

# 基于注意力机制的二次回路端子文本 检测与识别方法

钟 鸣<sup>1,2</sup>, 陶 军<sup>1,2</sup>, 阿敏夫<sup>1,2</sup>, 杨 逸<sup>1,2</sup>

(1. 内蒙古电力科学研究院, 内蒙古自治区 呼和浩特 010020; 2. 内蒙古自治区电力系统智能化电网仿真企业重点实验室,  
内蒙古自治区 呼和浩特 010020)

**摘 要:** 变电站二次回路是二次高级集成业务的基础, 采用图像识别技术对二次回路的自动特征识别、信息提取, 可实现二次回路的智能运维业务。而变电站采集的图片环境背景杂乱、分辨率低以及失真, 使得采用图像识别技术识别不规则文本极具挑战。因此, 提出一种基于注意力机制的二次回路端子文本检测与识别方法。该方法主要包含预处理、文本检测和文本识别 3 个部分, 其中文本识别部分提出一种时空嵌入编码方法, 可以更好利用图片的位置信息。在训练过程中, 相较未改进方法仅需要序列级的标注信息, 而无需额外细粒度的字符级别框或分割掩码。最后, 通过实际工作场景数据集证明该方法不仅易用、性能好, 且在识别精度上也优于其他方法。

**关 键 词:** 变电站二次回路; 图像识别; 注意力机制; 特征提取; 文本识别

DOI: 10.19781/j.issn.1673-9140.2023.03.014 中图分类号: TM863 文章编号: 1673-9140(2023)03-0132-08

## Attention mechanism-based text detection and recognition method for secondary circuit terminals

ZHONG Ming<sup>1,2</sup>, TAO Jun<sup>1,2</sup>, A Minfu<sup>1,2</sup>, YANG Yi<sup>1,2</sup>

(1. Inner Mongolia Power Research Institute, Hohhot 010020, China; 2. Inner Mongolia Autonomous Region Power System  
Intelligent Grid Simulation Enterprise Key Laboratory, Hohhot 010020, China)

**Abstract:** The secondary circuit of the substation is the basis of the secondary advanced integrated business. The automatic feature recognition and information extraction of the secondary circuit by image recognition technology can realize the secondary circuit's intelligent operation and maintenance business. However, the images collected by the substation have messy backgrounds, low resolution, and distortion, making it very challenging to identify irregular text using image recognition technology. Therefore, a text detection and recognition method of a secondary loop terminal based on an attention mechanism is proposed. This method mainly includes preprocessing, text detection, and text recognition. In the text recognition part, a spatiotemporal embedding encoding method is proposed, which can better use the picture's location information. Compared with the unimproved method, only the sequence-level annotation information is needed in the training process, and no additional fine-grained character level box or segmentation mask is needed. Finally, it is proved that the proposed method is not only easy to use and has good performance but is also better than other methods in recognition accuracy.

**Key words:** secondary circuit of substation; image recognition; attention mechanism; feature extraction; text recognition

收稿日期: 2021-05-06; 修回日期: 2021-07-06

基金项目: 内蒙古自治区科技关键技术攻关计划项目(2019GG373)

通信作者: 钟 鸣(1987—), 男, 硕士, 工程师, 主要从事继电保护等研究; E-mail: zhongning005@163.com

随着社会智能化的蓬勃发展,电力系统在日常生活和社会生产中发挥的作用与日俱增。变电站二次回路作为变电站的关键与核心部分之一,其存在设备屏蔽柜数量多、外部接线种类繁多等诸多问题,从而导致维修复杂度和维修成本高;同时其关系着继电保护设备能否正常跳闸出口、测控能否正常采集信号等基础功能,是变电站所有二次高级集成业务的基础。若能通过图像识别技术自动对二次回路进行特征识别、信息提取,可实现二次回路的智能运维业务,为变电站其他高级业务的安全可靠运行提供基础支撑。

