EILPR: Toward End-to-End Irregular License Plate Recognition Based on Automatic Perspective Alignment

Hui Xu, Xiang-Dong Zhou, Zhenghao Li, Liangchen Liu, Chaojie Li, and Yu Shi

***Abstract*— Automatic License plate recognition (ALPR) remains a challenging task in face of some difficulties such as multi-line character distribution and license plate (LP) deformation due to camera angles. Most existing ALPR methods either focus on single-line LP or perform horizontal multi-line LP detection and recognition with character-level annotations. In this paper, we propose a novel end-to-end irregular license plate recognition (EILPR) to detect and recognize the LP of multi-line text or arbitrary shooting angles, using only plate-level annotations for training. In EILPR, a coarse-to-fine strategy is adopted to extract the LP features accurately for sequence recognition. Firstly, a coarse rectangular box of the LP is located, along with the corresponding predicted LP class which is single-line or double-line. Then, considering the fact that a LP mainly generates perspective distortion in the image due to its rigid feature, we propose a new automatic perspective alignment network (APAN) to extract the fine LP features connecting the detection and recognition. For recognition, a location-aware 2D attention based recognition network is performed to recognize the multi-line and multinational LP based on the extracted features. Experiments on several datasets show that EILPR achieves the state-of-the-art performance, demonstrating the effectiveness of the proposed method.**

***Index Terms*— License plate detection and recognition, automatic perspective alignment, end-to-end training.**

I. Introduction

ALPR is of great significance to modern Intelligent Transportation System (ITS). With increasing demand of ALPR and increasing number of LP types, ALPR systems face greater challenges. LP images can be collected from surveillance cameras, high-speed snap cameras, cellphones, etc [1]. The plates in the images are likely to be distorted due to shooting angles, which directly affect the recognition of license number. Moreover, some complex LP types such as multi-line plates increase the challenges of the task. Since there are only single-line and double-line plates at present, the multi-line plate in this paper only refers to the double-line plate.

With the success of convolutional neural network (CNN)-based methods in text recognition, objection detection and recognition in the wild [2]–[4], CNN based approaches are widely used in ALPR systems, which contains character-level and sequence-level methods. Character segmentation based LPR methods [5]–[7] search for the location of each character and then classify them. Most multinational ALPR methods [8]–[10] are based on character detection/segmentation due to the differences in LP layouts among different countries. However, any incorrect character locating will result in the mis-recognition of the text, even with a strong recognizer. Character segmentation is unreliable to lighting, pose and noise in the image, which is easy to cause recognition failure. Sequence labelling based methods [2], [11]–[13] no longer require character segmentation, which improve the reliability of LPR. However, existing such methods almost only recognize the single-line LP taken from a frontal and horizontal angle. Cao *et al*. [14] attempted to recognize double-line LPs by the 1D sequence recognition method [15], cutting the feature map from the center line and concatenating the split feature maps in the horizontal direction. However, this method is implemented under the ideal condition that the LP is frontal and have no deformations. In fact, the LP images in real scenes are taken from arbitrary angles, so that the LP in the image may produces deformation. To detect and recognize the arbitrary-shaped LP in an end-to-end framework, a feature alignment module, which rectifies the deformed LP features into regular ones and connects the detection and recognition, is indispensable. However, the universal text alignment methods [16], [17], which are designed for various irregular text, cannot achieve good effectiveness in ALPR.

To overcome these problems, we propose a novel end-to-end irregular license plate recognition (EILPR) to detect and recognize the LP of multi-line text or arbitrary shooting angles. EILPR consists of these steps: LP detection, automatic perspective alignment of LP features and 2D attention based

Manuscript received February 1, 2021; revised May 26, 2021 and November 10, 2021; accepted November 17, 2021. Date of publication December 3, 2021; date of current version March 9, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61773325, Grant 61876154, and Grant 62106247; in part by the Natural Science Foundation of Fujian Province under Grant 2018J01574; and in part by the Chongqing Research Program of Technological Innovation and Application Demonstration under Grant cstc2019jscx-msxmX0424. The Associate Editor for this article was T.-H. Kim. *(Corresponding author: Xiang-Dong Zhou.)*

Hui Xu is with the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences (CAS), Chongqing 400714, China, and also with the Chongqing School, University of Chinese Academy of Sciences, Chongqing 400714, China (e-mail: xuhui@cigit.ac.cn).

Xiang-Dong Zhou, Zhenghao Li, and Yu Shi are with the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences (CAS), Chongqing 400714, China (e-mail: zhouxiangdong@cigit.ac.cn; lizh@cigit.ac.cn; shiyu@cigit.ac.cn).

Liangchen Liu is with the School of Computing and Information Systems, The University of Melbourne, Parkville, VIC 3010, Australia (e-mail: liangchen.liu@unimelb.edu.au).

Chaojie Li is with the School of Electrical Engineering and Telecommunications, UNSW Sydney, Kingsford, NSW 2032, Australia (e-mail: chaojie.li@unsw.edu.au).

Digital Object Identifier 10.1109/TITS.2021.3130898

LP recognition, where the first two modules are used as a coarse-to-fine process to extract LP features. LPR should be performed based on the vehicle detection, so we take the cropped vehicle images as the input. Firstly, the LP detection network predicts a coarse rectangle box and the corresponding LP class (single-line or double-line) for each LP. Based on the detection output, APAN is proposed to automatically calculate the aligned LP features, which is the key to make EILPR end-to-end trainable. Finally, the aligned features are fed into the 2D attention based recognizer to predict the LP number.

An early version of this work appeared in the conference paper [18], which is only a recognition method based on detected plate images. We extend it in numerous ways, (i) exploiting a novel end-to-end LPR framework based on cropped vehicle images, which includes both detection and recognition, (ii) developing an alignment layer to connect the detection and recognition for end-to-end training, (iii) and verifying the performance on multinational LP benchmarks. Hence the main contributions of the paper are summarized as follows:

1) An end-to-end license plate detection and recognition method for irregular plates such as multi-line plates and plates with arbitrary shooting angles is introduced.
2) Considering the characteristics of the license plate itself, a feature alignment layer APAN is proposed to calculate the aligned feature and connect the detection and recognition into a unified framework.
3) APAN is verified to be more effective than TPS-based alignment. Meanwhile, EILPR has achieved state-of-the-art performance on several license plate benchmarks.

The rest of the paper is structured as follows: Section II briefly introduces the related works of ALPR, which includes two-stage and end-to-end approaches. In Section III, we depict the proposed method in details, including the LP detection network, APAN and the 2d attention based recognizer. The Experimental evaluation is shown in Section IV, while the conclusion is in Section V.

