基于帧特征及维特比解码的手写体与印刷体分类

林琴[1*]，夏俊峰[2]，涂铮铮[2]，郭玉堂[1]
[1]合肥师范学院计算机学院，安徽 合肥 230601；
[2]安徽大学计算机学院，安徽 合肥 230039

摘要 为有效区分手写体与印刷体，提出了一种基于卷积神经网络隐层帧特征的分类方法。基于卷积神经网络，提取隐层帧特征，利用高斯混合模型结合隐马尔可夫模型的方法对该特征进行建模，再通过维特比解码算法判定每帧特征的类别。基于帧特征的识别结果，结合文本行图像信息对识别结果进行后处理，确定最终的手写体和印刷体的区域。在签名文书类文本行图像上，相比基线，所提方法对手写体与印刷体分类的识别率提升 10.8% 和 27.57%。在自然场景、表格和带噪文档行验证了其有效性。

关键词 图像处理；手写体与印刷体分类；卷积神经网络；隐马尔科夫模型；维特比解码
中图分类号 TP391.4 **文献标识码** A **doi**：10.3788/LOP56.061003

Discrimination of Handwritten and Printed Texts Based on Frame Features and Viterbi Decoder

Lin Qin[1*], Xia Junfeng[2], Tu Zhengzheng[2], Guo Yutang[1]
[1] *School of Computer Science Technology, Hefei Normal University, Hefei, Anhui* 230601, *China*;
[2] *College of Computer Science and Technology, Anhui University, Hefei, Anhui* 230039, *China*

Abstract To effectively distinguish the handwritten and printed texts, a discrimination method is proposed based on the hidden layer frame features of a convolutional neural network. The hidden layer frame feature is extracted by the convolutional neural network. The Gaussian mixture model is first combined with the hidden Markov model to model the features, and then the Viterbi decoding algorithm is used to determine the category of each frame feature. Based on the recognition results of the frame features, the recognition results are post-processed in combination with the image information. The final handwritten and printed text areas are determined. For the signature document line images, relative to the baseline, the discrimination accuracy of handwritten and printed texts by the proposed method increases by 10.8% and 27.57%, respectively. The effectiveness of the proposed method is verified with the natural scenes, tables and noisy documents.

Key words image processing; discrimination of handwritten and printed texts; convolutional neural network; hidden Markov model; Viterbi decoder
OCIS codes 100.2000; 100.3008; 100.5010

1 引 言

随着移动互联网技术以及数字技术的迅猛发展，视觉文字数据正以惊人的速度增长，迫切需要发展智能视觉文字分析技术。视觉文字分析是指从机器视觉的角度，抽取和理解文字信息，涉及图像处理、模式识别、计算机视觉、机器学习和心理学等一系列学科知识，一直以来都是相关领域的重要研究对象。在日常工作中，经常在文档图像中出现既包含印刷体又包含手写体的情况。针对此类图像数据，通常需要将手写体与印刷体分开处理，如在安全机构的信息校验、安全性检查中，需要将文档图像中

收稿日期：2018-08-21；**修回日期**：2018-09-28；**录用日期**：2018-10-10
基金项目：国家自然科学基金青年基金(61602006)、安徽省高校省级自然科学研究重点项目(KJ2017A934)、安徽省高校省级自然科学研究重点项目(KJ2013A217)

