A Key Point-Based License Plate Detection with Pyramid Network Structure

Lingjun Kong
College of Telecommunication & Information Engineering
Nanjing University of Posts & Telecommunications
Nanjing, China
ljkong@njupt.edu.cn

Yunchao Bao
College of Telecommunication & Information Engineering
Nanjing University of Posts & Telecommunications
Nanjing, China
1219012717@njupt.edu.cn

Lijun Cao
Center of Innovation Resource Suzhou Keda Technology Co., Ltd
Suzhou, China
caolijun@kedacom.com

Shengmei Zhao
College of Telecommunication & Information Engineering
Nanjing University of Posts & Telecommunications
Nanjing, China
zhaosm@njupt.edu.cn

2021 2nd Information Communication Technologies Conference (ICTC) | 978-1-6654-0474-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICTC51749.2021.9441496

***Abstract*—In this paper, a key point detection method of license plate based on convolution network is proposed. Traditional license plate detection methods use features like shape, texture and color to locate a license plate with defects such as pertinence, high time complexity, window redundancy and poor robustness. The license plate detection methods based on deep learning have been greatly improved in accuracy and real-time performance, but the detection results of license plates with large rotation angle, small size, less illumination and occlusion are poor. In our method, the rotation angle of the license plate is obtained by detecting four corners of the license plate, and the perspective transformation is used for correction. In order to improve the location accuracy of license plate object, this paper proposes a pyramid network structure to extract high-level and low-level semantic features. Experiments show that the proposed model can not only detect the license plate in general scenes, but also has good detection effect for license plate with large rotation angle.**

Keywords—deep learning, license plate detection, key point detection, pyramid network

I. Introduction

Advances in transportation infrastructure construction technology and an upsurge in the volume of vehicles on the streets have led to increased pressure on vehicle management systems, and license plate detection has therefore become a very important aspect of intelligent transportation. Traditional license plate detection methods based on prior information and deep learning-based license plate detection are the existing mainstream license plate detection methods.

The traditional license plate detection method with edge detection and morphological transformation is used in [1]-[3], but it is difficult to apply to complex environments. The development of deep learning has led to the creation of new methods for license plate detection. Li *et al.* [4], utilized the cascade framework to extract license plates from the detected character regions. Dong *et al.* [5], improved the Faster R-CNN [6] to generate candidate license plates from a light RPN network. The sampler extracted regions of interest from the original high-resolution images and inputed them into the R-CNN network to classify candidate license plates and return four corners of license plate. Silva *et al.* [7], introduced a novel CNN framework, which can detect and correct multi direction license plates in a cascade way. Onim *et al*. [8], presented a YOLOv4 [9] object detection model based on Convolutional Neural Network to detect vehicle license plates in Bangladesh. Xie *et al*. [10] introduced the MDYOLO method based on the YOLO framework to achieve multi-directional license plate detection. An end-to-end deep neural network proposed in [11], which designed two independent branches with different convolutional layers for vehicle detection and license plate detection. Wang *et al*. [12] designed VSNet for automatic license plate detection and recognition, and validated that taking advantage of vertex information is helpful for the CNN model.

The license plate detection method based on object detection obtain the position (the center point coordinates of the license plate) and the size (the length and width of the license plate) of the license plate, without the rotation angle information of the license plate. Recognizing the large-angle rotation license plate detected by this method will cause the phenomenon of overlap between characters, which makes it difficult to apply the common license plate detection method to complex scenes. Inspired by face key detection, this paper attempts to apply key point detection to license plate detection to improve the accuracy of license plate recognition. On the CCPD [13] dataset, the recognition accuracy of our model is 31.10% higher than Faster R-CNN. In particular, the contributions of this paper are as follows:

- A pyramid network structure is introduced, which fully combines high and low-level features to improve detection accuracy.
- We propose a new method of license plate detection, which obtain the rotation angle information by detecting the key point of the license plate, and improve the recognition accuracy of the license plate information.
- Compare the performance of perspective transformation (PT) and rotation transformation (RT) to correct rotating license plates.

The rest of the paper is organized as follows: Section II provides a detailed description of our method. Section III outlines the experiment details and results. Finally, the conclusions and future work are presented in Section IV.

II. METHOD

An RGB image is used as model input, and is represented as a set of feature maps through the backbone network. Then, the pyramid network structure is employed to combine high and low-level semantic features to further extract the feature maps to derive the class, location and key point prediction values of the bounding box. The block diagram of the license plate detection scheme is shown in Fig. 1.

