

大学生创新训练项目申请书

项目编号_____

项目名称_____基于深度级联卷积神经网络的
_____机器人控制目标位姿学习系统_____

项目负责人_____曾银雪_____联系电话_____13588218705_____

所在学院_____自动化与电气工程学院_____

学 号_____1170399065_____专业班级_____自动化 174_____

指导教师_____周乐_____

E-mail_____zengyinxue_email@163.com_____

申请日期_____2020 年 6 月 10 日_____

起止年月_____2020 年 6 月至 2021 年 6 月_____

浙江科技学院

填 写 说 明

1、本申请书所列各项内容均须实事求是，认真填写，表达明确严谨，简明扼要

2、申请人可以是个人，也可为创新团队，首页只填负责人。
“项目编号”一栏不填。

3、本申请书为大 16 开本(A4)，左侧装订成册。可网上下载、自行复印或加页，但格式、内容、大小均须与原件一致。

4、负责人所在学院认真审核，经初评和答辩，签署意见后，将申请书（一式两份）报送浙江科技学院创新创业学院。

一、基本情况

项目名称	基于深度级联卷积神经网络机器人控制目标位姿学习系统						
所属学科	学科一级门： 控制科学与工程 学科二级类： 控制理论与控制工程						
申请金额	10000 元		起止年月		2020 年 6 月至 2021 年 6 月		
负责人姓名	曾银雪	性别	女	民族	汉族	出生年月	1999 年 11 月
学号	1170399065	联系电话	宅： 手机:13588218705				
指导教师	周乐	联系电话	宅： 手机:13656648682				
负责人曾经参与科研的情况		参与浙江省自然科学基金面上项目“面向多采样率数据集的造纸废水处理过程故障检测研究”，承担基于多采样率数据集的线性回归建模及仿真验证工作，录用学术论文《基于多采样率因子分析回归的工业过程软测量建模》一篇（第一作者）。					
指导教师承担科研课题情况		1、主持，国家自然科学基金青年项目，随机噪声环境下的动态过程概率建模及故障检测方法 2、主持，中国博士后基金（一等资助），基于概率建模的复杂动态过程质量预测与监控研究 3、主持，浙江省自然科学基金面上项目，面向多采样率数据集的造纸废水处理过程故障检测研究 4、主持，浙江省教育厅一般项目，基于加权学习的工业过程概率建模与过程监测研究					
指导教师对本项目的支持情况		指导教师周乐一直从事数据驱动故障诊断与软测量、工业大数据分析、人工智能与深度学习理论及其应用等领域的研究工作，主持国家级、省部级及产学研项目多项，熟悉本领域国内外的研究前沿，在理论研究和工业过程应用方面积累了丰富的研究经验和科研成果。长期指导学生参加科研训练，可以完成对本项目的指导工作。					
项目组主要成员	姓 名	学号	专业班级		所在学院		项目中的分工
	曾银雪	1170399065	自动化 174		自动化与电气工程学院		代码编写
	夏王法	1170399091	自动化 174		自动化与电气工程学院		硬件调试
	余李新千	1190309024	自动化 191		自动化与电气工程学院		采集并处理数据

二、 立项依据（可加页）

（一） 项目简介

随着新一代信息技术与先进制造业的深度融合，以智能制造为代表的新一轮产业发生了巨大变革。而智能制造是新一轮产业变革的核心方向，目前国内很多传统制造业在转型升级中，已将智能制造列为重点转型的领域，持续为传统制造业创新发展创造新机遇。工业机器人作为衡量一个企业自动化水平高低的重要标志，机器视觉系统的优越性则是提高生产的柔性和自动化程度的保证。

目前工业生产过程中采用的颜色特征和纹理特征等基本机器视觉方法用于特征提取和目标检测，仅用于一些几何形状规则，颜色特征显著的目标更为合适，该局限性限制了工业机器人的普及应用，因此需要研究基于机器视觉的精确目标识别及分类方法，以实现对于工业机器人的精准的控制。并且随着生产自动化的要求越来越高，对更复杂物体的检测成为该领域迫切攻克的难题。

因此，本项目旨在开发出高效精确的目标识别及分类方法，提升工业机器人的“智能化”水平，提高工业机器人生产任务完成的可靠性及鲁棒性，提高生产的柔性和自动化程度，进而推动工业机器人在智能制造过程中的普及使用。项目理念迎合了工业 4.0 时代对传统制造业提出的新要求、新目标。随着技术的革新，将会为企业带来生产效率的提高和人力费用的节省。这是传统制造业必须要经历的生产革新，为我国制造业的发展搭桥铺路。



图 1.1 规则物体的抓取



图 1.2 复杂物体的抓取

（二） 研究目的

本项目旨在开发出高效精确的目标识别及分类方法，对工业机器人的控制目标特征提取进行研究。为去除无关背景信息对关键点检测的影响，根据由粗到细思想，提出基于残差卷积神经网络的光面瓶体位姿估计方法，级联网络通过区域定位、粗定位、精定位的过程逐级提升关键点定位精度，从而有效提高瓶体位姿估计精度。传统方法与新兴技术相融合，人工智能技术和互联网技术与机器人领域紧密融合，力求协同发展出硕果，开发出高效精确的目标识别及分类方法，提升工业机器人的“智能化”水平，提高工业机器人生产任务完成的可靠性及鲁棒性，提高生产的柔性和自动化程度，进而推动工业机器人在智能制造过程中的普及使用。

（三） 研究内容

随着信息时代的到来和世界经济竞争的不断加剧，工业机器人是衡量一个国家自动化水平高低的重要标志。机器人技术的广泛应用必将推动科学技术和社会经济产生另一次飞跃。在自动化生产的理念推动下，机器人已经在一些特定工业生产上展露人类无法比拟的能力，它将在未来的制造企业中扮演越来越重要的角色。一直以来，针对机器人抓取的位姿估计问题有许多相关研究，并取得了较好的成果。其中深度学习方法避免了人为选择设计特征的过程，检测精度不依赖匹配模板库，对比于基于模型的位姿估计方法更具优势。

