IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 7, 2019, accepted November 20, 2019, date of publication November 27, 2019, date of current version December 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956357

A Light CNN for End-to-End Car License Plates Detection and Recognition

WANWEI WANG, JUN YANG, MIN CHEN, AND PENG WANG
Tianjin Key Laboratory for Advanced Signal Processing, Civil Aviation University of China, Tianjin 3003000, China
Corresponding author: Wanwei Wang (wwwang@cauc.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0502405, and in part by the Fundamental Research Funds for the Central Universities of Civil Aviation University of China under Grant 3122013D020.

ABSTRACT License Plate Recognition (LPR) is of great significance due to its wide range of applications in the Intelligent Transportation System (ITS). It is an important and challenging research topic in image recognition fields. However, many of the current methods are still not robust in real-world complex scenario. The main contribution of this paper is to propose a multi-task convolutional neural network for license plate detection and recognition (MTLPR) with better accuracy and lower computational cost, and introduce a comprehensive data set of Chinese license plate. First, we train a Multi-task Convolutional Neural Networks (MTCNN) to detect license plate. Then we introduce an end-to-end method to recognize license plate information, which further improves the recognition precision. Last, We compare the experimental result with other state-of-the-art methods. The experimental result shows that our method achieves up to 98% recognition precision and is superior to other methods in the precision and speed of detection and recognition.

INDEX TERMS Object detection, optical character recognition, license plate recognition, convolutional neural network.

I. INTRODUCTION

License Plate Recognition (LPR) technology is an important component of modern intelligent transportation system, and it is a challenging and important task which is used in traffic management, digital security surveillance, vehicle recognition, parking management of large cities. The basic process of LPR is as follows: first, process and analyze the vehicle images or videos captured by the camera, then use digital image processing, pattern recognition or other technologies to obtain the license plate number and color information. In real-world scenario, the images are captured over days and nights in a highly complex environment with varying illuminations and different weather conditions, so localization, segmentation and recognition become challenging tasks.

In this paper, we propose a new robust real-time LPR method named multi-task convolutional neural network for license plate detection and recognition (MTLPR), which uses an end-to-end algorithm to recognize plate characters. The main contributions of this paper can be summarized as follows:

1) We propose a real-time end-to-end license plate recognition framework which is based on a cascaded network. This framework has advantages of simple structure, high accuracy and low computational cost.
2) We augment a Chinese license plate data set to make it richer and more comprehensive. It greatly improves the performance of LPR. Then we will open the source code and the data set.

This paper is organized as follows. We briefly review related work in Section II. Section III presents our proposed MTLPR algorithm. We report and discuss the results of our experiments in Section IV. Conclusions and future work are given in Section V.

II. RELATED WORK

A. STATE-OF-THE-ART METHODS

The research of LPR started earlier. More and more practical LRP systems have been applied to a lot of fields such as Electronic Toll Collection (ETC), traffic monitoring, traffic management, and other scenarios. Compared with foreign countries, the development of Chinese LPR research is relatively slow. First of all, the structure of Chinese license plates is quite different from foreign plates. In addition, Chinese

The associate editor coordinating the review of this manuscript and approving it for publication was Sohail Jabbar.

license plate has various character patterns, such as different sizes, fonts and colors. Therefore, foreign LPR technology can not be directly used in relevant scenes to recognize Chinese license plate.

However, the vast majority of the LPR products on the market adopt the technical scheme of detection, segmentation and recognition [1]. This scheme relies on the effect of image segmentation, which is particularly affected by blur, occlusion and adhesion.

There are a number of ways to implement LPR technology, for instance, Support Vector Machine (SVM) classifier [2], feature classifier [3], artificial neural network classifier [4], [5], cluster analysis, and other methods. At present, LPR algorithms is mainly divided into two categories: traditional algorithms and deep learning algorithms. Typical representatives of traditional algorithms include AdaBoost algorithm [6] based on Haar feature, SVM algorithm based on Hog feature, Cascade LBP algorithm and so on. For example, Hsieh *et al.* [7] uses the morphological method to significantly reduce the number of candidate areas, thereby speeding up the process of license plate detection. Yu *et al.* [8] proposes a robust method to locate license plate based on wavelet transform and empirical mode decomposition analysis. Wang *et al.* [9] uses the cascaded AdaBoost classifier and voting mechanism to obtain candidate license plate information. In algorithm [10], a new mode named Local Structure Patterns is introduced to detect the license plate area.

With the development of deep learning, the object detection model based on convolutional neural network (CNN) [11]–[16] has been widely used for LPR. Typical representatives are the R-CNN (Regions with CNN features) [11]–[13], the YOLO (You Only Look Once) [14], [15] and the SSD (Single Shot MultiBox Detector) [16].