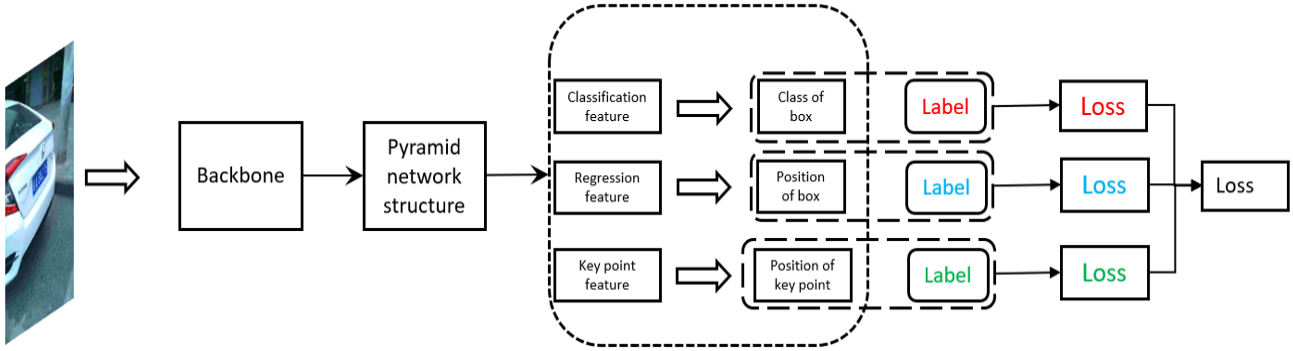

Fig. 1. License plate detection based on key point with pyramid network structure.

A. Pyramid Network

Our goal is to create a pyramid network that integrates high and low-level semantic features based on the backbone network to achieve multi-scale detection of license plates. Our network takes images of any size as input, and outputs proportionally sized feature maps at multiple levels, in a fully convolutional fashion. This process is based on the backbone network, and the MobileNetV2 [14] is used as the backbone network in this paper.

The feedforward calculation process of the backbone network includes many groups of convolutional layers with the same downsampling rate, and we say these layers are in the same network stage. For our featured pyramid, we define three pyramid levels for the bottom three layers, and we choose the output of the last layer of these stages as our reference set of feature maps, which we will enrich to create our pyramid. We call these three feature maps a feature map group, and they have more advanced semantic features from top to bottom. We resize the lowest-level and highest-level feature maps in the feature map set, so that they can be merged with the middle feature maps. The sequence of the fusion process is the fusion of low-level features and then the fusion of high-level features, and each fusion is followed by a layer of convolution and a layer of normalization. The feature map output by the pyramid network structure is used as the input of the model prediction module. Our proposed pyramid network structure is shown in Fig. 2.

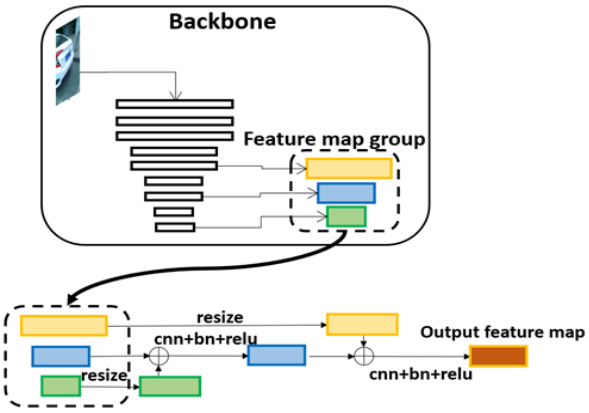

Fig. 2. Pyramid network structure.

B. Anchor Optimization

The concept of "anchor" is proposed by Ren et al. in Fast R-CNN and used for region proposal. The selection of a suitable anchor will make the model easier to converge, and different anchor selection strategies are often used for different object detection tasks. Unlike the hand-picked anchors in Faster R-CNN, Redmon et al. find good anchors by clustering on ground truth (GT) bounding.

Distinct from the method above, we obtain the frequency distribution of the minimum side length of the license plate and the frequency distribution of the aspect ratio of the license plate in the training dataset and the test dataset in the CCPD dataset, which are shown in Fig. 3. We use 16, 64, 128, 256 as the minimum side length of the anchor, and set the aspect ratio of the anchor to 1:2. The final optimized anchors size are (16,32), (64,128), (128,256), (256,512) .

C. Loss Function

In object detection, the total loss function of the bounding box is composed of two parts, one is the box class loss, the other is the box regression loss. In the key point detection framework proposed in this paper, in addition to the classification task and regression task, there is also a key point regression task.

