

Vehicle Detection at Night Based on Style Transfer Image Enhancement

Jianing Shen* and Rong Li

Abstract

Most vehicle detection methods have poor vehicle feature extraction performance at night, and their robustness is reduced; hence, this study proposes a night vehicle detection method based on style transfer image enhancement. First, a style transfer model is constructed using cycle generative adversarial networks (cycleGANs). The daytime data in the BDD100K dataset were converted into nighttime data to form a style dataset. The dataset was then divided using its labels. Finally, based on a YOLOv5s network, a nighttime vehicle image is detected for the reliable recognition of vehicle information in a complex environment. The experimental results of the proposed method based on the BDD100K dataset show that the transferred night vehicle images are clear and meet the requirements. The precision, recall, mAP@.5, and mAP@.5:.95 reached 0.696, 0.292, 0.761, and 0.454, respectively.

Keywords

BDD100K Dataset, CycleGAN, Image Enhancement, Style Transfer Model, Vehicle Detection at Night, YOLOv5s Network

1. Introduction

Traffic congestion has become a major problem that seriously affects living standards. Building intelligent traffic monitoring systems reduces traffic congestion and improves traffic efficiency [1]. The emergence of intelligent transportation has received extensive attention, and relevant transportation departments hope to solve road problems by developing intelligent transportation systems. Through statistics on traffic flow, managers can quickly understand road congestion and make corresponding decisions promptly. Using cameras to obtain traffic flow information is considerably more convenient and faster than obtaining the information manually. Simultaneously, the use of cameras can reduce the workload of traffic police and reduce the manpower input of traffic management [2].

With the development of artificial intelligence technology centered on deep learning, autonomous vehicles are constantly advancing toward commercialization. Computer vision technology is used in driverless systems for road detection and the detection of environment-related objects such as vehicle targets.

Traditional computer vision solutions for vehicle target detection often require the artificial design of features, which can easily cause problems such as high time complexity and window redundancy, and

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received December 7, 2022; first revision February 3, 2023; accepted February 12, 2023.

* Corresponding Author: Jianing Shen (shenj@wxu.edu.cn)

College of Computer Internet of Things Engineering, Wuxi Taihu University, Wuxi, China (shenj@wxu.edu.cn, Lily_78@163.com)

their detection accuracy and speed cannot meet this requirement. The development of artificial intelligence technology can compensate for this deficiency. Deep learning and other related algorithms can achieve fast and high-precision target detection [3]. During the day, owing to the obvious features of vehicles, existing target detection algorithms can perform reliable detection. However, in a complex nighttime environment, owing to insufficient light and large changes in distances and angles, there are detection problems such as the occlusion of dense targets and omission of small targets.

2. Related Works

Current vehicle detection algorithms can be divided into target-detection and deep-learning algorithms. Traditional target detection algorithms extract vehicle information through threshold processing, morphological processing, and other methods and formulate a vehicle detection threshold through a sliding window. Parvin et al. [4] proposed a night vehicle recognition algorithm for an intelligent transportation system that used segmentation technology and the double threshold method; the algorithm realized vehicle tracking through a centroid algorithm. Shu et al. [5] proposed a local enhancement fusion Gaussian mixture model for detecting small mobile vehicles using a satellite. Kavya et al. [6] proposed an efficient vision-based vehicle detection and tracking system that uses a train cascade target detector to detect vehicles from a mobile host; the system also incorporates nonlinear filters, such as a Kalman filter, for vehicle tracking. This method overcomes the complexities of light, background, perspectives, and weather conditions. In [7], the authors proposed a vehicle detection and tracking method based on visual traffic monitoring. The detection accuracy was improved by considering the background subtraction concept and threshold processing of mobile vehicle detection combined with the application of speckle analysis and an adaptive bounding box. Traditional target detection algorithms have problems, such as low feature generalization and complex operations.

