

Automatic target detection in vehicle images based on YOLOv10 deep learning algorithm

Davidson Gao

University of Southern California, 3551 Trousdale Pkwy, Los Angeles, CA, 98004, USA

hgao861@gmail.com

Abstract. In this paper, we explore the effect of its application in vehicle image detection based on the latest YOLOv10 model. YOLOv10, as the latest version in the YOLO series, inherits and optimises the advantages of the previous generations of models, aiming to provide higher detection accuracy and faster inference speed. In our experiments, we firstly divided the dataset reasonably and trained the YOLOv10 model with data from the training set. By analysing the Precision-Recall curve, we found that the area below the PR curve is larger, which indicates that the AUC value of the model is close to 1, thus reflecting the superiority of the model in classification performance. In addition, from the F1-confidence curve, the model is still able to maintain good classification results at higher confidence levels and maintains high F1 scores in most confidence intervals, which further validates the reliability of the model. Test results show that the trained YOLOv10 model is able to accurately recognise vehicles and effectively predict vehicle types. This enables the model to monitor the vehicle conditions on the road in real time and achieve accurate detection and classification of road traffic targets. Therefore, this paper demonstrates the powerful ability of YOLOv10 in the field of vehicle detection, which provides an important reference for the development of future intelligent transport systems. Overall, this study not only verifies the effectiveness of YOLOv10 in practical applications, but also lays the foundation for further improvement of target detection technology.

Keywords: YOLOv10, Vehicle image detection, AUC.

1. Introduction

With the development of intelligent transport systems, vehicle image target detection has become an important research direction in the field of computer vision and automatic driving [1]. Traditional traffic monitoring and management methods often rely on manual monitoring, which is not only inefficient but also susceptible to human factors. In recent years, with the rapid development of deep learning technology, especially the breakthroughs in the field of image processing, it makes automated vehicle detection possible [2]. In application scenarios such as urban traffic management, intelligent parking systems, and driverless cars, it is crucial to identify and locate vehicles accurately and in real time. This not only helps to improve the efficiency of traffic flow management, but also effectively reduces traffic accidents. Therefore, it is very important to use computer vision techniques to achieve automatic vehicle detection and identification.

Deep learning algorithms, especially Convolutional Neural Networks (CNNs), have demonstrated powerful capabilities in image processing tasks. YOLO (You Only Look Once) [3], as an efficient real-time target detection algorithm, plays an important role in target detection of vehicle images. The core concept of YOLO is to consider target detection as a regression problem, and through a single neural network directly predicts bounding box and category probabilities to achieve fast and accurate object detection.

Since its first release, YOLO has gone through several iterations, from the initial version to the current YOLOv7 and YOLOv8, with each update improving both accuracy and speed. For example, YOLOv4 introduced some new features such as CSPNet [4] and Mish [5] activation functions, which substantially improved the model performance. And some recent versions further optimise the model structure, making it lighter and more suitable for embedded devices and mobile platforms. In practical applications, YOLO is widely used for real-time recognition of objects such as pedestrians, vehicles, and traffic signs in self-driving cars. In addition, it can be used for urban traffic monitoring to optimise signal light control and improve road capacity by analysing vehicle flow data [6].

In the future, with the development of large-scale datasets and more powerful computational resources, vehicle image target detection will develop towards higher accuracy, faster speed, and stronger robustness. At the same time, multimodal data fusion (e.g., combining radar, LiDAR and camera data) will also become an important trend to improve detection performance. In this paper, based on the latest version of the YOLOv10 model, we apply the YOLOv10 model to vehicle image detection to explore the effect of target detection.

2. Sources of data sets

The dataset used in this paper is selected from the open source dataset, the dataset contains images of various categories (cars, bicycles, ambulances, trucks, and motorbikes), the images are divided into three parts, the training set, the validation set and the test set, the training set contains a total of 878 images, the validation set contains a total of 250 images, and the test set contains a total of 126 images. Some of the datasets are taken for presentation and the results are shown in Fig. 1.

