Highway Obstacle Recognition Based on Improved YOLOv7 and Defogging Algorithm

Mingliang Fan, Jing Liu(✉), and Jiaming Yu

Engineering School of Networks and Telecommunications, Jinling Institute of Technology, Nanjing 211169, China
liuj608@jit.edu.cn

Abstract. This study explores a haze removal algorithm based on an enhanced Laplacian operator and guided filtering to enhance image visibility and quality. This algorithm is applicable not only to images taken in hazy weather conditions but also incorporates intelligent transportation systems by employing YOLOv7 object recognition technology. This enables accurate detection and identification of traffic objects within the images. By integrating the improved Laplacian operator into the traditional guided filtering framework, this approach effectively mitigates issues of image blurring and decreased contrast caused by haze, thereby enhancing image transparency. Simultaneously, the Integration of YOLOv7 into the system allows for rapid and precise detection of traffic objects, providing accurate data support for intelligent transportation systems. As a result, this study not only explores novel haze removal techniques in the field of image processing but also brings forth new technological applications for the advancement of intelligent transportation systems.

Keywords: highway · obstacle detection · YOLOv7 · deep learning

1 Introduction

With the continuous development and progress of human science and technology, computers have been widely used in various industries. At the same time, computer vision technology and its supporting equipment are also updated in parallel with the rapid development of computer technology. Image recognition technology [1], image super-resolution technology [2], video surveillance equipment [3], aerial photography drones [4] and so on have penetrated into every aspect of people's lives. Highway is an important part of modern transportation, which has the characteristics of fast driving speed and high passage efficiency, providing convenience for people's travel and logistics. However, highways also have some safety hazards, such as traffic accidents, congestion, intruding obstacles, etc. These problems not only affect the efficiency of transportation, but also threaten the safety of people's lives and properties. Therefore, how to effectively monitor and manage highways and improve the safety and intelligence of highways is an important topic in the current intelligent transportation field [5].

X. Chen et al. (Eds.): IoTaaS 2023, LNICST 585, pp. 22–34, 2025.
https://doi.org/10.1007/978-3-031-70507-6_3

Obstacle detection is a core task of highway monitoring and management and automatic driving, which detects obstacles on the road, such as vehicles, pedestrians, animals, etc., in real time by analyzing and processing images or videos of highway scenes, and outputs their categories and location information. Highway road surface temperature during the day, large temperature difference between day and night, very easy to form a dense haze, prompting a sharp decrease in local visibility, seriously affecting the normal imaging of computer vision acquisition equipment, reducing the clarity and contrast of the image, increasing the noise and interference in the image, resulting in a decrease in the accuracy and robustness of the target detection of obstacles, vehicles, and other objects, which affects the subsequent work of target classification and recognition.

However, the current target detection algorithms are not accurate in recognizing various types of target markers on highways under foggy weather, and the false detection rate and misdetection rate are high. Therefore, accurate and fast recognition of highway targets under foggy conditions has important research and application value. Based on the above problems, this paper combines the image de-fogging algorithm and the target detection algorithm to propose a novel highway obstacle detection method under foggy weather. The contributions of this paper are as follows:

(1) Combining the dark channel defogging [6] with the YOLOv7 [7] framework, the fog is successfully removed to avoid the interference of foggy weather on the recognition accuracy, which ensures the timeliness and at the same time, it can realize efficient and accurate obstacle and vehicle detection under different weather conditions.
(2) An image segmentation method based on Laplace operator [8] is proposed to divide the image into foreground and background parts, calculate their transmittance separately, and optimize the fused transmittance by guided filtering [9], so as to improve the accuracy and stability of the transmittance.
(3) The WIoU loss function is used to replace the original loss function of YOLOv7 to improve the generalization ability and convergence speed of the network to detect obstacles.
(4) A highway scene dataset is constructed, which contains highway images and videos under different weather conditions, as well as the corresponding obstacle labeling information, providing data support for the research of highway obstacle detection.

