A License Plate Detection and Recognition Model in Complex Scenarios

Haiying Qi, Quanyan Gao, Shulong Wang

***Abstract*—The precise recognition and interpretation of vehicle LPs in variable and uncontrollable environmental conditions remain significant and enduring challenges for visual surveillance systems. While current approaches achieve acceptable performance in controlled conditions, conventional methods exhibit considerable performance degradation when applied to complex real-world situations. To address these limitations, we propose a novel computational architecture ICS-Net, to mitigate these issues. In the detection phase, the Model employs an improved IC-YOLOv5 algorithm based on YOLOv5 and introduces Channel Feature Fusion and Pairing (CFFI) between the backbone and neck networks. This enhancement strengthens the flow of channel information, leading to improved overall performance. Additionally, we apply the Channel and CBAM at the end of the backbone network. This effectively extracts key channel and spatial information while suppressing redundant features, allowing for more accurate LP detection and localisation. In the execution of character recognition for vehicle registration plates, ICS-Net employs an enhanced SE-LPRNet methodology rooted in neural network architecture. By analyzing the feature location of the SE module in the network, we use a standard strategy to insert the SE module, further optimizing network performance. Moreover, the input configuration of the network is restructured to enhance the initial feature set, which not only speeds up training convergence but also boosts training effectiveness. Extensive experiments show that the ICS-Net Model performs exceptionally well on the CCPD, PKUData, and CLPD datasets. Specifically, the average detection accuracy for LPs is 95%, and the average recognition accuracy is 99%. Compared to baseline methods, ICS-Net outperforms others in multiple complex scenarios, demonstrating its ability to effectively handle LP detection and recognition tasks in demanding and complex environments and meet practical application requirements. Compared to baseline methods, ICS-Net outperforms others in intricate environments, demonstrating its ability to effectively handle LP detection and recognition tasks in demanding and complex environments and meet practical application requirements.**

***Index Terms*—LP recognition, Convolutional network, attention mechanism, LPRNet, YOLOv5**

I. Introduction

License Plate Recognition technology serves as a critical component in Intelligent Transport Systems [1], providing significant value by automatically extracting vehicle information. This will improve traffic management efficiency, increase accuracy, and reduce human errors and operational costs. Over the past few years, a significant number of researchers worldwide have introduced diverse Models and approaches for LP recognition. The following section highlights the main research advances of these two types of methods.

Conventional LP recognition methods rely on image processing and machine learning techniques, typically organized into three fundamental stages: LP localisation, character segmentation and character recognition. Regarding number plate positioning, common methods are divided into colour-based [2] and edge features [3]. The colour method aids detection through rich visual information, but is susceptible to light changes and difficult to deal with number plates similar to the background colour; the edge method is efficient and fast but has low accuracy in fuzzy or obscured situations. To improve the positioning effect, in [4], the authors proposed the Edge-LPR system, but it may still miss the detection of LPs with blurred boundaries. For character segmentation, the projection method [5], the connected domain method [6] and the static boundary method [7]are commonly used to solve the character segmentation problem, but they are less effective when dealing with character sticking and complex backgrounds. In the character recognition stage, template matching [8], support vector machine [9] and neural network [10] are common methods, although, these systems are affected by background clutter and insufficient datasets, which compromises their recognition capabilities. Overall, traditional LP recognition methods perform more stably in ideal environments but have poor recognition accuracy and adaptability in practical applications with uneven lighting, blurred LPs or complex backgrounds. Given the growing demand for traffic management, the advancement of more resilient and efficient LP recognition technologies has emerged as a critical research focus. Deep learning-based number plate recognition methods have become the focus of research, the core of which automatically extracts high-level features of images through deep neural networks to achieve fast and accurate recognition in complex scenes. In contrast to conventional approaches, deep learning eliminates the need for character segmentation, circumvents segmentation-related errors, and autonomously learns the holistic features of LP images, thereby enhancing both recognition accuracy and robustness. In terms of LP localisatio [11] significantly reduces the reliance on manual features and improves the target detection performance; in [12], the research introduces a dual-stage algorithm that integrates YOLOv3 with an optimized LP recognition network, with experimental results demonstrating its exceptional performance in complex environments. To further improve the recognition effect in multi-lane and urban environments, in [13], the authors used YOLOv4 for integrated Model and LP recognition while in low-resolution and multi-vehicle environments; in [14],

Manuscript received September 13, 2024; revised April 10, 2025.

Teaching Reform Project: Liaoning Province Graduate Education Teaching Reform Research Funding Project (LNYJG2024138)

Haiying Qi is an associate professor at the School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian, China(email: qihaiying@126.com).

Quanyan Gao is a postgraduate student at School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian, China (corresponding author: tel: +086-15140396493; e-mail:15140396493@163.com).

Shulong Wang is a postgraduate student at School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian, China (email: 2224512824@qq.com).

