

doi: 10.11835/j.issn.1000-582X.2026.06.006

引用格式:何涛,冀毅.基于深度学习的铁路计算机联锁界面文本定位与识别方法研究[J].重庆大学学报, 2026,49(6): 59-70.



# 基于深度学习的铁路计算机联锁界面文本定位与识别方法研究

何涛<sup>a,b</sup>,冀毅<sup>c</sup>

(兰州交通大学 a. 自动控制研究所; b. 甘肃省轨道交通信号与控制评测行业技术中心;

c. 自动化与电气工程学院,兰州 730070)

**摘要:**为了有效解决铁路计算机联锁系统人工测试效率低下、准确性低等问题,提出了一种面向铁路联锁上位机界面文本定位与识别的深度学习模型。首先,通过对比基于CTPN(connectionist text proposal network)神经网络的文本定位模型与 ResNet50、AlexNet、ZF 以及 VGG16 四种卷积神经网络分别结合的策略,选择了最优的特征提取网络 VGG16,增强了卷积特征图表示高层语义信息的细节特征,以利于定位小文本区域。其次,对比常见的目标检测模型在文本定位效果上的表现以及采用 dropout 方法,提升了文本定位模型 CTPN 网络的泛化能力和鲁棒性。然后,采用水平投影和垂直投影相结合的分割方法,可有效避免联锁上位机界面文本粘连等问题。最后,采用改进的 AlexNet 网络,输出文本识别结果。通过在 TensorFlow 环境验证联锁上位机界面文本定位和识别数据集,结果显示,CTPN 网络在铁路联锁上位机界面文本定位精确率上达到了 87.98%,召回率 73.33%,调和平均数指标 80.39%;改进的 AlexNet 网络文本识别准确率达到了 89%。说明本文方法能够对铁路计算机联锁上位机界面文本实现准确定位和识别,并为联锁自动测试过程中自动办理进路及测试结果自动分析提供可靠的数据支持。

**关键词:**文本定位;文本识别;CTPN;计算机联锁上位机界面;AlexNet

中图分类号:TP391.1

文献标志码:A

文章编号:1000-582X(2026)06-059-12

## Deep learning-based text location and recognition for railway computer interlocking interfaces

HE Tao<sup>a,b</sup>, JI Yi<sup>c</sup>

(a. AutoMatic Control Institute; b. GanSu Industry Technology Center of Evaluation and Testing of Rail Transportation Signal and Control; c. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, P. R. China)

**Abstract:** To address the low efficiency and accuracy of manual testing in railway computer interlocking systems, this study proposes a deep learning-based method for text localization and recognition in interlocking interface images. First, a text localization model based on the connectionist text proposal network (CTPN) is developed. By

收稿日期:2021-04-11

基金项目:甘肃省科技计划项目(20CX9JA125, 20JR5RA407)。

Supported by Gansu Province Science and Technology Planning Project (20CX9JA125, 20JR5RA407).

通信作者:何涛(1977—),男,教授,主要从事轨道交通测试方向研究,(E-mail)1162483144@qq.com。

comparing multiple backbone networks (ResNet50, AlexNet, ZF and VGG16), VGG16 is selected as the feature extractor to enhance high-level semantic representation and improve the detection of small text regions. Second, the generalization ability and robustness of the CTPN model are improved through performance comparison with common object detection models and the incorporation of dropout. A projection-based segmentation method, combining horizontal and vertical projections, is further employed to address text adhesion issues in the interface. Finally, an improved AlexNet model is used for text recognition. Experimental results on a railway interlocking interface dataset in the TensorFlow environment show that the proposed method achieves a localization accuracy of 87.98%, a recall of 73.33%, and an F-score of 80.39%, while the recognition accuracy reaches 89%. These results demonstrate that the proposed approach can effectively locate and recognize interface text, providing reliable data support for automated routing and test result analysis in interlocking system testing.

**Keywords:** text location; text recognition; CTPN; railway computer interlocking interface; AlexNet

计算机联锁系统(computer-based interlocking system, CBI)是保证车站内行车安全的关键设备<sup>[1-2]</sup>,具有极其严格的安全性、可靠性、稳定性要求,在投入使用前需要进行全面、严格的仿真测试<sup>[3]</sup>。随着铁路事业的不断进步和快速发展,铁路网变得越来越密集,确保铁路安全的任务也必将越来越重。现阶段计算机联锁系统仿真测试主要依靠人工测试方法,然而人工测试效率低下,准确性依赖于测试人员的经验和知识储备,不易发现系统存在的潜在错误<sup>[4]</sup>。因此,为满足铁路行车安全要求,科研工作者展开了深入的研究,利用各种新技术来解决人工测试不足的问题。在此背景下,随着图像处理和模式识别等技术的发展,联锁自动测试应运而生。基于深度学习的联锁自动测试关键在于,如何利用联锁上位机界面实现自动办理进路和测试结果的自动分析。计算机联锁上位机显示界面包括站场图形显示、信号显示、操作按钮显示、设备状态信息显示、报警信息显示等几方面,所有显示内容、显示效果将直接影响车务人员识别和确认信息,从而影响列车安全高效的运行<sup>[5]</sup>。因此,如何定位联锁上位机界面中所需要的文本信息,快速、准确地识别显得尤为重要。

