



图学学报
Journal of Graphics
ISSN 2095-302X, CN 10-1034/T

《图学学报》网络首发论文

题目: 基于深度强化学习的虚拟手自适应抓取研究
作者: 伍一鹤, 张振宁, 仇栋, 李蔚清, 苏智勇
收稿日期: 2020-09-30
网络首发日期: 2021-01-04
引用格式: 伍一鹤, 张振宁, 仇栋, 李蔚清, 苏智勇. 基于深度强化学习的虚拟手自适应抓取研究. 图学学报.
<https://kns.cnki.net/kcms/detail/10.1034.t.20210104.1450.002.html>



网络首发:在编辑部工作流程中,稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定,且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件,可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定;学术研究成果具有创新性、科学性和先进性,符合编辑部对刊文的录用要求,不存在学术不端行为及其他侵权行为;稿件内容应基本符合国家有关书刊编辑、出版的技术标准,正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性,录用定稿一经发布,不得修改论文题目、作者、机构名称和学术内容,只可基于编辑规范进行少量文字的修改。

出版确认:纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约,在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版,以单篇或整期出版形式,在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z),所以签约期刊的网络版上网络首发论文视为正式出版。

基于深度强化学习的虚拟手自适应抓取研究

伍一鹤¹, 张振宁², 仇 栋¹, 李蔚清², 苏智勇¹

(1. 南京理工大学自动化学院, 江苏 南京 210094;

2. 南京理工大学计算机科学与工程学院, 江苏 南京 210094)

摘 要: 在计算机角色动画的抓取研究中, 生成动作序列的自然性、稳定性及自适应能力三者难以同时得到保证, 即自然又稳定的抓取控制器往往泛化能力有限, 无法适用于其他类型、尺寸物体的抓取任务。通过引入和抓取类型相对应的手部示教数据、设计回报函数, 构建了一种基于深度强化学习的虚拟手自适应抓取控制器。实验结果表明, 所设计的控制器能够生成兼具自然性和稳定性的抓取运动序列, 同时对素材库中不同尺寸、不同类型的基元物体也具备较好的自适应能力。

关 键 词: 深度强化学习; 示教学习; 运动生成; 虚拟手; 动作捕捉数据

中图分类号: TP 391

DOI:

文献标识码: A

文 章 编 号: 2095-302X(0000)00-0000-00

Research on adaptive grasping of virtual hands based on deep reinforcement learning

WU Yi-he¹, ZHANG Zhen-ning², QIU Dong¹, LI Wei-qing², SU Zhi-yong¹

(1. School of Automation, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, China;

2. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, China)

Abstract: In the grasping research of computer character animation, it is difficult to guarantee the naturalness, stability and adaptive ability of the generated action sequence at the same time. That is, the natural and stable grasping controller often has limited generalization ability and cannot be applied to other types of grabbing tasks. We construct a virtual hand adaptive grasping controller based on deep reinforcement learning by introducing hand teaching data corresponding to the grasping type and designing the reward function. Experimental results show that the designed controller can generate a grasping motion sequence with both naturalness and stability, and also has a good adaptive ability to different sizes and different types of primitive objects in the material library.

Keywords: deep reinforcement learning; demonstration learning; motion generation; virtual hands; mocap data

收稿日期: 2020-09-30; 定稿日期: 2020-12-24

Received: 30 September, 2020; Finalized: 24 December, 2020

基金项目: “十三五”装备预研项目(61409230104, 1017, 315100104); 中央高校基本科研业务费专项(30918012203); 上海航天科技创新基金(SAST2019009)

Foundation items: Thirteenth Five-Year Plan Equipment Pre-research Project (61409230104, 1017, 315100104); Fundamental Research Fund for Central Universities (30918012203); Shanghai Aerospace Science and Technology Innovation Fund (SAST2019009)

第一作者: 伍一鹤(1996-), 男, 新疆博州人, 硕士研究生。主要研究方向为计算机图形学、增强现实。E-mail: yihewu@njust.edu.cn

First author: WU Yi-he (1996-), male, master student. His main research interests cover computer graphics, augmented reality. E-mail: yihewu@njust.edu.cn

通信作者: 苏智勇(1981-), 男, 江苏泰州人, 副教授, 博士。主要研究方向为计算机图形学、机器学习等。E-mail: su@njust.edu.cn

Corresponding author: SU Zhi-yong (1981-), male, associate professor, Ph.D. His main research interests cover computer graphics, machine learning, etc. E-mail: su@njust.edu.cn

虚拟人手部的抓取运动生成技术广泛应用在计算机角色动画、机器人学、第一人称虚拟现实等领域。由于人手是一个多自由度的复杂层次结构^[1]、是人体最灵活的部位,为了实现虚拟手对不同形状、不同尺寸物体的自适应抓握交互,并保证生成的运动序列自然流畅,保证最终的抓取姿态稳定牢固,对控制器的设计提出了较高的要求。

