

单位代码	10476
学号	1708283033
分类号	TP391

河南师范大学

硕士学位论文

基于 CNN 的脱机手写体汉字识别研究

学 科 、 专 业 ： 软件工程
研 究 方 向 ： 汉字识别
申 请 学 位 类 别 ： 工学硕士
申 请 人 ： 韩梦云
指 导 教 师 ： 郑延斌 教授

二〇二〇年六月

**RESERCH ON OFFLINE
HANDWRITTEN CHINESE CHARACTER
RECOGNITION BASED ON CNN**

A Dissertation Submitted
to the Graduate School of Henan Normal University
in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering

June , 2020

摘要

随着科学技术的不断发展,脱机手写体汉字识别在日常生活中的应用变得越来越广泛。然而,由于脱机手写体汉字类别繁多、书写随意性大且相似汉字很多等问题,识别手写体汉字成为了模式识别领域中一个极具挑战的问题。因此,如何能够快速且准确的识别脱机手写体汉字,已经成为模式识别研究领域最重要的课题之一。

近年来,卷积神经网络(Convolutional neural network, CNN)在汉字识别方面取得了突破性的进展。CNN 是深度学习的代表算法之一,它具有自动提取特征、权值共享等特点。在实际应用中, CNN 在 手写体汉字识别方面表现了很好的识别性能。因此,采用 CNN 识别手写体汉字具有很好的理论基础和实践价值。本文采用 CNN 的方法识别手写体汉字,取得的主要成果如下:

(1) 针对 CNN 识别手写体汉字识别速度慢的问题,提出了一种基于二维主成分分析(Two dimensional principal component analysis, 2DPCA)与 CNN 相结合的手写体汉字识别方法。该方法通过减少图像中的冗余数据提高手写体汉字的识别速度。首先,利用 2DPCA 方法提取手写体汉字的主要特征;其次,将提取的特征矩阵作为 CNN 的输入进行分类。实验表明,所提出的方法能够在不损失精度的前提下,大大的降低计算成本和运行时间。

(2) 针对脱机手写体汉字识别率低的问题,提出了一种基于独立成分分析(Independent component analysis, ICA)与 CNN 相结合的手写体汉字识别方法。该方法通过增加图像特征提高手写体汉字的识别率。首先,利用 FastICA 算法提取手写体汉字的独立基图像和投影向量,得到特征向量;其次,将 FastICA 提取的特征向量与原图像一起放入 CNN 的输入层进行分类。该方案充分利用了 ICA 的特征提取能力和 CNN 的分类能力。在 CASIA-HWDB1.1 数据库中进行验证,手写体汉字的识别率有所提高,证明了 ICA 有助于优化卷积神经网络模型,提高手写体汉字的识别率。

关键词: 手写体汉字识别, 卷积神经网络, 独立向量分析, 二维主成分分析

ABSTRACT

With the development of science and technology, offline handwritten Chinese character recognition has become more and more widely used in daily life. However, the recognition of handwritten Chinese characters has become a very challenging problem in the field of pattern recognition due to the variety of offline handwritten Chinese characters, the randomness of writing and the similarity of Chinese characters. Therefore, how to quickly and accurately identify handwritten Chinese characters has become one of the most important research topics in the field of pattern recognition.

In recent years, deep learning (especially convolutional neural network) has made a breakthrough in image recognition. As one of the representative algorithms of deep learning, convolutional neural network has the characteristics of automatic feature extraction, weight sharing. It performs well in handwritten Chinese character recognition. Therefore, using convolutional neural network to recognize handwritten Chinese characters has a good theoretical basis and practical value. The main achievements of this study are as follows:

(1) aiming at the low recognition rate of offline handwritten Chinese characters, a handwritten Chinese character recognition method based on independent component analysis (ICA) and convolutional neural network (CNN) was proposed. This method improves the recognition rate of handwritten Chinese characters by adding image features. First, we extracted independent basis images and the projection vector of handwritten Chinese character recognition by using FastICA, and obtained the feature vector. Then, the feature vectors extracted by FastICA are put into the input layer of CNN for classification together with the original image. The project took full advantage of good feature extraction feature capability of ICA and strong classification ability of CNN. It was verified in

CASIA-HWDB1.1 database that the recognition rate of handwritten Chinese characters was improved, which proved that ICA was helpful to optimize the convolutional neural network model and improve the recognition rate of handwritten Chinese characters.

(2) To solve the problem of slow recognition speed of handwritten Chinese characters by convolutional neural network, a handwritten Chinese character recognition method based on two dimensional principal component analysis and convolutional neural network is proposed. This method improves the recognition speed of handwritten Chinese characters by reducing redundant data in the image. Firstly, the main features of handwritten Chinese characters were extracted by 2DPCA; Secondly, the extracted feature matrix is classified as the input of the convolutional neural network. Experiments show that the proposed method can greatly reduce the computational cost and running time of the convolutional neural network without losing the accuracy.

KEYWORDS: handwritten Chinese character recognition, convolutional neural network, independent component analysis, two dimensional principal component analysis.

目 录

摘 要.....	I
ABSTRACT.....	III
目 录.....	V
第一章 绪论.....	1
1.1 研究背景及意义	1
1.2 国内外研究现状	3
1.3 本文主要研究内容	5
1.4 文章结构安排	6
1.5 本章小结	7
第二章 相关理论基础	9
2.1 引言	9
2.2 预处理	9
2.2.1 二值化	10
2.2.2 平滑去噪	12
2.2.3 归一化	14
2.3 特征提取	14
2.3.1 基于结构的特征提取方法.....	14
2.3.2 基于统计的特征提取方法.....	15
2.4 卷积神经网络	16
2.4.1 卷积神经网络的概述.....	16
2.4.2 卷积神经网络的算法.....	17
2.4.3 卷积神经网络的结构.....	18
2.5 本章小结	21
第三章 基于 2DPCA 与 CNN 的手写体汉字识别方法	23
3.1 引言	23
3.2 二维主成分分析	23

3.2.1 2DPCA 算法思想	23
3.2.2 特征提取	24
3.3 2DPCA-CNN 算法	25
3.3.1 预处理	26
3.3.2 基于 2DPCA 的特征提取方法	26
3.3.3 基于特征矩阵的 CNN 识别	27
3.4 实验结果及分析	28
3.4.1 数据集	28
3.4.2 实验配置及平台	28
3.4.3 实验分析	29
3.5 本章小结	33
第四章 基于 CNN 与 ICA 的手写体汉字识别方法	35
4.1 引言	35
4.2 ICA-CNN 算法	35
4.2.1 预处理	36
4.2.2 基于 FastICA 的特征提取	36
4.2.3 CNN 分类	37
4.3 实验结果及分析	39
4.3.1 数据集	39
4.3.2 实验配置及实验平台	39
4.3.3 实验结果及分析	40
4.4 本章小结	42
第五章 总结与展望	43
5.1 总结	43
5.2 展望	44
参考文献	45
致 谢	51
攻读硕士学位期间的科研成果	53
独 创 性 声 明	55

关于论文使用授权的说明	55
-------------------	----

第一章 绪论

1.1 研究背景及意义

文字识别是指通过电子设备将纸张上的文字信息转换成计算机文字的过程，它的研究从上个世纪五十年代就已经开始出现，当时的计算机才刚刚起步，人们开始着手研究如何更好的将纸质文字保存至计算机。最开始，人们研究文字识别技术是为了减轻文字处理的工作量。由于纸质资料寿命有限且难以传播，为了将各种文本资料保存到电脑中并方便查找，人们花费了大量的时间和精力手工录入文字，但人工录入不仅效率低和成本高，有时还伴随着大量的错误。

随着计算机的快速发展与互联网的迅速普及，计算机的应用也变得越来越广泛。在许多情况下，人们都需要用到汉字识别技术，如银行票据识别、邮件分拣、办公自动化、快递信息扫描等^[1-3]。在信息化发展的今天，人们已经习惯性依赖计算机完成各种文字信息处理任务。然而，由于每天产生的文字信息数量巨大，传统的手工录入文字信息的方式已经无法满足人们想要快速获取信息的需求。因此，如何通过计算机技术快速的获取文字信息并精确的识别，是模式识别领域亟待解决的问题之一。

目前，汉字是世界上使用人口最多的文字，也是世界上保存最为古老的文字。我国的光辉历史和长久累积的知识都是依靠汉字才得以保存至今的。作为我国记录、保存、传播知识的独特工具，汉字对推动中国的社会发展和文明进步发挥了重大的作用。因此，对汉字识别技术进行研究，既有实用价值又有现实意义。

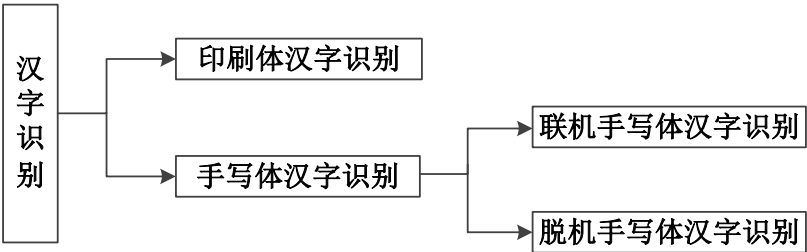


图 1-1 汉字识别技术分类

汉字识别根据识别的对象不同主要可以分为两类^[4-5]：一类是印刷体汉字识别；另一类是手写体汉字识别。根据采集方式不同，手写体汉字可以分为脱机手写体汉字和联

机手写体汉字两类,如图 1-1 所示。联机手写汉字识别所处理的手写文字是书写者通过物理设备(如数字笔、数字手写板或触摸屏)在线书写获取的文字信号,书写的轨迹可以通过定时采样即时输入到计算机中。而脱机手写汉字识别所处理的手写汉字是通过扫描仪或摄像头等图像捕捉设备采集到的手写汉字的二维图像^[6]。由于印刷体汉字与联机手写体汉字更容易开发,且产品已经趋于成熟,本文主要研究的是脱机手写体汉字字符的识别。

脱机手写体汉字识别作为模式识别的一个分支,是一个极具难度的文字识别问题。它的识别困难主要表现在^[7-9]:第一,与其他种类的语言相比,汉字字符种类特别多。在国标 GB2312-80 编码中,仅常用的一级、二级汉字就定义了 6763 类;第二,汉字的结构特别复杂。汉字的结构种类可以分为上下结构、左右结构、半包围结构和镶嵌结构等多种结构;第三,手写体汉字的书写存在一定随意性,相同的汉字由不同的人书写出来存在很大的差异性;第四,汉字中存在很多相似的汉字,且每个人书写这些相似汉字时会产生一定的形变,这使得汉字的识别更加困难。基于这些问题,虽然脱机手写体汉字已经研究了几十年,但识别效果仍然没有达到人们的心理预期。

尽管脱机手写体汉字识别十分困难,但由于脱机手写体汉字识别研究有重要的理论意义和实用价值,从研究脱机手写体汉字以来,人们对它的研究从未中断或停歇。几十年来,已经有很多关于脱机手写体汉字的研究成果陆续出现,但由于脱机手写体汉字自身的特点,传统的分类方法一直没有取得令人满意的结果^[10]。近年来,随着深度学习的出现,基于深度学习的手写体汉字识别在方法和性能上得到了突破性的进展,为手写体汉字识别提供了新的研究方向^[11]。

深度学习的概念源于人工神经网络的研究,它是由 Hinton 率先提出来的^[12-13]。与传统的人工神经网络相比,深度学习的核心思想是通过一系列的非线性变换,从汉字图像中自动提取多层次多角度的特征信息,从而使获取的特征具有更强的泛化能力和表达能力。近年来,随着深度学习技术的不断发展和数据处理能力的迅速提升,深度神经网络的训练变得更加简单快捷。因此,深度学习的研究受到了越来越多的关注,成为了脱机手写体汉字识别领域的主要研究方法。

典型的深度学习方法包括^[14]:深度置信网络(Deep belief network, DBN)、卷积神经网络(Convolutional neural network, CNN)、回归神经网络(Deep recurrent neural network, RNN)、S 层叠自动编码器(Stacked auto-encoder, SAE)等,其中 CNN 在脱机手写体汉字

识别方面取得的成果最为显著。与传统的手写体汉字识别方法相比, CNN 不仅能够有效的提取汉字图像的特征,并对不同风格、不同类型的手写体汉字有良好的适应性,因此,采用 CNN 识别手写体汉字有较好的识别效果。但是,由于不同的 CNN 模型适用于不同的识别对象,所以用 CNN 模型识别手写体汉字时,需要不断学习 CNN 模型的理论知识、不断研究 CNN 模型识别率高的原因和不断探索最佳的 CNN 模型,这对于解决脱机手写体汉字的识别有很好的实际应用价值^[15]。

1.2 国内外研究现状

文字识别的研究起源于 1929 年,最早是由德国的 Taushek 提出的,他采用光学模板匹配技术对西文字符进行识别,开创了通过机器识别文字的先河。后来,随着计算机的诞生,通过计算机进行字符识别的研究也相继出现。首先,IBM 公司的 Casey 和 Nagy 在 1966 年发表了第一篇印刷体汉字识别的文章,他们采用简单的汉字结构匹配方法对印刷体汉字进行识别,能够识别的汉字数量达到了 1000 个^[16]。同年,联机的汉字识别也取得了一些进展,如 Liu 等^[17]提出了一种基于汉字笔画特征的识别方法;Zobark^[18]提出了一种基于汉字笔迹特征的方法。受这些方法的启发,在 1977 年和 1984 年,日本研究者分别建立了第一个印刷体汉字识别系统和多体印刷汉字识别系统,其中多体汉字印刷系统的识别率高达 99.98%,达到了当时印刷体汉字的最高识别水平^[19-20]。

国内对汉字识别的研究相对较晚,七十年代末期,才开始对汉字识别技术进行研究。我国对汉字识别的研究可以分为三个阶段^[21]:第一阶段(从七十年代末到八十年代初)主要为汉字识别技术的探索阶段,在这十年左右的时间里,许多研究员对印刷体汉字识别方法的理论进行研究;第二阶段(八十年代初到九十年代初)为研制开发阶段,印刷体汉字和联机手写体汉字识别技术开始从实验室转向应用,并开始对脱机手写体汉字的识别进行研究;第三阶段(九十年代至今)为手写体汉字的飞速发展阶段,在该阶段,印刷体汉字和联机手写体汉字的研究逐渐成熟,且已经成功应用于市场。

然而,就脱机手写体汉字的识别而言,发展就相对缓慢的多。在深度学习兴起之前,汉字识别方法一直采用的是传统的汉字识别方法,该方法虽然也取得了一定的进步,但识别率一直无法达到人们的心理预期。最近几年,深度学习的方法在脱机手写体汉字取得了突破性的进展,为脱机手写体汉字的研究提供了新的思路。为此,本文把目前研究者提出的脱机手写体汉字识别方法分为传统的识别方法(非 CNN 方法)和基于 CNN 的

手写体汉字识别方法，并分别对这两类方法进行综述。

传统的手写体汉字识别方法

传统的脱机手写体汉字识别方法主要包括三个步骤：图像预处理、特征提取和分类识别^[22]。针对这些步骤，人们提出了许多有效的方法提高手写体汉字的识别性能。在图像采集过程中，光照不均、分辨率低或纸张质量差等都会影响汉字样本的使用，且由于汉字大小不一、字体各异，如果不预先进行预处理的步骤，手写体汉字的识别率将完全无法保证。因此，为了提高脱机手写体汉字的识别率，人们提取出一些图像预处理的方法。常用的如二值化、平滑去噪和样本归一化等。其中，一些代表性的方法有大津法、均值滤波、中值滤波、非线性归一化^[23]、伪二维归一化^[24]等，这些方法减少了脱机手写体汉字风格导致的字符结构不稳定问题，在一定程度上提高了脱机手写体汉字的识别性能。

