Received: 5 November 2021 | Revised: 12 December 2021 | Accepted: 10 January 2022 | Published online: 27 January 2022

Journal of Computational and Cognitive Engineering
2022, Vol. 1(3) 147–151
DOI: 10.47852/bonviewJCCE20514

BON VIEW PUBLISHING

RESEARCH ARTICLE

Research on Vehicle Automatic Driving Target Perception Technology Based on Improved MSRPN Algorithm

Min Yang[1,*]

[1]*College of Mechanical Engineering, Nanjing University of Science and Technology, China*

Abstract: Vehicle automatic driving technology can effectively improve the safety performance of vehicle driving. This research is aimed at the needs of vehicle automatic driving. Combined with vehicle perception technology, a better target recognition algorithm is proposed. By comparing the recognition and recall rate of ION algorithm, HYPERNET algorithm, R-CNN (Regions with CNN features) algorithm, and multi-strategy region proposal network (MSRPN) algorithm, it can be seen that MSPRN algorithm has better algorithm performance and is suitable for target detection and recognition in vehicle automatic driving.

Keywords: vehicles, automatic driving, MSPRN, target recognition, R-CNN

1. Introduction

Automobile is one of the indispensable means of transportation in people's life. At present, it has become a very important component in the human world, and the automobile industry has also developed into an important industry. With the development of modern computer technology and the progress of Internet technology, vehicle automatic driving technology has become a reality (Lee & Lee, 2020; Li et al., 2019; Kale et al., 2019; Ota et al., 2019). In the process of traffic driving, due to the limitations of self-reaction, human drivers are difficult to accurately identify and judge the road conditions in a very short time, may make mistakes in decision-making, and may not be able to take correct actions in operation. Therefore, the emergence of automatic driving technology can no longer need human to control the vehicle and can improve the driving safety of the vehicle (Bulgakov et al., 2020; Redzuan et al., 2019; Sun et al., 2021). In the process of vehicle automatic driving, the computer system must be able to accurately identify the driving environment and effectively identify the surrounding targets. In the traditional recognition technology, monocular vision target recognition technology is more adopted. This recognition technology has some shortcomings in both accuracy and efficiency (Sharma et al., 2019; Vaiyapuri et al., 2021). In the follow-up research, R-CNN (Regions with CNN features) algorithm is proposed for target detection and recognition. However, the algorithm is still rough in target detection and classification (Arnold et al., 2020; Cai et al., 2020; Marti et al., 2019). On this basis, this research has been optimized and improved, hoping to propose a vehicle automatic driving target recognition algorithm with excellent performance.

At the beginning of this research innovation, the main points are as follows: (1) a vehicle automatic driving target recognition algorithm integrating MSRPN and fast R-CNN algorithm is proposed, which can effectively solve some problems existing in traditional recognition algorithms. (2) Comparing the performance of the proposed multi-strategy region proposal network (MSPRN) algorithm with the traditional R-CNN algorithm, it can be seen that the algorithm proposed in this paper has higher recall and recognition rate.

2. Construction of Machine Vision Perception Platform for Vehicle Automatic Driving

In vehicle automatic driving, the perception of the surrounding environment is very dependent on machine vision technology. The parameters of the surrounding environment are obtained through the on-board camera, so as to serve automatic driving (Chen et al., 2020; Schulz et al., 2021; Sridhar & Eskandarian, 2019; Yang et al., 2020). As required, the machine vision perception platform for automatic vehicle driving is equipped with six cameras, which are arranged according to three different functions: front view, measuring view, and rear view. The front camera mainly captures the image information of the road in front of the vehicle, obtains the front environment according to the obtained information, and empties the information for the system's early warning and vehicle control through in-depth analysis (Christy, 2019; Merriman et al., 2021; Seo et al., 2021). The rear view camera is mainly used to obtain the rear

*Corresponding author: Min Yang, College of Mechanical Engineering, Nanjing University of Science and Technology, China. Email: yangmin082@njust.edu.cn

situation of the vehicle. Combined with the side view camera, the blind area of the vehicle can be eliminated. All cameras think of the 360° field of vision information acquisition around the vehicle, which can build the machine vision recognition function of the vehicle and provide services for the safe automatic driving of the vehicle.

