Received 10 December 2023, accepted 25 December 2023, date of publication 28 December 2023,
date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3347791

RESEARCH ARTICLE

Nighttime Vehicle Detection Algorithm Based on Improved Faster-RCNN

YAQING XU[1], **KAIBIN CHU**[2], **AND JI ZHANG**[1]
[1]School of Computer and Artificial Intelligence, Changzhou University, Changzhou, Jiangsu Province 213164, China
[2]School of Microelectronics and Control, Changzhou University, Changzhou, Jiangsu Province 213164, China
Corresponding author: Kaibin Chu (ckb910@cczu.edu.cn)

ABSTRACT Vehicle detection is important for the development of Intelligent Transportation Systems (ITS), which has made great strides in recent years. However, at night, vehicle detection faces many difficulties such as low illumination, street lights, and the appearances vehicle headlights, etc. In order to solve these problems, we propose an improved nighttime vehicle detection algorithm based on Faster R-CNN. Firstly, we combine the Deformable Convolutional Network with Faster R-CNN to improve the detection accuracy features of night vehicles of different sizes and shapes. Secondly, to improve the prediction accuracy of bounding box position information, we adopt Side-Aware Boundary Localization to replace the traditional bounding box prediction. It can further obtain more accurate position information. At the same time, aiming at the imbalance of samples in the training process, we use Oline Hard Example Mining(OHEM) to train samples with a high probability of error to improve the learning effect of a few classes; and to improve the accuracy of night vehicle detection. In addition, we use Soft Non-Maximum Suppression(Soft-NMS) to reduce the number of missed vehicles. The improved algorithm efficiently improves the night vehicle detection accuracy and reduces the model complexity. Furthermore, we verify the effectiveness of each innovation module through ablation experiments and comparison experiments. Finally, the advantages of the improved model in terms of nighttime vehicle detection accuracy are verified by experimenting on the open-source intelligent traffic dataset UA-DETRAC and the open-source diverse automated driving dataset BDD100K.

INDEX TERMS Nighttime vehicle detection, faster R-CNN, deformable convolutional network (DCNN), side-aware boundary localization (SABL), intelligent transportation system (ITS).

I. INTRODUCTION

Nighttime vehicle detection plays a crucial role in the Intelligent Transport System (ITS) [1], which has achieved remarkable achievements. Nighttime vehicle detection helps to monitor the condition of vehicles on the road section, detect the abnormal behaviour and illegal operation of the vehicles in time, and improve the level of road traffic safety. The number and speed of vehicles can be monitored in real time through night vehicle detection to optimize traffic flow prediction [2] and reduce traffic congestion.

However, the accuracy of vehicle detection at night is unsatisfactory. The main problems and challenges of night vehicle detection are as follows: (1) the quality of vehicle images taken under the condition of low light at night is poor, and it is difficult to detect vehicle features such as vehicle details, color, and shape; (2) there may be noise in night images taken under the condition of low light at night, which will confuse the detection algorithm with the background, increasing false positive and false negative samples; (3) vehicle lights such as headlights and taillights will produce bright areas in the image, which will blur vehicle boundaries or cause vehicle false detection. Therefore, in order to improve the accuracy of night vehicle detection to improve night traffic safety, optimize intelligent traffic management and promote the development of intelligent transportation system. We proposes an improved night vehicle detection model based on Faster R-CNN.

The main contributions of this paper are as follows: (1) in order to improve the detection accuracy of night vehicle features of different sizes and shapes, Deformable Convolutional Network [3] network combined with Faster

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Shorif Uddin.

R-CNN adaptively adjusts the shape and size of the receptor field, so that more vehicle feature information can be captured to improve the detection accuracy; (2) using edge sensing boundary positioning instead of traditional boundary box prediction can better capture vehicle body contour, window edge and other details, more accurately locate the target boundary, and improve the robustness and accuracy of the model; (3) to solve the problem of unbalanced class samples in the training stage, OHEM [4] was adopted to focus on training samples with high error probability, improve the learning effect of a few classes, reduce the false detection rate of the model, and improve the training speed and detection performance of the model.

II. RELATED WORK

Traditional machine learning algorithms start with the extraction of object candidate boxes by sliding windows. Then it depends on a convolutional neural network to extract the relevant features of the region proposals, such as Harr features [5], SIFT features [6], and HOG features [7]. Finally, classifiers such as Support Vector Machines (SVM) [8] and Adaboost [9] are used to predict the presence of the object within regions and to recognize object categories. R. Girshick et al. introduced a deep convolutional network to object detection in 2014 by proposing the Regions with CNN features (RCNN) [10]. Two-stage object detection algorithms represented by RCNN, Fast-RCNN [11], Faster R-CNN [12], Mask-RCNN [13]. One-stage object detection algorithms represented by SSD [14], RetinaNet [15], YOLOv3 [16], YOLO9000 [17], YOLOv5 [18].

