International Journal on Semantic Web and Information Systems
Volume 19 • Issue 1

MC-YOLO-Based Lightweight Detection Method for Nighttime Vehicle Images in a Semantic Web-Based Video Surveillance System

Xiaofeng Wang, School of Electric Power, Civil Engineering, and Architechture, Shanxi University, China
Xiao Hao, College of Physics and Electronic Engineering, Shanxi University, China*
Kun Wang, College of Physics and Electronic Engineering, Shanxi University, China

ABSTRACT

Semantic web-based video surveillance systems can provide strong decision-making support for managers, and they have high requirements for real-time and precision of vehicle detection models in complex night scenes. To address this issue, a lightweight nighttime vehicle detection method (MC-YOLO) integrating MobileNetV2 and YOLOV3 is proposed. Firstly, in the preprocessing stage, image enhancement is performed on nighttime images to facilitate model feature extraction. Then, the lightweight network MobileNetV2 is used to extract feature by replacing the backbone network DarkNet53 in YOLOv3, thus accelerating the speed of target detection. Finally, after the convolution operation of the backbone network, a convolution block attention module is added to enhance the important feature information and suppress the secondary features, thereby improving the detection precision. The experimental results on the BDD100K dataset show that the proposed MC-YOLO model has a precision of up to 92.75%, which is superior to several other advanced comparative models.

KEYWORDS

Convolution Block Attention Module, Image Enhancement, Lightweight Network, Mobilenetv2, Nighttime Vehicle Detection, Semantic Web-Based Video Surveillance, Yolov3

1. INTRODUCTION

With the technological advancements in 5G/6G basic networks (Kumar et al., 2021), fog computing (Al-Qerem et al., 2020), cloud computing (Peñalvo et al., 2022; Vijayakumar et al., 2022), big data (Stergiou et al., 2021), social network (Almomani et al., 2022; Zhang et al., 2023; Arowolo et al., 2023), information security (Gaurav et al., 2023; Alhalabi et al., 2023), Internet of Things (IoT) (Memos et al., 2018), Internet of Vehicles (IoV) (Prathiba et al., 2021; Sharma et al., 2022), smart electric vehicles (Akl et al. 2021), the implementation and utilization of semantic Web-based video surveillance systems have become widespread, resulting in massive video and image data. Vehicle detection technology in semantic Web-based video surveillance systems can provide strong decision

DOI: 10.4018/IJSWIS.330752 *Corresponding Author

support for managers. Compared to other detection technologies (Zhao et al., 2022; Ding et al., 2022; Zhang et al., 2021), semantic Web-based video surveillance has several advantages, including convenient installation and maintenance, no need to interrupt traffic, low cost, large amounts of analyzable information, and no impact on road life. Currently, this technology is experiencing rapid development (Tsai & Chen, 2021; Mo et al., 2022; Zhang et al., 2023).

Vehicle detection is mainly divided into daytime and nighttime vehicle detection (Cui et al., 2022). Due to poor lighting conditions and more complex scenes at night, less information about vehicles and the environment can be obtained. Therefore, the difficulty of nighttime vehicle detection is greater than that of daytime detection (Shao et al., 2021; Zhang et al., 2022; Bell et al., 2022). Numerous vehicle detection algorithms applied in normal weather conditions have limited performance at night. Therefore, there is an urgent need for a high-performance night vehicle detection algorithm (Yang, 2020; Alam et al., 2022; Li et al., 2022).

The traditional vehicle detection algorithms are based on a single light feature for detection, which is difficult to accurately detect with limited feature information (Dai, 2019; Zaarane et al., 2019). The existing deep learning-based detection algorithms are based on multiple features for detection. Hence, these algorithms are significantly superior to traditional detection algorithms in terms of detection accuracy and real-time performance (Alcantarilla et al., 2011; Parvin et al., 2021). However, due to the significant impact of lighting, the potential for misjudgment remains high, and achieving the necessary levels of detection precision and real-time performance to meet the requirements for fast and accurate nighttime vehicles continues to pose challenges (Shao et al., 2021; Huang et al., 2021).

Therefore, this paper proposes a lightweight nighttime vehicle detection model (called MC-YOLO) that integrates MobileNetV2 (Gupta et al., 2022) and YOLOv3 (Li et al., 2021). The innovations of the proposed model are as follows:

1. Nighttime images often have issues, such as low lighting and glare. Moreover, it is difficult to obtain sufficient information about vehicles and the environment, which significantly affects the feature extraction process of the backbone network. Therefore, to enhance the image quality, automatic white balance is used in the preprocessing stage to reduce interference from streetlight color and vehicle glare. In addition, the mosaic algorithm is used for enhancement and sample expansion of nighttime images, which will facilitate the backbone network to accurately extract vehicle information from nighttime images.
2. Many existing vehicle detection models use complex backbone networks with a large number of parameters, leading to lower model detection efficiency and affecting its practicality. Therefore, the backbone network DarkNet53 in the YOLOv3 algorithm is replaced with the lightweight network MobileNetV2 to extract features and accelerate the speed of target detection.
3. Numerous existing vehicle detection models face challenges in effectively extracting features across various channels or spaces during the feature extraction stage, leading to the problem of potentially disregarding important image features. Therefore, after the convolution operation of the backbone network, a convolution block attention module (CBAM) (Li et al., 2022) is added to enhance the important feature information and suppress the secondary features, thereby improving the detection precision.

The experimental results on the BDD100K dataset demonstrate that the proposed MC-YOLO model has a precision of up to 92.75%. This performance surpasses that of the original YOLOv3 as well as several other state-of-the-art comparative models.

2. RELATED WORKS

Driving at night is an inevitable part of people's daily lives. Therefore, a detection method that can provide early warning information for drivers and is suitable for the night driving environment is particularly important (Arora et al., 2022; Zhang et al., 2021). Vehicle detection is a key aspect of intelligent driving and awareness of the surrounding environment. Currently, numerous vehicle detection algorithms based on different detection principles are available (Cai et al., 2021).

The traditional vehicle detection methods include ultrasonic method, radar detection and infrared method. Lian et al. (2021) developed an external perception model based on a cluster of sensors, incorporating technologies like ultrasonic radar and laser radar to perceive the external driving environment. Wang et al. (2022) extracted the Region of Interest (ROI) of vehicles from infrared images and combined pseudo visual search, directional gradient histogram, and local binary mode feature methods to achieve vehicle detection. Huang et al. (2021) proposed a deep network combination framework (DAP-BCL) integrating directional attention pool (DAP) and Bayesian corner location to improve the detection performance of vehicles at night. However, the practicality of DAP-BCL is low and the original image is not pre-processed. The lighting characteristics of nighttime vehicle images can easily hinder the improvement of detection performance. Wang and Zhang (2022) used a clustering algorithm to analyze the image dataset uploaded by laser radar and employed the support vector machine (SVM) algorithm to detect vehicles. The ultrasonic method and LiDAR-based method need to introduce visual information to determine whether the target is a vehicle. However, the detection range is easily limited due to electromagnetic interference (Matsui & Oikawa, 2021). The infrared method has a high cost and weak noise suppression ability due to the interference of the vehicle body's heat source.

