IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2020, accepted November 15, 2020, date of publication November 24, 2020, date of current version December 8, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3040238*

# A Robust License Plate Recognition Model Based on Bi-LSTM

**YONGJIE ZOU[1], YONGJUN ZHANG[1], JUN YAN[2], XIAOXU JIANG[2], TENGJIE HUANG[2], HAISHENG FAN[2], AND ZHONGWEI CUI[3]**

[1]Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, College of Computer Science and Technology, Guizhou University, Guiyang 550025, China
[2]Zhuhai Orbita Aerospace Science and Technology Company, Ltd., Zhuhai 519000, China
[3]Big Data Science and Intelligent Engineering Research Institute, Guizhou Education University, Guiyang 550018, China

Corresponding authors: Yongjun Zhang (zyj6667@126.com) and Jun Yan (yan@myorbita.net)

This work was supported in part by the 2017 Zhuhai Innovation and Entrepreneurship Team Introduction Program under Grant ZH01110405170027PWC and Grant ZH0405-1900-01PWC, in part by the Cloud Service Platform and Applications based on "ZHUHAI No.1" Constellation and Remote Sensing Big Data under Grant ZDXK [2018]007, in part by the Key Supported Disciplines of Guizhou Province-Computer Application Technology under Grant QianXueWeiHeZi ZDXK [2016]20, and in part by the National Natural Science Foundation of China under Grant 61462013 and Grant 61661010.

**ABSTRACT** License plate detection and recognition are still important and challenging tasks in natural scenes. At present, most methods have favorable effect on license plate recognition under restrictive conditions, and most of such license plates are shot under good angle and light conditions. However, for license plates under non-restrictive conditions, such as dark, bright, rotated conditions etc. from the Chinese City Parking Dataset (CCPD), the performance of some methods of license plate recognition will be significantly reduced. In order to improve the accuracy of license plate recognition under unrestricted conditions, a robust license plate recognition model is proposed in this paper, which mainly includes license plate feature extraction, license plate character localization, and feature extraction of characters. First of all, the model can activate the regional features of characters and fully extract the character features of license plates. Then locate each license plate character through Bi-LSTM combined with the context location information of license plates. Finally, 1D-Attention is adopted to enhance useful character features after Bi-LSTM positioning, and reduce useless character features to realize effective acquisition of character features of license plates. A large number of experimental results demonstrate that the proposed algorithm has good performance under unrestricted conditions, which proves the effectiveness and robustness of the model. In CCPD-Base, CCPD-DB, CCPD-FN, CCPD-Tilt, CCPD-Weather, CCPD-Challenge and other sub-datasets, the recognition rates reach 99.3%, 98.5%, 98.6%, 96.4%, 99.3% and 86.6% respectively.

**INDEX TERMS** Character localization, license plate detection, license plate recognition.

## I. INTRODUCTION

As shown in Figure 1, for images taken in dark, rotated, snowy, light, tilted or challenged scenes, it would be challenging for character segmentation of license plates. In these complex scenarios from Chinese City Parking Dataset (CCPD), if the character segmentation is not accurate, the recognition accuracy of license plates would decrease. In order to reduce inaccurate character segmentation and susceptibility to environmental influences, many researchers have done abundant work in license plate recognition in recent years. They used convolutional neural networks to extract license plate features, and adopted methods without segmenting characters to avoid inaccurate character segmentation, which improved the speed and accuracy of license plate recognition. Some of these methods only work in normal license plate images. Some methods employed Spatial Transformation Networks (STN) [1] to correct tilted or distorted license plates, through OCR or convert them to serial labels, and identify license plate features through long- and short-term networks. Aiming at these complex scenes, this paper designs a robust license plate recognition method, which can effectively recognize normal license plate images and some challenging license plates. Inspired by the MobileNetV3 [2] and Xception [3]

The associate editor coordinating the review of this manuscript and approving it for publication was Byung Cheol Song.

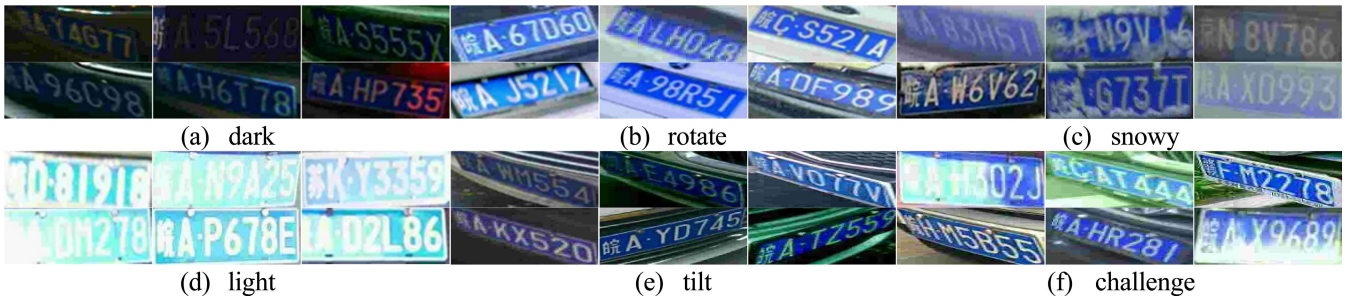



**FIGURE 1.** Scenes where the proposed algorithm successfully recognizes license plates.

networks, this paper proposes a license plate feature extraction network that can accurately activate the feature regions of license plate characters. The Bi-directional Long Short-Term Memory (Bi-LSTM) network is used to locate the character position, and 1D-Attention is used to extract features of license plate characters. This method can suppress useless features, enhance useful features, and effectively process regular and irregular license plate images in actual scenes. This method does not need to segment single license plate characters, yet extract the character features instead, thus avoiding inaccurate recognition caused by inaccurate character segmentation. In order to verify the robustness of the model, we conducted experiments on four datasets CCPD [4], PKUdata [5], CLPD [6] and AOLP [7]. Since PKUdata [5] does not include license plate labels, we manually labelled 2253 license plate images. Two datasets PKUdata and CLPD were only used for testing. CCPD contains about 290K real license plate images in different complex scenes. It is also currently the largest license plate dataset with the most scenes, which is challenging. In addition, we use the YOLOv3 [8] algorithm to locate license plates before license plate recognition.

This paper makes the following main contributions:

1. A license plate feature extraction network is designed based on spatial attention mechanism, which can activate the regions of license plate character features, and fully extract features of license plates.

2. The Bi-LSTM algorithm is adopted and visualized through heat maps, combined with the contextual position information of license plate characters. This method can accurately locate each character of license plates.

3. One-Dimensional (1D) attention is added to enhance useful character features and suppress useless character features in license plates, which can effectively extract character features, and improve the accuracy of license plate recognition.

By evaluating the reliability of the proposed algorithm on four public datasets, the experimental results prove that the proposed network model has a certain development in accuracy compared with other current algorithms, and has good generalization ability.