2 Related Work

According to the different processing methods, image defogging algorithms can be divided into three categories [10]: image enhancement, physical model defogging and deep learning model defogging. As shown in Fig. 1.

Image enhancement is a foggy image clarity means used in the early days, which is not based on the principle of image fogging recovery, only purely to improve the local or global contrast to alleviate the blurring of the image, so its recovery quality is general, and the effect on the real image is often not satisfactory; the physical model of defogging to the intrinsic principle of image degradation as a theoretical basis, through the tapped into the a priori knowledge or the use of the imaging model for the constraint processing to estimate the corresponding imaging parameters. Then, the estimated parameters are substituted into the model to recover a clear fog-free scene inversely. This type of method

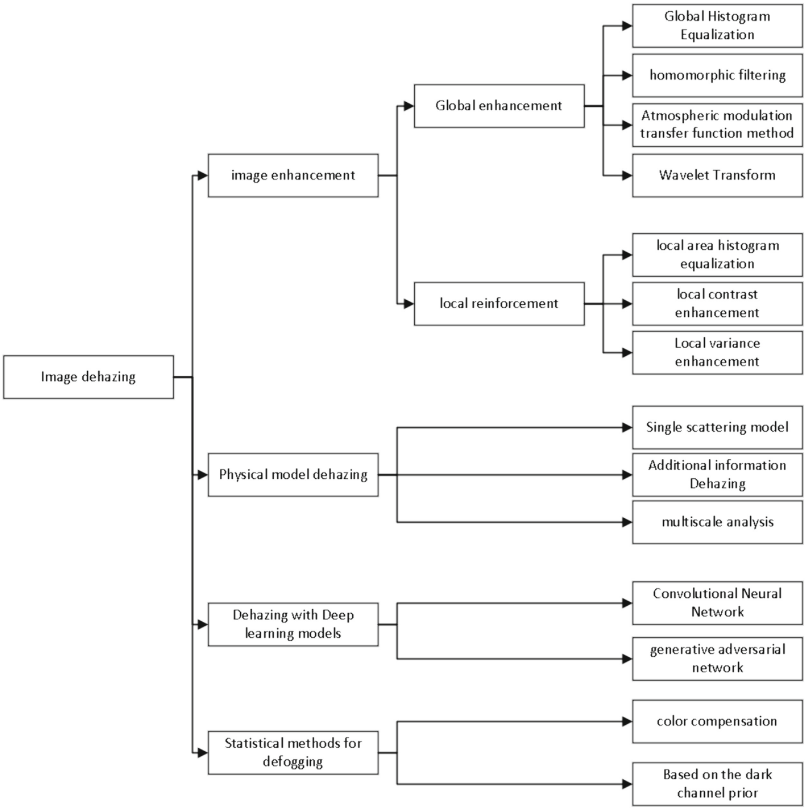

Fig. 1. Image defogging algorith

can obtain realistic defogging results for most images; deep learning model defogging benefits from the booming development of deep learning theory, and many scholars have designed a large number of deep learning model defogging methods in a data-driven manner.

