

Saliency Object Detection Method Based on Real-Time Monitoring Image Information for Intelligent Driving



Lei Yu^{1,2}, Hongwu Qin¹, Chao Zhang¹, Ju Wang¹, Ji Zou^{1*}

¹ Electronic Information Engineering College, Changchun University, Changchun 130000, China

² School of Automobile Electrical and Mechanical Engineering, Changchun Information Technology College, Changchun 130000, China

Corresponding Author Email: zouji@ccu.edu.cn

<https://doi.org/10.18280/ts.400318>

ABSTRACT

Received: 15 February 2023

Accepted: 27 May 2023

Keywords:

intelligent driving, real-time monitoring image, saliency object detection, YOLOv5s

Reducing traffic accident occurrences and enhancing road safety can be achieved through the processing of real-time surveillance image information for saliency object detection. Although existing saliency object detection methods based on real-time monitoring image information for intelligent driving have yielded certain results, there remain some shortcomings. In complex road environments, distinguishing between background and salient targets with existing methods proves difficult, resulting in false and missed detections. Consequently, this study investigates a saliency object detection method based on real-time monitoring image information for intelligent driving. The Visual Geometry Group (VGG) network discriminator in Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) is modified, and techniques such as spectral normalization (SN) are employed to improve the dynamic stability of training. Pixel-level image size amplification and feature enhancement are conducted on the salient objects in the dataset, providing a richer data foundation for subsequent real-time monitoring of saliency target detection and defect classification. The YOLOv5s algorithm is utilized as the identification network, and the original YOLOv5s backbone network is replaced with the MobileNetV2 network, significantly reducing network complexity and enhancing identification efficiency. The algorithm's performance in recognizing salient targets in real-time monitoring images for intelligent driving is further improved through network optimizer optimization and clustering algorithm adoption. The efficacy of the proposed method is substantiated by experimental results.

1. INTRODUCTION

With the rapid development of science and technology, intelligent driving technology has gradually become a research hotspot in the field of transportation [1-4]. The real-time monitoring image information plays a pivotal role in intelligent driving [5-11]. In order to ensure driving safety, intelligent driving needs to make rapid judgments and responses to the surroundings, therefore the processing speed and accuracy of real-time monitoring image information is crucial [12-18]. At the same time, the road environment is complex and changeable, and salient object detection technology can help intelligent driving systems better identify key information in complex environments. Through the processing of real-time monitoring image information, saliency object detection can reduce the incidence of traffic accidents and improve road safety [19-23]. Therefore, the research on the saliency object detection method based on real-time monitoring image information of intelligent driving will promote technological progress in related fields and provide more advanced auxiliary means for intelligent driving.

Li et al. [24] describes a target recognition algorithm applied to unmanned ground combat platforms for turning driving. A driving model for unmanned ground combat platforms turning at urban intersections has been established. The reliability of the driving model and the impact of radar

measurement accuracy on the algorithm were verified through simulation experiment. This algorithm solves the problem of target missing and target misidentification in adaptive cruise control (ACC) strategy during curve driving, and has the advantage of accurate positioning. Wang and Sun [25] proposed an intelligent transportation target recognition method based on support vector machines. The experimental results show that the target recognition method has strong classification and recognition capabilities.

Liu et al. [26] completed the design of vehicle target detection and statistical algorithms based on the characteristics of video images in green light duration. Firstly, a suitable image preprocessing algorithm was designed. Secondly, an improved mixed Gaussian background modeling algorithm is used for background subtraction to extract the target vehicle. At the same time, an external rectangular frame is used to identify the vehicle. Finally, the total number of vehicles in the entire video is calculated through calculating the number of vehicles running in the current frame. The experimental results show that the algorithm has a fairly high detection rate and certain practicality. *YOLO* is one of the object detection methods in deep learning, which has been used to detect traffic targets in real-time on conventional devices under general light to medium traffic density. Tian et al. [27] established a hybrid method that combines frame difference (FD) with *YOLO*. The results indicate that the accuracy of the bounding

box has been significantly improved, but the impact on detection speed is minimal. A real-time calculation model for target speed and distance based on a fixed monocular camera was also proposed. The method introduced can be used in the intersection-based traffic monitoring systems to warn drivers, and can also be used to help drivers avoid dangerous driving behaviors through on-board devices.

Although existing saliency target detection methods based on real-time monitoring image information of intelligent driving have achieved certain results, there are still some shortcomings. Many existing saliency target detection methods may not be fast enough to meet real-time requirements when processing large amounts of high-resolution image data. In complex road environments, existing methods may find it difficult to distinguish between background and salient targets, leading to false positives and missed detections. In order to overcome these shortcomings, this article conducted a research on saliency target detection methods based on real-time monitoring image information of intelligent driving. In section 2, the design of the *VGG* network discriminator in *ESRGAN* was designed and improved. The application of the techniques such as spectral normalization (SN) can enhance the dynamic stability of training, and pixel level image size amplification and feature enhancement can be performed on significant targets in the dataset, providing a richer data foundation for subsequent real-time monitoring of image saliency target detection and defect classification. In section 3, *YOLOv5s* algorithm is adopted as the identification network, and the original *YOLOv5s* backbone network is replaced by *MobileNetV2* network, which greatly reduces the network complexity and improves the network identification efficiency. In order to improve the performance of multi-class saliency target recognition, and further improve the algorithm's recognition performance of saliency targets in real-time monitoring images of intelligent driving, the network optimizers and clustering algorithms are optimized. The experimental results validate the effectiveness of the proposed method.

2. INTELLIGENT DRIVING REAL-TIME MONITORING IMAGE RECONSTRUCTION

Intelligent driving real-time monitoring images involve complex road environments with large degradation space, which requires powerful discriminators to distinguish between complex training outputs and real images. In order to better adapt to the characteristics of real-time monitoring images in intelligent driving, it is necessary to refer to existing mature methods and improve them to a more suitable network structure for this task. At the same time, as the saliency target recognition task requires accurate recognition of the edge information of the target and background, more accurate gradient feedback is needed to enhance the local details of the image generated by the network model.