II. Related Work

With the increasing development of CNN and recurrent neural network (RNN) [19], the main methods can be divided into two kinds: two-stage and end-to-end.

A. Two-Stage LPR Methods

The two-stage LPR methods cascade the detection and recognition models which are trained separately. Firstly, the license plate is located by the detection model, and then the recognition model extracts its feature maps to predict the sequence number. LP detection usually uses the object detection model, such as Faster-RCNN [20], YOLO [21] and so on [22], and LP recognition can be classified into two groups: bottom-up and top-down approaches.

Bottom-up approaches segment characters firstly and then classify each character. Chao *et al*. [23] presents a LPR method based on character-specific extremal regions (ERs) and hybrid discriminative restricted Boltzmann machines (HDRBMs). Hsu *et al*. Jain *et al*. [24] use a CNN classifier trained for individual characters along with a Spatial Transformer Network [25] for character recognition. In these approaches, accurate character segmentation plays a crucial role in the system, which however is unreliable for images taken in the wild.

Top-down approaches regard LPR as a sequence labelling problem. Thanks to the improvement of text recognition, segmentation-free methods of LPR have achieved great development. Li and Shen [26] extracted CNN features and employed RNNs with Connectionist Temporal Classification (CTC) to predict the sequential labels. Cheang *et al*. [27] propose a unified ConvNet-RNN model to recognize real-world captured license plate, using a Convolutional Neural Network (ConvNet) to perform feature extraction and a RNN for sequence recognition. Some arbitrary-shaped scene text recognition algorithms [16] can also be used for ALPR. However, the existing arbitrary-shaped text rectification networks are not designed specifically for LPR, which did not take the rigid attribute of LPs into account.

B. End-to-End LPR Methods

The end-to-end LPR methods unify the LP detection and recognition into an overall structure. Montazzolli and Jung [28] proposed an end-to-end DL-ALPR system for Brazilian license plates based on Convolutional Neural Network architectures. Laroca *et al.* [5] employed two CNN networks for character segmentation and recognition of Brazil LP with fixed 7 characters based on a yolo detector. Hui *et al.* [29] adopted a region proposal network for detection and Bidirectional RNNs (BRNNs) to capture the context information in both sides for recognition. HyperLPR [30] is an open source Chinese license plate detection and recognition framework with high speed. The framework use a mixture of deep neural networks and classic image processing algorithms to perform detection, segmentation and recognition. RPNet [31] is an end-to-end LP recognition model that first released the CCPD dataset. However, these end-to-end LPR methods can only handle LPs which are horizontal or small deformation. SLPNet [32] is proposed to detect and recognize the deformed LPs, which is a CNN-based lightweight segmentation-free ALPR. SEE-LPR [33] is an end-to-end multinational LPR based on semantic segmentation. Nevertheless, neither SLPNet nor SEE-LPR recognizes multi-line LPs.

To summarize, the end-to-end training framework developed recently performs well in joint optimition both detection and recognition. The CTC and 1D attention based methods [16], which perform well in single-line LP recognition, cannot handle the multi-line LP such as the double-line. And the multi-line LP recognition method [14] can only recognize the LP that have no deformations. In this paper, we propose a novel end-to-end LPR method for irregular plates based on automatic perspective alignment, which is verified to achieve the state-of-the-art performance.

III. Methodology

As depicted in Fig. 1, the proposed EILPR contains three modules: LP detection network, APAN and 2D attention based

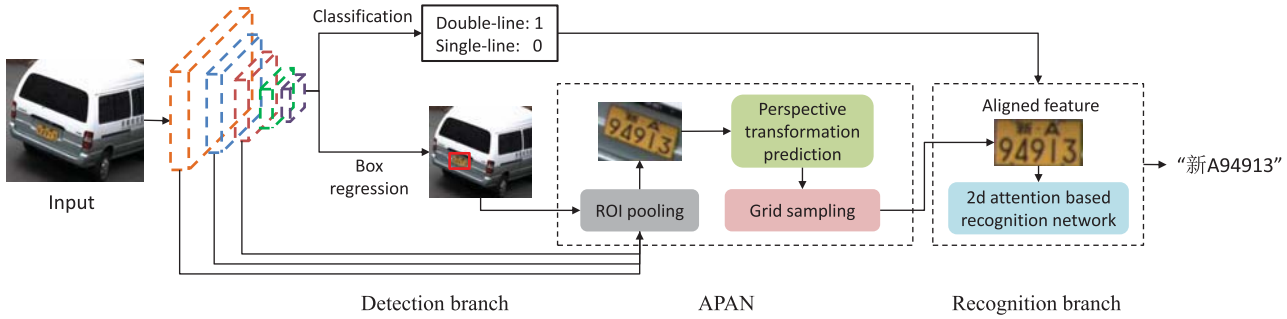


Fig. 1. Overall structure of our EILPR. The model is end-to-end trainable, which consists of a detection module, a feature alignment module and a 2d attention based recognition module. Given an input vehicle image, EILPR predicts the coarse bounding box and the LP number at the same time. The detection module has two task: bounding box regression and LP types classification. Depending on the coarse location of LP, the ROI pooling layer is used to obtain the ROI feature. Then, the automatic perspective alignment module is exploited to extract the LP feature for recognition. In the recognition network, the aligned LP feature is combined with the classification feature for sequence decoding. Finally, the LP number is predicted by the proposed 2D attention decoder.

recognition network. The detection module predicts a coarse rectangular box location and a LP class which is single-line or double-line. With the predicted bounding box, APAN firstly extracts the coarse ROI features from shared feature maps, and then calculates the fine LP features following the coarse-to-fine strategy as shown in Fig. 4. Finally, the aligned LP features are fed into a recognition network based on location-aware 2D attention mechanism to predict the license number.

A. LP Detection Network

The detection process is illustrated as the detection module in Fig. 1, which has two tasks: plate bounding box regression and plate classification.

1) Ground-Truth Generation: It is challenging to recognize the plate with serious perspective deformation and the multi-line text in the image. For the plate with serious perspective deformation, some methods [32], [34] detect it by locating the four corners. However, the corner regression loss function either L1 loss or IoU loss is not conducive to the plate recognition task as shown in Fig. 2(a). We can see that the green quadrilateral box has the similar detection loss to the blue one, but fail to be recognized due to the lack of part of text. Therefore, it is unwise to eliminate the influence of perspective deformation in the detection module.

As shown in Fig. 2(b), to ensure that the text can be fully included, the ground-truth box extends a certain margin based on the annotated box. The margin d_x, d_y are set to one-tenth of the width and the height respectively in the paper. The ground truth of detection module is expressed as $B = \{x, y, w, h, c\}$, where x, y, w, h represent the x and y coordinates of the center point, the width and the height of the bounding box respectively, and c is the plate class (single-line/double-line).

