License Plate Detection and Localization in Complex Scenes Based on Deep Learning

Tian Ying [1], Lu Xin [1], Li Wanxiang[2]

1. College of software, University of Science and Technology Liaoning, Anshan, 114051,China
E-mail: t_tianying@126.com

2. Software College, Northeastern University, Shenyang 110819, China
E-mail: asliyou@126.com

Abstract: In order to solve the problem of license plate detection and localization in complex scenes, a new algorithm based on Deep Learning is proposed. Applied the model of convolutional neural network(CNN), Recognized the license plate by converted object detection problem into a binary classification problem. The candidate regions are generated on the sliding window by using selective search algorithm, and the IoU overlap of license plate boundary box and the overlap degree of the candidate regions are used to evaluate the positive and negative samples, at last Support Vector Machine is used for classification. Compared with license plate detection in traditional machine learning, it has obvious advantages in detection accuracy, especially for complex scenes, the positioning accuracy has been improved significantly.

Key Words: Deep Learning; Convolutional Neural Network; License Plate Detection; Selective Search; Support

1 INTRODUCTION

The Automatic recognition of license plate is the basis of effective management in traffic, and it is one of the core technology in modern intelligent traffic management systems. It is widely used in road traffic monitoring. In license plate recognition, the automatic detection and localization of license plate is an important part. License plate detection and localization contain how to extract or segment the license plate region from the license plate image. After many years of development, the license plate detection and localization technology has gained quite a lot of research achievement[1-3]. Under some general conditions, whether accuracy and efficiency of recognition, the existing license plate detection and recognition systems can achieve satisfactory results[4][5]. But in complex scenarios, due to illumination, angle and the interference of external environment, the license plate in the picture or video will be influenced, therefore, the accuracy of license plate automatic detection and localization will be reduced greatly. In view of these difficulties, scholars have made much of attempts with the traditional machine learning method, and the results are not satisfactory.

Deep Learning[6] is a kind of feature learning method, the original data are changed through a simple and nonlinear model to a higher and more abstract expression. Finally, the ideal feature is formed to fit the pattern classification, which can improve classification accuracy. The convolutional neural network (CNN) was proposed by Lecun Yan et al[7] in 1998. It became the first deep learning method to truly successfully train a multilayer network structure, and avoids explicit feature extraction and design style, implicit learning from the training data. Its advantage is reflect extracting the recessive features and expansibility of data essence. CNN has displacement invariance, scaling invariance and other forms of distortion invariance in image target recognition. CNN can be used for almost all target recognition and classification areas[8], such as: face recognition[9], vehicle detection[10], Chinese character recognition[11], image understanding[12] and so on.

Therefore, a new deep learning network structure and training method based on CNN was designed, and designed network structure was used to detect and locate the license plate automatically. The proposed method can be used to automatic synthesize license plate feature for detection, from a large number of the license plate samples and negative samples in training set, and without any manual segmentation. It does not need to preprocess complex images like traditional algorithm, and has better detection robustness for images with different illumination changes, blur and complex scenes.

2 CONSTRUCTING A DEEP LEARNING NETWORK MODEL

2.1 Convolutional Neural Network

The advantage of convolutional neural network is that the display feature extraction can be avoided, and the neurons in each layer of the network can share the weight parameters. CNN mainly includes input layer, convolution layer, excitation layer, pooling layer and fully-connected layer [13].

The convolution layer main function is feature extraction. And the convolution layer is composed of multiple

This work is supported by National Nature Science Foundation under Grant 72472081

convolution kernels, each convolution kernel can be regarded as a filter. The excitation layer is to do the nonlinear mapping of the output result from the convolution layer, and increases the network's ability to express nonlinearity. The pooling layer is to compressed the number of data and parameters, reduce overfitting, each layer in the network contains multiple feature mapping, each feature mapping corresponds to a plane, and all neuron weights shared on the same plane. The convolution layer and the pooling layer are alternating, adding a fully-connected layer in the end .The structure is similar to the usual neural network, all neurons have weighted connections between two layers[14].

