Journal of Intelligent and Connected Vehicles
2023, **6**(3): 173–182
https://doi.org/10.26599/JICV.2023.9210022

Research Article

A deep learning method for traffic light status recognition

Lan Yang, Zeyu He✉, Xiangmo Zhao, Shan Fang, Jiaqi Yuan, Yixu He, Shijie Li, Songyan Liu

School of Information Engineering, Chang'an University, Xi'an 710064, China

Received: July 20, 2023; Revised: September 23, 2023; Accepted: October 10, 2023

ABSTRACT: Real-time and accurate traffic light status recognition can provide reliable data support for autonomous vehicle decision-making and control systems. To address potential problems such as the minor component of traffic lights in the perceptual domain of visual sensors and the complexity of recognition scenarios, we propose an end-to-end traffic light status recognition method, ResNeSt50-CBAM-DINO (RC-DINO). First, we performed data cleaning on the Tsinghua–Tencent traffic lights (TTTL) and fused it with the Shanghai Jiao Tong University's traffic light dataset (S2TLD) to form a Chinese urban traffic light dataset (CUTLD). Second, we combined residual network with split-attention module-50 (ResNeSt50) and the convolutional block attention module (CBAM) to extract more significant traffic light features. Finally, the proposed RC-DINO and mainstream recognition algorithms were trained and analyzed using CUTLD. The experimental results show that, compared to the original DINO, RC-DINO improved the average precision (AP), AP at intersection over union (IOU) = 0.5 (AP_{50}), AP for small objects (AP_s), average recall (AR), and balanced F score (F1-Score) by 3.1%, 1.6%, 3.4%, 0.9%, and 0.9%, respectively, and had a certain capability to recognize the partially covered traffic light status. The above results indicate that the proposed RC-DINO improved recognition performance and robustness, making it more suitable for traffic light status recognition tasks.

KEYWORDS: traffic light status recognition, autonomous vehicle, detection transformer with improved denoising anchor boxes (DINO), Chinese urban traffic light dataset (CUTLD)

1 Introduction

As China's economy has maintained rapid development and the continuous construction of transportation networks (Liu et al., 2021), the rapid acquisition and efficient utilization of traffic scenario information has emerged as the pivotal factor for the advancement of intelligent transportation systems (Lin et al., 2023a; Liu et al., 2022). The environmental perception system of autonomous vehicles captures and analyzes the status of traffic lights using onboard cameras, which provide reliable data support for the decision-making systems of autonomous vehicles (He et al., 2023; Zong et al., 2022b). This enables autonomous vehicles to perform reasonable speed optimization when approaching signalized intersections and reduces potential collisions, which can fully improve the efficiency and safety of autonomous vehicles (Liu et al., 2023; Zong et al., 2022a). However, several technical difficulties exist in accurately recognizing the status of traffic lights. Firstly, the traffic light images captured by vision sensors are susceptible to multiple factors, such as bad weather, exposure, and obstructions, which affect the imaging quality. Secondly, nocturnal interference at night from similar illuminants such as vehicle tail lights and streetlights. Thirdly, the extremely small percentage of pixels occupied by traffic lights in an image poses a significant challenge for the traffic light status recognition algorithm for feature extraction.

Currently, status-recognition methods for traffic lights include machine-learning and deep-learning-based methods (Zeng et al., 2023). The former method achieves traffic light status recognition by manual feature segmentation of region of interest (ROI) and combining it with classifiers. Although these methods demonstrate high recognition rates in specific scenarios, they often encounter challenges when applied to different scenarios, leading to poor portability. Deep-learning-based traffic light status recognition methods do not require manual feature design, which can significantly improve the recognition accuracy and inference speed. Nevertheless, the majority of current traffic light status recognition algorithms rely on foreign traffic light datasets, such as laboratory for intelligent and safe automobiles (LISA) and la route automatisée (LaRA). Few researchers have utilized domestic traffic light datasets to develop traffic light status recognition algorithms that align with the specific urban traffic characteristics in China. In addition, the recognition accuracy in complex scenarios is not high, and there are instances of false and missed detections. Furthermore, they lack the capability to solve the problem of identifying partially obscured lights.

Therefore, we propose an enhanced algorithm for traffic light status recognition called RC-DINO based on DINO (Zhang et al., 2022a), which uses ResNeSt50 (Zhang et al., 2022b) and combines the convolutional block attention module (CBAM) (Woo et al., 2018) to enhance the key information extraction capability of our algorithm (Mentasti et al., 2023). RC-DINO eliminates the requirement for manually designed features, such as non-maximum suppression, enabling the direct generation of object bounding boxes and category labels. We trained RC-DINO with mainstream traffic light status recognition algorithms on the Chinese urban traffic light dataset (CUTLD). The comparison experiments indicated that the RC-DINO had higher average

✉ Corresponding author.
E-mail: 2021124044@chd.edu.cn

precision (AP), average recall (AR), and balanced F score (F1-Score).

Drawing on existing research, the primary contributions can be summarized as follows:

1) Data cleaning was conducted on the Tsinghua–Tencent traffic lights (TTTL) dataset to filter out instances of traffic signs and relabel instances of traffic lights. It was then merged and filtered with the Shanghai Jiao Tong University's traffic light dataset (S2TLD) to create CUTLD, which is suitable for focused training of traffic light status recognition methods that align with the characteristics of Chinese road traffic lights.

2) An improved end-to-end traffic light status recognition method, RC-DINO, is proposed based on DINO. RC-DINO replaces the original backbone with ResNeSt50 and incorporates convolutional block attention module (CBAM) to improve the capability of the algorithm for feature information extraction.

3) Mainstream object detection models such as faster regions with CNN features (FRCNN) and you only look once-v8 (YOLOv8) were chosen for comparative analysis alongside the original DINO. These models were trained and evaluated based on the CUTLD. When all models achieved their optimal fits, the DINO model demonstrated exceptional performance with all common objects in context (COCO) evaluation metrics reaching their peaks, notably achieving an AP of 67.9% and an AP_{50} of 95.9%, proving that the DINO model has more advantages in traffic light status recognition tasks than traditional one-stage and two-stage object detection models.

4) The enhanced RC-DINO was trained based on the CUTLD and subsequently compared and validated against the first three methods, which exhibited superior performance in the DINO analysis experiments using various metrics. The experimental results clearly demonstrate RC-DINO's superiority, which has the highest AP, AR, and F1-Score. Compared to the original DINO, RC-DINO improved the AP, AP_{50}, AP_S, AR, and F1-Score by 3.1%, 1.6%, 3.4%, 0.9%, and 0.9%, respectively. Furthermore, we performed a horizontal comparison between the RC-DINO and the traffic light status recognition methods proposed by Chen et al. (2021) and Sathiya et al. (2015). The RC-DINO has the advantages of recognizing more statutes, higher recognition accuracy, and a certain capability to recognize a partially covered traffic light status. In summary, the RC-DINO is more suitable for the task of traffic light status recognition.