现有的研究方法主要分为两类:建模解析法和数据驱动法。建模解析法^[2-7]需要对虚拟手及交互对象进行严格的逆向工程,获得描述抓取姿态的解析方程^[4]。该方法的建模过程繁琐,且难以泛化^[5-7]。数据驱动法^[8-13]需要动作捕捉数据^[13]或大量如contact map^[8]、抓握标签^[9]等形式的先验数据,再结合机器学习等方法训练抓取模型。目前数据驱动的方法泛化能力有限,仅对单一形状物体能够取得较自然、稳定的抓取结果^[8-13],此外其中的部分方法依赖预训练,无法直接作用于目标物体^{[8][9]}。

为了解决计算机角色动画中虚拟手对不同形状、尺寸物体难以生成自然、稳定抓取运动的问题,本文利用深度强化学习方法,设计了一种虚拟手抓取控制器。通过手部示教序列的引入,我们的方法使得控制器能够参考真人数据,生成自然、拟人的抓取运动,简化了复杂的运动机理建模;通过模仿奖励和自适应罚项的设置,该控制器能够在模仿手部参考序列的同时,对不同形状、不同尺寸物体生成稳定的自适应抓取,并且对于不同的拓扑结构的手部,具备一定的重定向能力。

1 相关工作

1.1 虚拟手的抓取研究

按照实现方法,现有工作^[2-13]的方法可以分为建模解析法和数据驱动法两大类。

建模解析法是指通过对虚拟手的几何结构及物理特性进行分析,利用动力学或运动学方法对抓取进行建模优化的方法。Ferrari等^[2]最早提出力闭合概念,即一种外力可以被抓取的力平衡,并保持在手中的抓取状态。Miller等^[3]提出了度量抓取质量的 e 指标和 v 指标。在此工作的基础上,Miller等^[4]发布了开源机器人抓握模拟器GraspIt。Ciocarlie等^[5]度量手部预置点与物体距离,以此生成抓取姿态。蔡嫚娟等^[6]通过手势-特征映射的方法,利用三维鼠标,实现对不同物体的抓取操控。Peng等^[7]结合用户的交互,对抓取对象进行自定义的抓握。

数据驱动法是指以大量某种形式的抓取数据

作为样本,来获得抓取生成模型。该模型具备一定的泛化能力,即对于陌生的抓取测试样本,该模型同样能够生成有效的抓取姿态。Brahmbhatt等^[8]利用contact map作为输入,获得了既稳定又符合该物体功能性的抓握姿态。Tian等^[9]利用支持向量机与粒子群算法,优化了第一人称VR中对物体的抓取操作体验。Liu等^[10]将不同物体所对应的手部抓取位姿作为标签数据集,并设定映射损失函数和碰撞损失函数,最终利用深度神经网络实现了虚拟手对物体的抓取。Starke等^[11]给定抓取的动作捕捉数据,以此来训练自编码器,生成新的抓取姿态。在示教学习方面,Kopicki等^[12]利用少样本迁移学习方法,由深度相机记录的手部运动轨迹示范来学习抓取。王晓媛等^[13]利用基于超二次曲面模型的方法,将复杂物体分割成易于操作的物体元型。再利用模拟退火算法,结合参考示教数据,求得最优的抓取姿态。

总之,目前建模解析法的工作^[2-7]需要复杂的物理建模,数据驱动法的工作^[8-13]需要大量的先验数据。此外,这两种方法获得的虚拟手控制器的抓取自适应能力都较差。

1.2 基于强化学习的虚拟人运动生成研究

随着计算机硬件技术的发展以及深度学习算法的突破,学界逐渐出现了利用深度强化学习模型控制虚拟人的研究工作。Peng等^[15]利用PPO算法获得控制器,结合动作捕捉数据,实现了虚拟人的复杂运动控制,且对于指定的任务具有一定的自适应能力。在此基础上,Peng等^[16]利用基于视频姿态估计的方法,从RGB视频中得到参考运动序列,节省了动作捕捉数据的获得成本。Liu等利用DQN方法实现了虚拟人复杂的运动控制^[17],甚至篮球的运球控制^[18]等。Lee等利用DQN方法,结合视频恢复的参考运动序列,生成了花样滑冰运动员的运动控制器^[19]。基于人类足部的建模, Lee等利用PPO算法获得了足部控制器,该控制器可以自动调节虚拟人足部与地面之间相互作用,重现在运动捕捉数据中无法获得的足部姿势^[20]。方伟等^[21]基于Spark实现了虚拟人强化学习训练框架。Aravind等^[22]利用行为克隆算法,让虚拟手可以完成预设的任务。

现有的虚拟人强化学习工作^[15-21]虽然能够取得较好的自适应结果,但大多忽略了手部的运动控制;少量的虚拟手强化学习工作^[22]也仅是利用了任务驱动型的方法,即注重预设任务完成与否,而忽略了中间过程动画生成的自然性与稳定性。

2 系统概述

虚拟手自适应抓取系统的流程如图1所示，可划分为预处理阶段、训练阶段和推演阶段。在预处理阶段，我们先对虚拟手进行物理建模，然后将动

作捕捉数据绑定在用于示范的虚拟手上。在训练阶段，我们针对抓取任务类型，引入对应的训练物体，再进行深度强化学习的离线训练。在推演阶段，我们将推演物体作为输入，再利用训练阶段获得的强化学习模型，进行在线实时的抓取运动生成。