The front view camera is mainly used to detect the road conditions in front of the vehicle, including vehicles, traffic signs, pedestrians, etc. In order to ensure the accuracy of detection information, it is required that the camera applied to machine vision recognition monument should have a higher pixel camera. With the distance between the detection target and the camera getting farther and farther, the pixels of the obtained video image will gradually decline, and finally the target will be lost. In order to avoid rear-end collision, a safe distance model needs to be constructed.

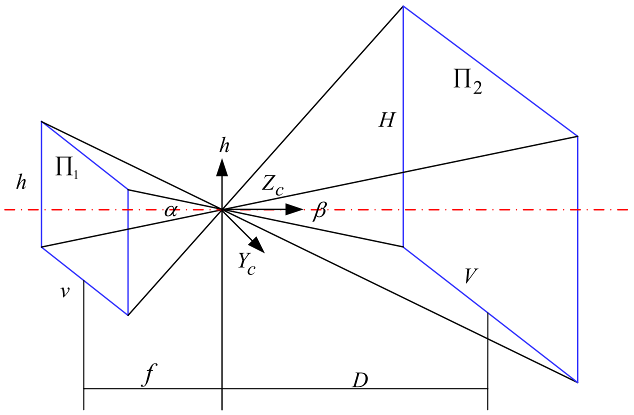
Suppose the speed and braking deceleration of the vehicle are respectively v_1, a_1, the speed and braking deceleration of the vehicle in front of the form are respectively v_2, a_2, the braking reaction time of the driver is t_z, the coordination time of the vehicle braking system when the driver takes braking action is t_x, and the minimum safety distance to be maintained when the two vehicles are stationary is d_0. Set the safe distance of the vehicle as d_w. Thus, the kinematic safety distance model during vehicle braking can be constructed, and the formula is

$$d_w = \frac{v_1^2}{2a_1} - \frac{v_2^2}{2a_2} + v_1(t_z + t_x) + d_0 \tag{1}$$

When the value range of d_0 is set to 1~2 m, it can be calculated that the distance of vehicle identification in front should be ≥110 m. When selecting a camera, it is necessary to select a pinhole camera with high definition, and its basic imaging model is shown in Figure 1. In Figure 1, v, V represent the imaging width and the width of the detection target, respectively, h, H represent the imaging height and the height of the detection target, respectively, and f, D represent the imaging plane and the distance from the detection target to the projection center, respectively.

When selecting a camera, it is necessary to select a pinhole camera with high definition, and its basic imaging model is shown in Figure 1. In the figure, v, V represent the imaging width and the width of the detection target, respectively, h, H represent the imaging height and the height of the detection target, respectively, and f, D represent the imaging plane and the distance from the detection target to the projection center, respectively.

Figure 1
Vehicle camera imaging model



According to the principle of similar triangle, we can get:

$$v = f\frac{V}{D} \tag{2}$$

$$h = f\frac{H}{D} \tag{3}$$

The image data collected by the camera can be calculated by the system, judge the state of the vehicle and the surrounding conditions, and control the vehicle as needed.

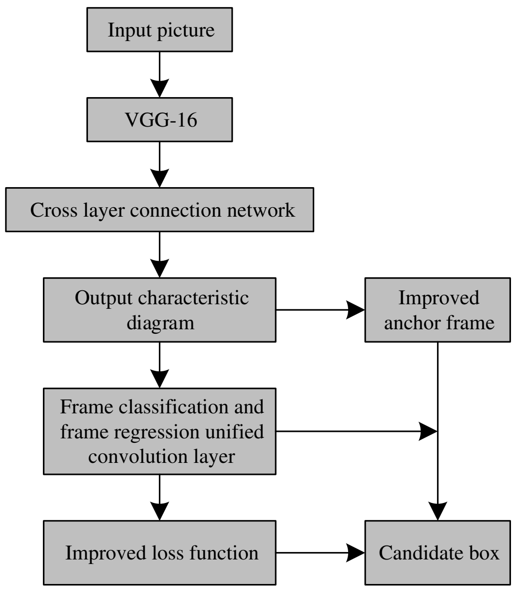
3. Design of Vehicle Automatic Driving Target Recognition Algorithm Based on MSPRN

3.1. Optimization and improvement strategy of MSPRN

In image detection and recognition technology, Faster R-CNN algorithm is a common detection technology. Using the RPN in the algorithm model, the height width ratio and area of the detection target can be scaled to generate regional candidate boxes (Lv et al., 2019). Although the algorithm can have good image detection effect, the number and quality of the generated region candidate boxes are not ideal, which will have an adverse impact on the detection effect of Faster R-CNN algorithm (Sathiskumar et al., 2020). In the process of vehicle automatic driving, the traditional Faster R-CNN algorithm cannot meet its target detection and surrounding environment judgment, so it needs to be optimized and improved. The regional candidate box network can be improved. This time, a multi-strategy regional candidate box (MSPRN) is proposed to be combined with Faster R-CNN algorithm.

