IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 14, 2020, accepted January 11, 2021, date of publication January 28, 2021, date of current version February 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3055243

A Single Neural Network for Mixed Style License Plate Detection and Recognition

QIUYING HUANG, ZHANCHUAN CAI, (Senior Member, IEEE), AND TING LAN
Faculty of Information Technology, Macau University of Science and Technology, Macau 999078, China

Corresponding author: Zhanchuan Cai (zccai@must.edu.mo)

This work was supported in part by the Science and Technology Development Fund of Macau under Grants 0052/2020/AFJ, 0038/2020/A, and 0069/2018/A2.

ABSTRACT Most existing methods for automatic license plate recognition (ALPR) focus on a specific license plate (LP) type, but little work focuses on multiple or mixed LPs. This article proposes a single neural network called ALPRNet for detection and recognition of mixed style LPs. In ALPRNet, two fully convolutional one-stage object detectors are used to detect and classify LPs and characters simultaneously, which are followed by an assembly module to output the LP strings. ALPRNet treats LP and character equally, object detectors directly output bounding boxes of LPs and characters with corresponding labels, so they avoid the recurrent neural network (RNN) branches of optical character recognition (OCR) of the existing recognition approaches. We evaluate ALPRNet on a mixed LP style dataset and two datasets with single LP style, the experimental results show that the proposed network achieves state-of-the-art results with a simple one-stage network.

INDEX TERMS ALPRNet, license plate recognition, object recognition, convolutional neural network.

I. INTRODUCTION

From paid-parking to traffic control and toll violations, automatic license plate recognition (ALPR) is one of the key components of many traffic-related applications [1]–[6]. There has considerable work on ALPR in the past decade, but most of the existing approaches focus on specific types of license plates (LPs), and there is little work supporting multiple or mixed LPs [7]–[9]. The growth of cross-boder communication and exchanges have promoted the demand of ALPR systems that support multiple and mixed style LPs. In the cities in the China's Pearl River Delta Zone, vehicles may carry up three LPs with different styles. This article focuses on multiple and mixed LPs recognition and introduces a new ALPRNet framework that integrates the LP detection and character recognition into a single neural network. ALPRNet treats LPs and characters as basic elements to detect and classify, and the tasks of LP and character detection and classification are conduct simultaneously in one pass.

Most of existing ALPR methods consider LP recognition as two independent tasks: LP detection and character recognition, which are implemented sequentially. The widely used CNN-based methods usually employ object detectors to detect LPs [10]–[15], such as you only look once (YOLO), faster region-based convolutional neural network (Faster R-CNN) and mask region-convolutional neural network (Mask R-CNN). For the training error accumulation, the independent sequential tasks result in a sub-optimization problem. Indeed, the tasks of LP detection and character recognition can work collaboratively by providing context and detail information to each other, and they can be conducted simultaneously to improve performance.

Recently, more and more end-to-end ALPR networks have been developed, which can complete the tasks of LP detection and character recognition by using a single unified deep neural network [16]. Since these methods alway combine the region-based convolutional neural network with an RNN-based sequential model, they have some problems: 1) the RNN-based part of the network consumes high computational cost, such that the detection part of the network is difficult to optimize; 2) the mainstream detectors of the region-based object detection, such as Faster R-CNN, single shot detector (SSD), and YOLO, are anchor-based detectors, they rely on a set of pre-defined anchor boxes. The

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

performance of these object detection networks is sensitive to their size, aspect ratios, and the number of anchor boxes, moreover, anchor boxes involve complicated computation of intersection over union (IoU) scored with ground-truth bounding boxes.

In this article, we propose an ALPR network called ALPRNet, wherein a single neural network conducts LP detection and recognition simultaneously, and the characters are considered as objects that need to be detect and classify. Such scheme predicts the bounding boxes and labels of LPs and characters directly and gets rid of the complicated RNN-based recognition procedure with the region of interest (RoI) pooling and crop procedures of object detection. And the contributions of this work are concluded as follows:

1). We design a novel one-stage network for LP detection and recognition, wherein two parallel branches of object detection and classification are introduced, they directly produce the LP type and the LP string excluding the redundant and intermediate steps.
2). We introduce the classification branch of LPs to support multiple and mixed style LPs.
3). We use the bounding box of LPs and characters to assemble the LPs, thus ALPRNet naturally supports multi-line and variable length of LPs.

The rest of the article is organized as follows: Section II gives a brief review of detection and related work. Section III presents a detail description of the proposed network. Section IV shows the experimental results of the proposed network. The conclusion of this work is given in Section V.