2 Related works

In machine learning-based methods, the initial step involves pre-segmenting the ROI of traffic lights (Reddy et al., 2023). Subsequently, various features such as signal color, morphology, histogram of oriented gradient (HOG), and scale invariant feature transform (SIFT) were extracted from the ROI. Finally, classifiers such as support vector machine (SVM) and decision tree (DT) are adopted for traffic light status recognition. Lee et al. (2018) adopted an approach involving threshold segmentation and morphologic manipulation to extract potential regions. Subsequently, an SVM was employed to recognize red and green lights. Although this method effectively mitigates the effects of light color variations, it suffers from low detection accuracy and slow processing speed. Gong et al. (2010) classified and tracked traffic lights by segmenting an acquired HSV space using machine learning and CAMSHIFT method. However, this method is susceptible to changes in lighting conditions, which can significantly affect its performance. Omachi and Omachi (2009) normalized the color space and extracted an ROI. They then combined Sobel edge detection and Hough transform (Chen et al., 2021) to detect circular traffic lights. However, this method is susceptible to color variations. Sathiya et al. (2015) utilized color threshold segmentation to extract the ROI for recognizing red and green traffic lights. They achieved this by comparing the counts of red and green pixels. Although this approach demonstrated high recognition accuracy, it suffered from a long inference time.

Machine learning-based traffic light status recognition methods primarily rely on human knowledge to extract relevant traffic light features and recognize their status. These methods demonstrate greater robustness in completing recognition tasks in specific scenarios. However, they are often unsuitable for other scenarios and suffer from poor real-time performance. Moreover, methods utilizing manual feature design tend to exhibit a lower recognition effect and a higher rate of leakage (Fang et al., 2023). With the development of hardware and software technologies, deep neural networks are progressively being used in the transportation field owing to their high reliability, applicability, and robustness (Lin et al., 2023b; Wang et al., 2023).

Numerous researchers have utilized deep learning methods for traffic light status recognition tasks. Saini et al. (2017) used color space transformation to segment the image and position the candidate region, after which an 8-layer convolutional neural network (CNN) was adopted to output the status. This method performed well in different lighting environments; however, there were cases of misrecognition of vehicle taillights. John et al. (2014) extracted texture, color, and shape features from candidate regions and employed a multilayer perceptron (MLP) for traffic light status recognition. Müller and Dietmayer (2018) proposed a TL-SSD method that utilizes Inception v3 as the base network for the SSD approach. It achieved 95% accuracy on a custom dataset. However, this method cannot ensure real-time performance. Qian et al. (2019) utilized an improved version of the YOLOv2 algorithm for traffic light status recognition and obtanied an mAP of 96.08% and a missing detection rate of 2.87%. Wang et al. (2021) improved the YOLOv4 algorithm and achieved an mAP of 82.15% on the LISA dataset, which was a 2.86% improvement over the original YOLOv4.

From existing research, it can be observed that the majority of recognition algorithms utilize foreign traffic signal datasets, such as LaRA and LISA, whereas there is limited research using domestic datasets. Furthermore, deep learning-based methods have significantly improved recognition accuracy and speed. However, they are susceptible to environmental factors such as vehicle obstacles, streetlights, and taillights, which can lead to certain levels of false and missed detections. Therefore, it is necessary to further enhance the recognition accuracy and complex scenario adaptability of traffic light status recognition algorithms. To overcome these problems, we constructed CUTLD and proposed an end-to-end traffic light status recognition algorithm called RC-DINO.

3 Chinese urban traffic light dataset

We cleaned and refused the domestic open-source traffic light dataset TTTL with S2TLD: (1) we filtered the traffic signal instances from the JSON file and deleted empty images; (2) we reannotated the traffic light instances with the category "Red left turn" and "Green forward" as "Red" and "Green", respectively; (3) the TTTL dataset was fused with S2TLD to form CUTLD in COCO format with the characteristics of Chinese urban traffic light.

CUTLD has 10,762 images with over 30,000 labeled-instances, including four traffic light status categories: "red", "yellow", "green", and "off". It covers traffic light scenarios, such as dense start-stop traffic, complex light interference, and low visibility. Traffic light scenarios in the CUTLD are shown in Fig. 1.

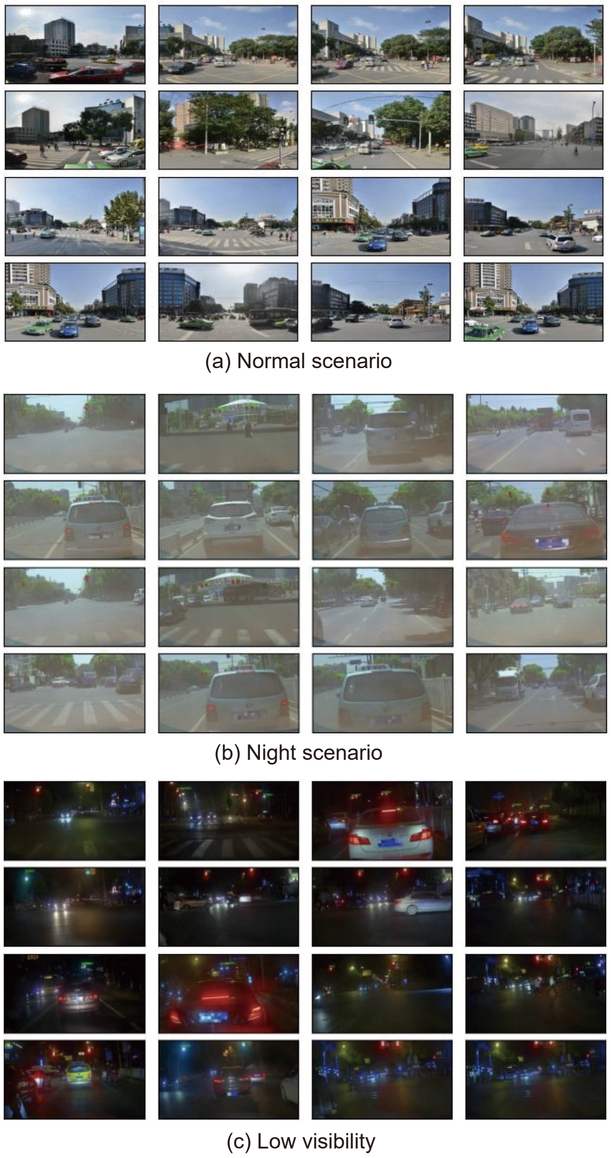


Fig. 1 Partly traffic light scenarios in CUTLD: (a) normal scenarios, (b) night scenarios, and (c) low visibility scenario.