To resolve these problems, object detection algorithms have been proposed based on deep learning, such as the YOLO series, SSD, and Fast R-CNN. Based on the detection network of cross-frame keypoints and a spatial motion information guidance and tracking network, Feng et al. [8] realized the accurate detection of mobile vehicles using complementary information between frames to enhance the detection of keypoints. Paidi et al. [9] proposed a multitarget vehicle detection method based on a deep learning algorithm. By processing images using a thermal imager, vehicle detection in various complex situations can be realized with high accuracy. Luo et al. [10] proposed a vehicle-detection model based on a Faster R-CNN. This model improves detecting robustness for challenging targets. The aforementioned deep learning methods can better extract relevant vehicle features; however, target detection accuracy is low, and the model is slow, making it unsuitable for practical real-time detection.

A night vehicle detection method based on style transfer image enhancement is proposed to solve these problems. The main innovations of the proposed method are summarized as follows:

- 1) To obtain an effective night dataset, a style transfer model is constructed using cycle generative adversarial networks (cycleGANs). The daytime dataset was transferred to the night dataset to obtain an ideal style dataset.
- 2) Owing to the complexity of the night environment, reliable vehicle detection using traditional methods is difficult. Therefore, the YOLOv5s network was used to learn and analyze style datasets and obtain high-quality detection results through multi-angle and multilevel learning.

3. Proposed Method

The architecture of the proposed night vehicle detection method based on style transfer image enhancement is shown in Fig. 1.

- 1) Construction of the style dataset: Using the cycleGAN network to build a style transfer model, the daytime data in the BDD100K dataset were transferred to the nighttime data for testing the vehicle detection methods.
- 2) Dataset division: According to the BDD100K dataset label method, the dataset was divided for model training and testing.
- 3) Vehicle target recognition: The nighttime data were detected based on the YOLOv5s network to obtain a nighttime picture of the vehicle.

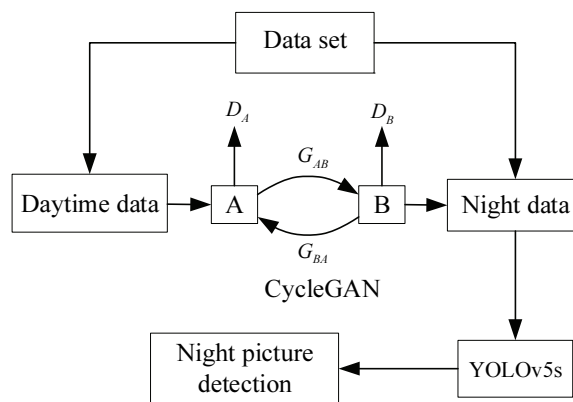


Fig. 1. Technical architecture of the proposed method.

3.1 BDD100K Dataset

The BDD100K dataset, released by the University of Berkeley in 2018, contains 100 thousand videos. Each video had a duration of approximately 40 seconds, high resolution (720 pixel), and high frame rate (30 fps). The videos also contained global positioning system/inertial measurement unit (GPS/IMU) information that recorded vehicle trajectory. The BDD100K dataset intercepts 100 thousand images from 100 thousand videos and labels the target objects in the images. This dataset contains data collected using a real driving platform under different weather conditions, scenes, and periods. Therefore, the BDD100K dataset has better data diversity than other datasets used in autonomous driving research.

There are approximately 1.84 million calibration boxes in the BDD100K dataset, of which the labels of the ground-truth calibration box have ten categories. In addition, the label file of BDD100K contains two parts: labels_images_train and labels_images_val, both of which use the JSON format.

3.2 Style Dataset Construction

There are currently no datasets for vehicle detection at night. Therefore, a fast style transfer model was used to generate the style dataset. The fast-style transfer model is composed of the cycleGAN model.

CycleGAN comprises two symmetrical mirror GANs forming a ring network, making its transformations cyclically consistent. Mathematically, if there are converters, $G_{AB}: A \rightarrow B$ and $G_{BA}: B \rightarrow A$, G_{AB} and G_{BA} should be reversed. The overall structure of the cycleGAN model is illustrated in Fig. 2.