* **E-mail**：linqin@hfnu.edu.cn

手写签名等信息单独提取。为了提高系统的识别率，通常需要采用不同的光学字符识别(OCR)系统分别处理手写体部分和印刷体部分。

目前，印刷体和手写体有两种分类方案：一是以块为基本处理单元的分布式方案；另一种是像素级别的一体化方案。本文研究的是第一种方案，该方案一般分为3个步骤：分割、特征提取和分类、后处理。分割的主要目的是从文档图像中切出稳定基本单元，这些单元是进行特征提取和分类的基础，通常可以通过算法标记为印刷体、手写体或噪声等；大多研究利用印刷体和手写体在视觉上的差别，提取梯度直方图、灰度邻接矩阵、Gabor 滤波响应[1]等特征，鲜有文献采用深度神经网络[2]特征的方法。特征分类可包含连通域、局部块和文本行3个层次[3-4]。基于连通域的手写体与印刷体分类方法，能够较好地显示字符的信息及位置，但其连通域分析依赖于图像的二值化效果，故在复杂背景的图像上难以发挥优势。Koyama 等[5]对局部块进行傅里叶变换，并将其作为特征进行手写体和印刷体的分类。当整个文本行都属于同一类时，基于文本行的手写体与印刷体分类方法，能够很好地描述手写体与印刷体的特性[6]，但当文本图像同时含有手写体字符和印刷体字符时，此分类方法效果不佳。

为了有效地进行分类，首先针对文本行图像只属于一类的情况，通过滑窗提取方向梯度直方图(HOG)特征；再利用支持向量机(SVM)进行分类[7]；根据投票的准则判断文本行类别。但该方法无法对既有手写体又有印刷体的文本行进行手写体定位。为进一步解决该问题，提出了一种基于卷积神经网络(CNN)隐层帧特征及其维特比解码的分类方法，该方法通过神经网络学习得到帧特征，与 HOG 特征相比，减少了人工干预的参数设置[8]，且帧特征大致相当于几个或十几个像素块，能够很好地细化手写体和印刷体分类，为确定其边界提供便利。

本文提取基于文本行的卷积神经网络帧特征，利用高斯混合模型(GMM)结合隐马尔科夫(HMM)模型对帧特征进行建模[9-10]，后端采用维特比解码判定帧特征的类别。维特比算法是在 HMM 上进行求解[11]，找到的最优输出状态序列，即为最优的手写体印刷体类别序列，该方法简称为 GMM+Viterbi。

2 基于帧特征的手写体印刷体分类

2.1 系统描述

基本流程如图1所示，对于输入文本行图片，首先经过光学字符识别系统的卷积神经网络提取帧特征。基于文本行的光学字符识别系统，以文本行作为系统输入，以文本行各帧的状态标签为目标，训练卷积神经网络[12-13]。该网络在深度神经网络(DNN-HMM)模型[14-15]基础上进行了改进，并以此网络为特征提取器，将文本行图像转换成一组帧特征，其值作为网络最后一个隐层的线性输出，特征维数则为该隐层的通道数。卷积神经网络的训练过程使隐层帧特征能够有效挖掘相邻像素之间的相关信息，形成紧凑的特征表示。相对于传统的图像特征，通过该方法提取的隐层帧特征具有更好的噪声稳健性和区分性。

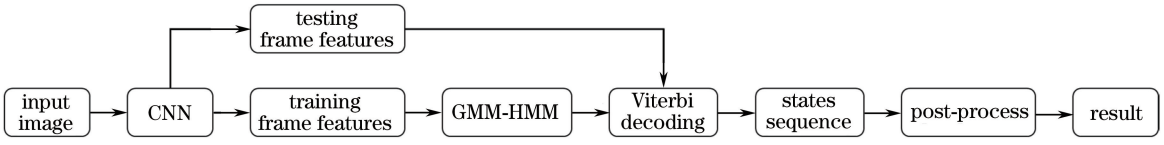

图1 所提算法流程图

Fig. 1 Flow chart of proposed algorithm

表1为基于文本行的光学字符识别卷积神经网络结构，表中 conv 为卷积层，pool 为池化层，采用最大值池化(max pool)，fc 为全连接层；卷积核参数中 d_{pad} 为卷积时边缘填充的大小，e_{stride} 为步长，$e_{\mathrm{stride\_h}}$ 为高度方向的步长，S 为卷积神经网络的类别数，即字符集合的状态总数。提取卷积层 3-2 线性输出作为文本行图像的帧特征，根据该卷积层的输出大小，可得文本行图像的帧特征维数。该卷积层的通道数为128维，文本行图像的帧数为$[(W-3)/2-3]$，其中：W 为文本行图像送入卷积神经网络识别时，将高度规整到48对应的宽度。一般而言，每帧特征对应横向约3 pixel，相比滑窗颗粒度更细；卷积神经网络特征具备图像上下文信息；相比传统图像特征，卷积神经网络可以增加各类场景的数据，并预先训练好，通过学习得到的特征更具稳健性。卷积神经网络每层卷积层代表输入图片不同层次的信息，Yosinski 等[16]将各卷积层进行可视化与理解分析，获取底层卷积层表征图像的边缘等信息，越高层越接近图像的语义信息。