In order to improve the accuracy and speed of saliency target recognition for real-time monitoring of intelligent driving, and improve the overall performance of the model, this article improves the *VGG* network discriminator design in *ESRGAN* to a more suitable *U-Net* network design for the task, ensuring that the model can still maintain high recognition performance in adverse weather or light changes. The optimized network structure for specific tasks may lead to increased training instability. The application of techniques

such as spectral normalization (SN) can enhance the dynamic stability of training, which improves model convergence speed and training effectiveness. The image size expansion and features enhancing at the pixel level for salient targets in the dataset provides a richer data foundation for subsequent real-time monitoring of salient target detection and defect classification in images.

In this section, in order to further improve the restored image quality of *GAN* network and achieve more realistic real-time monitoring image restoration for intelligent driving, this article eliminates all *BN* layers in the structure of the generator and fuses multi-level residual networks with densely connected residual dense blocks (RRDBs) to replace the original basic blocks in *SRGAN*. Figure 1 demonstrates the main structure of the generator.

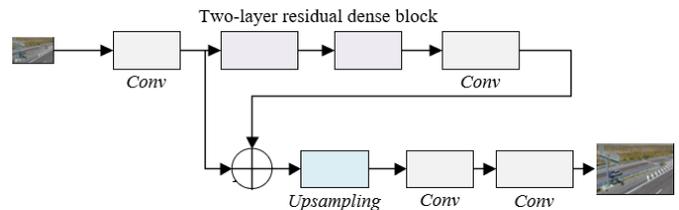


Figure 1. Main structure of generator

The reason for removing all *BN* layer operations is that the original *BN* (batch standardization) layer depends on the mean parameter and variance parameter of batch data during the characteristic regularization operation. During the testing process, it mainly relies on the estimated mean value and variance parameters of the test set. When there is a significant difference in data between the training set and the test set, the *BN* layer may introduce artifact interference, which affects image quality. At the same time, there are differences in the *BN* layer's processing of training and testing data, which may limit the model's generalization ability on different dataset. The removal of the *BN* layer can make the training process more stable and maintain consistent training results, thereby improving the quality of restored images while reducing the computational complexity of the model, simplifying the model structure, and reducing computational complexity.

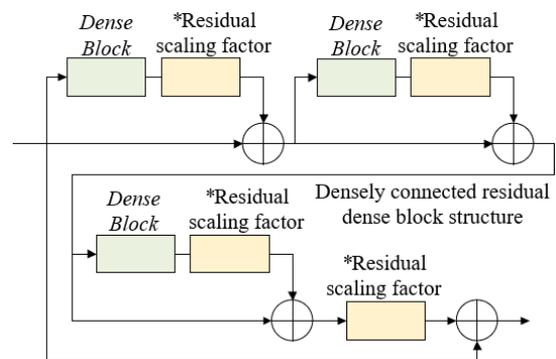


Figure 2. RRDB structure

Figure 2 and Figure 3 respectively show the *RRDB* structure and the *Dense Block* structure. There are four main reasons for fusing multi-level residual networks and densely connected residual-in-residual dense blocks (*RRDB*) to replace the original basic blocks in *SRGAN*. Firstly, *RRDB* has a deeper and more complex structure compared to the original remaining blocks in *SRGAN*, and model performance can be

improved through more hidden layers and network connections. Secondly, due to the dense connections in *RRDB*, its network capacity is more abundant and can capture more information and features, improving the model's representation ability. In addition, the introduction of residual scaling coefficients can adjust the residual before feeding it back to the forward function, effectively preventing non-stationary oscillations in the *GAN* network and improving model stability. In *ESRGAN*, the reduction of network space size and expanding of the channel size through the inverse operation of cancelling pixel through pixel can enable most calculations to be executed in small resolution space, thereby reducing computational resource consumption. The above operations can help improve model performance, enhance network capacity, prevent non-stationary oscillations, and simplify calculations, thus improving the quality of restored images and network stability, saving computational resources.

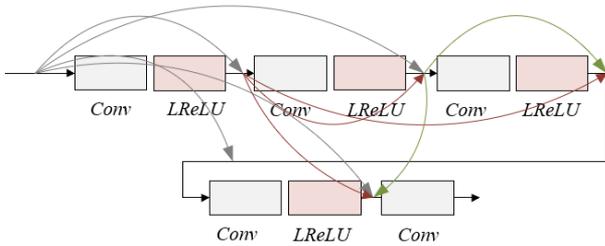


Figure 3. Dense Block structure

In the *W* section of the *GAN* network discriminator, the set of the output value of *W* is the probability where the real intelligent driving real-time monitoring image y_f tends to be more realistic than the false intelligent driving real-time monitoring image y_r .

$$\begin{aligned} RE: W(y_f) &= \xi(B(y_f)) \rightarrow 1 \\ FA: W(y_r) &= \xi(B(y_r)) \rightarrow 0 \end{aligned} \quad (1)$$

It is assumed that the probability of discriminative images being more realistic than false images is represented by $W_{Re}(y_f, y_r)$, and the probability of discriminative images being more false than real images is represented by $W_{Re}(y_f, y_r)$. The discriminator *W* probability calculation formula used in this article is as follows:

$$\begin{aligned} W_{Re}(y_f, y_r) &= \xi(B(y_f) - Q_{y_f}[B(y_r)]) \rightarrow 1 \\ W_{Re}(y_r, x_f) &= \xi(B(y_r) - Q_{y_f}[B(y_f)]) \rightarrow 0 \end{aligned} \quad (2)$$

