第 42 卷 第 8 期
2021 年 8 月

仪 器 仪 表 学 报
Chinese Journal of Scientific Instrument

Vol. 42 No. 8
Aug. 2021

DOI: 10.19650/j.cnki.cjsi.J2107892

基于三维检测网络的机器人抓取方法*

葛俊彦,史金龙,周志强,王 直,钱 强
(江苏科技大学计算机学院 镇江 212100)

摘 要:机器人抓取任务中面对的是不同形状和大小的物体,而散落在场景中的物体会有不同的姿态和位置,这对机器人抓取中计算物体位姿任务提出了较高的挑战。针对于此,本文设计了一种基于三维目标检测的机器人抓取方法,弥补了基于二维图像识别引导机器人抓取任务中对视角要求较高的缺陷。首先,设计了一种卷积神经网络在 RGB 图像中识别物体,并回归出物体三维包围盒、物体中心点;其次,提出一种计算机器人抓取物体最佳姿势的策略;最后,控制机器人进行抓取。在实际场景中,使用本文设计的三维检测网络,三维目标检测精度达到 88%,抓取成功率达到 94%。综上所述,本文设计的系统能有效找到机器人合适的抓取姿势,提高抓取成功率,满足更高的抓取任务要求。

关键词: 深度学习;三维检测;位姿计算;机器人控制
中图分类号: TP242 TH86 **文献标识码**: A **国家标准学科分类代码**: 510.40

A robotic grasping method based on three-dimensional detection network

Ge Junyan, Shi Jinlong, Zhou Zhiqiang, Wang Zhi, Qian Qiang
(*School of Computer Science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212100, China*)

Abstract: The robot faces different shapes and sizes of objects in the task of grasping. The scattered objects in the scene may have different poses and positions, which make the task of recognizing positions and poses of objects more difficult. In view of this, a three-dimensional scene recognition method for robotic grasping is proposed. It makes up a defect that the 2D detection method is sensitive to the field of view in robotic grasping task. Firstly, the convolutional neural network is designed to detect the object in the RGB image. Eight corner points of the three-dimensional bounding box of objects, and the center point of the object are generated. Secondly, a method is proposed to calculate the best position and pose for robotic grasping. Finally, the robot is controlled to grasp objects. In real scene, the detection accuracy reaches 88%, and the grasping success rate based on the designed three-dimensional recognition network is up to 94%. In summary, the designed network can effectively find a suitable grasping pose. The grasping success rate is improved. It is able to meet higher requirements.

Keywords: deep learning; 3D detection; pose calculation; robot control

0 引 言

近年来,机器人被广泛地应用于工厂、家居环境中,并发挥着越来越重要的作用[1]。因此,让机器人拥有对现实场景的感知能力有很高的研究价值。其中最普遍的感知方式是视觉感知,即在机器人任务中融合计算机视觉技术。随着计算机视觉的发展,传统的图像处理方法在机器人视觉抓取运用中已经较为成熟。自 2012 年以来,深度学习以其突出的目标检测精度和可靠性,在图像识别领域开始被应用[2],现阶段基于深度学习的视觉方法也被机器人抓取系统广泛应用。

在实际运用中,机器人会面对种类繁多、形状各异的物体,不准确的物体位置姿态计算会直接导致机器人抓取任务的失败。现阶段,基于深度学习的机器人抓取方法有如下 3 种:1)基于先验知识抓取点的方法。这类方法通过学习成功的抓取实例,结合特定场景进行抓取。首先进行物体位姿估计,其次利用点云配准算法将场景中物体与数据库中的模型配准,进行位姿调整,最后得到合适的抓取方式。Billings 等[3]提出了利用感兴趣区域

收稿日期:2021-05-08 Received Date: 2021-05-08
*基金项目:科技部国家重点研发计划(2018YFC0309104)项目资助

(region of interest, ROI)得到物体位姿和抓取点的方法,从抓取点数据库中找到合适的抓取位置。2)直接从图像中回归出抓取点。文献[4-7]先生成若干抓取选择,然后通过筛选,得到质量最高的抓取位置。3)通过网络估计或直接计算的方式得到物体在场景中的位置姿态信息。文献[8]中使用了吸盘式抓取工具,仅定位物体中心,不关注物体的姿态。文献[9]中提出通过 Faster R-CNN 目标识别方式得到物体位置,利用机械手爪进行抓取,该方法同样没有得到物体姿态信息。文献[10]利用加速鲁棒特征(speeded up robust features, SURF)重建场景,采用主成分分析法[11](principal component analysis, PCA)和迭代最近点法(iterative closest point, ICP)得到物体的姿态信息。文献[12]利用点云聚类得到物体的点云,将点云进行投影得到物体姿态信息。文献[13]使用 Mask R-CNN 和 PCA 联合的方法,得到物体的位置和姿态信息。该方法得到了物体位置姿态信息,但对图像采集视角非常敏感,可移植性不够高。