The training of convolutional neural network is the same as the traditional BP neural network, and is carried out by error back propagation.

2.2 Constructing License Plate Detection CNN Model

This paper uses convolutional neural network to construct deep learning network. The convolutional neural network structure constructed in this paper consists of 3 convolutional layers, 3 pooling layers and a fully-connected layer. The structure is shown in figure 1.

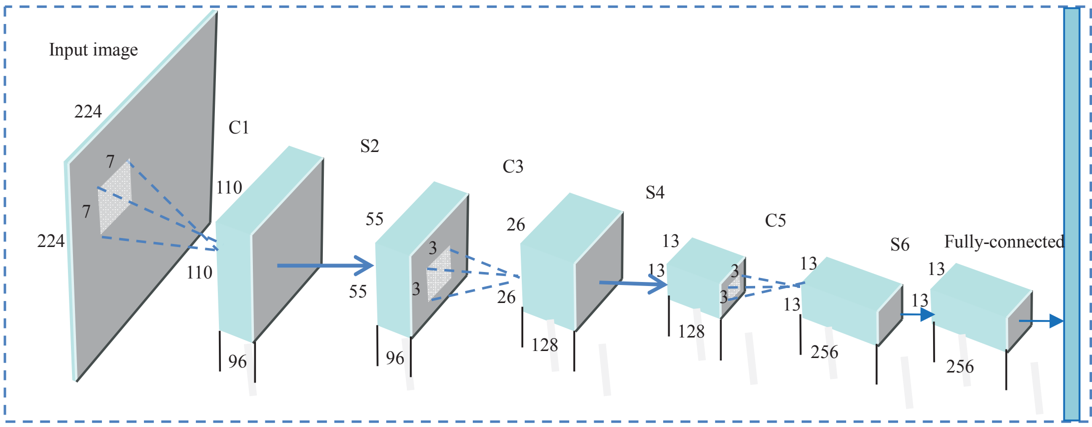

Fig 1. License plate detection and localization depth network structure

The first layer is the image data input. Before training the sample image, the image was simply preprocessed, and all the positive and negative samples of license plate need to be normalized to the size of 224×224.

Layer C1 is the first convolutional layer, the input image through different convolutional kernel operation to extract multiple feature extraction images. The number of convolutional kernels selected in this paper is 96, the size is 7×7, and the stride is 2. After convolution operation, 96 feature maps are obtained, the size of feature map is 110×110.

Layer S2 is the first sub-sampling layer, the convolutional map is first down sampled to obtain the compressed feature map. The size of pooling window is 3×3, and the stride is 2. After pooling, 96 feature maps are obtained, the size of feature map is 55×55.

Layer C3 is the second convolutional layer with 128 feature maps of size 26×26, and the feature map is convolutional again to extract the partial detection image. In this layer, 128 convolutional kernels are selected, the size is 5×5, stride is 2.

Layer S4 is the second sub-sampling layer. It continues to pooling sampled and compressed for feature map. The pooling window size is 3×3. After pooling, 128 feature maps are obtained, the size of feature map is 13×13.

Layer C5 is the third convolutional layer. It continues to do convolution to obtain multiple feature maps. In this layer, 256 convolutional kernels are selected, the size is 3×3, and 256 feature maps of size is 13×13 are obtained.

Layer S6 is the third sub-sampling layer. It is the third time to do down sampling compression. The pooling window is 3×3, and 256 feature maps of size 6×6 are obtained.

The last layer is fully-connected layer. It contains 9126 neurons. The final tag is obtained from the output layer.

In this paper, the expression of the sub-sampling layer is shown in formula (1):

$$X_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \tag{1}$$

In the formula, β_j^l and b_j^l is the weights and biases used for the output feature map of *j*, and the down(.) is the functions used for sub-sampling. Which down(.) function uses the region of continuous n * n of the input image to calculate the maximum value, This method not only increases the generalization ability and keeps the displacement invariance, but also can improve the convergence speed.