4 ResNeSt50-CBAM-DINO

RC-DINO is an end-to-end pipeline built on the DINO framework and comprises four main parts (Vaswani et al., 2017). The integrated architecture of the RC-DINO is depicted in Fig. 2. The backbone extracts important features from the traffic light images, which are then combined with their respective positional encodings. We employ a 6-layer transformer encoder for feature enhancement. Subsequently, we selected position-related information from the top-K (K stands for an integer value) features to set up the position queries (Zhang et al., 2022a). Then, multi-scale deformable attention was used to merge the output features of the encoder and perform layer-wise updates. Finally, an feedforward neural network (FFN) is utilized to output the predicted positions of traffic lights and their corresponding status classification results. Subsequently, multi-scale deformable attention was used to merge the encoder output features and perform layer-by-layer updates. Finally, the traffic light prediction object frame location and status classification results are output using FFN.

4.1 Backbone

ResNeSt is an improved version of the ResNet (He et al., 2016) that can be directly applied to downstream vision tasks and exhibits superior performance compared with ResNet. ResNeSt retains the basic architecture of ResNet while incorporating a split attention module that performs attention across groups of feature maps and assigns higher weights to important regions in the feature map. Therefore, the backbone network was replaced with ResNeSt50.

The ResNeSt50 Block is depicted in Fig. 3, where h, w, and c represent the number of channels in different feature maps. c' and c'' are the number of channels for the intermediate features. $i \in [1, r]$ represents the i-th branch for splitting, and $j \in [1, k]$ represents the j-th group. Position swapping of different modules in the ResNeSt50 Block did not affect the results. In other words, for the same value of r, the attention formation for different k values is the same. The radix-major approach was implemented in the experiments, which allowed for modular accelerated training using standard CNN operators.

Additionally, CBAM was added to ResNeSt50 to further facilitate information interaction between the feature maps. For the specific feature map $F \in R^{C \times H \times W}$, where C, W, and H represent the number of channels in the different feature maps, it calculates weights along both the channel and spatial dimensions independently. These weights were then multiplied by F to dramatically modify the feature map and efficiently restrain unrelated features. Consequently, CBAM achieves better performance than SENet, which focuses solely on the channel

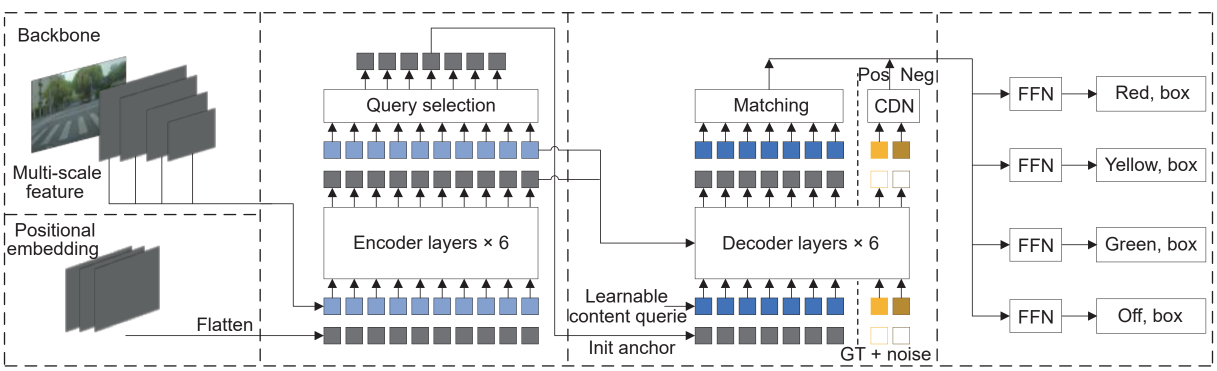


Fig. 2 Architecture of RC-DINO.

https://doi.org/10.26599/JICV.2023.9210022

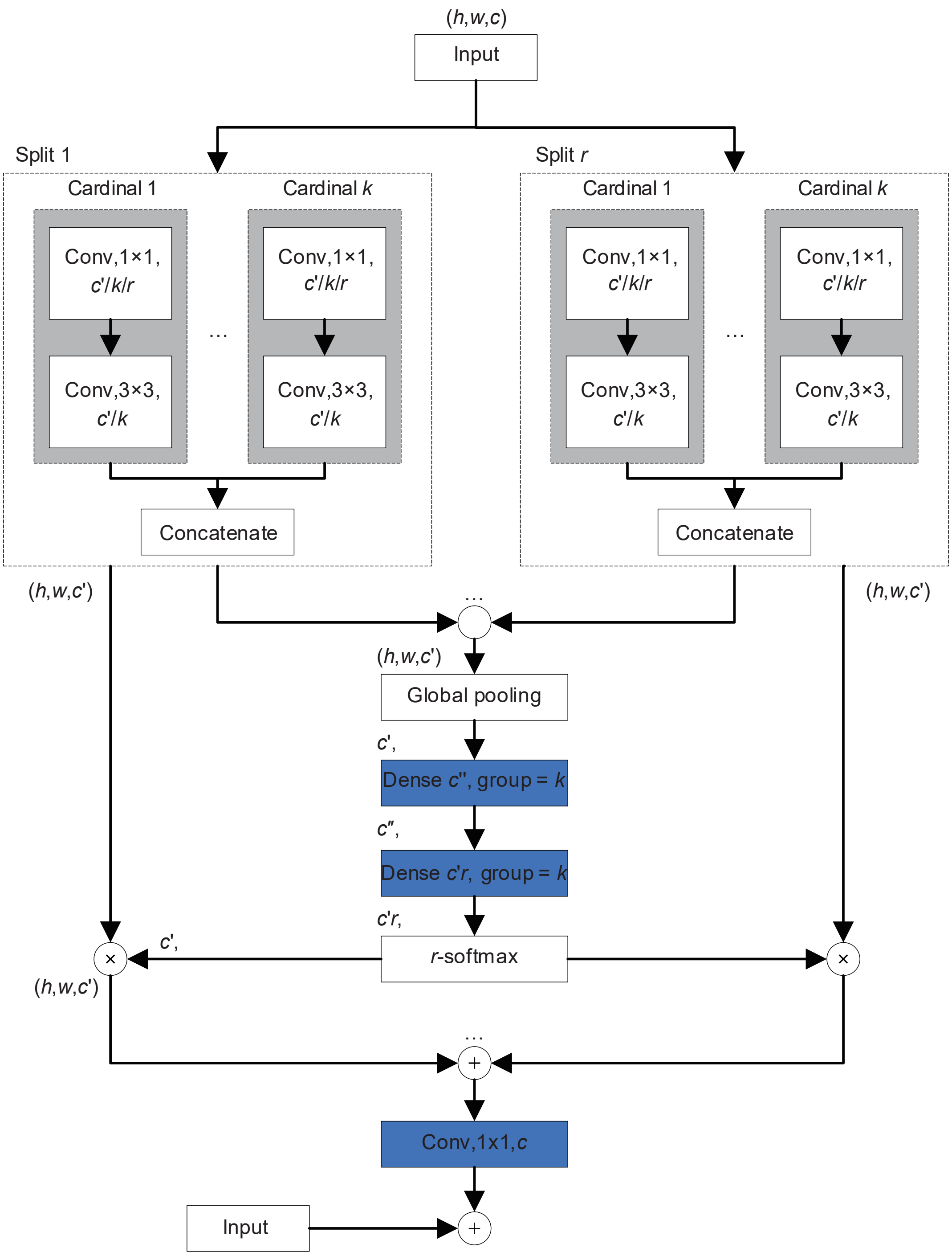


Fig. 3 ResNeSt50 block.

attention mechanism. Fig. 4 shows the architecture of CBAM.

