

# 深度学习在手写汉字识别中的应用综述

金连文<sup>1</sup> 钟卓耀<sup>1</sup> 杨钊<sup>2</sup> 杨维信<sup>1</sup> 谢泽澄<sup>1</sup> 孙俊<sup>3</sup>

**摘 要** 手写汉字识别 (Handwritten Chinese character recognition, HCCR) 是模式识别的一个重要研究领域, 最近几十年来得到了广泛的研究与关注, 随着深度学习新技术的出现, 近年来基于深度学习的手写汉字识别在方法和性能上得到了突破性的进展. 本文综述了深度学习在手写汉字识别领域的研究进展及具体应用. 首先介绍了手写汉字识别的研究背景与现状. 其次简要概述了深度学习的几种典型结构模型并介绍了一些主流的开源工具, 在此基础上详细综述了基于深度学习的联机 and 脱机手写汉字识别的方法, 阐述了相关方法的原理、技术细节、性能指标等现状情况, 最后进行了分析与总结, 指出了手写汉字识别领域仍需要解决的问题及未来的研究方向.

**关键词** 深度学习, 手写汉字识别, 卷积神经网络, 回归神经网络, 长短时记忆模型, 层叠自动编码器

**引用格式** 金连文, 钟卓耀, 杨钊, 杨维信, 谢泽澄, 孙俊. 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

## Applications of Deep Learning for Handwritten Chinese Character Recognition: A Review

JIN Lian-Wen<sup>1</sup> ZHONG Zhuo-Yao<sup>1</sup> YANG Zhao<sup>2</sup> YANG Wei-Xin<sup>1</sup> XIE Ze-Cheng<sup>1</sup> SUN Jun<sup>3</sup>

**Abstract** Handwritten Chinese character recognition (HCCR) is an important research field of pattern recognition, which has attracted extensive studies during the past decades. With the emergence of deep learning, new breakthrough progresses of HCCR have been obtained in recent years. In this paper, we review the applications of deep learning models in the field of HCCR. First, the research background and current state-of-the-art HCCR technologies are introduced. Then, we provide a brief overview of several typical deep learning models, and introduce some widely used open source tools for deep learning. The approaches of online HCCR and offline HCCR based on deep learning are surveyed, with the summaries of the related methods, technical details, and performance analysis. Finally, further research directions are discussed.

**Key words** Deep learning, handwritten Chinese character recognition (HCCR), convolutional neural network, recurrent neural network, long-short term memory (LSTM), stacked auto-encoder

**Citation** Jin Lian-Wen, Zhong Zhuo-Yao, Yang Zhao, Yang Wei-Xin, Xie Ze-Cheng, Sun Jun. Applications of deep learning for handwritten Chinese character recognition: a review. *Acta Automatica Sinica*, 2016, 42(8): 1125–1141

由于在拍照文档、支票、表单表格、证件、邮政信封、票据、手稿文书等光学字符识别 (Optical

character recognition, OCR) 图像识别系统以及手写文字输入设备中的广泛应用前景, 自从上个世纪 80 年代以来, 手写汉字识别 (Handwritten Chinese character recognition, HCCR) 一直是模式识别的一个重要研究领域, 得到了学术界的广泛研究和关注<sup>[1–10]</sup>. 手写体汉字识别由于数据采集方式不同可以划分为脱机 (Offline) 手写体汉字识别和联机 (Online) 手写体汉字识别两大类. 联机手写汉字识别所处理的手写文字是书写者通过物理设备 (如数字笔、数字手写板或者触摸屏) 在线书写获取的文字信号, 书写的轨迹通过定时采样即时输入到计算机中. 而脱机手写文字识别所处理的手写文字是通过扫描仪或摄像头等图像捕捉设备采集到的手写文字二维图片. 由于识别的对象不同, 使得这两类手写识别技术所采用的方法和策略也不尽相同. 前者的识别对象是一系列的按时间先后排列的采样点信息,

收稿日期 2015-11-04 录用日期 2016-04-18

Manuscript received November 4, 2015; accepted April 18, 2016  
国家自然科学基金 (61472144), 广东省科技计划 (2014A010103012, 2015B010101004, 2015B010130003, 2015B010131004) 资助

Supported by National Natural Science Foundation of China (61472144), Guangdong Science and Technology Plan (2014A010103012, 2015B010101004, 2015B010130003, 2015B010131004)

本文责任编辑 刘成林

Recommended by Associate Editor LIU Cheng-Lin

1. 华南理工大学电子与信息学院 广州 510641 2. 广州大学机械与电气工程学院 广州 510641 3. 富士通研究开发中心有限公司信息技术研究部 北京 100190

1. School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641 2. School of Mechanical and Electric Engineering, Guangzhou University, Guangzhou 510641 3. Information Technology Laboratory, Fujitsu Research & Development Center Co., Ltd, Beijing 100190

而后者则是丢失了书写笔顺信息的二维像素信息, 由于没有笔顺信息, 加之由于拍照扫描设备在不同光照、分辨率、书写纸张等条件下, 数字化会带来一定的噪声干扰, 一般来说, 脱机手写文字识别比联机手写文字识别更加困难.

手写汉字识别是一个极具挑战性的模式识别及机器学习问题, 主要表现在:

1) 中文汉字字符级别的类别繁多, 仅 1980 年制定的国标 GB2312-80 编码就定义了多达 6 763 类常用汉字, 远远高出英语、拉丁语等通常只有几十类数目的西方语言. 然而此编码标准仍然不能满足汉字信息处理的需求, 所以在 2000 年发布的国标 GB18010 第一版 (GB18010-2000) 大幅度扩展了字符集, 定义了 27 533 类汉字编码标准, 并成为汉字信息处理产品必须遵循的强制性执行标准; 此后 2005 年发布的国标第二版 (GB18010-2005) 更是新增加了 42 711 个汉字和多种少数民族文字等字符, 收录的字符类别高达 70 244 类汉字.

2) 相比较于印刷体汉字, 手写体字符的书写随意性大, 缺乏规范性, 人们经常使用的横、竖、撇、捺、点等笔划容易变形, 如直笔划变弯, 折笔划变成圆弧, 短横、短竖变成点, 点变成捺等, 笔划的倾斜角、笔划的相对长度不稳定; 连笔、笔划不清, 笔划之间的连接关系可能发生变化, 不该连的连上了, 该连的反而没有连; 笔划不清、缺笔等现象时有发生. 这些都给识别造成了很大的困难; 另外, 不同书写人书写风格多样和存在明显差异, 出自不同书写者的同一类汉字在汉字字形、汉字结构都会有明显的差异; 中国人口众多, 可以说几乎每个人都有自己的书写风格, 各种不同风格之间可能存在较大差异; 手写汉字的风格迥异, 从楷书、行书到草书, 识别难度逐渐提高.

3) 中文汉字中存在很多相似字和易混淆的汉字对, 例如: “己-已”、“口-囗”、“汨-汨-汨”、“睢-睢”、“余-余”、“晴-晴”、“海-海”、“绞-绞”、“莱-莱”等, 这些高度相似的字符给计算机自动识别带来极大挑战.

4) 充分的训练数据及其多样性是设计和训练一个手写汉字识别系统的关键, 但目前针对大类别多风格的无约束手写数据库仍显不足. 数据库的采集和整理工作需要消耗大量人力物力来兼顾规模性和准确性. 北京邮电大学发布的 HCL2000 脱机手写数据库<sup>[11]</sup> 及国家 863 中文手写评测数据<sup>[12]</sup>, 都是比较规范书写的数据库; 相比而言, 目前能表现真实书写情况的手写单字数据库有中国科学院发布的 CASIA-OLHWDB1.0-1.2<sup>[10]</sup> 联机单字, CASIA-HWDB1.0-1.2<sup>[10]</sup> 文本行数据集, 以及华南理工大学发布的涵盖单字、词组、文本行、数字、字母、

符号等综合手写数据集 SCUT-COUCH<sup>[13]</sup>, 但上述两个较新的数据集类别局限于 GB2312-80 标准的 6 763 类, 并且数据总量仍然不够大.

一般而言, 传统的手写中文单字识别系统主要包括数据预处理、特征提取和分类识别三部分: 1) 数据预处理主要包括样本归一化<sup>[14-15]</sup>、平滑去噪、整形变换<sup>[16]</sup>、伪样本生成<sup>[17-19]</sup>、添加虚拟笔画 (对联机数据)<sup>[20-22]</sup> 等; 2) 特征提取部分可以分为结构特征和统计特征两种, 结构特征主要对汉字结构、笔画或部件进行分析来提取. 但对手写字符而言, 目前最好的特征基本上都是统计特征, 例如方向特征<sup>[23-25]</sup>. 对脱机 HCCR 而言, Gabor 特征<sup>[26]</sup> 及 Gradient 特征<sup>[27]</sup> 是目前比较好的两种方向特征提取方法; 对联机 HCCR 而言, 8 方向特征<sup>[24]</sup> 是目前最有效的特征之一; 3) 分类器最常用的模型包括改进的二次判决函数 (Modified quadratic discriminant function, MQDF)<sup>[9, 28]</sup>、支持向量机 (Support vector machine, SVM)<sup>[29]</sup>、隐马尔科夫模型 (Hidden Markov model, HMM)<sup>[30]</sup>、鉴别学习二次判决函数 (Discriminative learning quadratic discriminant function, DLQDF)<sup>[31]</sup> 和学习矢量量化 (Learning vector quantity, LVQ)<sup>[32]</sup> 等. 对于文本行识别<sup>[9-12]</sup>, 主要有基于切分策略和无切分策略的两种识别方法, 分别利用投影法、连通域分析法等方法对文本行进行字符分割, 利用单字分类器对分割好的字符<sup>[33-36]</sup>, 或利用滑动窗口按一定步长滑窗, 利用单字分类器对滑动窗口内的字符进行识别<sup>[34]</sup>, 结合统计语言模型, 在 Bayes 等学习框架下对整个文本行的上下文关系进行建模<sup>[35]</sup>, 从而得出整体行识别结果.

经过四十多年来研究学者的不懈努力, HCCR 取得了极大进展, 例如文献 [10] 中使用鉴别特征提取方法 (Discriminative feature learning, DFE) 和鉴别学习二次判决函数 (Discriminative learning quadratic discriminant function, DLQDF) 分类器, 在极具挑战的联机 and 脱机手写汉字数据集 CASIA-OLHWDB 和 CASIA-HWDB<sup>[10]</sup> 的几个不同子集上, 联机手写单字符识别取得的最好识别率分别是: 95.28 % (DB1.0, 4 037 类汉字)、94.85 % (DB1.1, 3 926 类汉字) 和 95.31 % (ICDAR 2013 Competition DB, 3 755 类汉字), 脱机手写单字符识别取得的最好识别率分别是: 94.20 % (DB1.0), 92.08 % (DB1.1) 和 92.72 % (ICDAR 2013 Competition DB). 但是, 无论是联机还是脱机手写识别, 仍然没有达到足够高的精度. 最近的研究测试表明<sup>[37]</sup>: 目前市场上的主流手写识别软件的识别率并没有达到很高的精度, 不少知名品牌的手写输入法软件的识别率达不到 90 %, 最好的系统的识别率也

仅仅为 95 % 左右, 许多软件系统支持的字符集很不齐全, 在文本行/单字或重叠/单字混合书写识别模式下, 不少系统的识别率急剧下降, 相关软件产品的识别性能体验亟待改善. 因此, 可以说手写汉字识别仍然是一个还没有完全解决的具有挑战性的研究课题.

