Hindawi
Complexity
Volume 2020, Article ID 5137056, 11 pages
https://doi.org/10.1155/2020/5137056

Research Article

Predicted Anchor Region Proposal with Balanced Feature Pyramid for License Plate Detection in Traffic Scene Images

Hoanh Nguyen

Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Correspondence should be addressed to Hoanh Nguyen; nguyenhoanh@iuh.edu.vn

Received 26 December 2019; Revised 12 May 2020; Accepted 26 May 2020; Published 16 June 2020

Academic Editor: Hassan Zargarzadeh

License plate detection is a key problem in intelligent transportation systems. Recently, many deep learning-based networks have been proposed and achieved incredible success in general object detection, such as faster R-CNN, SSD, and R-FCN. However, directly applying these deep general object detection networks on license plate detection without modifying may not achieve good enough performance. This paper proposes a novel deep learning-based framework for license plate detection in traffic scene images based on predicted anchor region proposal and balanced feature pyramid. In the proposed framework, ResNet-34 architecture is first adopted for generating the base convolution feature maps. A balanced feature pyramid generation module is then used to generate balanced feature pyramid, of which each feature level obtains equal information from other feature levels. Furthermore, this paper designs a multiscale region proposal network with a novel predicted location anchor scheme to generate high-quality proposals. Finally, a detection network which includes a region of interest pooling layer and fully connected layers is adopted to further classify and regress the coordinates of detected license plates. Experimental results on public datasets show that the proposed approach achieves better detection performance compared with other state-of-the-art methods on license plate detection.

1. Introduction

License plate recognition plays an important role in intelligent transport systems, traffic control, vehicle parking, traffic management, and many other fields. A license plate recognition includes two stages: license plate detection and license plate recognition. License plate detection locates exactly license plates in image, while license plate recognition segments and identifies each character on the detected license plate. License plate detection plays a crucial role in the performance of the whole system because exactly locating license plate will increase the accuracy of recognition stage. Thus, many approaches have been proposed for license plate detection. Previous approaches can be divided into two groups: traditional approaches and deep learning-based approaches. Traditional approaches use handcraft features such as colour, edge, character, and texture to locate license plate in image. Traditional approaches work well under controlled conditions. However, in difficult conditions such as distortion, blurring, and complex backgrounds, the performance of these approaches is still limited.

Recently, deep learning-based object detectors such as faster R-CNN [1], SSD [2], and YOLOv3 [3] have achieved significant improvements on general object detection compared with traditional frameworks. However, these object detectors are based on single scale feature map for detecting objects with different scales or used multiscale feature map of the base network with less semantic information, thus limiting the detection performance of detecting multiscale objects and objects in difficult conditions. To further improve the detection performance, many frameworks which improve semantic information at each feature map such as FPN [4], RetinaNet [5], MS-CNN [6], and DSSD [7] have been proposed and achieved better detection results compared with the baseline framework. However, these deep networks have not yet been studied in license plate detection.

Motivated by the above research ideas, this paper proposes a novel deep learning-based framework based on faster R-CNN for license plate detection. In the first stage, a balanced feature pyramid generation module is used to generate balanced feature pyramid of which each feature level obtains equal information from other feature levels, thus enhancing the semantic information of each feature map. Furthermore, a novel multiscale region proposal network with predicted location anchor scheme is designed to generate good proposals. In the second stage, a detection network which includes a region of interest pooling layer and fully connected layers is used to further classify and regress the bounding box of detected license plates. Experimental results on public datasets show that the proposed framework achieves better detection accuracy than state-of-the-art methods on license plate detection. The main contributions of this paper can be summarized as follows:

(i) This paper proposes a balanced feature pyramid generation module to generate balanced feature maps, of which each integrated feature level possesses balanced information from each resolution, thus enhancing the semantic information of each feature map. In addition, this paper adopts ResNet-34 as the base network for generating the base feature maps, which can improve the detection performance compared with VGG architecture.

(ii) For generating proposals, this paper proposes a novel anchor generation scheme based on guided anchoring scheme for generating high-quality proposals. This scheme is integrated into each branch of the multiscale region proposal network to obtain a set of high-quality proposals.

(iii) Using a two-stage strategy with balanced feature maps and multiscale region proposal network, the proposed approach is evaluated on public datasets and obtains better detection accuracy than other state-of-the-art methods on license plate detection.

The remaining of this paper is organized as follows. Section 2 reviews the related work. Section 3 details the proposed framework. Section 4 provides the experimental results and comparison between the proposed method and other methods on public datasets. Finally, the conclusions and future works are drawn in Section 5.

2. Related Work

2.1. Deep Learning-Based Object Detection. With the fast development of deep learning, many deep learning-based object detectors have been proposed and achieved significant improvements compared with traditional methods. Those deep learning-based object detectors can be divided into two groups: two-stage framework such as fast R-CNN [8], faster R-CNN [1], and R-FCN [9] and one-stage framework such as SSD [2], YOLO [10], and YOLOv3 [3]. Faster R-CNN introduced Region Proposal Network (RPN), a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. R-FCN proposed a fully convolutional networks with almost all computation shared on the entire image. The position-sensitive score maps are established to address a dilemma between translation invariance in image classification and translation variance in object detection. SSD proposed a one-stage network that predicts category scores and box offsets for a fixed set of default bounding boxes using small convolutional filters applied to different feature maps at different scales. YOLOv3 proposed an improved framework which detects objects at different feature maps to increase the detection accuracy of small-scale objects. The above-mentioned object detection frameworks achieved better accuracy compared with traditional frameworks. However, they used single scale feature map for detecting objects with different scales or used multiscale feature map with less semantic information from the base convolution layers, thus limiting the detection performance of detecting multiscale objects or objects in difficult conditions.

