

DOI:10.12158/j.2096-3203.2022.05.026

# 基于深度学习的电力设备铭牌文本识别

王逸凡, 王佳宇, 仲林林, 高丙团

(东南大学电气工程学院, 江苏 南京 210096)

**摘要:**电力设备铭牌包含丰富的设备信息,通过图文识别技术获取设备铭牌信息,可更加高效快捷地完成电力设备的信息统计、台账校核等工作,也有利于提高电力系统的设备管理水平。针对电力设备铭牌与普通图像文本识别差异较大的特殊应用场景,文中提出一种基于深度学习的电力设备铭牌信息自动识别算法。该算法由铭牌检测、文本检测、文本识别三部分组成。通过改进损失函数设计、增加文本识别结果纠正、人工合成文本图像等方式,使得铭牌检测模型在测试集上的平均精度均值达到92.2%,文本检测模型在测试集上的 $F_1$ 值达到91.2%,文本识别模型的字符识别准确率达到94.0%,文本行识别准确率达到82.3%。

**关键词:**电力设备铭牌;计算机视觉;深度学习;目标检测;文本检测;文本识别

**中图分类号:**TM07;TP18

**文献标志码:**A

**文章编号:**2096-3203(2022)05-0210-09

## 0 引言

我国电网规模日益增长,电力设备的种类和数量快速增加,对电力设备信息的管理也日趋重要<sup>[1-2]</sup>。电力设备铭牌记录了设备的参数、型号等重要信息,但目前我国电力系统设备台账中普遍存在数据丢失、准确性低等问题。而电力系统日常运维中会产生大量包含铭牌信息的图像数据,因此采用图文识别技术快速准确识别设备铭牌信息,可提高电力设备统计、台账校核等工作的效率,提高电力系统的设备管理水平。

电力设备铭牌信息识别可细分为铭牌检测、文本检测和文本识别。其中铭牌检测指在图片中定位铭牌区域,去除背景干扰;文本检测指从铭牌图像中检测出文本连通域,生成文本行;文本识别指从文本行中识别出包含语义信息的文本字符。

传统的文本检测方法主要包括基于文本连通域的方法和基于滑动窗口的方法<sup>[3]</sup>。2类方法都能实现对一般文本的检测,但前者鲁棒性较差,难以检测背景复杂、对比度低等情况下的文本<sup>[4]</sup>,后者计算复杂,往往需要大量训练样本<sup>[5-6]</sup>。

近年来,基于深度学习的图像文本识别技术快速发展<sup>[7-8]</sup>。文献[9]提出一种基于细小候选框的文本区域建议网络(connectionist text proposal network, CTPN),通过同时预测候选框的位置和文本区域得分,得到文本区域并利用循环神经网络(recurrent neural network, RNN)合并得到文本检测结果。该方法只能检测基本水平的文本,而由于拍摄

等原因,电气设备铭牌文本往往有较大倾斜。片段连接(segment linking, SegLink)模型<sup>[10]</sup>同样基于小文字块检测文本,其在单步多盒检测器(single shot multibox detector, SSD)<sup>[11]</sup>的基础上加入方向信息,可以检测任意方向的文本,但无法检测间隔过大的文字块,无法适应电力设备铭牌文字间隔差异巨大的特点。兼具效率和精度的场景文本检测器(efficient and accuracy scene text detector, EAST)模型<sup>[12]</sup>直接对文本行进行预测,在速度、精度方面都有较好表现,且能适应倾斜文本,但在较长文本上会出现检测不完整的问题。

文本识别模型的输入一般为只包含文本信息的图片。文献[13]基于连接时序分类器(connectionist temporal classification, CTC)提出一种经典的文本识别模型,后续众多文本识别方法都沿用了类似结构。该模型使用卷积神经网络(convolutional neural network, CNN)提取文本图像的特征,使用RNN学习文本序列的标签分布,再通过CTC转换为识别结果。此外,注意力机制也被应用到文本识别领域,去耦注意力网络(decoupled attention network, DAN)<sup>[14]</sup>将注意力的对齐阶段从解码器中解耦,在手写和场景文本识别上都取得了极佳效果。

然而,电力设备铭牌图像存在其特殊性,有信息密度高,图像背景复杂,无统一格式,钢印文字模糊等特点,这些特点使得现有算法不能较好地适应该场景。针对设备铭牌这一对象,文献[15]在CTPN的基础上提出一种鲁棒文本行检测(robust text line detection, RTLD)模型,完成了对设备铭牌图像的文本检测,但无法检测间隔很小的文本行。

综上,目前缺少效果较好的设备铭牌文本信息

收稿日期:2022-04-13;修回日期:2022-06-16

基金项目:国家自然科学基金资助项目(92066106)

识别算法。因此文中针对电力设备铭牌的上述特点,提出一种电力设备铭牌信息自动识别算法,能够较好地适应电力设备铭牌这一特殊场景,实现较快、较准确的电力设备铭牌信息自动提取。

## 1 基于深度学习的电力设备铭牌文本识别模型

文中提出的电力设备铭牌文本识别模型包括铭牌检测、文本检测、文本识别三部分。首先对图像中的铭牌区域进行定位,然后将截取的铭牌图像作为文本检测的输入,框选出铭牌中的文字部分,最后对文本候选框中的内容进行文本识别,输出最终的铭牌文本信息,整个流程效果如图1所示。

