

基于卷积神经网络的手写体数字识别系统

陈 岩^{1,2}, 李洋洋¹, 余 乐^{1,2}, 王 瑶¹, 吴 超¹, 李阳光¹

(1 北京工商大学 计算机与信息工程学院, 北京 100048;

2 食品安全大数据技术北京市重点实验室, 北京 100048)

摘 要: 近年来, 卷积神经网络在图像的分类识别领域取得成功, 并逐渐在许多嵌入式终端设备上得到应用. 在 Linux 环境下采用 QT 开发框架, 设计实现了一个基于卷积神经网络的手写体数字识别系统. 该系统可以很方便的跨平台移植于各类嵌入式设备上. 测试结果表明, 该系统对手写体识别具有良好的识别效果.

关键词: 卷积神经网络; 手写体数字; Linux; QT

中图分类号: TP391

文献标识码: A

文章编号: 1000—7180(2018)02—0071—04

A System of Convolutional Neural Networks Based Handwritten Number Recognition

CHEN Yan^{1,2}, LI Yang-yang¹, YU Le^{1,2}, WANG Yao¹, WU Chao¹, LI Yang-guang¹

(1 School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China; 2 Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China)

Abstract: In recent years, Convolutional Neural Network (CNN) has been successfully used in the field of image classification and recognition. In this paper, based on the QT development framework, we designed and implemented a handwritten numeral recognition system based on CNN in Linux environment. The proposed system can be easily transplanted on various embedded devices. The test results show that the system has good recognition effect.

Key words: convolutional neural networks; handwritten number; Linux; QT

1 引言

近年来, 卷积神经网络(Convolutional Neutral Network, CNN)已经被广泛应用到图像识别领域, 比如: 人脸识别^[1]、基于航拍图像的目标检测^[2]. 随着计算机性能的提升和卷积神经网络结构的不断改进, 卷积神经网络在图像分类领域取得了突破性的进展. 与传统的模式识别方法不同^[3-4], CNN 不需要对图片进行复杂的预处理或者人工确定某些特征作为识别依据, 它可以将原始图片直接做为网络的输入, 自动提取特征, 再根据提取到的特征对图片进行分类, 而且在同等条件下相比传统方法具有显著的

准确度提升.

本文在 Linux 系统中采用 QT 设计了一个基于 CNN 的手写体数字识别系统. 该系统可实现手动输入数字 0~9、显示 CNN 中间特征提取的结果以及显示最终的识别结果. 测试结果表明该系统对手写体数字具有良好的识别效果.

2 卷积神经网络

卷积神经网络是一种带有卷积运算的神经网络, 采用局部感知域和权值共享的方法^[5], 大大减少了网络的参数的个数, 不仅降低网络模型的复杂程度, 而且缓解了网络模型过拟合的问题. 卷积操作对

平移、比例缩放以及其他形式的变形具有一定不变性。

典型的卷积神经网络主要由输入层、卷积层、池化层和全连接输出层构成,其结构图如图 1 所示。

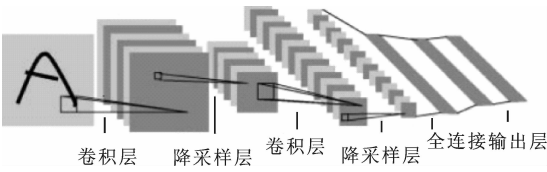


图 1 典型的卷积神经网络结构图

2.1 卷积层

对于卷积层,用一个参数可调的卷积核与上一层特征图(Feature map)进行滑动卷积运算再加上一个偏置量得到净输出,如式(1)所示,最后通过一个激活函数(Activation function)就能得到卷积的结果,即输出特征图,如式(2)所示。

$$u_j^l = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l \tag{1}$$

$$x_j^l = f(u_j^l) \tag{2}$$

式中, x_i^{l-1} 是 $l-1$ 层特征图被第 j 个卷积所覆盖的元素; k_{ij}^l 是 l 层卷积核中的元素; b_j^l 是第 j 个卷积结果的偏置量; M_j 是第 j 个卷积核覆盖的区域; u_j^l 是 l 层第 j 个卷积的净输出; $f(\cdot)$ 是激活函数,通常采用 Sigmoid; x_j^l 是 l 层第 j 个卷积操作的输出。