特征提取是传统汉字识别的关键环节。对于特征提取，人们希望提取的汉字特征可以很好的代表一类汉字，既能和其他汉字有所区分，又能识别相同汉字的不同样本。根据特征提取方式的不同，汉字特征提取方法可以分为两类^[25]：一类是基于结构的特征提取，一类是基于统计的特征提取。基于结构的特征主要是对汉字的笔画（横、撇、竖、捺）和特征点（端点、交点等）进行研究。常见的基于结构的特征提取方法包括轮廓特征^[26]、笔画密度特征^[27]等。不过，由于结构提取特征困难，且对噪声敏感，现在已经很少使用。相对于结构特征，基于统计的特征提取方法有更好的性能，如 Gabor 特征^[28]、ICA 特征^[29]和 Gradient 特征^[30]等都在脱机手写体汉字识别中获得了较高的识别率。

提取汉字的特征最终是为了对其进行分类。因此，关于分类器的设计，必须与选取的特征相适应才能得到最高的识别率。开始，人们常用单个分类器对脱机手写体汉字进行分类，分类方法有 K 近邻^[31]、支持向量机^[32]、贝叶斯^[33]和判别二次特征学习^[34]等。随着不断的研究发现，采用单个分类器的方法，并不能从根本上提高脱机手写体汉字的识别率。于是，人们开始引入基于多分类器集成的方法，该方法通过特定的组合方式，对单个分类器进行取长补短，使每个分类器可以发挥出最大的优势，以取得更高的识别率。然而，由于手工提取的特征具有有限的可表现性，传统的“预处理+特征提取+分类”的方法遇到了瓶颈^[35]，到现在为止，传统的脱机手写体汉字仍未有较大的进展。

基于 CNN 的手写体汉字识别方法

随着深度学习的不断发展和广泛应用，基于深度学习（尤其是 CNN）的脱机手写

体汉字识别方法逐渐取代了传统方法，成为识别脱机手写体汉字的主要方法。与传统的脱机手写体汉字识别方法相比，基于 CNN 的脱机手写体汉字识别方法可以自动学习汉字的深层特征，其识别性能大大优于传统方法，如 2011 年，基于 CNN 集成的脱机手写体汉字识别方法在 ICDAR(International Conference on Document Analysis and Recognition) 竞赛中获得了第 1 名^[36]；随后 2012 年，Ciresan^[37]等提出了一种端对端的多列 CNN 模型，在脱机手写体汉字识别中取得了 93.5% 的识别率，为当时最好的识别结果；另外，在 2013 的 ICDAR 中文脱机和联机手写汉字识别竞赛中，取得比赛冠军的方法也是基于 CNN 模型的^[38-39]。

近年来，通过对 CNN 模型进行加深和改进，基于 CNN 的手写体汉字识别方法的识别性能又有了进一步的提升。如：Chen 等^[40]提出了一个手写体汉字识别的自适应局部接受域卷积神经网络。该网络通过引入不同大小的局部接受域产生不同的特征映射集，并引入最大池化层从这些特征映射集中选择最佳神经元的方式提高了手写体汉字的识别精度。Yang 等^[41]提出了一种迭代求精模块，该模块由一个基于注意力的递归神经网络实现，可以迭代地使用之前的预测来更新注意力，然后细化当前的预测。通过这种方式，模型能够聚焦于输入图像的一个子区域，以区分视觉上相似的字符，同时为了克服某些字符之间的小差异，进一步提出了一种利用低层视觉线索和高层结构信息的新型卷积架构，结合提出的迭代求精模块，精确度有了进一步的提升。Bi 等^[42]提出了一个名为 GoogLeNet 的卷积神经网络在手写体汉字识别中的应用，该网络在保持网络计算资源不变的情况下，通过增加网络的宽度和深度并优化模型（通过更改模型的输入和输出层、配备批量标准化和卷积分解等方式），减少了参数的个数并提高了识别的准确率。

以上基于 CNN 的手写体汉字识别方法虽然有了巨大的进步和成功的应用^[43]，但它们只考虑了在 CNN 模型上进行调整和修改（大多数都是对 CNN 模型的层数、池化层及其函数操作进行改进），而未使用任何领域的特征信息。传统手写体汉字的特征提取方法有很多，例如，网格特征，ICA 特征，2DPCA 特征，Gabor 特征，梯度特征等。研究发现，将这些特征信息用于 CNN 分类，可以优化 CNN 模型，提高手写体汉字的识别性能。目前，将汉字领域的特征信息引入 CNN 的研究还比较少。

1.3 本文主要研究内容

本论文主要对脱机手写体汉字进行识别，首先针对手写体汉字的特点和各种深度学

习方法在手写体汉字的表现进行研究,选择出合适的机器学习模型。随着计算机能力的飞速增长、训练数据的不断积累以及非线性激活函数的不断完善,CNN 在手写体汉字识别中表现出较好的识别性能。针对现有的 CNN 模型识别率不高的问题,本文提出了一种 CNN 与独立成分分析(Independent component analysis, ICA)相结合的方法。该方法首先利用 ICA 提取汉字图像的特征,其次将提取的特征数据与原图像一起放入 CNN 中进行分类处理。通过增加手写体汉字的特征信息,可以优化 CNN 模型,进而提高手写体汉字的识别率。

增加手写体汉字特征信息的方法虽然在一定程度上提高了手写体汉字的识别率,然而,由于输入 CNN 的手写体汉字的特征越多,识别手写体汉字花费的时间越长,所需的存储空间也越大。因此,减少手写体汉字的无关特征,可以降低 CNN 识别所花费的时间^[44]。二维主成分分析(Two dimensional principal component analysis, 2DPCA)是一种基于二维图像的特征提取方法,利用 2DPCA 能有效的降低汉字图像的维度,提取汉字的主要特征。因此,针对 CNN 模型识别手写体汉字速度慢的问题,本文提出了一种基于 2DPCA 与 CNN 相结合的手写体汉字识别方法。第一,利用 2DPCA 提取手写体汉字的投影特征向量;第二,将手写体汉字的投影特征向量组成特征矩阵;第三,用组成的特征矩阵作为 CNN 的输入;第四,用 Softmax 函数进行分类。实验结果表明,本文所采用的识别方法花费的时间远远小于其他几种识别方法。

1.4 文章结构安排

第一章 绪论部分

该部分首先介绍脱机手写体汉字的研究背景与意义,然后对汉字识别和 CNN 的分类以及国内外研究现状进行分析,最后讲述了本文的主要研究内容和论文的结构安排。

第二章 对识别手写体汉字采用的相关方法进行详细阐述。

通过介绍传统的脱机手写体汉字识别过程,对预处理、特征提取和分类识别中常用的方法进行具体介绍。

第三章 基于 2DPCA 和 CNN 的手写体汉字识别。

本章首先介绍 2DPCA 的特征提取方法,然后具体描述 2DPCA 和 CNN 相结合的手写体汉字识别过程,最后在 CASIA-HWDB1.1 数据集上设计了三组对比实验,证明了该方法能够减少识别手写体汉字所花费的时间。

第四章 基于 ICA 和 CNN 的手写体汉字识别。

本章首先对 CNN 模型和 ICA 进行简单的介绍，然后介绍 ICA 和 CNN 相结合的手写体汉字识别方法，最后将该方法与其他几种识别方法在 CASIA-HWDB1.1 数据集上进行实验，实验结果证明 ICA 有助于优化 CNN 模型，提高手写体汉字的识别率。

第五章 总结与展望。

本章总结了本课题的研究内容，说明了目前的研究工作中遇到的问题并对下一步研究的重点和方向进行展望。

1.5 本章小结

本章主要对手写体汉字进行概述性的描述，首先介绍了脱机手写汉字识别的研究目的和意义，其次讲述了手写体汉字的分类和识别困难的原因，然后对神经网络和汉字识别的发展过程进行了阐述，最后简单介绍了本文的主要研究方法和通篇的结构安排。接下来本文将对 1.2.1 节中介绍的传统方法和 1.2.2 节中介绍的 CNN 模型展开描述，并重点研究将传统特征提取和 CNN 相结合的方法。

第二章 相关理论基础

2.1 引言

脱机手写体汉字识别是模式识别领域的研究热点之一，也是文字识别领域最为困难的问题之一^[45]。手写体汉字的识别涉及模式识别、图像处理、人工智能、组合数学和心理学等多类学科。由于在拍照文档、支票、表单表格、手稿文书等图像识别系统有广泛的应用前景，过去几十年，研究者做了大量的研究工作来提高手写体汉字的识别性能，但收效甚微^[46-47]。影响手写体汉字识别率的因素有很多，比如光照不均、分辨率低或书写纸张质量差等。另外，手写体汉字自身的特点（①字符集大②书写随意性大且缺乏规范③许多汉字相似度高④自身结构复杂）也为手写体汉字的识别带来了极大的难度。由于同一汉字不同样本间汉字结构不稳定，且相似的手写体汉字很难分辨，因此，识别手写体汉字需要良好的特征提取能力和泛化能力。

传统的脱机手写体汉字的识别过程主要分为三步：预处理、特征提取和分类识别^[48-49]。预处理的主要作用是尽量减少图像中的噪声信息，强化图像中汉字的特征并对不规则的汉字加以修正。特征提取的主要任务是提取出汉字图像中有效的特征信息，它是手写体汉字识别中很重要的一个阶段。由于汉字图像中包含大量的无关信息，对汉字图像直接处理时效率较低，且没有太大意义。在实际的应用中，大量的数据特征并不能刻画数据的本质特征，数据的冗余还可能会影响到后续的数据处理^[11]。因此，特征提取的好坏将直接影响手写体汉字的识别率。汉字分类识别是图像处理的目的。通过特征提取，汉字图像被提取出有效的特征信息，使用分类器对提取的特征进行分类训练，并根据识别结果寻找出最优的手写体汉字识别方法。

在脱机手写体汉字识别中，常用的预处理方法有二值化、图像平滑、归一化等；特征提取方法有 Gabor 特征提取、网格特征提取以及方向线素特征提取；分类方法包含最近邻分类、支持向量机(Support vector machine, SVM)、CNN 等。下面是关于这些方法的简单介绍。

2.2 预处理

2.2.1 二值化

二值化是图像处理过程中的一个重要的步骤，它广泛应用于图像识别、图像边缘提取和图像分割等领域^[50-52]。图像二值化的原理是通过划分阈值将灰度图像上的像素点（灰度值范围为 0-255）转换为只有 0 和 255 的黑白图像。根据划分阈值的方法不同，二值化大致可以分为全局阈值法与局部阈值法^[53]。

全局阈值法是对整个图像选取单一的阈值进行划分（大于阈值像素为 255，小于阈值像素为 0），相比局部阈值法，全局阈值法的优点是计算速度较快，对于汉字和背景灰度值相差较大的图像，采用全局阈值法具有很好的效果^[54]。常用的全局二值化算法有大津法、期望灰度算法等。

局部阈值法可以分为两种，一种是将整个图像划分为图像块，每个图像块都设定一个阈值进行二值化；另一种是根据每一个像素点邻域的灰度变化来设定阈值，然后逐点对图像进行二值化^[55]。局部阈值法的优点是对于光照不均匀或者低对比度的图像，它可以根据图像的局部灰度值自动调整阈值。相比全局阈值法，该方法具有更好的适应性^[56]。常用的局部二值化算法有 Bernsen 方法^[57-58]、Ni-black 方法^[59]等。下面对部分常用的二值化方法进行简要的介绍。

（1）大津法

大津法是一种自适应的阈值确定方法。它的基本原理是^[60]：设一个阈值将图像分成背景和目标两部分（一部分的图像灰度对应目标，另一部分的图像灰度对应背景），且该阈值可以使这两部分灰度值的类内方差最小，类间方差最大。

具体实现过程如下：

设一副图像的灰度级为 M ，阈值为 L ，则图像可分为 $C_0=\{1-L\}$ 和 $C_1=\{L+1-M\}$ 两部分。分别计算每部分的概率 w_0 、 w_1 和平均值 u_0 、 u_1 ，整幅图像的灰度平均值 u ，则两部分之间的方差可以由公式（2-1）计算。

$$G_2(L) = w_0(u_0 - u)^2 + w_1(u_1 - u)^2 \quad (2-1)$$

求出 $G_2(L)$ 最大时的 L 值， L 值即为图像的阈值。

（2）期望灰度法

设图像大小为 $M \times N$ ，灰度值为 L_1, L_2, \dots, L_k ，若用随机变量 X 表示图像中每个像素点的灰度值，则 X 有 K 个可能取值，分别为 L_1, L_2, \dots, L_k 。其实，概率分布也可以

描述图像的灰度分布情况。假设各个灰度值发生的概率分别为 $P_1=P(L_1)$, $P_2=P(L_2), \dots$, $P_K=P(L_K)$, 并且

$$\sum_{n=1}^K P_n = 1 \quad (2-2)$$

则随机变量可以用公式(2-3)表示, 从公式可以看出, 不同的图像可以由不同的密度矩阵来表示。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} L_1, L_2, \dots, L_n \\ P_1, P_2, \dots, P_n \end{bmatrix} \quad (2-3)$$

在统计随机变量 X 的过程中, 最重要的统计特征便是期望值, 它代表了变量 X 在整幅图像中的平均值, 随机变量 X 的取值经常集中在此^[61]。既然灰度值在图像当中是一个随机变量, 则从灰度“中心”进行分割应当是最佳的平衡点, 它使取黑像素的灰度值和白像素的灰度值均等^[62]。设 μ_{th} 表示阈值, 则有:

$$-\mu_{th} = \sum_{n=1}^K L_n P(L_n) \quad (2-4)$$

令 $P(L_n)$ 代表图像中灰度 L_n 出现的次数。则由公式(2-4)可以推出:

$$\begin{aligned} \mu_{th} &= \sum_{n=1}^K L_n P(L_n) \\ &= \sum_{n=1}^K L_n \frac{h(L_n)}{\sum_{m=1}^K h(L_m)} \\ &= \sum_{n=1}^K L_n h(L_n) / \sum_{n=1}^K h(L_n) \end{aligned} \quad (2-5)$$

公式(2-5)是一种基于全局阈值的分割方法, 该方法的适用性十分广泛, 大部分的灰度图像都可以使用, 摆脱了传统的基于直方图的阈值分割方法对图像模型的假定^[63]。而且, 该方法的运算过程十分简单, 能够大大提高图像的处理速度。由于汉字图像的背景与文字之间灰度相差较大, 采用期望灰度法能够把汉字很好的区分出来。

(3) Bernsen 法

Bernsen 法属于是一种动态选择阈值的自适应方法, 阈值的选取是随着像素位置变化而变化的, 它是一种局部阈值法。在光照不均匀的条件下, 它可以取得较好的处理效果。其主要思想如下:

人为设定两个值 S 与 TT (Bernsen 最初设 S 为 15, TT 为 128), 计算图像中以任意像

素点 P 为中心的、窗口大小为 $k \times k$ 的所有像素的最大值 Max 和最小值 Min 。求出两者的均值 T ，如果 $M-N$ 大于 S ，则当前像素点 P 的阈值为 T ；若小于 S ，则表示该窗口所在区域灰度级差别较小，那么窗口或是在目标区或是在背景区，随后判断 T 与 TT 的关系，若 $T < TT$ 则当前点的灰度值为 0，否则当前点的灰度值为 255^[64]。

(4) Niblack 方法

Niblack 方法是一种简单有效的动态阈值分割方法，修改得到最佳参数之后的效果比大津法要好，因为大津法是根据整个图像来确定一个阈值，而 Niblack 则是在不同的 $R \times R$ 区域会有不同的阈值。

Niblack 算法的二值化阈值是通过某一像素点及其邻域内像素点灰度值的均值和标准差计算得到的。在计算图像点 (x, y) 二值化阈值时，首先计算以 (x, y) 为中心的 $n \times n$ 大小的区域内像素点的灰度均值 m 和标准差 s 。灰度均值 m 和标准差 s 的计算公式如下：

$$m(x, y) = \frac{1}{n^2} \sum_{i=x-\frac{n}{2}}^{x+\frac{n}{2}} \sum_{j=y-\frac{n}{2}}^{y+\frac{n}{2}} f(i, j) \quad (2-6)$$

$$s(x, y) = \sqrt{\frac{1}{n^2} \sum_{i=x-\frac{n}{2}}^{x+\frac{n}{2}} \sum_{j=y-\frac{n}{2}}^{y+\frac{n}{2}} (F(i, j) - m(x, y))^2} \quad (2-7)$$

