分类号：TP391　　学校代码：10363

密　级：公开　　学　　号：2220110114

安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于双目视觉的零件精确识别与定位方法研究

论 文 作 者　王宜红

校 内 导 师　疏 达　（教授）

校 外 导 师　肖永强（高级工程师）

专业学位类别　机械

研究方向(领域)　机器视觉

论文提交日期：2025 年 5 月 20 日

分类号：TP391　　单位代码：10363

密　级：公开　　学　　号：2220110114

基于双目视觉的零件精确识别与定位方法研究

RESEARCH ON PRECISE RECOGNITION AND LOCALIZATION METHOD OF PARTS BASED ON BINOCULAR VISION

学生姓名：王宜红

指导教师：疏 达 （教授）

校外导师：肖永强 （高级工程师）

专　　业：机械

研究方向：机器视觉

论文答辩日期：2025 年 5 月 9 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：王宜红

日期：　2025　年 5 月 19 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：王宜红 指导教师签名：疏达

日期： 2025 年 5 月 19 日　　日期： 2025 年 5 月 19 日

基于双目视觉的零件精确识别与定位方法研究

摘 要

随着工业生产向智能化方向快速发展，工业机器人等自动化装备在生产制造中得到广泛应用。在机器人参与自动化生产的过程中，零部件的识别、定位与抓取是至关重要的环节，其识别与定位的准确性对生产作业的质量和效率有重要的影响。然而，面对散放、堆叠和遮挡的复杂码垛场景下，存在着零件的漏检、误检以及检测慢等问题，对零件的精准识别与定位提出了严峻的挑战。本文主要面向机器人码垛作业场景，针对复杂场景下的零件识别与定位进行了研究，主要研究内容如下：

(1) 针对在复杂场景下存在着零件的漏检、误检以及检测慢等问题，提出了一种基于改进 YOLOv8 零件识别算法。首先，通过线下拍摄采集零件图片，并将其划分为数据集；其次，改进 YOLOv8 零件识别算法，将改进后的算法命名为 YOLOv8-MSC；最后，通过对比不同算法进行实验，结果表明，改进后的算法 YOLOv8-MSC 相比与原始 YOLOv8 算法，其 mAP 值提升了 1.3%，精度提升了 1.1%，检测速度加快了 38%，解决了 YOLOv8 算法对零件的识别存在误检、漏检以及检测速度慢等问题。

(2) 在零件定位方面，采用 SGBM 立体匹配算法对零件开展研究。首先采用张正友标定法对双目相机进行初步标定，获取相机参数后进行后续的校正处理；其次通过对比分析 GC、BM 和 SGBM 三种立体

匹配算法开展了实验。结果显示，SGBM 立体匹配算法能获取更多零件深度信息，定位距离精度更高且误差更小。因此，本文选用 SGBM 立体匹配算法作为零件定位算法。

(3) 基于 YOLOv8-MSC 零件识别算法与 SGBM 立体匹配算法，本文设计了一种零件识别与定位方法的系统，利用 PyCharm 与 Qt 相结合，做出零件识别与定位的系统界面。首先对零件识别方法的可行性进行验证，对不同场景与距离下的零件进行识别实验，实验结果表明在不同距离与场景下，零件识别的精度随着距离的增加所下降，但其平均精度仍达到 85%以上，fps 值稳定在 18 至 21 之间；其次对零件的三维空间定位做了实验验证，由实验结果可知，X、Y、Z 三轴方向上的平均定位误差值分别为 1.6mm、1.4mm、1.3mm，其定位误差值与定位误差率均在合理范围之内，由此表明本文所采用的零件识别与定位方法具有较好的准确性与实时性，满足码垛机器人面对复杂环境下码垛作业任务。

关键词：零件识别与定位，YOLOv8-MSC，双目视觉，立体匹配，码垛作业

RESEARCH ON PRECISE RECOGNITION AND LOCALIZATION METHOD OF PARTS BASED ON BINOCULAR VISION

ABSTRACT

As industrial production rapidly evolves towards intelligent automation, robotic arms and other automated equipment are increasingly utilized in manufacturing processes. In the realm of automated production involving robots, the accurate identification, positioning, and grasping of components are critical steps. The precision of these tasks significantly impacts the quality and efficiency of production operations. However, in complex palletizing scenarios characterized by scattered, stacked, and occluded parts, challenges such as missed detections, false positives, and slow detection rates arise, posing formidable obstacles to the precise recognition and positioning of components. This paper primarily focuses on robotic palletizing scenarios, investigating the issue of precise visual recognition and positioning of components within such contexts. The main research contents are as follows:

(1) To address the issues of missed detection, false detection, and slow detection of parts in complex scenes, a modified YOLOv8 part recognition algorithm is proposed. Firstly, images of parts are collected through offline

photography and divided into datasets; secondly, the YOLOv8 part recognition algorithm is improved, and the modified algorithm is named YOLOv8-MSC; finally, experiments are conducted by comparing different algorithms. The results show that, compared to the original YOLOv8 algorithm, the improved YOLOv8-MSC algorithm has increased the mAP value by 1.3%, improved accuracy by 1.1%, and sped up the detection speed by 38%. This solves the problems of false detection, missed detection, and slow detection speed of parts by the YOLOv8 algorithm.

(2) In terms of part positioning, this paper adopts the SGBM stereo matching algorithm to conduct research on parts. Firstly, the Zheng you Zhang calibration method is used to calibrate the binocular camera, and the camera parameters are obtained for correction; Secondly, experiments are carried out by comparing the GC, BM and SGBM three-dimensional matching algorithms. The results show that the SGBM stereo matching algorithm can obtain more depth information about parts, has higher positioning distance accuracy and smaller errors. Therefore, this paper selects the SGBM stereo matching algorithm as the part positioning algorithm.

(3) Based on the YOLOv8-MSC part recognition algorithm and the SGBM stereoscopic matching algorithm, this paper designs a system for part recognition and positioning method, and creates a system interface for part recognition and positioning by combining PyCharm and Qt. Firstly,

the feasibility of the part recognition method is verified, and part recognition experiments are carried out under different scenarios and distances. The experimental results show that the accuracy of part recognition decreases with the increase of distance under different distances and scenarios, and the average accuracy still reaches more than 85%, with the fps value stabilizing between 18 and 21; Secondly, experimental verification is carried out for the three-dimensional positioning of parts. According to the experimental results, the average positioning error values in the X, Y and Z axis directions are 2.3mm, 2.1mm and 1.5mm respectively, and both the positioning error value and the positioning error rate are within a reasonable range. This indicates that the part recognition and positioning method adopted in this paper has good accuracy and real-time performance, and meets the palletizing operation tasks of palletizing robots in complex environments.

KEY WORDS: parts recognition and localization, YOLOv8-MSC, binocular vision, stereo matching, palletizing operation

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 视觉目标识别技术研究..... 2
1.2.2 双目视觉目标定位技术研究..... 5
1.3 本文的主要研究内容..... 6
第 2 章 双目视觉识别与定位的理论基础..... 7
2.1 基于深度学习的视觉目标检测理论基础..... 7
2.1.1 卷积神经网络..... 7
2.1.2 目标检测算法..... 10
2.2 双目视觉目标定位理论基础..... 12
2.2.1 四大坐标系及其转换关系..... 12
2.2.2 相机成像模型..... 15
2.2.3 双目相机测距原理..... 16
2.3 本章小结..... 17
第 3 章 基于改进 YOLOv8 目标识别研究..... 18
3.1 YOLOv8 算法原理..... 18
3.1.1 YOLOv8 网络结构..... 18
3.1.2 损失函数..... 22
3.2 YOLOv8 改进策略..... 24
3.2.1 MobileNetv3 轻量化网络结构..... 25
3.2.2 引入注意力机制..... 28
3.2.3 优化损失函数..... 31
3.3 实验结果分析..... 33
3.3.1 实验数据集采集..... 33
3.3.2 数据集的构建..... 34
3.3.3 实验环境配置..... 34
3.3.4 评价指标..... 35
3.3.5 网络轻量化改进与分析..... 36
3.3.6 不同注意力机制对比..... 36
3.3.7 优化损失函数结果分析..... 37
3.3.8 不同模型对比实验..... 39

3.3.9 实验结果分析.....40
3.4 本章小结.....41
第 4 章 基于双目立体视觉定位方法研究.....42
4.1 双目立体视觉系统的组成.....42
4.2 双目标定.....42
4.3 双目校正.....46
4.4 双目标定与校正实验.....46
4.5 立体匹配.....50
4.5.1 匹配基元.....51
4.5.2 约束准则.....52
4.5.3 BM 局部立体匹配算法.....52
4.5.4 GC 全局立体匹配算法.....53
4.5.5 SGBM 半全局立体匹配.....53
4.5.6 立体匹配算法对比与分析.....54
4.6 本章小结.....56
第 5 章 零件识别与定位实验分析.....57
5.1 实验环境.....57
5.2 零件识别与定位系统设计.....57
5.2.1 零件识别实验.....58
5.2.2 零件三维空间定位实验.....61
5.2.3 定位误差原因分析.....65
5.3 本章小结.....65
第 6 章 总结与展望.....66
6.1 总结.....66
6.2 展望.....66
参考文献.....68
攻读硕士期间发表学术论文和成果情况.....74
致谢.....75

第 1 章 绪 论

1.1 研究背景及意义

随着全球制造业向智能化、自动化方向加速转型，工业机器人技术成为推动生产效率提升的核心驱动力之一[1]。尤其在物流、汽车制造、食品饮料、医药等行业中，码垛作为生产流程中重复性高、劳动强度大的关键环节，其自动化需求日益迫切。据统计，2025 年全球码垛机器人市场规模预计突破 150 亿元，年复合增长率达 25%以上[2]。这一增长不仅源于劳动力成本的上升，更得益于人工智能、物联网(IoT)、机器视觉等技术的深度融合，使得码垛机器人从传统示教编程向智能化、自适应化方向演进[3]。

传统码垛作业依赖人工操作或固定程序控制的机械臂，存在效率低、灵活性差、无法适应复杂工件等问题[4]。如图 1-1 所示，在机床加工场景中，若工件尺寸、形状大小发生变化[5]，传统机器人需重新示教编程，导致生产线停机时间增加，直接影响企业产能。而现代工业生产线对柔性化、高精度和实时响应的需求，催生了零件识别与定位技术的快速发展，使其成为智能码垛系统的核心技术之一[6]。零件识别与定位技术是码垛自动化中的核心环节，其目标是通过高精度感知与智能决策，实现工件的快速分类、位置确定及抓取路径规划。早期的识别技术主要依赖光电传感器或简单的图像处理算法，但受限于环境光照、工件堆叠复杂度等因素，难以满足高精度要求。例如，在堆叠工件场景中[7]，传统边缘检测算法易受噪声干扰，导致定位偏差较大。近年来，深度学习与机器视觉的结合为这一领域带来了革命性突破。基于卷积神经网络(CNN)的目标检测模型(如 YOLO、Faster R-CNN)能够从复杂背景中提取工件的多维度特征，实现高精度分类与定位[8]。如图 1-2 所示，广州数控 GSKGR-C 机器人通过部署深度学习驱动的视觉系统，使码垛机器人对尺寸为 60mm×10mm 和 55mm×10mm 的钢制盘盖类零件的识别准确率提升至 99.5%以上，同时定位误差控制在±0.4mm 以内[9]。此外，迁移学习技术的应用进一步降低了模型训练对数据量的依赖，加速了工业场景的落地。

图 1-1 机器人机床加工

图 1-2 广州数控 GSKGR-C 机器人

尽管技术发展迅速，零件识别与定位技术在工业级码垛应用中仍然面临一些问题，如在零件散放、堆叠与遮挡等复杂场景下，存在零件的漏检、误检以及检测慢等。为了解决码垛作业中以上所面临的问题，针对码垛场景下零件的识别与定位任务，展开对零件的识别与定位方法研究。零件识别与定位的准确率将决定了机器人能否高效地完成码垛作业[10]。

1.2 国内外研究现状

近年来，随着计算机视觉与人工智能相结合为码垛领域注入了新活力。基于深度学习的视觉系统能够通过高分辨率相机、3D 传感器等多元数据融合，实时解析零件的几何特征、空间位姿及堆叠状态，并结合机器人运动控制算法实现自主抓取与码放[11]。

1.2.1 视觉目标识别技术研究

在现代制造业中，零件识别是生产流程中的关键环节，对于保证产品质量、提高生产效率以及实现制造过程的可追溯性具有至关重要的作用。随着计算机视觉和深度学习等前沿科技的迅猛发展，零件识别技术实现了突破性进展。在工业领域中零件识别研究对许多生产场景的工作效率有着显著的提升。比如机器人的零件抓取[12]、汽车制造领域、自动化装配、物流运输、码垛作业等众多应用场景。针对零件识别研究主要分为传统零件识别算法研究与深度学习零件识别算法研究。零件的识别研究经历了从传统人工特征提取到现代深度学习算法特征提取的演变过程。在传统方法中，研究人员需要手动设计关键特征描述符，通过提取这些预设特征进行比对来实现目标识别[13]。基于深度学习零件识别算法研究，通过大规模标注数据集进行多层次特征学习，构建从局部细节到全局结构的特征表征模型，从而实现目标的识别。

(1) 传统算法的目标识别研究

在零件识别研究领域，早期的技术路线主要依赖于传统目标检测算法。许鑫杰[14]等人利用 Canny 边缘检测算法对零件图像进行边缘提取，通过多尺度梯度计算与非极大值抑制技术获取精确的边缘轮廓信息，并结合边缘特征描述实现零件的精准识别与检测。孙小权[15]等人利用摄像头拍摄出料口位置信息，并对拍摄收集到的图片进行主要分析与小变化法特征提取，然后将提取到的特征输入 SVM 中识别出零件状态，最后结合机械臂实现自动送料过程。方祝平[16]等人先处理收集到的零件图片，然后使用灰度值和金字塔匹配算法模板，该算法成功实现了对多种工业零件的精确识别，其准确率达到 94%至 97%。TU-Cheng You[17]设计了一种零件分拣系统，首先利用机器视觉对零件进行统一预处理，接着提取零件的面积与弧度特征作为零件判别依据，从而实现对零件的类别预测。由于受光照影响薄壁零件在识别检测时效率较为低下，毛向向[18]等人提出了一种基于改进 Retinex 算法去除了光照对零件特征的影响，利用 Canny 边缘检测算法提取了零件的边缘轮廓，极大的提高了薄壁零件在光照环境下检测的效率。针对电子零件的识别，Lee[19]等人利用 Harris 与主元分析法对其进行识别。司小婷[20]设计了一种多特征融合技术的零件识别模型，通过几何特征、轮廓特征和形状特征的三维特征空间构建，建立了层次化的零件表征模型。首先，采集零件模板图像并提取特征建立零件库；其次，通过特征匹配算法将待测零件与零件库中的模板进行相似度计算，实现高精度零件识别。实验表明，该方法在复杂工况下的识别准确率达到 97.3%。xin[21]采用基于兴趣点(Interest Points)的检测技术，通过计算机视觉算法对机械零件进行自动化检测，提高了检测的效率和准确性。

传统识别算法虽然在一定程度上能够解决计算机视觉中的一些问题，但仍存在着特征提取不够智能、计算复杂度高、速度慢、对复杂环境适应性差以及数据需求大和可扩展性差等缺点。面对复杂的零件识别环境其识别精度与检测速度都无法满足机器人码垛作业要求。

(2) 深度学习算法的目标识别研究

相较于传统手工设计特征的方法，深度学习模型通过自主学习，能够自动提取多层次、高判别性的特征表达，显著提升复杂场景下的特征表征能力。在实际生产中，零件可能处于不同的光照条件、角度或存在部分遮挡的情况。深度学习

模型通过大量的数据训练，可以学习到在不同条件下零件的特征变化，面对复杂的工业环境，依然可以保持较高的识别准确率。

Tao[22]提出了一种名为多模式聚集特征增强网络(MAFENet)的网络，该网络基于 SSD 算法，其检测精度达到了 97.87%。宋栓军[23]等人利用 K-menns++聚类算法，改进了 YOLOv3 算法，添加了一个特征尺度，在识别小型零件时，其检测精度比原始值提升了 4.87%。面对多种工业零件，郭斐[24]等人提出了一种改进 Fast R-CNN 零件识别算法，简化了算法中的卷积层数，同时添加 Inception 结构，改进后的算法不仅减少了模型的参数量还使检测精度达到了 98.8%。余永维[25]等人在 SSD 网络框架中融合了 Inception 网络结构，并优化了损失函数，使用残差结构连接，提升了模型的鲁棒性和检测的准确率。在面对小目标检测时，韩强[26]提出改进 YOLOv8 目标检测算法，其在 YOLOv8 网络结构中引入了可变性卷积模快，加强了对小目标区域的检测，改善了误检和漏检，同时在颈部网络中引入卷积块注意力模块(CBAM)，通过通道注意力与空间注意力的协同作用，提升了模型对关键特征的提取能力，设计优化了损失函数，与原始实验相比其模型识别精度提升了 4.7%。

在现代工业智能化的背景下，零件识别检测的速度已成为决定生码垛作业的关键因素。高效的零件识别系统是实现工业自动化的重要基础，其关键在于实时高效地定位和准确分类众多零部件，从而确保码垛系统的运行效率与企业产能的提升。此外，随着物联网技术的快速发展，大量的应用场景需要网络模型部署在小型设备上，这对模型轻量化改进和计算效率提出了更高的标准。为了提升零件识别检测速度，曾毅星[27]提出了一种增加现实应用的轻量级零件识别研究。其对 YOLOv3 网络中的主干网络做出了可分离卷积操作的改进，同时引入了 MobileNetv2 与 v3 中的注意力机制和倒残差结构，在保证模型轻量化的同时其检测精度也有所提升。徐维[28]利用混合注意力机制增强特征提取能力，同时使用集成 Inception 模块的轻量化 MobileNetv2 替代 YOLOv4 的主干网络，显著减少了模型的计算复杂度，降低了对硬件资源的依赖，改进后的网络精度有所下降，为了弥补精度下降问题，作者融合了 ASFF 结构增强了 PANet 网络的检测能力，同时使用 K-means++优化聚类算法，使检测精度得到大幅提升。

综上所述，国内外研究人员不断致力于开发高精度与轻量化的零件识别算法，

但由于码垛任务中环境复杂性和机器人操作误差等挑战，传统识别方法已显现出局限性。相比之下，深度学习方法凭借其强大的特征提取能力和环境适应性，在复杂工况下展现出显著优势。因此，本文旨在通过结合基于深度学习的目标检测算法与双目立体视觉技术，以提高模型在复杂环境中的检测精度与速度。

1.2.2 双目视觉目标定位技术研究

在零件被准确识别之后，需要对零件进行三维空间定位为码垛机器人提供零件的位置信息，从而完成有效的抓取作业任务。在位姿检测方法的研究中，龚友彬[29]等人提出了一种新的工件空间定位方法，该方法深度融合了三维空间定位技术与双目视觉的立体感知优势，通过计算工件特征点在双目图像中的对应关系，实现了对工件空间姿态的准确测量。针对散乱工件识别与定位问题，王正波[30]等人提出了一种基于机器视觉的改进方法，通过优化 SURF(快速稳健特性)算法，结合视觉词袋(BOW)和 SVM 分类器实现工件识别与定位。此外，引入 Kinect 深度传感器进行辅助定位，实验结果表明其定位精度误差小于 5mm。Saleem[31]等人，通过对 KITTI 数据集的实验分析可以发现，在场景复杂度不高的情况下，双摄像头系统比单摄像头系统在目标定位方面表现更出色。针对 Ladybug3 相机的特殊分布结构，结合立体视觉原理，曹文君[32]设计了一种基于单目和双目视觉融合的目标测距方法。该方法通过实时分析目标物与视觉系统的空间位置关系，灵活选用单目或双目测量方式，这种动态调整方法不仅提升了测量的准确性，还显著降低了因图像畸变导致的测量误差。李佳[33]针对低亮度环境下的图像处理问题，提出了一种基于校正映射表和 BM 立体匹配算法的解决方案。该方法通过预先建立的坐标转换映射关系，可直接输出经过几何校正的图像坐标数据，并利用 BM 算法计算视差信息，最终结合双目测距原理实现了目标深度距离的精确测量。冯英旺[34]提出了一种基于 SURF 双目图像实时测距方法，通过匹配同一时刻左右图像中的 SURF 特征点，结合双目视觉几何关系，计算环境特征点的视差信息，进而获取目标的景深数据，最终实现了目标相对位置的实时测量。毛先胤[35]等通过 SGBM 算法计算双目视差，并利用 WLS 滤波算法进行优化处理。利用优化后的视差信息，结合重投影变换计算目标物体的深度值，从而实现目标在三维空间中的精确定位。

本文对双目视觉目标定位的研究现状展开分析，发现双目立体视觉系统在效

率、准确性以及适用性方面具有显著优势，且适用性更广。因此，本研究采用双目立体视觉技术作为主要架构，并结合改进后的 YOLOv8 目标识别检测算法，从而实现了零件的精确识别与定位。

1.3 本文的主要研究内容

本文以码垛场景中的木质零件为实验对象，通过融合深度学习检测算法与双目立体视觉技，实现了零件的准确识别与三维空间位置定位，能有效的为码垛机器人提供视觉信息，主要研究内容如下：

(1) 双目相机标定：在双目相机标定过程中，本文采用张正友标定法，利用线性约束完成双目校正，使双目相机标定结果更加准确。

(2) 零件的目标识别：以 YOLOv8 模型为基础做了以下三个方面改进，首先采用 MobileNetv3 替换 YOLOv8 的主干网络；其次在主干网络中第四层与第十层分别引入 CBAM 注意力机制；最后采用 SIoU 损失函数作为模型的损失函数。由对比消融实验结果表明，改进后的模型在检测精度和速度方面均有提升。

(3) 双目立体视觉定位：首先阐述了双目立体视觉的基本原理，随后针对零件图像特征开展了立体匹配研究。最后通过实验对比可知，在不同的场景下 SGBM 立体匹配算法匹配效果更好，且测量的距离更加准确。

(4) 零件的识别与定位实验：介绍了双目视觉识别与定位系统的流程，搭建了零件识别与定位平台，使用双目相机对零件在不同距离与场景下进行了识别与定位实验并对定位误差进行了分析。

第 2 章 双目视觉识别与定位的理论基础

本章详细介绍零件识别与定位算法的理论基础，针对码垛场景下，研究采用深度学习目标检测算法与双目立体匹配算法相结合的技术路线。为深入理解零件识别定位的技术理论基础，本章将系统性地讲解相关算法的核心原理与实现机制。

2.1 基于深度学习的视觉目标检测理论基础

2.1.1 卷积神经网络

卷积神经网络(CNN)是一种广泛应用于计算机视觉领域的深度学习模型。它通过卷积层、池化层和全连接层的组合，能够自动提取图像中的多层次特征，从而在物体检测、目标识别、图像分割等任务中表现出色，其结构如图 2-1 所示。

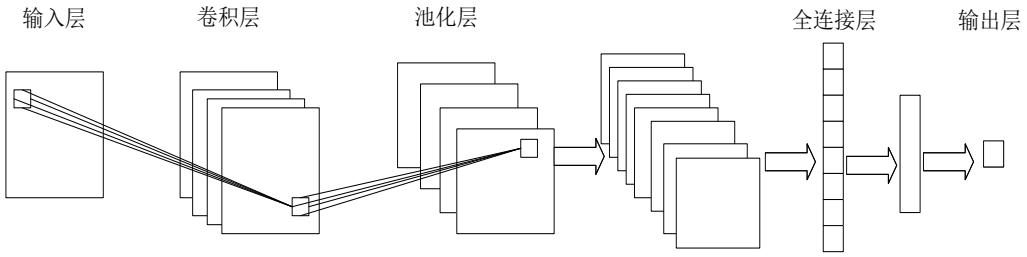


图 2-1 卷积神经网络结构图

1. 输入层

卷积神经网络(CNN)的输入层是数据进入网络的第一层，它负责接收原始数据并对其进行预处理，这些图像数据通常具备多个颜色通道，进而构成一个二维或者三维的矩阵形式。

2. 卷积层

在卷积神经网络中，卷积层是提取图像特征的核心部分，其主要依靠卷积操作来获取图像中的局部特征信息。该操作通过对输入图像的局部区域与卷积核，执行逐元素乘积累加，生成反映图像特征的特征图。卷积核是一个矩阵，其数值代表在卷积计算时采用的权重。卷积核在输入图像上滑动，每次移动的步长决定了它覆盖图像的区域大小。卷积层公式如 2-1 所示。

$$X_j^i = f\left(\sum_{i \in M_j} X_j^{i-1} \times K_{ij}^i + b_j^i\right) \tag{2-1}$$

式中：X_j^i为第i层中的第 j 个特征图；k 为卷积核；f 为激活函数，M_j 为输入

特征图集合；*b* 为偏置项。图 2-2 所示，是一个输入特征图尺寸为 5×5、卷积核尺寸为 3×3、步长为 1 的标准卷积运算过程。

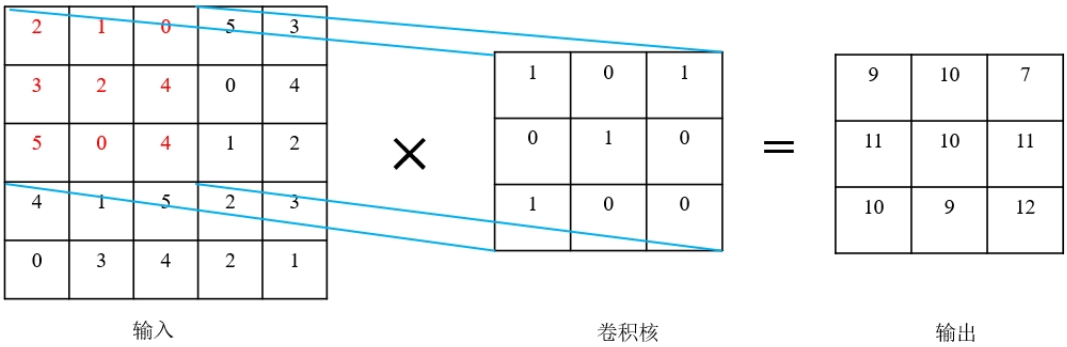


图 2-2 卷积操作示意图

3. 池化层

池化层的作用是降低特征图的空间维度，从而简化模型并减少计算需求，同时需要保持图像的关键特征不受损失。池化操作主要包括最大池化(Max pool)和平均池化(Avg pool)两种，其中，最大池化通过对特征图的局部区域进行最大值采样，选取该区域内的最大特征值作为输出，而平均池化则计算该区域中特征值的平均值。具体过程如图 2-3 所示。

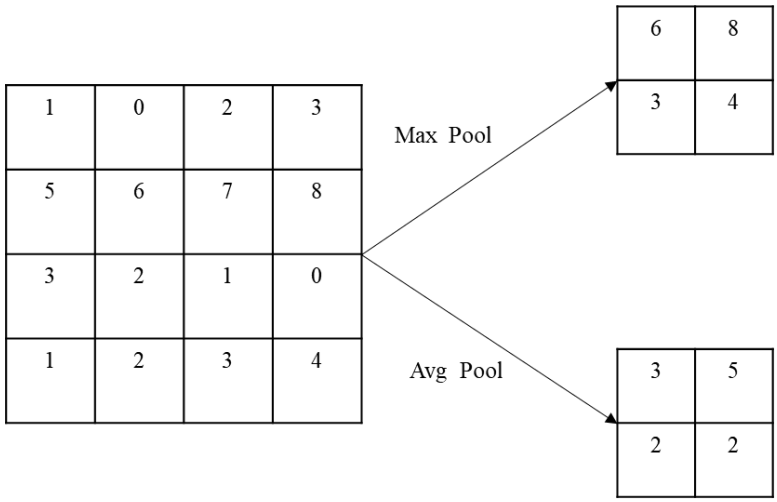


图 2-3 池化操作示意图

4. 全连接层

在卷积神经网络的最后阶段，通常会有一个或多个全连接层，其作用是将先前层级提取的特征转化为最终结果，如类别判定或数值预测。这类层的运行原理与传统神经网络类似，表现为层内每个节点均与前一层的全部节点相连。全连接层示意图如图 2-4 所示。