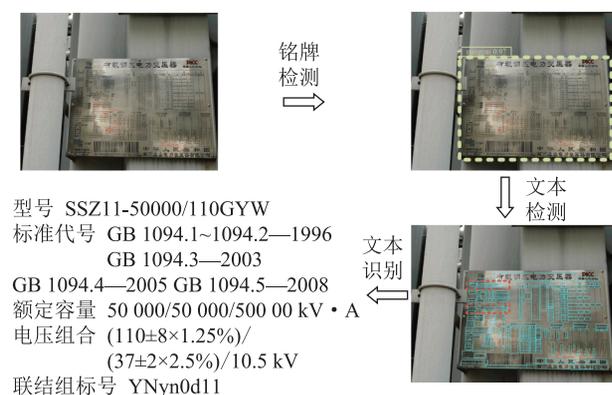


图1 算法模型的整体流程与效果

Fig.1 Process and effect of the algorithm model

### 1.1 基于YOLO的铭牌检测模型

YOLO模型是一种基于深度卷积神经网络的目标检测模型,在一个步骤中同时完成检测物体的定位与分类,具有效率高的优点<sup>[16]</sup>。文中基于YOLOv3模型设计了电力设备铭牌检测模型。YOLOv3不仅延续了YOLO高效的优点,同时提高了检测准确率,并可对不同尺度目标进行识别<sup>[17]</sup>。基于YOLO的铭牌检测模型框架如图2所示。

首先,利用Darknet-53作为主干网络提取输入图像不同尺度的特征<sup>[18]</sup>。Darknet-53中包含一系列卷积核为 $3 \times 3$ 和 $1 \times 1$ 的卷积层,并采用残差网络(residual network, ResNet)结构缓解网络深度带来的梯度消失问题,在性能和效率上均优于Darknet-19、ResNet101、ResNet-152等网络。然后,模型对提取到的不同尺度的特征进行上采样和张量连接以实现特征交互,最终输出3种不同尺度的特征向量,用于预测不同尺度目标的类别和位置信息。

#### 1.1.1 模型损失函数

YOLO原始模型的损失函数由三部分组成,即坐标损失、目标损失和分类损失。

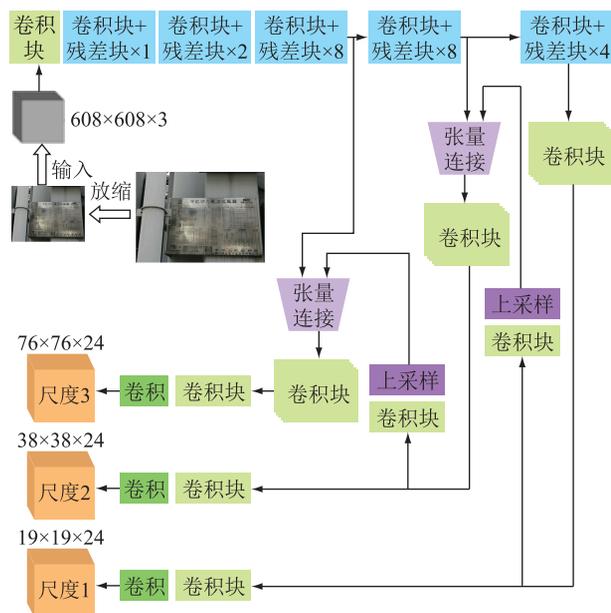


图2 基于YOLO的铭牌检测模型框架

Fig.2 Framework of the YOLO-based nameplate detection model

$$L = \lambda_{\text{coord}} L_{\text{coord}} + (\lambda_{\text{obj}} L_{\text{obj}} + \lambda_{\text{noobj}} L_{\text{noobj}}) + \lambda_{\text{cls}} L_{\text{cls}} \quad (1)$$

式中: $L_{\text{coord}}$ 、 $L_{\text{cls}}$ 分别为坐标损失和分类损失; $L_{\text{obj}}$ 、 $L_{\text{noobj}}$ 分别为候选框中存在目标的目标损失与不存在目标的目标损失; $\lambda_{\text{coord}}$ 、 $\lambda_{\text{obj}}$ 、 $\lambda_{\text{noobj}}$ 、 $\lambda_{\text{cls}}$ 为各损失相应权重。

在YOLO原始模型中,坐标损失采用均方误差(mean squared error, MSE)损失,目标损失采用二元交叉熵(binary cross entropy error, BCE)损失,文中针对2种损失分别进行了改进。

原始模型的坐标损失通过MSE衡量,在评估模型性能时将预测框与真实框的交并比(intersection over union, IoU)作为性能指标。但这在某些情况下并不合理,因为MSE低不一定代表IoU大,这意味着即使模型的损失函数收敛到很小也可能得不到性能很好的结果。以图3为例,3种情况下的损失都相同,但第一种情况的IoU显然高于另外2种,且主观观察结果也是第一个更好。

针对这个问题,文中将通用交并比(generalized intersection over union, GIoU)<sup>[19]</sup>损失应用到YOLO中代替原来的坐标损失。假设预测框和真实框分别为 $A$ 和 $B$ , $C$ 为 $A$ 和 $B$ 最小凸集(包围 $A$ 、 $B$ 的最小矩形框),则GIoU损失的计算公式为:

$$L_{\text{GIoU}} = L_{\text{IoU}} - \frac{|C \setminus (A \cup B)|}{|C|} \quad (2)$$

式中: $A \cup B$ 为矩形框 $A$ 、 $B$ 的并集; $C \setminus (A \cup B)$ 为矩形框 $C$ 与 $A$ 、 $B$ 不重叠的部分; $|C|$ 为矩形框 $C$ 的面

