Wuhan University
Journal of Natural Sciences

2023, Vol.28 No.2, 150-162

Article ID 1007-1202(2023)02-0150-13
DOI https://doi.org/10.1051/wujns/2023282150

Complex Traffic Scene Image Classification Based on Sparse Optimization Boundary Semantics Deep Learning

ZHOU Xiwei, WANG Huifeng†, LI Saisai, PENG Haonan, WU Jianfeng
School of Electronic & Control Engineering, Chang'an University, Xi'an 710064, Shaanxi, China

Abstract: With the rapid development of intelligent traffic information monitoring technology, accurate identification of vehicles, pedestrians and other objects on the road has become particularly important. Therefore, in order to improve the recognition and classification accuracy of image objects in complex traffic scenes, this paper proposes a segmentation method of semantic redefine segmentation using image boundary region. First, we use the SegNet semantic segmentation model to obtain the rough classification features of the vehicle road object, then use the simple linear iterative clustering (SLIC) algorithm to obtain the over segmented area of the image, which can determine the classification of each pixel in each super pixel area, and then optimize the target segmentation of the boundary and small areas in the vehicle road image. Finally, the edge recovery ability of condition random field (CRF) is used to refine the image boundary. The experimental results show that compared with FCN-8s and SegNet, the pixel accuracy of the proposed algorithm in this paper improves by 2.33% and 0.57%, respectively. And compared with Unet, the algorithm in this paper performs better when dealing with multi-target segmentation.
Key words: traffic scene; SegNet; image classification; simple linear iterative clustering(SLIC); conditional random field; boundary number
CLC number: U 495; TP 391.4

0 Introduction

With the continuous growth of the number of private cars, the accidents of malignant traffic occur frequently on expressways, which causes certain losses to the economy. In order to solve the traffic safety problems and ensure the safety of passengers' lives and property, the concept of Advanced Driver Assistance System (ADAS) has been proposed. ADAS mainly uses a variety of on-board sensors to obtain environmental information inside and outside the vehicle, and through appropriate information processing, analysis, fusion and the supplement of decision-making control, real-time perception of driver's own state, vehicle status and external environment, the ADAS can judge the potential dangers that the vehicles may face. Warn can be issued when it is

Received date: 2022-07-18
Foundation item: Supported in part by the Shaanxi Natural Science Basic Research Program(2022JM-298), the National Natural ScienceFoundation of China (52172324), Shaanxi Provincial Key Research and Development Program(2021SF-483) and the Science and Technology Project of Shaan Provincal Transportation Department(21-202K,20-38T)
Biography: ZHOU Xiwei, male, Ph.D. candidate, research direction: deep learning and embedded system. E-mail: zhouxiwei@chd.edu.cn
† To whom correspondence should be addressed. E-mail: conquest8888@126.com

necessary, or vehicle control systems are directly involved to perform part of the operation and improve vehicle driving safety[1,2].

According to the statistics of relevant literature, more than 90% of environmental information is acquired by visual means[3]. As one of methods of the most effective perception in ADAS system, visual perception can provide the most intuitive, reliable and abundant environmental information for ADAS system. At present, the main research of vision sensing technology of traffic environment is divided into object detection algorithm and image segmentation algorithm. The segmentation of road images is one of the most basic and important research fields, while the algorithm for target segmentation is mainly deep-learning algorithms based on convolutional neural network and the traditional machine learning algorithms.

Traditional image segmentation algorithms include methods based on image threshold[4-6], edge detection[7-10] and region image segmentation[11-14],etc. Ref.[4] proposed an adaptive gray enhancement and linear region threshold segmentation algorithm, which enhances the gray contrast of the target and background, avoids the bad influence of image segmentation method based on single threshold, and improves the accuracy of target recognition and measurement. Ref. [5] proposed a multi-level threshold optimization algorithm incorporating Kapur entropy, and this algorithm uses whale optimization algorithm (WOA) to improve the segmentation accuracy. Ref.[12] proposed an image segmentation method based on sector ring region, by which the object and background in the image can be accurately separated. Ref. [15] divided the image area, and used the improved Hough transform and the tangent relationship of the lane line model to realize the recognition and reconstruction of the lane line. However, due to the complexity of road scenes and the richness of target categories, these traditional image segmentation methods can not accurately distinguish target categories, and the probability of missed detection and false detection is high in practical application. Moreover, these algorithms have long calculation time, low real-time, and large limitations in real scenes. Since 2012, deep-learning algorithms have been rapidly applied to target recognition[16-19], target detection[20-22] and other tasks, and have achieved remarkable results. Deep learning is widely used in image semantic segmentation algorithms and intelligent vehicle assisted driving, providing reliable guidance and decision-making for intelligent vehicle assisted or active driving. Fully Convolutional Networks (FCN) proposed by Long *et al*[23] is used for image semantics segmentation and pixel level classification. When any size image is input and the full connection layer in traditional convolution neural network (CNN) is replaced by convolution layer, arbitrary size images can be classified by the fine-tuning model. However, the disadvantage of FCN is that the results are not precise enough and lack of spatial consistency. SegNet algorithm proposed by Badrinarayanan *et al*[24] is used for semantics image segmentation. Its central idea is based on FCN, which adopts symmetrical encode-decode structure. The encoder part uses the first 13 layers of convolutional network of VGG16. Each encoder layer corresponds to a decoder layer. Finally, the output of the decoder is sent to the soft-max classifier to generate class probability for each pixel independently, which improves the accuracy of image segmentation. Gan *et al*[25] proposed an improved Unet model of self attention mechanism, which had certain improvement in image segmentation of similar regions. Qian *et al*[26] improved the R-CNN network and introduced the characteristic pyramid network into the backbone network, which achieved higher accuracy in traffic sign segmentation. Ref. [27] fused polarization features and intensity images under low visibility conditions, and used depth neural network to segment and identify the vehicle road environment, which ultimately improved the perception effect under adverse weather conditions. Although deep learning has developed rapidly in the field of image segmentation, the difficulties of multi-target recognition and segmentation in complex traffic environments under changing conditions lie in the variety of targets, mutual occlusion between targets, and many similar regions. The existing network needs to have higher accuracy and robustness.

