Wuhan University
Journal of Natural Sciences

2021, Vol.26 No.3, 227-234

Article ID 1007-1202(2021)03-0227-08
DOI 10.19823/j.cnki.1007-1202.2021.0028

Attention U-Net with Multilevel Fusion for License Plate Detection

☐ YAO Yao[1], XIONG Yujie[1,2†], HUANG Bo[1], YANG Jing[1]

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. Shanghai Key Laboratory of Multidimensional Information Processing, East China Normal University, Shanghai 200241, China

Abstract: In recent years, license plate recognition system (LPRS) is widely used in various places. Fast and accurate license plate detection is the first and critical step in LPRS. In order to improve the performance of license plate detection in complex environment, we propose a novel attention U-net with multilevel fusion (AUMF). At first, input images are fed to the network. Then, the feature maps of each level are generated by convolution operations of the original images. Before the feature connection, there are multi-layer splicing and convolution to detect more features. The attention mechanisms are used to retain the information of important regions. In order to ensure that the size of the input and output images are the same, down-sampling and up-sampling are employed to resize the feature mappings between the upper and lower levels. In the complex environment, the AUMF can accurately detect the license plate. To validate the effectiveness of the proposed method, we conducted a series of experiments on the AOLP dataset. The experimental results show that our approach effectively improves the performance of license plate detection under the three different license plate environments of AOLP dataset.

Key words: attention U-net; multilevel fusion; license plate detection

CLC number: TP 391

Received date: 2020-02-28

Foundation item: Supported by the National Natural Science Foundation of China (62006150), Shanghai Young Science and Technology Talents Sailing Program (19YF1418400), Shanghai Key Laboratory of Multidimensional Information Processing (2020MIP001), and Fundamental Research Funds for the Central Universities

Biography: YAO Yao, female, Master candidate, research direction: computer vision. E-mail: y54yaoyao@163.com

† To whom correspondence should be addressed. E-mail: xiong@sues.edu.cn

0 Introduction

With the rapid development of society driven by science and technology, the use of vehicles has greatly improved the traffic problem. License plate detection has been more and more important in the field of traffic management automation. License plate detection is widely applied in electronic toll station, automatic vehicle access management of parking lot and other various fields. A typical license plate recognition system mainly consists of the following two parts: license plate detection and license plate character recognition. The license plate detection is greatly affected by the weather, light, and other factor. If the license plate position is not well detected, the following parts cannot be carried out effectively as well. However, in the harsh working conditions, the existing detection methods cannot detect the position of the license plate accurately. Attention U-net (AUnet)[1] is commonly applied in medical images segmentation. It focuses on the remarkable features that are useful for a particular task and suppress irrelevant areas in the input image. Inspired by AUnet, we present a novel attention U-net with multilevel fusion (AUMF) in this paper.

The remainder of this paper is arranged as follows. In Section 1 we review related work of the license plate detection. Then, we describe the AUMF architecture in Section 2. In Section 3 and 4, detailed implementation and experimental results are provided. Section 5 concludes the paper with the scope of the future work.

1 Related Work

1.1 Traditional Methods

As the first step of the system, license plate detec-

tion aims to locate the license plate clearly and accurately from images and videos. Thus, it always has the significant impact on the recognition results. Many impressive researches have been carried out by the researchers to deal with the task of license plate detection over the last two decades.

Previous related work on license plate detection mainly depends on features. According to these various hand-crafted features, traditional methods can be broadly divided into four categories[2-5]: color-based methods, edge-based methods, character-based methods, and texture-based methods.

Color-based methods are based on the observation that the special colors of license plate such as blue and yellow are different from the background. Azad *et al*[3] proposed a color-based method by converting the RGB images into the HSI space for license plate detection. Deb *et al*[4] used the HSI color model to identify the license plate candidate regions. Chang *et al*[5] proposed a method based on the character edge composed of different colors to localize Taiwan license plates. However, these color-based methods often cannot perform well in the images with uneven illumination.

