

Visual Grasp Control Optimization Method and System for Manipulator Based on Deep Reinforcement Learning

Bang Lin, Wei Zhou, Jiangbo Fan, Rui Li

Zhaotong Power Supply Bureau, Yunnan Power Grid Co., Ltd., Zhaotong Yunnan
Email: 624646985@qq.com

Received: Nov. 6th, 2018; accepted: Nov. 22nd, 2018; published: Nov. 29th, 2018

Abstract

In order to improve the compatibility between visual image features and optimal control, this paper proposes an optimization method for visual grasp control of manipulator arm based on deep reinforcement learning, which can extract features from visual images generated by interaction with the environment independently and directly apply the extracted features. By continuously grasping the experiment and optimizing the manipulator arm grabbing strategy, an excellent grasping strategy is finally obtained. And we can learn the grasping control strategy of the manipulator according to different grasping tasks and grasping environment.

Keywords

Deep Learning, Manipulator Arm, Machine Vision, Grasping Strategy

基于深度强化学习的机械臂视觉抓取控制优化方法

林 邦, 周 伟, 范江波, 李 锐

云南电网有限责任公司昭通供电局, 云南 昭通
Email: 624646985@qq.com

收稿日期: 2018年11月6日; 录用日期: 2018年11月22日; 发布日期: 2018年11月29日

摘 要

针对提高视觉图像特征与优化控制之间契合度的问题, 本文提出一种基于深度强化学习的机械臂视觉抓

文章引用: 林邦, 周伟, 范江波, 李锐. 基于深度强化学习的机械臂视觉抓取控制优化方法[J]. 人工智能与机器人研究, 2018, 7(4): 200-206. DOI: 10.12677/airr.2018.74024

取控制优化方法,可以自主地从与环境交互产生的视觉图像中不断学习特征提取,直接地将提取的特征应用于机械臂抓取控制。通过不断的抓取实验,优化机械臂抓取策略,最终获得优秀的抓取策略。并且可以针对不同的抓取任务和抓取环境,学习相应的机械臂抓取控制策略。

关键词

深度学习,机械臂,机器视觉,抓取策略

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

神经网络是人工智能研究领域[1]的核心之一,其本质是模拟人脑的学习系统,通过增加网络的层数让机器从数据中学习高层特征[2]。近年来,由于神经网络优异的算法性能,已经广泛应用于图像分析、目标检测等领域[3][4]。

视觉图像提供了丰富的特征信息[5],可以提供某个时刻对于研究问题所需要的环境的有效状态信息。正因为其具有的丰富的特征信息量,视觉图像被广泛应用于如机器人导航、机械臂控制等许多应用上。纯粹基于视觉图像的抓取控制,其质量极大地依赖于图像的特征提取的准确性和机械臂抓取控制的动力学模型的准确性。但是,传统的视觉特征提取算法受环境动态变化的影响,手工提取的特征往往费时费力,很难满足机械臂抓取控制的需求。同时,机械臂运动轨迹的求解与规划,正逆动力学模型的建立往往也十分困难。

目前国内外许多文章提出了不同的机械臂控制方法,近年来由于图像识别技术的发展以及人工智能得到广泛的运用[6],出现了将两者结合起来用于机械臂的控制,并且取得了良好的效果。文献[7]将机械臂避障路径规划问题置于强化学习的框架当中,采用深度Q学习的方法训练策略以规划路径,使得机械臂能够在空间中存在障碍物的情况下实现避障抓捕。文献[8]对图像进行了预处理并提取了目标边缘,结合目标特点对提取的边缘进行了椭圆拟合,并对椭圆拟合偏差进行了分析。针对常规方法设计视觉控制器的不足,基于BP神经网络对视觉控制器进行了设计,考虑了不同网络参数和训练算法对网络训练效果的影响。文献[9]深入分析了机械臂特性和目标特性,在建立机械臂、相机和目标的运动关系的基础上,基于立体视觉实现了对空间目标的位姿测量,控制机械臂进行抓取预定位。考虑立体相机的测量误差和机械臂的控制精度因素,在抓取末阶段采用单目相机,基于图像的视觉伺服策略控制机械臂进行精确位姿调整对目标进行抓取。文献[10]以二自由度机械臂为基础,用摄像机获取机械臂运动区域的图像,通过图像分析,求解机械臂末端执行器在直角坐标空间的坐标值,将该值作为系统的实际位置,并通过反馈与期望坐标值相减,将所获得的误差作为机械臂控制系统的补偿量,实现整个系统的全闭环控制,进一步解决了原半闭环控制带来的控制精度不高的问题。文献[11]针对多自由度机械臂的神经网络控制问题,将控制系统根据自由度分为两个子系统,提出全局神经网络与局部神经网络组合对机械臂系统进行控制。