鉴于传统算法对复杂场景的方法可移植性较差、二维图像识别技术对视角敏感的问题,本文设计了一种面向机器人抓取的场景三维检测方法。首先,利用深度相机采集场景 RGB 图像和深度图像,通过重建得到场景三维信息。其次,基于 YOLO[14] 目标检测网络,设计了一种直接回归出物体的中心点以及三维包围盒的三维检测网络,重新采集场景三维数据集,对网络模型进行微调训练,使网络对仿真场景有更好的检测效果;最后融合场景三维信息及优化后定位点的坐标,计算出物体在三维空间中的位置、姿态,控制机器人抓取。同时,设计对比实验,复现了文献[13]中二维目标检测分割方法,与本文提出的方法进行对比,验证了本文方法的有效性、优越性。

1　机器人抓取任务流程

机器人抓取流程如图 1 所示。首先,利用 Intel RealSense D435i 深度相机采集场景的 RGB 图像和深度图像,并将二者的像素对齐,重建出场景,得到场景中各个像素点在相机坐标系下的三维坐标;其次,使用三维目标检测网络回归出目标的中心点和三维包围盒,并对中心点作位置矫正;然后,利用包围盒的结构计算出物体主轴的朝向,并用旋转向量表示,即机器人抓取姿势;最后,通过手眼标定计算出相机坐标系与机器人基坐标系之间的旋转、平移矩阵,完成视觉引导机器人控制的抓取任务。

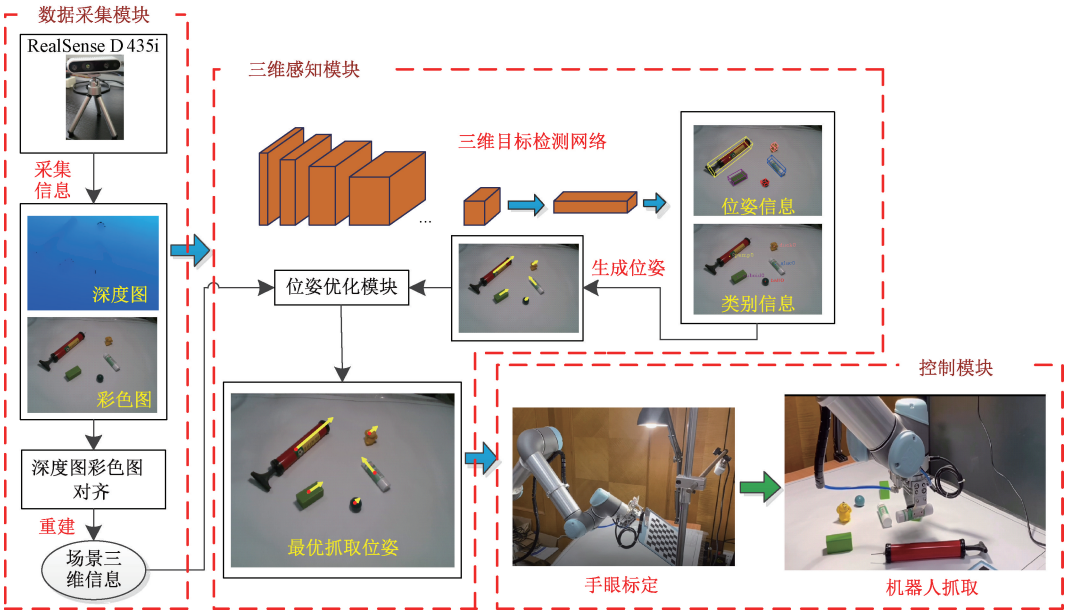


图 1　机器人抓取流程

Fig. 1　Work flow of robotic grasping

2　场景感知模块

2.1　基于 Mask-RCNN 和 PCA 的感知方法

Mask-RCNN[15]是一种基于深度学习的像素级目标检测与分割算法,能够精确地定位和分割图像中的实例。实例分割任务是判断出图像中每个像素属于哪个类别,利用实例分割任务,即可得到各个实例的像素分布。

PCA 是一种统计学方法。该算法旨在通过降维的思想,把数据投影到新的坐标空间中,使所有数据在某一个

维度上方差最大,即在这一维度上数据有最大的贡献度。由此,在得到 Mask-RCNN 实例分割的像素信息后,可以利用 PCA 算法计算出物体在平面上的方向。位置信息计算较为简单,在实例所属的所有像素中,分别计算横、纵坐标均值,即可作为物体的中心位置。基于 PCA 的实例方向估计算法具体如下:

1)取一个实例掩膜中所有的二维坐标,将二维点的两个坐标分别当作一个特征维度。

2)将两个像素维度的特征组成一个矩阵,求这个 $n \times 2$ 维矩阵的协方差矩阵 $\boldsymbol{C}$,$\boldsymbol{C} = \boldsymbol{X}^{\mathrm{T}} \cdot \boldsymbol{X}/(n-1)$。并计算 C 的特征值和对应的特征向量。

3)取特征值最大对应的特征向量方向,即像素分布方差最大方向,作为该实例的主轴朝向。

4)重复步骤 1),直到图像中所有检测到的实例被处理。

2.2 三维检测网络

PCA 算法解决了物体在图像中朝向的问题,但仍存在较为明显的局限性,即拍摄视角问题。具体地说,PCA 算法只适用于相机在场景正上方拍摄的情况,在视角倾斜的情况下,PCA 算法会出现明显的定位偏移错误,导致抓取失败。针对这个问题,本文设计了一种三维检测网络以及位姿计算方法,目的是在图像中直接回归出含有物体位置姿态信息的三维包围盒,并计算出物体位姿,解决二维识别分割任务可能导致的机器人抓取定位不准确问题。

1)数据集制作

根据应用场景,本文使用开源工具 Object Dataset Tools 重新采集了三维数据集。此工具包含纯 python 脚本,用于为 RGB-D 图像序列创建对象掩膜、边界框标签,同时进行目标物体三维模型重建。该工具可以为各种深度学习项目准备训练和测试数据,如 6D 目标姿态估计任务,以及许多目标检测和实例分割任务。

首先,将定位标签均匀分布在背景中,定位标签的作用是拼接点云并重建物体三维模型。其次,将目标物品散落在场景当中,如图 2 所示。利用 RealSense D435i 深度相机环绕场景一周并录像。本文在实验中,视频每 5 帧采样一次,作为数据集中的图像,对应的深度图像也同时保存。完成上述工作后,利用定位标签、深度图以及点云滤波的信息,通过重建得到物体的三维模型,如图 3 所示。得到三维模型后,Object Dataset Tools 会标注出当前目标物体在每张图像中的掩膜和该实例在每张图像中三维包围盒的归一化角点坐标。

2)网络结构设计

YOLO[16] 目标检测框架通过单步回归的方式直接完成目标检测,可达到较快的检测速度。不同于 R-CNN 系列的目标检测任务,YOLO 目标检测舍弃了候选区域的

图 2 数据集采集场景

Fig. 2 Scene in dataset collection

图 3 重建结果

Fig. 3 Reconstruction result

生成网络,将图像识别的任务转化为回归任务,直接在 RGB 图像中回归出包围盒的坐标。在本文的任务中,与 YOLO 的不同在于本文试图在 RGB 图像中直接回归出三维包围盒的角点。

因此,本文基于 YOLO 框架,设计了一个三维检测网络。网络输入为 $416 \times 416 \times 3$ 的 RGB 图像,输出为 $(13 \times 13) \times (9 \times 2 + 1 + 5)$ 的张量。其中,“13”表示图像被分成 13×13 个网格;9×2 表示网络将预测出每个实例的三维包围盒、中心点在图像中的投影坐标,且该坐标经过归一化处理;“1”表示每个预测网络会产生一个置信度,“5”表示五类物体的类别预测分支。此外,网络中融合了不同特征层的信息,提高检测的准确率。具体网络结构如表 1 所示。conv2d 表示二维卷积操作,BN 表示批量归一化,K_c 表示卷积核大小,K_m 表示池化核大小,S_m 表示池化步长。

3)物体位置姿态生成模块

本文基于文献[17],在预测信息时,引入了如下损失函数:

$$Loss = \alpha L_{corner} + \beta L_{confidence} + \gamma L_{class} \tag{1}$$

其中,L_{corner} 表示三维包围盒 8 个角点以及物体中心点坐标的损失,$L_{confidence}$ 表示置信度损失,L_{class} 表示分类损失。

在预测中心点和三维包围盒的角点时,采用了不同的方法。YOLO 中预测包围盒实质上预测的是包围盒相对于网格的偏移,物体中心点落到的网格单元负责该物体的检测。在本文的任务中,中心点的预测参考 YOLO