文献[1]中图像识别的对象是敞开式一、二次设备,其特征相对比较单一,业务信息内容较少,图像识别的应用也相较容易实现。二次回路由设备板卡、端子、端子排、压板、空开、二次电缆、光缆等连接模块和线缆组成,回路组成复杂,各模块特征多变,业务信息丰富,单纯应用图像识别技术难以简单地通过寻找固定的图像特征完成业务识别和信息解析,二次回路核心部件是汇转各线缆的端子排,研究通过图像识别技术高效实现对二次回路端子排等核心部件的连接关系、业务信息的识别读取是图像识别技术应用于二次回路智能运维的关键。需要指出的是,随着智能电网的提出,智能化的概念逐渐成为变电站的发展目标。目前国内针对变电站屏柜接线应用图像识别展开了研究,文献[2]通过训练屏柜标签字符定位网络和字符识别网络,可有效提高屏柜接线信息识别的准确率和识别时间的技术。尽管已有相关研究,但是对于图片环境杂乱情景下的二次回路端子文本识别精度还有待提高。因此本文提出一种变电站二次回路端子的文本识别方法,以期提高变电站检修和运维的工作效率。

近年来基于深度学习的方法在文本检测与识别领域均取得了不错的效果<sup>[3]</sup>。在文本检测领域,现有研究大都将文本区域视为对象,采用通用对象检测的框架,如 Faster R-CNN、单步多框目标检测(single shot multibox detector, SSD)等<sup>[4]</sup>。此类方法应用多个分支来预测偏移量和边界框的数量。为了解决文本检测中易出现形状不规则的难点,RRPN<sup>[5]</sup>将带有角度信息的倾斜锚点和旋转 RoI 池化层;PMTD<sup>[6]</sup>通过执行软掩码分割并在网络末端

添加平面聚类模块来改进 Mask R-CNN<sup>[7]</sup>。而对于文本识别算法,通常的思路是先用卷积神经网络(convolutional neural network, CNN)提取图像特征,再用长短期记忆(long short term memory, LSTM)获得字符序列,如 CRNN<sup>[8]</sup>、RobustScanner<sup>[9]</sup>。迄今为止,也有少量文献尝试将现有的文本检测和识别方法应用于电流互感器二次回路端子领域。文献[10]提出了一种电流互感器二次回路端子故障诊断方法。文献[11]提出变电站二次系统设备屏柜线标智能识别系统,但它采用了高效且准确的场景文字检测算法(an efficient and accurate scene text detector, EAST)<sup>[12]</sup>作为其文本识别方法,其识别结果仍有待提高。

针对变电站二次回路端子人工检测效率低且易漏检的难点,本文提出一种基于注意力机制的文本检测与识别算法。首先提出一个完整的文本检测和识别过程框架,可以获取输入图像中所有文本的位置和内容;然后针对基于注意力的文本识别模块,设计一种新颖的时空嵌入模块(spatial temporal embedding, STE)。最后通过实际工作场景数据集证明该方法不仅易用、性能好,而且在识别精度上也优于其他方法。

## 1 算法框架与原理

基于注意力机制的变电站二次回路端子文本识别方法由文本预处理模块、文本检测模块和文本识别模块 3 个部分构成,在预处理模块中,输入图像将被调整为固定比例并伴随随机的数据增强,其目的是为了丰富图像训练集,能更好地提取图像特征,泛化模型(防止模型过拟合)以及提升模型的鲁棒性。文本检测模块的目标是有效地检测出文本的位置。文本识别模块的目标则是识别检测区域中的字符。整个方法的流程如图 1 所示。

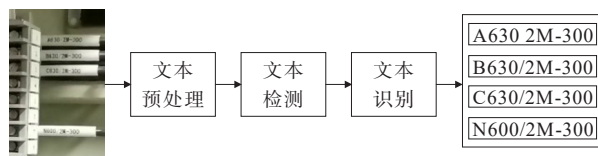


图 1 所提方法流程

Figure 1 Pipeline of our proposed method

## 2 文本检测模块

### 2.1 Mask R-CNN

Mask R-CNN是基于Faster R-CNN演进改良后的深度神经网络模型<sup>[13-14]</sup>。为了实现图像输入输出之间的像素对齐,Mask R-CNN增加了一个简洁非量化的层——RoIAlign层,能提高10%~50%的掩码精确度,目前在图像识别中对物体识别表现突出。

Mask R-CNN可分为图2中所示5个部分,其采用了与Faster R-CNN相同的2个阶段流程。第1阶段提出候选对象边界框;第2阶段在预测类和边界框偏移的同时,Mask R-CNN也为每个RoIPool输出一个二进制掩码。