The remaining of this paper is organized as follows. Section 2 reviews the related work of license plate detection and recognition; Section 3 introduces the proposed recognition network; Section 4 presents the datasets, training parameter settings and evaluation metrics; Section 5 analyzes the experimental results. Finally, Section 6 is the summary of the paper.

## II. RELATED WORK

This section briefly introduces the work related to the detection and recognition of license plates.

### A. LICENSE PLATE DETECTION

License plate detection includes traditional methods and deep learning methods. Traditional methods mainly extract features of objects according to the background color, edge and contour of the target. For example, Hsieh *et al.* [9] utilized morphological methods to significantly reduce the number of candidate images extracted from cluttered images, which speeds up the subsequent license plate recognition. This method improves the effectiveness and robustness of license plate detection. Yu *et al.* [10] proposed a robust method based on wavelet transform and Empirical Mode Decomposition (EMD) analysis, to locate the position of the license plate in images, aiming to solve some challenges in practical applications such as illumination changes, complex backgrounds, and perspective changes. The method can locate the position of license plates of various types with an accuracy of 97.91%. A license plate detection method based on multi-level information fusion was proposed by Yao [11], which aims to reduce the false alarm rate of the traditional Adaboost detector. The experimental results illustrate that the advancement of feature extraction and multi-level information fusion significantly enhances the performance of license plate detection. Compared with traditional methods, deep learning does not require manual extraction of features. At present, this method has been widely used in license plate detection. MASK-RCNN [12] proposes a simple, flexible, and universal instance segmentation model framework, which can generate a high-quality mask for each instance and segment the target accurately. SSD [13] eliminates the candidate frame for generating regional recommendations and the subsequent pixel or feature resampling stage, and encapsulates all calculations in a single network, making the network accurate, fast, easy to train, and sensitive to small targets. YOLO

and its upgraded versions [8], [14]–[16] regard detection tasks as regression tasks, and have been favored by many scientific research enthusiasts due to their high accuracy, fast speed and real-time detection in recent years. Therefore, YOLOv3 and YOLOv4 are adopted in this paper as the license plate detection frameworks. Due to their hardware performance and some uncertainties, as shown in Table 3, the comparison in equal conditions reveals that the performance of YOLOv4 is not as good as YOLOv3 in the CCPD dataset. So YOLOv3 is adopted to detect license plates.

### B. LICENSE PLATE RECOGNITION

License plate recognition methods are mainly divided into two categories: methods based on segmented characters [17], [18] and methods based on non-segmented characters [4], [19]–[23]. The character-split method segments the license plate into individual characters, and then uses the OCR model [18], [20], [24] to recognize segmented characters. By exploiting an improved YOLO network, the license plate characters were segmented and OCRed in [18]. The accuracy of character segmentation depends on segmentation performance, and is susceptible to external conditions, such as light intensity, license plate blurring, etc. These conditions may reduce the accuracy of license plate recognition. At present, most researchers adopt the method of not dividing characters. For example, [19] turned the character recognition into a sequential labeling problem, and recognized the sequence features of the entire license plate through a long and short-term memory recurrent neural network (RNN). The main advantage of this approach is that it does not require segmentation, and it can mine contextual information to avoid errors caused by segmentation. RPnet was proposed by Xu et al [4]. This model can quickly and accurately predict license plate bounding boxes and identify the corresponding license plate simultaneously, by extracting the ROI features in different convolutional layers. The model outperforms existing object detection and recognition approaches in both accuracy and speed with 98.5% accuracy. Li *et al.* [21] proposed a unified deep neural network for the detection and recognition of license plate images in natural scenes, which can recognize license plates while locating license plate characters. This method not only avoids the accumulation of intermediate errors, but also improves processing speed. Compared with previous work, our approach utilizes a Bi-LSTM algorithm based on 1D-Attention, which can locate characters through Bi-LSTM and extract character features by 1D-attention regardless of the appearance of LP images, thus enabling it to be applied to LPs in any direction.

### C. SCENE TEXT RECOGNITION

Scene text recognition has been a research interest in the field of computer vision. It is currently using an encoder-decoder framework with attention mechanism, to learn the mapping between input images and output sequences in a data-driven manner. License plate recognition can be seen as a special case of general scene text recognition tasks, for they share the similarity of continuous text. A flexible Thin-Plate Spline transformation was proposed by Shi *et al.* [32] for processing a variety of text irregularities. The recognition model predicts the character sequence from the rectified image. Cheng *et al.* [33] proposed Focusing Attention Network, which uses a focused attention mechanism to pull back the drifting attention and improve the accuracy of text recognition. Luo *et al.* [34] proposed a multi-object rectifier attentional network (MORAN) for general scene text recognition. MORAN consists of a multi-object rectifier network and an attention-based sequence recognition network, which can read both regular and irregular scene texts. The network reduces the difficulty of recognition and makes it easier for the attention-based sequence recognition network to read irregular text. Li *et al.* [35] achieved state-of-the-art performance on both regular and irregular scene text recognition with the encoder-decoder framework in LSTM and a two-dimensional attention module. This research is similar to ours, but in our framework, the Bi-LSTM and one-dimensional attention modules are adopted to accurately locate character features and enhance the recognition performance of the model.

## III. PROPOSED METHOD

### A. LICENSE PLATE RECOGNITION

As shown in Figure 2, the license plate recognition network is divided into three modules: license plate feature extraction module, license plate character localization module, and character feature extraction module. The license plate feature extraction module, as shown in Figure 3, uses deep separable convolutions and spatial attention mechanism to activate the feature regions of characters and fully extract features of license plates. The license plate character localization module adopts Bi-LSTM combined with the contextual position information of license plate characters to achieve accurate character localization. The character feature extraction module utilizes the 1D-Attention method, and sets the size of the convolution kernels to 1. Without changing the scale of the feature map, the features in different channels are merged. The dimension of the license plate feature ($38\times14\times256$) is reduced to ($38\times14\times1$), and non-linear excitation is added to enhance the abstract expression ability of local modules. Then the 1D-attention weights are used to reinforce the license plate character features, suppress useless character features, and multiply the original license plate features to extract character features. The license plate character classifier maps the dimension of the extracted character features to the predicted category, and uses the SoftMax function to obtain the correct characters in license plates.