R-CNN [11] uses the selective search algorithm to calculate the feature similarity of adjacent regions. By scoring similar regions, candidate with higher score is selected as an input of the convolutional neural network. The corresponding feature vectors are calculated according to the positive and negative features, and then classified by the SVM classifier. Finally, the bounding box regression algorithm is used for the candidate to achieve the purpose of object detection. Fast R-CNN [12] effectively solves the shortcoming that R-CNN algorithm must crop and scale the image area to the same size. The multi-task loss function is proposed to train classifiers and bounding box localizer. There is no need to store the features of the intermediate layer and back propagation routes derivatives through the Region of Interest (RoI) pooling layer. In order to solve the bottleneck of region proposal computation by Fast R-CNN, Faster R-CNN [13] introduces a Region Proposal Network (RPN) that shares full image convolutional features with the detection network and nearly cost-free to generate region proposals. RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. YOLO [14] predicts both bounding boxes and class probabilities for multiple objects of multiple classes in an image. It frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. However, there are problems such as inaccurate location, low recall rate, and poor detection effect of small objects. SSD [16] predicts the object region on the feature maps of different convolutional layers, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales. It not only maintains the fast detection speed of YOLO algorithm, but also guarantees the accurate boundary positioning effect close to the Faster R-CNN algorithm. However, it adopts multi-level feature classification, which makes it difficult to detect small targets.

In recent years, LPR algorithms based on convolutional neural network develop rapidly and have many problems at the same time. Li *et al.* [17] proposed method to jointly solve license plate detection and recognition by a single deep neural network. However, they adopt VGG to extract low level CNN features which has the disadvantages of slow training speed. Xie *et al.* [18] proposed a CNN-based MD-YOLO framework for multi-directional car license plate detection and elegantly managed rotational problems in real-time scenarios. But this method is difficult to detect small object which based on YOLO algorithm. Li *et al.* [19] proposed a cascade framework to improve performance on both detection and recognition, but the CNN classifiers can't support Chinese LPR.

LPR technology has made great progress, but there are still many difficulties [20]–[22], which mainly contains the following aspects: (1) Due to the noise influence of surrounding environmental factors, such as illumination, haze, rain, snow, etc. (2) Due to character distortion caused by capturing viewpoint, such as tilt, blur, as well as stuck, missing or broken characters. (3) Due to the particularity of Chinese license plates, such as the difficulty of recognizing complex Chinese characters over letters and numbers, and the diversity of license plate types and colors.

B. BASIC PROCESS

The traditional LPR systems typically have three stages [23]: license plate detection, character segmentation and character recognition. The earlier stages require higher accuracy, since failure of the license plate detection probably leads to a failure of the next stages.

First of all, we need to pre-process images captured by camera before the license plate recognition to avoid the impact of low-quality images caused by lighting, weather, camera position and other factors. Pre-processing operations include image binarization, edge detection, noise elimination, and image graying. License plate location is to locate the license plate position and obtain four corner coordinates. It is an important step in the license plate recognition process. Unsuccessful or incomplete location will directly lead to the final recognition failure.

Due to shooting angle, lens and other factors, the license plate in the image may have horizontal tilt, vertical tilt, trapezoidal distortion, or other deformation. After locating the license plate, the license plate correction processing is

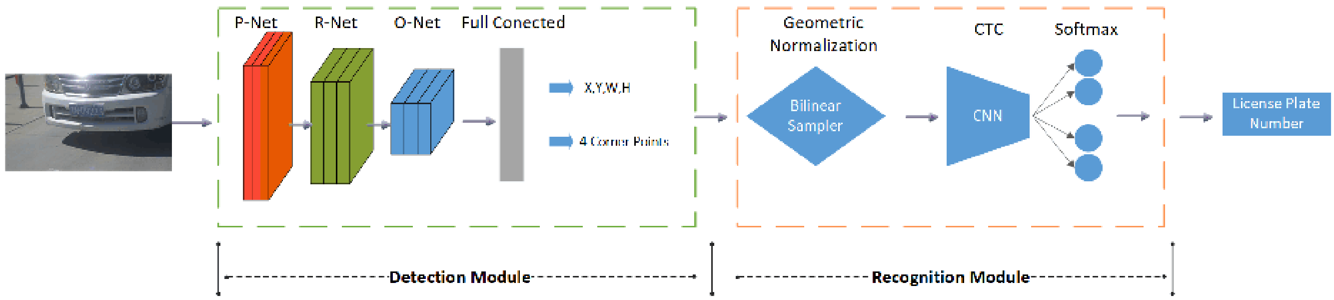


FIGURE 1. The architectures of MTLPR, where the green box means detection module and the red one means recognition module.

carried out, which is beneficial to remove the noise and other factors, and more conducive to character recognition.

