电 子 科 技 大 学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE

论文题目 **基于双目立体视觉的障碍物检测与定位技术**

专业学位类别	电子信息
学　　号	202152060120
作者姓名	石 功 林
指导教师	童 玲　　教 授
学　　院	自动化工程学院

分类号 TP391 密级 公开

UDC [注1] 004.4

学 位 论 文

基于双目立体视觉的障碍物检测与定位技术

（题名和副题名）

石 功 林

（作者姓名）

指导教师 童 玲 教 授

电子科技大学 成 都

（姓名、职称、单位名称）

申请学位级别 硕士 学科专业 电子信息

提交论文日期 2024 年 3 月 11 日 论文答辩日期 2024 年 5 月 6 日

学位授予单位和日期 电子科技大学 2024 年 6 月

答辩委员会主席

评阅人

注 1：注明《国际十进分类法 UDC》的类号。

Obstacle Detection and Localization Technology Based on Binocular Stereo Vision

A Master Thesis Submitted to
University of Electronic Science and Technology of China

Discipline	**Electronic Information**
Student ID	**202152060120**
Author	**Gonglin Shi**
Supervisor	**Prof. Ling Tong**
School	**School of Automation Engineering**

摘 要

随着人工智能和机器视觉技术的迅速发展，目标检测和双目立体技术在机器人领域的应用越来越广泛。然而，对于不断变化的外部环境信息，实现高精度的目标检测与定位始终是双目立体视觉领域的研究热点和挑战。这些挑战主要包括如何在日常环境中对小目标进行精确的检测，以及如何处理由于环境光照、纹理变化等因素导致的定位精度下降的问题。这些问题的解决在机器人智能化以及无人机导航等领域都具有重要意义。因此，本文将深入研究基于双目立体视觉的障碍物检测与定位技术，主要研究内容概括如下：

（1）本文提出一种基于注意力机制模块（CBAM）与多尺度特征检测模块相结合的方法实现对 YOLOv5 模型的改进，以增强 YOLOv5 算法的检测精度，改善对远距离小目标的识别效率，并极大的降低了漏检和误检的情况。最终实验结果表明，改进后的 YOLOv5 网络模型总体精确度提升 4.9%、召回率降低了 0.26%、mAP 提升 1.55%。

（2）针对 SGBM 算法对弱纹理，光照变化敏感而导致对于物体定位与测距精度不高，误差较大的问题。本文结合直方图均衡化、多尺度变换和滤波处理来优化 SGBM 算法，并利用改进后的 SGBM 算法对双目图像进行匹配，生成视差图，为双目定位和测距提供更加准确的结果。最终实验结果表明，本文所提优化方法可以显著改善 SGBM 算法的鲁棒性，并提高定位和测距的精度。

（3）本文采用了两段式标定分析法，用于解决针双目摄像机标定过程细节繁琐，标定参数不易确定的问题。该方法首先使用张正友标定法中的三维视图进行初步参数评估，然后通过分析内外参数的物理意义并与已知相机参数比较，进一步提高了标定参数的准确性和标定效率。

（4）本文设计并开发基于 YOLO 算法与双目定位算法相结合的原型系统。该系统包括图像基本的操作功能、图像预处理模块、障碍物检测模块、非实时物体定位测距模块以及实时物体定位测距模块，实现了方便快捷的障碍物的检测与定位。

关键词：目标检测，YOLOv5，SGBM，摄像机标定，双目定位

ABSTRACT

With the rapid development of artificial intelligence and machine vision technologies, the application of object detection and stereo vision technologies in the field of robotics is becoming increasingly widespread. However, achieving high-precision object detection and positioning amidst constantly changing external environmental information remains a research hotspot and challenge in the field of stereo vision. These challenges primarily include how to accurately detect small targets in everyday environments, and how to handle the decline in positioning accuracy due to factors such as environmental illumination and texture changes. The resolution of these issues holds significant implications in fields like intelligent robotics and drone navigation. Therefore, thesis will delve into the obstacle detection and positioning technology based on stereo vision. The main research content can be summarized as follows:

(1) Thesis proposes a method that combines the Convolutional Block Attention Module (CBAM) and multi-scale feature detection module to improve the YOLOv5 model, enhancing the detection accuracy of the YOLOv5 algorithm, improving the recognition efficiency of remote small targets, and greatly reducing the occurrences of missed detection and false detection. The final experimental results show that the overall accuracy of the improved YOLOv5 network model increased by 4.9%, the recall rate decreased by 0.26%, and the mAP increased by 1.55%.

(2) In response to the problem that the Semi-Global Block Matching (SGBM) algorithm is sensitive to weak textures and changes in illumination, leading to low accuracy and large errors in object positioning and ranging, thesis optimizes the SGBM algorithm by combining histogram equalization, multi-scale transformation, and filtering. The improved SGBM algorithm is used to match stereo images, generate disparity maps, and provide more accurate results for stereo positioning and ranging. The final experimental results show that the optimization method proposed in thesis can significantly improve the robustness of the SGBM algorithm and improve the accuracy of positioning and ranging.

(3) Thesis adopts a two-stage calibration analysis method to solve the problem of cumbersome details in the calibration process of binocular cameras and the difficulty in determining calibration parameters. This method first uses the three-dimensional view in

Zhang's calibration method for preliminary parameter evaluation, and then further improves the accuracy and efficiency of calibration parameters by analyzing the physical meaning of internal and external parameters and comparing them with known camera parameters.

(4) Thesis designs and develops a prototype system that combines the YOLO algorithm with the binocular positioning algorithm. The system includes basic image operation functions, image preprocessing modules, obstacle detection modules, non-real-time object positioning and ranging modules, and real-time object positioning and ranging modules, achieving convenient and quick obstacle detection and positioning.

Keywords: Object Detection, YOLOv5, SGBM Algorithm, Camera Calibration, Binocular Positioning

目　录

第一章 绪论

1.1 研究背景与意义

随着硬件技术和人工智能的飞速发展，机器人正在变得越来越智能化，它们开始在多个领域扮演关键角色，并逐步取代执行重复性高、劳动强度大甚至危险的工作[1]。机器人技术是一种模拟、增强或扩展人类能力的技术，包括硬件设计、软件开发和人工智能等众多领域。在这些领域中，目标检测与定位技术占据着重要地位，它们是实现机器人机械臂自主避障和拾取特定障碍物的基础。因此，一个机器人是否具备高效的目标检测与定位能力，已成为衡量其智能化水平的关键指标之一。

目标定位与测距技术是现代科技中不可或缺的一部分，它在机器人机械臂运动等领域中扮演着重要角色。这些技术的实现，需要使用传感器，从而感知并理解周围的世界，其中常用设备有视觉传感器和非视觉传感器[2]，如激光雷达、超声波传感器等属于非视觉设备，而结构光相机、双目立体相机等属于视觉设备这一类。然而，激光雷达和超声波等传感器虽然定位测距精度高，但价格昂贵，难以普及。相比之下，在光照良好、需要高精度深度信息的场景中，双目立体相机可能更为合适，并且因其低成本，小体积的优点广泛应用于目标的检测与定位任务中。

双目视觉系统通过成像设备从不同的位置获取被测物体的两幅图像，然后计算图像对应点间的视差，即两幅图像中同一物体的位置差异。这种视差信息可以被转化为物体的深度和距离信息，从而获取物体的三维信息[3]。双目立体视觉的系统结构简单，只需要两个相机和一些图像处理算法，相机的制造成本也相对较低，并且精度较高[4]。在实际场景中，障碍物作为移动机器人的检测目标之一，对障碍物精准高效的检测，以及在路径规划期间对障碍物的位置分析跟踪，不仅能及时预测潜在隐患，也能有效保证机器人避障工作顺利完成。

基于双目立体视觉的障碍物检测与定位技术是一种重要的机器视觉应用。这种技术使用双目相机，从两个不同的角度对障碍物进行成像。这两个图像被用来找到同名点进行立体匹配，然后使用三角测距原理对障碍物进行定位[5]。这种技术的主要目标是防止机器人与障碍物发生碰撞。然而在实际应用中，还存在一些挑战。例如，对于远离相机的障碍物，由于视差较小，可能导致测量误差增大。同样，如果障碍物被其他物体遮挡，或者光照条件不佳，也可能影响视差的计算和匹配，从而影响测距的准确性。此外，双目立体视觉的精度也受到相机硬件如相机的分辨率、镜头的质量的影响，这些都会降低测距的准确性。因此，双目立体视觉障碍物检测

与定位技术精度的提高对于机器人的智能化水平，以及保障其在复杂环境中的安全运行，都具有重要的价值。

1.2 国内外研究现状

目标检测技术和双目立体定位技术是机器人视觉中两个密切相关的步骤。目标检测的准确性直接影响到双目立体定位的精度和可靠性。如果目标检测不准确或精度较差，可能导致双目立体定位结果出现偏差或误差。因此，在进行双目立体定位前，必须进行准确的目标检测，以确保机器人可以正确地识别并定位其环境中的目标。本节在双目立体视觉障碍物检测与定位技术的基础上，将其分为目标检测和双目视觉定位两大技术，并针对这两大技术分别做国内外研究现状分析。

1.2.1 目标检测技术

传统的目标检测算法依赖于人工设计的特征提取器来从图像中提取有用的信息，这些信息被用来区分目标和非目标。在这些特征提取方法中，方向梯度直方图（Histogram of Oriented Gradient，HOG）[6]是一种常用的方法，它能够捕捉图像的局部形状信息。尺度不变特征变换（Scale-Invariant Feature Transform，SIFT）[7]是另一种常用的特征提取方法，它能够在尺度变化、旋转和仿射变换下保持稳定。哈尔特征（Haar-like features）[8]是一种在计算机视觉中用于物体检测的特征，它能够快速地在图像中计算出来。局部二值模式（Local Binary Pattern，LBP）[9]是一种用于纹理分类的简单而有效的特征描述符。

在特征提取之后，需要使用分类器来进行目标检测。支持向量机（Support Vector Machines，SVM）[10]是一种常用的分类器，它通过找到一个超平面来最大化正例和负例之间的间隔，从而实现分类。AdaBoost[11]是另一种常用的分类器，它是一种自适应的学习算法，通过将一系列的弱分类器组合成一个强分类器来实现分类。

（1）VJ 检测器[8]。VJ 检测器是 Paul Viola 和 Michael Jones 在 2001 年提出的一种用于实时人脸检测的算法，这个检测器的设计主要考虑到了实时性和准确性的需求。VJ 检测器引入了一种叫做积分图像（Integral Image）的概念。积分图像是一种预处理技术，它可以在常数时间内计算出任何大小的图像窗口的像素强度之和，这极大地加速了特征计算的速度。此外，VJ 检测器还采用了 AdaBoost 自适应算法用于选择和组合最有区分能力的特征，这使得检测器在保持高准确率[12]。然而，尽管 VJ 检测器在当时的人脸检测任务中表现出了优秀的性能，但它仍然存

在一些挑战。由于需要采用不同尺度的滑动窗口对图像进行全局扫描，这导致了模型在处理大规模或高分辨率的图像时，VJ 检测器会面临一些问题。

（2）HOG 检测器[6]。HOG 检测器是由 N.Dalal 和 B.Triggs 在 2005 年提出的，主要用于解决行人检测问题。HOG 检测器的工作流程如下：首先，计算归一化后的图像中每个像素的梯度。然后，将图像分割成小的连接区域，称为“细胞”。每个细胞通常包含 6-8 个像素。接下来，为每个细胞构建一个方向梯度直方图。然后，将这些细胞合并成较大的区域，称为“块”。每个块通常包含 2-3 个细胞。最后，对每个块内的所有细胞的梯度直方图进行归一化。通过以上步骤，每个块就可以得到一个特征向量，这个特征向量可以用于描述图像的局部形状信息。然后，这些特征向量可以输入到一个分类器进行目标检测。HOG 特征的优点使它可以捕捉图像的局部形状信息，而且对于图像的局部光照和尺度变化具有一定的鲁棒性，HOG 检测器的出现仍然为后来的目标检测技术的发展奠定了基础。

（3）可变形部件模型（Deformable Parts Model，DPM）[13]。可变形部件模型是一种目标检测算法，最初在 2008 年由 P.Felzenszwalb 等人提出，后来由 R.Girshick 进行了一系列的优化[14]。DPM 算法的主要思想是将目标视为由多个部件构成的，这些部件可以有一定的变形。DPM 算法通过对这些部件的检测和定位，可以更准确地检测和识别目标。虽然 DPM 算法在当下的目标检测任务中可能已经没有明显的性能优势，但是 DPM 算法中提出的混合模型、边框回归等思想对后续算法的发展具有重要影响。例如，混合模型的思想被用于处理目标的尺度和视角变化；边框回归的思想被用于精确地定位目标的边界。因此，DPM 算法仍然被看作是一个重要的基准模型，用于评估和改进新的目标检测算法。

传统的目标检测算法，通常采用滑动窗口的方式在图像中搜索目标。滑动窗口的思想是在图像上以不同的位置和尺度滑动一个窗口，然后使用分类器判断每个窗口内是否包含目标。这种方法会导致计算量非常大，并且滑动窗口的方法使用的分类器通常是浅层的，不能学习复杂的特征。这使得这种方法在处理复杂场景和变化多样的目标时，性能往往不尽如人意，而且在性能上已经无法满足复杂场景下的目标检测需求。因此，需要寻找一种新的检测方法，深度学习，特别是卷积神经网络（CNN）[15]的出现，为目标检测带来了革命性的改变。这种新的方法可以自动学习和提取图像特征，而无需人为设定规则或参数，大大提高了目标检测的精度和效率。更重要的是，基于深度学习的目标检测模型具有更好的稳定性和鲁棒性，能够更好地应对复杂的场景和变化多端的环境。

深度学习在目标检测领域的发展可以说是一次革命性的变革，其中具有代表性的一种是两阶段目标检测算法。2014 年，Girshick[16]等人提出了 R-CNN 算法，

它的提出将深度学习技术成功应用到目标检测任务中。R-CNN 工作流程可概述为：首先，通过选择性搜索生成约 2000 个候选区域；其次，每个区域经过规一化后，通过预训练的 CNN 提取特征；最后，使用 SVM 进行区域分类，线性回归模型实现边界框回归。2015 年，He Kaiming[17]等人提出了 SPP-Net 算法，解决了 R-CNN 无法处理任意大小输入图像的问题。SPP-Net 在 CNN 的卷积层后加入了一个空间金字塔池化层，可以对任意大小的特征图进行固定长度的特征提取。但是，SPP-Net 仍然保留了 R-CNN 的一些缺点，例如需要分阶段训练和使用 SVM 分类器等。为了解决这些问题，Girshick 在 2015 年提出了 Fast R-CNN[18]算法。Fast R-CNN 将 R-CNN 的多个步骤集成到一个网络中，形成了一个端到端的训练流程。Fast R-CNN 不仅提高了计算效率，而且通过使用多任务损失函数，将分类和边界框回归集成到一个统一的训练过程中，提高了检测的准确性。

一阶段目标检测算法是另一种目标检测方法，其主要特点是直接在特征图上预测目标的类别和位置信息，无需先生成候选区域。2016 年，Redmon[19]等人提出了 YOLO 算法，开创了一阶段目标检测算法的新篇章。YOLO 将整个图像分成 S×S 个网格，每个网格负责预测落在其内的目标的类别和位置信息。这种方法大大简化了目标检测的流程，使得 YOLO 在准确性和速度上都得到很大的提高。然而，YOLO 也存在一些问题，例如对小目标和密集目标的检测精度不高，以及每个网格只能检测一个目标的限制。为了解决这些问题，Redmon 在 2017 年提出了 YOLOv2 算法[20]。YOLOv2 在 YOLO 的基础上做了许多改进，例如引入了锚框来处理不同形状的目标，使用批量归一化来提高网络的收敛性，移除了全连接层以提高检测速度，以及使用联合训练来提高检测的准确性。YOLOv2 不仅提高了检测的精度，而且保持了 YOLO 的高速度。后来，Redmon 又提出了 YOLOv3 算法[21]。YOLOv3 在 YOLOv2 的基础上做了进一步的改进，例如使用更强大的 Darknet-53 网络替换了 Darknet-19[22]网络，引入了特征金字塔来处理不同尺度的目标，以及使用逻辑分类器替换了 Softmax[23]分类器，并引用 K-means[24]算法聚类 9 种先验框来检测大、中、小三种尺度大小的目标。这些改进使得 YOLOv3 在保持高速度的同时，进一步提高了检测的精度。2020 年，Bochkovskiy[25]等人提出了 YOLOv4 算法。YOLOv4 在 YOLOv3 的基础上引入了许多新的技术，例如 CSPNet[26]、空间金字塔池化、PANet[27]等。然而，尽管基于深度学习的目标检测算法在许多任务中具有优秀的效果，但在实际应用中仍然面临一些挑战。例如，目标的遮挡、信号的干扰、光照条件的变化等，这些因素可能导致大量小目标在检测过程中被忽略或误检，从而降低了检测的准确性。因此，我们需要进一步研究和改进深度学习目标检测算法，以提高其在复杂环境中的应用效果。

1.2.2 双目立体视觉定位技术

双目立体视觉技术是模拟人类视觉系统的一种技术，它通过使用两个摄像头（相当于人类的两只眼睛）来获取环境中的三维信息。随着工业化技术的快速发展，双目立体视觉技术逐渐成形并开始被广泛应用。这种技术主要包括五个环节，如图1-1 所示。

图 1-1 双目立体视觉技术主要步骤

立体匹配是双目立体视觉的核心部分，其任务是找出两个图像中对应的像素点，这是计算深度信息的基础。立体匹配的方法通常可以根据约束范围区分为全局匹配和局部匹配。全局匹配考虑整个图像的信息，寻找全局最优的匹配结果。局部匹配只考虑像素点的邻域信息，计算复杂度低。此外，立体匹配也可以按照生成的视差图类型分为稀疏匹配和稠密匹配[28]。全局匹配方法中，动态规划能找到全局最优解，但计算复杂度高；图像分割可以将图像划分为多个区域，缩小匹配范围，提高匹配的精度[29]；置信度传播则是一种迭代的方法，通过不断更新匹配的置信度来找到最优匹配。Leung 等人[30]针对动态规划计算复杂度高的问题，提出了一种对行、列分别进行迭代的方法，这种方法可以大大减少计算的复杂度，提高立体匹配的速度。与此同时，马瑞浩等人[31]注意到传统的像素级匹配方法容易受到噪声和纹理不足等问题的影响，提出了使用 SLIC 算法进行超像素分割，将相邻且颜色相似的像素聚合在一起，形成一个超像素，然后在超像素级别进行匹配。这样不仅可以减少匹配的计算量，而且可以提高匹配的准确性。陈星等人[32]提出了一种结合卷积神经网络（CNN）和特征金字塔结构（FPN）的多尺度融合立体匹配算法，这种方法能在多个尺度上进行匹配，从而得到更准确的视差图。另一名伦敦大学的研究者提出了一种将置信度传播与 PatchMatch 相结合的 PMBP 算法，这种算法能在保证匹配精度的同时，降低匹配的误差率[33]。

局部立体匹配是双目立体视觉中的一种主要方法，因其计算复杂度较低，适合实时应用，因此被广泛使用。Wang 等人[34]提出了一种改进的区域匹配方法，引入了区域间的合作与竞争机制，并实现了协作优化模式。这种方法在保证匹配速度的同时，也提高了匹配的准确性。蒋文萍等人[35]则从另一个角度出发，提出了一种基于 Tanimoto 系数与 Hamming 距离算法结合的改进 Census 变换自适应局部立体匹配算法。这种方法通过改进 Census 变换，提高了匹配的准确性，同时也保证了匹配的实时性。此外，基于图像特征的立体匹配方法也得到了广泛的研究。例如，有

研究将 SIFT 特征提取算法应用于立体匹配，特别是对边缘点的匹配[36]。边缘点往往是图像中信息丰富的区域，但由于其特殊的位置和结构，常常导致传统匹配方法在这些区域的匹配错误。另一方面，基于相位的匹配算法在立体匹配的效果上表现优秀，甚至被认为是目前效果最好的方法。然而，这种方法的应用并不广泛，相位点是指图像中具有明显相位变化的点，如何准确地找到这些相位点，并将其用于立体匹配，仍然是一个待解决的问题[37]。双目立体视觉定位技术则具有成本低、鲁棒性强等优点。此外，研究表明，双目立体视觉定位技术的测距精度已经可以满足实际应用的要求[38]。

计算机科学技术的发展为计算机视觉技术的进步提供了强大的动力。近年来，计算机视觉在无人驾驶汽车、智能交通、虚拟现实等领域的应用研究取得了重大进展[39]。无人驾驶汽车是近年来的热门研究领域，计算机视觉技术在其中起到了关键的作用。通过计算机视觉技术，无人驾驶汽车可以识别道路、行人、车辆等环境信息，实现自动驾驶。机场行人检测、智能交通等领域也在利用计算机视觉技术提高服务质量和效率。而虚拟现实则是通过计算机视觉技术，实现虚拟与现实的无缝连接，为用户提供沉浸式的体验。

总之，随着计算机视觉技术等一些技术的发展，不仅提高了双目视觉技术的性能，也降低了其成本，使得更多的消费者能够接触和使用双目视觉技术。因此，双目视觉技术在实际环境中的使用逐渐受到更多学者以及开发者的关注。这些变化为双目视觉技术的进一步发展和应用提供了更广阔的空间。

1.3 论文主要研究内容及创新点

本文主要研究集中在移动机器人智能系统中障碍物检测与定位任务的两个主要方面：障碍物检测和障碍物定位。研究的主要内容包括：首先，针对现有目标检测技术对远距离小目标检测精度低以及容易出现漏检和误检的问题，本文采用注意力机制模块（CBAM）与多尺度特征检测模块相结合的方法对 YOLOv5 算法模型进行改进，以增强 YOLOv5 算法的检测精度，改善对较小目标的识别效率。其次，在基于双目视觉的物体定位与测距方法中，由于特征点匹配算法 SGBM 的局限性，导致其对于物体定位与测距精度而言并不高，误差较大。针对 SGBM 算法对弱纹理，光照变化敏感的问题。本文提出结合直方图均衡化、多尺度变换和滤波处理来优化 SGBM 算法，并利用改进后的 SGBM 算法对双目图像进行匹配，生成视差图，为双目定位和测距提供准确数据。然后，针对双目相机标定过程较繁琐，标定参数准确性不易确定等问题。本文提出两段式标定分析法，以提高标定效率和得到准确度高的标定参数。最后，本文以双目摄像头为研究对象，提出了一种结合

改进 YOLOv5 障碍物检测算法和改进 SGBM 立体匹配算法的定位测距方法，对左目图像的特定障碍物目标进行识别检测，获取其类别、位置及距离信息。此外，本文还设计了一个原型系统，该系统不仅集成了基于双目摄像头的障碍物检测与定位算法，还实现了常用的图像操作功能。根据本文的研究内容，研究内容如图 1-2 所示，研究技术路线如图 1-3 所示。

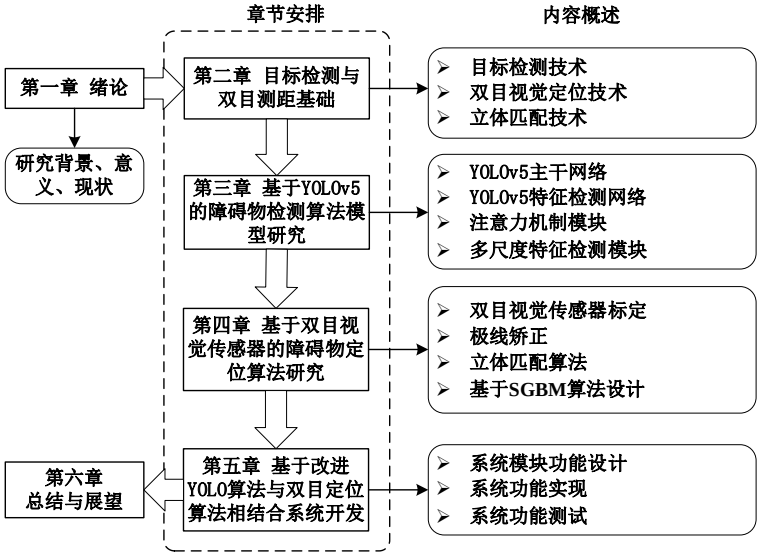


图 1-2 研究内容

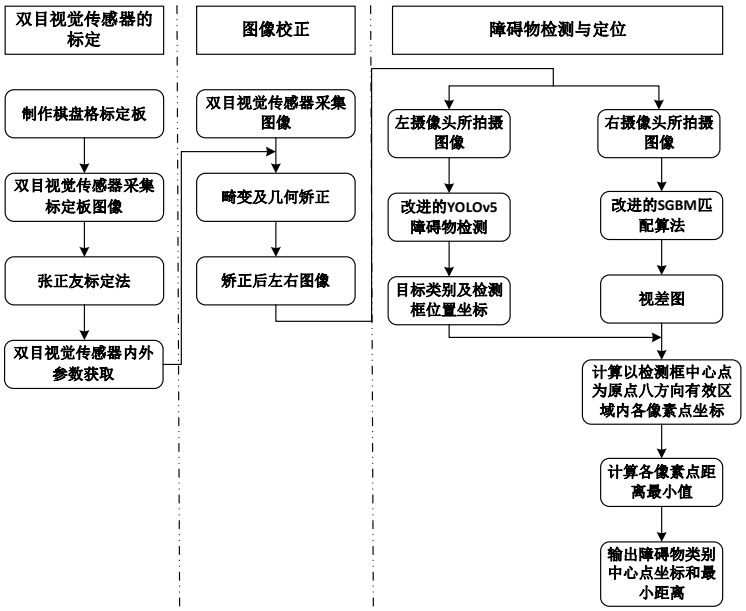


图 1-3 技术路线

1.4 论文组织结构安排

根据研究内容，论文共有六个章节，以下对每个章节的研究内容进行简要介绍：

第一章为绪论部分。首先介绍基于双目视觉的障碍物检测与定位技术的研究背景和意义，其次总结目标检测和双目定位技术的国内外研究现状，之后概述本文的主要研究内容及创新点，最后归纳总结论文结构和技术路线图。