### B. LICENSE PLATE FEATURE EXTRACTION

Inspired by MobileNetV3, Xception, and spatial attention mechanism, we design a license plate feature extraction network, the details of which are demonstrated in Figure 3. The convolutional layers of the model consist of deep convolutions and deep separable convolutions, and ReLU activation

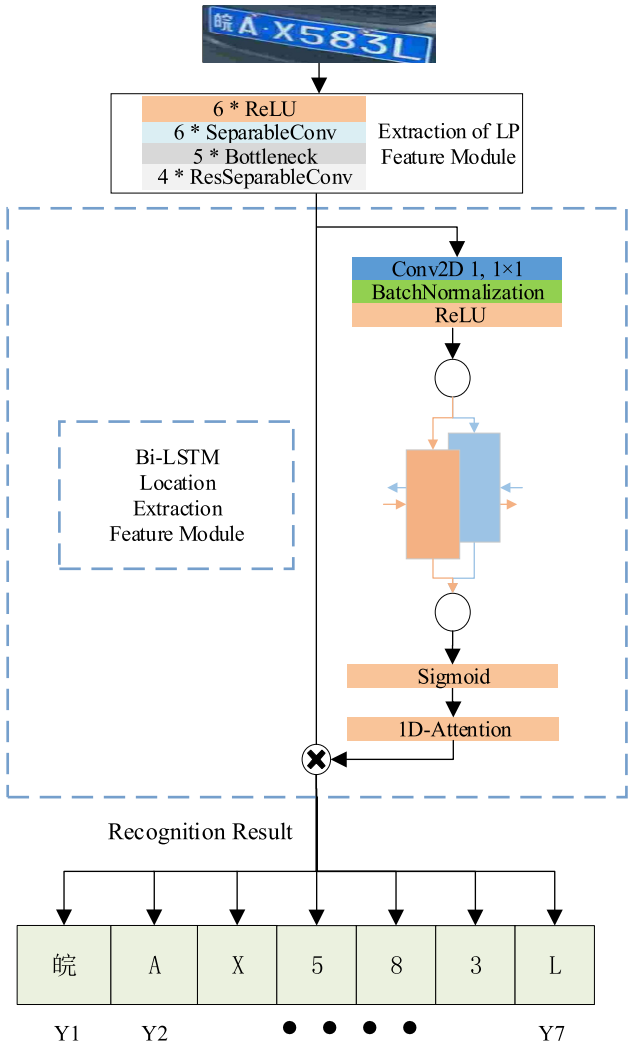



**FIGURE 2.** License plate recognition process.

function is added to each separated convolution to increase the nonlinear expression capability of the network. After every two separated convolutions, the network is fused with shallow features. Through that process, the network thus has deep and shallow features, which can prevent the vanishing gradient problem and speed up training process. Bottleneck block contains squeeze-and-excite [25] attention mechanism, which utilizes global pooling to obtain the global field of view. It also adopts global pooling as guide information, to enhance or suppress the responses of the characteristics of different channels according to their global response values, thus making the output features more distinguishable. Repeated ResSeparableConv blocks are chosen to extract deeper features of license plates. Then spatial attention is used to continuously compress the channel information through mean pooling and maximum pooling. Therefore, the network pays attention to the location information of the spatial characters of license plates. Finally, we use two separated convolutions to obtain a 256-dimensional feature map of the license plate characters (size ($38 \times 14 \times 256$)). The detailed network parameters are shown in Table 1.

**TABLE 1.** License plate feature extraction network parameters, where "exp size" and "out" denote expansion factor and output channel dimension respectively. "SE" represents whether there is a Squeeze-And-Excite in that block. "NL" denotes the nonlinear type used, "HS" denotes H-swish, "RE" denotes ReLU6 and "s" denotes stride.

| Input | Operator | exp size | out | SE | NL | s |
|---|---|---|---|---|---|---|
| 152 × 56 × 1 | conv2d, 3×3 | - | 64 | - | HS | 2 |
| 76 × 28 × 64 | SeparableConv, 3×3 | - | 128 | - | - | 1 |
| 76 × 28 × 128 | SeparableConv, 3×3 | - | 128 | - | - | 1 |
| 76 × 28 × 128 | bottleneck, 3×3 | 16 | 128 | True | RE | 2 |
| 38 × 14 × 128 | SeparableConv, 3×3 | - | 256 | - | - | 1 |
| 38 × 14 × 256 | SeparableConv, 3×3 | - | 128 | - | - | 1 |
| 38 × 14 × 128 | bottleneck, 3×3 | 72 | 256 | False | RE | 1 |
| 38 × 14 × 256 | ResSeparableConv, 3×3 | - | 256 | - | - | 1 |
| 38 × 14 × 256 | ResSeparableConv, 3×3 | - | 256 | - | - | 1 |
| 38 × 14 × 256 | ResSeparableConv, 3×3 | - | 256 | - | - | 1 |
| 38 × 14 × 256 | ResSeparableConv, 3×3 | - | 256 | - | - | 1 |
| 38 × 14 × 256 | bottleneck, 5×5 | 120 | 512 | True | HS | 1 |
| 38 × 14 × 512 | bottleneck, 5×5 | 144 | 512 | True | HS | 1 |
| 38 × 14 × 512 | bottleneck, 5×5 | 288 | 256 | True | HS | 1 |
| 38 × 14 × 256 | SeparableConv, 3×3 | - | 256 | - | - | 1 |
| 38 × 14 × 256 | SeparableConv, 3×3 | - | 256 | - | - | 1 |

ReLU uses x for linear activation in the $x \to 0$ region, which may cause the value after activation to be too large to keep the stability of the model. ReLU6 can offset the linear increase of the ReLU excitation function. Since sigmoid calculations are time-consuming, h-swith is adopted instead, which eliminates the potential accuracy loss when quantifying.

$$\mathrm{ReLU6} = \min\left(\max\left(x, 0\right), 6\right) \ni [0, 6] \tag{1}$$

$$\mathrm{h_{swith}} = x\frac{ReLU6\left(x+3\right)}{6} \tag{2}$$

### C. LICENSE PLATE CHARACTER LOCALIZATION

Figure 4 represents the detailed process of license plate character positioning. Taking the location process of the 'A' character features corresponding to Y2 as an example, from the visualization of the feature map (38, 14) output by the first convolutional layer of the localization module in the figure, it can be seen that the response value of the network to the features at the 'A' character position indicates the activation region of the network for the 'A' character feature. 38 license plate features are taken in the direction of width (w) of the feature map, and the Bi-LSTM method is adopted to combine the contextual location information of the license plate feature characters to locate the relative position of each character. Based on the response value region of the character 'A' feature, the full connection and Sigmoid activation functions are then used to obtain the 1D-Attention weight