The initial $F \in R^{H\times W\times C}$ undergoes global pooling operations along the width and height dimensions (Jiang and Huang, 2023), and we obtain two $1\times1\times C$ matrices. These two matrices were then input into MLP. Subsequently, the outputs of the MLP are combined with the obtained matrices to generate M_c. M_c is then combined with F to produce F'. Afterward, F' undergoes channel-based global pooling operations, resulting in two $H\times W\times 1$ matrices. The two matrices obtained were merged along the channel dimension. Subsequently, convolutional operations and a sigmoid function are applied to obtain spatial attention feature (M_s). Finally, the final output of CBAM F'' is obtained based on M_s and F' (Qi et al., 2023). The specific process is described by Eqs. (1)–(4):

$$M_c(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \tag{1}$$

$$F' = M_c(F) \otimes F \tag{2}$$

$$M_s(F') = \sigma\left[f^{7\times7}(\mathrm{AvgPool}(F'));(\mathrm{MaxPool}(F'))\right] \tag{3}$$

$$F'' = M_s(F') \otimes F' \tag{4}$$

4.2 Transformer architecture of RC-DINO

The SOTA results obtained by DINO in the COCO2017 (Lin et al., 2014) demonstrate that the DETR-like end-to-end concise pipeline performed well in computer vision tasks. As shown in Fig. 5, the architecture of RC-DINO is consistent with that of DINO.

In this study, the encoder consists of $N = 6$ identical encoder layers. A 1×1 CNN was used to dimension the output from the backbone into a one-dimensional representation, which was then combined with spatial position encodings and input into the encoder. Each encoder layer is composed of two sub-layers. The

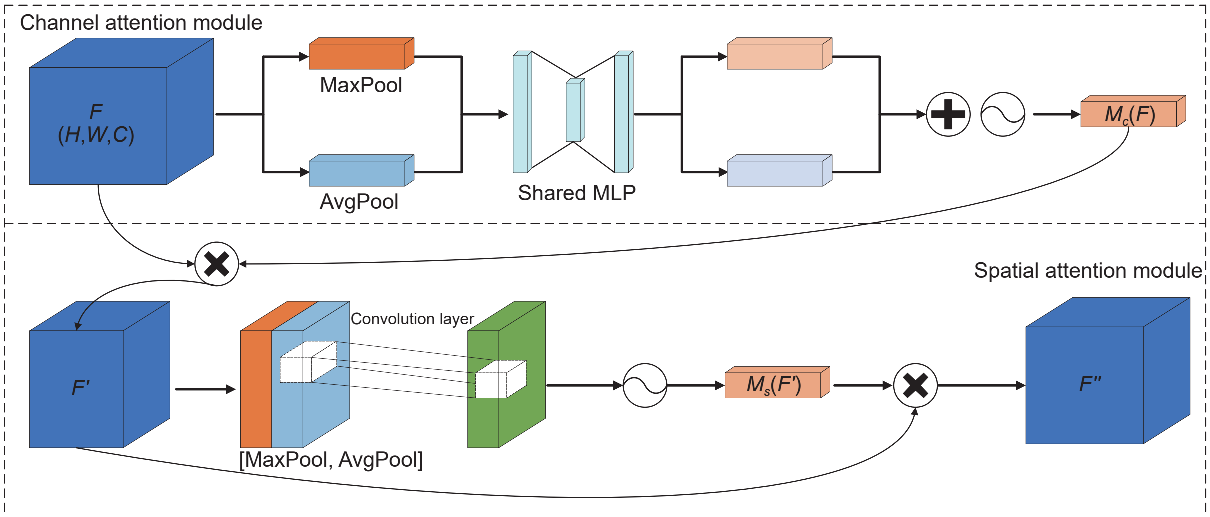


Fig. 4 Channel attention and spatial attention of CBAM.

