Vehicle License Plate Recognition In Complex Scenes

Zhuang Liu, Yuanping Zhu*
College of Computer and Information Engineering
Tianjin Normal University
China, Tianjin
e-mail: 18437961053@163.com, zhuyuanping@tjnu.edu.cn

Abstract—This paper studies the license plate recognition problem under the complex background and the license plate tilt. Existing methods cannot solve these problems well. This paper proposes an end-to-end rectification network based on deep learning. The model contains three parts: Rectification network, residual module and sequence module, which are responsible for distortion of license plate rectification, image feature extraction and license plate character recognition. In the experiments, we studied the effects of complex backgrounds such as light, rain and snow, and the inclination and distortion of license plates on the accuracy of license plate recognition. The experimental part of this article uses the Chinese Academy of Sciences CCPD dataset, which covers a variety of license plate data in natural scenes. The experimental results show that compared with the existing license plate recognition algorithm, the algorithm in this paper improves significantly the accuracy, and it averages 7.7% in complex scenarios of CCPD dataset.

Keywords-rectification network; convolutional neural network; license plate recognition; smart transportation

I. INTRODUCTION

License plate recognition technology is an important part of intelligent transportation. Traditional license plate recognition technologies are mostly based on character segmentation and SVM [1,2] to identify license plate numbers. They are simple to run and fast to recognize, but they often fail to appear in complex scenarios. With the rapid rise of deep learning in recent years, the license plate recognition technology has also brought huge innovations. Detection technologies such as YOLO [3], SSD [4], and FASTER RCNN [5] have emerged, with advantages such as fast speed and high accuracy; in the field of license plate recognition, single-word recognition based on convolutional neural networks, sequence recognition model CRNN [6], etc. have appeared. The common pre-processing methods for license plate recognition are the use of license plate segmentation, manual license plate tilt adjustment and license plate frame removal. The common methods for text recognition after cutting are template matching [7], transform feature recognition and projection histogram [8].The methods have been widely used in existing license plate recognition systems, but for some complex natural scenes, such as foggy, rainy and other harsh scenes, these methods have not shown a high recognition rate. Concerning unconventional license plate images such as tilt and distortion, deep learning-based convolutional networks cannot show excellent recognition results. This paper proposes to use the rectification-based license plate recognition technology to greatly improve the accuracy rate on the CCPD [9] dataset released by the Chinese Academy of Sciences. This paper proposes an end-to-end rectification method that not only can automatically correct the license plate automatically through a deep learning network, but also can achieve end-to-end recognition without a complicated pre-processing process, and reach a high level. The main contributions of this paper are as follows: 1. irregular license plate recognition using a rectification network; 2. proposing an end-to-end license plate rectification method without complicated preprocessing. 3. obtaining the highest accuracy rate on the CCPD dataset.

II. RELATED WORK

Previous work on license plate recognition usually first segmented the characters in the license plate, and then used optical character recognition (OCR) technology to identify each segmented character. For example, in [10], the extreme region algorithm (ER) was used to segment characters from roughly detected license plates and optimize the position of license plates, using a restricted Boltzmann machine to recognize characters. In [11], MSER is used for character segmentation. A linear discriminant analysis (LDA) classifier was used to extract and classify the local binary pattern (LBP) function for character recognition. However, character segmentation is a very difficult task in itself and is susceptible to uneven lighting, shadows, and noise in the image. It has a direct impact on plate identification. If the segmentation is not correct, even if we have a powerful recognizer, we will not be able to identify the plate correctly. With the development of deep neural networks, method for directly identifying the entire license plate without segmentation is proposed. In [12], segmentation and optical character recognition are jointly performed using Hidden Markov Model (HMM), where the most probable mark sequence is determined by the Viterbi algorithm. In [13], license plate recognition is considered as a sequence marking problem. A convolutional neural network (CNN) is used to extract a series of feature vectors from the license plate bounding box in a sliding window. A recurrent neural network (RNN) with connectionist temporal classification

978-1-7281-9409-7/20/$31.00 ©2020 IEEE

(CTC) [14] is used to label continuous data without character separation.

