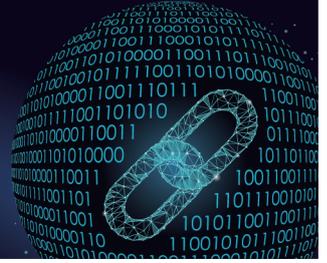




中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于语义分割与旋转目标检测的机器人抓取位姿估计

孙先涛, 闻勇, 陈文杰, 智亚丽, 胡祥涛, 陈伟海

引用本文:

孙先涛, 闻勇, 陈文杰, 智亚丽, 胡祥涛, 陈伟海. 基于语义分割与旋转目标检测的机器人抓取位姿估计[J]. 控制与决策, 2024, 39(9): 2913–2922.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0804>

您可能感兴趣的其他文章

Articles you may be interested in

[基于多层级特征的机械臂单阶段抓取位姿检测](#)

Single-stage grasp pose detection of manipulator based on multi-level features

控制与决策. 2021, 36(8): 1815–1824 <https://doi.org/10.13195/j.kzyjc.2019.1840>

[多目标小尺度车辆目标检测方法](#)

Multi-target and small-scale vehicle target detection method

控制与决策. 2021, 36(11): 2707–2712 <https://doi.org/10.13195/j.kzyjc.2020.0635>

[复杂背景下全景视频运动小目标检测算法](#)

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

[一种基于多层语义特征的图像理解方法](#)

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

[机器人抓取检测技术的研究现状](#)

Recent researches on robot autonomous grasp technology

控制与决策. 2020, 35(12): 2817–2828 <https://doi.org/10.13195/j.kzyjc.2019.1145>

基于语义分割与旋转目标检测的机器人抓取位姿估计

孙先涛¹, 闻勇¹, 陈文杰^{1†}, 智亚丽¹, 胡祥涛¹, 陈伟海²

(1. 安徽大学 电气工程与自动化学院, 合肥 230601;

2. 北京航空航天大学 自动化科学与电气工程学院, 北京 100191)

摘要: 针对姿态任意、尺寸不一的物体以及抓取角度离散性问题, 提出一种基于语义分割与旋转目标检测的单一位姿估计方法. 阶段 1: 首先利用 faster R-CNN (faster regions with convolutional neural networks features) 进行抓取检测获取候选抓取框; 其次利用语义分割网络获取待抓取物体的轮廓信息; 最后利用语义分割结果为每个待抓取物体筛选置信度最高的抓取框, 同时完成角度粗估计. 阶段 2: 利用旋转目标检测获取精细的抓取角度, 以修正阶段 1 抓取框的偏转角. 此外, 考虑到抓取物具有多尺度的特点, 提出一种多尺度特征融合模块, 使金字塔的所有层共享相似的语义特征. 针对智能算法求解逆运动学时出现迭代后期收敛速度慢的问题, 利用牛顿法收敛速度快的优点, 加快智能算法后期的收敛速度. 基于 V-REP 仿真以及实际抓取检测实验表明, 所提算法的抓取检测精度为 98.4%, 实际抓取成功率达到了 88.3%, 仿真抓取时的有害扭矩大小较修正前有所改善, 能够满足机械臂抓取要求.

关键词: 抓取检测; 语义分割; 特征金字塔; V-REP 仿真

中图分类号: TP242

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0804

引用格式: 孙先涛, 闻勇, 陈文杰, 等. 基于语义分割与旋转目标检测的机器人抓取位姿估计[J]. 控制与决策, 2024, 39(9): 2913-2922.

Robot grasping pose estimation based on semantic segmentation and rotating target detection

SUN Xian-tao¹, WEN Yong¹, CHEN Wen-jie^{1†}, ZHI Ya-li¹, HU Xiang-tao¹, CHEN Wei-hai²

(1. School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China; 2. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China)

Abstract: A monocular pose estimation method based on semantic segmentation and rotating target detection is proposed to solve the problem of arbitrary pose, different size objects and discrete grasping angles. In the first stage, faster R-CNN (faster regions with convolutional neural networks features) is used for grasping detection to obtain candidate grasping boxes. Then, the semantic segmentation network is used to obtain the contour information of the object to be captured. Finally, the semantic segmentation results are used to select the most reliable grasping box for each object to be grasped, and the coarse angle estimation is completed at the same time. In the second stage, the fine grasping angle is obtained by rotating target detection to correct the deflection angle of the grasping frame in the first stage. In addition, considering the multi-scale characteristics of the grab, a multi-scale feature fusion module is proposed to make all layers of the pyramid share similar semantic features. Aiming at the problem of slow convergence speed in the later stage of iteration when the intelligent algorithm solves the inverse kinematics, the advantage of fast convergence speed of the Newton method is used to speed up the convergence speed of the intelligent algorithm in the later stage. Based on V-REP simulation and actual grasping detection experiments, the grasping detection accuracy of the proposed algorithm is 98.4%, and the actual grasping success rate is 88.3%. The harmful torque during simulation grasping is improved compared with that before correction, which can meet the grasping requirements of the manipulator.

Keywords: grasp detection; semantic segmentation; feature pyramid; V-REP simulation

收稿日期: 2023-06-12; 录用日期: 2023-09-07.

基金项目: 国家自然科学基金项目(52005001).

责任编委: 瞿斌.

[†]通讯作者. E-mail: wjchen@ahu.edu.cn.

0 引言

近年来,机器人已广泛应用于工业、农业等场景中,并发挥着越来越重要的作用. 抓取作为机器人最常用的基础动作,是机器人完成取放任务的一项非常重要的能力. 由于非结构化环境和其他不确定性,对机器人而言,可靠地抓取物体仍然是一个巨大的挑战^[1]. 抓取任务不仅需要机器人准确地识别场景中的物体,还需要机器人准确地确定物体的位置和方向,以工业流水线上的牙刷或者长轴类零件的分拣机器人为例,视觉抓取技术能够帮助机器人准确地识别出物体的位置和方向,从而实现高精度的自动化操作.