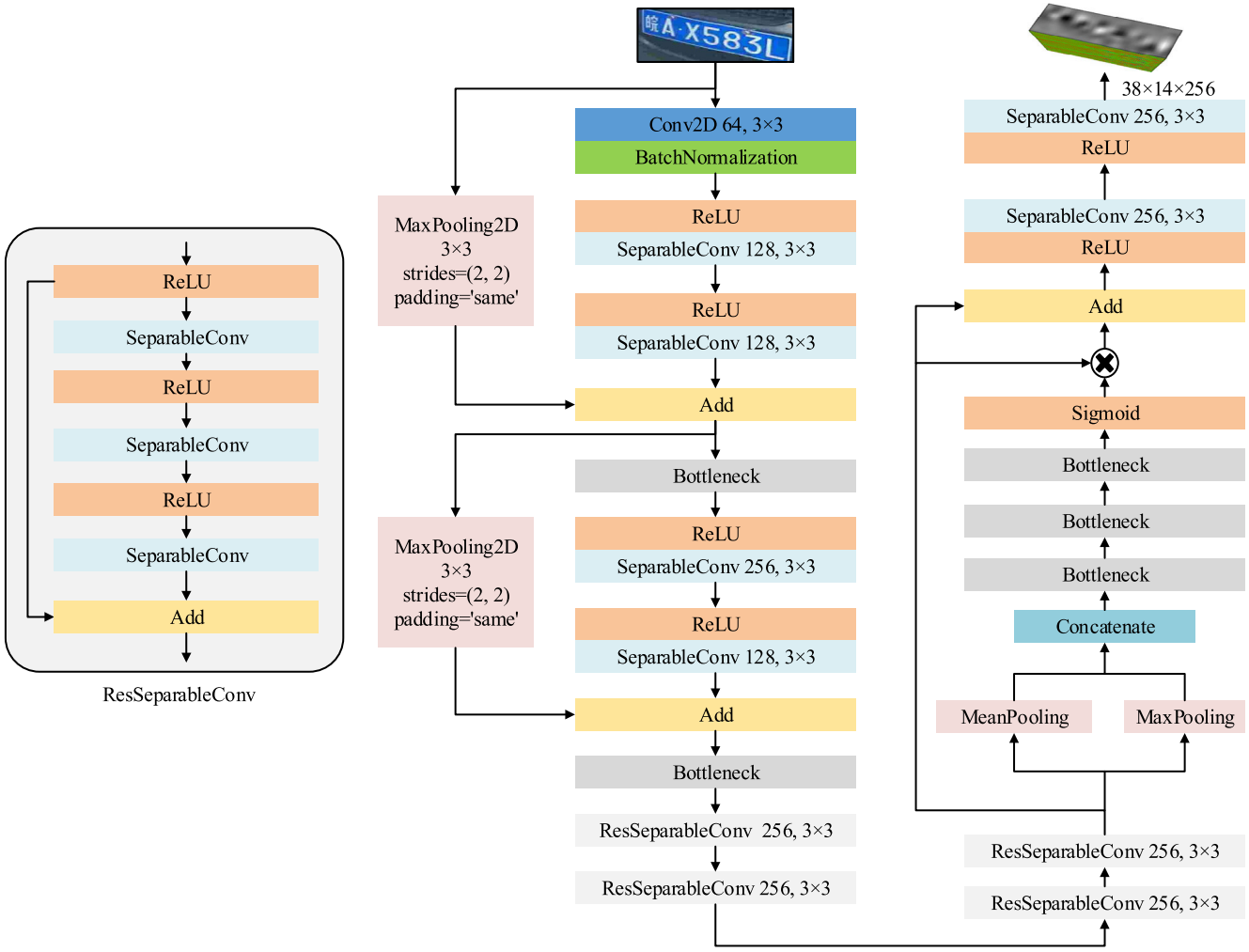



**FIGURE 3.** License plate feature extraction network.

of 'A' character value ranged between [0,1] corresponding to Y2. The localization methods of other characters are the same. One-way RNN infers information based on the previous information, but sometimes it is not enough to just focus on the information of the front characters. When the spacing information between the license plate characters cannot be accurately obtained, the license plate characters cannot be accurately located. The Bi-LSTM network is used to localize the license plate character features. The hidden layer of the bi-directional convolutional neural network has to store two values, A is involved in the forward computation and $A'$ is involved in the reverse computation. The final output value concatenates the hidden vectors of forward A and reverse $A'$ to obtain $O_t$. In forward computation, the $A_t$ of the hidden layer is related to $A_{t-1}$; in reverse computation, the $A_t$ of the hidden layer is relevant to $A_{t+1}$. The calculation formula is as follows.

$$O_t = g\left(VA_t + V'A'_t\right) \tag{3}$$
$$A_t = f\left(Ux_t + WA_{t-1}\right) \tag{4}$$
$$A'_t = f\left(U'x_t + W'A'_{t+1}\right) \tag{5}$$

### D. CHARACTER FEATURE EXTRACTION

The process of feature extraction of license plate characters is shown in Figure 5. Taking the character 'A' corresponded to Y2 position as an example, we obtain the 1D-Attention weight (38 × 1) of character 'A' by Bi-LSTM, which is multiplied by the output feature map (38 × 14 × 256) matrix of the license plate feature extraction network. In that way, the features related to character 'A' in the map can be enhanced and the features unrelated to character 'A' can be suppressed, and the features of the character 'A' in license plates are effectively extracted.

## IV. EXPERIMENTAL SETTING

### A. DATASETS

CCPD [4] is currently the largest public dataset that provides over 290k unique license plate images with detailed annotations. The data images have only one kind of license plates and each license plate consists of one Chinese character, one letter and five letters or numbers. The dataset is grouped by difficulty of recognition, illumination of the license plate regions, distance from license plates at the time

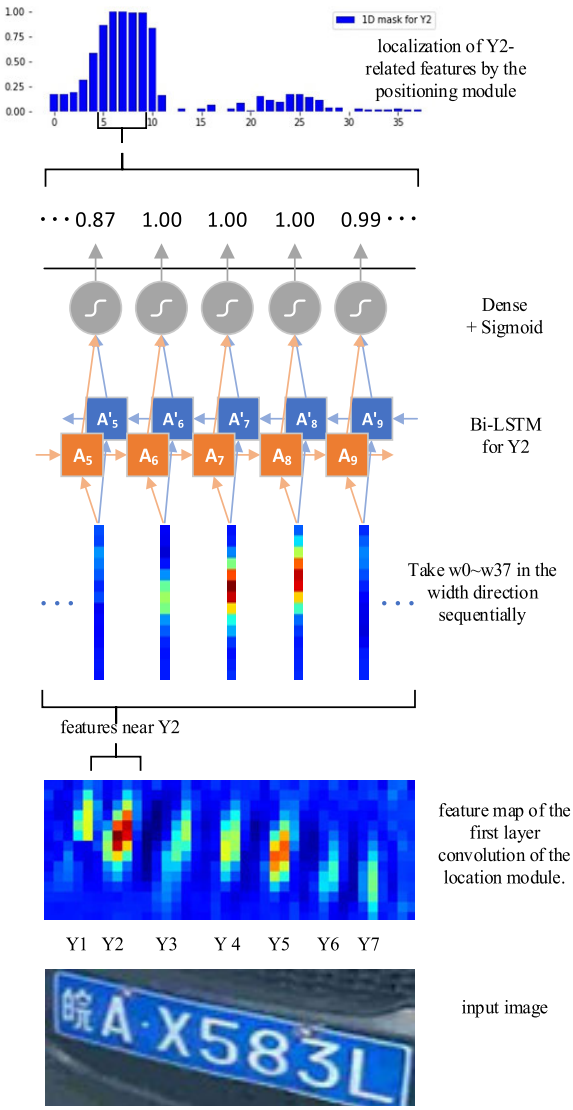



**FIGURE 4.** License plate character localization module.

of capture, degree of horizontal and vertical tilt, and weather (rain, snow, fog). It contains 9 data subsets: CCPD-base (200k), CCPD-db (20k), CCPD-fn (20k), CCPD-rotate (10k), CCPD-tilt (10k), CCPD-weather (10k), CCPD-challenge (10k), etc. The CCPD-base consists of about 200k images, half of which are used for training and the other half are for testing. The remaining subsets are all for testing.

