

分类号_____

学校代码 10487

学号_____

密级_____

華中科技大學

硕士学位论文

基于深度卷积神经网络的离线笔迹
鉴别研究

学位申请人:

学 科 专 业: 电路与系统

指 导 教 师:

答 辩 日 期:

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering**

**Research on Offline Writer Identification Based
on Deep Convolutional Neural Networks**

**Candidate :
Major : Circuit and System
Supervisor :**

**Huazhong University of Science & Technology
Wuhan 430074, P.R. China
May, 2016**

华中科技大学硕士学位论文

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到，本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐ ，在年解密后适用本授权书。
本论文属于 不保密 ☐ 。

（请在以上方框内打“√”）

学位论文作者签名：

日期： 年 月 日

指导教师签名：

日期： 年 月 日

摘要

笔迹鉴别作为一项重要的生物特征识别技术，在今天有着越来越广泛的应用，众多的研究者在笔迹鉴别这一问题上做出了突破性的成果。然而非受限情况下的离线笔迹鉴别仍是一项非常有挑战性的研究工作，有很多问题值得我们进一步深入研究。

深度卷积神经网络近年来蓬勃发展，解决了许多计算机视觉领域的难题，然而在离线中文笔迹鉴别领域却几乎没有看到这方面的研究。本文着重研究文本无关的离线中文笔迹鉴别，并把深度卷积神经网络应用于这一研究课题，取得了具有领先优势的笔迹鉴别效果。本文的主要研究内容如下：

首次把深度卷积神经网络提取的局部特征用于离线中文笔迹鉴别，并研究不同的特征编码方式，采用 Fisher Vectors 算法对已有的局部特征进行编码，进而生成全局的特征。实验表明，本文所采用的基于深度卷积神经网络的特征相比于传统的人工设计的特征更具有区分性。

研究不同网络结构，对网络结构进行不断地优化调整，从而达到更好地笔迹鉴别效果。研究深度卷积神经网络的训练，提出了基于滑动窗字符切割方法，评估笔迹图片归一化和数据扩充对最终鉴别结果的影响。为了选择出更有区分性的特征，采用了主成份分析对深度卷积神经网络提取的特征进行处理，并且对不同网络层的特征做了比较。

通过大量细致的实验对本文提出的方法进行评估，首先在中文数据集 CASIA-HWDB 上面进行验证，取得了领先的结果。接着在包含两种语言的外文数据集 ICDAR 2013 上面进行验证，也取得了有竞争力的结果。结果表明本文算法不仅对于中文笔迹鉴别有很好的效果，而且对于跨语言的笔迹鉴别也具有一定的优势，有很强的适应性和鲁棒性。

关键词：笔迹鉴别，深度卷积神经网络，特征提取，主成份分析，特征编码

Abstract

Writer identification, considered as an important biometric identification technology, been widely used nowadays. Although many researchers have made breakthroughs on this field, the challenge of offline text-independent writer identification still remains, and there are many problems worth our further study.

Deep Convolutional neural networks has been booming in recent years, and has solved many tough problems in computer vision, but little research has been reported on offline Chinese writer identification employing Deep convolutional neural networks. This paper focuses on offline text-independent Chinese writer identification and applies convolutional neural networks to this topic, which produces prominent results. The major contributions of the thesis are as follows:

Above all. It's the first time for use deep convolutional neural networks to extract local features used for offline Chinese writer identification, and study several kinds feature encoding methods, and a global descriptor is then formed by means of Fisher Vectors encoding. Experiments show features extracted form convolutional neural networks are more distinctive than traditional expert-designed features.

Then, several network structures is studied and optimized, so as to achieve better writer identification results. And the training of convolutional neural networks is probed into, a character segmentation method based on sliding window is proposed, and the effect of handwriting image normalization and data augmentation are evaluated to eventually achieve identification accuracy. In order to choose the more distinctive features, principal component analysis is adopted to select the feature extracted from convolutional neural networks, and the features of different network layers are compared.

Lastly, In order to evaluate the proposed method a lot of careful experiments have been done. First, the experiments are conducted on Chinese dataset CASIA-HWDB, and produces the state of art results. Then, our method is tested on dataset ICDAR 2013 which contains two foreign languages, and also yields competitive results. The results reveal that the proposed method not only has good performance on Chinese writer identification, but also has a certain advantage for cross-language handwriting identification, and has strong adaptability and robustness.

Key words: Offline writer identification, Deep convolutional neural networks, Feature extraction, Principal component analysis, Feature Encoding

目 录

摘 要.....	I
Abstract.....	II
1 绪论	
1.1 研究背景及意义	(1)
1.2 笔迹鉴别的分类	(2)
1.3 国内外的研究现状	(3)
1.4 本文的主要研究内容与组织结构	(6)
2 相关理论基础	
2.1 神经网络概述.....	(8)
2.2 深度学习简介.....	(11)
2.3 深度卷积神经网络	(11)
2.4 笔迹鉴别系统.....	(14)
2.5 本章小结.....	(15)
3 基于深度卷积神经网络的离线笔迹鉴别研究	
3.1 本文笔迹鉴别系统流程	(16)
3.2 深度卷积神经网络结构	(18)
3.3 深度卷积神经网络训练	(21)
3.4 深度卷积神经网络特征的选择	(26)
3.5 DCNN 特征的编码研究	(29)

3.6 分类器.....	(32)
3.7 本章小结.....	(34)
4 实验结果与分析	
4.1 笔迹鉴别评价标准	(35)
4.2 数据集介绍.....	(35)
4.3 本文方法对比实验	(38)
4.4 与其他方法的比较	(45)
4.5 实验平台及时间	(48)
4.6 本章小结.....	(49)
5 工作总结与展望	
5.1 全文总结.....	(50)
5.2 工作展望.....	(50)
致 谢.....	(52)
参考文献.....	(53)

1 绪论

1.1 研究背景及意义

笔迹鉴别,是通过手写文字信息来确认书写者身份的一种文件检验技术。它与人脸识别、指纹识别和虹膜识别等一样属于自动身份识别的研究领域。笔迹鉴别系统必须能够完成“一对多”的识别任务,即在给定一个笔迹材料库的基础上,输入一张笔迹材料,能够反馈与输入笔迹材料最接近的样本,从而确定输入材料书写人的身份信息。

人工智能和模式识别技术日新月异的发展,对笔迹鉴别这一颇具挑战性的问题起到了极大的推动作用。随着科学技术的发展与广泛的应用,笔迹鉴别在成为了一项越来越不可或缺的技术。笔迹鉴别的应用领域非常广泛,在金融领域,笔迹鉴别能够用来确认文件的真实性。在司法领域,笔迹鉴别能够解决犯罪学中的专业问题,能够应用在法医专家决策系统中。人工笔迹鉴定需要比较单个字的笔画的细微的差别,是一个非常耗时的过程,并且容易引入主观因素。通过笔迹鉴别系统能够提供一一个缩小的候选范围,极大地减少了工作量而且能够增加鉴别结果的可靠性。这一技术还可以应用在认证系统之中,可以用来监控对于机密网站或者机密数据的访问。这些机密网站通常会存储和管理诸如文本、表格、笔记、会议记录等资料,获取作者的身份往往能够提供重要的价值。笔迹鉴别还能够用在历史文本的分析^[1]、笔迹识别系统和手持设备上^[2]。非常确定的是随着近年来笔迹鉴别技术的发展和越来越高的准确度,笔迹鉴别已经成为了和 DNA 鉴定和指纹鉴别一样的非常重要的生物特征识别技术^[3]。

笔迹鉴别在今天的社会扮演着越来越重要的角色,有越来越多的研究者投入到这个科研领域之中。一个重要的原因是由于笔迹鉴别的需求激增,笔迹鉴别技术可以帮助鉴定司法文件从而来锁定罪犯^[4]。特别是当今社会恐怖主义抬头,笔迹鉴别和其他的生物识别技术结合起来用于发现恐怖分子^[5]。笔迹鉴别是根据书写人的书写习惯在文字中形成的各不相同的特征,来鉴别书写人的身份的。世界上的语言文字有成千上万种,而由于每种文字的轮廓、笔迹都有自己独特的特征,因此笔迹鉴定在不同的语言文字上面有不同的难点。

中国书法是一门古老的艺术,笔迹鉴别在我国更是能够发挥其独特的意义。中

国的书法文化博大精深，形成了草书，行书，楷书，隶书，小篆等书写风格。又由于中文文字的特殊性，使得文字笔画的轨迹变化极其繁复，可以说基本上找不到两个人的笔迹是完全一样的，这也为中文的笔迹鉴定提供了比较好的前提条件。笔迹鉴别在中国已经有很长时间的历史，在古代一个人的手书可以作为信用凭证，可以作为刑事侦查的证据。在今天，应用模式识别、人工智能的新技术和笔迹鉴定这一古老的艺术相结合，必定能够发挥出更大的作用。

1.2 笔迹鉴别的分类

通过笔迹材料来判别书写人身份有两种方式，分别为笔迹鉴别(Writer Identification)和笔迹验证(Writer Verification)。其中笔迹鉴别则是需要找出与给定笔迹材料最接近的样本。表 1-1 将从概念和预期目标对二者进行区分。

表格 1-1 笔迹鉴别和笔迹验证的不同

	笔迹鉴别	笔迹验证
定义	在给定笔迹材料库（已知书写人）的基础上，输入一张笔迹材料，从材料库中找出与输入笔迹材料最接近的样本。	通过比较两份给定的笔迹材料，考察二者的符合程度，从而判断两份材料是否为同一个人所书写。
预期目标	根据材料库的查询结果，得到一个已排序的候选人列表。从而确定输入材料书写人的身份信息。	通过距离度量法则来确定两份笔迹材料的距离，如果距离小于阈值，则说明两份材料是同一人书写，否则为不同人书写。

图 1-1 形象地说明了笔迹鉴别的概念，把一份待鉴别的笔迹（图左）材料输入笔迹鉴别系统（图中），笔迹鉴别系统通过提取输入材料的特征并和系统的笔迹材料数据库做比对，通过一定的相似度度量方法，从而锁定待查笔迹材料的书写人（图右）。

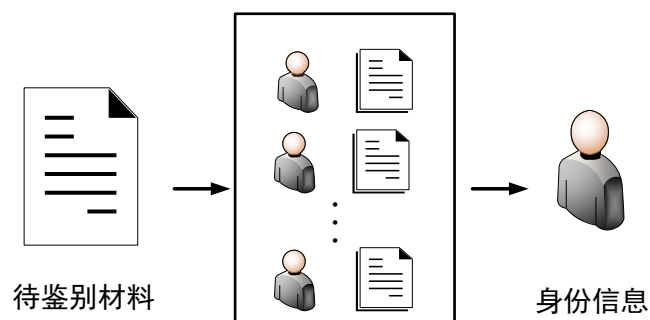


图 1-1 笔迹鉴别的定义

根据笔迹获取方式的不同，笔迹鉴定可以分为离线(Offline)笔迹鉴定和在线(Online)笔迹鉴定两类。在线笔迹鉴别的难度通常小一些，因为在线笔迹是通过专业的设备采集笔迹数据。能够采集包括书写的速度，角度，和书写的所用的压力以及写字的笔画顺序等丰富的信息。然而这些信息在离线的笔迹鉴别中是不能够获取的^[6]。离线笔迹材料就是用传统的书写工具在纸张上书写的，再通过照相机或扫描仪采集为图片的形式。由于离线笔迹材料的通用性和易获取性，离线笔迹鉴别具有更广泛的应用空间^[7]。

笔迹鉴别还可以进一步分为文本相关（Text-Dependent）和文本无关（Text-Independent）。文本相关的笔迹鉴别要求书写者提供内容一致的笔迹材料，这种方法一般需要人工来切割字符，然后再基于字符做鉴别。签名鉴定就是文本相关的笔迹鉴别的一个重要应用。而文本无关的笔迹鉴别对文本的内容没有要求，甚至不同语言之间的鉴别也属于这一范畴。因此文本无关的笔迹鉴别更加适合于实际应用，然而却很难获得前者那么高的准确率^[8]，本文主要研究文本无关的离线笔迹鉴别。

1.3 国内外的研究现状

由于不同的人书写风格不一样，笔迹材料的字形，字体，大小等必定会由于书写者的不同而有差异，笔迹鉴别就是要找出不同书写人的笔迹材料之间的差异。然而由于时间，书写人的状态，书写材质、书写环境等种种因素的变化，即使是同一个书写人的两份书写材料也会或多或少有所不同^[9]。这两个相互矛盾的因素使得笔迹鉴别成为了一项复杂的有挑战性的任务。在过去的几十年中，研究者在笔迹鉴定领域做出了许多有用的工作，提出了许多行之有效的方法。过去的研究主要集中在以英语为主的印欧语系，中文在这一方面还有一定的差距。文本无关的笔迹特征提

取方法主要有局部特征提取和全局特征提取。局部特征提取是对笔迹的局部结构、几何特征、轮廓、梯度等进行描述。全局特征提取通常是把笔迹材料作为图像来进行处理，进而提取其中的纹理信息。接下来，本文将分别介绍基于局部特征的和全局特征的笔迹鉴别的研究现状。

1.3.1 基于局部特征的笔迹鉴定

Jain 和 Doermann 提出了一种可重复的字符切割方法（Seam Cuts）来替代传统的字符切割方法^[10]。这种方法首先对笔迹材料提取连通区域，然后基于像素值和轮廓信息对于图像中每一个像素点算出一个“能量值”，通过寻找最小“能量值”来对连通区域进行划分。作者在这种字符切割方法基础上，提出的一种新颖的特征，这种特征使用轮廓梯度来描述字符的局部形状、笔画曲度。这个方法在 ICDAR 2013 笔迹鉴别比赛^[11]中取得了最好的结果。在 IAM 标准数据集上面的首选正确率达到了 96.5%，相比较于以前的研究错误率降低了 50%。

Jain 和 Doermann 基于之前的研究进一步提出了一种局部特征融合的方法。他们对三种特征进行了结合，从而取得了更加理想的效果^[12]。这三种特征分别是：K 临近分割（K-adjacent Segments）^[13]，这种特征能够很好地描述字符的笔画信息；SURF^[14]，能够提取笔迹图像的局部纹理和形状特征；轮廓梯度^[11]，提取笔迹图像的轮廓和曲度信息。实验中采用了交替训练的方法来确定三种特征的权重，例如，作者在 ICDAR 2013 实验数据集上测试，使用的是 ICFHR 2012 数据集来做训练，从而避免了测试集和训练集的交叠。实验结果表明这种融合特征的方法比传统的单一的特征更有区分性，能够取得比较领先的结果。

Nicolaou 和 Bagdanov 等人提出了一种对传统局部二值模式（LBP）改进的算法（RSR-LBP）^[16]。这种算法首先对图像的每个像素点提取 SRS-LBP 特征，然后过滤掉图像的前景只保留背景信息。接着对图像划分区，对于每个区间进行直方图统计，再对所有的直方图进行级联，形成图像的全局特征。最后还使用了 PCA 降维和 L2 归一化对特征进行了进一步的处理，这些处理已经被证明在图像识别特别是笔迹鉴别中有效^[17]。此方法是数据集 ICDAR 2013^[24]上比较领先的结果之一，并且是近年的研究中唯一采用密集（Dense）提取单一特征的笔迹鉴别方法。由于它简单有效，可以很方便得使用它和其他的方法结合来达到更理想的效果。

在国内，杨维信，金连文等人提出了一种字符级别的在线中文笔迹鉴别方法。

这种方法主要基于签字路径特征 (Path-signature Feature)、笔画丢弃 (Drop-stroke) 和深度卷积神经网络^[18]。签字路径特征是一种能够描述字符笔迹的特征, 将这种特征和深度卷积网络相结合能够达到全局的优化。所谓笔画丢弃, 就是随机删除一个字中的某些笔画, 从而极大得增加了训练数据样本数目, 并且能够提高深度卷积网络的泛化能力。文章所使用的是在线手写中文字符集 CASIA-OLHWDB1.0^[19], 这个数据集包含了 3866 个类, 作者随机挑选了 200 个类做训练, 3666 个类做测试, 保证了实验的文本无关性。然而, 由于离线笔迹鉴别难以准确获取字符的单一笔画, 这种方法并不直接适用于本文所研究的离线笔迹鉴别。

1.3.2 基于全局特征的笔迹鉴定

B. Helli 等人采用 Gabor 滤波器提取图像的纹理特征, 并且提出了一种新的分类方法用来做相似度计算^[20]。这种新的分类方法不同于一般的基于距离计算方法, 而是采用最长公共子序列 (LCS) 对已排序特征序列计算相似度。实验的数据包含 100 个书写人的 500 份笔迹材料, 平均准确率为 89%。B. Helli 等人在之前的基础上, 采用 Gabor 滤波器和 XGabor 滤波器相结合的方法, 对数据集中的笔迹图像提取了两组特征^[21]。进而通过特征之间关系生成特征关系图 (Feature Relation Graph, FRG), 这个有向无环的关系图也被称为 DAFRG (Directed Acyclic FRG)。在后续的分类阶段, 采用一种动态算法来计算两个 DAFRG 相似度。实验达到了 98% 的平均准确率。

鄢煜尘、陈庆虎等人提出了一种基于快速傅立叶变换的特征提取方法^[22]。根据手写材料纹理图像的构成原理, 采用一种估值方法来估计图像光谱特征的数学期望值。实验所使用的数据集有 100 个书写者, 每个书写者提供两份手写样本。采用欧氏距离和加权欧式距离 (Weighted Euclidian Distance, WED), 这种方法的 Top-10 的准确率为 98%, 首选准确率为 64%。这种方法的优势在于消除了光谱特征的随机性从而得到了相对稳定的特征, 并且对于大规模的数据集有较好的适应性。不足之处在于算法的计算量较大。

由于具有多尺度特征并且能够反映笔迹图像的细节信息, 小波变换在笔迹鉴别的研究中也有重要的应用。例如, 何震宇、尤新革等人提出了基于 Gabor 滤波器提取纹理特征, 并且结合小波域隐马尔可夫树的笔迹鉴别方法^[23]。先用小波变换对笔迹图像进行处理, 然后采用高斯混合模型 (Gaussian Mixture Model, GMM) 对小波系数进行建模。实验所使用的数据集包含 500 个书写者共 1000 份笔迹材料, 每份笔