In recent years, many researchers have proposed many methods to solve the problem of detecting vehicles at night. Cui et al. [19] proposed a self-supervised learning based multi-task automatic transformation (MAET) model to improve the object detection accuracy in dark environments. However, due to the limited number of dark target samples and insufficient image illumination, the multi-task automatic transformation MAET model cannot learn enough about the dark target and cannot effectively extract the key features of the dark target. In addition, the MAET model needs to handle multiple tasks at the same time, which increases the training time and computational resource requirements of the model. Yin et al. [20] proposed the combination of pyramid network and YOLOv3 for dark object target detection. However, the model using the image pagoda network is more complex and will consume a lot of computational resources. In addition, the robustness of the model is limited when facing dark object detection in different environments. Zhang et al. [21] proposed a dark target detection model based on the transformer network structure for low-light illumination. However, the dual-trunk transformer model cannot effectively adapt to the complex low-light background effectively, and it is easy to produce false detection or missing detection. Moreover, there are obvious differences between the images in the extremely dark environment and the images in the routine daytime scene, which will limit the generalization ability of the model and reduce the performance in practical applications.

Therefore, this paper studies and improves the nighttime vehicle detection algorithm based on the two-stage detection algorithm Faster R-CNN, which improves the detection accuracy of night vehicles at night.

III. METHOD

In this section, we elaborate on the proposed method of nighttime vehicle detection. The overall architecture of the proposed method is shown in Figure 1, which is based on the Faster R-CNN [16] algorithm.

A. DEFORMABLE CONVOLUTIONAL NETWORK

1) DEFORMABLE CONVOLUTION

The mesh size of the traditional convolution kernel is preset, and equation (1) is the definition of the traditional convolution structure. Each point of the output feature map corresponds to the center point of the convolution kernel. x represents the features of the input, y represents the features of the output, and P_n represents the offset of P_0 within the convolution kernel domain.

$$y(P_0) = \sum_{P_n \varepsilon R} w(P_n).x(P_0 + P_n) \tag{1}$$

However, in the environment of low light and limited visibility at night, the vehicle shape will produce nonlinear deformation and complex deformation. Deformable convolution adjusts the position of the sampling points in the convolution operation by introducing a learnable offset at each point on the input feature x, which makes the model better adapt to the vehicle deformation. Equation (2) is a deformable convolution formula with a learnable offset:

$$y(P_0) = \sum_{P_n \varepsilon R} w(P_n).x(P_0 + P_n + \Delta P_n) \tag{2}$$

Deformable convolution adaptively adjusts the shape and size of the receptive field according to the scale of the vehicle in the input image, and the model can capture the feature information from vehicles of different scales at night. The edge information of vehicles at night is fuzzy and the contrast is low, so it is difficult to extract the boundary information of vehicles accurately. By adjusting the position and the shape of the convolution sampling points, deformable convolution can capture the boundary features of vehicles more accurately and improve the detection accuracy of the night vehicle detection model. In addition, there is a lot of background noise and other interfering objects in the night image. The deformable convolution reduces the interference of noise and objects by introducing a learnable offset and receptive field and reduces the false detection rate of the model.

Figure 2 shows the implementation of a deformable convolution.

Therefore, as shown in Figure 1, we replace the 3 × 3 convolution of layers 3 to 5 of the backbone network ResNet50 [22] with a deformable convolution. The replaced

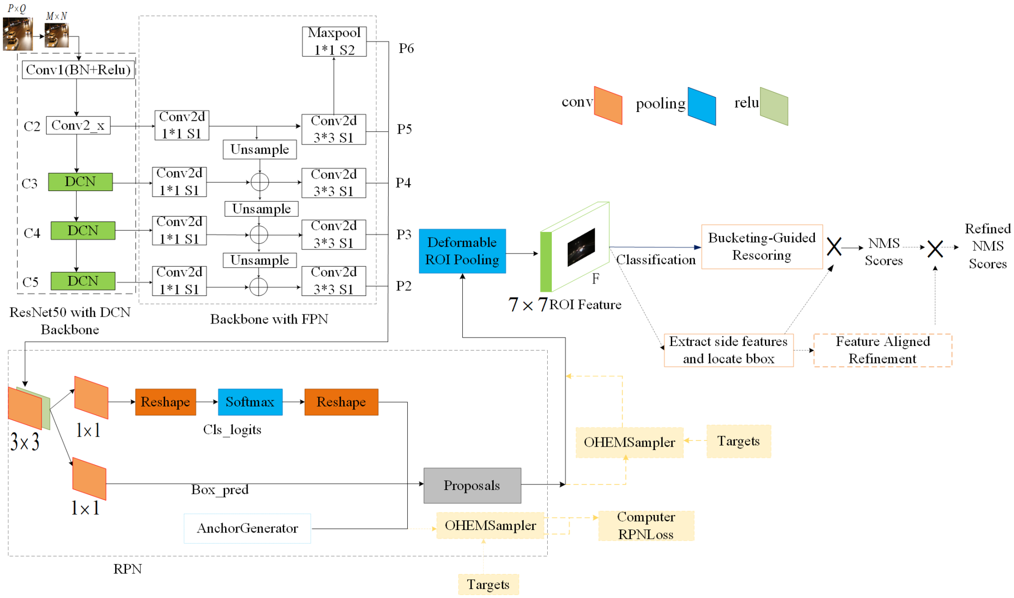


FIGURE 1. The overall framework of the proposed method.

