

基于多流卷积神经网络的中文笔迹鉴别研究^①

毛颖裕^{②**} 张怡龙^{*} 王海霞^{③*}

(* 浙江工业大学计算机科学与技术学院 杭州 310023)

(** 浙江工业大学信息工程学院 杭州 310023)

摘要 针对中文笔迹鉴别中局部特征与全局特征的提取与融合问题,本文提出一种基于多流卷积神经网络的中文离线笔迹鉴别方法。该网络共有 3 条支流结构,其中 1 条支流对整图进行全局特征提取,另外 2 条支流分别对原图采用水平分割和垂直分割方式来获取局部数据并进行特征提取,而后将提取到的特征与全局特征进行多尺度融合。本文采用了 top-1 和 top-5 2 种准确率进行性能评估并进行了作者级别的准确率对比分析。实验结果表明,本文所提出的结构不仅可以更好融合、利用局部和全局特征还能拓宽网络获取细节特征的来源,改善网络性能。

关键词 笔迹鉴别;特征提取;模式识别

0 引言

笔迹鉴别是通过手写字体信息来鉴别书写人身份的一种文件分析技术。它是一个典型的模式识别问题,在历史文献分析^[1]和司法鉴定^[2]中有着广泛的应用。根据笔迹样本的获取方式,可以将笔迹鉴别分成离线笔迹鉴别^[3,4]与在线笔迹鉴别^[5-7]2 种。在线笔迹鉴别要比离线笔迹鉴别相对容易,因为在线书写的过程中可以收集到更多的作者书写习惯信息,包括笔尖压力^[8]、书写速度^[9]、笔画顺序等,而这些动态信息在离线笔迹鉴别过程中无法获取。离线笔迹样本的获取,需要书写者使用书写工具(钢笔、铅笔等)在纸上书写完成后进行扫描或者拍摄。其相对于在线笔迹样本具有更强的通用性和易获取性,因此离线笔迹鉴别的应用范围更广、实用性更高。在数字时代,手写活动大幅减少,这就带来了基于少量可用文本来识别作者的需求。与英文字体相比,中文字体的形态结构通常更复杂,这意味着

中文字体在单个单词中比英文字体包含更多的判别信息,并且有可能用更少的文字被识别。本文研究的就是基于单个手写中文字体的离线笔迹鉴别问题。

离线笔迹鉴别受到多种因素的影响,譬如,书写人在书写时的现实物理环境、所使用笔的材料等。这对建模书写人的书写风格提出了挑战。基于局部数据特征提取^[10-12]和基于全局数据特征提取^[13-15]是离线笔记鉴别中常用的 2 类方法。为更好地对中文字体进行鉴别,本文将结合基于局部数据的特征提取与基于全局数据的特征提取为中文笔迹鉴别服务。然而现有的结合方法,即 FragNet 网络是应用在英文笔迹鉴别中,在中文笔迹鉴别应用中存在一定不足。FragNet 是一个双流的卷积神经网络,它的支流 1 输入是整张手写字体图片,负责提取整张图片的全局特征,形成图像金字塔;支流 2 的输入是原图片的切片,负责提取局部细节特征。两个支流在整个提取特征的过程中进行了多尺度的特征融合,最后采用 SoftMax 函数作为网络输出。该模型在中

① 国家自然科学基金(61976189, 61905218)资助项目。

② 男,1994 年生,硕士生;研究方向:图像处理与模式识别;E-mail: mirror_myy@163.com。

③ 通信作者,E-mail: hxwang@zjut.edu.cn。

(收稿日期:2021-12-08)

文笔迹鉴别使用中存在以下问题:(1)特征利用不充分。该结构只考虑了细节和全局特征在过程中的融合,并没有将支流1提取到的高级全局特征直接参与网络输出的决策;(2)原有的图片分割方式在中文手写字体中使用是不恰当的,导致其在中文数据集下细节特征获取的来源不够充分。因为中文和英文的书写习惯不同,所以英文数据集下采用的分割方式并不能直接用于中文数据集。

综上所述,本文基于FragNet神经网络,提出基于3条支流结构的卷积神经网络,采用了局部特征和全局特征融合的思想,1条支流是对整张图片进行全局特征提取,另2条支流以不同的方式对图片进行分割然后进行局部特征的提取。并且在提取的过程中,3种特征进行了多尺度的融合。本文改进了网络中图片的分割方式,增加了第3条网络支流,使得整个网络中的细节特征来源更加充分。并且让全局特征提取支流最终提取到的高级特征直接参与了网络最终输出的决策,使得特征融合得更加充分。实验结果表明,本文提出的策略可以有效提升网络模型最终的性能。

本文的主要贡献点有以下2点:(1)提出了一种多流卷积神经网络模型,该模型将基于局部数据的特征提取与基于全局数据的特征提取融合的方法引入到中文笔记鉴别问题中,并取得了不错的效果;(2)本文的网络结构采用2条支流对局部数据进行特征提取,可以提取出更多的细节特征,更充分地融合局部特征和全局特征,有效地提升了模型的性能。

1 相关研究

一般来说,离线笔记鉴别方法^[16-18]可以大致分为两大类:基于传统特征的方法^[19-21]和基于神经网络特征的方法^[10,13-14,22-23]。基于传统特征的方法主要采用人为定义的特征作为鉴别的依据。比如文献[19]将英文手写字体看成是一种纹理,使用Gabor和XGabor共同提取手写体样式的特征。文献[16]使用二维Gabor滤波提取特征,将笔记鉴别视为纹理识别。文献[20,21]更注重拐角特征,提取手写字体的形态信息,如笔画与墨水的宽度和曲率等。

文献[24]提出了基于混合码本与因子分析的文本独立笔迹鉴别算法,该方法采用加权的方向指数直方图法和距离变换法对每个书写模式进行双因子方差分析,并在英文数据集上取得了不错的成绩。文献[17]提出了一种结合小波变换和广义高斯模型的中文手写体文档识别方法,取得了良好的效果。使用传统算法进行特征提取的过程中,识别特征的选择或多或少会受到研究者主观意识的干扰,这将对最终的结果产生直接的影响。