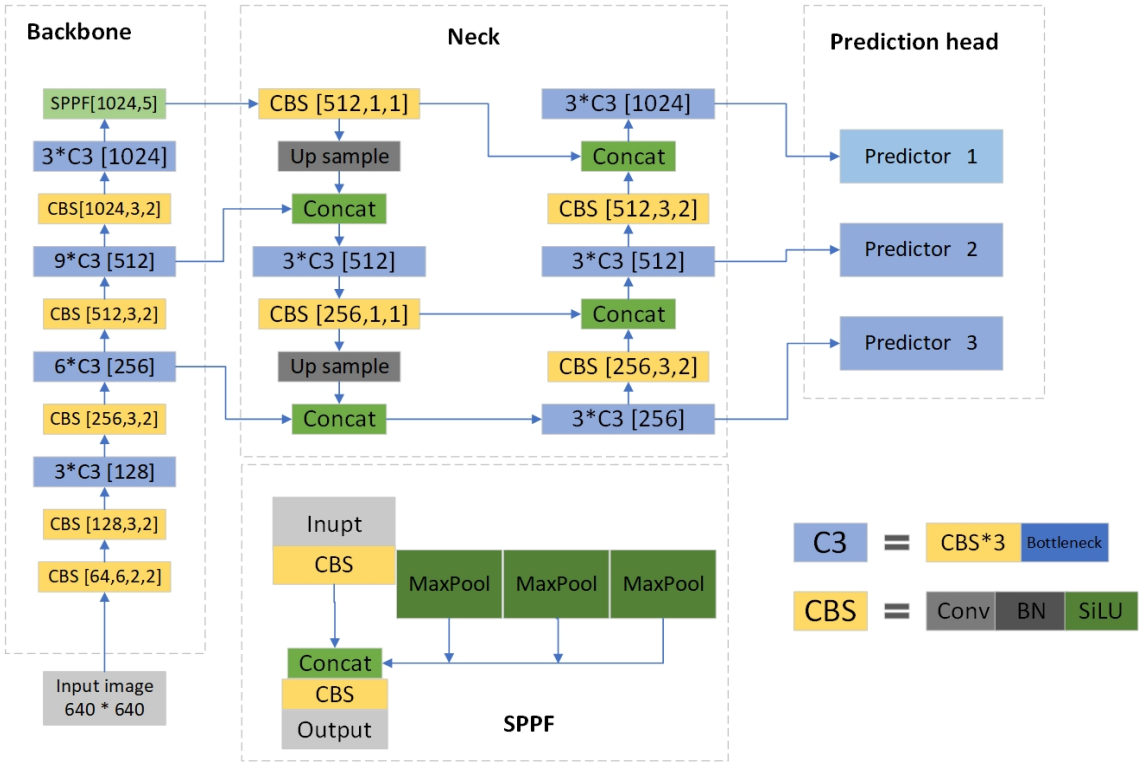


Fig. 1. YOLOv5 network structure

the authors proposed a method based on YOLOv5s, which improves the accuracy through data enhancement. In the domain of character recognition, addressing the segmentation challenge, the study referenced in [15] introduced an anchor-free approach by identifying boxes to isolate the target area attributes; concurrently, the work cited in [16] presented a streamlined CNN-driven ALPR framework, integrating image enhancement and morphological transformations to augment both image clarity and operational performance. In [17], the authors proposed a deep learning method combining Faster-RCNN and improved AlexNet, which significantly improved the recognition accuracy. In challenging environments, LP images frequently suffer from deformation, occlusion, and blurring. To address the problem of image tilt and deformation, in [18], the authors proposed a low-illumination LPR method and optimised the VGG architecture, which achieved better results than the traditional methods; in [19], the authors improved the perspective transformation accuracy by improving the CRNN Model and using an adaptive fusion feature segmentation network; and in [20], the authors proposed a de-fogging method based on MPGAN, combining the YOLOv3 and the LPRNet for LP localisation and recognition, achieving better results; in [21], the authors propose the YOLOv5-PDLPR Model, which further improves the recognition effect through a global information extractor. Despite the significant progress of deep learning-based number plate recognition methods, more resilient and efficient algorithms must be further developed to address the practical application demands posed by variable factors in complex environments.

In this paper, we propose a new Model, ICS-Net, which employs the improved YOLOv5 algorithm IC-YOLOv5 in the LP detection part. Channel information is enhanced by integrating the Channel Feature Fusion and Fitting in Pair (CFFI) module between the backbone and neck networks, thereby improving overall performance. Adding a CBAM module towards the conclusion of the backbone network's architecture can more efficiently extract key channels and spatial information, reduce redundant features, and improve the accuracy of localization.

During the LP character recognition stage, an enhanced version of the neural network algorithm SE-LPRNet is employed. By optimizing the insertion strategy of the SE module, the network's performance is further enhanced, and its ability to focus on key features is significantly improved. In addition, the improved input structure enriches the initial features, improves the training convergence speed, and enhances the Model's recognition ability in complex environments.

Overall, ICS-Net augments the dependability and efficacy of vehicle registration plate detection and interpretation by incorporating a range of cutting-edge advancements tailored to address the intricacies present in practical environments.

II. License Plate Detection

A. YOLOv5

In this study, YOLOv5 is selected as the experimental algorithm. The primary architecture of YOLOv5 is illustrated in Fig. 1.

B. Convolutional Attention Module (CBAM)

YOLOv5 backbone comprises the Conv module, CSP-Darknet53 (C3), and several SPPF modules. The Conv module applies 2D convolution, batch normalisation, and SiLU activation functions. The C3 module, based on CSPNet [22], is the core component for learning residual features. It consists of two branches: one with three Conv modules and

