

一种鲁棒的离线笔迹鉴别方法

陈使明¹ 王以松^{1,2}

摘 要 离线笔迹鉴别在司法鉴定与历史文档分析中有重要作用。当前的主要离线笔迹鉴别都是基于局部特征提取的方法, 其在笔迹检索中严重依赖于数据增强和全局编码, 在笔迹识别中需要较多的笔迹信息。针对这一问题, 本文提出一种基于统计的文档行分割与深度卷积神经网络相结合的离线笔迹鉴别方法 (DLS-CNN)。首先, 使用基于统计的文档行分割方法将笔迹材料分割成小的像素块; 然后, 用优化后的残差神经网络作为识别模型; 最后, 对局部特征使用取均值法进行编码。在 ICDAR2013 和 CVL 这两个标准数据集上的实验结果表明, 该方法能有效获得鲁棒的局部特征, 从而仅需要少量的笔迹信息就能取得较高的识别率, 而且不需依赖于数据增强和全局编码就能取得较好的检索效果。实验代码地址: <https://github.com/shiming-chen/DLS-CNN>。

关键词 笔迹鉴别, 笔迹检索, 文档行分割, 卷积神经网络, 特征提取

引用格式 陈使明, 王以松. 一种鲁棒的离线笔迹鉴别方法. 自动化学报, 2020, 46(1): 108–116

DOI 10.16383/j.aas.2018.c180441

A Robust Off-line Writer Identification Method

CHEN Shi-Ming¹ WANG Yi-Song^{1,2}

Abstract Off-line writer identification plays an important role in forensics and historical document analysis. The current well-known off-line writer identification approaches are based on local feature extraction. They rely heavily on data augmentation and global encoding for writer retrieval, and need a great number of written contents for writer recognition. This paper proposes a new off-line writer identification method, called DLS-CNN, which combines document line segmentation in terms of statistic and deep convolutional neural network. More precisely, handwriting documents are segmented into patches using document line segmentation at first. Secondly, an improved residual neural network serves as the identification model. Finally, the mean value of all local features vectors are used as final global features for writer identification. Experimental results on ICADAR2013 and CVL benchmark datasets show that, due to the extracted robust local features, DLS-CNN achieves higher identification rate with fewer written contents, and better retrieval result without data augmentation and global encoding. All the experiment codes are available at <https://github.com/shiming-chen/DLS-CNN>.

Key words Writer identification, writer retrieval, document line segmentation, convolutional neural network (CNN), features extraction

Citation Chen Shi-Ming, Wang Yi-Song. A robust off-line writer identification method. *Acta Automatica Sinica*, 2020, 46(1): 108–116

笔迹鉴别 (Writer identification) 是通过手写文字信息来鉴别书写人身份的一种文件检验技术。它作为机器视觉与模式识别领域的研究热点之一, 与人脸识别、语音识别、指纹识别等技术一样属于身

份识别的研究范畴, 在最近几十年受到广泛关注。它在司法鉴定与历史文档分析^[1–2] 方面发挥着重要作用。其中, ICDAR 与 ICFHR 会议以及他们组织的比赛一直在为这个研究方向的发展起着突出的推进作用^[3–5]。笔迹鉴别任务可分为笔迹识别与笔迹检索两个任务。笔迹识别是通过对比存储在数据库中的笔迹文档 (已知书写者) 来确定待识别文档属于哪一个书写者所写, 与说话人识别, 特定人脸识别等任务一致; 笔迹检索是通过度量笔迹材料的全局特征向量的相似度检索出查询库中与待查询的笔迹材料最相似的笔迹材料。依据笔迹材料的获取形式可将笔迹鉴别分为两种, 分别为在线笔迹鉴别与离线笔迹鉴别。在线笔迹鉴别可通过在线采集书写的速度、角度、书写者所使用的力量以及写字的笔画顺序等丰富信息进行笔迹鉴别。离线笔迹鉴别材料使用传

收稿日期 2018-06-21 录用日期 2018-10-11
Manuscript received June 21, 2018; accepted October 11, 2018
国家自然科学基金 (61370161, 61562009, 61976065), 贵州省优秀青年科技人才培养对象基金 (2015(01)) 资助
Supported by National Natural Science Foundation of China (61370161, 61562009, 61976065), Outstanding Young Talent Training Fund of Guizhou Province (2015(01))
本文责任编辑 金连文
Recommended by Associate Editor JIN Lian-Wen
1. 贵州大学计算机科学与技术学院 贵阳 550025 2. 贵州省智能医学影像分析与精准诊断重点实验室 贵阳 550025
1. School of Computer Science and Technology, Guizhou University, Guiyang 550025 2. Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, Guiyang 550025

统的书写工具在纸张上书写笔迹信息, 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与易获取性, 故其实用性较高, 然而离线笔迹材料包含的信息不如在线笔迹材料丰富, 故其鉴别难度更大。本文研究的对象即为离线笔迹鉴别。

离线笔迹鉴别受多种因素的影响, 例如, 1) 随着书写人年龄的增长, 其书写方式可能会有一定的变化; 2) 同一个书写人的材料可能来源于不同笔的使用; 3) 书写者的现实物理环境的变化会影响书写者的书写习惯; 4) 每个书写者的笔迹材料数量与笔迹信息都非常有限。这对于机器学习方法建模离线笔迹鉴别任务是很大的挑战, 特别是对于深度卷积神经网络 (Convolutional neural network, CNN) 方法, 如何利用这有限的训练数据训练一个可靠的神经网络模型是巨大难题。对于笔迹鉴别存在的这些挑战, 早期的研究者通常构造有效的手工特征方法对笔迹方向的角度、笔画的宽度等手写信息进行建模, 如用 Gabor 滤波器^[6] 与局部二值模式 (Local binary patterns, LBP)^[7] 提取笔迹材料的纹理特征和基于文字的柱状图特征; 近期的研究者充分利用卷积神经网络 (CNN) 的优势来自动提取特征^[8-15], 不断地提高在一些标准数据集上的评测结果, 使得笔迹鉴别提升到新的台阶。