采用神经网络算法可以很好地避免这一问题。当前,神经网络算法在笔迹鉴别中的应用大致可以归为2类:基于全局数据的特征提取方法^[13-15]和基于局部数据的特征提取方法^[10-12],这两者最大的区别就是在网络的输入部分。基于全局数据的特征提取方法将整张图片作为输入进行特征提取,是目前的主流方法。文献[13]提出了一种双流多任务的卷积神经网络模型提取显式信息和隐式信息,采用的是将整张图片作为网络的输入。文献[15]通过对传统VGG-16卷积神经网络模型进行改进,提出了一种CC-VGG网络模型,利用复合卷积层替换部分卷积层,实现了手写体英文笔迹的自动鉴别,提升了网络性能。文献[25]提出在对整图进行特征提取时,对输入的中文字体图片中的字体笔画进行了随机抹除的预处理方法,这既能对训练数据进行增强也提升了模型的泛化能力。基于局部数据的特征提取方法则是在原有的图片上进行分割,形成不同的切片,然后将切片送入神经网络进行局部特征提取。如文献[11]提出的基于行分割方式并结合Caf-feNet(convolutional architecture for fast feature embedding)提取局部特征的方法,然后通过取均值的编码方式将每张笔迹材料的局部特征编码为全局特征。该方法在CVL^[26]和ICDAR 2011^[27]数据集上取得了较好的成绩。文献[28]提出了将全局数据与局部数据进行特征提取与融合得到新的综合特征的FragNet网络模型,证明特征融合之后的模型具有更好的泛化能力。文献[10]提出了利用双输入网络对局部数据进行特征提取的算法,并且这2条支路以网络共享参数的方式来减少网络训练时的参数。为了可以使得不同局部数据的位置信息也参与

网络的训练和预测,文献[12]提出了GR-RNN(global-context residual recurrent neural network)网络,对不同局部数据中蕴含的作者的书写特征进行提取,进一步提升了网络在英文数据集上的识别率。

2 模型的网络结构

本文研究的是中文手写字体的离线笔记鉴别问题。本节主要从以下3个方面展开。(1)局部数据获取方式分析说明。对本文为什么要采取切片分割的方式进行以及本文在原有方式的改进进行分析说明。(2)网络结构介绍。对本文提出的网络基本构成与功能进行剖析介绍。(3)网络训练。对本文采用的实验环境与网络训练进行说明。

2.1 局部数据分割方式分析

基于局部数据的特征提取已被证明了其在笔迹鉴别中的作用。本节针对中文字体的形态结构,对局部数据的分割方式进行分析与设计,并对其理论依据进行论述。

本文将手写字体信息分为2类:字体基本结构单元和字体基本结构单元之间的连接,称为“连笔”。英文单词的基本构成单元为26个英文字母,中文字体的基本构成单元为32种基本笔画。如图1所示,图中列举了一些中英文字体及其特征信息,其中具有丰富书写人特征信息的连笔部分占据了比较大的比重。在网络特征提取过程中,这些特征有较大概率会被网络判别为“spotting region”(网络的有效感受野),即注意区域。

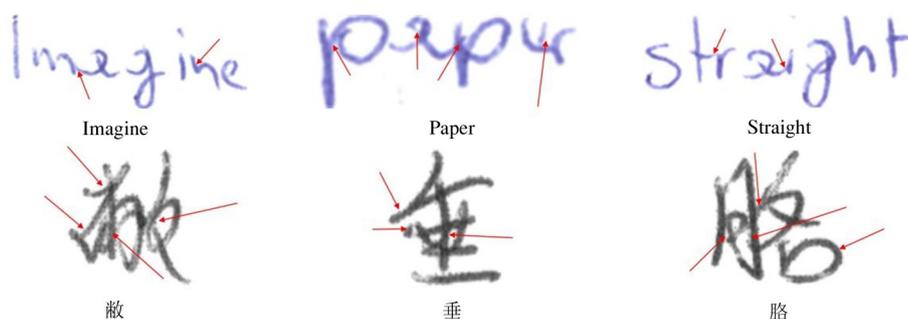


图1 字体特征信息示例(箭头指向处表示手写字体中具有丰富书写人特征信息的细节点)

文献[29]仅仅使用了书写时笔的活动坐标、落笔的位置以及最后提笔的位置这3项坐标数据序列,然后从中进行采样后送入循环神经网络(recurrent neural network, RNN)进行训练,该方法在中文公开数据集上可以获得99.46%的准确率,在英文公开数据集上可获得100%的准确率。文中通过逆向工程发现,使用少量采样笔画坐标序列就可以获得高准确率的笔画序列往往是笔画间或者单个字体间的连笔部分。这说明了连笔部分中包含了丰富的书写者的身份信息。虽然这是线上笔迹鉴别问题所采用的方法,但是,更偏向于使用字体笔画进行建模的离线笔迹鉴别问题想要达到这样高的准确度存在较大难度。因此,如何更好地引导网络关注“spotting region”是本文注重的。

采用对图片进行切片的方式送入网络可以使网络学习到的深层特征比从整张手写字体图像中学习

到的深层特征更有效。这点在提出FragNet的论文中已经得到了证明。其根本原因是网络在经过分割后局部数据块中可以更好地提取到更多的“spotting region”,得到更多的细节特征。中英文字体的展现形式不同,如图2所示,英文单词的书写顺序都是从左往右写,而中文字体的书写却没有这样的规律,从左往右、自上而下或者是两者结合的书写顺序都是存在的。原网络在英文数据中仅采用垂直分割方式,如图3(a)所示,因为英文单词从左往右的书写习惯确定了连笔只能发生在左右之间差不多高度的位置。但这在中文数据集中直接采用是不合适的,中文字体中可能同时存在自左而右、自上而下的写法,笔画间的连笔等特征信息可能发生在左右之间也有可能发生在上下之间,并且字体中其他形状也包含对作者识别有用的书写风格信息。因此,充分地结合它们蕴含的信息是可以提高网络性能的。在