Based on the above ideas, this paper proposes a target boundary optimization algorithm for complex vehicle road images. The LF-SegNet[28] network model is used to classify different target categories in the road images, but the network modeling ability is insufficient, resulting in poor network stability and target classification. The recognition accuracy is not high. Therefore, it is necessary to further optimize the classification to improve the overall performance of image classification. In this paper, two prior constraints are introduced by using the super-pixel feature. The first constraint is based on the correlation between adjacent pixels, which are likely

to belong to the same category. The second constraint is the label map and the original image. The boundary information is basically the same. With these two conditions, the road image boundary of the SegNet network segmentation can be optimized.

1 Overall Design

The proposed image boundary optimization algorithm feeds back the boundary and contour information of the original road image extracted by super-pixels to the convolution neural network classification image, further enhances and improves the preliminary model, and achieves the accurate classification of complex road targets. Firstly, the SegNet algorithm is used to extract the target pixel level features. Then, the simple linear iterative clustering (SLIC)[19-24,29] algorithm is used to extract the features of image super-pixel blocks and edge information. Then, the image boundary optimization is realized by combining the features of pixel level and super-pixel blocks. Finally, the precise edge recovery capability of Condition Random Field (CRF) is used to optimize the segmentation results. This paper optimizes the boundary of road image and the false segmentation of small area targets. The specific scheme framework is shown in Fig.1.

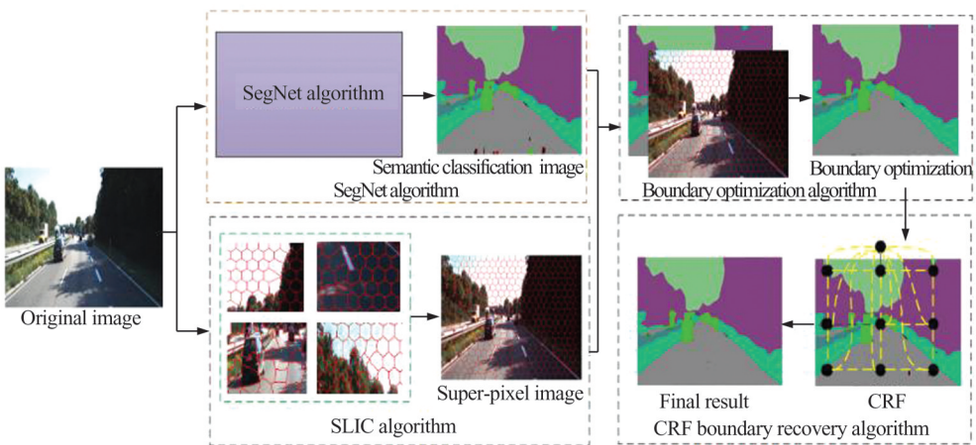

Fig.1 The flow chart of the proposed method

From Fig. 1, we can see that the road image is semantically classified and super-pixel segmented by SegNet algorithm and SLIC algorithm, respectively. The boundary optimization algorithm proposed in this paper is used to optimize the image boundary and to deal with local false segmentation. However, this method is not effective for the object with slender size, so the boundary recovery ability of conditional random field is used to deal with such object edges finally. Information is restored to further improve the effect of target classification.

2 Semantic Segmentation Algorithm Based on Super-Pixel Boundary Optimization

The boundary optimization algorithm based on super-pixel is mainly divided into three parts:

1) Pixel classification: using convolution neural network algorithm to obtain image features, and classify and recognize each pixel.

2) Region segmentation: combining the pixels with similar features in street scene image to form several representative regions and obtain the over- segmented regions of the image, and the classification of pixels in each region determining by combining the neural network.

3) Recognition of pixel categories: according to the pixel classification algorithm in each region, the pixels in the whole image are reclassified. The flow chart of the boundary optimization algorithm is shown in Fig.2.

2.1 Semantic Segmentation Algorithm Based on Boundary Optimization

In this paper, SegNet[28] network is used to classify images at the pixel level. SegNet network is mainly composed of encoder and decoder. The network model is shown in Fig. 3. The coding process is mainly based on the pretraining model VGG-16 network, which retains

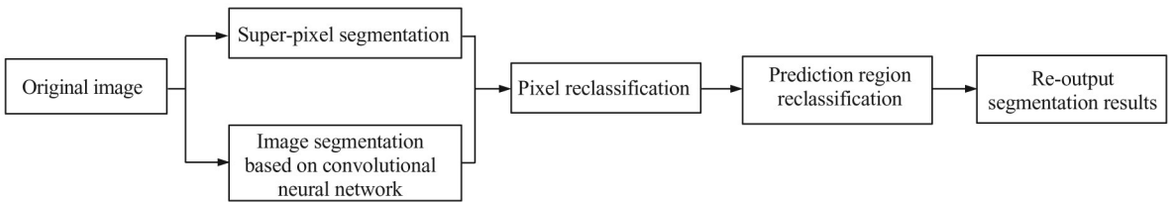

Fig.2 Image semantic segmentation based on super-pixel

13 convolution layers to extract image features. Only the last three full connection layers need to be discarded, which greatly reduces the number of learning parameters. SegNet network has 13 convolution layers, 5 pooling layers, 13 deconvolution layers and 5 upper sampling layers. The pool layer uses 2×2 window and strides 2 step size. Each pool layer is equivalent to a half-resolution reduction of the image. During each max pool, the location of the maximum value in each pooling window in feature maps is recorded. The decoding process mainly uses the largest pooled index recorded to sample the feature maps of the input image after convolution pooling. The up-sampling processes of SegNet and FCN are shown in Fig.4(a) and (b), respectively.