2) Network and Loss: The feed-forwarding result of the CNN-based detector is a 6-channel feature vector that include 2-channel single-line/double-line probabilities and 4-channel rectangle box parameters. The architecture of detection module has a total of 12 convolutional layers with ReLU and Batch Normalization as shown in Fig. 3. There are 4 max pooling layers of size 2×2 and stride 2. Finally, the detection block has two parallel convolutional layers: (i) one for inferring the probability, activated by a softmax function, and (ii) another for regressing the rectangle box parameters, without activation. In addition, the detection process described here is based on the detected vehicle image.

Fig. 2. (a) Description of corner regression errors. The red quadrilateral box represents the ground truth, while the blue and green quadrilateral boxes represent different detection and recognition results. (b) The blue rectangle box is the ground truth in this paper, while the red one is the minimum bounding box of the LP. d is the margin around ground truth box.

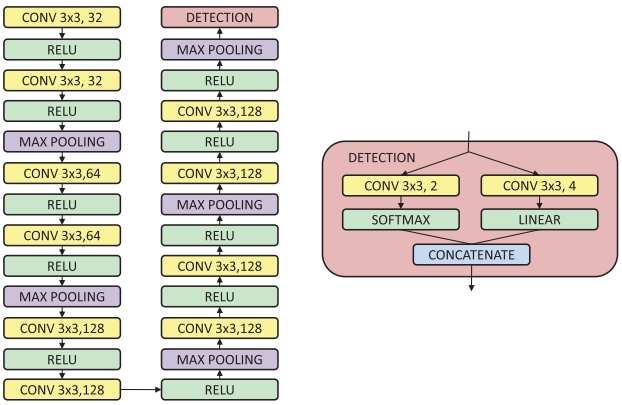


Fig. 3. Detailed network architecture of the detection module.

The training objective of detection network can be divided into two parts: the localization loss L_{loc} and the classification loss L_{cls}. Let N be the size of a mini-batch in training. L_{loc} is a Smooth L1 loss [35] between the predicted box (pb) and the ground truth box (gt), and L_{cls} is a cross-entropy loss. The

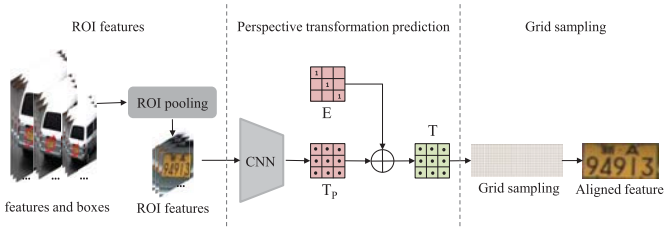


Fig. 4. Detailed architecture of APAN. We show the process on the LP images for better visualization, but actually the operation is on the feature maps.

detection loss function L_{det} is described as follows,

$$L_{loc} = \frac{1}{N}\sum_{N}\sum_{i\in\{x,y,w,h\}} smooth_{L1}(pb^i - gt^i) \quad (1)$$

$$L_{cls} = \frac{1}{N}\sum_{N} -[y \cdot log(p) + (1-y)\cdot log(1-p)] \quad (2)$$

$$L_{det} = L_{loc} + L_{cls} \quad (3)$$

where p denotes the prediction for the single-line LP while $(1-p)$ denotes the double-line LP. y represents the ground truth class, where single-line is 1 and double-line is 0.

B. Automatic Perspective Alignment Network (APAN)

The coarse location of the LP in the image is predicted by LP detection module. According to the coarse-to-fine strategy, APAN is developed to calculate fine LP features based on extracted coarse features as shown in Fig. 4. In APAN, we exploits region-of-interest (ROI) pooling layers [35] to extract coarse features of interest. Then, the perspective transformation prediction network and grid sampling are performed to calculate fine LP feature maps for recognition.

1) RoI Features: We know that feature maps from different layers within a network have different receptive field sizes, where the lower layers capture more fine details of the input objects. Because the area of the LP is expected to be very small relative to the entire image, feature maps from relatively lower layers are matter for recognizing LP. Inspired by [31], EILPR extracts feature maps at the end of three low-level layers: the second, fourth, sixth convolutional layer. The sizes of extracted feature maps are $(256\times ch)\times(256\times cw)\times 32$, $(128\times ch)\times(128\times cw)\times 64$, $(64\times ch)\times(64\times cw)\times 128$. After these feature maps are extracted, ROI pooling layers are used to extract each ROI from shared feature maps and convert it into a feature map with a fixed size $P_w \times P_h$(e.g.,16×50 in this paper). Afterwards, these three resized feature blocks $16\times 50\times 32$, $16\times 50\times 64$, $16\times 50\times 128$ are concatenated to one feature block of size $16\times 50\times 224$ for feature alignment.

2) Perspective Transformation Prediction: Thin-Plate-Spline [36] (TPS) transformation are mostly used for irregular text alignment as in [16], [37], which can handle various text deformations. TPS-based alignment method consists of 3 steps: control points prediction, a TPS transformation calculating and grid sampling. This kind method results in more complex deformation at higher degree of freedom, which is a burden to LPR. In comparison to TPS-based alignment, APAN introduces the prior information that the deformation of LPs is rigid deformation. It predicts perspective transformation parameters directly, which enhances geometric constraints of LP image according to the rigid body property of the LP. And control points are no longer needed.

TABLE I
ARCHITECTURE OF PERSPECTIVE TRANSFORMATION PREDICTION NETWORK

Layer Name	Configurations	Size
Input	-	$224\times 16\times 50$
Convolution	c:256,k:3 × 3, s:1 × 1, pad:1	$256\times 16\times 50$
MaxPooling	k:2 × 2,s:2 × 2	$256\times 8\times 25$
Convolution	c:256,k:3 × 3, s:1 × 1, pad:1	$256\times 8\times 25$
MaxPooling	k:2 × 2,s:2 × 2	$256\times 4\times 12$
Convolution	c:512,k:3 × 3, s:1 × 1, pad:1	$512\times 4\times 12$
AvgPooling	k:2 × 2,s:2 × 2	$512\times 1\times 1$
fc1	256	256
fc2	9	9

c,k,s,pad represent channel,kernel, stride and padding sizes respectively.

As shown in Fig. 4, the perspective transformation prediction network directly predicts a 3-by-3 offset matrix $T_p^{3\times 3}$. The architecture of the prediction network is showed in Table. I.