1) Class loss: The classifier of the model judges whether the bounding box is the object (license plate) or the background, and when the intersection over union (IoU) value of the bounding box and the GT is greater than 0.7, the bounding box is considered to be the object, less than 0.3 is considered to be the background, and other situations are ignored. The class loss can be computed as:

$$L_{cls}(p_i,\hat{p}_i) = -\sum_i \left[\hat{p}_i \log p_i + (1-\hat{p}_i)\log(1-p_i)\right] \quad (1)$$

where $\hat{p}_i$ represents the label of the sample i, the positive class is 1, and the negative class is 0. p_i represents the probability that sample i is predicted to be positive.

2) Bounding box regression loss: The bounding box regressor obtains the offset of the bounding box relative to the

anchor, and uses the label encoder to convert the output of RPN layer and the original label into the relative offset of the anchor. In the process of training, we only take the bounding box whose IoU value of the GT is greater than 0.7 as the training sample.

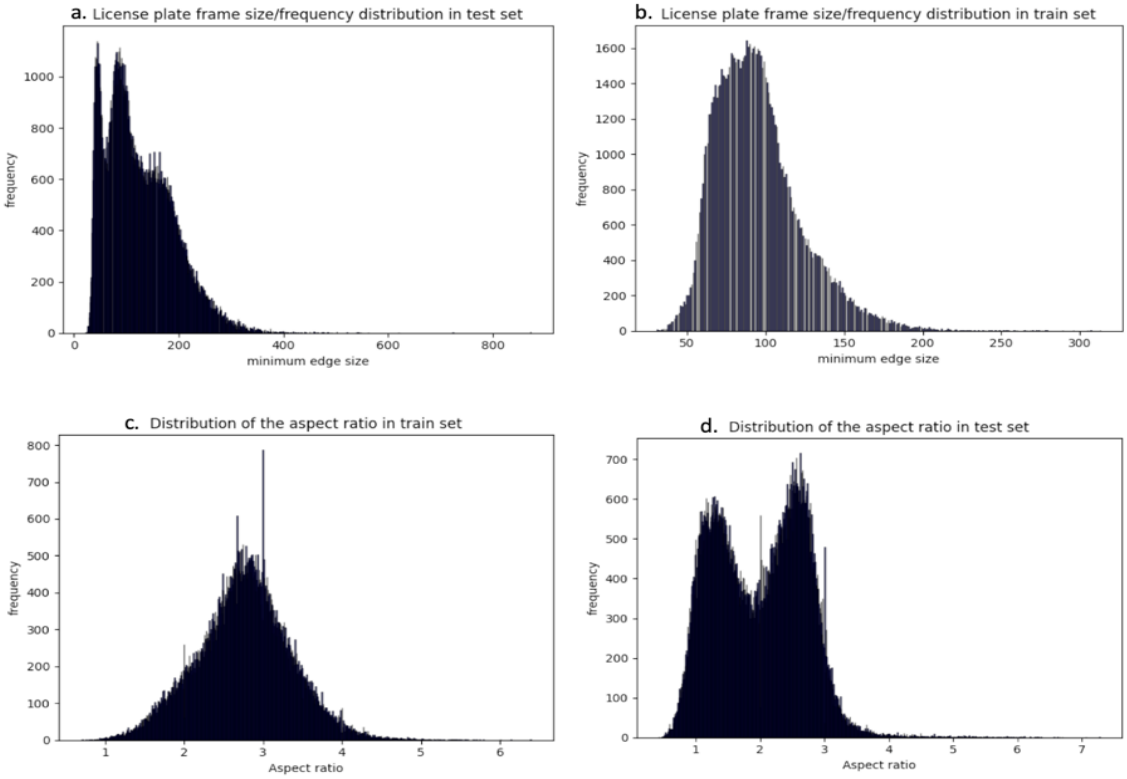

Fig. 3. The frequency distribution of the minimum side length and aspect ratio of the license plate.

We use the mean square error as the loss function of box regression which can be computed as:

$$L_{box}(t_i,\hat{t}_i)=\sum_i (t_i-\hat{t}_i)^2 \tag{2}$$

where t_i is composed of X-axis offset t_{ix}, Y-axis offset t_{iy}, width offset t_{iw} and height offset t_{ih} of the bounding box. $\hat{t}_i$ represents the GT of sample i. We take the sum of the mean square error losses of the four offsets as the final bounding box regression loss.

*3) Key point regression loss:*Key point regression uses the same loss function as bounding box regression, and the loss function can be computed as:

$$L_{pts}(l_i,\hat{l}_i)=\sum_i (l_i-\hat{l}_i)^2 \tag{3}$$

where l_i is composed of four corner points $l_{i1},l_{i2},l_{i3},l_{i4}$ of the license plate, and $\hat{l}_i$ represents the GT of sample *i.* Each corner point is represented by a set of two-dimensional coordinates as (x,y), and its loss is represented as the sum of the losses in each axis direction. We take the sum of the mean square error losses of the four corner points as the final key point regression loss.

