Hindawi
Mobile Information Systems
Volume 2022, Article ID 6389580, 11 pages
https://doi.org/10.1155/2022/6389580

Research Article

Road Sign Recognition Method Based on Segmentation and Attention Mechanism

Tianao Chen[1] **and Aotian Chen**[2]

[1]*University of Leeds, Southwest Jiaotong University, Chengdu 610000, China*
[2]*School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Aotian Chen; catchen2001@163.com

Received 12 April 2022; Revised 23 May 2022; Accepted 2 June 2022; Published 29 June 2022

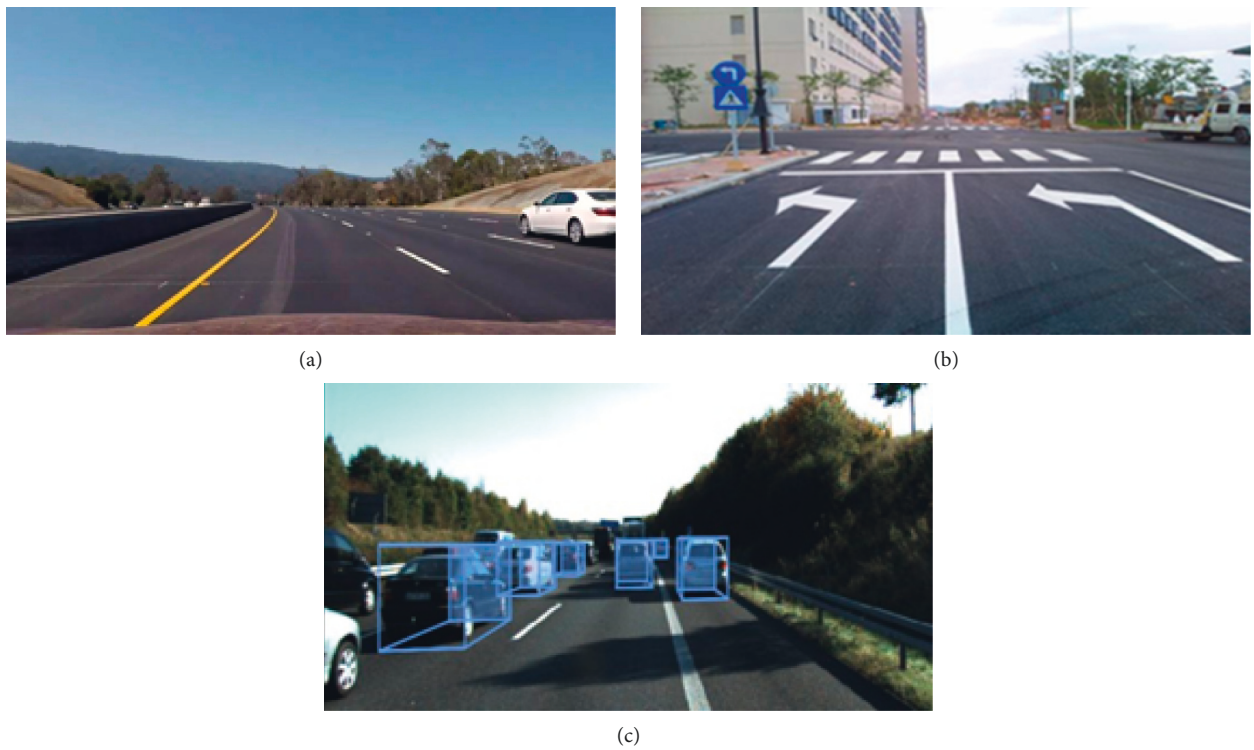
Academic Editor: Yajuan Tang

With the development of autonomous driving, low-cost visual perception solutions have become a current research hotspot. However, the performance of the pure visual scheme in unfriendly environments such as low light, rain and fog, and complex traffic scenes has a large room for improvement. Moreover, with the development and application of deep learning, the balance between the accuracy and real-time performance of deep learning models is a difficult problem for current research. Aiming at the problems of large differences in the target scale of pavement signs and the difficulty of balancing model accuracy and real-time performance, a ground semantic cognition method based on segmentation and attention mechanism is proposed. The lightweight semantic segmentation model ERFNet is used to realize the semantic segmentation of pavement signs and the instantiation of lane lines. When only lane line detection is required, the prediction branch of lane line existence is introduced based on the lightweight semantic segmentation model ERFNet to realize lane line instantiation cognition, solve the imbalance of positive and negative lane line detection samples, and obtain the final lane line detection result via postprocessing. Deep features were used to guide shallow layers to extract semantic features at high resolution, and the model performance was further optimized without increasing the inference cost.

1. Introduction

The number of motor vehicles is growing in tandem with the advancement of the economy and social growth. China will have 370 million motor cars by the end of 2020. The popularity of automobiles gives convenience to people's travel, but it also delivers a high number of traffic accidents. According to inadequate estimates, over 200,000 road accidents occur in China each year, resulting in up to 300,000 fatalities. According to studies, faulty driving actions cause more than 70% of traffic accidents. Traffic accidents are difficult to avoid because humans are subject to the natural restrictions of psychology and physiology. As unmanned driving technology becomes more and more mature, people hope to change this situation through autonomous driving. In recent years, with the continuous development of artificial intelligence technology, autonomous driving technology has also been applied in different fields, including intercity transportation, unmanned distribution, public transportation in the park, disaster relief, unmanned military equipment, and so on.