Li et al. (2022) proposed an integrated framework based on improved bioinspired multi-exposure fusion (IBIMEF) and improved the performance of nighttime vehicle detection. However, this model has a weak ability to identify important and secondary features, which hinders the improvement of precision.

In the daytime, the light is good, and the vehicle features are clear and easy to extract. Hence, the vehicle detection algorithms in the daytime scene are quite mature. However, research on vehicle detection algorithms in night scenes remains limited, and fewer detection models consider both accuracy and real-time. The vehicle image taken at nighttime has low definition due to the poor light and the interference of streetlamps, making it difficult to extract features, significantly reducing the detection performance (Ravindran et al., 2021).

Vehicle recognition algorithms developed in most daytime scenarios meet the detection requirements thanks to advances in deep learning. However, when the scene is nighttime, the detection performance of these algorithms is poor (Al-refai & Rawashdeh, 2020). Al-refai and Al-refai (2020) used Draknet-53 convolutional neural network (CNN) to identify and analyze street images, including pedestrians, vehicles, trucks, and cyclists. Mu et al. (2020) realized the detection and analysis of running vehicles based on the YOLOV4 network to reduce the risk of vehicle collision. Yi et al. (2021) adopted the multiple fusion YOLO network model to extract features of shallow convolution layer and enhance the ability of the model to detect targets.

However, it should be noted that there are many noise slots in the nighttime images. If the detection network model has a low ability to extract image features, it is difficult to extract and identify effective datasets. All the above-mentioned methods consider the difference of various feature information, which is insufficient to achieve targeted analysis of night images. Therefore, this paper integrates the multi-layer deep neural network model and optimizes the network from two aspects of network feature extraction and image analysis.

Table 1. Comparison of advantages and disadvantages

Reference	Advantage	Disadvantage
Lian et al., 2021	The driving environment of the vehicle is accurately perceived through sensor fusion technology.	As the sensing performance of the sensor changes, it is necessary to iteratively update the fusion of more advanced algorithms.
Wang et al., 2022	The surrounding environment of the vehicle can be accurately extracted from the infrared image.	It will be disturbed by the external natural environment, thus affecting the accuracy of detection.
Huang et al., 2021	It can simultaneously extract features from both object and pixel streams.	The practicality of this framework is low, and the original image is not preprocessed
Wang & Zhang, 2022	The clustering algorithm is used to accurately detect the vehicle through radar.	It will be affected by the external natural environment and needs to be continuously updated to improve the detection accuracy.
Li et al., 2022	It obtains accurate detection of images and objects by combining machine learning and computer vision algorithms.	It will be easily disturbed by the external natural environment and has high requirements for equipment.
Al-refai & Al-refai, 2020	This method preprocesses the image through image enhancement (IME) technology and enhances the model's feature extraction ability by combining multi-scale features with visual features.	It has a weak ability to identify important and secondary features.
Yi et al., 2021	The target object is accurately detected by vehicle thermal imaging technology.	It requires strong algorithm support, so the requirements for the algorithm are high and need to be updated iteratively.

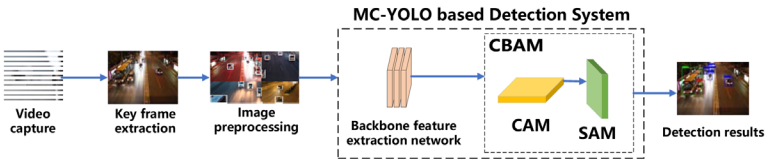
3. PROPOSED NIGHTTIME VEHICLE DETECTION METHOD MC-YOLO

The deep neural network accurately extracts effective vehicle features from the acquired nighttime vehicle image dataset and improves the nighttime vehicle detection performance. The pre-trained detector MC-YOLO is deployed in advance in a semantic web-based video surveillance system. When a large amount of nighttime vehicle videos are collected by video surveillance, keyframe extraction technology is employed to extract keyframes and nighttime images are pre-processed to enhance vehicle features. Then, vehicle targets are detected through the proposed detector MC-YOLO. Figure 1 illustrates the flow chart of nighttime vehicle image detection based on the proposed MC-YOLO.

To extract features, the proposed approach uses the backbone feature extraction network. The combination of MobileNetV2 and CBAM enables the network to enhance both channel and spatial feature information in the residual structure feature map, focus more easily on the targeted part, and improve the network's feature extraction capability.

In the enhanced feature extraction network, distinct operations are sequentially executed including max-pooling, convolution, up and down-sampling, and feature fusion. The structure of the MC-YOLO model is used to realize image analysis of multi-layer semantic features and multi-scale receptive fields.

Figure 1. Flow chart of vehicle image detection at night based on the proposed MC-YOLO



3.1 Image Preprocessing

The nighttime vehicle image has low definition, leading to challenges in extracting effective features from it. Therefore, the image quality of nighttime vehicle images is enhanced through image preprocessing. It can be seen from the existing datasets that the interference of streetlights at night makes the image color yellow, affecting the feature definition and the subsequent feature extraction process. Therefore, to reduce the interference of streetlights and enhance the image quality, an automatic white balance algorithm can be used to eliminate the impact of light sources on the image, making the features in the image more obvious.

To reduce the interference of streetlight color and vehicle glare on the backbone network feature extraction process, the Gray World Algorithm (GWA) is used to automatically balance the nighttime vehicle dataset, aligning the image color with the color of real-world objects in various light source environments and improve image quality.

The GWA first calculates the average values $\bar{R}$, $\bar{G}$, $\bar{B}$ of the three channels of the image, as shown in Formula (1):

$$aver = \frac{\bar{R} + \bar{G} + \bar{B}}{3} \tag{1}$$

The gain coefficients of channels R, G and B are calculated as shown in Formula (2):

$$\begin{cases} c_r = \dfrac{aver}{\bar{R}} \\ c_g = \dfrac{aver}{\bar{G}} \\ c_b = \dfrac{aver}{\bar{B}} \end{cases} \tag{2}$$

Finally, according to the diagonal model, the R, G and B channel components of each pixel in the image are corrected as shown in Formula (3):

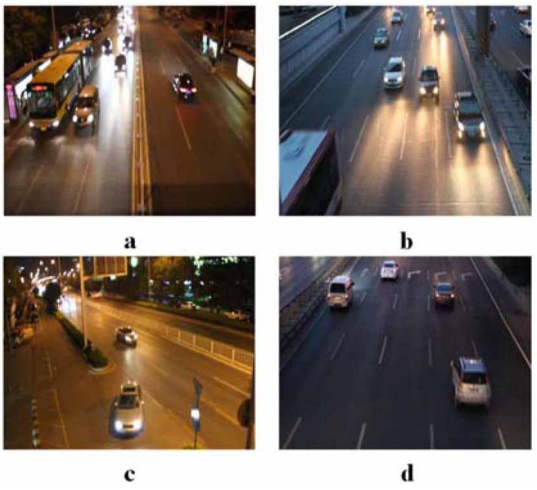
$$\begin{cases} \lambda(R') = \lambda(R) * c_r \\ \lambda(G') = \lambda(G) * c_g \\ \lambda(B') = \lambda(B) * c_b \end{cases} \tag{3}$$

3.2 Image Enhancement

Mosaic data enhancement will select four images and splice them. Each image has a corresponding box. After splicing, a new image and its corresponding box will be obtained, and then the new image will be input into the neural network for learning.

