



机器人抓取检测技术的研究现状

刘亚欣, 王斯瑶, 姚玉峰, 杨熹, 钟鸣

引用本文:

刘亚欣, 王斯瑶, 姚玉峰, 等. 机器人抓取检测技术的研究现状[J]. 控制与决策, 2020, 35(12): 2817–2828.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1145>

您可能感兴趣的其他文章

Articles you may be interested in

一种高匹配性的多层代价地图生成算法

A high matching layered costmap generation algorithm

控制与决策. 2020, 35(12): 2883–2888 <https://doi.org/10.13195/j.kzyjc.2018.0721>

凸优化与A*算法结合的路径避障算法

Convex optimization and A-star algorithm combined path planning and obstacle avoidance algorithm

控制与决策. 2020, 35(12): 2907–2914 <https://doi.org/10.13195/j.kzyjc.2019.0351>

基于免疫优化的平面Acrobot线性自抗扰鲁棒镇定

Robust stabilization of planar Acrobot using linear active disturbance rejection control with immune optimization

控制与决策. 2020, 35(12): 3053–3058 <https://doi.org/10.13195/j.kzyjc.2019.0289>

参数未知的离散系统Q-学习优化状态估计与控制

Q-learning optimal state estimation and control for discrete systems with unknown parameters

控制与决策. 2020, 35(12): 2889–2897 <https://doi.org/10.13195/j.kzyjc.2019.0180>

基于姿态估计的实时跌倒检测算法

Real-time fall detection algorithm based on pose estimation

控制与决策. 2020, 35(11): 2761–2766 <https://doi.org/10.13195/j.kzyjc.2019.0382>

机器人抓取检测技术的研究现状

刘亚欣^{1,2}, 王斯瑶¹, 姚玉峰^{1,2}, 杨 熹¹, 钟 鸣^{1,2†}

(1. 哈尔滨工业大学(威海) 船舶与海洋工程学院, 山东 威海 264209;

2. 威海市机器人与智能装备产业研究院, 山东 威海 264209)

摘 要: 作为机器人在工厂、家居等环境中最常用的基础动作,机器人自主抓取有着广泛的应用前景,近十年来研究人员对其给予了较高的关注,然而,在非结构环境下任意物体任意姿态的准确抓取仍然是一项具有挑战性和复杂性的研究. 机器人抓取涉及 3 个主要方面:检测、规划和控制. 作为第 1 步,检测物体并生成抓取位姿是成功抓取的前提,有助于后续抓取路径的规划和整个抓取动作的实现. 鉴于此,以检测为主进行文献综述,从分析法和经验法两大方面介绍抓取检测技术,从是否具有抓取物体先验知识的角度出发,将经验法分成已知物体和未知物体的抓取,并详细描述未知物体抓取中每种分类所包含的典型抓取检测方法及其相关特点. 最后展望机器人抓取检测技术的发展方向,为相关研究提供一定的参考.

关键词: 机器人; 抓取检测; 经验法; 未知物体; 深度学习

中图分类号: TP181

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1145

引用格式: 刘亚欣,王斯瑶,姚玉峰,等. 机器人抓取检测技术的研究现状[J]. 控制与决策, 2020, 35 (12): 2817-2828.

Recent researches on robot autonomous grasp technology

LIU Ya-xin^{1,2}, WANG Si-yao¹, YAO Yu-feng^{1,2}, YANG Xi¹, ZHONG Ming^{1,2†}

(1. School of Naval Architecture and Ocean Engineering, Harbin Institute of Technology, Weihai 264209, China;

2. Industrial Research Institute of Robotics and Intelligent Equipment at Weihai, Weihai 264209, China)

Abstract: As the most common and basic robotic movement in environment such as factory and household, the autonomous grasping has various application prospect. Although researchers have paid higher attention to the robot autonomous grasping in recent decades, grasping arbitrary objects accurately with arbitrary poses in non-structural environment is still a challenging and complex study. The robot grasping involves three main aspects: sensing, planning and controlling. The first step is detecting the targeted object and generating the grasping pose. It is the precondition of the successful grasp that helps achieve the grasp path planning and the entire grasp movement. The literature review in this study is mainly based on sensing. Firstly, the grasping technology is introduced from two aspects: analytical method and empirical method, and the empirical method is divided into the grasping of known objects and unknown objects from the perspective of whether there is prior knowledge of grasping objects. Moreover, the typical grasping methods and their related features contained in each category of unknown object grasping is detailly described. Finally, the development direction of robotic grasp detection technology is prospected with the aim of providing references for the relevant researchers.

Keywords: robot; grasp detection; empirical method; unknown object; deep learning

0 引 言

近年来,机器人已广泛应用于工厂、家居环境中,并发挥着越来越重要的作用. 机器人自主抓取是机器人获取、移动、运输目标物体的前提,作为机器人最常用的基础动作,自主抓取有着广泛的应用前景. 在工业生产中,自主抓取技术的进步加速了机器人技术

的发展,并应用于装配、包装、分拣等应用中. 然而,对于服务机器人在非结构环境下任意对象的准确抓取仍然是一项具有挑战性和复杂性的工作. 机器人抓取主要涉及 3 个方面:检测、规划和控制. 作为第 1 步,检测物体并生成抓取位姿是成功抓取的重要组成部分,抓取位姿的准确检测有助于规划抓取路径和整

收稿日期: 2019-08-09; 修回日期: 2019-12-26.

基金项目: 国家重点研发计划项目(2018YFB1309400); 山东省重点研发计划项目(2016GGX101013); 威海市大学共建项目(2016DXGJMS04).

责任编辑: 方勇纯.

†通讯作者. E-mail: zhongming@hit.edu.cn.

个抓取动作的实现.

本文回顾了目前所采用的机器人抓取检测技术的研究成果,并进行了分类、综述和整合,对未来抓取检测技术的研究方向及趋势进行分析,为抓取检测技术的研究以及实际应用提供更多有价值的参考.

1 机器人抓取检测技术简介

在阅读大量文献的基础上,本文将机器人抓取检测技术分为两大类^[1]:一类是传统的分析法,也称为硬编码方法^[2],指的是基于物体几何、物理模型、运动学和力学分析的抓取检测^[1]问题;另一类是基于经验的抓取方法,也称为数据驱动法,通过模仿人类抓取策略进行抓取检测,避免了数学和物理模型的计算.分析法在确定抓取时需要满足一些运动学和动力学公式,因为需要满足条件数量较多,所以计算复杂.图1为分析法在进行抓取检测时所采用的一般策略^[3].首先,基于环境限制和机械手与物体模型进行抓取检测,抓取检测是指找到满足与抓取任务相关的一组候选抓取位姿的过程;然后,基于一定

的选择标准和任务导向对场景输出的一组抓取位姿进行选择,输出最佳抓取位姿;最后,传递给机器人执行抓取.实际应用时不一定包含图1中介绍的所有元素.图2就分析法所涉及的动力学、运动学、几何、结构关系进行了详细的展开说明.在阅读文献的过程中发现,因为计算困难,当下应用的分析法多数没有考虑到任务约束,实际情况大多是基于任务导向的;分析法还有一个主要缺点,它要求对象参数是已知的,因此分析法不适合在非结构环境下广泛使用.

