基于卷积神经网络的手势识别控制系统

张行健 张建新*
(浙江理工大学机械与自动控制学院 浙江 杭州 310018)

摘 要 针对传统工业中机械手的人机交互方式不够直观的问题,利用卷积神经网络(CNN)设计一种基于手势动作的机器视觉型控制系统。采用 OpenCV 构建手势数据集,以 CNN 中的 AlexNet 结构为基础,改进和优化为一个更适合手势识别的 13 层 CNN 模型;通过串口通信技术将上位机的手势识别结果传给下位机,利用 STM32 单片机实现对机械手的相应控制。实验结果表明,该方式在测试集上的手势识别准确率平均为 98%,能直观且便捷地控制机械手作业。

关键词 手势识别 卷积神经网络 机械手控制

中图分类号 TP24 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2020.10.035

GESTURE RECOGNITION CONTROL SYSTEM BASED ON CNN

Zhang Xingjian Zhang Jianxin*
(*Faculty of Mechnical Engineering and Automation, Zhejiang Sci-Tech University, Hangzhou 310018, Zhejiang, China*)

Abstract Aiming at the problem that the man-machine interaction mode of the manipulator in the traditional industry is not direct, we use a convolutional neural network(CNN) to design a machine vision control system based on gesture. OpenCV was used to construct a gesture data set. Based on the AlexNet structure of CNN, a 13 layer CNN model which was more suitable for gesture recognition was improved and optimized. Through serial communication technology, the gesture recognition results of the upper computer were transmitted to the lower computer. Finally, the corresponding control of the manipulator was realized by STM32. The experimental results show that the accuracy of gesture recognition on the test set is 98% on average, which can directly and conveniently control the manipulator operation.

Keywords Gesture recognition CNN Manipulator control

0 引 言

随着机器学习和嵌入式等技术的发展,现代工业愈加智能化[1]。目前工业领域中传统的人机交互方式复杂,在操作上不够自然直观。为此,研究一种能实现对机械手的直观控制的新型人机交互方式很有必要,其中一种交互方式是让机械手直接模拟操作者的动作,实现直接控制,这一方式的关键是计算机能够对手势进行识别和判断。

目前手势识别方法主要分为三个类别。(1) 基于超声波的手势识别。Yang 等[2]提出的手势识别是利用多个超声波装置检测手势的位置,该方法的识别平均准确率可以达到 93%,但是在较复杂的噪声环境中容易被干扰。(2) 基于传感器的手势识别。谢小雨[3]提出一种利用手势控制臂带(MYO)传感器采集的肌电信息和加速度实现手势的识别,该方法的识别平均准确率可以达到 96%,但该类方法由于需要佩戴传感器,使用不够方便等。(3) 基于机器视觉的手势识别。贺航[4]基于 OpenCV 函数库,利用基于 HU 不变矩提取手势的图像特征值,再通过模板匹配法完成手势识别,但该方法对不同人的手势识别准确率差异较大,泛化性较差。孙玉等[5]利用 Leap Motion 设备获取手势的三维坐标信息,并结合长短期记忆网络模型进行动态手势识别,但 Leap Motion 设备较为昂贵,不利于工业的大规模使用。朱雯文等[6]提出了一种基于加速度

收稿日期:2019-07-10。浙江省高等教育"十三五"第一批教学改革研究项目(jg20180090);浙江理工大学教育教学改革研究项目(jgyl201704)。**张行健**,本科生,主研领域:嵌入式系统,机器学习。**张建新**,教授。

信号的卷积神经网络(CNN)模型,但是由于采用的是简单的 LeNet 网络结构,手势识别准确率只能达到 90%。石雨鑫等[7]提出了一种将 CNN 模型和随机森林(RF)相结合的手势识别算法,其识别准确率达到 97%,但 RF 会明显加大运算量,且容易在噪声较大的分类中出现过拟合现象。

本文将采用机器视觉和卷积神经网络的思路,提出一种以 AlexNet 结构为基础针对手势特征识别的 CNN 模型。改进后的 CNN 模型可以有效识别手势,通过 STM32 单片机实现对机械手的直观控制。本文改进的 CNN 模型的手势识别正确率在测试集中达到 98%。