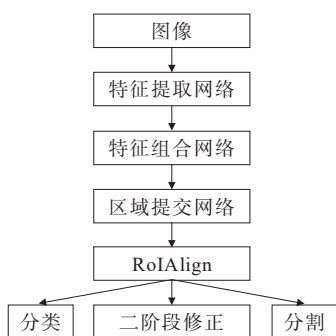


图2 Mask R-CNN 结构简化示意

Figure 2 Mask R-CNN structure

Mask R-CNN保留了Fast R-CNN的优点,并行应用边界盒分类和回归(这在很大程度上简化了原始R-CNN的多级管道)。图2中特征提取网络是整个神经网络中的核心网络,其计算量最大。本文以ResNet101<sup>[15]</sup>举例,取其输出的4个特征图,分别记作 $R_2$ 、 $R_3$ 、 $R_4$ 、 $R_5$ ,表示图像中不同深度的特征。随后在特征组合网络中将 $R_2$ 、 $R_3$ 、 $R_4$ 、 $R_5$ 重新组合生成新的特征图。其包含原图像中所有的特征信息。采用FPN来组合特征图 $R_2$ 、 $R_3$ 、 $R_4$ 、 $R_5$ 得到新特征图 $C_2$ 、 $C_3$ 、 $C_4$ 、 $C_5$ 、 $C_6$ 。对于 $i=5,4,3,2;U_6=0$ ,特征组合过程如下:

$$\begin{cases} C'_i = \text{sum}(U_{i+1}, \text{conv}(R_i)) \\ U_i = \text{upsample}(C'_i) \\ C_i = \text{conv}(C'_i) \end{cases} \quad (1)$$

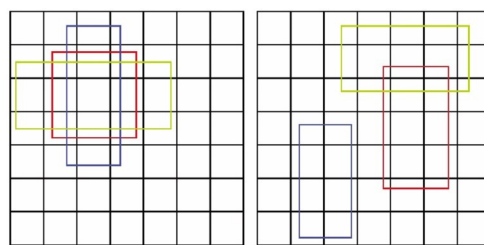
$$C_6 = \text{pooling}(C_5) \quad (2)$$

式中,sum为括号中元素进行对位求和;conv为卷积操作;upsample为将特征图 $C'_i$ 的长和宽扩大为原来2倍的升采样操作;pooling表示stride为2的最大池化操作。

区域提交网络的目的是采用Anchor技术在特征图上计算出能表示物体在图像中位置的候选框。区域提交网络是一个轻量的神经网络,通过滑动窗口扫描特征图进行卷积操作,结合不同的尺寸与长宽比,生成互相重叠区域,即Anchor,并给出每个Anchor默认预置的位置信息,以便于后续提取候选区域。例如对新特征图 $C_2$ 、 $C_3$ 、 $C_4$ 、 $C_5$ 、 $C_6$ 中每个点,都会以其坐标位置为中心,然后生成不同的宽、高的多个Anchor。随后以RPN网络回归得到的值来修正每个Anchor的中心、宽和高,从而得到新的边界框。Anchor修正示意如图3所示,图3(a)表示某特征图中的一个特征点预设的3个Anchor;图3(b)为前者中预设的3个Anchor经RPN网络回归值修正变换得到的结果。Anchor修正计算公式如下:

$$\begin{cases} x = (1 + \Delta x) \cdot x \\ y = (1 + \Delta y) \cdot y \\ w = \exp(\Delta w) \cdot w \\ h = \exp(\Delta h) \cdot h \end{cases} \quad (3)$$

式中, $x$ 、 $y$ 分别为Anchor中心的横坐标和纵坐标, $w$ 、 $h$ 分别为Anchor的宽和高。



(a) 预设图

(b) 修正图

图3 Anchor修正过程示意

Figure 3 Anchor correction process

传统方法会根据获得的候选框位置从原图中剪切出相应的区域,接着对此区域分类和分割。将特征图 $C_2$ 、 $C_3$ 、 $C_4$ 、 $C_5$ 、 $C_6$ 作为功能性网络的输入,利用RoIAlign算法直接裁剪得到与候选框对应位置

的特征,同时采用池化以及双线性插值使特征变换为均匀大小(Mask R-CNN使用输入 $7\times 7$ 大小进行两级分类和校正任务, $14\times 14$ 作为分割任务的大小)。双线性插值法是一种将离散前的插值过程转化为连续后的插值过程。在获得与该区域相对应的每个候选框的相同大小的特征后,它们被用作一系列称为头部(head)功能网络的输入,用于后续计算。分类头部通常由全连接层和 Softmax 层组成。采用 Faster R-CNN 对不同类别回归,用于候选框二次矫正。

