

基于深度强化学习的混杂场景目标物体 推抓协同策略

胡楷雄¹ 宋远航¹ 周勇^{1*} 李卫东²

1. 武汉理工大学交通与物流工程学院, 武汉, 430063

2. 上海理工大学机械工程学院, 上海, 200093

摘要:为提高机器人在混杂场景中抓取被遮挡目标物体的成功率和效率,提出一种基于深度强化学习的“推动”和“抓取”协同推抓策略。该策略利用两个深度 Q 网络,以 RGB-D 图像为输入来确定推动或抓取动作,并通过推动改变物体排列以优化抓取条件。该网络使用“抓-推-抓”三阶段模型训练方法显著提高了抓取能力。基于图像形态处理的方法识别并过滤低质量抓取动作,从而提高成功率和效率。实验结果表明,该方法有效提高了目标物体的抓取成功率和效率。

关键词: 机器人抓取; 混杂场景; 深度强化学习; 多动作协同

中图分类号: TG54; TP242.2

DOI:10.3969/j.issn.1004-132X.2025.01.014

开放科学(资源服务)标识码(OSID):



A Cooperative Strategy for Pushing and Grasping Target Object in Cluttered Scenes Based on Deep Reinforcement Learning

HU Kaixiong¹ SONG Yuanhang¹ ZHOU Yong^{1*} LI Weidong²

1. School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan, 430063

2. School of Mechanical Engineering, Shanghai University of Technology, Shanghai, 200093

Abstract: To improve the success rate and efficiency of robotic grasping for occluded target objects in cluttered scenes, a collaborative push-grasp strategy was proposed based on deep reinforcement learning. The strategy employed 2 deep Q networks and used RGB-D images as inputs to determine push or grasp actions, which optimized object arrangement for better grasping conditions. A “grab-push-grab” three-stage training method was introduced in the model to enhance grasping capabilities significantly. An image morphology-based assessment method effectively identified and filtered low-quality grasp actions to increase successful rates and efficiency. Experimental results confirm that this method significantly enhances the successful rate and efficiency of grasping target objects.

Key words: robot grasping; cluttered scene; deep reinforcement learning; multi-action collaboration

0 引言

机器人抓取物体的灵活性和熟练度是实现复杂操作任务的关键^[1]。通常的抓取方法可分为基于分析的方法和数据驱动的方法^[2-3]。基于分析的方法通过对目标物体形状的精确解析来规划合理的抓取姿态。这种方法需要已知物体的 3D 模型,且在寻找稳定的力闭合点时,往往依赖手工设计的特征,这限制了其应用^[4-6]。相比之下,数据驱动的方法允许机器人利用学习到的场景图像和深度信息执行抓取动作,无需明确的对象知识。基于深度强化学习的抓取策略能通过自我监督构建训练数据,具有更高的灵活性,因此在抓取领域得到广泛应用^[7-11]。

在混杂抓取场景中,物体可能相互堆叠或紧密排列,这增加了抓取操作的复杂性。在这种情况下,基于深度强化学习的抓取方法面临的如下两个核心问题成为机器人技术在复杂工业场景中应用的主要障碍:

1) 抓取成功率不高。仅依赖单一抓取动作往往难以有效抓取物体^[12]。为解决这一问题,一些研究引入改变物体排列的操作,如推动或吸取物体。ZENG 等^[13]提出了一种深度 Q 学习框架,通过学习推抓策略,实现推动与抓取动作的协调。DENG 等^[14]设计的复合机械手先通过吸盘从杂乱环境中吸起物体,再用夹爪抓住物体。TANG 等^[15]通过自我监督训练使机器人学习平面推动与六自由度抓取策略。上述研究扩展了抓取动作空间,并使机器人能处理各种杂乱物体场景,但主

要面向所有物体的抓取,无法针对特定对象,且多动作协调机制存在不足,抓取时机选择不够精准。

2) 混杂场景中的动作效率低下。动作效率主要体现在成功抓取指定物体所需的动作次数,如 YANG 等^[16]将抓取被遮挡目标物体的问题分为两个子任务,并结合贝叶斯策略和相关领域知识,利用二元分类神经网络输出抓取或推动动作的方法整体倾向于输出抓取动作,导致抓取动作过多。XU 等^[17]在面向目标的抓取任务中,采用全卷积神经网络 (fully convolutional networks, FCN) 计算各像素的 Q 值,并通过判断最大 Q 值是否达到阈值来决定动作类型。这种方法受 Q 值不准确的影响,容易导致抓取失败,增加动作次数。

上述两个问题影响了机器人在混杂场景中高效抓取目标物体的能力,亟需进一步的研究与改进。为解决混杂工作场景中目标物体的抓取问题,本文提出一种基于深度强化学习的“推动”和

“抓取”协同推抓策略。将深度相机获取的 RGB-D 图像作为输入,输出相应的推动或抓取动作。提出一种“抓-推-抓”三阶段训练方法,使网络逐渐学习到混杂场景有效抓取的策略,从而提高抓取能力。基于图像形态处理的方法可有效识别和过滤质量较差的抓取动作,提高成功率与动作效率。相较于其他方法,本文考虑了 Q 值在评估抓取动作优劣时的局限性,不再单纯依赖 Q 值确定动作类型,而是综合考虑抓取位置选择、物体干扰等多个状态信息来优化动作协同策略。