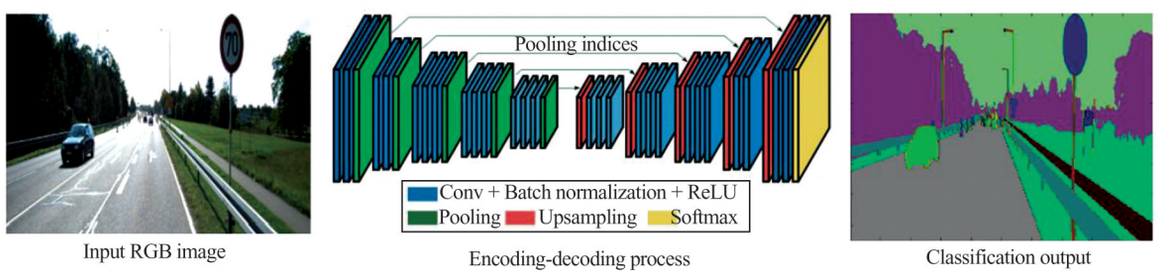

Fig. 3 The model of SegNet network structure

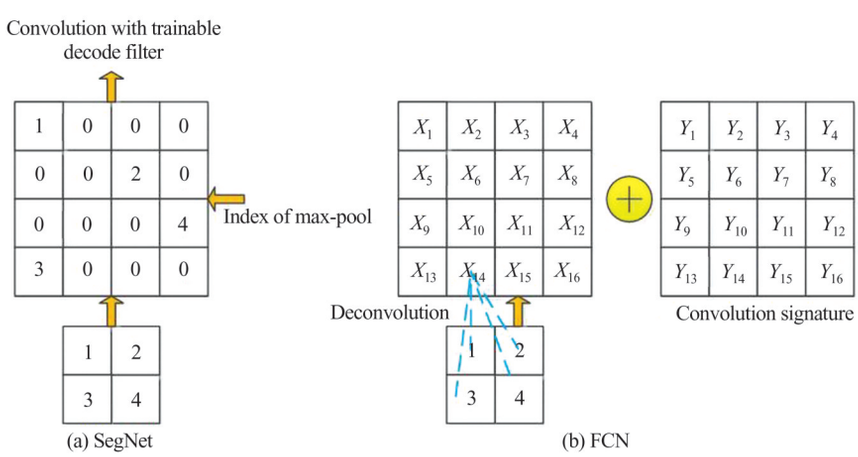

Fig. 4 Up-sampling differences between SegNet and FCN

In Fig. 4(a), the up-sampling process of SegNet is that characteristic graph values 1, 2, 3, 4 are mapped to a new feature graph by the maximum pooled coordinates previously saved; in Fig. 4(b), the up-sampling process of FCN is that by deconvoluting the eigenvalues 1, 2, 3 and 4, the new characteristic chart is added to the corresponding convolutional characteristic chart.

The test accuracy of the trained SegNet model is higher than that of the FCN algorithm. It is also used in the image segmentation of vehicle-road environment. By restoring the original image details hierarchically with multiple decoders, better boundary accuracy can be obtained to a certain extent.

Classification of each target is obtained by convolution layer, pooling layer and deconvolution layer, and each pixel is classified by Soft-max function. The Soft-

max function is

$$S_i^p = \frac{e^{c_i^p}}{\sum_1^N e^{c_i^p}} \tag{1}$$

In image classification, S_i^p denotes the probability that the pixel p belongs to class i, and N represents the total number of categories, and $e^{c_i^p}$ denotes the scoring value of point p belonging to class i in score graph. Then according to the error between the predicted value and the real value, the cross-entropy loss function is constructed:

$$l_p = -\sum_i y_i^p \ln S_i^p \tag{2}$$

where y_i^p denotes the true probability that the pixel p belongs to class i, and y_i^p is defined as follows:

$$y(1)=\begin{bmatrix}1\\0\\0\\\vdots\\0\end{bmatrix}, y(2)=\begin{bmatrix}0\\1\\0\\\vdots\\0\end{bmatrix}, y(3)=\begin{bmatrix}0\\0\\1\\\vdots\\0\end{bmatrix}, \cdots, y(k-1)=\begin{bmatrix}0\\0\\0\\\vdots\\1\end{bmatrix}, y(k)=\begin{bmatrix}0\\0\\0\\\vdots\\0\end{bmatrix} \tag{3}$$

If the pixel belongs to the third class, then the cross-entropy loss function represents the error between the predicted value and the real value, and the smaller l_p is, the higher the accuracy of the prediction is. By restoring the original image details hierarchically with multiple decoders in SegNet network, better boundary accuracy can be obtained to a certain extent. However, due to the complexity of road images and the ambiguity of some target boundaries, it is impossible to classify all targets directly by using SegNet network only.

2.2 SLIC Algorithm

In this paper, SLIC algorithm is used for image region segmentation. SLIC algorithm converts image from RGB color space to CIE-LAB color space. The color values of (l, a, b) and (x, y) coordinates of each pixel constitute a five dimensional vector $\boldsymbol{V}[l, a, b, x, y]$. The similarity of two pixels can be measured by their vector distance. The larger the distance, the smaller the similarity. The detailed flow of SLIC algorithm is as follows:

Step 1: Setting the number of super-pixels is K for the color image in CIE-LAB color space. Initial clustering centers are initialized by $\boldsymbol{C}_i=(l_i, a_i, b_i, x_i, b_i)^{\mathrm{T}}$ and the step length of clustering center of super-pixels is

$$S = \sqrt{\frac{N}{K}} \tag{4}$$

Step 2: Clustering. Class labels are assigned to each pixel in the neighborhood of each seed point. By calculating the distance between the seed point and each pixel, the lab color space distance between the seed point and the seed point is initialized and infinite, which represents the space coordinate distance and the comprehensive distance between the seed point and the seed point.

$$d_c = \sqrt{\left(l_j - l_i\right)^2 + \left(a_j - a_i\right)^2 + \left(b_j - b_i\right)^2} \tag{5}$$

$$D = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2} \tag{6}$$

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} = \sqrt{{d_c}^2 + \left(\frac{d_s}{S}\right)^2 m^2} \tag{7}$$

N_s is the maximum spatial distance within a class, which is defined as $N_s=S$ and it is suitable for each cluster. The maximum color distance is N_c. The final distance metric D' is as follows:

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} = \sqrt{{d_c}^2 + \left(\frac{d_s}{S}\right)^2 m^2} \tag{8}$$

Step 3: Iterative optimization. Repeat Step 2 until all clustering centers do not change.

Step 4: Remove outliers and enhance connectivity. When the number of input super-pixels K and the maximum color distance m are different, different segmentation effects will be produced. The effect is shown in Fig.5. It can be seen from the graph that using hundreds or thousands of super-pixels to optimize massive image data can obtain accurate image boundary, realize efficient image processing, understanding and expression, and serve for efficient and flexible perception of road environment information. Figure 5 shows that we can not only get clear edge areas of the image, but also greatly reduce the computational complexity of the pixel samples and improve the computational efficiency by using hundreds or thousands of super-pixels instead of massive image data .

