Research on License Plate Detection and Recognition Based on Deep Learning

Lei Gao
Institute of electro-optical
Nanjing University of Science and Technology
Nanjing, China
2679782712@qq.com

Weibin Zhang
Institute of electro-optical
Nanjing University of Science and Technology
Nanjing, China
13829668@qq.com

***Abstract*—With the advent of the era of intelligent transportation, the timeliness and accuracy of license plate(LP) recognition is very important for vehicle management. Traditional LP recognition algorithms rely on fixed scenes and complex image capture systems, no LP recognition algorithm can be widely used in a variety of scenarios; this paper proposes a LP recognition algorithm based on deep learning. First, you only look once(Yolo) v3-tiny with reducing the layers of Yolo v3 is used to roughly locate the LP in the video or image. Then with the landmark detection to precisely detect the LP, and finally recognition the LP with deep convolutional neural network(CNN) end to end. At the same time, in case of the difficulty of data collection, we propose an automatic LP generation algorithm, and we pre-trained base models first, then added real scene data fine-tuning the model for different scenarios to improve the portability and robustness of our models. Through experiments comparison proves that our method has significant advantages in real scenarios with timeliness and high accuracy.**

LP Recognition, Convolutional Neural Network, Model Fine-tuning. (key words)

I. INTRODUCTION

Automatic LP recognition system plays a very important role in the field of intelligent transportation, such as vehicle monitoring, traffic management, suspicious vehicle tracking, and many other fields. In recent years, with the development of the computer's computing power (thanks to the GPU), and the support of a very large amount of data, the method based on deep learning has achieved good performance in LP detection and recognition.

LP detection is the key technology and crucial steps of automatic LP recognition; It can work well under controlled conditions or in a complex image capture system. In specific applications, under uneven lighting conditions, background image diversity, the disturbance of snow and rain, and LP tilt have caused huge damage to the picture, which has seriously affected the detection accuracy of the LP. Therefore, it is still the research focus of the industry to propose a robust algorithm to solve the above-mentioned problems.

Traditional LP recognition is complicated, It is necessary to segmentation first and then single character recognition. The overall algorithm is robustness and difficult to meet the complex and varied outdoor environment, most papers on LP recognition often verify their methods on extremely limited datasets. Foreign LP don't have Chinese characters, the recognition difficulty is relatively low, and the start time is earlier. Hence, foreign research is more advanced and mature, but it cannot be directly applied to China's LP recognition.

In our LP recognition system, we use the method of deep learning, CNN consists of multiple layers of neurons, which can efficiently learn advanced features from a large number of labeled training data. After choosing the appropriate model, the workload will be transferred from challenging algorithm development to relatively mechanically simple sample labeling training and model parameter adjustment. When encountering new scenes, just select the LP image in the current scene as the training sample, which is an important reason for choosing deep learning methods for LP recognition in this paper.

The main contributions of this work are as follows:

- A new LP detection method is proposed, we use Yolo v3-tiny object detection algorithm to roughly locate the LP with K-means clustering algorithm to generate suitable LP detection anchors, then enlarge the roughly positioned LPs for landmark detection, and precisely locate the LP.
- To solve the problems of uneven distribution of LP in various provinces of China and the large workload of labeling, we created a method to automatically generate labeled LPs, and the appearance is close to the real LPs.
- we pre-train three base models first, and add real scene data fine-tuning the model for different scenarios to improve the portability and robustness of our models.

The structure of this paper is as follows: section 2 introduces related work in the field of LP detection and recognition. section 3 introduces our framework and algorithms , In section 4, we discuss the datasets and the results of the algorithm tested on the standard datasets. The article is summarized in section 5.

II. RELATED WORK

A. LP detection

LP detection locates the LP position in an image or video and generates a potential LP bounding box, existing algorithms can be roughly divided into four categories: based on edge information, color information, texture information, and machine learning. The detection method of based on edge information considers that the brightness change in the LP area is more significant than in other places, Tan et al. [1] use an edge detector and morphological operations to find