Automatic driving includes the technical links of perception, decision-making, control and positioning, among which ground semantic cognition and 3D object detection are important contents of automatic driving perception tasks. Ground semantic information contains different information such as lane lines and road signs, among which lane lines contain the extension direction of vehicle passable area (see Figure 1(a)). According to incomplete statistics, more than 50% of traffic accidents are related to the driver's active departure from the lane line. By detecting and recognizing lane lines, autonomous vehicles can drive safely in the original lane or make reasonable lane changes. At the same time, road marking is also one of the important



(a) (b)

(c)

Figure 1: Content of autopilot perception: (a) The lane line. (b) Ground marking. (c) 3D target detection.

research topics of automatic driving. Speed limit signs, directional arrows, stop lines, crosswalks, and other information (see Figure 1(b)) are examples of road signs that play a vital part in directing safe driving. Furthermore, in the urban road environment, there are numerous traffic participants, including motor vehicles, bicycles, and pedestrians, as well as other fixed obstructions. The 3D target detection technology combined with deep learning can collect categorization information and particular location of obstacle targets (as illustrated in Figure 1(c)), giving critical information for autonomous obstacle avoidance in self-driving vehicles. Therefore, we believe that it is of great significance to conduct cognitive and 3D object detection technology for semantic information on the ground including lane lines and road signs.

Research on autonomous driving originated in the United States. In the 1980s, the Army and Defense Advanced Research Projects Agency (DARPA), Defense Advanced Research Projects Agency) proposed the Autonomous Land Vehicle (ALV) plan [1], and successfully developed an eight-wheeled unmanned platform that can autonomously complete patrol tasks. Since then, DAPRA has held a number of driverless car competitions, attracting participation from universities including Carnegie Mellon and Stanford. Among them, in the cross-country race in 2015, Stanley [2] from Stanford University (see Figure 2(a)) successfully crossed the desert, tunnel, river bed, and other wild environments, reached the finish line first after 7 hours and won the championship. Google's Google X lab began developing self-driving cars in 2009, using sensors such as lidar, vision cameras, and millimeter-wave radar. In December 2016, Google announced the formation of Waymo, an autonomous driving company. In October 2019, Waymo announced the launch of Robotaxi, a driverless taxi service, in Phoenix, USA (see Figure 2(b)).