II. RELATED WORK

A. ALPR WITHOUT DEEP LEARNING

The existing ALPR methods can be classified into two categories: LP first and character first approaches. A traditional two-stage ALPR procedure uses the features of edge, color, and texture to detect the location of LP, and then performs character segmentation and features extraction followed by machine learning classifiers to identify each character. Yuan *et al.* [17] proposed a line density filter to connect regions by edge density from binary images. In [18], edges are clustered by Expectation-Maximization algorithm for LP detection. Ashtari *et al.* [19] developed a method of LP detection that analyzes pixels by a color geometric template via strip search. Yu *et al.* [20] employed a wavelet transform to produce horizontal and vertical features of an image and used empirical mode decomposition (EMD) analysis to determine the position of a LP. These methods are suitable for LPs that have distinguished features, but they are too sensitive to complex images.

Character first methods try to find character regions in the image and then cluster these regions based on the graph semantics to construct LPs. The graphic semantics includes the front and background color, orientation, characteristic scale, position, etc. In [21], maximally stable extremal region (MSER) is used to extract candidate characters, and then construct conditional random field (CRF) models on the candidate characters to represent the relationship among candidate characters. And it localize LPs through the belief propagation inference on CRF. In [22], stroke width transform (SWT) and MSER are combined to detect character regions and LPs by probabilistic Hough transform. Character first methods have received a high recall, but they are easily disturbed by background text.

B. ALPR WITH DEEP LEARNING

In order to achieve high accuracy, more and more ALPR work turns to use CNNs to detect and recognize LPs. Most of these methods are based on the traditional approach with two stages, i.e., the detection network based on the generic object detection methods (such as YOLO, Faster R-CNN, and SSD-based scheme) is used to detect LPs, and the recurrent neural network is used as the recognition network of OCR. In [14], the authors proposed a complete ALPR scheme that employs three subnetworks to conduct vehicle location, LP detection, and character recognition. A YOLOv2 [23] based subnetwork is used to detect vehicle, wherein a novel CNN called WPOD-NET is created for LP detection and affine transformation regress, and a modified YOLO network served as OCR module for character recognition. In [13], a real-time ALPR system is proposed, which uses a YOLO-based network for vehicle and LP detection, and a CR-NET [24] network is used for character segmentation and recognition. In [25], CNN-based text detectors are used to search character regions in digital images and cluster them by edge information to construct LP, and a RNN with connectionist temporal classification (CTC) is employed to label and recognize characters.

In addition, some end-to-end ALPR methods have been proposed by using united network to perform LP detection and character recognition, but they still include sequential stages for LP detection and OCR. In [26], a method uses VGG16 network [27] as backbone network and employs a region proposal network (RPN) on different layers of feature pyramid network (FPN) to generate LP proposals and regress the bounding box, and RNNs with CTC is used to recognize characters.

C. SCENE TEXT SPOTTING WITH DEEP LEARNING

Currently, many scene text spotting (STS) approaches based on deep learning have been proposed, but most of them employ generic object detectors (such as Faster R-CNN, SSD, and YOLO) to detect and regress text instances directly, or to predict text/non-text probability of each pixel by semantic segmentation methods (such as fully convolutional network (FCN) [28]. Li *et al.* [29] proposed an end-to-end approach that integrates text detection and recognition into a unified framework, which consists a tailored region proposal network for text detection and an attention-based recurrent neural network (RNN) encoder/decoder for text recognition. Zhou *et al.* [30] proposed an efficient and accurate scene text detector (EAST detector) to predict a word or line-level text,

