

基于深度学习的数显仪表字符识别

朱立倩

(中国石油大学(华东) 计算机科学与技术学院, 山东 青岛 266580)

摘要:在许多工业场景中,需要记录仪表的数据,将数据录入到电脑,这不仅耗时耗力,而且两次的转录可能导致错误的发生。为了提高监控效率,需要对数显仪表数据进行自动识别。针对传统字符分割方法适应性差,准确度低的不足,提出了一种基于深度学习的自动识别数显仪表字符的方法,由字符区域定位网络及字符识别网络构成。字符区域定位网络为改进的Faster R-CNN,将Faster R-CNN的骨干网络改为ResNeXt-101,感兴趣区域池化操作改为精确的感兴趣区域池化操作,以提高分类及定位的准确性。字符识别网络由卷积神经网络和加入注意力机制的长短时记忆网络构成,注意力机制的加入提高了字符识别的准确性。以变压器直流电阻测试仪为具体应用对象,实验结果显示,该方法可以达到95%的准确率。

关键词:数显仪表;卷积神经网络;注意力机制;字符检测;字符识别

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2020)06-0141-04

doi:10.3969/j.issn.1673-629X.2020.06.027

Character Recognition of Digital Display Instrument Based on Deep Learning

ZHU Li-qian

(School of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: In many industrial scenarios, it is necessary to record the data of the digital display instrument and then input it into the computer, which is not only time-consuming and laborious, but also two transcriptions may lead to errors. In order to improve the monitoring efficiency, it is necessary to recognize the data of the digital display instrument automatically. Aiming at the shortcomings of poor adaptability and low accuracy of traditional character segmentation methods, we propose a method of automatic recognition of digital display instrument characters based on deep learning, which is composed of character region location network and character recognition network. The character region location network is improved Faster R-CNN. The backbone network of Faster R-CNN is changed to ResNeXt-101, and the ROI-pooling is changed to precise ROI-pooling to improve the accuracy of classification and location. Character recognition network is composed of convolutional neural network and long short-term memory network with attention mechanism, which improves the accuracy of character recognition. Taking transformer DC resistance tester as a specific application object, the experimental results show that the proposed method can achieve 95% accuracy.

Key words: digital display instrument; convolutional neural network; attention mechanism; character detection; character recognition

0 引言

数显仪表在许多领域有着广泛的应用,但大多数的数显仪表不具备数据通信的功能,从而无法完成仪表数据的自动获取与存储。在实际应用中,需要大量的人力对仪表数据进行记录以及录入电脑,这不仅耗时耗力,而且多次的转录易导致错误发生,降低了系统的自动化水平。人们希望可以自动识别数显仪表中的数据,以提高工作效率及自动化水平。

机器视觉的不断发展,为数显仪表字符的自动识

别提供了可能。传统字符识别方法^[1-3]的一般步骤为对获取的原始图像进行预处理,字符定位,字符分割,字符识别,传统方法一般需要图片有较纯净的背景,略复杂的图片会导致识别失败,这表明传统方法适应性差,容错性低。近年来,随着深度学习技术^[4-6]的不断发展,文本检测和识别算法^[7-10]的准确度都大大提升,这为数显仪表字符识别提供了新的解决思路。

为了解决传统方法适应性差,容错性低的问题,采用深度学习技术,提出了一种基于深度学习的数显仪

收稿日期:2019-08-01

修回日期:2019-12-05

基金项目:国家自然科学基金(61402533);山东省自然科学基金资助项目(ZR2019MF049)

作者简介:朱立倩(1994-),女,硕士,研究方向为计算机视觉、图像处理。

表字符识别方法,记为 DDICRNet (digital display instrument character recognition net),该方法对图片内容有较大的容错性,可以应用于不同的数显仪表,只需要重新训练模型,同时可以保证识别的准确度。该方法以变压器直流电阻测试仪为具体应用对象,实验结果显示,该方法可以实现 95% 的准确率。