近两年来, 基于传统的“预处理+特征提取+分类器”的手写汉字识别框架似乎并没有特别大的研究新进展, 很少看到在识别性能上获得突破性进展的研究报道. 但是, 随着深度学习的兴起, 我们注意到深度学习对手写汉字识别难题带来了新的活力和极其有效的解决方法, 例如自 2011 年起, 连续两届的 ICDAR (International Conference on Document Analysis and Recognition) 手写汉字识别比赛的获胜者都是采用基于深度学习或神经网络的方法<sup>[38-39]</sup>. 值得一提的是, 在 2013 年的 ICDAR 手写汉字比赛中<sup>[39]</sup>, 来自英国华威大学的 Graham 利用深度稀疏卷积神经网络 (Spatially-sparse convolutional neural network) 的方法<sup>[40]</sup>, 获得了联机手写汉字识别第一名, 其识别率大幅度提升到 97.39 %, 而来自富士通公司的团队采用改进的 CNN (Convolutional neural network) 网络, 获得了脱机手写汉字识别的第一名, 识别率高达 94.77 %<sup>[39]</sup>, 无论是联机还是脱机 HCCR, 上述基于深度学习的方法所取得的结果, 均大幅度领先传统的方法, 显示出深度学习的强大潜力. 目前手写识别中, 深度学习方法相比于传统方法, 能达到更高的识别率, 但是其训练和测试时间较长, 字典存储偏大等问题, 仍待进一步研究改进.

最近几年来随着深度学习的兴起及不断发展<sup>[41-43]</sup>, 特别是深度卷积神经网络 (CNN)<sup>[44-45]</sup>、深度置信网络 (Deep belief network, DBN)<sup>[41]</sup>、层叠自动编码器 (Stacked auto-encoder, SAE)<sup>[46]</sup>、深度递归神经网络 (Deep recurrent neural network, DRNN)<sup>[47]</sup> 等深度模型在图像识别及计算机视觉各领域的大量突破性成果的涌现<sup>[48-59]</sup>, 相关的方法也被逐步应用到手写体文字识别领域<sup>[40, 60-82]</sup>, 并取得了不少突破性发展. 针对相关的学术及技术进展情况, 本文主要综述及分析基于深度学习的手写体汉字识别的最新进展.

本文组织结构如下: 第 1 节对深度学习模型及一些主流的开源工具进行简要介绍; 第 2 节将综述基于深度学习的手写汉字识别方法; 第 3 节总结全文并对未来研究方向进行讨论及展望.

## 1 深度学习模型及相关开源工具简介

深度学习作为神经网络模型的新发展<sup>[43, 55]</sup>, 其历史可以追溯到上个世纪 40 年代<sup>[83]</sup>, 但严格意义

上深度学习的早期模型如 CNN, 起源于上世纪 80 年代末及 90 年代初 LeCun 的工作<sup>[44-45, 84]</sup>, 而深度学习真正受到学术界的重视是从 2006 年加拿大学者 Hinton 等在 *Science* 上撰文介绍了一种新的深度神经网络模型 DBN 及其训练方法开始的<sup>[41]</sup>, 此后, 随着相关的一些深度学习模型如 DNN 及 CNN 在语音识别<sup>[52]</sup> 及图像识别<sup>[48, 85]</sup> 等领域的突破性进展报道, 深度学习逐渐获得了学术界及工业界的广泛重视. 在计算机视觉及图像识别领域, 自从 2012 年之后, 涌现出许多新的模型训练方法及实际应用技术<sup>[86-89]</sup>, 深度学习也成功地被用来解决许多极具挑战性的问题, 在语音识别<sup>[52]</sup>、图像识别<sup>[44-45, 48-59, 85-86, 90]</sup>、人脸识别<sup>[91-92]</sup>、姿态识别<sup>[93]</sup>、手写识别<sup>[60-82, 94-95]</sup>、文本检测和识别<sup>[96-97]</sup>和自然语言处理<sup>[98-102]</sup> 等诸多领域均获得了极大成功.

深度学习是传统机器学习的一个崭新分支, 其概念源于人工神经网络的研究, 与传统的人工神经网络相比, 深度学习是含有多隐藏层的深度神经网络结构, 通过组合低层特征形成更加抽象的高层来表示属性类别或特征, 以更好地发现数据的有效特征表示. 一般而言, 深度学习网络是一种含有多个隐藏层的神经网络, 具有特定的结构和相应的训练方法. 典型的深度学习结构包括: 深度置信网络 (DBN)、S 层叠自动编码器 (SAE)、卷积神经网络 (CNN)、回归神经网络 (Recurrent neural network, RNN) 等, 下面我们将做简要介绍及回顾.

深度置信网络 (DBN) 是 Hinton 教授等于 2006 年提出的一种崭新的神经网络模型<sup>[41]</sup>, 文中创造性地提出了含有多隐藏层的神经网络的训练方法及无监督逐层预训练的思想, 为模型参数找到一个合理的初始解, 使得后续的监督学习方法能够找到最优解, 同时这种逐层无监督训练可以学习出一定的层次性特征. 值得说明的是, 随着深度学习算法的发展, 如非线性激活函数的改进<sup>[48, 103]</sup>、更好的正则化方式<sup>[48]</sup> 和参数初始化方式<sup>[104]</sup> 的提出, 训练深层网络已经不再困难, 逐层预训练对参数进行初始化的方法已很少被利用.

CNN 的最早雏形理念是由 Fukushima 在 1980 年首次提出<sup>[105]</sup>, LeCun 等将反向传播算法 (Back-propagation, BP) 用于卷积结构, 提出了层与层之间局部连接且具有多层结构的卷积神经网络模型<sup>[44-45, 84]</sup>. 该结构由一系列的卷积层 (Convolutional layer) 和聚合层 (Pooling layer) 交替组成, 每一层的输出可以看出是一组特征图 (Feature map). 2012 年, Krizhevsky 等<sup>[48]</sup> 将 CNN 加深, 并利用 ReLU 非线性激活函数、Dropout 等一系列新的训练技术, 获得了 2012 年 Ima-

geNet<sup>[90]</sup> 比赛第一名. ReLU 及 Dropout 可以说是近年来对 CNN 发展的两个重要技术, 简要介绍如下: 1) ReLU 非线性激活函数: 对于传统的神经网络来说, 一个神经元的输出  $f$  是它的输入  $x$  的函数形式为:  $f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$  或者  $f(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}}$ . Krizhevsky 等<sup>[48]</sup> 提出了一种非线性的激活函数  $f(x) = \max(0, x)$ , 命名为 ReLU. 对于深层网络, 由于 sigmoid 或者 tanh 函数在接近饱和区时变化缓慢, 导数趋于 0, 所以在反向传播时, 很容易会出现梯度弥散的现象, 从而无法完成深度网络的训练. 而 ReLU 在大于 0 的部分导数为其本身, 能很好地适应深层网络的训练, 并且有助于加速网络的收敛. 2) Dropout 技术: Dropout 是 Krizhevsky 等<sup>[48]</sup> 提出的深度网络训练的一种重要的正则化方法, 能防止深度网络在训练过程中的过拟合现象. Dropout 正则化方法是指以一定的概率 (一个乘性因子) 抑制某一个隐藏层神经元, 即让某一隐藏层的神经元输出为 0, 也就是说, 被抑制的神经元在前向传播上输出为 0, 而且不参加反向传播的过程. 所以每一次训练样本进入网络, 网络结构其实都可以看作为不一样的, 但网络结构又是彼此共享权值的. 而且, 通过 Dropout 这种方法, 可以避免一种特征必须依赖另一种特征的存在而存在的现象. 在测试的时候, 并不让隐藏层神经元抑制, 而是对所有的隐藏层神经元的输出值乘上一个乘性因子, 这也进一步体现了模型平均的思想.

受到 DBN 思想的启发, Ranzato 等于 2007 年提出了层叠自动编码机的深度学习结构<sup>[46]</sup>. SAE 与 DBN 的主要区别在于组成深度结构的基本单元由 RBM 变为自动编码器 (Auto-encoder, AE), 它是一种典型的无监督的两层神经网络结构, 通过输出对输入进行重建使得隐藏层为输入层的压缩或稀疏表示, 可以看成是对输入数据的特征学习.

RNN 是一种具有储存记忆功能的神经网络<sup>[94]</sup>, 与前面介绍的 DBN、CNN 及 SAE 相比, 一个显著的特点就是网络的输出结果不仅仅依赖于当前的输入, 且与过去的输入相关联, 甚至与网络的未来输出也相关 (例如双向递归神经网络模型). 同样, RNN 可以使用反向传播算法进行训练, 这种跨时间的反向传播算法也被称为沿时间反向传播 (Back propagation through time, BPTT)<sup>[106]</sup>. 然而这种典型的 RNN 不具有长时间记忆的能力, Hochreiter 及 Schmidhuber 在 1997 年对 RNN 进行改进, 提出了长短时记忆模型 (Long short term memory, LSTM)<sup>[47]</sup>. 其主要思想是对 RNN 增加输入门、输出门、重置激活门, 能够实现对隐藏层输入、输出激活以及递归控制, 使得神经元在长时间范围内存储并传递信息.

近年来还涌现出不少别的深度学习模型, 其

中一类值得关注的模型是深度强化学习 (Deep reinforcement network, DRN)<sup>[107-108]</sup>, 例如 Google Deep Mind 团队 2015 年提出的深度 Q 网络 (Deep Q network, DQN)<sup>[108]</sup>, 其在模拟人的智能及在线强化学习方面表现出优异的性能. 但由于在手写汉字识别领域, 还鲜有深度强化学习网络的成功应用报道, 所以在本文中我们不对 DRN 进行详细介绍.

深度学习近年来的飞速发展及其在众多领域的广泛应用, 离不开许多优秀的实验仿真平台. ConvNet 是由 Krizhevsky 提供的基于 GPU 的开源 CNN 开源库, 在 2012 年 ImageNet 大规模视觉识别比赛 (Large Scale Visual Recognition Challenge 2012, ILSVRC2012)<sup>[90]</sup> 中获得了第一名, 此后还推出了支持多 GPU 的版本 Cuda-ConvNet2<sup>[109]</sup>; 近三年来, 涌现出了大量优秀的开源深度学习仿真工具, 例如 Caffe、Theano、Touch、TensorFlow 等. 表 1 总结了目前一些有代表性的深度学习开源工具及下载地址.

## 2 基于深度学习的手写汉字识别

传统的手写汉字识别一般包括预处理 (例如归一化等)、特征提取、特征降维、分类器设计等环节, 但卷积神经网络 (CNN) 的提出, 使得设计一个端到端 (End-to-end) 的手写字符识别系统成为可能, 而无需繁杂的预处理及特征提取、特征降维环节. 此外, 有研究学者发现, 尽管端到端的 CNN 手写识别方法能获得比传统方法更好的性能, 但结合传统领域知识, 仍然能使得常规的 CNN 手写识别系统性能得到进一步提升. 此外, 针对手写汉字识别的特征, 近年来还提出了不少改进的 CNN 训练方法. 下面我们将逐一进行详细介绍.