First, four images are randomly read from the dataset each time. Figures 2 (a), (b), (c), and (d) show the randomly selected first, second, third, and fourth images, respectively. Second, the four images are flipped (flip the original image left and right), zoomed (zoom the size of the original image), and their color gamut is altered (change the saturation, hue, and brightness of the original image). Finally, the images and boxes are combined. After the placement of the four images, the matrix method is employed to capture and retain the fixed areas in the four images, and then the images are spliced together to form a new image, which contains a series of contents such as boxes. The four

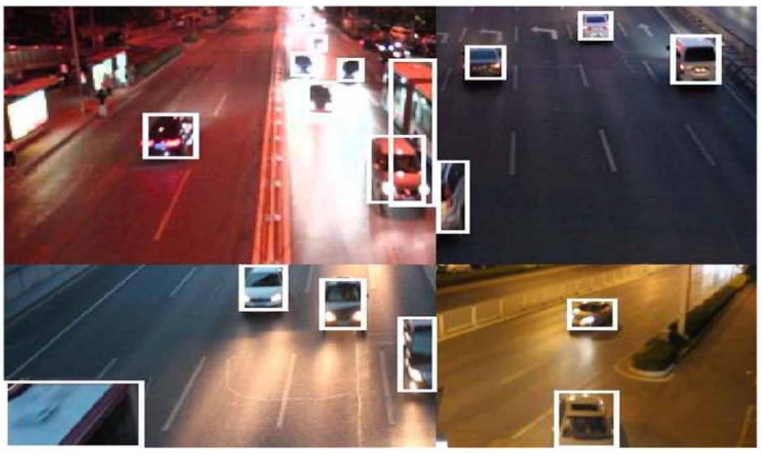
Figure 2. Original images



spliced images are shown in Figure 3. When the four images are spliced, they have clear edges. The split lines are horizontal and vertical. At the same time, the introduction of random scaling makes many small target vehicles appear in the spliced images, accentuating the distinctive features of these smaller target vehicles in the images.

The Mosaic algorithm is used for data augmentation and expanding the number of training samples to increase data diversity, which is beneficial for the detection model to distinguish between the background and foreground of the target object more easily. The Mosaic algorithm fuses four images with different semantic information to help the detection model extract better vehicle features than the conventional ones, thereby enhancing the robustness of the detection model. When the number of training samples is increased, the batch size in the batch normalization layer can also be increased. This adjustment results in the mean and variance of each feature layer approaching the mean and variance of the entire dataset, improving the detection performance. Moreover, it is often difficult for the model to accurately learn small targets due to size or clarity issues. However, the utilization of the

Figure 3. The enhanced image using the mosaic algorithm



Mosaic algorithm magnifies small targets in nighttime images through random scaling operations. This amplification proves advantageous for the detection model, allowing it to extract more features of small vehicles and significantly reduce the missed detection rate of small vehicles, thus improving the overall detection performance of the model.

3.3 MobileNetV2

MobileNet is a lightweight network, which mainly relies on depth-wise separable convolution to achieve lightweight. To extract feature maps, separable convolution uses a combination of depth-wise (DW) and pointwise (PW) methods. First, the feature image is convolved by each channel. Then, each resultant vector, obtained from channel-wise convolution, is subjected to further convolved using a 1x1 convolution kernel. The use of depth-wise separable convolution can significantly reduce model size and computational complexity. The feature map of $S_F \times S_F \times N$ can be obtained by convolution of the input feature map with the size of $S_F \times S_F \times U$.

The standard convolution uses N convolutions with a size of $S_E \times S_E \times N$, and its floating point operations per second (FLOPs) da_1 are expressed as:

$$da_1 = S_E \times S_E \times N \times S_F \times S_F \times U \tag{4}$$

Depth-wise separable convolution involves a two-step process. First, each channel undergoes convolution with a kernel size of $S_E \times S_E \times 1$ and then a pointwise convolution with a kernel size of $1 \times 1 \times N$. The FLOPs of depth-wise separable convolution are defined as:

$$da_2 = S_E \times S_E \times N \times S_F \times S_F + N \times N \times S_F \times S_F \tag{5}$$

Compared with ordinary standard convolution, the FLOPs of depth-wise separable convolution are reduced to the original:

$$\frac{S_E \times S_E \times N \times S_F \times S_F + N \times N \times S_F \times S_F}{S_E \times S_E \times N \times S_F \times S_F \times U} = \frac{1}{N} + \frac{1}{SE^2}$$

For example, using the same A * A convolution kernel, the computation of depth-wise separable convolution is 8-9 times less than that of standard convolution. However, there is an empty convolution kernel in the training process of depth-wise separable convolution, which will cause resource wastage. Therefore, MobileNetV2 is improved based on MobileNetV1. The overall structure of the MobileNetV2 is shown in Table 2, where T denotes the convolution layer's dilation rate, while C, N, and S represent the numbers of output channels, repeats, and steps, respectively.

3.4 CBAM

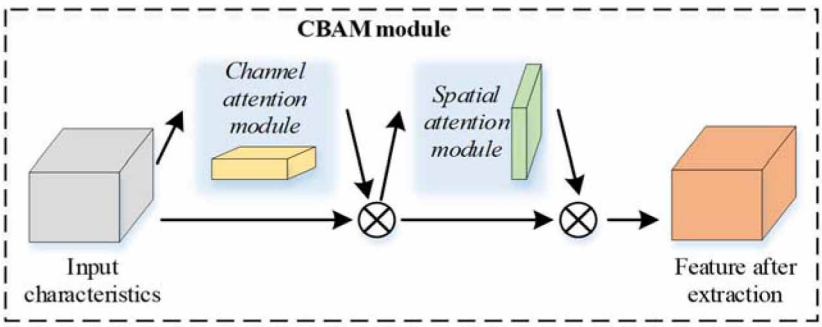
As an attention mechanism, the CBAM can be used to adaptively learn or extract weight distributions from features. These learned distributions are subsequently applied to the original features to optimize the distribution of the original features. This adaptive process enhances effective features while concurrently suppressing invalid features. The structure of CBAM is shown in Figure 4 (Huang et al., 2021).