CycleGAN aims to train the mapping function between two domains, namely, A and B . The model included two generators, $G_{AB}: A \rightarrow B$ and $G_{BA}: B \rightarrow A$, and two adversarial discriminators, D_A and D_B . D_A is used to distinguish between the image $\{a\}$ and transferred image sets $\{G_{AB}(b)\}$. Similarly, D_B is used to distinguish between the image $\{b\}$ and transferred image sets $\{G_{BA}(a)\}$.

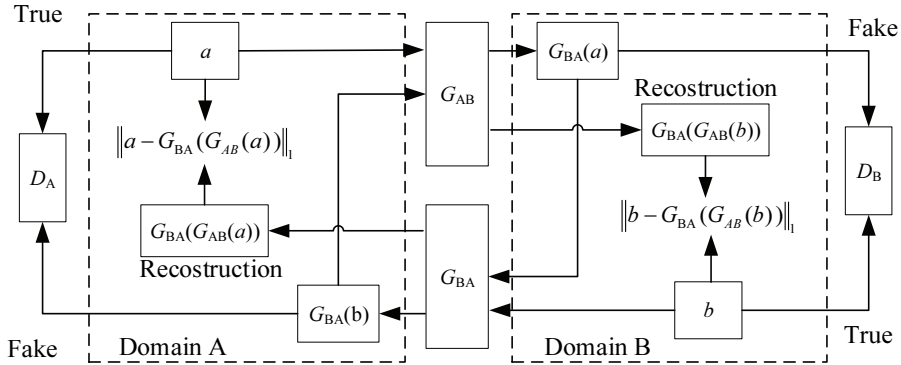


Fig. 2. Overall structure of the cycleGAN model.

There are three types of cycleGAN objective functions: adversarial loss, which is used to match the score of the generated image to the target domain; cycle consistency loss, which is used to prevent training G_{AB} and G_{BA} contradiction; and identity loss, which is used to save the color of the input image. The loss function is calculated as follows:

$$L(G_{AB}, G_{BA}, D_A, D_B) = L_{GAN}(G_{AB}, D_B, A, B) + L_{GAN}(G_{BA}, D_A, B, A) + \lambda L_{cyc}(G_{AB}, G_{BA}) + \lambda' L_{ide}(G_{AB}, G_{BA}), \quad (1)$$

where λ and λ' represent the importance of the cycle consistency and identity loss terms in the completed loss function, respectively. L_{GAN} , L_{cyc} , and L_{ide} represent the adversarial, cycle consistency, and identity loss functions, respectively.

Although the standard BDD100K dataset contains a large amount of vehicle information, vehicle information in the dark is lacking. Therefore, a style transfer model is constructed using the cycleGAN model. All the daytime data styles are transferred into the nighttime to obtain the nighttime dataset used to build a style dataset that meets nighttime conditions.

3.3 Dataset Division

The BDD100K dataset adopts JSON format; thus, “timeofday” in JSON is used to extract night pictures and daytime data from 700 million data points. The Python script `split_day_and_night.py` was used to divide the dataset. Part of the code is presented in Algorithm 1.

Algorithm 1. Partial code of dataset division

```

{
  "name": "0000f77c-6257be58",
  "frames": [
    {
      "timestamp": 10000,
      "objects": [
    ]
  },
  ],
  "attributes": {
    "weather": "clear",
    "scene": "city street",
    "timeofday": "daytime"
  }
}

```

3.4 Vehicle Recognition based on Style Transfer Image Enhancement

Because the features of vehicles at night are complex and difficult to observe, vehicle detection at night is currently difficult for driverless cars. The proposed method uses the YOLOv5s network model to process nighttime datasets and accurately identify vehicle images.

