基于卷积神经网络的手势识别初探①

蔡 娟[1], 蔡坚勇[1,2,3,4], 廖晓东[1,4], 黄海涛[1], 丁侨俊[1]

[1](福建师范大学 光电与信息工程学院, 福州 350007)
[2](福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)
[3](福建师范大学 福建省光子技术重点实验室, 福州 350007)
[4](福建师范大学 智能光电系统工程研究中心, 福州 350007)

摘 要: 提出一种用于手势识别的新算法, 使用卷积神经网络来进行手势的识别. 该算法避免了手势复杂的前期预处理, 可以直接输入原始的手势图像. 卷积神经网络具有局部感知区域、层次结构化、特征抽取和分类过程等特点, 在图像识别领域获得广泛的应用. 试验结果表明, 该方法能识别多种手势, 精度较高且复杂度较小, 具有很好的鲁棒性,也克服传统算法的诸多固有缺点.

关键词: 手势识别; 卷积神经网络; 局部感知; 特征抽取; 鲁棒性

Preliminary Study on Hand Gesture Recognition Based on Convolutional Neural Network

CAI Juan[1], CAI Jian-Yong[1,2,3,4], LIAO Xiao-Dong[1,4], HUANG Hai-Tao[1], DING Qiao-Jun[1]

[1](School of Electronic College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)
[2](Key Laboratory of Optoelectronic Science and Technology for Medicine Ministry of Education, Fujian Normal University, Fuzhou 350007, China)
[3](Fujian Provincial Key Laboratory for Photonics Technology, Fujian Normal University, Fuzhou 350007, China)
[4](Intelligent Optoelectronic Systems Research Centre, Fujian Normal University, Fuzhou 350007, China)

Abstract: The paper proposed a new algorithm used for hand gesture recognition which based on the convolutional neural network. The method not only avoids the hand gesture in the early period of the complex pretreatment, but also can directly input the gesture of original image. The convolutional neural network is characterized by local receptive field, hierarchical structure, global learning for feature extraction and classical. It has been applied to many image recognition tasks. Experimental results showed that the multi-class hand gestures can be recognized with high accuracy, small complexity and good robustness, while the inherent shortcomings of the traditional algorithm are overcame.

Key words: hand gesture recognition; convolutional neural network; local receptive; feature extraction; good robustness

手势是人机交互的方式之一, 通过这种方式可以实现人与机器的自然沟通[1], 手势识别是利用计算机分析每个手势的含义, 进而分析手势发出者的整个表达, 实现人机交互的自然与智能化. 手势识别的研究广泛应用于计算机和用户的交互, 与机器人的人机交互、手语识别等方面. 目前, 国内有关手势识别技术还处于起步阶段, 需要进一步的研究.

通常手势识别采用的方法[2]分为: (1)基于人工神经网络的手势识别. 该方法具有分类特性及抗干扰性, 能够实现复杂的非线性映射, 具有很高的计算速度、很强的容错性和鲁棒性, 目前广泛应用于静态手势识别. 然而由于神经元网络处理时间序列的能力不强, 对于动态手势的识别不够理想. (2)基于隐马尔科夫模型(HMM)的手势识别. 一般拓扑结构下的 HMM 具有非常强的描述手势信号的时空变化能力, 在用于描述与时间相关的随机过程方面具有很大的优势. 但由于要计算大量的状态概率密度, 需要估计的参数个数较多, 使得训练及识别的速度相对较慢. (3)基于几何特征的手势识别技术是利用手势的边缘特征和手势区域特征作为识别特征. 在手势识别中有良好的适应性及

① 收稿时间:2014-07-28;收到修改稿时间:2014-09-29

稳定性. 该方法不足之处在于学习能力不强, 学习效率不高, 随着样本量的不断增大, 其识别率没有提高的很明显.

上述有关于手势识别的算法都具有各自的适用范围. 为了满足实际所需要的实时性和准确性, 本文提出了一种基于卷积神经网络(Convolutional Neural Network, CNN)的手势识别算法.