Recently, many enhanced frameworks have been proposed to improve the detection performance by using different enhanced feature maps at different scales such as FPN [4], RetinaNet [5], MS-CNN [6], DSSD [7], and ION [11]. FPN proposed to augment a standard convolutional network with a top-down pathway and lateral connections so the network efficiently constructs a rich, multiscale feature pyramid from a single resolution input image. Each level of the pyramid can be used for detecting objects at a different scale. RetinaNet proposed novel Focal Loss function which focuses training on a sparse set of hard examples to address the class imbalance issue. MS-CNN consists of a proposal subnetwork and a detection subnetwork for detecting objects at multiple output layers. DSSD introduced deconvolution module to generate enhanced feature maps from input feature maps and improves the detection performance of small objects. ION presented an object detector that exploits information both inside and outside the region of interest.

2.2. License Plate Detection. Previous approaches on license plate detection can be divided into two groups: traditional approaches and deep learning-based approaches. Traditional approaches for license plate detection are usually based on handcraft features of license plate such as colour, edge, texture, and character to locate license plate in image. Raghunandan et al. [12] proposed a new Riesz fractional model to improve low quality license plate images affected by multiple factors. A modified MSER algorithm is then used for character candidate detection. Yuan et al. [13] proposed a novel image downscaling method for license plate detection which can substantially reduce the size of the image without sacrificing detection performance. In addition, a novel line density filter is designed for extracting license plate candidates. Gou et al. [14] used morphological operations, various filters, different contours, and validations for detecting coarse license plate. Then character-specific ERs are selected as character regions through a Real AdaBoost classifier with

decision trees. Ashtari et al. [15] proposed a vehicle license plate recognition system based on a modified template-matching technique by the analysis of target colour pixels to detect the location of a license plate, along with a hybrid classifier that recognizes license plate characters.

With the development of deep learning recently, many methods for license plate detection based on deep learning have been proposed. Kim et al. [16] used faster R-CNN framework to locate vehicle regions. Then, the hierarchical sampling method is used for generating license plate candidates from vehicle regions. Bulan et al. [17] proposed a weak sparse network of winnows classifier trained with successive mean quantization transform features to extract candidate regions and a strong readable/unreadable CNN classifier to classify those candidate regions. Xie et al. [18] proposed a preprocessing algorithm to strengthen the contrast ratio of original car image at the first stage. At the second stage, the integral projection method is used to verify the true plate. Finally, a new feature extraction model is designed to complete accurate recognition of the license plate characters. Zou et al. [19] proposed to use shallow CNN to quickly remove most of the background regions to reduce the computation cost. Deep CNN is then used to detect license plate in the remaining regions. Xie et al. [20] introduced a new MD-YOLO model for multidirectional car license plate detection. The proposed model could elegantly solve the problem of multidirectional car license plate detection and could also be deployed easily in real-time circumstances because of its reduced computational complexity compared with previous CNN-based methods. Han et al. [21] proposed novel and effective strategies to tightly enclose the multioriented license plates with bounding parallelograms and detect license plates with multiple scales. The proposed method outperformed existing approaches in terms of detecting license plates with different orientations and multiple scales.

3. Methodology

Figure 1 illustrates the overall architecture of the proposed framework. The proposed framework is based on faster R-CNN [1], a popular two-stage general object detector. As shown in Figure 1, a base network based on ResNet-34 [22] architecture is first adopted to generate the base convolution feature maps. Enhanced feature pyramid is then generated from the base feature maps as in FPN [4]. To balance semantic features of low-level and high-level information in each level of enhanced feature pyramid, a balanced feature pyramid generation module is added to generate balanced feature pyramid. In the multilevel region proposal network (RPN), a novel predicted location anchor scheme is designed to generate high-quality proposals. Finally, a detection network is used to further classify and regress the bounding box of detected license plates. Details of each module will be explained in the next sections.

3.1. Balanced Feature Pyramid Generation Module. The original faster R-CNN uses VGG-16 [23] as the base network. Ren et al. [1] showed that almost of the forward time is spent on the base network. Thus, using a faster base network can greatly improve the inference speed of the whole network. ResNet is an efficient architecture which presented a residual learning framework to ease the training of networks that are substantially deeper than previous networks. In [22], ResNet-34 achieved nearly as performance as ResNet-50 and ResNet-101 while being faster and simpler. Thus, this paper adopts ResNet-34 architecture as the base network to generate initial convolution feature maps. Compared with VGG-16, ResNet-34 is not only more accurate than VGG-16 but also faster than VGG-16 [24]. The architecture of ResNet-34 for ImageNet [25] is shown in Table 1.

As in [26], higher lever features in deeper layers of convolutional layers contain more semantic representation, while lower lever features in shallower layers could better describe the characteristics of the small-scale objects. However, shallow feature maps from the low layers of feature pyramid inherently lack fine semantic information for object recognition. Recently, many feature combination methods based on lateral connections such as FPN [4] and RetinaNet [5] have improved the performance of object detection over faster R-CNN and SSD. However, Pang et al. [27] showed that the sequential manner in above integration methods will make integrated features focus more on adjacent resolution but less on other resolutions. The balanced integrated features which possess balanced information from each resolution will significantly improve the detection performance. Thus, this paper proposes a balanced feature map generation module for generating balanced feature maps. The proposed module is based on FPN [4]. Figure 2 illustrates the architecture of the balanced feature map generation module.

First, let {C2, C3, C4, C5} represent the output of the last residual block for conv2, conv3, conv4, and conv5 block of the ResNet-34. The strides of these outputs are {4, 8, 16, 32} pixels, respectively, with respect to the input image. Following [4], a 1×1 convolutional layer is added on each output feature map of the base network to reduce channel depth. In the top-down path, coarser-resolution feature maps are upsampled by a factor of 2 by using the nearest neighbor upsampling operation. These upsampled features are then merged with the corresponding output feature maps of the base ResNet-34 by elementwise addition. Finally, to reduce the aliasing effect of upsampling, a 3×3 convolution layer is added on each merged feature map (except for M5) to generate the multiscale feature pyramid, denoted as {P2, P3, P4, P5}, which can be used for detecting objects at a different scale.