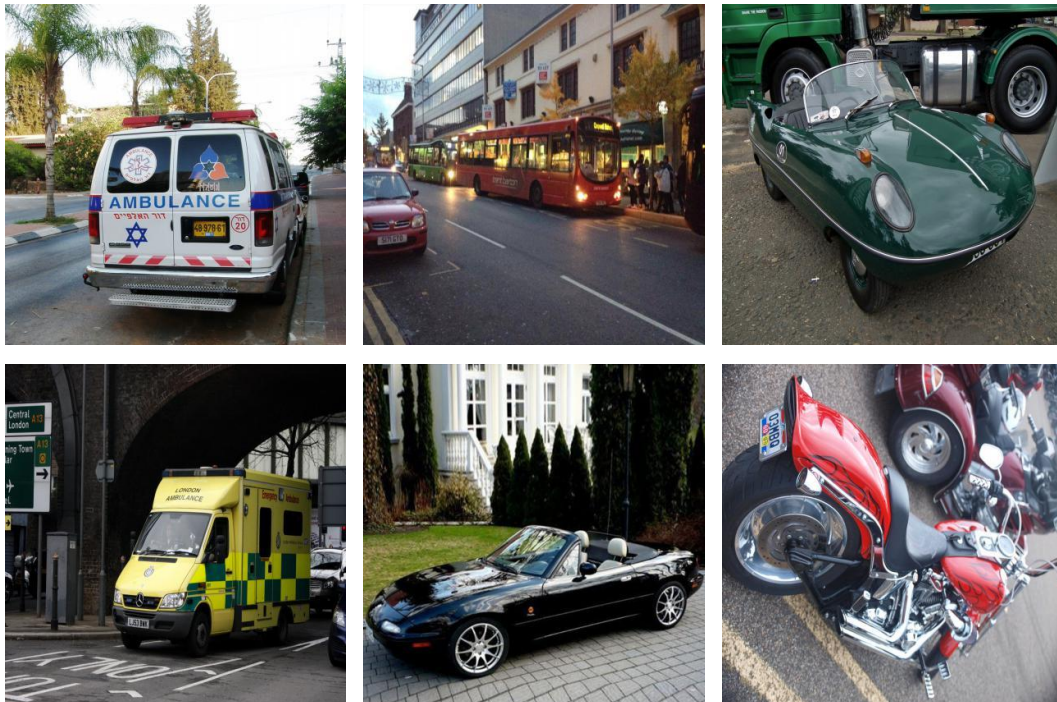


Figure 1. Selected data sets.

3. Method

3.1. Evolution and principles of the YOLO model

YOLO (You Only Look Once) is a real-time object detection system first proposed by Joseph Redmon et al. in 2015. Unlike traditional object detection methods, YOLO treats the object detection task as a regression problem, predicting bounding boxes and class probabilities directly from the image via a single neural network. This approach significantly improves speed and efficiency, making YOLO perform well in real-time applications [7].

The YOLO model divides the input image into an $S \times S$ grid. Each grid is responsible for predicting the object whose centre falls within that grid. When a grid detects an object, it outputs B bounding boxes and their corresponding confidence scores, and C class probabilities [8]. The confidence score reflects the likelihood of the presence of an object within the bounding box, and the extent to which the box overlaps with the true bounding box (typically measured using the IoU metric). In this way, the final score for each bounding box can be obtained by multiplying the confidence level with the class probability.

YOLO uses a Convolutional Neural Network (CNN) as its infrastructure, with initial versions using GoogLeNet or Darkne [9]. The front part of the network is used to extract features while the back part is used to generate predictions. During training, YOLO optimises the model by minimising a loss function that combines position error, confidence error and category error. This end-to-end training approach allows YOLO to learn better feature representations.

Since its first release, YOLO has gone through several iterations. From the original YOLOv1 to the now widely used YOLOv5 and YOLOv7, each version has improved in performance and accuracy. For example, the subsequent versions introduced multi-scale prediction, anchor mechanisms, and improved feature extraction networks to cope with detection tasks in complex scenarios. In addition, the latest version also enhances the detection of small objects and in dense scenes, making the model more comprehensive. In conclusion, YOLO, as an innovative object detection algorithm, has occupied an important position in the field of computer vision due to its unique methodology and evolving development history. The structure of the YOLOv5 model is shown schematically in Fig. 2, and that of the YOLOv8 model is shown schematically in Fig. 3.

