

文章编号: 2095-2163(2020)07-0040-05

中图分类号: TP391

文献标志码: A

# 基于深度学习的驾驶证识别方法研究

姚 砺, 王昭丽

(东华大学 计算机科学与技术学院, 上海 200000)

**摘要:** 依托不断发展的电子信息技术,对和车辆信息密切相关的驾驶证进行 OC 光学字符识别(OCR),成为实现智能交通的重要步骤。实际拍摄图像中,由于拍摄环境差别明显,光照不均匀和倾斜等情况普遍存在,识别时需要根据图像进行差异化处理,处理过程中很容易丢失文字结构信息,导致字符识别错误,降低系统识别率。本文针对驾驶证识别,采用“图片预处理+提取+识别”的整体思路,提出一种基于深度学习的机动车驾驶证检测与识别算法。首先,进行图片预处理,针对驾驶证图片存在的背景复杂、角度倾斜的问题,提出对边缘轮廓的位置进行定位以完成倾斜校正;其次,使用 CTPN 算法对图片中文字区域进行检测;最后,用 RCNN 算法对检测出的文字行进行文字检测,得到文字识别结果。

**关键词:** 驾驶证识别; 文字检测; CTPN; 文字识别; CRNN

## Research on driver license recognition method based on deep learning

YAO Li, WANG Zhaoli

(College of Computer Science and Technology, Donghua University, Shanghai 200000, China)

**[Abstract]** Relying on the continuous development of electronic information technology, OCR (Optical Character Recognition) for driving licenses closely related to vehicle information has become an important step to achieve intelligent transportation. In the actual captured images, due to obvious differences in shooting environment, uneven illumination and tilting, etc., differentiation needs to be performed according to the image during recognition. It is easy to lose text structure information during processing, resulting in incorrect character recognition and reduced system recognition rate. The main research content of this article is for the recognition of driver's license, and the overall idea adopted is the combination of picture preprocessing, extraction and recognition. This paper proposes a deep learning - based vehicle driving license detection and recognition algorithm. The algorithm first preprocesses the picture. In view of the problem of complex background and angle tilt of the driver's license picture, it is proposed to position the edge contour of the driving license to complete the tilt correction. Then use CTPN algorithm to perform text detection on the text area in the picture on the corrected image, and finally the RCNN algorithm is used to perform character recognition on the detected character lines to obtain the character recognition result.

**[Key words]** Driver's license recognition; Text detection; CTPN, Text recognition; CRNN

### 0 引言

在当前移动互联网快速发展的大环境下,使用手机管理身份证件信息成为一个比较热门而有价值的方向。机动车驾驶证是车辆管理的主要证件,随着社会信用体系不断完善,驾驶证的应用范围不再局限于交通领域,还能作为个人信用凭证,应用到保险、银行等部门<sup>[1]</sup>。通过手机拍照采集信息,进行驾驶证识别,可以解决传统的人工录入效率低、成本高、错误率难以控制的问题。同时方便了交通警察等执法人员进行户外执法检查。

近年来,对于证照识别的研究是一个比较热门的研究方向,相关工作可以分为图像预处理、文字检测和文字识别几个方面<sup>[2]</sup>。其中,图像预处理一般经过以下步骤:

(1) 倾斜校正: 主要有水平投影法、直线统计法、轮廓提取法。

(2) 二值化: 分为全局和局部二值化两种。

(3) 去噪: 常用中值滤波、高斯滤波等方法。关于文字检测,研究方法从以 MSER 和 SWT 为代表的连通域分析方法,发展到了目前基于深度学习技术的 CTPN, Seglink, EAST 等算法。针对文字背景形态不同,有不同的适合使用的算法。比如适合检测水平文本的 CTPN 算法,可检测弯曲文本的 SegLink 算法,可检测多尺度文本的 EAST 算法等,还有 Faster rcnn,它是由 fast rcnn 改进而来,而 fast rcnn 也是由基本的神经网络 cnn 发展而来<sup>[3]</sup>。文字识别方面,基于深度学习的端到端 OCR 技术有两大主流技术: CRNN OCR 和 attention OCR。其主要区别在于最后的输出层(翻译层),即如何将网络学习到的序列特征信息转化为最终的识别结果。这两大主流技术在其特征学习阶段都采用了 CNN + RNN 的网络结构,CRNN OCR 在对齐时采取 CTC 算