对于机器人抓取,常用的方法是确定在点云上的抓取姿态^[2-3],但实际应用中易受物体遮挡等因素的影响. 传统的目标检测一般根据具体场景手工设计特征,对场景变化的适应性较差,检测精度不足^[4]. 深度学习不需要手工设计特征,可以有效地处理非结构化环境,因此已经成为机器人视觉传感任务的主流方法. 在一般的目标检测任务中,深度学习使用方法使用一个4维表示的水平矩形框来表示检测位置. Girshick^[5]提出了一种基于区域的快速卷积网络(fast regions with convolutional neural networks features, fast R-CNN)用于目标检测,然而使用选择性搜索候选框时会导致计算量变大. 为了解决这一问题,Ren等^[6]提出了faster R-CNN,使用区域候选网络与检测网络共享图像卷积特征,减少了计算量. 然而,4维矩形框由于缺乏抓取角度信息,不适合抓取检测任务. 基于两阶段的网络结构^[7]将5维表示的抓取矩形用于机器人抓取任务. 其中,Chu等^[8]提出了一种基于faster R-CNN框架的抓取检测方法,将抓取检测分为坐标回归和角度分类. 但是,在预测抓取角度时,将 180° 量化为若干等长区间,再用分类方法获取的抓取角度会使抓取角度过于离散,导致抓手与物体之间产生较大偏角从而产生有害扭矩,而较大的有害扭矩不仅损伤待抓取物体表面,还会使抓手产生疲劳损伤,降低抓手的使用寿命. 李秀智等^[9]提出了一种基于目标检测的最优抓取姿态检测算法,利用目标检测框从一系列抓取框中为抓取对象筛选出抓取框与目标检测框IOU最大的框作为最优抓取的框. 但是,利用目标检测框筛选出的抓取框的抓取位置可能不在待抓取物体上而导致后续抓取失败.

针对复杂背景的干扰,可以利用图像分割获取待抓取物体的轮廓信息. Morrison等^[10]提出了一种生成抓取卷积神经网络(generative grasping convolutional neural network, GG-CNN),实时预测每

一个像素点的抓取质量和抓取位姿. 该网络相对简单,精度较低. 此外,Araki等^[11]提出了一种基于语义分割的抓取检测网络,利用关键点抓取,但忽略了抓取角度.

从视觉检测和机器人抓取需求出发,本文提出一种基于语义分割与旋转目标检测的单目位姿估计方法. 阶段1,利用抓取检测和语义分割结果获取最优抓取位置的抓取框,同时完成角度粗估计. 阶段2,针对阶段1角度分类时的离散性问题,利用旋转目标检测获取精细的抓取角度,从而监督阶段1输出的抓取框的偏转角,当偏转角度处于可修正范围内时,修正阶段1抓取框的偏转角. 针对复杂背景的干扰,采用背景差分法思想,降低抓取检测时的复杂背景干扰. 此外,针对遗传算法求解逆运动学时,出现迭代后期收敛速度慢的问题,本文利用牛顿法收敛速度快的优点,将牛顿法应用于遗传算法的迭代后期,以加快后期机械臂逆运动学求解的收敛速度.

1 机械臂抓取方法描述

1.1 问题描述

目前,抓取检测的表示方法主要有两种. 一种是Jiang等^[12]提出的7维抓取表示,主要针对抓取三维物体;另一种是Lenz等^[13]提出的二维对象的5维抓取表示方法,该方法简化了7维抓取表示^[12],不仅能充分地描述夹持器的相对位置、方向和开口宽度,还可以降低计算成本. 5维抓取表示方法在抓取检测方法中得到了广泛的应用,因此本文使用的是5维抓取表示方法. 5维抓取表示方法如图1所示.

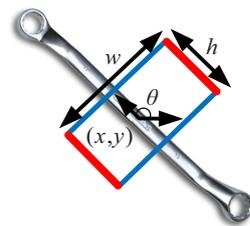


图1 5维抓取框

图1中: (x, y) 表示抓取框的中心坐标, (w, h) 表示抓取框的长度、宽度, θ 表示夹持器平面法线与抓取框水平轴的角度差. 5维抓取位姿的函数表示为

$$g = (x, y, w, h, \theta). \quad (1)$$

本文在抓取检测阶段利用回归方法和分类方法预测抓取矩形和旋转角度.

1.2 检测框映射到三维空间的抓取位姿

本文将RGB彩色图像中生成的二维抓取矩形作为机器人抓取姿态的参照. 其中,将检测框的偏转角

度 θ 作为机器人抓取时的角度,而检测框中心点的像素坐标要转化为机器人基坐标系下对应空间点的三维坐标值.

首先,由相机标定原理,计算出抓取框中心点的像素坐标在相机坐标系下的三维坐标值.设抓取框的中心点像素坐标为 (u, v) ,抓取框的中心点在相机坐标系下的坐标为 $({}^{\text{cam}}X, {}^{\text{cam}}Y, {}^{\text{cam}}Z)$,有

$$M_{\text{intri}} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

$${}^{\text{cam}}X = \frac{u - c_x}{f_x} \times {}^{\text{cam}}Z, \quad (3)$$

$${}^{\text{cam}}Y = \frac{v - c_y}{f_y} \times {}^{\text{cam}}Z. \quad (4)$$

式(2)中的 M_{intri} 为相机的内参矩阵,由相机标定获得.其中: f_x 和 f_y 分别表示相机在水平方向和垂直方向上的焦距, c_x 和 c_y 表示图像平面上的光学中心位置.式(3)和(4)中的 ${}^{\text{cam}}Z$ 为抓取点在相机坐标系下的 Z 轴坐标值,由抓取点在深度图像中对应的深度值转换所得.联立式(2)~(4)可以获得抓取中心在相机坐标系下的坐标.

其次,利用手眼标定原理,计算出抓取中心在机

器人基座坐标系下的坐标值 $({}^{\text{arm}}X, {}^{\text{arm}}Y, {}^{\text{arm}}Z)$,即机器人末端执行器抓取时的位置量. ${}^{\text{arm}}_{\text{cam}}T$ 是机械臂基坐标系到相机坐标系下的空间变换,表示为

$${}^{\text{arm}}_{\text{cam}}T = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}, \quad (5)$$

其中 R 和 T 为旋转矩阵与平移向量.最后,由下式计算出检测框中心点对应的三维空间点在机械臂基座坐标系下的坐标值:

$$\begin{bmatrix} {}^{\text{arm}}X \\ {}^{\text{arm}}Y \\ {}^{\text{arm}}Z \\ 1 \end{bmatrix} = {}^{\text{arm}}_{\text{cam}}T \begin{bmatrix} {}^{\text{cam}}X \\ {}^{\text{cam}}Y \\ {}^{\text{cam}}Z \\ 1 \end{bmatrix}. \quad (6)$$

2 网络模型

网络的整体架构可以分为两个阶段,如图2所示.阶段1:首先,利用抓取检测算法faster R-CNN^[6]获取初始抓取框;然后,利用获取的一系列初始抓取框和语义分割的结果,综合待抓取物体的轮廓信息和初始抓取框的几何信息,为每个待抓取对象筛选抓取物体上的位置,同时置信度最高的框作为初始抓取框,并完成角度粗估计.阶段2:利用角度预测分支获取精细的抓取角度,以监督阶段1抓取框的偏转角度.