Character segmentation is to extract a single character image from license plate, and then process the segmented character images to finally recognize the characters. The advantage of this method is that only a small number of character samples can be used to train the classifier, and better results can be obtained under the good illumination and camera conditions. Most commercial license plate recognition applications adopt this method.

However, in some complicated natural situations, the segmentation and recognition of license plate characters become particularly difficult. Different from traditional methods with slow computing speed and the heavily depending on the segmentation results, the detection and recognition algorithm in this paper is based on an end-to-end and segmentation-free recognition scheme, which has higher speed and accuracy.

III. APPROACH

In this section we describe our MTLPR network architectures in detail. The main architectures of end-to-end MTLPR network is shown in Figure 1.

A. DETECTION STRUCTURE

The deep neural network can perform nonlinear mapping to learn features from shallow to high level, so it can be applied to license plate detection. For example, the recognition result of Faster R-CNN mentioned above, is accurate but time consuming, and SSD and YOLO are relatively fast, but not suitable to recognize small objects. Therefore, Li et al. [24] proposed Cascaded CNN in 2015, which maintains the high precision of R-CNN, and creatively cascades several small networks with relatively low performance to form a strong classifier with excellent performance. Compared with the traditional method, the features used by each level classifier are completely generated by the neural network rather than manually processed, and have excellent fitting ability. Compared with the deep network, the small network greatly reduces the geometric growth.

Therefore, our method MTLPR is based on the improved Cascaded CNN [25] method, named the Multi-task Convolutional Neural Network (MTCNN) [26] algorithm. MTCNN is mainly used for face detection and face key point detection, and after improvements it can also be applied to LPR. The detection network can be divided into three layers of P-Net, R-Net, and O-Net.

The input image is scaled to different sizes according to different scaling ratios to form a feature pyramid, which is the input of the following three-stage cascaded framework. In the first phase, the P-Net network is used to obtain the license plate candidate and the regression vector of the bounding box. Then it merges the highly overlapping candidates by Non-Maximum Suppression (NMS). In the second phase, the candidate acquired in the first stage is trained using the R-Net, then the bounding box regression and NMS is used to fine tune the candidate and remove the overlapping windows. In the third phase, similar to R-Net, the O-Net locates the four corner points of the license plate while removing the overlapping candidate window. NMS [27] is used to extract the highest score window, which is an important part of the object detection process and widely used in many computer vision tasks. The candidates are scored by the object detection algorithm. Then the candidate with the highest score is selected, and other candidates overlapped obviously are suppressed.

The detailed network structure of detection is shown in the following tables, including:

P-Net (Proposal Network), as shown in Table 1. Three convolution layers are used to obtain candidate region classification results and boundary regression vectors. The proposal network has characteristics of small scale and low threshold to maximize positive sample recall. Its main purpose is to select as many license plate information as possible, and assign the more difficult classification tasks to the subsequent network. The main purpose of NMS is to eliminate the redundant license plate candidate and find the best detection position.

R-Net (Refine Network), as shown in Table 2. Compared to the first layer of P-Net, a fully connected layer is added. This network acts as a transitional network to filter out a large number of negative candidates with poor effects. Finally, Bounding Box Regression and NMS are performed to further optimize the prediction of the selected candidates.

O-Net (Output Network), as shown in Table 3. The basic structure is a more complex CNN, which has an additional convolutional layer compared with R-Net. This stage aims

TABLE 1. The architectures of P-Net.

Input	Operator	c	s
12 x 36 x 3	Conv 3 x 7	10	1
10 x 30 x 10	MP 2 x 2	10	2
5 x 15 x 10	Conv 3 x 7	16	1
3 x 9 x 16	Conv 3 x 7	32	1

The architectures of P-Net, where "MP" means max pooling and "Conv" means convolution. All layers in the same sequence have the same number c of output channels. The stride in convolution and pooling is 1 and 2, respectively.

TABLE 2. The architectures of R-Net.

Input	Operator	c	s
24 x 72 x 3	Conv 3 x 7	28	1
22 x 66 x 28	MP 3 x 3	28	2
11 x 33 x 28	Conv 3 x 9	48	1
9 x 25 x 48	MP 3 x 3	48	2
4 x 12 x 48	Conv 2 x 4	64	1
3 x 9 x 64	FC	128	1

The architectures of R-Net, where "MP" means max pooling and "Conv" means convolution, and "FC" means fully connect layer. All layers in the same sequence have the same number c of output channels. The stride in convolution and pooling is 1 and 2, respectively.

