

# 改进 CRNN 的车牌识别方法

林立雄<sup>1</sup>, 庄裕富<sup>2</sup>, 何洪钦<sup>3</sup>, 郑佳春<sup>1</sup>

(1. 集美大学海洋信息工程学院, 福建 厦门 361021; 2. 福州大学机械工程及自动化学院, 福建 福州 350108;  
3. 浙江大华技术股份有限公司, 浙江 杭州 310051)

[摘要] 针对传统车牌识别方法在车牌模糊和光照复杂的场景下难以快速准确识别车牌信息的问题, 为提高网络的特征提取能力, 将带残差的 ResNet 引入卷积循环神经网络(convolutional recurrent neural network, CRNN), 提出了 CRNN\_ResNet 车牌文本识别算法。该方法仅需 0.012 s 就能完成一张车牌图像的识别, 在 CCPD 公开数据集上的识别准确率达到 98.5%。

[关键词] 车牌识别; 深度神经网络; 卷积循环神经网络

[中图分类号] TP 391.41

## License Plate Recognition Based on CRNN

LIN Lixiong<sup>1</sup>, ZHUANG Yufu<sup>2</sup>, HE Hongqin<sup>3</sup>, ZHENG Jiachun<sup>1</sup>

(1. School of Ocean Information Engineerin, Jimei University, Xiamen 361021, China;

2. School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China;

3. Zhejiang Dahua Technology Co. Ltd., Hangzhou 310051, China)

**Abstract:** Aiming at the problem that the traditional license plate recognition method was difficult to recognize license plate information quickly and accurately under the scene of fuzzy license plate and complex illumination, a fast license plate recognition method based on improved convolutional recurrent neural network (CRNN) was proposed. By introducing ResNet with residual into CRNN to improve the feature extraction ability of the network, this new method was named CRNN\_ResNet license plate text recognition algorithm. This method only needed 0.012 s to recognize a license plate image, and the recognition accuracy reached 98.5% on the open data set of Chinese City Parking Dataset (CCPD).

**Keywords:** license plate recognition; deep neural network; convolutional recurrent neural network(CRNN)

## 0 引言

车牌识别技术能为智能交通管理提供数据支持, 为道路规划提供科学的指导; 还能够适时防范机动车辆被盗窃、盗抢、假牌、走私和黑市交易等犯罪行为<sup>[1-2]</sup>。传统车牌识别方法通过图像处理、特征提取方法获取图像中的车牌图像, 再利用轮廓提取<sup>[3]</sup>、垂直投影<sup>[4]</sup>、灰度统计<sup>[5]</sup>和模板匹配<sup>[6]</sup>等方法将车牌图像中的字符逐一分割, 最后通过边界特征<sup>[7]</sup>、HOG 特征<sup>[8]</sup>、模板匹配<sup>[9-10]</sup>、支持向量机<sup>[11]</sup>和模板

[收稿日期] 2022-09-09

[基金项目] 国家自然科学基金资助项目“作业型飞行机器人的动态飞行抓取运动动力学规划研究”(61803089); 福建省自然科学基金资助项目“未知环境下空地机器人协同导航多源信息融合技术研究”(2019J05024); 厦门市海洋智能物联终端研发与应用重点实验室项目(B18208)

[作者简介] 通信作者: 林立雄(1985—), 博士, 副教授, 从事无人控制系统、视觉处理等方向研究。E-mail: elelinlixiong@139.com.