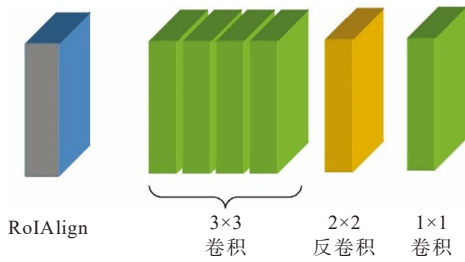


图 4 FCN 分割网络

Figure 4 FCN segmentation network

如图 4 所示,分割网络采用 $14\times 14$ 的 RoIAlign 输出特征作为输入,随后通过 4 个卷积核为 $3\times 3$ 的卷积层,紧接着通过 1 个卷积核为 $2\times 2$ 的反卷积层将输出变为 $28\times 28$ 的尺寸,最后通过 1 个卷积核为 $1\times 1$ 的卷积层与 Sigmoid 激活层,得到的输出值代表某类候选框形状的前后景置信度。

### 2.2 基于金字塔标签的 PMTD 模型

将类别 $\in\{0,1\}$ 的掩码硬标签细化为分数 $\in[0,1]$ 的软金字塔标签,以便金字塔掩码文本检测器(pyramid mask text detector, PMTD)可以从数据中捕获形状和位置信息。具体步骤为:将文本区域的中心指定为金字塔的顶点,理想值得分 $S_p=1$ ,将文本区域的边界指定为金字塔的底部边缘。使用线性插值来填充金字塔的每个三角形边,如图 5 所示。

给定四边形的 4 个角点 $A(x_a, y_a)$ 、 $B(x_b, y_b)$ 、 $C(x_c, y_c)$ 、 $D(x_d, y_d)$ ,点 $P(x_p, y_p)$ 的得分值 $S_p$ 可以按如下方式计算,首先文本区域的中心 $O(x_o, y_o)$ 可以计算如下:

$$x_o = (x_a + x_b + x_c + x_d) / 4 \quad (4)$$

$$y_o = (y_a + y_b + y_c + y_d) / 4 \quad (5)$$

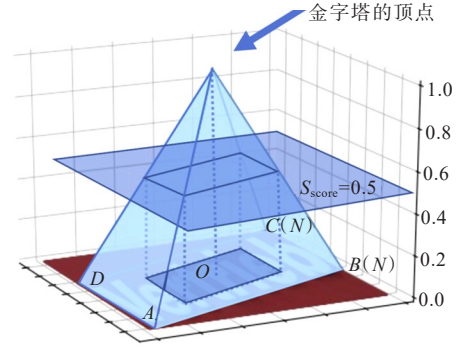


图 5 软金字塔标签的生成

Figure 5 Soft pyramid label generation

对于 $R_{OAB}$ 、 $R_{OBC}$ 、 $R_{OCD}$ 、 $R_{ODA}$ 的每一个区域 $R_{OMN}$ (两根射线 $OM$ 和 $ON$ 之间的区域, $MN$ 表示点 $ABCD$ 的任意两点), $\overrightarrow{OP}$ 都可被唯一地分解为

$$\overrightarrow{OP} = \alpha \overrightarrow{OM} + \beta \overrightarrow{ON} \quad (6)$$

$$\begin{bmatrix} x_p - x_o \\ y_p - y_o \end{bmatrix} = \begin{bmatrix} x_m - x_o & x_n - x_o \\ y_m - y_o & y_n - y_o \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (7)$$

由式(7)可以求得 $\alpha$ 和 $\beta$ 为

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} x_m - x_o & x_n - x_o \\ y_m - y_o & y_n - y_o \end{bmatrix}^{-1} \begin{bmatrix} x_p - x_o \\ y_p - y_o \end{bmatrix} \quad (8)$$

点 $P$ 所属的区域 $R_{OMN}$ 需要满足以下条件:

$$\alpha \geq 0 \text{ and } \beta \geq 0 \quad (9)$$

从而得到 $S_p$ 的计算公式如下:

$$S_p = \max(1 - (\alpha + \beta), 0) \quad (10)$$

本文所提方法的文本检测模块采用一种基于 Mask R-CNN 的 PMTD 模型,如图 6 所示。该模型不仅采用二进制文本掩码,而且还能在实现位置感知的同时执行像素级的回归,使每个文本生成软文本掩码,这种掩码模式相对于二进制文本掩码具有更多的有效信息。针对文本检测框的生成, PMTD 将二维软掩码投影至三维空间中,采用平面聚类算法可在三维形状的基础上预测出最优文本检测框。