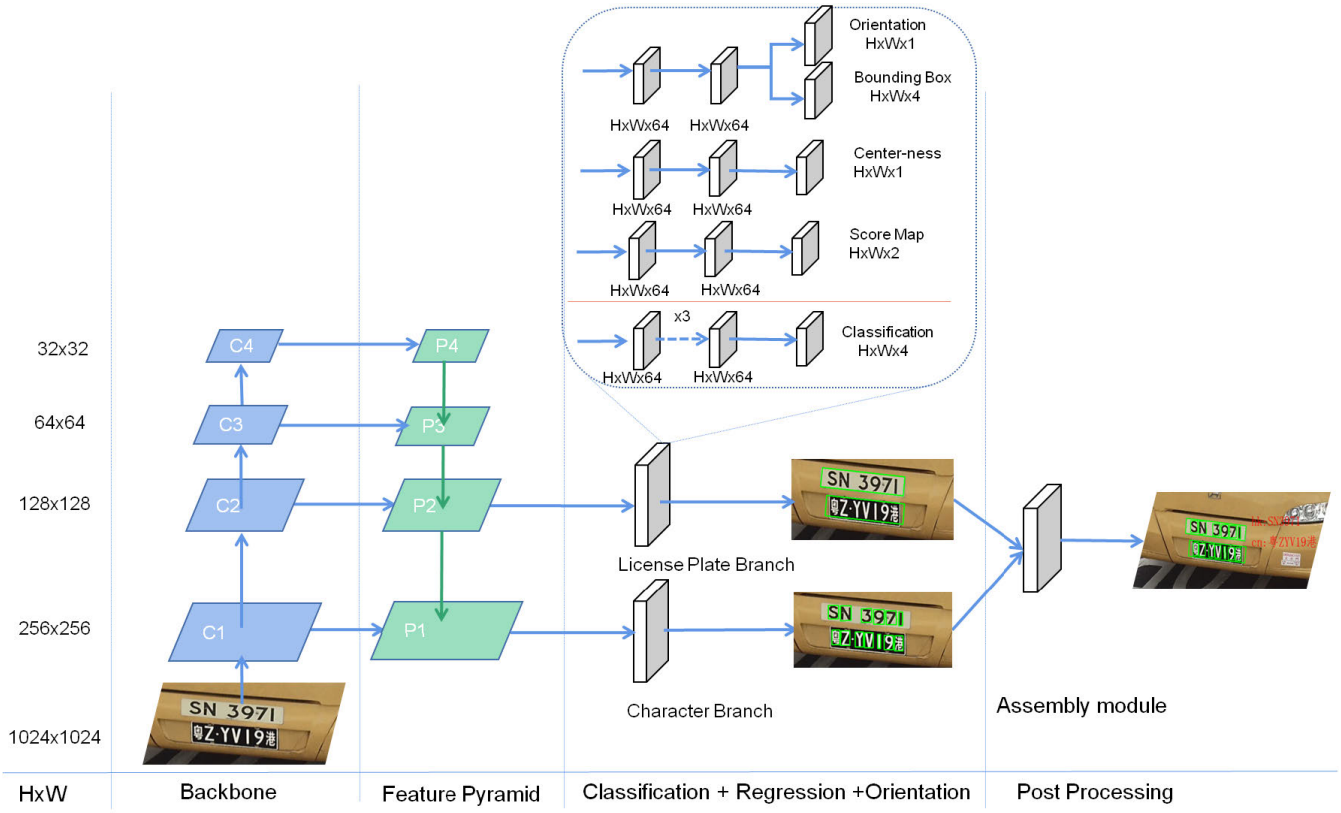


FIGURE 1. Framework of the proposed network: ALPRNet. Two fully convolutional one-stage object detectors are employed to detect bounding boxes and classify LPs and characters. The class number of LPs is 4 (Background, Mainland China, Hong Kong, and Macao), and the class number of characters is 73 (1 background, 26 English characters, 10 numeric characters, and 36 Chinese characters).

which consists of a fully convolutional network with non-maximum suppression (NMS) merging state. He *et al.* [31] proposed an end-to-end framework based on SSD by introducing an text attention module, which enables a direct text mask supervision and achieves strong performance improvements by training text detection and recognition jointly. Xing [32] proposed a one-stage model that processes text detection and recognition simultaneously.

The proposed network is a one-stage end-to-end model that treats LP detection and character recognition as object detection. We directly regress each object's bounding box and classify object types (LP type and character). It naturally supports multiple and mixed style LPs, and also supports LP that has multiple lines or variable length of characters.

III. METHODOLOGY

The proposed ALPRNet is a one-stage convolutional network consisting of two fully convolutional one-stage object detectors (FCOS detectors) [33] for the detection and classification of LPs and characters. These two detectors are treated equally and implemented in parallel on different levels of feature maps derived from backbone network, as shown in Fig. 1. We employ a ResNet-50 [34] with feature pyramid structure (FPN) to serve as backbone, design a assembly module to combine detected LPs and characters, and finally output the LP strings. Branches of LP and character are integrated seamlessly, resulting in an end-to-end trainable model.

The object detectors are fully convolutional, which are anchor free and get rid of the complicated RoI cropping and pooling operations of mainstream object detector methods. It directly classifies the detected objects, so it avoids the RNNs branch of OCR. As a result, the network is simple and easy to extend in practical use.

A. BACKBONE AND FPN

ResNet-50 is employed as backbone networks and applies a feature pyramid structure (FPN) by using a top-down structure to fuse features across multiple resolutions, and we use different layers of FPN to perform character and LP detection. The P2 layer of FPN has 1/8 resolution of the input image that provides a larger receptive field size, it is suitable for LP detection. The P1 layer of FPN has 1/4 resolution of the input image that provides more detail features, it is suitable for small object detection and is used for character branch.

B. FULL CONVOLUTIONAL ONE-STAGE OBJECT DETECTOR

We introduce two FCOS detectors to densely predict bounding box, center-ness, orientation, and classification of LPs and characters. Different from anchor-base detectors that regress target bounding box with anchor boxes as reference,

FCOS detector regards locations as training samples and directly regresses the target bounding box at each location. FCOS detector reduces intermediate steps, such as RoI proposals and segmentation. Moreover, it eliminates the sample imbalance problem between positive and negative anchor boxes during training, that causes the training to be inefficient. The regression targets of a location (x, y) are the distances to the borders of the bounding box, and it has the form:

$$\begin{aligned} l^* &= x - x_0, \quad t^* = y - y_0, \\ r^* &= x_1 - x, \quad b^* = y_1 - y \end{aligned} \tag{1}$$

where (x_0,y_0) and (x_1,y_1) are the top-left and right-bottom points of a bounding box, respectively.