作者简介: 姚 砺(1967-),男,博士,副教授,主要研究方向:图像处理;王昭丽(1995-),女,硕士研究生,主要研究方向:图像处理。

通讯作者: 王昭丽 Email: 1422183563@qq.com

收稿日期: 2020-05-11

法,而 attention OCR 采取 attention 机制。

对于驾驶证识别,拍摄驾驶证图像时的光照条件存在巨大差异,容易造成图像偏暗或偏亮,给背景分割造成很大困难。其次,拍摄的驾驶证图像通常具有一定程度的倾斜、偏移,算法需要确定驾驶证的轮廓区域,进行校正。在校正的过程中,还需要排除拍摄背景和驾驶证上覆膜的影响<sup>[4]</sup>。此外,驾驶证上的底纹,在光照偏暗的场景下,易误判为文字边缘,对文字识别准确率有一定的影响。针对上述问题,本文对机动车驾驶证识别算法做了进一步研究,基于背景校正技术,通过区域提取算法对驾驶证边缘轮廓的位置进行定位以完成倾斜校正等预处理,通过 CTPN 算法对图像进行文字区域检测,使用 RCNN 算法文字区域进行识别。

本文提出的驾驶证识别算法对于文字区域检测与文字识别均有较好的识别准确性。

## 1 基于深度学习的驾驶证识别方法

本文首先采用轮廓提取法对驾驶证照片进行倾



图1 基于深度学习的驾驶证识别方法

Fig. 1 Driver license recognition method based on deep learning

因为获取图片的方式是手机拍照,拍摄的驾驶证图像通常具有一定程度的倾斜、偏移,需要对图像进行倾斜矫正。

考虑到驾驶证边框比较清晰的特点,本文采用基于透视变换的图像倾斜纠正方法。首先,采用卷积核大小为  $3 \times 3$  的高斯滤波对图像去噪处理;其次,采用 Canny 边缘检测,提取出图像中证照的轮廓,再采用 Hough 直线检测获取证照的四条边框线;最后,采用透视变换进行图像倾斜纠正。

通用变换矩阵如公式(1)所示。

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (1)$$

转换后的横纵坐标计算公式(2)如下:

$$\begin{aligned} \hat{X} &= \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \\ \hat{Y} &= \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}. \end{aligned} \quad (2)$$

通过公式(1)与(2)计算可完成透视变换,得到最终的证照正视图。本文的倾斜纠正效果如图2所示。

斜矫正以得到水平的驾驶证。

传统的证照识别技术在图像预处理后,一般都要经过版面分析,模板匹配来进行文字区域检测,这就要提前制作模板。本文采用深度学习的方法来进行文字区域检测。对于驾驶证照来说,检测的都为水平文本,由 faster rcnn 改进而来的 CTPN 对于水平文本的检测精度高,因此本文基于 CTPN 技术进行文字检测。

传统的文字识别需从连通域分析、字符切割步骤开始,再到字符识别、语义校正,最后得到文字识别结果,这样中间过程较多会导致累计误差加大,影响了最终的识别精度。本文采用端到端的 CRNN 算法从文本图像直接识别出证照文字,减少了中间环节与累计误差。具体方法如图1所示。

### 1.1 图片预处理

图像预处理是为了解决或减少拍摄的驾驶证图片因倾斜,光照不均等因素对后续文字检测与识别造成的影响。

### 1.2 文字检测

文本检测的难点在于对一个完整的文本线的检测,同一文本线的字符可能存在差异大、距离远的情况,每条文本线的长度可能不同,难度比检测单个目标更大<sup>[5]</sup>。传统方法是先检测字符,再连成本文,其缺点是没有考虑到上下文,步骤复杂且检测精度易受拍摄条件影响。

