

学校代码:10363

学 号: 2210120174



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于双目的同质多目标跟踪与定位方法研究

论 文 作 者 刘 艺

校内导师 吴路路 副教授

校外导师 陈盟

专业学位类别	机 械
--------	-----

研究方向 (领域)	机器视觉
-----------	------

论文提交日期: 2024 年 5 月 28 日

分类号：TP391

单位代码：10363

密 级：公开

学 号：2210120174

基于双目的同质多目标跟踪与定位方法研究
Research on Homogeneous Multi-target Tracking and
Localization Method based on Binocular Vision

学生姓名：_____刘 艺_____

指导教师：_____吴路路 副教授_____

校外导师：_____陈 盟_____

专 业：_____机 械_____

研究方向：_____机器视觉_____

论文答辩日期：2024 年 5 月 28 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：刘艺

日期：2024 年 5 月 28 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：刘艺

指导教师签名：吴路路

日期：2024 年 5 月 28 日

日期：2024 年 5 月 28 日

基于双目的同质多目标跟踪与定位方法研究

摘 要

多目标跟踪与定位是视觉领域的重要内容，在视觉三维重建、视觉结构运动测量、交通监控、武器制导等多个领域中有广泛的应用，给各个行业带来了新的解决方案。由于受目标所处环境的光照、阴影、遮挡以及变化的背景等因素的影响，多目标检测、跟踪与精确定位仍然是视觉领域的重要研究议题。双目视觉测量能够获得被测目标的三维运动信息和更为丰富的数据，相比于单目视觉拥有独特的优势。本文针对机器人全局结构运动双目测量过程中多个相同人工标识的跟踪与精确定位问题，主要研究工作如下：

（1）针对机器人全局结构运动双目视觉测量过程中多目标跟踪与定位的实际需求，设计并搭建了双目视觉多目标跟踪与定位系统装置，分析了相机成像原理，推导了双目视觉三维运动测量模型，阐述了实现双目视角下多目标跟踪与定位的过程。

（2）提出了一种基于改进 YOLOv5 的小目标检测算法，算法针对机器人全局结构三维运动过程中，张贴在机器人结构上多个相同人工小标识的精确检测问题，通过引入 BottleneckCSP 模块减少网络计算参数，加入大尺度特征融合结构提升小目标特征检测范围，融合 SE 注意力机制提升网络效率，最终实现多个相同小目标精确检测。实验结果表明，改进的 YOLOv5 算法能够有效拓展小目标检测范围和检测率，目标像素占整个图像的比率由原算法的 0.01~0.0025 拓展至 0.0014~0.0008；小目标的检测成功率由原算法的 42%提升到 88%。

（3）针对运动测量过程中多目标的跟踪与精确定位问题，运用卡尔曼滤波算法对多目标轨迹进行预测、匈牙利算法对前后帧目标进

行匹配，从而实现多目标跟踪。通过对跟踪标识图像进行后处理，根据人工标识的直线特征，通过提取边缘的亚像素坐标再结合线性拟合求交方法实现标识特征点的精确定位。

关键词：双目视觉；多目标检测；YOLOv5；多目标跟踪；目标定位

RESEARCH ON HOMOGENEOUS MULTI-TARGET TRACKING AND LOCALIZATION METHOD BASED ON BINOCULAR VISION

ABSTRACT

Multi-target tracking and positioning is an important content in the field of vision. It has been widely used in many fields such as visual three-dimensional reconstruction, visual structure motion measurement, traffic monitoring, and weapon guidance, and has brought new solutions to various industries. Due to the influence of illumination, shadow, occlusion and changing background of the target environment, multi-target detection, tracking and precise positioning are still important research topics in the field of vision. Binocular vision measurement can obtain three-dimensional motion information and richer motion data of the measured target, which has unique advantages over monocular vision. In this paper, the tracking and precise positioning of multiple identical artificial markers in the process of binocular measurement of robot global structural motion is studied. The main research work is as follows:

(1) Aiming at the actual demand of multi-target tracking and positioning in the process of binocular vision measurement of robot global structure motion, a binocular vision multi-target tracking and positioning system is designed and built. The principle of camera imaging is analyzed, the three-dimensional motion measurement model of binocular vision is deduced, and the process of multi-target tracking and positioning from binocular perspective is expounded.

(2) A small target detection algorithm based on improved YOLOv5 is proposed. The algorithm aims at the problem of accurate detection of multiple artificial small marks posted on the robot structure during the

three-dimensional motion of the global structure of the robot. The BottleneckCSP module is introduced to reduce the network calculation parameters. The large-scale feature fusion structure is added to improve the detection range of small target features, and the SE attention mechanism is integrated to improve the network efficiency, and finally the accurate detection of multiple small targets is realized. The experimental results show that the improved YOLOv5 algorithm can effectively expand the detection range and detection rate of small targets, and the ratio of target pixels to the whole image is expanded from 0.01~0.0025 to 0.0014~0.0008. The success rate of small target detection is increased from 42% to 88%.

(3) Aiming at the problem of multi-target tracking and precise positioning in the process of motion measurement, the Kalman filter algorithm is used to predict the multi-target trajectory, and the Hungarian algorithm is used to match the front and rear frame targets, so as to realize multi-target tracking. By post-processing the tracking identification image, according to the linear characteristics of the artificial identification, the accurate positioning of the identification feature points is realized by extracting the sub-pixel coordinates of the edge and combining the linear fitting intersection method.

KEY WORDS: binocular vision; multi-target detection; YOLOv5; multiple target tracking; object location

目 录

摘 要.....	I
ABSTRACT.....	III
第 1 章 绪 论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 目标检测与特征点定位方法研究现状.....	2
1.2.2 多目标跟踪方法研究现状.....	4
1.3 本文的研究内容.....	5
1.4 论文的组织结构.....	6
第 2 章 基于双目同质多目标跟踪与定位理论基础.....	8
2.1 基于双目的同质多目标跟踪与定位系统.....	8
2.1.1 视觉系统构成.....	8
2.1.2 同质多目标跟踪与定位任务分析.....	9
2.2 双目视觉测量技术.....	11
2.2.1 摄像机成像几何模型.....	11
2.2.2 双目视觉测量基础.....	14
2.3 多目标检测跟踪与定位过程.....	17
2.3.1 图像预处理.....	18
2.3.2 神经网络基本原理.....	19
2.3.3 目标检测跟踪模型.....	21
2.4 本章小结.....	22
第 3 章 基于改进 YOLOv5 的小目标检测	23
3.1 YOLO 概述.....	23
3.2 YOLOv5 算法.....	24
3.2.1 网络结构.....	24
3.2.2 评价指标.....	26
3.3 基于改进 YOLOv5 的小目标检测算法	28
3.3.1 引入 BottleneckCSP 模块	28
3.3.2 增加大尺度特征融合结构.....	29
3.3.3 融合 SE 注意力机制.....	30
3.4 实验及结果分析.....	31
3.4.1 实验装置.....	31
3.4.2 实验数据准备.....	32
3.4.3 结果分析.....	33
3.5 本章小结.....	36

第 4 章 多目标跟踪与精确定位.....	37
4.1 跟踪算法概述.....	37
4.2 多目标跟踪算法.....	38
4.2.1 卡尔曼滤波轨迹预测方法.....	38
4.2.2 基于匈牙利算法的目标匹配.....	41
4.3 目标精确定位方法.....	43
4.4 实验结果与分析.....	47
4.5 本章小结.....	52
第 5 章 总结与展望.....	54
5.1 总 结.....	54
5.2 展 望.....	55
参考文献.....	56
攻读硕士学位期间主要成果.....	59
致 谢.....	60

第1章 绪 论

1.1 研究背景与意义

人工智能目前是世界产业变革的主要方向,是第四次科技革命的核心技术,对新世界的经济走向、社会发展和人类生活产生了重要的影响^[1]。人工智能开启了时代的变革,可以说它是对世界全新的一种思考和认识方式,是对传统生活、传统行业,甚至是传统世界的全新思考和认识^[2]。到目前为止,全世界已经有超过二十多个国家把人工智能确定为国家发展战略。为实现我国创新型国家建设,成为世界科技强国,国务院发布《新一代人工智能发展规划》明确规定部署人工智能发展的先发优势^[3]。人工智能将远远超出人们的想象给社会带来变革,未来几十年人工智能会不断深入社会生活的方方面面,改变传统服务业、能源、医疗、教育、制造业等,极大提高社会生产效率。

机器视觉结合了图像处理、模式识别、深度学习等多项技术,是人工智能的一个重要分支,致力于实现计算机对现实世界视觉信息的理解与分析。随着科技的不断进步和工业化进程的加速,机器视觉的研究和应用正日益受到广泛关注。计算机性能的增强和算法的优化,使得机器视觉技术应用越来越广泛,能够模拟人类的视觉感知过程,对图像和视频进行高效、准确的处理和分析。在自动化生产线上,机器视觉系统可以替代人工进行多个目标的质量检测、零件识别等任务,提高生产效率和产品质量^[4]。在交通管理中,视觉系统可以用于车辆检测、行人检测、交通违章检测等。通过实时监测和分析交通状况,提升交通维护效率,预防交通事故。在医疗领域,视觉技术可以进行细胞检测、病理切片分析,减轻医生的负担。通过对图像中多个目标的识别和分类,可以辅助医生进行疾病诊断和治疗^[5-7]。机器视觉的研究和应用具有显著的价值。而多目标检测跟踪,作为机器视觉领域的关键技术之一,更是为复杂场景下的目标识别、定位和行为分析提供了强大的支持。在算法优化、硬件设计等方面,机器视觉的研究将不断推动相关学科的技术进步和创新。

视觉技术根据使用成像相机的数目有单目视觉、双目视觉和多目视觉。以往的多目标检测跟踪大多数都是利用单目视觉完成的,单目视觉只追求多目标在图像或视频中的检测和跟踪,不考虑后续应用^[8,9]。双目视觉技术凭借其对空间深

度信息的准确捕捉,已成为多目标检测任务中的关键技术。随着现代社会的快速发展,多目标检测在自动驾驶、智能监控、机器人导航等领域的应用需求日益迫切,而双目视觉技术正是解决这些需求的重要工具。双目视觉系统通过模仿人类双眼的视觉模型,可以得到视场的深度信息,从而实现对目标的精确定位和识别。此外,双目视觉多目标检测算法在依据机器学习、深度学习等技术的发展条件下也在不断自我优化和完善,为实际应用提供了强大的技术支持^[10]。解决目标运动、物体遮挡、拍摄角度等问题,能够提高目标检测的精度和稳定性,减少误检和漏检的情况^[11]。双目视觉的多目标检测研究将有助于推动视觉技术的发展,该研究将不断推动机器视觉技术的进步和创新,为实际应用提供更高效、更可靠的解决方案。

本课题来源于国家自然科学基金(52005003)“工业机器人 3D 全结构刚柔耦合交互特性视觉测量方法研究”,旨在开展针对机器人全局运动结构测量中张贴在机器人结构上多个相同人工标识的跟踪与精确定位问题展开研究。视觉系统具有低成本、灵活性和全局测量等优势,可获取机器人空间运动的多点同步运动数据,进一步探究机器人动静态特性、结构柔性对机器人控制精度的影响,优化机器人系统精密运动控制。因此,探究机器人运动视觉高精度测量理论和方法,对破解机器人精密作业难题具有重要的理论和研究意义。

1.2 国内外研究现状

基于视觉的目标检测跟踪技术在国内外国都受到了广泛的关注和研究,具有重要的应用价值和研究意义。国内外知名高校和研究所如清华大学、上海交大、卡内基梅隆大学、加州大学等,针对双目视觉在多目标检测跟踪领域的应用进行了深入的研究。这些研究主要集中在算法优化、实时性提升以及鲁棒性增强等方面^[12,13]。

1.2.1 目标检测与特征点定位方法研究现状

目前根据目标检测的速度可以将检测方法大致分为两类,一是传统目标检测,二是基于深度学习的目标检测。由于软硬件的不成熟,目标检测跟踪领域最早提出来的传统检测算法虽然在现在看来还有一定的局限性,但我们不可以忽视这些算法对目标匹配跟踪领域所做出的贡献,并且一些经典的算法至今还被我们使用。有些优秀的传统方法具有代表性,如光流法、粒子滤波、模板匹配法、级联分类

器等方法。

(1) 光流法^[14]起源于 1950 年 Gibsion 的理论, 它描述了目标、场景或摄像机在连续图像帧中的运动导致的物体位移。这种位移表现为物体在观测平面上像素的“瞬时速度”。根据图像序列中像素的时间变化和帧间相关性确定像素位置的运动情况。

(2) 粒子滤波^[15]根据蒙特卡洛思想, 使用粒子集来表示概率, 适用于状态空间模型。初始化时提取目标特征, 搜索阶段在图像范围内均匀或高斯分布采样粒子, 计算粒子与目标相似度, 最终相似度最高位置预测为目标位置。

(3) 模板匹配^[16]核心思想是在给定的目标图像中搜索与模板子图像相似度最高的区域。模板匹配算法会按照滑窗的思路, 不断地移动模板图像, 并计算其与图像中对应区域的匹配度。通过均方差、归一化互相关等相似性度量方法来计算匹配度。

(4) 级联分类器^[17]采用了一种多阶段、逐步精细化的策略来进行目标检测。它通常由多个分类器组成, 每个分类器都经过训练, 用于在特定阶段对图像中的目标进行识别。

近些年来, 深度学习因为其卓越的特征建模能力, 使得它在目标检测跟踪领域取得很大的成绩。2012 年深度学习首次用于 ImageNet 比赛, 取得惊人效果后飞速发展, 并行计算资源持续更新迭代, 社会上不断有新的大规模数据库和评测标准建立起来, 促进了深度学习的发展^[18]。由于数据、算力、技术的不断铺垫, 视觉在目标检测方面取得了巨大的进步, 尤其是检测的效率和精度。和传统手工设计特征的方法相比较, 深度学习方法可以将手工设计特征这一繁琐的过程省略, 可以自动学习目标物体具体区分度的深度特征。现在基于深度学习的算法有 Two stage 和 One stage 两大类^[19]。Two stage 就是先产生预选框, 再进行分类, 主要的流程就是: 首先特征提取, 然后生成 RP, 最后进行分类和回归。R-CNN、Fast R-CNN、Faster R-CNN、SPP-Net 等是这类算法的代表作。One stage 这类算法就不用先产生预选框, 可以直接在网络中提取物体特征来预测目标物体的分类并且定位到图像中的位置, 算法流程: 特征提取后进行分类和定位回归。常见的算法有 YOLO^[20-22]、SSD 等。

徐守坤等人^[23]针对安全帽检测时存在遮挡、尺寸、准确率低、目标小的问题提出改进的 Faster R-CNN 的检测方法提高了检测准确性; 闫建伟等人^[24]为解决

果园中刺梨果实的识别速度问题,提出了一种基于改进的 Faster R-CNN 刺梨果实识别方法,有较高的实时性和正确率;顾超越等人^[25]针对无人机巡检架空线路销钉缺陷检测的问题,提出了一种基于改进 Faster R-CNN 的算法,提升了工作效率,并且相较于其他算法有优越性;裴伟等人^[26]提出了基于特征融合的航拍目标检测算法,在小样本漏检和重复检测问题上有较好的表现;鞠默然等人^[27]针对图像中小目标检测率低、虚警率高等问题,提出了一种 YOLOv3 的改进方法,并将其应用于小目标的检测,结果表明改进后的 YOLOv3 算法能有效检测小目标,对小目标的召回率和检测的平均准确率均值都有明显的提升;Peng 等人^[28]针对无人机巡检图像检测输电线路缺陷问题,提出了一种基于 YOLOv5 的缺陷识别方法 EDF-YOLOv5,以提高检测精度。现在目标检测还存在对一些形变较为严重的物体检测困难,对小目标的检测效果较差,容易出现漏检或误检的情况;对数据获取困难,泛化能力较差等问题。

特征点定位方法一般基于图像处理技术和相应的特征情况来制定对应的定位方案。根据实际的应用场景,研究人员提出了不同的解决方法。胡亮等人^[29]提出了一种非合作航天器关键特征的双目视觉位姿测量方法,首先改进弧支持线段方法提取目标表面的星箭对接环,然后根据短线段特征运用直线检测方法提取隔热屏十字线。张超等人^[30]根据煤矿井下低照度、高粉尘、高水雾等情况,设计了红外 LED 光源标志的双目立体视觉悬臂式掘进机位姿测量方法,采用高斯拟合法实现光斑特征中心点的快速精确提取。周凡贵等人^[31]设计了一种编码合作标志点,利用图像处理技术对标志点的特征进行提取与解码,为飞行器位姿动态测量提供解决方案;束安等人^[32]利用遍历方法,通过面积和曲率约束提取特征突出的标志点来实现空间非合作目标的连续位姿测量。

1.2.2 多目标跟踪方法研究现状

MOT (Multiple Object Tracking) 即多目标跟踪,或者 MTT (Multiple Target Tracking),指开始不知道具体的目标数量,能够对视频中的汽车、飞机、动物、行人^[33]等多个目标赋予 ID 进行轨迹跟踪。多个目标各自 ID 不同,方便后续路径预测和精确查找等工作^[34]。多目标跟踪在智能监控、汽车自动驾驶以及人体行为识别等众多领域发挥作用,是计算机视觉的核心技术^[35]。在理想状态下,目标被赋予的 ID 应在整个跟踪过程中保持唯一^[36]。然而,多目标跟踪在实际场景应

用中不仅需要面对变形、遮挡、移动速度快、尺度变化大、光照不同、运动模糊等单目标跟踪中的问题，还需要面对相似目标间的相互干扰、轨迹初始化与终止等问题^[37,38]。因此，多目标跟踪在图像处理领域仍然是一个具有挑战性的研究方向，持续吸引着大量研究人员投入其中，进行长期而深入的研究^[39]。

多目标跟踪研究的核心问题是如何对视频前后帧进行数据关联建立目标的前后联系。前期的研究者主要把目标轮廓、外观特征等方面设计的更好来提高跟踪效果，这些方法限制比较多。相关专家与学者通过不懈的努力研究出效果更好的跟踪算法来适应不同的场景。Feng 等人^[40]提出的多目标行人检测算法，根据目标检测框中的行人外观特征进行前后帧的特征匹配进行跟踪；Yu 等人^[41]研究了性能更好的检测并且依据深度学习方法的外观特征来提升多目标跟踪的效果；Bae 等人^[42]通过研究目标移动的轨迹置信度，关联轨迹置信度的方式来解决多目标的跟踪问题；Milan 等人^[43]为了解决多目标的跟踪问题，研究出运用两个神经网络来分别解决预测和匹配问题；李星辰等人^[44]提出一种融合了 YOLOv3 的多目标跟踪算法，通过数据关联的方式完成跟踪需求。