978-1-6654-3960-2/21/$31.00 ©2021 IEEE
DOI 10.1109/ICCEAI52939.2021.00081

2021 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI) | 978-1-6654-3960-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCEAI52939.2021.00081

candidate LPs, [2] use the canny operator to detect edges and then with hough transform to find strong vertical and horizontal lines as the LP boundaries, The edge-based method is very fast, but it’s too sensitive to unwanted edges, it can’t be applied to complex images. The color-based method based on the color of the LP is usually different from the color of the surrounding body [3] . Deb et al. [4] apply the HSI color model to detect candidate regions and achieve 89% accuracy on 90 images; The color-based method performs well on tilt or deformed LPs, but they cannot distinguish other objects in the image whose color and size are similar to the LP. Texture-based methods try to detect the LP based on the unconventional pixel intensity distribution in the LP area. In [5], a sliding concentric window(SCW) algorithm was developed to recognize the LP based on the local irregularities of the LP in the image texture; the texture-based method is computationally complex, although with more discriminative features than edges or colors. The machine learning-based method is widely used on object detection. Faster-R-CNN [6] utilizes the regional proposal network to generate high-quality prior boxes for detection, SSD [7] encapsulates all calculations in a network and eliminates the prior box generation and subsequent resampling stage, Yolo [8] and its improved version regard object detection as the bounding box regression problem.

B. LP recognition

LP recognition methods can divide into two categories: (1) first segmentation and then identify the single character, (2) non-segmentation method. The former should position the LP bounding box and shape correction before segmentation, In [12], Maglad use the connected component analysis(CCA) method to segment characters and J. Guo et al.[10] use projection-based methods with vertical and horizontal binary pixel projected images for character segmentation. Single character recognition based on template matching to identify each character by measuring the similarity between characters and templates [11] , deep learning based methods include support vector machines(SVM), artificial neural networks can handle characters with different fonts, lighting or rotation. The latter usually directly extracts characters to avoid segmentation, Y. Wen et al. regard LP recognition as optical character recognition(OCR), and it directly operates on the character level, [15] shows that the deep convolutional neural network performs well in multi-character text recognition. Xu Z et al. used two-way Long Short-Term Memory(LSTM) and Connectionist Temporal Classification(CTC) loss function methods to recognize LP characters end-to-end [13] .

III. MODEL

A. LP detection model

The pipeline of our license plate detection network is shown in Fig. 1. Yolo v3 has achieved good performance in object detection, However, it consists of too many layers and time-consuming, which is not suitable for real-time object detection. We use Yolo v3-tiny for LP rough detection, there are only 24 layers which is less than yolo v3 of 107 layers. The input is a 832 * 832 image, First, Conv*5 and Pool*4 is adopted here to extract low level CNN features, Yolo v3-tiny has two feature maps of different scales, for the 52 * 52 feature maps has much finer spatial details and is used for small license plate detection; for the 26 * 26 feature map has larger receptive fields, which benefits detection of larger license plates. We also set 3 anchors in each grid of the feature map at each scale, and each prediction is a 5-dimensional vector, which contains four box coordinates and one object category. Finally, with bounding box regression and NMS(non-maximum suppression) on the selected candidate boxes to further optimize the prediction results.

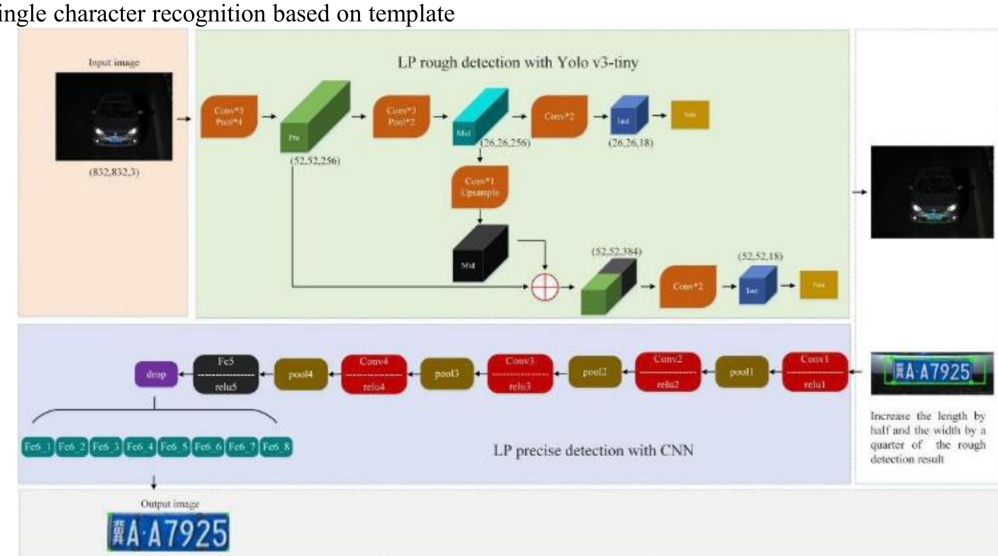

Figure 1. LP detection framework

For LP classifiction can be expressed into a two-class classifification problem, To determine whether input image contains the LP information we use the cross-entropy loss function:

$$L_i^{cl} = -y_i^{cl} \, log\,(p_i) + \left(1 - y_i^{cl}\right)\left(1 - log\,(p_i)\right) \qquad (1)$$

where $y_i^{cls} \in \{0,1\}$ represents the classification label, It's considered a positive sample while the intersection rate of the candidate box predicted by the network and the labeled ground truth (IoU) greater than 0.5, p_i means the probability of having a LP calculated by the network.

The localization loss can be formulated as a regression problem, and the Euclidean distance can be calculated as a loss function:

$$L_i^{box} = \left\| \hat{y}_i^{box} - y_i^{box} \right\|_2^2 \qquad (2)$$

Where $\hat{y}_i^{box}$ represents regression box predicted by the network, y_i^{box} represents the labeled ground truth, which is a four-dimensional vector (x, y, w, h) of left, top coordinates, width and height of the bounding box, the total loss can be formulated as:

$$L = \sum_{i=1}^{N} \left(L_i^{cls} + aL_i^{box}\right) \qquad (3)$$

where N is the number of training samples. We use a=1 in our implementation. Concerning the size of the LP, to improve the convergence speed and the accuracy of LP detection during training, we use the K-means algorithm to select the three aspect ratios, the k-means algorithm uses the IoU as the indicator, the steps of K-means algorithm to obtain anchors are as follows:

- The input of the K-means algorithm is the length and width of the LP:

$$C = \{d_1(x_1, y_1), d2(x_2, y_2,), \ldots, dn(x_n, y_n)\} \qquad (4)$$

  and the aspect ratio of the k candidate boxes.
- Randomly select one sample c_i in C and calculate the distance between this sample and all other samples:

$$\text{distapices} = 1 - \text{IoU}(d_i, c_i) \qquad (5)$$

  find the cluster core closest to the point and assign it to the corresponding cluster.
- After all points belong to clusters, samples are divided into K clusters. Then recalculate the center of gravity (average distance from the center) of each cluster and define it as the new "cluster core".
- Iterate steps 2-3 repeatedly until the samples in the k clusters are all don't change. Through experiments, three different aspect ratios (2, 3, 5) are obtained, and the adopted scale is [26, 52]. Fig. 2 shows the generation process of anchors.

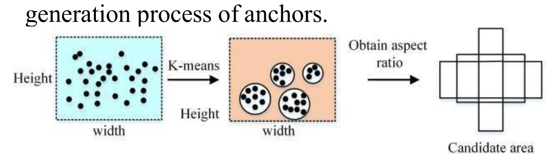

Figure 2. K-means algorithm to generate anchors

The LP box after rough detection by Yolo v3-tiny is not perfect, then increase the length by half and the width by a quarter of the rough detection result based on input image, then the landmark detection by CNN is used for LP precise detection. Our LP precise detection model consist 4 convolution layers, The classic CNN is usually used for image classification; the point regression of the convolutional neural network is to replace the final classification function (SoftMax) layer with a regression function (SumOfSquare):

$$J(\vartheta) = \frac{\sum_i^k (x_i - y_i)}{k} \qquad (6)$$

Where xi represents the i-th dimension of the predicted vector, yi represents the i-th dimension of the labeled vector, and is similar to Softmax, when xi and yi are close, the loss function J will be smaller. In the landmark problem, yi means the coordinate value of the landmark(a landmark has two coordinate values, the abscissa and the ordinate), xi corresponds to the predicted value of the coordinate. And there are four coordinates (upper-right corner, lower-right corner, lower-left corner and upper-left corner of the LP), each of which has two The dimension value (abscissa value and ordinate value), so a total of an 8-dimensional vector is used for training.

B. LP recognition model

After get the LP precise detection result, the LP image may have vertical tilt or horizontal tilt problems because of the shooting angle and lens, and we use the affine transformation to correct based on the four detected corner points of the LP.

