

基于深度学习的手写书法字体识别算法^①



许嘉谕, 林楚焯, 陈志涛, 邓卓然, 潘家辉, 梁 艳

(华南师范大学 软件学院, 佛山 528225)

通讯作者: 梁 艳, E-mail: liangyan@m.scnu.edu.cn

摘 要: 为解决因手写书法作品种类繁多而识别困难的问题, 降低人们观赏书法的门槛, 本文提出了基于深度学习的手写书法字体识别算法. 识别过程中首先使用投影法等图像处理方法对书法作品图像中的汉字进行定位和分割, 然后分别利用 GoogLeNet Inception-v3 模型和 ResNet-50 残差网络进行书体风格识别和字形识别. 实验结果表明, 本文算法能实现对整幅书法作品中楷书和篆书的书体风格以及字形的识别, 对楷书和篆书单字的识别率分别为 91.57% 和 81.70%, 达到了实用的需求.

关键词: 图像处理; 卷积神经网络 (CNN); 手写书法字体识别

引用格式: 许嘉谕, 林楚焯, 陈志涛, 邓卓然, 潘家辉, 梁艳. 基于深度学习的手写书法字体识别算法. 计算机系统应用, 2021, 30(2): 213-218. <http://www.c-s-a.org.cn/1003-3254/7819.html>

Handwritten Calligraphy Font Recognition Algorithm Based on Deep Learning

XU Jia-Yu, LIN Chu-Ye, CHEN Zhi-Tao, DENG Zhuo-Ran, PAN Jia-Hui, LIANG Yan

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: In order to solve the problem of difficult recognition due to the large variety of handwritten calligraphy fonts and reduce the threshold for people to appreciate calligraphy, we propose a handwritten calligraphy font recognition algorithm based on deep learning. In the process of recognition, image processing methods, such as projection method, are first used to locate and segment the Chinese characters in the calligraphy works. Then, the GoogLeNet Inception-v3 model and ResNet-50 residual network are used to recognize the styles and shapes of Chinese characters. Consequently, this algorithm can recognize the styles and shapes of regular script and seal script in an entire calligraphy work at single-character recognition rates of 91.57% and 81.70%, demonstrating its practicability.

Key words: image processing; Convolutional Neural Network (CNN); handwritten calligraphy font recognition

1 引言

1.1 手写书法字体识别研究背景

汉字是传承中华文化的重要载体, 而书法展现了中国汉字独特的书写艺术, 是中华传统文化的瑰宝. 然而古代书法作品大多数采用繁体汉字进行书写^[1], 且不同的书法风格导致字体形态复杂多变, 提高了人们阅

读和理解书法作品的门槛. 尤其是篆书风格的书法作品, 由于其汉字形态保留了古代象形文字的特点, 未经过专业学习的人士很难熟练地将各个篆体汉字准确辨别出来. 因此利用计算机实现识别书法字体可为书法学习者提供鉴赏指导, 同时有利于降低大众鉴赏书法的难度, 向大众传播优秀书法文化.

① 基金项目: 国家自然科学基金面上项目 (61876067); 广东省自然科学基金面上项目 (2019A1515011375); 广州市科技计划项目重点领域研发计划 (202007030005)

Foundation item: General Program of National Natural Science Foundation of China (61876067); General Program of Natural Science Foundation of Guangdong Province (2019A1515011375); Research and Development Plan of Key Areas, Science and Technology Plan of Guangzhou Municipality (202007030005)

收稿时间: 2020-06-25; 修改时间: 2020-07-27; 采用时间: 2020-08-19; csa 在线出版时间: 2021-01-27

目前造成书法字体识别困难的原因主要有3个方面. 第一, 汉字数量繁多且字体结构复杂. 据统计中国汉字字符数量超过 80 000^[2], 字体结构与数字和字母相比更加复杂, 同时还不乏字形相似的汉字, 使得识别汉字字形的难度大大增加.

第二, 书法汉字的字体风格多样化. 从古至今各书法名家创造了多种字体风格, 同一个汉字使用不同字体风格书写将展现不同的形态且差异较大, 例如楷体、隶书、篆书等.