当前的离线笔迹鉴别方法可归为两类: 基于局部特征提取方法与基于全局特征提取方法。基于局部特征提取方法^[8-12, 16-17] 是对笔迹材料的局部结构、梯度、轮廓、几何特征等进行特征描述, 并通过编码方式将局部特征编码为全局特征。基于全局特征提取方法^[7, 18-22] 直接对原始的笔迹材料进行全局的特征提取, 然后将提取的全局特征作为鉴别的依据。当然, 也有研究者将局部特征与全局特征组合起来构建更为强大的特征以提高鉴别效果^[23]。由于局部特征方法能有效地解决小样本问题, 提高算法对笔迹材料的局部信息的特征提取, 所以近几年来, 基于局部特征提取方法得到更大地重视, 而本文的方法亦是基于局部特征提取方法。局部特征提取方法的关键步骤有如下几步: 1) 将原始笔迹材料分成小的像素块; 2) 对这些小的像素块进行特征提取; 3) 通过全局编码方法将局部特征编码为全局特征向量。其中前两个过程是最为根本的两个步骤, 它体现了整体模型的泛化能力。Christlein 等^[24] 探讨了编码方式对笔迹鉴别的影响, 同时强调了前两个基本步骤的重要意义。

当前大部分的局部特征提取方法都是依赖于尺度不变特征变换 (Scale-invariant feature transform, SIFT)^[11-12, 24]、加速鲁棒性特征 (Speed up robust features, SURF) 等方法来提取局部像素块, 但是此方法需要组合编码能力强的编码方式 (Fisher

vectors, Vector of locally aggregated descriptors (VLAD)^[12], Gaussian mixture model (GMM)^[11] 等) 才能取得较好的鉴别结果。而另外的一种是基于文档行分割的方式将原始笔迹材料分割成小的像素块^[8, 10]。Fiel 等^[8] 首次提出了基于行分割的方式并组合 caffeNet 提取局部特征, 随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, 在 ICDAR2011^[25] 和 CVL^[26] 标准数据集上取得当时最好的结果。他使用 Deim 等^[27] 提出的行分割方法对笔迹材料进行分割, 用归一化方法纠正歪斜的行文本, 并进行数据增强。由于此模型使用的行切割方法对笔迹粘连或笔迹位于上下两行中间位置的笔迹材料切割的不准确性, 以及 8 层架构的 caffeNet 的特征学习能力不强等问题, 在 ICADRAR2013^[3] 这一多语言的高挑战性数据集上的表现并不理想。Xing 等^[10] 提出基于新的像素扫描策略与数据增强并组合多流并行 CNN 的方法, 在 IAM^[28] 和 HWDB^[29] 两个标准数据集上用少量的笔迹信息分别取得 99.01% 和 97.03% 的识别率^[10]。

虽然当前基于局部特征提取的离线笔迹鉴别方法取得很好的成果, 但它们在笔迹检索任务中还过度依赖于数据增强和编码方式, 在笔迹识别任务中需要较多的笔迹信息方可取得较高的识别率。主要由以下两个原因造成: 1) 当前基于局部特征提取的离线笔迹鉴别方法的数据预处理提取的小像素块不准确, 使得很多完整的笔迹信息被错误分割; 2) 当前基于局部特征提取的离线笔迹鉴别方法的识别模型的表征能力不足, 使得学习的特征还不能充分适应笔迹识别任务。这两个因素直接导致了模型提取的局部特征的鲁棒性和泛化能力差。

为解决以上问题, 并且鉴于图像分割方法准确性^[27, 30-32] 和 CNN 在笔迹鉴别领域的成功应用, 我们提出了一种基于统计的文档行分割与深度卷积神经网络方法 (Document line segmentation-convolutional neural network, DLS-CNN)。利用基于统计的文档行分割方法分割的准确性以及深度神经网络的强学习能力, 学习具有强鲁棒性的局部特征, 使得模型在笔迹检索任务中无需依赖于数据增强并通过取均值的编码方式就能够取得不错的鉴别效果, 在笔迹识别中仅需要少量的笔迹信息即可对笔迹进行较好地识别。在 ICDAR2013^[3] 与 CVL^[26] 两个标准数据集上实验结果表明: 1) DLS-CNN 相较于其他相似方法提取的局部特征具有更强的鲁棒性; 2) DLS-CNN 只需要极少的笔迹信息就能对笔迹进行较好的识别, 模型具有较强的泛化能力。

本文第 1 节对我们提出的方法进行详细的描述; 第 2 节给出在两个标准数据集上的实验结果与分析; 最后, 第 3 节对本文进行总结与展望。

1 基于文档行分割与卷积神经网络的局部特征提取方法 (DLS-CNN)

DLS-CNN (Document line segmentation-convolutional neural network) 通过基于统计的文档行分割与深度卷积神经网络的方法提取更具有鲁棒性的局部特征, 以提高笔迹鉴别模型的鉴别能力和泛化能力, 模型的整体结构如图 1 所示, 具体步骤如下: 1) 用基于统计的文档行分割方法将原始笔迹材料进行文档行分割; 2) 通过滑动窗口法将所有的行无重叠地切割成所需要尺度的小像素块; 3) 用修改和优化过的残差神经网络 (ResNet-50) 对这些准备好的小像素块进行训练; 4) 用训练好的模型对测试数据进行局部特征提取; 5) 将对应笔迹材料的所有局部特征取均值并计算余弦相似度作为后续检索评估的依据. 其中, 笔迹的识别只需要前 3 个步骤, 笔迹检索为 DLS-CNN 的整个过程.

