

学校代码: 10200
分 类 号: TP39

研究生学号: 2018102977
密 级:



东北师范大学

硕士学位论文

基于单目视觉的目标识别及抓取研究

Research on Target Recognition and Capture Based on Monocular Vision

作者: XXX

指导教师: XXX 副教授
专业学位类别: 工程硕士
专业学位领域: 计算机技术
学位类型: 专业硕士

东北师范大学学位评定委员会

2021 年 5 月

独 创 性 声 明

本人郑重声明：所提交的学位论文是本人在导师指导下独立进行研究工作所取得的成果。据我所知，除了特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。对本人的研究做出重要贡献的个人和集体，均已在文中作了明确的说明。本声明的法律结果由本人承担。

学位论文作者签名：_____ 日期：_____

学 位 论 文 使用 授 权 书

本学位论文作者完全了解东北师范大学有关保留、使用学位论文的规定，即：东北师范大学有权保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权东北师范大学可以将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、缩印或其它复制手段保存、汇编本学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名：_____ 指导教师签名：_____
日 期：_____ 日 期：_____

学位论文作者毕业后去向：

工作单位：_____ 电话：_____
通讯地址：_____ 邮编：_____

摘要

计算机视觉技术的不断进步不仅加速了工业自动化的进步，而且也大幅度推动了智能安全与人工智能等行业的发展。将计算机视觉技术应用于机器人抓取是将机器人控制和计算机视觉的有效结合，这种结合可能对工业生产的效率和其他方面产生重大影响。但是，市场上复合性较好的机器人产品大多存在着不开源、二次开发性较差、泛用性较差的问题，这些问题会导致机器人不能更为灵活的对环境进行适应。为了提高机器人的泛用性，本文在对机器人的目标识别、自主导航、自动抓取的相关算法的基础上，设计并实现了机器人的目标识别、自主导航、自动抓取等功能，为后续复杂系统的实现打下坚实的基础。

首先，本文对传统的与基于深度学习中的目标识别与抓取方法分别进行探讨；其次，阐述了深度学习的相关基础知识，并将深度学习的方法引入到目标识别技术中。为了满足在实际环境下对检测速度较快与检测效果较好的要求，在诸多基于深度学习的网络模型中，本文选择使用单映射的多框回归分类模型（Single Shot MultiBox Detector，SSD）在单目摄像头下实现对物体进行快速、有效的识别，并在 ROS 操作系统中通过机器人操作系统与 SSD 模型的节点进行数据交互，实现了对真实物体的实时识别；然后，对机器人的建立地图与自主导航原理与框架的相关内容进行研究，通过基于 Rao-Blackwellized 粒子滤波的 SLAM 方法对机器人周围环境进行有效建图，在 ROS 操作系统下，机器人可以在真实环境下建立地图，实现机器人的自主导航的功能。最后，为了实现对目标物体的抓取，对机器人进行手眼标定，确定机械夹手相对基坐标系的关系，进而进行了相关的实验。

本文主要对机器人的目标识别、自主导航、自动抓取的相关算法进行研究，并通过实验证明了算法的可行性，实验结果表明，在 ROS 操作系统下，可以对物体进行实时识别与检测、机器人自身进行定位与导航、对物体进行有效抓取，这为后续机器人泛用功能的提高奠定了研究基础，并对以后复杂机器人系统的设计与开发具有很好的前瞻性研究。

关键词：单目视觉；深度学习；目标检测与识别；定位与导航；目标抓取；ROS

Abstract

The continuous advancement of computer vision technology not only accelerates the progress of industrial automation, but also greatly promotes the development of industries such as intelligent security and artificial intelligence. The application of computer vision technology to robot grasping is an effective combination of robot control and computer vision. This combination may have a significant impact on the efficiency of industrial production and other aspects. However, most of the robot products on the market with better composite properties have the problems of not open source, poor secondary development, and poor versatility. These problems will cause robots to be unable to adapt to the environment more flexibly. In order to improve the versatility of the robot, this thesis designs and implements the robot's target recognition, autonomous navigation, and automatic grabbing functions based on the robot's target recognition, autonomous navigation, and automatic grabbing algorithms. The realization of the system lays a solid foundation.

First, this article introduces the traditional and deep learning-based target recognition and capture methods respectively; secondly, it explains the basic knowledge of deep learning and introduces deep learning methods into target recognition technology. In order to meet the requirements for faster detection speed and better detection effect in the actual environment, among many network models based on deep learning, this thesis chooses to use the single shot multibox regression classification model (Single Shot MultiBox Detector, SSD) in the single Under the eye camera, the object can be recognized quickly and effectively. In the ROS operating system, the robot operating system is used to interact with the nodes of the SSD model to achieve real-time recognition of real objects; then, the establishment of a map and autonomy for the robot The principle of navigation and the related content of the framework are introduced. The SLAM method based on Rao-Blackwellized particle filter is used to effectively map the surrounding environment of the robot. Under the ROS operating system, the robot can build a map in the real environment to realize the autonomous navigation of the robot. Function. Finally, in order to realize the grasping of the target object, the robot was calibrated by hand and eye, and the relationship between the mechanical gripper and the base coordinate system was determined, and then related experiments were carried out.

This thesis mainly studies the related algorithms of robot target recognition, autonomous navigation, and automatic grasping, and verifies the feasibility of the algorithm through experiments. The experimental results show that under the ROS operating system, objects can be recognized and detected in real time. Positioning and navigating by itself, and effectively grasping objects, lays a research foundation for the subsequent improvement of the general-purpose functions of robots, and has a good prospective study on the design and development of complex robot systems in the future.

Key words: monocular vision; deep learning; target detection and recognition; positioning and navigation; target grabbing; ROS

目 录

摘 要	I
Abstract	III
第一章 绪 论	1
1. 1 研究背景与意义.....	1
1. 2 机器人国内外研究现状.....	2
1. 2. 1 国内研究现状.....	2
1. 2. 2 国外研究现状.....	3
1. 3 论文主要内容及章节安排.....	4
第二章 目标识别及抓取方法概述	6
2. 1 目标检测.....	6
2. 1. 1 传统的目标检测.....	6
2. 1. 2 基于深度学习的目标检测.....	8
2. 2 目标抓取.....	9
2. 2. 1 传统目标抓取方法.....	9
2. 2. 2 基于深度学习的目标抓取方法.....	9
2. 3 本章小结.....	10
第三章 基于单目视觉的目标检测	13
3. 1 深度学习相关原理.....	13
3. 1. 1 深度学习.....	13
3. 1. 2 卷积神经网络.....	14
3. 2 SSD 目标检测算法.....	16
3. 2. 1 基础网络.....	16
3. 2. 2 SSD 网络.....	17
3. 2. 3 数据集.....	21
3. 3 实验验证.....	22
3. 4 本章小结.....	25
第四章 自主导航与目标抓取	27
4. 1 SLAM 建图.....	27
4. 1. 1 理论概述.....	27
4. 1. 2 获取深度信息的传感器.....	28
4. 1. 3 基于 Rao-Blackwellized 粒子滤波的 SLAM 方法.....	29
4. 2 机器人导航.....	30
4. 2. 1 理论概述.....	30
4. 2. 2 导航框架.....	31
4. 3 抓取方法总体概述.....	32
4. 3. 1 手眼标定.....	32
4. 3. 2 逆运动学分析.....	34
4. 4 实验验证.....	35
4. 5 本章小结.....	40
第五章 结论与展望	40

5.1 本文结论.....	40
5.2 未来展望.....	40
参考文献.....	41
在学期间公开发表论文及著作情况.....	44
致谢.....	45

第一章 绪 论

1.1 研究背景与意义

人工智能技术不仅吸引着许多科学家与研究人员的注意力，更是得到了多方媒体的关注。从目前的社会发展来看，有关于机器人的相关研究已经成为了一项全国性的技术与军备竞赛。机器人的可应用领域随着当今工业化的飞速发展在一定程度上得到了扩大，这是因为使用机器人代替人力资源在工业上具有巨大的优势：机器人可以长时间工作、机器人工作的准确性更高。用发展的眼光来看，现代工业技术需要具有更高适应性的、智能性更高、更为灵活的机器人。为满足当代工业化的需求，将感知技术引入到机器人技术是科技必然的发展，其中在感知技术中最为基础的便为视觉感知技术系统。

人类约有 80% 的信息是通过视觉获取得到的^[1]，这也可以说人类最为有力的感知方法便是视觉系统。视觉系统可以为人类提供大量有关周围环境的信息，由此人类就可以无需通过物理接触实现对物体进行直接操作。人类的视觉是通过作用在视觉器官上的光，由视觉神经系统处理后再产生视觉效果。机器视觉与其他感知方式相比具有更强的灰度分辨率、更高的空间分辨率、更强的环境适应性、更高的观察精度、更宽的光敏范围。与激光雷达与超声波相比，视觉传感器所收集到的信息比激光雷达和超声波收集到的信息更为精确，正所谓：“眼见为实，耳听为虚”；再者，视觉传感器相对于超声波和激光雷达的优势就像人眼相对于耳朵的优势一样，即视觉系统具有更低的延迟和更快的响应速度，由此可以说视觉系统更适用于实时工作，是以通过利用视觉来提高机器人的智能化水平具有非常重要的研究意义与价值^[2]。

随着人们生活水平的不断提高，每个人所追求的不仅是基础衣食住行的问题，更是追求整体生活水平的提高，与此同时人们对于自身安全的意识也得到了大幅的提高。为了避免由于剧烈运动、高压灰尘、噪音、重复性工作、高温等问题造成危险，机器人的技术自然而然的得到了大幅度发展^[3]。人类是一种擅长使用工具来制造特定工具的生物，在制造工具的同时人类还希望将生产效率最大化，机器人技术应运而生。就当今的工业化生产的发展过程来看，机器人在很大程度上取代了人力，对机器人的市场需求也在逐年增加。这种发展的原因是由于机器人的一次性投资自动化硬件的好处要远大于使用工人的装配线：通过在工业上使用机器人在提高生产效率的同时还可以防止劳资纠纷和其他问题的发生。

随着相关的机器人技术在近几十年来的不断发展，机器人已广泛应用于工业生产

与生活中的各个领域。机器人技术是一项较为完整且包括多种技术的复杂体系，其中所涉及的技术从实现角度来看都非常困难。但是由于整个社会对机器人的巨大的市场需求，越来越多的研究人员投身于机器人领域，解决其中的各类复杂问题。除个别区域，当今的市面上的机器人尤其特别适合于多个品种和批次的生产。通过应用机器人技术到工业生产中可以保证产品的质量、提高整体的生产效率、改进工作条件、解放人力资源、极大促进社会生产的发展，由此可见机器人技术在现代科技技术的发展过程中扮演着非常重要的角色。由于机器人所具有人力无可比拟的效率，机器人所成产的产品的更换周期可以得到大幅度的缩短，进而在产品降低价格的同时大幅度提高了产品在市场的竞争力，例如智能手机可以在用户可以接受的价格范围内尽可能多的发布新的产品版本。

机器人需要充分了解自身所处的环境，必须安装各种传感器才能感知环境，如视觉感知、距离感知、触觉感知和力量感知等感知信息，这些信息对机器人的感知有很大的影响，其中视觉感知是机器人最重要的感知方式，通过视觉系统机器人能够获取与周围环境相关的大量信息。机器人如果具有较强的视觉能力，可以使其通过视觉进行识别、收集、处理、理解等较为具体的决策。因此，自适应性较好的机器人不仅可以如同传统机器人一样可以执行较为机械化的基本动作，而且还可以独立的对问题进行决策。

除了视觉感知系统，机器人也会配置手臂进行具体任务的执行。举例来说，如果机器人周围的环境发生变化，需要调整相应设置与参数后，机器人才能执行命令的对应操作。为了提高机器人的自动化能力，让机器人变得更为智能。

1.2 机器人国内外研究现状

1.2.1 国内研究现状

国内关于计算机视觉机器人的相关研究起步时间相对于国外研究来说较晚，但是我国的科研人员在这方面也取得了一定的研究成果与技术突破。

湖南蓝天机器人科技公司推出的机器人系统不仅配置了 3D 相机，还搭载了与其相关的专用软件，通过这些硬件配置可以实现机器人对焊缝的自动识别和自动定位。这款三维视觉焊接机器人解决了传统焊接机器人在作业过程中可能出现的误差，如下料、装配和定位等，同时由于机器人搭载了可视化界面，在视觉感官上也大大降低了操作人员进行重复性编程的时间，极大的提高了整个焊接生产过程的效率和质量。

张朝阳设计了一款针对于废旧金属块分拣的基于计算机视觉的机器人分拣系统，系统首先利用传统的 OpenCV 视觉算法库对相机进行了标定，同时他也提出了一种基于金字塔形的迭代算法，进而完成了对运动中的目标物体进行抓取的行为，实现了对铜、铝等有色金属进行自动分拣处理的目标任务。北京理工大学的相关研究者开发了一款双足移动仿人机器人，此款机器人拥有的自由度为 32，身长 158cm，重 76kg，具有一系列识别功能，如语音感知、视觉感知、力觉感知等，同时此机器人还可以对太极拳进行模

仿，另外相关研究人员还在机器人的运动轨迹功能方面进行规划计算，对手臂抓取目标物等实际动作进行了进一步的研究。

根据透视成像原理^[4]，浙江大学相关研究人员提出一种只需要通过左右两个相机采集的图像中的特征点，然后对选取的特征点进行三维坐标的重建工作，最终实现检测机械动态结构，并可以对位姿进行精确的检测，这种方法的处理速度相对于传统的方法来说生产速度更快，生产效率更高。对于在生产线上的混杂分布的工件进行分拣工作的分拣机器人也有郝明等人进行了相关工作的研究，研究步骤包括对捕获到的工件图像信息进行平滑、去噪等处理，通过 Sobel 算法提取工件的边缘特征并对其进行二值化处理，最后通过 Hausdorff 算法模型进行模板匹配，进而得到工件的中心坐标信息，使得机器人实现了复杂工件的分拣任务。视觉技术同样可以应用于目标定位中，一种基于计算机视觉定位的脑外科机器人由中科院沈阳自动化研究所和东北大学研究提出，这项机器人技术的研究成果主要在医学领域进行应用，在这种脑外科机器人的帮助下可以大幅度提高手术的安全性与成功概率^[5]。

1.2.2 国外研究现状

在计算机视觉的相关研究中，美国等国家较国内来说较早将机器人与各行各业相结合，并且在发展速度方面来说要更快，随着商用计算机视觉系统与科技的不断进步而相关技术也取得了较大的进步。

斯坦福大学和美国加利福尼亚大学的研究者研发出一款机器人 4D 相机^[6]，这款相机是根据光场技术进行研究开发，不仅可以生成 4D 图像，也可以获得相机周围约 140 度的环境信息，由于这些优于当前机器人视觉系统的性能，此款相机可以广泛应用到虚拟现实、增强现实等技术中去。



图 1-1 4D 相机成像效果^[6]

德国卡尔斯鲁厄大学的研究人员开发了一种仿人双臂机器人，此款机器人可以较为准确的感知与模拟人类的执行能力，如在房间内执行抓取碗、碟等物品，这种机器人的手臂自由度为 7，且安装机器人身上的每个眼睛是彩色双目立体相机，这些相机可以在伺服电机的带动下进行转动。同时以该机器人为平台，卡尔斯鲁厄大学研究者提出了基于双目立体视觉的目标识别，视觉伺服抓取等相关算法。

大阪大学自主研发出一款自适应的双目视觉伺服系统，该系统通过雅克比矩阵对目标实时进行计算任务，进而可以对目标物体的运动方向进行预测，该方法相比于传统的视觉伺服跟踪系统，仅需要所捕获的两幅图像中有相对静止的参考物，整个过程中不需要任何摄像机的光学或摄像机的运动参数等数据信息^[7]。麻省理工学院提出通过将多种传感器融合并将其用于智能交通的领域中，整个过程首先使用雷达系统对目标的深度进