近年来, 卷积神经网络是广泛应用的一种高效识别方法, 已经成为许多科学领域的研究热点之一, 尤其是在二维图像处理、机器视觉和模式识别中. 卷积神经网络已成功应用手写字符识别[3,4], 人脸识别[5], 人眼检测[6], 行人检测[7], 机器人导航[8]中. 卷积神经网络可以识别有变化的模式, 具有对简单几何变形的鲁棒性. 卷积神经网络在国外的研究较多, 但是目前在国内的研究与应用还刚刚起步.

1　卷积神经网络

在传统的图像识别里, 传统的分类模型如图 1 所示. 首先输入图像, 通过一系列复杂的预处理来对图像进行特征提取, 然后根据提取出来的特征来分类, 并在输出端给出分类的结果. 而卷积神经网络分类模型的不同之处在于可直接将一幅二维图像输入模型中, 并在输出端给出分类的结果, 如图 2 所示. 其中, 特征提取是通过卷积层和降采样层来抽取图像特征. 可有多个卷积层和降采样层来进行自动的特征提取.

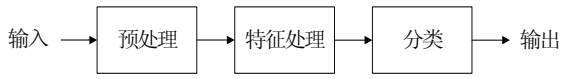

图 1　传统分类模型

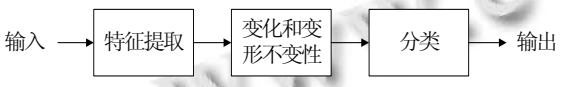

图 2　卷积神经网络模型

卷积神经网络可直接从原始图像中识别视觉模式, 该网络所需的预处理工作量非常少. 卷积神经网络能够识别有变化的模式, 对简单几何变形有一定的鲁棒性, 识别范围广, 因而得到了广泛的应用.

卷积神经网络通过结合三种方法来实现识别位移、缩放和扭曲不变性: 局域感受野、权值共享和降采样[9]. 局域感受野指的是每一网络层的神经元只与上一层的一个小邻域内的神经单元连接, 通过局域感受野, 每个神经元可以提取初级的视觉特征, 如方向线段、端点、角点等. 权值共享使得卷积神经网络具有更少的参数, 需要相对少的训练数据. 降采样可以减小特征的分辨率, 实现对位移、缩放和其它形式扭曲的不变性.

1.1　网络结构

图 3 是一个简单的卷积神经网络结构, 包括输入输出层、卷积层、降采样层和全连接层[10]. 从输入层输入 32×32 大小的图像, 通过交替出现的卷积层(C)、降采样层(S)和最后的全连接层(F), 在输出层给出网络的输出结果.

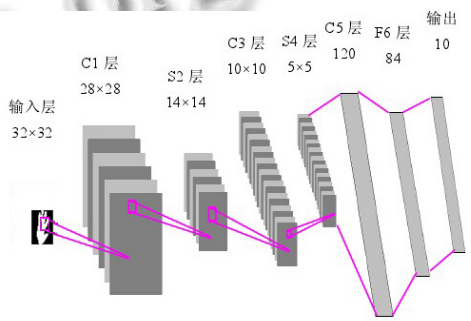

图 3　卷积神经网络的结构

通常, 卷积层后接有一个降采样层目的是减少计算时间和建立空间和结构上的不变性. 卷积层的作用是使原信号的特征增强, 降低噪音. 降采样层的作用是降低特征图的分辨率和网络输出对于位移及变形的敏感程度.

网络层 C1 是由 6 个特征图(见图 8)组成的卷积层. 每个神经元与一个大小为 5×5 的卷积核(又称滤波器模板), 进行卷积. 如图 4 为部分卷积核模板, 也就是滑动窗口. 在 32×32 的矩阵中, 滑动窗口为 5×5, 依次滑动, 根据卷积运算, 由于不考虑边界进行拓展, 卷积核中心滑动的范围只有 28×28, 因此每个特征图的大小为 28×28, 如图 5.

图 4　部分卷积核

网络层 S2 是由 6 个大小为 14×14 的特征图组成的降采样层, 是由 C1 层由 4 个点加权平均为 1 个点得到, 如图 6.

