

DOI: 10.19650/j.cnki.cjsi.2311531

基于多尺度特征融合和抓取质量评估的 抓取生成方法*

高翔, 谢海晟, 朱博, 徐国政

(南京邮电大学自动化学院、人工智能学院 南京 210046)

摘要:在非结构化场景中,物体的6-Dof抓取是智能服务机器人领域的一项极具挑战性的任务。在该场景中,机器人需要应对不同大小和形状的物体以及环境噪声等因素的干扰,因此难以生成准确的抓取姿态。针对此问题本文提出一种基于多尺度特征融合和抓取质量评估的6-Dof抓取姿态生成方法。首先,提出了自适应半径查询法,解决真实环境中点云采样不均匀导致的关键点查询异常的问题;其次设计了一种将多尺度特征和抓取质量评估融合的抓取生成网络,可以生成丰富的6-Dof抓取域;最后,定义了一种抓取质量评估方法,包含力闭合分数、接触面平整度、棱边分析和质心分数,并将这些标准应用在标准数据集上生成新的抓取置信分数标签,同时将这些标准融入抓取生成网络中。实验结果表明所述的方法与当前较为先进的方法FGC-GraspNet相比平均精度提升了5.9%,单物体抓取成功率提升了5.8%,多物体场景的抓取成功率提升了1.1%。综上所述,本文所提出的方法具备有效性和可行性,在单物体场景和多物体场景中具有较好的适应性。

关键词: 6-Dof 抓取; 关键点查询; 多尺度特征融合; 图卷积神经网络; 抓取质量评估

中图分类号: TP391 TH86 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Grasp generation method based on multiscale features fusion and grasp quality assessment

Gao Xiang, Xie Haisheng, Zhu Bo, Xu Guozheng

(College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

Abstract: In unstructured environments, the 6-DoF object grasp is a highly challenging task in the field of intelligent service robotics. In such scenarios, robots need to deal with interferences from objects of different sizes and shapes, as well as environmental noise, making it difficult to generate accurate grasp poses. To address this problem, this article proposes a grasp generation method based on multi-scale features fusion and grasp quality evaluation. Firstly, an adaptive radius query method is introduced to solve the issue of key points query anomalies caused by uneven point cloud sampling in real environments. Secondly, a grasp generation network is designed to fuse multi-scale features and grasp quality assessment, which enable the generation of rich 6-DoF grasp candidates. Finally a grasp quality assessment method is defined, which includes force closure score, contact surface flatness, edge analysis, and centroid score. These criteria are applied to generate new grasp confidence score labels on a standard dataset and incorporated into the grasp generation network. Compared with the current state-of-the-art method FGC-GraspNet, the experimental results show that the described method improves the average accuracy by 5.9%, the success rate of single-object grasp by 5.8%, and the success rate of multi-object scene grasp by 1.1%. In summary, the proposed method has effectiveness and feasibility, which has good adaptability in single-object scenes and multi-object scenes.

Keywords: 6-DOF grasp; key points query; multi-scale features fusion; graph convolutional neural network; grasp quality assessment

收稿日期: 2023-06-07 Received Date: 2023-06-07

* 基金项目: 国家自然科学基金(61603195)、江苏省自然科学基金(BK20140878)项目资助

0 引言

智能服务机器人在结构化场景中已经广泛应用,但随着人工智能和机器人技术的发展,机器人需要具备更智能化的能力,如快速、准确地抓取各种大小和形状的物体,以应对非结构化场景的需求。这种能力对机器人在各种环境中的应用至关重要。

当前主流抓取方法分为基于二维信息的平面抓取和基于三维信息的6自由度(6 degree of freedom, 6-Dof)抓取。传统的基于二维信息的平面抓取方法^[1]在结构化场景中,可以取得较好的效果,然而在非结构化场景中,由于彩色(red-green-blue, RGB)图片仅提供有限的场景信息因而无法完成一些灵巧操作且易受到阴影或颜色变化的干扰导致无法正确识别物体。因此,当前研究热点是将三维信息用于抓取任务,以克服上述问题。

点云是一种主要的三维场景信息表示方法,相对于其他表示方法如体素、多边形网格,点云具有更真实的几何特征、灵活的数据结构以及较小的数据存储成本。因此,以点云作为输入的6-Dof抓取越来越受到研究人员的关注。

在点云处理过程中,关键点的邻近点查询对于点云特征提取的质量有重要影响。目前常用的关键点查询方法包括K最近邻查询(K nearest neighbor query, KNN)和球查询法。然而,这些方法都存在一个共同问题,即对噪声和异常值敏感。为解决这一问题,本文提出了一种自适应半径查询法,以解决非均匀点云的关键点查询问题。

6-Dof抓取按照抓取生成方式可分为基于模型抓取和无模型抓取。在基于模型抓取中,物体的三维模型和形状信息已知,通过检测物体的六维姿态从预定义的抓取库中选择抓取^[2]。这种方法虽然避免了生成抓取的物理推导过程但是它需要获取物体完整的三维模型,在实际场景中这一步耗时且不稳定。

无模型的6-Dof抓取方法不需要关于物体的先验知识^[3]。该方法主要可以分为2类,一种是两阶段抓取方法,一种是单阶段抓取方法。两阶段抓取方法通常指先生成物体的抓取姿态,之后再行抓取质量评估并选择最优的抓取。Ten Pas等^[4]提出的方法通过点云体素化和感兴趣区域采样获取抓取点,并利用卷积神经网络(convolutional neural network, CNN)清洗抓取,以实现抓取的扩展和过滤。Peng等^[5]提出的方法首先生成抓取姿态,之后利用基于点网(PointNet)的抓取评估网络对抓取进行分类。由于涉及多个子模块协同工作,两阶段抓取方法存在拓展性差、效率低和鲁棒性不足等问题。单阶段抓取方法指可以在一个网络结构中完成抓取姿态生成和抓取质量评估,该方法简化了模型结构,降低了计算