1 提出的方法

此部分详细介绍了提出的数显仪表字符识别方法。整体设计思路为字符区域定位网络获得图片中的字符区域位置信息,利用位置信息裁剪字符区域,将裁剪后的字符区域送入字符识别网络得到识别结果。整个方法由两部分构成,分别为字符区域定位网络 CRLNet(character region location net)和字符识别网络 CRNet(character recognition net)。

1.1 字符区域定位网络

定位的准确性直接影响识别的准确性,所以字符区域定位的准确性至关重要。为了提高字符区域定位网络的准确性,在 Faster R-CCN^[5] 的基础上,从两方面对其进行改进:(1)使用更加高效的特征提取网络,

以保证分类的准确性;(2)使用精确感兴趣区域池化 (precise RoI-Pooling)^[11] 代替 RoI-Pooling,提高定位的准确性。通过这两个方面的提高,保证检测结果的准确性。

CRLNet 由 ResneXt-101^[12], 区域提案网络^[5], 分类和回归网络以及后处理算法^[5] 构成。网络结构如图 1 所示。ResneXt-101 由五个卷积组构成,每个卷积组由卷积层、池化层、激活层构成。卷积组 2-5 借鉴了 GoogLeNe^T^[13] 的 split-transform-merge 思想,在大卷积核层两侧加入 1×1 的卷积,控制核个数的同时减少参数个数。与 Resnet^[14] 相比,参数更少,效果更好,训练速度更快。由于以变压器直流电阻测试仪为具体应用对象,所以考虑到需要定位区域的大小和形状,将 RPN 中的候选框的大小设置为 $64 \times 64, 128 \times 128$ 。纵横比设置为 1:4。将 ResneXt-101 最后一层产生的特征图送入 RPN 网络中,得到区域提案。选取 256 个区域提案作为样本,正负样本比为 3:1,通过精确的感兴趣区域池化操作对区域提案的特征图统一尺寸,通过三个全连接得到分类和边框回归的预测结果。

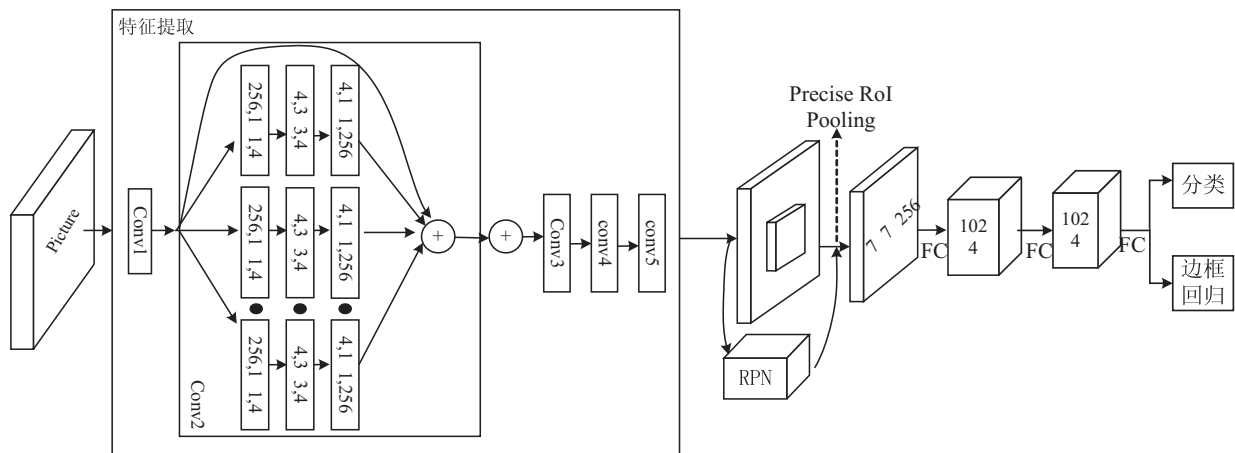


图 1 字符区域定位网络结构

对特征图上的感兴趣区域进行 RoI-Pooling 会导致定位不准确,这是因为 RoI-Pooling 进行的两次量化取整操作会带来精度损失。另外,RoI 区域中的像素是离散化的,没有梯度更新也就无法调节训练。精确的感兴趣区域池化相比 RoI-Pooling,没有任何坐标量化过程,整个区域的像素值通过积分计算得到,所有的像素值都有梯度的传递。所以能在特征图上对感兴趣区域提取更精准的特征。给定精确的感兴趣区域池化操作前的特征图 F , 设 w_{ij} 是该特征图上一个离散位置 (i, j) 处的特征,使用双线性插值,这个离散的特征图可以看作在任意连续的坐标 (x, y) 处都连续:

$$f(x, y) = \sum_{i,j} IC(x, y, i, j) \times w_{ij} \quad (1)$$

$$IC(x, y, i, j) = \max(0, 1 - |x - i|) \times \max(0, 1 - |y - j|)$$

1) 是插值系数。然后将 RoI 的一个 bin 表示为:

$$bin = \{ (x_1, y_1), (x_2, y_2) \} \quad (2)$$

其中, $(x_1, y_1), (x_2, y_2)$ 是左上角和右下角的连续坐标。给定 bin 和特征图 F , 通过计算一个二阶积分来执行池化:

$$PrPool(bin, F) = \frac{\int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y) dx dy}{(x_2 - x_1) \times (y_2 - y_1)} \quad (3)$$

1.2 字符识别网络

受文献[15]方法的启发,构建字符识别网络。网络可以视为编解码器结构,编码器由特征提取网络 ResneXt-50^[12] 和双向长短时记忆网络 (BiLSTM)^[16] 构成,解码器由加入注意力机制^[17] 的长短时记忆网络 (LSTM) 构成。网络结构如图 2 所示。

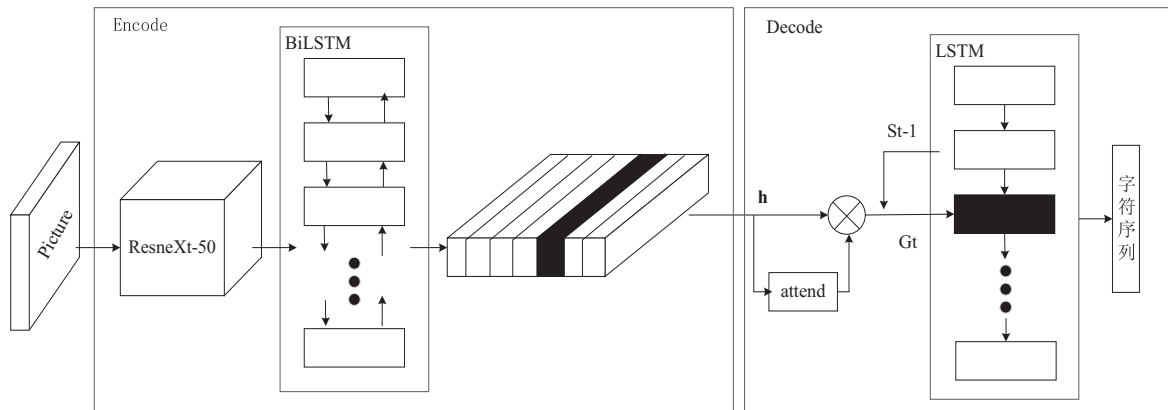


图 2 字符识别网络结构

由于 BiLSTM 的输入是向量,所以特征提取网络的输出高度应为 1 像素。而 ResNeXt-50 共有五个池化层,在高度上的步长均为 2,所以输入图片的高度应固定为 32 像素,为了不改变输入图像的纵横比,宽度为 $(W \times 32)/H$ 。其中 W, H 为裁剪的字符区域的宽度和高度。使用修改后的 ResNeXt-50 作为特征提取网络。所做修改如下:将 ResNeXt-50 最后三组卷积中的池化层步长由 2×2 改为 2×1 。这是因为字符是长条形的,即高度大于宽度,使用 2×1 的步长池化后得到的特征序列中一个像素点的感受野是长方形的,符合字符瘦高的特性。

