



License Plate Recognition via Attention Mechanism

Longjuan Wang^{1,2}, Chunjie Cao^{1,2}, Binghui Zou^{1,2}, Jun Ye^{1,2,*} and Jin Zhang³

¹School of Computer Science and Cyberspace Security Hainan University, Haikou, 570228, China

²Key Laboratory of Internet Information Retrieval of Hainan Province, Haikou, 570228, China

³Hilbert College, Hamburg, NY, 14075, USA

*Corresponding Author: Jun Ye. Email: yejun@hainanu.edu.cn

Received: 29 May 2022; Accepted: 26 August 2022

Abstract: License plate recognition technology use widely in intelligent traffic management and control. Researchers have been committed to improving the speed and accuracy of license plate recognition for nearly 30 years. This paper is the first to propose combining the attention mechanism with YOLO-v5 and LPRnet to construct a new license plate recognition model (LPR-CBAM-Net). Through the attention mechanism CBAM (Convolutional Block Attention Module), the importance of different feature channels in license plate recognition can be re-calibrated to obtain proper attention to features. Force information to achieve the purpose of improving recognition speed and accuracy. Experimental results show that the model construction method is superior in speed and accuracy to traditional license plate recognition algorithms. The accuracy of the recognition model of the CBAM model is increased by two percentage points to 97.2%, and the size of the constructed model is only 1.8 M, which can meet the requirements of real-time execution of embedded low-power devices. The codes for training and evaluating LPR-CBAM-Net are available under the open-source MIT License at: <https://github.com/To2rk/LPR-CBAM-Net>.

Keywords: License plate; detection; recognition; CBAM; YOLO v5

1 Introduction

The automobile license plate (ALP) is similar to the human ID number. With the development of science and technology, License plate recognition (LPR) is used widely in various situations, such as parking management, toll collection, automatic toll collection in high-speed sections, traffic violation punishment, etc. LPR has developed from pattern recognition technology to deep learning technology. The researchers have been working on recognizing license plates faster and more accurately. There are two main methods in the existing LPR. The first method: The LPR process, is divided into three parts. Part I: License plate detection (LPD). There are edge detection methods [1–4], texture feature detection methods [5], color feature detection methods [6,7], character feature detection methods [8,9], etc. With the strengthening of computer computing power, LPD has gradually changed from the traditional method to deep learning. Representative models include the R-CNN (Region based Convolutional Neural Network) series [10,11]. This algorithm first generates candidate regions in



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

detection and then classifies them based on candidate regions with relatively high accuracy. The other model algorithm does not need to generate candidate regions in detection and directly regresses the target category and boundary. It has fast detection speed and represents. Representative models include SSD (Single Shot MultiBox Detector) [12], YOLO (You only look once) series [13–17]. Part II: Each character was segmented from the detected LP picture. Huang [18] proposed to use the unimodal or bimodal characteristics of character projection to segment characters. Brillantes et al. [19] used the 3-Class Fuzzy Clustering with Thresholding and Connected Component Analysis and were recognized using Template Matching. Olmí et al. [20] Proposed segmentation method based on fuzzy logic, Stefanovi et al. [21] Proposed character segmentation using K-mean clustering analysis. Part III: Character recognition for each segment. Baten et al. [22] proposed character recognition using fuzzy matching. Mohamad et al. [23] proposed character recognition by feature matching, He et al. [24–26] proposed a character recognition model based on a convolutional neural network. The three-part method had errors in each training process, resulting in excessive error accumulation. So, many researchers had started to carry out end-to-end LPD in recent years. The second method: The end-to-end LPD. The deep semantic segmentation network was used to determine the LP region of the input image, and then the segmentation was performed through the depth encoder-decoder network architecture. Finally, two independent CNN models were used to recognize the numbers and city signs respectively [27]. Li et al. [28] Proposed using a deep neural network for license plate detection and recognition. Tourani et al. Proposed training two independent Yolo v3 unify the network architecture for unified identification of LPD and CR [29].

As seen from the kinds of literature stated above, with the development of GPUs (Graphics Processing Unit) that support high computing performance, CNNs are becoming more and more common in license plate recognition. The latest research trends in detecting and recognizing car license plates focus on deep learning methods. The deep learning method had achieved good results in license plate recognition. Most of the recognition rates in the literature were above 95%, but some real-time systems have higher recognition rate and detection speed requirements.

In recent years, relevant studies proposed to use the attention mechanism to act on the feature map in CNN, and obtained the available attention information in the special map, that achieve better task effect [30]. Hu et al. proposed SE-Net [31], a network architecture utilizing an attention mechanism. The architecture established the interdependence between the modeling feature maps, obtained the importance of unique maps through learning, and updated the original data according to the fundamental importance of increasing the reputation of valuable features and reducing the significance of useless features, improving efficiency and correctness of the task. Woo et al. proposed a convolutional attention module CBAM [32], which aggregated spatial and channel attention information. To obtain more comprehensive and reliable attention, they synthesized the information to a certain extent. Methods based on attention mechanisms have achieved excellent results in many fields in recent years.