第二章为目标检测与双目定位研究基础。首先介绍目标检测技术的研究基础，然后介绍双目定位技术的相关概念，最后对立体匹配的主要方法及各方法存在的问题进行介绍。

第三章为基于 YOLOv5 的障碍物检测算法模型研究。首先介绍神经网络模型训练数据集的制作技术，其次详细阐述 YOLOv5 模型结构，然后介绍基于卷积注意模块（CBAM）与多尺度特征检测模块对 YOLOv5 模型的改进，最后针对两种模型进行目标检测实验，并对实验结果进行对比分析，为第五章基于改进 YOLO 算法与双目定位算法相结合系统开发做铺垫。

第四章为基于双目视觉传感器的障碍物定位算法研究。本章先是对双目视觉传感器进行了标定与矫正，并针对双目相机标定过程繁琐，标定参数准确性不易确定等问题，提出两段式标定分析法，然后介绍三类常用的立体匹配算法，并对各类立体匹配算法进行视差图对比实验，分析各个算法的优缺点，最后针对 SGBM 匹配算法存在的问题，提出相应的改进方法，并将该方法用于视差图检测实验以及动态测距实验，对实验结果进行分析与评估。

第五章为基于改进 YOLO 算法与双目定位算法相结合的系统设计与实现。该系统不仅集成目标检测算法和双目定位算法，还实现一些常用的图像处理功能。最后对各模块功能进行了测试。

第六章为总结和展望。总结了本文的研究内容，并分析了本文研究中存在的问题，以及对后续工作的改进提出一些建议。

第二章 目标检测与双目定位研究基础

在基于双目视觉障碍物检测与定位技术的研究中，首先需要目标检测技术识别在图像中障碍物的种类和所处的位置，再通过立体匹配技术生成图像的视差图，最后结合双目定位技术进行距离的计算和准确地定位。本章将详细介绍以上三个技术的研究基础。

2.1 目标检测技术

目标检测是将图像中的目标识别出来的过程，它是许多视觉任务，如障碍物识别、机器人避障的基础，其准确性将直接影响这些任务结果的可靠性。因此，在深度学习和计算机视觉领域，目标检测的研究和应用具有至关重要的地位。

2.1.1 深度学习网络

人工神经网络的概念源自 1943 年，由 W.S.Mc Culloch 和 W.Pitts 首次提出[40]。这一概念的核心特征在于网络中含有大量的参数，这些参数的值并不是预先设定的，而是通过对大量数据的学习进行调整和优化，使得网络的输出能够尽可能地接近期望的结果。然而，由于当时计算机技术的限制，人工神经网络的发展受到了阻碍，进展缓慢。直到 2006 年，Hinton[41]提出了一种深层网络模型—深度置信网络（DBN），人工神经网络的发展才得以重启。深度学习的深度模型是对传统人工神经网络的自然扩展。与较浅层的网络模型相比，深度模型具有更多的层和参数，可以学习和表示更复杂的模式和结构，因此具有巨大的潜力。许多科研学者提出了不同的深度学习模型，这些模型各有特点，适应不同的任务和数据类型。例如，卷积神经网络（CNN）在处理图像和视频数据上表现出色，因为它能够有效地提取局部特征和空间结构。在语音识别、机器翻译等领域，循环神经网络（RNN）[42]通过捕捉时序中的信号和语义信息，对序列数据进行处理。

深度学习的核心思想，即通过构建深层次的、非线性的网络结构，来模拟和理解复杂的数据模式[43]。深度学习可以被看作是一个自动的、端到端的学习过程。对于给定的任务，比如图像识别或语音识别，深度学习模型可以自动地从原始数据中学习到有用的特征，而不需要人类专家的介入。在图像处理中，CNN 有着不俗的表现，接下来将介绍 CNN 的工作原理和网络结构。

2.1.2 卷积神经网络

随着深度学习技术的发展，CNN 已经被应用于各种任务中，包括图像识别、目标检测、语义分割等。在这些任务中，CNN 都取得了出色的表现，甚至超越了人类的表现。其结构如图 2-1 所示。

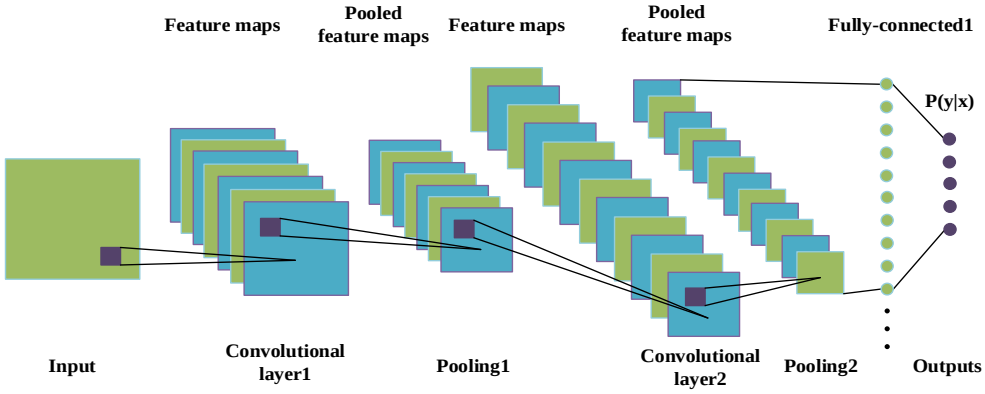


图 2-1 卷积神经网络

如图 2-1 所示，卷积神经网络包括多个卷积层、池化层，以及全连接层。卷积层和池化层负责从原始数据中提取有用的特征，全连接层则负责将这些特征进行整合，以进行最后的分类或回归任务。

（1）卷积层

卷积层的任务是对输入数据进行特征提取。在处理图像数据时，卷积层会用一组卷积核对输入图像进行卷积操作，每个卷积核都能够提取出图像的某种特定特征，比如边缘、角点、纹理等。理论上，通过设置更大的卷积核，可以扩大模型的感受野，从而提取更广泛的特征信息。然而，这个关系并不是线性的，因为感受野的大小也受到网络层数、池化操作等其他因素的影响。因此，设计有效的卷积层和选择合适的感受野，是深度学习网络设计中的重要任务。卷积层的设计中包括三项重要的超参数[44]，即卷积核尺寸、步长和填充。首先，卷积核尺寸决定了模型在进行特征提取时，能够关注到的输入图像的局部区域的大小。较大的卷积核尺寸能够帮助模型提取更复杂的特征。其次，步长是决定卷积核在图像上移动的步幅。步长的大小决定了特征图的大小，较大的步长可以降低特征图的尺寸，从而降低计算量。最后，填充是指在输入图像的边缘添加额外的像素，以允许卷积核在图像边缘进行卷积运算。填充的大小通常根据卷积核的尺寸和步长来确定，以保证特征图的尺寸不会因为卷积运算而减小。填充不仅可以帮助模型更好地提取边缘区域的特征，也可以防止特征图的尺寸过小，从而影响模型的性能。最常用的填充方法之一是全零填充，顾名思义，就是在输入图像的边缘添加零值像素。这种方式的优点是简单易实现，且不会引入任何额外的信息，从而避免了对原始图像信息的干扰。重复边界

填充是另一种填充方法，即根据图像的边界值进行填充。这种方法的优点是能够在一定程度上保留图像边缘的信息，从而提高模型在处理边缘区域特征时的性能。采用全零填充方式的一次卷积运算图如图 2-2 所示。

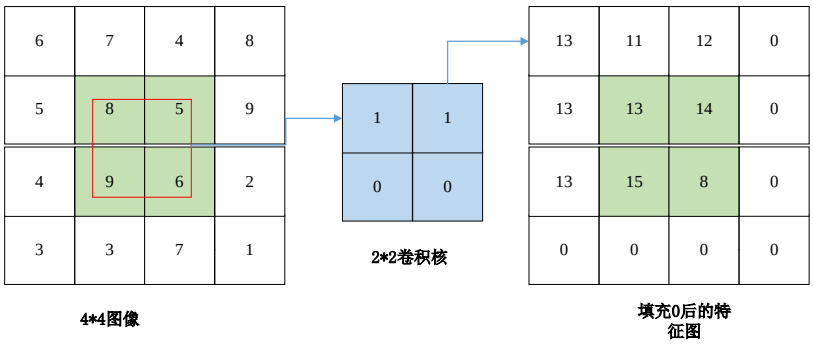


图 2-2 全零卷积填充示意图

（2）池化层

池化层在卷积神经网络（CNN）中的作用是对输入信息进行降采样操作。降采样是一种数据压缩技术，其目标是减小数据的规模，同时尽可能地保留原始数据的重要信息。卷积层和池化层的组合是 CNN 的一个基本构成单元。卷积层负责提取图像的局部特征，而池化层则对这些局部特征进行压缩，以提取更高层次的特征，同时也降低模型的计算复杂度。

池化层的实现方法常见的有三种，分别为最大池化、平均池化和最小池化。这些方法的核心思想都是在池化窗口内进行某种特定的操作，并将结果作为该窗口的输出。随着深度学习研究的深入，学者们也提出了其他更复杂的池化方法，如重叠池化和空间金字塔池化等。这些方法在一定程度上提高了模型的性能和灵活性。由图 2-3 可知三类池化的具体操作。

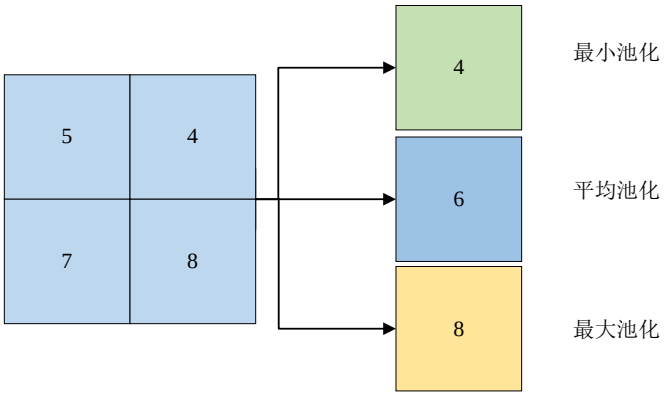


图 2-3 常见池化方式

（3）全连接层

在卷积神经网络中，图像信息经过卷积层和池化层的处理后，得到了深层次的特征表示。接下来，全连接层则负责对这些特征进行分类。全连接层的设计主要涉及三个关键参数：全连接层的层数，每个全连接层的神经元数量，以及选用的激活函数。首先，全连接层的层数决定了网络的深度，其直接影响模型的复杂性和学习能力。其次，全连接层的神经元数量则决定了网络的宽度。宽度是指每一层中神经元的数量，它决定了模型在每一层中可以学习的特征数量。最后，在神经网络中，激活函数的主要作用是引入非线性因素，使得模型能够学习和执行更复杂的任务。需要注意的是，全连接层的深度和宽度的增加可以提高模型的复杂度，从而增强模型的表达能力。然而，这也可能导致模型过拟合，即模型对训练数据过度适应，而在测试数据上的泛化能力下降。此外，增加全连接层的深度和宽度也会增加模型的计算复杂度，从而降低算法的运行效率。因此，在设计全连接层时，需要权衡模型的表达能力和泛化能力，以及计算效率。

2.1.3 目标检测算法

目标检测在许多实际应用中都有广泛的需求，如无人驾驶、视频监控、图像分析等。最初的目标检测算法主要采用两阶段（Two-stage）的方法。这种方法首先会对输入的图像数据生成一系列的候选框，然后对这些候选框进行分类或者回归分析，以确定其中是否包含目标对象以及目标的具体位置。这种算法的优点是检测精度较高，能够准确地定位出目标对象。然而，它的缺点是计算过程较为复杂，检测速度慢。常见的两阶段方法的模型包括 R-CNN、Fast-RCNN、Faster-RCNN[45]以及 Mask-RCNN[46]等。因此，为提高模型的检测速度，单阶段检测算法应运而生。这种方法不再分两步进行，而是直接对输入的图像数据进行预测，并一次性输出目标的类别与候选框的位置。这样的设计大大提升了目标检测的速度，使得模型能够满足实时性的要求。主流的单阶段检测算法有 YOLO 系列。常见的模型有 YOLO、YOLOv2、YOLOv3、YOLOv4、YOLOv5、YOLOX[47]、YOLOv6[48]、YOLOv7[49]等，它们具有速度快、精度高等优点。

2.2 双目视觉定位技术

双目立体视觉是利用两个相机从不同角度捕获图像，通过计算两个图像间的视差来恢复出场景的三维信息的过程。双目视觉是实现深度感知和物体定位的基础，其精度直接影响到深度和位置估计的准确性。因此，在计算机视觉等领域，双目立体视觉的研究和应用具有至关重要的作用。

2.2.1 摄像机成像模型

相机拍摄是将三维空间的物体投影到二维图像平面的过程。这个过程涉及到的物理和数学原理是非常复杂的，本节将对该过程进行详细阐述。在双目立体的研究中，有如下三个参考坐标系显得十分重要，它们分别是影像坐标系、相机坐标系和世界坐标系，下面将分别进行介绍。

（1）影像坐标系

在电子计算机中，图像被存储为一个二维矩阵，通常会将左上角的像素点作为坐标系的原点，这是因为在计算机的图像处理中，会从左上角开始扫描和处理像素点。然而，在相机的成像平面，投影点的坐标有两种表现形式：一种是以像素为单位，即像素坐标系；另一种是以物理长度为单位，即物理坐标系。物理坐标系是以相机的光心为原点，以物理长度为单位，用于描述投影点在成像平面上的实际位置。它们关系如图 2-4 所示。

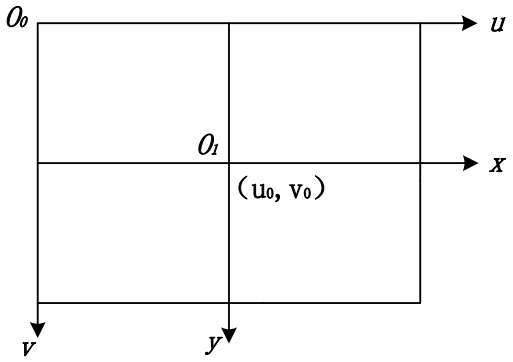


图 2-4 像素坐标系与物理坐标系关系图

像素坐标系的坐标原点为O_0，物理坐标系的坐标原点为图像像素坐标系的中心点$O_1(u_0,v_0)$，x轴与y轴分别与u轴和v轴平行。在物理坐标系中，假设每个像素点在图像中的相对位置是均匀分布的，考虑到图像中像素点的实际位置，每个像素点的位置是以物理长度为单位表示。设每个像素点坐标的单位长度分别为dx和dy。两个坐标系的转换关系可以通过一个线性变换来实现，转换关系见式(2-1)。

$$\begin{cases} u = \dfrac{x}{dx} + u_0 \\ v = \dfrac{y}{dy} + v_0 \end{cases} \tag{2-1}$$

当用齐次矩阵表示上式关系时，见式(2-2)。

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

在计算机视觉领域，三维重建是一个重要的研究方向，它的目标是从二维图像中恢复出三维的空间信息。这个逆向过程可由式(2-2)逆推得到式(2-3)。

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} dx & 0 & -u_0 dx \\ 0 & dy & -v_0 dy \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{2-3}$$

（2）相机坐标系和世界坐标系

相机坐标系示意图如图 2-5 所示：

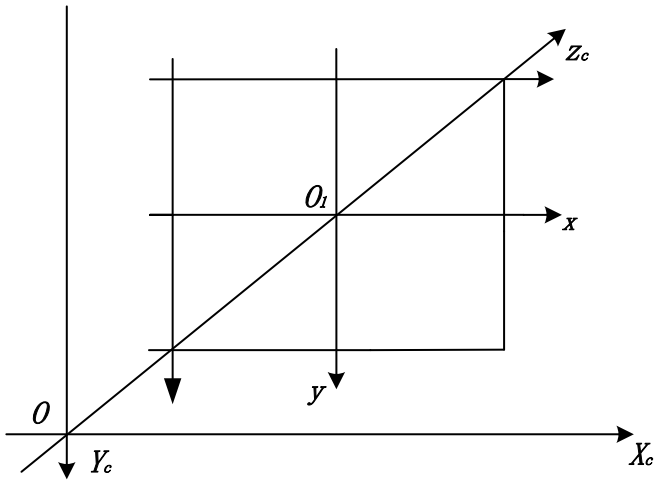


图 2-5 相机坐标系

在计算机视觉和图像处理的领域中，相机坐标系是一种常用的坐标系，它是以相机的光心为原点，光轴为Z轴，其成像平面的两个正交方向分别为X轴和Y轴。这个坐标系的定义使得我们可以方便地描述和计算相机成像过程中的几何关系。

然而，在实际的应用场景中，相机通常不是静止的，而是动态的。因此需要建立一个世界坐标系，如图 2-6 所示，它通常是一个固定的坐标系，用于描述三维空间中物体的绝对位置。

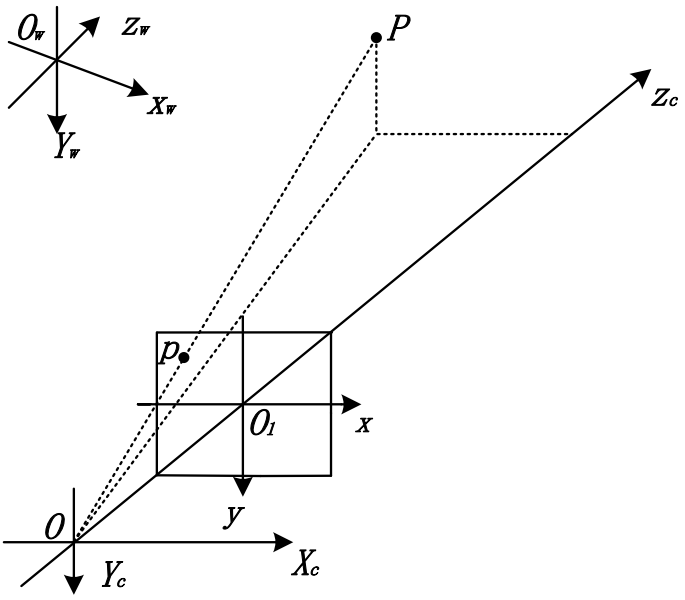


图 2-6 相机坐标系和世界坐标系

其中$O_wX_wY_wZ_w$为世界坐标系，空间点P在世界坐标系下的坐标为(X_w, Y_w, Z_w)，空间点的投影点p在相机坐标系下的坐标为(X_c, Y_c, Z_c)，世界坐标系与相机坐标系之间的关系可以通过一个刚体变换来描述，这个变换包括了旋转和平移两个操作。具体关系式见式(2-4)。

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2-4}$$

其中R为旋转矩阵，是 3×3 的单位正交阵，T是平移向量，为 3×1 的向量，0^T向量的值为$(0,0,0)^{\mathrm{T}}$，M_1为 4×4 的矩阵。在双目系统中，世界坐标系与左摄像机的相机坐标系重合。

（3）小孔成像模型

在计算机视觉和图像处理领域中，小孔模型是一种基本的相机成像模型，它为我们提供了一个理论框架，用以理解和描述相机成像过程中的基本几何和光学原理。小孔模型的基本假设是，三维空间中的物体表面反射的光线遵循直线传播的规律。这些光线通过相机的一个微小开口，然后在相机的成像平面上形成聚焦投影。如图 2-7 所示。

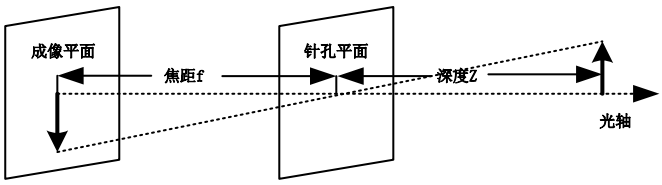


图 2-7 小孔成像模型

在小孔相机模型中，从小孔（或称为光心）到成像平面的距离被定义为焦距 f。这是一个重要的参数，因为它决定了相机的视场以及图像的放大或缩小程度。物体到小孔的距离则被表示为Z。根据光学原理，小孔模型的成像结果是一个倒立的图像，这是因为光线在通过小孔后会发生方向的改变，使得图像在成像平面上呈现出倒立的状态。通常会使用一个数学等价模型来代替原始的小孔模型。该等价模型通如图 2-8 所示。

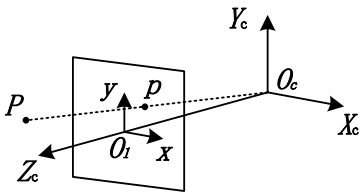


图 2-8 数学等价模型

在小孔模型中，将小孔平面视为相机平面，三维空间中的点(X_c, Y_c, Z_c)通过小孔投影到相机平面上，形成一个二维的投影图像，投影点坐标为$p(x, y)$。这个投影过程如图 2-9 所示：

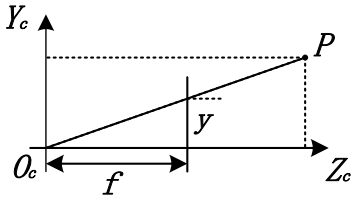


图 2-9 投影截面图

由图 2-9 可得两点之间的坐标转换关系，见式(2-5)。

$$x = \frac{fX_c}{Z_c}, y = \frac{fY_c}{Z_c} \tag{2-5}$$

表达为齐次矩阵见式(2-6)。

$$Z_c\begin{pmatrix}x\\y\\1\end{pmatrix}=\begin{pmatrix}f&0&0&0\\0&f&0&0\\0&0&1&0\end{pmatrix}\begin{pmatrix}X_c\\Y_c\\Z_c\\1\end{pmatrix} \tag{2-6}$$

在式(2-6)中，$(X_c,Y_c,Z_c,1)$代表相机坐标系下的投影点坐标，这是一个四维齐次坐标表示，它可以方便地表示和处理三维空间中点的变换。$(x,y,1)$则代表该点在物理坐标系下的坐标，f是相机的焦距。式(2-6)描述的是从三维空间映射到二维平面的过程。

将式(2-3)和式(2-4)代入式(2-6)，可以得到点P的坐标(X_w,Y_w,Z_w)。在从相机坐标系到世界坐标系的转换过程中，需要使用相机的内参和外参来描述和计算这个坐标转换，并得到三维空间点P与投影点p的坐标转换关系，见式(2-7)。

$$\begin{aligned}Z_c\begin{bmatrix}u\\v\\1\end{bmatrix}&=\begin{bmatrix}\frac{1}{dx}&0&u_0\\0&\frac{1}{dy}&v_0\\0&0&1\end{bmatrix}\begin{bmatrix}f&0&0&0\\0&f&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&T\\0^T&1\end{bmatrix}\begin{bmatrix}X_w\\Y_w\\Z_w\\1\end{bmatrix}\\&=\begin{bmatrix}f_x&0&u_0&0\\0&f_y&v_0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}R&T\\0^T&1\end{bmatrix}\begin{bmatrix}X_w\\Y_w\\Z_w\\1\end{bmatrix}=M_1M_2P\end{aligned} \tag{2-7}$$

式(2-7)中，焦距f是相机模型的已知参数，它决定了相机的视场和图像的放大缩小程度。M_1是相机的内参数矩阵，它包含了相机的物理属性。这些属性包括x轴和y轴上的焦距f_x和f_y（它们分别等于焦距f除以像素的物理尺寸dx和dy），以及图像的原点坐标(u_0,v_0)。这些参数通常需要通过标定过程来确定。M_2是相机的外参数矩阵，描述了相机坐标系和世界坐标系之间的关系。如果知道M_1和M_2的值，以及点P在世界坐标系中的坐标，我们就可以使用上式来计算出空间点P在图像上对应的投影点p的坐标。

（4）双目模型

常见的双目相机可以分为两种：相机光轴相交和光轴平行，两种双目相机结构的示意图如图 2-10 所示。

左侧的配置被称为光轴相交的结构，这种配置可以捕获到更大的视场，但在处理图像对齐和深度计算时可能会带来一些挑战，因为图像之间的几何关系可能会比较复杂。相对于光轴相交的结构，右侧的配置被称为光轴平行的结构，在这种配置下，两个相机拍摄的图像在水平方向上的偏移量可以直接用来计算物体的深度