PKUData [5] was released by Yuan *et al.* It provides license plate datasets and license plate detection images under different conditions and scenarios. However, they contain no license plate labels, so we labeled 2253 images in data subsets G1 (normal environment during daytime), G2 (daytime with sunshine glare), and G3 (nighttime), all of which are used to evaluate the generalization ability of the proposed method.

The CLPD [6] contains 1200 license plate images of different vehicle types from 31 provinces in mainland China, which are all real and covering various shooting conditions such as camera angle, time, resolution and background. However,

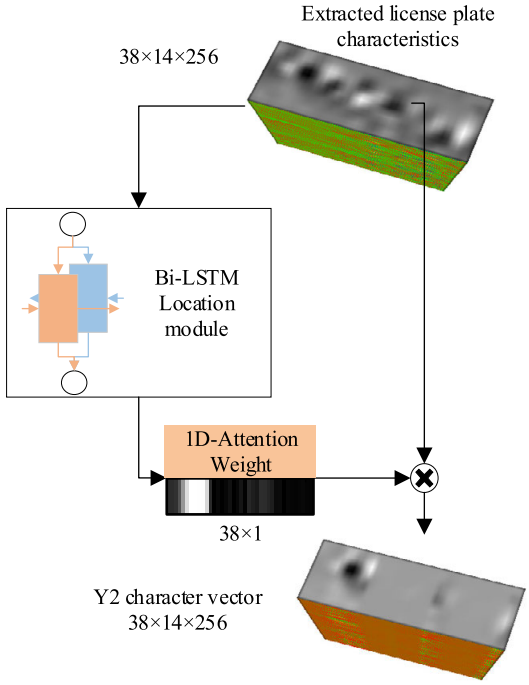



**FIGURE 5.** Character feature extraction module.

they are only used to evaluate the reliability of the proposed license plate recognition model.

AOLP [7] contains a total of 2019 images of Taiwan license plates, including three data subsets: Access Control (AC), Traffic Law Enforcement (LE), and Road Patrol (RP). AC contains 681 images, LE 757 images, and RP 611 images. 4/5 of the images are randomly selected as the training set and 1/5 as the test set. The dataset comparison is illustrated in Table 2.

### B. NETWORK TRAINING

License plate detection: The input image was resized to 736 × 736, and the Adam optimizer was adopted to train the detection network for 128 epochs. In the first stage, the initial batch size was set to 16, the learning rate to 1e-3, and the learning rate was multiplied by 0.1 every 3 epochs without reducing the loss. In the second stage, the initial learning rate was set to 1e-4, the learning rate strategy was the same as the first stage, and the batch size was set to 4.

License plate recognition: The input image was resized to 152 × 56, and the recognition network was trained for 200 epochs using category cross-entropy loss and Adam optimizer. During the training process, the initial batch size was set to 32, and the learning rate was 1e-4. When the loss did not decrease, the learning rate was multiplied by 0.5 every 3 epochs. The category cross entropy loss equation is as (6) demonstrates, where C represents the number of categories, $y_j$ denotes the true value, and $y_{predict_j}$ means the predicted value.

$$loss = -\sum_{j}^{C} y_j \log\left(y_{predict\,j}\right) \tag{6}$$

**TABLE 2. Comparison of license plate datasets.**

| | AOLP | PKUdata | CCPD | CLPD |
|---|---|---|---|---|
| Year | 2012 | 2016 | 2018 | 2019 |
| Number of images | 2049 | 2253 | 290k | 1200 |
| Chinese characters | 0 | 23 | 29 | 31 |
| Vehicle distance | Near | Far | Near Medium Far | Near Medium Far |
| LP size | 72×28 | 156×39 | 253×100 | 149×48 |
| LP angle | Frontal + oblique | Mostly frontal | Frontal + oblique | Frontal + oblique |

**TABLE 3. Comparison of the average precision (percentage) of license plate detection in different subsets when IoU = 0.7, where AP represents the average accuracy of the entire dataset.**

| Subset<br>Methods | AP | Base<br>(100k) | DB<br>(20k) | FN<br>(20k) | Rotate<br>(10k) | Tilt<br>（10k） | Weather<br>(10k) | Challenge<br>(10k) |
|---|---|---|---|---|---|---|---|---|
| Cascade classifier [31] | 47.2 | 55.4 | 49.2 | 52.7 | 0.4 | 0.6 | 51.5 | 27.5 |
| SSD300 [13] | 94.4 | 99.1 | 89.2 | 84.7 | **95.6** | **94.9** | 83.4 | 93.1 |
| YOLO9000 [15] | 93.1 | 98.8 | 89.6 | 77.3 | 93.3 | 91.8 | 84.2 | 88.6 |
| Faster-RCNN [26] | 92.9 | 98.1 | 92.1 | 83.7 | 91.8 | 89.4 | 81.8 | 83.9 |
| TE2E [21] | 94.2 | 98.5 | 91.7 | 83.8 | 95.1 | 94.5 | 83.6 | **93.1** |
| RPnet [4] | 94.5 | **99.3** | 89.5 | 85.3 | 94.7 | 93.2 | 84.1 | 92.8 |
| YOLOv4 | 95.1 | 96.8 | 93.7 | 93.1 | 93.5 | 94.7 | 96.6 | 85.5 |
| **YOLOv3** | **96.0** | 97.1 | **97.2** | **93.3** | 91.6 | 94.6 | **97.9** | 90.5 |

### C. EVALUATION METRIC

For fair comparison, the same evaluation criteria as [4] was used on the CCPD dataset. If and only if the IoU between the detected results and ground truth was greater than 0.6, and all the characters on the license plate (including the Chinese characters) are correctly recognized, the license plate recognition results were correct. As defined in formula (7), db denotes the detection box, gb represents the ground truth box, and area denotes the area. All experiments were performed on NVIDIA TITAN GPU with 12GB of GPU memory.

$$iou = \frac{area\,(db \cap gb)}{area\,(db \cup gb)} \tag{7}$$

## V. RESULTS

### A. EXPERIMENTS ON CCPD DATASET

We tested seven data subsets: CCPD-Base, CCPD-DB, CCPD-FN, CCPD-Rotate, CCPD-Tilt, CCPD-Weather, and CCPD-Challenge. AP represents the average accuracy on the whole test dataset. It can be observed from Table 3 that the average accuracy of YOLOv3 algorithm is incremented by 1.5% on the overall CCPD test set, and the average accuracy of the subsets CCPD-DB, CCPD-FN, and CCPD-Weather are increased by 5.1%, 8% and 13.8% respectively. From the last two rows, it is found that YOLOv3 is used as license plate detection instead of YOLOv4, and its accuracy is higher than YOLOv4.