迹材料包含 64 个中文字。首选准确率为 40%，Top-1，Top-10 的准确率分别为 82.4% 和 100%。

1.4 本文的主要研究内容与组织结构

本文采用理论分析和实验验证相结合的方法进行研究。首先通过分析前人在笔迹鉴别的特征提取、相似度度量方面的研究成果，结合实验深入研究已有算法，了解各种算法优势与不足。在此基础上，初步确定了自己的研究方案。通过大规模图像样本的学习，利用深度神经网络来提取笔迹图像的局部特征，在此过程中深入研究卷积神经网络的结构和其训练的改进方法。接着对卷积神经网络提取的特征进行选择，选取最具区分性的网络层的特征并去掉冗余的信息。而后对特征进行稀疏编码，形成进一步高阶的可分类的特征。最后研究选取了适合的相似性度量方法以及匹配准则，以达到最佳的笔迹鉴别效果。

本文的组织结构安排如下：

第一章首先介绍了笔迹鉴别的研究背景及意义，并指出了其应用领域与发展前景。简要介绍了笔迹鉴别的分类以及国内外研究现状，同时给出文章的主要研究内容与组织结构安排。

第二章介绍本文研究内容的相关理论基础。首先阐述了人工神经网络的发展及反向传播算法，这是基于深度卷积神经网络进行特征提取的理论基础，接着介绍了最近几年蓬勃发展的深度卷积神经网络。最后介绍了常见的笔迹鉴别系统，描述了笔迹鉴别的整体流程。

第三章是本文的重点研究内容。首先对本文所采用的深度卷积神经网络的网络结构进行了描述。接着研究了字符切割方式、训练数据加强以及主成份分析(Principal Component Analysis, PCA)对笔迹特征的提取影响，从而确定了怎样获得更具区分性的笔迹特征的方案。而后研究了笔迹鉴别的另外两项重要内容，即特征编码和分类器设计。通过对比两种常用的特征编码方式，确定较优的方案。描述了 K 临近分类方法原理，确定了适合本文笔迹特征的距离度量算法。

第四章是实验部分。首先对笔迹鉴别的评价准则进行了介绍，接着分别在中文数据集 CASIA-HWDB^[19]和外文数据集 ICDAR 2013^[23]上验证了本文算法的有效性，并和其他算法进行了对比。实验表明本文算法在中文数据集和外文数据集中均有一

定的优势，有很好的区分性和适应性。最后给出了本文的实验平台，以及各项时间统计。

第五章对本文的研究工作进行了总结，分析了本文提出的笔迹鉴别算法有待改进的地方，并展望了后续的研究工作。

2 相关理论基础

2.1 神经网络概述

人工神经网络通过模仿人的大脑的生理结构，进而模拟人脑处理信息的方式。“如果某种机制能够在自然生物中能够发挥作用，那么其在计算机系统中必然也能够起作用”，这是人工神经网络的思想源泉。人工神经网络是生物科学，数学和计算机科学相结合的产物，本质上是信息处理的数学模型。

2.1.1 神经网络的发展

早在 1943 年，美国芝加哥大学的数学家 Pitts 和心理学家 W.S.McCulloch^[25]率先提出了人工神经网络，并建立了其数学模型，即 MP 模型。在 1949 年 Donald Heb 指出神经通路会随着每一次使用得到加强，如果两个神经元在同一时间被触发，那么这两个神经元之间的连接就会得到加强。这一理论是解释人类具有学习能力的基础，给计算机科学家提供了灵感。人工神经网络第一次走向实际应用是在 1958 年，Rosenblat^[26]在前人的研究基础上进一步提出了感知器模型，从而在字符识别领域取得了一定的研究成果。当时，世界上许多实验室都在效仿感知机模型，有许多应用于语音识别，文字识别等方面的研究。然而，这一波研究神经网络的热潮并没有持续多久，这一时期基于冯·诺依曼体系的传统计算机研究正处于全盛时期，使得感知机的研究得不到重视。另外，1969 年 Marvin Minsky 和 Seymour Papert^[27]指出了神经网络研究存在的两个关键性问题。一是神经网络不能解决“异或”这样的逻辑问题，二是当时的计算机没有足够的计算能力处理大规模的神经网络，难以从单层的神经网络扩展到多层神经网络。六十年代末，人工神经网络的研究陷入了低潮。

直到 1982 年，加州理工学院的 John Hopfield^[28]提出了一种双向互连的神经网络，这一突破性的研究重新点燃了神经网络领域的研究热情。在 1986 年，多层神经网络成为了研究热点。Rumelhart 和 Hinton^[29]等人提出了反向传递算法（Backpropagation Algorithm, BP），这种算法把模式识别的误差分散到了整个网络，为多层神经网络的训练提供了可行性。如今，人工神经网络在硬件技术快速发展基础上，已经有了众多前沿的应用。如最近网络热门的“人机大战”，谷歌的机器选手“AlphaGo”在围棋这项比赛中战胜了人类，“AlphaGo”就是基于神经网络开发的。

2.1.2 反向传播算法

反向传播算法（Backpropagation Algorithm, BP）通常被用于多层前馈网络的训练，因此多层前馈网络也被称作 BP 网络。也就是说，神经网络是由神经元按层组成的，传递信号的时候是从前往后（Forward）一层一层传递的，计算误差的算法却是从后往前（Backwards）传递的。BP 算法解决了多层神经网络难以训练这一难题，是目前大多数神经网络模型中必不可少的算法。BP 网络从输入层（Input Layer）得到信号的出入，从输出层（Output Layer）获得结果，中间可能会有一个或多个隐含层（Hidden Layer）。

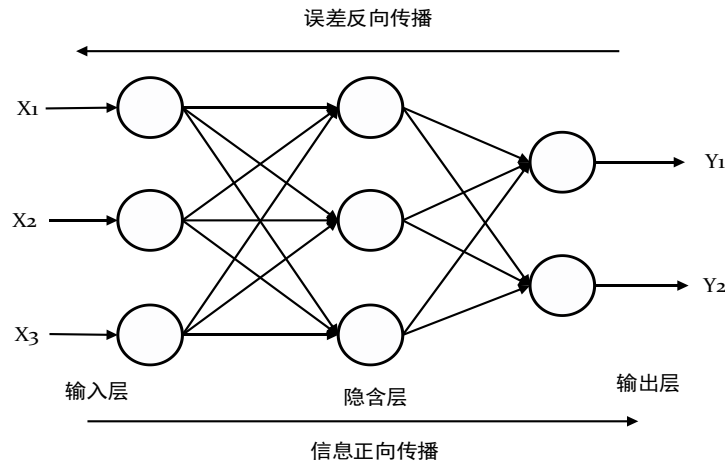


图 2-1 BP 网络模型模型

反向传递算法使用的是有监督学习（Supervised Learning）方法。对于给定的输入我们会计算网络的输出和正确值之间的误差，BP 算法就是通过更新整个网络的权重值，从而不断减小误差。BP 算法的理论基础是梯度下降法，如公式（2-1）所示。

$$x_{k+1} = x_k - \eta_k g_k \quad (2-1)$$

在以上公式中， x_k 表示当前权重值， g_k 表示当前函数的梯度， η_k 表示学习率（Learning Rate）。下面用一个三层网络距离来描述 BP 网络。我们用 x_i 来表示网络的输入，隐含层用 y_j 来表示，输出层的节点用 z_l 表示。输入层与隐含层的权重标记为 w_{ji} ，隐含层与输出的权重表示为 v_{lj} 。网络期望得到的正确的结果用 t_l 表示， $f(\cdot)$ 用来表示激活函数。

（1）网络的前向传播过程：

隐含层的输出如公式 (2-2) 所示:

$$y_j = f(\sum_i w_{ji}x_i - \theta_j) = f(net_j) \quad , \quad net_j = \sum_i w_{ji}x_i - \theta_j \quad (2-2)$$

整个网络的输出为:

$$z_l = f(\sum_j v_{lj}y_j - \theta_l) = f(net_l) \quad , \quad net_l = \sum_j v_{lj}y_j - \theta_l \quad (2-3)$$

网络误差为:

$$\begin{aligned} E &= \frac{1}{2} \sum_l (t_l - z_l)^2 = \frac{1}{2} \sum_l (t_l - f(\sum_j v_{lj}y_j - \theta_l))^2 \\ &= \frac{1}{2} \sum_l (t_l - f(\sum_j v_{lj}f(\sum_i w_{ji}x_i - \theta_j) - \theta_l))^2 \end{aligned} \quad (2-4)$$

(2) 反向传播过程:

采用梯度下降法来更新每一层的权重参数。输出层的平局误差函数的梯度为:

$$\frac{\partial E}{\partial v_{lj}} = \sum_{k=1}^n \frac{\partial E}{\partial z_k} \cdot \frac{\partial z_k}{\partial v_{lj}} = \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial v_{lj}} \quad (2-5)$$

E 这个函数包含有若干个 z_k , 但是所有的 z_k 是相互独立的, 只有 z_l 和 v_{lj} 相关。

因此, 上式中:

$$\frac{\partial E}{\partial z_l} = \frac{1}{2} \sum_k [-2(t_k - z_k) \cdot \frac{\partial z_k}{\partial z_l}] = -(t_l - z_l) \quad (2-6)$$

$$\frac{\partial z_l}{\partial v_{lj}} = \frac{\partial z_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial v_{lj}} = f'(net_l) \cdot y_j \quad (2-7)$$

权重在每一次训练时的修改值 Δv_{lj} , Δw_{ji} 和误差函数成比例关系, 并随着梯度的减小而减小, 输出层和隐含层之间的权值的更新函数为:

$$\Delta v_{lj} = -\eta \frac{\partial E}{\partial v_{lj}} = \eta \delta_l y_j \quad , \quad \delta_l = (t_l - z_l) \cdot f'(net_l) \quad (2-8)$$

在上式中, η 表示学习率。输入层和隐含层之间的权值的更新函数为:

$$\Delta w_{ji} = -\eta' \frac{\partial E}{\partial w_{ji}} = \eta' \delta'_j x_i \quad , \quad \delta'_j = f'(net_j) \cdot \sum_l \delta_l v_{lj} \quad (2-9)$$

2.2 深度学习简介

有研究表明,如果想通过学习一个复杂的函数来对语言、图片等内容高度抽象表示,那么就需要使用深度架构。人的大脑就是一个深度架构,并且能够高效地处理信息,从神经网络模仿人脑的意义上说,研究深度架构十分有必要。深度架构通常包含多层非线性操作,例如包含多个隐含层的网络结构或是多层嵌套的复杂公式。

深度学习的架构有若干种,深度卷积神经网络(Deep Convolutional Neural Networks, DCNN)^[30]和深度信念网络(Deep Belief Networks, DBN)^[31]是其中两个具有里程碑意义的网络架构,DCNN 更是现在神经网络领域应用最广泛的神经网络框架。在 2006 年以前,研究者们试了许多不同的方法来训练深层网络,除了 DCNN,其他都没有解决训练效率低下和准确率不高的问题。此后, Hinton 提出了深度信念网络,由多个受限玻尔兹曼机(Restricted Boltzmann Machines, RBM)组成,这一网络的提出催生了许多网络训练的新方法。近年来,衍生出了一些基于 DCNN 和 DBN 的变种网络。在文献[32]中,提出了一种 3-D DCNN 模型,其中的第三维是指时间信息。文献 [33]把 DCNN 和 DBN 相结合,提出了卷积深信网络(Convolutional Deep Belief Networks, CDBN),这个网络在原始的 DBN 网络的两个 RBM 层之间添加了池化层(Pooling Layer),它的作用是能够更有效得处理大维度的图片数据。下文将着重介绍本文所使用的深度卷积神经网络。

2.3 深度卷积神经网络

2.3.1 深度卷积神经网络的基本特征

深度卷积神经网络(DCNN)是前馈网络的一种,已经成为了模式识别领域的研究热点。DCNN 能够保证数据的旋转、平移不变性,这是由于其采用了三种优化网络结构的方法:

(1) 局部感知

DCNN 加强了相邻网络层神经元之间的局部连接,从而提取出图像的空间局部信息。网络的隐含层第 m 层只和第 $m-1$ 层的局部子集相连接,而且第 $m-1$ 层被连接的节点在空间上是相邻的。

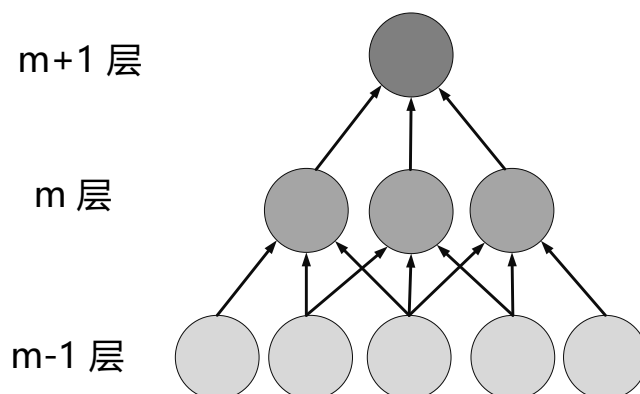


图 2-2 CNN 的局部感知

如图所示，假定 $m-1$ 层是整个网络的输入层， m 层的每个神经元与 $m-1$ 层的三个相邻的神经元节点相连。 $m+1$ 层和 m 层有着相似的关系，也是上一层与下一层的三个节点相连，它们的接受域大小是 3。但是输入层 $m-1$ 有 5 个节点，它的接受域大小却是 5。每个节点对于其接收域以外的下层节点的变化是没有响应的，这种网络结构很好得提取了输入图片的空间局部信息。多个局部感知的网络层叠加在一起就形成了一种非线性的滤波器，而且随着层数的变多也会变得“全局化”。如图 2-2 中， $m+1$ 层的每个节点间接和 $m-1$ 层的 5 个节点相连，从而提取这 5 个输入的特征。

(2) 权值共享

在 DCNN 中，整个感知域中的滤波器都是一样的，它们有相同那个的权重和偏置，从而形成特征图（Feature Map）。

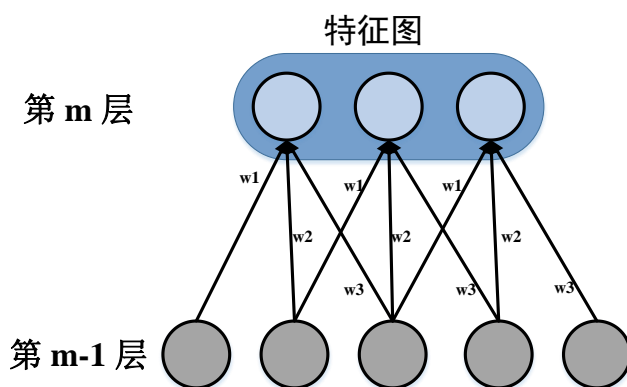


图 2-3 DCNN 的权值共享

在图 2-3 中，隐含层 m 的 3 个节点属于同一个特征图，每个节点分别都包含三个相等的权值： w_1, w_2, w_3 。通过共享权值，能够更好得提取感知域的特征。极大地减少了需要训练的网络参数，增加了网络训练的效率并且能够提升网络模型的泛化

能力。

(3) 下采样

在 DCNN 中一般采取的是最大下采样 (Max-pooling), 属于非线性的采样方式。首先把原始特征图分成若干个大小相等的矩形, 对于每一个子矩形区域取最大值作为输出。下采样能够缩小特征图的维度, 减少计算量同时保留有用的信息, 进一步抽象图像, 并形成新的特征图从而进行下一层卷积。

DCNN 包含一系列的卷积和下采样操作, 如图 2-4 所示: 卷积层 C_x 是输入图像经过滤波器 f_x 处理再加上偏置 b_x 得到的 (f_x 和 b_x 都是可训练的)。下采样操作对于每四个像素进行最大下采样并与权重 w_{x+1} 相乘, 再加上偏置 b_{x+1} , 通过 sigmoid 激活函数处理。从而得到更小的特征图 S_{x+1} 。

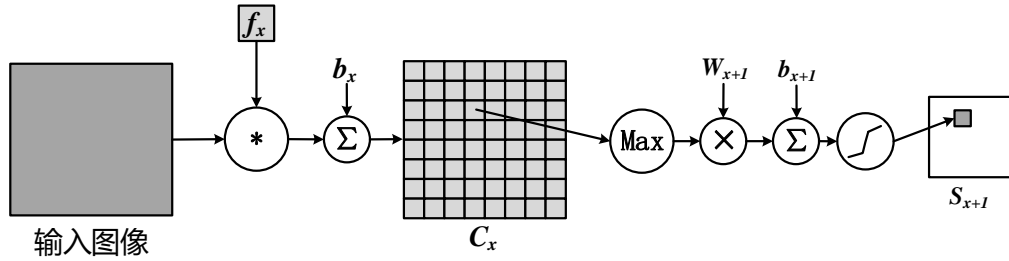


图 2-4 卷积处理和下采样

2.3.2 防止过拟合的策略

深度神经网络通常含有大量的参数, 这时防止过拟合就显得十分重要。我们主要采取以下两种方式来防止深度卷积神经网络过拟合。

(1) 数据扩展

数据扩展 (Data Augmentation) 是人为地对数据进行各种变化 (保留原始数据的标签), 以此来增大数据集, 这是防止过拟合最简单有效的方法。

一种常用的数据扩展方式是对于图像做随机切割和水平映射 (Horizontal Reflections)。假设原始的图片大小为 $w \times h$, 我们随机地从原图里面取大小为 $\alpha w \times \alpha h$ 的子图 (通常 α 取值为 0.9 左右), 从而可以用提取出来的子图和它们的水平映射来训练深度卷积神经网络。这样可以把数据集扩大若干倍 (具体倍数可以由随机的次数决定), 对于减少网络的过拟合起到了很好的效果。在测试阶段, 我们可以对于测试图像提取五个大小为 $\alpha w \times \alpha h$ 的子图 (分别对应原图的四个角和中心区域), 同

