基于深度学习的变电站钢结构图纸标题栏文字检测与识别

秦辞海　顾万里

（国网上海市电力公司，上海　200120）

【摘　要】为实现变电站工程建设中钢结构与电力设备的配套控制管理，需要从大量的钢结构图纸标题栏中识别相关信息，并与实物进行匹配。针对标题栏中字体模糊、表格形式多样、信息量混杂等问题，提出了基于深度学习 CNN + RNN 模型的文本检测和 CRNN 模型的文字识别方法。对现有钢结构变电站工程施工现场钢结构数据集的检测与识别显示，该方法的检测精确率达到 80% 以上，识别准确率达到 90% 以上，均优于其他文本检测与识别方法。工程应用结果表明，该方法有效解决了因文字的大小、字体、颜色与排列方式等差异引起的特征提取困难，提高了变电站钢结构图纸标题栏文字识别的准确率。

【关键词】变电站；钢结构；文本检测；文字识别；深度学习；图纸标题栏

【中图分类号】TU17　【文献标识码】A

引言

随着社会经济的发展，越来越多的变电站采用钢结构形式，在这些变电站工程建设中，钢结构施工占据了举足轻重的位置。目前，国内工程公司通过派驻代表长期驻守施工现场配合施工来实现变电站钢结构施工相关管理。在材料及进场验收、施工安装、后期维护等重要环节，由于信息采集不及时、不准确、标准不统一等因素，对相关管理工作造成了极大的困扰，因此有必要基于数字信息化实时反馈技术，对供货情况、出厂质量与现场验收情况进行实时掌握与反馈，通过数字信息规划卸货、吊装与安装，直观掌握钢构件框架完成后的全貌。识别变电站钢结构相关图纸所列钢构件清单，生成与清单上所列构件相对应的钢构件编码，并生成后台数据库是建立实时反馈控制管理的重要环节。由于部分变电站钢结构图纸标题栏图像结构复杂、扫描不清晰，文字大小不均匀，给文本的检测与识别带来困难，传统基于光学的检测识别方法难以满足需求。

近些年随着计算机算力的增加，深度学习方法在图像识别领域得到空前发展，成为了文本检测与识别的核心算法之一。文本检测与识别的研究一直在进行，从传统文字切割光学文字识别技术 OCR[1,2]，到现在的深度学习网络模型技术，检测与识别效果有了很大的进步。黄娜君等人[3]通过对采集到的图像进行处理，分割出交通标志所在的感兴趣区域，利用深度卷积神经网络模型进行一系列的卷积和池化处理，最后通过一个全连接的 BP 网络完成分类识别，输出结果。P. He 等人[4]采用级联多个卷积模型来实现准确预测，并在此基础上开发了分层模块，但是对弯曲或者倾斜文本检测效果欠佳。B. Shi 等人[5]提出的文本检测方法 SegLink，它将文字分解为片段和链接两个类别，片段是覆盖字符或文字行的一部分，通过链接能够组合多个定向框，然后采用全卷积网络进行端到端的训练，最后在多尺度上对两类元素进行密集检测，获得检测结果，但是对文本行相距太远的情况，该方法检测效果不好。Z. Zhang 等人[6]提出采用完全卷积网络以整体的方式预测文本区域，再结合显著图和字符

【作者简介】秦辞海（1977 - ），男，硕士研究生，主要研究方向：变电站工程建设数字化研究。

成分来估计文本行假设,最后使用分类器,去除假设,但是无法准确区分边框特征的敏感程度。Wang等人[7]将大型多层神经网络的表征能力与无监督特征学习结合起来,采用端到端的方法,在定位自然场景图像中的字符区域,识别出字符方面取得显著效果,然而对中文检测识别效果不明显。这些方法在某些场景检测识别效果好,但是针对钢结构图纸中格式多样、文字模糊、大小不均等场景,检测识别效果不甚理想。