The YOLOv5 algorithm integrates the advantages of many deep-learning target detection frameworks. It has four versions with different volumes according to the depth and width of the network structure: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The minimum volume YOLOv5s algorithm was selected for vehicle recognition. The network structure of YOLOv5s is shown in Fig. 3.

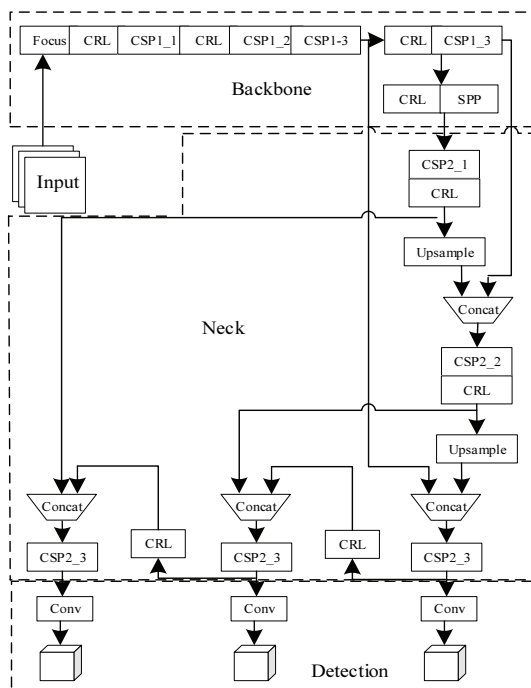


Fig. 3. Structure of the YOLOv5s network.

- 1) Input: YOLOv5s performs operations, splices the original data, enhances data into a convolutional neural network (CNN) for learning, and adopts an adaptive anchor frame calculation to obtain the appropriate anchor frame. Simultaneously, adaptive image scaling was used to reduce the calculations of the model and improve the detection of small targets.
- 2) Backbone: The backbone mainly includes the focus structure, cross stage partial (CSP) structure, and spatial pyramid pooling (SPP) module. The function of the focus module is to slice the image and then convolute it to obtain downsampled feature maps, which do not lose information. The YOLOv5s network model comprises two new CSP structures. The backbone network adopts the CSP1_1 and CSP1_3 structures, and the neck adopts the CSP2_1 structure. The CSP structure reduces computing bottlenecks and memory consumption. The SPP module mimics a pyramid. It enhances the expression ability of feature maps, which is beneficial for YOLOv5s to realize complex multitarget detection and improves detection accuracy.
- 3) Neck: Based on the feature pyramid network (FPN) structure and referring to the path aggregation network (PANet), a multiscale feature fusion network of the FPN+PAN structure is realized. After multiple downsamplings of the image to be detected using the backbone, FPN+PAN combine the target feature information at different scales.
- 4) Output: This output includes the bounding box loss function and non-maximum suppression (NMS). GIOU_LOSS, used in YOLOv5s, is the bounding box loss function. The NMS algorithm removes the prediction results whose GIOU_LOSS contribution is not the largest, and the prediction results with the largest probability are output to generate a bounding box and predict its category.

4. Experiments and Analysis

The experimental environment was based on Windows 10 Professional edition, and the program was written in Python. The network model was built using the deep-learning framework PyTorch 1.9.1+cu111. The hardware environment was an NVIDIA GeForce RTX 3060 12G.

4.1 Evaluation Index

This study evaluated the proposed method using three indices: precision, recall, and mean average precision (mAP). The larger the values of these indices, the better is the target detection. The indices were calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{mAP} = \frac{\sum_{i=1}^k AP_i}{k} \quad (4)$$

where AP_i is the area below the precision and recall curves and k is the AP category.

4.2 Style Transfer Effect Comparison

We selected approximately 1,000 daytime and 1,000 nighttime data to train the style transfer. According to nighttime data train A and daytime data train B under path `pytorch-CycleGAN- and pix2pix-master\datasets\day_night`, the style transfer model was used to transfer all daytime data styles into the nighttime style to obtain the nighttime training set. The effect of the style transfer is illustrated in Fig. 4.