编码器为双向长短时记忆网络,具有长期记忆功能,可以同时利用前向和后向的信息。解码器由加入注意力机制的长短时记忆网络构成。长短时记忆网络的长期记忆功能有限,当序列信息特别长时,通过多层后最初的信息会丢失,而注意力机制为特征序列的每一列分配权重,重新计算得到当前时刻的特征,此时的特征具有全文信息,有助于字符识别。

加入注意力机制的长短时记忆网络解码过程如图 2 所示,在时刻 t ,由特征序列和长短时记忆网络在时刻 $t-1$ 的状态 S_{t-1} 计算得到在时刻 t 特征序列的每一列 h_t 的权重 α ,计算公式为:

$$e_{t,i} = w^T \tanh(W_{s_{t-1}} + Vh_i + b) \quad (4)$$

$$\alpha_{t,i} = \exp(e_{t,i}) / \sum_{j=1}^n \exp(e_{t,j}) \quad (5)$$

其中, w, V, W 是训练得到的。通过式(5)得到时刻 t 输入的特征向量:

$$g_t = \sum_{i=1}^n \alpha_{t,i} h_i \quad (6)$$

t 时刻的输出为: $(S_t, h_t) = \text{rnn}(S_{t-1}, (g_t, h_{t-1}))$ 。 (g_t, h_{t-1}) 是指将两个向量拼接在一起。 h_t 用于预测 t 时刻的字符概率: $p(y_t) = \text{softmax}(Wh_t + b)$, W 和 b 是训练得到的。

最终 LSTM 的输出为 $1 \times w \times c$ 大小的概率矩阵, w 为特征图的宽度, c 为字典中字符的个数。在概率

矩阵中使用贪心搜索得到最后的字符序列。

2 实验分析

2.1 实验配置

处理器: Intel® Core™ i7-5930K CPU @ 3.50 GHz $\times 12$, 内存 16 G, 显卡 TITAN X (Pascal)。

2.2 数据集

Synth80 是人工合成的图片,可用于文字检测和识别。利用文本识别数据生成器合成类似于真实数据集的图片,共十万张。真实数据集为采集的数显仪表图片,训练集共 500 张,测试集共 50 张。

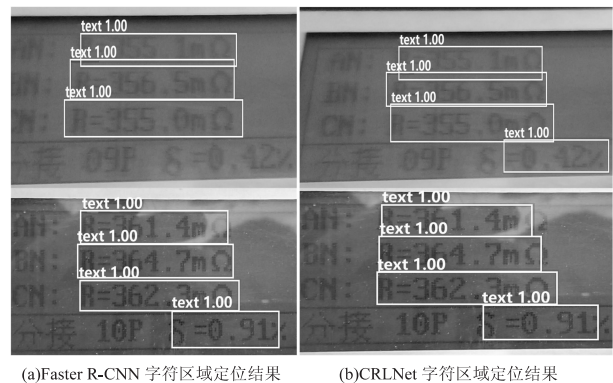
2.3 训练

训练字符区域定位模型:随机选取 synth80 中的十万张图片,制作成 Coco 格式的数据集,用于字符区域定位模型训练,得到预训练模型。在此模型基础上用真实数据集进行微调,得到最终的模型。

训练字符识别模型:使用合成的十万张图片进行预训练,使用真实的数据集进行微调,得到最终的字符识别模型。

2.4 实验结果与分析

由于图片较大,截取了图片中检测以及识别的部分用于展示结果。



(a)Faster R-CNN 字符区域定位结果 (b)CRLNet 字符区域定位结果

图 3 字符区域定位结果对比

FasterR-CNN 的部分检测结果如图 3(a) 所示,会出现漏检,定位不准确现象。改进后的算法 CRLNet

的部分检测结果如图 3(b) 所示,漏检现象减少且定位更加准确;表 1 中字符区域定位网络 CRLNet 的准确率达到 96%,比 Faster R-CNN 提高了 5 个百分点,进一步证明改进后的算法有效减少漏检现象。由实验结果发现模糊以及反光的图片易造成漏检和定位不准确现象,改进后的算法有效缓解了上述现象,但并不能完全消除。在采集图片时保证图片的清晰可以从根源上解决上述现象。

