

基于图神经网络的机器人抓取不可见的特定目标

朱家辉

摘要

在杂乱的场景中抓取物体是机器人操控的一个重要问题，当目标物体在完全遮挡或者周围没有合适的抓取位置时，基于深度强化学习的推和抓取动作的协同是一种有效的方法。在探索不可见的物体时，许多现有方法过分依赖模型设计和冗余的抓取动作，我们提出一个基于图的深度强化学习模型，以有效地探索不可见的物体，提高推和抓取的协同能力。我们的模型首先提取状态特征，然后根据是否发现目标物体，用不同的图 Q-Nets 估计 Q 值，图 Q-Nets 包含编码器、图推理模块和解码器，编码器用于整合状态特征，使一个区域的特征包括相邻区域的特征，图推理模块通过图卷积网络获取不同区域之间特征的内部关系，解码器将处理后的特征映射回原始状态特征大小，而且算法将目标物体的描述输入给 CLIPSeg 模型处理得到目标物体的掩码。我们的方法比原有的方法效果更好，在推抓协同任务中取得了 45% 的成功率。

关键词：机器人操控，深度强化学习，图神经网络，推抓协同，语言指导

1 引言

挑选杂质是工业生产线上典型场景之一，比如将煤炭生产线上的板岩与煤分开，但人工挑选既费时又费力，在这种情况下，机器人很有必要代替人类来完成工作。虽然机器人抓取在生产线上比以往得到更加广泛的应用，但现有大多数算法只能在结构化的环境中工作，无法处理动态情况，特别是在密集的混乱环境中抓取目标物体^[1]。

最近，面向目标的推动和抓取在机器人领域引起了极大的关注。^[2]主要关注于抓取一个特定的语义类别的物体上，这称为语义抓取。为了快速适应面向目标的任務，必须有效地结合目标信息，如目标掩码^[3]，但这通常是通过预训练的分割模块和昂贵的标签获得的，而在抓取杂质的问题上，很难预先定义所有杂质的类别来获取分割掩码。

抓取杂质的挑战不仅在于感知，还在于操控能力。在早期的研究中，操控能力的问题主要通过定制特征和预定义的动作序列来解决^[4]，硬编码的启发式方法限制了可以进行多个原始动作之间的协同动作。最近，深度强化学习被引入到机器人操控中，许多研究都集中在通过自监督的试验和错误来训练多种原始动作的策略，尽管已经取得了很大的进展，但协调不同原始动作后应用与面向目标抓取时，仍有许多挑战有待解决，例如，在没有目标的区域内进行无用的推动和抓取动作，导致效率变低。

在这篇报告中，我们提出了一种面向目标的推抓算法，该算法可以在密集的环境中主动发现并抓取目标，推抓动作之间的协同作用如图 1 所示，主要贡献总结如下：

- 我们提出了一种端到端的基于图的深度强化学习方法来学习探索完全被遮挡的目标对象的动作策略，即通过编码器、图推理模块和解码器建立图 Q-Nets (Graph Q-Nets) 来学习状态特征并预测 Q 值。在结果热图中，本方法比之前方法在目标对象周围的热图颜色更深，代表置信度更高，表明图 Q-Net 通过在不同区域中学到更精确的特征来帮助完成任务。

- 我们结合了一个用文本指导机器人抓取目标物体的分割模块，通过语言表达出我们想要抓取的目标，使系统更加清晰、方便了解到目标对象。
- 我们使用 EfficientNetV2 作为特征提取主干网络，提高推理速度，使机器人更快完成任务。

