

基于深度学习的机器人复杂动作控制技术研究

朱煜东¹, 万韬阮², 汤汶³, 薛涛¹

(1. 西安工程大学 计算机科学学院/陕西省服装设计智能化重点实验室, 西安 710048;

2. 布拉德福德大学; 3. 伯恩茅斯大学)

摘要: 本文提出了一种基于深度学习的手势识别来控制机器人做复杂动作的解决方案。本文提出的方案展现出有效的开放性和可扩展性, 无论是手势识别还是在机器人复杂动作方面, 都可以根据需求做相应调整。同时, 本研究工作创建了一个有效的基于人工智能网络的机器人控制实用实验平台。实验设计分别以运用 6 个基础手势控制机器人在危险环境中清除可疑物品为例, 验证了开发的平台可以让机器人实时、远程地完成预设复杂动作, 满足现实中自然人机交互的功能。

关键词: 深度学习; 手势识别; 机器人复杂动作; 人机交互; 实时远程控制

中图分类号: TP31

文献标识码: A

Research on Robot Complex Motion Control Technology Based on Deep Learning

Zhu Yudong¹, Wan Taoruan², Tang Wen³, Xue Tao¹

(1. Shaanxi Key Laboratory of Clothing Intelligence, School of Computer Science, Xi'an

Polytechnic University, Xi'an 710048, China; 2. University of Bradford; 3. Bournemouth University)

Abstract: In the paper, a solution based on deep learning gesture recognition to control the robot to do complex actions is proposed. The scheme proposed in this paper shows effective openness and expansibility, and can be adjusted according to the needs in both gesture recognition and complex robot actions. At the same time, this research also created an effective robot control experimental platform based on artificial intelligence network. The experiment design uses six basic gestures to control the robot to clear suspicious objects in dangerous environment. The experiment results show that the developed platform can enable the robot to complete preset complex actions in real time and remotely, and meet the function of natural human-computer interaction in reality.

Keywords: deep learning; gesture recognition; complex movements of robots; human-computer interaction; real-time remote control

0 引言

随着机器人在工业生产和日常生活的应用不断深入, 具有良好交互性^[5]的仿人机器人具有广阔的应用前景。在一些高危工作场景中, 如清除可疑爆炸物体、高辐射、充斥病毒细菌等场所, 机器人代替人完成一些工作具有重要意义, 而如何实时、远程、高效地人机交互是一个重要的研究领域。

人机交互技术^[4]是研究人与计算机之间通过相互的识别与理解的交流与通信, 在最大范围内为人们完成信息管理、信息服务和信息处理等功能的一门技术学科^[5]。而自然人机交互技术具有便捷、高效、实用等特点^[6], 逐渐成为未来几十年人机交互发展的新趋势。而基于手势的人机交互技术就是其中一种重要的人机自然交互方式。

进行手势识别是基于手势的人机交互技术的核心^[7]。

传统的识别是通过摄像头获取图像, 然后通过预处理和特征提取, 再经过特征选择, 最后进行推测或识别。深度学习是一种深层非线性网络结构, 通过网络中的多隐层节点将低层特征组合成更抽象的高层特征表示, 实现复杂函数逼近, 从少数样本学习数据集中本质特征的能力。景雨等^[8]提出了一种优化改进的多尺度深度学习网络, 通过运用提取的不同尺度特征来更准确地表征图像, 基于自适应多尺度特性实现了同一卷积层不同尺寸卷积核得到不同尺度特征, 从而提高卷积神经网络的识别率, 但文章没有提供识别图像, 另外自建的数据集中背景复杂, 在暗背景下, 很难从图像中提取手势。郝禹哲等^[9]基于 TensorFlow 深度学习平台构建 CNN 网络, 实验将手势图像作为输入, 通过 4 个卷积层, 2 个卷积层后接 1 个池化层, 再通过全连接后, 利用 Dropout 技术移除一些节点及其相关的