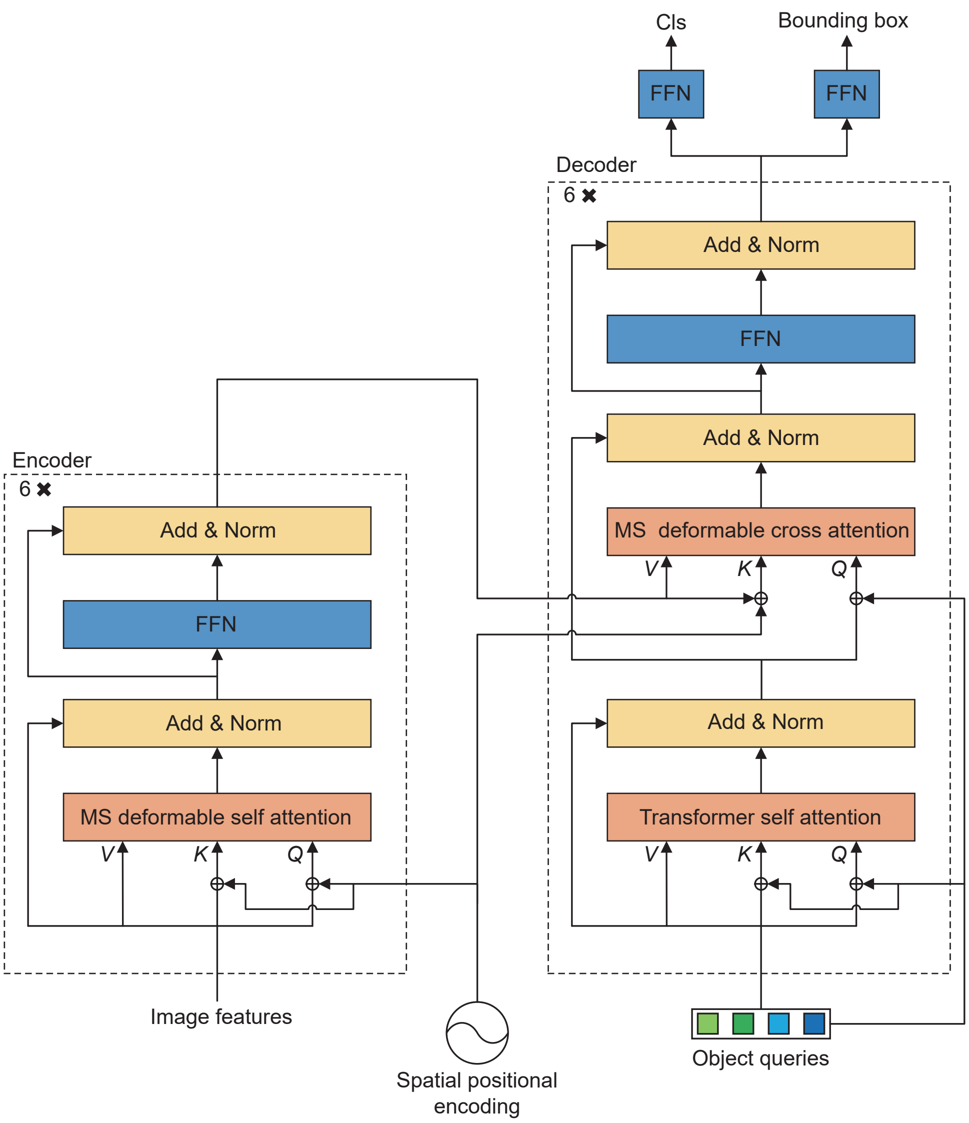


Fig. 5 Transformer architecture of DINO.

decoder consists of $M = 6$ identical decoder layers. The inputs of the decoder include spatial position encoding, object queries, and the encoder's outputs, where the content query is set to be learnable and the position queries are initialized with position-related information from the selected top-K features. Each decoding layer consists of three sub-layers. Residual connections were applied between every two sub-layers and layer normalization was performed.

5 Experiments and results

5.1 Training details

The specific hardware and software configurations used in these experiments are listed in Table 1. The CPU was an i9-12900KF, the GPU was an NVIDIA GeForce RTX 3090 Ti, the CUTLD was divided in a ratio of 8:1:1, and the starting learning rate was 0.0001. In addition, the pre-trained weights of ResNeSt50 were used.

Table 1 Experiment environment configuration

Configuration	Parameter
Operating system	Win 10
Central processing unit	i9-12900KF
Graphics processing unit	NVIDIA GeForce RTX 3090 Ti
CUDA	V11.1.105
CUDNN	V8.6.0
Initial learning rate	0.0001
Batch size	2
Weight decay	0.0001

5.2 Comparative experiments of DINO

We compared the original DINO with seven mainstream algorithms: YOLOv7, YOLOv8, Faster-RCNN (Ren et al., 2015), and DETR-like models (Carion et al., 2020). In our experiment, we replaced the backbone of these networks with ResNet50 and used several metrics to compare the performance of each algorithm.

As shown in Table 2, all types of AP (the AP mentioned in this paper are all mean average precisions) of DAB-DETR, DN-DETR, and DINO are substantially higher than the rest of the algorithms, which can be achieved better than the DETR and YOLO series algorithms by training for only a few epochs. DINO required only 36 epochs to achieve optimal results, which reached 95.9% AP_{50} and was considerably higher than those of Faster-RCNN (↑5.7%) and YOLOv7 (↑3.2%). Therefore, DINO was selected as the basic framework for the RC-DINO.

5.3 Comparative experiments of RC-DINO

To further illustrate the validity of the presented algorithm and showcase the intuitive experimental results, we compared the overall performances of DAB-DETR, DN-DETR, and DINO with that of RC-DINO. As illustrated in Fig. 6, RC-DINO exhibits higher precision at different recall rates, with its PR curve completely encompassing those of the other algorithms. Moreover, DINO and RC-DINO showed a stable AP (in Fig. 6, AP denotes AP_{50}) after the twelfth epoch, with RC-DINO converging to a significantly higher AP of 97.5% than with the other algorithms. Based on these findings, we conclude that RC-DINO exhibits the best performance and is the most suitable algorithm for traffic light status-recognition tasks.

To quantitatively analyze the superiority of the RC-DINO from multiple perspectives, we selected AP_{50}, AP_S, AR, FPS, and F1-Score as the main evaluation metrics. As shown in Table 3,

Table 2 Comparison of different algorithms in COCO evaluation metrics

Method	Epoch	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster-RCNN	20	62.4%	90.2%	76.6%	58.2%	69.3%	73.7%
YOLOv7	300	58.4%	92.7%	65.0%	52.4%	65.6%	73.9%
YOLOv8	100	52.4%	86.3%	57.0%	42.0%	67.5%	72.0%
DETR	500	54.1%	84.5%	63.0%	44.8%	63.5%	68.4%
Deformable-DETR	50	59.9%	93.0%	70.6%	50.7%	69.4%	69.4%
DAB-DETR	50	64.9%	94.2%	78.5%	58.6%	72.5%	75.8%
DN-DETR	50	67.4%	95.5%	80.3%	61.0%	74.4%	78.3%
DINO	36	67.9%	95.9%	82.5%	61.3%	75.1%	77.5%

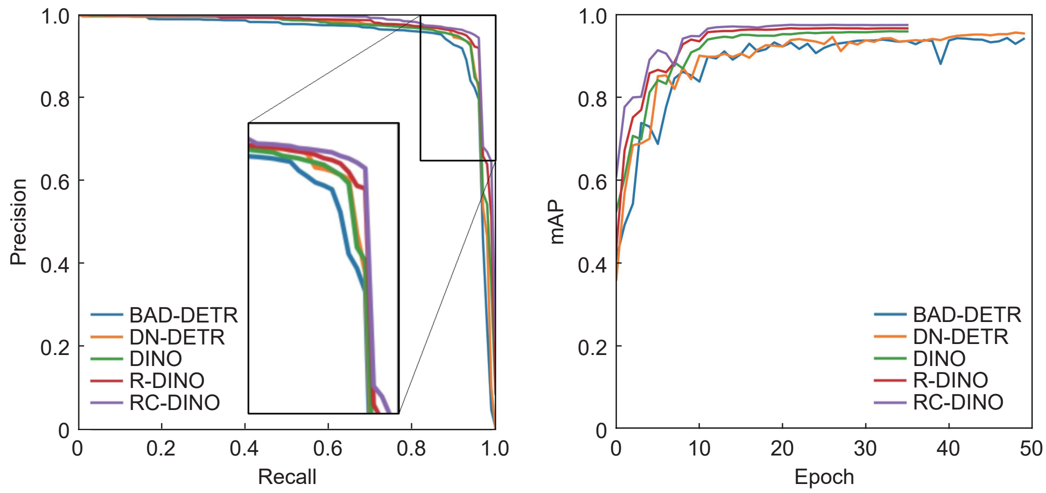


Fig. 6 PR curves and AP curves of different algorithms.