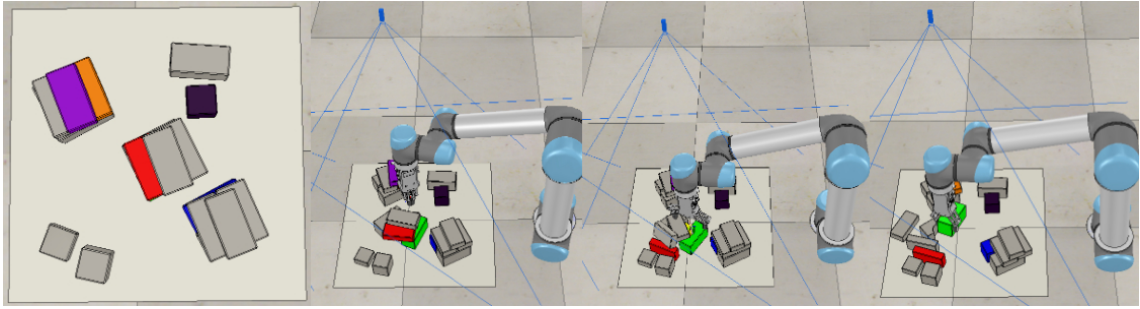


图 1: 本方法在测试场景中抓取绿色目标物体。左边第一张图为机器人视角下看到的工作空间，初始时看不到绿色物体，第二张图表示机器人通过推开其他物体来找到绿色物体，第三张图表示机器人找到了绿色物体，但是机械抓手没有足够的空间去抓取，所以还需要推开其周围物体来获取空间，第四张图代表机器人成功抓取绿色物体。

2 相关工作

2.1 机器人抓取物体

机器人抓取的传统解决方法是通过建立物体和抓手的物理动力学模型来找到稳定的力闭合大小和方向^[4]。然而，这些方法在很大程度上依赖于现实生活中不成立的假设，以及在现实世界中难以获取的物体的先验知识。近年来，由于深度学习在机器人视觉中的广泛应用^[5]，数据驱动的方法取得了很大的进步，与手工制作特征的方法相比，这些方法使机器人能够学习成功的抓取，并具有更强的泛化能力。由于现实世界的场景通常包含多个物体，这使得专注于抓取单个物体的数据驱动方法在现实世界中性能显著下降^[6]。为了解决这个问题，一些研究者将在密集环境中的抓取问题分解为依次抓取单独的物体。文献^[7]提出了只抓取的策略，该策略包括进行机器人抓取检测和操控关系推理，通过按照学习的顺序移除被覆盖的物体，最终抓取目标。然而，纯抓取策略无法抓取实际中杂乱场景中的物体，因为物体相互遮挡导致有效的抓取很少。

2.2 机器人推抓物体

在文献^{[8]、[9]}中已经探索了推和抓之间的协同作用，其目的是通过推的动作来重新调整杂乱的物体，使机器人有足够的空间进行抓取。文献^[8]提出了一个无模型的深度强化学习框架，以学习推和抓的联合策略。文献^[11]和^[10]通过分析指标或神经网络来评估目标对象是否可以抓取，否则执行推的动作来重新排列杂乱的物体。文献^[9]预测了推以及推动之后的状态，以便更直观地判断是否到了抓取的时机。

与上述没有目标物体的方法相比，面向目标的杂乱环境中推和抓取的研究相对更少。文献^[11]通过 Q-learning 学习推动策略，将目标对象从其周围的环境中分离出来；文献^[12]提出了一个双通道 DQN 框架，与^[11]相比，它加快了训练过程。文献^[13]的目标对象在初始状态下是不可见的，并使用贝叶斯策略来搜索它，我们的方法也是面向目标的推抓协同。

3 问题描述

给定目标物体的描述，比如颜色或形状，算法目标是在有限的推和抓动作序列内抓取到目标物体，目标物体可以在密集的环境中以任意姿态和被任意物体遮挡放置。为了解决目标物体以各种姿态和遮挡放置的问题，我们可以将抓取目标物体的问题分解为两个子任务。

第一个子任务是探索任务，如果目标被其他物体完全遮挡，那么机器人就会通过推动作来推开遮挡物，使目标可见。第二个子任务是协同任务，当目标物体可见时，不过目标物体可能周围被其他物体环绕，机械抓手没有足够的空间去抓取，所以如果不通过面向目标的推动作使周围空余空间增大的话，那么单一的面向目标的抓取动作是效率不高的。所以在协同任务中，需要协调目标导向的推和抓取动作顺序，才能以最高的效率抓住目标。

