IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 May 2023, accepted 13 June 2023, date of publication 19 June 2023, date of current version 22 June 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3287315

RESEARCH ARTICLE

Recognition of Front-Vehicle Taillights Based on YOLOv5s

HUAYUE ZHANG, JUNYOU ZHANG, SHUFENG WANG, QIAN ZHOU, AND XIAOLEI LI
Department of Transportation Engineering, College of Transportation, Shandong University of Science and Technology, Qingdao 266590, China
Corresponding author: Shufeng Wang (shufengwang@sdust.edu.cn)

This work was supported by the Natural Science Foundation of Shandong Province in China under Grant ZR2019MF056.

ABSTRACT In automatic driving, the recognition of Front-Vehicle taillights plays a key role in predicting the intentions of the vehicle ahead. In order to accurately identify the Front-Vehicle taillights, we first analyze the different characteristics of the vehicle taillight signal, and then propose an improved taillight recognition model based on YOLOv5s. First, CA(coordinate attention) is inserted into the backbone network of YOLOv5s model to improve small target recognition and reduce interference from other light sources; Then, the EIOU Loss is used to solve the class imbalance problem; Finally, EIOU-NMS is used to solve the problem of anchor box error suppression in the recognition process. We use the actual scene video and vehicle taillights dataset to conduct ablation experiments to verify the effectiveness of the improved algorithm. The experimental results show that the mAP value of the model is 9.2% higher than YOLOv5s.

INDEX TERMS Autonomous driving, vehicle taillights recognition, ablation experiment.

I. INTRODUCTION

In the past decades, automatic driving technology has developed to a more intelligent and autonomous stage [1]. Accurately identifying the intention of the vehicle in front is particularly critical for automatic driving. The Front-Vehicle taillights is a very important signal for the vehicle in front to transmit the driving intention. Therefore, accurately identifying the Front-Vehicle taillights information can greatly improve the safety of automatic driving. At present, many methods have been explored to detect and identify taillights, traditional methods and methods based on deep learning. Traditional methods mainly use manual extraction of visual information features, classifier classification, and finally combine with machine learning models. Regarding the machine learning classifiers, AdaBoost and SVM (Support Vector Machines) are the most commonly used ones [5]. In fact, these algorithms may not be able to detect the target accurately, because the manually extracted features are usually unable to adapt to different lighting conditions, as well as the confusion of other lights and the occlusion of obstacles. For example, Cui et al. [3] developed a layered method to detect daytime vehicles and taillights. First, deformable part model (DPM) [4] is used to detect vehicles, then HSV is used to pair taillights, and finally Sparse Dictionary Learning method is used to classify signal lights. Almagambetov et al. [5] proposed a stable and portable algorithm for vehicle vision system, which can track vehicle taillights and detect steering and braking signals through feature tracking and area tracking. Zhong et al. [6] proposed a method based on machine learning to predict the brake light status. This method first locates the vehicle and brake light area, and then uses SVM classifier to identify.

In recent years, due to the improvement of computing power, deep neural networks have been widely used. Compared with the traditional shallow methods, convolutional neural network (CNN) has achieved good performance by learning to extract features from images hierarchically, and has made significant improvements in image classification (such as ResNet [7], [8]), target detection (such as SSD [9] and Faster R-CNN [10]), etc. In general, deep learning based object detection methods are mainly divided into two categories, i.e., single-stage and two-stage detectors, respectively. The R-CNN series [10], [11], [12] represented two-stage detectors first generate a group of candidate regions that may contain objects through the region suggestion module, and then the detection head classifies the suggestions and predicts the anchor box coordinates. In contrast, single-stage

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

detectors regard target detection as a regression problem and directly predict the category labels and boundary box coordinates of each anchor position. Single stage detector integrates all calculations into a single network, making the model efficient and scalable. Representative single-stage detectors include SSD, YOLO series models [13], [14], [15], [16] and RetinaNet [17].

The taillight detection methods based on deep learning can be divided into three categories. In the first class, the vehicle is detected using a deep learning model, and then the taillights is detected using the artificial image features with a machine learning algorithm. For example, Nava et al. [18] proposed a method to detect vehicles and brake lights in real time, using lane detection algorithm and YOLO model to detect vehicles in front, and using SVM to identify brake lights, but the image features set manually were rough, resulting in insufficient detection accuracy and slow detection speed.

The second method uses independent deep learning sub network to detect vehicles and taillights. For example, Vancea and Nedevschi [19] designed a convolutional neural network workflow that can detect the driving direction of vehicles and taillights. In particular, Faster R-CNN was used to detect vehicles and classify their directions, and its signal state was recognized after the taillight pixels were segmented by a sub network, this method has greatly improved the detection accuracy, but its detection speed still needs to be improved.

The third type of method uses an end-to-end deep learning framework for vehicle detection and taillight detection. For example, Yi et al. [20] proposed a lightweight object detection model based on CA [21], which uses depthwise separable convolutions to optimize model computational complexity, For common traffic target categories, the mean average precision (mAP) is 0.96% higher than the original model, which is better than the mainstream detection models SSD300 and YOLOv4-Tiny, which are 9.95% and 11.66% higher, respectively. Zhang et al. [22] proposed an improved algorithm EIoU-YOLOV5 based on Yolov5, which improves the Loss function of prediction box by replacing CIOU Loss with EIOU Loss [23]. Compared with the original Yolov5 algorithm, the Recall of EIOU-YOLOV5 algorithm increased from 0.855 to 0.917, which increased by 6.2 percentage points.Therefore, the improved EIOU-yolov5 algorithm can better assist clinical diagnosis. Li et al. [24] proposed an improved target detection algorithm based on YOLOv3. First, the output layer is increased through feature fusion to enhance the small target detection capability. Second, the SPP module is introduced to extract multi-scale features. Finally, the focus loss is used to solve the class imbalance problem, and the taillights signal recognition is realized. Song et al. [25] this work proposes an action-state joint learning-based vehicle taillight recognition method on the basis of vehicles detection and tracking, which takes both taillight state features and time series features into account, consequently getting practicable results even in complex actual scenes. Pirhonen et al. [26] proposes a new brake light detection algorithm. A camera and YOLOv3 are used to detect the anchor box of the vehicle in front of the vehicle. L*a*b color space threshold preprocessing was performed on the anchor box. Then, the boundary box was adjusted to a resolution of 30 × 30 pixels and input into the random forest algorithm, and the accuracy of the algorithm reached 81.8%. Iftikhar et al. [27] the 2D convolutional neural network was used to estimate the achievement of the current 3D detection system in rejecting the real and fake pedestrian problem. The single-phase (YOLOv3 model) and two-phase (faster R-CNN) deep learning meta-structures under different image resolutions and attribute extractor (MobileNet) were evaluated to improve the accuracy of the real and fake pedestrian detection. And it still has real-time requirements. Iftikhar et al. [28] pedestrian detection problems and the latest progress in solving these problems with the help of distance learning technology are summarized. Also presented are informative discussions and future research work designed to provide readers with insights and inspire new research directions.