Table 3 Results of different algorithms under the main evaluation metrics

Method	Epoch	AP	AP_{50}	AP_S	AR	FPS	F1-Score
DAB-DETR	50	64.9%	94.2%	58.6%	75.7%	14	96.2%
DN-DETR	50	67.4%	95.5%	61.0%	78.0%	14	96.8%
DINO	36	67.9%	95.9%	61.3%	78.7%	13	97.3%
R-DINO	36	68.7%	96.7%	62.2%	79.0%	13	97.8%
RC-DINO	36	71.0%	97.5%	64.7%	79.6%	13	98.2%

compared with other algorithms, RC-DINO demonstrated improved evaluation metrics across all categories while exhibiting only a minor reduction of 1 FPS. As shown in Table 4, compared with DINO, the parameter and calculation quantities of RC-DINO only increased by 1.9 M and 10.5 G, respectively. However, the AP, AP_{50}, AP_S, AR, and F1-Score increased by 3.1%, 1.6%, 3.4%, 0.9%, and 0.9%, respectively. Thus, we conclude that the incorporation of ResNeSt50 and CBAM can significantly enhance the traffic light status recognition capability of RC-DINO, with only a minimal increase in computational overhead.

We chose AP, AP_{50}, AR, FPS, and F1-Score as the main evaluation metrics to draw the radar map. As shown in Fig. 7, owning to the increase in the number of model parameters and computational workload, the FPS of RC-DINO was slightly lower than those of DAB-DETR and DN-DETR, whereas other performance metrics exhibited significant improvements. In terms of the entire area of the radar map, the RC-DINO had the largest area, indicating that the comprehensive performance of the RC-DINO is the best.

5.4 Visual analysis of recognition results

Furthermore, we employed DINO and RC-DINO to generate visual results for complex traffic light scenarios beyond the scope of CUTLD. As shown in Figs. 8–10, in the daytime scenario, the DINO had redundant object boxes. R-DINO correctly recognized the green light, but the positioning accuracy of the traffic light background panel was not high. RC-DINO can correctly identify the traffic light status and position it accurately. In the low-visibility scenario, DINO and R-DINO could not recognize the red traffic light at the visual boundary, whereas RC-DINO could recognize it. In the nighttime scenario, although all of the algorithms demonstrated superior recognition capabilities, RC-DINO achieved the highest confidence level for recognizing green lights.

Table 4 Parameter quantities and calculation quantities of different algorithms

Method	Parameter	FLOP
DINO	45.2 M	119.0 G
R-DINO	47.1 M	129.5 G
RC-DINO	47.1 M	129.5 G

As shown in Fig. 11, the three models can recognize the green light covered by clutter telecommunications, but RC-DINO has the highest recognition confidence. As shown in Figs. 12 and 13, only the improved DINO model could recognize the green light partially covered by leaves in the images, and the recognition confidence of RC-DINO was higher. The visualization recognition results intuitively show that RC-DINO outperforms DINO in complex scenarios and exhibits the ability to recognize the status

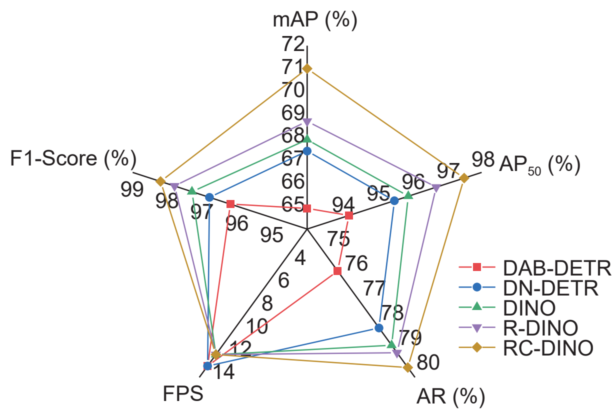
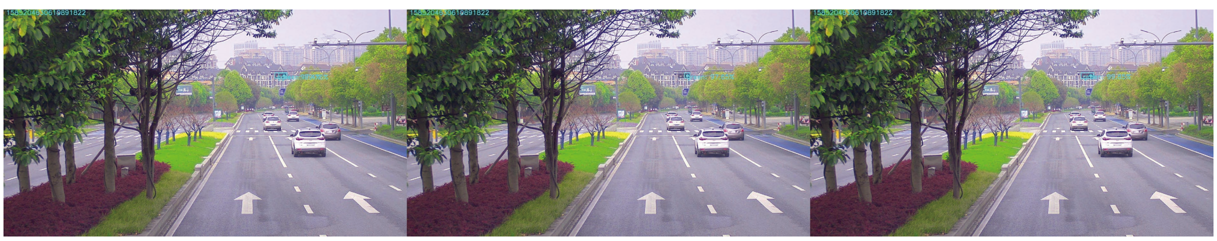
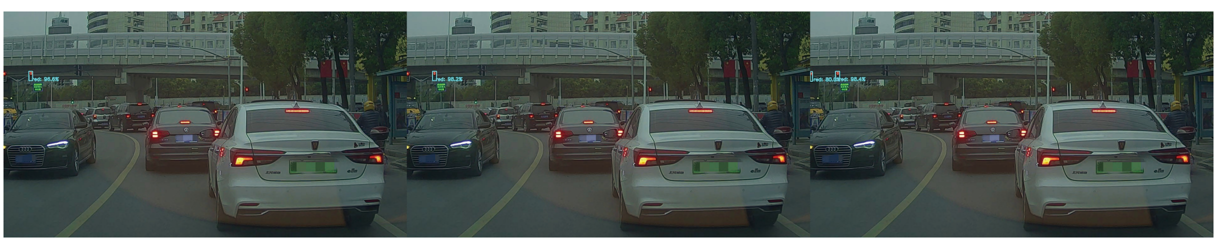


Fig. 7 Comparison radar map of different algorithms.



(a) DINO (b) R-DINO (c) RC-DINO

Fig. 8 Recognition results of different algorithms in daytime scenario.



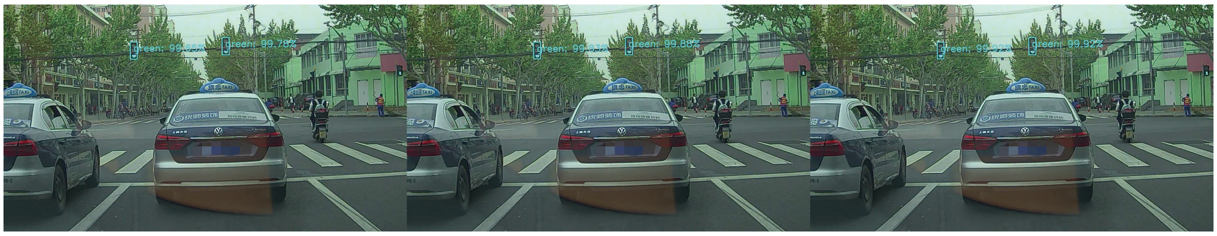
(a) DINO (b) R-DINO (c) RC-DINO

Fig. 9 Recognition results of different algorithms in low-visibility scenario.

