

# Nighttime vehicle detection algorithm based on image translation technology<sup>1</sup>

Yixun Wu, Taiyu Wang, Runze Gu, Chao Liu\* and Boqiang Xu

*School of Civil Engineering, Tongji University, Shanghai, PR China*

**Abstract.** In order to address the problem of decreased accuracy in vehicle object detection models when facing low-light conditions in nighttime environments, this paper proposes a method to enhance the accuracy and precision of object detection by using the image translation technology based on the Generative Adversarial Network (GAN) in the field of computer vision, specifically the CycleGAN, from the perspective of improving the training set of object detection models. This is achieved by transforming the existing well-established daytime vehicle dataset into a nighttime vehicle dataset. The proposed method adopts a comparative experimental approach to obtain translation models with different degrees of fitting by changing the training set capacity, and selects the optimal model based on the evaluation of the effect. The translated dataset is then used to train the YOLO-v5-based object detection model, and the quality of the nighttime dataset is evaluated through the evaluation of annotation confidence and effectiveness. The research results indicate that utilizing the translated nighttime vehicle dataset for training the object detection model can increase the area under the PR curve and the peak F1 score by 10.4% and 9% respectively. This approach improves the annotation accuracy and precision of vehicle object detection models in nighttime environments without requiring additional labeling of vehicles in monitoring videos.

**Keywords:** Vehicle object detection, CycleGAN, nighttime vehicle image dataset, deep learning, machine vision

## 1. Introduction

In recent years, the recognition of vehicles in surveillance videos has gradually become a focus of research and application in the fields of transportation and artificial intelligence. It has played a vital role in daily traffic flow detection [1], detection of vehicle behavior in motion [2], and real-time detection of vehicle loading [3]. Additionally, with the increasing severity of road traffic safety accidents, the demand for safety-assisted driving systems has become more urgent [4], which has raised higher requirements for vehicle recognition. In low-light environments, such

as at night, the recognition of vehicles is often interfered with by factors such as low illumination [5], resulting in poor operation of the detection system and safety issues such as traffic accidents. Therefore, establishing the correspondence between monitored images and vehicles in complex environments such as at night and achieving real-time and accurate detection of vehicles is a key issue in this research field [6, 7].

Object detection, as one of the main research directions in the field of computer vision, aims to classify and locate each target object in an input image [8]. This meets the need for vehicle detection using surveillance video images. The application of object detection algorithms in vehicle detection is a topic of widespread interest among scholars at home and abroad, and a widely used technology [9]. Apart from surveillance video images, object detection algorithms have also been applied in other types of vehicle video recognition research, includ-

<sup>1</sup>This work was supported by National Innovation and Entrepreneurship Training Project “Research on Real-time Monitoring Algorithm of Vehicle Load in Night Environment” (202210247066).

\*Corresponding author. Liu Chao, Associate Professor, School of Civil Engineering, Tongji University, PR China. E-mail: lctj@tongji.edu.cn

ing vehicle-mounted video image recognition [10], unmanned aerial vehicle-based vehicle detection [11], and perception and recognition of unmanned driving vehicles [12, 13].

A large number of dataset images are needed to construct vehicle object detection models and improve model accuracy, but the time and cost required for manual annotation of datasets can be enormous. Therefore, both academia and industry tend to use publicly available datasets, which provide annotated results for tens of thousands to millions of images. The most well-known datasets include the UA-DETRAC dataset [14], the BDD100K autonomous driving dataset [15], and the KITTI dataset [16], among others. Within this dataset system, there are also various excellent nighttime datasets, which have improved the effectiveness of nighttime vehicle detection to some extent. However, the quality of datasets obtained through shooting is not entirely satisfactory. They are mostly affected by low-light conditions at night and suffer from problems such as reduced details in the vehicle outline, complex light sources, dark lighting, and noise in the collected images [17]. Consequently, even after using these datasets to train object detection models for nighttime annotations, the improvement in nighttime detection results is limited, and there may still be some cases of false positives and false negatives. If there is a need to obtain large quantities of high-quality datasets under different lighting conditions, it is inevitable to carry out manual collection, selection, and annotation of dataset images, which will increase costs and reduce feasibility.

To address the existing problems with datasets, most studies have shifted to using the most prominent feature of nighttime vehicles, the paired headlights, as the basis for detecting vehicles [18]. However, in nighttime environments, complex situations such as motorcycle lights, street lamps, signage lights, and ground reflections must also be considered [19]. Industry experts have proposed various improvement methods, including using well-generalized light information to assist deep networks in nighttime vehicle detection [20], combining other sensors (such as radar) as auxiliary tools for vehicle detection [21, 22], using generative adversarial networks in conjunction with nighttime vehicle detection [23], using infrared image detection of vehicle contours [24], and using vehicle square wave pulse timing diagrams to detect traffic flow parameters [25]. These methods have all improved the accuracy and precision of nighttime vehicle detection to some extent. However, the

characteristics of object detection models lie in their real-time, fast, convenient, lightweight, and stable nature [26]. Making significant modifications to the original algorithms often leads to increased hardware requirements and decreased real-time performance, which is not conducive to practical applications.

An innovative approach to ensuring the lightweight and cost-effective nature of object detection models is the introduction of generative models to obtain high-quality nighttime vehicle datasets. This approach aims to improve the annotation accuracy of object detection models in nighttime environments without compromising their lightweight and convenient characteristics. Among the mainstream generative models, Generative Adversarial Network (GAN) [27] and Variational Autoencoder (VAE) [28] are the most commonly used. While VAE are simpler and more ingenious in their model design compared to GAN, they suffer from the drawback of generating slightly blurry images due to the adoption of the Mean Square Error (MSE) loss in their reconstruction loss function [29]. This limitation hinders the resolution of the problem of unclear vehicle outlines in the nighttime dataset addressed in this study. On the other hand, GAN, despite their higher hardware requirements during training, excel in achieving style transfer and ensuring clear outlines, thus demonstrating superior performance in acquiring datasets.

Therefore, this paper uses GAN as the dataset generation model, and uses CycleGAN [30] to establish a "day-to-night image translation model" that can convert the daytime style of a vehicle dataset into the nighttime style. Starting from the existing systematic daytime vehicle object detection datasets, this model is used to quickly obtain a large number of nighttime vehicle detection datasets, significantly reducing the time and cost required to construct datasets, while maintaining the lightweight and convenient nature and improving the accuracy and precision of nighttime vehicle detection, providing support for the development of intelligent transportation systems.