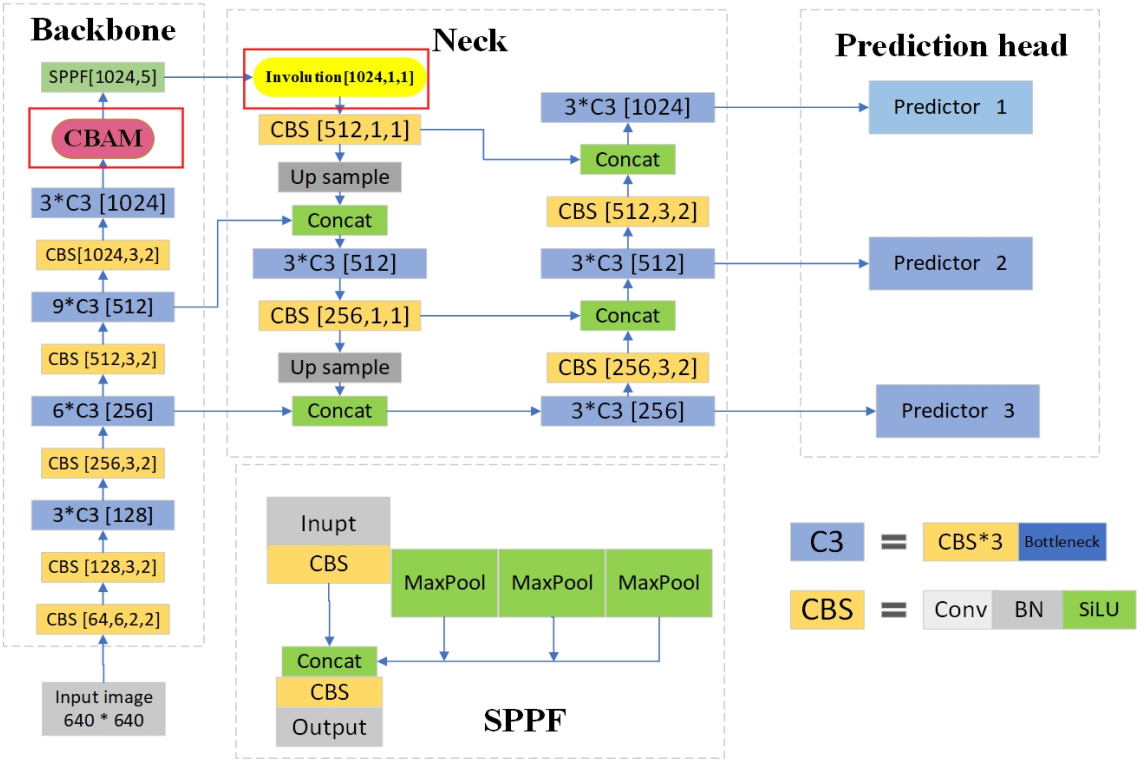


Fig. 2. IC-YOLOv5 network structure

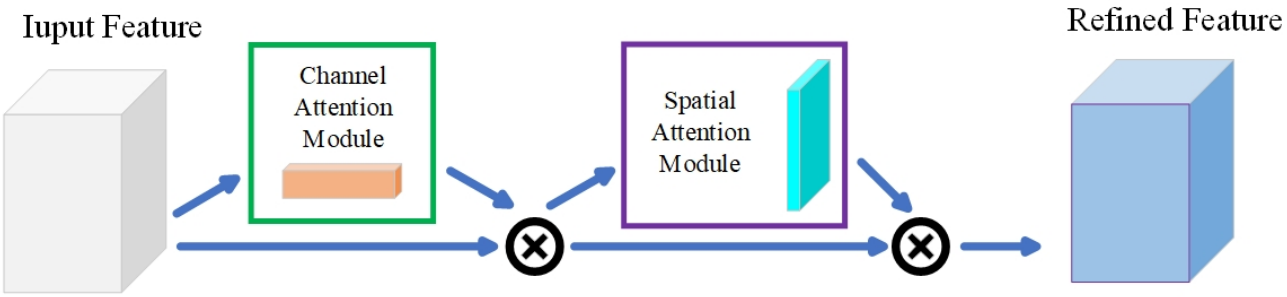


Fig. 3. CBAM network structure

several bottlenecks, and the other with a single Conv module. These branches are eventually merged and passed to the next layer.

The SPPF module, placed at the end of the backbone, is an enhanced version of the SPP module. Instead of using a large pooling kernel, SPPF uses multiple smaller cascading pooling kernels, which maintain the functionality of integrating feature mappings from diverse receptive fields, while simultaneously enhancing computational efficiency and enriching feature representation.

The structure of IC-YOLOv5 is depicted in Fig. 2. The conventional YOLOv5 framework comprises three main sections: the backbone dedicated to feature extraction, the neck focused on feature fusion and the trio of prediction heads. Based on the default Model, we propose two improvements: 1) the introduction of the pairwise fusion block (CFFI) at the beginning of the neck to improve the performance of the PANet, and 2) the inclusion of the CBAM in the backbone network. These improvements are designed to augment the feature extraction and fusion processes, thereby enhancing the Model's performance. With these changes, IC-YOLOv5 significantly enhances feature extraction and fusion capabilities while maintaining the lightweight nature of the original Model, thus improving target detection performance in diverse complex settings.

The CBAM employs a hybrid attention mechanism, consisting of two critical sub-modules, as depicted in Fig. 3. These modules process the input feature maps by integrating the resulting attention maps, thereby optimizing adaptive features. This mechanism enhances meaningful features across both channel and spatial dimensions while reducing redundant or irrelevant information. The channel attention module first performs global max pooling and global average pooling on each channel's feature map separately, then adds the results element-wise and applies the Sigmoid activation function to generate the channel attention vector$M_C(F)$ in Equ. (1).

$$M_C(F) = \delta(MLP(AvgPool(F)) + MLP(MaxPool(F))) = \delta\left(W_1\left(W_0(F^C_{avg})\right) + W_1(W_0(F^C_{max}))\right) \quad (1)$$

The spatial attention module conducts global max pooling and global average pooling operations on the pixel values at the same location of each feature map, subsequently concatenates these two pooled feature maps, and then performs a convolution operation followed by the Sigmoid activation function to generate the spatial attention vector, with the computational formula in Equ. (2):

$$M_S(\tilde{F'}) = \delta(f^{7\times7}([AvgPool(F'); MaxPool(F')]))$$

$$= \delta\left(f^{7\times7}([F'_{avg}; F'_{max}])\right) \quad (2)$$

By jointly applying the channel attention and spatial attention modules, CBAM can effectively enhance the CNN Model's focus on critical features and improve the precision of feature extraction. The integration of CBAM into IC-YOLOv5 significantly boosts the performance of LP detection and recognition.

C. Channel feature fusion method (CFFI)

The Neck component of YOLOv5 utilizes PANet as depicted in Fig.4. In the original YOLOv5, a 1×1 convolution is applied at the beginning of the Neck to reduce the number of channels, thereby enhancing computational efficiency. However, this method may lead to the loss of channel information, which can affect the performance of PANet. To tackle this problem, this study incorporates an Involution module between the backbone network and the neck network. This improvement mitigates information loss in the initial stage of FPN by enhancing and sharing channel information. The Involution block efficiently recovers and integrates feature information, thereby improving the performance of the FPN and, consequently, the overall effectiveness of the PANet.

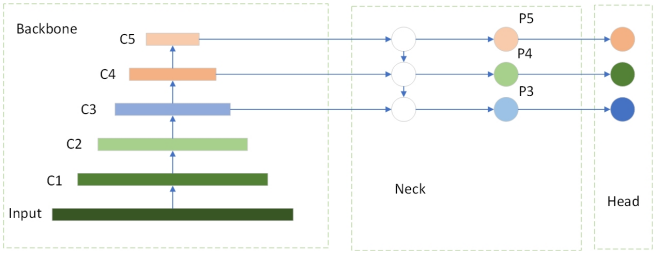


Fig. 4. FPN and PANet structures in YOLOv5

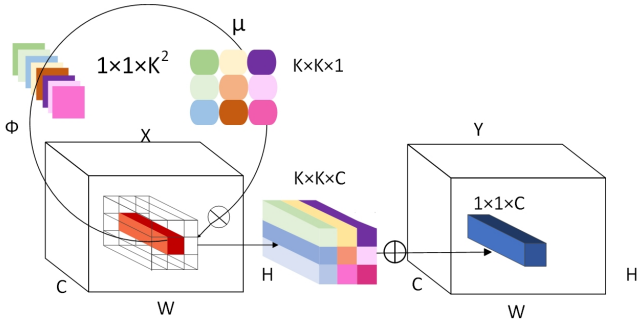


Fig. 5. Structure of Involution

D. SE_Block

Fig. 5 illustrates the structure of Involution. Its pairwise kernel is denoted as, where is the width of the feature map, K represents the kernel size, G indicates the number of groups, and each group shares the same pairwise kernel. Specifically, a particular pairwise kernel is designed for pixels shared between channels. The final output feature map of the pairwise combining operation can be expressed by the following Equation(3):

$$\mathbf{Y}_{i,j,k} = \sum_{(u,v)\in\Delta_K} \mathcal{H}_{i,j,u+\lfloor K/2\rfloor,v+\lfloor K/2\rfloor,\lceil kG/C\rceil}\mathbf{X}_{i+u,j+v,k} \quad (3)$$

Where $Y_(i,j,k)$ is at position (i,j) at which the channel on which the output is characterized. The pairwise kernel $H_(i,j,p,q,g)$provides transformations of inverse attributes in the spatial and channel domains, thus effectively integrating feature information from different regions. This design allows Involution to be more flexible in capturing and utilizing spatial and channel information in the feature map while maintaining computational efficiency, helping to improve Model performance and accuracy.