With the development of deep learning, non-segmentation character recognition has gradually become a trend, [13] use LSTM for LP recognition and effectively remove the steps of character segmentation, it's assumed that the text characters are composed of several column pixels. However, the characters of the LP are relatively independent and no previous characters to affect the probability of the next character, so it can't take advantage of remembering the previous characters. Deep learning networks have multiple hidden layers, [20]shows that the deep CNN performs well in multi-character text recognition, LP recognition can use the joint-training method in multi-task learning, we recognize seven LP characters simultaneously with 7 labels, the last layer of the network is changed to seven fully connected layers in parallel, and obtained the feature vector of the seven characters respectively, then output the result through softmax, because the characteristics of Chinese characters, letters and numbers in LP characters are unrelated, this process is equivalent to processing the characteristics of each character separately, which helps to improve the accuracy of LP character recognition.

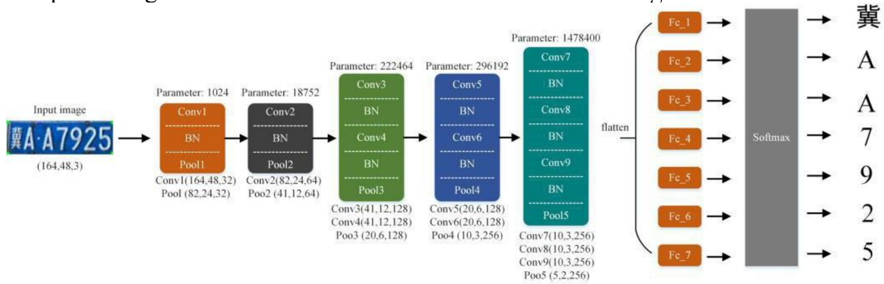

Figure 3. CNN multi-label LP recognition network

Our CNN multi-label LP recognition network with nine convolution layers is shown in Fig.3, we added BatchNormalization layer after each convolution layer to reduce calculation time and prevents overfitting, and considered the slow speed during gradient descent and more iterations of sigmoid and tanh activation function, we use the Relu as the activation function to improve speed and efficiency, the ReLU function is:

$$Relu\,(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (7)$$

Since loss comes from seven softmax, when the gradient of the neural network drops, back propagation results in a gradient value of about seven times that of classic picture classification problems, so in our training process we use SGD with a learning rate of 0.05, a momentum of 0.5, weight decay of 0.001. We need to process seven tasks, the front convolutional layer is shared, and the last layer is divided into seven output, calculate seven loss values separately, and optimize the total loss value.

IV. EXPERIMENTS

In this section, we conduct experiments to verify the effectiveness of the proposed method on both LP detection and recognition performance, our network is implemented using UBUNTU16.04 and Tensorflow with CUDA8.0 and CUDNN5.0, NVIDIA 1080ti GPU with 11GB memory, AMD r7-2700k CPU.

A. Datasets

We use two standard datasets to evaluate the effectiveness of our proposed method, the first dataset is OpenITS, which is a dataset built by Zhongshan University. OpenITS includes 1403 annotated LPs, covering all provinces and municipalities in mainland China, considering the small number of training images, data augmentation is implemented by rotation and affine transformation, and we got total 4000 images for training, 1500 images for test.

The second dataset is CCPD, which contains over 250k LP images with detailed annotations with resolution of 720 × 1160, CCPD also contains dark images, uneven or extremely bright images, tilt images, blurry pictures, images taken on a rainy day, snow day or fog day, and the first characters of CCPD is uneven distribution and mainly LP of AnHui province. We test the follow three datasets of CCPD, each category we have 10000 images for training and 3000 images for test.

TABLE I. DESCRIPTIONS OF TEST DATASETS IN CCPD

Subdataset	Description
Base	The only common feature of these photos is the inclusion of a license plate.
DB	Illuminations on the LP area are dark, uneven or extremely bright.
Tilt	Great horizontal tilt degree (15°~ 45° degrees) and vertical tilt degree (15°~ 45°).