文献[12]指出机械臂视觉抓取系统中基于深度学习的视觉识别可通过对深度卷积神经网络的训练实现多目标快速识别而不必更改识别算法。这需要搜集被检测目标的大量图像制作标准的训练集,利用前述基于深度神经网络的目标识别算法对数据集进行离线训练,将得到的模型用于在线识别。R-CNN系列算法[13][14]是目前主流的用于机械臂抓取的深度学习目标检测算法,但是速度上并不能满足实时的要求。

机械臂视觉抓取在常规的机器人学、控制理论的基础上,融合了计算机视觉理论、深度学习和人工智能等学科,具有重要的科研和应用价值。

如何提高视觉图像特征与优化控制之间的契合度是现有技术需要解决的问题。针对现有技术的缺点,本文提出了一种结合深度强化学习[15]的机械臂视觉抓取控制优化方法。在建立在人工智能算法深度强化学习的基础上,能自主地学习视觉图像的特征提取,并有效地规划机械臂的抓取控制轨迹。同时,根据不同的应用环境和抓取目标,自主地学习相应的机械臂抓取控制策略,提高了算法的泛化能力。深度强化学习的应用,避免了人为的特征提取和运动轨迹计算,提高了准确度,简化了复杂度,从而提升实际机械臂抓取控制的效率。

2. 深度学习神经网络模块设计

设计一种深度学习的神经网络模块,前端采用卷积神经网络[16],不断学习提取高维的视觉图像信息,并将其处理成低维的状态信息。然后经过一个长短时记忆网络,有效地处理时序信息,提取机械臂每个状态前后的隐含信息。后端采用强化学习的执行器-评价器[17]结构,获取低维的状态信息,从中学习提取有效信息,输出策略动作,从而控制机械臂的抓取。

由于整体框架采用神经网络,并通过反向梯度传播算法进行迭代学习,可以大大减少手工处理图像信息的难度。并且可以根据不同的机械臂工作环境,进行相应的抓取控制策略的学习,大大提升了该方法的实用性和泛化能力。此外,所提出的方法是端到端地进行机械臂抓取控制策略的学习,可以自主的学习有效的、必要的信息,提供更准确的抓取控制,达到基于视觉图像的机械臂抓取控制目的。

2.1. 执行器-评价器结构

执行器-评价器结构是指 ACTOR-CRITIC METHOD,一种结合了 Policy Gradient (Actor)和 Function Approximation (Critic)的方法。执行器(Actor)基于概率选行为,评价器(Critic)基于执行器(Actor)的行为评判行为的得分,执行器(Actor)根据评价器(Critic)的评分修改选行为的概率。该结构既有基于值的方法(Q-learning 等)可以进行单步更新的优点,提升了学习效率,又有基于策略的方法(Policy Gradient 等),能在连续动作空间进行决策的优势,扩大了应用范围。

2.2. 卷积神经网络

将机械臂视觉传感器获取的图像进行预处理,处理后的图像为一系列的 160×120 的红绿蓝图像 o_t , 输入到设计的神经网络,就可以获得通过卷积神经后的一组特征图 $F_t \in R^{m \times c \times w \times h}$, 其中 n 为一次机械臂抓取过程的阶段长度, c 为卷积神经网络的输出特征图通道数, w, h 分别代表为特征图的长和高,如公式所示:

$$F_t = \text{Convs}(o_t; \theta_t)$$

将卷积神经网络输出后的二维特征图,转变为一维特征 $F_t^f \in R^{1 \times V}$, 其中, $V = c \times w \times h$, 如公式所示:

$$F_t^f = \text{Flatten}(F_t)$$

将一维特征向量 $F_{t:n}^f$ 输入卷积层后的递归层长短时记忆网络 LSTM。在每个训练阶段开始,长短时记忆网络的隐含状态都为 0, 如公式所示:

$$H_t = \begin{cases} 0 & t = 0 \\ \text{LSTM}(F_t^f, H_{t-1} | \theta_t) & 0 < t \leq n \end{cases}$$