1 算法原理

1.1 策略整体架构

将面向目标的推送抓取问题建模为一个马尔可夫决策过程 (MDP)。MDP 的状态空间为 s , 动作空间为 a , 奖励为 r 。图 1 所示为协同策略整体架构。

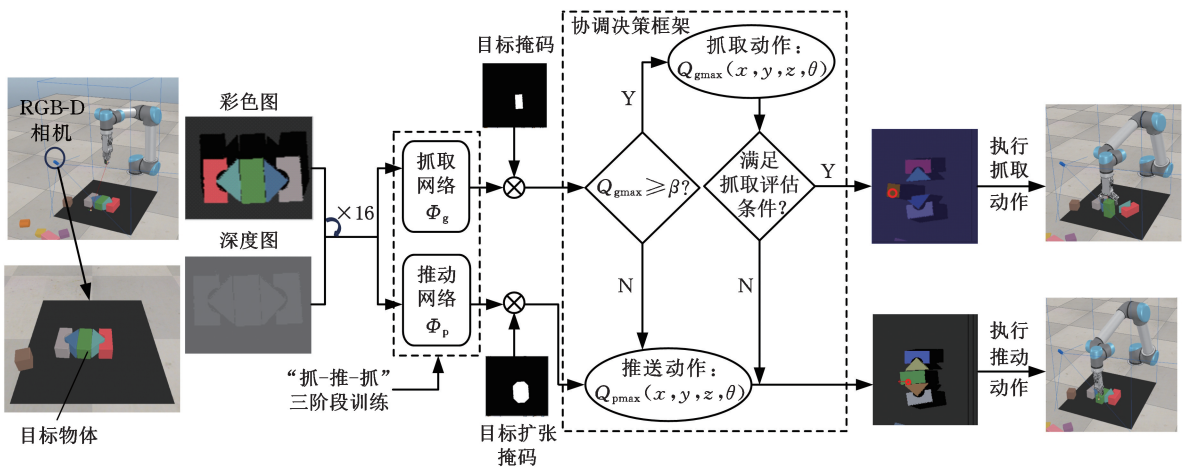


图 1 协同策略整体架构

Fig.1 Overall architecture of collaboration strategy

工作场景中,通过固定在斜上方的一个相机捕捉 RGB-D 图像,然后利用原始的 RGB-D 图像构建正交投影来表示时刻 t 的状态 s_t 。动作空间定义为元组 $a = (\psi, x_t, y_t, z_t, \theta_t)$, 其中, $\psi \in \{\text{grasp}, \text{push}\}$ 表示机器人的抓取动作 (grasp) 和推动动作 (push); (x_t, y_t, z_t) 为夹持器的空间坐标; θ_t 为夹持器绕 Z 轴旋转的角度。夹持器每次旋转的角度相同,均为 22.5° 。动作定义中,“推动”动作包括两个要素即起始位置和推动方向,而推动的长度是固定的。“抓取”动作包括两个要素即抓手平行爪的中间位置和方向,而夹爪张开的宽度是固定的。模型采用 2 个相同的神经网络架构分支来预测“推动”动作和“抓取”动作。将获得的彩色高度图和深度图分别旋转 16 个角度(每个角度为 22.5°)后,输入 2 个全卷积网络。这 2 个

全卷积网络的输出是与原始输入图像具有相同尺寸和分辨率的 Q 值密集像素映射,每个图像像素的 Q 值表示在对应 3D 位置执行动作的预期回报。随后,通过颜色识别获取目标物体的掩码。将该掩码与抓取网络和推送网络输出得到的 32 个 Q 值像素图进行 Hadamard 乘积处理,以筛选出有效的 Q 值。掩码与抓取网络相乘可得目标掩码(大小可根据需要灵活设计),掩码与推送网络相乘可得目标扩张掩码。上述两个操作有助于过滤掉非目标物体,消除不相关的像素 Q 值,规划机器人不同动作的执行范围。

系统由推送网络和抓取网络构成,输出包括推送动作和抓取动作。给定状态 s_t 时,机器人决定选择动作类型,以及在哪个位置和方向执行动作是至关重要的。本文从图像形态处理的角度入

手,构建包含评估判断方法的协调决策框架,输出执行动作。协调决策 a_t 的数学公式为

$$a_t = \begin{cases} \text{push} & Q_{\text{gmax}} < \beta \\ \text{grasp} & Q_{\text{gmax}} \geq \beta \text{ 且评估条件满足} \end{cases} \quad (1)$$

其中, β 为设定的阈值。目标物体最大抓取 Q 值 $Q_{\text{gmax}} < \beta$ 说明抓取动作质量不佳,此时抓取目标物体成功的概率较低,因此,协调策略直接选择推送动作。目标物体 $Q_{\text{gmax}} \geq \beta$ 表示目标物体有较大的概率抓取成功,但目标物体 Q_{gmax} 的预测并不完全准确,需要对抓取动作进行评估判断,评估判断方法见 1.3 节。如果评估条件满足,则在目标物体处于最大 Q 值 Q_{gmax} 时执行抓取动作,否则执行推送动作,以筛除质量较差的抓取动作。