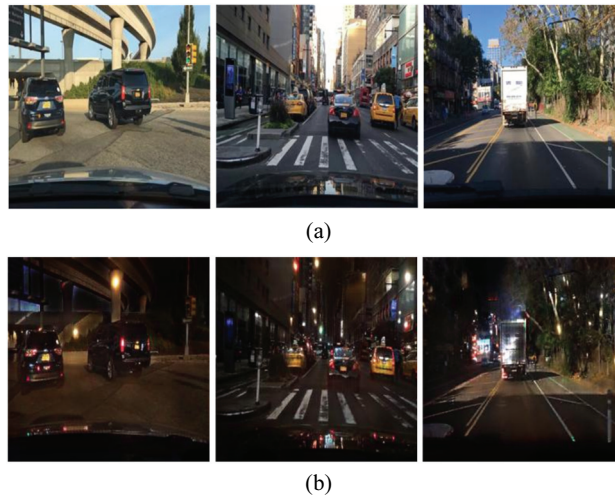


Fig. 4. Results of style transfer: (a) original daytime map and (b) transfer to a nighttime map.

As shown in Fig. 4, the cycleGAN-style transfer model can effectively transfer daytime data to nighttime data without changing other elements, such as vehicle and road information. Simultaneously, the image obtained at night has a good definition, providing a high-quality dataset for vehicle detection at night.

4.3 Comparison of Target Detection Results

To demonstrate the detection performance of the proposed method, it was compared with the YOLOv5 network model. The two methods were used to detect vehicle information in the BDD100K dataset simultaneously, and their performances were compared according to the calculation and analysis of the evaluation index.

4.3.1 Qualitative comparison

The detection results of the YOLOv5 network model and proposed method for the BDD100K dataset are shown in Fig. 5.

In Fig. 5, the YOLOv5 network model exhibits a missed detection rate. For example, vehicles are not detected in the first picture; however, the proposed method can detect all the vehicles in the dark. This is because the proposed method uses cycleGAN for image enhancement, which makes it convenient for the YOLOv5 network to obtain a better target image, demonstrating that the proposed method could detect vehicle information more comprehensively.

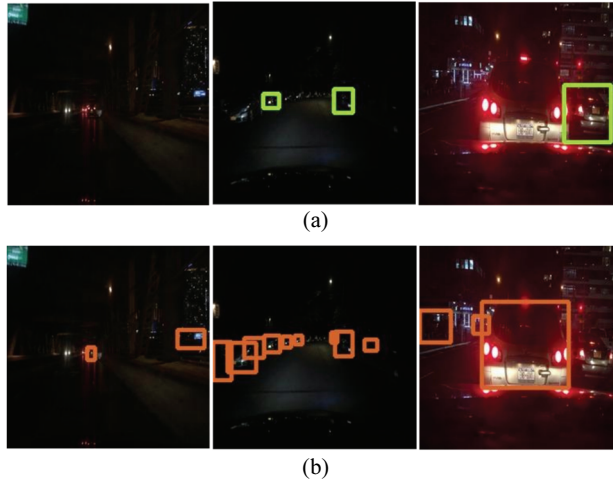


Fig. 5. Target detection results of different methods: (a) YOLOv5 and (b) proposed method.

4.3.2 Quantitative comparison

A quantitative analysis of the detection results obtained from the YOLOv5s network, and the proposed method is shown in Fig. 6.