TABLE 3. The architectures of O-Net.

Input	Operator	c	s
48 x 144 x 3	Conv 3 x 7	32	1
46 x 138 x 32	MP 3 x 3	32	2
23 x 69 x 32	Conv 3 x 9	64	1
21 x 61 x 64	MP 3 x 3	64	2
10 x 30 x 64	Conv 3 x 7	64	1
8 x 24 x 64	MP 3 x 3	64	2
4 x 12 x 64	Conv 2 x 4	128	1
3 x 9 x 128	FC	256	1

The architectures of R-Net, where "MP" means max pooling, "Conv" means convolution, and "FC" means fully connect layer. All layers in the same sequence have the same number c of output channels. The stride in convolution and pooling is 1 and 2, respectively.

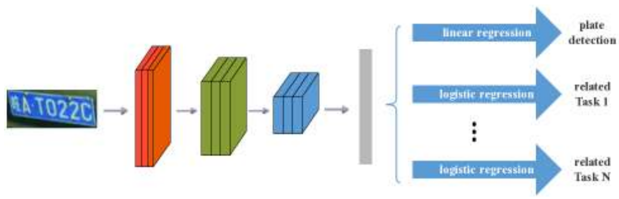


FIGURE 2. The architectures of multi-task computing.

to describe the license plate in more details. In particular, the network will output four corner positions.

The model adopts a well-designed three-level network, which successfully improves the detection result and achieves a good balance between detection precision and real-time performance. All three networks adopt end-to-end multi-task framework including license plate classification task, border regression task, corner point regression task, and color recognition task after the last fully connected layer. It also utilizes the correlation between tasks to improve the network performance during the training process.

B. MULTI-TASK TRAINING

Next, the multi-task collaborative computing of MTLPR is introduced in detail. We employ four tasks to train our CNN network simultaneously in the training process: plate classification, bounding box regression, plate landmark localization, and plate color recognition. There are different types of training images in the learning process, such as plate, non-plate and partially aligned plate. Multiple related tasks can be trained at the same time, to ensure both the independence and correlation. It solves problems caused by illumination and occlusion in the field of license plate location and detection, so as to improve the learning performance.

In our multi-task network, it produces candidate windows quickly through a shallow CNN. Then, it refines the windows to reject a large number of non-plate windows through a more complex CNN. The huge number of training samples waste the memory and time for training and testing. It is efficient to process multi-task training based on cascade networks which reject large number of easy examples at earlier stages.

We use cascaded networks and multi-task framework to train our CNN classifiers: license plate classification, bounding box regression, corner localization, and color recognition, as shown in Figure 2.

1) LICENSE PLATE CLASSIFICATION

The learning objectives of this task can be formulated into a two-class classification problem. We calculate the cross-entropy loss function for each sample x_i to determine whether input image include the license plate information, referring to (1):

$$L_i^{det} = -y_i^{det} log(p_i) + (1 - y_i^{det})(1 - log(p_i))) \quad (1)$$

where p_i represents the probability that indicates a sample being a license plate calculated by the network, the notation $y_i^{det} \in \{0, 1\}$ indicates the ground truth label.

2) BOUNDARY BOX REGRESSION

We predict the offset between each candidate and its nearest ground truth, translate and scale the bounding box of candidate window (i.e., the coordinate, width, and height). Therefore, the learning objective can be formulated as a regression problem to obtain the license plate area. For each sample x_i, the Euclidean distance can be calculated as a loss function:

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

where $\hat{y}_i^{box}$ represents regression target obtained through the network prediction, y_i^{box} represents the ground truth coordinate, which is a four-dimensional vector (x, y, w, h), including left, top, width, and height of the bounding box.