将递归神经网络输出 O^R ，通过柔性最大值传输函数变化映射成为机械臂的每个动作选择概率 P_i 和通过线性变化映射成为机械臂的状态值 V_i 值，如公式所示：

$$P_i = \text{Softmax}(O^R | \theta_i)$$

$$V_i = \text{Linear}(O^R | \theta_i)$$

本文所述的深度学习神经网络模块的优点是，可以自主地在大量数据中学习机械臂抓取控制的同时，应对不同场景不必重新设计算法，可以自主学习以提供不同的抓取策略，并且单纯的基于视觉图像。

3. 控制优化方法及实例

3.1. 学习流程

在图 1 所示学习流程中，智能体观察机械臂抓取的环境视觉图像，经过图 2 所示的神经网络的处理，应用深度强化学习，直接输出机械臂抓取控制策略。控制机械臂关节的移动，从而获得相应的来自于环境的奖励。智能体通过奖励的学习，不断优化机械臂的抓取控制策略，在下一次抓取时，提供更加优秀的抓取策略。通过不断地与环境交互的自主学习，机械臂渐渐的学会靠近抓取目标，最终抓取物体。

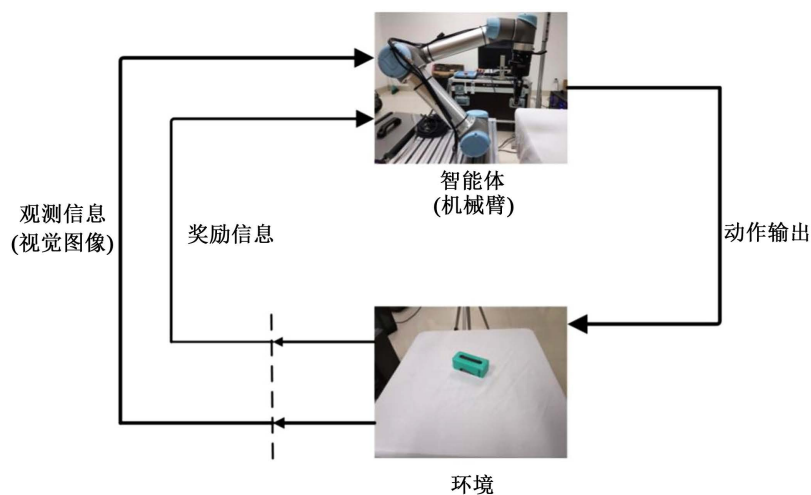


Figure 1. Learning process

图 1. 学习流程

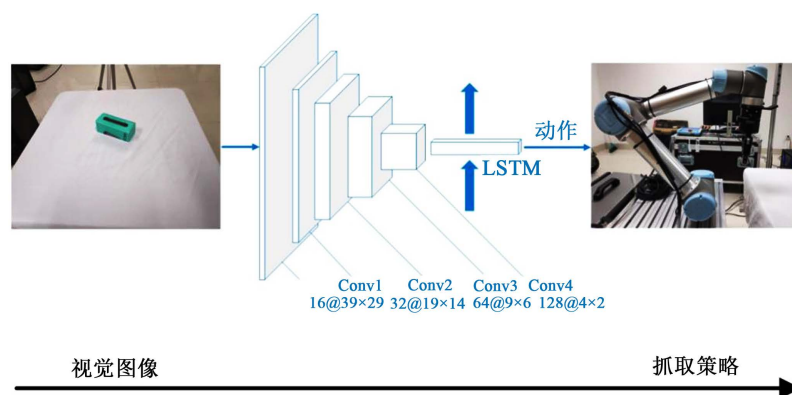


Figure 2. Neural network structure

图 2. 神经网络结构

3.2. 控制策略

神经网络架构获取机械臂环境的红绿蓝视觉图像。首先通过卷积神经网络，处理高维的视觉图像信息。然后经过全连接神经网络，提取有效环境状态特征信息。最后采用强化学习的执行器 - 评价器结构，直接输出机械臂的抓取控制策略。根据与环境交互产生的奖励，利用强化学习算法的贝尔曼方程，进行方向梯度传播进行梯度更新，优化神经网络参数，从而优化最终控制策略。

3.2.1. 贝尔曼方程

贝尔曼方程是对强化学习中值函数的一种迭代求解方法。首先定义累积回报 G_t (带衰减系数 γ)，如公式所示：

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

从而可以定义值函数为回报的期望：

$$v(s) = E[G_t | S_t = s]$$

将上式展开：

$$\begin{aligned} v(s) &= E[G_t | S_t = s] \\ &= E[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | S_t = s] \\ &= E[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + \dots) | S_t = s] \\ &= E[R_{t+1} + \gamma G_{t+1} | S_t = s] \\ &= E[R_{t+1} + \gamma v(S_{t+1}) | S_t = s] \end{aligned}$$