输入输出,然后用 SoftMax 层进行分类判决,手势识别准确率可达到 97% 左右,但在复杂背景下的动态手势识别存在不足,也缺乏手势图像实时采集功能。

本文采用基于飞桨 PaddlePaddle^[10] 框架实现手势判断,深度学习模型采用经典的卷积神经网络 LeNet-5^[11] 模型实验,进行了参数优化,分析比较模型在不同参数下的性能,从而找出合理的模型参数。通过 OpenCV 获取摄像头实时手势图像,然后利用训练后的模型进行识别,将识别结果传送给机器人,机器人得到指令后控制总线舵机完成相应动作,进而实现自然人机交互的功能。本方案的优势在于系统开放性和可扩展性强,手势识别可根据实际需求建立不同的训练库,从而实现手势的更改,另外机器人复杂动作也通过机器人上位机动作编辑软件编辑全新动作,可以说本设计同时也搭建了一个基于人工智能网络的机器人控制实用实验平台。

1 深度学习模型

CNN 是一种由卷积层、池化层和全连接层等部分组成的深度神经网络,输入数据通过卷积层提取其特征,然后通过池化层减少冗余数据,最大限度保留特征,最后通过全连接层将特征发送给分类层进行分类判断。

本文模型采用 CNN 中的经典 LeNet-5 模型,如图 1 所示。LeNet-5 共有 7 层,主要由 2 个卷积层、2 个池化层和 3 个全连接层组成,不包含输入,每层都包含可训练参数。

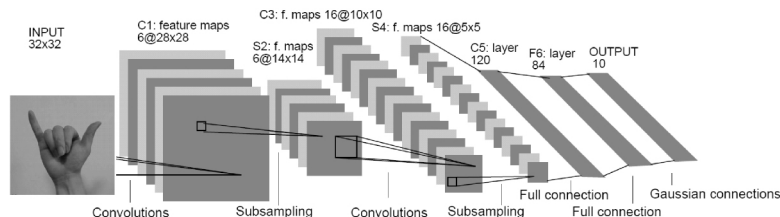


图 1 LeNet-5 模型

2 手势样本图像训练及结果分析

2.1 训练方式与过程

本实验采用自建数据集与百度 AI Studio 数据集相结合,其中自建数据集采集 6 种不同手势,如图 2 所示,每个手势 208 张图片,共包含 1248 张图片。训练和测试比例为 9:1。采用迭代次数终止,即迭代次数达到一定次数后停止训练。分别采取固定数据批处理大小与微调参数迭代次数参数训练方法,在多个基准数据集上进行系统性实验。

首先取数据批处理大小(batch size)为 128 进行训练。迭代次数(Epochs num)分别选取 20、40、80、100、120,训



图 2 数据集中的 6 种手势(依次为手势 1~6)

练损失值和模型校验平均准确率随着迭代次数的增加,模型校验平均准确率增加,当迭代次数达到 120 次时,数据集的准确率已稳定,基本不会发生太大改变,如表 1 所列。

然后取数据批处理大小(batch size)为 256,按不同迭代次数进行训练。训练损失值和模型校验平均准确率同样随着迭代次数的增加而增加,当迭代次数达到 120 次时,数据集的准确率已稳定,如表 2 所列。

表 1 模型校验平均准确率 1

迭代数	模型校验 平均准确率
20	0.796875
40	0.946022
60	0.963068
80	0.96875
100	0.9744318
120	0.97601944

表 2 模型校验平均准确率 2

迭代数	模型校验 平均准确率
20	0.85954213
40	0.9506183
60	0.96732956
80	0.9630682
100	0.97159094
120	0.97159094

训练准确率随着迭代次数的增加而逐渐增加,随后开始趋于稳定。损失值随着迭代次数的增加而逐渐减小,随后开始稳定在较小的值附近,说明训练好的网络模型具有

较好的稳定性。结合 batch size 为 128 时训练的数据可以得出,当 batch size 为 256 时训练的效率高。

2.2 模型验证

通过采集手势图片进行验证,分别对预设手势 1~6 进行验证,实验结果表明,训练后的模型能够有效、准确地进行手势识别。手势识别的实验结果同时也显示出算法对手势图像在平移、旋转等因素也有较强的鲁棒性,相比其他方法,基于深度学习的手势识别不需要专用图像处理硬件,即能满足人机交互过程中的实时性和自然性。