行估计，然后通过双目立体视觉系统对目标的具体深度信息进行精确计算，这种方法在高速的实时环境得到了优秀的效果^[8]。

卡内基梅隆大学与英特尔公司在 2010、2012 年共同合作开发了服务型机器人的 HERB1.0 版本和 HERB2.0 版本^[9]，该机器人搭载有视觉传感器，在软件系统方面选择使用主流 ROS 机器人操作系统与 OpenCV，此款机器人可以能够实现数据采集、信息交互、规划抓取算法等功能。

General Motors 公司于 1970 年研发出一款基于计算机视觉的检查系统，此系统对传送带上的工件进行识别。波士顿动力公司通过计算机视觉来获得关于周围的障碍物的相关距离信息而自主进行设计研发的仿生双目立体视觉四足机器人^[10]，在功能上可以实现自主的移动和有效的避障，同时可以有效的辨识机器人周围环境信息。PUMA/VC-100 在 1980 年进行国际发售，具有通过采用视觉算法辅助加工工件的能力，功能上可识别对象并可以较为准确的目标的位置坐标。

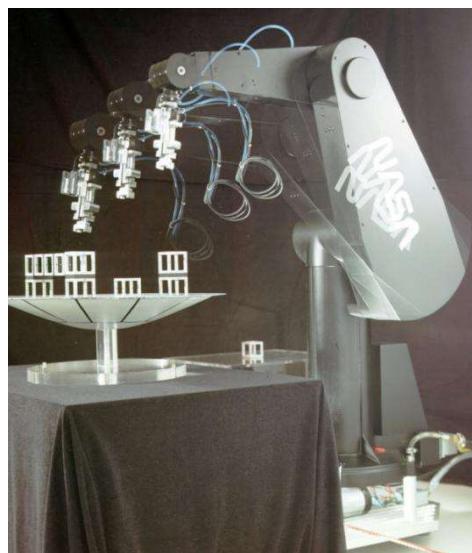


图 1-2 PUMA 机器人^[8]

在 1990 年后，整个日本的国家产业政策对于有高效、劳动密集型需求的生产线，如电子产业、汽车产业、制药产业、食品产业来说，工业视觉系统占有及其重要的地位，其中的部分功能实现了仅凭借人力不可达到的目标。波士顿公司于 2018 年发布 Atlas 的技能展示视频^[11]：Atlas 身长 1.5 米，重量为 75 公斤通过使用在其上搭载的视觉系统来对自身进行调整，同时还可以测量自身到障碍的精确距离，另外 Atlas 仅使用一条腿就可以跳过踏板。

1.3 论文主要内容及章节安排

本论文首先对机器人及视觉机器人的研究背景及描述相关的国内外技术进行阐述，接着对传统的目标识别方法与及基于深度学习的目标识别方法进行描述对比，然后分别

对基于单目视觉的目标检测、对目标周围进行地图构建、自主进行导航、物体抓取分别进行相关理论叙述与实验验证，实验结果均达到预期效果。

第一章是绪论部分，阐述计算机视觉与研究基于其的机器人研究的背景与意义，对相关技术与产品的国内外研究现状进行描述，论述主要章节内容的安排。

第二章是对传统方法与基于深度学习方法的目标识别与目标抓取方法分别进行阐述。

第三章是对基于单目视觉的物体识别进行研究。本章首先对基于目标识别算法进行了研究。接着重点研究了基于深度学习的 SSD 算法，并论述该算法在 ROS 框架下实现实时识别的相关内容，最后使用机器人在实际环境中进行实验验证。

第四章是对机器人建立地图的问题进行研究，对机器人的建图原理进行了分析，阐述了 ROS 操作系统下的建图方法，接着阐述视觉抓取系统的各个模块的构建，机器人手眼系统的标定以及机器人对目标物体的识别和抓取实验在此基础上论述机器人的导航原理，最后在实际环境中进行实验验证。

最后总结本文的全部工作内容，并对未来的工作进行展望。

第二章 目标识别及抓取方法概述

2.1 目标检测

目标检测不仅可以判断图像中存在哪些对象，还可以标记它们在图像中的位置。在机器人实现与外界交互时，最优先需要使用搭载在其自身或外在环境中的“眼”捕捉到外界的目标实例，这就是目标检测中所需要解决的问题。在解决更复杂的视觉任务，如图像分割、目标跟踪、图像描述和时间检测等任务时，目标检测更是关键的前期基础。目标检测的总目标是需要找到在给定图像中所关注的物体，包括两个基本的任务：定位与识别：目标识别负责对图像或所选择图像区域中是否存在任务所需的对象类别进行判断，同时生成一些标签。目标定位则对感兴趣的对象在输入图像或选定图像区域中的位置和范围，与此同时输出边界框、感兴趣的对象的中心和、闭合边界等。



图 2-1 目标检测效果

2.1.1 传统的目标检测

传统的目标检测算法主要分为 3 个部分：在给定图像上选择候选区域、对选定的区域进行特征提取、使用分类器进行分类，传统的目标检测流程如图 2-2 所示。为了定位目标实例在给定图像中的位置，首先需要选择一些子区域，其次在这些子区域中计算出类别概率值最大的子区域就是该目标实例的位置。在目标检测的算法的后续工作中，会对候选区域进行非极大值抑制以及修正，进而得到相对精确的目标位置信息。滑动窗口法

效率很低并且性能不优，为了解决滑动窗口所带来的效率低与性能低的问题，研究人员提出了候选区域算法，通过使用区域分割的算法来对潜在的物体进行识别。传统的目标检测算法中基于滑动窗口法首先存在着时间复杂度较大、窗口冗余的问题。其次，传统的目标检测方法没有反馈，只是单向地分析总结一幅图像的特征，没有实践反馈。



图 2-2 传统目标检测流程

滑动窗口法事先规定一个固定大小的窗口，使用这个窗口在原图中滑动，滑动到每个位置，候选区域由窗口与图像的重合部分构成，这个区域会为后续的检测任务服务。在 EdgeBoxes^[12]中可以看到这种方法未涉及到“深度学习”，采用纯图像的方法与当时的算法比较得到了更好的结果，但是在实际应用中这种方法对于给定图像中存在多目标的检测结果不理想。而且在滑动窗口法中，如果图像尺寸很大，会产生数量极多的候选区域。也就是说滑动窗口的尺寸设置需要与物体的尺寸相匹配才能带来好的效果，否则将严重影响后续特征提取和分类的速度和性能，因此对于检测任务来说，滑动窗口法效率很低并且性能不优。为了解决滑动窗口所带来的效率低与性能低的问题，研究人员提出了候选区域算法，通过使用区域分割的算法来对潜在的物体进行识别。通过不停的迭代，候选区域列表中的区域越来越大，就可以得到越来越大的候选区域。相比于滑窗法在不同位置和大小的穷举，候选区域算法将像素分配到少数的分割区域中，大大减少运行物体识别算法的次数。但是这种方法的速度不够快，无法满足实际应用需求。

通过算法所提取特征的优劣将直接影响到分类结果的准确性。由于目标实例个体本身所具有的形态多样性，所处周围环境的光照变化、背景等多方面的影响，提取鲁棒性的特征不容易。方向梯度直方图特征是一种特征描述子，这种特征描述种子被用于解决计算机视觉中的物体检测问题。这种方法的缺点是描述子生成过程冗长，这会导致速度变慢、实时性降低、遮挡的问题较难处理，同时因为梯度的相关性质，描述子对于噪点的存在敏感异常。该方法具有尺度和光照不变性，但是由于其过于依赖于局部区域像素的梯度方向，会造成后面特征匹配时放大误差，从而导致匹配不成功，而且这种方法忽略了色彩信息。

目标检测中的最后一个步骤是分类，主要有 SVM^[13]与 AdaBoost^[14]方法。SVM 是为了使得训练集中的正样本与负样本的间隔达到最大，于是在样本数据空间找到最佳的分离超平面。SVM 是一种经典的学习方法，这种学习方法是基于小样本的学习，且由于只由少数的支持向量决定决策函数，所以避免了“维数灾难”问题。但是 SVM 算法在遇到具有较大的训练样本与多分类问题时，分类性能不是十分理想。AdaBoost 算法的基本原理就是将多个弱分类器进行结合，使结合后的弱分类器成为一个新的强分类器。AdaBoost 算法简单，不必进行特征筛选或出现过度拟合的问题，但是容易受到噪声干扰，训练时间过长。

2.1.2 基于深度学习的目标检测

基于深度学习的目标检测方法根据设计思想可分为两大类：一阶段目标检测算法、二阶段目标检测算法，它们承载着不一样的算法思路：

(1) 一阶段目标检测算法：算法的中心思想是通过回归的方式对相机捕捉到的目标物体的图像信息进行操作，最后得到这个目标物体的边框和所属的类别。具体来说，这种检测算法首先会在一个大致的范围内进行分类，一般来说最开始都是从全图开始分类，通过多次迭代得到一个精细的位置数据信息。此类算法的最大优势在于速度较快，比较典型的算法有 YOLO^[15]、SSD^[16]。将待检测图像大小进行改动作作为输入是一阶段目标检测算法速度快的原因之一，但是若是使用这种方法，必定会损失许多的信息和一定的精度。

(2) 二阶段目标检测算法：二阶段目标检测算法的核心是将前景与后景区分开来，通过这种发哪个是就可以对样本进行选择性的挑选，这种方式的挑选可以均衡正样本与负样本的数量，训练范围的难度也比直接做混合的分类和回归简单很多。二阶段目标检测算法主要倾向于精确度高。其中最为代表性的即为 R-CNN 家族，特征图的使用是这个家族发展过程的主线。经典的算法包括有 R-CNN^[17]、Fast R-CNN^[18]、Faster R-CNN^[19]等。R-CNN 可以说是卷积神经网络用于目标检测方法的开山之作。Fast R-CNN 是将原来 R-CNN 中的串联式结构改为并行式结构，即在分类的同时，对边界框进行回归，最终实验效果表明 Fast R-CNN 不仅加快了预测的速度，而且大大提高了精度。在 Faster R-CNN 中首次提出了区域生成网络的概念，Faster R-CNN 可以利用神经网络自适应地生成候选区域，这也是此算法可以学习到更高层、抽象的特征原因。二阶段目标检测算法从最开始的串行结构发展到后来的并行结构，从最开始的单一信息发展到后来的三流信息，虽然可以保证物体识别的精确度。但是却无法达到一阶段目标检测算法的检测速度。

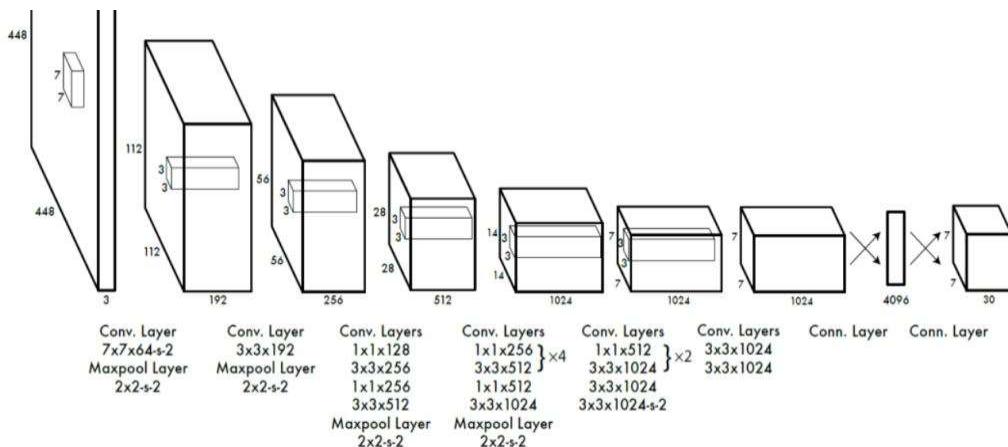


图 2-3 YOLO 网络结构^[15]



图 2-4 R-CNN 过程

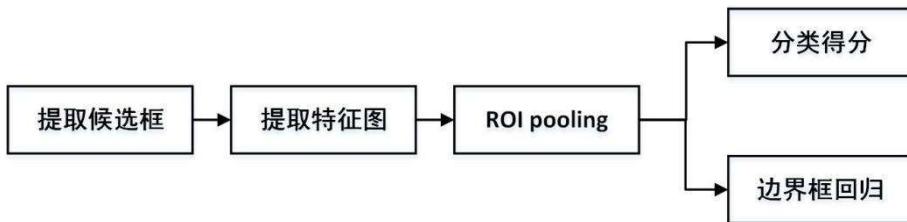


图 2-5 Fast R-CNN 过程示意图

目标检测过程中有两个重要的评价标准：速度和准确率。一阶段目标检测算法的优势是速度快；二阶段目标检测算法优点是准确率高，虽然没有一阶段目标检测算法那么快，但是也达到了每秒 6 帧的速度。两类目标检测算法都有其适用的范围，一阶段目标检测算法更适用于实时快速动作捕捉；而在复杂、多物体重叠的环境中，二阶段目标检测算法更加适用。因此，选择搭载在机器人“眼”中的算法要根据机器人所需处理的环境进行选择。

2.2 目标抓取

在经过机器人的“眼”识别物体与通过“眼”建立好空间物体与图像之间的关系后，就可以进行机器人与物体的交互，实现抓取等动作。

2.2.1 传统目标抓取方法

在传统的机器人抓取过程中，通过手眼标定得到物体实例与其图像之间的关系后，可进行下一步抓取动作的实现。在科研人员对抓取动作研究的早期，对稳定抓取的评价标准有如下两方面：一是抓取动作是否满足力封闭，二是通过放置手指是否可以限制物体的位置。

可以通过形态学提取并处理特征点来得到物体的中心抓取点位置^[20]。对于规则形状的物体，可以通过二值图像，使用基于区域和边缘的中心定位方法，使目标的中心、形心与质心重合^[21]。

2.2.1 基于深度学习的目标抓取方法

虽然使用传统方法能够实现机器人的抓取，但是想要设计出能够对周围环境进行观测，并可以在无论环境可知与否都可以做出应对的机器人的难度指数非常高。

现在的研究结果表明，通过结合深度学习与强化学习可以帮助机器人从周围环境中通过大量的数据与自身经验不断进行学习。深度学习在面对实际中的非结构化场景时有着良好的解决能力；强化学习可以通过算法让机器人在决策过程中做出鲁棒性更好的决策。

Google 设计了一种算法^[22]，此算法是通过大型数据集来获得最优解的。但是该算

法存在着无法对长期行为进行判断的问题，举例来说，若多个物体聚集在一起时，将其中的一个物体单独分离进行抓取是更容易的，但是将物体单独分离这一过程是一个长期训练与学习的过程。谷歌的研究人员提出 QT-Opt^[23]算法，在这个算法的实现过程中，研究人员首先会使用收集到的数据对模型进行离线的训练，由于在离线训练过程中未使用机器人实体，这使得该算法在分布式训练方面占有很大优势；接着，研究人员将离线训练好的模型部署到机器人实体上，同时对算法进行微调，而在对这个过程中使用新的数据集对算法进行训练。3D 神经网络 PointNet 可以实现对抓取器内部的点云，进行抓取质量估计^[24]。还可以通过使用大规模虚拟数据进行候选抓取位置的编码^[25]，通过这种方式，给定单视角下的点云可以解码出少量质量较高的候选抓取位置，然后方法使用 PointNet+ 网络对抓取器与单视角点云一起进行评估抓取质量评估；最后此算法还对最终的抓取位置进行了一定程度的优化。针对于 25DoF 的手形抓取器的算法也被提了出来^[26]，此算法特别依赖于物体的完整模型，此算法首先使用通过 GraspIt！生成的训练集，然后通过训练神经网络来得到抓取姿态。现阶段想要实现人形抓手的通用抓取的难度过高，目前只可实现类内的通用抓取。在实现抓取时也会首先构建基于深度图的抓取质量数据集，然后通过训练来得到关于抓取质量的评估网络^[27]，当进行在线使用时，会分割目标物体在当前视角下对应的深度图，图中每一个抓取位置都会得到一个抓取质量，算法会选择质量最高抓取位置对物体的进行抓取。实现抓取时也可使用形状补全技术^[28]，此算法的形状补全是通过 3D CNN 实现的，虽然此算法在相似的形状中有良好的表现，但是在现阶段无法做到任意通用。