In the training phase, all parameters are optimized by back propagation. The activation function for each convolution layer and the fully-connected layer is the ReLU (Rectified Linear Unit) function, defined as:

$$f(x) = \max(0, x) \tag{2}$$

Different from traditional neural network activation function(such as Sigmoid or hyperbolic tangent function), ReLU function has no saturation region, therefore, in the training process, the error can propagate smoothly through multiple layers without fading away. It has the characteristics of simple gradient and so on. Moreover, the network trained by this function has a moderate sparsity, which can avoids the problem of gradient vanishing in the back propagation process, and improves the the convolutional neural network convergence speed greatly .

3 LICENSE PLATE DETECTION ALGORITHM BASED ON CNN

3.1 Model Training

The deep learning network model requires proper training to be effective. A good robustness can be achieved by using the historical license plate data to train the model. Training samples selection is very important, and the main principle that the number is as large as possible and rich, such as different light intensity, rotation angle, fuzzy degree, background and so on. In this way, the training effect will be relatively good, and the missed detection rate will be reduced.

Before training, a series of preprocessing is performed on the positive sample at first, including gray the color images, the image enhancement and the image scale normalized.

The gray processing of color images is realized by the classical weighted average algorithm that shown in formula (3). According to the characteristics of car image, the image is enhanced by the gray-scale conversion enhancement method that shown in formula (4). Where the R_{max} and R_{min} are the gray maximum and minimum values of the histogram in the original image, S_{max} and S_{min} are the gray maximum and minimum values of the histogram in the transformed image, respectively.

$$\mathrm{Gray}(i,j)=0.299\times \mathrm{R}(i,j)+0.587\times \mathrm{G}(i,j)+0.114\times \mathrm{B}(i,j) \tag{3}$$

$$S = \frac{S_{\max} - S_{\min}}{R_{\max} - R_{\min}} f(i,j) + \frac{S_{\max} R_{\min} - S_{\min} S_{\max}}{R_{\max} - R_{\min}} \tag{4}$$

The candidate regions are generated by sliding window. At first, the random windows of different sizes are generated to slide on the input image, taking into account the different in target sizes and edge definition of the sample images, using selective search algorithm to generate the candidate region, which can avoid the huge number and uncertain number of sliding window and avoiding the problem of candidate regions high coincidence degree.

It's impossible to match the candidate region of the image generated by the sliding window and the manually annotated data region exactly. This paper uses the IoU overlap of the license plate bounding box and the candidate region to evaluate.

It is assumed that the bounding box area of the license plate is S_A, and the area of one candidate area is S_B, and the S_I represents the intersection area of two rectangular boxes. Then the coincidence degree of the two rectangular boxes can be expressed as the formula IoU=$S_I/(S_A+S_B-S_I)$.

After preprocessing the license plate image, the selective search algorithm searched 1000 candidate regions. The IoU of all searched candidate regions and the manually labeled detection regions is calculated, and if IoU>0.7 then the region is taken as the positive sample of the deep learning network, as well as the IoU<0.3 region is taken as the negative sample.

The generated positive and negative samples also need to be normalized as a uniform size used as sample input. In this paper, all positive and negative samples are normalized to the size of 224*224.

The training set includes positive samples and negative samples. After the positive sample set is finished, a number of non-license plate regions are segmented as the negative samples set for training. Some of the positive and negative samples are shown in figure 2 and figure 3.