信息，从而简化了深度计算和图像对齐的过程。本文基于光轴平行的成像模型进行研究。

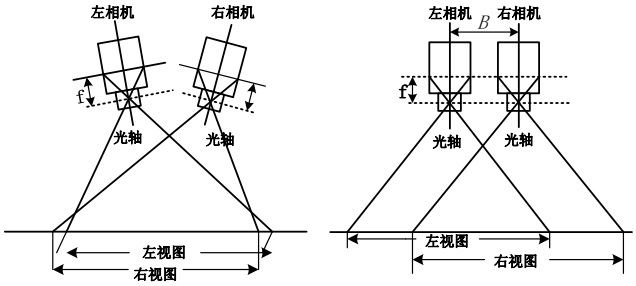


图 2-10 双目相机结构示意图

将光轴平行的双目模型简化，如图 2-11 所示。

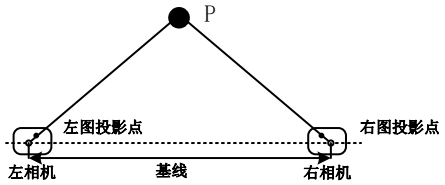


图 2-11 简化双目模型示意图

左右两相机都可以视作为小孔模型，两个相机的光心连线构成的线段称为基线。

2.2.2 双目相机参数标定

在双目立体视觉系统中，测距的基本原理是通过获取双目图像中目标点的位置，然后利用二维图像与三维空间的坐标系转换关系来计算目标点的三维坐标。这种转换关系的建立是通过摄像机标定来实现的。摄像机标定是一个关键的过程，该过程通常通过一组已知世界坐标系坐标和图像坐标系坐标的点来计算出相机的参数。

（1）双目摄像头标定

相机内参包括焦距、图像中心坐标以及镜头畸变参数等。在单摄像头标定中，内参误差可以通过一个 3×3 的矩阵来表示，有五个未知参数，包括焦距（f_x,f_y）、像素尺度因子 $skew$、以及图像中心坐标（C_x，C_y）。矩阵表示为 $[f_x, skew, C_x; 0, f_y, C_y; 0,0,1]$。

此外，因为相机使用透镜提高成像的采光效率，所以会引入径向畸变和切向畸变误差[50]。径向畸变可以通过三个参数k_1、k_2、k_3来描述，其中k_3的值一般为 0。

切向畸变是由于透镜不完全平行而产生的，它可以通过两个参数p_1、p_2来描述。在理想情况下，图像传感器的像素应该是规则排列的，其行和列应该是相互正交的。除此之外，由于制造过程中的误差、装配不精确等因素，图像传感器的像素可能不会完全正交，这就会导致 *skew* 参数的产生。

最后，通过摄像头拍摄到的二维图片标定得到外部参数。具体的参数如表 2-1 所示。

2-1 双目摄像头标定参数

标定参数	表达式	自由度
内参矩阵	$\begin{bmatrix} f_x & skew & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$	5
畸变系数	$\begin{bmatrix} k_1 & k_2 & p_1 & p_2 & k_3 \end{bmatrix}$	5
外参矩阵	$\mathrm{R} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$ $\mathrm{T} = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}$	6

（2）标定方法的选择

在摄像机参数标定中，传统标定方法和自标定方法[51]应用广泛。传统标定方法包括直接线性变换（DLT）方法[52]、非线性最优化法[53]、Tsai 两步法[54]、3D 立体标定法以及张正友平面标定法[55]等。

然而，传统标定方法的缺点是需要已知形状和尺寸的参照物，这在一些实际应用中可能不容易实现。例如，在需要频繁调整摄像机位置或者无法设置已知参照物的情况下，传统标定方法可能无法使用。因此，自标定方法通过分析摄像机在运动过程中拍摄多组图像之间的对应关系来标定摄像机的参数[56]来解决这个问题。

对各种标定方法进行对比，结果如表 2-2 所示[57]。

表 2-2 标定方法对比

标定方法	特点
直接线性变换法	没有考虑相机畸变，精度较低
非线性最优化法	影响标定因素多，准确性较低
Tsai 两步法	仅考虑径向畸变，鲁棒性低
3D 立体标定法	标靶成本高
张正友平面标定法	精度较高，操作简单，制作标定图案方便

张正友平面模板标定法于2000年提出。这种方法的主要优势在于只需要一个平面模板，而不需要像传统方法那样需要复杂的三维参照物。这大大简化了标定过程，同时也使得标定过程更加灵活和方便。在使用张正友标定法进行摄像机标定时，通过改变标定模板的方向和角度，对标定模板进行多次拍摄[58]。在这个过程中，只有摄像机的外部参数会随着模板的旋转和移动而变化。对获得的二维图像信息进行处理后，可以得到一个线性模型的初步解。然而，由于透镜的畸变效应，这个初步解可能并不准确。因此，张正友标定法进一步利用最大似然准则进行非线性优化，以对镜头产生的畸变进行校正。这个过程将得出摄像机参数的优化解，包括内部参数和外部参数。

在计算机视觉领域的实验研究中，摄像机标定是必不可少的一步，其目标是准确地确定摄像机的内部和外部参数。不同的标定方法具有不同的优点和缺点，选择适合的标定方法是实验设计的关键部分。

在本研究中，对比了多种摄像机标定方法，相比于其他的标定方法，张正友标定法具有显著的优势。首先，张正友标定法使用的平面标定模板易于获得，成本较低，这使得实验的准备工作更为简单和经济。其次，张正友标定法的操作过程简单，这有助于减少操作错误的可能性，提高实验的可重复性。最重要的是，张正友标定法能够提供较高的标定精度，这对于实验要求至关重要。因此，在考虑到这些因素后，本文决定在第四章的相机标定实验中选用张正友平面标定法对双目相机进行标定。

2.2.3 双目测距基本原理

（1）极线几何原理

极线几何是立体视觉研究中的一个核心概念，它描述的是两个相机图像平面间的几何关系。这种关系对于实现像素匹配、深度估计等关键任务具有重要意义。在双目相机模型中，基线定义了两个相机光心之间的距离和方向。在双目视觉系统中，极线约束是一个重要原理，它表明了一个三维空间点在两个图像平面上的投影点必然位于对应的极线上。最后是极点，它是基线与左右相机成像平面相交的点。极点是极线的交点，所有的极线都会通过极点。

这些概念之间的关系可以通过图 2-12 进行直观地展示。在这个图中，P_l和P_r分别代表了一个三维空间点在左右两个图像平面上的投影点。结合前文所提到的概念，O_lO_r即是基线，平面PO_lO_r即是极平面，e_l和e_r即是极点，而L_1和L_2即是极线。

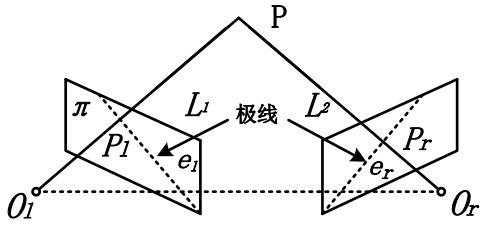


图 2-12 极平面、极点和极线的关系图

具体来说，当在一幅图像上确定了一个投影点，并知道相机的光心位置以及极点位置后，我们就可以通过连接投影点和极点来确定对应的极线。这条极线与确定的极平面相交，极平面是由三维空间中的某一点和两个相机的光心所确定的。这意味着，这个极平面在另一幅图像上与成像平面的交线，也就是另一条极线上，必然存在着这个投影点的对应点。这是由于所有的极线都会通过同一个极点，且极线的排列顺序与三维空间中点的排列顺序一致。这个极线约束的存在，使得在图像间的搜索匹配点由二维转化为沿极线的一维搜索，大大降低了计算复杂度。这是一个关键性的优化，因为在实践中，图像匹配是一个计算密集型的任务，任何可以减少计算量的方法都会对整体性能产生显著的影响。

（2）视差原理

人类的视觉系统是双目视觉的典型实例，它通过两只眼睛获取不同的视觉信息，然后在大脑中融合这些信息，形成对环境的三维感知。当人类同时用左右眼睛观察物体时，虽然每只眼睛接收到的图像略有不同，但大脑可以将这两个图像融合在一起，形成一个稳定的、连续的视觉感知，而不会感到物体位置的改变。这个过程在生理和神经层面上都是极其复杂的，但可以通过简单的模型进行模拟如图 2-13，以帮助我们理解视差的产生原理。

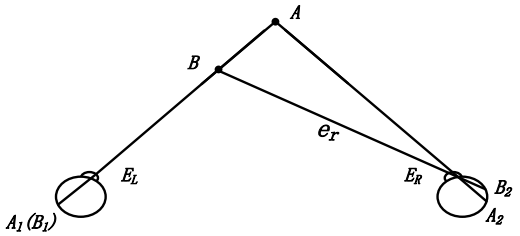


图 2-13 人眼视差模型

在三维空间中，设A和B分别为空间中两点，E_L和E_R分别代表左眼和右眼。当人的双眼同时观察这两个点时，左眼成像的投影点为A_1和B_1，同理，右眼成像的投影点为A_2和B_2。这种情况下，可以直观地观察到，同一点在左眼和右眼成像位置存

在差异，这种差异被称为双目视差。双目视差是我们感知三维空间和深度的重要机制。

在计算机视觉领域，我们通常使用两个相机模拟人的双眼，以获取物体的立体信息。图 2-14 是双目成像的数学几何模型。

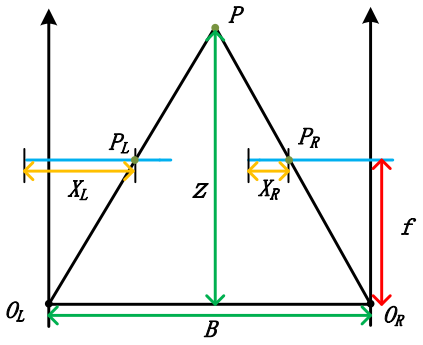


图 2-14 视差原理图

在图 2-14 中，O_L和O_R分别代表两个相机的光心，也就是光线进入相机的位置。f代表相机的焦距，这是一个固定的参数。B 代表两个相机的基线，即两个光心之间的距离，决定了系统对于深度的感知范围和精度。在三维空间中，P点是观察的物体点。当两个相机同时拍摄时，P点在左右两个相机的成像平面上产生的投影点分别是P_L和P_R。X_L和X_R分别代表这两个投影点在各自相机的图像坐标系中的横坐标。Z代表物体点P到基线B的距离，也就是通过双目视觉系统计算出来的P点的深度信息。

为了更好地理解双目视觉的原理，将线段$O_L P$沿着基线B的方向右移，使其与$O_R P'$重合，从而得到一个新的模型，如图 2-15 所示。

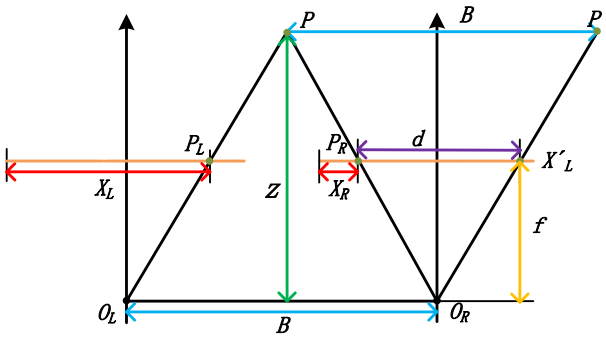


图 2-15 移动后的视差原理图

由图 2-15 可知，$X'_L - X_R$为视差d，通过相似三角形关系可得式(2-8)。

$$\frac{d}{b}=\frac{f}{Z} \tag{2-8}$$

因此所求点 P 的深度 Z 见式(2-9)。

$$Z=\frac{fb}{d}=\frac{fb}{X_l-X_r} \tag{2-9}$$

由公式（2-9）得知，三维世界中某一点的深度 Z 与其视差值 d 存在反比关系。即视差值 d 越大，对应的深度 Z 越小，反之亦然。这种现象与我们在实际生活中的感知规律相符，即我们对近处物体的深度变化更敏感，而对远处物体的深度变化则不太敏感。立体匹配算法从两幅图像中找到同名点，计算出视差值。进一步，根据视差值和已知的相机参数（如焦距和基线距离），就可以计算出这些点在三维空间中的深度信息。

2.3 立体匹配技术

立体匹配是指在双目视觉或多视角视觉中，找出不同图像间对应的像素或特征点的过程。立体匹配是进行深度估计和三维重建的基础，其精度将直接影响到三维信息的准确性和完整性。因此，在双目立体视觉和机器人导航等领域，立体匹配的研究和应用具有非常重要的地位。

2.3.1 基于区域的立体匹配算法

基于区域约束的立体匹配算法是一种主要通过比较图像中的相关区域关系来进行匹配的方法。这种方法的关键思想是，如果两个图像中的某一区域在三维世界中对应同一物体，那么这两个区域的像素值应该具有相似性[59]。根据用于度量区域相似性的方法不同，这类算法可以进一步分为 SAD[60]、SSD 和 ZSSD[61]三种。这些立体匹配算法的主要工作流程是，首先在待匹配的两幅图像中选择一个匹配窗口，然后通过计算窗口中的像素值的相似度度量值，找到对应的匹配窗口。通过这种方式，可以得到一个相对稠密的立体匹配视差图[62]。最后，通过优化相似度度量函数，可以找到每个待匹配区域的最优匹配结果。然而，这类算法面临一个主要的挑战，即如何选取合适的匹配窗口。如果窗口过大，可能会包含多个不同的物体，导致匹配错误；如果窗口过小，可能会受到噪声和纹理不足的影响，导致匹配不稳定。因此，如何根据图像的特性和应用需求，合理地选取和调整匹配窗口，是这类立体匹配算法研究和应用的一个重要问题。

2.3.2 基于特征的立体匹配算法

特征信息包括边缘、角点、拐点等各种图像特征。基于特征的立体匹配算法便是首先从图像中提取这些特征信息，并将其作为匹配的基元。常见的特征描述子类型包括 SIFT[63]特征算子和 SURF[64]特征算子等。这些特征算子能够在图像中检测出独特和稳定的特征点，为后续的匹配提供依据。然而这种算法获取的视差信息并不全面，只能反映出图像中特定特征的空间布局和深度信息。尽管基于特征的立体匹配技术在匹配速度上有优势[65]，因为它只需要处理图像中的特征点，而不是所有的像素，但所得到的视差图和三维重建的精度相对较低。这是因为这种方法只关注特定的特征信息，而忽略了图像中的其他信息，如区域的纹理和颜色信息等。这种信息的缺失可能导致在一些场景中，如特征稀疏或者特征重复的场景，匹配结果的不准确和不稳定。因此，如何在保持匹配速度的同时，提高匹配的全面性和精度，是基于特征的立体匹配算法研究和应用的一个重要问题。

2.3.3 全局立体匹配算法

前两种匹配算法主要通过寻找合适的匹配基元进行匹配。然而，这些算法所获取的视差信息不够全面和精确，因为它们通常只关注图像的部分信息，而忽略了其他可能有用的信息。为了解决这个问题，研究者提出了基于全局最优的立体匹配算法。该算法主要思路是，首先设定一个全局能量函数，这个能量函数通常包含数据项和平滑项。数据项度量了视差图的一致性，而平滑项度量了视差图的平滑性。然后，利用全局优化理论方法，如图割、置信传播和动态规划等，来最小化这个能量函数，从而获取最优视差。这种方法的优点是，它可以考虑到图像的全局信息，从而获得更全面的视差信息。同时，通过全局优化，可以提高视差的精确度，尤其是在纹理不足和深度不连续的区域。常见的基于全局最优的匹配算法包括基于图像分割的立体匹配算法[66]、基于置信传播的立体匹配算法[67]和基于动态规划的立体匹配算法[68]等。

基于全局最优的立体匹配算法的核心是构建并优化一个能量函数，如公式(2-10)所示：

$$E(d) = E_{data}(d) + E_{Smooth}(d) \tag{2-10}$$

在这个公式中，d代表视差图，E_{data}代表视差数据项，E_{Smooth}代表视差平滑项。视差数据项E_{data}度量了视差图与输入图像的一致性，即视差图中每个像素的视差值应与其在输入图像中的实际视差值尽可能接近。视差平滑项E_{Smooth}度量了视差图的平滑性，即视差图中相邻像素的视差值应尽可能接近。优化过程的目标是找到一个

视差图，使得能量函数达到最小。这个最小能量对应的视差图被认为是最优视差图，反映了物体表面的三维结构。

2.3.4 局部立体匹配算法

基于局部最优的立体匹配算法是利用局部优化的匹配方法来估计立体视差。步骤为：搜索待匹配的窗口图像特征点和最相似区域之间的视差值。这一步骤是基于图像特征的匹配，目标是找到在图像中具有相似特征的区域，并计算这些区域之间的视差值。然后，在窗口上选择合适像素，并基于这个像素特征点构建一个特征窗口。这一步骤是基于窗口的匹配，目标是在窗口内找到具有相似特征的像素，并以此为基础建立匹配窗口。目前常用的局部最优的匹配算法包括基于自适应匹配窗口的立体匹配算法[69]和基于自适应像素权重的立体匹配算法[70]等。这些算法的核心是剔除平滑项，采用自适应窗口。剔除平滑项是为了减小计算复杂度，使得算法更适用于实时应用。采用自适应窗口是为了能够更好地适应图像中的不同区域，特别是在目标大小和形状变化较大的情况下。这些算法的优点是计算速度快，适应性强，能够适应各种不同的场景和目标。然而，由于它们只考虑局部信息，可能会忽略一些全局信息，因此在一些复杂的场景中，其性能可能会受到影响。因此，如何在局部最优的立体匹配算法中融入更多的全局信息，以提高匹配的精度和鲁棒性，是当前的研究重点。

2.4 本章小结

本章主要介绍了目标检测与双目定位技术理论基础，首先对目标检测技术进行了详细的介绍，如深度学习网络，卷积神经网络等。其次对双目定位技术和测距过程中的核心原理进行了说明，并简单介绍了相机成像模型和参数标定，详细阐述了如何获得三维空间深度信息，以及从图像中计算获得深度信息、坐标信息的过程，并对双目测距过程进行了完整的数学推导。最后简单介绍了基于双目成像模型的立体匹配研究。

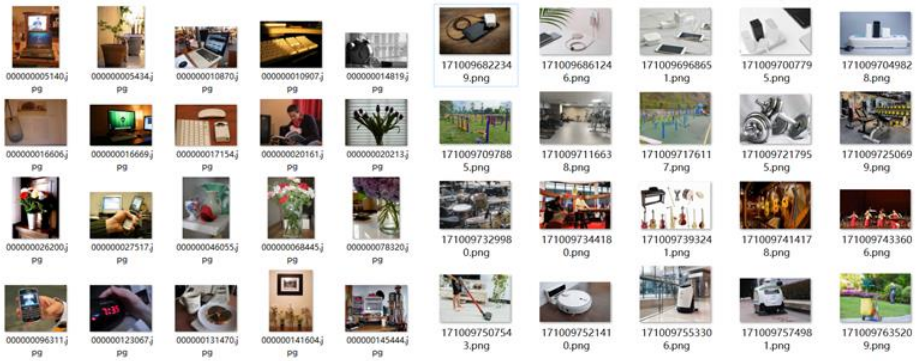
第三章 基于 YOLOv5 的障碍物检测算法模型研究

本章主要介绍 YOLOv5 算法的网络模型架构，并针对当前障碍物检测技术中由于距离摄像头的距离远而导致远距离小目标检测存在的误检和漏检问题，在 YOLOv5 模型中添加注意力机制模块和多尺度特征检测模块，从而提高模型的特征表达能力和检测精度，为后续的精准定位工作打下了坚实的基础。

3.1 模型训练数据集制作

3.1.1 图像数据源

本章的主要目标是识别日常图像中的特定障碍物。虽然开源数据集涵盖了大部分日常物体类别，但对于特定类别的障碍物样本却往往缺乏。因此，本文选择使用 COCO128 数据集，并结合部分自行标注的特定目标样本进行模型训练。COCO128 数据集包含了 118287 张图像，覆盖 80 个不同的类别。然而，随着社会的发展与科技的进步，现在我们日常生活中常见的障碍物已经不再局限于这 80 个类别，例如常见的充电器、健身器材、乐器以及智能清洁工具等。因此，本研究将对这些新出现的目标类别进行自行标注。图 3-1 展示了随机选取的一些训练集样本。



(a) (b)

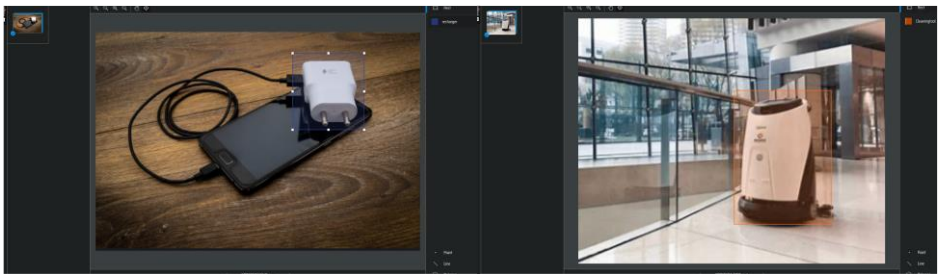
图 3-1 训练数据集示例。(a)COCO 数据集；(b)特定目标类别

3.1.2 制作标签数据

标签的质量对于监督学习模型的训练具有显著的影响。在深度学习领域，卷积神经网络是一种常见的模型，其包含大量的参数。为了优化这些参数并得到性能更优秀的模型，通常需要大量的训练数据。特别是在目标识别任务中，模型的性能很

大程度上依赖于训练数据的质量和数量。因此，确保标签数据的准确性是至关重要的。

为了获取高质量的标签数据，可以采用了人工标注的方式，但这种方式往往需要消耗大量的人力和物力。因此，标注工具的选取尤为重要。目前，有许多样本标注工具，如 Labelme、LabelImg、Make-Sense 等。在选择标注软件时，本文经过对各种软件的比较，发现相比其他的标注软件，Make-Sense 操作简单，并支持多种标注模式，可以多人协同，故选择其为本实验标注工具。标注界面如图 3-2 所示。其中蓝色框和橙色框为标注区域，右边为标签类型。



(a) (b)

图 3-2 Make-sense 软件界面。(a)充电器标注；(b)清洁工具标注

在人工标注样本的过程中，由于样本种类的差异，某些类别的样本数量较多，而有些类别的样本数量则相对较少。此外，样本的质量和数量也存在显著的差异性。这种情况可能会对模型的训练效果产生负面影响，因为模型可能会偏向于更频繁出现的类别，而忽视样本数量较少的类别。为了解决这个问题，我们需要对数据集进行增强处理。

数据增强主要包括对图像进行旋转、翻转、裁剪、缩放等操作，以增加模型的泛化能力。这种方法可以有效地扩大数据集的规模，并改善样本的差异性。此外，数据增强还可以通过引入一定程度的随机性，增加模型对于不同环境和视角变化的适应性。

3.1.3 训练数据增强

深度学习模型的训练过程涉及大量参数的优化，这需要对大量数据进行计算。PerezL[53]等人通过实验证明了数据增强技术在提升模型泛化能力方面的有效性。数据增强是一种在机器学习和深度学习中广泛应用的策略，旨在通过对原始数据进行特定的操作，如旋转、缩放、裁剪等，生成更多的训练样本。这种策略的目的是增加模型接触到的数据多样性，从而帮助模型更好地理解和学习样本间的共有

特征，提高模型的泛化能力。在本研究中，针对 YOLO 模型的特性，采用了三种数据增强技术：线性拉伸变换、空间几何变换以及马赛克数据增强。

（1）线性拉伸变换

线性拉伸变换是一种图像增强技术，它通过改变图像的灰度级别分布来提高图像的对比度，使得图像的细节更加清晰。这种技术对于处理因光照条件不均导致的图像质量问题尤其有效。通过线性拉伸，我们可以扩大图像的灰度范围，增强图像的对比度，使图像中的细节更加明显。在本研究中，若对对拍摄的图像进行灰度范围为[a,b]到[c,d]线性拉伸，公式可以表示为式(3-1)。

$$g(x,y)=\left[\frac{d-c}{b-a}\right]*f(x,y)+c \tag{3-1}$$

（2）空间几何变换

空间几何变换则包括旋转、平移、翻转等操作，它们可以生成不同角度或位置的图像，增加数据的多样性，使得模型能够从多个角度学习目标的特征。空间几何变换是增加训练样本的一个有效手段。

（3）马赛克数据增强

马赛克数据增强技术进一步扩展了 CutMix 概念，通过随机选择四张图像并进行随机缩放、排布和裁剪后进行拼接。这种方法的主要优点在于它可以极大地增加训练样本的数量和多样性，从而提高模型在处理各种类型输入数据时的泛化能力。图 3-3 展示了 Mosaic 数据增强的流程图。