表 1　网络结构

Table 1　Structure of network

	结构描述	参数	输出
#0	输入图像		416×416×3
#1	conv2d+BN	$K=3$	416×416×32
#2-4	conv2d+BN+MaxPooling	$K_c=3, K_m=2, S_m=2$	104×104×128
#5	conv2d	$K=3$	26×26×512
#6	conv2d	$K=3$	
#7	conv2d	$K=3$	
#8	conv2d+BN+MaxPooling	$K_c=3, K_m=2, S_m=2$	
#9	conv2d	$K=1$	26×26×64
#10	重构		13×13×256
	融合 conv(#5~8) & #9~10		
#11	conv2d	$K=3$	13×13×1 024
#12	conv2d	$K=1$	13×13×26

$$conf(x,y)=\begin{cases}\exp\left(\mu\left(1-\dfrac{\sqrt{(x-x')^2+(y-y')^2}}{\delta}\right)\right), & \sqrt{(x-x')^2+(y-y')^2}<\delta\\ 0, & \text{其他}\end{cases} \tag{3}$$

从损失函数来说,三维包围盒角点损失、物体中心点损失、置信度损失采用 MSE 均方根误差,分类损失采用交叉熵损失。

4)物体位姿生成

在 2.2 节中 1)内,已经回归出物体的三维包围盒以及对应的中心点坐标。而针对本文中特殊的平面抓取场景,不采用二维与三维对应点匹配的方法进行 6 自由度位姿估计[18-19],提出一种基于包围框结构的位姿估计方法。本节将介绍如何生成物体在场景中的位置和姿态。

首先描述位置信息。由于本文采用了平面抓取方式,那么抓取定位点应是物体正上方视角的中心点。网络虽然回归出了物体的中心点,但仍然存在较大的偏移,导致抓取失败,因此本文提出了一种中心点调整方案。三维包围盒中能够反映物体的结构信息,在机器人平面抓取时,应从物体结构的正上方视角定位,所以本文提出如下的物体定位点调整方法:

(1)找到构成物体结构上表面的四个角点。由于采集数据集的有序性,因此回归出的角点是有序的点序列,即上表面的点相对位置不会发生变化。

(2)取上表面的一个角点,与相邻两个角点计算欧式距离,找到上表面的两个短边,取两个短边连线中点作为物体定位点。经过定位点调整,能有效将定位点调整到物体正上方。

其次描述姿态信息。姿态即为物体在场景中主轴的朝向,只有得到了物体主轴朝向,才能调整机器人二指夹爪的转动进行抓取。虽然三维包围盒隐含了物体在三维中的朝向,但由于数据集中的数据误差,以及网络本身的误差,往往不能达到机器人的抓取精度要求。本文中,利用调整后中心点位置结合结构信息,改善了这个问题。本文使用了空间向量的形式表示姿态。在之前的描述中,已经得到了物体上表面的中心点以及一个短边的中点,将两个点分别作为向量的起始点和终点,就得到了主轴的姿态向量。特殊的,由于球体的对称性,本文不对球体设计姿态向量。具体来说,本文先规定了对称物体以及非对称物体集合,以物体类别索引的形式保存。本文的网络能够识别出物体的类别,根据类别索引即可判断该物体是否对称。至此,本文在三维层面上完成了 RGB 图像的三维包围盒检测任务、物体空间位姿计算任务。

从损失函数来说,三维包围盒角点损失、物体中心点损失、置信度损失采用 MSE 均方根误差,分类损失采用交叉熵损失。

4)物体位姿生成

在 2.2 节中 1)内,已经回归出物体的三维包围盒以及对应的中心点坐标。而针对本文中特殊的平面抓取场景,不采用二维与三维对应点匹配的方法进行 6 自由度位姿估计[18-19],提出一种基于包围框结构的位姿估计方法。本节将介绍如何生成物体在场景中的位置和姿态。

首先描述位置信息。由于本文采用了平面抓取方式,那么抓取定位点应是物体正上方视角的中心点。网络虽然回归出了物体的中心点,但仍然存在较大的偏移,导致抓取失败,因此本文提出了一种中心点调整方案。三维包围盒中能够反映物体的结构信息,在机器人平面抓取时,应从物体结构的正上方视角定位,所以本文提出如下的物体定位点调整方法:

(1)找到构成物体结构上表面的四个角点。由于采集数据集的有序性,因此回归出的角点是有序的点序列,即上表面的点相对位置不会发生变化。

(2)取上表面的一个角点,与相邻两个角点计算欧式距离,找到上表面的两个短边,取两个短边连线中点作为物体定位点。经过定位点调整,能有效将定位点调整到物体正上方。

其次描述姿态信息。姿态即为物体在场景中主轴的朝向,只有得到了物体主轴朝向,才能调整机器人二指夹爪的转动进行抓取。虽然三维包围盒隐含了物体在三维中的朝向,但由于数据集中的数据误差,以及网络本身的误差,往往不能达到机器人的抓取精度要求。本文中,利用调整后中心点位置结合结构信息,改善了这个问题。本文使用了空间向量的形式表示姿态。在之前的描述中,已经得到了物体上表面的中心点以及一个短边的中点,将两个点分别作为向量的起始点和终点,就得到了主轴的姿态向量。特殊的,由于球体的对称性,本文不对球体设计姿态向量。具体来说,本文先规定了对称物体以及非对称物体集合,以物体类别索引的形式保存。本文的网络能够识别出物体的类别,根据类别索引即可判断该物体是否对称。至此,本文在三维层面上完成了 RGB 图像的三维包围盒检测任务、物体空间位姿计算任务。