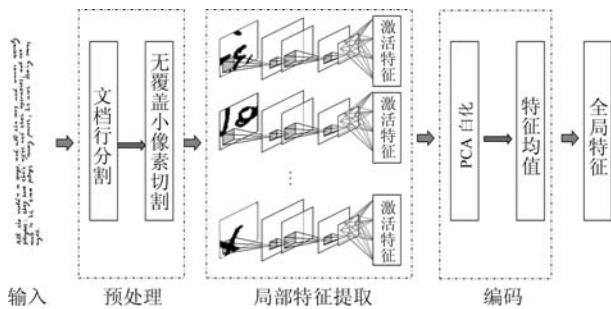


图 1 DLS-CNN 框架图

Fig. 1 The framework of DLS-CNN

1.1 基于统计的文档行分割方法

基于统计的文档行分割方法先由 Arivazhagan 等^[30] 提出解决文档行分割的问题. 此方法最大的优点就是对于行倾斜的文档亦能准确地将文档按行进行切割. 但是此方法有两个缺点: 第一, 当文档的笔迹分布于纸张的右侧时, 由于此算法找不到候选分割行而无法实现行分割; 第二, 当文档的上下行有较多的笔迹粘连或笔迹位于上下两行中间位置时, 由于笔迹成分确定算法的时间复杂度高, 使得其需要较长的时间来确定笔迹成分的归属. 为此, 我们对此方法做进一步的改进, 在保证算法准确进行行分割的同时拟解决以上两个问题. 基于统计的文档行分割方法包括如下几个过程. 1) 通过 OTSU 阈值法将图像二值化并获取文档笔迹轮廓投影映射; 2) 依据一张文档垂直方向投影映射的柱状图和自适应阈值的左右相邻柱状图低谷连接获取候选行; 3) 用二元混合高斯密度法将文档上下行成分进行确定; 4) 最后获取笔迹材料的行. 此算法的核心部分为步骤 2) 和 3). 步骤 2) 的关键是设计有效的函数寻找投影映射的柱状图低谷与高峰, 本文将垂直方向的投影

映射的柱状图分为 20 块. 步骤 3) 使用二元混合高斯模型对文档上下行有粘连笔迹或笔迹位于上下两行中间的成分进行确定, 这有助于将文档进行更加准确的文档行分割. 以下将详细描述如何通过二元混合高斯密度法对文档行成分进行建模. 我们使用前景像素的 x, y 坐标获取行成分的充分统计量 $\mu_A, \mu_B, \Sigma_A, \Sigma_B$ (分别表示 A, B 行的像素成分均值向量与协方差矩阵) 进行二元混合高斯密度建模. 计算 A 行的第 $N+1$ 个像素时, A 行像素成分的均值向量如式 (1) 所示:

$$\mu_{A(N+1)} = \frac{N-1}{N} \cdot \mu_{A(N)} + \frac{1}{N} \cdot p_{N+1} \quad (1)$$

其中, N 为当前已经计算过的成分数量, $p_{(N+1)} = \{x_{N+1}, y_{N+1}\}$ 为第 $N+1$ 个像素. 计算 A 行的第 $N+1$ 个像素时, A 行像素成分的协方差矩阵如式 (2) 所示:

$$\Sigma_{A(N+1)} = \frac{N-1}{N} \cdot \Sigma_{A(N)} + \frac{1}{N} \cdot (p_{N+1} - \mu_{A(N+1)})(p_{N+1} - \mu_{A(N+1)})^T \quad (2)$$

因此, A 行的第 i 个像素成分的二元高斯密度如式 (3) 所示:

$$P(p_i | \mu_A, \Sigma_A) = |2\pi\Sigma_A|^{\frac{1}{2}} (p_i - \mu_A) \Sigma_A^{-1} (p_i - \mu_A)^T \quad (3)$$

最后, A 行的像素成分概率如式 (4) 所示:

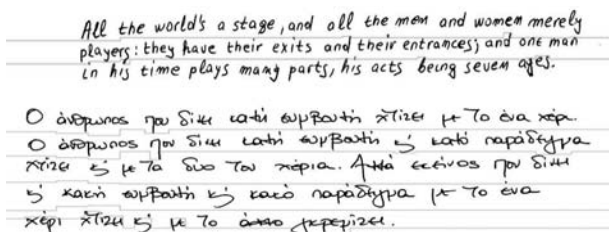
$$P(C | \mu_A, \Sigma_A) = P(p_1 | \mu_A, \Sigma_A) \cdot P(p_2 | \mu_A, \Sigma_A, p_1) \cdots P(p_N | \mu_A, \Sigma_A, p_1, p_2, \dots, p_{N-1}) \quad (4)$$

同理, B 行的像素成分概率与 A 行的计算方式相似. 最终我们依据上下两行的像素成分概率来确定像素属于哪一行的成分. ICDAR2013 的 029.1 和 002.4 样本的文档行分割结果如图 2 所示, 图 2(a) 为模型最终确定的两个样本文档行分割的结果, 图 2(b) 为样本 029.1 被分割后形成的行笔迹材料.

1.2 基于滑动窗口法的像素切割

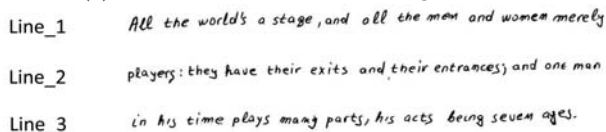
在完成笔迹材料的行切割之后, 将所有的行笔迹归一化, 然后做进一步的切割. 使用滑动窗口法, 依据所需要的尺度 (图像宽度像素值) 进行无重叠地切割. 由于尺度为 64 像素与 256 像素提取的像素块保留更完整的“字母”级与“单词”级笔迹, 所以实验中分别用到的尺度为 64 像素尺度与 256 像素尺度. 对于 64 尺度切割的像素块, 进一步用空白像素填充法将其扩充为 256×256 大小的像素块, 而对于 256 尺度切割的像素块, 则无需做进一步处理, 因为后续的特征提取网络的输入尺度为 224×224 .

最后再剔除对模型特征提取产生噪声的空白像素块和含笔迹信息较少的像素块(如只有一个点). 最终处理好的像素块如图 3(b) 和图 3(c) 所示, 图 3(b) 为 64 尺度切割的像素块, 图 3(c) 为 256 尺度切割的像素块. 对比其他用于笔迹鉴别的方法^[11] 提取的小像素块, 当上下行笔迹粘连或笔迹位于上下两行中间位置时(如笔迹中含有字母 *g, y, f* 等字母), 传统方法则不能较好的分割, 使得很多完整笔迹成分被错误分割, 如图 3(a) 所示.