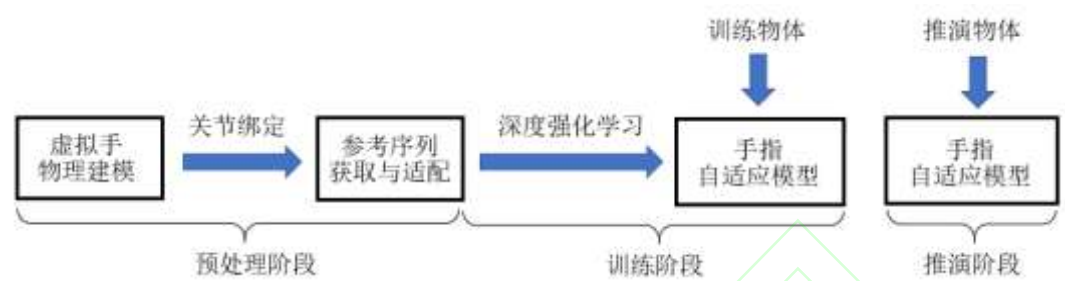


图1 系统流程图

Fig. 1 Pipeline of the system

3 虚拟手物理建模及数据绑定

作为强化学习智能体的虚拟手（白色）需要与物体进行交互，应严格遵循物理定律，即应包含关节链（Hinge Joint）属性、质量（Mass）属性、碰撞体（Collider）属性。通过对真实人手的解剖结构分析^[1]，我们以右手为例，定义了满足生理约束的虚拟手层次模型：每根手指的远指指段（指尖）连接并依附于中指指段；中指指段又依附于近指指段；手指根节点最终依附于手掌，如图2(a)所示。

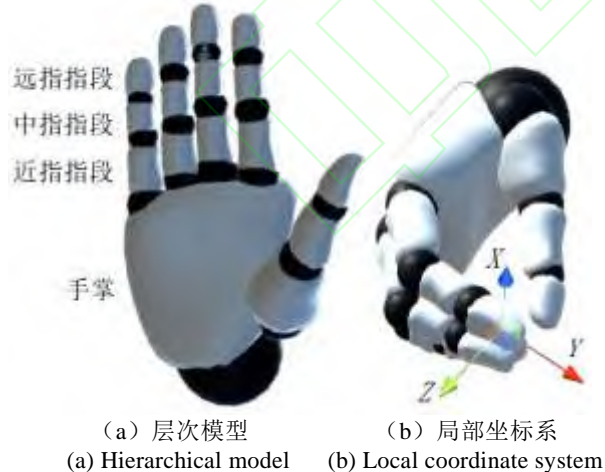


图2 虚拟手模型

Fig. 2 Virtual hand model

我们将指段-关节所组成的结构抽象为类杆-铰链结构，其对应的关节角度约束如表1所示：

表1 手部关节的角度约束

Table 1 Angle constraints of hand joints

关节类型	X旋转(°)	Y旋转(°)	Z旋转(°)
四指近关节	-5~90	locked	-20~20
四指中关节	0~90	locked	locked
四指远关节	0~45	locked	locked
拇指近关节	-60~30	locked	-50~50
拇指中关节	0~90	locked	locked
拇指远关节	0~90	locked	locked

表中 X 、 Y 、 Z 分别是沿手指关节轴方向、指段方向、以及二者平面的垂直向量方向，如图 2(b)所示。为了感知交互及避免穿透现象，我们给虚拟手的每个指尖部位施加了球形碰撞体。

本文中我们选用Mixamo角色公开数据集中的“Picking Up Object”动画^[23]，并将其定义为“横握”抓取运动的参考序列模板，作用于素材库中所有的目标类型。将其绑定在用于记录示教数据的虚拟手（红色）上的结果如图3所示。



图3 参考运动序列的绑定

Fig. 3 Binding of reference motion

该示教数据具体表示为由 t_l 至 t_n 时间内、第 i_l 至 i_m 个手指关节关于 X 、 Y 、 Z 轴的欧拉角 r_X 、 r_Y 、 r_Z 所组成的 $n \times m \times 3$ 的张量，其中 n 与 m 分别表示当前时刻帧数与手指关节的总数。

4 基于深度强化学习的手指自适应抓取运动生成

4.1 任务定义

在该强化学习任务中，我们定义智能体为第3节所述的虚拟手模型，环境中的交互对象为目标物体。我们可以用分类法，将生活中物体的主体部分抽象为特定的基础物体。本文中我们以球体、立方体、柱状体为例，加入目标物体素材库，部分抽象结果如图4所示。在推演阶段，我们定义任务为对物体素材库中的形状特定、尺寸任意的物体进行抓握，且要求生成的抓握运动序列自然而稳定。