The CBAM structure includes channel attention and spatial attention modules. Each channel of the channel attention module (CAM) can be regarded as an independent feature detector. Assuming

Table 2. Overall structure of MobileNetV2

Input	Operator	T	C	N	S
$268^2 \times 3$	conv2	-	32	1	1
$100^2 \times 32$	bottleneck	2	16	1	2
$124^2 \times 16$	bottleneck	6	24	2	2
$68^2 \times 24$	bottleneck	6	32	2	2
$46^2 \times 32$	bottleneck	6	64	3	1
$43^2 \times 64$	bottleneck	6	96	4	2
$24^2 \times 96$	bottleneck	6	160	3	1
$7^2 \times 160$	bottleneck	6	320	2	1
$7^2 \times 320$	bottleneck	-	640	1	-
$1 \times 1 \times 640$	bottleneck	-	-	-	-

Figure 4. CBAM module



that when the input feature is F and the size is $H \times W \times C$, two 1x1xC channel descriptions are obtained after the average-pooling and max-pooling of the space. Then the channel description is sent to a two-layer neural network. The numbers of neurons in the first and second layers of this two-layer neural network are C/r and C, respectively. Secondly, the weights are obtained by summing the features. Finally, the output features can be obtained by multiplying the original features with the obtained weights. The characteristics and calculation process of the spatial attention module (SAM) are similar to the CAM. Each channel can also be used as a feature detector. In the convolution block, the CBAM module can adaptively improve the intermediate feature map. Channel attention uses max-pooling and average-pooling to compress the feature map $J \in R^{C \times H \times W}$ on the spatial dimension to obtain two spatial background descriptions, $J^C_{\max} \in R^{C \times 1 \times 1}$ and $J^C_{avg} \in R^{C \times 1 \times 1}$, respectively. The

two feature maps outputs from the multi-layer perceptron (MLP) network are added and normalized using the Sigmoid function. The final channel attention feature map is $K_C \in R^{1\times1\times1}$, and its calculation is shown in Formula (6):

$$\begin{aligned} K_C(J) &= \delta(MLP(AvgPool(J)) + MLP(MaxPool(J))) \\ &= \delta(S_1(S_0(J^C_{avg})) + S_1(S_0(J^C_{\max}))) \end{aligned} \tag{6}$$

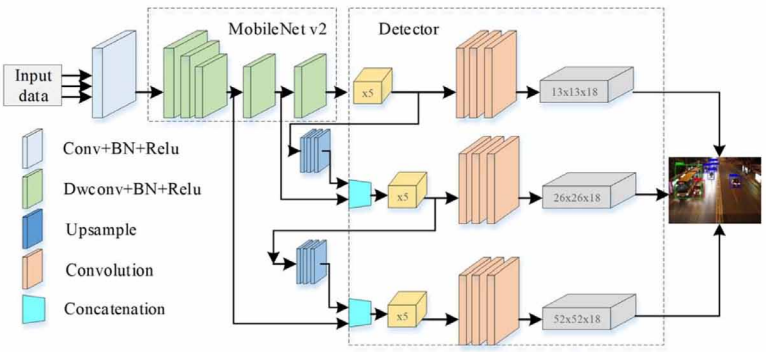
where δ represents the Sigmoid function, $S_0 \in R^{\frac{c}{c^*r}}$ and $S_1 \in R^{\frac{c^*c}{r}}$ represent the weights of the two layers of MLP, and r represents the compression ratio of the intermediate channel.

3.5 MC-YOLO

With the advancement of deep learning and the quest for a higher detection rate, a deeper and larger deep neural network is utilized by many traditional methods to increase the detection capabilities (Shang et al., 2021). However, the growing size of models also implies escalating demands on hardware resources, making it difficult to deploy them on hardware facilities with limited resources. Therefore, a more effective and rapid vehicle detection model, MC-YOLO, is built in this paper. Figure 5 shows the network structure of MC-YOLO.

Model compression includes two methods: shallow compression and deep compression. Shallow compression aims to reduce parameters and model hierarchy without changing the original network structure to improve the efficiency of detecting models. Common methods for shallow compression include model pruning and knowledge distillation. Model pruning can effectively compress network parameters. However, the process of iteratively testing thresholds is unstable, requiring substantial computational power to achieve ideal detection results. Knowledge distillation is suitable for learning small models. However, manually designing network structures can lead to significant subjectivity in the results, making the training effect unstable. Deep compression usually achieves better detection results than shallow compression. Deep compression usually requires modifications to the convolutional kernel and network hierarchy. Common deep compression methods include model quantization and lightweight network structure design. The model quantization method can achieve a significant reduction in model size while ensuring minimal accuracy loss. However, this approach presents significant challenges in terms of implementation, often leading to unstable detection accuracy and poor universality. These issues hinder the feasibility of deploying and transferring the model quantization approach onto embedded devices. Conversely, lightweight networks usually simplify the network by changing its structure or using efficient computing methods, making the

Figure 5. The network structure of MC-YOLO



network model compact while ensuring minimal accuracy loss. Lightweight networks are suitable for the transplantation and deployment of embedded devices.

Compared to traditional target detectors, the YOLOv3 network, constructed based on DarkNet-53, can significantly reduce the parameters and computational complexity of the detection model. The YOLOv3 effectively accelerates the operational speed of the detection model while maintaining improved performance. However, the network layers of DarkNet-53 are still very deep, including 53 convolutional layers, resulting in a large number of model parameters and computational complexity. Compared to DarkNet-53, the MobileNetV2 network has a smaller number of parameters. This reduction contributes to a reduction in the computational complexity of the detection model, improving the inference speed while maintaining detection performance. The MobileNetV2 network replaces the traditional standard convolutions with deep separable convolutions. Thus, it has strong feature extraction capabilities, especially for multi-scale object detection tasks. Therefore, the YOLOv3 is improved by replacing its backbone network DarkNet-53 with MobileNetV2.

The backbone network structure of MC-YOLO is shown in Table 3, where T indicates the convolution layer's dilation rate, while C, N, and S represent the numbers of output channels, repeats, and steps, respectively. The bottleneck in Table 3 is composed of 1×1 dilation layer, 2×2 depth-wise separable convolution, and 1×1 projection layer.

4. EXPERIMENT ANALYSIS

Vehicle target detection based on deep learning requires the storage and operation of large quantities of data. The experimental environment is listed in Table 4. In the training process, the Stochastic Gradient Descent (SGD) optimizer was used, the initial learning rate was 0.001, which was set to 0.0001 after 25, 000 iterations, and the batch size was 2.