在本项目中初步应用的对象是一种常见的容器---瓶体，其大量出现在工业生产和日常生活中，常用做液体产品的包装。而在各种相关的生产线中，对瓶体位姿的精确估计是保证瓶体抓取，液体灌装，瓶口封装等生产操作准确执行的前提。像一些高分子聚合物、玻璃因具有极佳的可塑性普遍用作瓶体的主要生产材料，其瓶体产物也具备复杂多样的外观形状，使得瓶体位姿估计具有较大的难度。光面瓶体没有明显的纹理，且具有反光和透光特性，瓶体表面特征容易受光照因素影响，进一步加大了位姿估计的难度。因此，研究面向机器人操作的光面瓶体位

姿估计方法，对实现工业机器人高效而精准的控制具有重要的意义，具有较广的应用市场。

本项目的研究内容主要针对工业机器人的控制目标识别难、现有目标特征提取识别方法精度低，鲁棒性差，无法满足工业应用要求等问题，研究基于机器视觉的工业机器人复杂工况下目标识别问题，利用先进的数据处理技术、深度学习技术，对机器人抓取目标的特征提取进行研究，双目视觉特征点定位方法的特点，从而提出基于残差卷积神经网络的机器人抓取目标的位姿估计方法。

（四） 国、内外研究现状和发展动态

（1） 机器视觉应用现状

机器人是一个正在高速崛起的产业，它逐渐成为促进社会化大生产发展的重要力量。随着信息时代的到来和世界经济竞争的不断加剧，工业机器人在未来的生产中必将担任更加重要的角色。如今，机器人已成为自动化生产线、柔性制造系统中必不可少的单元。相关统计显示，2001 年至 2013 年，全球机器人增速为 7.2%，2013 年产销量达到历史新高为 17.9 万台，2014 年突破 20 万台。其中中国机器人市场规模最大，增速也最快。2013 年，中国共销售 3.7 万台工业机器人，同比增长 60%。2008-2013 年，中国机器人销量年均增长率达 36%。目前，全球每销售 5 个机器人，其中就有 1 台是在中国安装。

习近平总书记在致 2015 世界机器人大会贺信中指出：“中国将机器人和智能制造纳入了国家科技创新的优先重点领域，我们愿加强同各国科技界、产业界的合作，推动机器人科技研发和产业化进程，使机器人科技及其产品更好为推动发展、造福人民服务。”因此，关于机器人技术的研究和应用推广将具有重要意义。在机器人相关技术中，机器视觉占据了重要地位。目前现有的工业应用尤其是在目标特征分析，定位等方面还只是采用较为简单或传统方法，精度和稳定性难以满足生产柔性化需求。

针对工业机器人控制目标的识别难题，目前国内外研究有如下几个方面：通过结构光技术与双目立体视觉技术获取目标的三维姿态信息从而进行目标识别和定位。这一方法能够可以得到较高的识别精度，但存在一定的局限性，多目立体视觉技术所需算法计算量大，识别定位的效率不高，并且由于相机与目标相对位

置固化,该种方法的鲁棒性较差; Doumanoglo 等提出了一种物体姿态的预测方法。取物体的多小块 RGB-D 数据,通过自编码器对数据进行降维,将降维后的特征用于训练 Hough Forest。在实际物体检测时通过在物体表面采样 RGB-D 数据,估计出位姿。其局限性在于,作为一个回归问题,回归精度还无法直接用于物体操作,因此还需要采用 ICP 等算法进行最后一步的匹配迭代。最后,实际机器人进行一次操作需要耗费时间、可能会造成损害、会破坏实验条件(需要人工恢复)等,在实际应用中还存在着较大的难度。

(2) 深度学习发展现状

深度学习的概念最早由多伦多大学的 G. E.Hinton 等于 2006 年提出。其动机在于建立、模拟人脑进行分析学习的神经网络,它模拟人脑的机制来解释数据,例如图像,声音和文本。它是当前图像处理与计算机视觉领域的研究热点。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,以发现数据的分布式特征表示。目前,深度学习算法已经在语音识别、视觉对象识别、对象检测等多个领域取得了显著运用,其常见的基本模型包括自动编码器、深度信任网络、卷积神经网络和递归神经网络等。

其中,卷积神经网络(CNN)由于其在处理图像识别与分类上的优越性能,已经被广泛应用于机器视觉领域。2012 年 ImageNet 比赛中 AlexKrizhevsky 等提出 AlexNet 首次将深度学习应用到大规模图像分类,取得了图像分类和目标定位任务的第一。在 ILSVRC-2014 比赛中,几乎所有的参赛队伍都采用了卷积神经网络及其变形方法,其中 GoogLeNet 小组采用卷积神经网络结合 Hebbian 理论提出的多尺度的模型,以 6.7%的分类错误,取得图形分类“指定数据”组的第一名。

(3) 深度学习在工业领域内、工业机器视觉中的应用

目前,深度卷积神经网络(CNN)已经在图像识别、图像分类、目标识别与定位等领域取得了丰富的理论研究和实际应用成果,但在工业机器人领域的运用还较少。2017 届亚马逊机器人大赛(原为 Amazon Picking Challenge,这一届更名为 Amazon Robotics Challenge,简称 ARC)展示了深度学习在工业领域应用的雏形。APC 比赛中物体类别多样,包含矿泉水瓶(透明)、网格笔筒(表面不平整)毛绒玩具(变形)等难以识别的物体。传统物体识别算法无法完全识别物体,因此几乎全部参赛队伍都采用了深度学习技术。虽然许多参赛队伍贡献了出色表现,但 2017ARC 展现的技术较实用还存在着难题亟待解决:几乎所有队伍都以吸盘作为

主要抓取方式，机械手作为辅助甚至完全不装备机械手。这种方式限制了可抓取物品的范围以及抓取的成功率，尤其是在工业零件上的适用性不足。

已经投入实际应用的工业机器人多是抓取特定种类的物体。以实际应用为驱动，生产设计特定功能的工业机器人。日本 Fanuc 公司造出来的 bin picking 工业机器人是将人工智能技术应用于工程作业的典型例子，2016 年 12 月份，Fanuc 在东京国际机器人展览会上展示了这台经强化学习训练的机器人。Fanuc 的机器人使用了一种深度强化学习的技术来训练它自己，可随时学习新的任务。它在尝试拾起物品的同时，能够抓取这个过程的录像。不管每次它是成功了还是失败了，它都会记住物品长什么样的，用它学到的知识改进控制它行动的深度学习模型或大型神经网络。大约经过 8 个小时的学习后，机器人的分捡成功率可以达到 90%，相当于一位专家为它编写程序所能达到的准确度。综上所述，本项目结合机器视觉技术、CNN 和级联卷积网络设计的机器人控制目标位姿学习系统将在一定程度上弥补相应的工业应用空白，具有不错的技术优势。

