Deep Learning-Based License Plate Recognition in Complex Environments

Zechao Dai

Computer Science and technology major, Guangzhou University, Guangzhou, China
zechao@uok.edu.gr

Abstract. The traditional method of license plate recognition has good recognition for objects in a relatively fixed position and objects with good lighting conditions, while it exhibits low robustness when dealing with varying light conditions and recognizing moving vehicles. In order to be able to accurately identify vehicles at high speeds and in complex environments, this review starts from the perspective of historical development. The deep learning algorithms represented by CNN, YOLO and Transformers and their combined application in the domain of license plate recognition are studied. By combining CNN with YOLO algorithms, the system achieves a 99.37% detection accuracy and a 98.43% overall recognition rate. The YOLO algorithm itself has an average accuracy of 98.56% for license plate detection, and Transformers can achieve an accuracy of 99% in SSIM under transformerRain100L.In conclusion, for CNN, it is recommended to develop towards optimizing their own parameters and combining them with other algorithms. High-accuracy YOLO and Transformer models can be integrated with other techniques to further enhance LPR performance.

Keywords: Deep Learning, CNN, Transformers, YOLO.

1 Introduction

In contemporary society, with the constant advancement and evolution of computer technology, computers have gradually entered people's daily lives, modern transportation has also experienced rapid development, the combination of computer technology and transportation has played a great role in urban transportation in real life. For example, traffic management, parking lot fees, urban monitoring, vehicle tracking, license plate recognition, and computer technology have been widely applied in these fields. Traditional license plate recognition relies on clear images, and strongly relies on manual feature extraction and rule matching. Traditional methods lack robustness in practical applications, there are many complex weather environments such as rainfall, snowfall, haze, and other severe weather conditions that seriously affect the accuracy of computer object recognition. Under this influence, errors and inaccuracies in license plate recognition will inevitably occur.

With the swift progression of deep learning technology in recent years, novel models and algorithms like convolutional neural networks (CNN) and recurrent neural

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_31

networks (RNN) have emerged. Object detection algorithm YOLO, R-CNN have entered public field of vision. They have exhibited outstanding performance in the areas of image processing and pattern recognition, making significant progress in LPR technology in demanding conditions and greatly improving the accuracy and real-time performance of lincese plate recognition systems (LPRS).

Despite the swift advancement of deep learning technology in the area of LPR, it continues to encounter numerous challenges in real-world applications. How to ensure the stability and accuracy of LPR in a complex environment, How to further optimize deep learning models for improved detection efficiency remains a crucial research topic. Deep learning-based LPR in complex environments offers both theoretical and practical value. The following is a brief description of LPR technology and its history.

The process of LPR technology has three steps.Firstly, after obtaining the vehicle image, the license plate position of the vehicle is located. After successful positioning, character segmentation of the license plate position image is performed. After segmentation is completed, character recognition is carried out. The final step involves character stitching and restoration according to the image sequence to reconstruct the complete plate number.

2 Deep learning based license plate recognition in complex environments

2.1 Traditional Method Period: Traditional methods mainly represented by KNN algorithm

The accuracy of Automatic LPR (ALPR) is readily influenced by a variety of factors.such as contrast, lighting, and moving images. KNN, or K-nearest neighbor is a supervised learning algorithm capable of handling both classification and regression tasks. And K-nearest neighbor is regarded as one of the most straightforward machine learning algorithms. The classification rules are shown in the Fig. 1.

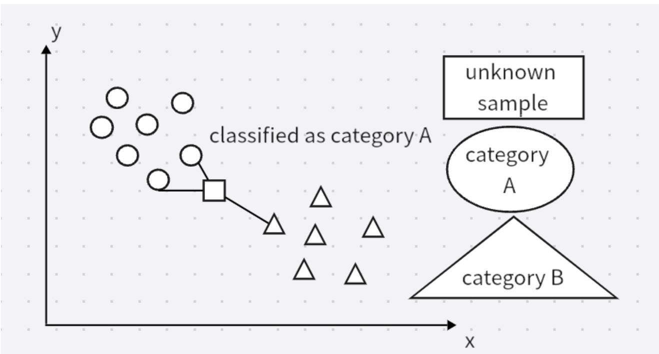

Fig. 1. KNN classifies images (Photo/Picture credit: Original)

As shown in Fig. 1 the choice of K value is 2, deviding the data into two categories: class A and class B. According to the distance formula (Euclidean distance, Manhattan distance (for continuous data), and Hamming distance (for categorizing data), the unknown sample is classified into the closer class A rather than class B.

