

【特别策划】

机器人无序分拣技术研究

翟敬梅, 黄乐

(华南理工大学 机械与汽车工程学院, 广州 510641)

摘要: **目的** 介绍机器人无序分拣技术最新进展和研究成果, 为复杂场景下的机器人自动化应用提供研究思路和技术支撑。**方法** 从机器人无序分拣过程中的 3 个关键技术任务展开论述, 具体包括散乱目标的检测与识别、目标的空间姿态估计与无序分拣的抓取决策, 并对各个任务中涉及的方法进行分析总结。**结论** 机器人无序分拣技术目前面临的主要挑战在于处理复杂环境下散乱堆叠物体的视觉感知、位姿估计和抓取决策。具体的分拣系统应当考虑实际的场景环境, 结合任务需求进行设计, 以求达到机器人尽可能替代人力劳动的目的。

关键词: 无序分拣; 目标检测与识别; 空间位姿估计; 抓取决策

中图分类号: TB472; TP242.6 **文献标识码:** A **文章编号:** 1001-3563(2022)08-0066-10

DOI: 10.19554/j.cnki.1001-3563.2022.08.009

Review of Unordered Picking Technology for Robots

ZHAI Jing-mei, HUANG Le

(South China University of Technology, School of Mechanical & Automotive Engineering, Guangzhou 510641, China)

ABSTRACT: This paper introduces the latest progress and research results of robotic unordered picking technology, and provides research ideas and technical support for robot automation application in complex scenarios. It discusses three key technical tasks in the robot unordered picking process, which including detection and recognition of cluttered objects, spatial pose estimation of objects, grasping strategy of unordered picking. Analyzes and summarizes the methods involved in each task. The main challenges of robotic unordered picking technology today are dealing with visual perception, pose estimation and grasping strategy of cluttered and stacked objects in complex environments. Specific picking systems should be designed considering the actual scenario environment and combined with the task requirements in order to achieve the goal of replacing human labor by robots as much as possible.

KEY WORDS: unordered picking; object detection and recognition; spatial pose estimation; grasping strategy

随着科学技术的发展和工业技术的进步, 现代社会的生产模式逐渐由传统的以人类劳动力为主导转变为以智能机器人为主导。目前机器人分拣已经广泛应用于生产制造、零件装配、搬运码垛与智能物流等领域中, 尤其在电商物流行业, 机器人分拣技术存在非常广阔的市场。2020 年中国社会物流总额达 300.1 万亿元, 总收入为 10.5 万亿元。物流仓储中货物依靠人工完成这些重复性高、强度大的任务不仅生产效率低下, 而且给劳动人员身心带来危害。许多大型物流企业引进机器人、物联网技术并研究智能分拣技术。亚马逊等大型电商公司举办了机器人无序分拣大

赛, 寻求机器人代替人工分拣的解决方案。图 1 分别展示了 2015 年—2017 年无序分拣挑战大赛获得冠军的团队。第一届来自德国柏林工业大学的 RBO 团队在硬件上采用 Barrett 机械臂、XR4000 移动平台和深度相机组成移动分拣平台, 在软件上基于直方图反向投影法实现目标识别算法^[1]; 第二届的 Delft 团队在视觉算法上引入了先进的目标检测算法 Faster R-CNN 对目标边界框提取, 大大提高目标检测识别效率, 并在安川 SIA20F 机械臂末端安装具有旋转自由度的吸盘进行抓取^[2-4]; 第三届的 MIT 队伍在机械结构和视觉算法上均做了较大改进, 机器人末端执行器

收稿日期: 2021-12-06

作者简介: 翟敬梅 (1967—), 女, 博士, 教授, 主要研究方向为机器人技术、机器智能、机器视觉。

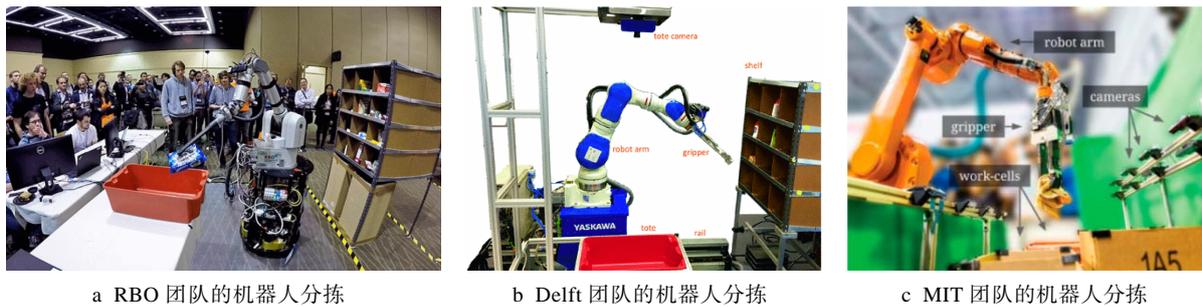


图 1 亚马逊无序分拣大赛
Fig.1 Amazon Picking Challenge

采用多功能二指夹爪, 由平行性夹爪和定制吸盘组成, 视觉方案采用 FCN 全卷积神经网络输出 4 个不同方向的置信度评分作为抓取依据, 在手眼操作上表现得更为协调^[5-6]。

机器人分拣应用场景主要有 3 种: 结构化、半结构化和非结构化, 非结构化环境下的目标分拣称为无序分拣。在非结构化环境下, 物料摆放的位置通常是无序且杂乱的, 复杂的分拣场景对于机器人实现自动化极具挑战性, 因此无序分拣问题也称为机器人操作中的“圣杯问题”。基本的无序分拣系统框架见图 2。由硬件平台和软件平台 2 部分组成, 硬件平台包含机器人、相机、计算机等设备, 软件平台包含目标识别、位姿估计、抓取规划等算法。为了实现机器人的无序分拣任务, 机器人首先需要感知物体的位置和方向, 常用的视觉传感器有 RGB 相机和深度相机, 其中深度相机能够提取场景中的深度信息, 在无序分拣平台中更为使用广泛。机器人的无序分拣过程大致分为目标识别、位姿估计、抓取规划 3 个部分。目标检测是检测图像中某一类目标的实例, 当场景中存在多类目标需要分拣时, 需判断抓取目标的类别。目前基于视觉的目标检测与识别技术逐渐趋于完善, 由于目标之间堆叠、遮挡和背景杂乱等环境因素, 均可能影响目标检测和识别的准确度。大多数生产流水线上的目标分拣场景较为简单, 只需确定分拣对象在二维平面下