Next, to integrate multilevel features and preserve their semantic hierarchy, interpolation and max pooling operation are used on P5 and {P2, P3}, respectively, to resize {P2, P3, P5} to the same size of P4. Because the features from different convolution layers have different scale of values, directly integrating them will lead to the domination of the larger values. Thus, this paper adds L2 normalization layer on each of rescaled features to keep the feature values from different convolution layers on the

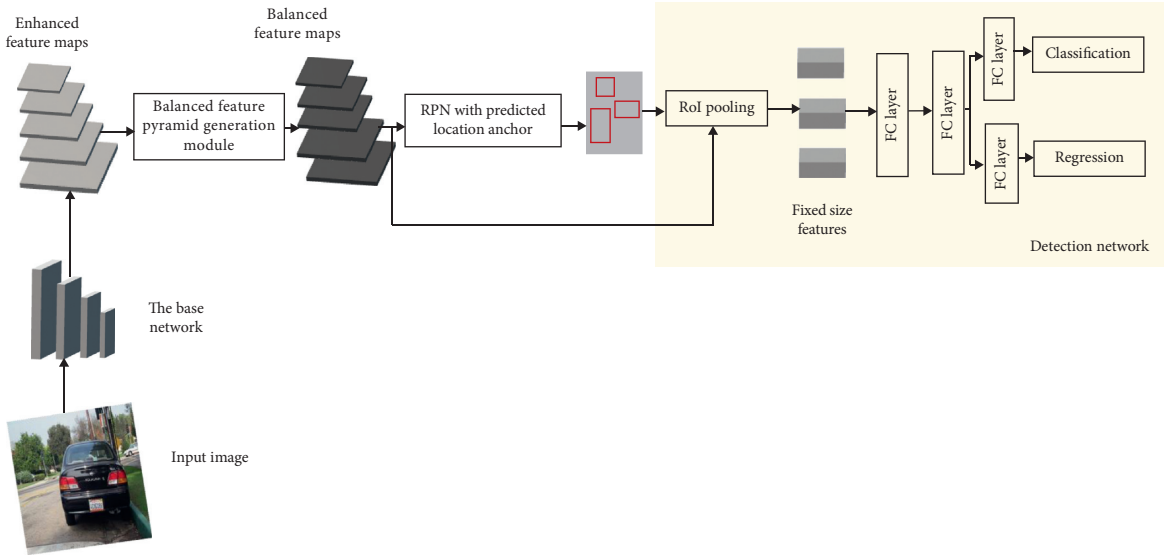


Figure 1: Overall pipeline of the proposed approach.

Table 1: The architectures of ResNet-34 for ImageNet.

Layer name	Kernel size	Output size
Conv1	$7\times7\times64$, stride 2 3×3 max pool, stride 2	112×112
Conv2	$\begin{bmatrix}3\times3\times64\\3\times3\times64\end{bmatrix}\times3$	56×56
Conv3	$\begin{bmatrix}3\times3\times128\\3\times3\times128\end{bmatrix}\times4$	28×28
Conv4	$\begin{bmatrix}3\times3\times256\\3\times3\times256\end{bmatrix}\times6$	14×14
Conv5	$\begin{bmatrix}3\times3\times512\\3\times3\times512\end{bmatrix}\times3$	7×7

Downsampling is performed by conv3-1, conv4-1, and conv5-1 with a stride of 2.

same scale. L2 normalization of a vector $y = \{y_1, y_2, \ldots, y_c\}$ is defined as follows:

$$\widehat{y} = \frac{y}{y_2} = \frac{y}{\left(\sum_{i=1}^{c} |y_i|^2\right)^{(1/2)}}, \quad (1)$$

where $\widehat{y}$ represents the normalized vector; y_2 represents the L2 normalization of y; and c represents the number of channels.

Based on the rescaled feature maps, balanced semantic feature map is generated as follows:

$$S = \frac{1}{n}\sum_{k=k_{\min}}^{k_{\max}} P_k', \quad (2)$$

where n represents the number of rescaled features ($n = 4$ in this paper); $k_{\min}$ and $k_{\max}$ represent the indexes of the lowest and the highest level in the rescaled features; P_k' represents the rescaled feature map at resolution level k.

Finally, the final balanced feature maps {F2, F3, F4, F5} are received by rescaling the semantic feature map in reverse procedure. More specifically, F5 is obtained by using max pooling operation on balanced semantic feature map, and {F2, F3} are obtained by using interpolation operation on balanced semantic feature map. With the proposed balanced feature pyramid generation module, each resolution in the final feature pyramid obtains equal information from other resolutions, thus balancing the information flow and leading the features more discriminative.

3.2. Multiscale Region Proposal Network with Predicted Location Anchor. In faster R-CNN, the RPN generates a set of anchor boxes at each location of the last convolution layer of the base network. The RPN then classifies these anchor boxes to object/background class and regresses the coordinates of these anchor boxes. There are 9 anchor boxes in total at each location of the feature map in original faster R-CNN framework. Each anchor box is associated with predefined scales and aspect ratios.

Wang et al. [28] showed that the uniform anchoring scheme in faster R-CNN can lead to significant computational cost because many anchor boxes are generated in regions where the objects of interest are unlikely to exist. In addition, a good of anchor box setting is needed for different problems to improve performance. Thus, this paper proposes a novel anchor generation scheme based on guided anchoring scheme [28] for generating anchor boxes.