1.2 “抓-推-抓”三阶段模型训练

模型的训练有 3 个阶段:① 抓取网络的初步训练,以训练基本的抓取能力;② 推送网络的训练,以扩展操作策略;③ 抓取网络的进阶训练。上述三个阶段的训练使模型从简单样本逐步深入到复杂样本。

1) 抓取网络的初步训练。此阶段是在稀疏场景下初步训练抓取网络,抓取目标物体相对容易。模型学习处理简单情境中的抓取任务,建立基本的抓取能力。在这个阶段只放置 5 个对象来随机生成稀疏场景,机器人只执行抓取动作,每次抓取对应一个回合。系统将目标物体掩码中 Q 值最大的像素点作为抓取位置。定义抓取奖励函数为

$$R_{g1} = \begin{cases} 0 & \text{抓取失败} \\ 1 & \text{抓取成功} \end{cases} \quad (2)$$

2) 推送网络训练。抓取网络的参数在此阶段不变,以使模型专注于学习推送动作。这个阶段的目标是提高机器人在场景中调整和移动物体的能力。系统推送目标物体扩张掩码中 Q 值最大的像素点位置。推送动作提高了抓取网络预测的 Q 值,将会获得奖励。定义推送奖励函数为

$$R_p = \begin{cases} 0.5 & \Delta Q_g > 0.1 \text{ 和 } \Delta R_b > 0.01 \\ 0 & \text{其他} \end{cases} \quad (3)$$

式中: ΔQ_g 为推送动作前后的目标物体最大抓取 Q 值之差; ΔR_b 为推送动作前后的目标物体周围空间被占用率的差值。

目标物体周围空间的被占用率通过以下方法获得:首先构建目标物体掩码扩张前(图 2a)和扩张后(图 2b)差值的外围轮廓掩码(图 2c 的红色区域);然后利用外围轮廓掩码中高于工作面的像素比例来衡量目标物体周围空间的拥挤程度。

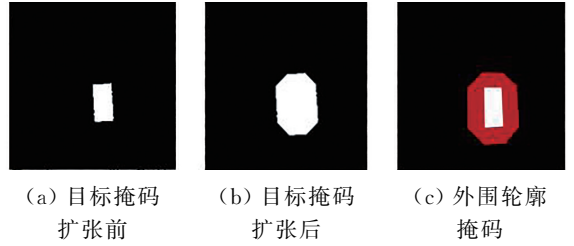


图 2 掩码示意图
Fig.2 Mask diagram

为提高推送网络训练效率,将每一回合的推送动作数限制在 5 以内,并在最后附加一个抓取动作。如果目标掩码内的最大 Q 值大于设定的阈值,则机器人立即执行抓取动作。

3) 抓取网络的进阶训练。此阶段是在复杂的场景中继续训练抓取网络,与第一阶段相比,场景中的物体从 10 个逐步增加到 30 个,抓取难度逐渐加大。利用第一阶段抓取网络的初步训练和第二阶段推送网络训练的成果,提高机器人在密集环境中抓取物体的能力。通过逐渐提高训练难度,更好适应真实工作环境中的复杂抓取任务。抓取动作的奖励函数设置为

$$R_{g2} = \begin{cases} -1 & \text{抓取失败} \\ 1.5 & \text{抓取成功} \end{cases} \quad (4)$$

在本阶段,固定推送网络参数,每一回合的推送数量仍限制在 5 以内。目标掩码内的最大 Q 值超过阈值时执行抓取动作。整体训练过程如图 3 所示,场景中的绿色物体为目标物体。

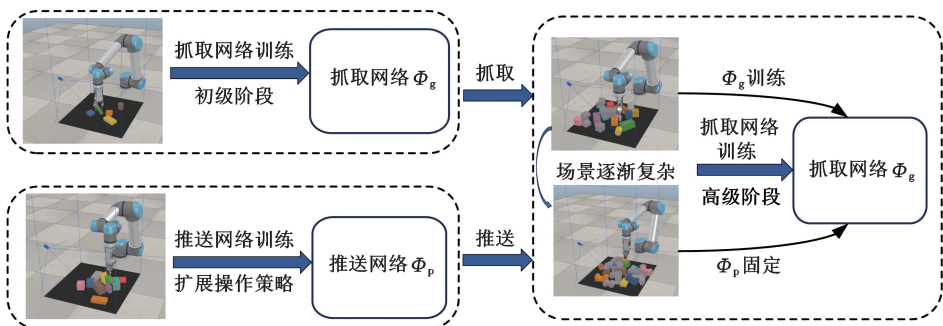


图 3 网络分阶段训练
Fig.3 Network training in stages

4) 训练细节。本文使用 2 个 121 层的 DenseNets^[18] 在 ImageNet^[19] 上进行预训练,提取视觉特征。这些特征表示馈入全卷积网络(包含 2 个 1×1 的卷积核和整流线性单元(ReLU)激活函数^[20])并进行批归一化处理^[21]。最后,对全卷积网络进行双线性上采样来生成逐像素的预测。协同模型使用 Adam 优化器进行网络训练^[22],学习率固定值为 10^{-4} ,权值衰减为 2^{-4} 。服务器的硬件包括 Ryzen Threadripper 3960X CPU 和 GeForce RTX 3080Ti 显卡。