在二维检测中,往往会使用交并比(intersection over union, IOU)指标作为置信度的值,而本文的方法基于三维检测,在三维空间中计算 IOU 较为繁琐,且会使训练过程变得缓慢,因此本文参考文献[17]设计如下的置信度函数。

$$conf(x,y)=\exp\left(\mu\left(1-\frac{\sqrt{(x-x')^2+(y-y')^2}}{\delta}\right)\right) \tag{2}$$

令 $dist=\sqrt{(x-x')^2+(y-y')^2}$,$dist$ 表示的是预测点(x,y) 和真实点(x',y') 的欧式距离。由式(2)可知,当欧式距离越小时,置信度越大。值得一提的是,δ 是一个截断值,当欧式距离大于截断阈值时,就不认为网络做出了正确的预测,此时将置信度设为 0。因此,可以将置信度函数补充为:

3　实验结果与分析

3.1　机器人控制

完成手眼标定[20-22]后,就可以把相机坐标系下的三维信息转换成机器人基坐标系下的三维信息。将相机坐标系下的物体中心点坐标记为(X_{cam_center}, Y_{cam_center}, Z_{cam_center}),则物体中心点在机器人基坐标下的坐标(X_{base_center}, Y_{base_center}, Z_{base_center}),表示为:

$$\begin{bmatrix}X_{base\_center}\\ Y_{base\_center}\\ Z_{base\_center}\end{bmatrix}=\begin{bmatrix}R_{cam2base} & t_{cam2base}\end{bmatrix}\begin{bmatrix}X_{cam\_center}\\ Y_{cam\_center}\\ Z_{cam\_center}\end{bmatrix} \tag{4}$$

姿态信息在机器人坐标系中的表示与中心点不同。得到物体的主轴三维向量后,不可直接将数据输入机器人坐标。本文使用的机器人工具夹爪姿态的表示方式是旋转向量,由于抓取先验的限制,首先将 R_z 固定为 0,即让工具夹爪只在两个维度上调整姿态。计算出的三维向量与夹爪平行的方向是正交关系,因此首先要将三维向量绕 z 轴旋转 90°,三维旋转向量要先利用罗德里格斯变换转化为旋转矩阵,具体过程见式(5)。

$$\begin{aligned} l &= \sqrt{Rx^2 + Ry^2} \\ \sin\alpha &= Rx/l \\ \cos\alpha &= Rx/l \\ Angle &= Rodrigues\left(\begin{bmatrix} 0 \\ \pi \\ 0 \end{bmatrix}\right)\begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \tag{5}$$

3.2 Mask-RCNN+PCA 位姿估计实验

为了与提出的三维检测网络效果对比,本文先使用了 Mask-RCNN+PCA 的方法估计物体位置、姿态。首先,使用 RealSense D435i 深度相机采集数据集,数据集中包含单个物体与混合物体场景,共 1 500 张 RGB 图像,其中,1 350 张图像作为训练集,150 张作为测试集。其次,使用工具 LabelMe 对 1 500 张图像进行掩膜标注。最后使用 Mask-RCNN 网络进行训练,训练参数设置为:batch_size=16,learning_rate=0.001,epoch=200。完成训练后,使用文献[15]的部分评价指标对训练结果进行评估,结果如表 2 所示,在 AP_{50} 指标上,网络检测任务以及分割均得到了较高的精度。图 4 是使用 2.1 节算法计算后的物体中心点和主方向结果,其中点表示物体中心,实线表示物体主方向。

表 2 检测分割任务准确率

Table 2 Accuracy of detection and segmentation task

指标	AP	AP50	AP75	APS	APM
检测任务	0.51	**0.89**	0.47	0.53	0.52
实例分割任务	0.47	**0.86**	0.42	0.51	0.43

上述实验中,相机位于场景正上方位置,位姿计算得到了较好的结果。然而,在相机视角倾斜时,物体定位任务会发生非常明显的错误。在视角倾斜的情况下,Mask-RCNN 输出的掩膜大多分布在物体的侧表面,或上表面与侧表面的交界处,这对于抓取任务是不利的。图 5 是倾斜视角下使用 Mask-RCNN+PCA 算法计算出的物体位姿结果,可以明显发现:长方体的定位点位置发生了严重的偏移,算法不能定位到物体上表面,且其他几类物体的定位点也发生了不同程度的偏移,严重影响抓取任务的实施。