## 2. Research methods and ideas

### 2.1. Algorithm Selection: From GAN to CycleGAN

CycleGAN is based on the Generative Adversarial Network (GAN) paradigm and is mainly applied in the field of domain transfer. GAN is a neural network composed of a generator (G) and a discrim-

inator (D), among others. G achieves style transfer between domains, such as converting original images in domain A into new images that are similar to those in domain B, with B's image features. It is evident that G is essentially a function, in which the independent variables and function values are both high-dimensional vectors (images). The discriminator D determines the domain to which a given image belongs and returns a value that can be used to evaluate the generator's ability. Specifically, if the generated image (denoted as b) resembles images in domain B, the discriminator should give a high score (up to 1), otherwise, it should give a low score (minimum 0).

A typical GAN network requires the coordinated action of a generator and a discriminator. During the iterative process, these two components function like predator and prey and compete with each other: the generator aims to create images that are indistinguishable from real ones, while the discriminator should effectively differentiate between real and generated images. When the parameters of the discriminator are fixed, the generator will strive to create more realistic images to deceive the discriminator. Conversely, when the parameters of the generator are fixed, the discriminator needs to enhance its ability to distinguish between real and generated images. Through the iterative learning process, these two components evolve gradually and eventually achieve the desired effect.

In traditional domain transfer models such as pix2pix [31], paired training data is generally required, which means images need to be of equal size and correspond pixel by pixel. However, it is often difficult to find a set of images that have identical content but different styles. To solve this problem, Zhu et al. [30] proposed the CycleGAN algorithm, which does not require strict correspondence between the data in the two domains, making it more widely applicable. Furthermore, lightweight network structures proposed by Wang Rongda et al. [32] make CycleGAN easier to use. In this study, we adopt the CycleGAN algorithm to train an image translation model.

CycleGAN has a "cyclic" network architecture, as shown in Figure 1, with two generators, G1 and G2, between domains A and B, which respectively convert images from domain A to domain B and from domain B to domain A. There are also two discriminators, D1 and D2, which use real images from domain A and domain B as a reference to evaluate the similarity between the generated images by G2 and G1,

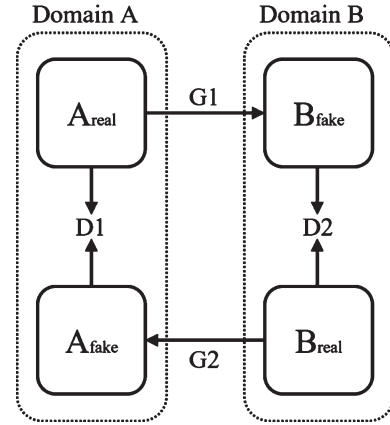


Fig. 1. CycleGAN translation process schematic diagram.

respectively, and give a score (0-1) to reflect the quality of the generated images. Specifically, for images in domain A, D1 should always give a full score, and for images in domain B, D2 should also always give a full score. This is evident that the original images in the dataset should be considered real. This characteristic can play a role in subsequent loss function construction.

## 2.2. Construction of Loss Functions

In the circular structure shown in Figure 1, it is necessary to introduce a loss function to evaluate the performance of the network, which is an important indicator for assessing the effectiveness of the training process [33]. In this study, a classic system of loss functions was adopted, starting with the conventional Loss\_GAN:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)}[\log D_Y(y)] + E_{x \sim p_{data}(x)}[\log(1 - D_Y(y))] \quad (1)$$

The presence of (1) increases the authenticity of the generated images. In this study, an exponential fitting was applied to this loss function to evaluate the authenticity of the translated images by the model, which will be presented in Section 4.

In CycleGAN, in addition to the loss function of the GAN network, Loss\_GAN, there are also Loss\_cycle and Loss\_identity.

Loss\_cycle, also known as cycle consistency loss:

$$L_{cycle}(G, F) = E_{x \sim p_{data}(x)}[||F(G(x)) - x||_1] + E_{y \sim p_{data}(y)}[||G(F(y)) - y||_1] \quad (2)$$

It characterizes the difference between the original domain and the new images generated through one generator and then another generator, thereby preventing the common mode collapse issue in traditional GANs (where generated samples converge, and all input images map to one output image). The goal of image translation is to achieve a style transfer effect while preserving the content, meaning that a bijective relationship should be established between the two domains. This is where  $\text{Loss\_cycle}$  comes into play: ideally, we should have  $G1(G2(x))=x$  and  $G2(G1(x))=x$ , meaning that the images returned after passing through the two generators should be as consistent as possible with the original images.

In contrast,  $\text{Loss\_identity}$  is much simpler:

$$L_{\text{Identity}}(G, F) = E_{y \sim p_{\text{data}}(y)}[||G(y) - y||_1] + E_{x \sim p_{\text{data}}(x)}[||F(x) - x||_1] \quad (3)$$

When images from domain A in the training dataset are input into G2, which is responsible for generating images with the style of domain B, the result should be the original image, i.e., the input and output images should be identical. Similarly, when images from domain B are input into G1, the output should also be identical to the input.  $\text{Loss\_identity}$  ensures that while the image style is transformed, other elements such as color tone are not excessively altered.

Normally, during the model training process, the loss function value for each iteration can be generated, and theoretically, the convergence of these values over time can be used to judge whether the training has achieved a good translation effect. However, this result cannot determine the quality of the generated images or guarantee that the generated images meet the desired criteria. Additionally, since this study only uses the generator (G) and not the discriminator (D), the loss function of the discriminator ( $\text{Loss\_D}$ ) will not be discussed here.

### 2.3. Core modeling ideas

As mentioned above, by providing two datasets of different style domains for training CycleGAN, it can achieve image style transfer while maintaining the content unchanged. Therefore, theoretically, using datasets from two different lighting environments, daytime and nighttime, to train the model can achieve the conversion of the daytime style of the vehicle dataset to the nighttime style, while keeping the element positions in the image unchanged. For the former case, as the daytime vehicle detection dataset has become a system, it can be used to

expand the nighttime dataset and serve as a supplementary training set for training the object detection model, directly improving the accuracy and precision of the object detection model in a nighttime annotation. Moreover, using the same image with two different styles indirectly enhances the robustness of object detection. In the latter case, since the annotated bounding box positions in the labeled dataset depend on the element positions in the image, this effect enables the direct use of the annotation box code as a training set for the object detection model, completely saving the significant human and time costs required for annotating the dataset.

