

基于深度学习的文本检测算法在银行运维中应用^①

马胜蓝

(福建省农村信用社联合社科技服务中心, 福州 350001)

摘要: 银行故障单中故障的截图常存在与自然场景中, 能够在该图中精确地进行文本检测, 将可以提高文本识别的精确度, 并提高案例库的搜索和主动运维能力. 为了提高自然场景文本检测的效率, 提出了一种基于深度学习的自然场景文本检测算法. 算法首先提取出图像中的最大稳定极值区域作为候选字母, 利用单链接层次聚类得到候选文本, 对候选文本进行中值滤波, 最后通过一个深度置信网络 DBN 来删除非文本候选. 实验结果表明, 基于 DBN 的方法能有效提高自然场景文本检测的准确率, 比传统方法具有更好的结果.

关键词: 主动运维; 文本检测; 深度学习; 深度置信网络; 自然场景

Application of Text Detection Based on Deep Learning Algorithm in Operation and Maintenance of Bank

MA Sheng-Lan

(Science and Technology Service Center, Fujian Rural Credit Cooperatives, Fuzhou 350001, China)

Abstract: Screenshots of bank fault often exit in natural scenes. If the text can be accurately detected in the screenshots, it will be able to improve the accuracy of text recognition and improve the case base search and active operation and maintenance capabilities. In order to improve the efficiency of the text detection of natural scenes, an algorithm based on deep learning in natural scene is proposed. Firstly, candidate letters are extracted from the maximum stable extreme region, and candidate texts are generated by single-link hierarchical clustering, then the algorithm makes median filter for the candidate text. Lastly, non-texts are removed by the deep confidence network DBN. Experimental results show that DBN-based approach can effectively improve the accuracy of the text detection of natural scenes, and has better results than traditional methods.

Key words: active operation and maintenance; text detection; deep learning; DBN; natural scene

随着信息化的快速普及与发展, 银行的系统复杂性也在不断增加, 为此对运维管理的投入也不断加大. 目前, 国内很多银行都通过了 ISO20000 IT 服务管理体系认证, 同时也建立了信息系统运维管理方面的案例库, 但这些案例基本都停留在案例展示阶段, 尚未实现基于已有案例的风险预测、应急决策等分析功能^[1]. 面对现代银行信息系统生产运维压力的急剧增加, 以及类似故障重复发生、处置差异较大的严峻形式, 迫切需要能够根据故障单进行案例库的自动匹配, 实现在故障预测方面的自动化主动案例的推理^[2].

为此本文针对银行主动运维中的工作所需, 针对日常运维中不同场景下的故障图像进行文本检测, 实

现文本区域的精确定位, 提高文本识别的能力, 进而通过文本匹配案例库的方式搜索到对应的案例经验, 提高主动运维能力.

日常运维工作中, 对于故障拍摄图像诸如电脑 PC、ATM 机、自助机具等, 常处于不同的自然场景中, 在这类场景中文本检测中主要面临三个问题: ①文本背景内容复杂多变; ②文本具有多样性, 包括大小、颜色、字体等; ③文本与背景具有相似性. 为了解决类似自然场景图像中的上述问题, 许多学者也进行了各方面的研究, 主要方法为份额为基于区域的方法^[3,4]和基于相连成分的方法. 然而银行的故障交易场景中文本具有多样性, 很难确定一组有效的特征集来训练分类

① 收稿时间:2016-05-31;收到修改稿时间:2016-07-17 [doi:10.15888/j.cnki.csa.005628]

器,因此本文通过引进深度学习算法提升文本检测的效率.深度置信网络 DBN 是一种深度学习网络,它结合了无监督学习过程和有监督学习过程,在图像处理领域已逐渐取得了一些成就.文献[5]通过差分深度置信网络来对图像进行分类,取得了较好的分类效果;文献[6]利用 SVM 作为 DBN 网络最后一层的分类器,对图像中的表情进行识别;文献[7]则是 DBN 在图像检测中的应用,通过 DBN 来检测图像中的车辆.