The network architecture consists of three convolutional layers. Each convolutional layer with a kernel of 3-by-3 is followed by a batch normalization layer, a ReLU layer and a pooling layer. With the input feature map I of size 16×50, the network predicts the offsets with 9 channels which corresponds to $T_p^{3\times 3}$. The rough perspective transformation $\hat{T}$ is combined by the two matrices as follows,

$$\hat{T} = T_p + E \quad (4)$$

where the unit matrix $E^{3\times 3}$ is used to keep the identity mapping when the predicted T_p is zero. The final perspective transformation T is normalized by $\hat{T}(3,3)$. $T^{3\times 3}$ is actually a 3D transformation, establishing the location relationship between the corresponding pixels of the aligned feature map I_r and the ROI feature map I.

We know that each member of the matrix T_p have a different meaning and different magnitude, so we do not design losses separately for APAN. Because all modules in APAN are differentiable, we combine detection network, APAN and 2D attention based recognition network for end-to-end training, requiring no manual annotations on the perspective transformation matrix. Inspired by STN which makes the spatial transformation learnable end-to-end, $T^{3\times 3}$ will be used to generate sampling grid for I_r.

3) Grid Sampling: A sampling grid $P = \{p_i \mid p_i \in R^{2\times 1}\}$ is built on I, applying T to each pixel position on I_r. To fulfill this, we have to complete the calculation on account of the 3D property of perspective transformation. Given a 3D point $\theta = [u, v, w]^T$ as an input, the output point $\theta' = [x, y, z]^T$ will be calculated by perspective projective mapping as follows,

$$\left[x', y', z'\right]^T = T \times \theta \quad (5)$$

$$x = \frac{x'}{z'}, \quad y = \frac{y'}{z'}, \quad z = \frac{z'}{z'} = 1 \quad (6)$$

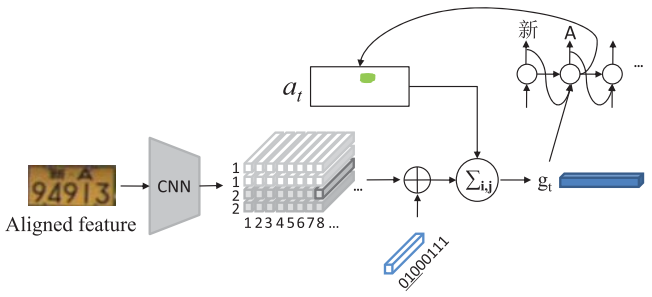


Fig. 5. Detailed network architecture of 2d attention based recognition module.

TABLE II
ARCHITECTURE OF THE ENCODER OF THE 2D ATTENTION RECOGNITION NETWORK

Layer Name	Configurations	Size
Input	-	$224 \times 16 \times 50$
Convolution	c:256,k:3 × 3, s:1 × 1, pad:1	$256 \times 16 \times 50$
Convolution	c:256,k:3 × 3, s:1 × 1, pad:1	$256 \times 16 \times 50$
MaxPooling	k:2 × 2,s:2 × 2	$256 \times 8 \times 25$
Convolution	c:512,k:3 × 3, s:1 × 1, pad:1	$256 \times 8 \times 25$
Convolution	c:512,k:3 × 3, s:1 × 1, pad:1	$256 \times 8 \times 25$
MaxPooling	k:2 × 2,s:2 × 2	$512 \times 4 \times 13$

c,k,s,pad represent channel,kernel, stride and padding sizes respectively.

where the vector $[x', y', z']^T$ is an intermediate variable. Pixel location θ indicates the pixel on I_r. u, v in the vector θ are the x-coordinate and y-coordinate of the pixel respectively, and w is a constant that is usually set to 1. Then, the 2D vector $p_i = [x, y]^T$ is the location of sampling point on I. The aligned features are extracted from the input feature maps with the set of sampling points, in which the interpolation method is bilinear interpolation. I_r is the same size as I and is generated as follows,

$$I_r = BilinearInterpolation(P, I) \tag{7}$$

Finally, the aligned feature map I_r is passed into the 2D attention based recognition network.

C. 2D Attention Based Recognition Network

The current segmentation-free ALPR methods, which extracted 1D sequence feature, can only recognize the single-line LP. By contrast, 2D attention based approaches, which have achieve good performance in scene text recognition [38], [39], can perform decoding in 2D space.

To predict a license number sequence directly from the 2D feature map extracted by CNN, we adopt 2D attention mechanism inspired by [40]. As depicted in Fig. 5, the recognition module is an attention-based encoder-decoder model. The encoder is a CNN structure, while the decoder is attention-based. In this paper, the encoder is a light-weight network shown in Table. II. With the input size of 16×50, the encoder outputs 2D feature maps with the size of 4×13. The feature maps are updated according to the plate classes, and then are fed into an attention based GRU (256 units) network.

We use $h = \{h_{i,j,c}\}$ to denote the feature map extracted by the CNN encoder, where i, j, c indicate the positions and channels in the feature map. Then, a gated recurrent unit (GRU) [41] is used to convert the feature maps into a character sequence $(o_1, o_2, \cdots, o_L)$, where L is the length of the predicted license number. At time step t, the final predicted distribution over characters is computed by

$$o_t = Softmax(W_o s_t + b_o) \tag{8}$$

where s_t is the hidden state of the GRU at time step t. To get s_t, we should generate the embedding vector $\hat{o}_{t-1} = Embedding(o_{t-1})$, which denotes the ground truth of the previous time step during training or the prediction of the previous time step during testing. The hidden state s_t is updated as,

$$s_t = GRU(Concat(\hat{o}_{t-1}, g_t), s_{t-1}) \tag{9}$$

where the glimpse vector $g_t = \sum_{i,j} \alpha_{t,i,j} h_{i,j,c}$ combines the image features and the spatial attention mask. $\alpha_t = \{\alpha_{t,i,j}\}$ denotes the spatial attention map at time step t. We predict the spatial attention map based on the previous GRU state as,

$$\alpha_{t,i,j} = \frac{exp(e_{t,i,j})}{\sum_{i,j}(exp(e_{t,i,j}))} \tag{10}$$

where $e_{t,i,j} = {V_a}^T tanh(H)$ and V_a is a trainable parameter. To make α_t sensitive to character arrangement especially for the single-line and double-line LPs, $h_{i,j}$ is concatenated with a one-hot encoding of the simplified spatial coordinates (i, r_j). r_j is used as the encoded coordinate instead of j as follows,

$$r_j = \begin{cases} 1, & single-line \\ \left[\frac{j}{2} + 0.5\right], & double-line \end{cases} \tag{11}$$

here $[\cdot]$, which is the same as function $floor(\cdot)$, returns the value of a number rounded downward to the nearest integer, and $j = 1, 2, 3, 4$. H is updated as follows,

$$H = W_h h_{i,j} + W_s s_{t-1} + W_i f_i + W_{r_j} f_{r_j} + b \tag{12}$$

where W_h, W_s, W_i, W_{r_j} and b are trainable parameters. f_i is a one-hot encoding of coordinate i, while f_{r_j} is that of r_j. More details of this process are referred to [18].