Therefore, the approach of this paper is to use datasets of different quantities from daytime and nighttime lighting environments of vehicles to establish a specific image translation architecture, improve the deep learning method of image feature matching, and train an image translation model that can convert the daytime vehicle dataset into nighttime. By comparing and testing different translation models, the optimal translation model is selected to achieve the desired style transfer of annotated images, providing a labeled nighttime dataset for training object detection algorithms suitable for nighttime environments, and ultimately improving the accuracy and precision of nighttime vehicle detection. The specific model, algorithm, and process used in this study are shown in Figure 2.

### 2.4. Evaluation metrics for vehicle dataset

Since the loss function cannot determine whether the generated nighttime dataset meets the requirements for nighttime vehicle detection, other evaluation metrics need to be introduced.

#### (1) Visual Inspection Evaluation

In this study, a multi-model comparison approach was adopted. Considering that GAN networks often produce distorted or unevenly toned images, comparing and evaluating the visual translation effects of different models is an effective way to conduct preliminary screening of the models.

#### (2) PR Curve

The PR curve is commonly used to describe the performance of a learning machine in machine learning. It is composed of precision (y-axis) and recall (x-axis) and is used to compare precision at different recall rates. Before calculating the two parameters, a standard is set for the detection evaluation function ( $\text{IoU}$ ) of the object detection model to calculate the data and evaluate the quality of the annotations.

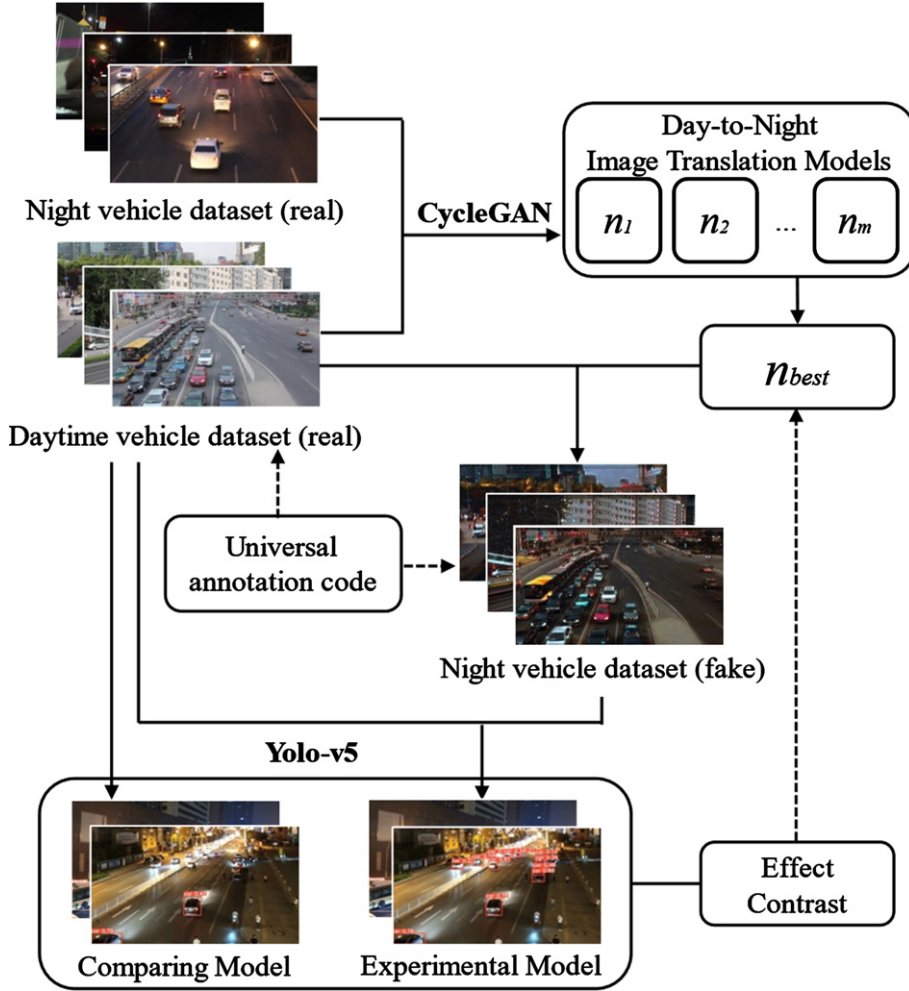


Fig. 2. Schematic diagram of the algorithm flow.

Precision and recall are calculated using the following formulas:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \quad (4)$$

Where  $P$  and  $R$  represent precision and recall, respectively. In this study,  $IoU=0.5$  was selected as the annotation standard, with  $TP$  representing the number of predicted boxes with  $IoU>0.5$ ,  $FP$  representing the number of predicted boxes with  $IoU\leq 0.5$ , and  $FN$  representing the number of boxes with no predictions. These parameters will be calculated during the testing of the object detection model. Generally, the closer the overall PR curve is to the upper right corner of the graph, or the larger its area, the better the annotation performance of the object detection model relative to the target.

### (3) F1 Curve

The F1 curve represents the numerical changes in the  $F1$  score (y-axis) across different confidence thresholds (x-axis). The  $F1$  score is calculated using the following formula:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (5)$$

Where  $P$  and  $R$  represent the calculated precision and recall. The  $F1$  score is in the range of  $[0,1]$ , and the closer it is to 1, the better the annotation performance of the object detection model. The F1 curve can visually represent the balance between precision and recall and can evaluate the performance of the object detection model by the peak value and corresponding confidence threshold. It can also be used to design the optimal confidence threshold for the model's operation.

### 3. Training of Daytime-to-Nighttime Translation Model

#### 3.1. Selection of vehicle datasets: BDD and UA-DETRAC

Adequate datasets are essential for building and testing the translation model. To train the translation model multiple times and achieve whole-vehicle recognition under different lighting and weather conditions, a large-capacity vehicle dataset is required. Therefore, this study selected two public vehicle datasets, BDD [15] and UA-DETRAC [14], and also collected nighttime vehicle images in locations such as YanJiao, Chongqing, and Xi'an to be used as datasets for model building and testing.