Edge-based methods are based on the edge information of license plates[6-10]. The shape of a license plate is usually rectangular with the special aspect ratio. And the edge density of license plate is higher than the other regions in the image, so researchers widely use the edge information to locate license plates. Chen *et al*[6] improved a prewitt arithmetic operator and then used an approach using horizontal and vertical projection to locate the upper and lower bounds of the edge position. Zheng *et al*[7] adopted the sobel arithmetic operator to extract out the vertical edges. When using edge-based methods, the sizes of the license plate regions are generally determined in advance. Though these kinds of methods often compute fast, they cannot be used to the images with inclined license plate as there are too many unwanted edges in images.

Character-based methods are used to detect license plate regions according to the rich character-based features in images. Because license plates are composed of a string of various characters and the character structure features are so obvious, many researchers have done a lot of work using character-based approaches on the task of license plate detection[8-11]. Lin *et al*[8] proposed an algorithm composed of two stages. They segmented out the characters on the license plate by using a significant map at first. At the second stage, sliding window approach was adopted to compute significant related features on these characters. Maximally Stable Extremal Regions (MSER) were used to extract the candidate characters. Wang *et al*[9] propose a novel character candidate extraction method based on superpixel segmentation and hierarchical clustering. Hao *et al*[10] extracted the regions of license plates based on MSER. At the same time, they classified these regions and finally located the license plates according to the SIFT features of extracted regions. Character-based methods are always reliable and tend to achieve a high recall. But the performances of these methods are also dependent on the presence of complex condition and interference factors in images.

Texture-based methods can also be used to detect license plates as plate regions generally have unconventional pixel texture distribution[12-14]. In this case, Deb *et al*[12] used a technique based on sliding concentric windows (SCW) for locating the candidate regions before their team used a HIS color model for identifying these regions. Texture-based methods often adopt more discriminative characteristics than edge-based methods and color-based methods, but they have high complexity in computing.

In recent years, with the advances in computing ability of computers, many methods based on machine learning are used in license plate detection including Boosting[15], Random Forest[16], Support Vector Machines (SVMs)[17, 18] and so on.

1.2 Deep Learning Methods

Traditional license plate detection methods are often based on hand-crafted image features. These hand-crafted image features have achieved good results in specific application scenarios, but are mostly dependent on the experience of feature designers. Furthermore, these traditional methods do not have good adaptability in complex environment. Fortunately, semantic segmentation technique with deep learning provides us a new way to achieve the goal of license plate detection.

Over the last decade, the deep learning technique based on convolutional neural network (CNN) has achieved great progress. Many problems in computer vision have been addressed by the methods using deep learning. For object detection and segmentation tasks, the methods based on deep learning perform so well that more and more researchers have started to design effective license plate detection algorithms based on deep learning. Thus U-net[19], Enet[20] and fully convolutional network (FCN)[21] are employed to the license plate detection and recognition tasks. Based on the high demand

of robustness, some methods based on CNN[22, 23] tend to employ the extracted image features instead of hand-crafted image features. Xie *et al*[24] improved a CNN-based framework named MD-YOLO network for positioning multi-directional license plates. They used a novel approach based on rotation angle prediction and a practical evaluation strategy to overcome the license plate problem in complicated scenarios. However, this method may perform badly when detecting the objects with small sizes. Xiang *et al*[21] proposed an improved fully convolutional network for Chinese license plate detection and their method achieved pretty good results in different application scenes. Li *et al*[25] used an approach based on the sliding window to detect the possible positions of license plate letters on all images through a CNN.

2 The Proposed Method

2.1 The Original AUnet

The main characteristics of the U-net[19] are a structure similar to a U-shaped and the skip connection. AUnet[1] is an improvement of the U-net model. The AUnet has a contraction path to get the context semantic information, and its corresponding is an extension path for precise feature location. Before the layer features of the contraction path are concatenated with the corresponding layer features of the extension path, the attention module is added to adjust the output characteristics of the contraction path.