（五）创新点与项目特色

机器视觉系统的优越性是提高生产的柔性和自动化程度的保证。在一些不适合于人工作业的危险工作环境或人工视觉难以满足要求的场合，常用机器视觉来替代人工视觉；同时在大批量工业生产过程中，用人工视觉检查产品质量效率低且精度不高，用机器视觉检测方法可以大大提高生产效率和生产的自动化程度。

目前现有的工业机器人应用普遍应对分拣、抓取等任务，而这些任务成功完成的前提是对目标正确的识别、分类及定位，传统方法是基于边缘特征或基于纹理特征通过特征匹配以实现目标识别与位姿估计。但随着生产自动化要求提高，更复杂物体的姿态估计逐渐成为机器视觉领域的研究热点。复杂物体由于光照、旋转、纹理不显著等因素的多样性，边缘特征和纹理特征等基本视觉特征存在应用的局限性，因此需要研究基于机器视觉的精确目标识别分类、位姿估计方法，以实现工业机器人高效而精准的控制。

在本项目中突出的创新点在于不同于传统方法的检测方法。传统基于神经网络的物体检测算法主要以滑动窗口结合分类卷积神经网络完成待检测物体分割与识别。该类方法首先使用滑动窗口获取输入图像子图区域，将子图输入分类卷积神经网络进行分类，最终根据将分类结果超过设定阈值的子图所对应的区域标记

为物体区域。为保证不同尺度下的检测正确率，通常预处理阶段需要构建图像的高斯空间金字塔或者采用不同尺度与宽高比的滑动窗口进行检测，该过程大幅度降低了算法的实时性。基于区域建议的物体检测网络改进了滑动窗口重复计算导致算法复杂度过大的缺陷，提取使用神经网络检测物体潜在区域，并完成端对端分类识别。减少了无关背景对关键点定位的干扰。随后，在获得区域结果基础上，基于粗定位网络定位目标的每个关键点。仅对若干个单目标位姿进行验证，当存在相机位置偏移扰动或应对试验数据集之外的目标位姿检测时，检测精度偏低，这表明该回归模型扩展性不足，系统鲁棒性差。

此外，基于传统工业机器人制造技术分析，本项目设计的基于级联卷积神经网络的机器人抓取目标位姿学习系统的核心竞争优势体系在以下几个方面：

自学习性，强鲁棒性。通过应用深度学习方法，自学习、自适应的提取对应特征用于位姿估计，避免了人为设计特征的繁琐过程；特别是涉及到光照对光面瓶体的检测的影响，瓶体表面反光会造成部分特征信息的缺失时，与其他传统方法相比，其优势更加明显。

本质特征，强适用性。通过特征点学习方法，根据学习到的特征点还原工业机器人抓取目标的位姿，所得到的是抓取目标的本质位姿特征，适用于不同作业类型的工业机器人，如吸盘、机械夹具等。而一些基于抓取框学习的目标检测方法仅适用于特定的机械抓取机器人。

深度学习，强迁移性。深度学习的机器人抓取目标位姿学习方法在应对不同的抓取对象时，不需要重新设计系统，可采用小样本数据，对级联卷积神经网络模型进行再训练，微调参数，就能应用于新的目标对象的位姿估计。

（六）技术路线、拟解决的问题及预期成果

在本项目先确定研究的主要内容和对象，再对相关的理论算法进行进一步的学习，分析国内外机器视觉应用和深度学习发展以及深度学习在工业领域内、工业机器视觉中的应用现状，进而设计项目方案。根据项目方案搭建相关硬件和软件平台，建立基本模型，编写程序以实现目标功能。程序主体框架搭建好之后，通过不断调试，进而提高测量的精度，完善模型的稳定性。整体完成之后，进行总结分析，收集在项目过程中所遇到的问题，积累经验并吸取在调试过程当中的失败教训。最终整理材料，以完成整个项目。

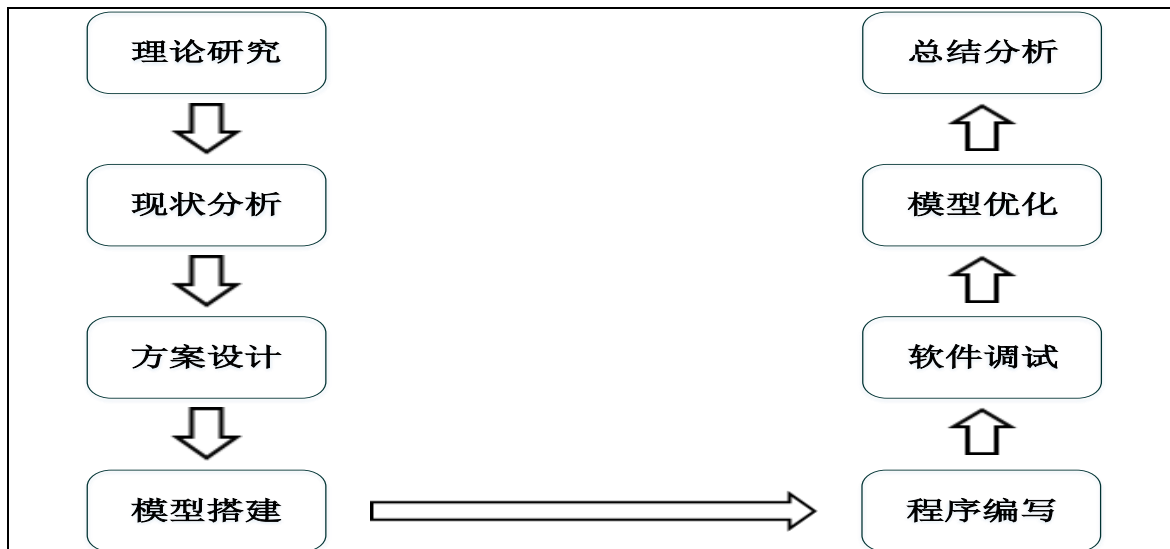


图 6.1 项目研发技术路线框图

在本项目当中对于机器人控制目标位姿学习系统主要采用三级级联卷积神经网络根据由粗到细的检测思想对目标体进行关键点定位。

技术路线主要如下：

第一级：区域检测网络：用于初步确定检测目标范围以去除背景干扰，即利用常规物体检测神经网络定位图像中的目标物体区域；

第二级：粗定位网络：利用图像全局信息粗略定位目标物体的全部关键点位置，特征点之间存在着几何约束，粗定位网络通过五个子网络来训练学习定位特征点的同时隐式地学习特征点之间的约束关系；