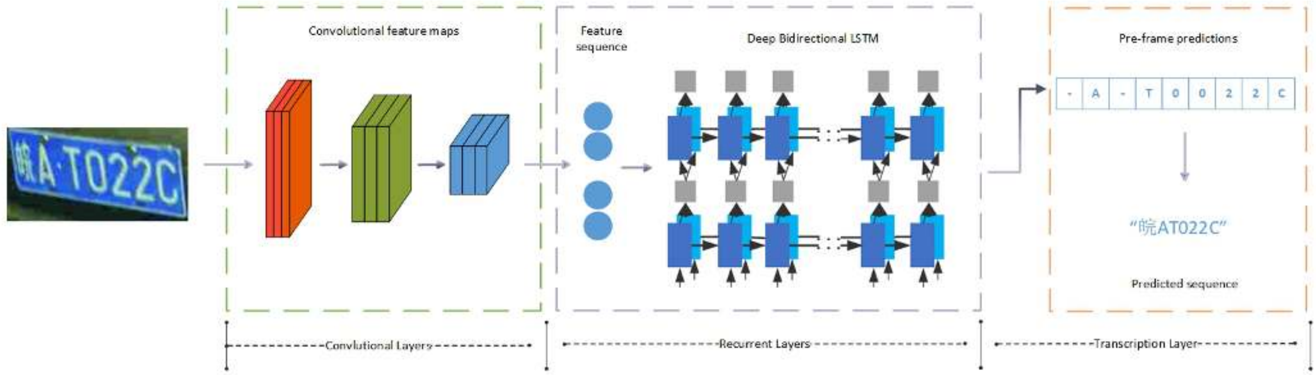


FIGURE 3. **The architectures of plate recognition. There are there layers including convolutional layers, recurrent layers, and transcription layers.**

3) FOUR-CORNER LOCALIZATION

This task can also be formulated as a regression problem, similar to bounding box regression task. We minimize the Euclidean loss:

$$L_i^{corner} = ||\hat{y}_i^{corner} - y_i^{corner}||_2^2 \tag{3}$$

where $\hat{y}_i^{corner}$ represents the coordinate of the four-corner points of the license plate obtained through the network prediction, y_i^{corner} is the ground truth coordinate, which is an eight-dimensional vector.

4) COLOR RECOGNITION

The Chinese license plates is divided into five main categories according to colors: blue, white, yellow, black and green. The learning objectives of this task can be formulated into a multi-class classification problem. We calculate the cross-entropy loss function for each sample x_i to determine the color of license plate, referring to:

$$L_i^{color} = -\sum_j y_i^{color} log(p_i^j) \tag{4}$$

where p_i^j represents the probability that indicates a sample being a certain color, the notation y_i^{color} indicates the ground truth label.

5) MULTI-TASK TRAINING

Since we employ different tasks in each CNN, the overall learning target can be formulated as:

$$L = \sum_{i=1}^{N}(\alpha_{det}L_i^{det} + \alpha_{box}L_i^{box} + \alpha_{corner}L_i^{corner} + \alpha_{color}L_i^{color}) \tag{5}$$

where N is the number of training samples. We use $\alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{corner} = 0.5, \alpha_{color} = 0.5$ in P-Net and R-Net, while $\alpha_{det} = 1, \alpha_{box} = 0.5, \alpha_{corner} = 1, \alpha_{color} = 0.5$ in O-Net to denote on the task importance.

C. LICENSE PLATE RECOGNITION

After the license plate detection, the license plate image may have horizontal tilt, vertical tilt or keystone distortion problems due to factors such as shooting angle and lens. Then the license plate correction process should be performed, which is beneficial to remove the influence of noise and more conducive to character recognition. In this paper, we adopt the perspective transformation method to realize the related matrix transformation on the four detected corner points of the license plate.

The traditional LPR frameworks which recognize plate after segmentation could not achieve a high precision. Therefore, we adopt Convolutional Recurrent Neural Network (CRNN) and Connectionist Temporal Classification (CTC) structure [28], which applies the end-to-end algorithm to LPR and outputs the license plate characters directly without the segmentation. The entire network can be divided into three parts, as shown in Figure 3. The bottom convolution layer automatically extracts the feature maps from each input image. Above the convolution network, a recurrent network is constructed to predict for each frame of the feature sequences. Finally per-frame prediction is converted into a predicted sequence through the transcription layer, as shown in Table 4:

(1) Convolutional Layers: We use the convolutional layer and the maximum pooling layer to construct convolutional layers and extract feature maps of input images. All input images should be adjusted to the same scale, and then the feature sequences are extracted from the feature maps and used as input of the recurrent layers.

(2) Recurrent Layers: This is a deep bidirectional long-term memory (LSTM) network [29] above convolutional layers. Based on the convolutional feature sequences, the text sequence is continuously extracted. LSTM is a special kind of Recurrent Neural Network (RNN), which is mainly used to solve the gradient disappearance and gradient explosion problems in long sequence training, and has better performance than ordinary RNN networks. There are three main advantages to use the recurrent layers. One is that LSTM has a strong ability to capture context information of a sequence. It is more stable and more helpful to use context

TABLE 4. The architectures of recognition network.