根据灰度均值和标准差计算得到点 (x, y) 的二值化阈值 T ，计算公式为 $T(x, y) = k \times s(x, y) + m(x, y)$ ，其中 k 为修正系数。最后根据计算得到的阈值 T 对该点进行二值化处理。将图像中所有的像素点按照此方法处理即可得到二值化图像。

$$b(x, y) = \begin{cases} 255, & f(x, y) \geq T(x, y) \\ 0, & f(x, y) < T(x, y) \end{cases} \quad (2-8)$$

不过，Niblack 算法虽然能够实现图像的二值化，但是如果选取的区域均为背景点时，该算法会将灰度值较高的点当做是目标点，导致伪噪声的引入。而且，用 Niblack 方法进行图像分割时，处理模板窗口 $R \times R$ 大小的选择十分重要，如果空间选择的太小，噪声抑制的效果不理想，目标主体不够突出，若空间选择的太大，则会丢失很多信息。

2.2.2 平滑去噪

由于图像在产生、传输和记录过程中，经常会受到各种噪声的影响，这些噪声可能会对图像的识别效果产生影响，因此，对图像进行降噪处理是预处理中一个重要的步骤^[65]。在手写体汉字识别过程中，图像平滑的目的是消除或减少噪声的影响，提高待识别

图像的质量,提高图像的识别率。常见的图像平滑算法有中值滤波、均值滤波和高斯滤波。

(1) 均值滤波

均值滤波是一种图像处理过程中常用的线性平滑滤波,它的主要思想是计算出邻域中所有像素的灰度平均值来代替每个像素的灰度值,即选择一个 $k \times k$ (k 通常为 3) 的窗口作为待处理像素点 (x,y) 的模板,该模板由其邻域的若干像素组成,求出模板中所有像素的均值,再把该均值赋予待处理的像素点 (x,y) ,处理后该点上的灰度值为 $g(x,y)$ ^[66]。求均值的公式为(2-9), m 为该模板中包含待处理像素点在内的像素的总个数。

$$g(x, y) = \frac{1}{m} \sum f(x, y) \quad (2-9)$$

该方法的主要优点是运算简单快速,同时又能够较为有效的去除高斯噪声,适用性非常广,缺点是破坏了图像的边缘,模糊了图像的轮廓。

(2) 中值滤波

中值滤波是常用的非线性滤波方法,也是图像处理技术中最常用的预处理技术。它在平滑噪声方面非常有效,同时也可以保留图像的边缘信息。与均值滤波相比,中值滤波能更好的保持图像的细节信息。它的基本思想是将窗口内所有像素的灰度中值代替某像素的灰度值。

标准的中值滤波算法是将待处理像素点及其邻域中所有灰度值比较大小后,依次排列,然后取其中值替换中心灰度值。以邻域 3×3 的中值滤波为例, $g(x, y)$ 与 9 个像素点的灰度值有关,9 个像素点分别为 $f(x-1, y-1)$, $f(x-1, y)$, $f(x-1, y+1)$, $f(x, y-1)$, $f(x, y)$, $f(x, y+1)$, $f(x+1, y-1)$, $f(x+1, y)$ 与 $f(x+1, y+1)$ 。则 $g(x, y)$ 取这 9 个数中灰度值大小排第 5 位的值。假设某个像素点与其邻域的所有灰度值分别为(10,20,20,20,15,20,20,25,100),排序后为(10,15,20,20,20,20,20,25,100),则 $g(x, y)=20$ 。

(3) 高斯滤波

高斯滤波是一种线性平滑滤波,适用于消除高斯噪声,广泛应用于图像处理的减噪过程。高斯滤波是对整幅图像进行加权平均的过程,每一个像素点的值,都由其本身和邻域内的其他灰度值经过加权平均后得到,它对于抑制服从正态分布的噪声非常有效^[67]。高斯滤波的具体操作是:用一个模板(或称卷积、掩模)扫描图像中的每一个像素,用模板确定邻域内像素的加权平均灰度值去替代模板中心像素点的值^[68]。在图像处理过程中,高斯滤波的实现方式一般有两种,一种是用离散化窗口滑窗卷积,另一种是通过

傅里叶变换。通常，最常用的是采用滑窗卷积的方式。不过，在离散化的窗口特别大的情况下，可能会考虑基于傅里叶变换的实现方法。

2.2.3 归一化

图像归一化就是把所有需要处理的汉字图像的大小、位置和笔画粗细等特征变换为统一的标准^[69]。汉字图像的归一化处理主要包括位置归一化和大小归一化。经过归一化处理后，图像数据会被限制在一定范围内。

图像归一化可以分为线性归一化和非线性归一化两种^[70]。线性归一化可以放大和缩小原始图像的长度和宽度，保留图像的线性性质，如图 2-1 所示，a 为归一化前的图像，b 为归一化后的图像。非线性归一化主要是通过将图像中的密度进行均衡来达到归一化的目的。这种方法可以有效的解决汉字图像的变形问题^[71]。在手写体汉字识别过程中，采用归一化是为了更方便的处理汉字图像，并加快程序运行的速度。有时候汉字图像可能不是所要求的形式，这时就要求对不同大小的汉字图像的中心位置做相应的更改，使它们统一到相同的位置上，故采用非线性归一化技术。

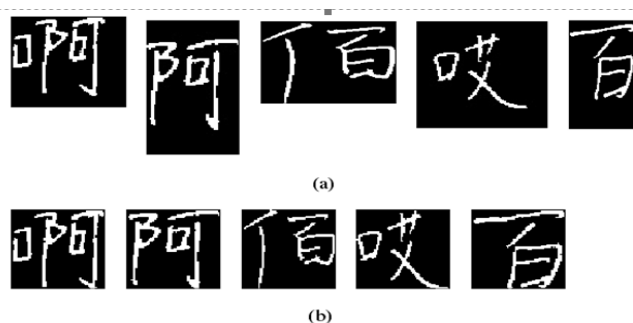


图 2-1 归一化图像前后对比

2.3 特征提取

在传统的脱机手写体汉字识别过程中，特征提取是一个很关键的步骤，它的好坏直接影响了脱机手写体汉字的识别性能。因此，为了有效的提取脱机手写体汉字的特征信息，人们提出了许多种特征提取的方法。根据提取特征的方式不同，人们将这些特征提取的方法分为了两类^[72-74]：一类是基于结构的特征提取；另一类是基于统计的特征提取。

2.3.1 基于结构的特征提取方法

基于结构的特征提取方法主要是提取汉字图像的结构信息。由于汉字的结构大部分都是由“横”、“竖”、“撇”、“捺”四种基本笔划和端点、拐点和交叉点等特征点组成，所以汉字的结构具有一定的稳定性。正因为汉字的结构稳定，所以，基于结构的特征提取方法对书写不规范的手写体汉字有较好的适应性。然而，基于结构的特征提取方法也有很大缺陷。首先，该方法需要经过大量的计算，因此耗费时间比较长，提取特征的速度十分缓慢。其次，当手写体汉字书写比较随意时，会对汉字的结构造成一定影响，在提取汉字图像的结构特征时，会因为受到干扰而导致提取结构的稳定性下降。基于结构的特征提取方法是最初研究汉字识别时采用的主流方法，由于结构提取特征困难，且对噪声敏感，现在已经很少使用。

2.3.2 基于统计的特征提取方法

基于统计的特征提取方法根据特征提取的区域不同可以分为全局统计特征和局部统计特征^[7]。局部统计特征的基本原理是：首先将汉字图像分割成块，其次统计分割后每部分图像的特征信息，最后，所提取的汉字的特征即为本次统计的结果。全局统计特征顾名思义是直接对整个汉字图像提取特征信息。与基于结构的特征提取方法相比，基于统计的特征提取方法具有速度快、识别率高等优点。在脱机手写体汉字识别中，广泛使用的统计特征提取方法有弹性网格特征、方向梯度直方图(Histogram of Oriented Gradients, HOG)特征、方向线素特征(Directional Element Feature, DEF)、Gabor 特征和 ICA 特征等。下面是一些常用特征提取方法的具体描述。

(1) 弹性网格特征

弹性网格就是根据汉字像素点的分布，用非均匀的线条将汉字图像划分所得到的非均匀的网格。根据区域进行划分，弹性网格可以划分为局部弹性网格和全局弹性网格两种。局部弹性网格由于需要对笔画的密度函数进行反复计算，因此计算复杂度较高，很少使用；与之相比，全局弹性网格的划分相对简单，应用比较广泛。

基于弹性网格的特征提取的基本原理是通过上述网格对汉字图像进行分块，然后通过变换或分析的方式对每一块的像素进行处理。采用该方法对汉字图像进行划分可以较好的反映汉字的结构特征，对汉字书写风格不同带来的字体形变等问题有较好的适应性。但该方法也有其缺陷，由于汉字划分导致了各个图像块之间互相不存在联系，因此，汉字的整体结构特征无法体现出来。

（2）HOG 特征

HOG 梯度特征是一种在计算机视觉和图像处理中用来进行物体检测的特征描述子，它一般通过统计图像局部区域的梯度方向直方图来表示汉字图像的特征。HOG 的主要思想是：在一副图像中，局部目标的表象和形状能够被梯度和边缘的方向密度分布很好的描述。与其他特征描述方法相比，HOG 有很多优点。首先，由于 HOG 是在汉字图像的局部单元上进行操作，所以它对图像几何的和光学的形变都能保持很好的不变性。

（3）Gabor 特征

Gabor 小波与人类视觉系统中简单细胞的视觉刺激响应非常相似，它在提取目标的局部空间和频率域信息方面具有良好的特性。与其它特征提取算法相比，Gabor 滤波器也有很多优点。首先，Gabor 小波对于图像的边缘敏感，能够提供良好的方向选择和尺度选择特性，其次，Gabor 小波对于光照变化不敏感，能够提供对光照变化良好的适应性。因此，采用 Gabor 滤波器提取脱机手写体汉字的特征具有很好的效果。在实际应用中，Gabor 滤波器可以在频域的不同尺度，不同方向上提取相关特征。

（4）ICA 特征

独立成分分析（Independent component analysis, ICA）是一种基于信号的高阶统计特性的分析方法，它充分地考虑了信号概率密度函数的统计独立性。与传统的多维信号分析方法相比，ICA 更加全面的揭示了数据间的本质结构。大量研究表明，图像中最重要的信息存在于图像像素的高阶统计量中。因此，ICA 可以提取图像的高阶统计特性，使变换后的图像各分量之间相互独立。

近年来，很多研究人员利用统计特征和结构特征的特点将两种特征提取方法相融合。它们以统计特征和结构特征的优点为出发点，研究出将两者融合的新的特征提取方法。目前，手写体汉字识别特征提取的主流方法是多特征融合方法，它似乎已经成为未来发展的趋势^[75]。

2.4 卷积神经网络

2.4.1 卷积神经网络的概述

CNN 是深度学习的代表算法之一。自 CNN 提出至今，无论是在目标定位、图像分类与识别等领域，还是在其他智能领域，CNN 都发挥了很大的作用并取得了极其优异

的成果^[72]。CNN 之所以有如此出色的表现，主要是因为它采用了局部连接和共享权值的方式。通过该方式，它不仅减少了权值的数量使得网络更易于优化，而且也降低了 CNN 过拟合的风险。作为神经网络的算法之一，CNN 另一个优点是可以直接对二维图像进行处理。该优点使汉字图像可以直接作为神经网络的输入，避免了将二维图像转换为一维向量时丢失数据的空间结构特征分布^[11]。

CNN 最早的原型理念是在上世纪八十年代由 Fukushima 提出的。九十年代，LeCun 等对其进行了开发。通过对卷积结构使用反向传播(BP)，提出了一种具有层与多层结构局部连接的卷积神经网络模型 LeNet。它由一系列卷积层和池化层组成，每一层的输出都可以看作是一组特征图。由于 LeNet 网络在手写体识别的成功应用，人们开始了 CNN 的研究。在 LeNet 网络之后，CNN 一直处于实验发展阶段，直到 2012 年 AlexNet 网络的提出，奠定了 CNN 在深度学习应用中的地位^[76]。

随着 AlexNet 的成功，研究者们对 CNN 模型进行了多种改进方法以提升 CNN 的性能。其中比较经典的几种 CNN 模型为 VGGNet^[77]，GoogleNet^[78]，ResNet^[79]。随着 CNN 模型不断的发展，人们已经开始研究更深更宽的神经网络模型。虽然更深更宽的神经网络可以得到较好的特征表示，但这也使得神经网络变得更加难以优化和花费的时间更长。为了解决这些问题，本文提出了两种优化 CNN 模型的方法。

2.4.2 卷积神经网络的算法

CNN 是一种模拟人脑神经元细胞的网络结构，它是由大量简单的神经元相互连接成的自适应非线性动态系统。相比于传统的多层感知器，本文采用整流线性单元(Rectified Linear Units)作为网络的激活函数，用 Softmax 函数作为神经网络的输出，并以交叉熵(Cross-entropy)衡量训练的损失。CNN 通过输入与一组核滤波器进行卷积来提取局部特征，然后把得到的卷积特征图进行子采样(表示为池化)，并过滤到下一层。下面是对 CNN 算法的简单介绍^[80]。

假设 $x_{l,i} \in \mathbf{R}^{M_l \times M_l}$ 表示第 l 层的第 i 个映射，第 l 层中的第 j 个核滤波器连接第 $(l-1)$ 层中的第 i 个映射表示为 $k_{l,j} \in \mathbf{R}^{k_l \times k_l}$ ，索引映射集 $M_j = \{i | \text{在第}(l-1)\text{层中的第}i\text{个映射连接第}l\text{层中的第}j\text{个映射}\}$ ，因此由公式(2-10)可以得出卷积运算。

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right) \quad (2-10)$$

其中 $f(\cdot)$ 为非线性激活函数， b_j^l 是偏置。常用的非线性激活函数包括 ReLU、sigmoid

和 Tanh, 由于 ReLU 具有简单、快速和避免“梯度消失”的优点, 因此采用 ReLU 作为非线性激活函数, 其计算公式如(2-11)所示:

$$f(z) = \max(0, z) \quad (2-11)$$

对于采样层来说, 就是将卷积的特征图降维, 方程式如(2-12)所示。

$$J(\theta) = -\frac{1}{N} \left[\sum_{i=1}^N \sum_{t=1}^T l(y_i = t) \log \frac{e^{\theta_t^T x_i}}{\sum_{l=1}^T e^{\theta_l^T}} \right] \quad (2-12)$$

其中 $\text{down}(\cdot)$ 是求和采样函数, 计算 $x(l-1)$ 映射中的每个 $n \times n$ 区域的最大值。

Softmax 主要用于多分类问题, 它可以将多个神经元的输出映射到(0,1)区间内, 从而将多分类的结果以概率的方式展现。假设有 T 类标签, 每个类别的训练数据用 (x_i, y_i) 表示, 其中 $i = \{1, \dots, N\}$, x_i 和 y_i 分别为特征向量和标签。CNN 的目标是最小化交叉熵损失函数:

$$J(\theta) = -\frac{1}{N} \left[\sum_{i=1}^N \sum_{t=1}^T l(y_i = t) \log \frac{e^{\theta_t^T x_i}}{\sum_{l=1}^T e^{\theta_l^T}} \right] \quad (2-13)$$