The specific optimization and improvement ideas of MSPRN are as follows: (1) a cross layer connection network is proposed to integrate the characteristics of multiple convolution layers, so as to improve the ability of pooling layer and the quality of regional candidate frame. (2) In order to improve the localization ability of target detection, an anchor frame with uniform scaling area is

Figure 2
MSPRN network structure diagram



introduced. (3) In order to improve the training speed and testing speed of the algorithm, the regression layer and classification layer are unified to obtain a convolution layer. (4) In order to improve the regression performance of anchor frame, the RPN task loss function is improved. According to the optimization and improvement strategy, the network structure diagram of MSPRN can be obtained, as shown in Figure 2. After image input, pre-training is carried out first, and then VGG-16 is used for feature extraction. Based on the cross layer connection network, the convolution layer features generated by VGG-16 model are fused, and finally the output feature map is calculated using the improved loss function.

3.2. Feature extraction method

In the traditional RPN method, the method of generating regional candidate frame through top-level convolution feature is slightly rough, and the resolution of the generated top-level convolution feature image is not high, so the quality of the final generated regional candidate frame is low (Cai et al., 2020). In order to solve these problems and increase the recognition ability of RPN for small targets, the multi-layer convolution feature fusion method (MLFC) is introduced for optimization and improvement. The cross layer connection network structure after optimization and improvement is shown in Figure 3.

Although low convolution features have low semantics, they also have high-resolution features. The fusion of high-level and low-level convolution features can improve the quality of regional candidate frames. The average pooling layer is an important supplement to the maximum pooling layer. At the same time, the average pooling layer and the maximum pooling layer are used to process the input features, and then the connection layer is used to fuse the pooling results of the two parts.

3.3. Optimization and improvement of anchor frame

In the original RPN, there are some problems in the design of the scaling area and aspect ratio of the anchor frame, which makes the target positioning ability insufficient. In the original design of anchor frame height width ratio, a convolution layer is used for feature extraction, and two 1×1 convolutions are added behind 2×2 convolution layer.

Let the height and width of the output characteristic graph be h_o, w_o, the height and width of the input characteristic graph be h_i, w_i, the size height and width of the convolution kernel are $\ker nel\_h, \ker nel\_w$, the step height and width are $stride\_h, stride\_w$, and the zero filling height and width are pad_h, pad_w. The height calculation formula of the output characteristic graph is shown in equation (4), and the corresponding width calculation formula is shown in equation (5).

$$h\_o = \frac{(h\_i + 2 * pad\_h - \ker nel\_h)}{stride\_h} + 1 \tag{4}$$

$$w\_o = \frac{(w\_i + 2 * pad\_w - \ker nel\_w)}{stride\_w} + 1 \tag{5}$$

In the VGG-16 model, the aspect ratio of the input image and the convolution feature image is consistent, which can reflect the size of the image. If there is a large difference between the aspect ratio and the feature image, the number of effective anchor frames will be reduced.

When improving the anchor frame, six anchor frames with different scaling areas are designed at each sliding frame position. The scaling areas from small to large are 50×50, 150×150, 250×250, 350×350, 450×450, and 550×550, respectively. The spacing of all scaling areas is evenly distributed. The improved anchor frame is shown in Figure 4.

In the training stage, the maximum anchor frame can be obtained by overlapping the training anchor frame with the real marked anchor frame. The anchor frame with an overlap value greater than 0.6 is set as a positive sample, that is, the anchor frame can cover the target to be detected. If the overlap value is less than 0.3, it is regarded as a negative sample. Since MSPRN adopts 50×50, 150×150 anchor frames, so it is easier to detect small targets, and its search ability has been enhanced. By increasing the number of effective anchor frames, its positioning ability can be comprehensively improved. For 1000×600 size input picture, its output characteristic image resolution is 125×75. According to the calculation of six anchor frames at each sliding frame position, a total of about 60,000 anchor frames can be generated.