### 2.1 基于 CNN 的端到端的识别方法

CNN 在上世纪 90 年代提出来后, 虽然在手写西文 (特别是英文) 识别中得到了极大应用推广<sup>[44-45, 84, 110-111]</sup>, 如在英文手写识别领域, LeCun 等在 1998 年提出的 LeNet5 CNN 模型<sup>[45]</sup> 在 MNIST 上的识别率已经高达 99.05%, 加上变形样本训练后识别率进一步可提升到 99.2%; 2003 年, 微软研究院的 Simard 等<sup>[111]</sup> 引入弹性变形 (Elastic distortion) 及仿射变形 (Affine distortion) 两种数据增加 (Data argumentation) 技术, 采用类似 CNN 的网络结构, 在 MNIST 上, 获得了 99.6% 的优异性能, 大幅度领先于包括 SVM、Boosting、多层感知器等在内的机器学习方法. 但 CNN 成功应用到手写中文识别的文献一直报道很少, 这主要源于手写中文识别的类别数及训练样本庞大, 用常规的 CPU 实验仿真平台很难在几天甚至几周内完成网络结构的训练. 直到 2011 年瑞士 IDSIA (Istituto Dalle



表 1 目前一些主流的深度学习开源仿真工具及其下载地址

Table 1 Some mainstream deep-learning open source toolboxes and their download address at present

工具名称	说明及备注	下载地址
Caffe <sup>[112]</sup>	UC Berkeley BVLG 实验室发布的深度学习开源工具, 是目前使用最为广泛的深度学习实验平台之一	<a href="https://github.com/BVLG/caffe">https://github.com/BVLG/caffe</a>
Theano <sup>[113-114]</sup>	基于 Python 语言的深度学习开源仿真工具	<a href="https://github.com/Theano/Theano">https://github.com/Theano/Theano</a>
Torch <sup>[115]</sup>	基于 Lua 脚本语言的工具, 支持 iOS、Android 等嵌入式平台	<a href="http://torch.ch/">http://torch.ch/</a>
Purine <sup>[116]</sup>	支持多 GPU, 提供线性加速能力	<a href="https://github.com/purine/purine2">https://github.com/purine/purine2</a>
MXNet <sup>[117]</sup>	由百度牵头组织的深度机器学习联盟 (DMCL) 发布的 C++ 深度学习工具库	<a href="https://github.com/dmlc/mxnet">https://github.com/dmlc/mxnet</a>
DIGITS <sup>[118]</sup>	由 NVIDIA 公司集成开发发布的一款基于 Web 页面的可视化深度学习仿真工具, 支持 Caffe 及 Touch 工程代码	<a href="https://github.com/NVIDIA/DIGITS">https://github.com/NVIDIA/DIGITS</a>
ConvNet <sup>[119]</sup>	最早的支持 GPU 的 CNN 开源工具之一, ILSVRC2012 比赛第一名提供的代码	<a href="https://code.google.com/p/cuda-convnet/">https://code.google.com/p/cuda-convnet/</a>
Cuda-ConvNet2 <sup>[109]</sup>	支持多 GPU 的 ConvNet	<a href="https://github.com/akrizhevsky/cuda-convnet2">https://github.com/akrizhevsky/cuda-convnet2</a>
DeepCNet <sup>[120]</sup>	英国 Warwick 大学 Graham 教授发布的开源 CNN 仿真工具, 曾获 ICDAR 2013 联机手写汉字识别竞赛第一名	<a href="https://github.com/btgraham/SparseConvNet">https://github.com/btgraham/SparseConvNet</a>
Petuum <sup>[121]</sup>	CMU 发布的一款基于多 CPU/GPU 集群并行化分布式, 机器学习开源仿真平台除了支持深度学习的常用算法之外, 还提供很多传统机器学习算法的实现. 可部署在云计算平台之中	<a href="https://github.com/petuum/bosen/wiki">https://github.com/petuum/bosen/wiki</a>
CURRENT <sup>[122]</sup>	支持 GPU 的回归神经网络函数库	<a href="http://sourceforge.net/projects/currentnt/">http://sourceforge.net/projects/currentnt/</a>
Minerva <sup>[123]</sup>	深度机器学习联盟 (DMCL) 发布的支持多 GPU 并行化的深度学习工具	<a href="https://github.com/dmlc/minerva">https://github.com/dmlc/minerva</a>
TensorFlow <sup>[124]</sup>	谷歌发布的机器学习可视化开发工具, 支持多 CPU 及多 GPU 并行化仿真, 支持 CNN、RNN 等深度学习模型	<a href="https://github.com/tensorflow/tensorflow">https://github.com/tensorflow/tensorflow</a>
DMTK <sup>[125]</sup>	微软发布的一套通用的分布式深度学习开源仿真工具	<a href="https://github.com/Microsoft/DMTK">https://github.com/Microsoft/DMTK</a>

Molle di Studi sull'Intelligenza Artificiale) 实验室的学者才第一次采用 GPU (GTX 580) 训练 CNN, 实现了其在大类别 (1000 类) 的手写汉字识别应用<sup>[126]</sup>. 文献 [75] 提出了采用多个 CNN 集成的方法 (不同的输入尺度), 在 NIST SD19 数据集上 (含 80 万样本), 取得了很好的结果 (89.12%), 在 MNIST 上的识别率为 99.72% (错误率 0.23%); 基于该方法, IDSIA 实验室提交 ICDAR 2011 脱机和联机手写汉字识别竞赛的系统分别获得了第一名 (识别率 92.12%) 和第四名 (识别率为 93.01%)<sup>[38]</sup>. 随后 2012 年, IDSIA 实验室提出了一种多列 CNN 模型 (Multi-column deep neural networks, MCDNN)<sup>[127]</sup>, 其基本思想是分别利用 GPU 训练多个 CNN 网络 (每个 CNN 网络含 4 个卷积层、4 个 Pooling 层、1 个全连接层), 然后用对所有 CNN 的输出进行简单的平均集成, 对联机及脱机手写样本均转换为图像输入到 CNN 网络进行训练, 不做任何特征提取或特征选择等方面的工作, CNN 的

输出直接作为最终的识别结果, 是一种典型的端到端 (End-to-end) 的解决方案. MCDNN 在脱机及联机中文识别均取得了当时最先进水平的识别结果<sup>[127]</sup>, 在 CASIA-OLHWDB1.1 联机手写中文数据集上, MCDNN 在只利用联机手写轨迹的二值图片信息 (忽略时序信息) 的情况下, 将错误率大大降低, 从 7.61% (当时传统方法取得的最好结果) 下降至 4.61%, 在脱机手写体汉字识别中取得了比 ICDAR 2011 脱机手写汉字竞赛第一名更好的识别性能 (93.50%). 另外, 在 ICDAR 2013<sup>[39]</sup> 中文脱机和联机手写汉字识别的比赛中, 取得第一名的队伍的方法均基于 CNN 模型.

此后, 基于 CNN 的深度学习方法逐渐吸引了手写中文识别领域学者们的重视, 近三年来不断有新的突破性结果报道. 下面将进一步介绍结合领域知识及其他改进技术的 CNN 手写体汉字识别方法.

## 2.2 结合领域知识的 CNN 识别方法

虽然以 MCDNN 为代表的端到端的 CNN 模型

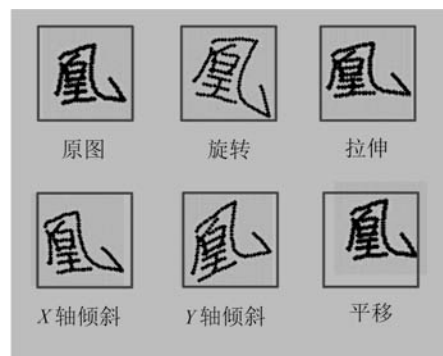
设计简单,也取得了优异的识别性能,但端到端的方法一个不足之处在于其没有很好地利用要手写文字的一些先验领域知识,例如联机时序信息、领域相关的数据生成技术、方向变换特征等,而这些领域信息是无法通过 CNN 网络自动学习而获取的.下面将介绍近年来传统领域知识和 CNN 相结合的手写汉字识别方法.

1) 数据生成技术: 为了避免在训练过程中出现过拟合 (Over fitting) 以及提升 CNN 模型的识别性能,除了采用 Dropout 等一些经典的方法之外,获取**充足的训练样本**是保障 CNN 等模型获得高性能的前提.但很多时候我们获得的训练样本是十分有限的,例如像 CASIA-HWDB/CASIA-OLHWDB 这样优秀的数据集,其书写者仍然有限 (不到 1080 人),且采集地区有限,不足以覆盖全国汉字书写者的统计分布,因此,**数据增加 (Data augmentation)** 技术是一种非常重要的提升 CNN 系统鲁棒性及推广能力的技术.早在文献 [45] 中 LeCun 等就提出了采用平移、尺度缩放、旋转、水平及垂直拉伸变形 (Squeezing, Shearing) 等方法对数据进行变形,有效地提升了识别性能.文献 [111] 在仿射变换的基础上,进一步提出了**弹性变形 (Elastic distortion)** 等两种数据生成技术,利用 CNN 在 MNIST 获得了 99.60% 的优异性能,文献 [16] **利用三角函数构成的非线性函数,提出了针对手写汉字的 24 种整形变换方法,也可看成是一种有效的手写伪样本数据生成方法<sup>[91]</sup>**;文献 [17-19] 提出了多种不同的手写伪样本生成技术,也被证明对识别性能的提升是有效的.文献 [128] 提出了 19 种手写伪样本生成方法,包括仿射变换、局部弹性形变、非线性拉伸、图像加噪、模糊化、对比度变化等,针对 NIST-19 手写数据集,生成了超过 8.19 亿的巨大样本,有效解决训练样本数据量不足问题.除了在手写字符识别领域之外,在自然场景文字检测与识别领域,数据生成技术也常常被证明是十分重要且有效提升深度学习模型性能的技术手段<sup>[96, 128]</sup>.

文献 [129] 采用更深的 CNN 网络 (15 层,其中 9 个卷积层),同时结合三维随机变形技术进行了大规模的数据生成,除文字模式图像上的 X 坐标、Y 坐标外,将各像素的灰度值作为 Z 坐标的参数,即可生成各种变形模式,模拟光影以及笔画轻重的变化,并借鉴 GoogLeNet 在网络中间层引入多个输出层以进行多层次监督学习的实现,在 ICDAR 2013 脱机手写汉字识别比赛测试集上用 5 个 CNN 网络投票的结果为 96.79%,是迄今为止在该数据集上公开报道的最好的识别结果.此外文献 [129] 的方法在 MNIST 手写数字上的识别结果为 99.82%,这也是迄今为止我们所知的在该数据集上公开报道的最好

的识别结果. Yang 等进一步基于 CNN 的分类器设计之中<sup>[61]</sup> 引入了伪样本生成技术<sup>[16]</sup>,应用于联机手写汉字识别,通过实验证实伪样本技术弥补了在 DCNN 训练中数据不足、泛化能力不够的问题.