Input	Operator	c	s	p
96 x 32 x 3	Conv 3 x 3	64	1	1
96 x 32 x 64	MP 2 x 2	64	2	0
48 x 16 x 64	Conv 3 x 3	128	1	1
48 x 16 x 128	MP 2 x 2	128	2	0
24 x 8 x 128	Conv 3 x 3	256	1	1
24 x 8 x 256	MP 1 x 2	256	2	0
24 x 4 x 256	Conv 3 x 3	512	1	1
24 x 4 x 512	MP 1 x 2	512	2	0
24 x 2 x 512	Conv 2 x 2	512	1	0
23 x 1 x 512	Map-to-Sequence	-	-	-
23 x 512	Bidirectional-LSTM	-	-	-
-	Transcription	-	-	-

The architectures of recognition network, where "MP" means max pooling, "Conv" means convolution. All layers in the same sequence have the same number "c" of output channels. Each layer has a stride "s" and padding "p".

TABLE 5. Descriptions of different sub-datasets in CCPD.

Sub-dataset	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.
FN	The distance from the LP to the shooting location is relatively far or near.
Rotate	Great horizontal tilt degree (20° - 50°) and the vertical tilt degree varies from −10° to 10°.
Tilt	Horizontal tilt degree (15° - 45°) and vertical tilt degree (15° - 45°).
Blur	Blur largely due to hand jitter while taking pictures.
Weather	Images taken on a rainy day, snow day or fog day.
Challenge	The most challenging images.
NP	Images of new cars without a LP.

TABLE 6. Plate Detection Precision (Percentage).

Model	FPS	AP	Recall
Cascade classifier [35]	32	47.2	81.6
SSD300 [16]	40	94.4	86.7
Faster-RCNN [13]	15	92.9	88.3
YOLO-V3 [15]	42	93.1	93.1
TE2E [17]	3	94.2	94.2
RPNet [31]	61	94.5	95.8
MTLPR	65	95.8	96.9
MTLPR + data augmentation	65	97.7	97.3

AP denotes average precision in the whole test set and FPS denotes frames per second.

for image-based sequence recognition than to process each symbol independently. Second, LSTM can propagate the loss error back to its input, and allow us to train the recurrent and convolutional layers together in one network. Third, the LSTM can handle sequences of any length.

(3) Transcription Layers: A temporal classification algorithm named Connectionist Temporal Classification (CTC) is used to convert the per-frame prediction of each frame into a final character sequence. CTC [30] focuses on solving the alignment problem of given labels and output data. Different from the traditional methods which need to align the input data with the given label in time to calculate cross entropy loss by frame, the CTC algorithm can be trained without label alignment. It cares about whether the output is consistent with the ground truth label as a whole, thereby reducing the tedious work of manual label.

The recognition algorithm can complete end-to-end training in LPR, without character segmentation and horizontal scaling operation. It can recognize plate characters of any length, including single-layer and double-layer license plate.

IV. EXPERIMENTAL RESULT

A. DATA SET

In this section, we mainly introduce the experimental results and compare the performance of our method with other state-of-the-art models on the Chinese City Parking Dataset (CCPD) [31]. The CCPD data set is the largest publicly available labeled license plate dataset by far in China. There are about 250K independent license plate images with different backgrounds, different shooting angles, different time, and different lighting levels, refer to Table 5. So we did not perform experiments on other data sets [32]–[34] because most existing license plate recognition data sets do not have richer diversity than CCPD data set.

1) DATA AUGMENTATION

Despite the diversity of CCPD, the data is relatively imbalanced. The number of license plate images in complex natural environment is small. Therefore, we performed data augmentation and expanded the CCPD data set. Data augmentation in machine learning refers to the techniques that synthetically expand a data set by applying transformations on the existing examples, in order to augment the amount of available training data. We generate about 500k license plates by the following augmentation methods:

1) Using image standardization.
2) Using geometric transform, such as translation, flip, and rotation.
3) Using random brightness and contrast adjustment to enhance the image.
4) Adding distortion noise and combining with the natural environment.

The samples of data augmentation are shown in Figure 4.

2) ABLATION EXPERIMENTS

We furthermore conduct different ablation experiments to evaluate the effect of our algorithm. It is significant to identify the precision and speed improvements of data augmentation. The experimental results of ablation prove that the data augmentation techniques help to improve precision significantly.

All of our training tasks are completed on the GPU, using the Tesla K80 GPU, and all of our prediction tasks run on the Intel i7-6700 processor with 8 cores of 3.40 GHz. Experiment results on CCPD data set show the robustness of LPR under complex conditions.

FIGURE 4. The plate samples of data augmentation results.

FIGURE 5. Some sample of recognized images on CCPD data set. The yellow border indicates the plate area, and the four red dots are corner point of the plate.

B. DETECTION RESULTS

First, we divide the CCPD data set into two parts, use one part as training data, and use the other part as testing data for detection and recognition. Then we add our augmented data set to train our model. Then we analyze and present the comparison of result on original data set and augmented data set. Table 6 shows the detection precision (percentage) of each state-of-the-art license plate detection algorithm including Cascade classifier, YOLO, SSD and Faster-RCNN.