Through the above formula, the standard discriminator *W* is replaced as the relative average discriminator W_{Re} . The standard discriminator in generative adversarial networks with super-resolution can be represented as $W(y) = \zeta(B(y))$, where the *Sigmoid* function is represented by ζ , the output of the discriminator without transformation is represented by $B(y)$, and the average operation on all false data in a small batch is represented by $Q_{y_r}[\cdot]$. The following formula gives the definition of loss function of identifier *W*:

$$\begin{aligned} GH_W^{Re} &= -Q_{y_f}[\log(W_{Re}(y_f, y_r))] \\ &- Q_{y_r}[\log(W_{Re}(y_r, y_f))] \end{aligned} \quad (3)$$

Assuming that the low-resolution image input into the *GAN* network is represented by y_j , the false image generated by the *GAN* network is represented by $y_r = T(y_j)$, and the loss function of generator *T* and discriminator *W* is symmetric, then there are:

$$\begin{aligned} GH_W^{Re} &= -Q_{y_f}[\log(1 - W_{Re}(y_f, y_r))] \\ &- Q_{y_r}[\log(W_{Re}(y_r, y_f))] \end{aligned} \quad (4)$$

In the actual network model confrontation generation training, the loss function of *T* is greatly affected by the gradient of y_f and y_r .

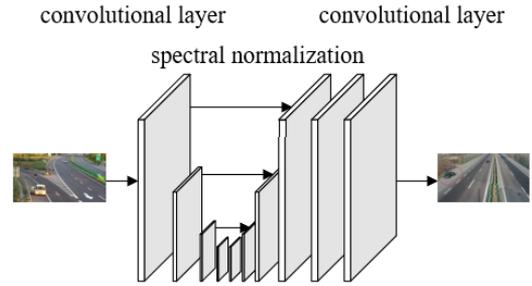


Figure 4. The discriminator structure used

Furthermore, the *VGG* discriminator is improved to a *U-Net* network design with skip connections, mainly because the *U-Net* network can output the true values of each pixel, and provide detailed per-pixel feedback to the generator, thus generating more realistic local textures. At the same time, through skipping connections to transfer information between different levels, local details of the image are enhanced, the quality of the restored image is improved, and more realistic saliency target restoration of real-time monitoring images for intelligent driving is achieved. The introduction of spectral normalization can stabilize training dynamics, alleviate boundary supersharpeness and artifacts caused by *GAN* training, and achieve a balance between local detail enhancement and artifact suppression, helping to generate more accurate gradient feedback, improve local details, improve image quality, enhance discrimination ability, and achieve network stability by introducing spectral normalization. Figure 4 shows the discriminator structure used.

Let $GH_1 = Q_{y_j} \|T(y_j) - x\|_1$, which is defined as the loss of 1-norm distance content between the restored image $T(y_j)$ and the target value, and the balance coefficients of different loss terms are expressed by η and γ . The following formula gives the total loss function of generator *T*, which is shown as follows:

$$GH_T = GH_{pe} + \eta GH_T^{Re} + \gamma GH_1 \quad (5)$$

3. SALIENT TARGET RECOGNITION FOR REAL-TIME MONITORING IN INTELLIGENT DRIVING

The method designed in section 2 of this paper enhances the feature representation of saliency targets through *ESRGAN*-based saliency target adversarial generation algorithm, which is limited in performance when the number of identification tasks is increased or categories is enriched. Multiple types of saliency targets may exist in the real-time monitoring image of

the intelligent driving. Therefore, it is necessary to adopt more powerful recognition network, such as *YOLOv5s*, to adapt to complex scene needs. Moreover, the processing of real-time monitoring images of intelligent driving demands high instantaneity, and reducing network complexity helps to speed up processing and meet real-time performance requirements. To this end, this paper adopts *YOLOv5s* algorithm as the identification network, and replaces the original *YOLOv5s*

backbone network with the *MobileNetV2* network to greatly reduce the network complexity and improve the identification efficiency of the network. In order to improve the recognition performance of multi-class saliency targets, the recognition performance of the algorithm for real-time monitoring image saliency targets of intelligent driving is further improved by optimizing network optimizers and clustering algorithms. Figure 5 shows the optimized network structure.