Figure 3
Improved cross layer connection network structure diagram

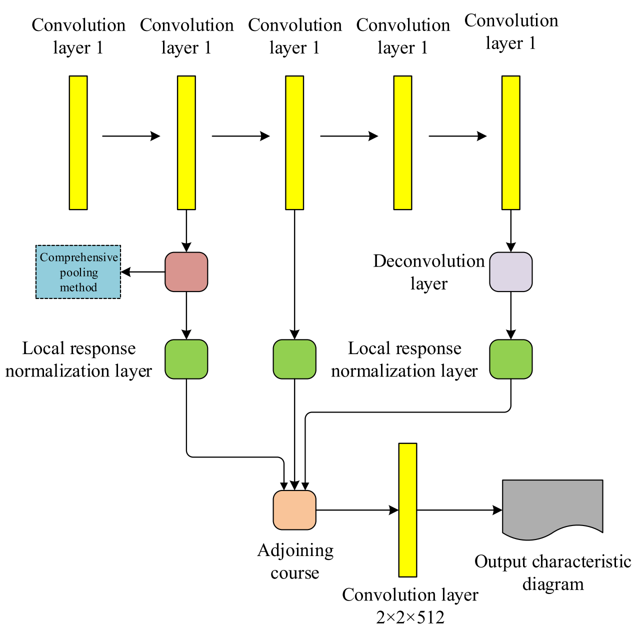


Figure 4
Schematic diagram of anchor frame improvement

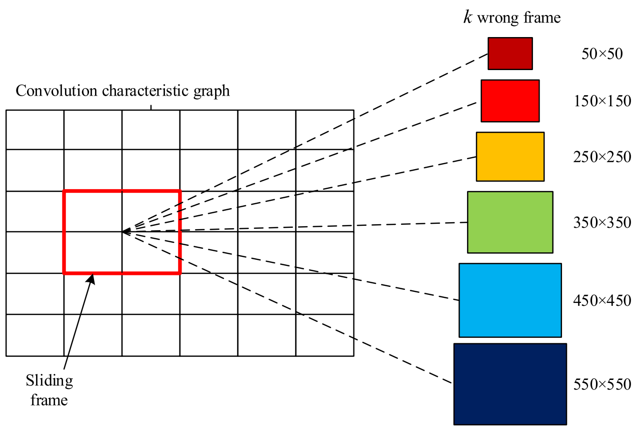
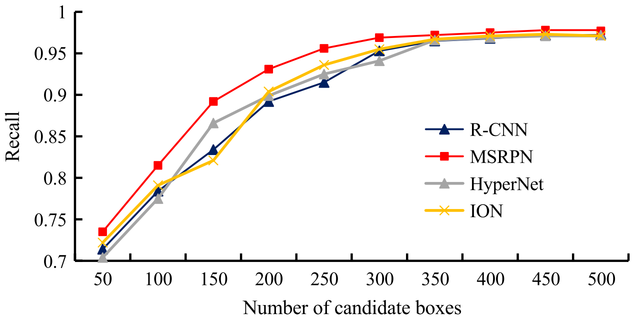


Table 1
Main parameter setting of MSRPN algorithm

Parameter name	Base_lr	lr_policy	Gamma	Stepsize	Momentum	Weight_decay	Nms_IoU
Numerical value	0.01	"step"	0.1	50000	0.9	0.0005	0.75

Figure 5
Comparative analysis of recall rate of four algorithms



In detail, set the short side of the input picture to 600, and its aspect ratio remains unchanged. The subsequent pictures used for training and testing will be adjusted to a fixed proportion. After the CGG-16 model is used to adjust the picture size, MSPRN is used for processing. During training and testing, reduce the number of overlapping area candidate boxes when discarding the out of bounds anchor box, set variable *NMS_IoU* to 0.75, and retain the top 150 area candidate boxes after NMS processing.

4. Simulation Analysis of Experimental Results

In order to verify the availability and comprehensive performance of the proposed algorithm, simulation experiments are carried out on the algorithm. The experimental data select the data set MS COCO for testing and training. The data have 3.28×10^5 pictures, including 80 categories of pictures. The parameter settings of MSRPN algorithm are shown in Table 1:

The simulation performance is compared and analyzed by recall rate and recognition rate. The number of pictures is 500. The pictures are all possible object targets during vehicle driving, such as pedestrians, vehicles, birds, and bicycles.