2.3 本章小结

本章对基于传统方法与基于深度学习的方法进的目标识别与目标抓取方法分别进行描述。传统的目标检测算法中基于滑动窗口法首先存在着时间复杂度较大、窗口冗余的问题。其次，传统的目标检测方法没有反馈，只是单向地分析总结一幅图像的特征，没有实践反馈。基于深度学习的目标识别方法虽然包含了训练与测试的过程，但是算法的鲁棒性较传统的目标检测方法来说要更好。传统的抓取方法与基于深度学习的方法分别在解决特定问题时有各自的用处。

第三章 基于单目视觉的目标检测

在现实中，同一幅图像中通常有一个、两个甚至多个目标对象，当不同类别的多个目标对象共同存在于同一幅图像时，计算机不可能将这个图像简单的只分类为同一个类别，所以此时就有必要对图像中的每个目标对象的类别进行区分，那么这个问题就变成了如何寻找目标对象的位置，也就是对目标物体进行定位。与定位单个目标对象的位置不同的是，多个目标的识别就涉及对图像中的所有对象进行定位和分类。本章采用的是在 ROS 操作系统下采用基于深度学习的目标检测方法 SSD 模型对物体进行识别，并进行了相关的实验验证。

3.1 深度学习相关原理

3.1.1 深度学习

深度学习其实是深度神经网络的代名词，起源于人工神经网络的相关研究。深度学习的相关网络结构随着研究的深入越来越复杂。深度学习尝试对所需处理的数据的高层次信息进行有效的建模^[29]，通过这种方式，深度学习就可以对事件进行有效的预测或对事件进行决断。深度学习的网络结构有前馈网络、卷积神经网络、递归神经网络三种。

Alex Krizhevsky 等研究人员在 2012 年将卷积神经网络这种结构成功应用到计算机视觉的相关领域，卷积神经网络由此也引起相关研究人员的关注。吴恩达将深度学习描述为：通过使用大脑模拟器，使用优化学习算法会在机器学习和人工智能领域取得革命性进步的技术成果，吴恩达还创立了“Google Brain”，通过使用“Google Brain”可以在谷歌的大量服务里面实现深度学习技术相关的产品化。自此，卷积神经网络凭借其在自动特征提取领域能力的优秀表现，在各个领域方向被广泛的应用。

神经网络的学习过程的过程可以表述为：权重的初始化也可以看作是对于初始化参数，也就是说，通过对已经定义好的隐藏层的数据信息进行变换操作来获得对于数据信息的预测值；实际值与预期值的误差的计算可以用于测量输出与期望值之间的距离；的损失值可以作为反馈信号来对权重进行调整，调整权重的目的是减少损失值，这种类型的调整是通过优化程序实现的。深度学习的整个过程就是通过连续的训练来降低损失值的大小同时对整个网络的权重进行调整，进而提高深度学习模型的准确性。

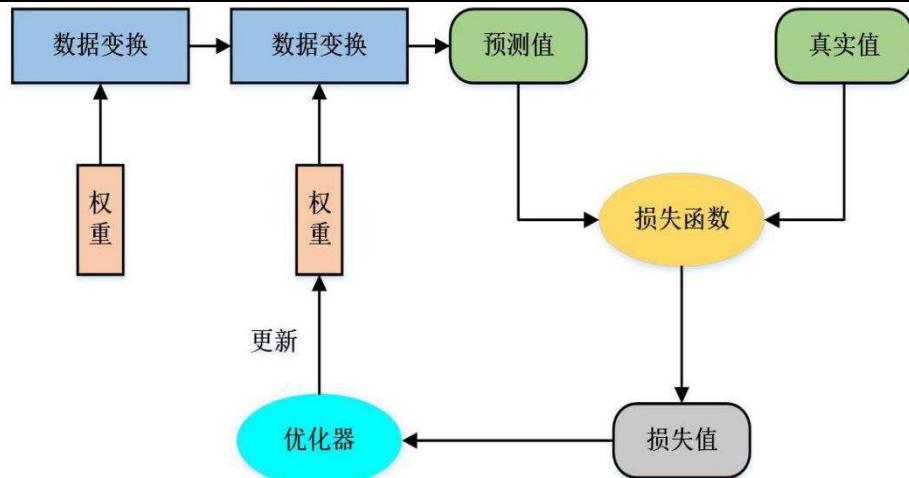


图 3-1 深度学习过程

3.1.2 卷积神经网络

受生物自然视觉认知机制的启发，卷积神经网络应运而生，卷积神经网络的核心是将复杂问题简单化，即对参数降维后再处理^[30]，卷积神经网络最后得到的结果不会受到降维的影响。卷积神经网络同时还对图像数据化后无法对特征信息进行保留的问题进行了解决：卷积神经网络利用类人视觉的方式对图像的特征信息进行了保留，也就是说当图像进行翻转、旋转、变换等操作时，卷积神经网络也可以对图像信息进行准确的识别^[31]。

从提出卷积神经网络到目前的相关应用来看，卷积神经网络共经历了三个阶段。首先在 1962 年，Hubel 提出了从视网膜到大脑的视觉信息在几个层面上都可以通过感受野来刺激来得到的方法；日本研究人员福岛于 1980 年在基于感受野的概念之上提出了一种神经认知机器，这种神经认知机器被认为是卷积神经网络的原型，核心是对视觉系统建模，这种建模方式不会受到视觉对象的位置和大小的影响。接着，计算机科学家 Yann LeCun 等研究人员于 1998 年提出 LeNet-5 网络^[32]，此网络通过基于梯度的反向传播算法进行学习，使用交替连接的卷积层和子采样层将原始图像逐渐转换为一系列特征图，研究人员对卷积神经网络的关注也始于 LeNet-5 网络的提出。直到 2012 年 AlexNet 网络^[33]提出后，卷积神经网络才开始逐渐在深度学习应用程序中逐渐兴起，在 AlexNet 网络后，一些新的 CNN 网络，如 VGG、ResNet^[34]、GoogleNet^[35]等也逐渐被提出，这也使得卷积神经网络逐渐地发展为商业应用。

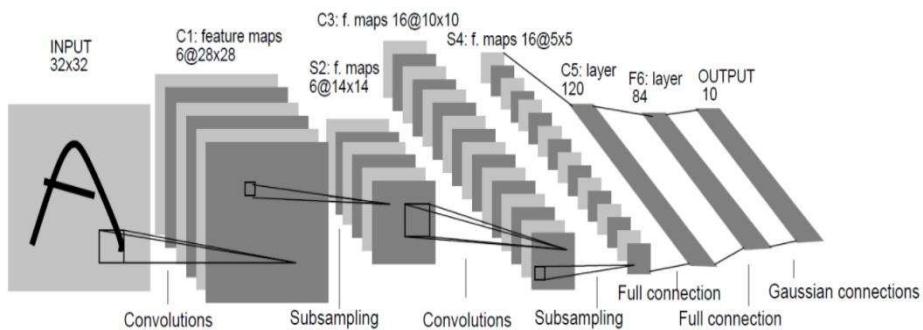


图 3-2 LeNet-5 网络^[31]

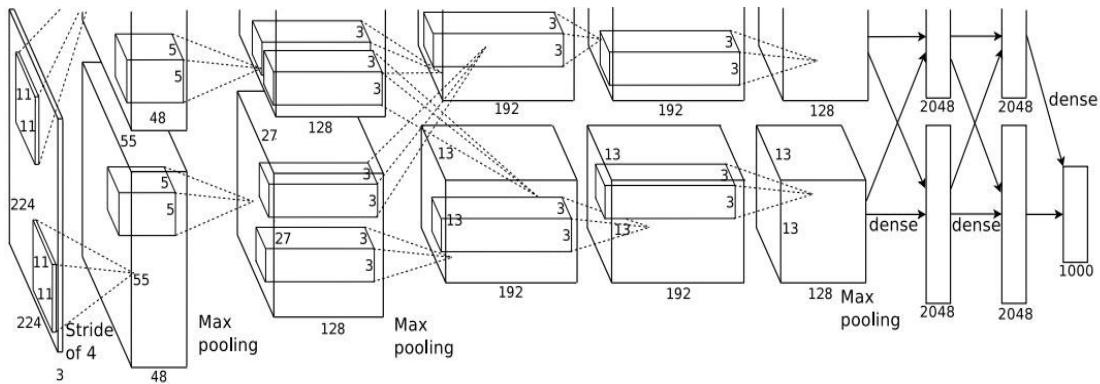


图 3-3 AlexNet 网络^[32]

神经网络模型可以对复杂的非线性函数进行表示，这依赖于多个神经元之间的层次关系来对输出信号和输入信号之间的关系进行表现。神经网络的基本单元是神经元，数个神经元并联并在不同的层级进行组合，就组合形成了基本的人工神经网络。人工神经网络中的隐藏层的层数和其中神经元数目是固定的，输入层和输出层的节点数也是固定的。人工神经网络中每层中神经元都与下一层完全进行连接，但是每一层中的神经元相互之间不进行连接。

卷积神经网络的基本结构包括三层：卷积层使用卷积内核对整个图像进行扫描来对图像的局部特征信息进行提取，这个过程类似于人类对于视觉信息的提取；在池化层的这个过程中不仅可以对数据的计算量进行降低，还可以防止过拟合发生；全连接层也就是对结果进行输出。

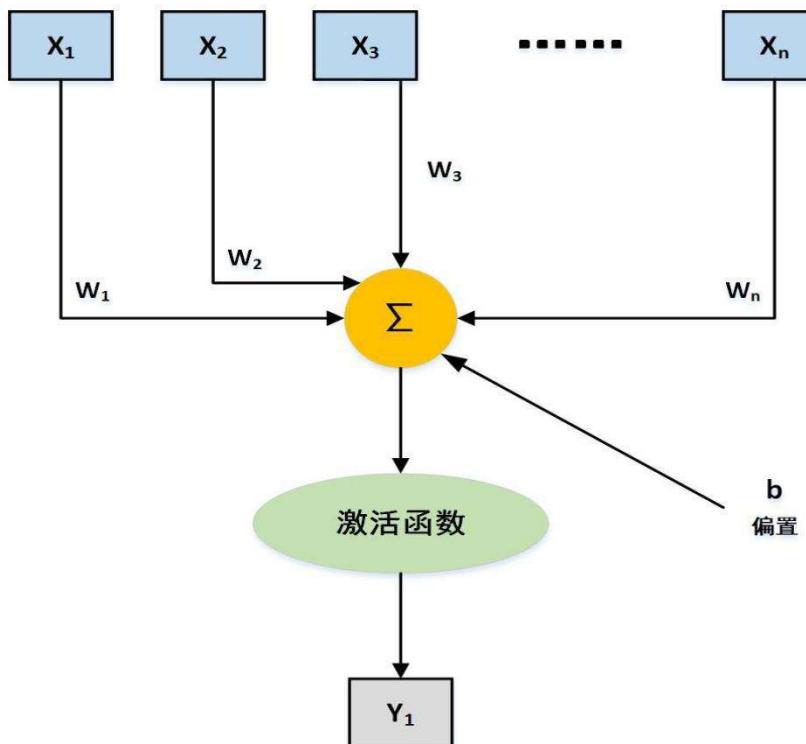


图 3-4 神经元

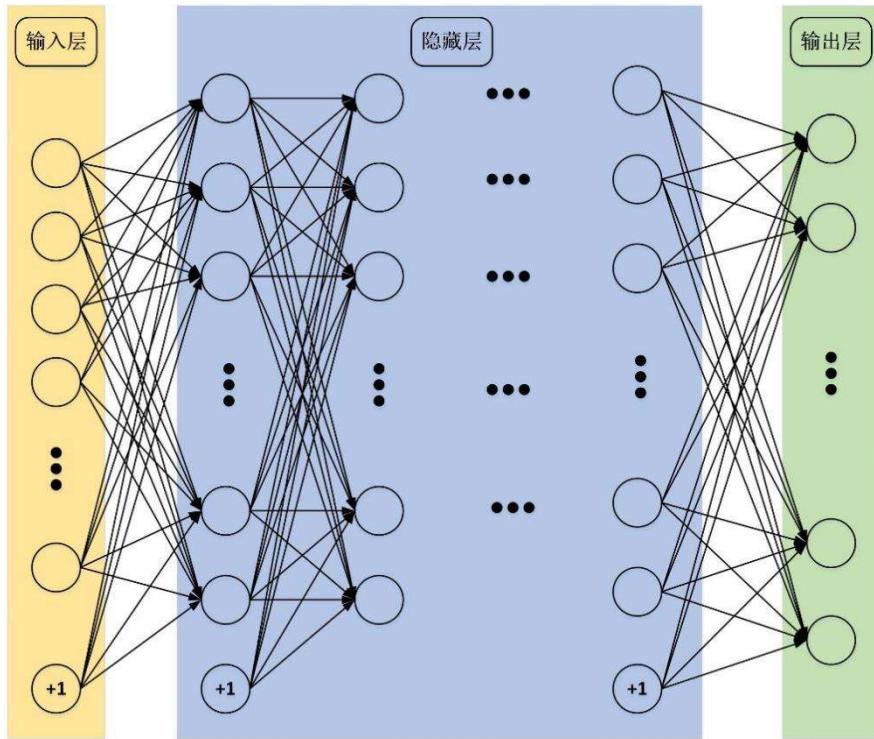


图 3-5 简单的神经网络

卷积神经网络在对图像信息和视频信息方面的处理能力较好，通过深度学习构建的模型可以有效对图像进行分类，即训练有效的模型来对图像中的物体进行识别。与传统的方法相比较而言，卷积神经网络的优势是避免了对图像的预处理，而在传统的算法层面来看是手动参与图像预处理是必要的步骤，也就是说，卷积神经网络可以对输入的原始图像直接进行处理。目前为止，卷积神经网络已经被广泛用于处理各种图像的相关问题。

3.2 SSD 目标检测算法

SSD 网络模型于 2016 年提出，SSD 模型也是当前主流的目标检测框架。SSD 网络模型在吸收了 YOLO 模型的回归思想的同时，借鉴了 Faster R-CNN 算法中的锚思想。SSD 模型是同时对多个尺度不同的特征图进行预测，同时 SSD 模型使用卷积层来对结果进行提取，通过使用这种方式可以节省参数，降低计算量。SSD 模型在识别速度以及识别准确率两方面都有较好的表现。

3.2.1 基础网络

VGG-16^[36]是一个比较基础的网络结构，与其他的基础网络相比 VGG-16 的优势有：分类性能好、结构规整、对数据集的适应能力较好。SSD 模型就是使用 VGG-16 作为基础网络进行目标检测。VGG-16 的结构有：卷积层、池化层、全连接层，其中卷积层与池化层对特征进行提取，全连接层进行分类，VGG-16 的结构如图 3-6 所示。

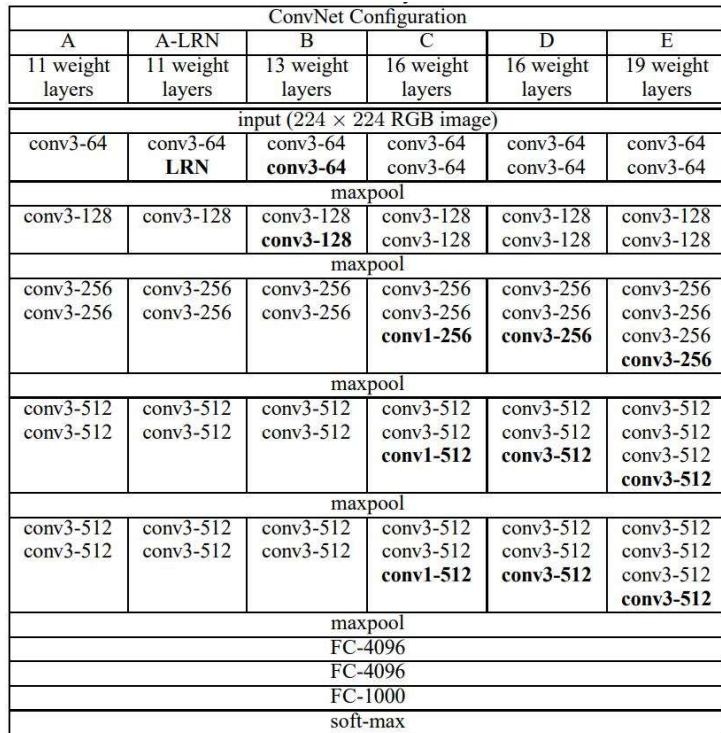


图 3-6 VGG 网络结构^[36]