FCOS detector predicts the center-ness $cn_{x,y}$ of each location (x, y) to depict the normalized distance from the location to the center of the object. In inference stage, center-ness is used to filter out low-quality predicted bounding boxes produced by locations far from the center of an object. Such that it improves detection performance and reduces the workload of NMS process. The center-ness of each location is defined as:

$$cn_{x,y} = \sqrt{\frac{\min(l^*, r^*)}{\max(l^*, r^*)} \times \frac{\min(t^*, b^*)}{\max(t^*, b^*)}} \tag{2}$$

where l_*, t_*, r_*, and b_* are the regression predictions of each location $b_{x,y}$.

C. LICENSE PLATE BRANCH

The LP branch is designed to detect and classify the type of LP in a higher level concept. This branch contains four sub-branches for LP segmentation (score map), bounding box and orientation regression (geometry), center-ness regression, and LP classification (class), respectively. We use the convolutional feature map from the P2 level of FPN to implement LP detection, and the input feature map has 1/8 spatial resolution of the input image.

The bounding box regression sub-branch employs 2 convolutional layers with filter size of 3×3, followed by 2 convolutional layers with filter size of 1×1 to produce the 5-channel feature maps, estimating a LP bounding box at each spatial location. Each bounding box is parameterized by five parameters, indicating the distances of current location to the top, bottom, left, and right sides of the bounding box, as well as the orientation of bounding box. The LP segmentation and center-ness sub-branches have 3 convolutional layers with filter sizes of 3×3, 3×3, and 1×1, respectively. The classification sub-branch of LP has four convolutional layers with one more 3×3 convolutional layer, it predicts 4-channel probability maps including 3 LP types and 1 background.

D. CHARACTER BRANCH

Similar with the LP branch, the character branch also divides into four sub-branches that perform character probability prediction (score map), bounding box regression,center-ness, and character classification (class). It uses the P1 output of FPN as input feature map that has 1/4 spatial resolution of the input image. The classification sub-branch predicts 73-channel probability maps including 1 background, 26 English characters, 10 numeric characters, and 36 Chinese characters.

E. ASSEMBLE MODULE

Assemble module is to combine the detected LPs and characters and finally outputs strings. The assembly operation depends on the calculation of the max overlap rate of character to each LP, which supports LPs with multiple lines or variant in character count. By IoU calculation, the predicted bounding boxes of LPs and characters are applied to manipulate the detected characters into LPs.

The IoU calculation orientated bounding box is more accurate than rectangle bounding box, especially for LPs with multi-lines. The use of orientated bounding box helps to calculate the correct order of each character of multi-line LP by rotating the LP to horizontal and sorting by the coordinate (x, y). Assemble module helps to handle non-LP characters in input image, only candidate characters in LP are treated as detected characters and only candidate LPs have characters are treated as detected LP. It filters out most of the non-LP characters. Finally we run the the outputs of assemble module against a regular expression matcher to find results that match LP patterns of the corresponding LP type.

F. LOSS FUNCTIONS

Corresponding to 4 sub-branches of LP detection and character recognition branches, the proposed train target is divided into four parts: score map loss (L_s), center-ness loss (L_{center}), geometry loss (L_g), and classification loss (L_c). The loss formula is defined as follows:

$$L = L_s + L_{center} + \lambda L_g + \beta L_c \tag{3}$$

In this article, λ is set to 1, and β is set to 10 to balance the losses of geometry and classification. In training, we find that characters and LPs are small objects, so the deviation from bounding box results bigger loss than the general object detection. This situation leads to a training stagnation, so the β value is set to 10 for solving this problem.

Score Map Loss: Most of the detection pipelines face to the class imbalanced problem, thus one should carefully process training image. However, it always introduces more parameters to tune and make the pipeline more complicated. The goal of dice coefficient loss is the maximization of these metrics, it performs better on the class imbalanced problems, which is defined as

$$L_s = 1 - 2 \times \frac{| X \cap Y | + \text{smooth}}{| X | + | Y | + \text{smooth}} \tag{4}$$

Center-ness Loss: This is L2 loss and is defined as

$$L_{center} = \frac{1}{N_{pos}} \sum_{k=1}^{n} (c - c^*)^2 \tag{5}$$

where N_{pos} denotes the number of positive samples, c and c^* represent the prediction and ground truth of center-ness, respectively.

Geometry Loss: The proposed training loss function of geometry is

$$L_g = L_{diou} + \lambda_\theta L_\theta \tag{6}$$

where L_{diou} is the loss of regression, L_θ is angle loss, and λ_θ is set to 1.