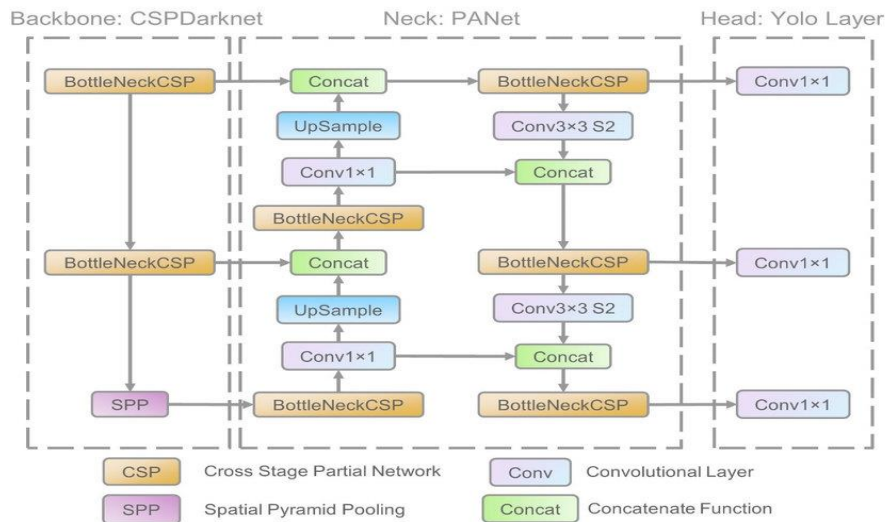


Figure 2. The structure of the YOLOv5 model.

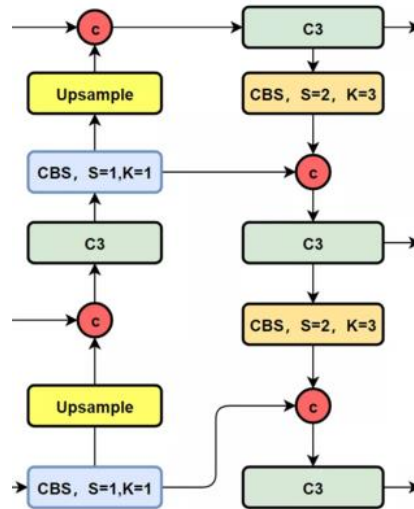


Figure 3. The structure of the YOLOv8 model.

3.2. YOLOv10

YOLOv10 is the latest object detection model in the YOLO family, which inherits and improves on the strengths of the previous versions, aiming to provide higher accuracy and faster inference. YOLOv10 has been made to perform even better in a wide range of application scenarios through a series of innovative techniques and optimisations, especially in real-time detection tasks.

Similar to earlier versions, YOLOv10 still treats object detection as a regression problem, predicting bounding boxes and their class probabilities directly from the image via a single neural network. The input image is divided into $S \times S$ grids, each of which is responsible for detecting objects whose centres fall within that grid. Unlike traditional methods, YOLOv10 employs an anchor point mechanism to improve detection in small objects and dense scenes. Multiple bounding boxes can be predicted for each grid, and each bounding box outputs confidence and category probabilities, enabling effective recognition of multiple objects.

YOLOv10 introduces a number of new network architecture designs, such as the use of deeper feature extraction networks (e.g., CSPNet) and attentional mechanisms to enhance the focus on important features. This improvement allows the model to better capture detailed information in complex scenes. In addition, YOLOv10 employs efficient convolutional modules, such as depth-separable convolution, to reduce computation and speed up inference [10].

During training, YOLOv10 uses a new loss function that not only takes into account the position error, confidence error and category error, but also introduces a specialised optimisation strategy for small objects. This allows the model to learn effectively at multiple scales, thus improving the overall detection performance.

Compared to previous versions, YOLOv10 offers significant improvements in several areas:

1. speed and efficiency: by optimising the network architecture and reducing the number of parameters, YOLOv10 achieves faster inference, enabling it to run with low latency requirements.
2. Accuracy: Newly introduced feature extraction techniques and loss function optimisation enable the model to perform better on a variety of datasets, especially in small object detection.
3. Flexibility: Supports a wide range of input sizes, which can be adjusted according to specific application requirements to achieve the best results.
4. Ease of use: provides a friendly API interface, enabling developers to quickly integrate into a variety of applications, whether embedded devices or cloud services.

4. Experiments and Results

After dividing the data, the YOLOv10 model was trained using the data from the training set, the summary of the training set is shown in Table 1.

Table 1. The summary of the training set.