Although the vehicle tail light detection based on deep learning has achieved some success [6], [19], [20], [22], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], it is still easy to have problems such as error identification and missing identification when detecting small targets such as steering lights or when there are many vehicles ahead.

In order to solve the above problems, and considering that the end-to-end deep learning framework has the characteristics of fast recognition speed and high accuracy, we propose an end-to-end deep learning framework based on YOLOv5s to identify the Front-Vehicle taillights. The work of this paper has three characteristics:

(1) The dataset used in this paper considers various road conditions (traffic jam, no traffic jam), weather conditions (rainy days, sunny days) and time (day, evening, night) to ensure the reliability of the model application.

(2) We have selected a variety of solutions (CA, EIOU Loss, EIOU-NMS) for the above problems and added them to the target network in turn, thus improving the recognition capability of YOLOv5s.

(3) After a lot of comparative experiments, our model has good target recognition ability under both day and night conditions.

II. RELATED WORK

A. WORKFLOW AND CHARACTERISTIC ANALYSIS

As an important means of communicating with each other during driving, the Front-Vehicle taillights have different use environments and color characteristics:

(1) The brake light is characterized by the combination of the high mounted brake lamp on the top of the rear and the brake lights on both sides that light up red at the same time. It is mainly used during deceleration and steering. Because of its obvious color and location characteristics, it is not easy to be interfered by light and other external factors.

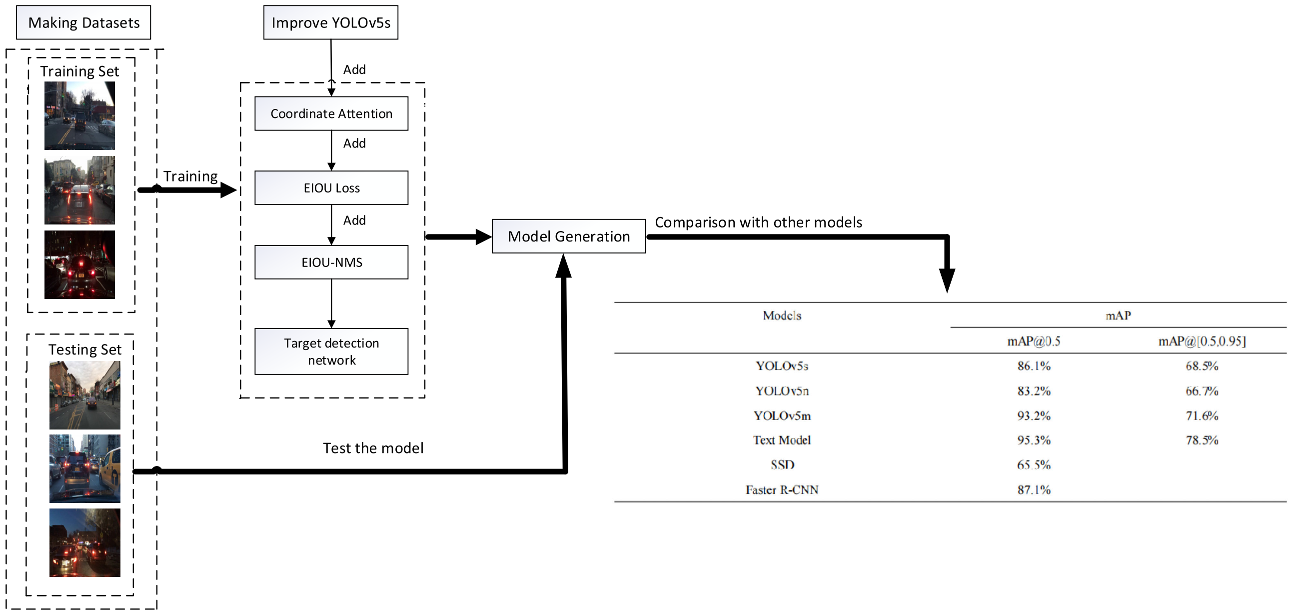


FIGURE 1. **Experimental procedure of this paper.**

(2) The left turn light is characterized by the amber flashing of the left taillight and the right taillight is characterized by the amber flashing of the right taillight. The main function of left turn and right turn is to remind the vehicles behind when the vehicle is turning. Because the turn signal light only lights up one taillight on one side and the color is light, it is easy to be affected by the light and the vehicles that are closer to each other. In addition, the target of the turn signal is small, and because it is on one side of the tail light, it is easy to miss detection in the case of occlusion.

In order to solve the above problems, we propose several improvement methods and compare them with other methods in the following. The workflow is shown in Figure 1, The model refers to the model trained at each stage of the ablation experiment.

B. PROPOSED ALGORITHM

This section mainly introduces the method of detecting and identifying Vehicle Taillights based on YOLOv5s. First, the network structure of YOLOv5s is introduced, and then three modification schemes for YOLOv5s are introduced:

(1) CA added to enhance the recognition ability of small targets and reduce other optical interference.

(2) Add EIOU-NMS to solve the error suppression of adjacent close targets.

(3) EIOU Loss is used to replace YOLOv5s' loss function, which effectively alleviates the problem of sample imbalance.

According to relevant papers and by comparing the results of ablation experiments, CA and EIOU Loss were selected, and EIOU-NMS was improved according to DIOU-NMS. The above three methods have improved the recognition accuracy of YOLOv5s, and are first applied in the Front-Vehicle Taillights recognition field.

III. ALGORITHM INTRODUCTION

A. INTRODUCTION OF YOLOV5S ALGORITHM

The YOLO series has become the most popular target recognition method for practical applications due to its excellent recognition accuracy and speed [42].

YOLOV5 consists of four components: Input, Backbone, Neck and Prediction. YOLOv5-6.1 contains five models, namely YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, which differ in the depth and width of the feature maps. YOLOv5x has high recognition accuracy but high calculation loss, which cannot meet the requirements of real-time performance. YOLOv5n meets the requirements of real-time performance, but its recognition accuracy cannot meet the requirements. YOLOv5s combines the advantages of YOLOv5x and YOLOv5n. It not only meets the requirement of real-time performance but also has good recognition accuracy. Therefore, this paper uses YOLOv5s as the basic model and optimizes it to achieve accurate vehicle taillights signal recognition.