联锁上位机界面文本识别过程一般分为文本区域定位和文本识别两步。目前,文本区域定位与识别相关研究有很多,尤其是基于深度学习的方法,取得了良好的效果。Tian等<sup>[6]</sup>应用CTPN(connectionist text proposal network)算法,将检测的小块文本连接形成文本行来检测文本;Ma等<sup>[7]</sup>开发了一种应用通用目标检测框架Faster R-CNN(faster region-based convolutional neural network)的基于候选区域的方法来检测文本;Liao等<sup>[8]</sup>为了解决各种长度文本行检测困难等问题,提出了一种采用长条形卷积核提取特征的文本检测模型TextBoxes;Deng等<sup>[9]</sup>通过采用语义分割和连通域相结合的思想,提出了PixelLink文本检测模型。Zhang等<sup>[10]</sup>引入了一种FCN(fully convolutional networks)算法的基于分割思想的方法来检测文本。刘澍等<sup>[11]</sup>提出了一种采用标准化差值平方和的方式判断图像相似程度的模板匹配法来识别文本;Chang等<sup>[12]</sup>提出了一种直接使用像素值排列作为特征向量的支持向量机(support vector machine, SVM)分类法来识别文本。Le等<sup>[13]</sup>提出了一种仅有两层卷积的LeNet-5网络来识别文本;Shi等<sup>[14]</sup>提出了一种基于序列的文本识别模型CRNN(convolutional recurrent neural network),网络直接提取文本行空间特征和序列特征来识别文本。

基于连接预选框网络的文本检测CTPN是目前流传最广、影响最大的开源文本检测模型,可以应用于检测水平或微斜的文本行<sup>[15]</sup>。传统的字符分割和改进的AlexNet神经网络相结合的方式分割及识别速度和精度较高,满足实时分割和识别的要求。

基于以上研究和分析,笔者研究将CTPN应用于联锁自动测试,检测联锁上位机界面中的文本信息,采用传统的字符分割和改进的AlexNet神经网络相结合的方式实现对检测文本的识别。经过算法改进和融合运用,文本定位精度和识别准确率均达到了良好的效果,为解决人工测试效率低下、准确性低等问题提出了一种新思路,为计算机联锁自动测试的实现奠定了基础。

# 1 基于CTPN的文本定位

## 1.1 CTPN简介

CTPN是在Faster R-CNN<sup>[16]</sup>的基础上加入了双向LSTM(long short-term memory)改进而来,ECCV 2016提出一种基于卷积神经网络和循环神经网络的文本检测模型。CTPN网络结构如图1所示,网络主要包括3个部分:卷积层、双向LSTM和全连接层。

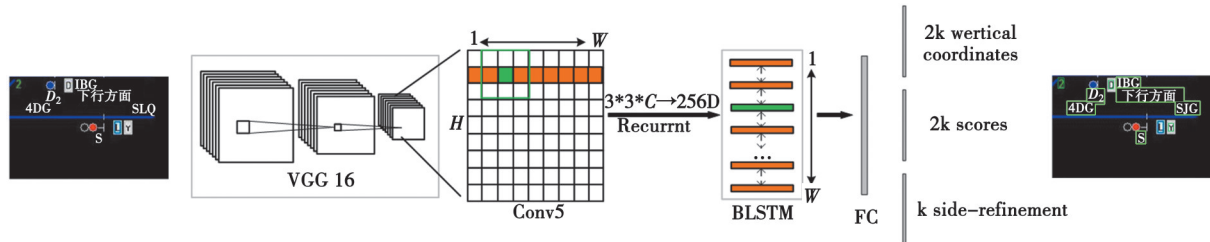


图1 CTPN网络结构

Fig. 1 CTPN network structure

CTPN神经网络类似于RPN,是一个全卷积网络,允许输入任意大小的图像<sup>[17]</sup>。输出的文本区域采用固定宽度16个像素,利用 $k$ 个垂直锚点来预测每个文本区域的 $y$ 坐标,从而获得极好的定位精度。通常 $k=10$ ,锚点高度在输入图像中从11个像素到273个像素变化(每次对输入图像的高除以0.7),10个先验框高度分别为 $\text{heights}=[11, 16, 23, 33, 48, 68, 97, 139, 198, 283]$ 。锚点边框相对垂直坐标的回归计算如式(1)所示

$$\begin{aligned} v_{\text{center}} &= (c_y - c_y^a) / h^a, & v_{\text{height}} &= \log(h / h^a) \\ v_{\text{center}}^* &= (c_y^* - c_y^a) / h^a, & v_{\text{height}}^* &= \log(h^* / h^a) \end{aligned} \quad (1)$$

式中, $v=(v_{\text{center}}, v_{\text{height}})$ 和 $v^*=(v_{\text{center}}^*, v_{\text{height}}^*)$ 分别是预测和真实的坐标。

对于小型文本行或文字的检测,当两文本区域的覆盖边被丢弃时,可能直接导致文本检测失败。为了解决这个问题,提出了一种边缘细化的方法,可以计算左右两侧水平方向上每个锚点(区域)的偏移量(称为边缘锚点或边缘区域),将用于优化最终的文本行边界框。通过式(2)分别计算左( $o_{\text{left}}$ )右( $o_{\text{right}}$ )两个边界的偏移值,

$$o = \frac{x_{\text{side}} - c_x^a}{w^a}, \quad o^* = \frac{x_{\text{side}}^* - c_x^a}{w^a}, \quad (2)$$

式中: $o$ 是 $x$ 轴上第 $k$ 个anchor预测的偏移值; $o^*$ 是 $x$ 轴上第 $k$ 个锚点的真实偏移值; $x_{\text{side}}$ 是回归得到的左边界或右边界; $x_{\text{side}}^*$ 是 $x$ 轴的真实边缘坐标; $c_x^a$ 是anchor中心横坐标; $w^a$ 是anchor固定宽度16像素。将 $x_{\text{leftside}}$ 与 $x_{\text{rightside}}$ 、 $x_{\text{leftside}}^*$ 与 $x_{\text{rightside}}^*$ 分别带入式(2)可得 $o_{\text{left}}, o_{\text{right}}$ 与 $o_{\text{left}}^*, o_{\text{right}}^*$ ,最终分别构成偏移量 $o_k=[o_{\text{left}}, o_{\text{right}}]$ , $o_k^*=[o_{\text{left}}^*, o_{\text{right}}^*]$ 。