Regarding the training of the translation model, this study mainly used the BDD dataset for training. To distinguish it from the dataset used for model training and to facilitate subsequent verification work, the UA-DETRAC dataset was also selected for testing the translation model.

#### 3.2. Model training environment

The model was trained on a platform consisting of an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 16GB of memory, an NVIDIA GTX 1660Ti graphics card, and the PyTorch deep learning framework. The CUDA version used was 11.6.

#### 3.3. Comparative experiments of multiple models

##### 3.3.1. Preparation for training

The collected dataset samples were divided into two groups, daytime and nighttime, and named trainA and trainB, respectively, as the training sets. The number of images in both training sets can be adjusted, and there are no requirements for image format, size, or correspondence between the two training sets. However, since the pixel composition of the images is diverse, with a minimum resolution of  $960 \times 540$ , and considering the hardware requirements of the image translation model training process, the images used for training were preprocessed and resized to a uniform size of  $640 \times 360$  pixels.

To ensure smooth testing of the model, a fixed set of test data was prepared by selecting some daytime and nighttime images, and testA and testB were created as two separate testing datasets. By applying different translation models, multiple translation results of



Fig. 3. Original image used in model testing.

the same image were obtained, allowing for a visual evaluation of the model's effectiveness by observing the images.

During training, the loss value of the aforementioned image translation model can be obtained by calculating the difference between the translated sample image and the predicted translated image. Visualizing and analyzing the loss values is one of the indicators for evaluating the translation effect and processing the data.

##### 3.3.2. Construction of comparative models

Considering that the effect of the image translation model mainly depends on the number of images in the training set, and the capacity of the training set can be adjusted, we fixed the number of images in trainA and trainB to be the same (capacity denoted as  $n$ , i.e., the number of images in both training sets is  $n$ ), and conducted multiple rounds of training by changing the value of  $n$  to obtain different translation models. Then, we compared their effectiveness to explore the optimal number of training set images for achieving better translation results.

Firstly, based on the reference of the training set capacity of other types of image translation models, training was conducted with  $n=50, 100, 200$ , and  $300$ , resulting in three image translation models. The translation effects were tested using the models and the test dataset in section 3.3.1, and the original and translated images are shown in Figures 3 and 4, respectively. Figures 4(a), (b), (c), and (d) show the translation results of the image translation model corresponding to  $n=50, 100, 200$ , and  $300$ , respectively.

Comparative tests showed that the clarity and authenticity of the images obtained decreased with smaller values of  $n$ , mainly due to the mixed color distribution, which had a lower impact on the vehicle color and contour. This issue can be addressed by



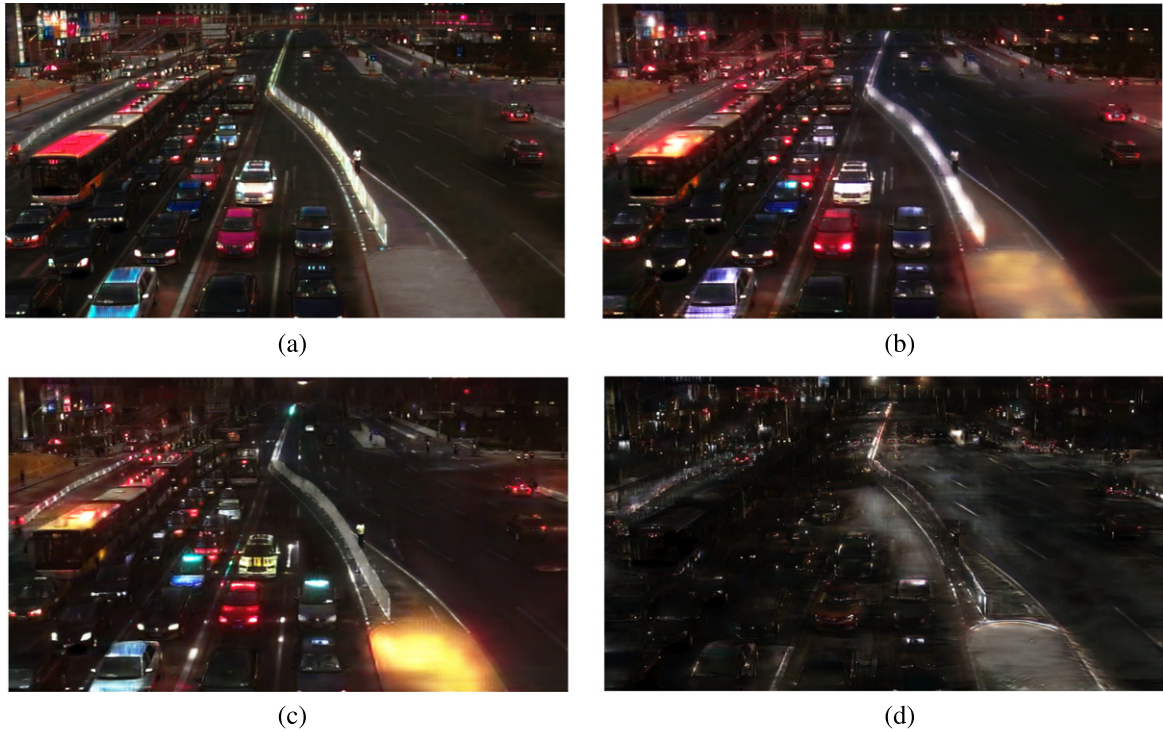


Fig. 4. Case image of the translation results of the first comparison experiment.



Fig. 5. Translated image presents "supersaturation" phenomenon.

reducing the depth of the translation model training or reducing the training set capacity. However, as  $n$  gradually increases, although the color distribution blur slightly decreases, the translated images gradually become oversaturated (as shown in Figure 5), and the internal composition elements of the images are gradually destroyed, mainly reflected in the vehicle's own features and image clarity. Even under the model with  $n=300$ , the image elements translated are completely destroyed, and such images are obviously not suitable for selection in the dataset.

