

基于注意力机制的手写体中文字符识别

黄婉蓉, 何凯*, 刘坤, 高圣楠

天津大学电气自动化与信息工程学院, 天津 300072

摘要 手写体中文的自动识别在文档数字化、手写笔记转录等方面有广泛应用。针对其具有的书写随意、结构复杂、数目众多等特点, 提出了一种基于注意力机制的手写体中文识别方法。在卷积神经网络(CNN)模型的基础上, 搭建了一种 AT(Attention)-CNN 网络模型, 利用注意力机制实现网络层之间的信息交互, 减少了因池化操作导致的信息丢失。在经典手写体中文数据集 HWDB 上进行实验, 结果表明, 该方法的识别准确率可以达到 95.05%, 相比其他模型有显著提升。

关键词 手写体中文; 卷积神经网络(CNN); 注意力机制; 注意力-CNN 模型(AT-CNN 模型)

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP57.081002

Handwritten Chinese Character Recognition Based on Attention Mechanism

Huang Wanrong, He Kai*, Liu Kun, Gao Shengnan

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Abstract Automatic recognition of handwritten Chinese has a wide range of applications in document digitization and handwritten note transcription. A method based on attention mechanism is proposed to recognize the handwritten Chinese characterized by their random writing, complex structure, and large number of features. Based on the traditional convolutional neural network (CNN) model, an attention-CNN (AT-CNN) model is proposed. The information interaction between each layer in the network is realized using attention mechanism, thus the information loss caused by pooling operations reduces. Experiments on the classical handwritten Chinese data set HWDB show that the recognition accuracy of this method can reach 95.05%, which is significantly improved compared with that by other models.

Key words handwritten Chinese; convolutional neural network(CNN); attention mechanism; attention-CNN model (AT-CNN model)

OCIS codes 100.3008; 100.4996; 100.2000

1 引言

手写体中文的自动识别是模式识别领域的热点问题, 具有重要的实际应用价值, 如: 识别信封上的地址、支票上的信息以及其他需要人工处理的繁琐任务^[1]。尽管手写体中文的自动识别在近些年取得了较大的进展, 但与识别手写数字和拉丁字母相比, 仍存在一些问题^[2-3]。主要原因: 1) 数目庞大, 常用的中文字符超过 3000 个^[4]; 2) 中文字符结构复杂、

笔画繁琐、书写随意, 部分中文字符字形特征十分相似, 人工甚至都难以识别。

手写中文识别可分为在线和离线识别 2 大类^[5]。前者是书写者通过物理设备(触摸屏或手写板等)在线书写后自动转化为文本, 后者是通过计算机将光学设备(扫描仪或照相机等)捕获的静态图像自动转换为文本。与在线识别系统相比, 离线识别系统由于缺少书写笔顺信息, 识别准确率往往较低^[6]。离线手写中文识别常用的分类器模型主要

收稿日期: 2019-07-05; 修回日期: 2019-08-30; 录用日期: 2019-09-10

基金项目: 国家自然科学基金(61271326)

* E-mail: hekai@tju.edu.cn

有:支持向量机、改进的二次判决函数^[5]等,这些模型的识别精度有限,无法满足实用化的要求。而深度学习技术的发展为手写中文识别带来了新的思路和解决方案。如:LeCun等^[7]在20世纪90年代提出用Lenet5网络模型识别手写数字,在MNIST数据集上的识别准确率超过99%。卷积神经网络(CNN)在手写英文、中文识别等领域也有所应用^[8-9]。如:富士通团队^[3]提出的深度卷积神经网络,吴等^[10]提出的松弛卷积神经网络(ATR-CNN),以及金连文等^[11]提出的Ensemble-GoogLeNet网络,都取得了较好的手写中文识别效果。与传统方法相比,深度学习方法虽然能获得更高的识别准确率,但大多是通过加深网络层数、扩大各层网络规模来提升网络表现,容易产生梯度爆炸和梯度弥散的问题。

为解决这些问题,本文在卷积神经网络模型的基础上,提出了一种基于注意力机制^[12]的AT(Attention)-CNN模型。实验结果表明,与传统CNN模型相比,该模型能够更好地保留中文字符的特征信息,获得更高的识别准确率。