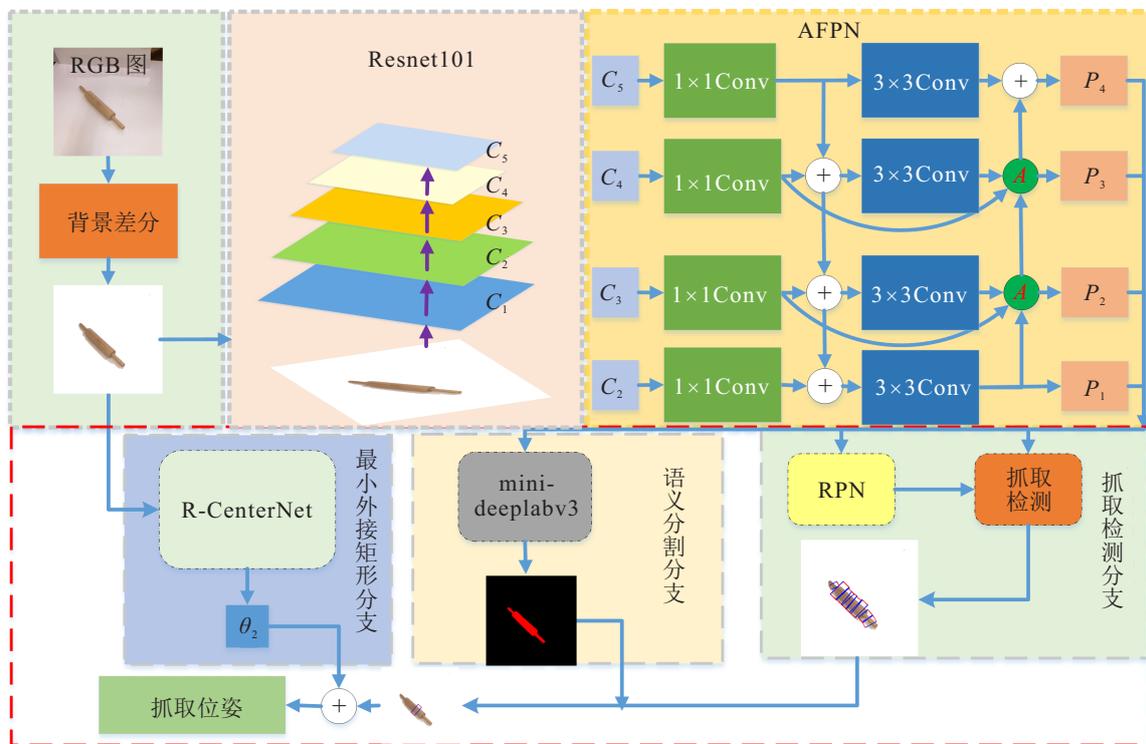


图2 抓取位姿估计网络结构

此外,在预测抓取位姿时,为了抑制复杂背景的干扰,可以采用基于RGB的背景差分法^[14]将物体从背景中分离出来,而对背景的抑制可能会在物体轮廓

内部引入噪声点,要用形态学开运算去除部分噪声点的影响.但是,此方法抑制不了场景中待抓取物体阴影的影响.

2.1 faster R-CNN 模型

本文采用 Resnet101^[15] 作为 faster R-CNN 模型中的主干网络以提取特征. 其中, faster R-CNN 模型由骨干网络、区域建议网络 (region proposal network, RPN) 和检测头部组成. 主干网络提取输入图像的特征并生成特征图 $C_1 \sim C_5$. 特征金字塔网络接收主干网络输出的特征图 $\{C_2, C_3, C_4, C_5\}$, 将不同层级的特征图进行特征融合生成 $P_1 \sim P_4$.

为了减小特征随着网络的加深而消失这一影响, 采用改进后的特征金字塔网络 (attention feature pyramid networks, AFPN), 它是一种基于注意力的自顶向下和自底向上的网络结构, 能够实现多尺度特征融合. 区域建议网络则利用 AFPN 产生的特征图生成一系列抓取候选框输出到抓取检测头部, 与文献 [8] 类似, 抓取检测头部对抓取框的偏转角度采用 18 分类, 而抓取框的中心点像素坐标和抓取框的宽高则利用回归操作以生成一系列抓取框.

2.2 特征金字塔模块

特征金字塔 (feature pyramid networks, FPN)^[16] 的低维特征图主要包含边缘线条等物体局部位置特征信息, 缺乏足够的语义信息; 高维特征图主要包含细节与轮廓等全局语义信息, 但位置信息比较粗糙, 通常丰富的高维语义信息能够获得更好的检测效果^[17]. 因此, FPN 将低分辨率高语义信息的高层特征和高分率低语义信息的低层特征进行融合, 使金字塔每个等级上的特征图都能包含丰富的语义信息.

对于抓取检测而言, 位置信息也很重要, 因此本文在 FPN 的基础上添加一个自底向上的结构, 在高维决策特征中加入低维特征, 从而获取物体更准确的位置信息. 本文提出的一种注意力特征金字塔网络 AFPN, 是一种基于自注意力机制的自顶向下和自底向上结构, AFPN 网络结构如图 2 所示, AFPN 中的 A 为余弦非局部注意力模块^[18]. 在 AFPN 中, 使用注意力模块和自底向上结构, 捕获不同尺度的精细化信息, 实现多尺度特征融合, 使金字塔的所有层共享相似的语义特征和精确的位置信息, 以提高抓取检测的精度.

2.3 语义分割模块

基于目标检测的抓取检测^[19] 并未考虑待抓取物体的轮廓信息, 只是利用目标检测框从一系列抓取框中筛选出一个与目标检测框有交集且置信度最高的抓取框作为最优抓取框, 而没有考虑抓取位置是否在物体上. 语义分割可以获取场景中待抓取物体的轮廓信息, 因此, 可以利用语义分割获取的物体轮廓信

息从一系列抓取框中筛选出物体上的抓取位置, 且置信度最高的框作为最优抓取位置的抓取框.

本文的语义分割网络结构如图 3 所示. 语义分割分支的具体过程分为 3 步: 首先, 利用语义分割模块接收 AFPN 的输出作为输入, 送入到 ASPP (atrous spatial pyramid pooling) 中^[20], 以获取工作台上所有待抓取物体的轮廓信息; 然后, 利用工作台上所有待抓取物体的轮廓信息从 faster R-CNN 生成的一系列抓取框中, 为每个待抓取物体筛选置信度最高的框作为初始抓取框, 同时完成角度粗估计.