复杂度且对于未知物体有着更好地适应性和泛化能力。因此,本文设计的抓取生成方法采用单阶段结构。近年来越来越多的科研人员将精力投入到设计单阶段抓取方法中。Fang等^[6]提出的方法采用两个并行阶段评估生成的抓取姿态,然而并没有考虑物理因素对抓取的影响。Alliegro等^[7]在预测出第1个接触点之后,回归出抓取角度和第2个抓取点,最后对抓取可执行性进行分类。由于该方法采用了简单的网络结构,所以在抓取生成时存在精度方面的问题。以上两阶段与单阶段抓取方法均采用PointNet++结构中的抽象集合(set abstraction, SA)层提取点云特征。从图卷积的角度看,它的边缘函数是 $e_{\theta}(x_i, x_j) = e_{\theta}(x_j)$,忽视了点间关系,不能很好地提取全局特征^[8]导致场景的空间信息贫乏,无法把握物体整体形状从而导致不准确的抓取。因此,本文采用多尺度点云特征提取结构解决PointNet++全局特征提取不充分的问题。

抓取质量评估是6-Dof抓取的重要部分。为评估抓取质量,传统的分析方法对物体本身和夹持器的几何形状进行物理分析。抓取旋量空间分析(grasp wrench space analysis, GWS)和力闭合度量是评估抓取质量的两种经典方法。力闭合度量要求夹持器和物体之间有摩擦,而GWS可以在没有摩擦的情况下评估抓取质量。Liang等^[9]对抓取姿态同时进行力闭合度量和GWS分析,以获得置信分数的加权和。Wen等^[10]通过观察随机抽样的抓取的稳定性来计算每个抓取的连续得分。由于被抓取物体形状的多样性和场景的复杂性,仅从力的角度去评估抓取质量是不充分的。因此,本文提出一种综合考虑物体自身物理特性和力学分析的抓取质量评估方法。

针对上述3个问题,本文提出一种基于多尺度特征融合和抓取质量评估的抓取生成方法。首先,本文设计了自适应半径查询法解决不均匀点云的关键点查询问题;其次改进了自适应图卷积(adaptive graph convolution, Adaptconv)^[11]算子并结合文献[12],设计了一种多尺度点云特征提取方法。最后设计了一种抓取质量评估方法,计算抓取姿态的置信分数。本文分别在公共数据集和真实场景中进行了实际测试,验证了上述方法的有效性。

1 抓取生成方法总体框架

抓取生成方法总体框架如图1所示。首先使用深度相机获取场景的RGB图片和深度图片,之后利用点云处理库将二者转换为点云;其次将该点云输入到本文设计的多尺度抓取质量评估抓取网络(multiscale-grasp quality assessment-GraspNet, MG-GraspNet)中生成抓取域(包含多个抓取的抓取接近方向、抓取宽度、抓取深度、平面内旋转以及抓取置信分数);最后利用抓取域中的信息解算抓取姿态。

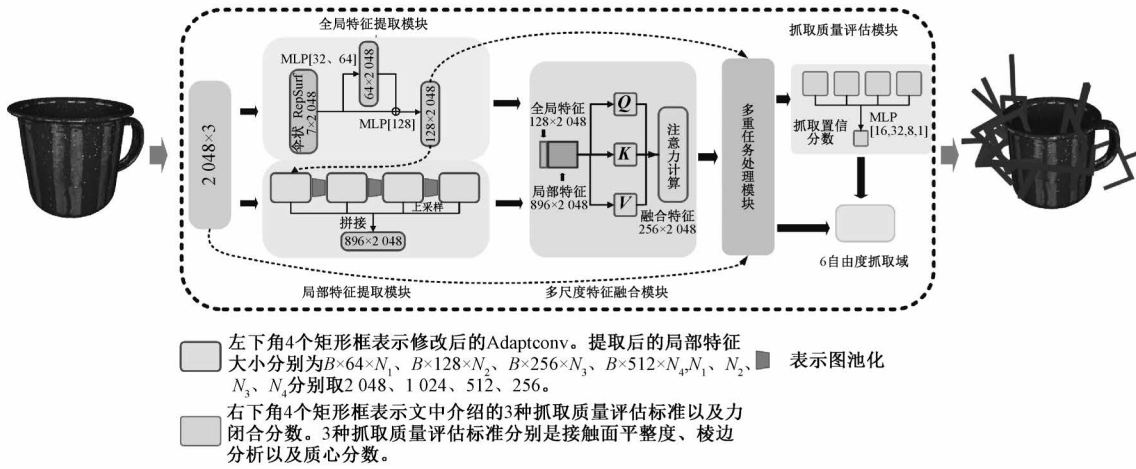


图1 抓取生成方法总体框架

Fig. 1 The overall framework of the grasp generation method

2 MG-GraspNet 抓取生成网络

本文设计的 MG-GraspNet 主要包含了全局特征提取模块、局部特征提取模块、多尺度特征融合模块、多重任务处理模块及抓取质量评估模块。其中，全局特征提取模块负责提取整个点云的全局特征，局部特征提取模块将点云划分为多个局部区域，提取每个区域的特征。该模块采用串联结构提取4层局部特征，并通过上采样和拼接操作将局部特征与全局特征融合。多尺度特征融合模块进一步融合不同尺度的特征信息，以获取更加全面的特征表示。最后，经过多重任务处理模块、抓取质量评估模块得到最终的抓取域。

在本文的2.1节中，将介绍自适应半径查询法来解决非均匀点云关键点查询问题。该方法在上述各个模块中都得到了应用，并发挥了重要的作用。

2.1 自适应半径查询法

本文针对真实环境中由于点云采样不均匀导致的关键点查询异常问题，提出了一种自适应半径查询法。该方法根据点云密度动态调整搜索半径，能够准确提取关键点并适应不同形状和复杂度的物体表面。通过考虑点云密度的变化，该方法能够有效降低噪点干扰，并捕捉到更完整的物体形状和结构信息。该自适应半径查询法在各模块中应用，为点云关键点查询提供了一种有效灵活的解决方案。

该方法的计算方式如下：首先计算出半径为 r 的球中，中心点与球内各点的距离之和。具体计算如下：

$$d_i = \sum_{j=1}^N \| \mathbf{p}_i - \mathbf{p}_j \|^2 \quad (1)$$

$$\hat{d}_i = d_i \times \frac{1}{\max dg}, \quad g = 1, 2, \dots, k \quad (2)$$

其中， \mathbf{p}_i 表示查询中心点， \mathbf{p}_j 表示半径 r 区域内的点； N 为半径 r 区域内共有 N 个点； K 表示 K 个中心点。之后，计算出一个归一化的密度 \hat{d}_i ， g 表示获取到的中心点。基于这个归一化的密度，将查询半径 r ($r < R$) 放缩为 \hat{r}_i ，其计算公式如下：

$$\hat{r}_i = \sqrt{r^2 \left(1 + \left(\text{sigmoid} \left(\hat{d}_i - \frac{1}{2} \right) - \frac{1}{2} \right) \right)} \quad (3)$$