Tu Jing et al. reported utilization of the K-nearest neighbor (KNN) classification algorithm in a car park LPRS that its accuracy is mainly affected by the K value. A K value that is too small may decrease classification accuracy, whereas a K value that is too large may amplify noise. Tu Jing et al. proposed applying the KNN (K-nearest neighbor) classification algorithm to character recognition. By extracting feature vectors from vectorized characters, the dimensionality and complexity of recognition features were reduced, making it easy to implement in code while ensuring retrieval efficiency and accuracy. In the experimental results, The training set achieved a recognition accuracy of 97.80%, while the test set had a recognition accuracy of 95.50%., the LPR rate was 92.80%, and the single sample recognition time was 0.36 seconds. Applying this method to the development of actual parking lot management systems, the experimental findings reveal that the KNN algorithm demonstrates superior reliability and accuracy in the processes of license plate localization and subsequent recognition, compared to SIFT and ANN neural network algorithms, enhancing the stability and practical applicability of LPR systems [1].

2.2 Convolutional neural network (CNN) period: Neural network method mainly represented by CNN algorithm

In addition to KNN, another widely adopted method is the CNN, which is structured as follows.CNN are mainly composed of five parts:

2.2.1 Input layer.

Receives the original image data and performs pre-processing operations, such as resizing the image, normalization processing and possible cropping or filling, to ensure the consistency of the input data and provide a suitable input format for subsequent processing of the convolutional layer.

2.2.2 Convolutional layer.

Receives the original image data and performs pre-processing operations, such as image resizing, normalization processing and possible cropping or filling,to guarantee the uniformity of the input data and furnish an appropriate input format for the subsequent processing by convolutional layers.

2.2.3 Activation function.

After the convolutional layer, nonlinear activation functions are usually introduced to increase the network's nonlinear expression ability. These activation functions can zero negative values in the feature map (ReLU) or perform other forms of nonlinear transformations, thereby enhancing the network's feature learning ability.

2.2.4 Pooling layer.

The pooling layer performs downsampling on the feature map by applying a local operation to select either the maximum value within a specified region (known as max pooling) or the average value of the elements in that region (known as average pooling) in the feature map to reduce its size while preserving important feature information. The pooling layer helps reduce computational complexity, to mitigate overfitting and enhance the robustness of the model.

2.2.5 Fully connected layer.

Subsequent to the convolutional and pooling layers, the architecture usually incorporates one or more fully connected layers to process the extracted features further. The fully connected layer maps the feature vectors extracted from the feature map onto the output category space, performs linear transformation through weight matrix and bias term, and combines with activation function for nonlinear transformation to achieve classification or regression tasks. Convolution and pooling operations are fundamental components to CNN.

Convolution operation is the core of CNN, which extracts local features by convolution checking input images or feature maps. The convolution kernel slides over the input data, performs dot product operations on each position, and generates feature maps, which contain feature information at different scales and directions of the image, providing key features for subsequent classification or regression tasks. As shown in Fig. 2.

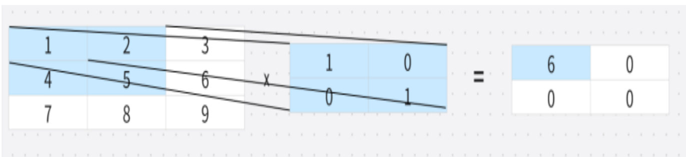

Fig. 2. Convolution operation specification (Photo/Picture credit: Original)

As shown in Fig. 2, the calculation method of convolution operation in the figure is 1x1+2x0+4x0+5x1=6, and then the left blue block moves to the right by one grid, that is, it contains 2, 3, 5 and 6. Continue convolution operation until all convolution operations on the left are completed.

Pooling operation is a downsampling technique used in CNN to decrease the dimensions and count of parameters in feature maps.This operation complements convolution by summarizing feature information effectively. It summarizes information by selecting local maxima (max pooling) or averages (average pooling) from feature maps, thereby preserving important features and reducing computational complexity, enhancing model robustness, and preventing overfitting. As shown in Fig. 3.