We train the model using minimum the negative log-likelihood of conditional probability as follows,

$$L_{reg} = -\sum_{n=1}^{N}\sum_{t=1}^{M} \log\ p(y_{n,t}|y_{n,1:t-1}, C_n, I_n, \delta) \tag{13}$$

where I_n, C_n and $Y_n = \{y_{n,t}\}$ are the input image, classification label and corresponding license number in the train set $B = \{I_n, C_n, Y_n\}$, $n = 1 \cdots N$. δ represents the parameters of the model. M is the maximal length of the predicted license number. Note that during training, we directly perform APAN based on the ground truth detection results.

D. Training

EILPR accomplishes LP bounding box detection and LP number recognition in a single forward. The training involves loss functions for detection and recognition performance, as well as pre-training the recognition module before training EILPR end-to-end.

1) Training Objective: The detection and recognition sub-networks can be connected by APAN for end-to-end training. No other input is required except for the vehicle image annotated with the plate bounding box, plate class and plate number. For end-to-end training of LP detection and recognition, the whole loss function can be formulated as:

$$L_{EILPR} = L_{det} + \lambda L_{reg} \tag{14}$$

where λ is the hyper-parameter to control the balance among detection and recognition task, which is set to 1 in this paper.

2) Pre-Training Recognition Module: EILPR is a end-to-end LPR system for multi-line LPs. Existing LP data sets for both detection and recognition do not uniformly cover various LP classes and character categories, which is greatly affect the recognition ability. To make the recognition module robust, we pre-train the recognition module with an integrated data set which consists of real single-line LP images and synthesized double-line LP images, and fine-tune the model end-to-end on different real data sets. In the pre-training phase, we extract the coarse features directly through the first three conventional layers of detection module as the ROI features in Fig. 4, ignoring ROI pooling layer. We pre-train the recognition module by batches of 64 for 30 thousand iterations, and evaluate the accuracy on the validation set every 500 steps. We choose the model of the highest accuracy in evaluations on validation set as the pre-trained model.

EILPR is trained end-to-end from scratch with ADADELTA optimizer. The learning rate is set to 1.0 at the beginning and decayed to 0.1 after 20 thousand iterations. Since not enough public double-line plate images are available, we fine-tune the model end-to-end on a private double-line plate dataset and evaluate the performance of double-line images on the SYSU-ITS [42] dataset. Due to the imbalance in the number of single-line and double-line LP image sets in real scene, we fine-tune the pre-trained model on the two datasets respectively to evaluate the end-to-end performances.

IV. Experiments

We evaluate end-to-end LPR performance of the proposed method on several standard benchmarks containing multinational LPs. Moreover, analysis of each module and comparisons with previous methods are also given to demonstrate the superiority and reasonableness of EILPR.

A. Datasets

Synthetic data are generated randomly by image processing, containing 20,000 multinational LP images with various brightness, chroma, clarity and angles of view as shown in Fig. 6(a). It contains the LP images of China, America, Italy and so on. And the characters on the plates have 66 categories, which comprise 32 Chinese characters, 24 letters and 10 numbers.

Chinese license plates dataset (CLPD) [43] dataset contains about 260,000 Chinese single-line LPs collected from different security and surveillance cameras as shown in Fig. 6(a). We randomly selected 150,000 images as training set and the rest as test set. CLPD and the synthetic dataset are combined into a pre-training LP dataset **(PLPD)** to pre-train the recognition module in EILPR.

CCPD [31] was collected in the parking lot of Hefei province of China as illustrated in Fig. 6(b), which is divided into several sub-datasets. We randomly select 50,000 ordinary blue plate images from CCPD for experiments. As usual, all the images are subsequently split into 2 subsets in the proportion of 9:1 respectively for training and testing.

Double-line Traffic-data **(DLTD)** is a private double-line plates dataset collected from various surveillance cameras of traffic crossroads, as shown in Fig. 6(d). DLTD contains about 1000 images, and we split them as 8:2 into training set and test set respectively. Since the number of images in DLTD is far fewer than that of CCPD, we separately verify the single-line and double-line plates.

SYSU-ITS dataset is a public LP image set, which is provided by OpenITS. SYSU-ITS includes 1402 images containing 958 single-line LPs and 84 double-line LPs, which are all selected from the HD images of road bayonets as shown in Fig. 6(c). We pick out double-line LP images as a test set. In addition, the accuracy of labels which are not in the training set is not taken into account in the evaluation. For the convenience of the experiments, we combine the double-line test sets of DLTD and SYSU-ITS into one test set named D-SYSU.

Open-ALPR consist of 114 Brazilian and 108 European LP images, which is used to verify the generalization of EILPR in multinational LP recognition.

The LP images above are annotated with rectangle boxes for detection and character sequence for recognition. We fine-tune the pre-trained model on training set CCPD and SYSU-ITS, and then verify the performance on Open-ALPR and CCPD for single-line plates. Also we fine-tune the pre-trained model on DLTD and verify the performance on D-SYSU for double-line plates.

B. Implementation Details

We implement the proposed approach under the framework of PyTorch [44]. We train the model on a work-station with two Intel Xeon(R) E5-2620 2.10GHz CPU, single NVIDIA GeForce GTX Titan X, and 64GB RAM. CUDA 9.0 and CuDNN v7 backends are used in our experiments so that the model is GPU-accelerated.

C. Ablation Studies

For better understanding the strengths of the proposed method, we first provide the ablation studies from three aspects. First, we demonstrate the benefits of end-to-end training. Second, we compare APAN and TPS-based alignment on recognition performance. Third, we compare our 2D attention based recognizer with the more popularized LSTM based one and 1D attention based one.

1) With vs. Without End-to-End Training: In end-to-end framework, the LP recognition is depended on the LP region predicted by detection module. Meanwhile, the recognition supervision can provide more detailed text stroke features for

(a) PLPD (b) CCPD (c) SYSU-ITS (d) DLTD (e) Open-ALPR

Fig. 6. Examples of different datasets.

