

# 复杂环境下基于推抓协同操作的目标物体抓取

孙先涛<sup>1</sup>, 唐思宇<sup>1</sup>, 陈文杰<sup>1†</sup>, 贺春东<sup>1</sup>, 智亚丽<sup>1</sup>, 陈伟海<sup>2</sup>

(1. 安徽大学 电气工程与自动化学院, 安徽 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:** 针对现有抓取技术在复杂环境下难以进行有效的目标导向性抓取的问题, 本文提出了一种基于深度强化学习的推动和抓取协同操作的方法. 相对于以往的抓取方法, 本方法使用深度学习来处理Intel-D435i相机所获得的RGB-D图像数据, 同时又在视觉网络中引入了注意力机制, 用来提高系统对工作区域内目标物体的敏感性. 其次, 使用深度Q网络来学习UR5机械臂与环境之间的交互过程, 提出了密集奖励策略来评判推动或抓取操作的好坏. 随着训练次数的不断增加, UR5机械臂在训练过程中不断地优化两种操作之间的协同策略, 从而更高效的进行决策. 最后, 在V-rep仿真平台上设计了仿真场景, 并进行测试, 平均抓取成功率达到92.5%. 通过与其他几种方法进行对比, 证明该方法可以在复杂环境下较好的完成目标物体的抓取任务.

**关键词:** 深度强化学习; 神经网络; 机械臂抓取; Q网络

**引用格式:** 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. 控制理论与应用, 2023, 40(10): 1713 – 1720

DOI: 10.7641/CTA.2023.20682

## Target object grasp based on push-grasp cooperative operation in complex environment

SUN Xian-tao<sup>1</sup>, TANG Si-yu<sup>1</sup>, CHEN Wen-jie<sup>1†</sup>, HE Chun-dong<sup>1</sup>, ZHI Ya-li<sup>1</sup>, CHEN Wei-hai<sup>2</sup>

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

**Abstract:** Aiming at the problem that the existing grasping technology is difficult to carry out effective target-oriented grasping in complex environments, this paper proposes a method of promoting and grasping collaborative operation based on deep reinforcement learning. Compared with the previous grasping methods, this method uses deep learning to process the RGB-D image data obtained by the Intel-D435i camera, and at the same time introduces an attention mechanism into the visual network to improve the system's ability to detect target objects in the work area. Second, a deep Q network is used to learn the interaction process between the UR5 manipulator and the environment, and a dense reward strategy is proposed to judge the quality of pushing or grasping operations. It can also continuously optimize the collaborative strategy between the two operations, so as to capture more efficiently. Finally, the simulation scene is designed and tested on the V-rep simulation platform, and the average grasping success rate reaches 92.5%. By comparing with several other methods, it is proved that this method can better complete the grasping task of target objects in complex environments.

**Key words:** deep reinforcement learning; neural network; manipulator grasping; Q-network

**Citation:** SUN Xiantao, TANG Siyu, CHEN Wenjie, et al. Target object grasp based on push-grasp cooperative operation in complex environment. *Control Theory & Applications*, 2023, 40(10): 1713 – 1720

## 1 引言

机器人抓取是机器人领域的一项重要分支, 是机器人完成各种抓取任务的基础和关键. 随着机器人的日益普及, 机器人已经从传统的工业应用领域走向千家万户. 在固定单一的工作场景中进行简单的抓取已

经无法满足人们的需求, 机器人在实际的抓取过程中经常会面临复杂的工作环境, 比如多个物体紧密相连或者物体之间存在遮挡, 在这种情况下, 抓取任务对于机器人来说就是一种挑战.

机器人抓取方法大致可以分为两类: 经典分析

收稿日期: 2022-08-01; 录用日期: 2023-03-08.

†通信作者. E-mail: wjchen@ahu.edu.cn; Tel.: +86 13146000753.

本文责任编辑: 赵冬斌.

国家自然科学基金项目(52005001)资助.

Supported by the National Natural Science Foundation of China (52005001).

法<sup>[1]</sup>和数据驱动法<sup>[2]</sup>。经典分析法需要精确的物体三维模型来找到稳定的抓取力闭合,但是在大多数情况下很难获得物体的精确模型数据,并且这种方法不能很好的应用于新的物体上。随着以卷积神经网络为代表的深度学习技术的飞速发展,数据驱动的机器人抓取方法也成为了近年来的研究热点。

数据驱动法可以根据是否基于模型再分成两类。基于模型的数据驱动法通常使用物体检测和抓取位姿估计的组合来采集抓取样本。比如文献[3]中就采用基于模型的方法训练了一个卷积神经网络,用来生成点云中的6自由度(6-degree of freedom, 6-DOF)抓取姿态。这类方法大都会用到指定对象的信息,如形状、姿态等,只需要提前将模型数据发送给机器人,然后在抓取的时候对物体进行检测,机器人就可以进行抓取。这种方法在实际抓取中仅需要少量的运算,具有运行速度快的特点,缺点是推广到未知对象的能力有限,并且在建立数据模型时较为繁琐。

另外一种不需要模型的数据驱动法探索了与模型无关的抓取策略,这种策略将视觉信息和抓取直接联系起来。文中的方法基于无模型深度强化学习,使用深度学习来处理所获得的图像数据,同时使用强化学习来控制机器人进行决策,该方法可以很好的推广到新对象上。例如文献[4]中就设计了一个新的抓取系统,可以针对已知或者未知的物体预测多组抓取动作,然后选择成功率最高的一组动作执行,具有较好的泛化性。