4.1 Datasets

The BDD 100K dataset was released by the University of California, Berkeley in 2018 containing 100,000 videos. In BDD 100K, the video recording frame rate is 30FPS, the average video

Table 3. MC-YOLO backbone structure

Input	Operator	T	C	N	S	Down-Sample Rate
$422^2 \times 3$	conv2	-	64	1	1	1
$206^2 \times 16$	bottleneck	2	32	1	2	2
$198^2 \times 32$	bottleneck	6	16	2	2	2
$168^2 \times 64$	bottleneck	6	32	2	2	4
$368^2 \times 16$	bottleneck	6	64	3	1	8
$98^2 \times 128$	bottleneck	6	128	4	2	16
$368^2 \times 8$	bottleneck	6	256	3	1	16
$168^2 \times 64$	bottleneck	6	512	2	1	32

Table 4. Experimental environment

Project	Parameter
CPU processor	Intel Core i7-1265U
GPU	Tesla V100
Memory	32G
Operating system	Windows 10
Development language	Python 3.6
Image processing library	OpenCV 4.2.0
CUDA Version	10.2
CUDNN Version	7.4.1

duration is 40s, the resolution is 720p, and the GPS/IMU track information is included. The BDD 100K dataset consists of 100,000 images and corresponding annotations, which are used to label the 10th-second image of each video. The night vehicle images selected from the BDD 100K dataset were used as the experimental dataset. Figure 6 shows some samples from the experimental dataset.

4.2 Evaluation Metric

The widely recognized Average Precision (AP) and Intersection over Union (IoU) were used to measure performance. IoU represents the intersection ratio between the areas covered by the real and the predicted bounding boxes. IoU is calculated as:

$$IoU = \frac{A_{pre} \cap A_{gt}}{A_{pre} \cup A_{gt}} \tag{7}$$

The Precision and Recall are shown in Equations (8) and (9), respectively:

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

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Figure 6. Samples of night vehicle images from the BDD 100K dataset

where TP is the correct forecast of the number of positive samples, FP is the number of negative samples predicted to be true, and FN is the number of missed inspections (the number of positive samples predicted to be false).

4.3 Model Training

To verify the feasibility of the proposed MC-YOLO, 70% of the samples were used as the training dataset to build the detection model. Figure 7 shows the training loss and IoU curves of the proposed MC-YOLO model after 25,000 iterations.

In Figure 7, after 10,000 iterations, the value of the loss function of the model remains below 0.05, while the IoU maintains a better value after 12,500 iterations. This pattern of results underscores that the proposed method's accuracy and recognition rate produced satisfactory results in the vehicle target detection test conducted on remote sensing images.

To compare the proposed MC-YOLO model with YOLOv3, Figure 8 shows the relationship curve between Precision and Recall, and Figure 9 shows the visualization results of vehicle detection for the two models.

It can be observed from Figures 8 and 9 that the proposed MC-YOLO achieves better performance results than the original YOLOv3. Analysis of the reasons shows that automatic white balance can improve image quality and reduce the impact of lighting and glare on the feature extraction process. After the introduction of the Mosaic algorithm, the semantic information of images is enriched and small targets are easier to extract. After the introduction of CBAM, important features are highlighted and secondary features are suppressed. Therefore, considering these advantages, the feature extraction ability of the proposed MC-YOLO has been significantly improved compared with the original YOLOv3.

4.4 Comparison and Analysis

To further verify that the proposed MC-YOLO has better image analysis and identification performance, it is compared with DAP-BCL (Huang et al., 2021), IBIMEF (Li et al., 2022), and several benchmark models: YOLOV3 (Li et al., 2021), SSD300 (Park et al., 2020), and Fast RCNN (Lyu et al., 2021). All methods were run in the same scene. Table 5 and Figure 10 compare the night vehicle detection results of the different models.

Table 5 shows that the Recall of the proposed MC-YOLO is 91.04%, which is 1.5% and 3.46% higher than that of DAP-BCL and IBIMEF, respectively. Moreover, the Precision of the proposed MC-YOLO is 92.75%, which is 2.71% higher than that of DAP-BCL.