Therefore, based on the existing research. We proposed a license plate recognition method based on attention mechanism (LPR-CBAM-net). The method achieves efficient license plate detection and recognition suitable for Chinese vehicles. Our main contributions can be summarized as follows:

- 1) We propose to utilize the object detection model YOLOv5 as the detector. We fuse the attention mechanism CBAM with the detection network. To strengthen the ability of the model to extract features, to achieve the purpose of the model paying more attention to the detected target itself.
- 2) We propose a deep neural network method for automatically recognizing detected license plates without character segmentation. By comparing the recognition results of the two models

(LPRnet, Light-MutiCNN), the impact of the CBAM module incorporating the attention mechanism on the model is compared. Experiments show that the recognition rate of the license plate recognition model added to the CBAM model is 2% higher than that of the original model, and the parameters and size of the model remain unchanged.

- 3) We use CTC Loss as the loss function improves the problems of low positioning accuracy and slows recognition and regression during training.

The remaining of this paper is organized as follows. Section 2 reviews the related works. 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 Works

Deep learning has the characteristics of fast speed and high accuracy in LPDR. In recent years, the research trend of researchers has been basically on the method of deep learning. Li et al. [28] introduced CCPD, a large and comprehensive LP dataset, and proposed end-to-end license plate detection and recognition based on a deep neural network. This method was a single unified depth neural network, which could directly detect the license plate from the image and recognize the license plate simultaneously. The recognition speed exceeded 61fps, and the LPR accuracy rate reached 98.5%. Shivakumara et al. [33] proposed a CNN-RNN-based license plate recognition method. This method utilized the powerful recognition ability of a convolutional neural network to achieve feature extraction, and a recurrent neural network could extract contextual information based on past information. This method explored the combination of convolutional and recurrent neural networks to achieve the optimal combination, thereby improving the recognition performance. Omar et al. [27] proposed license plate detection and recognition based on cascaded deep learning. This method used the deep semantic segmentation network to determine the three license plate regions of the input image and then performs segmentation through the depth encoder-decoder network architecture. The determined license plate region was input into two independent convolutional neural network models for Arabic numeral recognition and city recognition. Three CNN models were used independently, one for license plate detection and two for city and digital character recognition. Silva et al. [34] proposed real-time license plate detection and recognition based on a deep convolution neural network. This method recognized the APL area through two scans on the same CNN and then used the second CNN to recognize the license plate characters. Vaiyapuri et al. [35] proposed the SSA-CNN model. The model was a deep learning model that combined the squirrel search algorithm and the convolutional neural network. Pustokhina et al. [36] proposed an intelligent transportation system based on an evolutionary neural network for license plate recognition. The improved Bernsen algorithm and connected component analysis CCA model were used for license plate detection and location in this method. OKM clustering technology was used for license plate segmentation, and the CNN model was used for license plate recognition. He et al. [37] proposed robust automatic recognition of Chinese License plates in natural scenes. The method mainly handled and detected severely distorted car license plates. First, an affine transformation on the warped image was performed to obtain a new image and then used the existing EfficientDet [38] as the vehicle detection network. A distortion-corrected LPD network was used for license plate detection, and a robust three-layer LPR network architecture was reconstructed for license plate recognition. Wang et al. [39] proposed Chinese license plate recognition in complex environments. This method directly recognizes license plates by constructing an end-to-end deep learning network. They extracted basic features through a residual network, and extracted multi-scale features through multi-scale networks. They were locating license plates and characterized

via a regression network, and Charactered recognition by classification network. A practical scheme based on batch normalization was used during training to speed up the training and achieved better execution efficiency and higher recognition accuracy. Liu et al. [40] proposed an improved YOLOv3 network to solve the problem of multiple license plate detection in complex scenes. Kessentini et al. [41] proposed a two-level deep neural network for license plate detection and recognition in complex scenes (multi-norm). The method consisted of two deep learning stages of networks. They used the YOLO v2 detector to complete the license plate detection in the first stage. They compared the segmentation-free method using a convolutional recurrent neural network and the joint detection and recognition method for license plate recognition in the second stage.

SENet is a channel attention mechanism proposed by Hu et al. [31] The network enhances important features and weakens unimportant features by paying attention to the relationship between channels and recalibrating channel features to enhance the representation ability of the network. The structure of SENet is shown in Fig. 1. Fsq compressed the $C \times H \times W$ vector into a $1 \times 1 \times C$ vector through global average pooling, which has a global receptive field; Fex is the excitation operation. The result of Fsq is nonlinearly transformed through two fully connected layers; Fscale is a feature redirection operation, and the result of Fex in the previous step is used as a weight to multiply the original feature vector $C \times W \times H$. Finally, the feature map after feature enhancement is output.