2.2 The AUnet with Multilevel Fusion

The architecture of our proposed AUMF is shown in Fig. 1. There are five levels in AUMF. The left side of U shape is the contraction path, which is equivalent to the encoder in the network. Each convolution is added with a filling layer to ensure that the size of the images remains unchanged, and there is a positive linear Relu function after each convolution operation. The Relu function can be expressed as:

$$f(x)=\begin{cases}0, & x<0\\ x, & x>0\end{cases} \tag{1}$$

where x is the input and $f(x)$ the output. After the Relu function, max-pooling with a 2×2 window and stride 2 is performed. In each down-sampling process, the number of characteristic channels is doubled to prevent the loss of features in the down-sampling process. The max-pooling process can be expressed as

$$O=\frac{(\text{inwidth}+2\times\text{pad}-k)}{\text{str}}+1 \tag{2}$$

where O is the size of the output image, inwidth the size of the input image, pad the parameter of padding setting, k the size of the convolution kernel, and str the stride.

On the right side is the extended path, which is equivalent to the decoder, and 2×2 up-convolution is used for up-sampling operation layer by layer. The up-convolution operation is formulated as

$$O=\text{str}\times(\text{inwidth}-1)+k-2\times\text{pad} \tag{3}$$

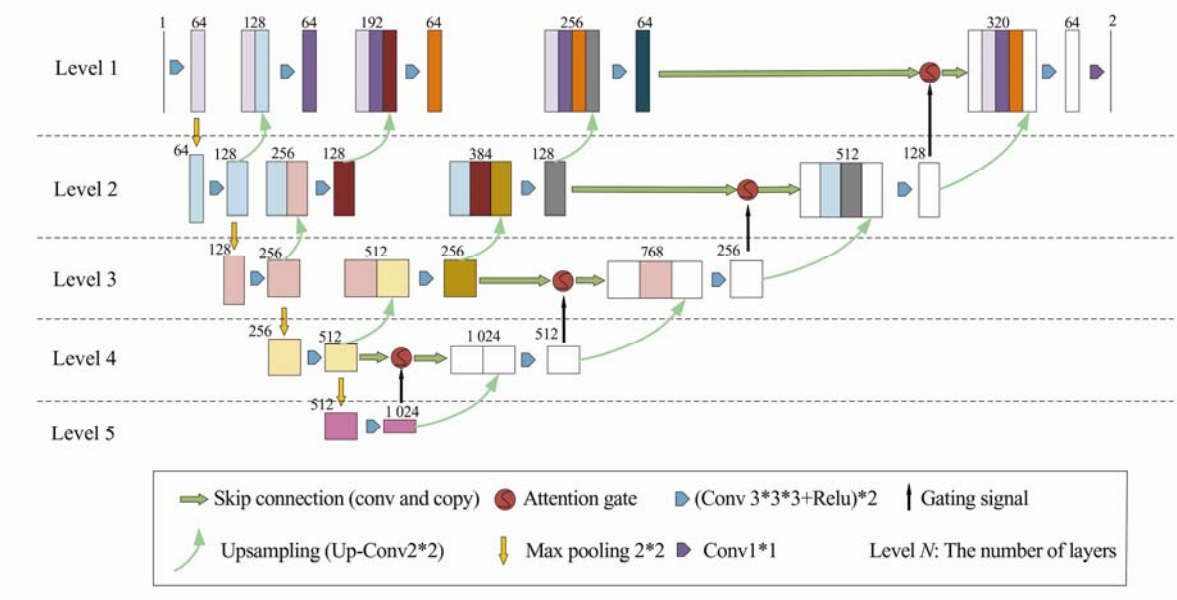

Fig. 1 AUMF architecture

In the same level, the rectangular boxes with the same color and shape denote the same feature maps; In different levels, the rectangular boxes with the same color and different shapes are obtained by up-sampling

A part of the feature connection graph is obtained from the same layer feature graph of contraction path and the upper layer feature graph of expansion path through the attention gate, and the other part is obtained by the up-sampling (2×2 up-convolution) of the upper feature map. After feature splicing, the number of channels is doubled, and then it goes through 3×3×3 convolution and a Relu function.