YOLOv5s uses Mosaic data enhancement and adaptive anchor box calculation at the output; The Focus structure is used in Backbone to slice the input image to preserve more fine-grained features; Using FPN+PAN+CSP2-X structure in Neck to enhance network feature fusion capability; CIOU-Loss [43] selected at output computes the IOU as a Loss function of the target bounding box, and filters the anchor box using a weighted NMS operation. Although YOLOv5s has a great improvement in recognition speed, there are still some problems such as inadequate accuracy and target frame error suppression in the recognition of vehicle taillights, which are subject to more external interference. Therefore, this paper puts forward three improvement methods to further improve the recognize accuracy of

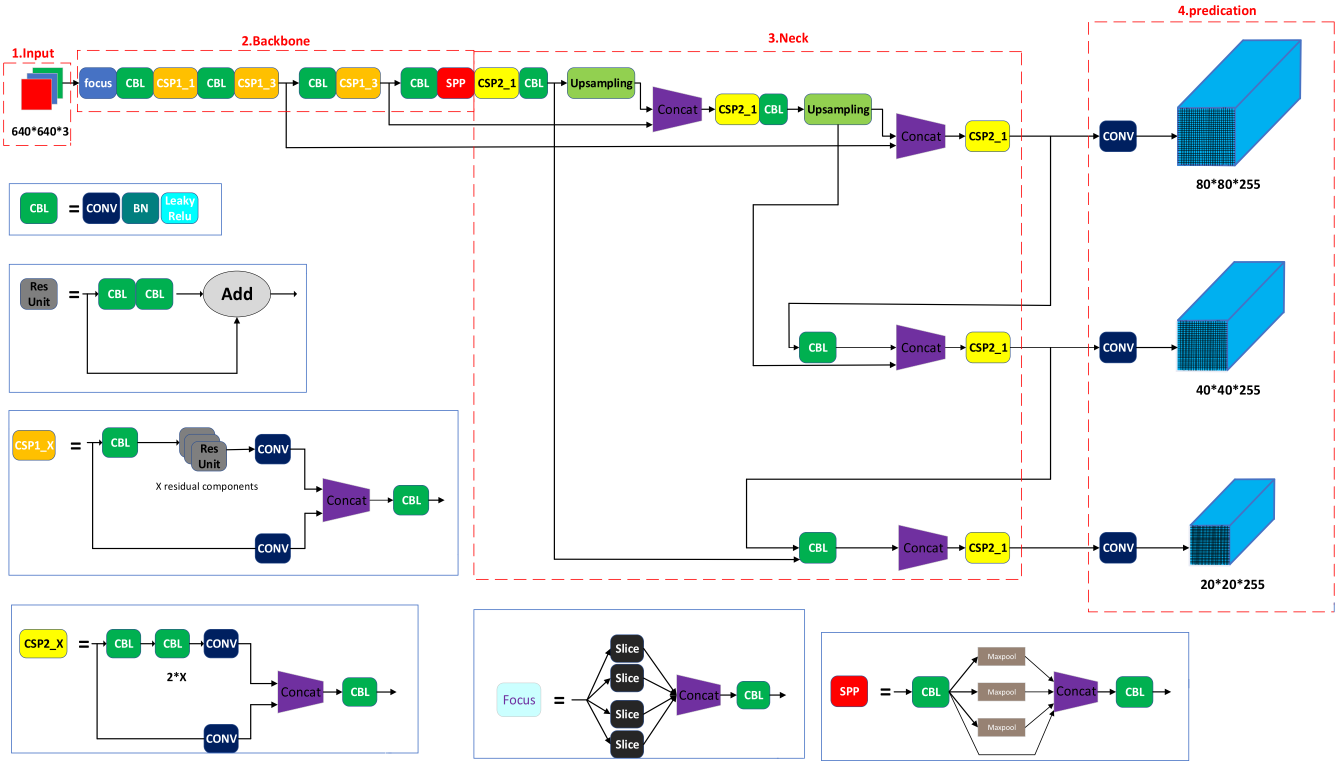


FIGURE 2. YOLOv5s network structure.

vehicle taillights. The structure flow of YOLOv5s is shown Figure 2.

B. COORDINATE ATTENTION

The function of attention mechanism is to make recognition algorithm ignore irrelevant information in data and focus on key information. Generally speaking, it can be divided into channel attention mechanism, spatial attention mechanism and attention mechanism combining the two. At present, the popular attention mechanisms in deep learning include SENet [44], CBAM [45] and CA.

CA encodes channel relationships and long-term dependencies through accurate location information. The specific operation is divided into two steps: embedding of coordinate information and generating of coordinate attention. Its structure is shown in Figure 3.

(1) Coordinate information embedding: Because the global pooling method compresses the global information into the channel descriptor, it is difficult to save the location information. In order to enable the attention module to capture accurate position information, the CA decomposes the global pooling according to the following formula, which is converted into a pair of one-dimensional feature encoding operations:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \tag{1}$$

Specifically, given an input X, each channel is first encoded along the horizontal and vertical coordinates using a pooling kernel of size $(H, 1)$ or $(1, W)$, respectively. Thus, the output of the cth channel with height h can be expressed as:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i) \tag{2}$$

Similarly, the output of the cth channel with width w can be expressed as:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w) \tag{3}$$

The two transformations described above allow the attention module to capture long-term dependencies along one spatial direction and to preserve location information along another spatial direction, facilitating the network to locate the target of interest more accurately.

(2) Coordinate Attention generation: To exploit the representations generated by Coordinate information embedding, Coordinate Attention proposes a second transformation called Coordinate Attention generation. It can make full use of the captured location information so that the region of interest can be captured accurately and also capture the relationship between channels effectively.

By means of the transformation of the information embedding, the second transformation performs a linking operation on the above transformation, which is then transformed using the 1×1 convolutional transformation function F_1:

$$f = \delta(F_1([z^h, z^w])) \tag{4}$$

The equation $[z^h, z^w]$ is a linking operation along the spatial dimension, and δ is the nonlinear activation function, and f is the intermediate feature mapping that encodes the spatial information in the horizontal and vertical directions, respectively f^h and f^w. Using the other two 1×1 convolutional transform F_h and F_w respectively, the f^h and f^w transform to a tensor with the same number of channels to the input X to obtain the equation:

$$g^h = \sigma(F_h(f^h)) \tag{5}$$
$$g^w = \sigma(F_w(f^w)) \tag{6}$$

Usually an appropriate reduction ratio r (e.g., 32) is used in this step to reduce the f the number of channels, thus reducing the complexity and computational overhead of the model. Afterwards, the modeling of g^h and g^w are extended and used as attention weight values, respectively, and the output of the final Coordinate Attention Y can be expressed as:

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{7}$$

SENet belongs to channel attention mechanism, CBAM belongs to channel attention mechanism combined with spatial attention mechanism, CA belongs to channel attention mechanism plus x direction attention mechanism and y direction attention mechanism. The addition of CA helps to identify small targets and reduce the interference of other light during recognition.