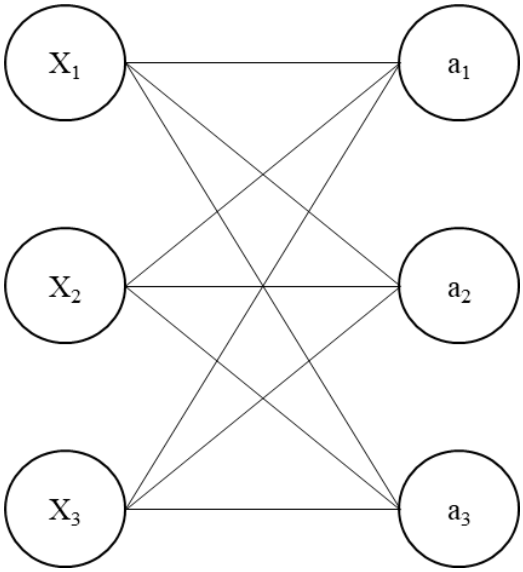


图 2-4 全连接层示意图

5. 激活函数

在人工神经网络中，激活函数作为核心组件，其主要功能是引入非线性变换特性，使神经网络具备学习和建模复杂数据模式的能力。目前广泛使用的激活函数主要包括以下三类：Sigmoid 函数(Logistic 函数)、双曲正切函数(Tanh)以及线性整流函数(ReLU)。其中，Sigmoid 函数将输入映射到(0,1)区间，Tanh 函数输出范围为(-1,1)，而 ReLU 函数则通过保留正输入值、抑制负输入值的方式实现非线性转换，其公式分别如 2-2、2-3 与 2-4 所示，函数图像如图 2-5 所示。

$$f(x)=\frac{1}{1+e^{-x}} \tag{2-2}$$

$$f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2-3}$$

$$f(x)=\max(0,x) \tag{2-4}$$

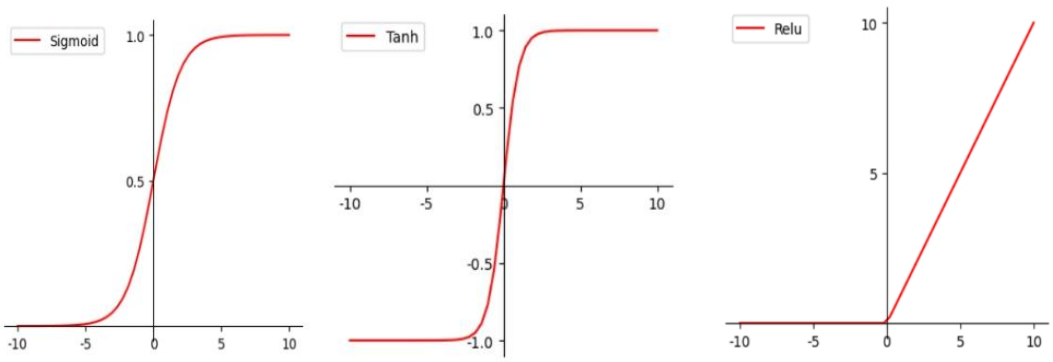


图 2-5 激活函数图像

2.1.2 目标检测算法

经典的目标检测算法主要分为单阶段目标检测算法和双阶段目标检测算法两大类别。其中双阶段目标检测算法主要代表有 RCNN、R-FCN[36]和 Faster RCNN[37],它们检测图像数据时分为两步，第一步要预选出目标所在的区域，也是粗略的检测出目标所在的区域；第二步设置检测条件，进一步的对目标所在的区域进行细化分，从而更加准确检测出目标所在的区域。双阶段目标检测算检测准确率高但检测的速度慢。单阶段目标检测算法主要有 YOLO 系列和 SSD。它们只需要进行单次识别检测就能输出检测的结果。单阶段目标检测算法省去了预选这一步骤，这使得检测速度更快但检测精度要略逊于一些双阶段目标检测算法。下面重点讲述 Faster RCNN、SSD 和 YOLO 系列目标检测算法。

1. Faster RCNN 算法

Faster RCNN 是一种基于深度学习的目标检测算法，其整体架构由三个主要部分组成：卷积神经网络(CNN)用于特征提取；区域建议网络用于生成候选区域；边界框回归和分类器用于精确定位和识别对象。其网络结构图如图 2-6 所示。Faster RCNN 作为两阶段经典目标检测算法，需要获取目标所在的候选框，在进行检测时不仅耗时，而且准确率较低。

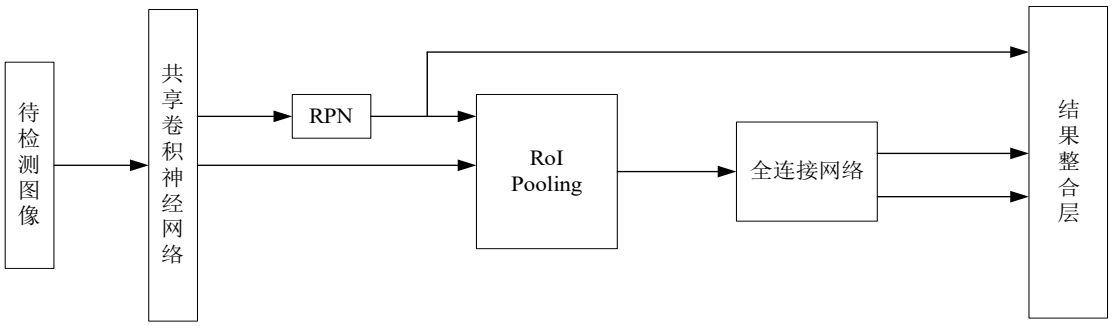


图 2-6 Faster RCNN 网络结构图

2. SSD 算法

SSD[38]算法由 Wei Liu 等人于 2016 年提出，它通过单次前向传递即可实现多类别目标的检测与定位。SSD 的网络结构通常基于 VGG16 或 ResNet 等经典网络，但会去除最后的全连接层，并添加额外的卷积层以获取更丰富的空间特征。网络结构图如图 2-7 所示。SSD 属于单阶段目标检测算法，其在检测速度上快于 Faster R-CNN，但精度低于 Faster RCNN，无法兼顾高效性与高精度。

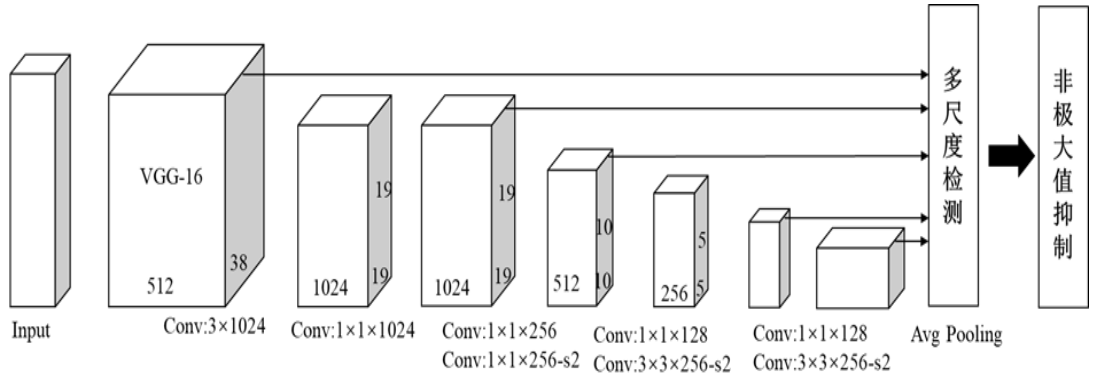


图 2-7 SSD 网络结构图

3. YOLO 系列

YOLO 系列代表的单阶段模型通过卷积网络直接提取特征，直接在特征图上生成边界框，同步执行分类与回归任务，从而实现对输入图像的实时预测输出[39]，其网络结构图如图 2-8 所示。下文将介绍 YOLO 系列算法，其中 YOLOv1 开创了这一系列产品，YOLOv3 进行了重大改进，YOLOv5 为 YOLOv3 的训练过程引入了许多训练技术，YOLOv8 通过改变 YOLOv5 网络的一些结构实现了更好的检测效果。本章将涵盖 YOLOv1、YOLOv3 和 YOLOv5 算法的基本原理，下一章将介绍 YOLOv8 的算理。

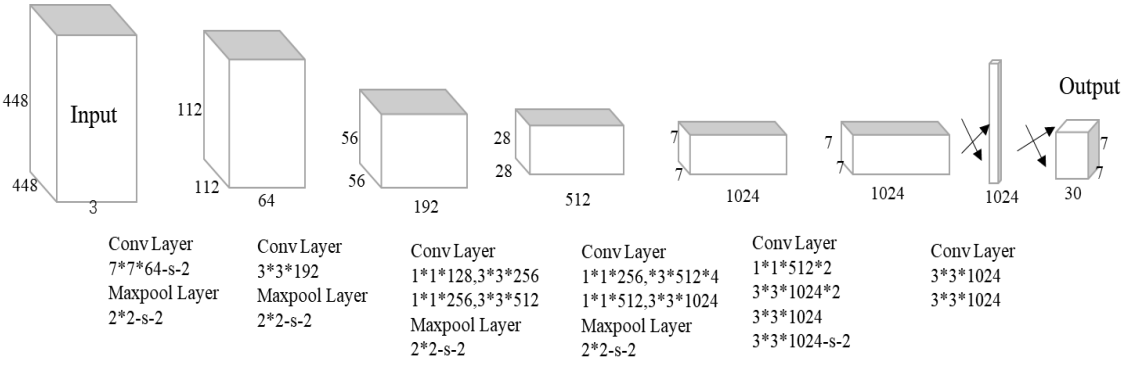


图 2-8 YOLO 网络结构图

(1) YOLOv1

在目标检测领域，YOLO(You Only Look Once)系列算法以其高效、快速的特点，成为近年来的研究热点。其中，YOLOv1 作为该系列的开山之作，其独特的原理和设计思想为后续版本的优化提供了坚实的基础。YOLOv1 采用回归思路处理检测任务，首先将输入图像划分为 S×S 的网格结构，随后，每个网格单元负责预测 B 个边界框以及每个边界框的类别概率。这些边界框是通过卷积神经网络(CNN)提取的特征图生成的，它们包含了位置信息与及置信度。同时，每个网格

还需要预测 C 个条件概率分布，用于表示该网格中各个类别的概率分布。

(2) Yolov3

相较于 YOLOv1，YOLOv3 借鉴了 ResNet 的设计理念，并采用了 Darknet-53 作为其核心网络架构，增加了网络的深度和复杂度。输出三种尺度的特征图(13x13，26x26，52x52)，每种尺度负责不同大小的物体检测，提高了模型对小物体的检测效果。YOLOv3 还采用了残差连接来提高梯度流动，防止梯度消失问题，并且支持多尺度训练，能够处理不同大小的物体。

(3) YOLOv5

与 YOLOv3 相比 YOLOv5 在算法架构上进行了多项改进。首先，它采用了 Backbone+Neck+Head 的结构，其中 Backbone 负责提取特征，Neck 负责特征融合，Head 负责预测。这种结构使得 YOLOv5 能够更好地利用多尺度信息，提高检测精度；其次，YOLOv5 引入了 Anchor Free 机制，不再依靠预设锚框的方法，而是直接预测边界框的位置和大小，从而减少了模型的复杂度和计算量。此外，YOLOv5 整合了多项数据增强策略，包括 Mosaic DAR 和 Auto Labeling 等技术，以此增强模型的泛化能力和鲁棒性。

2.2 双目视觉目标定位理论基础

双目视觉理论是计算机视觉领域的一项核心技术，它通过模仿人类双眼的工作原理，采用双摄像头从不同视角同步采集场景图像，进而重建出三维世界的信息[40]。

2.2.1 四大坐标系及其转换关系

双目立体视觉成像技术通过建立世界坐标系、相机坐标系、图像坐标系和像素坐标系之间的映射关系，利用透视投影模型和相机标定参数，实现空间点从三维世界坐标到二维像素坐标的投影变换，它们之间的关联如图 2-9 所示。

(1) 图像标系(x，y)：该坐标系以 O_i 为原点，其 x 轴和 y 轴分别对应图像的水平和垂直方向，图像像素在 x 轴和 y 轴上的长度分别为 dx 和 dy[41]。

(2) 像素坐标系(u，v)：以图像左上角为 O_o 原点，u 轴与 v 轴分别表示图像的宽度和高度方向，用于定位图像中的像素位置。

(3) 相机坐标系(X_C，Y_C，Z_C)：该坐标系以光心 O_c 为原点，X 轴和 Y 轴平行于成像平面，Z 轴与光轴重合并垂直于像平面。

(4) 世界坐标系(X_W，Y_W，Z_W)：该坐标系是一个全局的三维直角坐标系，用于描述物体在真实世界中的绝对位置。此坐标系的原点与轴向可根据具体应用场景的需求灵活设定。

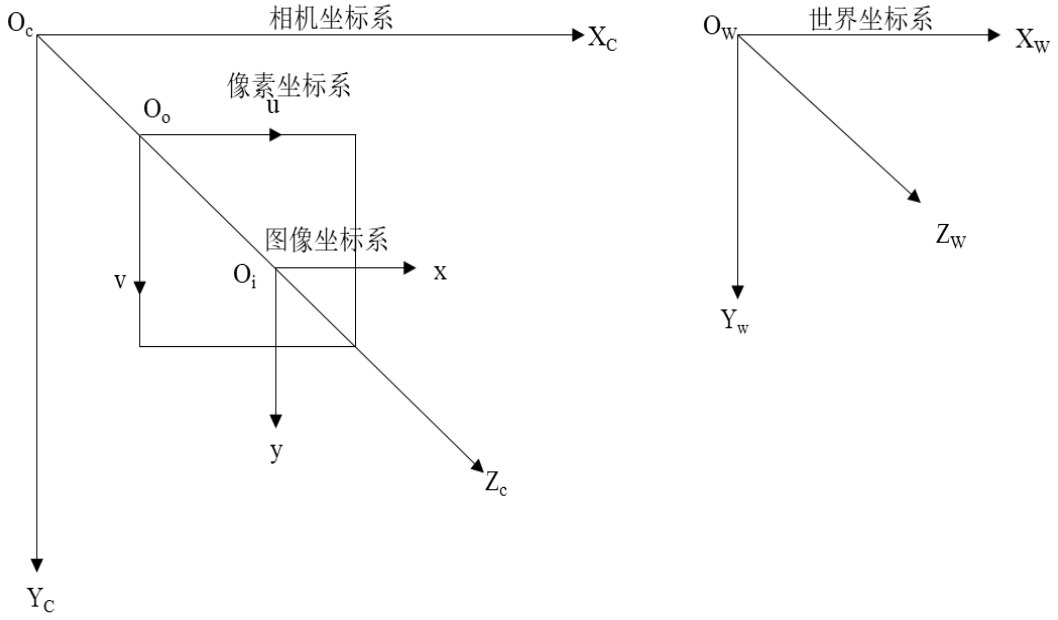


图 2-9 四大坐标关系示意图

(5) 四大坐标系之间的转换关系，描述了真实世界中某点在世界坐标系中的坐标与其在像素坐标系中的坐标之间的对应转换系[42], 具体过程如图 2-10 所示：

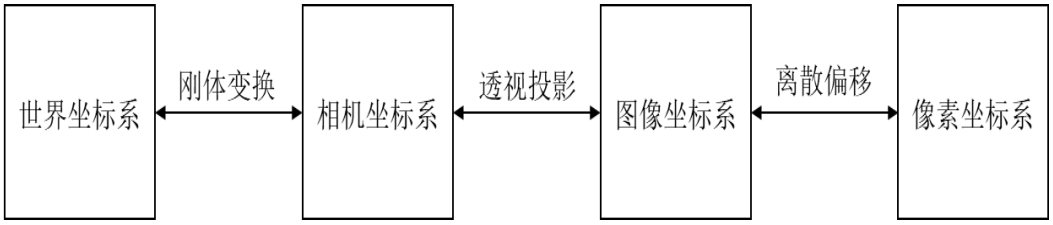


图 2-10 四大坐标系关系转换图

设图像坐标系的原点 O_i 位于(u_o，v_o)，则图像平面上的任意点 $M(x，y)$在图像坐标系中对应的 x 轴和 y 轴的长度分别为 dx 和 dy，其相互之间转换关系如公式 2-5 所示。

$$\begin{cases} u = \dfrac{x}{dx} + u_o \\ v = \dfrac{y}{dy} + v_o \end{cases} \tag{2-5}$$

通过矩阵转换如公式 2-6 所示：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2-6}$$

设点 $P(X_C，Y_C，Z_C)$为世界坐标系下的一点，即空间上的一点，由相似三角形可知，点 M 与点 P 的对应关系如公式 2-7 所示:

$$\begin{cases} x = f\dfrac{X_C}{Z_C} \\ y = f\dfrac{Y_C}{Z_C} \end{cases} \tag{2-7}$$

将公式进行其次线性转换形式可得公式 2-8：式中 f 为相机焦距。

$$Z_C \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \tag{2-8}$$

将公式带入可得，空间上的 P 点与像素坐标系之间的坐标转换关系，如公式 2-9 所示：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{Z_C} \begin{pmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} = \frac{1}{Z_C} \begin{pmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \tag{2-9}$$

世界坐标系与相机坐标系之间可以通过右手坐标系规则进行相互转换，这种转换仅需通过一个平移向量 $\boldsymbol{T}$ 和一个旋转矩阵 $\boldsymbol{R}$ 来实现[43]，其转换公式如 2-10 所示：

$$\begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} = \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix} \tag{2-10}$$

再将公式 2-9 代入公式 2-10 得到，世界坐标系到像素坐标系的转换，具体公式如 2-11 所示：

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{Z_C} \begin{pmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & T \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix} \tag{2-11}$$

化简可得公式 2-12 如下：

$$Z_C \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M_1 M_2 \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix} \tag{2-12}$$

式中：Z_C 为尺度因子，$\boldsymbol{M_1}$，$\boldsymbol{M_2}$ 分别为相机的内外参矩阵。

2.2.2 相机成像模型

相机成像模型是描述了三维空间中的点如何通过相机投影到二维图像平面上，其主要分为线性模型和非线性模型两类[44]。其中线性模型是处于理想环境下的模型，因为线性模型忽略相机镜头畸变因素对成像模型产生的影响，但在实际生活中，由于相机镜头包含了各种透镜，在现实成像模型中是需要考虑透镜畸变因素对成像的影响，这就是非线性模型。

1. 线性模型是一种简化的光学成像模型. 它假设光线在穿过镜头时没有发生色散，并且相机内部不存在畸变等因素。这种模型也被称为针孔相机模型[45]，如图 2-11 所示。

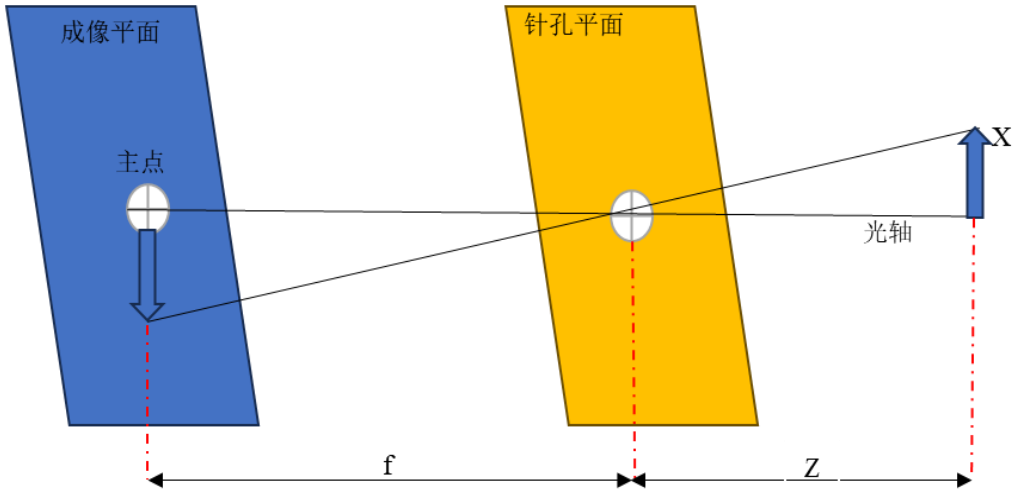


图 2-11 小孔成像模型

2. 非线性模型。与线性模型相比，非线性模型能够更精确地表示复杂系统的

行为，尤其是在处理光学系统中相机镜头透镜畸变等非线性问题时。这类模型不仅包含了线性部分来描述基本行为，还通过添加非线性成分来考虑并校正额外的影响因素，比如透镜畸变。这使得非线性模型在结构上更为复杂，但同时也提供了更高的精度和更好的适应性。相机的非线性模型主要包含了径向畸变模型和切向畸变模型。

(1) 径向畸变主要由镜头的光学设计所引起的，这些镜头设计使得光线在穿过透镜时发生不同程度的折射，进而引发畸变现象。径向畸变修正公式如 2-13 所示

$$\begin{cases} x_r = x\left(1+k_1r^2+k_2r^4+k_3r^6\right) \\ y_r = y\left(1+k_1r^2+k_2r^4+k_3r^6\right) \end{cases} \tag{2-13}$$

式中：$(x_r，y_r)$表示修正后的径向畸变坐标，$(x，y)$表示原始坐标，d 表示点到光心的距离，k_1、k_2、k_3 表示径向畸变参数。

(2) 切向畸变主要由摄像机生产制造过程中的误差所引起的，如透镜与成像平面不平行，或镜头组装时存在偏差。这些因素使得光线在进入传感器前发生偏转，从而产生切向畸变。切向畸变修正公式如 2-14 所示。

$$\begin{cases} x_t = x+2p_1xy+p_2\left(d^2+2x^2\right) \\ y_t = y+p_1\left(d^2+2y^2\right)+2p_2xy \end{cases} \tag{2-14}$$

式中：$(x_t，y_t)$表示修正切向畸变后的坐标，$(x，y)$表示原始坐标，d 表示点到光的中心点之间的距离，p_1、p_2 为切向畸变参数。

2.2.3 双目相机测距原理

双目测距，也称为立体视觉或深度感知，是一种通过两个视角捕捉同一场景的方式来测量物体距离的方法。这项技术模仿了人类的双眼视觉，通过计算两个不同视角之间的视差来确定物体的距离，从而获得物体的空间三维坐标信息[46]。工作原理如下：在目标测距中，设 P 为空间的一个点，点 P 的位置由左右相机的光心 O_l 和 O_r 确定，两者之间的基线距离为 T，相机的焦距为 f，点 P_l 和 P_r 在图像坐标系中的最左和最右距离分别为 X_l 和 X_r，而点 P 到相机的垂直距离为 Z。其原理图如图 2-12 所示。

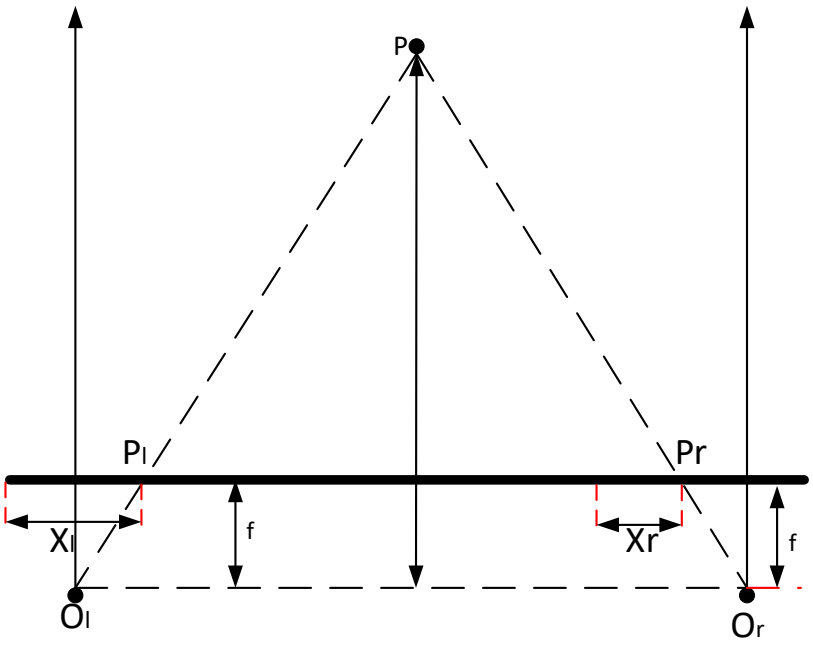


图 2-12 双目视觉测距原理

由图可知，$\Delta PP_lP_r \sim \Delta PO_lO_r$，根据相似三角形可得公式 2-15：

$$\frac{T-(X_l-X_r)}{T}=\frac{Z-f}{Z} \tag{2-15}$$

从而可以推出距离 Z 的公式如 2-16 所示：

$$Z=\frac{fT}{X_l-X_r}=\frac{fT}{d} \tag{2-16}$$

式中：$d=X_l-X_r$，d 代表左右相机的视差，基线 T 和焦距 f 可由相机标定获得。因此，只需求得相机的视差 d 数值，即可以计算出目标点 Z 的距离信息。

2.3 本章小结

本章主要围绕目标检测与双目视觉的理论基础展开。在目标检测识别部分，首先系统介绍了卷积神经网络(CNN)的核心组成部分及其工作原理；其次还重点讲解了 Faster RCNN、SSD 和 YOLO 系列三大经典的目标检测算法原理，并通过相互对比，最终选择 YOLO 系列算法作为本文的检测识别算法。双目视觉定位方面主要介绍了相关坐标系及其转换关系、相机成像模型以及双目测距原理，并推导出双目测距的公式。

第 3 章 基于改进 YOLOv8 目标识别研究

近几年来随着机器视觉快速发展，以及硬件性能方面不断的提升。研究人员提出了许多高效且性能出色的目标检测算法，这些算法凭借其优异的表现逐渐在码垛行业中受到重视，并被广泛运用于实际场景。零件识别作为码垛过程中的核心环节，其算法的研究对推动码垛行业智能化发展具有重要作用。YOLOv8 作为当前目标检测领域比较高效的算法，在 YOLO 系列中展现了显著的性能优势。因此，本章将以改进的 YOLOv8 目标检测算法为基础对零件识别展开研究。

3.1 YOLOv8 算法原理

YOLOv8 采用 CSPDarknet53 作为特征提取网络。该网络结合了卷积神经网络和空间金字塔池化的优势，通过引入深度可分离卷积和残差连接，显著提高了特征提取能力。CSP 模块有效减少了梯度消失问题，增强了特征图的表示能力。C2f 模块是 CSPDarknet53 的重要组件，通过结合倒残差结构和卷积算子，提升了特征提取效果。输入特征图被分成两部分处理后相加，形成新的特征图，增加了网络深度，减少了参数量，提高了计算效率。

3.1.1 YOLOv8 网络结构

YOLOv8 是 YOLO 系列推出的目标检测模型之一。YOLOv8 提供了 5 个版本，分别为 YOLOv8n、YOLOv8m、YOLOv8s、YOLOv8x 和 YOLOv8l。在这几个版本中，虽然 YOLOv8n 模型识别精度稍低，但是其模型架构最轻量化，检测速度最快。基于这一特性，本章研究经过综合考虑，选择 YOLOv8n 作为零件识别的基础模型。图 3-1 为 YOLOv8 网络结构图。YOLOv8 的网络结构主要由四个核心部分组成：输入模块、Backbone 主干网络、Neck 颈部网络以及 Head 头部网络。Backbone 主干网络由 C2f、Conv 和 SPPF 三个模块组成。Neck 颈部网络采用了路径聚合网络(path aggregation network, PANet)，与特征金字塔网络(feature pyramid network, FPN)相比，PANet 引入了一种自下向上的路径机制，使得底层信息可以更为顺畅地传递到网络结构顶部[47]。Head 头部采用解藕头结构将分类和定位预测进行了分离，这样做使模型收敛速度更快，检测效果更好。