2.2.1 The multilevel fusion module

The information combination of the AUnet is completed by a simple crop and copy operation, which integrates the semantic and location information into the high dimensional feature map. The simple concatenating operation causes the loss of the meaningful information. However, there is still an abundant of texture and location information, so it is necessary to make improvements to address this problem. In order to extract more features, we add multi-layer splicing and convolution operations into the skip connection.

The process of the multilevel fusion is described as follows: the high-dimensional feature maps (Level N+1) are up-sampled, and combined with the low dimensional feature maps (Level N), whose combination is the new characteristic maps. And then a series of convolution operations are conducted on the new characteristic maps of multichannel number. As the number of layers increases, the number of feature maps merging is less. With the increase of the number of attributes to extract features, the depth-wise separable convolutions can save more parameters. The usage of batch normalization layer is to reduce parameters. Therefore, these improvements can obtain better generalization ability and reduce the calculation cost.

Those problems are well solved by adding multi-layer splicing and convolution operations before the skip path of AUMF, and the more features are properly combined into the depth feature maps.

2.2.2 The attention module

The attention module is shown in Fig. 2. x is the last feature map generated by level N, and g is the gate signal generated by the feature map of level N+1. x and g are convoluted by 1×1×1, so that the size of channels is the same, and keep the size of the images remains unchanged. x and g with the same number of channels are spliced and accumulated, and then through Relu function, a 1×1×1 convolution and Sigmoid function are used to obtain feature maps composed of 0 to 1.

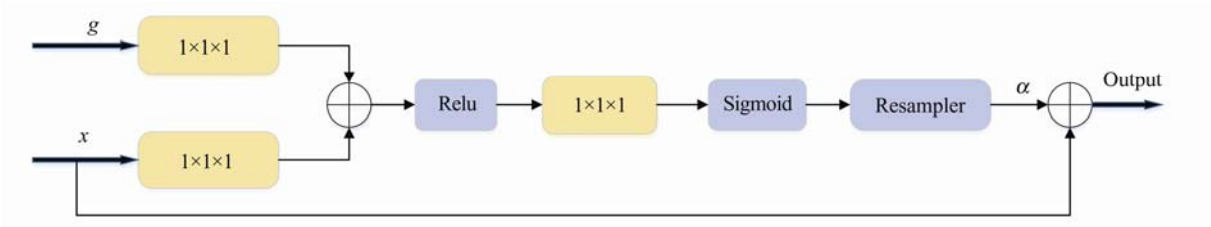

Fig. 2 An example of attention module

The feature map is multiplied by the input x of the skip connection to obtain the final output feature image. There is an attention coefficient α which gets a large value in the target area and a small value in the background area, which helps to improve the accuracy of image segmentation.

2.2.3 The cross-entropy loss function

In the training stage, the cross-entropy loss function is used to evaluate the error between the predicted value and the ground truth. The cross-entropy loss function is formulated as

$$\text{Loss} = -\frac{1}{n}\sum_{k}^{n}(y_k \log(\hat{y}_k) - (1-y_k)\log(1-\hat{y}_k)) \qquad (4)$$

The more obvious the differences between the predicted value and the y_k value are, the more nonlinearly the value of loss increases. The advantage of using cross-entropy loss function is that the model can make the predicted output value closer to the ground truth.

3 Experiments

3.1 Experimental Setups

The proposed method is all performed on software and hardware environment with i9-9900K 3.60GHz CPU, 32GB memory and an NVIDIA GeForce RTX 3080 10GB memory, Windows 10 with 64-bit system.

