计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.010120] [CSTR: 32024.14.csa.010120]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

复杂环境下机械臂目标导向的推抓协同①

孟军英[1,2], 宋子涵[1], 于平平[1], 赵晓东[1], 张　丹[1]

[1](河北科技大学 信息科学与工程学院, 石家庄 050018)
[2](石家庄学院 未来信息技术学院, 石家庄 050035)
通信作者: 于平平, E-mail: yppflx@aliyun.com

摘　要: 在复杂堆叠环境中, 引入推动动作辅助机械臂抓取可以提升抓取成功率. 然而, 现有推抓协同方法中存在网络特征提取能力不足与推动策略低效等问题. 针对上述问题, 本文提出一种改进的基于深度 Q 网络 (DQN) 的推抓协同算法. 该方法在感知-动作策略网络中引入高效多尺度注意力 (efficient multi-scale attention, EMA) 机制, EMA 模块通过通道分组与跨空间建模增强对物体边缘、物体表面等关键任务特征的提取能力; 同时, 设计基于图像频域能量变化与能量质心位移的推动有效性评估机制, 构建更具判别力的奖励函数, 以引导智能体学习有效的推抓协同策略. 在 CoppeliaSim 仿真环境平台上的实验表明, 本文方法相较于 METOVPG 等基线方法, 在抓取成功率和动作效率方面均有显著提升. 其中, 在仿真环境下测试抓取成功率提升 21.2%, 验证了所提注意力机制与奖励设计在复杂场景下的有效性与协同优势.

关键词: 深度 Q 网络; 机械臂抓取; 图像频域能量; 注意力机制

引用格式: 孟军英,宋子涵,于平平,赵晓东,张丹.复杂环境下机械臂目标导向的推抓协同.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/10120.html

Cooperative Pushing and Grasping of Manipulators with Target Orientation in Complex Environments

MENG Jun-Ying[1,2], SONG Zi-Han[1], YU Ping-Ping[1], ZHAO Xiao-Dong[1], ZHANG Dan[1]

[1](School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China)
[2](College of Future Information Technology, Shijiazhuang University, Shijiazhuang 050035, China)

Abstract: In complex, cluttered environments, introducing pushing actions to assist robotic grasping can improve the success rate. However, existing push-grasp collaboration methods face issues such as insufficient feature extraction capability and inefficient pushing strategies. To address these problems, this study proposes an improved push-grasp collaborative algorithm based on deep Q-network (DQN). The proposed method integrates an efficient multi-scale attention (EMA) mechanism into the perception-action policy network. The EMA module enhances the extraction of key task features, including object edges and surfaces, through channel grouping and cross-spatial modeling. In addition, a push effectiveness evaluation mechanism is designed based on changes in image frequency energy and the displacement of the energy centroid, which contributes to constructing a more discriminative reward function. This function guides the agent to learn an effective push-grasp collaborative strategy. Experiments conducted on the CoppeliaSim simulation platform show that the proposed method significantly outperforms baseline methods such as METOVPG in terms of grasp success rate and action efficiency. Specifically, it achieves a 21.2% relative improvement in the grasp success rate in simulation tests, validating the effectiveness and collaborative advantages of the proposed attention mechanism and reward design in complex scenarios.

Key words: deep Q-network (DQN); manipulator grasping; image frequency energy; attention mechanism

① 基金项目: 国家重点研发计划 (2024YFD2402205); 河北省高等学校科学技术研究项目 (QN2025371)
收稿时间: 2025-08-26; 修改时间: 2025-10-10, 2025-11-03; 采用时间: 2025-11-17; csa 在线出版时间: 2026-03-02

机械臂的抓取操作是机器人实现物理交互的核心能力之一, 其性能直接决定了工业分拣、家庭服务等场景的应用潜力. 机器人在非结构化场景中对未知物体的鲁棒抓取仍是一项具有挑战性的任务. 现有的抓取方法可以分为传统的几何分析方法与数据驱动的方法. 传统的几何分析方法[1,2]需要对周围环境以及目标物体进行精确建模, 在已知对象场景中表现良好, 但其对于未知场景的决策能力有一定的局限性. 而数据驱动的方法通过从大量的数据中学习策略, 能够更好地适应复杂多变的环境.

在数据驱动方法的研究中, 研究者们[3–6]利用目标检测网络和位姿估计网络对目标物体进行识别并估计物体抓取位姿. 此方法需要大量的数据集进行标注, 标注成本高且对于复杂堆叠场景中的遮盖物体适应性差. 近年来, 基于深度强化学习 (deep reinforcement learning, DRL) 的抓取方法被提出, 将深度学习的特征提取能力与强化学习的序列决策能力结合, 使得机器人能够通过试错自主学习环境交互策略. 在复杂堆叠场景中, 当物体被相邻物体紧密包围时可抓取空间小. 因此引入推动动作改变物体空间分布, 使目标物体暴露在可操作区域. Zeng 等人[7]首次提出基于 DRL 的机械臂视觉抓取与推动框架 (VPG). 通过智能体与环境的不断交互, 学习到较优的推动和抓取策略, 其利用前后深度图像差值作为判断推动的有效性条件. Yu 等人[8]利用动作前后 RGB 图像的差值来判断推动是否有效. Phan 等人[9]引入推出惩罚项, 以补偿仅依赖图像差分判断环境变化的不足. 但以上 3 种方法均未能有效解决物体分离状态的判定问题. Xu 等人[10]设计了 3 阶段目标导向交替训练机制的分离推抓策略网络, Yang 等人[11]通过训练基于贝叶斯的探索策略以及基于动作分类器的策略选取当前状态下的动作. 文献[10,11]通过判断物体周围障碍物面积占比判断推动是否有效, 其并不能高效评判目标物体是否与障碍物体做分离运动. Sarantopoulos 等人[12]提出 Split-deep-Q-learning 的方法, 通过利用两个 Q 网络来学习推动动作, 使得目标物体从复杂环境中分离出来. 但其奖励只考虑推动动作将目标分离, 未优化物体的推动行为. Yu等人[13]设计了一个结合推动预测与抓取评估方法, 通过预测推动后的物体布局, 实现对后续抓取动作的动态优化与协调选择. Yang 等人[14]通过设计一个推动评估网络来判断当前推动动作是否使物体分离. Gao 等人[15]设计了一个基于全局目标分布的评估网络, 用于在推动奖励函数中分析动作前后物体是否分离. 文献[13–15]的方法通过引入评估网络提升推抓协同性能, 但该评估网络需要收集大量有效数据样本进行训练. 左国玉等人[16]通过引入坐标注意力机制来动态增强物体处的 Q 值热力图, 但其并未增强模型自身特征提取的能力. 综上所述, 现有基于深度强化学习的推抓方法虽在复杂场景中展现出一定的自适应性, 但仍存在感知与动作策略网络特征提取能力不足以及推动有效性评估低效的问题. 为此, 本文针对感知-动作网络的特征提取能力不足与推动有效性判断低效这两个关键问题进行了改进.