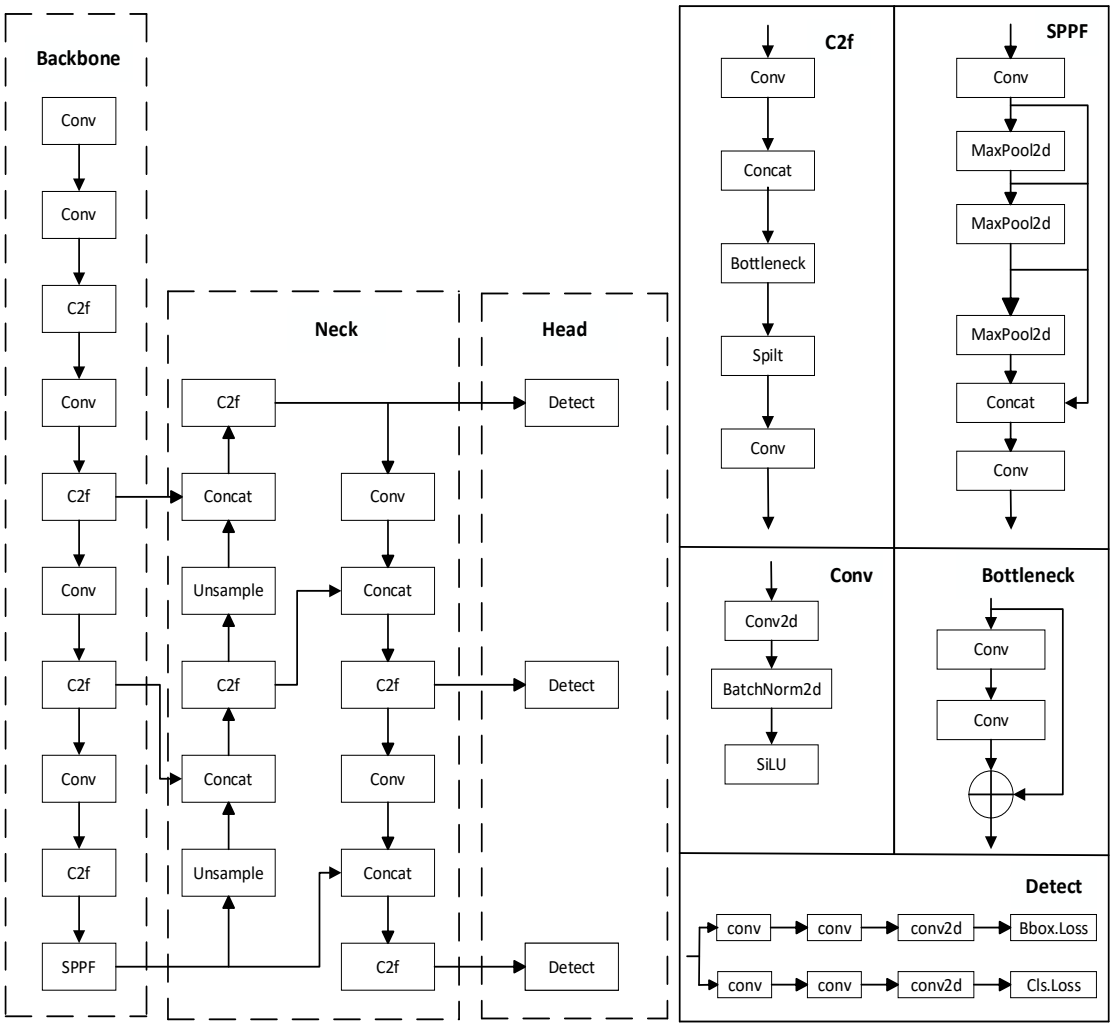


图 3-1 YOLOv8 网络结构图

1. 输入层(Input)

YOLOv8 的输入层是接收和预处理图像数据的关键部分，其核心功能是解析原始视觉信息。该模块处理的数字图像采用标准 RGB 色彩模式存储，每个像素由红、绿、蓝三个色彩通道构成。在特征提取阶段启动前，系统会通过自适应缩放算法将输入样本统一转换为预设分辨率(如 640×640 像素或 416×416 等标准化规格)，以确保所有输入数据具有一致的形状。这一步骤对于后续的卷积操作至关重要，因为它保证了特征图的空间维度统一。

2. 主干网络(Backbone)

YOLOv8 的主干网络在目标检测中扮演了核心角色，它承担着从输入图像中提取关键特征的任务。Backbone 主干网络的结构包含三个关键模块：C2f 模块、Conv 模块以及 SPPF 模块，下面将详细介绍这三个模块。

(1) C2f 模块

在 YOLOv8 中，C2f 模块是一个残差块(Residual Block)，其核心思想是通过引入残差连接(Residual Connection)，使得网络能够学习输入和输出之间的直接映射关系，从而优化网络的训练过程。C2f 模块的结构如图 3-2 所示：从图中可以看出，C2f 模块由两个卷积层(Conv)组成,中间通过残差连接(Residual Connection)进行衔接。在残差连接中，输入信号直接传递到输出端，同时通过卷积操作学习输入与输出之间的映射关系。这种设计能够有效缓解深度神经网络中的梯度消失问题，并增强模型的表示能力。

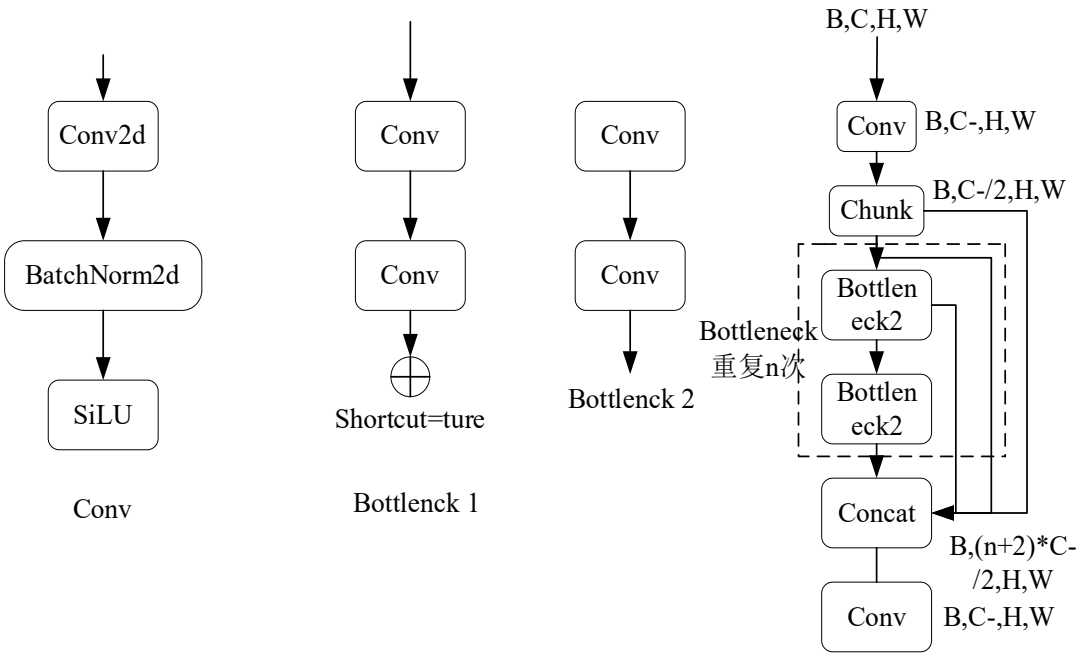


图 3-2 C2f 模块结构图

(2) Conv 卷积模块

Conv 卷积模块是架构中最基本的模块，主要包括 Conv2d 层、BatchNorm2d 层和 SiLU 激活函数。Conv 模块的结构图如图 3-3 所示。在运行过程中，首先 Conv2d 模块通过卷积操作以提取目标物体的关键特征信息；其次通过 BatchNorm2d 层来提高模型训练的稳定性，同时加快训练过程中的收敛速度；最后引入 SiLU 激活函数层，用以提高模型的检测识别能力。Conv 模快不仅有效的提取了目标的特征，同时也加快了模型训练过程中的收敛速度。

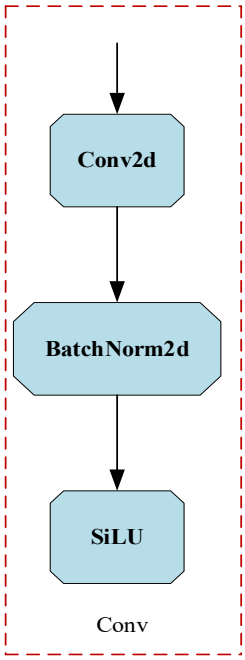


图 3-3 Conv 模型结构图

(3) SPPF 模块

YOLOv8 中的 SPPF 模块，即空间金字塔池化融合模块，是一种用于目标检测的特征提取和融合模块。结构图如图 3-4 所示，其工作原理是通过多尺度池化操作提取特征信息。具体而言，它会对输入特征图分别进行 1x1、5x5 和 9x9 等不同尺度的池化操作，然后将这些池化后的特征图进行拼接，形成一个包含多尺度信息的特征向量。

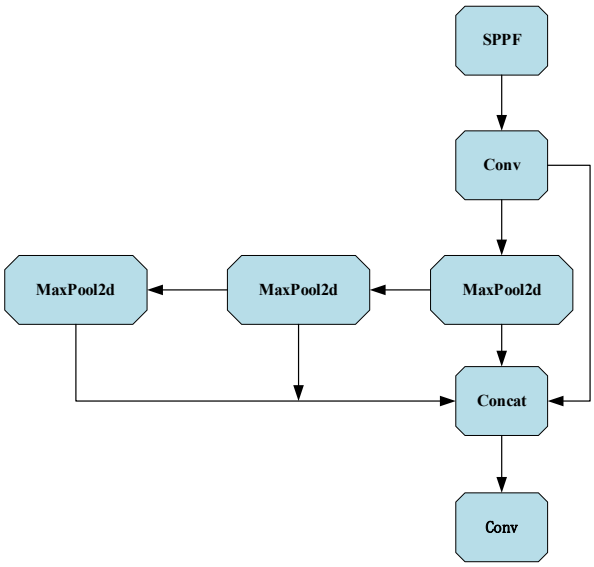


图 3-4 SPPF 模块示意图

3. 颈部网络(Neck)

YOLOv8 的颈部网络采用了优化后的 PAN-FPN(路径聚合网络-特征金字塔

网络)结构[48]。这种结构通过聚合，以增强特征表示能力骨干网络的多尺度特征图，为后续的目标检测提供更丰富的信息。与 YOLOv5 相比 YOLOv8 还在 PAN-FPN 的基础上进行了以下改进：首先删除了卷积结构；其次替换了 C3 模块：YOLOv8 将 PAN-FPN 中的 C3 模块替换为了 C2f 模块。C2f 模块具有更少的参数量和更优秀的特征提取能力，有助于实现网络的轻量化。

4. 头部网络(Head)

YOLOv8 的头部网络继承了 YOLO 系列模型的传统，采用了 Decoupled Head(解耦头)设计。这种设计将分类任务和回归任务分别交给不同的网络层来处理，从而提高了各自的精度。结构图如图 3-5 所示。具体来说，YOLOv8 的头部网络包含三个输出分支，每个分支又进一步细分为分类和回归两部分。分类头主要负责检测目标的类别；回归头主要负责预测目标的边界框坐标。YOLOv8 在回归头中采用了 DFL(Distributed Focal Loss)策略，这是一种优化的损失函数，用于解决边界模糊的问题。DFL 将坐标预测从一个确定性的单值转变为一个概率分布，从而提高了边界框预测的准确性。综上所述，YOLOv8 的头部网络通过采用解耦设计和 DFL 策略等先进技术，实现了高效的目标检测和分类任务。这些特点使得 YOLOv8 在准确性、鲁棒性和效率等方面均表现出色。

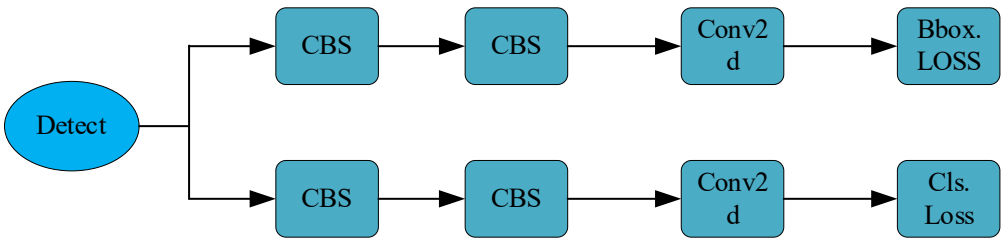


图 3-5 解耦头结构图

3.1.2 损失函数

YOLOv8 中的损失函数主要由分类损失函数(BCE Loss)和回归损失函数(Distribution Focal Loss 与 CIoU Loss)三个部分组成具体公式 3-1 如下：

$$Loss = \lambda_1 BCE + \lambda_2 DFL + \lambda_3 CIOU \tag{3-1}$$

1. 分类损失函数(BCE Loss)

YOLOv8 中的分类损失函数采用二元交叉(BCE)公式如 3-2 所示：它是一种常用于二分类问题的损失函数，其核心是通过计算模型预测的概率分布与真实标签之间的差异来衡量模型的准确性。

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i.\log\left(p_i\right)+\left(1-y_i\right).\log\left(1-p_i\right)\right] \tag{3-2}$$

式中：N 为样本总数；y_i 和 p_i 分别表示第 i 个样本的真实标签（0 或 1）和预测概率。

2. 回归损失函数(CIoU Loss)

YOLOv8 的 CIoU 损失函数是目标检测任务中用于量化模型预测边界框与真实边界框之间差异程度的评估函数。其公式 3-5 如下：

$$v = \frac{4}{\pi^2}\left(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h}\right)^2 \tag{3-3}$$

$$\alpha = \frac{v}{\left(1-iou\right)+v} \tag{3-4}$$

$$CIOU\ Loss = 1-\left(IOU - \frac{l^2}{c^2} - \alpha\times v\right) \tag{3-5}$$

式中：w^{gt} 与 h^{gt} 分别为真实框的宽高，w 与 h 表示两者长宽比差异，α 为权重系数，l 为两个矩形框中心点之间的距离，c 表示最小闭合区域的对角线距离。IoU 为交并比。

3. 回归损失函数(Distribution Focal Loss)

在进行图像检测时，通常会存在目标重叠和遮挡等复杂的情况，若出现这些复杂的情况，此时目标检测框与标注框无法反映出真实的图像情况。面对这些复制的情况，YOLOv8 引入 Distribution Focal Loss 损失函数,简称 DFL[49]。其公式 3-6 如下：

$$\hat{y} = \int_{-\infty}^{+\infty} P(x)xdx = \int_{y_0}^{y_n} P(x)xdx \tag{3-6}$$

通过离散分布概率可得公式 3-8 如下：

$$\hat{y} = \sum_{i=0}^{n} P(y_i)y_i \tag{3-7}$$

$$DFL\left(S_i, S_{i+1}\right) = 1-\left(\left(y_{i+1}-y\right)\log\left(S_i\right)+\left(y-y_i\right)\log\left(S_{i+1}\right)\right) \tag{3-8}$$

式中：$\hat{y}$网络预测输出值，S 为概率输出值，当 y_{i+1} 与 y_i 接近时，其对应的 $P(y_i)$值越大，此时 DFL 值越小，使得分布向标注中心集中靠拢。

3.2 YOLOv8 改进策略

面对散放、堆叠和遮挡的复杂码垛场景下，为了提高识别的准确率和检测的效率。本文以 YOLOv8 模型为基础做了以下三个方面改进：

(1) 将 YOLOv8 模型的主干网络替换为轻量型 MobileNetv3。

(2) 在模型的主干网络中第四层和第十层分别引入 CBAM 注意力机制。

(3) 将 YOLv8 模型中的损失函 CIoU 数替换成 SIoU。

将改进后的模型命名为 YOLOv8-MSC，其主要由 4 个部分构成，输入层，主干网络、颈部网络和头部网络组成，其网络结构图如图 3-6 所示。

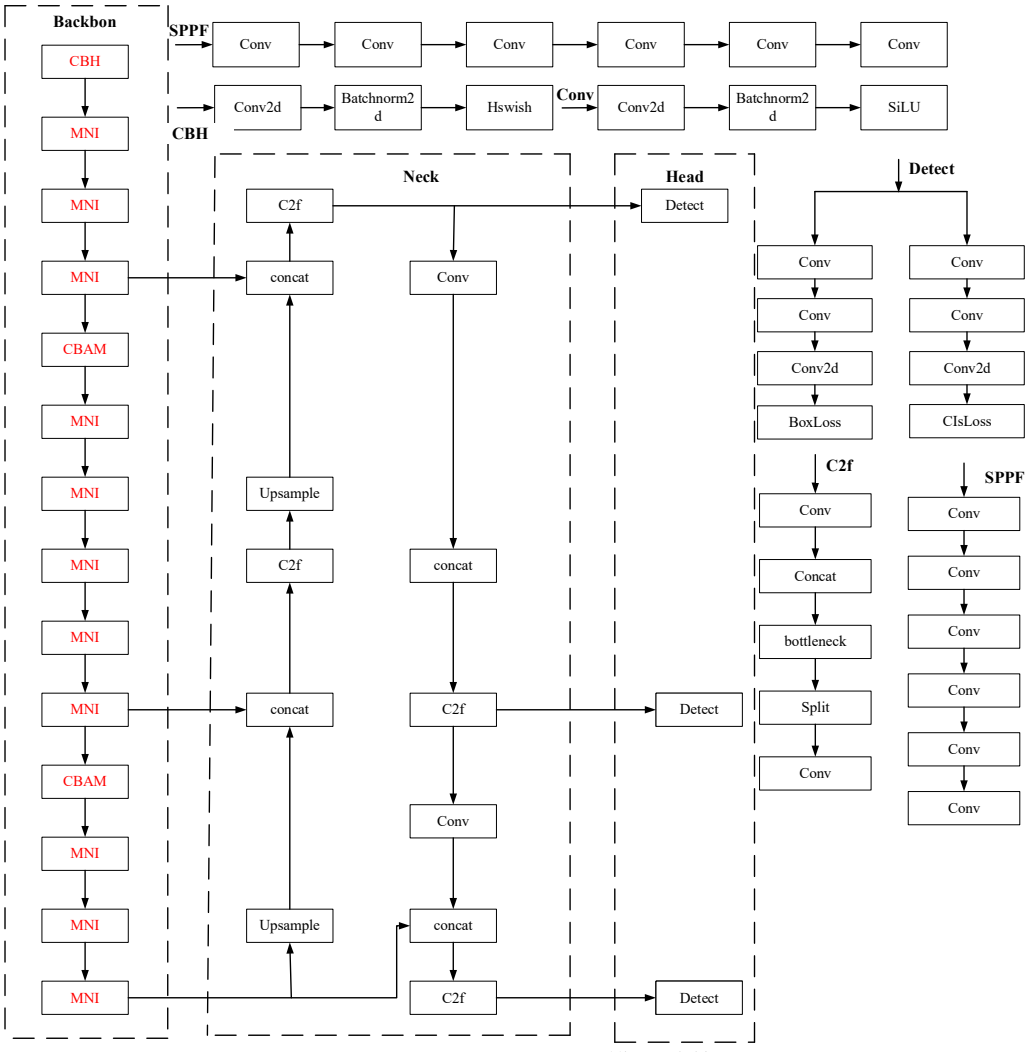


图 3-6 YOLOv8-MSC 网络模型结构图

YOLOv8 模型采用 SPDarknet-53 作为其主干网络架构，其网络模型在识别检测精度方面表现出优异的性能，但同时其网络模型结构参数量较大，且计算量较为复杂，不易部署在高速移动且小型设备上。为此针对 YOLOv8 网络模型过于繁琐复杂的问题，从而进行了网络模型轻量化结构设计。从轻量化网络结构来

看，目前常见的轻量化网络模型主要有SqueeZeNet[50]、ShuffleNet[51]、MobileNet[52]网络等。SqueezeNet 是一种卷积神经网络(CNN)模型，由 UC Berkeley 的研究人员在 2016 年提出。该模型的主要特点是具有较少的参数和计算复杂性，同时在图像分类任务上仍能取得良好的性能。但其使用了 Fire 模块对模型的参数量进行压缩，导致其特别受限于资源环境；ShuffleNet 是由 Face++在 2017 年提出，其核心创新包括逐点群卷积和通道混洗操作，这些技术有助于减少计算复杂度并增强特征通道间的信息流动，但 ShuffleNet 训练模型时比较耗时，实际运行起来运行速度并没有达到理想的状态。相比于前两个轻量化网络模型结构而言，MobileNetv1 采用了可分离卷积有效的减少了计算量，MobileNetv2 引入 ResNet 的残差结构，设计了倒残差模块，MobileNetv3 在此基础上加入 SE 注意力机制，显著提升了模型的表征能力。因此选用 MobileNetv3 作为本文的主干网络。

3.2.1 MobileNetv3 轻量化网络结构

MobileNet 系列自提出以来经历了多次迭代，每一代都在前一代基础上进行了优化和改进。MobileNetv1 首次引入了深度可分离卷积，将标准卷积分解为深度卷积和逐点卷积，大幅减少了计算量。MobileNetv2[53]进一步优化了这种结构，引入了线性瓶颈和反向残差结构，提升了特征提取能力。MobileNetv3[54]则结合神经架构搜索技术，优化了网络的整体结构,并在激活函数和注意力机制上进行了创新。使模型的整体上的性能得到了大幅度的提升。MobileNetv3 网络模型结构参数如表 3-1 所示。

表 3-1 MobileNetv3 网络模型结构参数

输入	操作	扩展输出通道数	输出通道	SE	步长
224×224×3	Conv2d	-	16	√	2
112×112×16	bneck,3×3	16	16	×	2
112×112×16	bneck,3×3	72	24	×	2
56×56×24	bneck,3×3	88	24	×	1
56×56×24	bneck,5×5	96	40	√	2
28×28×40	bneck,5×5	240	40	√	1
28×28×40	bneck,5×5	240	40	√	1
28×28×40	bneck,3×3	240	80	×	1
14×14×80	bneck,3×3	120	80	×	1
14×14×80	bneck,3×3	144	80	×	2
14×14×80	bneck,3×3	288	80	√	1
14×14×80	bneck,3×3	576	112	√	1

续表 3-1 MobileNetv3 网络模型结构参数

输入	操作	扩展层输出通道数	输出通道数	SE	步长
14×14×112	bneck,3×3	576	112	√	1
14×14×112	bneck,5×5	672	160	√	1
7×7×160	bneck,5×5	960	160	√	2
7×7×160	bneck,5×5	960	160	√	1
7×7×960	Conv2d,1×1	-	960	√	1
7×7×960	Pool,7×7	-	-	×	1
1×1×960	Conv2d,1×1,NBN	-	1280	×	1
1×1×1280	Conv2d,1×1,NBN	-	k	×	1

该表格采用分层结构展示 MobileNetv3 的网络配置：第一列详细列出了各层输入特征向量的空间维度；第二列描述了每层所采用的卷积运算类型；第三列和第四列分别记录了 Bottleneck 结构中扩展层的输出通道数和特征层的最终输出通道数；SE 列用于标识是否集成了通道注意力机制；最后一列表示卷积操作的步长。下面是对 MobileNetv3 网络结构的详细讲解。

1. 深度可分离卷积

常规卷积通过一个卷积核同时处理输入特征图的所有通道，生成一个新的输出通道。深度可分离卷积分为两步，首先是深度卷积(Depthwise Convolution)采用分解式卷积策略，将标准卷积分解为两个独立的计算阶段：第一阶段为深度卷积(Depthwise Convolution)，采用通道隔离的计算方式，即每个输入通道分别与对应的二维卷积核进行空间卷积运算，实现空间特征的独立提取；第二阶段为逐点卷积(Pointwise Convolution)，通过 1×1 卷积核实现跨通道的特征融合，将深度卷积输出的多通道特征图投影到新的特征空间，构建具有特定通道维度的输出特征表示。其原理如下：

假设输入特征图的尺寸为 $H\times W\times N$,卷积核大小 $K\times K$，输入通道数为 N，输出通道数为 M。常规卷积计算量公式 3-9 如下：

$$H\times W\times N\times K\times K\times M \tag{3-9}$$

深度可卷积计算量公式 3-10 如下：

$$K\times K\times N\times H\times W+M\times H\times W\times N \tag{3-10}$$

将两种计算量作对比可得公示 3-11：

$$\frac{K\times K\times N\times H\times W+M\times H\times W\times N}{H\times W\times N\times K\times K\times M}=\frac{1}{M}+\frac{1}{k^2} \tag{3-11}$$

将两种卷积作对比可知，深度可分离卷积操作的计算量是常规卷积操作计算量的$\frac{1}{M}+\frac{1}{K^2}$倍。与常规卷积相比，深度可分离卷积显著降低了计算复杂度。

2. 倒残差结构

倒残差结构是深度学习中卷积神经网络的一种重要设计模块，旨在提高模型的效率和性能。该结构首先对输入特征图进行轻量级扩张卷积以增加通道数，随后通过深度可分离卷积提取特征，并引入线性激活函数以增强非线性表达能力，最后通过逐点卷积将通道数恢复至输出特征图，其结构图如图 3-7 所示。

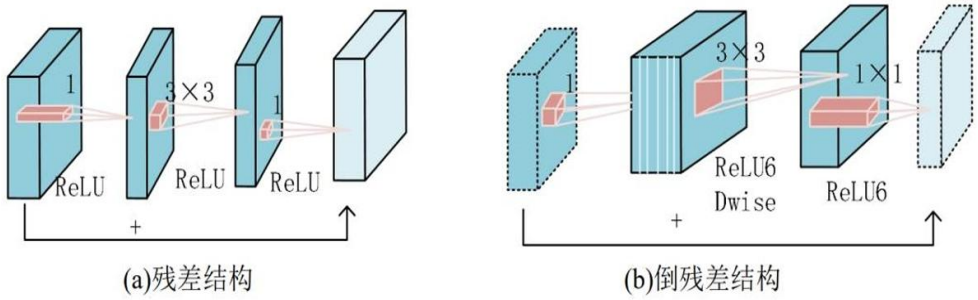


图 3-7 残差和倒残差的结构

3. SE 模块

SE 模块(Squeeze-and-Excitation)[55]是一种注意力机制，旨在增强网络的特征表达能力。它通过对特征通道的重要性进行动态重加权，从而提升模型的性能和效率。其结构图如图 3-8 所示。

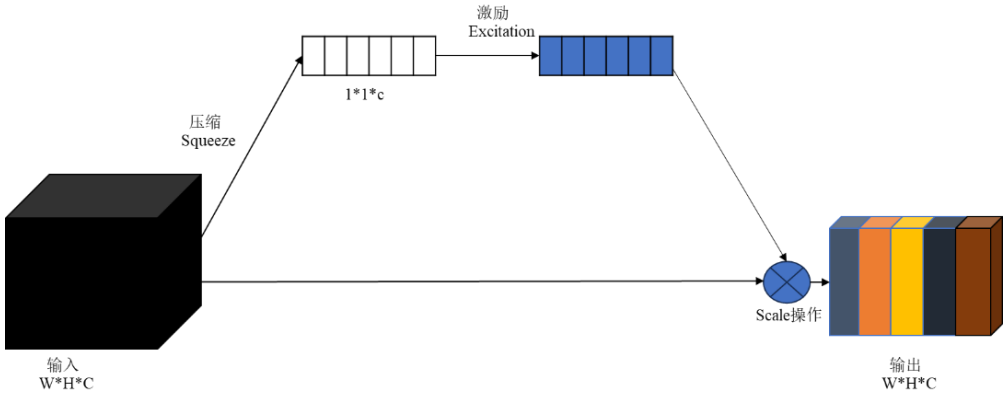


图 3-8 SE 模块结构图

在 MobileNetv3 中，SE 模块被集成到每个深度可分离卷积层之后，以实现对不同特征通道的自适应权重调整。SE 模块的核心机制是通过全局平均池化将特征图的空间维度压缩 1×1，随后利用两个全连接层(一个降维，一个升维)学习通道间的依赖关系，生成一组权重系数。这些系数与原始特征图相乘，从而突出重要特征并弱化不相关特征。这种设计使得 MobileNetv3 能够在保持轻量化的同