1　总体方案设计

本文设计的手势识别控制系统是先用摄像头保存操作者的手势动作,然后用 PC 机中预先训练好的 CNN 模型去识别手势动作,再将手势动作的识别结果通过串口通信传输给 STM32 单片机,最后通过单片机控制机械手,实现操作者手势直接控制机械手的功能。总体方案设计如图 1 所示。

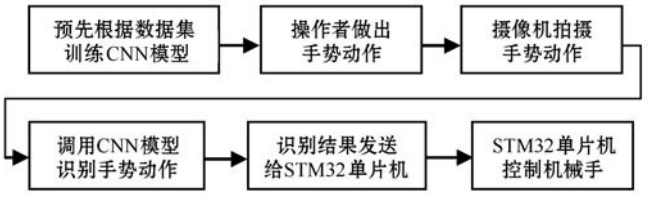

图 1　总体方案设计图

2　卷积神经网络及其改进

2.1　卷积神经网络设计

CNN 的典型结构如图 2 所示。CNN 发展历史中的里程碑事件是 Alex 提出 AlexNet 结构,获得了 ILSVRC-2012 大赛冠军[8]。AlexNet 结构通过 6 000 万个参数和 65 万个神经元,成功地把 120 万幅高分辨率图像分成 1 000 个不同的类别[8]。

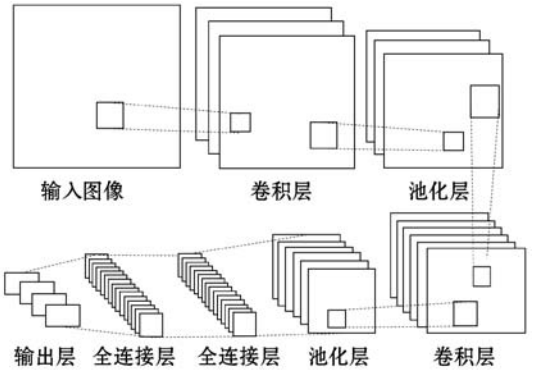

图 2　CNN 典型结构示意图

本文提出的 CNN 结构以 AlexNet 结构为基础,并根据实际手势识别的情况做了以下改进(具体结构参数如表 1 所示):

(1) 删除 LRN 层。AlexNet 结构使用局部响应归一化(LRN)用于正则化。但是 Simonyan 等[9]发现 LRN 的效果十分细微,反而会大量占据存储量和花费许多计算时间。本文在有无 LRN 层的 CNN 模型上进行测试,结果并无区别,说明 LRN 层确实性价比太低,于是本文删除 LRN 层,以加速训练过程。

(2) 改小卷积核大小。AlexNet 结构的第一个卷积核的大小为 11×11,是为了适应 1 000 种图像的多分类问题,让输入的第一个卷积层尽可能包含大的图像特征。而本文改进的 CNN 模型是用于手势识别,手势的特征分布相对区分度较小,所以较小的卷积核能够更好地获取这些特征分布。本文把第一个卷积核大小改为 3×3,并且使用较小的卷积核可以很明显地减少训练参数。

(3) 加深网络深度。AlexNet 结构的卷积核过大导致随着网络层数加深,网络的参数指数型上升,会发生明显的过拟合现象。而本文把卷积核减小了很多,所以可以适当地加深网络的深度,因为卷积层的深度对 CNN 识别准确率有很重要的影响。本文分别在第一个和第二个池化层前增加了一个和前一个卷积层规模大小一样的卷积层,即一共增加了 2 层卷积层。

(4) 更换优化函数。AlexNet 结构使用的是随机梯度下降法(SGD),虽然 SGD 比标准的梯度下降算法在运算速度上有所提高和更不容易收敛到局部最优值,但是由于 SGD 频繁的更新和波动会导致存在一定的超调量。本文使用自适应时刻估计算法(Adam),该算法能计算每个参数的自适应学习率[10]。Adam 可以计算和存储每个参数的对应动量变化,可以有效缓解学习率消失、收敛过慢和损失函数波动较大问题。