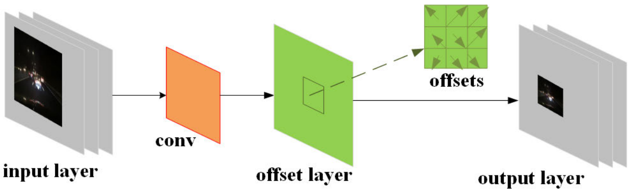


FIGURE 2. The structure of deformable convolution.

convolution can learn an offset for each position of the input feature, and can accurately position various deformed objects, thus improving the detection accuracy of the model.

2) DEFORMABLE ROI POOLING

For the given input feature map and region of size $w \times h$, the RoI pooling will divide the feature map x into $k \times k$ bins, and outputs the feature map y of size $k \times k$, as shown in the following equation (3):

$$y(i,j) = \sum_{p\varepsilon bin(i,j)} x(P_0 + p)/n_{ij} \tag{3}$$

where n_{ij} is the number of pixels in the bin.

Compared to traditional ROI pooling, deformable pooling adds an offset, which can be seen in the following equation (4):

$$y(i,j) = \sum_{p\varepsilon bin(i,j)} x(P_0 + p + \Delta P_{ij})/n_{ij} \tag{4}$$

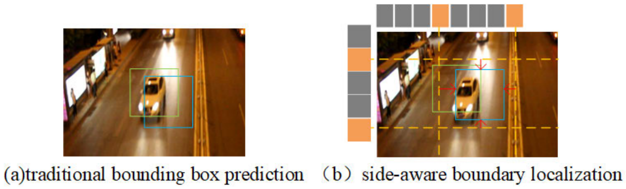
In the top path of deformable pooling, the feature map is still generated by conventional RoI pooling. Then, a fully connected layer generates a normalized offset ΔP_{ij}, which makes the offset learning independent of the RoI size. Finally, deformable convolutional pooling is implemented in the bottom path to output the rectangular region as a feature map with improved offsets. In night vehicle detection, the night light is weak and the field of view is limited, the vehicle will undergo non-rigid deformation, and the detailed information such as the texture, color, and logo of the vehicle is difficult to extract. Deformable pooling can learn offsets and flexibly adjust sampling positions to better adapt to vehicle deformations and capture vehicle details. In addition, deformable pooling can adapt to different vehicle sizes by adjusting the sampling position and size. Therefore, we replace the RoI Pooling in Faster R-CNN with Deformable RoI pooling in Figure 1. In this way, the pooling operation can be deformed according to the shape of different targets, which can improve the adaptability to the receptive field, to improve the accuracy and robustness of the night vehicle detection model.

B. SIDE-AWARE BOUNDARY LOCALIZATION

The mainstream method of bounding box prediction is shown in Figure 3. The green box indicates a suggestion box, and the blue box indicates a prediction box. It is based on the offset between the bounding box and the center of the proposal to determine the target position. Figure 3 shows Side-Aware Boundary Localization. To improve positioning accuracy, SABL divides the target space into multiple buckets by calculating each edge of the bounding box. The gray rectangle represents the bucket, and the orange rectangle represents the prediction and estimation of the bucket.

The pipeline of SABL for the Faster R-CNN network is shown in Figure 4. The specific implementation steps are as follows:

First, four boundary features F_{left}, F_{right}, F_{top} and F_{down} are obtained from the ROI feature extracted from rpn by two 1×1 convolution and normalization. Then, for a given proposal box, its boundary is enlarged by a factor of $\sigma\,(\sigma > 1)$



(a)traditional bounding box prediction （b）side-aware boundary localization

FIGURE 3. Bounding box prediction method.

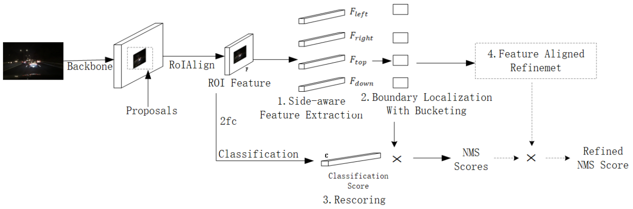


FIGURE 4. Bounding box prediction method.

so that it covers the whole object. The candidate region is divided into 2k bucket regions, with each boundary centerline corresponding to a bucket. Binary classification is used to determine which bucket the bounding box boundary is closest to. Then, the exact regression positioning of the boundary is performed. Finally, the resulting boundary coefficient is multiplied by the value of the classification score as the NMS. The fourth part is the feature reuse module, which aims to reduce the amount of computation.