1.3 评估判断方法

夹爪的简化三维模型(尺寸已知)如图 4a 所示。执行具有最大 Q 值的抓取动作之前,已预先确定抓取点的精确坐标及抓取动作的具体方向,因此可以确定在执行该抓取动作时夹爪在工作场景中的正向投影,如图 4b 所示。

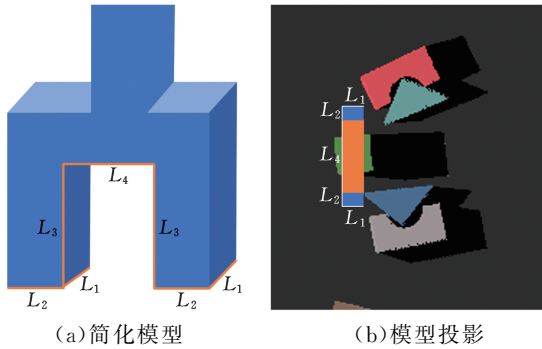


图 4 夹爪的简化示意图

Fig.4 Simplified schematic diagram of gripper

通过计算夹爪两端投影范围(图 4b 中两个 $L_1 \times L_2$ 的蓝色矩形区域)内的深度来评估夹爪在执行抓取操作时是否具有足够的空间进行插入,进而判断是否存在与物体碰撞的风险,避免抓取失败。此外,计算夹爪完全展开后所覆盖区域(图 4b 中的橙色区域)的高度,并减去目标物体在投影区域的深度,有助于判断夹爪闭合时是否仅捕获目标物体,还是可能同时夹取多个物体。如果 $Q_{\text{gmax}} \geq \beta$ 且夹爪两端有足够的插入空间(见图 5),则执行抓取动作;反之,则存在其他物体的潜在干扰,需要灵活调整策略,将执行抓取动作改为执行推动动作,防止夹爪在展开下落过程中与物体发生碰撞,避免抓取失败。推送动作能有效分散目标抓取物体,为后续的抓取操作创造更理想的环境和条件。通过评估判断方法,系统可在执行抓取动作前进行评估判断,及时规避不合适的抓取动作,从而提高抓取动作的准确性和成功率。

仿真实验发现,抓取网络对不同方向的抓取动作的 Q 值不敏感,这导致 Q_{gmax} 对应的抓取方

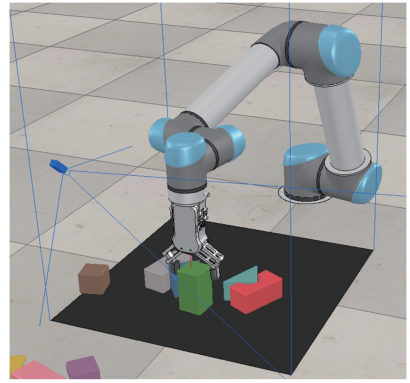


图 5 执行抓取动作

Fig.5 Performing the grasp action

向可能并非最优,较小 Q 值对应的抓取方向更可能成功。如果仅依据当前选择的抓取动作进行判断,可能会错过其他成功的抓取机会。因此,当最初选择的具有最大抓取 Q 值的抓取动作不满足判断条件时,在剩余的不同抓取方向中选择 Q 值最大的抓取动作进行评估判断,机器人根据动作是否满足判断条件来决定执行抓取动作还是推送动作,如图 6 所示。该处理方式扩大了有效的评估样本,避免了过早舍弃可能成功的抓取动作。然而,在选择剩余不同方向的抓取动作样本时,由于抓取动作的 Q 值较小,满足判断条件的可能性较低,这会增加计算时间,降低系统的实时性和效率。此外,较大偏差对应的抓取方向容易导致抓取失败,不利于提高抓取成功率和动作效率。

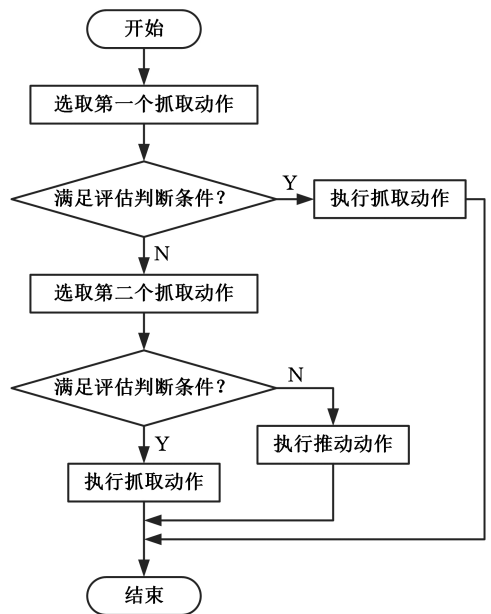


图 6 动作协同流程

Fig.6 Action collaboration process

2 实验分析

为评估本文所提方法,在仿真环境和现实环境中进行了一系列的实验,以验证本文方法的优

越性。

2.1 仿真实验及结果分析

为验证模型效果,将本文所提策略与 Grasping the Invisible(GI)^[16]、Efficient Push-grasping (EPG)^[17] 进行比较。比较过程中,仿真实验环境和评估指标与对比方法均保持一致,以确保公正性和准确性。比较采用以下 3 个实验评估指标。