Class	Images	Instances	P	R	mAP50	m
All	250	454	0.658	0.57	0.606	0.469
Ambulance	250	64	0.751	0.781	0.859	0.746
Bus	250	46	0.691	0.696	0.746	0.616
Car	250	238	0.648	0.403	0.464	0.312
Motorcycle	250	46	0.626	0.522	0.51	0.346
Truck	250	60	0.574	0.449	0.453	0.324

Output the confusion matrix predicted by the training set, which records the prediction results for the five types of vehicles, as shown in Fig. 4.

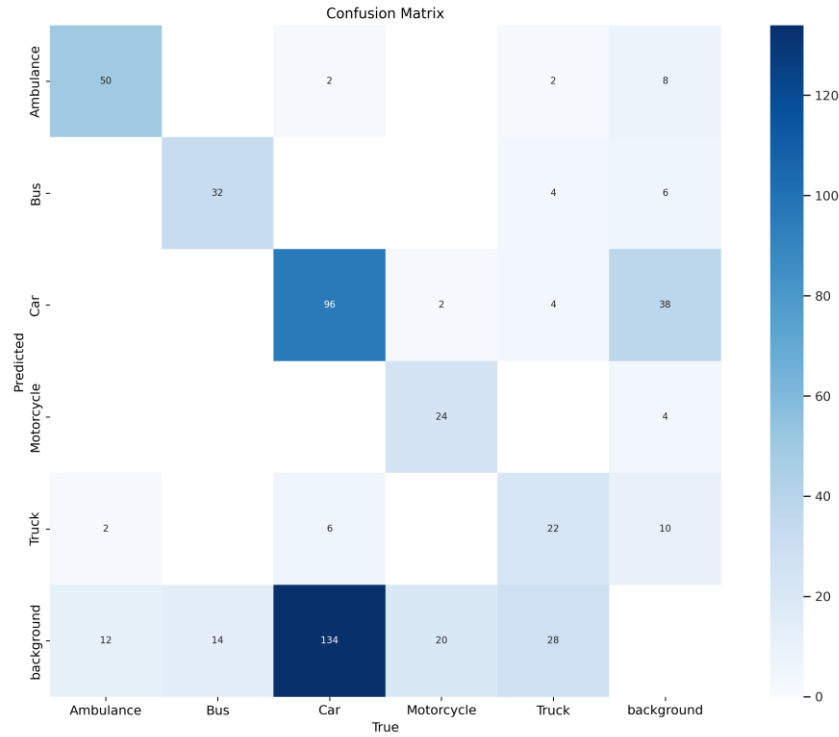


Figure 4. Classification of electrical faults.

The Precision-Recall curve is output, as shown in Fig. 5, and the F1-confidence curve is output, as shown in Fig. 6.

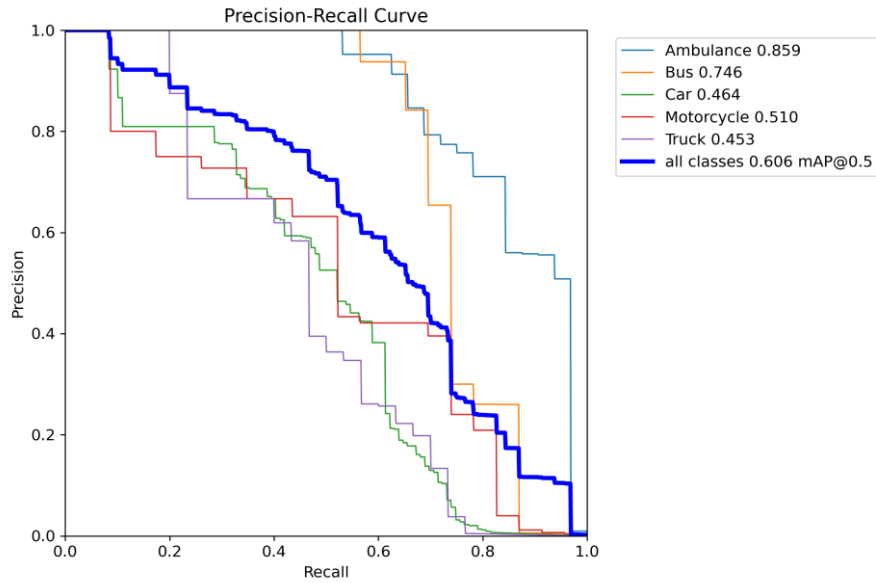


Figure 5. Classification of electrical faults.