When the light is insufficient at night, the vehicle's boundary is not clear enough, and the traditional boundary frame cannot accurately locate the target vehicle. By extracting the boundary information of the target vehicle, SABL can more accurately locate the target boundary and better adapt to the non-rigid deformation of the vehicle. The accuracy and robustness of the detection model are improved. By extrapolating and locating the boundary edge of the night vehicle, the details of the body contour, window edge, etc. can be captured, and the detection accuracy of the model can be improved. Therefore, we adopt the SABL to replace the traditional bounding box regression in the Faster-RCNN network. By replacing the way of predicting the bounding box regression with a SABL after output proposals of rpn of Faster R-CNN. Each edge of the bounding box is positioned separately by SABL to improve the accuracy of target boundary positioning.

C. ONLINE HARD EXAMPLE MINING

Vehicle detection performance will decrease in low light and weak light conditions at night. During the training process, OHEM will focus on difficult samples. It can improve the detection ability of the model in difficult scenes such as low light at night, and improve the robustness and accuracy of the model. Due to the lack of light at night, the detailed features of vehicles are difficult to capture and identify, and there are many incorrect samples in training. OHEM will focus on training these false samples to reduce the false detection rate of the model. Aiming at the problem that the selected night vehicle data set is large, which reduces the training efficiency of the detection model. OHEM can eliminate simple samples that are easy to classify, and train limited samples that are more difficult to avoid repeated training. In addition, OHEM selects difficult samples to train during the training process to improve the adaptability of the model in different complex scenarios.

To improve the detection speed and accuracy of Faster R-CNN on night vehicles, OHEM is applied to Faster R-CNN in this paper. OHEM is applied to the classification loss part of rpn, that is to say, all the region proposals are sent dierctly to the ROI pooling when calculating the classification loss of RCNN. OHEM can automatically select some samples with diversity and hard examples for training. It can effectively improve the training speed and detection performance of the model for night vehicle detection.

IV. EXPERIMENTS

A. EXPERIMENTAL SETTING

1) DATASET AND IMPLEMENTATION DETAILS

The UA-DETRAC [23] dataset contains 82085 images in the training set and 56,167 images in the test set. 878 night scene training images and 220 night scene test images are extracted from the training and test sets, respectively. The UA-DETRAC dataset was divided into four categories: car, bus, van, and other. There are 100,000 images in the road target boundary box of the BDD100K [24] dataset. 10,500 night vehicle images are selected from the 70,000 images in the training set as the training set, and 1,500 night vehicle images are selected from the 20,000 images in the test set as the test set. We keep only cars, buses, bicycles, trucks, motorcycles, and trains in six categories.

Our detection models are built on the MMDetection framework, which is a library of target detection tools based on PyTorch. MMDetection includes dozens of advanced models and methods and provides a very suitable high-level interface for researchers to perform secondary development. This makes it easier for us to experiment. We choose ResNet50 with FPN [25] as the backbone of Faster R-CNN. All the experiments were done on a single NVIDIA GeForce RTX3090 GPU.

2) EVALUATION METHOD

We use the following commonly used vehicle detection indicators to evaluate the experimental results.

(1) The accuracy rate P refers to the probability that the positive sample is correctly predicted among all the predicted samples. The calculation equation is as follows:

$$Precision = TP/(TP + FP) \tag{5}$$

(2) Recall ratio R refers to the probability of being correctly predicted as a positive sample among all real positive samples. The calculation equation is as follows:

$$Recall = TP/(TP + FN) \tag{6}$$

(3) Average accuracy (mAP) is the sum of AP values of all categories of vehicles and then calculate the average value. mAP is an indicator to measure the average accuracy of target detection, and its calculation equation is as follows:

$$mAP = \frac{\sum P_{classAve}}{N_{classes}} \tag{7}$$

(4) F1-score is an evaluation index that integrates precision and recall. It is the harmonic average of precision and recall, and its value ranges from 0 to 1. The calculation equation is as follows:

$$F1 - score = \frac{2.precision.recall}{precision + recall} \tag{8}$$

AP is averaged across all categories. AR refers to the maximum recall in a given fixed number of detection results in each image, which is averaged over all IoUs and all categories. TP is the positive sample that was correctly predicted; FP is the positive sample that is incorrectly predicted; TN is the negative sample that was correctly predicted; FN is the negative sample that was incorrectly predicted; $N_{classes}$ is the total number of target classes; $\sum P_{classAve}$ is the sum of the average accuracy of all classes.

B. THE RESULTS OF COMPARISONS

To verify the superiority and validity of the proposed method on the selected BDD100K dataset and UA-DETRAC dataset. We use the most advanced testing methods, including Faster R-CNN, RetinaNet, Cornernet [26], Centernet [27], FCOS [28], and Detr [29]. The learning strategy of Centernet, FCOS, and RetinaNet is 1x. The optimization algorithm is SGD, and the learning rate is set to 0.005, the momentum factor to 0.9, and the weight attenuation factor to 0.0001. The initial learning rate is 0.001, which gradually increases linearly over the first 500 iterations and decreases at the 8th and 11th epochs. In total, 24 epochs were trained. The learning strategy of an SSD is 2x, which means that the learning rate changes to 0.0025. The learning rate decreases in the 16th and 22nd epochs. The optimization algorithm for the Detr model is AdamW, where the learning rate is set to 0.0001 and the total number of training rounds is set to 14. The optimization algorithm of the Cornernet model is Adam, the learning rate is 0.0005, and the total number of training rounds is 24.