但更多时候机械臂仅仅依靠抓取操作并不能完成复杂环境下的抓取任务,这就需要定义更加高级的操作。例如文献[5]中提出了一种Slide-to-wall动作,针对偏平物体不易抓取的问题,先将扁平物体移动到墙边,然后借助墙面将物体抓起来,但该方法应用场景有限。除此之外,推动操作可以有效的将紧密相连的物体分开,为抓取创造空间。推动与抓取操作的协同配合,极大地提高了机械臂在复杂环境下的抓取效率。为了能在复杂环境下减少杂波对抓取成功率的影响,文献[6-9]研究了推动等操作,以达到分离物体的目的,来提高抓取成功率。虽然目前有大量研究都致力于推动和抓取操作,但他们主要是单独研究的,比如文献[10]中使用Deep Q-learning算法让机器人将目标物体推动到一个位置。文献[11]中提出了Split-deep Q-learning的方法,使用两个Q网络结构学习推动操作,推动目标物体或者推动干扰物体,完成了复杂环境下对于目标物体的分离任务,并没有使用到抓取操作。

目前,有一部分学者专注于推抓协同操作的研究。比如文献[12-14]侧重于从背景中分割物体,然后根据背景中的信息来选择推动动作。但是在复杂环境中,物体的大部分很有可能被遮挡,这会导致物体与周围

环境的分割变得更加困难。文献[15]通过在环境上方的RGB-D(red green blue-depth)相机直接对深度图像进行处理,避免了场景分割的繁琐操作。文献[16]提出了一个无模型的基于深度强化学习的深度Q网络算法,将图像信息作为输入,使用两个并行的网络分别来预测抓取和推动动作,并选择一个最优的动作来执行。但是他们的研究并没有对抓取顺序做出规划,只是完成了桌面清理任务。文献[17]中研究了面向目标的抓取,但他们的环境是简单的,稀疏的,并不能很好的应用于复杂环境中。

针对上述提到的抓取顺序以及稀疏环境的问题,本文以复杂环境作为出发点,对抓取顺序做出规划,依次抓取目标物体。在本文的复杂环境中,这些物体大小不同,形状不一。本文还设置了多个干扰物体以增加复杂程度,并且这些物体没有任何排列规律,相互之间可能会存在遮挡、重叠、接触等情况,这会大大增加机械臂的识别与抓取难度。为了解决这些问题,文章以深度强化学习技术作为理论支撑,提出了一种基于深度Q网络的抓取方法,直接对RGB-D相机采集到的深度图像进行处理,然后确定目标物体,将图像输入到预测网络中,最终选择执行动作。本文主要贡献包括:1)在深度Q网络模型的基础上,引入注意力网络,提升了系统对目标物体的感知程度,在机器人与环境进行交互的过程中,注意力网络可以优化提取的工作空间特征,更高效地学习抓取策略,进而获得更好的抓取表现;2)针对机械臂在学习过程中难以获得有效奖励的问题,提出了密集奖励策略,提高了系统对奖励的敏感性,加快了系统的收敛速度。

## 2 基于深度Q网络的目标抓取

### 2.1 系统综述

系统由一个固定安装在工作空间上方的RGB-D相机采集RGB和深度图像数据。然后将采集到的RGB图像输入到预先训练好的语义分割模块中,来预测目标物体掩码。接着将RGB图像、深度图像和掩码图像在垂直方向上进行投影,来获得RGB高度图、深度高度图以及目标物体掩码高度图。将获得的3个高度图以22.5的角度依次旋转16次后输入到注意力网络中进行特征提取,然后再输入到PushNet和GraspNet网络中来预测所有像素点的期望Q值,得到分别对应于抓取和推动动作的像素级Q值表。最终采用贪婪策略将具有最大Q值的点转换成动作执行的位置点 $(x, y, z)$ ,同时获得执行动作的角度,选择出当前状态下最优的操作来执行。如果选择推送动作,系统将启动一个主动推送机制,以在推送动作的像素预测中使用最高Q值点执行推送动作。否则,系统将使用抓取动作Q值表中的最高Q值点执行抓取动作。重复上述过程,直到成功捕获工作区中的所有对象,如图1所示。