匹配结合神经网络<sup>[12]</sup>等方法识别车牌字符,但其模型存在设计复杂、泛化能力差、效率低等问题。

近年来,得益于丰富的数据集和优异的深度神经网络训练技术,基于深度学习的文本识别技术也逐渐被应用于车牌识别<sup>[13]</sup>。深度学习方法通过减少繁琐的步骤和计算复杂度,简化文本检测的整体流程,使研究人员能够从繁重的人为特征设计任务中解脱出来,从而推动文本识别进一步发展。文本识别方法分为单字符识别和文本行识别。单字符识别通常包含三个关键步骤:正则化、分割和识别<sup>[14-15]</sup>。正则化的目的是去除各种类型的噪声,将不规则的输入图像转化为正则图像;分割是将正则化的文本图像切割成多个包含单个字符的子图像;识别是以连续的方式逐个处理单个字符子图像,识别单个字符子图像的内容并将它们连接成文本字符串。由于单字符识别不适用于模糊及字符粘连的文本图像,因此目前更主流的方法是直接对文本行进行识别。根据文本行具有序列特征这一性质,通过学习文本行的序列特征,再解码成具体文本内容,这样效率更高,这种方法包括图像编码、特征编码和文本解码三部分。图像编码模块采用卷积提取输入图像的关键特征,并将其编码为向量序列,每个向量对应图像的局部区域;特征编码模块利用 RNN (recurrent neural network) 将得到图像的特征向量进一步映射到潜在空间;文本解码模块利用另一个 RNN 将潜在空间中编码的信息解码为扩展字母表上的一系列概率分布,使用概率向量对输出文本序列进行预测。Shi 等<sup>[16]</sup>提出的 CRNN (convolutional recurrent neural network) 先通过卷积层提取图像的空间特征,再利用双向 LSTM (long short-term memory) 提取序列特征,最后通过转录部件得到标签序列,占用的内存空间小,效率高。Wojna 等<sup>[17]</sup>提出了一种基于注意力机制的文本识别架构,利用注意力机制对 CNN (convolutional neural network) 提取的特征加权后再传入 RNN,模型简单,通用性强。Shi 等<sup>[18]</sup>提出的 ASTER (attentional scene text recognizer with flexible rectification) 包括矫正网络和识别网络两个部分,矫正网络负责把图像中不规则的文本区域变成规则的文本区域图像,再利用识别网络对文本进行检测识别。Luo 等<sup>[19]</sup>提出的 MORAN (multi-object rectified attention network) 由矫正网络 MORN (multi-object rectified network) 和识别网络 ASRN (attentionbased sequence recognition network) 组成,其中 MORN 摆脱了几何变换的限制,但该方法仅能矫正垂直方向的畸变,且训练难度大。Shi 等<sup>[20]</sup>提出的 RARE (robust text recognizer with automatic rectification) 能够识别形状不规则的文本,通过空间变化网络对不规则文本图像进行矫正,从而得到可读性更强的文本图像,再对文本进行识别。

传统车牌识别方法依赖人为设计的特征,泛化能力差;而 CRNN 网络占用的内存空间小、效率高,理论上适用于任意长度的文本序列。本文针对 CRNN 文本识别方法在车牌识别上精度不足的问题展开研究,研究经典 CNN 网络与 ResNet 网络,并对 CRNN 网络卷积层进行改进,以期提出精度更高的车牌字符识别方法。

### 1 CRNN\_ResNet 车牌识别网络

CRNN 网络共包含卷积层、循环层和转录层三个部分。卷积网络层使用的是 CNN 网络,输入图像大小为 (32, W, 3), 经过 CNN 网络提取特征后,会生成大小为 (1, W/4, 512) 的特征图。循环网络层使用两层双向 LSTM 网络,用于提取文字的序列特征。循环网络层采用的最大时间长度  $T = W/4$ 。即,双向 LSTM 的输入  $\mathbf{x} = (x^1, x^2, \dots, x^T)$ , 其中每一列  $\mathbf{x}^t = (x^t_1, x^t_2, \dots, x^t_{512})$ ; 输出  $\mathbf{y} = (y^1, y^2, \dots, y^T)$  是一个概率矩阵, 每一列  $\mathbf{y}^t = (y^t_1, y^t_2, \dots, y^t_n)$ , 其中  $n$  表示需要识别的字符集合长度。转录层使用 CTC (connectionist temporal classification) 算法将双向 LSTM 的输出结果转换为标签序列 (见图 1)。

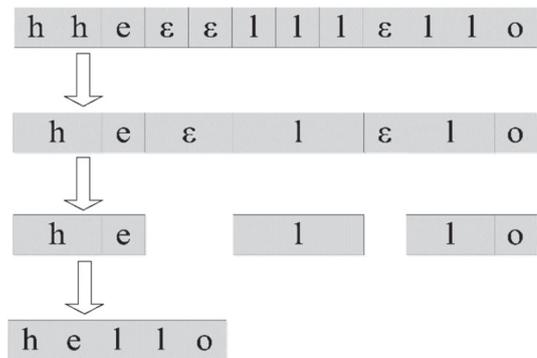


图 1 CTC 去重示意图  
Fig.1 Schematic diagram of CTC deduplication

CTC 算法引入符号 “ε” 用于表示字符间的空白处, 得到一个包含字符和空白占位符的集合:  $L^T \in L'$ 。其中:  $L' = L \cup \{\text{blank}\}$ ;  $L$  为字符集; blank 为空白占位符集。以空白占位符为边界进行去重操作, 得到最终的输出结果, 且最终结果的最大长度不超过循环层的时间序列  $T$ , 并将这一过程定义为 B 变换:  $L^T \rightarrow L \leq T$ 。