图 1 列举了几种常用的数据增广技术所生成的样本.可见在手写识别问题中,只要在不改变原字内部结构关系的情况下,能够丰富样本空间的样本个数的数据增广技术都值得考虑使用.值得注意的是,在使用数据增广技术时,**需要控制样本变换的程度,使得生成样本尽量符合实际样本分布.**



(a) 文献 [19] 采用的样本生成方法

(a) Sample distortion methods mentioned in [19]



(b) 文献 [16] 采用的样本生成方法

(b) Sample distortion methods mentioned in [16]

图 1 几种常用的手写汉字数据增广技术示意图

Fig. 1 The influences of the controller parameters on the tracking errors

2) 方向变换特征: 在文献 [40] 中,除了提出了稀疏 CNN 模型之外, Graham 引入一种新的领域知识路径积分特征 (Signature of path)<sup>[130-131]</sup> 作用于 CNN 的输入层,被证明是一种非常有效的联机时序信息特征提取方法,由于该路径积分特征图的引入,使得其在 CASIA-OLHWDB 数据集上的识别结果明显优于 MCDNN 的方法,并且取得了 ICDAR 2013 年联机手写中文汉字识别比赛的第一名. Graham 在文中还验证了路径积分特征的有效

性, 在 CASIA-OLHWDB1.1 数据集上, 若不采用路径积分特征图, CNN 最佳错误率为 5.12%, 而引入路径积分特征图之后, 错误率大幅降低到 3.82%, 相对错误率下降了 25.39%, 充分说明了路径积分特征图这一领域知识的有效性。

值得一提的是, 路径积分特征图是近年来提出的一种很好的描述联机时序信号的特征, 文献 [40, 60–61] 将之引入与 CNN 结合, 在联机手写中文识别中取得了很好的效果, 文献 [62–63] 应用其来进行书写者鉴别, 相比传统方法, 大幅度提升了识别精度, 在 NLPR 手写签名数据集上取得了迄今为止最好的识别率。路径积分特征是通过路径的多重累积积分获得多阶特征, 文献 [131] 中给出了路径积分的具体数学定义描述, 文献 [40] 中则举例说明了路径积分的计算过程, 文献 [62] 对其在 handwritten 汉字的提取进行了可视化, 如图 2 所示。从物理意义上看, 第 0 阶路径积分代表轨迹的二值图像信息, 第 1 阶代表轨迹采样点的位移信息, 第 2 阶代表轨迹内采样点的曲率信息, 以此类推。相比较之前在联机手写汉字识别被广泛采用八方向特征<sup>[24]</sup>, 可以看到路径积分特征图提取的方向信息及梯度变化信息更为丰富完整。

此后, Yang 等进一步在基于 CNN 的分类器设计之中加入非线性归一化<sup>[14–15]</sup>、虚拟笔画技术<sup>[20–22]</sup> (对联机数据)、八方向特征<sup>[24]</sup> 和路径积分特征图<sup>[40]</sup> (对联机数据) 等相关领域知识作为先验知识, 进一步提高联机手写汉字的识别性能<sup>[61]</sup>。通过实验得到以下观点: 1) 虚拟笔画技术在 DCNN 中能够很好地权衡真实笔画和虚拟笔画之间的权重分配, 明显改进识别结果; 2) 8 方向特征和路径积分特征的互补可以取得更好的结果。

在脱机手写汉字方面, 文献 [77] 借鉴在计算机视觉领域参加 2014 年 ImageNet 大规模图像识别竞赛 (ILSVRC 2014) 获得了物体识别第一名的 GoogLeNet 的网络结构, 对其进行了精简和改进, 并引入脱机手写汉字识别领域中常用的几种领域特征提取方法如多尺度多方向的 Gabor 特征、八方向梯度特征等到 GoogLeNet, 提出了一种改进的 HCCR-GoogLeNet 模型, 在 ICDAR 2013 比赛数据集上取得了 96.74% 的识别率, 首次在该数据库上超过了人类的识别水平 (96.13%)<sup>[39]</sup>, 该工作同时表明手动提取的汉字特征图作为先验知识嵌入到 CNN 网络之中, 能够帮助 CNN 网络更好学习到汉字的辅助特征, 有效提高网络的识别性能。

### 2.3 其他改进的 CNN 识别方法

在 handwritten 汉字识别领域, 为了进一步提升深度学习模型的性能, 研究人员除了加大模型深度和宽度及将传统领域知识与深度模型相结合之外, 还对深度学习的网络结构和训练方法进行了各种改进。CNN 中常见最大值聚合 (Max-pooling, MP) 操作的大小和步长一般都是整数 (例如大小为  $3 \times 3$ , 步长为 2), Graham<sup>[132]</sup> 对此提出了更一般化的分数最大值聚合方法 (Fractional max-pooling, FMP)。FMP 使用的聚合区域大小是 1 到 2 之间的一个随机分数, 因此需要比以往更多的隐层来确保特征图缩小到合适大小, 因此 FMP 可以防止传统 MP 中卷积层特征图快速缩小造成的信息丢失过快的问题。除此之外, FMP 因为引入了随机聚合区域大小, 因此在测试阶段可以对每个样本进行多次不同区域大小的 FMP 从而得到多个测试结果, 集成后可以提高识别率。实验证明<sup>[132]</sup>, 使用 FMP 思想设计的单个

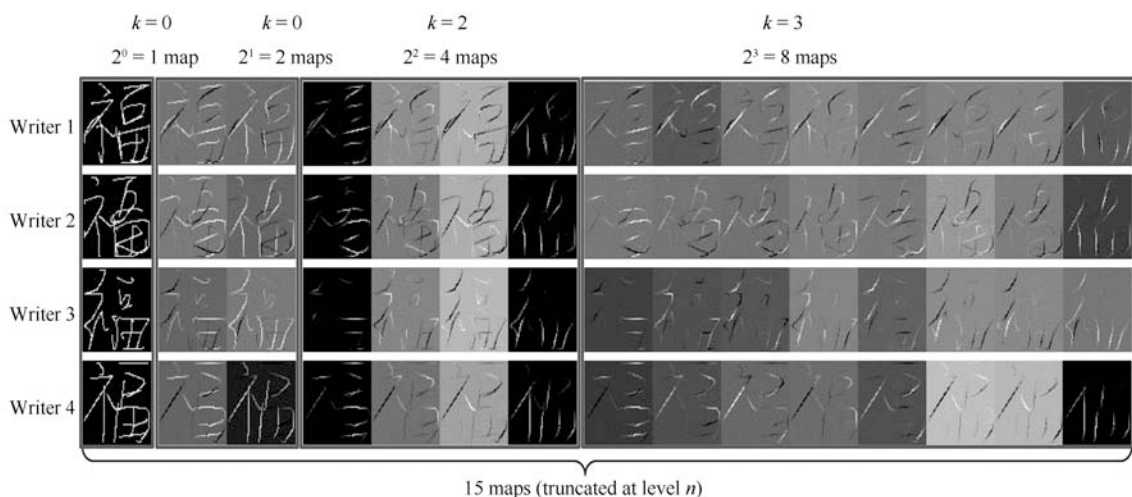


图 2 手写汉字的路径积分特征图可视化

Fig. 2 Path signature feature map visualization of handwritten Chinese characters

网络在 CASIA-OLHWDB1.1 上单次测试错误率为 3.26%，在不需增加模型存储量情况下，多次（12 次）测试集成结果能有效降低 2.97%。

除此之外，研究者对深度模型的样本训练方式也做出改进。传统最简单有效的样本训练方式是使用 Mini-batch 模式，每个 Mini-batch 内的样本是从总样本集合中随机抽取，每个样本被抽取的机会均等。Yang 等<sup>[61]</sup>从心理学对人的记忆遗忘规律的研究中受到启发，提出 DropSample 训练技术。因为在大规模的手写样本中，CNN 理解每个样本的难易程度不同，在训练的后期大量已充分学习的样本无需反复训练，而错误标记的样本或严重噪声干扰的样本应该从训练集中剔除，对还未很好学习理解的样本则需要频繁进入网络训练。DropSample 训练方法给每个样本设置一个配额，根据 CNN 识别输出的置信度为依据<sup>[82]</sup>，对当前网络输出的置信度高低来动态调整该样本被选入 Mini-batch 训练的概率，使得网络训练更加高效；DropSample 对未很好掌握的样本反复学习，能够提高网络的泛化能力。由于 DropSample 作为一种深度模型训练方法，只影响每次样本抽取的概率，因此能够很好地和其他神经网络结构或改良技术相结合，从而进一步提高性能。文献 [61] 在多个数据集上实验，使用 DropSample 训练并集成多种领域知识的 CNN，最终单个网络在 CASIA-OLHWDB1.1 上取得 3.43% 的错误率，9 个不同领域知识的网络集成的结果为 2.94%。在 ICDAR 2013 联机手写中文竞赛的数据集上，DropSample 训练 CNN 集成后最好结果是 2.49%，是目前我们所知的公开文献报道中最好的结果。另外，文献 [76] 提出了两种改进的 CNN 模型用于脱机手写汉字识别：Relaxation CNN (R-CNN) 及 Alternately training CNN (ART-CNN)，R-CNN 网络与传统的 CNN 相比，主要区别在于其不进行卷积层的权值共享，这样做的一个潜在动机是为了使每个相应的神经元都能独立地学习有效的特征，并且防止过拟合，但此网络的缺点是将导致存储量明显增大；而 ART-CNN 在 R-CNN 的基础上，进一步提出了一种新的动态调整学习率的训练策略，在训练过程中，每隔  $N$  次（例如 10 次）随机将某个权值矩阵学习率设置为 0，从而对相应的权值不进行学习和更新，采用此策略可以极大提升训练速度。文献 [76] 利用 10 个 ART-CNN 网络进行集成，在 ICDAR 2013 比赛数据集上，获得了 96.06% 的性能。此外，ART-CNN 网络在 MNIST 手写数字识别实验上，也取得了 0.25% 的较低的错误率。此后，富士通研究开发中心进一步改进了 R-CNN 模型，并结合大数据生成基础，单个网络的识别率已经提升至 96.79%<sup>[129]</sup>。

表 2 总结对比了不同的深度学习方法在联机手写中文单字符数据集 CASIA-OLHWDB1.1 上的识别结果。表 3 总结对比了几种典型的深度学习方法在 CASIA-OLHWDB1.0、CASIA-OLHWDB1.1、ICDAR 2013 联机手写中文单字符识别竞赛数据集识别率情况。表 4 为不同方法在 ICDAR 2013 脱机手写汉字竞赛集上的识别性能对比。由于某些方法并没有在文献中提及其训练时间和测试速度，所以在比较中更关注识别性能和字典大小的对比。

表 2 不同方法在 CASIA-OLHWDB1.1 联机手写中文单字符数据集上的识别结果对比

Table 2 Comparison with different methods on the CASIA-OLHWDB1.1			
方法	准确率 (%)	伪样本变形	模型集成 (模型数量)
传统最佳方法: DFE+DLQDF <sup>[10]</sup>	94.85	×	×
HDNN-SSM-MCE <sup>[66]</sup>	89.39	×	×
MCDNN <sup>[127]</sup>	94.39	✓	✓(35)
DeepCNet <sup>[40]</sup>	96.42	✓	×
DeepCNet-8 方向直方图特征 <sup>[40]</sup>	96.18	✓	×
DCNN (4 种领域知识融合) <sup>[60]</sup>	96.35	✓	×
HSP-DCNN (4 种领域知识集成) <sup>[64]</sup>	96.87	✓	✓(8)
DeepCNet-FMP (单次测试) <sup>[132]</sup>	96.74	✓	×
DeepCNet-FMP (多次测试) <sup>[132]</sup>	97.03	✓	✓(12 test)
DropSample-DCNN <sup>[61]</sup>	96.55	✓	×
DropSample-DCNN (集成) <sup>[61]</sup>	<b>97.06</b>	✓	✓(9)

表 3 不同深度学习方法在 CASIA-OLHWDB1.0-1.1 以及 ICDAR 2013 竞赛数据集上的识别结果 (%)

Table 3 Comparison with different methods on the CASIA-OLHWDB1.0-1.1 and ICDAR 2013 Online CompetitionDB (%)			
	CASIA-OLHWDB1.0	CASIA-OLHWDB1.1	ICDAR 2013 竞赛数据集
MQDF 传统方法 <sup>[10]</sup>	95.28	94.85	92.62
MCDNN <sup>[127]</sup>	94.39	—	—
DeepCNet <sup>[40]</sup>	—	96.42	97.39 <sup>1</sup>
DropSample-DCNN <sup>[61]</sup>	96.93	96.55	97.23 <sup>1</sup>
DropSample-DCNN (集成) <sup>[61]</sup>	<b>97.33</b>	<b>97.06</b>	<b>97.51</b>