2.2 降采样层

对于降采样层,它把输入的特征图用窗口划分成多个不重叠的 $n \times n$ 图像块,再对每个图像块采用最大值法或者均值法,从而使得输出图像缩小了 n 倍。加上偏置量后,最后通过激活函数得到输出。最大值法如公式(3),均值法如公式(4),输出函数如公式(5)所示。

$$u_j = \max_{i \in N_j} (x_1^l, \dots, x_i^l) \tag{3}$$

$$u_j = \frac{1}{n \times n} \sum_{i \in N_j} x_i^l \tag{4}$$

$$x_j^{l+1} = f(u_j^l + b_j) \tag{5}$$

式中, N_j 是第 j 个 $n \times n$ 的图像块; x_i^l 是 N_j 图像块所包含的元素; b_j 是 N_j 图像块的偏置量;激活函数 $f(\cdot)$ 通常采用 Sigmoid; x_j^{l+1} 是 l 层 N_j 图像块降采样的输出,也是 $l+1$ 层特征图的某个元素。

2.3 全连接输出层

经过卷积和降采样后,原始图片的高级特征已经被提取出来了。全连接层的作用就是使用这些特征对原始图片进行分类。全连接层的输出可以通过对这些输入的特征做加权求和,再加上偏置量,最后

通过激活函数获得最终的输出,如式(6) 所示。

$$y^l = f(w^l x^{l-1} + b^l) \tag{6}$$

式中, x^{l-1} 是前一层的输出特征图,里面的元素都通过卷积和降采样提取出来的高级特征; w^l 是全连接层的权重系数; b^l 是全连接层 l 的偏置量。

3 系统设计与实现

本文基于 QT 开发手写体字符识别系统,其中应用程序交互界面如图 2 所示。Hand Input 区域可以拖动鼠标画出数字 0~9,CNN Forward-propagation 区域显示卷积层、降采样层、全连接层各自运算后的结果,Recognition Result 区域显示系统最终的识别结果。

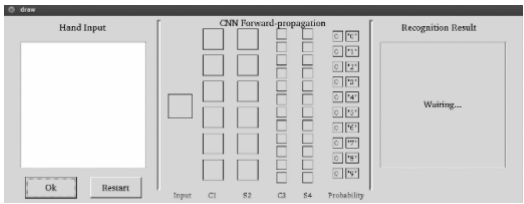


图 2 应用程序交互界面

3.1 数字输入

数字输入采用鼠标绘制的方式以实现,绘图区窗口大小为 $288 \times 288 \text{ pixel}$,手绘数字保存成图片格式为 BMP,记做 P_{original} ,以供后续处理。

3.2 预处理

由于真彩色图片包含信息量大,处理速度较慢,图片的颜色很容易受光照等因素的影响,导致同类物体的颜色有很多种变化,所以彩色图片本身难以提供关键信息。本文中手写体数字的位置、倾斜度等因素对 CNN 识别的准确度也有影响,综上所述,待识别图片的预处理是必不可少的。

预处理流程图如图 3 所示。第一,读取 P_{original} 的文件头、信息头和位图数据并保存

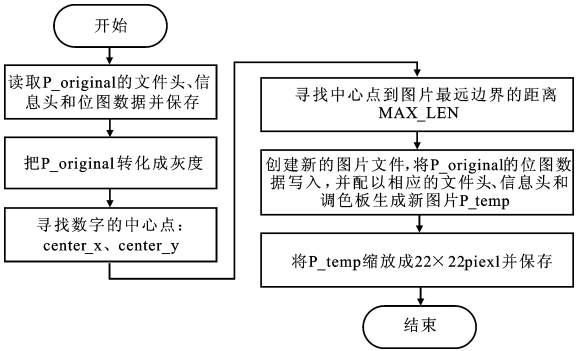


图 3 图片预处理流程图

存;第二,把 P_original 转化成灰度,以降低计算量,并且让关键信息(如:边缘信息)更容易被提取;第三,寻找手写数字的中心点以及中心点到图片最远边界的距离 MAX_LEN;第四,创建新的图片文件,并将之前保存的位图信息写入文件并配上相应的文件头,信息头等,产生新的图片(信息头中位图的长宽均为 $2\times\text{MAX_LEN}$),记做 P_temp,其目的是将字符移至图片中央.由于 MNIST 字符集中的图片都是 28×28 pixel,每张图片的上下左右三行的灰度值都很大,使得图片相应的部分接近白色.最后,将新的图片转化成灰度,并缩放成 22×22 pixel,记做 P_final.通过上述预处理以提高 CNN 识别的准确率,经过预处理后的图片如图 4 所示.