清華大学出版社 Tsinghua University Press | IEEE

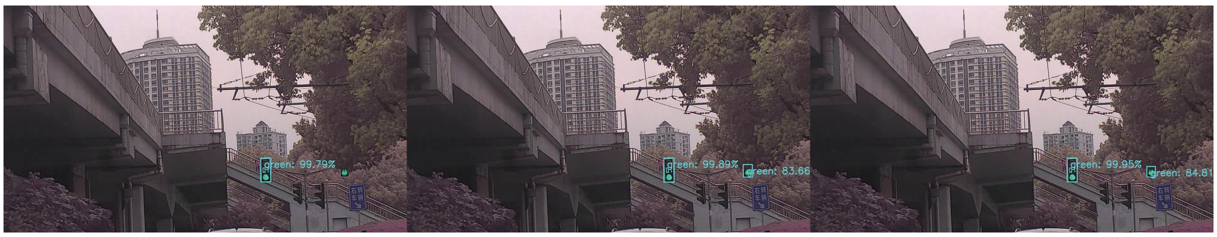
(a) DINO (b) R-DINO (c) RC-DINO

Fig. 10 Recognition results of different algorithms in nighttime scenario.



(a) DINO (b) R-DINO (c) RC-DINO

Fig. 11 Recognition results of different algorithms in covered scenario.



(a) DINO (b) R-DINO (c) RC-DINO

Fig. 12 Recognition results of different algorithms in covered scenario.

(a) DINO (b) R-DINO (c) RC-DINO

Fig. 13 Recognition results of different algorithms in covered scenario.

of covered traffic lights. In addition, it should be noted that there was no image data of covered traffic lights in the CUTLD used to train the algorithms. However, the RC-DINO has a certain capability to recognize the partially covered traffic light status, which verifies the superior generalization of the algorithm proposed in this study.

5.5 Horizontal comparison of RC-DINO

To further demonstrate the superiority of our proposed method, we compared RC-DINO with the methods proposed by Chen et al. (2021) and Sathiya et al. (2015). As shown in Table 5, RC-DINO can recognize four statuses: green, red, yellow, and off. In contrast, Chen et al. (2021) and Sathiya et al. (2015)'s methods can only identify regular colors without determining whether the traffic light is off. Additionally, Sathiya et al. (2015)'s method is limited to recognizing static images rather than analyzing dynamic video streams. Furthermore, it can only recognize the status under favorable weather conditions. On the other hand, our proposed method excels in recognizing the status in continuous dynamic scenarios under various weather conditions. Compared to Sathiya et al. (2015)'s method, our algorithm exhibits a shorter execution time while achieving higher recognition accuracy. Although the execution time of our algorithm was similar to that of Chen et al. (2021)'s method, its recognition accuracy improved by approximately 5%. Overall, the RC-DINO performed better.

Table 5 Traffic light recognition algorithm comparison

Metric	This study	Chen et al. (2021)	Sathiya et al. (2015)
Kinds of status	GRY+Off	GRY	GRY
Dynamic identification	√	√	×
Weather condition	Variety	Variety	Fine day
AP	97.5%	92.76%	96%
Avg. exec time (ms)	76	67.35	> 2,000

6 Conclusions

In this study, we constructed a Chinese urban traffic light dataset.

Subsequently, an end-to-end traffic light status recognition algorithm called RC-DINO was proposed based on DINO. It replaces the original DINO backbone with ResNet50 and accesses the CBAM to enhance the feature information extraction ability of the backbone for traffic lights. Lastly, we conducted numerous comparative experiments to confirm that the RC-DINO demonstrates superior comprehensive performance and robustness for the recognition of traffic light status. Furthermore, it exhibits a certain capability to recognize the partially covered traffic light status. Owing to the large number of parameters and computational requirements of deep learning algorithms, along with the requirement for certain hardware equipment, the FPS of these algorithms may be lower than that of lightweight recognition algorithms. In future research, we will focus on lightweight algorithms that can significantly improve the FPS of the model without reducing the accuracy, enabling its successful application to vehicles. In addition, high-precision maps will be incorporated to achieve lane-level traffic light status recognition and better guide the safe driving of autonomous vehicles.

Replication and data sharing

Tsinghua–Tencent traffic lights (TTTL) is available at https://cg.cs.tsinghua.edu.cn/traffic-light. Shanghai Jiao Tong University's traffic light dataset (S2TLD) is available at https://github.com/Thinklab-SJTU/S2TLD.

Acknowledgements

This research was supported by the National Key R&D Program of China (2021YFB2501200), the Key Program of the National Natural Science Foundation of China (52131204), and the Shaanxi Province Key Research and Development Program (2022GY-300).

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S., 2020. End-to-end object detection with transformers. In: European Conference on Computer Vision. Cham: Springer, 2020, 213–229.

Chen, X., Chen, Y., Zhang, G., 2021. A computer vision algorithm for locating and recognizing traffic signal control light status and countdown time. J Intell Transp Syst, 25, 533–546.

Fang, S., Yang, L., Zhao, X., Wang, W., Xu, Z., Wu, G., *et al.*, 2023. A Dynamic Transformation Car-Following Model for the Prediction of the Traffic Flow Oscillation. IEEE Intell Trans Syst Mag.

Gong, J., Jiang, Y., Xiong, G., Guan, C., Tao, G., Chen, H., 2010. The recognition and tracking of traffic lights based on color segmentation and CAMSHIFT for intelligent vehicles. In: 2010 IEEE Intelligent Vehicles Symposium, 431–435.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

He, Y., Liu, Y., Yang, L., Qu, X., 2023. Deep adaptive control: Deep reinforcement learning-based adaptive vehicle trajectory control algorithms for different risk levels. IEEE Trans Intell Veh. https://doi.org/10.1109/TIV.2023.3303408

Jiang, J., Huang, H., 2023. Semantic segmentation of remote sensing images based on dual channel attention mechanism (DCAM). https://doi.org/10.21203/rs.3.rs-3006101/v1

John, V., Yoneda, K., Qi, B., Liu, Z., Mita, S., 2014. Traffic light recognition in varying illumination using deep learning and saliency map. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2286–2291.

Lee, S. H., Kim, J. H., Lim, Y. J., Lim, J., 2018. Traffic light detection and recognition based on Haar-like features. In: 2018 International Conference on Electronics, Information, and Communication (ICEIC), 1–4.

Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Zitnick, C., 2014. Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, 740–755.

Lin, H., He, Y., Liu, Y., Gao, K., Qu, X., 2023a. Deep demand prediction: An enhanced conformer model with cold-start adaptation for origin–destination ride-hailing demand prediction. IEEE Intell Transp Syst Mag, 2–15.