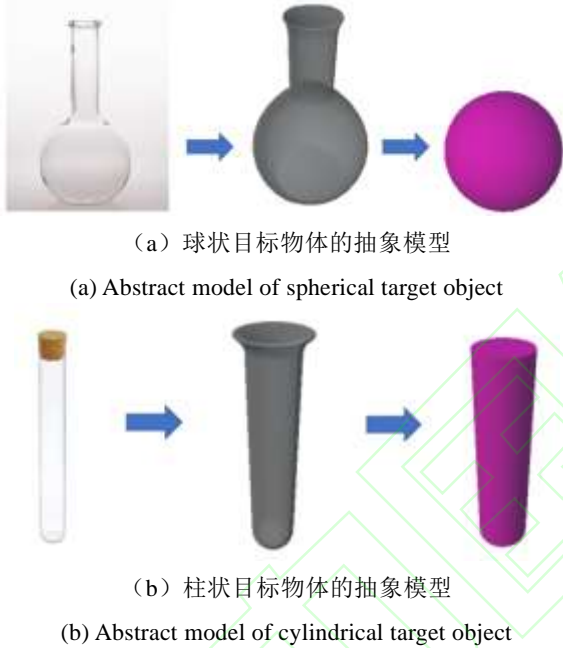


图4 素材库中的基础物体

Fig. 4 Basic objects in the material library

智能体模仿参考运动序列的一个完整流程称之为的一幕 (Episode)。在每一幕的每一时刻 t ，智能体（白色虚拟手）通过观察参考运动序列（红色虚拟手）和目标物体，更新状态信息。后根据训练阶段习得的策略调整最合适的手部姿态与物体进行交互，获得反馈。重复观察-交互-反馈的步骤，直至这一幕结束，即生成了一个完整的抓取序列。

4.2 网络结构

在本文中，我们采用PPO (Proximal Policy Optimization, PPO) 算法^[24]将参数化的策略进行迭代训练。我们使用了一个结构如图5所示的4层神经网络作为策略参数的逼近器。该网络的输入为长度为199的向量，表示智能体的观测值；输出为一个长度为20的向量，表示智能体的行为值。

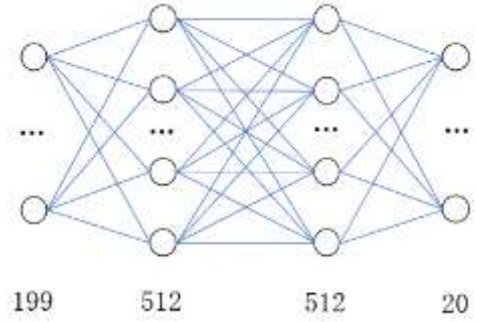


图5 神经网络结构

Fig. 5 Neural network structure

4.3 观测值

在强化学习方法中，观测值指智能体所感知的环境信息。在图5所示神经网络中，输入端表示的观测值组具体表示为：

$$s_t = \{\phi_t, \lambda, d_t^p, d_t^c, p_t, r_t, v_t, \omega_t\} \quad (1)$$

其中 ϕ_t 为 t 时刻归一化后的相位值，并定义 ϕ_0 为此运动序列的初始相位， $\phi_{t_{\max}}$ 为此运动序列的最终相位； λ 是智能体所观测到的目标物体的尺寸放大倍数； d_t^p 是 t 时刻指尖与目标物体距离的欧式距离和，即：

$$d_t^p = \sum_i \|\Delta p_i\|_t \quad (2)$$

d_t^c 是 t 时刻五指质心与物体质心的欧式距离偏差，即：

$$d_t^c = \|\Delta c\|_t \quad (3)$$

p_t 、 r_t 、 v_t 、 ω_t 四组观测值分别是 t 时刻手部15个关节的坐标、角度、速度、角速度矩阵。

4.4 行为值

行为值是智能体根据观测值，结合训练阶段所习得的策略，做出的行为的量化值。图5中神经网络输出的行为值用 v_t 表示，指 t 时刻15个指关节绕X轴或Z轴的速度矩阵。

4.5 回报函数

回报函数是强化学习策略在训练过程中需要最大化的指标函数。本文在DeepMimic^[15]的基础上，设计了适用于手部自适应抓取的回报函数：

$$R_t = w^I R_t^I + w^G R_t^G \quad (4)$$

其中我们分别定义 R_t^I 和 R_t^G 为模仿奖励和自适应罚项， w^I 和 w^G 分别为两者所对应的权值。该回报函数意味着智能体的策略需要在模仿任务和自适应任务中做权衡，最终生成自然且稳定的抓取序列。

受文献[25]的启发，为了缓解初始手型罚项过大的问题，我们定义 w^G 为如下形式：

$$w^G = \begin{cases} \frac{\phi_t}{\phi_{trans}}, & 0 \leq \phi_t \leq \phi_{trans} \\ 1, & \phi_{trans} \leq \phi_t \leq 1 \end{cases} \quad (5)$$