Currently, the mainstream method is to calculate the atmospheric illumination and transmittance, and establish a physical model of atmospheric scattering for defogging. Chinese researchers He et al. proposed the dark channel prior principle in 2009 [6]. This method proposed the concept of dark channel for the first time, and compared with some complex image defogging algorithms, dark channel a priori principal defogging is easier to calculate. However, this type of method cannot deal with multi-scale or complex scenes, Luo Huilan et al. proposed an image defogging method based on multi-scale Retinex algorithm, by combining multi-scale Retinex and color recovery, this method can better deal with the details and color information in the image, so as to achieve a more accurate and natural image defogging effect. In 2013, He et al. introduced the concept of guided filter concept in combination with defogging processing to further enhance the defogging effect. These methods have the advantages of simplicity and real-time, but have limited effect on complex scenes and detail recovery in images, and such methods have insufficient edge protection may lead to blurring or distortion of edge information when performing smoothing operations. In 2011, ESL Gastal et al. proposed a domain transform-based method for realizing edge-aware image and video processing, which is more effective than the traditional bootstrap filtering method, this method can better preserve and enhance the edge information of the image. In 2011, Yu Jing et al. proposed a fast physical model-based defogging method. This algorithm performs haze removal by modeling the scattering model in the image using parameters such as transmittance and atmospheric light. This method can more accurately reproduce the real scene and has better results for complex images. However, this type of method has difficulties in calculating accurate transmittance. With the rise of deep learning, many researchers have started to apply it in haze removal algorithms. In 2019, Hodges C proposed about single image haze removal method using deep neural networks, which achieved better results, but there are limitations in coping with a wide range of haze types and the problem of haze removal under different environmental conditions.

Although the above dehaze algorithms can achieve good results in specific areas, when used in highway scenarios, there is a large deviation in the transmittance estimation, and the dehaze effect is poor, with obvious distortions, which needs to be further improved.

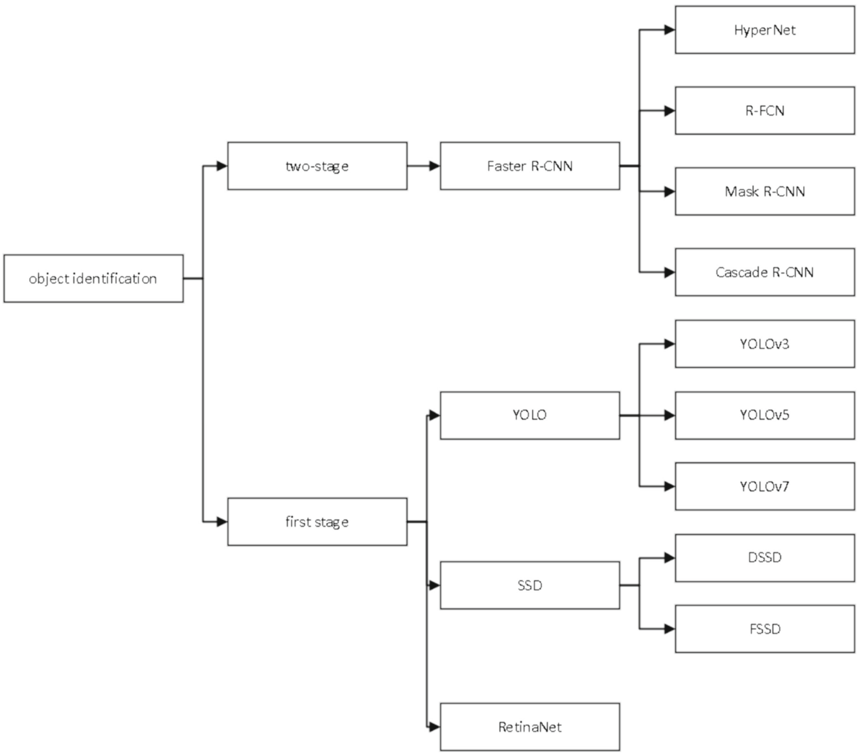

Fig. 2. Classification of target detection

According to the different detection processes, target detection algorithms can be divided into two categories: one-stage detection algorithms and two-stage detection algorithms. As shown in Fig. 2.

One-stage detection algorithm is to predict the category and location of the target directly on the image, without the generation of candidate regions, its advantage is fast, the disadvantage is low accuracy, especially for the detection of small targets. Representative algorithms include YOLO, SSD, etc. The two-stage detection algorithm is to generate candidate regions on the image first, and then classify and regress each candidate region, the advantage of which is high accuracy, the disadvantage of which is slow, especially for the processing of a large number of candidate regions. Representative algorithms include Faster R-CNN, Mask R-CNN, and so on.