在 CTC 中, 给定输入  $\mathbf{x}$ , 则输出  $l$  的概率为:  $P(l|\mathbf{x}) = \sum_{\pi \in B^{-1}(l)} P(\pi|\mathbf{x})$ 。其中:  $\pi$  表示经过 B 变换后输出为  $l$  的路径。对于任意一条路径  $\pi$  有:  $P(\pi|\mathbf{x}) = \prod_{t=1}^T y_{\pi_t}^t, \forall \pi \in L^T$ 。其中  $y_{\pi_t}^t$  表示时刻为  $t$  时路径  $\pi$  的成员  $y$ 。将图 1 所示的这条路径定义为  $\pi_1$ , 那么路径  $\pi_1 = (\text{hhεεlllεllo})$ , 即:

$$P(\pi_1|\mathbf{x}) = y_h^1 y_h^2 y_\epsilon^3 y_l^4 y_l^5 y_l^6 y_l^7 y_\epsilon^8 y_l^9 y_l^{10} y_l^{11} y_o^{12}。$$

网络训练时, 通过梯度  $\frac{\partial P(l|\mathbf{x})}{\partial \omega}$  对循环层的参数  $\omega$  进行调整。理论上加深网络可以提高网络的特征提取能力, 但随着网络加深到一定层数后, 无法保证更深层网络提取到的特征是对整个网络更有用的, 因此整个网络在训练时很容易出现梯度爆炸和梯度消失的现象, 使精度不升反降。为解决上述问题, 拟采用 ResNet<sup>[12]</sup> 网络。并在 ResNet 网络中加入残差连接, 每个单元的输入为  $\mathbf{x}$ , 主路上  $\mathbf{x}$  会经过两次权重处理和一次 Relu 激活函数激活, 然后得到  $F(\mathbf{x})$ , 残差链接上  $\mathbf{x}$  会跨越两层与  $F(\mathbf{x})$  进行叠加, 最后经过 Relu 激活函数激活, 使得网络能够保留上一层学习到的特征, 转而去学习一个残差。

CRNN 网络的卷积网络层一共包含 7 个卷积层和 4 个最大池化层。这种网络结构的参数量较大, 占用的内存较多, 训练难度也较高, 且 7 个卷积层提取到的特征十分有限, 会直接影响字符识别的精度。增加网络层数可在一定程度上加强对特征的提取能力。但是卷积层的堆叠存在瓶颈, 达到一定的层数后, 特征提取能力不升反降。针对上述问题, 本研究对 CRNN 的卷积网络层进行改进, 引入残差块, 使用更深层的 ResNet 进行特征提取, 从而提出了 CRNN\_ResNet 车牌字符识别算法。由表 1 可见, 引入 ResNet 的卷积网络共有 20 层, 每个 Res\_block 中包含两个卷积进行权重处理, 其中 Res\_block2\_1、Res\_block3\_1 和 Res\_block4\_1 的输入通道数均与输出通道数不一致, 而残差的叠加操作要求输入的两个特征图的通道数相同。所以在这三层结构的残差链接上, 使用一个  $1 \times 1$  的卷积将输入特征图的通道数分别提升到 128、256 和 512, 而卷积步长与主路上的第一个卷积操作相同就能保证特征图尺寸大小一致。