第三级：精定位网络：对每一个特征点进行精确定位，即利用各个关键点周围小范围区域特征信息，微调该关键点坐标，对关键点进行精确定位。

最后进行重复精定位过程进一步提高精度。在定位方式当中主要采用了取平均值的思想，即用多个网络对同一个特征点进行重复预测，把重复预测的特征点进行结果平均，可以提高网络的稳定性、防止预测特征点的位置偏差过大，从而提高测量的精度。

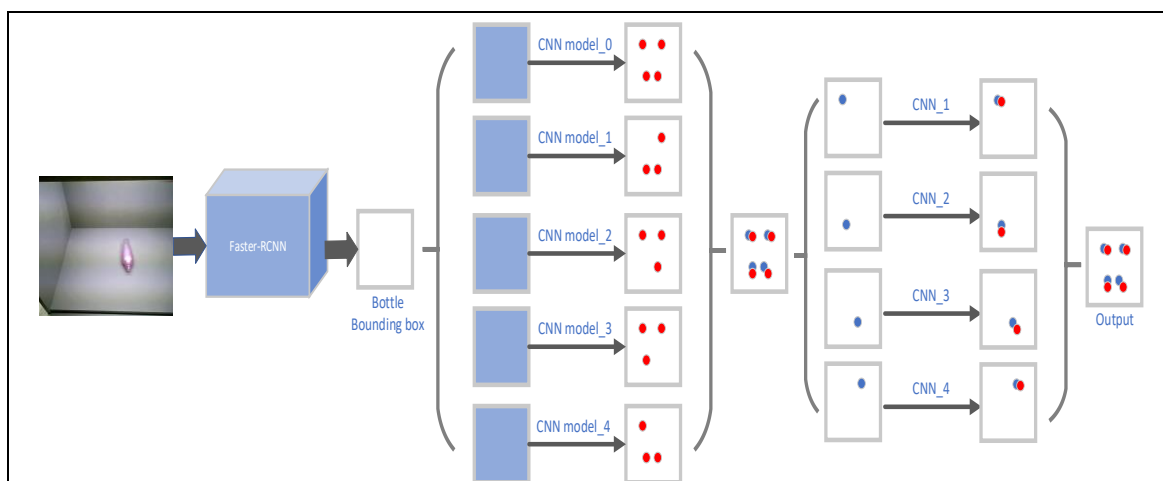


图 6.2 三级级联卷积神经网络

具体流程:

第一级.区域检测网络。Faster RCNN 检测算法分两个阶段：RPN 区域建议和 FastRCNN 物体检测。原始图像经过 VGG、ResNet 等训练模型提取高层特征，RPN 及 Fast RCNN 网络与前置特征抽取器互连共享特征图。Faster RCNN 模型主要步骤如下：

(1) 特征提取：以整张图片为输入，利用卷积层提取输出图片的特征图，用于后续的区域提名网络层和全连接层。

(2) 区域提名：在最终的卷积特征图上为每个点利用 k 个不同的矩形框(Anchor Box)进行提名， k 一般取 9。

(3) 区域判定和回归：区域提名网络 (Region Proposal Networks, RPN) 对生成的矩形框对区域进行 (目标对象/非目标对象) 二分类，并用 k 个回归模型 (各自对应的不同的矩形框)，微调候选框位置和大小。RPN 网络仅检测存在物体的位置，不做具体类别的识别。

(4) 分类与回归：对区域提名网络给出的区域结果进行筛选，进行目标分类和回归物体边界框。

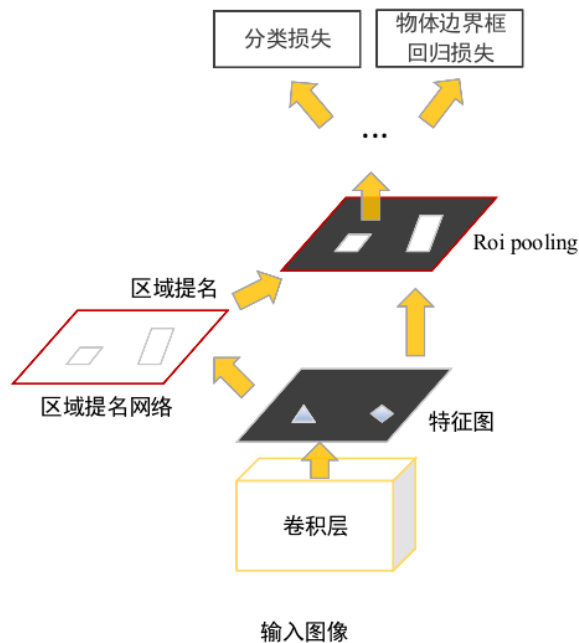


图 6.3 Faster RCNN 网络

第二级.粗定位网络。基于 Faster-RCNN 区域检测的方法获得目标体边界框，减少了无关背景对关键点定位的干扰。随后，在获得区域结果基础上，基于粗定位网络定位目标体的每个关键点。关键点之间存在着几何约束，粗定位网络通过五个子网络来训练学习定位关键点的同时隐式地学习关键点之间的约束关系。粗定位级联网络的输入为去除无关背景的目标体图片，因此，首先需要基于区域检测网络定位目标体位置，中心裁剪为 224×224 尺寸大小，然后输入粗定位网络，网络由五个子网络(CNN model)组成，多方面预测多个目标体关键点，其中第零号网络(CNN model_0)用于预测全部 4 个目标体关键点，其余子网络(CNN model_1~4)分别预测其中 3 个关键点。最后将五个子网络预测结果取平均用于下一级精定位网络。对于每个关键点，多个子网络的预测取平均可减少方差，提高检测的鲁棒性。