III. LICENSE PLATE RECOGNITION NETWORK

In this section, the license plate recognition network design proposed in this paper will be described in detail. In recent studies, some powerful classification networks such as VGG [15], ResNet [16] or GoogLeNet [17] are often used as backbones to complete their tasks through feature extraction. However, this is not the best choice for building fast and lightweight networks, so we redesigned the backbone network. The method in this paper consists of three parts. The first part is a rectification network, which can be run independently of the other parts. The second part is a residual module for the extraction of license plate features. The third part is the sequence recognition part.

A. Rectification Network

The rectification network part of the model is an improvement of the spatial transformation network [18] method. This method first returns the position of the license plate through the convolution layer, and then corrects the obtained license plate area through the TPS transformation method. Horizontal license plate image is used as input information in the next part of the model. The rectification network architecture is shown in Fig. 1.

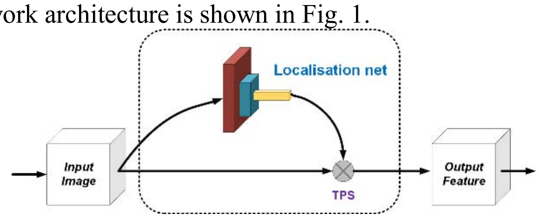

Figure 1. Rectification network

Among them, put license plate image into the rectification network. Location net is a convolution layer of the regression network consisting of five layers of 3 * 3 convolution kernels. In the Linear Layer of the function, the predicted θ value is corrected on the original image by TPS, and finally a corrected Feature Map is output.

In this process, θ is predicted by the positioning network, and then through the following affine transformation formula:

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (1)$$

Among them, x, y represent the coordinate information of the pixel. We get the transformed coordinate information through the mapping formula. And through the thin plate spline (TPS) transformation, a differentiable image sampling method.

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c \max(0,1 - |x_i^s - m|\max(0,1 - |y_i^s - n|)) \quad (2)$$

Among equation (2), U represents the license plate images, V represents the output feature map. After transformation, and finally get the corrected license plate image.

B. Residual Module

Yann Lecun [19] proposed in 1998 a convolutional neural network with weight sharing and local connection. This network can take the image as input directly, and does not need to carry out image feature extraction and data reconstruction like traditional methods, which simplifies the image processing process and has great advantages.

Moreover, its generalization ability is significantly better than other methods, convolutional neural network has been applied to pattern classification, object detection and object recognition, etc. The residual module method used in it is shown in the left figure of Fig. 2. Based on this, the residual module of the model in this paper is improved to be shown in the right figure of Fig. 2.

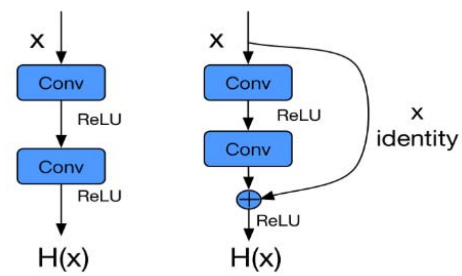

Figure 2. Residual module

The residual blocks used in the encoder of the model in this paper include two convolution layers with ReLU, and a jump connection between the input and output of the second convolution layer.

Where x represents the input to the block, H(x) is the complex function we want to fit.

The objective of ordinary convolution block is to find the appropriate parameters of the convolution layer for direct approximation. The function formula is:

$$H(x) = F(x, W) \quad (3)$$

The residual convolution block has a special jump connection, so as to find the appropriate parameters of the convolution layer to approximate the residual function. The fitted function formula is:

$$H(x) = F(x, W) + x \quad (4)$$

Although both modules should be approximate, the degree of difficulty varies.

In order to avoid the degradation caused by adding more ordinary convolution blocks, this paper USES multiple residual convolution blocks as the basic components in the feature extraction

Figure 3. License plate images from from different positions and angles under a variety of natural weather.