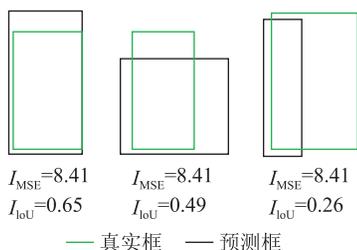


图3 不同情况下预测框与真实框的IoU

Fig.3 IoU of bounding box and ground truth in different cases

积; $L_{IoU}$ 为矩形框A、B的交并比,其计算见式(3),其中 $A \cap B$ 为矩形框A、B的交集。

$$L_{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

由式(2)可知, $L_{GIoU}$ 的取值范围为 $(-1, 1]$ ,当A、B框不重叠时:

$$L_{GIoU} = -1 + \frac{|A \cup B|}{|C|} \quad (4)$$

显然, $|A \cup B|$ 为定值,若要最大化 $L_{GIoU}$ 则要最小化 $|C|$ ,即令预测框与真实框不断靠近,即令交并比更大,从而统一训练时的损失函数与评估时的性能指标,使得训练效果更好的模型在评估时也能获得更好的结果。

YOLO模型通过生成大量的候选框进行目标检测,但大部分候选框中都是简单易分的背景样本,使得训练中负样本占主导,不利于模型学习目标特征。为解决该问题,在YOLO模型中引入焦点损失(Focal Loss)函数<sup>[20]</sup>作为目标损失,降低了易识别的简单样本在训练中所占的损失比例,加大了难识别的困难样本的损失比例,平衡了正负样本。

YOLO原始模型中的目标损失为BCE损失函数,计算如下:

$$L = -y \ln y^* - (1 - y) \ln(1 - y^*) = \begin{cases} -\ln y^* & y = 1 \\ -\ln(1 - y^*) & y = 0 \end{cases} \quad (5)$$

式中: $y^*$ 为模型输出; $y$ 为样本标签值, $y=1$ 为正样本, $y=0$ 为负样本。

可见,对正样本而言,输出概率越大,BCE损失越小;对负样本而言,输出概率越小,BCE损失越小。此时,训练过程受大量易识别的简单样本主导而导致迭代过程缓慢,可能无法得到满意的优化结果。Focal Loss函数在BCE的基础上作如下改进:

$$L_{FL} = \begin{cases} -(1 - y^*)^\gamma \ln y^* & y = 1 \\ -(y^*)^\gamma \ln(1 - y^*) & y = 0 \end{cases} \quad (6)$$

式中: $\gamma$ 为正值权重因子,是一个可调整的超参数。引入正值权重因子 $\gamma$ 后,易分类样本的损失降低,

模型更加关注难以分类、易错的困难样本。

此外,为解决样本正负比例不平衡问题,文中在Focal Loss函数中进一步增加了平衡因子 $\alpha$ :

$$L_{FL} = \begin{cases} -\alpha(1 - y^*)^\gamma \ln y^* & y = 1 \\ -(1 - \alpha)(y^*)^\gamma \ln(1 - y^*) & y = 0 \end{cases} \quad (7)$$

### 1.1.2 锚参数聚类

为适应电力设备铭牌的大小特征,文中对数据集中所有目标框进行K-means聚类以获取合适的锚参数。传统K-means算法基于欧氏距离分类,会导致较大的框相比于较小的框产生更大的分类误差,而文中的目标是使得到的锚框与目标框有更大的IoU,这一目标与锚框的大小无关<sup>[21]</sup>。因此,文中使用基于IoU的距离计算公式进行聚类:

$$D(B_{box}, B_{center}) = 1 - L_{IoU}(B_{box}, B_{center}) \quad (8)$$

式中: $D$ 为计算得到的距离; $B_{box}$ 为数据集中的样本框; $B_{center}$ 为聚类中心的样本框; $L_{IoU}(B_{box}, B_{center})$ 为二者的交并比。

如图4所示,设置9个聚类中心,不同颜色的框线代表不同大小的锚框,最终得到铭牌数据集锚参数为:(301, 282)、(165, 136)、(30, 27)、(55, 45)、(601, 402)、(463, 274)、(267, 177)、(84, 72)、(124, 91)。

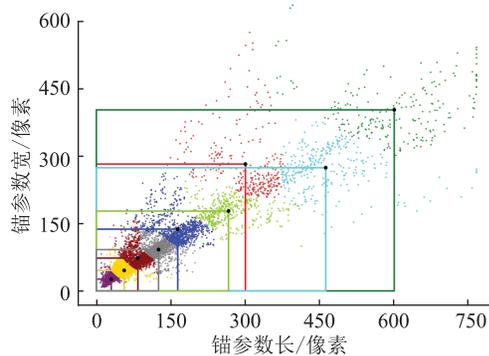


图4 基于K-means聚类的锚参数

Fig.4 Anchor parameters based on K-means clustering

### 1.2 基于EAST的文本检测模型

EAST是一种基于深度学习的端到端的文本检测方法,省去中间阶段,直接对文本行进行预测,大大简化了文本检测过程。文中基于EAST模型设计电力铭牌文本检测模型,其模型框架如图5所示。

该模型首先利用一个通用的主干网络提取图像特征,文中选择常用的ResNet50提取不同尺度的特征图<sup>[22]</sup>,分别为输入图像的1/32,1/16,1/8,1/4,用于预测不同大小的文本行。然后,对上层特征进行上采样并经过卷积层,逐渐与下层特征融合。最后,输出层预测文本置信度以及文本位置信息,并

