



学 校 代 码 10459

学号或申请号 201922202012908

密 级

郑 州 大 学

专业硕士学位论文

改进 SSD 的平面态散乱小零件分拣研究

作 者 姓 名: 张志浩

导 师 姓 名: 邱 益 副教授

梁 杰 副教授

专 业 学 位 名 称: 工程硕士

培 养 院 系: 机械与动力工程学院

完 成 时 间: 2022 年 5 月

A thesis submitted to
Zhengzhou University
for the degree of Master

**Research on improved SSD plane state scattered small parts
sorting**

By Zhihao Zhang

Supervisor: Prof. Yi Qui

Associate supervisor: Prof. Jie Liang

Master of engineering

School of Mechanical and Power Engineering

May, 2022

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

学位论文作者：张志浩

日期：2022年5月31日

学位论文使用授权声明

本人在导师指导下完成的论文及相关的职务作品，知识产权归属郑州大学。根据郑州大学有关保留、使用学位论文的规定，同意学校保留或向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅；本人授权郑州大学可以将本学位论文的全部或部分编入有关数据库进行检索，可以采用影印、缩印或者其他复制手段保存论文和汇编本学位论文。本人离校后发表、使用学位论文或与该学位论文直接相关的学术论文或成果时，第一署名单位仍然为郑州大学。保密论文在解密后应遵守此规定。

学位论文作者：张志浩

日期：2022年5月31日

摘要

目前,生产线上大量使用的工业机器人仅能进行重复动作,本质上属于示教型机械臂,不具备自主感知和决策能力,不能适应散乱、无序零件抓取和分拣工作。将现有工业机器人和机器视觉技术相结合可以解决以上问题,能够使得分拣机器人具有更高自主性与灵活性。本文系统研究了基于深度学习算法的平面态散乱小零件分拣所需的目标检测算法、细长类零件深度估计与姿态角回归算法以及龙门式三轴机械臂搭建与控制方法。在算法方面,本文完成了算法设计、训练、部署以及测试评价等开发任务,实现了平面态散乱小零件的类型识别、位置回归、深度估计以及姿态角回归等功能。在抓取执行机构设计实现方面,本文完成了三轴龙门式机械臂硬件搭建和控制系统设计,实现了机械臂手眼标定和运动控制。本文主要研究内容如下:

(1) 基于感知哈希算法的目标零件存在判定算法设计:通过感知哈希算法计算两张图片相似度系数,进而判断工作台中是否有待分拣零件。

(2) 基于 SSD 的目标检测算法优化设计:以 SSD 算法为主干,将其根据小零件分拣的特定应用场景引入算法检测头精简、内卷积运算子、特征金字塔等技巧提高该目标检测算法对本数据集的适用性,实现降参增效目的。

(3) 细长类零件深度估计算法设计:根据不同深度所对应不同姿态在工业相机中成像具有极大差异的特点,设计基于深度卷积神经网络的图像分类算法,实现对该类零件不同姿态的识别来达到对细长类零件深度估计的目的。

(4) 细长类零件姿态角回归算法设计:设计并实现基于传统图像处理算法的细长类零件姿态角回归算法,达到对细长类零件在工作台上姿态角检测的目的。

(5) 机械臂搭建与手眼标定:搭建基于步进滑台的三轴龙门式机械臂,并采用九点标定方法实现工业相机与机械臂的标定,得到像素坐标到工作台平面之间坐标变换矩阵。

关键字: 深度学习; 机器视觉; 内卷积; 姿态角回归; 机械臂

Abstract

At present, the industrial robots widely used in the production line can only perform repetitive actions. In essence, they belong to the teaching manipulator. They do not have the ability of independent perception and decision-making, and can not adapt to the grabbing and sorting of scattered and disordered parts. The combination of existing industrial robots and machine vision technology can solve the above problems and make the sorting robot have higher autonomy and flexibility. This paper systematically studies the target detection algorithm required for the sorting of planar scattered small parts based on depth learning algorithm, the depth estimation and attitude angle regression algorithm of slender parts, and the construction and control method of gantry three-axis manipulator. In terms of algorithm, this paper completes the development tasks of algorithm design, training, deployment and test evaluation, and realizes the functions of type recognition, position regression, depth estimation and attitude angle regression of planar scattered small parts. In the design and implementation of grasping actuator, this paper completes the hardware construction and control system design of three-axis gantry manipulator, and realizes the hand eye calibration and motion control of manipulator. The main research contents of this paper are as follows:

(1) Design of target part existence judgment algorithm based on perceptual hash algorithm: calculate the similarity coefficient of two pictures through perceptual hash algorithm, and then judge whether there are parts to be sorted in the workbench.

(2) Optimization design of target detection algorithm based on SSD: take SSD algorithm as the backbone and introduce it into the skills of algorithm detection head simplification, inner convolution operator and feature pyramid according to the specific application scenario of small part sorting, so as to improve the applicability of the target detection algorithm to this data set and achieve the purpose of reducing parameters and increasing efficiency.

(3) Design of depth estimation algorithm for slender parts: according to the characteristics that different attitudes corresponding to different depths have great differences in imaging in industrial cameras, an image classification algorithm based

on depth convolution neural network is designed to recognize different attitudes of such parts to achieve the purpose of depth estimation for slender parts.

(4) Design of attitude angle regression algorithm of slender parts: design and implement the attitude angle regression algorithm of slender parts based on traditional image processing algorithm, so as to detect the attitude angle of slender parts on the workbench.

(5) Manipulator construction and hand eye calibration: build a three-axis gantry manipulator based on the stepping slide, use the nine point calibration method to calibrate the industrial camera and manipulator, and obtain the coordinate transformation matrix from the pixel coordinates to the workbench plane.

Keywords : deep learning; machine vision; internal convolution; attitude angle regression; mechanical arm

目录

摘要	I
Abstract.....	II
目录	IV
1 绪论	1
1.1 研究的背景与意义	1
1.2 国内外研究现状	3
1.2.1 机器视觉技术现状	4
1.2.2 智能分拣机器人发展现状	7
1.3 本文主要研究内容	8
2 平面态散乱小零件分拣系统设计	10
2.1 系统整体架构	10
2.2 系统硬件构成	12
2.3 算法运行平台	13
2.4 图像采集模块与通信方式	15
2.5 机械臂设计方案与控制方式	16
2.6 本章小结	17
3 目标检测与姿态识别算法的设计	18
3.1 算法整体框架	19
3.2 目标物体判定算法	20
3.3 基于内卷积的目标检测算法优化设计	22
3.3.1 目标检测算法整体框架及优化方向总结	22

3.3.2 内卷积计算子	25
3.3.3 特征金字塔设计	27
3.3.4 先验框分布重设计	29
3.4 机械手抓取姿态预测算法设计	34
3.4.1 细长类零件深度估计算法	34
3.4.2 细长类零件姿态角回归算法	36
3.5 本章小结	36
4 算法训练与部署	38
4.1 数据集制作	38
4.1.1 目标检测数据集	38
4.1.2 图像分类数据集	40
4.2 算法训练	41
4.2.1 目标检测算法训练	41
4.2.2 图像分类算法训练	43
4.3 Tensorrt 加速部署	45
4.4 本章小结	46
5 实验研究	47
5.1 机械臂搭建	47
5.1.1 机械臂硬件构成与搭建	47
5.1.2 机械臂控制系统搭建	48
5.1.3 手眼标定	49
5.2 算法性能实验	50
5.2.1 目标检测算法实验	50
5.2.2 细长类零件深度预测算法实验	52

目录

5.2.3 细长类零件姿态角回归算法实验	53
5.3 机械臂执行实验	54
5.4 本章小结	60
6 总结与展望	61
6.1 全文总结	61
6.2 展望	62
参考文献	63
个人简历、在学期间发表的学术论文与研究成果	66
个人简历	66
在学期间发表的学术论文与研究成果	66
致谢	67

1 绪论

1.1 研究的背景与意义

分拣工作是制造业中最为常见的工作。由于分拣工作存在很强的重复性，这也是工业机器人为什么在分拣工位上大批量应用的重要原因^[1-2]。传统制造业中，分拣工位多采用人工，该方法不但效率低下且成本高昂，难以实现不间断工作。采用工业机器人可以在一定程度上实现分拣作业自动化，能够完成大多重复性劳动，进而在保证工作效率的前提下提高产品质量和标准化^[3]。图 1.1 为工业机器人分拣的一般应用场景。

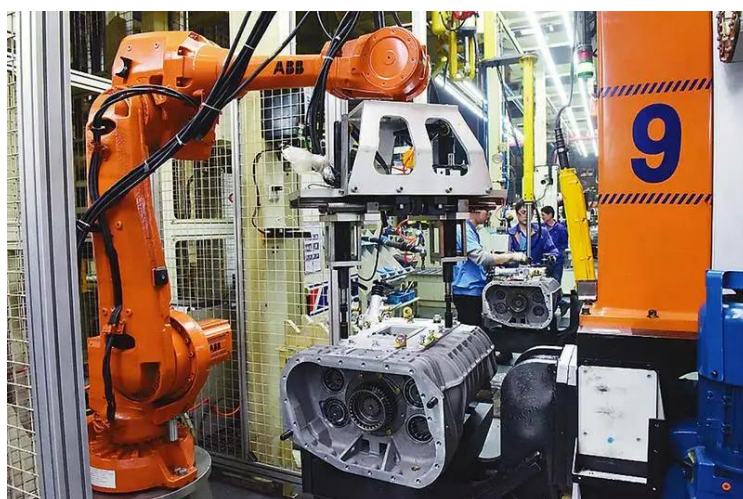


图 1.1 工业机器人分拣的一般应用场景

目前，生产线上主要使用的工业机器人还是只能进行重复动作的示教型机械臂。这类机器人的应用虽然大大提高了生产线的执行效率并保证了产品质量，但是这类机器人不具有适应性，不能根据工位上不同类别工件或者不同姿态工件进行自主地调节抓取位置与抓取角度^[4-5]。以上不足就需要生产线首先将工件按照需求进行摆正才能进行后续操作，然而在实际情况下往往零件类型多样且姿态杂乱，这就极大地限制了现有机器人的应用范围且增加了用工成本^[6-10]。将现有多自由度机器人技术和机器视觉技术相结合可以解决以上问题，能够使得工业机器人具有更高的自主性与灵活性。分拣工作还可以引申到分类码垛、分类

包装、分类装配等等。其应用场景虽多样，但从工作过程上看其具有一致性，都可以用图 1.2 概述。

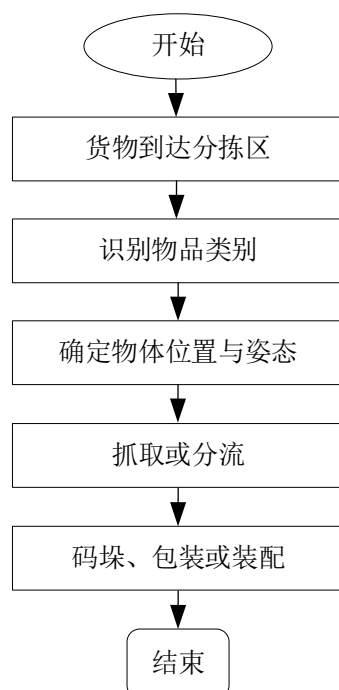


图 1.2 工业分拣流程图

图 1.2 中的分拣流程代表了工业产线中主要工作内容。其中识别物品类别和确定待分拣物体位置与姿态的步骤，长期受困于传统机器视觉技术难以在复杂应用场景下保证识别准确性而难以实现智能化、无人化。如何解决该问题，是实现智能化无人工厂的关键所在，也是人工智能在制造业领域的发力点所在。近些年基于深度学习的视觉算法开始应用于工业生产中，采用传统视觉算法和基于深度学习视觉算法相结合的视觉解决方案在很大程度上兼顾了运算效率和识别精度，这些技术的发展使实现平面态散乱小零件分拣成为可能。将基于人工智能的机器视觉技术与机器人技术结合起来，实现制造领域中分拣工作的完全智能化和无人化，将会提高现有生产线上的生产效率并降低用工成本^[11-12]。将基于深度神经网络的视觉算法应用于工业产品分拣工位中，不但能够提高机器视觉算法的识别精度和稳定性，还可以发挥深度学习端到端的优势，使得分拣工位的可移植性、产线的柔性化能力得到极大提高^[13]。

本课题提出的基于目标检测的平面态散乱小零件分拣研究，其意义在于将最新的基于深度卷积神经网络的目标检测技术、图像分类技术和三自由度龙门

式机械手集合起来,实现平面态散乱零部件的智能化分拣,验证现有基于深度学习的机器视觉技术和传统机器人技术相结合的应用可行性。打通深度学习和制造业相结合的技术路径,进一步提高工业自动分拣系统的分拣质量及鲁棒性。

1.2 国内外研究现状

基于视觉的智能化分拣任务中最重要的工作就是搭建精确且稳定的目标检测和图像分类算法。该算法的性能直接决定智能分拣设备的应用可能性。最初的机器视觉算法基本上都是人工构造目标物体特征,采用传统机器学习或者模板匹配的方法实现图像数据的目标检测,该方法的精度和鲁棒性较差,难以实现工业化应用^[14-16]。多伦多大学的 Hintou 等人在 ImageNet 中率先提出了 AlexNet 网络并在图像分类大赛中取得了第一名的成绩,该网络在 ImageNet 竞赛中超出第二名 41%的准确率。该网络促进了深度神经网络在机器视觉领域中的应用^[17-18]并奠定了其在该领域的统治地位。

随着近几年深度神经网络技术的发展,以深度卷积神经网络为基础的目标检测技术取得了巨大进展。目标检测技术是计算机视觉中一项重要任务,它用于检测可视对象实例属于某一类并给出位置标记。以深度学习为基础的目标检测技术以其优越的检测识别精度和较高的鲁棒性,已经广泛地被工程技术人员应用于实际生产当中,如智慧医疗、智能驾驶等领域。深度学习为基础的目标检测技术在工业自动化产线、城市安全等行业有着越发重要的地位,深度学习算法能够减轻这些行业的人工成本,产生可观的经济效益^[19-22]。深度学习算法的提升主要依靠数据驱动,将样本数据与误差反向传播相结合来寻找最优的算法参数^[23-26],所以使用分布更加全面而广泛的数据训练出来的算法比少量数据训练得到的结果在泛化性能上有更好表现,这也使得深度学习为基础的目标检测技术得以取得更好的应用。目前,国内外的制造业、高科技公司等纷纷投入精力、物力在该领域,致力于将传统制造业和深度学习相结合来实现更加智慧的工厂。如库卡机器人、埃尔深智能科技、梅卡曼德等科技公司都先后推出了各自的智能分拣机器人以及对应的 2D、3D 视觉定位方案。埃尔深智能科技实现分拣任务的演示如图 1.3 所示。



图 1.3 智能分拣机器人

1.2.1 机器视觉技术现状

图像分类技术是机器视觉发展的开端,其发展历史可以追溯到 20 世纪 50 年代,最开始的计算机视觉仅在实验室里被研究人员进行探索,直到 20 世纪 80 年代才形成独立的学科,此时工程人员才逐渐将理论从实验室迁移到工程应用。机器视觉技术在过去几十年的发展历程中,根据其所采用的技术特征可以将其分为两大阶段,一个阶段是 2014 年以前占主导地位的,采用人工设计特征的传统方法,一个阶段是当前广泛使用的,采用深度卷积神经网络的现代方法。其中传统方法中具有代表性的有基于特征的图像识别算法和基于梯度直方图目标检测算法。David Lowe 发表的题为《基于局部尺度不变特征 (SIFT 特征) 的物体识别》,标志着基于特征的图像识别算法开始被研究人员所采用。该方法将采用高斯模糊的方法实现尺度空间变换,采用高斯差分函数完成极值检测,最后得到具有一定空间、明暗不变性的关键点。该方法算法复杂度较高,限于计算能力,运算速度较低^[27-28]。Dalal 等人提出将图像的方向梯度直方图 (HOG) 应用到行人检测上。该方法是目前机器视觉、模式识别技术领域用于图像局部纹理特征描述的方法^[29-30]。

2012 年 Alex 等人在 ILSVRC 大赛中率先设计并训练了一个大型的卷积神经网络 AlexNet, 该网络使其赢得了本次大赛的冠军。它第一次向世人展示了卷积神经网络在图像识别领域中的强大性能。该网络在 ImageNet 数据集上的识别准确率达到了 84%, 与人眼的识别能力相当, 它的出现极大促进了卷积神经网络在计算机视觉的应用。在今后的几年中, 先后出现了性能更为优秀的卷积神经网络,

如 2014 年牛津大学的计算机视觉组提出了新的卷积神经网络 VggNet, 该网络在 ILSVRC2014 比赛中的图像分类项目中取得了第二名和定位项目的第一名。VggNet 很好的探索了深度卷积神经网络的卷积层深度和算法性能之间的关系^[31-32]。

2015 年, 何凯明提出了全新的深度卷积神经网络 ResNet, 该网络很好的解决了卷积神经网络层的深度和由卷积层增加而导致的梯度消失问题, 缓解了深度卷积神经网络的难训练问题。引入残差结构的卷积神经网络 ResNet 是目前使用最为广泛的卷积神经网络, 包括最新的目标检测或语义分割网络, 都采用了残差结构的思想^[33-34]。

2017—2018 年, 深度学习框架逐渐走向成熟, 经过多年的发展与验证, Pytorch 和 TensorFlow 成为目前应用最为广泛的两大框架。国内各大科技公司与智能制造公司逐渐采用以上两种开源框架实现深度学习的实际应用。

在深度卷积神经网络逐渐优化改进以及开源深度学习框架逐渐走向完善的背景下, 采用卷积神经网络作为主干特征提取网络的目标检测算法开始被提出。该类算法网络不再使用人工设计特征点的方法, 而是采用多层连接的神经网络提取图像底层到高层的特征, 该方法能够在没有人工干预的前提下提取图像中有价值的特征, 能够实现端到端的训练和部署工作。总的来说, 基于深度学习的目标检测算法分为两个类别, 一种是首先生成感兴趣区域框并在每个区域框内进行分类判定的目标检测方法, 该方法的典型代表就是 R—CNN 系列算法。另一种是直接采用主干特征网络在整个图像数据中提取全局特征, 再分别在多层次特征图上进行全卷积计算, 进而回归出目标区域位置和所属类别的置信度, 该技术路径的典型代表有 SSD 和 Yolo^[35-37]。

2013 年, R—CNN 系列的算法开始被提出, 该类算法在业内又常被称为两步法, 顾名思义就是采用串联的两大步完成目标区域的检测。其中第一步是采用 RPN 网络生成感兴趣区域并用卷积神经网络提取特征, 第二步就是在感兴趣区域上设计分类器对其进行分类并修正目标框位置。以 R-CNN 网络为代表的两步法先后经历了 3 次大的改进, 第一次是在 2014 年由何凯明博士提出了 SPP-Net, 其目的是采用一个卷积神经网络对整张图进行特征提取并返回 RIO 区域, 该方法虽大幅减少计算量, 但仍没有解决网络占用过多内存空间的不足。在此基础上, RBG 又将 SPP-Net 进行了改进, 提出了 Fast R-cnn 网络, 该方法合并了 SPP-Net 中的分类运算和边界框回归运算, 同时还引入了 SVD 分解, 进一步降低了计算