1) 完成率即所有有效测试运行中成功完成任务的百分比均值。如果机器人在一轮测试中能成功抓起目标物体且没有连续 10 次抓取失败,则该任务被视为完成。完成率 R_C 的计算公式为

$$R_C = \frac{1}{T_{\text{valid}}} \sum_{i=1}^{T_{\text{valid}}} n_i \quad (5)$$

$$T_{\text{valid}} = T - T_{\text{nonvalid}} \quad (6)$$

$$n_i = \begin{cases} 1 & \text{抓取完成} \\ 0 & \text{抓取未完成} \end{cases} \quad (7)$$

式中: n_i 表示在第 i 轮测试中是否抓取完成; T 为总的测试轮数; T_{valid} 为有效的测试轮数,即抓取动作次数大于 0 的测试次数; T_{nonvalid} 为无效的测试轮数,即抓取动作次数为 0 的测试次数。

2) 抓取成功率。所有有效测试运行中目标物体的抓取成功率为

$$R_{GS} = \frac{1}{T_{\text{valid}}} \sum_{i=1}^{T_{\text{valid}}} m_i \quad (8)$$

$$m_i = \begin{cases} 1/g_i & n_i = 1 \\ 0 & n_i = 0 \end{cases} \quad (9)$$

式中: m_i 、 g_i 分别为第 i 轮测试中的抓取成功率和抓取次数。

3) 动作数量,即有效测试运行中机器人执行推动和抓取操作的平均次数,它反映机器人动作的效率,以获得较小值为目标。动作数量 N_M 的计算公式为

$$N_M = \frac{1}{T_{\text{valid}}} \sum_{i=1}^{T_{\text{valid}}} k_i \quad (10)$$

$$k_i = g_i + p_i \quad (11)$$

式中: k_i 为第 i 轮测试中的动作次数,包括推动次数与抓取次数; p_i 为第 i 轮测试中的推动次数。

为评估抓取网络的抓取效果,分别在场景中随机放置 3 个、5 个、8 个物体,测试 30 回合,统计仅通过一个抓取动作就能成功抓取目标物体的成功率,如表 1 所示,本文方法在一个抓取动作的成功率最大,超过了 GI 和 EPG。

表 1 抓取网络的抓取成功率

物体数量	GI	EPG	本文方法
3	83.3	83.3	93.3
5	80.0	76.7	83.3
8	70.0	70.0	73.3

本文方法在训练完抓取网络和推送网络后,采用固定推送网络参数的方法对抓取网络进行进阶训练。进阶训练阶段的场景设置更复杂,物体数量从 10 逐步递增至 30,抓取难度逐步提高。借助抓取网络第一阶段的训练成果和推动动作的协同作用,机器人已具备在更密集复杂的环境中进行抓取训练的能力。相对于 EPG 中固定 10 个物体的简单训练场景,本文方法通过渐进提高训练难度,使机器人能更好地适应复杂多变的工作环境,从而显著提升其抓取网络的性能。

为全面评估所提方法,在 2 个不同测试场景中进行实验。第一个测试场景是结构化场景,其中所有物体的排列方式均提前设定好,如图 7 所示,目标物体(绿色物体)被其他物体紧紧包围,直接抓取目标不可行,因此,机器人必须通过推动动作将目标物体从场景中分离出来才能进行抓取。这对机器人的推抓协同配合能力提出了较高要求。每个测试场景都进行 30 次测试,然后求取 10 个案例的平均值,实验结果见表 2。

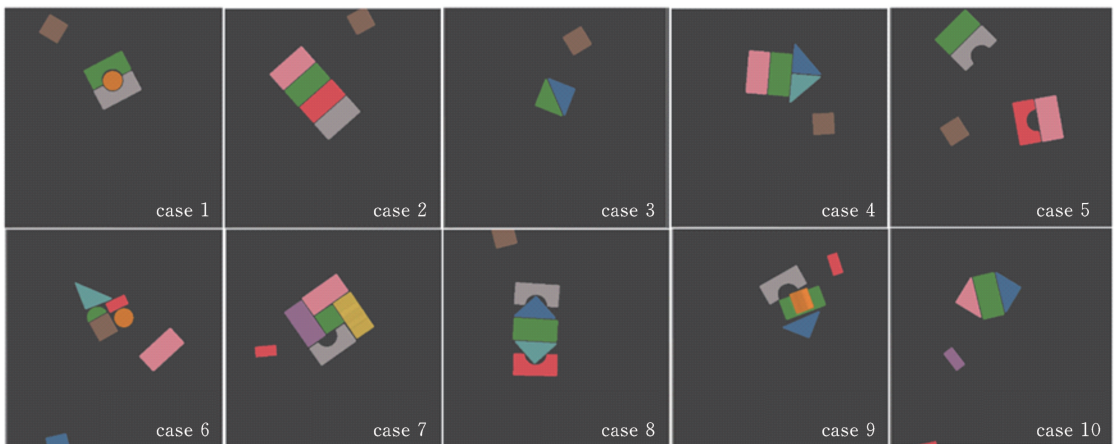


图 7 结构化测试场景

Fig.7 Structured test scenario

表 2 结构化场景的测试结果

Tab.2 Test results of structured scenario

方法	完成率/%	抓取成功率/%	动作数量
GI	95.0	70.4	4.33
EPG	99.0	90.0	2.77
本文方法	100	95.0	2.55

10 个结构化测试场景中,本文方法的完成率达到 100%。GI 与 EPG 的完成率稍低的主要原因在于,部分测试案例中,GI 与 EPG 在尝试抓取目标物体失败后会将抓取目标推出工作场景,导致抓取任务未能完成,从而降低整体的完成率。