The first set of experiments showed that for image translation models trained on vehicle datasets, when

the training set capacity  $n$  was large, the translation effect and image quality were actually unsatisfactory. Therefore, additional reference groups with smaller numbers of images ( $n=10, 20, 30$ , and  $40$ ) were added for further comparative experiments, with other parameters remaining the same as described above. The translation effects are shown in Figure 6, where (a), (b), (c), and (d) represent the translation results of the image translation models corresponding to  $n=10, 20, 30$ , and  $40$ , respectively.

It is easy to observe that as the training set capacity decreases, the phenomena of color distribution blur in the translated images gradually weaken, and the clarity and authenticity of the images are significantly improved. After comparison, the relatively optimal translation model was selected as the  $n=10$  group, which is suitable for style transfer in the dataset. Ideally, the training effect of CycleGAN should satisfy a loss function approaching 0. Therefore, the three loss functions generated during the model training process were plotted as a line chart. The Loss\_GAN function was smoothed using the Savitzky-Golay filtering method (SG smooth), and exponential curve fitting ( $y = ae^{bx}$ ) was used to analyze the loss\_GAN function and evaluate the model's performance through trend analysis and convergence analysis. Figure 7



Fig. 6. Case image of the translation results of the second comparison experiment.

shows the results, where Figure (a) shows the three evaluation indicators of the loss function, and their specific meanings are as described in section 2.2, and Figure (b) shows the loss evaluation and fluctuation trend of the model's image generator authenticity.

By observing and fitting the loss function curve, it can be concluded that the loss function of the  $n=10$  model has a clear convergence trend, indicating the effectiveness of the training and translation.

Furthermore, to further demonstrate the reasonableness of adopting the  $n=10$  model, the Loss\_GAN function graphs of the corresponding translation models for  $n=20, 30, 40$ , and  $50$  were subjected to the same smoothing and fitting processes, as shown in Figure 8. Figures (a), (b), (c), and (d) show the processing results for the  $n=20, 30, 40$ , and  $50$  translation models, respectively. By comparing the trend of the smoothed curves and the positivity or negativity of the independent variable coefficients in the exponential regression formula, it can be concluded that the Loss\_GAN functions of these four models did not converge. Moreover, compared to the  $n=10$  model, as the size of the training set increased, the loss function curve of the translation model fluctuated more dramatically, and there was a sudden change in the value of the loss function in the  $n=50$  model. These phe-

nomena indirectly confirmed the conclusion that the training and testing effects of the translation model were poorer when the size of the training set was larger.

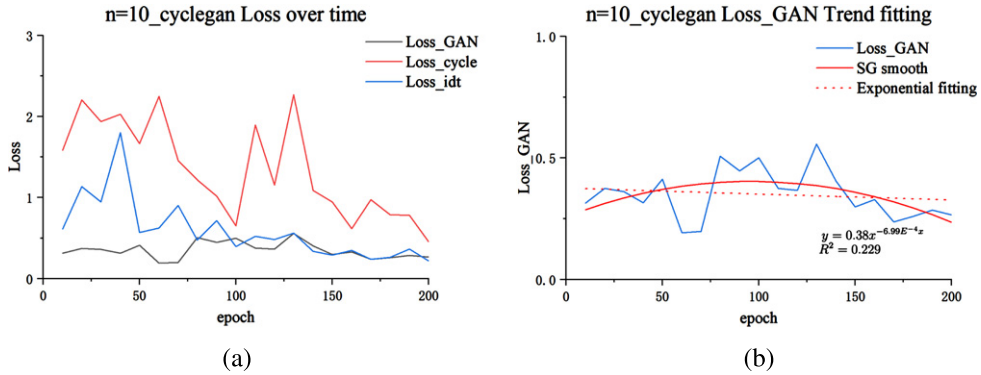
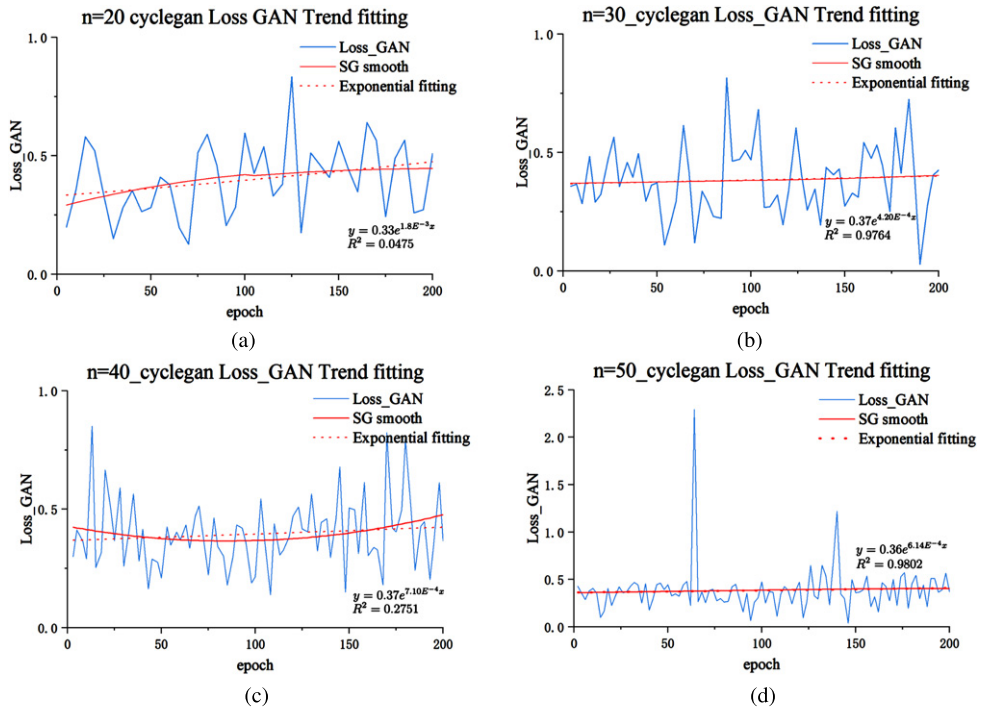
## 4. Model evaluation

### 4.1. Intuitive performance evaluation

Through the above comparative experiments, a good day-to-night image translation model was obtained. To further ensure the reliability of the model's translation effect, the experiment again used the model to perform style translation on other images in the BDD dataset, resulting in a large amount of nighttime vehicle dataset, as shown in Figure 9, where (a) and (c) are the original daytime dataset images, and (b) and (d) are the nighttime dataset images obtained through translation.