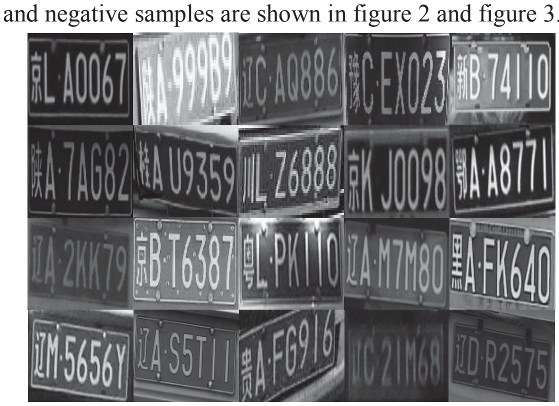

Fig 2. Partial training positive samples

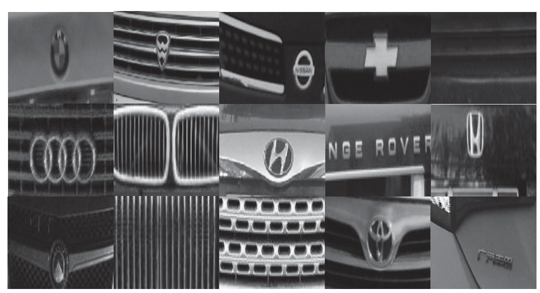

Fig 3. Partial training negative samples

3.2 SVM Classifier

For linearly separable binary classification problems, the goal is to design a classifier that can effectively classify the unobserved instances, i.e., it has good generalization ability. For linearly inseparable problems, we can promote optimal hyperplane concept by introducing the method of slack variables and penalty function, a more general method is to map the input vector to a high dimensional space by introducing a nonlinear mapping which based on the kernel function theory of integral operator space, support vector machines can gives the best classification hyperplane in high dimensional space[15].

More specifically, the given training sets $\{y_i, x_i\}$,

$x_i \in \kappa$, $y_i \in \{-1,+1\}$,i=1,…,l, then constructing the SVM with a kernel function $K\{y_i, x_i\}$ can be attributed to solving the conditional constraint optimization problem as formula(5):

$$\bar{\alpha} = \arg\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i \bullet x_j) - \sum_{i=1}^{l}\alpha_i \qquad (5)$$

Constraint condition is $0 \le \alpha_i \le C$, i=1,…,l, $\sum_{i=1}^{l}\alpha_i y_i = 0$.

In solution of the formula (5), x_i corresponded to all non-zero Lagrange multipliers $\bar{\alpha}_i$ constitutes SVM (Support Vectors Machine), and the classifier constructed by this method is as formula (6):

$$f(x) = sign\left[\sum_{x_i \in SVs}\bar{\alpha}_i y_i K(x_i, x) + \bar{b}\right] \qquad (6)$$

Where: $\bar{b} = -\frac{1}{2}\sum_{x_i \in SVs}\bar{\alpha}_i\, y_i[K(x_r, x_i) + K(x_s, x_i)]$, x_r and x_s are support vectors belonging to different classes.

In order to convenient for the target detection and localization, the output range of formula (6) is usually mapped to the interval of [0,1] during classification, as shown in formula(7):

$$f(x) = \frac{1}{1+e^{-s(x)}} \qquad (7)$$

Where: $s(x) = \sum_{x_i \in SVs}\bar{\alpha}_i y_i K(x_i, x) + \bar{b}$

4 EXPERIMENT RESULTS AND ANALYSIS

In order to verify the algorithm effectiveness, the simulation experiment is carried out, the experimental environment is: CPU, i7-5830K; internal memory is 32G, the capacity of SSD hard disk is 256G, the video card is NVIDIA GTX 1080; Caffe is selected as network training framework.

The data used in the experiment is a sample image library that collected by ourselves which contain license plate images. There are nearly 10000 sample images from different scenes, different weather and different shooting angles. Figure 4 is a partial correct result of the license plate detection by using the algorithm that presented in this paper.

The presented results in figure 4 include static shooting in the parking lot, dynamic shooting on the road, the shooting in strong light , dark light and also in night . The picture includes front license plate, rear license plate, single license plate and multiple license plate, as well as complex environment. Most of the license plates have rotation angles. From figure 4, the probability of the detection area for license plate is higher than 0.95.