尽管本文方法也存在将目标物体推出工作场景的问题,但在此之前规避了不合适的抓取动作,相应的抓取动作次数为 0。根据评估指标的定义,该轮测试视为无效测试,不纳入完成率的计算。因此,本文方法的完成率达到 100%。此外,本文方法的抓取成功率为 95%,在 3 种方法中最高,同时动作数最小,进一步表明本文方法的优越性。在执行推动或抓取动作时,GI 的机器学习模型输入的相关参数与动作类型之间的关系不够明确,因此抓取成功率最低。目标物体抓取问题涉及目标物体的相对位置、周围环境的拥挤程度、物体的分布状态等因素,不能简单依赖某些参数的数值来确定动作类型。GI 在目标物体很难直接抓取的情况下仍然倾向于选择抓取动作。EPG 抓取成功率较低的原因在于,尽管一些抓取动作的 Q 值达到阈值,但夹爪落点受物体干扰,导致抓取失败。这表明抓取动作的 Q 值在一定范围内波动并不能准确反映抓取动作的优劣。对于具有不同 Q 值的抓取动作,不能简单认为 Q 值较大的动作必然比 Q 值较小的动作更利于成功抓取目标。直接依据 Q 值大小选择并执行抓取动作可能导致抓取任务的失败。因此,在选择执行抓取动作时,除了考虑 Q 值大小,还需综合考虑抓取位置的选择、环境干扰等因素。

第二种测试场景是随机场景,如图 8 所示,依次在场景中随机放置 30 个位置和角度都随机生成的物体,其中,绿色物体为目标物体。30 次测试结果的均值见表 3。

随机测试场景中,本文方法在各项评价指标

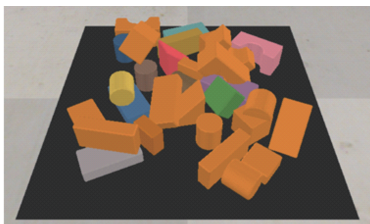


图 8 随机化测试场景

Fig.8 Randomization test scenario

表 3 随机化场景测试结果

Tab.3 Test results of randomized scenarios

方法	完成率/%	抓取成功率/%	动作数量
GI	96.7	54.6	5.37
EPG	97.8	90.0	4.82
本文方法	100	95.3	2.48

上仍然是最优的。GI 的抓取成功率和完成率最低,表明其系统对高度混乱的场景适应能力较差。本文方法通过评估判断筛选与排除质量较差的抓取动作,及时选择推送动作,减少了抓取失败带来的冗余动作。EPG 的动作数量远远高于本文方法,这是因为随机测试场景中的物体过于密集,抓取难度较大,需要较多推动使物体的 Q_{gmax} 达到抓取阈值。本文方法对抓取动作进行了筛选评估,抓取阈值较 EPG 的设置值小。这样既可以剔除质量较差的抓取动作,又不会错过抓取 Q 值较小但适合抓取的动作,能显著提高动作效率。

最后通过消融实验探讨评估判断方法能否提高抓取成功率和动作效率。如表 4 所示,与无评估判断方法相比,本文方法在所有指标上都更好,完成率达到 100%,抓取成功率提高到 95.0%,动作数量减少到 2.55,这说明引入评估判断的方法有助于提高系统的抓取成功率和动作效率。

表 4 消融实验测试结果

Tab.4 Results of ablation experiments

方法	完成率/%	抓取成功率/%	动作数量
无评估判断	95.1	87.3	3.49
本文方法	100	95.0	2.55

2.2 真实环境实验及结果分析

为测试模型在现实环境中的工作效果,在真实的环境中搭建了抓取实验平台,如图 9 所示。该抓取实验平台使用 1 个 UR5 机器人和 1 个大寰 AG95 二指夹爪,并通过斜上方的 Kinect V2 相机获取 RGB-D 图像。