Figure 4. License plate data after rectification. The first row in the picture represents the license plate image, and the second row represents the corrected license plate image

module, and constructs a set of extremely deep feature extractor with 18 convolution layers.

C. *Loss Function*

During the model building process, a loss function suitable for the model is selected. The connectionist time classification (CTC) proposed by Graves et al. [20] is used here. In this paper, the CTC method is used to perform a comparison test with a loss function based on the attention mechanism.

CTC interprets the output probability distribution sequence as a conditional probability over a possible label sequence. Defining L as including all character sets, we get the final label space where {blank} represents a class that does not observe characters. Given a probability distribution, the conditional probability of order π is:

$$p(\pi|y) = \prod_{t=1}^{w} y_{\pi_t}^{t} \qquad (5)$$

Here is the probability of a label appearing at t. In the process of calculating Loss, a training set is given, where and represent the word image and corresponding ground truth labels, respectively.

The objective function is formulated as the sum of the negative log-likelihoods of the probabilities of the target labels:

$$loss = -\sum_{(l_i, I_i)\in D} log\, p\,(l_i|y_i) \qquad (6)$$

Minimizing the objective function is equal to the ability to maximize the probability of generating the target label, which can be solved by dynamic programming. Using CTC, we can process sequences of any length without the need for pre-divided training data. The loss function training set based on the attention mechanism. The conditional probability of minimizing the negative logarithm of the logarithm to D is as follows:

$$loss = -\sum_{i=1}^{N}\sum_{t=1}^{|Y_i|} log\, p\,(Y_{i,t}|I_i;\theta) \qquad (7)$$

IV. EXPERIMENT

The CCPD dataset is a complex scene dataset open source by the Chinese Academy of Sciences and contains 50W license plate image information. Each image contains only one license plate information. Each license plate consists of a Chinese character, a letter, and five letters or numbers. The accuracy of the license plate, that is, the recognition of each character in the license plate is an important indicator of the recognition of the license plate.

The shooting time of the picture is from 7:30 am to 10 pm, and pictures are collected in each time period. The only requirement in the picture collection is to include the license plate. PFC (Picture Collector) shoots from different positions and angles under a variety of natural weather such as rainy, foggy weather and different lights. There will even be a slight vibration to make the collected license plate. The data has different states such as distortion and tilt. The obtained picture is shown in Fig. 3. In Table Ⅰ, CCPD-base represents license plate data with the same characteristics; illuminations on the LP area are dark, uneven or extremely bright in CCPD-DB; the distance from the LP to the shooting location is relatively far or near in CCPD-FN.

During the experiment, the target license plate picture is first transmitted to the rectification network, the original license plate image is corrected, and then the converted horizontal license plate is obtained. Then the license plate is sent to the convolutional neural network to extract features, and finally the extracted features are transferred. Into the sequence layer, the string result is obtained through the processing of the prediction function, and the license plate information expected to be obtained. The model was trained and tested on Ubuntu 16.04, Cuda 9.0, Cudnn 7.5, and GTX 1080. The experimental training of each group took about 6-12 hours. Using the Adadelta optimizer, the learning rate gradually decreased from 1.0 to 0.01. The experiment uses the same training operation and finally completes the experiment. Through experimental comparative analysis in the Fig. 4, we can find that when the model uses the loss