3.2 Datasets

In order to evaluate the performance of the network proposed in this paper, two license plate datasets are used to test the performance of the algorithm in license plate detection to better verify the license plate detection algorithm in complex background, non-uniform illumination conditions and bad weather. AOLP[26] dataset includes the horizontal angle and different angle plate, also in-

volves the city traffic plates under complicated background. AOLP dataset is a widely used open-source public license plate detection dataset consisting of 2 049 license plate images. AOLP dataset is divided into three sub datasets: AC (Access Control) datasets, LE (Law Enforcement) dataset and the RP (Road Patrol) dataset. Figure 3 shows the samples of the license plate images in three different environments. The above sub-datasets contain 681, 757 and 611 samples, respectively. In the AC dataset, almost all samples are license plate images in horizontal direction, which are collected when the speed limit passes through the intersection. The samples of the LE dataset are from urban traffic vehicles, including the interference factors of pedestrians, street lights and road signs in the complex road background. Samples of the RP dataset are slanted.

In this experiment, the samples of AOLP dataset were randomly divided into training set (85%) and test set (15%). In order to ensure the consistency of the distribution of the three sub-datasets of the training set and the test set, the data of the sub-datasets were randomly divided in the same proportion to ensure that the data distribution of the three sub-datasets of the training set and the test set was roughly the same.

Fig. 3 Image samples of the AOLP

3.3 Evaluation Criteria

In this paper, Mean-IOU (the mean intersection over union), Accuracy, Recall, Precision and F-score were used to evaluate the experimental results. These evaluation criteria are widely used in image segmentation.

1) Mean-IOU

Mean-IOU is an assessment method based on category calculation, which is not only used in semantic segmentation, but also can be used as a target detection index. The IOU of each category is calculated first, and then the average value is accumulated and the final evaluation is obtained.

$$\text{Mean-IOU} = \frac{1}{n+1}\sum_{i}^{n}\frac{(\text{TP})_i}{(\text{FN}+\text{FP}+\text{TP})_i} \tag{5}$$

where TP is the true positive, FN the false negative and FP the false negative.

2) Accuracy

The Accuracy refers to the ratio of the number of correctly predicted samples to the total number of predicted samples.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{6}$$

where TN is the true negative.

3) Recall

The Recall refers to the ratio of the number of correctly predicted positive samples to the total number of true positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

4) Precision

The Precision means the ratio of the number of correctly predicted positive samples to the number of predicted positive samples.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{8}$$

5) F-score

F-score is an evaluation index combined by Precision rate and Recall rate, which is equivalent to the har-

monic average of these two values. Any numerical change of Recall rate and accuracy will cause the change of F-score, and its expression is as follows:

$$\text{F-score}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precison+Recall}} \tag{9}$$

4 Results and Analysis

4.1 Analysis of Our Method

In the comparative experiments, we know that other methods also put the IOU as the standard of license plate detection, in which the IOU is set to a fixed value. The proposed method can carry out the category average directly according to the IOU value rather than setting fixed IOU values. The experimental results show that the license plate detection can achieve a good segmentation performance without setting the specified IOU. Table 1 shows the Mean-IOU and Accuracy of our approach in three different sub-datasets.

Table 1 The Mean-IOU and Accuracy of our method in three different license plate datasets

%

Dataset	Mean-IOU	Accuracy
AOLP-AC	98.26	99.89
AOLP-LE	91.53	98.91
AOLP-RP	97.55	99.87

4.2 Comparison of Different Methods

The AC dataset contains only horizontal license plate images, in which 103 images are used as the test set and 578 images are used as the training set. As shown in Table 2, the Precision of the proposed method on AC dataset is 99.34%, Recall is 98.88% and the F-score is 99.11%. All of them are higher than other methods. Figure 4 shows the detection results of AUMF on the AOLP-AC dataset.

In the LE dataset, there are license plate images with complex road background including urban road pedestrians, street lamps and road signs, in which 114 images are used for testing and 643 are used for training. As shown in Table 3, the precision rate of the proposed method on LE dataset is 97.09% and F-score is 95.44%. Figure 5 is the detection results of AUMF on the AOLP-LE dataset.