(a) 文档行分割结果

(b) The result of document line segmentation



(b) 分割后的行笔迹材料

(c) The line handwritten content after segmentation

图 2 文档行分割样例

Fig. 2 The example of document line segmentation



(a) 传统方法提取笔迹样本

(b) The patches extracted with conventional method



(b) 64 尺度样本

(c) Samples of 64 patch size



(c) 256 尺度样本

(d) Samples of 256 patch size

图 3 分割好的像素块

Fig. 3 The segmented patches

1.3 深度残差神经网络 (ResNet)

深度残差神经网络 (ResNet) 由 He 等^[33] 提出, 他们将网络引入残差学习单元保护了输入信息的完整性, 使得随着网络的加深模型不会出现退化的现象, 而是进一步提升学习能力. 他们在 ILSVRC 和 COCO2015 比赛中的目标定位、分割、识别、探测任务中均取得第一名的优异成绩. 随后, 他们进一

步探索深度残差神经网络的学习能力, 并通过实验表明残差神经网络是当前特征学习能力最强的网络模型^[34]. Christlein 等^[24] 也通过实验表明残差神经网络在笔迹鉴别任务中比其他网络具有更强的表征能力. 为此, 在此模型中采用 ResNet 学习局部特征. 为了平衡时间复杂度、资源消耗与学习效率, 我们使用 ResNet-50 这个模型框架. 由于笔迹识别任务的特殊性, 将 ResNet-50 做了相应的优化和改进, 其结构组成如表 1 所示. 为了获取更具全局性与鲁棒性的局部特征, 我们将原始的平均池化层改为全局池化层. 此外, 由于训练集相对较小与网络较深的原因, 我们在全连接层后面添加了 relu 激活特征层与 dropout 层, 并将 dropout 率设置为 0.5, 以增强模型的正则化, 防止过拟合. 在笔迹检索任务中, 学习率初始设置为 0.1, 并在训练 30 步后将其降为 0.01; 在笔迹识别任务中, 学习率初始设置为 0.1, 训练 20 步之后将其缩小为 0.02, 在训练 30 步之后, 进一步将其缩小为 0.01. 模型的输入图像大小为 224×224 , 梯度下降的动量因子设置为 0.9, softmax 层的大小设置应根据任务(笔迹识别、笔迹检索)与数据集而定. 我们在 ICDAR2013 数据集上做笔迹检索任务时, 将其大小设置为 100; 而在 CVL 数据集上做笔迹识别任务时, 我们将其大小设置为 310.

表 1 ResNet-50 结构

Table 1 The structure of ResNet-50

Layer name	Layers	Output size
Conv1	$7 \times 7, 64$, Stride 2	112×112
Conv2-x1	3×3 Max pool, Stride 2	56×56
Conv2-x2	$1 \times 1, 64$	$\times 3$ 56×56
	$3 \times 3, 64$	
	$1 \times 1, 256$	
Conv3-x	$1 \times 1, 128$	$\times 4$ 28×28
	$3 \times 3, 128$	
	$1 \times 1, 512$	
Conv4-x	$1 \times 1, 256$	$\times 6$ 14×14
	$3 \times 3, 256$	
	$1 \times 1, 1024$	
Conv5-x	$1 \times 1, 512$	$\times 3$ 7×7
	$3 \times 3, 512$	
	$1 \times 1, 2048$	
Global average pool		1×1
Fc, Relu, Dropout, Softmax		1×1

1.4 局部特征提取

对于笔迹检索任务, 需将测试集中的小像素块进行局部特征提取. 本文分别对全局池化层与全连接层的特征进行了不同特征数 (128, 512, 1024, 2048) 的提取, 以寻找模型最佳局部特征. 在特征提

取过程中, 残差神经网络的最后一层可以舍弃. 随后, 将提取的局部特征进行 PCA 白化处理, 以降低特征维度和降解特征间的相互关系. 这个白化操作被证明在笔迹与图像检索中具有积极作用^[35-36].

1.5 特征编码与相似度计算

特征编码作为基于局部特征提取的笔迹检索中必不可少的一步, 对笔迹检索的提升具有重大意义. Christlein 等^[24]通过实验表明编码能有效促进笔迹检索效果, 并且编码方式有优劣之分. 当前应用于笔迹鉴别的编码方式有很多, 如取均值法^[8]、Fisher Vectors 编码法^[37]、GMM 编码法^[11, 38]、VLAD 编码法^[12, 24]等, 其中后三种编码方式是最有效的编码方式, 这几种编码方式充分利用局部特征的相关信息对笔迹检索效果的促进作用较大. 而取均值法由于信息损失过多, 使得其编码效果得不到较好的提升. 但本文为了验证 DLS-CNN 模型的学习能力与检验提取的局部特征的鲁棒性, 并与 Fiel 等^[8]提出的模型进行有效对比, DLS-CNN 型采用和 Fiel 等^[8]一致的取均值的编码方式. 这种取均值的方式是将一份原始笔迹材料的所有小像素块的特征向量的均值作为这份原始笔迹材料的全局特征. 因此第 j 份原始笔迹材料的全局特征 \mathbf{V}_j 为:

$$\mathbf{V}_j = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i^{(j)} \quad (5)$$

其中, N 为第 j 份原始笔迹材料的小像素块数量, $\mathbf{x}_i^{(j)}$ 为第 j 份原始笔迹材料的第 i 个小像素块的特征向量. 因为当模型使用信息损失较大的编码方式都能取得较好的效果时, 说明该模型学习的局部特征具有更强鲁棒性与泛化能力. 随后, 将编码好的全局特征求取待查询笔迹文档的余弦相似度作为检索标准.