<sup>1</sup> DeepCNet 模型参数大约为 590 万个，DropSample-DCNN 的模型参数为 380 万个。

2.4 基于其他深度模型的识别方法

拥有多个隐层的全连接神经网络一般可以统称为 DNN，例如含有多个隐藏层的多层感知器神经网络，传统的多层神经网络的训练容易过拟合、容易陷入局部最优，而且随着层数加深会出现梯度弥散等问题。2014 年，Du 等<sup>[65]</sup>将传统方法和 DBN 方法结合，使用了三种分类器：原型距离、DNN 和



Bottleneck 特征 (Bottleneck feature, BNF) 分类器. 原型分类器由聚类算法构建, 使用基于最大样本分离边界 (Sample separation margin, SSM) 的最

表 4 不同深度学习方法及部分典型的传统方法在 ICDAR 2013 脱机手写汉字竞赛集上的识别性能

Table 4 Comparison with different traditional and deep-learning based methods on ICDAR 2013 Offline CompetitionDB

方法	Top1 (%)	Top5 (%)	Top10 (%)	模型存储量
HCCR-Gradient-GoogLeNet <sup>[77]</sup>	96.28	99.56	99.80	27.77 MB
HCCR-Gabor-GoogLeNet <sup>[77]</sup>	96.35	99.60	99.80	27.77 MB
HCCR-Ensemble-GoogLeNet <sup>[77]</sup> (average of 4 models)	96.64	99.64	99.83	110.91 MB
HCCR-Ensemble-GoogLeNet <sup>[77]</sup> (average of 10 models)	<b>96.74</b>	<b>99.65</b>	<b>99.83</b>	277.25 MB
CNN-Fujitsu <sup>[39]</sup>	94.77	—	99.59	2460 MB
MCDNN-INSIA <sup>[74]</sup>	95.79	—	99.54	349 MB
MQDF-HIT <sup>[39]</sup>	92.61	—	98.99	120 MB
MQDF-THU <sup>[39]</sup>	92.56	—	99.13	198 MB
DLQDF <sup>[39]</sup>	92.72	—	—	—
ART-CNN <sup>[76]</sup>	95.04	—	—	51.64 MB <sup>2</sup>
R-CNN Voting <sup>[76]</sup>	95.55	—	—	51.64 MB <sup>2</sup>
ATR-CNN Voting <sup>[76]</sup>	96.06	—	—	206.56 MB <sup>2</sup>
MQDF-CNN <sup>[78]</sup>	94.44	—	—	—
Multi-CNN Voting <sup>[129]</sup>	<b>96.79</b>	—	—	—

<sup>2</sup> 根据文献 [76] 给出的模型参数 (CNN 层数、各层卷积核大小及数量、聚合层大小及数量、全连接数量), 按照每个参数以浮点数存储 (占用 4 个字节) 方式推算而得.

小分类误差 (Minimum classification error, MCE) 准则<sup>[67]</sup> 进一步改进, 是一个小型且高效的分类器. DNN 分类器则是对每个样本后验概率进行建模, 它将每两层连续的神经元作为一个 RBM, 从而使用 RBM 的分层预训练方法<sup>[68]</sup>, 之后进行有监督的精细微调训练. BNF 分类器则是用到了训练好的 DNN 模型中拥有较小神经元个数的一层 (Bottleneck layer) 作为降维后的特征, 输入原型分类器训练. 这三种分类器集成, 可以在提高大类别联机手写汉字识别率的同时, 有效控制模型大小和测试时间. 文献 [66] 提出了将训练 DNN 作为特征转换模型, 在此基础上进一步提出多级 DNN (Hierarchical DNN, HDNN), 采用传统的特征提取方法, 输出单元分解为多个子空间, 每一个用 DBN 进行训练. 一般的网络集成方法是训练多个完整的网络后对结果进行集成, HDNN 则是将输出层的向量分成多个子向量, 每个子向量由一个基本的 DNN 单独训练, 这样兼顾网络的深度和宽度, 使得每个网络的训练更加容易, 在 CASIA-OLHWDB1.1 上最好的识别结果为 89.39%. 与文献 [66] 类似的方法被用来进行手写汉字书写者自适应<sup>[69]</sup>, 采用的深度 DBN

以 RBM 作为基本单元, 从 DBN 的最后一层的输出作为特征, 分类器采用 SSM-MCE 训练多原型距离分类器, 在 15 000 个超大类别的手写中文识别实验中, 未进行书写者自适应前, 分类器识别率约为 81.21%; 进行书写者自适应学习, 采用风格转移学习的方法, 最高识别率可达到 91.86%. 由于文献 [65, 69] 并没有公开其实验的数据库, 也未在公开的数据集如 CASIA-OLHWDB 上进行实验, 所以其方法与别的方法暂时不具备可比性.

两个主流的深度神经网络 CNN 和 DBN 相比较而言, CNN 不需要预训练, 则相比于 DBN 的逐层预训练要省时, 同时 CNN 因为加入卷积操作, 能够更好处理图像信息, DBN 则主要处理一维的数据, 因此在手写体汉字识别中, CNN 比 DBN 略胜一筹. 此外, SAE 等其他深度模型在大类别手写中文字识别中的成功应用报道尚不多见.

## 2.5 基于 RNN/LSTM 的文本行识别方法

如前文所述, 采用深度学习等技术之后, 目前手写单字符识别问题已经在某种程度上获得了较好的解决, 但相对单字符识别而言, 含序列信息的手写文本行识别仍然是此领域还未解决的难点问题之一. 针对此问题, 一个有极大潜力的解决办法是应用深度学习方法中回归神经网络 (RNN) 模型, 特别是 LSTM 和 BLSTM (Bidirectional long short term memory) 等模型, 由于对序列数据有很好的建模能力, 因此它们更适合于解决诸如包含时序先后顺序信息的文字行识别问题. 基于 LSTM-RNN<sup>[47]</sup> 和 CTC (Connectionist temporal classification)<sup>[133]</sup> 的方法在英文、拉丁文等西方语言的脱机文本行识别中取得了很好的结果<sup>[133-137]</sup>, 包括英文<sup>[70]</sup>、孟加拉文<sup>[72]</sup> 和蒙古文<sup>[73]</sup> 等语言的联机手写识别中也已经取得了很多成功应用的报道, 但在大类别手写汉字识别中的应用报道很少. 文献 [138] 是首个将 LSTM-RNN 和 CTC 训练框架应用于大类别 (3 755 类) 脱机中文文本行识别的工作, 提出了多维 LSTM 回归神经网络 (Multi-directional long short term memory-recurrent neural network, MDLSTM-RNN) 的方法和 CTC 训练框架, 来进行端到端的手写中文文本行识别, 并结合语言模型的无切分策略方法进行中文文本行识别. 未采用语言模型时, 在 ICDAR 2013 脱机手写体文本行识别竞赛数据集上 CER 指标最好可达到 10.6%, AR 指标达到 90.4%, 已基本上达到目前此领域先进水平. 基于 MDLSTM 与传统方法在 ICDAR 2013 竞赛数据集上的识别结果对比如表 5 所示.

此外, 近年来我们看到了越来越多应用 RNN 及 LSTM 的方法并应用到文档分析与识别多个其

他领域,包括语言分类<sup>[139]</sup>、文档版面分析<sup>[140-141]</sup>、中英文语言模型建模<sup>[98-102]</sup>、自然场景文字检测及识别<sup>[142-143]</sup>等方面.其中值得一提的是近年来有不少学者把 RNN 与 CNN 结合,成功应用于解决 OCR 文字识别问题.例如文献 [142] 是利用滑动窗口方法,将滑动窗口区域的字符输入 CNN,得到有序的高度抽象的特征序列,将得到的序列作为 RNN 的输入,得到最终识别结果.文献 [143] 先利用 CNN 对整个输入图像提取特征,将最后一个卷积层特征图的不同维度的同一列组成一个特征矩阵,不同列构成了有序的特征矩阵序列,作为 RNN 的序列输入.上述两种 CNN+RNN 相结合的方法,都是可训练的端到端的深度学习模型,我们期待类似的网络结构在不久的将来也能被应用于解决大类别手写文本行识别问题并取得突破性进展.

表 5 不同研究方法在 ICDAR 2013 Offline Text CompetitionDB 数据对比记录表 (%)

Table 5 Comparison with different methods on the ICDAR 2013 Offline Text CompetitionDB (%)

方法/系统名称	CR	AR	CER
HIT-MQDF+LM <sup>[39]</sup>	88.8	86.7	13.3
THU-MQDF+DP <sup>[39]</sup>	86.1	83.6	<b>6.4</b>
MQDF+Multiple Contexts <sup>[35]</sup>	<b>89.3</b>	90.2	10.7
MDLSTM-RNN <sup>[138]</sup>	—	83.5	16.5
MDLSTM-RNN+LM 4-gram <sup>[138]</sup>	—	<b>90.4</b>	10.6

### 3 总结、讨论与展望

本文简要回顾了手写体汉字识别的研究背景,介绍了传统手写汉字识别系统中各个主要步骤所涉及到的主要技术,指出当前研究的主要困难及所面临的挑战.然后介绍了 CNN、RNN、LSTM、SAE、DBN 等几种主要的深度学习模型,总结了一些主流的深度学习开源仿真工具.在此基础上,详细综述及分析了近年来采用基于深度学习的方法在联机及脱机手写体汉字识别的最新进展,总结了各种深度学习模型在手写体汉字识别的研究现状.

近年来,由于智能手机、平板电脑等触屏智能设备以及以 Microsoft Surface Pro4、iPad Pro、三星 Note4 等为代表的手写笔交互的移动互联网智能设备的迅猛发展,并逐渐在人们日常生活中占据重要地位.随着触屏智能手机代替传统键盘手机,笔交互设备的第二次复兴,文字输入从原来纯键盘的拼音或五笔输入方式将逐渐变成虚拟键盘、手写和语音等多种输入结合的方式.艾媒咨询 2015 年第二季度中国市场调研数据显示<sup>[144]</sup>:输入方式的使用比例中,手写输入方式占 13.1%,仅次于九宫格拼音(占 47.2%)和全键盘拼音输入(占 24.8%),并远大于语

音输入(占 5.8%)和五笔输入(3.6%),手写输入用户连续三年呈现平稳增长态势,手写输入作为一个重要的触屏交互应用也逐渐流行并广受重视,每天将产生大量的各种各样手写样本.因此,中文手写识别技术仍然值得更多的关注和更深入的研究.

手写体汉字识别经历了四十多年的长足发展,在单字和文本行识别性能上有了很大的提高,特别是以 CNN 为代表的一系列深度学习模型的出现,手写单字符中文识别问题已经基本上得到了很好解决,无论是联机还是脱机手写中文字符识别,目前基于 CNN 及其改进模型的方法均取得了接近甚至是超过人眼识别性能的高识别率.然而,在书写汉字识别领域,仍然很多值得研究的问题有待解决,例如:

1) 手写文本行识别问题:目前基于深度学习模型的联机及脱机手写文本行识别的成功报道很少,自从 ICDAR 2013 中文手写文本行竞赛以来,近两年在此方向上仍然没有突破性进展,对于联机中文手写文本行识别,目前最好的识别率指标(CR 或者 AR)均才达到 95% 左右<sup>[36,38]</sup>,而脱机手写文本行识别最好的 AR 及 CR 指标仅为 90% 左右<sup>[35,138]</sup>,特别是以整行为单位来评价识别率,行级别的识别率将会很低,仍然有很大的提升空间.可以说联机及脱机手写文本行识别仍然是未解决的难题.一些值得关注的研究方向包括:a)融合基于切分+基于 CNN 的单字识别+路径优化算法,这当中有不少关键技术需要解决,例如如何提升 CNN 的置信度的可靠性、能否用 CNN 的方法进行字符分割等;b)基于无切分的 RNN/LSTM/BLSTM 的滑窗式识别方法,此方面对于小类别的拉丁文字已经成为公认的最佳方法,但对于大类别的汉字识别问题是否最优仍然值得商榷及探讨;c)基于 CNN+LSTM 的端到端的纯深度学习解决方案.