虽然近些年来多目标跟踪算法一直在发展，但是大多数都是运用在诸如行人、车辆等常规物体的，对于工业应用基础的标识点跟踪相对较少。目前在对工业机器人的标识点上进行的多目标匹配跟踪研究还有很大发展空间。本课题基于此提出，旨在利用双目相机系统对于机器人结构上张贴的多个相同人工标识进行准确跟踪匹配，进而得出其像素坐标和空间坐标，可以获得机器人的位置状态和相关振动信息，准确估计机器人运动精度、进行性能分析。

1.3 本文的研究内容

准确测量机器人的运动是评价机器人动静态性能的重要组成部分。机器人的运动精度、承载能力、变形、振动、稳定性、结构模态等性能指标都可以根据机器人的运动位移来计算。因此，改进机器人运动测量方法，提高测量精度，丰富测量内容，受到了国内外学者的广泛关注。本文的主要研究工作是为实现机器人的全局结构运动测量的重要内容，通过准确检测跟踪张贴在机器人本体上的多个人工标识，实现机器人全局结构运动三维重建，为后续机器人运动精确测量、性能分析等做准备。本文的主要研究内容如下：

（1）设计并搭建了双目视觉多目标跟踪系统

根据机器人全局结构运动测量的实际需求,设计并搭建了双目视觉多目标跟踪系统装置,分析了本研究所涉及的多目标跟踪与定位的任务,推导了相机成像原理以及双目视觉三维运动测量模型,阐述了多目标跟踪与定位的工作过程。

(2) 提出了一种基于改进 YOLOv5 的小目标检测算法

针对机器人全局结构运动测量中,张贴在机器人结构上的人工标识小目标精确检测问题,研究基于改进 YOLOv5 的小目标检测算法,引入 BottleneckCSP 模块减少网络计算参数,加入大尺度特征融合结构提升小目标特征检测范围,融合 SE 注意力机制提升网络效率,实现多个小目标精确检测。

(3) 探寻了实现多目标跟踪与精确定位方法

针对多目标的跟踪与精确定位问题,运用卡尔曼滤波算法对目标运动轨迹进行预测,通过匈牙利算法进行前后运动帧之间目标的匹配,达到了 ID 信息不变的跟踪效果;最后通过对跟踪图像后处理,根据人工标识的直线特征,通过提取边缘的亚像素坐标结合线性拟合求交方法实现标识特征点的精确定位。

1.4 论文的组织结构

本文针对机器人全局运动结构测量中张贴在机器人结构上多个人工标识的跟踪与精确定位问题展开研究,在双目模型基础上,通过改进 YOLOv5 算法实现目标检测,利用卡尔曼滤波和匈牙利算法完成双目视角下的多目标跟踪,采用亚像素边缘检测和线性边缘拟合求交的方法实现目标的精确定位。

本论文共分为五个部分进行写作,论文各章节主要安排如下:

第一章:绪论。阐述了本论文选题的研究背景以及意义,讲述了目标检测与特征点定位以及多目标跟踪的国内外研究现状,详细说明了本文的主要研究内容与论文的组织结构和各章节的内容安排。

第二章:基于双目的同质多目标跟踪与定位理论基础介绍。面向机器人全局结构运动测量过程中多目标跟踪与定位的任务需求,搭建了双目视觉多目标跟踪、定位与测量系统装置;推导了相机成像模型,探讨了双目视觉三维重建与空间运动测量的理论基础和重要技术内容;阐述了双目视觉多目标跟踪与定位的过程以及关键的算法。

第三章:小目标检测算法研究。针对双目视觉测量过程中小目标检测的任务需求,通过引入 BottleneckCSP 模块并且加入大尺度特征融合结构,融合 SE 注

注意力机制，改进现有的 YOLOv5 模型结构，减少网络计算参数，提升小目标特征检测范围，提升网络效率，实现多小目标精确检测。

第四章：多目标跟踪与定位算法研究。在前一章的基础上，通过研究卡尔曼滤波跟踪算法对运动目标轨迹进行预测，运用匈牙利算法融合外观信息和运动信息进行目标匹配，实现多目标的跟踪。在分析整个运动过程中，初始帧到结尾帧的目标 ID 不变，即为跟踪成功。根据人工标识的直线特征，通过提取边缘的亚像素坐标结合线性拟合求交方法实现标识特征点的精确定位。

第五章：总结和展望。对本论文整体进行总结，分析研究有待改进的地方，并对未来的研究方向进行展望。

第2章 基于双目同质多目标跟踪与定位理论基础

双目立体视觉是模仿人眼的一种视觉成像技术，其原理是利用两个型号、参数等相同的摄像机从不同的视角去拍摄同一个目标场景。通过摄像头的感光组件可以得到两张二维平面数字图像，然后通过相关算法和数字图像处理计算同一空间点在左右图像上的像素坐标，通过两个像素坐标和相应的转换关系就可以得到三维空间的真实信息^[45,46]。本章主要介绍了双目同质多目标跟踪与定位的系统组成；多目标跟踪与定位的任务内容；双目视觉测量的基本原理；目标检测及跟踪相关的基础理论；实现多目标跟踪与定位的过程。

2.1 基于双目的同质多目标跟踪与定位系统

本文针对机器人全局结构运动双目测量过程中多个相同人工标识目标的跟踪与精确定位问题，进而实现机器人全局结构跟踪与精确定位，实现机器人的全局结构运动三维重建，为后续机器人运动精确测量、性能分析等做准备。

机器人是现代智能制造的代表和制造业升级发展的焦点，其研发水平、应用水平等成为衡量国家科技水平的主要标志，是一个国家科技水平的集中体现。工业机器人的应用可以提高生产效率、改变传统工业人力为主的模式、降低成本，因而被广泛使用。一般工业机器人是由连杆和移动关节组成，在机器人运行过程中如何保证其运动精度一直是研究重点。测量机器人的运动方式很多，传统的接触式测量是在工业机器人上安装传感器、非接触式用激光干涉进行测量。随着计算机技术等视觉传感技术的发展，视觉测量展现出强大的优越性。

视觉测量方法的设备结构更加灵活，可以满足全局、多维度的测量需求。通过两个甚至多个视觉传感器可以获得工业机器人在空间中的三维运动信息、变形信息等，可以测量到更多、更丰富的信息。双目视觉可以获取工业机器人上多点同步运动、变形信息。

2.1.1 视觉系统构成

研究基于双目视觉系统下工业机器人的空间运动信息和力学行为可视化检测。通过对高速摄像机拍摄的视频信号进行连续多帧的分析，实现工业机器人全局运动测量，进而分析工业机器人运动特性和运动精度。为实现此目的需要从双目摄像机拍摄的二维图像上进行工业机器人的三维重建，进而获取工业机器人各

结构的运动轨迹和形变等信息。所以，准确的、高精度的三维重建直接影响后续的研究过程^[47]。为了简化三维重建的过程和减低难度，在三维重建中用一系列点状特征群来代替工业机器人，利用这些特征点的运动和重建信息来反应工业机器人的真实运动信息。通过对这些特征点群的提取，利用双目视觉系统数学原理，构建出它们的三维坐标，进而重构出工业机器人的三维模型和运动信息。

如图 2-1 所示双目系统主要包括：两台高速摄像机、工业机器人运动控制系统、视频处理系统、辅助系统。



图 2-1 双目同质多目标跟踪与定位系统示意图

2.1.2 同质多目标跟踪与定位任务分析

在实际过程中为了简化构建出工业机器人的三维信息，通过把工业机器人本体上张贴或喷涂上特定的多目标特征点集，通过跟踪和定位这些多目标特征点集的运动信息来反应工业机器人的运动信息。针对不同场景和应用情况，多目标特征点的选取一般不同，例如张贴圆形特征、矩形标志、带编码的标志、在特定位置喷涂散斑、在一些黑暗条件下安装发光标志等。如图 2-2 所示为前人所用的标识多目标。

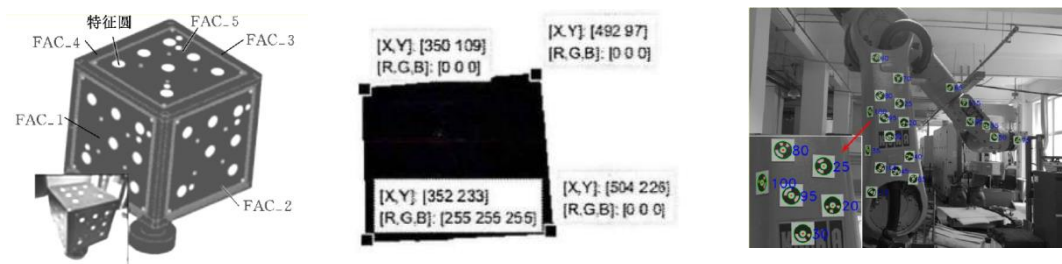


图 2-2 标识多目标

这些多目标主要作用是提供位置和边缘的信息，便于摄像机进行识别与定位，并为后续跟踪定位提供帮助。本研究通过研究实际应用环境和条件，自主设计一种圆形特征标识，标识中等分为四份、黑白相间，在圆形标识的中心有明确的黑白相交特征标志，方便后期图像处理的中心点位。标识如图 2-3 所示。

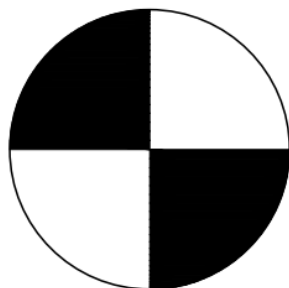


图 2-3 本文标识

通过把这些标识张贴在所需测量的工业机器人本体上，后续应用相应设计的检测跟踪算法对这些标识多目标进行检测跟踪，运用图像处理定位算法计算出目标的中心像素坐标后通过双目视觉原理即可测量工业机器人的位移、变形等重要信息，为研究工业机器人做基础。传统的研究基于图像处理技术进行目标的检测，例如模板匹配、特征点检测等方式，这些方法运行时间长、处理速度慢，完全达不到实时检测的要求。如何设计出快速准确的多目标检测方法是后期跟踪定位的第一步，处理速度快、检测准确率高成为本研究的重中之重。如图 2-4 所示为本研究张贴在工业机器人本体上的标识多目标示意图。



图 2-4 张贴标识多目标示意图

2.2 双目视觉测量技术

双目视觉模型是利用两个型号相同的摄像机在不同角度针对同一物体进行拍摄，然后通过双目视觉模型准确还原三维坐标信息。本节从摄像机的成像模型开始介绍，呈现二维图像中的一点是如何通过双目视觉系统、坐标系转换、三角重建等推算出三维坐标信息的。

2.2.1 摄像机成像几何模型

摄像机成像的基本模型就是俗称的“小孔成像”。该模型是视觉领域的基础模型，在光学研究中被称为是中心投影，根据光沿直线传播理论，清晰地解释了相机成像过程，成像平面所成的像是由现实物体表面反射的光线经过小孔投射形成的。如图 2-5 所示，在空间中的点 P ，经过投影过程后，在成像平面上对应的点是 P' 。此处，成像平面与小孔（光心）平面之间的距离叫做焦距，表示为 f 。

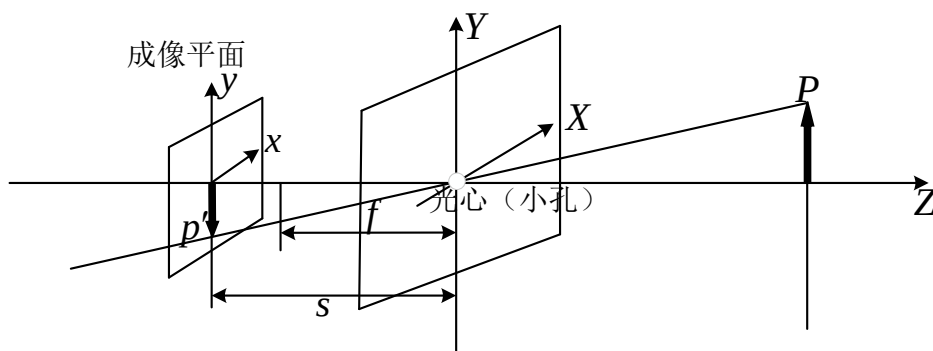


图 2-5 小孔成像模型

以小孔成像为基础，通过像素坐标系、图像坐标系、相机坐标系以及世界坐标系建立四个坐标系之间的关系来描述相机的成像过程。

（1）像素坐标系

像素坐标系是用于描述图像在计算机内存中不同像素之间相对位置关系的参考体系，在图像的左上角设定为坐标原点 O_p 来建立像素坐标系， u 轴平行于图像坐标系的 x 轴， v 轴平行于图像坐标系的 y 轴，而且它们的方向相同。像素坐标系是为了描述物体在成像后像点在数字图像（即照片）上的具体坐标位置而被引入的。像素坐标系以像素数目为单位来反应从相机内部实际读取到的信息。如图 2-6 所示为像素坐标系与图像坐标系位置关系。

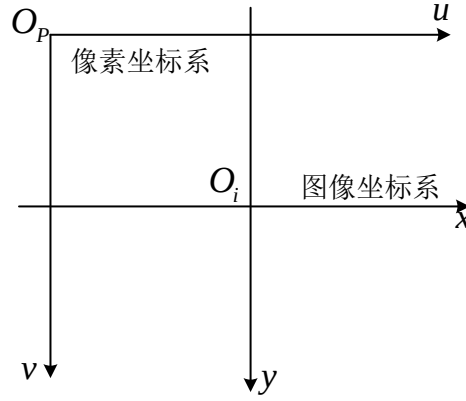


图 2-6 图像和像素坐标

(2) 图像坐标系

图像坐标系是一个采用物理长度来量化图像中不同点之间相对位置关系的参照体系。图像坐标系的位置在相机的感光片上，其中，光轴在感光片上的垂直投影点被定义为坐标原点 O_i ，从而确立了图像上各点的精确位置，图像坐标系中的 x 轴与图像边缘平行，向右为正方向； y 轴则垂直于 x 轴，向下为正方向。图像坐标系能够描述物体在成像过程中从相机坐标系到图像坐标系的投影透射关系，方便进一步获取像素坐标系下的坐标信息。 $P=(x,y,1)$ ，单位 mm。通过假设相机主点在像素坐标系中的坐标为 $O_i(u_0, v_0)$ 推导像素坐标系与图像坐标系之间的转换关系，相机主点一般位于图像的中心位置，而单个像素在 x 轴方向的物理尺寸为 d_x ，在 y 轴方向的物理尺寸为 d_y ，基于此得出转换关系为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-1)$$

(3) 相机坐标系

相机坐标系作为沟通世界坐标系与图像坐标系和像素坐标系之间的桥梁，从相机的视角描述物体位置，起到了关键的中间转换作用， $P_c(X_c, Y_c, Z_c)$ 。相机坐标系以相机的光心 O_c 处为原点，横坐标与成像平面的 x 轴平行，纵坐标与成像平面的 y 轴平行，而 z 轴则与相机的光轴重合，且垂直于图像平面。这种设定使得相机坐标系能够准确描述物体在相机视角下的三维位置。相应的，图像坐标系与相机坐标系之间的数学转换关系为：

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2-2)$$

(4) 世界坐标系

世界坐标系是用于描述目标物体位置的参考体系,其坐标轴包括 (X_w, Y_w, Z_w) ,在实际应用中,世界坐标系的放置可以根据运算的便利性和实际需求进行灵活调整。在双目视觉系统中,世界坐标系扮演着多重角色。首先,它在标定过程中用于确定标定物的精确位置。其次,世界坐标系是双目视觉系统中能够描述两个摄像机相对位置关系,进而求出相机之间的相对位置和姿态的核心参考系。最后,世界坐标系还作为存储容器,用于存放通过重建过程得到的三维坐标,从而实现对物体三维结构的完整描述。世界坐标系是我们在视觉运算中引入的第一个参考系,用于纳入所观测到的物体。相机坐标系和世界坐标系通过欧式变换原理进行转换:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2-3)$$

在上式中, R 为旋转矩阵,大小是 3×3 , t 为平移向量,大小是 3×1 。

点 $P(X_w, Y_w, Z_w)$ 在世界坐标系中与点 $q(u, v)$ 在像素坐标系中的转换关系可以表述为:

$$\begin{aligned} s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \alpha_x & 0 & u_0 & 0 \\ 0 & \alpha_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \underbrace{A[R|t]}_M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \end{aligned} \quad (2-4)$$

上式 s 是比例因子,与前面公式中 Z_c 等价;相机尺度因子,有时也被称为归一化焦距: $\alpha_x = f/d_x$ 、 $\alpha_y = f/d_y$; A 是相机的内参矩阵,数值取决于尺度因子 α_x 、 α_y 和像素坐标 u_0 、 v_0 ; R 、 t 是相机的外参矩阵,由相机和世界坐标系的相对位置决

定；内参矩阵与外参矩阵合并就得到相机的投影矩阵 M ，即为三维空间坐标信息到二维图像信息的过程转化，就是相机的成像模型。坐标系转换关系如图 2-7 所示。

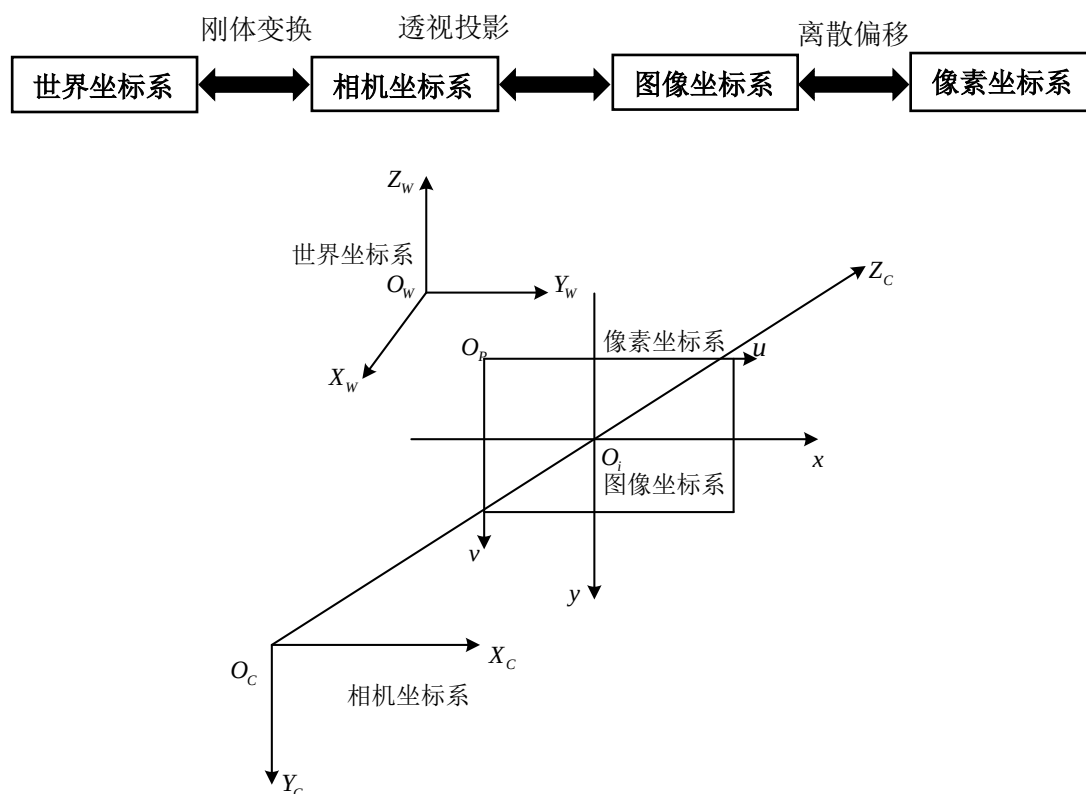


图 2-7 坐标系转换关系

2.2.2 双目视觉测量基础

双目视觉系统根据两个摄像机的空间位置结构差异，一般分为两种类型：平行式双目相机和汇聚式双目相机。平行式双目相机中，两个摄像机的光轴相互平行，便于进行水平视差计算；而汇聚式双目相机则允许两个摄像机的光轴相交于一点，适用于更广泛的场景和视角。实际应用中需根据具体需求以及平衡两种配置的优缺点进行具体选择。