表 1　CNN 具体结构参数

名称	卷积核大小	卷积核步长	卷积核个数	填充方式
卷积层 1	3×3	2	64	SAME
卷积层 2	3×3	2	64	SAME
池化层 1	3×3	2	无	VALID
卷积层 3	3×3	1	128	SAME
卷积层 4	3×3	1	128	SAME
池化层 2	3×3	2	无	VALID
卷积层 5	3×3	1	256	SAME
卷积层 6	3×3	1	512	SAME
卷积层 7	3×3	1	512	SAME
池化层 3	3×3	2	无	VALID

本文改进的 CNN 结构的池化层都选择最大值池化方式,再接着 3 个全连接层,每个全连接层都有 1 024 个神经元,dropout 设置为 0.5。在全连接层 3 后面有一个 Softmax 函数,将预测结果分为剪刀、石头、布、GOOD 和 OK 这 5 类手势动作。

2.2 过拟合现象解决

过拟合现象是困扰卷积神经网络模型发展的重要因素。过拟合是指模型在训练集上学习的特征过多,以至于不具有泛化能力,虽然可以在训练集上达到 100% 的正确率,但是在测试集上的正确率却不尽如人意。本文主要通过以下 4 种方法来缓解甚至避免过拟合现象,以提升模型预测的准确率,具体过程与结果如表 2 所示。

表 2 优化过程

优化方法	训练集照片数量	测试集照片数量	优化前预测正确率/%	优化后预测正确率/%
数据增强	2 500	2 250	50	70
Dropout	6 650	2 250	70	85
调整池化	6 650	2 250	85	90
调整参数	6 650	2 250	90	98

(1) 数据增强。数据增强是避免过拟合现象最简单的方法。数据增强有很多方法,比如翻转、平移、水平反射和改变图像 RGB 通道的强度等方法。本文采用最原始的方式,更多地拍摄手势照片,因为一般只用标准手势动作。最开始本文的训练集是每种手势 500 幅照片,即一共只有 2 500 幅照片,模型在测试集上识别正确率只有 50%。当本文的训练集增加到每种手势 1 300 ~ 1 450 幅,一共有 6 650 幅照片时,模型在测试集上识别正确率可以达到 70%。

(2) Dropout 技术。Dropout 是一种简单但非常有效的避免过拟合的技术[11]。Dropout 是指在模型每次随机挑选部分神经元不参与训练,减弱神经元的协同效应,继而让神经元不能依赖其他神经元存在,被迫学习与其他神经元不同的随机子集,获得更健壮的特性。本文在最后三层的全连接层都使用了 Dropout 技术,并且设置随机“脱落”神经元的概率为 50%。

(3) 调整池化方式。池化层的填充有 VALID 和 SAME 两种方式。当池化层的卷积核根据步长移动到图像数据外面时,SAME 是自动在周围填 0 补齐,而 VALID 是忽略,所以相对于 SAME 池化方式,VALID 方式会输入周围更少的特征。重叠池化是指让卷积核移动的步长小于卷积核自身的大小,让图像数据中间的特征更丰富。因为训练集中的手势动作基本都在图像的中间,于是本文调整池化方式为 VALID 和重叠池化。

(4) 选择合适参数。训练批次大小、训练迭代步数、学习率和训练图像大小等参数都会影响模型在测试集上的识别准确率。通过多轮单一变量对比实验发现,在以本文的数据集、模型结构和卷积大小数量为训练背景下,代价函数用交叉熵函数,其形式如下:

$$C = \frac{1}{n}\sum_{x}[y\ln a + (1-y)\ln(1-a)] \quad (1)$$

式中:C 表示代价;x 表示样本;n 表示样本的总数;y 表示实际值;a 表示预测值。

激活函数用非线性 ReLU 函数。因为 ReLU 函数可以通过单侧抑制,使神经网络中的神经元具有稀疏激活性,其形式如下:

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leqslant 0 \end{cases} \quad (2)$$