From an intuitive perspective, the translated dataset has achieved a relatively ideal conversion of lighting style, well simulating the monitoring perspective under nighttime scenes, and the image quality is good. Moreover, the elements in the images such as vehicles, non-motorized vehicles, obstacles, and



Fig. 7.  $n=10$  Image translation model loss function value image.Fig. 8.  $n=20, 30, 40, 50$  Image translation model Loss\_GAN function value image.

lane markings are basically preserved (which is also a characteristic of the image translation model), meeting the requirement of directly applying the annotation box code from the original daytime style dataset to the nighttime dataset.

#### 4.2. A Nighttime Vehicle Detection Model Based on YOLO-v5

The ultimate goal of the translation model is to optimize the vehicle object detection model, that is, to improve the recognition accuracy of the vehicle

detection model based on the YOLO object detection algorithm. In addition, since CycleGAN only changes the style of the image without changing the position of the elements in the image, the position of the vehicle bounding box in the translated monitoring image remains unchanged. Therefore, after translating the annotated image data into nighttime style, it can be directly used for training the object detection model. Therefore, in this paper, we choose to use the daytime dataset and the translated nighttime dataset to train the YOLO-v5 model, and compare it with the model trained only on the daytime dataset.

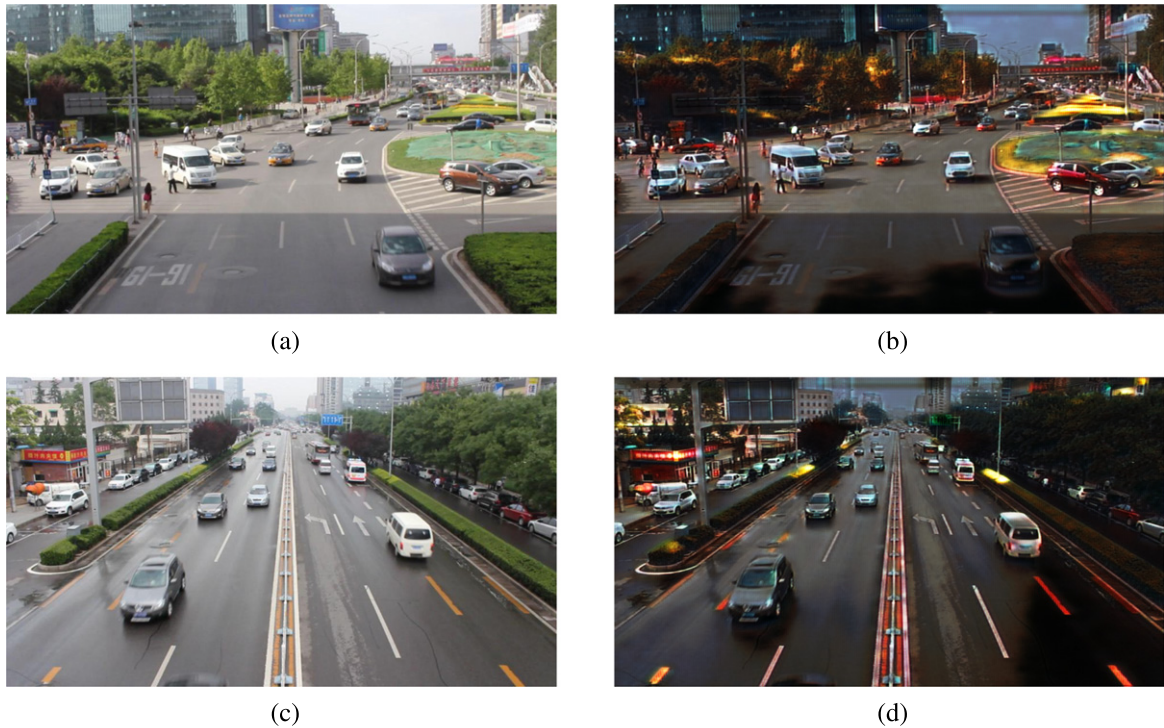


Fig. 9. Model translation effect diagram.

Firstly, the project team used the translation model to translate the UA-DETRAC vehicle dataset mentioned above and obtained the DETRAC dataset in a nighttime style. Then, using the YOLO-v5 algorithm and controlling for the same size of the training set, the team trained 30,000 original UA-DETRAC daytime vehicle dataset images to obtain the **comparative model** for object detection. The team also trained a model on a mixed vehicle dataset consisting of 30,000 images, including some daytime images and some translated nighttime images, resulting in the **experimental model** for object detection. Comparing the nighttime annotation effects of the comparative and experimental models allows for a direct evaluation of the quality of the dataset.

Using the two models, the same nighttime road vehicle dataset was annotated and the results were compared, resulting in partial schematic diagrams as shown in Figure 10. Figures (a) and (c) show the annotation effects of the comparative model, while Figures (b) and (d) show the annotation effects of the experimental model.

The rectangular annotation boxes in the figures represent the annotation effects of the object detec-

tion model on the images, while the labels above the annotation boxes show the category of the annotated objects and their confidence score. The comparison results are shown in Figure 11. Figures (a) and (b) respectively show the PR curves and F1 curves of the two models.

The PR curve graph records the area enclosed by the PR curve and the coordinate axis of the two models, while the F1 curve graph records the highest F1 score achieved by the two models and its corresponding confidence score, with the peak points labeled on the curve. Compared to the comparative model, the experimental model's AP and F1 score peak values in the PR curve and F1 curve respectively increased by 10.4% and 9%, and both curves showed a significant overall improvement, successfully demonstrating the improved effect of the translated nighttime dataset on the object detection model's nighttime annotation.