2.3 Reclassification

For the reclassification of boundary and mis-segmented pixels in each region, the specific steps are as follows: All the pixels in each super-pixel S_p are $C=\{C_1, C_2, \cdots, C_S\}$, the number of pixel labels n^k for each class in this super-pixel is counted:

$$n_p = \{n_1, n_2, n_3, \cdots, n_K\} \tag{9}$$

n_i indicates the number of labels of the pixels owned by class i in region P, then the maximum value n_j in n_p is found, and the pixels in region P are classified as class j. But there is also a case that when the number of maximum number tags n_i is close to the number of sub- maxi-

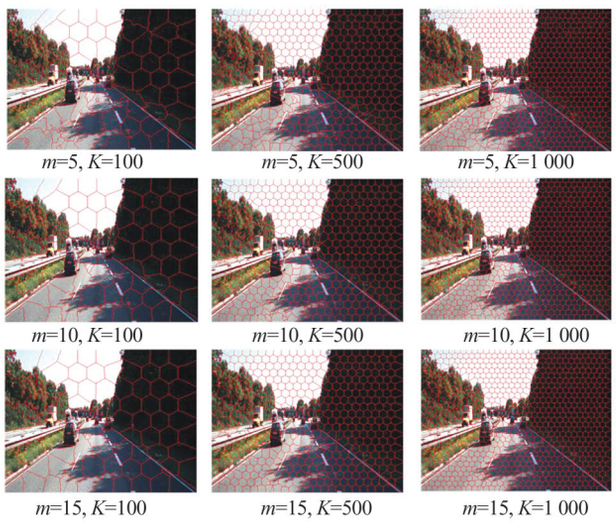

Fig.5 Results of super-pixel segmentation with different parameters

mum number tags n_j in n_p, it is impossible to determine whether the region P belongs to class i or j, but in the neural network, it is defined as the highest probability of classification, so there will be a phenomenon of false segmentation. Thus we need to define a threshold T.

$$T = \frac{n_i - n_j}{S} \tag{10}$$

In this paper, T is 0.2. If T is larger than 0.2, the pixels in region P are classified as class i. Otherwise, they are still classified according to the result of semantic segmentation of convolutional neural network.

2.4 Specific Flow of Boundary Optimization Algorithms

There are many ways to realize the optimization of the image boundary, and the method used in this paper is to combine super-pixels with convolutional neural network to achieve the optimization of the image boundary. Firstly, SegNet image semantics segmentation algorithm based on VGG-16 network model is used to realize the rough segmentation of images and extract rough features. SLIC algorithm is used to process the image of image segmentation generated super-pixel. Then rough features are optimized by the boundaries of these super-pixel objects. This method can improve the accuracy of object boundary segmentation in a way. The key algorithm of boundary optimization has been fully proved in Algorithm 1.

The flow chart of the Algorithm 1 is as follows:

Step 1: Input original image I and rough feature graph L extracted by SegNet algorithm.

Step 2: After segmenting the original image with SLIC algorithm, K super-pixels $S_p = \{S_1, S_2, \cdots, S_K\}$ are obtained, and the area of each super-pixels is marked with label i.

Step 3: for $i = 1{:}K$

1) All the pixels in the super-pixel S_i are $S_i = \{C_1, C_2, \cdots, C_N\}$, where C_j corresponds to one of the pixels in class j in the feature map;

2) Initialize the number of occurrences of the same label to $n = 0$;

3) for $j = 1{:}K$

The feature label of C_j is saved as L_{C_j}, and the number of pixels with the same label is calculated, and the whole super-pixel is traversed.

$$n_j = \sum C_j \tag{11}$$

W_{C_j} is the proportion of pixels in the same label L_{C_j}.

$$W_{C_j} = \frac{n_j}{N} \tag{12}$$

4) Redistribution of labels.

If $W_{C_j} > 0.8$, mark the super-pixel with the corresponding label $L_{C_{\max}}$ of W_{C_j} and jump to Step 4.

Else search for maximum $W_{\max}$ and sub-maximum W_{sub}; If $W_{\max} - W_{\text{sub}} > 0.2$, mark the super pixel with $L_{C_{\max}}$ corresponding to $W_{\max}$, and jump to Step 4.

Else use the class of the segmentation result of DeepLabV2 to mark the super pixel.

Step 4: Use $L_{C_{\max}}$ to reassign the current super-pixel classification and output the image I'.

Through this algorithm, we can optimize the boundary of the road image and the scene of small area of mis-segmentation in road image, which shows that there are many small "spots" in the image. Through SLIC algorithm processing, these "spots" can be eliminated by its similarity with the block labels of surrounding super-pixels to a certain extent. The specific effect is shown in Fig.6.

The two images on the left side of Fig.6 are superimposed by the vehicle road image of SegNet segmentation and super-pixel segmentation, but its boundary is not optimized. As shown in the enlarged part 1,2,3,4 of the image, we can see that the boundary of the image and some parts of the local area are not classified. Therefore, the optimization of the algorithm in this chapter can show that the boundary of the image is optimized and the error of local small. Figure 6 is the effect diagram of local optimization. Thus, area classification can be eliminated. For example, 1 and 2 of the local graph show that the boundary is optimized, and 3 and 4 show

that some small "spots" are correctly classified. The specific optimization diagram is shown in Fig.7.