2) 无约束的手写文字识别问题:其中一个值得关注的研究问题是旋转无关的手写识别问题,根据 2010 年发布的国家标准 GB/T18790-2010“联机手写汉字识别系统技术要求与测试规程”<sup>[145]</sup>,手写输入软件及设备必须要能识别  $\pm 45^\circ$  的手写样本,然而目前市场上的绝大部分主流输入法产品均无法满足此要求.尽管一些研究人员注意到此问题,并开展了不少前期探索工作<sup>[67,140,146-148]</sup>,但总体而言,此问题仍然远未得到有效解决.相信深度学习新技术的出现,将为解决此问题提供崭新的思路及技术手段.另外,目前的研究工作绝大部分局限于解决简单的问题,例如单字符识别或简单的文本行识别,对于联机手写重叠文本行识别、混合手写单字/文本行/重叠以及来自整屏任意无约束书写的手写汉字识别的研究工作仍然鲜有报道,这是一个值得研究的课题.

3) 超大类别手写汉字识别问题:目前手写汉字

识别研究报道所能识别的文字类型基本上以国标一级字库 3 755 类汉字为主, 针对实际应用场景下要能识别 10 000 个以上类别的实用化手写识别研究的报道还不多, 且缺乏公开的超大类别 (例如支持 GB8010-2000 标准的 27 533 类) 训练及测试数据集. 在如此大类别的情况下, 如何研究一个处理速度快、模型参数足够小的可实用化的基于深度学习的解决方案将变得极具挑战性, 从 ICDAR 2013 手写中文竞赛的评测结果来看, 基于深度学习的解决方案无论是在存储量还是识别速度上均还没有达到实用化的要求. 因此, 超大规模数据集的构建、针对各种深度学习模型的快速算法、深度学习模型的参数压缩技术等问题仍然需要研究人员的继续探索及解决. 此外, 超大类别还有一个很大的应用是古籍文献及手稿的内容识别及检索, 中国古籍的字符类别超过 3 万类, 而且有更多的变体字和相似字, 如何解决训练样本不足的问题以及如何设计超大类别分类器都是值得关注的问题.

4) 新的深度学习模型在手写汉字识别中的应用研究: 目前在手写汉字识别领域能取得比传统方法明显好的深度学习模型主要是基于 CNN 及其各种改进方法, 其他的深度学习模型如 DBN、RNN、LSTM/BLSTM/MDLSTM 以及深度强化学习 (DRN) 模型在大类别手写汉字识别中的研究工作开展得还不多, 各种深度学习模型之间的相互联系及融合应用的研究并不深入, 我们十分期待其他的深度学习模型以及未来能有更新更好的针对文字识别的深度模型能提出来, 并在手写汉字识别领域能取得突破性进展, 从而促进此领域的研究及发展.

5) 自然场景中的文字检测与识别: 近年来, 随着大量的互联网图片爆炸式增长, 自然场景中的文字检测及识别成为文字识别乃至计算机视觉领域一个极其重要和广受关注的热点研究课题<sup>[96-97]</sup>, 深度学习理论及技术的出现和发展为解决这一极具挑战的难题提供了很好的解决方案, 近年来已经出现了大量研究成果<sup>[96-97, 128, 142-143]</sup>. 但与传统的 MSER 框架的方法相比, 深度学习的方法处理速度慢、模型参数存储量大等也亟待解决. 在检测识别精度方面, 从 ICDAR 2015 场景文字检测及识别竞赛的结果来看<sup>[149]</sup>: 非受限环境下的自然场景文字 (Incidental scene text) 检测及识别性能还远未得到有效解决, 文字检测任务 (Task 4.1) 最好的召回率 (Recall) 才 46.94%, 单词识别任务 (Task 4.3) 最好的正确识别率仅为 32.93%. 在无语料信息的真实环境下的端到端识别任务 (Task 4.4), 最好识别率仅 34.96%, 可见此问题极具挑战性, 仍然有极大的提升空间. 此外, 目前绝大多数研究工作是针对英文语言, 针对中

文的自然场景文字检测及识别的研究报道还不多见, 而自然场景图像中的手写汉字检测及识别方面的研究工作开展的还很少, 要解决这些问题任重道远.

总体而言, 深度学习为解决手写汉字识别提供了新的理念及技术, 近几年来也在此领域诸多方面取得了大量的研究成果, 但仍然有不少研究问题值得进一步研究. 本文通过对相关领域的研究进展的回顾及分析讨论, 希望能够给该领域的研究人员带来新的信息及研究思路, 共同促进手写体汉字识别及相关文档分析与识别领域的进一步发展及繁荣.

## References

- 1 Hildebrandt T H, Liu W T. Optical recognition of handwritten Chinese characters: advances since 1980. *Pattern Recognition*, 1993, **26**(2): 205-225
- 2 Suen C Y, Berthod M, Mori S. Automatic recognition of handprinted characters — the state of the art. *Proceedings of the IEEE*, 1980, **68**(4): 469-487
- 3 Tai J W. Some research achievements on Chinese character recognition in China. *International Journal of Pattern Recognition and Artificial Intelligence*, 1991, **5**(01n02): 199-206
- 4 Liu C L, Jaeger S, Nakagawa M. Online recognition of Chinese characters: the state-of-the-art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(2): 198-213
- 5 Cheriet M, Kharma N, Liu C L, Suen C Y. *Character Recognition Systems: a Guide for Students and Practitioners*. USA: John Wiley & Sons, 2007.
- 6 Plamondon R, Srihari S N. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(1): 63-84
- 7 Dai R W, Liu C L, Xiao B H. Chinese character recognition: history, status and prospects. *Frontiers of Computer Science in China*, 2007, **1**(2): 126-136
- 8 Liu C L. High accuracy handwritten Chinese character recognition using quadratic classifiers with discriminative feature extraction. In: *Proceedings of the 18th International Conference on Pattern Recognition*. Hong Kong, China: IEEE, 2006. 942-945
- 9 Long T, Jin L W. Building compact MQDF classifier for large character set recognition by subspace distribution sharing. *Pattern Recognition*, 2008, **41**(9): 2916-2925
- 10 Liu C L, Yin F, Wang D H, Wang Q F. Online and offline handwritten Chinese character recognition: benchmarking on new databases. *Pattern Recognition*, 2013, **46**(1): 155-162
- 11 Zhang H G, Guo J, Chen G, Li C G. HCL2000 — a large-scale handwritten Chinese character database for handwritten character recognition. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Barcelona, Spain: IEEE, 2009. 286-290

- 12 Qian Yue-Liang, Lin Shou-Xun, Liu Qun, Liu Yang, Liu Hong, Xie Ying. Design and construction of HTRDP corpora resources for Chinese language processing and intelligent human-machine interaction. *Chinese High Technology Letters*, 2005, **15**(1): 107–110  
(钱跃良, 林守勋, 刘群, 刘洋, 刘宏, 谢紫. 863 计划中文信息处理与智能人机接口基础数据库的设计和实现. 高技术通讯, 2005, **15**(1): 107–110)
- 13 Jin L W, Gao Y, Liu G, Liu G Y, Li Y Y, Ding K. SCUT-OUCH2009—a comprehensive online unconstrained Chinese handwriting database and benchmark evaluation. *International Journal on Document Analysis and Recognition*, 2011, **14**(1): 53–64
- 14 Liu C L, Sako H, Fujisawa H. Handwritten Chinese character recognition: alternatives to nonlinear normalization. In: *Proceedings of the 7th International Conference on Document Analysis and Recognition*. Edinburgh, UK: IEEE, 2003. 524–528
- 15 Liu C L, Marukawa K. Pseudo two-dimensional shape normalization methods for handwritten Chinese character recognition. *Pattern Recognition*, 2005, **38**(12): 2242–2255
- 16 Jin L W, Huang J C, Yin J X, He Q H. Deformation transformation for handwritten Chinese character shape correction. In: *Proceedings of the 3rd International Conference on Advances in Multimodal Interfaces*. Beijing, China: Springer, 2000. 450–457
- 17 Miyao H, Maruyama M. Virtual example synthesis based on PCA for off-line handwritten character recognition. In: *Proceedings of the 7th International Workshop on Document Analysis Systems VII*. Nelson, New Zealand: Springer, 2006. 96–105
- 18 Chen G, Zhang H G, Guo J. Learning pattern generation for handwritten Chinese character using pattern transform method with cosine function. In: *Proceedings of the 2006 International Conference on Machine Learning and Cybernetics*. Dalian, China: IEEE, 2006. 3329–3333
- 19 Leung K C, Leung C H. Recognition of handwritten Chinese characters by combining regularization, Fisher's discriminant and distorted sample generation. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Barcelona, Spain: IEEE, 2009. 1026–1030
- 20 Okamoto M, Nakamura A, Yamamoto K. Direction-change features of imaginary strokes for on-line handwriting character recognition. In: *Proceedings of the 14th International Conference on Pattern Recognition*. Brisbane, QLD: IEEE, 1998. 1747–1751
- 21 Okamoto M, Yamamoto K. On-line handwriting character recognition using direction-change features that consider imaginary strokes. *Pattern Recognition*, 1999, **32**(7): 1115–1128
- 22 Ding K, Deng G Q, Jin L W. An investigation of imaginary stroke technique for cursive online handwriting Chinese character recognition. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition*. Barcelona, Spain: IEEE, 2009. 531–535
- 23 Jin L W, Wei G. Handwritten Chinese character recognition with directional decomposition cellular features. *Journal of Circuits, Systems, and Computers*, 1998, **8**(4): 517–524
- 24 Bai Z L, Huo Q. A study on the use of 8-directional features for online handwritten Chinese character recognition. In: *Proceedings of the 8th International Conference on Document Analysis and Recognition*. Seoul, Korea: IEEE, 2005. 262–266
- 25 Liu C L, Zhou X D. Online Japanese character recognition using trajectory-based normalization and direction feature extraction. In: *Proceedings of 10th International Workshop on Frontiers in Handwriting Recognition*. La Baule, France: IEEE, 2006.
- 26 Ge Y, Huo Q, Feng Z D. Offline recognition of handwritten Chinese characters using Gabor features, CDHMM modeling and MCE training. In: *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Orlando, FL, USA: IEEE, 2002. I-1053–I-1056
- 27 Liu C L. Normalization-cooperated gradient feature extraction for handwritten character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(8): 1465–1469
- 28 Kimura F, Takashina K, Tsuruoka S, Miyake Y. Modified quadratic discriminant functions and the application to Chinese character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, **PAMI-9**(1): 149–153
- 29 Mangasarian O L, Musicant D R. Data discrimination via nonlinear generalized support vector machines. *Complementarity: Applications, Algorithms and Extensions*. US: Springer, 2001. 233–251
- 30 Kim H J, Kim K H, Kim S K, Lee J K. On-line recognition of handwritten Chinese characters based on hidden Markov models. *Pattern Recognition*, 1997, **30**(9): 1489–1500
- 31 Liu C L, Sako H, Fujisawa H. Discriminative learning quadratic discriminant function for handwriting recognition. *IEEE Transactions on Neural Networks*, 2004, **15**(2): 430–444
- 32 Jin X B, Liu C L, Hou X W. Regularized margin-based conditional log-likelihood loss for prototype learning. *Pattern Recognition*, 2010, **43**(7): 2428–2438
- 33 Srihari S N, Yang X S, Ball G R. Offline Chinese handwriting recognition: an assessment of current technology. *Frontiers of Computer Science in China*, 2007, **1**(2): 137–155
- 34 Su T H, Zhang T W, Guan D J, Huang H J. Off-line recognition of realistic Chinese handwriting using segmentation-free strategy. *Pattern Recognition*, 2009, **42**(1): 167–182
- 35 Wang Q F, Yin F, Liu C L. Handwritten Chinese text recognition by integrating multiple contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(8): 1469–1481
- 36 Zhou X D, Wang D H, Tian F, Liu C L, Nakagawa M. Handwritten Chinese/Japanese text recognition using semi-Markov conditional random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(10): 2413–2426