首先, 对于感知-动作策略网络关键特征提取能力较低, 动态环境中需要实时调整特征关注区域以匹配策略需求的问题, 本文在感知与动作策略网络中引入高效多尺度注意力机制 (efficient multi-scale attention, EMA)[17]. 通过通道分组与多尺度建模, 动态增强具有高语义价值的通道响应, 从而提升模型对关键特征的提取能力. 同时, 利用 EMA 并行多尺度卷积分支提取全局空间信息生成空间注意力图, 利用跨空间建模机制将不同分支输出的空间注意力图进行聚合, 调整动作决策对感知特征的关注区域. 针对判断推动动作有效性低的问题, 通过对图像频域能量差异量化来分析物体运动模式, 构建环境动作敏感的奖励函数以增强模型对有效推动动作的识别能力.

本文的研究聚焦面向复杂堆叠场景的机械臂推抓协同策略优化, 主要贡献如下.

(1) 提出一种融合 EMA 注意力机制的感知-动作策略网络, 增强特征提取与决策耦合的自适应性.

(2) 设计基于频域能量变化的推动奖励函数, 实现推动动作有效性的动态量化评估.

1 相关理论

目标导向的推抓协同任务中, 机械臂推抓问题被建模为一个马尔可夫决策过程 (MDP)[18]. 机械臂在当前状态S_t下, 根据策略π选择相应的动作a_t, 转移到下一状态S_{t+1}获得对应的奖励$R(a_t)$.

策略π的目标是通过最大化累积折扣奖励$G_t=\sum_{k=0}^{\infty}\gamma^k R_{t+k}G_t$来优化整个决策序列.

本文采用 off-policy 的 Q-learning[19]来训练策略π, 通过最大化动作值函数$Q_\pi(S_t,a_t)$来选择动作.

$$y_t = R_{a_t}(S_t, S_{t+1}) + \gamma Q(S_{t+1}, \text{argmax}_{a'}(Q(S_{t+1}, a'))) \quad (1)$$

其中, a'代表所有可用的动作集合, γ表示折扣因子.

2 机械臂目标导向推抓协同算法

2.1 系统概述

本文提出一种面向机械臂目标导向操作的深度强化学习算法, 用于实现推抓协同控制. 其系统整体架构如图 1 所示.

该系统首先将固定摄像头采集的图像数据通过转换为 3D 点云数据后垂直方向正交投影得到高度图像; 同时, 基于语义分割模型获取的目标掩码生成掩码高度图. 上述 3 类图像以固定角度旋转生成多视角状态表示. 将高度图输入 EMA-DenseNet121 感知网络提取多尺度特征, 随后分别输入引入 EMA 模块的抓取与推动的全卷积策略网络以预测像素级 Q 值热力图. 两个 Q 值热力图分别表示在各像素位置执行抓取或推动动作的预期收益. 最终, 根据最优动作对应的最大 Q 值对应的抓取位姿驱动机械臂执行, 直至目标物体被成功抓取为止. 本方法在 DenseNet121 感知网络的前 3 个密集块之后嵌入 EMA 注意力机制, 并在全卷积策略网络的输入层同样引入该模块, 以增强对高层图像特征的提取与融合能力. 智能体通过与环境交互获取奖励信号, 并基于频域能量变化评估推动动作的有效性. 由于堆叠物体在推动过程中会引起边缘等频域敏感区域的显著响应, 该机制能够辨识并保留有效的推动动作, 从而引导模型学习高效的推抓协同策略, 实现从感知到执行的端到端优化.

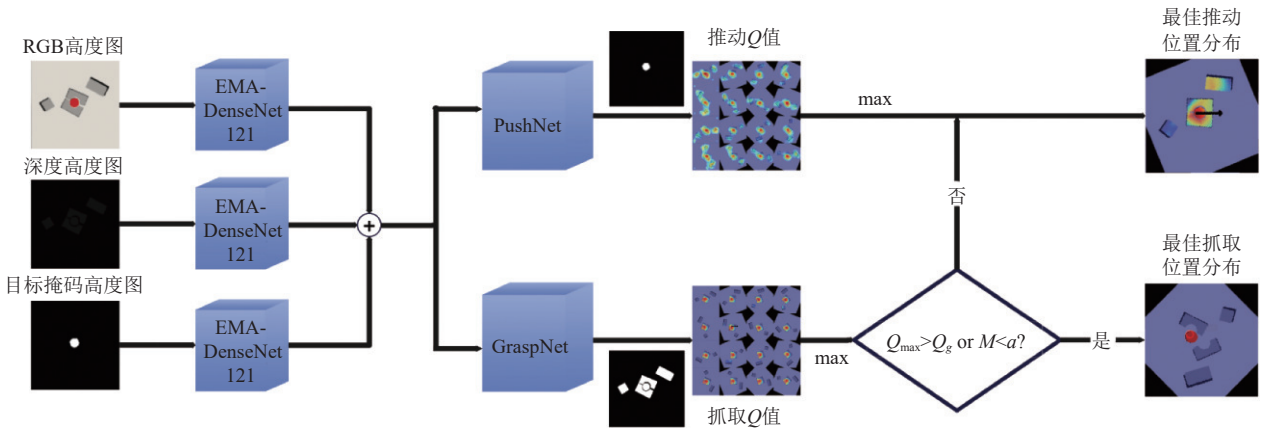


图 1 系统框图

2.2 状态表示

将固定在工作区域上方的摄像头捕获的 RGB 图与深度图数据转换至 3D 点云后, 再将其进行垂直方向的投影, 获得相应的 RGB 高度图像与深度高度图像, 紧接着利用预训练的颜色语义分割模型分割出目标掩码高度图. 将这 3 种高度图以 22.5°的间隔旋转生成多视角状态表示 S.

2.3 动作设计

定义每个操作动作a_t是由两个动作原语 (push, grasp) 组成的动作类型φ与动作位姿ψ这两部分组成. 机械臂的每个操作动作位姿定义为:

$$\psi_r = (P, \theta_r, Q) \quad (2)$$

其中, $P = (x, y, z)$是夹持器中心位置, θ_r是夹持器绕 z 轴的旋转角度, Q 是对应动作的置信度分数. 图像中的动作位姿ψ_i按照像素位置定义为:

$$\psi_i = (x, y, \theta_i, Q) \quad (3)$$

其中, x, y为图像坐标中的动作姿态中心, θ_i为图像旋转角度, Q 为动作对应的置信度分数. 定义高级动作原语如下.

推动动作$\psi_{\text{p}} = (x, y, \theta_i, Q)$为机械臂在夹爪关闭的状态下, 从$(x, y)$起自上而下沿着方向$\theta_i$水平移动 10 cm.