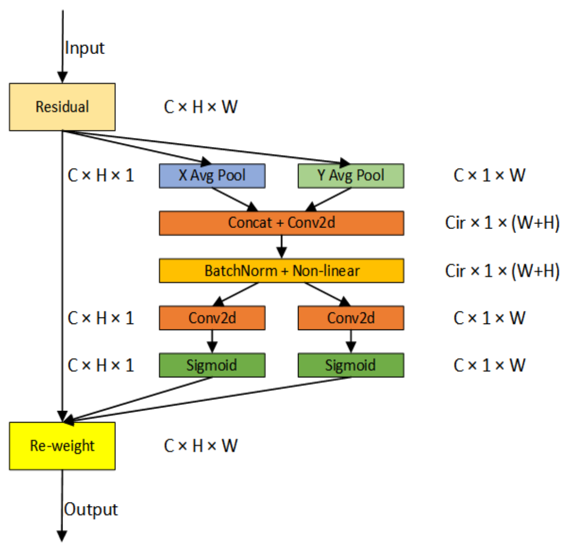


FIGURE 3. **Coordinate attention structure diagram.**

C. EIOU LOSS

Object detection uses bounding box regression prediction to locate targets in the image, and early object detection work used IOU [46] as the localization loss. The definition of IOU Loss is to first calculate the ratio of the intersection and union between the predicted box and the real box, and then calculate the negative logarithm. However, in practical use, we often write IOU Loss as $1-IOU$. This has inspired several improved IOU based loss designs, including CIOU Loss and EIOU Loss. In YOLOv5s, the loss of confidence and the loss of classification probability are calculated by BCE Loss,

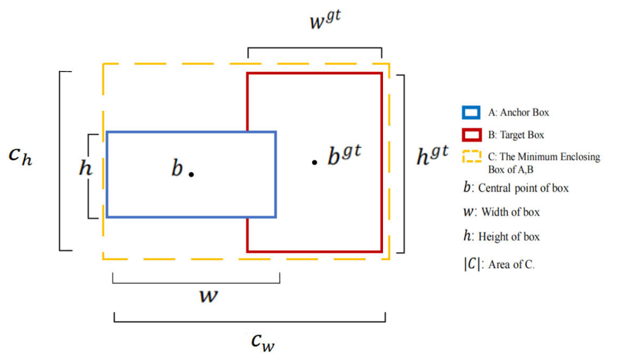


FIGURE 4. **IOU loss diagram.**

TABLE 1. **Number of each category of taillight identification dataset.**

Type	brake	Left	Right	Car
Training	4125	1352	1481	10432
Testing	500	500	500	3200

while the loss of rectangular box is calculated by CIOU Loss. Figure 4 illustrates the IOU Loss.

CIOU Loss considers the overlapping area, center point distance and aspect ratio of anchor box regression, but it reflects the difference of aspect ratio through the formula v instead of the difference between width and height and their confidence, so it sometimes hinders the similarity optimization of the model. To solve this problem, EIOU Loss takes apart the aspect ratio based on CIOU Loss and adds Focal Loss, solving the problem of penalty failure caused by proportional changes in aspect ratio in CIOU Loss.

The formula of CIOU Loss is as follows:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{8}$$
$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \tag{9}$$

where IOU denotes the overlap area of the bounding box regression, i.e., the overlap area of the predicted box and the actual box. The α in formula (8) is the weight function. Formula (9) is used to measure the similarity of aspect ratio. The purpose of $\frac{\rho^2(b,b^{gt})}{c^2}$ is to calculate the Euclidean distance between the center points of two anchor boxes.

The formula for EIOU-Loss is as follows:

$$L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \tag{10}$$

where c_w and c_h is the width and height of the minimum bounding box that covers both bounding boxes.

In summary, there are two main advantages of EIOU Loss:

(1) Splitting the Loss item of aspect ratio into the difference of predicted width and minimum outer frame width, accelerates convergence and improves regression accuracy;

(2) Focal Loss was introduced to optimize the sample imbalance in the bounding box regression task, eliminating

FIGURE 5. Sample image of taillight identification dataset ((a) to (f)).

the prediction boxes that overlap the target boxes less, and focusing the regression process on high-quality anchor boxes.

D. EIOU-NMS

NMS (Non Maximum Suppression) is used to eliminate redundant anchors. The suppression process is an iterative ergodic elimination process. The main process is as follows:

(1) Sort the scores of all anchor boxes and select the box with the highest score;

(2) Traverse the rest of the anchor boxes. If the overlap area with the current highest score anchor box is larger than a certain threshold, the anchor box will be deleted;

(3) Continue to select the one with the highest score from the unprocessed anchor box, and repeat the above process.

The NMS of YOLOv5s uses IOU as the threshold to suppress redundant anchor boxes, but IOU only considers overlapping areas, which is easy to lead to error suppression.

The use of DIOU Loss as the standard of NMS not only considers the overlapping area, but also considers the distance of the center point, which makes the processing result of NMS more reasonable (DIoU-NMS thinks that the anchor box with two far center points may be located on different objects and should not be deleted). However, although DIOU-NMS considers the distance of the center point, it can still cause false suppression when facing two close targets. DIOU-NMS formula is as follows:

$$s_i = \begin{cases} s_i, & IoU - R_{DIoU}(M, B_i) < \varepsilon \\ 0, & IoU - R_{DIoU}(M, B_i) \geq \varepsilon \end{cases} \quad R_{DIoU} = \frac{\rho^2(b, b^{gt})}{c^2} \tag{11}$$

where s_i is the classification confidence level, and ε is the NMS threshold (generally set to 0.6), and M is the box with the highest confidence, and B_i is the overlap between each box and the box with the highest confidence, and R_{DIoU} is the distance between the centroids of the two anchor boxes.

We modified DIOU-NMS and got EIOU-NMS. Based on DIOU, the aspect ratio of two anchor boxes is considered, which greatly improves the problem of error suppression for two close targets. The EIOU-NMS formula is as follows:

$$s_i = \begin{cases} s_i, IoU - R_{EIoU}(M, B_i) < \varepsilon \\ 0, IoU - R_{EIoU}(M, B_i) \geq \varepsilon \\ R_{EIoU} = \dfrac{\rho^2(b, b^{gt})}{c^2} + \dfrac{\rho^2(w, w^{gt})}{c_w^2} + \dfrac{\rho^2(h, h^{gt})}{c_h^2} \end{cases} \tag{12}$$

where s_i is the classification confidence level, and ε is the NMS threshold (generally set to 0.6), and M is the box with the highest confidence, and B_i is the overlap between each box and the box with the highest confidence, and R_{EIoU} is the distance between the center points of the two anchor boxes plus the aspect ratio of the two anchor boxes. The addition of EIOU-NMS helps to suppress the anchor box correctly during target recognition.

IV. VEHICLE TAILLIGHTS RECOGNITION EXPERIMENT

In this section, we first introduce the characteristics of taillights and the dataset of taillights used. and then verify the effectiveness of this method through experiments. Finally, we use ablation experiments to compare the identification results of different improvement stages and other target

TABLE 2. Comparison of three attention models.

Attentional Model	Precision	mAP	
		mAP@0.5	mAP@[0.5,0.95]
YOLOv5s	81%	86.1%	68.5%
YOLOv5s+SENet	85.9%	87.1%	69.3%
YOLOv5s+CBAM	87.2%	88.5%	72.1%
YOLOv5s+CA	90.8%	91.3%	73.8%

TABLE 3. Comparison results of three loss functions.

Loss function	Precision	mAP	
		mAP@0.5	mAP@[0.5,0.95]
YOLOv5s+CIOU Loss	81%	86.1%	68.5%
YOLOv5s+DIOU Loss	84.5%	87.2%	69.2%
YOLOv5s+EIOU Loss	89.9%	88.1%	70.1%

detectors. In addition, we also apply the improved model to real scenes to verify the robustness and practicality of the model.