时，提高模型的识别能力和泛化性能。MobileNetv3 模块结构图如图 3-9 所示。

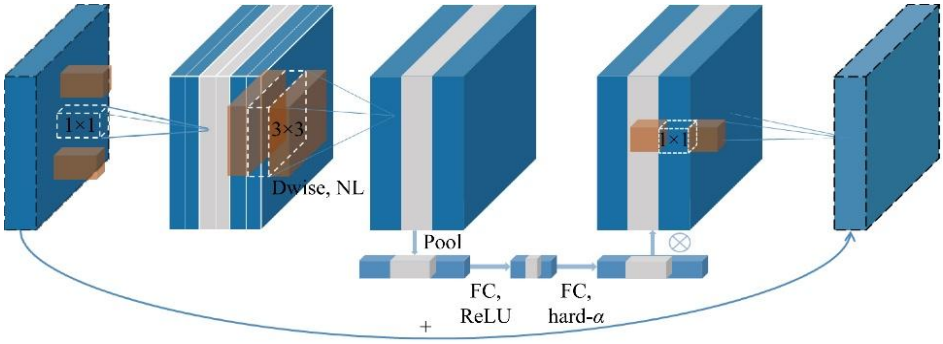


图 3-9 MoileNetv3 模块结构

4. h-swish 激活函数

MobileNetv3 引入了 h-swish 激活函数。h-swish 激活函数的工作原理是基于 Swish 激活函数的优化改进，它通过分段线性函数替代 Sigmoid 函数，以简化计算并保持性能优势。其公式如 3-13 所示：函数图如图 3-10 所示：

$$Swish(x) = x \bullet sigmoid(x) = \frac{x}{1+e^{-x}} \tag{3-12}$$

$$h-swish(x) = x \bullet \frac{ReLU(x+3)}{6} \tag{3-13}$$

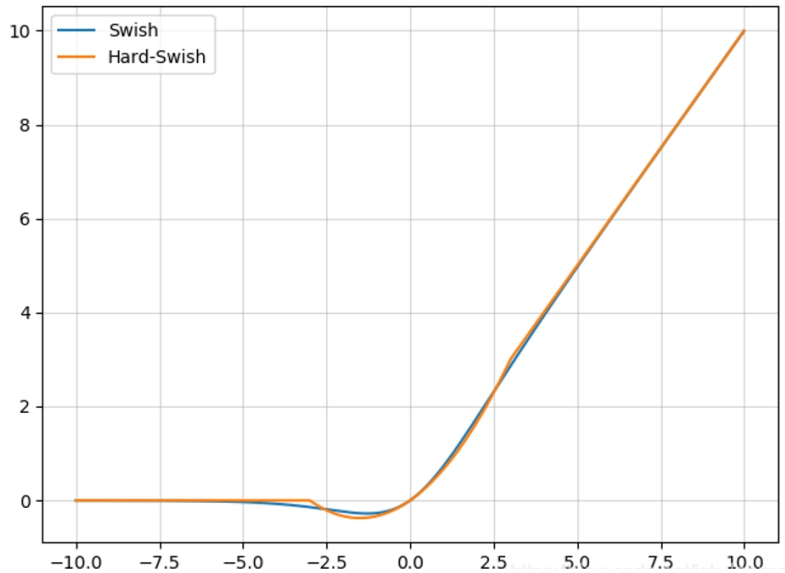


图 3-10 h-swish 激活函数

3.2.2 引入注意力机制

注意力机制是一种模仿人类视觉选择性关注能力的模型，它允许神经网络在处理输入数据时动态调整注意力权重，突出重要信息并抑制不重要的信息。其工作原理主要通过计算输入数据的注意力权重，将关注点放在重要的特征上，然后将这些特征进行加权求和，最后生成一个注意力向量，作为模型的输出。以下是

常见的几个注意力机制。

1. SA 注意力机制

SA[56]是一种高效的注意力模块能够同时结合空间和通道注意力机制，其工作原理是，SA 先将通道维度分组为多个子特征，接着在每个子特征上使用 Shuffle Unit 操作，将空间和通道注意力机制集成到一个模块中，最后通过 Shuffle 单元融合两种注意力机制的结果，生成最终特征表示，这种并行处理提高了计算效率。其结构图如图 3-11 所示。

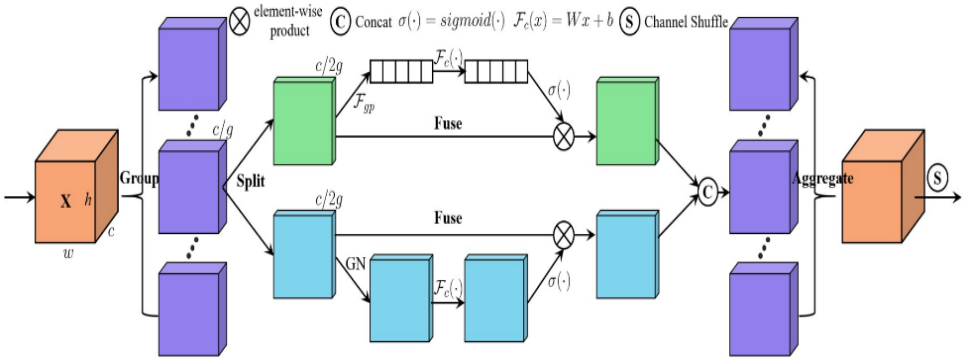


图 3-11 SA 注意力机制模块结构图

2. CBAM 注意力机制

CBAM[57]注意力机制是一种结合通道注意力机制与空间注意力机制而成的模块。它帮助主干网络强化关键特征并抑制非关键特征，它有效提升了模型在遮挡和重叠目标场景下的识别精度。其结构图如图 3-12 所示。

空间注意力机制(Spatial Attention Module)[58]：空间注意力机制模块通过以下步骤实现特征图的空间重加权：首先，对输入特征图，分别执行全局平均池化和最大池化操作；随后，将两种池化结果沿通道维度相加，得到两个一维向量；然后，通过点积运算计算注意力权重矩阵；最后，将该矩阵与原始输入特征图相乘，得到经过空间注意力机制调整的特征图。计算公式 3-14 如下：

$$M_s(F)=\sigma\left(f^{7\times7}\left[F^s_{avg};F^s_{max}\right]\right) \tag{3-14}$$

式中：$M_s(F)$表示空间注意力值、$f^{7\times7}$ 表示 7×7 的卷积运算、F^s_{avg} 表示空间上平均池化映射特征、F^s_{max} 表示空间上最大池化映射特征。

通道注意力模块(Channel Attention Module)[59]：通道注意力机制模块通过以

下步骤实现特征图的通道重加权：首先，对输入特征图，分别执行全局平均池化和最大池化操作；接着，通过多层感知机(MLP)学习每个通道的权重；最后，将学习到的通道权重与输入特征图相乘，生成经过通道注意力机制调整的特征图。这种机制能够有效地增强重要通道的特征表示，提升模型的特征提取能力。计算公式 3-15 如下。

$$M_c(F)=\sigma\left(W_1W_0\left(F_{avg}^c\right)\right)+W_1\left(W_0\left(F_{max}^c\right)\right) \tag{3-15}$$

式中：$M_c(F)$表示通道注意力值、σ表示 sigmoid 激活函数、F_{avg}^c表示平均池化后在通道上映射特征、F_{max}^c表示最大池化后在通道上映射特征、W_0表示第一个全连接层权重矩阵、W_1表示第二个全连接层权重矩阵。

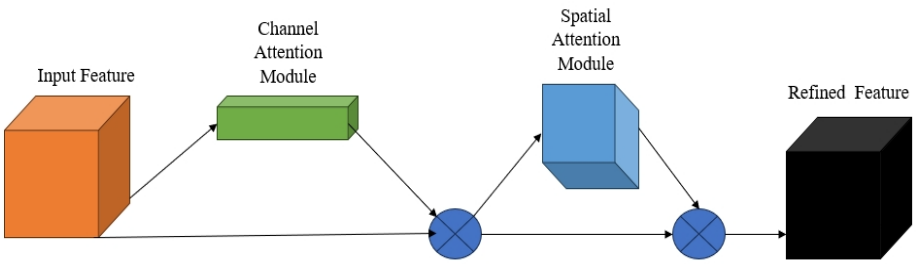


图 3-12 CBAM 注意力机制模块结构图

3. GAM 注意力机制

GAM[60]注意力机制是一种全局注意力机制(Global Attention Mechanism)，旨在减少信息损失并增强跨维度交互。在通道注意力机制模块中，GAM 采用三维排列保留信息，通过双层级 MLP 增强跨维度通道，同时在空间注意力单元中采用双重卷积层实现空间特征整合，并取消了最大池化处理。这种机制有效提升了深度神经网络的性能。结构图如图 3-13 所示。

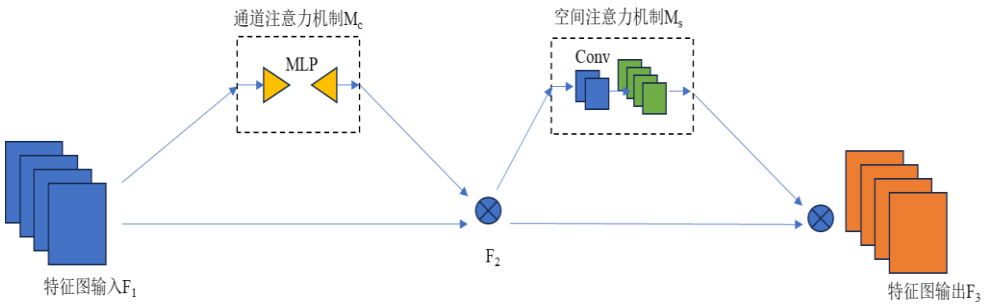


图 3-13 GAM 注意力机制模块结构图

3.2.3 优化损失函数

传统的目标检测损失函数有 GIoU[61]、CIoU、EIou[62]和 SIoU[63]，YOLOv8 采用的损失函数是 CIoU。传统的损失函数通常采用预测框与真实框之间的距离、重叠面积和纵横比为边界回框的回归指标。然而，传统的损失函数在评估目标检测模型的性能时，通常未能充分考虑预测边界框与真实边界框在空间位置上的具体偏差问题。这一局限性导致模型在训练过程中出现预测框不稳定的现象，表现为预测框在真实框周围反复波动，不仅降低了模型的收敛速度，还影响了最终检测精度。

针对以上问题，本文将 YOLOv8 模型的原始损失函数替换为 SIoU 损失函数。SIoU 不仅关注于边界框的重叠程度即(IoU)，还引入了角度损失、距离损失和形状损失等组件，以更全面地评估预测框与真实框的匹配程度。其优点主要体现在以下三个方面：第一提高精度：通过综合考虑多个因素，SIoU 损失函数能够更准确地评估预测框与真实框的匹配程度，从而提高目标检测的精度；第二加快收敛速度：引入角度损失等组件有效促进模型快速收敛至最优解，从而提高了训练效率；第三增强鲁棒性：SIoU 损失函数对于不同大小、形状和位置的目标具有更好的适应性，能够在复杂场景下保持较高的检测性能。SIoU 主要由四个模块组成，具体为角度损失、距离损失、形状损失和交并比损失。

(1) 角度损失：通过引入预测框与真实框之间的向量角度约束，有效解决了目标检测中预测框定位不稳定的问题。公式 3-16 与示意图 3-14 如下：

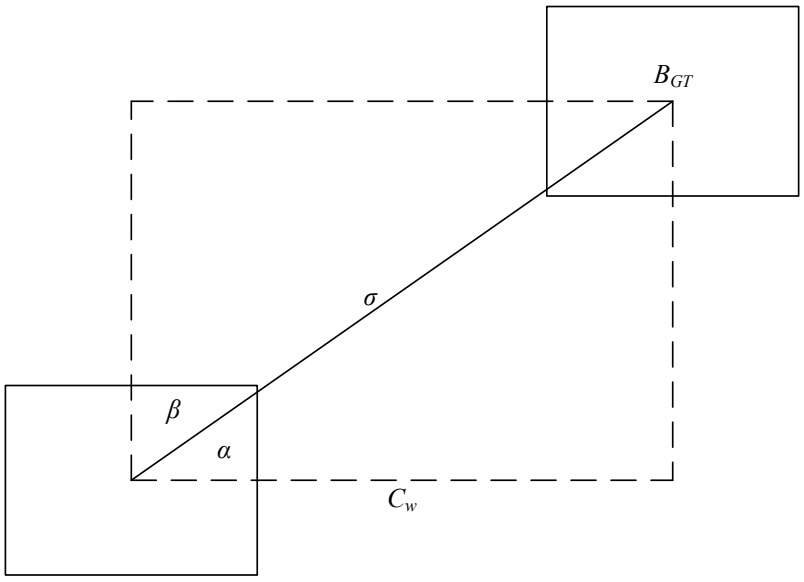


图 3-14 角度损失函数图

$$\begin{aligned}
&\Lambda = 1 - 2 * \sin^2\left(\arcsin(x) - \frac{\pi}{4}\right) \\
&x = \frac{c_h}{\sigma} = \sin(\alpha) \\
&\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \\
&c_h = max\left(b_{c_y}^{gt}, b_{c_y}\right) - min\left(b_{c_y}^{gt}, b_{c_y}\right)
\end{aligned} \tag{3-16}$$

式中：Λ为角度损失，$\sin(\alpha)$为真实框与预测框之间的距离和高度比。

$(b_{c_x}^{gt}, b_{c_y}^{gt})$为真实框中心点坐标，$\left(b_{cg}, b_{cy}\right)$为预测框中心点坐标。

(2) 距离损失：基于预测框和真实框中心点之间的距离进行惩罚，鼓励模型将预测框移动到更接近真实框的位置。公式 3-17 与示意图 3-15 如下：

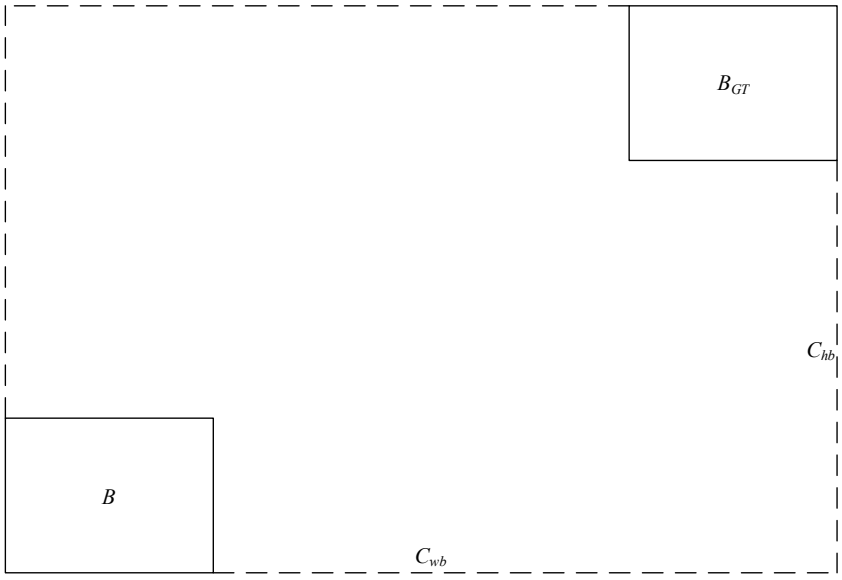


图 3-15 距离损失函数图

$$\begin{aligned}
\Delta &= \sum_{t=x,y} (1 - e^{-\gamma\rho_t}) \\
&= 2 - e^{-\gamma\rho_x} - e^{-\gamma\rho_y}
\end{aligned} \tag{3-17}$$

式中：Δ为距离损失，$p_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)$，$p_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)$，$\gamma = 2 - \Lambda$，$C_w$，$C_h$表示真实矩形框和预测矩形框最小宽度和高度。

(3) 形状损失：考虑预测框和真实框的形状差异，确保预测框在形状上尽可能接近真实框。公式 3-18 如下：

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta} \tag{3-18}$$

$$\omega_w = \frac{\left|w - w^{gt}\right|}{\max\left(w, w^{gt}\right)}, \omega_h = \frac{\left|h - h^{gt}\right|}{\max\left(h, h^{gt}\right)} \tag{3-19}$$

式中：ω为形状损失,(w,h)和(w^{gt},h^{gt})分别表示预测框和真实框的宽度与高度，θ为模型对形状损失的关注程度。

(4) 交并比：传统的交并比损失，通过计算预测框与真实框交集区域面积与两者并集区域面积的比值来表示。公式 3-20 如下：

$$IoU = \left|\frac{A \cap B}{A \cup B}\right| \tag{3-20}$$

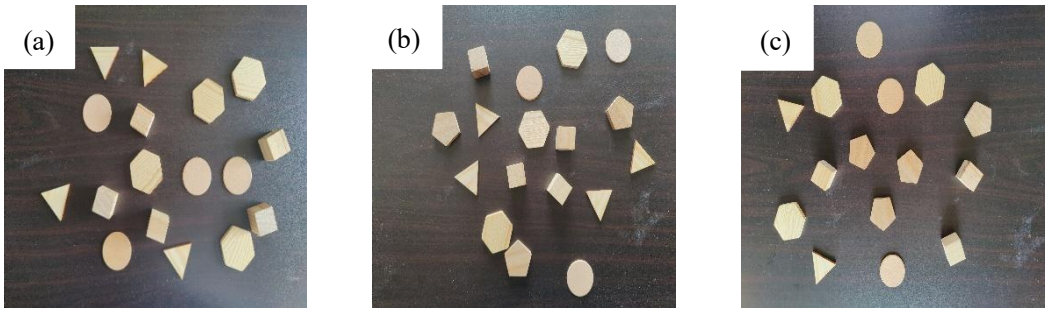
综上所述，最终 SIoU 函数表达式如下 3-21 所示：

$$L_{box} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{3-21}$$

3.3 实验结果分析

3.3.1 实验数据集采集

本文通过深度学习目标检测识别算法来实现对码垛材料零件的识别，为后续码垛机器人的码垛作业任务做良好的铺垫。为了验证改进后算法模型的有效型，本文选取了三角形、圆形、正方形、五边形、六边形 5 种不同形状零件木块作为码垛识别研究对象，考虑到机器人实际做业环境，通过不同的摆放方式模拟了码垛零件识别的应用场景，其中包括散放、堆叠和遮挡等应用场景。针对以上多种情况采用不同角度与不同距离拍摄获取实验数据，共拍摄图片 1000 张。部分零件材料数据集如图 3-16 所示，(a)、(b)、(c)三图为零件散放，(d)、(e)、(f)三图为零件的堆叠与遮挡。



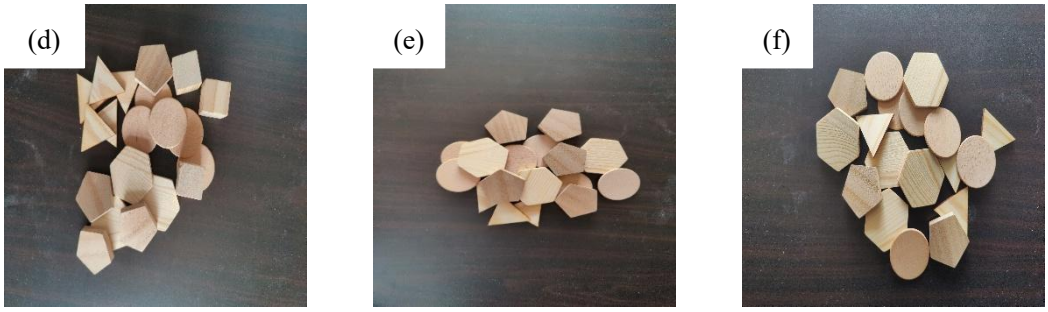


图 3-16 不同场景与角度下实验零件图：(a) 零件散放；(b) 零件散放；(c) 零件散放；(d) 零件堆叠与遮挡；(e) 零件堆叠与遮挡；(f) 零件堆叠与遮挡

3.3.2 数据集的构建

在码垛零件的数据集完成构建后，需要对目标数据进行标注，在进行目标检测识别时，对图像数据集进行标是非常重要的一步。目前常见的标注工具有，Labelimg、Labelme、Rectlabel 等。本文选用 Labelimg 软件对数据集进行标注划分，标注过程如图 3-17 所示，数据集的信息选用 YOLO 格式储存在 txt 文件中，在使用 Labelimg 标注时，将对应的 5 种零件圆形、三角形、正方形、五边形、六边形分别命名为 yuan、san、fang、wu、liu。具体标注过程如下，利用随机分配的方法，将整个数据集按 4:1:1 的比率划分为训练集、验证集以及测试集，分别为 800、200、200 张图片。在划分数据集时会删除无效标注、去除低分辨率图像，导致最终数据图片减少。

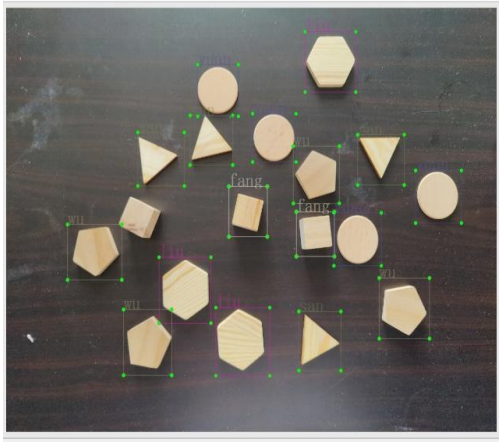


图 3-17 Labelimg 软件标注图

3.3.3 实验环境配置

为了验证本文所改模型的有效性，使用改进后 YOLOv8-MSC 网络模型与其它模型进行对比。实验环境基于 Windows 11 操作系统，使用 Pytorch 作为网络架构。为了提高模型训练速度，使用 CUDA11.3 与 cuDN8.3.1 来调用 GPU 进行加

速训练。具体实验平台参数如表 3-2 所示，训练模型采用的参数如表 3-3 所示。

表 3-2 实验平台参数

平台参数	配置
操作系统	Windows11
GPU	RTXA2000
显存	12G
网络架构	Pytorch1.11
开发语言	Python

表 3-3 训练模型参数

训练参数	参数设置
训练批次	200
批次大小	30
学习率	0.01
非极大值抑制阈值	0.5

本文输入码垛零件图片经过图片缩放处理其分辨率均为 640*640，训练批次大小设置为 30、训练轮次为 200 轮，其它参数均为 YOLOv8 模型中的原始默认参数。

3.3.4 评价指标

本研究采用的性能评价指标包括准确率(Precision)、召回率(Recall)和平均精确率均值(mAP)。此外，检测速度作为衡量模型效率的关键指标，定义为模型识别单张图片所需的时间，以毫秒(ms)为单位进行量化。准确率代表预测的正样本中有多少是真正的正样；召回率代表正样本的样本中，有多少被模型正确地检测出。mAP 是多个类别上的平均精度的平均值。AP 是衡量单个类别检测效果好坏的一种指标，它综合考虑了该类别上不同召回率水平下的精度。mAP 是通过计算所有类别的 AP 值的平均值得出的，主要用于衡量目标检测模型的整体性能。准确率、召回率与 mAP 值计算公式 3-22、3-23 与 3-24 如下：

$$Precision = \frac{TP}{TP + FP} \tag{3-22}$$

$$Recall = \frac{TP}{TP + FN} \tag{3-23}$$

$$mAP = \frac{\sum_{i=1}^{n} \int_{0}^{1} Precision\left(Recall\right) d\left(Recall\right)}{n} \tag{3-24}$$

式中：*TP* 代表模型准确识别并分类正确的样本数量；*FP* 为模型误将其他类别的样本归类为目标类别的数量；*FN* 则是模型未能识别出本应属于目标类别的

样本数量；*Precision* 为准确率；*Recall* 代表召回率；*n* 表示类别个数。

3.3.5 网络轻量化改进与分析

本文基于 YOLOv8 为模型基础，分别使用 MobileNetv3,shuffentnev2[64],作为 YOLOv8 的主干网络，对这 3 个模型的参数量和检测速度进行对比，对比结果如表 3-4 所示：

表 3-4 不同模型参数与检测速度对比

模型	权重大小(mb)	参数量(G)	检测速度(ms)
YOLOv8	12.73mb	8.2G	11.4
YOLOv8-shuffentnev2	6.12mb	5.7G	8.2
YOLOv8-MobileNetv3	2.61mb	2.8G	6.1

由表可知相比于其它两个模型，YOLOv8-MobileNetv3 网络模型的权重、参数量与检测速度都是最小，其权重大小为 2.61mb；参数量为 2.8G，检测速度达到了 6.1ms。其值相比与原始 YOLOv8 分别减少了 79.5%、65.9%与 46.5%。综上所述，本文选取 MobileNetv3，作为 YOLOv8 的主干网络。YOLOv8-MobileNetv3 模型识别结果图 3-18 如下：

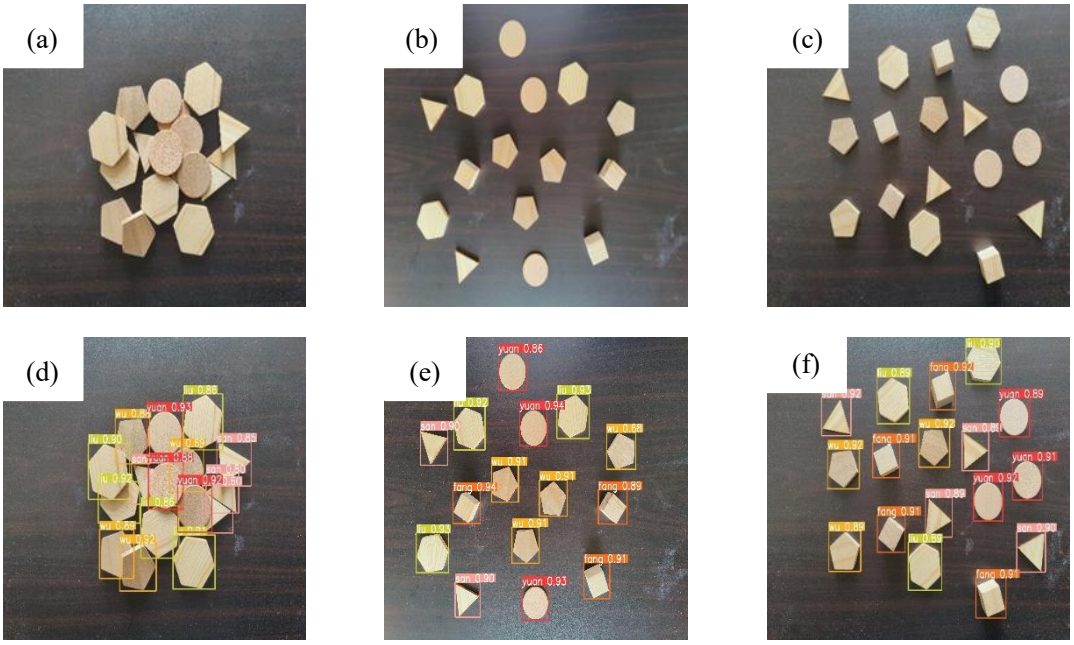


图 3-18 不同场景与角度下 YOLOv8-Mobilenetv3 模型检测识别图：(a) 原图堆叠与遮挡；(b) 原图散放；(c) 原图散放；(d) 堆叠与遮挡识别图；(e) 散放识别图；(f) 散放识别图

3.3.6 不同注意力机制对比

本文以 YOLOv8 为基础，在其主干网络中分别引入 CBAM，SA, GAM 三种注意力机制。为了直观地表示三种注意力机制，本文分别通过 Grade-CAM[65]中

(b) (c)

的激活热力图和 mAP 值对 YOLOv8 模型进行验证，实验对比结果如图 3-19 和表 3-5 所示。其中红色鲜艳区域表示模型提取特征越多，黄色次之，蓝色最少。