3.2.2 SSD 网络

SSD 网络是在 VGG-16 网络的基础上进行改进得到的，其中 VGG-16 网络是作为特征的提取器。SSD 网络在 VGG-16 网络后添加了卷积层，并通过卷积核实现预测的过程，具体来说，SSD 网络将 VGG-16 的全连接层的前面两个连接层变换为卷积层，同时将全网络连接层前的池化层的采样核尺寸由 2×2 变换为 3×3 ，另外还去掉了 Dropout 层与全连接层，新增 8 个卷积层，最后对所选取的特征图进行预测。

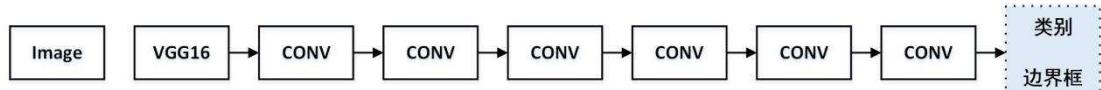


图 3-7 SSD 网络识别过程

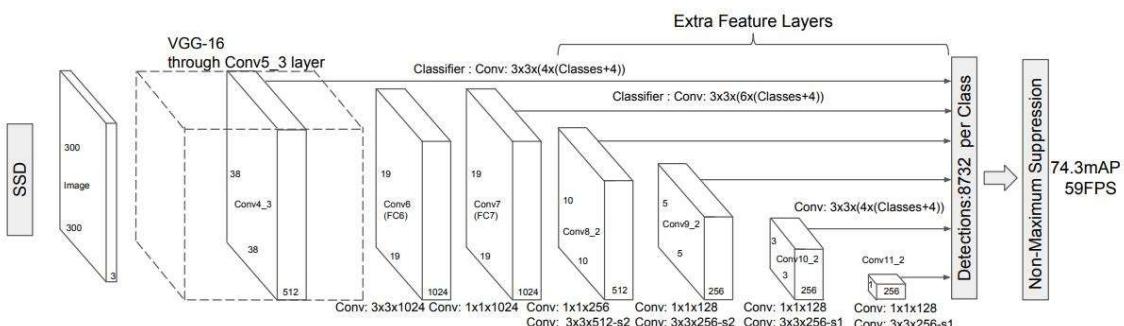


图 3-8 SSD 网络结构^[16]

SSD 网络的核心理念如下：

(1) 使用不同尺度的特征图进行预测。一般来说在卷积神经网络前面的特征图尺寸较大，SSD 网络首先会降低这些特征图的尺寸大小，即以步长为 2 进行卷积计算或进行池化，通过这种方法无论是大的特征图还是小的特征图，都可以进行目标检测。不同

尺寸的特征图对目标进行识别的优点是：小目标可以使用尺寸大的特征图识别，大目标可以使用尺寸小的特征图识别。在图 3-10(a)中，使用浅层网络可以很好对猫这一目标进行识别，同时用蓝框对物体进行框定；但如果对狗这一目标进行识别时，若选择尺寸较小的特征图进行识别，选框就无法很好对目标进行识别；深层网络中，特征图在经过池化层后特征图的尺寸已经大幅度减小，同时感受野变大，这意味着红色选框可以更准确地对狗这一目标进行识别。产生这个结果是因为每一个特征图只能用尺寸相同的选框，如果选框尺寸与所要进行识别的目标尺寸相差过大时，模型就无法进行有效的检测

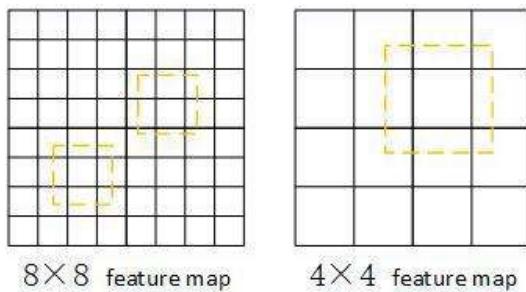
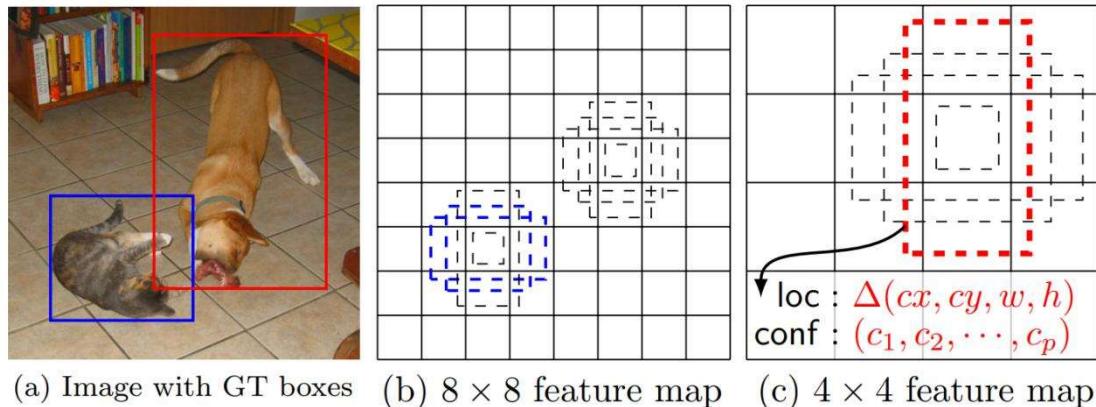


图 3-9 多尺度特征图

图 3-10 多尺度特征图检测^[16]

(2) 设置尺寸相同的默认选框。SSD 模型预测得到的边界框是以这些默认框为标准计算得到的，通过使用默认框可以大幅降低网络在训练过程中的难度。对应于上一小节选用不同尺度的特征图进行预测的方法，可以判断得到默认框就是在特征图上每一个点上选取的具有不同的长宽比的选框。与 YOLO 模型不同的是，YOLO 模型在每一个点只选取正方形的选框，但现实中所需识别的目标的形状大多是不同的，这也就需要 YOLO 模型在其训练时进行自适应调整。特征图的默认框尺寸大小可以按照公式 (3-1) 计算，其中 S_{min} 设定为 0.2，即最底层的尺寸大小为 0.2； S_{max} 设定为 0.9，即最高层的尺寸大小是 0.9，由此可以得出 SSD 模型提取的 6 个特征图的默认框的尺寸大小分别为：30、60、111、162、213、264。通过公式(3-2)、(3-3)可以对默认框的宽度 w 和高度 h 进行计算。另外，SSD 模型还会设置一个先验框，这个先验框的尺寸大小可以由公式(3-4)得到。SSD 模型总计可以得到 8732 个预测的默认框，由此可以说 SSD 模型的计算过程在本质上来说是一个进行密集采样的过程。

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), k \in [1, m] \quad (3-1)$$

$$\mathbf{w}_k^a = s_k \sqrt{a_r} \quad (3-2)$$

$$\mathbf{h}_k^a = \frac{s_k}{\sqrt{a_r}} \quad (3-3)$$

$$\mathbf{s}_k^r = \sqrt{s_k + s_{k+1}}, ar = 1 \quad (3-4)$$

(3) 使用卷积对目标进行识别。SSD 模型直接使用卷积对不同尺寸的特征图来对图像中的物体进行识别。分别用两个尺寸大小为 3×3 的卷积核进行输出，输出内容包括两个：一个输出分类用的置信结果，每个默认框可以生成 21 个置信结果，这个结果是对应于 VOC 数据集中包含的 20 类别与对于背景的分类；另一个输出回归用的位置信息，每个默认框都可以生成 4 个坐标值。

SSD 模型对目标进行识别可以描述为：首先对默认框进行匹配，确定训练图像中的哪个真实目标与哪个先验框进行匹配，匹配原则主要有两点：寻找与每一个真实目标框有最大重叠度的默认框，这样就可以保证每一个真实目标物体至少与一个默认框进行一一对应；SSD 模型还会将还没有配对的默认框与真实目标物体进行配对，只要重叠度大于阈值，就会对两者进行匹配，重叠度定义可见公式(3-5)。默认框若可以与真实目标匹配就可认为是正样本，反之则称之为负样本。每个真实目标尽管可以与多个先验框进行匹配，但是真实目标在数量上来说相对与先验框的数量相比还是过少，会导致负样本的数量要比正样本多。为了保证正负样本在数量上尽可能的保持平衡，SSD 模型采用了非极大值抑制，非极大值抑制的过程就是对负样本进行抽样，在整个抽样的过程中会进行降序排列，排列的标准就是置信度误差值的大小，通过选取排列顺序中的 top-k 做为训练过程中负样本的方式，来保证正样本与负样本比例接近 1:3。

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (3-5)$$

在训练样本确定后，就是需要设置损失函数。损失函数的定义是：位置误差与置信度误差的加权和，如公式(3-6)所示，其中 L_{conf} 是置信度误差、 L_{loc} 是位置误差、 N 是先验框的正样本数量、 c 为类别置信度预测值、 l 是先验框所对应边界框的位置预测值、 g 是真值的位置参数。通过交叉验证将权重系数 α 设置为 1、对位置误差进行设置，对其的定义如公式(3-7)所示。对于置信度误差，采用 softmax loss，定义如公式(3-8)所示，从公式中可以看出，置信度的误差包含正样本的误差和负样本的误差两个部分。

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (3-6)$$

$$L_{loc}(x, l, g) = \sum_{i \in Pos}^N \sum_{m \in [cx, cy, w, h]} x_{ij}^k smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (3-7)$$

$$L_{conf}(x, c) = -\sum x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Pos}^N \log(\hat{c}_i^0) \quad (3-8)$$

实验结果显示，通过数据增强的方式可以明显对算法的性能在一定程度上进行提高。对数据增强更为直观的理解是通过数据增强的操作可以有效在数量上对训练样本进行增加，与此同时可以构造出更多的形状不同、尺寸不同的目标，将其输入到网络模型中去，可使模型学习到更加鲁棒性的特征。

SSD 网络模型通过使用 Hole 算法在对模型进行微调的同时，又可以通过对卷积网络结构进行改变的同时，获得相对于原图来说更为稠密的得分图，Hole 算法的原理如图 3-11 所示，由图我们可以看到 Hole 算法在增加特征图尺寸的同时也对感受野进行了扩展。

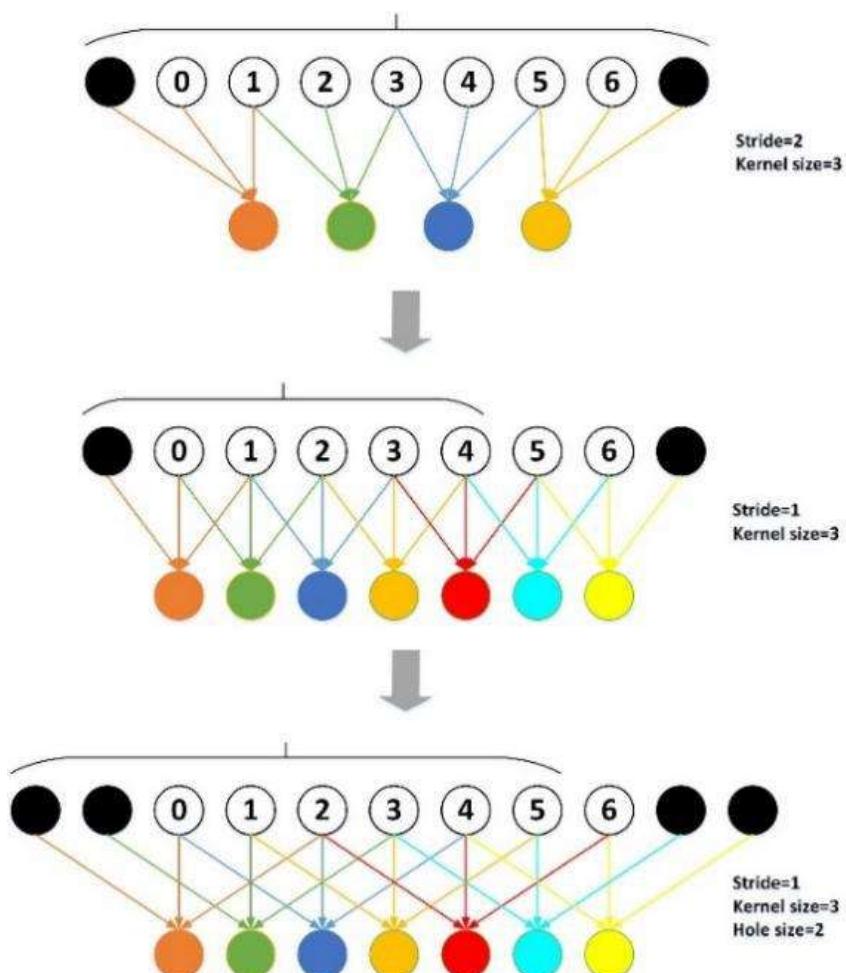


图 3-11 Hole 算法扩大感受野

如图 3-12 所示，在普通的卷积过程或池化过程中，过滤器中的相邻权重作用在特征图上的位置在物理层面上来看都是连续的，为了让感受野不改变，某一层的步长由 2 变成 1 后，后面的层需要采用 Hole 算法，Hole 算法就是根据 hole 的尺寸大小将连续的连接关系变成跳跃连接。

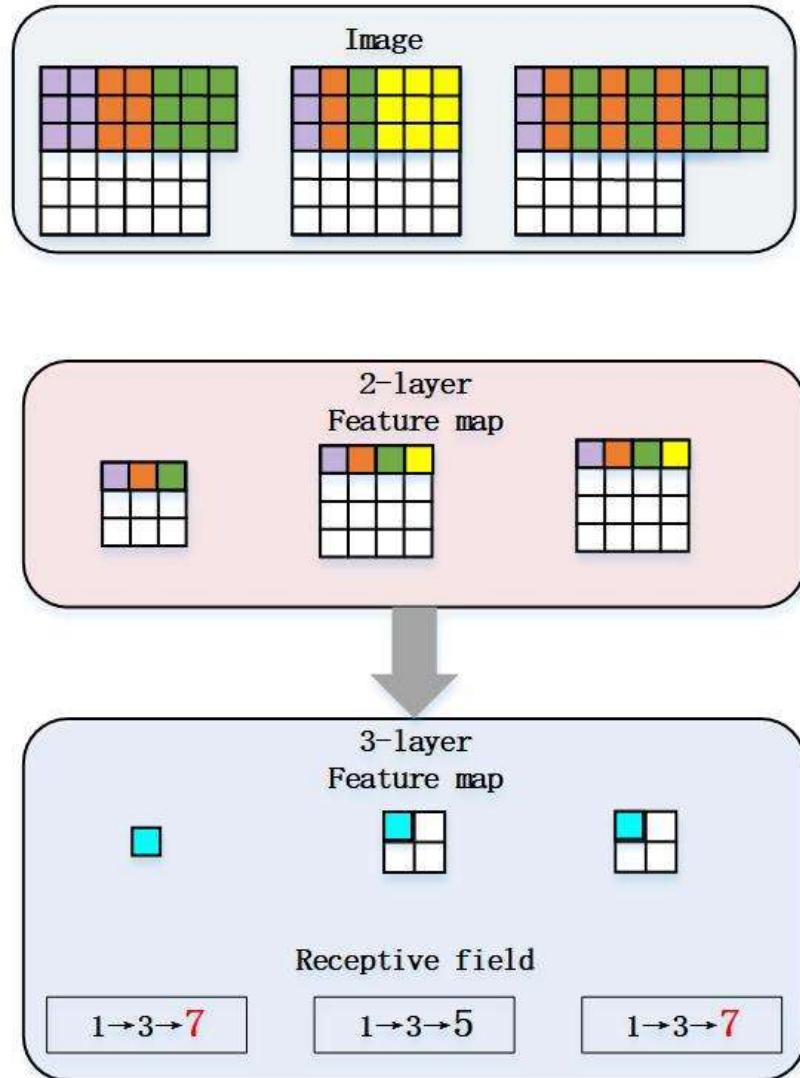


图 3-12 Hole 算法增加特征图尺寸

同时，SSD 网络通过非极大值抑制来实现物体识别，在 SSD 模型中，非极大值抑制的不可取代，这是由于在识别过程中会产生多个特征图，这些特征图会在网络模型中产生大量的边界框，所产生的边界框代表着最终检测结果，但是这些边界框中大多存在错误、重叠、不准确的问题，这些问题不仅会使得计算量大幅增加，而且如果不能对这些问题的边界框进行有效的处理会对算法本身的性能造成负面影响。但是如若对边界框的处理仅依赖于重叠度是不现实的：重叠度的值设置过大，可能就会导致一部分识别目标的丢失，即漏检情况的发生；重叠度值设置过小，则会发生重叠检测，重叠检测也会对检测器的性能产生负面影响。但是即通过重叠度可以处理大部分边界框，此时还需要通过使用非极大值抑制的机制对网络模型进行迭代与优化。