抓取动作$\psi_{\text{g}} = (x, y, \theta_i, Q)$为机械臂运动到$(x, y)$, 打开夹爪, 夹爪旋转角度$\theta_i$, 自上而下抓取.

2.4 感知-动作策略网络设计

为了实现基于视觉输入的像素级动作决策, 本文采用全卷积网络 (fully convolutional network, FCN)[20] 结构与深度 Q 网络 (deep Q-network, DQN) 相结合的方式, 对输入状态图像进行逐像素 Q 值预测. FCN 能

够在保持空间分辨率的同时输出与输入图像尺寸一致的 Q 值映射, 从而实现对每个像素位置执行动作的价值评估.

2.4.1 感知网络

感知网络基于预训练的 DenseNet121 网络构建. DenseNet121 通过密集连接机制实现了特征的高效复用, 但其对空间-通道联合信息的建模能力有限, 难以在复杂场景中灵活聚焦于关键区域. 为此, 本文在预训练的 DenseNet121 的骨干网络基础上引入了一种高效多尺度注意力机制 EMA. EMA 模块通过通道分组与特征重塑建模跨通道依赖, 并结合跨空间学习捕获像素级关系, 同时利用多尺度分支动态增强语义判别性和空间信息表达能力. 具体而言, 本文将 EMA 模块插入 DenseNet121 的前 3 个 Dense block 输出之后, 并在最终分类网络前保持原始结构不变. 该设计可在高层特征阶段增强通道判别与空间结构感知能力, 为后续策略预测提供更具判别性的特征表示.

2.4.2 动作策略网络结构

推动与抓取的 Q 值预测网络共享相同的解码结构, 主要包括 1 个 EMA 注意力层、3 个级联的 1×1 卷积层、ReLU 激活函数及批量归一化层; 每组卷积与归一化后接一个双线性插值上采样操作, 以恢复至输入图像分辨率. 注意力机制放置在输入层, 使得动作策略网络可以强化其相应动作对应的特征, 提高输出 Q 值图像对应点位的精准性.

经 EMA-DenseNet121 层提取的特征通过在通道维度上进行简单拼接后, 输入至抓取和推动的策略网络中, 输出 32 个像素级别的 Q 值映射图 (16 个用于推动操作, 16 个用于抓取操作), 此像素级别的 Q 值映射图表示在相应位置执行推动和抓取动作的预期未来回报. 为了聚焦关键的目标位置, 将各动作像素级别的 Q 值图像与相应动作对应的掩码进行 Hadamard 乘积, 使动作聚焦于有意义的区域. 感知-动作策略网络结构如图 2 所示.

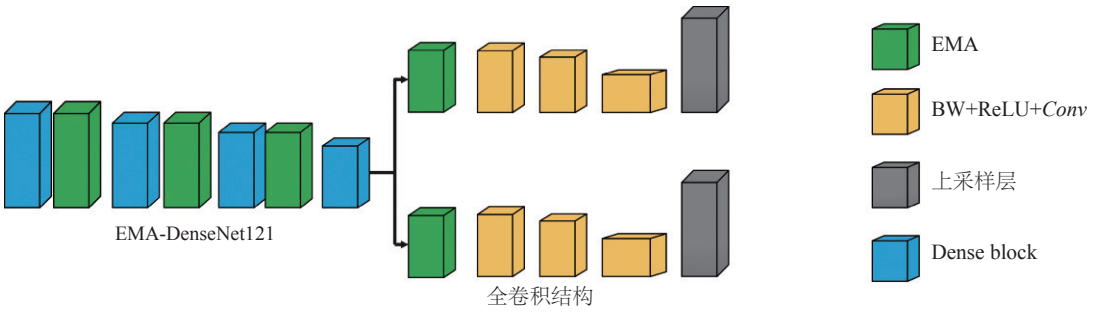


图 2 感知-动作策略网络结构

2.4.3 EMA 注意力机制

注意力机制能够有效提升网络的特征表达能力. 在感知-动作策略网络中, 所采用的注意力模块既要具备通道与空间维度的建模能力, 又需保持较高的计算效率. 传统通道注意力机制 SE (squeeze and excitation network)[21]通过显式建模通道之间的依赖关系, 实现对通道维度特征响应的自适应调整, 但未引入空间信息. CBAM (convolutional block attention module)[22]建立了跨通道和跨空间信息, 并在特征图中建立了空间和通道维度之间的语义相互依赖关系. 因此, CBAM 在跨维注意力权重集成到输入特征方面表现出了巨大的潜力. 然而, 全局池化操作涉及复杂的处理, 这会带来一些计算开销. CA (coordinate attention)[23]注意力机制将精确的位置信息嵌入到通道中, 并在空间上捕获远程交互, 从而实现明显的性能提升, 但它忽略了整个空间位置之间相互作用的重要性. 此外, 1×1 卷积核的有限感受野阻碍了局部跨通道交互建模和上下文信息利用. Triplet attention[24]三重注意力将交叉通道和空间信息与旋转操作混合到 3 个平行分支中, 以学习越来越抽象的特征. 然而, 捕获的注意力权重直接通过简单平均的方式进行聚合, 不利于提高深度特征的可分辨性.

与上述机制相比, EMA 在结构和效率上均具有优势. 其核心思想是通过通道分组与特征重塑, 将输入特征划分为 G 个子组:

$$X=[X_0,X_1,\cdots,X_{G-1}],\ X_i\in R^{C//G\times H\times W},\ i=0,1,\cdots,G-1 \tag{4}$$

其中, C 为输入通道数, H、W 分别为特征图的高和宽.

在分组注意力的设计中, 通常要求分组数 G 明显

小于通道数 C, 以避免每组通道过少而削弱语义建模能力. 在此基础上, 本实验采用 G=32, 既保证了每组通道具有足够的表达能力, 又避免了过多分组带来的额外计算开销.

此分组方式保证了子组内空间语义特征的均匀分布, 有效缓解了全局通道建模中因通道压缩带来的信息丢失. 同时, 分组计算显著降低了复杂度, 使 EMA 在推动与抓取任务中既能保持判别力, 又具备更高的计算效率.

如图 3 所示, EMA 模块包含 3 条并行路径: 两条 1×1 卷积分支和一条 3×3 卷积分支.