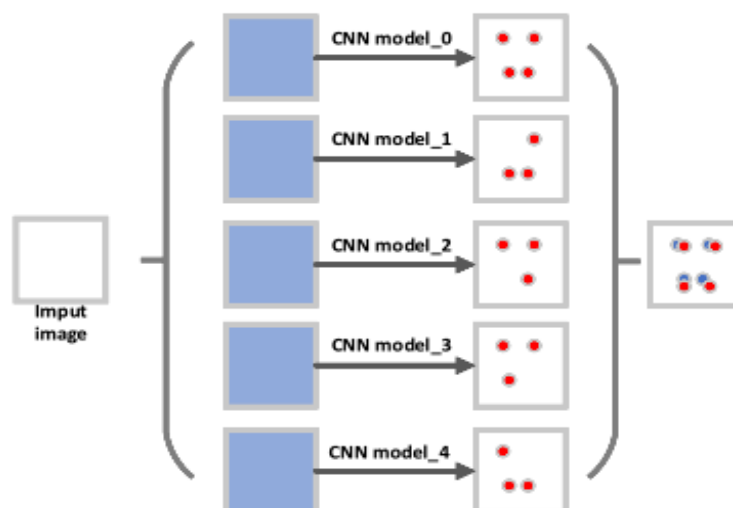


图 6.4 粗定位网络

粗定位子网络的结构，包含七个卷积层，两个池化层以及两个上采样层。将尺寸为 $224 \times 224 \times 3$ 的图像输入卷积神经网络，经过两个卷积加池化的组合提取图像特征，获得 $56 \times 56 \times 80$ 的特征矩阵，然后通过 2 个卷积层进一步提取图像特征信息，随后通过上采样加卷积的组合将特征图尺寸映射至 $224 \times 224 \times 10$ ，最后一个卷积层也是输出层，用于输出关键点预测热图，根据子网络预测关键点的数目不同，第零号网络(CNN model_0)输出尺寸为 $224 \times 224 \times 4$ 预测结果，包含 4 个关键点的坐标信息，第 1-4 号网络(CNN model_1~4)输出尺寸为 $224 \times 224 \times 3$ 预测结果，包含 3 个关键点的坐标信息。

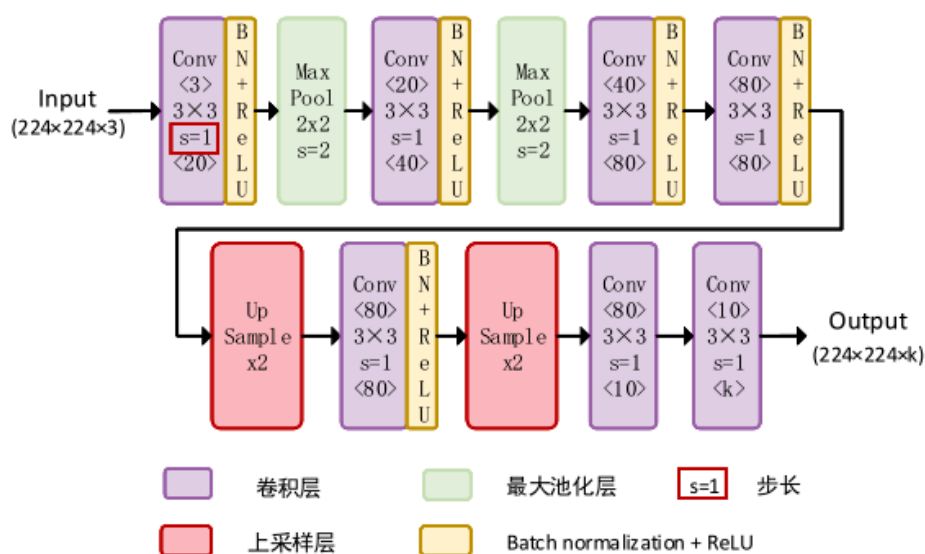


图 6.5 粗定位子网络

第三级.精定位网络。在粗定位网络预测结果的基础上，以预测关键点中心裁剪 40×40 尺寸图像，得到四个关键点对应的局部图像。精定位网络由四个子网络组成，每个子网络分别输入四个关键点所对应的裁剪图像，分别训练各自的参数，用于分别定位。尽管对于单一关键点的预测，子网络被设计为具有相同的结构，在相同映射上局部共享卷积权重有助于提高性能和效率，但是对于目标体的关键点定位并不适用。传统卷积网络假设相同的特征可能出现在图像的任何地方，将局部区域有效的卷积核应用于图像全局提取相同的特征是合理的。然而，全局共享权重在具有固定空间布局的图像上并不适用。因此，精定位网络的四个子网络互不共享权重，独立的训练预测各自关键点。

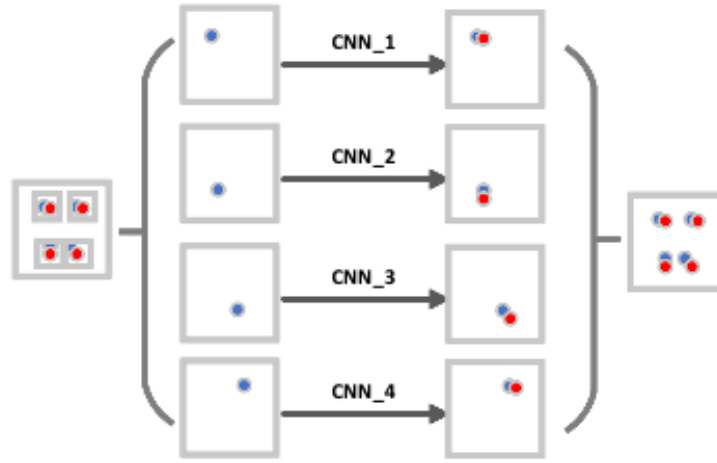


图 6.6 精定位网络

精定位子网络的结构组成大体结构与粗定位子网络类似，包含七个卷积层，两个池化层以及两个上采样层。网络输入为上一步粗定位网络关键点预测结果中心裁剪得到尺寸为 $40 \times 40 \times 3$ 的图像，经过两次卷积加池化的组合操作得到 $10 \times 10 \times 80$ 的特征图，两个卷积层进一步提取图像特征信息后通过两次上采样加卷积的组合将特征图尺寸映射至 $40 \times 40 \times 1$ ，最后一个卷积层用于输出 $40 \times 40 \times 1$ 的关键点预测热图，精细预测目标体的一个关键点。