III. License Plate Recognition

A. LPRNet

LPRNet is a groundbreaking lightweight network designed specifically for LP number recognition, which does not rely on recurrent neural networks (RNNs) [23]. It supports recognizing variable-length character sequences and exhibits strong performance in challenging conditions such as low light, poor viewing angles, and bright glare.

The network structure of LPRNet comprises several key components: the Feature Extraction Network (Backbone), Feature Extractor, Sequence Modeling Module, and Connected Timing Classification (CTC) Loss. The Feature Extractor utilizes Convolutional Neural Networks to extract relevant features from the LP image. During the feature extraction process, LPRNet progressively refines and compresses image features through multiple convolutional and pooling layers, aiming to retain essential information while reducing computational costs.

Additionally, LPRNet incorporates a sequence Modelling module that employs convolutional layers and batch normalization to Model character relationships. This module generates probability predictions for each character position. For training, LPRNet uses a CTC loss function, which addresses the challenges of sequence length mismatches between the input and target sequences and accommodates duplicate and blank characters in the predictions.

B. The SE_Block

The SE_Block(Squeeze-and-Excitation Block structure is shown in Fig.6. In this configuration, the input is denoted as X, where $X \in R^{H'\times W'\times C'}$, and the output is U, with $U \in R^{H\times W\times C}$. Here, we treat F_{tr} as a basic convolution operation, which produces the output U. For ease of representation, the result of the convolution operation is expressed as a vector $V = [v_1, v_2, \ldots, v_n]$, where v_c represents the output of the convolution kernel. The corresponding output

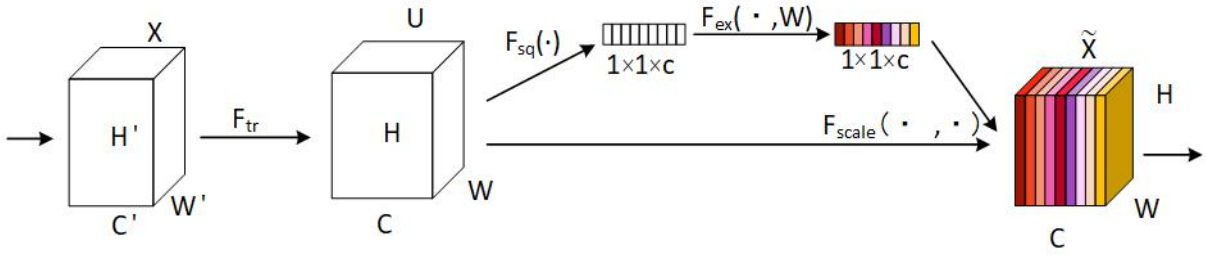


Fig. 6. Structure of SE_Block

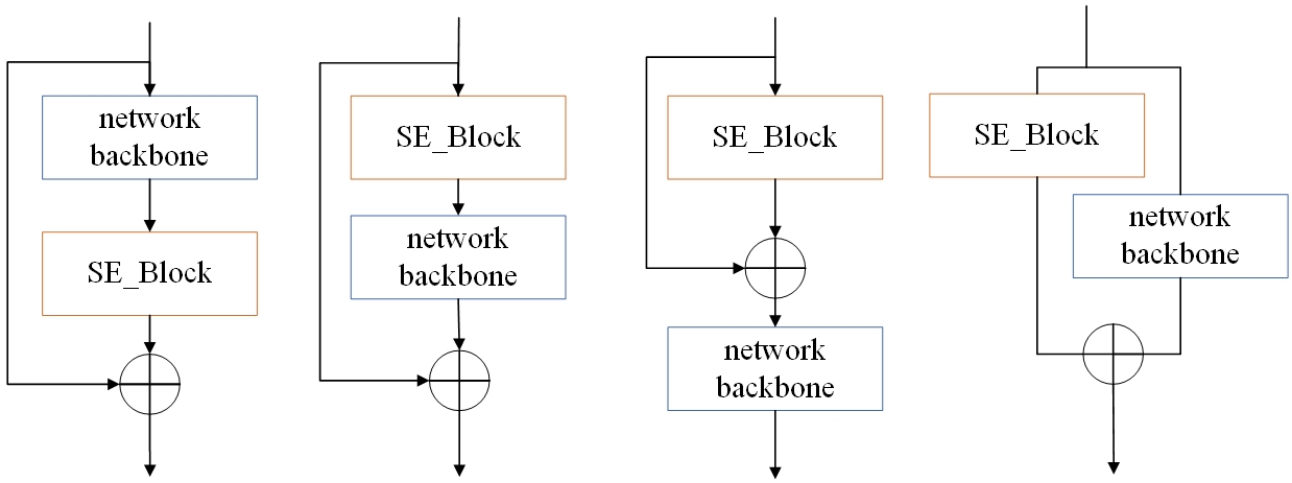


Fig. 7. Four integration designs

vector is $U = [u_1, u_2, \ldots, u_n]$,from which the following relationship is derived as in Equation(4):

$$u_c = v_c * X = \sum_{s=1}^{C'} v_c^8 * x^8 \tag{4}$$

This formulation provides a clearer understanding of the SE_Block's underlying working mechanism, thereby offering a theoretical foundation for subsequent analysis, optimization, and potential Model improvements.

To address inter-channel dependencies, we enhance the Model's context awareness by compressing global spatial information into each channel's descriptor. Traditional convolution operations typically rely only on local receptive fields, limiting the Model's ability to capture global context. To overcome this, we apply global average pooling to generate channel-wise statistical features, integrating global information. Specifically, the statistics are derived by downscaling the spatial dimensions of U, where each element in the statistic vector z is calculated by contracting the HXW dimensions, as shown in Equation(5). This approach improves feature representation and provides richer information for the subsequent Excitation operation, ultimately boosting the Model's expressiveness and performance.