4 种测试案例如图 10 所示,目标物体通过黄色星号标记。每个测试案例进行了 15 次测试,并

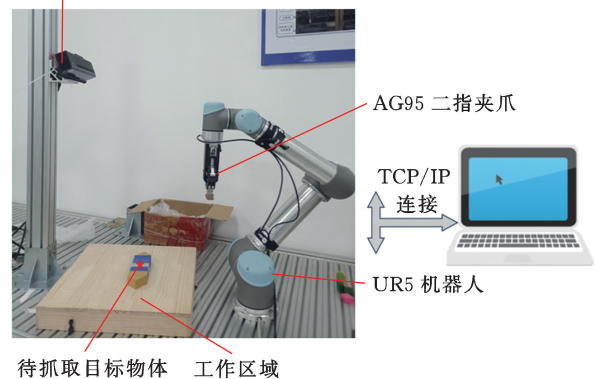


图 9 抓取实验平台

Fig.9 Grasping the experimental platform

计算平均值。从表 5 所示的实验结果看,与其他方法相比,本文方法在各项评价指标上都表现出

最优的性能,这表明本文方法在实际环境中能有效地完成目标物体抓取任务。

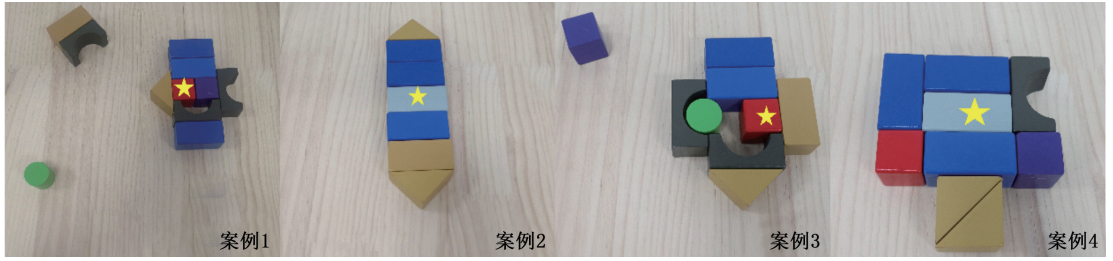


图 10 真实环境下四种测试案例

Fig.10 Four test cases in real environment

表 5 真实环境下四种测试案例实验结果

Tab.5 Experimental results of four test cases in real environment

模型	完成率/%	抓取成功率/%	动作数量
GI	85.0	70.3	6.81
EPG	95.0	86.6	4.62
本文方法	100	93.3	2.89

图 11 展示了案例 1 的一次完整实验过程。初始状态下,机器人首先通过对场景的判断选择执行推动动作,将目标物体(红色物块)推到一侧,为抓取创造条件。随后,机器人对打乱的场景再次进行判断,选择执行抓取动作,即在合适的抓取位置和抓取方向上成功抓取目标物体。仅仅通过两步动作,机器人就把被紧紧包围的目标物体成功抓取出来。

“抓取”协同推抓策略。该策略通过 2 个深度 Q 网络算法实现推动和抓取的相互配合。通过对抓取网络进行“抓-推-抓”分阶段训练,提高抓取网络的抓取能力,同时提出一种基于图像形态处理的评估判断方法来筛除质量较差的抓取动作。该方法在结构化测试场景中的完成率达到 100%,抓取成功率达到 95.0%,所用动作数量为 2.55;在随机化测试场景中的完成率达到 100%,抓取成功率达到 95.3%,所用动作数量为 2.48。与典型方法 GI(Grasping the Invisible)、EPG(Efficient Push-grasping)相比,本文方法的各项指标均最优,有效提高了目标物体的抓取成功率和动作效率。但本文中的抓取对象结构相对规整简单,后续工作将考虑使用更多种类和形状更复杂的物体进行实验,以扩展模型的适用场景,并验证其泛化性。

参考文献:

[1] DUAN Haonan, WANG Peng, HUANG Yayu, et al. Robotics Dexterous Grasping: the Methods Based on Point Cloud and Deep Learning[J]. Frontiers in Neurorobotics, 2021, 15:658280.

[2] BOHG J, MORALES A, ASFOUR T, et al. Data-driven Grasp Synthesis—a Survey[J]. IEEE Transactions on Robotics, 2014, 30(2):289-309.

[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] LIANG Hongzhuo, MA Xiaojian, LI Shuang, et al. PointNetGPD: Detecting Grasp Configurations from Point Sets [C] // 2019 International Conference on Robotics and Automation(ICRA). Montreal, 2019: 3629-3635.

[5] RODRIGUEZ A, MASON M T, FERRY S. From Caging to Grasping[J]. The International Journal of Robotics Research, 2012, 31(7):886-900.

[6] 贺辉腾,周勇,胡楷雄,等. 基于深度强化学习的机器人多动作协同抓取策略[J]. 计算机集成制造系统, 2024, 30(5):1789-1797.