表 1 基于文本行的 OCR 卷积神经网络结构

Table 1 Convolutional neural network structure of OCR based on text line

Layer name	Output size	Convolution kernel
conv1-1	$24\times[(W-3)/2+1]$	$32@3\times3$, $d_{\text{pad}}=1$
conv1-2		$64@1\times1$
conv1-3		$64@3\times3$, $d_{\text{pad}}=1$
pool1		3×3 Max pool, $e_{\text{stride}}=2$
conv2-1	$6\times[(W-3)/2-3]$	$64@1\times1$
conv2-2		$128@3\times3$, $e_{\text{stride\_h}}=2$
conv2-3		$64@1\times1$
conv2-4		$128@3\times3$, $d_{\text{pad}}=1$
pool2		3×3 Max pool, $e_{\text{stride\_h}}=2$
conv3-1	$1\times[(W-3)/2-3]$	$256@3\times1$, $d_{\text{pad}}=1$
conv3-2		$128@3\times1$
fc		$S@1\times1$

基于上述卷积神经网络帧特征，对手写体和印刷体的训练集特征，分别训练 GMM，GMM 为产生式模型，其描述的是特征的分布特性。对于 C 维的特征 x，用含有 K 个高斯混合数的 GMM 建模，其公式表达为

$$p(\boldsymbol{x}\mid\Lambda)=\sum_{k=1}^{K}\pi_k N(\boldsymbol{x}\mid\mu_k,\Sigma_k),\tag{1}$$

式中：π_k 为第 k 个高斯的权重，且满足 $\sum_{k=1}^{K}\pi_k=1$，$\pi_k\geqslant0$；μ_k 和 Σ_k 分别为对应高斯的均值和方差；p 表示特征分布概率。GMM 模型参数 $\Lambda=\{\pi_k,\mu_k,\Sigma_k\}_{k=1}^{K}$，采用最大似然准则进行参数估计，并用期望最大化(EM)算法[17]优化模型参数。本文采用两个 GMM 模型，分别描述手写体和印刷体帧特征的分布特性；对测试文本行图像提取出每帧特征，通过手写体 GMM 模型 $\Lambda_{\text{handwritten text}}$ 和印刷体 GMM 模型 $\Lambda_{\text{printed text}}$，分别计算其属于手写体与印刷体的概率，即给定观测特征的后验概率。

HMM 模型如图 2 所示，其中观测序列 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_T)$ 为卷积神经网络隐层帧特征序列，y_i 为观测样本，i 为序号，T 为序列的总长度。隐含状态序列 $\boldsymbol{S}=(s_1,s_2,\cdots,s_T)$ 为每帧特征对应的手写体或印刷体类别，s_i 为隐含状态。概率最大的隐含状态序列满足

$$(s_1,s_2,\cdots,s_T)=\operatorname{argmax}p(s_1,s_2,\cdots,s_T\mid y_1,y_2,\cdots y_T),\tag{2}$$

根据隐马尔科夫假设，(3)式等价于

$$(s_1,s_2,\cdots,s_T)=\operatorname{argmax}\prod_{t=1}^{T}p(s_t\mid y_t)p(s_t\mid y_{t-1}),\tag{3}$$

式中：$p=(s_t\mid y_t)$ 通过 GMM 模型计算得到，即给定帧特征 y_t 时 s_t(手写体或者印刷体)的后验概率。

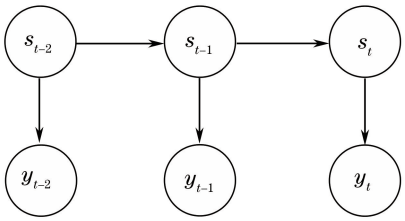

图 2 隐马尔可夫模型示意图

Fig. 2 Schematic of hidden Markov model

将手写体和印刷体类别分别设定为隐马尔科夫模型 HMM 的状态，记为 s_S 和 s_Y，则每个时刻的状态 s_t 为 s_S 或 s_Y，状态之间转移概率示意图见图 3。p_{ij} 表示状态 s_i 到 s_j 的状态转移概率，其值通过训练得到，本文表示帧特征在手写体和印刷体类别之间的转移概率。