TABLE 1. The comparisons of different detection approaches on the selected BDD100K dataset.

Method	mAP	AP@0.75	F1-score	AR	AR@S
Faster R-CNN	32.3	32.4	37.1	43.5	20.8
RetinaNet	32.8	33.0	38.8	47.6	23.8
Centernet	27.7	27.9	34.4	45.4	22.8
Cornernet	25.1	24.8	31.6	42.5	27.8
Detr	21.9	19.5	26.7	34.3	10.1
Ours	**35.4**	**36.5**	**41.6**	**50.5**	**25.1**

As is shown in Table 1, the model Detr based on transformer architecture has the lowest accuracy values and recall rates. Due to the self-attention mechanism of the transformer model, Detr is relatively poorly adapted to complex shapes and unconventional targets. As a result, Detr has poor detection effectiveness when dealing with target vehicles of different sizes or shapes. Compared with Faster R-CNN in terms of detection accuracy, Centernet and Cornernet are not particularly ideal. Because Cornernet is a target detection model specifically designed to detect corner points of objects. For small and medium-sized targets, complex shapes and targets with unconventional shapes, the Cornernet model makes it difficult to correctly detect the corner points of these targets. Its detection effect is far less than the traditional bounding box representation method, so there will be false detection or missing detection. Centernet is a model based on central point detection. There are lots of overlapping or occluded targets in the BDD100K dataset, so the model is not accurate in locating and classifying the center points of such targets. Compared with Detr, Cornernet, and Centernet, RetinaNet realizes the correct detection of small targets by using the focal loss function to solve the problem of category imbalance and difficult detection of small targets in target detection. Meanwhile, multi-scale feature maps and specific regression branches are used to improve the accuracy of target localization and classification. However, compared with our method, the results of RetinaNet are not very satisfactory in all aspects. Our method uses OHEM and SABL to solve the problem of sample class imbalance and border positioning so that the detection accuracy of small targets can be achieved. Our method also uses the deformable convolutional network to solve the problem of poor adaptability of the model to complex shapes and unconventional images and improves the detection accuracy of the model to objects with different shapes. So our method has good performance in night vehicle detection and can improve the detection accuracy of night vehicles. In addition,our method also achieves the maximum F1-score.

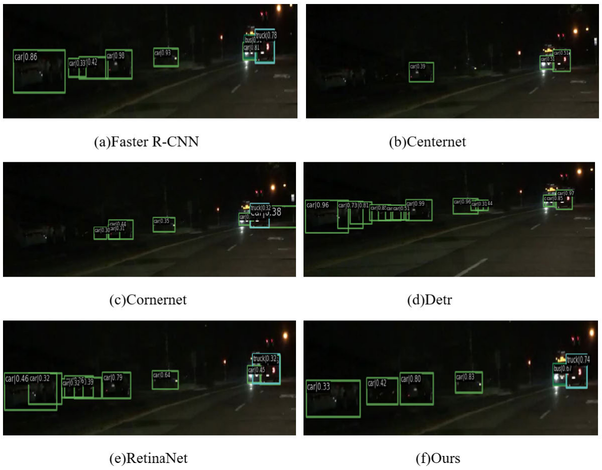
TABLE 2. The comparisons of different detection approaches on the selected UA-DETRAC dataset.

Method	mAP	AP@0.75	F1-score	AP@S	AR@S
Faster R-CNN	80.7	94.1	83.3	57.4	68.4
Detr	65.7	75.2	69.2	27.0	48.7
FCOS	78.1	90.7	80.3	58.4	66.2
RetinaNet	78.9	92.2	80.7	57.1	65.3
SSD	53.7	59.2	57.3	25.1	42.0
Ours	**83.3**	**95.0**	**85.8**	**60.8**	**73.5**

As shown in Table 2, our method achieves the maximum F1-score compared to other models, indicating that our model has the best quality. Compared with other detection methods, RetinaNet obtains the maximum mAP value and the maximum AP@0.75, but the detection performance is poor under small and medium vehicle sizes. The use of multi-scale feature maps by RetinaNet will overlap and block different targets, resulting in poor detection effects on dense targets. FCOS is better than RetinaNet in $AP@S$ and $AP@M$. FCOS can improve the adaptability of targets with different shapes by adaptive prediction of target center point and boundary frame size. Detr is an end-to-end object detection model based on transformer. The performance of Detr is not very good compared to other methods. However, our method shows the best performance in all aspects, especially in the detection performance of small and medium objects. As shown in Table 2, the best performance can be obtained when these three modules are applied to the baseline.

C. QUALITATIVE ANALYSIS

To verify the validity of the proposed method, we performed visual comparisons on the selected BDD100K dataset and the selected UA-DETRAC dataset. Some representative scenarios are selected for detection. Figure 5 and Figure 6 show the detection results of dataset BDD100K, Figure 7 and Figure 8 show the detection results of dataset UA-DETRAC. The number on the outside of the target detection box indicates confidence. A higher confidence level indicates the model's estimate of the sample mean and the population mean. The closer the confidence level is to 1, the more accurately the model estimates the sample mean and the population means.