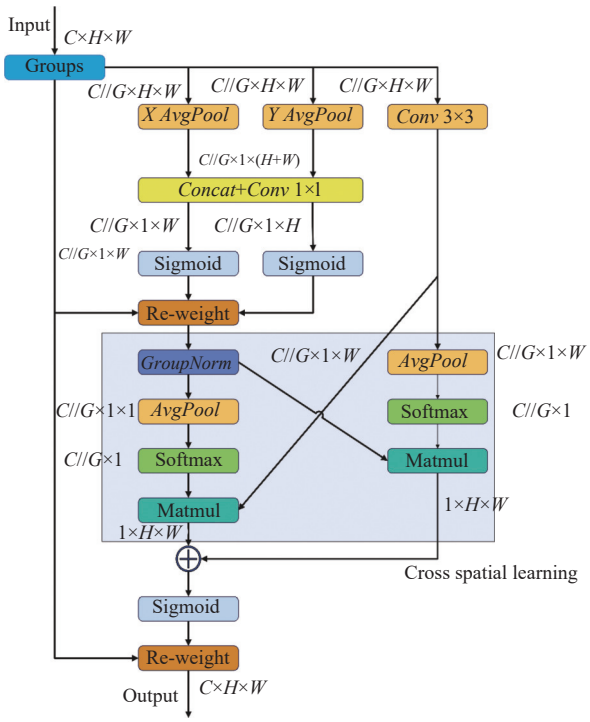


图 3 EMA 网络结构

两条 1×1 卷积分支用于捕获扩展通道关系. 分别进行水平与垂直方向的全局平均池化得到X_{CH}、X_{CW}:

$$X_{\mathrm{CH}} = AvgPool_H(X_C) \tag{5}$$

$$X_{\mathrm{CW}} = AvgPool_W(X_C) \tag{6}$$

其中, X_{CH}、X_{CW}分别表示沿着高度和宽度方向的经过平均池化后的向量.

将池化的结果按高度方向拼接后通过 1×1 卷积整合, 将整合后的输出分解为两个向量后通过 Sigmoid 函数捕获两个空间方向上的通道注意力权重图, 利用点积操作将两个方向上的通道注意力权重聚合, 从而实现组内通道动态强化. 最后通过组归一化输出聚合后的特征.

$$(A_H, A_W) = Split(Conv_{1x1}(Concat[X_{\mathrm{CH}}, X_{\mathrm{CW}}])) \tag{7}$$

$$Y_1 = GroupNorm(X_C \cdot \sigma(A_H) \cdot \sigma(A_W)) \tag{8}$$

其中, $\cdot$为逐元素相乘, σ为 Sigmoid 函数, $GroupNorm$为组归一化操作.

一条 3×3 卷积分支用于引入局部上下文信息和更大感受野, 补充多尺度特征.

$$Y_2 = Conv_{3\times3}(X_C) \tag{9}$$

其中, Y_2为 3×3 卷积分支输出.

因此, EMA 不仅对通道间的信息进行编码以调整不同通道的重要性, 而且将精确的空间结构信息保留到了通道中.

跨空间建模过程中, 1×1 和 3×3 的卷积分支的输出分别进行二维全局平均池化进行全局空间信息编码, 二维全局池化公式如下:

$$Z_1 = \frac{1}{H\times W}\sum_{j=1}^{H}\sum_{i=1}^{W} Y_1(i,j) \tag{10}$$

$$Z_2 = \frac{1}{H\times W}\sum_{j=1}^{H}\sum_{i=1}^{W} Y_2(i,j) \tag{11}$$

其中, H 为输入特征图高度, W 为输入特征图宽度. Z_1为1×1分支的输出, Z_2为3×3分支的输出.

1×1分支的输出Z_1经过 Softmax 拟合线性变换后, 与Y_2进行矩阵点积运算生成第 1 个空间注意力图. 同理, 将Y_1与Z_2进行矩阵点积运算生成第 2 个空间注意力图. 两个空间注意力图聚合后通过 Sigmoid 函数形成空间注意力权重图, 并与输入特征图进行逐像素点乘, 捕捉像素级的成对关系, 并突出显示所有像素的全局上下文.

2.4.4 EMA 在推抓任务中的适应性

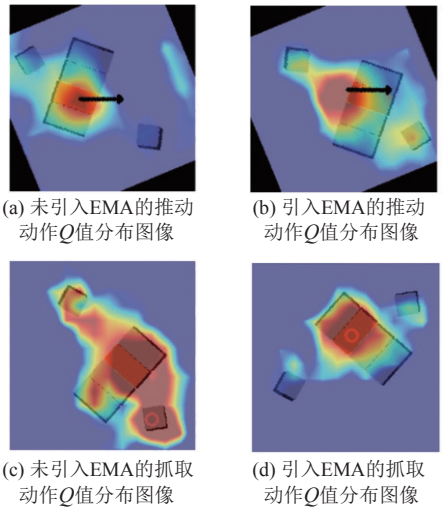
在机械臂感知-动作策略网络中, EMA 模块能够突出与当前任务相关的语义通道, 并抑制冗余信息, 使特征表示更具判别性与任务相关性.

实验中, 我们将引入 EMA 前后的 Q 值分布绘制成热力图 (如图 4 所示). 在此热力图中红色较深的区域代表高 Q 值区域.

推动任务中, EMA 引导的注意力更多集中在物体的边缘与接触可能性较高的位置, 使得策略网络能更

准确地判断推动方向与力的施加点.

抓取任务中, EMA 将注意力分布于物体表面的高置信区域, 有助于提升抓取成功率.



(a) 未引入EMA的推动动作Q值分布图像　(b) 引入EMA的推动动作Q值分布图像

(c) 未引入EMA的抓取动作Q值分布图像　(d) 引入EMA的抓取动作Q值分布图像

图 4　引入 EMA 前后的 Q 值分布

2.4.5　EMA 模块计算复杂度对比分析

本文以 SE 模块作为比较分析, SE 模块在处理较多通道时开销较大, 主要因为其对每个通道都进行了压缩与扩展操作. EMA 模块通过特征分组、方向池化和点乘融合, 有效降低了通道维度上的计算压力, 其计算量主要由特征图的空间尺寸决定, 因此整体更易于控制.

EMA 模块整体计算复杂度为:

$$FLOPs_{\mathrm{EMA}} = \mathrm{O}\left(B\cdot H\cdot W\cdot C^2/G^2\right)+\mathrm{O}(B\cdot H\cdot W\cdot C/G) \tag{12}$$

SE 模块的计算复杂度为:

$$FLOPs_{\mathrm{SE}} = \mathrm{O}\left(B\cdot\frac{C^2}{r}\right)+\mathrm{O}(B\cdot C\cdot H\cdot W) \tag{13}$$

其中, B 为批量维度, C 为通道维度, H 为图像高度, W 为图像宽度, r 为压缩比.

2.5　奖励函数

在强化学习中, 奖励函数决定了策略优化的方向和效率. 奖励函数通过对智能体采取的动作给出即时反馈, 引导其朝着最大化长期累积奖励的目标进行学习. 对于推抓协同任务, 奖励函数能够引导机械臂更快地找到合适的推抓动作序列, 提升推抓效率.