In the past, scholars tended to improve only the deep learning methods or only the de-fogging algorithms, but did not consider combining the two algorithms closely. So the detection effect still cannot meet the demand when facing especially extreme weather.

3 Methods

3.1 General Flow of the Algorithm

In this paper, the defogging algorithm and the target detection algorithm are combined for obstacle detection. The detection flow is shown in Fig. 3.

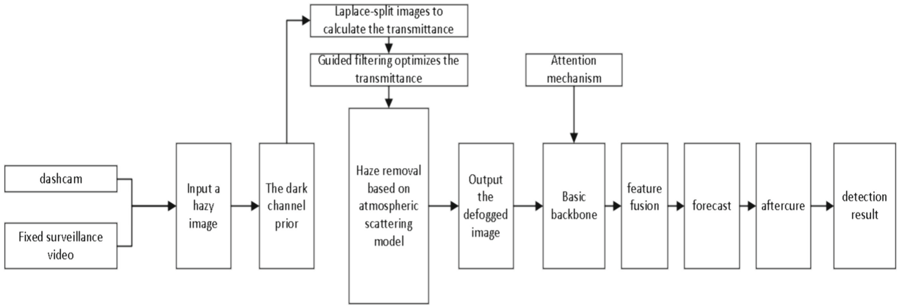

Fig. 3. General flowchart of the algorithm

As shown in the figure, the images captured by the camera on the highway will be transmitted to the computer for preprocessing followed by de-fogging, then advanced image features are extracted using deep neural networks, and finally, the detected obstacles can be visualized.

In this paper, highway obstacles are detected by de-fogging preprocessing combined with YOLOv7. The transmittance and atmospheric illumination are estimated by the dark channel a priori algorithm [10]. However, due to the special characteristics of the highway, most of the upper part of the image is the sky region, and the dark-channel a priori algorithm tends to distort the transmittance when facing the sky region. The use of Laplace operator to separate the sky and non-sky regions, respectively, calculate the transmittance of the two, and finally take the fusion of the two transmittances, to achieve a moderate transmittance, to achieve a more realistic de-fogging effect.

3.2 Image Preprocessing

For the transmittance of the sky region, because the brightness of the sky region is much larger than the brightness of the road surface, and the distribution of brightness is positively correlated with the depth of the scene, so we use the luminance intensity to estimate the transmittance, based on the luminance intensity model of the transmittance representation:

$$t_L(x) = \exp[-\beta L(x)] \tag{1}$$

$$L(x) = \frac{l(x)}{l^*} \times D \tag{2}$$

$l(x)$ is the brightness of the input image, l^* is the range of brightness in the current environment. D is the real scene depth range.

For non-sky regions: since the brightness of the road surface is much smaller than the area of the sky region, the transmittance can be estimated by the dark channel a priori algorithm.

$$t_{nosky'}(x) = 1 - \omega \min_{y\in\Omega(x)}\left[\min_c \frac{I^c(y)}{A^c}\right] \tag{3}$$

Assume that the overall luminance average of the input map is defined as L_1, the luminance average of the sky region is L_2, and the area of the number of pixels in the input map that are smaller than L_1 is S_1, and the area of the number of pixels that are smaller than L_2 is S_2. Fusion coefficient ρ and fusion transmittance $t(x)$ are defined as follows:

$$\rho = \frac{S_1}{S_2} \tag{4}$$

$$t(x) = \rho t_L(x) + (1 - \rho)t_{nosky'}(x) \tag{5}$$

When optimizing the transmittance map by using a fogged image as the guiding map and applying guiding filters, it results in the transmittance map containing a large amount of detailed information, although it is able to adequately capture the details of the changes in the guiding map. The transmittance represents the atmospheric scattering coefficient, which is usually considered constant over a localized region. Therefore, the transmittance is largely dependent on the depth of field of the scene. And in most cases, the depth of field is smooth, i.e., it has a similar and smooth depth of field in small localized areas that do not contain depth of field jumps. Therefore, the ideal transmittance should be similar and smooth in regions having the same depth of field, not containing detail information, and varying more in regions where the depth of field jumps or is discontinuous, i.e., retaining significant variations at object edge features. It can be seen that although the use of a grayscale map of a fogged image as a guiding map preserves clear edge features, it also results in too much detail and not enough smoothing in regions with the same depth of field. Therefore, this kind of guiding map still needs further improvement.