Firstly, the recall rate is compared and analyzed. ION algorithm, HYPERNET algorithm, R-CNN algorithm, and MSRPN algorithm are selected for comparison and analysis. The comparison and analysis results of algorithm recall rate of 500 candidate boxes are shown in Figure 5. As can be seen from Figure 5, with the increase of the number of candidate boxes, the recall rates of the four algorithms continue to increase. The recall rate increases faster in the early stage and tends to be stable in the later stage. In contrast, MSPRN algorithm has excellent recall rate, which shows that the improvement effect is significant.

Further analyze the recognition rate of the four algorithms, that is, detect the target in 500 pictures. The recognition rate of the two algorithms is shown in Table 2. It can be seen from Table 1 that the recognition rate of MSPRN algorithm is better than that of R-CNN algorithm and the other two algorithms. When the number of candidate boxes reaches 500, the recognition rate of MSPRN algorithm is 80.15%, which is 10.58%, 7.51%, and 1.67% higher than that of R-CNN algorithm, ION algorithm, and HYPERNET algorithm, respectively.

5. Conclusion

In this study, an optimized and improved multi-strategy region candidate box network target recognition and detection algorithm is proposed for vehicle automatic driving technology. In order to verify the performance of the algorithm, 500 images are selected to simulate and compare ION algorithm, HYPERNET algorithm, R-CNN algorithm, and MSRPN algorithm. Through simulation analysis, it can be seen that MSPRN algorithm has better recall rate and its algorithm recognition rate is higher than the other three algorithms. From the comprehensive analysis, MSPRN algorithm has better recall and recognition rate and is suitable for vehicle automatic driving technology.

Acknowledgement

The work described in this paper was supported by the General Natural Science Foundation in 2016 of Jiangsu Educational Committee, PR China, "Research on driving stability and control of tilt steering three wheel ENV" (16KJD460004).

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

References

Arnold, E., Dianati, M., & Temple, R. D. (2020). Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors. *IEEE Transactions on Intelligent*

Table 2
Comparative analysis of recognition rates of two algorithms (%)

Algorithm	50	100	150	200	250	300	350	400	450	500
R-CNN	91.53	90.76	87.24	85.18	83.26	80.17	78.55	74.39	72.04	69.57
MSRPN	98.24	97.55	94.18	90.24	88.99	86.25	85.18	84.56	82.97	80.15
ION	92.45	89.45	88.77	86.25	84.17	83.25	80.16	78.11	76.25	74.13
HyperNet	94.15	92.56	90.18	87.56	85.24	84.15	82.54	81.19	79.54	78.81

Transportation Systems, *23*(3), 1852–1864. https://doi.org/10.1109/TITS.2020.3028424.

Bulgakov, V. M., Kuvachov, V. P., Holovach, I. V., & Ihnatiev, Y. I. (2020). Investigation of automatic driving of a mobile wide span agricultural vehicle in the wake of a constant technology track. *Mechanization and Electrification of Agricultural*, *110*(11), 12–26. https://doi.org/10.37204/0131-2189-2020-11-1.

Cai, P., Wang, S., Sun, Y., & Liu, M. (2020). Probabilistic end-to-end vehicle navigation in complex dynamic environments with multimodal sensor fusion. *IEEE Robotics and Automation Letters*, *5*(3), 4218–4224. https://doi.org/10.1109/LRA.2020.2994027.

Chen, G., Wang, F., Li, W., Hong, L., & Konll, A. (2020). NeuroIV: Neuromorphic vision meets intelligent vehicle towards safe driving with a new database and baseline evaluations. *IEEE Transactions on Intelligent Transportation Systems*, *23*(2), 1171–1183. https://doi.org/10.1109/TITS.2020.3022921.

Christy, A. (2019). Artificial intelligence based automatic decelerating vehicle control system to avoid misfortunes. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(6), 3129–3134. https://doi.org/10.30534/ijatcse/2019/75862019.

Kale, B. S., Gorale, G. L., Khalkar, U. K., & Patil, S. V. (2019). Smart vehicle with automatic pneumatic bumper system and video surveillance. *International Journal of Electronics Engineering Research*, *11*(1), 1–8. http://ripublication.com/ijeer19/ijeerv11n1_01.pdf

Lee, J. W., & Lee, B. R. (2020). Design of the driving controller of automatic guided vehicle based on fuzzy inference system. *The Journal of Korean Institute of Information Technology*, *18*(6), 11–16. https://doi.org/10.14801/jkiit.2020.18.6.11.