The total loss function is shown as follow:

$$L=L_{cls}(p_i,\hat{p}_i)+\lambda_1\hat{p}_iL_{box}(t_i,\hat{t}_i)+\lambda_2\hat{p}_iL_{pts}(l_i,\hat{l}_i) \tag{4}$$

where L_{cls}, L_{box}, L_{pts} represent the class loss, bounding box regression loss, and key point regression loss of the license plate, respectively. When the GT of the bounding box is the object, the model produces bounding box regression loss and license plate key point regression loss. λ_1 and λ_2 represent the training loss weights of the two types of losses, which are set to 0.1 and 0.5 in this paper.

D. Correction Method

This paper uses perspective transformation and rotation transformation to correct the image, and compare their effects through experiments.

The process of rotation transformation is shown in Fig. 4. After the license plate is positioned to the four-point coordinates, (x_1,y_1), (x_2,y_2), (x_3,y_3), (x_4,y_4) are obtained, which respectively represent the upper left, upper right, lower left, and lower right coordinates of the license plate. Then find the coordinates of the midpoint on the left and the midpoint on the right as:

$$(x_{left},y_{left})=\frac{(x_1,y_1)+(x_3,y_3)}{2} \tag{5}$$

$$(x_{right},y_{right})=\frac{(x_2,y_2)+(x_4,y_4)}{2} \tag{6}$$

The slope of the line passing through the two midpoints can be expressed as:

$$k=\frac{y_{right}-y_{left}}{x_{right}-x_{left}} \tag{7}$$

Then the rotation angle of the license plate is:

$$\theta=(\tanh k)*180/\pi \tag{8}$$

Rotate the license plate by angle "θ" to obtain the corrected license plate.

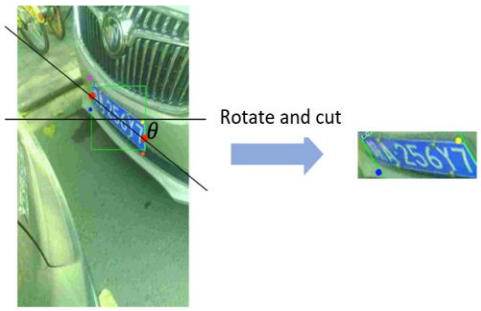

Fig. 4. Rotation correction of test results.

Perspective transformation maps the original image to the transformed image through the projection matrix. According to Fig. 5 which is shown the process of perspective transformation, we define the projection matrix as:

$$P=\begin{pmatrix} \frac{2n}{r-l} & 0 & \frac{r+l}{r-l} & 0 \\ 0 & \frac{2n}{t-b} & \frac{t+b}{t-b} & 0 \\ 0 & 0 & -\frac{f+n}{f-n} & -\frac{2nf}{f-n} \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{9}$$

where A and B represent the image before and after the perspective transformation, respectively. n and f represent the distances between the origin and the A and the B, respectively. t and b represent the ordinates of the top and bottom edges of the A, respectively. l and f represent the horizontal ordinates of the left and right edges of the A, respectively.

III. Experiments

The dataset used in this paper is the CCPD dataset released by the University of Science and Technology of China, in which about 190,000 images are used for training, and about 140,000 images are used for testing. We conduct two evaluation experiments to verify the pyramid network structure and key point network proposed in this paper.

We test the accuracy of the pyramid network structure proposed in this paper on the license plate detection under the condition that the key point-based license plate detection method has not been proposed. The results of the comparative experiment are shown in Table I. We use the classic Faster R-CNN for license plate detection and the average precision (AP) value is only 0.7023. After replacing its backbone network with MobileNetV2 to form Mobile-Faster R-CNN, the AP value drops to 0.6978. The AP value of our model using Mobile-Faster R-CNN and this paper proposed pyramid network structure reaches 0.7572, which is 7.82% and 8.51% higher than the Faster R-CNN and Mobile-Faster R-CNN respectively.

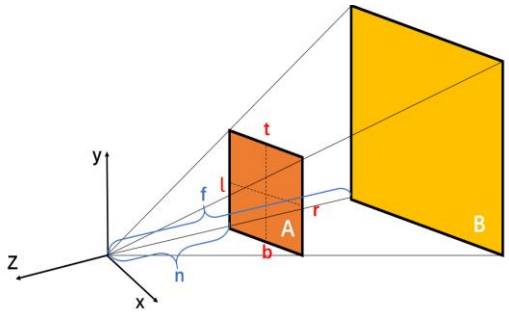

Fig. 5. The process of image perspective transformation.