4 本文方法

4.1 本文方法概述

图 2 展示了我们提出的方法框架。首先，在特征提取器中，从工作空间中的固定深度摄像机捕获 RGB-D 图片，特定的目标物体的文本描述通过 CLIPSeg 模型处理用来分割目标物体的掩码。将 RGB 和深度信息重建成三维点云，通过点云重力方向的投影生成彩色高度图，深度高度图和目标掩码高度图。三种类型的高度图由 Efficient v2 网络处理，以输出状态特征 X 。然后，针对上一节提出的两个子任务，使用模型结构相同，但参数不同的图神经网络 Graph Q-Net，用于根据状态特征 X 估计 Q 值。在探索任务中，机器人只使用推动作，而在协同任务中，机器人执行推或者抓取动作，当动作执行后，算法得到奖励。我们将在下面详细讨论这些步骤。

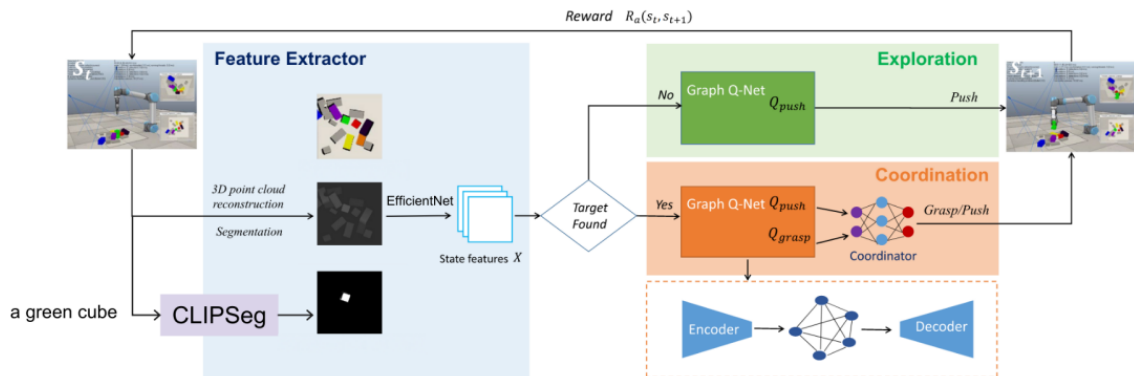


图 2: 方法示意图

4.2 CLIPSeg 模块

CLIPSeg 模型主要作用是分割出目标物体的掩码，输入是工作空间的 RGB 图和关于目标物体的文本描述（颜色或形状），输出是目标物体的掩码^[14]。当存在多种颜色的物体时，CLIPSeg 模型通过文本描述确定目标物体，并且能够拓展到更多训练时未见过的颜色和形状。本方法使用的文本描述是物体的颜色，比如“green”，输入给 CLIPSeg 模型后得到绿色物体的掩码后，机器人根据掩码所在像素点的位置执行动作。CLIPSeg 模型能够分割出在训练时未见过的物体，比如，在测试时，在工作空间中加入粉红色物体，输入文本描述“pink”，CLIPSeg 模型也能得到出粉红色物体的掩码。这是因为 CLIPSeg 模型的使用大图文配对模型 CLIP 作为编码器，它具有非常强的图片和文本配对能力，所以 CLIPSeg 模型甚至不需要对本方法使用的物体进行训练，直接使用其预训练模型就能达到非常好的分

割效果。

算法可以通过语言转成文本，进而输入给 CLIPSeg 模型，用语言作为输入是为了能在以后日常生活中更加通用，提高人机交互能力，而不是向机器人展示一张目标物体的图片，再让机器人去找我们需要的东西。所以语言能够更加方便快捷地表达需求，而且能够更加清晰、无歧义地描述物体，比如当出现两个绿色物体，但其中一个圆柱体，一个是正方体，那我们可以输入文本描述 “a green cube” 来表示目标物体。语言能比以往用图片来表达目标物体更加清晰和方便。