The results of the license plate recognition on CCPD can be seen from Table 4. The proposed algorithm outperforms other algorithms on both Overall and most subsets. The only exception is that the method of Zherzdev *et al.* [22] is better than ours on Rotate subset. The reason might be that Zherzdev *et al.* [22] preprocessed input images via a Spatial Transformer Layer [1], which is specifically designed for correction of rotated images. The comparison demonstrates that the other methods do not perform well on Rotate, and there might be some character features which are not sufficiently extracted in license plates, resulting in inaccurate character recognition. For Rotate dataset, our algorithm can also benefit from the inclusion of this technique, which is expected to further improve the accuracy of the proposed algorithm on Rotate dataset.

In irregular license plate image subsets such as CCPD-Weather and CCPD-Challenge, the proposed algorithm has significant advantages compared to other algorithms. The accuracy of the CCPD-Weather subset reaches 99.3%, and the accuracy of the CCPD-Challenge subset 86.6%, achieving an increase of 3.9% and 3.5% respectively compared with other best results. It can be proved that our model has better robustness towards deformed license plates. In other subsets CCPD-Base (100k), CCPD-DB, and CCPD-FN, the accuracy rates are 99.3%, 98.5% and 98.6%, achieving an improvement of 0.2%, 1.6% and 1.3 respectively. The accuracy on all subsets is 97.8%, which has been lifted by 1.2%.

The heat maps of some extremely challenging license plate (dark, rotated, snowy, light, tilted, challenged) characters are visualized, which allows us to observe the position of the characters localized by the network, and extract the corresponding features for recognition. As shown in Figure 6, each

**TABLE 4.** Comparison of the license plate recognition accuracy in different subsets when IoU = 0.6 (percentage).

| Subset<br>Methods | Overall | Base<br>(100k) | DB<br>(20k) | FN<br>(20k) | Rotate<br>(10k) | Tilt<br>（10k） | Weather<br>(10k) | Challenge<br>(10k) |
|---|---|---|---|---|---|---|---|---|
| Ren et al. (2015) [26] | 92.8 | 97.2 | 94.4 | 90.9 | 82.9 | 87.3 | 85.5 | 76.3 |
| Liu et al. (2016) [13] | 95.2 | 98.3 | 96.6 | 95.9 | 88.4 | 91.5 | 87.3 | 83.8 |
| Joseph et al. (2016) [15] | 93.7 | 98.1 | 96.0 | 88.2 | 84.5 | 88.5 | 87.0 | 80.5 |
| Li et al. (2017) [21] | 94.4 | 97.8 | 94.8 | 94.5 | 87.9 | 92.1 | 86.8 | 81.2 |
| Zherzdev et al. (2018) [22] | 93.0 | 97.8 | 92.2 | 91.9 | 79.4 | 85.8 | 92.0 | 69.8 |
| Xu et al. (2018) [4] | 95.5 | 98.5 | 96.9 | 94.3 | 90.8 | 92.5 | 87.9 | 85.1 |
| Zhang et al. (2019) [27], [22] | 93.0 | 99.1 | 96.3 | 97.3 | **95.1** | 96.4 | 97.1 | 83.2 |
| Wang et al. (2020) [28] | 96.6 | 98.9 | 96.1 | 96.4 | 91.9 | 93.7 | 95.4 | 83.1 |
| **Ours (Real Data Only)** | **97.8** | **99.3** | **98.5** | **98.6** | 92.5 | **96.4** | **99.3** | **86.6** |

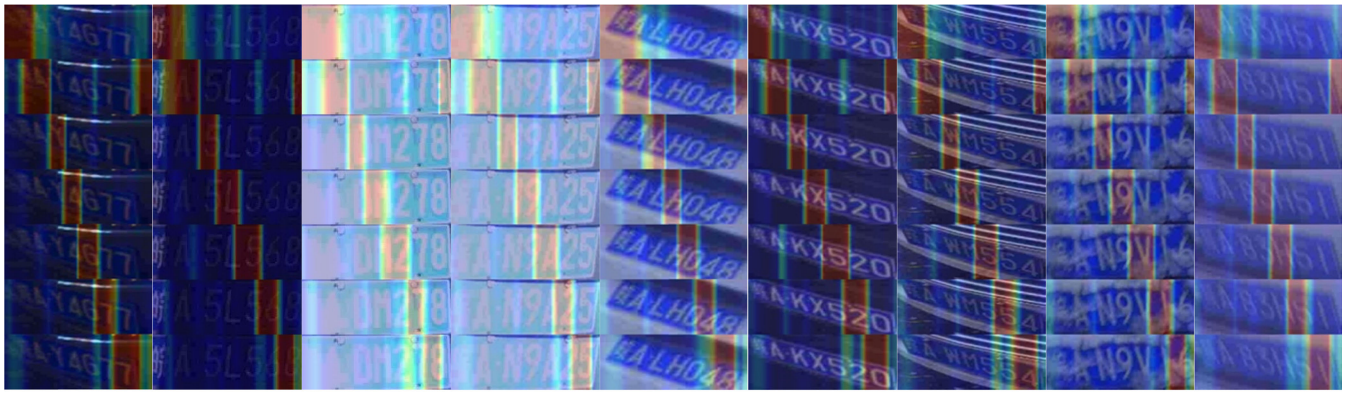



**FIGURE 6.** Heat maps of character locations in CCPD datasets.

column is composed of seven license plates. The darker the color of the characters on each license plate position indicates that the network focuses more on the character features at this position and extracts these features for recognition. For example, in the license plate ''皖AY4G77'' in Figure 6, '皖' character is darker in color in the first row, meaning that the network is concerned more with the features in position of '皖' character. The darker color of the 'A' character in the second row denotes that the network pays more attention on the features on the location of the 'A' character. In the same way, the network is able to locate the features on all the character locations on the license plate, and therefore can accurately identify characters in license plates.

Table 5, 7 and 9 compares the accuracy of ground truth and YOLOv3 for license plate detection. It can be observed that the accuracy when adopting YOLOv3 as a detector is close to the accuracy of true license plate labels, regardless of whether Chinese characters are included or not.

### B. EXPERIMENTS ON PKUdata DATASET

The model trained on the above CCPD-Base dataset was used to test PKUdata. The results of license plate recognition are compared in two cases of including and excluding Chinese characters, as shown in Table 6. In the case where Chinese characters are included, the proposed model outperforms the [6] model by 8.3%. By Excluding Chinese characters, the performance is improved by 6.1%. The PKUdata dataset has not been trained before, and the accuracy gain is even more pronounced (about 8%), which indicates that the generalization ability of our model is reliable, and the license plate recognition model owns robustness.