Figure 7. Loss and IoU curves in the training process

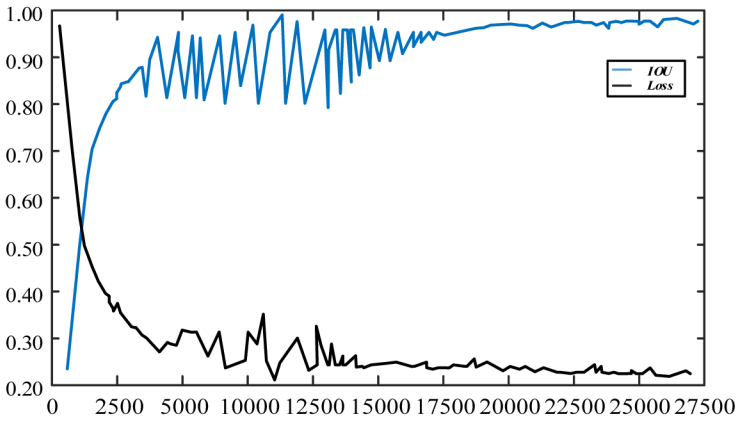


Figure 8. Precision curves of YOLOv3 and MC-YOLO with changes in recall

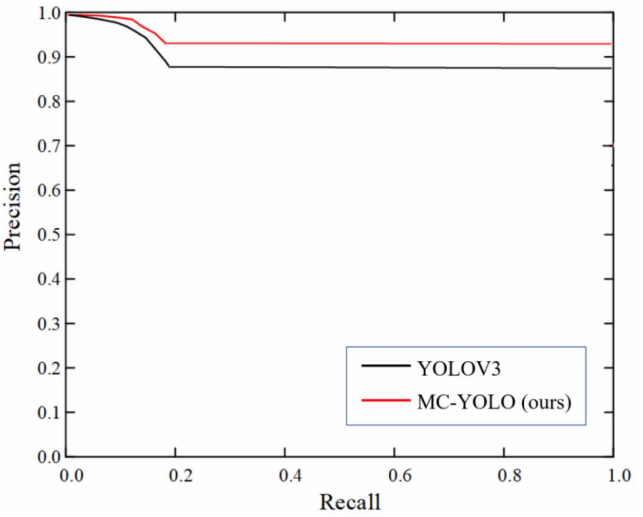


Figure 9. The visualization results of vehicle detection of YOLOv3 and MC-YOLO

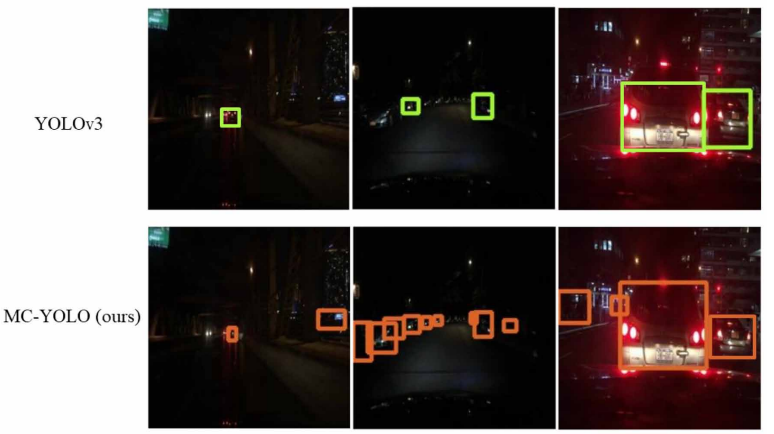
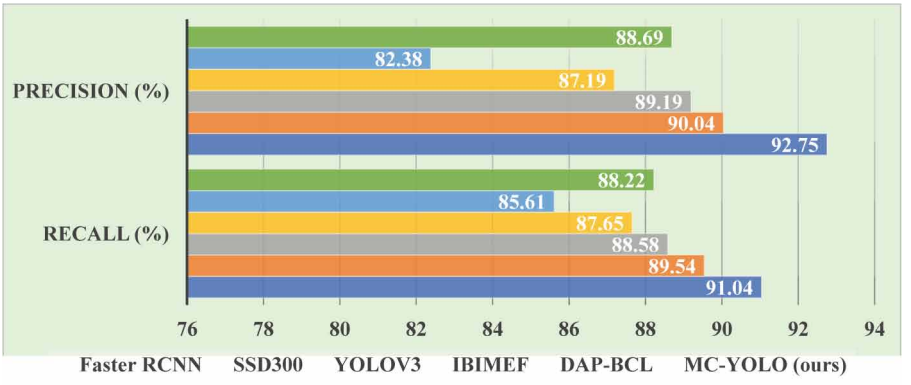


Table 5. Night vehicle detection results of different models

Model	Recall (%)	Precision (%)	FPS (Frame/s)
DAP-BCL	89.54	90.04	24
IBIMEF	88.58	89.19	26
YOLOV3	87.65	87.19	36
SSD300	85.61	82.38	38
Faster RCNN	88.22	88.69	28
MC-YOLO (ours)	91.04	92.75	34

Figure 10. Night vehicle detection histogram of different models



Compared with the DAP-BCL, the white balance and Mosaic algorithm in the proposed MC-YOLO are used to pre-process nighttime images, which will be more helpful for feature extraction of the backbone network. Compared with the IBIMEF, the CBAM in the proposed MC-YOLO can effectively enhance important features and suppress secondary features, directly improving the feature extraction ability of the backbone network.

Furthermore, compared to benchmark models, such as YOLOv3, SSD, and Fast RCNN, the proposed MC-YOLO model has a slightly higher detection time. However, this increased detection time is still significantly lower than that of the DAP-BCL and IBIMEF models. Although the proposed MC-YOLO model replaces the backbone network in YOLOv3 with the lightweight network MobileNetV2, the addition of the CBAM module increases the time consumption to a certain extent. Fortunately, the introduction of MobileNetV2 in YOLOv3 still holds a significant advantage in detection time over DAP-BCL and IBIMEF. In terms of balancing detection precision and efficiency, the proposed MC-YOLO is significantly superior to other comparative models.

4.5 Ablation Experiment

The ablation experiment was conducted on the proposed MC-YOLO model and the results are listed in Table 6. The comparative models include YOLOv3, M1 (YOLOv3+ MobileNetV2), M2 (IME+ YOLOv3+ MobileNetV2), and M3 (YOLOv3+ MobileNetV2+CBAM).

Table 6 shows that after replacing the backbone network DarkNet53 with MobileNetV2, the detection precision of the M1 model only lost 0.14% compared with the original YOLOv3, while the detection efficiency improved by 6 frames/s. The reason for this is that the introduction of the lightweight network MobileNetV2 simplified the model structure and reduced the number of parameters, significantly improving the detection efficiency. With the introduction of image

Table 6. Ablation experimental results of the proposed MC-YOLO

Models	Backbone	Recall (%)	Precision (%)	FPS (Frames/s)
YOLOv3	DarkNet53	87.65	87.19	36
M1	MobileNetV2	86.92	87.05	42
M2	MobileNetV2	87.76	88.30	40
M2	MobileNetV2	89.82	90.76	37
MC-YOLO (ours)	MobileNetV2	91.04	92.75	34

enhancement and CBAM modules, the detection precision of the model gradually increased to the highest value of 92.75%, while the detection efficiency gradually decreased to 34 frames/s, only 2 frames/s lower than the original YOLOv3. This indicates that the proposed MC-YOLO can effectively improve the detection precision of YOLOv3 while maintaining efficiency.

5. CONCLUSION

In the web video surveillance system, a novel lightweight nighttime vehicle detection method MC-YOLO is proposed to fulfil the high-precision and real-time requirements of nighttime vehicle detection. The proposed MC-YOLO network is optimized based on two aspects: image feature acquisition and image analysis. The CBAM module is introduced into the model to significantly enhance the ability of the network model to acquire image features. Moreover, the proposed MC-YOLO network also enhances the ability of the model to capture and analyse image features. Simulation experiments were conducted on the BDD 100K dataset. The experimental results demonstrate and validate that the proposed method can efficiently extract and recognize element information in a complicated image dataset as well as extract vehicle information. The comparison with the existing algorithms indicates that the vehicle detection model constructed in this paper can fulfil the requirements of high detection accuracy. After improving the precision of nighttime vehicle detection, the semantic web-based video systems can achieve accurate monitoring of traffic congestion, control traffic lights, precise vehicle identification, and counting, as well as the detection of vehicle offsets. These advancements are instrumental in assisting traffic management personnel with accurate judgments and informed decision-making. This has widespread significance for the development and progress of smart parks, smart cities, and smart transportation.

However, the proposed MC-YOLO network has some limitations. For example, only the BDD 100K dataset was used to evaluate the proposed MC-YOLO model. Therefore, the proposed MC-YOLO model will need to be applied to other larger, more diverse nighttime vehicle datasets to better explore its scalability in larger semantic web-based video surveillance systems. The proposed MC-YOLO does not specifically incorporate a dedicated solution for occlusion scenarios. However, in the future, technologies such as DIoU-NMS will be considered to address this issue. Furthermore, the proposed MC-YOLO adopts the lightweight model MobileNetV2 as a backbone network, which can improve the detection efficiency of the model to a certain extent. However, with the transformation of the Web 3.0 era and Metaverse (Deveci et al. 2022), the real-time requirements for vehicle detection algorithms in web video surveillance systems are also increasing. Therefore, in future work, strategies like substituting smaller backbone networks, and implementing pruning techniques will be explored to improve the detection efficiency of the MC-YOLO model. This ongoing refinement aims to further enhance the model's practicality and applicability.

DATA AVAILABILITY

The data used to support the findings of this study are included within the article.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

FUNDING STATEMENT

This work was supported by Fundamental Research Program of Shanxi Province (202103021223029) and Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi, 2021 (STIP).