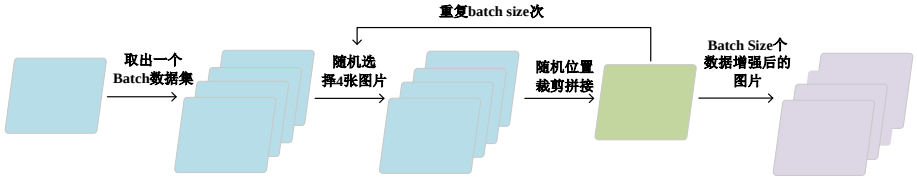


图 3-3 Mosaic 数据增强流程图

3.2 YOLOv5 障碍物检测算法

YOLOv5 是一种高效、准确、简单易用且轻量级的目标检测算法。它在速度和准确性方面都具备优势，适用于各种实际应用场景，如实时目标检测、自动驾驶、工业检测等。本节将详细介绍 YOLOv5 网络的模型架构。

3.2.1 YOLOv5 主干网络

本研究采用了 YOLOv5 网络，该网络的主干部分由 Focus 和 CSP 模块构成。下面将介绍 Focus 和 CSP 模块。

（1）Focus 模块

输入图像在进入主干网络前，由 Focus 模块对其进行切片操作，将分辨率高的图像进行拆分，得到多个低分辨率图像。Focus 模块如图 3-4 所示。

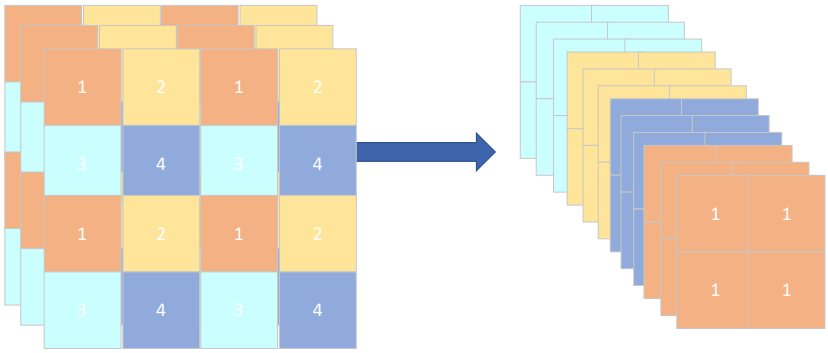


图 3-4 Focus 操作示意图

Focus 模块是一种在卷积神经网络中用于特征提取的有效方法。该模块的优点在于，保持特征图大小不变的情况下，提高特征图的通道数，从而提高模型的表达能力。同时，由于特征图的大小没有增加，因此卷积操作的计算量并没有增加，这有助于加快模型的运算速度。此外，这种方法还可以保留更多的特征信息。Focus 模块通过在通道维度上进行拼接，可以将不同尺度的特征信息融合在一起，从而保留更多的特征信息。最终，经过 Focus 模块处理后的输出结果是一个二倍下采样的特征图。

（2）CSP 结构

在深度神经网络的优化过程中，存在梯度重复计算的问题，这会导致在推理阶段计算量过大，而影响模型的效率。为解决该问题，YOLOv4 引入了 CSP 模块，引入 CSP 模块后，由于特征图在通道维度上的分割和合并，使得网络可以学习到更丰富的特征表示，从而提高模型的性能。

在 YOLOv4 和 YOLOv5 的主干网络中，都采用了 CSP 结构，这是一种旨在解决梯度重复计算问题的网络结构设计。然而，两者在实际应用中有所不同，主要体现在 YOLOv5 设计了两种不同的 CSP 模块，即 CSP1_X 和 CSP2_X。结构如图 3-5 和图 3-6 所示。CSP1_X 模块首先将输入的特征图按照通道维度拆分为两部分。一部分进行常规的卷积操作，以提取特征；另一部分构建残差组件，以增强特征的非线性表达能力。最后，这两部分通过特定的方式合并，得到新的特征图。而 CSP2_X 模块则采用了一种不同的设计，它用卷积层代替了残差组件，从而保留了更多的原始图像信息。这种设计可以在一定程度上提高模型的性能，同时也降低了模型的复杂度。

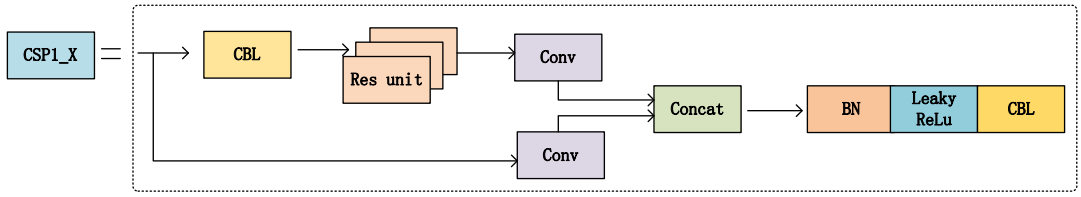


图 3-5 CSP1_X 结构模型图

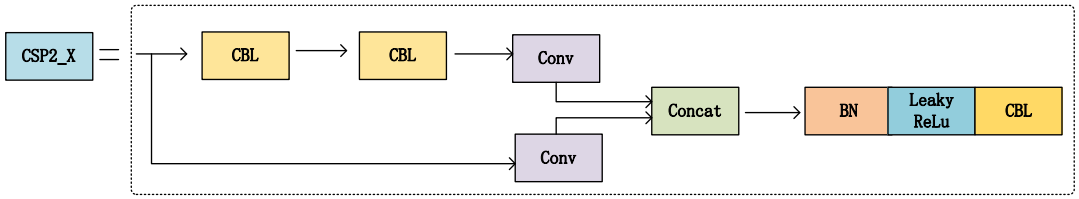


图 3-6 CSP2_X 网络结构图

具体而言，CSP1_X 模块中，特征图的第一部分经过 CBL 模块、残差模块以及卷积运算。CBL 模块用于进行特征提取，残差模块则用于增强特征的非线性表达能力，卷积运算则用于进一步提取特征。而在 CSP2_X 模块中，特征图的第一部分则经过两个 CBL 模块以及卷积运算，这样的设计有助于提取更丰富的特征。对于特征图的第二部分，在两种模块中都仅经过卷积运算。在两部分的运算结束后，将其在对应的通道维度上进行拼接，得到融合的特征结果。这种设计，CSP 结构能够有效地降低模型的计算量，由于特征图在通道维度上的分割和合并，使得网络可以学习到更丰富的特征表示，从而提高模型的性能。

3.2.2 YOLOv5 特征检测网络

YOLOv5 的检测网络由损失函数以及 NMS 非极大值抑制构成，通过 CIoU 损失函数[71]可以减少预测精度的损失，以及通过 NMS[72]非极大值抑制去除重复框。

（1）损失函数

传统的 IoU 计算方式是求真实框与预测框之间的交集。如图 3-7 所示，而当预测框与真实框不存在交集时，此时的 IoU 值为 0，模型则无法计算梯度，从而无法对参数进行优化。

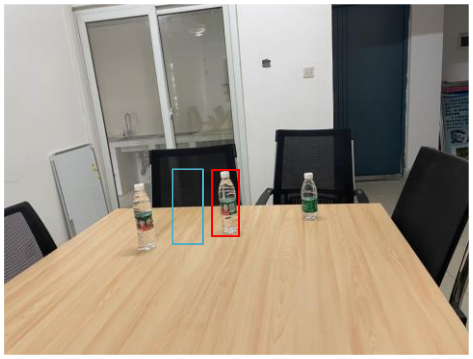


图 3-7 未交集的预测结果与真实标签示意图

YOLOv5 对传统 IoU 计算存在的缺陷进行了改进。在传统的 IoU 计算中，主要考虑的是预测框和目标框的重叠区域，但这种方式忽略了框的中心点位置和长宽比等其他重要信息，可能导致目标检测的精度降低。为了解决这个问题，YOLOv5 采用 CIoU_Loss 损失函数。这种损失函数的设计使得模型在学习过程中能够更好地关注到目标的位置和形状，从而提高了目标检测的精度和速度。CIoU Loss 的计算方式如式(3-2)至(3-4)所示。

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

$$\alpha = \frac{v}{1 - IoU + v} \tag{3-3}$$

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{3-4}$$

其中α参数通过IoU系数选择平衡优先级，v参数防止预测框与真实框的长宽比相同，$\arctan w/h$为边界框对角线的倾斜角度，范围为$0\sim\pi/2$，$(\arctan(w^{gt}/h^{gt}) - \arctan(w/h))^2$范围为$0\sim\pi^2/4$，最后乘以$4/\pi^2$进行归一化。

（2）NMS 非极大值抑制

在目标检测中，网络模型会输出大量的候选检测框，其中一些候选框会包含同一个目标。然而，对于每一个目标，我们最终的目标是得到一个最精确的检测结果，这就需要对这些候选框进行筛选和优化。一种常用的方法是在后处理阶段，利用非极大值抑制 NMS 算法找到所有的局部最大值，并抑制非局部最大值，从而实现候选框的筛选，得到一个最佳的检测结果。

3.3 YOLOv5 网络模型改进

前文对 YOLOv5 网络模型的主干网络和特征检测网络进行了详细的介绍，本节主要介绍 YOLOv5 网络模型的改进方法，通过对卷积注意模块与多尺度特征检

测模块的性能分析，从而得出模型改进的具体方法，改进后的模型提升了特征提取能力以及小目标的识别能力。

3.3.1 注意力机制模块

近年来，注意力机制引起了计算机视觉和深度学习领域研究者的广泛关注。研究表明，引入注意力机制[73]能有效改善网络提取关键信息的能力，提高模型性能。卷积注意力模块 CBAM 是由 S Woo 等人于 2018 年提出的一种的注意力机制，其设计的初衷是为了提高卷积神经网络的表达能力，以此来提高模型的性能。CBAM 模块由通道注意力模块（Channel Attention Module，CAM）和空间注意力模块（Spatial Attention Module，SAM）组成。CAM 模块主要负责对输入特征的各个通道进行权重分配，以此来决定哪些通道的信息更重要；而 SAM 模块则是对输入特征的各个空间位置进行权重分配，以此来决定哪些空间位置的信息更重要。其结构图如图 3-8 所示。

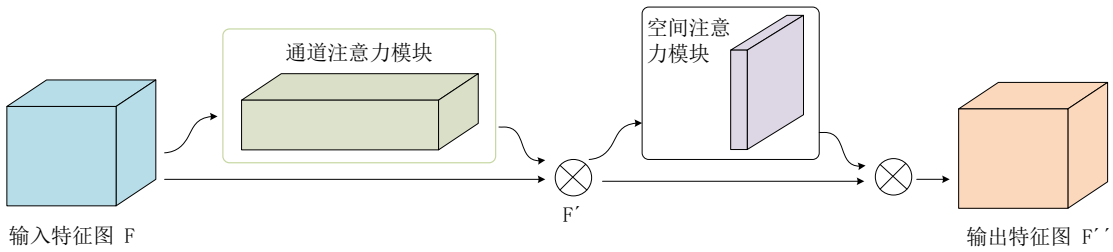


图 3-8 CAM 与 SAM 结构图

在图 3-8 中，卷积注意力模块的处理流程被详细地展示出来。首先，通道注意力模块 CAM 对输入特征图F进行自适应修正，得到修正后的特征图F'。这个过程的主要目标是确定输入特征图F中的哪些通道包含了更重要的信息，然后通过自适应的方式对这些通道进行加权，以提高这些通道在后续处理中的影响力。接下来，空间注意力模块 SAM 对修正后的特征图F'进行进一步矫正，得到最终的特征图F''。这个处理流程可以用公式(3-5)和公式(3-6)来表述：其中，$\otimes$表示元素级别的相乘，输入特征图F的维度为$\mathbb{R}^{C\times H\times W}$，通道注意力特征图$M_c$的维度为$\mathbb{R}^{C\times 1\times 1}$，空间注意力特征图$M_s$的维度为$\mathbb{R}^{1\times H\times W}$。

$$F' = M_c(F) \otimes F \tag{3-5}$$

$$F'' = M_s(F') \otimes F' \tag{3-6}$$

注意力机制的引入，是基于这样一个观察：在处理视觉任务时，人类并不是平等地处理所有的信息，而是会选择性地关注那些对当前任务更重要的信息。卷积注

意力模块 CBAM 设计精巧且高效，其优雅的设计使得它可以无缝地集成到任何卷积神经网络中，并且可以与基础模型进行端到端的训练。此外，CBAM 的计算复杂度极低，因此其引入对模型的计算开销几乎没有影响。因此，这可以显著提升网络的特征处理能力，从而提高模型的性能。

在 YOLOv5 的网络结构中，特征处理的核心部分为 Neck 的结构。其在网络的主干和输出层之间起桥接作用。主干网络的主要任务是从输入图像中提取出一系列的特征，这些特征包括了图像的各种局部和全局信息。然而，这些特征通常是高维度的，并且包含了大量的冗余信息。通过在 Neck 结构中引入各种复杂的操作，如卷积、池化、非线性变换等，可以将主干网络提取的特征进行有效的聚合和处理，从而提取出对目标任务有用的信息。因此，Neck 网络的特征处理能力对于整个网络的性能有着至关重要的影响。

基于上述考量，本文在 YOLOv5s 的 Neck 组件中引入了卷积注意力模块 CBAM。新的网络结构如图 3-9 所示，本节在 Neck 部分的每个 C3 模块之后，以及卷积操作之前，都插入了一个 CBAM 模块，以提高网络对小目标的关注度。

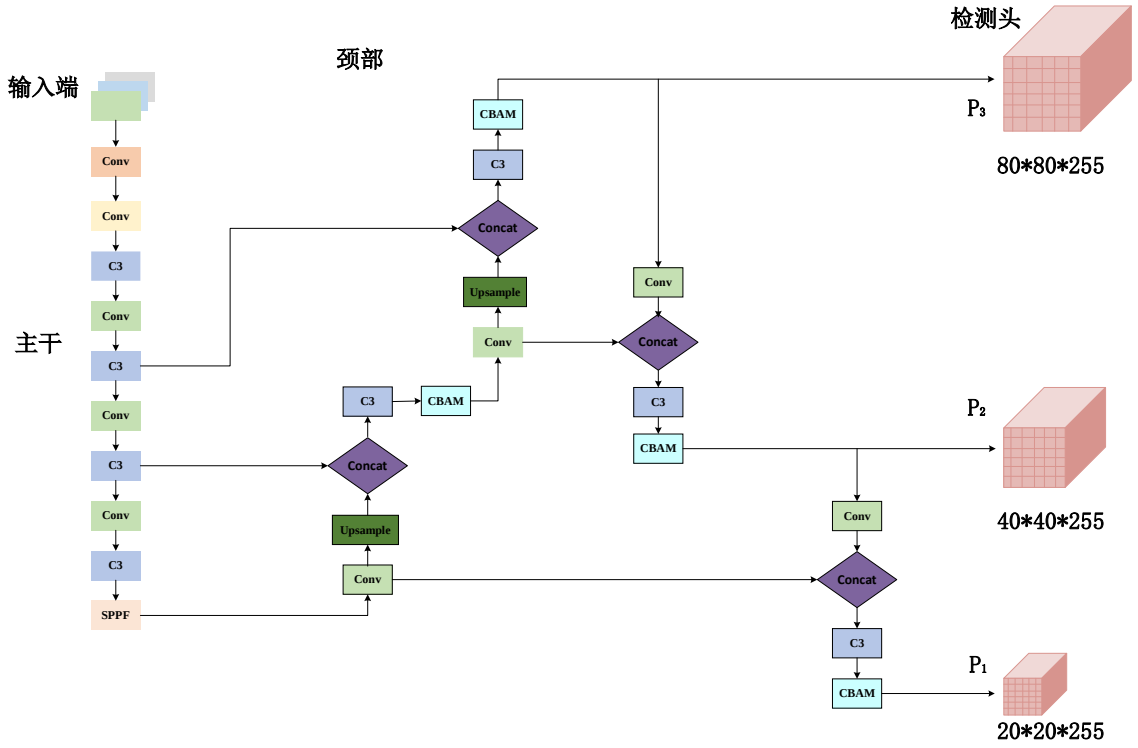


图 3-9 YOLOv5-CBAM 模型结构

如图 3-9 所示，通过在 Neck 部分引入卷积块注意力模块，特征处理流程经历了以下改变：在特征金字塔网络（FPN）结构中，深层特征图通过上采样并与浅层特征图融合，然后经过 C3 模块处理以降低计算量。CBAM 注意力机制用于增强对

关键特征的关注，接着执行卷积和上采样以准备下一次融合。在金字塔注意网络（PAN）结构中，CBAM 同样用于强化对重要特征的关注，并将浅层几何特征融入深层特征。然后，特征图经过下采样并与小分辨率特征图融合。

由此可见，引入 CBAM 模块的改变，使得网络在处理特征的过程中，更加关注于关键的通道和区域，这有助于提高网络的性能。同时，这种改变也使得网络在处理特征时，能够更好地平衡计算量和性能之间的关系。

3.3.2 多尺度特征检测模块

在使用卷积神经网络提取图像特征时，随着卷积层深度的增加，特征图的分辨率降低，而感知范围扩大。深层特征对全局信息的关注度较高，但在小目标检测中效果不佳。相反，浅层特征包含更多位置和细节信息，但仅依赖它们可能导致误检和漏检。因此，研究者提出将浅层和深层特征融合，以在不同尺度的特征图上预测物体位置，这种多尺度检测方式能有效提升算法性能[74]。

目标检测算法 YOLOv3 引入了多尺度目标检测技术以更准确地预测不同大小的目标。这一思想在 YOLOv4 和 YOLOv5 中得以继承并优化。在 YOLOv5 中，对于小目标，8 倍下采样可能会导致位置和细节信息的丢失。此外，选取的实验数据集 COCO128 中包含了大量的小目标，这对模型在学习分辨率小于 8×8 的目标特征信息时带来了挑战，也是 YOLOv5 在该数据集上表现不尽如人意的一个重要原因。

基于此，本节对 YOLOv5 的检测头进行改进，增加一个可用于检测更小分辨率的微小目标检测头，加上上节所改进的 CBAM 模块，最终改进后的模型结构如图 3-10 所示。

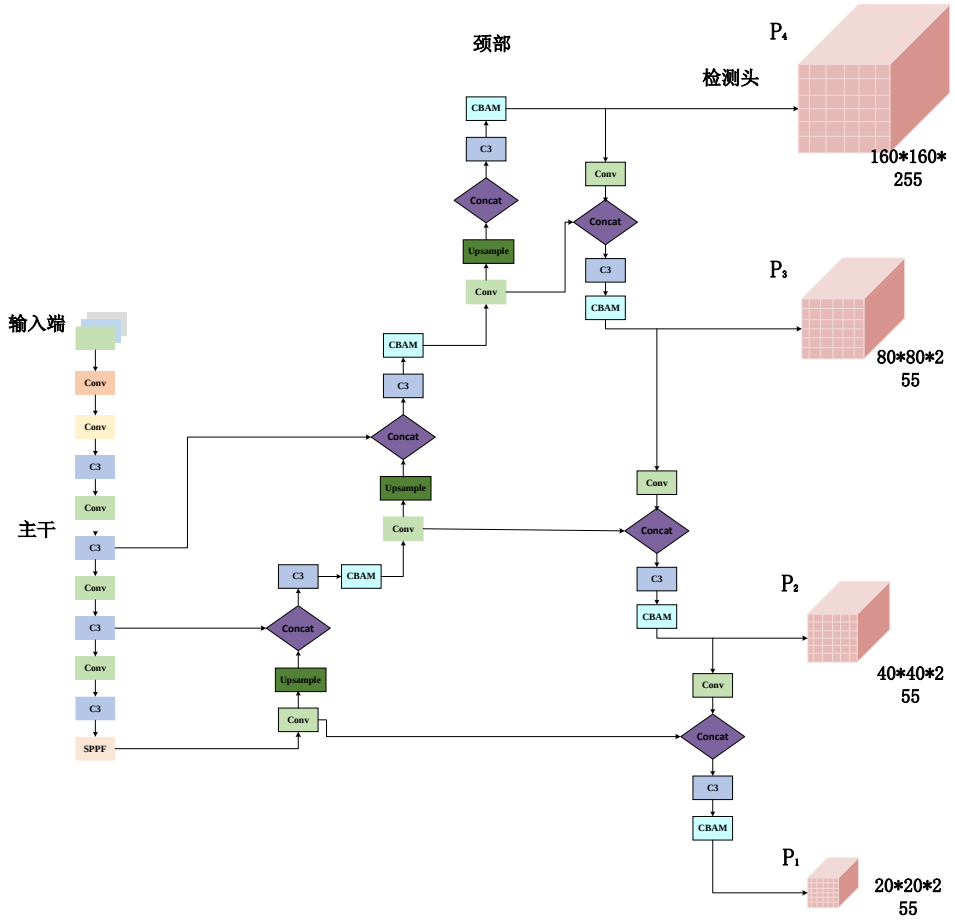


图 3-10 YOLOv5-CBAM+SmallTarget 模型结构

由于增加了一个检测头，原模型的 Neck 网络结构也相应发生了变化。变化后的 YOLOv5-SmallTarget 模型结构如图 3-11 所示。

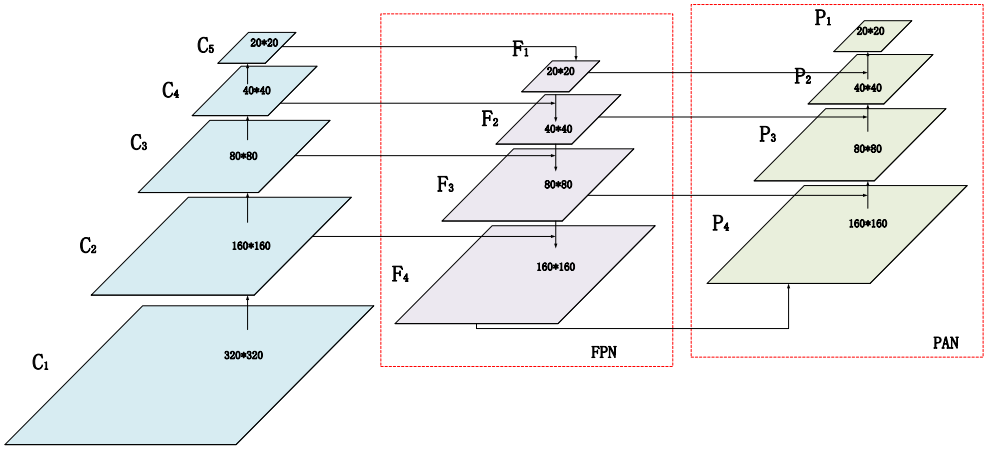


图 3-11 YOLOv5-SmallTarget 模型特征融合过程图

从图 3-11 的特征融合结构图中可见，主干网络结构并未改变，与基准模型一致，依次经过 5 次下采样操作得到 5 个不同尺度的特征图，分别是 C1、C2、C3、C4 和 C5，其尺寸从下到上逐步减小。然后在 Neck 网络中，为了将深层特征图中的语义信息传递到浅层，F1 被不断上采样并与主干网络中对应尺寸的特征图进行特征融合，从而得到 4 个尺度的特征图，分别是 F1、F2、F3、F4，其尺寸从上到下逐步增大。之后，为了将浅层特征图中的位置信息传递到深层，P4 被不断下采样，并与 FPN 结构中对应尺寸的特征图进行特征融合，得到 4 个不同尺度的特征图，分别是 P4、P3、P2 和 P1，随着特征图尺寸的不断减小，其感受野不断增大。P4 是新增的用于检测微小目标的特征头，其尺寸为 16×160×255，感受野为 4×4。值得注意的是，P4 融合了来自主干网络中特征图 C2 的信息，由于 C2 更接近输入原图，因此包含的位置信息更丰富，这对小目标的定位非常有利，使得网络对小目标更敏感，从而提升模型对小目标的检测效果。并且，由于 P3、P2 和 P1 是通过 P4 下采样得到的，因此它们也间接获得了 C2 中的位置特征信息，这提升了 YOLOv5s 原有的三个检测层对目标的定位能力。

综上所述，通过在基准模型 YOLOv5s 中添加一个用于检测微小目标的检测头，一方面提升模型对小目标的检测效果；另一方面，使得其它检测层的特征图能够融合更多浅层的特征。

3.4 实验结果与分析

3.4.1 实验设置

本实验采用 Pytorch 框架来构建优化后的 YOLOv5 网络模型，数据集采用 3.1 节中的数据集。考虑到任务需求，选取 COCO128 中常见的障碍物进行训练，主要选取的障碍物有 10 类，分别为花瓶，背包，书本，水瓶，纸杯等，划分训练集 5856 张图片，测试集 953 张图片。参数设定如下：预训练模型采用 YOLOv5 提供的 yolov5s.pt，输入图像的尺寸设定为 640，训练的总轮次（Epoch）设定为 200，同时使用 8 个线程进行计算。在优化器的选择上，本研究采用了随机梯度下降（SGD）[75]算法。在学习率的更新策略上，本实验采用了余弦退火策略[76]，初始学习率设定为 0.01。此外，每个批次（batch）的数据大小设定为 4。在损失函数的选择上，本实验采用 CIoU 作为目标函数。在数据增强方面，除了常规的图像翻转、旋转、HSV 增强外，还引入了马赛克数据的数据增强方式。

在测试阶段，图像输入尺寸设定为 640。模型的置信阈值被设定为 0.25，非极大值抑制（NMS）的 IoU 阈值为 0.45，最大检测对象数为 1000，每个批次（batch）的数据大小设定为 1。本实验的软硬件环境如表 3-1 所示。