的位姿即可抓取, 而对无序杂乱的目标而言, 实施分拣变得更加困难, 确定目标在三维空间下的位姿是机器人抓取成功的关键, 也是无序分拣区别于普通情况下机器人分拣的重要体现。视具体分拣场景, 根据分拣目标的位姿进行机器人抓取的过程中, 机器人与料框及其他目标之间均有可能发生干涉和碰撞, 因此合理的抓取规划有利于避免一些异常情况的发生, 提高无序分拣的成功率。

机器人视觉在无序分拣技术中发挥着更加重要的作用, 国外的一些公司也在逐渐将自主研发的视觉系统嵌入机器人中, 实现无序分拣过程的自动化。日本的 FANUC 公司推出了 iRVision 视觉系统以识别和定位零件, 并将其紧密耦合在机器人控制器中, 以确保视觉伺服过程中的手眼协调^[8]。德国的 ISRA VISION 公司也推出了 Shape Scan 3D 系统与 IntelliPICK3D 系统, 该视觉传感器系统对复杂零部件采用 CAD 文件读取, 可以不受零件的角度限制, 同时集成多种监测手段, 对比较细小或者表面比较肮脏的零部件都可以识别, 加快了生产线的运转, 实现更高的生产节拍^[9]。国内的一些企业一直致力于无序分拣系统的研发, 各种人机协作机器人的出现也为此带来契机, 利用协作机器人代替人类, 提高了企业的劳动力水平, 同时提高自身竞争力。爱尔森智能科技有限公司(Alsontech)是国内首家自主研发机器人 3D 视觉引导系统的公司。针对散乱无序工件的分拣, 该系统通过 3D 快速成像技术对工件表面轮廓数据进行扫描, 获取工件点云数据, 辅以机器人路径规划和防碰撞技术, 计算出当前工件的实时坐标, 并发送指令给机器人实现自动分拣^[10]; 阿丘科技(Aqrose)基于团队核心的 3D 视觉技术和机器学习, 通过对目标物体的识别和定位, 实现工业机器人对复杂分拣环境处理能力, 自动完成识别分拣的相关任务^[11]。无序分拣技术是未来智能机器人研究的一个重要方向, 在现代化社会生产中逐步实现机器换人的大趋势下, 具有广阔的应用前景和巨大的经济价值。

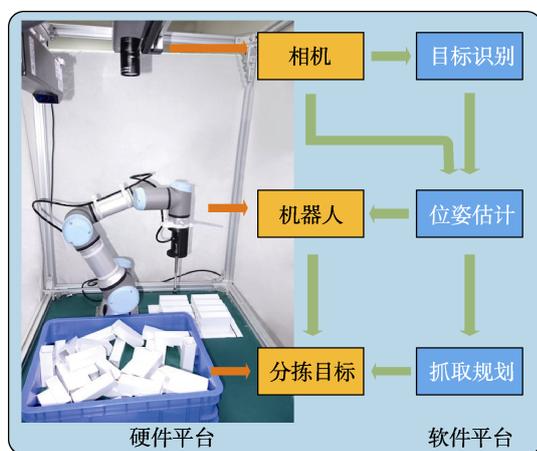


图 2 无序分拣系统框架图^[7]
Fig.2 Unordered picking system framework diagram^[7]

1 散乱目标的检测与识别

目标的检测识别是机器人抓取领域中常见的技

术手段,快速准确地对目标进行识别,对机器人无序分拣速度和精度的提高起到重要作用。然而在目标散乱堆叠的环境下,除了环境因素的干扰外,还需考虑目标之间遮挡带来的特征检测难题,目标检测识别的成功率将直接影响后续抓取的成功率。目前目标的检测与识别方法总体可以归纳为以下几类:基于模板匹配的方法、基于特征匹配的方法、基于霍夫森林的方法和基于深度学习的方法,见图3。

基于模板匹配的方法是从已有的数据中制作模板,通过测量待匹配图像与模板之间的相似性来确定目标所属类别。此类方法由于不需预先训练,鲁棒性强,常用于散乱堆叠目标的识别中。文献[12]制作了两个抓取位置检测的模板去扫描相机捕获的图像,利用差分进化算法高效搜索目标函数最大的抓取位置,快速实现散乱螺栓的识别。但该算法要求目标与模板具有相同的尺寸、方向和图像,并且根据模板在搜索目标的过程中只能进行平行移动,若原图像中目标发生旋转或大小变化,则算法容易失效。针对模板搜索过程中的大小和方向局限性,文献[13]通过以参数化的方式构造了模板,基于灰度匹配识别待抓取的铆钉,这种方式可以扩展到多尺寸铆钉类型的识别。虽然基于模板匹配的方法具有简单快速的特性,但还存在着对光照、亮度等环境因素敏感等问题需要解决。

基于特征匹配的方法是通过寻找两幅图像之间的特征对应关系进行目标识别定位,这类特征通常为手工设计的描述符。二维图像检测中典型的特征描述符有 SIFT (Scale Invariant Feature Transform)^[14]、SURF (Speeded Up Robust Feature)^[15]、ORB (Oriented

FAST and Rotated BRIEF)^[16]等,除此之外还有一些学者针对特定场景中的目标识别而设计一些描述符,其基本原则为外观相似的特征应该有相似的描述子。SIFT 描述子在图像的不变特征提取方面具有一定的优势,常用于盒子书本等堆叠物体的识别中。文献[17]提取目标图像的 SIFT 特征点并将其按极大特征值点与极小特征值点进行分类,根据 SIFT 特征点区域方向直方图计算的角度信息限制匹配范围进行识别。当目标在待识别图像中发生局部遮挡、旋转、尺度变化或者弱光照等情况下,该算法仍能完成目标的快速识别。但 SIFT 描述子还存在着实时性不高、特征点较少、对边缘模糊的目标无法准确提取特征点等不足。为了改进 SIFT 描述子的一些缺陷,文献[18]采用改进的自适应阈值的 Canny 边缘提取算子提取目标的边缘信息,采用 SURF 特征与 FREAK(Fast Retina Keypoint)特征结合的方式对目标进行描述,加快模型的建立,缩短目标识别的时间。

优化后的特征点描述方法在匹配速度和鲁棒性方面都有改善,但构造出足够多的特征点需要图像具备丰富的纹理,否则构建的特征区别性不大,容易造成误匹配,甚至完全失效。对工业上用于分拣的目标,大多为弱纹理或者无纹理的物体,此类物体常通过检测目标点云的三维特征描述子进行识别。常用的三维特征描述子可分为局部特征描述子和全局特征描述子,局部特征包括 PFH (Point Feature Histogram)^[19]、FPFH (Fast Point Feature Histograms)^[20]、SHOT (Signature of Histogram of Orientation)^[21]和 Spin image 等,全局特征描述子包括 VFH (Viewpoint Feature Histogram)^[22]、