图 4 Mask-RCNN+PCA 位姿计算结果

Fig. 4 Results of Mask-RCNN+PCA

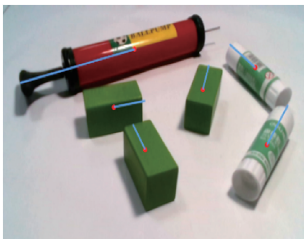


图 5 含有错误的位姿计算结果

Fig. 5 Pose calculation results with error

3.3 三维检测网络效果及评估实验

在训练三维检测网络时,设置批量大小为 16,学习率为 0.001,迭代轮次为 300,使用 GTX-1080Ti 显卡进行训练。

图 6 是不同物体的三维检测结果,其中白色表示真实三维框,黑色表示经神经网络预测的三维框。图 7 是

图 6 各类物体三维检测结果

Fig. 6 3D detection results of different objects

混合多种物体的仿真场景三维检测结果,为了便于显示,图中颜色相同的包围盒表示同一类物体。

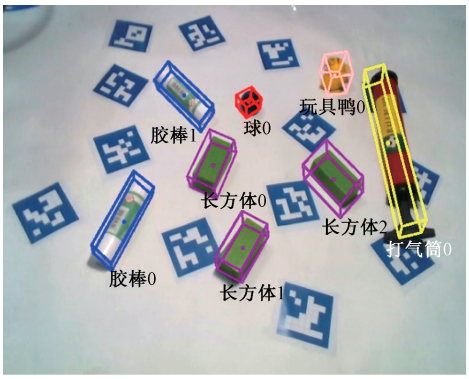


图 7　混合场景三维检测结果

Fig. 7　3D detection results of mixed scene

为了验证抓取点调整的合理性,图 8 对比了调整前后的实际抓取场景。由图可见,抓取点调整后可以成功抓取物体。

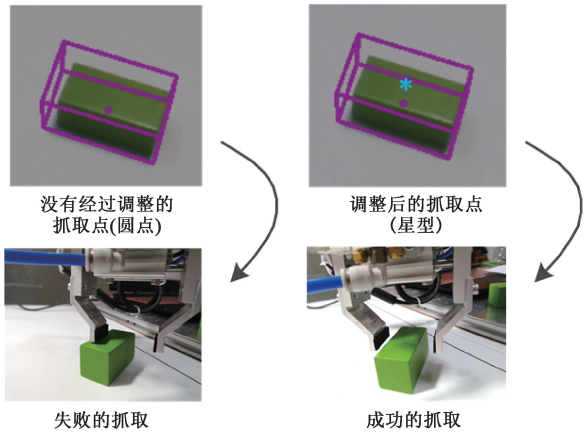


图 8　抓取点调整前后对比

Fig. 8　Comparison before and after grasping point adjustment

此外,本文还使用了基于像素间欧式距离的精确度评估方法对预测结果进行评估,具体方法为:计算预测点与真实点的二维像素欧式距离,并设置不同的像素阈值,本文阈值设置为 5、10、15 和 20 pixel,如果像素欧式距离小于像素阈值,则认为该点为正确的预测。图 9 中的 5 条曲线分别表示在不同分辨率误差范围内正确预测点数占所有预测点数的比例。其中,横轴为不同的误差范围阈值,纵轴为对应比例值,当阈值取 10 pixel 时,各个物体的平均检测准确率已经到达 88%,且结合相机内参计算出的实际误差距离不超过 4.6 mm,能够满足机器人抓取要求。为了更好地评估检测准确率对抓取的影响,本文将不同像素阈值下的指标做了更详细的量化,数据见表 3。

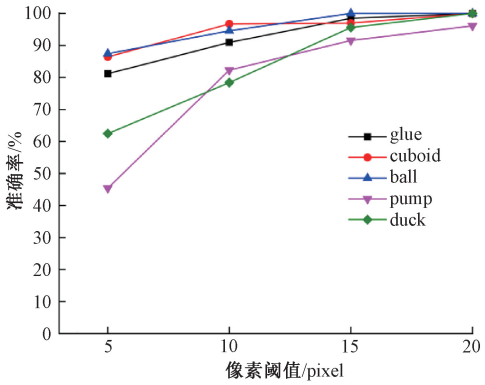


图 9　5 类物体在不同像素阈值下的三维框预测准确率

Fig. 9　Accuracy of 5 objects in different pixel thresholds

表 3　不同像素阈值对抓取影响量化指标

Table 3　Effects for grasping in different thresholds

	mAP/%	实际误差/mm	是否适合抓取
5 pixel	72	<2.3	√
10 pixel	**88**	**<4.6**	√
15 pixel	96	<6.9	×
20 pixel	98	<9.2	×