表 3-1 实验环境配置

名称	信息
CPU	Inter（R）core i5-12500H
GPU	GeForce RTX 3050
内存	16G
显存	4G
操作系统	Windows11
Python	3.9.12
Pytorch	1.11.0
CUDA	11.3
CUDNN	8.2.0

3.4.2 实验结果评价指标

评估目标检测算法性能的主要标准通常包括分类精度和边界框回归。分类精度主要是对算法对目标类别判断的准确性进行评估，而边界框回归则是对算法预测目标位置的准确性进行评估。具位置的判断则是基于预测框与真实框的重叠度来进行的。这种重叠度通常被称为交并比（IoU），其计算方式如公式(3-7)所示。在这个公式中，area(D)代表预测框的面积，area(G)而代表真实框的面积。

$$IoU = \frac{area(D) \cap area(G)}{area(D) \cup area(G)} \tag{3-7}$$

混淆矩阵（Confusion Matrix，CM）是一种常用的工具，用于评估和分析分类模型的性能。矩阵行代表真实类别，列代表预测类别。对角线上的值越大，表明模型分类正确的概率越高，反映出较好的分类效果。在目标检测中，真阳性（TP）是预测框与真实框（GT）IoU 大于阈值且数量最大的情况。假阳性（FP）是非 TP 的预测框数量，而假阴性（FN）是未被模型检测到的真实目标数量。

在目标检测任务中，常用于衡量算法性能的评估指标如下所示：

（1）准确率（Accuracy）：Accuracy 是一个全局性的指标，它评估的是模型对所有类别的预测正确性，如公式(3-8)所示。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3-8}$$

（2）精确率（Precision）：Precision 是针对每个类别的预测结果进行评估，衡量的是正确预测为正例的样本占模型预测为正例样本的比例。计算方式如公式(3-9)所示。

$$Precsion = \frac{TP}{TP + FP} \tag{3-9}$$

（3）召回率（Recall）：Recall 是针对每个类别进行评估，它衡量的是被模型正确预测为正例的样本占所有真实为正例的样本的比例。计算方式如公式(3-10)所示。

$$Recall = \frac{TP}{TP + FN} \tag{3-10}$$

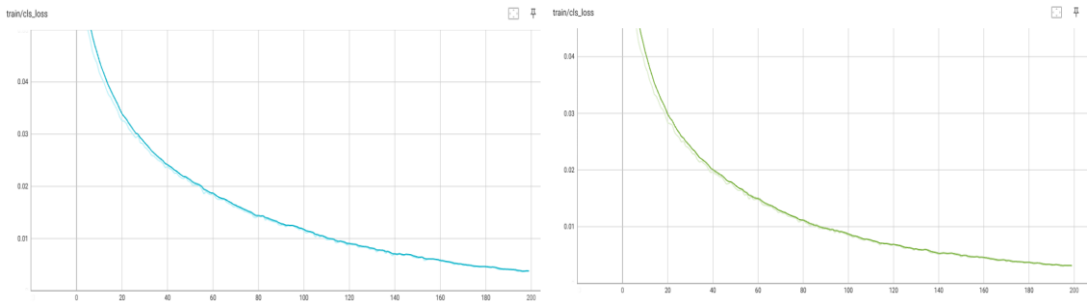
（4）平均精度（AP）：AP 是在不同召回率水平下精确率的平均值，它能够更全面地反映模型的性能。

（5）平均精度均值（mAP）：mAP 则是所有类别的 AP 的平均值，它是一个全局性的指标，能够反映模型对所有类别的整体预测性能。

精确率（Precision）和召回率（Recall）是一对经常被用于评估分类模型性能的指标，然而，它们都存在单点值的局限性，即只能反映模型在特定阈值下的性能，无法全面评估模型的性能。因此，为了得到一个更全面的性能指标，引入了平均精度 AP 和平均精度均值 mAP。AP 是通过计算精确率和召回率构成的 P-R 曲线下的面积来得到的，其中，精确率作为纵坐标，召回率作为横坐标。这个面积的大小反映了模型在不同召回率水平下的平均精确率，因此，AP 能够更全面地评估模型的性能。mAP 则是在整个数据集中，对所有类别的 AP 值进行平均得到的。它是一个全局性的指标，能够反映模型对所有类别的整体预测性能。在计算 AP 和 mAP 时，通常将 IoU 的阈值设置为 0.5，即只有当预测框与真实框的 IoU 大于 0.5 时，该预测框才被认为是有效的。除此之外，目标识别中还用 FPS（Frame Per Second，每秒帧率）来衡量算法的运算速度。

3.4.3 结果分析与评价

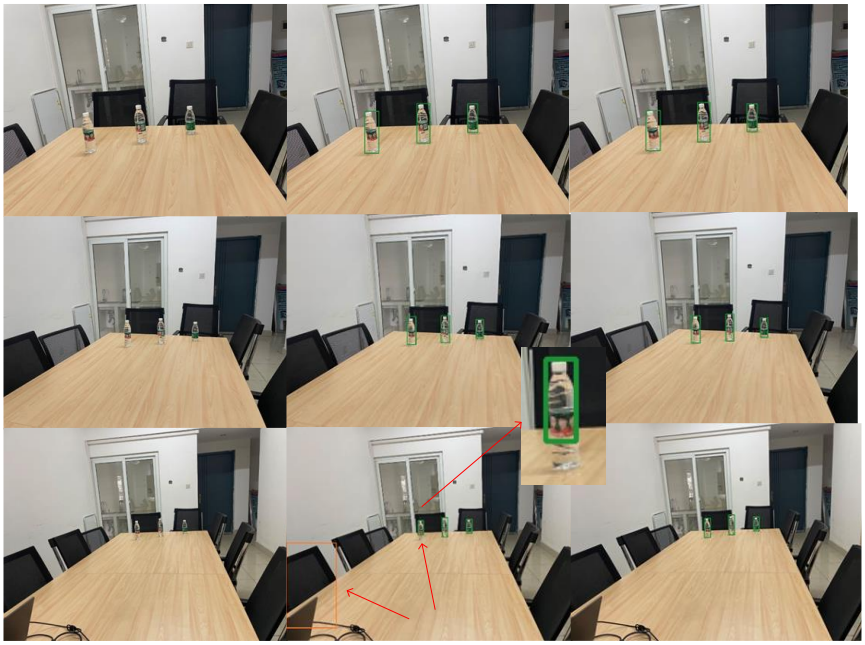
为了验证本本提出的改进算法，分别将测试集数据在训练好的模型进行测试，如图 3-12 所示，经过 200 轮的 Epoch 训练后，Loss 曲线成功收敛。图 3-12 中(a)为 YOLOv5 模型，(b)为改进后 YOLOv5 模型。



(a) (b)

图 3-12 模型训练过程 loss 变化

在小目标中，本实验选择常见的水瓶和水杯作为测试目标，下面展示在这两种类别下的实验结果。在水瓶目标上，基于 YOLOv5 与本文改进后的 YOLOv5 模型识别效果如图 3-13 所示。

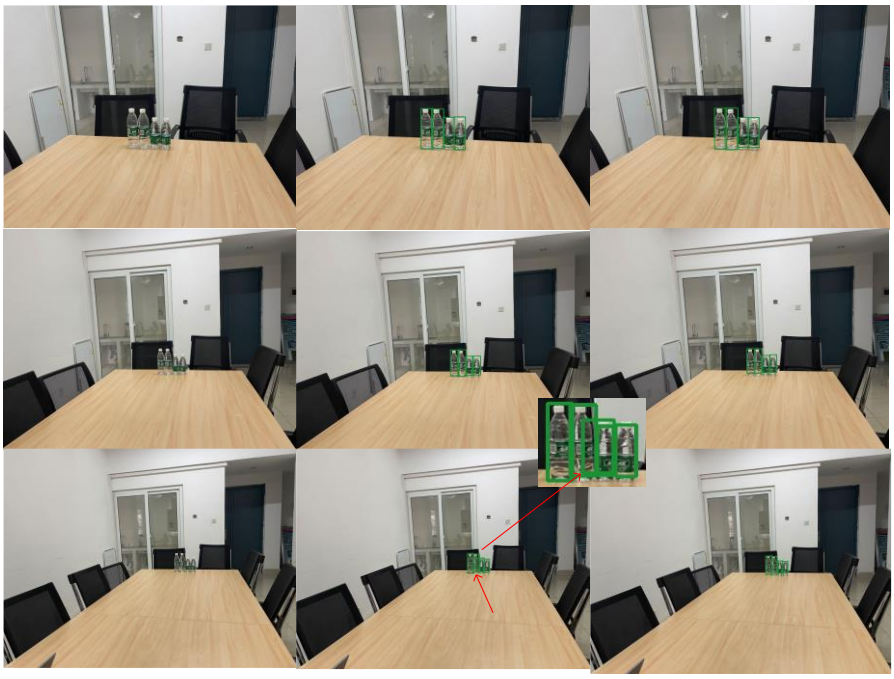


(a) (b) (c)

图 3-13 水瓶目标识别结果图。（a）测试图像；（b）yolov5 实验结果；（c）改进后 yolov5 实验结果

图 3-13 展示了两种模型的目标检测结果。尽管两种模型在这三幅图的识别中都没有出现漏识别的现象，但从图 3-13(b)和图 3-13(c)中可以明显看出，原始的 YOLOv5 模型存在误识别的情况，且在小目标的识别中出现了检测框未能完全包围住物体的情况。相比之下，改进后的 YOLOv5 模型并未出现这种情况，表明其在小目标检测方面有所优化。

在不同距离障碍物检测实验的基础上，将障碍物密集化，对两种模型做出对比，实验结果如图 3-14 所示。



(a) (b) (c)

图 3-14 密集水瓶目标识别结果图。（a）测试图像；（b）yolov5 实验结果；（c）改进后 yolov5 实验结果

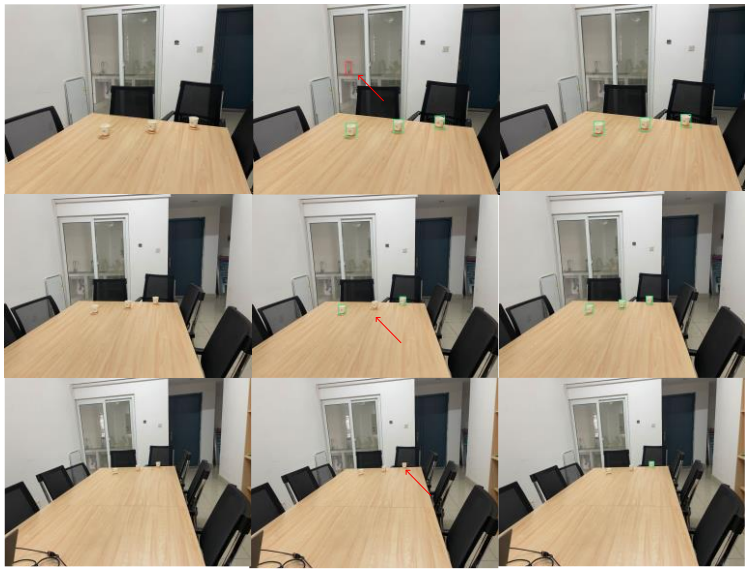
从图 3-14 的测试结果中可以看出，未改进的 YOLOv5 模型在小目标靠近的情况下存在检测框重叠的现象，而改进后的 YOLOv5 模型则不存在这种情况。从上面的两个实验可以明显看出改进后的 YOLOv5 模型均优于未改进的 YOLO 模型，除此之外本节对两种模型进行定量的评价，依照 3.4.2 节计算性能指标，计算结果如表 3-2 所示。

表 3-2 水瓶目标评价指标

Model	Size	Precision	Recall	mAP	FPS
YOLOv5	640	56.9%	60.3%	58.4%	32.5
SCBAM_YOLOv5	640	63.9%	59.3%	62.0%	31.2

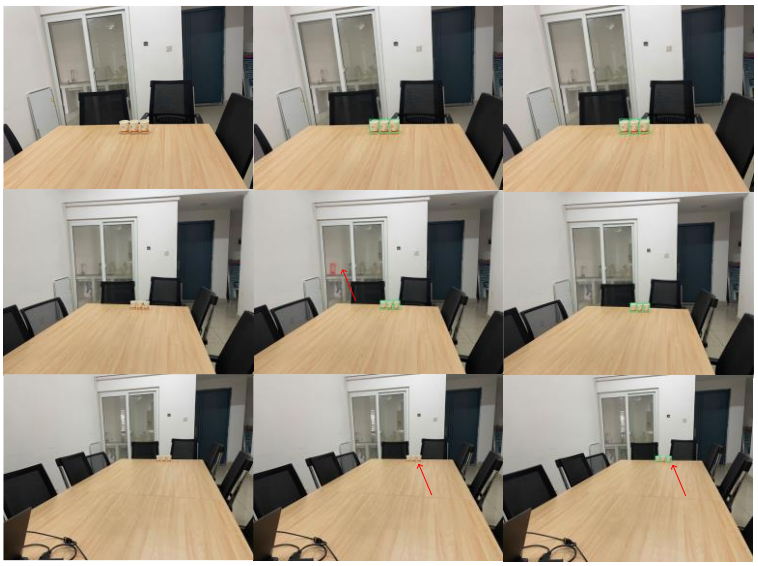
由表 3-2 可知，SCBAM_YOLOv5 模型比原 YOLOv5 模型相对于水瓶目标精确率提高了 7%，召回率降低了 1%，mAP 值提升了 3.6%。尽管模型的速度稍微降低了 1.3FPS，但仍能达到工程要求。

针对水杯种类障碍物，本节也做了同样的实验用以验证，图 3-15 和图 3-16 为对水杯目标的识别效果图：



(a) (b) (c)

图 3-15 水杯目标识别结果图。（a）测试图像；（b）yolov5 实验结果；（c）改进后 yolov5 实验结



(a) (b) (c)

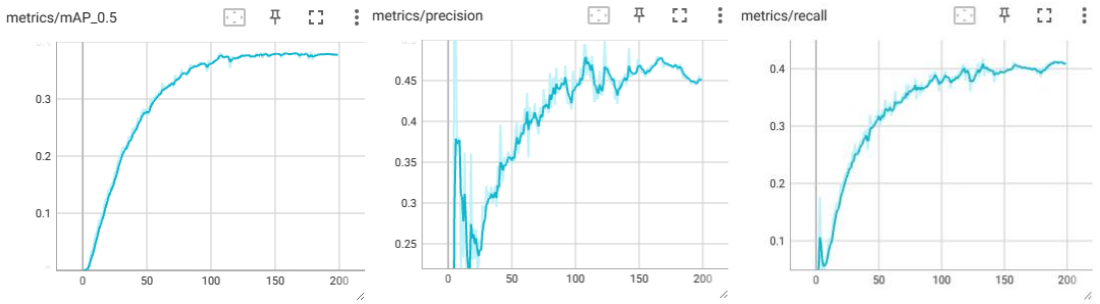
图 3-16 密集水杯目标识别结果图。（a）测试图像；（b）yolov5 实验结果；（c）改进后 yolov5 实验结果

由于水杯相对水瓶体积较小，其识别难度增大，实验结果显示存在识别不全、误识别及遗漏识别的情况。但同样可以看出改进后的 YOLOv5 模型的识别效果仍然优于未改进的 YOLOv5 模型，对水杯识别效果进行定量分析，结果如表 3-3 所示。

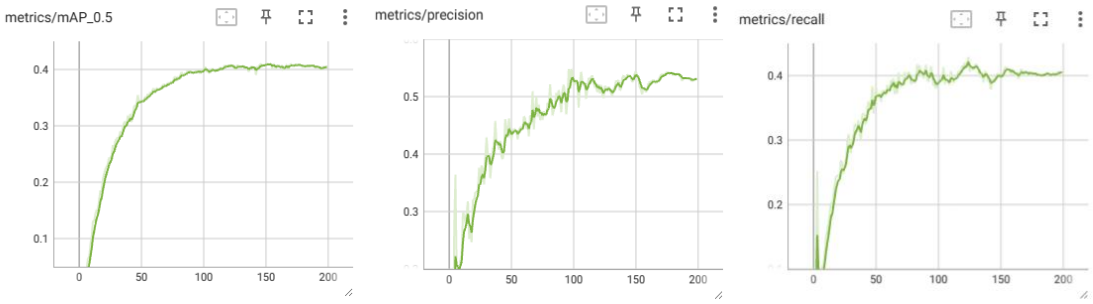
表 3-3 水杯目标评价指标

Model	Size	Precision	Recall	mAP	FPS
YOLOv5	640	52.9%	43.2%	44.4%	32.5
SCBAM_YOLOv5	640	63.2%	40.6%	44.8%	31.2

通过定量分析，SCBAM_YOLOv5 模型比 YOLOv5 模型对于水杯目标精确率提高了 10.3%，召回率降低了 2.6%，mAP 值提升了 0.4%。图 3-17 为 YOLOv5 模型与 SCBAM_YOLOv5 模型的实验结果对比图。



(a)



(b)

图 3-17 两种模型的实验结果对比图

从实验结果来看，通过对 YOLOv5 网络模型中的检测 Neck 模块进行改进，并且添加 CBAM 以及多尺度特征检测模块，可以进一步提升模型的性能。具体来说，这些改进使得网络能够提取到更多待识别目标，增强了对小目标识别的精确性，目标未识别、识别错误的现象明显减少，网络鲁棒性明显提升。

3.5 本章小结

本章对 YOLOv5 目标检测算法对远距离的小目标存在误检、漏检的问题进行了分析，针对这些问题本文在 YOLOv5 的结构基础上提出了两个改进方案。首先

在模型中添加了注意力机制，有效改善了网络提取关键信息的能力；其次添加了多尺度特征检测模块，改善了对远距离较小目标的识别效率，极大降低了漏检和误检的情况。实验表明，基于 YOLOv5 改进的 SCBAM_YOLOv5 模型显著提升了目标识别准确性，总体精度增长 4.9%，召回率微降 0.26%，mAP 值提升 1.55%。

第四章 基于双目视觉传感器的障碍物定位算法研究

结合前述模型改进，本章将采用双目相机实现障碍物定位，首先利用张正友标定法获取相机内外参数。并实现对双目图像的畸变和极线矫正，实验过程中，本章对在标定中所出现的各类问题做出总结并给出解决方法，提出两段式标定分析法以提高标定效率，获取更好的标定结果。其次为获得障碍物的三维坐标位置，本文以使用双目摄像机中左侧摄像头的相机坐标系作为世界坐标系，并对双目图像进行立体匹配，从而得到准确的三维坐标值。立体匹配是双目摄像机障碍物定位研究中的关键步骤，因此，选择一个既快速又准确的立体匹配算法是至关重要的。在移动机器人的应用场景中，尽管传统的 SURF 和 SIFT 等立体匹配算法被广泛应用，但由于其运算速度较慢且匹配精度不够高，因此无法满足实际需求。为此，本文通过实验对比了一些常用的匹配算法，并最终选择了半全局立体匹配 SGBM 算法进行图像立体匹配。最后，基于 SGBM 算法的一些缺点，如对光照敏感等问题，对该匹配算法进行了优化设计，并且给出静态测距的数据验证。通过本章的研究，不仅提高了相机标定的工作效率，改善了 SGBM 算法的鲁棒性，而且提高了定位和测距的精度。

4.1 双目视觉传感器系统

双目立体视觉系统由硬件和软件系统两部分组成。硬件系统包括双目立体相机和个人笔记本电脑。软件系统则具备图像采集、图像处理、障碍物检测识别以及障碍物定位测距等多项功能。在本文中，所使用的硬件平台是由一台型号为汇博视捷 1080P 的双目相机（如图 4-1 所示）和一台笔记本电脑组成。在软件方面，本文主要使用 Pycharm 作为开发环境，Python 作为编程语言，并搭配 pytorch 深度学习框架进行障碍物检测模型的训练。此外，本文还使用 Matlab 进行图像处理，其内部集成了丰富的图片处理函数以及专门用于双目标定的工具箱。双目摄像头部分功能参数如表 4-1 所示。

图 4-1 双目立体相机

表 4-1 双目摄像头部分功能参数

主要参数	指标
产品型号	HBVCAM-4M2214HD-2 V11
感光芯片	OV4689 CMOS
基　　线	60mm
视场角/焦距	80°/焦距 3mm
格　　式	MJPEG
分辨率	MJPEG 3840×1080 30FPS
通信/供电	USB2.0
支持系统	Windows/Linux/MAC OS/Android
测距范围	40cm-400cm
工作温度	-20℃-70℃

4.2 双目视觉传感器的标定与矫正

本节以双目视觉传感器为研究对象，首先通过对视觉传感器的标定，获取其内外参数，并对拍摄的图片进行矫正，从物理意义上分析了标定参数，然后在多次标定过程中，总结了相应的注意事项及解决方法，最后针对标定过程繁琐等问题，提出了两段式标定分析法，提高了标定效率。

4.2.1 双目视觉传感器的标定

本节采取张正友标定法[78]对双目像机进行标定。Matlab 的 Stereo Camera Calibrator 基于 Zhang 标定法，通过双目摄像头拍摄不同角度的棋盘格图像进行标定，图像导入指定路径后，选择适当的标定方式进行操作。在标定过程结束后，工具箱会输出摄像头的内外参数。内参数描述了摄像头本身的特性，如焦距和镜头畸变等。外参数则包括旋转与平移矩阵，这些参数描述了摄像头相对于世界坐标系的位置和姿态。图 4-1 展示了实验中所选取的双目摄像头。

本文双目摄像头的具体标定步骤如下[78]：

（1）制作棋盘格标定板

如图 4-2，本节使用 6×9 布局，格子尺寸 21mm×21mm 的标准棋盘格标定板进行双目摄像头标定，棋盘格标定板的图像质量要求清晰，标定板平面必须平整。为了满足这些要求，本研究选择使用平板电脑来显示和使用标定棋盘格，这样可以确保在物理意义上的平整性，从而提高标定的精度和可靠性。

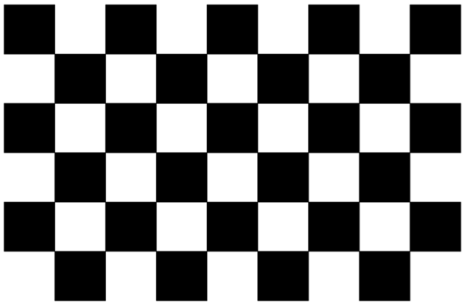
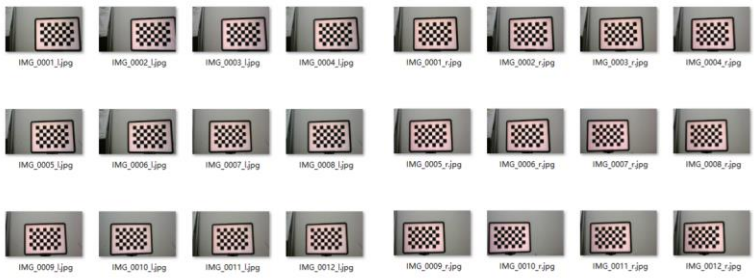


图 4-2 标准棋盘格标定示意图

（2）双目视觉传感器采集棋盘格图像

在标定过程中，本实验选择固定标定板的位置，而移动双目摄像头在标定板前变换角度和位置。这种方法可以确保从多个不同的视角和距离捕获到标定板的图像，从而提高标定的精度。在拍摄过程中，尽量使标定板图像占据整个图像的 1/3 到 2/3，为了准确地标定相机的畸变系数，需要在相机视图中均匀采集标定图像。这是因为摄像头的边缘区域通常会有更大的畸变，因此需要在这些区域也采集到足够的标定图像。在相机标定时，至少需要采集 5 对左右图像对。然而，为了提高标定的精度和可靠性，本研究选择采集 20 对符合要求的棋盘格双目图像进行实验。这些图像包括了各种不同的角度和位置，可以提供更全面和准确的标定信息。部分图像如图 4-3 所示。



(a) (b)

图 4-3 棋盘格双目图像。(a)左相机图像采集结果；(b)右相机图像采集结果

（3）提取角点

将摄像头拍摄的 40 张棋盘格图像导入到 Matlab 中，利用其提供的 Stereo Camera Calibrator 工具箱进行处理，功能包括自动识别和标记出图像中的角点等。这些角点是棋盘格图像的关键特征点，对于后续的标定过程至关重要。在这个过程中，需要关注的一个关键参数是角点标定的不确定性误差。一般来说，如果这个误差可以控制在 1%以内，那么就可以认为标定的结果是良好的。检测结果如图 4-4 所示。