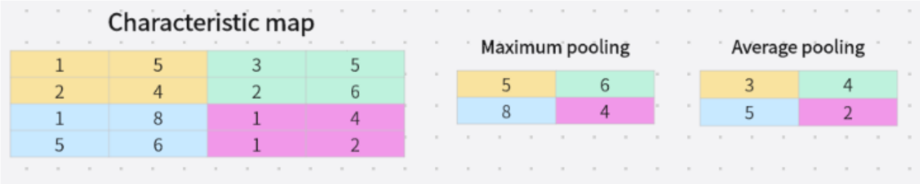

Fig. 3. Pooling operation description (Photo/Picture credit: Original)

As shown in Fig. 3,Maximum pooling is to select the largest value in the region and put it into a new region. As shown in Fig. 3, number 1, 2, 4 and 5, the maximum value 5 is the result of maximum pooling. The output of average pooling is the mean value of the input elements within the pooling window, 1, 2, 4, and 5 in the calculation area is 3.

Selmi, Z et al. reported a deep learning-based framework for detecting and recognizing license plates automatically. Before applying the first CNN model to classify boards/non boards, many preprocessing steps should be performed. Subsequently, they performed preprocessing steps to segment the license plate (LP), and ultimately utilized a second CNN model with 37 classes to identify all uppercase letters (A-Z) and digits (0-9) on it. The proposed system's performance was evaluated on two datasets comprising images captured under diverse conditions, such as images with poor quality, distorted perspective, captured during bright daylight, and at night, and challenging environments. A significant portion of the results demonstrated the accuracy of the proposed system. Experimental results for LP detection in recall rates and F-score accuracy performed on the Caltech dataset gave high rates: 93.8% and 91.3%, respectively. For character recognition, they tested their system with two datasets, The experimental results obtained from the Caltech dataset yielded an accuracy of 94.8%,but in the three subsets of the AOLP dataset, they achieved 96.2% accuracy (AC), 95.4% localization success (LE), and 95.1% recognition precision (RP). It is proved that CNN has great improvement compared with KNN [2].

2.3 End-to-end period: end-to-end method mainly represented by YOLO algorithm

YOLO is a real-time object detection framework that converts object detection into a regression problem, the input image is processed by a single CNN to directly output bounding boxes and category probabilities. YOLO has fast speed and low background false detection rate, however its positioning accuracy and recall rate still require enhancement, particularly in complex environments. It is widely applied in fields like autonomous driving. Video surveillance, and industrial quality inspection, and has continuously developed improved versions such as YOLOv2, YOLOv3, and YOLOv4 to enhance detection accuracy and speed.

Tourani, A, et al. Observed that license plates are differ in terms of resolutions, layouts, the scarcity of numbers/characters, a variety of background colors, and varying font sizes. In this context, the system leverages a distinctively ne-tuned YOLOv3 platform for each of the above stages and automatically extracts Persian characters from the input image in two steps. During the training and testing phase, images of various vehicles under a variety of challenging and urgent conditions have been gathered. From the actual system installed as a monitoring application. The experimental results demonstrate the end-to-end accuracy of 5719 images is 95.05%. The test data includes color and grayscale images, which contain vehicles captured at varying distances, angles, brightness levels, and resolutions. In addition, for actual data, the system can perform LPD and CR tasks within an average of 119.73 milliseconds, which demonstrates the real-time performance and applicability of the system. The system is fully automated and does not require pre-processing, calibration, or program configuration [3].

Liu Lingyuan, et al. developed an efficient LPRS tailored for complex scenarios by integrating yolov5 deep learning model and Tesseract-OCR library, this study developed an efficient LPR system suitable for complex scenes. The system is divided into two parts: license plate detection and character recognition, which greatly improves the recognition performance and the robustness of the system under harsh conditions. The experimental findings reveal that the mean accuracy of license plate detection is reaches 98.56% and the average accuracy of character recognition reaches 96.56%, which demonstrates the efficacy and outstanding performance of the proposed approach. Developed an efficient system for localizing and recognizing license plates, built upon YOLOv5 and Tesseract OCR. Through comprehensive testing of 5000 license plate images from different scenarios, it has been proven that the system has high accuracy and robustness in standard, tilted, and severe weather conditions. Despite the challenges in low light and long-distance conditions, the system's high accuracy in other scenarios demonstrates its great potential in practical applications [4]. It can be observed that the YOLO algorithm has only a slight improvement in accuracy compared to CNN, but in research, automation has been achieved. Here's a look at the study of CNN combined with the YOLO algorithm [3].