2.5.1　基于频域能量变化的推动奖励函数

推动动作的有效性评估是推动奖励设计中的核心挑战. 一个理想的推动应能实现目标物体与周围障碍物的有效分离, 从而为后续的抓取创造机会. 然而, 推动的效果高度依赖于接触点的位置: 当作用于物体边缘时, 更易产生旋转与平移, 从而引发物体间的分离; 而作用于物体中心的推动则往往导致物体分离效果不明显.

基于上述物理特性, 本文提出从图像频域能量变化的角度来量化推动动作的有效性. 其核心动机在于: 物体间的分离必然伴随着其接触区域边缘结构的断裂或形变, 这一物理过程在图像域中直接表现为高频成分 (即边缘信息) 的显著变化. 因此, 通过监测推动前后目标物体周围区域的高频能量变化, 可以鲁棒地捕捉到由有效推动引发的场景结构改变.

当障碍物与目标物体紧邻时, 其高频能量集中于连接处, 当障碍物推离时, 接触区域的边缘可能断裂或形变, 导致高频能量 (边缘强度) 发生显著变化. 通过计算两幅图像中目标物体周围区域 (包含目标物体) RGB 图像的高频能量变化率, 可检测目标物体周围障碍物是否被推动.

定义两帧图像的高频能量变化率为:

$$\varphi = \frac{E_t - E_{t-1}}{\max(E_{t-1},\varepsilon)} \tag{14}$$

其中, $E_t = \sum_{i,j}[\nabla^2 I_t(i,j)]^2$ 表示第 t 帧图像的高频能量, ∇^2 为拉普拉斯算子, ε 为防止除 0 的小常数.

当 φ 超过预设阈值 ς 时, 认为障碍物与目标物体发生显著运动. 然而, 在实验中发现, 当目标物体在受到推动时, 可能与邻近障碍物发生整体旋转或平移, 此类运动虽然会引起边缘结构的形变, 从而导致高频能量显著变化, 但并未实质性改变物体间的接触关系. 为避免此类误判, 可通过比较高频能量质心相对于目标物体中心的位置变化, 可判定目标物体是否与障碍物体分离, 增强判定的鲁棒性.

定义高频能量质心为:

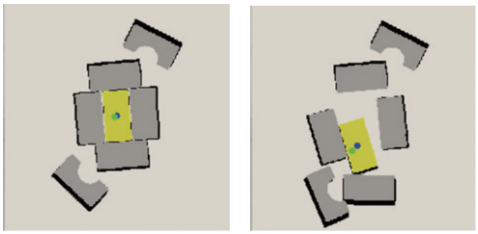
$$C_t = \left(\frac{\sum_{i,j} i\sum_{i,j}[\nabla^2 I_t(i,j)]^2}{E_t}, \frac{\sum_{i,j} j\sum_{i,j}[\nabla^2 I_t(i,j)]^2}{E_t}\right) \tag{15}$$

其中, $I_t(i,j)$ 为图像在位置 (i,j) 的灰度值. 假设目标物体中心通过传统图像处理方法计算为 $O_t=(x_t,y_t)$, 则能量质心与物体中心的相对距离为:

$$d_t = \|C_t - O_t\|_2 \tag{16}$$

若推动后 d_t 显著大于推动前 d_{t-1}, 则进一步确认障碍物已脱离接触区域.

图 5 展示了动作前后频域质心的变化, 其中蓝色为目标物体中心, 绿色为高频能量质心.



(a) 推动前 (b) 推动后

图 5 动作前后频域能量质心变化图

同时, 为了引导推动动作朝向目标物体, 首先根据策略网络输出的最佳推动像素构建推动向量. 推动向量以动作起点为起点, 沿预测的推动方向延伸至终点, 形成一条连续的像素级直线. 随后, 将该向量与目标物体掩码进行几何交集判断, 统计推动路径上与目标掩码重合的像素数量. 当重合像素数量超过设定阈值时, 说明该推动动作能够有效与目标物体发生交互, 从而给予相应的奖励. 因此设计推动奖励函数如下:

$$R_{\text{push}}=\begin{cases}0, & \text{推动无效}\\ 0.25, & \text{推动向量经过目标物体上方}\\ 0.5, & \varphi>\varsigma\text{且}d_t>d_{t-1}\end{cases} \tag{17}$$

2.5.2 抓取奖励函数

当机械臂抓取点位于目标物体上方时, 给予一定奖励. 当机械臂执行抓取动作后, 夹爪指尖距离大于零时, 认为是一次成功的抓取. 基于上述方法, 设计抓取奖励函数为:

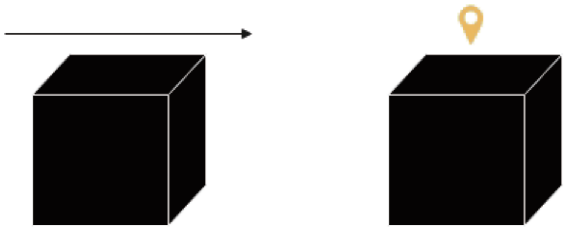
$$R_{\text{grasp}}=\begin{cases}0, & \text{抓取失败}\\ 0.5, & \text{抓取点位于目标物体上方}\\ 1, & \text{抓取成功}\end{cases} \tag{18}$$

推动奖励函数中的推动向量与抓取奖励函数中的抓取点可视化如图 6 所示.

2.6 训练细节

在抓取推动训练过程中, 依据推动与抓取预测值的大小来决定动作, 同时抓取奖励值大于推动奖励值, 会导致推动动作较少, 模型难以学习到有效动作策略. 据此, 受 Xu 等人[10]的启发, 本文提出一种两阶段的训练方法. 在第 1 个阶段 (前 1 000 步), 场景中目标物体周围几乎没有遮挡, 障碍物数量为 3 个, 此时抓取判断条件遵循最大 Q 值原则, 依据抓取和推动的最大 Q 值大小来推断最优动作.

第 2 个阶段, 障碍物数量增至 8 个, 取第 1 阶段成功抓取的 Q 值平均值作为区分目标对象抓取是否处于合适抓取状态的固定阈值Q_{g}, 当预测抓取值超过上一阶段学习到的阈值时, 执行抓取动作, 否则执行推动动作. 此外结合目标物体周围障碍物密度 M[11]一起作为抓取条件进行判断, 若其周围障碍物像素密度小于阈值或抓取值超过固定抓取阈值, 则判定抓取. 该阈值机制旨在动态平衡推抓动作的选择, 通过抑制在目标物体被严重遮挡、抓取成功率低时的无效抓取, 引导智能体优先执行推动动作以改变环境布局, 鼓励模型探索为抓取提供更多空间的有效推动策略.