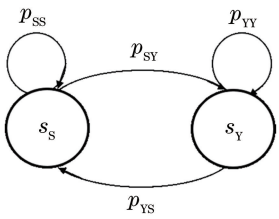

图 3 隐马尔科夫模型的状态转移示意图

Fig. 3 Schematic of state transition of hidden Markov model

给定 HMM 模型参数以及观测序列，采用维特比算法求解(3)式最优隐含状态序列，即每帧特征对应的手写体印书体类别序列。维特比算法是一种动态规划算法，利用动态规划求解概率最大的路径，见图 4，图中每条从 $t=1$ 到 $t=T$ 的路径都对应一条可能的状态序列。从 $t=1$ 时刻开始，不断向后递推寻找一个状态路径的最大概率，直到到达最终的最优路径终点，然后依据终点回溯到起始点，即得最优

路径。

维特比算法结合 HMM 状态转移概率的解码结果，相当于对 GMM 模型的输出概率的平滑效果。对帧级类别之间的跳转给定不同的惩罚因子，并非仅依据手写体和印刷体 GMM 模型的概率结果。测试过程为对测试文本行图像的卷积神经网络帧特征进行维特比解码，从而获得每帧特征的手写体或印刷体类别。

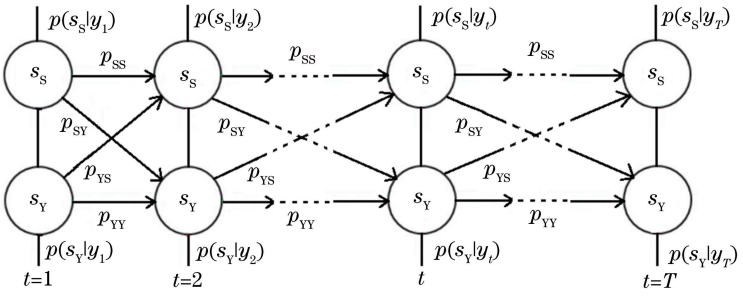

图 4 所有可能的维特比解码路径

Fig. 4 All possible Viterbi decoding paths

2.2 后处理

通过维特比解码获得帧特征的类别，进一步分析发现，字符过渡区域以及空白部分容易被误识。因此，需对帧特征的识别结果进行后处理。

1）根据文本行图像信息，将图像纵向分割成从左到右的一系列具有一定字符意义的区域；根据投票准则，认为该区域中帧数最多的类别，即为该区域的类别。例如，某区域一共包含 N 帧，其中半数以上为手写体，则判定该区域为手写体区域。

2）利用大津法对文本行图像作二值化处理；再结合图像文本的分布特征，将图像按纵向分割，得到一系列和原图等高的区域。这些区域可能是某个字符、某个字符的偏旁部首或者粘连在一起的连续字符。

3）根据提取帧特征的卷积神经网络结构，将帧特征的手写体印刷体识别结果反推映射到输入的文本行图像的像素上，且同一纵向上的像素具有相同类别。具体而言，文本行图像上的手写体印和刷体分类，即为手写体和印刷体区域左右边界的确定。

4）根据反推得到的像素上的手写体和印刷体类别，在纵向分割出来的区域内进行投票，判定其类别，继而获得整行图像上手写体和印刷体区域的左右边界。

3 实验及结果分析

选取签名文书类文本行图像进行实验验证。其中，训练集共 10 万行图像，文本行中含有手写体图像、印刷体图像以及两者均包含的图像。不对数据进行任何去噪、去下划线等预处理，直接送入基于文本行的卷积神经网络字符识别模型，提取帧特征。测试集在真实应用场景中收集得到，共 4019 行图像；采用与训练集相同的处理方法，将类别区域块的正确率作为系统性能的衡量指标。