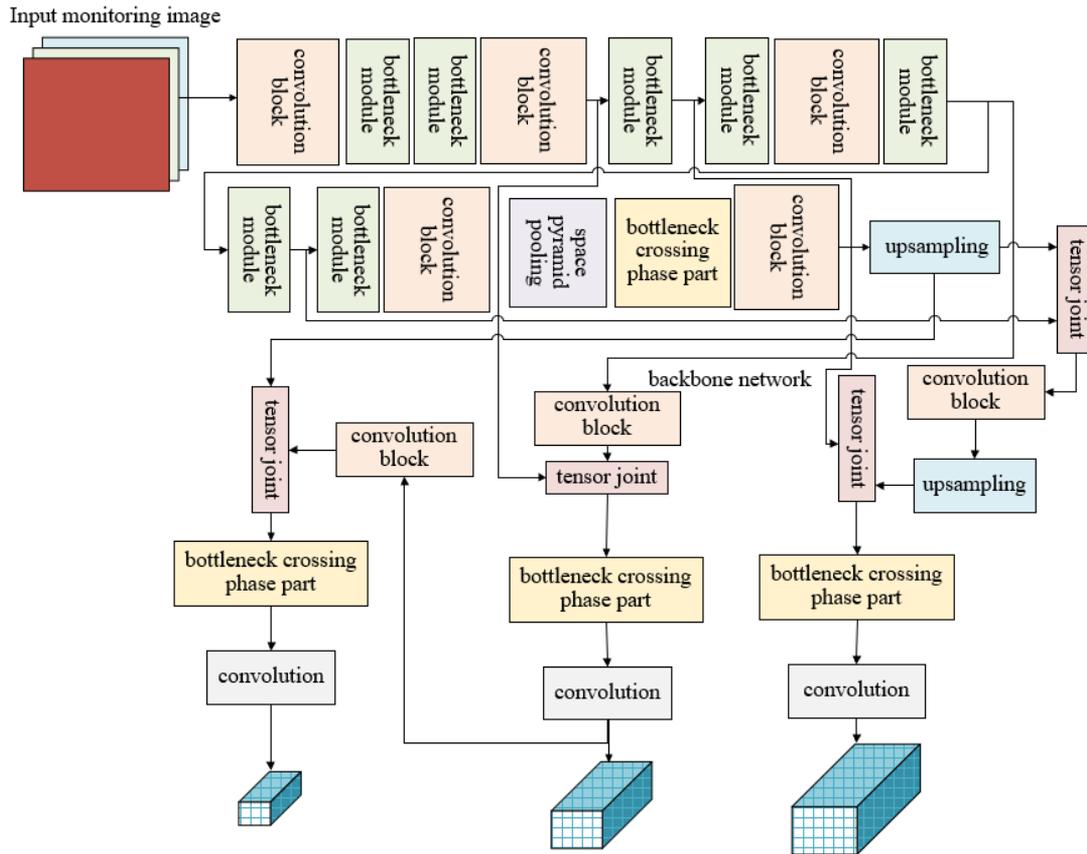


Figure 5. Optimized network structure

Assuming that the minimum closure area of the predicted box and the real box is represented by G_b , the area in G_b that does not belong to two boxes is represented by Y , and the ratio of the intersection of the two boxes to the union of the two boxes is represented by IoU . The following formula gives the calculation formula of *GIoU* loss function commonly used in *YOLOv5* output layer:

$$GIoU = IoU - \frac{|G_b - Y|}{|G_b|} \quad (6)$$

The operation of decomposing standard convolution into deep convolution and floating-point convolution can be considered as a deep separable convolution operation. Assuming that the depth of the input is represented by N , the depth of the output is represented by M , the convolution kernel is represented by S , the depth of the convolution kernel is represented by W_s , the input feature map is represented by R , and the size is (W_R, W_R, N) . The standard convolution S used is (W_s, W_s, N, M) , the output feature map is represented by Z , and the size is represented by (W_Z, W_Z, M) . The following formula provides the calculation formula for the standard convolution:

$$Z_{s,j,m} = \sum_{j,i,n} S_{j,i,n,m} \cdot R_{s+j-1,j+i-1,n} \quad (7)$$

Assuming that the number of input channels is represented by H and the number of output channels is represented by K , the following equation provides the corresponding calculation formula:

$$\psi = W_s \cdot W_s \cdot H \cdot K \cdot W_R \cdot W_R \quad (8)$$

It is assumed that deep convolution is represented by \hat{S} , and the convolution kernel is represented by $(W_s, W_s, 1, H)$. The calculation formula for deep convolution is given by the following equation:

$$\hat{Z}_{s,j,m} = \sum_{j,i,n} \hat{S}_{j,i,n,m} \cdot \hat{R}_{s+j-1,j+i-1,n} \quad (9)$$

The calculation amount of deep convolution and point by point convolution can be calculated using the following equation:

$$\hat{\psi} = W_s \cdot W_s \cdot H \cdot K \cdot W_R \cdot W_R + H \cdot K \cdot W_R \cdot W_R \quad (10)$$

By comparing the calculation amount of the two, the following can be got:

$$\frac{W_S \cdot W_S \cdot H \cdot W_R \cdot W_R + H \cdot K \cdot W_R \cdot W_R}{W_S \cdot W_S \cdot H \cdot K \cdot W_R \cdot W_R} = \frac{W_S^2 + K}{KW_S^2} \quad (11)$$

As can be seen from the above equation, the deep separable convolution operations can greatly reduce computational complexity.

Intelligent driving real-time monitoring image processing requires high instantaneity. Simplifying network structure and obtaining smaller network models can help accelerate processing speed and meet real-time performance requirements. Therefore, this paper removes the *Focus* layer of the backbone network and replaces the *CSP* module with the bottleneck module and replaces the feature fusion part with the bottleneck cross phase part. By optimizing the model structure and changing the convolutional hierarchy, the network can better adapt to the needs of different scenarios, and improve the accuracy of saliency target recognition. By performing deep separable convolution operations, smaller network models can be obtained, changing the convolutional hierarchy of the original network and rebuilding the model structure, thereby reducing computational resource consumption and improving processing speed, which helps to improve network performance and adapt to different scenarios.

It is assumed that the current model parameter is demonstrated by μ_g , the model parameter for the next stage is demonstrated by μ_{g+1} , the learning rate is represented by λ_g , the gradient is represented by $\nabla_{\mu_g} D(\mu_g)$, the total number of batch training is represented by m , and the samples for each input and output are represented by $p^{(j)}$ and $q^{(j)}$, respectively. A randomly selected gradient direction is represented by $(j: j+m)$. The following equation provides the calculation formula for the *SGD* optimizer used in *YOLOv5s*:

$$\mu_{g+1} = \mu_g - \lambda_g \cdot \nabla_{\mu_g} D(\mu_g; p^{j:j+m}; q^{j:j+m}) \quad (12)$$

It is supposed that the accelerated speed accumulated at the moment of d is represented by f_d and the magnitude of the momentum is represented by ε . The calculation formula of *SGDM* optimizer after the introduction of momentum is given as follows:

$$\begin{aligned} f_d &= \varepsilon f_{d-1} + \gamma \nabla_{\mu_{d-1}} D(\mu_{d-1}; p^{(j:j+m)}; q^{(j:j+m)}) \\ \mu_{d+1} &= \mu_d - f_d \end{aligned} \quad (13)$$

Due to the high instantaneity required for the task of recognizing saliency target in the real-time monitoring image from intelligent driving, the saliency targets may feature diversity and complexity. Compared to the *SGD* optimizer used in *YOLOv5s*, the *Adam* algorithm has a faster convergence speed, and the *Adam* algorithm automatically adjusts the learning rate of each parameter. Combining the information of first-order momentum and second-order momentum, different parameters have different learning rates during the training process, thereby improving the efficiency of optimization. Moreover, using *Adam* algorithm with self-adaptive learning rate adjustment ability can help better adapt to the needs of different scenarios. Therefore, choosing *Adam* algorithm instead of *SGD* optimizer for recognizing saliency target in the real-time monitoring image from intelligent

driving has obvious advantages and necessity. These advantages help to improve training efficiency, improve model performance, and better adapt to the needs of different scenarios, thus meeting the needs of processing real-time monitoring image from intelligent driving.