量。以上改进仍然都是使用选择搜索的方法生成候选区域，计算量仍然很大。基于以上考虑，研究人员又提出了 Faster R-CNN 目标检测网络，该网络率先提出了 RPN 网络并将其应用于选择搜索中。该网络的应用标志着端到端的计算方法已经引入到了目标检测算法中。两步法系列的目标检测算法在检测精度上具有优势，但是由于结构上是串联的，导致其运算速度不高，难以在视频流上实现实时检测。

2015 以后，针对两步法运算速度低的问题，研究人员尝试了新的算法思路，即采用全卷积的方法直接完成目标物体的位置回归和类别识别。该类算法最具代表性的算法有 SSD 系列和 YOLO 系列算法。从前向传播的流程看，One-Stage 目标检测算法都是首先将输入数据标准化为固定尺寸，然后再在固定尺寸图像上引入一个较深的卷积块用于特征提取，一般称该卷积块为主干网络，主干网络用于提取从浅到深的特征图。从检测逻辑上看，该类网络的检测方法是在上述主干网络得到的多个特征图上布置多组检测矛筐，所述矛筐以特征图的像素点为中心点，每个中心点分布多个矛筐，其尺寸和长宽比不相同。每个检测头在对应特征图上做全卷积运算，得到上述矛筐中存在待检测目标物体的概率和实际坐标位置。由于该方法是利用整张图像的全局特征，直接通过全卷积神经网络得到对应的结果，所以检测速度相对于两步法有很大的优势。以 2020 年发布的 YOLO-V5 目标检测算法为例，其检测精度已经可以满足工业化的应用需求。图 1.4 为 YOLO-V5 目标检测算法网络结构图。

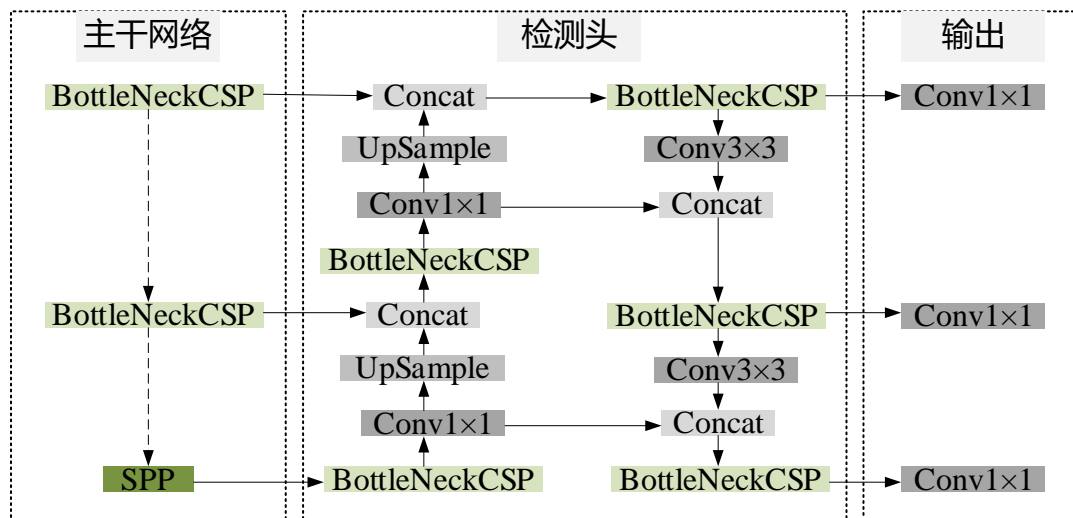


图 1.4 YOLO-V5 目标检测算法网络结构图

基于深度学习的目标检测技术主要由国外科研机构提出，并且目前用于搭建深度学习算法的框架，也是由美国谷歌公司和 Facebook 人工智能研究院所提供。国内目前对于该方面的开创性研究成果数量不足，但我国在应用方面做的较好。目前，我们国家很多制造企业都开始将基于深度学习的目标检测技术应用实际生产中，例如，国内最大的生猪养殖、屠宰公司——牧原集团已经将该技术深度融合到生猪的养殖、健康检测、疾病快速诊断、无人化猪舍、以及智能化屠宰等业务中。国内的科技公司，如阿里巴巴、百度、科大讯飞等已经研发出了性能极为优越的算法框架，并将该算法通过 API 接口的方式向社会提供对应的深度学习计算服务。我国对于该领域的深度应用，必将极大促进我国生产力的发展，对于我国制造业能力的提升起到举足轻重的作用。

1.2.2 智能分拣机器人发展现状

智能分拣机器人是智能制造背景下重要产物，其发展水平在很大程度上代表了智能制造水平。由于在工业产线上，抓取和分拣作业是主要工作，其往往是由生产工人完成，如果将该部分工作用智能机器人代替，就能将该批工人从枯燥乏味的重复性工作中解放出来。智能分拣机器人的应用领域非常广泛，由于各个行业有其特殊的工艺需求，所以各个领域智能分拣机器人的结构形式、控制方法以及视觉算法所采用的解决方案都有差异。比如水果生产企业，其所采用的智能分拣机器人要能够分辨出水果成熟度、水果大小、水果表面缺陷等指标。例如屠宰企业中用于智能包装的分拣机器人，其要求能根据图像识别出目标物体类别。在制药领域，分拣机器人要具备识别药片颜色、表面缺陷以及计数等功能^[38-41]。

2014 年以后，在智能分拣视觉算法设计领域中，工程人员往往采用传统图像处理 and 深度学习相结合的方法。基于深度学习的视觉算法发展至今，由于抗干扰能力强、环境适应性好以及算法鲁棒性高等优点，工程人员已经开始利用该工具去解决更为复杂应用领域中。同时，由于传统图像处理算法在运算速度上具有更低延迟率，且由于其发展多年而积累多种开发途径，所以其应用兼容性和开发周期具有更大优势。在实际应用中，往往可以将深度学习方法和传统图像处理方法相结合，形成高低档搭配来发挥各自优势，去更高效地解决复杂的视觉应用场景。

国内近几年在智能分拣领域大力发力，国内很多企业都希望能用该技术改造自身生产工位，实现减人增效目的。国内伍锡及其团队研发了基于机器视觉的

工件分拣机器人，其采用的算法就是将传统图像处理算法和深度学习相结合。采用传统图像处理方法确定出工作区域中工件位置，然后将存在目标物体区域图像剪裁下来，再通过卷积神经网络做多类别图像分类来进一步判别该区域中目标物体是什么。国内牧原肉食集团智能化部门利用基于深度学习的目标检测技术实现肉制品气调包装分拣。随着近些年电商大力发展，国内京东、三通一达每天将处理非常多快递包裹，为提高分拣效率，物流智能分拣机器人发挥了关键作用。

在国际社会中，特别是以美、日、英为代表的发达国家，由于其人力资源相对短缺，所以其对智能分拣机器需求量会更大。例如日本 FANUC、EPSON 以及德国库卡机器人等公司，都在各个应用领域中设计研发智能分拣机器人。例如 FANUC 公司开发的 M-1iA 型智能分拣机器人，其具有结构紧凑、灵活等特点，该机器人内置视觉运算模块能够实现智能分拣、装配、码垛等功能。图 1.5 为食品包装生产领域智能分拣机器人。



图 1.5 智能包装机器人

1.3 本文主要研究内容

本文在基于机器视觉智能分拣研究背景下，分析传统机器视觉优缺点以及该类算法在智能分拣机器人中的应用现状，分析其在现有应用场景下的优缺点。同时，本文还分析基于深度学习技术的一系列图像处理算法，指出其相较于传统图像处理算法的优点以及对应的技术特点、解决方案等，以得出其在智能分拣任务中应当扮演什么角色。在算法设计方面，本文为解决传统图像处理算法在智能

化分拣中表现出鲁棒性不足等缺点，设计了一套采用深度学习和传统图像处理的机器视觉算法，该算法包含目标物体存在判定功能、目标物体位置回归功能、目标物体类别识别功能、目标物体中细长类零件深度与姿态角度预测功能。该算法相互配合能够根据工业相机采集到的图像数据计算出目标物体在图像中中心点坐标、包围矩形框、以及机械手对于细长类零件抓取姿态角。上述运算结果通过机械手与工业相机之间手眼标定矩阵进行变换，即可得到能够直接控制机械手抓取的位置信息。最后上位机再将位置信息通过串口通信发送给下位机，下位机控制机械臂即可完成小零件抓取。论文的主要研究内容如下：

（1）第一章，绪论；本章首先说明了本课题的研究意义以及研究目的，在此基础上得出本课题需要解决的技术问题以及对应需要进一步优化方向。然后介绍了机器视觉技术近 20 多年的发展历程，得出深度学习在 2014 年以后开始超越传统图像处理方法等结论，给本课题进一步研究指明方向。

（2）第二章，平面态散乱小零件分拣系统设计；本章分析了平面态散乱小零件分拣系统的整体解决方案，给出了该分拣系统整体结构图，并在整体系统结构基础上分析了系统硬件构成并给出硬件构成图。同时，本章还分析了本系统中算法运行平台、工业相机选型、通信方式以及机械臂机构形式等基本设计要素。

（3）第三章，目标检测与姿态识别算法的设计；本章主要分析并设计了该分拣系统所用识别算法，包括目标检测算法优化设计、细长类零件抓取姿态识别算法设计等基本算法模块。

（4）第四章，算法训练与部署；本章首先介绍了数据集搜集、以及标注工作，分析了目标检测算法、细长类零件深度预测算法的训练方法与训练过程。分析了算法训练阶段如何找到最优的网络权重，也即是如何训练出最好的算法，最后还着重分析了算法的部署过程以及利用 TensorRT 进行算法结构加速优化。

（5）第五章，实验研究；本章分析了机械臂构成与搭建过程、基于九点标定原理的机械臂手眼标定以及目标检测算法、深度预测算法以及姿态角回归算法在验证数据集上的测试精度。本章还通过单片机向机械臂发送运动控制指令，记录目标位置 and 实际运动位置以及二者误差，得出机械臂运动精度和引起误差的原因。最后测试了给机械手在算法引导下的实际运动精度和抓取效果。

（6）第六章，总结与展望；对本文的技术方案进行总结以及提出了自己的不足，并对以后进行了展望。

2 平面态散乱小零件分拣系统设计

2.1 系统整体架构

本实验平台以工业标准件抓取为研究背景，采用深度卷积神经网络实现了标准件识别、定位以及抓取姿态预测。该实验平台将算法理论和应用结合，采用开源的神经网络框架实现算法，采用 STM32 配合步进滑台完成硬件搭建。

自动分拣系统是将机器视觉技术和工业机器人技术相结合的一种智能化设备，它包括了算法模块和机械臂模块。其中，算法模块的设计工作包含算法结构的设计、算法实现、数据集制作、算法训练、算法部署等工作，机械臂模块的设计工作包含机械臂的搭建、机械臂控制系统设计、机械臂与相机的手眼标定等工作。在此基础上，整个系统的整体架构如图所示。

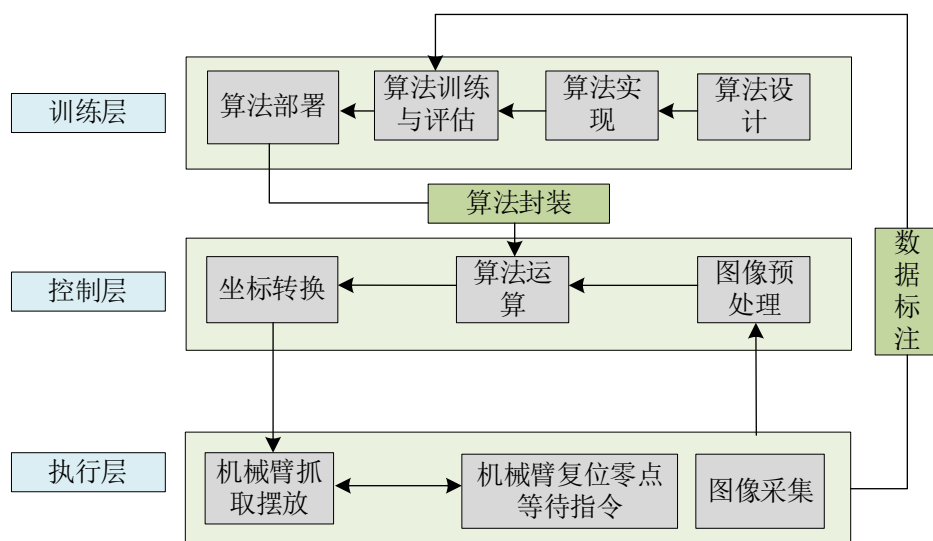


图 2.1 系统整体框架

该分拣系统从业务逻辑上可以分成三个层面，分别是训练层、控制层和执行层。

训练层：训练层是在分拣系统开始执行分拣任务之前需要进行的工作，即要让系统知道需要抓取分拣的目标物体类别并具备目标物体识别和定位的能力。从工作内容上看，训练层要根据待抓取目标物体的特征以及复杂程度设计合适

的识别算法并采用深度学习框架完成算法的搭建。同时，还要大量的采集目标物体在工作台上图像，在图上进行数据特征标注并形成数据集，最后用数据集去训练设计的算法。当算法在数据集上的识别效果满足应用要求时即认为完成了训练层面的工作。

控制层：控制层的作用是将图像数据应用训练层得到的算法进行识别并给出机械臂控制信号。在控制层，摄像头完成图像数据的采集并进行图像的前期处理，然后将预处理的图像传输给训练层得到的算法，得到目标物体在图像中像素坐标与对应的类别，最后再根据手眼标定的坐标转换矩阵完成像素坐标到机械手的世界坐标的转换。控制层是分拣系统在工作阶段循环调用的模块，该部分的工作要高效、准确。

执行层：执行层属于硬件层面的工作，首先通过工业摄像头采集图像数据，然后将采集的数据传输给控制层，控制层将根据图像计算得到的结果返回给执行层，执行层将得到的世界坐标数据转换为机械臂控制的脉冲信号，使得机械臂末端的机械手运行到目标物体的位置中心点，然后根据目标物体类别信息，将其抓取到对应位置。这里需要指出的是，在训练阶段，摄像头采集的数据要保存下来进行数据标注，作为训练集的数据来源。

该自动分拣系统中所用到的算法都是采用的深度卷积神经网络作为算法的基础方法。相较于传统图像处理方法的特点是去除了人为构造待检测物体特征以及构造匹配模板的过程。这样做在前期模型训练阶段具有的好处就是能够实现端到端的训练，即可以从数据直接得到想要的识别算法模型。即是当完成好算法设计之后，操作人员就可以将算法模型本身完全当作一个黑箱，当产线需要更改待识别物体的类别时，操作人员只要完成数据的采集和标注工作，然后将数据“喂”给算法模型即可得到想要的模型。对于操作人员来说，几乎不用修改算法本身，最多修改一下算法的配置文件即可。另外，值得一提的是，基于深度卷积神经网络的视觉算法由于是由数据驱动的，通过合适的数据增强手段，可以做到对于背景变化具有适应性的能力，这大大提高了识别算法的鲁棒性，对于设备的工作环境要求更低。

2.2 系统硬件构成

基于机器视觉的智能分拣系统是计算机与机械相结合的现代化产物，由于基于深度学习视觉算法的数学模型往往具有很多的参数，这也就要求计算机的计算性能与内存容量足够高。计算能力在模型训练阶段将会影响到训练迭代的周期长度，进而影响算法的训练时间，而在部署应用阶段，计算能力将会直接影响分拣系统对一张图片的识别时间，这会直接影响到机械手的分拣效率。计算机内存容量将会制约算法模型的大小，也即视觉算法的数学模型参数量，由于算法模型的参数量决定算法的深度，进而会直接影响算法的识别性能。当然，一切都要坚持适度原则，过高的计算性能如果得不到有效的发挥也是浪费，所以要在满足需求的基础上做到不浪费。

分拣系统主要由单目工业摄像机、龙门式抓取平台、搭载 GTX1070 显卡的计算机、STM32 单片机等部分组成。单目工业摄像机用于采集工作台上待抓取零件的图像信息以引导机械手准确抓取。三轴龙门式抓取平台负责执行抓取动作，分拣系统的末端执行器是两指式气动机械爪，能够实现物体的夹取。在算法的训练阶段，搭载 GTX1070 显卡的计算机负责处理摄像机所采集到的图像，即是将收集制作的数据集传输给所设计并搭建好的算法中，进行多次的迭代运算并通过一定的评价手段选择出满足使用要求的算法模型。在分拣系统的部署运行阶段，Jetson TX2 嵌入式计算卡通过运行训练阶段得到目标检测算法和抓取姿态预测算法来返回待抓取目标的类别、位置和机械手的抓取姿态。STM32 单片机负责将计算机运算的结果转换为龙门式抓取平台的动作指令来控制机械手完成对应的抓取动作。分拣系统的硬件设计架构如图 2.2 所示。

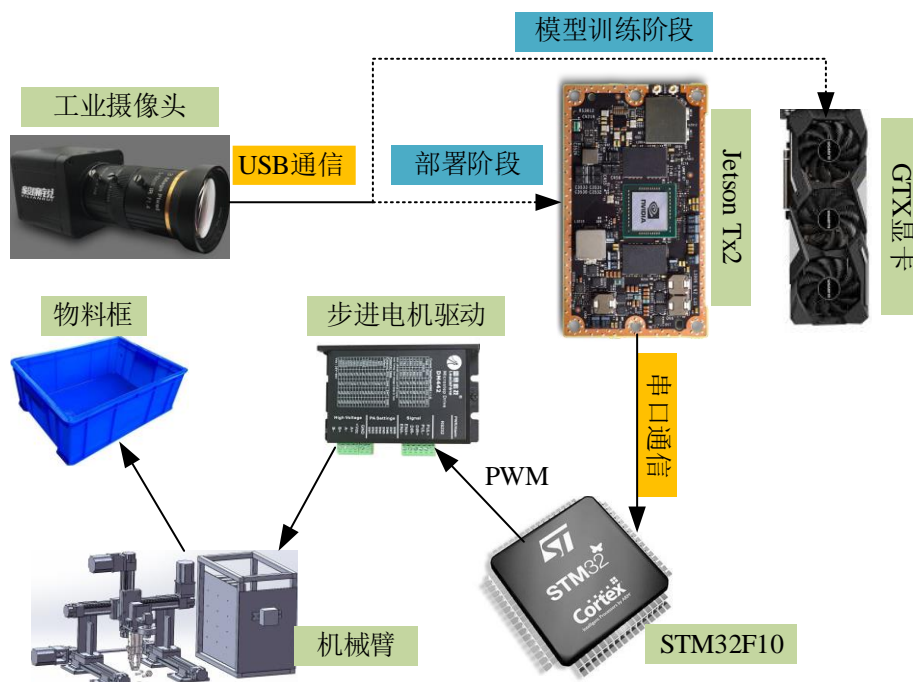


图 2.2 硬件设计架构

分拣平台的硬件构成从整个业务逻辑上可以分为两条主线，一条是模型的训练阶段，一条是部署运行阶段。在模型训练阶段，工业摄像头采集待抓取零件在工作台上不同位置、不同姿态下的图像，将图像数据保存下来并人工进行标注后形成数据集，然后将数据用于算法模型的训练进而得到最优的算法。在模型训练阶段，只有数据采集和算法的训练，不进行上位机和机械臂控制器的通信与动作控制。在部署应用阶段，不再进行算法的训练工作，仅在 Jetson TX2 上运行训练好的算法，将摄像机采集到的图像数据进行处理并得到机械臂的控制信号，然后控制机械臂将对应的目标物体进行夹取、摆放到对应的物料框中。