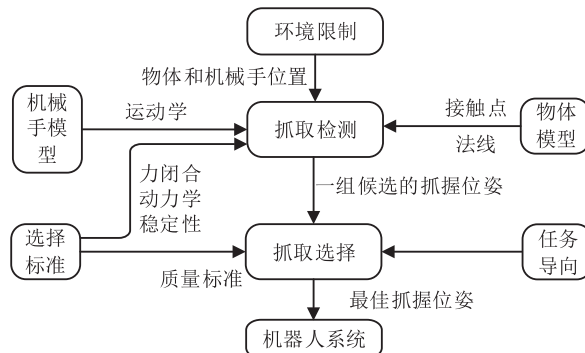


图1 使用分析法抓取检测的策略^[3]

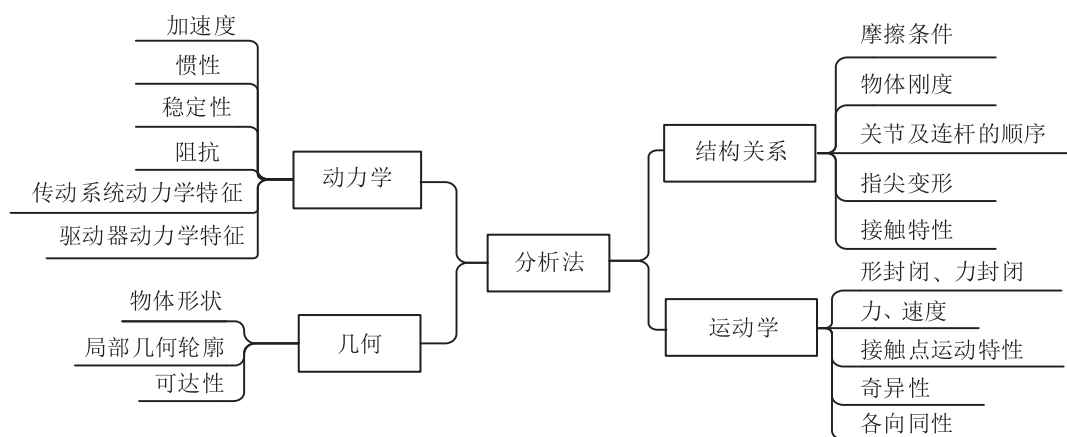


图2 分析法研究所涉及内容

为了避免上述缺点,研究人员将经验法引入到抓取问题上,经验法将为机器人提供更高的感知和适应能力,减少或完全消除手动建模的工作^[4].

经验法依赖于先前已知的成功抓取经验,基于一些度量标准对候选抓取位姿进行排序,参考基于分类和学习方法的技术,避免分析计算的复杂性,Kamon等^[5]称其为比较法,Shimoga^[6]称其为以知识为基础的方法.经验法的分类方式多种多样,按照算法将其分为学习式和启发式,如图3所示.学习式从人类示范、标记示例或反复试验中学习,启发式依赖于各种启发式方法将感知到的物体结构数据直接与候选抓取姿态比对.也可以按照感知信息的不同将经验法分为依赖物体模型的抓取姿态估计和不依赖物体模型的抓取姿态检测.此外,按照抓取对象的先验知识

又可以分类为已知物体的抓取检测和未知物体的抓取检测.这种分类方法更能捕捉到这些方法的多样性,下文将详细介绍.

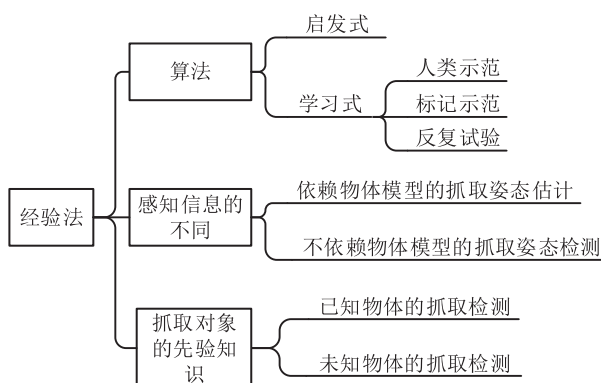


图3 经验法的分类方式

2 已知物体抓取检测技术研究现状分析

已知物体通常是指具有完整的几何三维模型并且有抓取姿态库的物体。机器人可以访问包含许多良好候选抓取位姿的几何对象模型数据库,此数据库通常是离线构建的,称为抓取经验数据库。图4为已知物体抓取的典型流程,在离线阶段,对物体模型数据库的每个物体进行处理,生成候选抓取位姿,基于一定的选择标准进行评分排序,将抓取位姿存入抓取经验数据库,与物体模型数据库进行一一对应,以便检索。在在线阶段,将场景进行分割,识别出物体,找到物体模型数据库中的物体与之对应,然后进行物体姿态估计,根据位姿检索找到抓取经验数据库中的抓取位姿,最后进行抓取选择和可达性过滤,执行抓取。

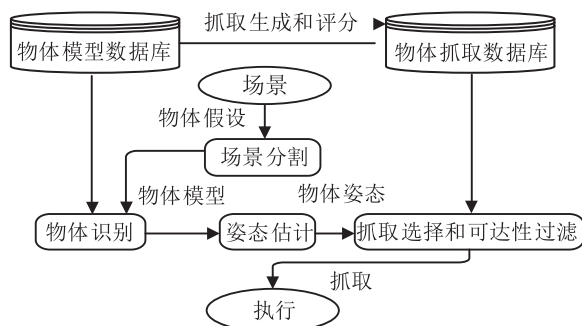


图4 已知物体抓取的典型流程

对于已知物体抓取,需要解决两大问题:一是如何建立已知物体的抓取经验数据库;二是如何进行待抓物体识别与位姿估计的问题。根据抓取经验数据库建立方法的不同,面向已知物体抓取的方法主要分为以下几类:

1) 基于三维网格模型抓取点分析的方法。此类别中许多方法都假设对象的三维网格模型可用^[7-18],该方法的难点是如何自动生成一组良好的抓取假设,涉及的关键技术有如何对可能存在合适手部抓取位姿的物体表面的无限空间进行采样、如何设定量化标准对生成的候选抓取位姿进行评分与排序等。

2) 基于示教的方法。产生抓取假设的另一种方法是观察人类如何抓住一个物体^[19-24]。机器人观察人类抓住已知物体,两个感知处理同时进行,一边识别对象,一边提取或识别所展示的抓取位姿,最后将对象模型与抓取模型存储在一起。

3) 基于试错的方法。这种方法不是对已知物体采用固定的一组抓取候选位姿,而是试图通过反复试验以改进它们^[19,23-25]。在这种情况下,离线学习和在线开发之间没有分离。分割和识别场景中的已知物体,检索该对象过去的抓取经验,并且在已经测试的

抓取经验库中生成或选择新的抓取假设。在执行所选择的抓取后,评估性能,并更新对象的经验库。

现实世界中,很难获得精确的物体三维模型,所以对未知物体的研究具有更高的研究价值,这里不详细讲解已知物体抓取,更多详细信息参见文献^[1]。

3 未知物体抓取检测技术研究现状分析

未知物体是指不可以访问物体模型、没有抓取经验的物体,本文将其分为基于感知的方法和基于学习的方法两大类。

3.1 基于感知的方法

在机器人抓取过程中,大多数情况下机器人只能感知到来自某些传感器的不完整信息,例如颜色信息或深度信息、触觉信息等,这使得抓取问题更具挑战性,因为机器人必须能够使用不完整信息来生成良好的抓取位姿。基于感知的方法专注于识别感知数据中的结构或特征、生成和评估排序候选抓取位姿,这通常基于传感器感知物体的局部或全局特征。图5为该方法常见的抓取检测流程。对于未知对象,该方法基于启发式将感知数据中的结构与候选抓取姿态直接链接起来,根据使用数据特征的不同将这类方法分为两种:一是提取分割后的点云或图片数据中的二维或三维特征,基于启发式进行抓取规划^[26-29];二是根据已经分割好的点云直接拟合或估计基本几何形状,进而根据该几何形状进行抓取规划^[30-31]。