It is assumed that the gradient of the current parameter is represented by v_d , the gradient of the first order momentum is represented by n_d , the gradient of the second order momentum is represented by f_d , the mean value of the correction gradient is represented by \hat{n}_d and \hat{f}_d , and the learning rate is represented by u^* . The two hyperparameters that control the first order momentum and the second order momentum are represented by ω_1 and ω_2 . The smoothing term to avoid the denominator to be 0 is represented by τ . The following provides the calculation formula for the *Adam* algorithm.

$$v_d = \nabla_{\mu_d} D(\mu_{d-1}) \quad (14)$$

$$n_d = \omega_1 n_{d-1} + (1 - \omega_1) v_d \quad (15)$$

$$f_d = \omega_2 f_{d-1} + (1 - \omega_2) v_d^2 \quad (16)$$

$$\hat{n}_d = n_d / (1 - \omega_1^d) \quad (17)$$

$$f_d = f_d / (1 - \omega_2^d) \quad (18)$$

$$\mu_d = \mu_{d-1} - u^* \hat{n}_d / \left(\sqrt{\hat{f}_d} + \tau \right) \quad (19)$$

In practical applications, the real-time surveillance image in intelligent driving may be interfered by various noises and outlier. Since *DBSCAN* algorithm can handle clusters with irregular shapes, it helps to better distinguish different types of significant targets, thus improving the target recognition effect of real-time surveillance image of intelligent driving. The *DBSCAN* algorithm can recognize and process irregularly shaped clusters, which is particularly important in recognition of saliency target in real-time monitoring image of intelligent driving, as the target shapes may have great diversity. Moreover, the *DBSCAN* algorithm does not require a predetermined number of clusters, which makes the algorithm more flexible and able to better cope with clustering needs in different scenarios. Therefore, it is more appropriate to select *DBSCAN* clustering algorithm for saliency target recognition in real-time monitoring image of intelligent driving. Its advantages help to improve recognition accuracy, enhance model robustness, and better distinguish different types of targets, thus meeting the needs of processing the real-time monitoring image from intelligent driving.

Because intelligent driving real-time monitoring image recognition needs to be completed in a short time, the backbone network in this paper is adjusted to *MobileNetV2*, which helps to improve recognition speed and meet instantaneity requirements. Moreover, running the saliency target recognition algorithm for intelligent driving real-time surveillance image on embedded devices or edge computing devices may be limited by hardware resources. *MobileNetV2* has a small model volume and relatively fast reasoning speed. Adjusting the backbone network to *MobileNetV2* has great advantages in resource-constrained environments, such as embedded devices or edge computing devices.

4. EXPERIMENT RESULTS AND ANALYSIS

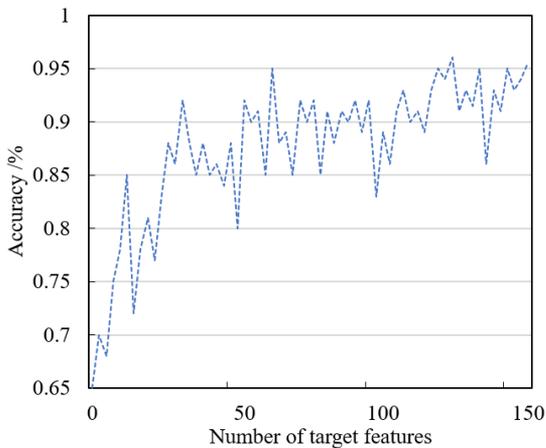
According to the error data of significant target positioning provided in Table 1, it is possible to analyze the significant target positioning error based on real-time monitoring image information of intelligent driving. It is necessary to pay attention to the error between the theoretical coordinates and the calculated coordinates in the table. Based on these error values, the positioning accuracy can be evaluated. From the data in the table, it can be seen that the error value is relatively

small, with an error range between 0.01mm to 0.09mm, which indicates that the constructed model has high accuracy in target localization. The main reason is that optimization strategies such as *MobileNetV2* replacing backbone network, deep separable convolution and *DBSCAN* clustering algorithm are adopted in the process of model construction, which improves the performance of the model when dealing with significant target location tasks. This makes the model highly practical in applications and helps to improve the performance and safety of intelligent driving systems.

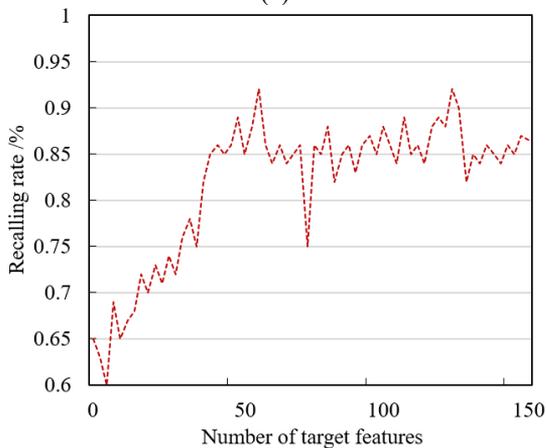
Table 1. Error analysis of salient target location