进行非极大抑制 (non-maximum suppression, NMS) 操作, 删去冗余的文本候选框。

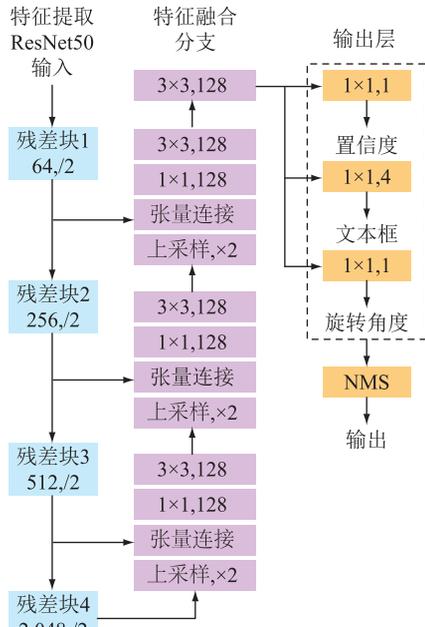


图5 基于 EAST 的文本检测模型框架  
Fig.5 Framework of the EAST-based text detection model

EAST 模型的损失函数由两部分组成, 即文本得分损失和几何形状损失:

$$L = L_s + \lambda_g L_g \quad (9)$$

式中:  $L_s$ ,  $L_g$  分别为文本得分损失和几何形状损失;  $\lambda_g$  为几何形状损失的权重, 文中设置其值为 1。

几何形状损失包括轴平行矩形框损失  $L_{AABB}$  和倾斜角度损失  $L_\theta$ , 设置倾斜角度损失权重  $\lambda_\theta$  为 10。

$$L_g = L_{AABB} + \lambda_\theta L_\theta \quad (10)$$

$$L_{AABB} = -\ln L_{IoU}(R, R^*) = -\ln \frac{|R \cap R^*|}{|R \cup R^*|} \quad (11)$$

$$L_\theta = 1 - \cos(\theta - \theta^*) \quad (12)$$

式中:  $R$ ,  $\theta$  分别为预测的轴平行矩形框区域及其倾斜角度;  $R^*$ ,  $\theta^*$  为相对应的真实标签。

在原始的 EAST 模型中, 为简化训练过程, 文本得分损失采用一种类平衡交叉熵损失<sup>[23]</sup>。文中为解决铭牌图像中前景与背景比例不平衡的问题, 采用 Dice 损失函数<sup>[24]</sup> 衡量文本得分损失。

$$L_s = 1 - \frac{2|Y \cap Y^*|}{|Y| + |Y^*|} \quad (13)$$

式中:  $Y$  为文本得分的预测矩阵;  $Y^*$  为文本得分对应的标签矩阵。  $|Y \cap Y^*|$  为  $Y$  和  $Y^*$  之间的重叠部分, 其计算为两矩阵相乘得到的新矩阵中元素之和;  $|Y|$ ,  $|Y^*|$  分别为  $Y$  和  $Y^*$  中元素的个数。

此外, EAST 模型在较长文本上会出现检测不完整的问题, 文中针对电力设备铭牌多为水平长矩形文本的特点, 设计一种非极大抑制算法<sup>[25]</sup>, 以适应电力设备铭牌中的长文本检测。

### 1.3 基于卷积循环神经网络的文本识别模型

卷积循环神经网络 (convolutional recurrent neural network, CRNN) 是目前主流的文本识别方法之一<sup>[26]</sup>。该算法借鉴语音识别的思路, 利用文本的序列特征, 将 RNN 与 CTC 引入到图像信息识别中, 能够有效识别任意长度文本。文中基于 CRNN 设计了电力铭牌文本识别模型, 模型框架如图 6 所示。

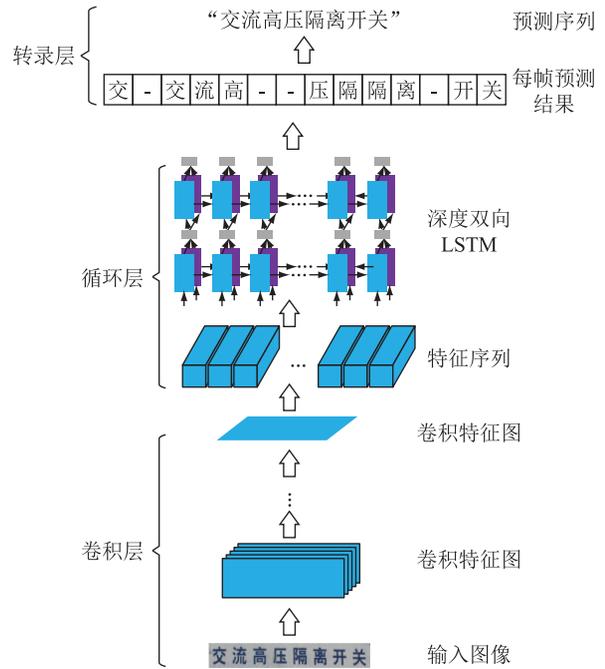


图6 基于 CRNN 的文本识别模型框架  
Fig.6 Framework of the CRNN-based text recognition model

首先, 将输入的铭牌文本图像转换成灰度图并缩放到固定高度, 文中设置该高度为 32 像素, 即输入尺寸为  $32 \times W \times 1$ , 其中  $W$  为缩放后输入图像的宽度。接着, 利用卷积层提取铭牌文本图像特征, 得到尺寸为  $32 \times (W/4) \times 512$  的特征序列。然后, 将特征序列作为循环层的输入, 利用长短期记忆网络 (long short-term memory network, LSTM)<sup>[27]</sup> 学习并预测文本序列的标签分布。最后, 在转录层中使用 CTC 将标签分布转录为文本识别结果。

实验中发现, 训练后的模型对字符的识别准确率较高, 而对文本行的识别准确率相对较低, 这说明文本行中的识别错误大多发生在少量字符上, 因此文中设计一种基于字典的文本识别结果纠正方法。针对电力设备铭牌中存在的高频词汇, 如“额定容量”、“标准代号”等建立字典库, 统计每个词汇