表 1 CRNN\_ResNet 网络结构

Tab.1 CRNN\_ResNet network structure

| 详细结构            | Kernel   | Cin  | Cout                                       | Stride                                 |
|-----------------|--|--|--|--|
| Conv2d( ReLu)   | 3  | 3  | 64   | 1                                      |
| Res_block1      | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 64 \\ 64 \end{Bmatrix}$   | $\begin{Bmatrix} 64 \\ 64 \end{Bmatrix}$   | $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ |
| Res_block2_1    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 64 \\ 128 \end{Bmatrix}$  | $\begin{Bmatrix} 128 \\ 128 \end{Bmatrix}$ | $\begin{Bmatrix} 2 \\ 1 \end{Bmatrix}$ |
| Res_block3_1    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 128 \\ 256 \end{Bmatrix}$ | $\begin{Bmatrix} 256 \\ 256 \end{Bmatrix}$ | $\begin{Bmatrix} 2 \\ 1 \end{Bmatrix}$ |
| Res_block3_2    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 256 \\ 256 \end{Bmatrix}$ | $\begin{Bmatrix} 256 \\ 256 \end{Bmatrix}$ | $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ |
| Res_block3_3    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 256 \\ 256 \end{Bmatrix}$ | $\begin{Bmatrix} 256 \\ 256 \end{Bmatrix}$ | $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ |
| Res_block4_1    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 256 \\ 512 \end{Bmatrix}$ | $\begin{Bmatrix} 512 \\ 512 \end{Bmatrix}$ | $\begin{Bmatrix} 2 \\ 1 \end{Bmatrix}$ |
| Res_block4_2    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 512 \\ 512 \end{Bmatrix}$ | $\begin{Bmatrix} 512 \\ 512 \end{Bmatrix}$ | $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ |
| Res_block4_2    | $\begin{Bmatrix} 3 \times 3 \\ 3 \times 3 \end{Bmatrix}$ | $\begin{Bmatrix} 512 \\ 512 \end{Bmatrix}$ | $\begin{Bmatrix} 512 \\ 512 \end{Bmatrix}$ | $\begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$ |
| AvgPool2d       |  |  |  |  |
| Conv2d( ReLu)   | $2 \times 2$   | 512  | 512  | 2                                      |
| BatchNorm2d     |  |  |  |  |
| Map-to-Sequence |  |  |  |  |
| 双向 LSTM 512→256 |  |  |  |  |
| 双向 LSTM 256→256 |  |  |  |  |

## 2 训练与实验结果分析

使用中国城市车牌数据集 (Chinese City Parking Dataset, CCPD) 制作用于车牌字符识别训练的数据集。根据 CCPD 数据集中每张车牌的位置信息, 利用 OpenCV 工具截取车牌图像, 并提取车牌号信息制作标签, 共制作了 10 万张训练集和 1 万张测试集。

在 Pytorch 框架上分别构建 CRNN 和 CRNN\_ResNet 网络模型, 并构建由省份简称、字母及数字组成的字典。通过已制作的 10 万张数据集分别训练 CRNN 和 CRNN\_ResNet 网络模型, 训练 30 epoch。图 2 描述了两种网络的训练损失。从图 2 可以看出, CRNN\_ResNet 在训练过程中, 损失收敛得更快, 说明该模型训练难度更低。由表 2 可见, 改进后的网络模型虽然有所增大, 但 CRNN\_ResNet 最终的训练损失值比 CRNN 低了约 0.06, 精度更是由原来的 94.8% 提升到 98.5%。在显卡型号为 NVIDIA GeForce RTX 3080, CPU 型号为 Inter (R) Xeon (R) CPU E5-2670 v3 的硬件环境下, 测试两种网络的车牌识别速度。测试结果发现 CRNN\_ResNet 网络每张车牌的处理时间比 CRNN 网络仅增加了 0.005 s, 并不会对实时性造成太大影响。

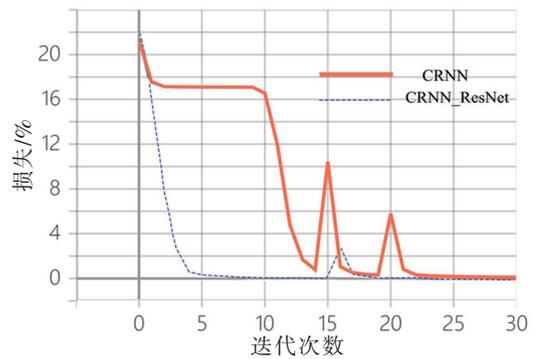


图 2 训练损失对比  
Fig.2 Training loss comparison

表 2 CRNN 与 CRNN\_ResNet 网络模型训练结果

Tab.2 CRNN and CRNN\_ResNet network model data

| 网络模型        | 损失     | 精度    | 模型大小/MB | 识别速度/(张 · s <sup>-1</sup> ) |
|-------------|--------|-------|---------|-----------------------------|
| CRNN        | 0.1045 | 0.948 | 31.8    | 0.007                       |
| CRNN_ResNet | 0.0428 | 0.985 | 79.5    | 0.012                       |

以自然场景下的车辆图片测试本算法的识别效果。使用 CRNN\_ResNet 对车牌框选区域进行识别, 并利用 PIL 工具将车牌识别结果打印在图像左上角。识别结果如图 3 和图 4 所示。