平行式双目相机中，左右相机的光轴是平行的，也就是说，左右相机的视点在同一水平线上。这意味着相机的基线距离保持不变，不随深度变化而变化。在平行式双目相机中，左右相机的光轴平行，并且通过调节其水平位置来控制基线距离。两个相机在水平方向上对准，成像平面也一致。平行式双目相机的配置比较简单，不需要对两个目镜进行调焦调整。相机的基线距离保持不变，适用于对深度范围较大的场景进行精确的深度测量。平行式双目相机，作为双目视觉系统

的一个经典模型，已被广泛采纳为标准范式。如图 2-8 所示，生动地描绘了平行式双目相机的系统架构，直观展示了其工作原理和组成部件，为我们深入理解这一视觉系统提供了有力的视觉支持。

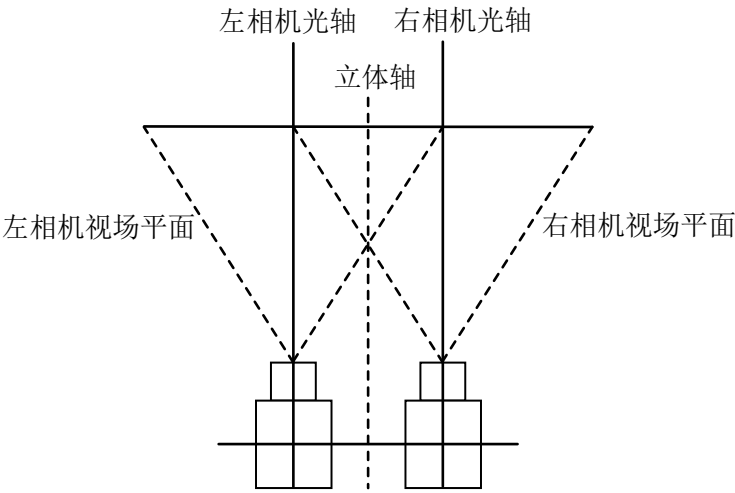


图 2-8 平行式双目相机系统

汇聚式双目相机中，左右相机的光轴是会汇聚到一个焦点上的，也就是说，左右相机的视点会聚在一个点上。这意味着相机的基线距离会随深度变化而变化。在汇聚式双目相机中，左右相机的光轴汇聚到一个焦点上，形成一个夹角，该夹角控制了基线距离。两个相机在垂直方向上对准，焦点与成像平面重合。汇聚式双目相机基线距离随着深度变化而变化，适用于近距离或纵向较大的场景。可以利用视差信息来估计物体的深度。汇聚式双目相机系统如图 2-9 所示。

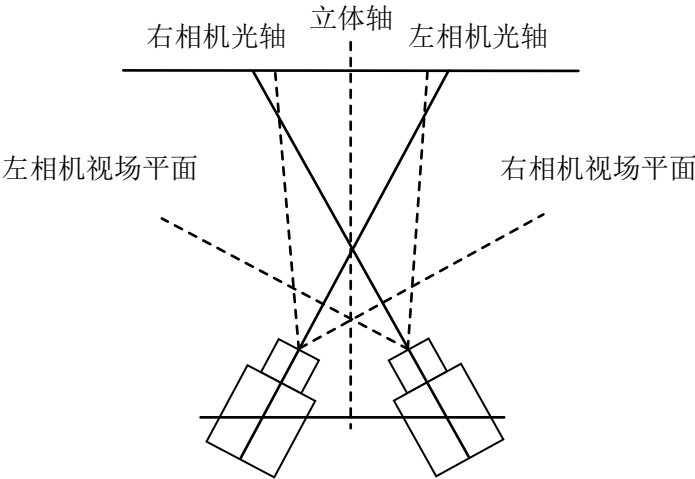


图 2-9 汇聚式双目相机系统

这两种双目相机配置适用于不同的应用场景，配置方式各有优劣。平行式双目相机适用于需要在深度范围较大的场景中进行精确深度测量的任务，例如室外

距离测量和立体视觉导航。汇聚式双目相机特别适用于那些需要精确测量近距离物体或深度变化较小的场景。比如，在近距离物体检测任务中，它能够提供清晰且准确的图像数据，为物体的识别与定位提供有力支持。此外，这款相机的摆放位置十分灵活，不受到严格的空问约束，这使得两个相机能够拥有较大的共同视野，从而捕捉到更广泛的场景细节，进一步提升了其在各种应用场景中的实用性和性能。通过对比和实验需求，本研究最终选择了汇聚式双目模型。

双目视觉系统是模拟人类的双眼来感知三维世界的。根据三角重建的基本原理，当三维空间中的点 P 在左右两个相机的相平面上分别形成投影点 q_l 和 q_r 时，已经获得了两个相机的内部参数，并且明确知道它们之间的相对位置关系，可以利用两图像上的对应点 q_l 和 q_r 求解出该空间点 P 的三维坐标。如图 2-10 为三角重建示意图。

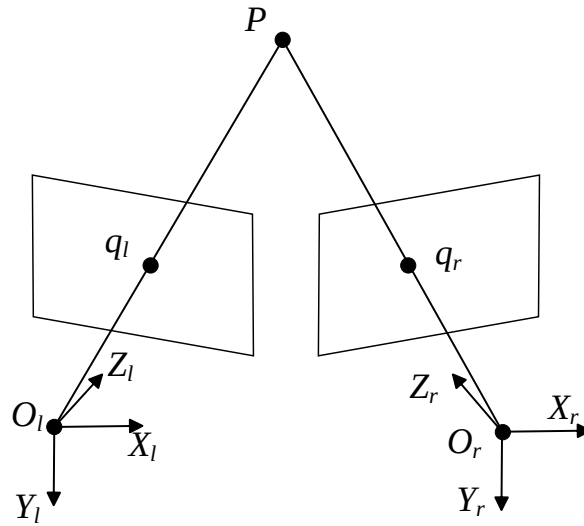


图 2-10 三角重建示意图

通过设定世界坐标系与左相机坐标系重合，可得：

$$s_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_{lx} & 0 & u_{l0} \\ 0 & \alpha_{ly} & v_{l0} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2-5)$$

$$s_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_{rx} & 0 & u_{r0} \\ 0 & \alpha_{ry} & v_{r0} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} \quad (2-6)$$

上式各参数含义为： s_l 、 s_r 是比例因子； α_{lx} 、 α_{ly} 和 α_{rx} 、 α_{ry} 代表左右相机归一化焦距； (u_{l0}, v_{l0}) 、 (u_{r0}, v_{r0}) 代表左右相机主点坐标； (X, Y, Z) 、 (X_r, Y_r, Z_r) 代表空

间点 p 的左右相机坐标；左右图像中的投影点 q_l 、 q_r 对应的像素坐标是 (u_l, v_l) 、 (u_r, v_r) ；左右相机的坐标系 $O_l-X_lY_lZ_l$ 和 $O_r-X_rY_rZ_r$ 相对位置关系可以通过欧式变换转化：

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \quad (2-7)$$

上式 R 代表旋转矩阵， t 代表平移向量，是由相机标定得出；通过以上公式，可以求得左右相机的像点 q_l 和 q_r 的转换关系：

$$s \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_{rx} & 0 & u_{r0} \\ 0 & \alpha_{ry} & v_{r0} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} Z(u_l - u_{l0}) / \alpha_{lx} \\ Z(u_l - u_{l0}) / \alpha_{ly} \\ Z \\ 1 \end{bmatrix} \quad (2-8)$$

由此，可以计算得到三维空间点 p 的坐标为：

$$\begin{cases} X = Z \frac{u_l - u_{l0}}{\alpha_{lx}} \\ Y = Z \frac{u_l - u_{l0}}{\alpha_{ly}} \\ Z = \frac{t_x - \frac{u_r - u_{r0}}{\alpha_{rx}} t_z}{\frac{u_r - u_{r0}}{\alpha_{rx}} (Ar_{31} + Br_{32} + r_{33}) - (Ar_{11} + Br_{12} + r_{13})} \end{cases} \quad (2-9)$$

通过以上分析可知，要求得一点的空间三维坐标需要知道：相机的内部参数和一对匹配的像点像素坐标，以及左右相机的位置参数。准确获取匹配点的像素坐标值对求得一点的空间三维坐标有重要影响。因此，研究出如何准确快速检测标识点的方法，对与后期机器人全局结构测量有决定性作用。通过多目标检测、跟踪，最后定位出标识点中心坐标来代替工业机器人关键结构，达到机器人全局结构运动测量效果。

2.3 多目标检测跟踪与定位过程

通过上述研究基础了解到，想要准确研究工业机器人的运动信息，首先得检测出工业机器人本体上的人工标识，实时准确的标识检测算法是处理这一问题的关键所在。由于传统的目标检测方法在速度和精度等方面达不到要求，所以基于

深度学习的目标检测方法成为主流。目标跟踪是在检测的基础上对目标赋予一个唯一 ID，并且在后续的视频中能够保持目标 ID 不变。最后运用图像处理技术，对检测跟踪的目标进行处理，得到标识的中心点坐标。深度神经网络是深度学习的基础内容，本节将介绍神经网络相关知识以及深度学习的一些模型。为准确得到标识坐标，从相机拍摄到最后结果需要经过多部操作，处理过程示意图如图 2-11 所示。

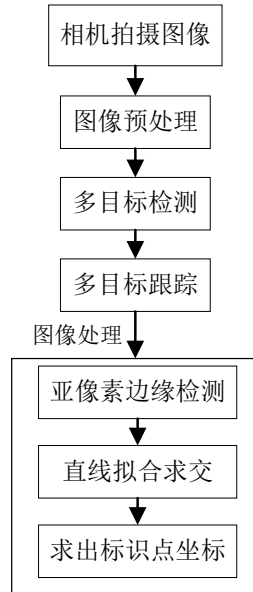


图 2-11 处理过程

2.3.1 图像预处理

计算机视觉的本质是对拍摄的图像或视频进行相应的处理，通过算法来得到有用信息。图像预处理是为了提高图像数据的质量和适应性而进行的一系列操作。在实际的拍摄过程中会因为实际场景、光照、相机质量等各种因素导致图像质量不佳，影响后期的算法进行信息提取。通过图像预处理，可以对图像进行一些常见的操作，以消除图像中的噪声、调整图像的尺度和对比度、增强图像中的关键特征等，以便计算机视觉算法可以更好地理解和处理图像，获得更好的图像质量和更具信息量的数据。

图像滤波是图像预处理的方式之一。在设备进行拍摄的过程中，会受到外界暗电流噪声、光子噪声等环境的干扰而产生椒盐噪声污染图像。椒盐噪声会模糊图像中的细节和信息，使图像看起来不清晰。在图像的边缘检测、纹理和特征点提取等方面造成影响。

在去除图像噪声的方式中，它是去除图片中不重要信息的从而凸显重要信息的方法。在尽可能保留图像细节特征的前提下，有效抑制椒盐噪声。这一处理步骤的效果优劣，将直接决定后续目标检测的精确性。图 2-12 展示了处理前后的对比效果。



图 2-12 图像处理前后对比图

2.3.2 神经网络基本原理

深度学习的核心目的在于构建并模拟人类大脑的分析机制，以形成神经网络。它借鉴人类大脑的工作方式来解析各类数据，包括但不限于图像、声音和文本。简而言之就是，深度学习就是如何让机器能够自己掌握学习能力。

人工神经网络(Artificial Neural Network, 即 ANN)起源于现代神经科学，它建立简单的数学模型模拟人脑的神经元网络形式来进行信息处理。在上世纪由心理学家 W.S.McCulloch 和数理逻辑学家 W.Pitts 提出著名的 M-P 模型。人脑通过数百亿个神经元细胞前方的树突来接收信号。神经元的工作方式主要依靠判断树突所受的刺激大小，刺激大会将信号传递到下一级，刺激小则终止。传递的方式主要依靠神经递质和电信号。神经元结构如图 2-13 所示。

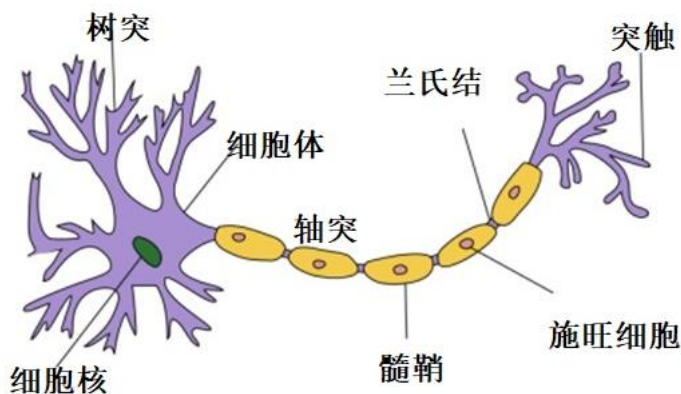


图 2-13 神经元结构图

神经元模型是神经网络中的最基本单元，神经元都是相互连接在整个神经网络中的，当神经元兴奋时就会通过化学信号传递到下一个神经元。简化为数学关

系如下, 神经元数学模型如图 2-14 所示。

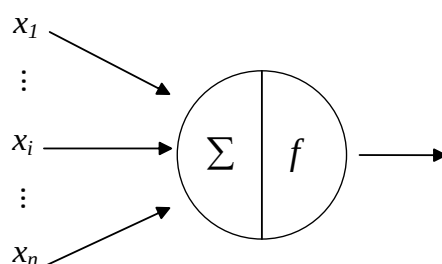


图 2-14 神经元数学模型

简化神经传递模型如下式所示, $x_1 \cdots x_n$ 为输入的信号, 最后的值通过 sigmoid 等函数进行激活。数值大完成激活, 传递到下一级, 数值小即终止。

$$y = \omega_1 x_1 + \omega_2 x_2 \cdots + \omega_n x_n + b \quad (2-10)$$

输入层、隐藏层和输出层构成了一个完整的神经网络, 每层中包含多个神经元。输入层接受外部信号输入, 隐藏层执行计算过程, 输出层产生输出得到结果。如图 2-15 所示为一个完整的神经网络。

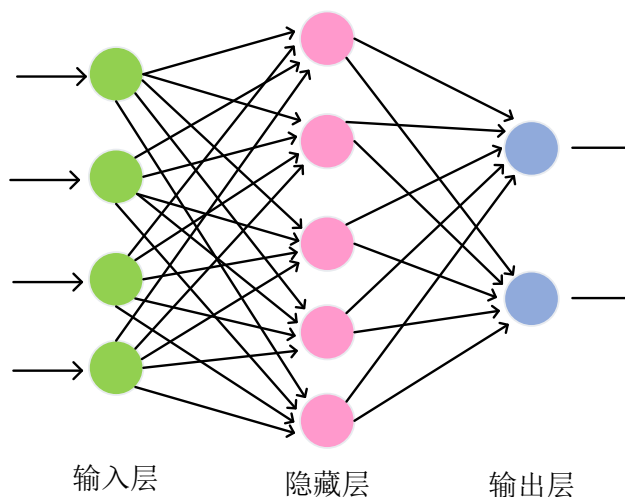


图 2-15 神经网络模型

卷积神经网络 (Convolutional Neural Networks CNN) 作为深度学习的代表算法, 它包含卷积计算且有深度结构的前馈神经网络 (Feedforward Neural Networks)。卷积神经网络作为神经网络的一种特殊类型, 在神经网络的基础上引入了卷积层和池化层, 用于处理如图像数据之类的网格结构的数据, 因而广泛应用于目标检测、图像分类和语义分割等任务。

输入层、卷积层、激活函数、池化层、全连接层和输出层构成了基本的卷积神经网络。卷积神经网络, 把全连接变成一个个小的局部连接, 多层网络结构特征提取, 实现了高效的特征提取。通过模拟人眼的一层视神经进行卷积、池化、

激活三个步骤得到图像的细节特征。然后，第二层的输入以第一层输出开始行第二次卷积、池化、激活，得到范围更大的图像特征，以此类推。

卷积层的功能就是提取输入图像的特征，是卷积神经网络的主要层。通过卷积核在图像表面一定规则滑动，该过程为像素的加权求和，即对应像素相乘再相加。通过卷积操作可以得到一张基于原图像的特征图。卷积操作如下图 2-16 所示。

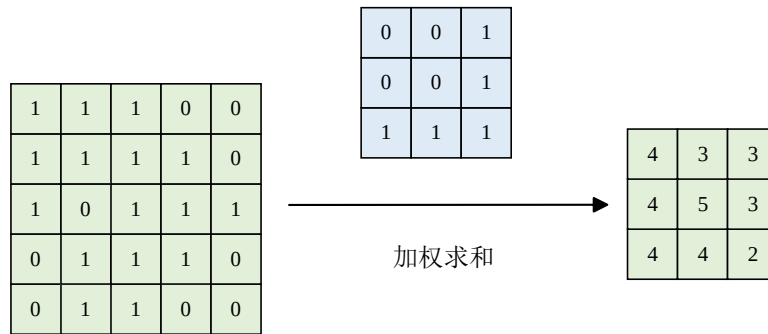


图 2-16 卷积操作

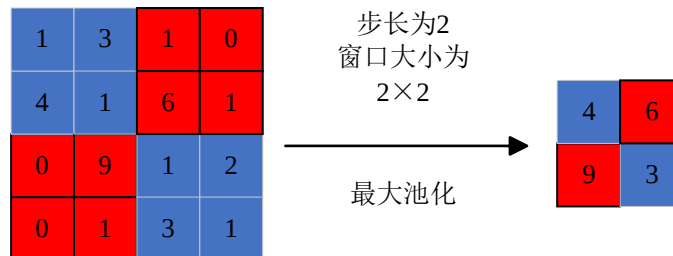


图 2-17 最大池化

激活层作用是引入非线性特性，使得神经网络能够学习更为复杂的模式和特征，Sigmoid、ReLU、Tanh 等都是常见的激活函数。

2.3.3 目标检测跟踪模型

深度学习经过多年的发展，出现了一批优秀的算法模型。根据检测形式的不同可以分为两类：两阶段检测模型和单阶段的检测模型。

两阶段模型代表作有 R-CNN、Fast R-CNN、Faster R-CNN 等，一阶段模型代表作有 YOLO、SSD 等。因为一阶段检测模型在处理速度上具有实时性的优