Izidio, D. M, et al.,propose a system design method that designs and utilizes convolutional neural networks (CNNs) to recognize Brazilian license plates, specifically tailored for embedded systems. The system effectively detects license plates within the captured images,using the Tiny YOLOv3 architecture and accurately recognizes the characters on them, and a second convolutional network trained on the composite image and fine-tuned using the real license plate image .The proposed architecture has exhibited robustness against variations in angle, lighting, and noise, with only one forward pass required per network , thus allowing for faster processing compared to other deep learning methods. This paper method was verified using real license plate images, achieving a detection rate of 99.37% and an overall recognition rate of 98.43% under different environmental conditions. The average time for processing 1024 × 768 images utilizing a single license plate image in the Raspberry Pi3 is 2.70 seconds. In order to enhance the accuracy of recognition, a series of CNN models were tested, rather than a single CNN model, the average processing time per image was 4.88 seconds, and a boost in recognition rate up to 99.53% as a result. Finally, they discussed the impact of using a set of CNNs while taking into account the balance between accuracy and performance when designing an embedded system for LPR. Overall, the system exhibits robust changes in angle and lighting, and has achieved state-of-the-art results. The proposed method is segmentation free, this minimizes errors stemming from inadequate character segmentation and necessitates an assessment of each license plate. Additional feedback is anticipated to enhance the overall accuracy of the system [5].

Wu Yuanyuan et al. introduced an end-to-end LPR algorithm. It is capable of detecting and recognizing license plates with a high degree of accuracy in complex environments addressing issues of slow speed and low accuracy in traditional models. The YOLOv7 target detection model is employed as the detector to locate the LP, and the CNN algorithm is used to recognize the license plate in the complex environment with strong light, weak light or uneven illumination, which improves the accuracy of LPR. CBAM channel attention mechanism is added in the output process of the feature layer of YOLOv7 network layer, which improves the model's capability for feature extraction. The improved YOLOv7 algorithm is used to detect license plates in complex environments. The detected license plate area is preprocessed, and the processed license plates are input into the improved CNN recognition model for character recognition. The experimental results show that the average detection the precision of the YOLOv7 detection model after adding the attention mechanism reaches 87.5%, and the recognition accuracy of the enhanced model attains 97.16%, marking a significant improvement over traditional LPR technology, and the recognition performance is good in complex environments, which has practical application value [6].

Two studies have demonstrated the effectiveness of combining CNN and YOLO in enhancing accuracy. Research further explored parameter adjustments, contributing to improved recognition precision and highlighting a promising future direction [5].

This is followed by other end-to-end methods such as Faster Region-based Convolutional Neural Network algorithm (Faster R-CNN). The key modification in Faster R-CNN compared to traditional CNN is the integration of the Region Proposal Network (RPN).The network generates potential object regions and refines each one through Region of Interest (RoI) pooling and subsequent classification. In addition, Faster R-CNN adopts an anchor mechanism, which improves the flexibility and accuracy of detection, achieves end-to-end object detection, and significantly improves detection speed and accuracy.

A study by Khan, K et al., proposes that a two-step approach for plate positioning under challenging conditions. In the first step, the Faster R-CNN is designed to detect all types of vehicles in the image, thereby obtaining zooming information to identify the position of the license plate. In the second step, morphology is used to reduce the non-plate area. At the same time, geometric properties are utilized to pinpoint the location of plates within the HSI color space. This method enhances accuracy while decreasing processing time.For character recognition, use an adaptive boosting lookup table (LUT) classifier and a modified Census Cross (MCT) as the feature extractor. The proposed system for license plate detection and character recognition methods are significantly outperforms traditional methods in terms of accuracy and recall rate in multi LPR. The proposed method exhibits a significantly higher detection rate compared to existing methods, achieving a total detection rate of 96.72% and a recognition rate of 98.02% across multiple license plates and varying lighting conditions. The proposed algorithm is potentially suitable for real-time ITS applications. Furthermore, future research efforts could concentrate on creating parallel implementations of existing algorithms. The authors believe that this will further reduce the execution time for recognising licence plates [7].

The accuracy of the experimental results reached 96.72%, which was improved compared with CNN algorithm, indicating that CNN also has a huge space for development in algorithm improvement. Research on the combination of YOLO and Mask R-CNN. Mask R-CNN is an instance segmentation algorithm builds upon Faster R-CNN. It adds segmentation tasks and outputs classification, regression, and segmentation results.