No. of Feature Point	Theoretical coordinate/mm	Coordinate calculated after reconstruction/mm	Error (Absolute value)/mm
1	(1,1)	(-0.02,0.14)	(0.01,0.06)
2	(12,12)	(15.26,17.48)	(0.05,0.09)
3	(23,23)	(23.69,27.41)	(0.01,0.05)
4	(35,35)	(31.52,39.15)	(0.03,0.07)
5	(41,41)	(42.58,47.39)	(0.05,0.01)
6	(56,56)	(51.28,56.14)	(0.07,0.09)
7	(68,68)	(58.47,59.61)	(0.05,0.04)

Based on the constructed network, the significance targets in the test set are identified and their categories and confidence coefficient are labelled. Figure 6 shows the iterative curve during network training. From the given Figure 6 it can be observed that as the number of target features increases, the significant changes in target recognition accuracy, recall rate, and *mAP* value based on real-time monitoring image information of intelligent driving can be observed.



(a)



(b)

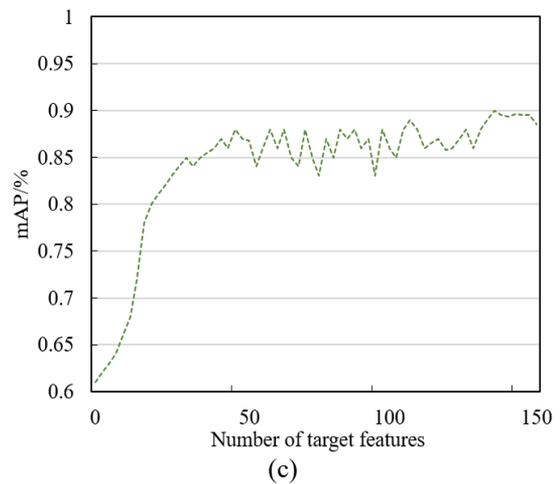


Figure 6. Iteration curve during network training

As shown in the figure, as the number of target features increases, the accuracy, recalling rate, and *mAP* values of the model as a whole show an upward trend. This indicates that the model has high adaptability in identifying significant targets, which verifies the effectiveness of the constructed model. The three indicators of the model exhibit significant fluctuations under certain feature quantities, possibly due to uneven samples and varying training accuracy. These factors may affect the stability and generalization ability of the model. The performance of the constructed model in the task of recognizing significant target in real-time surveillance image from intelligent driving is generally good, which mainly depends on the replacement of the backbone network with *MobileNetV2*, deep separable convolution and other technologies, and the selection of more appropriate optimizers (such as *Adam*) and clustering algorithms (such as *DBSCAN*). The performance of the model in processing significant target recognition task is improved.

This article will also compare the experimental results using the *Faster-RCNN* method and federated learning method for salient target recognition based on real-time monitoring image information of intelligent driving with the results using this method. Figure 7 shows a comparative analysis curve of the average candidate salient target extraction methods.

According to the recall data provided in Figure 7, the performance of different algorithms in recognizing the salient target in the real-time monitoring image information of intelligent driving can be compared. It can be seen that the method in this article presents a recall rate between 0.9 and 0.95, showing good overall performance, and indicating that the model has high accuracy in identifying significant targets. The recall rate of the *Faster R-CNN* method ranges from 0.71 to 0.91, and its overall performance is better than that of the federated learning method, but slightly inferior to that of the proposed method. The recall rate of the federated learning method ranges from 0.55 to 0.9, and the overall performance is relatively poor. Based on the above comparative analysis, the performance of our method in significant target recognition tasks is superior to that of *Faster R-CNN* and federated learning methods, which further validated the effectiveness of the constructed model adopting a series of optimization strategies.

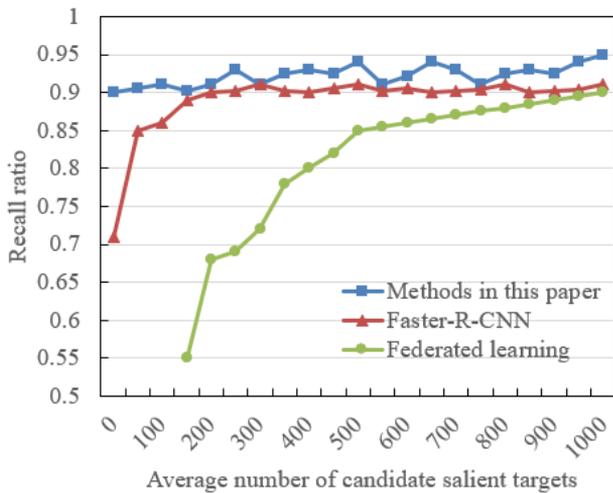


Figure 7. Comparative analysis curve of average candidate salient target extraction methods

Table 2. Performance comparison of different methods

Method	Recall ratio	Precision ratio
HOG+LDA	82.15%	81.27%
Efficient CNN	86.26%	91.67%
Federated learning	91.24%	95.38%
Faster-R CNN	93.47%	94.17%
Detection method in this paper	91.04%	98.62%

Table 2 shows the comparison results of performance of different methods. According to the data given in the table, it can be observed that the detection method in this paper is superior in terms of the precision ratio, reaching 98.62%, which is obviously superior to other algorithms. This means that the constructed model has a high accuracy in identifying significant targets, which can effectively reduce misinformation. Although the recall rate using the detection method proposed in this paper is slightly inferior to that of *Faster-RCNN* and federated learning methods, the gap is not large, and the accuracy rate using the proposed method is significantly higher. This indicates that the constructed model can significantly improve the recognition accuracy while maintaining a high recall rate when detecting targets.

Figure 8 shows an example diagram of recognition results using the proposed method before and after image reconstruction. As can be seen from the figure, the proposed

method shows excellent performance in the task of recognizing significant target in real-time monitoring images of intelligent driving. By comparing the recognition results before and after image reconstruction, it could be clearly found that after image reconstruction, the proposed method can more accurately extract and retain the key features of the target, so as to achieve higher accuracy and recall rate in salient target recognition.