Through a comparative experiment of the object detection model, this study successfully evaluated the quality of the nighttime dataset obtained through translation and concluded that the translated nighttime dataset can be used as a training set for vehicle

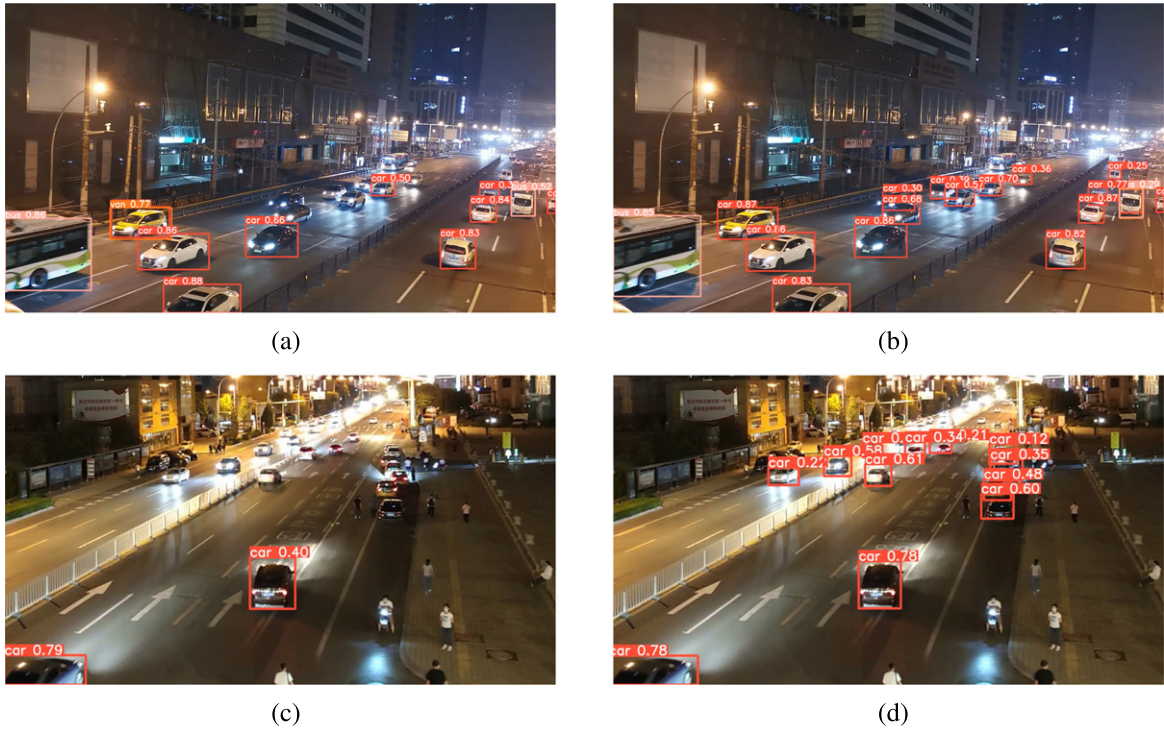


Fig. 10. The annotation effect diagram of object detection comparative model and experimental model.

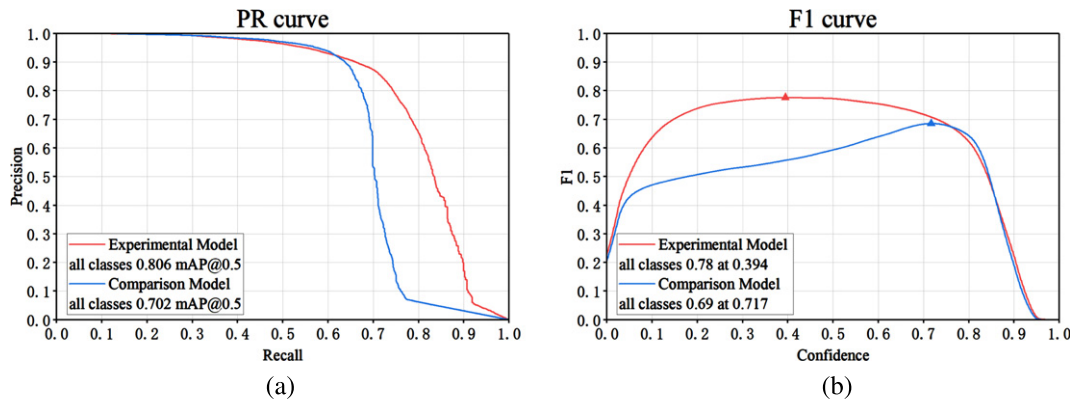


Fig. 11. PR curves and F1 curves for different models.

object detection models and can improve nighttime vehicle object detection effectiveness.

## 5. Conclusion

This paper adopts a comparative experimental approach, by changing the number of images in the training set of the image translation model and conducting translation tests separately, to select an image translation model and its corresponding training set

capacity that can effectively convert the daytime style of the vehicle dataset into nighttime style.

By using image translation models to perform style transfer on images, and thus obtaining nighttime datasets, this method has the advantages of simple model operations, fast running speed, and consistent image elements with the original dataset images. By simply changing the image names, the translated images can be applied to the training of object detection models. This satisfies the need to obtain a large amount of nighttime vehicle datasets

in a short period, greatly reducing human and time costs.

By using the translated nighttime vehicle dataset for training in the object detection algorithms, the detection confidence and accuracy of the generated object detection models have significantly improved. This validates the effectiveness of the translation model and demonstrates the optimization of object detection model detection in nighttime environments.

In addition, by adjusting and training the image translation model using training sets of different style types, the model can be applied again to obtain vehicle detection datasets in different styles or other complex environments, demonstrating good versatility and portability.

It should be noted that the experimental approach used in this paper is to adjust the number of images in the training set. However, for the translation model, the content of the training set images will also have an impact on the final translation effect, which can be a new approach to improving the model's translation performance. At the same time, this paper also verifies another conclusion: for vehicle detection datasets, increasing the pixel and quantity of training set images will lead to a "saturation" phenomenon in the model, which results in a decline in model performance instead of an improvement. Therefore, it is not advisable to optimize the quality of the translation model by increasing the training set capacity and image quality (including image pixels).

This study still has room for improvement. Due to the diversity of vehicle dataset images, the translation effect may vary for different input images based on the current translation model. For example, for daytime-style vehicle images with strong light conditions such as reflections from rain or direct sunlight, the translated nighttime dataset may have issues such as high contrast and mixed color distribution. If these images are used as datasets, they may affect the accuracy of the object detection model. Nevertheless, the improvement effect of constructing a dataset through image translation on nighttime annotation is still highly advantageous.