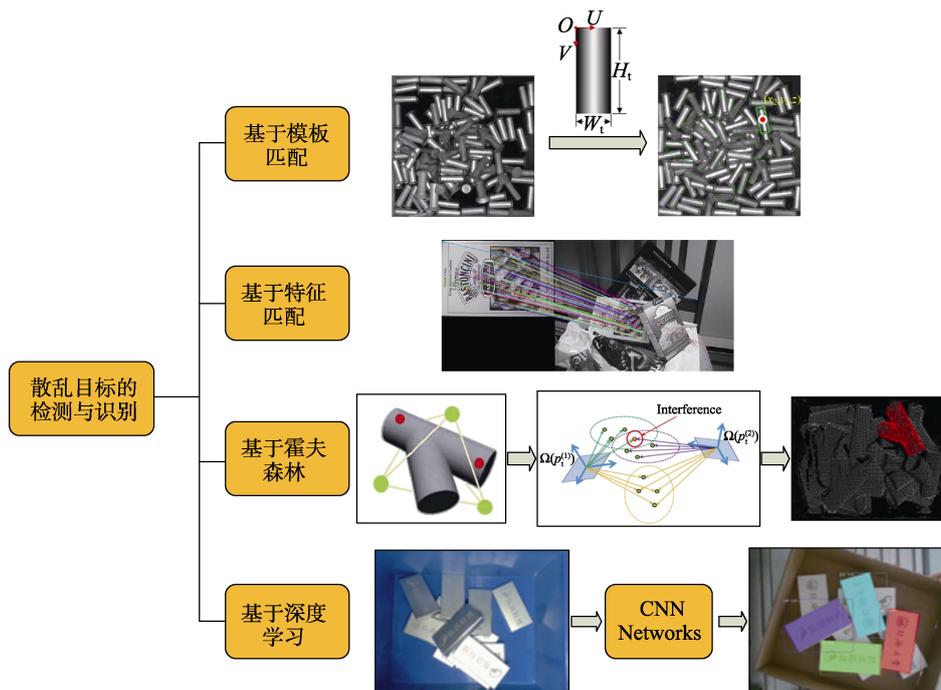


图3 散乱目标的检测与识别方法

Fig.3 Detection and identification method of cluttered objects

OUR-CVFH (Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram)^[23]和 ESF (Ensemble of Shape Functions)^[24]等, 根据 2 组点云之间的特征描述子进行匹配, 以此作为识别的重要依据。文献[25]和文献[26]分别提取 Spin image 描述符和 SHOT 特征进行匹配识别。文献[27]设计了一个高度描述性、鲁棒性和计算效率高的局部形状描述符来建立模型点云和场景点云之间的对应关系, 利用关键点的局部参考帧进行聚类的方法来选择正确的对应点。此外, 堆叠目标整体提取三维特征描述子耗时过长, 因此有学者提出先将散乱堆叠的目标点云进行分割, 对分离的目标提取三维特征再进行匹配识别, 可以有效提高识别的效率。为了防止对散乱目标点云产生过分割或欠分割, 文献[28]采用超体聚类的过分割方法将点云数据分割成许多小块, 在超体聚类的基础上再聚类生成各个目标点云, 结合 VFH 特征和 FPFH 特征对点云对象进行识别。三维特征描述子相对于二维特征描述子在杂乱的环境中有更强大的鲁棒性, 并且除了用于目标识别外, 还可以估计目标的位置和方向, 更有利于物体的抓取操作。

基于霍夫森林的方法是通过构建随机森林从图像上提取图像块, 在构建的随机森林中的每个决策树上进行判断处理并在霍夫空间进行投票, 根据投票结果实现目标识别。针对杂乱和遮挡的场景中模型的多个实例识别, 文献[29]利用三维特征检测和描述来计算三维模型与当前场景之间的一组对应关系, 将每个特征点与其相对于模型质心的位置相关联, 使其在三维霍夫空间内可以同时对所有特征进行投票识别。为了可以处理更加稀疏的点云对象识别, 文献[30]提出了一种基于面向方向的点对特征创建全局模型描述的方法, 并使用快速投票方案对模型进行局部匹配, 在简化的二维搜索空间中利用有效的投票方案进行局部识别, 具备更好的识别性能。为了提高霍夫投票识别性能, 文献[31]利用给定三维场景的局部特征生成一组参考点集合, 参考点集合由几个参考点组成, 每个参考点用绿色或红色表示, 其中红色的参考点用于验证假设, 绿色参考点用于霍夫投票, 通过将六维的投票空间分解为几个三维空间实现更加有效的三维目标识别。

基于近些年来深度学习的方法发展迅速, 由于深度神经网络的强特征提取能力, 此类方法在复杂堆叠场景下目标的检测识别具有明显的优势, 大致可以分为两阶段 (two-stage) 和一阶段 (one-stage) 两大类。two-stage 方法首先生成一系列样本的候选区域, 利用卷积神经网络对候选区域分类进行识别, 包括 R-CNN^[32]、Faster R-CNN^[3]和 Mask R-CNN^[33]等方法。为了解决深度学习训练需要大量人工标注数据集的问题, 文献[34]采用物体和背景的各种组合生成图像增强数据集的技术并自动标注, 采用 Faster R-CNN 在生成的数据集上进行训练, 识别出各个堆叠散乱的

目标。Faster R-CNN 识别出的目标采用矩形包围框的形式表示目标的边界框, 但矩形包围框的精度较低, 在需要机器人精密操作的场合下并不适用, 因此文献[35]和文献[36]采用了实例分割网络 Mask R-CNN 确定分拣箱中目标的种类和形状, 实例分割的方式可以精确到物体的边缘信息, 但识别过程更为耗时。

One-stage 方法不用产生候选框, 直接将目标框定位的问题转化为回归问题处理, 识别速度比 two-stage 方法更快, 适用于散乱目标实时检测识别场景, 包含 YOLO^[37]、SSD^[38]和 RetinaNet^[39]等方法。文献[40]为实现工业机器人对复杂工件的自动分拣, 采用边界像素检测算法和深度神经网络识别训练算法对目标进行精确定位与图像分割, 利用 CNN 模型识别检测目标, 能够快速准确地识别与定位复杂目标物体。针对散乱目标中小对象的检测识别, 文献[41]基于改进的 YOLO-V3 算法, 对网络结构和多尺度检测进行了改进, 并引入了空心卷积的思想, 利用残差密集块替代原算法中的残差密集块, 增强网络对工件特征信息的提取, 提高对小对象的检测能力。文献[42]基于 SSD 检测算法设计了堆叠目标检测定位模块, 通过优化目标函数和修改网络结构, 解决小面积堆叠的目标识别问题, 在不损失检测速度的前提下, 保证了目标的准确识别定位。