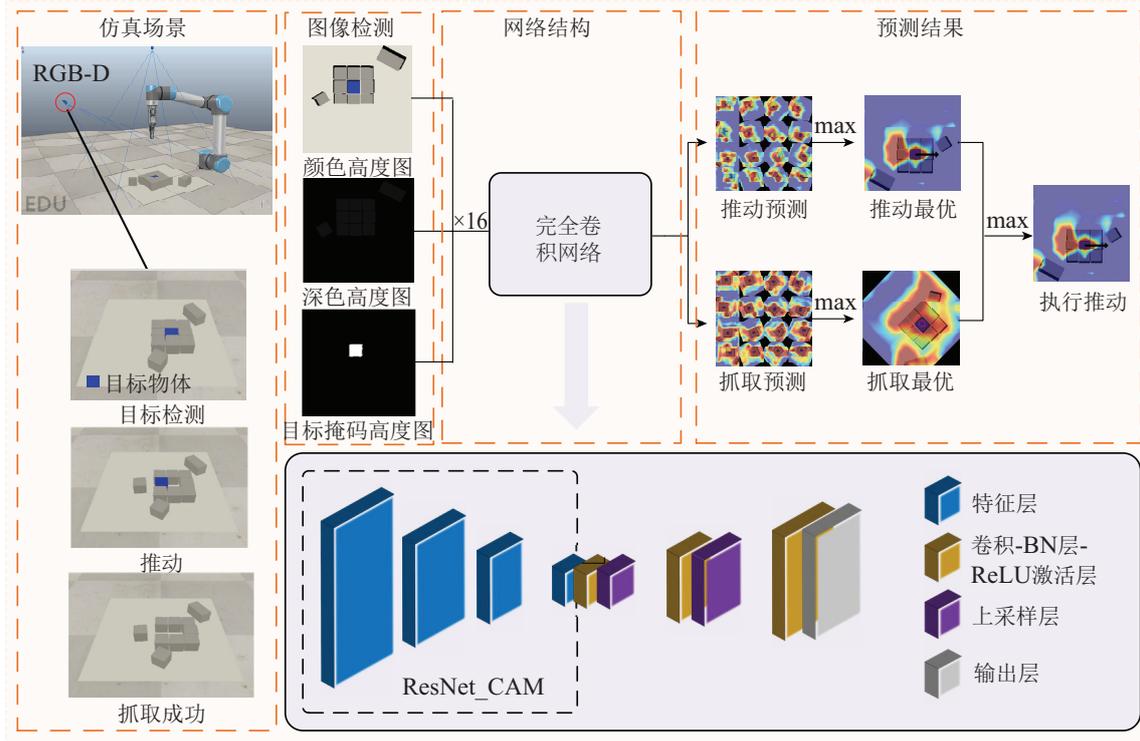


图 1 系统结构图

Fig. 1 System structure diagram

本文将这个问题建模为离散马尔科夫决策过程 (morkov decision process, MDP), 给定一个时间  $t$ , 机械臂通过工作空间上方的相机获得当前工作空间中物体的状态  $s_t$ , 并根据策略  $\pi$ , 选择一个执行动作  $a_t$ , 动作执行后, 工作空间的状态由  $s_t$  进入到  $s_{t+1}$ , 并获得相对应的奖励  $R_{at}(s_t, s_{t+1})$ , 将随机策略与深度 Q 网络进行结合, 得到状态-动作值函数  $Q_{t+1}$ , 计算公式如下:

$$Q_{t+1} = R_{at}(s_t, s_{t+1}) + \gamma \max_a (s_t, a', \theta_{t+1}), \quad (1)$$

其中:  $Q_{t+1}$  是执行动作的预测值;  $\gamma$  是折扣因子定值 0.5, 旨在减小当前奖励对未来决策的影响;  $R_{at}$  是执行动作  $a_t$  后获得的奖励值。

## 2.2 动作原语

为了协调推动和抓取操作, 文章将每一个动作  $a_t$  参数化为一个原始动作  $\psi$  (推动或抓取), 每次动作后的位置  $q$  会包含在状态  $s_t$  下的一个像素点  $p$  中, 表示为  $a = (\psi, q) | \psi \in \{\text{push, grasp}\}, q \rightarrow p \in s_t$ . 对于推动动作,  $q$  表示在  $k = 16$  个方向上推动的起始位置, 每次推动定义为 10 cm, 路径是直的, 由闭合夹爪来完成. 对于抓取操作,  $q$  表示在  $k = 16$  个方向上自上而下运动的二指抓手中心位置, 在做出抓取决策后, 抓手移动到目标位置, 然后竖直向下移动 3 cm 进行抓取. 在这两个动作中, 机械臂的路径规划是自动选择的, 不会发生碰撞. 之所以令  $k = 16$ , 是因为本文将平面上的场景分割成了 16 个方向, 每个方向依次旋转  $22.5^\circ$ ,

刚好满足  $16 \times 22.5^\circ = 360^\circ$ , 且在每个方向上只考虑机械臂在水平方向上的运动。

## 2.3 注意力网络模型 Resnet\_CAM

Resnet 网络是 He 等人<sup>[18]</sup>于 2015 年提出来的, 解决了卷积神经网络 (convolutional neural network, CNN) 随着深度的增加而导致分类精度下降的问题, 他们采用残差学习的方法来加速 CNN 的训练过程, 有效避免了梯度消失和梯度爆炸的问题。

本文为了进一步提高 Resnet50 网络的特征提取和表达能力, 将通道注意力模块 (channel attention model, CAM)<sup>[19]</sup>引入到 Resnet50 网络中, 网络结构如图 2 所示. 利用通道注意力机制进行残差拼接, 对每个特征通道的重要程度进行建模, 针对不同的任务需求来加强或者抑制不同的通道, 这个操作虽然增加了少量的参数计算, 但是获得了较大的性能提升. 通道注意力机制更加注重图像输入通道的信息, 通过提取通道特征信息来提高特征分类的精度. 首先将输入的特征分别进行全局最大池化和全局平均池化, 然后使用多层感知机 (multilayer perceptron, MLP) 拼接进入全连接层, 经过 Sigmoid 激活函数后进行输出, 最后通过乘法运算加权到之前的特征通道上, 以完成在通道维度上的特征重标定. 计算公式如下:

$$M_c(F) = \sigma(\text{MLP}(\text{Avg}(F)) + \text{MLP}(\text{Max}(F))), \quad (2)$$

式中:  $F$  表示输入网络的特征图;  $\sigma$  表示 Sigmoid 激活函数;  $M_c$  表示在 channel 维度上做 attention 提取操作。