3.2.3 数据集

(1) VOC 数据集^[16]：目标检测中非常经典的数据集之一就是 VOC 数据集。VOC 数据集共有 20 个类别。自 2007 年始，VOC 数据集的层级结构包括四个大类：车辆、

家庭日常、动物、人类，共计 20 个小类，预测时只输出图中黑色粗体的类别。VOC 数据集主要的关注点在分类和检测，至于其他任务如分割、动作识别等，它们的数据集一般是分类和检测数据集的子集。

(2) COCO 数据集^[16]：COCO 数据集是由微软团队提供用来进行物体识别，全称是 Common Objects in Context，数据集不仅包括图像，还包括对图像的 3 种不同标注。其中的图像共包括 91 类目标、328000 个影像、2500000 个标签。数据集包括：训练集、验证集、测试集。

3.3 实验验证

识别算法运行在 ubuntu 系统下，需要安装 ubuntu 18.04，并搭载 ROS 操作系统，为了保证功能的全面实现，安装了 ROS Melodic。在 ROS 操作系统下对于摄像头所捕获的图像处理的具体方法如下：

- (1) 使用 cv_bridge 将 ROS 中的图像消息转换成所需要使用的图像格式；
- (2) 进行图像识别，并且将识别到的物体用矩形框标注出来；
- (3) 转换成 ROS 的图像消息进行发布，提供给 ROS 中的订阅者。

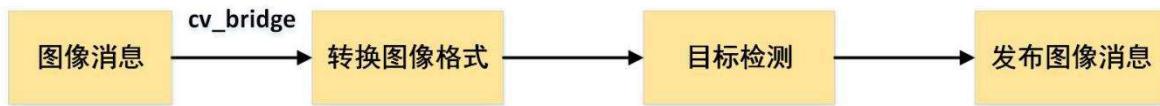


图 3-13 ROS 下的目标检测

算法中 name 与 label 的部分映射如表 3-1 所示。

表 3-1 name 与 label 的映射

Name	id	display_name
/m/01g317	1	person
/m/015qff	10	traffic light
/m/07bgp	20	sheep
/m/080hkjn	31	handbag
/m/03grzl	40	baseball glove
/m/0cmx8	50	spoon
/m/0jy4k	60	donut
/m/09g1w	70	toilet
/m/01k6s3	80	toaster
/m/012xff	90	toothbrush

在 ROS 操作系统下进行实时目标检测的实验结果如图 3-14、3-15、3-16、3-17、3-18、3-19、3-20、3-21、3-22、3-23 所示，由于 label 中没有木块这一类别所以无法显示出准确的对应类别，图 3-24 表示在 ROS 操作系统下以识别瓶子类别为例时所获得的实时坐标。

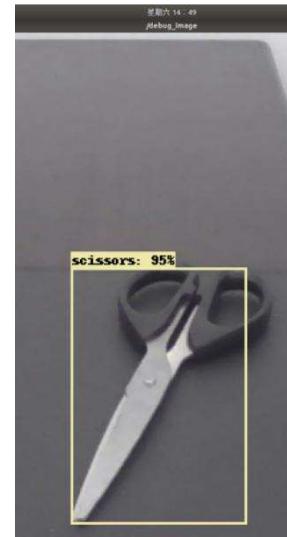
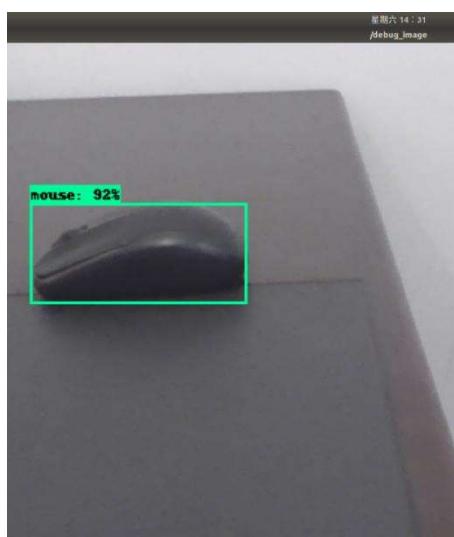


图 3-14 识别结果 a

图 3-15 识别结果 b

图 3-16 识别结果 c

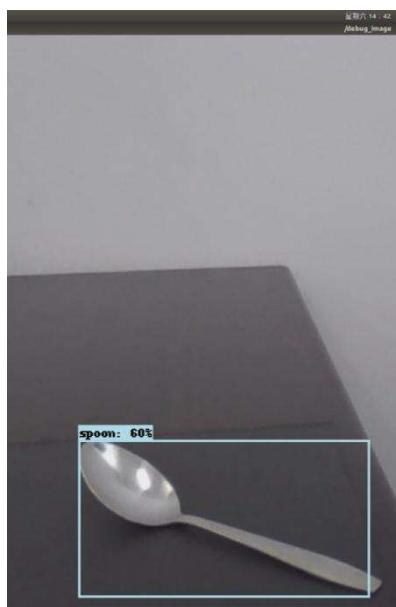


图 3-17 识别结果 d

图 3-18 识别结果 e

图 3-19 识别结果 f

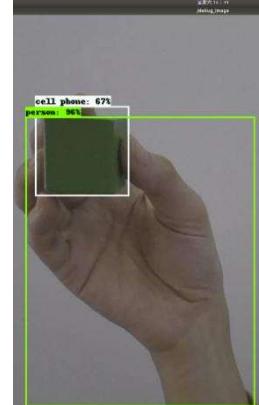
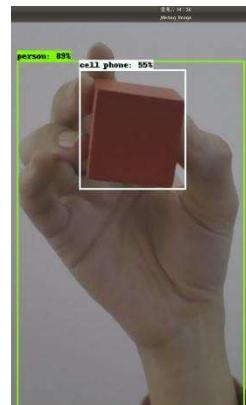
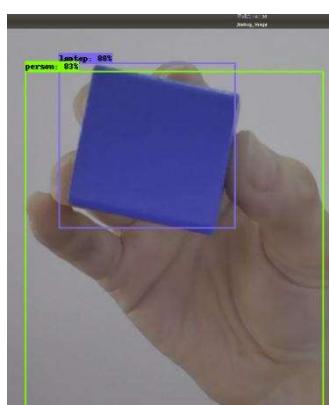


图 3-20 识别结果 g

图 3-21 识别结果 h

图 3-22 识别结果 j



图 3-23 识别结果 10

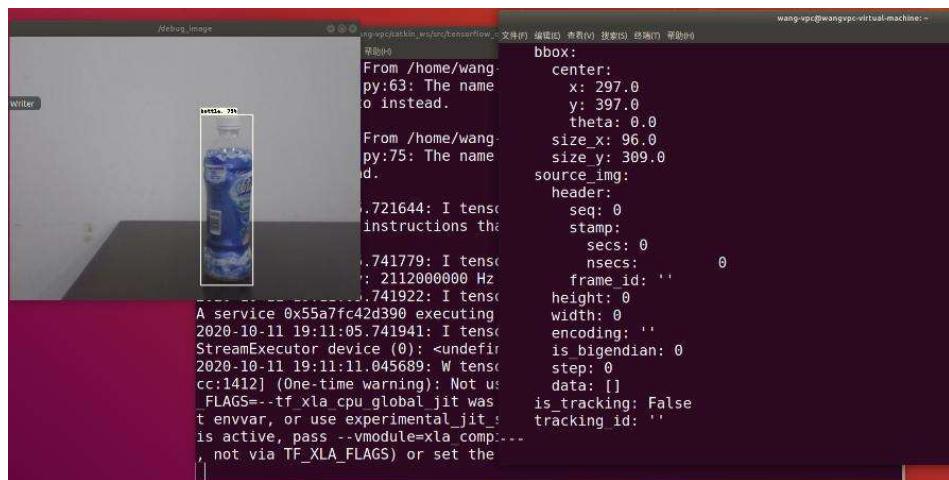


图 3-24 识别结果 2

在识别图像的同时所获得的与识别框相关的坐标信息如下表：

表 3-2 与标注框相关的坐标信息

size_x	size_y	center_x	center_y
96.0	309.0	297.0	397.0

核心代码段如下：

```
def image_cb(self, data):
    objArray = Detection2DArray()
    try:
        cv_image = self.bridge.imgmsg_to_cv2(data, "bgr8")
    except CvBridgeError as e:
        print(e)
    image=cv2.cvtColor(cv_image,cv2.COLOR_BGR2RGB)
    image_np = np.asarray(image)
    image_np_expanded = np.expand_dims(image_np, axis=0)
    image_tensor = detection_graph.get_tensor_by_name('image_tensor:0')
    boxes = detection_graph.get_tensor_by_name('detection_boxes:0')
    scores = detection_graph.get_tensor_by_name('detection_scores:0')
    classes = detection_graph.get_tensor_by_name('detection_classes:0')
    num_detections = detection_graph.get_tensor_by_name('num_detections:0')
    (boxes, scores, classes, num_detections) = self.sess.run([boxes, scores, classes, num_detections],
        feed_dict={image_tensor: image_np_expanded})
    objects=vis_util.visualize_boxes_and_labels_on_image_array(
        image,
        np.squeeze(boxes),
        np.squeeze(classes).astype(np.int32),
        np.squeeze(scores),
        category_index,
        use_normalized_coordinates=True,
        line_thickness=2)
```

3.4 本章小结

本章研究了深度学习与物体目标检测的基础，讨论了 SSD 网络模型，并通过 ROS 的话题订阅与发布的通讯机制实现 SSD 模型在机器人操作系统下对物体的实时识别，并在机器人设备上进行了实验验证。

第四章 自主导航与目标抓取

在机器人对物体进行识别后，需要行进到目标物体的位置后再进行抓取动作的实现。机器人接近目标物体时，即到达目的地需要机器人对目标物体周围的地图进行构建并实现自主导航；在行进到目标物体附近时候就会对目标物体进行抓取，在实现抓取动作时就需要对抓取过程进行设计。

4.1 SLAM 建图

4.1.1 理论概述

随着机器人技术的飞速发展，机器人逐渐进入千家万户的生活中去，这使得机器人的相关产品受到了大众的广泛关注。然而，在机器人落地成为可以更智能的为人们的生活服务之前仍有许多问题亟待解决，其中定位和导航是最为重要的问题之一。在研究这类问题前，必须先要理解三个要点：第一个是对地图进行精确的建模；第二个是机器人可以准确的实现定位；第三个是对路线进行实时的规划，相关的研究人员针对上述三个关键问题提出了多种有效的解决方案。GPS 可以用于室外的定位和导航，但是室内定位与导航问题更加复杂。许多技术不断涌现，以实现定位和姿态，其中 SLAM 技术逐渐脱颖而出。

“SLAM”的全称为“Simultaneous Localization And Mapping”，即是“即时定位与地图构建”。SLAM^[37-39]最早由 Smith、Self 和 Cheeseman 于 1988 年提出，相关研究人员对其中的理论与实际应用价值进行评估后认为：SLAM 技术是实现真正全自主移动机器人的核心技术。SLAM 技术可以更准确的描述：机器人在自身所处的未知环境中自一个未知点开始移动，机器人根据位置估计和地图在整个移动过程中对机器人自身进行定位，在定位的同时对地图进行增量式的构建，通过这种方式对机器人的自主定位与导航进行功能上的实现^[40]。

想象一个处在陌生环境中的盲人，如果这位盲人想要知道自己周围环境的一般情况时，他必须让他的手作为他的“传感器”不断向周围进行伸展探索，进而对自己周围的障碍进行探索。当然，盲人的“传感器”范围必须每时每刻都在不断变化，同时盲人本身还在大脑中将“传感器”感知到的信息进行整合处理。如果新探索到的环境是之前遇到的环境的时候，盲人不仅会对嵌入在脑海中的地图进行更新，同时还会对当前自身的位置信息进行迭代。当然，在上述情况中，这是一位感知能力十分有限的盲人，因此他

搜索的环境信息相对于整体的环境信息来说不一定是完全正确的，会存在一定的误差，他会根据确定性程度大小来设置要搜索的障碍物的概率值，这个概率值越高，那么在此处与障碍物发生碰撞的可能性也就越大，这也就意味着盲人探索未知环境的时间越多。从本质上来说，上述对于盲人探索的整个过程基本就可以能代表 SLAM 算法的主要过程。

室内机器人的主要应用场景是购物中心、火车站等环境，机器人需要在这些应用场景中实现四处走动的功能并对特定的任务进行执行，此时就会要求机器人具有自主运动与定位能力，这些能力就统称为自主导航。SLAM 与自主导航通常密不可分，这是因为机器人进行自主运动的蓝图就是通过 SLAM 技术生成的。这种类型的问题可以归结为：从起始状态找到目标状态的最佳路径，这也就是说需要在机器人的工作空间中选择一条可以避开障碍物的最优路径。

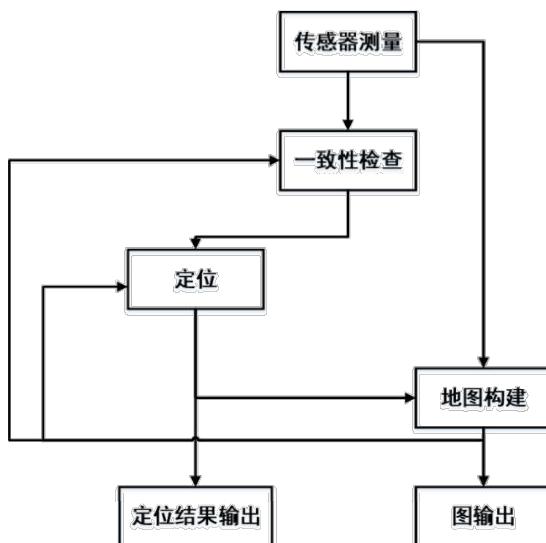


图 4-1 SLAM 过程

4.1.2 获取深度信息的传感器

为了实现机器人的 SLAM 映射，机器人必须首先具有对自身周围环境进行感知的能力，特别是对周围环境的深度信息的感知，因为这些深度信息对于障碍物的检测来说是极为重要的数据。

(1) 激光雷达

激光雷达是目前机器人领域研究最深入、最成熟的深度传感器，通过激光雷达，机器人可以获得与周围环境障碍物之间的距离信息，现在市面上许多常见的扫地机器人都配备了性价比极高的激光雷达传感器。激光雷达具有精度高、响应速度快、数据量小的优点，由于激光雷达存在以上优点就可以使得机器人可以实时执行 SLAM 任务。但是激光雷达缺点是成本较高，进口的精密激光设备的价格超过 10000 元。如今，许多国内公司都专注于研发高精度雷达，并且在当前市场上有许多出色的相关激光雷达产品。