2 分拣目标的空间姿态估计

目标的空间 6D 位姿估计可以帮助机器人感知要抓握抓取的物体的位置和方向。目标的 6D 位姿是指相机坐标系下目标的 3D 位置和 3D 姿态, 或者说确定物体坐标系相对于相机坐标系的平移和旋转变换关系^[43]。位姿估计方法大致可以分为四种, 分别是基于 CAD 模型、基于模板、基于投票与基于回归的方法, 见图 4。

基于 CAD 模型的方法主要针对具有标准 CAD 模型的目标, 通过寻找输入点云和已有模型之间的 3D 特征描述符的对应关系进行位姿估计。文献[44]通过虚拟相机采集工件三维点云数据作为 CAD 模型数据库, 从模型数据的点云计算出所有的点对特征描述符构建全局模型描述, 利用特征匹配对场景目标进行位姿估计。堆叠目标整体点云与模型匹配获取位姿过于耗时, 为了更有针对地对单个目标进行位姿估计, 文献[45]基于 Mask R-CNN 算法检测并分离图像中的不同对象, 同时为每个对象生成高质量的分割掩膜, 利用结构光系统的三维三角剖分计算出每个对象的三维点云, 基于 ICP (Iterative Closest Point) 算法对模型和目标点云配准得到位姿信息。结合目标点云信息里的其他属性进行位姿估计可以进一步提高位姿估计的精度, 文献[46]在匹配策略上引入物体的轮廓、正态分布和曲率等几何属性进行性能优化, 利用数据库中的目标和真实场景进行比较调度, 生成姿态聚

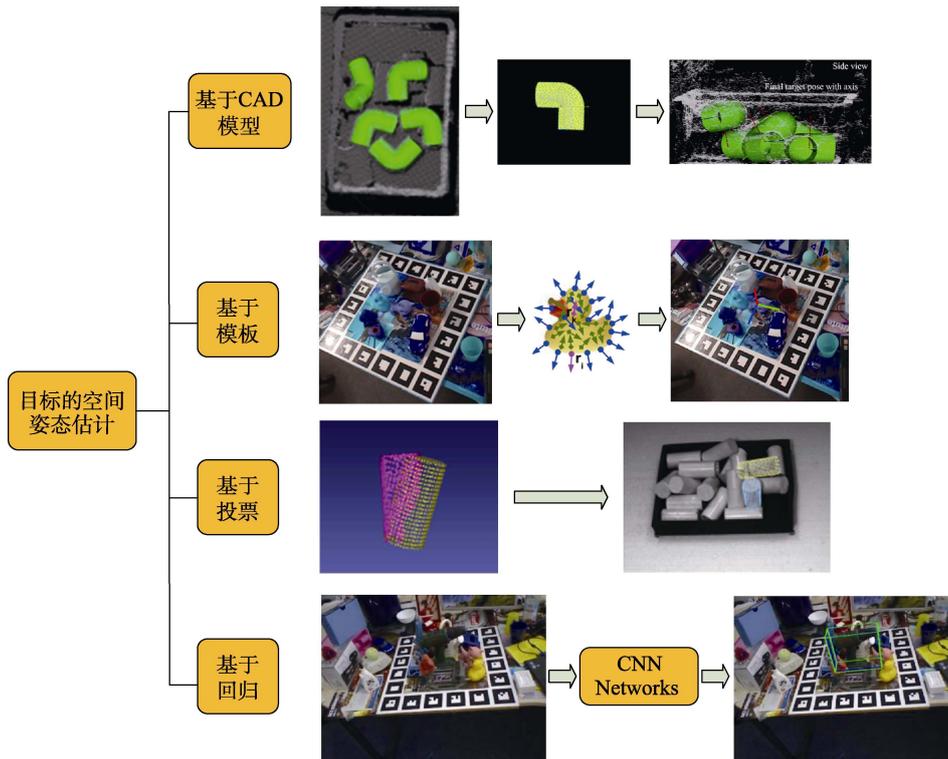


图4 目标的空间姿态估计方法

Fig.4 Spatial pose estimation methods of objects

类，用于对不同形状物体的位姿估计。根据现有的CAD模型通常可以获得较好的位姿估计性能，但在没有专业的扫描设备或者目标不易建模的情况下，很难构建出精细的三维模型，还需通过其他方式进行位姿估计。

基于模板的方法是从标记好6D位姿的模板中，选择最相似的模板，将其作为当前物体的6D位姿。针对无纹理目标的位姿估计，文献[47-49]提出了一种基于模板匹配的物体位姿估计方法LineMod，通过计算输入RGB图像的梯度信息并结合深度图的法向特征作为模板匹配的依据，能够在实时情况下实现两千个模板的匹配，解决了在复杂背景下无纹理物体的实时位姿估计问题。但LineMod算法的特征提取只在输入图像的强边缘进行，仅使用LineMod模板匹配可能导致匹配失败，特别是对具有类似局部轮廓的对象。因此，文献[50]在基于位姿模板的基础上，利用ICP的点云配准方法对模型点云与场景点云进行优化，以提高LineMod位姿检测的结果。为了提高机器人在非结构化环境中的感知能力，文献[51]基于SSD神经网络框架实现初始检测目标并生成感兴趣区域，应用LineMod模板匹配来提供候选模板，通过设计的聚类算法对计算出来的模板进行分组，去除相似结果后采用ICP算法估计目标的位姿，可以同时检测多个目标并估计其位姿。

基于投票的方法每个像素或者每个3D点通过投票，贡献于最终的6D位姿估计。为了获取精确的目标位姿估计，文献[52]将参数最小化的遗传算法与霍

夫投票相结合，使用多视角数据融合，进行配准获取三维位姿。针对三维点云数据的不连续性，文献[53]在物体表面上的点与法线组成的几何特征基础上，引入了边缘抑制损失函数用于约束物体的三维形状，同时在投票框架中加入了边缘信息这一紧凑特征进行估计，获取更好的位姿估计性能。在复杂环境中目标存在遮挡和截断的情况下，文献[54]在二维图像的每个像素上都预测一个指向关键点的单位向量，使用RANSAC对关键点的位置进行投票，实现了利用物体在图像中可见部分的局部信息来检测关键点的目的，最后对2D关键点进行PnP求解和优化得到目标的位姿。为了提高场景中多个物体同时进行位姿估计的能力，文献[55]对提前定义好的关键点、每个点的分类、中心点偏移进行预测，应用聚类算法区分具有相同语义标签的不同实例，根据相同实例上的点投票并聚类出该物体的3D关键点，使用最小二乘法估算出目标的6D位姿参数。