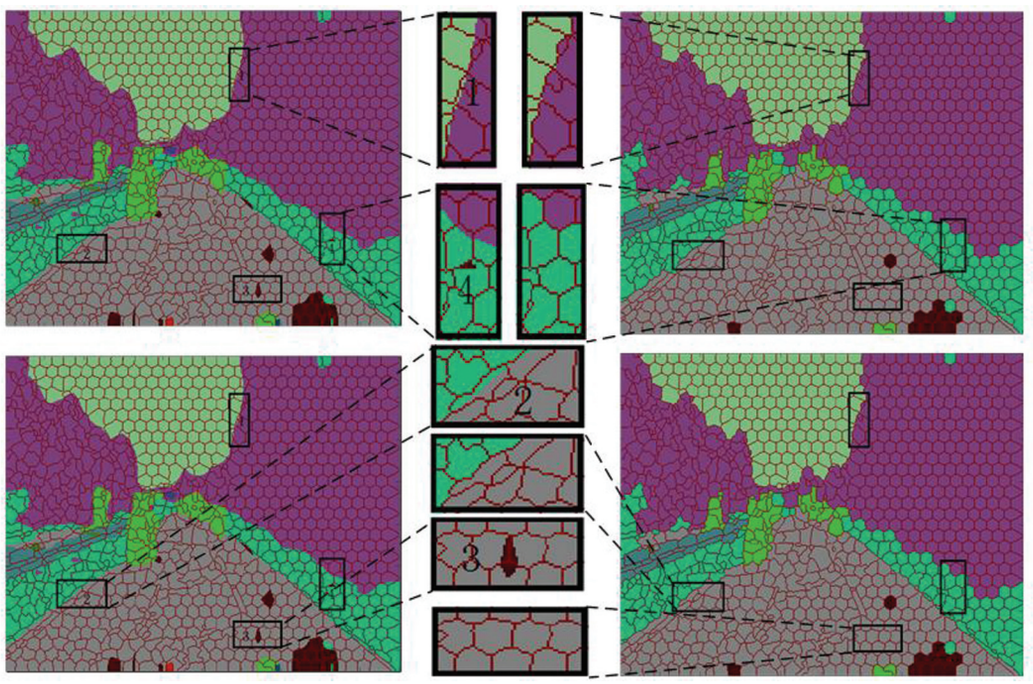

Fig. 6 The effect diagram of local optimization

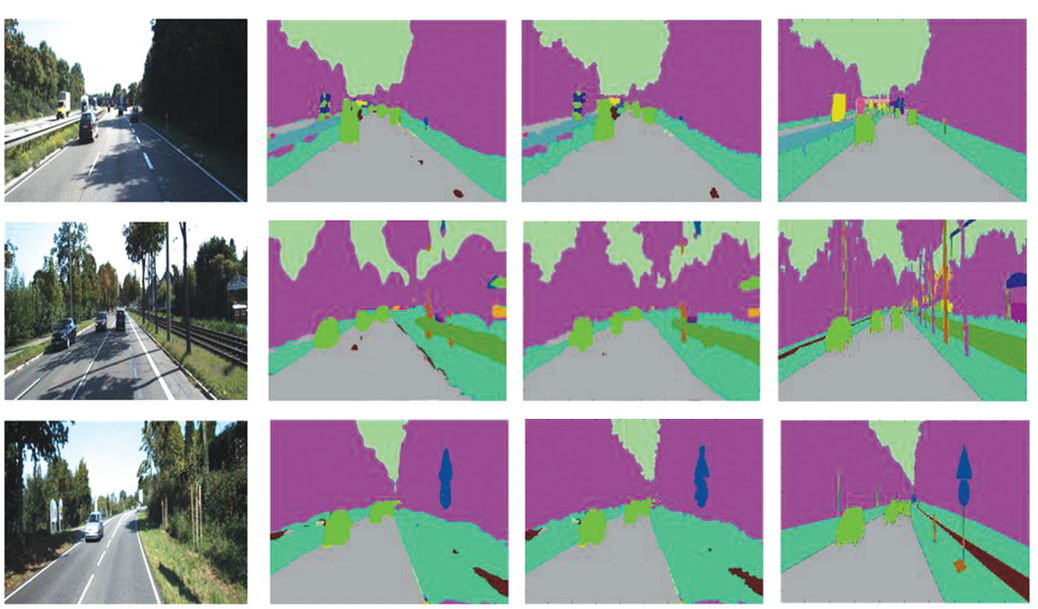

Fig. 7 Specific optimization diagram

2.5 Boundary Restoration

Although the super-pixel boundary optimization algorithm can optimize the road image boundary and the wrong segmentation area, the optimized image boundary is no longer smooth, and it is difficult to segment into several subblocks for such objects as poles, fences and so on. Therefore, CRF model is used to refine the image boundary in the subsequent optimization in this section to restore the boundary more accurately.

CRF[30] is a typical discriminant model, which is similar to the probabilistic indirected graph model. If the pixel label in the image is regarded as a node, the weight

relationship between adjacent two pixels is regarded as an edge. The graph $G=(V,E)$ is given by the method of graph theory, where $V=\{V_1, V_2, \cdots, V_N\}$ is the vertex set and $E=\{E_1, E_2, \cdots, E_N\}$ is called the edge set. Each edge E in E_i connects two vertices in V. Let's assume that the class label of the whole picture is $L=\{l_1, l_2, \cdots, l_n\}$, where i denotes a random pixel representing class l_i and an input pixel of image I is N, plus normalization, and the final conditional probability of (I, L) is

$$P(L=l|I)=\frac{1}{Z(I)}\cdot\exp(-E(l|I)) \tag{13}$$

where $E(l)$ is the marked Gibbs energy $l\in L^N$, $Z(I)$ is the partition function and $E(l|I)$ is the energy function. Fully connected CRF model using Energy Function is

$$E(l)=\sum_i \psi_i(l_i)+\sum_{i,j}\psi_{i,j}(l_i,l_j) \tag{14}$$

Among them, $\sum_i \psi_i(l_i)$ is the one-dimensional potential function of the probability of taking label l_i for pixel i, which comes from the optimized output of front-end super-pixels; $\sum_{i,j}(l_i,l_j)$ is the binary potential function of assigning label l_i, l_j to pixel i, j at the same time, and similar pixels are assigned the same label, while the pixels with large differences are assigned different labels. In this section, the one-dimensional potential can be regarded as boundary optimization feature mapping, which can help improve the performance of CRF model. The two-dimensional potential function usually simulates the relationship between adjacent pixels and weights them by color similarity. The expression of the binary potential function is as follows:

$$\psi_{i,j}(l_i,l_j)=\mu(l_i,l_j)[\omega_1\exp(\frac{\left\|\boldsymbol{P}_i-\boldsymbol{P}_j\right\|^2}{2\sigma^2\alpha}-\frac{\left\|\boldsymbol{I}_i-\boldsymbol{I}_j\right\|^2}{2\sigma^2\beta})$$
$$+\omega_2\exp(-\frac{\left\|\boldsymbol{P}_i-\boldsymbol{P}_j\right\|^2}{2\sigma^2\gamma}] \tag{15}$$

Among them, $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ are color vectors, $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ are pixel positions, so the value of binary potential function depends on the pixel position and color information. $\sigma\alpha$, $\sigma\beta$ control the proximity and similarity between two pixels. As shown in Potts model[31], if $l_i=l_j$, then $\mu(l_i,l_j)$ is 1, otherwise 0, indicating that adjacent similar pixels should be assigned different labels. The similar pixels are assigned the same label.