(2) 摄像头

在 SLAM 技术中使用的摄像头有以下两种：单目摄像头与双目摄像头。单目摄像头如字面意思所示，即使用一个独立的摄像进行 SLAM，单目摄像头的传感器较为简单，适用性更强，但是实现 SLAM 的复杂度更高，而且单目摄像头在静止状态下是无法进行距离测量的，只有在目标在运动状态下才能根据三角测量等原理进行距离信息的感知；双目摄像头相比于单目摄像头来说，双目摄像头在运动状态与静止状态下，都可以进行距离信息的感知，但是对于双目摄像头来说，双目摄像头的标定较单目摄像头而言更复杂，这样双目摄像头所得到的图像数据信息会导致运算量更大。

(3) RGB-D 摄像头

近年来兴起的一种更为新型的传感器是 RGB-D 摄像头，RGB-D 摄像头不仅可以和普通摄像头一样获取摄像头周围环境中的彩色图像信息，也可以使用红外结构光、飞行时间等原理得到每个像素的深度信息，RGB-D 摄像头所捕获的数据不仅可用于 SLAM 技术，还可用于图像处理、物体识别等多种应用。但是与此同时 RGB-D 摄像头也具有测量视野窄、盲区大、噪声大等缺点。RGB-D 摄像头的成本与单目摄像头、双目摄像头相比要更低，RGB-D 摄像头也是目前室内服务机器人的主流视觉传感器，较为常见的 RGB-D 摄像头有：Kinect vlv2、Xtion Pro 等。

4.1.3 基于 Rao-Blackwellized 粒子滤波的 SLAM 方法

近几年，与 SLAM 技术相关的很多研究工作被相继提出^[41-46]。SLAM 的核心思想是根据机器人观测得到的值和里程计所测量到的信息来对联合后验概率密度函数进行估计，由此可以看出，机器人所行进的轨迹和所建立的地图需要同时进行计算，很多 SLAM 的相关工作是基于滤波的方法实现的^[47-50]。其中基于 Rao-Blackwellized 的粒子滤波的算法最为经典有效，算法通过公式(4-1)对联合概率密度函数进行因式分解。

$$p(x_{1:t}, m | z_{1:t}, u_{0:t}) = p(m | x_{1:t}, z_{1:t}) p(x_{1:t} | z_{1:t}, u_{0:t}) \quad (4-1)$$

由地图的概率密度函数可以看出，地图对机器人的位姿的依赖性强，通过已知的机器人位姿对地图的概率密度函数进行建图，即“Mapping with known poses”来进行计算得到，通过使用粒子滤波的方式来对后验概率密度函数进行估计。

效果最好粒子滤波的算法即为重要性重采样(Sampling Importance Resampling, SIR)，重要性采样通过以下四个步骤完成：

(1) 预测阶段：粒子滤波会生成采样数信息，这些数据的信息量是由状态预测生成且数量庞大，采样得到的数据信息就被称为粒子，对粒子进行加权和逼近计算就可以得到后验概率密度。

(2) 校正阶段：粒子会循环往复的按观测值顺序对自身的重要性权值进行计算，在这个阶段过程中越有可能观测得到的粒子，这个粒子所获得的权重越高。

(3) 重采样阶段：根据权值的比例将采样粒子重新进行分布，此步骤重要不可略过，这是因为近似逼近连续分布的粒子在数量上非常有限

(4) 地图估计：对每个粒子的轨迹进行观测计算得到的。

SIR 算法需要在新的观测值到达时对粒子的权重进行评估，随着时间的推移，这个过程的计算复杂度将越来越高。因此 Doucet 等学者通过式(4-2)对重要性概率密度函数进行限制来获得一个递归公式，通过这个递归公式来对重要性权值进行计算，权值可通过公式进行(4-3)计算。

$$\pi(x_{1:t} | z_{1:t}, u_{1:t-1}) = \pi(x_t | x_{1:t-1}, z_{1:t}, u_{1:t-1}) \pi(x_t | x_{1:t-1} | z_{1:t-1}, u_{1:t-2}) \quad (4-2)$$

$$w_t^{(1)} = p(x_{1:t}^{(1)} | z_{1:t}, u_{1:t-1}) / \pi(x_{1:t}^{(1)} | z_{1:t}, u_{1:t-1}) \quad (4-3)$$

4.2 机器人导航

4.2.1 理论概述

在上一小节中，通过使用激光雷达对地图进行了建立，所建立的地图为栅格化地图，这种栅格化地图是在导航过程中的机器人常用的地图格式。所谓栅格化地图就是在二维空间上的一种描述机器人周围环境中的障碍物分布状况的地图形式。栅格化地图将机器人周围的整个环境划分为水平和垂直的网格空间，每个网格空间都使用不同的颜色进行标记。如果将每个网格空间放大并进行仔细观看，就可以看到 2D 地图中的每个详细信息都是由一个个小的正方形网格组成，其中白色表示网格空间中没有障碍物，机器人可以通过；黑色表示网格空间中有障碍物，机器人无法通过；灰色则表示机器人未搜索网格空间，有通过的可能性是未知的。

有了栅格地图，机器人的相关地图导航问题就变成机器人在栅格化的地图中寻找可以通行的区域，然后通过电机驱动机器人从起点位置移动到终点位置，其中包含了以下两个任务：

(1) 机器人定位：机器人需要知道自己在地图上当前时刻的位置才能确定导航。在机器人的移动过程中，机器人还必须确定其位置是否仍与预期路径匹配。

(2) 路径规划与导航：路线规划算法会在 2D 的栅格地图中为机器人寻找到一条可遍历的路径，该连续可遍历的栅格从机器人的当前位置可以延伸到导航的目标边缘。

在机器人的实际行进的过程中，机器人不仅要考虑自身可步行的白色网格的连通性，还要对机器人所占据的空间进行考虑，只有白色网格的空间大于机器人框架的直径，机器人才可以通过。因此，在发生碰撞的情况下的安全极限通常是在有黑色障碍物的区域中进行扩展，该限制的宽度极限是机器人的半径，这也就意味着当机器人超过此安全极限时，它将撞到障碍物，同时机器人的可行路径避开了这个安全限制，并为机器人提供了其他的可行进区域。

4.2.2 导航框架

导航主要是由机器人自身定位与机器人的路径规划两大部分组成，图 4-2 对导航的各部分组成进行了阐述说明。

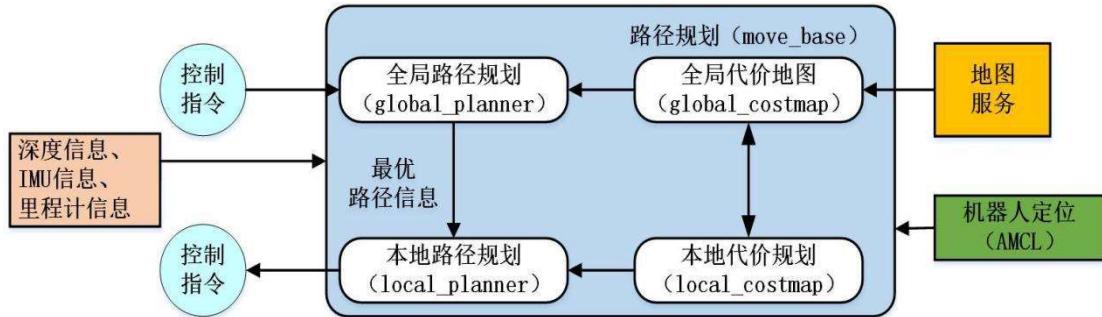


图 4-2 导航总体框架

导航主要由里程计、激光雷达或者深度摄像头的信息、已知的先验地图、坐标系变换信息几个部分组成。整个过程中首先需要对机器人上所搭载的传感器进行信息的采集，进而实现实时避障，为了实现这个目标，机器人需要发布相关的二维激光信息 sensor_msgs/LaserScan 或相关的三维点云 sensor_msgs/PointCloud 的消息；机器人还需要发布里程计信息，即 nav_msgs/Odometry 格式的信息，同时还需要发布与之相对应的 TF 变换；机器人还需要机器人的控制结点具有对中线速度、角速度的进行解析控制的能力，通过控制才可使得机器人进行相关功能的实现，其中定位和路径规划所进行的内容如下：

(1) 机器人定位：通过使用蒙特卡洛自适应定位算法进而实现了机器人的定位功能。蒙特卡洛自适应定位算法是一种使用概率论的方法来对机器人在已知地图中进行自定义位置估计的方法。该方法是在机器人可能运动的位置周围进行多个位置的假设，然后在机器人移动时根据如激光雷达、里程表等之类的信息对假设的位置进行过滤，并逐渐将不值得信任的假设位置进行消除，并将有可能的信息进行保留，同时下载具有更高可能性的位置。机器人会在所有的假想位置散布许多绿色箭头。这些绿色的箭头会随着机器人的不断移动，箭头将逐渐收敛，最终这些绿色箭头会收敛为单个箭头，这个单个箭头也就是机器人最可靠的定位位置。

(2) 路径规划：使用 move_base 进行机器人导航中的最优路径规划的问题研究。move_base 将机器人在导航过程中所需要使用到的地图、坐标、路径、行为规划器等信息进行了有效的连接，同时 move_base 还为机器人提供了设置导航参数的接口，基于 move_base 的导航框架如图所示，具体方法为：

① “/map” 中的 map_server 会为机器人提供有关全局地图的信息，global_costmap 从 “/map” 里获得关于全局的地图信息，全局的地图性信息再结合从激光雷达和点云等其他的信息，融合成一个全局的栅格代价地图，其中激光雷达和点云的信息是从 sensor sources 获得的。

② Global_planner 则进行全局路径规划，全局的路径规划则是通过全局的栅格化地图完成，外部节点会将之结合，并发送移动终点中，移动终点是存在于

“/move_base_simple/goal”，最后机器人会得出全局的移动路径。

③ Global_costmap 中会从 local_costmap 中截取在机器人周围一定范围内的地图，同时这个地图会结合机器人得到的传感器信息生成一个局部的代价地图。

④ Global_planner 会从 local_planner 得到一个全局规划路径，这个全局规划路径会结合机器人获得的局部代价地图与蒙特卡洛自适应定位算法所提供的位置信息，会计算出机器人当前移动的速度，这个速度信息会发送到 “/cmd_vel” 主题中去，接着就会驱动机器人进行移动。

4.3 抓取方法总体概述

本文所建立的抓取方法设计思路如图所示，系统由读取图像、二维像素位姿、手眼标定、最终位姿、运动学逆解、通讯、执行器执行几个部分组成。

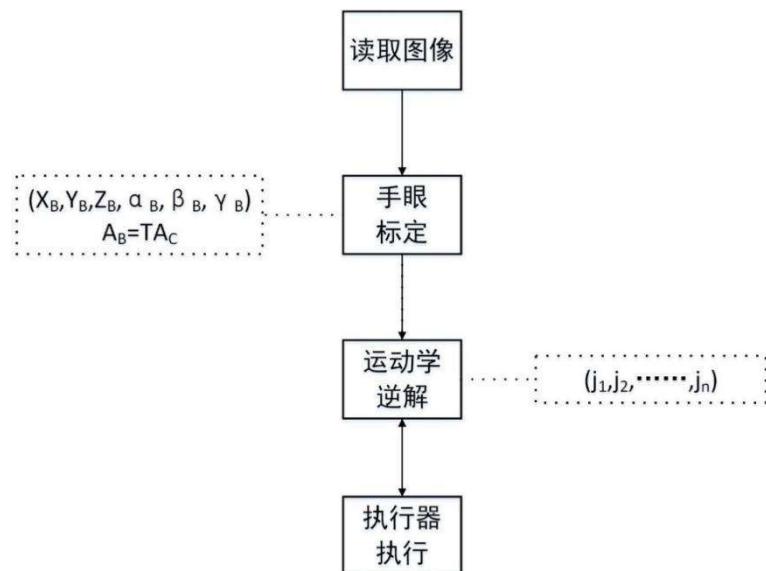


图 4-3 抓取方法设计思路

4.3.1 手眼标定

手眼标定的核心是为了实现真实物体在两个坐标系下的转换：世界坐标系、机器人坐标系。在标定的整个过程中，一般会在平面上先设置一个坐标系，这个坐标系就是世界坐标系，且世界坐标系不重合于机器人坐标系，通过对相机的内参与外参进行标定就可以计算出世界坐标系下物体的具体位置。如果机器人要与视觉进行联动，则需要物体在机器人坐标系下的坐标，根据相机固定的位置分为两种：相机固定在机器人末端，这种情况就称为“眼在手”，如图 4-4 所示；相机固定在机器人外，则称为“眼在外”，如图 4-5 所示。

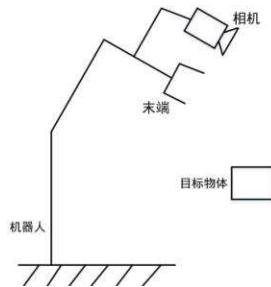


图 4-4 摄像机固定在机器人末端

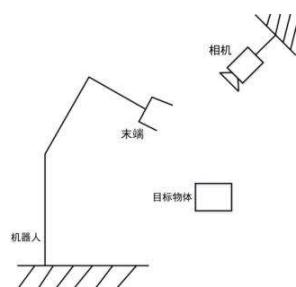


图 4-5 摄像机固定在机器人外面

从机器人的角度来看，摄像头的位置不一定与操纵器的位置相关，并且必须进行转换以允许机器人定位摄像头所占据的位置。成像之间的相关坐标有世界坐标系、相机平面坐标系、成像平面坐标系、图像坐标系，具体如下：

(1) 世界坐标系：任意选择环境中一个基准坐标系，这个基准坐标系就作为参考来对摄像机的位置进行描述，一般选择标定板坐标系作为相机标定中的世界坐标系。

(2) 相机平面坐标系：相机平面坐标系是以针孔模型的聚焦中心 O_C （光心）为原点，摄像机光轴为 Z 轴建立三维坐标系，其中 X_C 轴、 Y_C 轴与成像平面 x、y 轴平行。光心到图像平面的距离为有效焦距 f 。

(3) 成像平面坐标系：成像平面坐标系以主点为原点，主点是光轴与成像平面交点，成像平面坐标系是以物理单位表示的平面直角坐标系。

(4) 图像坐标系：图像坐标系是固定在图像上的平面直角坐标系，它是以像素为单位，且通常来说以像素原点设定为图像的左上角，此时会定义一个直角笛卡尔坐标系： $u-v$ 。图像的信息是以数列的形式进行存储，其中每一像素的 (u, v) 分别对应于该像素在图像数组中的行数和列数。图 4-6 即可对图像中的笛卡尔坐标系进行表示。