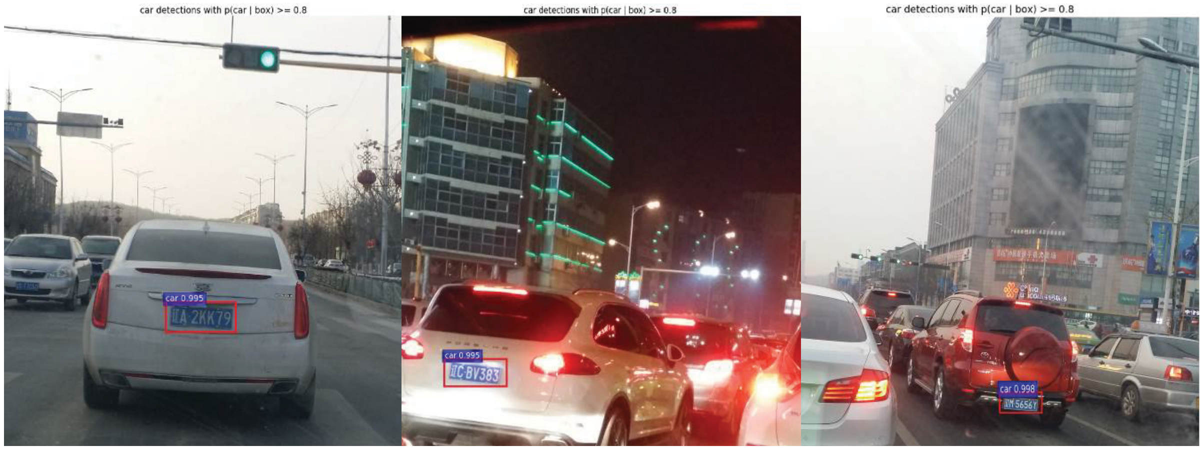

Fig 4. Some detection and localization results of license plate

For all the photos we have taken, given different probability thresholds, the false negatives rate(FNR) and false positives rate(FPR) are shown in table 1.

Table 1 FNR and FPR under different probability thresholds

Thresholds	FNR(%)	FPR(%)
0.6	0	9.36
0.65	0.01	8.27
0.7	0.01	2.35
0.75	0.07	2.02
0.8	0.12	1.16
0.85	0.17	1.02
0.9	1.34	0.73
0.95	3.36	0.01

From table 1, under the condition of given threshold≥0.85, correct detection and localization rate can reached 98.81%. It proves that the algorithm in this paper can detect and locate the correct license plate part. Besides, the accuracy of detection and localization is very high, and also has strong robustness and adaptability.

When the rotation angle is too large, or the license plate is too far away from the camera, most of the license plate feature information is lost, it is difficult to locate by using this small amount of information in detection process, there will be false negatives or false positives, as shown in figure 5.

Figure 5 is part of the inaccurate detection results, the license plate probability of the picture on the left which detected is 0.965, actually the license plate part has contained in the detection box, only non-license plate part in the box is too much, it classified as a false detection range. The right detection result is missed detection. The right picture, from the detection results, due to the lack of feature information, undetected is normal. Even if the license plate recognize correctly, the license plate character recognition can not be carried out later. Because of the character that in the license plate part is not clear at all. Thus, the license plate detection and localization method proposed in this paper has a very high positioning accuracy.

Fig 5. Inaccurate license plate localization

This article also implemented two common localization algorithms: gray level jump feature algorithm and multi-class feature fusion based license plate detection algorithm. The mainly process of gray level jump as follows: At first, convert license plate images into binary images, calculate the number of gray level jump for each row of 0 and 1 values, when frequency hopping is greater than a certain threshold, is marked as candidate row, and then project the initial license plate localization vertically in the candidate row. According to the literature[2], a convolutional neural network algorithm fusing multi-class features is implemented. Then can get the advantages and disadvantages of two algorithms are compared and analyzed in the same data set with the proposed method.

Table 2 is the experimental results which use the license plate localization algorithm comparing with other methods in the same image database.