(a) 推动向量经过目标物体 (b) 抓取点位于目标物体上方

图 6 奖励函数可视化

2.7 损失函数

整体网络通过最小化时间误差 (temporal difference error, TD error) 进行训练:

$$\delta_t=Q^{\theta_t}(S_t,a_t)-y^{\theta_t^-} \tag{19}$$

通过 Huber 损失函数来计算每步损失以更新模型权重.

$$L_\delta=\begin{cases}\dfrac{1}{2}\delta_t^2, & |\delta_t|<1\\ \left|\delta_t-\dfrac{1}{2}\right|, & \text{其他}\end{cases} \tag{20}$$

其中, θ_t为时间 t 的网络参数, 目标参数θ_t^-在迭代中保持恒定. 在时间 t, 只通过执行运动原语的单个像素传递梯度, 而所有其他像素反向传播, 损失为 0.

本实验训练使用梯度下降法进行训练, 具体参数设置如下: 学习率为2×10^{-4}, 未来奖励折扣γ为 0.5, 初始化贪婪因子ε为 0.5, 贪婪值衰减率为 0.998.

3 实验分析

3.1 仿真实验配置

实验使用系统配置为 CPU 采用 Intel i7-11800H, GPU 为 RTX4090 24G, 操作系统选用 Ubuntu 18.04, 并使用 PyTorch 1.6 作为运行框架. 仿真平台选择 Coppe-

liaSim, 该仿真环境搭建配备有 RG2 夹爪的 UR5 机械臂, Intel D435i 的相机, 0.224 m×0.224 m 的工作平台. 仿真环境如图 7 所示. 该仿真环境旨在尽可能精确地模拟现实世界条件, 以确保实验结果的有效性和可迁移性. 所有环境参数均经过精心调整, 以匹配实际情况, 从而增强实验结果的实际使用价值.

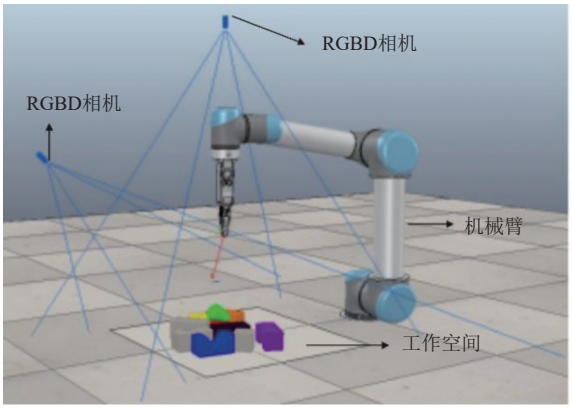


图 7 仿真环境

3.2 评估指标

实验设计包括两类主要评估指标, 用于全面衡量算法在推抓任务中的有效性与执行效率.

抓取成功率 (grasp success, *GS*): 衡量算法完成目标物体抓取的能力.

动作次数 (action efficiency, *AE*): 评估策略执行任务的效率, 以单次任务消耗的动作次数表征.

在训练过程中抓取成功率以间隔步长统计, 反映算法实时学习效果.

$$GS_{\text{train}}(t)=\frac{N_{\text{gs}}(t)}{N_{\text{interval\_steps}}} \tag{21}$$

其中, t 为当前训练步数, $N_{\text{gs}}(t)$表示最近$N_{\text{interval_steps}}$步内抓取动作中抓取成功的次数. 此设计可直观展示算法在长期训练中的稳定性.

测试阶段以轮次为单位进行, 每轮任务在限定的动作次数内完成抓取操作, 若成功抓取目标物体, 则视为该轮测试成功并终止当前轮次. 最终抓取成功率定义为成功轮次占总测试轮次数的比例.

$$GS_{\text{test}}=\frac{N_{\text{ns}}}{N_{\text{nt}}} \tag{22}$$

其中, N_{ns}表示所有测试中成功抓取目标物体的轮次数, N_{nt}表示总的测试轮次数.

动作次数 (*AE*) 为所有测试中所有成功抓取目标物体的轮次中所需的最小动作次数的平均值, 用于评估策略的执行效率.

$$AE=\frac{\sum_{i=1}^{N_{\text{ns}}}N_i}{N_{\text{ns}}} \tag{23}$$

其中, N_{ns}表示所有测试中成功抓取目标物体的轮次数, N_i 表示第 i次成功测试中实际消耗的动作次数.

3.3 训练结果与分析

为验证所提方法的性能, 选取 METOVPG (mask-enhanced target-oriented VPG), GII (grasp in the invisible)[11]进行对比实验. 同时为确保比较的公平性, 本研究中算法所涉及的模拟环境与 METOVPG 和 GII 保持一致. 仿真场景中加入目标物块和干扰物块. 通过语义分割模型来进行目标物体选取, 一旦当前空间中的目标物体抓取完成后, 仿真环境自动重置, 开始下一轮的任务循环.

METOVPG (mask enhanced target oriented VPG): 加入目标掩码增强的目标导向 VPG[7]. 原始 VPG 算法中仅通过 RGB 图与深度图来实现无目标的任意物体抓取, 且环境复杂度后期未提升, 本文通过引入目标掩码使得 VPG 实现目标导向的抓取操作, 且引入分阶段训练, 在第 2 阶段中提升环境复杂度.

GII (grasping in the invisible): 该方法利用深度图像、RGB 图像与目标物体掩码图像进行目标物体抓取, 其训练过程设计为分阶段训练. 在第 2 阶段环境复杂度上升后, 通过多层感知机 (MLP) 神经网络作为动作分类器进行动作选择.

以上对比方法所使用的基础网络框架相同. 图 8 展示了不同方法在训练过程中抓取目标物体的成功率变化趋势, 其中抓取成功定义为目标物体被完整提起. 抓取成功率以每 200 步为统计间隔进行评估. 从整体趋势来看, 本文方法在两个训练阶段中均展现出优越的性能.

从训练曲线图 (图 8) 来看, METOVPG 方法的训练表现最差, 抓取成功率始终在 40%–60% 之间. 其原因在于, 该方法的推动行为的策略倾向于为改变环境而推动, 而非为后续抓取创造有利条件. 这导致了大量未能有效改变目标物体与障碍物之间遮挡关系的无效推动动作.

GII 方法通过引入一个二元分类器进行动作选择, 并设计了基于目标物体周围障碍物密度变化的推动评

估机制, 对策略进行了协同优化, 使其抓取成功率提升至 60%–70% 的区间.