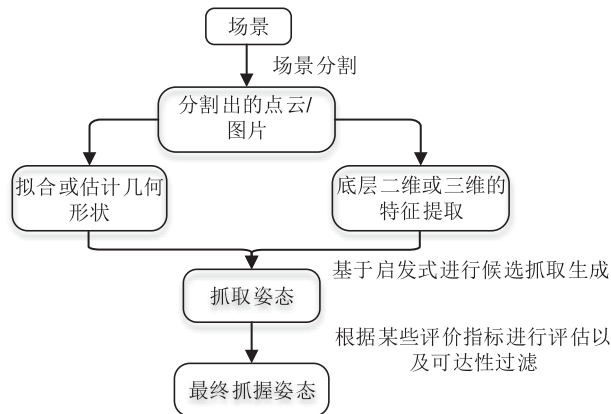


图5 基于感知方法对未知物体抓取检测的流程

3.2 基于学习的方法

机器学习方法已被证明对广泛的感知问题有效^[32-36],允许感知系统学习从一些特征集到各种视觉特性的映射^[37]。研究人员更是将深度学习引入抓取检测中,将学习方法应用于视觉中,引入学习方法对抓取质量进行评分^[38]。近期的文献采用了更丰富的特征和学习方法,允许机器人抓取部分被遮挡的已知物体^[39]或已知物体未知姿态^[40]以及系统之前未见过的全新物体(未知物体)^[41],本文将讨论后一种情

况. 早期的工作集中在仅从2D部分视图数据中检测单个抓取点,使用启发式方法基于此点确定机械手姿态^[41]. 由于只有二维信息,第三维度的距离必须是固定的,抓取具有局限性,而3D数据的使用显著改善了这些结果^[42]. 随着诸如Kinect等低成本RGB-D传感器的出现,深度数据在机器人抓取中的使用变得更加普遍.

使用学习的方法检测抓取可分为两大类,如图6所示:一是基于抓取检测的抓取方法(需要一个单独抓取规划控制系统),通过抓取检测方法生成抓取位姿,再使用单独的规划控制系统基于轨迹规划生成轨迹,实现完整的抓取;二是基于视觉运动控制策略的端到端抓取方法,实现从图像直接到抓取动作的映射. 其中,第1类按照学习内容不同又包含两种方

法,一种是学习抓取表示的结构化输出,例如抓取框等;另一种是学习抓取鲁棒性评估^[2]. 表1列出了基于学习的抓取检测方法的分类以及相关文献的检测准确率、速度等结果,并展示了文献所采用的数据集和算法,下文将紧紧围绕表1进行详细说明.

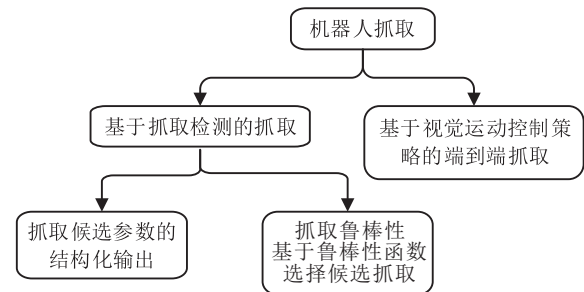


图6 基于学习的方法进行抓取的分类^[2]

表1 基于学习的抓取检测方法总结

类别	文献	检测准确率/%		检测速度 (fps) / ms	数据集	相机	算法
		image-wise	object-wise				
基于学习的抓取检测	[43]	60.50	58.30	0.02	Cornell	双目	快速搜索
	滑窗 [37]	73.90	75.60	0.07	Cornell	RGB-D	两级 SAE, SVM
	[44]	93.20	89.10	—	Cornell	RGB-D	ZF-net, 9 种预选框, hybrid network
	口法 [45]	94.10	93.30	—	Cornell	RGB-D	参考 LeNet-5 和 AlexNet 的网络结构, 空间金字塔池化方法
	[46]	93.80	91.30	平面物体 17.50	Cornell	RGB-D	R-FCN-4, angle-Net-6
	[47]	88.00	87.10	3.31	Cornell	RGB-D	AlexNet-5, 多抓取模型
	[48]	89.21	88.96	16.03	Cornell	RGB-D	ResNet-50*2, ReLU - RGB-D, ReLU, ReLU, 多模态
	结构化检测 [49]	成功率 65.60		—	Cornell	RGB-D	YOLOv3, 残差模块
	[50]	84.30		—	CIFAR-10, Cornell	RGB-D	ResNet101, mask RCNN, RPN
	输出法 [51]	抓取成功率 92.00		—	典型物体	RGB-D	Faster R-CNN
基于视觉运动控制策略的端到端抓取	[52]	97.00	95.80	23	Cornell	—	RetinaNet, SENet, 平衡金字塔
	[53]	96.00	96.10	8.33	Cornell	RGB-D	ResNet-50, 深度抓取, 9 种预选框
	[54]	97.74	96.61	28	Cornell	RGB-D	SSD, ResNet-101, 6 个默认方向
	[55]	93.00		1.25	Dex-Net	RGB-D	GQ-CNN 模型
	[56]	77.00		—	Bigbird	RGB-D	几何算法、力闭合分析等, CNN 分类学习
	鲁棒性评估 [57]	单物体 71.30 多物体 89.60		单物体 13.6 多物体 23.0	Cornell	RGB-D	STN 和 ResNet
	[58]	75.60		—	YCB, Bigbird	RGB-D	CNN 分类学习, 多指, 多通道
	[59]	88.90		4	Dex-Net, YCB	RGB-D	深度图像归一化算法
	[60]	到达目标点成功率 51		—	—	单目	DQN, 合成图像
基于视觉运动控制策略的端到端抓取	[61]	92.40 (抓取成功率 70)		—	Dex-Net	RGB-D	GQ-CNN, 迁移学习
	[62]	93.50		—	Dex-Net	RGB-D	吸力接触模型
	[63]	抓取和吸取精度均大于 90		60	—	RGB-D	抓取框架, 跨域图像分类框架, FCN

3.2.1 学习抓取表示的结构化输出

最常用方法是滑动窗口法,如图7所示。该方法将点取场景信息与机械手几何参数输入到卷积神经网络(convolutional neural networks, CNN)中,使用分类器确定图像每个小部分是否构成良好抓取,得分高的部分被认为是很好的候选抓取^[47]。

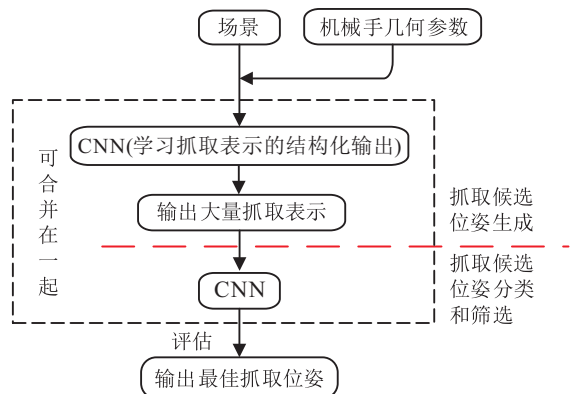


图7 学习抓取表示结构化输出的抓取检测流程

Saxena等^[41]通过监督学习进行训练,使用合成图像作为训练集,可在1.2s内识别单个抓取点,但没有使用深度图像,获得87.5%的抓取成功率。Rao等^[64]采用监督定位方法选择抓取点,估计形状后找到一对抓取点,同样获得了87.5%的抓取成功率。随后, Jiang等^[43]提出了七维抓取框的抓取表示,取代了原来的抓取点表示方法,明确地模拟了夹具的物理尺寸(即宽度),仅使用单个抓取点或一对抓取点无法表示该尺寸,而抓取矩形表示方法却能严格限制夹具的抓取边界。此外,抓取点表示方法具有任意的支撑区域(如围绕这些点的一定半径的区域),该支撑区域与抓取手指将占据的物理空间可能出现不匹配的情况。