利用 HOG 特征的窗口位置进行手写体或印刷体定位。在 HOG-SVM 方法中，将文本行图像高度规整到 40 pixel，HOG 特征窗口大小设定为 40×40，最终得到 576 维特征；再利用 libsvm[18] 实现线性 SVM 进行分类。

3.1 本文方法与 HOG-SVM 方法的手写体与印刷体分类效果的对比

利用深度卷积神经网络提取帧特征 128 维，即为该卷积层的通道数。将 GMM 模型的高斯混合数设定为 128，对比测试集两个系统的识别效果，如表 2 所示。从实验结果可以看出，本文提出的方法，无论在手写体还是印刷体的识别率上，都有较明显的提升。

表 2 基于帧特征及维特比解码的实验验证结果

Table 2 Experimental test results based on frame features and Viterbi decoder %

Method	Handwritten Accuracy	Printed Accuracy
HOG+SVM	67.24	61.55
GMM+Viterbi	72.90	88.65

将类别比作 HMM 的状态，利用类别之间的转换概率，使维特比解码在手写体与印刷体分类的应用中起到了平滑作用。将帧特征的位置映射到文本行图像上，并将类别区域画在图上进行分析，如图 5(a)所示，黑色区域框表示印刷体，灰色区域框表示手写体。从图 5(a)可以看出，字符的过渡区域和空白区域容易被误识。根据文本行图像信息以及

X-Y cut 方法[19]，将图像分割成具有一定字符意义的区域，如图 5(b)所示。在分割出来的区域内，对识别结果进行后处理，根据投票准则得到区域结果，如图 5(c)所示。从图 5(c)可以看出，后处理对字边界区域进行了有效修正，如图 5(a)最后一个过渡区域帧特征的识别结果为印刷体，后处理后根据结果被判定为手写体。表 3 为测试集经过后处理的识别结果。

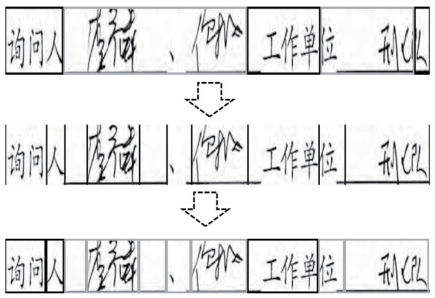

图 5 手写体和印刷体的分类结果。(a)帧特征解码结果映射到文本行图像；(b)纵向分割图像；(c)各区域重新判定的结果

Fig. 5 Discrimination results of handwritten and printed texts. (a) Frame feature decoding results mapped to text line images; (b) longitudinal image segmentation; (c) re-determination results in each region

实验结果显示，根据文本行图像本身的信息以及字符的特性对识别结果进行后处理，有利于最终手写体和印刷体区域的判断。

表 3 经后处理的基于帧特征及维特比解码的实验结果

Table 3 Experimental results based on frame features and Viterbi decoding followed by post-processing

Method	Handwritten accuracy /%	Printed accuracy /%	Frame /s
GMM+Viterbi	72.90	88.65	502
GMM+Viterbi+post-processing	78.04	89.12	496
BiLSTM	79.28	89.91	39

表 3 给出了基于双向长短时记忆网络(LSTM)的帧级别一体化方案，该方法不需要将深度神经网络帧级特征提取出来，也不需要进行维特比解码，深度神经网络的输出即为各帧手写体与印刷体的分类结果，但该方法模型结构复杂且效率慢。考虑到实验室资源以及应用效率，本文主要研究深度神经网络帧特征及维特比解码的方法。相同签名文书测试集下，两种方法的效率对比如表 3 最后一列，由此可以看出，本文方法在识别效果存在有限损耗时效率能够达到一体化方案的 10 倍以上。

3.2 不同手写体与印刷体分类方法对应的字符识别效果的对比

利用基于文本行的光学字符识别系统作为识别器，其网络结构如表 1 所示。将不同手写体和印刷体分类方法得到的区域块作为文本行，分别送入识别器中进行识别，通过手写体和印刷体的识别率验证所提方法的有效性，识别结果如表 4 所示。