2.2 局部特征提取模块

文献[11]中，作者介绍了 Adaptconv 算子提取点云的局部特征。该算子可以动态调整邻点权重，提高算法的适应性和泛化能力。然而，该算子仅计算了特征核而没有计算坐标核，这会导致模型无法准确感知点云的空间结构和几何特征。因此，本文对文献中的 Adaptconv 算子进行了一些调整：1) 将关键点的查询方式从 KNN 替换为 2.1 节中所介绍的自适应半径查询法；2) 对 Adaptconv 算子的结构进行了修改。如图 2 所示，首先根据点的特征生成一个特征核，之后依据点的坐标再为这些点生成一个坐标核，将这两个核心进行点积得到一个全新的自适应核。之后将该自适应核作用在点对 $(\mathbf{x}_i, \mathbf{x}_j)$ 上，以获取图中每条边的权值。修改的部分可以表示如下：

$$\boldsymbol{\alpha}_{ijm} = \gamma_j(\Delta \mathbf{x}_{ijm}), \quad jm \in N(i), m = 1, 2, \dots, k \quad (4)$$

$$\Delta \mathbf{x}_{ijm} = \text{contact}[\mathbf{x}_i \cdot \mathbf{x}_{jm} - \mathbf{x}_i] \quad (5)$$

$$\mathbf{d}_{ijm} = \boldsymbol{\alpha}_{ijm} \cdot \boldsymbol{\beta}_{ijm} \quad (6)$$

$$\mathbf{f}_i' = \max_{j \in N(i)} (\text{Relu}(\mathbf{d}_{ijm})) \quad (7)$$

式中： $\boldsymbol{\alpha}_{ijm} \cdot \boldsymbol{\beta}_{ijm}$ 分别表示坐标核和特征核； $\gamma_j(\cdot)$ 表示特征映射函数； $N(i)$ 表示的是第 i 个点周围的邻点； m 表示周围的第 m 个点；因为式(6)采用的是点乘操作，所以需要坐标核和特征核的大小进行调整。最后经过激活函数和聚合函数得到最终的输出特征 \mathbf{f}_i' 。

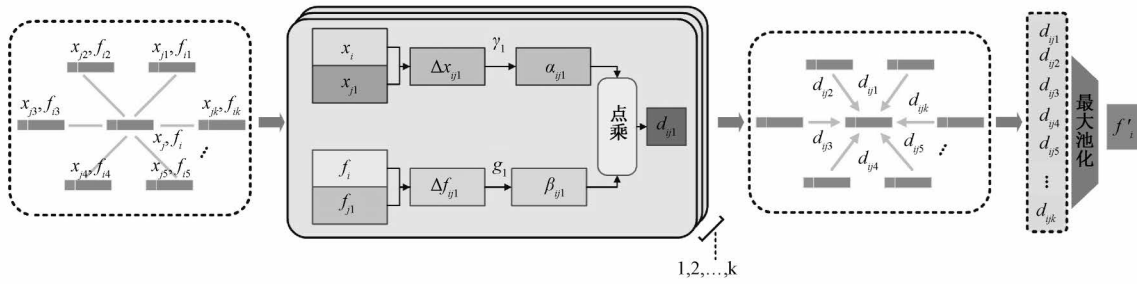


图2 局部特征提取模块

Fig. 2 Local features extraction module

本文采用一种串联结构提取点云的多尺度局部特征。对于输入点云需要提取4层局部特征:第1层局部特征 F_{l1} 大小为 $B \times 64 \times N$,第2层局部特征 F_{l2} 大小为 $B \times 128 \times N_1$,第3层局部特征 F_{l3} 大小为 $B \times 256 \times N_2$,第4层局部特征 F_{l4} 大小为 $B \times 512 \times N_3$ 。 N, N_1, N_2, N_3 分别为 2 048、1 024、512、256。

2.3 全局特征提取模块

为了从点云中提取出全局特征,本文希望从点云的物理性质出发,并将其映射到高维空间后得到全局特征。文献[12-13]中介绍了多种点云物理特征表示方法,本文采用了文献[12]中的方法,将点云表示为伞状表面表示(surface representation, RepSurf)坐标。该方法首先需要已知点云的法向量,通过法向量和点的坐标计算出经过该点的平面,将平面方程中的常数项与法向量一起定义为点云的三角 RepSurf 坐标。具体计算过程如下^[12]:

给定一组法向量 $v_i = (a_i, b_i, c_i)$, 和点 $x_{pi} = (x_i, y_i, z_i)$, 定义表面位置 $p_i = a_i x_i + b_i y_i + c_i z_i$, 则定义的三角 RepSurf 为:

$$t_i = (a_i, b_i, c_i, p_i) \tag{8}$$

之后获取该点周围 K 个点的坐标,并将它们与三角 RepSurf 坐标 t_i 连接,经过转换函数(使用多层感知机)和聚合函数(使用的是最大池化层)最终得到伞状 RepSurf 坐标。伞状 RepSurf 坐标计算可表示为:

$$u_i = \sum_{j=0}^K \{T(x'_{ij}, t_{ij})\} \tag{9}$$

该方法可以显式地表达物体的局部形状,伞状 RepSurf 坐标 $u_i \in R^7$ (7 是指 3 个 3 维坐标、3 个法向量以及表面位置)。本文将伞状 RepSurf 坐标理解为全局特征的表示方法,而不是像文献[12]中那样将其理解为物体的坐标。因此在获取伞状 RepSurf 坐标 u_i 之后,使用两个一维卷积层首先将 u_i 映射到 32 维,之后映射到 64 维,并将该 64 维特征与 u_i 拼接,最终经过一个一维卷积层,获得 128 维全局特征。

2.4 多尺度特征融合模块

多尺度特征融合模块如图3所示,将多尺度局部特征和全局特征融合,可以有效提取物体大小和形状信息。在获取到多尺度局部特征和全局特征之后,首先需要将二者拼接。其中关键问题是将每层特征点进行上采样。