2 实验结果与分析

由于 ICDAR2013^[3] 和 CVL^[26] 这两个具有挑战性的多语言标准数据集都是近些年笔迹鉴别研究使用最多的数据集之一, 故我们同时使用这两个标准数据集评估 DLS-CNN 提取特征的强鲁棒性和模型的学习能力. 其中, 我们在 ICDAR2013 数据集上做笔迹检索任务, 在 CVL 数据集上做笔迹识别任务.

2.1 评估标准

为评估模型的鲁棒性与泛化能力, 我们使用平均准确率均值 (Mean average precision, mAP)、Soft top- k 、Hard top- k 三个评估标准用于笔迹检索任务中, 这几个评估标准是广泛应用于

信息检索任务的标准. 我们使用 Top- k 评估标准用于笔迹识别任务中.

平均准确率均值: mAP 是平均准确率的均值, 其是反映模型在所有相关文档上的性能指标. 假设我们有 N 个待查询笔迹材料, 则第 i 个查询笔迹材料的平均准确率 $AP(i)$ 为:

$$AP(i) = \frac{\sum_{k=1}^M P(k) \cdot rel(k)}{R} \quad (6)$$

其中, M 表示查询库 (待查询笔迹材料库) 中的笔迹材料数量, 此论文中, 我们将数据集 ICDAR2013 的测试集作为一个查询库. R 表示查询库中与第 i 个查询笔迹材料相关 (同一笔迹) 的文档总数. $P(k)$ 表示查询结果中前 k 个结果的准确率, 即前 k 个查询结果中出现相关笔迹材料的个数与 k 的比值. $rel(k)$ 是一个指示函数, 如果第 k 个查询结果是相关文档, 则其值为 1, 否则为 0. 从而 mAP 定义为:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (7)$$

Soft top- k 与 Hard top- k : Soft top- k (S- k) 的评估方法与累计匹配特性值 (Cumulative matching characteristic, CMC) 等价, 其表征模型在单个相关文档上的性能指标. 当每次查询的前 k 个查询结果含有一个 (及以上) 相关文档时, 我们将其定义为一次正确查询, Soft top- k 表示为所有正确查询的次数与总查询次数的百分比. 而在 Hard top- k (H- k) 评估标准中, 当返回的前 k 个查询结果全是相关文档时, 我们将其定义为一次正确查询, Hard top- k 表示为正确的查询次数与总查询次数的百分比, 其刻画模型在多个相关文档上的性能指标. 很明显, Soft top- k 随着 k 的值增加而升高, Hard top- k 随着 k 的值增加而降低.

Top- k : Top- k 评估指标用于笔迹识别任务, 与 Soft top- k 评估指标意义一致, 都是反应模型在单个相关文档上的性能指标. 当每次识别的前 k 个最相似的类中含有与目标类一致的类, 我们将其定义为一次正确识别, 则 Top- k 表示为所有正确识别的次数与总识别次数的百分比.

2.2 ICDAR2013 数据集

ICDAR2013^[3] 标准数据集的训练集和测试集分别包含了 100 和 250 个书写者的笔迹材料, 每个书写者贡献 4 张笔迹材料, 其中 2 张的笔迹内容为英语, 另外两张为希腊语. 由于混合语言, 这个数据集的笔迹识别具有较大的挑战. 对数据集分别进行 64 与 256 尺度的像素切割之后, 64 尺度切割的小像

素块数量为: 训练集 70 911 个, 测试集 218 999 个; 256 尺度切割的小像素块数量为: 训练集 21 160 个, 测试集 64 820 个. 我们分别对这两个尺度切割的数据进行了实验, 结果如表 2 所示. 结果显示: 基于词级别大小数据提取的特征比基于字母级别大小数据提取的特征, 在所有的指标上均取得更好的表现, 表明基于词级别大小的数据更适合于 DLS-CNN, 因为基于词级别大小的数据包含更多的完整笔迹. 并且, 256 尺度切割的像素块比 64 尺度切割的像素块少, 使得模型在 256 尺度的数据上进行实验的运行时间相对较短. 为此, 接下来的所有实验均基于 256 尺度切割的数据.

表 2 不同像素块大小的对比 (%)
Table 2 Comparison of different patch sizes (%)

	S-1	S-5	S-10	H-2	H-3	mAP
64 尺度	87.8	94.7	97.0	57.3	36.7	76.5
256 尺度	95.0	98.4	99.3	70.3	49.5	84.7

表 3 不同特征层的对比 (%)
Table 3 Comparison of different feature layers (%)

	S-1	S-5	S-10	H-2	H-3	mAP
全局池化层	95.4	97.9	98.5	63.1	41.2	79.7
全连接层	95.0	98.4	99.3	70.3	49.5	84.7

表 4 特征数目的对比 (%)
Table 4 Comparison of feature numbers (%)

	S-1	S-5	S-10	H-2	H-3	mAP
128	95.2	98.7	99.0	70.1	48.6	84.3
512	95.0	98.4	99.3	70.3	49.5	84.7
1 024	95.0	98.4	99.0	70.0	48.8	84.1
2 048	96.0	98.4	98.6	67.1	45.8	83.0

此外, 我们探讨了不同特征层提取特征的表现能力. 在实验中分别提取了全局池化层与全连接层的特征作为局部特征, 实验结果如表 3 所示. 结果显示: 全连接层提取的特征仅 Soft top-1 指标比全局池化层提取的特征表现差, 在其他的指标中都有更加优异的表现. 这主要归因于: 1) 全连接层提取的特征对前面层的特征进一步的抽象成为高阶的描述特征, 更能全局地描述输入图像的特点; 2) 分类层以全连接层的特征作为分类的依据, 分类层所产生的损失将直接作用于全连接层, 使得全连接层更能体现分类信息. 在此基础上, 我们对提取的特征数做了进一步的探讨, 分别对 128, 512, 1 024, 2 048 个特征数目进行了实验. 实验表明, 当取 512 个特征数时, 模型具有更好的表现, 并且模型所需的训练时间与检索时间也相对较短, 结果如表 4 所示.

在 DLS-CNN 模型中, 主成分分析也对模型产生了积极作用, 我们将提取 95 % 的主成分特征作为

最终的局部特征. 结果显示, PCA 白化不仅提高模型的泛化能力, 而且还通过压缩特征节约了笔迹检索时间, 结果如表 5 所示.