- 37 Qiu L Q, Jin L W, Dai R F, Zhang Y X, Li L. An open source testing tool for evaluating handwriting input methods. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 136–140
- 38 Lin C L, Yin F, Wng Q F, Wang D H. ICDAR 2011 Chinese handwriting recognition competition. In: Proceedings of the 11th International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 1464–1469
- 39 Yin F, Wang Q F, Zhang X Y, Liu C L. ICDAR 2013 Chinese handwriting recognition competition. In: Proceedings of the 12th International Conference on Document Analysis and Recognition. Washington, DC, USA: IEEE, 2013. 1464–1470
- 40 Graham B. Spatially-sparse convolutional neural networks. arXiv: 1409.6070, 2014.
- 41 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504–507
- 42 Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(8): 1798–1828
- 43 Schmidhuber J. Deep learning in neural networks: an overview. *Neural Networks*, 2015, **61**: 85–117
- 44 LeCun Y, Boser B, Denker J S, Howard R E, Habbard W, Jackel L D, Henderson D. Handwritten digit recognition with a back-propagation network. In: Proceedings of Advances in Neural Information Processing Systems 2. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990. 396–404
- 45 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324
- 46 Ranzato M A, Poultney C, Chopra S, LeCun Y. Efficient learning of sparse representations with an energy-based model. In: Proceedings of the 2007 Advances in Neural Information Processing Systems. USA: MIT Press, 2007. 1137–1144
- 47 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, **9**(8): 1735–1780
- 48 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 2012 Advances in Neural Information Processing Systems 25. Lake Tahoe, Nevada, USA: Curran Associates, Inc., 2012. 1097–1105
- 49 Ouyang W L, Wang X G, Zeng X Y, Qiu S, Luo P, Tian Y L, Li H S, Yang S, Wang Z, Loy C C, Tang X O. Deepidnet: Deformable deep convolutional neural networks for object detection. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 2403–2412
- 50 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- 51 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2014.
- 52 Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada: IEEE, 2013. 6645–6649
- 53 Xu K, Ba J, Kiros R, Cho, Courville A, Salakhutdinov R, Zemel R, Bengio Y. Show, attend and tell: neural image caption generation with visual attention. arXiv: 1502.03044, 2015.
- 54 Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: a neural image caption generator. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 3156–3164
- 55 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, **521**(7553): 436–444
- 56 Tang Y C, Mohamed A R. Multiresolution deep belief networks. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics. La Palma, Canary Islands, Spain: Microtome Publishing, 2012. 1203–1211
- 57 Srivastava N, Salakhutdinov R. Multimodal learning with deep Boltzmann machines. In: Proceedings of the 2012 Advances in Neural Information Processing Systems. Tahoe, Nevada, USA: Curran Associates, Inc., 2012. 2222–2230
- 58 Shao J, Kang K, Loy C C, Wang X G. Deeply learned attributes for crowded scene understanding. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 4657–4666
- 59 Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 1717–1724
- 60 Yang W X, Jin L W, Xie Z C, Feng Z Y. Improved deep convolutional neural network for online handwritten Chinese character recognition using domain-specific knowledge. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 551–555
- 61 Yang W X, Jin L W, Tao D C, Xie Z C, Feng Z Y. DropSample: a new training method to enhance deep convolutional neural networks for large-scale unconstrained handwritten Chinese character recognition. arXiv: 1505.05354, 2015.
- 62 Yang W X, Jin L W, Liu M F. Character-level Chinese writer identification using path signature feature, dropstroke and deep CNN. arXiv: 1505.04922, 2015.
- 63 Yang W X, Jin L W, Liu M F. DeepWriterID: an end-to-end online text-independent writer identification system. arXiv: 1508.04945, 2015.
- 64 Su T H, Liu C L, Zhang X Y. Perceptron learning of modified quadratic discriminant function. In: Proceedings of the 2011 International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 1007–1011

- 65 Du J, Hu J S, Zhu B, Wei S, Dai L R. A study of designing compact classifiers using deep neural networks for online handwritten Chinese character recognition. In: Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm, Sweden: IEEE, 2014. 2950–2955
- 66 Du J. Irrelevant variability normalization via hierarchical deep neural networks for online handwritten Chinese character recognition. In: Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition. Heraklion, Greece: IEEE, 2014. 303–308
- 67 Du J, Huo Q, Chen K. Designing compact classifiers for rotation-free recognition of large vocabulary online handwritten Chinese characters. In: Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan: IEEE, 2012. 1721–1724
- 68 Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527–1554
- 69 Du J, Hu J S, Zhu B, Wei S, Dai L R. Writer adaptation using bottleneck features and discriminative linear regression for online handwritten Chinese character recognition. In: Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition. Heraklion, Greece: IEEE, 2014. 311–316
- 70 Liwicki M, Graves A, Bunke H. A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks. In: Proceedings of the 9th International Conference on Document Analysis and Recognition. Curitiba, Paraná, Brazil, 2007. 367–371
- 71 Frinken V, Bhattacharya N, Uchida S, Pal U. Improved BLSTM neural networks for recognition of on-line Bangla complex words. *Structural, Syntactic, and Statistical Pattern Recognition*. Berlin Heidelberg, German: Springer, 2014. 404–413
- 72 Wu W, Gao G L. Online cursive handwriting Mongolia words recognition with recurrent neural networks. *International Journal of Information Processing and Management*, 2011, **2**(3): 20–26
- 73 Graves A. Generating sequences with recurrent neural networks. arXiv: 1308.0850, 2013.
- 74 Cireşan D, Meier U. Multi-column deep neural networks for offline handwritten Chinese character classification. In: Proceedings of the 2015 International Joint Conference on Neural Networks. Killarney, Ireland: IEEE, 2015. 1–6
- 75 Cireşan D C, Meier U, Gambardella L M, Schmidhuber J. Convolutional neural network committees for handwritten character classification. In: Proceedings of the 2011 International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 1135–1139
- 76 Wu C P, Fan W, He Y, Sun J, Naoi S. Handwritten character recognition by alternately trained relaxation convolutional neural network. In: Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition. Crete, Greece: IEEE, 2014. 291–296
- 77 Zhong Z Y, Jin L W, Xie Z C. High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. In: Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR). Tunis: IEEE, 2015. 846–850
- 78 Wang Y W, Li X, Liu C S, Ding X Q, Chen Y X. An MQDF-CNN hybrid model for offline handwritten Chinese character recognition. In: Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition. Heraklion, Greece: IEEE, 2014. 246–249
- 79 Gao Xue, Wang You-Wang. Recognition of similar handwritten Chinese characters based on CNN and random elastic deformation. *Journal of South China University of Technology: Natural Science Edition*, 2014, **42**(1): 72–76 (高学, 王有旺. 基于 CNN 和随机弹性形变的相似手写汉字识别. 华南理工大学学报: 自然科学版, 2014, **42**(1): 72–76)
- 80 Yang Zhao, Tao Da-Peng, Zhang Shu-Ye, Jin Lian-Wen. Similar handwritten Chinese character recognition based on deep neural networks with big data. *Journal on Communications*, 2014, **35**(9): 184–189 (杨钊, 陶大鹏, 张树业, 金连文. 大数据下的基于深度神经网络的相似汉字识别. 通信学报, 2014, **35**(9): 184–189)
- 81 Feng B Y, Ren M W, Zhang X Y, Suen C Y. Automatic recognition of serial numbers in bank notes. *Pattern Recognition*, 2014, **47**(8): 2621–2634
- 82 He M J, Zhang S Y, Mao H Y, Jin L W. Recognition confidence analysis of handwritten Chinese character with CNN. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 61–65
- 83 Bengio Y, Goodfellow I J, Courville A. Deep learning [Online], available: <http://www.deeplearningbook.org>, May 11, 2016
- 84 LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W, Jackel L D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, **1**(4): 541–551
- 85 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 1–9
- 86 Lin M, Chen Q, Yan S C. Network in network. arXiv: 1312.4400, 2013.
- 87 Orr G B, Müller K R. *Neural Networks: Tricks of the Trade*. German: Springer, 1998.
- 88 Hinton G E, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov R R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv: 1207.0580, 2012.
- 89 Wan L, Zeiler M, Zhang S X, LeCun Y, Fergus R. Regularization of neural networks using dropConnect. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta, USA, 2013. 1058–1066

- 90 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, Huang Z H, Karpathy A, Khosla A, Bernstein M, Berg A C, Li F F. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, **115**(3): 211–252
- 91 Sun Y, Chen Y H, Wang X G, Tang X O. Deep learning face representation by joint identification-verification. In: *Proceedings of Advances in Neural Information Processing Systems 27*. Montréal, Canada: MIT, 2014. 1988–1996
- 92 Taigman Y, Yang M, Ranzato M A, Wolf L. DeepFace: closing the gap to human-level performance in face verification. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014. 1701–1708
- 93 Toshev A, Szegedy C. Deeppose: Human pose estimation via deep neural networks. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014. 1653–1660
- 94 Williams R J, Zipser D. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1989, **1**(2): 270–280
- 95 Graham B. Sparse arrays of signatures for online character recognition. arXiv: 1308.0371, 2013.
- 96 Jaderberg M, Simonyan K, Vedaldi A, Zisserman A. Synthetic data and artificial neural networks for natural scene text recognition. arXiv: 1406.2227, 2014.
- 97 Jaderberg M, Vedaldi A, Zisserman A. Deep features for text spotting. In: *Proceedings of the 13th European Conference Computer Vision*. Zurich, Switzerland: Springer, 2014. 512–528
- 98 Wu Y C, Yin F, Liu C L. Evaluation of neural network language models in handwritten Chinese text recognition. In: *Proceedings of the 13th International Conference on Document Analysis and Recognition*. Tunis: IEEE, 2015. 166–170
- 99 Bengio Y, Schwenk H, Senécal J S, Morin F, Gauvain J L. Neural probabilistic language models. *Innovations in Machine Learning*. Berlin Heidelberg, Germany: Springer, 2006. 137–186
- 100 Chen X, Tan T, Liu X, Lanchantin P, Wan M, Gales MJF, Woodland PC. Recurrent neural network language model adaptation for multi-genre broadcast speech recognition. In: *Proceedings of the 2015 International Speech Communication Association Interspeech*. Dresden, Germany, 2015. 3511–3515
- 101 Sak H, Senior A, Rao K, İrsoy O, Graves A, Beaufays F, Schalkwyk J. Learning acoustic frame labeling for speech recognition with recurrent neural networks. In: *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing*. South Brisbane, QLD: IEEE, 2015. 4280–4284
- 102 De Mulder W, Bethard S, Moens M F. A survey on the application of recurrent neural networks to statistical language modeling. *Computer Speech & Language*, 2015, **30**(1): 61–98
- 103 He K M, Zhang X Y, Ren S Q, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: *Proceedings of the 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE, 2015. 1026–1034
- 104 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv: 1502.03167, 2015.
- 105 Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 1980, **36**(4): 193–202
- 106 Werbos P J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 1990, **78**(10): 1550–1560
- 107 Littman M L. Reinforcement learning improves behaviour from evaluative feedback. *Nature*, 2015, **521**(7553): 445–451
- 108 Mnih V, Kavukcuoglu K, Silver D, Rusu A A, Veness J, Bellemare M G, Graves A, Riedmiller M, Fidjeland A K, Ostrovski G, Petersen S, Beattie C, Sadik A, Antonoglou I, King H, Kumaran D, Wierstra D, Legg S, Hassabis D. Human-level control through deep reinforcement learning. *Nature*, 2015, **518**(7540): 529–533
- 109 Cuda-ConvNet2 [Online], available: <https://github.com/akrizhevsky/cuda-convnet2>, May 11, 2016
- 110 Bengio Y, LeCun Y, Nohl C, Burges C. LeRec: a NN/HMM hybrid for on-line handwriting recognition. *Neural Computation*, 1995, **7**(6): 1289–1303
- 111 Simard P Y, Steinkraus D, Platt J C. Best practices for convolutional neural networks applied to visual document analysis. In: *Proceedings of the 7th International Conference on Document Analysis and Recognition*. Edinburgh, UK: IEEE, 2003. 958–963
- 112 Caffe [Online], available: <http://caffe.berkeleyvision.org/>, May 11, 2016
- 113 Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow I, Bergeron A, Bouchard N, Warde-Farley D, Bengio Y. Theano: new features and speed improvements. arXiv: 1211.5590, 2012.
- 114 Bergstra J, Breuleux O, Bastien F, Lamblin P, Pascanu R, Desjardins G, Turian J, Warde-Farley D, Bengio Y. Theano: a CPU and GPU math expression compiler. In: *Proceedings of the 9th Python for Scientific Computing Conference*. Austin, TX, USA, 2010. 1–7
- 115 Torch [Online], available: <http://torch.ch/>, May 11, 2016
- 116 Lin M, Li S, Luo X, Yan S C. Purine: a bi-graph based deep learning framework. arXiv: 1412.6249, 2014.
- 117 MXNet [Online], available: <https://github.com/dmlc/mx-net>, May 11, 2016
- 118 DIGITS [Online], available: <https://developer.nvidia.com/digits>, May 11, 2016