In this paper, the accuracy of transmittance can be improved by the guiding map with the following characteristics:

(1) Having the characteristics of foggy images at the edges.
(2) Smoothing at similar depth of field.
(3) As close as possible to the guided filter input.

The transmittance of foggy images is optimized using guided filtering by calculating the guided map based on the atmospheric light curtain model.

First, the RGB channel of the foggy image is minimized to obtain the image W. Then the local mean image T is obtained by applying bilateral filtering to W, so that T maintains the smoothness and retains the edge features. Calculate the local standard difference image between the image and the local mean image, perform bilateral filtering on this difference image and then differ from image T to obtain the quadratic difference image G. Use image G to obtain the atmospheric light curtain image, and finally derive the guiding map I_{guide}, the specific steps are as follows:

$$T(x, y) = Bilateral(W(x, y)) \tag{6}$$

$$G(x, y) = T(x, y) - Bilateral(|W - T(x, y)|) \tag{7}$$

$$\mathrm{I_{guide}} = 1 - \frac{max(\mathrm{min(G,W),0)}}{\mathrm{A}} \tag{8}$$

Secondly, using I_{guide} as the guiding map, the optimization of the transmittance $t(x)$ by using the guiding filter can make the optimized transmittance maintain the edge characteristics of the foggy image and tend to be smooth at the similar depth of field. The guiding filter is a local linear model between the guiding map I_{guide} and the filtered output $t_1(x)$, which can be considered as a linear transformation of all pixels in a window of size w_k centered at pixel k, as follows.

$$t_1(\mathrm{x}) = a_k I_{guide} + b_k, \forall i \in w_k \tag{9}$$

($\mathrm{a_k}$,$\mathrm{b_k}$) are linear transformation coefficients that are constant within the window w_k. Due to $\nabla t_2 = a_k \nabla I_{guide}$, which ensures that $t_1(x)$ has the same gradient information as I_{guide}. . The linear coefficients ($\mathrm{a_k}$,$\mathrm{b_k}$) are determined by minimizing the cost function that minimizes the difference between the output image of the guided filter and the input image. The cost function is given in the following equation.

$$E(a_k, b_k) = \sum\left(\mathrm{a_k}\left[\mathrm{I_{guide}}\right] + b_k - [\mathrm{t(x)}]_\mathrm{i}^2 + \varepsilon \mathrm{a_k^2}\right) \tag{10}$$

After defogging preprocessing as described above, the defogged image is subjected to target recognition by YOLOv7.

3.3 Image Classification and Recognition

The yolov7 network consists of four distinct components, input, backbone, neck and head.

The input module performs a series of preprocessing operations including resizing the input image to a standard size and augmenting the input data; the backbone module consists of several BConv modules, Extended ELAN (E-ELAN) module, and MPConv module, where the BConv module consists of convolutional layers, batch normalization (BN), and SiLU activation functions for extracting image features at different scales. The E -ELAN module guides the learning of different features through different computational blocks by operations such as expanding, shuffling, and merging bases without changing the original gradient paths. The MPConv layer enhances the generalization performance of the network by adding a Maxpool layer on top of the BConv layer; and between the head and the neck, there is a SPPCSPC module. In the SPPCSPC module, the inputs are split into multiple branches and undergo a MaxPool operation, and then merged by a cat operation to prevent image distortion; the Neck module employs the Path Aggregation Feature Pyramid Network (PAFPN) architecture, which is characterized by the efficient integration of multilevel semantic graphs using a pyramid-based framework. Finally, the Head module passes the REP structure and outputs predictions through subsequent convolutional operations.