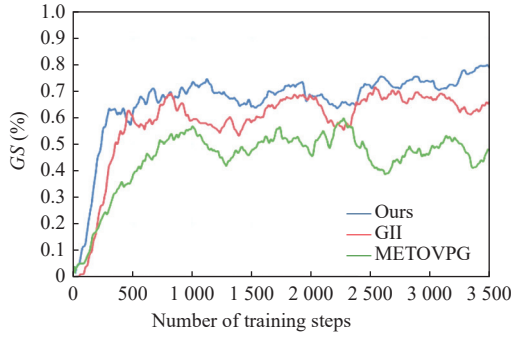


图 8 目标物体抓取训练表现

从图 8 中训练曲线可知, 本文方法的抓取成功率在整个训练过程中均优于上述两种方法. 模型在约 3 300 步左右时实现了约 80% 的抓取成功率. 在第 1 阶段, 得益于 EMA 注意力机制对关键特征的增强提取能力, 模型能快速学习到精准的抓取位姿, 表现为成功率上升较快. 进入第 2 阶段后, 动作选择由上一阶段计算得到的抓取阈值以及目标周围障碍物面积来判断. 在面对显著提升的环境复杂度, 基于频域能量的推动评估机制开始发挥关键作用, 它能有效判别出有助于物体分离的推动行为, 同时注意力机制强化模型对于相应动作的关键特征的提取. 注意力机制与推动评估机制的协同工作, 共同驱动本文方法学习到了以成功抓取为最终目标的、高效的推抓协同策略.

为深入探究性能优势的来源, 本文从模型训练更新机制角度对两模块协同机制进一步分析, 模型通过最小化时间差分误差 (TD error) 并结合 Huber 损失进行稳定训练, 由于梯度仅在实际执行的动作点位上传递, 网络的更新重点集中在与动作相关的区域. 推动有效性评估机制在此过程中捕捉有效推动动作, 提升该动作点位的奖励值, 从而加大这些位置在 TD 目标中的权重, 放大其对参数更新的贡献; 与此同时, 注意力机制则在特征提取和反向传播阶段引导模型突出这些高奖励对应的特征区域. 二者协同作用, 使得模型在迭代过程中不断强化对关键特征和有效推动行为的敏感性, 从而实现更高效的策略学习.

为了验证本文所提方法的有效性, 并进一步从各模块作用的独立性角度进行分析, 本文设计了消融实验: 通过单个模块作用效果的训练曲线来评估其对整体性能的影响. 图 9 展示了各功能模块在单独引入时的消融实验训练表现.

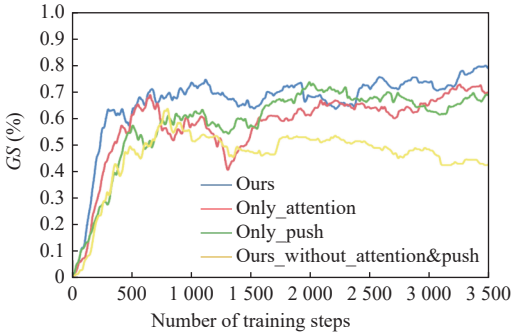


图 9 消融实验

(1) Only_attention: 仅引入注意力机制与抓取奖励.

(2) Only_push: 仅引入推动奖励与抓取奖励.

(3) Ours: 整合所有方法.

(4) Ours_without_attention&push: 仅抓取奖励.

从消融实验训练曲线图 (图 9) 可以看出, Only_push 在训练表现中的抓取成功率曲线上在环境复杂度上升后一定步数内优于 Only_attention, 说明推动奖励可以使模型学习到有效动作, 从而改变物体布局, 使目标物体更容易抓取. 在模型训练后期 Only_attention 高于 Only_push, 这是因为 EMA 注意力机制通过增强对物体表面、几何轮廓等关键特征的提取能力, 显著提升了抓取成功率. 同时 Only_push 的推动行为可能使物体处于比较好的抓取环境中, 但抓取并未成功. Only_attention 在训练表现中的抓取成功率显著高于 Ours_without_attention&push, 说明注意力机制能提高策略的判别能力, 使模型在抓取任务中更关注关键特征. Ours_without_attention&push 由于没有其他模块作用, 模型较难学习到更有效的策略导致模型抓取成功率较低.

通过以上消融实验的训练表现, 验证了本文所提方法对模型抓取性能提升的有效性.

图 10 对比了不同 EMA 分组数量 (Ours_EMA16 和 Ours_EMA32) 的训练表现. 从图中的训练过程曲线可以看出, 不同分组数量对应的训练表现有所差异, 其中分组数量为 32 时, 其训练过程表现整体优于分组数量为 16 时的训练表现, 并且其训练过程中最高达到了 80% 的抓取成功率, 而分组数量为 16 时, 训练过程中最高达到了约 76% 的抓取成功率. 其可能原因为: 分组数过小会导致每组特征范围过宽, 难以突出局部差异; 而适度增加分组数后, 特征被划分得更细, 网络能够更有针对性地捕捉到关键信息, 同时在计算开销上仍保持合理, 从而获得更优的性能表现.

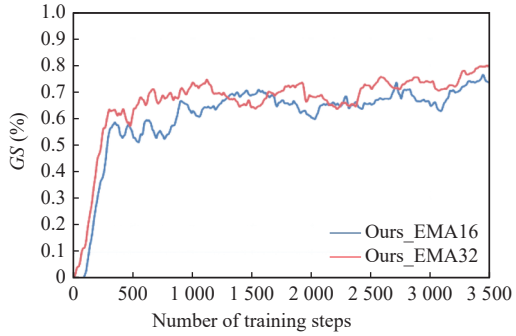


图 10 不同分组数量的训练表现

3.4 仿真环境结构化场景测试

为验证所提方法的泛化性能, 手动设计了 8 种不同的挑战性测试案例. 对所提方法进行仿真测试. 放置方式如图 11 所示. 每个测试场景均进行 30 组独立实验, 单次实验中设定最大动作次数为 5 次. 实验结果如表 1 所示.

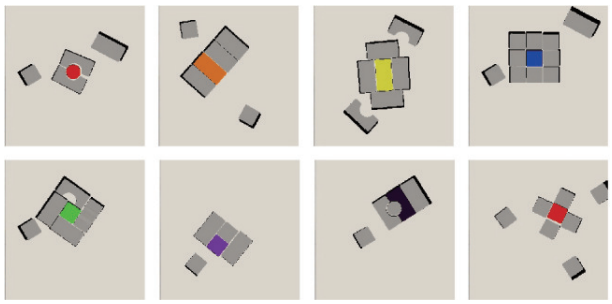


图 11 仿真环境挑战性排列

表 1 仿真环境结构化场景测试结果

测试场景	测试指标	METOVPG	GII	Ours	Ours without push	Ours without attention
场景1	*GS* (%)	56.7	93.3	93.3	90	90
	AE	3.59	3.82	3.33	3.93	3.33
场景2	*GS* (%)	76.7	86.7	100	93.3	96.7
	AE	3.57	3.42	2.4	3.39	2.59
场景3	*GS* (%)	80	83.3	90	80	86.7
	AE	3.83	3.84	3.93	4.79	4.15
场景4	*GS* (%)	83.3	93.3	93.3	86.7	90
	AE	4.12	2.75	3.39	4.15	3.56
场景5	*GS* (%)	36.7	86.7	80	76.7	80
	AE	4.73	4.58	3.63	4.00	3.83
场景6	*GS* (%)	70	76.7	83.3	80	83.3
	AE	3.76	4.22	4.04	4.58	4.16
场景7	*GS* (%)	73.3	93.3	93.3	90	93.3
	AE	4.14	2.64	2.42	3.33	2.53
场景8	*GS* (%)	83.3	86.7	96.7	90	93.3
	AE	3.8	4.00	3.34	3.74	3.82
所有场景	*GS* (%)	70	87.5	91.2	85.84	89.17
	AE	3.94	3.66	3.31	3.99	3.49