Regression Loss: Since the size of the LPs and characters vary widely, the loss function of regression should be scale-invariant. Otherwise, it will causes loss bias. Distance-IoU (DIoU) loss [35] is invariant to the scale of regression, and it provides moving direction for bounding box when there is no overlap with ground truth box. It considers on the overlap area and central point distance of bounding boxes. DIoU is also used in NMS.

$$L_{diou} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} \tag{7}$$

where b and b^{gt} indicate central points of bounding box, $\rho(\cdot)$ is the Euclidean distance, c is the diagonal length of the box covering the two boxes, and IoU is the rate of intersection/union areas between b and b^{gt}.

Angle Loss: The loss of rotation angle of bounding box is computed by

$$L_\theta = 1 - \cos(\theta - \theta^*). \tag{8}$$

where θ and θ^* represent the prediction and ground truth of the rotation angle of a bounding box, respectively.

Classification Loss: This loss has the form:

$$L_c = \frac{1}{N_{pos}} \sum_{x,y} L_{cls}(c_{x,y}, c^*_{x,y}) \tag{9}$$

where L_{cls} is the focal loss of the position (x, y), $c_{x,y}$ and $c^*_{x,y}$ represent prediction and ground truth of class of the position (x, y), respectively, and N_{pos} is the number of positive samples.

G. INFERENCE

In ALPRNet, two different feature map layers (P1 and P2) are used to predict characters and LPs. During the inference, we forward the image through ALPRNet to obtain the object scores $sore_{x,y}$, regression prediction $geo_{x,y}$, center-ness $center_{x,y}$, and classification $class_{x,y}$ of each location on the feature map.

Only the locations with $score_{x,y} > 0.95$ are considered as positive samples, but there are still a large number of bounding boxes that increase the workload of the following NMS process. The quality of bounding boxes produced by the locations that far away from the center of object is poor. Therefore, we use center-ness to filter out these locations that have $center_{x,y} < 0.3$. The combination of using center-ness and score helps to filter out most of the low quality predictions. Moreover, since the bounding boxes from nearby pixels are highly correlated, we use Locality-Aware NMS [30] to merge the bounding boxes row by row, thus reducing the computation cost of NMS.

H. TRAINING STRATEGY

The training strategy is an end-to-end learning process, wherein the LP detection and character recognition are trained concurrently on the same network. The joint training of these two tasks in a unified framework can avoid error accumulations among cascade models. The amount of annotation is critical to the accuracy of the proposed network, but it is not practical to annotate it only by a human operator. Therefore, we choose a two steps strategy. The training process contains two stages: pre-trained on the synthetic dataset and fine-tuned on the real-world dataset.

IV. EXPERIMENTS

In this section, we conduct experiments to verify the effectiveness of ALPRNet, and then we summary the testing results on the Hong Kong-Zhuhai-Macao (HZM) multi-style dataset and compare the proposed network with some state-of-the-art end-to-end ALPR methods on the AOLP [18] and PKU [17] datasets. These tasks are carried out on a digital computer with one Nvidia RTX2070 GPU (PCIe 8 GB), wherein the used CPU is Intel Corel i7-6700.

We implement ALPRNet in PyTorch1.0, and ResNet-50 is used as the backbone networks. The model is trained with ''Ranger'' optimizer, and the initial learning rate is 0.0002, it is reduced by the formula

$$lr = 0.94^{(epoch-num/8)} \times lr_0. \tag{10}$$

The learning rate is reduced after about 10 K iterations. The size of input images is 1024×1024. To support variant size of LPs and orientations, the training images are cropped, resized, and rotated randomly, and then they are padded to the size of 1024×1024 before feed to train.

A. HZM MULTI-STYLE DATASET OF LICENSE PLATES

The HZM multi-style dataset includes three styles of LPs: Mainland China LP, Hong Kong LP with white background, and Macao LP with black background, which is a private dataset including 1376 images collecting from the real-world system running on the Hong Kong-Zhuhai-Macao Bridge. The images can be divided into 4 groups: Mainland China+Macao LPs, Mainland China+Hong Kong LPs, Macao+Hong Kong LPs, and Mainland China+Macao+Hong Kong LPs. The resolution of these images is 1190×500 pixels. Fig. 2 shows the examples of vehicle running on the Hong Kong-Zhuhai-Macao Bridge.

B. ANNOTATION

In this work, we choose VGG Image Annotator (VIA) to edit image annotations, the annotations of ground truths of all images of a dataset are saved in a single file. There are 74 object types to be annotated, including 3 LP types and 71 characters. The bounding boxes of objects are defined by

(a) (b) (c)

FIGURE 2. Example images from the HZM multi-style dataset. (a) Example image A. (b) Example image B. (c) Example image C.