And to train LP character recognition and precise detection model, it is necessary to label a large number of LP datasets, so we designed an algorithm to automatically generate labeled LP data. First of all, formulate a LP template, and then find the relative position of each character, we collected all character fonts as a total character library, randomly select characters from the character library, for example, the first character is randomly selected from the Chinese characters of all provinces, the second character is selected from all uppercase English letters, and the third to the seventh characters are selected from all capital English letters and numeric characters. Then embed the characters in the corresponding position of the template to get the artificial original LP, and select a real LP data from the LP database, and map the pixels of the artificial original LP to the real LP through perspective transformation according to the four marked points. Finally, to increase the complexity of the generated LPs, we added color transformation, salt noise and pepper noise; the final artificial LP renderings are shown in Fig. 4.

Figure 4. generated dataset

we have trained three base models, all models used in experiments are fine-tuned the base models with adding real scene datasets, for the rough detection base model, we used 2500 public data sets from CCPD for training, and use LabelImg(data labeling software) to label the datasets. For the precise detection base model, we generated 50k labeled data sets for training, compared with the generated data mentioned above, this dataset increase the length of the LP template by half and the width by a quater. For the character recognition base model, we generated 100k labeled datasets fortraining.

B. Evaluation Criterion

To evaluate the effectiveness of our proposed method, we use IoU as detection accuracy metric; the bounding box is considered to be correct if its IoU with the ground-truth bounding box is more than 70% (IoU > 0.7). And LP recognition is correct if the IoU is greater than 0.6 and all characters in the LP are correctly recognized.

C. Performance Evaluation on OpenITS

We compared the detection-only performance; end-to-end accuracy and speed of our method with other state-of-the-art methods on OpenITS datase, our fine-tuned models only change the last 1-2 layers because of the small number of OpenITS dataset. As shown in Table 2, Tan et al. use edge detector method for LP detection[3], it achieves high speed but imperfect accuracy, as the above mentioned, we also Yolo v3 method, but it's not as fast or accurate as our method. [20] use Faster R-CNN for LP detection and improved Alexnet for LP recognition(FA), this method has also been tried in this paper, and achieve high detection accuracy (90.8%) and end-to-end accuracy (87.2%), but it’s time-consuming. We also compared our base model(Ours_base) with our fine- tuned model(Ours_ft), and the result shows that the model fine-tuning can obviously increase the performance.

TABLE II. EXPERIMENTAL RESULTS ON OPENITS DATASET

Method	detection-only accuracy(%)	end-to-end accuracy(%)	detection speed(ms)	end-to-end speed(ms)
Tan et al.[3]	78.2	—	60	—
Yolo v3	93.9	—	80	—
FA	90.8	87.2	170	240
Xu et al.[18]	95.4	91.3	95	200
Ours_base	92.3	84.3	60	150
Ours_ft	95.6	91.6	60	150

D. Performance Evaluation on CCPD

In this section, we test the detection-only performance; end-to-end accuracy and speed of our method of different scenes on CCPD. As shown in Table 3, compared with speed, our detection speed and end-to-end speed is faster, our LP detection model achieved the best performance on all four subsets, and averagely 2% higher than the Yolo v3 direct detection and 4% higher than the Faster R-CNN method [6]. RPnet towards end-to-end LP detection and recognition and did effective research, in different datasets, the accuracy of our models is averagely 3% higher than the RPnet. In [13], Xu et al. use LSTM and CTC loss for LP recognition which considered a state-of-the-art recognition model also achieved high accuracy, but compared with speed, ours is faster. Ours the end-to-end performance on DB dataset is worse than Xu et al.[13], that may because the DB dataset’s illuminations on the LP area are dark, uneven or extremely bright, our method haven’t targeted handle it, which is a future work.

TABLE III. EXPERIMENTAL RESULTS ON CCPD DATASET

Method	detection-only accuracy(%)			end-to-end accuracy(%)			detection speed(ms)	end-to-end speed(ms)
	Base	*DB*	*Tilt*	*Base*	*DB*	*Tilt*		
Yolo v3	93.8	92.1	92.8	—	—	—	80	—
FA	91.5	87.5	87.9	85.7	83.2	81.6	170	240
Xu et al.[18]	95.3	92.3	92.9	90.8	89.1	89.8	95	200
RPnet	91.2	87.1	87.3	88.4	86.2	85.9	85	160
Ours_ft	95.6	93.1	93.5	91.5	88.5	90.1	60	150

E. Model analysis

Example results for LP detection and recognition produced by evaluating our models on CCPD are shown in Fig. 5, We set a higher IoU boundary in the detection accuracy metric because a higher boundary can filter out imperfect bounding boxes and thus better evaluates the detection performance. We also considered only use rough detection or direct landmark detection for LP detection, experiments shows our rough detection with Yolo v3-tiny first and then precise detection with landmark detection has a better performence with same settings and other external conditions, and the precise detection also results a higher accuracy of subsequent character recognition.