Table 2 Comparison of different methods in AOLP-AC

%

Method	Precision	Recall	F-score
Li *et al* [25]	98.53	98.38	98.45
Hsu *et al* [26]	91.00	96.00	93.43
Selmi *et al* [27]	92.60	96.80	94.65
Ours (AUMF)	**99.34**	**98.88**	**99.11**

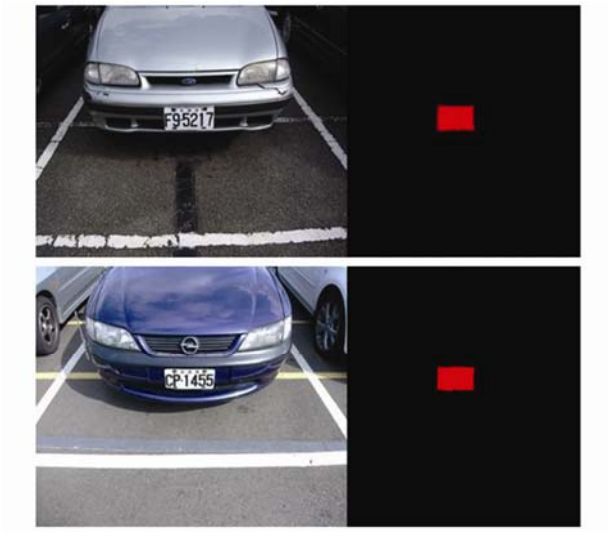

Fig. 4 Detection results of our method in AOLP-AC

The RP dataset contains slant license plate images from different angles. 92 images are used for testing and 519 images are used for training. As shown in Table 4, the Precision, Recall and F-score of our method on the RP dataset is 97.27%, 99.17%, and 98.21%. Figure 6 shows the detection results of AUMF on the AOLP-RP dataset.

Table 3 Comparison of different methods in AOLP-LE

%

Method	Precision	Recall	F-score
Li *et al* [25]	**97.75**	**97.62**	**97.68**
Hsu *et al* [26]	91.00	95.00	92.96
Selmi *et al* [27]	93.50	93.30	93.39
Ours (AUMF)	97.09	93.84	95.44

Fig. 5 Detection results of our method in AOLP-LE

Table 4 Comparison of different methods in AOLP-RP

%

Method	Precision	Recall	F-score
Li *et al* [25]	95.28	95.58	95.43
Hsu *et al* [26]	91.00	94.00	92.48
Selmi *et al* [27]	92.90	96.20	94.52
Ours (AUMF)	**97.27**	**99.17**	**98.21**

Fig. 6 Detection results of our method in AOLP-RP

5 Conclusion

The AUnet with multilevel fusion is an improvement of U-net. We apply the proposed method into the field of license plate detection. Our AUMF architecture can get more characteristic and obtain significant features under the effect of attention module and multilevel fusion. The experimental results show that the proposed method improves the performance of license plate detection in complex environments effectively.

References

[1] Oktay O, Schlemper J, Folgoc L L, *et al*. Attention U-Net: Learning where to look for the pancreas [EB/OL]. [2021-02-20]. *https://arxiv.org/abs/*1804.03999.

[2] Chowdhury D, Mandal S, Das D, *et al*. An adaptive technique for computer vision based vehicles license plate detection system [C]// *International Conference on Opto-Electronics and Applied Optics*. Piscataway: IEEE, 2019: 1-6.

[3] Azad R, Davami F, Azad B. A novel and robust method for automatic license plate recognition system based on pattern recognition [J]. *Advances in Computer Science*: *An International Journal*, 2013, **2**(3): 64-70.

[4] Deb K, Jo K H. HSI color based vehicle license plate detection [C]// *International Conference on Control*, *Automation and Systems*. Piscataway: IEEE, 2008: 687-691.

[5] Chang S L, Chen L S, Chung Y C, *et al*. Automatic license plate recognition [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2004, **5**(1): 42-53.

[6] Chen R, Luo Y J. An improved license plate location method based on edge detection [J]. *Physics Procedia*, 2012, **24**(part B): 1350-1356.