TABLE III

PERFORMANCE ON CCPD AND D-SYSU. *Multi-line* REPRESENTS WHETHER THE METHOD IS APPLICABLE TO MULTI-LINE LICENSE PLATE RECOGNITION. *End-to-End* REPRESENTS WHETHER THE METHOD IS TRAINED END-TO-END

Method	Multi-line	End-to-End	Accuracy (CCPD)	Accuracy (D-SYSU)	Accuracy (Open-ALPR)	
					EU	BR
HyperLPR[30]	×	×	79%	-	-	-
MTCNN+LPRNet[45]	×	×	92.1%	-	-	-
RPNet[31]	×	✓	93.8%	-	-	-
SLPNet[32]	×	✓	98.2%	-	-	-
MTCNN+Yu[14]	✓	×	89.7%	86.3%	-	-
MTCNN+Attention[40]	✓	×	95.0%	89.7%	-	-
SEE-LPR[33]	×	✓	-	-	93.52%	94.73%
Soghadi[10]	✓	✓	-	-	98.54%	98.28%
Our Two-Stage	✓	×	95.6%	90.5%	96.3%	96.18%
with TPS-based alignment	✓	✓	97.7%	93.3%	98.62%	98.21%
with LSTM	×	✓	98.4%	-	98.73%	98.31%
with 1D-attention	×	✓	98.5%	-	98.68%	98.25%
EILPR(ours)	✓	✓	**98.8%**	**93.6%**	**98.9%**	**98.43%**

Fig. 7. Results of aligned LP and recognition. *ROI* represents the detected LP from the input. *TPS-based* represents the aligned LP and the corresponding predicted LP number with TPS-based alignment method. *APAN* represents the aligned LP generated by APAN and the corresponding LP number predicted by EILPR.

(a) (b)

Fig. 8. Detection results with and without end-to-end training. From left to right in (a, b): detection without guidance from recognition and EILPR.

LP detection. To demonstrate the performance of end-to-end training, we evaluate a variant of our method in which LP detection and recognition are trained separately.

Some qualitative results are shown in Fig. 8, which reveal the detection performance with and without end-to-end training. Apparently, the detection results with end-to-end training is more accurate. In Fig. 8(a), with end-to-end training, LP whose feature is not salient could also be detected accurately, because the recognition supervision has a correction effect on the detection network. Without end-to-end training, LP detection may miss some LP regions or misclassify LP-alike background. As shown in Fig. 8(b), the model without end-to-end training mis-classified the text which has similar structured with LP. For the recognition performance, we can see in Table. III, the end-to-end training based methods (including EILPR and other configurations) outperform our two-stage method significantly in recognition performance. It can be seen that in the end-to-end training framework, detection and recognition modules are mutually optimized.

2) APAN vs. TPS-Based Alignment: Aster [16], which is a state-of-the-art irregular text recognition method, consists of a text alignment module with TPS transformation and a recognition module with 1D attention decoder. The TPS-based

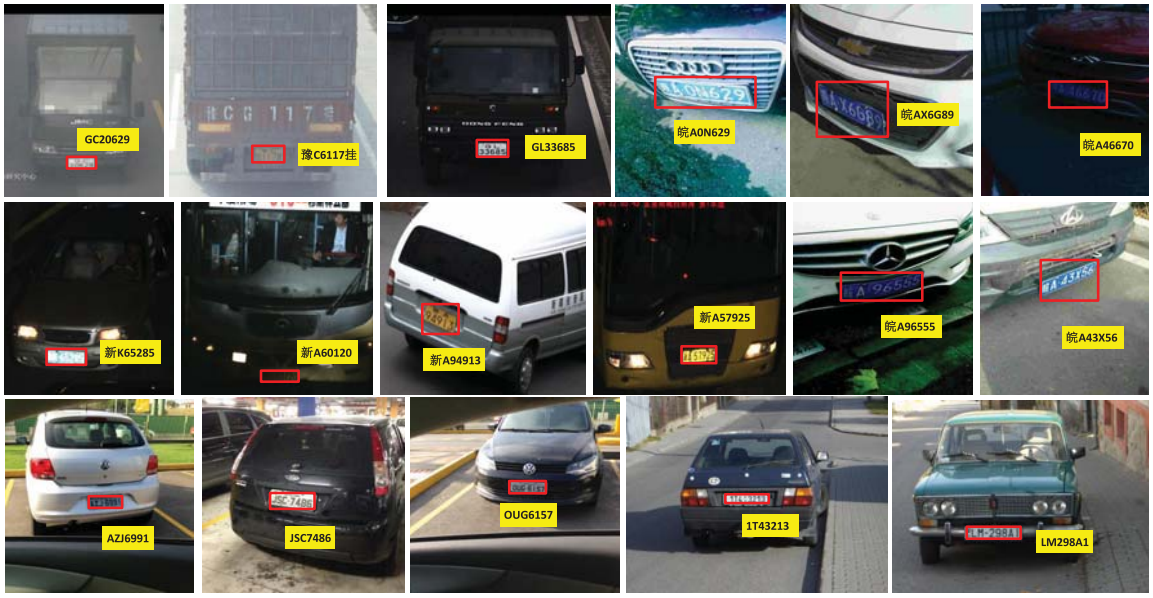


Fig. 9. Examples of detection and recognition results with EILPR.

alignment aims to transform features with TPS transformation by STN. Through analyzing the end-to-end recognition results, we argue that TPS-based alignment may be unsuitable for license plates. Some examples are shown in Fig. 7. The red characters in Fig. 7 are mis-classified with TPS-based method, as they are not properly aligned. TPS-based methods depends on the prediction of dozens of control points, and for the image with large deformation, it may cut off part of the text. APAN aims to reduce text deformation, but it does not completely eliminate it due to the presence of predictive bias. As shown in Fig. 7, the aligned plates generated by APAN is significantly superior to TPS-based method and are sufficient for accurate recognition, although the alignment is not perfect.

To further explore the impact of TPS-based alignment on LPs of different shapes, we evaluate a variant of our method which replace APAN with TPS-based alignment. Table. III show that TPS-based method for large deformation LPs will reduce the end-to-end recognition performance, indicating that TPS transformation parameters are not suitable for irregular license plate recognition. Because the LPs in D-SYSU have small deformations, the recognition accuracy is not much different on D-SYSU. It can be seen that APAN, which is based on perspective transformation, is better to depict the deformation of LP images.

3) 2D Attention Based Recognizer vs. Others: Most previous ALPR methods are based on CTC or 1D-attention decoder, which predict LP number from 1D sequence features. However, these methods cannot recognize the multi-line license plates. In contrast, our method adopts the 2d attention to decode the multi-line LP number. To verify the effectiveness of 2D attention based recognizer, we evaluate a variant of our method which consists of the CNN in our recognition module, a bi-directional LSTM with 256 output channels per direction, a fully-connected layer and a CTC decoder. Also, we evaluate another variant of our method which consists of the CNN in our recognition module and a 1D-attention based decoder. As shown in Table. III, adopting CTC or 1D-attention have little influence on single-line LPR performance. As a result, the 2D attention based recognizer far outperforms CTC based and 1D attention based recognizer on double-line LP end-to-end performance.