2 卷积神经网络与注意力机制

卷积神经网络是一种从传统神经网络方法演化而来的多层网络结构,在自然语言处理、计算机视觉等方面有着重要的应用。卷积神经网络通常由卷积层、池化层以及全连接层组成,通过训练进行特征提取完成特定的任务,常用于预测和分类^[13]。

2.1 卷积层

卷积层由多个特征图组成,拥有局部连接和共享特征的权重,其目的是获得输入图像的特征表示。卷积核是卷积层的核心,包括输入通道数、输出通道数和卷积核的大小等参数。卷积层首先将输入特征图与卷积核进行计算,再将结果传递给激活函数,可表示为

$$\mathbf{x}_j^l = f\left(\sum \mathbf{x}_i^{l-1} * \mathbf{k}_{ij}^l + \mathbf{b}_j^l\right), \quad (1)$$

式中, l 为卷积层数, i 为特征图的个数, \mathbf{k} 为卷积核, \mathbf{M}_j 为第 j 个特征图, \mathbf{b} 为偏置矩阵, f 为激活函数。典型的激活函数主要有:sigmoid,tanh和relu等函数^[14]。

2.2 池化层

池化层具有二次特征提取的效果,通常放在两个卷积层之间,可以降低卷积层输出的特征图像维数,提高特征提取的鲁棒性,减少参数个数,降低计算复杂度。池化层中特征图像的大小一般根据内核的移动步长来确定。典型的池化方式主要有:平均

池化^[15]和最大池化^[16]等。平均池化是将感受野内所有像素值相加后取平均值作为输出,而最大池化则是直接选择感受野内的最大值作为输出。

2.3 全连接层

卷积神经网络的分类器是由一个或多个全连接层组成,全连接层可以看作一种卷积核大小为 1×1 的特殊的卷积层,目的是将前一层中的所有神经元连接到当前层对应的神经元。全连接层中没有保留空间信息,通常在全连接层中使用dropout或L1、L2等正则化方法,防止产生过拟合现象。

2.4 注意力机制

注意力机制是指有选择性地关注视觉输入中的部分内容,而忽略其他信息,在序列学习的任务中发挥着重要作用^[12,17]。通过训练模型,关联选择的输入和输出序列项目,有效克服编码-解码器架构中的限制,即输出序列中的每一项都取决于输入序列中的选择项。目前,注意力机制已广泛用于不同类型的序列预测,如:文本翻译,语音识别。

3 本文方法

3.1 AT模块

卷积神经网络虽然在图像分类领域取得了良好的效果,但其本身仍然存在一些缺陷。如:池化计算在降维的过程中,容易忽略特征中的某些细节信息,而手写汉字风格的千变万化,以及汉字之间本身的强相似性,这些信息被忽略后往往会对最终识别的准确率产生较大影响。为解决该问题,提出了一种基于卷积神经网络和注意力机制的交互思想,利用AT模块,实现了两层池化层之间的信息交互,一定程度上避免了有价值信息的丢失。AT模块的具体结构如图1所示。

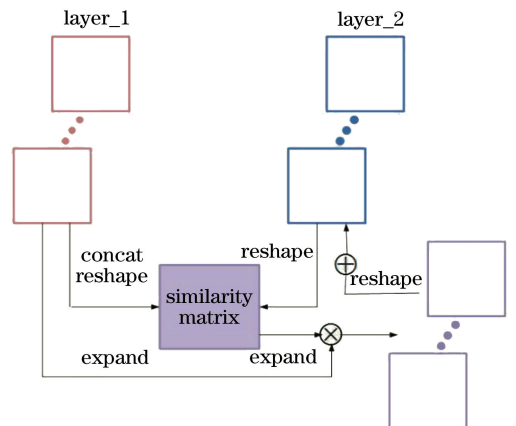


图1 AT模型网络结构

Fig. 1 AT model network structure

1) 层级表示的预处理。CNN 由多层卷积构成,用 L_{layer_1} 和 L_{layer_2} 表示任意相邻的两层图像。AT 模块首先将这两个二维表示用一维表示为 $L_{layer_1}(X_{height}, Y_{width}) \rightarrow L_{layer_1}(X_{height} \times Y_{width})$, 其中, X_{height} 和 Y_{width} 分别为 L_{layer_1} 的高和宽。