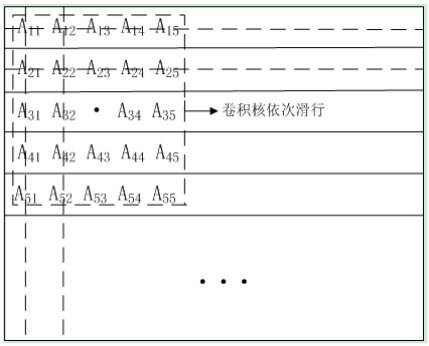

图 5 不考虑边界进行拓展

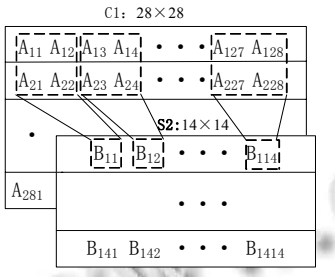

图 6 网络层 S2 大小的由来

网络层 C3 也是卷积层, 是由 16 个大小为 10×10 的特征图组成的. 原理同 C1 层一样. 图 7 显示了 S2 层的特征图如何结合形成C3 的每个特征图, 其中的每一列表示 S2 层的哪些特征图结合形成 C3 的一个特征图. 例如, 从第二列可知, S2 层的第 1 个, 第 2 个, 第 3 个特征图结合得到 C3 的第 1 个特征图.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

图 7 S2 层特征图与 C3 层特征图的连接方法

网络层 S4 同 S2, 为第二个降采样层, 在 C3 的基础上进行降采样, 大小减小到 5×5. 也是起到再一次特征提取的作用.

网络层 C5 对 S4 进行卷积的操作, 采用全连接的方式, 即每个 C5 中的卷积核均在 S4 的所有特征图上都有卷积的操作, C5 层包含了 120 个大小为 1×1 的特征图.

网络层 F6 有 84 个单元, 与 C5 层全相连, F6 层计算输入向量和权重向量之间的点积, 再加上一个偏置, 最终得到一个 1×10 的输出结果, 即输出编码采用“one-of-c”的方式.

关于特征图, 例如图 8 为卷积神经网络分类过程中各层特征图示意图. 输入图像为数字 4, 将原图与 6 个卷积核进行卷积运算, 得到 6 个特征图, 即为网络层C1层, 这些特征图包含了图片通过各个卷积核后所获得的特征. 接着 C1 层经过一个 2×2 到 1 的降采样操作得到 S2 层. S2 层比 C1 层减少了特征图大小, 并且一定程度增强了噪音和轻微扰动的鲁棒性. 卷积神经网络一直重复这个过程, 直到 C5 层, 有 120 个 1×1 的特征图, 将这 120 个特征图通过全连接的方式传播到大小为 10×1 的输出层, 输出的分类结果为数字 4.

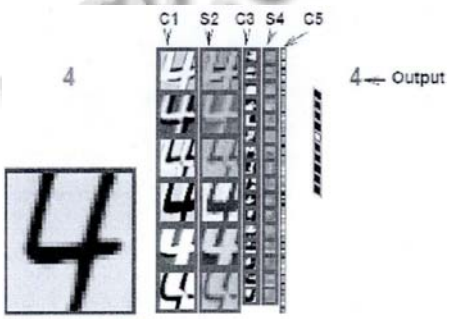

图 8 卷积神经网络分类过程中各层特征图示意图

1.2 卷积层

在卷积层, 前一层的特征图与一个可学习的核(卷积核)进行卷积, 然后通过激活函数, 输出这一层的特征图. 每一个输出的特征图可能包含多个输入图的卷积. 一般地, 卷积层的形式如式(1)[11]所示:

$$x_j^l = f\left(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (1)$$

其中, x_j^l 代表在卷积层 l 层的第 j 个特征图, f(·)表示的是激活函数, k 是卷积核, m_j 表示输入图的集合, 卷积操作*, 以及偏置 b.

1.3 降采样层

降采样层对输入进行抽样操作. 如果输入的特征图为 n 个, 那么输出的特征图也为 n 个, 但是输出的特征图相对于输入图要变小. 降采样层的一般形式如式(2) [11]所示:

$$x_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \quad (2)$$

其中, down(·) 表示降采样函数. 降采样函数一般是对该层输入图像的一个 n×n 大小的区域加权求和, 因此, 输出图像的大小是输入图像大小的 1/n. β 为加权系数, b 为偏置.