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j). \tag{5}$$

To optimize the utilization of information obtained from the squeeze operation, an additional step is introduced to Model the interdependencies among channels. This function must fulfill two critical criteria: first, it should demonstrate flexibility and the capacity to learn nonlinear channel interactions; second, it should support non-mutually exclusive relationships, allowing the simultaneous emphasis of multiple channels. To achieve this, we use a simple sigmoid activation gating mechanism as in Equation(6) where σ refers to the ReLU function $\mathbf{W}_1 \in R^{\frac{C}{r} \times C}$ and $\mathbf{W}_2 \in R^{C \times \frac{C}{r}}$.

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(g(\mathbf{z}, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})), \tag{6}$$

C. SE-LPRNet

To strike an optimal balance between recognition precision and real-time processing efficiency, this study incorporates the SE (Squeeze-and-Excitation) module into the LPRNet neural architecture, leading to the creation of the SE-LPRNet Model. Drawing on prior research, the SE module is implemented using a standardized integration approach, termed the SE standard strategy. Additionally, the study investigates four alternative integration configurations, as depicted in Fig. 7. A comparative evaluation of these strategies demonstrates that the SE standard strategy substantially boosts the neural network's operational efficiency. As a result, this research adopts the SE standard strategy for integrating the SE module into LPRNet, achieving significant enhancements in overall performance.

This study introduces an improved methodology to tackle the problem of redundant convolution kernels and excessive pooling parameters in the LPRNet recognition framework. The proposed solution focuses on reconfiguring the convolution kernel structure within the neural network, eliminating

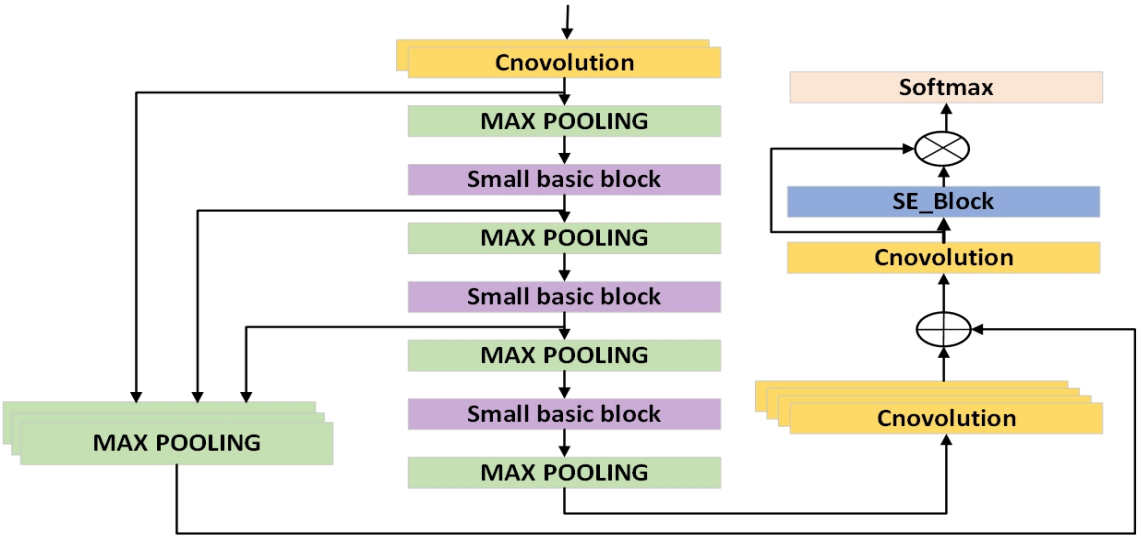


Fig. 8. SE-LPRNet structure

TABLE I
THE COMPREHENSIVE DETAILS OF THE LP DATASETS

Datasets Information	Year	Number of images	LP colors	Chinese province codes	Image size	Sequence length
CCPD	2019	10.4GB	blue	29	720×1160	7
PKUData	2016	395MB	blue+ yellow	27	1082 × 727	7
CLPDs	2019	84.1MB	blue + yellow + green	31	220 × 165/4596 × 2388	7/8

TABLE II
A SUMMARY OF THE INDIVIDUAL SUB-DATASETS WITHIN THE CCPD DATASET

Sub-Dataset	Description of LP styles
CCPD-Base	An ordinary LP image.
CCPD-FN	The position of the LP may vary, being either near or distant from the camera's capture area.
CCPD-Blur	LP images that are blurred as a result of camera lens shake.
CCPD-DB	LP areas may appear brighter, darker, or unevenly illuminated.
CCPD-Tilt	LP with a larger slant angle.
CCPD-Rotate	The LP at a smaller angle.
CCPD-Challenge	The LP is certain images present greater difficulties and challenges for accurate identification.
CCPD-Weather	Images of LPs captured during harsh weather conditions may be significantly affected.

the deactivation layer, and refining both the dataset and learning strategy. By substituting the deactivation layer's role with alternative techniques, the method effectively mitigates issues related to network parallelism and enhances computational efficiency. Furthermore, the SE (Squeeze-and-Excitation) module is incorporated into the neural architecture using a standardized SE strategy, substantially boosting the network's performance, especially in demanding scenarios. As a result, the research presents an advanced SE-LPRNet Model, which exhibits enhanced architectural and functional capabilities, leading to superior accuracy and robustness in LP recognition tasks.

IV. EXPERIMENTS

The experimental configuration is as follows: this experimental platform is based on the Ubuntu 18.04 operating system with Python 3.8 as the main programming language. The CPU used is Intel(R)Xeon(R) Platinum um8260CPU@2.30GHz with 12 available CPU cores and 60GB of available memory. The GPU is NVIDIA Tesla P100-16 GB, with 16GB of video memory and 732.16GB/s of video memory bandwidth, supporting 5.18TFLOPS of floating point operations per second. The algorithm is implemented by PyTorch, and the CUDA version used for acceleration is 11.3, the cuDNN version is 8, and NVCC is

also supported for the compilation and acceleration of CUDA code.

A. Datasets

1) CCPD dataset

CCPD2019 [24] is a large and diverse open-source dataset specifically designed for research on Chinese urban LPs, containing 290,000 uniquely annotated LP images. For the Model evaluation experiment, a portion of the data from the CCPD-base subset was randomly chosen for training, with the remainder allocated for validation. Moreover, six subsets were utilized for model testing, as detailed in Table 2.