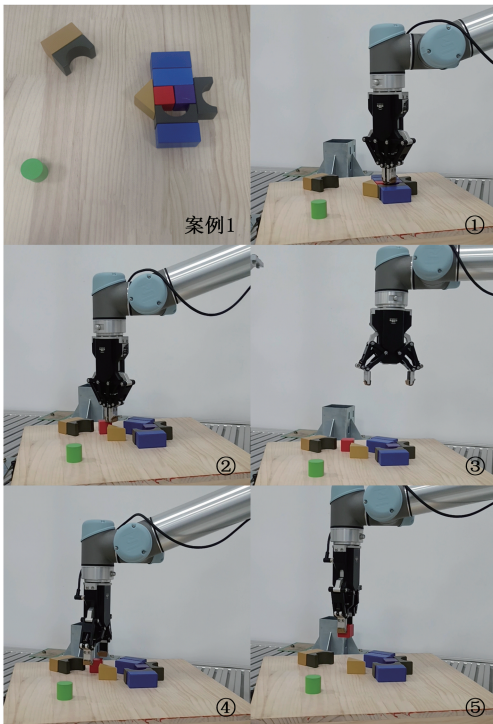


图 11 完整的抓取实验过程

Fig.11 A complete grasping experiment process

3 结语

本文提出一种基于深度强化学习的“推动”和

- HE Huiteng, ZHOU Yong, HU Kaixiong, et al. Robot Multi-action Cooperative Grasping Strategy Based on Deep Reinforcement Learning[J]. Computer Integrated Manufacturing Systems, 2024, 30(5):1789-1797.
- [7] 张胜文, 周曦, 李滨城, 等. 基于图像深度学习的零件加工特征信息提取方法[J]. 中国机械工程, 2022, 33(3):348-355.
- ZHANG Shengwen, ZHOU Xi, LI Bincheng, et al. Information Extraction Method of Part Machining Features Based on Image Deep Learning[J]. China Mechanical Engineering, 2022, 33(3):348-355.
- [8] 王勇, 陈荟西, 冯雨齐. 基于改进 CenterNet 的机械臂抓取检测[J]. 中南大学学报(自然科学版), 2021, 52(9):3242-3250.
- WANG Yong, CHEN Huixi, FENG Yuqi. Robotic Grasping Detection Based on Improved CenterNet[J]. Journal of Central South University (Science and Technology), 2021, 52(9):3242-3250.
- [9] 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 Hanbo, LAN Xuguang, ZHOU Xinwen, et al. Robotic Grasping in Multi-object Stacking Scenes Based on Visual Reasoning[J]. Scientia Sinica(Technologica), 2018, 48(12):1341-1356.
- [10] 孙先涛, 唐思宇, 陈文杰, 等. 复杂环境下基于推抓协同操作的目标物体抓取[J]. 控制理论与应用, 2023, 40(10):1713-1720.
- SUN Xiantao, TANG Siyu, CHEN Wenjie, et al. Target Object Grasp Based on Push-grasp Cooperative Operation in Complex Environment[J]. Control Theory & Applications, 2023, 40(10):1713-1720.
- [11] 李鑫, 沈捷, 曹恺, 等. 深度强化学习的机械臂密集场景多物体抓取方法[J/OL]. 计算机工程与应用, 2019:1-9[2023-10-23]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20231008.1644.008.html>.
- LI Xin, SHEN Jie, CAO Kai, et al. Deep Reinforcement Learning for Manipulator Multi-object Grasping in Dense Scenes[J/OL]. Computer Engineering and Applications, 2019:1-9[2023-10-23]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20231008.1644.008.html>.
- [12] TEN PAS A, PLATT R. Using Geometry to Detect Grasp Poses in 3D Point Clouds[M]//Robotics Research. Cham:Springer, 2018:307-324.
- [13] ZENG A, SONG Shuran, 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). Madrid, 2018:4238-4245.
- [14] DENG Yuhong, GUO Xiaofeng, WEI Yixuan, 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). Macau, 2019:619-626.
- [15] TANG Bingjie, CORSARO M, KONIDARIS G, et al. Learning Collaborative Pushing and Grasping Policies in Dense Clutter[C]//2021 IEEE International Conference on Robotics and Automation(ICRA). Xi'an, 2021:6177-6184.
- [16] YANG Yang, LIANG Hengyue, CHOI C. A Deep Learning Approach to Grasping the Invisible[J]. IEEE Robotics and Automation Letters, 2020, 5(2):2232-2239.
- [17] XU Kechun, YU Hongxiang, LAI Qianen, et al. Efficient Learning of Goal-oriented Push-grasping Synergy in Clutter[J]. IEEE Robotics and Automation Letters, 2021, 6(4):6337-6344.
- [18] HUANG Gao, LIU Zhuang, van der MAATEN L, et al. Densely Connected Convolutional Networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu, 2017:2261-2269.
- [19] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: a Large-scale Hierarchical Image Database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009:248-255.
- [20] NAIR V, HINTON G E. Rectified Linear Units Improve Restricted Boltzmann Machines[C]//Proceedings of the 27th International Conference on Machine Learning(ICML-10). Haifa, 2010:807-814.
- [21] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. 32nd International Conference on Machine Learning. Lille, 2015:448-456.
- [22] KINGMA D P, BA J. Adam: a Method for Stochastic Optimization[EB/OL]. 2014; arXiv:1412.6980. <http://arxiv.org/abs/1412.6980>

(编辑 张洋)

作者简介:胡楷雄,男,1985年生,副教授。研究方向为智能制造。发表论文20余篇。E-mail: kaixiong.hu@whut.edu.cn。
周勇*(通信作者),男,1973年生,副教授、博士研究生导师。研究方向为机器人技术及应用、物流装备协同作业调度与智能化。发表论文50余篇。E-mail:zhoyo@163.com。

本文引用格式:

胡楷雄,宋远航,周勇,等.基于深度强化学习的混杂场景目标物体推抓协同策略[J].中国机械工程,2025,36(1):133-140.

HU Kaixiong, SONG Yuanhang, ZHOU Yong, et al. A Cooperative Strategy for Pushing and Grasping Target Object in Cluttered Scenes Based on Deep Reinforcement Learning[J]. China Mechanical Engineering, 2025, 36(1):133-140.