Detection precision is followed the standard protocol in object detection Intersection-over-Union (IoU). When the overlap between the detection area and the ground truth exceeds 70% (IoU > 0.7), the detection result is considered correct [36]. As shown in Table 6, the Cascade classifier is difficult to accurately locate the license plate, and the performance is poor, especially when dealing with the inclined license plate. YOLO has a higher detection precision overall, but only a low detection precision on the CCPD-FN data set, so the detection performance of the small object is bad. The method in this paper is optimized in both detection and recognition algorithms, and it is faster and more robust than RPNet and YOLO-V3 in performance.

C. RECOGNITION RESULTS

We compare the detection and recognition result with other model. The final recognition results consider both the IoU and the recognition precision. The license plate recognition is considered as correct only when the IoU is greater than

TABLE 7. Plate Recognition Precision (Percentage).

Model	FPS	AP	Recall
Cascade classifier + HC	29	58.9	74.2
SSD300 + HC	35	95.2	82.6
Faster-RCNN + HC	13	92.8	86.8
YOLO-V3 + HC	36	93.7	91.9
TE2E	3	94.4	94.2
RPNet	61	95.5	95.4
MTLPR	64	97.2	96.7
MTLPR + data augmentation	64	98.8	97.7

AP denotes average precision in the whole test set and FPS denotes frames per second. HC denotes Holistic-CNN.

TABLE 8. MTLPR Comparison to Popular Models.

Model	Million Parameters
Cascade classifier	0.12
Faster-RCNN	251
SSD300	25
YOLO-V3	23
MTLPR	1.5

0.6 and all the characters of the license plate in the image are correctly recognized. The recognition results are shown in Table 7.

In addition to MTLPR, TE2E and RPNet, we add a high-performance model to other object detection models for license plate recognition, where HC represents the Holistic-CNN [24]. These combined models can achieve high recognition speed and precision, but our method still has higher performance than other models. The license plate detection and recognition results are shown in Figure 5.

D. MODEL ANALYSIS

Table 8 compares our proposed method with other state-of-the-art methods. Cascade classifier is traditional method, which has fewer parameters and less accuracy. SSD and Faster-RCNN are all based on VGG16, and have large scale parameters. MTLPR is more accurate than YOLO-V3 while being 15 times smaller. Therefore, the network size and computation cost of our proposed method is much smaller than other models.

V. CONCLUSION

In this paper, we propose a lightweight method for end-to-end car license plates detection and recognition. We realize the multi-task license plate detection algorithm of three-layer cascaded network, to complete the license plate detection, corner detection and color recognition task using an entire network structure. Compared with the traditional detection algorithm, multi-task parallel processing improves the detection precision and speed. Then the license plate recognition adopts CRNN and CTC model to realize end-to-end object recognition algorithm with high precision, which can support single-layer and double-layer license plates with different color. Compared with traditional algorithms and other neural network algorithms, our method has a higher recognition precision on the CCPD dataset. In the future, we will open our source code and data set.

REFERENCES

[1] S. M. Silva and C. R. Jung, "Real-time Brazilian license plate detection and recognition using deep convolutional neural networks," in *Proc. 30th SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Oct. 2017, pp. 55–62.

[2] A. E. Ghahnavieh, A. Amirkhani-Shahraki, and A. A. Raie, "Enhancing the license plates character recognition methods by means of SVM," in *Proc. 22nd Iranian Conf. Elect. Eng. (ICEE)*, May 2014, pp. 220–225.

[3] B. Hong, and C. Yang, "An approach to license plate locating in intelligent transportation system," in *Proc. 2nd IEEE Int. Conf. Pervasive Comput. Appl.*, Jul. 2007, pp. 319–322.

[4] K. Venkatesan and A. Shamir, "Application of neural networks in character recognition," *Int. J. Comput. Appl.*, vol. 52, no. 12, pp. 1–6, 2012.

[5] S. H. Park, K. I. Kim, K. Jung, and H. J. Kim, "Locating car license plates using neural networks," *Electron. Lett.*, vol. 35, no. 17, pp. 1475–1477, Aug. 1999.

[6] M. S. Al-Shemarry, Y. Li, and S. Abdulla, "Ensemble of adaboost cascades of 3L-LBPs classifiers for license plates detection with low quality images," *Expert Syst. Appl.*, vol. 92, pp. 216–235, Feb. 2018.

[7] J.-W. Hsieh, S.-H. Yu , and Y.-S. Chen, "Morphology-based license plate detection from complex scenes," in *Proc. Object Recognit. Supported Interact. Service Robots*, Aug. 2002, pp. 176–179.