4.3 基于图的 Q-learning

基于图的 Q-learning 使算法能从过去的任务中学习到更多知识^[15]，使其能够适应不同的场景，更好地完成抓取任务。我们使用 DQN 作为基础上增加了图 Q-Net (graph Q-Net) 用于预测 Q 值。图 Q-Net 包含编码器，图推理模块和解码器，图 3 展示了其处理过程，首先降低状态特征 X 的维度，以减少计算复杂度，然后通过编码器对当前状态特征进行编码，形成新的特征 V 。图推理模块对特征 V 进行处理，形成一组新的特征 Z ，解码器将特征 Z 映射到原始状态特征大小，并通过卷积来估计 Q 值。图 Q-Net 网络结构都用于探索和协同任务，下面将详细介绍该网络。

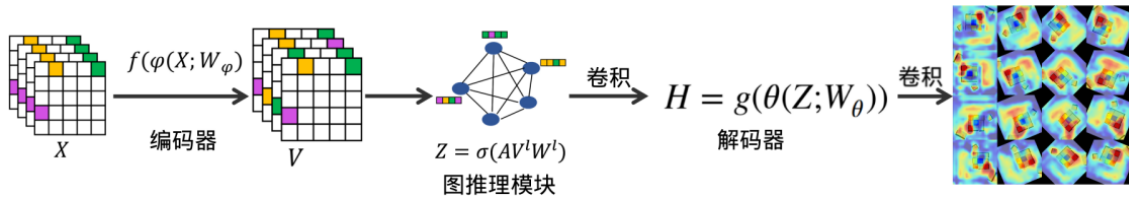


图 3: 图 Q-Nets 数据流

4.3.1 编码器

在本方法中，通过编码器可以集成状态特征 X ，使一个区域的特征包含其他区域的特征，以便更好地进行特征学习。首先网络需要学习一个权重矩阵 W_φ 来线性组成降维后的特征，考虑到状态特征之间的融合与状态特征本身密切相关，我们对特征 X 进行一维卷积来学习 W_φ ，即利用图像特征跨通道融合的结果来整合不同区域的状态特征。这样，全局的信息就会相互作用，编码后的特征 V 就可以包含多个区域的特征，编码过程可以式 1 表示。

$$W_\varphi = \text{Conv 1D}(X), V = f(\varphi(X; W_\varphi)) = X \cdot W_\varphi \quad (1)$$

4.3.2 图推理模块

图推理模块用于集成相邻区域的特征。之前的方法不能适应不同场景的一个可能的原因是它们没有充分利用状态特征之间的信息，导致它们会执行下一时刻的最佳动作，而不考虑完成整个任务的动作序列。在本方法中，我们使用图卷积来获取这些特征之间的内部关系。具体而言，状态特征 X 经过编码器后得到特征 V ，我们用图结构的形式来表示它，在这个构造的图中，我们将特征 V 的维度 N 作为节点数，每个维度都是一个节点的特征。因此，构造的图结构可以表示为 $G = (A, V)$ ，其中 A 是 $N \times N$ 个节点的邻接矩阵，用于在节点之间传播信息，并在训练过程中通过随机初始化参数和梯度下降进行学习，我们通过式 2 更新特征。

$$Z = \sigma(AV^l W^l) \quad (2)$$

其中 W^l 是需要学习的权重矩阵, l 是图卷积网络的层数。这样一个区域的特征不仅包含本身的节点特征, 而且融合了相邻区域的特征。

4.3.3 解码器

为了使处理后的特征对应于特定的像素, 我们使用卷积层进行特征分析, 分析后的特征 H 可以更好地用于以后的网络层估计 Q 值。所以我们的目标是学习一个函数 g , 并用式 3 解码这些特征,