### C. EXPERIMENTS ON CLPD DATASET

On CLPD, the same approach as PKUdata above was chosen to test CLPD, using only the model trained on the CCPD-Base dataset. The experimental results in Table 6 reveal the advantages of the proposed model. The accuracy is the highest with or without the Chinese characters. The accuracy is enhanced by 3.9% when the Chinese characters is included and by 6.4% if the Chinese characters is excluded.

### D. EXPERIMENTS ON AOLP DATASET

For fair comparison, we did not use any synthetic data during model training. Rotating, offset, scaling and other methods were adopted to expand the training data set. The results in Table 8 show that the proposed method outperforms other methods on two subsets AC and RP, which reflects the superiority of our method. The accuracy is increased

**TABLE 5.** Comparison of the license plate recognition accuracy of different bounding boxes in different scenarios. YOLOv3 Detection indicates the license plate bounding boxes predicted by YOLOv3, ground truth indicates the real license plate bounding boxes, Without Chinese Characters ACC means the accuracy of license plate recognition without Chinese Characters, and ACC indicates the accuracy of correct recognition of all characters on license plates.

| Subsets | Bounding boxes | Without Chinese Characters ACC (%) | ACC (%) |
|---|---|---|---|
| Base (100k) | Yolov3 Detection | 99.5 | 99.3 |
| | **Ground truth** | **99.7** | **99.6** |
| DB (20k) | Yolov3 Detection | 98.8 | 98.5 |
| | **Ground truth** | **99.3** | **99.0** |
| FN (20) | Yolov3 Detection | 98.9 | 98.6 |
| | **Ground truth** | **99.3** | **99.1** |
| Rotate (10k) | Yolov3 Detection | 93.3 | 92.5 |
| | **Ground truth** | **98.2** | **97.6** |
| Tilt (10k) | Yolov3 Detection | 97.0 | 96.4 |
| | **Ground truth** | **98.7** | **98.3** |
| Weather (10k) | Yolov3 Detection | 98.1 | 97.6 |
| | **Ground truth** | **98.6** | **98.2** |
| Challenge (10k) | Yolov3 Detection | 88.2 | 86.6 |
| | **Ground truth** | **89.9** | **88.5** |

**TABLE 6.** Comparison of the recognition accuracy of license plates with and without Chinese Characters in the CLPD and PKUdata datasets.

| Methods | CLPD | | PKUdata | |
|---|---|---|---|---|
| | Without Chinese Characters ACC | ACC | Without Chinese Characters ACC | ACC |
| Masood et al. (2017) [29] | 85.2 | - | 89.3 | - |
| Xu et al. (2018) [4] | 78.9 | 66.5 | 78.4 | 77.6 |
| Zhang et al (Real Data Only) (2020) [6] | 86.1 | 70.8 | 86.5 | 84.8 |
| Zhang et al (Real + Synthetic Data) (2020) [6] | 87.6 | 76.8 | 90.5 | 88.2 |
| **Ours (Real Data Only)** | **94.0** | **80.7** | **96.6** | **96.5** |

**TABLE 7.** Comparison between the recognition accuracy of license plates with and without Chinese characters of different license plate bounding boxes in CLPD and PKUdata datasets.

| Bounding boxes | CLPD | | PKUdata | |
|---|---|---|---|---|
| | Without Chinese Characters ACC | ACC | Without Chinese Characters ACC | ACC |
| Yolo3 Detection | 94.0 | 80.7 | 96.6 | 96.5 |
| **Ground truth** | **94.2** | **81.2** | **96.7** | **96.6** |

**TABLE 8.** Comparison of the recognition accuracy of the license plate characters and the overall license plate in AOLP dataset.

| Methods | AC | | LE | | RP | |
|---|---|---|---|---|---|---|
| | Char (%) | Plate (%) | Char (%) | Plate (%) | Char (%) | Plate (%) |
| Li et al. [21] | - | 95.29 | - | 96.57 | - | 83.63 |
| Li et al. [19] | - | 94.85 | - | 94.19 | - | 88.38 |
| Hsu et al. [7] | 96 | 88.5 | 94 | 86.6 | 95 | 85.7 |
| DenseNet [30] | 99.08 | 96.61 | **99.65** | **97.8** | 97.22 | 91 |
| **Ours** | **99.51** | **97.06** | 98.39 | 96.55 | **98.9** | **93.39** |

by 0.45% on AC and 2.39% on RP. The accuracy of the license plate characters is increased by 0.43% on AC and 1.68% on RP. The RP subset is mainly composed of rotated license plates, and our method has the greatest improvement on performance on these license plates. This result proves the effectiveness of the proposed method in identifying

**TABLE 9.** Comparison of the recognition accuracy of the license plate characters and the overall license plate of different bounding boxes. YOLOv3 Detection represents the license plate bounding boxes predicted by YOLOv3, and Ground truth represents the real license plate bounding boxes.

| Bounding boxes | AC | | LE | | RP | |
|---|---|---|---|---|---|---|
| | Char (%) | Plate (%) | Char (%) | Plate (%) | Char (%) | Plate (%) |
| Yolo3 Detection | 99.51 | 97.06 | 98.39 | 96.55 | 98.90 | 93.39 |
| **Ground truth** | **99.88** | **99.27** | **99.68** | **98.71** | **99.04** | **95.08** |


**FIGURE 7.** Recognition results of license plates images from CCPD datasets.

irregular license plates. Figure 7 illustrates the experimental results.

## VI. CONCONCLUSION

This paper proposes a robust model based on license plate recognition in complex scenarios, which is mainly composed of three modules: license plate feature extraction, license plate character positioning, and character feature extraction. The license plate feature extraction module of this model is based on Xception, MobileNetV3, and spatial attention mechanism, which can fully extract the character features of license plates and establish a sound foundation for the license plate character localization. License plate character localization module adopts the method of Bi-LSTM combined with context location information of license plates. From heat maps of license plate positioning in this paper, it can be found that Bi-LSTM can accurately locate the characters of each license plate. Character feature extraction module uses the designed 1D-Attention to enhance the useful character features, and suppress useless character features after Bi-LSTM localization, which plays a key role in character recognition. The accuracy of character recognition directly relates to the quality of character extraction. Through the evaluation of multiple datasets, the experimental results demonstrate the superiority of our proposed method. Whether it is a regular or irregular license plate, or in normal or complex scenes, the method can accurately locate and recognize characters in each license plate. In future work, the GAN network might be leveraged to simulate license plates of various regions to compensate for the imbalance of Chinese characters categories in the dataset, which may help improve the accuracy. Secondly, we can feed back to the network according to the accuracy of localization of the character features in license plates, which may lead to better results.

## REFERENCES

[1] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, 2015, pp. 2017–2025.

[2] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," presented at the Int. Conf. Comput. Vis., 2019, doi: 10.1109/ICCV.2019.00140.

[3] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, doi: 10.1109/CVPR.2017.195.

[4] 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," presented at the Eur. Conf. Comput. Vis., 2018, doi: 10.1007/978-3-030-01261-8_16.

[5] 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.