图 3 车牌识别结果 1  
Fig.3 License plate recognition result 1



图 4 车牌识别结果 2  
Fig.4 License plate recognition result 2

### 3 结论

本文围绕车牌模糊和光照复杂场景下的车牌识别问题对 CRNN 算法展开研究。针对 CRNN 卷积层特征提取能力弱的问题, 将带残差的 ResNet 引入 CRNN 以改进特征提取层, 提出了 CRNN\_ResNet 车牌文本识别算法, 并在 CCPD 公开数据集上进行测试。测试结果表明精度由原来的 94.8% 提高到了 98.5%。本文所提出的车牌识别方法要求输入完整的车牌图像, 后续研究可以结合车牌定位方法, 从而实现自然场景下的车牌识别。

## [ 参考文献 ]

- [1] 王飞,李在铭.一种新的车牌实时识别系统[J].电子测量与仪器学报,2004,18(1):12-16,11.
- [2] 丁进超,张伟伟,吴训成.基于双向长短期记忆网络的车牌识别算法[J].电子测量与仪器学报,2018,32(6):173-179.
- [3] TA T D, LE D A, LA M T, et al. Automatic number plate recognition on electronic toll collection systems for Vietnamese condition[C]//Proceedings of the 2015 International Conference on Ubiquitous Information Management and Communication. New York: ACM, 2015:1-5. DOI:10.1145/2701126.2701187.
- [4] 马腾飞,郑永果,赵卫东.基于边缘检测与Hough变换的车牌字符分割算法[J].系统仿真学报,2006,18(增刊1):391-392.
- [5] WEN Y, LU Y, YAN J Q, et al. An algorithm for license plate recognition applied to intelligent transportation system[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(3):830-845.
- [6] 陈寅鹏,丁晓青.复杂车辆图像中的车牌定位与字符分割方法[J].红外与激光工程,2004,33(1):29-33.
- [7] INGOLE S K, GUNDRE S B. Characters feature based Indian vehicle license plate detection and recognition[C]//2017 International Conference on Intelligent Computing and Control. Coimbatore: IEEE, 2017:1-5. DOI:10.1109/I2C2.2017.8321953.
- [8] IALAM R, ISLAM M R, TALUKDER K H. An efficient method for extraction and recognition of bangla characters from vehicle license plates[J]. Multimedia Tools and Applications, 2020, 79(1):20107-20132.
- [9] XING J J, LI J, XIE Z F, et al. Research and implementation of an improved radon transform for license plate recognition [C]//2016 International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). Hangzhou: IEEE, 2016:42-45. DOI:10.1109/IHMSC.2016.52.
- [10] KHAN M A, SHARIF M, JAVED M Y, et al. License number plate recognition system using entropy-based features selection approach with SVM [J]. IET Image Processing, 2018, 12(2):200-209.
- [11] 徐小力,蒋章雷,吴国新,等.基于拓扑特征和投影法的东巴象形文识别方法研究[J].电子测量与仪器学报,2017,31(1):150-154.
- [12] 芮润华,苏婷婷,马晓伟. BP神经网络联合模板匹配的车牌识别系统[J].清华大学学报(自然科学版),2013,53(9):1221-1226.
- [13] 曹泽卫,欧阳,林冬婷,等.基于CNNs识别反馈的点阵字符检测方法[J].电子测量与仪器学报,2020,34(8):159-166.
- [14] 王栋.人工智能OCR技术的应用研究[J].电子技术与软件工程,2022(1):122-125.
- [15] 王珂,杨芳,姜杉.光学字符识别综述[J].计算机应用研究,2020,37(增刊2):22-24.
- [16] 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.
- [17] WOJINA Z, GORBAN A N, LEE D S, et al. Attention-based extraction of structured information from street view imagery [C]//14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto: IEEE, 2017:844-850. DOI:10.1109/ICDAR.2017.143.
- [18] SHI B, YANG M, WANG X, et al. Aster: an attentional scene text recognizer with flexible rectification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(9):2035-2048.
- [19] LUO C, JIN L, SUN Z. A multi-object rectified attention network for scene text recognition[J]. arXiv preprint arXiv, 2019:1901.03003.
- [20] SHI B, WANG X, LYU P, et al. Robust scene text recognition with automatic rectification[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016:4168-4176. DOI:10.1109/CVPR.2016.452.

(责任编辑 朱雪莲 英文审校 黄振坤)