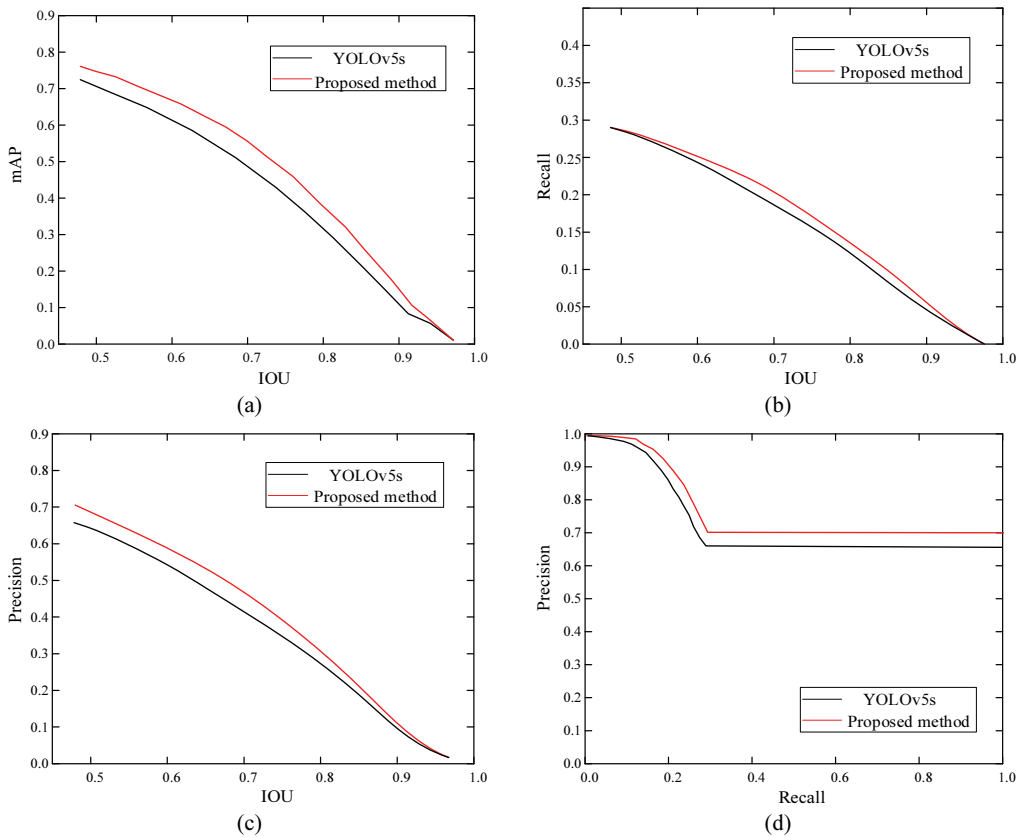


Fig. 6. Quantitative comparison of vehicle detection results: (a) mAP, (b) recall, (c) precision, and (d) P-R curve.

In Fig. 6, with an increase in the intersection over union (IOU), the mAP, recall, and precision all decline. However, for the same IOU, the values of the mAP, recall, and precision of the proposed method were higher, indicating better detection results. Compared to the YOLOv5s network model, the P-R curve area of the proposed method was larger. This is because the proposed method integrates the cycleGAN network into the YOLOv5s target detection network for image enhancement for a better detection effect.

Simultaneously, the indices of the detection results of the two detection methods are shown in Table 1, where $mAP@.5:.95$ is the average of $mAP@.5$ – $mAP@.95$.

Table 1. Indices of test results

	YOLOv5s	Proposed method
Precision	0.653	0.696
Recall	0.286	0.292
$mAP@.5$	0.723	0.761
$mAP@.5:.95$	0.408	0.454

Table 1 compares the proposed method with the YOLOv5s network, and the proposed model had a precision, recall, $mAP@.5$, and $mAP@.5:.95$ of 0.696, 0.292, 0.761, and 0.454, respectively. The proposed method can achieve better vehicle detection at night using style transfer image enhancement, which provides a certain theoretical basis for developing driverless technology.

5. Conclusion

A night vehicle detection method based on style transfer image enhancement was proposed. A style dataset was obtained and divided based on the cycleGAN network. The style dataset was used to train and test the YOLOv5s network to obtain reliable nighttime vehicle image detection. The experimental results based on the BDD100K dataset showed that the transferred nighttime vehicle images were clear, and the vehicle information detected by the proposed method was comprehensive and accurate. Therefore, the detection effect was ideal. However, the YOLOv5s network has some limitations, such as the inaccurate detection of small targets. Therefore, future research must optimize the YOLOv5s network to improve its learning ability in complex environments and for small targets.