3 机器人复杂动作控制

3.1 机器人动作规划

本文的实验机器人采用国内幻尔公司的 TonyPi 机器人,如图 3 所示。TonyPi 机器人是一款优秀的仿人形智能机器人,机器人搭载树莓派 3B 微型计算机,机器人内置代码采用 Python 语言编程,运动单元采用串口总线舵机控制单元。控制器与舵机之间采用总线方式通信,通信波

特率为 115 200。每个舵机拥有一个 ID 号,用户可以用 Python 语言编写串口通信程序,发出包含 ID 信息的指令,匹配 ID 号的舵机就能完整接收这个指令,并按照指令执行动作。动作编辑软件界面如图 4 所示。



图 3 TonyPi 机器人

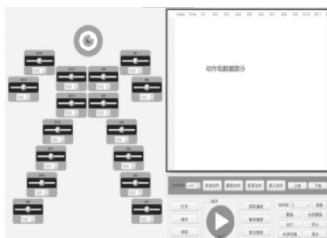


图 4 动作编辑软件界面

幻尔 TonyPi 机器人内置了很好的上位机软件,在上位机软件中可以编辑动作,给每个舵机设置转动参数,通过调试设计出符合预期的动作组,然后保存动作,保存后的动作存放在固定的动作组文件夹中,可供 Python 程序代码调用。本文通过动作编辑软件编辑了前进、后退、左移动、右移动、搬起物品、放下物品和搬运物品 7 个动作组,如图 5~图 7 所示。

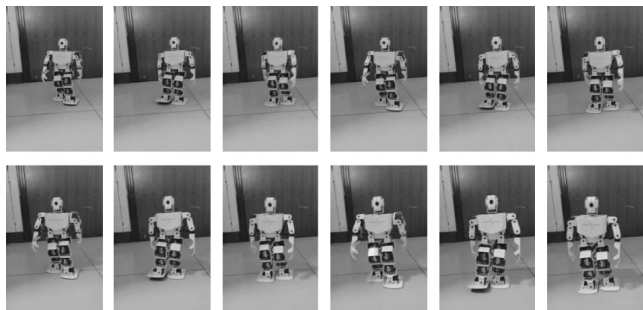


图 5 前进(后退)动作关键帧图示

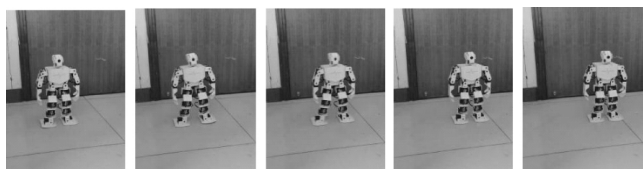


图 6 左移动作关键帧图示

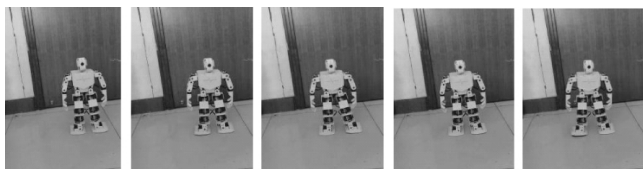


图 7 右移动作关键帧图示

3.2 人机交互控制实现

3.2.1 实验设计

机器人复杂动作控制系统以 TonyPi 仿人机器人为平台,TonyPi 机器人身高 36 cm,共有 16 个总线舵机、18 个

自由度,能够灵活地做出各种复杂动作。硬件配置:64 位 1.5 GHz 四核博通 CPU,2.4 GHz/5 GHz 双频 Wi-Fi, Raspbian GNU/Linux 10 操作系统,支持跨平台、多语言编程。

在深度学习训练手势库后,通过 Python 代码的 OpenCV 从电脑摄像头实时获取人的手势动作图像,获取图像后进行预处理,图片处理完成后用训练后的模型进行手势预测,然后将预测后的编码通过 socket 库函数发送,相应的 TonyPi 机器人中的 Python 程序通过 socket 库函数接收手势编码,根据手势编码调用相应基础动作,从而机器人作出响应动作,进而实现人机交互控制。实验流程图如图 8 所示。