点，所以本文主要研究一阶段检测模型中的 YOLO 模型。

YOLO (You Only Look Once) 算法模型是由 Redmon 等人在 2015 年提出，它的独特之处在于将目标检测任务视为一个单一的回归问题，通过在图像的全局范围内同时预测目标的类别和位置来实现实时目标检测。其核心思想是将图像划分为 $S \times S$ 的网格，每个网格负责预测一个或多个边界框以及对应的目标类别概率。这样，每个网格单元的预测信息包括边界框的位置（坐标值）和目标的类别概率。

YOLO 使用综合的损失函数，包括定位误差、目标存在与否的二值分类误差以及目标类别的分类误差。这有助于模型在训练过程中平衡各个任务。其损失函数如下：

$$\begin{aligned}
 & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B II_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\
 & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B II_{ij}^{\text{obj}} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\
 & + \sum_{i=0}^{S^2} \sum_{j=0}^B II_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \\
 & + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B II_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 \\
 & + \sum_{i=0}^{S^2} II_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2
 \end{aligned} \tag{2-11}$$

2.4 本章小结

本章通过分析机器人全局结构运动测量过程中多目标跟踪与定位的任务需求，明确跟踪定位的多目标类型，搭建了双目视觉多目标跟踪、定位与测量系统装置；推导了相机成像模型，探讨了双目视觉的基本类型以及三维重建与空间运动测量的理论基础和重要技术内容；阐述了双目视觉多目标跟踪与定位的过程以及关键的算法。

第3章 基于改进 YOLOv5 的小目标检测

在目标检测与定位方法的研究中，发现传统的模板匹配、SIFT 算法虽然能实现单个目标的检测和定位。但是这两种方法的局限性太大，只能够在一些实验环境相对较好地情况下实现，而且目前只能实现单个目标检测。当涉及多个目标需要同时检测时，并且目标旋转偏移角度过大、形变、亮度不均等情况发生时，目标检测成功率大大下降。同时在目标过小时，由于目标可以利用的特征较少，导致检测率不高，检测效果差。在进行相关调研后发现，基于深度学习的目标检测方法可以实现同质多目标的同时检测，并且具有更高的鲁棒性^[48]。

深度学习是机器学习技术分支之一，通过搭建深层的人工神经网络来进行知识的学习。基于深度学习的目标检测流程如下所述，主要分为以下步骤：

(1) 数据准备：依据检测任务类型，准备相应的数据，同时对准备好的数据进行预处理来提其高平衡性和显著性；

(2) 定义神经网络模型：通过确定模型结构和神经元之间的连接方式来确定数据的计算规则；

(3) 定义损失函数与优化方法：确定误差的计算规则，选择合适的优化方式；

(4) 模型训练：将准备好的用于模型的迭代训练，获得最优的模型结构；

(5) 性能检测：用训练好的模型结构对实际检测目标进行验证并进行性能评级。

3.1 YOLO 概述

YOLO (You Only Look Once) 是一种基于深度神经网络的目标检测算法。从算法的名称可以看出，YOLO 的最大特点是处理速度快,可以进行实时检测处理^[49,50]。以前目标检测任务主要由两个部分组成：1) 对物体进行分类；2) 确定物体在图像中的具体位置。基于回归的 YOLO 目标检测算法只有一个步骤，直接在图像级别上回归边界框和类别信息，实现了实时的目标检测。

本文使用的是 YOLOv5 模型，是 YOLO 系列算法的第五代版本，因其整体性能的大幅提升和实时检测效果能够满足更多的要求，因此在工业检测中有着更多的使用场景，例如缺陷检测、目标识别等领域。

3.2 YOLOv5 算法

在 YOLOv5 版本中又细分为四个型号，分别是 YOLOv5x、YOLOv5l、YOLOv5m、YOLOv5s，其网络结构的宽度和深度不断减小。根据本论文的研究场景需求，考虑实际的目标运动情况选择了模型结构相对较小的 YOLOv5s。在实际工作中由于 YOLOv5s 消耗计算资源最少，处理速度最快，因此是最合适的选择。其模型结构如图 3-1 所示。

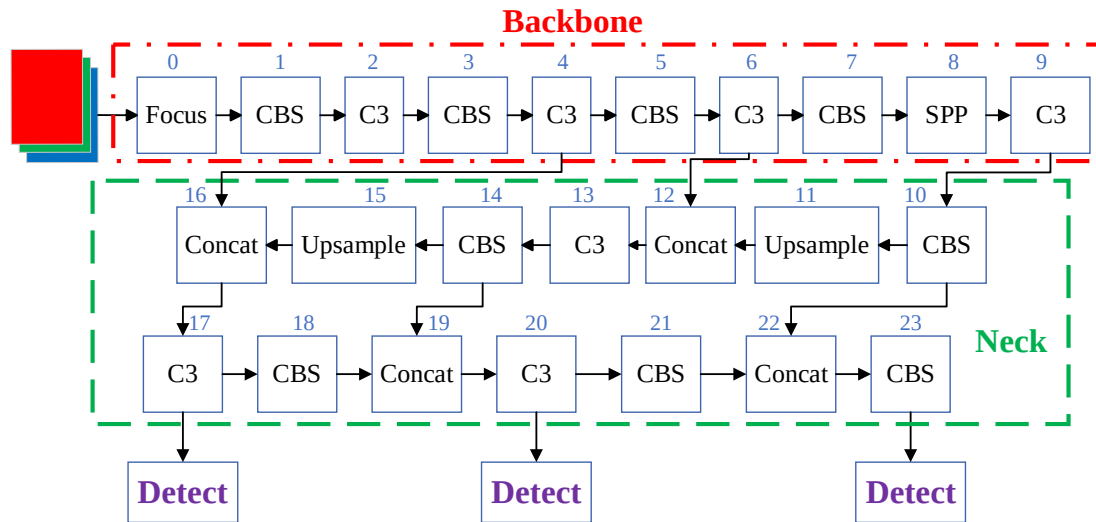


图 3-1 模型结构图

YOLOv5 目标检测流程大致为：

第一步，输入端把待检测的图像进行尺度变化操作，得到 640×640 像素大小的图片；

第二步，主干网络 CSPDarknet53 进行特征提取，逐渐提取图像低级的纹理信息到高级的语义信息等不同层级的特征；颈部网络采用 FPN+PAN 结构融合不同尺度的特征信息来处理不同大小和位置的目标；

第三步，输出检测到的目标类别和概率。

3.2.1 网络结构

YOLOv5 算法主要分为以下几大结构：输入端（Input）、骨干网络（Backbone）、颈部网络（Neck）、检测层（Detect）。

（1）输入端（Input）

YOLOv5 主要由 Mosaic 图像增强、自适应锚框计算和自适应图片缩放组成输入端。

Mosaic 图像增强是通过组合多个不同的图片来生成新图片提供训练的数据增强方法。它通过随机挑选四张不同的图像，然后把这四张图像拼接成一张大的图像，并且四个小图像的位置和大小都是随机的。如图 3-2 所示为本文对标识图像进行的 Mosaic 图像增强操作。

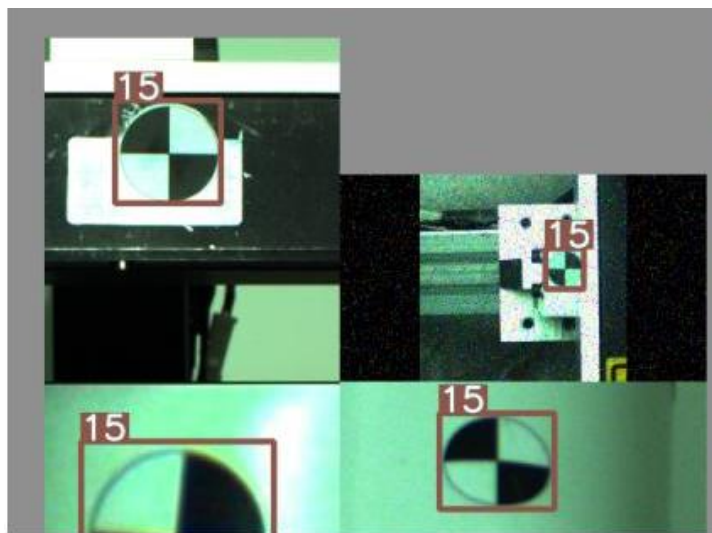


图 3-2 Mosaic 图像增强

自适应锚框计算可以很好地解决传统目标检测算法中锚框大小和宽高比等参数需要手动设置的问题。其核心思想是运用学习的方法来自动计算出最适宜的锚框参数，以此提高目标检测的精度和鲁棒性。

自适应图片缩放是一种基于目标尺度的图像缩放方式，它可以把输入的图像自适应地缩放到合适的尺寸以适应不同的尺度目标检测。

(2) 骨干网络 (Backbone)

骨干网络是目标检测任务中的核心组成部分，它负责从输入图像中提取特征、生成多尺度特征图、提供语境信息以及增加模型的感知能力，从而帮助模型高效准确地检测图像中的目标物体。骨干网络中含有一些主要模块，如 Focus 结构、CBS 结构、C3 结构等。

Focus 结构的作用是在图片进入骨干网络之前对图片进行下采样操作，将输入的图片分成四个子图，并将这个四个子图在通道维度上进行拼接，拼接完成后图像的尺寸是原图像的一半，并且通道数翻倍。Focus 结构通过对输入特征图实行下采样操作，达到减少计算量和参数数量的效果，并且保留了输入特征图中相关重要信息，有助于提高模型的特征提取能力。

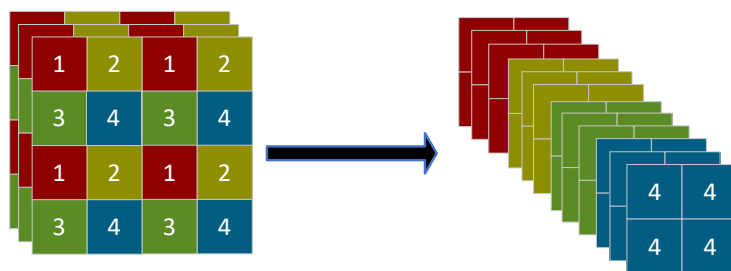


图 3-3 Focus 结构示意图

CBS 结构由卷积操作 (Conv), 批量标准化操作 (BN) 以及激活函数 (SiLU) 三个部分组成, 结构如图 3-4 所示。



图 3-4 CBS 结构示意图

C3 结构为更好地提取图像特征, 通过在输入特征图的不同尺度上进行卷积运算。其结构原理是通过对输入特征图的不同尺度进行卷积运算来提取不同尺度下的图像特征, 从而更好地适应不同大小和比例的目标物体。

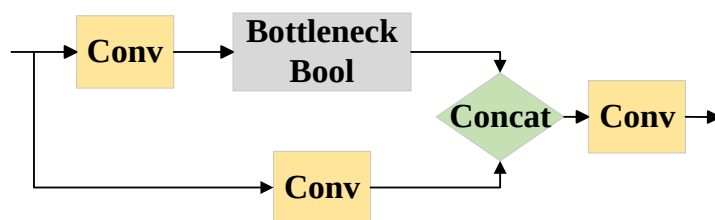


图 3-5 C3 结构示意图

(3) 颈部网络 (Neck)

颈部网络采用 FPN+PAN 结构, 通过上采样和下采样操作把不同层次的特征图融合在一起, 以此生成多尺度特征金字塔。FPN 自底向顶把语义特征 (纹理、颜色、类别) 进行传递; PAN 将目标类别的位置信息、形状信息通过下采样的方式对定位特征进行传递。进一步整合和处理骨干网络提取的特征以及引入多尺度的信息, 从而提高目标检测的性能。

(4) 检测层 (Detect)

检测层对输入的图像进行预测, 在这个阶段, 模型会依据输入图像的特征预测出图中目标的位置、类别以及置信度等信息。

3.2.2 评价指标

(1) 准确率 (Precision) 表示分类的准确性:

$$Precision = \frac{TP}{TP + FP} \quad (3-1)$$

(2) 召回率(Recall): 召回率的含义是表示(正类分类正确)与(正类分类正确和把正类判断成负类)的比值。召回率表示没有检测到的物体的比例, 如果图片中的所有目标都被检测出来, 召回率就高; 只检测出部分目标则召回率低。

$$Recall = \frac{TP}{TP + FN} \quad (3-2)$$

式中, TP 、 FP 、 FN 、 T 、 F 、 P 、 N 含义;

T 代表 True 正确的判断;

F 代表 False 错误的判断;

P 代表 Positives 正类(表示需要检测的目标);

N 代表 Negatives 负类;

TP 代表 True Positives(表示判断正确, 把正类判断成正类);

FP 代表 False Positives(表示判断错误, 把负类判断成正类);

FN 代表 False Negatives(表示判断错误, 把正类判断成正类);

TN 代表 True Negatives(表示判断正确, 把负类判断成负类)。

(3) 平均精度 (AP, Average Precision)

平均精度用来衡量评估目标检测算法的性能指标。平均精度主要用于衡量模型在不同类别的目标上的准确性和召回率, 使得评估更加准确。AP 的取值范围在 0 和 1 之间, 值越高表示模型在目标检测任务中的性能越好。AP 的优点之一是它对不同类别之间的不平衡进行了处理, 因为它考虑了每个类别的独立性。

(4) 整体平均精度均值 (mAP, mean Average Precision)

整体平均精度均值是目标检测任务中常用的性能度量指标。对不同类别的平均精度 (AP) 取平均得到的值即为整体平均精度均值, 用来评估目标检测模型在多个类别上的综合性能。整体平均精度均值的取值范围在 0 和 1 之间, 数值越高代表模型在目标检测任务中的表现性能越好。mAP 的优势在于它能够反映出模型对多个类别的整体性能, 而不仅仅是对某个特定类别的性能。计算为单个类别的精确度的总和除以整体类别数, 从整体上评价目标的检测效果。

(5) FPS: 画面每秒传输帧数

一般定义为视频中每一秒含有的画面数。(FPS 表示模型每秒识别或分类的图像数量, 它可以衡量模型处理图像的平均速度。表示一个网络的检测速度, 值越大则表明检测速度越快)。FPS 用来衡量算法处理图像的速度大小, 其中包括图像预处理 (Pre)、推理 (Inference) 和非极大值抑制 (NMS) 时间。单位为赫兹 (Hz)。

$$FPS = \frac{1000}{Pre + Inference + NMS} \quad (3-3)$$

3.3 基于改进 YOLOv5 的小目标检测算法

针对现有目标检测算法进行小目标检测, 效果不理想、漏检率高的问题, 本文提出一种改进的 YOLOv5 目标检测算法, 提升小目标检测的效果。在原有模型基础上, 引入 BottleneckCSP 模块并增加大尺度特征融合结构, 提升模型小目标特征捕捉能力; 同时在网络结构中融合 SE 注意力机制, 使得网络自主学习更关注小目标特征通道, 增强网络模型对小目标的检测效果。

3.3.1 引入 BottleneckCSP 模块

为了更好地提取图像的深层特征, 简化 YOLOv5 模型参数, 引入 BottleneckCSP 模块来替换颈部网络中的 C3 模块。BottleneckCSP 模块相对于 C3 模块, 具有更强的特征表达能力和感受野扩展能力, 能够有效提升模型对小目标检测能力。BottleneckCSP 模块由 Bottleneck 和 CSP 两部分组成; 如下图 3-6 所示 Bottleneck 采用经典的残差结构, 主干通过卷积核尺寸是 1×1 的卷积层与卷积核尺寸是 3×3 的卷积层相连接, 然后残差结构把主干部分的输出与输入相加作为整个 Bottleneck 的输出。

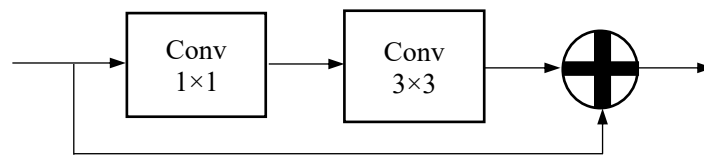


图 3-6 Bottleneck 结构图

CSP 连接是 BottleneckCSP 模块的核心特性, 它通过将特征在通道维度上分割为两个部分, 分别进行卷积操作, 并在最后将它们合并起来。这种跨阶段部分连接可以增强特征的传递性和多样性, 促进信息在网络中的流动, 有助于提高模型的检测性能。

BottleneckCSP 模块结构如图 3-7 所示，输入数据经过模块两个的分支，每个分支分别对一半的通道进行卷积操作，其中一分支通过 Bottleneck 进行操作，减少参数量。最后两个分支通过 concat 融合在一起，使得 BottleneckCSP 的输入与输出保持同样的通道数。此过程可以让模型学习更多的特征，在增加网络深度的同时，减少参数量和计算复杂度，并提高特征表示的能力。

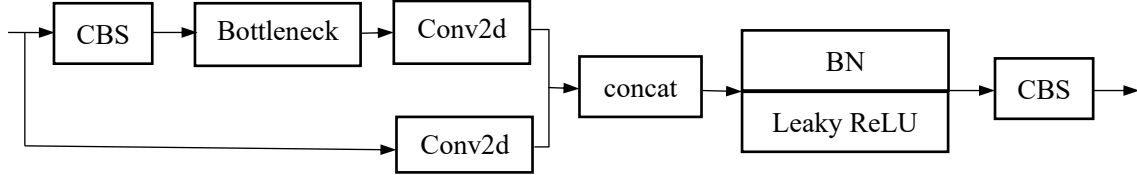


图 3-7 BottleneckCSP 模块结构图

3.3.2 增加大尺度特征融合结构

在 YOLOv5 中采用 FPN+PAN 特征金字塔结构提取不同尺度的特征图，然后将多尺度特征进行融合，最后输入进检测层。原特征金字塔中有三层用于检测层，分别是特征金字塔的 P3 层，对应检测层特征图大小为 80×80 ，用来检测小目标；P4 层，对应检测层特征图大小为 40×40 ，用来检测中目标；P5 层，对应检测层特征图大小为 20×20 ，用来检测大目标。

为提升 YOLOv5 小尺寸特征的检测范围，在原网络基础上融合大尺度特征，将金字塔的 P2 层加入特征融合，浅层特征图中含有相关位置信息，可以提升小目标检测效果。由于浅层网络尺寸大，包含更多小目标信息，因此更有利于小目标的检测，调整后结构如图 3-8 所示。

层数越低的特征图其含有的分辨率高。高分辨率意味着可以包含详细的目标位置和细节信息。而深层的特征图含有更强的语义信息，即图像的纹理、颜色、类别等；分辨率低，对图像的细节感知能力差，通过把深层和浅层特征图进行融合可以改善模型的小目标检测性能。