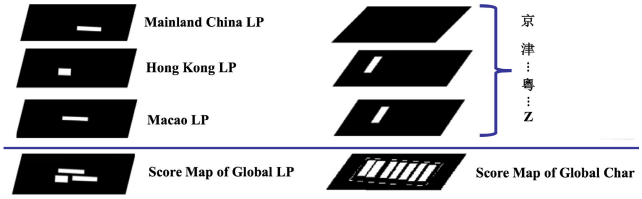


FIGURE 3. Score and classes map.

polygons that are more accurate than rectangles and are easy to rotate the bounding box when we apply data augmentation.

C. GROUND TRUTH GENERATION

We follow EAST to generate geometry map for each training object. For each training object, we calculate a rotated rectangle (RBOX) that covers the object with minimal area, and then generate a 4-channel (left, right, top, and bottom) geometry map that depicts the distance to 4 boundaries of RBOX of each pixel of positive score. Different from EAST, we construct two sets of score maps and geometry maps that are used for LP detection and character recognition. To support object classification, the score map has a number of channels corresponding to the number object classes to detect, such as 4 (3 LP types and 1 background) channels for LPs and 72 (71 characters and 1 background) channels for characters. To reduce the memory used for training, the channels of score map are only generated when calculating loss. Fig. 3 indicates the score and class maps for LPs and characters.

D. SHRINKING SCORE MAP

EAST has proved that the shrinking score map is critical to dense object detection, it means that the center part of object is more important for training and predicting than edge parts. In this article, we use a simpler method that shrinks the endpoint along the line from the endpoint to the center point of quadrangle. The method is defined as follows: For a quadrangle $V = \{v_i | i = 1, 2, 3, 4\}$, $v_i = \{x_i, y_i\}$ represents the vertices on the quadrangle. To shrink V, we calculate the center point of the quadrangle and move its endpoints inward along the line between endpoint and center point, as follows:

$$\begin{aligned} \text{center} &= \frac{\sum_{i=0}^{4} V(x_i, y_i)}{4}, \\ V_i^{'} &= \text{center} + (V_i - \text{center}) \times (1 - r \div 2) \end{aligned} \tag{11}$$

TABLE 1. Performance of the Proposed Network on the HZM Multi-Style Dataset, Wherein 176 Images Includes 280 LPs.

\	LPDA	LPDR	E2E	Passing Rate
ALPRNet	100%	100%	98.21%	99.43%

where r is the shrink rate, which is set to 0.3 by following EAST.

E. EXPERIMENTAL RESULTS

The testing subset of the HZM multi-style dataset consists of 176 images that are continuous in captured time. Totally, there are 280 LPs in these images, about 1.5 LPs per image. As shown in Table 1, the LP detection accuracy (LPDA) is 100%, and the LP detection recall (LPDR) is 100%, there is no false alarm after post handle. The LP recognition accuracy rate (E2E) is 98.21%, and there are 5 LPs that have errors with only one character misclassification.

The passing rate means only images that all LPs are misrecognized causing the vehicle to be rejected. According to the rules of the Hong Kong-Zhuhai-Macao Bridge, a vehicle with multiple LPs can be identified by any of these LPs. Passing rate is an index for this specialized scenario, if one LP has been correctly recognized in an image, it will be counted as passing. In the testing dataset, only one image has not been correctly identified, so the passing rate is 99.43%. Fig. 4 lists some images of LPs that have been detected and read.

F. PERFORMANCE ON THE AOLP DATASET

The AOLP dataset is a dataset of Taiwan LPs, which has 2049 images in total. The AOLP dataset is categorized into AC, LE, and RP subsets, wherein the RP subset of AOLP is the most challenging subset. We use images from other two subsets to train the model of each subsets, such as we use images from LE and AC to train the model for RP subset. Then, we apply data augmentation by cropping, resize, rotation, and affine transformation. Annotation of LPs and characters are generated based on the ground truths information from the AOLP dataset. We train the model with the synthesized dataset that is generated based on Taiwan LP regulation first and trained it with the AOLP dataset.

Table 2 shows that the proposed network surpasses in LP detection on the three subsets. For the end-to-end recognition,

FIGURE 4. LPs have been detected and recognized, wherein "hk" means Hong Kong, "mo" means Macao, and "cn" means Mainland China.

(a) AC (b) LE (c) RP

FIGURE 5. Example results for LP detection and recognition based on the AOLP dataset.

FIGURE 6. Example results for LP detection and recognition based on the PKU dataset with the G5 subset.

the proposed network surpasses on all the subsets, especially in the RP subset, the proposed network exceeds by 3 points.

Fig. 5 shows some images of the AOLP dataset that LPs have been detected and recognized. The results prove the

TABLE 2. Comparison Results Based on the AOLP Dataset.