The ablation experiment is equivalent to the control variable method, which means that when the author proposes a new scheme, the scheme changes multiple parameters at the same time, and the authors take turns keeping a parameter constant, and finally compare the results to identify which parameter has greater impact on the results.

In this ablation experiment, three attention mechanisms, three loss functions and three NMS were added to YOLOv5s respectively, and only one method was added to YOLOv5s each time; Then added YOLOv5s in turn according to one attention mechanism, one loss function and one NMS random combination; After training the above two experimental methods with the taillight dataset, the model is tested in the actual scene video, and the test results are obtained. And repeat the above training and testing process to get the test results. Finally, the best three methods of test results are added into YOLOv5s in order to repeat the above training and testing process to verify the effectiveness and practicality of the three methods.

A. DATASET AND EVALUATION INDEX

1) PRODUCTION OF DATASETS

This dataset(Taillight Dataset: 20% of the data is internet search, and 80% of the data is self-shooting) collects ten videos taken by drive recorder in different time periods and different road conditions, including videos in different environments and different road conditions during the day, at dusk and at night. After the video is transformed into frames, it is filtered to make a dataset and labeled with LabelImg(YOLO Format) to ensure the applicability and robustness of the model.

TABLE 4. Comparison results of three NM.

NMS Metrics	Precision	mAP	
		mAP@0.5	mAP@[0.5,0.95]
YOLOv5s+NMS	81%	86.1%	68.5%
YOLOv5s+DIOU-NMS	85.9%	86.5%	69%
YOLOv5s+EIOU-NMS	89.7%	87.6%	71.2%

It can be seen from many road videos that the taillight does not necessarily light up when the car is driving, and the process of turning is often accompanied by braking and deceleration, and the double flash almost does not appear. Therefore, braking is the most important semantic information released by vehicles running on the road, accounting for 67%; Left turn and right turn account for 17% and 14% respectively according to different road conditions in different countries and regions; Double flash only accounts for 2% and there are few application scenarios. Therefore, the semantic features of vehicle taillights in this dataset are brake most, left and right less. The dataset includes three types:

(1) 40 minutes of daytime video, including 20 minutes of congestion video and 5 minutes of rain video. The daytime video is characterized by less light interference and clear images;

(2) Dusk video lasts for 20 minutes, including 10 minutes of congestion video. Dusk video is characterized by light interference but less than that at night;

(3) There are 50 minutes of night video, including 35 minutes of congestion video. Night video is characterized by more light interference and unclear images; Congestion video is characterized by close vehicles and multiple recognition targets. Then, according to the characteristics of the taillights analyzed above, the images are labeled as brake, Left, Right and Car to better distinguish the different semantic information of the taillights. Figure 5 shows an example of the image from the dataset.

There are 7000 images in the vehicle taillights identification dataset in this paper, and the number of objects in each category is shown in Table 1. The dataset is divided into training set, validation set and test set(The data pictures included vary). Training set is used to train model parameters; The validation set is used to adjust the model hyperparameters to prevent the model from over fitting; Test sets are used to evaluate model accuracy. The dataset is divided into training sets, validation sets and test sets according to the proportion of 6:2:2, that is, 80% of the data is used for training(Both the training set and the validation set are applied to the training, and the test set is applied to the experimental part of the model after the training).

2) EVALUATION INDICATORS

The most common model evaluation indicators (model performance) are mainly precision (P), recall (R), average precision (AP), and mean average precision (mAP). Precision

TABLE 5. **Ablation experiment results of front-vehicle taillights recognition.**

Models	Precision	Recall	F1-curve	mAP	
				mAP@0.5	mAP@[0.5,0.95]
YOLOv5s	81%	91%	83%	86.1%	68.5%
YOLOv5s+CA	90.8%	93%	93%	91.3%	73.8%
YOLOv5s+EIOU Loss	89.9%	90%	89%	88.1%	70.1%
YOLOv5s+EIOU-NMS	89.7%	89%	88%	87.6%	71.2%
YOLOV5s+CA+EIOU Loss	93.2%	95%	94%	93%	75.5%
YOLOv5s+CA+EIOU-NMS	92.2%	94%	93%	94.2%	75.7%
YOLOv5s+EIOU Loss+EIOU-NMS	93.1%	95%	93%	94.7%	76%
YOLOV5s+CA+EIOU Loss+EIOU-NMS	97%	99%	95%	95.3%	78.5%

TABLE 6. **Comparison results of front-vehicle taillights recognition.**

Models	mAP		AP				FPS
	mAP@0.5	mAP@[0.5,0.95]	Car	brake	Left	Right	Average Value
YOLOv5s	86.1%	68.5%	95%	93.2%	75.4%	80.7%	80
YOLOv5s+CA+EIOU Loss+EIOU-NMS	95.3%	78.5%	99.3%	97.6%	87.9%	96.4%	70
SSD	65.6%	\	93.2%	63.3%	45.2%	30%	20
Faster R-CNN	87.1%	\	96%	85.9%	76%	74.4%	16
YOLOv5n	83.2%	66.7%	93%	89.1%	71.4%	75.2%	90
YOLOv5m	93.2%	71.6%	96.5%	94.2%	78.6%	79.4%	74