Concat 融合只改变通道数，不改变特征图的尺寸，Concat 进行图像融合的原理如式 (3-4) 所示。下式中 X_1 表示一个特征图， X_2 表示另一个特征图， (X_1, X_2) 表示将特征图 1 与特征图 2 按照通道数进行拼接。其中特征图 1 的形状为 (B, C_1, H, W) ，特征图 2 的形状为 (B, C_2, H, W) ，Concat 融合后的形状为 (B, C_1+C_2, H, W) ，其中 B 代表样本数量， C_1 和 C_2 分别代表两个特征图的通道数， H 和 W 代表特征图的高度和宽度。

$$\text{Concat}(X_1, X_2) = (B, C_1, H, W) + (B, C_2, H, W) = (B, C_1 + C_2, H, W) \quad (3-4)$$

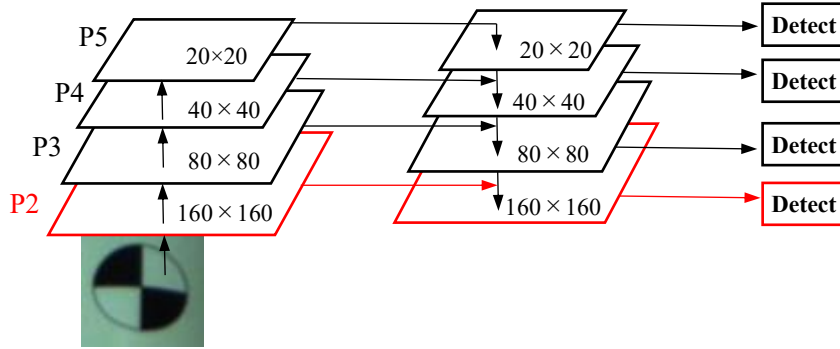


图 3-8 增加的大尺度特征融合结构

为适应增加的大尺度融合，需要对原始锚框（Anchor）做出调整，并增加一层 Detect。在 YOLOv5 模型中自带的初始锚框为：[10, 13, 16, 30, 33, 23]、[30, 61, 62, 45, 59, 119]、[116, 90, 156, 198, 373, 326]。配合本文所融合的大尺度特征图，将锚框由原来的三层增加为四层：[4, 5, 8, 10, 22, 18]、[10, 13, 16, 30, 33, 23]、[30, 61, 62, 45, 59, 119]、[116, 90, 156, 198, 373, 326]。通过使用小尺寸的 Anchor，模型能够更敏锐地捕捉到图像中微小特征和细节，从而提高小目标的检测和识别能力。

3.3.3 融合 SE 注意力机制

SE（Squeeze and Excitation）是一种用于动态地学习每个通道的重要性的方法；然后依据学习到的重要程度来抑制用不到的特征，与此同时将有用的特征权重提升；以提高小目标检测的一种注意力机制。

SE 注意力机制可以在不增加模型复杂度的情况下提升模型对于特征的表征能力和性能。如图 3-9 所示为 SE 注意力机制模块， X 代表特征图，尺寸大小为 $C \times H \times W$ ， C 是通道数， H 和 W 分别是高度和宽度。

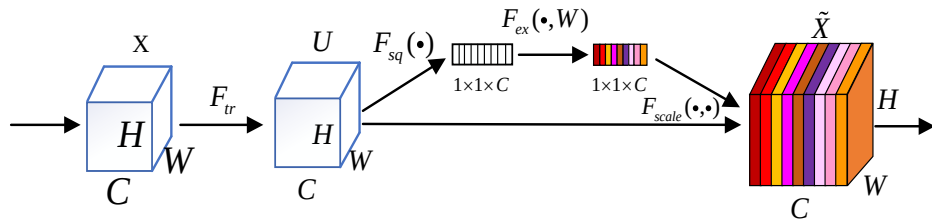


图 3-9 SE 模块

首先将输入的特征图进行压缩处理，将原特征图大小 $C \times H \times W$ 压缩到 $C \times 1 \times 1$ ，从而达到减少计算量和降低参数数量的作用；

$$Z_C = F_{sq}(U_C) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_C(i, j) \quad (3-5)$$

通过挤压操作而聚合得到消息，并且捕获通道之间的依赖关系，最后得到权重 S 。 W_1 、 W_2 为全连接层参数， δ 为 ReLU 函数， σ 为 sigmoid 函数：

$$S = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3-6)$$

将上式求得的注意力权重加权到图像的每个通道特征上，从而得到含有通道注意力的特征图：

$$\tilde{X}_C = F_{scale}(u_C, s_C) = s_C \cdot u_C \quad (3-7)$$

通过 SE 注意力机制对通道的重要性进行建模，从而提高模型的表现性能。将 SE 注意力机制融合在骨干网络（Backbone）后。Backbone 被用来对输入图像的特征进行提取，而将 SE 注意力机制融合在 Backbone 后面可以对提取到的特征进行汇总，抑制不重要特征，从而将重要特征信息传入下一阶段。SE 注意力机制在 Backbone 和 Neck 之间起到桥梁作用，将无用信息过滤的同时减少模型参数量。如图 3-10 所示为最终模型结构。

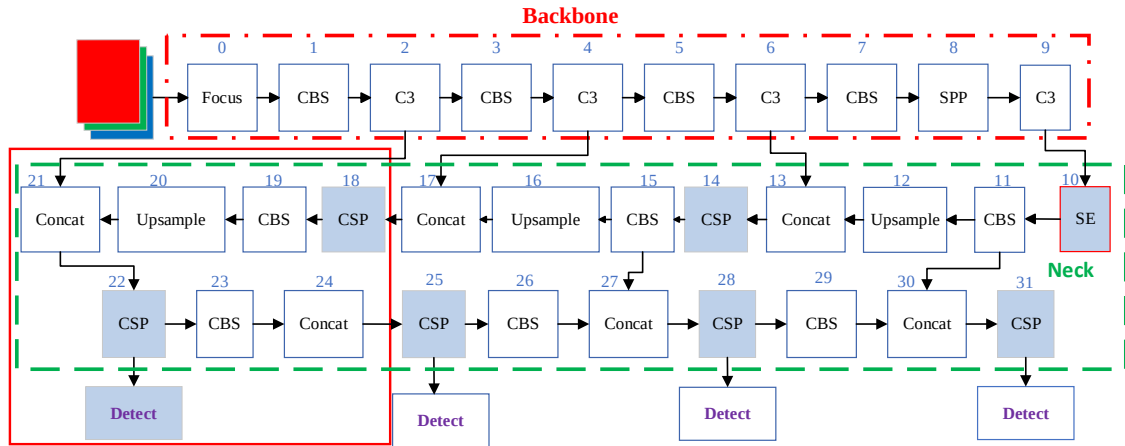


图 3-10 本文模型结构

3.4 实验及结果分析

3.4.1 实验装置

实验装置包括：（1）六自由度工业机器人，带动人工标识运动，实现动态测量的环境；（2）高速相机及其所用镜头，对应参数如表 3-1 和表 3-2 所示；（3）安装相机所用的三脚架；（4）辅助光源；（5）算法处理平台，其参数如表 3-3 所示。

表 3-1 相机参数

传感器	分辨率 (H×V)	帧率 (s ⁻¹)	位深	像素尺寸 (um ²)	快门模式
COMS	2368×1720	<13000	8 bit	7×7	全局

表 3-2 镜头参数

光圈范围	过滤器螺纹直径 (mm)	焦距 (mm)	放大倍数	视角
F/16-F/1.4	77	24	1:5.3	84.1°

表 3-3 算法处理平台

操作系统	显卡配置	开发环境	框架版本	加速库版本
Windows11	NVIDIA A30	Python3.9	Pytorch2.0.0	CUDA12.0



图 3-11 搭建的实验装置系统

3.4.2 实验数据准备

训练所用图像均由搭建的实验平台采集而来。为提高模型泛化能力，图像数据采集过程中，不断改变光照强度、标识倾斜和旋转角度，同时将标识贴至其他物体上，以此来采集不同状态的样本。这样做可以达到丰富图像数据，防止训练集过小产生过拟合的效果。本实验采用的标注工具为 Labelimg。在对图片特征进行标注时，标注类别只有一类，标注名称为“biaoshi1”。数据集整体数量为 1000 张图片，如图 3-12 所示为所制备的数据集中的部分图像数据。训练集、验证集以 9:1 的比例划分。

现在有两种定义方式来描述小目标：绝对定义和相对定义。绝对定义指出当目标物的像素尺寸小于 32×32 像素时，即称为小目标；相对定义是指当目标物的像素尺寸相较目标物所在图像的占比小于 0.1 时，此目标定义为小目标。在实际应用中，由于相机拍摄图像尺寸不固定，相对定义的方式应用更为广泛，本文主要针对相对定义下的小目标检测问题。如图 3-11 所示实验系统，相机采集图像尺寸为 2368×1720 ，相机和机器人保持上述距离时，图像中目标像素尺寸在 100×100 像素左右，根据相对定义，目标尺寸比上图像尺寸得到的比率约为 0.0024，数值远小于小目标的相对定义 0.1，本实验中的目标为小目标。

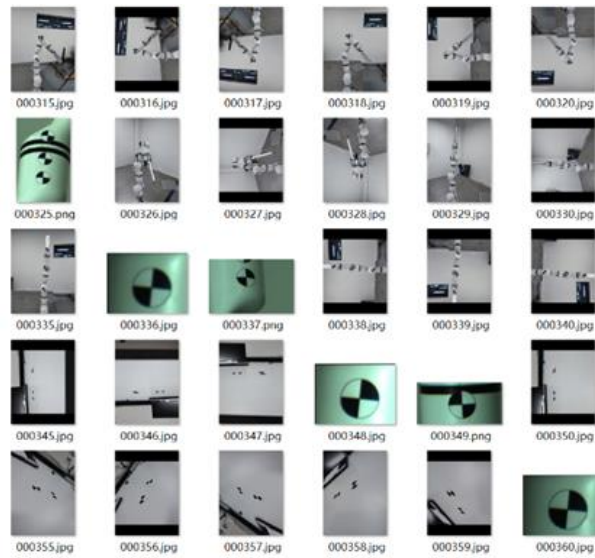


图 3-12 人工标识数据集

3.4.3 结果分析

通过消融实验来验证所提改进算法的效果，即原始 YOLOv5 算法、引入 BottleneckCSP 模块的 Y-BP 算法、增加大尺度特征融合结构 Y-U 算法、融合 SE 注意力机制 Y-SE 算法以及融合所有改进的 Y-ALL 算法。消融实验结果如表 3-4 所示。可以看出改进算法各方面性能稳步提升，证明单个改进方法的有效性。

表 3-4 消融实验

模型	精确率 (P)	召回率 (R)	mAP (0.5:0.95)
YOLOv5	0.91	0.93	0.92
Y-BP	0.92	0.93	0.92
Y-U	0.92	0.94	0.93
Y-SE	0.94	0.95	0.94
Y-ALL	0.95	0.97	0.96

对所提算法、原始 YOLOv5 算法以及前人针对小目标检测引入 CA 注意力机制的 YOLOv5+CA 算法同时加入 SSD 和 Faster R-CNN 算法进行训练和验证，试验小目标检测范围以及目标平均检测率。YOLO 网络设置输入图像大小为 640×640，训练（Epochs）共计 100 轮。各模型性能对比如表 3-5 所示。

表 3-5 模型性能对比

模型	精确率（P）	召回率（R）	mAP（0.5:0.95）
YOLOv5	0.91	0.93	0.92
YOLOv5+CA	0.92	0.94	0.93
SSD	0.89	0.88	0.64
Faster R-CNN	0.90	0.89	0.66
Y-ALL	0.95	0.97	0.96

为进一步验证所提出改进小目标检测算法的有效性，对比算法的小目标检测范围以及平均检测率。比对小目标检测范围，改变待测目标和相机之间的距离，从而改变目标的大小及其在图像中占比的大小。实验过程中，相机和待测目标之间的距离由 1.25m 不断调整至 2.0m，目标比率由 0.1 变化至 0.0008，检测结果如表 3-6 所示，其中“✓”在该范围内能够实现目标检测。由表 3-6 所列结果可以看出，原始 YOLOv5 算法的小目标检测最小范围在 0.0025 至 0.0018 区间，YOLOv5+CA 算法的小目标检测最小范围在 0.0018 至 0.0014 区间，SSD 和 Faster R-CNN 算法表现不佳，所提改进算法的小目标检测最小范围在 0.0014 至 0.0008 区间，相比较而言，所提改进算法能够有效提升小目标检测范围。

表 3-6 小目标检测范围

模型	比率					
	0.1~	0.05~	0.01~	0.0025~	0.0018~	0.0014~
	0.05	0.01	0.0025	0.0018	0.0014	0.0008
YOLOv5	✓	✓	✓	✓		
YOLOv5+CA	✓	✓	✓	✓	✓	
SSD	✓	✓	✓			
Faster R-CNN	✓	✓	✓	✓		
Y-ALL	✓	✓	✓	✓	✓	✓

为比对所提算法的小目标平均检测率，开展多个小目标检测对比实验，如图 3-11 所示，在机器人本体上张贴 5 个人工标识，改变机器人的姿态，连续获得 10 帧包含 5 个人工标识图像，比对前述训练后的算法小目标识别的检测率，其中某

一帧的实际检测效果如图 3-13 所示，连续 10 帧图像的检测结果如表 3-7 所示。

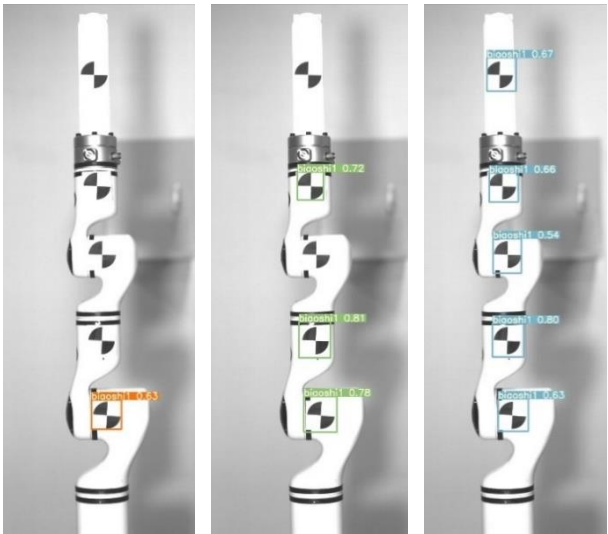


图 3-13 实际检测效果对比图

表 3-7 平均检测率对照表

模型	实验											平均检 测率
		1	2	3	4	5	6	7	8	9	10	
YOLOv5	检测个数	3	4	3	3	1	2	2	0	1	2	42
	检测率/%	60	80	60	60	20	40	40	0	20	40	
YOLOv5+CA	检测个数	4	3	3	3	3	3	4	3	2	3	62
	检测率/%	80	60	60	60	60	60	80	60	40	60	
SSD	检测个数	1	0	1	2	0	1	0	1	1	0	14
	检测率/%	20	0	20	40	0	20	0	20	20	0	
Faster R-CNN	检测个数	2	1	0	2	1	0	2	1	0	1	20
	检测率/%	40	20	0	40	20	0	40	20	0	20	
Y-ALL	检测个数	5	4	5	3	5	4	5	5	3	5	88
	检测率/%	100	80	100	60	100	80	100	100	60	100	

由表 3-7 可以看出，对于同一张包含 5 个目标的图片，所提算法检测出小目标的个数远超其他算法。连续 10 帧图像的小目标平均检测率为 88%，对比原始 YOLOv5 算法的 42%和 YOLOv5+CA 算法的 62%，SSD 算法 14%和 Faster R-CNN 的 20%，检测性能提升明显。

3.5 本章小结

本章提出一种改进的 YOLOv5 目标检测算法，解决现有目标检测算法进行小目标检测时，检测效果不理想、漏检率高的问题。通过在原始的 YOLOv5 目标检测框架中引入 BottleneckCSP 模块、增加大尺度特征融合结构同时嵌入 SE（Squeeze and Excitation）注意力机制来提升小目标检测能力。通过实验验证得出：（1）改进算法能够提升原始 YOLOv5 目标检测算法的 mAP 值和训练过程中的收敛速度；（2）有效拓展小目标检测范围，由原始算法的 0.01~0.0025，拓展至 0.0014~0.0008；（3）提高小目标检测性能，相比原始 YOLOv5 算法，平均检测率提升 46%；改进算法能够有效提升小目标检测的能力。

第 4 章 多目标跟踪与精确定位

在目标检测后进行目标跟踪步骤，目标在前一帧被检测到并标记后，后面的视频帧中能够连续捕捉同一个目标或者保持同一个 ID。目标跟踪在计算机视觉里有着广泛的应用，主要是识别和跟踪视频中的运动目标，如行人、车辆、飞机等目标。主要应用于视频监控、自动驾驶、医学图像分析等领域。主要研究内容为实时获取目标的位置和轨迹，以满足应用需求。目标跟踪通常有单目标跟踪和多目标跟踪。

常用的跟踪方法有 Meanshift、KCF 等算法。Meanshift 在单目标、实验环境简单、目标运动缓慢等情况下跟踪成功率高，当目标运动速度过快时跟踪容易失败。KCF 算法在目标运动过程中发生较大尺度变化、目标偏转角度过大等情况发生时，跟踪框容易发生偏移导致跟踪失败。本章主要对第三章中检测到机器人本体上张贴的人工标识进行轨迹跟踪，卡尔曼滤波和匈牙利算法是核心，确保多目标能够成功跟踪。

4.1 跟踪算法概述

本章内容是基于上一章 YOLOv5 目标检测算法所检测到的目标框的坐标(x, y, w, h)信息进行推理。通过目标检测得到的当前帧的目标位置坐标、大小等信息，卡尔曼滤波算法依据此时刻的位置信息进行估计，得出下一时刻的目标位置信息。然后进行相似度计算，计算前一帧和当前帧的目标的匹配程度。通过匈牙利算法进行数据匹配，为多个追踪到的目标分配 ID。匈牙利算法可以解决分配的问题，通过匹配目标检测算法的检测框与卡尔曼滤波算法的预测框，卡尔曼滤波可以通过这个方法找到与预测框最匹配的检测框，从而达到跟踪效果。

算法流程如下：



图 4-1 跟踪算法流程

4.2 多目标跟踪算法

4.2.1 卡尔曼滤波轨迹预测方法

卡尔曼滤波算法可以对系统状态进行最优估计。它是一种线性系统状态方程，通过系统的输入输出数据来完成估计。由 Kalman 在 1960 年代提出并在阿波罗登月计划中得到应用，后来在机器人、自动驾驶、制导与控制等领域也被广泛应用。由于观测数据中含有系统中的干扰和噪声，所以该方法也被看作是滤波的过程。该滤波方法的核心思想是不断融合系统动态模型（预测）和测量的实际观测值（校正）。在有噪声和不确定的情况下对系统提供最优估计。