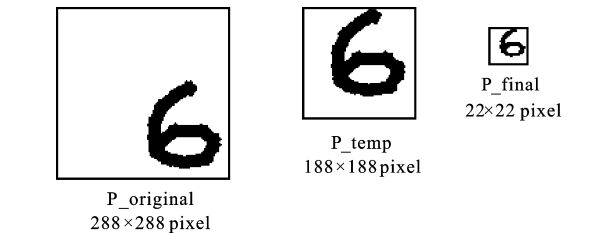


图 4 预处理前后图片对比

3.3 CNN 参数训练

本文采用 LeNet-5 模型的卷积神经网络对手写体进行识别,使用 MNIST 数据集(70 000 张图片)对网络进行训练,其中 60 000 张做训练集,10 000 张做测试集. MNIST 数据集中训练集格式如表 1 所示,测试集格式如表 2 所示.

表 1 图像文件的数据格式

偏移量	数据类型	值	描述
0000	32 bit integer	2051	幻数
0004	32 bit integer	60000	图片数量
0008	32 bit integer	28	图片行数
00012	32 bit integer	28	图片列数
00016	unsigned byte		像素值
00017	unsigned byte		像素值
.....			
XXXX	unsigned byte		像素值

注:像素的值为 0~255.

表 2 标签文件的数据格式

偏移量	数据类型	值	描述
0000	32 bit integer	2049	幻数
0004	32 bit integer	10000	图片数量
0008	unsigned byte		标签值
0009	unsigned byte		标签值
.....			
XXXX	unsigned byte		标签值

注:标签值为 0~9.

网络的训练与测试均在 Matlab 中进行并保存训练好的权重.如图 5 所示,随着训练次数的不断增加,误差越来越小,最终误差为 3.32%.最后将网络的前向识别过程在 QT 中用 C++实现,并载入训练好的权重,从而对手写体数字进行识别.

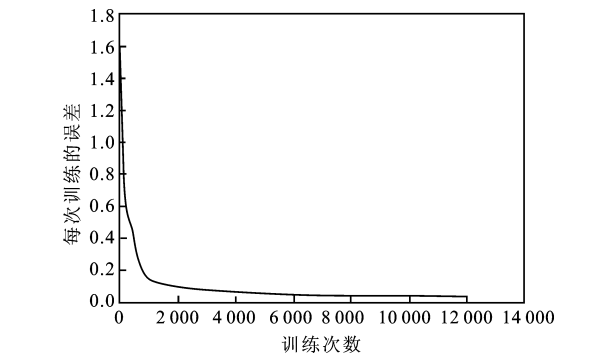


图 5 CNN 训练次数与误差的关系

3.4 LeNet-5 模型介绍

LeNet-5 模型的卷积神经网络共有 6 层,网络结构如图 6 所示.

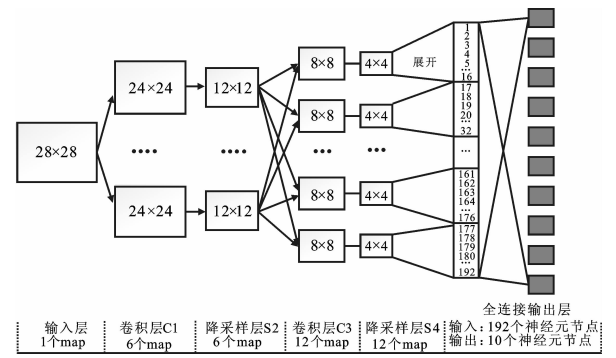


图 6 LeNet-5 模型的网络结构

输入层:MNISTS 数据集中每张图片都是 28×28 pixel,即一个 28×28 的矩阵.所以,首先要读取 P_final 的位图数据得到一个 $22\times 22\times 3$ 的矩阵,接着对每个像素点进行灰度处理得到一个 22×22 的矩阵,最后用 0 对矩阵上下左右三行进行填充得到一个 28×28 的矩阵作为 CNN 的输入.

卷积层:卷积运算的本质是每个像素点的加权求和,再加上一个偏置量,最后求 sigmoid 获得,卷积核是一个 5×5 的矩阵,步长为 1.卷积层的 map 的大小由卷积核和上一层输入 map 的大小决定,假设上一层的 map 大小是 $n\times n$ 、卷积核的大小是 $k\times k$,则该层的 map 大小是 $(n-k+1)\times (n-k+1)$.卷积操作最主要的任务就是提取图片的特征信息.