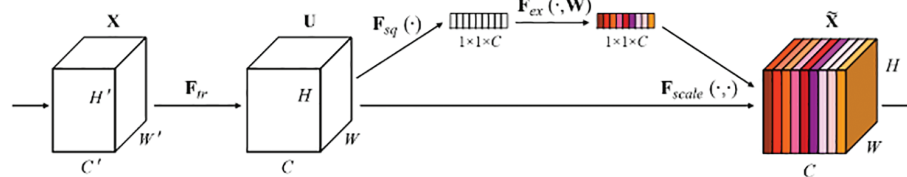


Figure 1: The overview of SENet

Woo et al. Proposed the convolutional block attention module (CBAM) [32], which constructed two sub-modules: The spatial attention module (SAM) and the Channel Attention Module (CAM). The model summarized the attention information of space and channel and synthesized the information to a certain extent, which obtained more comprehensive and reliable attention information and achieved a more reasonable allocation of computing resources. As shown in Fig. 2 below.

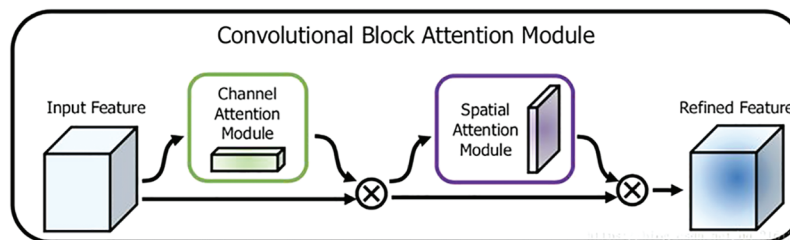


Figure 2: The overview of CBAM

The channel attention module passed the input feature map $F (H \times W \times C)$ through global max pooling and global average pooling based on width and height, respectively, and obtains two $1 \times 1 \times C$. The feature maps were sent to a two-layer neural network (MLP). The number of neurons in the first layer was C/r (r is the reduction rate), the activation function was Relu, the number of neurons in the second layer was C , and the two-layer neural network was shared. Finally, the element-wise based

summation was performed on the output of the features by the MLP. Then the sigmoid activation operation generated the final channel attention feature, namely $M_c(F)$ in formula (1). Among them, F_{avg} was the result of the global average pooling of feature maps, and F_{max} was the result of maximum global pooling.

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (1)$$

CBAM is a lightweight general module that can be seamlessly integrated into any CNN architecture with negligible overhead and could be trained end-to-end together with the base CNN.

He et al. [42] proposed an end-to-end approach to text recognition using alignment and attention, introducing a character attention mechanism that utilized spatial character information as additional supervision. The force was encoded into the model so that the RNN could focused on the current attention features when decoding to improve the efficiency of word recognition. Furthermore, through experiments, it was verified that the method of introducing the attention mechanism could indeed improved the recognition rate of words. Zhang et al. [43] proposed a multi-feature learning model with enhanced local attention for vehicle re-identification (MFELA). The model consisted of global and local branches. The global branched leverages the mid-to-high-level semantic features of ResNet50 to enhance the global representation. Furthermore, multi-scale pooling operations were used to obtain multi-scale information. Whereas the local branch utilizes the proposed Region Batch Dropblock (RBD), which encourages the model to learn discriminative features of different local regions and simultaneously batch randomly drop the corresponding identical regions during training to enhance the local region attention. Then, the features of the two branches are combined to provide a more comprehensive and unique feature representation. Sun et al. [44] propose a vehicle Re-ID model based on optimized DenseNet121 with joint loss. This model applied the SE block to automatically obtained the importance of each channel feature and assigned the corresponding weight to it. Then, features were transferred to the deep layer by adjusting the corresponding weights, which reduced the transmission of redundant information in the process of feature reuse in DenseNet121.

Based on the above research, this paper proposes to use the target detection model YOLOv5 as the license plate detection, combined with CNN, to perform license plate recognition on the car license plate to achieve end-to-end detection and recognition. The attention mechanism is integrated into the detection and recognition model. The model pays more attention to the detected target itself. Use the loss function in CTC loss [43] to improve the problems of low localization accuracy and slow regression during training.

3 Modeling Algorithms

To avoid significant accumulated error in typical three-step LPR methods and improve recognition accuracy, a deep convolution neural network recognition method based on an attention mechanism named LPR-CBAM-Net is proposed in this paper. LPR-CBAM-Net consists of two parts: license plate location module and character recognition module. The license plate localization module adopts the improved YOLOV5 as the detector, and the character recognition module consists of Light-MutiCNN of LPR-net [44] fused with the CBAM module. The input of Light-MutiCNN is a gray or color image of any size, and the output is the license plate number of the input image. The process of license plate detection and recognition is shown in Fig. 3. The picture containing the license plate is enhanced by mosaic and then sent to the Yolov5-CBAM module to detect the position of the license plate in the picture. Then the license plate image is obtained through the cropping operation, and

then the cropped license plate image is used as the input of Light-MutiCNN, and finally recognize the license plate number.