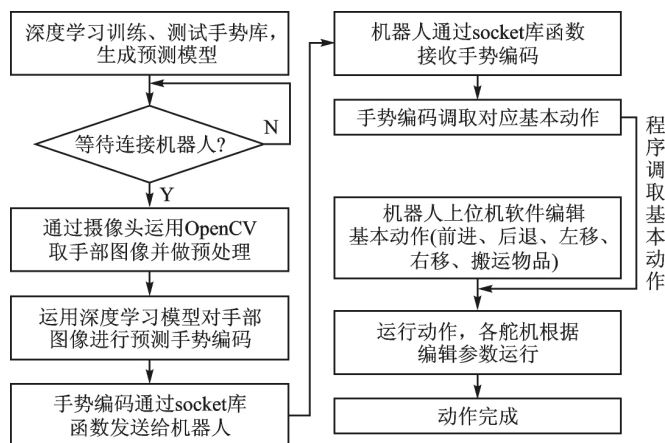


图 8 实验流程图

3.2.2 实验任务

实验分为两部分:一部分为通过手势 1~6 控制机器人行走,本部分只运用了其中 4 个手势,其中手势 1 关联机器人前进,手势 2 关联机器人左移,手势 4 关联机器人右移,手势 6 关联后退,实验者可以通过以上 4 个手势控制机器人行走或离开目标区域;实验的另一部分为同样通过手势 1~6 控制机器人搬运物品,其中手势 1 为机器人弯腰取物品,手势 2 为搬运物品,手势 4 为弯腰放物品,手势 6 为松手并立正,通过 4 个手势控制机器人完成搬运物品的动作。当然手势对应动作也可以分解为更加细致的动作单元,如弯腰、抓取物品、起立、移动等,本文考虑到动作及手势的简洁性,搬运物品中运用了一些合成动作,如图 9~图 11 所示。

实验结果为:通过几个手势可以远程、实时地控制机器人搬运物品,满足现实中自然人机交互的功能。由于手势识别运用深度学习模型,后期还可以按需求收集其他手势集完成更多手势识别,对应编辑需求的机器人动作,按需求进行自由组合,完成更多的动作。