$$H = g(\theta(Z; W_\theta)) \quad (3)$$

其中 W_θ 是学习的权重矩阵, H 是解码后的状态特征。最后, 我们通过两个卷积层预测解码特征 H 的 Q 值。

4.4 动作策略

以下是算法抓取不可见目标物体的策略执行的细节, 重复执行直到机器人抓取目标物体。对于每次迭代, 算法决定根据是否找到目标物体, 即目标掩码 M 是否存在, 来执行探索或协同任务。在探索任务中, 当寻找目标物体时, 机器人只使用推动作, 因此, 我们用一个图 Q-Net 进行探索, 机器人通过这个图 Q-Net 在最佳 Q 值的位置执行这个动作。在协同任务中, 当目标物体可见时, 机器人需要推和抓取两个动作的协同。因此, 我们使用两个图 Q-Nets 来分别构建协同任务中的推和抓取的 Q-Nets, 可以用式 4 表示,

$$y = f_a(q_p, q_g, r_b, n_b, c_g) \quad y \in \{m_{\text{grasp}}, m_{\text{push}}\} \quad (4)$$

其中 f_a 是一个有三个全连接层组成的函数预测器, q_p 和 q_g 分别表示这两个图 Q-nets 的估计 Q 值, r_b 和 n_b 代表目标周围的凌乱程度, r_b 是目标目标周围边界的占用率, n_b 是目标边界占用的模, c_g 是连续抓取失败的次数, $m_{\text{grasp}}, m_{\text{push}}$ 分别表示抓取动作和推动作。

4.5 损失函数定义

机器人抓取目标物体的任务可以用马尔科夫决策过程。在 t 时刻, 机器人根据当前状态 s_t 做出决策并执行动作 a_t , 然后生成一个新的状态 s_{t+1} 。算法的目标是找到一个最优策略来最大化回报价值, 所以损失函数可用 5 表示。

$$L_{\text{exp}} = |Q_{\text{exp}}(s_t, a_t) - (R_a(s_t, s_{t+1}) + \gamma \max Q_{\text{exp}}(s_{t+1}, a'_t))| \quad (5)$$

$$L_{\text{coo}} = |Q_{\text{coo}}(s_t, a_t) - (R_a(s_t, s_{t+1}) + \gamma \max Q_{\text{coo}}(s_{t+1}, a'_t))| \quad (6)$$

其中 L_{exp} 是探索任务的损失函数, L_{coo} 是协同任务的损失函数, $Q(s_t, a_t)$ 是当前状态 s_t 下估计的最大 Q 值, $R_a(s_t, s_{t+1})$ 是行动后的奖励, $Q(s_{t+1}, a'_t)$ 是下一次迭代中所有可能动作对应的预测 Q 值, γ 是折扣因子。

5 复现细节

5.1 与已有开源代码对比

与已有开源代码对比, 本方法主要有 4 点改进。第一, 测试场景的物体的改变, 本实验存在多种不同颜色的物体, 而已有开源代码的测试场景中只有灰色和目标物体的颜色的其中一种, 它的算法设置是抓取除了灰色之外的另一种颜色的物体, 但这种实验环境无法应用于实际生活中, 因为现实存在

多种颜色的物体，所以我们需要额外的输入来确定我们的目标物体。

第二，算法使用文本和图像作为输入，并用 CLIPSeg 模型分割出目标物体的掩码。文本作为输入，能够清晰地、直观地和方便地表达出我们的需求，从多种颜色物体中确定我们想要的颜色的物体，提高人机交互能力，而之前的代码只有图像作为输入，所以无法确定多种颜色物体场景下的目标物体，只会随机抓取其中一个有颜色的物体；CLIPSeg 模型不需要微调模型，也就是不用人工收集训练分割模型的数据，而已有开源代码需要收集实验环境下的物体的图片作为分割模型的训练数据，并且需要人工进行标注，才能训练出只能识别特定类别的分割模型，这不仅耗费时间去人工标注数据，而且无法拓展到新颜色的物体，比如当出现粉红色物体时，由于分割模型没有粉红色物体的训练数据，导致无法识别这个物体，所以其分割模型拓展性很差，相反，CLIPSeg 模型使用大量的图文配对的数据进行训练，我们直接用其预训练模型即可达到较好的分割效果，而且能够拓展到新颜色的物体。