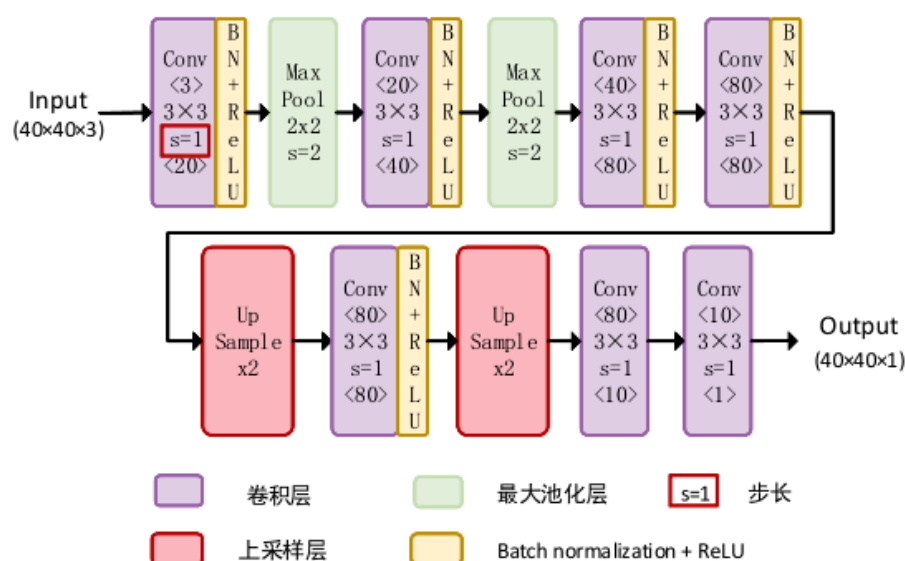


图 6.7 精定位子网络

在本项目当中拟解决的问题为需要进一步提高目标位姿估计定位的精确度以及降低识别目标的出错率。初期以瓶体的位姿估计为例近似代替机器人控制目标。而级联卷积神经网络能够提取传统机器视觉较难处理的光滑瓶体图像特征并能较好地估计瓶体的位姿信息。

(七) 项目研究进度安排

本项目研究主要分为以下四个阶段：

项目初期(2020 年 6 月至 2020 年 7 月)：资料收集整理，理论学习并搭建代码框架；

项目前期(2020 年 7 月至 2020 年 9 月)：完成目标体的选取和编写代码进行调试以及建立模型进行模型训练和完成基本位姿估计；

项目中期(2020 年 9 月至 2021 年 2 月)：依照项目中的试验方案进行试验测试并在此过程当中调试不断进行优化，以达到更好的效果，并实现相应的功能；

项目后期(2020 年 3 月至 2021 年 6 月)：整理项目资料，撰写报告，准备结题结辩。

(八) 已有基础

1. 与本项目有关的研究积累和已取得的成绩

(1) 已基本完成与本项目相关的理论学习和基本模型框架的搭建，初步建立位姿估计学习系统；

机器人控制目标位姿学习系统估计结果准确性很大程度取决于关键点定位的精度，通常关键点检测问题通过点坐标回归的方式输出预测点，回归损失仅考虑

欧式距离，即预测点和真实点之间的欧式距离。但是在训练过程中两点欧式距离所能提供的监督信息较少，导致网络收敛速度慢，误差也较大。故将样本标签变换为热度图标签，即根据图像中每个像素位置与关键点的距离生成 0~1 之间的概率，从而生成概率分布图。每个像素位置都提供了监督信息，为输出预测提供更加可靠的指导。此外为了减少无关背景对检测的影响，提高检测精度，结合人脸关键点检测带来的思路启发，基于级联卷积神经网络的机器人控制目标位姿学习系统估计新方法，可以获取更高的检测精度。

本项目为基于级联卷积神经网络用于机器人控制目标位姿学习系统，建立的位姿估计学习系统的模型框架如图 8.1 所示。

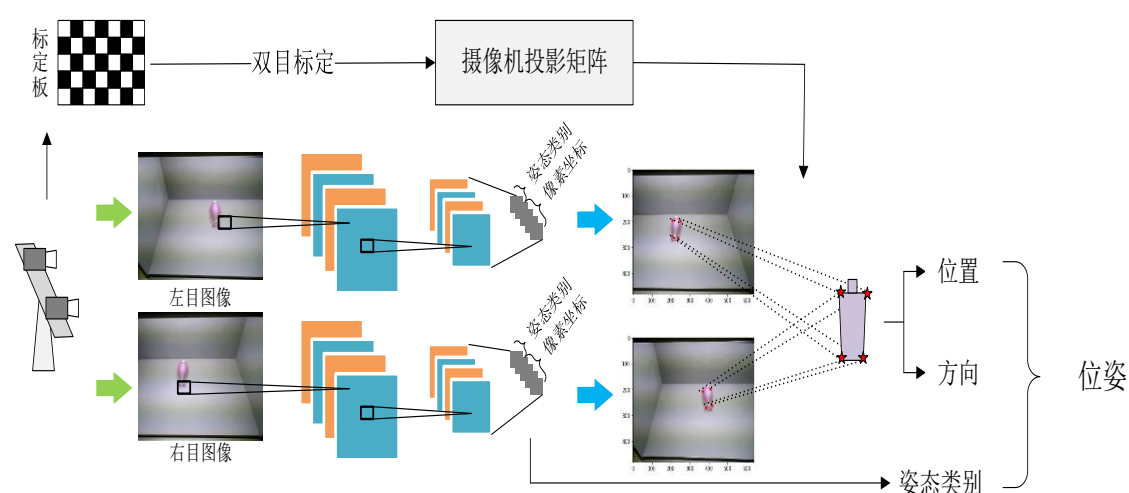


图 8.1 位姿估计学习系统的原理流程图

利用双目视觉平台采集目标瓶体左目和右目图像数据，基于残差卷积神经网络学习输出瓶体姿态类别和特征点像素坐标，同时通过摄像机标定获取摄像机投影矩阵；然后，通过双目立体视觉技术三维反求特征点三维坐标，还原特征点的空间位置；最后，根据获得特征点的三维坐标构造相对距离和向量表征瓶体的所在空间位置和方向，对位姿估计结果的进行度量。