基于回归的方法通过直接从输入图像映射到参数表示的位姿信息，是一种端到端的学习方法。文献[56]首次提出了一种用于直接6D目标位姿估计的卷积神经网络PoseCNN，目标的三维平移是通过定位图像中心和预测到相机的距离来估计的，三维旋转是通过回归到四元数表示来计算的，并且引入ShapeMatchLoss函数，使PoseCNN能够处理对称对象。为了将6D姿态问题转化为2D图像中坐标点检测问题，文献[57]提出了一种用于6D目标位姿估计的端到端深度网络YOLO-6D+，利用目标检测框架预

测目标的 2D 坐标点, 然后使用 PNP 算法估计 6D 位姿, 这样做可能导致原本在 2D 图像上较小的检测误差, 映射到 3D 空间中会变大, 但此思路仍然值得借鉴。针对高度混乱场景中的目标位姿估计, 文献[58]使用一个异构网络去分别处理输入 RGB-D 图像中的颜色和深度信息, 以一种稠密的像素级融合方式将 RGB 数据的特征和点云的特征以一种更合适的方式进行了整合, 通过端到端的网络训练出位姿估计模型。为了处理给定类别中不同的和从未见过的物体实例的位姿估计, 文献[59]引入标准化物体坐标空间 NOCS (Normalized Object Coordinate Space) 将同一个类别中的所有物体实例使用一个共享的标准模型来表示, 通过训练神经网络来推断观察的像素与共享标准模型的对应关系和其他信息, 将预测图像与深度图相结合, 共同估计杂乱场景中多个物体的 6D 位姿和尺寸。基于回归的方法需要大量地标注位姿的训练数据, 对数据集中不存在的目标泛化性能较差。

3 无序分拣中的抓取决策

目标的检测识别和位姿估计过程为机器人的无序分拣提供了视觉上的保障, 但在抓取和放置目标的过程中还需考虑机械臂的运动规划, 用以规避障碍物和优化分拣时间来提高分拣过程的安全性及效率。文献[60]基于 RRT 算法构建一个可达和无碰撞配置树, 结合寻找可行的抓取、求解逆运动学和搜索无碰撞轨迹, 生成潜在抓握位置, 并计算接近该位置的动作, 使机械臂达到抓取姿态。但是 RRT 是概率完备且不是最优的, 对环境类型不敏感, 当机器人操作空间中包含大量障碍物或狭窄通道约束时, 算法的收敛速度慢, 效率会大幅下降。为了提高规划速度, 文献[61]采用人工势场法引导随机树向目标方向生长, 减少 RRT 算法在路径搜索过程中的盲目性。这种双向的 RRT 技术具有良好的搜索特性, 比原始 RRT 算法的搜索速度、搜索效率都有显著的提高, 且被广泛应用。

随着现有机器人抓取技术应用场景的不断扩展, 如何在无目标精确模型的情况下实现机器人抓取行为的通用性和在与环境交互中提高机器人的自适应学习抓取姿态的能力, 对机器人无序分拣的抓取规划具有重要的现实意义。基于深度强化学习的算法为机械臂开辟了新的发展方向, 深度强化学习兼具深度学习的神经网络图像处理能力, 并具备强化学习的策略自主学习能力, 由于深度强化学习的控制过程不依赖于精确的环境模型, 而强调在与环境的交互中获取反映真实目标达成度的反馈信号, 有利于提高机器人抓取过程中的自适应控制能力。文献[62]提出了一种基于强化学习的改进 DDPG (Deep Deterministic Policy Gradient) 算法对煤矸石进行分拣, 根据相应传感器返回的煤矸石位置及机械臂状态进行决策, 并向相应运

动控制器输出一组关节角状态控制量, 根据煤矸石位置及关节角状态控制量控制机械臂运动, 比传统 DDPG 算法更快收敛于探索过程中所遇最大奖励值, 策略泛化性更好。文献[63]采用非策略强化学习框架来学习抓取策略, 使用双深度 Q 学习框架和 GraspQ-Network 用于学习抓取的概率, 提高抓取成功率。但深度强化学习的方法存在探索效率低, 训练时间长等问题, 尝试将前面建立的目标检测和位姿估计算法与强化学习算法进行融合, 对机械臂的探索范围和接近目标后的抓取位姿进行限制, 提高训练过程中的探索效率和最终抓取的成功率, 从而提高训练的效率。

此外, 复杂场景下多目标的抓取规划还需考虑到最优抓取策略的问题, 保证机器人抓取过程中不会发生干涉和碰撞。目前主要的策略是将散乱堆叠的目标分开, 根据各个目标的位姿关系设定评价指标评估各个目标的可操作性, 以此确定各个目标的抓取顺序。文献[64]构建分割后物体三维点云的抓取空间和物体空间, 通过分析实际抓取空间和物体空间的关系得到目标的抓取可操作性, 根据抓取评价函数计算所有抓取方案下的目标可操作性评价价值, 以高评价价值对应的抓取方案作为机器人对场景中多目标物的最佳抓取策略。该方法有效解决了场景中障碍物存在障碍物的多目标物抓取规划问题, 并提高机器人对多目标场景的处理效率。文献[65]通过精细分割散乱堆叠目标的点云数据, 基于位置密度函数确定上层目标的抓取得分, 根据得分高低从上至下分拣目标。文献[66]基于深度强化学习建立抓取作业的马尔可夫决策过程, 通过深度 Q 网络完成图像信息与机械臂抓取动作之间的映射, 根据 Q 值的大小确定机械臂的最佳抓取位置, 使机械臂能够自主完成抓取作业。由于使用了放射性区域回报奖励和经验优先级采样方式的应用, 使机械臂优先抓取放射性区域内具有高放射性活度的物块, 同时增加了机械臂抓取无序紧密摆放物体时的效率。

在一些更加复杂的情况下, 目标之间紧密连接甚至需要抓取的目标被其他障碍物遮挡, 此时可以采用联合的推动与抓取策略在复杂环境中对目标物体进行抓取。文献[67]将大规模分布式优化与一个新型拟合深度 Q 学习算法相结合, 从之前收集的数据完全离线训练出一个模型, 再将该模型部署到真正的机器人上进行微调, 机器人根据最近的观察不断更新其抓取策略。该抓取策略能够应对一组紧密贴合目标的分开和抓取情况, 并且能够泛化到不同种类的物体抓取上, 即使这些物体并没有在训练时遇到过。文献[68]实现了在机器人操作中抓取原本不可见的目标物体, 通过深度 Q 学习将视觉观察和目标信息映射到推动和抓取的预期未来奖励中, 将原问题划分为推寻和抓取 2 个子任务。基于贝叶斯理论的策略考虑了过去的行动经验, 执行推寻目标; 一旦找到目标, 基于分类器的策略配合面向目标的推送抓取, 进而实现对原本