鉴于上述各种方法的局限性,本文提出了通过卷积神经网络与循环神经网络相结合的文字检测与识别方式。并将该方法运用到变电站建筑工地钢结构图纸标题栏图像文本检测识别中,解决钢结构图纸标题栏图像难检测、难识别的问题。然后将识别出的信息传输到建立的数据库中,方便变电站钢结构施工质量管控。本文研究内容主要分为两部分,一是文本检测,二是文字识别。在文本检测之前,要对输入的变电站钢结构图纸标题栏图像进行尺寸缩放、去噪等预处理; 检测出文字后,将文字区域分割出来,传入到文字识别模型中,进而得出图像文字信息,传输到指定数据库。在结果分析中,本文设计了对比实验。一是对同一训练集,将本文检测方法与其他方法的检测结果进行对比; 二是采用不同训练集训练本文的文字识别方法,最终识别结果进行对比。最后,得出检测识别钢结构图纸标题栏图像中文字的最优模型,为智能实时管控做前期准备。

1 本文模型

1.1 文本检测方法

文本检测需在图像中定位检测到文字,传统的文本检测方法在OCR领域中获得不错的效果,但是对于复杂场景的检测中完全落后基于深度学习的检测方法。基于深度学习的文本检测方法模型,泛化能力强,对高层语义特征的识别更加稳定。卷积神经网络(Convolutional Neural Networks,CNN)[8]是常用的特征提取网络,在图像文字特征提取方面取得了很好的效果。本文采用VGG16模型[9],对预处理过的图像进行文字特征提取,使用循环神经网络(Recurrent Neural Network,RNN)[10]编码水平行的文字位置信息,全连接层进行分类,再经过非最大值抑制处理,得到最终检测结果。文本检测主要目的是为了检测出文字所在的区域,生成文字图像片段,作为后续文字识别的输入。

图1是VGG16网络的结构,它所有卷积层采用3像素×3像素的小尺寸卷积核,卷积步长为1。为保证卷积后图像大小不变,图像四周各填充一个像素。所有池化层都是2像素×2像素的核,步长为2。所有隐层的激活单元都是ReLU函数。

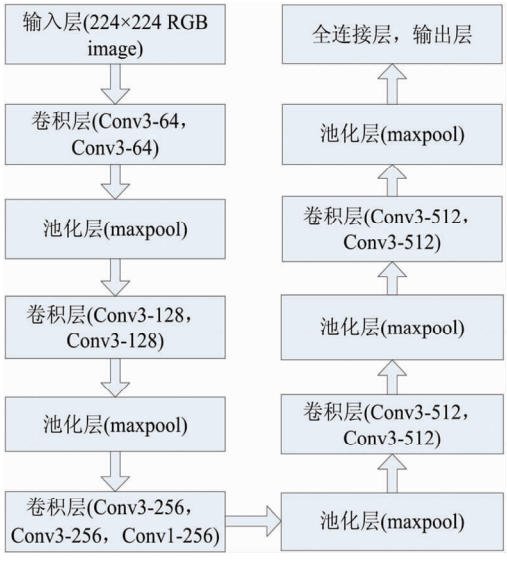


图1 VGG16网络结构

循环神经网络由输入层、循环层和输出层构成,其具有记忆功能,会记住网络在上一时刻运行时产生的状态值,并将该值用于当前时刻输出值的生成。循环神经网络的输入可抽象表示为向量序列:

$$x_1, x_2, \cdots, x_i, \cdots, x_t, \cdots \tag{1}$$

其中,x_i是向量,下标i表示时刻。网络每个时刻接收一个输入x_i,并产生一个输出y_i。y_i由之前的输入序列共同决定。在循环层中,t时刻的输出值是h_t,它由上一时刻的输出值h_{t-1}以及当前时刻的输入值x_t共同决定,即

$$h_t = f(h_{t-1}, x_t) \tag{2}$$

其中,f为激活函数,一般选用tanh函数,形如

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

标准的RNN会有梯度消失或者梯度爆炸问题,为避免这个问题,本文采用比较主流的LSTM方法[10]。LSTM是对循环层单元进行改造,主要由三个门组成:输入门、遗忘门和输出门。LSTM的网络单元结构如图2所示。图中,h_{t-1}是循环层中t-1