Zhang Xiaohui et al. proposed an algorithm that integrates multiple approaches through fusion of improved Yolov3 vehicle recognition and Mask R-CNN license plate detection, the approach detected vehicles and license plates respectively in two steps, and finally built YOL3-RCNN multi-LPR model. The model firstly integrates large, medium and small image features based on upsampling algorithm [8].

Then, the morphological clustering algorithm is used to adaptively calculate the IOU threshold, improving the vehicle detection rate and accuracy; Next, the detected vehicles are cropped, SOFTNMS optimized, and affine transformed to solve the problems of license plate occlusion, tilt, and complex background; Finally, the ResNet50 network is utilized for recognize license plate characters and numbers, and RPN is used for regression verification. The simulation results of multiple sets of baseline algorithms show that on UA-DETRAC vehicle detection data set, compared with the traditional LPR algorithm, the accuracy rate and recall rate of YOL3-RCNN algorithm are increased by 7.7% and 9.6%, respectively, and have a higher license plate detection rate [8].

Table 1. Performance parameters of license plate detection model [8]

Detection model	Performance parameter Training set/Test set P (%)	Performance parameter Training set/Test set R (%)	Performance parameter Training set/Test set F (%)	Modeling time T(%)	Detection speed V(%)
Fast R-CNN	72.06/71.43	73.47/72.52	72.76/71.97	76.8	114.7
R-FCN	78.41/75.68	79.68/77.89	79.04/76.77	85.2	89.0
Yolov3	88.58/88.86	86.25/87.23	87.40/88.04	82.7	105.6
SSD	75.33/76.73	72.73/74.15	74.01/75.42	91.4	93.2
YOLO-RCNN	96.25/95.52	95.84/95.09	96.04/95.30	100.0	100.0

As shown in Table 1, With YOL3-RCNN as 100% modeling time and detection speed, YOL3-RCNN has a modeling time standardized to 100% and a median detection speed, but its overall performance surpasses other models significantly. In summary, the YOL3-RCNN multi-LPR model constructed has high detection accuracy, timeliness, and adaptability to various environments. The combination of YOLO and Mask R-CNN algorithm can improve the performance significantly at the same detection rate [8].

2.4 Transformers period: New method mainly represented by transforemers algorithm

Transformers is a neural network architecture that relies on the self-attention mechanism, used for processing sequential data. It achieves global modeling and efficient processing of sequence data through techniques such as encoder decoder architecture, multi head attention mechanism, and positional encoding. It has wide applications in natural language processing, speech recognition and other fields, with advantages such as good parallel performance and short training time.

Wu Sinan proposed a network for single image snow removal, which is based on a multi-scale generative adversarial network. Which incorporates several innovative features. In this method, the residual module is used as the backbone of the generator, and the context-interactive Transformer module is introduced to carry out the appearance prior reconstruction of the generation of snowless images, focusing on the features of snow lines, and using the context information of the snow line neighborhood to recover the background information covered by snow marks. Simultaneously adopting learnable RN-L, automatically detecting potential damaged and undamaged regions for separate normalization, and performing global affine transformation to enhance their fusion. This method uses a multi-scale discriminator on the discriminator to make the discrimination more comprehensive and preserve more details. This article proves the effectiveness of the algorithm in snow removal through comparative experiments, and the ablation experiment proves that each module has the effect of improving snow removal. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), PSNR is a commonly used image quality evaluation index in the field of image rain removal, while SSIM is more in line with the evaluation standards of the human eye. Under Rain100L, the accuracy of SSIM can reach 0.99, and the accuracy of SSIM under Rain100H is far superior to other methods [9].

Table 2. Quantitative experimental evaluation of three synthetic data sets

Methods	Parameters	Time 512x512	Rain100H		Rain100L		Rain12	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
VRGNet	0.242 M	0.079 s	25.22	0.79	33.30	0.95	29.58	0.92

DiG-CM	4.103 M	0.607 s	19.44	0.68	30.98	0.94	32.43	0.94
PReNet	0.169 M	0.080 s	27.81	0.89	36.26	0.98	34.80	0.96
JORDER	4.169 M	0.222 s	25.69	0.83	35.71	0.97	32.80	0.95
DDN	0.064 M	0.186 s	17.26	0.55	29.48	0.92	33.43	0.94
Ours	0.176 M	0.076 s	29.02	0.90	39.36	0.99	35.50	0.97

As shown in Table 2, PSNR is a commonly used index in image rain removal, whereas SSIM aligns better with human visual evaluation. Under Rain100L, the accuracy of SSIM can reach 0.99, and the accuracy of SSIM under Rain100H is far superior to other methods [9].