From the Precision-Recall curve, it can be seen that the area below the PR curve is larger, indicating that the AUC is closer to 1 and the model has better classification performance.

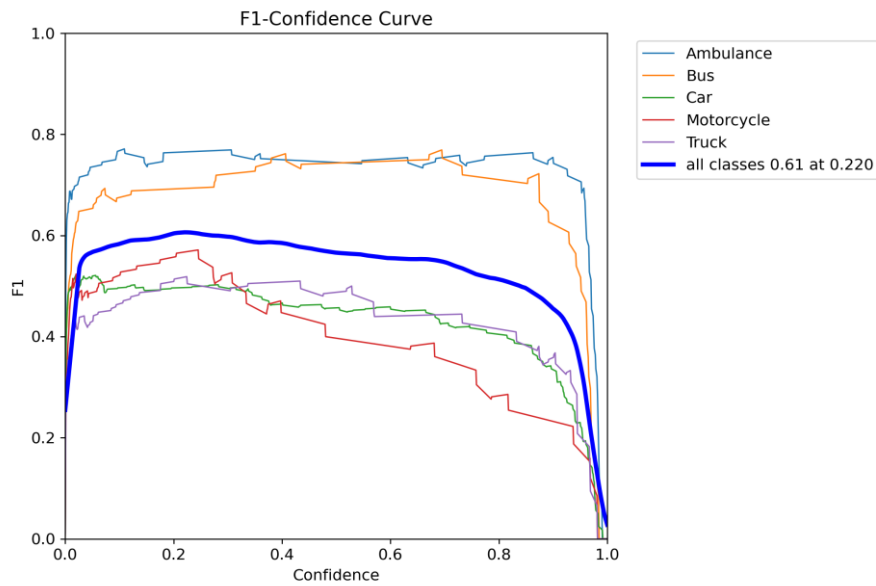


Figure 6. Classification of electrical faults.

From the F1-confidence curve, it can be seen that the model is still able to maintain a better classification performance at higher confidence levels, and secondly, the model is able to maintain a higher F1 score at most confidence levels, and the model has a better reliability.

Finally, the model is tested using the test set, the test is divided into two parts, the first part is target detection of vehicle images, followed by classification of the cars detected by the target to identify the car class to which the vehicle belongs, the results are shown in Figure 7.

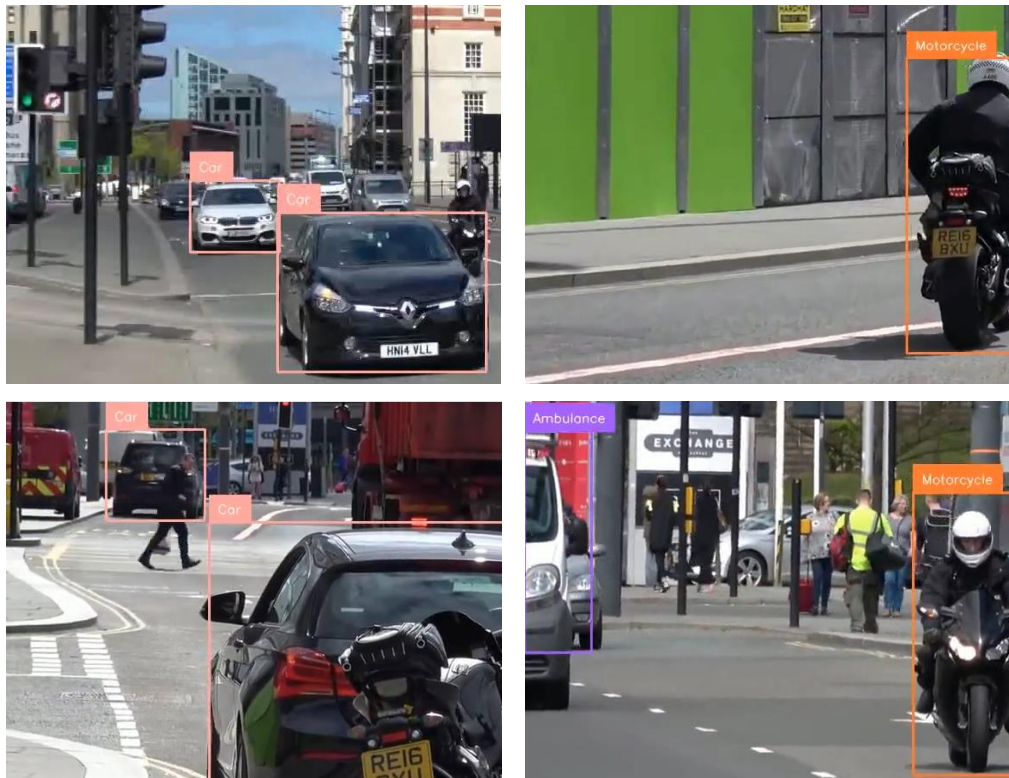


Figure 7. Classification of electrical faults.