其中 ϕ_{trans} 指过渡相位值。

4.5.1 模仿奖励

模仿奖励是通过逆向强化学习^[26]的思想，使得智能体能够模仿手部参考运动序列的奖励函数。相较于简单地取完成任务，该奖励的引入使得生成的手部的抓取序列更接近于真人示范，更具自然性，有：

$$R_t^I = w^p r_t^p + w^v r_t^v + w^e r_t^e \quad (6)$$

其中 r_t^p 、 r_t^v 分别表示虚拟手的角度奖励和速度奖励，有：

$$r_t^p = \frac{\max(\|\Delta p\|) - \sum_i \|\Delta p\|_i}{\max(\|\Delta p\|)} \quad (7)$$

$$r_t^v = \frac{\max(\|\Delta v\|) - \sum_i \|\Delta v\|_i}{\max(\|\Delta v\|)} \quad (8)$$

上两式意味着当且仅当手部各个关节 i 与参考运动序列中各个关节 i 在每个时刻 t 角度偏差 Δp 较小、速度矢量偏差 Δv 较小时，两奖励函数才能分别取得较高的回报值。

r_t^e 表示虚拟手的末端坐标奖励。由于手部是一个具父子关系的层级结构，从手掌根节点起越接近末端，执行器的坐标越容易出现偏差。我们定义：

$$r_t^e = \frac{\max(\|\Delta p\|) - \sum_i \|\Delta p\|_i}{\max(\|\Delta p\|)} \quad (9)$$

其中 ρ_t 指 t 时刻手指末端5个远指指段的坐标矩阵，满足 $\rho_t \in \mathbf{p}_t$ 。末端坐标奖励使得在每个时刻 t 内，手部5个指尖与参考运动序列指尖在相对坐标系内位置尽可能接近。

4.5.2 自适应罚项

自适应罚项是为了在同一抓取任务下，满足对不同形状、不同尺寸的目标物体都能做到稳定抓取的罚函数。需要指出的是，若只采取4.5.1节中的模仿奖励，即单纯地模仿参考运动序列，最终生成的抓取姿态并不能满足上述自适应能力，具体为：

$$R_t^G = w^p r_t^p + w^c r_t^c + w^a r_t^a \quad (10)$$

r_t^p 是在 t 时刻虚拟手的指尖距离罚项^[5]。特别地，我们仅计算手指末端预置点与物体距离。该罚项的存在使得虚拟手指尖会尽量接触目标物体，而不是完全地模仿参考运动序列。对于虚拟手每个手指的指尖 i ，有：

$$r_t^p = \begin{cases} -\frac{\|\Delta p\|_i}{\max(\|\Delta p\|)}, & \text{指尖与物体未接触时} \\ 0, & \text{指尖与物体接触时} \end{cases} \quad (11)$$

r_t^a 是在 t 时刻虚拟手指尖的悬空罚项。为了进一步保障指尖与物体表面的接触，避免不必要的悬空现象，对于每个手指的指尖，我们取：

$$r_t^a = \begin{cases} 0, & \phi_t \leq 0.6 \text{ 时} \\ -0.2, & 0.6 \leq \phi_t \text{ 且指尖仍未接触物体时} \end{cases} \quad (12)$$

r_t^c 是在 t 时刻虚拟手指尖的质心罚项。之前所述的 r_t^d 和 r_t^a 仅解决了指尖接触与否的问题，但接触表面指尖的分布仍需讨论，即何种分布能够更具稳定性。受工作^[7]的启发，我们认为“稳定的抓取”要求五指质心与目标物体质心的空间距离尽可能小，即：

$$r_t^c = -\frac{\|\Delta c\|_i}{\max(\|\Delta c\|)} \quad (13)$$

t 时刻五指质心的坐标为：

$$c_t = \frac{\sum_i (m^i \rho_t^i)}{\sum_i m^i} \quad (14)$$

其中 m^i 为第 i 个指尖的质量。

5 实验结果及分析

本文实验的硬件为一台配置Intel i7-9700KF处理器、NVIDIA GTX 2070显卡、16GB内存的服务器；软件平台为基于Unity3D及ML-Agents^[27]插件所搭建的虚拟实验环境。在训练阶段，我们设定训练物体的尺寸因子 $\in [0.5, 1.0]$ ，并定义权值参数 $w^p=0.4$ ， $w^v=0.2$ ， $w^e=0.4$ ， $w^p=0.25$ ， $w^a=0.25$ ， $w^c=0.5$ ，过渡相位值 $\phi_{trans}=0.4$ ，定义手部每个指段质量 m^i 为0.03kg，手掌的质量 m^{palm} 为0.5 kg。

下面分别从三个不同的角度，通过消融实验、自适应抓取实验和手部重定向实验，对实验结果进行定性和定量的分析。

5.1 消融实验

消融实验 (Ablation) 指用控制变量法，将实验因素逐一排除后再进行的对比实验。本小节以球体的抓取任务为例，对模仿奖励和自适应罚项两者进行变量控制。通过对“仅模仿”、“仅自适应”、“DeepMimic”和“模仿加自适应”四种不同的回报函数所得的训练模型分别进行消融实验，分析公式(4)中总回报函数每项的作用。