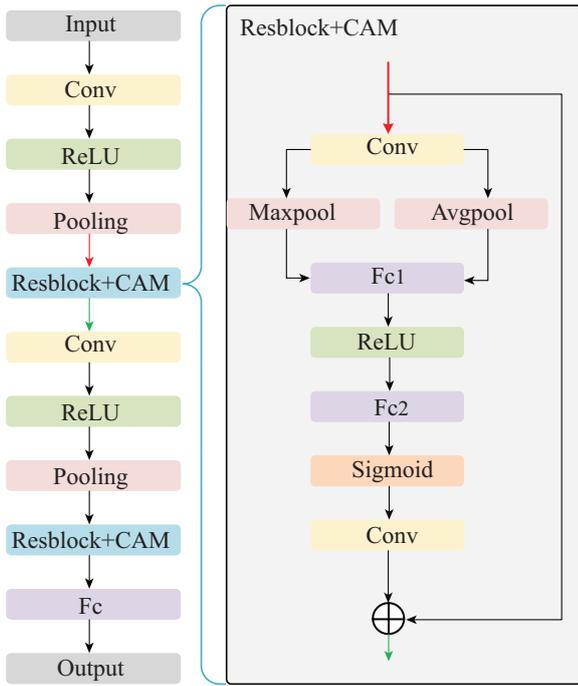


图2 基于CAM的Resnet网络模型  
Fig. 2 Resnet-CAM model

### 2.4 状态-动作值函数选择FCN

完全卷积网络 (full convolutional networks, FCN) 的出现改变了图像的像素级分类, 解决了语义级的图像分割问题. 本文采用前馈全卷积网络进行Q函数建模, 同时将FCN引入到DQN (deep Q network) 中, 做为全卷积作用值函数. 首先, 将动作空间离散为与场景图像像素相对应的像素级运动基元, 然后将机器人操作任务描述为一个像素级分割问题. FCN经过训练以后, 可以将工作空间的图像作为输入, 计算像素级预测Q值, 具有最高Q值的动作将在相应的像素位置执行. 与深度强化学习的端到端策略训练相比, 该方法降低了复杂性, 更有利于训练.

本文使用由两个前馈全卷积神经网络(GraspNet和PushNet)组成的深度Q网络分别对两个Q函数进行评估. 其中, GraspNet用于评估抓取Q函数, PushNet用于评估推动Q函数. 首先经过注意力网络对高度图像进行特征提取, 以产生运动不可知性, 然后将提取到的特征输入到GraspNet和PushNet网络中, 分别预测抓取Q值 $Q_g$ 和推动Q值 $Q_p$ . 最终在两张Q值图中选择Q值最高的像素点作为动作点.

### 2.5 密集奖励函数

在机械臂强化学习的任务中, 奖励函数发挥着重要的作用, 为了达到所需要的策略, 设计合适的奖励函数通常需要大量的领域专业知识以及反复试验. 为了最大限度地减少设计奖励函数所涉及的工作量, 提出了一种密集奖励函数的方法. 与最先进的高维奖励函数的方法相比, 本文不使用对抗性训练, 因此不容易出现训练的不稳定性. 实验结果也表明, 用密集奖励函数训练的策略具有更好的性能和更快的收敛速

度.

文章的奖励函数由动作前奖励和动作后奖励两个部分组成. 在非目标导向性抓取任务中, 机械臂只要抓取到工作空间内任意一个物体, 系统都会对其动作进行一个奖励. 而在本文抓取任务中, 机械臂每次抓取的目标只有一个, 只有当抓到这个目标物体时, 系统才会对其动作进行一个奖励, 机械臂更难获得有效奖励, 这样就会导致学习效率低下, 系统难以收敛或者需要成倍的训练次数才能收敛. 因此经过多次训练, 最终确定了密集奖励函数, 意在加速网络的收敛. 对于推动操作, 动作前会奖励0.25, 当推动成功时, 奖励0.5, 其中推动成功与否是根据目标物体周围的环境是否发生一定阈值的变化来确定的. 对于抓取操作, 动作前奖励0.5, 当抓取成功时, 奖励1. 综合以上所述, 可得奖励函数为

$$R(s_t, s_{t+1}) = \begin{cases} \begin{cases} 0.25, & \text{推动,} \\ 0.5, & \text{推动成功,} \end{cases} \\ \begin{cases} 0.5, & \text{抓取,} \\ 1, & \text{抓取成功.} \end{cases} \end{cases} \quad (3)$$

### 2.6 训练过程

FCN在每次迭代中都要使用Huber损失函数进行训练; 即

$$L_t = \begin{cases} \frac{1}{2} (Q^{\theta_t}(s_t, a_t) - y_t^{\theta_t})^2, & |Q^{\theta_t}(s_t, a_t) - y_t^{\theta_t}| < 1, \\ |Q^{\theta_t}(s_t, a_t) - y_t^{\theta_t}| - \frac{1}{2}, & \text{其他,} \end{cases} \quad (4)$$

如式(4)所示:  $t$ 为迭代次数;  $y_t^{\theta_t}$ 为目标值;  $\theta_t$ 是迭代时Q网络的参数;  $\theta_t^*$ 是目标Q网络的参数, 该目标Q网络周期性复制Q网络的参数, 在Q网络每一次更新时参数并不会发生变化, 以此来提高训练时网络的稳定性. 网络参数更新公式如下:

$$\theta_{t+1} = \theta_t + \alpha [r + \gamma \max_{a'} Q(s, a, \theta) - Q(s, a, \theta)] \nabla Q(s, a, \theta). \quad (5)$$

本文使用动量梯度下降的方法来进行训练, 具体参数设置如下: random seed为1234, 动量为0.9, 学习率为 $10^{-4}$ , 权重衰减为 $2^{-5}$ , 未来奖励折扣为0.5, 使用优先经验回放法. 机械臂遵循 $\epsilon$ -贪婪策略, 贪婪因子 $\epsilon$ 的最初取值为0.5, 随着训练次数的增加逐步衰减到0.1, 折扣因子恒定为0.5. 整体训练流程图及算法流程如图3和表1所示.