- 119 ConvNet [Online], available: <https://code.google.com/p/cuda-convnet/>, May 11, 2016
- 120 DeepCNet [Online], available: <http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/graham/>, May 11, 2016
- 121 Xing E P, Ho Q R, Dai W, Kim J K, Wei J L, Lee S, Zheng X, Xie P T, Kumar A, Yu Y L. Petuum: a new platform for distributed machine learning on big data. *IEEE Transactions on Big Data*, 2015, 1(2): 49–67
- 122 Weninger F, Bergmann J, Schuller B. Introducing CUR-RENT: the Munich open-source CUDA recurrent neural network toolkit. *The Journal of Machine Learning Research*, 2015, 16(1): 547–551
- 123 Minerva [Online], available: <https://github.com/dmlc/minerva>, May 11, 2016
- 124 TensorFlow [Online], available: <https://github.com/tensorflow/tensorflow>, May 11, 2016
- 125 DMTK [Online], available: <https://github.com/Microsoft/DMTK>, May 3, 2016
- 126 Cireşan D C, Meier U, Schmidhuber J. Transfer learning for Latin and Chinese characters with deep neural networks. In: Proceedings of the 2012 International Joint Conference on Neural Networks. Brisbane, QLD: IEEE, 2012. 1–6
- 127 Cireşan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, Rhode Island: IEEE, 2012. 3642–3649
- 128 Bastien F, Bengio Y, Bergeron A, Boulanger-Lewandowski N, Breuel T, Chherawala Y, Cisse M, Côté M, Erhan D, Eustache J, Glorot X, Muller X, Lebeuf S P, Pascanu R, Rifai S, Savard F, Sicard G. Deep self-taught learning for handwritten character recognition. arXiv: 1009.3589, 2010.
- 129 Chen L, Wang S, Fan W, Sun J, Naoi S. Beyond human recognition: a CNN-based framework for handwritten character recognition. In: Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition. Kuala Lumpur, Malaysia: IEEE, 2015. 695–699
- 130 Chen K T. Integration of paths-A faithful representation of paths by noncommutative formal power series. *Transactions of the American Mathematical Society*, 1958, 89(2): 395–407
- 131 Lyons T. Rough paths, Signatures and the modelling of functions on streams. arXiv: 1405.4537, 2014.
- 132 Graham B. Fractional max-pooling. arXiv: 1412.6071, 2014.
- 133 Graves A, Fernández S, Gomez F, Schmidhuber J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, Pennsylvania, USA: ACM, 2006. 369–376
- 134 Graves A, Schmidhuber J. Offline handwriting recognition with multidimensional recurrent neural networks. In: Proceedings of the 2009 Advances in Neural Information Processing Systems 21. Vancouver, B.C., Canada: Curran Associates, Inc., 2009. 545–552
- 135 Zhang X, Wang M, Wang L J, Huo Q, Li H F. Building handwriting recognizers by leveraging skeletons of both of-line and online samples. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 406–410
- 136 Simistira F, Ul-Hassan A, Papavassiliou V, Gatos B, Katsourous V, Liwicki M. Recognition of historical Greek polytonic scripts using LSTM networks. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 766–770
- 137 Frinken V, Uchida S. Deep BLSTM neural networks for unconstrained continuous handwritten text recognition. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 911–915
- 138 Messina R, Louradour J. Segmentation-free handwritten Chinese text recognition with LSTM-RNN. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 171–175
- 139 Mioulet L, Garain U, Chatelain C, Barlas P, Paquet T. Language identification from handwritten documents. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 676–680
- 140 Huang S M, Jin L W, Lv J. A novel approach for rotation free online handwritten Chinese character recognition. In: Proceedings of the 10th International Conference on Document Analysis and Recognition. Barcelona, Spain: IEEE, 2009. 1136–1140
- 141 Moyssset B, Kermorvant C, Wolf C, Louradour J. Paragraph text segmentation into lines with recurrent neural networks. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 456–460
- 142 He P, Huang W L, Qiao Y, Loy C C, Tang X O. Reading scene text in deep convolutional sequences. arXiv: 1506.04395, 2015.
- 143 Shi B G, Bai X, Yao C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. arXiv: 1507.05717, 2015.
- 144 iiMedia Research. 2015Q2 Report of input methods for mobile phone in China market [Online], available: <http://www.iimedia.com.cn/>, May 11, 2016
- 145 General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. GB/T18790-2010 Requirements and test procedure of on-line handwriting Chinese character recognition system. 2011  
(中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T18790-2010 联机手写汉字识别系统技术要求与测试规程. 2011)



- 146 Long T, Jin L W. A novel orientation free method for on-line unconstrained cursive handwritten Chinese word recognition. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa, FL, USA: IEEE, 2008. 1–4
- 147 He T T, Huo Q. A character-structure-guided approach to estimating possible orientations of a rotated isolated online handwritten Chinese character. In: Proceedings of the 10th International Conference on Document Analysis and Recognition. Barcelona, Spain: IEEE, 2009. 536–540
- 148 Huang S. A Study on Recognition for Rotated Isolated Online Handwritten Chinese Character [Master dissertation], South China University of Technology, China, 2010 (黄盛明. 联机手写汉字的旋转无关识别研究 [硕士学位论文]. 华南理工大学, 2010)
- 149 Karatzas D, Gomez-Bigorda L, Nicolaou A, Ghosh S, Bagdanov A, Iwamura M, Matas J, Neumann L, Chandrasekhar V R, Lu S J, Shafait F, Uchida S, Valveny E. ICDAR 2015 competition on robust reading. In: Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis: IEEE, 2015. 1156–1160



**金连文** 华南理工大学电子与信息学院教授. 1991 年在中国科学技术大学获得学士学位, 1996 年在华南理工大学获得博士学位. 主要研究方向为模式识别, 深度学习, 文字识别, 图像处理, 计算机视觉. 本文通信作者.

E-mail: [lianwen.jin@gmail.com](mailto:lianwen.jin@gmail.com)

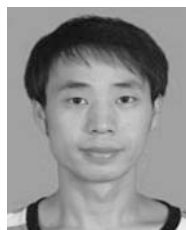
(**JIN Lian-Wen** Professor at the School of Electronic and Information Engineering, South China University of Technology. He obtained his B.S. degree from the Department of Electronics Engineering, University of Science and Technology of China and Ph.D. degree in communication and information system from South China University of Technology in 1991 and 1996, respectively. His research interest covers pattern recognition, deep learning, character recognition, image processing and computer vision. Corresponding author of this paper.)



**钟卓耀** 华南理工大学电子与信息学院博士研究生. 2015 年获得华南理工大学工学学士学位. 主要研究方向为机器学习, 模式识别, 自然场景文字检测与识别. E-mail: [z.zhuoyao@mail.scut.edu.cn](mailto:z.zhuoyao@mail.scut.edu.cn)

(**ZHONG Zhuo-Yao** Ph.D. candidate at the School of Electronic and Information Engineering, South China

University of Technology. He received his B.S. degree in electronics and information engineering from South China University of Technology in 2015. His research interest covers machine learning, pattern recognition, text detection and recognition in natural scene.)



**杨 钊** 广州大学机械与电气工程学院讲师. 2014 年获得华南理工大学信息与通信工程专业博士学位. 主要研究方向为机器学习, 模式识别, 计算机视觉.

E-mail: [15989023646@126.com](mailto:15989023646@126.com)

(**YANG Zhao** Lecturer at the School of Mechanical and Electric Engineering, Guangzhou University. He re-

ceived his Ph.D. degree from South China University of Technology in 2014. His research interest covers machine learning, pattern recognition, and computer vision.)



**杨维信** 华南理工大学电子与信息学院博士研究生. 2013 年获得华南理工大学工学学士学位. 主要研究方向为机器学习, 手写分析和识别, 计算机视觉和智能系统. E-mail: [wxy1290@163.com](mailto:wxy1290@163.com)

(**YANG Wei-Xin** Ph.D. candidate at the School of Electronic and Information Engineering, South China Uni-

versity of Technology. He received his B.S. degree in electronics and information engineering from South China University of Technology in 2013. His research interest covers machine learning, handwriting analysis and recognition, computer vision, and intelligent systems.)



**谢泽澄** 华南理工大学电子与信息学院博士研究生. 2014 年获得华南理工大学工学学士学位. 主要研究方向为机器学习, 文档分析与识别, 计算机视觉和人机交互. E-mail: [xiezcheng@foxmail.com](mailto:xiezcheng@foxmail.com)

(**XIE Ze-Cheng** Ph.D. candidate at the School of Electronic and Information Engineering, South China Uni-

versity of Technology. He received his B.S. degree in electronics and information engineering from South China University of Technology in 2014. His research interest covers machine learning, document analysis and recognition, computer vision, and human-computer interaction.)



**孙 俊** 富士通研究开发中心有限公司信息技术研究部部长. 2002 年获得清华大学模式识别和智能系统博士学位. 主要研究方向为图像处理、机器学习和模式识别. E-mail: [sunjun@cn.fujitsu.com](mailto:sunjun@cn.fujitsu.com)

(**SUN Jun** Director of Information Technology Laboratory, Fujitsu Research & Development Center Co.,

Ltd. He received his Ph.D. degree in pattern recognition and intelligent system from Tsinghua University in 2002. His research interest covers image processing, machine learning, and pattern recognition.)