

# 基于神经网络的准考证号码识别研究

邓立

**摘要：**本文对卷积神经网络（CNN）以及 BP 神经网络进行了简单介绍，比较了在准考证号码识别的应用场合下，BP 神经网络与卷积神经网络的识别准确率。针对准考证号码识别这一特定场景，对 NMIST 训练数据进行有针对性的数据增强，实验证明能有效提高识别准确率。本文根据准考证号码识别的应用场景提出了两种图像预处理方法，并通过实验验证了该方法对识别准确率的提升效果。最后，本文分析了两种网络模型的参数大小以及所需运算量。

**关键词：**卷积神经网络、BP 神经网络、准考证号识别、字符识别

## Study of admission ticket number recognition system based on neural network

DENG LI

**Abstract:** This paper briefly introduces the convolutional neural network and BP neural network, and compares the recognition accuracy of BP neural network and convolutional neural network in the application of admission ticket number identification. For the specific scene of the admission ticket number identification, The targeted data enhancement of NMIST training data has been carried out. The experiment proves that the recognition accuracy can be effectively improved. In this paper, two image preprocessing methods are proposed according to the application scenario of the admission ticket number identification, and the accuracy of the method is verified by experiments. Finally, this paper analyzes the parameter size of the two network models and the required amount of computation.

**Keyword:** Convolution neural network; BP neural network; Handwritten number recognition; Character Recognition;

### 0 引言

在日常考试中，大部分试卷都要求考生填写准考证号码以识别考生身份，阅卷老师需要输入准考证号码以录入考生成绩。然而，通过人工的方式录入考生准考证号不仅效率低下而且还耗费改卷老师大量的时间和精力。如果能够将机器视觉应用到准考证号识别上能极大地方便阅卷老师登记成绩。传统的数字识别方法是 OCR 和模板匹配，但由于手写数字的笔迹因人而异，OCR 与模板匹配只适合识别规范的印刷数字，并不适合用于识别手写数字<sup>[1]</sup>。对手写数字的研究有基于特征提取的方式与基于分类器决策两种方式，基于特征提取的手写数字识别有点，线，连通域等基本特征，通过抽取手写数字的骨架和轮廓，提取其几何特征，如：矩、傅里叶描绘子、样条曲线拟合等几何特征，可用于反映手写数字的细节<sup>[2]</sup>，但这种方法识别率并不高。

近年来,得益于 GPU 芯片性能的飞速发展,越来越多的学者使用人工神经网络进行模式识别及分类,本质上,手写数字的识别问题可以归为图像分类问题。基于神经网络的手写数字识别得到越来越多研究者的关注。而 BP 神经网络(Back Propagation Neural Networks)具有良好的容错能力、强大的分类能力、自适应和自学习等特点,备受人们的重视,在字符识别领域得到了广泛的应用<sup>[3]</sup>。卷积神经网络在图像分类中表现优异,相比较其他深度、前馈神经网络,卷积神经网络需要考量的参数更少,使之成为一种颇具吸引力的深度学习结构。

本文通过在真实的手写准考证测试样本下研究对比两种神经网络的识别准确率。

## 1 系统的总体设计方案

用神经网络实现准考证识别有两种方法,一种是先根据先验知识,人工选择某些特征并对待识别数字进行特征提取,然后将提取出的特征输入到 BP 神经网络或支持向量机(SVM)进行训练,这种方法的识别效果与人工挑选特征的好坏有关,且需要人工选取特征,检测性能很大程度上取决于人工所选用的特征是否准确,现在已经很少采用。另一种方法是利用神经网络的学习特性,直接将待识别图像输入到神经网络中,让神经网络的输出经过若干次训练后收敛于理想值,这种方式不需要人为指定特征,容错性更强,故本文采用第 2 种方法识别手写准考证号。

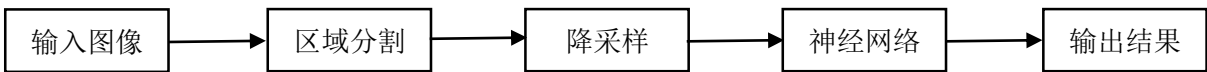


图 1 系统原理图

## 2 本文采用的实验数据

为了使神经网络具备对任意手写字体的泛化能力,在选择训练数据时应考虑识别对象的特性,尽量选用样本较多的数据集,本文选用 NMIST 手写数字训练集。MNIST 数据集来自美国国家标准与技术研究所, National Institute of Standards and Technology (NIST)。训练集由来自 250 个不同人手写的数字构成,其中 50% 是高中学生,50% 来自人口普查局 (the Census Bureau) 的工作人员。共 60000 张,每张大小为 28\*28 像素的灰度图。