其中 θ 是模型参数, $\sum_{l=1}^T e^{\theta_l^T x_i}$ 是归一化因子, $l(\cdot)$ 是指示函数。

在 CNN 的训练过程中, 使用随机梯度下降(SGD)算法可以使 $J(\theta)$ 的损失函数最小化。

2.4.3 卷积神经网络的结构

经典的 CNN 模型结构包括输入层、卷积层、池化层、全连接层和输出层。输入层神经元主要负责接收信号, 与传统的神经网络一样, CNN 在接收信号之前也需要对图像进行预处理; 卷积层为特征提取层, 它的作用是提取图像的局部特征, 第一层卷积通常只能提取一些低级的特征信息, 但随着层数的增加, 则能从低级的特征信息中提取更高级、更复杂的特征, 且一旦该特征被提取, 则与其他特征间的位置关系也会随之确定; 池化层一般与卷积层相连, 由于卷积层得到的图像特征维度较大, 池化层通过对特征切块的方法, 提取每块特征的最大值或平均值, 然后达到降低计算复杂度的作用; 全连接层则负责将所有的局部特征通过权值矩阵连接起来, 然后通过 Softmax 函数或其他函数进行输出。下面是对每一层作用的详细介绍^[81]。

(1) 输入层

输入层的主要作用就是将数据输入 CNN，数据输入之后 CNN 开始运行。在识别汉字图像的过程中，被输入的是一幅图像的像素矩阵。在该矩阵中，长和宽为输入汉字图像的大小，它的深度为图像的色彩通道。比如，黑白图像的深度为 1，CNN 的输入是一个二维神经元，如图 2.2 中 a 所示。而对于 RGB 格式的图片，图像的深度则为 3，CNN 的输入则是一个三维神经元（RGB 中的每一个颜色通道都有一个的矩阵），如图 2.2 中 b 所示。

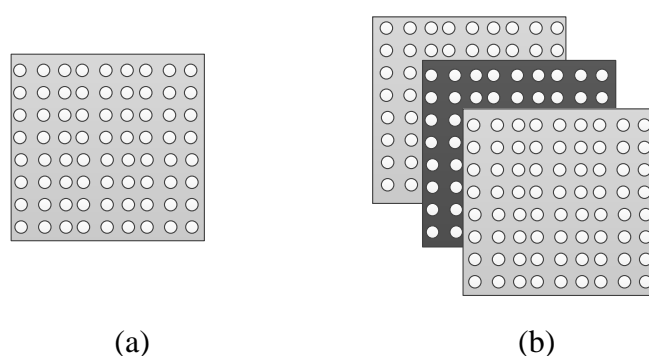


图 2-2 图像通道

(2) 卷积层

卷积层是构建 CNN 的核心层，它产生了网络中大部分的计算量。在卷积运算时，会人工设定一个大小为 $K \times K$ 的方阵，称为过滤器。每个过滤器在宽度和高度上都比较小，但是过滤器的深度 d 必须和输入层的深度 d 保持一致，因此过滤器的大小为 $K \times K \times d$ 。从数学的角度出发，过滤器为 d 个 $K \times K$ 的方阵。在实际的操作中，不同的模型会确定不同数量的过滤器，其个数记为 n ，每一个 n 包含 d 个 $K \times K$ 的方阵，并且计算生成一个输出矩阵。

(3) 池化层

池化在 CNN 中有非常重要的作用。池化之后，图像相比输入前的图像会明显缩小，因此更容易处理。池化层的引入是仿照人的视觉系统对输入的对象进行降维和抽象。在 CNN 工作过程中，研究者普遍认为池化层有三个优点。首先，池化相当于在空间范围内做了维度约减，从而使模型可以抽取更加广范围的特征。同时，它减小了下一层的输入大小，进而减少了计算量和参数个数；其次，池化使 CNN 模型具有特征不变性，因为池化操作更加关注是否存在一些特征而不是特征存在的位置；最后也是最重要的，池化层可以有效地防止 CNN 发生过拟合。

池化层主要分为两类，最大池化层和平均池化层。在池化层中，不采用复杂的加权计算，而是直接选取滤波器中的最大值或平均值作为提取的特征。如果提取的是滤波器中的最大值，则被称为最大池化(max pooling)操作，若提取的是平均值，则被称为平均池化(average pooling)操作。

与上边卷积层一样，池化层同样需要人工设定滤波器的大小，并考虑是否需要填充 0 和设置滤波器移动的步长等问题。与卷积层不同的是，池化层的滤波器仅仅影响一个深度上的节点，而卷积层采用的滤波器是横跨整个深度的。因此，池化层的滤波器不仅需要在长和宽这两个维度上移动，另外，它也需要在深度维度上移动。

如图 2-3 所示，在连续的卷积层之间通常会周期性地插入一个池化层。如果池化层选择的是最大池化层，最常见的形式是：池化层选取尺寸为 2×2 的滤波器，然后以步长为 2 来对每个深度切片进行降采样。最大池化层操作是从 4 个像素值中取最大值，去掉其中 75% 的激活信息（也就是从深度切片中某个 2×2 的区域中取一个值），深度保持不变。

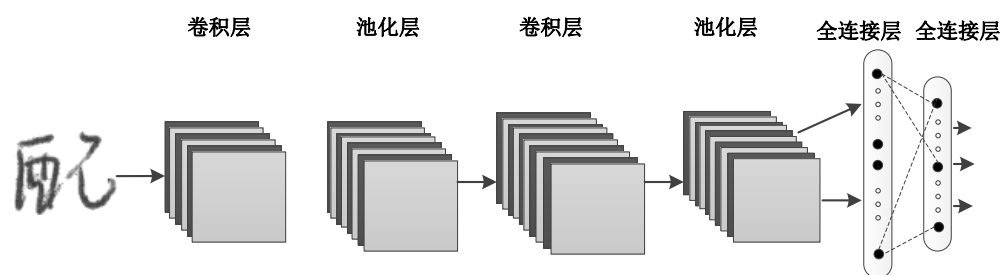


图 2-3 CNN 模型

(4) 全连接层

全连接层在 CNN 过程中起到了分类器的作用。经过不断的卷积和池化之后，更深层次的特征已经被提取出来，全连接层需要将每一个结点都同上一层的所有结点相连接，然后将提取到的特征信息通过权值矩阵做一个合并。

(5) Softmax 层

Softmax 层可以理解为归一化。假设汉字图像的类别为 10，Softmax 层的输出则是一个 10 维的向量。向量中的第一个值为待识别图像为第一类的概率，第二个值为待识别图像为第二类的概率，以此类推，它们的概率和为 1。

2.5 本章小结

本章对传统的手写体汉字识别步骤和方法进行了简要的阐述，阐述的内容包含三个部分。第一部分是对于手写汉样本的预处理，分别介绍了二值化、平滑去噪和归一化三种方法，并对每种方法的作用进行了简要描述。特征提取是手写体汉字识别的重要过程，特征提取的好坏往往决定了手写体汉字识别的效果。因此，第二部分对常用的特征提取方法进行介绍。常用的特征提取方法包括方向像素特征、网格特征和基于独立成分分析的特征提取等，对每种方法的优缺点进行分析和了解，以便更好地对其应用。由于本文采用基于 CNN 的识别方法，因此第三部分描述的是 CNN 相关知识。该部分不仅对 CNN 模型进行了简要概述，也对它的模型结构和算法实现过程进行了简单的说明。

第三章 基于 2DPCA 与 CNN 的手写体汉字识别方法

3.1 引言

随着深度学习时代的到来,手写体汉字作为模式识别研究中的一个热点问题,引起了学者们的广泛关注。CNN 作为深度学习的经典算法之一,在 手写体汉字识别方面表现优异。然而,随着 CNN 层数的加深,识别手写体汉字所花费的时间也不断增加。因此,针对 CNN 识别手写体汉字速度慢的问题,本文提出了一种先降维后识别的方法——基于 2DPCA 的手写体汉字识别方法。该方法在保持手写体汉字识别率的情况下,可以有效的提升手写体汉字的识别速度。它主要分为两个阶段:第一阶段为图像特征提取,利用 2DPCA 提取手写体汉字的二维图像特征;第二个阶段为图像分类,将第一阶段提取的二维图像特征矩阵放入 CNN 的输入层中进行分类。通过对实验结果的分析与评价,该方法显著降低了 CNN 的运行时间,表明了该方法的合理性和有效性。

3.2 二维主成分分析

3.2.1 2DPCA 算法思想

设 \mathbf{X} 表示一个 n 维的单位列向量,2DPCA 的思想是通过公式(3-1)将图像 \mathbf{A} (一个 $m \times n$ 的图像矩阵)投影到 \mathbf{X} 上^[82-84]:

$$\mathbf{Y} = \mathbf{AX} \quad (3-1)$$

从而得到一个 m 维的投影向量 \mathbf{Y} ,投影向量 \mathbf{Y} 为图像 \mathbf{A} 的投影特征向量。然而,投影向量 \mathbf{X} 应该如何求得呢?实际上,投影样本的总体散布矩阵可以用来测量投影向量 \mathbf{X} 的判别能力,而投影样本的总体散布矩阵可以利用投影特征向量的协方差矩阵的迹来表示^[85]。从这一点来看,可采用以下准则:

$$J(\mathbf{X}) = \text{tr}(\mathbf{S}_x) \quad (3-2)$$

其中,训练样本投影特征向量的协方差矩阵由 \mathbf{S}_x 表示,且 \mathbf{S}_x 的迹用 $\text{tr}(\mathbf{S}_x)$ 表示。公式(3-2)中准则最大化的物理意义是求出一个投影方向 \mathbf{X} ,所有的样本都投射到它上面,使产生的投射样本的总体散射是最大的。协方差矩阵 \mathbf{S}_x 可以表示为^[86]:

$$\begin{aligned}
 S_x &= E(Y - EY)(Y - EY)^T \\
 &= E[AX - E(AX)][AX - E(AX)]^T \\
 &= E[(A - EA)X][(A - EA)X]^T
 \end{aligned} \tag{3-3}$$

因此,

$$tr(S_x) = X^T [E(A - EA)^T (A - EA)] X \tag{3-4}$$

定义总体散布矩阵 G_t 如公式(3-5)所示:

$$G_t = E[(A - EA)^T (A - EA)] \tag{3-5}$$

由定义可知, G_t 是一个 $n \times n$ 的非负定矩阵, 可以直接使用训练图像样本来评估 G_t 。假设训练图像样本的总数为 L , 第 j 个训练图像用一个 $m \times n$ 的矩阵 $A_j (j=1, 2, \dots, M)$ 表示, 所有训练样本的均值图像用 \bar{A} 表示, 则 G_t 可以被评估通过

$$G_t = \frac{1}{L} \sum_{i=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \tag{3-6}$$

则(3-7)中的准则可以表示为

$$J(X) = X^T G_t X \tag{3-7}$$

其中 X 是一个单位列向量。这个准则称为广义总体散射准则。使准则最大化的单位向量 X 称为最优投影轴。当将图像矩阵投影到 X 上之后, 投影样本的总散射将达到最大化。

若 X 的最优投影轴用 X_{opt} 表示, 则 X_{opt} 是最大化 $J(X)$ 的单位向量, 即 G_t 的特征向量对应于最大的特征值。一般来说, 只有一个最优投影轴是不够的。通常需要选择一组投影轴 $X_1; X_2; \dots; X_d$, $X_1; X_2; \dots; X_d$ 受正交约束和最大化准则 $J(X)$ 的管制, 如公式(3-8)和公式(3-9)所示^[87]:

$$X_i^T X_j = 0, i, j = 1, 2, \dots, d \tag{3-8}$$

$$\{X_1, X_2, \dots, X_d\} = \arg \max J(X) \tag{3-9}$$

事实上, 最优投影轴 $X_1; X_2; \dots; X_d$ 是 G_t 的标准正交特征向量对应的前 d 个最大的特征值。

3.2.2 特征提取

2DPCA 的最优投影向量 $X_1; X_2; \dots; X_d$ 被用作特征提取。对于给定的图像样本 A , 令

$$Y_k = AX_k, k=1,2,\dots,d. \quad (3-10)$$

然后,得到一组投影特征向量 $Y_1; Y_2; \dots; Y_d$, $Y_1; Y_2; \dots; Y_d$ 为样本图像 A 的主成分。需要注意的是, 2DPCA 的每个主成分都是一个向量。所有的主成分构成一个 $m \times d$ 的矩阵 $B=[Y_1; Y_2; \dots; Y_d]$, B 称为图像样本 A 的特征矩阵或特征图。

3.3 2DPCA-CNN 算法

为了提升 CNN 模型的识别性能,人们提出许多方法提高手写体汉字的识别率,主要包括数据增加和采用更深的 CNN 网络。数据增加是一种非常重要的提升 CNN 鲁棒性及推广能力的技术,通常采用平移、尺度缩放、旋转、仿射变换和弹性形变等方法对手写体汉字图像数据进行扩充。然而,数据增加和更深的 CNN 网络模型,都会导致计算成本增加、训练和识别时间过长。为了提高 CNN 的识别速度,采用了一种基于 2DPCA 与 CNN 的手写体汉字识别方法,该方法可以保持手写体汉字的识别精度并提高识别效率。

利用 2DPCA 与 CNN 相结合方法识别手写体汉字,首先输入样本图像并对其进行预处理。预处理的第一步是统一图像的尺寸大小,即将所有样本转为大小相同的图像。第二步是将相同大小的图像归一化处理,归一化之后所有图像的像素值都在 $[0, 1]$ 区间内。预处理之后,利用 2DPCA 提取总体散布矩阵的特征向量对应的前 d 个最大特征值,即最优投影向量;将图像投影到最优投影向量即可得到图像的特征矩阵。最后,将特征矩阵放入 CNN 的输入层,通过卷积、池化、全连接等一系列操作,求取识别结果。识别的算法描述过程如下:

输入: 样本集 A

低维空间维数 d .

过程:

对样本集 A 进行预处理;

根据式(3-1)求 A 的投影向量 Y ;

利用式(3-3)求协方差矩阵 S_x ;

根据 S_x 求出协方差矩阵的迹;

求广义总体散射准则函数: $J(X)=X^T G_t X$;

求最优投影向量 $X_k(k=1,2,\dots,d)$ 。最优投影轴 X_1, X_2, \dots, X_d 是 G_t 的标准正交特征向量对应

的前 d 个最大的特征值 X_1, X_2, \dots, X_d ;

将图像 A 投影到 X_k , 求投影特征向量 Y_1, Y_2, \dots, Y_d ;

特征矩阵 $B=[Y_1, Y_2 \dots, Y_d]$;

令训练集 $A_l=\{(B_k, F_k)\} | k=1$; 测试集 $A_m=\{(B_{l+1}, F_{l+1})\} | m k=l$;

用 A_l 训练一个 CNN 模型;

输出:用 CNN 对 A_m 中的样本进行预测, 预测结果与其标签作比较, 得到一个预测精度。

通过 2DPCA 与 CNN 相结合的方法识别手写体汉字, 首先求汉字图像的二维特征, 每个汉字图像对应一个特征矩阵, 利用此矩阵, CNN 可以实现图像的分类。整个识别过程在 3.3.1-3.3.3 节中详细介绍。

3.3.1 预处理

(1) 尺寸统一化

设整个手写体汉字的数据集为 A , 由于 A 中汉字图像的尺寸大小各不相同, 为了便于处理, 通过尺寸统一化将所有汉字图像设置成相同的大小。统一尺寸后的图像高为 m , 宽为 n 。

$$\text{Resize}(A)=[m,n]$$

(2) 图像归一化

图像归一化, 就是通过一系列变换, 将待处理的原始图像转换成相应的唯一标准形式。在提取特征之前, 对汉字图像进行归一化处理, 归一化之后所有的像素值都在 $[0, 1]$ 区间内。

3.3.2 基于 2DPCA 的特征提取方法

预处理之后, 汉字图像为 $m \times n$ 的归一化图像。2DPCA 算法的目的是将每一张 $m \times n$ 的汉字图像降维成 $m \times d$ 的特征矩阵 (d 远远小于 n), 且将 $m \times d$ 的特征矩阵放入 CNN 中训练, 手写体汉字的识别率不会明显降低。