Finally, the study of the combination of transformers and CNN algorithm. Jun, K developed a model that uses Vision Transformer (ViT) as the backbone to detect the corner position of license plate correction, a non-convolutional model that has received a lot of attention recently, and compared the performance of the proposed method with that of existing CNN-based methods ViT was initially applied in the field of visual classification, and its application has been expanded to encompass the fields of detection and segmentation. ViT exhibits versatility in its composition and can be utilized in a hybrid form when combined with other established models. They crafted a corner prediction model, utilizing ViT as its foundational backbone and compared the performance of their model with that of existing models that utilize ResNet and MobileNet. Therefore, considering its size,ViT exhibits notable performance, yet ResNet holds a superior position in terms of absolute performance . These results were derived from a performance analysis of 600 ViT backbone models, which were generated through a combination of four key structural determinants of ViT. The results reveal that simply replacing CNNtoViT's backbone will not achieve performance improvements. In short, ViT is a more challenging model to work with compared to CNN; In this case,it requires more careful handling when importing images, and the training process is more intricate [10].

Although transformers have strong performance, they may face issues such as unclear integration with CNN and the need for parameter adjustments. These challenges provide valuable references for future research directions

3 Conclusion

From the perspective of historical development, this review starts from the period of traditional methods, to the period of CNN, to the end-to-end period, and finally to the transformers period. In the traditional methods era, the KNN algorithm achieved a recognition rate of 92.80%. Subsequently, the combination of CNN, YOLO, and transformer algorithms demonstrated superior accuracy. Achieving the detection rate reached 99.37%, with an overall recognition rate of 98.43% under different environmental conditions. Moreover, The recognition system's performance can be further improved through parameter optimization, and it can be studied in the direction of optimization of the algorithm itself and combination with other algorithms to improve accuracy. As the main algorithm in the end-to-end period, the YOLO algorithm achieves an average accuracy rate of 98.56% for LPR.It can be developed in the direction of algorithm optimization. Transformers, as the cutting-edge technology, can achieve 99% accuracy in SSIM under Rain100L and move forward to algorithm optimization. Finally, for CNN, it is recommended to develop towards optimizing their own parameters and combining them with other algorithms. High-accuracy YOLO and Transformer models can be integrated with other techniques to further enhance LPR performance.

References

1. Tu, J., Liu, D., Zhong, X.: 'Application of KNN classification algorithm in parking lot license plate recognition system', Journal of Huainan Normal University, 2021, 2, 143-148
2. Selmi, Z., Halima, M.B., Alimi, A.M.: 'Deep learning system for automatic license plate detection and recognition', Proc. 14th IAPR Int. Conf. Document Analysis and Recognition (ICDAR), Kyoto, Japan, November 2017, 1132-1138
3. Tourani, A., Shahbahrami, A., Soroori, S., Khazaee, S., Suen, C.Y.: 'A robust deep learning approach for automatic iranian vehicle license plate detection and recognition for surveillance systems', IEEE Access, 2020, 8, 201317-201330
4. Liu, L.: 'License plate detection and recognition based on deep learning', Modern Information Technology, 2024, 23, 53-57
5. Izidio, D.M., Ferreira, A.P., Medeiros, H.R., Barros, E.N.D.S.: 'An embedded automatic license plate recognition system using deep learning', Design Automation for Embedded Systems, 2020, 24, (1), 23-43
6. Wu, Y., Shi, Q.: 'License plate recognition algorithm based on deep learning', Modern Computer, 2024, 17, 7-12
7. Khan, K., Imran, A., Rehman, H.Z.U., Fazil, A., Zakwan, M., Mahmood, Z.: 'Performance enhancement method for multiple license plate recognition in challenging environments', EURASIP Journal on Image and Video Processing, 2021, 2021, 1-23
8. Zhang, X., Song, O., Zheng, W., Yu, X.: 'Automatic recognition of multiple license plates at high traffic intersections in complex road environments', Computer Simulation, 2024, 8, 150-154
9. Wu, S.: 'Research on Image Rain and Snow Removal Algorithm and Application'. Master's thesis, Shenyang University of Technology, 2023
10. Jun, K.: 'Is vehicle plate corner prediction by vision transformer better than CNNs?', Scientific Programming, 2023, 1, 4301632