More specifically, CCPD has a large number of difficult samples with severe perspective deformation, while D-SYSU and Open-ALPR have less. As shown in Table. III, the recognition module based on 2D attention decoder performs well on both single-line plates and double-line plates. The attention-based methods outperforms CTC-based methods, because attention-based decoders are more better at handling the irregular text. It can be concluded that, our model performs best on both multi-line LPs and largely deformed LPs.

D. Comparison With the State-of-the-Art

In this subsection, we compare with previous methods on several benchmarks to verify the superiority of our work. We evaluate the performance of the proposed EILPR with other publicly reported models on Chinese LP recognition. The rule for the calculation of the recognition accuracy is described as follows: only when all the characters of each LP on an image are correctly recognized, the result is considered to be correct. We compare with several publicly available models that are HyperLPR [30], MTCNN+LPRNet [45], RPNet [31], SLPNet [32], SEE-LPR [33] and so on. However, these models can only recognize single-line license plates or recognize multi-line plates based on character segmentation, so we integrate MTCNN and the method [14] as the comparison model for multi-line LPR.

1) Experiments on Single-Line LP: As shown in Table. III, the proposed method can achieve state-of-the-art performance on CCPD. Because CCPD has a large number of samples with severe perspective deformation, EILPR and SLPNet, which

take into account the LP deformation in the image, outperform others significantly. For the single-line LP recognition, the two methods have the similar effectiveness.

2) Experiments on Double-Line LP: As shown in Table. III, the proposed method can achieve state-of-the-art performance on D-SYSU, which consists of double-line LPs. With the help of 2D attention based recognizer, EILPR outperforms other methods by a large gap. Due to the limitation of data volume and data balance, the performance on double-line LPs is not as good as that of single-line LPs.

3) Experiments on Multinational LP: EILPR can also be used for multinational LP detection and recognition, which is verified on benchmark Open-ALPR as shown in Table. III. SEE-LPR can only recognize single-line LP because of its CTC-based decoding. The method [10] recognizes multi-line LP based on character detection, which is limited by image quality and manual annotation. By comparison, EILPR has a good generalization ability, because it is not limited by the length of LP number and the spatial position of character.

Fig. 9 shows detection and recognition results on benchmarks with our EILPR model. In addition, in terms of the recognition speed, the model based on 2D attention is about two times slower than the CTC-based model, which needs further optimization. Fortunately, the Inference Engine from Intel® OpenVINO [46] has been applied to accelerated CTC-based or segmentation-based LPR models [6], [43]. Acceleration of 2D attention based LPR model will be implemented in the future research. In addition, the synthesis data of LPs will be improved by GAN-based methods [47] and the ALPR method can be applied in urban network with related methods [48]. And we will improve our method in difficult conditions such as occlusion and low-resolution inspired by relevant methods [49], [50].

V. Conclusion

In this paper, we proposed EILPR, a novel end-to-end irregular ALPR method based on automatic perspective alignment. It can detect and recognize multi-line license plates from the detected vehicle images accurately. In order to weaken the influence of detection error on the recognition results, we adopt a coarse-to-fine strategy to extract the LP features precisely. Therefore, we propose automatic perspective alignment network (APAN) to implement the feature extraction strategy and connect detection and recognition for end-to-end training. Considering the rigid body property and 2D text distribution of the LP, EILPR is verified to have good performance on the largely deformed and multi-line LP. Experiments conducted on challenging multinational benchmarks verify the effectiveness of our method. As for future work, we will improve the performance of the system in difficult environment such as foggy and snowy. In addition, model acceleration based on OpenVINO is also in the future research.

References

[1] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 377–391, Sep. 2008.

[2] 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.

[3] S. Ge, J. Li, Q. Ye, and Z. Luo, "Detecting masked faces in the wild with LLE-CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2682–2690.

[4] S. Ge, C. Zhang, S. Li, D. Zeng, and D. Tao, "Cascaded correlation refinement for robust deep tracking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 99, pp. 1276–1288, Apr. 2020.

[5] R. Laroca *et al.*, "A robust real-time automatic license plate recognition based on the Yolo detector," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–10.

[6] R. D. Castro-Zunti, J. Yépez, and S. Ko, "License plate segmentation and recognition system using deep learning and OpenVINO," *IET Intell. Transp. Syst.*, vol. 14, no. 2, pp. 119–126, Feb. 2020.

[7] C. L. P. Chen and B. Wang, "Random-positioned license plate recognition using hybrid broad learning system and convolutional networks," *IEEE Trans. Intell. Transp. Syst.*, early access, Aug. 4, 2020, doi: 10.1109/TITS.2020.3011937.

[8] M. A. Raza, C. Qi, M. R. Asif, and M. A. Khan, "An adaptive approach for multi-national vehicle license plate recognition using multi-level deep features and foreground polarity detection model," *Appl. Sci.*, vol. 10, no. 6, p. 2165, Mar. 2020.

[9] C. Henry, S. Y. Ahn, and S.-W. Lee, "Multinational license plate recognition using generalized character sequence detection," *IEEE Access*, vol. 8, pp. 35185–35199, 2020.

[10] Z. T. Soghadi and C. Y. Suen, *License Plate Detection and Recognition by Convolutional Neural Networks* (Pattern Recognition and Artificial Intelligence). Cham, Switzerland: Springer, 2020.

[11] L. Zhang, P. Wang, H. Li, Z. Li, and Y. Zhang, "A robust attentional framework for license plate recognition in the wild," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 99, pp. 1–10, Nov. 2020.

[12] S.-L. Chen, C. Yang, J.-W. Ma, F. Chen, and X.-C. Yin, "Simultaneous End-to-End vehicle and license plate detection with multi-branch attention neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3686–3695, Sep. 2020.

[13] 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.

[14] Y. Cao, H. Fu, and H. Ma, "An end-to-end neural network for multi-line license plate recognition," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3698–3703.

[15] C.-Y. Lee and S. Osindero, "Recursive recurrent nets with attention modeling for OCR in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2231–2239.

[16] 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. 2018.

[17] L. Qiao *et al.*, "Text perceptron: Towards end-to-end arbitrary-shaped text spotting," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, Apr. 2020, pp. 11899–11907.

[18] H. Xu, Z.-H. Guo, D.-H. Wang, X.-D. Zhou, and Y. Shi, "2D license plate recognition based on automatic perspective rectification," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 202–208.