在训练阶段，四种模型所对应的累积回报增长曲线如图6所示。“仅模仿”是指回报函数 R_t 中自适应权值 w^G 取0， w^I 取1所对应的情况。由于此时不存在罚项，所以累积回报初始值为0，且累积回报上升空间较小。“仅自适应”指模仿权值 w^I 为0， w^G 为1的情况，此阶段只有罚项，且累积回报理论最大值为0。“DeepMimic”指 w^I 为1， w^G 也为1情况，以此代表文献[15]的学习策略。本文方法采用的“模仿+自适应”指 w^I 为1，且 w^G 满足式(5)的情况。后两种情况下训练初期罚项作用大，累积回报为负值；随着迭代步数的增长，累积回报逐渐变为正值；最后进入缓慢增长的阶段。相较于DeepMimic方法，由于式(5)中的动态权值缓解了初始手型罚项过大的问题，因此本文方法训练收敛速度更快。

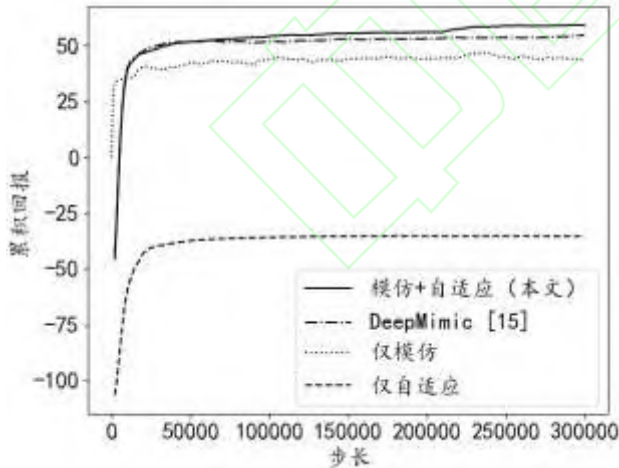


图6 训练阶段的累积回报增长曲线

Fig. 6 Cumulative return curve of training phase

“仅模仿”、“仅自适应”、“模仿加自适应”三种模型所对应的抓取效果如图7所示。图片中的红色小立方体表示目标物体的质心，蓝色小立方体表示虚拟手的五个指尖的质心。由于“DeepMimic”和“模仿加自适应”模型抓取效果区分度不大，这里省略该模型的后续结果分析。



(a) “仅模仿”策略

(a) Imitation-only strategy



(b) “仅自适应”策略

(b) Adaption-only strategy



(c) “模仿+自适应”策略

(c) Imitation- adaption-only strategy

图7 消融实验结果

Fig. 7 Ablation experiment results

上述三种模型对应的抓取结果指标如表2所示。表中的 $Sim(X, Y)$ 为序列相似度^[28]指标，本文以该指标反映生成序列的自然性，具体有：

$$sim(X, Y) = \frac{l}{\max(n, m)} \quad (15)$$

其中 X 为智能体 (白色) 生成的手部运动序列， Y 为示教数据 (红色) 参考运动序列， n 和 m 分别为两者对应的序列长度， l 为其公共子序列长度。我们定义 $d_{min}=0.5$ 为阈值，当序列间某些帧的姿态的最小二乘距离小于 d_{min} 时，由这些帧所构成的序列即为两序列的公共序列。 η 是质心离差比，有：

$$\eta = \frac{\|\Delta c\|_{t'}}{\lambda} \quad (16)$$

其中 t' 指五指均接触到物体的时刻。本文以该质心偏差的相对值反映生成抓取姿态的牢固程度。每组实验结果的指标取连续10幕的均值，后续实验将沿用上述实验量化指标。

表2 消融实验结果表

Table 2 Ablation experiment results

策略名称	$Sim(X, Y)$	η
“仅模仿”	89.70%	8.97%
“仅自适应”	75.43%	1.32%
“模仿+自适应”	86.42%	1.59%

由分析易得知，“仅模仿”情况下虚拟手会尽量逐帧地保持和参考运动序列的同步性。该策略虽然能够获得较高的序列相似度，但抓取稳定性指标较差，有时甚至不能保证指尖接触物体，如图7(a)所示。“仅自适应”情况下的虚拟手具备较稳定的抓取，但生成的抓取姿态自然性无法得到保障，例如如图7(b)所示指段交错的情况。最适合的模型为“模仿+自适应”的情况，此时生成的序列兼具较高的自然性和稳定性，如图7(c)所示。

5.2 自适应抓取实验

为了探究虚拟手的抓取自适应能力，本节先对物体素材库中的物体进行变量控制，进行形状自适应实验；再对尺寸因子进行变量控制，进行尺寸自适应实验。

5.2.1 形状自适应实验

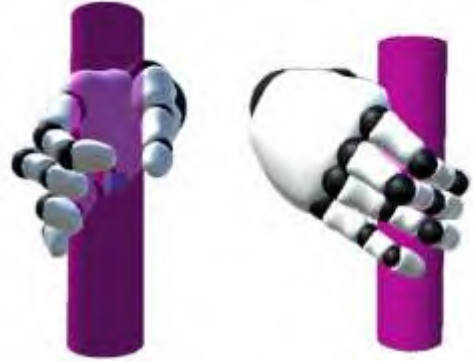
本小节将球体的训练模型分别作用于立方体、柱状体和球体的抓取任务，分析强化学习策略对不同形状分类物体的自适应能力。