针对文本和非文本的分类,采用交并比(intersection over union, IOU)作为指标,即模型预选框和实际预选框的交叠率。若anchor对应的预选框和实际预选框的IOU指标最大或大于0.7,则标记为正样本;若anchor对应预选框和实际预选框的IOU指标小于0.5,则标记为负样本。通过式(3)计算IOU指标。

$$\text{IOU} = \frac{S_{\text{anchorBox}} \cap S_{\text{groundTruth}}}{S_{\text{anchorBox}} \cup S_{\text{groundTruth}}}, \quad (3)$$

式中: $S_{\text{anchorBox}}$ 表示预测的预选框; $S_{\text{groundTruth}}$ 表示实际的预选框。

CTPN有3个输出共同连接到最后的FC层,这3个输出同时预测文本/非文本分数、垂直坐标和边缘细化偏移。为了满足同时预测以上3个参数,分别引入了 $L_s^{\text{cl}}, L_v^{\text{rc}}$ 和 $L_o^{\text{rc}}$ 3种损失函数,总体目标函数如式(4)所示。

$$L(s_i, v_j, o_k) = \frac{1}{N_s} \sum_i L_s^{\text{cl}}(s_i, s_i^*) + \frac{\lambda_1}{N_v} \sum_j L_v^{\text{rc}}(v_j, v_j^*) + \frac{\lambda_2}{N_o} \sum_k L_o^{\text{rc}}(o_k, o_k^*), \quad (4)$$

式中: $s_i$ 表示预测第 $i$ 个anchor的有效概率, $s_i^* \in \{0, 1\}$ 表示真实值; $v_j$ 和 $v_j^*$ 表示第 $j$ 个anchor预测和真实的 $y$ 坐标; $o_k$ 和 $o_k^*$ 表示第 $k$ 个anchor的 $x$ 轴预测和真实偏移量; $L_s^{\text{cl}}$ 表示区分文本和非文本的分类损失; $L_v^{\text{rc}}$ 和 $L_o^{\text{rc}}$ 表示

回归损失;  $\lambda_1$  和  $\lambda_2$  表示损失权重, 用来平衡不同的任务, 先验值分别设置为 1.0 和 2.0。  $N_s$ 、 $N_v$  和  $N_o$  分别表示  $L_s^{cl}$ 、 $L_v^{rc}$  和  $L_o^{rc}$  的 anchor 总数。

## 1.2 特征提取网络

特征提取作为 CTPN 神经网络的基础, 用于提取输入图片的表示高层语义信息。为了更好地验证特征提取的效果对后续检测结果的影响, 针对联锁上位机界面文本区域的定位, 本实验在相同数据集和相同定位模型 CTPN 下, 将常见的特征提取网络 VGG16、ResNet50、AlexNet、ZF 进行对比实验。结果如表 1 所示。文本常以行形式出现, 定位和识别要求较高, 必须定位和识别出完整的文本行。相较常规目标检测, 其难度更大, 评价指标也更加复杂, 通常用以下 3 个评价指标: 精确率(Precision)、召回率(Recall)和调和平均数(Hmean)。其中, 精确率=正确定位的预选框数量/检测框的总数量; 召回率=正确定位的预选框数量/真实框的总数量; 调和平均数=精确率 $\times$ 召回率 $\times$ 2/(精确率+召回率)。

实验结果表明, 使用 VGG16 作为 CTPN 的特征提取网络, 可以取得不错的效果, 调和平均数为 78.43%。而 ResNet50 网络使用了残差跳跃连接防止出现梯度消失现象, 调和平均数提升了 1.42%, 其性能表现略优于 VGG16。为了减少网络改进带来的不必要麻烦, 本文仍使用 VGG16 网络来提取图像特征。

表 1 不同特征提取网络对定位精度的影响

特征提取网络	评价指标		
	精确率	召回率	调和平均数
ResNet50	86.45	69.39	79.85
AlexNet	80.66	73.67	76.32
ZF	78.89	75.93	77.11
VGG16 (本文网络)	85.37	71.88	78.43

为了弥补未使用 ResNet50 网络带来的定位精度损失, 本文在训练阶段采用 dropout 的方法, 提高网络的泛化能力, 寻找网络的最高精度值。训练过程中, 通常初始化 dropout 值为 0.5。本次实验中, 选择 dropout=0.4、0.45、0.5、0.55、0.6 分别进行实验, 实验结果如表 2 所示。

表 2 不同 dropout 值对网络定位精度的影响

评价指标	dropout 取值				
	0.4	0.45	0.5	0.55	0.6
精确率	83.14	83.55	85.37	87.98	86.54
召回率	70.55	71.02	71.88	73.33	72.52
调和平均数	76.68	77.25	78.43	80.39	79.21

由表 2 结果可知, 当网络选择 dropout=0.55 时, 取得了最佳的定位效果。其中, 准确率为 87.98%, 召回率为 73.33%, 调和平均数达到了 80.39%, 均高于使用 ResNet50 网络的 3 项评价指标, 有效弥补了定位精度损失。

## 1.3 文本定位

基于 CTPN 的铁路计算机联锁上位机界面文本区域定位步骤如下:

第 1 步: 将联锁上位机界面图像输入卷积层, 用 VGG16 前 5 个卷积层提取特征, 得到含有空间特征等表示高层语义信息大小为  $N \times C \times H \times W$  ( $N$  为图像的数量,  $C$  为图像的长度,  $H$  为图像的高度,  $W$  为图像的宽度) 的特征图。

第2步:CNN和LSTM的中间层将VGG16最后一层卷积层输出的特征图转化为长度 $3 \times 3 \times C$ 的特征向量,得到输出为 $9C \times H \times W$ 的特征图。

第3步:双向LSTM根据卷积层输出的每一行所有窗口对应的 $3 \times 3 \times C$ 的特征向量得到含有空间特征和序列特征大小为 $N \times H \times W \times 256$ 的特征图。