时刻的输出值；c_{t-1}是 t－1 时刻的状态值；x_t 是 t 时刻的输入值；h_t 是 t 时刻的输出值；输入门表示为 $I_t=\sigma(W_i[h_{t-1},x_t]+b_i)$，控制着当前 t 时刻的输入 x_t 有多少可以进入当前状态 Ct；遗忘门为 $F_t=\sigma(W_f[h_{t-1},x_t]+b_f)$，决定了上一时刻的值 h_{t-1}有多少会被传递到当前状态 Ct；输出门为 $O_t=\sigma(W_o[h_{t-1},x_t]+b_o)$，根据当前状态 Ct、上一时刻的值 h_{t-1}和当前时刻的输入 x_t 来决定当前时刻的输出 h_t；σ 是 sigmoid 函数，W_i、W_f、W_o 是权重参数，b_i，b_f，b_o 是偏置参数。通过“门”的结构，有选择性的影响循环神经网络每个时刻的状态，输入门作用于当前时刻输入值，遗忘门作用于上一时刻的输出值，二者加权和，得出当前时刻状态，最后与输出门共同决定输出值，预测出文字序列位置信息，生成文字图像片段。

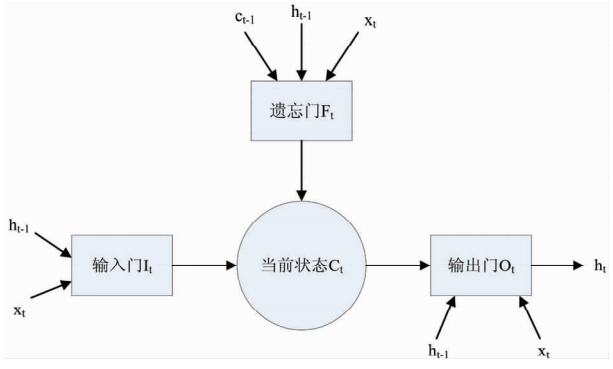


图 2　LSTM 单元结构

1.2　文字识别方法

文字识别模型复杂多样，本文主要运用基于深度学习的卷积循环神经网络（Convolutional Recurrent Neural Network，CRNN）[11]，进行钢结构图纸标题栏图像内容的文字识别，它是通过端对端地解决基于文本检测结果图像的不定长的文字序列识别问题。CRNN 将 CNN 与 RNN 相结合，把特征提取、序列建模和转录整合到统一的神经网络架构中。CRNN 网络结构如图 3 所示，CRNN 模型主要分为两个部分：一部分为特征提取，由多个卷积层、池化层和激活函数组成；一部分是序列预测，由循环层 RNN 和转录层 CTC 模型组成。CTC 是不定长字符识别，具有归纳字符连接间的特性，可解决输入序列和输出序列对应关系未知的问题。本文识别模型是使用 CNN 从文本检测结果图像中提取特征序列，传递给循环层 RNN 预测出特征序列的时序和种类，最后 CTC 是把时序和种类通过去重整合等操作转换成最终的识别结果。

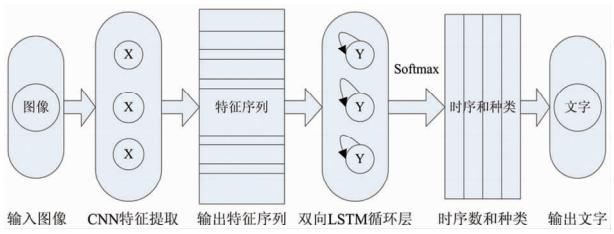


图 3　CRNN 网络结构

（其中 X、Y 代表不同的网络神经元，Softmax 是分类函数）

2　模型训练

文本检测训练模型，数据集采用 ICDAR 2017[12]，损失函数采用交叉熵与定位损失相加权的方式，即

$$L(p,u,t^u,v)=L_{cls}(p,u)+\alpha L_{loc}(t^u,v) \tag{4}$$

其中，交叉熵为：

$$L_{cls}(p,u)=-\sum p\log(u) \tag{5}$$

定位损失为：

$$L_{loc}(t^u,v)=\sum smooth(t^u-v) \tag{6}$$

式中，p 是类别概率值；u 是类别，t^u 是第 u 类预测矩形框，v 是真实矩形框，α 是人工设定的参数；$smooth$ 是一个光滑分段函数，表达式为：

$$smooth(x)=\begin{cases}0.5x^2, & |x|<1\\ |x|-0.5, & |x|\geq 1\end{cases} \tag{7}$$