Lin, H., Liu, Y., Li, S., Qu, X., 2023b. How generative adversarial networks promote the development of intelligent transportation systems: A survey. IEEE/CAA J Autom Sin, 10, 1781–1796.

Liu, Y., Jia, R., Ye, J., Qu, X., 2022. How machine learning informs ride-hailing services: A survey. Commun Transport Res, 2, 100075.

Liu, Y., Lyu, C., Zhang, Y., Liu, Z., Yu, W., Qu, X., 2021. DeepTSP: Deep traffic state prediction model based on large-scale empirical data. Commun Transport Res, 1, 100012.

Liu, Y., Wu, F., Liu, Z., Wang, K., Wang, F., Qu, X., 2023. Can language models be used for real-world urban-delivery route optimization? Innovation, 4, 100520.

Mentasti, S., Simsek, Y. C., Matteucci, M., 2023. Traffic lights detection and tracking for HD map creation. Front Robot AI, 10, 1065394.

Müller, J., Dietmayer, K., 2018. Detecting traffic lights by single shot detection. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 266–273.

Omachi, M., Omachi, S., 2009. Fast detection of traffic light with color and edge information. Electronics Engineers of Japan, 38, 673–679.

Qian, H., Wang, L., Mou, H., 2019. Fast detection and identification of traffic lights based on deep learning. Comput Sci, 46, 272–278. (in Chinese)

Qi, F., Yang, C., Shi, B., Ma, S., 2023. Micro-expression recognition based on DCBAM-EfficientNet model. J Phys: Conf Ser, 2504, 012062.

Reddy, V. P., Raja, A. R., Polasi, P. K., Ponnuri, R. T., Kumar, G. K., Ahamed, S. F., 2023. Design and Development of Traffic Light Recognition method for Autonomous vehicles using V2I Communication. In: 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), 1–6.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell, 39, 1137–1149.

Saini, S., Nikhil, S., Konda, K. R., Bharadwaj, H. S., Ganeshan, N., 2017. An efficient vision-based traffic light detection and state recognition for autonomous vehicles. In: 2017 IEEE Intelligent Vehicles Symposium (IV), 606–611.

Sathiya, S., Balasubramanian, M., Priya, D. V., 2015. Real time recognition of traffic light and their signal count-down timings. In: International Conference on Information Communication and Embedded Systems (ICICES2014), 1–6.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., *et al.*, 2017. Attention is all You need. Advances in Neural Information Processing Systems, 2017, 30.

Wang, K., Xiao, Y., & He, Y., 2023. Charting the future: Intelligent and connected vehicles reshaping the bus system. J Intell Connect Veh. http://doi.org/10.26599/JICV.2023.9210024

Wang, Q., Zhang, Q., Liang, X., Wang, Y., Zhou, C., Mikulovich, V. I., 2021. Traffic lights detection and recognition method based on the improved YOLOv4 algorithm. Sensors, 22, 200.

Woo, S., Park, J., Lee, J. Y., Kweon, I. S., 2018. CBAM: convolutional block attention module. In: European Conference on Computer Vision. Cham: Springer, 3–19.

Zeng, S., Wang, R., Li, M., Guan, L., 2023. N_ResNet: A real-time traffic light recognition network using object detection. https://doi.org/10.21203/rs.3.rs-2526425/v1

Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., *et al.*, 2022a. DINO: DETR with improved DeNoising anchor boxes for end-to-end object detection. https://doi.org/10.48550/arXiv.2203.03605

Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., *et al.*, 2022b. ResNeSt: split-attention networks. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2735–2745.

Zong, F., He, Z., Zeng, M., Liu, Y., 2022a. Dynamic lane changing trajectory planning for CAV: A multi-agent model with path preplanning. Transp B Transp Dyn, 10, 266–292.

Zong, F., Wang, M., Tang, J., Zeng, M., 2022b. Modeling AVs & RVs' car-following behavior by considering impacts of multiple surrounding vehicles and driving characteristics. Phys A Stat Mech Appl, 589, 126625.

Lan Yang received the Ph.D. degree from Chang'an University, Xi'an, China, in 2013. Currently, she is an Associate Professor with the School of Information Engineering, Chang'an University. Her research interests include the cooperative eco-driving of CAVs, vehicle active safety technology, and vehicle behavior cognition.

Zeyu He received the B.S. degree in network engineering from Tianjin University of Finance and Economics, Tianjin, China, in 2021. He is currently pursuing the M.S. degree in the School of Information Engineering at Chang'an University, Xi'an, China. His research interests include autonomous vehicle environment perception, object detection, and deep learning.

Xiangmo Zhao (Member, IEEE) received the Ph.D. degree from Chang'an University, Xi'an, China, in 2006. He is currently the President of Xi'an Technology University and a Professor with the School of Information Engineering, Chang'an University. He is also the Vice President of the Joint Laboratory for Connected Vehicles, Ministry of Education and China Mobile, and the Shaanxi Road Traffic Intelligent Detection and Equipment Engineering Technology Research Center. His research interests include connected vehicles, automated vehicles, and intelligent transportation systems. He is the Director of the Information Professional Committee, a member of the Advisory Expert Group of the China Transportation Association, a member of the National Motor Vehicle Operation Safety Testing Equipment Standardization Committee and the Leading Group of the National Traffic Computer Application Network, the Vice Chairperson of the Institute of Highway Association on Computer Professional Committee, and the Deputy Director of the Institute of Computer in Shaanxi Province.

Shan Fang received the B.S. degree in computer science from Jilin University, Changchun, China, in 2016. He is currently pursuing the Ph.D. degree in transportation information engineering and control in the School of Information Engineering, Chang'an University, Xi'an, China. His research interests include traffic flow simulation and cooperative eco-driving strategy for vehicles.

Jiaqi Yuan received the B.S. degree in software engineering from the School of Computer Science, Chengdu University, Chengdu, China, in 2021. She is currently pursuing the M.S. degree in the School of Information Engineering at Chang'an University in Xi'an, China. Her research interests include autonomous driving test and evaluation, deep learning, and intelligent transportation systems.

Yixu He received the B.S. degree in computer science in the School of Information Engineering, Chang'an University, Xi'an, China, in 2021. He is currently pursuing the M.S. degree in the School of Information Engineering, Chang'an University, Xi'an, China. His research interests include transportation big data analysis and modelling, machine learning, data mining, and intelligent transportation systems.

Shijie Li received the B.S. degree in Internet of Things engineering from the School of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an, China, in 2022. He is currently pursuing the M.S. degree in the School of Information Engineering at Chang'an University, Xi'an, China. His research interests include autonomous vehicle testing and evaluation.

Songyan Liu received the B.S. degree in computer science and technology from the School of Information Engineering, Chang'an University, Xi'an, China, in 2021. She is currently pursuing the M.S. degree at the School of Information Engineering of Chang'an University, Xi'an, China. Her research interest is transportation engineering.