[19] X. He, C. Li, T. Huang, C. Li, and J. Huang, "A recurrent neural network for solving bilevel linear programming problem," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 824–830, Apr. 2014.

[20] 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.

[21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," in *Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Comput. Soc., 2018.

[22] M. Mittal *et al.*, "An efficient edge detection approach to provide better edge connectivity for image analysis," *IEEE Access*, vol. 7, pp. 33240–33255, 2019.

[23] 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.

[24] V. Jain, Z. Sasindran, A. Rajagopal, S. Biswas, H. S. Bharadwaj, and K. R. Ramakrishnan, "Deep automatic license plate recognition system," in *Proc. 10th Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*, 2016, pp. 1–8.

[25] M. Jaderberg *et al.*, "Spatial transformer networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2015, pp. 2017–2025.

[26] 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. 2016.

[27] T. K. Cheang, Y. S. Chong, and H. T. Yong, "Segmentation-free vehicle license plate recognition using ConvNet-RNN," in *Proc. Int. Workshop Adv. Image Technol.*, 2017, pp. 1–5.

[28] S. Montazzolli and C. 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.

[29] 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. 2017.

[30] Zeussees. *High Performance Chinese License Plate Recognition Framework*. Accessed: Oct. 2020. [Online]. Available: https://github.com/zeusees/HyperLPR

[31] Z. Xu, W. Yang, A. Meng, N. Lu, and L. Huang, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 255–271.

[32] W. Zhang, Y. Mao, and Y. Han, "SLPNet: Towards end-to-end car license plate detection and recognition using lightweight CNN," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2020, pp. 290–302.

[33] D. Tang, K. H., X. Meng, R. Z. Liu, and T. Lu, *SEE-LPR: A Semantic Segmentation Based End-to-End System for Unconstrained License Plate Detection and Recognition*. Cham, Switzerland: Springer, 2020.

[34] S. Montazzolli and C. R. Jung, "License plate detection and recognition in unconstrained scenarios," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 580–596.

[35] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[36] F. L. Bookstein and P. Warps, "Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, p. 585, Jun. 1989.

[37] M. Yang *et al.*, "Symmetry-constrained rectification network for scene text recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9147–9156.

[38] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: A simple and strong baseline for irregular text recognition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 8610–8617.

[39] X. Yang, D. He, Z. Zhou, D. Kifer, and C. L. Giles, "Learning to read irregular text with attention mechanisms," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, p. 3.

[40] Z. Wojna *et al.*, "Attention-based extraction of structured information from street view imagery," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 844–850.

[41] K. Cho, B. V. Merrienboer, C. Gulcehre, F. Bougares, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.

[42] OpenITS. *SYSU-ITS*. Accessed: Oct. 2020. [Online]. Available: https://www.openits.cn/openData4/569.jhtml

[43] S. Zherzdev and A. Gruzdev, "LPRNet: License plate recognition via deep neural networks," 2018, *arXiv:1806.10447*.

[44] A. Paszke *et al.*, "Automatic differentiation in pytorch," in *Proc. NIPS-W*, 2017, pp. 1–4.

[45] X. Xue. *A Two Stage Lightweight and High Performance License Plate Recognition in MTCNN and LPRNet (2019)*. Accessed: Nov. 2019. [Online]. Available: https://github.com/xuexingyu24/License Plate Detection Pytorch.

[46] *Intel Openvino Toolkit/Intel Software*. Accessed: Sep. 2020. [Online]. Available: https://software.intel.com/en-us/articles/OpenVINO-InferEngine.

[47] T. Wang, T. Zhang, L. Liu, A. Wiliem, and B. Lovell, "Cannygan: Edge-preserving image translation with disentangled features," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 514–518.

[48] K.-J. Pai, R.-S. Chang, R.-Y. Wu, and J.-M. Chang, "A two-stages tree-searching algorithm for finding three completely independent spanning trees," *Theor. Comput. Sci.*, vol. 784, pp. 65–74, Sep. 2019.

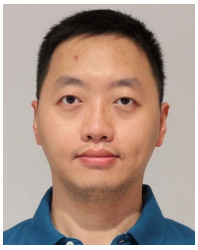
[49] S. Ge, C. Li, S. Zhao, and D. Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3387–3397, Oct. 2020.

[50] S. Ge, S. Zhao, C. Li, Y. Zhang, and J. Li, "Efficient low-resolution face recognition via bridge distillation," *IEEE Trans. Image Process.*, vol. 29, pp. 6898–6908, 2020.

Hui Xu received the B.Sc. and M.Sc. degrees in automation from the Beijing University of Posts and Telecommunications in 2009 and 2012, respectively. She is currently pursuing the Ph.D. degree with the Chongqing School, University of Chinese Academy of Sciences. She also works as an Engineer with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences. Her research interests include scene text detection and recognition, license plate detection and recognition, and intelligent transportation systems.

Xiang-Dong Zhou received the B.S. degree in applied mathematics and the M.S. degree in management science and engineering from the National University of Defense Technology, Changsha, China, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1998, 2003, and 2009, respectively. He is currently a Professor with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences. His research interests include handwriting recognition and ink document analysis.

Zhenghao Li received the B.S. degree in telecommunications engineering and the Ph.D. degree in instrumentation science and technology from Chongqing University, China, in 2003 and 2009, respectively. In 2019, he has been with the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, where he is currently an Associate Professor. His current research interests include image processing, pattern recognition, and edge computing.



Liangchen Liu received the B.Eng. degree in information engineering and the M.Sc. degree in instrument science and technology from Chongqing University, Chongqing, China, in 2009 and 2012, respectively, and the Ph.D. degree from The University of Queensland, Brisbane, QLD, Australia, in 2017. He is currently a Research Fellow with The University of Melbourne. His current research interests include unsupervised learning, detection, segmentation, and visual attribute and its related applications.

Chaojie Li received the B.Eng. degree in electronic science and technology and the M.Eng. degree in computer science from Chongqing University, Chongqing, China, in 2007 and 2011, respectively, and the Ph.D. degree from RMIT University, Melbourne, Australia, in 2017. He is currently a Senior Research Associate with the School of Electrical Engineering and Telecommunications, UNSW Sydney, Sydney. His current research interests include graph representation learning, distributed optimization and control of energy storage, neural networks, and their application.

Yu Shi received the B.S. degree in computer science and technology and the M.E degree in software engineering from Wuhan University, Wuhan, China, in 2003 and 2007, respectively. He is currently a Senior Engineer with CIGIT, CAS. He is also the Director of the Research Center for Intelligent Security Technology, CIGIT. He has published more than 20 patents and obtained four patent licenses. He is the West Light A Class awarded by the Chinese Academy of Sciences.