FragNet 的网络框架下,本文进一步探索对输入手写字体图像中所有形状信息使用不同的局部切片进行特征提取与集成。如图 4 所示,聚合局部块会在输入图像中产生多个“spotting region”,从而使分类证据更加稳健。基于这一观点,本文提出的了采用水

平和垂直分割方式对输入图像进行分割切片,以迫使网络学习手写风格信息或在每个片段中找到定位区域。结合这些片段的结果使网络具有多个定位区域,避免了对输入图像中某些特定部分的过度拟合,从而提高了网络性能。

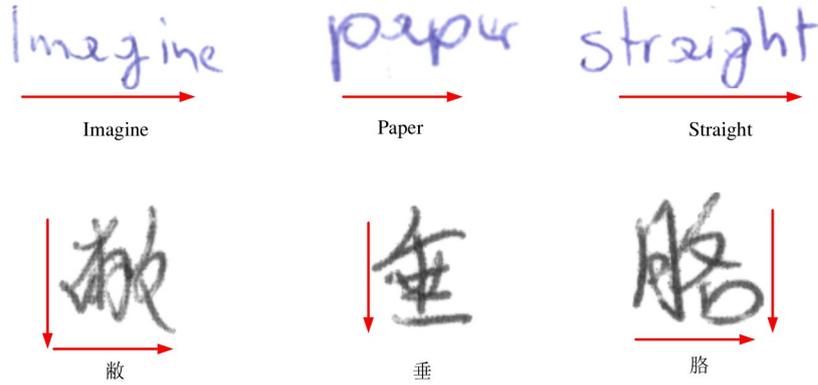
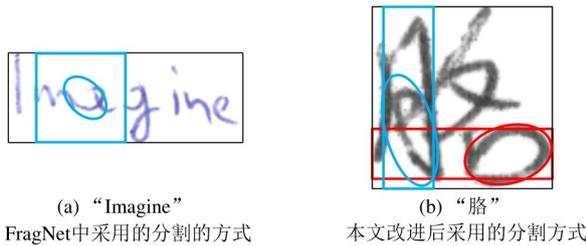


图 2 中英文字体的书写方向差异(箭头代表字体的书写方向)



(a) “Imagine” FragNet中采用的分割的方式 (b) “胳膊” 本文改进后采用的分割方式

图中圆圈部分表示字体的连笔部分

图 3 中英文字体中采用的分割方式的对比

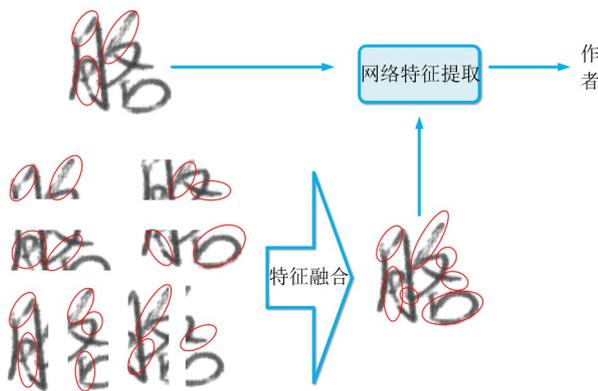


图 4 分割切片与整图输入进行特征提取的差异

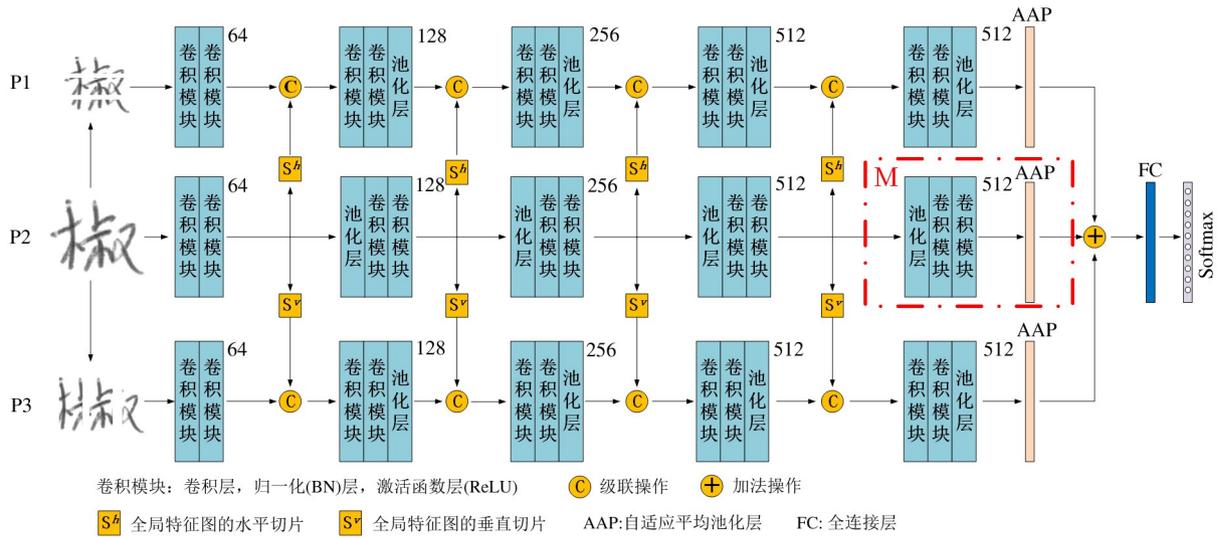
根据文中之前的分析,本文在图 3(b)垂直切片的基础上,新增滑动窗口进行水平切片,然后进行特征提取。同时采用水平切片和垂直切片的方式来增加送入网络的局部图像块的形式,迫使网络可以在不同的局部图像块上产生不同的“spotting region”,

即提取到更加充分的作者身份信息,减低网络过拟合的风险,提升网络的泛化能力。这样做的目的就是为了拓宽网络细节特征获取的来源,使得网络更好适应目标字体变化带来的变化。