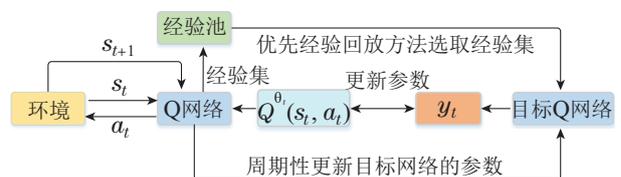


图3 系统结构图

Fig. 3 System structure diagram

表 1 伪代码

Table 1 Pseudocode

算法流程
输入: 迭代次数 <i>i</i> , 贪婪策略, 状态 <i>s</i> , 图像 <i>I</i> , 目标 <i>T</i> , 掩码 <i>M</i> , 动量, 学习率, 权重, 奖励折扣, 推动和抓取网络
输出: 动作 <i>a<sub>t</sub></i> , 执行位置 <i>p</i>
$i \leftarrow 0$
While true do
重置仿真场景, 并随机掉落物块
随机选择一个目标物体 <i>T</i>
$i \leftarrow i + 1$
$M \leftarrow \text{Object segmentation}(I, T)$ 计算目标物体掩码
$s_t \leftarrow \text{Heightmap projection}(I, M)$ 计算系统当前状态
$Q_p \leftarrow \psi_p(s_t), Q_g \leftarrow \psi_g(s_t)$ 计算推动 <i>Q</i> 值和抓取 <i>Q</i> 值
$a_t \leftarrow \arg \max_a Q_y$ 执行具有最大 <i>Q</i> 值的动作
$R_{at} \leftarrow \text{Reward function}$ 计算奖励值
保存 <i>s<sub>t</sub></i> , <i>a<sub>t</sub></i> , <i>R<sub>at</sub></i> 到经验池中
网络参数更新
End while

### 3 仿真训练与测试

#### 3.1 仿真训练

为了验证本文所提出方法在复杂环境下抓取时的优越性, 文章在Ubuntu16.04上搭建了仿真平台, 所使用的CPU为Intel i7-11800 H, GPU为NVIDIA RTX-3050, 运行环境为Pytorch1.9. 训练在V-Rep仿真软件上进行, 配备有UR5机械臂, RG2抓手, Intel D435i相机, 以及一个大小为 $0.224 \times 0.224 \text{ m}$ 的工作平台. 为了模拟实际工作中的情况, 会随机选择7个候选目标对象(即语义分割模块可以检测出来的对象)和*n*个基本对象(即干扰对象). 让这些随机产生的对象自由落到工作平台上. 机械臂需要通过一系列的操作将目标对象抓起来, 一旦机械臂成功抓取目标物体, 系统就对工作空间内的物体重新进行检测, 随机选择下一个目标对象. 当工作空间内没有目标对象时, 系统会清空工作平台, 并重新随机产生对象进行下一次训练.

训练分为两个阶段进行, 前1000次训练视为第1个阶段, 将*n*设置为3, 系统主要学习如何成功的将目标物体抓取起来. 在第2个阶段, 将*n*设置为8, 增加工作场景的杂乱程度, 这一过程中系统主要学习推动和抓取操作之间的协同关系, 如图4所示.

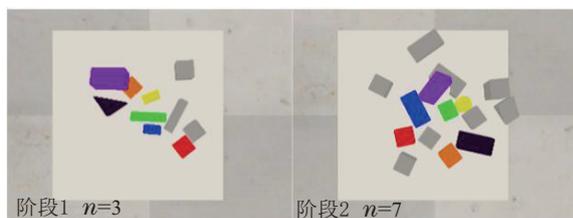


图 4 训练阶段1, 2

Fig. 4 Training phases 1, 2

#### 3.2 训练结果与分析

为了验证本文所提出方法的有效性, 和其他4种不同的策略进行了对比. 每组训练进行2500次. 在训练完成后绘制每组训练的抓取成功率曲线图. 具体如下.

1) Grasping-only<sup>[4]</sup>是一种仅采用抓取策略的方法, 使用了全卷积网络, 直接对输入的图像进行视觉观测, 输出一个稠密的网格图, 表示从每个位置预测4个抓取动作后成功的概率, 机器人最终执行成功率最高的抓取动作.

2) VPG (visual pushing grpsping)-Target<sup>[16]</sup> 给定目标物体, 也是通过两个全卷积动作网络来预测*Q*值, 学习推动和抓取之间的操作关系.

3) GIT (grasping invisible target)<sup>[20]</sup>是一种基于分类器的策略来协调推动和抓取之间的关系, 以在杂乱中抓住目标物体.

4) VAR (visual attribute reasoning)<sup>[21]</sup>提出了一种具有局部关系理解的分层机器人策略, 可以在复杂环境中抓住目标物体.

训练结果如图5所示.