2) PKUData dataset

The PKUData dataset [25], contains photographs taken in various settings, enabling the evaluation of LP recognition systems across different real-world scenarios. By encompassing a broad spectrum of environmental factors, the dataset provides valuable resources for testing the robustness and generalizability of recognition Models.

3) CLPD dataset

The CLPD dataset [26], made available by Zhang et al., contains 1,200 LP images collected from 31 provincial-level administrative regions across China, excluding Taiwan, Hong Kong, and Macau. This dataset captures a diverse set of environmental conditions, providing a comprehensive representation of real-world scenarios. Together with the PKUData dataset, the CLPD dataset is frequently used to assess the performance of LP recognition Models. These datasets are valuable benchmarks, offering a broad range of challenges to evaluate the robustness and accuracy of such systems across different environments.

B. Network Model Evaluation Indicators

In this research, the proposed method is evaluated through experiments conducted on multiple public datasets, including CCPD, CLPD, and PKUData. Detection performance is measured using the Intersection over Union (IoU) metric, which evaluates the overlap between predicted bounding boxes and ground truth boxes. The IoU threshold is typically set at 0.5, and when IoU exceeds this value, the detection result is considered accurate. We calculate both IoU and recognition results, and a prediction is deemed correct when IoU exceeds the threshold and the recognition result matches.

C. Experiment of License Plate Recognition Algorithm

1) Detection

This study conducts a thorough performance comparison of the IC-YOLO (Algorithm Model 5) LP recognition Model against Faster-RCNN (Algorithm Prototype 1), C-RCNN (Algorithm Prototype 2), YOLOv3 (Algorithm Prototype 3), and YOLOv4 (Algorithm Prototype 4) using the CCPD dataset. Recognition performance is shown in Table 3. ”Detection performance is evaluated exclusively based on bounding box outcomes. Experimental results indicate that the proposed IC-YOLO (Model 5) achieves superior accuracy across most sub-datasets. Model 5 excels in both accuracy and speed among all Models, with an average accuracy of 94.9%. Its performance is particularly remarkable across all sub-datasets. It is worth noting that Model 5 achieved an identification accuracy of 96.0% on both the Base and Rotate sub-datasets, demonstrating its excellent capability in processing standard LP images and rotated images. Additionally, the accuracy of Model 5 on the Challenge sub-dataset was 94.0%, and it performed well even with more challenging data. Besides its outstanding accuracy, Model 5 also boasts a processing speed of 217.2 FPS, giving it a significant advantage in real-time processing tasks. Compared to other Models such as Prototype 4 (85.4 FPS), Prototype 3 (45.3 FPS), and Prototype 2 (26.2 FPS), the high speed of Model 5 ensures its rapid response and ability to handle higher frequency input data in practical applications. In summary, not only does Model 5 have significant advantages in accuracy, especially in complex or more challenging sub-datasets, but its high speed of 217.2 FPS makes it the best choice for real-time LP recognition tasks, showcasing its excellent comprehensive performance.

2) Recognition

We integrated the proposed LP detection method with recognition algorithms and conducted comparative experiments, including Faster-RCNN (Prototype 1) + LPR, C-RCNN (Prototype 2) + LPR, YOLOv3 (Prototype 2) + LPR, YOLOv4 (Prototype 3) + LPR, and our algorithm IC-YOLO (Model 5) + LPR. Recognition performance is shown in Table 4. The results show that Model 5 + LPR performs excellently, achieving an average accuracy of 96.7%, far exceeding other Models. It performs outstandingly across all sub-datasets, reaching 99.0% on the Base sub-dataset and 98.5% on the Rotate sub-dataset, demonstrating its strong capability in handling normal and rotated LP images. Additionally, Model 5 + LPR maintains high accuracy in challenging sub-datasets such as DB (97.0%), FN (96.5%), Tilt (95.5%), Weather (95.0%), and Challenge (95.5%), showcasing its robustness in different environments and conditions. Notably, Model 5 achieved an accuracy of 96.0% on both the Base and Rotate datasets, demonstrating its excellent ability to manage standard LP images and rotationally distorted images. In contrast, Prototype 1 + LPR has the lowest accuracy and speed, with a processing speed of only 34 FPS. Other Models (Prototype 2 + LPR, Prototype 3 + LPR, Prototype 4 + LPR) perform stably but are inferior to Model 5 + LPR in terms of accuracy and speed. Therefore, Model 5 + LPR is an efficient Model that performs excellently under various complex conditions and is highly suitable for practical applications, making it the best choice for LP recognition on the CCPD dataset. Example results are shown in Figs 9, 10, and 11.

D. Performances on mixed dataset

The generalization ability of convolutional neural networks has always been an important criterion for evaluating their performance. To verify this, we combined the training portions of the CCPD and CLPD datasets, trained an ensemble model, and embedded it into the test set for validation. The experimental results presented in Table 5 demonstrate that the ensemble model can effectively adapt to various types of LPs. By directly applying this ensemble Model to the PKU dataset, its effectiveness was further verified. Example results are shown in Figs 12, 13. In general, the experimental results show that after training on multiple datasets, the ensemble Model can accommodate various LP layouts, but the diversity

TABLE III
CCPD PERFORMANCE OF LP DETECTION ON DIFFERENT SUB-DATASETS (EXPRESSED AS PERCENTAGES). IT WAS COMPARED WITH THE FASTER-RCNN (ARCHETYPE 1) OBJECT DETECTION ALGORITHM, THE C-RCNN (ARCHETYPE 2) ALGORITHM, THE YOLOV3 (ARCHETYPE 3) MODEL, AND THE YOLOV4 (ARCHETYPE 4) METHOD. THE RESULTS SHOW THAT OUR METHOD (MODEL 5) IS MORE ACCURATE ON MOST SUB-DATASETS. ADDITIONALLY, OUR METHOD CAN RUN IN REAL TIME.

Method	Avg	Base	DB	FN	Rotate	Tilt	Weather	Challenge	Speed(FPS)
Prototype 1	89.0	92.5	93.0	89.5	85.5	90.5	89.5	83.5	17.5
Prototype 2	91.5	93.5	91.0	89.5	92.5	93.5	89.0	93.5	26.2
Prototype 3	91.5	94.0	93.5	89.0	89.0	90.5	91.0	93.5	45.3
Prototype 4	91.6	95.0	94.0	88.5	91.0	90.0	91.5	91.5	85.4
Model 5	**94.9**	**96.0**	**95.0**	**96.0**	**95.5**	**94.5**	**93.0**	**94.0**	**217.2**

TABLE IV
PERFORMANCE OF LP RECOGNITION ON DIFFERENT SUB-DATASETS OF CCPD (EXPRESSED AS PERCENTAGES). IT WAS COMPARED WITH FASTER-RCNN (PROTOTYPE 1) + LPR, C-RCNN (PROTOTYPE 2) + LPR, YOLOV 3 (PROTOTYPE 2) + LPR, AND YOLOV 4 (PROTOTYPE 3) + LPR. THE RESULTS SHOW THAT OUR METHOD (IC-YOLO (MODEL 5) + LPR) IS MORE ACCURATE ON THE SUB-DATASETS.