第4步:全连接层将输入的特征转化为 $512 \times H \times W$ 的预选框特征图。

第5步:分类和回归预选框,得到检测框和得分。

第6步:使用文本行构造算法,得到需要检测的联锁界面文本行信息。

## 2 基于字符分割和改进的 AlexNet 的文本识别

### 2.1 字符分割简介

字符分割是指将原始图像分割成单个字符并提取保存为图像的过程<sup>[18]</sup>。常用的字符分割有投影分割法和连通域分割法<sup>[19]</sup>。针对不同厂家的联锁上位机界面分辨率不同、部分文本密集分布、文本重叠(分辨率低)、文本体积小等问题,使用连通域分割法对字符的分割能力有限,无法解决问题。因此,在相同数据集和相同定位及识别网络下,通过对常用文本分割算法进行对比实验,总结归纳,寻找适用于联锁上位机界面字符分割方法。结果如表3所示。其中,评价标准准确率=正确识别的文本数量/文本总数量。

表3 不同字符分割方法对识别精度的影响

Table 3 Influence of different character segmentation methods on recognition accuracy

字符分割方法	测试样本总数	识别错误数量	准确率/%
连通域分割法	300	47	84
水平投影分割法	300	54	82
垂直投影分割法	300	42	86
聚类分割法	300	51	83
水平投影+垂直投影分割法 (本文方法)	300	30	90

由表3可以看出,研究采用单一字符分割方法,虽然准确率都在80%以上,但均未达到良好的识别效果。而采用水平投影+垂直投影分割法,准确率达到90%,均优于其他字符分割方法。因此,研究采用水平投影+垂直投影分割法对联锁界面中的字符进行分割。

字符分割过程分为5步:

第1步:读取文本定位过程中得到的字符原图如图2所示。

第2步:为方便后期处理,对字符原图进行预处理,包括灰度化、二值化、去除孤立噪声、去除多余空白区域,得到如图3的字符图像。



图2 字符原图

Fig. 2 Character original image



图3 预处理后的字符图像

Fig. 3 Preprocessed character image

第3步:首先进行第1级分割(垂直投影分割),分割完毕后进行得分判定。

第4步:对不满足分割要求的图像,进行二级分割(水平投影分割),得到单个字符图像如图4所示,然后归一化每个字符图像尺寸,最后保存图像,完成分割任务。字符图像垂直和水平投影直方图如图5所示。



图4 分割结束后的字符图像

Fig. 4 Character image after cutting

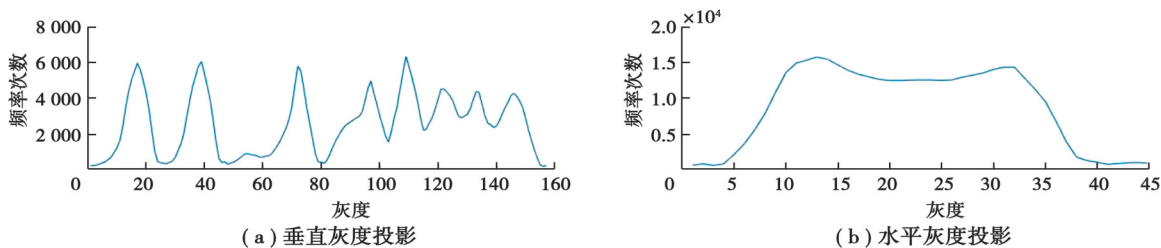


图5 字符图像垂直和水平投影直方图

Fig. 5 Character image vertical and horizontal projection histogram

## 2.2 改进的 AlexNet 网络简介

### 2.2.1 AlexNet 网络简介

第1个典型的卷积神经网络是 LeNet-5 网络,作为识别的开山之作,对后续卷积神经网络的发展打下了良好的理论基础。AlexNet 网络是在 LeNet-5 网络的基础上,主要采用卷积的方式提取特征,改变了原来模型的采样过程,增加池化操作,层数达到了较深的8层<sup>[20]</sup>,其网络结构如图6所示。AlexNet 采用双 GPU 进行计算,大大提高了运算效率,使得分类效果有了较大的提升。

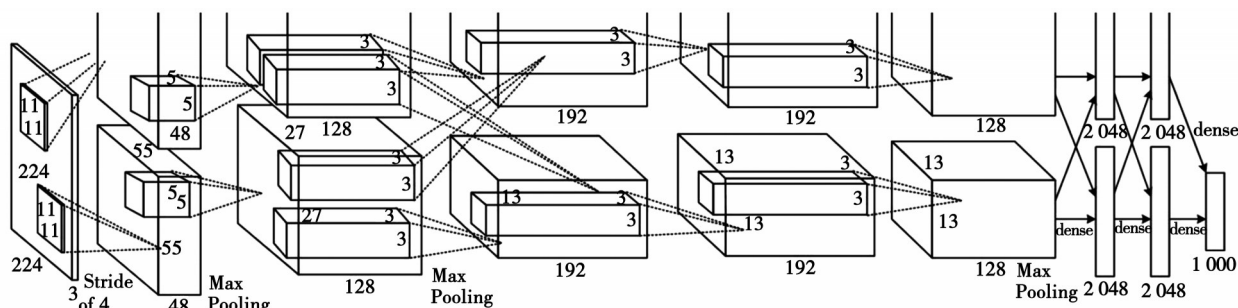


图6 AlexNet 网络结构

Fig. 6 AlexNet network structure

AlexNet 网络共包含5个卷积层,3个全连接层。第1和第2个卷积层后连有 LRN 层,每个 LRN 及第5个卷积层后是最大池化层,各个权重层均连有 ReLU 激活函数,有效解决了梯度弥散问题。全连接层使用 dropout,防止出现过拟合现象。