Compared with developed countries in Europe and the United States, the research on automatic driving technology in China is carried out late. In the 1990s, China's first autonomous vehicle, ATB-1 (Autonomous Test Bed1), was jointly developed by universities including the National University of Defense Technology, Beijing Institute of Technology, and Zhejiang University. During the ninth Five-Year Plan period, the second generation of unmanned platform ATB-2 has been successfully developed, and its performance has been improved compared with that of the first generation. In 2005, ATB-3 was successfully developed, and its environmental perception ability and motion control ability were further improved [3]. In addition to major universities and research institutions, domestic Internet giants, artificial intelligence enterprises, and major automakers have a layout in the field of autonomous driving, participating in the research boom of autonomous driving. In 2013, Baidu started relevant research on autonomous driving, and in December 2015, it conducted fully automatic driving tests on expressways and urban roads in Beijing (see Figure 2(c)). In April 2017, Baidu announced Apollo to provide an open, complete, and secure software platform to partners in the automotive industry and autonomous driving. At this point, the domestic autonomous driving research boom came to an unprecedented height.

There are now two camps based on different perceptual sensor methods. On one side, Waymo represents the Robotaxi camp, which opts for a sensor scheme that includes

(a) (b)

(c)

Figure 2: Self-driving cars from around the world: (a) Stanly from stanford. (b) Robotaxi from waymo. (c) Apollo from baidu.

rather pricey rotational lidar, as well as multichannel cameras and millimeter-wave radar. The image information package provides the target's texture and color information, whereas the point cloud information package contains the target's location information. The two complement each other well, allowing for the direct landing of L4 automated driving. Tesla, on the other hand, represents the progressive camp, which uses sensor schemes based on multichannel cameras reinforced by multichannel millimeter-wave radar. The cost of the sensor scheme is low, and relying on deep learning model and massive data, its business model determines that the pure vision scheme is the optimal solution for both driving experience and cost. At the 2019 CVPR conference, Baidu presented a purely visual solution Apollo Lite. So far, Baidu Apollo Lite has become the only pure visual L4 autonomous driving solution for urban roads in China, and Apollo Lite solutions are also used for autonomous parking products AVP and pilot-assisted driving products ANP, realizing the commercialization of L4 capability reduction.

2. Related Work

Ground semantics include lane lines, directional arrows, crosswalks, and other different information. According to the difference between detection needs and semantic information, pavement semantic cognition can be divided into two tasks: lane line detection and pavement marker detection.

2.1. Lane Detection. According to the different methods, lane detection can be divided into traditional method and deep learning method. According to different detection algorithms, lane line detection can be divided into feature-based and model-based methods.

The feature-based method uses image feature information such as color, edge, width, and so on, to segment the road surface and extract lane lines. Lane lines have simple qualities that can be separated into straight lines and curves with obvious edge and contour elements. The color information is in sharp contrast with the gray road surface, which is white and yellow, respectively. Cheng et al. [4] proposed to extract lane lines based on color information. Since there are vehicles with the same color as lane lines, other characteristics such as the size and shape of lane lines should be used to distinguish them, so as to eliminate the influence of road vehicles on lane lines. Deng and Wu [5] proposed a lane detection method based on constrained Hough transform bilateral extraction. First, Canny edge detection was used to extract lane line edge information, and lane lines were divided into straight line part and curve part. In the curve part, the least square method is used to fit the projectile line model. In order to improve the robustness of the algorithm under different lighting conditions, Hao et al. [6] proposed a graying method based on gradient enhancement. The grayscale transformation vector is transformed according to lighting, and after transformation, there is a large gradient transformation at the edge of the lane line, which can reduce