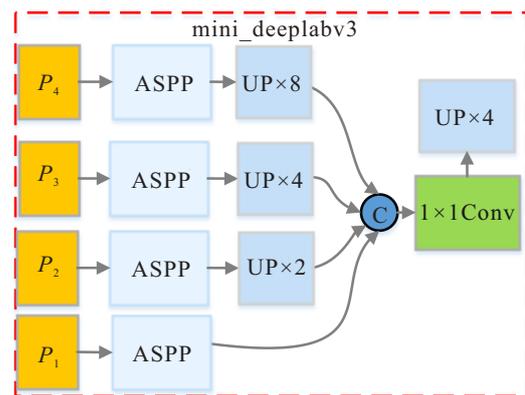


图3 语义分割网络

2.4 基于 R-CenterNet 的位姿估计模块

获取抓取位置后, 需进一步获取精细的抓取角度. 旋转目标检测算法 R-CenterNet^[21] 多用于遥感检测, 是针对方向及形状不确定性物体进行精确检测的算法, 其无需设置 anchor, 避免了阶段 1 输出的抓取框偏转角离散性问题. 因此, 为提高实际抓取的成功率, 可利用旋转目标检测输出的抓取角 θ_2 来监督阶段 1 输出的抓取框偏转角 θ_1 .

由旋转目标检测获取的旋转矩形框如图 4 所示, 本文先根据旋转目标检测框来求抓取框的监督角度 θ_2 , 再判断哪个角度为最佳抓取角度.

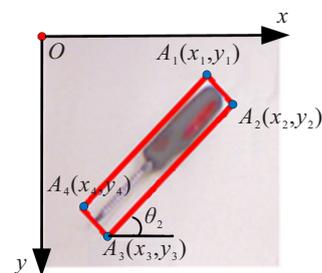


图4 旋转矩形框

首先, 由旋转矩形框 4 个顶点的坐标 A_1, A_2, A_3, A_4 , 求出旋转矩形框的任意相邻两条边的长度, 如下所示:

$$L_1 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \quad (7)$$

$$L_2 = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2}. \quad (8)$$

其次,对边长 L_1 、 L_2 值进行大小判断. 若 $L_2 > L_1$ 且 $x_2 \neq x_3$, 则 $\theta_2 = \alpha$; 若 $L_2 < L_1$ 且 $x_2 \neq x_3$, 则 $\theta_2 = \beta$.

$$\alpha = \frac{180^\circ}{\pi} \times \arctan \left| \frac{y_2 - y_3}{x_2 - x_3} \right|, \quad (9)$$

$$\beta = 90^\circ + \frac{180^\circ}{\pi} \times \arctan \left| \frac{y_2 - y_3}{x_2 - x_3} \right|. \quad (10)$$

最后,由下式确定机器人的最优抓取角度 θ :

$$\theta = \begin{cases} \theta_1, & \text{otherwise;} \\ \theta_2, & |\theta_2 - \theta_1| \leq T \cap (u, v) \subseteq S. \end{cases} \quad (11)$$

其中: T 为容错阈值,根据阶段1的角度粗分类间隔确定; (u, v) 为阶段1输出的抓取位置; S 为阶段2的旋转矩形框区域.

3 机械臂自主抓取仿真与实验

3.1 抓取数据集

实验采用的数据集是 Cornell 开源数据集,该数据集共有 240 个不同样本的 885 张图像,每张图像的分辨率为 640×480 ,专门为机器人抓取设计.但是,由于数据集的规模相对较小,训练时易发生过拟合现象.因此,为了避免过拟合,本文采用数据增强技术来扩大数据集.采用了旋转变换、翻转变换、平移变换、亮度以及对比度变换、颜色变换等方法增大数据集规模.

3.2 抓取检测网络对比实验

对 Cornell 数据集抓取检测的评估,前人主要使用了两种不同的指标,分别为点度量与矩形度量.点度量是计算从预测抓取框的中心到真实抓取框的中心的距离,若距离小于某个阈值,则认为成功抓取.但点度量没有考虑角度信息,因此本文使用矩形度量作为指标.当预测的矩形框满足以下两个条件时,该抓取框可用于抓取物体:

1) 预测的抓取角度与真值标定的抓取角度之差在 30° 以内.

2) 下式中定义的 Jaccard 系数大于 0.25:

$$J(g_p, g_t) = \frac{|g_p \cap g_t|}{|g_p \cup g_t|}. \quad (12)$$

其中: g_p 为预测的抓取框面积, g_t 为真实抓取框面积.

本文网络在 Pytorch 深度学习框架下实现抓取位姿检测任务的训练与测试,在 GPU 为 NVIDIA GeForce RTX 2080 Ti 上对网络进行训练,初始学习率设为 0.01,权值衰减设为 0.000 1,采用随机梯度下降 (stochastic gradient descent, SGD) 作为优化器,动量因子设为 0.9,批量设为 12.

将本文的抓取检测网络 (AFPN) 与近年来的抓取检测算法^[22-24] 进行比较.从表 1 可知:文献 [24] 的检测速度是最快的,但是检测的准确率低于本文算法 1.6%. 而对比文献 [22-23],本文算法的检测准确率分别高出了 4.8% 和 9.7%,而检测速度也较文献 [22-23] 快.此外,抓取检测网络采用改进后的特征金字塔网络 AFPN,较原始 FPN 的检测精度提升了 0.2%. 实验结果表明,在抓取检测网络以 RGB 图像作为输入时,本文算法在保证精度的同时,检测速率也能满足机器人抓取的要求.

表 1 在 Cornell 数据集上的抓取检测比较

方法	输入	准确率/%	时间/s
文献 [22]	RGB	93.6	0.039
文献 [23]	RGB	88.7	0.500
文献 [24]	RGB	96.8	0.024
本文 (FPN)	RGB	98.2	0.034
本文 (AFPN)	RGB	98.4	0.037

3.3 位姿估计与动力学仿真实验

图 5 是网络采用背景差分法前后的抓取检测结果对比,第 1 行是未采用差分法抑制背景而直接进行位姿估计的结果,第 2 行是采用差分法以抑制背景的位姿估计结果.在抑制背景的影响前,抓取检测会受到周围背景的干扰,产生的抓取框会覆盖更多的背景.抑制背景后,抓取检测受到背景的干扰小,产生的抓取框更贴近待抓取物体.

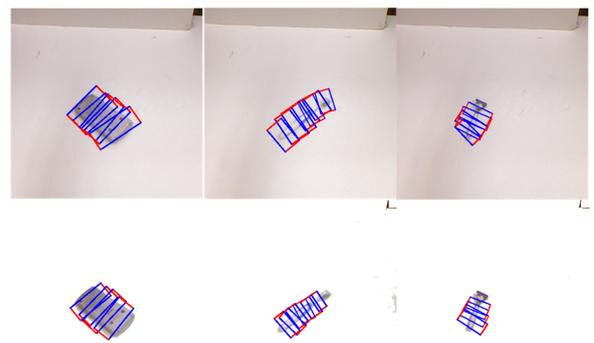


图 5 抑制背景前后的抓取检测

图 6 为语义分割获取初始抓取框.从左到右,第 1 列为抓取检测结果;第 2 列为语义分割的结果,每种颜色代表一个特定的类别;第 3 列为场景中待抓取物体最优抓取位置的抓取框,红线表示夹持器的平行板,蓝线表示开口宽度.利用语义分割获取的物体轮廓信息,从抓取检测输出的一系列抓取框中为场景中的待抓取物体分配一个物体上的抓取位置,同时置信度最高的框作为角度待修正的初始抓取框.