文字识别训练数据集是在不同图纸中截取的文字，训练集 800 张，验证集 300 张。损失函数采用负对数条件似然函数，即

$$loss=\sum_{(I_i,l_i)\in\chi}\log p(l_i|y_i) \tag{8}$$

式中，χ 表示训练数据集，I_i 代表训练数据集中的一张图像，l_i 是对应标签序列标注，y_i 表示对应预测概率分布序列。

3　评测指标

3.1　文本检测评测指标

目标检测常见的评测指标[13]是精确率（Precision）、召回率（Recall）和 F1－度量（F1－score）。根据预测标签与实际标签定义的关系，精确率为：

$$\text{Precision}=\frac{TP}{TP+FP} \tag{9}$$

召回率为：

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

F1 – 度量为：

$$F1 = \frac{\text{Precision} \times \text{Recall} \times 2}{\text{Precision} + \text{Recall}} \tag{11}$$

其中，TP 为实际标签为正例，预测标签为正例；FP 为实际标签为负例，预测标签为正例；FN 为实际标签为负例，预测标签为负例。$F1$ – 度量越高，说明实验方法越有效。

3.2 文字识别评测指标

在文字识别中，识别效果的评测指标主要是单字识别准确率 AR（Accurate Rate）和整行识别率 LA（Line Accurate）。单字识别准确率 AR 是文本行中正确识别的字符个数占总字符个数的比例。整行识别准确率 LA 是文本行中完全识别正确行的比例，即

$$LA = \frac{N_t}{N_a} \tag{12}$$

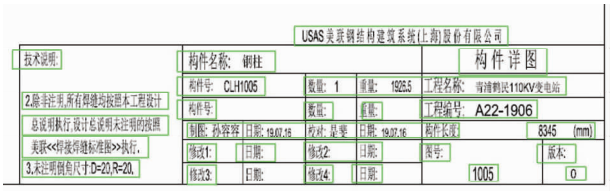
其中，N_t 指整行识别正确行数量，N_a 指整个图像测试集行数量。

4 工程应用结果与分析

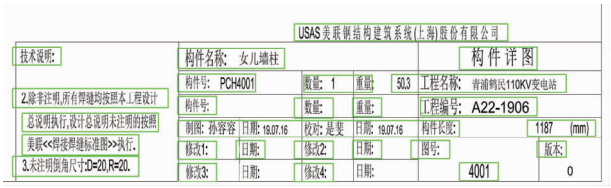
4.1 图像文本检测结果分析

本文是在 NVIDIA GeForce GTX 1070Ti 的 GPU 工作台，基于谷歌公司 Tensorflow 平台进行开发测试，训练采用 Adam 优化器，初始学习率 0.001，每 1000 次迭代，学习率下降为之前的 0.95，其他参数默认。在 ICDAR 2017 数据集上，分别对 SARI FDU[14]、Wenhao He et al[15] 和本文模型进行训练。根据训练完成的模型分别对大量经预处理的蕴藻浜—闸北 220kV 线路装设统一潮流控制器工程变电站的钢结构图纸标题栏模糊图像和清晰图像进行测试。图 4 绿色框是本文模型的检测结果，从图中可以看出，本文模型钢结构图纸标题栏模糊图像（图 4(a)）与图像清晰（图 4(b)）文本检测结果相近，基本可以正确检测出文本文字片段。

对于不同训练集训练的模型，根据上文 3.1 文本检测评测指标，其评测结果如表 1 和表 2 所示。从表中可以看出，同一数据集，使用不同的方法进行训练，最终预测的结果存在一定差异，本文检测模型的评测指标优于 SARI FDU、Wenhao He et al. 两个方法，在清晰图片上的文本评测对比更加明显。



（a）变电站钢结构图纸标题栏模糊图像文本检测结果



（b）变电站钢结构图纸标题栏清晰图像文本检测结果

图 4 本文模型变电站钢结构图纸标题栏图像文本检测结果

表 1 不同方法在 ICDAR2017 数据集上的模糊图片文本评测对比

Method	Precision(%)	Recall(%)	F1 – score(%)
SARI FDU	71.17	55.50	62.37
Wenhao He et al.	74.65	61.32	67.33
本文检测模型	80.78	69.44	74.68