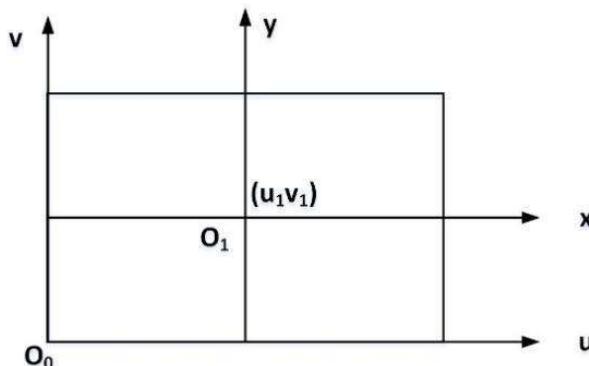


图 4-6 图像的笛卡尔坐标系

摄像机的成像模型如图 4-7 所示：

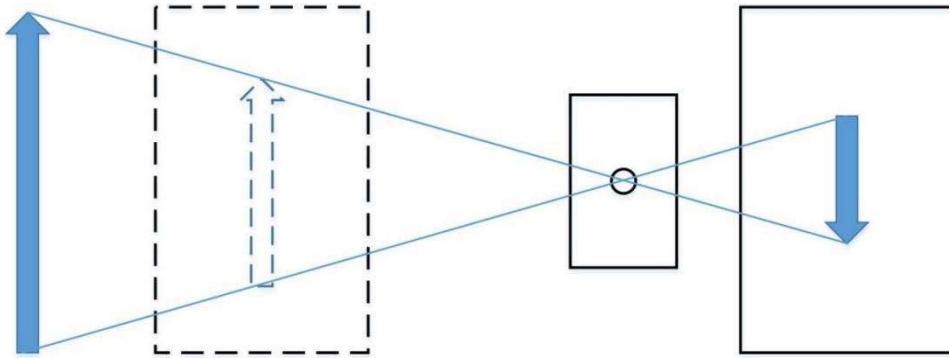


图 4-7 摄像机成像模型

4.3.2 逆运动学分析

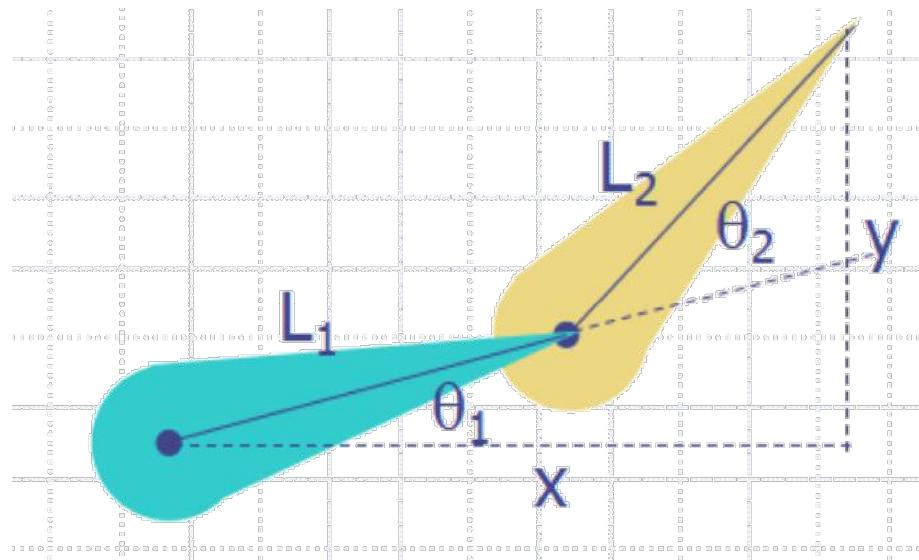
机械臂实际应用中，需要解算到达目标位姿时，各关节所处角度，即本节将分析的机械臂逆运动学。若将机器人正运动学用公式表示为：

$$\mathbf{x} = f(\mathbf{q}) \quad (4-4)$$

则机械臂逆运动学可以表示为：

$$\mathbf{q} = f^{-1}(\mathbf{x}) \quad (4-5)$$

求解机械臂逆运动学普遍采取的两种方案分别是：数值解法，也称迭代解法、解析解法。机器人大学的教材上一般只对运动学逆解的解析解法进行探讨，对数值解也称迭代解法提及较少，原因如下：一是相对于解析解来说数值解所需要的运算量更大，不适用于实时性要求较高的场合；二是一般在对机器人结构进行设计时会考虑其逆解的可解性，如相邻的三个关节轴线相交即存在解析解。下面以最简单的二连杆结构为例子，对机器人运动学逆解的数值解法进行解析，如图 4-8 所示，很容易就可以得到运动学正解，雅克比矩阵 \mathbf{J} 就可以由此进行计算得到。

图 4-8 二连杆结构^[51]

$$\mathbf{x} = L_1 \cos \theta_1 + L_2 \cos(\theta_1 + \theta_2) \quad (4-6)$$

$$y = L_1 \sin \theta_1 + L_2 \sin(\theta_1 + \theta_2) \quad (4-7)$$

$$J = \begin{bmatrix} \frac{\partial x}{\partial \theta_1} & \frac{\partial x}{\partial \theta_2} \\ \frac{\partial y}{\partial \theta_1} & \frac{\partial y}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} -L_1 \sin \theta_1 - L_2 \sin(\theta_1 + \theta_2) & -L_2 \sin(\theta_1 + \theta_2) \\ -L_1 \cos \theta_1 + L_2 \cos(\theta_1 + \theta_2) & L_2 \cos(\theta_1 + \theta_2) \end{bmatrix} \quad (4-8)$$

雅克比矩阵可比作为函数 $f(x)$ 的一阶导数，即可以说是线性近似。可以通过解析法和数值法对雅克比矩阵进行计算，当通过解析法和数值法都无法得到雅克比时可以使用差分法对微分代替进行求解。Jacobian Transpose 的思想是最速降法也称梯度法，此方法利用目标函数在迭代点的局部性态，每步搜索都沿着函数值下降最快的方向，即负梯度方向进行搜索。整体的迭代具有以下优点：几何概念直观、方法简单，程序实现简单；缺点则是在每次的迭代过程中需要沿着迭代点按照负梯度方向进行搜索，这就会导致整体搜索路径更为曲折，收敛速度更慢。

通常的，解析法需要对机器人构建各自相关的解析表达式，但工业机器人却可以通过对的参数化描述得到相似结构机器人的统一的求解模型。此方法常见于各类机器人学教材，主要通过几何分析、代数运算等手段，如三角变换、消元法等方法进行求解。举例来说 ABB4600 和 KUKA240 的 DH 参数实际上是相同的，进一步地讲，理论上其实是可以对 6 轴机器人逆运动学进行统一的求解，前提是将所有参数进行有效的描述，该工作实际上已经在 Peter Corke 的 Robotics Toolbox 上完成了初步实现。IKfast 算法于 2010 年由 Rosen Diankov 提出，虽然此算法也是基于解析法，但其适用范围更为广阔，适用于大多数的机器人。析法的优点有：速度快、精度高、能得到所有解。但是同时也存在一些缺点：通用性差、适用性弱即只针对于一些特定结构可解。

4.4 实验验证

建图与导航算法运行在 ubuntu 系统下，需要安装 ubuntu 16.04，为了方便后期在移动小车上验证整个算法的可行性，搭载 ROS 操作系统。为了保证功能的全面实现，安装了 ROS Kinetic。

节点的流程框架如下图所示，机器人的底盘的控制驱动程序会将机器人底盘的相关的里程计信息发布到平台中，在机器人底盘的移动终点位置设置后，将发含有终点位置的信息话题发布到 ROS 中，接着负责导航的节点将订阅此话题已经发布的相关的里程计信息的话题，并根据里程计数数据信息进行路径规划。为了可以控制机器人的底盘的运行速度，导航的功能节点将发布期望速度的话题，机器人底盘将使得机器人的驱动程序对这个话题进行订阅。于是，机器人底盘就可以按照算法规划好的路径进行行进运动。机器人底盘模型、机器人底盘控制驱动与机器人底盘的导航功能包也将发布坐标变化

话题。

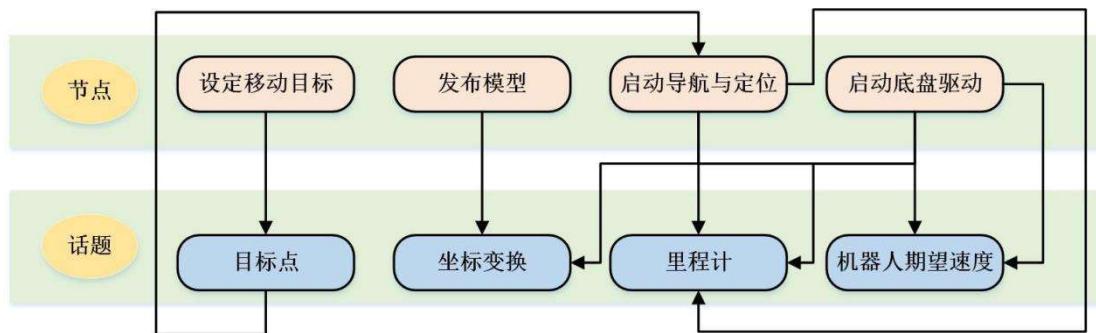


图 4-9 节点流程

具体方法如下：

- (1) 将机器人的底盘与上位机连接到一起。
- (2) 打开机器人在底盘上设置的电源开关。
- (3) 旋转打开底盘的急停开关，搭载在底盘上的电机会处于抱死状态，如果人力强行对机器人底盘作用力，机器人会受到阻力。
- (4) 启动建立地图的 launch 文件。
- (5) 移动机器人即可实时进行建图，实时建图过程如图 4-10 所示，这个过程中会建立好栅格化地图，最终得到的地图如图 4-11 所示。



图 4-10 实时建图过程

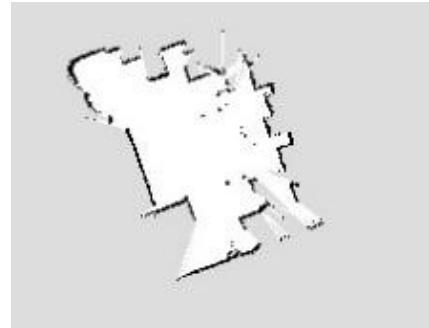


图 4-11 建立好的栅格化地图

核心代码段如下：

```
<node name="joint_state_publisher" pkg="joint_state_publisher" type="joint_state_publisher">
<rosparam file="$(find wpb_home_bringup)/config/wpb_home.yaml" command="load" />
</node>
<node name="robot_state_publisher" pkg="robot_state_publisher" type="robot_state_publisher"/>
<node name="rviz" pkg="rviz" type="rviz" args="-d $(arg rvizconfig)" required="true" />

<node pkg="hector_mapping" type="hector_mapping" name="hector_mapping" output="screen">
<!-- Frame names -->
<param name="pub_map_odom_transform" value="true"/>
<param name="map_frame" value="map" />
<param name="base_frame" value="base_footprint" />
<param name="odom_frame" value="base_footprint" />
```

导航的具体方法如下：

- (1) 对上位机与机器人底盘进行有线连接。
- (2) 打开机器人在底盘上设置的电源开关。
- (3) 启动 Rviz。
- (4) 给机器人设定一个初始位置，程序可以自主为机器人导航到的目标地点，选择完机器人的朝向后，全局规划器会如图 4-12 所示规划出紫色的路径，这条路径从机器人初始位置出发，到移动目标点结束。
- (5) 路径规划完毕后，现实世界里的机器人会如图 4-13、4-14、4-15 所示开始沿着规划好的路径进行移动，一直运行到终点位置。

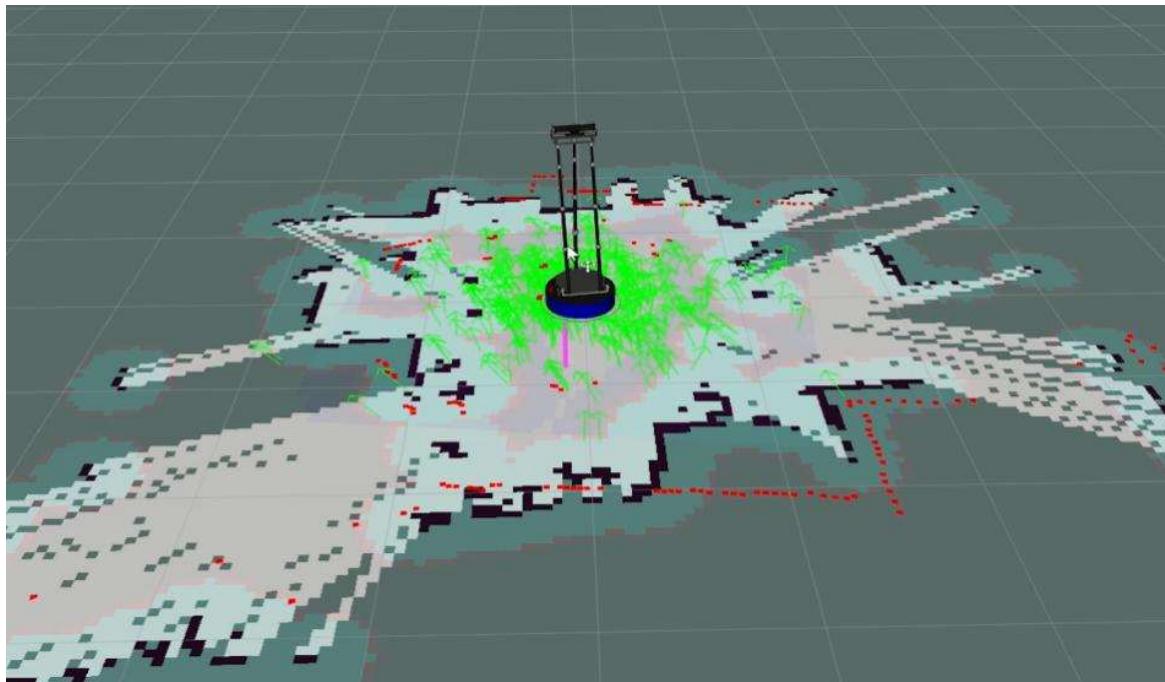


图 4-12 在 Rviz 中的导航过程



图 4-13 实际导航过程 1

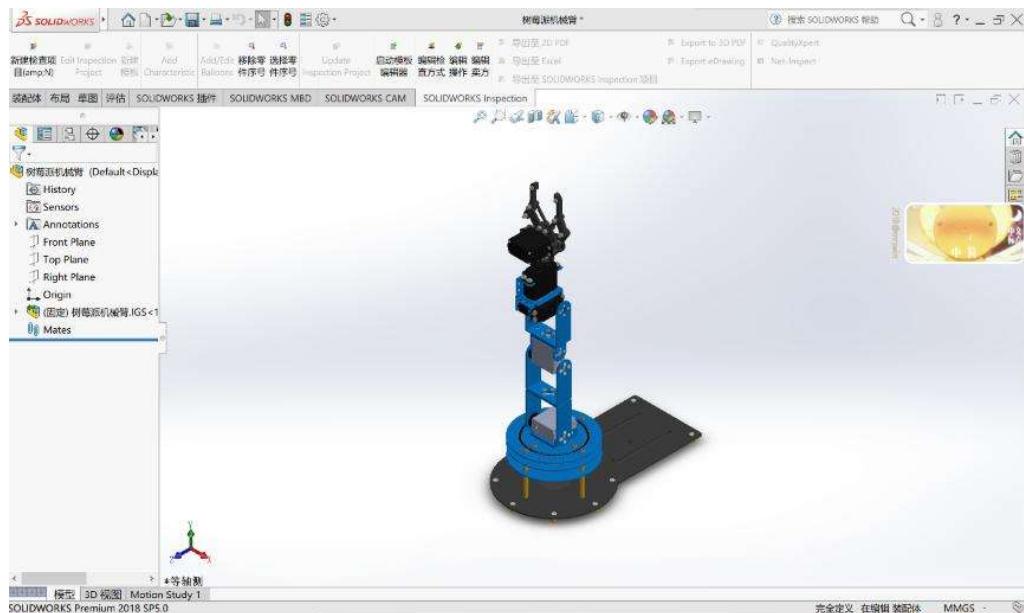


图 4-14 实际导航过程 2



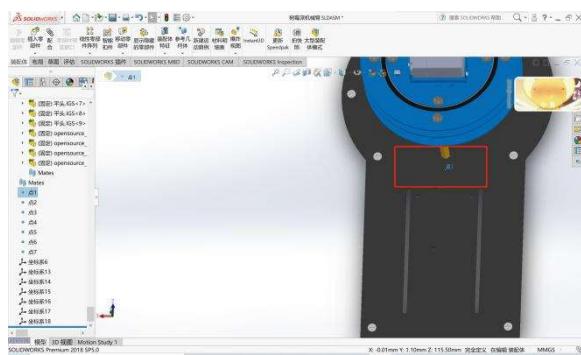
图 4-15 实际导航过程 3

由以上实验验证，可以使得机器人可以实现对目标物体的周围环境进行建图，尔后通过机器人所建立的地图还可以进行自主导航，通过自主导航可以使得机器人到达目标所在位置，达到目标所在位置后就可以通过机械臂对目标物体进行抓取。进行抓取时首先需要在 SolidWorks 中对机械臂进行坐标的构建，SolidWorks 中的机械臂模型如图 4-16 所示。