从而得到： $v(s) = E[R_{t+1} + \gamma v(S_{t+1}) | S_t = s]$ 。

上式即为贝尔曼方程，它表明当前状态的价值和下一步的价值以及当前的反馈回报有关，说明值函数是可以通过迭代来进行计算的。

3.2.2. 神经网络参数选取

图 2 所示的神经网络结构里，神经网络的输入为同一系列中的 4 张 160×120 的红绿蓝三通道图像，输出为机械臂 7 个关节的角速度和夹具的动作(打开或关闭)。感知网络由四层卷积神经网络和一层长短期记忆单元组成。第一个卷积层使用了 16 个 10×10 的卷积核，步长为 4，填充为 1，卷积后得到 16 个 39×29 的特征图。第二层使用 2 个 5×5 的卷积核，步长为 2，填充为 1，卷积后得到 32 个 19×14 的特征图。第三层使用 2 个 5×6 的卷积核，步长为 2，填充为 1，卷积后得到 64 个 9×6 的特征图。第四层使用 2 个 5×6 的卷积核，步长为 2，填充为 1，卷积后得到 128 个 4×2 的特征图。将这些特征图送入长短期记忆单元，输出一个 8 维向量，前 7 维表征机械臂关节角速度，最后一维表征夹具动作。卷积网络使用 kaiming 高斯初始化方法。强化学习部分使用基于随机排列的经验次序回放。探索策略使用 ϵ 贪婪策略，初始化为 0.5，然后在训练中退火到 0.1。衰减系数 γ 为 0.5。

3.3. 抓取控制实例

机械臂在仿真环境中机械臂学习抓取控制，抓取其中一个物块，将其叠放在另外一个物块纸上，比如红色物块叠放在绿色物块之上，如图 3 所示。

机械臂在真实环境中的抓取控制实例。上半图所示为机械臂学习抓取控制，抓取物块移动到黄色目标位置。下半图所示为机械臂学习抓取，成功检测到桌上物块的位置，准确定位并实施抓取，如图 4 所示。