2.2 网络结构分析

本文以 FragNet 作为基本结构,提出了局部数据特征提取与全局数据特征提取相结合的多流卷积神经网络结构,如图 5 所示,神经网络由 P1、P2、P3 3 条支流组成。P2 支流结构的输入是一整张手写字体图片,它负责提取整张图片的全局特征,形成图像金字塔。P1 与 P3 支流的输入是原始图片的横向与纵向切片图,负责从局部图像块中提取出更多的“spotting region”。P1、P2 和 P3 支流在整个提取特征的过程中进行了多尺度的特征融合。最终 3 条支流都通过一个自适应平均池化层生成 512 维的向量,将这 3 个向量以直接相加的形式进行融合后送入全连接层进行结果预测,最后通过 SoftMax 函数输出。

组成这 3 条支流的基本结构单位是相似的,都由池化层和 2 个卷积层组成的卷积组构成。3 条支流在第 1 次提取特征后都没有使用池化层,这是为了避免在初级特征阶段丢失过多的信息。而后 P1、P3 与 P2 支流结构在池化层作用的位置上发生了变化。因为 P1、P3 都进行特征图的级联,所以在级联



P1、P2、P3 代表它们所在位置的 3 条网络支流结构

图 5 多流卷积神经网络结构

之后需要进一步地特征提取,故将池化层放在卷积组之后。而 P2 支流没有这一步操作,所以直接进行了池化操作。P2 支流在每一阶段提取的全局特征图,通过滑动窗口分割得到水平切片 S^h 和垂直切片 S^v 。特征图切片 S^h 和 S^v 分别与 P1、P3 支流的同级特征级联。需要注意的是,因为网络中有池化层的存在,所得到的特征图的尺寸不断地以 2 的倍数方式在减小。因此,生成 S^v 与 S^h 的分割参数是不断变换的。在分割时,所使用的滑动窗口的尺寸以及采用的步长也都将以 2 的倍数方式进行减小,以保证所得到的全局特征图的切片与局部特征图具有相同的尺寸进行融合。还需要注意的是,P1 和 P3 支流结构的卷积组不会改变输入特征图的通道数,即当输入特征图的尺寸为 $H \times W \times C$ (高 \times 宽 \times 通道数),输出特征图的尺寸依旧为 $H \times W \times C$ 。而 P2 支流则会对输入特征图的通道数进行翻倍,即输入特征图的尺寸为 $H \times W \times C$,输出特征图的尺寸为 $H \times W \times 2C$ 。换句话说,P2 支流中卷积通道数的增加是通过自身增加卷积核的数目来完成的,而 P1 和 P3 支流则是通过特征图的级联来完成的。

本文网络的改进包含 2 个部分。首先,增加 P1 支流从而丰富网络提取细节特征(“spotting region”)的来源。本文采用水平分割和垂直分割 2 种图片切片方式来增加网络获取细节特征的渠道,确保 P1 和 P3 支流可以提取到足够多不同的细节特

征与 P2 中提取的全局特征进行多尺度融合,形成更具综合性的高级新特征,增强网络的泛化能力。其次,在 P2 支流中,增加 M 模块,从而让 P2 支流中提取到的全局高级特征能直接参与网络的最终决策,更加充分地利用 P2 所提取的高级特征信息。

2.3 网络训练与环境

本文最终所采用的分割方式如图 6 所示,P1 支流采用横向滑动窗口的水平分割方式,高为 64,宽为 112,移动步长为 16,自上向下滑动。它更专注于提取水平方向的连笔细节。P3 支流采用纵向滑动窗口的垂直分割方式,高为 112,宽为 64,移动步长为 16,自左向右滑动。它更专注于提取垂直方向的连笔细节。

本文实验硬件环境包含 CPU 处理器 Intel i9-10900X,显卡 NVIDIA RTX2080Ti,显存 11 GB。软件环境包含 Python3.7, Pytorch 深度学习框架。本文提出的网络模型训练时初始学习率设为 0.000 08,在第 20 轮除以 4,第 25 轮除以 3,第 30 轮设为 0.000 003,一共进行 40 轮训练。本文网络本质上是一个分类任务,故采用交叉熵作为损失函数,采用 Adam^[30]进行模型优化:

$$Loss = - \sum_{i=1}^n q_i \cdot \log(p_i) \quad (1)$$

其中 n 表示数据集中一共含有的作者数目, p 表示网络的输出结果, q 表示对应图片的标签。

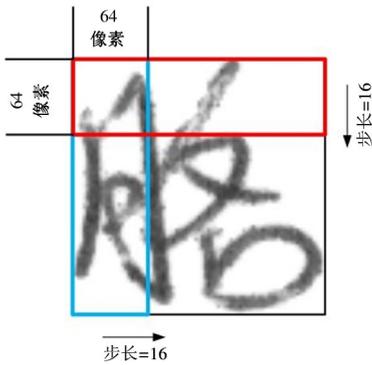


图6 本文所采用的图片分割方式

3 实验结果分析

本节将对实验的结果进行对比分析。本节内容分为3个部分:第1部分对本文所用的公开数据集进行说明;第2部分对网络进行消融实验对比,证明



图7 HWDB1.1数据集中不同作者书写的相同汉字的示例

3.2 实验结果与分析

本节对所提的网络结构进行消融实验,并结合实验数据进行分析。本节共进行了5组实验,如表1所示,具体包括 P2-M + P3 (FragNet)、P2 + P3、P1 + P2、P1 + P3 以及 P1 + P2 + P3。并分成3部分进行分析,分别为对加入 M 模块有效性的证明、对加入 P1 支流结构有效性的证明以及对全局特征在笔迹鉴别问题中的不可或缺性的验证。本文采用 top-1 (前1)准确率和 top-5 (前5)准确率这两个指标来对模型的性能进行评估。top-1 准确率是指将网络输出中概率最大的一个作为网络的最终预测结果,并且与标签是一致的,则认为模型预测正确。top-5 准确率是指网络输出中概率最大的5类别中包含有标签对应类别,则认为模型预测正确。除此之外,为了能