The loss function is the most important aspect in the field of deep learning as it quantifies the difference between the predicted and actual outputs of the model and improves the performance of the model by optimizing the model parameters to effectively reduce this error. The precise selection and configuration of the loss function has a significant impact on the learning efficiency and overall results of the model.

The loss function of the YOLOv7 model consists of three distinct components: localization loss, confidence loss, and classification loss. These three losses are harmoniously integrated to form the total loss, which is obtained by weighted summation of the above components. The binary cross-entropy loss function is used to calculate the confidence loss and classification loss. Meanwhile, the localization loss is calculated utilizing the CIoU loss function.

$$LOSS = \alpha_{box} \times L_{box} + \alpha_{cls} \times L_{cls} + \alpha_{obj} \times L_{obj} \quad (11)$$

where the weight values of the three loss functions are respectively. $\alpha_{\{box\}}$

4 Experimental Setup

4.1 Introduction to the Dataset

At this stage, datasets for highway target detection are not comprehensive and specific, so in this paper, we use our own labeled road camera dataset to test the performance of the algorithm. We have collected a large number of road camera datasets and used software to label different targets. Our labeled dataset includes six types of labels, i.e., cars, trucks, buses, motorcycles, pedestrians, and obstacles. The specific data categories and tag numbers are shown in Table 1.

Table 1. Dataset introduction

Form	Data volume
enclosed carriage	12800
trucks	5880
tourist bus	3220
motorcycle	3350
pedestrians	3150
obstacle	1600

4.2 Assessment Metrics

Transmittance describes the degree of light attenuation in a hazy environment. It reflects the degree of influence of haze on the propagation of light, and areas with higher transmittance indicate less light attenuation and relatively clearer images, while areas with lower transmittance indicate greater light attenuation and blurrier images.

Peak Signal-to-Noise Ratio (PSNR) [11] and Structural Similarity Index (SSIM) [12] are used to evaluate the image processing algorithms as well as to assess the effectiveness of image compression, noise reduction and restoration techniques.

Mean Average Precision (mAP) is used to evaluate the overall performance of the target detection algorithm.

Precision (Precision) refers to the proportion of detected positive samples that are true positive samples. That is, the ratio of correctly detected targets to all detected targets.

Recall (Recall): recall is the proportion of true positive samples that are correctly detected. That is, the ratio between detected targets and all true targets.

Transmittance is generally used as a judgment metric for estimating fog concentrations. In the case of different degrees of fog warning, the transmittance estimation value can be different according to the actual situation and the method used. In general, the higher the transmittance estimate, the lighter the degree of fog in the image and the clearer the target objects in the image. A high transmittance (close to 1) indicates that the haze in the image is very slight or almost absent, and the details and colors of the target objects can be clearly observed; a medium transmittance (between 0.3 and 0.7) indicates that there is a certain degree of haze in the image, but it can still be recognized and observed, which is equivalent to an orange fog warning; and a low transmittance (close to 0) indicates that the details and colors of the target objects are very dense. When the details and colors of the target object have low transmittance (close to 0), it means that the haze in the image is very thick, and the details and colors of the target object are almost unobservable, and the whole image shows a fuzzy and grey effect, which is equivalent to the fog red warning.

4.3 Hardware Environment

In order to in order to validate the effectiveness of the proposed model, we conducted the training and evaluation of the neural network on the basis of computer hardware and software configurations, and used specific model hyperparameters. The specific configurations and hyperparameters are shown in Tables 2 and 3.