不可见目标的有效抓取。为了更好地学习有效的推送和抓取策略,文献[69]将推送策略和抓取策略分别视为一个生成器和一个判别器,在抓取判别器的监督下对推送策略进行训练,从而使推送奖励更加密集,以更少的运动次数实现了最优的任务完成率和目标抓取成功率。为了更好地发挥推动和抓取之间的协同作用,文献[70]引入离散度指标描述物体在环境中的分布情况,根据目标的动作价值和分散程度设计协调机制,运用不同动作之间的协同作用,使抓握更加有效,大大提高机器人在目标密集杂乱环境下的抓取成功率,并且具有推广到新场景的能力。

4 结语

机器人的无序分拣广泛应用于家庭服务、工业制造和物流仓储等领域,近年来许多学者对此展开了研究,但由于场景的复杂性与目标的多样性,使这项技术目前仍处于初步发展阶段。而且大多数机器人分拣系统的分拣对象是单类目标物体或者相互分离的多类目标物体,对于非结构化场景中存在堆叠或遮挡的多类目标物体尚未有良好的解决方案。此外现有的抓取技术对已知物体的抓取效果较好,但很难将现有的抓取能力扩展到新的物体上。因此,机器人无序分拣系统应当考虑对新对象具有较强的泛化能力,更好地处理复杂环境中的对象,提高系统的鲁棒性。

未来机器人无序分拣技术的发展方向包括视觉感知、位姿估计和抓取决策几个方面。利用深度学习的方法训练更多的数据对象来增强机器人的视觉感知能力,同时解决不同物体相互作用所产生的数据不完整性问题,克服物体形状和纹理上的一些局限性。普遍来说目标的位姿估计依赖于标准的CAD模型,在未知模型的情况下,通过融合实例分割和目标检测等手段从二维图像或三维点云数据中获取位姿,同时提高同类物体的位姿估计能力,但是由于物体对称性带来的位姿估计难题还需进一步解决。结合视觉感知和位姿估计的结果对复杂环境下堆叠目标的抓取顺序、抓取路径进行决策,保证无序分拣过程中机械臂做到无碰撞无干涉地抓取,发挥深度强化学习在自适应学习抓取姿态中的优势,使机器人无序分拣技术更加智能和高效。

参考文献:

- [1] CORRELL N, BEKRIS K E, BERENSON D, et al. Analysis and Observations from the First Amazon Picking Challenge[J]. IEEE Transactions on Automation Science and Engineering, 2018, 15(1): 172-188.
- [2] HERNANDEZ C, BHARATHEESHA M, KO W, et al. Team Delft's Robot Winner of the Amazon Picking Challenge 2016[J]. Springer, 2017, 9776: 613-624.
- [3] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [4] MELLADO N, AIGER D, MITRA N J. Super 4PCS Fast Global Pointcloud Registration via Smart Indexing[J]. Computer Graphics Forum, 2014, 33(5): 205-215.
- [5] FUJITA M, DOMAE Y, NODA A, et al. What are the Important Technologies for Bin Picking? Technology Analysis of Robots in Competitions Based on a Set of Performance Metrics[J]. Advanced Robotics, 2020, 34(7-8): 560-574.
- [6] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [7] 李晓菲. 库柏特: 让工业机器人更聪明[J]. 支点, 2018(1): 78-81.
- [8] LI Xiao-fei. COBOT: Make Industrial Robots Smarter[J]. Pivot, 2018(1): 78-81.
- [9] CONNOLLY C. A New Integrated Robot Vision System from FANUC Robotics[J]. Industrial Robot: an International Journal, 2007, 34(2): 103-106.
- [10] RASOLZADEH B, BJÖRKMAN M, HUEBNER K, et al. An Active Vision System for Detecting, Fixating and Manipulating Objects in the Real World[J]. The International Journal of Robotics Research, 2010, 29(2/3): 133-154.
- [11] 爱尔森智能科技有限公司. 3D视觉引导连杆料框无序分拣上料[EB/OL]. (2016-05-16) [2021-11-25]. <https://www.alsontech.com/col.jsp?id=203>.
- [12] Alsontech Co., LTD. 3D vision guided connecting rod for unordered sorting[EB/OL]. (2016-05-16) [2022-03-02]. <https://www.alsontech.com/col.jsp?id=203>.
- [13] 北京阿丘科技有限公司. 混合拆垛工作站[EB/OL]. (2017-05-25) [2021-11-25]. <https://www.aqrose.com/aipick.html>.
- [14] Aqrose Technology Co., LTD. Hybrid palletizing workstation[EB/OL]. (2017-05-25) [2021-11-25]. <https://www.aqrose.com/aipick.html>.
- [15] 喻强, 庄志炜, 田威, 等. 基于单目视觉的散堆铆钉识别与定位技术[J/OL]. 计算机集成制造系统. (2021-11-01) [2021-11-25]. <https://kns.cnki.net/kcms/detail/11.5946.TP.20211029.1927.016.html>.
- [16] YU Q, ZHUANG Z W, TIAN W, et al. Identification and Localization of Scattered Rivets Based on Monocular Vision[J/OL]. Computer Integrated Manufacturing Systems (2021-11-01) [2021-11-25]. <https://kns.cnki.net/kcms/detail/11.5946.TP.20211029.1927.016.html>.
- [17] FURUKAWA H, SATO J, YAMADA T, et al. Grasping Position Detection Using Template Matching and Differential Evolution for Bulk Bolts[C]// IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society. Lisbon: IEEE, 2019: 5622-5627.