Method	AC (%)		LE (%)		RP (%)	
	LPDA	E2E	LPDA	E2E	LPDA	E2E
Hsu *et al.* [18]	96	-	95	-	94	-
Li *et al.* [25]	98.38	94.85	97.62	94.19	95.58	88.38
Li *et al.* [26]	99.56	95.29	99.34	96.57	98.85	83.63
ALPRNet	**99.82**	**95.78**	**99.66**	**96.62**	**99.50**	**91.58**

TABLE 3. Comparison Results Based on the PKU Dataset.

Method	Detection Performance (%)					
	G1	G2	G3	G4	G5	Average
Zhou *et al.* [36]	95.43	97.85	94.21	81.23	82.37	90.22
Yuan *et al.* [17]	98.7	98.42	97.72	96.23	97.32	97.69
Li *et al.* [26]	99.88	**99.86**	99.60	**100**	99.31	99.73
ALPRNet	**100**	**99.86**	**99.86**	**100**	**99.65**	**99.83**

effectiveness of the proposed network, it shows that the proposed network has advantage in LPDA, which comes from two collaborative parallel branches of LP detection and character recognition.

G. PERFORMANCE ON THE PKU DATASET

The PKU dataset contains 3977 images of Mainland China LPs. The dataset is divided into 5 groups (i.e., G1, G2, G3, G4, and G5). Since there is only ground-truth information of bounding boxes of LPs, we only use it to evaluate the performance of LP detection. Based on the training results of the synthesis dataset, we use G1 to train the model for G2 subset and apply data augmentation on it. Then, the G1 and G2 subsets are used to train the model for G3, G4, and G5, The G2 subset is used to train the model for G1. Table 3 shows that the proposed network achieves good results when compared with some state-of-the-art methods. Fig. 6 shows that LPs have been detected and recognized.

V. CONCLUSION

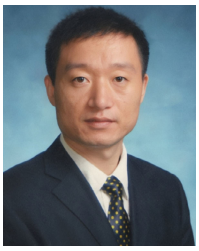
In this article, we present a one-stage ALPRNet for multiple and mixed style LP recognition, which equally treats LPs and characters as objects to detect and classify, and it conducts these two tasks simultaneously. This results in a one-stage fully convolutional framework that solves LP detection and recognition tasks in a integrated framework without any RNNs branches. By sharing the convolutional feature maps, ALPRNet is compact with less parameters, and these two tasks can be trained more effectively and collaboratively. In the experiments, ALPRNet achieves 98.21% accuracy rate on the HZM multi-style dataset, and the results on the datasets with single LP style also show that the proposed network achieves state-of-the-art recognition accuracy.

REFERENCES

[1] C. Henry, S. Y. Ahn, and S. Lee, "Multinational license plate recognition using generalized character sequence detection," *IEEE Access*, vol. 8, pp. 35185–35199, 2020.
[2] M.-X. He and P. Hao, "Robust automatic recognition of Chinese license plates in natural scenes," *IEEE Access*, vol. 8, pp. 173804–173814, 2020.
[3] W. Weihong and T. Jiaoyang, "Research on license plate recognition algorithms based on deep learning in complex environment," *IEEE Access*, vol. 8, pp. 91661–91675, 2020.
[4] I. V. Pustokhina, D. A. Pustokhin, J. J. P. C. Rodrigues, D. Gupta, A. Khanna, K. Shankar, C. Seo, and G. P. Joshi, "Automatic vehicle license plate recognition using optimal K-means with convolutional neural network for intelligent transportation systems," *IEEE Access*, vol. 8, pp. 92907–92917, 2020.
[5] A. Tourani, A. Shahbahrami, S. Soroori, S. Khazaee, and C. Y. Suen, "A robust deep learning approach for automatic iranian vehicle license plate detection and recognition for surveillance systems," *IEEE Access*, vol. 8, pp. 201317–201330, 2020.
[6] Y. Zou, Y. Zhang, J. Yan, X. Jiang, T. Huang, H. Fan, and Z. Cui, "A robust license plate recognition model based on bi-LSTM," *IEEE Access*, vol. 8, pp. 211630–211641, 2020.
[7] H. Seibel, S. Goldenstein, and A. Rocha, "Eyes on the target: Super-resolution and license-plate recognition in low-quality surveillance videos," *IEEE Access*, vol. 5, pp. 20020–20035, 2017.
[8] S. Zhang, G. Tang, Y. Liu, and H. Mao, "Robust license plate recognition with shared adversarial training network," *IEEE Access*, vol. 8, pp. 697–705, 2020.
[9] B. B. Yousif, M. M. Ata, N. Fawzy, and M. Obaya, "Toward an optimized neutrosophic k-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR)," *IEEE Access*, vol. 8, pp. 49285–49312, 2020.
[10] W. Wang, J. Yang, M. Chen, and P. Wang, "A light CNN for end-to-end car license plates detection and recognition," *IEEE Access*, vol. 7, pp. 173875–173883, 2019.
[11] Hendry and R.-C. Chen, "Automatic license plate recognition via sliding-window darknet-YOLO deep learning," *Image Vis. Comput.*, vol. 87, pp. 47–56, Jul. 2019.
[12] Z. Selmi, M. B. Halima, U. Pal, and M. A. Alimi, "DELP-DAR system for license plate detection and recognition," *Pattern Recognit. Lett.*, vol. 129, pp. 213–223, Jan. 2020.
[13] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Goncalves, W. R. Schwartz, and D. Menotti, "A robust real-time automatic license plate recognition based on the YOLO detector," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, Art. no. 18165770.
[14] S. M. Silva and C. R. Jung, "License plate detection and recognition in unconstrained scenarios," in *Proc. Conf. Comput. Vis.* Munich, Germany: Springer, 2018, pp. 593–609.
[15] 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
[16] 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)*, Beijing, China, Aug. 2018, pp. 3698–3703.
[17] Y. L. Yuan, W. B. Zou, Y. Zhao, X. Wang, X. F. 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. 2016.
[18] 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.
[19] A. H. Ashtari, M. J. Nordin, and M. Fathy, "An iranian license plate recognition system based on color features," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 4, pp. 1690–1705, Aug. 2014.
[20] 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.
[21] B. Li, B. Tian, Y. Li, and D. Wen, "Component-based license plate detection using conditional random field model," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1690–1699, Dec. 2013.
[22] D. F. Llorca, C. Salinas, M. Jimenez, I. Parra, A. G. Morcillo, R. Izquierdo, J. Lorenzo, and M. A. Sotelo, "Two-camera based accurate vehicle speed measurement using average speed at a fixed point," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 2533–2538.
[23] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525.