(a)Faster R-CNN

(b)Centernet

(c)Cornernet

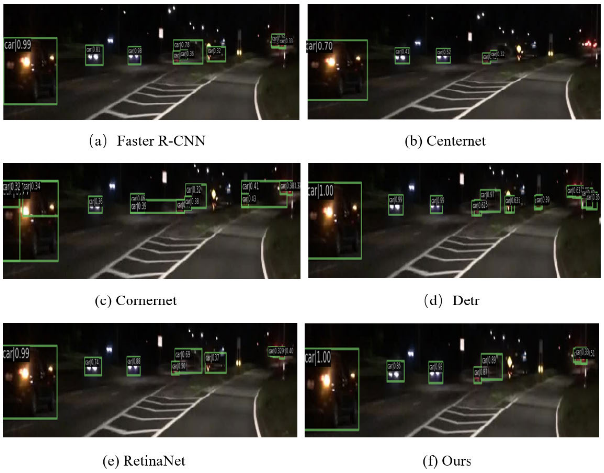
(d)Detr

(e)RetinaNet

(f)Ours

FIGURE 5. Detection results of dark scene from BDD100K. (a) Faster R-CNN, (b) Centernet, (c) Cornernet, (d) Detr, (e) RetinaNet, (f) Ours.

Figure 5 is an extremely dark scene with small dark targets at a distance. Figure 6 is a relatively dark scene with interference from street lights and car lights. As shown in Figure 5, there are redundant detection cases on the left side of Figure 5(a), and there are false detections and more detection cases on the right side of the distant dark small target. The left side of Figure 5(b) has a significant missing detection. Figure 5(c) also has missed and false checks. Figure 5(d) and (e) have a large number of redundant tests. Figure 5(f) shows the detection results of our proposed model. The proposed model can accurately detect vehicles. As shown in Figure 6, Figure 6(a), Figure 6(c), Figure 6(d) and Figure 6(e) all have cases of false detection and redundancy detection, while Figure 6(b) has cases of missing detection. Figure 6(f) shows the detection results of our proposed model, which achieves accurate detection and has the highest confidence.



(a) Faster R-CNN

(b) Centernet

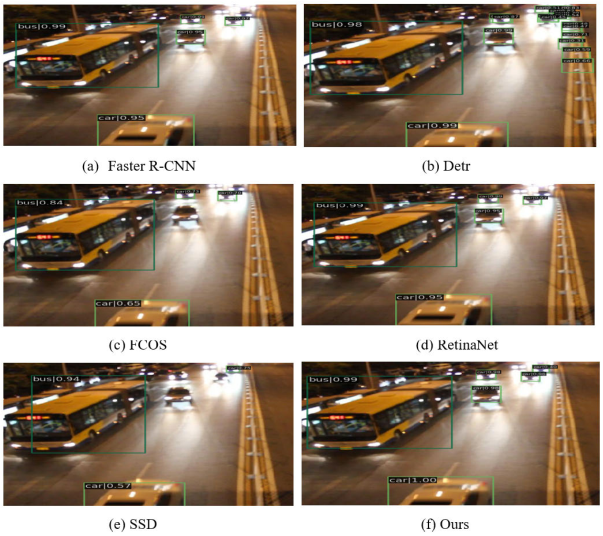
(c) Cornernet

(d) Detr

(e) RetinaNet

(f) Ours

FIGURE 6. Detection results of dark scene from BDD100K. (a) Faster R-CNN, (b) Centernet, (c) Cornernet, (d) Detr, (e) RetinaNet, (f) Ours.



(a) Faster R-CNN

(b) Detr

(c) FCOS

(d) RetinaNet

(e) SSD

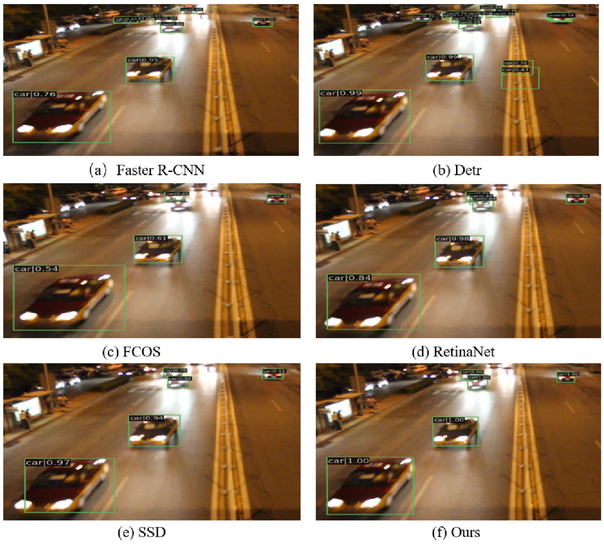
(f) Ours

FIGURE 7. Detection results of the headlight scene from UA-DETRAC. (a) Faster R-CNN, (b) Detr, (c) FCOS, (d) RetinaNet, (e) SSD, (f) Ours.