在铭牌中出现的频率并进行排名,频率最高的排名记为1。文本识别结果的纠正流程如图7所示。

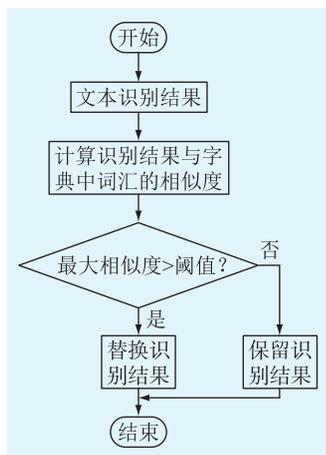


图7 文本识别结果纠正流程

Fig.7 Flow chart of the correction of text recognition results

每个文本识别结果与字典中词汇的相似度的计算如下:

$$D(w_1, w_2) = \frac{D_{\text{Edit}}(w_1, w_2)}{N_{w_1} + N_{w_2}} + \lambda \frac{R_{w_2}}{N_{\text{dict}}} \quad (14)$$

$$S(w_1, w_2) = \frac{1}{D(w_1, w_2)} \quad (15)$$

式中: $w_1$ 为文本识别结果字符串,长度为 $N_{w_1}$ ;  $w_2$ 为字典中待匹配的字符串,长度为 $N_{w_2}$ ;  $R_{w_2}$ 为其在字典中的频率排名;  $N_{\text{dict}}$ 为字典中的词汇总数;  $\lambda$ 为字典排名权重,文中设置为0.1;  $D_{\text{Edit}}(w_1, w_2)$ 为编辑距离;  $D(w_1, w_2)$ 为差异度;  $S(w_1, w_2)$ 为相似度。

得到相似度后,选择最大相似度及其对应的字典中词汇,若相似度大于设定的阈值,则把对应的字典中词汇作为纠正后的识别结果,否则保留原结果,文中设定阈值为5。

## 2 铭牌文本识别结果与分析

### 2.1 铭牌检测模型

#### 2.1.1 数据预处理

由于电力设备铭牌数据相对较少,数量级在几百到一千,而包含电力设备种类较多,涵盖了变压器、开关、断路器等各种电力设备,因此已有的数据量级不足以支撑深度学习算法的训练。文中采用随机旋转、放缩、剪裁、色彩增强等方式对已有的图像样本进行数据增广。

#### 2.1.2 实验结果

文中采用随机参数搜索的方式选择模型的超参数,首先根据经验设置超参数的初始值,对模型进行预训练,训练少量迭代次数,采用平均均值精

度(mean average precision, mAP)和 $F_1$ 的加权作为性能指标评价训练结果,根据训练结果选取性能最好的一组超参数,进行随机改动,更新超参数。其中, $F_1$ 平衡了召回率和准确率,计算方法见式(16)。重复上述步骤直至达到最大迭代次数。

$$F_1 = 2 \frac{PR}{P + R} \quad (16)$$

式中: $P$ 为准确率; $R$ 为召回率。

模型训练过程如图8所示。

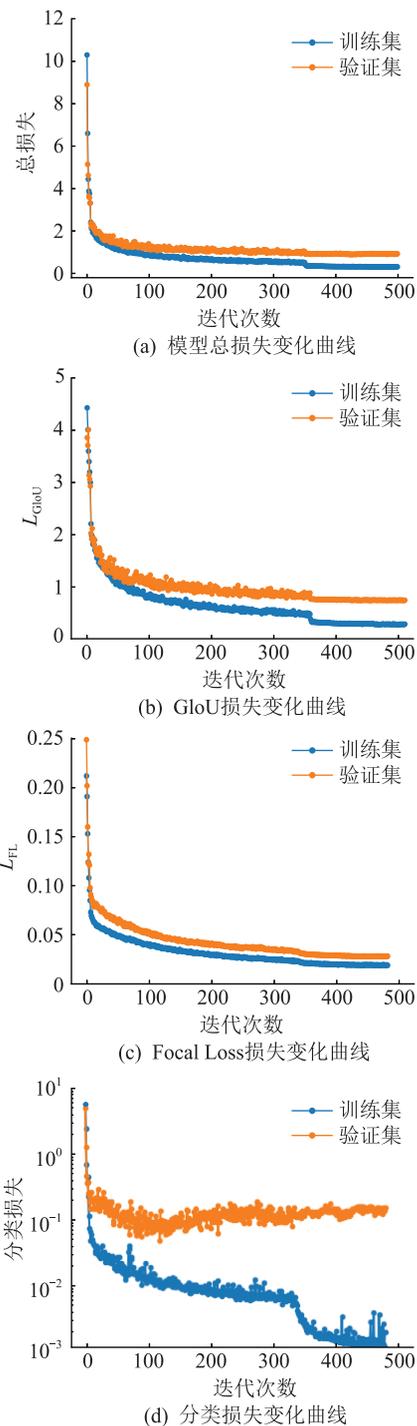


图8 铭牌检测模型的训练损失曲线

Fig.8 Curves of training loss for nameplate detection model

由图 8 可见,模型随着迭代次数上升最终趋于收敛,且没有出现过拟合的现象。模型在测试集上的 mAP 值、准确率和召回率分别为 92.2%,84.5%,88.3%。将损失函数改进后的模型与原始模型在同一测试集上进行对比,结果如表 1 所示。

表 1 铭牌检测模型改进前后对比

Table 1 Comparison between the original and improved nameplate detection models %

模型	mAP 值	准确率	召回率
原始模型	89.9	82.9	86.6
仅使用 GIoU	92.1	84.2	88.6
仅使用 Focal Loss	90.0	82.1	87.2
使用 GIoU 与 Focal Loss	92.2	84.5	88.3