本文在文字检测前对图像进行倾斜纠正预处理,所需检测的文字都是水平文本,因此本文采用 CTPN 算法来进行文字检测。

CTPN 采用与 RPN 类似的 anchor 机制,固定 width 而只预测 anchor 的 y 坐标和高度,达到比较精准的 text proposal 效果。同时,引入了 RNN,使用 BLSTM 使得预测更加精准。CTPN 在自然场景上下文提取的效果较好,不同于传统的 bottom-up 检测单个字符再去连接文本线的方法,其准确性主要依赖于单个字符的识别,错误会累积,其使用的仅仅是 low-level 的 feature;而 CTPN 提取的是深度特征,采用 anchor 机制做精准预测,用循环神经网络对 anchor 识别的区域进行连接,精度要高很多。CTPN 对于水平文本识别目前来说准确率相对是很高的,但将它用于驾驶证识别的研究并不多。



(a) 原始拍摄图像  
(b) 倾斜纠正图像  
(a) Original captured image  
(b) Tilt correction image

图2 倾斜纠正效果

Fig. 2 Effect of Tilt correction

本文基于CTPN的驾驶证文字检测具体实现流程包含3个部分:检测小尺度文本框,循环连接文本框,文本行边细化。

本文所采用的CTPN网络结构如图3所示。

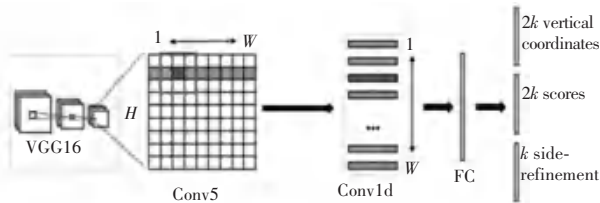


图3 CTPN网络结构图

Fig. 3 Network structure diagram of CTPN

具体的实现步骤如下:

(1)使用VGG16作为base net提取特征,得到conv5\_3的特征作为feature map,大小是 $W \times H \times C$ ,这里 $W$ 和 $H$ 都为输入图像的 $1/16$ ;

(2)在feature map上本文使用大小为 $3 \times 3$ 的滑动窗进行滑动,每个窗口都能得到一个长度为 $3 \times 3 \times C$ 的特征向量,每个滑动窗口中心都会预测 $k$ 个

相对于anchor的偏移;

(3)使用一个双向的LSTM,将上一步得到的特征输入到这个双向LSTM中,得到长度为 $W \times 256$ 的输出,接一个512的全连接层,准备输出;

(4)输出层部分主要有3个输出。 $2k$ 个垂直坐标(vertical coordinate),因为一个anchor用的是中心位置的高( $y$ 坐标)和矩形框的高度两个值表示的,所以一个用 $2k$ 个输出,这里输出的是相对anchor的偏移; $2k$ 个评分(score),因为预测 $k$ 个text proposal,所以有 $2k$ 个分数,text和non-text各有一个分数; $k$ 个边缘细化(side-refinement),用来精修文本行的两个端点,表示的是每个proposal的水平平移量;

(5)使用一个标准的非极大值抑制算法来滤除多余的text proposal;

(6)使用基于图的文本行构造算法,将得到的文本段合并成文本行。

CTPN算法识别驾驶证结果如图4所示。



(a) 证照原始图  
(b) ctpn 文本检测效果  
(a) Original photo  
(b) Text detection effect of ctpn

图4 CTPN文本检测结果

Fig. 4 Text detection results of CTPN



由图4(b)可见,检测出一些无用的小文本框。针对这个问题,本文先根据检测出的证照四周边框高宽进行归一化处理,根据有效文本区域的高度设置过滤条件,从而将文本高度过小的无用文本框过滤掉。而对于证照左下部的交警支队红章,由于驾驶证的版面布局固定,各个信息区域在驾驶证中的坐标比例也固定,采用区域位置过滤,文本检测效果如图5所示。