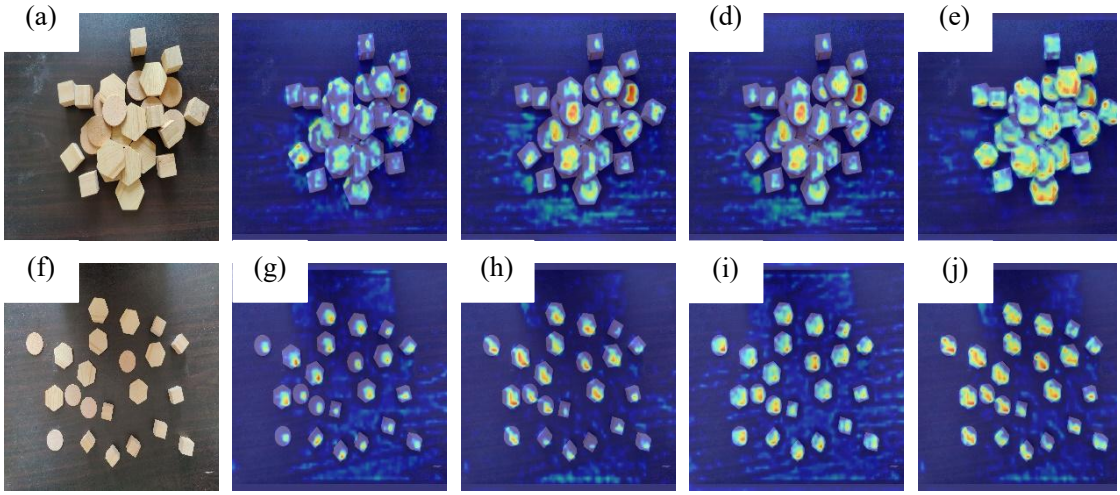


图 3-19 不同注意力机制的激活热力图：(a) 零件遮挡与堆叠；(b) YOLOv8；(c) YOLOv8-SA；(d) YOLOv8-GAM；(e) YOLOv8-CBAM；(f) 零件散放；(g) YOLOv8；(h) YOLOv8-SA；(i) YOLOv8-GAM；(j) YOLOv8-CBAM

表 3-5 不同注意力机制 mAP 值

模型	mAP
YOLOv8	0.971
YOLOv8+ SA	0.975
YOLOv8+ GAM	0.980
YOLOv8+CBAM	0.987

由图 3-20 可知，YOLOv8 主干网络引入三种不同注意力机制，其中 CBAM 注意力机制对码垛零件关注度程度最高，提取材料特征最多，能够更快的识别出目标。由表 3-5 可知 YOLOv8 在引入了 CBAM 注意力机制后，其 mAP 值最高，达到了 98.7%。综述所述 CBAM 注意力机制符合 YOLOv8 网络，所以本文采用 CBAM 注意力机制识别码垛所用的零件。

3.3.7 优化损失函数结果分析

由于 YOLOv8 采用 CIoU 损失函数，这种损失函数存在收敛速度慢，检测精度低等问题，为了解决这些问题，本文分别采用 GIoU、EIoU 和 SIoU 三种损失函数代替 YOLOv8 中的 CIoU 损失函数并行对比，对比如图 3-20 所示。

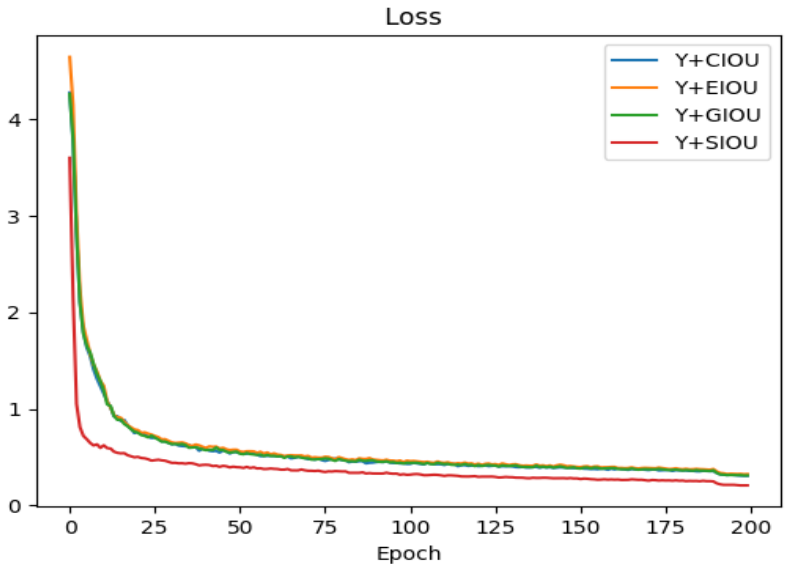


图 3-20 不同损失函数对比图

从图 3-20 可知，YOLOv8 采用 SIoU 损失函函数时，模型收敛速度最快，且损失函数收敛值小于采用其它损失函数收敛值，这表明采用 SIoU 损失函数使得模型预测框与真实框之间的差异也变小了，所以本文采用 SIoU 损失函数作为 YOLOv8 算法的损失函数。

为了进一步研究改进模型对码垛材料的识别，采用改进 YOLOv8-MSC 模型与原始 YOLOv8 模型进行对比消融实验，在相同的训练参数、设置及环境部署下得到的实验结果如表 3-6 和图 3-21 所示。表 3-6 中 M 代表 mobileNetV3 网络、S 代表 SIoU 损失函数、C 代表 CBAM 注意力机制、P 代表准确率(Precision)、R 代表召回率(Recall)、mAP0.5 代表在 50%的 IoU 阈值下的 mAP 平均值、mAP0.5-0.95 代表在 50-95%的 IoU 阈值范围内的 mAP 平均值。图 3-21 中 Y 代表 YOLOv8、Y+M 代表 YOLOv8+MobileNet、Y+M+C 代表 YOLOv8+MobileNetV3+CBAM、Y+M+S 代表 YOLOv8+MobileNetv3+SIoU 、Y+M+S+C 代表 YOLOv8+MobileNetv3+SIoU+CBAM。

表 3-6 YOLOv8-MSC 对比消融实验

M	S	C	P	R	mAP0.5	mAP0.5-0.95	权重参数(mb)	检测速度(ms)
×	×	×	0.962	0.940	0.972	0.786	5.96	9.83
√	×	×	0.960	0.938	0.970	0.777	2.53	6.04
√	×	√	0.970	0.971	0.981	0.792	2.57	6.08
√	√	×	0.971	0.972	0.983	0.801	2.54	6.05
√	√	√	0.973	0.975	0.985	0.819	2.60	6.11

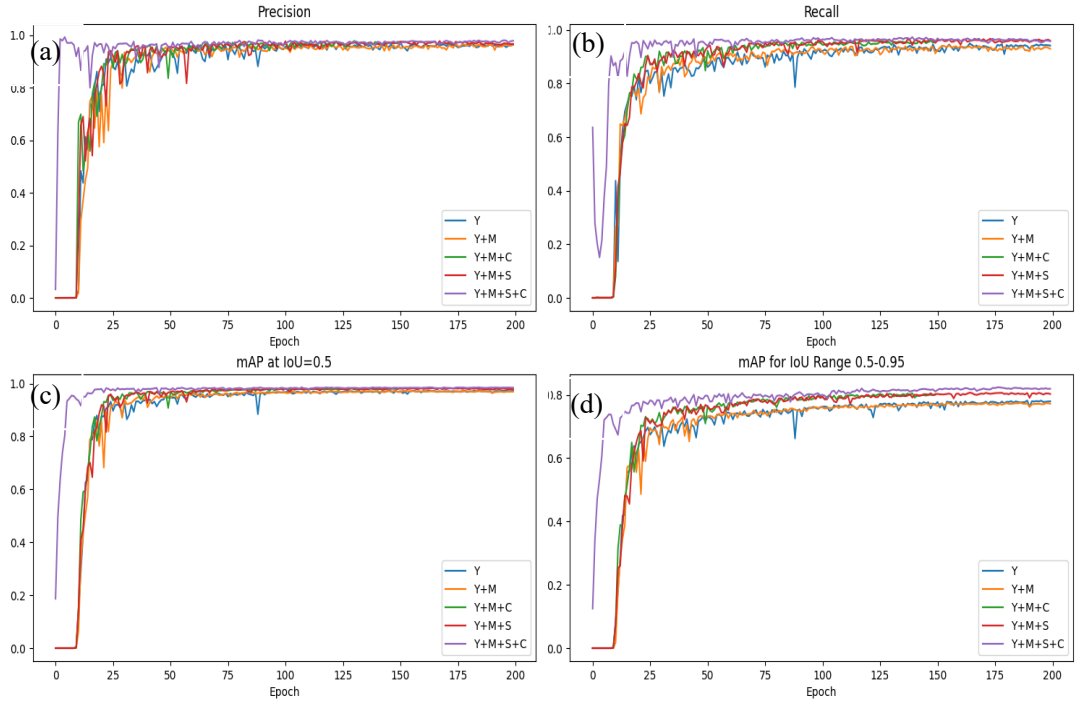


图 3-21 YOLOv8-MSC 对比消融实验图：(a) 准确率；(b) 召回率；
(c) 50%的 IoU 阈值下的 mAP 平均值；(d) 50-95%的 IoU 阈值范围内的 mAP 平均值

从表 3-6 可知，原始 YOLOv8 检测速度为 9.83ms，mAP 值为 0.972，权重参数为 5.96mb，在将 YOLOv8 主干网络替换成 MobileNetv3 后，检测速度为 6.04ms，比原 YOLOv8 减少了 3.79ms，说明 MobileNetv3 对原 YOLOv8 模型轻量化改进有明显提升效果。在 MobileNetv3 主干网络第 4 层和第 10 层分别引入 CBAM 注意力机制，与原模型和主干网络 MileNetv3 对比，mAP 值分别提升了 0.9%和 1.1%，说明 CBAM 可以更有效，更多地提取材料特征。在 MobileNetv3 主干网络下同时引入 SIoU 损失函数和 CBAM 注意力机制时，其 mAP 值最高，与原模型和主干网络 MileNetv3 对比，mAP 值分别提升 1.3%和 1.5%。

3.3.8 不同模型对比实验

为了验证改进算法的有效性，将本文 YOLOv8-MSC 算法与当前主流目标检测算法 Faster-rcnn、YOLOv5、YOLOv6、YOLOv8 做对比实验，选取检测速度和 mAP 值作为评价指标，实验结果如表 3-7 所示。

表 3-7 不同模型对比实验

模型	权重参(mb)	mAP0.5	检测速度(ms)
Faster-rcnn	118.27	0.853	220.98
YOLOv5n	5.16	0.935	7.83
YOLOv7n	8.50	0.949	11.31
YOLOv8n	5.96	0.972	9.83
YOLOv8-MSC	2.60	0.985	6.11

从表 3-7 可知 Faster-rcnn，权重参数和检测速度都是最大且 mAP 值最小，YOLOv8-MSC 相对于其它主流 YOLO 算法其参数量和检测速度都有所减少且 mAP 值最大，其权重参数为 2.60mb，检测速度为 6.11ms，mAP 值为 98.5%。由此可知，改进后 YOLOv8-MSC 算法在精度和检测上均有所提升。

3.3.9 实验结果分析

为了检测改进后的模型，本文采用 YOLOv8-MSC 与原始 YOLOv8 作比较，在面对散放、堆叠与遮挡场景下进行对比。设置参数 IoU=0.5，置信度为 0.25，两种模型检测结果如图 3-22 与 3-23 所示：

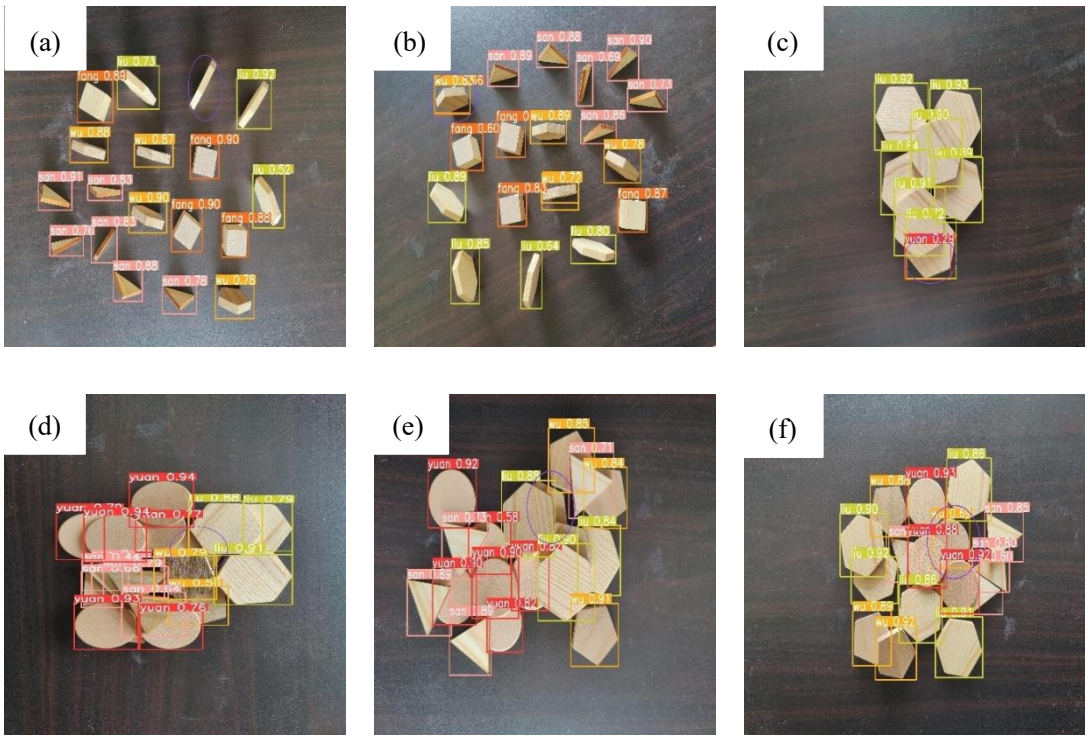
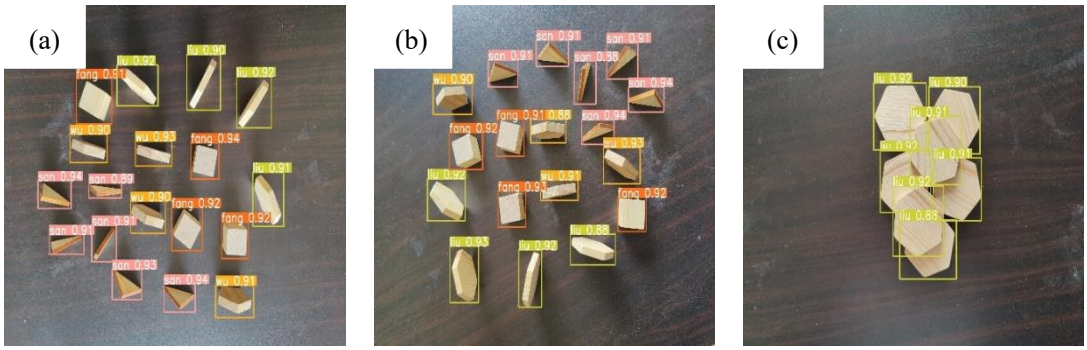


图 3-22 不同场景与角度下 YOLOv8 模型检测识别图：(a) 零件散放；(b) 零件散放；(c) 零件堆叠与遮挡检；(d) 零件堆叠与遮挡；(e) 零件堆叠与遮挡；(f) 零件堆叠与遮挡



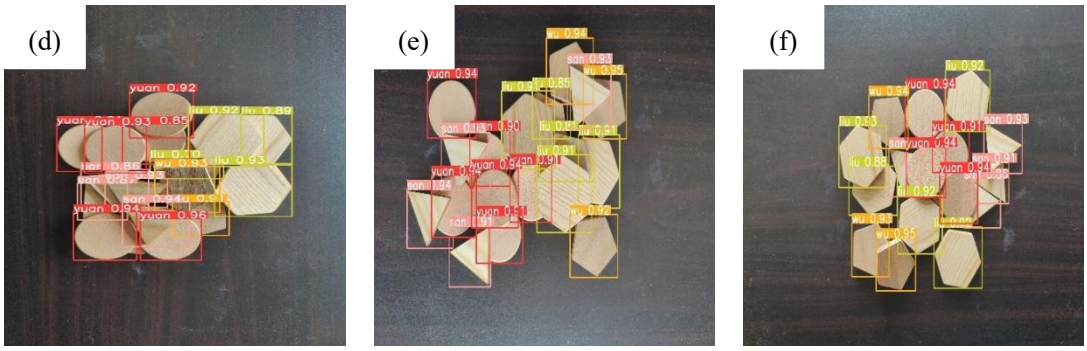


图 3-23 不同场景与角度下 YOLOv8-MSC 模型检测识别图。(a) 零件散放；(b) 零件散放；(c) 零件堆叠与遮挡检；(d) 零件堆叠与遮挡；(e) 零件堆叠与遮挡；(f) 零件堆叠与遮挡

由上述实验结果表明，YOLOv8 模型在检测识别中存在误差与漏检等情况，图 3-22 中已用紫色圆圈标注出现。在面对散放零件场景下，由图 3-22 中的(a)和(b)与图 3-23 中的(a)和(b)相比较可以发现，YOLOv8 模型检测漏检了一个六边形与重复误检了一个五边形；在面对堆叠与遮挡场景下，由图 3-22 中的(c)与图 3-23 中的(c)对比可知，模型错误的将六边形检测识别成圆形；由图 3-22 中的(d)和(e)与图 3-23 中的(d)和(e)对比可知，YOLOv8 模型漏检了一个六边形；由图 3-22 中的(f)与图 3-23 中的(f)对比可知，YOLOv8 模型错误的将圆形检测识别成五边形。而相比较于 YOLOv8 模型，改进后 YOLOv8-MSC 模型在面对散放与堆叠的场景下，均没有出现误检与漏检等情况，在检测识别精度上表现的更加优异。

3.4 本章小结

本章首先介绍了 YOLOv8 算法原理，然后讲解了零件木块图片的采集与划分；其次对 YOLOv8 算法做出相应的改进，将改进后的模型命名为 YOLOv8-MSC；最后通过 5 组对比消融实验可知，其 mAP 值提升了 1.3%，精度提升了 1.1%，检测速度提升了 38%。改进后的算法模型能够更有效快速准确的识别散放、堆叠和遮挡场景下码垛材料，可以满足码垛机器人快速精准识别的要求。

第 4 章 基于双目立体视觉定位方法研究

上文针对码垛零件的识别检测做了研究，但仅得到零件的目标检测信息，并不能满足码垛机器人对零件的抓取作业。只有得到了目标零件的三维坐标完成对目标零件的定位，才能使得码垛机器人实行下一步做业。因此本章将基于 SGBM 双目立体匹配算法，完成对零件的定位方法研究，从而获取目标零件的三维坐标。

4.1 双目立体视觉系统的组成

本文双目视觉系统的组成包括计算机与双目摄像头两个部分，目标图片获取通过汇博视捷双目摄像头，如图 4-1 所示，其具有成本低、读取速度快、体积小等优点，非常适合部署在小型移动设备上。因此使用双目视觉摄像头，通过 USB 接口与计算机相连接。相机的详细配置如下：具备 400 万像素，其帧率与基线分别为 30fps 和 3mm，支持最大图像分辨率为 3840×1080，拍摄范围为 80°。为了保证后续定位实验的精度与准确性，左右相机采用平行的安装方式。双目相机的具体图片如 4-1 所示。

图 4-1 汇博视捷双目相机

4.2 双目标定

双目相机标定方法依据标定过程中是否使用参照物，可分为传统标定方法和自标定方法两种主要类型[66]。在传统标定方法中，通常需要借助精确设计的外部标定物，如棋盘格或圆点格，这些标定板的设计使得能够从多角度拍摄图像，从而提供丰富的数据用于计算相机的内外参数。传统标定方法的优点包括高精度、广泛的适用性以及成熟的理论基础。其缺点则体现在操作复杂性、对设备依赖性

强以及计算复杂度高，因此标定效率较低。自标定方法是一种不依赖于外部参照物或特定场景特征，仅通过图像间的对应关系即可完成标定，因而具有显著优势：其灵活性高、操作便捷，且适用于更广泛的应用场景，但由于自标定方法依赖于图像之间的对应关系，导致其标定精度通常低于使用标定板的传统方法，同时该方法还具有对场景依赖性较强，鲁棒性不足计算复杂度高等方面的缺点。

本文采用张正友标定法[67]作为双目相机的标定方法，该方法是传统标定方法与自标定方法相结合，既不需要复杂的标定设备，同时又能保证较高的标定精度。张正友标定法在标定过程中只需要一张尺寸规定棋盘格图片，该方法的核心是通过改变双目相机与棋盘格标定板之间的相对位置和角度进行多组拍摄，从而获取棋盘格的图片，然后所拍摄的图片导入 MATLAB 软件中进行操作，最后就可以获取相应的相机参数。具体原理如下：首先，在三维空间中选定一个 P 点，该点的坐标为(X，Y，Z)，该点在像素坐标系下的投影坐标为 p(u，v)，为简化计算，假设标定过程在 $Z=0$ 的平面上进行，并将平面棋盘格的坐标系定义为世界坐标系。由相机针孔原理可知：空间坐标系下的 P 点与像素坐标系下的 p 点对应转换关系如公式 4-1 所示：

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=A\begin{bmatrix}r_1 & r_2 & r_3 & t\end{bmatrix}\begin{bmatrix}X\\Y\\0\\1\end{bmatrix}=A\begin{bmatrix}r_1 & r_2 & t\end{bmatrix}\begin{bmatrix}X\\Y\\1\end{bmatrix}=A\begin{bmatrix}R & t\end{bmatrix}\begin{bmatrix}X\\Y\\1\end{bmatrix} \tag{4-1}$$

式中：s 代表尺度因子，表示是点投影到图像平面上时所采用的比例系数；[$\boldsymbol{R}$　$\boldsymbol{t}$]代表相机坐标系相对于世界坐标系下的旋转向量与平移向量；$\boldsymbol{A}$ 代表相机的相关参数矩阵，具体公式 4-2 如下：

$$A=\begin{bmatrix}f_x & \alpha & u_0\\0 & f_y & v_0\\0 & 0 & 1\end{bmatrix} \tag{4-2}$$

式中：$f_x=\frac{f}{dx}$，$f_y=\frac{f}{dy}$，f_x，f_y 分别表示像素坐标系中 u 轴与 v 轴方向上的尺度变换因子，α 为坐标轴间的扭曲参数，在理想成像条件下（即像素阵列完全对齐）其值为零，（u_0，v_0）表示像素坐标系的中心点坐标。

P 点与 p 点之间可以通过单应性矩阵表示，具体表示公式 4-3 如下：

$$H = A\begin{bmatrix} r_1 & r_2 & t \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} \tag{4-3}$$

带入公式可得公式 4-4 如下：

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{4-4}$$

由公式可知，***H*** 是一个 3×1 的矩阵，其中包含了 8 个待求解的未知参数，因此需要八个方程组才能求解出 ***H*** 中的八个未知变量，即需要四组相互对应的投影点坐标。上述求解得出理想状态下的 ***H***，其中包含了相机的内外参数。实际状态下的矩阵 ***H*** 为理想转态下的 λ 倍，可得公式 4-5 如下：

$$\begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} = \lambda A\begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \tag{4-5}$$

旋转矩阵 ***R*** 中存在向量相互正交关系，即如公式 4-6 所示：

$$\begin{cases} r_1^T r_2 = 0 \\ r_1^T r_1 = r_2 r_2^T = 1 \end{cases} \tag{4-6}$$

可推出公式 4-7 如下：

$$\begin{cases} h_1^T A^{-T} A^{-1} h_2 = 0 \\ h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-T} h_2 \end{cases} \tag{4-7}$$

再令：

$$B = A^{-T}A^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \tag{4-8}$$

$$B = \begin{bmatrix} \dfrac{1}{\alpha^2} & -\dfrac{\gamma}{\alpha^2\beta} & \dfrac{v_0\gamma - u_0\beta}{\alpha^2\beta} \\ -\dfrac{\gamma}{\alpha^2\beta} & \dfrac{\gamma^2}{\alpha^2\beta^2} + \dfrac{1}{\beta^2} & -\dfrac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \dfrac{v_0}{\beta^2} \\ \dfrac{v_0\gamma - u_0\beta}{\alpha^2\beta} & -\dfrac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \dfrac{v_0}{\beta^2} & \dfrac{(v_0\gamma - u_0\beta)^2}{\alpha^2\beta^2} + \dfrac{v_0}{\beta^2} + 1 \end{bmatrix} \tag{4-9}$$

式中：*α=1/dx*，*β=1/dy*，矩阵 ***B*** 含有 6 个未知参数，将表示为 6×1 的向量形式，如公式 4-10 所示：

$$b = \begin{bmatrix} B_{11} & B_{12} & B_{22} & B_{13} & B_{23} & B_{33} \end{bmatrix} T \tag{4-10}$$

矩阵 ***H*** 第 *i* 列向量如 4-11 所示：

$$h_i = \begin{bmatrix} h_{i1} & h_{i2} & h_{i3} \end{bmatrix} \tag{4-11}$$

故得公式 4-12 如下：

$$h_i A^{-T} A^{-1} h_j = h_j b h_j = V_{ij}{}^T b \tag{4-12}$$

式中 $v_{ij} = \begin{bmatrix} h_{i1}h_{j1} & h_{i1}h_{j1} + h_{i2}h_{j1} & h_{i2}h_{j2} & h_{i3}h_{j1} + h_{i1}h_{j1} & h_{i3}h_{j2} + h_{i2}h_{j3} & h_{i3}h_{j3} \end{bmatrix} T$

从任何一张标定图像可以推导出公式 4-13：

$$\begin{cases} h_1^T A^{-1} A^{-1} h_2 = 0 \\ h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2 = 1 \end{cases} \rightarrow \begin{cases} v_{12}{}^T b = 0 \\ v_{11} b = v_{22} b \end{cases} \tag{4-13}$$

用矩阵的形式表示如公式 4-14：

$$\begin{bmatrix} v_{12}{}^T \\ \left(v_{11} - v_{22}\right)^T \end{bmatrix} b = 0 \tag{4-14}$$

求解可得到相机的参数值如公式 4-15 所示：

$$\begin{cases} v_0 = \left(B_{12}B_{12} - B_{11}B_{23}\right) / \left(B_{11}B_{22} - B_{12}{}^2\right) \\ \lambda = B_{33} - \left[B_{13}{}^2 + v_0\left(B_{12}B_{13} - B_{11}B_{23}\right)\right] / B_{11} \\ \alpha = \sqrt{\lambda / B_{11}} \\ \beta = \sqrt{\lambda B_{11} / \left(B_{11}B_{22} - B_{12}{}^2\right)} \\ \gamma = -B_{12}\alpha^2\beta / \lambda \\ u_0 = \gamma v_0 / \alpha - B_{13}\beta^2 / \lambda \end{cases} \tag{4-15}$$

由上式可求出相机的内参矩阵 $\boldsymbol{A}$ 的全部未知变量，综上所得结合内参矩阵求解出相机外参矩阵，如公式 4-16 所示：

$$\begin{cases} r_1 = \lambda A^{-1} h_1 \\ r_2 = \lambda A^{-1} h_2 \\ r_3 = r_1 r_2 \\ t = \lambda A^{-1} h_3 \end{cases} \tag{4-16}$$

式中：$\lambda = \dfrac{1}{\left\| A^{-1} h_1 \right\|} = \dfrac{1}{\left\| A^{-2} h_2 \right\|}$，由此可以求出相机的外参矩阵，完成双目相机的标定。

4.3 双目校正

双目校正是一种利用将双目摄像头拍摄的图像进行对齐和校准的过程，以便它们能够更准确地测量视差并计算物体的深度信息。在理想条件下，双目相机的成像平面共面且对齐，此时极点位于无穷远处。这种配置下，同一物点在两个相机成像面上形成的像点仅在水平位置(x 轴)上有差异，而垂直位置(y 轴)保持不变，从而形成所谓的立体视差。然而在实际环境中，由于生产和安装等情况双目相机往往无法达到理想状态，因此需要对双目相机进行立体校正。如图 4-2 所示为双目校正的示意图。双目相机拍摄到的原始图像分别位于两个不共面且行未对准的图像平面上。为了进行有效的处理，需要对这两幅图像进行校正，以确保校正后的图像平面共面并对准。本文采用 Bouguet[68]算法对双目相机进行校正，它通过旋转矩阵和平移向量将两个相机坐标系调整到共面且行对齐的状态。