The "distance" refers to the relevant color space distance and the actual distance. Therefore, the accuracy of image boundary optimization can be improved by using CRF algorithm. Figure 8 is an enlarged image of local details optimized by CRF algorithm.

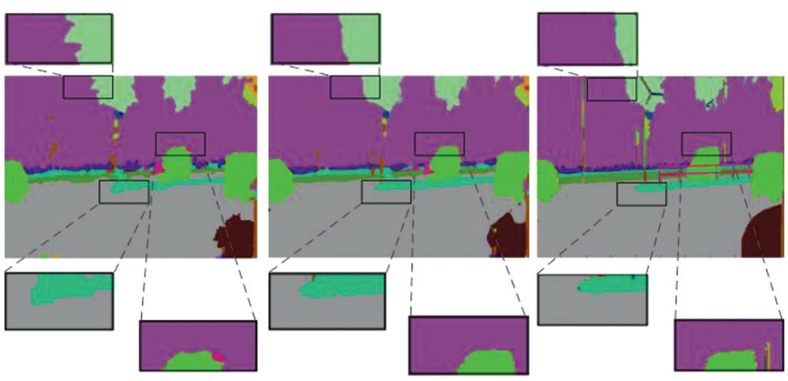

Fig. 8 CRF for edge detail recovery

From Fig. 8, we can see that although the super-pixel boundary optimization algorithm can eliminate the local error segmentation and optimize the image edge, it still needs to strengthen the edge optimization for thin edge and complex overlapping area. After adding CRF algorithm, we can see that the image boundary details can be optimized, and more effective information can be applied.

3 Experimental Results and Analysis

In this section, we evaluate and analyze the experimental results from subjective performance and objective performance, respectively, to verify the effectiveness and performance of our proposed algorithm.

3.1 Evaluation and Analysis of Subjective Performance

In this paper, KITTI data set is used to train and simulate the network. After many times of network iteration training, the network training results are obtained. The classification results are optimized by the fusion of super-pixel and convolution neural network. The segmentation results of road map tested by training model are shown in Fig.9.

From Fig. 9, we can see that the segmentation results of our algorithm and SegNet's algorithm are not very different from each other from a visual point of view because the image boundary pixels are relatively small. However, the phenomenon of misjudgment in small area in the image has been significantly improved. From the right grassland in the first line image, the lane part in the second line image, the traffic signs in the third line image, and the right grassland in the fifth line image, there are misjudged pixels in these areas. By using this algorithm, some areas can be classified correctly.

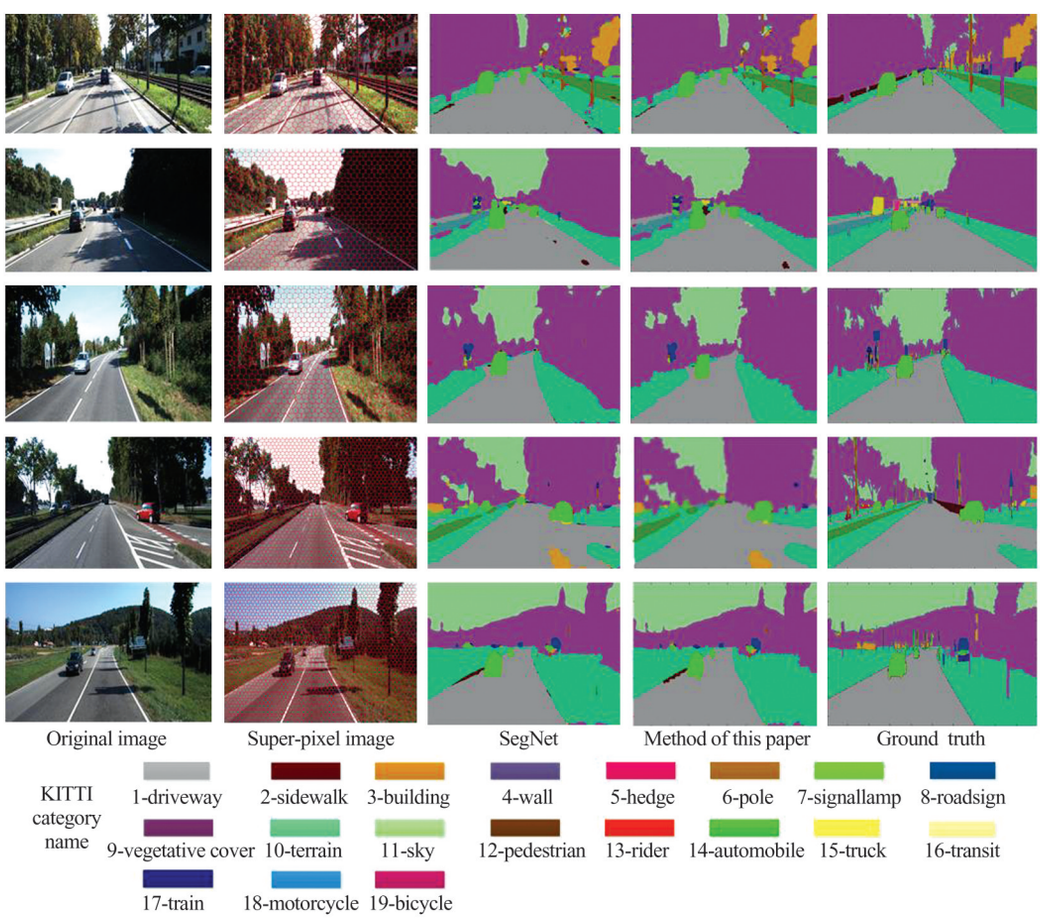

Fig. 9 Contrast chart of subjective evaluation

3.2 Evaluation and Analysis of Objective Performance

In image segmentation, many criteria are usually used to measure the accuracy of the algorithm. These criteria are usually variations of pixel accuracy and intersection-over-union (IoU). In the formula for evaluating indicators, the basic principle is that an image has $k+1$ categories, p_{ij} indicates the number of pixels that should belong to class i but are mis-predicted to class j.