由表 1 可见,使用 GIoU 损失后,模型的各项指标均有所提高;使用 Focal Loss 损失后,模型拥有更高的召回率,但准确率有所下降;同时使用 2 种改进损失函数后,模型的准确率和召回率均较高,mAP 值和准确率均为最优。

## 2.2 文本检测模型

### 2.2.1 数据预处理

由于原始铭牌图像数据中的文本大多为不规则四边形标注,为了满足 EAST 模型的文本标注要求,文中对原始数据进行了预处理。首先将原始标注的四边形边长收缩 0.3 倍,以使标注更加精确,然后生成收缩后标注框的轴平行矩形框标签以及倾斜角标签。此外,文中还对图像数据进行了数据清洗和图像增广操作,删去不规范的标注框,扩大训练集规模。

### 2.2.2 实验结果

训练过程中使用 Adam 优化器优化模型参数,迭代前期设置学习率为 0.000 1,迭代后期设置学习率为指数衰减。模型训练过程如图 9 所示,各项损失在训练中都逐渐减小并收敛至较小的值。

原始 EAST 模型在测试集上的  $F_1$  值为 88.4%,而改进后模型最终在测试集上的  $F_1$  值达到 91.2%,可以较为精确地检测出电力设备铭牌图像中不同类型、不同尺度的文本信息。

## 2.3 文本识别模型

### 2.3.1 铭牌文本生成

由于已有的铭牌图片数据中文本数量过少,文中根据电力设备铭牌文本多样、背景多样、排版多样、形变多样、图片质量多样的特点,设计铭牌文本生成器,人工合成大量训练样本。

首先,收集与电力设备相关的文本信息,包括国家标准、相关设备教材定义等,建立电力设备铭

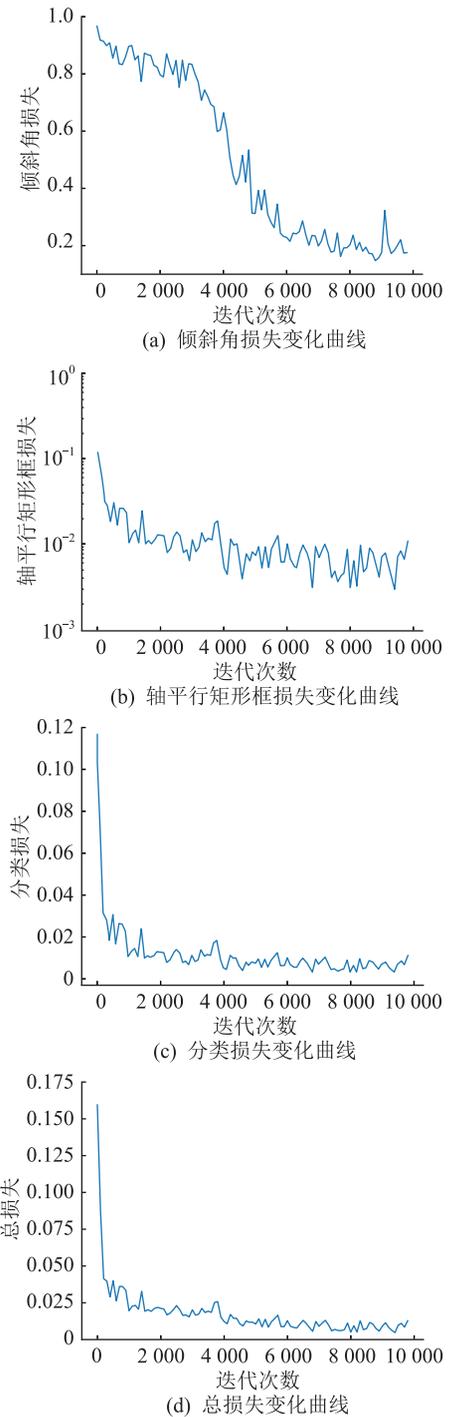


图 9 文本检测模型训练损失曲线

Fig.9 Curves of training loss for text detection model

牌文本语料库,利用库中的文本生成文本图片;其次,根据真实铭牌图像中截取的背景,建立铭牌文本背景库;最后,从语料库中随机提取文本,并进行字体、形变的随机变换,以模拟实际样本中的文本情况,加入背景库中随机提取的背景,完成样本的合成。具体文本图像合成流程如图 10 所示。

最终,根据实际铭牌文本图像特征,人工合成 100 万张训练样本,包含清晰、正常文本,有遮挡的文本,模糊的文本等各种情况下的文本图像。部分

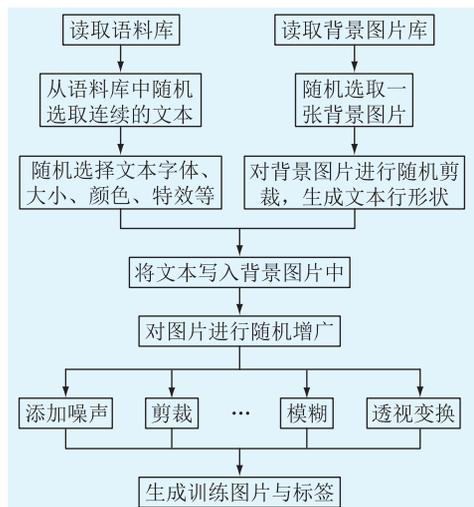


图 10 文本图像合成流程

Fig.10 Process of text image synthesis

合成文本图像如图 11 所示。合成的文本图像采用和 CRNN 开源数据集中一致的尺寸质量,统一裁剪为 280×32 的尺寸。



图 11 部分合成文本图像

Fig.11 Partial results of the synthetic text images

2.3.2 实验结果

训练过程中使用 RMSProp 优化器优化模型参数,整个过程中学习率保持定值 0.000 1。

训练过程分为在 CRNN 公开数据集上的预训练以及在人工合成样本上的微调,训练损失如图 12 所示。模型在预训练中的损失已经下降到很小的值后,在微调训练中损失进一步下降。