[7] Zheng D, Zhao Y, Wang J. An efficient method of license plate location [J]. *Pattern Recognition Letters*, 2005, **26**(15): 2431-2438.

[8] Lin K, Tang H, Huang T S. Robust license plate detection using image saliency [C]// *IEEE International Conference on Image Processing*. Piscataway: IEEE, 2010: 3945-3948.

[9] Wang C, Yin F, Liu C. Scene text detection with novel superpixel based character candidate extraction [C]// *International Conference on Document Analysis and Recognition*. Piscataway: IEEE, 2017: 929-934.

[10] Hao W L, Tay Y H. Detection of license plate characters in natural scene with MSER and SIFT unigram classifier [C]// *IEEE Conference on Sustainable Utilization and Development in Engineering and Technology*. Piscataway: IEEE, 2010: 95-98.

[11] Llorca D F, Salinas C, Jimenez M, *et al*. Two-camera based accurate vehicle speed measurement using average speed at a fixed point [C]// *International Conference on Intelligent Transportation Systems*. Piscataway: IEEE, 2016: 2533-2538.

[12] Deb K, Chae H U, Jo K H. Vehicle license plate detection method based on sliding concentric windows and histogram [J]. *Journal of Computers*, 2009, **4**(8): 771-777.

[13] Zhang H F, Jia W J, He X, *et al*. Learning-based license plate detection using global and local features [C]// *International Conference on Pattern Recognition*. Piscataway: IEEE, 2006: 1102-1105.

[14] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection [C]// *International Conference on Image Processing- Rochester*. Piscataway: IEEE, 2002: 900-903.

[15] Cantarini G, Noceti N, Odone F. Boosting car plate recognition systems performances with agile re-training [C]// *International Conference on Image Processing*, *Applications and Systems*. Piscataway: IEEE, 2020: 102-107.

[16] Breiman L. Random forests [J]. *Machine Learning*, 2001,

45(1): 5-32.

[17] Miyata S, Oka K. Automated license plate detection using a support vector machine [C]// *International Conference on Control, Automation, Robotics and Vision*. Piscataway: IEEE, 2016: 1-5.

[18] Ho W T, Hao W L, Yong H T. Two stage license plate detection using gentle adaboost and SIFT-SVM [C]// *Asian Conference on Intelligent Information and Database Systems*. Piscataway: IEEE, 2009: 109-114.

[19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [EB/OL]. [2021-01-18]. *https://arxiv.org/abs/*1505.04597.

[20] Paszke A, Chaurasia A, Kim S, *et al*. ENet: A deep neural network architecture for real-time semantic segmentation [EB/OL]. [2021-02-07]. *http://www.arXivpreprintarXiv*: 1606.02147.

[21] Xiang H, Zhao Y, Yuan Y L, *et al*. Lightweight fully convolutional network for license plate detection [J]. *Optik-International Journal for Light and Electron Optics*, 2018, **178**: 1185-1194.

[22] Wang Q, Gao J Y, Yuan Y. Embedding structured contour and location prior in siamesed fully convolutional networks for road detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**(1): 219-224.

[23] Wang Q, Gao J Y, Yuan Y. A joint convolutional neural networks and context transfer for street scenes labeling [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**(5): 1457-1470.

[24] Xie L L, Ahmad T, Jin L W, *et al*. A new CNN-based method for multi-directional car license plate detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**(2): 507-517.

[25] Li H, Shen C H. Reading car license plates using deep convolutional neural networks and LSTMs [EB/OL]. [2021-01-18]. *https://arxiv.org/pdf/*1601.05610*.pdf*.

[26] Hsu G S, Chen J C, Chung Y Z. Application-oriented license plate recognition [J]. *IEEE Transactions on Vehicular Technology*, 2013, **62**(2): 552-561.

[27] Selmi Z, Halima B H, Alimi A M. Deep learning system for automatic license plate detection and recognition [C]*// International Conference on Document Analysis and Recognition*. Piscataway: IEEE, 2017: 1132-1138.

□