样对这五个子图分别做水平映射，最后把这十个子图的预测平均值作为最终的结果。

（2）Dropout

有研究表明，结合多个网络模型来做预测可以很大程度上提高预测的准确率，但是这样的做法往往导致训练时间太长以及算法过于复杂。Dropout 策略能够达到同样的效果，而且其训练成本也可以维持在可控的范围内。所谓 Dropout 是指在每一次训练的过程中随机地删除网络中一半节点，我们把被删除节点的输出值设为 0，让这些节点在网络的前向传递和反向传递中都不起作用。使得对于每一次输入，神经网络都呈现出不同的网络结构。在测试阶段，我们让每一个网络节点的输出都乘以 0.5，这等同于由 Dropout 策略所产生的多个网络的预测平均值。Dropout 策略减小了网络节点相互适应的复杂度，增强了网络的泛化能力，能有效地防止过拟合。

2.4 笔迹鉴别系统

笔迹鉴别系统通常可以拆分为四个步骤：笔迹材料图片获取，图像预处理，特征提取和分类（相似度度量）^[38]。图 2-5 给出了笔迹鉴别的流程图。

（1）笔迹材料图片获取

通常是指把纸质文档等实物材料转化为笔迹图片过程，扫描器是这个过程中最常用的设备，因为相比与其他的设备扫描器在处理图像的时候带入的噪声比较少。笔迹材料特征提取的效果往往会受到图片获取的质量的影响，低分辨率的图像会使后续图像二值化的效果变差。

（2）图像预处理

未经处理的笔迹图像可能会含有很多无关的信息，如空白，下划线等。去除这些无关的信息能够更好得提取特征。笔迹图像预处理包括：图像去噪，图像二值化，文字行定位等。

（3）特征提取

特征提取是笔迹鉴别系统的关键性步骤，也是本文笔迹鉴别研究的重点。特征提取关键问题是怎样提取有区分性的特征，从而取得高准确率的分类结果。

（4）分类

笔迹鉴别是要评估查询的材料和库里面的保存的材料的相似性，最常用的分类方法是 K 近邻分类，一般会通过计算样本特征的之间的距离来进行相似度度量。

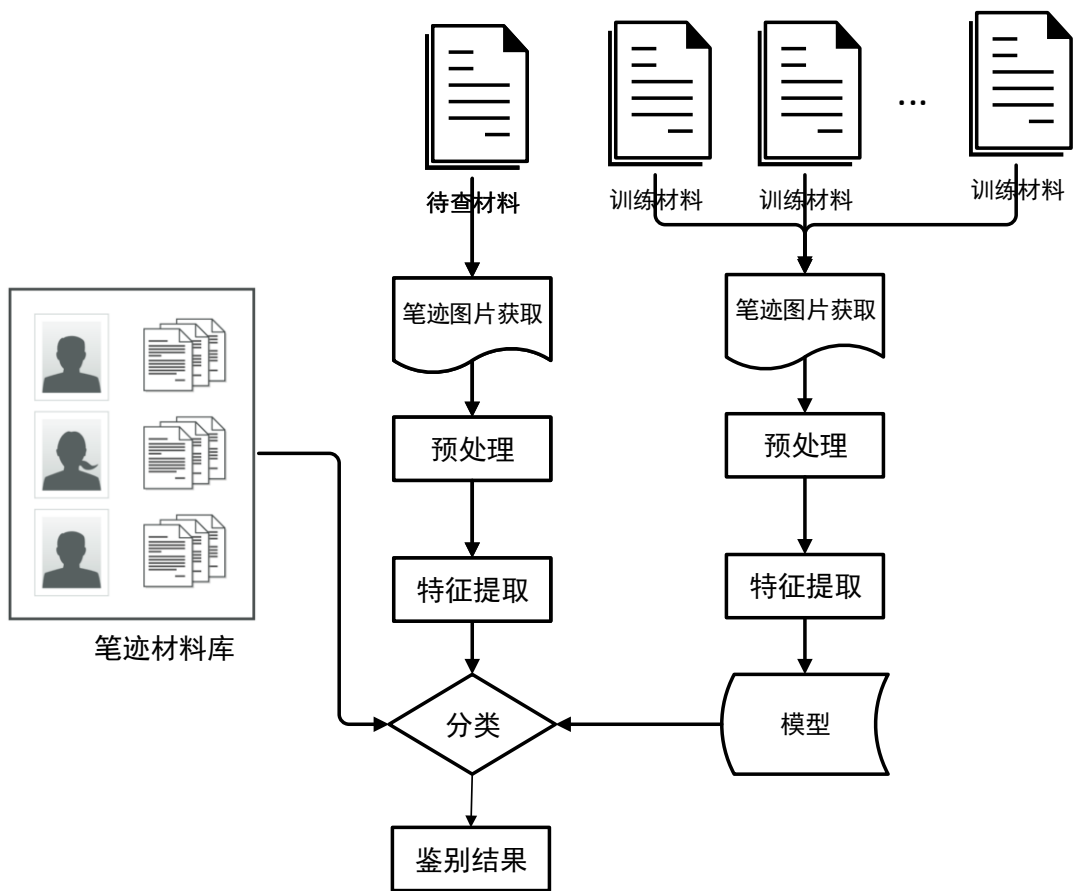


图 2-5 笔迹鉴别系统流程

2.5 本章小结

人工神经网络是本文研究的理论基础，本章首先简要介绍了人工神经网络的发展，概述了人工神经网络的关键性内容反向传播算法。接着介绍了本文所使用的深度卷积神经网络，最后给出了常用笔迹鉴别系统的框架。

3 基于深度卷积神经网络的离线笔迹鉴别研究

深度学习近年来在模式识别和计算机视觉领域许多困难问题上被证明是有效的方法。例如：人脸识别、图像分类、图像识别和手写文本识别。在中文手写文本识别领域，Graham^[35]对深度卷积神经网络进行了改进，提出了 DeepCNet，从而取得了突破性的成果。尽管深度学习在文字识别领域有比较多的研究成果，但是在与文字识别很相近的文本无关离线笔迹鉴别领域，深度学习在这一方面的应用才刚刚开始。特别是离线中文笔迹鉴别，目前还没有看到这方面的研究成果。一个重要的原因是因为，在笔迹鉴别研究中，训练数据集和测试数据集是分离的（测试数据和训练数据不能包含同一类别），因此就不能直接使用 DCNN 来做分类。

与以往的研究不同，本文没有使用人工设计的特征，如 SIFT^[43], SURF^[14], LBP^[44]等，而是采用 DCNN 来提取图像的特征。DCNN 提取特征本质上是数据驱动，网络的每一个卷积层都是对上一层特征的抽象，经过多层卷积之后能够得到更具有区分性的特征。

3.1 本文笔迹鉴别系统流程

本文系统流程和传统的基于人工设计的特征的笔迹鉴别略有不同，后者在提取特征的阶段通常不需要训练，其训练往往存在于分类阶段（训练分类器）。由于本文方法是基于深度卷积神经网络来提取笔迹图像的局部特征，因此需要首先训练 DCNN 模型，再基于训练好 DCNN 模型来提取特征。本文的系统流程如图 3-1 所示，分为训练部分和测试部分。

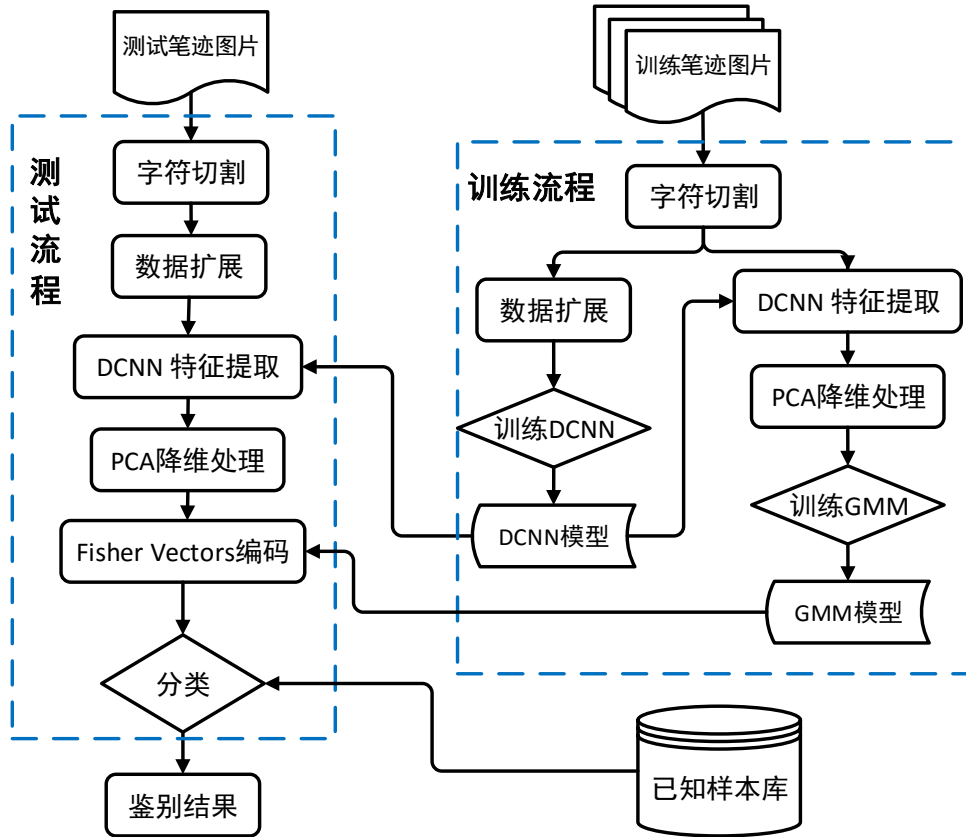


图 3-1 本文系统流程

(1) 训练流程:

首先需要训练 DCNN 网络模型，这是提取笔迹材料局部特征的必要准备。对于训练集里面的笔迹材料做字符切割，数据扩展等一系列处理，从而得到若干个笔迹材料的小块图像，用这些小块图像来训练 DCNN。

接着，利用上一步训练得到的 DCNN 模型对所有字符切割得到的小块图像提取特征，并且对所得的特征进行 PCA 处理。再利用处理后的特征来训练高斯混合模型（Gaussian Mixture Model, GMM），这是测试阶段做 Fisher Vectors 特征编码需要用到的模型（详见 3.5.1 小节）

(2) 测试流程:

同训练流程一样，首先对于输入的笔迹材料进行字符切割、数据扩展（可选）等处理，得到若干个笔迹材料的小块图像。然后利用训练得到 DCNN 模型对于切割出来的小块图像提取特征，并进行 PCA 去掉冗余信息。再基于 GMM 模型利用 Fisher Vectors 算法对以上特征进行编码，进而得到笔迹图像的全局特征。用此特征来和已

知样本库中的笔迹材料的特征进行比较（通过欧式距离计算），找出与之最接近的实例，从而得出鉴别结果。

3.2 深度卷积神经网络结构

本文使用 DCNN 来提取笔迹图像单个字的特征，单个字图像都是有标签的（标记这个字的书写者），因此，DCNN 网络的最后一层是 SoftMax，网络节点数目和训练数据的类别数目相同。训练完成后，选用网络最后几层的输出来提取图像的特征。一般来说，网络结构越大，网络的层数越多，需要学习的参数越多，过拟合的可能性越大，训练的速度也越慢。但如果网络的层数太少，训练的效果可能就没有那么好。综合考虑，本文使用具有 3 层卷积的网络结构（下文称 Net-WI），另外再采用经典的 AlexNet^[52]做对照实验。

3.2.1 Net-WI 网络结构

本文使用含有 3 个卷积层的 DCNN，具体的网络结构如图 3-2 所示。除卷积层外，还有 2 个全连接层，每个卷积层之后都接有一个池化层（下采样层）。

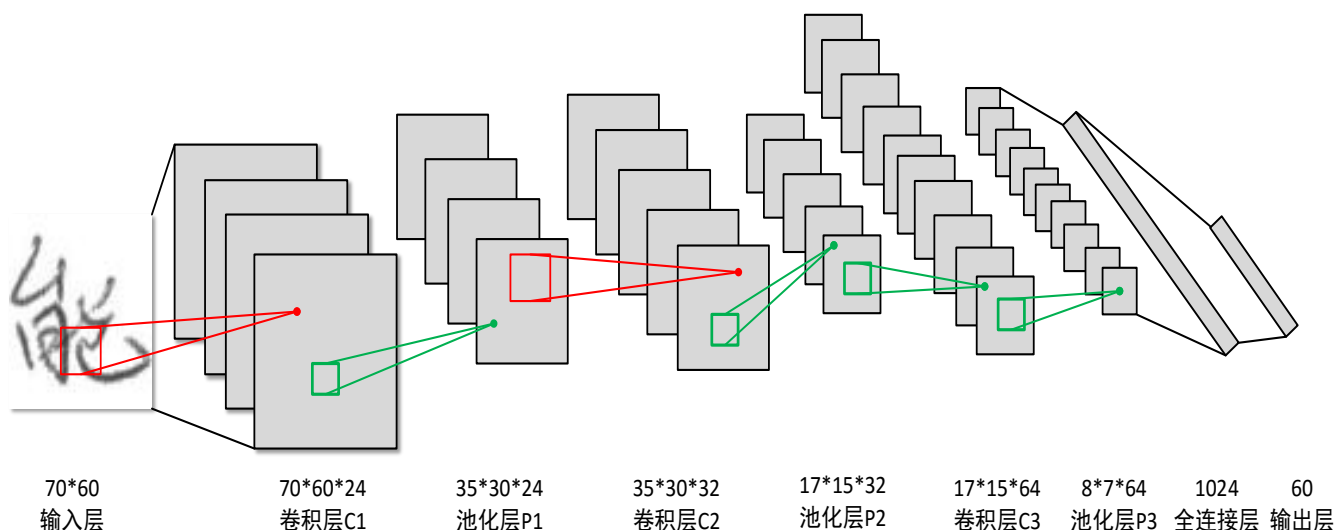


图 3-2 深度卷积神经网络结构图

本文在两个数据集上进行了实验，而这两个数据集的图片大小并不一样，下面以中文数据集 CASIA-HWDB^[19]为例来对 DCNN 的网络结构进行分析。如图 3-2 所示，输入图片是大小为 70*60 的单个字的图片。DCNN 的第一层是卷积层，与输入

相连,记为 C1。输入的单个字图像通过和 24 个滤波器卷积,从而产生 24 个特征图,采用高斯滤波器。网络中共使用两种不同大小的卷积核,图中较大的矩形框大小为 5*5,较小的矩形框对应卷积核大小为 3*3,本层卷积核的大小为 5*5。卷积步长为 1,为了使卷积函数能够作用图片边缘的像素,我们在图片的上下左右各填补 2 个像素宽度 (Pad)。因而 C1 层的特征图大小为:

$$\left(\frac{70+4-5}{1} + 1\right) * \left(\frac{60+4-5}{1} + 1\right) = 70 * 60 \quad (3-1)$$

和输入图片大小一样。由于 C1 总共有 24 个卷积核,而每个卷积核需要 5*5,即 25 个参数,另外还有一个偏置参数,所以 C1 的参数总数为 26*24,即 624。

DCNN 的第二层是第一个池化 (Pooling) 层,即图 3-2 的 P1 层。池化层是用来减少特征图的维度,进而极大地减少了计算量。池化层采取最大化采样 (Max-pooling),即对于子采样区域,取其中的最大值作为输出。采样区域大小为 3*3,采样间隔设为 2,这样特征图的大小就会减小 4 倍。P1 的特征图分辨率为 35*30。

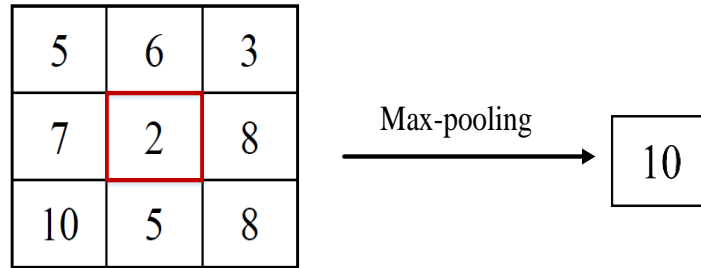


图 3-3 最大化采样示意图

第三层卷积层 C2,本层卷积核的大小和上层一样,依然是 5*5。这一层的卷积核个数设置为 32,卷积步长为 1,依然向图片的四周 pad 两个像素宽度。C2 的输入是 P1,采用和 C1 类似的计算方法得出本层特征图大小为 35*30,本层需要的参数个数为 $(5 * 5 + 1) * 32$,即 832 个参数。

第四层是池化层 P2,与卷积层 C2 相对应,特征图的数目为 32。采样区域大小和 P1 一致,采样间隔为 2。同样是采取最大化采样,最后得到的特征图大小为:

$$\left(\frac{35-3}{2} + 1\right) * \left(\frac{30-3}{2} + 1\right) = 17 * 15 \quad (3-2)$$

第五层是卷积层 C3,这一层卷积核的个数设置为 64,卷积核的大小设置为 3*3,向图片的四周 pad 一个像素宽度。卷积步长和前两层一样,还是设置为 1。需要的参数个数为 $(3 * 3 + 1) * 64$,即 640 个参数。本层的特征图大小和 P2 一样为: 17*15。

具体计算如下：

$$\left(\frac{17+2-3}{1} + 1\right) * \left(\frac{15+2-30}{1} + 1\right) = 17 * 15 \quad (3-3)$$