将本文设计的方法在结构化场景中进行测试, 其中 Ours without push 代表去除推动奖励, Ours without attention 代表去除注意力机制. 测试结果表明, 本文方法 Ours 测试成功率与动作效率均优于 METOVPG, 其中测试成功率高于基线方法 21.2%. 同时, 引入推动奖励对于抓取成功率的影响较大, 注意力机制与推动奖励二者同时作用进一步提升了抓取成功率与动作效率. 进一步说明推动奖励可以使模型学习到有效的推动动作, 同时注意力机制使模型更有效地关注到动作对应关键特征, 进而提升抓取成功率与动作效率.

4 结束语

本文提出了一种基于深度强化学习的推抓协同控制方法, 实现了机械臂在复杂堆叠场景下对目标物体的高效抓取. 该方法引入基于图像频域能量变化的推动有效性评估机制, 以提升模型在训练过程中识别有效推动动作的能力, 从而准确判断推动动作是否成功实现了目标物体与障碍物的分离. 在感知-动作策略网络方面, 本文采用 EMA 注意力机制对原有网络结构进行优化, 提高了特征提取与利用效率, 从而提升整体预测精度.

本文方法在抓取成功率方面均优于现有基线方法, 为复杂环境下的机械臂操作任务提供了可靠的解决思路. 由于本实验的抓取动作与推动动作是通过固定抓取与推动位姿进行动作训练, 未能很好地利用 6 自由度的优势, 因此在后续将会考虑进行 6 自由度的动作训练并完成在真实环境高精度实验平台上的测试.

参考文献

1 Guo N, Zhang BH, Zhou J, *et al*. Pose estimation and adaptable grasp configuration with point cloud registration and geometry understanding for fruit grasp planning. Computers and Electronics in Agriculture, 2020, 179: 105818. [doi: 10.1016/j.compag.2020.105818]

2 Wu YW, Li WZ, Liu ZY, *et al*. Autonomous learning-free grasping and robot-to-robot handover of unknown objects. Autonomous Robots, 2025, 49(3): 18. [doi: 10.1007/s10514-025-10201-y]

3 Ye XH, Qin XY, Zhan LM, *et al*. Research on a fusion technique of YOLOv8-URE-based 2D vision and point cloud for robotic grasping in stacked scenarios. Applied Sciences, 2025, 15(12): 6583. [doi: 10.3390/app15126583]

4 Sun RH, Wu CD, Zhao X, *et al*. Object recognition and grasping for collaborative robots based on vision. Sensors, 2023, 24(1): 195. [doi: 10.3390/s24010195]

5 Nguyen T, Vu MN, Huang B, *et al*. Language-driven 6-DoF grasp detection using negative prompt guidance. Proceedings of the 18th European Conference on Computer Vision. Milan: Springer, 2024. 363–381.

6 Bai JM, Cao GH. G-RCenterNet: Reinforced CenterNet for robotic arm grasp detection. Sensors, 2024, 24(24): 8141. [doi: 10.3390/s24248141]

7 Zeng A, Song SR, Welker S, *et al*. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning. Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Madrid: IEEE, 2018. 4238–4245.

8 Yu S, Zhai DH, Xia YQ, *et al*. An efficient robotic pushing and grasping method in cluttered scene. IEEE Transactions on Cybernetics, 2024, 54(9): 4889–4902. [doi: 10.1109/TCYB.2024.3381639]

9 Phan LAD, Phan DQ, Bui TT, *et al*. Approaching collaborative manipulation by pushing-grasping fusion. IEEE Access, 2025, 13: 97693–97707. [doi: 10.1109/ACCESS.2025.3574018]

10 Xu KC, Yu HX, Lai QE, *et al*. Efficient learning of goal-oriented push-grasping synergy in clutter. IEEE Robotics and Automation Letters, 2021, 6(4): 6337–6344. [doi: 10.1109/LRA.2021.3092640]

11 Yang Y, Liang HY, Choi C. A deep learning approach to grasping the invisible. IEEE Robotics and Automation Letters, 2020, 5(2): 2232–2239. [doi: 10.1109/LRA.2020.2970622]

12 Sarantopoulos I, Kiatos M, Doulgeri Z, *et al*. Split deep Q-learning for robust object singulation. Proceedings of the 2020 IEEE International Conference on Robotics and Automation. Paris: IEEE, 2020. 6225–6231.

13 Yu S, Zhai DH, Xia YQ. A novel robotic pushing and grasping method based on vision Transformer and convolution. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(8): 10832–10845. [doi: 10.1109/TNNLS.2023.3244186]

14 Yang YX, Ni ZH, Gao MY, *et al*. Collaborative pushing and grasping of tightly stacked objects via deep reinforcement learning. IEEE/CAA Journal of Automatica Sinica, 2022, 9(1): 135–145. [doi: 10.1109/JAS.2021.1004255]

15 Gao J, Li YF, Chen YM, *et al*. An improved SAC-based deep reinforcement learning framework for collaborative pushing and grasping in underwater environments. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 2512814.

16 左国玉, 赵敏, 黄高, 等. 基于坐标注意力的杂乱环境中机器人推抓协同学习. 北京工业大学学报, 2024, 50(6): 674–682. [doi: 10.11936/bjutxb2022090024]

17 Ouyang DL, He S, Zhang GZ, *et al*. Efficient multi-scale attention module with cross-spatial learning. Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes Island: IEEE, 2023. 1–5.

18 Jaakkola T, Singh SP, Jordan MI, *et al*. Reinforcement learning algorithm for partially observable Markov decision problems. Proceedings of the 8th International Conference on Neural Information Processing Systems. Denver: MIT Press, 1994. 345–352.

19 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. 计算机工程与应用, 2024, 60(23): 325–332. [doi: 10.3778/j.issn.1002-8331.2307-0326]

20 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.

21 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.

22 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

23 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 13708–13717.

24 Misra D, Nalamada T, Arasanipalai AU, *et al*. Rotate to attend: Convolutional triplet attention module. Proceedings of the 2021 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3138–3147.

(校对责编: 李慧鑫)