第三, 不同书法作者具有不同的书写习惯, 因此书法手写体的呈现形态和排布方式会因人而异. 比如楷书的字体风格又分成颜体、柳体等, 它们同为楷书却都有各自的特点. 与此同时古代书法作品与现今汉字印刷体的排布方式有着较大的差异, 因此汉字印刷体的识别方法不适用于书法字体识别.

1.2 研究现状

在图像识别和模式识别领域中, 汉字识别是重要的研究课题之一. 日常的汉字手写体识别经过多年的研究, 如今识别率已经达到了 90% 以上甚至更高^[3]. 然而目前针对书法汉字的识别技术发展比较滞后, 但也已经有了许多研究成果.

浙江大学硕士研究生顾刚的学位论文《汉字识别关键算法研究与应用》^[2]中提出了基于卷积神经网络的书法字识别方法: 首先利用计算机系统标准字库和汉字图书字库训练卷积神经网络模型, 使之能够判断出待识别图像的书体风格, 然后使用 MQDF 算法在对应字体下的特征库进行识别. 经过实验分析可知卷积神经网络能有效提取图像的深度特征, 在识别书法字体的速度和正确率等方面具有很大的优势. 此外西安理工大学张福成的硕士学位论文《基于卷积神经网络的书法风格识别的研究》^[4]和温佩芝等的期刊论文《基于卷积神经网络的石刻书法字识别方法》^[5]也提供了针对书法风格和字体的识别方法. 然而这些研究成果都是针对计算机中的标准书法字库或石刻书法图像进行识别, 未解决识别手写书法字体问题.

2014年浙江大学硕士研究生林媛的学位论文《中国书法字识别算法研究及应用》^[6]中提出了基于检索的书法字识别, 能够高效识别楷书、篆书、行书等书法风格的单个手写书法字体图片. 这些研究成果为书法字体识别的技术实现提供了很多理论依据, 但都未提出识别完整手写书法作品图像的解决方法, 目前识别手写书法作品图片面临着诸多挑战. 因此本文算法

结合图像处理方法和深度学习技术实现完整手写书法作品图像的书体风格和内容识别, 为大众学习和欣赏书法提供解读参考.

2 书法字体识别算法描述

本文算法对输入图像进行预处理和识别, 最终输出包含书法风格和字体内容的识别结果, 算法描述流程如图 1 所示. 首先对待识别图像进行预处理, 消除图片噪声并去除框线; 然后进行目标分割, 对书法作品中的单字进行定位并分割出目标单字; 随后利用 GoogLeNet Inception-v3 模型^[7]和 ResNet-50 残差网络^[8]分别实现书体风格和单字字形的识别, 在判断出书法作品图像的书法风格后选择对应书体的单字字形识别模型进行识别, 最后输出书法作品中的书体风格和汉字识别结果.

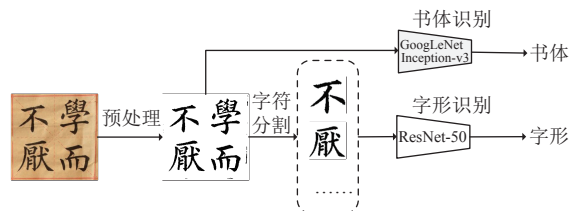


图 1 书法字体识别算法流程框架图

3 图像预处理

3.1 图像去噪声和二值化

由于部分书法作品年代久远, 拍摄得到的图片中可能会有噪点. 本文方法采用图像形态学的方法, 使用开运算腐蚀图像中的噪点, 可在不影响书法字体的情况下达到去噪效果.

随后对待识别图像进行二值化处理, 图像二值化是指将像素值分为黑白两种元素的集合, 以黑色像素作为前景, 白色像素作为背景, 以达到区分目标和背景像素的目的. 由于书法作品形式多样, 不同作品图像的对比如各异^[9], 为了降低图像对比如带来的影响, 采用最大类间方差法^[10]对图像进行二值化处理.