2) 计算层级间的相似性矩阵。由于不同层级表示包含信息不同,且这些信息相互关联,因此,通过注意力机制,利用多层层级之间的交互结果增强层级表示。相似性矩阵计算公式为

$$S_{ij} = \alpha(x_i, y_j), \quad (2)$$

$$\alpha(m, n) = \omega^T[m \oplus n \oplus (m \times n)], \quad (3)$$

式中, S_{ij} 为 x 的第 i 个向量和 y 的第 j 个向量之间的相似度, α 为一个可训练的标量函数, ω 为随机初始化的矩阵, \oplus 为级联操作。取 x 为 L_{layer_1} 、 y 为 L_{layer_2} , 得到相似性矩阵 S 。

3) 计算注意力机制交互结果。获得相似性矩

阵 S 后,将此交互结果作用于层级表示 L_{layer_1} 上,即:用矩阵乘法与 L_{layer_1} 层完成特征加权操作,再与 L_{layer_2} 层相加。可将新的 \tilde{L}_{layer_2} 层表示为

$$\tilde{L}_{layer_2} = S \times L_{layer_1} + L_{layer_2} \circ \quad (4)$$

S 既与原先 L_{layer_1} 层的特征相关,又与 L_{layer_2} 层的特征相关,从而避免了计算 \tilde{L}_{layer_2} 过程中因池化操作而导致的信息丢失。

3.2 AT-CNN 模型

所提出的 AT-CNN 模型由 CNN 模型和 AT 模块 2 部分组成。搭建一个包含 5 个卷积层、4 个池化层、1 个全连接层和 1 个输出层的卷积神经网络模型,具体的结构参数如表 1 所示。表中, C1~C8 为卷积层, MP2~MP9 为最大池化层, FC 为全连接层, Output 为输出层,因全连接层和输出层的特征图大小为 1×1 ,没有卷积核大小和步长这两个参数。

表 1 AT-CNN 模型结构和参数

Table 1 Structure and parameters of the AT-CNN model

Sequence of layer	Feature map size	Number of feature maps	Filter size	Step size
C1	64×64	32	3×3	1
MP2	32×32	32	2×2	2
C3	32×32	64	3×3	1
MP4	16×16	64	2×2	2
C5	16×16	128	3×3	1
MP6	8×8	128	2×2	2
C7	8×8	256	3×3	1
C8	8×8	256	3×3	1
MP9	4×4	256	2×2	2
FC	1×1	1024	—	—
Output	1×1	3755	—	—

AT-CNN 模型的具体结构如图 2 所示。为了在不增加网络深度的前提下提升分类准确度,AT-CNN 模型仅在 CNN 模型现有层级之间加入注意力机制,通过筛选传递特征,使保留的信息更有利于降低训练损失,从而提高分类的准确度。在网络的前端,利用 CNN 自动提取输入图片的相关特征,根据神经网络上一时刻的输出,利用 AT 模块在 CNN 模型中 MP2 最大池化层和 MP4 最大池化层之间加入注意力机制,计算 MP2 和 MP4 的注意力权重;将卷积神经网络 MP2 输出的特征图与注意力权重相结合,计算特征图的加权平均,再反馈给卷积神经网络 MP4;最终通过卷积神经网络全连接层实现分类输出。

4 实验结果及分析

4.1 数据集

实验使用 HWDB 数据集进行训练和测试,该数据集是中国科学院自动化研究所(CASIA)模式识别国家重点实验室公布的脱机手写样本数据库中的数据之一。由 300 位作者书写,包含 3755 个中文类别^[18],图 3 为其中两个中文字符的示例,可以看出,手写中文字符比较随意,同一中文字符不同样本之间存在的差异较大。