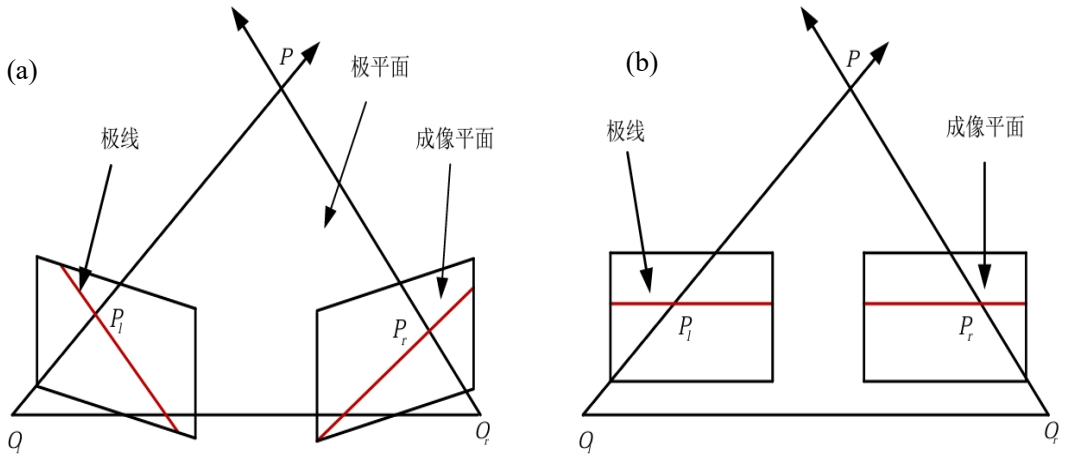


图 4-2 校正前后示意图：(a) 校正前示意图；(b) 校正后示意图

4.4 双目标定与校正实验

本文采用汇博视捷双目相机进行标定与校正实验，相机的相关参数配置，在 4.1 小节已详细介绍过，本小节不再做具体介绍。标定过程中使用的软件包括 MATLAB2022b、OpenCV 和 PyCharm。双目相机的标定与校正流程如下：(1) 使用双目相机分别拍摄 40 张棋盘格图像，并将这些图像分别存储在两个文件夹中；(2) 将双目相机拍摄的图像同时导入 MATLAB 软件中进行标定，在完成双目相机后获取相机的相关参数并导出。双目相机标定与校正的流程图 4-3 如下：

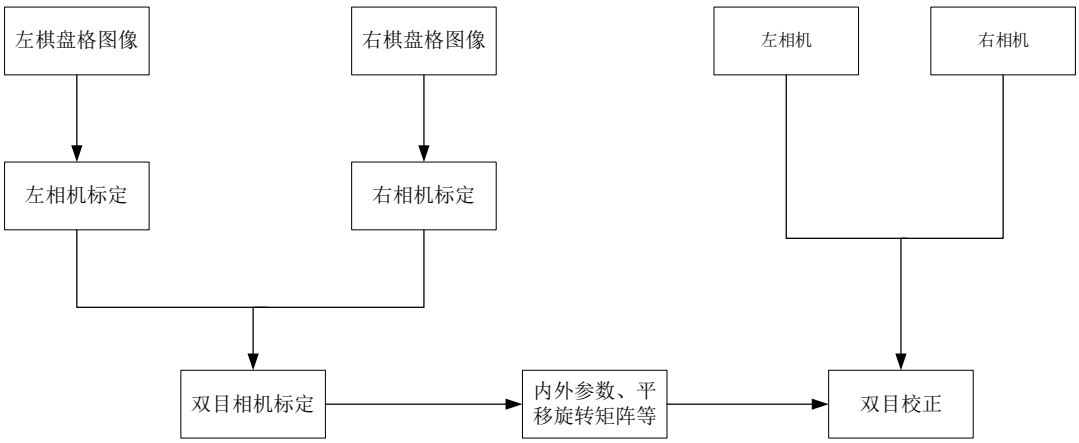


图 4-3 双目相机标定与校正流程图

本文采用张正友棋盘格标定法，并结合 MATLAB 软件中的相机标定工具箱“Stereo Camera Calibration Toolbox”进行双目标定，所采用的标定板为黑白棋盘格，其尺寸规格为 13×17，每个方格大小均为 10×10mm，标定所用的棋盘格如图 4-4 所示。在标定前首先使用双目相机从不同角度与位置对标定板进行拍摄，共获得照片 40 张，分别将左右双目相机拍摄的照片储存在两个不同的文件夹中。

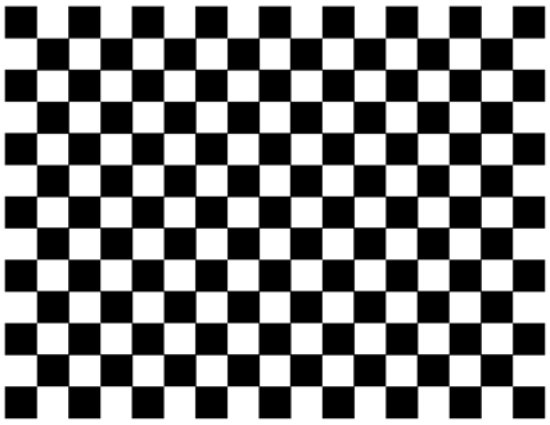


图 4-4 棋盘格标定板图

1. 标定流程

首先，打开 MATLAB 软件，在工具栏中找到并点击“Stereo Camera Calibration Toolbox”工具箱以启动标定功能。接着，将左右双目相机拍摄的棋盘格图像导入该工具箱中。在导入图像之前，需输入棋盘格的尺寸 10×10mm。完成导入后，点击工具箱中的角点提取按钮，具体操作如图 4-5 所示。对于未识别或识别错误的角点，可以通过右键点击删除，同时对应的棋盘格图像也会被移除。

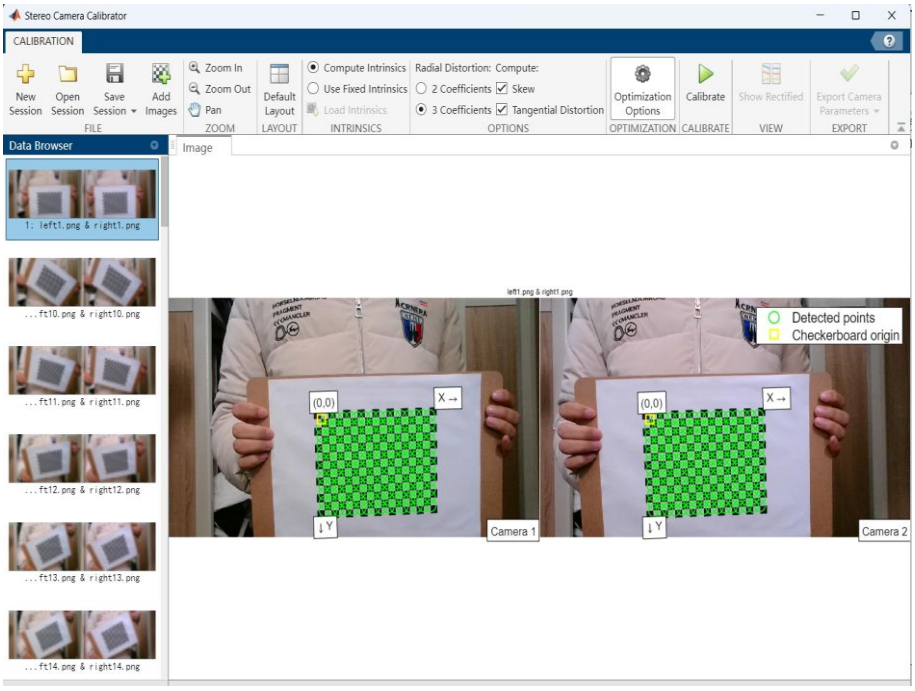


图 4-5 标定预处理示意图

点击软件工具箱中的 Calibrate 按钮，生成左右相机的重投影误差图，如图 4-6 所示，同时展示相机与标定板在空间中的相对位置关系，如图 4-7 所示。对于重投影误差较大或异常的照片，可以将其删除并重新计算误差，如图 4-8 所示。通过删除误差较大的照片，将重投影误差控制在 0.5 以内。随后，通过代码导出左右相机的内参、外参及畸变系数等参数，具体代码如图 4-9 所示，相关参数结果如图 4-10 和表 4-1 所示。

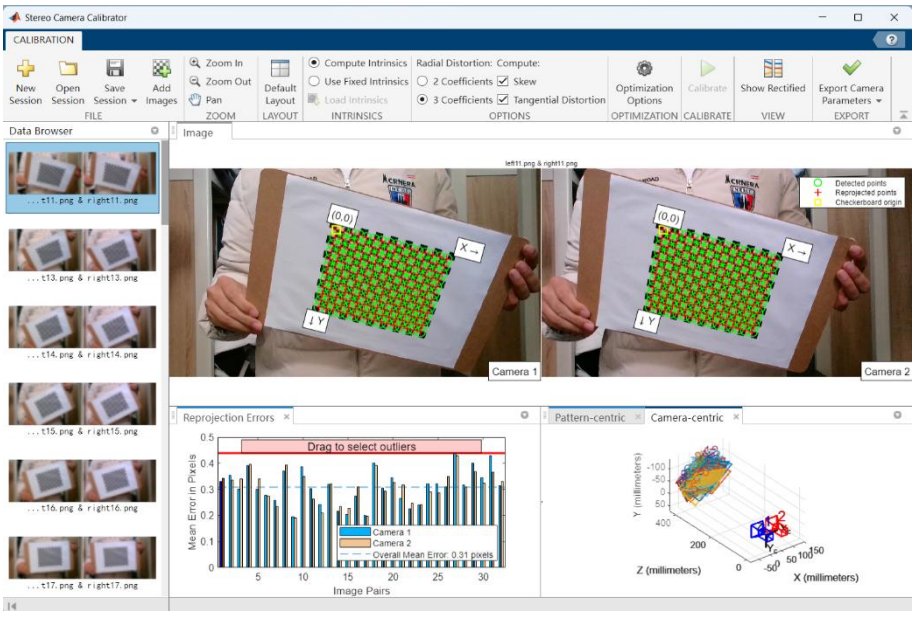


图 4-6 标定处理结果图

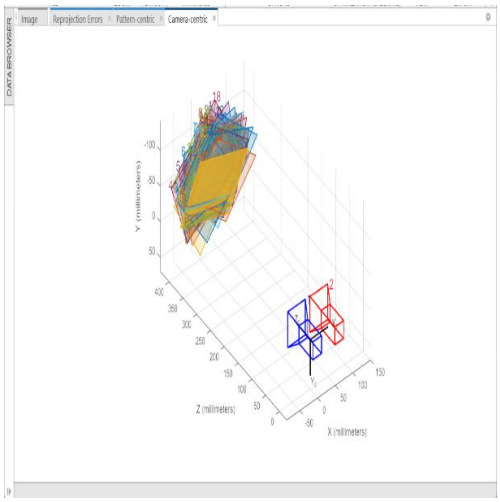


图 4-7 相机空间位置示意图

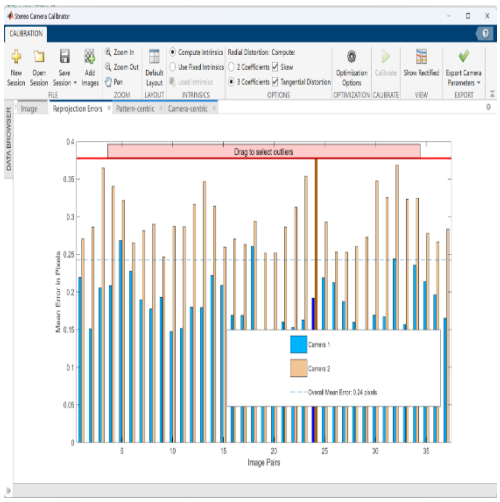
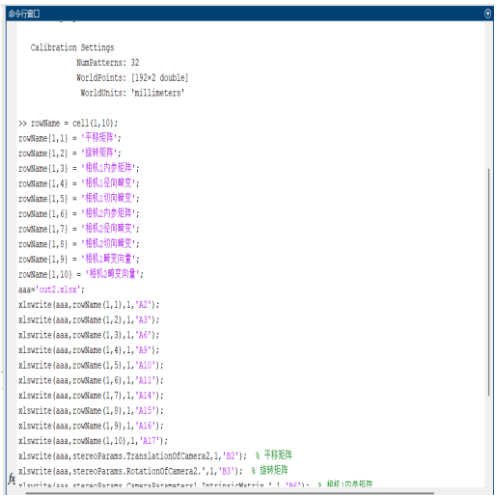


图 4-8 重投影误差



```
Calibration Settings
        NumPatterns: 32
        WorldPoints: [192×2 double]
        WorldUnits: 'millimeters'

>> rowName = cell(1,10);
rowName{1,1} = '平移矩阵';
rowName{1,2} = '旋转矩阵';
rowName{1,3} = '相机1内参矩阵';
rowName{1,4} = '相机1径向畸变';
rowName{1,5} = '相机1切向畸变';
rowName{1,6} = '相机2内参矩阵';
rowName{1,7} = '相机2径向畸变';
rowName{1,8} = '相机2切向畸变';
rowName{1,9} = '相机1畸变向量';
rowName{1,10} = '相机2畸变向量';
aaa='out2.xlsx';
xlswrite(aaa,rowName(1,1),1,'A2');
xlswrite(aaa,rowName(1,2),1,'A3');
xlswrite(aaa,rowName(1,3),1,'A6');
xlswrite(aaa,rowName(1,4),1,'A9');
xlswrite(aaa,rowName(1,5),1,'A10');
xlswrite(aaa,rowName(1,6),1,'A11');
xlswrite(aaa,rowName(1,7),1,'A14');
xlswrite(aaa,rowName(1,8),1,'A15');
xlswrite(aaa,rowName(1,9),1,'A16');
xlswrite(aaa,rowName(1,10),1,'A17');
xlswrite(aaa,stereoParams.TranslationOfCamera2,1,'B2');  % 平移矩阵
xlswrite(aaa,stereoParams.RotationOfCamera2.',1,'B3');  % 旋转矩阵
```

图 4-9 代码示意图

平移矩阵	-61.8755	0.007234	1.006606		
旋转矩阵	0.999974	0.001514	-0.00711		
	-0.00149	0.999994	0.003201		
	0.00711	-0.00319	0.99997		
相机1内参	973.3437	1.8545	657.6625		
	0	976.4326	467.2264		
	0	0	1		
相机1径向	-0.13108	0.215049	1.47718		
相机1切向	-0.00592	0.003162			
相机2内参	974.081	-1.67712	671.56		
	0	977.1881	465.2005		
	0	0	1		
相机2径向	-0.06286	0.02793	0.515028		
相机2切向	-0.00248	-0.00136			
相机1畸变	-0.13108	0.215049	-0.00592	0.003162	1.47718
相机2畸变	-0.06286	0.02793	-0.00248	-0.00136	0.515028

图 4-10 双目相机相关参数

表 4-1 双目相机相关参数

参数	左相机	右相机
内参矩阵	$\begin{bmatrix} 973.344 & 1.855 & 657.633 \\ 0 & 976.433 & 467.226 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 974.081 & -1.677 & 671.560 \\ 0 & 977.188 & 465.2010 \\ 0 & 0 & 1 \end{bmatrix}$
畸变系数	[-0.131 0.215 -0.006 0.003 1.477]	[-0.063 0.028 -0.002 -00001 0.515]
外参系数	$R=\begin{bmatrix} 1.000 & 0.002 & -0.007 \\ 0 & 1.000 & 0.003 \\ 0.007 & -0.003 & 1.000 \end{bmatrix}$	$T=\begin{bmatrix} -61.879 & 0.007 & 1007 \end{bmatrix}$

2. 双目校正流程

在完成双目标定后，将获取的相机相关参数输入 PyCharm 中，利用 OpenCv 行双目标定的步骤如下：首先加载相机参数，调用 cv2.stereoRectify()函数计算校正变换参数；然后将结果输入 cv2.initUndistortRectifyMap()函数中，生成校正映

射参数 map x 和 map y。最后利用 cv2.remap()函数完成图像校正。校正后的图像基本实现了行对准和共面要求。校正前后图像对比如 4-11 与 4-12 图所示：

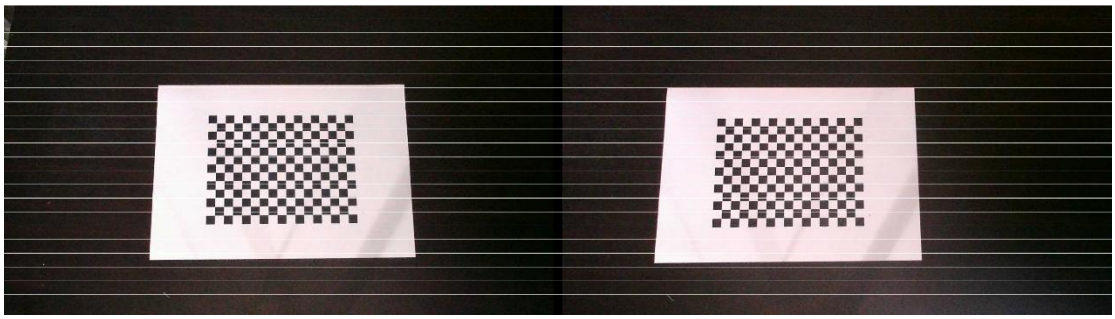


图 4-11 校正前图像

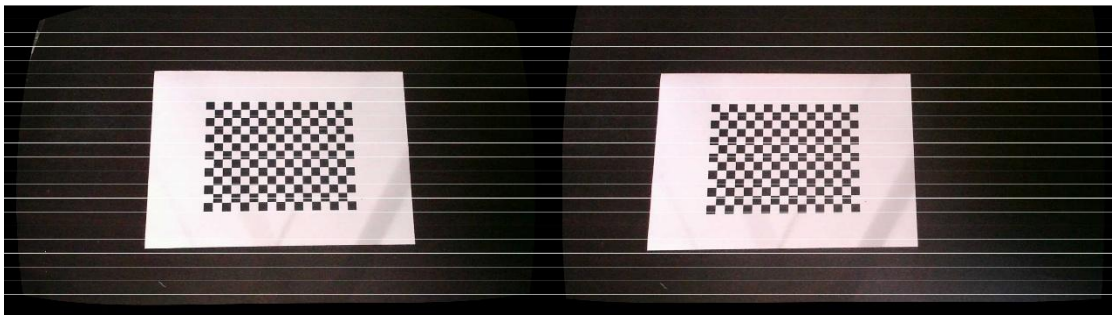


图 4-12 校正后图像

4.5 立体匹配

立体匹配是计算机视觉中的一项核心技术，旨在将二维图像对转换成三维场景。这一过程通过计算两幅图像之间的视差，获取目标物体图像中的深度信息，进而实现三维场景的重建和目标定位等功能。立体匹配的核心在于找出左右视图中对应的像素点，这通常通过计算每个像素的视差值来实现，即两个视图之间同一物体点的水平和垂直位移差。立体匹配算法主要有：SBGM[69]、BM[70]、GC[71]这三种算法，其算法的主要步骤如下：

(1) 匹配代价计算。其目的是为了衡量待匹配像素与候选像素之间的相似度。无论这两个像素是否确实为同名点，都可以利用匹配代价函数来计算它们之间的代价值；该值越低，则表明这两像素间的关联性越强，成为同名点的概率也相应增加。在匹配同名点时，通常为每个像素定义视差搜索区间 D(Dmin~Dmax)，并搜索范围被限制在该区域内，通过构建一个大小为 W×H×D 的三维矩阵 $\boldsymbol{C}$(其中 W 表示影像的宽度，H 表示影像的高度)，记录每个像素在视差区间内各视差级别对应的匹配代价。该矩阵 $\boldsymbol{C}$ 通常称为视差空间图像 DSI(Disparity Space Image)。常见的匹配代价计算有灰度绝对值法(AD)[73]、灰度差平方法(SD)具体公式如 4-

17 与 4-18 所示：

$$Cdata_{AD}\left(dx\right)=abs\left(I_L\left(x\right)-I_R\left(x+dx\right)\right) \tag{4-17}$$

$$Cdata_{SD}\left(dx\right)=\left(I_L\left(x\right)-I_R\left(x+dx\right)\right)^2 \tag{4-18}$$

(2) 匹配代价聚合。匹配代价聚合是在计算每个像素的视差代价后，进一步提高立体匹配精度的过程。常见的聚合方法有两种：局部代价聚合与全局代价聚合。局部代价聚合策略通常考虑每个像素邻域内的视差代价，将近邻像素的代价加到当前像素上，以增强像素间的匹配一致性。全局代价聚合是指将整个图像或者较大范围区域的信息整合起来确定每个像素点的最优视差。

(3) 视差计算。通过分析聚合后的代价矩阵，确定每个像素的最佳视差值，即选取最小匹配代价对应的视差作为最优解，这个过程就叫视差计算[74]。

(4) 视差优化。在完成立体匹配计算生成视差图后，往往可能会存在一些匹配错误和噪点等问题，会对后续的计算产生误差。需对初步生成的视差图进行优化处理，这一过程包括涵盖平滑处理、错误视差剔除以及像素级精度优化等步骤，从而提高视差图的精确度。

4.5.1 匹配基元

在计算机视觉和图像处理中，匹配基元有多种类型，以下是一些常见的匹配基元：

(1) 点特征匹配基元主要包括角点和边缘点。角点是图像中方向变化明显或曲率较大的点，如墙角、道路交叉点等，具有高可重复性和稳定性，适用于不同视角和光照条件，常用于立体匹配。边缘点是灰度值急剧变化的点，多位于物体边缘或轮廓上，对灰度变化和光照条件不敏感，是重要的图像特征。通过检测边缘点，可获取物体边缘信息，为立体匹配提供支持。

(2) 线特征匹配基元主要包括边缘线段和轮廓线。边缘线段由连续的边缘点组成，能够反映物体的边缘轮廓，比单个边缘点包含更多空间信息，更适合描述物体的形状和结构。在立体匹配中，通过提取图像中的边缘线段并在左右图像中寻找对应线段来实现匹配。轮廓线是物体的外部边界线，能够完整勾勒物体形状，具有较高的稳定性和可识别性，适用于区分不同物体和场景。在立体匹配中，可利用轮廓线确定物体的位置和姿态。

(3) 区域特征匹配基元主要包括灰度区域和纹理区域。灰度区域由图像中灰

度值相似的像素组成；纹理区域则指具有特定纹理模式的区域，如砖墙、木地板等，包含丰富的表面信息，能反映物体的材质和表面特性。在立体匹配中，利用纹理区域的纹理特征可提高匹配的准确性和可靠性，但纹理区域易受遮挡、光线变化和重复纹理等因素的影响。

4.5.2 约束准则

立体匹配约束准则是确保匹配准确性和效率的关键规则。以下是一些常见的立体匹配约束准则：

(1) 极线性约束。它由一个三维点及其在两个相机成像平面上的投影点共同定义，这三个点确定了一个包含基线(连接两相机中心的直线)的平面，这个平面被称为极平面。其作用在于大幅减少匹配点的搜索范围，进而提高匹配效率，是立体匹配中最基本且最重要的约束准则之一。

(2) 唯一性约束。在立体匹配中，每幅图像中的任意一个像素点仅与另一幅图像中的一个像素点对应，即每个基元只能匹配另一幅图像中的唯一基元，因此每个匹配基元最多对应一个视差值。这种唯一性约束显著简化了匹配过程，消除了多匹配结果可能导致的歧义，从而提高了匹配结果的准确性和可靠性。

(3) 相似性约束：在不同视角下，三维物体的投影匹配基元(如点、块、线)需具备相同或相似的属性。由于光照等因素可能限制其适用性，但该约束确保了匹配像素点在属性上的一致性，从而提升匹配的准确性和可靠性。

本文采用极线性约束，对双目相机进行校正，使双目相机达到行对准且共面的需求，提升了后续零件定位精度。

4.5.3 BM 局部立体匹配算法

BM 算法的核心思想是通过滑动窗口的方式，在一幅图像中选取一个固定大小区域作为为参考块，然后在另一张图像中寻找与该参考块最相似的对应区域。相似度的度量通常采用绝对误差和(SAD)或平方误差和(SSD)。具体来说，对于左目图像中的每个像素点，算法会在右目图像中以该像素点为中心，定义一个与左目图像相同大小的搜索区域，并在该区域内寻找与左目图像块最相似的块，相似度最高的块的中心像素即为匹配点。BM 算法简单直观，易于实现，匹配效果较好，但也存在对复杂场景处理不足、计算量大等缺点。

4.5.4 GC 全局立体匹配算法

GC 全局立体匹配算法通过建立一个全局能量函数来优化视差估计，该能量函数通常由数据项和平滑项两部分构成。数据项用于评估匹配效果，体现两个图像块之间的相似度；平滑项则引入了场景的约束，确保匹配结果在空间上保持平滑和一致。在具体实现中，算法会计算左右图像对应像素之间的匹配代价，并通过最小化全局能量函数来确定最佳视差值。这个过程与图割(Graph Cuts)算法密切相关，即在一个加权无向图上寻找最小割，以实现能量函数的最小化。该算法具有高准确性和灵活性，但由于其计算复杂度较高，难以满足复杂场景下的码垛要求。

4.5.5 SGBM 半全局立体匹配

SGBM 半全局立体匹配算法结合了局部匹配算法的高效性和全局匹配算法的高精度优势，旨在提高视差图的质量并降低计算复杂度。其主要是，首先选择视差值的方式为图像中的每个像素生成视差从而形成初始的视差图，然后预先设定的全局能量函数对初始视差图进行处理，最后通过求解出该函数的最小值得到最终视差图[76]，该函数具体如公式如 4-19 所示：

$$E(D)=\sum_{p}\left\{C(p,D_p)+\sum_{q\in N_p}p_1 I\left[\left|D_p-D_q\right|=1\right]+\sum_{q\in N_p}P_2 I\left[\left|D_p-D_q\right|>1\right]\right\} \quad (4\text{-}19)$$

式中：D 代表原始视图差，$E(D)$为全局能量函数，其中 p 和 q 分别表示图像中的两个像素点，NP 表示像素点 p 的八邻域像素集合。$C(p，D_P)$像素代表点 p 在视差 D_P 下的匹配代价值，p_1 和 p_2 分别表示像素点 p 与其邻域像素点视差值相差 1 和大于 1 时的惩罚系数[77]。上述公式在进行图像最优化求解中时，过程比较复杂，为了简化求解过程。通过一维分解公式，具体步骤如下：选择任意像素点 P，对其八个邻域点包括（上、下、左、右等）进行一维分解，再对分解后的动态规划进行求解，具体公式如 4-20 所示：

$$L_r(p,d)=C(p,d)+min\begin{Bmatrix}L_r(p-r,d)\\L_r(p-r,d-1)+p_1\\\min_i L_r(p-r,i)+p_2\end{Bmatrix}-\min_k L_r(p-r,k) \quad (4\text{-}20)$$

式中：$L_r(p，d)$表示像素点 p 在路径 r 上视差值为 d 时的能量值。当八个路径的能量值计算完毕后，通过累加得到像素点 p 的总能量值，选取最小的视差值作为像素点 p 的最优视差值[78]，如公式 4-21 所示：

$$S(p,d)=\sum_{r}L_r(p,d) \tag{4-21}$$

4.5.6 立体匹配算法对比与分析

通过描述对比三种立体匹配算法可知，GC 全局立体匹配算法计算过程复杂程度高，不适应于码垛材料零件识别的应用场景。因此，本章只选用 SGBM 立体匹配算法与 BM 立体匹配算法进行实验对比分析。立体匹配对象为码垛材料零件，针对码垛所遇见的场景，本章模拟了散放、堆叠与遮挡两种场景下的匹配信息，具体如图 4-13 与 4-14 所示。本实验首先利用两种不同算法分别对零件进行立体匹配，其次计算两个算法的匹配时间并进行对比，具体如表 4-2 所示。