Figure 3 illustrates the difference between the RPN in faster R-CNN and the proposed RPN with novel anchor generation scheme. In the proposed RPN, this paper first applies a 1×1 convolution layer at each feature scale of the balanced feature maps to create objectness score map. Elementwise sigmoid function is then adopted to convert the objectness score map to probability map. Based on the probability map on each scale, the positive regions where license plate candidate may possibly exist can be determined by selecting those locations whose corresponding probability values are above a predefined threshold. As in [28], this predefined threshold is set at 0.01 in this paper. Finally, a 3×3 convolution layer filter is applied across each sliding

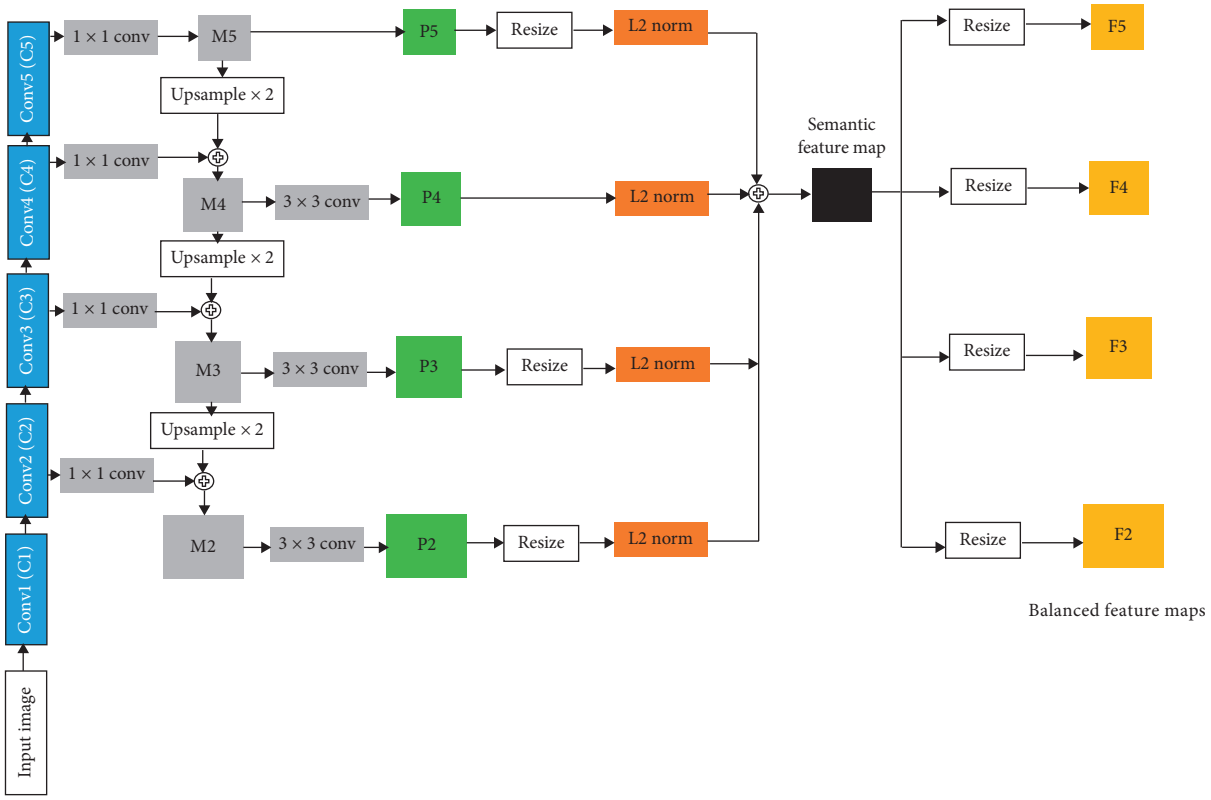


Figure 2: The architecture of the balanced feature map generation module.

position on the input feature map. At each position on the input feature map corresponding with positive regions on the probability map, the local features are extracted and concatenated along the channel axis and form a 256-d feature vector, which is then fed into two separate fully convolutional layers for license plate/background classification and box regression. The probability map can eliminate almost negative regions while still maintaining the same recall. Figure 4 shows the example results of the original RPN and the proposed RPN. Because the proposed RPN slides over all positive locations in all balanced pyramid levels, it is not necessary to have multiscale anchors on a specific level. Instead, this paper assigns anchors of a single scale to each level of the balanced pyramid according to the size and aspect ratio of license plates in the dataset summary table (Table 2). More specifically, this paper defines the anchors to have the height of {5, 10, 15, 20} pixels with an aspect ratio width/height = 5 on {F2, F3, F4, F5}, respectively.

3.3. Detection Network. Although the proposed multiscale RPN could work as a detector itself, it is not strong, since its sliding windows do not cover objects well. To increase detection accuracy, a license plate detection network is added. License plate detection network is used to classify proposals generated by the proposed RPN to license plate and background class and further refine the coordinates of detected license plate. The license plate detection network has a region of interest (RoI) pooling layer and two fully connected (FC) layers as shown in Figure 1.

Based on proposals generated by the multiscale RPN, RoI pooling layer is used to extract the fixed size feature patches from the balanced feature pyramid. As in [4], this paper selects the balanced feature map layer in the most proper scale to extract the feature patches based on the size of each proposal. More specifically, a proposal of width w and height h is assigned to the level F_k of the proposed feature pyramid with k being calculated as the following formula:

$$k = k_0 + \log_2\left(\frac{\sqrt{wh}}{224}\right), \tag{3}$$

where 224 is the canonical ImageNet pretraining size and k_0 is the target level on which a proposal with $w \times h = 224 \times 224$ should be mapped into. This paper sets $k_0 = 4$ as in [4].

The fixed size patches are then flattened into a vector and passed through the two 1024-d FC layers followed by ReLU. The encoded features are then fed into two separate linear transformation layers: license plate classification layer and bounding box regression layer. The license plate classification layer has two outputs, which indicate the softmax probability of each proposal as license plate/background. The license plate regression layer produces the bounding box coordinate offsets for each proposal.