表 5 PCA 白化的评估 (%)
Table 5 Evaluation of PCA Whitening (%)

	S-1	S-5	S-10	H-2	H-3	mAP
无 PCA 白化	88.9	97.1	98.0	63.9	47.6	82.1
有 PCA 白化	95.0	98.4	99.3	70.3	49.5	84.7

最后, 也将 DLS-CNN 与其他当前表现较好的相似模型进行对比. 特别是与 Fiel 等^[8] 提出的模型对比, 因为 DLS-CNN 的处理过程与 Fiel 等提出的模型最为相似. 由于 Fiel 等提出的方法对于上下行粘连的笔迹和位于上下行中间的笔迹不能正确分割, 使得其在更具挑战性的 ICDAR2013 数据集上的鉴别表现不佳. 实验结果显示, 我们提出的模型在大部分的评估指标都取得明显更优的表现, 并且在 Hard top-2 与 Hard top-3 上分别有 29.8 % 与 33.7 % 的提升, 说明 DLS-CNN 弥补了基于行分割与 CNN 模型的不足. 此外, DLS-CNN 模型与 Christlein 等^[24] 提出的模型在同样使用取均值编码的情况下进行对比, DLS-CNN 在 Soft top-1 与 mAP 两个评估指标上分别提升了 8.2 % 与 5.8 %, 表明基于统计的文档行分割与深度卷积神经网络的笔迹鉴别方法的模型能学得更具鲁棒性的特征, 具有更强的泛化能力. 但是我们提出的方法在 Soft top-k 的指标上并没有取得更加优异的表现. 我们通过进一步实验找出那些未能正确查询的笔迹材料, 发现我们的方法对粗笔所写的笔迹材料识别效果差. 细笔/粗笔像素块的样例如图 3(c) 所示. 可能原因有以下两个: 1) 由于数据集中粗笔写的笔迹材料极少, 导致模型不能较好地学习粗笔所写笔迹材料的相应特征; 2) 由于粗笔所写笔迹材料不能较好地被正确分割, 使得所获取的像素块不能保留更多完整的笔迹信息. 具体实验对比如表 6 所示.

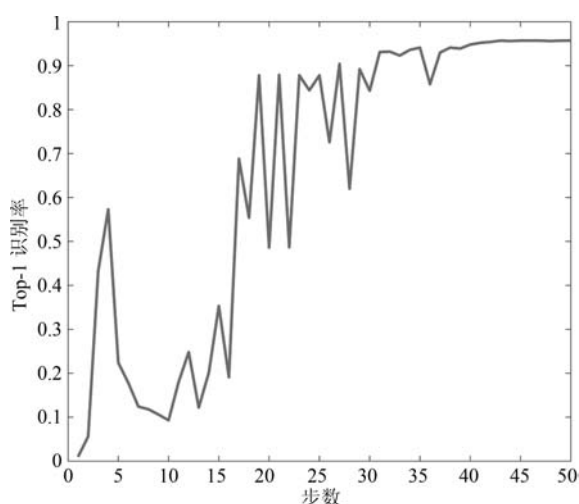
2.3 CVL 数据集

CVL^[26] 数据集的训练集与测试集分别包含 27 与 283 个书写者, 在训练集与测试集中每个书写者分别贡献 7 份与 5 份笔迹材料, 每个书写者写一份德语笔迹材料, 其他均为英语笔迹材料. 在实验中, 在训练集中每个作者取 5 份与测试集笔迹材料内容相同的笔迹材料. 最终将 310 位书写者贡献的 1 550 份笔迹材料用 256 尺度分割成 153 315 个小的像素块, 并按 8:1:1 的比例分为训练集、验证集和测试集. 在此数据集上, 为了检验模型在笔迹识别任务的鲁棒性与泛化能力, 我们直接对 256 尺度大小的小像素块进行训练与测试.

表 6 与其他模型的对比 (%)

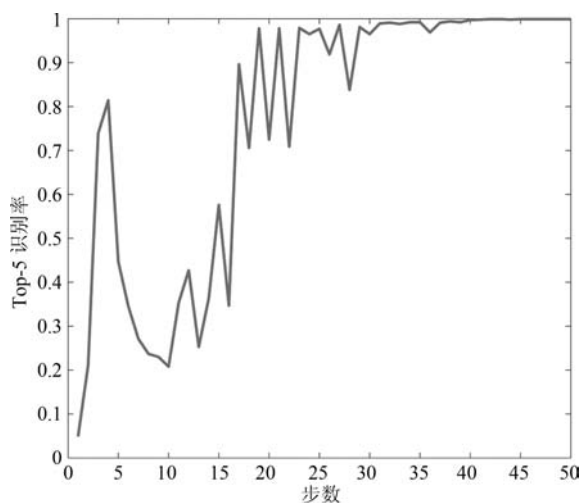
Table 6 Comparison with other models (%)

	S-1	S-5	S-10	H-2	H-3	mAP
CS-UMD-a ^[3]	95.1	98.6	99.1	19.6	7.1	N/A
CS-UMD-b ^[3]	95.0	98.6	99.2	20.2	8.4	N/A
HIT-ICG ^[3]	94.8	98.0	98.3	63.2	36.5	N/A
TEBESSA-a ^[3]	90.3	96.7	98.3	58.2	33.2	N/A
TEBESSA-b ^[3]	93.4	97.8	98.5	62.6	36.5	N/A
Christlein ^[11]	97.1	98.8	99.1	42.8	23.8	67.7
Wu ^[9]	95.6	98.6	99.1	63.8	36.5	N/A
Nicolaou ^[14]	97.2	98.9	99.2	52.9	29.2	N/A
Fiel ^[8]	88.5	96.0	98.3	40.5	15.8	N/A
Christlein ^[24]	86.8	N/A	N/A	N/A	N/A	78.9
DLS-CNN	95.0	98.4	99.3	70.3	49.5	84.7