As can be seen from the model test results, the YOLOv10 model in this paper can accurately identify the vehicles, and at the same time predict the type of vehicles, and identify the road vehicle conditions in real time, which is a good implementation of the road vehicle target detection and classification.

5. Conclusion

In this paper, based on the latest version of the YOLOv10 model, we deeply discuss the effectiveness of its application in vehicle image detection. Firstly, we reasonably divided the data set into a training set and a test set to ensure that the model can be effectively learnt and evaluated. After training YOLOv10 using the training set, by analysing the Precision-Recall (PR) curve, we found that the area under the PR curve (AUC) is close to 1, which indicates that the model performs well in classification performance. A high AUC value usually implies that the model is able to efficiently distinguish between different categories, thus improving the reliability of target detection.

Further, by analysing the F1-confidence curve, we can see that YOLOv10 still maintains good classification performance at higher confidence levels. This result indicates that the model is still able to maintain a high F1 score even when facing complex environments or blurred images, reflecting its strong robustness and reliability. In addition, the F1 scores remain high at most confidence levels, which further strengthens our confidence in the model's performance.

A comprehensive analysis of the test results confirms that the YOLOv10 model has the ability to accurately identify vehicles and their types. This feature makes it possible to monitor the status of vehicles on the road in real time, providing strong support for traffic management and the development of intelligent transport systems. In particular, YOLOv10 shows a wide range of application potential in application scenarios such as urban traffic management, autonomous driving technology, and intelligent surveillance systems.

In summary, this study shows that the YOLOv10 model performs well in the vehicle target detection and classification task, and its accuracy and real-time performance meet the expected goals. This not only validates the importance of YOLOv10 as an advanced object detection technique, but also lays the

foundation for future development in related fields. In the future, we will continue to explore how to further optimise the model to cope with more complex and changing real-world scenarios, and we also look forward to extending it to more fields and realising a wider range of application values.

References

- [1] Zhao, Chenao, et al. "SatDetX-YOLO: A More Accurate Method for Vehicle Target Detection in Satellite Remote Sensing Imagery." *IEEE Access* (2024).
- [2] Li, Yan, et al. "A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved YOLOV3." *Sensors* 20.17 (2020): 4885.
- [3] Dang, Bo, Danqing, Ma, Shaojie, Li, Zongqing, Qi, Elly, Zhu. "Deep learning-based snore sound analysis for the detection of night-time breathing disorders". *Applied and Computational Engineering* 76. (2024): 109-114.
- [4] Danqing Ma, Meng Wang, Ao Xiang, Zongqing Qi, Qin Yang. "Transformer-Based Classification Outcome Prediction for Multimodal Stroke Treatment." (2024).
- [5] Hedi Qu, , Danqing Ma, Zongqing Qi, Ni Zhu. "Advanced deep-learning-based chip design enabling algorithmic and hardware architecture convergence." *Third International Conference on Algorithms, Microchips, and Network Applications (AMNA 2024)*. SPIE, 2024.
- [6] Zongqing Qi, , Danqing Ma, Jingyu Xu, Ao Xiang, Hedi Qu. "Improved YOLOv5 Based on Attention Mechanism and FasterNet for Foreign Object Detection on Railway and Airway tracks." (2024).
- [7] Tan, Li, et al. "YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm." *Computers & Electrical Engineering* 93 (2021): 107261.
- [8] Wu, Wentong, et al. "Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image." *PloS one* 16.10 (2021): e0259283.
- [9] Xu, He, et al. "Unmanned aerial vehicle perspective small target recognition algorithm based on improved yolov5." *Remote Sensing* 15.14 (2023): 3583.
- [10] Du, Shuangjiang, et al. "Weak and occluded vehicle detection in complex infrared environment based on improved YOLOv4." *IEEE Access* 9 (2021): 25671-25680.