在全部由真实电力设备铭牌组成的数据集上,对模型进行测试,部分文本识别结果如表 2 所示。

实验发现,在 CRNN 开源数据集上预训练的原始模型的字符识别准确率为 87.1%,而文本行识别

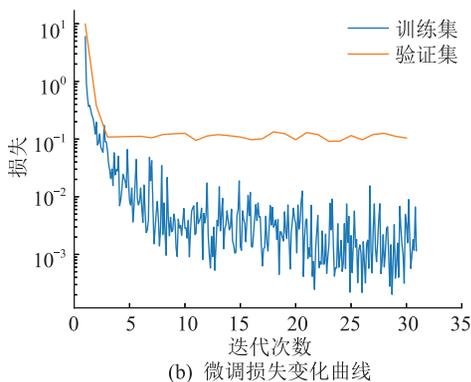


图 12 文本识别损失曲线

Fig.12 Curves of text recognition loss

表 2 文本识别部分结果

Table 2 Partial results of the text recognition

铭牌文本图片	识别结果
GB1094.3-2003	GB 1094.3—2003
变压器用波纹储油柜	变压器用波纹储油柜
A.C相充气运输重:	4.C 相充气运输重:
高压绕组连接示意图	高压绕组连接示意图
242000	240200
0.5 Fs ≤ 5	0.5 Fs ≤ 5
ONAN/ONAF 67%/100%	ONAN/ONAF67%100%
载调压电力变压器	载调压电力变压器
C2007122	C2007122

准确率仅有 51.7%;微调训练完成的模型对字符的识别率高达 93.2%,文本行识别准确率提升至 82.3%,为进一步提升模型效果,文中使用 1.3 节中所述纠正方法对结果进行后处理,结果如表 3 所示,字符、文本行识别准确率均有所提升。

表 3 纠正前后的实验结果

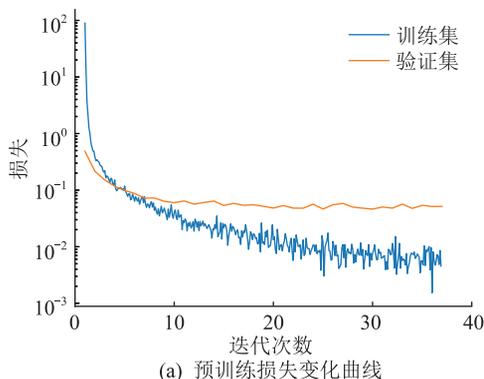
Table 3 Experimental results before and after correction

实验	字符识别准确率	文本行识别准确率	%
纠正前	93.2	82.3	
纠正后	94.0	83.3	

3 结语

文中设计了一种基于深度学习的电力设备铭牌文本自动识别算法,可有效实现对电力设备铭牌图像中文本信息的自动识别。

该算法分为铭牌检测、文本检测、文本识别三部分。首先通过铭牌检测模型,检测并框选出图像中的铭牌部分;然后将铭牌图像作为输入,进行文本检测,得到文本行的位置、角度等信息;最后截取铭牌中的文本行,进行文本识别,得到最终的识别结果。在现有模型的基础上改进损失函数的设



(a) 预训练损失变化曲线

计、增加文本识别结果纠正环节、人工合成文本图像以丰富电力设备铭牌文本库等,得到文中铭牌检测模型在测试集上的 mAP 值、准确率和召回率分别可达 92.2%,84.5%和 88.3%;文本检测模型在测试集上的  $F_1$  可达到 91.2%;文本识别模型在测试集上拥有 94.0%的字符识别准确率和 82.3%的文本行识别准确率。