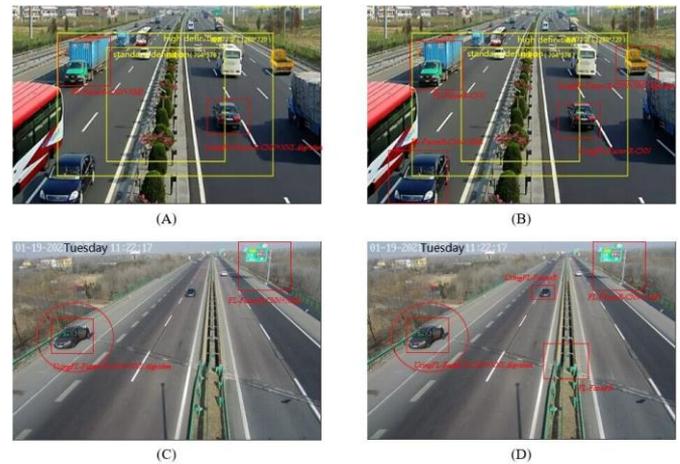


Figure 8. Example of recognition results of the method before and after image reconstruction

As can be seen from the figure, prior to image reconstruction, existing methods may be affected by image noise and background interference, resulting in inaccurate and unstable target recognition. After image reconstruction processing, these problems have been effectively solved. The reconstructed image can not only retain the original target information, but also reduce the noise and background interference, making the salient target more prominent and clearer in the real-time monitoring image. This provides better input for subsequent target recognition tasks and is conducive to improving the performance and safety of intelligent driving system in practical applications.

5. CONCLUSIONS

This article conducts research on saliency target detection methods based on real-time monitoring image information from intelligent driving. The design of *VGG* network discriminator in *ESRGAN* is improved, and spectral normalization (*SN*) and other techniques are applied to enhance the dynamic stability of training, and pixel-level image size amplification and feature enhancement are carried out on the saliency object in the dataset, which provides a richer data basis for the subsequent real-time monitoring of saliency target detection and defect classification. The *YOLOv5s* algorithm is used as the identification network, and the original *YOLOv5s* backbone network is replaced with the *MobileNetV2* network, which greatly reduces the network complexity and improves the identification efficiency of the network. In order to improve the recognition performance of multi-class salient objects, network optimizer and clustering algorithm are used to further improve the performance of recognition of salient objects in real-time monitoring images of intelligent driving. According to the error data of salient target location provided by experiments, the salient target

location error based on real-time monitoring image information of intelligent driving is analyzed. The results show that the constructed model has high accuracy in target location. The iterative curve of network training is given. The results show that the model has high adaptability in identifying significant targets, which verifies the effectiveness of the proposed model. The experimental results of *Faster-R CNN* method and federated learning method, which are also used to carry out salient target recognition based on real-time monitoring image information of intelligent driving, are compared with the results of the method in this paper. The results show that the model can significantly improve the recognition accuracy while maintaining a high recall rate. Examples of recognition results of the proposed method before and after image reconstruction are given, and it is proved that the influence of image noise and background interference are effectively solved after image reconstruction with this method.

FUNDINGS

This work was supported by the project of Jilin Provincial Science and Technology Department (Grant No.: 20210402081GH), the project of Jilin Provincial Development and Reform Commission (Grant No.: 2023C042-4), the Innovation and Entrepreneurship Talent Funding Project of Jilin Province (Grant No.: 2023RY17), the Changchun Science and Technology Bureau (Grant No.: 23ZGM31), the project of Jilin Provincial Science and Technology Bureau (Grant No.: 212558JC010686261), and the project of Changchun Science and Technology Bureau (Grant No.: 21ZGM31).

REFERENCES

[1] Gupta, P., Coleman, D., Siegel, J.E. (2022). Towards physically adversarial intelligent networks (PAINs) for safer self-driving. *IEEE Control Systems Letters*, 7: 1063-1068. <https://doi.org/10.1109/LCSYS.2022.3230085>

[2] Yang, N., Yuan, R., Li, W., Tan, X., Liu, Z., Zhang, Q., Li, C., Ge, L. (2023). Magnetic-driving giant multilayer polyelectrolyte microcapsules for intelligent enhanced oil recovery. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 664: 131107. <https://doi.org/10.1016/j.colsurfa.2023.131107>

[3] Zhou, J., Zheng, H., Wang, J., Wang, Y., Zhang, B., Shao, Q. (2019). Multiobjective optimization of lane-changing strategy for intelligent vehicles in complex driving environments. *IEEE transactions on vehicular technology*, 69(2): 1291-1308. <https://doi.org/10.1109/TVT.2019.2956504>

[4] Gao, F., Duan, J., Han, Z., He, Y. (2020). Automatic virtual test technology for intelligent driving systems considering both coverage and efficiency. *IEEE Transactions on Vehicular Technology*, 69(12): 14365-14376. <https://doi.org/10.1109/TVT.2020.3033565>

[5] Xiong, Y., Gao, H., Li, J. (2020). Intelligent assisted driving system based on Multi-MCU. In *Journal of Physics: Conference Series*, 1676(1): 012186. <https://doi.org/10.1088/1742-6596/1676/1/012186>

[6] Cheng, S., Li, L., Chen, X., Fang, S.N., Wang, X.Y., Wu, X.H., Li, W.B. (2020). Longitudinal autonomous driving

based on game theory for intelligent hybrid electric vehicles with connectivity. *Applied energy*, 268: 115030. <https://doi.org/10.1016/j.apenergy.2020.115030>

[7] Huang, H., Zhao, L., Huang, H., Guo, S. (2020). Machine fault detection for intelligent self-driving networks. *IEEE Communications Magazine*, 58(1): 40-46. <https://doi.org/10.1109/MCOM.001.1900283>