卷积层 C1 输入  $224 \times 224 \times 3$  的图像,使用96个  $11 \times 11$  的卷积核,提取图像表示高层语义信息的特征;输入到 ReLU 激活函数中;使用  $3 \times 3$  的池化单元;局部归一化后最终输出2组尺寸为  $27 \times 27 \times 48$  的特征图。卷积层 C2、C3、C4、C5 与 C1 类似,分别输出2组,每组大小分别为  $13 \times 13 \times 128$ 、 $13 \times 13 \times 192$ 、 $13 \times 13 \times 192$ 、 $6 \times 6 \times 128$ ,在此不再赘述。全连接层 FC6 和 FC7 均输出的4096个数据。FC7 与 FC8 进行全连接,输出预测结果。AlexNet 模型权重的设置可以减少训练误差,其权重调整规则如式(5)所示。

$$v_{i+1} = 0.9 \cdot v_i - 0.0005 \cdot \varepsilon \cdot w_i - \varepsilon \cdot \left\langle \frac{\partial L}{\partial w} \middle| w_i \right\rangle_{D_i}, \quad (5)$$

$$w_{i+1} = w_i + v_{i+1},$$

式中: $i$ 、 $\varepsilon$ 、 $w$  和  $v$  分别表示迭代次数、学习率、权重以及动量; $\left\langle \frac{\partial L}{\partial w} \middle| w_i \right\rangle_{D_i}$  表示目标函数关于导数  $D_i$  的平均值。

### 2.2.2 AlexNet模型改进

根据联锁上位机界面字符识别的特征,模型改进如下:

- 1)将单个字符尺寸统一设置为国家标准尺寸 $45 \times 90^{[21]}$ ;
- 2)适当减小卷积核的大小和池化区域数量,并增加一层卷积来获取更多表示高层语义信息的特征;
- 3)AlexNet网络模型具有1 000个输出,而标准站联锁上位机界面的汉字有48个左右,数字有10个,字母有25个,特殊字符有2个,总的分类数为85。为了防止不同站场图中汉字的差异,将改进模型的输出设置为100。改进的AlexNet网络模型如图7所示。

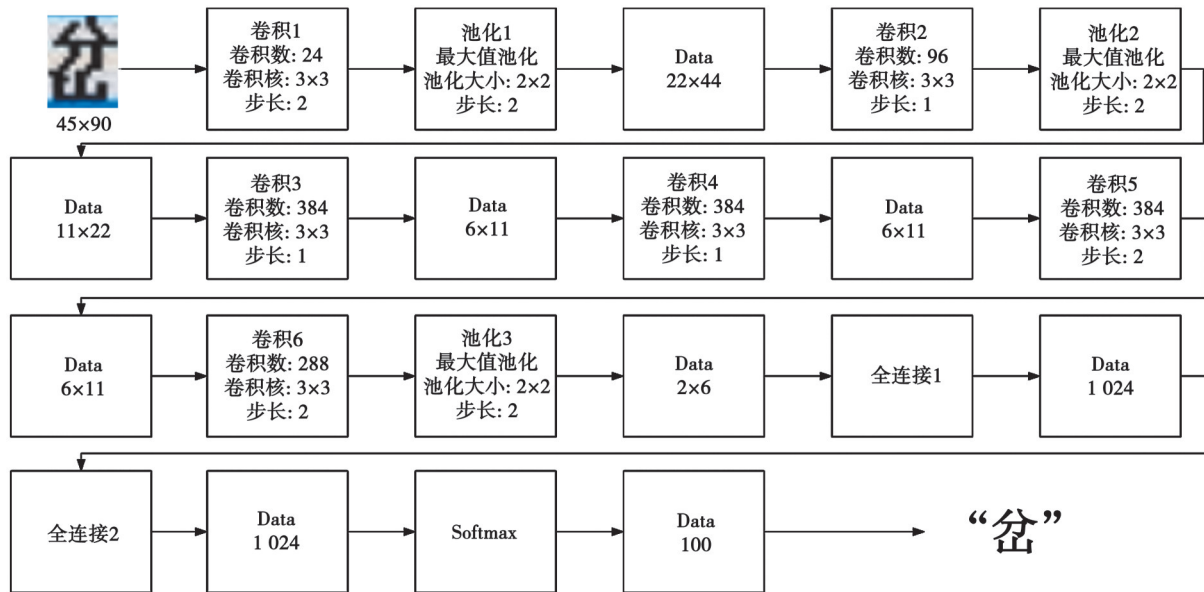


图 7 改进的 AlexNet 网络模型

Fig. 7 Improved AlexNet network structure

改进的网络输入图像尺寸为 $45 \times 90$ 的3通道RGB图像,随后经过6个卷积层,其中第1层和第6层卷积核的大小为 $3 \times 3$ ,步长为2,边界无填充。其余4层卷积核的大小为 $3 \times 3$ ,步长为1,边界填充1。第1~3层池化区域大小为 $2 \times 2$ ,步长为2。最终得到288个 $2 \times 6$ 的图像,将这288个图像全部展开,得到第一个全连接层,包含1 024个神经元,然后再连接一个包含1 024神经元的全连接层,最后是100个神经元的输出层,并用Softmax回归最后的输出。

## 3 模型的搭建与训练

### 3.1 数据集制作

为了增加图像的多样性,办理一条 $D_1 \sim D_0$ 的进路,然后办理自动模拟行车,采用录屏的方式采集视频,采用提取所有帧的方式进行图像采集。对采集的图像进行预分割,长4等分宽3等分,从 $1364 \times 736$ 大小的图像中获取 $341 \times 245$ 的图像,办理进路视频总的图像为144张,每个图像分为12张,最终获得1 728张图像作为联锁上位机界面数据集,每一张图像大小在10 kB到20 kB之间。部分联锁上位机界面数据集如图8所示。