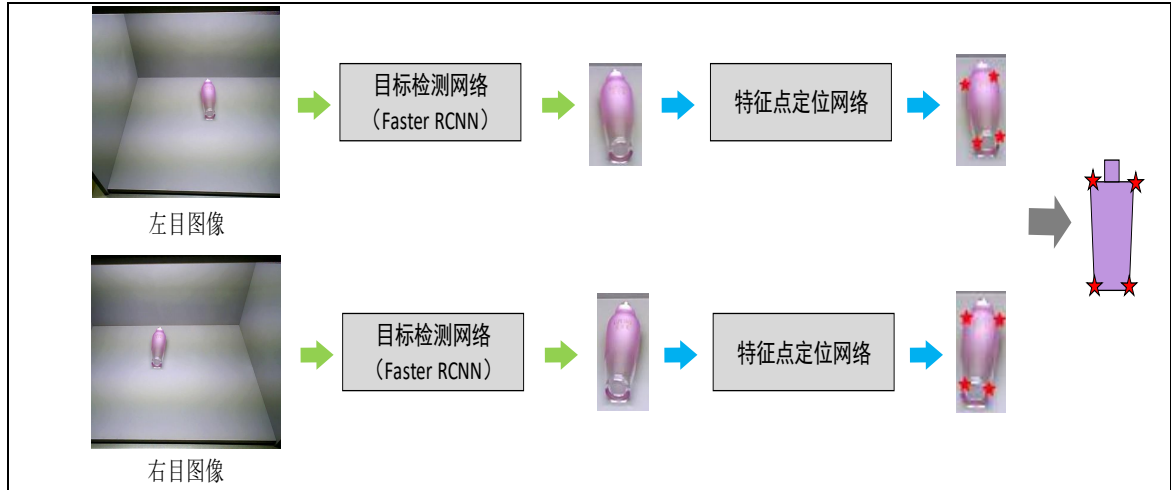


图 8.2 检测框架

对于相机标定，双目摄像机所拍摄到的左右图像中，同一特征点的左右两个二维像素坐标唯一对应着世界坐标下一个空间点。这样的对应关系是由摄像机成像的几何模型所决定的，这样的几何模型就是摄像机的参数，参数主要分为内部参数和外部参数，获得这些参数的过程就是摄像机标定。

常用坐标系。摄像机标定涉及 4 种坐标系，分别为：

- (a) 世界坐标系 $O_w - X_w Y_w Z_w$ ：空间场景的绝对坐标系；
- (b) 摄像机坐标系 $O_c - X_c Y_c Z_c$ ：以摄像机光心为原点，以垂直于成像平面的摄像机光轴为 Z 轴建立三维直角坐标系左摄像机坐标系 $O_{c1} - X_{c1} Y_{c1} Z_{c1}$ ，右摄像机坐标系 $O_{c2} - X_{c2} Y_{c2} Z_{c2}$ ；
- (c) 图像物理坐标系 $O_p - XY$ ：原点为透镜光轴与成像平面的交点， X 与 Y 轴分别平行于摄像机坐标系的 x 与 y 轴，是平面直角坐标系，长度单位为毫米；左摄像机图像物理坐标系 $O_{p1} - X_1 Y_1$ ，右摄像机图像物理坐标系 $O_{p2} - X_2 Y_2$ ；
- (d) 图像像素坐标系 $O_{pix} - uv$ ：为固定在图像上的以像素为单位的平面直角坐标系，其原点位于图像左上角,坐标轴平行于图像物理坐标系的 X 和 Y 轴。

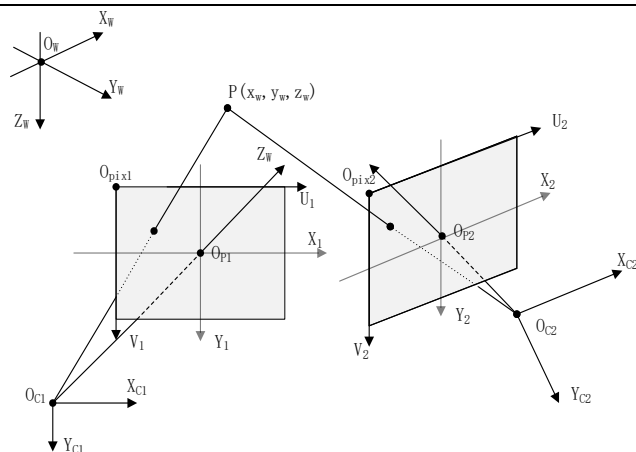


图 8.3 双目定位原理图

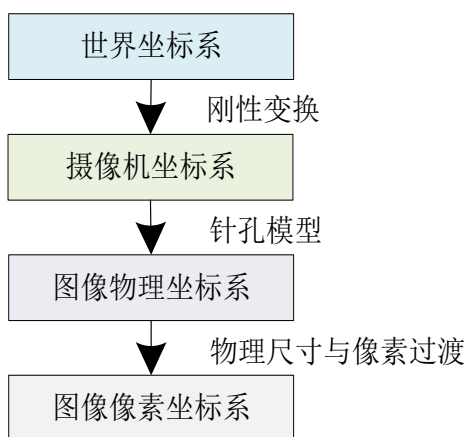


图 8.4 坐标变换图

坐标变换。各坐标系之间具有变换关系，变换关系满足特定规律如图 8.4 所示，这样就可建立像素坐标与空间坐标之间的变换。

(2) 在项目前期工作完成目标体的选取，初步以瓶体位姿估计为例。基本完成目标体的热度图处理。

针对网络收敛速度慢，无法精确回归的问题，利用关键点热度图代替关键点的坐标。热度图是一个大小为 $h \times w \times k$ 的三维矩阵，其中 k 为关键点的个数。一个关键点生成大小为 $h \times w \times 1$ 热度图，多个关键点的热度图堆叠生成 $h \times w \times k$ 的热度图，作为标签替换原来的像素坐标标签，如图 8.5 所示