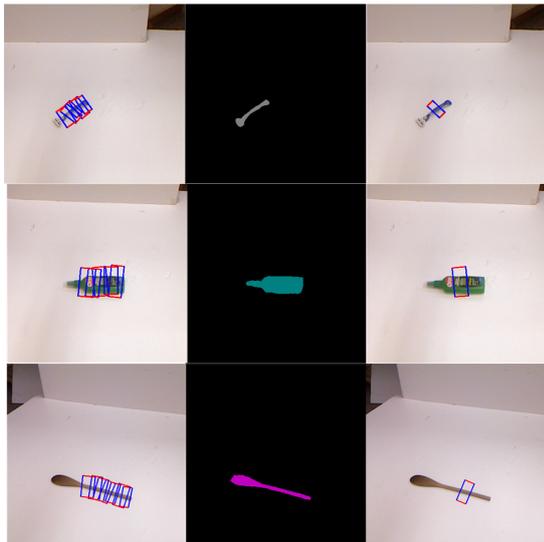


图6 语义分割获取初始抓取框

图7是网络添加角度预测模块前后的抓取检测结果,第1行是阶段1直接进行抓取预测的结果,第2行是添加角度预测模块来监督阶段1输出的抓取框偏转角后得到的结果.检测结果表明,添加角度预测分支以监督阶段1获取的抓取框偏转角时,最终获取的抓取框更贴近与物体轴线垂直,因此在抓取时不仅降低了抓手与物体的碰撞几率,也使抓手在抓取物体时更稳定,增加了抓取时的成功率.

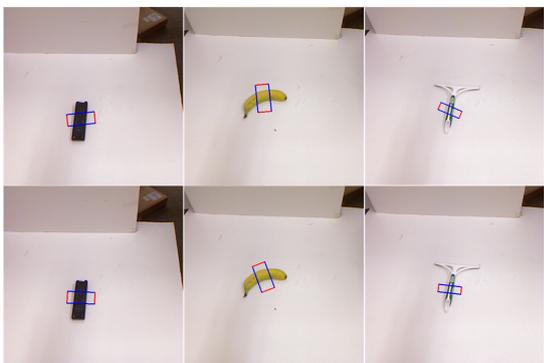


图7 角度监督初始抓取框

为了验证加入角度预测模块来监督阶段1抓取框的偏转角以获取最优抓取角度的优越性,本文在V-REP中搭建了一个如图8所示的仿真环境,并通过检测姿态矫正前后抓取时的扭矩大小是否改善进行验证.硬件部分包括UR5机械臂、二指抓手、深度相机、六维力传感器等.其中,将二指抓手连接到机械臂的腕关节处,工作平台与力传感器连接在一起,放置物体后把抓取对象固定在平台上,相机固定于平台正上方,负责采集待抓取物体的RGB-D信息.

本文选取4种生活中常见的物品进行抓取仿真实验,分别是螺丝刀、香蕉、锤子和扳手.首先,通过RGB彩色图获取抓取位姿;然后,规划一条运动轨迹,控制机械臂从初始位姿逐步到达抓取位姿执行抓取

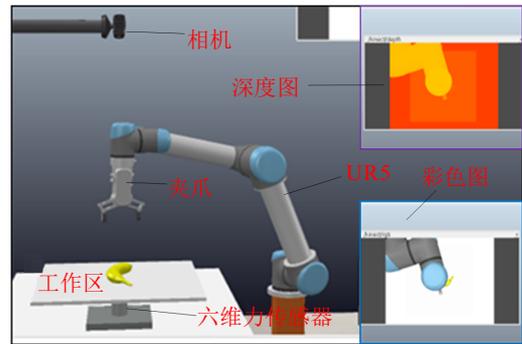


图8 抓取仿真系统

操作;最后,通过测定抓取时平台在Z轴上产生的扭矩大小进行验证.

在获取位姿后,需要进行逆运动学运算,将位姿转化为6个关节转角,进而驱动机械臂.传统的牛顿迭代算法具有收敛速度快的优势,但在求解机器人逆运动学问题时,此类方法受设定初值的影响较大,容易出现迭代发散的情况.采用遗传算法求解机器人逆运动学的主要问题在于迭代后期收敛速度较慢,但是其可以通过初始化一定组数的6关节转角来求解适应度,进而迭代产生趋于收敛的解,这样便容易获得适合牛顿迭代算法的初始值,避免了迭代发散的现象.因此,结合上述两类算法的优点,先使用遗传算法求解满足一定精度的关节角作为牛顿迭代算法的初始值,然后采用牛顿迭代法^[25]进行迭代求解,获取能同时满足目标位置与目标姿态精度的6个关节转角.组合算法流程如图9所示.

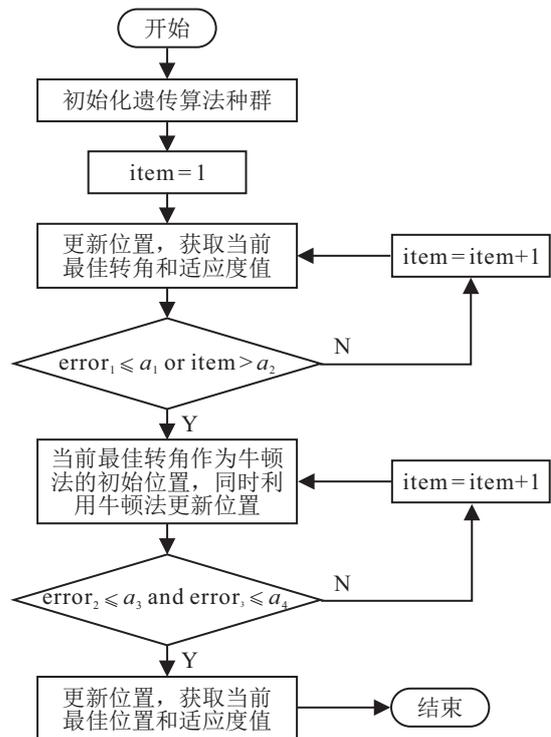


图9 遗传-牛顿算法流程

对于遗传算法的个体基因编码方式,本文采用

的是浮点数编码. 该编码方式的基因多样性好, 被优化参数的数值精度高. 每个基因被初始化到 $[0, 1]$ 区间内, 基因数量取决于被优化的参数个数, 本文为6个. 为了增加遗传算法的局部搜索能力, 本文还采用了贪心算法^[26]进行种群初始化. 选择操作的目的是筛选出优良个体并将其保存在种群中, 为了更好地模拟生物优胜劣汰的自然选择, 本文引入了精英保留策略生成种群1, 从而在后续的交叉和变异过程中, 这些优秀个体的基因信息可以被传递给下一代, 提高整个种群的适应度. 交叉操作的目的是通过交换基因重组来产生适应度更高的子代个体, 本文利用交叉概率对种群1进行筛选, 获取待交叉的种群2, 然后从种群2中依次选择两个父辈个体通过交叉进行基因组交换. 本文采用了双点交叉^[27], 在两个个体中随机生成两个交叉点, 然后交换2个交叉点之间的基因部分. 变异操作可以为种群提供多样性, 因此本文采用逆转变异^[28]以增加种群多样性. 同时, 利用下式进行局部搜索:

$$x = \begin{cases} x + (x_{\text{best}} - x) \times \text{rand}_1, & \text{rand} < 0.5; \\ x + \text{rand}_2, & \text{otherwise.} \end{cases} \quad (13)$$

其中: x 为变异的个体, x_{best} 为最好的个体, rand_1 为 $[0, 1]$ 之间的均匀随机数, rand_2 为 $[-0.01, 0.01]$ 之间的均匀随机数. 为了增强遗传算法的局部搜索能力, 本文还引入了模拟退火机制^[29], 对每一代最好的一组角度进行局部搜索. 最后利用精英策略保留下来的种群1, 以及交叉、变异和模拟退火获取的个体形成新的种群.

遗传算法的参数分别为: 种群规模为50, 交叉概率为0.8, 变异概率为0.2, 6个关节转角的上下限为 $[-\pi, \pi]$. error_1 为遗传算法的适应度函数, 可以表示为

$$\text{error}_1 = \delta_1 + \delta_2, \quad (14)$$

$$\delta_1 = t(x_0 - x_1)^2 + (y_0 - y_1)^2 + (z_0 - z_1)^2, \quad (15)$$

$$\delta_2 = (\alpha_0 - \alpha_1)^2 + (\beta_0 - \beta_1)^2 + (\gamma_0 - \gamma_1)^2. \quad (16)$$

式(15)为位置误差平方和, 式(16)为姿态误差平方和. 在执行遗传-牛顿算法时, a_1 为执行遗传算法阶段的最小适应度值, 本文取为1200; a_2 为执行遗传算法阶段的最大迭代次数, 本文取为16.

执行牛顿迭代算法阶段, error_2 为位置误差, error_3 为姿态误差, 可以由下式得到:

$$\text{error}_2 = |x_0 - x_1| + |y_0 - y_1| + |z_0 - z_1|, \quad (17)$$

$$\text{error}_3 = |\alpha_0 - \alpha_1| + |\beta_0 - \beta_1| + |\gamma_0 - \gamma_1|. \quad (18)$$

a_3 为位置误差精度, 大小为0.001 mm; a_4 为姿态误差精度, 大小为0.1度.

图10为遗传算法迭代过程中的适应度曲线, 在

执行遗传算法时, 只要遗传算法获取的位置精度能达到0.01 mm且姿态精度能达到0.1°时, 则跳出迭代运算. 从图10的收敛曲线可以看出, 适应度曲线在前期快速衰减, 但是后期收敛速度慢, 一直达到最大迭代次数也未满足位置精度与姿态精度的要求. 图11为遗传-牛顿算法迭代过程中的适应度曲线, 可以看出采用遗传-牛顿法的迭代次数在25次以内就可以达到位置精度和姿态精度的要求.

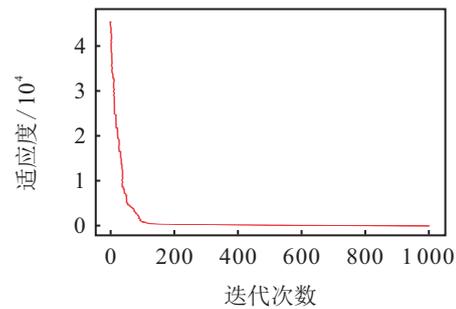


图10 遗传算法适应度曲线

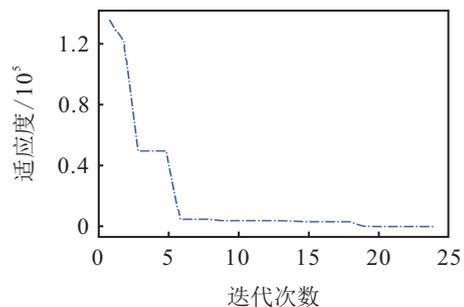


图11 遗传-牛顿算法适应度曲线

图12为遗传-牛顿算法的误差实验结果, 横轴为实验次数, 纵轴分别为位置误差和姿态误差. 从图12可知, 位置误差在 $0 \sim 0.001$ mm 之间波动, 而姿态误差在精度要求为0.1°时, 一直为0. 因此, 遗传-牛顿算法在位置精度要求在0.001 mm以内, 姿态精度要求在0.1°以内的作业环境下, 能够满足作业要求.

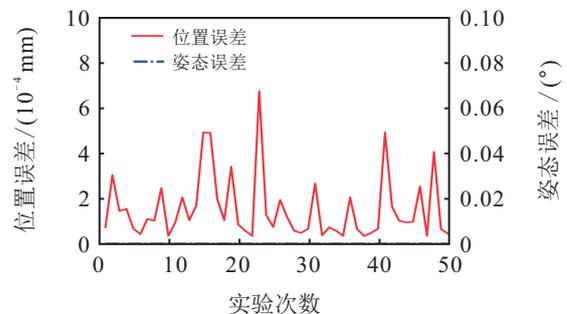


图12 遗传-牛顿算法误差

在一台2.30 GHz主频的笔记本电脑上对1200组角度进行逆运动学求解测试, 测试的计算时间如图13所示. 由图13可知, 6R机器人的逆运动学求解的单次计算时间大部分在120 ms以内波动, 而平均求解时间为46 ms. 在实际的工业应用中, 可以采用并行

计算技术等其他加速方法加速运算,进而进一步降低机械臂逆运动学求解的计算时间.