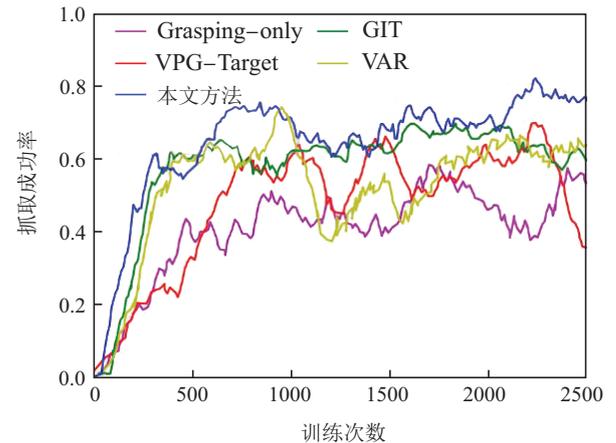


图 5 训练结果图

Fig. 5 Training process curves

从训练结果曲线图中不难看出, Grasping-only方法表现最差, 抓取成功率只有40% ~ 60%, 这也证明了在面对复杂环境时, 仅仅使用抓取的方法难以完成目标抓取任务. VPG-Target学习了推动与抓取之间的协同关系, 但也仅是为了改变环境而使用推动操作, 有时会执行一些无效推动, 学习效率低下, 训练数据也并不理想. GIT和VAR方法通过分类器来选择推动或者抓取操作, 提高了两种动作的协同工作效率. 本文中所提出的方法整体上优于上述几种方法, 在训练2500次左右时就可以达到80%的抓取成功率. 在训练第1阶段, 而本文方法抓取成功率上升的速度明显比另外几种方法快, 并且由于注意力机制的引入, 在训练进入到第2阶段时, 训练数据曲线在稳定性上的表现也优于其他几种方法.

### 3.3 测试结果与分析

本文使用8个场景进行测试, 每个场景执行20次抓取任务, 若机械臂在5次动作以内将目标物体抓起来则视为抓取成功, 抓取成功率=抓取成功的次数/测试总次数. 同时记录机械臂所操作的次数, 平均操作次数=(推+抓)总操作数/测试总次数. 用于测试的8个场景如图6所示. 测试结果如表2和图7-8所示.

图7反映了上面提到的几种方法在8个测试场景下的抓取成功率, 从图中可以看出, Grasping-only方法仅仅使用抓取操作, 难以适应复杂的测试场景, 抓取成功率都比较低. 而其他几种方法使用了推抓协同的操作, 在成功率上都取得了较大的提升, 这表明推动操作的引入可以有效的解决复杂环境下的抓取问题.

图8反映了上面提到的几种方法在8个测试场景下的平均动作数. 从上面两张图中可以看出, 本文所提出的方法可以使用较少操作次数的同时取得较高的抓取成功率, 明显优于其他几种方法.

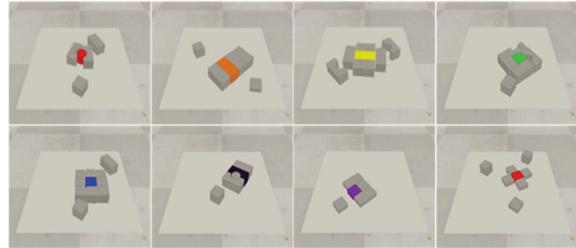


图6 测试场景图

Fig. 6 Test scenario

表2 不同场景测试数据表

Table 2 Different scenario test data table

方法 测试场景	Grasping-only		VPG-Target		GIT		VAR		本文方法	
	成功率/%	动作次数	成功率/%	动作次数	成功率/%	动作次数	成功率/%	动作次数	成功率/%	动作次数
场景1	40	4.4	70	4.2	95	3.8	95	3.8	90	3.4
场景2	20	3.8	65	4.2	75	4.1	85	3.4	100	2.5
场景3	35	4.8	70	4.0	85	3.8	85	3.9	85	4.2
场景4	50	3.8	80	3.2	90	3.2	95	2.6	100	2.8
场景5	20	4.6	60	4.4	85	3.7	85	4.1	80	4.7
场景6	25	4.6	60	4.2	85	4.2	80	3.9	100	3.5
场景7	40	4.8	70	4.2	95	3.2	85	2.8	95	3.3
场景8	50	3.8	80	3.8	90	3.4	90	3.6	100	2.6