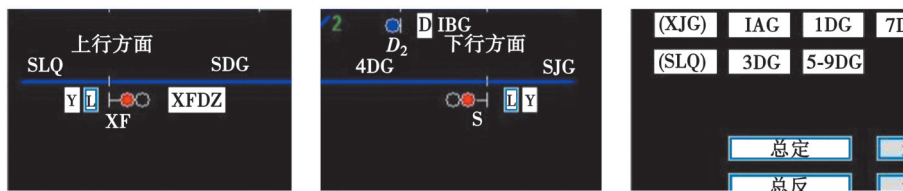


图 8 部分联锁上位机界面数据集

Fig. 8 Interface data set of partial interlocking upper computer

分析联锁上位机界面中的文本信息可知,存在着大写字母A~Z(为了区分字母O和数字0,将不采用字母O)、数字0~9、符号\(\道岔表示符号\)和\(-\(\区段表示符号\)\),除使用MINIST数据集和开源的大写字母数据集部分数据外,使用LabelImg制作其他字符数据集。部分字符数据集如图9所示。



图9 部分字符数据集

Fig. 9 Partial character data set

联锁上位机界面中存在部分汉字,汉字基本为印刷字体,且整体字体一致,因此数据集采用开源的印刷字体数据集为主,使用LabelImg制作联锁上位机界面中含有的汉字数据集为辅。部分汉字数据集如图10所示。



图10 部分汉字数据集

Fig. 10 Some Chinese character data sets

### 3.2 图像预处理

图像质量将直接影响文本定位和识别的效果,本文对图像进行RGB图像预分割、灰度化等预处理。

### 3.3 模型的构建与训练

CTPN模型采用0.01的学习率作为初值,在迭代40000步之后,学习率以0.1的衰减随迭代一步步减小,在经过60000步之后,模型的total\_loss收敛在0.08,即模型达到了87%以上的精确率。训练过程total\_loss收敛情况如图11所示。

AlexNet模型和改进AlexNet模型选取0.01作为学习率的初值,设置分类总数为100,经过200 epoch之后,识别准确率分别达到了86.2%和89.0%。两模型的识别准确率变化曲线如图12所示,损失函数变化曲线如图13所示。

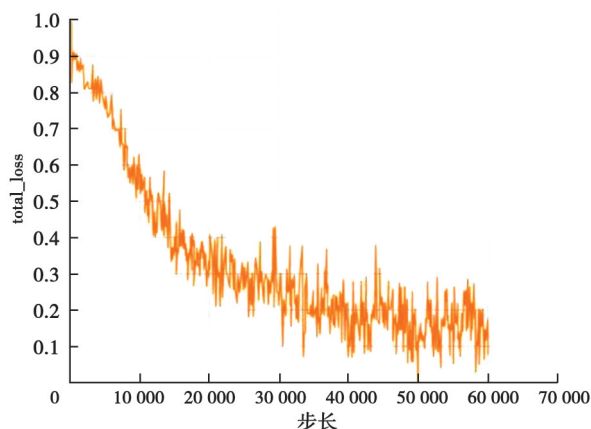


图11 CTPN模型损失变化曲线

Fig.11 CTPN model loss curve

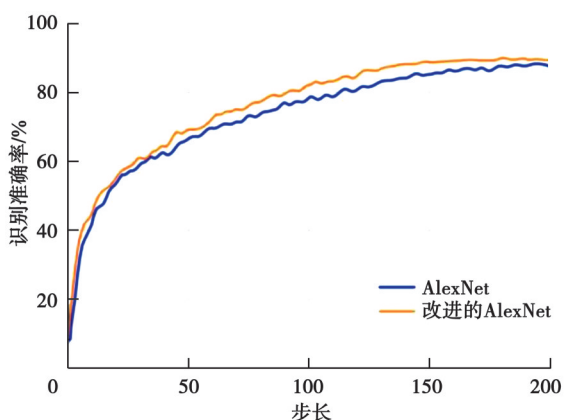


图12 改进的AlexNet模型和AlexNet模型识别准确率变化曲线

Fig. 12 Improved AlexNet model and AlexNet model recognition accuracy change curves

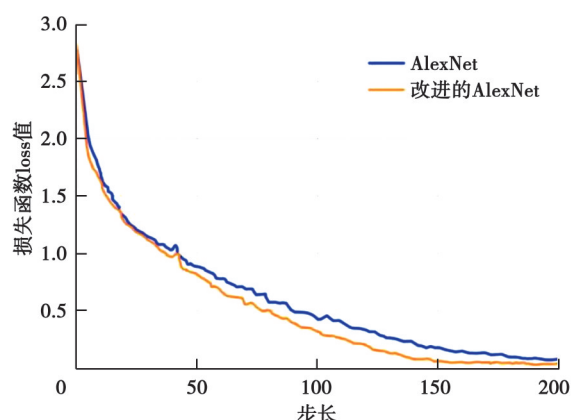


图13 改进的AlexNet模型和AlexNet模型损失函数变化曲线

Fig. 13 Improved AlexNet model and AlexNet model loss function change curves

## 4 实验结果与分析

### 4.1 系统配置

本文深度学习模型系统硬件配置为:深度学习环境在 anaconda 下,搭建 python3.6+cuda10.0+opencv3.3+tensorflow\_gpu-1.13.0+pytorch,CPU 为 Intel(R)Core(TM) i7-10750 CPU @ 2.60 GHz,显卡为 NVIDIA GeForce GTX 1650i,内存为 16 GB,外存为西数 WDC PC SN730 SDBPNTY-512G-1101。

### 4.2 文本区域定位不同方法对比实验

文本在长宽比、尺度和方向上都存在不同程度的变化,将直接导致定位难度增大,为了更好地适应联锁上位机界面文本的特点,本实验在相同数据集和相同特征提取网络 VGG16 下,将常见的目标检测网络进行对比实验,以寻找适合文本定位的网络。结果如表 4 所示。