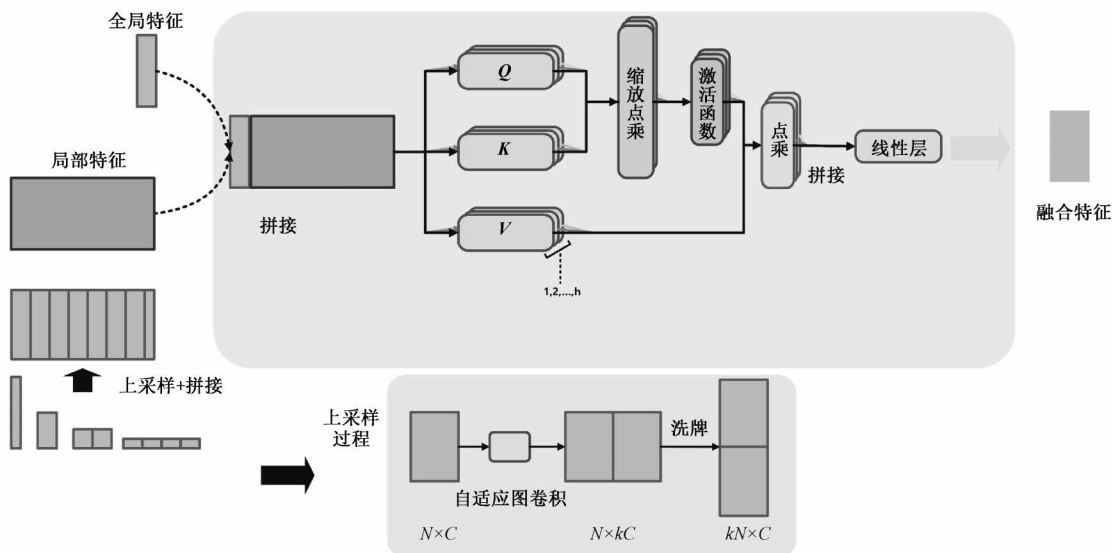


图3 多尺度特征融合模块

Fig. 3 Multi-scale features fusion module

文献[14]中采用图卷积神经网络的方式拓展它的特征维度,再将它们洗牌后得到上采样后的特征值。本文借鉴了文献[14]中的思想,利用图卷积网络对特征进行上采样。本文将文章中使用的图卷积神经网络更换为2.2节中介绍的修改后的 Adaptconv 算子,同时在上采样时只需要使用坐标核作为其自适应核。

得到拼接的特征之后,采用多头 Transformer 结构来更好地融合点云的特征,如图3所示。此结构以拼接后的大小为 $B \times (128 + 64 + 128 + 256 + 512) \times N$ 的特征 (F^-) 为输入,经过多头 transformer 之后,再经过一个一维卷积层得到最终输出 F_a ,其大小为 $B \times 256 \times N$ 。多头注意力可以形成多个子空间,使模型能够关注空间与几何中不同方面的信息。具体公式如下^[15]:

$$\begin{aligned} Q_m &= F^- W_{QPm} \\ K_m &= F^- W_{KPm} \\ V_m &= F^- W_{VPm} \\ M_m &= \text{softmax} \left(\frac{QK^T}{\sqrt{\frac{d}{n}}} \right) \\ F_m &= M_m V, F = \text{concat}(F_1, \dots, F_n) \\ F'_a &= \text{LayerNorm}(F^- + \varepsilon(F_m)) \\ F''_a &= \text{LayerNorm}(F^- + \varepsilon(F'_a)) \\ F_a &= 1D\text{conv}(F''_a) \end{aligned} \quad (10)$$

式中: W_{QPm} 、 W_{KPm} 、 W_{VPm} 为可学习权重矩阵, m 为第 m 个头,共有 n 个头, B 是可学习的位置编码矩阵, $\varepsilon(\cdot)$ 指的是线性层。

2.5 多重任务处理模块

该部分的输入是经过降采样后的点云、全局特征 F_c 以及融合后的特征 F_a 。该网络首先利用降采样后的点云和全局特征经过3个一维卷积层输出旋转矩阵大小为 $B \times 2048 \times 3 \times 3$ 。

之后利用降采样后的点云、输出特征对抓取姿态进行预测。在该步骤中,以降采样点云 $B \times 2048 \times 3$ 为中心在其自身中利用上文中使用的自适应半径查询法查询周围的 k 个点;将得到的查询序列在输出特征中查找到与之对应的特征,其大小为 $B \times 2048 \times k \times 256$ 。之后输入到 SharedMLP^[16] 层中解决点云的无序性问题,输出的维度为 $B \times 2048 \times k \times 256$ 。同时,为了加强对局部特征的关注,本文在 SharedMLP 层之后采用了多头 transformer 结构来提取该区域内的上下文信息,最后通过最大池化层得到该关注区域内最明显的特征,将其再经过3个一维卷积层以得到最终输出,该模块的输出大小为 $B \times 2048 \times 16$ 。16 分别表示力闭合分数 S_f (1 维)、质心分数 $S_{gravity}$ (1 维)、抓取宽度 (1 维)、抓取深度 (1 维)、抓取中心点坐标 (3 维)、旋转矩阵 (9 维)。

2.6 抓取质量评估模块

本模块的输入来自多重任务处理模块的输出,包括力闭合分数 S_f 、质心分数 $S_{gravity}$ 、抓取宽度、抓取深度、抓取中心点坐标、旋转矩阵,并将这些参数解析为左右抓取点和抓取姿态。利用本文提出的抓取质量评估方法可以计算出 $S_{flatness}$ 和 S_{cosdis} 两种评估标准的分数。具体计算方法将在第3章中介绍。然后,将 $S_{flatness}$ 、 S_{cosdis} 以及预测得到的 $S_{gravity}$ 、 S_f 作为输入送入4层一维卷积结构中,从而获得最终的抓取置信分数。这些一维卷积层将4个输入维度的分数分别映射到16维、32维、8维和1维。通过多层一维卷积结构,可以细化权重分配,学习各个物理特性之间的内在联系,从而提高抓取置信分数的准确性。网络最终输出的大小为 $B \times 2048 \times 15$ 。15 维分别表示抓取置信分数 (1 维)、抓取宽度 (1 维)、抓取深度 (1 维)、抓取中心点坐标 (3 维)、旋转矩阵 (9 维)。