第六层是池化层 P3，参数设置和前面两个池化层一样，卷积核个数与 C3 保持一致，即 54 个卷积核。本层的特征图大小为 8*7。

第七层是全连接层，该层的节点数目为 1024。该层的每个节点都和 P3 层的每个节点两两相连，因此总的连接数为：

$$(8 * 7 * 64) * 1024 = 3670016 \quad (3-4)$$

对于全连接层，需要计算的权重个数和连接数目一样，因此权重个数也为 36700016。全连接层是给网络加入更多的非线性因素，从而更好得描述为了中文笔迹的复杂性。最后一层是输出层，采用 SoftMax 分类器。不同于逻辑回归只能处理非线性的二分类问题，SoftMax 可以处理多分类问题，并且一个样本只能对应一个类别，类别之间是互斥的，笔迹鉴别的数据样本符合这样的要求，即每份笔迹材料的标签（Label）是某一个书写人。节点数目与训练数据类别数目一样，对于 CASIA-HWDB 中文数据集为 60 个节点。这一层的连接数和权重参数个数为：

$$1024 * 60 = 61440 \quad (3-5)$$

每一个卷积层和全连接层的激活函数都是 RELU(Rectified linear units)^[36]。RELU 函数是一个非线性的激活函数，被证明比传统 Sigmoid、Softpuls、Tanh 等更有效^[37]，而且 RELU 函数相比于其他函数来说计算量小很多。使用 RELU 函数会使得很多网络几点输出 0，因而对于过拟合有一定的抑制作用。RELU 的公式如下：

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (3-6)$$

3.2.2 AlexNet 网络结构

AlexNet 获得了 2012 年 ImageNet 图像分类比赛的冠军，是一个经典的网络结构。它有 5 层卷积层，3 个全连接层（包括最后的输出层）。

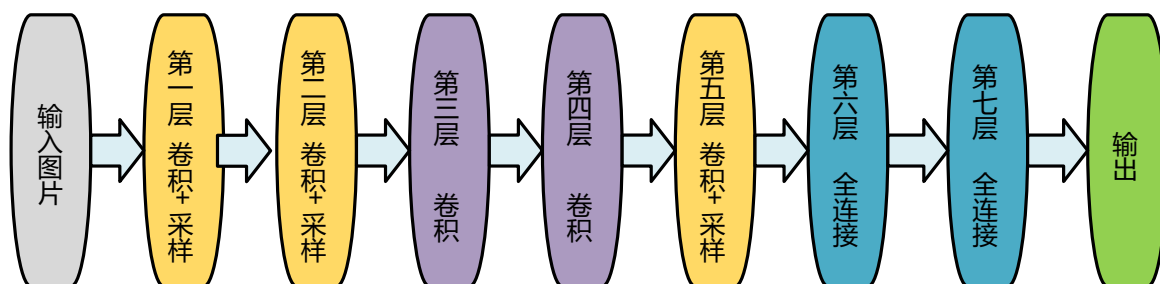


图 3-4 AlexNet 网络结构图

如图 3-4 所示，第一个卷积层有 96 个卷积核，卷积核的大小是 11×11 ，卷积步长是 4。第二个卷积层有 256 个卷积核，卷积核大小为 5，卷积步长为 1。接下来的三个卷积层的卷积核大小均为 3，卷积步长均为 1。第三个和第四个卷积层的卷积核个数为 384，第五个卷积层卷积核个数为 256。最后连接两个全连接层节点数目 4096，输出层节点数为 1000。

由于本文笔迹鉴别图像比较小，只有 70×60 ，如果按照原始 AlexNet 的网络结构直接训练，那么特征图很快就会缩小为 1×1 ，后面的操作卷积也就无法进行了。因此我们对 AlexNet 做出一下改进：

(1) 把第一层卷积层的卷积核大小改成 7×7 ，因为原始 11×11 的卷积核对于切割出来的笔迹材料子图的来说太大。再把这一层的卷积步长改为 1。

(2) 把倒数第二个全连接层的节点数改成 1024，因为如果节点数是 4096，经过 Fisher Vectors 编码（详见 3.5 节）以后特征维度太大，会显得很稀疏，不利于做检索。

3.3 深度卷积神经网络训练

本节介绍和网络训练相关的一些数据预处理，数据归一化，样本集扩充等相关的研究，以及深度卷积神经网络在训练时的参数设置。

3.3.1 滑动窗字符切割

本文通过实验发现，提取单个“字”图像或者“字母”来训练深度卷积神经网络，比用整张笔迹页面图像，或者行图像来训练更有效果。因此对于笔迹材料需要首先做字符图像提取，而相较于字符识别所需要的字符切割而言，笔迹鉴别对字符切割并没有很高的精准度要求。

针对以上情况，本文提出了一种基于滑动窗的字符切割方法，实验证明这对于笔迹鉴别来说是一种行之有效的字符切割方法，相比于传统的方法计算量小很多，并且能在一定程度上增加了训练样本数量。

我们首先对比笔迹材料提取“词”，再在“词”的基础上进行下一步的切割。笔迹材料的背景一般，我们采用提取连通区域（Connected Components）的方法来做“词”的切割。下面拿数据集 ICDAR2013^[24]的笔迹材料来举例，切割的效果如图 3-5 所示。

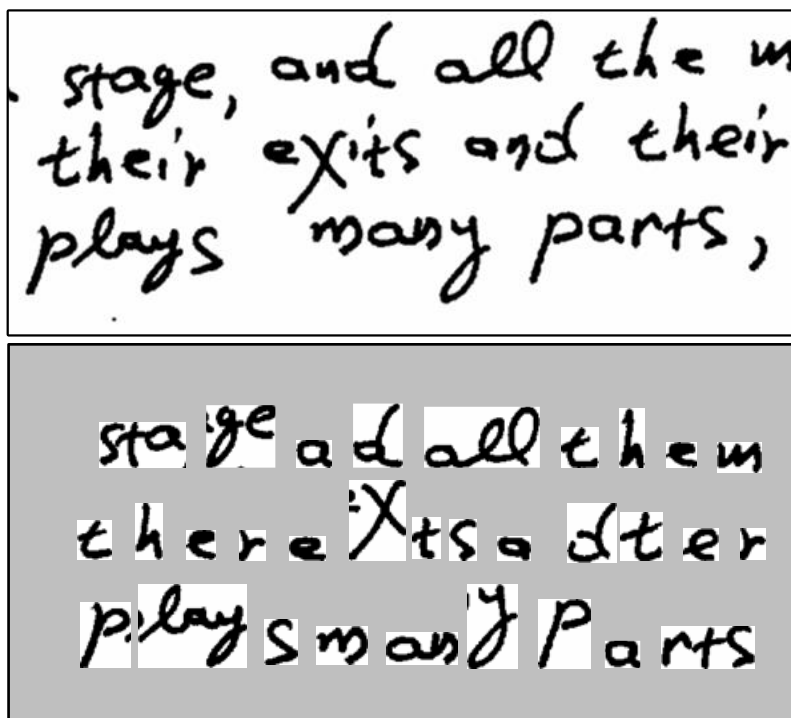


图 3-5 “词”切割效果图，上图是原始笔迹材料，下图是切割后的笔迹材料

本文在切割“词”的过程中滤掉了是长度或者宽度非常小的图片，这样的图片对于训练 DCNN 是没有帮助的。如图 3-5 中上图的逗号，字母“i”等。由于训练 DCNN 的图片大小必须统一的，所以我们获取的笔迹材料图片的长宽比最好集中在一定的范围。通过统计数据集中所有的图片宽高比的平均值，我们通过进一步的滑动窗切割来切割那些宽度比较大的连通区域，使所有的“字”图片的宽高比尽量靠近这一均值。图 3-6 是在上一步的基础上，对连通区域进一步切割的结果。

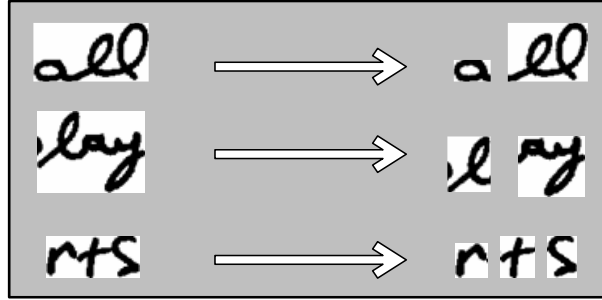


图 3-6 连通区域的进一步切割

本文的切割算法是对于二值图像来进行的，如果要处理的图像不是二值图像，则首先要对图像进行二值化处理，然后再对二值图像进行切割，再把切割后的结果映射到原图上面去。对于二值图像而言，黑色笔迹部分的像素值为 0，空白部分的像素值为 1。我们首先对待切割图像的每一个像素定义一个能量值：

$$E(x, y) = \begin{cases} 1, & \text{if } I(x, y) = 0 \\ 0, & \text{if } I(x, y) = 1 \end{cases} \quad (3-7)$$

上式中 $I(x, y)$ 是表示图片像素值，我们把黑色部分的能量值设为 1，白色部分能量值设为 0。因此图像每一列的能量值为：

$$E(x) = \sum_{i=1}^H E(x, i) \quad (3-8)$$

这里的 H 表示图像的高度，接下来我们使用滑动窗的方法来对能量值最低的列进行切割。这里用 A_h, A_w 分别表示数据集中所有切割出来的连通区域的高和宽的平均值。我们对那些宽高比大于平均宽高比 2.1 倍或者宽度大于平均宽度 2.3 倍的连通区域才进行切割，即：

$$\frac{w}{h} > 2.1 * \frac{A_w}{A_h}, \text{ or } w > 2.3 * A_w \quad (3-9)$$

我们设滑动窗的宽度为 A_w 的两倍，用 W 表示滑动窗的宽，即 $W = 2 * A_w$ 。首先我们把滑动窗至于待切割图像的最左边，对于滑动窗最前面的 $1/4$ 的区域我们是不做切割的（这样做是为了防止切割出过小的子区域），在剩余的滑动区域内选取列能量值最小的列进行切割。在一次切割完成后滑动窗移动到此次的切割点，从而对剩余的区域进行下一轮切割，直到剩余的区域不再需要切割，即不满足 3-9 式。

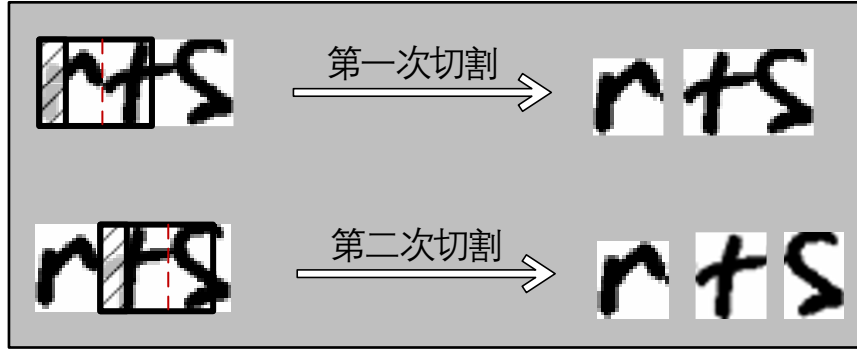


图 3-7 基于滑动窗的切割

滑动窗切割过程如图 3-7 所示，图中滑动窗左边的阴影部分就是保留不做切割的区域，虚线是切割点。第二次切割的时候滑动窗移动到了第一次的切割点。

3.3.2 数据的归一化和扩充

(1) 数据归一化

由于训练 DCNN 的图片必须是大小一样的。因此我们首先要把数据集中的样本归一化的统一的尺度，对于 CASIA-HWDB 数据集，我们的输入图片宽高是 60*70。比较简单的做法是直接把数据缩放到想要的尺度，具体计算公式如下：

$$\alpha = \frac{60}{W}, \quad \beta = \frac{70}{H} \quad (3-10)$$

$$W' = \alpha W, \quad H' = \beta H \quad (3-11)$$

如果采用上述缩放方法，会导致字迹图像变形。而对于笔迹鉴别而言字迹的宽高比本身就是一项具有区分性的特征，这样的做法会使后期提取特征的效果变差。本文采取缩放加填白的方法，使得图像归一化的过程中不会发生形变。具体公式如下：

$$\alpha = \min\left(\frac{60}{W}, \frac{70}{H}\right) \quad (3-12)$$

$$W' = \alpha W, \quad H' = \alpha H \quad (3-13)$$

按以上公式处理图片可以保证较长的一边达到缩放要求，另外一边会小于要求缩放的长度，我们就在这一边的两端填入宽度相等的白色区域，从而使缩放后图片的宽高分别为 70 和 60。

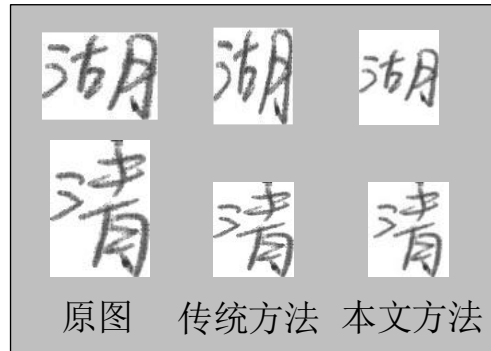


图 3-8 样本归一化

(2) 样本集扩充

DCNN 的训练一般是基于大规模的数据集，一般来说训练集越充分训练出来的网络越好。而当数据样本比较少而网络的参数较多时，就容易出现过拟合。对于 CASIA-HWDB 数据集，我们的训练数据只有 60 个类别，每个类别才 5 份笔迹材料。总共是 300 份笔迹材料，切割完以后也才 75000 个字。这时对于数据的扩充就显得十分必要。另外，由于需要对 DCNN 提取的特征做编码（详见 3.5 节），对于单一的笔迹材料提取的 DCNN 特征越多的编码时的重建误差就会越小。因此我们对于测试集也做扩充，从而增加了每页笔迹材料包含的 Patch，也就增加了 DCNN 特征的数量。

我们首先把 70*60 的笔迹图片缩放到 76*66，即宽高各增加 6 个像素。然后用滑动窗的形式在放大后的图像中采集 70*60 的子区域，滑动窗的步长为 3，这样总共可以得到 9 个子区域。为了引入一定的随机性，这 9 个子区域不全部选取，而是对于每个子区域必须满足以下条件才能被选中：

$$rand() < 0.3 \quad (3-14)$$

这里的 $rand()$ 是 0 到 1 的随机函数，因此每个子区域被选中的概率为 0.3。接着我们再把上下左右各 pad 3 个像素值，pad 的方式是复制图片边缘的像素，再把这 pad 以后的图片缩放到 70*60，然后保存。原图依然保留在扩充后的数据集里面，这样一来我们的数据集就会扩充 4.7 倍左右 ($9*0.3+2$)。本文并没有对图片进行旋转扩充，因为每个人写字都有自己的倾斜习惯，对样本进行旋转有可能使之丢失固有的特征。图 3-9 展示了本文数据集扩充的效果：

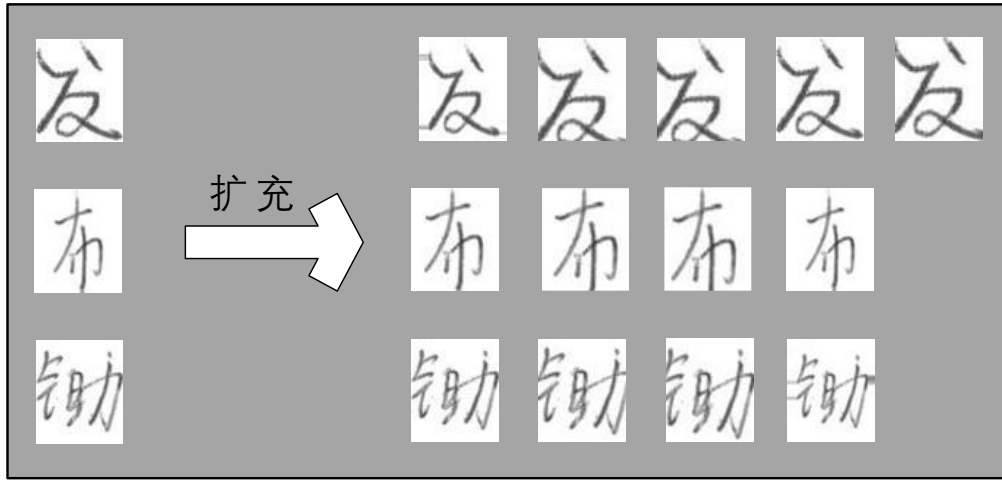


图 3-9 样本集的扩充

3.3.3 网络参数的设置

学习率 (Learning Rate) 对于训练 DCNN 来说是一个重要的参数, 这个参数关系到梯度下降法的快慢。如果设置得太小, 那么网络学习的速度就会很慢, 如果设置得过大, 当训练达到一定的次数之后可能导致学习的停滞, 无法收敛。本文采取的策略是隔段下降法, 网络总的训练次数设为 8 万次, 在 2 万次之前学习率设置为 0.001, 在 2 万到 4 万次阶段学习率设置为 0.0001, 在 4 万到 8 万次之间设置为 0.00001。这样设置能够比较合理地训练 DCNN, 并且能够加快学习的速度。

DCNN 训练时采用梯度下降法更新权重, 冲量 (Momentum) 设置为 0.9, 权重衰减 (Weight Decay) 设置为 0.001。我们把训练数据划分为训练集和验证集 (Validation Set), 拿 90% 的数据来训练, 拿 10% 的数据来测试。每训练两千次, 做一次验证测试 (Validation), 并且计算出准确率和网络损失 (Loss), 以此来观察网络的训练状态是否正常。如果损失能够持续不断地下降, 说明网络的学习率设置得比较合适, 如果损失总是来回震荡、忽高忽低说明学习率有可能偏大, 如果损失总是维持在很低的水平, 说明已经接近收敛。在训练的时候我们把每 64 张图片组成一个数据组 (Batch), 每次拿一个数据组来训练, 梯度下降也是根据这一组数据的平均损失来进行, 提高了梯度下降法的稳定性。

3.4 深度卷积神经网络特征的选择

本节主要研究深度卷积神经网络特征的选择。首先研究了不同网络层特征的选

取，然后采用主成份分析对 DCNN 特征进行进一步处理，去掉冗余信息。

3.4.1 不同网络层特征的选取

由于 DCNN 不同网络层对图片抽象程度不一样，进而提取的特征的区分能力也不一样。研究表明，越深层次的特征区分性和不变性越强^[51]，因此本文分别选取 DCNN 网络的最后四层输出来做实验，经过实验发现 Fc2 层提取的 DCNN 特征可以取得更好的结果，因而选用这一层特征作为最终的方案（具体实验见 4.3.4）。