图 2 NMIST 手写数字集

### 2.1 数据增强方法

在实际的准考证识别情况下，考生可能会因为粗心大意把数字填歪，填写的数字大小不一或者偏离中心等情况。因此，为了使识别系统的在个别异常笔迹下仍能较好的识别出数字，本文采用数据增强(data augmentation)的方法对上述情况进行模拟，如此不仅更贴合准考证号识别的应用场景，也可以获得更多样本以供训练。

针对上述在实际中可能会出现的情况，对训练集进行如下扩充:

- 1) 对训练样本随机进行缩放(0.8-1.2 倍)，模拟字迹大小不一的情况。
- 2) 对训练样本随机加入小角度旋转(15 度)，模拟数字歪扭的情况
- 3) 对训练样本随机加入位置扰动，模拟数字出格或偏离中心的情况。

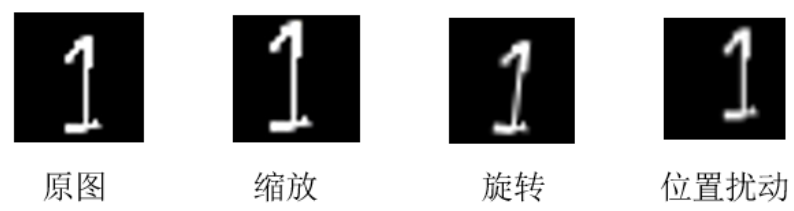


图 3 数据增强示意图

2.2 测试数据集

由于试卷上用于填写准考证号的位置固定，并且在附近标有用于方便机器识别的标记。因此，对手写字符的定位和分割非常简单,进行分割后得到每个手写支付的图像块(patch)，然后送入神经网络进行识别。不失一般性地，本文采用以下扫描图片进行测试。

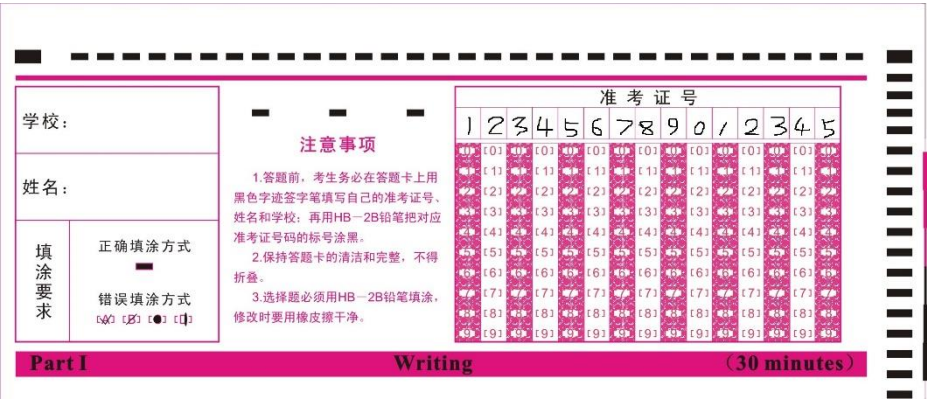


图 4 用于测试的图片

由于准考证号填写在方框内，因此只需事先设定好 ROI 区域，根据读入图片相对位置便可找出每个方格对应数字。由于训练样本是黑底白字而测试样本的白底黑字，找好 ROI 后还需将其进行图片反色，再对其进行大小压缩、归一化处理，将处理后的图片送进神经网络进行识别<sup>[4]</sup>。