2.7 损失函数

由于预测同时存在回归问题和分类问题,所以需要设计一种混合损失函数来学习这种预测结果。多重任务损失可以表示如下:

$$L = \mu(L_{obj} + L_v) + \theta(L_w + L_s + L_r + L_d) \quad (11)$$

式中: L_{obj} 表示是否是一个物体的分类损失, L_v 、 L_w 、 L_s 、 L_r 、 L_d 表示接近方向、抓取宽度、抓取分数、平面内旋转以及抓取深度的回归损失。上述的回归损失只在相关点在物体上时计算,本文中的分类损失采用两类的 softmax 损失^[6],回归损失采用 smooth-l1 损失,并采用 Adam 优化器优化模型参数。

3 抓取质量评估方法

为了全面评估抓取质量,本文提出了一种综合考虑物体自身物理特性和力学分析的抓取质量评估方法,包括以下评价指标:接触面平整度、棱边分析、质心分数以及力闭合^[6]分数 S_f 。

3.1 接触面平整度

接触点附近的接触面越平缓,抓取成功的可能性就越高。因此本文利用该特征作为抓取质量评估方法的判别标准之一。首先需要从接触点及其附近的点中拟合出一个平面,之后计算出查询范围内所有点到拟合平面的距离,最后取距离的平均值作为接触面平整度分数,记为 $S_{flatness}$ 。该计算步骤如下:

首先,在左右两个抓取点附近采用2.1节中介绍的自适应半径查询法获取抓取点周围的 k 个点。之后使用随机抽样一致性 (random sample consensus, Ransac) 算法在 k 个点中拟合出一个平面 P_c ,并计算出这些点到平面

P_c 的距离,最后取得平均值,具体计算公式如下:

$$P_{ci} = \text{Ransac}(A(\mathbf{p}_i)), \mathbf{p}_i \in (\mathbf{p}_l, \mathbf{p}_r) \quad (12)$$

$$S_{flatness} = \frac{1}{2k} \sum_{i=1}^2 \sum_{j=1}^k \|\mathbf{p}_j, P_{ci}\|_2 \quad (13)$$

式中: $A(\cdot)$ 表示自适应半径查询法, $(\mathbf{p}_l, \mathbf{p}_r)$ 表示左右抓取点, $\|\cdot, \cdot\|_2$ 表示点到平面的距离。

3.2 棱边分析

真实环境中抓取点附近存在棱边会影响抓取的稳定性,因此,本文将该特性作为抓取质量评估的判别标准。

本文采用法线的余弦距离来度量抓取点附近棱边特性并生成相应余弦分数 S_{cosdis} 。首先获取左右抓取点的邻近点序列,之后通过用主成分分析法 (principal component analysis, PCA) 计算出点云的法向量集合,并在法向量集合中依据先前计算的序列获取邻近点的法向量同时计算出各个点之间的法向量余弦距离,最后为了使得分数为正值,使用了归一化函数,具体的公式如下:

$$g_{ci} = A(\mathbf{p}_i), \mathbf{p}_i \in (\mathbf{p}_l, \mathbf{p}_r)$$

$$\mathbf{n}_{ci} = \text{PCA}(g_{ci})$$

$$S_{cosdis_dir} = \frac{2}{k(k-1)} \sum_{i=0}^k \sum_{j=0}^k \left(\frac{1}{2} + \frac{1}{2} \cdot \frac{\mathbf{n}_i \cdot \mathbf{n}_j}{\|\mathbf{n}_i\| \cdot \|\mathbf{n}_j\|} \right),$$

$i \geq j$

$$S_{cosdis_dir} \in (S_{cosdis_l}, S_{cosdis_r})$$

$$S_{cosdis} = \frac{1}{2} (S_{cosdis_l}, S_{cosdis_r}) \quad (14)$$

式中: g_{ci} 为获取到的邻近点序列。

3.3 质心分数

当抓取点距离物体重心越近时抓取越稳定,因此本文利用欧氏距离将这种关系显式表达。本文使用两步来量化这种物理关系。该方法通过计算抓取点和重心之间的夹角以及连接线的欧氏距离来评估该特性。对于规则物体而言当重心夹角越大时,重心距离两个抓取点越近;同时直接计算点到连接直线的距离来衡量距离的远近。首先获取到重心角 θ_g , 计算出一个角度分数 $S_\theta = (1 - \cos(\theta_g))$; 然后计算出重心点到左右抓取点连接直线的欧氏距离 S_{dis} , 并使用乘法的形式将这两个计算标准进行组合。其具体计算过程如下:

$$\cos\theta_g = \frac{\|\mathbf{p}_l - \mathbf{p}_c\|_2^2 + \|\mathbf{p}_r - \mathbf{p}_c\|_2^2 - \|\mathbf{p}_l - \mathbf{p}_r\|_2^2}{2 \cdot \|\mathbf{p}_l - \mathbf{p}_c\|_2 \cdot \|\mathbf{p}_r - \mathbf{p}_c\|_2}$$

$$S_\theta = (1 - \cos(\theta_g)) \quad (15)$$

$$S_{dis} = 1 - \frac{\|(\mathbf{p}_l - \mathbf{p}_c) \times (\mathbf{p}_r - \mathbf{p}_c)\|_2}{\|\mathbf{p}_l - \mathbf{p}_r\|_2}$$

$$S_{gravity} = S_\theta \cdot S_{dis}$$

式中: $\mathbf{p}_l, \mathbf{p}_r, \mathbf{p}_c$ 分别为左抓取点、右抓取点以及重心的三维坐标, $\|\cdot\|_2$ 为向量的二范数。由于在实际情况下,真实场景中获取的点云是不完整的,在线测量质心会存在误差,因此该计算方法只在生成抓取标签时使用,在网络中选择直接预测。

3.4 抓取质量评估标准融合

本文设计的抓取质量评估方法融合了力闭合度量分数和 3.1~3.3 节介绍的多种几何度量标准,最终置信分数如下:

$$S = \alpha \cdot S_f + \beta \cdot S_{flatness} + \gamma \cdot S_{cosdis} + \delta \cdot S_{gravity} \quad (16)$$

本文利用固定的 $\alpha, \beta, \gamma, \delta$ 在 Graspnet-1billion 数据集中生成利用以上抓取质量评估方法的抓取置信分数标签;同时在 MG-GraspNet 中的抓取质量评估模块中利用 4 层一维卷积层针对不同场景不断细化这些参数。