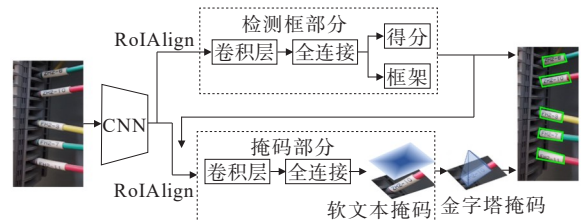


图 6 文本检测网络结构

Figure 6 Text detection network structure

### 3 文本识别模块

文本识别模块采用编码器—解码器架构。其编码器由一个深度卷积神经网络和一个基于LSTM的序列模块组成,旨在将输入图像映射到固定长度的整体特征<sup>[16]</sup>。解码器是另一个序列模块,它基于二维注意力机制逐步恢复字符序列。

#### 3.1 编码器

如图7所示,输入图像首先输入一个深度卷积神经网络,得到一个大小为 $H \times W \times D$ 的二维特征图,其中 $H$ 、 $W$ 分别表示高度和宽度, $D$ 是通道数。需要注意的是,特征图不仅是LSTM编码器的输入,也是解码过程中二维注意力网络的输入。

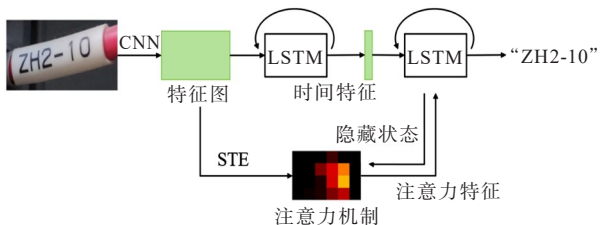


图7 文本识别网络结构

Figure 7 Text recognition network structure

为了更好地利用序列信息,将特征图输入到LSTM模型,神经单元大小设置为512。在每个时间 $t$ ,LSTM编码器接收一系列二维特征图,然后沿垂直轴进行最大池化,并更新其隐藏状态 $h$ 。经过 $W$ 步后,得到最终的隐藏状态 $h$ ,它被视为输入图像的整体表示并输入到解码器。

#### 3.2 解码器

解码器是另外一个与编码器具有相同架构的LSTM模型。在解码过程中,在每个时间步 $t$ ,解码器LSTM模型首先根据最后一个隐藏状态 $h_{t-1}$ 和当前输入 $x$ 生成一个隐藏状态 $h_t$ 。

$$h_t = \text{LSTM}(h_{t-1}, x_t), x_t = \begin{cases} \langle \text{star } t \rangle, & \text{if } t = 0 \\ y_{t-1}, & \text{if } t > 0 \end{cases} \quad (11)$$

式中, $\langle \text{star } t \rangle$ 代表字符序列开始的特殊标记; $y$ 为最后解码步骤的输出。

#### 3.3 注意力机制

输入图像映射得到的不同特征值对神经网络的准确识别影响是不同的,因此为了提高识别准确率,在文本识别模块引入注意力机制。其分2步进行计算:一是在所有的输入值上计算注意力分布值 $A'_{ij}$ ;二是对单个输出值计算输入信息的加权平均。具体步骤为:一般将 $h$ 视为一个查询向量,将特征图的每个特征向量作为一个键值,来计算一个注意力机制图。

$$A'_{ij} = \text{softmax}(h_i^T \cdot F_{ij}) \quad (12)$$

其中, $F_{ij}$ 表示位置在 $(i,j)$ 的特征向量。但是这种计算注意力的方式实际上忽略了位置信息,为使得模型能够在不同的时间步长上关注特征图的不同位置。 $A'_{ij}$ 表示位置在 $(i,j)$ 的注意力机制图结果。采用Transformer<sup>[17-18]</sup>中的位置编码方式,本文为解码器设计了一种新颖的时空嵌入模块。具体步骤为:首先将空间嵌入(spatial embedding)添加到特征图,将时间嵌入(temporal embedding)添加到 $h$ ,然后再进行线性变换:

$$P_{E(\epsilon_{\text{pos}}, 2j)} = \sin(\epsilon_{\text{pos}} / 10\,000^{2j/d_{\text{model}}}) \quad (13)$$

$$P_{E(\epsilon_{\text{pos}}, 2j+1)} = \cos(\epsilon_{\text{pos}} / 10\,000^{2j/d_{\text{model}}}) \quad (14)$$