训练批次大小设置为 64,训练迭代步数设置为 799 步,学习率设置为 0.000 1,训练图像大小压缩为 227 × 227,RGB 图像为输入。

3 机械手控制系统实现

3.1 硬件设计

本文使用的机械手系统如图 3 所示,其动力系统由 6 个 MG996R 舵机组成,可以实现机械手的上下、左右及前后抓取搬运等动作演示。转向关节处均采用的是杯式轴承,可以使转向更加灵活,同时使舵机的转向在同一圆心。底盘采用 4 mm 厚度的铝制圆盘形式,使机械手左右转动更加灵活顺畅并且稳固。机械手控制模块是一个搭载 STM32F103 核心处理器的 6 路舵机控制模块,该模块是一种高效的微伺服电机控制器,可以控制 6 个舵机协同动作。机械手系统硬件的总体设计框图如图 4 所示。STM32 单片机和 PC 机通过串口进行通信,实物设计图如图 5 所示。

图 3 机械手实物图

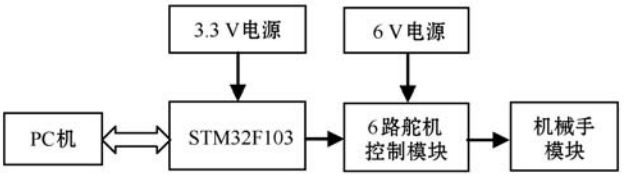

图 4 机械手系统总体设计图

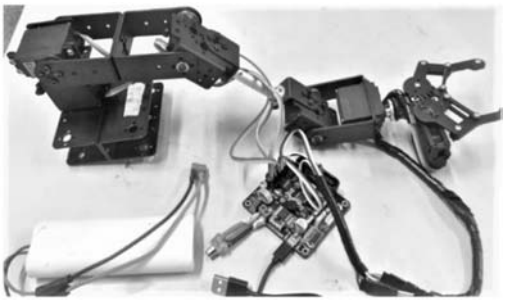

图 5　机械手系统实物图

机械手系统选用 USART 作为 STM32 单片机的串口通信寄存器,通过 USB 转 TTL 设备完成上位机和下位机的通信,该设备的一端连接 STM32 单片机的四个 IO 口,另一端连接电脑的 USB 口。