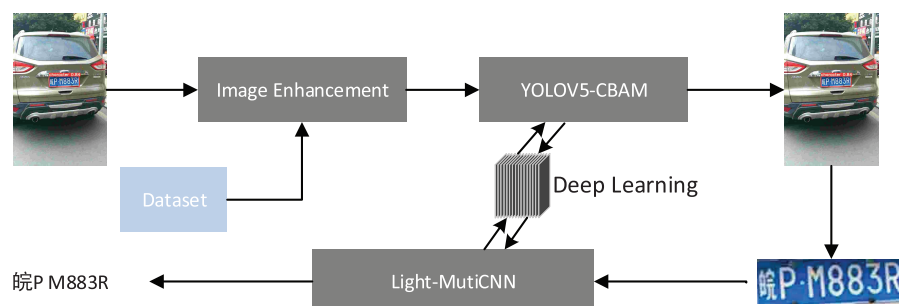


Figure 3: The process of license plate detection and recognition

3.1 YOLO-v5 License Plate Detection

The structure of YOLOv5s was mainly divided into four parts: Input, Backbone network, Neck network, and Head output, as shown in Fig. 4. The input was mainly data preprocessing, including Mosaic data enhancement, adaptive image filling, etc. The backbone network extracted different levels of features from images through deep convolution operations. The Neck network layer contained the feature pyramid FPN and the path aggregation structure PAN. The FPN transmitted semantic information from top to bottom in the network, PAN transmitted positioning information from bottom to top, and fused information from different network layers in Backbone to further improve detection capabilities. As the final detection part, the head output mainly predicted targets of different sizes on feature maps of different sizes.

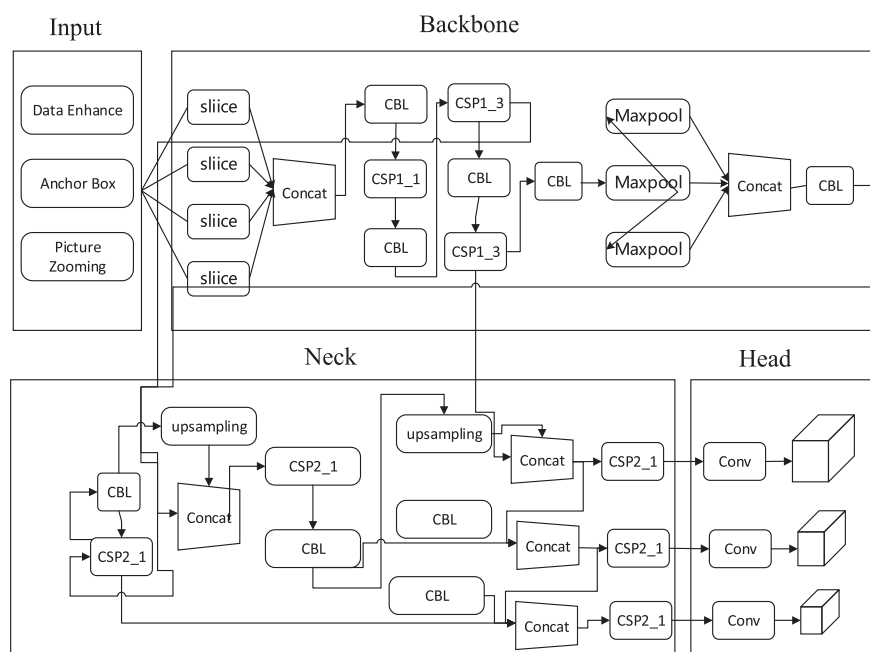


Figure 4: Structure of YOLO v5

In this paper, CBAM [32] is integrated into the Backbone module, where CBAM performs attention reconstruction, which plays a linking role. The specific structure diagram is shown in Fig. 5.