$$T_{\text{SE}_{ij}} = P_{E(\epsilon_{\text{pos}}, 2j)}(x_{ij}) \quad (15)$$

$$T_{\text{TE}(t)} = P_{E(\epsilon_{\text{pos}}, 2j+1)}(x_t) \quad (16)$$

式中, $\epsilon_{\text{pos}}$ 表示位置, $j$ 是维度, $d_{\text{model}}$ 对应通道数,其中 $P_{E(\epsilon_{\text{pos}}, 2j)}$ 表示位置 $\epsilon_{\text{pos}}$ 、维度为 $2j$ 的位置编码结果, $T_{\text{SE}_{ij}}$ 、 $T_{\text{TE}(t)}$ 分别表示嵌入空间特征图、时间特征图的结果。

$$F'_{ij} = W_s \cdot (F_{ij} + T_{\text{SE}_{ij}}) + b_s \quad (17)$$

$$h'_t = W_t \cdot (h_t + T_{\text{TE}(t)}) + b_t \quad (18)$$

式中, $F'_{ij}$ 和 $h'_t$ 分别为经空间嵌入和时间嵌入后的位置增强表示, $W_s$ 、 $W_t$ 为线性变换的权重, $b_s$ 和 $b_t$ 为线性变换的偏置。

接下来,将这2个位置增强表示连接起来,并使用回归层对注意力图进行回归:

$$A'_{ij} = \text{softmax}(W_{\text{att}} \cdot \text{concat}(F'_{ij}, h'_t) + b_{\text{att}}) \quad (19)$$

式中,softmax表示softmax函数,concat表示矩阵的拼接, $W_{\text{att}}$ 和 $b_{\text{att}}$ 为归一化的权重和偏置。

然后,将注意力特征向量计算为特征图的加权和  $g_i$ :

$$g_i = \sum_{ij} A_{ij} \cdot F_{ij} \quad (20)$$

最后,使用带有权重和偏置参数 ( $W_{\text{cls}}, b_{\text{cls}}$ ) 的线性层将向量分类为一个字符:

$$y_i = \text{softmax}(W_{\text{cls}} \cdot g_i + b_{\text{cls}}) \quad (21)$$

## 4 算例分析

### 4.1 数据集与评估准则

本文采用某地变电站一个由 1 502 张图片和 10 120 个人工标注的文本实例组成的数据集来验证所提方法的有效性。在数据集中随机选择 1 200 张图片作为训练集,其余的图片作验证集。对于文本检测任务,设定检测出的文本与真实文本的交并比 (intersection-over-union, IoU) 阈值作为检测结果是否正确的判断指标,接着采用精确度 (precision)、召回率 (recall) 和 F-score 评分来衡量算法性能。对于文本识别任务,以对整个文本字符识别的准确率作为评估指标。

### 4.2 仿真结果分析

本文所提出的方法是在 PyTorch 中实现的,实验在一个 Nvidia GeForce RTX3090 GPU 上进行,显存为 24 GB。对于文本检测模型,总共训练了 40 轮次 (epoch),使用 SGD 作为优化器,批次大小为 64。初始学习率为 0.05,并在第 20 个、第 32 个轮次中衰减至之前的十分之一。对于文本识别模型,在训练时使用 Adam 优化器,批处理量为 32,学习率最初设定为 0.001,每 100 次迭代的衰减率为 0.8,直到误差达到  $10^{-5}$ 。

为验证本文所提方法的有效性和优势,同样采用本文数据集进行对比实验,选择 CTPN<sup>[19]</sup>、EAST<sup>[12]</sup>、RRPN<sup>[5]</sup> 这 3 种方法与本文方法进行对比,均采用 ResNet101 结构模型进行训练,得到在 IoU 阈值分别为 0.7、0.8、0.9 这 3 种情况下的对比实验结果,如图 8 所示,可知在不同 IoU 下本文方法最优,且在 0.7 时取得最优 F-score,此时检测结

果如表 1 所示,选取 CRNN<sup>[8]</sup> 和 RobustScanner<sup>[9]</sup> 2 种文本识别方法与本文方法进行对比,识别结果如表 2 所示。