3.2　软件设计

机械手系统的软件控制总流程图如图 6 所示,STM32 单片机通过串口中断接受 PC 机的手势识别结果,再通过定时器中断控制机械手。

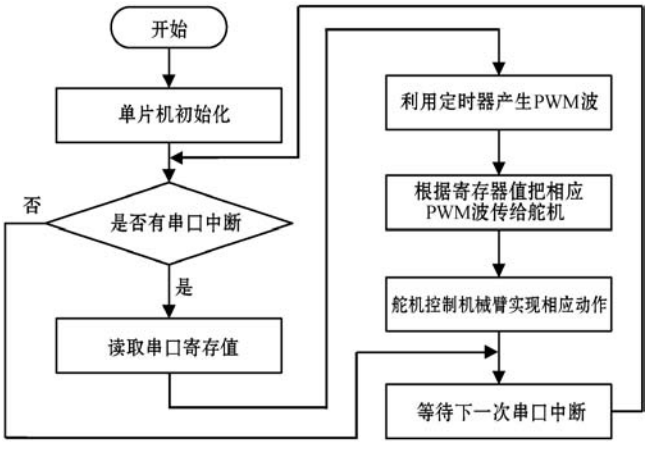

图 6　机械手系统控制流程图

STM32 单片机通过定时器产生周期为 20 ms,即 50 Hz,高电平的脉冲宽度的最小值为 1 ~ 2 ms 的 PWM,具体流程图如图 7 所示。

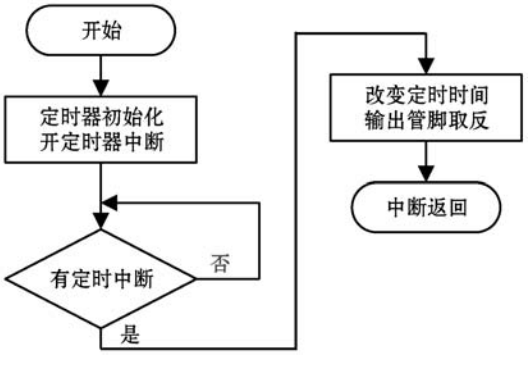

图 7　定时器生成 PWM 波

由于系统一共识别 5 种手势,所以 STM32 单片机存储了 5 个 PWM 波组,每个 PWM 波组都有 6 个脉冲宽度不同的 PWM 波,使机械手系统可以做 5 个不同的动作,每个动作都有 6 个自由度可以设置。

4　实验结果分析

4.1　数据集构建与测试

本文利用 OpenCV 建立数据集,并分为训练集和测试集。图片数据集中有 5 种基本手势:剪刀、石头、布、GOOD 手势和 OK 手势;采集自 9 位大学生志愿者,其中 5 位男性、4 位女性。训练集中 OK 手势图片有 1 450 幅,其他四种手势图片有 1 300 幅;测试集中每种手势的图片数量都为 450 幅。采集的数据集如图 8 和图 9 所示。

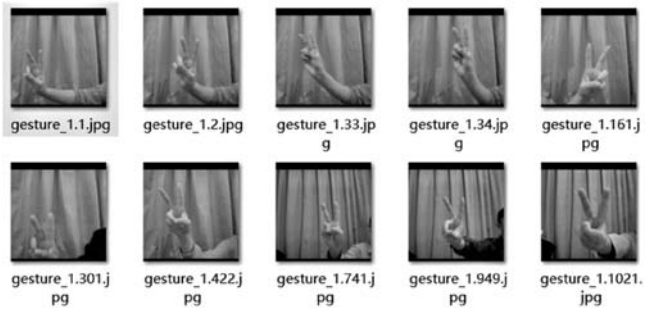

图 8　剪刀手势数据图

图 9　五种手势数据图

拍摄手势的照片是按照手势动作名加上当前帧数命名,这样后期不需要再手动进行标注设置。同时由于只需要识别手势动作,并不需要摄像头拍到的完整画面,于是在保存图像的时候自动设置感兴趣区域的大小为 300 ×300,切割后可以有效地减少拍摄背景等干扰。

本文利用 TensorFlow 搭建的 CNN 模型在训练集训练的效果如图 10 所示,实际训练步数是每个 step 值的 20 倍。

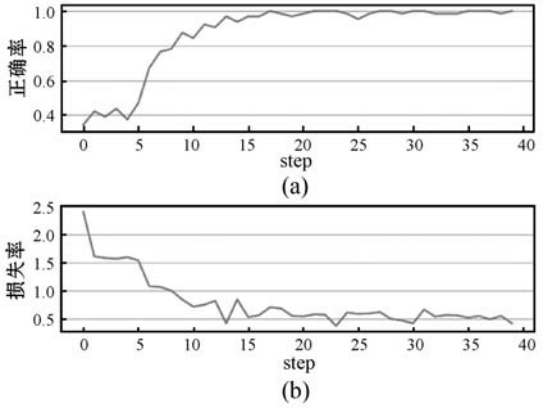

图 10　CNN 在训练集的正确率和损失率

改进后的 CNN 模型在测试集中的平均准确率能达到 98%,且用交叉熵度量的损失率仅为 0.14。

4.2 手势实时测试

手势实时测试是调用保存在 PC 机中预先训练好的 CNN 模型进行在线手势判断。本文在线测试了 5 种手势,在背景为灰色窗帘有人脸干扰的情形下测试手势识别效果如图 11 所示。