(a) (b) (c)

Figure 5. Examples results for LP detection and recognition in different datasets of CCPD: (a)Base dataset; (b)DB dataset; (c)Tilt dataset

we carried out some controlled experiments to examine the number of hidden layer of our recognition models affects recognition performance, experiments show that the performance of the model increases with depth of CNN, The process of neural network training can be regarded as a process of parameter adjustment, the more parameters you can adjust, you will get higher accuracy, but to a certain extent, the effect is no longer improved, experiments show that with the number of nine hidden layers achieves the best performance.

V. CONCLUSION

This article introduces a method of LP detection and character recognition based on deep learning. Our network is suitable for images of any size as input to avoid image distortion, we designed an algorithm to automatically generate LPs to avoid the problem of datase t labeling, at the same time, we pre-train base models and add real scene data fine-tuning the model to improve the portability and robustness of our models, experiment shows our method performs well in real scenes. However, our work also has some limitations, we haven't use the end-to-end training method, which will be a direction for our future research.

ACKNOWLEDGMENT

This research is sponsored by National key research and development program (Grant 2018YFB1601101 of 2018YFB1601100), and National Natural Science Foundation of China (Grant 71971116).

REFERENCES

[1] J. Tan, S. Abu-Bakar, M. Mokji, LP localization based on edge-geometrical features using morphological approach, Proc. IEEE Int. Conf.Image Process, 2018. pp. 4549–4553.

[2] S. Rasheed, A. Naeem, O. Ishaq, Automated number plate recognition using hough lines and template matching, Proc. World Cong. Engin. Comp. Sci, 2019.pp. 199-203.

[3] Deb, K. Jo, K.H, Hsi color based vehicle LP detection. In: Control, Automation and Systems, 2008. ICCAS 2008. International Conference on, IEEE (2018) 687-691

[4] Deb, K. Jo, HSI color based vehicle LP detection, Proc. Int. Conf.Cont. Autom. Syst., 2008. pp. 687-691.

[5] W. Jia, H. Zhang, X. He, Region-based LP detection, Jour. Netw.Comp. Appl. 30 (2015) 1324–1333.

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

[7] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2015.

[8] Redmon J, Divvala S, Girshick R, Farhadi A: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2016) 779–788

[9] Al-Shemarry, M S, Li Y, Abdulla, S:Ensemble of adaboost cascades of 3l-lbps classifiers for LPs detection with low quality images. Expert Systems with Applications 92 (2018) 216-235

[10] S. Rasheed, A. Naeem, O. Ishaq, Automated number plate recognition using hough lines and template matching, Proc. World Cong. Engin. Comp. Sci., 2012.pp. 199-203.

[11] S. Du, M. Ibrahim, M. Shehata, W. Badawy, Automatic LP recognition (ALPR): a state- of-the-art review, IEEE Trans. Circuits Syst. Video Technol.23 (2) (2013) 311-325.

[12] Maglad, K.W: A vehicle LP detection and recognition system. Journal of Computer Science 8(3) (2012) 310

[13] Xu Z, Yang W, Meng A, et al. Towards End-to-End LP Detection and Recognition: A Large Dataset and Baseline: 15th European Conference, Munich, Germany, September 8- 14, 2018, Proceedings, Part XIII[J]. 2018.

[14] Redmon J,Farhadi A.YOLOv3:An Incremental Improvement[J]. 2018.

[15] Goodfellow I J, Bulatov Y, Ibarz J, et al. Multi-digit number recognition from street view imagery using deep convolutional neural networks[J]. arXiv preprint arXiv:1312.6082, 2013.

[16] G. Hsu, J. Chen, Y. Chung, Application-oriented LP recognition. IEEE Trans. Veh. Technol. 62 (2) (2013) 552–561.

[17] Yucheng Zhang, Application Research of Deep Learning in LP Recognition [D]. 2018.

[18] Lu Q, Liu Y, Huang J, et al. LP detection and recognition using hierarchical feature layers from CNN[J]. Multimedia Tools & Applications, 2018.