降采样层:将采样的窗口大小为 2×2 ,步长为 2,每个元素值都是 1/4,将前一层 map 中的值与

窗口内对应位置值相乘并求和,即得到对应位置的输出元素值,相当于对窗口内元素做均值运算.降采样操作在保持图片重要信息的同时降低了图片的维度,减少了运算量.

全连接层:在全连接层的输出层采用 Softmax 激活函数确让识别结果以概率的形式输出,本文最终提取了原始图片 192 个特征,有 0~9 共 10 个类别.

3.5 实验结果

实验结果如图 7 所示,在 Hand Input 输入区用鼠标写数字“6”,CNN 计算图片上的数字是“6”的概率最大(94%),故判定图片上是数字“6”并在 Recognition Result 区域显示. CNN Forward-propagation 区域显示的是经预处理后的原始图片和卷积层、降采样层、全连接层各自计算后的结果.

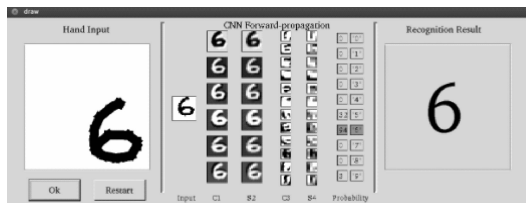


图 7 实验结果

4 结束语

本系统先将手绘数字转成图片,再采用卷积神经网络的进行识别处理.这样在保证准确识别的情况下,避免了复杂的预处理和人为的特征提取过程.

本设计中可以对数字的位置与大小进行平移、缩放校正,位置和大小偏差对于网络识别精度的影响很小.综上所述,本文设计的系统对手写体数字有良好的识别效果.

参考文献:

- [1] Kim Y H, Kim H, Kim S W. Illumination normalisation using convolutional neural network with application to face recognition [J]. Electronics Letters, 2017, 53(6):399-401.
- [2] Igor Ševo, Aleksej Avramović. Convolutional neural network based automatic object detection on aerial images [J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(5):740-744.
- [3] 张伟,王克俭,秦臻.基于神经网络的数字识别的研究[J].微电子学与计算机,2006,23(8):206-208.
- [4] 申家振,张艳宁,刘涛.基于形状上下文的形状匹配[J].微电子学与计算机,2005,22(4):144-146.
- [5] 丁小刚. BP 神经网络与卷积神经网络在文字识别中的应用研究[D]. 武汉:华中科技大学,2014.

作者简介:

陈 岩 女,(1963-),博士,教授.研究方向为无线传感器网络技术应用、数字通信抗干扰技术.

余 乐(通讯作者) 男,(1983-),博士,讲师.研究方向为类脑计算机与人工智能. E-mail:ladd_u@163.com.

(上接第 70 页)

- [2] 张琰,李吴松,张荆沙,等.物联网环境下基于 PLC 的远程控制系统设计[J]. 微电子学与计算机, 2016, 33(2):130-134.
- [3] Li F, Yang J, Wu J, et al. Configuration analysis and recommendation: Case studies in IPv6 networks [J]. Computer Communications, 2014, 53(53):37-51.
- [4] 贾锐,蔡皖东. Linux 平台双协议栈主机网络管控系统设计与实现[J]. 微电子学与计算机, 2016, 33(6): 14-17.
- [5] 阎军,咸爱国. 一种低成本串口服务器的设计[J]. 电子设计工程, 2014, 21(14):190-192.
- [6] 吴荣斌. 基于 ARM9 的以太网串口服务器的设计与实现[D]. 兰州:兰州交通大学, 2014.
- [7] 张从鹏,赵康康. 基于 STM32 的串口服务器系统开

发[J]. 仪表技术与传感器, 2016, 52(1):73-75.

- [8] 崔二勇,简献忠,蔡晶晶,等. 基于 LPC2103 的串口服务器设计与实现[J]. 电子世界, 2014, 35(7): 161-162.
- [9] 孙艳琴. 基于串口服务器的水温远程通信系统设计与实现[J]. 无线互联科技, 2016, 22(11):1-3.

作者简介:

刘姝黎 男,(1991-),硕士研究生.研究方向为智能仪器、目标跟踪.

史健芳(通讯作者) 女,(1966-),博士,教授.研究方向为智能仪器及检测技术、智能信息处理、目标跟踪.

E-mail:sxtyshijianfang@163.com.

杨 静 男,(1970-),博士,副教授.研究方向为信息融合.