4 实验结果与分析

4.1 抓取平台搭建

实物验证环境搭建如图 4 所示,使用 Kinova Jaco2 机械臂, Kinect V2 深度相机,网络训练和评估所使用的数据集为 GraspNet-1Billion 抓取数据集。实验所使用的硬件平台的 CPU 为英特尔 Xeon (R) Silver 4110 CPU 2.10 GHz x 32, GPU 为 Quadro P4000;软件环境为 CUDA 10.2。实验中,本文将参数 $\alpha, \beta, \gamma, \delta, \mu, \theta$ 分别设置为 0.8、0.08、0.06、0.06、0.95 和 0.5。



图4 实验环境搭建

Fig. 4 Setting up the experimental environment

4.2 抓取网络生成实验

本实验使用了 Graspnet-1billion 数据集,并采用与文献[6]相同评估标准。首先通过姿态非最大抑制和碰撞检测技术筛选出排名前 50 的抓取样本并获取它们的真实分数标签。将 Precision@k 作为评估指标,表示前 k 个抓取中的正确抓取占比。之后计算在摩擦系数 μ 条件下, k 范围从 1~50 的 Precision@k 值,并计算它们的平均值,表示为 AP μ 。摩擦系数 μ 的设定方法与文献[5]保持一致。最后,计算所有 AP μ 的平均值,得到平均精度 mAP。网络的输入是在完整场景点云中随机选择 15 k

个点,初始学习率设置为0.001,批次大小设置为2,学习率在第12个epoch后下降到0.0001,第16个epoch后下降到0.00001,第20个epoch后下降到0.000001。本文模型与多种模型对比的实验结果如表1所示。模型*a*

表示本文设计的方法,模型*b*表示文献[6]中方法,模型*c*表示文献[17]中方法,本文设计的网络模型*a*相较于模型*b*、*c*在3个类别的平均精度方面平均提升分别为25.1%、5.9%,证明了本文模型的有效性。

表1 不同模型在数据集中的表现

Table 1 Different models behaving in the dataset

%

方法	已知			相似			未知		
	mAP	AP0.3	AP0.7	mAP	AP0.3	AP0.7	mAP	AP0.3	AP0.7
文献[6]方法,模型 <i>b</i>	31.34	28.21	11.14	28.23	19.35	7.23	21.21	8.42	2.83
文献[17]方法,模型 <i>c</i>	41.33	40.23	24.12	31.79	28.95	13.83	23.22	11.52	3.21
本文方法,模型 <i>a</i>	44.01	40.11	25.81	33.36	30.17	13.30	24.73	12.28	4.59

注:数据采集设备选用 Kinect,GQA 表示抓取质量评估

4.3 消融实验

为了证明本文所设计的自适应半径查询法、抓取质量评估模块和多尺度特征融合模块的有效性,本文设计了具有针对性的消融实验。消融实验结果如表2

所示。由表2可以得出,多尺度特征融合模块、自适应半径查询方法以及抓取质量评估模块均会对本文提出的网络模型产生正面的影响且影响力依次增大。

表2 消融实验结果

Table 2 Ablation experiment results

%

方法	已知			相似			未知		
	mAP	AP0.4	AP0.8	mAP	AP0.4	AP0.8	mAP	AP0.4	AP0.8
本文方法(传统球查询)	40.31	37.24	24.73	31.32	26.34	13.12	24.37	10.29	2.87
本文方法(直接拼接特征)	42.63	39.47	22.24	32.11	28.92	13.97	23.12	11.44	2.71
本文方法(不使用GQA)	40.07	38.31	23.34	31.26	27.37	12.95	24.32	10.58	3.42
本文方法	44.01	40.11	25.81	33.36	30.17	13.30	24.73	12.28	4.59

4.4 单物体场景抓取实验

为了验证本文设计的抓取生成方法的有效性,首先进行了针对单一物体的抓取实验。同时为了评估模型在日常复杂环境下的鲁棒性,本文将抓取背景分别设置为平整的桌面和粗糙的毛毯。本实验中,采用6种生活中常见的物品,每一种物品进行25次实验,每次实验时,物品分别放置为不同的位置和不同的姿态。模型*a*、*b*和*c*对比实验结果如表3所示,单物体抓取姿态生成部分对比结果如图5所示。本文将生成的抓取映射到二位平面,图5中显示了评分最高的抓取。实验结果表明,在无背景的条件下,本文所设计的网络抓取成功率总体优于目前先进的抓取模型;在有背景的条件下,网络总体表现出较强的鲁棒性。但是洗发水类的抓取成功率较低,其原因是洗发水横向放置时,该场景中的部分毛毯褶皱高于洗发水横置高度且与洗发水贴合,模型将毛毯褶皱误认为是洗发水的一部分,导致抓取成功率下降。本文设计的模型在无背景和有背景场景下的平均抓取成功率相较于模型*b*、*c*分别提升14.0%、5.8%。

4.5 多物体场景抓取实验

为验证本文方法在多物体场景下的抓取有效性,进行了多物体场景下的抓取实验。共设5种场景,每个场景加入4个不同的物品,并以将物体抬高10cm作为抓取是否成功为标准。实验中每个场景进行10次测试,每次测试物品的位置和姿态都不同。每次测试生成40个抓取,计算每个物体的抓取成功率,并利用加权平均的方式计算该场景的抓取成功率。将同一场景下10次测试的抓取成功率平均值作为该场景的抓取成功率。3种模型的对比实验结果如表4所示,每个场景排名前15的抓取结果如图6所示。对比结果显示,本文设计的网络模型在多物体场景中总体抓取成功率较高,相较于模型*b*、*c*分别提升2.9%、1.1%。原因是本文的网络模型在提取特征信息时充分考虑了全局信息,增强了点与点之间的相关性,从而满足了抓取任务的需求。具体的物体编号及对应关系如表5所示。

表3 单物体场景抓取实验结果

Table 3 Results of single-object scene grasp experiment

物体名称	苹果			水杯			香蕉			盒装牛奶			洗发水			洗洁精		
模型	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
无背景成功次数	21	19	20	19	18	19	20	19	20	22	20	21	19	17	19	22	17	19
有背景成功次数	20	17	19	19	17	17	18	15	17	19	16	16	17	15	19	17	15	15
无背景抓取成功率/%	84	76	80	76	72	76	80	76	80	88	80	84	76	68	76	88	68	76
有背景抓取成功率/%	80	68	76	76	68	68	72	60	68	76	64	64	68	60	76	68	60	60