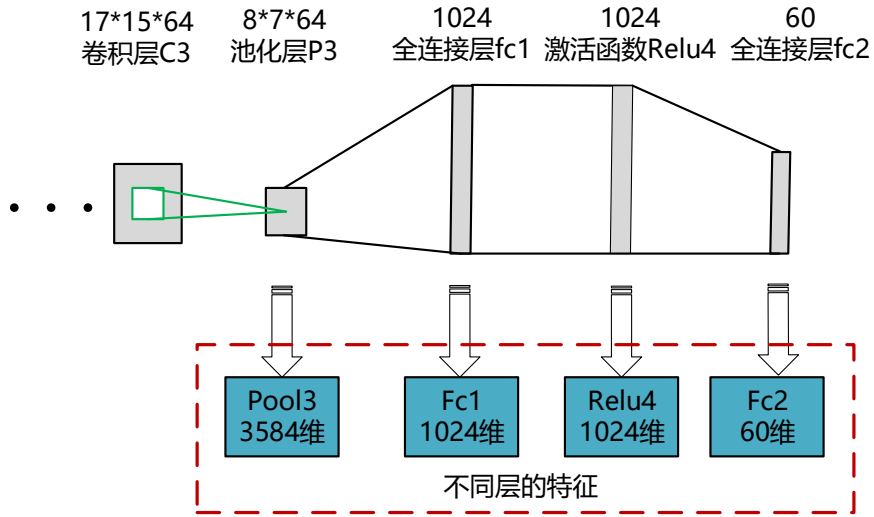


图 3-10 不同网络层特征的提取

3.4.2 特征的主成份分析

主成分分析（Principal Component Analysis，PCA）^[45]是数据降维的通用做法，并且已经被应用在了许多计算机视觉问题上面，如目标识别^[39]、特征选择^[40]和人脸识别^[41]。在文献[42]中，Yan Ke 指出把 PCA 和 SIFT 特征结合起来使用比单纯使用 SIFT 特征在图像检索方面效果更突出，并且在存储和速度方面更有优势。受此启发，本文 PCA 算法对于提取的 DCNN 特征进行了重组，过滤掉了噪声，增加了笔迹鉴别的准确率并且加快了算法的速度。下面将简述 PCA 算法的流程。

首先采用 DCNN 对于数据集中的所有训练样本提取特征，记为 A_1, A_2, \dots, A_N ，这里假设训练样本的总数为 N ， A_i 表示第 i 个样本的特征，特征的维度身为 k ，即 $A_i = (a_{i1}, a_{i1}, \dots, a_{ik})$ ， $i = 1, 2, \dots, N$ 。则有：

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i \quad (3-15)$$

从而得到所有训练数据的协方差矩阵，记为 Cov：

$$\text{Cov} = \frac{1}{N} \sum_{i=1}^N (A_i - \bar{A})(A_i - \bar{A})^T \quad (3-16)$$

我们需要进一步求出 Cov 的特征值和特征向量，这是矩阵论里面的基础知识，这里就不再赘述。把特征值从大到小排序，记排序后的特征值为： $\lambda_1, \lambda_2, \dots, \lambda_k$ ，对应的特征向量为： $\Psi_1, \Psi_2, \dots, \Psi_k$ 。这里的 k 和原始的特征维度一样大，选取前 t 个特征向量，记为：

$$W = (\Psi_1, \Psi_2, \dots, \Psi_t), \quad t \leq k \quad (3-17)$$

对于一个 N 维度的向量 X_i ，则对 X_i 应用 PCA 降维后的数据可以表示为：

$$Y_i = W^T (X_i - \bar{A}) \quad (3-18)$$

上式中 Y_i 的维度为 t，通过调整 t 的大小就可以达到降维目的。选择 t 值的大小有以下常见的几种方法：

(1) 保留所有的非 0 特征值相对应的特征向量，这时 $t = k$ 。这种方法在原始的特征维度比较大的时候计算量会比较大，通常降维效果也不明显，在很多时候甚至起不到降维的作用，去除冗余数据的效果也不好。

(2) 从后面去除一定比例的特征向量，即 $t = \alpha k$ 。因为特征值从大到小排序的，这种方法实际上是去除了信息量最少的那一部分特征向量。优点是降维程度可控，但是信息量丢失程度不可控。

(3) 通过选取一定量的特征值，使得信息量达到一定的比例，从而得出 t 值的大小。即，使得 t 满足：

$$\frac{\sum_{i=0}^t \lambda_i}{\sum_{i=0}^k \lambda_i} > \beta \quad (3-19)$$

本文使用的是这种方法，我们把 β 设为 0.995，这样原始特征在经过 PCA 降维后能够保留绝大部分的信息量，并且去除冗余数据的效果较好。对于冗余数据比较多的情况下也能取得很好的降维效果。

3.5 DCNN 特征的编码研究

CASIA-HWDB 数据集^[19]中的每份笔迹材料经过切割之后通常会有两三百个“字”的子图，这些小块子图我们称之为 Patch，那么在应用数据扩充之后，每份笔迹材料会有上千个 Patch，我们把 Patch 数目记为 N 。DCNN 对于每个 Patch 提取特征，就会得到 N 个特征向量。这 N 个特征向量描述的是笔迹材料的局部（Local）特征，我们需要把这 N 个特征组织起来形成全局（Global）的特征。具体的方式有以下两种：

（1）简单的做法是对所有的 N 个特征向量的每一维求平均值，把平均后的向量作为笔迹材料的全局特征。这种方法对于特征的信息损失太大，最终的结果也比较差。文献^[46]中就采用了这种方法聚合 DCNN 提取的特征，但是其结果并不好，甚至还不如往年的一些传统的方法。

（2）另外一种更加有效的方式是基于词袋模型（Bag of Words, BOW）对所有的特征进行编码（Coding），矢量量化（Vector Quantization, VQ）是编码的一种常用方式，但是这种编码方式采取的是硬编码（Hard Quantization），通过直方图统计来获取特征，每一个特征编码后只有一个非零值，因而丧失了很多高阶的统计特征。最近的研究中提取了几种新的编码方式，把硬编码替换成了选择性的编码方式，即软编码。这些编码方式可以保留更多图像原始特征的信息。本文采取近年来提出的新的编码方式来对 DCNN 特征进行编码，即 Fisher Vectors 编码（Improved Fisher Kernel）^[15]，并且采用局部约束线性编码（locality-constrained linear encoding, LLC）^[47]进行对比实验。下文将对 Fisher Vectors 和 LLC 算法做简要介绍，并给出了适合本文笔迹鉴别算法的参数设置。

3.5.1 Fisher Vectors 特征编码

Fisher Vectors 是 Perronnin 等人在 2010 年提出来的，在目标识别和图像检方面有许多应用^[15]。Fiel 等人采用 SIFT 特征建立基于 Fisher Vectors 的词袋模型，在笔迹鉴别的研究中取得了比较大的进展^[48]。Jain 和 Doermann 基于之前的研究提出了一种局部特征融合的方法，并利用 Fisher Vectors 对所获得的特征进行编码，也取得了不错的研究成果^[12]。受此启发，本文应用 Fisher Vectors 来对 DCNN 提取的局部特征进行编码，从而形成维度一样的具有全局特性的特征。

第一步我们需要对训练集中所有的 Patch 提取 DCNN 特征，并且对其进行 PCA 降维。然后采用上述特征训练一个高斯混合模型（Gaussian Mixture Model, GMM），如等式所示：

$$u_\lambda(x) = \sum_{k=1}^K w_k u_k(x) \quad (3-20)$$

$$u_k(x) = \frac{1}{2\pi^{D/2} * |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-u_k)' * \Sigma_k^{-1} * (x-u_k)} \quad (3-21)$$

上式中， $\lambda = (w_1, u_1, \Sigma_1, \dots, w_K, u_K, \Sigma_K)$ 是 GMM 模型的参数向量，含有 K 个高核。 u_k 表示第 k 个高斯单元的平均值，是一个 D 维的向量， Σ_k 表示协方差矩阵（ $D * D$ ）。 w_k 表示第 k 个高斯单元的权重，并且满足 $\sum_{k=1}^K w_k = 1$ ，并且每个 w_k 都大于 0 小于 1。

通过式 4-3 可以计算输入特征 x_t 映射到高斯核 k 的权值：

$$\gamma_\lambda(k) = \frac{w_k u_k(x_t)}{\sum_{j=1}^K w_j u_j(x_t)} \quad (3-22)$$

进而通过累计直方图构建 BOW 模型，如 3-23 和 3-24 所示，

$$G_{u,k}^x = \frac{1}{T \sqrt{w_k}} \gamma_\lambda(k) \left(\frac{x_t - u_k}{\sigma_k} \right) \quad (3-23)$$

$$G_{\sigma,k}^x = \frac{1}{T \sqrt{2 * w_k}} \gamma_\lambda(k) \left(\frac{(x_t - u_k)^2}{\sigma_k^2} - 1 \right) \quad (3-24)$$

假定 x_t 的维度为 D ， $G_{u,k}^x$ 表示均值 u_k 的梯度（偏导数）， $G_{\sigma,k}^x$ 表示方差的梯度，二者的维度均和 x_t 的维度一样，都是 D 维。计算方差和均值的梯度是为了获取高斯模型对于输入特征的变化，Fisher Vectors 是把 K 个 $G_{u,k}^x$ 和 $G_{\sigma,k}^x$ 向量级联起组成了一个 $2 * K * D$ 维的特征。

在本文的研究中，还应用了在文献[15]中提出的两种可以提高 Fisher Vectors 效果的方法。第一是对其进行 L2 归一化，由于每份笔迹材料包含的 patch 数目并不一样，L2 归一化可以减轻这一影响，L2 归一化还对解决图片之间背景不同有帮助。

第二个提高方法是能量归一化（Power Normalization），随着高斯核的增多，Fisher

Vectors 也就会变得越来越稀疏。因为随着高斯核的增多, x_t 映射到每个高斯核的权重就会减小, 如果高斯核 i 的权值过小, 那么 $G_{u,k}^X$ 和 $G_{\sigma,k}^X$ 就会是近似于 0 向量。对于 L2 归一化以后的两个向量求 L2 距离, 就相当于求这两个向量的内积, 而内积用来做稀疏向量的相似度度量效果并不好, 因此我们需要使用功率归一化来减轻向量的稀疏性, 减弱那些出现频率很高的局部特征的影响。功率归一化的定义如下:

$$f(z) = \text{sign}(z)|z|^\alpha, \quad 0 \leq \alpha \leq 1 \quad (3-25)$$

这里的 α 是功率归一化参数, 其值应该根据高斯核的个数 K 来设定。实验中发现设定 $K = 256$ 时, 设置 $\alpha = 0.5$ 是比较合理的。我们把 L2 归一化和功率归一化结合起来使用, 先对 Fisher Vectors 做功率归一化, 再对其做 L2 归一化。

3.5.2 局域性约束线性编码

本文采用局部约束线性编码(Locality-constrained Linear Coding, LLC)^[47]来和 Fisher Vectors 进行对比实验。LLC 是最早为了做图像分类而提出的, 有着非常广泛的应用。LLC 应用局部约束性把输入特征映射到相应的码本, 相比于其他的稀疏编码(Sparse Coding)方式它更强调编码的局部性。文献^[49]指出 LLC 比传统的矢量量化编码或者类似的稀疏编码方式有更低的特征重建误差。

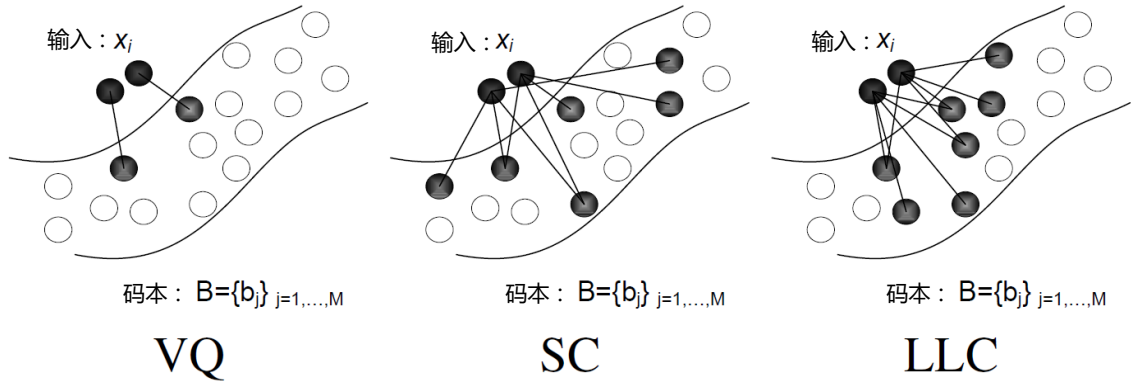


图 3-11 VQ, SC 与 LLC 的比较

假设 X 是一图像样本的特征集, 每个特向量的长度为 D , 即, $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$, B 是视觉单词的码本(Codebook), 也就是 $B = [b_1, b_2, \dots, b_M] \in R^{D \times M}$ 。LLC 可以公式 3-26 描述:

$$\begin{aligned} \min_c \sum_{i=1}^N ||x_i - Bc_i||^2 + \lambda ||d_i \odot c_i||^2 \\ \text{s.t. } 1^T c_i = 1, \forall i \end{aligned} \quad (3-26)$$

其中 \odot 表示向量元素之间的相乘运算， c_i 表示 x_i 编码后的结果，是一个 M 维的向量。 d_i 是权重，用来表示输入特征 x_i 到视觉单词的距离， $d_i \in R^M$ 。定义如下：

$$d_i = \exp\left(\frac{\text{dist}(x_i, B)}{\sigma}\right) \quad (3-27)$$

上式中 $\text{dist}(xi, B) = [\text{dist}(xi, b_1), \dots, \text{dist}(xi, b_M)]$ ，其中 $\text{dist}(xi, bi)$ 表示特征 x_i 和 b_i 的L2距离。 σ 用来调节 d_i 的大小，是权值衰减参数。我们通常会求得 $\text{dist}(xi, bi)$ 的最大值，从而对每个做 $\text{dist}(xi, bi)$ 归一化。对于3-26式，如果仅仅以向量含有0元素的个数来衡量稀疏性，那么其并不具有稀疏性。但是它的参数向量中往往只有少数几个大的参数作用，我们可以设定一个阈值，把那些非常小的系数设为0，以此增加其稀疏性。式3-28给出了LLC的解析解。

$$\begin{aligned} \tilde{c}_i &= C_i + \lambda \text{diag}(d) \\ c_i &= \tilde{c}_i / 1^T \tilde{c}_i \end{aligned} \quad (3-28)$$

上式中， $C_i = (B - 1x_i^T)(B - 1x_i^T)^T$ 为数据的协方差矩阵。这样就提高了计算的速度。作者与此同时还提出了一种LLC的快速近似算法，大幅提高了算法的效率。使用 x_i 的 K ($K < D < M$)最近邻 B_i 来替换原始的码本 B ，同时使用一个更加线性的公式来进行编码：

$$\begin{aligned} \min_{\tilde{c}} \sum_{i=1}^N ||x_i - c_i B_i||^2 \\ \text{s.t. } 1^T \tilde{c}_i = 1, \forall i \end{aligned} \quad (3-29)$$

这样一来时间复杂度就从 $O(M^2)$ 降低到了 $O(M+K^2)$ ，这里的 K 远小于 M 。本文在实验中采取的是LLC的快速算法，设 $K = 5$ 。

3.6 分类器

3.6.1 K近邻分类

本文的分类器选用K近邻分类(K Nearest Neighbor, KNN)，这是一种简单而

且应用广泛的分类方法。即给定一个已知样本标签的数据集，假定待分类的样本为 X ，KNN 首先需要找出 X 到已有数据集中每一实例的距离，然后选出距离最短的 K 个实例，那么 X 的类别就是这 K 个实例的最大类。

在笔迹鉴别的研究中，由于笔迹材料每个类的样本数目很少，有的可能只有一个样本。因此我们采取的是最近邻分类，即设定 $K=1$ 。求出 X 于笔迹库中距离最短的 1 个实例，然后把这个实例的类别作为 X 的类别。

3.6.2 距离度量

K 近邻分类的基于距离度量的，两个样本的距离越小，说明这两个样本越有可能属于同一类。距离度量的方法有很多，每一种都有不同的特性和使用范围。下面简要介绍几种比较常用的相似度量方法。为了表示方便，我们定义两个 N 维的向量， $A = (a_1, a_2, \dots, a_N)$ ， $B = (b_1, b_2, \dots, b_N)$ 。

(1) 欧式距离 (Euclidean Distance)

欧式距离，也称 L_2 距离，源自平面集合里面的两点距离计算，是最简单最常用的距离计算方法。定义如下：

$$\text{Dist}_{ED} = \sqrt{\sum_{i=1}^N (a_i - b_i)^2} \quad (3-30)$$

(2) 加权欧式距离 (Weighted Euclidean Distance)

欧式距离中向量每一维所占的权重是一样的，加权欧式距离就是使笔迹特征每一维所占的权重不一样，权重为训练样本特征在每一维的标准差。

$$\text{Dist}_{WED} = \sqrt{\sum_{i=1}^N \left(\frac{a_i - b_i}{\sigma_i} \right)^2} \quad (3-31)$$

(3) 余弦距离 (Cosine Distance)

余弦距离也是源自平面几何，在二维空间中，如果两个向量的余弦距离越大说明这两个向量越接近。

$$\text{Dist}_{cos} = \frac{\sum_{i=1}^N a_i b_i}{\sqrt{\sum_{i=1}^N a_i^2} \sqrt{\sum_{i=1}^N b_i^2}} \quad (3-32)$$

(4) 曼哈顿距离 (Manhattan Distance)

$$\text{Dist}_{MAN} = \sqrt{\sum_{i=1}^N |a_i - b_i|} \quad (3-33)$$

(5) 卡方距离 (χ^2 Distance)

$$\text{Dist}_{CHI} = \sqrt{\sum_{i=1}^N \frac{(a_i - b_i)^2}{a_i + b_i}} \quad (3-34)$$