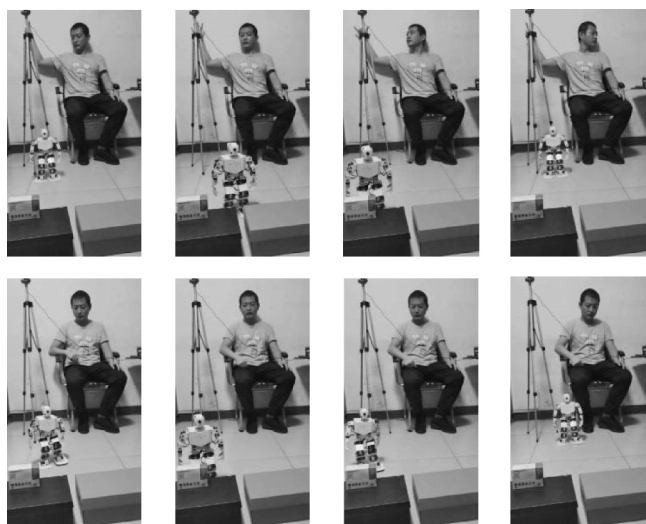


图9 手势控制机器人完成行走动作
(上为手势,下为相应动作)关键帧图示

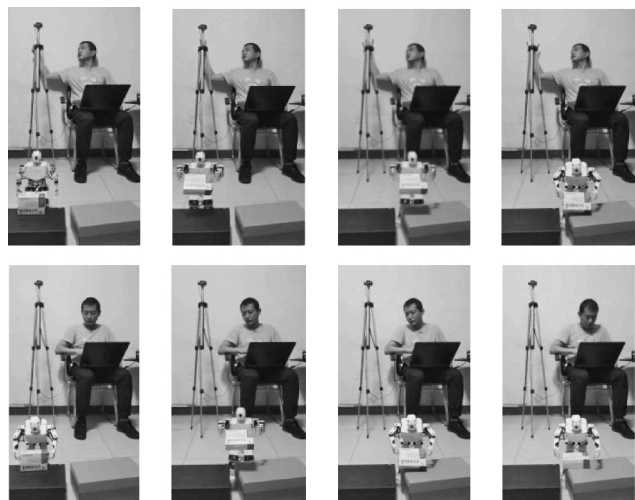


图10 手势控制机器人完成搬运物品动作
(上为手势,下为相应动作)关键帧图示

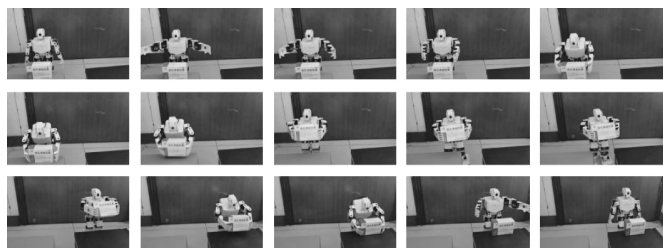


图11 机器人完成搬运物品动作关键帧图示

4 结 语

手势识别技术为人机交互提供一种新的非接触式交互,本文通过深度学习模式获取手势信息,利用卷积神经网络的学习性对数据集进行分析,提高了识别正确率,测

试中手势识别的精度达到 97% 以上,结合 TonyPi 机器人强大的运动控制能力可以很好地完成大量复杂动作,实验验证了手势识别系统的可行性,满足机器人复杂动作依靠自然手势连续实时地控制,实现更好的自然人机交互,同时也创建了一个基于人工智能网络的机器人控制实验平台。

本方案进一步可能改进的工作如下:增加和设计高效的手势与动作配对数据库,以扩大可执行动作的复杂性,这样可以满足不同应用场合;引进广义手势概念,比如加入识别人体上肢甚至加腿部动作来进行控制,如实验者作抱取动作,机器人也相应作抱取动作,更能贴近自然交互。ME

参考文献

- [1] 李健平,郑煜辉,罗威,等. 基于 Kinect 体感控制的机器人全动作跟随系统设计[J]. 电子世界,2019(7):131-133.
- [2] 李佳. 工业机器人,推动智能制造—访中国机械工业联合会执行副会长、中国机器人产业联盟执行理事长兼秘书长宋晓刚[J]. 制造技术与机床,2015(6):56-57.
- [3] WenTang, Marc Cavazza, Dale Mountain, et al. A constrained inverse kinematics technique for real-time motion capture animation[J]. The Visual Computer, Springer-Verlag, 1999, 15(7-8):413-425.
- [4] 罗泽仁. 科学技术哲学视域下的人机交互技术研究[J]. 科技资讯,2020(15):9-10.
- [5] 杨涛. 基于手势人机交互专利技术综述[J]. 中国科技投资, 2017(30):319.
- [6] 张莹莹,郭星. 基于 Kinect 动态手势识别算法的研究与实现[J]. 计算机技术与发展,2017,27(12):11-15.
- [7] 刘玉月,吴东苏,顾宏斌. 基于视觉伺服的手部跟踪仿真研究[J]. 信息技术,2012(5):9-15.
- [8] 景雨,祁瑞华,刘建鑫,等. 基于改进多尺度深度卷积网络的手势识别算法[J]. 计算机科学,2020(6):180-183.
- [9] 郝禹哲. 基于卷积神经网络的手势识别方法[J]. 传感器与微系统,2020(5):48-50.
- [10] 马艳军,于佃海,吴甜. 飞桨:源于产业实践的开源深度学习平台[J]. 数据与计算发展前沿,2019,1(1):105-115.
- [11] 张万征,胡志坤,李小龙. 基于 LeNet-5 的卷积神经图像识别算法[J]. 图像处理,2020,35(5):486-490.
- [12] 闻力生. 人工智能在服装智能制造中的应用[J]. 纺织高校基础科学学报,2020,33(2):30-36.

通信作者:朱煜东,324694556@qq.com。

(责任编辑:薛士然 收稿日期:2021-03-08)