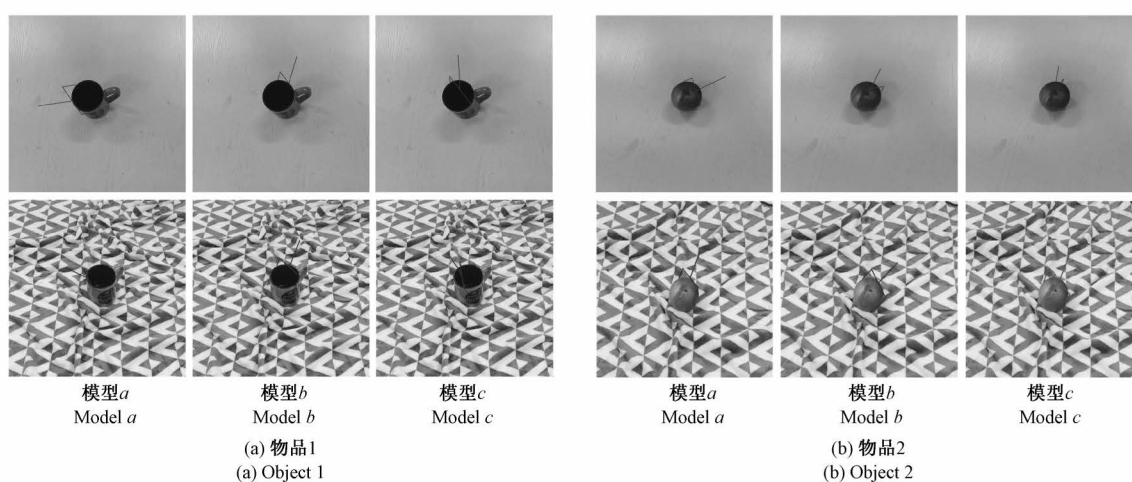


图5 单物体场景抓取实验

Fig. 5 Single object grasp experiment

表4 多物体场景抓取实验结果

Table 4 Results of multi-object scene grasp experiment

场景编号 (物品编号)	场景1 (1,2,3,4)			场景2 (12,13,14,15)			场景3 (3,4,5,6)			场景4 (5,6,7,8)			场景5 (1,9,10,11)		
模型	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
无背景抓取成功率/%	80.5	78.5	78.5	82.3	79.5	81.3	82.3	78.5	79.5	80.5	78.5	81.5	82.5	79.8	81.5
有背景抓取成功率/%	78.3	76.8	77.5	81.8	78.8	81.3	79.3	78.3	80.5	80.0	77.5	78.5	79.3	77.8	78.3

表5 物体编号对应关系

Table 5 Object number correspondence

ID	物品名称	ID	物品名称
1	咖啡冻	9	水杯(含把手)
2	香飘飘奶茶	10	苹果
3	桶装薯片	11	盒装牛奶
4	盒装面包	12	抽纸
5	喷水壶	13	橘子
6	瓶装饮料	14	眼镜盒
7	袋装薯片	15	药品盒
8	洗洁精		

4.6 抓取分数对比实验

为了得到预测分数与实际抓取分数的差距,本文在单物体场景下进行了该实验。该实验一共使用了两个不同的物体,每个物体进行10次测试(物体放置于不同的位置和姿态),每次测试中获取最高得分的抓取姿态。利用该抓取姿态计算出实际抓取分数,并与预测分数进行比较,结果如表6所示。

4.7 遮挡抓取实验

为了证明本文所提出的抓取生成网络模型仅利用场景点云提取特征,本实验将物体分别放置在毛毯上以及毛毯下进行抓取实验。本实验使用3个形状大小各不相

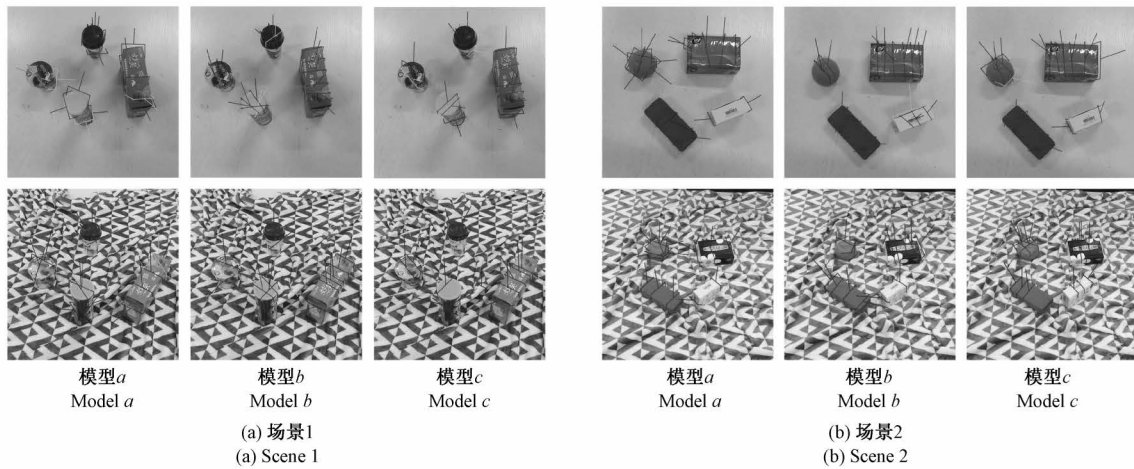


图 6 多物体场景抓取实验
Fig. 6 Multi-object grasp scene experiments

表 6 抓取分数对比实验

Table 6 Grasp score comparative experiment

物体名称	苹果	洗洁精
实验次数	10	10
预测分数均值	0.444 3	0.492 6
真实分数均值	0.421 2	0.522 4

表 7 遮挡抓取实验结果

Table 7 Occlusion grasp test results

物体名称	苹果	桶装薯片	面包包装盒
实验次数	25	25	25
地毯上抓取成功率/%	88	84	92
地毯下抓取成功率/%	80	76	80

同的物体,在毛毯上和毛毯下分别放置 25 种不同的位置与姿态,选择抓取分数最高的抓取并计算抓取成功率。抓取生成结果如图 7 所示,实验数据如表 7 所示。结果表明在毛毯上和毛毯下的成功率差距较小,均可以较好地抓取物品,不会受到毛毯颜色以及图案的影响。