从数据集中随机选取 80% 的数据(约 901200 张图片)作为训练数据集,剩下的 20% 数据(约 225300 张图片)作为测试数据集。

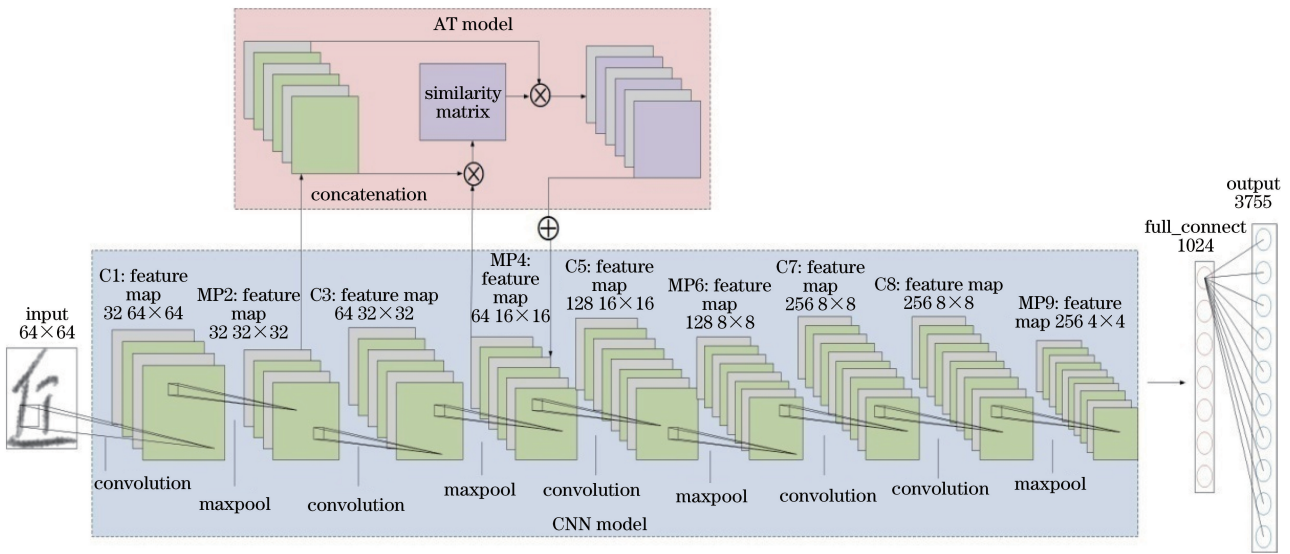


图 2 AT-CNN 模型网络结构

Fig. 2 AT-CNN model structure



图 3 HWDB 数据集示例

Fig. 3 Examples of HWDB dataset

4.2 实验结果及分析

实验环境:操作系统为 Ubuntu16.04, GPU 型号为 GTX1080TI, 内存为 11 G, 采用 TensorFlow 深度学习框架。其中, 批量大小 (Batch size) 设为 32, 随机失活 (Dropout) 率设为 0.8, 对学习率进行动态调整, 初始学习率设为 0.0002, 迭代 2000 次, 衰减至 97%。采用 AT-CNN 模型, 在训练集上准确率 (Accuracy) 曲线和损失 (Loss) 曲线如图 4 所示。可以看出, AT-CNN 模型经过 80000 步训练之后, 准确率曲线和损失曲线都逐渐趋于平稳; 140000 步之后, 准确率基本达到 97% 以上, loss 接近为 0, 证明了 AT-CNN 模型的有效性和准确性。

为进一步对比 CNN 和 AT-CNN 的两种模型, 随机选取 4 篇不同笔迹的中文文本进行测试, 如图 5 所示。先对文本图片做二值化处理, 然后利用直方图投影技术^[19]将其分割为字符图片, 最后使用训练好的模型对字符图片进行识别。

图 5 中 4 篇中文文本共 840 个中文字符。利用

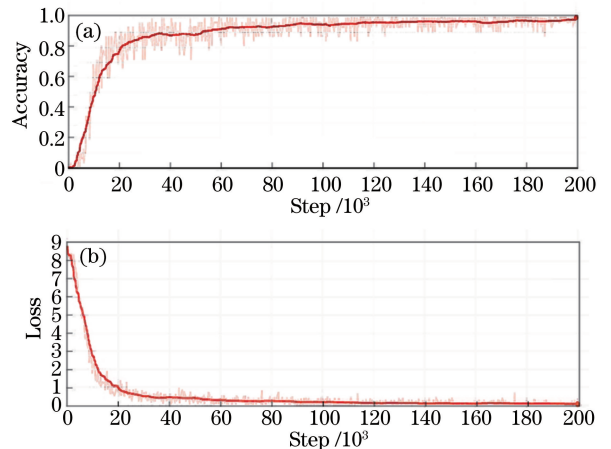


图 4 AT-CNN 模型训练结果。(a) Accuracy 曲线图; (b) Loss 曲线图

Fig. 4 AT-CNN model training result.