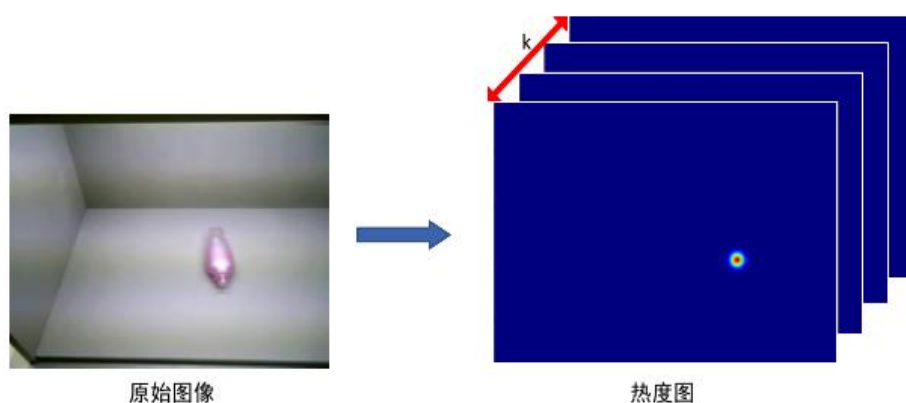


图 8.5 热度图示意图

单个热度图的生成方法是在图片对应关键点所在位置生成一个固定方差的二维高斯分布，生成公式如下：

$$Y = e^{\frac{-4\ln 2 \left[(x-x_0)^2 + (y-y_0)^2 \right]}{\sigma^2}}$$

其中， (x_0, y_0) 表示预测关键点坐标， (x, y) 热度图上像素点的坐标， Y 表示目标热度图。 (x, y) 离关键点越近，数值越接近 1， (x, y) 离关键点越远，数值越接近 0。生成效果如图 8.6 所示。

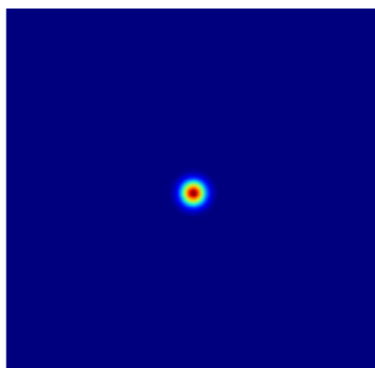


图 8.6 关键点热度图

热度图通过对图像中每个像素位置赋予概率值解决了特征空间到像素坐标 (x, y) 的映射难以学习的问题，每个像素位置都一定程度地提供监督信息，使网络能较快的收敛，同时对每一个像素位置进行预测能够为点回归提供有效的指导，从而提高关键点的定位精度。

(3) 对于级联卷积神经网络的代码框架已经基本完成搭建，目前仍在调试和修改

当中。

算法数据集双摄像机采集左右各 134 张光面瓶体图像，其中 126 张为训练集，包含倒置光面瓶体各旋转角度的样本图像，基 labelim 平台手工标注瓶体目标检测框及关键点坐标标签。左右各 8 张为测试集，旋转角度分别为 $0^{\circ}\sim 315^{\circ}$ ，每张间隔 45° 旋转角度。基于级联卷积神经网络的关键点定位具体步骤如下：

- (a) 将原始图像输入 Faster-RCNN 获得瓶体所在区域的检测框；
- (b) 根据获得检测框确定裁剪范围，裁剪去除大部分无关背景，获得包含所有关键点的光面瓶体图像，同时更新关键点像素坐标，生成对应热图标签；
- (c) 将获得裁剪图像输入粗定位网络，独立训练 5 个粗定位子网络模型；
- (d) 将 5 个粗定位子网络结果取平均，得到粗定位结果；
- (e) 根据粗定位网络四个关键点的检测结果为中心，确定裁剪范围，获得四个关键点的光面瓶体局部图像，同时更新像素坐标标签。
- (f) 将获得裁剪图像输入精定位网络，独立训练 4 个精定位子网络模型；
- (g) 整合 4 个精定位子网络结果，得到最终关键点定位结果。

三级级联网络通过由粗到细的思想精确定位瓶体关键点的位置，通过对每一步的裁剪信息的记录，可反推确定原始图像中各个关键点的像素坐标位置，最终获得关键点预测结果。

2. 已具备的条件，尚缺少的条件及解决方法

- (1) 项目负责人所在实验室设备齐全，包含了多台超算工作站和 AMAX 超算服务器 ServMax™ XG-48202G 等，为本项目开展实验室仿真研究提供了硬件、软件和计算平台的保障。深度学习网络模型训练基于 Anaconda 平台的 Python 语言，Tensorflow-gpu 版本为 1.12.0，CUDA 版本为 9.0。项目负责人和成员对深度学习算法有一定的基础和代码编写能力，对本项目已经完成了初步的模型结构的搭建，和代码调试的部分。
- (2) 本项目已经具备了开展研究工作所需的实验仪器设备、仿真平台与数据处理及运算设备，无需另外添置设备，仅需要申请少量经费购置数据存储和设备借口改造等材料。
- (3) 目前尚缺少对其他各类深度学习算法的深入研究，建立的模型测量结果的精度还不够高。
- (4) 解决办法：继续研究其他算法以及加深对模型的改进，从而搭建精度更高，

测量更稳定的优化模型。

三、 经费预算

开支科目	预算经费 (元)	主要用途	阶段下达经费计划 (元)	
			前半阶段	后半阶段
预算经费总额	10000		5500	4500
1. 业务费	5500		3000	2500
(1) 计算、分析、测试费	0		0	0
(2) 能源动力费	0		0	0
(3) 会议、差旅费	3000	参加学术会议	1500	1500
(4) 文献检索费	0		0	0
(5) 论文出版费	2500	论文版面费	1500	1000
2. 仪器设备购置费	1500	小型设备购买	1000	500
3. 实验装置试制费	0		0	0
4. 材料费	3000	试验耗材购买	1500	1500
5.其他				
学校批准经费				

四、 指导教师意见

曾银雪同学学业基础扎实，实践动手能力强，有一定的科研项目经验，科研能力出色，能够独立查阅相关学术文献，开展学术研究工作。项目来源于智能制造行业的实际需求，研究目标定位准确，研究内容合理，技术路线和时间安排可行性高。项目具有一定的创新性，能够有效锻炼大学生的创新能力。

同意项目申报，并承诺按时给予学生指导。

导师（签章）：

年 月 日

五、 院系大学生创新创业训练计划专家组意见

专家组组长（签章）：

年 月 日

六、 学校大学生创新创业训练计划专家组意见

负责人（签章）：
年 月 日

七、 大学生创新创业训练计划领导小组审批意见

负责人（签章）：
年 月 日