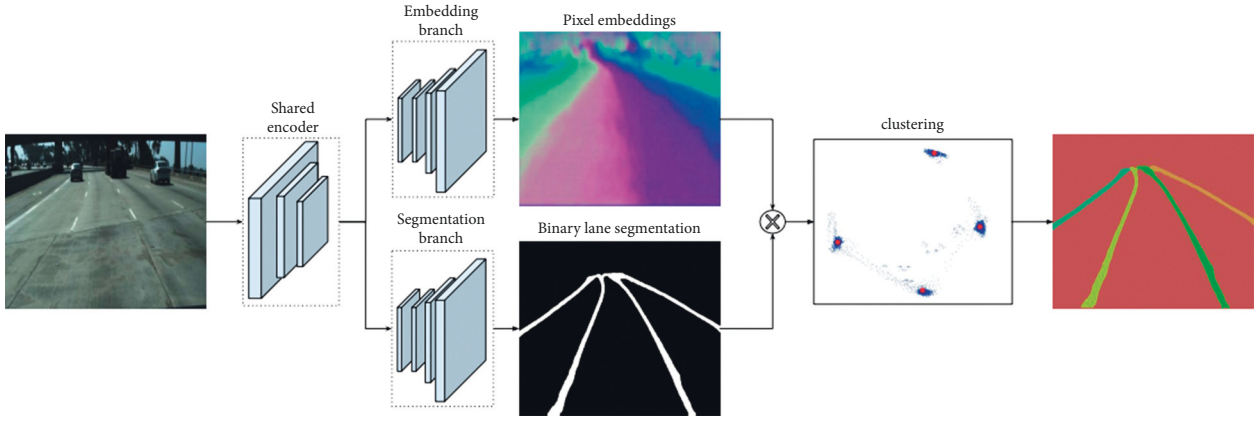


Figure 3: LanNet system framework diagram.

the influence of lighting changes on detection. The feature-based method is appropriate for environments with basic road conditions and obvious lane lines, but it cannot be applied to scenes with varying conditions in practice. Shadow, occlusion, rainy day, fog, and other unfriendly environment will affect the recognition effect, and the robustness is poor.

Model-based method refers to the process of transforming lane line detection into model parameters based on preset parameters of lane lines. Lane lines can be detected when the calculated results are the same or similar to preset parameters. Tabelini et al. [7] proposed a lane line detection algorithm based on particle filter, which segmented the lane and estimated the vanishing point. Finally, the hyperbolic model was used to fit the lane line. Haris and Glowacz [8] used linear model to fit lane boundaries and extract lane line features. Finally, traffic lines were classified by classifier, including dotted line, solid line, dotted line, dotted line and double solid line, with good illumination robustness. Yenİaydin and Schmidt [9] proposed a Gaussian probability density function fitting method, through which the histogram of the left and right areas of the image is obtained from the aerial view through Gaussian probability density function fitting, and lane lines are modeled in the region of interest. This approach can recognize lanes accurately in complicated settings, such as worn or curved lanes. When compared to the feature-based method, the model-based method can lessen the impact of other road disturbances and has greater robustness. However, its iterative operation procedure is complex, and the calculation amount is enormous, and it cannot detect situations that are not in the template.

Because the traditional methods cannot be applied to different scenarios, the robustness cannot meet the practical requirements, and the processing and calculation process is complex, with the rapid development of deep learning, people begin to try to use convolutional neural network to detect lane lines. DVCNN [10] proposed by Yan et al. uses a dual-perspective (front-view and top-view) CNN model for lane line detection. Front-view images can eliminate misjudgments caused by moving vehicles, fences, and road boundaries, while top-view images can be obtained through reverse perspective transformation, which can remove rod-shaped structures, such as arrows and characters on the road. De Brabandere et al. [11] proposed an end-to-end lane line detection method, which is divided into two parts: a deep network that predicts the piecewise weight diagram of each lane and a differentiable least square fitting module used to return fitting parameters. In addition, some scholars regard lane line detection as a segmentation problem.