4-16 Solidworks 中的机械臂模型

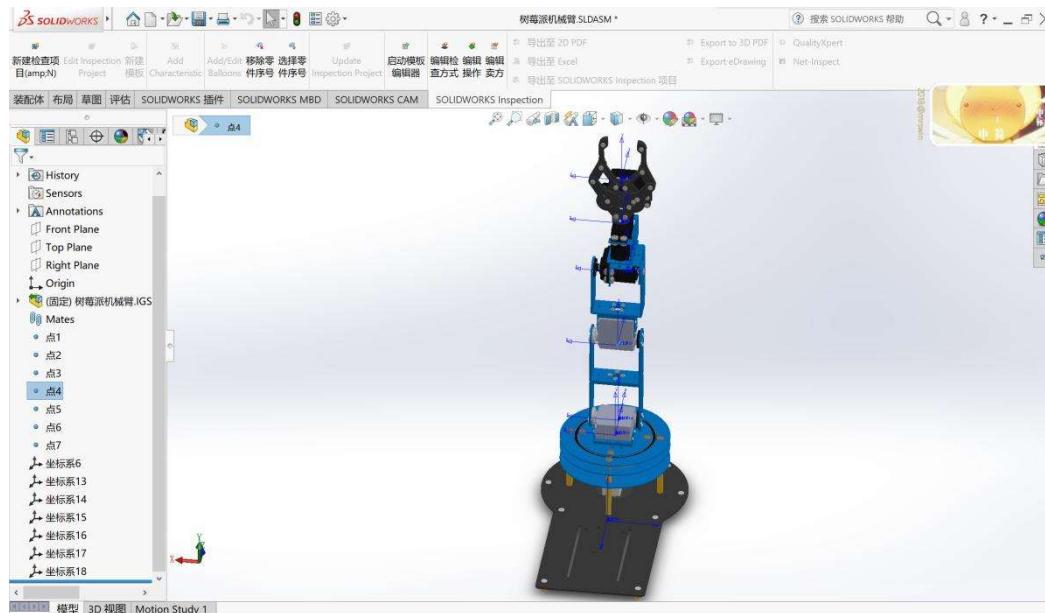
SolidWorks 中进行坐标的标定如图 4-17、4-18、4-19 所示。



4-17 Solidworks 中的机械臂模型



4-18 Solidworks 中的机械臂模型



4-19 Solidworks 中的机械臂模型坐标系

至此关于机械臂抓取的所有准备工作已经完成，对机械臂抓取目标物体进行测试，以此来验证抓取方法的正确性，抓取结果如图 4-20、4-21、4-22 所示：

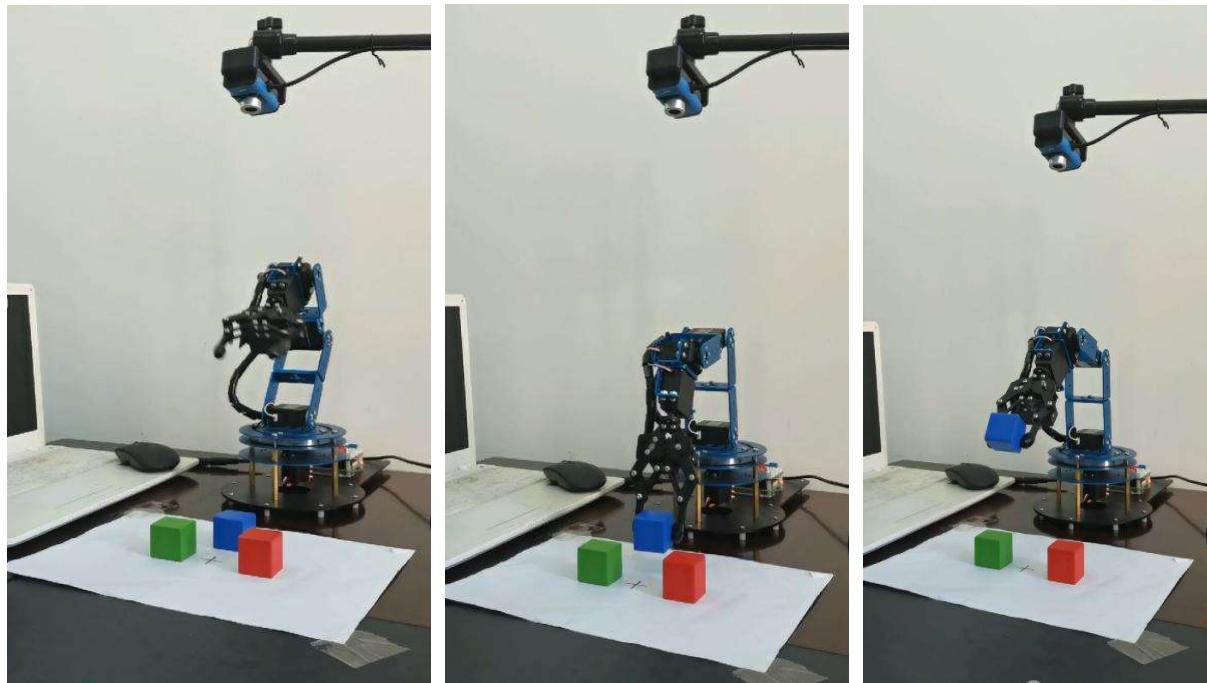


图 4-20 抓取过程 1

4-21 抓取过程 2

4-22 抓取过程 3

4.5 本章小结

本章研究了机器人实时建图与自主导航的基础，讨论了基于 Rao-Blackwellized 的粒子滤波的 SLAM 方法和导航的具体过程，在机器人设备上进行了实验验证，验证了在具体设备上建图与导航算法的可实施性。研究了机械臂抓取物体的基础，研究了手眼标定与机器人运动学建模的相关方法，并在具体设备上进行了设计方法的可行性验证。

第五章 结论与展望

5.1 本文结论

针对本文的研究工作进行以下的总结：

(1) 对深度学习下的目标物体识别进行研究。分析了深度学习下的物体识别的原理，并对深度学习、卷积神经网络的相关内容进行了阐述，最终根据需求选择使用 SSD 网络进行目标物体的识别，对 SSD 网络详细的阐述，描述了其基础网络 VGG-16 及 SSD 本身网络结构，同时对相关数据集进行研究，在 ubuntu 系统上进行了实验，实验结果显示算法达到预期效果。

(2) 分析了移动机器人建图与导航的相关原理，根据利用一系列随机样本的加权和近似后验概率密度函数，通过求和来近似积分操作进行建图；应用概率理论在已知地图中对机器人自定位置进行估计的方法，即蒙特卡洛自适应定位算法，实现了机器人的定位。在 Ubuntu 系统下进行了相关实验验证。设计目标物体抓取系统。阐述了手眼标定的相关内容，并对手眼标定所涉及到的坐标转换进行研究，并在实际机械臂设备上进行了实验，实现了机器人对目标物体的识别抓取。

5.2 未来展望

有关于机器人的研究是漫长且具有挑战的过程，所以对研究做以下的展望：

(1) 在建图与导航过程中，如若使用单一的传感器进行感知，机器人感知到的环境就有一定的局限性，在未来可以考虑到将雷达与其他传感器进行有效的结合，将会得到更加稳定、可靠、准确的结果。

(2) 机器人产品在移动、抓取、物体识别等方面封装过于死板，无法在多平台实现共通，且在解决多任务问题时候存在框架不同，环境不同，无法在同一设备上解决的问题，应将视觉与机械臂、底盘等设备统一到同一环境中，并在此基础上实现多任务多拆分，真正意义上实现机器人的“开源”。

参考文献

- [1] 马颂德, 张正友. 计算机视觉: 计算理论与算法基础[M]. 科学出版社, 1998.
- [2] 蔡自兴. 机器人学的发展趋势和发展战略[J]. 高技术通讯, 2001, 000(004): 11-16.
- [3] 蔡鹤皋. 机器人技术的发展与在制造业中的应用[J]. 机械制造与自动化, 2004, 33(1): 6-7.
- [4] 梁峰. 三维场景实时重建技术在遥操作机器人上的实现[D]. 西南科技大学.
- [5] 潘峰, 武威, 杨铁璐, 等. 视觉定位脑外科手术机器人系统的坐标映射[J]. 东北大学学报(自然科学版), 2005(05): 413-416.
- [6] Dalibard, S, et al. Dynamic Walking and Whole-Body Motion Planning for Humanoid Robots: an Integrated Approach[J]. International Journal of Robotics Research, 2013, 32(910): 1089-1103.
- [7] Vahrenkamp N, Asfour T and Dillmann R. Efficient inverse kinematics computation based on reachability analysis[J]. International Journal of Humanoid Robotics, 2012, 9(4): 1-25.
- [8] Weng J, Cohen P, Herniou M. Camera calibration with distortion models and accuracy evaluation[J]. IEEE Transaction on pattern analysis and machine intelligence, 1992, 14(10): 965-980.
- [9] Muller J, Frese U, Rofer T. Grab a mug-Object detection and grasp motion planning with the Nao robot[J]. European Journal of Endocrinology, 2012, 132(6): 349-356.
- [10] Borenstein J, Everett H R, Feng L, al. Mobile robot positioning-sensors and techniques [J]. Journal of Robotics Systems, 1997, 14(4): 231-249.
- [11] Qu Liping, WANG Hongjian. An overview of robot SLAM problem[C]. International Conference on Consumer Electronics, Communications and Networks (CECNet). Xianning, China, 2011: 1953-1956.
- [12] Zitnick, C. Lawrence, and Piotr Dollár. "Edge boxes: Locating object proposals from edges." European conference on computer vision. Springer, Cham, 2014: 391-405.
- [13] V. N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 1999, 10, 988-999
- [14] Freund, Yoav, and Robert E. Schapire. "A desicion-theoretic generalization of on-line learning and an application to boosting." European conference on computational learning theory. Springer, Berlin, Heidelberg, 1995: 23-37.
- [15] Redmon, J. , Divvala, S. , Girshick, R. , & Farhadi, A. . (2016). You Only Look Once: Unified, Real-Time Object Detection. Computer Vision & Pattern Recognition. IEEE.
- [16] Liu W , Anguelov D, Erhan D , et al. SSD: Single Shot MultiBox Detector[J]. 2016.
- [17] Lipton Z C , Berkowitz J , Elkan C . A Critical Review of Recurrent Neural Networks for Sequence Learning[J]. Computer Science, 2015.
- [18] Girshick, Ross. Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 1440-1448.
- [19] Ren, S. , He, K. , Girshick, R. , & Sun, J. . (2017). Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(6), 1137-1149.
- [20] Tsinnere, Granade. Whiteheadca. Autonomous Rendezvous and Docking Sensor Suite[J]. SPIE. 2003: 5086.

东北师范大学硕士学位论文

- [21] 赵月. 单目位姿测量目标中心定位算法研究[D]. 哈尔滨: 哈尔滨工程大学信息与通信工程学院, 2011.
- [22] Levine, Sergey, et al. "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection." *The International Journal of Robotics Research* 37.4-5 (2018): 421-436.
- [23] Kalashnikov, Dmitry, et al. "Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation." arXiv preprint arXiv:1806.10293 (2018).
- [24] Liang, Hongzhuo, et al. "Pointnetgpd: Detecting grasp configurations from point sets." *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019:3629-3635.
- [25] Mousavian, Arsalan, Clemens Eppner, and Dieter Fox. "6-dof graspnet: Variational grasp generation for object manipulation." *Proceedings of the IEEE International Conference on Computer Vision*. 2019:2901-2910.
- [26] Liu, M., Pan, Z., Xu, K., Ganguly, K., & Manocha, D. (2019). Generating grasp poses for a high-dof gripper using neural networks. arXiv preprint arXiv:1903.00425.
- [27] Mahler, Jeffrey, et al. "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics." arXiv preprint arXiv:1703.09312 (2017).
- [28] Varley, Jacob, et al. "Shape completion enabled robotic grasping." *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017:2442-2447.
- [29] 李超波, 李洪均, 徐晨. 深度学习在图像识别中的应用 [J]. 南通大学学报(自然科学版), 2018 (1) .
- [30] 祝彬, 郑娟. 惯性导航与制导技术的新发展[J]. 中国航天, 2008(1): 43-45.
- [31] 匡青. 基于卷积神经网络的商品图像分类研究[J], 天津, 软件导刊.
- [32] LECUN, Yann, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86.11: 2278-2324.
- [33] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, 60.6: 84-90.
- [34] HE, Kaiming, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 770-778.
- [35] SZEGEDY, Christian, et al. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. p. 1-9.
- [36] SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- [37] 辛江慧, 李舜酩, 廖庆斌. 基于传感器信息的智能移动机器人导航评述[J]. 传感器与微系统, 2008(04):4-7.
- [38] 蔡自兴, 贺汉根, 陈虹. 未知环境中移动机器人导航控制研究的若干问题[J]. 控制与决策, 2002(04):385-390+464.
- [39] Borenstein J, Everett H R, Feng L, al. Mobile robot positioning-sensors and techniques [J]. *Journal of Robotics Systems*, 1997, 14(4):231-249.
- [40] 李磊, 叶涛, 谭民, 陈细军. 移动机器人技术研究现状与未来 [J]. 机器人, 2002 (05) :475-480.
- [41] QU Liping, WANG Hongjian. An overview of robot SLAM problem[C]. *International Conference on Consumer Electronics, Communications and Networks (CECNet)*. Xianning, China, 2011: 1953-1956.
- [42] M. Montemerlo, S. Thrun, D. Koller etc. FastSLAM: A factored solution to the simultaneous localization and mapping problem[C]. *Proceedings of the National Conference on Artificial Intelligence*. Edmonton: American Association for Artificial Intelligence Menlo Park, 2002, 593-598.
- [43] D Hahnel, W Brugard, D Fox etc. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements[C]. *Proceedings of IEEE/RSJ*

- International Conference on Intelligent Robots and Systems.2003 (1).206-211.
- [44] G Grisetti, C Stachniss, W Burgard. Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Selective Resampling[C]. Proceedings of IEEE International Conference on Robotics and Automation (ICRA).2005, 2423-2437.
- [45] Lenard J, Durrant-Whyte H. Dynamic map building for an autonomous mobile robot [J]. The International Journal on Robotics Research, 1992, 11(4):286-298.
- [46] Mei, C.; Sibley, G.; Cummins, M.; Newman, P.; Reid, I. RSLAM: A System for Large-Scale Mapping in Constant-Time Using Stereo. Int. J. Comput. Vis. 2011, 94, 198–214.
- [47] 潘泉, 杨峰, 叶亮, 梁彦, 程咏梅. 一类非线性滤波器——UKF 综述[J]. 控制与决策, 2005(05):481-489+494.
- [48] Kalman, R. E. A. New Approach to linear Filtering And Prediction Problems [J]. Journal of Basic Engineering. 1962, 82: 35-45.
- [49] Boris M. Miller, Wolfgang, J. Runggaldier. Kalman filtering for linear systems with coefficients driven by a hidden Markov jump process [J]. Systems & Control Letters, 31(2): 93-102.
- [50] Sameni R et al. A Nonlinear Bayesian Filtering Framework for ECG Denoising [J]. IEEE Transactions on Biomedical Engineering, 2007, 54(12): 2172-85.
- [51] MCKERROW, Phillip John; MCKERROW, Phillip. Introduction to robotics. Sydney: Addison-Wesley, 1991.

在学期间公开发表论文及著作情况

文章名称	发表刊物(出版社)	刊发时间	刊物级别	第几作者
Multi-Modality Video Representation for Action Recognition	Journal on Big Data	2020年10 月13		2

致谢

为期三年研究生的学习生涯转瞬即逝，感谢所有老师在我学习期间给予的建议和指导。幸运的是，在学习和研究过程中，我遇到了许多志同道合的知己，这些都是我周围最好的榜样和催化剂，我从中受益匪浅。

在东北师范大学读研的三年中，对所学机器人与深度学习理论有了一定的理解和掌握，其中离不开 XXX 老师的指导和解惑；同时，我也要感谢所有同门师兄师姐在学习过程中提供的有关内容与帮助，在与师兄师姐的交流和讨论中，加深了对各种理论知识的理解。

感谢多年来学习中遇到的所有同窗，由于各位同窗好友的存在，我可以看到自己的许多缺点并不断进行调整。感谢我的母亲多年来为我提供的照顾，只有不懈的努力才能回馈。