3.4　实际场景抓取实验

在抓取实验中,本文使用 UR5 轻量型协作式工业机器人,有效载荷为 5 kg,满足本次抓取实验要求。本文在单类物体及混合多种物体的实际场景中,进行了抓取实验,并比较三种方法的抓取成功率,如表 4 所示。

表 4　实际场景抓取成功率

Table 4　Success grasping rate of real scene　%

	Mask-RCNN+PCA	基于三维检测	基于三维检测+位姿调整(本文)
胶棒	68	88	92
长方体	48	60	88
球	80	92	100
打气筒	60	80	96
玩具鸭	40	88	96
多类别场景	59	81	94

由表 4 实验结果可知,基于 MaskRCNN 和 PCA 的二维检测分割引导的机器人抓取,在不规则物体和有明显棱角的物体上,表现出较差的性能。而基于三维检测网络引导的抓取成功率有了小幅度的提升,在此基础上采用本文提出的位姿调整方案后,抓取成功率有了大幅度的提升。

同时,本文对文献[3,7,10]的方法进行基本思路复现,并在多种类物体场景中做对比实验。结果如表 5 所示。3 篇文献分别对应了机器人抓取任务的 3 种不同思路,由于方法不同,本文不对算法本身的精度做出比较,而比较实际场景中抓取成功率和时间性能。文献[3]使用深度学习的方法,直接回归出物体的姿态。文献[7]使用了回归抓取框的方式得到机器人抓取姿态。文献[10]利用了点云多次配准的方法得到物体的姿态信息,引导机器人抓取。

表 5 抓取成功率与时间性能比较

Table 5 Grasping success rate and time performance comparison

	基于位姿估计(文献[3])	基于抓取框(文献[7])	基于点云配准(文献[10])	基于三维检测+位姿调整(本文)
抓取成功率/%	< 80	87		94
平均每个物体检测耗时/s	0.6	0.6	1.2	0.5

本文实验视频见:https://www.bilibili.com/video/BV1cT4y1u7Ta

对比实验结果分析如下:

1)文献[3]的方法,由于重建模型的质量不够高,导致物体姿态的检测精度较低,各类物体姿态预测精度都低于 80%,不能达到稳定抓取要求,故不另外对其进行实际抓取实验。

2)文献[7]的方法,基于 Faster-RCNN 网络,时间性能较好,但本文的方法基于单阶段网络,比 Faster-RCNN 时间性能略好。且在有些极限场景下,该方法不能有效地预测抓取框。

3)文献[10]虽然得到了较好的抓取成功率,但点云配准消耗时间较长,且每个物体需要单独执行配准过程,不适合抓取任务。这种方法中,点云匹配后仍需在模型上预定义物体抓取点,与本文工作无关,故不进行实际抓取实验。而本文的方法得益于单阶段神经网络,推理时间大大缩短,时间性能远高于点云匹配方法。

综上所述,本文提出的方法在保证抓取成功率的基础上,时间性能也得到了较好的结果。

4 结 论

本文针对机器人平面抓取的特殊场景,设计了一种面向机器人的三维感知定位方法,并制作三维数据集,突破了视觉引导机器人抓取任务中感知模块局限于二维平面检测的瓶颈,且解决了传感器布局要求苛刻的问题,即大幅扩大了场景感知视野,降低了因场景不确定性带来的抓取失败可能,提高任务效率。本文实验中引入许多误差,来源于传感器的精度,算法的可靠性等等。在今后的研究中,预采用直接从网络中回归出物体位姿的方法,并且精确建立物体模型,减少多个步骤的算法带来的不稳定性,使解决方案更加可靠。

参考文献

[1] 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.
LIU Y X,WANG S Y,YAO Y F,et al. Research status of robotic grasping detection technology[J]. Control and Decision,2020,35(12):2817-2828.

[2] 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 ZH,LI J H,ZHANG X Y,et al. Detection method of robot optimal grasping posture based on deep learning[J]. Chinese Journal of Scientific Instrument,2020,41(5):108-117.

[3] BILLINGS G,JOHNSON-ROBERSON M. Silhonet: An rgb method for 3d object pose estimation and grasp planning [J]. ArXiv Preprint, 2018, ArXiv: 1809.06893,2018.

[4] 成超鹏,张莹,牟清萍,等. 基于改进型抓取质量判断网络的机器人抓取研究[J]. 电子测量与仪器学报,2019,33(5):80-87.
CHENG CH P,ZHANG Y, MOU Q P,et al. Research on robotic grasping based on improved grasping quality judgment network[J]. Journal of Electronic Measurement and Instrumentation,2019,33(5):80-87.