Table 2. Hardware configuration and environment

Software	Matrix
CPU	12th Gen Intel(R) Core i9-13900K 3.0 GHz
GPU	NVIDIA GeForce RTX4090 (24G)
Operating System	Windows11
CUDA	11.7
Python	3.10.9
Torch	1.13.1

Table 3. Experimental parameters

Parameter name	Parameters
Momentum	0.937
Weight decay	0.0005
Batch size	32
Learning rate	0.01
Image size	640 × 640
Epochs	300

4.4 Experiments and Results

The objective of the algorithm in this paper is to explore and determine the best combination of de-fogging detection algorithms and to demonstrate the facilitation of the de-fogging algorithm proposed in this paper on the task of de-fogging detection, hence the experiments are carried out using the control variable approach.

Preprocessing Module:

In this paper, the effect of defogging is analyzed by comparing the difference in transmittance as well as PSNR and SSIM values under different defogging algorithms.

The comparison of the transmittance of fogged and fogged pictures by different defogging algorithms is shown in Fig. 4.

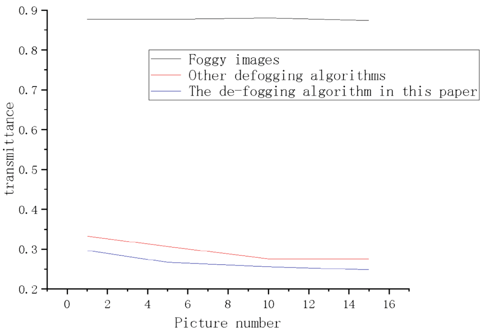

Fig. 4. Transmissivity values

The PSNR and SSIM values under different defogging algorithms are shown in Figs. 5 and 6:

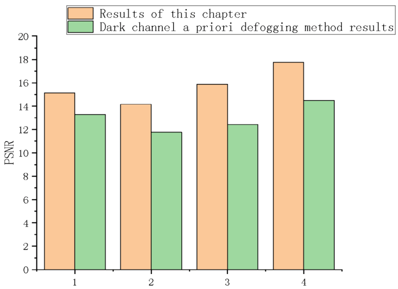

Fig. 5. PSNR values under different defog

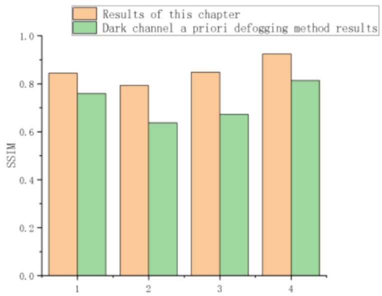

Fig. 6. SSIM values under different defogg

It can be seen that the de-fogging algorithm improved in this paper has better defogging effect than other de-fogging algorithms.

The de-fogging method proposed in this chapter is used for target recognition of highway video pictures, and the original picture, the picture before de-fogging, and the picture after de-fogging are compared, and the comparison experiments are shown in Figs. 7, 8, and 9. The data in the table is obtained on a test set of 1000 samples. The images after performing the defogging process are compared to the images before defogging, where the accuracy P after defogging is improved by 9.859% compared to the accuracy P before defogging, and the recall R after defogging is improved by 9.5238% compared to the recall R before defogging, Average AP after defogging is 12.008% higher compared to average AP before defogging.

As shown in Figs. 7, 8 and 9. In fogged images, the accuracy, recall and average AP value are lower than the values of the three data compared to the de-fogged images for target recognition after going through the de-fogging method in this chapter, compared to the fogged images, the de-fogging algorithm in this chapter has a good effect on target recognition.