2　仿真实验

2.1 手势图像的简单预处理

简单的预处理能够提高整个识别系统的实时性以及准确性. 手势图像的预处理包括手势区域检测、分割和手势的规范化处理.

首先将手势区域检测出来. 将所拍摄的所有彩色图像转为灰度图像, 再利用明显的灰度值差, 选取一个阈值, 进行二值化处理, 手势部分为 1, 背景部分为 0. 本文的手势的规范化即为几何归一化. 将二值化图像按照一定的比例, 保持高宽比, 然后置于一个 120×120 大小区域的中心位置, 其余的像素值补 0. 最后统一缩放到 32×32 的区域内, 以此作为实验的输入数据. 如图 9 所示.

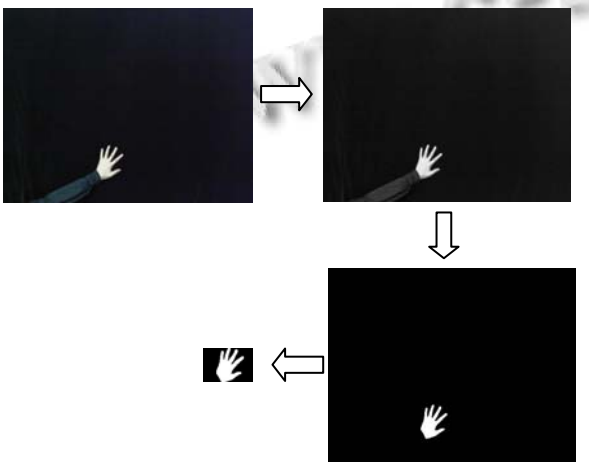

图 9　手势图像简单的预处理

2.2 仿真实验分析

对于本文所讨论的算法, 作者就自己拍摄了一组手势数据库进行试验. 手势数据库分为五种手势, 分别是 1、2、3、4、5, 如图 10 所示. 一共有十个人拍摄五种手势的五段视频, 一段视频为 1 分钟, 将视频格式转化成帧的格式, 并转成灰度图像, 以便下一步有关于手势提取的操作. 拍摄时, 手与摄像头的位置基本不变, 手势可随意摆放, 可旋转, 上下左右任意.

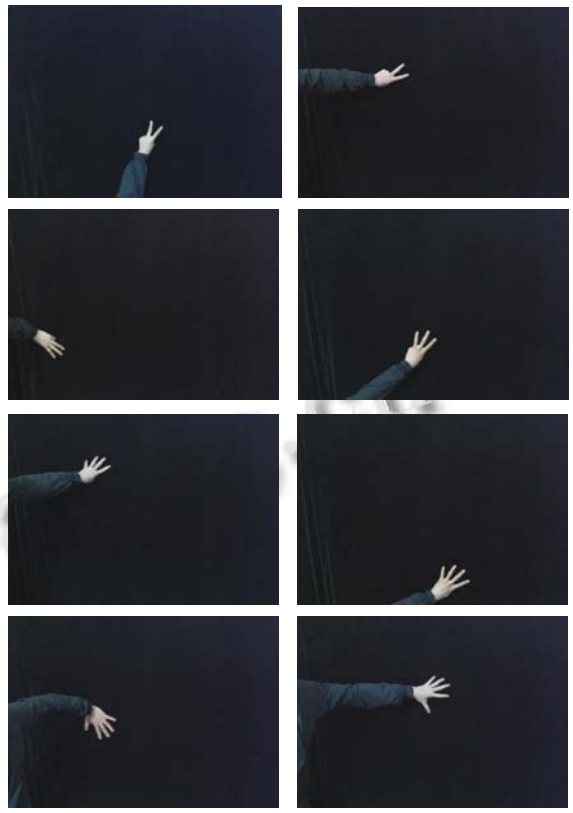

图 10　依次为手势 1、2、3、4、5 的图像

在本实验之前, 本课题找了一个公共手势库, 名称为 Thomas Moeslund 的手势识别数据库, 选取五种手势, 每种手势只有一百张的图片, 测试集和训练集各 50 张, 得到 33%的准确率. 卷积神经网络是建立在大量样本的基础, 故拍摄大量的样本来测试卷积神经网络的算法. 每种手势随机选取了三千个样本, 其中随机的选取测试集 500 个, 训练集 2500 个, 迭代 10 次, 取平均值, 作为该种方法的识别率.