如何求特征矩阵呢? 设汉字图像的最优投影向量为 X_1, X_2, \dots, X_d , 通过 $Y_k=AX_k$ ($k=1, 2, \dots, d$) 对 $m \times n$ 的汉字图像 A 做线性变换, 投影特征向量 Y_1, Y_2, \dots, Y_d 即可构成特征矩阵 B , 特征矩阵 $B=[Y_1, Y_2 \dots, Y_d]$ 。

通过 2DPCA 算法求特征矩阵的具体步骤如下:

步骤 1: 根据式(3-3)求得协方差矩阵 S_x , 由 3.2.1 节知 $S_x=E[(A-EA)X][(A-EA)X]^T$ 。

步骤 2: 根据迹的定义求出协方差矩阵 S_x 的迹, $\text{tr}(S_x)=X^T[E(A-EA)^T(A-EA)]X$ 。

步骤 3: 求广义总体散射准则函数。定义总体散布矩阵 $G_t=E(A-EA)^T(A-EA)$, 由 3.2.1 节公式(3-1)的准则函数可得 $J(X)=X^TG_tX$ 。

步骤 4: 求最优投影向量 $X_k(k=1,2,\dots,d)$ 。最优投影轴 X_1, X_2, \dots, X_d 是 G_t 的标准正交特征向量对应的前 d 个最大的特征值。

步骤 5: 将图像 A 投影到 X_k 上, 由公式(3-1)可知投影特征向量为 Y_1, Y_2, \dots, Y_d 。

步骤 6: 特征矩阵 $B=[Y_1, Y_2, \dots, Y_d]$ 。

3.3.3 基于特征矩阵的 CNN 识别

特征矩阵作为 CNN 的输入在手写体汉字识别方面展现了很好的性能, 它不仅没有降低手写体汉字的识别率, 还明显提高了 CNN 的识别速度。本文采用三种 CNN 模型验证该方法的有效性。它包括 AlexNet (Alex Krizhevsky) 模型与文献[10]和文献[11]提出的两种深度卷积神经网络模型。文献[10]和文献[11]的模型分别命名为 ACNN 模型和 DCNN 模型。

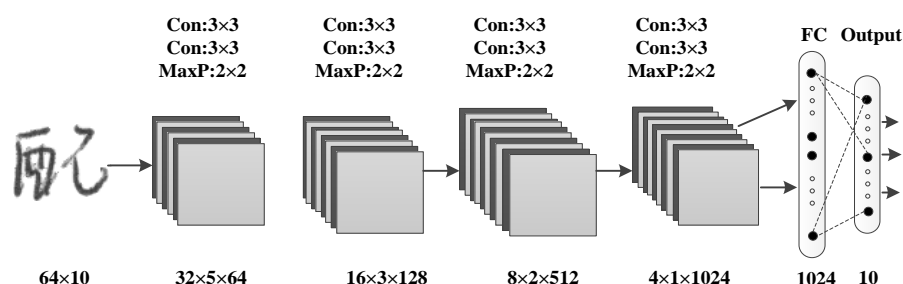


图 3-1 ACNN 的模型架构

文献[10]的 CNN 结构包含 8 个卷积层、4 个池化层、1 个全连接层和 1 个输出层。全连接层连接上一层的所有激活, 输出层用 Softmax 函数生成分类。如图 3-1 所示, 图中每一部分包含两个卷积层和一个池化层。

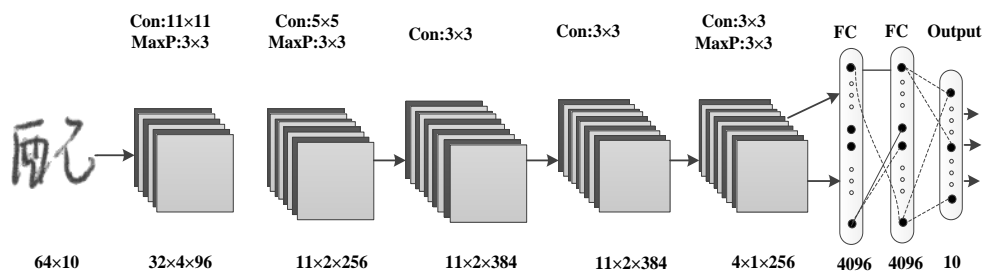


图 3-2 AlexNet 的模型架构

AlexNet 采用与文献[76]相同的体系结构, 模型如图 3-2 所示。它由 8 个加权层组成;

前五层包括三组卷积层和最大池化层, 以及两个单独的卷积层; 其余三层是完全连接的层。在该模型中, 所有滤波器的大小均为 3×3 , 卷积步长为 1, 池化层的大小为 2, 步长也为 2。

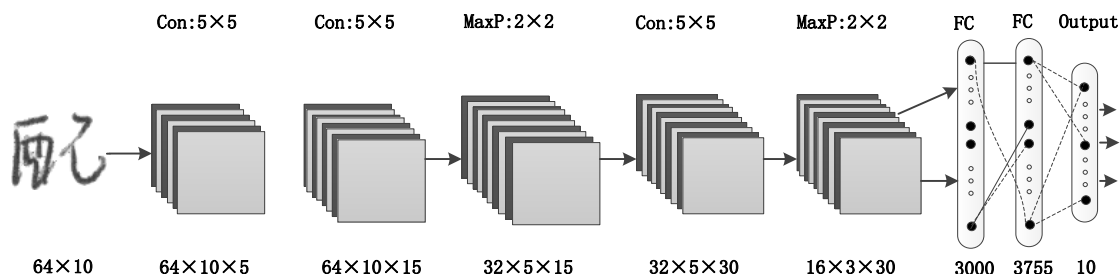


图 3-3 DCNN 的 CNN 模型架构

文献[11]提出的 CNN 架构如图 3-3 所示, 主要由 3 个卷积层、2 个池化层、1 个全连接层和一个 Softmax 回归层组成。其中, 网络的前 6 层用于特征提取, 最后一层用于分类。

3.4 实验结果及分析

3.4.1 数据集

本研究采用中国科学院自动化研究所提供的脱机手写体汉字数据集 CASIA-HWDB1.1 进行实验。CASIA-HWDB1.1 数据集包含了 3755 个常用的 GB2312 一级汉字, 由 300 个不同的编写者书写。为了验证本文提出的方法, 选取 CASIA-HWDB1.1 中的 15 组相似手写体汉字进行验证, 每组包含 10 个相似样本, 每个汉字的样本库包含 240 个训练样本和 60 个测试样本。部分相似手写体汉字样本如图 3-4 所示。



图 3-4 CASIA-HWDB1.1 的相似数据样本

3.4.2 实验配置及平台

在预处理中,将所有的汉字图像统一为 64×64 的大小,并对 64×64 的汉字图像进行归一化处理。提取特征矩阵时,选择 d 的个数为 10,即每张图像的特征矩阵大小为 64×10 。提取特征矩阵之后,对每个特征矩阵进行标准化处理,保证每个维度的特征数据方差为 1,均值为 0。

特征矩阵输入 CNN 之前,首先打乱整个训练样本,以减少训练时的过拟合。为了得到更好的手写体汉字识别率,对 3 个 CNN 模型进行了多次重复实验进行调参。选取部分参数作对比,不同 CNN 模型的识别率如表 3-1 所示,对 3 个 CNN 模型的平均识别率作比较,可知第三组数据的识别率最高,效果最好。因此, CNN 模型的参数配置如下:批大小设置为 48,正则化大小为 0.8,学习率为 0.0002,每循环 200 次学习率降低 10%。

实验采用的平台是基于 python 语言的深度学习框架 tensorflow,硬件环境为 Intel i5,操作系统为 windows10 64 位。

表 3-1 不同参数的手写体汉字识别率对比

网络模型参数				网络模型识别率			
批大小	正则化	学习率	循环降低率	ACNN	AlexNet	DCNN	平均值
48	0.7	0.0002	10%	94.65%	93.70%	88.26%	92.20%
60	0.7	0.0002	10%	93.73%	93.36%	88.54%	91.88%
48	0.8	0.0002	10%	95.69%	94.69%	92.51%	94.30%
60	0.8	0.0002	10%	95.30%	95.60%	91.28%	94.06%
48	0.7	0.0001	10%	94.58%	93.52%	90.46%	92.85%
60	0.7	0.0001	10%	94.33%	93.16%	90.33%	92.60%
48	0.8	0.0001	10%	94.85%	93.84%	91.06%	93.25%
60	0.8	0.0001	10%	95.02%	93.85%	91.44%	93.44%
48	0.7	0.0001	20%	94.38%	95.31%	90.05%	93.25%
60	0.8	0.0001	20%	95.27%	94.16%	90.08%	93.17%

3.4.3 实验分析

将 2DPCA 与 CNN 相结合的方法用于手写体汉字识别,并分别与 3.3.3 节中的三种 CNN 模型进行对比。2DPCA 与三个 CNN 模型相结合的方法分别命名为 2DPCA-ACNN、2DPCA-AlexNet 和 2DPCA-DCNN,三种方法统称为 2DPCA-CNN。通过对比三种模型与 2DPCA 结合前后的识别率和识别时间,验证本文方法的合理性和有效性。

(1) 比较 2DPCA-ACNN 与 ACNN

将 2DPCA-ACNN 与文献[10]中的 ACNN 方法作对比,ACNN 的架构如 3.3.3 节的图 3-1 所示。在识别过程中,任选一组相似手写体汉字分别采用本文方法与文献[10]中的方法进行训练和识别,随着迭代次数的增加,相似手写体汉字中测试集的识别率变化

曲线如图 3-5 所示。其中横坐标为迭代次数,纵坐标为测试集的识别率,迭代次数设置为 5000 次。

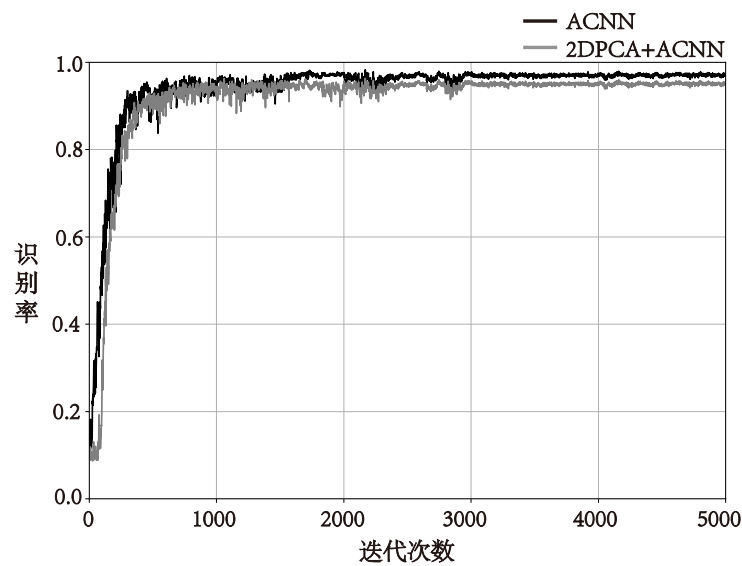


图 3-5 2DPCA-ACNN 与 ACNN 识别率变化

从图中可以看出,随着迭代次数的增加,测试集的识别率逐渐上升并趋于稳定,且 2DPCA-ACNN 与 ACNN 识别率分别为 95.12%和 94.88%, 2DPCA-ACNN 与 ACNN 的识别方法在本次实验中识别率相差不大。

表 3-2 2DPCA-ACNN 与 ACNN 的识别率和识别时间对比

组别	网络模型	识别率	识别时间
第一组	ACNN	95.60%	211942s
	2DPCA-ACNN	94.77%	38972s
第二组	ACNN	92.76%	212635s
	2DPCA-ACNN	93.30%	43024s
第三组	ACNN	93.03%	210769s
	2DPCA-ACNN	92.89%	40069s
第四组	ACNN	94.32%	213856s
	2DPCA-ACNN	94.07%	42057s
第五组	ACNN	92.85%	213650s
	2DPCA-ACNN	93.10%	42033s
平均值	ACNN	93.71%	212570s
	2DPCA-ACNN	93.63%	41231s

为了避免实验结果的偶然性,选择 5 组相似手写体汉字进行实验,并求出 5 组相似手写体汉字的平均识别率和识别时间。对比 2DPCA-ACNN 与 ACNN,得到的测试集的识别率和识别时间如表 3-2 所示。从表 3-2 可以看出,ACNN 和 2DPCA-ACNN 中每一组手写体汉字的识别率波动不超过 1%,2DPCA-ACNN 识别手写体汉字所花费的时间相

比 ACNN 所花费时间却减少了 80%。

(2) 比较 2DPCA-AlexNet 与 AlexNet

该实验采用的 CNN 模型是文献[76]中的 AlexNet 模型, AlexNet 的模型架构如 3.3.3 节的图 3-2 所示。在本次识别过程中, 随着迭代次数的增加, 相似手写体汉字中测试集 的识别率变化曲线如图 3-6 所示。由图中可以看出, AlexNet 与 2DPCA-AlexNet 的识别 率分别为 92.45% 和 94.29%, 2DPCA-AlexNet 的识别率略大于 AlexNet。(由于每次实 验初始化参数值都是随机的, 样本打乱顺序也是随机的, 因此一定程度的波动属于正常 范围)。

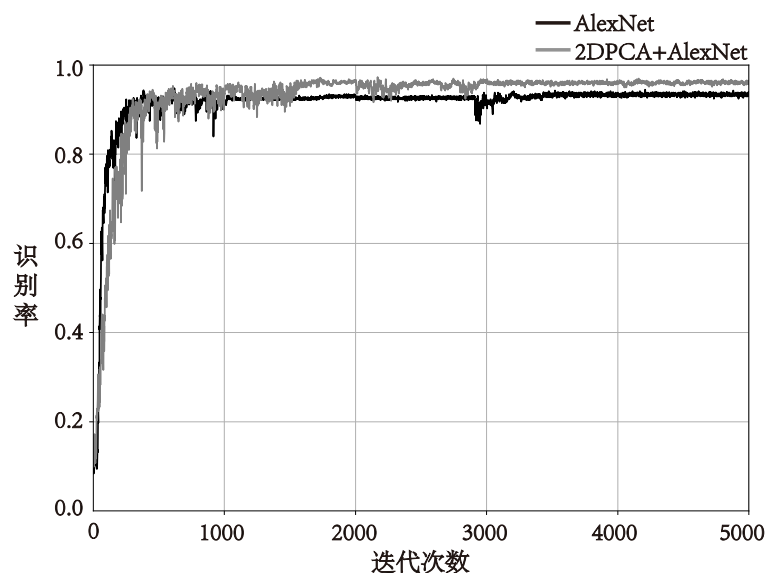


图 3-6 2DPCA-AlexNet 与 AlexNet 识别率变化

为了避免本次实验结果的偶然性, 仍然选择这 5 组相似手写体汉字进行实验。对比 2DPCA-AlexNet 与 AlexNet, 得到的测试集的识别率和识别时间如表 3-3 所示。从表 3-3 可以看出, 随着 CNN 深度的增加, 手写体的识别率也逐渐增加, 且 2DPCA-AlexNet 与 AlexNet 中每组手写体汉字的识别率略小于 2DPCA-ACNN 与 ACNN 的识别率。此外, 与 AlexNet 算法相比, 2DPCA-AlexNet 识别手写体汉字花费的时间减少了 78%。

表 3-3 2DPCA-AlexNet 与 AlexNet 的识别率和识别时间对比

组别	网络模型	识别率	识别时间
----	------	-----	------

第一组	AlexNet	94.60%	110952s
	2DPCA-AlexNet	95.28%	24524s
第二组	AlexNet	90.32%	118647s
	2DPCA-AlexNet	89.57%	24696s
第三组	AlexNet	92.03%	117027s
	2DPCA-AlexNet	91.88%	24754s
第四组	AlexNet	93.02%	115955s
	2DPCA-AlexNet	93.67%	24638s
第五组	AlexNet	92.26%	115821s
	2DPCA-AlexNet	91.07%	24775s
平均值	AlexNet	92.45%	115680s
	2DPCA-AlexNet	92.29%	24677s