表 4 不同定位方法对精度的影响

定位方法	评价指标		
	精确率	召回率	调和平均数
RPN	73.25	65.87	67.54
YOLOv3	85.57	74.68	79.87
Faster R-CNN	78.32	70.28	73.69
TextBoxes	83.52	73.38	78.46
PixelLink	86.54	73.12	79.61
CTPN (本文方法)	87.98	73.33	80.39

由表 4 可以看出,研究采用的 CTPN 模型在文本定位精度(精确率)上达到了 87.98%,均优于其他 5 种算法,满足联锁自动测试对于文本定位精度的要求。因此,CTPN 模型能够较好地解决联锁上位机界面文本检测困难的问题。使用训练好的 6 种网络模型进行实验,定位结果如图 14 所示,绿色框表示文本定位正确,红色框表示文本定位失败。

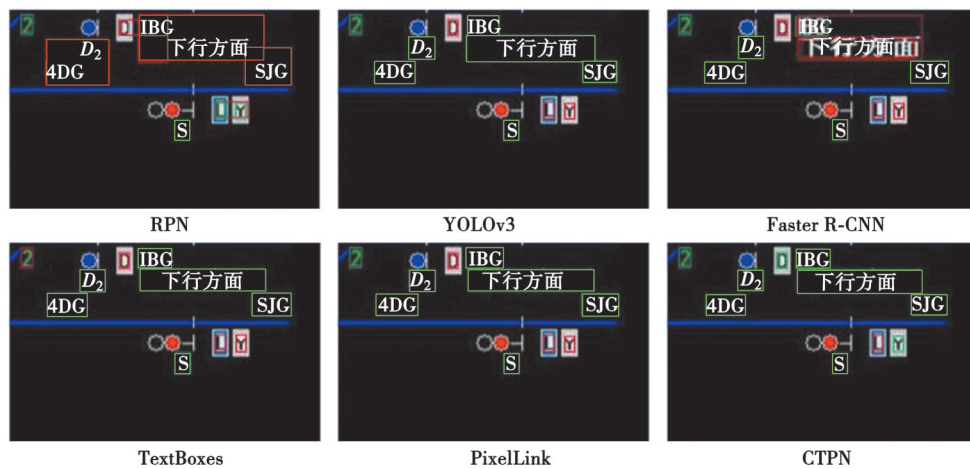


图 14 不同定位方法定位结果

Fig.14 Positioning results of different positioning methods

由图 15 可以看出,由于联锁上位机界面存在诸多干扰,基于 RPN 的方法,图像中很多文本定位失败,不具有良好的鲁棒性。Faster R-CNN 网络作为传统的目标检测模型,应用在文本定位过程中表现不佳,定位精度低于 CTPN 网络模型。YOLOv3 作为端到端实时目标检测方法,以速度快见长,但针对文本等小目标检

测,虽然定位精度达到了85.57%,高于Faster R-CNN网络,但定位整体效果仍然不好。作为一个快速而精确的文本检测器TextBoxes,采用基于全连接的卷积神经网络,文本定位过程中存在少于3个字符的文本行被忽略,检测效果不好。PixelLink采用实例分割的方法提高了定位精度,但相较于CTPN检测精度不够。CTPN模型采用双向LSTM提取文本上下文特征,提高了文本定位精度。因此,本文采用定位精度相对较高、应用相对较多的CTPN模型用于定位文本。

#### 4.3 不同文本识别方法对比实验

为了验证文本识别方法的有效性,在相同数据集和相同定位网络及字符分割方法下,对比了传统的模板匹配法、卷积循环神经网络(CRNN)方法、支持向量机(SVM)分类法、LeNet-5卷积神经网络方法、AlexNet卷积神经网络方法以及改进的AlexNet卷积神经网络方法。结果如表5所示。精度较高的四种深度学习模型识别准确率变化曲线如图15所示。

表5 不同文本识别方法对识别精度的影响

Table 5 Influence of different text recognition methods on recognition accuracy

方法	测试样本总数	识别错误数量	准确率/%
模板匹配	500	92	81.6
CRNN	500	61	87.8
SVM	500	81	83.8
LeNet-5	500	72	85.6
AlexNet	500	69	86.2
改进的AlexNet(本文方法)	500	55	89.0

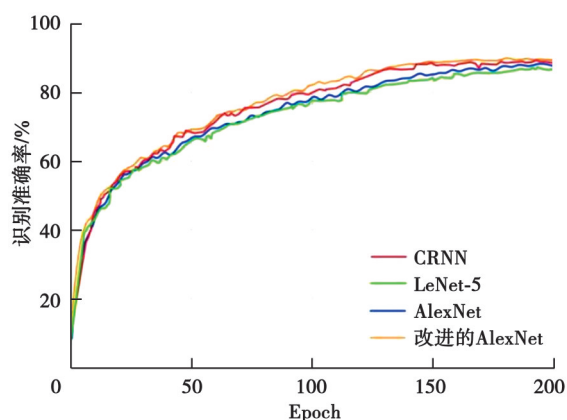


图15 4种深度学习模型识别准确率变化曲线

Fig. 15 Four kinds of deep learning model recognition accuracy rate change curves

由表5可以看出,传统的模板匹配方法和基于SVM的机器学习方法表现比较稳定,但识别准确率较低;LeNet-5网络和AlexNet网络识别准确率均低于改进的AlexNet网络;CRNN网络虽然在传统的数据集上表现良好,但针对联锁上位机界面,文本识别准确率为87.8%,略低于本文方法。通过对比实验,可知改进的AlexNet网络作为本文的文本识别模型,文本识别准确率为89.0%,满足联锁自动测试对于文本识别准确率的要求,可有效解决由于各种干扰带来的文本识别准确率较低的问题。