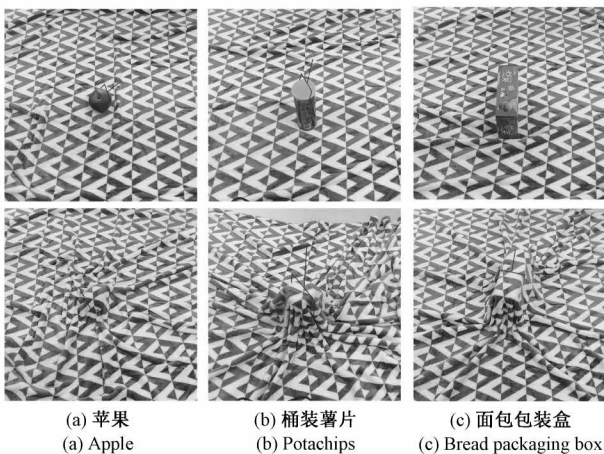


图 7 遮挡抓取实验
Fig. 7 Occlusion grasp test results

5 结 论

本文提出了一种基于多尺度特征融合和抓取质量评估的抓取生成网络。该网络旨在解决机器人在面对不同大小、形状的物体以及环境噪声等干扰因素时,难以生成准确的抓取姿态的问题。该网络使用多尺度特征融合模块可以有效提取物体大小和形状信息,同时将多尺度特征融合模块与抓取质量评估模块结合可以有效解决环境噪声的干扰。通过在公共数据集和真实环境中进行实验验证,证明了本文所设计的网络具有良好的泛化能力,并能够成功地完成对场景中未知物体的抓取生成任务。实验结果验证了本文模型在不同环境和未知物体上的鲁棒性和适应性。

综上所述,本文提出的基于多尺度特征融合和抓取质量评估的抓取生成方法可以较好地完成抓取任务。但是本文所提出抓取生成网络利用改进的 Adaptconv 算子提取局部信息,由于需要计算节点之间的距离并确定邻点之间的权重,在处理大规模点云时可能会导致额外的计算复杂度。未来将借鉴分组卷积和参数共享的思想,改进该算子,实现更高效的抓取生成方式。

参考文献

- [1] 朱江,杜瑞,李建奇,等. 基于注意力机制的曲轴瓦盖上料机器人视觉定位和检测方法[J]. 仪器仪表学报,2021,42(5):140-150.
- ZHU J, DU R, LI J Q, et al. Visual positioning and detection method of crankshaft cover feeding robot based on attention mechanism[J]. Chinese Journal of Scientific Instrument, 2021,42(5):140-150.
- [2] DENG X, XIANG Y, MOUSAVIAN A, et al. Self-supervised 6d object pose estimation for robot manipulation [C]. IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020: 3665-3671.
- [3] PENG G, LIAO J, GUAN S, et al. A pushing-grasping collaborative method based on deep Q-network algorithm in dual viewpoints [J]. Scientific Reports, 2022, 12(1): 3927.
- [4] 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.
- [5] PENG G, REN Z, WANG H, et al. A self-supervised learning-based 6-DOF grasp planning method for manipulator [J]. IEEE Transactions on Automation Science and Engineering, 2021, 19(4): 3639-3648.
- [6] FANG H S, WANG C, GOU M, et al. Graspnet-1 billion: A large-scale benchmark for general object grasping[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11444-11453.
- [7] ALLIEGRO A, RUDORFER M, FRATTIN F, et al. End-to-end learning to grasp via sampling from object point clouds[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 9865-9872.
- [8] 黎碧玉,张旭东,孙锐,等. 基于椭球拟合的点云旋转不变网络[J]. 电子测量与仪器学报,2022,36(9): 111-117.
- LI SH Y, ZHANG X D, SUN R, et al. Point cloud rotation invariant network based on ellipsoid fitting[J]. Journal of Electronic Measurement and Instrumentation, 2022,36(9):111-117.
- [9] LIANG H, MA X, LI S, et al. Pointnetgpd: Detecting grasp configurations from point sets [C]. International Conference on Robotics and Automation (ICRA), IEEE, 2019: 3629-3635.
- [10] WEN B, LIAN W, BEKRIS K, et al. Catgrasp: Learning category-level task-relevant grasping in clutter from simulation [C]. International Conference on Robotics and Automation (ICRA), IEEE, 2022: 6401-6408.
- [11] ZHOU H, FENG Y, FANG M, et al. Adaptive graph convolution for point cloud analysis[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4965-4974.
- [12] RAN H, LIU J, WANG C. Surface representation for point clouds [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 18942-18952.
- [13] XIANG T, ZHANG C, SONG Y, et al. Walk in the cloud: Learning curves for point clouds shape analysis[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 915-924.
- [14] QIAN G, ABUALSHOUR A, LI G, et al. Pu-gcn: Point cloud upsampling using graph convolutional networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11683-11692.
- [15] ENGEL N, BELAGIANNIS V, DIETMAYER K. Point transformer[J]. IEEE Access, 2021, 9: 134826-134840.
- [16] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [17] LU Y, DENG B, WANG Z, et al. Hybrid physical metric for 6-DoF grasp pose detection[C]. International Conference on Robotics and Automation (ICRA), IEEE,

2022: 8238-8244.

作者简介



高翔, 1990年于东南大学获得学士学位, 2005年于东南大学获得硕士学位, 2007年于东南大学获得博士学位, 现为南京邮电大学教授, 主要研究方向为机器人传感技术、服务机器人。

E-mail: gaox@njupt.edu.cn

Gao Xiang, received her B. Sc. degree in 1990 from Southeast University, received her M. Sc. degree from Southeast University in 2005, received her Ph. D. degree from Southeast University in 2007. She is currently a professor at Nanjing University of Posts and Telecommunications. Her main research interests include robot sensing technology and service robotics.



朱博(通信作者), 2004年于山东大学获得学士学位, 2007年于河海大学获得硕士学位, 2014年于东南大学获得博士学位, 2016年于新加坡国立大学担任博士后研究员, 现为南京邮电大学讲师, 主要研究方向为机器人环境感知、机器人视觉和语义映射。

E-mail: zhuboseu@163.com

Zhu Bo (Corresponding author), received his B. Sc. degree in 2004 from Shandong University, received his M. Sc. degree in 2007 from Hohai University, received his Ph. D. degree from Southeast University in 2014, worked as a postdoctoral researcher at the National University of Singapore in 2016. He is currently a lecture at Nanjing University of Posts and Telecommunications. His main research interests include robot environment perception, robot vision, and semantic mapping.