图 11 的横坐标为各类训练样本的个数, 纵坐标表示误识率, 由图我们可获得整体误识率为 3.5%. 经过简单的预处理后, 迭代一次平均所需的时间为 51 秒.

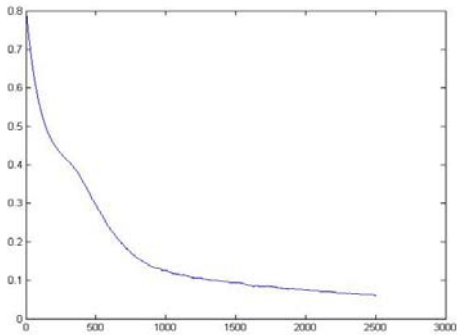

图 11　整体误识率

表 1 为各类手势相应的误识率. 由表 1 知, 识别错误的手势只要集中在手势 3 和手势 4 上. 这主要是用为手势 3 和手势 4 的中指和无名指的距距离过近, 或者是由于角度的关系, 有拇指隔得较近, 训练所用

的样本模式较少, 不能覆盖所有模式, 导致于此类的手势无法识别, 最终得到错误的输出结果. 当然, 只要增加此类手势的训练集, 也会提高该手势的识别.

表 1 各个手势相应的误识率

	1	2	3	4	5
误识率	0	0	8.25%	9.25%	0
准确率	100%	100%	91.75%	90.75%	100%

3 结论

本文对卷积神经网络用于手势识别进行了初步的研究. 鉴于卷积神经网络在处理二维图像时有许多独特的优点, 可以直接将二位图像输入到神经网络中, 大大减少了预处理的难度; 局域野和权值共享技术减少了参数空间, 大幅度降低了算法的复杂度; 降采样技术增强了网络鲁棒性, 能容忍图像一定程度的畸变.

因此采用经典卷积神经网络进行手势识别, 解决了传统识别方法中训练方法复杂度高、训练参数多、耗费时间多等问题, 而且无需对输入手势进行特征提取和分类. 该算法仅需少量的预处理即可对手势进行分类识别.

参考文献

1 张凯.基于立体视觉的自然手势识别[学位论文].北京:北京大学,2005.

2 殷涛.基于几何矩的手势识别算法[学位论文].上海:上海海运学院,2004.

3 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. USA: IEEE, 1998: 2278–2324.

4 Lauer F, Suen CY, Bloch G. A trainable feature extractor for handwritten digit recognition. Pattern Recognition, 2007, 40(6):1816–1824.

5 Lawrence S, Giles CL, Tsoi AC, Back AD. Face recognition: A convolutional neural network approach. IEEE Trans. on Neural Networks, 1997, 8(1): 98–113.

6 Tivive FHC, Bouzerdoum A. An eye feature detector based on convolutional neural network. Proc. 8th Int. Symp. Signal Process. Applic. Sydney, New South Wales, Australia. IEEE, 2005: 90–93.

7 Mate S, Akira Y, Munetaka Y, Jun O. Pedestrian detection with convolutional neural networks. IEEE Intelligent Vehicles Symposium Proceedings. USA: IEEE, 2005: 224–229.

8 Cun YL, Muller U, Ben J, Cosatto E, Flepp B. Off-road obstacle avoidance through end-to-end learning. Advances in Neural Information Processing Systems. USA: MIT Press, 2005.

9 赵志宏,杨绍普,马增强.基于卷积神经网络 LeNet-5 的车牌字符识别研究.系统仿真学报,2010,22(3):638–641.

10 徐姗姗,刘应安,徐昇.基于卷积圣经网络的木材缺陷识别.山东大学学报(工学版),2013,43(2):23–28.

11 许可.卷积神经网络在图像识别上的应用研究[学位论文].杭州:浙江大学,2012.