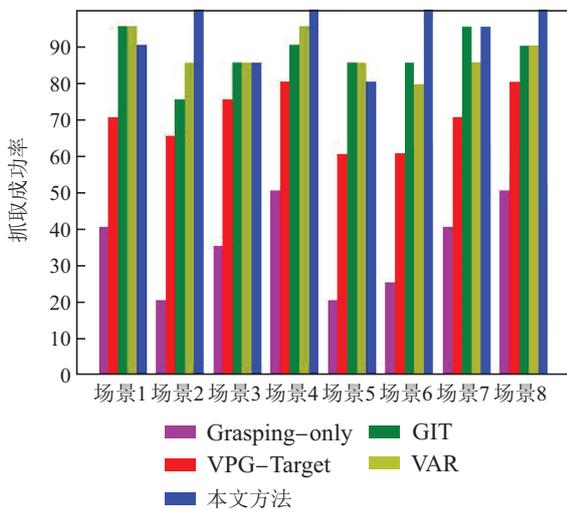


图7 抓取成功率柱状图

Fig. 7 Histogram of capture success rate

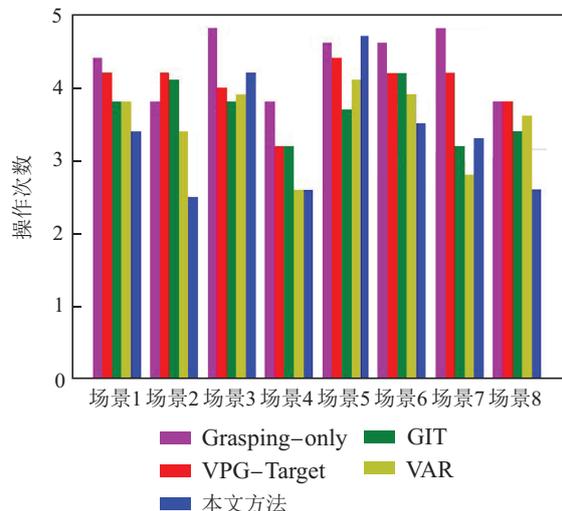


图8 抓取动作数柱状图

Fig. 8 Histogram of grabbing action number

从表格3中可以看出, 本文所提出的方法平均抓取成功率达到了92.5%, 平均抓取所需次数仅为3.375, 无论是在抓取成功率还是在运动次数上的表现都明显优于另外几种方法. 这表明本方法可以有效完成杂乱环境中的目标物体抓取任务.

### 3.4 消融实验

为了证明本文引入注意力机制和提出密集奖励策略的有效性, 文章进行了消融对比研究, 结果如下表4所示.

消融研究结果如图9-10及表4所示, 从中不难看出

出, 注意力网络和密集奖励策略的使用使得系统Q值预测更为准确, 在训练过程中的损失也更小. 在抓取成功率和平均操作次数上也有较大提升.

表 3 测试数据表

Table 3 Test data table

方法	成功率/%	动作次数
Grasping-only	35.0	4.325
VPG-Target	70.0	4.025
GIT	87.5	3.675
VAR	87.5	3.5125
本文方法	92.5	3.375

表 4 消融研究数据表

Table 4 Ablation study data sheet

方法	成功率/%	动作次数
注意力+密集	92.5	3.375
注意力+稀疏	84.0	4.175
无注意力+密集	75.0	4.825

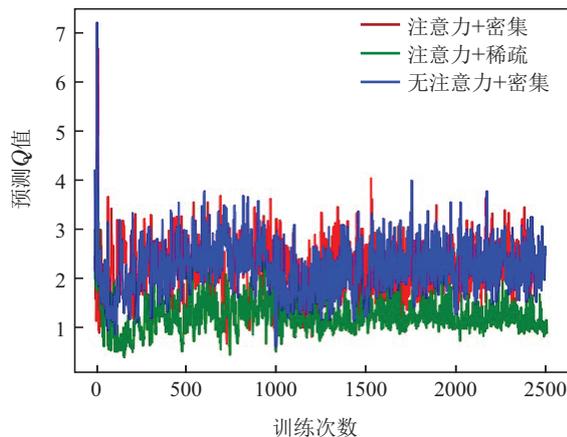


图 9 预测Q值

Fig. 9 Predicted Q value

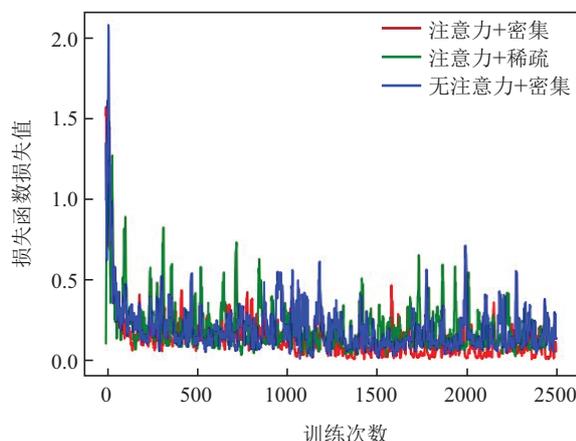


图 10 损失值

Fig. 10 Loss value

此外, 在表5中还能看出, 当系统使用稀疏奖励策略训练时, 机械臂更倾向于使用推动操作. 在2500次的训练过程中, 系统使用了1056次推动操作. 而使用密集奖励策略训练时, 系统仅使用了约650次的推动操作. 其中的原因是系统在使用稀疏奖励策略时, 获得奖励的可能性降低, 故采用更为保守的推动操作以获得稳定而较小的奖励值.

表 5 训练数据表

Table 5 Ablation training data

方法	注意力+密集	无注意力+密集	注意力+稀疏
抓取目标物体	1011	967	767
抓取非目标物体	15	21	27
抓取失败	792	873	650
推动成功	435	265	621
推动失败	247	374	435
合计	2500	2500	2500

#### 4 结束语

本文使用深度Q学习的方法, 通过引入推动的操作, 克服了机器人面对复杂环境时抓取效率低, 学习速度慢以及推抓动作配合不理想的问题, 提升了机器人在复杂环境下的目标抓取性能.

本文提出了视觉注意力网络, 生成更为精确的推抓动作功用性图; 还提出了密集奖励策略, 加快系统的收敛速度, 使得系统在更短的时间内获得更大的性能提升. 在训练阶段提升推抓动作的效率, 在测试阶段提升抓取成功率, 降低平均动作次数. 此外, 还设计了消融对比实验, 以验证视觉注意力网络和密集奖励策略的有效性.

即便文章已经充分考虑到实际环境中的一些影响因素, 但模型的输入数据来源于仿真场景, 不可避免的会和现实场景存在一些偏差. 因此, 仿真训练的结果模型迁移至真实场景中也会存在一定偏差. 为了提升模型性能和迁移能力, 笔者在后续研究中会搭建真实的机器人抓取平台, 以验证所提的方法在现实生活场景中的可靠性, 从而促进机器人抓取技能在人类生产生活中的应用.