实验发现, 加权欧式距离, 曼哈顿距离, 卡方距离对本文所提出的特征度量效果并不理想。欧式距离和余弦距离的效果比较理想, 且二者的实验结果完全一样。经过分析, 发现我们使用 Fisher Vectors 对 DCNN 特征进行编码以后做了 L2 归一化, 而对于 L2 归一化后的两个向量求欧式距离和余弦距离本质上一样的, 都和这两个向量的内积成比例关系。

3.7 本章小结

本章是本文的重点内容。首先着重介绍了本文所使用的 NET-WI 的网络结构, 并引入了 AlexNet 作为对比的 DCNN 网络结构。接着介绍了本文对于 DCNN 训练进行的改进, 包括本文提出的滑动窗字符切割, 数据集的归一化和扩充, 以及训练时参数的设置。然后介绍了 DCNN 特征的选择, 通过确定提取特征的 DCNN 网络层和主成份分析来选择 DCNN 特征。最后介绍了笔迹鉴别的另外两项重要内容, 即特征编码和分类器设计。通过对比两种常用的特征编码方式, 确定较优的方案。描述了 K 临近分类方法原理, 确定了适合本文笔迹特征的距离度量算法。

4 实验结果与分析

4.1 笔迹鉴别评价标准

笔迹鉴别的评价标准主要有三个，即平均正确率均值（Mean Average Precision, MAP）以及 Soft TOP-N 和 Hard TOP-N。下面将分别给出它们的定义。

4.1.1 平均正确率均值

MAP 是平均准确率（Average Precision, AP）的均值。我们假设有 N 个查询文档，则第 i ($i \in [1, N]$) 个查询文档的平均准确率 $AP(i)$ 为：

$$AP(i) = \frac{\sum_{k=1}^M P(k) * rel(k)}{R} \quad (4-1)$$

上式中， M 表示查询库中文件的总数， R 表示库中与查询文件相关的（属于同一类）文件总数。 $P(k)$ 表示查询结果中前 k 个结果的准确率，即前 k 个中相关文件的数目除以 k ， $rel(k)$ 是一个二值函数，如果第 k 个文件为相关文件，其值为 1，反之为 0。从而可以得到 MAP 的定义：

$$MAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (4-2)$$

4.1.2 Soft TOP-N 和 Hard TOP-N

Soft TOP-N 的评价方法和累计匹配特性值（Cumulative Matching Characteristic, CMC）是一样的。当返回前 N 个结果中有一个是相关的文档，我们就认为这是一次正确的查询，Soft TOP-N 就是所有正确的查询次数占查询总次数的百分比。而在 Hard TOP-N 的评价标准中，只有当返回前 N 个结果全部是相关的文档，我们才认为这是一次正确的查询。因此 Soft TOP-N 随着 N 递增，Hard TOP-N 随着 N 递减。

4.2 数据集介绍

4.2.1 CASIA-HWDB 中文数据集

CASIA-HWDB^[19] 是一个离线的中文笔迹数据集，是由中科院自动化研究所，模式识别国家实验室建立的。这个数据集包含了三个子集（DB2.0-2.2），本文使用其中

的 DB2.1 来做实验。

DB2.1 数据集又分为“DB2.1-A”和“DB2.1-B”两个子集。“DB2.1-A”是其中比较小的，包含有 60 个书写者，每个书写者有 5 分笔迹材料，总共包含 300 份笔迹材料，这个数据集用来做训练。“DB2.1-B”里面包含 240 个书写者，每个书写者同样提供 5 份笔迹材料，共 1200 份笔迹材料，这个数据集被用来做测试。在测试数据中，只使用每个书写者的前两份材料，一份用作查询材料一份用作库材料，这样保证了每个书写人在材料库中有唯一一份书写材料。这种划分方法和文献[50]的做法是一致的，从而为实验对比提供了方便。图 4-1 给出了两份书写人不同测试用本，图 4-2 给出了两份书写人相同测试样本。

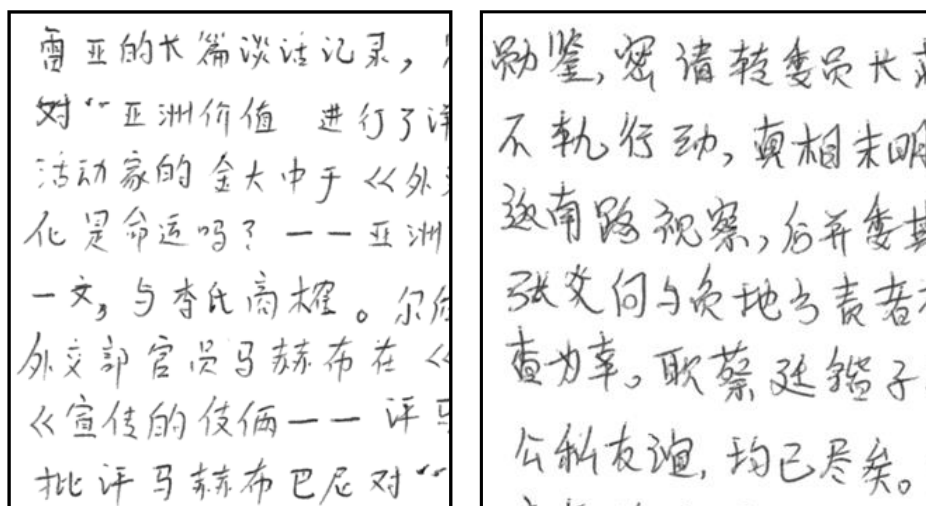


图 4-1 CASIA-HWDB 数据集中两份笔迹材料样本的局部（书写人不同）

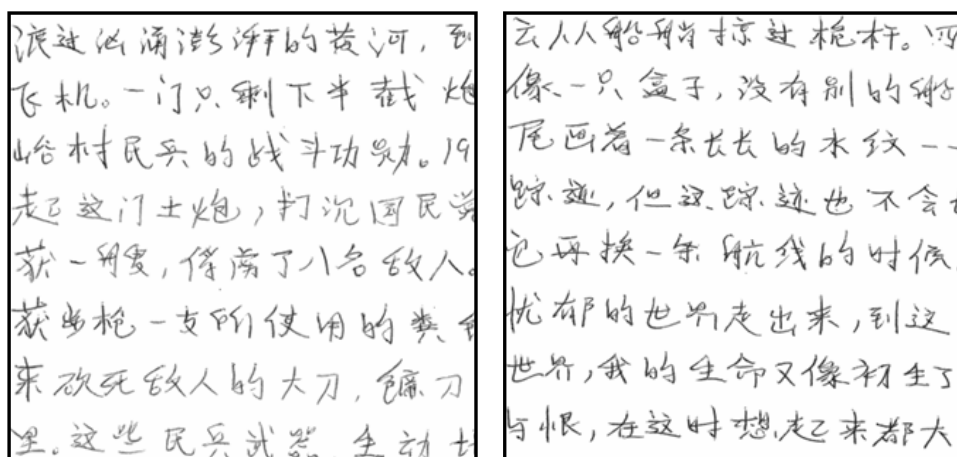


图 4-2 CASIA-HWDB 数据集中两份笔迹材料样本的局部（书写人相同）

可以看出想要分辨两份材料是否来自同一书写者并不是一件简单的事情，而笔迹鉴别需要从上百份笔迹材料中查找和待鉴别材料是同一书写人的样本，因而笔迹鉴别是一项很难的任务。

4.2.2 ICDAR 2013 数据集

ICDAR 2013 数据集^[24]是 2013 年国际文档分析与识别大会（International Conference on Document Analysis and Recognition, ICDAR）笔迹鉴别比赛所使用的数据集。训练集总共 100 个书写者，测试集总共有 250 个书写者，每个书写者提供四份笔迹材料（两份希腊文和两份英文），总共 1400 份笔迹材料。这些材料已经经过二值化处理，每份材料包含三至四行文本。这个数据集可以作为一个整体来使用，也可以把英文和希腊文拆分开来使用。这个数据集的难点在于跨语言的笔迹鉴别。以下两图是 ICDAR 2013 数据集中笔迹材料的展示。

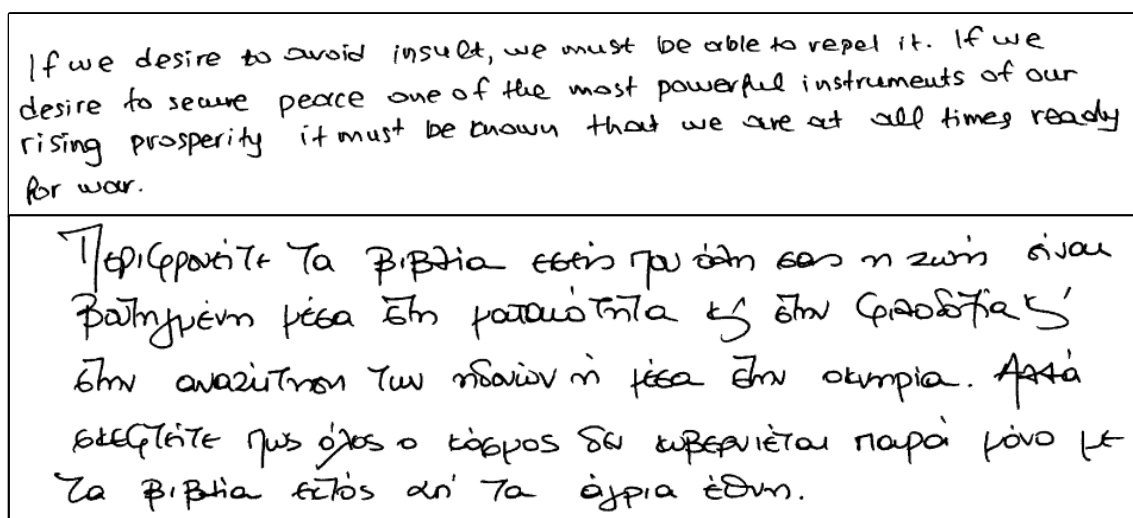


图 4-3 两位书写者的笔迹材料，分别为英文和希腊文图

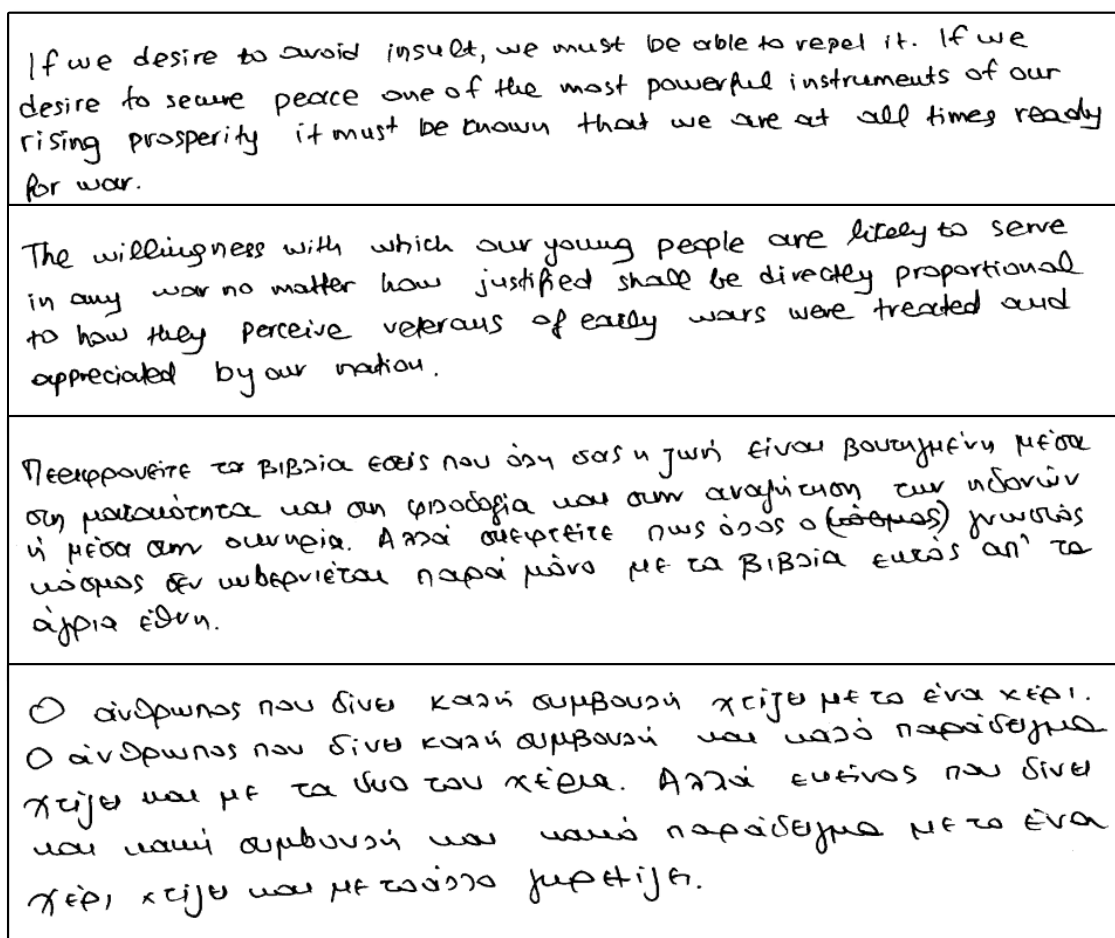


图 4-4 同一位书写者的 4 份笔迹材料，上两份为英语，下两份为希腊语

可以看出，同一书写人的四份笔迹材料的从直观上看很难找到共同点，不同语言的笔迹材料混合在一个数据集中增加了笔迹鉴别的难度。

4.3 本文方法对比实验

本节内容主要是为了验证本文提出的方法的有效性，以及对比不同方法的效果。本节给出了在中文数据集 CASIA-HWDB 上的对比实验结果，所有的对比实验都保证了单一的变化量，即除了需要比较的选项，实验其他的设置项完全一致。因为每个测试样本在测试集中只有一个与查询项相关的实例，本节所有的测试不统计 Hard TOP-N。下文表格数据中的最高得分会用粗体标记。

4.3.1 不同网络结果对比

本节主要比较本文所使用的网络 Net-WI 和改进的 AlexNet。AlexNet 比本文提出

的 Net-WI 大很多， AlexNet 的网络模型大小是 265M， 而 Net-WI 的模型大小仅为 14M， 可见 AlexNet 相应的权重参数也多很多。

本实验所采用的数据集是 CASIA-HWDB 中文数据集， Net-WI 和 AlexNet 均选择网络的输出层来提取特征， PCA 参数 β （详见式 3-19）均设为 0.995。其中 Net-WI 的特征经 PCA 降维后维度为 51， AlexNet 经 PCA 降维后维度为 48。均采用改进的 Fisher Vectors 编码， 设定 $K = 256$ ， $\alpha = 0.5$ 。二者均没有采用数据集扩展， 而是使用原始的数据集进行训练和测试。测试结果如表格 4-1 所示：

表格 4-1 Net-WI 和 AlexNet 结果比较

网络	TOP-1	TOP-2	TOP-5	TOP-10	MAP	模型大小
AlexNet	0.967	0.975	0.980	0.988	0.973	265 M
Net-WI	0.958	0.979	0.988	0.992	0.971	14 M

由表 4-1 可知， Net-WI 仅仅在 TOP-1 比 AlexNet 差了 1 个百分点， 其它的还略微高于 AlexNet。平均正确率均值也仅仅比 AlexNet 低 0.2%， 可以说两个网络的效果相差非常小。从图 4-5 中我们也可以看出两个网络的效果相当。

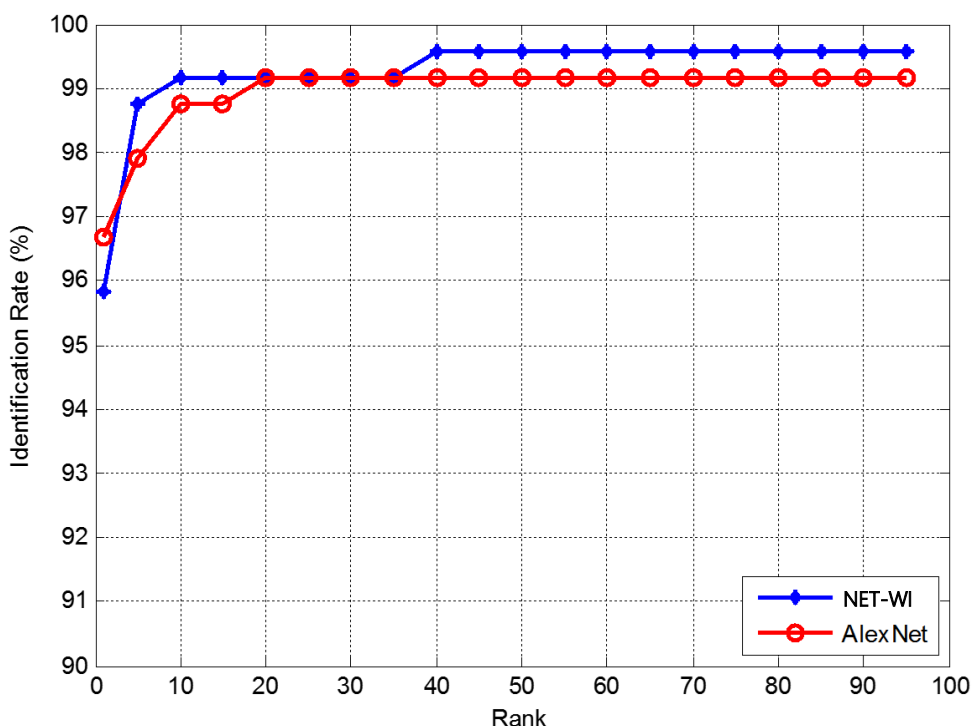


图 4-5 Net-WI 和 AlexNet 结果比较

由以上实验分析可知, AlexNet 对于笔迹鉴别的效果相比于 Net-WI 并没有明显的优势。原因是本文利用 DCNN 来对笔迹图像的 Patch 提取局部特征, 由于笔迹图像的 Patch 尺寸很小, 而 AlexNet 又有多层卷积和下采样, 并不适合对这种小图像提取特征。而本文提出的 Net-WI 虽然网络层数没有 AlexNet 多, 却对于用来提取笔迹图像的局部特征有很好的效果。另一个可能原因是 AlexNet 的参数太多, 而训练的样本较少, 容易产生过拟合。考虑到 Net-WI 在速度和存储方面的优势, 本文选用 Net-WI 来进行接下来的实验。