2.3 算法运行平台

如上所述，算法的运行平台要兼顾效率和成本这两方面的设计要求。由于模型的训练阶段需要做的工作较多，这包括算法的搭建、调试、训练结果可视化、训练结果的保存与转换、数据集的制作等等工作，涉及到的工作面较为宽泛，所以在模型训练阶段需要通用性较强的计算机以便于处理一些较为复杂的工作。同时，在深度学习模型的训练过程中，需要常常多次尝试不同的算法结构和较为

复杂的算法去寻找最优的算法方案，这往往需要大量的计算和存储资源，所以，模型训练用的计算设备需要配 GPU，用来加速训练的整个过程。根据以上应用需求，这里采用搭载 GTX1070 显卡的台式机作为模型训练阶段运算平台。该算法平台的硬件资源和配套的软件资源如表 2.1 所示。

表 2.1 算法平台硬件资源与软件配置

名称	配置	名称	配置
CPU	AMD 3500X	操作系统	Win10
GPU	GTX1070(8G 显存)	虚拟环境	Anaconda
内存	16G (DDR4)	深度学习框架	Tensorflow&Pytorch
SSD	512G	编程环境	Python&C

在算法部署阶段，算法平台所做的计算任务相对单一化，即是将训练阶段所得到的算法进行前向传播计算得到识别结果即可，所以为了节约成本，在算法的部署应用层面可以选择具有针对深度学习单独优化的计算平台。另一方面，在部署应用层面，算法的计算速度越快越好，并且是部署阶段并不做算法的改进优化，对算法模型参数的精确度可以适当的降低，这样能够在满足计算精度的前提下尽可能提高运算速度进而提高分拣系统的实时性。为满足以上性能需求和成本要求，在部署应用阶段采用 Jetson TX2 嵌入式深度学习计算卡作为运算平台。Jetson TX2 是英伟达公司开发的一款针对嵌入式计算应用需求的高性能、低功耗的计算平台。该计算卡可在功耗仅有 7W 的计算模块上实现快速的人工智能运算，该计算卡基于 NVIDIA Pascal 架构，支持 8GB 内存和 59.7GB/s 内存带宽。同时，该计算卡还具有多种标准的硬件接口且体积极小，可以将其轻松应用于多种产品当中。它已经在开源社区被开发者们广泛应用于机器人、无人机、智能便携式设备等场景。Jetson TX2 嵌入式深度学习计算卡的硬件参数如表 2.2 所示，Jetson TX2 计算卡在本项目所需的软件资源如表 2.3 所示。

表 2.2 Jetson TX2 硬件参数

名称	配置
CPU	Dual-Core NVIDIA Denver 2 64-Bit CPU & Quad-Core ARM® Cortex®-A57 MPCore
GPU	256-core NVIDIA Pascal™ GPU architecture with 256 NVIDIA CUDA cores

Memory	8GB 128-bit LPDDR4 Memory 1866 MHx - 59.7 GB/s
Storage	32GB eMMC 5.1
Power	7.5W / 15W

表 2.3 Jetson TX2 软件参数

名称	软件环境的功能介绍
操作系统	Ubuntu 16.04
TensorRT	TensorRT 是英伟达推出的一款基于 CUDA 和 cudnn 的神经网络推断加速引擎
CUDA	英伟达为图形处理单元（GPU）上开发的通用并行计算平台和编程模型。利用 CUDA，开发人员可以利用 GPU 的强大功能，大大加快计算应用程序的速度。
cuDNN	由英伟达打造的针对深度神经网络的加速库，是一个用于深度神经网络的 GPU 加速库
OpenCV	跨平台的计算机视觉和机器学习软件库

Jetson TX2 计算卡将训练阶段得到的模型，通过 TensorRT 库加速部署在本地并封装成动态链接库，在业务逻辑层面调用该算法模型的动态链接库即可。该方法将 Python 代码搭建并训练好的模型转换成二进制的可执行文件，实现了算法的安全保护与计算的加速，是目前较为可靠的解决方案。从业务层面看，Jetson TX2 通过 USB 通信，接收工业摄像机采集到的图像，并将识别的结果通过手眼标定的转换矩阵，完成目标物体中心点的像素坐标系到世界坐标系的转换，即可控制机械手完成对应的分拣工作。

2.4 图像采集模块与通信方式

由于小零件分拣系统需要将摄像头采集到的图像数据通过手眼标定方法转换为世界坐标，进而才能够控制机械手动作，这个过程中就要求摄像头具有较小的图像畸变。为满足该要求，采用小畸变工业摄像机较为合适。考虑数据传输低延时性以及 Jetson TX2 实际外接设备资源，采用 USB3.0 作为摄像头和计算机之间通信传输方式，具体信号为 TX-USB-1200W。图 2.3 为 USB 通信工业摄像机。



图 2.3 工业摄像机

2.5 机械臂设计方案与控制方式

机械臂是运动执行装置，负责在 X 、 Y 、 Z 三个方向上运动，来将末端夹取机构运送到指定的抓取位置上去。由于本抓取系统是面向于平面态的小零件抓取，为尽可能的降低成和控制复杂度，采用三自由度龙门式机械手作为运动执行装置。三轴龙门式机械臂由相互串联的三个滑台作为机械臂本体，同时作为辅助装置，机械臂还包括单片机控制器、限位开关、步进电机驱动器等装置。步进电机是一种运动相对精确的驱动电机，其运动距离和速度严格按照控制脉冲执行。采用龙门式三自由度机械臂这种结构形式，能够尽最大程度降低机械臂运动控制部分的设计难度。

该机械臂采用上位机和下位机相互配合的控制方式，上位机和下位机采用串口通信进行信息交互。在训练测试期和部署执行期，上位机采用台式电脑和 Jetson TX2 嵌入式计算机，由于该类计算机具备较强运算能力和存储能力，比较适合运行大型算法来处理图像数据，而由于该类计算机并不具备 IO 口等控制信号接口，所以不适合直接用来驱动步进电机。为解决该问题，采用 STM32 单片机作为下位机，辅助上位机将算法结果转换为机械臂直接控制信号。

2.6 本章小结

本章分析了平面态散乱小零件分拣系统的整体解决方案，给出了该分拣系统整体结构图，并在整体系统结构基础上分析了系统硬件构成并给出硬件构成图。同时，本章还分析了本系统算法运行平台，分析不同阶段对运算平台需求特点，在此基础上确定出训练阶段和应用阶段的算法运行平台。最后，本章给出工业相机选型、通信方式以及机械臂机构形式等基本设计要求。

3 目标检测与姿态识别算法的设计

平面态散乱小零件分拣系统的研发过程中，目标检测算法与姿态识别算法是最主要的设计工作，该部分的工作最具挑战性，且该部分工作将直接影响到分拣系统的工作可靠性和效率。

机械手要想完成对零件的分拣工作，首先要解决两个问题，一个是如何识别零件的类别，一个是如何确定零件的位置。解决这两个问题可以采用目标检测技术。目前主流的目标检测技术已经能够在视频流上实现实时的高精度多分类检测，能够准确的回归出目标物体在图像上最小包围矩形框和类别信息。

采用训练完成的目标检测算法对图像进行识别，能够在图像上给出目标物体的包围框和其属于某一类别的置信度，一般在算法中会设置一个阈值，当阈值在某一类别下达到时就认为该区域就属于该物体。将得到的像素坐标通过手眼标定转换为世界坐标即可实现抓取和分拣。

目标检测算法虽然解决了类别识别和定位，但不可忽略的是，在实际机械手抓取过程中，还有一个抓取姿态问题。所述目标检测算法仅给出了目标零件在图像中的包围框的中心点，但是并未给出目标物体在图像中的姿态。以细长杆为例，要想稳定的抓起，仅仅知道中心点坐标还是不够的，这是因为细长杆沿零件轴向方向抓取显然是极其困难的，如果机械手平面的垂线和细长杆的轴线垂直，那么抓取起来将会非常容易。

以细长类零件螺栓为例，可将其在工作台上的姿态离散的分为 5 大个类别，分别是竖直一类和平躺 4 类（按照姿态角分）。分类情况如图 3.1 所示。要想准确的完成抓取，算法要能后准确回归出细长类零件在工作台上的角度。同时，如姿态 5 和姿态 6，在特殊情况下，一旦螺栓处于竖直状态，其高度和处于平躺状态的螺栓将差别很大，此时机械手的抓取高度需做出调整才能适用该姿态，由于竖直姿态和平躺姿态是离散的真值姿态空间，且二者在图像上的差异较大，所以可将平躺姿态估计和竖直姿态估计问题转化为图像分类问题，即可设计对应的具备图像分类能力的卷积神经网络对工作台上的零件进行姿态预测。而对于平躺姿态，其角度回归计算可采用传统图像处理方法完成。



图 3.1 螺栓姿态

根据以上对智能化机械手抓取所需控制信息的分析,可以得到如下结论,散乱小零件分拣系统在算法层面需要完成目标物体的识别、定位以及对于细长类零件的姿态判断这三部分工作。根据各个需求特点和现有算法理论,从技术层面上分析,算法工作包括基于深度卷积神经网络的目标检测算法设计以及图像分类算法设计。所以,本章主要完成目标检测算法和图像分类算法的优化设计。为保证算法在满足精度要求的前提下,尽可能的轻量化,在目标检测算法设计中引入了内卷积的计算单元,即采用内卷积计算子替换掉目标检测算法中的标准卷积,同时该部分还引入了特征金字塔,矛盾聚类概念来使得通用目标检测算法更好的适应于本项目的数据集。针对图像分类算法,引入特征融合的思想与原理,强化了现有残差卷积神经网络的特征提取性能,有效的保证了细长类零件的姿态识别精度。

3.1 算法整体框架

算法的整体流程图如图 3.2 所示。

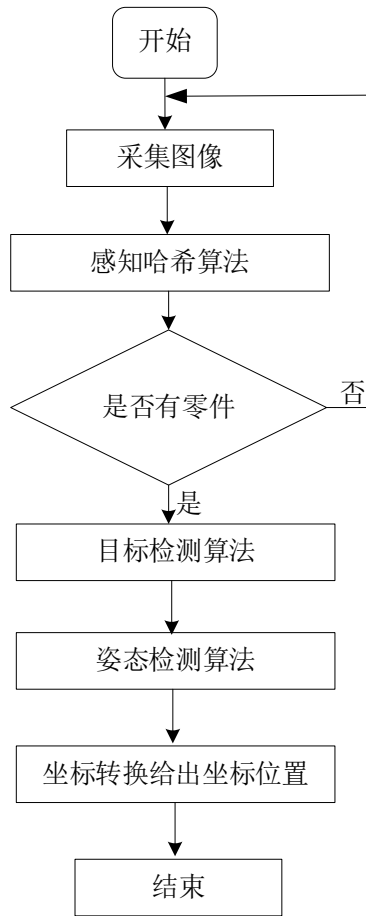


图 3.2 算法整体流程图

考虑到基于深度卷积神经网络算法所需要占用的计算资源较大，如果在整个视频流上都运行目标检测算法进行计算，将会造成大量的资源浪费。所以，为避免这个问题，算法逻辑上首选采用一个占用运算与内存资源极少的算法，先对工作平面的图像数据进行一个初检，先判定目标区域是否有物体出现，当没有物体在工作平面上时，算法控制逻辑将不再调用后面的目标检测算法做进一步判断。当发现工作台平面上存在物体的时候，再去调用后面的目标检测算法去进一步判断工作台平面上物体的位置、类别以及对应的姿态。

3.2 目标物体判定算法

目标物体判定算法是整个分拣系统视觉识别算法的第一步，其所做的工作非常简单，就是对采集到的工作台上的图像数据做一个简单有与无的判定，并不

对目标物体本身做细节判定。所以要求算法应当尽可能的轻量化,要占用资源少,速度快。为满足以上设计要求,可采用感知哈希算法完成该部分工作。

感知哈希算法^[42-43]是对一类算法的总称,它包括平均值哈希算法、感知哈希算法、差异值哈希等算法。其中,差异值哈希算法以其算法运算速度快,精确度高而被广泛应用于图像的相似度判别领域。上述感知哈希算法,其技术原理就是针对每张图片,通过图像降采样、灰度化等处理后采用对应的差异值计算方法计算生成一个字符串,将该字符串作为该图片的“指纹”。最后,对不同的图片做同样的运算,得到各个图片的指纹信息,再去比较不同图片生成的指纹字符串。比较的方法是计算不同图片间的汉明距离,汉明距离越小,说明指纹间的字符串越相近,也就是说明图片越相似。由于差异值哈希算法具有一定的技术优势,所以这里采用该方法,基于差异值的感知哈希算法的实现步骤总结如下:

(1) 图像降采样,缩放成固定尺寸;由于对该算法的技术要求是高效、轻量化,所以计算量要尽可能的少,所以第一步要进行降采样。通过降采样,能够将大尺寸图片中的冗余信息去除,隐藏该算法中不需要的细节部分,做到只见“森林”,不见“树木”。一般做法是将其缩放为 9×8 的像素尺寸,理论上虽然可以将其缩放为任意大小尺寸,但是缩放成这个值是比较合理的。这是因为后期做差异值计算时,宽度为九有利于转换为 hash 值。

(2) 灰度化,将彩色图像转换为 0—255 的具有明暗信息的灰度图;差异值哈希算法的“指纹”生成是通过计算灰度图像中每一行相邻像素之间的颜色强度差异而得到的。虽然经过大比例降采样后的图像已经隐藏了图像中的色彩、纹理等细节信息,但是,由于它是彩色图像,所以图像信息量还是比较大的,还有进一步压缩的空间。这是因为,彩色图像是由 RGB 值组成三通道张量,其中白色表示为 (255,255,255),黑色表示为 (0,0,0),值越大颜色越亮,越小则越暗。如果直接采用彩色图像进行颜色强度计算,计算的结果将是三通道的,运算量将会变得很大。如果将三通道图像转换为灰度图,计算量将会减少三倍,且对于结果影响不大。

(3) 差异计算;差异值计算是通过计算上述降采样得到的 9×8 个像素的特征图而得到的。计算方法是逐行进行的,即每行的相邻两个像素值做差,如果前一个像素的颜色强度大于第二个像素,那么差异值就设置为 True,也即是设置为 1,如果比后一个像素灰度值小,那么差异值就设置为 False,也即是 0,以上计算得到的这个 64 位的二进制值就是这张图片的指纹。举个例子,假设 9×8 的

降采样特征图数据表示如下： a_1 — a_9 、 b_1 — b_9 、 c_1 — c_9 …… h_1 — h_9 。即计算 a_2 — a_1 、 a_3 — a_2 、 a_4 — a_3 …… a_9 — a_8 并记录下来。 b 行， c 行等以此类推，一共记录 64 个二值化值，即为该图像的“指纹”。

(4) 将指纹图转换为哈希值；为计算方便，这里将差异数组中的每一个值都看作一个 bit，再将每 8 个数据组成一个 16 进制的数值，最后再将 16 进制的值相互连接到一起组成字符串，这样运算就可以得到包含一张图像身份信息的哈希值。

(5) 计算汉明距离；通过步骤一到步骤四的计算可以得到一张图片的身份指纹信息，那么，对于两张图片，如何判断这两张图片是相似的呢，评价方法就是汉明距离。所述汉明距离是信息论中的概念，表示两个等长的字符串在对应位置上字符有不同的数目。例如“abcdef”和“ascrdf”的汉明距离为 3。一般来说，两张图片指纹信息的汉明距离小于 5 时，其基本就是同一张图片。实际测试可以根据图像样本的实际情况，判断汉明距离临界值为多少。

感知哈希算法本身是用于计算两张图片相似度系数的算法，在本分拣系统中，可以通过计算工业摄像头采集到的前后几帧图片的相似度来间接判断图片中是否放置上去了待分拣的零件。

3.3 基于内卷积的目标检测算法优化设计

3.3.1 目标检测算法整体框架及优化方向总结

目标检测技术作为机器视觉领域中极具挑战性的研究方向，在神经网络和强大算力支撑下已经开始应用于各个领域。但是，现有通用目标检测算法普遍存在缺陷，即对视野中占据像素较少的小目标检测效果不理想。图 3.3 为 SSD 目标检测算法^[44-45]结构图，该算法是 Liu W 等人提出的一类单阶段目标检测算法，该类算法相较于两步法目标检测算法具有很大的检测效率优势。该目标检测算法分为两大部分，一部分是主干网络，由 VGG16 特征提取网络充当。另一部分为检测头，也是由全卷积神经网络所构成的用于目标物体位置回归和类别识别的算法块。检测头分布在 VGG16 网络中不同深度的六个特征图上，从浅层到深层一共选择了六个特征图作为检测头的输入，采用这个布置结构能够产生语义信息丰富的多尺度特征图。其中，浅层特征图具有较大的分辨率，用于检测小目

标, 高层特征图虽具有较低分辨率, 但是语义信息丰富, 主要用来检测占据像素较多的大目标物体。

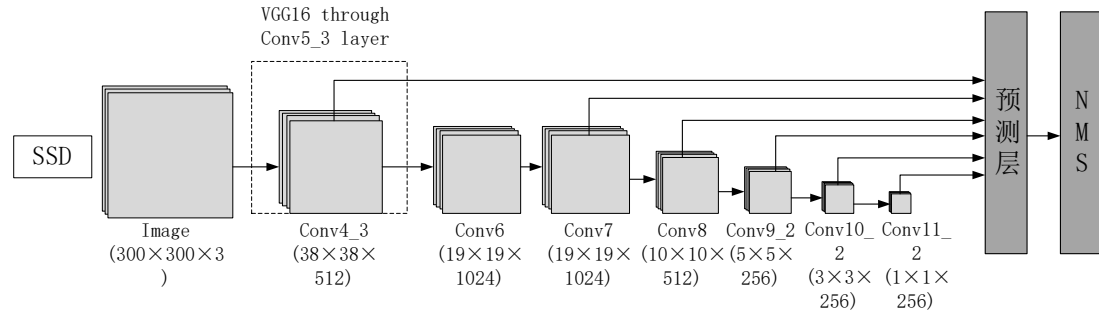


图 3.3 SSD 算法结构图

SSD 目标检测算法虽然在通用 VOC 数据集上取得了不错的检测效果, 但是该检测精度还是难以应用于工业分拣环境。同时, 由于待分拣物体中均为不具备纹理信息的金属零件, 特别是小螺钉这样的微小金属件, SSD 目标检测算法难以实现稳定的检测效果。总的来讲, 这是由于 SSD 的主干卷积神经网络的整个结构是金字塔形状, 浅层特征图的尺寸大, 降采样量少, 所包含的像素信息较丰富, 这比较适合小目标物体的检测。由于浅层特征图所包含的语义信息较少, 直接将其用于目标物体检测也难以取得较好效果, 如果将浅层特征图和深层特征图的信息进行融合后形成新的特征图, 对于小目标物体的检测将会有利, 这是因为深层特征图包含较多的语义信息。所以, 将特征金字塔的思想引入这部分优化能够实现该目的, 特征金字塔的网络结构, 能在增加极小计算量的情况下, 处理好物体检测中的多尺度变化问题。