- [14] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [15] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded up Robust Features[C]// *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*. Berlin: Springer-Verlag, 2006: 58-67.
- [16] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An Efficient Alternative to SIFT or SURF[C]// *2011 International Conference on Computer Vision*. Barcelona: IEEE, 2011: 2564-2571.
- [17] 聂海涛, 龙科慧, 马军, 等. 采用改进尺度不变特征变换在多变背景下实现快速目标识别[J]. *光学精密工程*, 2015, 23(8): 2349-2356.
NIE Hai-tao, LONG Ke-hui, MA Jun, et al. Fast Object Recognition under Multiple Varying Background Using Improved SIFT Method[J]. *Optics and Precision Engineering*, 2015, 23(8): 2349-2356.
- [18] 彭刚, 廖金虎. 插件元器件分拣机器人系统设计与实现[J]. *华中科技大学学报(自然科学版)*, 2020, 48(1): 108-114.
PENG Gang, LIAO Jin-hu. Design and Realization of Sorting Robots System for Plug-in Components[J]. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2020, 48(1): 108-114.
- [19] RUSU R B, BLODOW N, MARTON Z C, et al. Aligning Point Cloud Views Using Persistent Feature Histograms[C]// *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Nice: IEEE, 2008: 3384-3391.
- [20] RUSU R B, BLODOW N, BEETZ M. Fast Point Feature Histograms (FPFH) for 3D Registration[C]// *2009 IEEE International Conference on Robotics and Automation*. Kobe: IEEE, 2009: 3212-3217.
- [21] SALTI S, TOMBARI F, DI STEFANO L. SHOT: Unique Signatures of Histograms for Surface and Texture Description[J]. *Computer Vision and Image Understanding*, 2014, 125: 251-264.
- [22] RUSU R B, BRADSKI G, THIBAUX R, et al. Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram[C]// *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Chinese Taipei: IEEE, 2010: 2155-2162.
- [23] ALDOMA A, TOMBARI F, RUSU R B, et al. OUR-CVFH – Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation[J]. *Lecture Notes in Computer Science*, 2012, 7476: 113-122.
- [24] WOHLKINGER W, VINCZE M. Ensemble of Shape Functions for 3D Object Classification[C]// *2011 IEEE International Conference on Robotics and Biomimetics*. Phuket: IEEE, 2011: 2987-2992.
- [25] KUO Hao-yuan, SU Hong-ren, LAI Shang-hong, et al. 3D Object Detection and Pose Estimation from Depth Image for Robotic Bin Picking[C]// *2014 IEEE International Conference on Automation Science and Engineering*. Chinese Taipei: IEEE, 2014: 1264-1269.
- [26] 邱焱. 基于 3D 目标识别的工业机器人无序分拣技术研究[D]. 西安: 西安理工大学, 2019.
QIU Yao. Research on Disordered Sorting Technology of Industrial Robot Based on 3D Target Recognition[D]. Xi'an: Xi'an University of Technology, 2019.
- [27] TAO Wu-yong, HUA Xiang-hong, YU Ke-gen, et al. A Pipeline for 3-D Object Recognition Based on Local Shape Description in Cluttered Scenes[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(1): 801-816.
- [28] 龚学健. 基于 RealSense 的散乱零件三维目标识别[D]. 哈尔滨: 哈尔滨工业大学, 2018.
GONG Xue-jian. 3D Object Recognition of Scattered Parts Based on RealSense[D]. Harbin: Harbin Institute of Technology, 2018.
- [29] TOMBARI F, DI STEFANO L. Object Recognition in 3D Scenes with Occlusions and Clutter by Hough Voting[C]// *2010 Fourth Pacific-Rim Symposium on Image and Video Technology*. Singapore City: IEEE, 2010: 349-355.
- [30] DROST B, ULRICH M, NAVAB N, et al. Model Globally, Match Locally: Efficient and Robust 3D Object Recognition[C]// *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco: IEEE, 2010: 998-1005.
- [31] EJIMA T, ENOKIDA S, KOUNO T, et al. 3D Object Recognition Based on the Reference Point Ensemble[C]// *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*. Lisbon: IEEE, 2014: 261-269.
- [32] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// *2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014: 580-587.
- [33] HE Kai-ming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// *2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017: 2980-2988.
- [34] JO H J, MIN C H, SONG J B. Bin Picking System Using Object Recognition Based on Automated Synthetic Dataset Generation[C]// *2018 15th International Conference on Ubiquitous Robots (UR)*. Honolulu: IEEE, 2018: 886-890.
- [35] SHIN H, HWANG H, YOON H, et al. Integration of Deep Learning-Based Object Recognition and Robot Manipulator for Grasping Objects[C]// *2019 16th International Conference on Ubiquitous Robots (UR)*. Jeju: IEEE, 2019: 174-178.
- [36] 王德明, 颜熠, 周光亮, 等. 基于实例分割网络与迭代优化方法的 3D 视觉分拣系统[J]. *机器人*, 2019, 41(5): 637-648.