第三，使用 EfficientNet V2 网络作为提取特征的主干网络。当得到彩色高度图，深度高度图和目标掩码高度图后，如何快速得到有效的特征呢？我们使用的是 Efficient v2 网络，因为 EfficientNetV2 训练速度更快、精度更高和参数量更少，所以比之前用 Densenet 网络能够更快速学到有效的特征。第四，本方法使用图神经网络学习 Q 值，而开源代码使用 DQN 网络预测 Q 值。图结构化表示和迭代信息可以用来发现和预测场景之间的关系，而开源代码会选择下一个最有可能的抓取姿势，而不是在整个抓取任务中预测整个动作序列，导致执行很多重复的动作，而且它不能从训练过程中学到额外的知识来适应不用的场景，而我们使用编码器、图推理模块和解码器来学习抓取和推动作策略，所以本方法能够在了解周围环境并识别目标物体后执行一个动作序列来完成任务。

以上就是我在已有开源代码上改进的四点内容，将其整合到系统中，能够实现在多种颜色物体的场景下抓取目标物体，而且实验结果比原有代码效果好。

5.2 实验环境搭建

在仿真环境 V-REP 中，采用带有 RG2 抓手的 UR5 机器臂对模型进行训练，用两个摄像头获取图像信息，其中正上方的摄像头获取 RGB 图像，斜上方的摄像头的图像信息仅用于构建高度图，而不知道物体的颜色。特征提取器 EfficientNetV2 用了在 ImageNet 数据集的预训练模型，图 Q-Nets 是从头开始训练，而且使用不同的训练数据集对两个用于探索和协同任务的图 Q-Nets 进行训练，也就是训练前 1000 次迭代训练探索任务，然后开始训练协同任务。训练探索任务的数据集有 10 个场景，训练协同任务有 8 个场景，测试场景有 8 个，但是因为目标物体可以是不同颜色，所以算法可以产生足够多的抓取成功轨迹来训练网络，图 4 分别展示的是训练探索任务、协同任务和测试的数据。

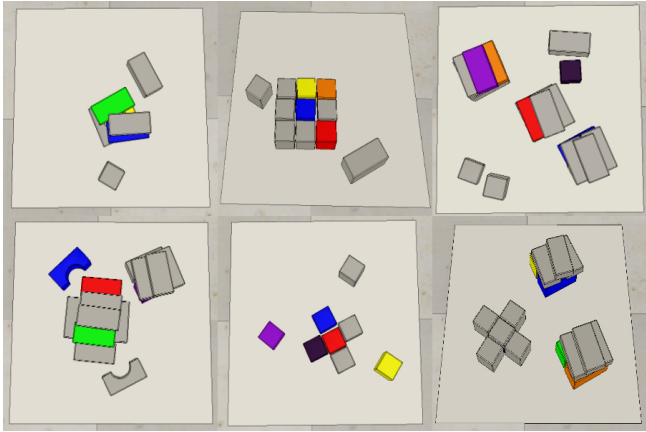


图 4: 数据集的场景

表 1: 两种方法在测试场景上的表现性能

6 实验结果分析

我们将之前的开源代码称为 Grasping-Invisible(GI), 评价指标使用抓取成功率 Success Rate(SR) 和完成任务的总动作次数 Motion number(MN)。SR 表示只关心抓取动作的次数, 即是否抓取到目标物体而且抓取是否牢固。在测试场景中 (1、3、4、8) 上, 对于 SR 指标, 我们的方法平均比 GI 高 20.3%, 对于 MN 指标, 我们的方法平均对 GI 少 2.7 步。在 4 个测试场景结果表明本方法比原方法效果要好, 因为本方法通过 CLIPSeg 模型分割出目标物体的掩码, 而原有代码设置上虽明确了目标物体的颜色, 但是它不能分割出目标物体, 而是分割出所有带颜色的物体的掩码, 所以它会一直抓取有颜色的物体, 直到抓到目标物体, 这样导致抓取成功率下降, 动作次数增多。结果表明, 我们的方法能够以较少的动作成功地抓取目标物体。