在目标跟踪研究中，如何准确预测目标的行进轨迹是研究重点。是算法能够准确跟踪目标的前提。卡尔曼滤波本质为线性最小均方差估计器，当系统存在噪声和干扰的状态下进行系统状态的最优估计。根据现在时刻视频帧中的测量值和上一帧对本帧的预测值和误差值进行加权平均操作，进而得到本视频帧的最优值，并对下一视频帧进行预测。卡尔曼滤波可以很好地在含有噪声的数据过程中估计状态。

系统状态方程：

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (4-1)$$

上式中， A 是状态转移矩阵，表示此时刻 k 的状态和上一时刻 $k-1$ 状态之间的转换关系； B 是控制输入模型（矩阵）， u_{k-1} 为上一时刻 $k-1$ 外部的输入信号， x_k 代表系统此时刻 k 的状态变量， x_{k-1} 代表系统上一时刻 $k-1$ 的状态变量， w_{k-1} 代表 $k-1$ 时刻系统过程噪声； $X=[Cx, Cy, r, h, Vx, Vy, Vr, Vh]^T$ 为系统状态； Cx 代表目标矩形框中心点的 x 坐标； Cy 代表目标矩形框中心点的 y 坐标； r 代表目标矩形框的宽高比； h 代表目标矩形框的高； Vx, Vy, Vr, Vh ：是各个参量的速度；

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & dt & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & dt & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & dt & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & dt \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4-2)$$

控制输入模型 $B=0$

过程噪声 $p(w) \sim N(0, Q)$, Q 表示过程噪声协方差矩阵; 过程噪声源自目标移动过程中的一些不确定性, 比如突然减速、加速、转弯等等。

观测方程:

$$z_k = Hx_k + v_k \quad (4-3)$$

H 是观测矩阵, 代表当前状态到测量的转换, z_k 表示系统 k 时刻的观测变量, v_k 表示系统观测噪声;

需要在 N 个检验框中初步寻找到观测值。 k 时刻的最优估计值与 $k+1$ 时刻所有框, 匹配度最高的为 $k+1$ 时刻的观测值。得到所有观测值: x, y, w, h, dx, dy

$dx=x-x^-$ (当前时刻-上一时刻), $dy=y-y^-$ (当前时刻-上一时刻)

状态观测模型:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4-4)$$

观测噪声 $p(v) \sim N(0, R)$, R 为观测噪声的协方差矩阵, 观测噪声来自检验框的实际情况, 比如检验框的重叠、丢失、不准确等等。

当匹配失败, 则使用上一帧的最优估计值作为观测值;

$$x_k = Ax_{k-1} \quad (4-5)$$

$$\begin{pmatrix} Cx \\ Cy \\ r \\ h \\ Vx \\ Vy \\ VR \\ Vh \end{pmatrix}_k = \begin{pmatrix} 1 & 0 & 0 & 0 & dt & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & dt & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & dt & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & dt \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Cx \\ Cy \\ r \\ h \\ Vx \\ Vy \\ VR \\ Vh \end{pmatrix}_{k-1} \quad (4-6)$$

此时更新一次状态, 直到在 $k+n$ 帧里面寻找到匹配框。

整个卡尔曼滤波可以分为预测和更新两个过程:

(1) 状态预测过程:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \quad (4-7)$$

$$P_k^- = AP_{k-1}A^T + Q \quad (4-8)$$

由公式 (4-7) 可知当前时刻 k 的状态量是依据上一时刻 $k-1$ 的状态后验估计值预测得到的。 $\hat{x}_k$ 代表此时刻 k 的状态先验估计, A 代表目标的状态转移模型 (矩阵), $\hat{x}_{k-1}$ 代表上一时刻 $k-1$ 目标状态的后验估计值; 公式 (4-8) 表示由上一时刻的协方差得到此时刻的协方差, P_k^- 表示 k 时刻的协方差矩阵先验估计值, P_{k-1} 代表 $k-1$ 时刻的协方差后验估计值, Q 表示过程噪声的协方差矩阵。

(2) 状态更新过程:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \quad (4-9)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H\hat{x}_k^-) \quad (4-10)$$

$$P_k = (I - K_k H) P_k^- \quad (4-11)$$

上述公式中, K_k 表示此时刻 k 的卡尔曼增益, 用于估计误差的重要程度; $\hat{x}_k$ 代表 k 时刻的状态后验估计值, P_k 代表 k 时刻协方差矩阵后验估计值。通过状态更新过程, 模型获得了此时刻 k 系统状态的最终估计值 $\hat{x}_k$ 和此时刻 k 的协方差 P_k , 在下一时刻的预测过程中这两个值将用来进行新的迭代。

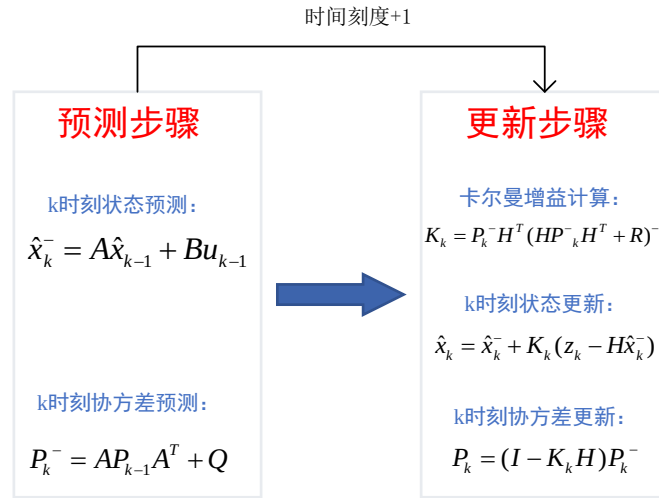


图 4-2 卡尔曼滤波步骤

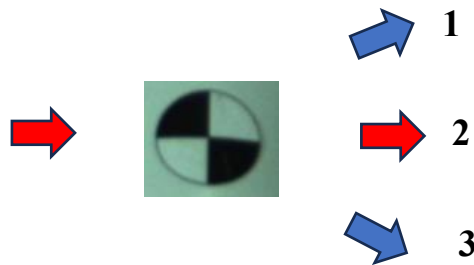


图 4-3 卡尔曼预测

如图 4-3 所示，当目标以直线方向从左边向右边运动的过程中，经过卡尔曼轨迹预测后会给出红色箭头所指路径。

4.2.2 基于匈牙利算法的目标匹配

通过目标检测器对导入的视频流进行检测，得到目标的检测框，与此同时卡尔曼滤波对目标的运动进行预测，生成目标的预测框，同时提取框内的目标特征。通过计算检测框与预测框的相似度来对连续帧出现的目标进行数据匹配，最后将 ID 分配给计算后的目标。

在检测视频中会出现多个目标，当检测完毕后会为每个目标分配一个唯一 ID，例如“Biaoshi-1, Biaoshi-2, Biaoshi-3...”等等，在视频后面的每一帧中如果每个目标的 ID 都能一直保持不变，则说明目标跟踪成功，反之跟踪失败。确保 ID 前后保持不变，这里需要匈牙利算法。

如图 4-4 所示为目标跟踪匹配示意图。

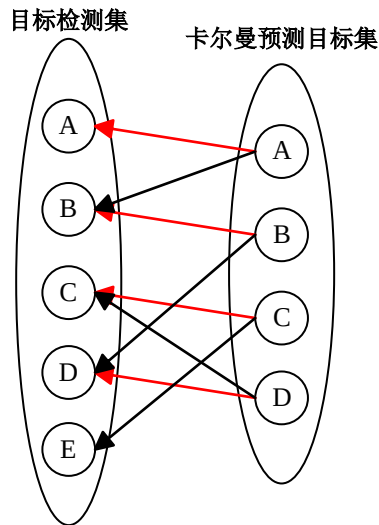


图 4-4 目标跟踪匹配示意图

匈牙利算法是一种解决分配问题的优化算法，它在目标跟踪中用于分配每个检测结果给跟踪器，以确定哪些检测结果属于同一个目标。也被称为二分图的最佳匹配问题，在这个问题中，一侧的节点代表当前帧的检测结果，另一侧的节点代表前一帧中的跟踪结果，权重表示将一个检测结果与一个跟踪器关联的成本（距离或相似性度量）。匈牙利算法的任务是找到最佳匹配，即最小化总成本，以确定该检测结果与该跟踪器关联。

通过匈牙利算法，能够将当前帧中的每个检测结果分配给前一帧中的跟踪器，

从而建立目标的跟踪关系。这确保了在连续帧之间，相同的目标被正确地跟踪。但是在匹配时还需加上目标的运动信息和外观信息来增加匹配的成功率，确保匹配的稳定性。

匈牙利算法通过加权融合外观特征的相似度量度和运动特征的距离度量，获得匈牙利算法的代价矩阵，从而实现数据的全局最优匹配。

(1) 运动特征信息匹配

利用上一帧跟踪框与当前帧的检测框间的马氏距离来计算目标在像素位置维度的相似度，从而实现目标在运动特征上的信息匹配。本文采用马氏距离来衡量目标跟踪框与目标检测框之间的相似度，是因为马氏距离可以度量目标样本间的距离，同时不受目标尺寸和变量相关性的干扰。马氏距离可以起到去掉错误分配的重要作用。

$$d_1(i, j) = (Z_j - X_i)^T S_i^{-1} (Z_j - X_i) \quad (4-12)$$

如上公式(4-12)所示， i 表示目标的跟踪框序号， j 表示目标的检测框序号， $d_1(i, j)$ 代表跟踪框与检测框之间的马氏距离， S_i 表示在卡尔曼滤波器中第 i 个协方差矩阵，代表目标位置的容错性。 Z_j 表示第 j 个检测框的观测模型（矩阵）， X_i 代表第 i 个跟踪框状态估计值。马氏距离的值越小，代表目标检测框和跟踪框之间的运动特征相似度越高，检测框与跟踪框也就越匹配。

(2) 外观特征信息匹配

在检测视频前后帧之间，同一目标物体一般运动距离较短，对这两个目标物体进行特征提取，然后通过比较两个物体的特征相似度来判断是否为同一目标物，相似度越高则可能是同一物体，反之则不是。

通过模型对第 i 个跟踪框和第 j 个检测框进行特征提取，获得外观特征信息值 r^i, r_j ，计算两者之间的余弦距离来衡量彼此之间的相似程度。

$$d_2(i, j) = \min\{1 - r_j^T r_k^i \mid r_k^i \in R_i\} \quad (4-13)$$

式(4-13)中， r_j 代表第 j 个目标的特征值， r_k^i 代表第 k 时刻的轨迹特征， R_i 代表整个轨迹 i 的特征集合， $d_2(i, j)$ 表示两个值之间的余弦距离，余弦距离值越小说明跟踪框和检测框的相似度越高，两者的匹配度也就越大。

$$C(i, j) = \lambda d_1(i, j) + (1 - \lambda) d_2(i, j) \quad (4-14)$$

通过超参数 λ 把马氏距离与余弦距离进行融合，从而解出匈牙利算法的代价

矩阵，如式（4-14）所示，来完成数据间的匹配。

4.3 目标精确定位方法

通过 YOLOv5 目标检测，然后跟踪算法进行目标跟踪，可以准确在视频帧中检测跟踪机器人本体上张贴的人工标识。但是这不够，需要精确定位出双目相机中的同一标识的一点，这样才能利用双目相机原理还原三维世界的坐标。通过分析标识的特殊性，决定利用图像处理等操作提取标识的中心点坐标。

首先针对检测跟踪到的人工标识进行提取，然后通过图像裁剪进一步处理人工标识图像；之后进行感兴趣区域提取，得到人工标识中心的局部坐标；最后通过亚像素边缘检测找到人工标识的纵横交叉线，通过直线拟合得到两条直线的交点即为人工标识的中心点坐标。

（1）检测跟踪标识提取

在处理视频帧时，算法会实时检测跟踪人工标识，通过前期设定，可以将检测跟踪到的人工标识单独保存下来进行图像后处理。如下图 4-5 所示为保存的人工标识图。然后对人工标识进行裁剪，避免后期进行边缘检测时产生影响。



图 4-5 保存的人工标识



图 4-6 裁剪的人工标识

(2) 亚像素边缘检测

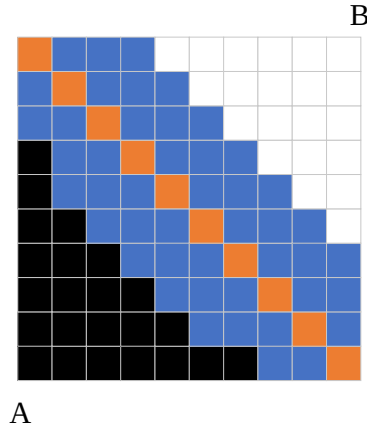


图 4-7 亚像素边缘示意图

如图 4-7 所示，边缘两侧的强度值为 A 和 B ，设 A 和 B 已知（ A ：白色； B ：黑色）像素点强度按照公式计算：

$$G(i, j) = B + \frac{A - B}{h^2} E_{i,j} \quad (4-15)$$

式中， $G(i, j)$ 代表 (i, j) 处的灰度值， h 代表边缘长度，单位为 1， $E_{i,j}$ 代表边缘点 (i, j) 内边缘线下的面积。

求解边缘线的表达式，在目标边缘提取满足求解要求最小区域，设 S_L 、 S_M 、 S_N 是子窗口左、中、右三列像素强度和，表达式为：

$$\begin{cases} S_L = \sum_{n=j-2}^{j+2} F_{i-1,n} = nB + \frac{A-B}{h^2} E_L \\ S_M = \sum_{n=j-2}^{j+2} F_{i,n} = nB + \frac{A-B}{h^2} E_M \\ S_N = \sum_{n=j-2}^{j+2} F_{i+1,n} = nB + \frac{A-B}{h^2} E_N \end{cases} \quad (4-16)$$

其中， E_L 、 E_M 、 E_N 代表左、中、右列边缘线的下方面积，由公式可以求出：

$$\begin{cases} E_L = \int_{-3h/2}^{-h/2} (ax^2 + bx + c + nh/2) dx = ch - bh^2 + \frac{13}{12} ah^3 + \frac{n}{2} h^2 \\ E_M = \int_{-h/2}^{h/2} (ax^2 + bx + c + nh/2) dx = ch + \frac{1}{12} ah^3 + \frac{n}{2} h^2 \\ E_N = \int_{h/2}^{3h/2} (ax^2 + bx + c + nh/2) dx = ch + bh^2 + \frac{13}{12} ah^3 + \frac{n}{2} h^2 \end{cases} \quad (4-17)$$

$$\begin{cases} a = \frac{S_L + S_N - 2S_M}{2(A - B)} \\ b = \frac{S_N - S_L}{2(A - B)} \\ c = \frac{2S_M - 5(A + B)}{2(A - B)} - \frac{1}{12}a \end{cases} \quad (4-18)$$

解出边缘线的方程即可把像素边缘转换为亚像素,可以提升标识的边缘检测精度。

对裁剪得到的人工标识进行边缘检测以便提取标识点的中心坐标。首先对图像进行相关滤波处理,去除图像的杂质来提高图像的质量。然后亚像素边缘检测算法对裁剪的人工标识进行边缘检测提取轮廓。处理结果如图 4-8 所示。

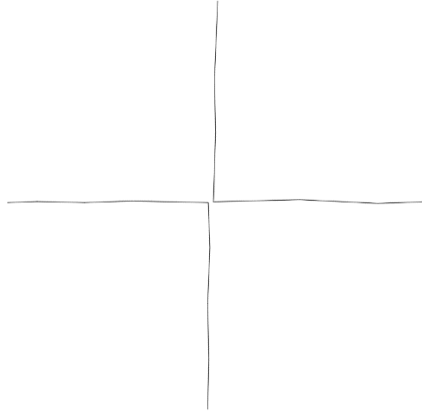


图 4-8 边缘检测结果

(3) 直线拟合

求标识的中心点坐标一个很好的办法就是通过标识的特征求出两条明显的直线,通过直线相交的位置可以判断为标识图像的中心点坐标。首先对图像进行亚像素边缘检测,然后进行直线拟合;将 $y=kx+b$ 作为直线方程,然后求出 y_i 为加权平方和最小观测值偏差:

$$f_{\min}(k, b) = \sum_{i=1}^N [y_i - (b + kx_i)]^2 \quad (4-19)$$

式中, k 为直线方程斜率, b 为直线截距, 边缘数据点为 (x_i, y_i) , 边缘数据点序号为 N 。

对上式的 k, b 求导:

$$\begin{aligned}\frac{\partial}{\partial b} \sum_{i=1}^N [y_i - (b + kx_i)]^2 &= -2 \sum_{i=1}^N (y_i - b - kx_i) = 0 \\ \frac{\partial}{\partial k} \sum_{i=1}^N [y_i - (b + kx_i)]^2 &= -2 \sum_{i=1}^N [(y_i - b - kx_i)x_i] = 0\end{aligned}\quad (4-20)$$

整理得：

$$\begin{cases} \hat{b} = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{N(\sum x_i^2) - (\sum x_i)^2} \\ k = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{N(\sum x_i^2) - (\sum x_i)^2} \end{cases} \quad (4-21)$$

横、纵直线的解析式为：

$$\begin{cases} y_1 = k_1 x_1 + b_1 \\ y_2 = k_2 x_2 + b_2 \end{cases} \quad (4-22)$$

式中， x_1 、 y_1 表示第一条直线的点， x_2 、 y_2 表示第二条直线的点； k_1 表示第一条直线的斜率； b_1 表示第一条直线的截距； k_2 表示第二条直线的斜率； b_2 表示第二条直线的斜截距；

求得两直线的交点，即为标识图像的中心点坐标：

$$\begin{cases} x = \frac{b_2 - b_1}{k_1 - k_2} \\ y = k_1 \times \left(\frac{b_2 - b_1}{k_1 - k_2} \right) + b_1 \end{cases} \quad (4-23)$$