- WANG De-ming, YAN Yi, ZHOU Guang-liang, et al. 3D Vision-Based Picking System with Instance Segmentation Network and Iterative Optimization Method[J]. Robot, 2019, 41(5): 637-648.
- [37] REDMON J, DIVVALA S, GIRSHICK R, et al. You only Look Once: Unified, Real-Time Object Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [38] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector[C]// Computer Vision – ECCV 2016. Cham: Springer, 2016: 21-37.
- [39] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [40] 伍锡如, 黄国明, 孙立宁. 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(6): 711-719.
WU Xi-ru, HUANG Guo-ming, SUN Li-ning. Fast Visual Identification and Location Algorithm for Industrial Sorting Robots Based on Deep Learning[J]. Robot, 2016, 38(6): 711-719.
- [41] PENG Jin-min, LIU Wen-yu, YOU Tong-fei, et al. Improved YOLO-V3 Workpiece Detection Method for Sorting[C]// 2020 5th International Conference on Robotics and Automation Engineering (ICRAE). Singapore City: IEEE, 2020: 70-75.
- [42] 田立勋, 刘雄飞, 张彩芹, 等. 基于深度学习的低成本堆叠物料定位系统[J]. 北京理工大学学报, 2020, 40(9): 963-969.
TIAN Li-xun, LIU Xiong-fei, ZHANG Cai-qin, et al. A Low-Cost Location System Based on Deep Learning for Stacked Materials Sorting[J]. Transactions of Beijing Institute of Technology, 2020, 40(9): 963-969.
- [43] DU Guo-guang, WANG Kai, LIAN Shi-guo, et al. Vision-Based Robotic Grasping from Object Localization, Object Pose Estimation to Grasp Estimation for Parallel Grippers: A Review[J]. Artificial Intelligence Review, 2021, 54(3): 1677-1734.
- [44] WU Cheng-hei, JIANG S Y, SONG Kai-tai. CAD-Based Pose Estimation for Random Bin-Picking of Multiple Objects Using a RGB-D Camera[C]// 2015 15th International Conference on Control, Automation and Systems (ICCAS). Busan: IEEE, 2015: 1645-1649.
- [45] TRAN V L, LIN H Y. 3D Object Detection and 6D Pose Estimation Using RGB-D Images and Mask R-CNN[C]// 2020 IEEE International Conference on Fuzzy Systems. Glasgow: IEEE, 2020: 1-6.
- [46] YAN W, XU Z H, ZHOU X F, et al. Fast Object Pose Estimation Using Adaptive Threshold for Bin-Picking[J]. IEEE Access, 2020(8): 63055-63064.
- [47] HINTERSTOISSER S, HOLZER S, CAGNIART C, et al. Multimodal Templates for Real-Time Detection of Texture-less Objects in Heavily Cluttered Scenes[C]// 2011 International Conference on Computer Vision. Barcelona: IEEE, 2011: 858-865.
- [48] HINTERSTOISSER S, LEPETIT V, ILIC S, et al. Model Based Training, Detection and Pose Estimation of Texture-less 3D Objects in Heavily Cluttered Scenes[C]// Computer Vision – ACCV 2012. Berlin: Springer, 2012: 548-562.
- [49] HINTERSTOISSER S, CAGNIART C, ILIC S, et al. Gradient Response Maps for Real-Time Detection of Textureless Objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(5): 876-888.
- [50] MIAO Chun-xiao, ZHONG Xun-gao, ZHONG Xun-yu, et al. Detection and Grasping of Texture-less Objects Based on 3D Template Matching[C]// 2021 40th Chinese Control Conference (CCC). Shanghai: IEEE, 2021: 3943-3948.
- [51] CHEN Shi-li, HONG Jie, LIU Xi-neng, et al. A Framework for 3D Object Detection and Pose Estimation in Unstructured Environment Using Single Shot Detector and Refined LineMOD Template Matching[C]// 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation. Zaragoza: IEEE, 2019: 499-504.
- [52] LIN H Y, LIANG S C, CHEN Yu-kai. Robotic Grasping with Multi-View Image Acquisition and Model-Based Pose Estimation[J]. IEEE Sensors Journal, 2021, 21(10): 11870-11878.
- [53] CHOI C, TAGUCHI Y, TUZEL O, et al. Voting-Based Pose Estimation for Robotic Assembly Using a 3D Sensor[C]// 2012 IEEE International Conference on Robotics and Automation. Saint Paul: IEEE, 2012: 1724-1731.
- [54] PENG Si-da, LIU Yuan, HUANG Qi-xing, et al. PVNet: Pixel-Wise Voting Network for 6DoF Pose Estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Angeles: IEEE, 2019: 4556-4565.
- [55] HE Yi-sheng, SUN Wei, HUANG Hai-bin, et al. PVN3D: A Deep Point-Wise 3D Keypoints Voting Network for 6DoF Pose Estimation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 11629-11638.
- [56] XIANG Yu, SCHMIDT T, NARAYANAN V, et al. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes[J/OL]. arXiv preprint, 2017(1): 3[2018-05-26]. <https://arxiv.org/abs/1711.00199>.
- [57] KANG Jia, LIU Wen-jun, TU Wen-zhe, et al. YOLO-6D+: Single Shot 6D Pose Estimation Using Privileged Silhouette Information[C]// 2020 International Conference on Image Processing and Robotics (ICIP). Negombo, Sri Lanka. IEEE, 2020: 1-6.

- [58] WANG Chen, XU Dan-fei, ZHU Yu-ke, et al. DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 3338-3347.
- [59] WANG He, SRIDHAR S, HUANG Jing-wei, et al. Normalized Object Coordinate Space for Category-Level 6D Object Pose and Size Estimation[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 2637-2646.
- [60] VAHRENKAMP N, DO M, ASFOUR T, et al. Integrated Grasp and Motion Planning[C]// 2010 IEEE International Conference on Robotics and Automation. Anchorage: IEEE, 2010: 2883-2888.
- [61] 马灼明, 朱笑笑, 孙明镜, 等. 面向物流分拣任务的自主抓取机器人系统[J]. 机械设计与研究, 2019, 35(6): 10-16.
MA Zhuo-ming, ZHU Xiao-xiao, SUN Ming-jing, et al. Autonomous Picking Robot System for Logistics Sorting Task[J]. Machine Design & Research, 2019, 35(6): 10-16.
- [62] 张永超, 于智伟, 丁丽林. 基于强化学习的煤矸石分拣机械臂智能控制算法研究[J]. 工矿自动化, 2021, 47(1): 36-42.
ZHANG Yong-chao, YU Zhi-wei, DING Li-lin. Research on Intelligent Control Algorithm of Coal Gangue Sorting Robot Arm Based on Reinforcement Learning[J]. Industry and Mine Automation, 2021, 47(1): 36-42.
- [63] JOSHI S, KUMRA S, SAHIN F. Robotic Grasping Using Deep Reinforcement Learning[C]// 2020 IEEE 16th International Conference on Automation Science and Engineering. Hong Kong: IEEE, 2020: 1461-1466.
- [64] 颜培清, 何炳蔚, 雷阿唐, 等. 基于深度信息的多目标抓取规划方法研究[J]. 电子测量与仪器学报, 2016, 30(9): 1342-1350.
YAN Pei-qing, HE Bing-wei, LEI A-tang, et al. Research on Grasp Planning of Multi-Objects Based on Depth Data[J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(9): 1342-1350.
- [65] 王宏远. 基于区域逆生长分割的堆叠混杂目标识别及抓取定位技术[D]. 杭州: 浙江大学, 2020.
WANG Hong-yuan. Research on Grasping Technology of Stacked Objects Based on Inward Region Segmentation[D]. Hangzhou: Zhejiang University, 2020.
- [66] 周祺杰, 刘满禄, 李新茂, 等. 基于深度强化学习的固体放射性废物抓取方法研究[J]. 计算机应用研究, 2020, 37(11): 3363-3367.
ZHOU Qi-jie, LIU Man-lu, LI Xin-mao, et al. Research on Solid Radioactive Waste Grasping Method Based on Deep Reinforcement Learning[J]. Application Research of Computers, 2020, 37(11): 3363-3367.
- [67] KALASHNIKOV D, IRPAN A, PASTOR P, et al. QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation[J/OL]. arXiv preprint, 2018(1):23[2018-11-28].
<https://arxiv.org/abs/1806.10293>.
- [68] YANG Yang, LIANG Heng-yue, CHOI C. A Deep Learning Approach to Grasping the Invisible[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 2232-2239.
- [69] XU Ke-chun, YU Hong-xiang, LAI Qia-nen, et al. Efficient Learning of Goal-Oriented Push-Grasping Synergy in Clutter[J]. IEEE Robotics and Automation Letters, 2021, 6(4): 6337-6344.
- [70] LU Ning, LU Tao, CAI Ying-hao, et al. Active Pushing for Better Grasping in Dense Clutter with Deep Reinforcement Learning[C]// 2020 Chinese Automation Congress (CAC). Shanghai: IEEE, 2020: 1657-1663.

责任编辑: 陈作