Figure 7 is the scene of the vehicle's headlights refracting on the road. Figure 8 shows the scene of a small target vehicle with headlights at a distance. As shown in Figure 7(a), Figure 7(c), Figure 7(d), and Figure 7(e) all have a lack detection due to the influence of vehicle's headlights. Figure 8(b) shows a large number of false detections. Figure 8(f) shows the detection results of our proposed model, which can be accurately detected and has the highest confidence. As shown in Figure 8, there are redundancy detection and false detection in Figure 8(a), Figure 8(b) and Figure 8(d). In Figure 8(c), there is a case of missing detection. Although the vehicle is detected in Figure 8(e), its confidence is not as high as that in Figure 8(f). Our proposed method not only accurately detects vehicles, but also has a confidence level of 1 for each.

D. ABLATION STUDY

To investigate the contribution of different components to the overall network, we performed ablation experiments



(a) Faster R-CNN (b) Detr

(c) FCOS (d) RetinaNet

(e) SSD (f) Ours

FIGURE 8. Detection results of the headlight scene from UA-DETRAC. (a) Faster R-CNN, (b) Detr, (c) FCOS, (d) RetinaNet, (e) SSD, (f) Ours.

on the selected dataset BDD100K and the selected dataset UA-DETRAC. We use Faster R-CNN as the baseline, which uses ResNet50 with FPN as the backbone. We apply DCNN, SABL, and OHEM to the baseline respectively, and the performance is shown in Table 3 and Table 4.

TABLE 3. Ablation results of each module on the selected BDD100K dataset.

Baseline	DCNN	SABL	OHEM	mAP	AP@0.75	AR
✓				32.3	32.4	43.5
✓	✓	✓		35.3	36.8	51.5
✓	✓		✓	34.5	35.6	50.4
✓		✓	✓	34.3	35.8	49.9
✓	✓	✓	✓	**35.4**	**36.5**	**50.5**

TABLE 4. Ablation results of each module on the selected UA-DETRAC dataset.

Baseline	DCNN	SABL	OHEM	mAP	AP@0.75	AR
✓				80.7	94.1	86.1
✓	✓	✓		82.8	93.5	87.6
✓	✓		✓	82.2	94.6	87.9
✓		✓	✓	82.0	92.9	86.0
✓	✓	✓	✓	**83.3**	**95.0**	**88.4**

As is shown in Table 3, it can be seen that DCNN improves the baseline. Firstly, this paper combines the DCNN with Faster R-CNN to improve the ability of the model to extract night vehicle features of different sizes and shapes. ResNet50 with FPN is introduced to perform multi-scale feature fusion of night feature mapping. By replacing the traditional bounding box regression with SABL, the accuracy of the bounding box can be improved. As a result, a 3.0% improvement in *mAP* and a 4.4% improvement in *AP*@0.75 improvement are obtained in Table 3, illustrating the advantage of adding the two components. Similarly, *mAP* increased by 2.1% and *AR* increased by 1.5% in Table 4. Secondly, we combine DCNN with Faster R-CNN, and we also use OHEM to train samples with high error probability to improve the learning effect of a few classes, to improve the night vehicle detection accuracy. As a result, a 2.2% improvement in *mAP* and a 3.2% improvement in *AP*@0.75 are obtained in Table 3, which illustrates the advantage of adding the two components. In the same way, *mAP* is improved by 1.5% and *AR* is improved by 1.8% in Table 4.

Thirdly, we apply the SABL to the baseline to improve the detection accuracy of vehicle targets. To improve the detection speed and accuracy of Faster R-CNN on night vehicles, OHEM is applied to baseline in this paper. As a result, a 2.0% improvement in *mAP* and a 3.4% improvement in *AP*@0.75 are obtained in Table 3, which illustrates the advantage of adding the two components. Meanwhile, *mAP* increased by 1.3% in Table 4. Finally, we add these three modules to the baseline. In Table 3, these three modules can effectively improve *mAP* by 3.1% and *AP*@0.75 by 4.1% compared to the baseline. In Table 4, these three modules can effectively improve *mAP* by 2.6% and *AR* by 2.3% compared to the baseline. In conclusion, different components contribute to the improvement of detection accuracy.

V. CONCLUSION

This paper proposes an improved nighttime vehicle detection method based on Faster-RCNN, which can obtain accurate vehicle detection results. It can be seen from Table 1 that our method has excellent detection performance on the BDD100K dataset. Compared to baseline, RetinaNet, Centernet, Cornernet, and Detr, the detection accuracy of our method was improved by 3.1%, 2.6%, 7.7%, 10.3%, and 13.5%, respectively. The recall rate increased by 7%, 2.9%, 5.1%, 8%, and 16.2%, respectively. The recall rate increases by 7%, 2.9%, 5.1%, 8%, and 16.2%, respectively. As shown in Table 2, the detection performance of our method on the UA-DETRAC dataset is also excellent. Compared to basline, Detr, FCOS, RetinaNet, and SSD, the detection accuracy of our method increased by 2.6%, 17.6%, 5.2%, 4.4%, and 29.6%, respectively. The recall rate increased by 7.4%, 13.5%, 8.3%, 7.2%, and 18.2%, respectively. By integrating the three modules into the baseline, the method can effectively improve the detection accuracy of night-time target vehicles and partially occluded vehicles photographed at a distance. Compared with other advanced detection methods, our method has strong advantages.