3.4. Loss Function. The proposed framework is trained in an end-to-end fashion using a multitask loss function. Beside the conventional classification loss L_{cls} and regression loss L_{reg}, this paper adds additional loss function for the anchor

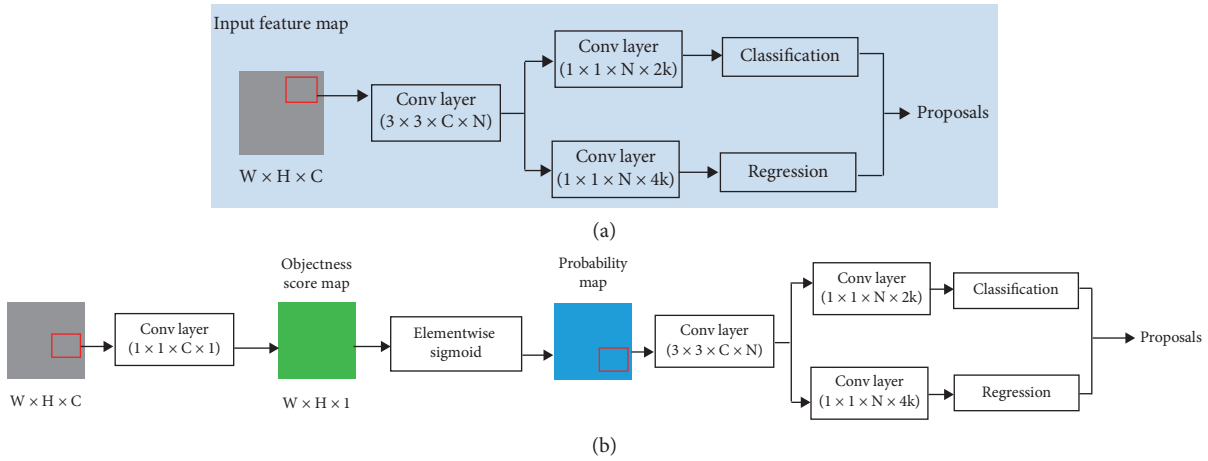


Figure 3: The architecture of the original RPN (a) and the proposed RPN with predicted location anchor (b). Multilayer RPN is used in this paper.

Figure 4: Example results of the original RPN (a) and the proposed RPN (b).

Table 2: Dataset summary.

Dataset		Number of images	Number of license plates	Image resolution	License plate height (in pixels)
PKU vehicle dataset	G1	810	810	1082 × 728	35–57
	G2	700	700	1082 × 728	30–62
	G3	743	743	1082 × 728	29–53
	G4	572	572	1600 × 1236	30–58
	G5	1152	1438	1600 × 1200	20–60
AOLP dataset	AC	681	681	352 × 240	25–70
	LE	757	757	640 × 480	28–80
	RP	611	611	320 × 240	30–70

box location prediction L_{loc}. Thus, the multitask loss function is defined as follows:

$$L = \sum L_{\text{cls}} + \sum L_{\text{reg}} + L_{\text{loc}}. \tag{4}$$

In (4), the binary logistic loss is used for box classification, and smooth L1 loss [1] is adopted for box regression. For training the anchor box location prediction branch in the proposed RPN, this paper follows the training scheme designed in [28]. More specifically, this paper denotes the ground-truth bounding box as $(x_{\text{gt}}, y_{\text{gt}}, w_{\text{gt}}, h_{\text{gt}})$, where $(x_{\text{gt}}, y_{\text{gt}})$ represents the center coordinates and $(w_{\text{gt}}, h_{\text{gt}})$ represents the size of the ground-truth bounding box. The ground-truth bounding box is mapped to the corresponding balanced feature map scale to obtain $(x'_{\text{gt}}, y'_{\text{gt}}, w'_{\text{gt}}, h'_{\text{gt}})$. Based on the obtained bounding box, the center box (CB), ignore box (IB), and outside box (OB) are defined as follows:

$$\text{CB} = \left(x'_{\text{gt}}, y'_{\text{gt}}, \partial_1 w'_{\text{gt}}, \partial_1 h'_{\text{gt}}\right), \tag{5}$$

$$\text{IB} = \left(x'_{\text{gt}}, y'_{\text{gt}}, \partial_2 w'_{\text{gt}}, \partial_2 h'_{\text{gt}}\right) - \text{CB}, \tag{6}$$

$$\text{OB} = \left(x'_{\text{gt}}, y'_{\text{gt}}, w'_{\text{gt}}, h'_{\text{gt}}\right) - \text{CB} - \text{IB}, \tag{7}$$

where $\partial_2 > \partial_1$. Pixels inside CB are assigned as positive locations, while pixels inside OB are assigned as negative locations. Otherwise, pixels inside IB are discarded in training samples. In the end, for each image in the training set, a binary label map where 1 represents a positive location and 0 represents a negative location is generated for training the anchor box location prediction branch. Note that each level of the balanced feature map should only assign objects of a specific scale range, so CB is only assigned on a feature map that matches the scale range of the targeted object. The same regions of adjacent levels in the balanced feature pyramid are set as IB. Finally, focal loss function [5] is adopted to train the anchor box location prediction branch for solving sample level imbalance problem.

4. Results and Discussion

In order to compare the effectiveness of the proposed approach with other state-of-the-art approaches on license plate detection, this paper conducts experiments on two public datasets: PKU vehicle dataset [13] and Application Oriented License Plate (AOLP) dataset [29]. The proposed approach is implemented on a Window system machine with Intel Core i7 8700 CPU, NVIDIA GeForce GTX 1080 GPU, and 16 Gb of RAM. TensorFlow is adopted for implementing deep CNN frameworks.

4.1. Dataset and Evaluation Metric. Two public license plate datasets are adopted to evaluate the performance of the proposed method in this paper, including PKU vehicle dataset [13] and AOLP dataset [29].

PKU vehicle dataset includes 3828 vehicle images captured from various scenes under diverse environment conditions. The image in this dataset is divided into five groups (G1-G5) corresponding to different configurations. More specifically, all images in G1, G2, and G3 group were taken on highways, while images in G4 group were taken on city roads, and images in G5 group were taken at intersections with crosswalks. The image in G4 group is captured during nighttime, while the image in other groups is captured during daytime. There is one Chinese license plate in each image of G1-G4 group, while multiple Chinese license plates are captured in each image of G5 group. For training the proposed network, this paper adopts CarFlag-Large dataset [30], which contains 46,0000 images with Chinese license plates.