图5 最终文本检测结果

Fig. 5 Final text detection result

### 1.3 文字识别

识别的图像是拍照得来的,而拍出的照片往往会存在光照不均等问题,图像清晰度和光照等问题一直是文字识别的一大难点<sup>[6]</sup>,本文采用CRNN算法进行文字识别。

CRNN把CNN做图像特征工程的潜力与LSTM做序列化识别的潜力结合,既提高了鲁棒特征,又避免了传统算法中难度极高的单字符切分与单字符识别,同时序列化识别也嵌入时序依赖。CRNN将语音识别领域的CTC-LOSS引入图像,采用CTC模型实现序列标签转录。CTC按照RNN的每帧输出预测相应的标签,忽视了标签所在的具体位置。因此,CTC预测序列标签避免了标注单个字符具体位置的工作。目前使用CRNN算法进行文字识别较多,但将其与CTPN结合用于特定驾驶证识别研究还不够成熟。

CRNN由三大部分架构而成,分别对应卷积层、循环层和转录层。CNN部分用于底层的特征提取,RNN采取了BiLSTM,用于学习关联序列信息并预测标签分布,CTC用于序列对齐,输出预测结果。

本文第一部分CNN采取了经典的VGG16,在VGG16的第3第4个max pooling层采取的是1×2的矩形池化窗口,这有别于经典的VGG16的2×2的正方形池化窗口,从2×2改为1×2是为了保证提出的特征具有横向的长度,有利于识别较长的文本。VGG16部分还引入了BN模块加速模型收敛。本文CRNN的输入是灰度图像,即图像深度为1。针对

字符拉伸导致识别率降低的问题,保持输入图像尺寸比例,但是图像高度必须统一为32个像素,卷积特征图的尺寸动态决定LSTM时序长度。

输入图像,做如下处理:

(1)将图像缩放到 $32 \times W \times 1$ 大小,经过CNN后变为 $1 \times (W/4) \times 512$ ;

(2)针对LSTM,设置 $T = (W/4)$ , $D = 512$ ,即可将特征输入LSTM;

(3)LSTM有256个隐藏节点,经过LSTM后变为长度为 $T \times nclass$ 的向量,再经过softmax处理,列向量每个元素代表对应的字符预测概率,再将这个T的预测结果去冗余合并成一个完整识别结果即可。

## 2 实验结果与分析

### 2.1 实验环境与数据

本文的实验平台采用Linux系统下的Ubuntu16.04, GTX1060,3.2G Hz,内存为16 GB。本文训练CTPN使用的数据集为ICDAR17MLT标准数据集,训练CRNN使用的数据集为合成的中文数据集。

### 2.2 结果分析

采用倾斜矫正、去噪对图片进行预处理,将驾驶证图片通过CTPN识别,得到的实验结果如图5所示。由图5可见,姓名,性别,证号,地址等关键字区域使用CTPN算法基本可以正确检测。

为了验证CTPN算法的有效性,实验使用280张驾驶证图片,用于计算关键区域的检测准确率。统计正确检测出的文本框个数,错误检测到的文本框个数以及未检测到的文本框个数,通过和实际文本框的比值得到算法的准确率。统计结果见表1。

表1 驾驶证文字检测准确率

Tab. 1 Accuracy of text detection of driver's license

驾驶证文字区域检测	检测准确率/%
姓名	96.5
证号	97.8
地址	95.4

因为识别的驾驶证上中文很多,所以训练CRNN使用的数据集为合成的中文数据集,利用图像处理技术手工大批量生成文字图像,一共360万张图像样本,划分了训练集和测试集(10:1),并单独存储为两个文本文件,对该数据集进行lmdb格式转化,得到训练集和测试集的lmdb。接下来进行网络训练,文字类别数为6732,一共训练20个epoch, batch\_Size设置为64,所以一共是51244次迭代/epoch。测试时需初始化CRNN网络,载入训练好的

(下转第48页)