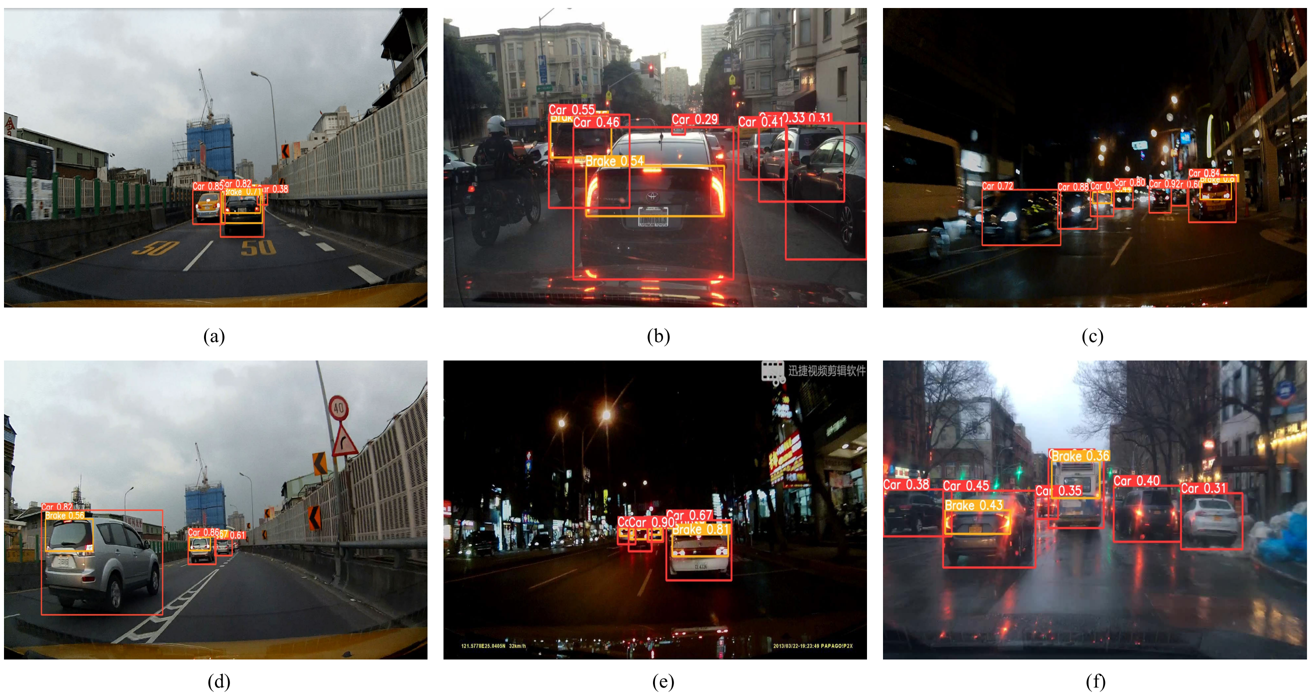


FIGURE 6. **Identification results of some actual scene images ((a) to (f)).**

represents the ability of the model to identify related objects; it is the percentage of correct predictions. Recall refers to the ability of the model to find all relevant objects; it is the percentage of true positives detected in all ground truths. *AP* measures the performance of the classifier, namely, the area of the $P - R$ curve.

The formula of precision, recall and mAP is as follows:

$$P = \frac{TP}{TP + FP} \tag{13}$$

$$R = \frac{TP}{TP + FN} \tag{14}$$

$$mAP = \int_0^1 P(R)dR \tag{15}$$

TP (True Positive) means to correctly classify positive examples as positive examples; FN (False Negative) indicates that positive cases are wrongly classified as negative cases; TN (True Negative) means correctly classifying negative cases as negative cases; FP (False Positive) indicates that negative cases are incorrectly classified as positive cases.

B. MODEL TRAINING

1) EXPERIMENTAL ENVIRONMENT

Deep learning model occupies a lot of computing resources and memory in the training process, which requires high machine performance. Good experimental environment support and high configuration hardware can improve the training efficiency. Therefore, this paper builds a deep learning network environment based on the GPU(Graphics Processing Unit), and the detailed configuration is as follows: The operating system is WIN10; The Deep Learning Framework is Pytorch 1.11; The CPU(Central Processing Unit) is Intel Core i9-10850K; The GPU is NVDIA Geforce RTX 3070; The Operating Platform is Cuda11.3 and Cudnn8.4.0.27; The Programming Language is Python.

The hardware configuration, experimental environment and parameter settings of the training model used in the experiment are as follows: The Weights is YOLOv5s.pt; The Epoch is 300; The Weight_decay is 0.0005; The batch-size is 32; The Initial Learning Rate is 0.01; The Momentum is 0.937.

C. EXPERIMENTAL RESULTS AND ANALYSIS

1) ABLATION EXPERIMENT BASED ON VEHICLE TAILLIGHT DATASET

When the tail light of the front vehicle lights up, which does not affect the normal driving of the vehicle behind, this model only detects the vehicle in front. Only when the taillights of the vehicle in front show signals that affect the normal running of the vehicle behind, can the semantic information in the taillights area be recognized. The purpose of identifying a vehicle is to pair the taillights with the vehicle to prevent false recognition.

We used taillight dataset to conduct ablation experiments on the improved YOLOv5s, and the comparison results are shown in Table 2, Table 3 and Table 4.

We compared YOLOv5s with different attention models in Table 2, where the accuracy of CA was 5.4% higher than SENet, 4.1% higher than CBAM, and 6.1% higher than the original YOLOv5s. According to the proportion classification in the image, Car belongs to large objects, the brake becomes a medium object due to the presence of high mounted brake lights, while Left and Right belong to small objects with different features, which increases the difficulty of training and recognition. In order to improve the performance of the model, this article has decided to add CA in the recognition process to improve the accuracy of recognition, as CA can accurately locate the region of interest.

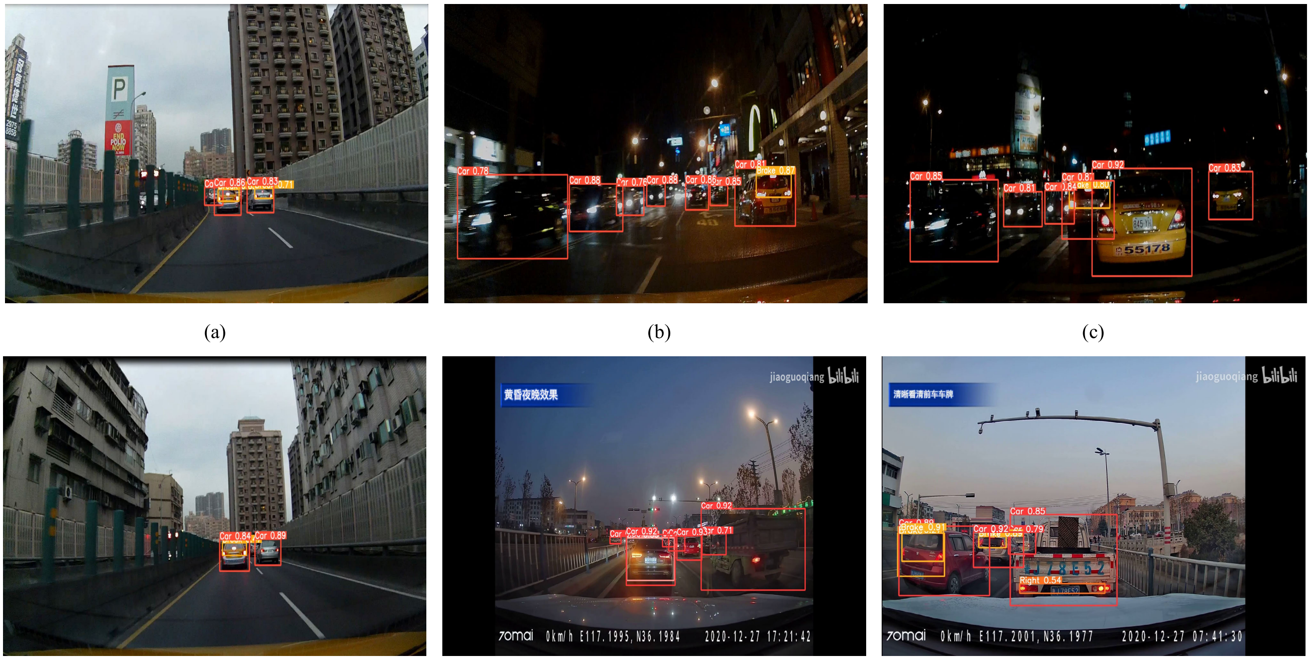
The comparison results of loss function in Table 3 show that the accuracy of EIOU Loss is 4.7% higher than that of CIOU Loss used in the original YOLOv5s and 5.4% higher than that of DIOU Loss. As the most common sample, cars and brakes account for a large proportion in the data set, while turn signal account for a small proportion, which often leads to sample imbalance. In order to solve the problem of sample imbalance and improve the training speed of the model, we have decided to add EIOU Loss.