取多个分割阈值 T 区分图像中的前景与背景, 设前景像素点占比为 m_0 , 其平均灰度为 n_0 ; 背景像素点占比为 m_1 , 其平均灰度为 n_1 , 计算图像前景和背景类间方差 g , 如下所示:

$$g = m_0 \times m_1 \times (n_0 - n_1)^2 \quad (1)$$

类间方差越大说明类间差异越大, 因此最后以类间方差计算结果最大的阈值来区分图像前景与背景.

3.2 去除文档边框线

部分书法作品可能具有文档边框线, 将会对后续的单字分割和识别产生影响, 因此需要去除文档边框线. 本文算法使用矩形结构化元素分别得到图像水平边框线和垂直边框线后, 融合获取图像中所有多余边框线, 再利用帧差法去除边框线. 设原图像为 $f_1(x,y)$, 边框线图像为 $f_2(x,y)$, 进行求差得到差分图像 $F(x,y)$, 即去除文档边框线后的书法作品图像, 计算公式如下所示:

$$F(x,y) = |f_1(x,y) - f_2(x,y)| \quad (2)$$

4 书法汉字的定位与分割

书法汉字的定位与分割是书法单字识别前的重要处理步骤, 如图 2 所示, 一共分为 4 个步骤: 列分割、行分割、单字轮廓提取及切割和扩大画布.



图 2 书法汉字的定位与分割流程

4.1 列分割

对图像进行垂直投影^[11], 由于书法作品一般从上至下进行书写, 列与列之间较少有黏连, 可依据竖直方向上背景点的数量获取竖直切割线的位置, 同时得出列的平均宽度, 若待切割的某列的列宽度大于平均宽度的两倍, 表示列与列之间存在黏连, 需对该列再次进行切割. 列分割效果如图 3 所示.



图 3 列分割效果

4.2 行分割

对切割好的列图像进行水平投影, 由于汉字结构中存在上下结构的字形, 进行行切割时有可能将上下结构的字体分割开^[12]. 因此对初次行分割后得到的图像进行检测, 若分割后的图像宽度小于平均宽度 1/3, 对该图像的相邻分割图像重新进行检测, 若均小于 1/3,

则说明已将单字分割开, 需要进行图像合并^[13]. 行分割效果如图 4 所示.



图 4 行分割效果

4.3 单字轮廓提取及切割

切割后的单字图像可能还会存在细小的噪点, 因此需利用 OpenCV 的 findContours 函数提取单字图像的轮廓并进行再一次切割, 能够有效消除噪声并切除多余的背景像素.

4.4 扩大画布

对画布上、下、左、右 4 个方向上分别进行 20% 的背景像素填充, 使得识别目标处于图片正中心后, 统一将单字图片处理成 144×144 的尺寸, 便于进行后续的数据增强处理.

5 书法作品的风格识别

书体识别是对图像中的书法风格进行识别. 书体识别模型是字形识别模型的前置模型, 用户输入待识别图片后, 经书体识别模型处理后得出此作品所用书体并根据书体识别结果调用对应字形识别模型, 因此书体识别模型对后续的字形识别至关重要.

本文算法利用 GoogLeNet Inception-v3 模型^[7]实现书体识别, 该模型以稀疏连接的方式扩增网络的深度与宽度, 能够在扩大网络规模的同时, 节约参数数量, 防止网络计算量过大. GoogLeNet Inception-v3 网络主要由多个 Inception-v3 模块组成, 模块内部结构如图 5 所示, GoogLeNet Inception-v3 网络整体结构如图 6 所示.

原始 Inception 模块使用多个不同大小的卷积核, 对输入图像执行并行的卷积操作以及最大池化操作, 以提取出抽象的书法风格特征. 随后将该模块的卷积输出拼接起来, 并输入至下一个网络模块中. Inception-v3 模块对原始的 Inception 模块作出改进, 利用 1×1 卷积操作对图片进行降维, 随后将原始 Inception 模块中较大的二维卷积核拆分成两个较小的一维卷积核, 尺寸分别是 1×n 和 n×1, 以达到节约参数和加速网络运算

的目的.与此同时 Inception-v3 仍然保留了 4 个卷积分支的设置,使得卷积神经网络保持一定的宽度,能更高效地提取书法作品风格的抽象信息.