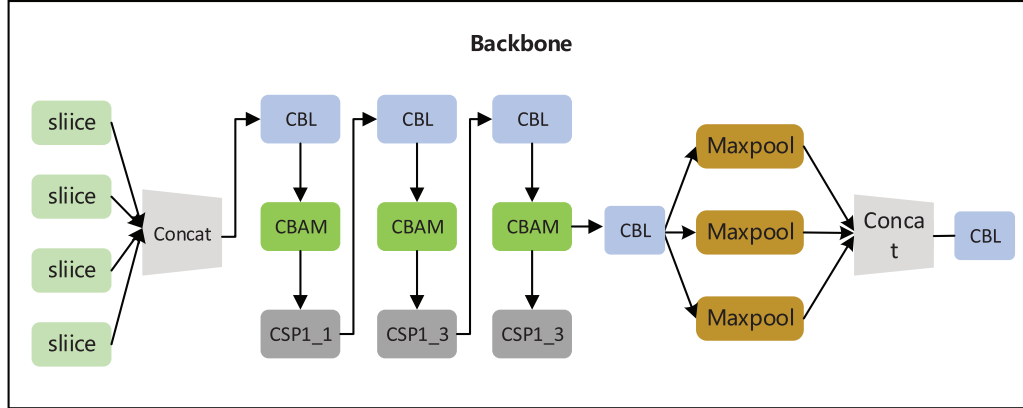


Figure 5: Structure of backbone integrates CBAM

3.2 Light-MutiCNN

Traditional license plate character recognition generally required pre-segmentation before recognition. However, when the license plate is tilted, blurred, and reflective, it is difficult for each character on the license plate to be accurately segmented. In order to perform license plate recognition more accurately, we designed an end-to-end LPR model. The model name is Light-MutiCNN, based on the existing LPRnet model and adds the attention mechanism CBAM. The size of the Light-MutiCNN model is only 1.8 MB, and the accuracy rate reaches 97.2%. The recognition speed is 3.7 ms/frame on NVIDIA GeForce RTX 3060 (GPU) and 3.9 ms/frame on AMD Ryzen 5 3600 (CPU).

We add CBAM after the first max-pooling layer, then add two CBAMs consecutively after the second max-pooling layer, which are connected by batch normalization and ReLU activation function. (The structure diagram is shown in Fig. 6.)

The backbone of Light-MutiCNN has 22 layers, including the convolution layer, the max-pooling layer, the batch normalization layer, the ReLU layer, the CBAM layer, and the Dropout layer. The structure is shown in Table 1. With the help of the global context modeling network GCNet [45], we fuse the features extracted from layers 2, 6, 13, and 22 of the backbone network, and the extracted features of these four layers are sent to the volume after global average pooling. Layer, and finally obtain the output tensor logits through dimension transformation.

4 Evaluation

The experimental configuration environment of this paper is as follows: Debian operating system, CPU for AMD Ryzen 5 3600, GPU for AMD Ryzen 5 3600 (8 GB memory), 16 GB memory, using Python 3.9.6, deep learning framework for Keras, and cuda10.0 installed to accelerate the calculation. TensorFlow is adopted for implementing deep CNN frameworks.

4.1 Dataset

We use the CCPD [46] dataset for model training. CCPD provides over 250k unique license plate images and detailed annotations. The resolution of each image is 720 (w) * 1160 (h) * 3 (channel).

A total of 19,996 ccpd_base data sets in CCPD are used for experiments, of which 90% are used as training sets, and 10% are used as test sets.