综上所述, 为提高现有框架在工业零件抓取应用中的短板并尽可能的提高其检测速度, 本文综合应用以下 4 种技巧提高现有框架的性能。

(1) 为降低卷积运算的参数量, 将内卷积作为基本运算符替代 SSD 目标检测算法中原有的卷积运算, 实现主干特征提取网络的重构。

(2) 为缓解小目标零件在浅层检测时语义信息不足而深层检测时特征信息丢失严重的矛盾, 本文采用双线性插值将高层特征融合于低层次特征后再作为有效特征层进行位置回归和分类。

(3) 为轻量化目标检测框架并提高运算速度, 在主干特征提取网络引入 Focus 模块并剪裁掉后三层卷积块。

(4) 为保证先验框更加契合本项目的实际数据集, 采用 K-means 聚类数据集, 然后将聚类结果作为依据重新设计先验框的分布。

优化后的目标检测模型网络结构如图 3.4 所示。

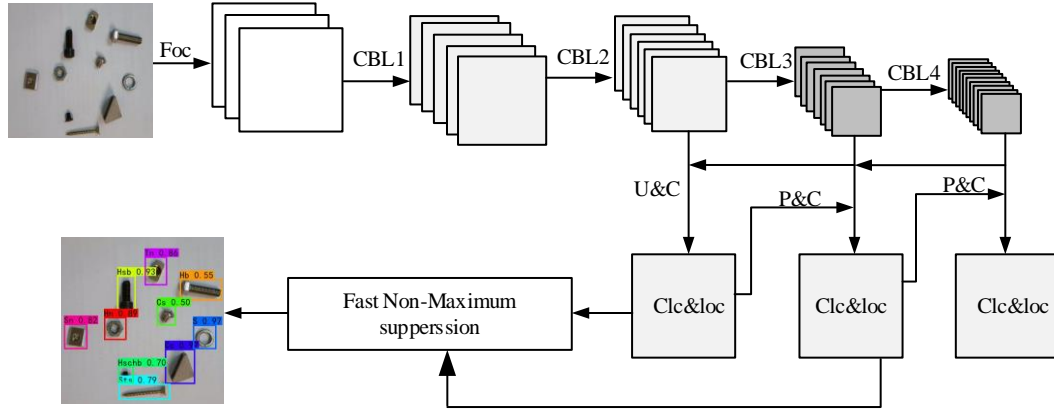


图 3.4 目标检测模型网络结构

图 3.5 中 Foc 代表 Focus 模块, CBL 代表有内卷积—批标准化—relu 激活所构成的计算块, U&C 代表上采样和张量沿通道方向上的拼接, P&C 代表下采样和张量沿通道方向上的拼接, Clc&loc 代表目标物体分类运算和目标物体包围框回归预测。Focus 模块置于输入张量和卷积块之间, 其目的是在不丢失输入张量信息的前提下实现下采样^[46-47]。Focus 模块的计算过程可以描述如下: 首先对输入三通道张量进行切片采样, 其采样规则采用间隔采样, 即每隔一个像素进行一次采样。用这样的方式采样出四个小的张量。然后再将采样后的张量沿通道方向上拼接, 拼接起来的图片相对于原先的 RGB 三通道模式变成了 12 个通道, 即最终得到了没有信息丢失情况下的二倍下采样特征图。该模块的计算过程描述如图 3.5 所示。

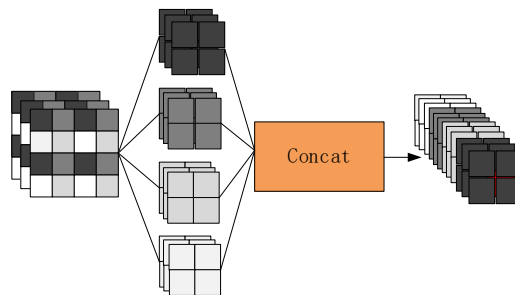


图 3.5 Focus 模块计算图

原版 SSD 目标检测算法的主干特征提取网络采用的是 VGG16 网络，该网络存在参数量大且特征提取能力不强的缺点。为实现增效降参的目的，采用内卷积替换掉原版中标准卷积运算子并采用特征金字塔的设计思想，实现多层次特征图融合。主干特征提取网络的结构如图 3.6 所示。

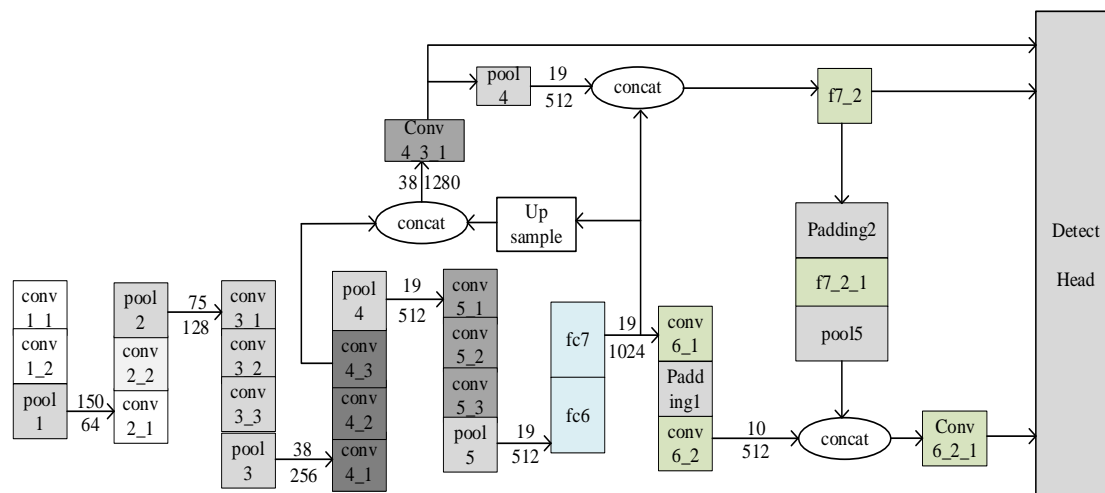


图 3.6 主干特征提取网络图

3.3.2 内卷积计算子

由 Duo Li 等人共同完成的文献^[48]中提出了一种新的卷积计算子，其结构具有通用性，能够作为一种新的神经算子而嵌入到其他的神经网络算法中。该新算子相较于标准卷积运算更加轻量化，所引入的可训练参数更少。相较于现有的自注意力机制，形式上更加简洁，能够用统一的视角去理解自注意力操作。

内卷积（Involution）是对现有标准卷积运算的重构，内卷积与标准卷积所不同的是：它实现了卷积核在通道上的共享，即实现了通道不变性和空间特异性。采用该计算子可以在保证特征提取性能的同时大幅减少卷积神经网络的参数冗余，进而实现目标检测框架轻量化。内卷积可用公式 3.1 表示：

$$Y_{i,j,k} = \sum_{(u,v) \in \Delta k} H_{i,j,u+[k/2],v+[k/2],[kG/c]} \times X_{i+u,j+v,k} \quad (3.1)$$

$$\Delta k = [-[k/2], \dots, [k/2]] \times [-[k/2], \dots, [k/2]] \quad (3.2)$$

式中 H 为内卷积计算子的卷积核， X 为输入张量， k 代表后续 involution 卷积核大小， G 代表 involution 操作的分组数， C 为特征图通道数。

内卷积运算的卷积核并非固定，而是依赖于输入张量。内卷积计算子的卷积核生成公式为：

$$H_{i,j} = W_1 \times \beta(W_0 \times X_{i,j}) \quad (3.3)$$

式中 w_0 、 w_1 代表对输入张量的一系列线性变换，这里采用 1×1 的卷积核实现对输入通道的缩放。函数 β 为批标准化和非线性激活。

内卷积的运算过程如图 3.7 所示。

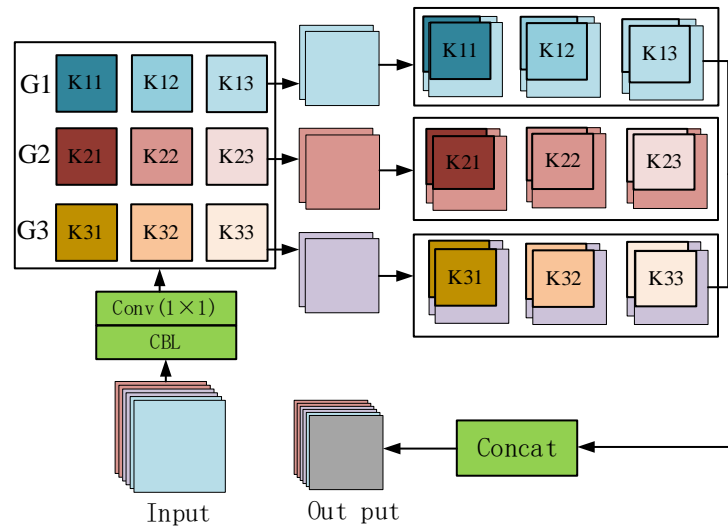


图 3.7 内卷积计算图

简单来讲，内卷计算子的运算过程就是针对特征图上的一个坐标点及其邻域所构成的特征向量做生成卷积核运算，从而得到该坐标点上所对应的卷积核。然后再将输入特征图中该点的邻域特征向量与得到的卷积核做乘加运算，最终可以得到内卷积运算的输出特征图。当然，在实际使用中，以上操作要扩展到整个输入特征图上。

为评价内卷积计算子的实际使用性能和应用特点，将内卷积计算子引入到目前最常用的残差神经网络中并将其应用于具有挑战性的图像分类中。为保证实验的可靠性，采用开源的残差神经网络结构并都采用 Tensorflow2 来搭建标准卷积计算子和内卷积计算子。在训练阶段，均不引入迁移学习，即均从最原始的条件下开始进行训练，并采用同样的数据增强手段。由标准卷积计算子搭建的深度残差神经网络和由内卷积计算子搭建具有同样网络结构的深度残差神经网络，在

同样数据集上实现图像分类任务，二者所取得的分类效果以及网络参数的汇总见表 3.1。

表 3.1 网络参数对比表

网络结构		权重文件大小 (M)	浮点运算次数 (G)	最好分类准确率(%)
标准卷积	ResNet-26	13.7	2.4	73.6
内卷积	In-ResNet-26	9.2	1.7	76
标准卷积	ResNet-38	19.6	3.2	76
内卷积	In-ResNet-38	12.4	2.2	77.6
标准卷积	ResNet-50	25.6	4.1	76.8
内卷积	In-ResNet-50	15.5	2.7	78.4
标准卷积	ResNet-101	44.6	7.9	78.5
内卷积	In-ResNet-101	25.6	4.7	79.1

通过表 3.1 可以看出，在不同深度的残差网络结构中，采用内卷积替换标准卷积后，在保证检测精度优于采用标准卷积的残差网络的前提下，内卷积残差网络具有更少的参数量和更低的计算开销。

3.3.3 特征金字塔设计

由 SSD 算法结构图可以知道，该目标检测算法从浅层到深层一共选择了六个特征层作为检测头的输入数据。该算法之所以这样设计是为了缓解小目标在深层次特征图上信息丢失的问题。换句话说就是，原版 SSD 算法的作者希望在浅层特征图上去检测小目标，在高层特征图上去检测大目标，尽量做到分而治之。这一结论可以通过分析 SSD 算法候选框的分布可以知道。

SSD 算法候选框在上述六个检测头的部署是不一样的，每个候选框的尺寸可用公式 3.4 计算。

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1} (k-1), k \in [1, m] \quad (3.4)$$

公式 3.4 中， m 代表检测头对应特征图编号， s_k 代表先验框相对于输入图片的比例， s_{\max} 和 s_{\min} 分别代表比例的最大值和比例的最小值，一般取 0.2 和 0.9。可以看出，分布在深层特征图所对应检测头上的候选框，所占据整个像素的比例较大，而浅层特征图所对应检测头上的先验框比较小。该设计虽然可以利用多尺

度特征图上的信息，但是为了避免使用浅层的特征图，该算法的第一个检测头，采用第四个卷积块的第三层卷积所输出的特征图作为输入。而为了满足后面的检测头具有足够的语义信息，其又在 VGG16 网络基础上又添加了几个卷积块。该网络结构和候选框的设计在该平面态散乱小零件分拣系统的应用存在以下两点不足：

(1) 按照公式 3.1 和原版 SSD 算法的输入图像尺寸，可以计算得到候选框的尺寸分布如表 3.2 所示。

表 3.2 候选框的尺寸分布

	最小尺寸	最大尺寸
Conv4_3	30	60
Conv7(fc7)	60	111
Conv8_2	111	162
Conv9_2	162	213
Conv10_2	213	264
Conv11_2	264	315

从表中可以看出，Conv9_2、Conv10_2、Conv11_2 这三个位于深层的检测头上分布的先验框非常大，而本文所研究小零件分拣系统中所待识别零件都比较小，不会存在那么大的待检测目标物体。所以为了提高算法的检测效率，可以将后三个位于深层的检测头剪裁掉。这可以为引入图像特征金字塔提供一定的算力支撑。

(2) 算法中引入的六个检测头虽然处于不同的深度，但深层的卷积块没有和浅层的卷积块共享信息。如果将深层的信息重新引入到浅层的检测头中，将在一定程度上弥补浅层特征图语义信息不足的缺陷，进而可以尽最大可能的提高图像分类检测的置信度评分。

为提高原版 SSD 算法在本应用场景下的适应度，可以引入图像特征金字塔。由于小目标物体在经过多次卷积和下采样后会丢失其信息，特别是对本数据集中的小沉头螺钉，经过多次下采样后仅剩几个像素。很自然的想法是，将浅层的特征作为有效特征图进行检测，但是浅层特征图的语义信息不足，难以满足检测要求。为缓解以上矛盾，设计了如下计算方法：

$$Y_i = \begin{cases} B[C(X_i + U(X_{i+1}))] & i = 1 \\ B[C(X_i + P(X_{i-1}))] & 1 < i \leq 3 \end{cases} \quad (3.5)$$

式中 X_i 为有效特征层，从浅层到高层一共将主干网络的三处节点作为有效特征层。其中，函数 U 为双线性插值运算，函数 P 为最大池化运算。以上两个函数其目的是调节输出特征图大小使得两层特征图可进行张量拼接。函数 C 为卷积核大小为 1×1 的卷积运算。函数 B 为类别预测计算和包围框回归运算，均采用标准卷积运算实现。

3.3.4 先验框分布重设计

由于本分拣系统所采用的基本算法框架是基于先验框的目标检测算法，在该类算法中，先验框设置的合理与否将直接影响到算法性能。所述先验框就是在待检测图像上预先设定的一系列矩形框，先验框分布于整个图像中，且具备多尺寸、多长宽比的特点。先验框在图像中某一个位置上分布的矩形框如图 3.8 所示。

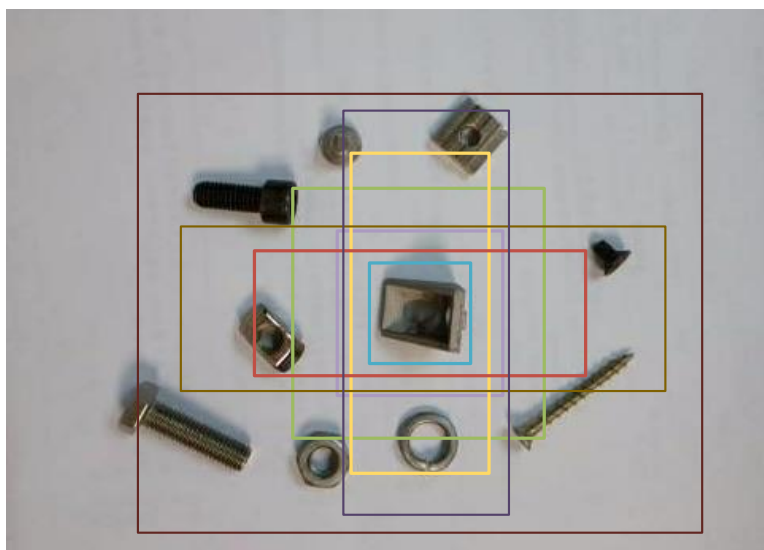


图 3.8 先验框在图像中某个位置的分布示例

以图 3.8 为例，图中有多个待检测的物体，其中在角码的位置上设置了 5 个矩形框，该五个矩形框的中心点正好和角码所在的像素中心点是重合的，以上五个先验框中都包含了角码，但是蓝色的正方形框包围的是最贴和的，即其在完全包含的基础上做到了背景信息量最低。在实际算法中，图中的先验框会非常密集的分布于整张图像。

在实际搭建算法网络时，上述矩形框并不直接布置于待检测图像上，而是在算法的主干网络中，选择由浅到深的几个层级的特征图，分别独立布置上去。上述几个层级特征图是和检测头一一对应的。也即是在多层卷积后的特征图上设置先验框，但设置的位置和尺寸是以输入图像为基准，也即是能一一映射回原图像上。

在算法的训练阶段，样本集中用于训练的图像中手工标注出来的置信框是需要和上述的先验框绑定的，也即需要将置信框赋予到预先设置好的具有固定位置、固定尺寸、固定长宽比的先验框上。然后再传入到训练网络中进行训练。

如图 3.9 所示，每个像素中心点部署了四个不同的先验框，图 a 中猫和狗分别匹配到了与其最匹配的先验框，也即将该先验框用于对猫和狗的识别训练中。如果在每个特征图上每个像素点中心部署 k 个先验框，那么对于每个先验框都需要预测 c (包括背景) 个类别置信度评分和 4 个预测框与先验框的坐标误差值。因此，如果一个特征图的大小是 $m \times n$ ，那么这个特征图上就一共有 $K \times m \times n$ 个先验框，每个先验框需要预测 4 个坐标相关值和 C (包括背景) 个类别置信度评分，则在 $m \times n$ 的特征图上面将会产生 $(c+4) \times k \times m \times n$ 个输出。这些输出个数的含义可以理解为：采用尺寸为 3×3 的卷积核，对该层特征图进行卷积时，卷积核的个数应当包含两部分，一部分是数量为 $c \times k$ 的卷积核，该部分卷积计算的结果对应类别置信度评分，也即是表示每个先验框包含某个类别物体的置信度，也可以理解每个先验框包含某个类别物体的概率，概率越高则认为该先验框中存在该物体。其中，数量为 $4 \times k$ 的卷积核对应预测框和先验框坐标误差值的输出，表示每个先验框要想最优包含该物体时，应当做的位置和尺寸上的调整。原版 SSD 目标检测算法的作者在实验中也发现了，先验框设计的个数和尺寸变化数量越多，对于目标物体的检测效果也就越好。如图 b 中的几个虚线框，表示不同尺度的先验框。但是，过多的先验框的设置也即意味着需要开支更多的算力和内存空间。