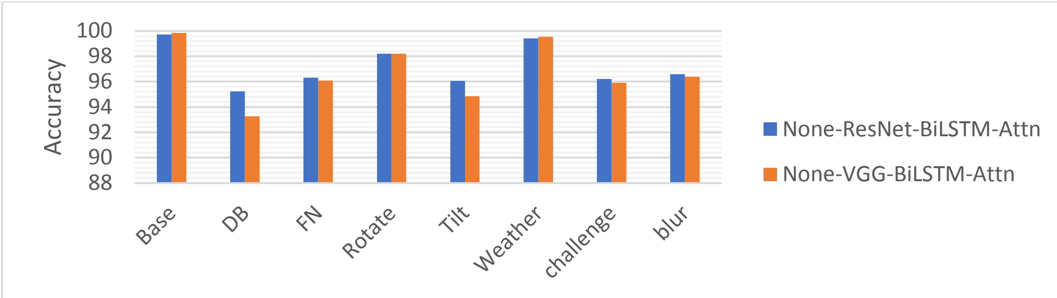

Figure 5: Effect of using different feature extraction layers on results

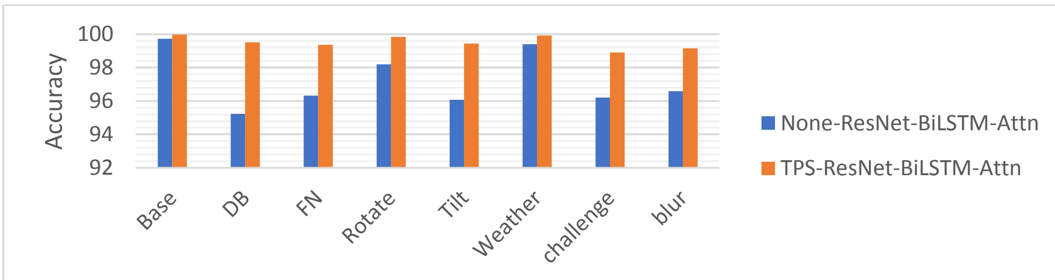

Figure 6: Network impact of using rectification networks

TABLE I. COMPARISON OF LICENSE PLATE RECOGNITION RESULTS IN DIFFERENT SCENARIOS

	Base	DB	FN	Rotate	Tilt	Weather	Challenge	Average w/o Base
TE2E[21]	97.8	94.8	94.3	87.9	92.1	86.8	81.2	90.7
RPnet [9]	95.5	96.9	94.3	90.8	92.5	87.9	85.1	91.8
Ours	99.9	99.5	99.3	99.8	99.4	99.9	98.9	99.5

function of the attention mechanism, its accuracy in each scene is better than that of CTC.

In Fig. 5, VGG and Residual Module are used for comparison. In Fig. 6, whether to use the correction network is compared. For complex license plate recognition scenarios, we use single-channel images for training to reduce the impact of noise on deep networks. During training, random [-10, 10] rotation augmentation is performed, and no processing is performed during prediction. In the experiment, all input license plate images are uniformly converted into gray images with a size of [32, 100], as shown in the first column of Fig. 4. After inputting the processed CCPD image information to the rectification network, the grayscale image shown in the second column of Fig. 4 was obtained, and its size was still [32, 100]. By comparing up and down, it can be clearly found that the corrected license plate is more prominent license plate area, and most of the license plates are horizontal.

Impact of different feature extraction layers on experimental results: In recent image recognition models, the network that uses ResNet as the feature extraction layer is better than the network of VGG, and this effect is particularly obvious in complex scenes. In this paper, a comparison experiment is performed between a network using ResNet and a network using VGG. The results show that the accuracy of the ResNet network is 2-3% higher than the average of the VGG network. The experimental results are shown in Fig. 5. The comparison of using rectification networks is Fig. 6. This shows that the rectification network is effective.

TE2E [21] and RPnet are end-to-end license plate recognition networks. They include two parts: detection and recognition. These two methods use CCPD data sets and achieve better accuracy. In the evaluation of this paper, TE2E and RPnet are compared with the recognition model proposed in this paper. In this paper, the training set images generated from the license plate area are extracted from the training set, and on the test set produced in the same way, the recognition accuracy of the contrast attribute commonly used in the previous algorithms of the rectification network model is more than 99%, which is higher than both. The results are shown in Table I. In CCPD-Challenge dataset, get the highest improvement, an increase of 13.8%. In the last column of Table I, we calculate the average of other scenarios except CCPD-base, and our results are still the best. In addition, it is has better performance on CCPD-Rotate, CCPD-Weather, and especially CCPD-Challenge in ours method.