REFERENCES

Akl, N. A., El Khoury, J., & Mansour, C. J. (2021). Trip-based prediction of hybrid electric vehicles velocity using artificial neural networks. *2021 IEEE 3rd International Multidisciplinary Conference on Engineering Technology (IMCET)*, 60-65. doi:10.1109/IMCET53404.2021.9665641

Al-Qerem, A., Alauthman, M., Almomani, A., & Gupta, B. (2020). IoT transaction processing through cooperative concurrency control on fog–cloud computing environment. *Soft Computing*, *24*(8), 5695–5711. doi:10.1007/s00500-019-04220-y

Al-refai, G., & Al-refai, M. (2020). Road object detection using Yolov3 and Kitti dataset. *International Journal of Advanced Computer Science and Applications*, *11*(8), 48–53. doi:10.14569/IJACSA.2020.0110807

Al-refai, G., & Rawashdeh, O. A. (2020). Improved Candidate Generation for Pedestrian Detection using Background Modeling in Connected Vehicles. *International Journal of Advanced Computer Science and Applications*, *11*(3), 649–660. doi:10.14569/IJACSA.2020.0110381

Alam, A., Jaffery, Z. A., & Sharma, H. (2022). A cost-effective computer vision-based vehicle detection system. *Concurrent Engineering, Research and Applications*, *30*(2), 148–158. doi:10.1177/1063293X211069193

Alcantarilla, P. F., Bergasa, L. M., Jimenez, P., Parra, I., Llorca, D. F., Sotelo, M., & Mayoral, S. S. (2011). Automatic LightBeam controller for driver assistance. *Machine Vision and Applications*, *22*(5), 819–835. doi:10.1007/s00138-011-0327-y

Alhalabi, W., Gaurav, A., Arya, V., Zamzami, I. F., & Aboalela, R. A. (2023). Machine learning-based distributed denial of services (DDoS) attack detection in intelligent information systems. *International Journal on Semantic Web and Information Systems*, *19*(1), 1–17. doi:10.4018/IJSWIS.327280

Almomani, A., Alauthman, M., Shatnawi, M. T., Alweshah, M., Alrosan, A., Alomoush, W., Gupta, B., Gupta, B. B., & Gupta, B. B. (2022). Phishing website detection with semantic features based on machine learning classifiers: A comparative study. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–24. doi:10.4018/IJSWIS.297032

Arora, N., Kumar, Y., Karkra, R., & Kumar, M. (2022). Automatic vehicle detection system in different environment conditions using fast R-CNN. *Multimedia Tools and Applications*, *81*(13), 18715–18735. doi:10.1007/s11042-022-12347-8

Arowolo, M. O., Misra, S., & Ogundokun, R. O. (2023). A machine learning technique for detection of social media fake news. *International Journal on Semantic Web and Information Systems*, *19*(1), 1–25. doi:10.4018/IJSWIS.326120

Bell, A., Mantecón, T., Díaz, C., del-Blanco, C. R., Jaureguizar, F., & García, N. (2022). A novel system for nighttime vehicle detection based on foveal classifiers with real-time performance. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 5421–5433. doi:10.1109/TITS.2021.3053863

Cai, B. X., Wang, Q. D., Chen, W. W., Zhao, L. F., & Wang, H. (2021). Research on vehicle detection based on the regional feature fusion. *Proceedings of the Institution of Mechanical Engineers. Part D, Journal of Automobile Engineering*, *236*(8), 1795–1808. doi:10.1177/09544070211046673

Cui, Y., Xie, S., Xie, X., Zhang, X., & Liu, X. (2022). Dynamic probability integration for electroencephalography-based rapid serial visual presentation performance enhancement: Application in nighttime vehicle detection. *Frontiers in Computational Neuroscience*, *16*, 1006361. doi:10.3389/fncom.2022.1006361 PMID:36313812

Dai, X., Liu, D., Yang, L., & Liu, Y. (2019). Research on headlight technology of night vehicle intelligent detection based on hough transform. In *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)* (pp. 49-52). IEEE. doi:10.1109/ICITBS.2019.00021

Deveci, M., Pamučar, D., Gokasar, I., Koppen, M., & Gupta, B. B. (2022). Personal mobility in Metaverse with autonomous vehicles using Q-Rung Orthopair fuzzy sets based OPA-RAFSI model. *IEEE Transactions on Intelligent Transportation Systems*, 1–10. doi:10.1109/TITS.2022.3186294

Ding, Y. X., Qu, Y. C., Du, D. K., Jiang, Y., Zhang, H., Song, B., Zhou, X., & Sun, J. (2022). Long-distance vehicle dynamic detection and positioning based on Gm-APD Lidar and LIDAR-YOLO. *IEEE Sensors Journal*, *22*(17), 17113–17125. doi:10.1109/JSEN.2022.3193740

Gaurav, A., Gupta, B. B., & Panigrahi, P. K. (2023). A comprehensive survey on machine learning approaches for malware detection in IoT-based enterprise information system. *Enterprise Information Systems*, *17*(3), 2023764. doi:10.1080/17517575.2021.2023764

Gupta, P., Pareek, B., Singal, G., & Rao, D. V. (2022). Edge device based military vehicle detection and classification from UAV. *Multimedia Tools and Applications*, *81*(14), 19813–19834. doi:10.1007/s11042-021-11242-y

Huang, D., Zhou, Z., Deng, M., & Li, Z. (2021). Nighttime vehicle detection based on direction attention network and bayes corner localization. *Journal of Intelligent & Fuzzy Systems*, *41*(1), 783–801. doi:10.3233/JIFS-202676

Kumar, N., Poonia, V., Gupta, B., & Goyal, M. (2021). A novel framework for risk assessment and resilience of critical infrastructure towards climate change. *Technological Forecasting and Social Change*, *165*, 120532. doi:10.1016/j.techfore.2020.120532

Li, F., Jiang, Z., Zhou, S., Deng, Y. T., & Bi, Y. F. (2022). Spilled load detection based on lightweight YOLOv4 trained with easily accessible synthetic dataset. *Computers & Electrical Engineering*, *100*, 107944–107956. doi:10.1016/j.compeleceng.2022.107944

Li, J., Xiao, D., & Yang, Q. (2022). Efficient multi-model integration neural network framework for nighttime vehicle detection. *Multimedia Tools and Applications*, *81*(22), 32675–32699. doi:10.1007/s11042-022-12857-5

Li, Y., Wu, Z., Li, L., Yang, D., & Pang, H. (2021). Improved YOLOv3 model for vehicle detection in high-resolution remote sensing images. *Journal of Applied Remote Sensing*, *15*(2), 026505. doi:10.1117/1.JRS.15.026505

Lian, H. J., Pei, X. F., & Guo, X. X. (2021). A local environment model based on multi-sensor perception for intelligent vehicles. *IEEE Sensors Journal*, *21*(14), 1–10. doi:10.1109/JSEN.2020.3018319

Lyu, W., Lin, Q., Guo, L., Wang, C., Yang, Z., & Xu, W. (2021). Vehicle detection based on an improved faster R-CNN method. *IEICE Transactions on Fundamentals of Electronics, Communications & Computer Sciences, E104/A*(2), 587-590.