Neven et al. [12] proposed an end-to-end lane line detection method, which is composed of LaneNet and lane line fitting module. The system block diagram is shown in Figure 3. LaneNet is made up of two branches: segmentation and embedding. The segmentation branch produces the binary lane line mask, but the embedded branch produces a multidimensional embedding for each lane pixel, making the embedding from the same lane close to the manifold and the embedding from different lanes far from the manifold. LaneNet's output results transform lane pixels by using the transformation matrix output by H-NET, fit a third-order polynomial for each lane, and re-project the lane onto the image. Pan et al. proposed Spatial CNN [13], which converts the traditional convolution layer-by-layer connection form into the form of strip-by-strip convolution in feature graph, so that information can be transmitted between pixel rows and columns in the graph. It is suitable for detecting long-distance continuous shape targets or large targets, and has good extensibility in lane line detection. Hou et al. [14] proposed self-attentional distillation (SAD) for lane detection, and realized further learning by implementing top-down and hierarchical attention distillation networks within the network. Qin et al. [15] regard lane line detection as a row selection problem based on global features. Lane lines are encoded, positioned, and classified based on row direction, and lane lines are modeled through structural loss to achieve lane line detection of 300+ frames. Liu et al. [16] proposed a lane-line detection model based on transformer. The transformer network constructed uses self-attention mechanism to model nonlocal interaction and can learn richer structural information and context information compared with other models. For the scenario of lane line detection in curves, Huawei proposed the lane-sensitive architecture search framework Curvelane-NAS [17], which

can capture long-distance coherent and short-distance accurate curve information, extract features through feature fusion search module and elastic trunk search module, and complete postprocessing through adaptive point mixture module.

2.2. Road Marking Detection. Road signs refer to signs painted on the road surface, such as directional arrows, speed limit signs, crosswalks. Pavement sign detection can also be divided into traditional methods and deep learning methods.

The traditional method of pavement sign detection is usually based on feature or model. Xu et al. [18] used the corner detector of FAST to detect a group of interest points and used the positions and feature vectors of interest points extracted from all template images to construct the template dataset. Finally, the structure-matching algorithm was used to test whether the subset of matched interest point pairs formed road signs matching the road signs in the die plate image. Ahmed et al. [19] used the improved Hu invariant moment to construct the feature vector of the image and used THE SVM classifier to classify three typical road traffic signs, including straight-ahead signs, straight-right turn signs, and left-right turn signs, but the recognition categories were few and real-time detection could not be achieved. There are also some people use the combination of lane lines and road signs. Yao et al. [20] integrated information between lane lines and road signs to strengthen the association between semantic information on the ground.

Deep learning is also widely used in pavement sign detection. He et al. [21] proposed VPGNet, which is a joint end-to-end trainable multitask network that uses quench-point information to supplement features and can simultaneously detect and recognize roads and pavement signs under extreme weather conditions.

Many research accomplishments have been made in lane line detection, road sign detection, and other ground semantic cognition methods, including standard and deep learning methods. After parameter adjustment, the traditional technique performs well in a given case, but it has weak robustness and practical utility in other contexts. Methods based on deep learning although gradually with the increase of training samples can meet the requirements of robustness, but about how to balance the ground semantic cognition the accuracy and real-time performance of deep learning model, and the semantic ground target dimension differences, the friendly environment detection have some challenging difficulties, such as has not been targeted research. If we can design some functional modules on the basis of a lightweight detection model, supplemented by the data enhancement method of specific environment, we can solve the above difficulties to a certain extent.

2.3. Image Semantic Segmentation. Image semantic segmentation is the basic task in image segmentation, which means that each pixel in the image is labeled with the corresponding category, without distinguishing individuals. Before the popularity of deep learning methods, semantic segmentation methods based on traditional machine learning classifiers such as gray segmentation and random forest were commonly used. With the development of deep learning, people begin to use deep learning model to complete semantic segmentation. According to different model principles, image semantic segmentation based on deep learning can be divided into semantic segmentation based on candidate regions and semantic segmentation model based on full convolution.