(a) Accuracy graph; (b) loss graph

CNN 模型正确识别个数为 807, 准确率为 96.1%, 用时 247 s; 利用 AT-CNN 模型正确识别个数为 814, 准确率为 96.9%, 用时 222 s。可以看出, 基于

CNN 的模型和 AT-CNN 的模型在 HWDB 训练集上训练获得的识别模型,在实际手写中文样本上也获得了良好的识别效果,对于同一测试样本,AT-

CNN 模型的识别准确率和识别速度都优于 CNN 模型,证明了 AT-CNN 模型具有较强的鲁棒性和泛化能力。

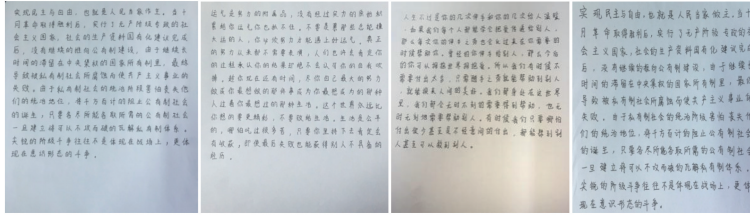


图 5 手写中文文本

Fig. 5 Handwritten Chinese text

采用不同方法对 HWDB 数据集进行识别,其准确率如表 2 所示。可以看出,CNN 模型和 AT-CNN 模型,识别准确率明显优于传统方法;而 AT-

CNN 模型相比于 CNN 模型,识别准确率提升了约 2 个百分点,有效提高了识别准确率,表中的空白项是由于对比的文献中未给出该参数。

表 2 几种方法在 HWDB 数据集上识别准确率比较

Table 2 Comparison of the recognition accuracy of different methods on HWDB dataset

Recognition method	RAM/MB	Accuracy / %
MQDF ^[20]	—	89.55
CCPR-2010 champion: HKU ^[21]	339. 10	89. 99
MQDF+CNN ^[20]	—	92.03
ICDAR-2011 champion: IDSIAnn-2 ^[22]	71. 35	92. 18
ICDAR-2013 champion: Fujitsu ^[3]	2460. 00	94. 77
HCCR-Ensemble-GoogLeNet ^[11]	277. 25	96. 74
CNN	108.30	93.01
AT-CNN	115.30	95.05

5 结 论

基于 CNN 模型和注意力机制思想,提出一种 AT-CNN 模型,通过网络层之间的信息交互,有效避免了部分有价值信息的丢失。基于 AT-CNN 模型对手写中文字符进行识别,实验结果表明,相比其他神经网络模型,AT-CNN 模型识别准确率有明显提高,且使用的参数较少,有利于实际应用。但本模型目前只实现了 3755 个常用汉字的识别,未来工作中将增加字符种类,扩大数据集,继续增强本模型的表现力和泛化能力。

参 考 文 献

[1] Maidana R G, dos Santos J M, Granada R L, et al. Deep neural networks for handwritten Chinese character recognition[C] // 2017 Brazilian Conference on Intelligent Systems (BRACIS), October 2-5, 2017, Uberlandia, Brazil. New York: IEEE, 2017: 192-197.

[2] Wu Y C, Yin F, Liu C L. Improving handwritten Chinese text recognition using neural network

language models and convolutional neural network shape models [J]. Pattern Recognition, 2017, 65: 251-264.

[3] Yin F, Wang Q F, Zhang X Y, et al. ICDAR 2013 Chinese handwriting recognition competition [C] // 2013 12th International Conference on Document Analysis and Recognition, August 25-28, 2013, Washington, DC, USA. New York: IEEE, 2013: 1464-1470.

[4] Leng W Y, Shamsuddin S M. Writer identification for Chinese handwriting [J]. International Journal of Advances in Soft Computing & Its Applications, 2010, 2(2): 142-173.

[5] Chen P F. Research on offline handwritten Chinese character recognition based on deep learning [D]. Jiangmen: Wuyi University, 2018.

陈鹏飞. 基于深度学习的脱机手写汉字识别技术研究 [D]. 江门: 五邑大学, 2018.

[6] Plamondon R, Srihari S N. Online and off-line handwriting recognition: a comprehensive survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(1): 63-84.