[8] S. Yu, B. Li, Q. Zhang, C. Liu, and M. Q.-H. Meng, "A novel license plate location method based on wavelet transform and EMD analysis," *Pattern Recognit.*, vol. 48, no. 1, pp. 114–125, 2015.

[9] R. Wang, "License plate detection using gradient information and cascade detectors," *Optik*, vol. 125, no. 1, pp. 186–190, 2014.

[10] Y. Lee, T. Song, B. Ku, S. Jeon, D. K. Han, and H. Ko, "License plate detection using local structure patterns," in *Proc. 7th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug./Sep. 2010, pp. 574–579.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[12] R. Girshick, "Fast R-CNN" in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[13] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[15] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Apr. 2018, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[17] H. Li, P. Wang , and C. Shen, "Toward end-to-end car license plate detection and recognition with deep neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1126–1136, Mar. 2019.

[18] L. Xie, A. 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.

[19] 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.

[20] F. D. Kurpiel, R. Minetto, and B. T. Nassu, "Convolutional neural networks for license plate detection in images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3395–3399.

[21] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): A state-of-the-art review," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 2, pp. 311–325, Feb. 2013.

[22] Y. Yuan, W. Zou, Y. Zhao, X. Wang, X. Hu, and N. Komodakis, "A robust and efficient approach to license plate detection," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1102–1114, Mar. 2017.

[23] O. Bulan, V. Kozitsky, P. Ramesh, and M. Shreve, "Segmentation- and annotation-free license plate recognition with deep localization and failure identification," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2351–2363, Sep. 2017.

[24] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5325–5334.

[25] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[26] H. Qin, J. Yan, X. Li, and X. Hu, "Joint Training of Cascaded CNN for face detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3456–3465.

[27] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. IEEE 18th Int. Conf. Pattern Recognit.*, Aug. 2006, pp. 850–855.

[28] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 369–376.

[31] Z. Xu, W. Yang, A. Meng, N. Lu, H. Huang, C. Ying, and L. Huang, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *Proc. 15th Eur. Conf. Comput. Vis.*, 2018, pp. 261–277.

[32] *Caltech: Caltech Licese Plate Dataset*. Accessed: Mar. 20, 2019. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html

[33] *Zemris: Zemris License Plate Dataset*. Accessed: Mar. 20, 2019. [Online]. Available: http://www.zemris.fer.hr/projects/LicensePlates/hrvatski/rezultati.shtml

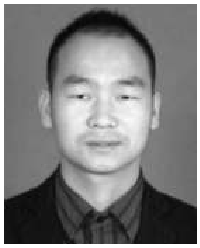
[34] J. Špaňhel, J. Sochor, R. Juránek, A. Herout, L. Maršík, and P. Zemcík, "Holistic recognition of low quality license plates by CNN using track annotated data," in *Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug./Sep. 2017, pp. 1–6.

[35] S.-Z. Wang and H.-J. Lee, "A cascade framework for a real-time statistical plate recognition system," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 2, pp. 267–282, Jun. 2007.

[36] J. Han, J. Yao, J. Zhao, J. Tu, and Y. Liu, "Multi-oriented and scale-invariant license plate detection based on convolutional neural networks," *Sensors*, vol. 19, no. 5, p. 1175, 2019.

WANWEI WANG received the M.S. degree in signal and information processing from the Civil Aviation University of China, in 2010. He is currently the Deputy Director of the Flight Tracking and Surveillance Technology Research Center, the Assistant Director of the Tianjin Key Laboratory for Advanced Signal Processing, Civil Aviation University of China, where he has also been a Lecturer with the Institute of College of Electronic Information and Automation, since 2010. His current research interest includes signal and image processing.

JUN YANG received the M.S. degree in signal and information processing from the Civil Aviation University of China, Tianjin, in 2014. Since 2014, he has been a Research Assistant with the School of Electronic Information and Automation, Civil Aviation University of China. His research interests include deep learning and human behavior recognition.



MIN CHEN received the M.S. degree in signal and information processing from Shenzhen University, in 2013. Since 2013, he has been a Research Assistant with the Institute of College of Electronic Information and Automation, Civil Aviation University of China. His research interests include the development of air traffic control systems and ADS-B surveillance systems, ADS-B, and radar signal process and simulation.

PENG WANG received the M.S. degree in signal and information processing from the Civil Aviation University of China, Tianjin, in 2013. Since 2013, he has been a Research Assistant with the School of Electronic Information and Automation, Civil Aviation University of China. His research interests include the development of air traffic control systems and signal processing.