In future work, the existing night vehicle datasets are small in scale and single in scene. We plan to build a larger, more diverse set of scenarios and a more refined vehicle annotation dataset. The real word scene is complex and diverse, there are some extremely dark and distant night scenes such as almost no light, and it is difficult to achieve correct and fast detection. This paper plans to use infrared images for night vehicle detection in the next step to further improve the detection performance and robustness of the model. To improve the detection performance of the model in snow, fog, rain, and other complex environments.

REFERENCES

[1] L. Figueiredo, I. Jesus, J. T. Machado, J. R. Ferreira, and J. M. De Carvalho, "Towards the development of intelligent transportation systems," in *Proc. IEEE Intell. Transp. Syst. (ITSC)*, Aug. 2001, pp. 1206–1211.

[2] J. Azimjonov and A. Özmen, "A real-time vehicle detection and a novel vehicle tracking systems for estimating and monitoring traffic flow on highways," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101393.

[3] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.

[4] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 761–769.

[5] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[6] L. David, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.

[8] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, p. 1.

[10] 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.

[11] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1137–1149.

[13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.

[15] T.-Y. 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.

[16] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

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

[18] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 2778–2788.

[19] Z. Cui, G.-J. Qi, L. Gu, S. You, Z. Zhang, and T. Harada, "Multitask AET with orthogonal tangent regularity for dark object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2533–2542.

[20] X. Yin, Z. Yu, Z. Fei, W. Lv, and X. Gao, "Pe-YOLO: Pyramid enhancement network for dark object detection," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2023, pp. 163–174.

[21] B. Zhang, J. Suo, and Q. Dai, "A complementary dual-backbone transformer extracting and fusing weak cues for object detection in extremely dark videos," *Inf. Fusion*, vol. 97, Sep. 2023, Art. no. 101822.

[22] 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.

[23] L. Wen, D. Du, Z. Cai, Z. Lei, M.-C. Chang, H. Qi, J. Lim, M.-H. Yang, and S. Lyu, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Comput. Vis. Image Understand.*, vol. 193, Apr. 2020, Art. no. 102907.

[24] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.

[25] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[26] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 734–750.

[27] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6568–6577.

[28] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9626–9635.

[29] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," 2020, *arXiv:2010.04159*.

YAQING XU received the master's degree in computer science and technology (majoring in computer vision) from Changzhou University.

KAIBIN CHU received the B.E. degree in applied electronic technology from the Department of Electronics, Jiangsu Institute of Technology, in July 1997, and the M.E. degree in materials from the School of Materials and Engineering, Changzhou University, in May 2014.

He was a Teaching Assistant with the Jiangsu Commercial Technical School, from August 1997 to February 2000, and has been a Lecturer with Changzhou University, since February 2000. He has published papers, such as 1). Dsp-Based Battery Capacity Performance Tester Design, Power Technology, in 2012, Peking University Core; 2). Design of High-Precision Continuously Adjustable High-Voltage Switching Power Supply, Microcomputer and Application, in 2012, Peking University Core; and 3). Development of Switching Power Supply with Precision Continuously Adjustable High Voltage, ICEEP2012, International Conference. His current project direction is the design of electronic instruments. The projects that have been completed or/are being completed are: the research of automatic test technology of batteries, the research of voltage test equipment, and the research of impedance test projects.

JI ZHANG received the B.Sc. degree in information and computer science from the School of Science, Nanjing University of Science and Technology, in July 2004, and the M.E. degree in pattern recognition and intelligent control (major) from the School of Computer, Nanjing University of Science and Technology, in July 2006.

He was an Assistant Professor (July 2006–October 2009), a Lecturer (October 2009–September 2021), and an Associate Professor (since October 2021) with Changzhou University. His publications are: 1). J. Zhang, H.-Y. Wang, and F.-H. Chen, "Efficient Euclidean local-structural based sparse coding for robust visual tracking," in *Proc. IScIDE*, in Lecture Notes in Computer Science, vol. 8261, 2013, pp. 561–569.(EI); 2). J. Zhang, H.-Y. Wang, and F. Chen, "An improved Fisher discriminant dictionary learning for video object tracking," in *Proc. IScIDE*, in Lecture Notes in Computer Science, vol. 7751, 2013, pp. 700–710.(EI); and 3). J. Zhang and W. Hong-Yuan, "Fast sparse coding tracking algorithm for video targets," *J. Comput. Sci. Explor.*, vol. 6, no. 8, pp. 760–768, 2012. His main research areas are pattern recognition, image processing, and computer vision.