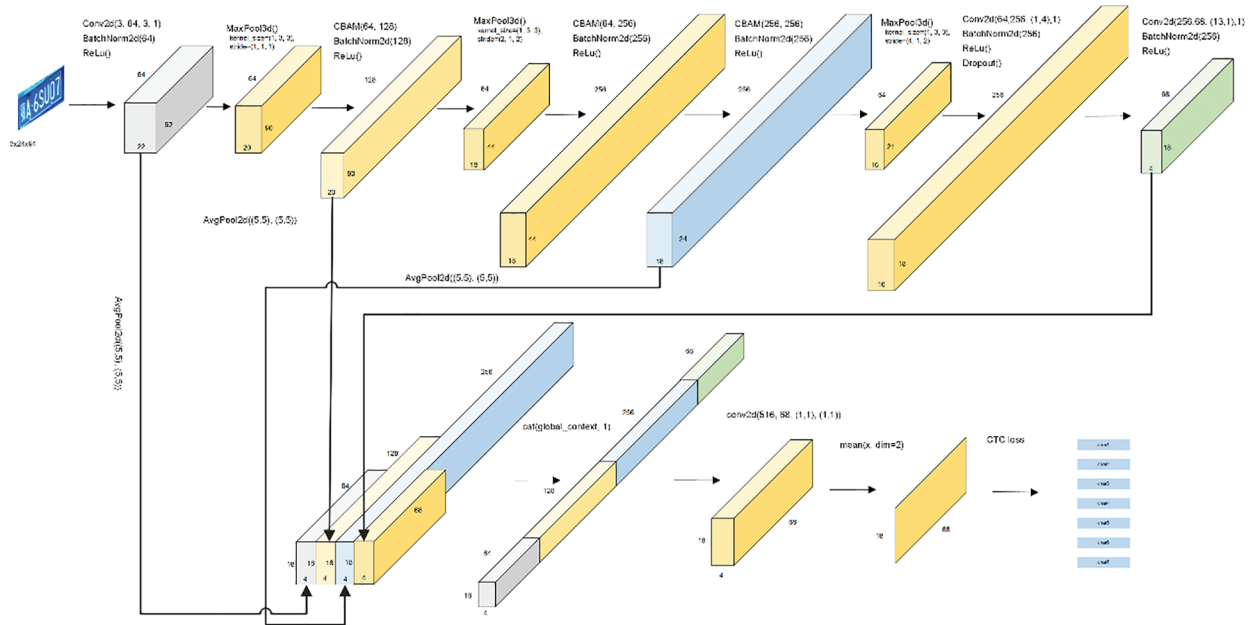


Figure 6: Light-MutiCNN structure diagram

Table 1: Description of Light-MutiCNN backbone architecture

Layer type	Filters	Stride	Output
Input	$3 \times 94 \times 24$ pixels RGB image		
Convolution	3×3	1	$64 \times 22 \times 92$
MaxPooling	3×3	1	$64 \times 20 \times 90$
CBAM	-	-	$128 \times 20 \times 90$
MaxPooling	3×3	(2, 1)	$64 \times 18 \times 44$
CBAM	-	-	$256 \times 18 \times 44$
CBAM	-	-	$256 \times 18 \times 44$
MaxPooling	3×3	(2, 1)	$64 \times 16 \times 21$
Dropout	0.5 ratio		
Convolution	1×4	1	$256 \times 16 \times 18$
Dropout	0.5 ratio		
Convolution	13×1	1	$68 \times 4 \times 18$
Output	$64 \times 4 \times 18$		

4.2 Loss Functions

The loss function of license plate detection includes location loss, confidence loss, and class loss. The position loss uses Eq. (2) [47]. The target box regression function considers three important geometric factors: overlapping area, center point distance, and aspect ratio. By adding a penalty factor, the aspect ratio of the prediction box and the real target box are considered. v is a parameter used to measure the consistency of the aspect ratio.

$$CIOU_Loss = 1 - CLOU = 1 - \left(IOU - \frac{Distance_2^2}{Distance_C^2} - \frac{v^2}{(1 - IOU) + v} \right) \quad (2)$$

Confidence loss and class loss adopt the FocalLoss proposed in the literature [48], Eq. (3). FocalLoss loss considers a solution strategy for the severe imbalance of positive and negative samples in target detection. The design idea of this loss function is similar to boosting, which reduces the impact of easy-to-classify examples on the loss function and focuses on the training of difficult-to-classify samples. p_i formally represents the confidence that the corresponding correct class is predicted. $(1-p_i)$ is a variable balance factor. γ is the adjustment factor, and $\gamma \geq 0$.

$$FL(p_i) = -(1 - p_i)^\gamma \log(p_i) \quad (3)$$

The character recognition loss function uses CTC loss [41]: Connectionist Temporal Classification loss, which is used to solve the classification of time series data.

4.3 Experimental Results and Comparisons

4.3.1 License Plate Test Results and Comparisons

Table 2 shows the detection results of each proposed experiment on the CCPD dataset. The addition of CBAM to YOLOV5 improves the average detection by 0.7% and 2.7% compared to the commonly used detection model RPNNet [46].

Table 2: Performance comparison with other detection method

Models	FPS	CCPD_base
YOLOV5	120	0.99
RPNNet	240	0.97
YOLOV5-CBAM	125	0.99742

4.3.2 Symbol Recognition Results and Comparison

In training Light-MutiCNN, we use the RMSprop optimizer to update the network parameters. The whole process is trained for 300 epochs, divided into three training sessions, each training 100 epochs, and the three initial learning rates are 0.001, 0.0001, and 0.00001, respectively. In order to obtain the optimal global solution, we use the cosine annealing strategy to adjust the learning rate in each training. Fig. 7 shows the change in the learning rate during the third training process.

Tables 3 and 4 list the correctly and incorrectly recognized license plates in several license plate recognition.

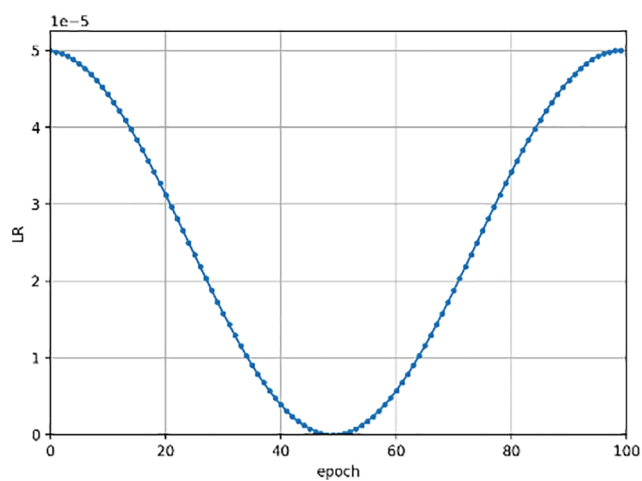


Figure 7: Change in learning rates

Table 3: Examples of successful license characters identification

License Plates	License Characters Identification
	皖 A818DB
	皖 A15926
	皖 A516X1
	苏 FE560P
	皖 A81A92

Table 4: Examples of successful license characters identification





License Plates	License Characters Identification
	皖 AVE902
	皖 AZ218F
	苏 L00184
	皖 AZ30S3

Table 5 shows that the Light-MutiCNN model with CBAM is improved by 2% compared to LPRNet [40] without the attention mechanism. The model's size is only 1.8 M, which can meet the real-time execution requirements of embedded low-power devices.