Li, B., Yu, Z., Huang, L., & Guo, B. (2019). Vehicle departure pattern and queue length prediction at an isolated intersection with automatic vehicle identity detection. *IET Intelligent Transport Systems*, *13*(12), 1804–1813. https://doi.org/10.1049/iet-its.2019.0117.

Lv, B., Sun, R., Zhang, H., & Yue, R. (2019). Automatic vehicle-pedestrian conflict identification with trajectories of road users extracted from roadside LiDAR sensors using a rule-based method. *IEEE Access*, *7*, 161594–161606. https://doi.org/10.1109/ACCESS.2019.2951763.

Marti, E. D., Miguel, M., Garcia, F., & Perez, J. (2019). A review of sensor technologies for perception in automated driving. *IEEE Intelligent Transportation Systems Magazine*, *11*(4), 94–108. https://doi.org/10.1109/MITS.2019.2907630.

Merriman, S. E., Plant, K. L., Revell, K., & Stanton, N. A. (2021). Challenges for automated vehicle driver training: A thematic analysis from manual and automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *76*, 238–268. https://doi.org/10.1016/j.trf.2020.10.011.

Ota, T., Iwasaki, Y., Nakano, A., Yoshinaga, K., & Higashide, T. (2019). Development of small automatic guided vehicle by contact detection to hydroponics cultivation system. *Japan Agricultural Research Quarterly*, *53*(1), 31–40. https://doi.org/10.6090/jarq.53.31.

Redzuan, S. M., Nuraida, M. H., Norzalina, O., & Noorhelinahani, A. B. (2019). Development of ultra sonic sensor driven automatic vehicle reverse safety system. *International Journal of Scientific & Technology Research*, *8*(1), 19–21.

Sathiskumar, S., Navean, G. V., Hari Prakash, R., & Vishnu Praveen, S. (2020). Automatic vehicle speed control system in a restricted zone. *International Journal of Scientific & Technology Research*, *9*(2), 2207–2210.

Schulz, Y., Mattar, A. K., Hehn, T. M., & Kooij, J. F. P. (2021). Hearing what you cannot see: Acoustic vehicle detection around corners. *IEEE Robotics and Automation Letters*, *6*(2), 2587–2594. https://doi.org/10.1109/LRA.2021.3062254.

Seo, K., Lee, J., Lee, J. Y., Lee, J., & Yi, K. (2021). Fail safe process of vehicle localization for reliability improvement of LV3 autonomous driving. *International Journal of Automotive Technology*, *22*(2), 529–535. https://doi.org/10.1007/s12239-021-0049-8.

Sharma, P., Singh, A., Raheja, S., Raheja, S., & Singh, K. (2019). Automatic vehicle detection using spatial time frame and object based classification. *Journal of Intelligent and Fuzzy Systems*, *37*(3), 1–11. https://doi.org/10.3233/JIFS-190593.

Sridhar, S., & Eskandarian, A. (2019). Cooperative perception in autonomous ground vehicles using a mobile-robot testbed. *IET Intelligent Transport Systems*, *13*(10), 1545–1556. https://doi.org/10.1049/iet-its.2018.5607.

Sun, T., Gao, Z., Gao, F., Zhang, T., & Chen, S. (2021). Intelligent vehicle automatic stop-and-go task based on humanized learning control model. *Advances in Civil Engineering*, *2021*(26), 1–11. https://doi.org/10.1155/2021/8867091.

Vaiyapuri, T., Mohanty, S. N., Sivaram, M., & Shankar, K. (2021). Automatic vehicle license plate recognition using optimal deep learning model. *Computers, Materials and Continua*, *67*(2), 1881–1897. https://doi.org/10.32604/cmc.2021.014924.

Yang, P., Duan, D., Chen, C., Cheng, X., & Yang, L. (2020). Multi-Sensor Multi-Vehicle (MSMV) localization and mobility tracking for autonomous driving. *IEEE Transactions on Vehicular Technology*, *69*(12), 14355–14364. https://doi.org/10.1109/TVT.2020.3031900.

How to Cite: Yang, M. (2022). Research on Vehicle Automatic Driving Target Perception Technology Based on Improved MSRPN Algorithm. *Journal of Computational and Cognitive Engineering* 1(3), 147–151, https://doi.org/10.47852/bonviewJCCE20514