Semantic segmentation based on candidate regions first generates candidate regions in the image in a free form, then gradually extracts features from the candidate regions and classifies them, and lastly translates the region-based classification prediction into pixel-level prediction. Maskrcnn [21] proposed by He et al. team is a branch that adds a predictive dividing mask to Fasterrcnn [22]. Like Faster R-CNN, MaskRCNN adopts a two-stage method, including candidate region generation and classification regression prediction. As shown in Figure 4, first, the candidate regions are generated by RPN, Region Proposal Network), and then each candidate region is classified and located by using a unique regional feature aggregation method (RoI Align). At the same time, the mask branch realizes the decoupling between mask and category prediction and the semantic segmentation task at pixel level. However, because the semantic segmentation method based on candidate regions will generate a large number of candidate regions, it has certain redundancy, which will increase the computational cost and cannot meet the real-time requirements to some extent.

The full convolution-based semantic segmentation model adopts the full convolution network without the full connection layer and achieves the semantic segmentation results by convolution and deconvolution. Its classic representatives include FCN [23] and DeepLab series [24–26]. The proposal of FCN [27] changed the previous idea that the semantic segmentation task needs to be transformed into the image classification task through candidate regions, and features were extracted through the convolution layer and refined semantic segmentation results were obtained through deconvolution. Using a deconvolution layer instead of full connection layer can avoid the loss of spatial information caused by the compression of two-dimensional matrix into one-dimensional matrix, which is more suitable for semantic segmentation. Plabv1 [24] combines deep convolutional neural network (DCNNs) and probability graph model (Dense CRF), and uses Dense CRF as the postprocessing method of the network, which makes the boundary of semantic segmentation results clearer. Because repeated pooling operation and downsampling process will reduce the resolution of feature map, the enlargement of receptive field is also very important for semantic segmentation task, and there is a certain contradiction between them.

To address this issue, DeepplaBV2's hole convolution method [25] widens the receptive area of the feature extraction process while maintaining the feature map's resolution and increasing Atreus spatial pyramid pooling (ASPP). Structure, extract features by using multiple expanded convolutions with different sampling rates, and then fuse different features to obtain context information of

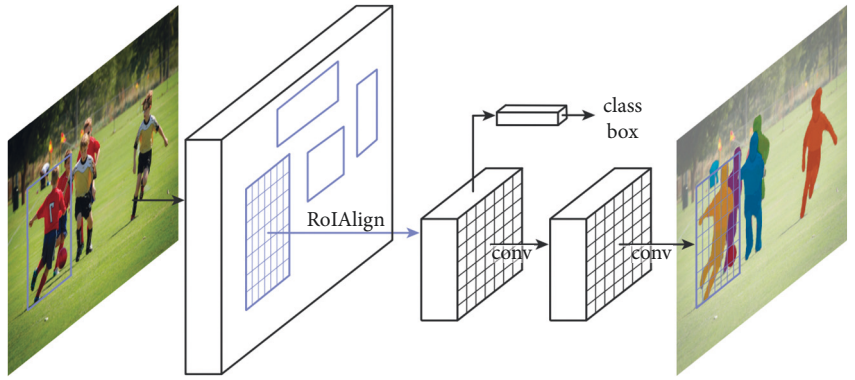


Figure 4: MaskRCNN model frame diagram.