为了尽可能模拟准考证号码识别的应用场合，笔者制作了 1083 个笔迹潦草的手写数字用于测试评估。其中包括字体偏离方框中心，字体出格，字体大小不一等情况。

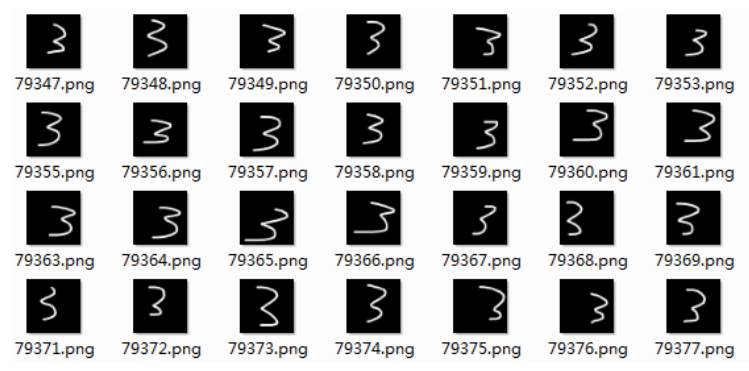


图 5 部分用于测试的准考证数字

### 3 手写数字识别神经网络

#### 3.1 BP 神经网络

BP 网络能学习和存贮大量的输入-输出模式映射关系，而无需事前揭示描述这种映射关系的数学方程。。BP 神经网络模型拓扑结构包括输入层（input）、隐含层(hidden layer)和输出层(output layer)。典型的一层 BP 神经网络结构如图 6 所示。

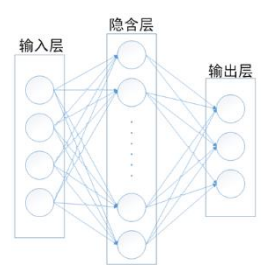


图 6 BP 神经网络示意图

已有理论证明，含有一个隐含层的 BP 神经网络可以实现对连续的非线性函数进行任意精度的拟合，因此采用一个隐含层的 BP 神经网络作为训练网络。采用 NMIST 手写数字训练集，共 60000 张，每张大小为 28\*28 像素的灰度图。因此，输入层的节点数选择为 784。

对于输出层结点，由于只需识别 0~9 的手写数字，输出状态较少，可直接使用 10 个输出层结点，每个输出层结点输出值对应为该数字的可能性。由此可知，手写体数字 0 的正确输出应该为(1, 0, 0, 0, 0, 0, 0, 0, 0, 0)<sup>T</sup>。要预测手写体数字，只需找出值最大的输出节点即可。

对于隐含层节点数的选择，并没有理论公式指导，本文隐含层选用 200 个神经元。

在送入神经网络前，先将图片的灰度值统一进行归一化。由于 BP 神经网络会自动收敛，因此随机初始化输入层权重以及隐含层权重。

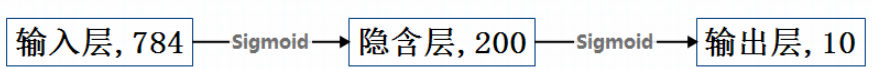


图 7 BP 神经网络结构简图

采用批量梯度下降法更新权值，每批 200 个样本，迭代 500 次，即 mini\_batch=200，epoch=500，采用 Adam 优化算法。

3.2 卷积神经网络

卷积神经网络（Convolutional Neural Network，CNN）是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元<sup>[5]</sup>，对于大型图像处理有出色表现。卷积神经网络由一个或多个卷积层和顶端的全连通层（对应经典的神经网络）组成，同时也包括关联权重和池化层（pooling layer）。这一结构使得卷积神经网络能够利用图像的二维结构<sup>[6]</sup>。

本文采用的卷积神经网络结构如下图所示：

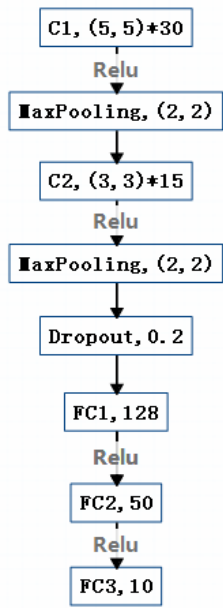


图 8 本文所用卷积神经网络结构图

注意到，不同于 BP 神经网络，卷积神经网络对 28x28 的输入图像进行局部的卷积操作，C1 层采用 30 个 5x5 大小的卷积核用于特征提取，得到 30 通道，23x23 的特征图后进行最大值池化降采样为 30 通道，11x11 的特征图，再送入 C2 层进行卷积得到 15 通道 8x8 的特征，再次最大值池化为 15 通道 4x4 的特征，传入 Dropout 层，以 20% 的概率随机断开输入神经元，展开送入全连接层 FC1，FC2，FC3，最后用 softmax<sup>[7]</sup> 输出对应数字概率。