直线拟合后求交点坐标如图 4-9 所示：

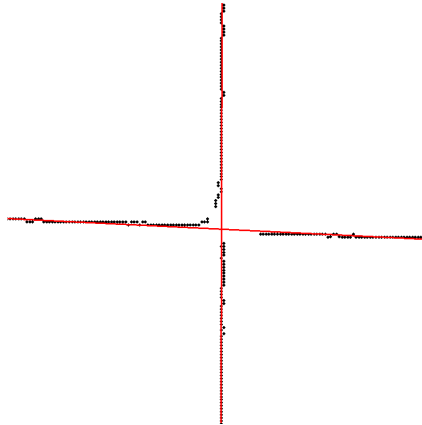


图 4-9 求交点

两直线交点就是人工标识的中心点坐标，就可以求出标识的像素坐标，之后反推到整个图像中，就可以得出该标识所代表机器人的坐标。标识中心点坐标如图 4-10 所示。

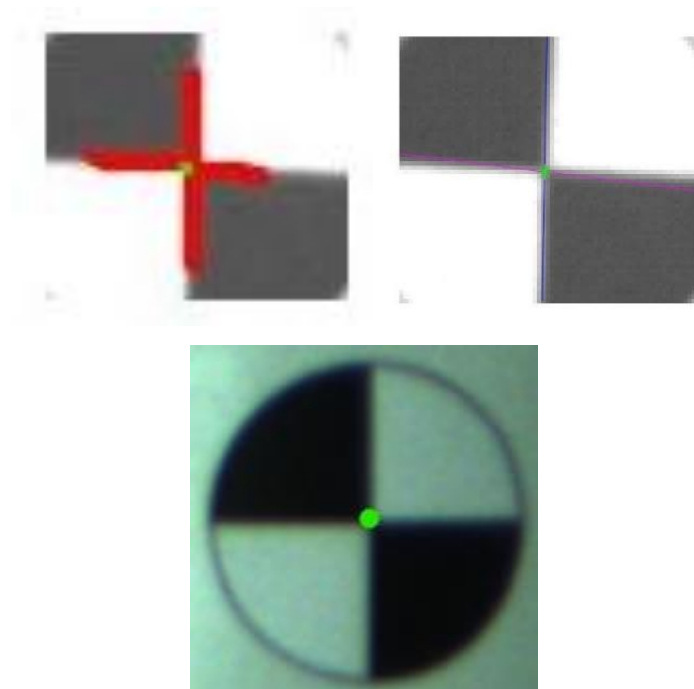


图 4-10 标识中心显示

4.4 实验结果与分析

为验证跟踪算法设计的有效性，现在根据第三章检测结果进行跟踪实验。将检测视频导入算法中，查看算法针对人工标识的跟踪情况。首先测试单个目标下跟踪效果，然后再进行多目标的跟踪实验。

（1）单目标跟踪实验

在机器人本体上只张贴一个人工标识，检验跟踪效果；

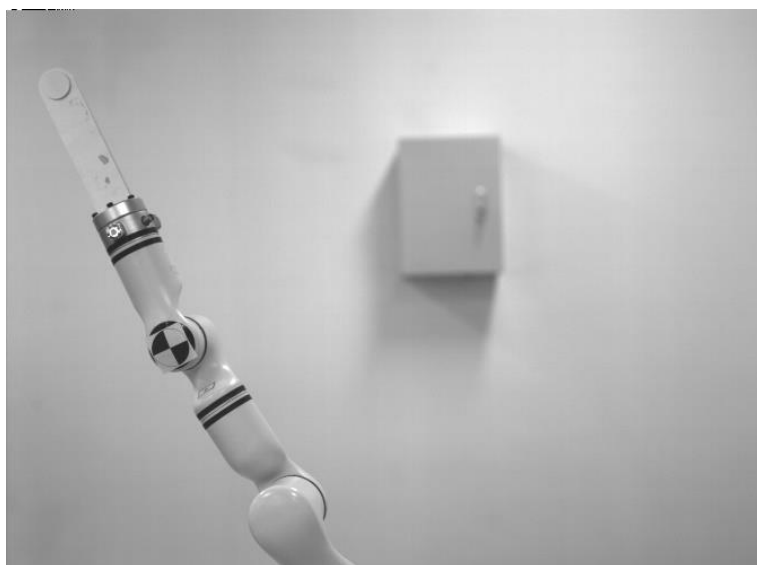


图 4-11 实验图

导入视频，目标检测器对第一帧画面进行目标检测，将检测到的目标结果输

送给目标跟踪器进行目标跟踪，第二帧开始将预测目标与检测目标进行匹配保持同一个 ID，往后视频帧中如此循环，以达到跟踪效果。

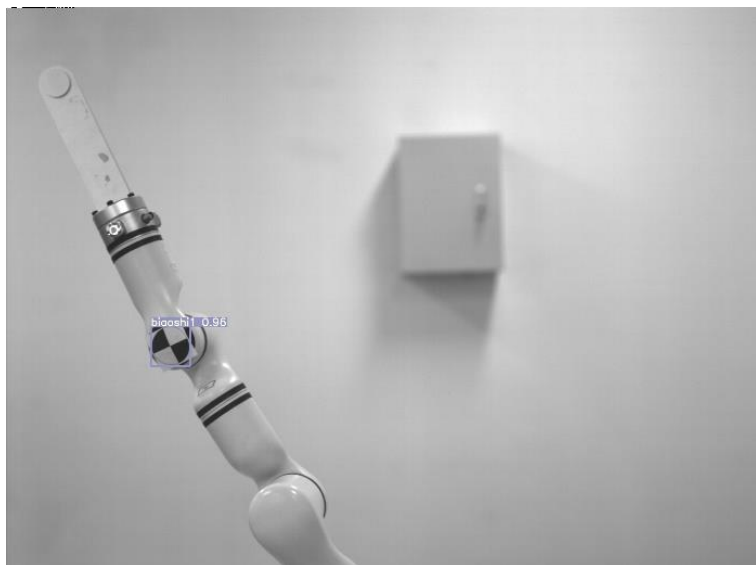


图 4-12 检测结果

将检测结果赋予跟踪 ID，如图 4-13 所示为“BiaoShi-1”，如果在后续的视频帧中，每一帧都能检测跟踪到人工标识并且保持 ID 一致，即可认为跟踪成功。

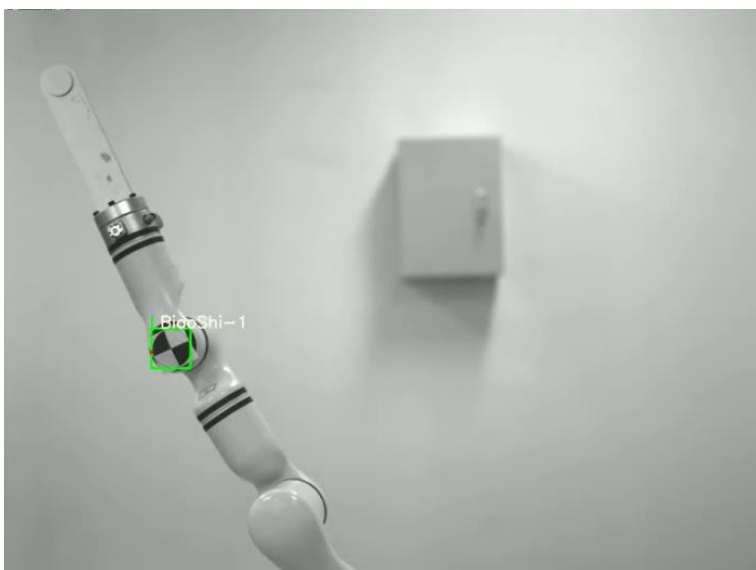
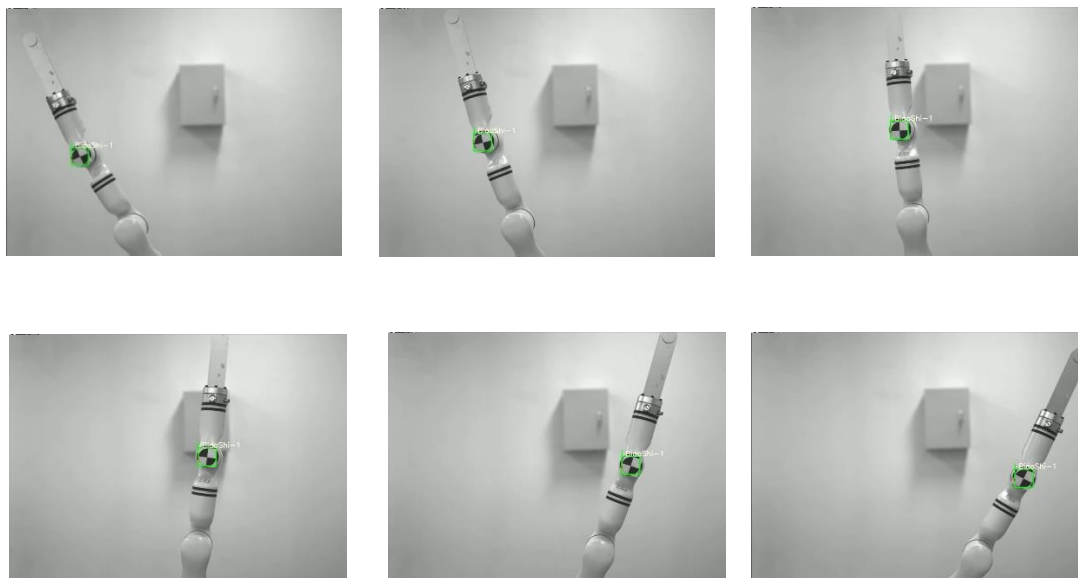


图 4-13 赋予 ID

查看目标在视频中的跟踪情况：



4-14 单目标跟踪实验图

如图 4-14 所示是在机器人从左往右地运动过程中截取的跟踪图像，从整个跟踪结果可以看出，在单目标跟踪实验中，目标可以准确跟踪并且整个跟踪过程中 ID 保持不变，说明跟踪实验成功。

(2) 多目标跟踪实验

当只有一个目标时检测跟踪比较容易，不会存在 ID 混乱、交叉等现象。为验证跟踪模型的稳定性，在机器人本体上张贴三个人工标识来进行多目标跟踪实验并观察检测跟踪的结果。