1) Pixel Accuracy (PA): it represents the proportion of correctly marked pixels to total pixels.

$$\mathrm{PA}=\frac{\sum_{i=0}^{k}p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}} \tag{16}$$

2) Mean Pixel Accuracy (MPA): it represents the proportion of the number of pixels correctly classified in each class, then the average of all classes is calculated.

$$\mathrm{MPA}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}} \tag{17}$$

3) Mean Intersection over Union (MIoU): it means calculating the ratio of the intersection and union of two sets. In the problem of semantic segmentation, the two sets are ground truth and predicted segmentation. This ratio can be transformed into the ratio of positive and true numbers to the sum of true, false negative and false positive (union). IoU is calculated on each class and then averaged.

$$\mathrm{MIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}\left(p_{ji}-p_{ii}\right)} \tag{18}$$

The indicators overall performance of the three different models studied in this paper are shown in Table 1 below:

Table 1 Overall performance index for different algorithms

Segmentation algorithm	PA	MPA	MIoU
FCN-8s algorithm	0.906 9	0.740 1	0.610 6
Unet algorithm	0.935 8	0.600 8	0.533 1
SegNet algorithm	0.916 3	0.759 8	0.628 2
Algorithm in this paper(without CRF)	0.916 4	0.762 1	0.629 2
Algorithm in this paper	0.923 4	0.782 3	0.633 9

The model proposed in this paper is mainly composed of three parts. SegNet model is used to roughly segment the network, SLIC algorithm is used to fine segment the segmentation result under super pixel, and finally CRF algorithm is used to restore the image boundary. To prove the effectiveness of the proposed model in this paper, the SegNet model, SegNet+SLIC algorithm (no CRF algorithm added) are compared with it. The experimental results are shown in Table 1. Table 1 shows that before adding CRF to the algorithm in this paper, the pixel accuracy of this algorithm is very similar to that of SegNet model, but after adding CRF algorithm, the pixel accuracy of our algorithm is improved by 0.71% and the average pixel accuracy is increased by 2.25%. The MIoU increased by 0.10% without CRF and increased by 0.57% with CRF compared with SegNet. Compared with Unet, the MPA and MIoU of this algorithm are 18.15% and 10.08% higher, respectively. Experiments show that the proposed algorithm improves the overall performance of road image segmentation, but the segmentation accuracy of slender objects such as poles and fences still needs to be improved.

Using the KITTI data set to test the value of IoU for each class, the contrastive values are shown in Table 2.

From Table 2, compared with FCN-8s and SegNet, it can be seen that IoU of road, sidewalk, building, fence, pole, traffic light, traffic sign, vegetation, terrain, sky, rider, car, truck, bus, motorcycle and bicycle has been improved, and the performance of wall remains unchanged. The performance of person and train has declined compared with FCN-8s, because there are not many human images in the data set, and most of them are concentrated in the distance of the scene. There is also the phenomenon of occlusion, it is easy to classify them into other categories by using the boundary optimization algorithm based on super-pixel. After adding CRF algorithm to the image boundary restoration, the IoU of all target categories has been basically improved, which shows that CRF algorithm has a strong ability to restore image boundary details. In short, the algorithm in this paper is helpful to image boundary optimization.

Comparing the proposed algorithm with the Unet model, it can be found that although the total pixel accuracy of Unet is one percentage point higher than that of the proposed algorithm , the MPA and MIoU of Unet are significantly lower than that of the proposed model . It can be seen from Table 3 that although Unet has strong recognition ability for target categories that account for a large proportion of pixels such as road and terrain, its performance for multi-target segmentation tasks is not as good as that of the algorithm proposed in this paper. The comparison experiment shows that the algorithm in this paper has better performance for multi segmentation tasks.

4 Conclusion

This paper proposes a segmentation method of semantic redefine segmentation using image boundary region. First, the rough features of the target image are extracted using the SegNet model. Then, the SLIC algorithm is used to extract the contour information of the image edge at the super pixel level, and the super pixel feature is applied to the edge information to improve the segmentation accuracy of the target image. In addition, the CRF algorithm is used to restore the image boundary, further refining the segmentation effect. At the same time, the ablation experiment proves that the use of CRF

Table 2 IoU indicators for various targets under different models

%

Target category	FCN-8s	SegNet algorithm	Algorithm in this paper (without CRF)	Algorithm in this paper
road	93.2	**95.1**	**95.0**	**95.2**
sidewalk	80.3	80.2	80.2	**80.4**
building	86.3	87.9	87.5	**88.1**
wall	**56.8**	56.8	56.9	**56.8**
fence	40.1	43.6	43.0	44.3
pole	27.1	28.2	27.9	28.5
traffic light	37.3	37.4	37.7	**37.7**
traffic sign	51.1	52.6	53.0	**53.1**
vegetation	90.5	92.5	92.6	**92.9**
terrain	71.0	76.9	77.0	**77.6**
sky	95.2	96.3	96.7	**97.0**
person	**52.1**	50.7	50.9	51.3
rider	55.4	56.5	56.8	**57.1**
car	87.9	92.0	92.1	**92.3**
truck	45.0	47.3	47.0	48.1
bus	50.5	59.2	59.5	**60.5**
train	43.9	42.0	42.4	**42.9**
motorcycle	41.8	40.2	40.6	41.9
bicycle	54.6	58.3	58.7	59.1

Note: The bold number indicates the most effective indicator among all comparative experiments

Table 3 IoU differences of different models in segmenting key vehicle road environmental objects

%

Target category	FCN-8s	SegNet algorithm	Unet algorithm	Algorithm in this paper (without CRF)	Algorithm in this paper
road	93.2	**95.1**	**95.0**	**95.0**	**95.2**
building	86.3	87.9	48.2	87.5	**88.1**
pole	27.1	28.2	19.5	27.9	28.5
traffic sign	51.1	52.6	12.0	53.0	**53.1**
vegetation	90.5	92.5	92.0	92.6	**92.9**
terrain	71.0	76.9	**82.7**	77.0	**77.6**
sky	95.2	96.3	96.1	96.7	**97.0**
car	87.9	92.0	77.5	92.1	**92.3**
bus	50.5	59.2	22.7	59.5	**60.5**

Note: The bold number indicates the most effective indicator among all comparative experiments

algorithm improves the performance of the algorithm proposed in this paper.

Through the algorithm in this paper, the object segmentation result in the image is more accurate, and the recognition accuracy of each category is higher in the multi object segmentation scene. The final experiment in

Table 1 shows that the MIoU of this algorithm on KITTI dataset can reach 63.39%, 2.33% higher than FCN-8s, 0.57% higher than SegNet. At the same time, compared with Unet, the MPA and MIoU of this algorithm are 18.15% and 10.08% higher, respectively.