Table 5: Performance comparison with other recognition method

Models	Recognition accuracy	Size (MB)
RPnet	0.96	210.0
LPRNet	0.95	1.8
Light-mutiCNN	0.97248	1.8

The license plate image has a large background area, and the key feature is the pixel part occupied by the outline of the character. Compared with the background area, this part occupies little space in the whole image. The spatial attention in CBAM first takes the feature map obtained by the channel attention module as input, performs average pooling and maximum pooling in the channel dimension, combines the two results, and then undergoes a convolution operation. Thus, the compression of the channel dimension is achieved, and finally, the spatial attention map is obtained. To achieve the purpose of improving the license plate recognition rate.

5 Conclusion

In this work, we have shown that for LPR-CBAM-Net, one can utilize pretty small convolutional neural networks. Effective fusion of attention mechanism and CNN can improve recognition accuracy by as little as 2% without increasing parameters and model size Rate. As a future research direction, we also plan to extend our solution to the license plate detection of electric bicycles to solve the problem of illegal traffic control of electric bicycles.

Funding Statement: This paper was supported in part by the Natural Science Foundation of Hainan Province under Grant 621MS017 and the National Natural Science Foundation of China under Grant U19B2044.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] P. S. Ha and M. Shakeri, "License plate automatic recognition based on edge detection," in *Proc. Artificial Intelligence & Robotics*, Rawalpindi, Pakistan, pp. 170–174, 2016.
- [2] A. M. Al-Ghaili, S. Mashohor, A. R. Ramli and A. Ismail, "Vertical-edge-based car-license-plate detection method," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 26–38, 2013.
- [3] S. García, S. Aubert, S. Iraqui, I. Janbon, G. Ghigo *et al.*, "Candida albicans biofilms: A developmental state associated with specific and stable gene expression patterns," *Eukaryotic Cell*, vol. 3, no. 2, pp. 536–545, 2004.
- [4] M. R. Asif, Q. Chun, S. Hussain, M. S. Fareed and S. Khan, "Multinational vehicle license plate detection in complex backgrounds," *Journal of Visual Communication and Image Representation*, vol. 46, no. 1, pp. 176–186, 2017.

- [5] M. S. Al-Shemarry, Y. Li and S. Abdulla, "An efficient texture descriptor for the detection of license plates from vehicle images in difficult conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 553–564, 2020.
- [6] E. R. Lee, P. K. Kim and H. J. Kim, "Automatic recognition of a car license plate using color image processing," in *Proc. 1st Int. Conf. on Image Processing*, Austin, TX, USA, pp. 301–305, 1994.
- [7] M. A. Khan, M. Sharif, M. Y. Javed, T. Akram, M. Yasmin *et al.*, "License number plate recognition system using entropy-based features selection approach with SVM," *IET Image Processing*, vol. 12, no. 2, pp. 200–209, 2018.
- [8] J. Matas and K. Zimmermann, "Unconstrained licence plate and text localization and recognition," in *Proc. IEEE Intelligent Transportation Systems*, Vienna, Austria, pp. 225–230, 2005.
- [9] Q. Li, "A geometric framework for rectangular shape detection," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 4139–4149, 2014.
- [10] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580–587, 2014.
- [11] 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 & Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [12] L. Wei, A. Dragomir, E. Dumitru, S. Christian, R. Scott *et al.*, "SSD: Single shot multibox detector," in *Proc. European Conf. on Computer Vision*, Cham, Switzerland, Springer, pp. 483–498, 2016.
- [13] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.
- [14] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, Honolulu, Hawaii, USA, pp. 6517–6525, 2017.
- [15] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
- [16] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
- [17] G. Jocher, "YOLOv5," 2021. [Online]. Available: <https://github.com/ultralytics/yolov5>.
- [18] J. Huang, "Research on license plate image segmentation and intelligent character recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 18, pp. 2050–2062, 2019.
- [19] A. Brillantes, A. A. Bandala, E. P. Dadios and J. A. Jose, "Detection of fonts and characters with hybrid graphic-text plate numbers," in *Proc. TENCON 2018–2018 IEEE Region 10 Conf. IEEE*, Jeju, Korea, pp. 0629–0633, 2018.
- [20] H. Olmí, C. Urrea and M. Jamett, "Numeric character recognition system for Chilean license plates in semicontrolled scenarios," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 405–418, 2017.
- [21] H. Stefanovi, R. Veselinovi, G. Bjelobaba and A. Savi, "An adaptive car number plate image segmentation using K-means clustering," in *Proc. 2018-Int. Scientific Conf. on Information Technology and Data Related Research*, Singidunum University, Sinteza, pp. 74–78, 2018.
- [22] R. A. Baten, Z. Omair and U. Sikder, "Bangla license plate reader for metropolitan cities of Bangladesh using template matching," in *Proc. 2014 8th Int. Conf. on Electrical and Computer Engineering (ICECE)*, Dhaka, Bangladesh, pp. 776–779, 2015.
- [23] M. A. Mohamad, H. Hassan, D. Nasien and H. Haron, "A review on feature extraction and feature selection for handwritten character recognition," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 2, pp. 45–69, 2015.
- [24] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.