因此,本文将深度置信网络 DBN 作为分类器,通过深度学习的自学习能力得到文本的特征,从而有效地提高分类的准确率.此外,为了降低候选文本图像中的噪声影响,对候选文本进行了中值滤波,归一化后作为 DBN 的输入.实验表明,本文提出的方法能有效地提高文本检测的准确率.

本文后续部分的结构如下:第二节介绍深度置信网络 DBN 的文本检测算法;第三节阐述银行主动运维中文本检测算法的应用;第四节结合 ICDAR 数据集进行实验结果比较,并展示部分银行运维中文本检测的效果图.

1 深度置信网络 DBN 的文本检测算法

本算法主要包括以下几个基本步骤:1)提取图像中的最大稳定极值区域 MSER 作为候选字母^[8];2)训练字母分类器以删除非字母候选;3)利用单链接聚类将候选字母聚类,得到候选文本^[9];4)对候选文本进行中值滤波,去除候选文本图像中的噪声;5)对候选文本进行归一化,利用深度学习来删除非文本候选,得到最终的结果.

1.1 对候选文本进行中值滤波并归一化

由于训练图像中存在较多的噪声,会对笔画宽度等特征计算造成很大的影响,引入中值滤波能够有效地抑制噪声,而且能有效地保护原图像的边缘细节^[10].峰值信噪比(PSNR)是最普遍、最广泛使用的评鉴图像质量的客观量测法,PSNR 值较大的图像质量较高^[11].

1.2 利用深度置信网络 DBN 删除非文本

本文采用深度置信网络 DBN 进行非文本的删除.DBN 网络是由若干层无监督的受限玻尔兹曼机(RBM)和一层有监督的反向传播网络(BP)组成的一种深层神经网络,是属于深度学习的一种机器学习模型^[12].

DBN 在训练模型的过程中主要分为两步,首先分别单独无监督地训练每一层 RBM 网络,确保特征

向量映射到不同特征空间时,都尽可能多地保留特征信息;在 DBN 的最后一层设置 BP 网络,接收 RBM 的输出特征向量作为它的输入特征向量,有监督地训练实体关系分类器.

本文得到的候选文本为二值化图像,在经过滤波后,对其进行归一化.基于文献[13]的经验,对于隐层单元个数的选择,可以先估算一下用一个模型描述一个数据所需的比特数,再用其乘上训练集容量,基于所得的数,选择比其低一个数量级的值作为隐层单元个数.因此,DBN 的隐层数选择了通用的三层,综合考虑图像整体质量与 DBN 计算速度,本文将归一化的大小设为 50*25,每幅图像 1250 个像素点,即 DBN 可视层的单元数为 1250 个.实验中,DBN 共包含 3 个隐层,各隐层单元数分别为 200、150、100.DBN 输出值为 1 表示该图像为文本图像,为 0 表示为非文本图像.

2 银行主动运维文本检测算法的应用

在建立基于文本检测算法的银行主动运维体系中,故障发现者将故障单和描述通过 ITIL 平台提交故障信息,电子化工具执行基于深度学习的文本检测算法,获取到故障单图像中的文本区域,之后调用 OCR 进行文本识别,在获取到文本信息后进入案例库进行检索匹配.对于匹配出的已有的问题解决方案的,直接通过 ITIL 平台返回给故障发现者进行处理,对于类似需要运维人员处理的,将故障单和案例库信息提交给运维人员进行进一步的处理.具体的过程如图 1 所示.