表 4 不同分类方法下手写体与印刷体的字符识别率

Table 4 Character recognition accuracy of different discrimination methods of handwritten and printed texts

%

Method	Handwritten accuracy		Printed accuracy	
	Sent	Word	Sent	Word
Artificial segmentation	64.92	73.01	84.67	92.10
GMM+Viterbi+post-processing	**61.02**	**69.18**	**82.31**	**90.56**
HOG+SVM	57.85	66.43	79.62	87.95

由于测试集中的手写体大多为签名，相比印刷体识别难度较大，因此整体字符识别率较低。对比人工切分的识别情况，分类后再识别的效果存在些许损耗，一方面是分类方法的误检和漏检引起的，另一方面是分类边界不精准、边界图像字符识别错误引起的。从表 4 可以看出，在使用同一字符识别器的情况下，所提方法相比 HOG-SVM 更精准。

3.3 不同场景下的手写体与印刷体的分类效果及对应的字符识别率

为了验证所提方法的有效性，在不同场景进行实验验证，包括签名文书类、自然场景类、表格类和带噪文档类等。各场景的测试集数据均在真实应用场景中收集，各随机选取 5000 行图像。其中自然场景数据是随机拍照数据，背景多样复杂；表格数据是文书中带表格的数据，主要考虑表格线以及倾斜的影响；带噪文档主要是文档类中噪声干扰较大的数据。各场景的手写体与印刷体的分类结果如表 5 所示。

将各场景分类得到的手写体与印刷体区域分别送入识别器进行识别，其句识别率与字识别率如表 6 所示。

从表 5 和表 6 可以看出，无论在什么场景下本文方法都明显优于基线方法，利用卷积神经网络提取的帧特征更加稳健，而且相对 HOG+SVM 方法中通过滑窗来确定手写体和印刷体边界，所提方法中的帧特征对应的颗粒度更细，确定边界时也更加精确。

表 5 经后处理的各场景的手写体与印刷体分类准确率

Table 5 Classification accuracy of handwritten and printed texts after post-processing in each scene %

Scene	HOG+SVM		GMM+Viterbi+post-processing	
	Handwritten	Printed	Handwritten	Printed
Signed document	67.24	61.55	78.04	89.12
Natural scene	63.81	57.49	76.32	86.71
Table	65.29	57.43	72.66	86.36
Noisy document	60.31	55.23	71.48	82.23

表 6 不同场景中手写体与印刷体的字符识别率

Table 6 Character recognition accuracy of handwritten and printed texts in different scenes %

Scene	HOG+SVM				GMM+Viterbi+post-processing			
	Handwritten		Printed		Handwritten		Printed	
	Sent	Word	Sent	Word	Sent	Word	Sent	Word
Signed document	57.85	66.43	79.62	87.95	61.02	69.18	82.31	90.56
Natural scene	53.05	60.92	72.29	78.72	55.59	64.96	78.44	82.86
Table	54.61	61.98	73.89	78.73	55.16	65.01	78.60	85.21
Noisy document	45.35	54.87	66.40	72.56	48.21	56.52	68.73	76.67

4 结 论

基于深度卷积神经网络提取的隐层帧特征，提出了一种手写体和印刷体分类方法。该方法通过对帧特征进行 GMM-HMM 建模，构建手写体和印刷体特征分布模型；并利用维特比解码，获得帧特征的手写体和印刷体类别序列；最后，结合图像信息对帧特征结果进行后处理，得到最终的手写体和印刷体区域。通过签名文书类文本行图像上的实际应用，进一步证明了该方法的有效性。下阶段工作主要针对基于深度卷积神经网络的隐层帧特征的提取方法进行优化，将深度卷积神经网络的多层特征融合考虑到特征表达上，以期获得文本行图像不同层次的信息。

参 考 文 献

[1] Ye Z, Bai L. Hyperspectral image classification algorithm based on Gabor feature and locality-preserving dimensionality reduction[J]. Acta Optica Sinica, 2016, 36(10): 1028003.
叶珍, 白璘. 基于 Gabor 特征与局部保护降维的高光谱图像分类算法[J]. 光学学报, 2016, 36(10): 1028003.