损失函数使用多分类交叉熵,设 y_i 为模型预测目标 x 属于第 i 类的概率,符号 t_i 标记目标样本 x 是否属于第 i 个类别,即:

$$t_i = \begin{cases} 1, & x \text{ 属于第 } i \text{ 类} \\ 0, & x \text{ 不属于第 } i \text{ 类} \end{cases} \quad (3)$$

那么,损失函数可按照下列公式进行计算:

$$E = - \sum_{i=1}^n t_i \log(y_i) \quad (4)$$

书体识别模型通过随机梯度下降的方式来进行迭代训练,模型训练时将学习率设置成 0.01,迭代次数设置 10000 并保存具有最优测试结果的模型参数.

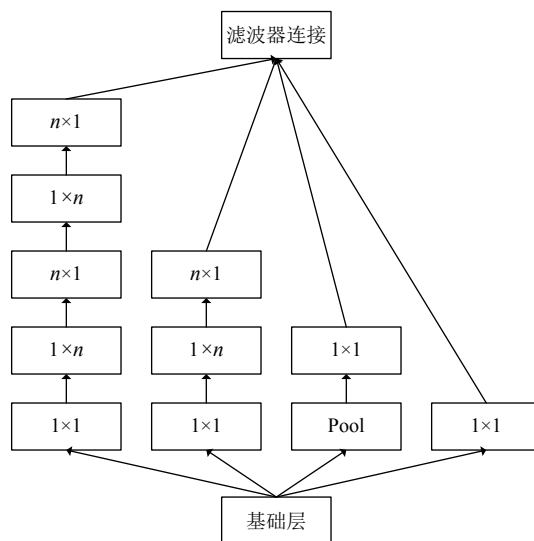


图 5 Inception-v3 模块结构图

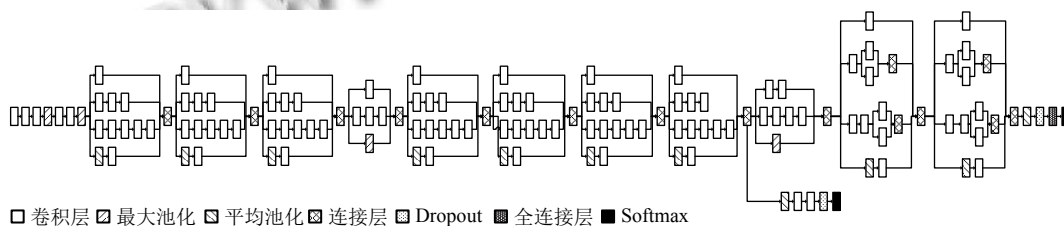


图 6 GoogLeNet Inception-v3 网络结构图

6 书法汉字的字形识别

6.1 数据增强

深度神经网络拥有大量的参数,其数量级可达到数百万以上.如果训练数据不充足,将可能会出现数据过拟合的现象.因此需要在原有数据集基础上进行数据增强,对图像数据进行旋转、裁剪、移位等操作,制作出更丰富的数据,从而使得网络模型具有更好的泛化能力和鲁棒性.考虑到书法作品图像的特点,本文采用的数据增强方法有以下 4 种:对目标单字进行上下左右 4 个方向上 0~20% 以内的随机平移、任意角度旋转、高斯模糊.经过数据增强后每个书法单字有 500 张图像,大大增加了训练数据量.

6.2 书法字形识别模型

由于书法单字分类具有三千多种,网络模型需要有足够的深度才能满足目标识别效果.然而网络深度到达一定程度之后,单纯叠加神经网络层数可能会出现网络退化的问题^[8],导致模型检测效果比浅层网络更差.因此本文算法选择 ResNet-50 网络模型来实现书法

单字字形识别功能,残差网络能够在网络通过增加残差块来解决梯度爆炸的问题^[8].ResNet-50 网络主要由残差学习模块和卷积模块组成,这两个模块均会对输入特征图进行卷积和拼接处理,提取图像中的书法单字字形特征,其模块内部结构如图 7 所示,网络整体结构如图 8 所示.