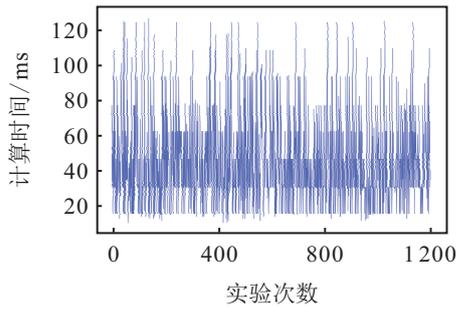
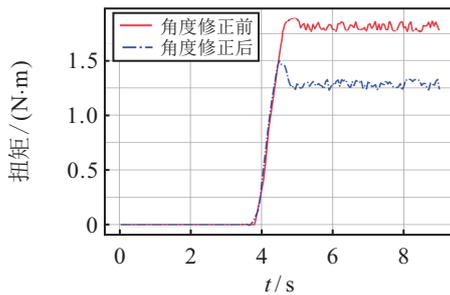
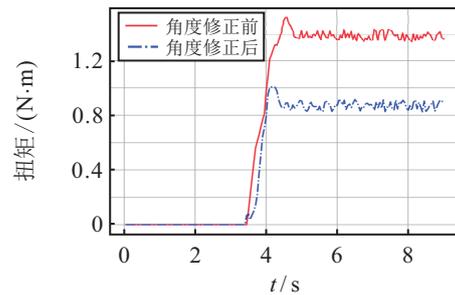


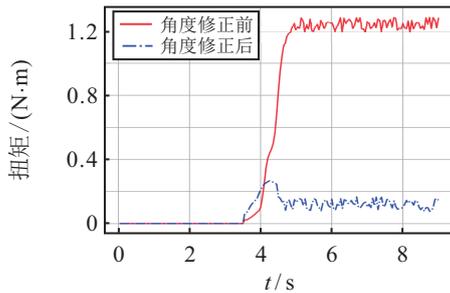
图13 1200组逆解测试时间



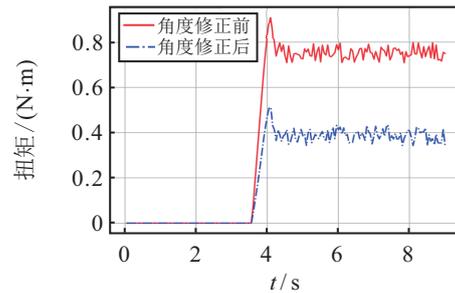
(a) 仿真1



(b) 仿真2



(c) 仿真3



(d) 仿真4

图14 抓取扭矩

3.4 机器人实际抓取实验

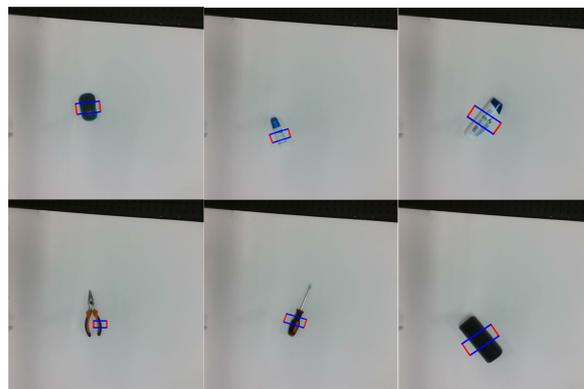
本文搭建如图15(a)所示的机械臂抓取检测系统,该系统由UR5机械臂、二指机械抓手和D435深度相机等组成.具有6个自由度的UR5机械臂手腕处装有二指抓手作为末端执行器,可以进行异状物体抓

取.将仿真过程中的扭矩与时间关系绘制成图14,图14中的同一仿真实验是针对同一物体,红色曲线是在抓取角度修正前采集的抓取扭矩大小,蓝色点划线是在抓取角度修正后采集的抓取扭矩大小.由图14可知,添加角度预测分支,以对抓取检测网络生成的抓取框的偏转角进行修正,其中角度修正后抓取时的扭矩大小较角度修正前抓取时的扭矩大小要小.仿真结果表明,添加角度预测模块来监督抓取检测网络输出的抓取框偏转角,能够有效降低抓取时有害扭矩大小.

取.相机固定于工作台的正上方,安装方式采用“眼在手外”的模式,以便获取良好的RGB-D信息.RGB彩色图被送入到预先训练好的网络中进行抓取检测,获取带有抓取位置和抓取角度的抓取矩形框,部分物体的实际抓取检测效果如图15(b)所示.



(a) 实际抓取位姿检测场景



(b) 实际抓取位姿检测结果

图15 实际抓取检测系统和抓取检测效果

实际实验中选取6种常见的物体对模型性能进行评估,每个对象以10个不同的位置随机放置在工作区中分别抓取,记录抓取成功的次数.实际抓取实验结果如表2所示,结果表明,在有限的实际抓取实验中,本文设计的算法能达到88.3%的平均抓取成功率.

表2 抓取姿态检测与抓取实验结果

物体	抓取次数	成功次数	成功率/%
钳子	10	8	80
香波	10	10	100
鼠标	10	7	70
眼镜盒	10	10	100
固体胶	10	8	80
螺丝刀	10	10	100
平均			88.3

在有限的实验数据中,本文的模型用于实际抓取时依然存在抓取失败的情况.对于固体胶和鼠标,该类物体表面过于光滑,摩擦系数小,导致抓手与物体表面之间的摩擦阻力不够,因此未能成功夹起物体.对于钳子,由于抓取位置不在质心处,抓取时作用力合力未落于摩擦锥内使物体被提起的过程中出现滑动,导致抓取失败.以上现象说明了检测出来的抓取位置未必适合实际抓取需求,需进一步根据物体质心和抓取时摩擦锥的影响来限制抓取检测的位置.

4 结论

针对机械臂抓取位姿检测任务,本文提出了基于语义分割网络与旋转目标检测的单目位姿估计方法.在Cornell数据集上进行网络对比实验,本文提出的网络抓取检测精度可以达到98.4%.利用V-REP搭建机械臂抓取检测系统仿真实验平台,仿真结果表明,抓取时的有害扭矩大小较修正前有所改善.此外,在有限的实际抓取实验中,平均抓取成功率达到了88.3%.但是,由抓取实验失败案例来看,后期可以将物体的质心和摩擦锥的影响考虑进抓取检测中,此为未来定位抓取算法的改进重点.

参考文献(References)

[1] Sintov A, Shapiro A. Dynamic regrasping by in-hand orienting of grasped objects using non-dexterous robotic grippers[J]. *Robotics and Computer-Integrated Manufacturing*, 2018, 50: 114-131.

[2] Zapata-Impata B S, Gil P, Pomares J, et al. Fast geometry-based computation of grasping points on three-dimensional point clouds[J]. *International Journal of Advanced Robotic Systems*, 2019, 16(1):

172988141983184.

[3] Mousavian A, Eppner C, Fox D. 6-DOF GraspNet: Variational grasp generation for object manipulation[C]. *2019 IEEE/CVF International Conference on Computer Vision*. Seoul, 2019: 2901-2910.