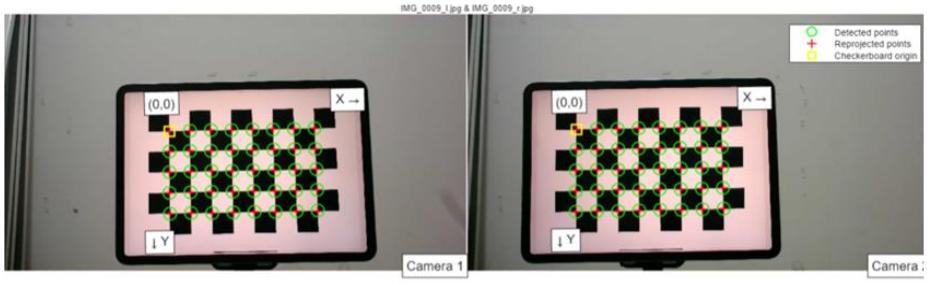


图 4-4 角点检测图像

（4）双目视觉传感器标定

在本节的相机标定实验中，选取 20 对图片进行标定。在标定过程中剔除了 7 对效果较差的图片，标定的重投影误差为 0.10pixels，如图 4-5 所示。

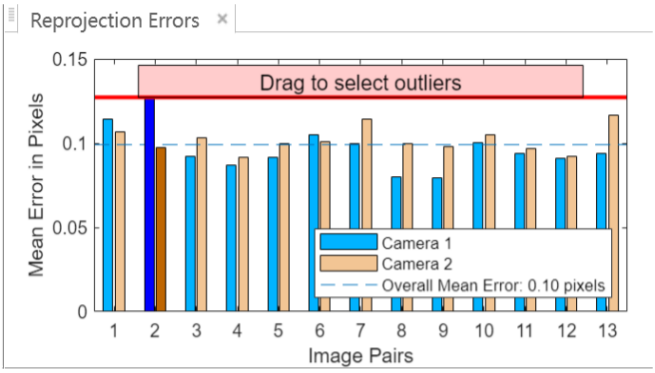


图 4-5 相机重投影误差分析

标定结果如表 4-2 和表 4-3 所示。可以看出，旋转矩阵与单位矩阵近似，从平移向量中第一个参数得到左、右摄像头之间的水平距离 59.61mm，近似于相机的基线距离 60mm，并且重投影误差在 0.1 个像素以内，认为标定的结果是良好的。

表 4-2 双目摄像头内参标定结果表

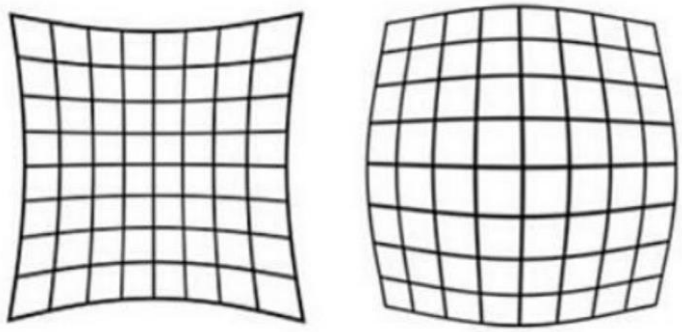
标定参数	左摄像头	右摄像头
内参矩阵	$\begin{bmatrix} 1538.7309 & -0.0399 & 892.5084 \\ 0 & 1538.2709 & 582.9248 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1531.1903 & -0.6140 & 926.7897 \\ 0 & 1530.2196 & 582.3756 \\ 0 & 0 & 1 \end{bmatrix}$
畸变系数	$\begin{bmatrix} -0.0438 \\ 0.1238 \\ 0.0006 \\ -0.0007 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0.0481 \\ 0.1367 \\ 0.0003 \\ -0.0002 \\ 0 \end{bmatrix}$

表 4-3 双目摄像头外参标定结果表

标定参数	外参参数
旋转矩阵	$\begin{bmatrix} 0.9999 & -0.0005 & -0.0129 \\ 0.0005 & 0.9999 & -0.0049 \\ 0.0129 & 0.0049 & 0.9999 \end{bmatrix}$
平移向量	$[-59.6114 \quad 0.1518 \quad -1.5300]$

4.2.2 图像畸变校正

在相机的生产过程中，透镜与成像平面之间的差异会影响光线的传播，从而影响最终的成像效果。这种影响通常表现为图像的畸变，即实际物体在成像后所呈现出的失真和变形。畸变主要分为两类：径向畸变和切向畸变。径向畸变是由于镜头镜片的放大率和形状差异引起的。另一类是切向畸变，这种畸变通常是由透镜的光轴倾斜所引起，这种畸变会使图像看起来像被拉向一个方向，从而造成图像的扭曲。为了更直观地展示这两种畸变的效果，在图 4-6 中给出了示例。



(a) (b)

图 4-6 畸变示意图。(a)枕型畸变图；(b)桶型畸变图

对于切向畸变，平面上点P可以表示为$[x, y]^T$或极坐标$[r, \theta]^T$。其中r和θ分别表示点P到原点的距离和水平夹角。如公式(4-1)，对径向畸变进行消除。

$$\begin{cases} x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \tag{4-1}$$

其中，特征点P在归一化平面上的坐标$[x, y]^T$矫正后为$[x_{\text{corrected}}, y_{\text{corrected}}]$。图像校正主要用$k_1$消除边缘畸变较小的问题，用$k_2$矫正边缘畸变较大的问题，$k_3$常用于消除鱼眼相机等大畸变镜头。

对于切向畸变，如公式(4-2)，对畸变进行矫正。

$$\begin{cases} x_{corrected} = x + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{corrected} = y + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases} \tag{4-2}$$

根据畸变参数k_1，k_2，k_3，p_1和p_2，计算点P在像素平面的坐标，并将其三维空间坐标投影到图像平面。归一化坐标设为$[x, y]^T$，对其进行径向和切向畸变消除，如公式(4-3)所示。

$$\begin{cases} x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases} \tag{4-3}$$

将由公式(4-3)矫正后的坐标$(x_{\text{corrected}}, y_{corrected})$乘以x、$y$方向上的像素尺寸，得到像素平面的投影点(u, v)。如公式(4-4)所示。

$$\begin{cases} u = f_x x_{\text{corrected}} + c_x \\ v = f_y y_{\text{corrected}} + c_y \end{cases} \tag{4-4}$$

4.2.3 极线矫正

基于第二章的理论，正交和平移矩阵描述了双目相机镜头的位姿关系，即相机的相对位置和朝向，这是双目视觉系统的关键参数。基于此，我们进行极线校正，也称立体校正，采用 Bouguet 方法，使相机光轴平行。该方法首先获取双目相机内参矩阵，旋转使成像平面平行，然后通过矫正矩阵将非共面的左右相机坐标系矫正为平行且共面。

通过对左右相机分别绕轴旋转特定角度，我们得到新的旋转矩阵，如公式(4-5)所示。

$$\begin{cases} R_l = R^{1/2} \\ R_r = R^{-1/2} \end{cases} \tag{4-5}$$

构造矫正矩阵R_{rect}使得基线与成像平面平行，矫正矩阵公式如(4-6)所示。

$$R_{rect} = \begin{bmatrix} (e_1)^T \\ (e_2)^T \\ (e_3)^T \end{bmatrix} \tag{4-6}$$

其中，$e_1 = T/\| T \|, e_2 = [-T_y \quad T_x \quad 0]/\sqrt{T_x^2 + T_y^2}, e_3 = e_1 \times e_2$,$T$为左右镜头间的偏移矩阵，$\mathrm{T} = [T_x \quad T_y \quad T_z]^T$。将构造的矫正矩阵与左右相机的新旋转矩阵相乘，得到双目相机的整体旋转矩阵R_l'和R_r'。如公式(4-7)所示，从而实现极线矫正。

$$\begin{cases} R_l' = R_{\mathrm{rect}} \times R_l \\ R_r' = R_{\mathrm{rect}} \times R_r \end{cases} \tag{4-7}$$

4.2.4 实验结果与分析

1．双目视觉传感器的标定

针对双目视觉传感器的标定，本节先后标定 30 余次，将在标定过程中存在问题及注意事项总结如下，并给出两段式标定分析法，显著提高标定效率。

双目摄像头标定的总结：

（1）标定板的选择。对于高精度的测距，需向标定板制作厂家定制符合要求的标定板。本文在已知的双目摄像头的硬件条件下，要求 4m 的测量距离下，误差不大于 1%，在本节实验过程中，使用 11 寸平板电脑标定板，标定板的选择上要求平整光滑。

（2）拍摄环境的选择。在对双目相机的标定过程中，环境光以及标定板自身的反光程度都对标定结果有影响。尽量选择在温和均匀光线条件下的环境中进行标定，除此之外，在用平板电脑作为标定板的过程中，调节平板电脑的屏幕亮度，使拍摄的照片更贴近真实场景。

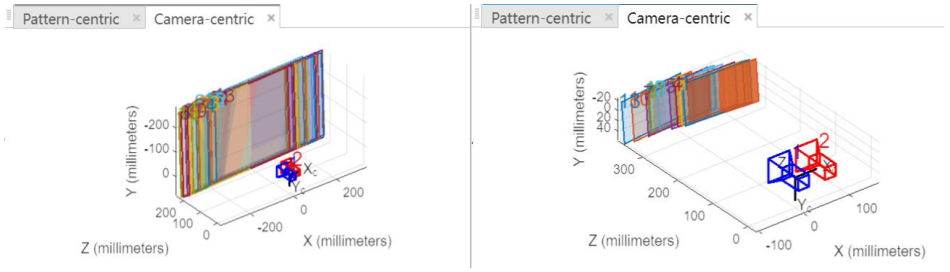
（3）拍摄方式的选择。总体而言，拍摄方式分为两种，一种是固定相机，移动标定板，另一种是固定标定板，移动相机。两种方式均可，但在拍摄过程中，标定板与摄像头的夹角不能超过 30 度，并且标定板在图幅中占据 1/3 及以上的区域，均匀分布在图片的各个位置。由于需要满足在图幅中占据相应的比例，故拍摄距离的范围也从而确定。

（4）标定结果的精确度分析。拍摄照片过程中需要 10-20 组照片，将其导入 matlab 的标定工具箱中，选择合适的参数后，进行标定，理论上讲，当标定的重投影误差小于 0.1 个像素时，认为标定的结果良好。

在标定结束后，需将得到的内外参数作为输入给到双目定位的代码中，进而得到视差图，通过深度图分析具体物体的定位精度。相机的准确标定结果需要进行多次标定才能确定，若每次标定后都需要对测距精度进行实时分析，效率非常低下，本节通过对标定参数的分析，提出两段式标定结果分析，以提高标定效率。

（1）使用 matlab 标定工具箱过程中

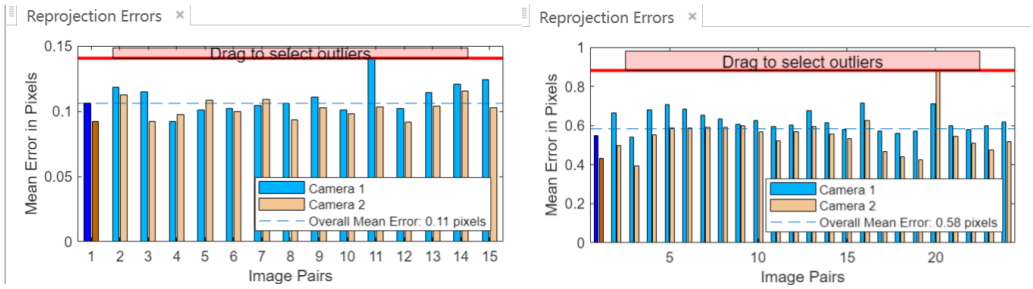
首先载入拍摄好的图片，选择合适的标定参数，可以从 Camera-centric 或 Pattern-centric 的窗口中初步判断后续生成相机内外参数的好坏，如图 4-7 所示。



(a) (b)

图 4-7 相机与标定板的位置关系。(a)存在问题的位置关系图；(b)良好的位置关系图

其次判断图片的重投影误差是否满足 0.1 个像素左右，可以对误差较大的照片进行剔除，以满足要求，否则重新拍摄图片，再一次进行标定，如图 4-8 所示。



(a) (b)

图 4-8 相机重投影误差分析。(a)较好的重投影误差图；(b)较差的重投影误差图

从图 4-7 中可以看出拍摄时的三维视图，在 Z 轴上近似表示拍摄图片时相机与标定板之间的距离，因此可以初步对后续的参数输出做一个预判，若轴所显示的距离与实际拍摄距离出入较大，则可以重新拍摄图片，再次进行标定。

（2）导出相机参数后的分析

导出内外参数后，得到双目相机中右相机相对于左相机的位置关系，即旋转平移矩阵，如表 4-3 所示。

其中平移矩阵为一个向量，向量中的第一个值代表左右相机光心的距离，即基线的距离，单位为毫米。旋转矩阵理想情况为单位阵，可通过对比上述两个参数，从而决定是否是要重新标定。

最后，使用 MATLAB 进行双目摄像头标定时，可以选择模型的复杂度，包括径向畸变的系数（2 Coefficients 或 3 Coefficients），切向畸变，以及像素的非正方形（Skew）。其中鱼眼相机需要选择三参数模型，一般双目相机只需选择两参数模

型，并且勾选上切向畸变（Tangential Distortion），以及像素的非正方形（Skew）等两个选项，本文的标定结果已在表 4-2 和 4-3 中展示。

2．图像畸变矫正

如图 4-9 和图 4-10 所示，(a)图均为未校正的原图，从(b)图可以看到图片的四周均有黑边，为对原图进行畸变矫正后的结果。

(a) (b)

图 4-9 左相机镜头畸变矫正图。(a)校正前；(b)矫正后

(a) (b)

图 4-10 右相机镜头畸变矫正图。(a)校正前；(b)矫正后

3．极线矫正

如图 4-11 所示，为左右图像的角点提取示意图，该图是未进行极线矫正的左右影像，在图 4-12 中，为极线矫正后的图像，在横线上，左右影像的同名点满足共线，因此特征点匹配的搜索范围从二维搜索降为一维线性搜索，极大的提高了搜寻效率，为下节的立体匹配提供了约束条件。

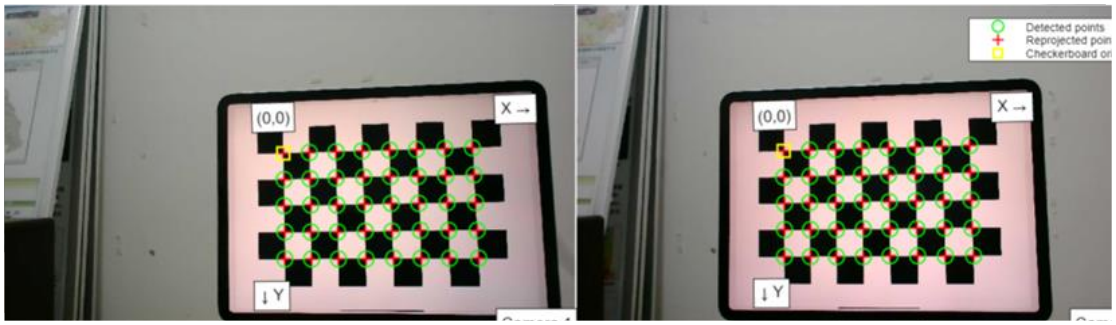


图 4-11 角点提取图

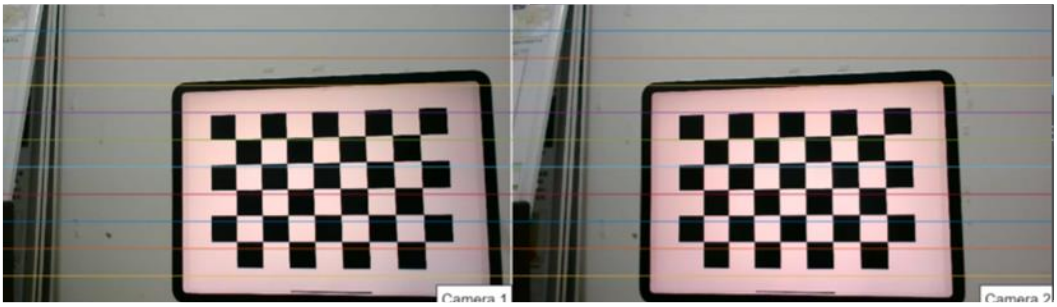


图 4-12 极线矫正图

4.3 立体匹配算法的分析

针对双目立体图像，通过立体匹配可以得到的同一特征点的位置，从而计算同名点的视差，并根据双目相机的焦距，求解物体在其表面的三维信息。常用的立体匹配算法有图像匹配算法 SAD、局部匹配算法 BM 和半全局匹配算法 SGBM。

4.3.1 SAD 算法

SAD 匹配算法是一种基于像素级差异的图像匹配方法，其数学表达式为 $SAD(u,v) = \text{Sum}\{|L(u,v) - R(u,v)|\}$。这个公式描述了在双目视觉系统中，左图像$L$和右图像$R$之间的像素差异的绝对值之和。在这里，$u$和$v$分别代表了图像的行和列坐标。SAD 算法的基本思想是将匹配问题分解为左右视差图中每个像素值之差的绝对值的求和问题。这种方法将每个像素或者图像块与其在另一图像中的对应像素或图像块的数值差的绝对值进行求和，从而得到一个量化的匹配误差。这个匹配误差可以用来评估两个图像块之间的相似性，从而实现图像匹配。SAD 匹配算法因其简单和计算高效的特性，广泛应用于多级图像处理的初步图像甄别阶段。然而，SAD 算法的匹配精度并不高，这是因为它仅仅考虑了像素级的差异，而没有考虑到像素之间的空间关系和图像的全局信息。因此，尽管 SAD 算法的运算速度快，但其匹配结果的准确性有待提高。

SAD 匹配算法的基本流程首先包括接收两个输入图像，即左图像left_image和右图像right_image。在这个过程中，首先设定一个 SAD 窗口，以左图像中的某个坐标点(x, y)作为锚点，生成一个覆盖区域。同样的操作也应用于右图像，生成一个对应的覆盖区域。在这两个覆盖区域中，定义左右图像的像素差为difference。然后，对这个difference值进行绝对值运算并求和，得到一个量化的匹配误差。这个匹配误差可以用来评估在当前窗口位置下，左右图像的匹配程度。接下来，将 SAD 窗口在右图像中滑动，重复上述操作，每次滑动都会得到一个新的difference值。这样，就可以得到一个difference值的矩阵，这个矩阵描述了在不同窗口位置下，左右图像的匹配误差。最后，从这个difference值的矩阵中找出最小的值，这个最小值就代表了左右图像的最佳匹配位置，也就是视差值。这个视差值可以用来估计物体的深度信息。SAD 算法计算简单且高效，但准确性不高。

4.3.2 BM 算法

BM 算法是一种经典的局部立体匹配算法，主要用于构建视差图。其基本思想是在左右两幅图像中，对每一个像素，找到一个固定大小的窗口，然后在另一幅图像中寻找最相似的窗口，从而得到视差。BM 算法的具体步骤如下：首先，在左图像中选择一个窗口，然后在右图像中的搜索范围内，计算该窗口与所有可能窗口的相似度。相似度计算通常基于像素强度的差异，例如总和绝对差 SAD 或平方差 SSD。最后，选择相似度最高（或差异最小）的窗口，其对应的视差作为当前像素的视差值。BM 算法的优点包括：简单、直观、计算速度快。它不需要复杂的优化过程，因此在实时系统中经常被使用。然而，BM 算法也存在一些缺点：首先，它假定窗口内的所有像素具有相同的视差，这在物体边缘或纹理稀疏区域可能导致误匹配。其次，窗口的大小选择对结果影响较大，窗口过大可能忽视细节，窗口过小可能导致匹配不稳定。

相比于 SAD 算法，BM 算法通常包含更多的预处理和后处理步骤，如滤波、截断处理等，这些步骤可以有效地提高匹配效果，尤其是在低亮度、纹理稀疏的区域。此外，BM 算法在视差判定时，考虑了所有相同像素的对应函数及其绝对值之和，这可以防止在处理超近距离图像时，将具有较低亮度纹理的物体误分类。然而，尽管 BM 算法在视差图的生成上相比 SAD 算法有所优势，但由于它是一种局部匹配方法，因此生成的视差图仍然具有较大的“非视差区域”。这意味着在这些区域，由于缺乏足够的纹理信息，BM 算法无法准确地估计视差。

4.3.3 SGBM 算法

SGBM 是一种半全局匹配算法。它通过为每个像素选择视差值，并设定全局能量参数，来获取视差映射关系。值得注意的是，SGBM 算法在获取视差的过程中并不设置堆栈函数，这与一些其他的立体匹配算法有所不同。在代价计算过程中，SGBM 首先对输入图像进行水平方向的梯度滤波，这可以有效地提取图像的边缘信息，从而提高匹配的准确性。接下来，SGBM 将计算出的 BT 代价值进行融合，然后通过 SobelX 处理图像计算 BT 代价。SobelX 是一种用于检测图像中水平方向边缘的算子，通过这种处理，可以进一步提取和强化图像的边缘信息。最后，SGBM 对结果得到的代价值进行成块处理。这种处理方式可以有效地降低计算复杂性，同时也能够保持匹配的准确性。

SGBM 算法的能量函数形式如公式(4-8)所示：

$$E(D)=\sum_{P}\left(C\left(P,D_p\right)+\sum_{q\in N_p}P_1I\left[\left|D_p-D_q\right|=1\right]+\sum_{q\in N_p}P_2I\left[\left|D_p-D_q\right|>1\right]\right) \tag{4-8}$$

在 SGBM 算法中，$C\left(P,D_p\right)$表示像素点p对应的视差为D_p时，所有像素匹配代价之和。这是一个关键参数，用于评估不同视差值的匹配质量。P_1和P_2是在临近像素点N_p处的惩罚常数，它们是 SGBM 算法中的重要参数，用于约束视差图的平滑性和连续性。P_1仅在视差值等于 1 个像素时起作用，与图像的平滑度有关。P_2与视差图的边缘有关，其值越大，生成的视差图的边缘越差。这个参数通常用于处理视差图中的不连续区域，从而保持视差图的边缘清晰。通常，会使用式(4-9)来设置P_2的阈值，以便在保持视差图边缘清晰的同时，避免过度平滑或过度不连续的问题。

$$P_2=\left(\frac{p_2'}{\left|I_p-I_q\right|}\right) \tag{4-9}$$

设$L_r(p,d)$表示点p在r方向上，视差为d的匹配代价总和，则可以求出匹配代价总和，如公式(4-10)所示：

$$\begin{aligned}L_r(p,d)=C(p,d)+min(L_r,(p-r,d),L_r(p-r,d-1)+P_1,\\ L_r(p-r,d-1)+P_1+\min_i L_r(p-r,i)+P_2)-\min_K L_r(p-r,k))\end{aligned} \tag{4-10}$$

上式中，$\mathrm{C(p,d)}$为像素代价值，设像素p的灰度值为$\mathrm{I_{(p)}}$。将$\mathrm{I_{(p)}}$，$\left(\mathrm{I_{(p)}}+\mathrm{I_{(P-1)}}\right)/2$，$\left(\mathrm{I_{(p)}}+\mathrm{I_{(P+1)}}\right)/2$，分别与$\mathrm{I_{(p)}}$做差，其中插值最小的即为$\mathrm{d(p,p-d)}$，然后再从$\mathrm{I_{(p)}}$，$\left(\mathrm{I_{(p)}}+\mathrm{I_{(P-1)}}\right)/2$，$\left(\mathrm{I_{(p)}}+\mathrm{I_{(P+1)}}\right)/2$这三个值中选出与$\mathrm{I_{(p)}}$插值最小的，即为$\mathrm{d(p,p-d)}$。以上两值中的最小值，就为像素p的代价值$\mathrm{C(p,d)}$，该过程如下公式(4-11)和公式(4-12)所示：

$$C(p,d) = min\big(d(p,p-d,I_L,I_R), d(p-d,p,I_L,I_R)\big) \tag{4-11}$$

$$d(p,p-d,I_L,I_R) = \min_{(p-d-0.5,p-d+0.5)} |I_L(p) - I_L(q)| \tag{4-12}$$

公式(4-10)中的第二项表明，当点p像素差值为d，$d-1$，$d+1$以及其他的值时，所对应的最小代价值分别为$L_r(p-r,d)$，$L_r(p-r,d-1)+P_1$，$L_r(p-r,d+1)+P_1$以及$L_r(p-r,i)+P_2$。

该算法对每个像素从相邻路径方向计算代价并累加，选择累加值最小的为最终视差值，如公式(4-13)所示：

$$S(p,d) = \sum_r L_r(p,d) \tag{4-13}$$

4.3.4 实验结果与分析

本节通过相关实验来验证各个算法的效果与可行性，选取 Middlebury 官方给出极线矫正后的 teddy 左右影像作为分析各个立体匹配算法的输入，输入如图 4-13 所示，从图 4-14 看出左右两幅图像已实现了极线矫正，即左右图像同名点共线。

(a) (b)