（3）比较 2DPCA-DCNN 与 DCNN

本次实验比较方式同 3.4.3 中的（1）、（2）一致，DCNN 识别手写体汉字的架构如 3.3.3 节中的图 3-3 所示。在识别过程中，随着迭代次数的增加，相似手写体汉字中测试集的识别率变化曲线如图 3-7 所示。此时测试集在 2DPCA-DCNN 与 DCNN 模型上的变化曲线基本重合，证明了两种方法的识别率相同。

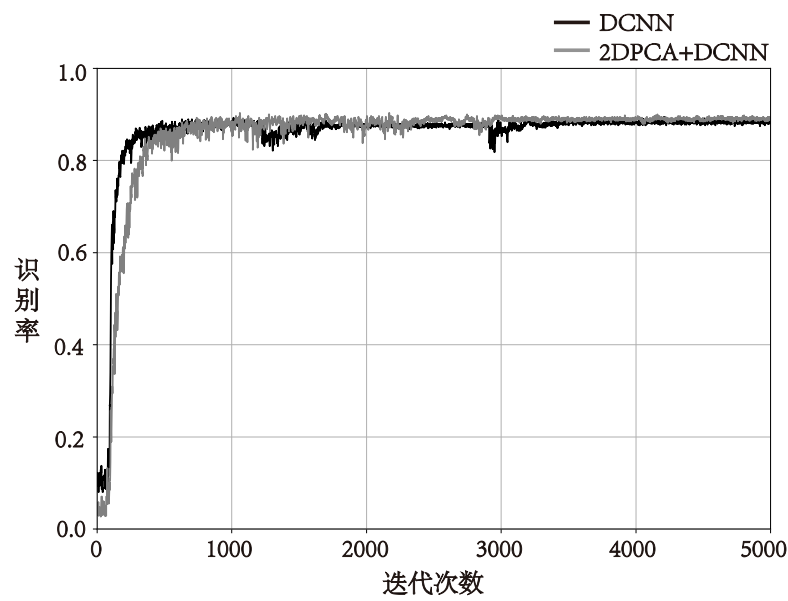


图 3-7 2DPCA-DCNN 与 DCNN 识别率变化

本次在 2DPCA-DCNN 与 DCNN 模型上进行对比，得到每组的测试集的识别率和识别时间如表 3-4 所示。在本次实验中，两种方法的手写体汉字的识别率波动不大，且与 DCNN 相比，2DPCA-DCNN 识别手写体汉字所花费的时间降低了 73%。

比较（1）与（2）中的实验可知，随着 CNN 层数的增多，不仅手写体汉字的识别率显著提高，CNN 识别手写体汉字的时间也成倍增长。比较每次实验中 CNN 与

2DPCA-CNN 所花费的时间,可以发现 CNN 层数越深,本文方法节省的时间也越多。比较 (1)、(2) 和 (3) 中的三次实验,由图 3-5、图 3-6 和图 3-7 可以直观的看出,随着迭代次数的增加,2DPCA-CNN 与 CNN 的识别率趋于稳定并且识别率大小没有明显的差异。比较表 3-2、表 3-3 和表 3-4 中的每组数据和其平均值,可以看出每次实验 2DPCA-CNN 识别手写体汉字的时间远远小于 CNN 所花费的时间,证明了 2DPCA 与 CNN 相结合的方法与 CNN 方法相比具有明显的时间优势。

表 3-4 2DPCA-DCNN 与 DCNN 的识别率和识别时间对比

组别	网络模型	识别率	识别时间
第一组	DCNN	92.51%	17672s
	2DPCA-DCNN	91.26%	4524s
第二组	DCNN	89.32%	18647s
	2DPCA-DCNN	90.52%	4696s
第三组	DCNN	91.03%	17027s
	2DPCA-DCNN	90.84%	4454s
第四组	DCNN	90.62%	15955s
	2DPCA-DCNN	91.37%	4638s
第五组	DCNN	91.26%	15821s
	2DPCA-DCNN	91.37%	4775s
平均值	DCNN	90.95%	17024s
	2DPCA-DCNN	91.07%	4617s

3.5 本章小结

近年来,在手写体汉字识别领域比传统方法好的深度学习模型主要是基于 CNN 及其改进方法的。但是 CNN 也有其缺陷,比如需要大量的训练样本、更深的 CNN 模型保持其精度。由于增加样本和采用更深的 CNN 模型需要大量的参数计算,这无疑增加了 CNN 的训练时间。针对 CNN 训练时间长的问题,本文采用了 2DPCA 与 CNN 相结合的手写体汉字识别方法。该方法首先去除图像中的冗余信息,得到手写体汉字的特征矩阵;其次将特征矩阵放入 CNN 中进行识别。由于特征矩阵的参数变少,因此 CNN 识别手写体汉字所花费的时间减少。

第四章 基于 CNN 与 ICA 的手写体汉字识别方法

4.1 引言

脱机手写体汉字的识别问题是模式识别研究领域的难点之一,为了证明传统的特征提取方法——独立成分分析有助于优化 CNN 的模型,提高手写体汉字的识别率,提出了一种 ICA 和 CNN 相结合的手写体汉字识别方法。首先,利用 FastICA 算法提取手写体汉字的独立基图像和投影向量,得到特征向量;其次,将 FastICA 提取的特征向量与原图像一起放入 CNN 的输入层进行分类。该方案充分利用了 ICA 的特征提取能力和 CNN 的分类能力。实验表明,该方法能够优化 CNN 模型,提高手写体汉字的识别率。

4.2 ICA-CNN 算法

CNN 是一种深度监督学习下的机器学习模型。近年来,CNN 在手写体汉字识别方面取得了很好的识别效果。特征提取是传统手写体汉字识别的重要步骤之一,在手写体汉字识别中,特征提取的方法主要可以分为两类:一类是基于结构的特征提取方法,另一类是基于统计的特征提取方法。ICA 是一种广泛使用的基于统计的特征提取方法,利用 ICA,能够提取到 CNN 无法提取的特征信息。将 ICA 提取的特征向量与原图像相结合,可以增加图像的特征信息,使后续的分类在精度上有所提升。

ICA 与 CNN 相结合识别手写体汉字的流程图如图 4-1 所示。首先输入汉字图像并对汉字图像进行预处理。预处理的第一步是对汉字图像的尺寸归一化处理,即转为大小相同的图像。归一化之后,每张图像可以看作是大小相同的二维矩阵。第二步是将每张图像矩阵转化为一维向量,则所有图像数据可以组成一个包含所有图像的二维矩阵,二维矩阵的每行是一张图像的所有数据信息。预处理之后,通过 FastICA 算法提取二维矩阵的独立基图像,将待处理图像投影到独立基图像的空间中得到图像的特征向量,将得到的特征向量与原图像相结合,得到特征图像。最后,将特征图像作为 CNN 的输入图像,通过卷积、池化、全连接等操作,得出识别结果。

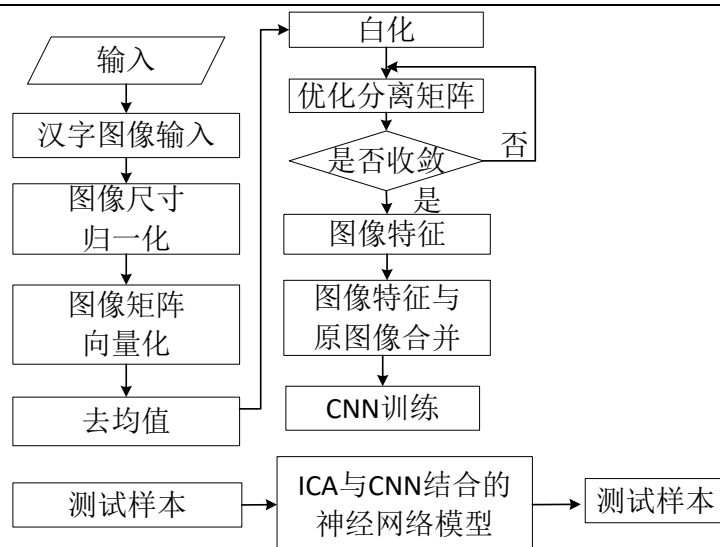


图 4-1 ICA 与 CNN 相结合的手写体汉字识别流程图

用上述方法识别手写体汉字，主要分为三个步骤：图像预处理、基于 FastICA 算法的特征提取和 CNN 分类，其内容在 4.2.1-4.2.3 节中详细介绍。

4.2.1 预处理

(1) 尺寸归一化

设手写体汉字的数据集为 \mathbf{X} ， \mathbf{X} 中的图像是由不同尺寸的手写体汉字图像组成，为了更好地处理这些汉字图像，通过尺寸归一化将汉字图像设置成相同的尺寸，归一化后的高为 M ，宽为 N 。

$$\text{Resize}(\mathbf{X}) = [M, N]$$

(2) 图像向量化

由于 FastICA 只能处理一维的样本数据，而汉字图像是一个二维矩阵，因此需要先将二维的汉字图像矩阵转化为一维向量。如果将归一化后的数据集 \mathbf{X} 表示为 $\mathbf{X} = \{\mathbf{A}_1(M, N), \mathbf{A}_2(M, N), \dots, \mathbf{A}_n(M, N)\}$ ，则每张图像 $\mathbf{A}_i(M, N)$ 向量化后可由 $\mathbf{A}_i(M * N)$ 表示。 $\mathbf{A}_i(M, N)$ 是矩阵， $\mathbf{A}_i(M * N)$ 是一维向量。 n 张图像组成的样本矩阵为 $\mathbf{X} = n \times (M * N)$ 。

4.2.2 基于 FastICA 的特征提取

如 4.2.1 节所述，每张汉字图像被向量化为一维向量，通过对所有汉字图像向量化得到了一个 n 行 $M * N$ 列的样本矩阵 \mathbf{X} ，样本矩阵 \mathbf{X} 的每行代表一张图像。本节由 FastICA 算法求解每张图像的分离矩阵 \mathbf{W} ，通过分离矩阵 \mathbf{W} 得到每张图像的基图像和特征系数，

特征系数即为汉字特征。假设独立分量个数为 $a*N$ ，则经过处理的图像有 $a*N$ 个特征值。对每一张图像进行特征提取，即可得到每一张图像的特征。将每张图像的特征与原图像相结合，即可得到特征图像 Y ，特征图像的大小 $(M+a)*N$ 。

通过 FastICA 算法提取图像特征的具体步骤如下所示：

(1) 特征中心化。即每一维的数据都减去该维的均值。

$$X = X - \bar{X}$$

(2) 白化处理：白化的主要目的是去除数据的相关性。数据的白化处理可以使随后的计算大为简化。我们通常使用特征值分解的方法进行数据的白化，白化处理的过程分为三步，首先通过 X 求协方差矩阵 C ，然后由 C 求出其特征值 D 和特征向量 E ，最后由 E 求出白化矩阵 R ，公式如下所示。

$$C = \text{cov}(X', 1) \quad \% \text{计算 } X \text{ 的协方差矩阵 } C$$

$$[E, D] = \text{eig}(C) \quad \% \text{计算协方差矩阵 } C \text{ 的特征值和特征向量}$$

$$R = D^{(-1/2)} * E' \quad \% \text{得到白化矩阵 } R, \text{ 白化矩阵 } R \text{ 经过线性变}$$

换可得到正交矩阵 Z ，通过 Z 可以降低后续工作的复杂度。

$$Z = RX \quad \% \text{生成正交矩阵 } Z$$

(3) 用迭代法分离矩阵 W 。

基于负熵的迭代格式计算分离矩阵 W 的公式如下：

$$W(k+1) = E[Zg(W^T(k)Z)] - E[g'(W^T(k)Z)]W(k)$$

(4) 求每张图像的独立基图像。

$$S = W^T X$$

(5) 求每张图像的特征，即投影系数 $F(A)$ 。

$$F(A) = X * \text{pinv}(S)$$

(6) 原图像与该图像的特征相结合，得到特征图像 Y 。

$$Y = [X, F(A)]$$

4.2.3 CNN 分类

CNN 可以处理大量的图像数据，由 4.2.2 提取的特征图像而不是原图像作为 CNN 的输入，增加了图像的信息输入，可以训练出更好的 CNN 模型。

如图 4-2 所示，CNN 主要包含 6 个卷积层 3 个池化层和 2 个全连接层。前三个卷积层后面都有一个最大池化层。最后一个最大池化层之后是三个卷积层和两个全连接层，

两个全连接层包含 1024 个神经元。最后一个输出层包含 10 个神经元,用来执行最后的分类。在该模型中,所有滤波器的大小均为 3×3 ,卷积步长为 1,池化层的大小为 2,步长也为 2。

另外,本文还加入了许多其他方法优化该模型。比如:为了降低学习计算成本,避免通过噪声更新产生局部极小值,该模型利用分批处理的方式处理图像;为了减少过拟合,引入了修正线性单元(ReLU)激活函数;为了验证 CNN 结果的准确性,选择了交叉验证的方式。由于完全连接层是相当冗余的,在两个完全连接的层之间添加 dropout 层进行正则化。

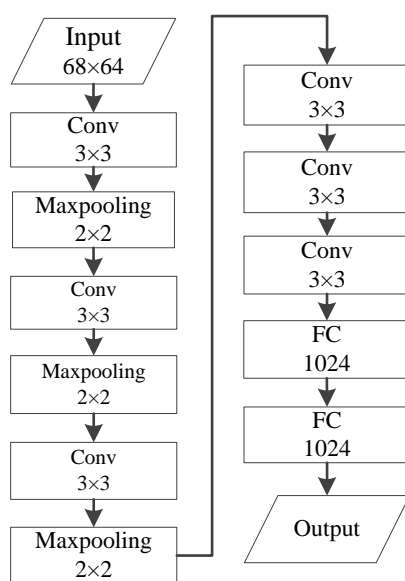


图 4-2 识别脱机手写体汉字的 CNN 结构

用上述 CNN 模型的学习过程和识别过程如下:

步骤 1 输入特征图像,对该图像进行正向传播(经过了卷积、ReLU、池操作、全连接层和 Softmax 中的正向传播)。

步骤 2 用交叉熵计算 Softmax 的总误差。

步骤 3 首先用反向传播算法计算各个权值的误差梯度,其次采用改进的梯度下降算法(Adam Optimizer)更新滤波权值和参数值,以最小化输出误差为目标。换句话说就是根据权重对网络分类误差的贡献,按照比例进行调整。

步骤 4 对训练集中的所有图像重复上述步骤多次。在对 SoftMax 输出向量的概率(选择概率最高的状态)进行 ArgMax 评估后,通过计算正确预测的次数来评估 CNN 的准确性。

步骤 5 将测试数据输入到训练好的 CNN 模型中, 计算识别率。

4.3 实验结果及分析

4.3.1 数据集

公开的脱机手写体汉字识别数据库有很多, 但大多数书写比较规范, 识别准确率已经很高, 而 CASIA-HWDB1.1 汉字识别库中的汉字, 很多字迹书写潦草, 书写风格多样, 因此识别准确率较低^[88]。由于手写体汉字识别精度低的主要原因是相似汉字区分困难和书写随意性大, 因此, CASIA-HWDB1.1 中相似手写体汉字的识别率提高, 它的整个汉字数据库的识别率也会相应提高。

本文选取了 CASIA-HWDB1.1 中的 10 组相似汉字验证该方法的有效性。部分样本如图 4-3 所示。CASIA-HWDB1.1 包含了 3755 个汉字, 每个汉字由 300 个不同的人书写, 每个汉字的样本库包含 240 个训练样本和 60 个测试样本。



图 4-3 CASIA-HWDB1.1 的相似数据样本

4.3.2 实验配置及实验平台

在预处理中, 将图像归一化为 64×64 的大小作为原图像; 对原图像进行特征提取时, 采用的独立分量的个数为 16×64 , 即每张图像提取 16×64 个特征, 将原图像与图像特征相结合, 特征图像的大小为 80×64 , 即 CNN 的输入图像大小为 80×64 的特征图像。将特征图像输入 CNN 之前, 首先打乱整个训练样本, 以减少训练时的过拟合。CNN 模型参数配置如下: 批处理大小为 48, 正则化大小为 0.8, 学习率为 0.0002, 每循环 200 次学习率降低 10%。