## 5 结 语

1)针对现有铁路联锁人工测试存在的不足,结合铁路计算机联锁上位机界面的文本特征,采用深度学习算法对其界面文本进行定位和识别,期望将深度学习应用到联锁自动测试中,实现联锁测试自动点击和结果自动分析的功能。

2)铁路计算机联锁上位机界面文本受各种干扰因素的影响,文本识别准确率低。本文采用水平投影和垂直投影相结合的字符分割方法,分割出单个字符图像,有效解决了由于分辨率低导致的字符粘连等问题。实验结果表明,通过传统的字符分割方法,优化后的字符图像,文本识别准确率较为理想。

3)针对铁路计算机联锁上位机界面文本数量和种类具有确定性的特征,结合 AlexNet 卷积神经网络在字符分类中表现出来的良好性能,对 AlexNet 网络的特征提取网络、卷积核、连接方式等结构进行改进,以达到适应联锁上位机界面文本特性。实验结果表明,通过对 AlexNet 网络的改进,文本识别准确率得到了较大提升。

4)本文对特征提取网络、文本定位网络、字符分割方法、文本识别方法都设计了一组对比实验,实验结果均证明了本文采用方法的有效性。

## 参考文献

- [ 1 ] 赵志熙. 微机联锁系统技术[M]. 北京: 中国铁道出版社, 1995.  
Zhao Z X. Microcomputer interlocking system technology[M]. Beijing: China Railway Publishing House, 1995. (in Chinese)
- [ 2 ] 徐洪泽. 车站信号计算机联锁控制系统: 原理及应用[M]. 北京: 中国铁道出版社, 2005.  
Xu H Z. Computer interlocking control system of station signal: principle and application[M]. Beijing: China Railway Publishing House, 2005. (in Chinese)
- [ 3 ] 曹爽. 航空软件测试用例自动生成方法研究[D]. 南京: 南京航空航天大学, 2010.  
Cao S. Research on automatic generation method of test case for aviation software[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2010. (in Chinese)
- [ 4 ] 孙婧鑫. Web应用的自动化测试研究[J]. 电脑与信息技术, 2020, 28(5): 43-44  
Sun J X. Research on automated testing of web applications[J]. Computer and Information Technology, 2020, 28(5): 43-44. (in Chinese)
- [ 5 ] 梁艺凡. A\*算法在计算机联锁上位机软件中的应用研究[D]. 兰州: 兰州交通大学, 2013.  
Liang Y F. Research on A\* algorithm application in upper computer software of computer interlocking[D]. Lanzhou: Lanzhou Jiaotong University, 2013. (in Chinese)
- [ 6 ] 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, 2016: 56-72.
- [ 7 ] 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.
- [ 8 ] Liao M H, Shi B G, Bai X, et al. Textboxes: a fast text detector with a single deep neural network[C]//AAAI Conference on Artificial Intelligence. San Francisco, CA, USA: AAAI, 2017: 4161-4167.
- [ 9 ] Deng D, Liu H F, Li X L, et al. PixelLink: detecting scene text via instance segmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 6773-6880.
- [ 10 ] Zhang Z, Zhang C Q, Shen W, et al. Multi-oriented text detection with fully convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016. Las Vegas, NV, USA: IEEE, 2016: 4159-4167.
- [ 11 ] 刘澍. Research on key technologies of license plate character recognition system[D]. 武汉: 华中师范大学, 2019.  
Liu S. Research on key technologies of license plate character recognition system[D]. Wuhan: Central China Normal University, 2019.
- [ 12 ] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27.
- [ 13 ] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [ 14 ] 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.

- [15] 艾鑫, 邹琪, 罗常津. 面向精确定位的列车车号文本定位与识别[J]. 计算机辅助设计与图形学学报, 2020, 32(11): 1863-1870.  
Ai X, Zou Q, Luo C J. Precise localization and recognition of train characters[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(11): 1863-1870. (in Chinese)
- [16] 桑军, 郭沛, 项志立, 等. Faster-RCNN 的车型识别分析[J]. 重庆大学学报, 2017, 40(7): 32-36.  
Sang J, Guo P, Xiang Z L, et al. Vehicle detection based on faster-RCNN[J]. Journal of Chongqing University, 2017, 40(7): 32-36. (in Chinese)
- [17] 朱意星. 基于深度学习的文本与遥感图像目标检测研究[D]. 合肥: 中国科学技术大学, 2020.  
Zhu Y X. Research on deep learning based object detection for text and aerial image[D]. Hefei: University of Science and Technology of China, 2020. (in Chinese)
- [18] Wang X, Du H, Wen X P. Research on segmentation and recognition of printed Chinese characters[J]. Journal of Physics: Conference Series, 2019, 1237(2): 022011.
- [19] 李祥鹏, 闵卫东, 韩清, 等. 基于深度学习的车牌定位和识别方法[J]. 计算机辅助设计与图形学学报, 2019, 31(6): 979-987.  
Li X P, Min W D, Han Q, et al. License plate location and recognition based on deep learning[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(6): 979-987. (in Chinese)
- [20] 纪荣嵘, 林绍辉, 晁飞, 等. 深度神经网络压缩与加速综述[J]. 计算机研究与发展, 2018, 55(9): 1871-1888.  
Ji R R, Lin S H, Chao F, et al. Deep neural network compression and acceleration: a review[J]. Journal of Computer Research and Development, 2018, 55(9): 1871-1888. (in Chinese)
- [21] 罗月童, 黄文, 饶永明, 等. 基于深度学习的芯片表面字符识别方法[J]. 合肥工业大学学报(自然科学版), 2020, 43(4): 463-467.  
Luo Y T, Huang W, Rao Y M, et al. Method of chip surface character recognition based on deep learning[J]. Journal of Hefei University of Technology (Natural Science), 2020, 43(4): 463-467. (in Chinese)

(编辑 侯 湘)