网络设计的有效性;第3部分将本文提出的网络模型的性能与其他网络模型的性能进行对比。

3.1 数据集说明

本文使用的数据集 HWDB1.1^[31]是中国科学院自动化研究所收集并发布的中文手写字体数据集。它包含由300位不同作者编写的1 172 907张中文字体图像。每个作者大约有3755个不同的汉字,它们以8位灰度图像的形式存储。为了节省训练的时间,本文对 HWDB1.1 数据集进行了采样,每个作者随机抽取300个词。因此,形成了一个由300位作者和90 000张图像组成的子数据集 HWDB1.1-Sub。将这90 000张图像随机分成3等份,2份用于训练,剩下的1份用于测试。如图7所示,本文随机挑选了其中的3位作者书写的5个字,很明显,它们拥有各自的书写风格。

够更细致地分析模型性能间的差异,本文还对每个模型中每一位作者的 top-1 准确率进行了统计,然后将需要对比的2个模型的数据绘制成散点图进行作者级别的准确率变化分析。

首先是对增加 M 模块有效性的证明,即本文提出原本网络中全局特征与局部特征融合的不够充分,对 P2 支流提取到的高级特征没有充分利用的问题而提出增加 M 模块的有效性的证明。根据表1中第1、2组实验结果可知,未加入 M 模块以前网络的 top-1 准确率仅为 85.06%,top-5 准确率为97.53%。当加入了 M 模块之后,也就是整个 P2 + P3 支流组成的网络,它的 top-1 准确率提升到了 87.91%,相比原来增长了 2.85%。top-5 准确率也增长了近0.4%。这充分说明了将 P2 支流提取到的高级全局特征融

合入网络参与网络最终决策是十分有效的,原来网络确实存在融合不充分,没有充分利用已提取到的高级特征的问题。此外,为进一步分析 M 模块的加入对网络性能的影响,本文对每一位作者在加入 M 模块前后的准确率进行了统计分析。如图 8 所示,图中的每个点都代表一个作者,共 300 个点。横坐标代表 $P2-M+P3$ (FragNet) 支流结果中每位作者的准确率,纵坐标代表 $P2+P3$ 支流结果中每位作者的准确率。横纵坐标的表示范围都是 $0 \sim 1$ 。从图中不难发现,大多数点都是位于图中虚对角线的上方,说明这些作者的字在加入 M 模块之后其鉴别的准确率都得到了提升。特别是在 $0.6 \sim 0.8$ 范围内,其中的各位作者的准确率相较于原来得到了较大的提升。

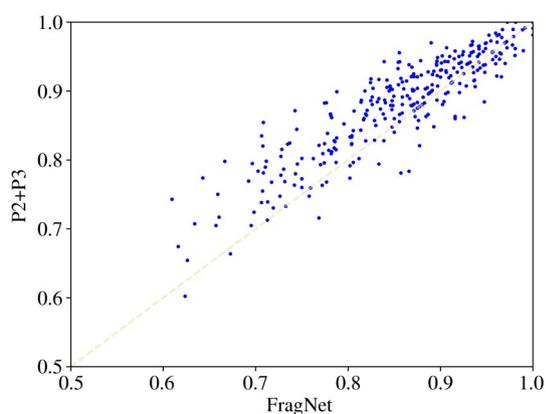


图 8 $P2+P3$ 与 FragNet 网络中每位作者 top-1 准确率情况对比

其次是对加入 $P1$ 支流结构有效性的证明,即对增加分割方式,拓宽细节特征来源有效性的证明。对于这部分的实验,本文分 2 步进行证明。第 1 步需要证明新增加的采用水平分割方式的 $P1$ 支流具有与原本存在的垂直分割方式的 $P3$ 支流具有相近的特征提取能力。根据表中的第 2、3 组实验结果可知, $P2+P3$ 支流网络 top-1 准确率与 $P1+P2$ 支流网络的 top-1 准确率仅相差 0.17% , top-5 的准确率几乎相等。这 2 组实验证明了新增加的采用水平分割方式的 $P1$ 支流具有与采用垂直方式的 $P3$ 支流相近的特征提取能力 ($P3$ 支流略高于 $P1$ 支流)。第 2 步需要证明同时采用水平分割方式和垂直分割方式进行特征提取能否增加网络获取细节特征来

源,提升网络性能的问题。根据表中第 2、3 和 5 组实验结果可知,第 2 组和第 3 组分别单独采用了垂直分割 ($P2+P3$) 和水平分割方式 ($P1+P2$),而第 5 组实验是两者的结合 ($P1+P2+P3$)。其中 $P1+P2+P3$ 网络结构的 top-1 和 top-5 准确率都是最高的,分别达到了 88.18% 和 98.07% 。虽然最终的结果比采用单个支流的结果最多仅高出了大约 0.44% (top-1) 和 0.17% (top-5)。但这依然可以说明增加采用水平分割方式的 $P1$ 支流确实拓宽了网络获取细节特征的能力,提升了网络的性能。最终结果之所以没有拉开很大的差距,可能是原本采用单个分割方式的网络已经具有了很强的特征提取能力了,增加不同分割方式的支流进入网络所能提供新的特征信息对于整个网络性能提升的帮助有限。同样,本文也进一步分析了它们在每个作者准确率上的差异,如图 9 和 10 所示。从图 9 展示的结果来看,300 位作者的准确率分布在对角线两边大致是均匀的,没有发生很明显的偏向,这与总的准确率结果是一致的。图 10 中将采用单一分割方式与采用 2 种分割方式的每位作者的准确率进行了对比,总的来说是在对角线上半部分的点多些,但是因为它们的总体的准确率差距不是很大,图中表现得并不明显。可能的原因在前面的分析中已经提过。

最后是对全局特征在笔迹鉴别问题过程中的不可替代性的验证。表 1 中的第 4 组实验没有采用中间的 $P2$ 支流结构,只有 $P1$ 和 $P3$ 支流结构。根据第 2~4 组实验结果的对比可以发现,在没有全局特