[2] Wang D D, Li Y N. Video fingerprint algorithm based on spatio-temporal deep neural network[J]. Laser & Optoelectronics Progress, 2018, 55(1): 011006.
汪冬冬, 李岳楠. 基于时空深度神经网络的视频指纹算法[J]. 激光与光电子学进展, 2018, 55(1): 011006.

[3] Ding H, Zhang X F. Connected handwritten and printed text discrimination in uneven lighted images [J]. Computer Engineering and Design, 2012, 33(12): 4634-4638.
丁红, 张晓峰. 非均匀光照图像中粘连手写体和印刷体的辨别[J]. 计算机工程与设计, 2012, 33(12): 4634-4638.

[4] Yu X Y, Guo Y B, Chen G, *et al*. Real-time point feature extraction based on connected components labeling and distributed computing[J]. Acta Optica Sinica, 2015, 35(2): 0210001.
于潇宇, 郭玉波, 陈刚, 等. 基于点目标连通域标记的实时特征提取及其分布式运算[J]. 光学学报, 2015, 35(2): 0210001.

[5] Koyama J, Hirose A, Kato M. Local-spectrum-based distinction between handwritten and machine-printed characters [C] // 2018 Conference on Image Processing, October 12-15, 2008, San Diego, CA, USA. New York: IEEE, 10422955.

[6] Kavallieratou E, Stamatatos S. Discrimination of machine-printed from handwritten text using simple structural characteristics [C] // 2004 Conference on Pattern Recognition, August 26-26, 2004, Cambridge, UK. New York: IEEE, 8213163.

[7] Bristow H, Lucey S. Why do linear SVMs trained on HOG features perform so well?[EB/OL]. (2014-06-10) [2018-08-20]. https: // arxiv. org/abs/1406.2419.

[8] Jiang B, Song Y, Wei S, *et al*. Deep bottleneck features for spoken language identification[J]. PLoS

One, 2014, 9(7): e100795.

[9] Torres-Carrasquillo P A, Singer E, Kohler M A, *et al*. Approaches to language identification using Gaussian mixture models and shifted delta cepstral features[C] // 2002 Conference on Spoken Language Processing, September 16-20, 2002, Denver, Colorado, USA. [S. l. : s. n.], 2001: 89-92.

[10] Guan D J, Huang H J. Off-line recognition of realistic Chinese handwriting using segmentation-free strategy[J]. Pattern Recognition, 2009, 42(1): 167-182.

[11] Li R, Zhuo Z, Li H. The research of speaker diarization based on BIC and G_{PLDA} [J]. Journal of University of Science and Technology of China, 2015, 45(4): 286-293.
李锐, 卓著, 李辉. 基于 BIC 和 G_{PLDA} 的说话人分离技术研究[J]. 中国科学技术大学学报, 2015, 45(4): 286-293.

[12] Russakovsky O, Deng J, Su H, *et al*. Imagenet large scale visual recognition challenge [J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[13] Haykin S, Kosko B. Gradient-based learning applied to document recognition [M]. New Jersey: Wiley-IEEE Press, 2009: 306-351.

[14] Hinton G, Deng L, Yu D, *et al*. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups[J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.

[15] Du J, Wang Z R, Zhai J F, *et al*. Deep neural network based hidden Markov model for offline handwritten Chinese text recognition [C] // 2016 Conference on Pattern Recognition, December 4-8, 2016, Cancun, Mexico. New York: IEEE, 16835646.

[16] Yosinski J, Clune J, Nguyen A, *et al*. Understanding neural networks through deep visualization[EB/OL]. (2015-06-22)[2018-08-20]. https://arxiv.org/abs/1506.06579.

[17] Bishop C M. Pattern recognition and machine learning[M]. New York: Springer, 2006: 452-472.

[18] Chang C C, Lin C J. LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 1-27.

[19] Chen Y L, Wu B F. A multi-plane approach for text segmentation of complex document images [J]. Pattern Recognition, 2009, 42(7): 1419-1444.