#### 参考文献:

- [1] WEISZ J, ALLEN P K. Pose error robust grasping from contact wrench space metrics. *IEEE International Conference on Robotics and Automation*. Saint Paul, USA: IEEE, 2012: 557 – 562.
- [2] BOHG J, MORALES A, ASFOUR T, et al. Data-driven grasp synthesis—a survey. *IEEE Transactions on Robotics*, 2013, 30(2): 289 – 309.
- [3] LI Huijun, QU Xiaochang, YE Bin. Six-degree-of-freedom robot grasping based on three-dimensional point cloud features of unknown objects. *Control Theory & Applications*, 2022, 39(6): 1103 – 1111. (李会军, 瞿孝昌, 叶宾. 基于未知物体三维点云特征的机器人6自由度抓取. *控制理论与应用*, 2022, 39(6): 1103 – 1111.)

- [4] LOU X, YANG Y, CHOI C. Collision-aware target-driven object grasping in constrained environments. *IEEE International Conference on Robotics and Automation*. Xi'an, China: IEEE, 2021: 6364 – 6370.
- [5] LIANG H, LOU X, YANG Y, et al. Learning visual affordances with target-orientated deep Q-network to grasp objects by harnessing environmental fixtures. *IEEE International Conference on Robotics and Automation*. Xi'an, China: IEEE, 2021: 2562 – 2568.
- [6] EITEL A, HAUFF N, BURGARD W. Learning to singulate objects using a push proposal network. *Robotics Research*. Cham: Springer, 2020: 405 – 419.
- [7] BAUZA M, RODRIGUEZ A. A probabilistic data-driven model for planar pushing. *IEEE International Conference on Robotics and Automation*. Singapore: IEEE, 2017: 3008 – 3015.
- [8] LLOYD J, LEPORA N F. Goal-driven robotic pushing using tactile and proprioceptive feedback. *IEEE Transactions on Robotics*, 2021, 38(2): 1201 – 1212.
- [9] DANIELCZUK M, MAHLER J, CORREA C, et al. Linear push policies to increase grasp access for robot bin picking. *IEEE the 14th International Conference on Automation Science and Engineering*. Munich: IEEE, 2018: 1249 – 1256.
- [10] WU J, SUN X, ZENG A, et al. Spatial action maps for mobile manipulation. *ArXiv Preprint*, arXiv: 2004.09141, 2020.
- [11] SARANTOPOULOS I, KIATOS M, DOULGERI Z, et al. Split deep Q-learning for robust object singulation. *IEEE International Conference on Robotics and Automation*. Paris: IEEE, 2020: 6225 – 6231.
- [12] KATZ D, VENKATRAMAN A, KAZEMI M, et al. Perceiving, learning, and exploiting object affordances for autonomous pile manipulation. *Autonomous Robots*, 2014, 37(4): 369 – 382.
- [13] EITEL A, HAUFF N, BURGARD W. Learning to singulate objects using a push proposal network. *Robotics Research*. Cham: Springer, 2020: 405 – 419.
- [14] BOULARIAS A, BAGNELL J A, STENTZ A. Learning to manipulate unknown objects in clutter by reinforcement. *The 29th AAAI Conference on Artificial Intelligence*. Austin, USA: AAAI, 2015: 1336 – 1342.
- [15] ZENG A, SONG S, WELKER S, et al. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Madrid: IEEE, 2018: 4238 – 4245.
- [16] JANG E, DEVIN C, VANHOUCKE V, et al. Grasp2vec: Learning object representations from self-supervised grasping. *ArXiv Preprint*, arXiv: 1811.06964, 2018.
- [17] FANG K, BAI Y, HINTERSTOISSER S, et al. Multi-task domain adaptation for deep learning of instance grasping from simulation. *IEEE International Conference on Robotics and Automation*. Brisbane: IEEE, 2018: 3516 – 3523.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770 – 778.
- [19] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation. *IEEE/CVF Conference on Computer Cision and Pattern Recognition*. Long Beach: IEEE, 2019: 3146 – 3154.
- [20] YANG Y, LIANG H, CHOI C. A deep learning approach to grasping the invisible. *IEEE Robotics and Automation Letters*, 2020, 5(2): 2232 – 2239.
- [21] YANG Y. *Target-driven robotic manipulation with visual attribute reasoning*. Minnesota: University of Minnesota ProQuest Dissertation Publish, 2022.

#### 作者简介:

**孙先涛** 博士, 副教授, 目前研究方向为视觉抓取、欠驱动机械手,

E-mail: xtsun@ahu.edu.cn;

**唐思宇** 硕士研究生, 目前研究方向为智能机器人抓取, E-mail:

937599854@qq.com;

**陈文杰** 教授, 博士生导师, 目前研究方向为智能机器人灵巧抓取、协作机器人共融操作、助力外骨骼及蛇形机器人, E-mail: wjchen@ahu.edu.cn;

**贺春东** 博士, 目前研究方向为声信号处理、噪声与振动控制, E-mail: chundong\_09@126.com;

**智亚丽** 博士, 目前研究方向为故障诊断、间歇控制、人工智能, E-mail: zhiyali@ahu.edu.cn;

**陈伟海** 教授, 目前研究方向为机器人视觉、自主移动机器人控制、仿生机器人、康复机器人, E-mail: whchenbuaa@126.com.