[8] Shao, C., Cheng, F., Xiao, J., Zhang, K. (2023). Vehicular intelligent collaborative intersection driving decision algorithm in Internet of Vehicles. *Future Generation Computer Systems*, 145: 384-395. <https://doi.org/10.1016/j.future.2023.03.038>

[9] Wu, W., Fu, F. (2023). Federated learning-based driving strategies optimization for intelligent connected vehicles. In *Green, Pervasive, and Cloud Computing: 17th International Conference, GPC 2022, Chengdu, China*, pp. 67-80. https://doi.org/10.1007/978-3-031-26118-3_5

[10] Palazzo, M., Vollero, A., Siano, A. (2023). Intelligent packaging in the transition from linear to circular economy: Driving research in practice. *Journal of Cleaner Production*, 388: 135984. <https://doi.org/10.1016/j.jclepro.2023.135984>

[11] Sun, Y., Yang, X., Xiao, H., Feng, H. (2020). An intelligent driving simulation platform: architecture, implementation and application. In *2020 International Conference on Connected and Autonomous Driving (MetroCAD), Detroit, MI, USA*, pp. 71-75. <https://doi.org/10.1109/MetroCAD48866.2020.00019>

[12] Zhang, Y., Li, K., Li, W. (2020). Design and realization of unmanned multi - mode collaborative intelligent driving system. *SID Symposium Digest of Technical Papers*, 51: 201-204. <https://doi.org/10.1002/sdtp.13789>

[13] Park, H., Piamrat, K., Singh, K., Chen, H.C. (2020). Data analysis for self-driving vehicles in intelligent transportation systems. *Journal of Advanced Transportation*, 2020: 9386148. <https://doi.org/10.1155/2020/9386148>

[14] Hu, J., Liu, D., Du, C., Yan, F., Lv, C. (2020). Intelligent energy management strategy of hybrid energy storage system for electric vehicle based on driving pattern recognition. *Energy*, 198: 117298. <https://doi.org/10.1016/j.energy.2020.117298>

[15] Gao, F., Duan, J., He, Y., Wang, Z. (2019). A test scenario automatic generation strategy for intelligent driving systems. *Mathematical Problems in Engineering*, 2019: 3737486. <https://doi.org/10.1155/2019/3737486>

[16] Zhang, T., Zhang, S., Chen, Y., Xia, C., Chen, S., Zheng, N. (2019). Mixture modules based intelligent control system for autonomous driving. In *Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece*, pp. 92-104. https://doi.org/10.1007/978-3-030-19823-7_7

[17] Xiao, K., Shi, W., Gao, Z., Yao, C., Qiu, X. (2020). DAER: A resource preallocation algorithm of edge computing server by using blockchain in intelligent driving. *IEEE Internet of Things Journal*, 7(10): 9291-9302. <https://doi.org/10.1109/JIOT.2020.2984553>

[18] Zhang, Z. (2020). Driving, Monitoring and Protection Technology for SiC Devices Using Intelligent Gate Drive: An Overview. In *2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia), Nanjing, China*, pp. 735-740. [1032](https://doi.org/10.1109/IPEMC-</p>
</div>
<div data-bbox=)

- ECCEAsia48364.2020.9367887
- [19] Wang, Y., Zeng, X., Song, D. (2020). Hierarchical optimal intelligent energy management strategy for a power-split hybrid electric bus based on driving information. *Energy*, 199: 117499. <https://doi.org/10.1016/j.energy.2020.117499>
- [20] Sun, Q., Wang, C., Guo, Y., Yuan, W., Fu, R. (2020). Research on a cognitive distraction recognition model for intelligent driving systems based on real vehicle experiments. *Sensors*, 20(16): 4426. <https://doi.org/10.3390/s20164426>
- [21] Lai, Y., Wang, Q., Chen, R. (2020). Improved single image haze removal for intelligent driving. *Pattern Recognition and Image Analysis*, 30: 523-529. <https://doi.org/10.1134/S1054661820030177>
- [22] Chang, B.J., Chiou, J.M. (2019). Cloud computing-based analyses to predict vehicle driving shockwave for active safe driving in intelligent transportation system. *IEEE transactions on intelligent transportation systems*, 21(2): 852-866. <https://doi.org/10.1109/TITS.2019.2902529>
- [23] Gu, L., Zhang, S., Yang, Y. (2019). Research on machine learning method for intelligent driving. *IOP Conference Series: Materials Science and Engineering*, 631(5): 052017. <https://doi.org/10.1088/1757-899X/631/5/052017>
- [24] Li, J., Han, L., Dong, Z., Li, Y., Lang, P., Shang, T. (2017). A target recognition algorithm applied to the unmanned ground combat platform in curve driving. In 2017 IEEE International Conference on Unmanned Systems (ICUS), Beijing, China, pp. 192-196. <https://doi.org/10.1109/ICUS.2017.8278339>
- [25] Wang, L.P., Sun, D.C. (2014). Research on SVM-based intelligent traffic target recognition algorithm. *Applied Mechanics and Materials*, 556: 2981-2985. <https://doi.org/10.4028/www.scientific.net/AMM.556-562.2981>
- [26] Liu, C., Ren, Y., Hong, L. (2019). Dynamic vehicle target detection and traffic statistics algorithm research. In Eleventh International Conference on Digital Image Processing (ICDIP 2019), Guangzhou, China, pp. 595-602. <https://doi.org/10.1117/12.2540070>
- [27] Tian, S., Yu, H., Yang, Z., Jing, X., Zhang, Z., Shi, M., Wang, Y. (2019). An improved target detection and traffic parameter calculation method based on YOLO with a monocular camera. In 19th COTA International Conference of Transportation Professionals, Nanjing, China, pp. 5696-5708. <https://doi.org/10.1061/9780784482292.490>