(a) Top-1 识别率

(a) Top-1 identification rate



(b) Top-5 识别率

(b) Top-5 identification rate

图 4 256 尺度大小的识别率

Fig. 4 The identification rate of 256 patch size

实验结果如图 4 所示. 当训练步数达到 35 步之后, DLS-CNN 的识别率基本稳定, Top-1 的最高识别率为 95.8%, Top-5 的最高识别率达到 99.9%. 此外, 由于前 20 步的学习率较大, 使得模型不易收敛并且未找到较好的局部最优点, 在 20 步后将学习率缩小 5 倍, 模型相对比较稳定, 在 30 步之后进一步将学习率缩小 2 倍, 模型慢慢收敛并趋于稳定. 具体结果如图 4 所示, 图 4(a) 表示 Top-1 的识别率, 图 4(b) 表示 Top-5 的识别率. 与当前在 CVL 数据集获得识别率效果最好的算法的相比, DLS-CNN 仅以一个 256 尺度大小的像素块在 Top-5 的识别率相较于其他基于 1 整张笔迹材料的 Top-5 识别率从 99.8% 提升到 99.9%, 取得当前最高的识别率, 如表 7 所示. 但我们的方法在 Top-1 评估指标上的表现不理想, 原因在于我们的方法对粗笔所写的笔迹材料不能提取比较有效的特征对其进行更好地识别. 实验表明, DLS-CNN 能够以很少量的笔迹信息对笔迹进行较准确地识别, 体现了 DLS-CNN 能在笔迹识别任务中具有较强的鲁棒性与泛化能力.

表 7 与其他模型的对比 (%)

Table 7 Comparison with other models (%)

	输入笔迹材料	Top-1	Top-5
TSINGHUA ^[26]	1 页	97.7	99.0
Fiel ^[8]	1 页	98.9	99.3
Wu ^[9]	1 页	99.2	99.5
Nicolaou ^[14]	1 页	99.0	99.4
Christlein ^[38]	1 页	99.4	N/A
Tang ^[13]	1 页	99.7	99.8
DLS-CNN	256 像素块	95.8	99.9

3 总结与展望

本文提出了一种基于统计的文档行分割与深度卷积神经网络结合的离线笔迹鉴别方法, 以学习具有强鲁棒性的局部特征和较强泛化能力的模型为目标, 解决了其他基于局部特征的笔迹鉴别方法提取的局部特征过度依赖数据增强与全局编码的问题和模型泛化能力不足的问题. 这得益于基于统计的文档行分割方法进行文档行分割的准确性, 以及改进的残差神经网络强学习能力的优越性. 在 ICDAR2013 与 CVL 两个含多语言的标准数据集上的实验结果表明了 DLS-CNN 是一种具有强鲁棒性的离线笔迹鉴别方法.

在未来的工作中, 基于统计的文档行分割方法可以做进一步的改进, 通过更有效的方法对柱状图低谷和高峰进行更准确的确定, 使得模型在初始行的寻找更加准确, 从而得到更加精确的行分割. 在实验中发现该模型对于粗笔写的笔迹材料不能较好的鉴别, 致使我们提出的方法不能在 Soft top-k 指标

上取得较好的表现, 我们将对此问题做更进一步的研究. 此外, 寻求一种合适的正则化方法对 CNN 模型进行正则化, 防止模型的过拟合且使模型更加稳定, 并在多个包含其他语言 (中文、拉丁文、法语等) 的数据集上对模型做进一步的验证.