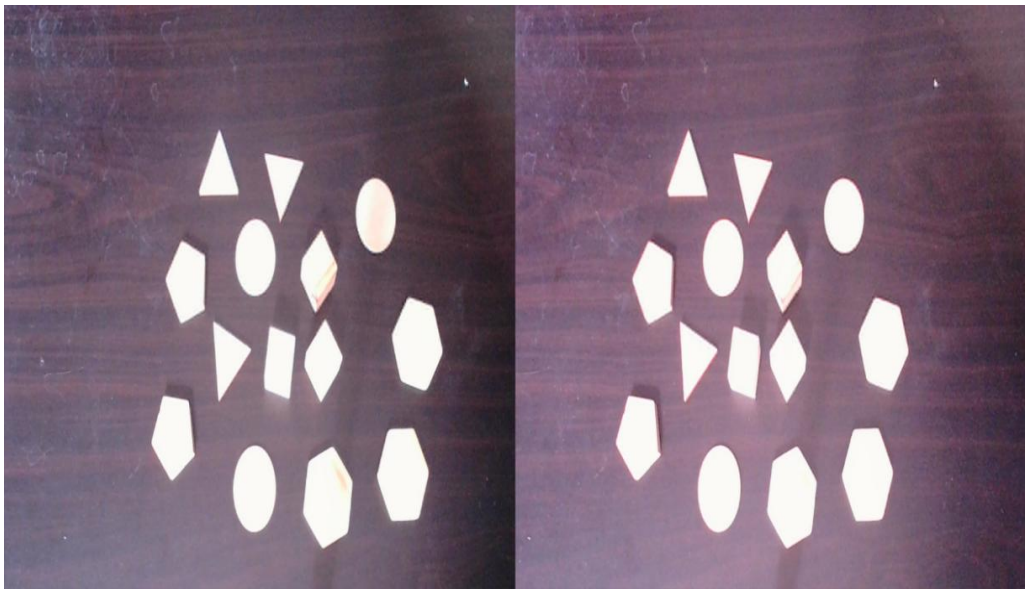


图 4-13 零件散放图

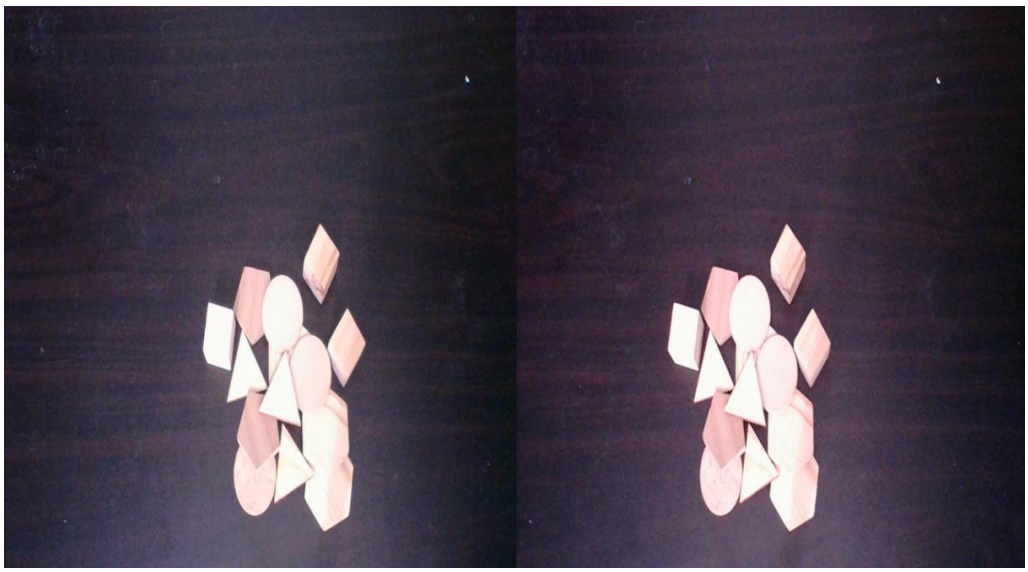


图 4-14 零件堆叠与遮挡图

表 4-2 两种算法匹配时间

立体匹配算法	匹配时间/fps
SGBM	18.09
BM	21.52

由表可知，BM 算法的匹配时间快于 SGBM 算法，同时满码垛需求，因此本文选择对两种算做进行进一步的视差对比实验。具体视差图如图 4-15 与 4-16 所示。

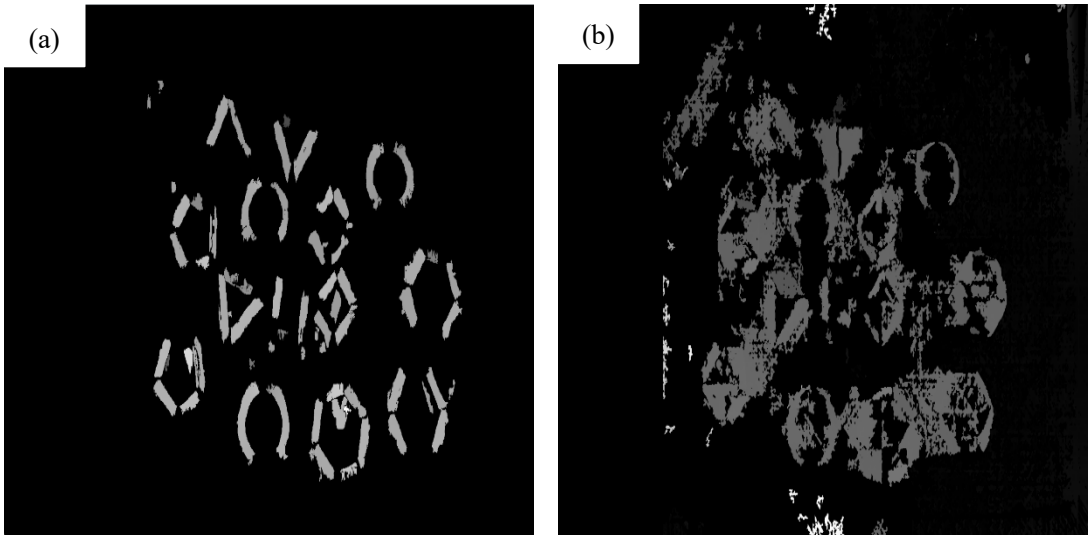


图 4-15 零件散放下视差图：(a) BM 算法视差图；(b) SGBM 算法视差图

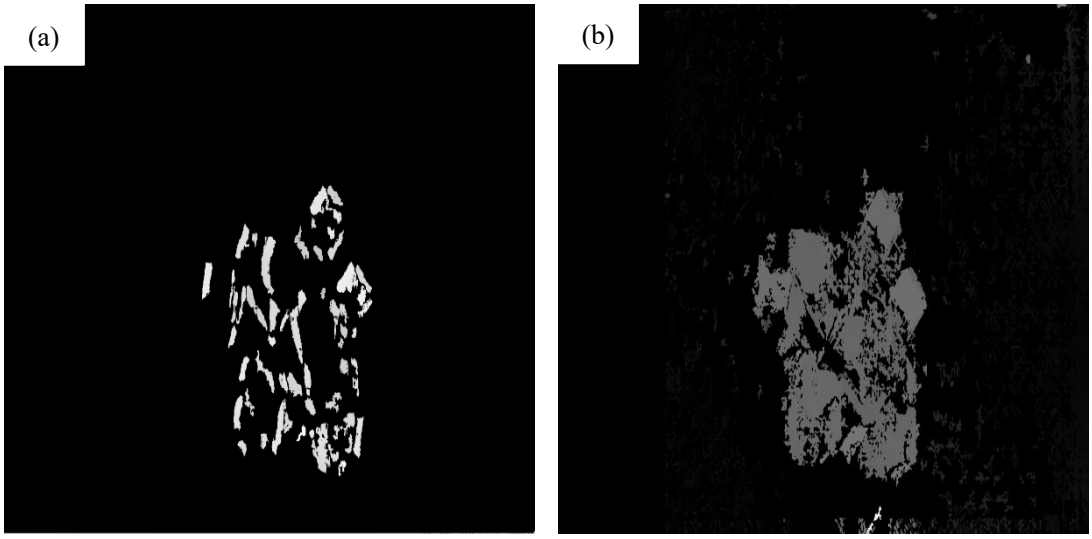


图 4-16 零件堆叠与遮挡下视差图：(a) BM 算法视图差；(b) SGBM 算法视差图

由图 4-15 与 4-16 可知 BM 算法的视差图在散放、堆叠与遮挡的场景下只匹配出了零件的轮廓信息，在许多区域缺乏深度信息的情况下，SGBM 算法生成的视差图不仅清晰地匹配出了零件的轮廓，而且在包含深度信息的区域表现更为丰富。这表明 SGBM 算法生成的视差图中空洞较少，匹配效果优于 BM 算法。两种算法的目标三维坐标如图 4-17 所示，详细的测距距离如表 4-3 所示：

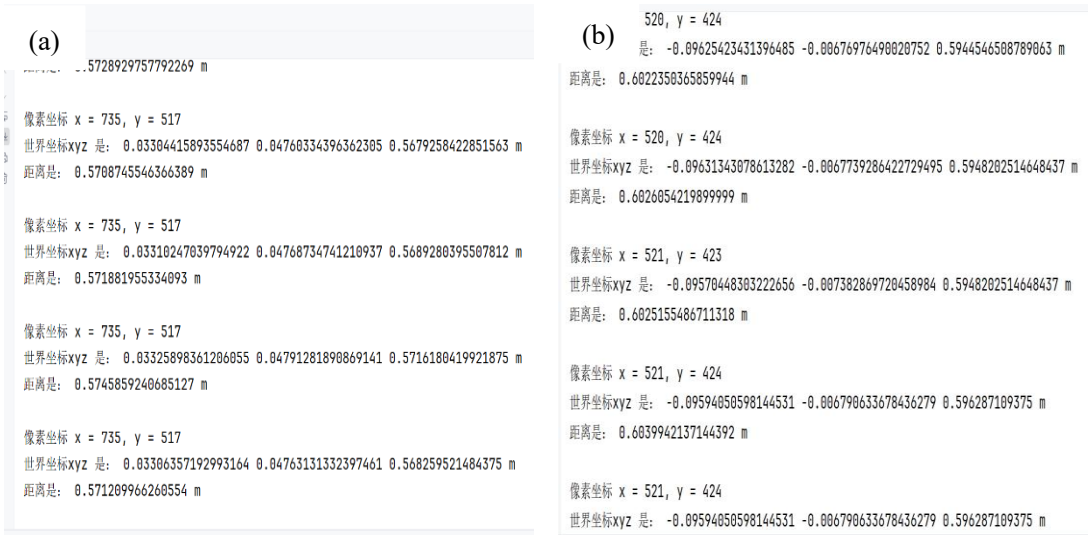


图 4-17 不同算法视差图三维坐标图：(a) BM 算法；(b) SGBM 算法

表 4-3 不同算法测距对比

立体匹配算法	测量距离(m)	实际距离(m)
BM	0.571	0.600
SGBM	0.603	0.600

由图 4.17 与表 4-3 可知，SGBM 算法匹配测距的精度误差比 BM 算法更小，其误差值控制在 3mm 以内。综上所述，虽然 SGBM 算法检测速度低于 BM 算法，但是仍能满足码垛机器人的做业，其检测速度达到了 18.09fps。在同等实验条件下，SGBM 算法的精度要远高于 BM 算法，其能获取更多零件深度信息。所以本文选用 SGBM 立体匹配算法作为零件定位算法。

4.6 本章小结

本章首先阐述了双目立体视觉系统的基本组成，重点介绍了张正友棋盘格标定法的原理及其应用，并深入分析了双目校正的核心原理。随后，利用 MATLAB 软件对本文采用的双目相机进行标定，并将标定所得的相机参数导入 PyCharm 环境中，开展双目校正实验，对比分析了校正前后的图像效果，使双目相机共面且行对准；其次详细介绍了三种常见的立体匹配算法，并对其中的 BM 与 SGBM 算法进行了对比实验，由实验结果表明 SGBM 立体匹配算法匹配的效果更高，所以选用 SGBM 立体匹配算法作为零件的定位方法，利用视差图获取零件的三维坐标信息。

第 5 章 零件识别与定位实验分析

本文第三章与第四章分别完成了零件的识别与定位的算法研究，提高了零件识别的准确性与定位的实时性。本章将首先介绍零件识别与定位算法的实验环境与总体设计方案，详细阐述零件识别与定位方法的实现过程。随后，将针对不同距离与场景下的零件进行一系列实验验证，最后对实验结果进行总结与分析。

5.1 实验环境

零件识别与定位算法基于 Pytorch 为框架，结合 YOLOv8-MSC 目标检测算法与 SGBM 双目立体匹配算法并通过双目视觉相机完成零件的识别与定位。为了提高实验精度与实时性，本文选择租用云服务器来进行实验。其系统环境为 Windows10，显卡为 4090Ti，具体详细实验平台环境如下表 5-1 所示：

表 5-1 实验环境与配置

环境	配置版本
操作环境	Windows10
开发环境	PyCharm
深度学习框架	Pytorch1.11
编程语言	Python3.8
内存	200G
GPU	NVIDIA GeForce RTX 4090
CUDA	11.3
双目相机	汇博视捷

5.2 零件识别与定位系统设计

零件识别与定位系统是由 YOLOv8-MSC 识别检测算法与 SGBM 立体匹配算法共同完成的，其中对于零件的识别是由双目相机的左目完成的，将识别后的零件中心点像素坐标进行记录。定位方面是由 SGBM 双目立体匹配算法完成，其通过双目相机对零件进行立体匹配获取零件的深度信息并结合像素坐标通过坐标系转换获取零件的空间三维坐标，即完成零件的识别与定位，具体流程图如图 5-1 所示：

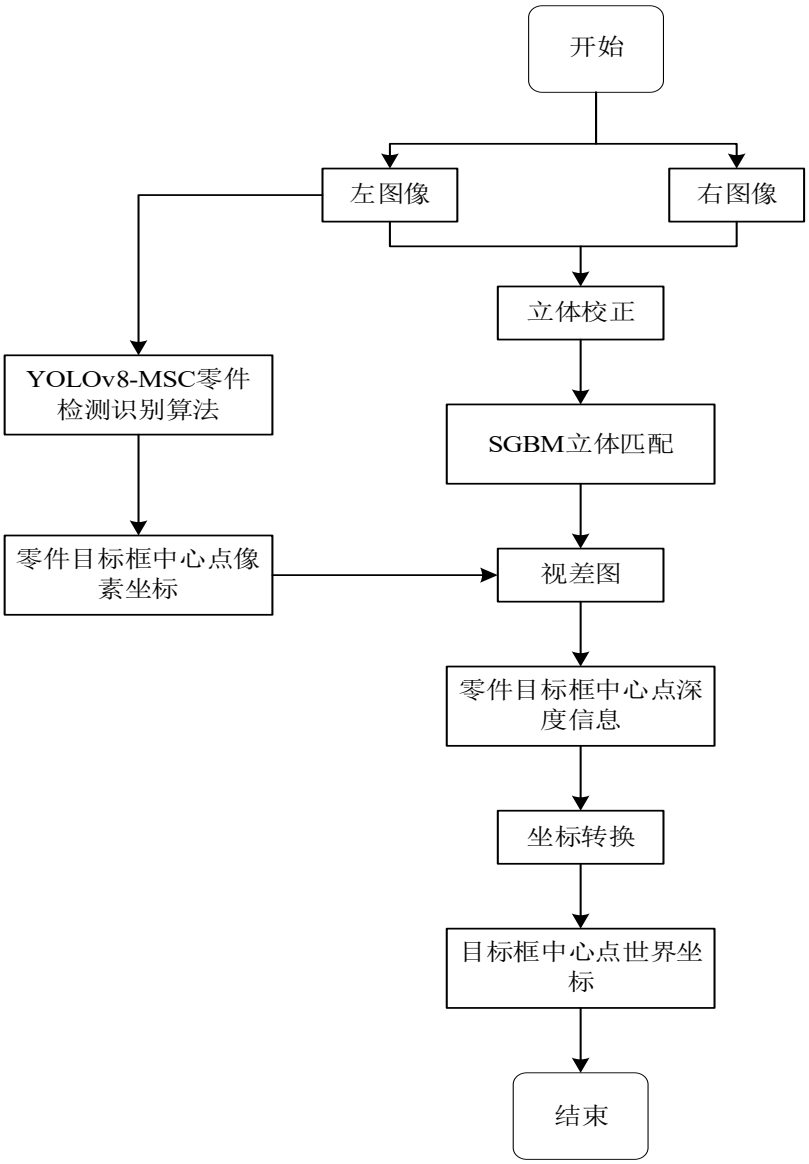


图 5-1 双目立体视觉零件识别与定位系统流程图

零件识别与定位方法的详细过程为：首先输入双目相机拍摄的图片，左目相机使用 YOLOv8-MSC 算法进行零件的检测识别并获取零件目标框中心点像素坐标；其次左右双目相机同时进行双目校正，校正后通过 SGBM 双目立体匹配算法获取零件的视差图和目标框中心点深度信息；最后通过坐标转换获得零件的空间三维坐标信息，完成对零件的识别与定位实验。

5.2.1 零件识别实验

本实验将设计 6 组对比实验，分别在散放、堆叠与遮挡场景下进行识别，距离分别为 0.5m、0.6m、0.7m、0.8m、0.9m、1m 等六个距离下进行识别与定位实验，实验平台如图 5-2 所示，其主要由双目相机，卷尺，支架、导轨等组成，软件层面利用 PyCharm 工具结合 Python 调用 Qt 包，做出双目识别与定位系统界

面，如图 5-3 所示。实验结果如图 5-4 所示。

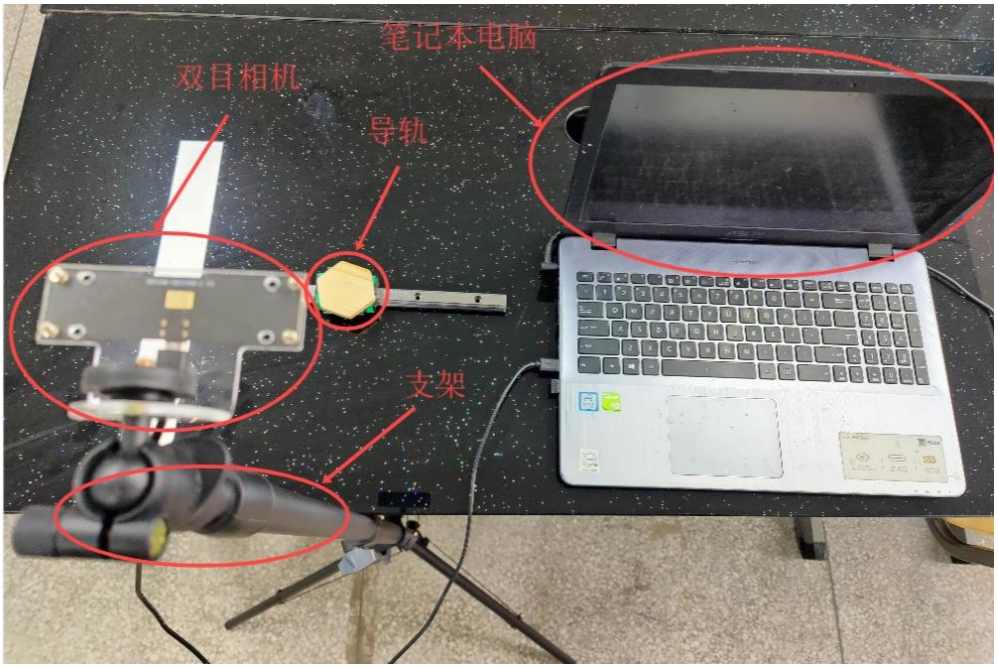
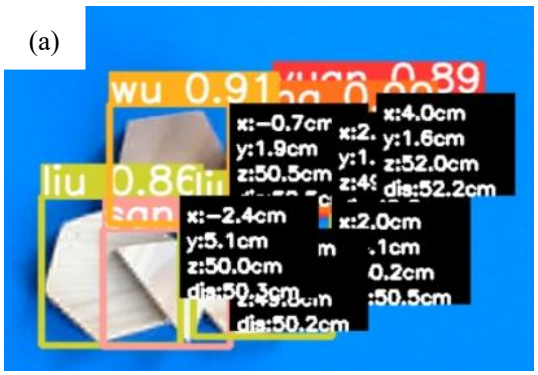
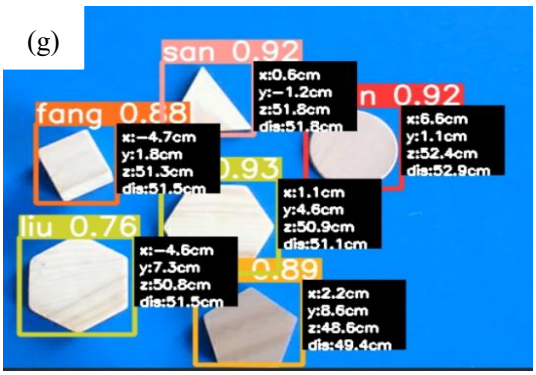


图 5-2 零件识别与定位平台

图 5-3 零件识别与定位系统界面



0.5m

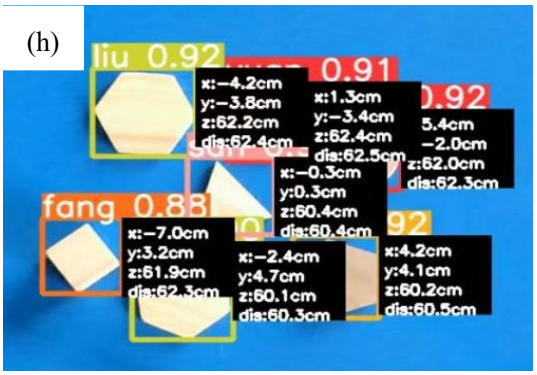


0.5m

(b)

0.6m

(h)

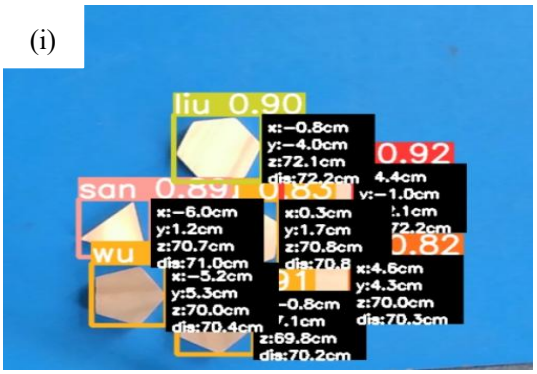


0.6m

(c)

0.7m

(i)

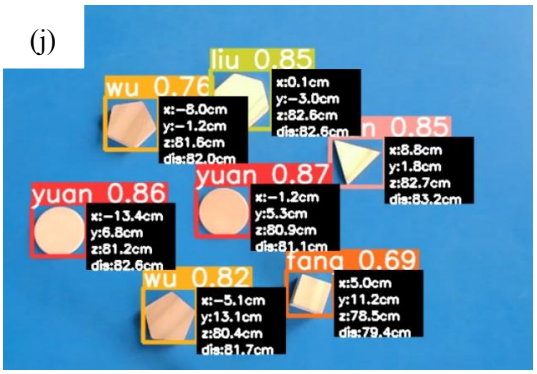


0.7m

(d)

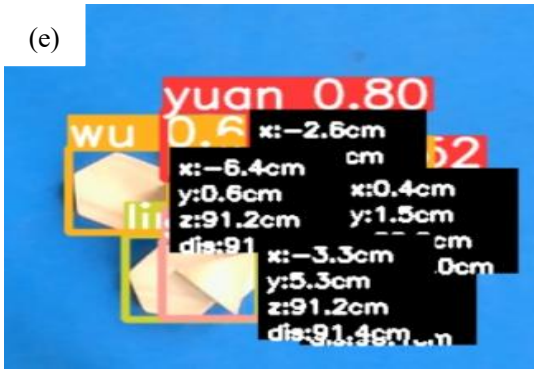
0.8m

(j)



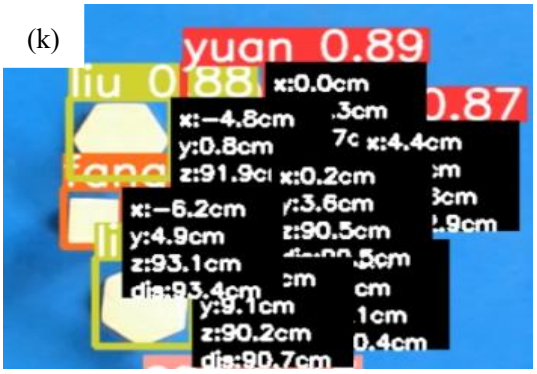
0.8m

(e)

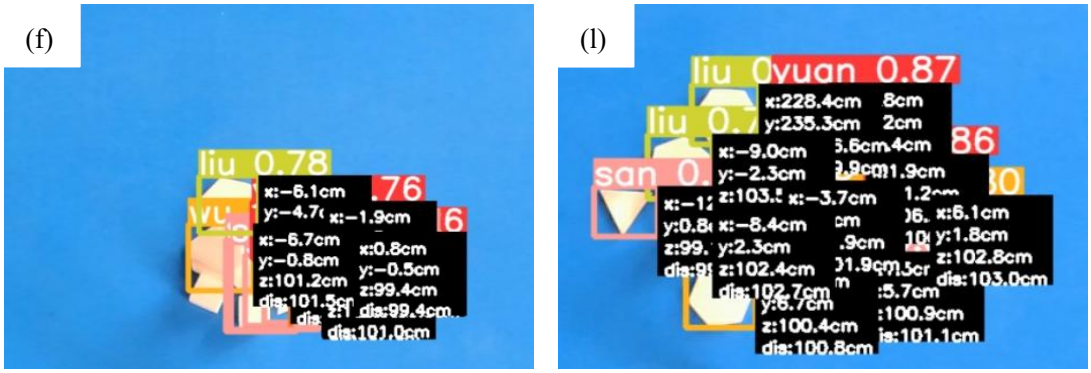


0.9m

(k)



0.9m



1.0m　　1.0m

图 5-4 不同距离与场景下零件识别实验结果图。(a) 0.5m 零件散放；(b) 0.6m 零件散放；(c) 0.7m 零件散放；(d) 0.8m 零件散放；(e) 0.9m 零件散放；(f) 1.0m 零件散放；(g) 0.5m 零件堆叠与遮挡；(h) 0.5m 零件堆叠与遮挡；(i) 0.5m 零件堆叠与遮挡；(j) 0.5m 零件堆叠与遮挡；(k) 0.5m 零件堆叠与遮挡；(l) 0.5m 零件堆叠与遮挡

由图 5-4 可知，在不同距离与场景下，零件识别的精度随着距离的增加其精度有所下降，其平均精度仍达到 85%以上，fps 值稳定在 18 至 21 之间，由此表明本文所采用的零件识别与定位方法具有较好的准确性与实时性，满足码垛机器人面对复杂环境下码垛作业任务。

5.2.2 零件三维空间定位实验

为了验证零件在三维空间定位的准确性，根据图 5-1 流程对零件进行三维空间定位实验来获取零件的空间坐标。由于双目相机在使用 SGBM 立体匹配算法时，无法直接对零件定位的三维空间坐标精度进行验证，因此本文采用相对定距测量方法，对定位坐标的精度进行验证。本次实验以一个零件为实验对象，以双目相机首次拍摄并计算出的初始三维空间坐标作为定位实验的原始坐标，每次平移 3cm 作为定位误差比较。分别对 X、Y、Z 三个坐标轴方向进行定位误差实验，具体实验步骤如下：

(1) 将双目相机固定在支架上，相机平行安装，尽量让双目相机垂直于零件平面。零件通过水平移动进而来进行误差计算。

(2) 每次尽量控制导轨以 3cm 的距离进行水平移动，每次移动后界面检测值都存在一定的数值波动，等波动值趋于稳定之后再记录取值，作为定位检测结果。

(3) 在实验中记录数据时，将相邻两个定位结果之间的差值视为实验误差。该误差的绝对值，即与真实位移值之间的差异，被称为定位误差。此外，相邻误差率定义为相邻位置间的定位误差与实际位移值的比值。