Matsui, Y., & Oikawa, S. (2021). Pedestrian detection before motor vehicle moving off maneuvers using ultrasonic sensors in the vehicle front. *Stapp Car Crash Journal*, *65*(1), 163–187. PMID:35512788

Memos, V. A., Psannis, K. E., Ishibashi, Y., Kim, B. G., & Gupta, B. (2018). An efficient algorithm for media-based surveillance system (EAMSuS) in IoT smart city framework. *Future Generation Computer Systems*, *83*, 619–628. doi:10.1016/j.future.2017.04.039

Mo, X. L., Sun, C. P., Zhang, C. Y., Tian, J. P., & Shao, Z. S. (2022). Research on expressway traffic event detection at night based on Mask-SpyNet. *IEEE Access : Practical Innovations, Open Solutions*, *10*(1), 69053–69062. doi:10.1109/ACCESS.2022.3178714

Mu, C. Y., Kung, P., Chen, C. F., & Chuang, S. C. (2022). Enhancing front-vehicle detection in large vehicle fleet management. *Remote Sensing (Basel)*, *14*(7), 1–18. doi:10.3390/rs14071544

Park, J. M., & Lee, J. W. (2020). Combination of SSD and a rule-based approach for nighttime vehicle detection on roads. *Journal of Institute of Control*, *26*(6), 493–498. doi:10.5302/J.ICROS.2020.20.0010

Parvin, S., Rozario, L. J., & Islam, M. E. (2021). Vision-based on-road nighttime vehicle detection and tracking using taillight and headlight features. *Journal of Computational Chemistry*, *9*, 29–53.

Peñalvo, F. J. G., Sharma, A., Chhabra, A., Singh, S. K., Kumar, S., Arya, V., & Gaurav, A. (2022). Mobile cloud computing and sustainable development: Opportunities, challenges, and future directions. *International Journal of Cloud Applications and Computing*, *12*(1), 1–20. doi:10.4018/IJCAC.312583

Prathiba, S. B., Raja, G., Bashir, A. K., AlZubi, A. A., & Gupta, B. (2021). SDN-assisted safety message dissemination framework for vehicular critical energy infrastructure. *IEEE Transactions on Industrial Informatics*, *18*(5), 3510–3518. doi:10.1109/TII.2021.3113130

Ravindran, R., Santora, M. J., & Jamali, M. M. (2021). Multi-object detection and tracking, based on DNN, for autonomous vehicles: A Review. *IEEE Sensors Journal*, *21*(5), 5668–5677. doi:10.1109/JSEN.2020.3041615

Shang, J., Guan, H. P., Liu, Y., Bi, H. B., Yang, L., & Wang, M. (2021). A novel method for vehicle headlights detection using salient region segmentation and PHOG feature. *Multimedia Tools and Applications*, *11*(15), 1–21. doi:10.1007/s11042-020-10501-8

Shao, X., Wei, C., Shen, Y., & Wang, Z. (2021). Feature enhancement based on CycleGAN for nighttime vehicle detection. *IEEE Access : Practical Innovations, Open Solutions*, *9*(1), 849–859. doi:10.1109/ACCESS.2020.3046498

Sharma, R., Sharma, T. P., & Sharma, A. K. (2022). Detecting and preventing misbehaving intruders in the Internet of Vehicles. *International Journal of Cloud Applications and Computing*, *12*(1), 1–21. doi:10.4018/IJCAC.295242

Stergiou, C. L., Psannis, K. E., & Gupta, B. (2021). InFeMo: Flexible big data management through a federated cloud system. *ACM Transactions on Internet Technology*, *22*(2), 1–22. doi:10.1145/3426972

Tsai, W. K., & Chen, H. J. (2021). High-accuracy vehicle lamp detection for real-time night-time traffic surveillance. *IET Intelligent Transport Systems*, *14*(1), 1923–1934.

Vijayakumar, P., Rajkumar, S. C., & Deborah, L. J. (2022). Passive-awake energy conscious power consumption in smart electric vehicles using cluster type cloud communication. *International Journal of Cloud Applications and Computing*, *12*(1), 1–14. doi:10.4018/IJCAC.297108

Wang, H., & Zhang, X. D. (2022). Real-time vehicle detection and tracking using 3D LiDAR. *Asian Journal of Control*, *24*(3), 1459–1469. doi:10.1002/asjc.2519

Wang, Z., Zhan, J., Duan, C., Guan, X., & Yang, K. (2022). Vehicle detection in severe weather based on pseudo-visual search and HOG–LBP feature, fusion. *Proceedings of the Institution of Mechanical Engineers. Part D, Journal of Automobile Engineering*, *236*(7), 1607–1618. doi:10.1177/09544070211036311

Yang, J., Wang, C., Wang, H., & Li, Q. (2020). A RGB-D based real-time multiple object detection and ranging system for autonomous driving. *IEEE Sensors Journal*, *20*(20), 11959–11966. doi:10.1109/JSEN.2020.2965086

Yi, S., Zhou, S. Y., Shen, L., & Zhu, J. M. (2021). Vehicle-based thermal imaging target detection method based on enhanced lightweight network. *Infrared Technology*, *43*(3), 237–245.

Zaarane, A., Slimani, I., Al Okaishi, W., Atouf, I., & Hamdoun, A. (2019). An automated night-time vehicle detection system for driving assistance based on cross-correlation. In *2019 International Conference on Systems of Collaboration Big Data, Internet of Things & Security (SysCoBIoTS)* (pp. 1-5). IEEE. doi:10.1109/SysCoBIoTS48768.2019.9028038

Zhang, B. L., Qin, H. R., Jiang, S., Zheng, J. Y., & Wu, Z. H. (2021). A method of vehicle detection at night based on RetinaNet and Optimized Loss Functions. *Automotive Engineering*, *43*(8), 1195–1202.

Zhang, K., Wang, C., Yu, X., Zheng, A., Gao, M., Pan, Z., Chen, G., & Shen, Z. (2022). Research on mine vehicle tracking and detection technology based on YOLOv5. *Systems Science & Control Engineering*, *10*(1), 347–366. doi:10.1080/21642583.2022.2057370

Zhang, Q., Guo, Z., Zhu, Y., Vijayakumar, P., Castiglione, A., & Gupta, B. (2023). A deep learning-based fast fake news detection model for cyber-physical social services. *Pattern Recognition Letters*, *168*, 31–38. doi:10.1016/j.patrec.2023.02.026

Zhang, S., & Tong, Y. (2023). Nighttime vehicle detection algorithm enhanced by NightvisionGAN. *2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 672-676. doi:10.1109/CVIDL58838.2023.10165840

Zhang, X., Story, B., & Rajan, D. (2021). Night time vehicle detection and tracking by fusing vehicle parts from multiple cameras. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 8136–8156. doi:10.1109/TITS.2021.3076406

Zhao, K., Liu, L., Meng, Y., Liu, H., & Gu, Q. (2022). 3D detection for occluded vehicles from point clouds. *IEEE Intelligent Transportation Systems Magazine*, *14*(5), 59–71. doi:10.1109/MITS.2021.3064897

Kun Wang, Associate Professor, received his PhD from South China University of Technology in 2017. He currently works at Shanxi University. His research interests include intelligent sensing and ASIC/hardware system and the Internet of Things.