#### 参考文献:

- [1] 李思妍,台升,张宇航,等. 基于轻量化 YOLOv3 和 Tesseract OCR 的电力设备标志牌识别技术[J]. 智慧电力,2021,49(7):79-85.  
LI Siyan,TAI Sheng,ZHANG Yuhang, et al. Electrical sign recognition technology based on simplified YOLOv3 and Tesseract OCR[J]. Smart Power,2021,49(7):79-85.
- [2] 方静,彭小圣,刘泰蔚,等. 电力设备状态监测大数据发展综述[J]. 电力系统保护与控制,2020,48(23):176-185.  
FANG Jing,PENG Xiaosheng,LIU Taiwei, et al. Development trend and application prospects of big data-based condition monitoring of power apparatus[J]. Power System Protection and Control,2020,48(23):176-185.
- [3] 李翌昕,马尽文. 文本检测算法的发展与挑战[J]. 信号处理,2017,33(4):558-571.  
LI Yixin,MA Jinwen. The developments and challenges of text detection algorithms[J]. Journal of Signal Processing,2017,33(4):558-571.
- [4] MEMON J,SAMI M,KHAN R A, et al. Handwritten optical character recognition (OCR): a comprehensive systematic literature review (SLR)[J]. IEEE Access,8:142642-142668.
- [5] MITTAL R,GARG A. Text extraction using OCR: a systematic review[C]//2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). Coimbatore, India. IEEE:357-362.
- [6] TIAN S X,PAN Y F,HUANG C, et al. Text flow: a unified text detection system in natural scene images[C]//2015 IEEE International Conference on Computer Vision. Santiago, Chile. IEEE:4651-4659.
- [7] LIU X Y,MENG G F,PAN C H. Scene text detection and recognition with advances in deep learning: a survey[J]. International Journal on Document Analysis and Recognition (IJ DAR), 2019,22(2):143-162.
- [8] 杨传凯,孔志战,谢倩楠,等. 基于 DeepLab v3+深度卷积网络的输电导线图像识别方法[J]. 电力工程技术,2021,40(4):189-194.  
YANG Chuankai,KONG Zhizhan,XIE Qiannan, et al. Image recognition method for transmission line based on the DeepLab v3+ deep convolutional network[J]. Electric Power Engineering Technology,2021,40(4):189-194.
- [9] TIAN Z,HUANG W L,HE T, et al. Detecting text in natural image with connectionist text proposal network [M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016:56-72.
- [10] SHI B G,BAI X,BELONGIE S. Detecting oriented text in natural images by linking segments[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. IEEE:3482-3490.
- [11] LIU W,ANGUELOV D,ERHAN D, et al. SSD: single shot MultiBox detector[M]//Computer Vision-ECCV 2016. Cham: Springer International Publishing,2016:21-37.
- [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. Honolulu, HI, USA. IEEE:2642-2651.
- [13] 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.
- [14] LEE C Y,OSINDERO S. Recursive recurrent nets with attention modeling for OCR in the wild [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. IEEE:2231-2239.
- [15] LAI J Y,GUO L Q,QIAO Y, et al. Robust text line detection in equipment nameplate images [C]//2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). New York: ACM,2019:889-894.
- [16] REDMON J,DIVVALA S,GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. IEEE:779-788.
- [17] REDMON J,FARHADI A. YOLOv3: an incremental improvement[EB/OL]. 2018. <https://arxiv.org/abs/1804.02767>.
- [18] 王立辉,秦成帅,杨贤彪,等. 基于深度学习的发电站制冷水管焊缝图像检测[J]. 电力工程技术,2020,39(5):191-196.  
WANG Lihui,QIN Chengshuai,YANG Xianbiao, et al. Image detection on welding area of cooling water pipe in power station based on deep learning[J]. Electric Power Engineering Technology,2020,39(5):191-196.
- [19] REZATOFIGHI H,TSOI N,GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE:658-666.
- [20] LIN T Y,GOYAL P,GIRSHICK R, et al. Focal loss for dense object detection [C]//2017 IEEE International Conference on Computer Vision. Venice, Italy. IEEE:2999-3007.
- [21] 郭敬东,陈彬,王仁书,等. 基于 YOLO 的无人机电力线路杆塔巡检图像实时检测[J]. 中国电力,2019,52(7):17-23.  
GUO Jingdong,CHEN Bin,WANG Renshu, et al. YOLO-based real-time detection of power line poles from unmanned aerial vehicle inspection vision [J]. Electric Power,2019,52(7):17-23.
- [22] SHARMA N,JAIN V,MISHRA A. An analysis of convolutional neural networks for image classification [J]. Procedia Com-

- puter Science, 2018, 132:377-384.
- [23] XIE S N, TU Z W. Holistically-nested edge detection [ C ] // 2015 IEEE International Conference on Computer Vision. Santiago, Chile. IEEE: 1395-1403.
- [24] LI X Y, SUN X F, MENG Y X, et al. Dice loss for data-imbalanced NLP tasks [ EB/OL ]. 2019. <https://arxiv.org/abs/1911.02855>.
- [25] HOSANG J, BENENSON R, SCHIELE B. Learning non-maximum suppression [ C ] // 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. IEEE: 6469-6477.
- [26] TONG G F, LI Y, GAO H S, et al. MA-CRNN: a multi-scale attention CRNN for Chinese text line recognition in natural scenes [ J ]. International Journal on Document Analysis and Recognition ( IJDAR ), 2020, 23(2) : 103-114.
- [27] XIE Z C, SUN Z H, JIN L W, et al. Fully convolutional recurrent network for handwritten Chinese text recognition [ C ] // 2016 23rd International Conference on Pattern Recognition (ICPR). Cancun. IEEE: 4011-4016.

作者简介:



王逸凡

王逸凡(1999),男,硕士在读,研究方向为人工智能技术在高电压与放电等离子体中的应用(E-mail: ee\_wyf@seu.edu.cn);

王佳宇(1995),男,硕士,研究方向为人工智能技术在电力系统中的应用;

仲林林(1990),男,博士,副研究员,博士生导师,研究方向为高电压与放电等离子体数值计算以及人工智能交叉技术。

## Text recognition of power equipment nameplates based on deep learning

WANG Yifan, WANG Jiayu, ZHONG Linlin, GAO Bingtuan

(School of Electrical Engineering, Southeast University, Nanjing 210096, China)

**Abstract:** There is abundant information in power equipment nameplates. Extracting information from nameplates through image and text recognition technology enables the high effective and quick works on statics and account check of power equipment, which is also beneficial to improve the equipment management level of power system. Considering the great difference of text recognition between power equipment nameplates and ordinary images, an algorithm of automatic recognition of power equipment nameplates based on deep learning is proposed in this paper. This algorithm consists of three parts, namely nameplate detection, text detection and text recognition. By improving the design of loss functions, adding the correction of text recognition, and synthesizing text images, the mean average precision of the nameplate detection model on the test set reaches 92.2%, the  $F_1$  of the text detection model on the test set reaches 91.2%, and the character recognition accuracy rate of the text recognition model reaches 94.0%, the text line recognition accuracy rate of the text recognition model reaches 82.3%.

**Keywords:** power equipment nameplate; computer vision; deep learning; object detection; text detection; text recognition

(编辑 方晶)