图 4-13 teddy 原始左右影像

图 4-14 teddy 极线矫正图

（1）将图 4-12 左右影像作为 SAD 匹配算法的输入，得到的视差图如图 4-15 所示：

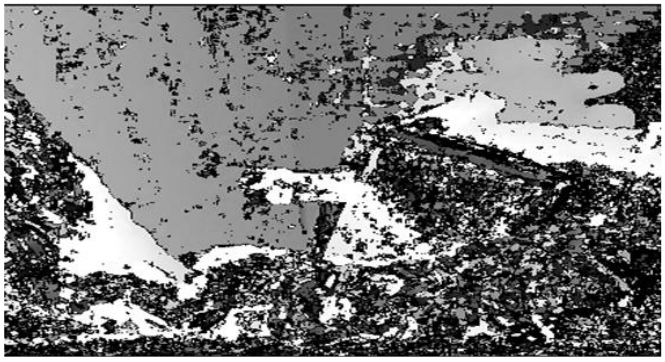


图 4-15 SAD 匹配后的视差图

（2）将图 4-12 左右影像作为 BM 匹配算法的输入，得到的视差图如图 4-16 所示：

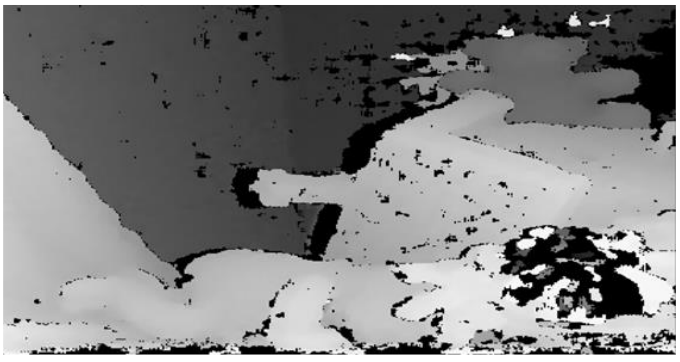


图 4-16 BM 匹配后的视差图

（3）将图 4-12 左右影像作为 SGBM 匹配算法的输入，得到的视差图如图 4-17 所示：

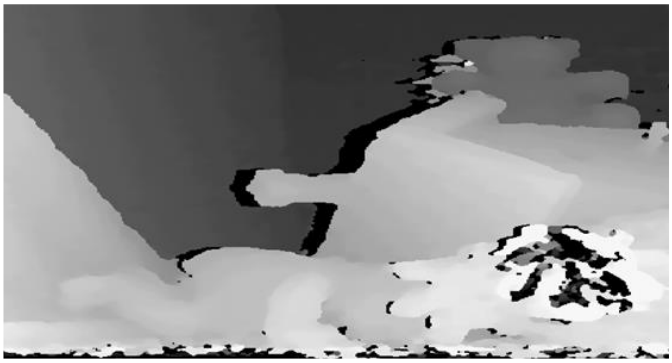


图 4-17 SGBM 匹配后的视差图

本实验分别对 SAD、BM 以及 SGBM 算法进行测试，匹配耗时如表 4-4 所示：

表 4-4 各算法匹配耗时表

	SAD	BM	SGBM
耗时（ms）	236.7	376.4	468.8

实验测试图为 Middlebury 中的泰迪标准测试图，分辨率为 375×450。实验表明，SAD 匹配算法用时最快，但所得到的视差图空洞点多且视差图不连续，导致测距结果不够精确，如图 4-15 所示；BM 算法用时稍长于 SAD 算法，但在视差图的质量上好于 BM 算法，仍存在一些噪点，如图 4-16 所示；SGBM 算法在计算时间上略长于 BM 算法，但匹配的结果更加连续，视差图的质量明显更好，如图 4-17 所示。结果表明，SGBM 算法能够得到更为连续且有层次的视差图。

4.4 基于 SGBM 的算法设计

在常用的匹配算法中，SGBM 算法以其能够有效地处理纹理丰富的区域，具有较好的准确性和鲁棒性的特点，被广泛应用于计算机视觉领域中的立体匹配任务，如三维重建、环境感知等领域。基于 4.3 节的不同立体匹配算法的实验结果分析，在满足时间要求的基础上，SGBM 算法有最高的匹配精度，因此，本文采用 SGBM 作为匹配算法。但是 SGBM 算法仍有一些缺点，下面将介绍 SGBM 算法的缺点以及基于 SGBM 的算法设计，技术路线如图 4-18 所示：

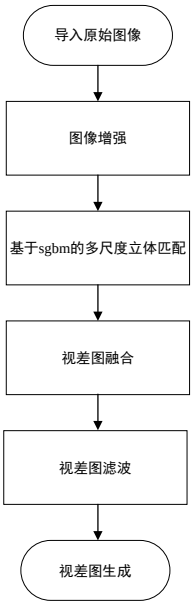


图 4-18 基于 SGBM 的算法设计流程图

4.4.1 SGBM 算法的分析

（1）对纹理匮乏的区域处理效果较差：SGBM 算法依赖于图像的纹理信息来进行匹配，因此在处理纹理匮乏的区域，会产生较大的匹配误差。所谓的纹理，是指图像中重复出现的基本元素和它们的排列规律，例如树叶、砖墙、织物等都有丰富的纹理。在有丰富纹理的区域，每个像素点的上下文信息都是独特的，使得 SGBM 能够找到准确的匹配。但在纹理匮乏的区域，例如平滑的墙面、单一颜色的物体等，像素点的上下文信息可能都非常相似，使得 SGBM 难以找到准确的匹配。因此，在处理纹理匮乏的区域时，SGBM 可能会产生较大的匹配误差，导致计算出的深度信息不准确。

（2）对光照和视差变化敏感：光照和视差变化是计算机视觉中的两个重要因素，它们对 SGBM 算法的影响较为明显。理解这些因素对立体匹配的影响，对于改进算法和提高匹配准确性具有重要意义。首先，光照变化是一个挑战。在实际应用中，光照条件可能会随着时间、天气、环境等因素的变化而变化。例如，当光照条件变化时，相同的物体可能会产生不同的像素值。这是因为像素值反映的是物体表面反射的光线的强度，而这个强度是受到光照条件的直接影响的。因此，当光照条件变化时，即使是同一个物体，也可能会在图像中产生不同的像素值，这可能会导致 SGBM 算法无法找到正确的匹配，从而导致匹配失败。其次，视差变化也是一个挑战。视差是由于相机的立体视觉引起的，它反映了一个物体在两个相机视图中的位置差异。当视差变化较大时，例如当物体离相机非常近或非常远时，可能会产生匹

配歧义。这是因为在这些情况下，相同的视差可能对应多个可能的物体位置，这使得 SGBM 难以确定正确的匹配。因此，当视差变化较大时，SGBM 可能会产生错误的深度估计。

4.4.2 基于 SGBM 的算法改进

为了解决上面提出的关于 SGBM 算法对弱纹理，光照变化和视差变化敏感的问题，本节分别在立体匹配前、中、后等三个阶段做出改进。

（1）立体匹配前，增加图像对比度和直方图均衡化，这两种方法都可以增强图像的纹理信息，使得纹理匮乏的区域更容易被匹配。增加图像对比度可以使图像中的纹理特征更加明显，而直方图均衡化则可以改善图像的亮度分布，使得不同的纹理特征更容易被区分。

（2）立体匹配过程中，采用多尺度策略，如前面所述，多尺度策略可以在不同的尺度（即不同的分辨率）上处理图像，以更好地处理纹理匮乏和视差不连续的问题。在较低的分辨率上，每个像素包含的是一个较大的区域的信息，因此可以更好地处理纹理匮乏的区域；在较高的分辨率上，可以更精细地处理视差的变化，因此可以更好地处理视差不连续的区域。

（3）立体匹配后，对生成的视差图进行滤波处理，首先由于图像采集和处理过程中可能会引入一些噪声，这些噪声会导致视差图中出现不必要的细小波动。通过滤波处理，可以有效地去除这些噪声，使得视差图更加平滑和稳定。其次滤波处理可以帮助消除视差图中的异常值和离群点，从而提高视差图的精度和准确性。这对于后续的深度估计、三维重建等任务非常重要。经过滤波处理的视差图更加清晰、平滑，可以提升人眼的观感体验，也更适合用于图像展示和可视化。最后，经过滤波处理的视差图更容易被后续的图像处理算法识别和利用，例如利用视差图进行物体分割、场景重建等任务时，滤波处理可以帮助提高后续处理的效果和稳定性。

4.4.3 实验结果评价指标

本实验分为两部分，其一是对物体的静态测距，以验证改进立体匹配算对测距精度的影响。本实验将距离作为评价指标有两点原因：一方面，坐标不易对比且较难获取准确的坐标数值，而距离易获取，且与实际距离容易作出对比分析；另一方面，在试验设计过程中，已尽量保持待测物体平面与镜头平面平行，且镜头光心垂直与待测物体平面，因此，坐标的只在 Z 方向上变化，而 Z 值即为待测物体与镜头间的距离，故该实验将实测距离作为定位实验精度分析的指标。所测的数据包括

实测距离值、未改进算法所测距离值以及改进后算法测距离值，实验结果于 4.4.4 节进行展示与分析。

其二是基于目标检测及定位算法相结合系统对障碍物的检测与定位的实验。该实验主要评价指标为检测正确度、测距精度以及测距范围。选择测距范围作为评价指标的主要原因是尽管双目摄像头本身有测距范围参数，但在该范围内障碍物识别效果并不得知，因此需要在保证障碍物识别率的基础上重新定义测距范围，具体实验结果将在 5.2.4 节展示。

4.4.4 实验结果与分析

本次实验分为两部分，一部分是和上节提到的 SAD，BM，SGBM 等匹配算法对于 Middlebury 提供 tiddy 图像所生成的视差图做静态对比；另一部分是利用双目摄像头分别使用普通 SGBM 算法和本节改进的算法进行实际距离的测量，将测距精度作为动态对比。

1.针对 tiddy 图像的静态对比

结果如图 4-19 所示。

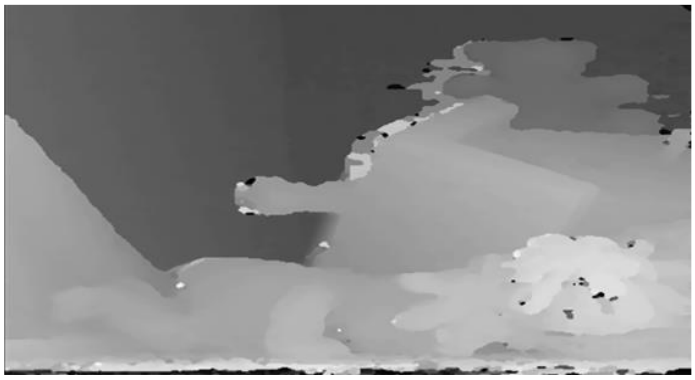


图 4-19 改进 SGBM 匹配后的视差图

从图 4-19 可以看出，无论从噪点个数还是视差图的连续性上看，效果都优于普通的 SGBM 算法所得到的视差图，但运行时间由 468.8ms 增加到 520.58ms。

2.实际测距的动态对比

在实际测距过程中，本实验使用的测距工具为激光测距仪，具体参数如表 4-5 所示：

表 4-5 激光测距仪具体参数

参数名称	参数数值
测量范围	0.05m-50m
测量精度	±3mm
测量单位	m
激光等级	2 类
激光类型	630-670mm 红光

实验地点为会议室如图 4-20 所示，测距距离为 50cm-400cm，从距离摄像头 50cm 处开始测量，待测物体为表面水平的茶盒，并且调节了待测物体的高度，使双目摄像头的左相机镜头光束与待测物体平面垂直，因此，在实际用激光测距仪测距中，结果会更加准确。在移动待测物体过程中，将沿着卷尺方向移动，尽可能保证待测物体平面与左相机镜头平面平行。测量过程图如 4-21 所示，两种算法的测量结果如表 4-6 和表 4-7 所示。

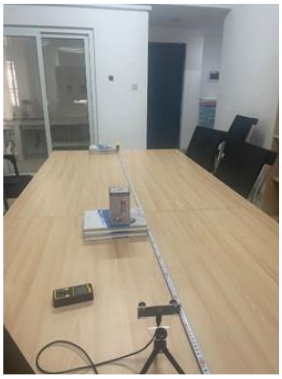


图 4-20 测距地点

(a)

(b)

图 4-21 测量过程图。(a)50cm 和 100cm 测距图；(b)150cm-200cm 测据图

表 4-6 各算法测距精度表

组数	实际距离/cm	改进 SGBM 计算距离/cm	SGBM 计算距离/cm
1	50	49.69	50.78
2	100	99.20	100.92
3	150	149.84	151.20
4	200	200.30	201.01
5	250	249.56	251.49
6	300	299.14	302.23
7	350	349.85	352.21
8	400	400.34	403.19

将上表转化为相对误差分析表如 4-7 所示：

表 4-7 各算法测距相对误差表

组数	实际距离/cm	改进 SGBM 相对误差	SGBM 相对误差
1	50	-0.62%	1.56%
2	100	-0.80%	0.92%
3	150	-0.10%	0.80%
4	200	-0.15%	0.50%
5	250	-0.17%	0.59%
6	300	-0.28%	0.74%
7	350	-0.04%	0.63%
8	400	0.08%	0.79%

从上面的统计数据中可以看出改进的 SGBM 算法在实测数据中相对于 SGBM 算法所测得的数据均有提升，说明改进的思路对弥补 SGBM 本身缺点有正向帮助，这为对双目匹配算法的改进提供了有效的思路。

4.5 本章小结

本章首先对双目摄像机相机进行标定和校正，并给出标定的经验和方法，提出了提高标定效率的两段式标定分析法，其次对常用的立体匹配算法进行了实验分析，分别研究了 SAD，BM 和 SGBM 算法的各自特点，并给出具体的实验结果，最后针对 SGBM 算法的对于光照、纹理敏感的问题，提出了改进的算法，并且对新旧算法进行了静态和动态的比较，实验证明本节提出的算法不论是在视差图的生成上还是实际的定位精度上都较 SGBM 算法有提升，这为双目立体匹配算法的改进提供了思路。

第五章 基于改进 YOLO 算法与双目定位算法相结合的系统开发

为了实现障碍物检测与定位的自动化，本章设计并开发了一套基于优化后的 YOLO 算法和双目定位算法的系统软件。该软件整合了障碍物检测算法、障碍物定位与测距算法，以及一系列常用的图像处理功能，包括图像拉伸、图像增强和图像裁剪等。为保持软件的灵活性、可维护性和可扩展性，该软件在设计时严格遵循了高内聚低耦合原则。

5.1 系统模块功能设计

障碍物检测与定位软件主要实现了一些常用的图像操作及处理功能。本章将这些功能具体划分为四大模块，分别为可视化、图像预处理、障碍物识别、非实时障碍物定位测距及实时障碍物定位测距。该软件功能架构如图 5-1 所示。

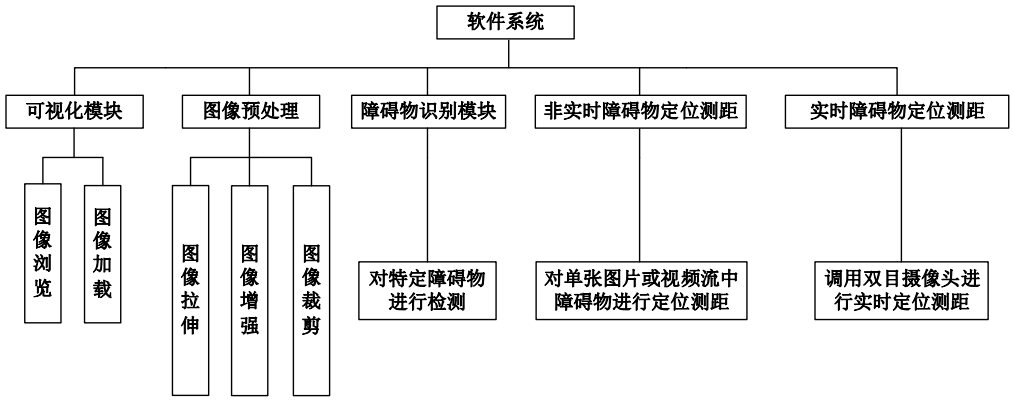


图 5-1 软件功能架构图

（1）可视化模块：该模块主要负责图像的加载、浏览等功能。此外，它还提供了图像缩放和漫游等功能，从而使用户能够更方便、更直观地操作图像数据，提升用户体验。

（2）图像预处理模块：该模块提供了包括图像拉伸、图像增强以及图像裁剪等图像处理功能。这些功能可用于对图像进行预处理，从而为后续的模型训练和双目定位提供更优质的输入数据，有助于提高系统的性能和准确性。

（3）图像识别模块：这是软件系统的功能模块之一。通过调用这个算法模型，我们可以自动地识别图像中的障碍物类型，并获取这些障碍物的检测框信息。这些信息将被用于后续的双目定位模块，从而确保更为精准的目标定位。

（4）非实时定位测距模块：该模块的是对静态图片或预先录制的视频流中的障碍物进行分类检测、定位以及距离估计。这一模块具有针对非实时数据进行深度分析的能力，为进一步的研究和应用提供了重要的功能支持。

（5）实时定位测距模块：该模块的主要功能是通过笔记本电脑调用双目摄像头，对动态移动的障碍物进行类型识别，以及位置定位和距离测量。这种设计使得系统能够更好地适应移动机器人在实际工作环境中可能遇到的复杂情况，提高其在不断变化的环境中的避障能力。

5.1.1 系统实现环境

本系统是在 Windows 平台上，以 Microsoft Visual Studio 2019 和 Qt Creator 4.10.0（Qt）作为开发工具，使用 C++语言并结合第三方开源库 OpenCV 开发的。计算机软硬件配置具体如表 5-1 所示。

表 5-1 计算机软硬件配置

运行环境	配置
操作系统	Windows 11
CPU	Inter core i5-12500H
GPU	GeForce RTX 3050
内存容量	16GB
分辨率	2560×1440

5.1.2 系统功能实现

（1）可视化模块

该模块的主要职责包括图像的浏览、加载、显示以及缩放和漫游等功能。软件系统的用户界面是基于 Qt 框架构建的，而图像的输入、显示以及缩放漫游等功能则是在 QGIS 的基础上进行二次开发实现的。此系统能够支持常见的图像格式的加载。系统的主界面如图 5-2 所示。

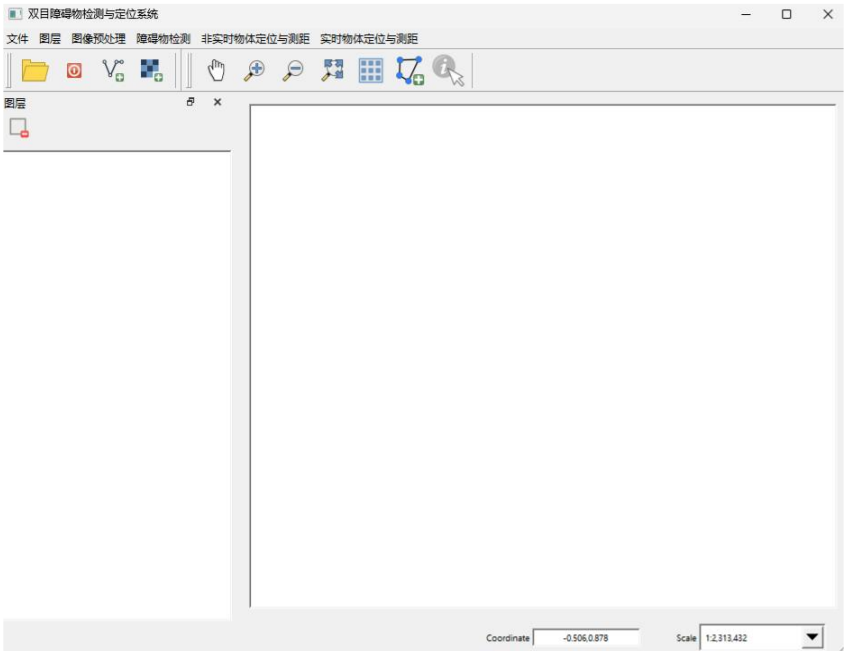


图 5-2 系统主界面

图像加载的具体流程如图 5-3 所示，图像的加载与显示、图像的缩放漫游分别如图 5-4 所示。

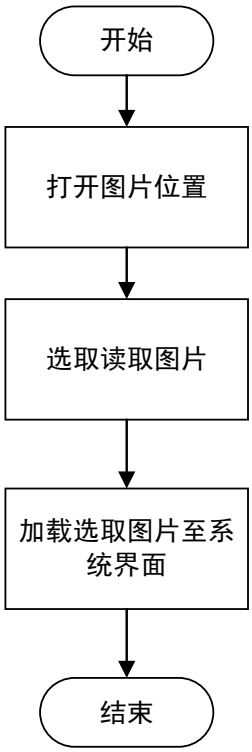
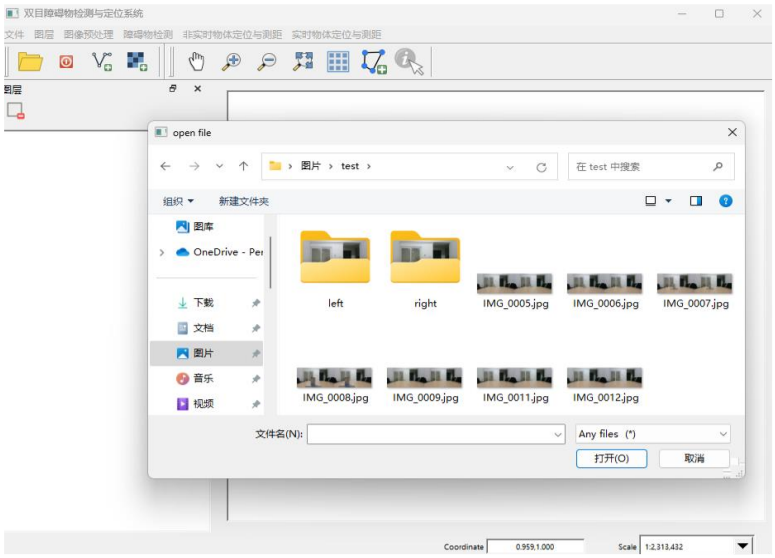
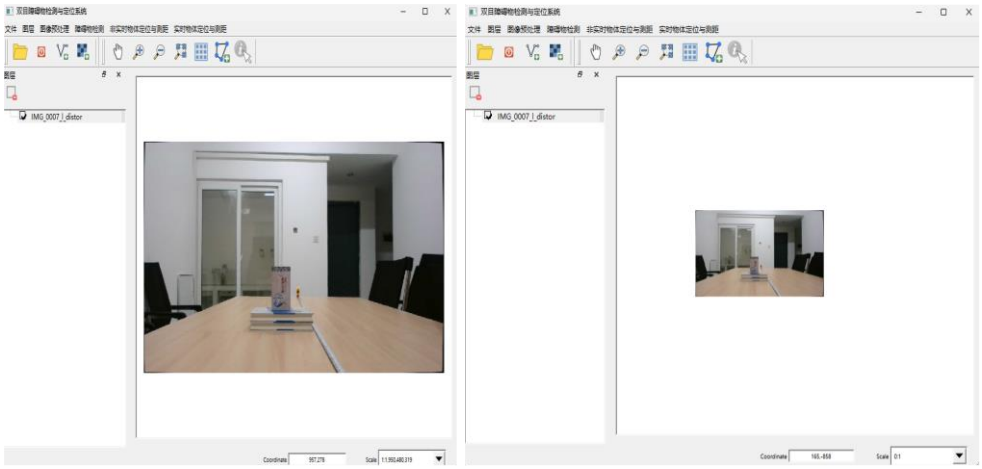


图 5-3 图像读取流程图



(a)



(b)

图 5-4 图像的加载与缩放。(a)图像加载功能；(b)图像缩放功能

（2）图像预处理模块

图像预处理模块是双目视觉任务中至关重要的一部分，主要包含图像增强、图像拉伸、和图像裁剪三个核心功能。这些功能的主要目的是优化原始图像数据，以提升后续图像分析和处理的效果和效率。图像拉伸方法主要包括标准值拉伸、线性拉伸、和百分比拉伸三种模式。这些方法主要通过调整图像的亮度和对比度，以增强图像的清晰度并提升图像质量。具体来说，标准值拉伸是基于统计学原理，通过调整图像的亮度和对比度使其符合某一标准分布；线性拉伸和百分比拉伸则是通过直观的线性变换或比例调整，来改善图像的视觉效果。图像增强主要是为后续的图像分析和处理提供更好的输入数据。图像裁剪功能主要应用于大规模图像数据的处理。由于计算机硬件资源有限，对大规模图像的处理通常需要消耗大量的计算资源，并且处理速度较慢。通过图像裁剪，可以将大规模图像分割成多个小规模的

子图像，然后对这些子图像进行单独处理，从而大大提升图像处理的效率。图 5-5 展示了预处理模块的用户界面，用户可以通过这个界面选择和调整各种预处理方法，以满足不同的图像处理需求。