[5] REDMON J, ANGELOVA A. Real-time grasp detection using convolutional neural networks[C]. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2015: 1316-1322.

[6] CHU F J, XU R, VELA P A. Real-world multiobject, multigrasp detection[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3355-3362.

[7] 陈丹,林清泉. 基于级联式 Faster RCNN 的三维目标最优抓取方法研究[J]. 仪器仪表学报,2019,40(4):229-237.
CHEN D,LIN Q Q. Research on optimal capture method of 3D targets based on cascaded Faster RCNN [J]. Chinese Journal of Scientific Instrument,2019,40(4):229-237.

[8] 沈博. 电动吸盘式分拣机械手自动控制系统设计[J]. 计算机测量与控制,2018, 26(6):77-80.
SHEN B. Design of automatic control system for electric suction cup sorting manipulator [J]. Computer Measurement and Control,2018,26(6):77-80.

[9] 张亚辉. 基于 Faster R-CNN 目标检测的机器人抓取系统研究[D]. 北京:中国科学院大学(中国科学院深圳先进技术研究院),2019.
ZHANG Y H. Research on robotic grasping system for Faster R-CNN target detection[D]. Beijing:University of Chinese Academy of Sciences (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), 2019.

[10] 梁飒. 机器人抓取物体的三维位姿识别系统研究[D]. 天津:天津科技大学,2019.
LIANG S. Research on 3D pose recognition system of robot grabbing objects[D]. Tianjin:Tianjin University of Science and Technology,2019.

[11] WOLD S, ESBENSEN K, GELADI P. Principal component analysis[J]. Chemometrics and Intelligent Laboratory Systems,1987,2(1-3):37-52.

[12] 孙自飞. 服务机器人动态环境下定位及物体抓取技术[D]. 南京:东南大学,2017.
SUN Z F. Positioning and object grasping technology in service robotic dynamic environment[D]. Nanjing: Southeast University,2017.

[13] CHEN Z,LIM M,JIA Z,et al. Towards generalization and data efficient learning of deep robotic grasping[J]. ArXiv E-prints,2020,ArXiv:2007. 00982.

[14] REDMON J,DIVVALA S,GIRSHICK R,et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[15] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-Cnn[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[16] REDMON J, FARHADI. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.

[17] TEKIN B,SINHA S N,FUA P. Real-time seamless single shot 6d object pose prediction[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018: 292-301.

[18] WANG C,XU D,ZHU Y,et al. Densefusion:6d object pose estimation by iterative dense fusion[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:3343-3352.

[19] PENG S, LIU Y, HUANG Q, et al. Pvnet: Pixel-wise voting network for 6dof pose estimation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 4561-4570.

[20] PARK F C, MARTIN B J. Robot sensor calibration: Solving AX = XB on the euclidean group[J]. IEEE Transactions on Robotics and Automation, 1994, 10(5): 717-721.

[21] ROTH W E. The equations AX-YB=C and AX-XB=c in matrices[J]. Proceedings of the American Mathematical Society, 1952, 3(3): 392-396.

[22] SHIU Y C, AHMAD S. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB[J]. IEEE Transactions on Robotics and Automation, 1989, 5(1): 16-29.

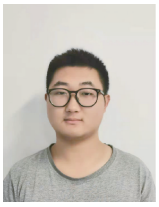
作者简介

葛俊彦,2019 年于南京大学金陵学院获得学士学位,现为江苏科技大学硕士研究生,主要研究方向为计算机视觉、机器人抓取。
E-mail: gjy@ stu. just. edu. cn

Ge Junyan received his B. Sc. degree from Nanjing University Jinling College in 2019. He is currently a master student at Jiangsu University of Science and Technology. His main research interests include computer vison and robotic grasping.

史金龙(通信作者),2012 年于复旦大学获得博士学位,现为江苏科技大学教授、硕士生导师,主要研究方向为计算机视觉和三维重建。
E-mail: shi_jinlong@ 163. com

Shi Jinlong (Corresponding author) received his Ph. D. degree from Fudan University in 2012. He is currently a professor and a master advisor at Jiangsu University of Science and Technology. His main research interests include computer vison and 3D reconstruction.



周志强,2019 年于江苏科技大学获得学士学位,现为江苏科技大学硕士研究生,主要研究方向为计算机视觉。
E-mail: 13255237623@ 163. com

Zhou Zhiqiang received his B. Sc. degree from Jiangsu University of Science and Technology in 2019. He is currently a master student at Jiangsu University of Science and Technology. His main research interest is computer vison.