表 1 网络准确率

模型	字符区域定位 准确率/%	字符识别 准确率/%	定位+识别 准确率/%
FasterR-CNN	91	-	-
CRLNet	96	-	-
CRNet	-	98	-
DDICRNet	-	-	95

由表 1 可知,单纯的字符识别网络 CRNet 可以达到 98% 的准确率。将字符区域定位和识别模型集成在一起构成 DDICRNet,得到最终的识别准确率为 95%。部分识别结果如图 4 所示。实验结果显示,字符区域定位的准确性直接影响字符识别的效果,在保证定位准确清晰的情况下,基本可以识别正确。在实际应用中,最好的措施是在采集图片时注意遮挡,避免抖动,尽量保证图片的清晰,以提高字符区域定位算法的准确性,进而保证识别的准确性。

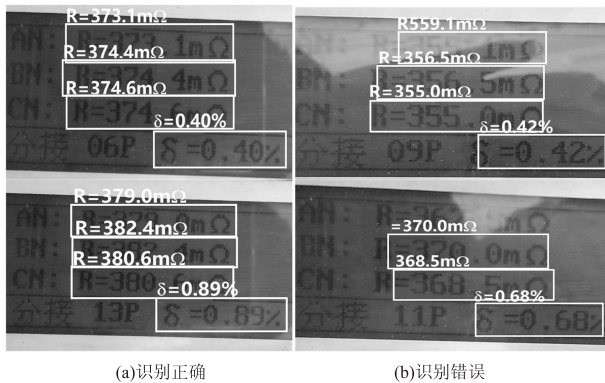


图 4 字符识别效果图

3 结束语

提出利用深度学习技术解决数显仪表字符识别的问题,取得了良好的效果,但由于网络结构较复杂,图片尺寸大(1 440 × 1 080),检测和识别一张图片需要 0.2 秒左右,无法满足实时性要求,下一步的工作将研究提高网络的识别速度。

参考文献:

[1] 董文超,舒小华,龙永红,等. LED 显示仪表的字符识别方法[J]. 湖南工业大学学报,2014,28(1):67-70.

- [2] 刘宗林,田莹,李雪滢. 一种数显仪表字符快速自动识别方法[J]. 现代计算机,2015(12):37-40.
- [3] 刘金清,喻丽春,刘引,等. 仪表字符识别中的图像处理算法研究[J]. 湖南城市学院学报:自然科学版,2016,25(2):91-93.
- [4] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Santiago, Chile: IEEE,2015:1440-1448.
- [5] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(6):1137-1149.
- [6] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Venice, Italy: IEEE,2017:2961-2969.
- [7] 刘华春. 卷积神经网络在车牌识别中的应用研究[J]. 计算机技术与发展,2019,29(4):128-132.
- [8] 周成伟. 基于卷积神经网络的自然场景中数字识别[J]. 计算机技术与发展,2017,27(11):101-105.
- [9] 陈利. 基于深度学习的车牌识别系统设计[J]. 计算机技术与发展,2018,28(6):85-89.
- [10] 缪裕青,刘水清,张万桢,等. 自然场景图像中的中文文本检测算法[J]. 计算机工程与设计,2018,39(3):804-807.
- [11] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]//Proceedings of the European conference on computer vision (ECCV). Munich, Germany: Springer,2018:784-799.
- [12] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, HI: IEEE,2017:1492-1500.
- [13] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston, Massachusetts: IEEE,2015:1-9.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA: IEEE,2016:770-778.
- [15] SHI Baoguang, YANG Mingkun, WANG Xinggang, et al. ASTER: an attentional scene text recognizer with flexible rectification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,41(9):2035-2048.
- [16] ZHOU J, XU W. End-to-end learning of semantic role labeling using recurrent neural networks[C]//Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing. Beijing, China: ACL,2015:1127-1137.
- [17] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Advances in neural information processing systems. [s. l.]: [s. n.],2014:2204-2212.