4.3.2 数据集扩充的效果

本实验采用 Net-WI 网络来测试本文 3.3.2 所述的数据集扩充的效果。实验均采用网络的 Fc2 层作为输出特征, 输出的特征维度为 60。PCA 参数 β 为 0.995, 经 PCA 降维后的维度均为 51 维。均采用 Fisher Vectors 编码, 其参数和上一个实验保持一致。本实验分别对不做样本集扩充, 只扩充测试样本集, 训练样本集和测试样本集都扩充做了详细的测试, 测试的具体结果如表格 4-2 所示。

表格 4-2 数据集的扩充

数据集是否扩充	TOP-1	TOP-2	TOP-5	TOP-10	MAP
不扩充	0.921	0.963	0.983	0.992	0.950
测试集扩充	0.95	0.967	0.971	0.983	0.963
训练集和测试集扩充	0.971	0.975	0.983	0.988	0.977

分析以上测试结果我们可以知道, 对于测试集的扩充对于效果的提升比较明显, TOP-1 提高了 2.9 个百分点, MAP 提高了 1.3 个百分点, 说明了测试样本集的扩充对于减小编码的重建误差有一定的作用。训练集和测试集一起扩充, 整体的效果得到了进一步的提升, TOP-1 的准确率比完全不扩充提高了 5%, MAP 提高了 2.7%。由此可见, 本文样本集扩充的方法对于笔迹鉴别的整体效果有比较明显提升作用, 图 4-6 是上述表格数据更为直观的展示。

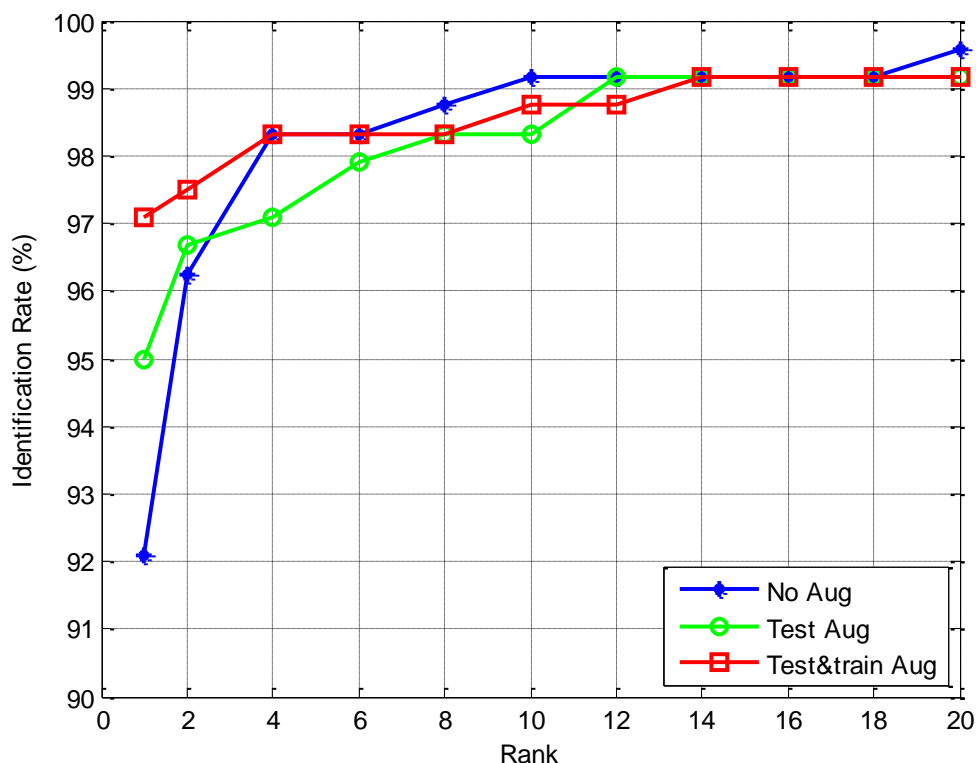


图 4-6 数据集的扩充

4.3.3 主成份分析的影响

本实验主要研究主成份分析对最终笔迹鉴别效果的影响。依然采用 CASIA-HWDB 中文数据集，所用的网络是 Net-WI，训练和测试样本集均做了扩充。并且均采用网络的 Fc2 层作为输出特征。输出的特征维度跟训练集类别一样为 60 维，PCA 参数 β 为 0.995，经 PCA 降维后为 51 维。经过 Fisher Vector 编码后特征长度分别为：30720，26112。实验结果如表格 4-3，图 4-7 所示。

表格 4-3 PCA 对实验结果的影响 (Fc2 层)

比较项	TOP-1	TOP-2	TOP-5	TOP-10	MAP	特征长度
NO PCA	0.963	0.979	0.988	0.988	0.973	60
PCA	0.971	0.975	0.983	0.988	0.977	51

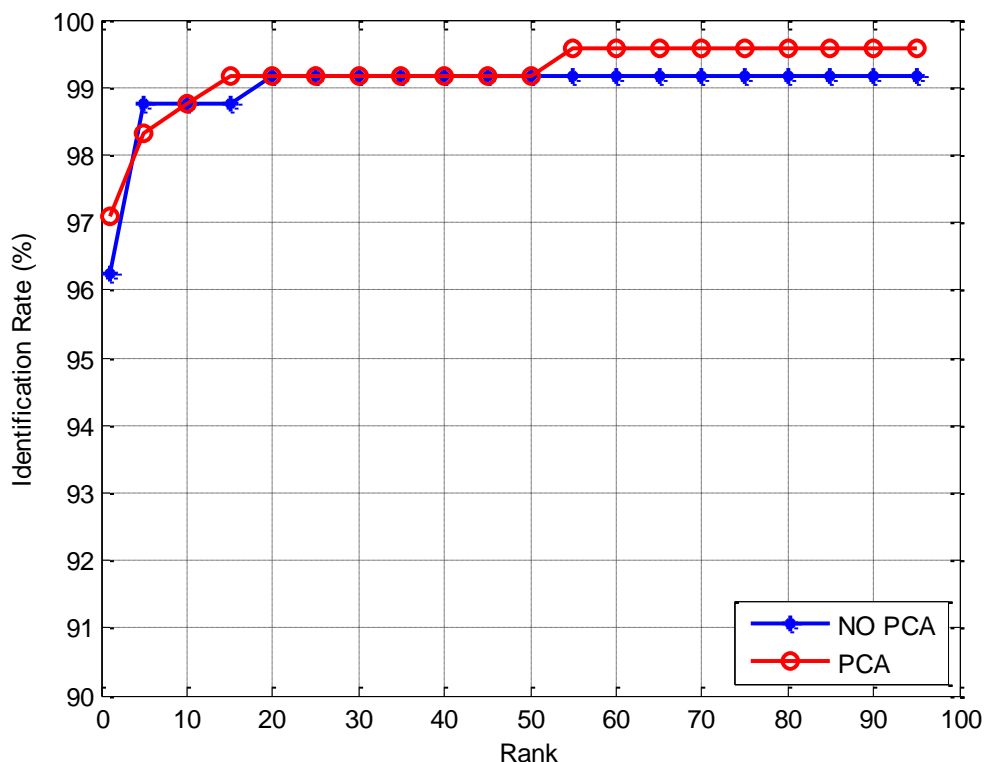


图 4-7 PCA 对实验结果的影响 (Fc2 层)

有以上实验数据可知，使用 PCA 之后对笔迹鉴别的整体效果有一定提升，其中 TOP-1 的准确率提高了 1.2%，MAP 提高了 0.4%。由于 DCNN 特征本身维度才 60 维，含有的冗余信息比较少，所以降维的效果不是很明显（15%）。但是如果采用 Fc1 层（倒数第二个全连接层）特征，PCA 的参数 β 同样设置为 0.995，降维效果就十分明显，原始 Fc1 层特征维度为 1024，降维后维度仅仅为 172。降维幅度达 82.3，如果实际应用中特征库规模比较大的话，能很大程度上减少存储空间并且加快检索的速度。表格 4-4 为 Fc1 层采用 PCA 的效果对比。

表格 4-4 PCA 对实验结果的影响 (Fc1 层)

比较项	TOP-1	TOP-2	TOP-5	TOP-10	MAP	特征长度
NO PCA	0.954	0.979	0.983	0.988	0.969	1024
PCA	0.946	0.979	0.988	0.992	0.966	172

4.3.4 不同网络层的特征对比

为了研究 DCNN 网络的哪一层提取的特征更适合做笔迹鉴别, 本实验分别提取 Net-WI 最后四层网络 (Pool3, RELU4, Fc1, Fc2) 的特征。所有测试所采用了一样的参数, PCA 参数 β 设为 0.995, 均采用改进的 Fisher Vectors 编码, 设定 $K = 256, \alpha = 0.5$ 。全部都采用扩展的数据集训练和测试。实验结果如表格 4-5 所示。

表格 4-5 不同层网络特征对比

网络层	TOP-1	TOP-2	TOP-5	TOP-10	MAP	DCNN 维度	PCA 维度
Pool3	0.692	0.746	0.792	0.825	0.742	3584	2517
Relu4	0.871	0.942	0.975	0.983	0.919	1024	650
Fc1	0.946	0.979	0.987	0.992	0.966	1024	172
Fc2	0.971	0.975	0.983	0.988	0.977	60	51

可以看出 Pool3 层提取特征的整体表现很差, 可能的原因有两个, 一是 Pool3 网络层深度不够, 且这层之前没有全连接层, 提取的特征区分性不够。二是 Pool3 提取的特征维度过大, 经 PCA 降维之后 2517 维, 经 Fisher Vectors 编码以后维度扩大到 1288704, 这么高的维度并不利于做检索。整体表现最好的是 Fc2 层 (最后一个全连接层), 它的 MAP 比 Fc1 高 1.1%, TOP-1 比 Fc1 高 2.5%。除此之外, Fc1 的特征维度最短, 在笔迹鉴别的速度上面更具有优势。因此, 本文确定选用 Fc2 层特征。图 4-8 给出了更为直观的对比效果。

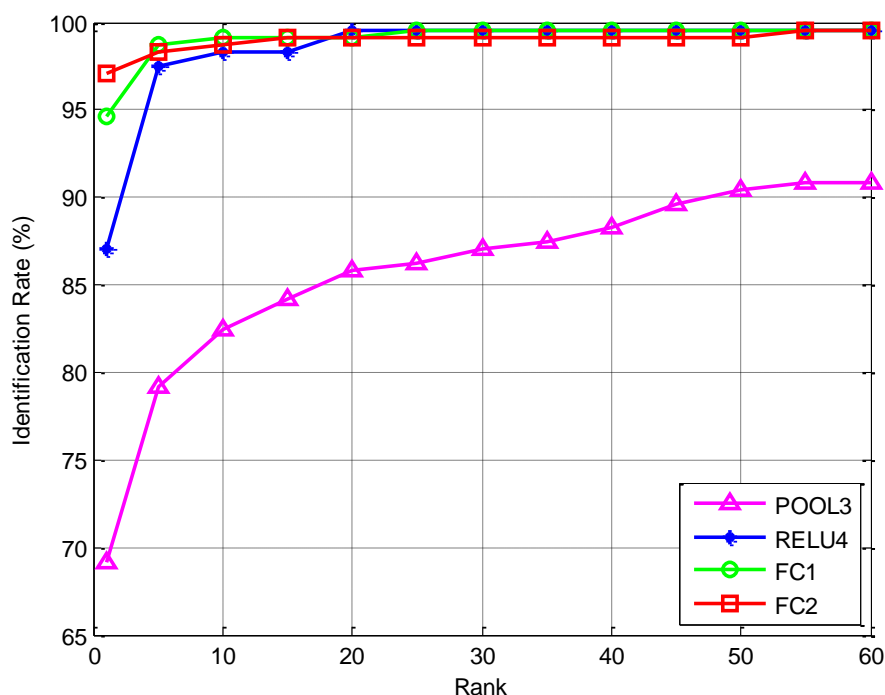


图 4-8 不同层网络特征对比

4.3.5 两种不同编码方式的比较

本节对比 LLC 和 Fisher Vectors 对 DCNN 特征编码的效果，LLC 编码和 Fisher Vectors 都是先对每个 Patch 提取 DCNN 特征，然后再对所有编码得到的特征求均值（Average Pooling）。Fisher Vectors 的设置和之前实验的设置一样，LLC 采用 K-means 聚类，聚类中心设置为 1024，则编码后得到的特征为 1024 维。为了尽量维持比较的单一性，二者均没有使用 PCA，取得都是 DCNN 最后一层的特征（Fc2）。二者在 CASIA-HWDB 中文数据集上的表现如表格 4-6 所示。

表格 4-6 Fisher Vectors 和 LLC 对比

编码方式	TOP-1	TOP-2	TOP-5	TOP-10	MAP
LCC	0.867	0.913	0.963	0.971	0.907
Fisher Vectors	0.954	0.979	0.983	0.988	0.969

可以看出在笔迹鉴别这一研究领域，Fisher Vectors 与 DCNN 特征的结合相比于 LLC 与 DCNN 特征结合有比较大的优势，TOP-1 比 LLC 高 8.7%，MAP 也有 6.2% 领先。图 4-9 给出了更为直观的对比效果。

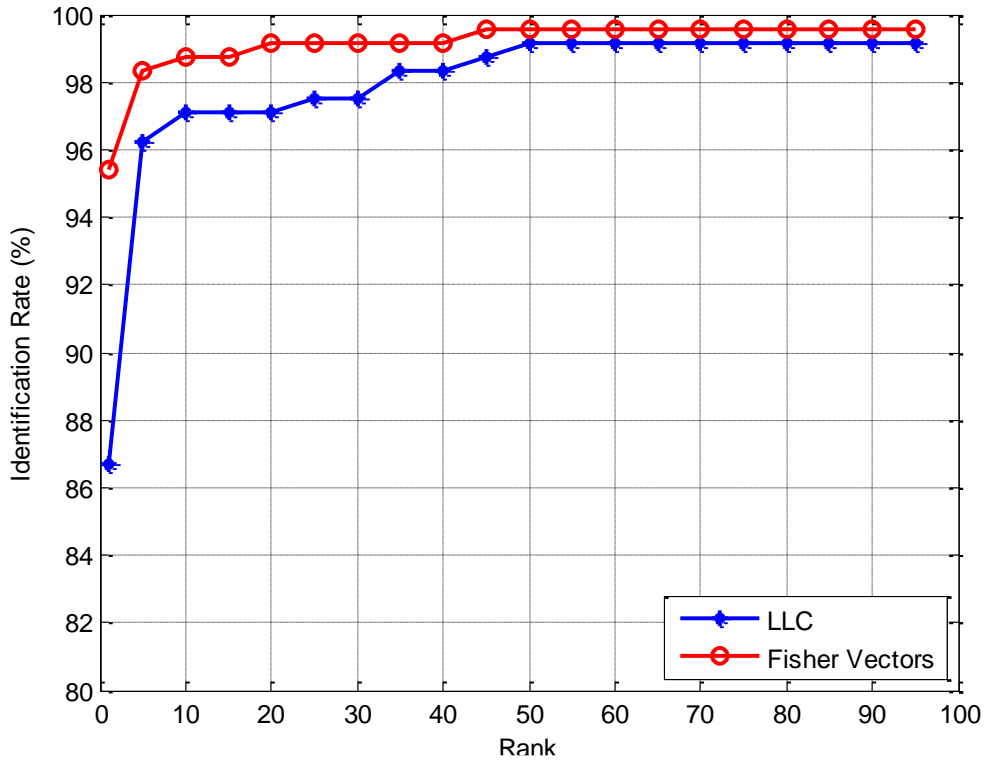


图 4-9 Fisher Vectors 和 LLC 对比

4.4 与其他方法的比较

本节所有实验都采用相同的设置，所用的网络是 Net-WI（最后一层的节点数目和训练数据的类别一样），并且均采用网络的 Fc2 层作为输出特征。PCA 参数 β 为 0.995，均采用改进的 Fisher Vectors 编码，设定 $K = 256$ ， $\alpha = 0.5$ 。下述表格数据中的最高得分会用粗体标记。

4.4.1 CASIA-HWDB 实验结果与比较

为了验证本文方法的有效性，本节将本文的方法和近年来提出的文本无关的中文笔迹鉴别的其它方法进行了比较。本文在 CASIA-HWDB 上训练集和测试集的划分是和文献^[50]采取的划分方法一样的，表格 4-7 给出了本文方法与他人方法的对比结果。

表格 4-7 在 CASIA-HWDB 数据集上的比较

比较项	书写人数	TOP-1	TOP-2	TOP-5	TOP-10	MAP
Zhu ^[53]	17	0.957	<i>null</i>	<i>null</i>	<i>null</i>	0.886
He ^[54]	500	0.364	0.436	0.604	0.746	<i>null</i>
Du ^[55]	50	0.670	0.810	<i>null</i>	0.970	<i>null</i>
He ^[56]	10	0.800	<i>null</i>	<i>null</i>	<i>null</i>	<i>null</i>
Li ^[57]	240	0.900	<i>null</i>	<i>null</i>	0.971	<i>null</i>
Hu ^[50]	240	0.963	<i>null</i>	<i>null</i>	0.996	<i>null</i>
本文方法	240	0.971	0.975	0.983	0.988	0.977