三种物体所对应的抓取效果如图8所示，对应的抓取结果指标如表3所示。为了生成最佳的抓取运动序列，现有工作^{[6][13]}通常按照严格的分类标准，让模型去抓取对应形状分类的物体；欲对不同类型、不同尺寸的物体进行抓取，得到的抓取效果较差。通过分析可知，本文方法生成的抓取运动生成算法对于不同形状分类的物体同样能生成自然、稳定的抓取姿态，且量化指标均较为稳定。



(a) 立方体抓取任务

(a) Cube-grasping task



(b) 柱状体抓取任务

(b) Cylinder-grasping task



(c) 球体抓取任务

(c) Sphere-grasping task

图8 形状自适应实验结果

Fig. 8 Shape adaptation experiment results

表3 形状自适应实验结果表

Table 3 Shape adaptation experiment results

任务名称	$Sim(X, Y)$	η
立方抓取任务	83.77%	1.89%
柱状体抓取任务	82.08%	1.69%
球体抓取任务	86.42%	1.59%

5.2.2 尺寸自适应实验

本节以球体的抓取任务为例，对物体的尺寸因子进行变量控制，分析强化学习策略对训练阶段外不同尺寸物体的抓取自适应能力。在训练阶段，训练物体的尺寸因子 λ 满足 $\lambda \in [0.5, 1]$ ；在推演阶段，我们分别将尺寸因子 λ 设置为0.3和1.2。

两种尺寸的球体抓取任务效果如图9所示，对应的抓取结果指标如表4所示。通过分析可知，本文方法生成的抓取运动生成算法对于尺寸适中（ $\lambda = 0.75$ ）的物体具有较稳定的抓取效果。对于较大尺寸（ $\lambda = 1.2$ ）或较小尺寸（ $\lambda = 0.3$ ）物体也能做到一定程度的抓取。由于此时生成的抓取姿态已经和初始参考运动序列差别较大，因此 $Sim(X, Y)$ 指标偏低， η 指标偏高。



(a) 尺寸因子取0.3时
(a) Size factor equals 0.3



(b) 尺寸因子取1.2时
(b) Size factor equals 1.2



(c) 尺寸因子取0.75时
(c) Size factor equals 0.75

图9 尺寸自适应实验结果

Fig. 9 Size adaptation experiment results

表4 尺寸自适应实验结果表

Table 4 Size adaptation experiment results

λ	$Sim(X, Y)$	η
0.3	74.21%	7.52%
1.2	67.96%	6.73%
0.75	85.88%	1.30%

5.3 手部重定向实验

本小节在正常比例的虚拟手策略的基础上，将其直接作用于“较长指节”（1.3倍）、“较短指节”（0.7倍）和“长短指节”（0.7和1.3倍混合）三种不同拓扑比例的虚拟手模型上，探究该抓握算法的重定向能力。由于不同手部模型之间存在一定的拓扑差异，从最初的参考运动序列到最终虚拟手抓握的整个过程中，对算法的重定向能力提出了较高的要求。

三种虚拟手模型所对应的抓取效果如图10所示。经分析可知，对于手部拓扑结构的改变，本文提出的抓取运动生成算法具备一定的重定向能力。



(a) 较长指节
(a) Longer fingers



(b) 较短指节
(b) Shorter fingers



(c) 正常指节
(c) Standard fingers

图10 手部重定向实验结果

Fig. 10 Hand retarget experiment results

6 总结

本文提出了一种虚拟手的自适应抓取运动生成方法，设计了一种基于真人手部示教序列的离线深度强化学习控制器。结果表明，对于物体形状、尺寸在训练数据集以外的目标物体，该方法同样能够生成自然且稳定的抓取运动序列，且具备一定的重定向能力。

本文方法的局限性体现在虚拟手的运动仅由运动学算法生成，未考虑其与物体进行交互时的力闭合属性；此外本文算法也仅适用于由物体主体部分抽象出的简单基元。未来的工作可以考虑使用基于RGB视频的手部姿态估计方法获取参考运动序列，减少动作捕捉数据的获取成本。

参考文献 (References)