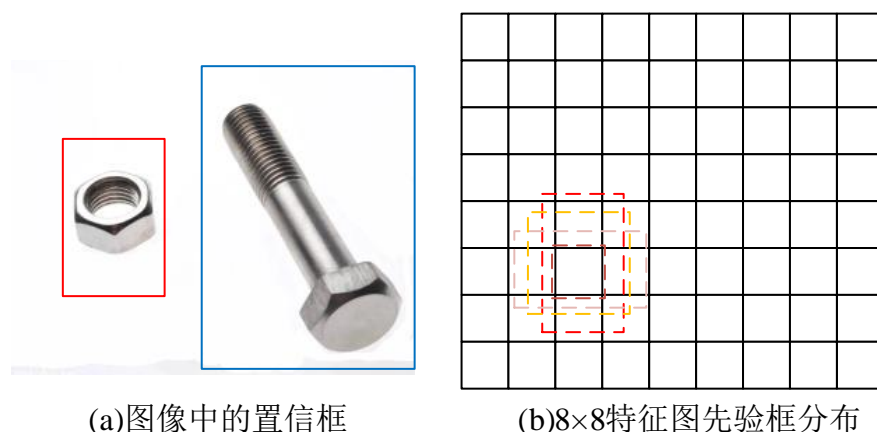


图 3.9 先验框布置

以上分析可以知道，对于基于先验框的目标检测算法，其先验框的布置非常重要，布置的原则应当是在先验框数量尽可能少的前提下对待检测数据集做到全覆盖。采用聚类的方式先对数据集做聚类，再根据聚类的结果去设置先验框是一种较为理想的先验框布置方式，因为这样能尽最大可能保证设置的先验框贴合本数据集。聚类的方法可以采用无监督的 K -means^[49-50] 聚类。 K -means 聚类目的是将数据集划分成 K 个簇，使得相同簇之内的数据相似性高，不同簇之间的数据相似性低。该算法的计算步骤描述如下：

(1) 初始化 K 个簇中心；即在数据集中随机挑选 K 个值，作为聚类的簇中心点；

(2) 相似度计算；即首先定义一种评价方法，该方法要能够评价两个数据之间的相似度，最为常用的是欧式距离。通过计算待学习数据集中每一个数据与初始化簇中心点之间的相似度值，即可将每个值分别分配给其对应的相似度值最大的簇中心，然后将其作为一个类别；

(3) 计算每个类别的中心点，将每个类别的中心点作为新的簇中心点；

(4) 重复步骤 2 和步骤 3，不断的更新簇中心点，直到簇中心点不在做较大的变化即停止更新；

(5) 返回停止更新时的簇中心即为聚类的最终结果；

K -means 聚类算法的计算示意如图 3.10 所示。

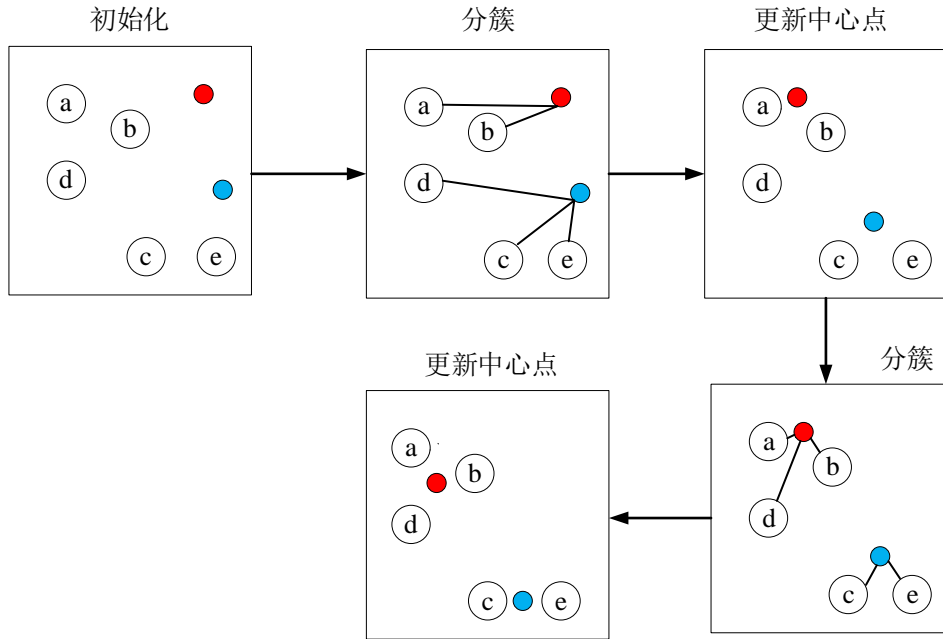


图 3.10 K-means 聚类算法计算示意图

根据 K-means 聚类算法可知，在进行聚类前应当先搜集数据集以及根据数据的性质来定义评价两个数据之间距离的方法。通常，先验框是由矩形框左上角顶点坐标和右下角顶点坐标表示的，其形式是 $[x1, y1, x2, y2]$ 。所以，这里可以采用先验框的宽度和高度作为数据集，并且由于数据集中图片的大小可能不同，还需要先使用图片的宽和高对先验框宽度和高度做归一化。归一化的方法采用如下公式：

$$w = \frac{w_{box}}{w_{img}}, h = \frac{h_{box}}{h_{img}} \quad (3.5)$$

由于待训练样本是先验框宽度和高度数据，所以直接采用欧式距离将难以客观的评价两个先验框的相似程度，所以，这里引入先验框的交并比（IOU）作为两个先验框相似程度的评价标准。交并比的计算方法如图 3.11 所示。图中点 A 和点 B 为红色矩形框和蓝色矩形框相交矩形的两个顶点，交并比（IOU）的计算即为上述两个矩形框交集与并集的面积比。

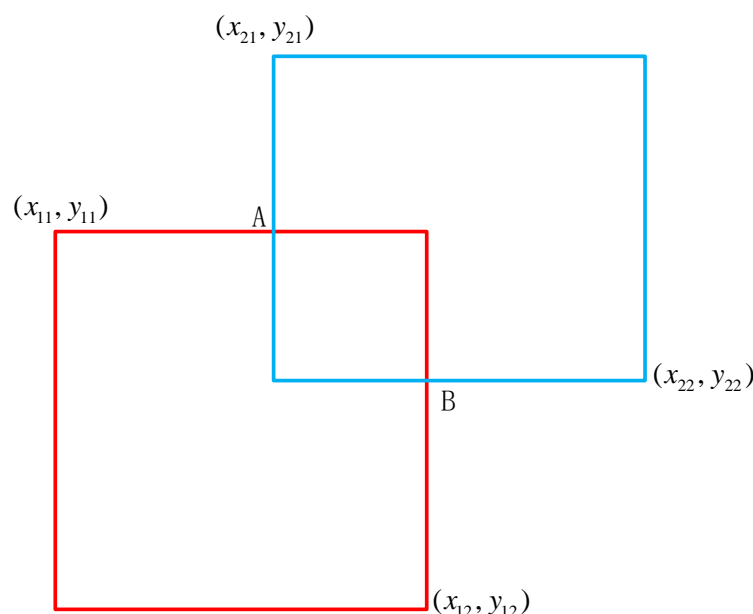


图 3.11 交并比计算示意图

基于交并比的相似度采用如下公式计算：

$$D = 1 - IOU \quad (3.6)$$

采用基于交并比作为相似度度量的先验框聚类方法采用如下步骤完成：

(1) 数据收集：将数据集中所有置信框的坐标提取出来，一般情况下一张图中有多个物体类别的多个矩形框，但是这里仅需要无区别的将所有图片的所有矩形框提取出来，放在一起即可。

(2) 数据处理获得所有训练数据置信框的宽高数据；标注好的训练数据往往是其置信框的 4 个坐标，但是需要聚类处理的是先验框的宽高大小。所以需要 将坐标数据转换为框的宽高大小，计算方法采用如下形式：长=右下角横坐标-左上角横坐标、宽=右下角纵坐标-左上角纵坐标。

(3) 初始化 k 个先验框；通过在所有的先验框中随机选取 k 个值作为先验框的初始值。

(4) 数据归类；计算每个置信框与每个先验框的相似度差值，将每个数据归类到对应的类簇中；

(5) 中心点更新；分别计算每个类簇中矩形框数据，求得这些矩形框宽高的中值，将计算得到的中值作为该先验框的新尺寸。

(6) 重复第 4 和第 5 个步骤, 直到先验框的尺寸不再发生比较大的改变就停止迭代, 将该值作为最后聚类结果。

采用聚类的方法在三个检测头上布置的先验框的数据汇总见表 3.3。

表 3.3 候选框的尺寸分布

卷积层名	特征图尺寸	先验框尺寸	单个像素布置先验框个数	该层总个数
Conv4_3	38×38	min:20 max:50	4	5776
Fc7	19×19	min:60 max:90	6	2166
Conv6_2	10×10	min:100 max:130	6	600
总先验框个数				8542

3.4 机械手抓取姿态预测算法设计

3.4.1 细长类零件深度估计算法

由于细长类零件螺栓在工作台上可能存在竖直的状态, 该姿态和平躺姿态存在较大的深度差异, 所以针对该情况需要设计对应的算法去估计该类零件的深度, 进而引导机械手进行抓取深度的调整。由于该类零件在竖直姿态和平躺姿态在图像上存在较大差异, 并且由于深度姿态空间是离散的真值姿态空间, 所以可将姿态估计问题转化为图像分类问题, 设计如图 3.12 所示的卷积神经网络对工作台上的零件进行姿态预测。

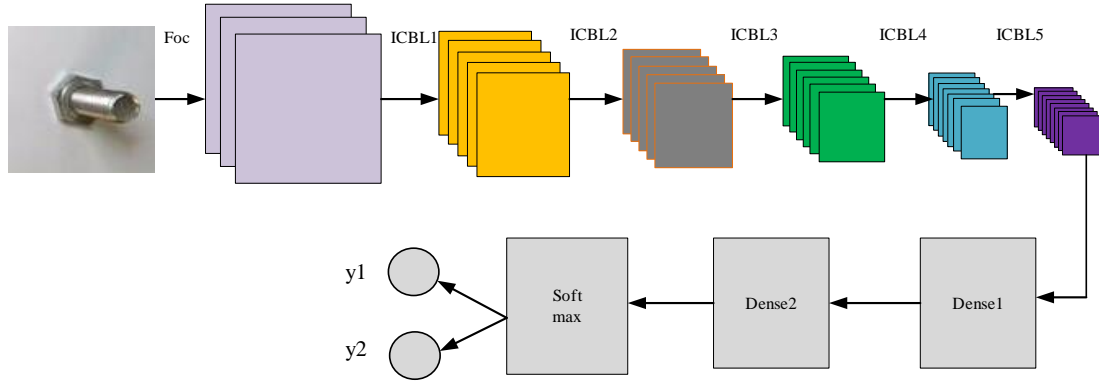


图 3.12 标准件姿态预测网络

该深度姿态预测神经网络采用在内卷积运算子和激活函数之间布置批归一化，以保证每个卷积块的运算结果处于激活函数的敏感区，避免由于激活函数饱和而出现神经网络模型在训练过程中难以收敛^[51-53]。

$$\begin{cases} \mu_c^d = \frac{1}{m^d} \sum_{i=1}^{m^d} X_c^d(i) \\ \sigma_c^d = \frac{1}{m^d} \sum_{i=1}^{m^d} (X_c^d(i) - \mu_c^d)^2 \end{cases} \quad (3.6)$$

$$\begin{cases} \hat{X}_c^d(i) = \frac{X_c^d(i) - \mu_c^d}{\sqrt{(\sigma_c^d)^2 + \varepsilon_c^d}} \\ Y_c^d(i) = \gamma_c^d(i) \hat{X}_c^d(i) + \beta_c^d \end{cases} \quad (3.7)$$

公式 3.6 中 μ_c^d 、 σ_c^d 分别为第 d 层卷积的计算结果中的第 c 个批量特征图的均值和方差，m 为该批样本的个数。公式 3.7 中 \hat{X}_c^d 为批量样本数据中第 i 个样本的第 d 个卷积层的第 c 个特征图张量。 Y_c^d 为特征图张量所对应的标准化后的结果。

为保证该姿态预测神经网络的训练效果，采用交叉熵损失函数作为该神经网络训练的损失函数。由于该标准件数据集中零件的姿态可以离散的分 2 种，将这 2 种姿态进行独热编码后即可引入交叉熵损失。

$$Loss = \sum_{i=1}^k l_i \times \ln(y_i) \quad (3.8)$$

式中 k 为分类的种类数， l_i 为样本独热编码标签， y_i 为神经网络前向传播运算结果。

3.4.2 细长类零件姿态角回归算法

由于目标检测算法仅能够识别目标物体类别和回归最小包围框，针对抓取的工况要求，目标物体相对于机械手的角度对抓取成功率的影响也是巨大的。特别是针对细长类零件，机械手的抓取角度异常重要。以细长类零件螺栓为例，如果采用两指式的机械爪进行抓取，机械爪抓取平面和细长类零件的轴线平行时才最有利于抓取。

该算法的设计采用传统的图像处理方法即可实现，其核心思想是对灰度化后的螺栓图像做固定阈值分割，在分割图像寻找闭合区域的最小包围矩形，根据最小包围矩形姿态计算螺栓姿态。该算法的详细计算流程如图 3.13 所示。

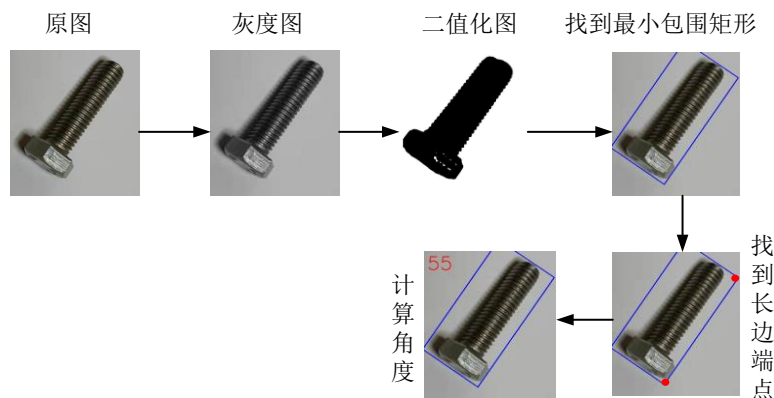


图 3.13 姿态回归算法执行流程

姿态角回归算法采用 OpenCV 开源图像计算库实现图 3.13 中各个操作，OpenCV 中内置的最小包围矩形寻找算法，返回最小包围矩形结果中仅有描述该矩形的四个顶点的坐标，且坐标顺序是杂乱无章的。为准确找到能够描述零件姿态角度的坐标点，算法先利用该坐标点生成一组用于描述所有两点之间距离的向量，利用矩阵长边长度值大于短边长度，且短于对角线长度的特点来确定用于描述姿态角的两个坐标点。通过计算两点间连线角度即可得到细长类零件姿态角。

3.5 本章小结

本章主要分析并设计了该分拣系统所用识别算法，包括目标检测算法优化设计、细长类零件抓取姿态识别算法设计。本章首先给出算法整体架构，即分析各个算法模块在整个系统中发挥的作用和其配合关系。在此基础上，分别单独给

出各个子算法模块组成原理以及实现方法。目标物体判断算法采用基于差异值的感知哈希算法实现，目标检测算法在 SSD 算法框架基础上做进一步优化，采用了内卷积计算子提高运算效率，采用特征金字塔和检测头重构等手段进一步提高算法效率和对小目标检测效果。对于细长类零件抓取姿态问题，采用细长类零件深度估计算法和姿态角回归算法相互配合解决。其中深度估计问题本质上是判定零件在工作台上平躺状态还是竖直状态，所以将该问题转化为图像分类问题，采用深度卷积神经网络解决该问题。对于细长类零件在工作台平面上角度预测问题，采用传统图像处理方法即可高效迅速解决该问题。

4 算法训练与部署

算法的设计是包含算法结构设计以及算法训练两大部分的，在实际操作中，二者相互关联，并且往往是一个循环往复的过程，即一个能够满足工程应用的算法往往需要经过很多次重复的训练和修改的。所以，良好的算法训练策略也是非常重要的，它将直接决定算法是否能够发挥最大潜力。

算法完成训练后并不能应用于实际的工业应用场景，算法训练过程是网络不断对训练数据集进行学习的过程。训练包括前向传递和后向传播两个阶段，前向传递用于预测标签，然后再通过预测标签与真实标签之间的误差进行后向传播不断修改网络的权重，直到网络权重能够使得算法取得好的推理效果或者达到预先设定好的迭代次数。所以，在训练阶段更加注重的是如何找到最优的网络权重，也即是如何训练出最好的算法。而在部署应用阶段，算法的权重参数不会变化，仅做前向传播，所以该阶段注重算法的计算效率。所以，本章着重分析算法的训练和部署过程。

4.1 数据集制作

由于本分拣系统所采用的算法都是基于卷积神经网络的，都属于有监督学习，有监督的算法在训练阶段需要大量有标注样本作为训练的评价标准。从训练过程中看就是通过标注好的样本和算法前向传播结果，根据定义好的损失函数进行损失结果的计算，然后计算损失函数各个节点的梯度，最后再采用合适的梯度下降算法对神经网络的各个参数进行一个微小的更新。重复上述过程，直到损失函数的结果达到最小即得到网络的最优解。所以，在进行算法训练前，需要根据该分拣系统的目标物体进行数据采集和数据标注。

4.1.1 目标检测数据集

首先在工作台上方部署好工业相机，然后在摄像机下方的工作台上随机将目标物体放置。通过上述过程，采集目标物体在工作台上各个位置，各个姿态下的图像。采集到的图像样张如图 4.1 所示，数据集信息汇总见表 4.1。