AOLP dataset includes 2049 images of Taiwan license plates in various locations, time, traffic, and weather conditions. This dataset is categorized into three subsets: access control (AC) with 681 images; traffic law enforcement (LE) with 757 images; and road patrol (RP) with 611 images. AC refers to the cases that a vehicle passes a fixed passage at a reduced speed or with a full stop. LE refers to the cases that a vehicle violates traffic laws and is captured by a roadside camera. RP refers to the cases that the camera is installed or handheld on a patrolling vehicle, which takes images of vehicles with arbitrary viewpoints and distances. Each image contains one license plate. Since there is no standard split for AOLP dataset, this paper follows the same strategy as in [30] for training the proposed network. More specifically, this paper uses images from different subsets for training and test separately. In addition, data augmentation is conducted by rotation and affine transformation to increase the number of training images. In this paper, PKU vehicle dataset and AOLP dataset are adopted to evaluate the performance of the proposed approach and compare the detection results with the results of other state-of-the-art approaches. Table 2 shows the detailed descriptions of each dataset used in this paper.

For the evaluation metric, this paper follows the criterion used in [13] to evaluate the performance of the proposed method and other methods on the PKU vehicle dataset and AOLP dataset. More specifically, a detection is considered to be correct if the license plate is totally encompassed by the bounding box and the IoU between the detected license plate and the ground-truth license plate is at least 0.5.

4.2. Experimental Results on PKU Vehicle Dataset. In order to show the effectiveness of the proposed approach, this paper compares the performance results of the proposed method with the results of state-of-the-art license plate detection methods on PKU vehicle dataset, including the methods proposed by Zhou et al. [31], Li et al. [32], Yuan et al. [13], and Li et al. [30]. Zhou et al. [31] proposed to discover the principal visual word characterized with geometric context for each license plate character. With a new license plate image, the license plates are extracted by matching local features with principal visual word. Li et al. [32] used maximally stable extremal region detector to extract candidate characters in images. The exact bounding boxes of license plates are estimated through the belief propagation inference on conditional random field which are constructed on the candidate characters in neighborhoods. Yuan et al. [13] proposed a novel line density filter approach to extract license candidate regions, and a cascaded license plate classifier based on linear support vector machines using colour saliency features is designed to identify the true license plate from among the candidate regions. Li et al. [30] proposed an approach to address both detection and recognition of license plate using a single deep neural network.

Table 3 shows the comparison of detection results on PKU vehicle dataset. As shown in Table 3, the proposed approach achieves the best detection accuracy on PKU vehicle dataset. More specifically, in terms of average detection performance, the performance of the proposed method is improved by 9.53%, 8.23%, 2.06%, and 0.24% compared with methods proposed by Zhou et al. [31], Li et al. [32], Yuan et al. [13], and Li et al. [30], respectively. It should be noted that the performance of the proposed method surpasses the best of the reference methods proposed by Li et al. [30] by a significant margin on G5 group. Images in G5 group contain multiple license plates in difficult conditions such as large variance of scales, reflective glare, and blurry and are affected by defects. This result shows a strong ability of the proposed framework on detecting license plate in difficult conditions with a large

Table 3: Comparison of detection results on PKU vehicle dataset.

Method	Detection ratio (%)					
	G1	G2	G3	G4	G5	Average
Zhou et al. [31]	95.43	97.85	94.21	81.23	82.37	90.22
Li et al. [32]	98.89	98.42	95.83	81.17	83.31	91.52
Yuan et al. [13]	98.76	98.42	97.72	96.23	97.32	97.69
Li et al. [30]	99.88	99.71	99.46	99.83	98.68	99.51
Proposed approach	99.88	99.86	99.73	99.83	99.44	99.75

variance of scales. Figure 5(a) shows some examples of detection results of the proposed method on PKU vehicle dataset. As shown in Figure 5(a), the proposed algorithm is effective to detect license plates with different scales under different situations.

4.3. Experimental Results on AOLP Dataset. To further evaluate the effectiveness of the proposed framework, the performance of proposed approach is tested on AOLP dataset. Table 4 shows the comparison of detection results of the proposed method and methods proposed by Hsu et al. [29], Li et al. [33], and Li et al. [30]. Experimental results in Table 4 show that the proposed method achieves the best detection ratio on all three subsets compared to the previous methods. More specifically, in terms of average detection, the performance of the proposed method is improved by 4.42%, 2.23%, and 0.62% compared with methods proposed by Hsu et al. [29], Li et al. [33], and Li et al. [30], respectively. The experimental results demonstrate that the proposed balanced feature pyramid and predicted location anchor can effectively enhance feature representation power and boost the performance of license plate detection in difficult conditions. Figure 5(b) shows some examples of detection results of the proposed method on AOLP dataset. As can be observed, the proposed method can accurately locate small license plates as well as medium or large ones.