We improved EIOU-NMS on the basis of DIOU-NMS, and EIOU is used as the standard of NMS. In this experiment, we set the threshold value of NMS to 0.6. Table 4 shows that compared with the original NMS, the accuracy of EIOU-NMS is improved by 4.5% and that of DIOU-NMS is improved by 3.8%. Because there are many vehicles running on the current road and they are often close to each other. So we decided to add EIOU-NMS to prevent false suppression when identifying Front-Vehicle taillights.

Based on the comparison results of ablation experiments in Table 2, Table 3 and Table 4, we decided to select CA, EIOU Loss and EIOU-NMS as the methods to add to YOLOv5s. Table 5 shows the comparison results of the above three ablation experiments, and Table 6 shows the comparison results between the model in this paper and other target detectors.

As shown in Table 6, compared with YOLOv5s, the AP gains of this model on Car, brake, Left and Right are 4.3%, 1.4%, 12.5% and 15.7% respectively. After adding CA, the recognition accuracy of the turn signal lamp has increased significantly, because CA divides the spatial attention mechanism into the attention mechanism in the X direction and the Y direction, allowing the model to accurately locate the area of interest during recognition, improve the recognition ability of small targets, and reduce the interference of other light. EIOU LOSS makes the model pay more attention to the categories with fewer samples in the training process, optimizes the problem of sample imbalance, calculates the difference between the predicted value and the real value, and improves the training speed. After that, EIOU-NMS solves the problem of anchor box error suppression caused by close targets in complex traffic environment and improves the recognition accuracy.

We tested two representative target detectors SSD and Faster R-CNN using the taillight dataset, and finally the model in this paper (96.1% mAP@0.5) is proved it has better performance than SSD (65.6% mAP) and Faster R-CNN (87.1% mAP). The test results are shown in Figure 6((a) to (f)). In Table 6, the recognition speed of each



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

FIGURE 7. Partial identification results of actual scene video in this model ((a) to (f)).

model has been compared. Although the calculation amount has been increased to improve the accuracy of the model, the real-time performance of the model in this paper can still be guaranteed.

2) THE VEHICLE TAILLIGHTS VIDEO RECOGNITION EXPERIMENT

In this section, the real scene video from the vehicle perspective is used to evaluate the model. Figure 7((a) to (f)) shows some of the test results. The test results show that the model has high recognition accuracy for two common targets, vehicle and brake. At the same time, we accurately identify the left turn signal and the right turn signal. Although these two targets are easy to identify errors or missing recognition in the case of occlusion, the model in this paper is mainly used to identify the front taillights of vehicles in the same lane, and there is almost no occlusion.

V. CONCLUSION

Aiming at the problems in the recognition process, such as small light signal from the rear of the front vehicle, easy to be affected by other lights, unbalanced samples, and easy misrecognition caused by too close distance between targets, we proposed an improved target recognition model based on YOLOv5s.

(1) We first add CA to the backbone network to effectively improve the model's ability to identify small targets and improve other optical interference problems. This is mainly due to its ability to accurately locate areas of interest;

(2) Replacing the CIOU Loss of the original YOLOv5s with EIOU Loss effectively alleviates the problem of sample imbalance, improves the training speed and recognition ability of the model;

(3) EIOU-NMS is a modification we made based on DIOU-NMS. Compared with the default NMS used by the author in YOLOv5s, the anchor box error suppression problem has been greatly improved, and the recognition accuracy has been improved.

The algorithm improves the feature extraction module, bounding box regression loss function and anchor box suppression process to ensure the recognition capability of the model. The experimental results show that the mAP value of the algorithm reaches 95.3%, 9.2% higher than the original YOLOv5s mAP value, which greatly improves the recognition accuracy. In the future, we will improve the recognition speed and study the recognition of occluded objects without sacrificing the recognition accuracy to obtain better recognition results.

REFERENCES

[1] S. Garg, K. Kaur, S. H. Ahmed, A. Bradai, G. Kaddoum, and M. Atiquzzaman, "MobQoS: Mobility-aware and QoS-driven SDN framework for autonomous vehicles," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 12–20, Aug. 2019, doi: 10.1109/MWC.2019.1800521.

[2] W. Chen, T. Liao, Z. Li, H. Lin, H. Xue, L. Zhang, J. Guo, and Z. Cao, "Using FTOC to track shuttlecock for the badminton robot," *Neurocomputing*, vol. 334, pp. 182–196, Mar. 2019, doi: 10.1016/j.neucom.2019.01.023.

[3] Z. Cui, S. Yang, and H. Tsai, "A vision-based hierarchical framework for autonomous front-vehicle taillights detection and signal recognition," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 931–937, doi: 10.1109/ITSC.2015.156.

[4] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2241–2248, doi: 10.1109/CVPR.2010.5539906.

[5] A. Almagambetov, S. Velipasalar, and M. Casares, "Robust and computationally lightweight autonomous tracking of vehicle taillights and signal detection by embedded smart cameras," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3732–3741, Jun. 2015, doi: 10.1109/TIE.2015.2400420.

[6] G. Zhong, Y.-H. Tsai, Y.-T. Chen, X. Mei, D. Prokhorov, M. James, and M.-H. Yang, "Learning to tell brake lights with convolutional features," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1558–1563, doi: 10.1109/ITSC.2016.7795765.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[8] J. Guo, Y. Liu, Q. Qiu, J. Huang, C. Liu, Z. Cao, and Y. Chen, "A novel robotic guidance system with eye gaze tracking control for needle based interventions," *IEEE Trans. Cogn. Developmental Syst.*, vol. 13, no. 1, pp. 179–188, Dec. 2020, doi: 10.1109/TCDS.2019.2959071.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 91–99, doi: 10.1109/TPAMI.2016.2577031.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.

[12] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, CL, USA, Dec. 2015, pp. 1440–1448, doi: 10.1109/ICCV.2015.169.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[14] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[15] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Apr. 2018, pp. 1–6.

[16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Apr. 2020, pp. 1–17.

[17] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[18] D. Nava, G. Panzani, and S. M. Savaresi, "A collision warning oriented brake lights detection and classification algorithm based on a mono camera sensor," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 319–324, doi: 10.1109/ITSC.2019.8916961.

[19] F. I. Vancea and S. Nedevschi, "Semantic information based vehicle relative orientation and taillight detection," in *Proc. IEEE 14th Int. Conf. Intell. Comput. Commun. Process. (ICCP)*, Sep. 2018, pp. 259–264, doi: 10.1109/ICCP.2018.8516631.

[20] Y. Li and F. Yu, "CDMY: A lightweight object detection model based on coordinate attention," in *Proc. IEEE 10th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, vol. 10, Chongqing, China, Jun. 2022, pp. 1258–1263, doi: 10.1109/ITAIC54216.2022.9836670.

[21] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717, doi: 10.1109/CVPR46437.2021.01350.