- [1] 付宜利, 刘诚. 虚拟装配中基于生理约束的虚拟手建模与抓持规划[J]. 计算机集成制造系统, 2009, 15(4): 681-684.
Fu Yili, Liu Cheng. Hand modeling and motion controlling based on lay figure in virtual assembly[J]. Computer Integrated Manufacturing Systems, 2009, 15(4): 681-684 (in Chinese).
- [2] Ferrari C, Canny J F. Planning optimal grasps[C]. //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos, 1992: 2290-2295.
- [3] Miller A T, Allen P K. Examples of 3D grasp quality computations[C]. //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos, 1999: 1240-1246.
- [4] Miller A T, Allen P K. Graspit! a versatile simulator for robotic grasping[J]. IEEE Robotics & Automation Magazine, 2004, 11(4): 110-122.
- [5] Ciocarlie M, Goldfeder C, Allen Dimensionality reduction for hand-independent dexterous robotic grasping[C]. //Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos, 2007: 3270-3275.
- [6] 蔡娴娟, 程成, Alhazmi等. 虚拟制造环境中虚拟手的行为构造[J]. 计算机辅助设计与图形学学报, 2015, 27(3):499-507.
Cai Xianjuan, Cheng Cheng, Alhazmi Marwah, et al. Research on Behavior Simulation of Virtual Hand in Virtual Manufacturing Environment[J]. Journal of Computer-Aided Design & Computer Graphics, 2015, 27(3):499-507 (in Chinese).
- [7] Peng Song, Zhongqi Fu, Ligang Liu. Grasp planning via hand-object geometric fitting[J]. The Visual Computer, 2018, 34(2): 257-270.
- [8] Brahmabhatt, Handa, Hays, et al. ContactGrasp: functional multi-finger grasp synthesis from contact[C]. // IEEE International Conference on Intelligent Robots and Systems. Macau, 2019: 2386-2393.
- [9] Tian, Wang, Manocha, et al. Realtime hand-object interaction using learned grasp space for virtual environments[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(8): 2623-2635.
- [10] Liu, Pan, Xu, et al. Generating grasp poses for a high-dof gripper using neural networks[C]. // IEEE RSJ International Conference on Intelligent Robots and Systems. Macau, 2019:1518-1525.
- [11] Starke, Eichmann, Eichmann, et al. Synergy-based, data-driven generation of object-specific grasps for anthropomorphic hands[C]. // International Conference on Humanoid Robots. Beijing, 2018: 327-333.
- [12] Kopicki, Detry, Adjigble, et al. One shot learning and generation of dexterous grasps for novel objects[J]. International Journal of Robotics Research, 2016, 35(8):959-976.
- [13] 王晓媛, 田浩, 王长波. 虚拟手自然抓取动作生成研究[J]. 计算机辅助设计与图形学学报, 2020, 32.
Wang Xiaoyuan, Tian Hao, Wang Changbo, et al. Research on Natural Grasp Generation of the Virtual Hand[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32 (in Chinese).
- [14] Liu Libin, Van De Panne, Yin KK. Guided learning of control graphs for physics-based characters[J]. ACM Transactions on Graphics, 2016, 35(3).
- [15] Xue Bin Peng, Abbeel, Pieter, Levine, et al. DeepMimic: example-guided deep reinforcement learning of physics-based character skills[J]. ACM Transactions on Graphics, 2018, 37(4).
- [16] Xue Bin Peng; Kanazawa, Malik, et al. Reinforcement learning of physical skills from videos[J]. ACM Transactions on Graphics, 2018, 37(6):178-192.
- [17] Libin L, Hodgins J. Learning to schedule control fragments for physics-based characters using deep q-learning[J]. ACM Transactions on Graphics, 2017, 36(3).
- [18] Libin Liu, Hodgins Jessica. Learning Basketball Dribbling skills using trajectory optimization and deep reinforcement learning [J]. ACM Transactions on Graphics, 2018, 37(4):1-14.
- [19] Yu R, Park H, Lee J. Figure skating simulation from video[J]. Computer Graphics, 2019, 38(7): 225-234.
- [20] Park H, Yu R, Lee J. Multi-segment foot modeling for human animation[J]. Computer Graphics Forum, 2019, 39(1):637-649.
- [21] 方伟, 黄增强, 徐建斌等. 基于Spark的分布式机器人强化学习训练框架[J]. 图学学报, 2019, 40(5): 852-857.
Fang W, Huang ZQ, Xu JB. Training Framework of Distributed Robot Reinforcement Learning Based on Spark[J]. Journal of Graphics, 2019, 40(5): 852-857 (in Chinese).
- [22] Aravind R, Vikash K, Abhishek G, et al. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations[C]. //Robotics: Science and Systems. Pittsburgh, 2018.
- [23] Mixamo[EB/OL]. [2020-01-12]. <https://www.mixamo.com/#/?page=1&query=grab&type=Motion%2CMotionPack>.
- [24] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[EB/OL]. 2019 [2019-09-15]. <https://arxiv.org/abs/1707.06347>.
- [25] Antotsiou D, Garcia G, Tae K, et al. Task-oriented hand motion retargeting for dexterous manipulation imitation[C]. //Computer Vision - ECCV 2018 Workshops, Munich, 2018: 287-301.
- [26] 刘乃军, 鲁涛, 蔡莹皓等. 机器人操作技能学习方法综述[J]. 自动化学报, 2019, 45(3):458-470.
Liu Naijun, Lu Tao, Cai Yinghao. A Review of Robot Manipulation Skills Learning Methods. Acta Automatica Sinica, 2019, 45(3):458-470 (in Chinese).
- [27] Juliani A, Berges V P, Vckay E, et al. Unity: a general platform for intelligent agents[EB/OL]. 2019, [2019-09-15]. <https://arxiv.org/abs/1809.02627>.
- [28] 陈昌伟. 基于Kinect的人体动作比对分析及生物力学分析[D]. 天津, 天津大学, 2014.
Chen Changwei. Comparative analysis and biomechanical analysis of human motion based on Kinect[D]. Tianjin, Tianjin University, 2014 (in Chinese).