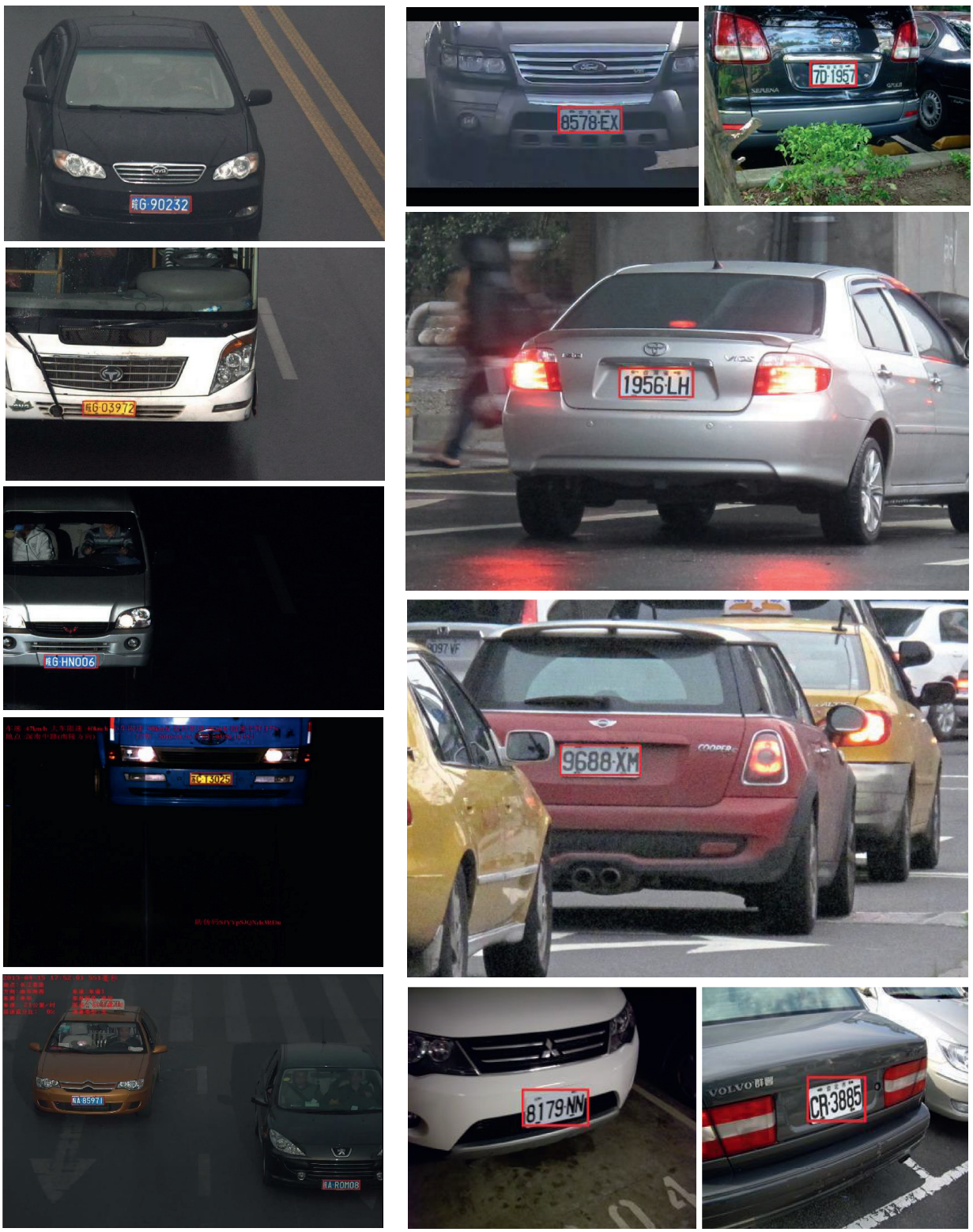
4.4. Ablation Experiments. To evaluate the effectiveness of each module in the proposed approach, this paper conducts several experiments on the Chinese City Parking Dataset (CCPD) [34] and compares the detection results with the results of original faster R-CNN [1] and FPN with faster R-CNN baseline [4] framework. CCPD dataset is a large publicly available labeled license plate dataset. It contains 25k independent license plate images under diverse illuminations, environments, and backgrounds. Each image has resolution of 720×1160 and contains one license plate. All images containing license plate are divided into 8 groups based on different conditions: CCPD-Base with 200k images; CCPD-FN with 20k images; CCPD-DB with 20k images; CCPD-Rotate with 10k images; CCPD-Tilt with 10k images; CCPD-Weather with 10k images; CCPD-Challenge with 10k images; CCPD-Blur with 5k images. As in [34], this paper adopts 100k images in CCPD-Base subset to train the proposed network and then evaluates the results on CCPD-Base, CCPD-DB, CCPD-FN, CCPD-Rotate, CCPD-Tilt, CCPD-Weather, and CCPD-Challenge.

In the first experiment, this paper replaces VGG-16 network in original faster R-CNN by the proposed balanced feature pyramid generation module. The RPN network is kept unchanged in this experiment. To show the effectiveness of the L2 normalization, L2 normalization layer in the balanced feature pyramid generation module is discarded in the second experiment. In the third experiment, this paper adds the proposed RPN network with the predicted location anchor module to replace the original RPN network. The VGG-16 is kept unchanged as the base network in this experiment. In the fourth experiment, this paper adds both the proposed RPN network with the predicted location anchor module and the proposed balanced feature pyramid generation module with L2 normalization layer to replace the original RPN network and VGG-16 architecture.

Table 5 shows the detection results of each proposed experiment on the CCPD dataset. As shown in Table 5, comparing with the original RPN in faster R-CNN framework, the proposed predicted location anchor scheme improves the average detection by 0.2%. By generating good proposals, the features for the detection network are more discriminative, thus improving the detection results. Comparing with the feature pyramid in FPN with faster R-CNN baseline, the proposed balanced pyramid generation module improves the average detection by 1.5%. It should be noted that there is no parameter added in the proposed module. With the proposed module, each level in the balanced feature pyramid obtains equal information from other levels, thus improving the detection performance of the detection network. Comparing with faster R-CNN and FPN with faster R-CNN baseline, the proposed approach improves the average detection by 4.5% and 1.9%, respectively. The comparison results indicate that the proposed framework is superior to both single scale and multiscale features for a region-based object detector. Furthermore, with L2 normalization layer added on each of rescaled features in the balanced feature pyramid generation module, the average detection is improved by 0.6% compared with the balanced feature pyramid generation module without L2 normalization. This result shows the effectiveness of the L2 normalization layer, which keeps the feature values from different convolution layers on the same scale.

5. Conclusions and Future Work

This paper proposes a novel deep learning-based framework for license plate detection. In the proposed framework, a balanced feature pyramid generation module based on ResNet-34 architecture is used to generate enhanced balanced feature pyramid of which each feature level obtains equal information from other feature levels. In addition, a multiscale region proposal network with predicted location anchor scheme is introduced to generate good proposals from each level of the balanced feature pyramid. With good proposals generated from balanced feature maps, the proposed approach shows significant improvements compared



(a) (b)

Figure 5: Examples of detection results of the proposed method on PKU vehicle dataset (a) and AOLP dataset (b).

with other approaches on license plate detection. The good performance of the proposed approach on license plate detection has a high reference value in the field of intelligent transport systems. For the future work, this paper will explore and compare more feature combination and multiscale detection methods, such as DeepLabv3+ [35] and MOSI-

Table 4: Comparison of detection results on AOLP dataset.