Jiang等^[43]提出了一种多模态深度学习结构,首先进行无监督分层学习,用于RGB和深度模态的特征提取;然后通过组合RGB和深度特征开发共享层;最后将无监督分层学习用作监督特征分类器以进行最终决策。Lenz等^[37]在七维抓取框的基础上提出了五维抓取框,以此为抓取表示,基于二维图像提出了一个具有两级深度网络的级联系统。如图8所示,首先通过一个较浅的卷积神经网络找到所有可能的抓取矩形,保留一些评分较高的抓取矩形;然后通过一个深的卷积神经网络在保留的抓取矩形中找到评分最高的,即最佳抓取矩形框,得到最佳抓取矩形框后,将矩形框中心的点云法线方向作为机械手的接近向量,检测准确率达到75%,每个图像的处理时间为13.5s^[2]。该检测算法处理时间较慢,也仅依赖于局部特征确定抓取位置。由于采用类似于穷举法的搜索机制,需要在不同大小的图像块上使用分类器进行重

复计算,计算量非常大,且十分耗时。

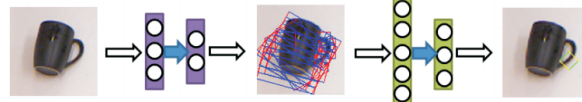


图8 两阶段检测过程

Guo等^[44]提出了一种视觉和触觉传感混合的深层结构,使用参考矩形方法识别图像的可抓取区域,该方法改编自区域候选网络(region proposal network, RPN)^[65],实质是基于滑窗的无类别物体检测器,检测准确率达到89.1%。喻群超等^[45]提出了一种3级串联卷积神经网络,第1级用来定位物体,为下一级卷积神经网络进行搜索抓取框做准备,第2级使用较浅的网络进行抓取框的搜索,去除掉不能实现的抓取框,第3级用一个深的卷积神经网络精确搜索抓取框,检测准确率达到93.3%,但是当物体背景颜色变化较大时,该方法不能很好地使用,鲁棒性不好。夏晶等^[46]为了快速检测平面物体的抓取姿态,提出了一种级联卷积神经网络,第1级基于R-FCN定位抓取框并进行了抓取框角度的粗估计,第2级使用angle-Net进行抓取框角度的精细估计,使得机器人抓取平均成功率达到92.5%。使用滑动窗口方法识别参考矩形位置的方法,以遍历搜索获得最优解,时间代价大。亟需解决的问题是减少抓取定位时间消耗和提升姿态估计精度。

鉴于滑动窗口法存在时间长的问题,一次检测法应运而生。该方法在图像上应用CNN直接预测抓取坐标,不用遍历搜索,该网络层数较多,但因为只对图像应用一次,实时性能有所提升。Redmon等^[47]利用大型卷积网络强大的学习能力,对物体的完整图像进行全局抓取预测。该网络在不使用标准滑动窗口或区域候选网络的情况下对可抓取的边界框直接执行单阶段回归,达到88%的准确性,在GPU上以每秒13帧的速度运行。该网络可以同时执行分类,以便在一个步骤中识别对象并找到一个良好的抓取矩形,对该模型的修改通过使用局部约束的预测机制来预测每个对象的多个抓取位姿。对于有多种方式抓取的对象,局部约束模型表现得更好。每张图像的检测准确率为84.4%,处理速度为76ms。因为该数据集没有对图像中的多个抓取进行适当的评估,所以局部预测模型不能对预测的每个图像的多个抓取能力进行定量评估。该方法使用类似于AlexNet的卷积神经网络模型实现单阶段检测,以更快的速度达到更高的检测精度,但由于卷积神经网络结构的复杂性仍然存在模型较大的缺陷。Kumra等^[48]使用深度卷积神经网络从

图片中提取特征,并使用浅卷积神经网络预测感兴趣物体的抓取位姿,该多模态模型在标准Cornell Grasp数据集上的检测准确率达到89.21%,每个图像的处理速度达到100 ms,仅在RGB图像上训练并且不具有深度特定特征,其采用网络结构更复杂、特征提取能力更强的ResNet50提取抓取特征,用SVM预测抓取配置的参数,虽然可以达到比较好的检测精度,但是由于模型采用层数较深的残差网络,导致网络模型和计算量都比较大.可以看出,一次性检测中的大部分工作都采用深度迁移学习技术来使用预先训练的神经网络架构^[66-67].

我国一些研究人员也采用该方法进行研究.张凯宇^[49]提出了基于卷积神经网络的栅格化抓取位置检测方法,使用YOLOv3算法以及残差模块,抓取成功率为65.6%,但当前模型对物体点云信息利用得不够充分,使用的训练数据量也较小,模型的准确性还有较大提升空间.唐博恒^[50]提出了一个两级抓取网络,第1级基于Mask R-CNN进行目标分类和生成掩膜图,第2级网络使用深度、灰度、掩膜图进行抓取预测.该网络最终得到84.3%的泛化性实验准确率,但在小目标物体抓取检测方面准确率不高.张亚辉^[51]基于Faster R-CNN提出一种多物体分类及摆放算法,旨在解决工业生产问题,该抓取系统的总成功率为92%左右,但没有考虑到物体的姿态、避障等功能.卢智亮等^[52]提出一种基于嵌入通道注意力结构SENet的一阶抓取检测网络(squeeze and excitation networks-RetinaNet used for grasp, SE-RetinaGrasp)模型的机器人抓取区域实时检测方法,该方法采用快速的一阶目标检测模型RetinaNet作为基本结构,在其特征提取网络中嵌入通道注意力模块SENet以提升重要特征通道的权重,确保检测精度,而且结合平衡特征金字塔(balance feature pyramid)思想,充分融合高低层的特征信息,加强检测小抓取框的能力,每张图片与每个物体的检测准确率为97.0%与95.8%,时间为23 ms,但未考虑多物体堆叠的情况.

尽管一次检测的首选方法是抓取表示的直接回归,但是在很多情况下,分类和回归技术被用于一次检测中,将方向预测视为分类问题^[2].一部分研究人员在一次检测法的基础上,细致地研究了抓取框的方向.Chu等^[53]将回归问题转换为区域检测和方向分类问题(零假设竞争)的组合,使用具有红色、绿色、蓝色和深度(RGB-D)图像输入的深度神经网络预测单个对象或多个对象的多个抓取候选者,在单个对象处理时,该方法优于Cornell数据集的最先进

方法,分别在图像和对象分割上具有96.0%和96.1%的检测准确率.对多对象数据集的评估表明了该体系结构的泛化能力.抓取实验在一组家用物体上实现了96.0%的抓取定位和89.0%的抓取成功率.从图像中的抓取检测到抓取动作规划,实时过程花费不到0.25 s,该方法在多对象多抓取时存在改进空间,可以结合先验知识提高检测准确率.该模型适用于多物体的抓取场景,并且达到了较高的抓取检测准确率,因为模型较深且类似于级联系统会导致模型较大.Zhou等^[54,68]在Guo等^[44]的基础上,提出了一种定向锚箱机构模型,并在训练过程中采用端到端的完全卷积神经网络.网络由两部分组成:特征提取器和多抓取预测器.特征提取器是深度卷积神经网络.多抓取预测器从预定义的定向矩形(称为定向锚框)中回归抓取矩形,并将矩形分类为可抓取和不可抓取.在标准的Cornell Grasp数据集中,该模型分别在图像分割和逐个对象分割上实现了97.74%和96.61%的检测准确率,该方法没有在杂乱环境下进行.可见,一次检测方法在速度以及检测准确率方面具有更好的性能,但是网络结构较为复杂.

3.2.2 学习抓取鲁棒性评估

学习抓取鲁棒性评估也是许多深度抓取检测研究的核心思想.研究人员使用此功能来识别具有最高评分的抓取位姿作为输出,该方法主要评估抓取而不学习控制器如何抓住物体.抓取鲁棒性描述了某个位置或图像区域的抓取概率^[55],对于这种方法,二元分类是一种很好的研究方法,它将候选抓取位姿分类为有效或无效(1或0).方法流程见图9,采用几何等方法生成候选抓取,再通过卷积神经网络基于鲁棒函数进行学习评估,输出最佳抓取.