Table 2 Comparison of different algorithms

Algorithm	FNR(%)	FPR(%)
Grays level jump	7.4	9.36
Multiple feature fusion	1.76	2.05
Algorithm in this paper	0.17	1.02

As shown in Table 2, compared with other algorithms, the license plate detection and localization algorithm based on convolutional neural network has the lowest missed detection rate and false detection rate, which shows that the algorithm has stronger robustness. It is shown that the method of license plate detection and localization based on convolutional neural network has more persuasive than traditional machine learning method indirectly, and it can make the part of the license plate detection more accurate.

5 CONCLUSION

License plate localization is an important part of license plate recognition, which has a great impact on the following character recognition. This paper proposes a deep learning network model based on convolutional neural network, which is a new attempt on license plate detection and localization. The proposed method can detect and locate the license plate without any manual extraction. The method can automatically synthesize a feature extractor with license plate feature from a large number of license plate positive samples and non-license plate negative samples, and can detect and locate the license plate without any manual extraction. Compared with the traditional method of license plate detection, the robustness and accuracy of this method is better. It has strong resistance in illumination change, angle change and complex background. It not only can effectively improve the accuracy of license plate detection and localization, no need for complex image preprocessing like traditional algorithms, but also can avoid tedious and one-sided for artificial selected feature.

REFERENCES

[1] TANG Yufeng, LUO Bin,TANG Jin, WU Fubu. New heuristic algorithm of license plate location based on statistical feature.[J] Computer Engineering and Application, 2015,51(6): 154-158.

[2] Abolghasemi V, Ahmadyfard A. An edge-based color-aided method for license plate detection[J]. Image & Vision Computing, 2009, 27(8):1134-1142.

[3] Comelli P,Ferragina P,Granieri M N. Optical Recognition of Motor Vehicle License Plates[J], Vehicular Technology, IEEE Transactions on, 1995,44: 790-799.

[4] SI Qing-ya. The design and implementation of License Plate Recognition System based on robust illumination[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2013.

[5] Deneen K M V. An Algorithm for License Plate Recognition Applied to Intelligent Transportation System[J]. IEEE Transactions on Intelligent Transportation Systems, 2011, 12(3):830-845.

[6] Y lecun,Y Bengio,G Hinton.Deep learning. Nature 2015, 512(7553): 436-444.

[7] Y Lecun, L Bottou, Y Bengio, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278- 2324.

[8] Zheng Yin,Chen Quanqi,Zhang Yujin. Deep learing and its new progress in object and behavior recognition[J]. JOURNAL OF IMAGE AND GRAPHICS,2014, 2(19): 175-184.

[9] WANG Jimin,LU Jianfeng. Face gender recognition base on convolutionalal heural network[J]. Modern Electronic technique, 2015, 48(7): 81-84.

[10] LI Lin-hui,LIU Zhi-mei,LIAN Jing et al.Convolutionalal neural network-based vehicle detection method[J]. Journal of Jilin University(Engineering and Technology Edition), 2017,47(2): 384- 391.

[11] YAN Xiliang, WANG Liming. Handwritten Chinese character recognition system based on neural network convolutionalal depth[J]. Computer Engineering and Applications, 2016(9).

[12] CHANG Liang,DENG Xiao-Ming, ZHOU Ming-Quan et al.Convolutionalalal Neural Networks in Image Understanding[J]. Acta Automatica Sinica. 2016, 42(9): 1300-1312.

[13] CHEN Xian Chang. Research on algorithm and application of deep learning based on convolutional neural network[D]. Hangzhou, Zhejiang Gongshang University. 2013.

[14] SUN Zhi-yuan, LU Cheng-xiang, SHI Zhong-zhi et al. Research and Advances on Deep Learning[J]. Computer Science, 2016, 43(2):1-7.

[15] LI Jintao. Face Recognition Using Support Vector Machines[J]. COMPUTERSCIENCE.2003,30(4):11-15.