Method	Detection ratio (%)			
	AC	LE	RP	Average
Hsu et al. [29]	96.0	95.0	94.0	95
Li et al. [33]	98.38	97.62	95.58	97.19
Li et al. [30]	99.12	99.08	98.20	98.8
Proposed approach	99.41	99.34	99.51	99.42

Table 5: Detection results of each proposed network on CCPD dataset.

Network	Detection performance (%)							
	Base	DB	FN	Rotate	Tilt	Weather	Challenge	Average
Faster R-CNN	98.1	92.1	83.7	91.8	89.4	81.1	83.9	88.6
FPN with faster R-CNN baseline	99.2	95.4	87.5	93.0	91.3	85.4	86.3	91.2
Faster R-CNN + balanced feature pyramid	99.5	96.2	89.1	93.2	91.6	88.9	90.1	92.7
Faster R-CNN + balanced pyramid without L2-norm	99.3	96.2	88.9	93.1	91.0	87.6	88.3	92.1
Faster R-CNN + predicted anchor RPN	98.5	92.6	84.0	91.5	89.4	81.4	84.4	88.8
Faster R-CNN + balanced pyramid with L2-norm + predicted anchor RPN	99.5	96.4	90.1	93.2	91.8	89.7	91.2	93.1

LPD [21]. In addition, this paper will adopt the nonlocal module [36] to further refine the balanced semantic features. This step may enhance the integrated features and further improve the detection results.

Data Availability

The codes used in this paper are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2015.

[2] W. Liu, D. Anguelov, D. Erhan et al., "Single shot multibox detector," 2016, https://arxiv.org/abs/1512.02325.

[3] J. Redmon and F. Ali, "Yolov3: an incremental improvement," 2018, http://arxiv.org/abs/1804.02767.

[4] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944, Honolulu, HI, USA, July 2017.

[5] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999–3007, Venice, Italy, October 2017.

[6] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," 2016, http://arxiv.org/abs/1607.07155.

[7] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg, "Dssd: deconvolutional single shot detector," 2017, http://arxiv.org/abs/1701.06659.

[8] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, December 2015.

[9] J. Dai, Yi Li, K. He, and J. Sun, "R-fcn: object detection via region-based fully convolutional networks," *Advances in Neural information Processing Systems*, pp. 379–387, Mit Press, Cambridge, MA, USA, 2016.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," 2016, http://arxiv.org/abs/1506.02640.

[11] S. Bell, C. L. Zitnick, K. Bala, and R. Girshick, "Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2874–2883, Las Vegas, NV, USA, June 2016.

[12] K. S. Raghunandan, P. Shivakumara, H. A. Jalab et al., "Riesz fractional based model for enhancing license plate detection and recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 9, pp. 2276–2288, 2018.

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

[14] C. Gou, K. Wang, Y. Yao, and Z. Li, "Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1096–1107, 2016.

[15] A. H. Ashtari, M. J. Nordin, and M. Fathy, "An Iranian license plate recognition system based on color features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1690–1705, 2014.

[16] S. G. Kim, H. G. Jeon, and H. I. Koo, "Deep-learning-based license plate detection method using vehicle region extraction," *Electronics Letters*, vol. 53, no. 15, pp. 1034–1036, 2017.

[17] O. Bulan, V. Kozitsky, P. Ramesh, and M. Shreve, "Segmentation- and annotation-free license plate recognition with deep localization and failure identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 9, pp. 2351–2363, 2017.

[18] F. Xie, M. Zhang, J. Zhao, J. Yang, Y. Liu, and X. Yuan, "A robust license plate detection and character recognition

algorithm based on a combined feature extraction model and BPNN," *Journal of Advanced Transportation*, vol. 2018, Article ID 6737314, 14 pages, 2018.

[19] L. Zou, M. Zhao, Z. Gao, M. Cao, H. Jia, and M. Pei, "License plate detection with shallow and deep CNNs in complex environments," *Complexity*, vol. 2018, Article ID 7984653, 6 pages, 2018.

[20] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 507–517, 2018.

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

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, http://arxiv.org/abs/1409.1556.

[24] J. Huang, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3296-3297, Honolulu, HI, USA, July 2017.

[25] O. Russakovsky, J. Deng, H. Su et al., "Imagenet large scale visual recognition challenge," 2014, http://arxiv.org/abs/1409.0575.

[26] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*, Springer, Berlin, Germany, 2014.

[27] J. Pang, "Libra r-cnn: towards balanced learning for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019.

[28] J. Wang, K. Chen, S. Yang, C. C. Loy, and D. Lin, "Region proposal by guided anchoring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2965–2974, Long Beach, CA, USA, June 2019.

[29] G.-S. Hsu, J.-C. Chen, and Y.-Z. Chung, "Application-oriented license plate recognition," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 552–561, 2013.

[30] H. Li, P. Wang, and C. Shen, "Toward end-to-end car license plate detection and recognition with deep neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 1126–1136, 2019.

[31] W. Zhou, H. Li, Y. Lu, and Q. Tian, "Principal visual word discovery for automatic license plate detection," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4269–4279, 2012.

[32] B. Li, B. Tian, Y. Li, and D. Wen, "Component-based license plate detection using conditional random field model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1690–1699, 2013.

[33] H. Li and C. Shen, "Reading car license plates using deep convolutional neural networks and LSTMs," 2016, https://arxiv.org/abs/1601.05610.

[34] Z. Xu, W. Yang, A. Meng et al., "Towards end-to-end license plate detection and recognition: a large dataset and baseline," *Computer Vision—ECCV 2018*, Springer, Berlin, Germany, pp. 261–277, 2018.

[35] L.-C. Chen, Y. Zhu, P. George, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818, Munich, Germany, September 2018.

[36] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794–7803, Honolulu, HI, USA, July 2018.