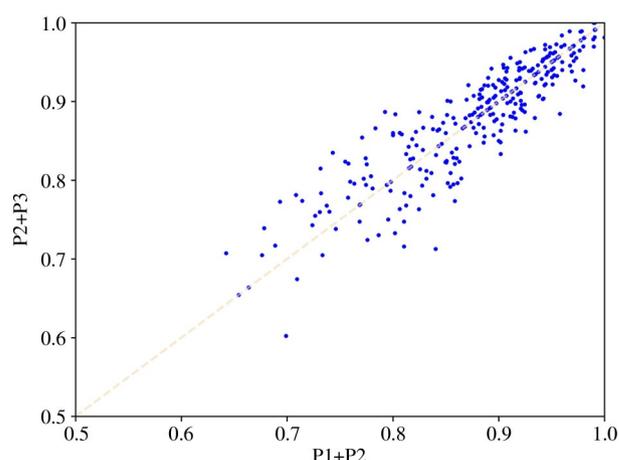


图 9 水平分割与垂直分割方法中每位作者 top-1 准确率情况对比

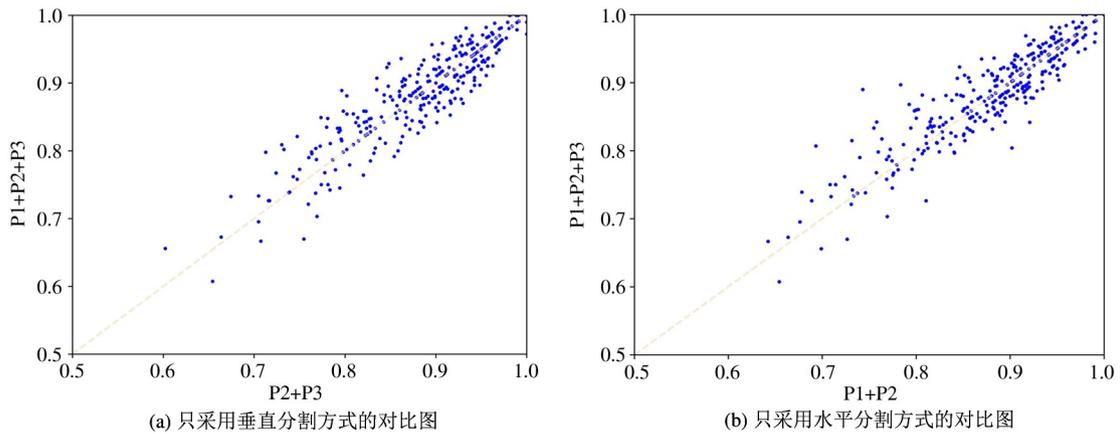


图 10 只采用单一分割方式和两种分割方式的方法中每位作者 top-1 准确率情况对比

征的参与下,仅有 P1 和 P3 支流结构组成的网络性能明显低于任何一个有全局特征参与的网络结构。这也进一步验证了全局特征在笔迹鉴别问题中是不可或缺的。而在之前的分析中同样也证明了细节特征在笔迹鉴别中的重要性。因此,可以得出结论:在笔迹鉴别问题中,手写字体的细节特征和全局特征都是需要的,单独依靠某一个特征无法取得很好的效果,提取 2 种特征进行融合处理才能获得更好的

效果。这 3 个网络模型的作者准确率变动情况如图 11 所示。可以直观地发现,有全局特征参与,无论是对于水平分割还是垂直分割方式来说,都可以很好地提升作者被识别的准确率。图 11 中 2 幅图都存在明显的点向下偏移的情况。

3.3 与其他算法的性能对比

本节在 HWDB1.1 数据集上与其他网络模型方法进行性能比较。研究中国文字体笔迹鉴别的论文比

表 1 网络模型结构消融实验结果

组号	模型结构	准确率 (top-1)	准确率 (top-5)
1	P2-M + P3 (FragNet)	85.06%	97.53%
2	P2 + P3	87.91%	97.90%
3	P1 + P2	87.74%	98.03%
4	P1 + P3	85.63%	97.52%
5	P1 + P2 + P3 (本文方法)	88.18%	98.07%

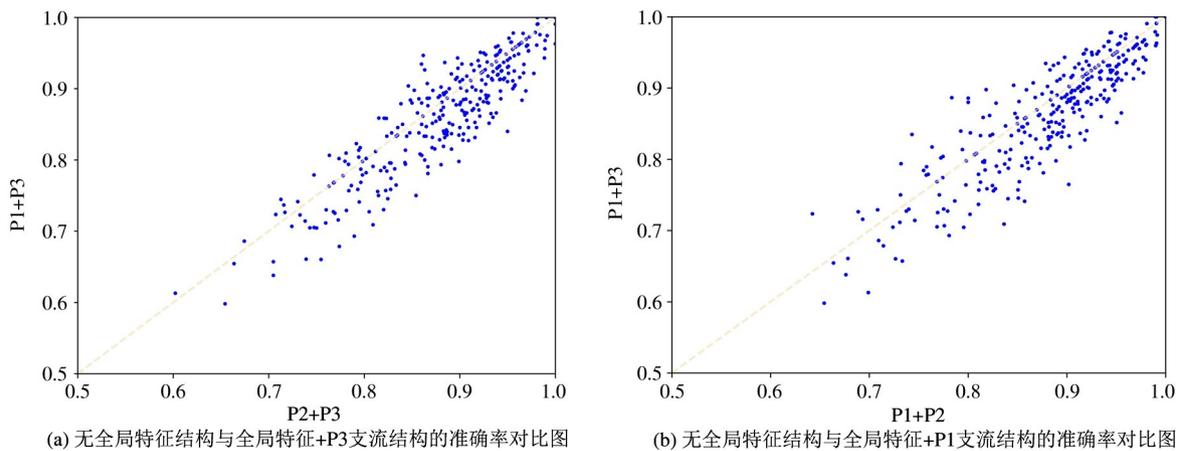


图 11 有无全局特征参与的方法中每位作者 top-1 准确率情况对比

研究英文字体的要少。本文关注的是基于单一手写字体的离线笔记鉴别问题,相应的论文更少。由于使用相应数据集进行比较的论文很少,因此本文进行了2方面的对比。

首先使用了研究单一手写字体笔迹鉴别的网络来进行比较。本文使用 ResNet18^[32]、研究英文笔迹鉴别的 FragNet^[28] 和 Deep Adaptive^[13] 进行比较。对于 ResNet18、FragNet 和 Deep Adaptive 网络,为了消除不同图片大小带来的影响,将输入图像的大小设置为 112×112 ,并使用 HWDB1.1-Sub 进行训练和测试。实验使用60 000张的样本进行训练,30 000 样本进行测试。采用 top-1 准确率和 top-5 准确率这2个指标来对模型的性能进行评估。其结果如表2中前4行所示。在 top-1 和 top-5 的这2个准确率指标上,本文提出的方法都是最佳的。