Method	Avg	Base	DB	FN	Rotate	Tilt	Weather	Challenge	Speed(FPS)
Prototype 1 + LPR	89.9	91.5	92.0	89.5	85.5	91.5	90.5	88.5	34
Prototype 2 + LPR	91.2	91.0	91.0	90.0	92.5	92.0	90.0	92.5	40
Prototype 3 + LPR	92.5	94.5	93.5	91.5	90.0	92.5	93.0	92.5	45
Prototype 4 + LPR	92.6	95.0	94.5	90.5	91.5	92.0	93.5	91.0	26
Model 5 + LPR	**96.7**	**99.0**	**97.0**	**96.5**	**98.5**	**95.5**	**95.0**	**95.5**	**60**

TABLE V
PERFORMANCES (AS PERCENTAGES) ON MIXED DATASET.

Performance(%)	Unified Model(%)	Single Model(%)	Gap
CCPD	96.5	96.0	0.5
CLPD	95.5	94.5	1
PKUData	95.0	93.0	2

and distribution of the training dataset will to some extent affect the overall efficacy of the Model.

E. Ablation study

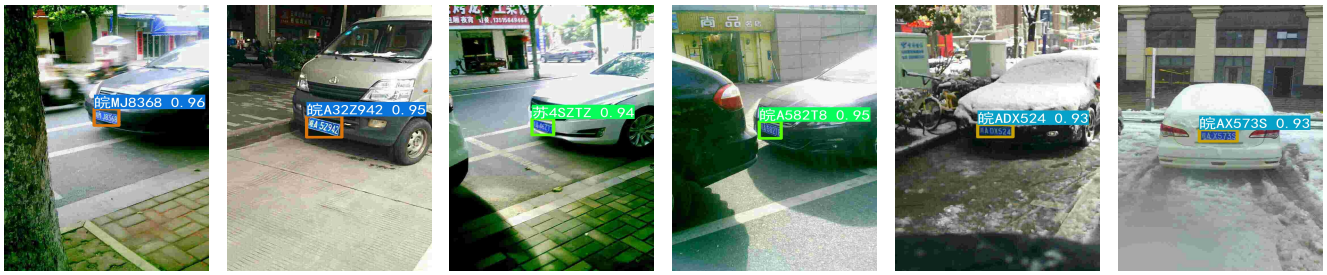
To verify the effectiveness and rationality of the proposed method, ablation study experiments were conducted entirely based on the CCPD dataset. This dataset is large and diverse, thus avoiding the influence of accidental factors. To assess the contributions of various modules, we designed a series of comparative experiments, with the results presented in Table 6. First, we analyzed the impact of introducing Channel feature fusion and CFFI on Model performance through comparative experiments. The experimental outcomes revealed that the Model's efficacy exhibited a significant enhancement on the Rotate sub-dataset (95.5% vs. 92.0%) and Weather (94.5% vs. 93.0%) sub-datasets, but slightly decreased on the Challenge sub-dataset (93.0% vs. 94.0%). To evaluate the improvement of CBAM on Model performance, we compared Models with and without CBAM. The experimental results showed that CBAM significantly improved Model performance, especially on the Base (96.5% vs. 93.0%) and Rotate (95.5% vs. 92.0%) sub-datasets, and also achieved significant accuracy improvements in complex scenarios such as Challenge (95.0% vs. 91.0%).

TABLE VI
AN ABLATION STUDY OF THE PROPOSED METHOD WAS CONDUCTED. WE COMPARED THE DETECTION ACCURACY OF MODELS WITH AND WITHOUT CHANNEL FEATURE FUSION (CFFI), AS WELL AS THE RECOGNITION ACCURACY OF MODELS WITH AND WITHOUT THE SE AND CBAM MODULES.

Performance(%)	Base	FN	DB	Rotate	Weather	Tilt	Challenge
without channel feature fusion and CFFI	95.0	94.0	90.5	92.0	93.0	93.5	94.0
with channel feature fusion and CFFI	**95.5**	**94.0**	**94.0**	**95.5**	**94.5**	**95.5**	**93.0**
without SE_Block	94.5	93.0	92.5	92.0	93.0	94.5	94.0
with SE_Block	**96.0**	**95.0**	**94.5**	**93.5**	**95.5**	**94.0**	**93.0**
without CBAM	93.0	93.0	92.5	92.0	93.5	92.5	91.0
with CBAM	**96.5**	**95.0**	**94.0**	**94.5**	**95.5**	**94.5**	**95.0**

(a) (b) (c) (d) (e) (f)

Fig. 9. ICS-Net in CCPD-Base recognition results as in figs. (a) and (b); in CCPD-DB recognition results as in figs (c) and (d); in CCPD-FN recognition results as in figs (f) and (g).



(a) (b) (c) (d) (e) (f)

Fig. 10. ICS-Net in CCPD-Rotate recognition results as in (a) and (b); in CCPD-Tilt recognition results as in (c) and (d); in CCPD-Weather recognition results as in (f) and (g).

(a) (b)

Fig. 11. ICS-Net in CCPD-Challenge recognition results as shown in (a) and (b).

This indicates that CBAM excels in capturing spatial and channel information, effectively enhancing feature extraction capabilities. Finally, to verify the role of SE Block, we compared Models with and without SE Block, showing that adding SE Block significantly improved Model performance on multiple sub-datasets, particularly on Base (96.0% vs. 94.5%) and FN (95.0% vs. 93.0%), indicating that SE Block helps the Model better focus on key features. Overall, the three methods effectively improved Model performance on all sub-datasets, especially in handling challenging data, demonstrating stronger robustness.