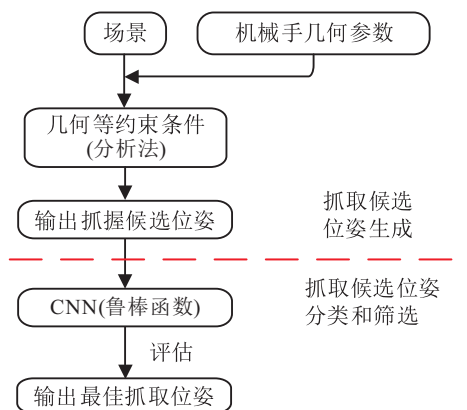


图9 学习抓取鲁棒性评估的抓取检测流程

Mahler等^[55]提出了一种学习最佳抓取鲁棒性函数的方法,对抓取点表示方法进行了更深一步的研究,将鲁棒性视为 $[0, 1]$ 范围内的标量概率,并创造了

Dex-Net 2.0数据集,其中包括670万点云和分析抓取质量指标,针对2指平行夹爪,使用对1500个3D对象网格模型的数据集进行稳健的准静态平行夹爪空间分析,以每个3D模型表面法向量为接近向量随机生成大量抓取位姿,然后将生成的抓取位姿与点云对应存储,投入抓取质量网络(grasp quality convolution neural network, GQ-CNN)进行训练. 利用Dex-Net 2.0数据集测试了CNN,抓取的检测准确率达到93%,但需要进行大量计算工作. Ten等^[56]对提取到的点云信息进行处理优化,基于几何信息计算曲率、法线等,生成大量六自由度抓取候选位姿(图10显示了基于该算法生成候选抓取位姿的3个示例),继而基于深度学习法对大量六自由度抓取候选位姿进行分类,根据机器人可以轻松达到的程度选择最佳抓取位姿,以识别密集物体杂乱中的可抓取区域. 该方法针对物体点云数据能达到77%的检测准确率,主动点云可以达到93%抓取成功率,此种方法在标记时必须记录至少两个点云.

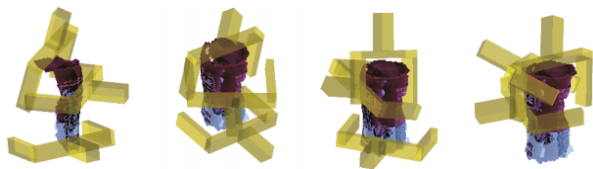


图10 生成候选抓取位姿

Park等^[57]使用基于原始空间变换网络(spatial transformer networks, STN)^[69]和深度残余网络(deep residual learning for image recognition, ResNet)^[70]的多级空间变换网络预测抓取候选,与其他基于回归的抓取检测方法不同,该方法允许部分观察中间结果,如抓取位置和方向. 每张图像的抓取检测准确率为89.60%,处理速度为23 ms. Lu等^[58]使用了反向传播算法在神经网络内部使用梯度上升优化有效地进行抓取概率研究,以使新物体的检测准确率达到75.6%. 抓取检测算法使用解析的方法生成候选抓取姿态,需要大量计算工作. 王斌^[59]在Mahler等^[55]的基础上,设计了一种基于深度图像和深度学习的归一化算法,以更好地适应各种夹具宽度和物体的距离,并基于GQ-CNN和改进的Dex-Net 2.0结构,设计了改进的GQ-CNN,在仿真环境下单阶段抓取检测算法的抓取检测准确率达到88.9%,平均计算用时为4 ms,但在实际抓取中存在大量噪声,抓取效果不好,而且没有考虑到多物体、杂乱环境有遮挡、非水平背景下的抓取检测情况,有待进一步改进. 此类方法能够观察抓取检测的中间步骤,但计算复杂.

3.2.3 基于视觉运动控制策略的端到端抓取

近年来,人们探索了新的方法,可以不使用单独的抓取规划控制系统,而使用深度学习训练视觉运动控制器. 该学习器迭代地校正抓取点,直到在抓爪之间成功抓取物体为止.

Zhang等^[60]提出了一种使用强化学习^[59]确定将机器人末端执行器扩展到2D图像平面中某个点的动作的方法. 以Mnih等^[71]提出的深度Q网络(DQN)为基础,使用合成图像使其适应机器人系统,在测试中系统达到目标点的成功率为51%. 这些结果在很大程度上不能从合成场景应用到现实场景,领域适应性较差. 此项目使用单目相机进行视觉传感,作者指出使用立体摄像机或RGBD传感器应该会获得更好的结果. Mahler等^[61]提出了一种可以在杂乱环境下利用在合成数据上训练CNN挑选物体的方法. 使用离散时间部分可观察的马尔可夫决策过程进行建模,通过将迁移学习技术与之前的方法^[55]结合使用,实现了92.4%的检测准确率,以及70%的抓取成功率. 来自Dex-Net 2.1的GQ-CNN高平均精度得分表明,当模型不具有较高的置信度时,可以通过引入替代动作(如探测或使用吸盘)提高性能. Mahler等^[62]提出了一种吸力接触模型,该模型可以计算吸盘与局部目标表面之间的密封质量,并衡量吸力抓取抵抗外部重力扳手的能力. 在Dex-Net 3.0上训练的GQ-CNN在约552000个数据点的验证集上具有93.5%的检测准确率,对分类为“基本”(棱柱形或圆柱形)、“典型”(更复杂的几何)和“对抗性”(具有很少的吸点)的未知对象进行评估时,抓取成功率分别达到98%、82%和58%. 该方法将夹爪换成吸盘,方法新颖,抓取成功率在“基本”“典型”物体中具有较高的检测准确率,但是对于“对抗性”物体的抓取成功率有待改进.

Zeng等^[63]提出了一种能够在杂乱的环境中识别抓取物体的视觉控制方法. 首先使用一个与类别无关的抓取框架来完成从视觉观察到动作的映射;推断出4个不同的抓取原始动作负担的密集像素级概率图;然后以最高的承受能力执行操作,并使用跨域图像分类框架识别拾取的对象,该框架将观察到的图像与产品图像进行匹配. 由于产品图像可轻松用于各种对象(如从网上获取),该系统无需任何其他数据收集或重新训练. 检测准确率为88.6%,时间为0.06 fps,抓取和吸取精度均大于90%,其中抓取略好于吸取,此方法限于无障碍抓取或者吸取. 加强学习以对原始动作进行排序的策略是本方法值得探索的.

替代方法. 此类方法不需单独的抓取系统, 可以进行分类, 但速度较慢.

4 机器人抓取检测技术总结

当前机器人抓取检测技术可按图 11 进行分类, 主要分为分析法和经验法. 其中, 分析法在 2005 年之前应用较为广泛, 使用分析法仍然可以准确地从图像

中检测物体的抓取配置问题, 但是由于面向任务存在复杂性高以及依赖已知参数的缺点, 后来逐步被经验法所取代. 经验法可应用于针对已知、未知物体的抓取检测, 其中针对已知物体的抓取检测方法必须具有完整的三维模型, 导致在实际应用中具有较大局限. 因此, 针对未知物体的抓取检测方法成为当前的研究热点.