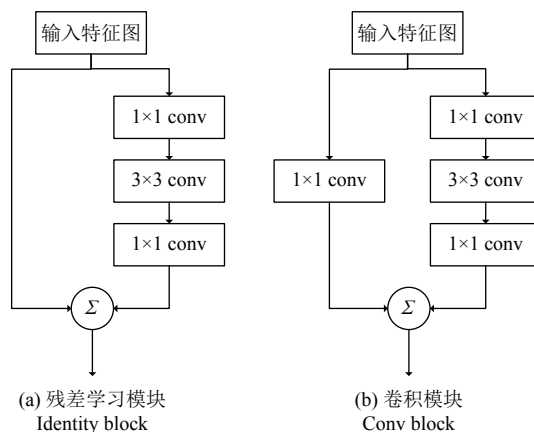


图 7 残差学习模块和卷积模块内部结构图

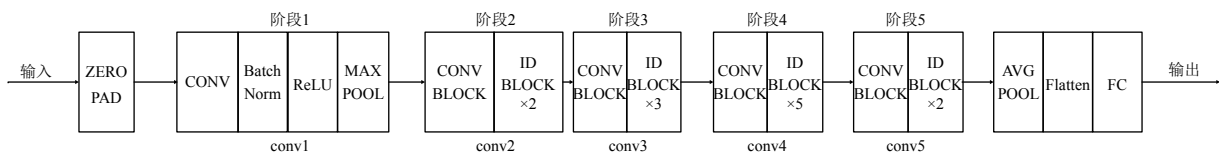


图8 ResNet-50网络整体结构图

残差学习模块有两个分支,其中一条分支通过短路连接来学习残差来进行优化,另一分支中的两个 1×1 卷积核的作用是分别对特征图进行降维和升维,模块的输出可定义为:

$$y = F(x, \{W_i\}) + x \quad (5)$$

其中, x 表示输入特征图, y 表示输出特征图, $F(x, \{W_i\})$ 表示对 x 进行多层卷积.残差学习模块的作用是在网络训练过程中,当低层网络提取的特征已经足够成熟时,更深层网络在训练时能通过中间隐层推导实现恒等映射,因而避免了网络退化的问题.

另外,卷积模块的输出可定义为:

$$y = F(x, \{W_i\}) + W_s x \quad (6)$$

其中, W_s 表示 1×1 卷积操作,目的是调整特征图 x 的通道数,使其与隐层计算结果的维度相匹配,能有效减少参数规模.

字形识别模型的损失函数仍然选择多分类交叉熵,形式与式(3)、式(4)相同.模型训练时学习率设置为0.0001,并按照指数衰减法逐步减小学习率,迭代次数为14 000.

7 实验结果与分析

本次实验利用分布式爬虫技术,针对书法字典网站中的楷书和篆书类别,获取常用3500字对应的单字图片以及各书法家整幅作品集图片,作为训练模型和测试效果的数据,数据规模如表1所示.

表1 数据集规模表

书体	作品图片总数	预处理后单字种类数量	切割后单字图片数(张)
楷书	5872	3479	116340
篆书	4470	2700	34828

将数据集按照7:3的比例划分训练集和测试集,进行数据增强并实现书体识别和字型识别模型,在测试集上的测试效果如表2、表3所示.实验结果显示本文方法针对图像书法风格的识别率能达到85.53%,对楷书和篆书单字的识别率分别达到91.57%和81.70%.

能够准确识别出大部分图像的书法风格以及字形,并且识别一幅书法作品图片的平均时间为2.5 s.