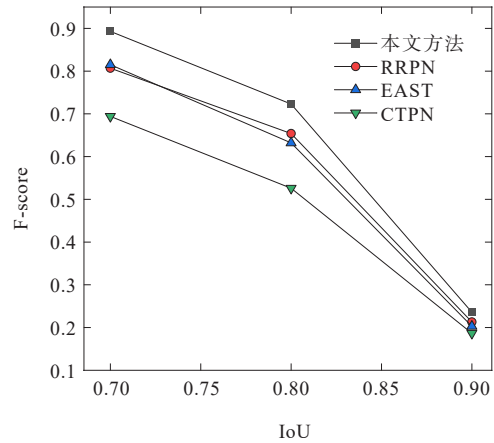


图 8 不同 IoU 下的 F-score 对比

Figure 8 Comparison of F-score under different IoU

表 1 文本检测结果

Table 1 Text detection performance

方法	精确度	召回率	F-score
CTPN	0.75	0.63	0.69
EAST	0.83	0.80	0.82
RRPN	0.82	0.79	0.80
本文检测方法	0.91	0.87	0.88

从表 1 可知,本文检测方法优于上述其他方法。根据表 1 中结果综合评价,对比方法中 CTPN 表现最差,EAST、RRPN 这 2 种方法的检测效果相差不大,而本文所提方法相较此 3 种方法在精确度、召回率、F-score 评分这 3 种评价指标上均有提高。

表 2 文本识别结果

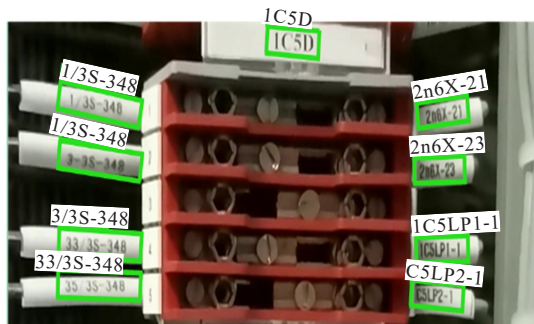
Table 2 Text recognition performance

方法	准确率/%
CRNN	81.54
RobustScanner	89.19
本文识别方法	90.08
本文识别方法(添加 STE)	91.92

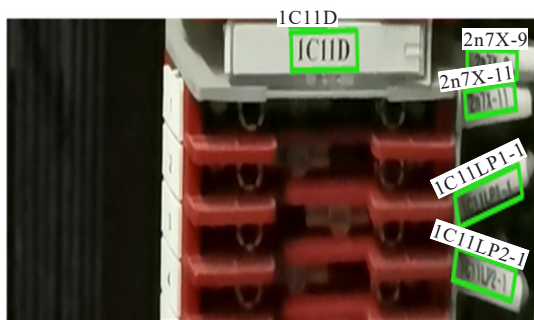
根据表 2 中的识别结果可知,本文所提识别方法表现最佳,分别在准确率上超过了对比方法 CRNN 和 RobustScanner 10.38% 和 2.73%。除此之

外,可以得知模型增加了STE模块后在实际数据集上检测准确率提高了大约1.84%,此结果证明了本文提出的STE模块的有效性。

图9展示了经本文方法识别的二次回路端子文本检测与识别的结果,可以看到检测框可以很好地适应文本不同角度的倾斜与变形,最终得到正确的文本识别结果。



(a) 示例1



(b) 示例2

图9 文本检测与识别结果示例

Table 9 Example of text detection and recognition results

## 5 结语

本文提出了一种基于注意力的变电站二次回路终端的文本检测和识别方法,可以自动获得文本的位置和字符。整个算法流程可分为预处理模块、文本检测模块和文本识别模块3个步骤。具体步骤如下:采用金字塔掩码文本检测器作为文本检测模块,基于LSTM的序列模块作为基本文本识别模型。在解码器中提出了一种新的空间时间嵌入层,以提高网络捕捉重要信息的能力。最后通过实际场景中采集的数据集验证了本文所提