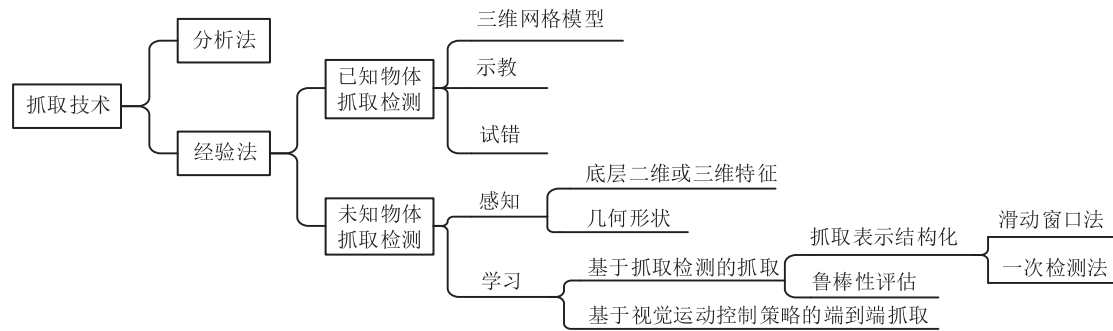


图 11 分析法研究所涉及内容

表 2 未知物体抓取检测技术总结

类别		物体特点	最优结果	特点	文献
感知	感知底层特征	某种形状	抓取成功率为 91.6 ^[28] ; 平均用时 0.2140 s ^[30]	只能抓取一部分物体	[26-31]
	感知形状	类别的未知物体		具有局限性, 计算复杂, 检测准确率低, 耗时	
学习	滑动窗口法	未知物体 (除文献 [46])	检测准确率: 图像分割法 94.1%, 物体分割法 93.3% ^[45] ; 速度 0.07 fps ^[47]	模型简单, 耗时, 训练较为耗时	[37, 43-46, 65]
	一次检测法	未知物体	检测准确率: 图像分割法 97.74%, 物体分割法 96.61% ^[54] ; 速度 16.03 fps ^[48]	计算简单, 模型较大, 训练较为耗时	[47-54]
	基于鲁棒函数学习对抓取表示的评估	未知物体	检测准确率: 图像分割法 93% ^[55] ; 速度 1.25fps ^[55]	计算复杂, 能够观察抓取检测的中间步骤	[55-56, 58-59, 70]
	基于视觉运动控制策略的端到端抓取	未知物体	检测准确率: 图像分割法 93.5% ^[61] ; 速度 60 ms ^[63]	不需单独的抓取系统, 可以进行分类, 速度较慢	[60-63]

未知物体抓取检测技术的分类及其特点如表 2 所示. 可见, 未知物体的抓取检测分为基于感知的抓取检测和基于学习的抓取检测. 基于感知的抓取检测没有结合现有的流行技术, 只针对特定形状物体, 且计算量较大, 检测准确率稍低, 耗时较长, 实际应用具有局限性. 基于学习的抓取检测方法目前研究较多, 不限于被抓物体形状, 其中基于抓取检测的抓取需要独立配置规划控制系统, 根据学习内容不同, 又可将其分为学习抓取表示的结构化输出和学习抓

取鲁棒性评估两种方法. 学习抓取表示的结构化输出方法不允许观察中间步骤, 其中的滑动窗口法多采用以 LeNet-5 和 AlexNet 为基础的网络模型, 结构简单, 检测准确率较好, 能泛化性抓取, 主要局限性在于以遍历搜索获得最优解, 不适合实时操作, 亟需解决的问题是减少抓取定位时间消耗, 以及提升姿态估计精度; 一次检测法能够实时进行检测, 但是采用的网络多以 ResNet 为基础, 检测准确率有所提高, 耗时较短, 但网络结构较为复杂, 需大型数据集进行训

练且较为耗时,亟需解决的问题是精简网络结构. 学习抓取鲁棒性评估的方法能够观察抓取检测的中间步骤,检测准确率较好,但计算复杂,大多没有给出检测时间,部分文献给出的时间也较长,理论上源于计算的复杂性,未来的研究方向可聚焦于提升检测速度方面. 一次检测法虽然训练时间较长,但是好在使用过程中速度快,而学习抓取鲁棒性评估的方法与其相反,但是大多情况下,期望能够在检测过程中提升速度. 与基于抓取检测的抓取不同的是基于视觉运动控制策略的端到端抓取,该方法不需独立配置规划控制系统,可以实现从图像直接到抓取动作的映射,能够用于分类但速度较慢,未来可提升速度并探索其在实际中的应用.

由前文可知,视觉传感方面使用RGB-D相机较多,相比于双目和单目相机,较为廉价,使用较为方便,且检测准确度较高. 建议研究学者们使用RGB-D相机,当然如果需要在室外进行抓取,双目相机也不失为一个好的选择. 同样可以看出,抓取点表示方法^[41,55,64]使用较少,抓取框的表示方法能进一步限制夹具的抓取情况,使用更为普遍.

总体而言,基于深度学习的抓取检测技术方兴未艾,正在逐步代替传统的人工设计、提取图像特征的方式. 目前常用的抓取检测方法主要有如下两种:

1) 基于一次检测法的抓取方法. 由前文可见,研究人员大多采用一次检测进行抓取检测,在图像上应用CNN直接预测抓取坐标. 相比于滑动窗口法,一次性检测法引入了更深的网络. 深度增加时,可表达更为复杂的模型,能从图像中学习更多特征,因此对物体的检测准确率显著上升. 但是也存在一些弊端:一是训练数据需求量大,较大的数据集将导致要提取的特征增加,而有限的数据集将导致过度拟合,同时也需要训练数据之间具有较大差异,即使是人工标注也存在较大难度;二是可训练参数的数量将随着深度的增加而增加,需要大量的计算和存储资源,对硬件的要求大幅提高. 图像检测准确率的提高依赖于更新的算法、改进的网络体系结构、强大的硬件支持、更大的训练数据集以及更大的学习模型,因此仍然具有挑战性. 未来可以尝试着设计特定于应用的深层卷积神经网络,使用轻量化模型,或者使用稀疏连接的深层网络以使识别速度更快,再在其中加入残差模块等以保持较高的检测准确率,使用网络预训练避免出现过拟合,此外,为了便于与国内外研究成果进行比较,Cornell数据集不失为一个好的选择. 相信随着科技的发展,硬件设备更好,训练可无需依赖NVIDIA

显卡,不只使用CUDA进行加速,亦可尝试使用其他框架进行训练,更加便于移植.

2) 基于视觉运动控制策略的端到端抓取方法. 通过使用深度学习训练视觉运动控制器,可以避免独立配置规划控制系统,实现从图像直接到抓取动作的映射. 近期已有研究人员开始探索将强化学习应用在该方法中,强化学习是一种基于奖励的方法,不需要预先标记数据集,但涉及到更长的训练过程,类似于反复试验过程,获得必要数量的有效数据实例所需时间较长,需要进一步的工作来确定更省时的方法. 迄今为止,研究人员较为回避此方法在实际产业中的应用,以减轻对机器人设备的潜在损害以及需要较长训练时间的问题. 然而,随着领域自适应技术的发展,可以在仿真环境下训练强化学习算法,再在实际产业中应用,避免对机器人的潜在损害,此类方法值得进一步探索. 譬如在工厂中,可按照实际的需要以及特定物体的特点,选用吸取或者抓取的方式分拣物体,此时,针对杂乱环境中物体的分层抓取以及语义理解,将成为研究的重点与难点.

5 结 语

随着机器人在工厂、家居等环境的快速普及,机器人抓取技术已经成为最近几年的研究热点,近年来我国有很多的研究人员和机构聚焦于抓取检测领域,但整体而言我国抓取检测技术成果与国外相比仍有较大差距. 然而,当前的国内外研究都存在一些共性问题:1) 高检测准确率下检测抓取框时间过长,不满足机器人抓取检测的实时性要求;2) 容易忽略目标物中可用于抓取的小部位信息,检测出来的抓取框偏大、精确度不足;3) 技术载体多为实验室样机,很多是在仿真环境下执行抓取,未能投入到生产生活中进行实际使用;4) 只有部分文献考虑到多物体、杂乱环境情况下的抓取,建议研究人员们可以从单一物体无遮挡情况下的检测逐渐转变为多物体杂乱环境下的抓取检测,同时也应关注小目标的检测以及抓取. 无论采用哪种方法,本文能够提供一定的参考价值,使得机器人抓取检测准确率更高,速度更快,进而提升抓取成功率,以增强机器人在复杂多变环境下的作业能力,满足实际生活、生产中的各种抓取任务需求,提高生活质量与生产效率.