其次,本文与使用相同了数据集的 Deep Writer^[7] 算法进行对比。Deep Writer 的结果选自文献[7],由于它使用 HWDB1.1 全部的数据集进行划分训练,本文也按照它使用的划分方法重新对数据集进行了划分与训练。使用了大约 780 000 张样本进行

训练,大约 260 000 样本进行测试。实验结果如表2最后2行所示,本文提出的方法在 top-1 准确率上比 Deep Writer 高了 4.32%。

除此之外,本文还用了 torchstat 工具对网络实际的计算复杂度和参数量进行了分析比较。从整体上来看,虽然无论是与目标字体为英文字体的网络还是目标文字为中文字体的 Deep Writer 网络相比,本文提出的模型性能都是最佳的,但本文提出的网络在参数量、计算训练的时间等3个指标上表现较弱。如何在不特别影响性能的前提下减少网络中的参数和计算量是今后可以进一步研究的方向。

本节也对表2中各个网络模型中每位作者的准确率情况做了统计,如图12所示。需要说明的是,由于本文直接采用了 Deep Writer 论文中的数据,而不是进行重新训练得到的数据,因此无法对其结果中的作者准确率进行统计。对于 FragNet 来说,较低准确率的作者在本文方法中提升的幅度很大,随着准确率的提高这一效果变得越来越弱。而对于其他2个网络来说没有这一现象,主要原因就是它们整体的识别率相较于本文方法来说差距大,因而使

表2 本文方法与其他算法的性能对比

模型结构	准确率(top-1)	准确率(top-5)	模型计算复杂度 /Floats	模型参数量 /MB	模型训练时间 /h	训练数据量 / $\times 10^4$ 条
FragNet ^[28]	85.06%	97.53%	9.98 G	18.15	9.24	6
ResNet18 ^[32]	74.52%	93.72%	486.86 M	11.34	1.55	6
Deep Adaptive ^[13]	66.78%	89.22%	5.81 G	32.74	6.33	6
本文方法	88.18%	98.07%	17.63 G	36.19	17.81	6
Deep Writer ^[10]	93.45%	/	1.38 G	13.81	/	78
本文方法	97.77%	99.80%	17.63 G	36.19	105.00	78

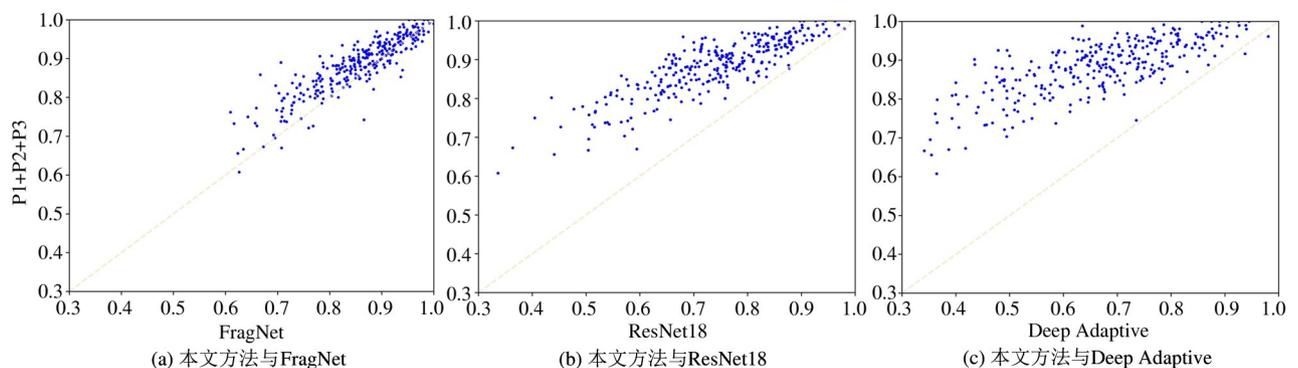


图12 本文提出方法与其他方法中每位作者 top-1 准确率情况对比

得几乎所有的点都位于对角线上方且偏离的幅度很大。可见本文提出的方法对于特征提取的能力都优于其他3种方法。

4 结论

针对现有局部特征与全局特征融合网络在特征使用不充分、提取到的全局高级特征没有直接参与网络最终决策的问题,本文提出了在原来的网络结构上增加卷积模块与全连接层来保证提取到的高级特征能得到充分利用;针对原网络中分割方式不适用中文字体,导致在中文数据集下细节特征获取的来源不够充分的问题,本文以增加水平分割的方式来拓宽细节特征来源。通过消融实验数据的对比分析,证明了本文方法的有效性,促使网络整体的性能得到了明显提升。同时,本文还验证了细节特征和全局特征对于笔迹鉴别问题具有同等的重要性,两者的结合能获得更好的识别结果。最后,本文提出的模型在与其他模型的对比中也取得了最好的成绩。

本文提出的方法虽然采用了水平分割和垂直分割的方式对手写字体图片进行分割,但对于切片序列之间的空间关系建模、增强切片局部特征的识别能力没有涉及。针对该问题,未来考虑在切片序列空间关系建模算法方面展开研究工作。