(4) Y、Z 轴两个方向上的定位坐标实验重复以上 1、2、3 三个步骤，并记录

数值。按照上述实验步骤进行零件的定位实验，图 5-5 为 Qt 界面显示的零件三维坐标。定位结果如图所示，表 5-2 是沿 X 轴方向上的定位结果及定位误差，在实验时保持 Y，Z 两轴方向上的坐标不变，零件每次平移 3cm，共统计 9 组数据进行实验误差分析，表 5-3、5-4 分别为 Y 轴，与 Z 轴两个方向上的定位结果。

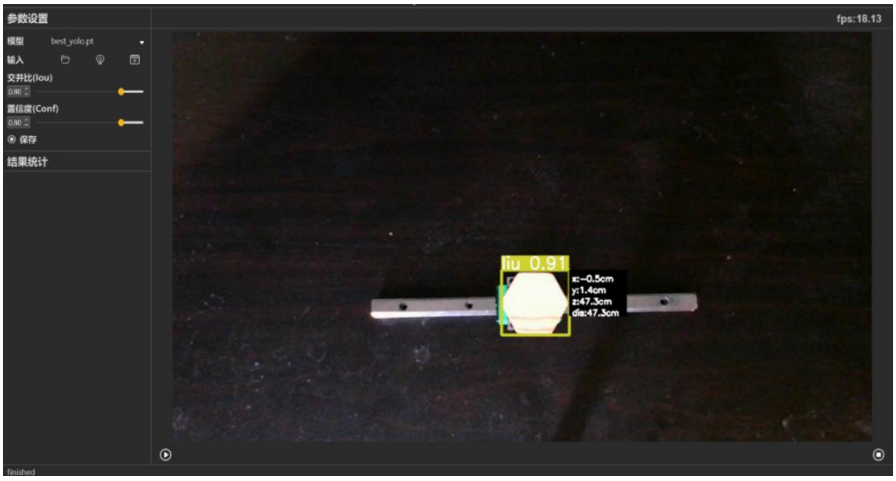


图 5-5 零件三位空间定位图

表 5-2 零件的 X 轴方向上的定位误差

序号	三维空间坐标			相邻位置定位差/cm	相邻位置定位误差/cm
	X/cm	Y/cm	Z/cm		
0	-8.8	2.6	42.9	-	-
1	-5.9	2.5	41.5	2.9	0.1
2	-2.8	2.5	41.5	3.1	0.1
3	0.1	2.4	42.9	2.9	0.2
4	3.3	2.4	42.6	2.6	0.2
5	6.7	2.5	43.7	2.7	0.4
6	9.6	2.5	44.2	2.8	0.1
7	12.7	2.6	44.2	3.1	0.1
8	15.6	2.6	42.1	2.9	0.1

表 5-3 零件的 Y 轴方向上的定位误差

序号	三维空间坐标			相邻位置定位差/cm	相邻位置定位误差/cm
	X/cm	Y/cm	Z/cm		
0	-2.1	6.2	42.5	-	-
1	-2.1	3.2	43.9	3.0	0
2	-2.2	0.3	44.9	2.9	0.1
3	-2.3	-3.2	44.2	3.3	0.1
4	-2.3	-5.4	44.7	2.6	0.2
5	-2.3	-8.2	43.6	2.6	0.2
6	-2.2	-11.3	43.6	3.1	0.1
7	-2.1	-14.1	45.1	3.1	0.2
8	-2.1	-17.3	45.2	3.2	0.2

表 5-4 零件的 Z 轴方向上的定位误差

序号	三维空间坐标			相邻位置定位差/cm	相邻位置定位误差/cm
	x/cm	Y/cm	Z/cm		
0	2.4	1.6	51.2	-	-
1	2.3	1.4	48.3	2.9	0.1
2	2.3	1.1	45.3	3.0	0
3	2.1	1.3	42.5	2.8	0.2
4	2.0	1.1	39.6	2.9	0.1
5	2.0	0.8	36.7	2.9	0.1
6	1.5	0.6	33.5	3.4	0.2
7	1.7	0.3	30.4	2.9	0.1
8	2.1	-0.3	27.2	3.2	0.2

以上数值均由卷尺测量得出，由以上表可知 X、Y、Z 三个坐标轴方向定位误差结果，为了更加直观的显示出定位实验结果，对以上三个坐标轴方向上误差率绘制误差曲线图，如图 5-6、5-7、5-8 所示：

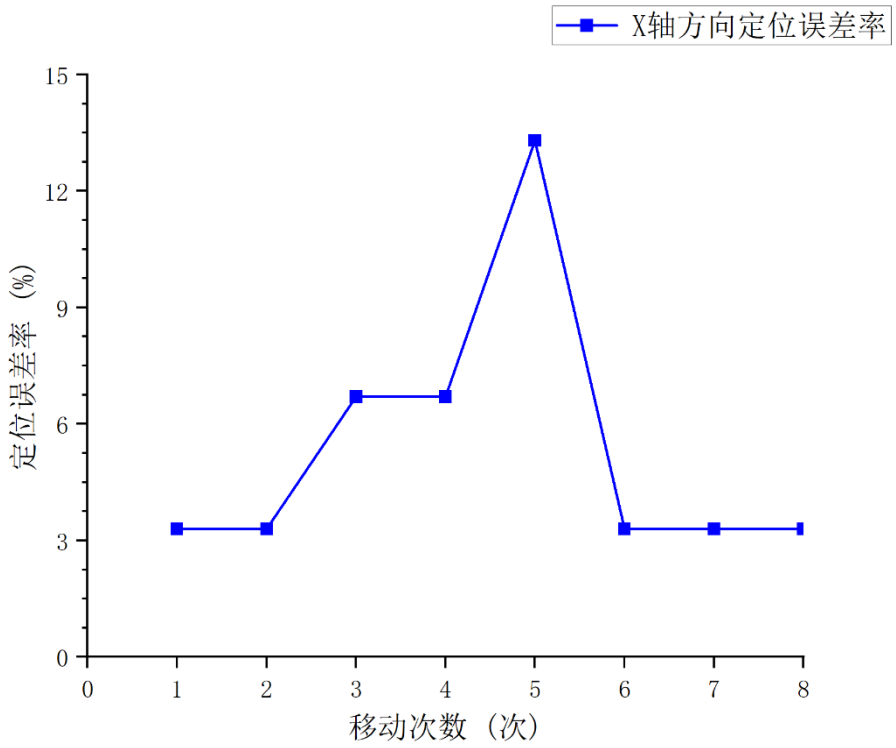


图 5-6 X 轴定位误差率

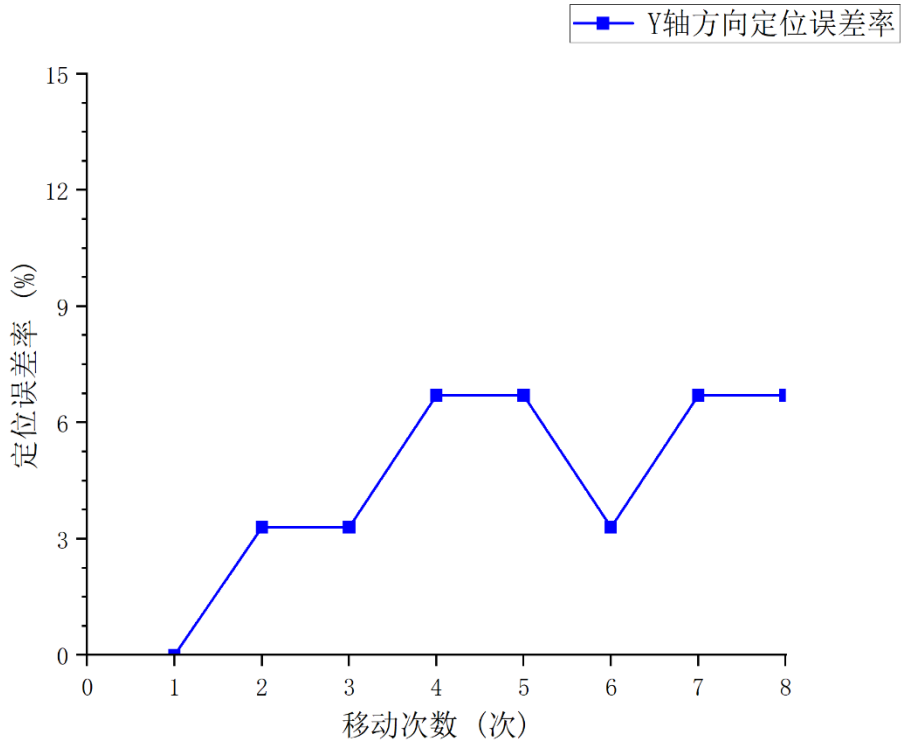


图 5-7 Y 轴定位误差率

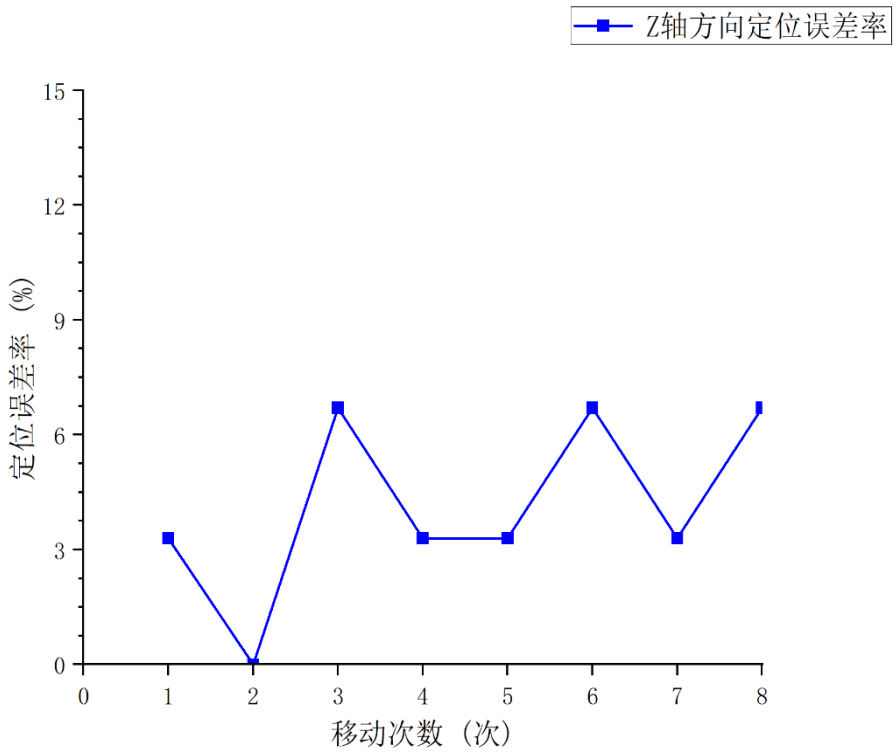


图 5-8 Z 轴定位误差率

由上述实验数据可知，相邻位置出现个别误差会较大数值。根据实验分析，通常是由外界环境因数导致的实验结果不准确。虽然结果如此，本文的零件定位方法仍能保持一定的稳定性。由表格计算实验结果可知 X、Y、Z 三轴的平均定

位误差为 1.6mm、1.4mm、1.3mm，相邻位置误差值基本都误差合理范围之内，符合定位精度要求。由图 5-6、5-7 以及 5-8 可知其 X、Y、Z 三轴的平均定位误差率为 5.4%、4.6%与 4.2%。由此说明本文的定位方法能准确的定位零件，满足码垛机器人所需要的定位精度。

5.2.3 定位误差原因分析

由于双目相机的相关坐标系本质上是一种数学模型，其不可见性导致其在实际操作实验中很难进行控制，因此影响定位误差的主要因素包括以下几点：

(1) 零件通过手动平移且通过平移距离也是通过手动卷尺进行测量，每次平移和测量都存在相应的误差，这些误差会导致定位实验结果出现偏差。

(2) 实验平台与相机安装位置可能并没有达到完全水平状况，两者之间的夹角也会对定位误差产生影响。

(3) 双目相机自身的相关参数，如相机的畸变、标定精度等因素，均会对定位误差产生影响。

(4) 实验现场环境中的各种干扰因素，如光线变化、噪声以及背景复杂性等，都会对零件在空间坐标系中的定位产生影响。

5.3 本章小结

本章将 YOLOv8-MSC 零件检测识别算法与 SGBM 立体匹配算法相结合，搭建了双目视觉识别与定位系统平台。首先介绍了零件识别与定位的实验环境与总体方案设计；其次验证了不同距离与场景下零件识别与定位方法，实验结果表明该方法在面对散放、堆叠与遮挡和不同距离的场景下，具有良好的实时性与准确性，满足码垛机器人的码垛作业任务。

第 6 章 总结与展望

6.1 总结

本文提出并设计一种零件识别与定位方法，并将该方法应用在码垛机器人作业中。作为机器人视觉模块信息，其中包括设计了零件识别与定位算法的总体方案、以及采集零件木块的数据集、零件识别算法模型训练、双目标定与校正、零件的三维空间定位等内容。本文主要的研究工作与成果如下：

(1) 提出了一种基于改进 YOLOv8 零件识别算法。针对复杂场景下，存在着零件的漏检、误检以及检测慢等问题，本文首先采用 MobileNetv3 网络结构替换 YOLOv8 的主干网络以加速模型的检测速度；其次在主干网络中引入 CBAM 注意力机制以提升模型对零件特征的提取；最后采用 SIoU 损失函数作为模型的损失函数，优化模型的训练过程并提高收敛效率。将改进后的算法命名为 YOLOv8-MSC 算法。实验结果表明优化改进后的模型，在检测精度与速度均有所提升。其 mAP 值达到了 98.7%，参数量减少到 2.8G，检测速度减少至 6.11ms，更适合部署在小型移动的设备上。

(2) 零件视觉定位。首先对双目相机进行标定，以获取其内部和外部参数，随后基于这些参数对双目相机进行校正，保证双目相机共线且行对准；其次对双目立体匹配算法进行理论介绍，包括匹配基元、约束准则与立体匹配流程。经实验分析，局部立体匹配算法(BM)的效率高、匹配速度快，但其匹配精度低、性能不稳定，不能保证匹配的准确性。因此本文选半局部立体匹配算法(SGBM)作为零件的定位匹配算法，其具有匹配精度高、性能稳定等优点，虽然匹配速度有所下降，但仍达到了 18.09fps，符合视觉定位要求。

(3) 零件识别与定位方法的系统设计。将改进后的 YOLOv8-MSC 算法与 SGBM 双目立体匹配算法相结合，利用 Qt 做出系统界面。使用该方法在不同距离与场景下进行零件识别与定位实验，并分析了定位误差原因，由实验结果表明，该方法能够达到码垛机器人对零件的抓取作业任务。

6.2 展望

(1) 在双目相机标定方面。本文采用张正友标定法，从获取的双目参数可知，

相机的标定仍存在相应的误差，可以改进相机标定方法从而减小误差，使双目相机拍摄的图像更加精准。

(2) 在零件的目标识别方面，虽然本文对零件检测识别算法做出了一些改进，但对于一些损坏的零件与其它障碍物并未进行考虑，下一步需要添加相应的数据科技，扩大检测识别范围，以实现对零件更加准确的识别。

(3) 在零件的空间三维定位方面。本文采用半目立体匹配算法(SGBM)，来实现零件的定位任务。然而半目立体匹配算法立体匹配算法极易受到外部环境因素的干扰，例如光照变化和噪声等，因此后续研究可以针对半目立体匹配算法做出相应的改进，增强模型的鲁棒性，从而提高零件的定位准确率。

参考文献

[1] 冯枫. 工业机器人自动化生产技术的实践研究[J]. 山西电子技术, 2025, (01): 106-109.

[2] 黄文玉. 码垛机器人智能码垛解决方案行业市场发展现状及前景趋势全景调研与战略咨询报告[R]. 四川: 中研普华产业研究院, 2024

[3] 王绍存, 毕玉强. 码垛机器人的现状与发展综述[J]. 锻压装备与制造技术, 2024, 59(05): 98-101.

[4] 谈剑峰, 罗卢洋, 李德衡, 等. 基于机器视觉的码垛工件自动识别和定位技术研究[J]. 家电维修,2024, (06): 56-58.

[5] 梁继海. 不规则零件的数控机床加工工艺研究[J]. 现代制造技术与装备, 2024, 60(07): 173-175.

[6] Khazaei N B, Hashjin T T, Ghassemian H, et al. Application of machine vision in modeling of grape drying process[J]. Journal of Agricultural Science & Technology, 2018, 15(6): 1095-1106.

[7] 郭北涛, 刘瀚齐, 张丽秀. 基于双目视觉的堆叠零件识别与定位研究[J]. 组合机床与自动化加工技术, 2024, (08): 43-47.

[8] 王龙. 基于深度学习的工业零件识别方法研究[D]. 兰州理工大学, 2023.

[9] 李辉, 施卫保, 虎鹏. 智能识别分拣码垛工业机器人工作站系统的设计与应用[J]. 机械工程师, 2019, (12): 110-112.

[10] 郭忠峰, 王健鹏, 杨钧麟, 等. 基于改进 YOLOv8-Pose 的码垛快速识别与抓取点检测[J]. 组合机床与自动化加工技术, 2024, (11): 125-129.

[11] 刘玮. 基于 3D 点云的盘类元件识别与定位技术研究[D]. 长春工业大学, 2022.

[12] 麻方达. 基于深度学习的热流道零件识别与抓取研究[D]. 青岛大学, 2023.

[13] 刘承逸. 基于 YOLOv8 的加工零件识别方法研究[D]. 南昌大学, 2024.

[14] 许鑫杰, 王秀锋, 鲁文其, 等. 基于边缘检测的零件轮廓识别系统开发[J]. 机电工程, 2019, 36(02): 201-205.

[15] 孙小权, 邹丽英. 基于 SVM 的图像识别在零件分拣系统中的应用[J]. 机电工程, 2018, 35(12): 1353-1356.

[16] 方祝平, 张博, 王曦, 等. 应用于机器装配的多目标视觉识别与分类系统[J]. 制造业自动化, 2021, 43(07): 24-28.

[17] You F C, Zhang Y B. A mechanical part sorting system based on computer vision[C]//2008 International Conference on Computer Science and Software

Engineering. IEEE, 2008: 860-876.

[18] 毛向向, 王红军.薄壁零件复杂光照情况下的轮廓特征识别[J]. 电子测量与仪器学报, 2021, 35(03): 137-143.

[19] Lee J, Lee W. A new recognition and pose estimation of tiny electronic parts using principal component analysis and harris corner features[J]. Journal of Electrical Engineering & Technology, 2020, 15(1) :43-51.

[20] 司小婷. 基于视觉的零件特征识别与分类方法研究与实现[D]. 中国科学院研究生院(沈阳计算技术研究所), 2016.

[21] Xin D U. A mechanical parts inspection model based on image classification technology[J]. Academic Journal of Manufacturing Engineering, 2016, 14(3): 14-21.

[22] Tao Y, Jun Z, Zhi-hao Z, et al. Fault detection of train mechanical parts using multi-mode aggregation feature enhanced convolution neural network[J]. International Journal of Machine Learning and Cybernetics, 2022, 13(6): 1781-1794.

[23] 宋栓军, 侯中原, 王启宇, 等. 改进 YOLOv3 算法在零件识别中的应用[J]. 机械科学与技术, 2022, 41(10): 1608-1614.

[24] 郭斐, 靳伍银, 王猛. 基于改进的 Faster R-CNN 算法的机械零件图像识别[J]. 机械设计, 2019, 36(09): 113-116.

[25] 余永维, 韩鑫, 杜柳青. 基于 Inception-SSD 算法的零件识别[J]. 光学精密工程, 2020, 28(08): 1799-1809.

[26] 韩强. 面向小目标检测的改进 YOLOv8 算法研究[D]. 吉林大学, 2023.

[27] 曾毅星. 面向增强现实应用的轻量级零件识别算法研究[D]. 电子科技大学, 2021.

[28] 徐维. 基于改进 YOLOv4 算法的工业机器人零件识别与抓取研究[D]. 重庆理工大学, 2022.

[29] 龚友彬, 康曦, 王军. 基于双目视觉的工件位姿检测[J]. 信息技术与信息化, 2024, (11): 141-144.

[30] Wang Z B, MA X, MU C Y, et al. Research on adhesive workpiece recognition and positioning based on machine vision[C]. Control and RoboticsEngineering, 2019: 135-140.

[31] Saleem N H, Klette R. Accuracy of free-space detection: monocular versusbinocular vision[C]//2016 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2016: 1128-1136.

[32] 曹文君. 全景视觉环境避障测距方法研究[D]. 华南农业大学, 2016.

[33] 李佳. 基于车载双目相机的行人测距技术研究[D]. 沈阳工业大学, 2021.

[34] 冯英旺. 双目被动测距系统设计及算法研究[D]. 南京理工大学, 2019.
[35] 毛先胤, 吕黔苏, 马晓红, 等. 基于双目视觉的电力线巡检机器人障碍物定位测距[J]. 自动化与仪器仪表, 2021(9): 244-248.
[36] R-FCN D J. Object detection via region-based fully convolutional networks[C]// Proceedings of IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 67-72.
[37] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems, 2017,39(6): 1137-1149
[38] Liu W, Angelov D, Erhan D, et al. SSd: Single shot multibox detector[C]// /European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
[39] 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.
[40] 李科, 毋涛, 刘青青. 基于深度图与改进 Canny 算法的人体轮廓提取[J]. 计算机技术与发展, 2021, 31(05): 67-72.
[41] 刘阳阳. 无人机双目视觉目标检测与测距方法研究[D]. 重庆大学, 2019.
[42] 杨伟民, 杨光熙, 孙双. 摄像机标定技术在油田监控设备中的应用[J]. 信息系统工程, 2020(05): 90-104.
[43] 廖秉旺. 基于机器视觉的工件分拣系统研究与设计[D]. 广东工业大学, 2021.
[44] Zhang R, Ma C Y, Liu J J, et al. Stereo imaging camera model for 3D shape reconstruction of complex crystals and estimation of facet growth kinetics[J]. Chemical Engineering Science. 2017, 160(8): 171-182.
[45] 张竞艺. 基于双目视觉的运动目标实时追踪与测距[D]. 南京邮电大学, 2020.
[46] 颜坤. 基于双目视觉的空间非合作目标姿态测量技术研究[D]. 中国科学院大学(中国科学院光电技术研究所), 2018.
[47] 陈思涵, 刘勇, 何祥. 基于改进 YOLOv8 的工厂行人检测算法[J]. 现代电子技术, 2024, 7(09): 1-7
[48] Wang W, Xie E, Song X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8440-8452.
[49] Li X, Wang W, Wu L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems, 2020,17(63): 21002-21012.
[50] Qiang B H, Zhai Y J, Zhou M L, et al. SqueezeNet and Fusion Network-Based

Accurate Fast Fully Convolutional Network for Hand Detection and Gesture Recognition[J]. IEEE Access, 2021, 9: 77661-77674.

[51] Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.

[52] Qin Y Y, Cao J T, JI X F. Fire detection method based on depthwise separable convolution and YOLOv3[J]. International Journal of Automation and Computing, 2021, 18 (2): 300-310.

[53] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[54] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019:1320-1326.

[55] Hu J, Shen L, Sun G. Squeeze-and-excitation net-works[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[56] Fuchs F, Worral D, Fischer V, et al. Se(3)-transformers: 3d roto-translation equivariant attention networks[J]. Advances in Neural Information Processing Systems, 2020, 33: 1970-1981.

[57] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

[58] 尚潘.基于改进 YOLOv8n 的无人机航拍图像目标检测[J].无线互联科技, 2025, 22(04): 84-87.

[59] 贾志洋, 许兆, 冷艳梅, 等. 基于并行优化 CBAM 的轻量级故障诊断模型[J]. 应用科学学报, 2025, 43(01): 94-109.

[60] Liu Y, Shao Z, Hao N, et al. Global attention mechanism: Retain information to enhance channel-spatial interactions. [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2023: 6-20.

[61] Rezatofighi H, Tsoi N, Gwak J Y ,et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:658-666.

[62] Wang S Y, Qin Z, Gao L Y. Multi-spatial pyramid feature and optimizing focal loss function for object detection[J]. IEEE Transactions on Intelligent Vehicles, 2023, 9(1): 1054-1065.

[63] Yu Q F, Han Y D, Han Y, et al. Enhancing YOLOv5 Performance for Small-Scale Corrosion Detection in Coastal Environments Using IoU-Based Loss Functions[J]. Journal of Marine Science and Engineering. 2024,12: 16-18

[64] Ma N, Zhang X, Zheng H T, et al. Shufflenetv2: Practical guidelines for efficient cnn architecture design[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 116-131.

[65] Li Y, Wu Y H, Sun G, et al. Vision transformers with hierarchical attention[J]. Machine Intelligence Research, 2024, 21: 1-14

[66] 朱凌涛. 基于双目视觉的钢轨紧固件中心定位系统[D]. 湖南大学,2019.

[67] Zhang Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.

[68] Tsai R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses[J]. IEEE Journal on Robotics and Automation, 1987, 3(4): 323-344.

[69] Setyawan R A, unoko R, Choiron M A, et al. Implementation of stereo vision semi-global block matching methods for distance measurement[J]. Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 2018, 12(2): 585-591.

[70] Kanade T, Okutomi M. A stereo matching algorithm with an adaptive window: Theory and experiment[J]. IEEE transactions on pattern analysis and machine intelligence, 1994, 16(9): 920-932.

[71] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1222-1239.

[72] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1222-1239.

[73] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]//International Conference on Machine Learning. PMLR, 2015: 448-456.

[74] 汪小愉. 基于双目视觉的图像深度信息数据集的研究[D]. 杭州电子科技大学, 2018.

[75] 陈莉. 基于机器视觉的双摄像头测距系统的研究与实现[D]. 河北科技大学,2019.

[76] Hirschmuller H. Accurate and efficient stereo processing by semi-global matching and mutual information[C]//2005 IEEE Computer Society Conference on

Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005: 807-814.
[77] 陈秋梦. 基于双目视觉的目标检测与测距系统[D]. 东北石油大学, 2022.
[78] 严邓涛. 基于卷积神经网络的立体匹配算法研究[D]. 南京邮电大学, 2019.

攻读硕士期间发表学术论文和成果情况

王宜红,夏敏胜,疏达,吴路路. 基于改进 YOLOv8 无序码垛识别研究[J]. 湖南工程学院学报, 2024. (一作，已录用)

致谢

时光荏苒，我的研究生求学生涯即将结束。回首这段旅程，有艰辛，有收获，更有无数值得铭记的温暖与感动。在此，我谨向所有给予我帮助、支持和鼓励的老师、同学、家人和朋友致以最诚挚的谢意。

首先，我要深深感谢我的导师疏达教授。从论文选题、研究方法到最终定稿，您始终以渊博的学识、严谨的治学态度和耐心的指导引领我前行。您不仅教会我如何做学问，更以自身的学术品格和人格魅力影响着我，让我明白何为真正的学者风范；其次，非常感谢吴路路老师、在论文修改过程中提出的宝贵建议，使我的研究更加完善；同时要感谢课题组的王建平老师、王建彬老师、王安恒老师和王风涛老师，是他们在每次开组会时都会给予我很大的班助。

其次，还要对课题组的师兄李维汉、赵垒、汤勇、刘恩智、吴文专；同门李永昆和刘少辉；师弟张彬、杨强、李文瑞、殷硕、吴郁、罗红宁、王赟、夏敏胜表示真诚的感谢。难忘我们一起讨论学术问题、互相修改论文的日日夜夜，也难忘在科研遇到瓶颈时你们的鼓励与帮助。这段并肩奋斗的时光，将成为我研究生生涯最珍贵的回忆。

最后，特别感谢我的家人。父母数十年如一日的默默付出，是我坚持求学的最大动力，你们的理解与支持，让我在迷茫时总能找到前行的勇气。

研究生生涯的结束，亦是人生新篇章的开始。我将带着这段经历赋予我的知识、能力和感恩之心，继续探索、成长，不负所有关心与帮助过我的人。

尚德教学 唯实惟新