参考文献(References)

- [1] Bohg J, Morales A, Asfour T, et al. Data-driven grasp synthesis — A survey[J]. IEEE Transactions on Robotics, 2014, 30(2): 289-309.
- [2] Caldera S, Rassau A, Chai D. Review of deep learning

- methods in robotic grasp detection[J]. *Multimodal Technologies and Interaction*, 2018, 2(3): 57.
- [3] Sahbani A, El-Khoury S, Bidaud P. An overview of 3D object grasp synthesis algorithms[J]. *Robotics and Autonomous Systems*, 2012, 60(3): 326-336.
- [4] Konidaris G, Kuindersma S, Grupen R, et al. Robot learning from demonstration by constructing skill trees[J]. *The International Journal of Robotics Research*, 2012, 31(3): 360-375.
- [5] Kamon I, Flash T, Edelman S. Learning to grasp using visual information[C]. *Proceedings of IEEE International Conference on Robotics and Automation*. Rehovot: Weizmann Science Press of Israel, 1996: 2470-2476.
- [6] Shimoga K. Robot grasp synthesis algorithms: A survey[J]. *The International Journal of Robotics Research*, 1996, 15(3): 230-266.
- [7] Morales A, Asfour T, Azad P, et al. Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Beijing: IEEE, 2006: 5663-5668.
- [8] Miller A T, Knoop S, Christensen H I, et al. Automatic grasp planning using shape primitives[C]. *Proceedings IEEE International Conference on Robotics and Automation*. Taipei: IEEE, 2003: 1824-1829.
- [9] Goldfeder C, Allen P K, Lackner C, et al. Grasp planning via decomposition trees[J]. *Proceedings IEEE International Conference on Robotics and Automation*. Roma: IEEE, 2007: 4679-4684.
- [10] Borst C, Fischer M, Hirzinger G. Grasping the dice by dicing the grasp[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Las Vegas: IEEE, 2003: 3692-3697.
- [11] Ciocarlie M T, Allen P K. Hand posture subspaces for dexterous robotic grasping[J]. *International Journal of Robotics Research*, 2009, 28(7): 851-867.
- [12] Rosen D, Paul R. Automated construction of robotic manipulation programs[M]. Pittsburgh: Robotics Institute Carnegie Mellon University, 2010: 1-263.
- [13] Balasubramanian R, Xu L, Brook P D, et al. Physical human interactive guidance: Identifying grasping principles from human-planned grasps[J]. *IEEE Transactions on Robotics*, 2014, 28(4): 899-910.
- [14] Przybylski M, Asfour T, Rüdiger Dillmann. Planning grasps for robotic hands using a novel object representation based on the medial axis transform[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco: IEEE, 2011: 1781-1788.
- [15] Roa M A, Argus M J, Leidner D, et al. Power grasp planning for anthropomorphic robot hands[C]. *IEEE International Conference on Robotics and Automation*. Saint Paul: IEEE, 2012: 563-569.
- [16] Huebner K, Welke K, Przybylski M, et al. Grasping known objects with humanoid robots: A box-based approach[C]. *International Conference on Advanced Robotics*. Munich: IEEE, 2009: 1-6.
- [17] Brook P, Ciocarlie M, Hsiao K. Collaborative grasp planning with multiple object representations[C]. *IEEE International Conference on Robotics and Automation*. Shanghai: IEEE, 2011: 2851-2858.
- [18] Tegin J, Ekvall S, Kragic D, et al. Demonstration-based learning and control for automatic grasping[J]. *Intelligent Service Robotics*, 2009, 2(1): 23-30.
- [19] Detry R, Baseski E, Popovic M, et al. Learning object-specific grasp affordance densities[C]. *Development and Learning*. Shanghai: IEEE, 2009: 1-7.
- [20] Balasubramanian R, Xu L, Brook P D, et al. Physical human interactive guidance: Identifying grasping principles from human-planned grasps[J]. *IEEE Transactions on Robotics*, 2012, 28(4): 899-910.
- [21] Pastor P, Righetti L, Kalakrishnan M, et al. Online movement adaptation based on previous sensor experiences[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco: IEEE, 2011: 365-371.
- [22] Romero J, Hedvig Kjellström, Kragic D. Modeling and evaluation of human-to-robot mapping of grasps[C]. *International Conference on Advanced Robotics*. Munich: IEEE, 2009: 1-6.
- [23] Kroemer O B, Detry R, Piater J, et al. Combining active learning and reactive control for robot grasping[J]. *Robotics and Autonomous Systems*, 2010, 58(9): 1105-1116.
- [24] Stulp F, Theodorou E, Kalakrishnan M, et al. Learning motion primitive goals for robust manipulation[C]. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco: IEEE, 2011.
- [25] Detry R, Kraft D, Buch A G, et al. Refining grasp affordance models by experience[J]. *IEEE International Conference on Robotics and Automation*, 2010, 5509126: 2287-2293.
- [26] Detry R, Henrik C, Marianna E, et al. Generalizing grasps across partly similar objects[C]. *IEEE International Conference on Robotics and Automation*. Saint Paul: IEEE, 2012: 3791-3797.
- [27] Kraft D, Pugeault N, Baseski E, et al. Birth of the object: Detection of objectness and extraction of object shape through object-action complexes[J]. *International Journal of Humanoid Robotics*, 2008, 5(2): 247-265.
- [28] Hsiao K, Chitta S, Ciocarlie M. Contact-reactive grasping