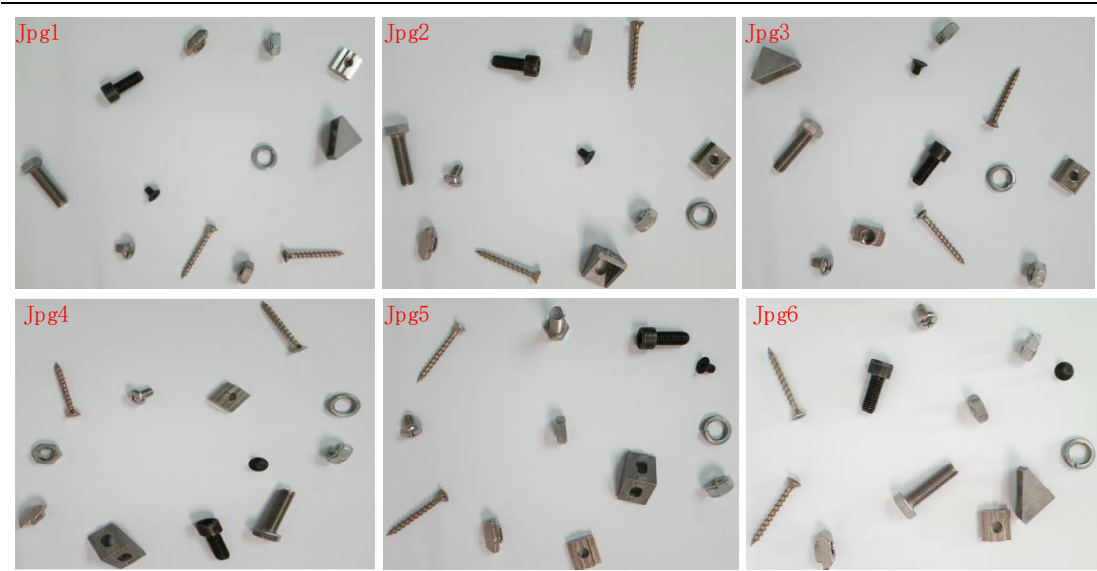





图 4.1 图像数据集样张

表 4.1 数据集信息汇总

编号	零件名称	标记名称	样例	采样数量
1	螺母	Hn		900
2	螺栓	Hb		600
3	内六角螺栓	Hsb		600
4	自攻螺钉	Sts		900
5	沉头螺钉	Hschb		600
6	圆头螺钉	Cs		600
7	弹簧垫片	S		600

8	滑块螺母	Tn		600
9	角码	Cc		600
10	T 型滑块螺母	Sn		600

目标检测算法在训练过程中是通过对输入图片的像素信息进行学习的，学习的特征包括目标物体的类别信息和位置信息，所以在数据集标注的时候要包含目标物体的类别和物体的最小包围框。标注的工具采用开源 LabelImg 软件，标注的过程如图 4.2 所示。





图 4.2 数据集标注

4.1.2 图像分类数据集

由于只有螺栓和内六角螺栓存在竖立起来而造成深度差异的可能，所以图像分类数据集包含这两个类别即可。对于图像分类算法，在算法训练阶段，引导算法学习的信息仅有图像所属的类别标签，所以该数据集只需给出标签信息即可。对于标签的分类，从工程应用角度来看，算法只要能区分出来待抓取的物品是处于平躺还是竖直的状态即可，所以数据的分类仅也包含平躺零件样张和竖

直零件样张。在实际操作中，只要将上述两类图像放置于不同文件夹下即可。图像分类数据集汇总见表 4.2。

表 4.2 图像分类数据集汇总表

编号	样本名称	编码	样例	采样数量
1	竖直态	0		80
2	平躺态	1		100

4.2 算法训练

4.2.1 目标检测算法训练

在目标检测算法的训练过程中，需要设计能够评估算法模型输出结果和置信结果之间误差的损失函数。由于目标检测算法需要识别物品类别和物品的位置，所以该算法的损失函数要能够评估以上两方面的误差。根据以上要求，损失函数中采用加权求和的方法来将上述二者结合起来，其中评估物品分类识别的误差采用交叉熵损失，其中评估目标物体预测位置 and 实际位置误差值的方法采用 smooth L1 公式。最终的损失函数^[54-56]计算采用公式 4.1

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c)) + \alpha L_{loc}(x, l, g) \quad (4.1)$$

公式 4.1 中：
$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{i,j}^k \times smooth_{L1}(l_i^m - \hat{g}_j^m),$$

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

其中 N 是本次训练过程中，一个批次数据中先验框匹配到置信框的个数，如果先验框没有匹配到置信框，则该值为零，对应的损失也为零。其中 α 是一个用于调节置信度损失和定位损失的权重系数，一般取为 1。公式中 l_i^m 为预测框参数， \hat{g}_j^m 为置信框参数。在置信度损失函数部分中， i 指代搜索框序号， j 指代真实框序号， p 指代类别序号， $p=0$ 时表示背景。其中， $x_{i,j}^p$ 是一个条件函数，当第 i 个先验框和第 j 个置信框匹配，且这个先验框的类别为 p 时取 1，否则取 0， c_i^p 表示第 i 个搜索框对应类别 p 的预测概率，采用 softmax 分类器计算其概率值。

利用该损失函数完成目标检测算法训练的具体流程如下：

- (1) 从数据集中读取图像信息和标签信息。
- (2) 对图像数据进行数据增强操作，包括图像随机翻转、拼接、尺寸调整等操作。
- (3) 将图像数据传给主干特征提取网络，得到检测头对应的三个特征图。
- (4) 利用先验框生成规则，在每个特征图上布置先验框。
- (5) 利用 IOU 评价指标，计算先验框和置信框之间的相似度矩阵。
- (6) 根据先验框匹配原则，将置信框和先验框进行匹配，将匹配到置信框的先验框设置为正样本，如果先验框不与任何置信框匹配，则认为该先验框为负样本。由于一张图片中没有匹配到任何置信框的先验框有非常多，而匹配到置信框的先验框一般只有十几个，这会造成样本不均衡进而导致算法学习不到有用的信息，所以应当舍弃绝大部分负例样本，使得正例样本和负例样本满足三比一的比例关系。
- (7) 根据匹配结果，为每个先验框分配目标类别，同时计算正例样本先验框和其匹配上的置信框之间的位置误差，也即神经网络需要预测的定位结果。
- (8) 检测头在对应特征图上做全卷积运算，预测每个边框和对应目标物体类别的位置调整量。
- (9) 通过随机采样和困难采样的方法，将样采样得到的正例样本和负例样本带入到损失函数中，计算分类损失和位置损失。
- (10) 计算损失函数的梯度，利用误差反向传播实现卷积神经网络的参数更新。
- (11) 重复上述过程，直到参数不再较大的变化时即停止。

训练平台采用 GTX1070 显卡，经过 500 轮的训练即可达到较为理想的识别效果。训练过程参数的变化如图 4.3 所示。

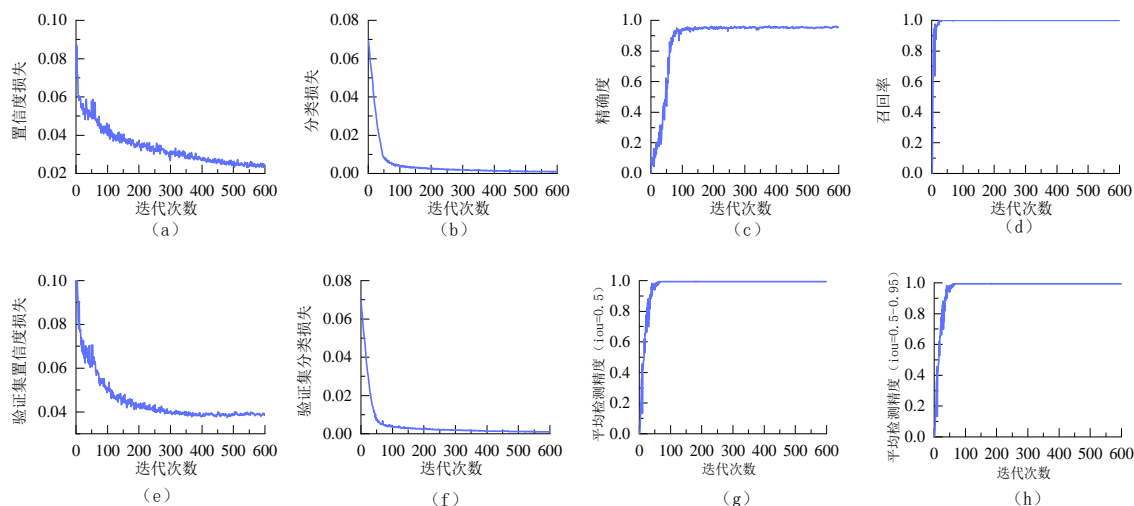


图 (a):置信度损失函数随迭代次数的变化曲线、图 (b):分类损失函数随迭代次数的变化曲线、图 (c):检测精确度随迭代次数的变化曲线、图 (d):召回率随迭代次数的变化曲线、图 (e):置信度损失函数在验证集上随迭代次数的变化曲线、图 (f):分类损失函数在验证集上随迭代次数的变化曲线、图 (g):验证集上的平均检测精度 (交并比阈值为 0.5) 随迭代次数的变化曲线、图 (h):验证集上的平均检测精度 (交并比在 0.5 至 0.95) 随迭代次数的变化曲线

图 4.3 训练结果汇总

4.2.2 图像分类算法训练

图像分类算法的训练，只需要传入算法中图像数据和对应图像分类信息即可。为提高算法训练的效果，采用图像增强算法首先对图像进行扩充。采用的方法是通过仿射变换来增加图像多样性。即用各种类型的仿射变换矩阵对训练图像进行平移、缩放、旋转、剪切等操作，增加训练样本的数量以及多样性，能提高模型通用性。部分样本仿射变换效果如图 4.4。



图 4.4 图像增强中用到的仿射变换效果

在分类网络的训练过程中，采用遵循高斯正态分布随机数作为初始化参数，卷积层采用 Relu 的激活函数，采用随机动量梯度下降优化算法并将初始学习率设置为 0.001。卷积层后连接全连接神经网络，采用二分类的 SoftMax 分类器输出最后的分类预测概率值。以上训练过程在 GTX 1070 显卡上进行 150 迭代，训练的损失函数和验证集准确率变化曲线如图 4.5 所示。根据训练结果可知，当迭代次数达到 135 次左右时，验证集识别精度已经可以达到 100%。

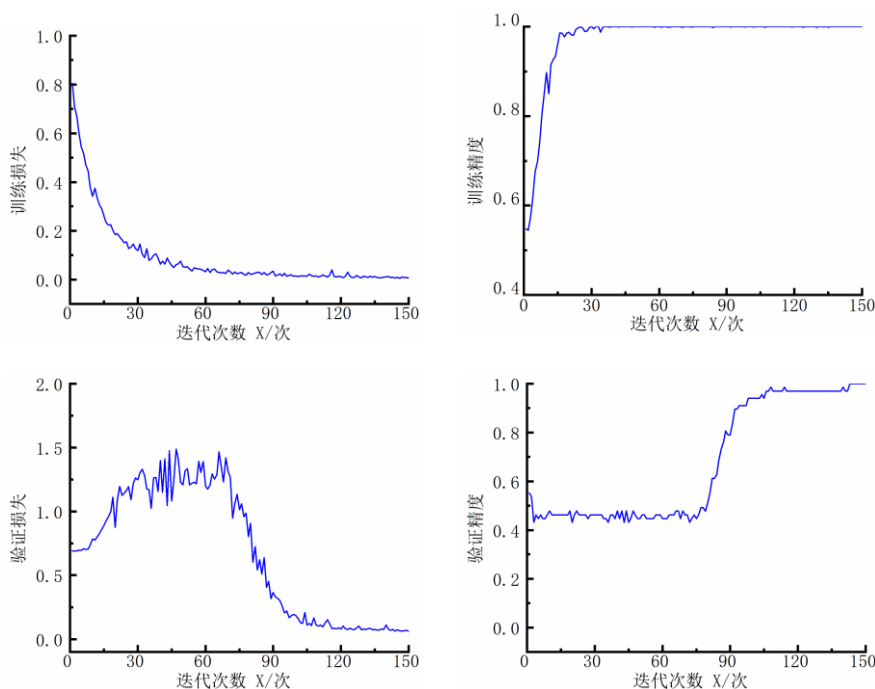


图 4.5 训练结果汇总

4.3 Tensorrt 加速部署

在算法的训练阶段，注重的是如何取得最优的权重参数使得算法取得最好的效果，所以在训练阶段需要灵活、通用的框架完成算法模型的搭建，以便于在训练阶段能够简单、高效地完成对算法框架结构的调整。算法训练阶段所采用的计算平台往往性能较为宽裕，而实际部署阶段考虑到成本问题和能耗问题，往往需要采用边缘计算设备或者嵌入式计算设备，而现有开源的深度学习框架都是基于 Python 语言的，由于该语言是交互式的语言，所以代码执行效率上难以满足速度要求。并且，由于训练时的显卡对于浮点运算的精度较高，这对边缘计算平台或者嵌入式平台也是不友好的^[57]。为解决以上问题，英伟达公司推出了 TensorRT 框架，它是一个高性能的深度学习推理（Inference）优化器，可以支持用 C++ 编程语言为深度学习应用提供低延迟、高吞吐率的部署推理。

TensorRT 框架优化神经网络的前向传播过程主要采用卷积层层间融合或张量融合以及降低浮点运算精度^[58-60]。层间融合就是将卷积计算层进行纵向或者横向合并，使得卷积层的数量大大降低，这可以在显卡计算中大大降低占用的 CUDA 核心数，能够更高效的利用显卡的计算资源。深度学习框架在训练神经网络时，一般采用 32 位浮点数作为网络中张量数据，这是因为在误差反向传播时会取得更高的精度，当算法完成训练之后，在部署阶段由于只做前向的推理，所以可以适当降低数据精度，比如降为 FP16 或 INT8 的精度。更低的数据精度将会使得内存占用和延迟更低，模型体积更小。TensorRT 框架优化神经网络结构的方法如图 4.6 示。

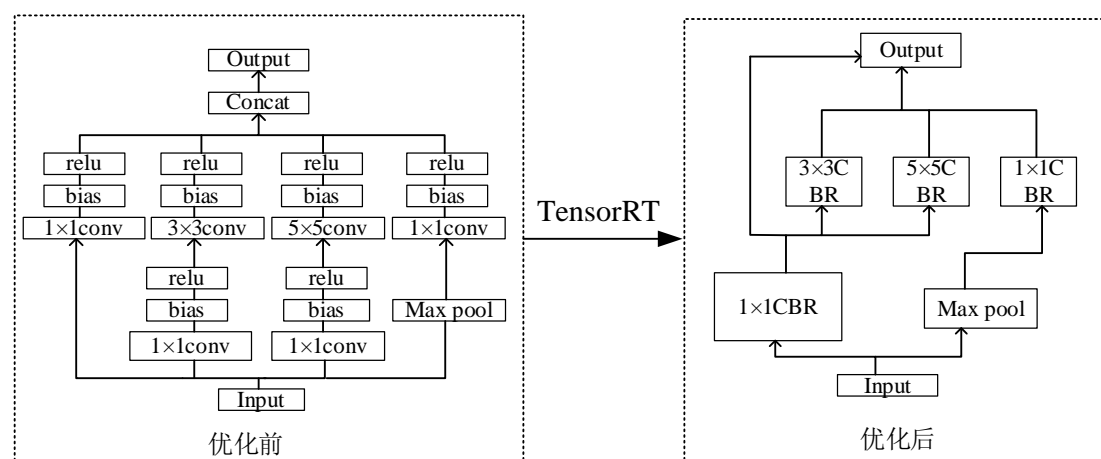


图 4.6 神经网络优化方法原理图

4.4 本章小结

本章首先介绍数据集搜集、以及标注工作，由于目标检测算法和细长类零件深度预测算法是两个独立算法流程，且二者在整个算法结构上是串联的，所以数据集的制作也是分开进行的。本章还分析了目标检测算法、细长类零件深度预测算法的训练方法与训练过程。分析了在算法训练阶段如何找到最优的网络权重，也即是如何训练出最好的算法。而在部署应用阶段，算法的权重参数不会变化，仅做前向传播，所以该阶段注重算法的计算效率，所以，本章在最后还着重分析了算法的部署过程以及利用 **TensorRT** 进行算法结构加速优化。

5 实验研究

5.1 机械臂搭建

5.1.1 机械臂硬件构成与搭建

机械臂是运动执行装置，负责在 X、Y、Z 三个方向上运动，来将末端夹取机构运送到指定的抓取位置上去。由于本抓取系统是面向于平面态的小零件抓取，为尽可能降低成本和控制复杂度，采用三自由度龙门式机械手作为运动执行装置。三轴龙门式机械臂由相互串联的三个滑台作为机械臂本体，同时作为辅助装置，机械臂还包括单片机控制器、限位开关、步进电机驱动器等。三轴龙门式机械臂机械模型如图 5.1 所示。

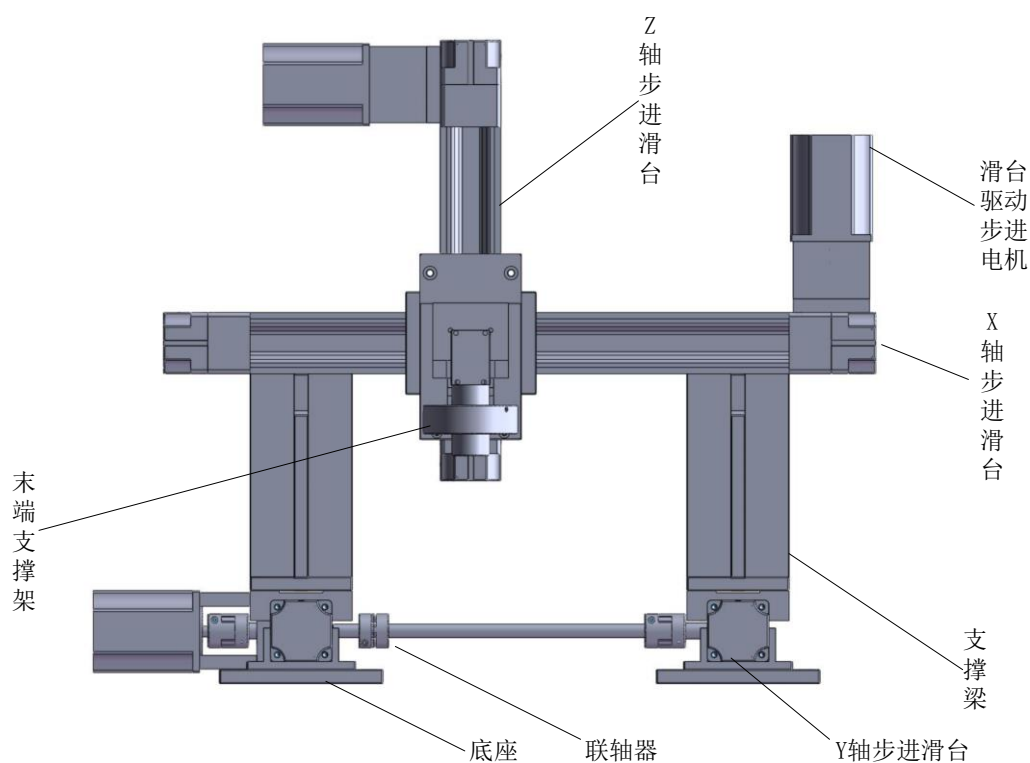


图 5.1 龙门式机械臂模型图

三轴龙门式机械臂实物如图 5.2 所示。

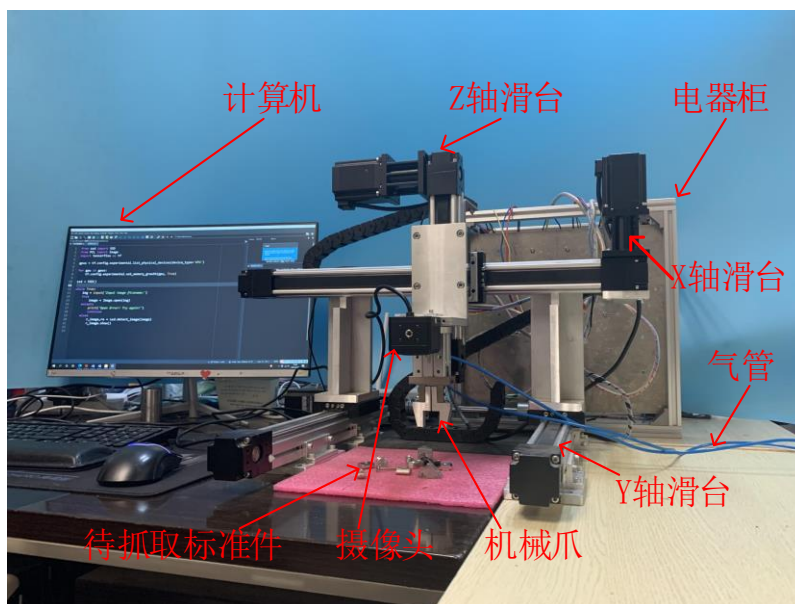


图 5.2 龙门式机械臂实物图

(1) X、Y、Z 轴滑台：该三个滑台是整个机械臂的主体，是由步进式电机配合滑动轨道而构成，能够实现较为精确的运动控制，负责执行机械手在三个自由度方向上的运动控制；