方法的可行性和有效性。通过目前的研究继续探索二次回路故障预警与检测将是未来工作的重点方向。

## 参考文献:

- [1] 律方成,牛雷雷,王胜辉,等.基于优化YOLOv4的主要电气设备智能检测及调参策略[J].电工技术学报,2021,36(22):4837-4848.  
LÜ Fangcheng, NIU Leilei, WANG Shenghui, et al. Intelligent detection and parameter adjustment strategy of major electrical equipment based on optimized YOLOv4 [J]. Transactions of China Electrotechnical Society, 2021, 36 (22): 4837-4848.
- [2] 王磊,黄力,张礼波,等.基于联合训练方法的变电站屏柜接线故障诊断技术[J].供用电,2020,37(5):85-90.  
WANG Lei, HUANG Li, ZHANG Libo, et al. Substation panel cabinet wiring fault diagnosis technology based on joint training method[J]. Distribution & Utilization, 2020, 37 (5): 85-90.
- [3] 苏玉刚,阳剑,戴欣,等.基于TensorFlow神经网络的MCR-WPT系统负载与互感识别方法[J].电力系统自动化,2021,45(18):162-169.  
SU Yugang, YANG Jian, DAI Xin, et al. Tensorflow neural network based load and mutual inductance identification method for magnetic coupling resonant wireless power transfer system[J]. Automation of Electric Power Systems, 2021, 45(18): 162-169.
- [4] 杨剑锋,秦钟,庞小龙,等.基于深度神经网络的输电线路异物入侵监测和识别方法[J].电力系统保护与控制,2021,49(4):37-44.  
YANG Jianfeng, QIN Zhong, PANG Xiaolong, et al. Foreign body intrusion monitoring and recognition method based on Dense-YOLOv3 deep learning network[J]. Power System Protection and Control, 2021, 49(4): 37-44.
- [5] MA J Q, SHAO W Y, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [6] LIU J C, LIU X B, SHENG J, et al. Pyramid mask text detector[EB/OL]. [2021-03-02]. <https://arxiv.org/abs/1903.11800>.
- [7] 彭秋辰,宋亦旭.基于Mask R-CNN的物体识别和定位[J].清华大学学报(自然科学版),2019,59(2):135-141.  
PENG Qiuchen, SONG Yixu. Object recognition and

- localization based on Mask R-CNN[J]. Journal of Tsinghua University (Science and Technology), 2019, 59(2): 135-141.
- [8] SHI B G, 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, 39(11): 2298-2304.
- [9] YUE X Y, KUANG Z H, LIN C H, et al. RobustScanner: dynamically enhancing positional clues for robust text recognition[C]//European Conference on Computer Vision, Glasgow, uk, 2020.
- [10] 袁滢钦. 智能变电站二次回路检测技术研究[D]. 广州: 广东工业大学, 2020.
- YUAN Ganqin. Research on secondary circuit detection technology of intelligent substation[D]. Guangzhou: Guangdong University of Technology, 2020.
- [11] 周健, 李岩, 张克声. 变电站二次系统设备屏柜线套标签智能识别系统[J]. 机械与电子, 2018, 36(11): 67-70.
- ZHOU Jian, LI Yan, ZHANG Kesheng. Intelligent identification system for equipment screen cabinet line label of substation secondary system[J]. Machinery & Electronics, 2018, 36(11): 67-70.
- [12] ZHOU X Y, YAO C, WEN H, et al. EAST: an efficient and accurate scene text detector[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA, 2017.
- [13] 楚瀛, 李帅奇. 基于联合特征的变电站设备图像识别[J]. 电力科学与技术学报, 2018, 33(1): 102-107.
- CHU Ying, LI Shuaiqi. Image recognition of substation equipment based on joint feature[J]. Journal of Electric Power Science and Technology, 2018, 33(1): 102-107.
- [14] 吴军, 白梁军, 董晓虎, 等. 基于 Cascade R-CNN 算法的输电线路小目标缺陷检测方法[J]. 电网与清洁能源, 2022, 38(4): 19-27+36.
- WU Jun, BAI Liangjun, DONG Xiaohu, et al. Transmission line small target defect detection method based on cascade R-CNN algorithm[J]. Power System and Clean Energy, 2022, 38(4): 19-27+36.
- [15] 李俊, 任景, 王晔琳, 等. 基于改进 U-net 和 CNN 的绝缘子自爆检测方法研究[J]. 智慧电力, 2021, 49(8): 98-103.
- LI Jun, REN Jing, WANG Yelin, et al. Insulator self-explosion detection method based on improved U-net and CNN[J]. Smart Power, 2021, 49(8): 98-103.
- [16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017.
- [17] 王彦彪, 陈振勇, 郭文萍, 等. 基于双注意力机制优化 CNN 架构的 GIS 局部放电模式识别[J]. 电力科学与技术学报, 2022, 37(2): 22-29.
- WANG Yanbiao, CHEN Zhenyong, GUO Wenping, et al. PD pattern recognition for GIS based on CNN architecture optimized by the double attention mechanism[J]. Journal of Electric Power Science and Technology, 2022, 37(2): 22-29.
- [18] TIAN Z, HUANG W L, HE T, et al. Detecting text in natural image with connectionist text proposal network[C]//European Conference on Computer Vision, Amsterdam, The Netherland, 2016.