[4] Gao K, Zhang Y D, Zhang W, et al. Mining concise and distinctive affine-stable features for object detection in large corpus[J]. *International Journal of Computer Mathematics*, 2011, 88(18): 3953-3962.

[5] Girshick R. Fast R-CNN[C]. *2015 IEEE International Conference on Computer Vision*. Santiago, 2015: 1440-1448.

[6] 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.

[7] Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps[J]. *International Journal of Robotics Research*, 2015, 34(4/5): 705-724.

[8] Chu F J, Xu R N, Vela P A. Real-world multiobject, multigrasp detection[J]. *IEEE Robotics and Automation Letters*, 2018, 3(4): 3355-3362.

[9] 李秀智, 李家豪, 张祥银, 等. 基于深度学习的机器人最优抓取姿态检测方法[J]. *仪器仪表学报*, 2020, 41(5): 108-117.

(Li X Z, Li J H, Zhang X Y, et al. Detection method of robot optimal grasp posture based on deep learning[J]. *Chinese Journal of Scientific Instrument*, 2020, 41(5): 108-117.)

[10] Morrison D, Leitner J, Corke P. Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach[J/OL]. 2018, arXiv: 1804.05172.

[11] Araki R, Onishi T, Hirakawa T, et al. MT-DSSD: Deconvolutional single shot detector using multi task learning for object detection, segmentation, and grasping detection[C]. *2020 IEEE International Conference on Robotics and Automation*. Paris, 2020: 10487-10493.

[12] Jiang Y, Moseson S, Saxena A. Efficient grasping from RGBD images: Learning using a new rectangle representation[C]. *2011 IEEE International Conference on Robotics and Automation*. Shanghai, 2011: 3304-3311.

[13] Lenz I, Lee H, Saxena A. Deep learning for detecting robotic grasps[J]. *International Journal of Robotics Research*, 2015, 34(4/5): 705-724.

[14] 吴青青, 许廷发, 闫辉, 等. 复杂背景下的颜色分离背景差分目标检测方法[J]. *兵工学报*, 2013, 34(4): 501-506.

(Wu Q Q, Xu T F, Yan H, et al. An improved color separation method for object detection in complex

- background[J]. *Acta Armamentarii*, 2013, 34(4): 501-506.)
- [15] He K M, Zhang X Y, Ren S Q, et al. Identity mappings in deep residual networks[C]. *European Conference on Computer Vision*. Cham: Springer, 2016: 630-645.
- [16] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, 2017: 936-944.
- [17] 李明, 鹿朋, 朱龙, 等. 基于RGB-D融合的密集遮挡抓取检测[J]. *控制与决策*, 2023, 38(10): 2867-2874. (Li M, Lu P, Zhu L, et al. Densely occluded grasping objects detection based on RGB-D fusion[J]. *Control and Decision*, 2023, 38(10): 2867-2874.)
- [18] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 7794-7803.
- [19] Wu G B, Chen W S, Cheng H, et al. Multi-object grasping detection with hierarchical feature fusion[J]. *IEEE Access*, 2019, 7: 43884-43894.
- [20] Porzi L, Bulò S R, Colovic A, et al. Seamless scene segmentation[C]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019: 8269-8278.
- [21] Jiang Y H, Li W W, Liu L. R-CenterNet+: Anchor-free detector for ship detection in SAR images[J]. *Sensors*, 2021, 21(17): 5693.
- [22] Zhang H B, Lan X G, Bai S T, et al. ROI-based robotic grasp detection for object overlapping scenes[C]. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Macau, 2019: 4768-4775.
- [23] Karaoguz H, Jensfelt P. Object detection approach for robot grasp detection[C]. *2019 International Conference on Robotics and Automation (ICRA)*. New York: ACM, 2019: 4953-4959.
- [24] Xu R N, Chu F J, Vela P A. GKNet: Grasp keypoint network for grasp candidates detection[J]. *International Journal of Robotics Research*, 2022, 41(4): 361-389.
- [25] Liu W Z, Zhao C, Liu Y, et al. Sim2real kinematics modeling of industrial robots based on FPGA-acceleration[J]. *Robotics and Computer-Integrated Manufacturing*, 2022, 77: 102350.
- [26] 叶闻语, 王春阳, 于金阳, 等. 基于贪心遗传算法的液晶光学相控阵多光束扫描[J]. *兵工学报*, 2023, 44(9): 2650-2660. (Ye W Y, Wang C Y, Yu J Y, et al. Multi-beam scanning of liquid crystal optical phased array based on greedy algorithm[J]. *Acta Armamentarii*, 2023, 44(9): 2650-2660.)
- [27] 常宁东, 冯春, 程鹏达, 等. 基于Bekker理论改进遗传算法的野外路径优化方法研究[J]. *光学精密工程*, 2023, 31(5): 767-775. (Chang N D, Feng C, Cheng P D, et al. Research on off-road path optimization algorithm based on Bekker theory improved genetic algorithm[J]. *Optics and Precision Engineering*, 2023, 31(5): 767-775.)
- [28] 阎哲, 汪民乐, 汪江鹏, 等. 基于混合遗传算法的海军航空兵场站物资配送车辆调度智能优化[J]. *系统工程与电子技术*, 2023, 45(12): 3908-3914. (Yan Z, Wang M L, Wang J P, et al. Intelligent optimization of vehicle scheduling for material distribution in naval aviation station based on hybrid genetic algorithm[J]. *Systems Engineering and Electronics*, 2023, 45(12): 3908-3914.)
- [29] 牛志华, 苑璨, 孔得宇. 计算周期序列k-错线性复杂度的混合遗传算法[J]. *上海交通大学学报*, 2020, 54(6): 599-606. (Niu Z H, Yuan C, Kong D Y. A hybrid genetic algorithm for computing the k-error linear complexity of periodic sequences[J]. *Journal of Shanghai Jiao Tong University*, 2020, 54(6): 599-606.)

作者简介

孙先涛(1985—), 男, 副教授, 硕士生导师, 从事欠驱动机械手、机器视觉等研究, E-mail: xtsun@ahu.edu.cn;

闻勇(1995—), 男, 硕士生, 从事机器视觉的研究, E-mail: 1446964956@qq.com;

陈文杰(1964—), 男, 教授, 博士生导师, 从事欠驱动机械手、外骨骼等研究, E-mail: wjchen@ahu.edu.cn;

智亚丽(1987—), 女, 讲师, 硕士生导师, 从事时滞系统控制理论、故障诊断等研究, E-mail: zhiyali@ahu.edu.cn;

胡祥涛(1982—), 男, 教授, 博士生导师, 从事机器学习、机器人化智能制造等研究, E-mail: hust_hoo@163.com;

陈伟海(1955—), 男, 教授, 博士生导师, 从事康复机器人、仿生机器人等研究, E-mail: whchen@buaa.edu.cn.