- [7] LeCun Y, Jackel L D, Bottou L, et al. Learning algorithms for classification: A comparison on handwritten digit recognition[J]. *Neural Networks: The Statistical Mechanics Perspective*, 1995: 261-276.
- [8] Sun H W. Research on handwritten English letter recognition algorithm based on deep learning[D]. Xi'an: Xidian University, 2018.
孙华伟. 基于深度学习的手写英文字母识别算法研究[D]. 西安: 西安电子科技大学, 2018.
- [9] Fang D B, Feng G, Cao H Y, et al. Handwritten formula symbol recognition based on multi-feature convolutional neural network [J]. *Laser & Optoelectronics Progress*, 2019, 56(7): 072001.
方定邦, 冯桂, 曹海燕, 等. 基于多特征卷积神经网络的手写公式符号识别[J]. *激光与光电子学进展*, 2019, 56(7): 072001.
- [10] Wu C P, Fan W, He Y, et al. Handwritten character recognition by alternately trained relaxation convolutional neural network [C] // 2014 14th International Conference on Frontiers in Handwriting Recognition. Crete, Greece: IEEE, 2014: 291-296.
- [11] Jin L W, Zhong Z Y, Yang Z, et al. Applications of deep learning for handwritten Chinese character recognition: a review [J]. *Acta Automatica Sinica*, 2016, 42(8): 1125-1141.
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[J]. *自动化学报*, 2016, 42(8): 1125-1141.
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C] // *Advances in Neural Information Processing Systems*, December 4-9, 2017, Long Beach, CA, USA. Canada: NIPS, 2017: 5998-6008.
- [13] Wang M, Fan T F, Yun W G, et al. PFWG improved CNN multispectra remote sensing image classification[J]. *Laser & Optoelectronics Progress*, 2019, 56(3): 031003.
王民, 樊潭飞, 负卫国, 等. PFWG改进的CNN多光谱遥感图像分类[J]. *激光与光电子学进展*, 2019, 56(3): 031003.
- [14] Wang H X, Zhou J Q, Gu C H, et al. Design of activation function in CNN for image classification [J]. *Journal of Zhejiang University (Engineering Science)*, 2019, 53(7): 1363-1373.
王红霞, 周家奇, 辜承昊, 等. 用于图像分类的卷积神经网络中激活函数的设计[J]. *浙江大学学报(工学版)*, 2019, 53(7): 1363-1373.
- [15] Zubair S, Yan F, Wang W W. Dictionary learning based sparse coefficients for audio classification with max and average pooling [J]. *Digital Signal Processing*, 2013, 23(3): 960-970.
- [16] Hang S T, Aono M. Bi-linearly weighted fractional max pooling[J]. *Multimedia Tools and Applications*, 2017, 76(21): 22095-22117.
- [17] Zhu M K, Lu X L. Human action recognition algorithm based on Bi-LSTM-Attention model [J]. *Laser & Optoelectronics Progress*, 2019, 56(15): 151503.
朱铭康, 卢先领. 基于Bi-LSTM-Attention模型的人体行为识别算法[J]. *激光与光电子学进展*, 2019, 56(15): 151503.
- [18] Liu C L, Yin F, Wang D H, et al. CASIA online and offline Chinese handwriting databases [C] // 2011 International Conference on Document Analysis and Recognition, September 18-21, 2011, Beijing, China. New York: IEEE, 2011: 37-41.
- [19] Wang L L, Yu Y. Text image character segmentation method based on bidirectional projection [J]. *Digital Technology and Application*, 2017(5): 74-75.
王莉丽, 于印. 一种基于双向投影的文本图像字符分割方法[J]. *数字技术与应用*, 2017(5): 74-75.
- [20] Wang Y W, Li X, Liu C S, et al. An MQDF-CNN hybrid model for offline handwritten Chinese character recognition [C] // 2014 14th International Conference on Frontiers in Handwriting Recognition, September 1-4, 2014, Heraklion, Greece. New York: IEEE, 2014: 246-249.
- [21] Liu C L, Yin F, Wang D H, et al. Chinese handwriting recognition contest 2010 [C] // 2010 Chinese Conference on Pattern Recognition (CCPR), October 21-23, 2010, Chongqing, China. New York: IEEE, 2010: 11694086.
- [22] Liu C L, Yin F, Wang Q F, et al. ICDAR 2011 Chinese handwriting recognition competition [C] // 2011 International Conference on Document Analysis and Recognition, September 18-21, 2011, Beijing, China. New York: IEEE, 2011: 1464-1469.