V. CONCLUSION

This research introduces an innovative algorithm called ICS-Net, specifically developed to address the challenges of LP detection and recognition. Unlike traditional feature extraction techniques, ICS-Net uses IC-YOLOv5 for LP detection. To further optimize network performance, a Channel Feature Fusion and Inversion (CFFI) module is incorporated between the backbone and neck networks, enhancing the transmission and processing of channel information, thereby significantly improving overall detection accuracy. In addition, the Channel and Spatial Attention Module (CBAM)

(a) (b) (c) (d)

Fig. 12. ICS-Net in PKUData recognition results are shown in (a) and (b) and (c) and (d)

(a) (b) (c) (d)

Fig. 13. ICS-Net in CLPD recognition results are shown in (a) and (b) and (c) and (d)

was introduced into the backbone network, enabling it to precisely extract key channel and spatial information while automatically filtering out redundant features, thereby achieving more accurate LP detection and localization. For LP character recognition, ICS-Net utilizes an advanced version of the SE-LPRNet neural network algorithm. This approach integrates the Squeeze-and-Excitation (SE) module, strategically positioned within the network to prioritize key features and suppress irrelevant information effectively. By adhering to the standard SE module insertion strategy, the network's recognition capability is substantially enhanced, while training performance is also optimized. Furthermore, the input structure of the network is modified to enrich the diversity of initial input features, accelerating Model convergence and improving the overall efficiency of the algorithm.

To validate the effectiveness and robustness of the ICS-Net algorithm, initial training was performed on the CCPD-base dataset, followed by testing on six subsets of this dataset to comprehensively evaluate the model's performance across diverse scenarios. To further evaluate the algorithm's detection and recognition capabilities in complex environments, additional testing was performed using the PKUData and CLPD datasets. Experimental results demonstrate that ICS-Net maintains a high level of recognition accuracy across various challenging environments, highlighting its exceptional robustness and efficiency. These research results indicate the method's substantial potential for application in LP detection and recognition tasks, particularly in real-world scenarios characterized by diverse and complex conditions.

REFERENCES

[1] Sasikala, S., R. Neelaveni, and P. Sweety Jose. "An Intelligent Traffic Analysis and Prediction System Using Deep Learning Technique." International Journal on Artificial Intelligence Tools 33.01 (2024): 2350055.

[2] Bhargav, Rohit, and Parag Deshpande. "LP characters recognition using colour, scale, and rotation independent features." IETE Journal of Research 69.11 (2023): 8006-8018.

[3] Hussain, Bydaa Ali, and Mohammed Sadoon Hathal. "Development of Iraqi LP recognition system based on Canny edge detection method." Journal of Engineering 26.7 (2020): 115-126.

[4] Leng, Jiancai, et al. "A Light Vehicle License-Plate-Recognition System Based on Hybrid Edge–Cloud Computing." Sensors 23.21 (2023): 8913.

[5] Han, Chunhui, and Zhao Chen. "Research on Blob Analysis Algorithm-based Segmentation of Characters in LP." Academic Journal of Computing & Information Science 2.2 (2019): 39-48.

[6] Yang, Yuping. "LP character segmentation algorithm based on improved regression Model." Journal of Physics: Conference Series. Vol. 1453. No. 1. IOP Publishing, 2020.

[7] Huang, Qiuying, Zhanchuan Cai, and Ting Lan. "A new approach for character recognition of multi-style vehicle LPs." IEEE Transactions on Multimedia 23 (2020): 3768-3777.

[8] Lin, Guocong, et al. "LP recognition based on mathematical morphology and template matching." 2019 Chinese Automation Congress (CAC). IEEE, 2019.

[9] Pirgazi, J., A. G. Sorkhi, and M. M. P. Kallehbasti. "An efficient robust method for accurate and real-time vehicle plate recognition. J. Real-Time Image Proc. 18 (5), 1759–1772 (2021).

[10] Kibaara, Peter Muthuri, Edna C. Too, and David Gitonga Mwathi. "Convolutional neural networks and support vector machines for hybrid number plate recognition Model." International Journal of Hybrid Intelligence 2.2 (2023): 128-150.

[11] Wang, Zhichao, et al. "Research and Implementation of Fast-LPRNet Algorithm for LP Recognition." Journal of Electrical and Computer Engineering 2021.1 (2021): 8592216.

[12] Zou, Yongjie, et al. "LP detection and recognition based on YOLOv3 and ILPRNET." Signal, Image and Video Processing 16.2 (2022): 473-480.

[13] Park, Se-Ho, et al. "An all-in-one vehicle type and LP recognition system using YOLOv4." Sensors 22.3 (2022): 921.

[14] Zhu, Qianqian, et al. "Research on LP location algorithm based on yolov5." Journal of Physics: Conference Series. Vol. 2278. No. 1. IOP Publishing, 2022.

[15] Qin, Shuxin, and Sijiang Liu. "Towards end-to-end car LP location and recognition in unconstrained scenarios." Neural Computing and Applications 34.24 (2022): 21551-21566.

[16] Kaur, Parneet, et al. "Automatic LP Recognition System for Vehicles Using a CNN." Computers, Materials & Continua 71.1 (2022).

[17] Xu, Fan, et al. "A CRNN-based method for Chinese ship LP recognition." IET Image Processing 18.2 (2024): 298-311.

[18] Zheng, Yujie, Lei Guan, and Haohong Li. "The Low-light LP Recognition via CNN." Journal of Physics: Conference Series. Vol. 2424. No. 1. IOP Publishing, 2023.

[19] Rao, Zhan, et al. "LP recognition system in unconstrained scenes via a new image correction scheme and improved CRNN." Expert Systems with Applications 243 (2024): 122878.

[20] Zhang, Weiguo, et al. "Research on the algorithm of LP recognition

based on MPGAN Haze Weather." IEICE TRANSACTIONS on Information and Systems 105.5 (2022): 1085-1093.
[21] Tao, Lingbing, et al. "A Real-Time LP Detection and Recognition Model in Unconstrained Scenarios." Sensors 24.9 (2024): 2791.
[22] Chauhan, Anshumaan, et al. "LPRNet: a novel approach for novelty detection in networking packets." International Journal of Advanced Computer Science and Applications 13.2 (2022).
[23] Chauhan, Anshumaan, et al. "LPRNet: a novel approach for novelty detection in networking packets." International Journal of Advanced Computer Science and Applications 13.2 (2022).
[24] Xu, Zhenbo, et al. "Towards end-to-end LP detection and recognition: A large dataset and baseline." Proceedings of the European conference on computer vision (ECCV). 2018.
[25] Yuan, Yule, et al. "A robust and efficient approach to LP detection." IEEE Transactions on Image Processing 26.3 (2016): 1102-1114.
[26] Zhang, Linjiang, et al. "A robust attentional framework for LP recognition in the wild." IEEE Transactions on Intelligent Transportation Systems 22.11 (2020): 6967-6976.