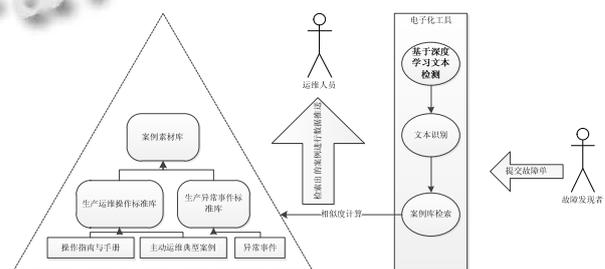


图 1 银行主动运维文本检测的过程

虽然 DBN 具有较好的学习能力,但 DBN 的运算复杂度较高.基于该算法的分析,主要的耗时在训练的过程中,一旦训练模型形成,分类的速度仍然可以满足实时性要求.在银行主动运维的应用场景中,对于训练模型可以在夜间通过跑批来生成,而在日间业

务中仅需实时调用模型来识别即可。

3 实验

本文的实验分为两部分，一部分是在被广泛使用的 ICDAR2011 测试集上进行的，性能评估也采用同样的方法^[14]，采用召回率 (recall rate)，准确率 (precision rate) 以及 f-measure 值作为评价指标^[15]；第二部分是对于银行故障截图进行文本检测的展示。

3.1 ICDAR 实验比较与分析

本文进行了两组实验，A 组为采用几何特征训练 adaboost 分类器进行文本分类；B 组为采用纹理特征训练 adaboost 分类器进行文本分类；C 组为使用 DBN 自动学习特征来进行文本分类。其中 A 组的特征为文本宽高比，几何特征值，共 15 个特征；B 组文本分类器特征为：文本宽高比，水平、垂直、对角三个方向上共生矩阵的熵、能量、对比度和相关性值，共 13 个特征。实验分类结果准确率对比如表 1、图 2 所示。

表 1 分类准确率对比表

图片	候选文本数	实际文本数	DBN			几何分类器			纹理分类器		
			检测数	文本数	准确率	检测数	文本数	准确率	检测数	文本数	准确率
1	5	4	4	4	1	4	3	0.75	5	4	0.8
2	16	7	8	7	0.875	10	7	0.7	9	7	0.78
3	9	5	4	4	1	7	4	0.57	6	4	0.67
4	5	4	4	4	1	5	4	0.8	4	3	0.75
5	4	4	4	4	1	4	4	1	4	4	1
6	6	4	4	3	0.75	5	3	0.6	4	3	0.75
7	9	3	3	3	1	6	3	0.5	5	3	0.6
8	4	4	4	4	1	4	4	1	4	4	1
9	5	2	3	2	0.67	4	1	0.25	5	1	0.2
10	13	5	6	4	0.67	8	3	0.375	7	3	0.429

表 1 中，候选文本数表示经过单链接聚类得到的候选文本数量，作为分类器的输入；实际文本数为图像中的文本个数；检测数表示经分类器分类后的检测个数；文本数为在检测结果中文本的个数；准确率则表示检测结果中文本所占的比例。

图 2 中的实线表示采用几何特征训练的 adaboost 分类器在每幅图像中的分类准确率；虚线表示采用纹理特征训练的 adaboost 分类器在每幅图像中的分类准确率；点状虚线表示深度置信网络 DBN 在每幅图像中的分类准确率。由表 1 和图 4 可以看出，使用 DBN 进行文本