V. CONCLUSIONS

In this work, it is effective to correct the license plate information in complex scenes such as rain, snow, fog, obstructions, distorted license plates, and blurred scenes. In the CCPD data set that contains multiple scenarios, the performance is better, and its accuracy rate is higher than the existing recognition methods. In terms of speed, the recognition speed of this article can reach 125FPS, while RPnet only reaches 60FPS.

For future work, we hope to further optimize the recognition speed of the network, apply our research to realistic intelligent transportation and other systems, and combine advanced detection algorithms to achieve a complete end-to-end solution from license plate detection and rectification to recognition.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Tianjin (Grant No. 18JCYBJC85000).

REFERENCES

[1] Wang Wei, Ma Yongqiang and Peng Qiang, Application of SVM multi-class classifier in character recognition of license plate. Computer Engineering and Design, (9): 262-265.

[2] Huang Zhibin and Chen Dingsheng, Application of Support Vector Machine in License Plate Character Recognition. Computer Engineering, (5): p. 200-202.

[3] Redmon J, Farhadi A. YOLO9000: better, faster, stronger.Proceedings of the IEEE conference on computer vision and pattern recognition. 2017:p. 7263-7271.

[4] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. Ssd: Single shot multibox detector.European conference on computer vision. Springer, Cham, 2016:p. 21-37.

[5] Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015. 39(6): p. 1137-1149.

[6] Shi B, 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, 2016, 39(11):p. 2298-2304.

[7] Xun Runhua, Su Tingting and Ma Xiaowei, BP neural network combined template matching license plate recognition system. Journal of Tsinghua University (Natural Science Edition), (9): Page 3-8.

[8] Wang, T. H., Ni, F. C., Li, K. T., & Chen, Y. P. Robust license plate recognition based on dynamic projection warping.IEEE International Conference on Networking, Sensing and Control, 2004. IEEE, 2004, 2:p. 784-788.

[9] Xu, Z., Yang, W., Meng, A., Lu, N., Huang, H., Ying, C., & Huang, L. Towards end-to-end license plate detection and recognition: A large dataset and baseline.Proceedings of the European Conference on Computer Vision (ECCV). 2018:p. 255-271.

[10] Gou, C., Wang, K., Yao, Y., & Li, Z. Vehicle License Plate Recognition Based on Extremal Regions and Restricted Boltzmann Machines. IEEE Transactions on Intelligent Transportation Systems, 2016. 17(4): p. 1096-1107.

[11] Hsu, G.S., J.C. Chen and Y.Z. Chung, Application-Oriented License Plate Recognition. 2013. 62(2): p. 552-561.

[12] Bulan, O., Kozitsky, V., Ramesh, P., & Shreve, M. Segmentation- and Annotation-Free License Plate Recognition With Deep Localization and Failure Identification. IEEE Transactions on Intelligent Transportation Systems, 2017: p. 1-13.

[13] Li, H., Wang, P., You, M., & Shen, C. Reading Car License Plates Using Deep Neural Networks. Image & Vision Computing: p.14-23.

[14] Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., & Schmidhuber, J. A Novel Connectionist System for Unconstrained Handwriting Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009. 31(5): p. 855-868.

[15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv, 2014:p. 1409-1556,.

[16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 :p. 770-778.

[17] Szegedy C, Wei Liu, Yangqing Jia, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Andrew A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:p. 1-9.

[18] Bartz C, Yang H, Meinel C. STN-OCR: A single neural network for text detection and text recognition, arXiv preprint arXiv 2017:1707.08831.

[19] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation. 1(4): p. 541-551.

[20] Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks, Proceedings of the 23rd international conference on Machine learning. 2006:p. 369-376.

[21] Li H, Wang P, Shen C. Toward end-to-end car license plate detection and recognition with deep neural networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(3): 1126-1136.