为了进一步说明本文提出的算法在字形识别任务中的有效性,我们将本文算法与SIFT-MQDF方法^[2]进行对比,结果如表4所示.从表4可以看到,与SIFT-MQDF方法相比,本文算法的楷书字体识别率略低(相差0.23%),但本文算法的篆书识别率远高于SIFT-MQDF方法的79.10%(相差2.60%).值得注意的是,SIFT-MQDF方法所用的数据集为CADAL数字图书馆汉字图书子库^[2],字符背景单一,噪声少,而本文所采用的数据集为自建数据集,图片从各网站收集而来,样本数据存在噪声较多、图片分辨率低、背景复杂等问题,对算法的挑战度更大.因此,测试结果的数据表明,本文算法更能有效地提取书法字体特征,有利于提高识别的性能.

表2 书法风格识别测试结果统计表

识别风格种类	测试集数量	正确识别数量	错误识别数量	识别率(%)
楷书	325	556	94	85.53
篆书	325			

表3 字形识别测试结果统计表

单字风格	测试集数量	正确识别数量	错误识别数量	识别率(%)
楷书	593	543	50	91.57
篆书	82	67	15	81.70

表4 字形识别测试结果对比(%)

识别方法	楷书单字识别率	篆书单字识别率
SIFT-MQDF ^[2]	91.80	79.10
本文算法	91.57	81.70

8 结束语

本文针对手写书法字体识别困难的问题提出了书法字体识别算法,利用图像处理方法与深度神经网络技术对图像中的书法字体进行定位和识别,实验表明本文算法能够有效判断图像中的书法风格并识别出楷书和篆书字体,识别率分别是91.57%和81.70%,识别结果可为大众提供解读书法作品的参考.下一步的工

作是提高单字检测和分割的准确度,通过实现端到端的单字检测与识别,进一步提高识别书体风格和字形的精度,并且在未来增加可识别的字体种类,例如隶书和行书等多种字体风格.

参考文献

- 1 毛天骄. 书法字书体风格识别技术 [硕士学位论文]. 杭州: 浙江大学, 2014.
- 2 顾刚. 汉字识别关键算法研究与应用 [硕士学位论文]. 杭州: 浙江大学, 2017.
- 3 金连文, 钟卓耀, 杨钊, 等. 深度学习在手写汉字识别中的应用综述. 自动化学报, 2016, 42(8): 1125–1141. [doi: [10.16383/j.aas.2016.c150725](https://doi.org/10.16383/j.aas.2016.c150725)]
- 4 张福成. 基于卷积神经网络的书法风格识别的研究 [硕士学位论文]. 西安: 西安理工大学, 2018.
- 5 温佩芝, 姚航, 沈嘉炜. 基于卷积神经网络的石刻书法字识别方法. 计算机工程与设计, 2018, 39(3): 867–872. [doi: [10.16208/j.issn1000-7024.2018.03.047](https://doi.org/10.16208/j.issn1000-7024.2018.03.047)]
- 6 林媛. 中国书法字识别算法研究及应用 [硕士学位论文]. 杭州: 浙江大学, 2014.
- 7 Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 2818–2826.
- 8 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 9 雷鑫, 李俊阳, 宋宇, 等. 用于手写汉字识别的文本分割方法. 智能计算机与应用, 2018, 8(2): 126–128. [doi: [10.3969/j.issn.2095-2163.2018.02.029](https://doi.org/10.3969/j.issn.2095-2163.2018.02.029)]
- 10 苏敏. 古籍手写文字分割算法的研究与应用 [硕士学位论文]. 兰州: 兰州交通大学, 2018.
- 11 张一夫. 碑帖图像文字的分割与提取 [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2013.
- 12 吴相锦, 张忠林, 周生龙. 基于可变窗口的古代手写汉字多步切分方法. 计算机工程与设计, 2016, 37(4): 1102–1106. [doi: [10.16208/j.issn1000-7024.2016.04.049](https://doi.org/10.16208/j.issn1000-7024.2016.04.049)]
- 13 王序哲, 曹卫. 一种手写汉字投影切分方法的改进算法. 电脑知识与技术, 2011, 7(25): 6205–6206. [doi: [10.3969/j.issn.1009-3044.2011.25.055](https://doi.org/10.3969/j.issn.1009-3044.2011.25.055)]