同样地，batch\_size=200，epoch=500，采用 Adam 优化算法

3.3 对比测试

网络结构	有否使用数据增强	NMIST 测试集准确率	准考证号码测试集准确率
BP	无	98.28%	45.52%
CNN	无	99.30%	71.01%
BP	有	97.51%	87.35%

CNN	有	99.22%	96.68%
-----	---	--------	--------

表 1 不同实验下的识别准确率

测试发现，卷积神经网络的识别准确率普遍高于 BP 神经网络。而且，有针对性地根据场景进行数据增强能够大幅提升神经网络的识别准确率，其中使用数据增强对 BP 神经网络的识别准确率有很大的提升。

4 改进提升

在上一章节中，我们注意到两种神经网络的准考证号码识别准确率相比 NMIST 准确率要低得多，通过观察截取的准考证号码，我们注意到以下特点：

- 1. 准考证号码中的字体位置不固定
- 2. 准考证号码中的字体大小不一

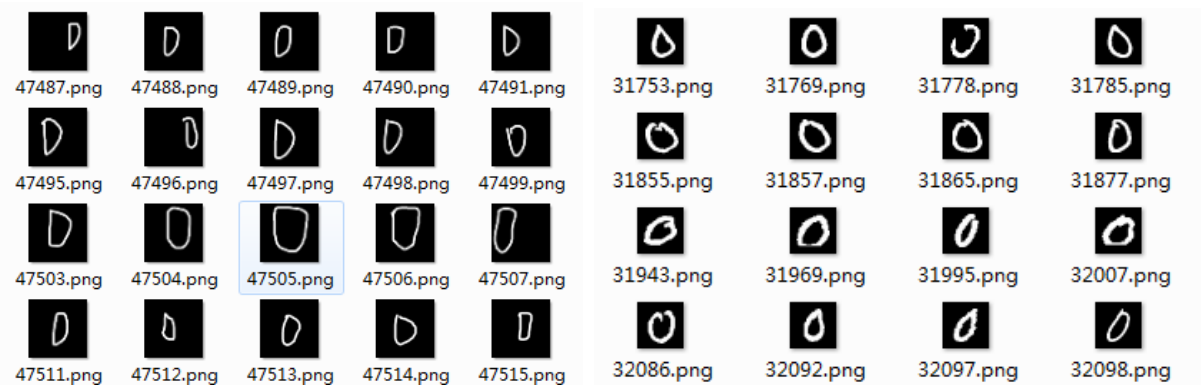


图 9 准考证号码(左)与 NMIST 测试集(右)

根据上述观察，我们尝试对输入的准考证号码先进行某些预处理，然后再送入神经网络。

4.1 字体居中

由于考生在填写准考证数字时，位置未必处于图像中央，而 NMIST 训练集的数字往往位于图像中央。因此，在把图片送入神经网络检测前，先求字体重心，然后根据重心整体平移字体到图片中央。

4.2 字体尺度调整

填写准考证号码的数字大小，不同考生往往有不同的习惯，因此填写数字有可能大也有可能小。但 NMIST 训练集中的数字较为均匀。为了确保准考证数字在图片中的尺度，在进行字体居中后，将字体方框的最长边等比例缩放至 200 像素并居中（整副图像大小为 280x280）。最后统一缩放至 28x28。



图 10 预处理前(左)与预处理后(右)

4.3 实验结果

使用训练好的模型进行测试，结果如下：

网络结构	预处理方法	准考证数字识别准确率	准考证数字识别准确率
------	-------	------------	------------

		(使用数据不增强模型)	(使用数据增强模型)
CNN	无	71.01%	96.68%
CNN	居中	95.20%	96.49%
CNN	调整尺度	92.89%	97.14%
BP	无	45.52%	87.35%
BP	居中	86.52%	89.75%
BP	调整尺度	70.54%	92.89%

表 2 实验结果

可以看到，识别效果最好的模型是使用数据增强训练 CNN 模型+调整尺度，在此情况下识别率达到了 97.14%。在未经数据增强训练出的网络模型下，使用数据增强能够大幅提升其识别准确率。同时注意到，由于经过数据增强训练出的模型已经对数字的位置扰动、尺度等有了一定的抗干扰性，因此使用数据预处理对经数据增强训练的模型提升有限。