[6] L. Zhang, P. Wang, H. Li, Z. Li, C. Shen, and Y. Zhang, "A robust attentional framework for license plate recognition in the wild," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 18, 2020, doi: 10.1109/TITS.2020.3000072.

[7] G.-S. Hsu, J.-C. Chen, and Y.-Z. Chung, "Application-oriented license plate recognition," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 552–561, Feb. 2013.

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

[9] J. Hsieh, S. Yu, and Y. Chen, "Morphology-based license plate detection from complex scenes," presented at the Int. Conf. Pattern Recognit., 2002, doi: 10.1109/ICPR.2002.1047823.

[10] 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, Jan. 2015.

[11] Z. Yao and W. Yi, "License plate detection based on multistage information fusion," *Inf. Fusion*, vol. 18, no. 1, pp. 5–78, Jul. 2014.

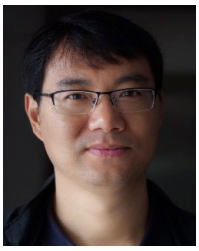
[12] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," presented at the Int. Conf. Comput. Vis., 2017, doi: 10.1109/ICCV.2017.322.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "SSD: Single shot multibox detector," presented at the Eur. Conf. Comput. Vis., 2016, doi: 10.1007/978-3-319-46448-0_2.

[14] J. Redmon, S. K. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, doi: 10.1109/CVPR.2016.91.

[15] R. J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," presented at the IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, doi: 10.1109/CVPR.2017.690.
[16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. [Online]. Available: http://arxiv.org/abs/2004.10934
[17] G. R. Gonçalves, S. P. G. da Silva, D. Menotti, and W. R. Schwartz, "Benchmark for license plate character segmentation," *J. Electron. Imag.*, vol. 25, no. 5, Oct. 2016, Art. no. 053034.
[18] S. M. Silva and C. R. Jung, "License plate detection and recognition in unconstrained scenarios," presented at the Eur. Conf. Comput. Vis., 2018, doi: 10.1007/978-3-030-01258-8_36.
[19] H. Li and C. Shen, "Reading car license plates using deep convolutional neural networks and LSTMs," 2016, *arXiv:1601.05610*. [Online]. Available: http://arxiv.org/abs/1601.05610
[20] C. Gou, K. Wang, Y. Yao, and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1096–1107, Apr. 2016.
[21] 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.
[22] S. Zherzdev and A. Gruzdev, "LPRNet: License plate recognition via deep neural networks," 2018, *arXiv:1806.10447*. [Online]. Available: http://arxiv.org/abs/1806.10447
[23] S. Duan, W. Hu, R. Li, W. Li, and S. Sun, "Attention enhanced ConvNet-RNN for Chinese vehicle license plate recognition," presented at the Pattern Recognit. Comput. Vis., 2018, doi: 10.1007/978-3-030-03335-4_36.
[24] 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.
[25] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
[26] 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.
[27] 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.
[28] T. Wang, Y. Zhu, L. Jin, C. Luo, X. Chen, Y. Wu, Q. Wang, and M. Cai, "Decoupled attention network for text recognition," presented at the Amer. Assoc. Artif. Intell., 2020, doi: 10.1609/aaai.v34i07.6903.
[29] S. Zain Masood, G. Shu, A. Dehghan, and E. G. Ortiz, "License plate detection and recognition using deeply learned convolutional neural networks," 2017, *arXiv:1703.07330*. [Online]. Available: http://arxiv.org/abs/1703.07330
[30] C. Wu, S. Xu, G. Song, and S. Zhang, "How many labeled license plates are needed," presented at the Chin. Conf. Pattern Recognit., 2018, doi: 10.1007/978-3-030-03341-5_28.
[31] 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.
[32] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: An attentional scene text recognizer with flexible rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2035–2048, Sep. 2019.
[33] Z. Cheng, F. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou, "Focusing attention: Towards accurate text recognition in natural images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5086–5094.
[34] C. Luo, L. Jin, and Z. Sun, "MORAN: A multi-object rectified attention network for scene text recognition," *Pattern Recognit.*, vol. 90, pp. 109–118, Jun. 2019.
[35] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: A simple and strong baseline for irregular text recognition," 2018, *arXiv:1811.00751*. [Online]. Available: http://arxiv.org/abs/1811.00751

**YONGJIE ZOU** received the B.S. degree in computer science and technology from Guizhou Education University, in 2018. Then, he majored in computer technology in the School of Computer Science and Technology, Guizhou University. His research includes machine learning, computer vision, as well as remote sensing image processing based on deep learning.



**YONGJUN ZHANG** received the master's and Ph.D. degrees in software engineering from Guizhou University, Guiyang, China, in 2010 and 2015, respectively. He is currently an Associate Professor with Guizhou University. From 2012 to 2015, he is a joint training doctoral student of Peking University and Guizhou University, and studying in the Key Laboratory of Integrated Microsystems, Shenzhen Graduate School, Peking University. His research interests mainly focus on the intelligence image algorithms of computer vision, such as scene target detection, extraction, tracking, recognition, and behavior analysis.

**JUN YAN** was born in Gaomi, Shandong, China, in October 1962. He received the Ph.D. degree from Dublin City University. He is the Founder and Chairman of Zhuhai Orbita Aerospace Science and Technology Company, Ltd. He is a Standing Director of Guangdong Western Returned Scholars Association, a part-time Professor of the Harbin Institute of Technology, and the Honorary President of the BigData College of Qingdao University of Science and Technology, and the Chief Designer of the "Zhuhai No. 1" satellite constellation which consists of 34 satellites. Since 2000, he has participated in and directed the research and development of S698 series of rad-hardened chips based on SPARC architecture, SIP cubic package modules, satellite spatial information platform, remote sensing satellite constellation, and satellite big data projects.

**XIAOXU JIANG** received the B.S. degree in environmental engineering from the Guilin University of Electronic Technology, in 2015. Then, he worked as an Artificial Intelligence Algorithm Engineer at ORBITA. His researches include machine learning, computer vision, as well as remote sensing image processing based on deep learning.

**TENGJIE HUANG** received the B.S. degree in computer science and technology from the Beijing Institute of Technology, Zhuhai, in 2018. Then, he worked as an Assistant Director of the Artificial Intelligence Research Institute, ORBITA. His research includes machine learning, computer vision, as well as remote sensing image processing based on deep learning.

**HAISHENG FAN** is currently a Ph.D. President with the Division of Satellite Big Data Solutions. Expert at GIS and RS software design and development, researcher on image processing methods, and manager of remote sensing application projects. One of the major pioneers of first domestic commercial RS service platform based on cloud-computing system. Designer and Executor of ORBITA satellite big data service platform which is under construction.

**ZHONGWEI CUI** received the master's degree in computer application technology from Guizhou University, Guiyang, in 2008, where he is currently pursuing the Ph.D. degree. He is currently an Associate Professor with the School of Mathematics and Big Data, Guizhou Education University, Guiyang, China, since December 2013. He has 11 years of teaching experience. His research interests include machine vision and wireless networks.