References

- [1] R. Matsuo, T. Tanigawa, K. Tomooka, A. Ikeda, H. Wada, K. Maruyama, and I. Saito, “The importance of screening for sleep disordered breathing in the prevention of motor vehicle crashes: the Toon Health Study,” *Juntendo Medical Journal*, vol. 66, no. 6, pp. 476-477, 2020. <https://doi.org/10.14789/jmj.2020.66.JMJ20-A03>
- [2] A. D. Puzanau and D. S. Nefedov, “Synthesis of algorithm of unmanned aerial vehicle detection by acoustic noise,” *Doklady BGUIR*, vol. 19, no. 2, pp. 65-73, 2021. <https://doi.org/10.35596/1729-7648-2021-19-2-65-73>

- [3] J. Lei, Y. Dong, and H. Sui, "Tiny moving vehicle detection in satellite video with constraints of multiple prior information," *International Journal of Remote Sensing*, vol. 42, no. 11, pp. 4110-4125, 2021. <https://doi.org/10.1080/01431161.2021.1887542>
- [4] S. Parvin, L. J. Rozario, and M. E. Islam, "Vision-based on-road nighttime vehicle detection and tracking using taillight and headlight features," *Journal of Computer and Communications*, vol. 9, no. 3, pp. 29-53, 2021. <https://doi.org/10.4236/jcc.2021.93003>
- [5] M. Shu, Y. Zhong, and P. Lv, "Small moving vehicle detection via local enhancement fusion for satellite video," *International Journal of Remote Sensing*, vol. 42, no. 19, pp. 7189-7214, 2021. <https://doi.org/10.1080/01431161.2021.1944694>
- [6] T. S. Kavva, Y. M. Jang, T. Peng, and S. B. Cho, "Vehicle detection and tracking from a video captured by moving host," *Indian Journal of Computer Science and Engineering*, vol. 11, no. 3, pp. 226-235, 2020. <https://doi.org/10.21817/indjcse/2020/v11i3/201103187>
- [7] L. Anuj, M. T. Gopalakrishna, C. Naveena, and Y. H. Sharath Kumar, "V-DaT: a robust method for vehicle detection and tracking," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 2, pp. 2492-2505, 2021. <https://doi.org/10.17762/turcomat.v12i2.2092>
- [8] J. Feng, D. Zeng, X. Jia, X. Zhang, J. Li, Y. Liang, and L. Jiao, "Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 116-130, 2021. <https://doi.org/10.1016/j.isprsjprs.2021.05.005>
- [9] V. Paidi, H. Fleyeh, and R. G. Nyberg, "Deep learning-based vehicle occupancy detection in an open parking lot using thermal camera," *IET Intelligent Transport Systems*, vol. 14, no. 10, pp. 1295-1302, 2020. <https://doi.org/10.1049/iet-its.2019.0468>
- [10] J. Q. Luo, H. S. Fang, F. M. Shao, Y. Zhong, and X. Hua, "Multi-scale traffic vehicle detection based on faster R-CNN with NAS optimization and feature enrichment," *Defence Technology*, vol. 17, no. 4, pp. 1542-1554, 2021. <https://doi.org/10.1016/j.dt.2020.10.006>



Jianing Shen <https://orcid.org/0000-0003-4050-0299>

He received master's degree of Computer Science and graduated from Jiangnan University in 2008. He is worked in Wuxi Taihu University as a lecturer. His research interests include artificial intelligence and software engineering.



Rong Li <https://orcid.org/0000-0002-3165-5195>

She received master's degree of Control Engineering and graduated from Jiangnan University in 2011. She is worked in Wuxi Taihu University as an associated professor. Her research interests include pattern recognition and cloud computing.