References

[1] Lee J, Cho C W, Shin K Y, *et al*. 3D gaze tracking method using Purkinje images on eye optical model and pupil[J]. *Optics and Lasers in Engineering*, 2012, **50**(5): 736-751.

[2] Biondi F, Rossi R, Gastaldi M, *et al*. Beeping ADAS: Reflexive effect on drivers' behavior[J]. *Transportation Research Part F Psychology & Behaviour*, 2014, **25**: 27-33.

[3] D'Amato A, Pianese C, Arsie I, *et al*. Development and onboard testing of an ADAS-based methodology to enhance cruise control features towards CO2 reduction [C]// *IEEE International Conference on Models and Technologies for Intelligent Transportation Systems*. New York: IEEE, 2017: 503-508.

[4] Wang W D, Xin B J, Deng N, *et al*. Single vision based identification of yarn hairiness using adaptive threshold and image enhancement method [J]. *Journal of the International Measurement Confederation*, 2018, **128**: 220-230.

[5] Yan Z, Zhang J, Yang Z, *et al*. Kapur's entropy for underwater multilevel thresholding image segmentation based on whale optimization algorithm[J]. *IEEE Access*, 2021, **9**: 41294-41319.

[6] Kumar V, Lal T, Dhuliya P, *et al*. A study and comparison of different image segmentation algorithms[C]// *Proceedings of the* 2016 2*nd International Conference on Advances in Computing*, *Communication*, *& Automation* (*ICACCA*). New York: IEEE, 2016.

[7] Yan S J, Tang G A, Li F Y, *et al*. An edge detection based method for extraction of loess shoulder-line from grid DEM [J]. *Geomatics and Information Science of Wuhan University*, 2011, **36** (3): 363-367.

[8] Zhou W, Du X, Wang S. Techniques for image segmentation based on edge detection[C]// 2021 *IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology.* New York: IEEE, 2021: 400-403.

[9] Sun Q D, Qiao Y M, Wu H, *et al*. An edge detection method based on adjacent dispersion[J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2016, **30** (10): 943-951.

[10] Guan W L, Wang T T, Qi J Q, *et al*. Edge-aware convolution neural network based salient object detection[J]. *IEEE Signal Processing Letters*, 2019, **26** (1): 114-118.

[11] Yang Y F, Peng H , Jiang Y, *et al*. A region-based image segmentation method under P systems [J]. *Journal of Information and Computational Science*, 2013, **10** (10): 2943-2950.

[12] Xu S Y, Peng C L, Chen K, *et al*. Measurement method of wheat stalks cross section parameters based on sector ring region image segmentation[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, **49**(4): 53-59.

[13] Chen Y L , Ma Y D ,Kim D H, *et al*. Region-based object recognition by color segmentation using a simplified PCNN [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26** (8): 1682-1697.

[14] Su L, Fu X J, Zhang X D, *et al*. Delineation of carpal bones from hand X-ray images through prior model, and integration of region-based and boundary-based segmentations[J]. *IEEE Access*, 2018, **6**: 19993-20008.

[15] Wang H, Wang Y, Zhao X, *et al*. Lane detection of curving road for structural highway with straight-curve model on vision[C]// *IEEE Transactions on Vehicular Technology*, 2019, **68**(6): 5321-5330.

[16] AI-Kofahi Y, Zaltsman A, Graves R, *et al*. A deep learning-based algorithm for 2-D cell segmentation in microscopy images[J]. *BMC Bioinformatics*, 2018, **19** (1):6-8.

[17] Li L H, Qian B, Lian J, *et al*. Traffic scene segmentation based on RGB-D image and deep learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19** (5) : 1664-1669.

[18] Garcia-Garcia A, Orts-Escolano S, Oprea S, *et al*. A survey on deep learning techniques for image and video semantic segmentation[J]. *Applied Soft Computing Journal*, 2018, **70**: 41-65.

[19] Chen L C, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40** (4): 834-848.

[20] Lu R T, Shen T, Yang X G, *et al*. Infrared dim moving target detection algorithm assisted by incremental inertial navigation information in highdynamic air to ground background (Invited) [J]. *Infrared and Laser Engineering*, 2022, **51** (4): 50-60.

[21] Mammeri A, Zuo T, Boukerche A. Extending the detection range of vision-based vehicular instrumentation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2016, **65** (4): 856-873.

[22] Li J N, Liang X D, Shen S M, *et al*. Scale-aware fast R-CNN for pedestrian detection[J]. *IEEE Transactions on Multime-*

dia, 2018, **20** (4): 985-996.

[23] Long Z R, Wei B, Feng P, *et al*. A fully convolutional networks (FCN) based image segmentation algorithm in binocular imaging system[C]// *International Conference on Optical Instruments and Technology*, 2018: 10621.

[24] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(12): 2481-2495.

[25] Gan J H, Zhang R Q. Ultrasound image segmentation algorithm of thyroid nodules based on improved U-Net network [C]// *Proceedings of the* 2022 3*rd International Conference on Control*, *Robotics and Intelligent System* (*CCRIS* ' 22). New York: Association for Computing Machinery, 2022: 61-66.

[26] Qian H, Ma Y, Chen W, *et al*. Traffic signs detection and segmentation based on the improved mask R-CNN[C]//2021 40*th Chinese Control Conference* (*CCC*). New York: Association for Computing Machinery, 2021: 8241-8246.

[27] Wang H F, Shan Y H, Hao T, *et al*. Vehicle-road environment perception under low-visibility condition based on polarization features via deep learning[C]// *IEEE Transactions on Intelligent Transportation Systems*, 2022, **23**(10): 17873-17886.

[28] Mittal A, Hooda R, Sofat S. LF-SegNet: A fully convolutional encoder-decoder network for segmenting lung fields from chest radiographs,wireless personal communications[J] *Wireless Personal Communications*, 2018, **101**(1): 511-529.

[29] Shan J C, Li X Z, Song M, *et al*. Semantic segmentation based on deep convolution neural network[C]// *Journal of Physics*: *Conference Series*, 2018, **1069**(1): 012169.

[30] Gadde R, Jampani V, Kiefel M, *et al*. Super pixel convolutional networks using bilateral inceptions[C]// *Lecture Notes in Computer Science.* Berlin: Springer-Verlag, 2016, **9905**: 597-613.

[31] Wu F Y. The Potts model[J]. *Reviews of Modern Physics*, 1982, **54**(1): 235-268.

□