(2) 电器柜：负责支撑电源、接线端子、步进电机驱动、STM32 单片机等电器组件；

(3) 摄像头：负责采集图像；

(4) 机械爪：由气动式两指机构充当，负责夹取待分拣的物体；

(5) 计算机：负责处理数据，根据图像结果给出抓取坐标；

(6) 气管：负责驱动机械爪；

(7) 限位开关：由于驱动滑台采用步进电机，该电机属于开环控制，在机械臂驱动期间可能会丢步。为避免机械臂在运行期间存在严重积累误差，需要在机械臂运行一个周期后返回原点以进行校准，来保证机械臂连续运行精度。

5.1.2 机械臂控制系统搭建

机械臂的控制部分仅是接收上位机发送坐标指令，然后通过向步进电机驱动器发送一定数量脉冲信号来控制其按照需要动作，附加的控制也仅是机械臂回原点动作，所以该部分的控制比较简单，需要的算力很低。但是，由于脉冲信号控制需要电压为 5V 的 IO 口来实现，所以这里采用 STM32 单片机比较合适。

STM32 单片机属于微控制器，具有体积小、功耗低、价格低廉以及可靠性强等诸多优点。STM32 单片机内部集成了机械臂控制所需要的相关资源，如逻辑总线、IO 口、定时器、外部中断、串行通信协议等常用模块以及支撑单片机工作的外围接口电路，该分拣系统所采用单片机是具有 Cortex 架构的 STM32F4 系列微控制器^[61-63]。为构成完整的控制系统，机械臂控制系统还包括电源模块、步进电机驱动模块、光耦隔离模块、串口通信模块以及光电传感器。由该单片机组成的机械臂控制系统结构如图 5.3 所示。

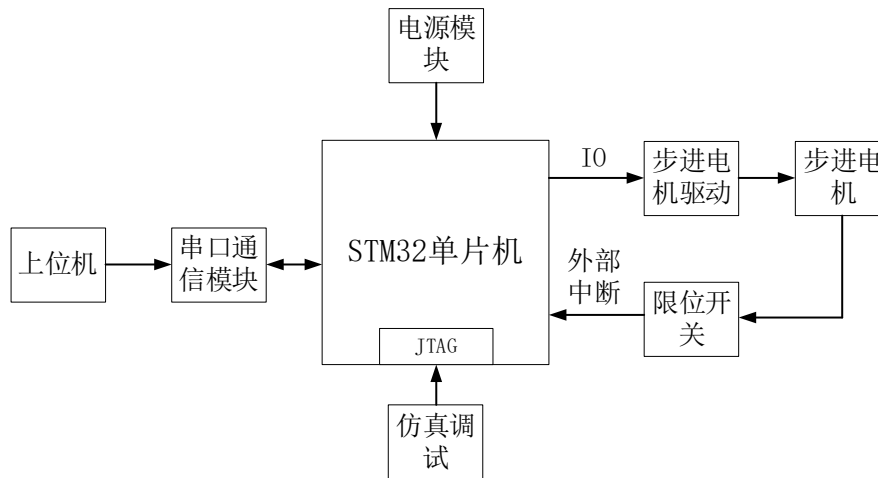


图 5.3 机械臂控制系统结构图

5.1.3 手眼标定

手眼标定的目的就是找到像素坐标系和空间机械手坐标系的坐标转化关系，以实现机械手在相机的引导下做对应动作。手眼标定方法从空间维度上可以将其划分为平面手眼标定和空间手眼标定。由于本课题是研究平面态小零件分拣，采用单目工业相机作为机械臂的“眼睛”，所以对应的手眼标定方法也为平面手眼标定。平面手眼标定最常用的解决方法即是九点标定法，即通过在工作台平面上布置一个标准板，该标准板是一个具有较好刚性的平板且其上具有九个点，再通过摄像机对标定板采样和机械手进行示教，得到标定结果。

九点标定就是为了确定像素坐标系和机械手坐标系的矩阵变换关系。在实际应用中，相机检测到目标物体在图像中像素位置后，通过标定好的坐标转换矩阵，将相机像素坐标变换到机械手的空间坐标系中，然后根据机械手坐标系，计算出各个电机该如何运动，从而控制机械手到达指定位置。九点标定法的直接作用就是计算出像素坐标与实际坐标的变换矩阵，本质上也是相机的外参。由于九

点标定法应用于平面的标定，所以待求解的变换矩阵本质上是一个平面仿射变换矩阵。

平面九点标定法可以根据仿射变换矩阵进行推导，首先约定标定板上某一点像素坐标为 (x, y) ，该点对应的空间坐标为 (x', y') 。根据相机的成像原理，可以认为标定板在相机中的图像和空间坐标系中某一平面中图像是经过仿射变换过来的。所以针对以上过程可以根据仿射变换列出 5.1 方程。

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x' \\ y' \end{bmatrix} \quad (5.1)$$

方程 5.1 中一共有 6 个未知数，为了解开这个方程，最少需要 6 组独立的方程构成方程组才能求解。所以标定板上至少要求 3 个点（因为每个点可列两个方程）。当采用大于 3 个点的时候，可以采用最小方差的方法减少误差，优化目标函数为公式 5.2。

$$\min \sum_{i=1}^n (AX_i - b) \quad (5.2)$$

实际操作过程认为相机畸变很小，可以通过标定相机内外参进行畸变矫正，对于工业相机而言，畸变矫正这一步可以忽略，标定的精度也可满足使用要求，但应当注意，标定位置和使用位置尽量靠近图像的中心。方程 5.1 的求解可以采用 OpenCV 中仿射变换方程求解函数（`estimateRigidTransform`）。

5.2 算法性能实验

5.2.1 目标检测算法实验

目标检测算法在训练集和测试集上的表现只能在训练阶段对算法进行一个初评，实际验证算法的工作性能还需要在验证集上进行验证。为验证目标检测算法的实际表现，单独采集 200 张图像作为验证集，验证集中每张图像包含所有类别目标物体，且姿态和位置随机变化。在验证集上，目标检测算法进行待抓取零件位置和姿态识别效果如图 5.4 所示。



图 5.4 目标检测算法识别效果

目标检测算法在测试集上对每个待抓取目标物体的识别精度汇总见表 5.1。

表 5.1 零件识别实验结果汇总表

实验类别	实 验 次 数	检出次数	类别预测正确次 数	包围框回归 准确次数	准确率 (%)
螺母	200	200	200	200	100
螺杆	200	200	200	200	100
内六角螺栓	200	200	200	200	100
自攻螺钉	200	199	199	196	98
沉头螺钉	200	198	198	197	98.5
圆头螺钉	200	200	200	200	100
弹簧垫片	200	200	200	200	100
滑块螺母	200	200	200	200	100
角码	200	200	200	200	100

T型滑块	200	200	200	200	100
------	-----	-----	-----	-----	-----

根据测试结果可以得到以下结论：对于长宽比不大且尺寸不过于小的物体，目标检测算法对其识别准确率可到 100%，对于类似于自攻螺钉这类细长类零件，由于其长宽比大，进而导致其最小包围矩形也相对较大，这就导致其检测框内包含的无关信息过多，一旦附近零件和其靠近的距离较近的时候，会导致其检出度较低且定位不准确。对于自攻螺钉这类零件，由于其体积比较小，这就导致其包含的有效像素信息相对较低，且在随机放置零件时容易和其他零件贴近，这就导致其难以和相邻零件区分开，在进行先验框匹配的时候也更困难，这是导致其检出误差率较高的主要原因。同时，还应当指出的是，采用单目工业相机进行数据采集时，对于相互堆叠的零件的检出率比较低，所以在实际使用时应当避免零件相互堆叠的情况出现。该情况如图 5.5 所示，图中垫片和自攻螺钉的堆叠导致自攻螺钉难以实现准确定位。



图 5.5 零件堆叠情况

5.2.2 细长类零件深度预测算法实验

细长类零件深度预测本质上是通过卷积神经网络设计的分类识别算法，来识别细长类零件是处于竖直状态还是平躺状态。首先采集一百张该零件在工作台不同位置上的竖直状态和平躺状态的图像数据，采用训练好的算法对采样的数据集做分类识别，统计其识别准确率。竖直姿态和平躺姿态的检测效果如图 5.6 和 5.7 所示。



图 5.6 竖直姿态识别效果



图 5.7 平躺姿态识别效果

根据一百张测试数据可知，该分类识别算法对于上述两种姿态的分类识别精度可达 100%。

5.2.3 细长类零件姿态角回归算法实验

细长类零件姿态角回归算法采用 150 张平躺姿态螺栓作为验证数据集，该 150 张数据包含了该零件在工作台上的各个位置、多个姿态角的图像数据。采用调试完毕的姿态角回归算法对测试数据集进行姿态角识别，根据测试数据发现，算法给出的姿态角和实际螺栓轴线的姿态角在个别样本上存在一定的差异，但整体角度误差可以控制在 20 度以内，也即是该误差完全满足抓取的精度要求，说明该算法具有应用能力。

姿态角识别出现误差的原因分析如下：根据该算法工作原理，将螺栓从背景中分离开的方法采用了固定阈值分割，由于固定阈值分割是在图像灰度图上进行，所以其对图像明暗度非常敏感，根据采样数据可以看出，测试数据集中不同样张存在较为明显的明暗度区别，这就会导致固定阈值分割难以准确的找到每一张图片的精确边界，这是导致角度预测偏差的主要原因。所以，解决该问题最简单方法是在工作台上布置合理补光光源，保证每一张样图像都具有差别不大的亮度即可解决该问题。姿态角回归算法识别效果如图 5.8 所示。

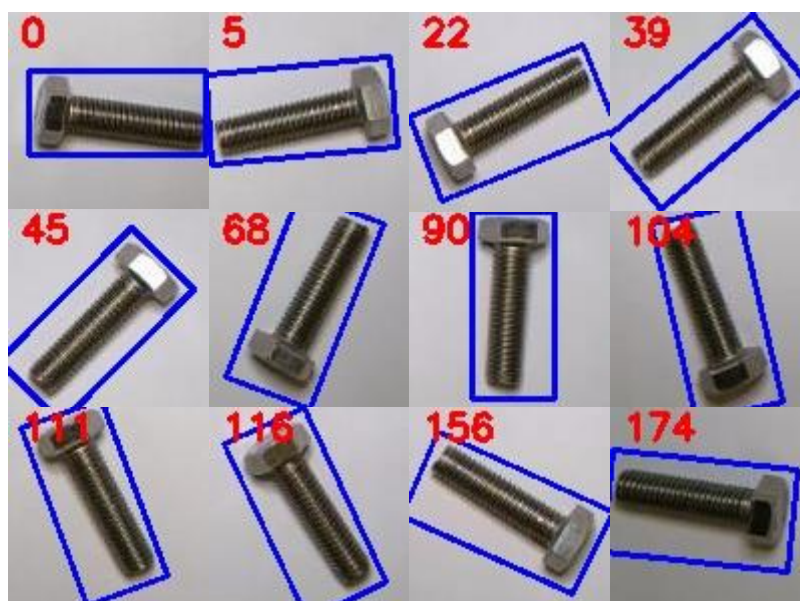


图 5.8 姿态角回归算法效果

5.3 机械臂执行实验

机械臂是末端执行装置，其运行的精度将直接影响到最终抓取效果，由于本三轴龙门式机械臂采用开环步进式驱动方式，其运动精度需要做测试才能了解其具体工作性质，进而才能更好应用该机械手。分析可知，引起机械手动作误差的主要来源是步进电机没有配备编码器，这造成无法形成完整的闭环控制进而在机械手启动时，由于存在较大加速度而造成步进电机丢步进而导致运动误差。实验方法采用，记录并对照步进电机目标运动位置 and 实际运动位置的偏差，进而评价其运动精度。为客观评价该龙门式机械臂运动误差的大小、形式以及形成误差的原因，该实验通过单片机向机械臂发送直线运动控制指令和随机运动控制指令，记录目标位置 and 实际运动位置以及二者误差，实验结果见表 5.2。

表 5.2 机械手运行结果

编号	目标位置 (mm)	实际位置(mm)	误差(mm)
1	(41, 32)	(42, 31)	1.4
2	(60, 12)	(62, 11)	2.2
3	(100, 25)	(101, 23)	2.2
4	(129, 41)	(129, 40)	1
5	(47, 66)	(49, 68)	2.8
6	(69, 48)	(70, 49)	1.4
7	(100, 47)	(100, 48)	1
8	(76, 87)	(77, 88)	1.4
9	(82, 65)	(83, 67)	2.2
10	(111, 65)	(110, 64)	1.4

从机械臂动作记录结果看,采用开环方式控制机械臂存在一定误差,但整体运动误差可以限定在 3 毫米以内,考虑到机械手具有一定的位置容错性,所以该误差可以满足抓取精度要求。根据机械臂运动误差分布特点,可以知道该机械臂误差是由于在启动阶段步进电机丢步所引起的。所以为了降低机械臂运动误差,机械臂在连续运行几个点后就需要回零点来去除运动的积累误差。

为验证算法引导机械手执行抓取的实验效果,将作为样本的小零件放置于机械手工作平面,利用工业相机进行图像采集并调用算法计算小零件位置 and 对应姿态,对于每个小零件重复做五次实验,记录每次机械臂定位误差和机械夹爪的抓取是否成功情况。实验结果汇总于表 5.3—5.12。

表 5.3 T 型滑块螺母抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差 (mm)	细长类零件姿 态识误差	是否成 功抓取
1	(158, 67)	3	20 度	是
2	(77, 80)	2.7	5 度	是
3	(47, 144)	1.4	35 度	否
4	(105, 119)	2.4	10 度	是
5	(95, 77)	1.6	5 度	是

5 实验研究

表 5.4 螺栓抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差 (mm)	细长类零件姿 态识误差	是否成 功抓取
1	(101, 48)	2.3	竖直识别成功	是
2	(141, 95)	2	0 度	是
3	(136, 94)	1.5	10 度	是
4	(122, 102)	1.2	15 度	是
5	(109, 133)	1.4	8 度	是

表 5.5 内六角螺栓抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差 (mm)	细长类零件姿 态识误差	是否成 功抓取
1	(110, 72)	1.8	竖直识别成功	是
2	(108, 120)	1.7	5 度	是
3	(90, 118)	1	9 度	是
4	(49, 120)	2.9	12 度	是
5	(133, 42)	1.8	10 度	是

表 5.6 自攻螺钉抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差 (mm)	细长类零件姿 态识误差	是否成 功抓取
1	(92, 70)	2.3	0 度	是
2	(135, 50)	2.6	3 度	是
3	(67, 158)	1.8	5 度	是
4	(112, 71)	2.9	8 度	是
5	(55, 154)	2.8	5 度	是

表 5.7 螺母抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差(mm)	是否成 功抓取
1	(114, 148)	2.6	是

5 实验研究

2	(58,42)	1.2	是
3	(139, 44)	1.7	是
4	(153, 147)	1.2	是
5	(124, 114)	0	是

表 5.8 沉头螺钉抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差(mm)	是否成功抓取
1	(70, 123)	1.5	是
2	(45, 154)	1.9	是
3	(36, 40)	1.7	是
4	(105, 54)	1	是
5	(73, 95)	3	是

表 5.9 圆头螺钉抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差(mm)	是否成功抓取
1	(52, 74)	2.1	是
2	(56, 58)	2.3	是
3	(134, 103)	2.8	是
4	(111, 44)	2.1	是
5	(67, 93)	1.5	是

表 5.10 弹簧垫片抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差(mm)	是否成功抓取
1	(90, 49)	2.6	是
2	(113, 82)	3	否
3	(112, 71)	1	是
4	(120, 107)	2.1	否
5	(143, 88)	2.1	是

5 实验研究

表 5.11 滑块螺母抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差(mm)	是否成功抓取
1	(81, 157)	2.4	是
2	(144, 84)	1.3	是
3	(76, 124)	2.9	是
4	(118, 84)	2.1	是
5	(145, 90)	1	是

表 5.12 角码抓取实验结果

编号	算法识别位置 (mm)	机械手定位误差(mm)	是否成功抓取
1	(69, 128)	2	是
2	(119, 91)	1.5	否
3	(84, 71)	3	是
4	(71, 66)	2.4	是
5	(117, 157)	2.8	是

根据抓取实验结果可知，目标检测算法对于小零件定位和类别识别的精度较高，实验中未发现较大误差，机械手在工业相机引导下能够实现较为精准的位置控制，运动误差能够控制在 3 毫米以下。对于细长类零件 T 型滑块螺母，由于其表面存在较大反光且该零件长宽比较小，姿态角回归算法对其姿态角识别偶尔会存在较大误差，进而导致机械夹爪在夹取过程中可能出现掉落的情况。其中弹簧垫片和角码在抓取过程也能出现“夹不起”的情况，其原因在于本实验所采用的两指式机械爪开合行程较短，导致机械爪没有夹住该零件。通过提高机械夹爪开合行程，或者采用更多“手指”的夹爪能够很好解决该问题。图 5.9 为机械手定位的实际情况，a、b、c、d 中矩形框为目标检测算法给出的目标零件包围框，红色点为机械爪定位的中心点。图 5.10 为机械手抓取效果图。