## References

- [1] C. Liyun, X. Yancong, H. Hongcheng and W. Weiping, Time series optimization prediction model of traffic flow in expressway service area, *Highway* **67**(06) (2022), 212–217.
- [2] L. Dong, H. Jia, S. Lukui, W. Song and L. Ji, Intelligent recognition of vehicle speed and vehicle type based on moving object detection and ResNet, *Journal of Beijing Jiaotong University* **43**(05) (2019), 10–19.
- [3] W. Chao, Y. Qingxiang, Z.Z. Sheng and Y. Zhongtao, Spatio-temporal parameter identification of vehicle based on machine vision, *Science Technology and Engineering* **23**(09) (2023), 3950–3955.
- [4] T. Haiqing, Night vehicle object detection based on yolov5s-se and data augmentation, *South China University of Technology* (2021).
- [5] C. Liu and B. Xu, A night pavement crack detection method based on image-to-image translation[J], *Computer-Aided Civil and Infrastructure Engineering* **37**(13) (2022), 1737–1753.
- [6] K. Xuan, Z. Jie, D. Lu and L. Yingkai, Research advances on vehicle parameter identification based on machine vision, *China Journal of Highway and Transport* **34**(04) (2021), 13–30.
- [7] H. Shuanhai, W. Anhua, Z. Zhao and Z. Yu, Research progress on intelligent detection technologies of highway bridges, *China Journal of Highway and Transport* **34**(12) (2021), 12–24.
- [8] X. Cong, Research on Vehicle Detection Algorithm Based on Improved R-FCN and SVM, Jilin University, 2021.
- [9] M. Chaochao and B. Hong, Vehicle detection research overview[C]// China Computer Users Association network application branch, *Proceedings of the 22nd Annual Conference on New Network Technologies and Applications*, 2018, 4.
- [10] Z. Su, Z. Xuelei, L. Feibin, N. Hao, J. Lianxin and W. Nan, Research on vehicle detection and tracking algorithm based on onboard video image, *Journal of Tongji University(Natural Science)* **47**(S1) (2019), 191–198.
- [11] L. Xu, S. Shiqi and Y. Xiaoqing, Real-time vehicle detection technology for UAV imagery based on target spatial distribution features, *China Journal of Highway and Transport* **35**(12) (2022), 193–204.
- [12] T. Qianwen, H. Zhaozheng, C. Hao, H. Gang and W. Xianglong, A study of vehicle perception and positioning localization: An overview of the 29th IEEE Intelligent Vehicles symposium, *Journal of Transport Information and Safety* **37**(02) (2019), 1–9.
- [13] M. Zhihui, Z. Jiali, W. Xin and L. Jun, Review of YOLO based target detection for autonomous driving, *Computer Engineering and Applications* **58**(15) (2022), 68–77.
- [14] 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, *Computer Vision and Image Understanding* **193** (2020), 102907.
- [15] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen and F. Liu, Bdd100k: A diverse driving dataset for heterogeneous multitask learning[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2636–2645.
- [16] A. Geiger, P. Lenz, C. Stiller and R. Urtasun, Vision meets robotics: The kitti dataset, *The International Journal of Robotics Research* **32**(11) (2013), 1231–1237.
- [17] Z. Ying, L. Wei, L. Yanyan and L. Shouxin, Research on vehicle detection algorithm based on low illumination environment, *Mechanical Engineering* **49**(07) (2022), 66–74.
- [18] R. O'Malley, E. Jones and M. Glavin, Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions, *IEEE Transactions on Intelligent Transportation Systems* **11**(2) (2010), 453–462.
- [19] Y.-L. Chen, B.-F. Wu, H.-Y. Huang and C.-J. Fan, A real-time vision system for nighttime vehicle detection and traffic surveillance, *IEEE Transactions on Industrial Electronics* **58**(5) (2011), 2030–2044.



- [20] H. Danyang, Z. Zhiheng, D. Ming and L. Zhihao, Night-time vehicle detection based on direction attention network and bayes corner localization, *Journal of Intelligent and Fuzzy Systems* **41**(1) (2021), 783–801.
- [21] X. Zhou and S.R. Chen, A vehicle detection algorithm based on multi-feature fusion and cascade classifier, *Modern Computer* (17) (2018), 38–43.
- [22] J. Lisheng, C. Lei and C. Bo, Leading vehicle detection at night based on millimeter-wave radar and machine vision, *Journal of Automotive Safety and Energy* **7**(02) (2016), 167–174.
- [23] L. Hui, B. Pengfei, L. Shixiao, Q. Yuanfeng, Y. Zichuan and Z. Guofu, Night vehicle detection method based on vehicle light characteristics, *Electronic Measurement Technology* **43**(14) (2020), 89–95.
- [24] W. Zhangu, Research on Vehicle Detection Method under Adverse Weather Conditions, Jilin University, 2022.
- [25] G. Zhen, H. Yulin, Z. Jiangyu and W. Xuesong, Real-time detection algorithm of traffic flow parameters based on sequence diagram of vehicle square wave pulse, *Journal of Tongji University(Natural Science)* **48**(08) (2020), 1143–1148.
- [26] Y. Lei, T. Hai, C. Yanrong, G. Ren and W. Wenhuan, Improved YOLOv5 for road object detection in complex environments, *Computer Engineering and Applications* 1-14[2023-06-20]. <https://kns.cnki.net/kcms/detail/11.2127.TP.20230607.1151.004.html>.
- [27] I. Goodfellow, J. Pouget-Abadie and M. Mirza, Generative adversarial nets, *Neural Information Processing Systems* MIT Press, 2014.
- [28] P.D. Kingma and M. Welling, Auto-encoding variational bayes, *arXiv.org*, (2013), 1312.6114.
- [29] G. Bredell, K. Flouris and K. Chaitanya, Explicitly minimizing the blur error of variational autoencoders, *arXiv preprint arXiv*, (2023), 2304.05939.
- [30] J.Y. Zhu, T. Park, P. Isola and A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *IEEE* (2017).
- [31] P. Isola, J.Y. Zhu, T. Zhou and A.A. Efros, Image-to-image translation with conditional adversarial networks, *IEEE* (2016).
- [32] W. Rongda, L. Ningzhong, L. Qiangyi and S. Jiaquan, A lightweight image-to-image translation model based on GAN, *Computer Technology and Development* **31**(11) (2021), 52–57.
- [33] Z. Bingli1, Q. Haoran1, J. Shang, Z. Jieyu and W. Zhenghai, A method of vehicle detection at night based on retinanet and optimized loss functions, *Automotive Engineering* **43**(08) (2021), 1195–1202.