参考文献

- [1] BRINK A A, SMIT J, BULACU M L, et al. Writer identification using directional ink-trace width measurements [J]. *Pattern Recognition*, 2012,45(1):162-171.
- [2] ASI A, ABDALHALEEM A, FECKER D, et al. On writer identification for Arabic historical manuscripts[J]. *International Journal on Document Analysis and Recognition*, 2017,20(3):1-15.
- [3] NEWELL A J, GRIFFIN L D. Writer identification using oriented basic image features and the Delta encoding[J]. *Pattern Recognition*, 2014,47(6):2255-2265.
- [4] HE S, SCHOMAKER L. Writer identification using curvature-free features[J]. *Pattern Recognition*, 2017,63:451-464.
- [5] VENUGOPAL V, SUNDARAM S. Modified sparse representation classification framework for online writer identification[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021,51(1):314-325.
- [6] VENUGOPAL V, SUNDARAM S. Online writer identification with sparse coding-based descriptors [J]. *IEEE Transactions on Information Forensics and Security*, 2018,13(10):2538-2552.
- [7] VENUGOPAL V, SUNDARAM S. Modified sparse representation classification framework for online writer identification[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018,51,314-325.
- [8] 柴田若梅. 离线笔迹鉴定中动态特征的分析与应用 [J]. *广东公安科技*, 2020,28(4):21-23.
- [9] 洪晓倩,刘福林. 快写签名字及慢写样本笔迹的检验 [J]. *法制博览*, 2021(28):88-89.
- [10] XING L, QIAO Y. Deepwriter: a multi-stream deep CNN for text-independent writer identification[C]//2016 15th International Conference on Frontiers in Handwriting Recognition. Shenzhen: IEEE, 2016: 584-589.
- [11] FIEL S, SABLATNIG R. Writer identification and retrieval using a convolutional neural network[C]// Proceedings of the 16th International Conference on Computer Analysis of Images and Patterns. Berlin, Heidelberg: Springer-Verlag, 2015:26-37.
- [12] HE S, SCHOMAKER L. GR-RNN: global-context residual recurrent neural networks for writer identification[J]. *Pattern Recognition*, 2021,117:107975.
- [13] HE S, SCHOMAKER L. Deep adaptive learning for writer identification based on single handwritten word images [J]. *Pattern Recognition*, 2019,88:64-74.
- [14] WEI P, LI H, HU P. Inverse discriminative networks for handwritten signature verification[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019:5757-5765.
- [15] 何凯,马红悦,冯旭,等. 基于改进 VGG-16 模型的英文笔迹鉴别方法[J]. *天津大学学报(自然科学与工程技术版)*, 2020,53(9):984-990.
- [16] 陈使明,王以松. 一种鲁棒的离线笔迹鉴别方法[J]. *自动化学报*, 2020,46(1):108-116.
- [17] HE Z, YOU X, TANG Y Y. Writer identification using global wavelet-based features [J]. *Neurocomputing*, 2008,71(10-12):1832-1841.
- [18] HU Y, YANG W, CHEN Y. Bag of features approach for offline text-independent Chinese writer identification[C]//2014 IEEE International Conference on Image Processing. Paris: IEEE, 2014:2609-2613.
- [19] HELLI B, MOGHADDAM M E. A text-independent Persian writer identification based on feature relation graph (FRG)[J]. *Pattern Recognition*, 2010,43(6):2199-2209.

- [20] BULACU M, SCHOMAKER L. Text-independent writer identification and verification using textural and allographic features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007,29(4):701-717.
- [21] BRINK A A, SMIT J, BULACU M L, et al. Writer identification using directional ink-trace width measurements [J]. Pattern Recognition, 2012,45(1):162-171.
- [22] CHEN S, WANG Y, LIN C T, et al. Semi-supervised feature learning for improving writer identification [J]. Information Sciences, 2019,482:156-170.
- [23] REHMAN A, NAZ S, RAZZAK M I, et al. Automatic visual features for writer identification: a deep learning approach [J]. IEEE Access, 2019, 7:17149-17157.
- [24] 阿依夏木·力提甫, 鄢煜尘, 肖进胜, 等. 基于混合码本与因子分析的文本独立笔迹鉴别 [J]. 自动化学报, 2021,47(9):2276-2284.
- [25] YANG W, JIN L, LIU M. Chinese character-level writer identification using path signature feature, DropStroke and deep CNN [C] // 2015 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015:546-550.
- [26] KLEBER F, FIEL S, DIEM M, et al. CVL-database: an off-line database for writer retrieval, writer identification and word spotting [C] // 2013 12th International Conference on Document Analysis and Recognition. Washington: IEEE, 2013:560-564.
- [27] LOULLOUDIS G, STAMATOPOULOS N, GATOS B. IC-DAR 2011 writer identification contest [C] // Proceedings of the 11th International Conference on Document Analysis and Recognition. Beijing: IEEE, 2011:1475-1479.
- [28] HE S, SCHOMAKER L. FragNet: writer identification using deep fragment networks [J]. IEEE Transactions on Information Forensics and Security, 2020,15:3013-3022.
- [29] ZHANG X, XIE G, LIU C, et al. End-to-end online writer identification with recurrent neural network [J]. IEEE Transactions on Human-Machine Systems, 2017,47(2):285-292.
- [30] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. (2017-1-30) [2022-03-29]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [31] LIU C, YIN F, WANG D, et al. CASIA online and offline Chinese handwriting databases [C] // 2011 International Conference on Document Analysis and Recognition. Beijing: IEEE, 2011:37-41.
- [32] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:770-778.

Multi-stream convolutional neural network for offline Chinese writer identification

MAO Yingyu **, ZHANG Yilong *, WANG Haixia *

(* College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023)

(** College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023)

Abstract

Aiming at the extraction and fusion of local and global features in Chinese handwriting identification, this paper proposes a multi-stream convolutional neural network based offline Chinese writer identification method. The network has three tributary structures, one of which extracts global features from the entire image while the other two perform feature extraction on image segments cut horizontally and vertically from the original image, respectively. Multi-scale fusion is then performed on the extracted features. This paper uses top-1 and top-5 accuracy rates for performance evaluation and carries out the comparative analysis based on accuracy rate of each author. The experimental results show that the proposed method can not only integrate local and global features, but also broaden the source of the network to obtain detailed features, consequently improving the network performance.

Key words: handwriting identification, feature extraction, pattern recognition