The key point based license plate detection model proposed in this paper can detect the position information of the license plate and perform rotation correction, our significant experiment is to verify the effectiveness of this method. This experiment uses the latest open source text recognition application programming interface of Baidu to identify the results of the traditional license plate detection method and the proposed scheme, and compares the results of perspective transformation and rotation transformation. Our key point based model-RT and our model-PT employ the pyramid network for license plate detection. And they use rotation transformation and perspective transformation for rotation correction, respectively. Table II shows the results of the comparative simulations, and it can be seen that the recognition accuracy of Faster R-CNN, our model-RT, and our model-PT are 0.6382, 0.7584 and 0.8367 respectively. And the recognition accuracy of our model-RT and model-PT is 18.83% and 31.10% higher than Faster R-CNN, respectively. At the same time, the recognition accuracy of model-PT is 10.32% higher than our model-RT, because perspective transformation can better restore the license plate according to the ratio of the length and width of the license plate.

IV. Conclusion

In this paper, we have presented a key point-based method for license plate detection, which includes a pyramid network structure. The experiments have verified that our method (perspective transformation) consistently outperforms the Faster R-CNN on CCPD dataset, in terms of recognition accuracy. In particular, even with the rotation transformation, our method can achieve competitive performance. Improving the robustness of the model and reducing its complexity are our future directions.

TABLE I. Comparative Experiment Results of Faster RCNN and MobileNetV2 with Pyramid Network Structure on License Plate Detection

Model	AP
Faster R-CNN	0.7023
Mobile-Faster R-CNN	0.6978
Our model	**0.7572**

TABLE II. Comparative Experiment Results of Traditional License Plate Detection Method and Table Detection Method Based on Key Point, and Compares the Results of Perspective Transformation and Rotation Transformation

Model	Recognition accuracy
Faster R-CNN	0.6382
Our model-RT	0.7584
Our model-PT	**0.8367**

ACKNOWLEDGMENT

This work is supported by China Postdoctoral Science Foundation Funded Project (Grant No. 2020M671595), Postdoctoral Science Foundation of Jiangsu Province (Grant No. 2020Z198), NSFC (Grant No. 61871234).

REFERENCES

[1] H. Lin, J. Zhao, S. Li and G. Qiu, "License plate location method based on edge detection and mathematical morphology," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2020, pp. 853-857.

[2] K. P. P. Aung, K. H. Nwe and A. Yoshitaka, "Automatic License Plate Detection System for Myanmar Vehicle License Plates," 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 2019, pp. 132-136.

[3] H. Li, P. Wang, M. You, and C. Shen, "Reading car license plates using deep neural networks," Image Vis. Comput., vol. 72, pp. 14–23, Apr. 2018.

[4] M. Dong, D. He, C. Luo, D. Liu, and W. Zeng, "A CNN-based approach for automatic license plate recognition in the wild," in Proc. Brit. Mach. Vis. Conf., 2017, pp. 1–12.

[5] Ren, Shaoqing , et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis & Machine Intelligence 39.6(2017):1137-1149.

[6] S. M. Silva and C. R. Jung, "License plate detection and recognition in unconstrained scenarios," in Proc. Eur. Conf. Comput. Vis., Sep. 2018, pp. 593–609.

[7] Redmon, Joseph , and A. Farhadi . "YOLO9000: Better, Faster, Stronger." (2017):6517-6525.

[8] Onim, Md, et al. "Traffic Surveillance using Vehicle License Plate Detection and Recognition in Bangladesh." arXiv preprint arXiv:2012.02218 (2020).

[9] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).

[10] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," IEEE Trans. Intell. Transp. Syst., vol. 19, no. 2, pp. 507–517, Feb. 2018.

[11] S. -L. Chen, C. Yang, J. -W. Ma, F. Chen and X. -C. Yin, "Simultaneous End-to-End Vehicle and License Plate Detection With Multi-Branch Attention Neural Network," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 9, pp. 3686-3695, Sept. 2020.

[12] Wang, Yi, et al. " Rethinking and Designing a High-performing Automatic License Plate Recognition Approach. " arXiv preprint arXiv:2011.14936 (2020).

[13] Xu, Zhenbo , et al. "Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline." European Conference on Computer Vision Springer, Cham, vol. 11217,.2018.

[14] Sandler, Mark et al. "Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation." ArXiv abs/1801.04381 (2018): n. pag.