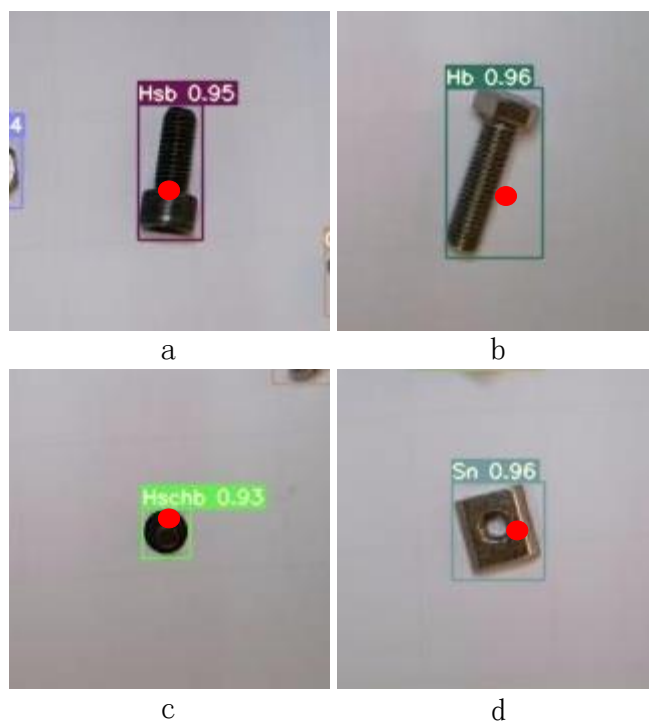


图 5.9 机械手定位情况

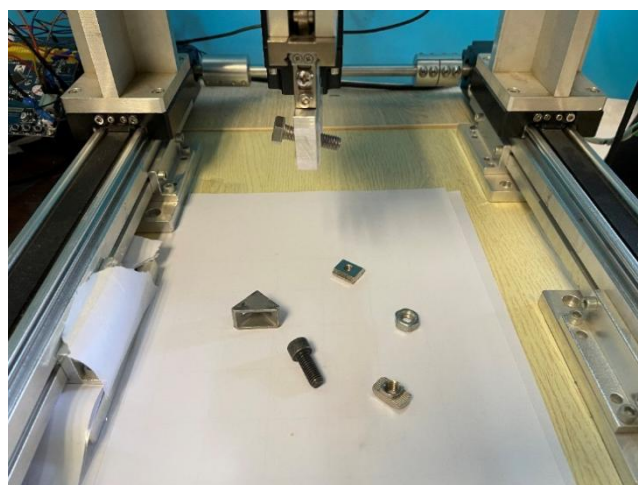


图 5.10 机械手抓取效果图

5.4 本章小结

本章首先分析了机械臂构成与搭建过程，同时分析用于执行抓取动作的机械臂所需要的各个功能模块以及各个模块所承担的功能，还分析了机械臂所需控制系统。为了将机械臂和工业相机相互协调，本章给出基于九点标定原理的机械臂手眼标定过程。为在实际应用场景下评价本文所设计并训练的算法性能，本章分别针对目标检测算法、深度预测算法以及姿态角回归算法制作三个验证数据集，将上述三个算法分别在验证数据集上进行测试，得到其实际识别精度。本章还通过单片机向机械臂发送运动控制指令，记录目标位置 and 实际运动位置以及二者误差，得出机械臂运动精度和引起误差的原因。最后应用该机械手分别对10个类别的零件进行抓取实验，得到各个零件抓取效果和引起抓取失败的原因，并给出对应的解决方案。

6 总结与展望

6.1 全文总结

在人口老龄化逐渐严重的背景下,分拣作业作为一种重复性简单劳动,其在工业生产线中属于最耗费人力、财力的工种,由于其工作性质具有重复性强、作业场景单一等特点,所以该工作较容易通过智能化改造来实现自动化生产。基于此本文提出了基于目标检测技术的平面态小零件智能分拣系统开发,首先分析传统机器视觉优缺点以及其在智能分拣中应用现状,根据现有应用现状和其技术特点得出其在复杂应用环境下鲁棒性不足的缺点。同时,本文还分析了基于深度学习技术的一系列图像处理算法,指出其相较于传统图像处理算法的优点以及对应的技术特点、解决方案等。

为解决传统图像处理算法在智能化分拣中表现出鲁棒性不足等缺点,本文设计了一套采用深度学习和传统图像处理的机器视觉算法,该算法包含目标物体存在判定功能、目标物体位置回归功能、目标物体类别识别功能、目标物体中细长类零件深度与姿态角度预测功能。该算法相互配合能够根据工业相机采集到的图像数据计算出目标物体在图像中中心点坐标、包围矩形框、以及机械手对于细长类零件抓取姿态角。上述运算结果通过机械手与工业相机之间手眼标定矩阵进行变换,即可得到能够直接控制机械手抓取的位置信息。最后上位机再将位置信息通过串口通信发送给下位机,下位机控制机械臂即可完成小零件抓取。本文所做的主要工作总结如下:

(1) 将内卷积计算子代替目标检测算法主干网络中标准卷积,提高算法主干网络的执行效率。

(2) 剪裁掉 SSD 网络中深层的三个检测头,并引入特征金字塔思想做特征融合,达到在提高运算效率前提下实现对小目标物体检测性能提升的目的。

(3) 设计基于深度卷积神经网络的图像分类算法,实现对细长类螺栓零件深度估计,进而更好引导机械手抓取该类零件。

(4) 设计基于传统图像处理算法的细长类零件姿态角预测算法,实现对细长类零件在工作台上姿态角检测。

(5) 搭建基于步进滑台的三轴龙门式机械臂, 采用 STM32 单片机搭建机械臂控制系统, 实现具有较精确运动控制的机械臂。

(6) 采用九点标定方法实现工业相机与机械臂的标定, 得到像素坐标到工作台平面之间坐标变换矩阵。

6.2 展望

本文针对改进 SSD 的平面态散乱零件分拣系统做了大量相关工作和实验, 采用新方法、新理论研究并实现了智能分拣相关算法, 尽管取得了一定的研究成果, 但由于课题时间有限, 仍然存在很多需要完善的部分, 本课题需要在以下几个方面需要进一步研究。

(1) 本课题研究范围限定在平面态零件分拣, 对于相互堆叠的应用场景没做相关研究和思考, 所以该课题进一步需要研究小零件在工作台上如何才能实现平面态分布和算法应当如何处理特殊堆叠的情况, 这样才能将该系统可靠地接入生产线中。

(2) 本课题所采用的视觉传感器是单目相机, 这就导致该课题在理论上仅能在二维平面上实现抓取功能, 零件相对于相机的深度信息无法直接获取。对于细长类零件可能存在的深度变化也仅是根据其在二维相机中成像区别去做判断。进一步的, 为提高该分拣系统应用范围, 需要采用双目相机或者激光雷达实现对零件的实际深度信息获取。

(3) 通过实验发现, 该三轴式龙门机械臂由于是采用开环控制, 其驱动步进电机在起步阶段存在丢步可能, 这导致机械臂在连续运动时存在积累误差。为进一步提高其运动精度, 需要采用配备编码器的伺服电机作为机械臂驱动电机。

参考文献

- [1]. 韩峰涛.工业机器人技术研究与发展综述[J].机器人技术与应用, 2021, (05):23~26
- [2]. 宋博仕.工业机器人自动化生产技术的实践研究[J].现代制造技术与装备, 2021, 57(09):179~183
- [3]. 伍锡如, 黄国明.基于深度学习的工业分拣机器人快速视觉识别与定位算法[J].机器人, 2016, 38(06):711~719
- [4]. 赵丽君, 李冰冰.基于视觉检测的工业机器人快速分拣控制方法[J].制造业自动化, 2021, 43(10):86~90
- [5]. 王鸿博, 李建东.基于工业机器人的分拣生产线群控通信系统设计[J].制造技术与机床, 2016(02):170~173
- [6]. 彭林鹏.基于 RGB-D 的多尺寸箱体智能分拣与码垛规划[D].浙江: 浙江大学, 2021
- [7]. 黄万军.关于特定目标的智能识别和分拣技术研究[D].长沙: 电子科技大学, 2020
- [8]. 李策. 基于机器视觉的多机器人协作分拣技术研究[D].天津: 天津大学, 2019
- [9]. 张彬. 基于机器视觉的目标分拣算法研究[D].郑州: 郑州大学, 2019
- [10].朱雨贺. 基于深度学习目标检测的机械臂分拣系统研发[D].浙江: 浙江大学, 2019
- [11].彭杰, 孟祥印.基于机器视觉的工件分拣及上下料系统[J].机床与液压, 2021, 49(21):38~42
- [12].程子华.基于机器视觉的残缺饼干分拣系统开发[J].现代食品科技, 2022, 38(02):313~318
- [13].王诗宇. 智能化工业机器人视觉系统关键技术研究 [D].沈阳: 中国科学院大学(中国科学院沈阳计算技术研究所), 2021
- [14].陈啸宇. 多目标检测理论与方法研究[D].成都: 电子科技大学, 2021
- [15].夏超群. 基于局部和全局特征表示的红外小目标检测算法研究[D].浙江: 浙江大学, 2019
- [16].尚晋霞. 显著性目标检测方法及应用研究[D].四川: 四川大学, 2021
- [17].Krizhevsky, Sutskever. Imagenet classification with deep convolutional neural networks [J].Advances in neural information processing systems,2012,25.
- [18].何江, 张科兴.基于机器视觉的 AlexNet 网络煤矸石检测系统[J].煤炭技术, 2022, 41(03):205~208
- [19].董文轩, 梁宏涛. 深度卷积应用于目标检测算法综述[J].计算机科学与探索, 2022, 3(24):1~20
- [20].张振伟, 郝建国.小样本图像目标检测研究综述[J].计算机工程与应用, 2022, 3(24):1~14
- [21].科岑, 王晓强.深度学习中的单阶段小目标检测方法综述[J].现代计算机科学与探索, 2022, 16(01):41~58
- [22].李柯泉, 陈燕. 基于深度学习的目标检测算法综述[J].计算机工程, 2022, 3(24):1~17
- [23].崔磊, 仇晓兰.一种基于误差反向传播优化的多通道 SAR 相位误差估计方法[J].雷达学报, 2020, 9(05):878~885

- [24].赵虎, 杨宇.基于迭代式 MapReduce 的误差反向传播算法[J].计算机应用, 2016, 36(04):923~926
- [25].范秀娟.基于自适应误差反向传播信号调整的 BP 网络研究与应用[J].计算机应用与软件, 2010, 27(08):117~119
- [26].朱奕坤, 郭从洲.误差反向传播卷积神经网络的权值更新 [J]. 信息工程大学学报, 2021, 22(05):537~544
- [27].王孝义, 邢素霞.基于 SITF 算法的乳腺肿块分类[J].计算机仿真, 2022, 39(01):446~450
- [28].宋琳. 基于局部特征的人体行为识别算法研究 [D].南京: 南京邮电大学, 2021
- [29].罗亮, 吕俊杰.基于改进的 NC-HOG 特征的工程车车型自动识别算法[J].计算机工程与设计, 2021, 42(11):3164~3173
- [30].宋一言, 唐东林.改进穿线法与 HOG+SVM 结合的数码管图像读数研究[J].计算机科学, 2021, 48(02):396~399
- [31].Wang L, Guo S, Places205-vggnet models for scene recognition [C].arXiv preprint arXiv,2015,6(08)
- [32].Simonyan K, Zisserman A Very deep convolutional networks for large-scale image recognition [C].arXiv preprint arXiv,2014,6(12)
- [33].He K, Zhang X, Deep residual learning for image recognition [C]. Proceedings of the IEEE conference on computer vision and pattern recognition,2016,770-778
- [34].汪凯旋, 陈光化.基于改进的 ResNet 手指静脉识别[J]. 激光与光电子学进展, 2021, 58(20):100~106
- [35].王鑫鹏, 王晓强.深度学习典型目标检测算法的改进综述[J].计算机工程与应用, 2022, 58(06):42~57
- [36].赵立新, 邢润哲.深度学习在目标检测的研究综述[J].科学技术与工程, 2021, 21(30): 12787~ 12795
- [37].Zou Z, Shi Z, Guo Y, Object detection in 20 years: A survey [J]. arXiv preprint arXiv,2019,3(1)
- [38].王成军, 韦志文.基于机器视觉技术的分拣机器人研究综述[J]. 科学技术与工程, 2022, 22(03):893~902
- [39].廖凯, 陈坚泽.基于机器人视觉传感技术及应用研究综述[J].现代信息科技, 2020, 4(11):159~162
- [40].王成军, 严晨.机器视觉技术在分拣系统中的应用研究综述[J]. 制造技术与机床, 2020, 05(02):32~37
- [41].龙慧, 朱定局.深度学习在智能机器人中的应用研究综述[J].计算机科学, 2018, 45(02):43~47
- [42].刘帆, 王颖. 基于差分的图像感知哈希算法[J].计算机工程与设计, 2021, 42(03):782~789
- [43].赫泽宏.图像感知哈希算法及其应用研究[D].天津: 天津工业大学, 2020
- [44].Liu W, Anguelov D, Ssd: Single shot multibox detector [J]. European conference on computer vision. Springer, Cham,2016, 43~47
- [45].洪哲昊, 陈东方.基于多任务分支 SSD 的目标检测算法[J].计算机工程与设计, 2022, 58(04):134~142

- [46].王静, 孙紫云.改进 YOLOv5 的白细胞检测算法[J].计算机工程与应用, 2022, 43(11):117~119
- [47].褚文杰.基于 YOLOv5 的坦克装甲车辆目标检测关键技术的研究[D].北京: 北京交通大学, 2021
- [48].Li D,Hu J, Wang C Involution: Inverting the inference of convolution for visual recognition [C].Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021:12321-12330
- [49].凤冰霞. K-means 算法的一种新解法及应用[D].南昌: 华东交通大学, 2021
- [50].薛丁文. 基于核心集的 k-means 聚类算法研究[D].哈尔滨: 哈尔滨工业大学, 2019
- [51].宋雅婷. 基于卷积神经网络的垃圾检测算法研究与应用[D].南昌: 南昌大学, 2021
- [52].Yamashita R, Nishio M. Convolutional neural networks: an overview and application in radiology [J].Insights into imaging, 2018,9(4)
- [53].Albawi S, Mohammed T A. Understanding of a convolutional neural network [C].international conference on engineering and technology (ICET),2017,6(1)
- [54].Bochkovskiy A, Wang C. Y Yolov4: Optimal speed and accuracy of object detection [J]. arXiv preprint arXiv,2020,2004.10934
- [55].Girshick R. Fast r-cnn [C].Proceedings of the IEEE international conference on computer vision,2015,1440-1448
- [56].夷德. 基于 YOLO 的目标检测优化算法研究[D]. 南京: 南京邮电大学, 2021
- [57].舒志鸿. 基于边缘计算的联邦学习方法的研究[D]. 南京: 南京邮电大学, 2021
- [58].高媛媛. 基于深度学习的移动端图像识别研究与实现[D].宁夏: 宁夏大学, 2021
- [59].申世博. 深度神经网络结构性学习及压缩[D].浙江: 浙江大学, 2021
- [60].周立君, 刘宇, 等. 使用 TensorRT 进行深度学习推理[J].应用光学, 2020, 41(02):337~341
- [61].黄诚. 仿人机械手设计及其控制算法研究[D].[硕士学位论文].成都: 电子科技大学, 2019
- [62].祝鹏. 基于 ARM 的注塑机专用机械手控制系统设计[D].广州: 华南理工大学, 2015
- [63].王宇. 多轴机械手控制系统的研究与开发[D].浙江: 浙江工业大学, 2012

个人简历、在学期间发表的学术论文与研究成果

个人简历

张志浩，男，生于 1994 年 11 月，河南驻马店人。

2019 年 9 月至今，就读于郑州大学机械与动力工程学院，机械工程专业，获得工学硕士学位。

在学期间发表的学术论文与研究成果

时间	参赛作品	赛事名称	获奖荣誉
2021 年 6 月	工业零件智能分拣机器人	郑州大学第三届“研究生机器人创新设计大赛”	二等奖

- [1]. 邱益, 张志浩, 梁杰. 基于深度学习的无序件抓取系统开发[J]. 实验室研究与探索, 2022, 02 (中文核心, 已录用刊出)

致谢

天波易谢,寸暑难留,三年的研究生时光如白驹过隙。回首在郑州大学的三年学习生活,感慨颇多!回首这三年,忙碌的日子总是那么令人难忘,不能忘怀和同门以及师弟们在实验室讨论问题的画面,不能忘怀老师给予我的指导与教诲,不能忘怀与同学、舍友一起畅饮谈笑的场景。还记得和老师、实验室师兄们一起去食堂打饭的那段日子,每天三点一线的生活虽然枯燥,但那是进步最快的一段时期。研究生三年我收获满满,有使我立身安命的知识、有给予我无私帮助、谆谆教诲的导师、有活泼乐学,互帮互助的师兄们以及给予我支持和陪伴的女友。

在即将走向社会之际,回顾多年来的学习生涯,也曾不懈追求过许多,亦曾遗憾失去过许多。感念来时的路,一句时光太过匆匆似乎难以描述此时波澜复杂的心情。饮水思源,学成念师。感谢我的导师邱益教授和梁杰教授在研究生学习生涯中给予我的帮助与教导。在这三年中,二位导师在生活、学习以及就业中给予我的帮助与指导使我受益匪浅。二位导师宽阔的视野与胸怀,严谨求是的工作态度使我如沐春风,这将深刻影响我对于今后生活和工作的态度,我也定将积极向上,秉持导师的作风与态度,做个合格的社会主义接班人。感谢风趣幽默、平易近人的杨炯老师,杨老师在我课题开展期间给予了我很多具有建设性的意见。感谢杨老师给予我的锻炼与指导。

感谢研途中与我相伴的同门师兄弟,感谢你们在研究生期间给予我的帮助。因为有你们,在那些乏味、单调的学习日子里才不会觉得无聊。感谢张健借阅给我的书籍,他的文学素养以及对美的追求令人钦佩。感谢崔松涛、王帅普带领我们参加多次学科竞赛,他们积极的学习态度值得学习。感谢张志强日常的鞭策与监督。特别感谢张康宁同门,他对人、对事严谨的态度以及积极乐观的精神值得学习。感谢他在我实习期间多次帮助我处理我不能及时办的事情,感谢他帮我做论文格式的校对与修改。感谢曹若彦师弟帮我校对论文直到深夜。感谢万宝元师弟帮助处理入党材料的整理。还要感谢孙光耀日常提醒我健康打卡,感谢何宜斌师弟在生活中的帮助。感谢袁鹏、王孟建、马洋洋三位室友,感谢你们在生活和学习上的支持与帮助,你们的优秀催我奋进。

春晖寸草,山高海深。二十载漫漫求学路,感谢父母无私的付出与悉心栽培。父母虽才疏学浅,但始终教导我知识的重要性,鼓励我在求学道路上取得一些成就。家境虽不富裕,但始终没有放弃对我教育上的支持。大恩不言谢,但我知道父亲在烈日下在挥洒汗水、我知道母亲在遥远的北京思念家人。感谢你们对家的支撑,等我毕业一定让你们放下生活的重担。

最后,感谢我的对象周倩丽,你乐观积极的生活态度时常鼓励我勇敢面对工作和学习中的挑战。感谢你的陪伴与鼓励,感谢你在我不开心时耐心的开导。愿我们喜乐有分享,共度日月长。

愿疫情早日结束,愿春天和自由重返大地。向抗疫勇士致敬!