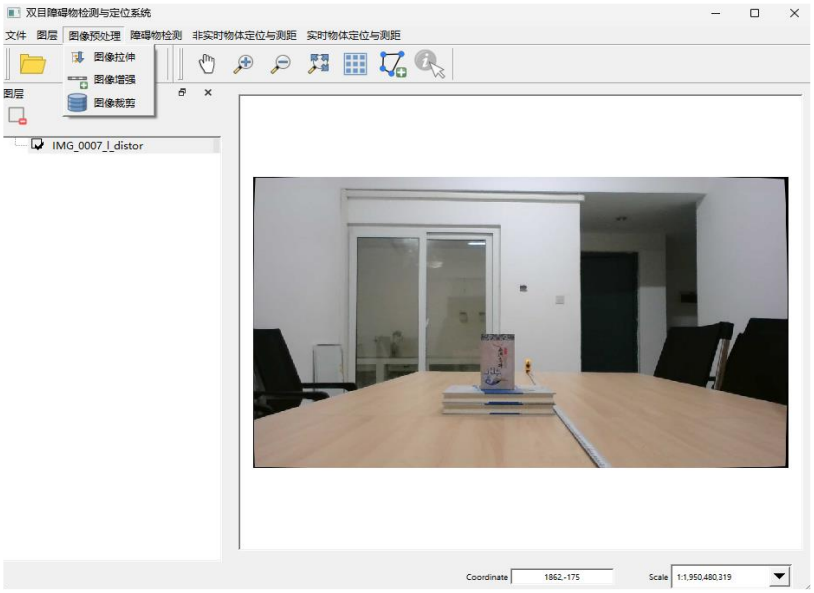


图 5-5 图像预处理主界面

在系统中，图像预处理包含图像拉伸、图像裁剪、图像增强三个功能。如图 5-6 所示，执行图像拉伸时，选择对应功能，选取待处理图片和适当的算法即可。

图 5-6 图像拉伸主界面

（3）图像识别模块

该模块主要设计了目标识别子界面及模型调用功能，包含图像和模型选择框，并设定了结果保存路径选择框。同时，为满足模型更新和扩展，设计了新模型添加接口。图 5-7 为目标识别子界面，其中有图像选择栏、算法类型选择栏、结果保存栏等。

图 5-7 目标识别子界面

（4）非实时障碍物定位与测距

此模块同样需要设计一个定位测距子界面。本文提出的基于 SGBM 的改进算法需要利用改进后的目标识别算法获取的信息，包括障碍物类别信息、障碍物坐标信息等，然后，给出障碍物相对于双目摄像头的位置信息。图 5-8 展示了测距模块的子界面设计。

图 5-8 测距子界面

（5）实时障碍物定位与测距

该模块将直接调用双目摄像头进行实时的障碍物检测与测距，并实时显示障碍物的类别，坐标以及相对距离，具体流程如图 5-9 所示，其中在计算障碍物最小距离时，本文采用从目标检测框的中心点向八个方向均匀采样，如图 5-10 所示，对比所采样的像素点的深度值，取最小值作为障碍物到左目摄像头的距离，相比于扫描法逐像素比较取最小值，时间复杂度更低，效率更高。

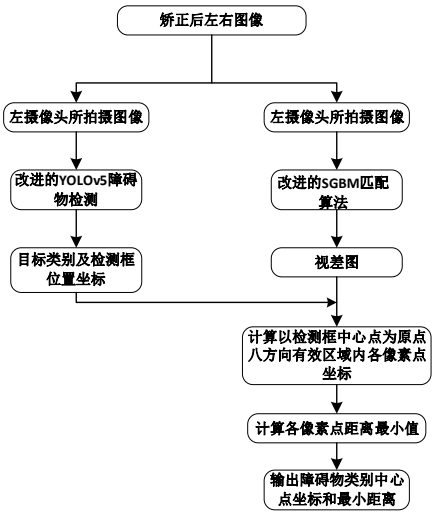


图 5-9 实时障碍物测距流程图

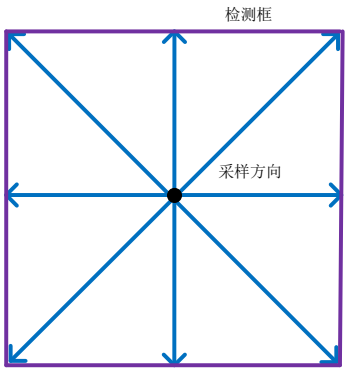


图 5-10 八方向像素点采样图

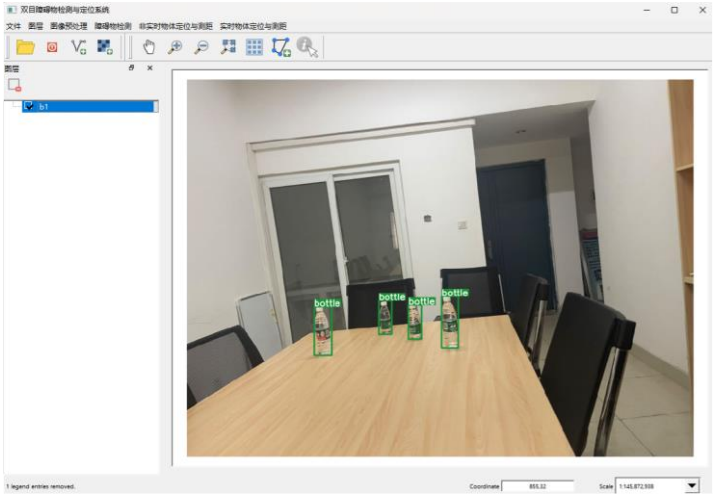
5.2 系统功能测试

在系统搭建完成后，本节对原型系统进行了功能测试。障碍物识别算法的测试数据采用第三章中的测试数据集。对于非实时障碍物定位测距，本节也沿用了障碍物识别算法的测试数据。而实时障碍物定位测距则通过调用摄像头自动生成测试数据。

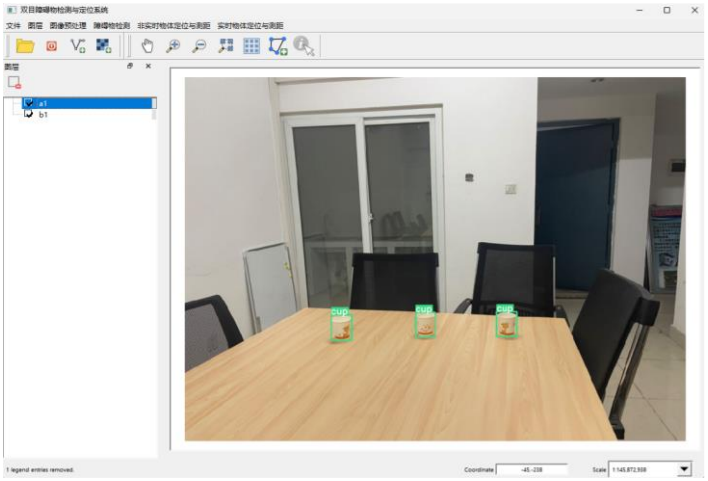
5.2.1 特定障碍物目标识别功能测试

在障碍物检测的子界面中，输入图像部分点击“打开”按钮，选择需要检测的图像。然后，在“算法类型”下拉框中选择所需要使用的检测模型，最后在输出图像栏中，设置输出图像文件名，并设置其保存路径。完成上述操作后，点击“运行”按钮即可进行障碍物检测。检测后的结果会储存到保存路径下，同时也会在界面上显示。障碍物检测功能测试如图 5-11 所示。

(a)



(b)

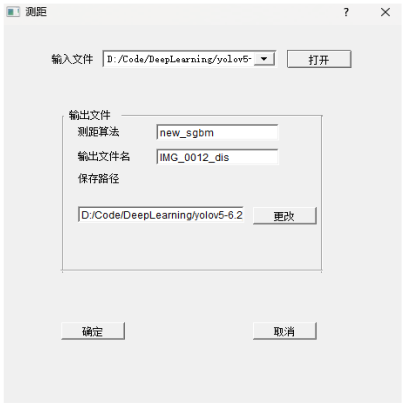


(c)

图 5-11 障碍物检测功能测试。(a)障碍物检测主界面；(b)水瓶目标识别结果；(c)水杯识别结果

5.2.2 非实时障碍物定位功能测试

在测距主界面“输入文件”栏中，选择需要进行障碍物检测的图片或者视频文件，点击“输出文件”栏中的“更改”按钮，为输出结果选择合适的保存路径，也可直接使用默认路径，然后点击“确定”按钮，触发障碍物检测与测距算法，直至出现“运行完成!”提示界面，即完成障碍物检测与测距，得到检测与测距结果。障碍物检测与定位功能测试如图 5-12 所示。该功能展示了在实际的应用场景中，即被检测物体在摄像头的不同方向上，仍能有效地对物体进行检测与定位，并且在与实测结果的比较中，物体定位精度也满足要求。



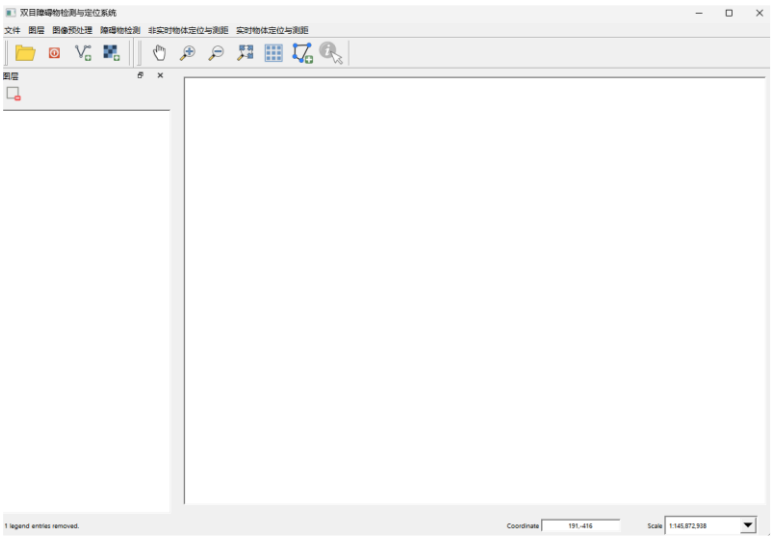
(a)

(b)

图 5-12 非实时障碍物定位测距功能测试。(a)非实时障碍物测距检测主界面；(b)测距结果展示

5.2.3 实时障碍物定位功能测试

在主界面中，点击“实时物体测距”按钮，即可调用双目摄像头进行实时的障碍物检测与测距，并且在实时显示在视频窗口中。实时障碍物检测与定位功能测试如图 5-13 所示。



(a)

(b)

图 5-13 实时障碍物定位测距功能测试。(a)主界面；(b)测距结果展示

5.2.4 性能指标测试

为验证系统对特定障碍物检测与定位精度，本实验选取常见的水杯为研究对象，如图 5-14 所示，在相同的环境中，分别沿与摄像头同一直线方向移动水杯至不同距离，每次以 50cm 的间隔移动，统计其目标检测率、测距精度以及定位与测距范围，其中目标检测率是指系统识别多个水杯中能正确识别的比值，具体结果如表 5-2 所示

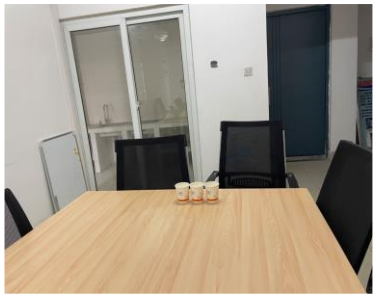


图 5-14 性能指标测试图

表 5-2 水杯性能指标测试结果表

组数	实测距离/cm	系统测距距离/cm	目标检测率
1	40	Nan	100%
2	90	88.84	100%
3	140	139.10	100%
4	190	199.04	100%
5	240	240.45	100%
6	290	291.11	100%
7	340	341.45	33.33%
8	390	392.13	0

实验过程为分别将三个紧靠的水杯沿直线方向移动不同的距离，得到系统对目标的识别结果以及测距距离。从上表中，可以看出共做了八组实验，当水杯放置距离 40cm 时，尽管系统均能识别出水杯，但未给出距离值，而当距离为 280cm 左右时，水杯不能被完全识别出来，只能识别出其中某一个，但实测距离与真实距离误差较小，当距离为 390cm 时，系统对三个水杯均识别不出。为准确给出系统在保证检测正确率基础上的测距范围，对水杯移动距离以 5cm 为步长在两个临界值进行实验，最终实验结果表明，在 45cm-320cm 左右的范围内，系统保证水杯检测率为 100%，并且能得到相对准确的测距结果。

除此之外，在相同的实验环境下，使用相同方法，即只改变物体的位置，分别对水瓶、背包、键盘三类目标进行性能指标测试。结果如表 5-3 所示：

表 5-3 不同类型障碍物性能指标测试结果表

类型	测距范围低值/cm	测距范围高值/cm	目标检测率
Cup	45	320	100%
Bottle	48	334	100%
Backpack	70	388	100%
Keyboard	82	280	100%

从表中数据分析可以得出，在满足对目标检测率 100%的要求下，对于背包等大体积目标而言，测距范围的低值相较于小目标会高一些，测距范围的高值也会更高，这是由于背包能否完整显示在成像图片中，对于目标检测尤为重要，而对于键盘这类物体，双目摄像机能否将其所有特征拍摄显示在图片中，对后续目标检测也有极大影响。因此单一双目摄像头很难完成在指定范围内对障碍物的检测和定位双重任务。

5.3 本章小结

本章主要介绍了障碍物检测算法与定位算法的系统设计与实现。该系统整合了障碍物检测算法、障碍物定位与测距算法，以及一系列常用的图像处理操作，除此之外，本章针对各项功能进行了测试，以保证系统的正常运行。

第六章 总结和展望

6.1 总结

在 21 世纪，人工智能的快速发展为机器人技术提供了强大的推动力。机器人不再仅仅是执行预设任务的自动化设备，而是具有学习和适应环境能力的智能实体。它们可以在复杂的环境中进行导航，执行精细的操作，甚至进行人类的社交交互。这一切都得益于深度学习和机器视觉等关键技术的发展。目标检测和双目立体视觉技术的进步，使得机器人能够更好地理解和交互其周围的环境。例如，通过这些技术，机器人机械臂可以识别并定位障碍物，进行抓取和各种操作。因此，本研究以 YOLOv5 目标检测技术和双目视觉定位技术为基础，对移动机器人机械臂避障或抓取任务中的障碍物检测与定位技术进行了深入研究。以下是本文的主要研究成果和结论：

（1）针对当前目标检测技术中，远距离小目标的检测精度较低，容易出现漏检和误检的问题。本文采用注意力机制模块（CBAM）与多尺度特征检测模块相结合的方法对 YOLOv5 算法模型进行改进，以增强 YOLOv5 算法的检测精度，改善对较小目标的识别效率。

（2）在当前基于双目视觉的物体定位与测距方法中，特征点匹配多用 BM，SGBM 算法，此类算法对于物体定位与测距精度而言并不高，误差较大，针对 SGBM 算法对弱纹理，光照变化敏感的问题，本文提出结合直方图均衡化、多尺度变换和滤波处理来优化 SGBM 算法，并利用改进后的 SGBM 算法对双目图像进行匹配，生成视差图，为双目定位和测距提供准确数据。

（3）针对双目相机标定过程注意事项较繁琐，标定参数准确性不易确定等问题，本文采用两段式标定分析法，以提高标定效率，得到准确度高的标定参数。除此之外，本文提出了一种结合改进 YOLOv5 障碍物检测算法和改进 SGBM 立体匹配算法的定位测距方法对左目图像特定障碍物目标进行识别检测，获取类别信息与位置信息，并满足实时性要求，能够便于实际的应用。

（4）为了方便、快捷地实现障碍物检测与定位，本文在 Windows 系统下，基于 Qt、OpenCV 等开源库，开发了一款障碍物检测与定位软件。该软件除了能对图像进行浏览、缩放漫游外，还具有图像预处理、障碍物检测、障碍物定位测距等功能。

6.2 展望

本文通过研究相机标定、立体匹配算法和目标检测算法，对双目定位任务有了更深的了解，但在学习过程中仍有一些不足，现将不足之处的改进方法从以下几方面展开：

（1）在目标检测任务中，本研究采用了基于 YOLOv5 网络模型的方法对图像中的障碍物进行识别，并对 YOLOv5 模型进行了改进。尽管改进后的模型在远距离小目标识别方面取得了一定的提升，但仍存在一些识别不准确或者无法识别的情况。因此，未来的研究可以在此模型的基础上进行优化，或者探索使用其他深度学习模型，以进一步提高目标识别的精度和准确性。

（2）在障碍物定位任务中，本文对 SGBM 立体匹配算法的改进虽然提升了匹配效果，但还需要注意，匹配效果的优劣并非仅由算法决定，其各个参数的设定也起着至关重要的作用。当前，大部分参数的调整都是手动进行的，以观察匹配效果的变化。然而，对于不同的图像，参数的最优设定可能会有所不同，因此这种方法的鲁棒性相对较差。在未来的研究中，可以考虑实现参数的自适应调整，以提高算法的鲁棒性和效果。

（3）本文成功地将障碍物检测与定位功能系统化，然而，这一系统尚未被集成到硬件平台中。在未来的研究中，可以考虑将这一系统架构移植到各类硬件设备上，例如移动机器人、嵌入式平台或无人机等，以实现其在工程领域的实际应用。

参考文献

[1] Perrone D, Iocchi L, Antonello P C. Real-time Stereo Vision Obstacle Detection for Automotive Safety Application[J]. IFAC Proceedings Volumes, 2010, 43(16): 240-245.

[2] Broggi A, Caraffi C, Fedriga R I, et al. Obstacle detection with stereo vision for off-road vehicle navigation[C].2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops. IEEE, 2005: 65-65.

[3] Huh K, Park J, Hwang J, et al. A stereo vision-based obstacle detection system in vehicles[J]. Optics and Lasers in engineering, 2008, 46(2): 168-178.

[4] Broggi A, Caraffi C, Porta P P, et al. The single frame stereo vision system for reliable obstacle detection used during the 2005 DARPA grand challenge on TerraMax[C].2006 IEEE Intelligent Transportation Systems Conference. IEEE, 2006: 745-752.

[5] Ashida T, Yamashita H, Yoshida M, et al. Signal processing and automatic camera control for digital still cameras equipped with a new type CCD[C].Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V: Vol. 5301. SPIE, 2004: 42-50.

[6] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C].2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05): Vol. 1. IEEE, 2005: 886-893.

[7] David L. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60: 91-110.

[8] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C].Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001: Vol. 1. IEEE, 2001: I-I.

[9] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on pattern analysis and machine intelligence, 2002, 24(7): 971-987.

[10] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20: 273-297.

[11] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of computer and system sciences, 1997, 55(1): 119-139.

[12] Rätsch G, Onoda T, Müller K R. Soft margins for AdaBoost[J]. Machine learning, 2001, 42: 287-320.

[13] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C].2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008: 1-8.

[14] Felzenszwalb P F, Girshick R B, McAllester D. Cascade object detection with deformable part models[C].2010 IEEE Computer society conference on computer vision and pattern recognition. IEEE, 2010: 2241-2248.

[15] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.

[16] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[17] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[18] Girshick R. Fast r-cnn[C].Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[19] 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.

[20] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[21] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[22] Moore D, Rid T. Cryptopolitik and the Darknet[J]. Survival, 2016, 58(1): 7-38.

[23] Gao B, Pavel L. On the properties of the softmax function with application in game theory and reinforcement learning[J]. arXiv preprint arXiv:1704.00805, 2017.

[24] Krishna K, Murty M N. Genetic K-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29(3): 433-439.

[25] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[26] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C].Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 390-391.

[27] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[28] 陈华, 王立军, 刘刚. 立体匹配算法研究综述[J]. 高技术通讯, 2020, 30(2): 9.

[29] Taniai T, Matsushita Y, Sato Y, et al. Continuous 3D label stereo matching using local expansion moves[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(11): 2725-2739.

[30] Leung C, Appleton B, Sun C. Iterated dynamic programming and quadtree subregioning for fast stereo matching[J]. Image and Vision Computing, 2008, 26(10): 1371-1383.

[31] 马瑞浩, 朱枫, 吴清潇, 等. 基于图像分割的稠密立体匹配算法[J]. 光学学报, 2019, 39(03): 240-248.

[32] 陈星, 张文海, 候宇, 等. 改进的基于多尺度融合的立体匹配算法[J]. 西北工业大学学报, 2021, 39(04): 876-882.

[33] Besse F O. PatchMatch Belief Propagation for Correspondence Field Estimation and Its Applications[D]. UCL (University College London), 2013.

[34] Wang Z F, Zheng Z G. A region based stereo matching algorithm using cooperative optimization[C].2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-8.

[35] 蒋文萍, 汪凌阳, 韩文超, 等. 基于改进Census变换的自适应局部立体匹配[J]. 电子测量技术, 2022, 45(13): 82-87.

[36] 郭龙源, 孙长银, 杨万扣, 等. SIFT 特征点引导的区域立体匹配算法[J]. 计算机工程与应用, 2013, 49(4): 23-25.

[37] Schaffalitzky F, Zisserman A. Viewpoint invariant texture matching and wide baseline stereo[C].Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001: Vol. 2. IEEE, 2001: 636-643.

[38] Gong M, Peng T, Zhang Z. A new non-parallel binocular stereo vision ranging system using combinations of linear and nonlinear methods[C].Sixth International Conference on Optical and Photonic Engineering (icOPEN 2018): Vol. 10827. SPIE, 2018: 193-198.

[39] 邹庆华, 张月雷. 计算机视觉技术应用[J]. 信息通信, 2015(12): 2.

[40] Epstein H T, Rapoport A. A note on the McCulloch-Pitts neural net for heat-cold discrimination[J]. The bulletin of mathematical biophysics, 1951, 13: 21-22.

[41] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.

[42] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv:1409.2329, 2014.

[43] Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal processing magazine, 2012, 29(6): 82-97.

[44] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.

[45] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[46] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C].Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[47] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[48] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.

[49] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C].Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.

[50] 侯一民, 洪梁杰. 双目摄像机标定及校正算法[J]. 中国新技术新产品, 2021(07): 1-3.

[51] 黄学然. 基于双目立体视觉的三维重建技术研究[D]. 西安电子科技大学, 2018.

[52] Abdel-Aziz Y I, Karara H M, Hauck M. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry[J]. Photogrammetric engineering & remote sensing, 2015, 81(2): 103-107.

[53] Li R, Di K, Matthies L, et al. Photogrammetric Engineering & Remote Sensing[J]. 2004.

[54] Tsai R Y. An efficient and accurate camera calibration technique for 3D machine vision[C].Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1986. 1986: 364-374.

[55] Zhang Z. A flexible new technique for camera calibration[J]. IEEE Transactions on pattern analysis and machine intelligence, 2000, 22(11): 1330-1334.

[56] 张宏. 基于双目立体视觉的三维重建技术研究[D]. 华中科技大学, 2007.

[57] 黄春燕. 双目立体相机标定算法研究与实现[D]. 中北大学, 2015.

[58] 岳晓峰, 祁欢. 基于张正友平面模板法的双目立体视觉系统标定[J]. 机械工程师, 2014(2): 3.

[59] 周芳. 双目视觉中立体匹配算法的研究与实现[D]. 大连理工大学, 2013.

[60] Anandan P. A computational framework and an algorithm for the measurement of visual motion[J]. International Journal of Computer Vision, 1989, 2(3): 283-310.

[61] Rosenholm D. Multi-point matching using the least-squares technique for evaluation of three-dimensional models.[J]. PHOTOGRAMM. ENG. REMOTE SENS., 1987, 53(6): 621-626.

[62] Hirschmuller H. Stereo processing by semiglobal matching and mutual information[J]. IEEE Transactions on pattern analysis and machine intelligence, 2007, 30(2): 328-341.

[63] Wu J, Cui Z, Sheng V S, et al. A Comparative Study of SIFT and its Variants[J]. Measurement science review, 2013, 13(3): 122-131.

[64] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features[C].Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9. Springer, 2006: 404-417.

[65] 周东翔, 孙茂印. 一种基于特征约束的立体匹配算法[J]. 中国图象图形学报, 2001(07): 43-46.

[66] 何俊, 葛红, 王玉峰. 图像分割算法研究综述[J]. 计算机工程与科学, 2009, 31(12): 4.

[67] Papadakis N, Caselles V. Multi-label depth estimation for graph cuts stereo problems[J]. Journal of Mathematical Imaging and Vision, 2010, 38: 70-82.

[68] Xu Z, Ma L, Kimachi M, et al. Efficient contrast invariant stereo correspondence using dynamic programming with vertical constraint[J]. The Visual Computer, 2008, 24: 45-55.

[69] Zhang K, Lu J, Lafruit G. Cross-based local stereo matching using orthogonal integral images[J]. IEEE transactions on circuits and systems for video technology, 2009, 19(7): 1073-1079.

[70] Yoon K J, Kweon I S. Adaptive support-weight approach for correspondence search[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(4): 650-656.

[71] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE transactions on cybernetics, 2021, 52(8): 8574-8586.

[72] Bodla N, Singh B, Chellappa R, et al. Soft-NMS–improving object detection with one line of code[C].Proceedings of the IEEE international conference on computer vision. 2017: 5561-5569.

[73] 赵婉月. 基于 YOLOv5 的目标检测算法研究[D]. 西安电子科技大学, 2021.

[74] 郭磊, 薛伟, 王邱龙, 等. 一种基于改进 YOLOv5 的小目标检测算法[J]. 电子科技大学学报, 2022, 51(2): 8.

[75] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088): 533-536.

[76] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.

[77] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[J]. arXiv preprint arXiv:1710.09412, 2017.

[78] 达星宇. 基于双目视觉的路面障碍物检测技术与算法研究[D]. 长安大学, 2022.