References

- 1 Fiel S, Kleber F, Diem M. ICDAR2017 Competition on Historical Document Writer Identification (Historical-WI). In: Proceedings of the 14th International Conference on Document Analysis and Recognition. Kyoto, Japan: IEEE, 2018. 1377–1382
- 2 Asi A, Abdalhaleem A, Fecker D. On writer identification for Arabic historical manuscripts. *International Journal on Document Analysis and Recognition*, 2017, **2017**(3–4): 1–15
- 3 Louloudis G, Gatos B, Stamatopoulos N. ICDAR 2013 Competition on Writer Identification. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, DC, USA: IEEE, 2013. 1397–1401
- 4 Cloppet F, Eglis V, Kieu V C. ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script. In: Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition. Shenzhen, China: IEEE, 2017. 1371–1376
- 5 Chawki D, Somaya A M, Imran S, Abdeljalil G, He Sheng. ICFHR 2018 Competition on Multi-Script Writer Identification. In: Proceedings of the 16th International Conference on Frontiers in Handwriting Recognition. Niagara Falls, USA: IEEE, 2018. 506–510
- 6 Helli B, Moghaddam M E. A text-independent Persian writer identification based on feature relation graph (FRG). *Pattern Recognition*, 2010, **44**(6): 229–240
- 7 Bertolini D, Oliveira L S, Justino E. Texture-based descriptors for writer identification and verification. *Expert Systems with Applications*, 2013, **40**(6): 2069–2080
- 8 Fiel S, Sablatnig R. Writer Identification and Retrieval Using a Convolutional Neural Network. In: Proceedings of the 16th International Conference on Computer Analysis of Images and Patterns. Springer International Publishing, 2015. 26–37
- 9 Wu Xiang-Qian, Tang You-Bao, Wei Bu. Offline Text-Independent Writer Identification Based on Scale Invariant Feature Transform. *Information Forensics and Security*, 2014, **9**(3): 526–536
- 10 Xing Lin-Jie, Qiao Yu. DeepWriter: A Multi-Stream Deep CNN for Text-independent Writer Identification. In: Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition. Shenzhen, China: IEEE, 2017. 584–589
- 11 Christlein V, Bernecker D, Honig F. Writer Identification Using GMM Supervectors and Exemplar-SVMs. *Pattern Recognition*, 2017, **63**: 258–267
- 12 Christlein V, Gropp M, Fiel S, Maier A. Unsupervised Feature Learning for Writer Identification and Writer Retrieval. arXiv preprint arXiv: 1705.09369, 2017
- 13 Tang You-Bao, Wu Xiang-Qian. Text-Independent Writer Identification via CNN Features and Joint Bayesian. In: Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition. Shenzhen, China: IEEE, 2017. 556–571
- 14 Nicolaou A, Bagdanov A D, Liwicki M. Sparse radial sampling LBP for writer identification. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis, Tunisia: IEEE, 2015. 716–720
- 15 Chen Shi-Ming, Wang Yi-Song, Lin Chin-Teng, Ding Wei-Ping, Cao Ze-Hong. Semi-supervised Feature Learning For Improving Writer Identification. arXiv preprint arXiv: 1807.05490, 2018
- 16 Li Xin, Ding Xiao-Qing, Peng Liang-Rui. Writer identification based on improved microstructure features. *Acta Automatica Sinica*, 2009, **35**(9): 1199–1208
(李昕, 丁晓青, 彭良瑞. 一种基于微结构特征的多文种文本无关笔迹鉴别方法. *自动化学报*, 2009, **35**(9): 1199–1208)
- 17 Zou Jie, Sun Bao-Lin, Yu Jun. Online handwriting matching algorithm based on stroke features. *Acta Automatica Sinica*, 2016, **42**(11): 1744–1757
(邹杰, 孙宝林, 於俊. 基于笔画特征的在线笔迹匹配算法. *自动化学报*, 2016, **42**(11): 1744–1757)
- 18 Khan F A, Tahir M A, Khelifi F. Novel geometric features for off-line writer identification. *Pattern Analysis and Applications*, 2016, **19**(3): 699–708
- 19 Bertolini D, Oliveira L S, Sabourin R. DeepWriter: Multi-script writer identification using dissimilarity. In: Proceedings of the 23rd International Conference on Pattern Recognition. Cancun, Mexico: IEEE, 2017. 3025–3030
- 20 Shaus A, Turkel E. Writer Identification in Modern and Historical Documents via Binary Pixel Patterns, Kolmogorov–Smirnov Test and Fisher’s Method. *Journal of Imaging Science and Technology*, 2017, **61**(1): 104041–104049
- 21 He S, Schomaker L. Writer identification using curvature-free features. *Pattern Recognition*, 2017, **63**: 451–446
- 22 Khan F A, Tahir M A, Khelifi F. Robust off-line text independent writer identification using bagged discrete cosine transform features. *Expert Systems with Applications*, 2017, **71**(C): 404–415
- 23 Bulacu M, Schomaker L. Text-Independent Writer Identification and Verification Using Textural and Allographic Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(4): 1–17
- 24 Christlein V, Maier A. Encoding CNN Activations for Writer Recognition. arXiv preprint arXiv: 1712.07923, 2017
- 25 Louloudis G, Stamatopoulos N, Gatos B. ICDAR 2011 Writer Identification Contest. In: Proceedings of the 11th International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 1475–1479
- 26 Kleber F, Fiel S, Diem M, Sablatnig R. CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington DC, USA: IEEE, 2013. 560–564

- 27 Diem M, Kleber F, Sablatnig R. Text Line Detection for Heterogeneous Documents. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. In: Proceedings of International Conference on Document Analysis and Recognition. Washington DC, USA: IEEE, 2013. 743–747
- 28 Marti U V, Bunke H. The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 2002, **5**(1): 39–46
- 29 Liu Cheng-Lin, Yin Fei, Wang Da-Han, Wang Qiu-Feng. CASIA online and offline Chinese handwriting databases. In: Proceedings of the 11th International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 37–41
- 30 Arivazhagan M, Srinivasan H, Srihari S. A statistical approach to line segmentation in handwritten documents. *Document Recognition and Retrieval XIV*, 2007, **6500**(T): 1–11
- 31 Liu Ji-Ming, Tang Yuan-yan. Adaptive Image Segmentation with Distributed Behavior-based Agents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, **21**(6): 544–551
- 32 You Xin-Ge, Peng Qin-Mu, Yuan Yuan, Cheung Yiu-Ming, Lei Jia-Jia. Segmentation of Retinal Blood Vessels Using the Radial Projection and Semi-supervised Approach. *Pattern Recognition*, 2011, **44**(10–11): 2314–2324
- 33 He Kai-Ming, Zhang Xiang-Yu, Ren Shao-Qing, Sun Jian. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, United States: IEEE, 2016. 770–778
- 34 He Kai-Ming, Zhang Xiang-Yu, Ren Shao-Qing, Sun Jian. Identity mappings in deep residual networks. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer International Publishing, 2016. 630–645
- 35 Christlein V, Bernecker D, Angelopoulou E. Writer identification using VLAD encoded contour-Zernike moments. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis, Tunisia: IEEE, 2015. 906–910
- 36 Spyromitros-Xioufis E, Papadopoulos S, Kompatsiaris I Y. A Comprehensive Study Over VLAD and Product Quantization in Large-Scale Image Retrieval. *IEEE Transactions on Multimedia*, 2014, **16**(6): 1713–1728
- 37 Fiel S, Sablatnig R. Writer Identification and Writer Retrieval Using the Fisher Vector on Visual Vocabularies. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington DC, USA: IEEE, 2013. 545–549
- 38 Christlein V, Bernecker D, Maier A. Offline Writer Identification Using Convolutional Neural Network Activation Features. In: Proceedings of the 37th German Conference on Pattern Recognition. Aachen, Germany: Springer International Publishing, 2015. 540–552



陈使明 贵州大学计算机科学与技术学院硕士研究生. 主要研究方向为模式识别, 计算机视觉, 机器学习.

E-mail: gchenshiming@gmail.com

(**CHEN Shi-Ming** Master student at the College of Computer Science and Technology, Guizhou University. His research interest covers pattern recognition, computer vision, and machine learning.)



王以松 贵州大学计算机科学与技术学院教授. 主要研究方向为知识表示与推理, 机器学习, 人工智能. 本文通信作者.

E-mail: yswang@gzu.edu.cn

(**WANG Yi-Song** Professor at the College of Computer Science and Technology, Guizhou University. His research interest covers knowledge representation and reasoning, machine learning and artificial intelligence. Corresponding author of this paper.)