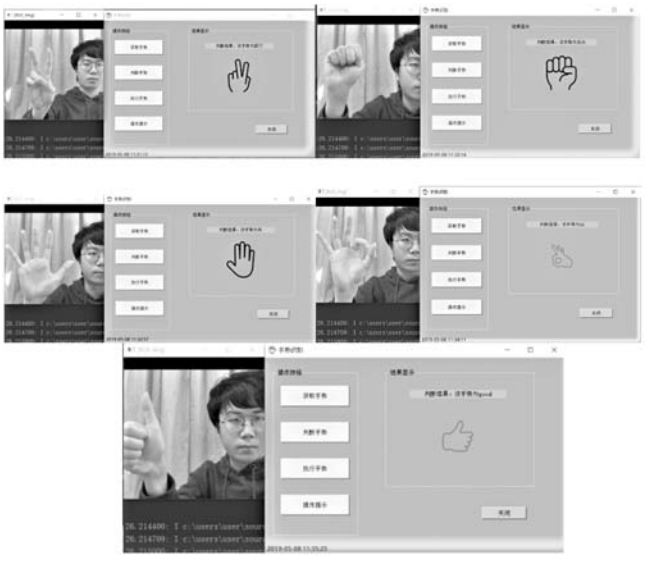

图 11 五种手势实时识别

在手势实时测试过程中每种手势都测试了 10 次,每次 CNN 模型都能正确识别;一位女性测试者在相同的环境进行手势识别测试,她的 5 种手势也都可以被准确识别。实验证明该 CNN 模型在手势识别中具有较高的准确性和较强的泛化性。

4.3 不同模型实验对比

本文用 LeNet-5 结构、AlexNet 结构、VGG-16 结构和改进的 CNN 模型在测试集上预测,得出的实验结果如表 3 所示。结果表明本文提出的改进后的 CNN 模型在手势识别领域比 AlexNet 结构具有更高的识别准确率,能满足手势识别的需求。

表 3 实验结果对比

结构名称	训练集照片数量	测试集照片数量	测试集平均准确率/%
LeNet-5	6 650	2 250	65
AlexNet	6 650	2 250	80
VGG-16	6 650	2 250	95
本文改进的 CNN 模型	6 650	2 250	98

5 结 语

本文提出改进的 CNN 模型的正确率在测试集中可以达到 98%,在背景不是很复杂的情形下(如只有人脸干扰)可以达到 100% 的识别准确率。实验证明该 CNN 模型具有较强的泛化能力,可以很好地完成手势动作识别任务,从而直观地控制机械手做出相应动作。未来可以考虑如下两个方面的改进:

(1) 数据增强。本文的数据集还远不够,要拍摄更多不同背景下的手势动作和更多志愿者的手势动作,并且使用多种数据增强的手段,如:旋转、缩放、膨胀和平移等。

(2) 多结果融合。同时用多个不同的 CNN 模型算出结果概率,然后将这些结果概率取平均得到最大预测结果概率。

参 考 文 献

[1] 张静,周献军,张翔,等. 工业自动化控制的现状和发展趋势分析[J]. 中国新通信,2018,20(20):152.

[2] Yang Q F, Tang H, Zhao X B, et al. Dolphin: Ultrasonic-based gesture recognition on smartphone platform[C]//2014 IEEE 17th International Conference on Computational Science and Engineering. IEEE,2014: 1461 - 1468.

[3] 谢小雨. 基于表面肌电信号和惯性测量单元的手势动作识别的研究[D]. 太原:太原理工大学,2018.

[4] 贺航. 基于 OpenCV 的手势识别的设计改进与实现[D]. 广州:华南理工大学,2018.

[5] 孙玉,袁贞明,孙晓燕. 基于 Leap Motion 的动态手势识别[J]. 计算机工程与应用,2019,55(13):151 - 157.

[6] 朱雯文,叶西宁. 基于卷积神经网络的手势识别算法[J]. 华东理工大学学报(自然科学版),2018,44(2):260 - 269.

[7] 石雨鑫,邓洪敏,郭伟林. 基于混合卷积神经网络的静态手势识别[J]. 计算机科学,2019,46(s1):165 - 168.

[8] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012,1:1097 - 1105.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB]. [2019 - 07 - 10]. arXiv:1409.1556,2014.

[10] Kingma D P, Ba J. Adam: A method for stochastic optimization[C]//International Conference on Learning Representations,2015.

[11] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research,2014,15(1):1929 - 1958.