- of objects with partial shape information[C]. Proceedings of IEEE International Conference on Intelligent Robots and Systems. Taipei: IEEE, 2010: 1228-1235.
- [29] Klingbeil E, Rao D, Carpenter B, et al. Grasping with application to an autonomous checkout robot[C]. IEEE International Conference on Robotics and Automation. Shanghai: IEEE, 2011: 2837-2844.
- [30] Richtsfeld M, Vincze M. Grasping of unknown objects from a table top[J]. Workshop on Vision in Action: Efficient Strategies for Cognitive Agents in Complex Environments, 2008: 1-11.
- [31] Morales A, Sanz P J, Pobil A P D, et al. Vision-based three-finger grasp synthesis constrained by hand geometry[J]. Robotics and Autonomous Systems, 2006, 54(6): 496-512.
- [32] Viola P. Rapid object detection using a boosted cascade of simple features[C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai: IEEE, 2001: 511.
- [33] Hinton G E. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [34] Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]. Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM Press, 2009: 609-616.
- [35] Socher R, Huval B, Bhat B, et al. Convolutional-recursive deep learning for 3D object classification[C]. Proceedings of Advances in Neural Information Processing Systems (NIPS). Stanford: Lake Tahoe, 2012: 665-673.
- [36] Bo L, Ren X, Fox D. Unsupervised feature learning for RGB-D based object recognition[M]. Berlin: Springer Heidelberg, 2013: 387-402.
- [37] Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps[J]. International Journal of Robotics Research, 2015, 34(4/5): 705-724.
- [38] Kamon I, Flash T, Edelman S. Learning to grasp using visual information[M]. Rehovot: The Weizman Institute of Science, 1994: 7-14.
- [39] Glover J, Rus D, Roy N. Probabilistic models of object geometry for grasp planning[C]. Robotics: Science and Systems IV. Cambridge: Eidgenössische Technische Hochschule Zürich, 2008: 25-28.
- [40] Detry R, Baseski E, Popovic M, et al. Learning object-specific grasp affordance densities[C]. Development and Learning. Shanghai: IEEE, 2009: 1-7.
- [41] Saxena A, Driemeyer J, Kearns J, et al. Robotic grasping of novel objects[C]. Proceedings of the 2006 Conference on Advances in Neural Information Processing Systems. British Columbia: MIT Press, 2006: 1209-1216.
- [42] Saxena A, Wong L L S, Ng A Y. Learning grasp strategies with partial shape information[C]. National Conference on Artificial Intelligence. Chicago: AAAI Press, 2008: 1491-1495.
- [43] Jiang Y, Moseson S, Saxena A. Efficient grasping from RGBD images: Learning using a new rectangle representation[C]. IEEE International Conference on Robotics and Automation. Shanghai: IEEE, 2011: 9-13.
- [44] Guo D, Sun F, Liu H, et al. A hybrid deep architecture for robotic grasp detection[C]. IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 1609-1614.
- [45] 喻群超, 尚伟伟, 张驰. 基于三级卷积神经网络的物体抓取检测[J]. 机器人, 2018, 40(5): 762-768.
(Yu Q C, Shang W W, Zhang C. Object grab detecting based on three-level convolution neural network[J]. Robot, 2018, 40(5): 762-768.)
- [46] 夏晶, 钱堃, 马旭东, 等. 基于级联卷积神经网络的机器人平面抓取位姿快速检测[J]. 机器人, 2018, 40(6): 794-802.
(Xia J, Qian K, Ma X D, et al. Fast detection of robot plane grab position based on concatenated convolution neural network[J]. Robot, 2018, 40(6): 794-802.)
- [47] Redmon J, Angelova A. Real-time grasp detection using convolutional neural networks[DB/OL]. 2014, arXiv: 1412.3128.
- [48] Kumra S, Kanan C. Robotic grasp detection using deep convolutional neural networks[DB/OL]. 2016, arXiv: 1611.08036.
- [49] 张凯宇. 基于RGB-D图像的机械臂抓取位姿检测[D]. 杭州: 浙江大学机械工程学院, 2019: 1-89.
(Zhang K Y. Robotic grasp posture detection based on RGB-D image[D]. Hangzhou: School of Mechanical Engineering, Zhejiang University, 2019: 1-89.)
- [50] 唐博恒. 基于改进Mask RCNN的不规则3D物体抓取点识别[D]. 石家庄: 河北科技大学电气工程学院, 2019: 1-85.
(Tang B H. Grasping point recognition of irregular 3D objects based on improved mask RCNN[D]. Shijiazhuang: School of Electrical Engineering, Hebei University of Science and Technology, 2019: 1-85.)
- [51] 张亚辉. 基于Faster R-CNN目标检测的机器人抓取系统研究[D]. 深圳: 中国科学院大学深圳先进技术研究院, 2019: 1-65.
(Zhang Y H. Research on robot grasping system based on faster R-CNN target detection[D]. Shenzhen: Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences University, 2019: 1-65.)
- [52] 卢智亮, 林伟, 曾碧, 等. 机器人目标抓取区域实时检测方法[J]. 计算机工程与应用, 2019(9/10/11/12):

- 1-10.
(Lu Z L, Lin W, Zeng B, et al. Real-time detection method for robot target grabbing area[J]. Computer Engineering and Applications, 2019(9/10/11/12): 1-10.)
- [53] Chu F J, Ruinian X, Patricio V. Real-world multi-object, multi-grasp detection[J]. IEEE Robotics and Automation Letters, 2018, 1802: 1.
- [54] Zhou X W, Lan X, Zhang H, et al. Fully convolutional grasp detection network with oriented anchor box[DB/OL]. 2018, arXiv: 1803.02209.
- [55] Mahler J, Liang J, Niyaz S, et al. Dex-Net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics[C]. Robotics Science and Systems (RSS). Cambridge: MIT, 2017: 58.
- [56] Ten Pas A, Gualtieri M, Saenko K, et al. Grasp pose detection in point clouds[J]. The International Journal of Robotics Research, 2017, 36(13/14): 1455-1473.
- [57] Park D, Chun S Y. Classification based grasp detection using spatial transformer network[DB/OL]. 2018, arXiv: 1803.01356.
- [58] Lu Q, Chenna K, Sundaralingam B, et al. Planning multi-fingered grasps as probabilistic inference in a learned deep network[DB/OL]. 2017, arXiv: 1804.03289.
- [59] 王斌. 基于深度图像和深度学习的机器人抓取检测算法研究[D]. 杭州: 浙江大学机械工程学院, 2019: 1-85.
(Wang B. Research on robot grab detection algorithm based on depth image and deep learning[D]. Hangzhou: School of Mechanical Engineering, Zhejiang University, 2019: 1-85.)
- [60] Zhang F, Leitner J, Milford M, et al. Towards vision-based deep reinforcement learning for robotic motion control[C]. Proceedings of the Australasian Conference on Robotics and Automation. Canberra: IEEE, 2015: 2-4.
- [61] Mahler J, Goldberg K Y. Learning deep policies for robot bin picking by simulating robust grasping sequences[C]. Proceedings of the 1st Annual Conference on Robot Learning. California: Mountain View, 2017: 515-524.
- [62] Mahler J, Matl M, Liu X, et al. Dex-Net 3.0: Computing robust robot vacuum suction grasp targets in point clouds using a new analytic model and deep learning[DB/OL]. 2018, arXiv: 1709.06670.
- [63] Zeng A, Song S, Yu K, et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain Image matching[C]. Proceedings of the IEEE International Conference on Robots and Automation (ICRA). QLD: Brisbane, 2018: 21-26.
- [64] Rao D, Le Q, Phoka T, et al. Grasping novel objects with depth segmentation[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Taipei: IEEE, 2010: 2578-2585.
- [65] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington: IEEE, 2014: 580-587.
- [66] Basalla M, Ebert F, Tebner R, et al. Grasping for the real world(greifen mit deep learning)[DB/OL]. (2017-02-13)[2019-10-25]. <https://www.frederikebert.de/abgeschl-ossene-projekte/greifen-mit-deep-learning/>.
- [67] Watson J, Hughes J, Iida F. Real-world, real-time robotic grasping with convolutional neural networks[C]. Towards Autonomous Robotic Systems, Lecture Notes in Computer Science. Cham: Springer, 2017: 617-626.
- [68] 张翰博, 兰旭光, 周欣文, 等. 基于视觉推理的机器人多物体堆叠场景抓取方法[J]. 中国科学: 技术科学, 2018, 48(12): 1341-1356.
(Zhang H B, Lan X G, Zhou X W, et al. Robotic grasping multi-object stacking scenes based on visual reasoning[J]. Scientia Sinica: Technologica, 2018, 48(12): 1341-1356.)
- [69] Jaderberg M, Simonyan K, Zisserman A, et al. Spatial transformer networks[J]. Advances in Neural Information Processing Systems, 2015, 28(6): 2017-2025.
- [70] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [71] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

作者简介

刘亚欣(1981—), 女, 副教授, 博士, 从事液体微操作机器人技术、医疗检测微系统、医疗服务机器人等研究, E-mail: liuyaxin@hit.edu.cn;

王斯瑶(1996—), 女, 硕士, 从事医疗机器人的研究, E-mail: wangsiyao1996@126.com;

姚玉峰(1978—), 男, 教授, 博士生导师, 从事数字化医疗检测装备、医疗微系统及医疗机器人等研究, E-mail: yyf1023@163.com;

杨熹(1998—), 男, 本科生, 从事医疗机器人的研究, E-mail: yangxi_hit@163.com;

钟鸣(1981—), 男, 副教授, 博士, 从事医疗机器人、特种机器人、农业机器人等研究, E-mail: zhongming@hit.edu.cn.

(责任编辑: 郑晓蕾)