[24] M. A. Rafique, W. Pedrycz, and M. Jeon, "Vehicle license plate detection using region-based convolutional neural networks," *Soft Comput.*, vol. 22, no. 19, pp. 6429–6440, Oct. 2018.
[25] 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
[26] 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.
[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–14.
[28] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.
[29] H. Li, P. Wang, and C. Shen, "Towards end-to-end text spotting with convolutional recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 5248–5256.
[30] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "EAST: An efficient and accurate scene text detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2642–2651.
[31] P. He, W. Huang, T. He, Q. Zhu, Y. Qiao, and X. Li, "Single shot text detector with regional attention," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 3066–3074.
[32] L. Xing, Z. Tian, W. Huang, and M. Scott, "Convolutional character networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 9125–9135.
[33] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, Oct. 2019, pp. 9626–9635.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[35] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," 2019, *arXiv:1911.08287*. [Online]. Available: http://arxiv.org/abs/1911.08287
[36] W. Zhou, H. Li, Y. Lu, and Q. Tian, "Principal visual word discovery for automatic license plate detection," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4269–4279, Sep. 2012.

QIUYING HUANG received the M.S. degree in software engineering from the Beijing Institute of Technology, Beijing, China, in 2013. He is currently pursuing the Ph.D. degree with the Faculty of Information Technology, Macau University of Science and Technology, Macau, China. His research interests include image processing and artificial intelligence.



ZHANCHUAN CAI (Senior Member, IEEE) received the Ph.D. degree from Sun Yat-sen University, Guangzhou, China, in 2007.

From 2007 to 2008, he was a Visiting Scholar with the University of Nevada at Las Vegas, NV, USA. He is currently a Professor with the Faculty of Information Technology, Macau University of Science and Technology, Macau, China, where he is also with the State Key Laboratory of Lunar and Planetary Sciences. His research interests include image processing and computer graphics, intelligent information processing, multimedia information security, and remote sensing data processing and analysis.

Dr. Cai is a member of the Association for Computing Machinery, the Chang'e-3 Scientific Data Research and Application Core Team, and the Asia Graphics Association. He is also a Distinguished Member of the China Computer Federation. He was a recipient of the Third prize of the Macau Science and Technology Award-Natural Science Award, in 2012, the BOC Excellent Research Award from the Macau University of Science and Technology, in 2016, the Third Prize of the Macau Science and Technology Award-Technological Invention Award, in 2018, and the Second Prize of the Teaching Achievement Award from the Macau University of Science and Technology, in 2020.

TING LAN received the M.S. degree from the University of Macau, Macau, China, in 2014, and the Ph.D. degree from the Macau University of Science and Technology, Macau, in 2019.

He is currently a Postdoctoral Fellow with the Faculty of Information Technology, Macau University of Science and Technology. His research interests include image processing, data processing and analysis, and computer graphics.

Dr. Lan was a recipient of the First Prize at the 14th China Postgraduate Mathematical Contest in Modeling, China Academic Degrees, and the Graduate Education Development Center and China Graduate Mathematical Contest in Modeling Committee, in 2017, and the Third Prize of the Macau Science and Technology Award-Technological Invention Award, in 2018.