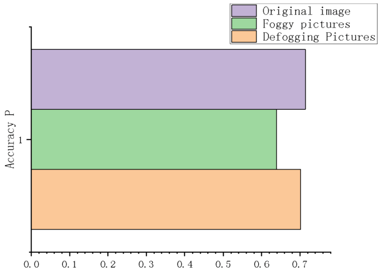

Fig. 7. Accuracy P

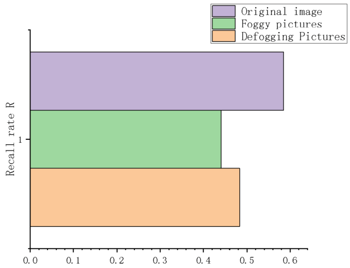

Fig. 8. Recall rate R

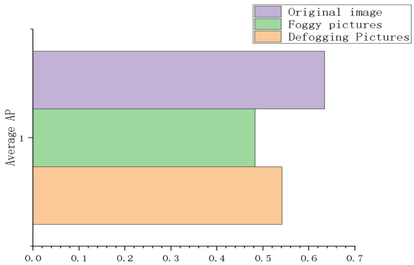

Fig. 9. Average AP

5 Conclusion

In this paper, we take foggy highway images as the object, and conduct in-depth research on the problems of inaccurate transmittance calculation, inaccurate target recognition, and low universality, which are commonly found in the current image defogging algorithms. In order to improve these problems, in the de-fogging algorithm, the method of dividing the image to calculate the transmittance and fusion respectively, and the method of optimizing the transmittance by guided filtering are proposed. In the target detection algorithm, it is proposed to use Wiou loss function to replace the loss function of YOLOv7 itself. After several experiments, the data processed by defogging using the method of this paper are closer to the expected data than those processed by other defogging algorithms. It fully demonstrates that the method in this paper is efficient and superior, and can be adapted to the needs of intelligent transportation.

Although the method based on deep learning and atmospheric scattering model proposed in this paper can obtain ideal results in most situations, the de-fogging effect for complex scenes and strong haze still needs to be improved, which is due to the insufficiently rich dataset. In the future, richer datasets will be collected and trained. In addition, the algorithm in this paper only de-fogs a single foggy image, while in modern smart city applications, real-time de-fogging is often required, and the future will study how to realize real-time video de-fogging.

Acknowledgement. This work was supported in part by the 2023 Jiangsu Province College Student Innovation Training Program (202313573086Y), and in part by the Natural Science Foundation of Higher Education of Jiangsu Province (Grant No. 17KJB510020).

References

1. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. **25**, 2 (2012)
2. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. IEEE IEEE (2015)
3. Alomari, A.H., Lebdeh, E.A.: Smart real-time vehicle detection and tracking system using road surveillance cameras. J. Transp. Eng. Part A Syst. (2022). https://doi.org/10.1061/JTEPBS.0000728
4. Kotaro, et al.: Advantages of unmanned aerial vehicle (UAV) photogrammetry for landscape analysis compared with satellite data: A case study of postmining sites in Indonesia. Cogent Geoscience (2018)
5. Cui, C.A.I.: Status analysis and Suggestions for the development of intelligent transportation in China. Highway Transp. Sci. Technol. Appl. Technol. Edn. **9**(6), 4 (2013)
6. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2010)
7. Redmon, J., et al.: You only look once: unified, real-time object detection. Comput. Vis. Patt. Recogn. IEEE (2016)
8. Wang, B., Yang, S.: Edge detection of gray image based on gauss-laplace operator. Comput. Eng. Appl. **39**(26), 3 (2003)
9. Chen, S., Ren, Z., Lian, Q.: Single image dehazing algorithm based on improved dark channel and guided filtering. Acta Automat. ICA Sin. **42**(3), 11 (2016)
10. Zhang, D., Ju, M., Qian, W.: Research status and Prospect of Image Dehazing algorithms. J. Nanjing Univ. Posts Telecommun. Natl. Sci. (2020)
11. Huang, H.F., Qing-Xin, C., JianKang, X.: Design of power supply noise and radiation free power—Ground plane for modern system in package. In: 2008 Asia- Pacific Symposium on Electromagnetic Compatibility and 19th International Zurich Symposium on Electromagnetic Compatibility. IEEE, 2008
12. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)