实验采用的平台是基于 python 语言的深度学习框架 tensorflow, 硬件环境为 Intel i5,

操作系统为 windows10 64 位。

4.3.3 实验结果及分析

(1) 本文识别手写体汉字的过程如第三章所述, 实验采用的网络结构及详细配置如 4.2.3 节和 4.3.2 节。在识别过程中, 将其中一组手写体汉字的特征图像放入 CNN 模型中进行训练和识别, 随着迭代次数的增加, 相似手写体汉字中测试集的认识率变化曲线如图 4-4 所示, 其中纵坐标为识别精度, 横坐标为迭代次数。从图中可以看出, 随着迭代次数的增加, 测试集的认识精度逐渐上升。本文将迭代次数设置为 5000 次, 训练完成后保存最佳模型, 此时测试集在该模型上的识别率为 95.8%。

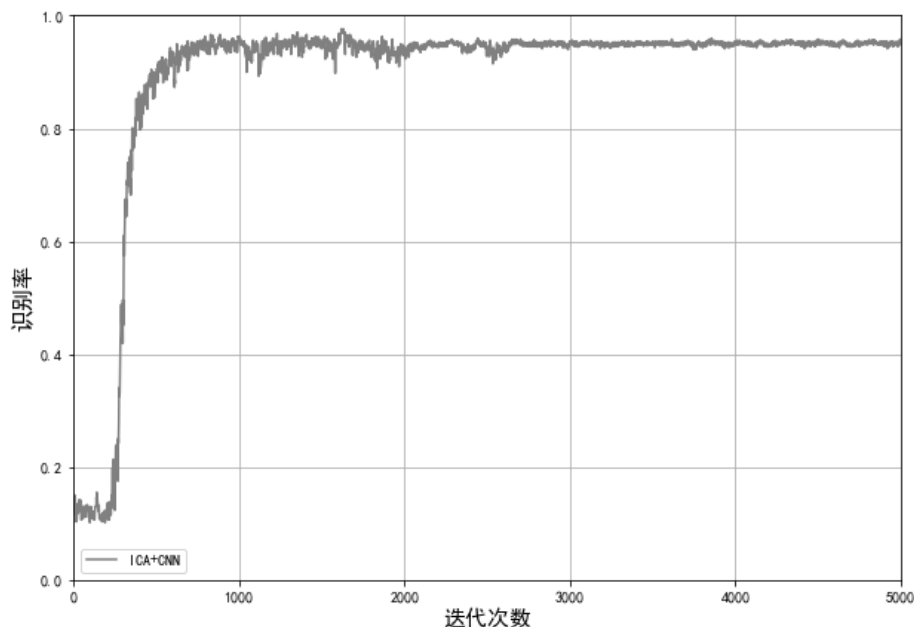


图 4-4 识别率变化曲线

(2) 为了验证本文方法的合理性和有效性, 采用了 3 种不同的手写体汉字识别方法与之做比较。第一种为基于 CNN 的手写体汉字识别方法, 该方法直接将图像作为 CNN 的输入, 简称 CNN; 第二种为文献[80]中的 HOG 与 CNN 相结合的方法, 简称 HOG+CNN; 第三种为文献[73]提出的基于多尺度梯度及深度神经网络的汉字识别方法, 简称 MGD+CNN, 本文方法简称 ICA+CNN。

为了更直观的看出每种方法的识别率变化, 本文从 10 组相似手写体汉字中随机选择一组数据进行实验, 分别与上述 3 种方法进行比较, 实验结果如图 4-5 所示。此时 CNN 的识别率为 92.87%, HOG+CNN 的识别率为 91.26%, MGD+CNN 的识别率为 94.03%, 本文方法的识别率为 95.80%, 比较实验结果可以得出, 本文方法具有较高的识别率。

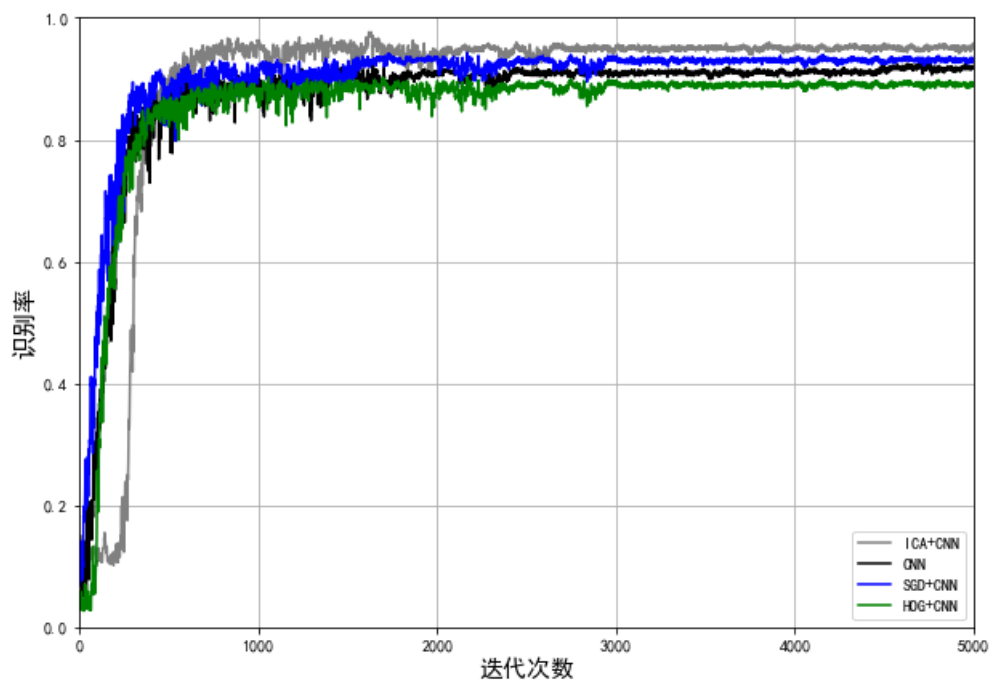


图 4.5 比较不同手写体汉字识别方法的识别率

为了避免识别结果的偶然性，本文对 10 组相似手写体汉字进行实验，并求出 10 组相似手写体汉字平均识别率，通过平均识别率比较 3 种算法与本文算法的性能。另外，三种方法在每一组相似手写体汉字上均采用上述 CNN 模型进行训练和识别，由此得到的测试集的识别率结果如表 4-1 所示。（每组训练集和测试集都包含了 240 张训练图片和 60 张测试图片）。

表 4-1 三种手写体汉字识别方法的正确率比较

方法 组别	CNN	HOG+CNN	MGD+CNN	本文方法
第一组	93.50%	90.60%	93.45%	95.60%
第二组	94.70%	91.80%	92.20%	93.70%
第三组	91.00%	91.20%	91.80%	93.40%
第四组	92.05%	89.69%	94.92%	96.80%
第五组	91.08%	92.06%	92.76%	92.20%
第六组	92%	90.80%	93.10%	93.40%
第七组	92.30%	91.96%	92.45%	94.47%
第八组	91.60%	90.18%	93.36%	93.85%
第九组	94.88%	90.58%	90.96%	94.64%
第十组	92.00%	89.69%	91.27%	93.70%
平均值	92.49%	90.86%	92.63%	94.18%

从表 4-1 可以看出，本文方法在每组相似手写体汉字得出的识别率，与基于 CNN 的手写体汉字识别方法做比较，本文方法在多组相似汉字上的识别率都比较高，且平均

识别率提高了 1.69%；与文献[73]中 HOG+CNN 的方法作对比，本文方法的平均识别率提高了 3.32%；与文献[80]中 MGD+CNN 的方法作对比，本文方法的平均识别率提高了 1.55%。

由此可知，在 CNN 识别前加入不同的特征信息，对手写体汉字的识别结果不一定是促进的，加入 HOG 特征可能对 CNN 的识别结果有抑制作用。MGD 是对 HOG 提取特征的改进，将 MGD 与 CNN 结合，识别率确有提升，但是仍小于本文方法的识别率。该实验证明了本文方法的合理性和有效性。

另外，本文采用的 CNN 模型仅有 13 层，属于轻型的神经网络结构，如果在更深、更宽的网络结构上进行训练和测试，手写体汉字的识别率将会进一步得到提升。

4.4 本章小结

ICA 是最近几年发展起来的一种线性变换方法，通过对汉字图像进行 ICA 变换，可以得到一种新的数据表示。CNN 是一种端到端的方法，近几年来在手写体汉字识别领域比传统方法好的深度学习模型主要是基于 CNN 及其改进方法的。为了验证 ICA 特征用于 CNN 分类可以提高手写体汉字的识别性能，本文提出了一种基于 CNN（CNN）和独立成分分析（ICA）相结合的手写体汉字识别方法。

虽然加入 ICA 提取的汉字特征用于 CNN 分类，提高了手写体汉字的识别率，但是识别手写体汉字所花费的时间也有了相应的提高。因此，今后的研究重点将集中在减少时间复杂度的同时，增加手写体汉字的识别率。

第五章 总结与展望

5.1 总结

由于近年来科学技术的发展,为 CNN 识别手写体汉字识别提供了技术保障,近年来, CNN 成为了识别手写体汉字的主流方法。为了提高手写体汉字的识别效果,研究者提出了很多 CNN 模型,如近几年提出的 GoogleNet 模型和 ResNet 模型,都促使了手写体汉字识别率的提升。

通过研究 CNN 的各种模型及 CNN 在汉字识别中的应用发现,与传统的手写体汉字识别相比, CNN 具有自动提取特征和识别率高的优点。然而,基于 CNN 的手写体汉字识别方法虽然有了巨大的进步和成功的应用,但也存在很多问题,比如需要大量的数据和更深的网络,识别花费时间较长,相似手写体汉字的识别率仍然有待提高等。针对时间较长和识别率不高的问题,本文通过两方面进行探索研究。第一,由于现在大多数的改进只考虑在 CNN 模型上进行调整和修改,而未使用任何领域特征信息。传统的特征提取方法有很多,利用这些特征提取方法提取特征信息,和原图像一起作为 CNN 的输入信息进行分类,是否可能会提高手写体汉字的识别率。第二, CNN 识别手写体汉字需要的特征数量相当大(图像的特征维度可能达到几百甚至几千),因此,直接存储和处理图像时效率比较低。在实际的应用中,大量的数据特征并不能刻画数据的本质特征,数据的冗余还可能会影响到后续的数据处理。在不影响识别率的情况下去除这些冗余信息,可以提取出更高效的特征信息,提升 CNN 的运算速度。因此,为了提高手写体汉字的识别速度和识别精度,本文在 CNN 识别汉字之前,通过引入不同的特征,达到了识别的目的。其主要工作概括如下:

(1) 针对 CNN 识别手写体汉字识别速度慢的问题,将 2DPCA 与 CNN 相结合识别手写体汉字。首先,利用 2DPCA 提取手写体汉字的投影特征向量;其次,将手写体汉字的投影特征向量组成特征矩阵;然后,用组成的特征矩阵作为 CNN 的输入;最后,用 Softmax 函数进行分类。实验结果表明,该方法在不降低识别精度的同时,可以减少识别手写体汉字的运行时间。

(2) 为了证明传统的特征提取方法——独立成分分析有助于优化 CNN 的模型,提高

手写体汉字的识别率，提出了一种基于独立成分分析和 CNN 相结合的手写体汉字识别方法。首先，利用 FastICA 算法提取手写体汉字的独立基图像和投影向量，得到特征向量；其次，将 FastICA 提取的特征向量与原图像一起放入 CNN 的输入层进行分类。该方案充分利用了 ICA 的特征提取能力和 CNN 的分类能力。从实验中可以看出，加入在汉字图像中加入 ICA 特征，识别率有了一定的提升。

5.2 展望

CNN 在手写体汉字识别方面的研究是模式识别的热点问题。本文对手写体汉字的识别研究也取得了一些成果，不过，仍有一些缺陷需要改进。首先，在 CNN 中加入 ICA 特征虽然在一定程度上提高了手写体汉字的识别率，但是该方法也伴随着 CNN 运行时间的增加，且识别率提升有限；其次，将 2DPCA 提取的特征图作为 CNN 的输入，虽然对手写体汉字的速度得到了极大的提升，但是随着迭代次数的增加，采用特征图作为 CNN 的输入，识别率的波动不如原图像的平稳，中间会有小幅度的波动。因此，今后的研究重点为：

(1) 加入 ICA 特征可以优化 CNN 模型，但是为了得到更好的识别结果，需要对图像特征提取方法进一步的研究。对一些好的特征提取方法进行归纳总结，并通过实验评估各种特征提取方法与 CNN 模型的契合度，寻找一种可以大幅度提高手写体汉字识别率的方法。

(2) 现在的手写体汉字识别方法已经取得了很大的进步，但是 CNN 识别率的提升通常需要大量的数据集和很深的网络模型。然而，无论是增加数据集还是加深网络模型，都会导致运行时间和存储容量的增加，且当 CNN 达到一定深度时，会达到瓶颈，进而引发神经网络退化和过拟合等问题。因此，下阶段的重点会放在对 CNN 结构的优化上面。

(3) CNN 中包含很多超参数信息，如学习率、dropout 等。目前，调整超参数基本都是依赖于已有的学习经验，因此，如何科学的调参也是个值得思考和研究的问题。

参考文献

- [1] Wang T W, Xie Z C, Li Z, et al. Radical aggregation network for few-shot offline handwritten Chinese character recognition[J]. Pattern Recognition Letters, 2019, 821-827.
- [2] Zhou M K, Zhang X Y, Yin F, et al. Discriminative quadratic feature learning for handwritten Chinese character recognition[J]. Pattern Recognition, 2016, 49: 7-18.
- [3] 喻方元. 导出核学习算法及其在手写体汉字识别上的应用[J]. 湖北汽车工业学院学报, 2016, 30(04): 74-76.
- [4] Zhang X, Bengio Y, Liu C, et al. Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark[J]. Pattern Recognition, 2017, 61(61): 348-360.
- [5] Zhu Y, Zhuang F, Yang J, et al. Adaptively Transfer Category-Classifer for Handwritten Chinese Character Recognition[M]// Advances in Knowledge Discovery and Data Mining. 2019.
- [6] 张秀玲, 周凯旋, 魏其珺, 董逍鹏. 多通道交叉融合的深度残差网络脱机手写汉字识别[J]. 小型微型计算机系统, 2019, 40(10): 2232-2235.
- [7] 朱程辉, 甘恒, 王建平. 基于 FSVM 脱机手写体汉字分类识别研究[J]. 计算机工程与应用, 2014, 50(23): 189-193.
- [8] 王芳元, 王建平. 仿反馈机制的智能识别方法研究与应用[J]. 合肥工业大学学报(自然科学版), 2018, 41(03): 333-341.
- [9] Liu X, Hu B T, Chen Q C, et al. Stroke Sequence-Dependent Deep Convolutional Neural Network for Online Handwritten Chinese Character Recognition[J]. IEEE transactions on neural networks and learning systems, 2020: 1-12.
- [10] 毛晓波, 程志远, 周晓东. 基于特征图叠加的脱机手写体汉字识别[J]. 郑州大学学报(理学版), 2018, v.50(03): 81-85.
- [11] 柴伟佳, 王连明. 卷积神经网络的多字体汉字识别[J]. 中国图象图形学报, 2018, 23(03): 410-417.
- [12] 蹇宇澄, 刘昭策. 深度学习的实现与发展——从神经网络到机器学习[J]. 电子技术与软件工程, 2017(11): 30-31.
- [13] 赵星宇, 丁世飞. 深度强化学习研究综述[J]. 计算机科学, 2018, 45(07): 1-6.
- [14] 陈站, 邱卫根, 张立臣. 基于一种改进 Inception 的脱机手写汉字识别[J/OL]. 计算机应用研究: 1-4[2020-04-06]. <https://doi.org/10.19734/j.issn.1001-3695.2018.09.0784>.
- [15] 赵晓娟. 手写体数字及英文字符的识别研究[D]. 东北师范大学, 2010.
- [16] Casey R, Nagy G. Recognition of Printed Chinese Characters[J]. IEEE Transactions on Electronic Computers, 1966, 15(1): 91-101.
- [17] J.-H. Liu. real time Chinese handwriting recognition[D]. University of Pittsburgh, Pennsylvania,

1966.