Test	Performance	Method	
Case	Metrics	GI	Our
a	SR(%)	30.0	45.0
	MN	12.6	8.6
b	SR(%)	18.0	71.0
	MN	13.0	6.5
c	SR(%)	30.0	31.0
	MN	9.0	9.6
d	SR(%)	21.0	33.0
	MN	10.4	9.6
Average	SR(%)	24.7	45.0
	MN	11.2	8.5

7 总结与展望

本报告提出了一种基于图的深度强化学习方法来抓取不可见的目标物体, 其中用 CLIPSeg 模块处理关于目标物体的描述来分割出其掩码, 而且构建了图 Q-Nets 来学习状态特征并预测 Q 值, 这解决了以往学习特征时缺乏全局上下文信息的问题。实验结果表明, 本方法的抓取成功率是 45.0%, 优于之前的方法。在以后的研究中, 考虑加入更多物体, 比如水果, 瓶子等, 因为机器人在这些场景中需要考虑更多, 比如推动物体过程中会不会导致杯子破碎。所以将算法泛化到各种形状和复杂场景的物体是我们之后工作的重点。

参考文献

- [1] DENG Y, GUO X, WEI Y, et al. Deep reinforcement learning for robotic pushing and picking in cluttered environment[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2019: 619-626.
- [2] DANIELCZUK M, KURENKOV A, BALAKRISHNA A, et al. Mechanical search: Multi-step retrieval

of a target object occluded by clutter[C]//2019 International Conference on Robotics and Automation (ICRA). 2019: 1614-1621.

- [3] DANIELCZUK M, ANGELOVA A, VANHOUCKE V, et al. X-ray: Mechanical search for an occluded object by minimizing support of learned occupancy distributions[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2020: 9577-9584.
- [4] GUPTA M, SUKHATME G S. Using manipulation primitives for brick sorting in clutter[C]//2012 IEEE International Conference on Robotics and Automation. 2012: 3883-3889.
- [5] LENZ I, LEE H, SAXENA A. Deep learning for detecting robotic grasps[J]. The International Journal of Robotics Research, 2015, 34(4-5): 705-724.
- [6] MOUSAVIAN A, EPPNER C, FOX D. 6-dof graspnet: Variational grasp generation for object manipulation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 2901-2910.
- [7] ZHANG H, LAN X, BAI S, et al. A multi-task convolutional neural network for autonomous robotic grasping in object stacking scenes[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2019: 6435-6442.
- [8] ZENG A, SONG S, WELKER S, et al. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2018: 4238-4245.
- [9] HUANG B, HAN S D, BOULARIAS A, et al. Dipn: Deep interaction prediction network with application to clutter removal[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). 2021: 4694-4701.
- [10] CHEN Y, JU Z, YANG C. Combining reinforcement learning and rule-based method to manipulate objects in clutter[C]//2020 International Joint Conference on Neural Networks (IJCNN). 2020: 1-6.
- [11] KIATOS M, MALASSIOTIS S. Robust object grasping in clutter via singulation[C]//2019 International Conference on Robotics and Automation (ICRA). 2019: 1596-1600.
- [12] SARANTOPOULOS I, KIATOS M, DOULGERI Z, et al. Split deep q-learning for robust object singulation[C]//2020 IEEE International Conference on Robotics and Automation (ICRA). 2020: 6225-6231.
- [13] YANG Y, LIANG H, CHOI C. A deep learning approach to grasping the invisible[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 2232-2239.
- [14] LÜDDECKE T, ECKER A. Image segmentation using text and image prompts[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 7086-7096.
- [15] ZUO G, TONG J, WANG Z, et al. A Graph-Based Deep Reinforcement Learning Approach to Grasping Fully Occluded Objects[J]. Cognitive Computation, 2022: 1-14.