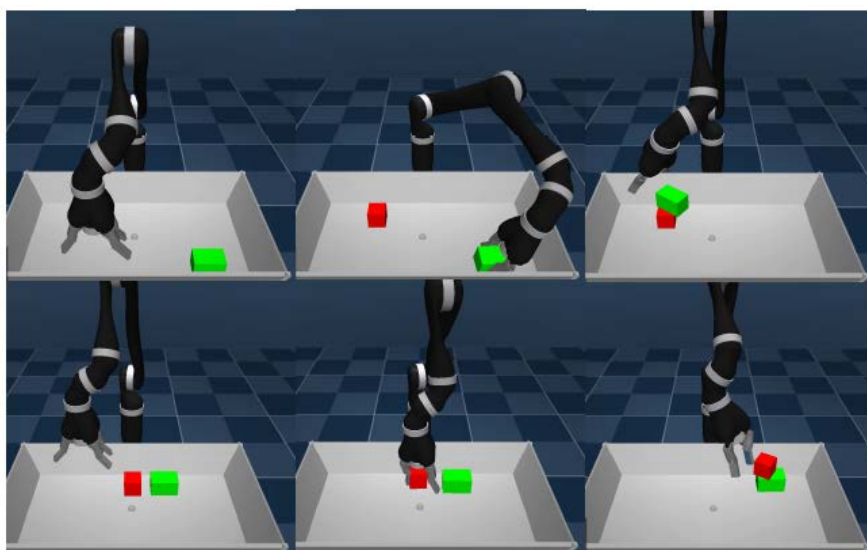


Figure 3. Example of grab control of the robot arm in the simulation environment

图 3. 机械臂在仿真环境中的抓取控制实例

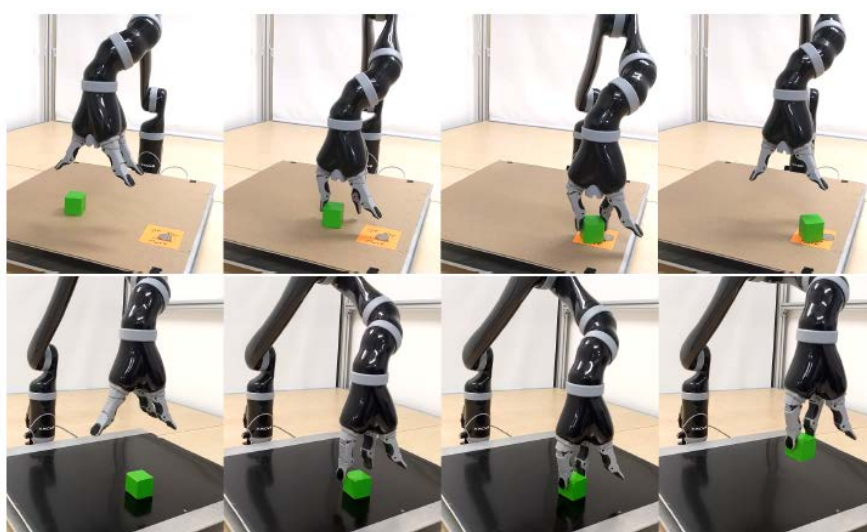


Figure 4. Example of grab control of a robot arm in a real environment

图 4. 机械臂在真实环境中的抓取控制实例

4. 结束语

本文提出了一种结合深度强化学习的机械臂视觉抓取控制优化方法。在建立在人工智能算法深度强化学习的基础上，能自主地学习视觉图像的特征提取，并有效地规划机械臂的抓取控制轨迹。同时，根据不同的应用环境和抓取目标，自主地学习相应的机械臂抓取控制策略，提高了算法的泛化能力。深度强化学习的应用，避免了人为的特征提取和运动轨迹计算，提高了准确度，简化了复杂度，从而提升实际机械臂抓取控制的效率。

参考文献

- [1] 曾毅, 刘成林, 谭铁牛. 类脑智能研究的回顾与展望[J]. 计算机学报, 2016, 39(1): 212-222.

- [2] 朱虎明, 李佩, 焦李成, 等. 深度神经网络并行化研究综述[J]. 计算机学报, 2018, 41(8): 1861-1881.
- [3] Kang, L., Ye, P., Li, Y. and Doermann, D. (2014) Convolutional Neural Networks for no-Reference Image Quality Assessment. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1733-1740. <https://doi.org/10.1109/CVPR.2014.224>
- [4] Ren, S.Q., He, K.M., Girshick, R., et al. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Proceedings of the Neural Information and Processing System*, Montreal, 7-12 December 2015, 91-99.
- [5] 马玉芳. 基于智能视觉的微型高精度图像采集系统设计[J]. 现代电子技术, 2018, 41(19): 67-70.
- [6] 孔峻. 深度卷积神经网络在计算机视觉中的应用[J]. 电子技术与软件工程, 2018(21): 130-131.
- [7] 王翌, 胡立生. 基于深度 Q 学习的工业机械臂路径规划方法[J]. 化工自动化及仪表, 2018(2): 141-145.
- [8] 李文博. 基于视觉伺服的机械臂控制系统研究[D]: [硕士学位论文]. 南京: 南京理工大学, 2016.
- [9] 李宇飞, 高朝辉, 申麟. 基于视觉的机械臂空间目标抓取策略研究[J]. 中国科学: 技术科学, 2015, 45(1): 31-35.
- [10] 王军南, 丛爽. 基于视觉定位的二自由度机械臂控制系统[J]. 机械与电子, 2008(2): 49-52.
- [11] 杨马英, 葛铁众. 基于 BP 神经网络的机械臂视觉伺服控制[J]. 计算机应用, 2017(a02): 279-282.
- [12] 白成超, 晏卓, 宋俊霖. 结合深度学习的机械臂视觉抓取控制[J]. 载人航天, 2018, 21(3): 299-307.
- [13] Redmon, J., Divvala, S., Girshick, R., et al. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 26-30 June 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [14] Redmon, J. and Farhadi, A. (2016) YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525.
- [15] 薛天. 深度强化学习原理及其在机器人运动控制中的运用[J]. 通讯世界, 2018(8): 240-241.
- [16] 任秉银, 魏坤, 吴卓琦. 机械臂视觉伺服路径规划研究进展[J]. 哈尔滨工业大学学报, 2018, 50(1): 1-10.
- [17] 陈学松, 杨宜民. 基于执行器-评价器学习的自适应 PID 控制[J]. 控制理论与应用, 2011, 28(8): 1187-1192.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2326-3415, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: airr@hanspub.org