[22] Z. Zhang, Z. Deng, Z. Wu, and G. Lai, "An improved EIoU-YOLOv5 algorithm for blood cell detection and counting," in *Proc. 5th Int. Conf. Pattern Recognit. Artif. Intell. (PRAI)*, Chengdu, China, Aug. 2022, pp. 989–993, doi: 10.1109/PRAI55851.2022.9904093.

[23] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient iou loss for accurate bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Jan. 2021, pp. 146–157.

[24] Q. Li, S. Garg, J. Nie, X. Li, R. W. Liu, Z. Cao, and M. S. Hossain, "A highly efficient vehicle taillight detection approach based on deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4716–4726, Jul. 2021, doi: 10.1109/TITS.2020.3027421.

[25] W. Song, S. Liu, T. Zhang, Y. Yang, and M. Fu, "Action-state joint learning-based vehicle taillight recognition in diverse actual traffic scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18088–18099, Oct. 2022, doi: 10.1109/TITS.2022.3160501.

[26] J. Pirhonen, R. Ojala, K. Kivekäs, J. Vepsäläinen, and K. Tammi, "Brake light detection algorithm for predictive braking," *Appl. Sci.*, vol. 12, no. 6, p. 2804, Mar. 2022, doi: 10.3390/app12062804.

[27] S. Iftikhar, M. Asim, Z. Zhang, and A. A. A. El-Latif, "Advance generalization technique through 3D CNN to overcome the false positives pedestrian in autonomous vehicles," *Telecommun. Syst.*, vol. 80, no. 4, pp. 545–557, Aug. 2022, doi: 10.1007/s11235-022-00930-1.

[28] S. Iftikhar, Z. Zhang, M. Asim, A. Muthanna, A. Koucheryavy, and A. A. Abd El-Latif, "Deep learning-based pedestrian detection in autonomous vehicles: Substantial issues and challenges," *Electronics*, vol. 11, no. 21, p. 3551, Oct. 2022, doi: 10.3390/electronics11213551.

[29] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, "WiFi CSI based passive human activity recognition using attention based BLSTM," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2714–2724, Nov. 2019, doi: 10.1109/TMC.2018.2878233.

[30] J.-G. Wang, L. Zhou, Y. Pan, S. Lee, Z. Song, B. S. Han, and V. B. Saputra, "Appearance-based brake-lights recognition using deep learning and vehicle detection," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 815–820, doi: 10.1109/IVS.2016.7535481.

[31] J. Wang, L. Zhou, Z. Song, and M. Yuan, "Real-time vehicle signal lights recognition with HDR camera," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Dec. 2016, pp. 355–358, doi: 10.1109/iThings-GreenCom-CPSCom-SmartData.2016.84.

[32] H.-T. Chen, Y.-C. Wu, and C.-C. Hsu, "Daytime preceding vehicle brake light detection using monocular vision," *IEEE Sensors J.*, vol. 16, no. 1, pp. 120–131, Jan. 2016, doi: 10.1109/JSEN.2015.2477412.

[33] K. She, G. Bebis, H. Gu, and R. Miller, "Vehicle tracking using online fusion of color and shape features," in *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, Washington, DC, USA, Mar. 2004, pp. 731–736, doi: 10.1109/ITSC.2004.1398993.

[34] Y.-M. Chan, S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao, "Vehicle detection under various lighting conditions by incorporating particle filter," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2007, pp. 534–539, doi: 10.1109/ITSC.2007.4357745.

[35] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 453–462, Jun. 2010, doi: 10.1109/TITS.2010.2045375.

[36] M. Casares, A. Almagambetov, and S. Velipasalar, "A robust algorithm for the detection of vehicle turn signals and brake lights," in *Proc. IEEE 9th Int. Conf. Adv. Video Signal-Based Surveill.*, Sep. 2012, pp. 386–391, doi: 10.1109/AVSS.2012.2.

[37] A. Almagambetov, M. Casares, and S. Velipasalar, "Autonomous tracking of vehicle rear lights and detection of brakes and turn signals," in *Proc. IEEE Symp. Comput. Intell. Secur. Defence Appl.*, Jul. 2012, pp. 1–7, doi: 10.1109/CISDA.2012.6291543.

[38] D. Koller, J. Weber, and J. Malik, "Robust multiple car tracking with occlusion reasoning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Berlin, Germany, Jan. 1994, pp. 189–196, doi: 10.1007/3-540-57956-7_22.

[39] P. Thammakaroon and P. Tangamchit, "Predictive brake warning at night using taillight characteristic," in *Proc. IEEE Int. Symp. Ind. Electron.*, Seoul, South Korea, Jul. 2009, pp. 217–221, doi: 10.1109/ISIE.2009.5218254.

[40] Q. Ming and K.-H. Jo, "Vehicle detection using tail light segmentation," in *Proc. 6th Int. Forum Strategic Technol.*, vol. 2, Aug. 2011, pp. 729–732, doi: 10.1109/IFOST.2011.6021126.

[41] S. Nagumo, H. Hasegawa, and N. Okamoto, "Extraction of forward vehicles by front-mounted camera using brightness information," in *Proc. IEEE Can. Conf. Elec. Comput. Eng. (CCECE)*, vol. 2, Montreal, QC, Canada, Sep. 2003, pp. 1243–1246, doi: 10.1109/CCECE.2003.1226124.

[42] Z. Cao, T. Liao, W. Song, Z. Chen, and C. Li, "Detecting the shuttlecock for a badminton robot: A YOLO based approach," *Exp. Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 113833, doi: 10.1016/j.eswa.2020.113833.

[43] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, Nov. 2019, pp. 12993–13000.

[44] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.

[45] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[46] Y. Jiahui, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "UnitBox: An Advanced Object Detection Network," in *Proc. 24th ACM Int. Conf. Multimedia*. New York, NY, USA: Association for Computing Machinery, pp. 516–520, 2016, doi: 10.1145/2964284.2967274.

HUAYUE ZHANG was born in Shandong, China. He received the bachelor's degree in mechanical design, manufacturing and automation from Qingdao Agricultural University (Haidu University), Qingdao, China, in 2021. He is currently pursuing the master's degree in transportation. His research interests include deep learning and computer vision, especially traffic image recognition.

JUNYOU ZHANG was born in Zibo, China. He received the Ph.D. degree from China Agricultural University, Beijing, China, in 2005. His research interests include transportation planning and management, intelligent transportation, autonomous vehicle control decisions, and traffic safety.

SHUFENG WANG received the Ph.D. degree from China Agricultural University, Beijing, China, in 2002. She is currently an Associate Professor with the Shandong University of Science and Technology. Her research interests include image processing, self-driving vehicle, and intelligent transportation.

QIAN ZHOU was born in Shandong, China. She received the B.S. degree in industrial engineering from the Shandong University of Science and Technology, Qingdao, China, in 2020, where she is currently pursuing the master's degree in transportation planning and management. Her research interests include deep learning and computer vision, especially traffic image recognition.

XIAOLEI LI was born in Dezhou, China. He received the B.S. degree in logistics engineering from the Shandong University of Science and Technology, Qingdao, China, in 2020, where he is currently pursuing the master's degree in transportation planning and management. His research interests include deep learning and computer vision, especially human action analysis.