#### 4.4 模型大小及运算量分析

在本文中，假设所有权值以 float32 的数据格式存储，即每个权重占 4 字节。通过计算，我们可以得出 BP 神经网络的参数占用空间为： $(784*200+200*10)*4=620\text{kb}$ 。而本文所用卷积神经网络的参数占用空间为： $(5*5*30+30*3*3*15+15*4*4*128+128*50+50*10)*4=165\text{kb}$ 。卷积神经网络的“权值共享”特性使其参数个数大大减少，约为 BP 神经网络的四分之一。

在计算量方面，本文所用 BP 神经网络只需进行  $784*200*10=15680000$  次浮点运算即可，而本文所用卷积神经网络则需进行  $1*5*5*30*23*23+30*3*3*15*8*8+15*4*4*128*50*10=16015950$  次浮点运算，与 BP 神经网络接近，考虑到两者在识别准确率上的表现，应该选用卷积神经网络

综上，对准考证号码识别来说，应用卷积神经网络在各方面均要优于 BP 神经网络。

#### 5 讨论

BP 神经网络自 1986 年问世以来，基于神经网络的研究进度日新月异，在其之后的经典分类网络有 LeNet<sup>[8]</sup>，AlexNet，GoogLeNet 等，但权值更新过程中的梯度反向传递原理仍与 BP 神经网络大同小异。BP 神经网络具备的强大学习特性令其至今仍用于最新的深度学习网络中，但在实验中，我们也注意到 BP 神经网络在训练中容易对训练数据“过拟合”，具体表现为对识别数字的位置以及大小非常敏感。此外，BP 神经网络的“全连接”特性，使得其更容易对局部值的变化敏感。因此不难解释，为什么 BP 神经网络常常用于对深度神经网络提取到的高层抽象特征进行学习。这就不难理解为什么卷积神经网络在应用全连接层之前先使用卷积层进行特征提取，因为卷积层提取到的是局部区域的特征<sup>[9]</sup>，每个像素点只对与其进行卷积的锚点有影响，从而避免了过拟合。另外，卷积神经网络添加了用于降采样的 MaxPooling 层，只保留响应值最强的值，减少了噪音，因而有着比 BP 神经网络更好的泛化能力。要令 BP 神经网络减少对数据过拟

合，提高泛化能力，可以在代价函数中加上权重约束项，惩罚过于复杂的网络，也可以参考 AlexNet<sup>[10]</sup> 的”Dropout”，以 50% 的概率随机屏蔽神经元的输出，减少神经元之间相互依赖程度，防止出现神经元参数之间出现互适应的情况。

## 5 结束语

准考证号识别等手写数字识别在日常生活中非常普遍，通过神经网络识别准考证号无疑是目前最流行的方法之一。本文从实际需求出发，根据应用场景设计对应的数据增强环节，对两种常用的神经网络在字迹潦草的测试集上进行了比较研究，并得出结论。通过分析准考证号码的特殊性，提出使用居中文字以及尺度调整两种数据预处理方法。结果表明，进行数据预处理能使检测准确率有所提高。最后，定量分析了卷积神经网络与 BP 神经网络的内存及计算消耗，同时定性分析了卷积神经网络之所以优于 BP 神经网络的原因，为其他类似的应用场合提供一定的参考。

## 参考文献

- [1] 杨淑莹编著, 图像模式识别[M].北京交通大学出版社, 2005
- [2] Lam L, Suen C Y. Structural classification and relaxation matching of totally unconstrained handwritten zip-code numerals. Pattern Recognition [J]. 1988, 21(1):19-31.
- [3] 王晓娟, 白艳萍. 基于 BP 神经网络的手写体数字的识别方法研究[J]. 数学的实践与认识. 2014, (7): 112-116.
- [4] 荣辉, 张济世. 人工神经网络及其现状与展望[J]. 电子技术应用. 1995(10)
- [5] Convolutional Neural Networks (LeNet) - DeepLearning 0.1 documentation. DeepLearning 0.1. LISA Lab. [31 August 2013].
- [6] 龚声蓉, 刘纯平, 王强. 数字图像处理与分析[M]. 北京: 清华大学出版社, 2006, (1).
- [7] 庄镇泉等编著. 神经网络与神经计算机[M]. 科学出版社, 1992
- [8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 86(11):2278-2324, November 1998.
- [9] Yanowitz, S. D., & Bruckstein, A. M. (1988, November). A new method for image segmentation[J]. Pattern Recognition, 1988., 9th International Conference on (pp. 270-275). IEEE.
- [10] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks[J]. NIPS, 2012.