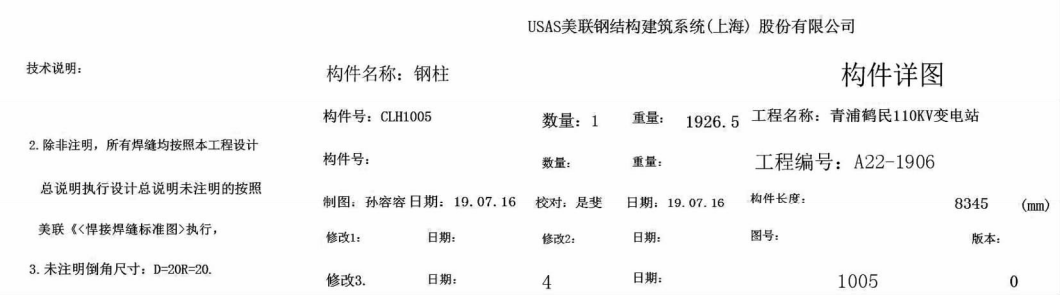
表 2 不同方法在 ICDAR2017 数据集上的清晰图片文本评测对比

Method	Precision(%)	Recall(%)	F1 – score(%)
SARI FDU	73.91	57.42	64.63
Wenhao He et al.	78.23	64.98	70.99
本文检测模型	83.10	71.54	76.89

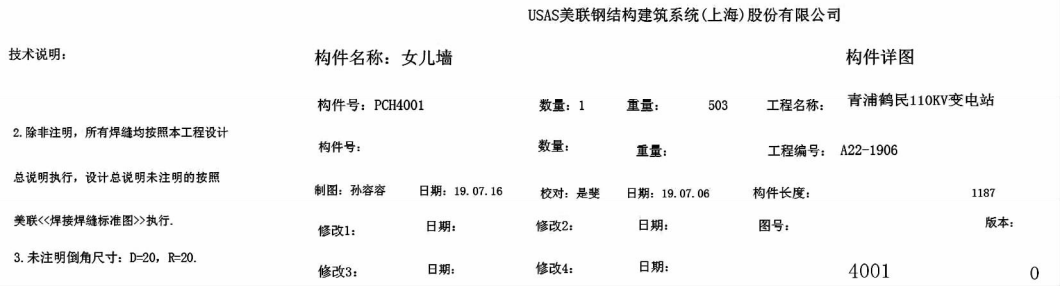
4.2 图像文本检测结果分析

检测出图像文字后，采用 CRNN 在不同数据集上训练后的模型，分别对变电站钢结构图纸标题栏模糊图像和清晰图像进行文字识别，识别结果如图 5 所示。从图中可以看出，本文识别方法基本可以正确识别出模糊图片（图 5(a)）和清晰图片（图 5(b)）的文字信息。

通过大量图像测试，根据上文 3.1 文字识别评测指标，其评测结果如表 3 和表 4 所示。从表中可以看出，通过本文数据集训练的 CRNN 模型测试钢结构图纸标题栏图像的单字识别准确率和整行识别准确率是高于 MSRA – TD500、ICPR2018 数据集。主要原因是针对测试图像，进行了针对性的训练，特别是训练集中含有大量的钢结构图纸标题栏图像。



(a)模糊图像文字识别测试结果



(b)清晰图像文字识别测试结果

图 5　CRNN 模型在本文数据集训练测试图像文字识别结果

表 3　CRNN 模型在不同数据集训练模糊图片文字识别测试结果

数据集	单字识别准确率 AR%	整行识别准确率 LA%
MSRA – TD500[16]	83.37	79.29
ICPR2018[17]	89.12	80.37
本文数据集	93.26	90.87

表 4　CRNN 模型在不同数据集训练清晰图片文字识别测试结果

数据集	单字识别准确率 AR%	整行识别准确率 LA%
MSRA – TD500[16]	89.67	82.50
ICPR2018[17]	92.12	84.45
本文数据集	96.47	91.31

5　结论

本文提出了一种 CNN 与 RNN 相结合的变电站工程钢结构图纸标题栏文本检测与文字识别模型。在文本检测中,同一训练集下分别对模糊和清晰的图像进行测试,工程应用结果表明本文所采用的方法检测精确率达到 80% 以上,整体优于文中提到的 SARI FDU、Wenhao He et al. 两个方法。

在文字识别中,通过不同的训练集训练 CRNN 模型,分别对已经检测定位的模糊图像和清晰图像文字进行识别,工程应用结果表明在测试图像中,本文数据集模型识别准确率达 90.87% 以上,识别效果高于 MSRA – TD500、ICPR2018 数据集。

本文模型有效检测、识别出了变电站钢结构图纸标题栏图像的主要信息,可有效应用于变电站建筑工地钢结构数据库的建立与实时管理。

参考文献

[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] 周翔,陈会, 张锴,等. 复杂背景下的图像文本区域定位方法研究[J]. 计算机工程与应用, 2013, 49(12): 101-105.