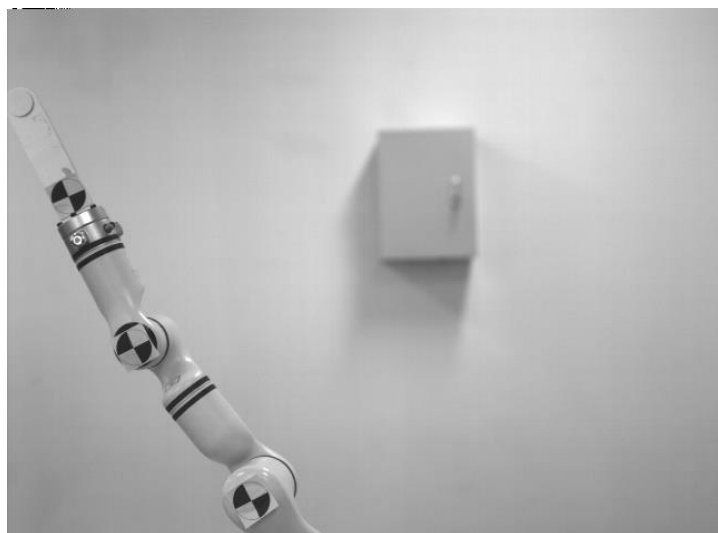


图 4-15 实验图

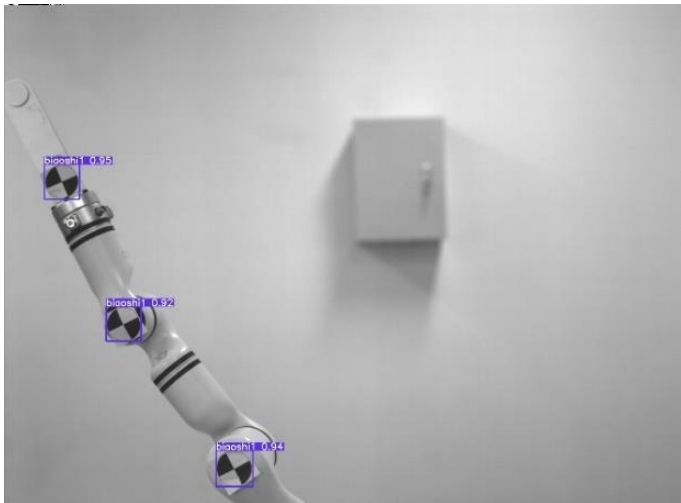


图 4-16 检测结果

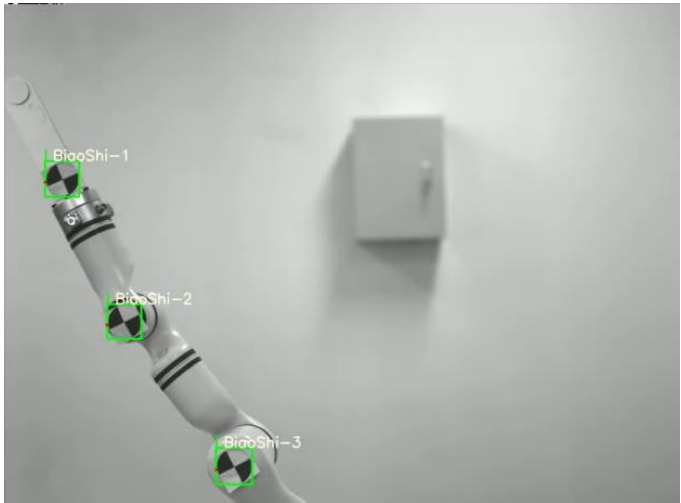


图 4-17 赋予 ID

查看多目标在视频中的跟踪结果：

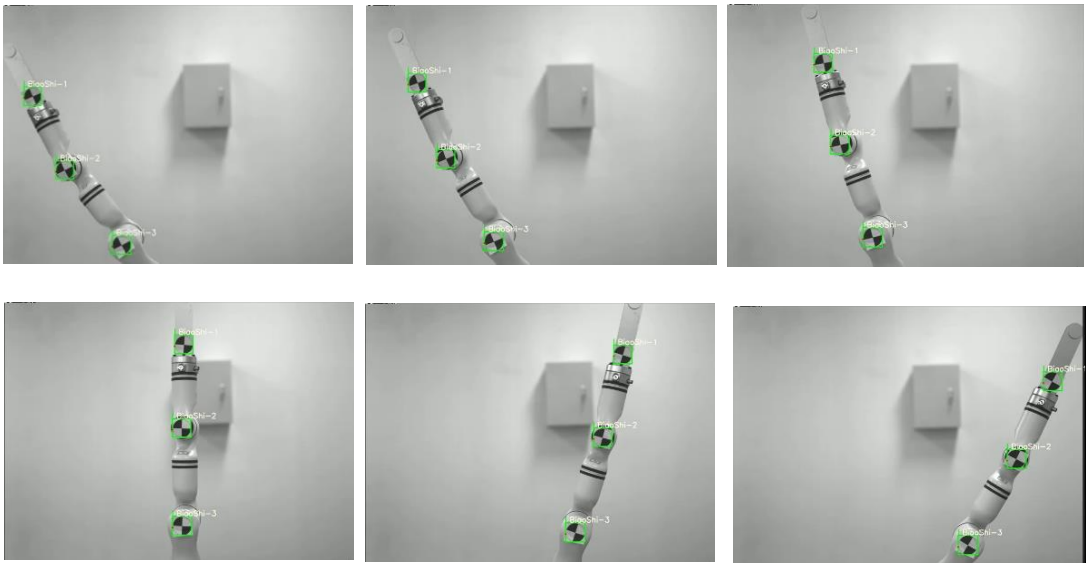


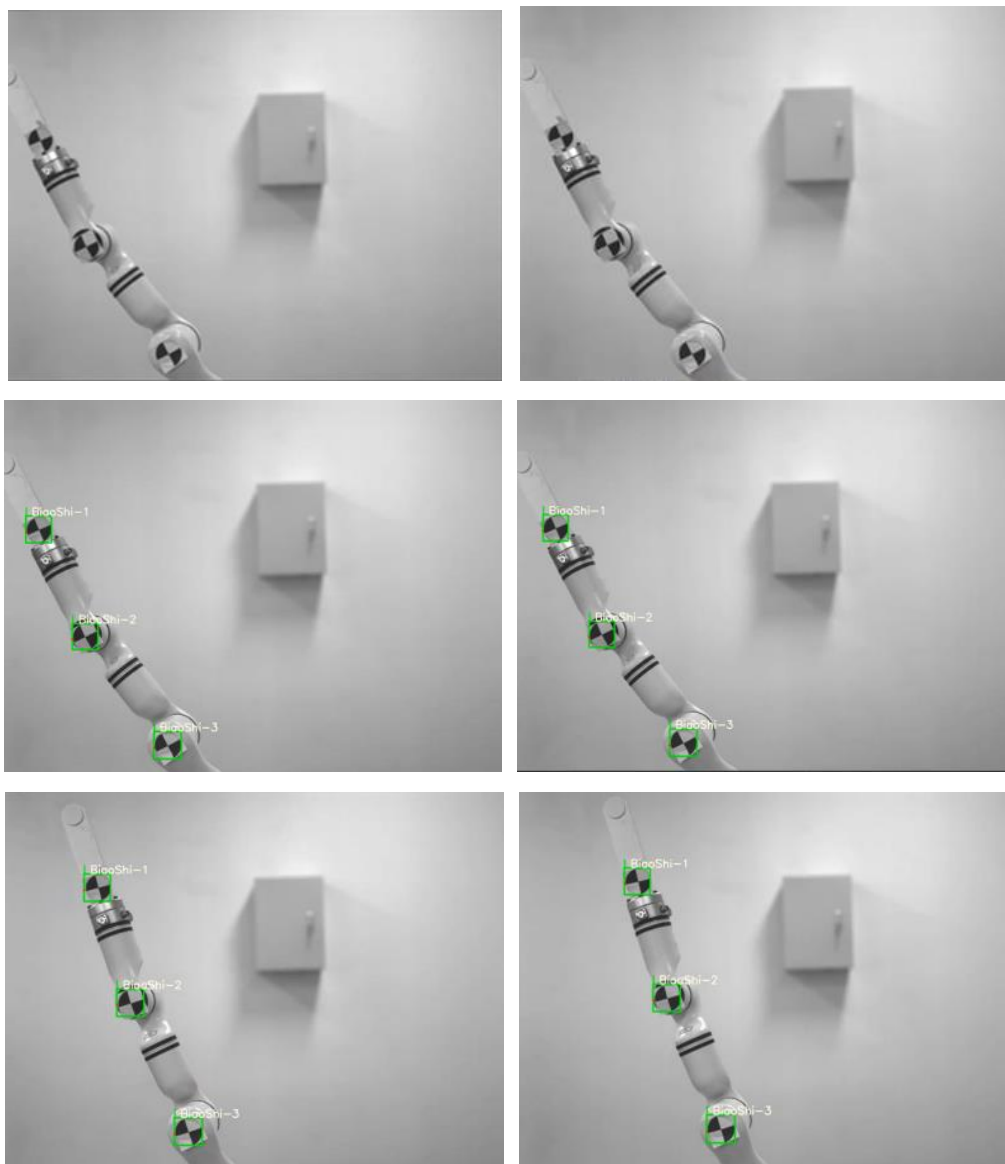
图 4-18 多目标跟踪实验图

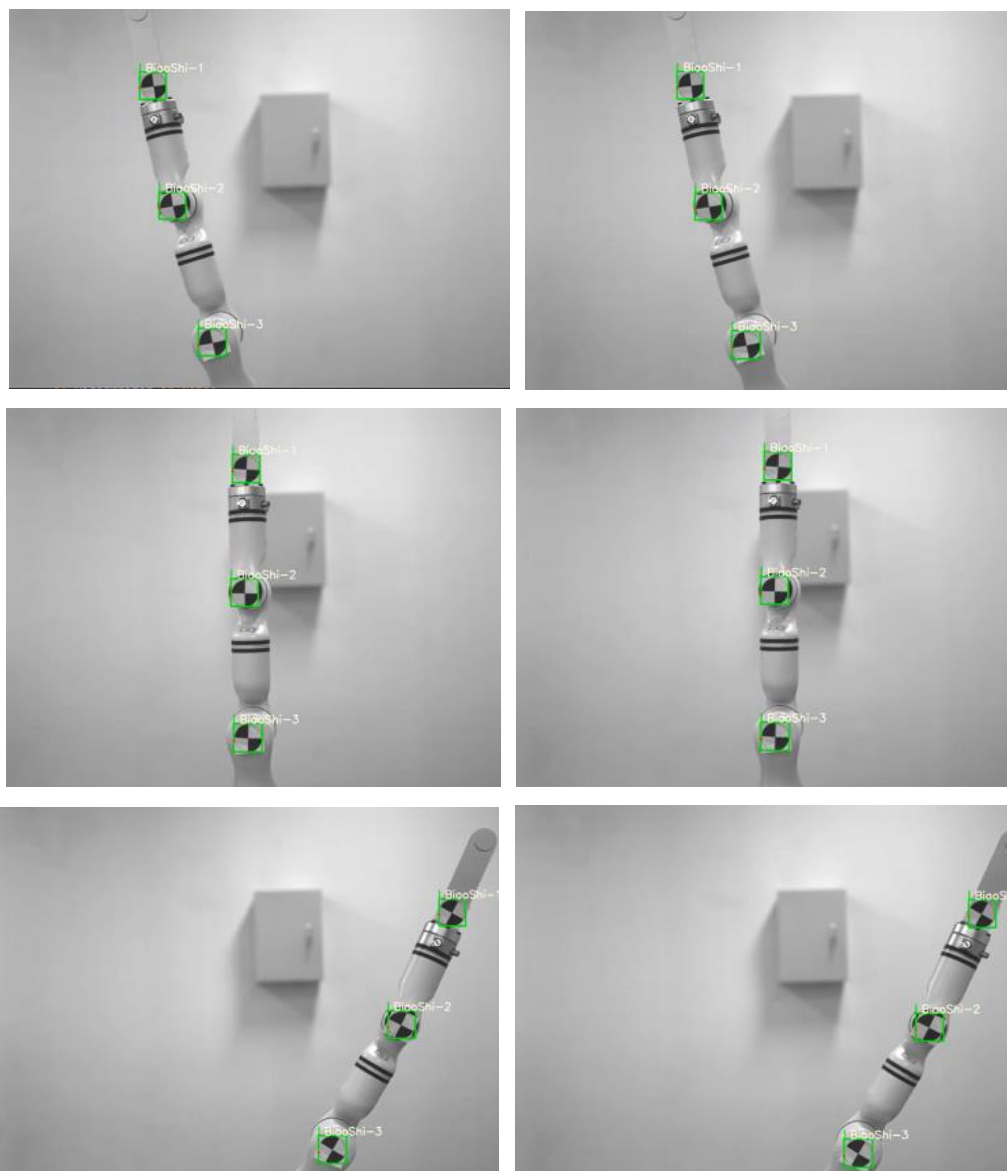
从图中可以看出在机器人从左至右地运动过程中,可以准确地检测并跟踪机器人本体上的三个目标,并且从开始到结束 ID 都没有发生变化,依旧保持着初始帧的顺序,说明多目标检测跟踪实验成功。

通过两个实验可以看出,本文设计的模型可以准确地实现目标的检测跟踪。并且,从跟踪效果来看,没有发生跟踪丢失、ID 混乱的现象。

(3) 双目多目标跟踪实验

首先在程序中导入左右相机对机器人拍摄的视频。查看在双目视觉系统中,目标的跟踪的效果。展示效果如下系列图所示。





从完整的视频检测完成结果可以看出，本文设计的算法在双目视觉系统中，依然可以保证多目标的跟踪定位效果。上图是截取检测视频中的几帧图片，在左右视频中目标检测并标注 ID 从上到下依次为“Biaoshi-1,Biaoshi-2,Biaoshi-3”，在整个检测序列中同一目标的 ID 依然保持不变，匹配效果好，说明双目同质多目标跟踪实验成功。

4.5 本章小结

针对多目标跟踪问题，本文设计了以卡尔曼滤波算法为核心的轨迹预测方法，对多目标的运动轨迹进行预测跟踪，提高跟踪效率；通过匈牙利算法，对前后两帧的多个目标进行匹配，同时运用外观信息和运动信息提高匹配准确性，保证了目标前后关联，确保了跟踪的准确性和连续性。通过实验证明算法的有效性。最

后，通过图像后处理技术运用亚像素边缘检测和线性拟合求交点的方法，对跟踪到的人工标识图像进行了中心点定位。

第5章 总结与展望

5.1 总 结

本文以机器视觉技术为载体, 研究探索在机器人全局结构运动双目测量过程, 实现张贴在机器人本体上的多个相同人工标识精确跟踪与定位的方法。根据跟踪定位的要求, 搭建了一套双目视觉检测实验平台。通过相关理论学习和分析, 制定符合实际情况的检测流程和方法。基于双目视觉理论基础的相应成像模型, 搭建了双目检测装置; 根据检测的速度要求, 确定了用深度学习方法为检测模型, 并分析了相关理论基础, 通过实验验证出检测的可行性; 采用估计预测和加权外观信息和运动信息的方式来进行目标跟踪; 利用提取边缘的亚像素坐标再结合线性拟合求交方法实现标识特征点的精确定位。最后, 采用 Python 语言结合成熟的 OpenCV 库及 Pytorch 深度学习框架完成相关程序编制。本文主要工作内容可以概括如下:

(1) 搭建了一套基于双目的同质多目标跟踪定位实验装置, 并自主设计一套用于检测跟踪的标识图, 便于后期定位处理。

(2) 重点研究和分析了多目标跟踪与定位的任务需求; 阐述了视觉系统的相关理论模型包括相机成像的几何模型和四大坐标系的相互转换关系; 并且针对双目视觉系统: 如双目视觉类型、双目视觉三维重建等方面理论做出详细介绍; 在目标检测跟踪方面, 阐述了图像预处理的重要性; 基于深度学习模型, 说明神经网络的组成部分和工作原理; 最后分析了本文研究的 YOLO 系列模型结构和损失函数。

(3) 为了实现机器人本体上人工标识的实时检测, 研究了以 YOLOv5 算法为基础的检测模型, 发现该目标检测算法在小目标检测时效果不理想、漏检率较高; 提出了在原有模型基础上引入 BottleneckCSP 模块并增加大尺度特征融合结构, 并且在模型结构中融合 SE 注意力机制以增强模型针对小目标的检测能力。最后实验发现, 改进后的算法对小目标的检测范围有所提升, 并且平均检测率较原始模型有 46% 的提高效果。

(4) 针对目标跟踪问题, 研究了以卡尔曼滤波算法为跟踪基础的多目标跟踪模型。将 YOLOv5 目标检测结果输入, 通过卡尔曼滤波的轨迹预测, 强化目

标跟踪效果；同时，为了增强目标跟踪的准确性和连续性，加入以外观信息和运动信息为辅助的匈牙利目标匹配方法。最后，在单目标跟踪实验以及多目标跟踪实验中可以看出目标的 ID 信息在整个视频序列中保持一致，证明算法的可靠性。在目标定位方面，采用亚像素边缘检测为基础，通过线性拟合直线求取直线交点坐标的方法求得人工标识的中心点坐标。

5.2 展 望

本文针对基于双目视觉的多目标跟踪定位方法进行了仔细研究。在目标检测阶段提出了基于 YOLOv5 模型为基础的改进提升小目标检测算法；在跟踪阶段通过卡尔曼滤波算法与匈牙利匹配算法相结合并且加入运动信息和外观信息提升跟踪效果。但是由于时间等因素，相关研究还有些可以提升或改进的空间，具体如下：

（1）针对目标检测模型，本文在原始 YOLOv5 基础上提出改进模型，检测效果虽然得到提升，能够满足本研究的需求。但是，为了达到更快更稳的检测效果，还需继续优化调整模型。

（2）在多人工标识的跟踪工作中，本文主要采用卡尔曼滤波方法探索多目标跟踪的方法，在目标跟踪时，不存在复杂的背景和遮挡等问题，能够达到目标跟踪的要求，但在其他跟踪任务中，遮挡和背景复杂等是较为普遍的现象，算法在此中状态下，还有待继续优化。

参考文献

- [1] 李杰, 李响, 许元铭, 等. 工业人工智能及应用研究现状及展望[J]. 自动化学报, 2020, 46(10): 2031-2044.
- [2] 袁雪. 计算机视觉[M]. 北京:清华大学出版社, 2021. 9.
- [3] 国务院. 《新一代人工智能发展规划》[Z]. 2017. 7.
- [4] Hou X y, Li C q, Fang Q. Computer vision-based safety risk computing and visualization on construction sites[J]. Automation in Construction, 2023, 156.
- [5] Noya M, S J P. The Influence of Phosphene Synchrony in Driving Object Binding in a Simulation of Artificial Vision[J]. Investigative ophthalmology visual science, 2023, 64(15): 5-5.
- [6] A M J, Qiu J w, L C D, et al. An Artificial Intelligence Platform for the Radiologic Diagnosis of Pulmonary Sarcoidosis: An Initial Pilot Study of Chest Computed Tomography Analysis to Distinguish Pulmonary Sarcoidosis from a Negative Lung Cancer Screening Scan[J]. Lung, 2023, 201(6): 611-616.
- [7] Wang Y b, Neil W d, Michael D c, et al. A brain MRI dataset and baseline evaluations for tumor recurrence prediction after Gamma Knife radiotherapy[J]. Scientific data, 2023, 10(1): 785-785.
- [8] Zhen X z, Zhen W s, Yu H g, et al. Object detection in 20 years: A survey[J]. IEEE, 2023, 111(3): 257-276.
- [9] 杨锋, 丁之桐, 邢蒙蒙. 深度学习的目标检测算法改进综述[J]. 计算机工程与应用, 2023, 59(11): 1-15.
- [10] 徐梦莹, 刘文波, 郑祥爱, 等. 基于深度学习的图像特征匹配方法[J]. 传感器与微系统, 2022, 41(07): 61-64.
- [11] 任条娟, 陈鹏, 陈友荣, 等. 基于深度学习的多目标运动轨迹预测算法[J]. 计算机应用研究, 2022, 39(01): 296-302.
- [12] 许德刚, 王露, 李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(08): 10-25.
- [13] Zhang Y, Hao Y s. A Survey of SAR Image Target Detection Based on Convolutional Neural Networks[J]. Remote Sensing, 2022, 14(24): 6240.
- [14] 袁国武, 陈志强, 龚健, 等. 一种结合光流法与三帧差分法的运动目标检测算法[J]. 小型微型计算机系统, 2013, 34(03): 668-671.
- [15] Chang D c, Fang M c. Bearing-Only Maneuvering Mobile Tracking with Nonlinear Filtering Algorithms in Wireless Sensor Networks[J]. IEEE Syst. J., 2014, 8(1): 160-170.
- [16] 王世勇, 乾国康, 李迪, 等. 面向边缘特征的实时模板匹配方法[J]. 华南理工大学学报(自然科学版), 2023, 51(09): 1-10.
- [17] 徐辉, 李海翔, 唐世轩, 等. 基于级联特征分类器的行人检测算法[J]. 实验室研究与探索, 2021, 40(02): 127-132.
- [18] 曹家乐, 李亚利, 孙汉卿, 等. 基于深度学习的视觉目标检测技术综述[J]. 中国图象图形学报, 2022, 27(06): 1697-1722.
- [19] 张博, 宋元斌, 吴冰, 等. 两阶段 Faster R-CNN 用于施工现场人车碰撞危险自动分析[J]. 制造业自动化, 2022, 44(06): 24-27+31.
- [20] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object

- Detection[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition , 2016: 779-788.
- [21] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]. CVPR, 2017: 6517-6525.
- [22] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. ARXIV, 2018.
- [23] 徐守坤, 王雅如, 顾玉宛, 等. 基于改进 Faster RCNN 的安全帽佩戴检测研究[J]. 计算机应用研究, 2020, 37(03): 901-905.
- [24] 闫建伟, 赵源, 张乐伟, 等. 改进 Faster-RCNN 自然环境下识别刺梨果实[J]. 农业工程学报, 2019, 35(18): 143-150.
- [25] 顾超越, 李喆, 史晋涛, 等. 基于改进 Faster-RCNN 的无人机巡检架空线路销钉缺陷检测[J]. 高电压技术, 2020, 46(09): 3089-3096.
- [26] 裴伟, 许晏铭, 朱永英, 等. 改进的 SSD 航拍目标检测方法[J]. 软件学报, 2019, 30(03): 738-758.
- [27] 鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(07): 253-260.
- [28] Peng H x, Liang M, Yuan C, et al. EDF-YOLOv5: An Improved Algorithm for Power Transmission Line Defect Detection Based on YOLOv5[J]. Electronics, 2023, 13(1): 148.
- [29] 胡亮, 段慧仙, 裴浩东, 等. 非合作航天器关键特征的双目视觉位姿测量[J]. 中国激光, 2022, 49(17): 56-64.
- [30] 张超, 张旭辉, 杜昱阳, 等. 基于双目视觉的悬臂式掘进机位姿测量技术研究[J]. 煤炭科学技术, 2021, 49(11): 225-235.
- [31] 周凡桂, 王晓光, 高忠信, 等. 双目视觉绳系支撑飞行器模型位姿动态测量[J]. 航空学报, 2019, 40(12): 44-54.
- [32] 束安, 裴浩东, 丁雷, 等. 空间非合作目标的双目视觉位姿测量方法[J]. 光学学报, 2020, 40(17): 107-117.
- [33] 徐涛, 马克, 刘才华. 基于深度学习的行人多目标跟踪方法[J]. 吉林大学学报(工学版), 2021, 51(01): 27-38.
- [34] 张星. 运动目标的稳定跟踪算法研究[D]. 中国科学院大学(中国科学院西安光学精密机械研究所), 2019.
- [35] 刘彩虹, 张磊, 黄华. 交通路口监控视频跨视域多目标跟踪的可视化[J]. 计算机学报, 2018, 41(01): 221-235.
- [36] 任珈民, 宫宁生, 韩镇阳. 基于 YOLOv3 与卡尔曼滤波的多目标跟踪算法[J]. 计算机应用与软件, 2020, 37(05): 169-176.
- [37] Xu Y, Ban Y, Delorme G, et al. TransCenter: Transformers with dense representations for multiple-object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 45: 7820-7835.
- [38] Chu P, Wang J, You Q, et al. Transmot: Spatial-temporal graph transformer for multiple object tracking[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 4870-4880.
- [39] 周辉. 基于深度学习的多目标跟踪算法研究[D]. 电子科技大学, 2018.
- [40] Yu F w, Li W b, Li Q q, et al. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature. [J]. ECCV Lecture Notes in Computer Science, 2016, 9914: 36-42.
- [41] Yu S, Meng D, Zuo W, et al. The Solution Path Algorithm for Identity-Aware Multi-object Tracking[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and

- Pattern Recognition, 2016: 3871-3879.
- [42] Bae S h, Yoon K j. Robust Online Multi-object Tracking Based on Tracklet Confidence and Online Discriminative Appearance Learning[C]. Computer Vision & Pattern Recognition. IEEE, 2014: 1218-1225.
- [43] Milan A, Rezatofighi H s, Dick R a, et al. Online multi-target tracking using recurrent neural networks. [C]. AAAI, 2017: 4225-4232.
- [44] 李星辰, 柳晓鸣, 成晓男. 融合 YOLO 检测的多目标跟踪算法[J]. 计算机工程与科学, 2020, 42(04): 665-672.
- [45] 张展. 基于双目视觉的三维重建关键技术研究[D]. 中国科学院大学(中国科学院沈阳计算技术研究所), 2019.
- [46] 宋子豪. 基于双目立体视觉的汽车测距避障和目标识别研究[D]. 华中科技大学, 2020.
- [47] 张小迪. 面向机器人末端位姿的视觉测量跟踪技术研究[D]. 南京航空航天大学, 2020.
- [48] 董文轩, 梁宏涛, 刘国柱, 等. 深度卷积应用于目标检测算法综述[J]. 计算机科学与探索, 2022, 16(05): 1025-1042.
- [49] 俞军, 贾银山. 改进 YOLOv5 的小目标检测算法[J]. 计算机工程与应用, 2023, 59(12): 201-207.
- [50] 陈范凯, 李士心. 改进 yolov5 的无人机目标检测算法[J]. 计算机工程与应用, 2023: 1-9.

攻读硕士学位期间主要成果

发表的学术论文

- [1] Wu L l, Liu Y, Wang Y, et al. Robot 3D spatial motion measurement via vision-based method [J]. Measurement Science and Technology, 2023, 34(6): 065003.(SCI, 导师一作, 本人二作)
- [2] 刘艺, 吴路路. 基于改进 YOLOv5s 小目标检测算法研究 [J]. 安徽科技学院学报, 2024. (已录用, 本人一作)

致 谢

二十载求学路将尽，叹时光易逝，惜韶华难追。我的研究生生活，始于 2021 年，终于 2024 年。遥想初遇，江城盛夏，满怀期待踏入安工程；三度春秋，步履未停，目光所及皆是回忆。花开花落万物道，聚散离别终有时。人生如逆旅，感恩每一次相遇与经历。

落其实者思其树，学其成时念吾师。感谢我的导师吴路路副教授。读研期间，吴老师在科研方面给予了我极大的帮助和指导，不仅常常给我分享经典文献，督促我认真学习，还会鼓励我独立思考和发现问题，悉心指导我论文写作。此次毕业论文从选题到最终定稿，吴老师耐心地给我提出许多修改建议，不断督促我完善精进，其认真严谨的治学态度和对待学术研究的满腔热忱深深地影响着我。片纸有尽，而感激之情无尽。再次向我的导师致以深深的感谢和由衷的祝福！

时光清浅，有幸相遇。感谢我的女朋友余欣悦从大学到研究生的陪伴和照顾，感谢你对我的耐心和理解，感谢你带给我乐观和阳光，感谢我们互相鼓励携手并进的日子。愿未来，心存希冀，目有繁星，追光而遇，沐光而行。愿往后余生，冷暖有相知，喜乐有分享，同量天地宽，共度岁月长。

春晖寸草，山高海深，父母之爱，得以成才。感谢我的父母这么多年的无私付出与爱护。感谢你们给予我无尽的爱与包容，因为有你们，我才能在真诚与幸福中快乐成长。感谢你们给予我充分的支持和肯定，因为有你们，我才能在一次次迷茫中重拾信心勇往直前。愿你们身体健康，平安喜乐。

言辞有尽，谢意难全。三载寒暑，无论喜悦还是酸楚，所有经历，皆是礼物。风有起止，人有聚散，感恩所有遇见。行文至此，落笔为终。回首看，轻舟已过万重山；向前看，前路漫漫亦灿灿。愿祖国山河无恙，愿你我未来可期。

尚德敏学 唯实惟新