分类能有效地提高文本分类的准确率。

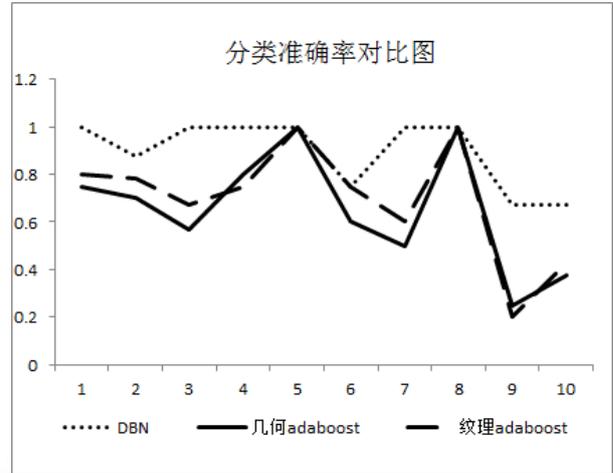


图 2 分类准确率对比图

如下将基于 MSER 的自然场景文本检测算法，通过与其他同类和不同类算法进行对比：文献[16]算法采用笔画宽度 SWT 的检测算法。文献[17]提出了一种基于边缘加强的 MSER 文本检测算法，在提取 MSER 的基础上加强区域边缘，再利用笔画宽度的特征过滤掉非文本。文献[18]提出了一种有效的剪枝方法来删除非字母候选，并且通过训练级联分类器来删除非文本候选，得到最终的检测结果。文献[15]和文献[19]提出了一种基于 MSER 及几何特征的文本检测算法，利用图像中文本具有相似几何特征这一特性，对候选文本进行分类，得到最终文本。文献[15]则提出了一种面向聚焦爬虫的文本检测方法，该方法利用角点强度增强技术和三值化方法将文本与背景分离。通过实验对比发现，本文的算法在检测准确率上具有明显的优势，在整体上表现出了较好的性能。图 3 列出了部分实验对比结果。



图 3 部分实验结果对比图

参考文献

- 1 谢新洲,夏晨曦.网络事件案例库建设与案例数据分析.情报学报,2012,31(1):72-81.
- 2 杨健,赵秦怡.基于案例的推理技术研究进展及应用.计算机工程与设计,2008,29(3):710-712.
- 3 Chen XR, Yulie AA. Detecting and reading text in natural scenes. Proc. IEEE Conference CVPR. Washington DC, USA. 2004. 366-373.
- 4 Lee JJ, Lee PH, Lee SW, et al. Adaboost for text detection in natural scene. Proc. ICDAR. Beijing. 2011. 429-434.
- 5 Zhou SS, Chen QC, Wang XL. Discriminative deep belief networks for image classification. 2010 17th IEEE International Conference on Image Processing (ICIP). Hong Kong. 2010. 1561-1564.
- 6 Wang H, Cai YF, et al. A vehicle detection algorithm based on deep belief network. The Scientific World Journal, 2014, (5): 1-7.
- 7 Huang CC, Gong W, et al. A research of speech emotion recognition based on deep belief network and SVM. Mathematical Problems in Engineering, 2014, 8(1).
- 8 Matas J, Chum O, Urban M, Pajdla T. Robust wide baseline stereo from maximally stable extremal regions. Proc. Brit. Mach. Vis. Conference. 2002. 384-393.
- 9 Chen X, Yin XW. Robust text detection in natural scene images. Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2014, 36(5): 970-983.
- 10 吴建华,李迟生,周卫星.中值滤波与均值滤波的去噪性能比较南昌大学学报,1998,20(1):32-36.
- 11 王敏,周磊,等.基于峰值信噪比和小波方向特性的图像奇异值去噪技术.应用光学,2013,34(1):85-89.
- 12 Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Computer, 2006, 18(7): 1527-1554.
- 13 Kim Y, Lee H, Provost EM. Deep learning for robust feature generation in audio visual emotion recognition. Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. 3687-3691.
- 14 Shahab A, Shafait F, Dengel A. ICDAR 2011 robust reading competition challenge 2: Reading text in scene images. Proc. ICDAR. 2011. 1491-1496.
- 15 彭浩,陈继锋,刘琼,等.一种面向聚焦爬虫的自然场景文本定位技术.小型微型计算机系统,2014,35(9):2014-2018.
- 16 Epshtein B, Ofek E, Wexler Y. Detecting text in natural scenes with stroke width transform. Proc. IEEE Conference CVPR. San Francisco, CA, USA. 2010. 2963-2970.
- 17 Chen HZ, Tsai SS, Schroth G, et al. Robust text detection in natural images with edge enhanced maximally stable extremal regions. Proc. IEEE International Conference Image Process. 2011. 2609-2612.
- 18 Chen X, Yin XW. Robust text detection in natural scene images. Proc. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2014, 36(5): 970-983.
- 19 Yin XW, Yin XC, Hao HW, et al. Effective text localization natural scene images with MSER, geometry-based grouping and adaboost. Proc. International Conference Pattern Recognit. Tsukuba, Japan. 2012. 725-728.