[3] 黄娜君,汪慧兰, 朱强军,等. 基于 ROI 和 CNN 的交通标志识别研究[J]. 无线电通信技术, 2018, 44(2): 160-164.

[4] He P, Huang W, He T, et al. Single Shot Text Detector with Regional Attention[C]. 2017 IEEE International Conference on Computer Vision(ICCV). IEEE, 2017.

[5] Shi B, Bai X, Belongie S. Detecting oriented text in natural images by linking segments[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2550-2558.

[6] Zhang Z, Zhang C, Shen W, et al. Multi-oriented Text Detection with Fully Convolutional Networks[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2016.

[7] Wang T, Wu D J, Coates A, et al. End - to - end text recognition with convolutional neural networks[C]. Proceedings of the 21st international conference on pattern recognition(ICPR2012). IEEE, 2012:3304-3308.

[8] Lecun Y, Bottou L. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998,86(11): 2278-2324.

[9] Dubey A K, Jain V. Automatic facial recognition using VGG16 based transfer learning model[J]. Journal of Information and Optimization Sciences, 2020,1-8.

[10] Gers F A, Schraudolph N N, Schmidhuber J. Learning Precise Timing with LSTM Recurrent Networks[J]. Journal of Machine Learnig Research, 2003, 3(1): p. 115-143.

[11] Shi B, Bai X, Yao C. An End - to - End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 1-1.

[12] Shi B, Yao C, Liao M, et al. ICDAR2017 Competition on Reading Chinese Text in the Wild(RCTW-17)[C]// 2017 14th IAPR International Conference on Document Analysis and Recognition(ICDAR). IEEE, 2017.

[13] 孙凯,姚旭峰,黄钢. 基于机器学习的白细胞六分类研究[J]. 软件,2020,41(10):98-101,134.

[14] Nayef N, Yin F, Bizid I, et al. ICDAR2017 Robust Reading Challenge on Multi-Lingual Scene Text Detection and Script Identification-RRC - MLT[C]//2017 14th IAPR International Conference on Document Analysis and Recognition(ICDAR). IEEE, 2017.

[15] Wenhao He, Xu - Yao Zhang, Fei Yin, Cheng-Lin Liu. Deep Direct Regression for Multi-Oriented Scene Text Detection[J]. arXiv preprint arXiv:1703.08289v1.

[16] Yao C, Bai X, Liu W, et al. Detecting texts of arbitrary orientations in natural images[C]. Computer Vision & Pattern Recognition. IEEE, 2012.

[17] Song Y, Cui Y, Hu Han, et al. Scene Text Detection via Deep Semantic Feature Fusion and Attention-based Refinement[C]. 2018 24th International Conference on Pattern Recognition(ICPR). 2018.

Text Detection and Recognition of Drawing Title Bar of Substation Steel Structure Based on Deep Learning

Qin Cihai, Gu Wanli

(*State Grid Shanghai Municipal Electric Power Company*, *Shanghai* 200120, *China*)

Abstract: In order to realize the control and management of steel structure and power equipment in the substation engineering construction, it is necessary to identify the relevant information from the title bar of a large number of steel structure drawings, and subsequently contrast them with the real structures. To deal with the blurriness of word, diversity of table and confusion of information, a deep learning method combining the CNN + RNN text detection model and the CRNN character recognition model is proposed. Carrying out the detection and recognition experiments in the existing data set of steel structures, the detection precision reaches over 80% and the recognition accuracy reaches over 90%, which is superior to other detection and recognition methods. The results of the engineering application show that this method can effectively reduce the difficulty in feature extraction caused by the differences in arrangement, size, font and color of text, which improving the accuracy of text recognition in title bar of steel structure drawings of substations.

Key Words: Substation; Steel Structure; Text Detection; Text Recognition; Deep Learning; Drawing Title Bar