- [25] S. Angara and M. Robinson, "License plate character recognition using binarization and convolutional neural networks," in *Science and Information Conf.*, Cham, Springer, pp. 272–283, 2019.
- [26] A. Abdussalam, S. Sun, M. Fu, Y. Ullah and S. Ali, "Robust model for Chinese license plate character recognition using deep learning techniques," in *Proc. Int. Conf. in Communications, Signal Processing and Systems*, Singapore, Springer, pp. 121–127, 2018.
- [27] N. Omar, A. Sengur and S. G. S. Al-Ali, "Cascaded deep learning-based efficient approach for license plate detection and recognition," *Expert Systems with Applications*, vol. 11, no. 3, pp. 149–163, 2020.
- [28] 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, 2018.
- [29] A. Tourani, A. Shahbahrami, S. Soroori, S. Khazaei 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, no. 1, pp. 201317–201330, 2020.
- [30] Z. L. Zhu, Y. Rao, Y. Wu, J. N. Qi and Y. Zhang, "Research progress of attention mechanism in deep learning," *Journal of Chinese Information Processing*, vol. 43, no. 11, pp. 111–122, 2019.
- [31] J. Hu, L. Shen and G. Sun, "Squeeze-and-excitation networks," in *Proc. the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7132–7141, 2018.
- [32] S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proc. the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 3–19, 2018.
- [33] P. Shivakumara, D. Tang, M. Asadzadehkaljahi, T. Lu, U. Pal *et al.*, "CNN-RNN based method for license plate recognition," *CAAI Transactions on Intelligence Technology*, vol. 3, no. 3, pp. 169–175, 2018.
- [34] S. M. Silva and C. R. Jung, "Real-time license plate detection and recognition using deep convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 71, pp. 1072–1082, 2020.
- [35] T. Vaiyapuri, S. N. Mohanty, M. Sivaram, I. V. Pustokhina, D. A. Pustokhin *et al.*, "Automatic vehicle license plate recognition using optimal deep learning model," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1881–1897, 2021.
- [36] I. V. Pustokhina, D. A. Pustokhin, J. J. Rodrigues, D. Gupta, A. Khanna *et al.*, "Automatic vehicle license plate recognition using optimal K-means with convolutional neural network for intelligent transportation systems," *IEEE Access*, vol. 8, no. 1, pp. 92907–92917, 2020.
- [37] M. X. He and P. Hao, "Robust automatic recognition of Chinese license plates in natural scenes," *IEEE Access*, vol. 8, no. 1, pp. 173804–173814, 2020.
- [38] M. Tan, R. Pang and Q. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, USA, pp. 10781–10790, 2020.
- [39] D. Wang, Y. Tian, W. Geng, L. Zhao and C. Gong, "LPR-Net: Recognizing Chinese license plate in complex environments," *Pattern Recognition Letters*, vol. 130, no. 1, pp. 148–156, 2020.
- [40] D. Liu, Y. Wu, Y. He, L. Qin and B. Zheng, "Multi-object detection of Chinese license plate in complex scenes," *Computer Systems Science and Engineering*, vol. 36, no. 1, pp. 145–156, 2021.
- [41] Y. Kessentini, M. D. Besbes, S. Ammar and A. Chabbouh, "A two-stage deep neural network for multi-norm license plate detection and recognition," *Expert Systems with Applications*, vol. 136, no. 1, pp. 159–170, 2019.
- [42] T. He, Z. Tian, W. Huang, C. Shen, Y. Qiao *et al.*, "An end-to-end textspotter with explicit alignment and attention," in *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, Salt Lake City, UT, USA, pp. 5020–5029, 2018.
- [43] A. Graves, S. Fernández, F. Gomez and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *Proc. the 23rd Int. Conf. on Machine Learning*, New York, NY, USA, pp. 369–376, 2006.
- [44] S. Zherzdev and A. Gruzdev, "Lprnet: License plate recognition via deep neural networks," arXiv preprint arXiv:1806.10447, 2018.

- [45] Y. Cao, J. Xu, S. Lin, F. Wei and H. Hu, "Gcnet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. the IEEE/CVF Int. Conf. on Computer Vision Workshops*, Long Beach, USA, pp. 234–246, 2019.
- [46] Z. Xu, W. Yang, A. Meng, N. Lu, H. Huang *et al.*, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *Proc. the European Conf. on Computer Vision (ECCV)*, Munich, Germany, pp. 255–271, 2018.
- [47] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye *et al.*, "Distance-IoU loss: Faster and better learning for bounding box regression," *The AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 12993–13000, 2020.
- [48] T. Y. Lin, P. Goyal, R. Girshick and K. He, "Focal loss for dense object detection," in *Proc. the IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 2980–2988, 2017.