-2024-08-003).
- [3] 大跨度钢结构农业大棚智能控制设备研发与应用(GXXT-2019-020)

攻读学位期间参与的学科竞赛

- [1] 第九届安徽省“互联网+”大学生创新创业大赛《去危就安——面向工业安全生产的防护设备佩戴 AI 检测》, 铜奖.
- [2] 第九届安徽省“互联网+”大学生创新创业大赛《基于深度学习的汽车轮胎参数云边协同检测系统》, 铜奖.
- [3] 第九届安徽省“互联网+”大学生创新创业大赛《MVIC——面向医疗产业的 3D 人脸重建及变形技术》, 铜奖.

致 谢

在这论文最后完成的时候，我要向在课题研究和论文撰写过程中所有给予我帮助和支持的人表达最诚挚的谢意。

首先，最需要感谢的是教导了我三年的导师汪军教授。在本研究的整个过程中，汪军老师给予了我悉心的指导和帮助。从选题、开题、研究方法的确立，到论文的撰写和修改，汪军老师都给予了我无私的帮助和支持。他治学严谨、学识渊博，且诲人不倦，这些优秀的品质使我受益良多，并将终身受益。

其次，我要感谢严楠和戴家树老师。在视觉研究方面，严楠和戴家树老师给予了我重要的指导和帮助，为本研究的顺利进行提供了重要的支持。

此外，我还要感谢 MVIC 实验室的各位同学。在研究过程中，我们互相帮助、共同探讨，营造了良好的学术氛围。他们的帮助和鼓励，使我能够克服研究中遇到的困难，顺利完成本研究。

最后，我由衷地感谢我的家人和朋友。每当我遭遇困难与挫折时，他们都会给予鼓励与安慰，给予我无私的关爱与支持。

再次由衷感谢所有曾给予我帮助和支持的人！