- [18] M.J. Zobrak, A method for rapid recognition of hand drawn line patterns[D]. University of Pittsburgh, Pennsylvania, 1966.
- [19] Wang Y, Ding X, Liu C.MQDF Discriminative Learning Based Offline Handwritten Chinese Character Recognition[J]. 2011:1100-1104
- [20] 朱小燕, 史一凡, 马少平. 手写体字符识别研究[J]. 模式识别与人工智能, 2000(2):342-346
- [21] 肖斌. 基于 SVM 的脱机手写体汉字识别研究[D]. 成都, 西华大学, 2009.
- [22] 高彦宇, 杨扬. 脱机手写体汉字识别研究综述[J]. 计算机工程与应用, 2004(07):74-77.
- [23] H. Yamada, K. Yamamoto, T. Saito, A nonlinear normalization method for handprinted Kanji character recognition-line density equalization, Pattern Recognit. 24 (9) (1990) 1023–1029.
- [24] C.-L. Liu, K. Marukawa, Pseudo two-dimensional shape normalization methods for handwritten Chinese character recognition, Pattern Recognit. 38(12)(2005) 2242–2255.
- [25] He Z, Zhong Y, Cao Y. High Accuracy Handwritten Chinese Character Recognition Based on Support Vector Machine and Independent Component Analysis[M]// Informatics and Management Science V. Springer London, 2013.
- [26] Kobayashi K, Yoda F, et al. Recognition of Handprinted Kanji Characters by the Stroke Matching Method[J]. Pattern Recognition Letters, 1983, 1(5-6): 481-488.
- [27] 高学, 金连文, 尹俊勋. 基于笔划密度的弹性网络特征提取方法[C]. 中国自动化学会. 2001 年中国智能自动化会议论文集(上册). 广州, 华南理工大学电子与通信工程系, 2001.470-475.
- [28] Rong C, Hong-Bo D, Lian-Wen J. Handwritten Chinese character recognition based on local gabor filter bank[J]. journal of computer applications, 2007, 27(5):1222-1224.
- [29] 黄璞, 陈才扣. 基于局部人脸图像的 ICA 人脸识别方法[J]. 计算机工程与设计, 2010, 31(11):2550-2553.
- [30] Liu C L. Normalization-Cooperated Gradient Feature Extraction for Handwritten Character Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(8):1465-1469.
- [31] Xu Y, Wang L. K-nearest Neighbor-based Weighted Twin Support Vector Regression[M]. Applied Intelligence, 2014, 41(1): 299-309.
- [32] 徐小丽, 范程华, 樊敏. 一种动态剪枝二叉树 SVM 的脱机手写体汉字识别方法研究[J]. 合肥师范学院学报, 2018, 36(03):11-14.
- [33] 蔺志青, 郭军. 贝叶斯分类器在手写汉字识别中的应用[J]. 电子学报, 2002(12):1804-1807.
- [34] Long T, Jin L. Building Compact MQDF Classifier for Large Character Set Recognition by Subspace Distribution Sharing[J]. Pattern Recognition, 2008, 41(9): 2916-2925.
- [35] 金连文, 钟卓耀, 杨钊, 等. 深度学习在手写汉字识别中的应用综述[J]. 自动化学报, 2016, 42(8):1125-1141.
- [36] Ciresan D C, Meier U, Gambardella L M, et al. Convolutional neural network committees for

- handwritten character classification[C]//Proceedings of 2011 International Conference on Document Analysis and Recognition. Beijing, China:IEEE, 2011: 1135-1139.
- [37] Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providen, RI:IEEE, 2012:3642-3649.
- [38] Yin F, Wang Q F, Zhang X Y, Liu C L. ICDAR 2013 Chinese handwriting recognition competition[J]. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, DC, USA: IEEE, 2013.1464-1470.
- [39] Zhiyuan Li, Qi Wu, Yi Xiao, Min Jin, Huaxiang Lu. Deep Matching Network for Handwritten Chinese Character Recognition[J]. Pattern Recognition, 2020.
- [40] Chen L, Wu C, Fan W, et al. Adaptive local receptive field convolutional neural networks for handwritten Chinese character recognition[C]//Chinese Conference on Pattern Recognition. Springer, Berlin, Heidelberg, 2014: 455-463.
- [41] Yang X, He D, Zhou Z, et al. Improving offline handwritten Chinese character recognition by iterative refinement[C]//2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2017, 1: 5-10.
- [42] Bi N, Chen J, Tan J. The Handwritten Chinese Character Recognition Use Convolutional Neural Networks with the GoogLeNet[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(11):1940016.
- [43] Liu C L, Yin F, Wang D H, et al. Online and offline handwritten Chinese character recognition: benchmarking on new databases[J]. Pattern Recognition, 2013, 46(1): 155-162.
- [44] 赵秀红. 基于主成分分析的特征提取的研究[D].西安, 西安电子科技大学, 2016.
- [45] 赵继印, 郑蕊蕊, 吴宝春, 等. 脱机手写体汉字识别综述[J]. 电子学报, 2010, 38(2): 405-415.
- [46] Liu C L. Normalization-Cooperated Gradient Feature Extraction for Handwritten Character Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(8):1465-1469.
- [47] Zhou M K, Zhang X Y, Yin F, et al. Discriminative quadratic feature learning for handwritten Chinese character recognition[J]. Pattern Recognition, 2016, 49(1):7-18.
- [48] He Z, Zhong Y, Cao Y. High Accuracy Handwritten Chinese Character Recognition Based on Support Vector Machine and Independent Component Analysis[M]// Informatics and Management Science V. Springer London, 2013.
- [49] 王建平, 徐奇. 基于过程神经网络的汉字笔段提取算法研究[J]. 计算机应用研究, 2010, 27(05):1965-1967.
- [50] Wen Y, Lu Y, Yan J, et al. An Algorithm for License Plate Recognition Applied to Intelligent Transportation System[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(3):830-845.
- [51] Su Q, Huang Z, Hu Z, et al. Binarization algorithm based on differential evolution algorithm for

- gray images[C]. fuzzy systems and knowledge discovery, 2012: 2611-2615.
- [52] Zheng L , He X , Samali B , et al. An algorithm for accuracy enhancement of license plate recognition[J]. Journal of Computer and System Sciences, 2013, 79(2):245-255.
- [53] 李了了, 邓善熙, 丁兴号.基于大津法的图像分块二值化算法[J].微计算机信息, 2005 (14): 76-77.
- [54] 辛迈,王平,曾瑞鹏,耿健,冯天保.基于线性插值的图像分块分割算法[J].工业仪表与自动化装置,2016(01):118-120.
- [55] 龙钧宇,金连文.一种基于全局均值和局部方差的图像二值化方法[J].计算机工程,2004(02):70-72.
- [56] Chen Q, Sun Q S, Heng P A, et al. A double-threshold image binarization method based on edge detector[J]. Pattern Recognition, 2008, 41(4):1254-1267.
- [57] 张洁玉.基于图像分块的局部阈值二值化方法[J].计算机应用,2017,37(03):827-831.
- [58] 王骏飞,黄樟灿,Ahmed Mahgoub Ahmed Talab.一种针对图像细节优化的 BM 二值化方法[J].武汉理工大学学报,2014,36(08):127-132.
- [59] 申森,李艾华,姚良,马治明.基于小波包和 Niblack 法的枪号图像二值化算法[J].光子学报,2013,42(03):354-358.
- [60] Zhao-Xia T. Image Segmentation Algorithm based on Improved GA and OTSU[J]. Jilin Normal University Journal(Natural Science Edition), 2011,1:70-72.
- [61] 董庆霞. 图像处理技术在指示表精度检定中的应用[D]. 北京, 北京工业大学,2001.
- [62] 刘畅,钱盛友,邹孝,孙福成,张激.基于阈值法和数学形态学的 HIFU 治疗区域确定[J].计算机工程与应用,2009,45(23):224-226.
- [63] 姚玉花,成渝.一种身份证扫描图像的二值化方法[J].机械制造与自动化,2002(05):65-68.
- [64] 张善文,黄文准,师韵.基于改进 Bernsen 二值化算法的植物病害叶片病斑检测[J].广东农业科学,2016,43(12):129-135.
- [65] 一种基于新型小波包阈值的图像去噪方法[J]. 宁波大学学报(理工版), 2009, 022(004):454-458.
- [66] 张德保,刘普吉,姜寿山.激光条纹图像处理算法的研究[J].微型机与应用,2011,30(02):47-50.
- [67] 严文培.基于 MatLab 不同平滑处理对边缘检测效果的影响[J].现代计算机,2013(25):39-43.
- [68] 张仲一,杨成,吴晓雨.基于 Kinect 的隔空人手键盘输入[J].中国传媒大学学报(自然科学版),2013,20(03):68-71.
- [69] 周彬彬.复杂环境下车牌识别系统中关键技术的研究与实现[D]. 南昌, 南昌航空大学, 2016.
- [70] 王庆,赵荣椿.手写体汉字的规范化处理及评价[J].数据采集与处理,2001,16(2): 227-232.
- [71] 刘妍. 基于 Gabor 双弹性网格特征提取的手写体汉字识别的研究[D]. 天津, 河北工业大学,2015.
- [72] Qian Y, Ying L, Pingping J. Speech emotion recognition using supervised manifold learning

- based on all-class and pairwise-class feature extraction[C]//2013 IEEE Conference Anthology. IEEE, 2013.
- [73] 潘炜深,金连文,冯子勇.基于多尺度梯度及深度神经网络的汉字识别[J].北京航空航天大学学报,2015,41(04):751-756.
- [74] 叶锋,邓衍晨,汪敏,廖茜,郑子华,林晖.部分级联特征的离线手写体汉字识别方法[J].计算机系统应用,2017,26(08):134-140.
- [75] 赵继印, 郑蕊蕊, 吴宝春,等. 脱机手写体汉字识别综述[J]. 电子学报, 2010, 38(2): 405-415.
- [76] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// Proceedings of the 2012 Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, USA: Curran Associates, Inc, 2012: 1097-1105.
- [77] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. computer vision and pattern recognition, 2014.
- [78] Szegedyc, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]//Proceedings of the IEEE Conference on computer vision and pattern recognition.IEEE,2015:1-9.
- [79] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C]// Proceedings of the IEEE Conference on computer vision and pattern recognition.IEEE, 2016:770-778.
- [80] Zhong Z, Jin L, Xie Z. High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps[C]// 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015.
- [81] 李彦冬,郝宗波,雷航.卷积神经网络研究综述[J].计算机应用,2016,36(09):2508-2515.
- [82] Liu K, Cheng Y Q, Yang J Y. Algebraic feature extraction for image recognition based on an optimal discriminant criterion[J]. Pattern Recognition, 1993, 26(6): 903-911.
- [83] Yang J. From image vector to matrix: A straightforward image projection technique—IMPCA vs. PCA[J]. Pattern Recognition, 2002, 35(9): 1997-1999.
- [84] 孙瑞霞,汪亚明,黄文清.基于 GA 的 2DPCA 在人脸识别中的应用[J].计算机工程与设计,2007,28(10):2398-2400.
- [85] 陈伏兵,陈秀宏,高秀梅,等.二维主成分分析方法的推广及其在人脸识别中的应用[J].计算机应用,2005,25(08):1767-1770.
- [86] 胡娜,马慧,湛涛.融合 LBP 纹理特征与 B2DPCA 技术的手指静脉识别方法[J].智能系统学报,2019,14(03):533-540.
- [87] 闫荣华,彭进业,汶德胜.PCA 与 2DPCA 的关系[J].计算机科学,2017,44(S2):202-206.
- [88] 覃朝勇,郑鹏,张骁.基于 MQDF-DBM 模型的脱机手写体汉字识别[J/OL].计算机工程与应用: 1-8[2019-12-03]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20190326.1831.027.html>.

致 谢

时光如剑，岁月如梭，美好的研究生学习生活已经进入尾声。回首研究生生活，不得不感叹时光的流逝。在这三年中，河南师范大学以其优良的学习风气、严谨的科研氛围教我求学，以其博大包容的情怀胸襟、浪漫充实的校园生活育我成人，使我收获良多。临毕业之际，在此对那些曾经给予我帮助的老师 and 同学们表示真诚的感谢。

本学位论文是在导师郑延斌教授的亲切关怀和悉心指导下完成的。郑老师严谨细致、一丝不苟的作风一直是我工作、学习中的榜样，虽身负教学、科研重任，郑老师仍一直关心论文的进展情况，他循循善诱的教导和不拘一格的思路给了我很多启迪。正是郑老师的无私帮助与鼓励，我的毕业论文才能顺利完成。在此，向郑延斌老师表示崇高的敬意和由衷的感谢！

同时，我也要感谢研一期间教导我的任课老师，你们的教导为我以后的科研学习打下了夯实的基础；我还要感谢我的校外实习老师刘永战老师，在他的带领下，我学习到了很多书本上没有的知识；感谢史春华老师和徐天鹤老师在我入党过程中提供的帮助。

在这三年的研究生期间，505 实验室为我提供了良好的学习氛围和空间。在实验室中度过的每一天，都令我终生难忘，是我人生中最值得回忆的美好时光。在实验室做项目和撰写论文期间，王林林和席鹏雪师姐提了很多有用的建议，给了我很多帮助。在研二期间，和樊文鑫同学一起们花了大半年的时间做项目，由于我自学能力较差，在项目期间，我的同门樊文鑫对我帮助巨大，教会了我很多编程知识。通过这个项目使我的专业知识得到了很大的提升。

另外最应该感谢的是我的父母和室友们，正是他们默默的支持、理解和奉献，才使我得以顺利完成学业；正是他们对我无私的爱，使我有勇气和动力克服生活和学习中所遇到的各种困难，不断进步才取得了今天的成绩。父母对我的爱我无以为报，在即将完成学业踏入工作岗位之际，我把最真诚的祝愿送给我的父亲母亲，祝他们身体健康！

最后，感谢参加论文评选和答辩的各位专家老师！

韩梦云

2020 年 6 月

攻读硕士学位期间的科研成果

发表论文：

[1]郑延斌,韩梦云,樊文鑫. 基于 2DPCA 与 CNN 的手写体汉字识别[J/OL].计算机应用 : 1-8[2020-05-06].<http://kns.cnki.net/kcms/detail/51.1307.TP.20200429.1626.006.html>.

[2]郑延斌,樊文鑫,韩梦云,陶雪丽.基于博弈论及 Q 学习的多 Agent 协作追捕算法[J/OL]. 计 算 机 应 用 :1-10[2020-03-28].<http://kns.cnki.net/kcms/detail/51.1307.TP.20200109.1438.026.html>.

[3]郑延斌,王林林,席鹏雪,樊文鑫,韩梦云.基于蚁群算法及博弈论的多 Agent 路径规划算法[J].计算机应用,2019,39(03):681-687.

[4]郑延斌,席鹏雪,王林林,樊文鑫,韩梦云.基于人工势场法的多智能体编队避障方法[J].计算机应用,2018,38(12):3380-3384+3413.

[5]郑延斌,王林林,席鹏雪,樊文鑫,韩梦云.动态环境下改进蚁群算法的多 Agent 路径规划[J].计算机工程与科学,2019,41(06):1078-1085.

[6]郑延斌,席鹏雪,王林林,樊文鑫,韩梦云.基于模糊人工势场法的多智能体编队控制及避障方法[J].计算机工程与科学,2019,41(08):1504-1511.