如表格 4-7 所示, 本文的方法相对于以往的离线文本无关的中文笔迹鉴别方法有比较大的优势。但是本文方法和文献[50]的方法总体差距最小, 在 TOP-1 本文的方法比它高 0.8%, 在 TOP-10 本文比它低 0.8%。由于每个查询文档在数据集中只有一个与之相关的文档, 因而 TOP-1 对于 MAP 的贡献率是 TOP-10 的 10 倍, 总体来说应该是本文方法占优(文献[50]未给出 MAP 值)。另外, 文献[50]所用的特征是人工设计的局部特征(SIFT), 不需要训练, 而本文所使用的基于 DCNN 的特征是基于大规模数据训练的, 然而训练集只有 60 个类别(测试集 240 个类别), 这对于训练 DCNN 来说数据是比较缺乏的。因此可以合理推测, 如果加大训练数据量, 本文的方法能够取得更好的结果。

4.4.2 ICDAR2013 实验结果与比较

本数据集测试时严格遵照 ICDAR2013 笔迹鉴别比赛^[24]的规则, 采用留一法(Leave-One-Out), 即每次从测试集里面取出一张图片来对剩余的图片查询。我们对于实验结果统计 Soft TOP-N, Hard TOP-N 和 MAP。实验结果如所示。一共进行三组不同的实验, 第一组实验使用 ICDAR2013 数据集中的希腊文(500 份笔迹材料), 第二组实验使用 ICDAR2013 数据集中的英文(500 份笔迹材料), 第三组实验使用数据集的全部数据(1000 份笔迹材料)。前两组实验, 由于每个书写者只有两份笔迹材料, 因只统计 Soft TOP-N, 对于第三组实验统计 Soft TOP-N 和 Hard TOP-N。前两组实验

结果如表格 4-8，表格 4-9 所示。

表格 4-8 在 ICDAR2013 希腊文数据集上的测试结果(Soft TOP-N)

方法	TOP-1	TOP-2	TOP-5	TOP-10	MAP
SIFT+FV [48]	0.884	0.920	0.968	0.978	<i>null</i>
SIFT+SOH [58]	0.938	0.964	0.972	0.978	<i>null</i>
Edge + Runs [59]	0.926	0.960	0.980	0.984	<i>null</i>
CS ^[10]	0.956	0.982	0.986	0.992	<i>null</i>
K&S&C* [12]	0.992	0.996	0.998	0.998	0.995
本文方法	0.994	0.996	1	1	0.996

表格 4-9 在 ICDAR2013 英文数据集上的测试结果(Soft TOP-N)

方法	TOP-1	TOP-2	TOP-5	TOP-10	MAP
SIFT+FV [48]	0.922	0.946	0.964	0.966	<i>null</i>
SIFT+SOH [58]	0.914	0.942	0.958	0.972	<i>null</i>
Edge + Runs [59]	0.912	0.934	0.962	0.966	<i>null</i>
CS ^[10]	0.946	0.970	0.984	0.988	<i>null</i>
K&S&C* [12]	0.974	0.978	0.986	0.988	0.979
本文方法	0.984	0.986	0.990	0.992	0.987

可以看出，每种方法在希腊文的表现都要英文要好，说明希腊文的笔迹鉴别难度要小于英文。表格 4-9 中 CS 方法是 ICDAR2013 笔迹鉴别比赛的最好结果，K&S&C 的是基于特征融合的方法，它也使用了 Fisher Vectors 进行编码。可以看出本文的方法在各项评价指标中均占优，在希腊文数据集上面，本文方法和 K&S&C 差别不大，但是在较难的英文数据上，本文方法有比较明显的优势。说明了 DCNN 特征比传统的特征要好。第三组实验结果如表格 4-10 所示，本文方法也取得了不错的结果，再一次验证了其有效性。

表格 4-10 ICDAR 全部数据测试，英文和希腊文一起（Hard TOP-N）

方法	TOP-1	TOP-2	TOP-3	MAP
CS ^[10]	0.951	0.196	0.071	<i>null</i>
SIFT+SOH ^[59]	94.8	0.632	0.365	<i>null</i>
SV ^[17]	0.971	0.428	0.238	0.671
本文方法	0.990	0.667	0.465	0.865

4.5 实验平台及时间

4.5.1 训练实验平台及时间

DCNN 的训练在一台专用的服务起上面进行，操作系统是 Linux(CentOS release 6.4)。具体的软硬件环境如下：

表格 4-11 训练实验平台环境

处理器	Intel(R) Xeon(R) CPU E5-2407 v2 @ 2.40GHz
显卡	Nvidia GeForce GTX 760 （4G 显存）
GCC 版本	4.4.7 20120313

Net-WI 和 AlexNet 的训练时间如表格 4-12 所示（训练次数都为 80000），可以看出文本所使用的 Net-WI 所需训练时间较少。

表格 4-12 DCNN 的训练时间

网络	时间
AlexNet	19 小时 26 分钟
Net-WI	9 小时 29 分钟

4.5.2 测试实验平台及时间

测试机器是 windows7 系统，程序基于 C++语言开发，具体的开发环境如表格 4-13 所示。

表格 4-13 测试实验平台环境

操作系统	Windows 7 旗舰版 64 位 SP1 (DirectX 11)
处理器	Intel(R) Xeon(R) E3-1230 V2 @ 3.30GHz
显卡	Nvidia GeForce GTX 650 (1G 显存)
编译器	VS2010

测试的各项时间统计如下：

表格 4-14 测试时间统计

提取单张笔迹图片特征的平均时间	4.5 秒
单次查询时间（库中笔迹样本 240 份）	12.2 毫秒
单次计算欧式距离时间	0.05 毫秒

4.6 本章小结

本章首先给出了笔迹鉴别的常用评价标准，接着介绍了本文所使用的中文数据集和外文数据集。然后通过实验验证了本文所提出的方法的有效性，并且把本文的方法与他人的方法做了充分的比较，证明了本文方法的先进性。

5工作总结与展望

5.1 全文总结

笔迹鉴别的应用领域非常广泛，随着人工智能和计算机视觉技术日新月异的发展，笔迹鉴别在成为了一项越来越不可或缺的技术。近年来，众多的研究者在笔迹鉴别这一问题上做出了许多突破性的工作，然而文本无关的离线笔迹鉴别仍然有很多不成熟的地方，有很多问题有待于我们进一步研究。因此，本文基于深度卷积神经网络对文本无关的离线笔迹鉴别进行了研究，主要的研究内容包括以下几个方面：

(1) 由于笔迹鉴别这一研究问题的复杂性，传统的笔迹鉴别大多是往往是把几种不同的人工设计的特征融合在一起来形成更有区分性的特征。但是这样的做法会存在计算量较大，算法设计复杂，各种特征之间难以调和等问题。本文首次把基于深度卷积神经网络的特征用于文本无关的离线中文笔迹鉴别，并通过实验证明，本文所采用的基于深度卷积神经网络的特征相比于传统特征融合的方法能获得更高的准确率。

(2) 对比不同 DCNN 的网络结构对于笔迹鉴别结果的影响。研究 DCNN 的训练。提出了滑动窗字符切割的方法，研究数据归一化，数据集扩展对最终结果的影响。为了选择出更有区分性的特征，采用了主成份分析都 DCNN 的特征进行处理，并且对 DCNN 不同网络层的特征做了比较。

(3) 研究特征编码算法，采用 Fisher Vectors 算法对提取的 DCNN 特征进行编码，并和 LLC 编码技术做了对比。

(4) 首先通过大量细致的实验对本文提出的方法进行验证，评估不同的算法对最终结果的影响。接着在中文数据集上面对本文方法进行验证，并取得了有竞争力的结果。然后在包含两种语言的外文数据集上进行验证，也取得了领先的结果，证明了本文算法有很强的适应性和鲁棒性。

5.2 工作展望

深度学习的研究正处于蓬勃发展的阶段，近年来，深度卷积神经网络在图像分类，文字检测，文字识别等领域更是取得的许多令人瞩目的成就。而本文对于把深度卷积神经网络的应用在笔迹鉴别上面做了初步的探索，但是依然存在许多不足的

地方。因而把深度学习和文本无关的离线笔迹鉴别相结合还有很多有用的工作可以做，可以从以下几个方面进行更深层次的研究。

(1) 进一步研究相似度度量方法，可以采用分类的思路来替换简单的欧式距离计算。在用 DCNN 提取特征的基础上，可以再训练一个 DCNN 模型用来区分两份笔迹材料的书写者是否为同一人，从而可以得到一个分类的得分，用这个得分作为两份笔迹材料的距离。

(2) 研究改进已有的网络结构，借鉴深度学习领域最新的研究成果，从而帮助提取更具区分性的特征。尝试新的训练方法，可以采取无监督学习和 DCNN 结合，先对训练数据做聚类（如 k 均值聚类），以聚类的结果作为笔迹图片的标签来训练 DCNN 网络。

(3) 如果需要鉴别的笔迹材料包含的字数太少，那么采用 Fisher Vectors 编码的误差就会比较大，进而影响鉴别的效果。针对这一问题，在编码方式上还值得做进一步的研究。

致 谢

参考文献

- [1] A. Fornes, J. Lladós, G. Sanchez. Writer Identification in Old Handwritten Music Scores. In: 8th IAPR Workshop on Document Analysis Systems, 2008. 347-353
- [2] J. Sas. Handwriting Recognition Accuracy Improvement by Author Identification. Artificial Intelligence and Software Computing, 2006(4029): 682-691
- [3] L. Schomaker. Advances in Writer Identification and Verification. In: 9th International Conference on Document Analysis and Recognition, 2007(2): 1268-1273
- [4] M. Tapiador, JA. Sigüenza. Writer Identification Method Based on Forensic Knowledge - Biometric Authentication. First International Conference, 2004(3072): 555-561
- [5] A. Schlappbach, H. Bunke. A Writer Identification and Verification System Using HMM Based Recognizers. Pattern Anal. 2007,10(1): 33-43
- [6] A. Schlappbach, L. Marcus, H. Bunke. A writer identification system for on-line whiteboard data. Pattern Recognition Journal. 2008,41(7): 2381-2397
- [7] L. Schomaker, Advances in Writer identification and verification. In: Ninth International Conference on Document Analysis and Recognition. 2007(2): 1268-1273
- [8] B. Marius, S. Lambert. Text-Independent Writer Identification and Verification Using Textural and Allographic Features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007,29(4): 701-717
- [9] R. Huber, A. Headrick. In Handwriting Identification: Facts and Fundamentals. The Fundamentals of the Identification Process. CRC Press, 1999
- [10] R. Jain, D. Doermann. Writer Identification Using an Alphabet of Contour Gradient Descriptors. International Conference on Document Analysis and Recognition (ICDAR), 2013. 550-554
- [11] G. Louloudis, B. Gatos, N. Stamatopoulos. "ICDAR 2013 Competition on Writer Identification," International Conference on Document Analysis and Recognition

- on (ICDAR), 2013. 1397-1401
- [12] R. Jain, D. Doermann. Combining local features for offline writer identification. In: 14th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE, 2014
- [13] V. Ferrari, L. Fevrier, F. Jurie, et al. Groups of adjacent contour segments for object detection. *Pattern Analysis and Machine Intelligence. IEEE Transactions.* 2008,30(1): 36-51
- [14] H. Bay, T. Tuytelaars, L. Van Gool. Surf: Speeded up robust features. Springer, 2006. 404-417
- [15] F. Perronnin, J. Sanchez, T. Mensink. Improving the fisher kernel for large-scale image classification. In: *Computer Vision—ECCV*, Springer, 2010. 143-156
- [16] A. Nicolaou, AD. Bagdanov, M. Liwicki, et al. Sparse Radial Sampling LBP for Writer Identification. In: *International Conference on Document Analysis and Recognition (ICDAR)*, 2015. 716-720
- [17] V. Christlein, D. Bernecker, F. Hoenig, et al. Writer identification and verification using gmm supervectors. In: *Winter Conference on Applications of Computer Vision*. IEEE, 2014
- [18] Weixin Yang, Lianwen Jin, Manfei Liu. Chinese Character-level Writer Identification using Path Signature Feature, DropStroke, and Deep CNN. In: *International Conference on Document Analysis and Recognition (ICDAR)*, 2015
- [19] C.L. Liu, F. Yin and D.H. Wang et al. CASIA online and offline Chinese handwriting databases. In: *ICDAR*, 2011
- [20] B. Helli, M.E. Moghaddam. A writer identification method based on XGabor and LCS. *IEICE Electron. Express*, 2009(6): 623-629
- [21] B. Helli, M.E. Moghaddam. A text-independent Persian writer identification based on feature relation graph (FRG). *Pattern Recogn*, 2010(43): 2199-2209
- [22] Y. Yan, Q Chen, W. Deng, F Yuan. Chinese Handwriting Identification Based on Stable Spectral Feature of Texture Images. *International Journal of Intelligent Engineering and Systems*, 2009
- [23] Z.He, X. You, Y.Y. Tang. Writer identification of Chinese handwriting documents

- nts using hidden Markov tree model. Pattern Recognition Journal,2008(41): 1295–1307
- [24] G. Louloudis, B. Gatos, N. Stamatopoulos , et al. ICDAR 2013 Competition on Writer Identification. In: International Conference on Document Analysis and Recognition (ICDAR), 2013. 1397-1401
- [25] W. McCulloch, W. Pitts. A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 1943(7): 115–133
- [26] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 1958, 65(6): 386
- [27] M. Minsky, S. Papert. An Introduction to Computational Geometry. MIT Press, 1987
- [28] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the USA, 1982,79(8): 2554-2558
- [29] D.E. Rumelhart, G.E. Hinton, R.J. Williams. Learning representations by back-propagating errors. Nature, 1986 (323) : 533-536
- [30] Y. LeCun, Y. Bengio. Convolutional networks for images, speech, and time series. handbook of brain theory and neural networks, 1997
- [31] G.E. Hinton, S. Osindero, T. YW. A Fast Learning Algorithm for Deep Belief Nets. 2006, 18(7): 1527–1554
- [32] J. Shuiwang, X. Wei , Y. Ming, et al. 3D Convolutional Neural Networks for Human Action Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231
- [33] L. Honglak, G. Roger, R. Rajesh, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning, 2009. 1-8
- [34] S. Fiel, R. Sablatnig. Writer Retrieval and Writer Identification Using Local Features. In: International Workshop on Document Analysis Systems, 2012. 145-149
- [35] B. Graham. Sparse arrays of signatures for online character recognition. Encyclopedia

- of Astrobiology, 2013, 135(3410): 89-90
- [36] V. Nair, G.E. Hinton. Rectified linear units improve restricted boltzmann machines. In: Proceeding of the 27th International Conference on Machine Learning, 2010. 21-24
- [37] X. Glorot, A. Bordes, Y. Bengio. Deep sparse rectifier networks. Learning/Statistic and Optimisation, 2011(15): 315-323
- [38] H.E. Said, T.N. Tan, K.D. Baker. Personal identification based on handwriting. Pattern Recognition, 2000, 33(1): 149-160
- [39] H. Murase, S. Nayar. Detection of 3D objects in cluttered scenes using hierarchical eigenspace. Pattern Recognition Letters, 1997, 18(4): 375-384
- [40] K. Fukunaga, W. Koontz. Application of the Karhunen-Loeve expansion to feature selection and ordering. IEEE Trans. Communications, 1970, 19(4): 311-318
- [41] M. Turk, A. Pentland. Face Recognition Using Eigenfaces. In: Proceedings of Computer Vision and Pattern Recognition, 1991
- [42] Y. Ke, R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In Proc. CVPR, 2004. 506-513
- [43] D. G. Lowe. Object recognition from local scale-invariant features. In: Proceedings of International Conference on Computer Vision, 1999, 1150-1157
- [44] T. Ojala, M. Pietikainen, T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions 2002, 24(7): 971-987
- [45] I.T. Joliffe. Principal Component Analysis. Springer-Verlag, 1986
- [46] S. Fiel, R. Sablatnig. Writer Identification and Retrieval Using a Convolutional Neural Network Computer. In: Analysis of Images and Patterns: 16th International Conference, 2015. 26-37
- [47] J. Wang, J. Yang, K. Yu, et al. Locality-constrained linear coding for image classification. Proc. CVPR, 2010
- [48] S. Fiel and R. Sablatnig, Writer Identification and Writer Retrieval using the Fisher Vector on Visual Vocabularies. In: International Conference on Document Analysis and Recognition (ICDAR), 2013. 545-549

- [49] J. Yang, K. Yu and Y. Gong, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. In: CVPR, 2009
- [50] Y. Hu, W. Yang and Y. Chen. Bag of Features Approach for Offline Text-Independent Chinese Writer Identification. ICIP, 2014.
- [51] M.D. Zeiler, R. Fergus. Visualizing and understanding convolutional networks. In: Proc. ECCV, Springer, 2014. 818-833
- [52] G.E. Hinton, N. Srivastava, A. Krizhevsky, et al. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580, 2012
- [53] Y. Zhu, T. Tan, Y. Wang. Biometric Personal Identification Based on Handwriting. In: International Conference on Pattern Recognition, 2000
- [54] Z. He, X. You, Y.Y. Tang. Writer identification of Chinese handwriting documents using hidden Markov tree model. Pattern Recognition, 2008(41): 1295-1307
- [55] L. Du, X. You, H. Xu, et al. Wavelet Domain Local Binary Pattern Features for Writer Identification. In: ICPR, 2010. 3691-3694
- [56] Z. He, B. Fang, J. Du, et al., A Novel Method for Off-line Handwriting-based Writer Identification. In ICDAR, 2005(1): 242-246
- [57] X. Lin, X. Ding. Writer Identification of Chinese Handwriting Using Grid Microstructure Feature. In: Third International Conference of Advances in Biometrics, 2009(5558): 1230-1239
- [58] X. Wu, Y. Tang, W. Bu. Offline Text-independent Writer Identification Based on Scale Invariant Feature Transform. IEEE Transactions on Information Forensics and Security, 2014,9(3): 526-536
- [59] C. Djeddi, I. Siddiqi, L. Souici-Meslati, et al. Text-independent writer recognition using multi-script handwritten texts. Pattern Recognition Letters, 2013, 34(10): 1196-1202