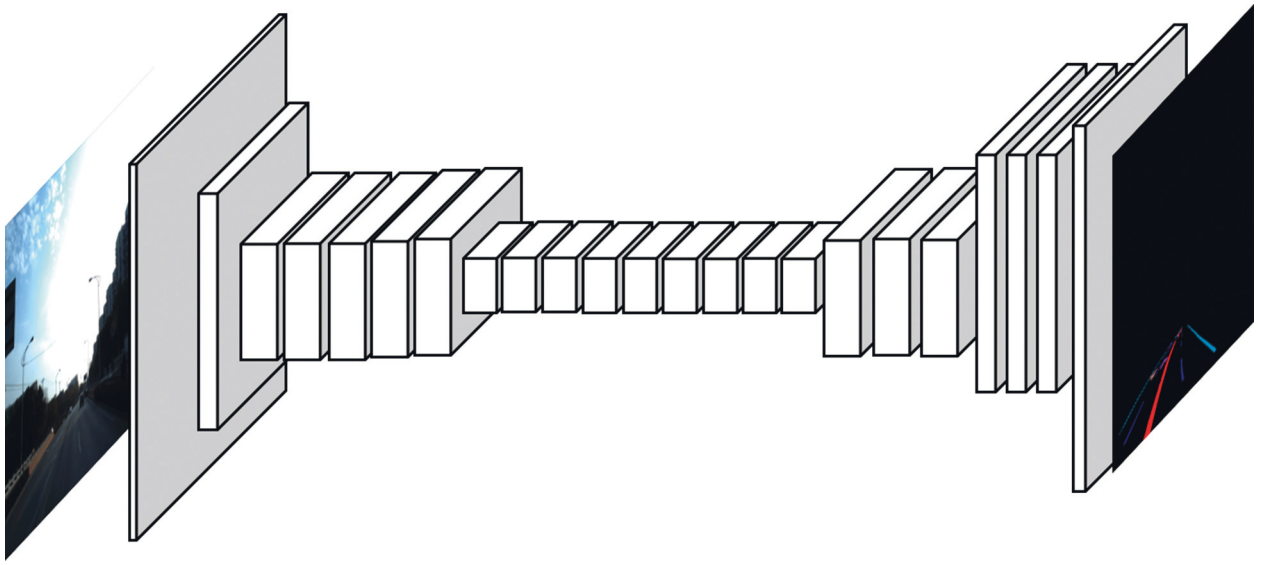


Figure 5: Framework diagram of erfnet model.

different sizes. Plabv3+ [26] adds a decoding and coding module to restore the original pixel information, so that the details of semantic segmentation can be better retained, and rich context information is encoded at the same time.

3. Ground Semantic Segmentation Based on the Coding and Decoding Model

Road sign recognition adopts semantic segmentation method based on deep learning. Semantic segmentation model adopts lightweight semantic segmentation model ERFNet, which follows the network structure of encoding and decoding, and completes the work of ground semantic segmentation on the premise of ensuring real-time.

3.1. Lightweight Coding and Decoding Model: ERFNet. To meet the real-time requirements of the ground semantic segmentation task, this topic adopts a lightweight semantic segmentation model ERFNet as the model benchmark, whose core is residual connection and 1D convolution kernel, which can alleviate the problem of gradient disappearance and reduce the amount of computation to a certain extent. ERFNet follows the network structure of encoding and decoding, and its model frame is shown in Figure 5. The encoder extracts and encodes the feature information of the ground semantics, and gradually obtains a multiscale down sampled feature map. The feature map is sampled by the decoder, which is consistent with the resolution of the input image, and a finer semantic segmentation result is obtained.

The introduction of residual layer can promote feature learning, which is used to alleviate the problems of gradient disappearance, gradient explosion, and model degradation caused by too deep network structure. The relationship between its output vector y and a layer vector input x is

$$y = F(x, W_i) + W_s x. \tag{1}$$

Among them, W_s is characteristic mapping and W_i is residual mapping to be learned. The original residual layer of is nonbottleneck structure (nonbottleneck) and bottleneck structure (bottleneck), as shown in Figure 6(a) and Figure 6(b). Nonbottleneck contains two 3×3 convolution kernels. In contrast, bottleneck only contains a 3×3 convolution kernel, which can achieve the accuracy similar to that of nonbottleneck with less computational cost.

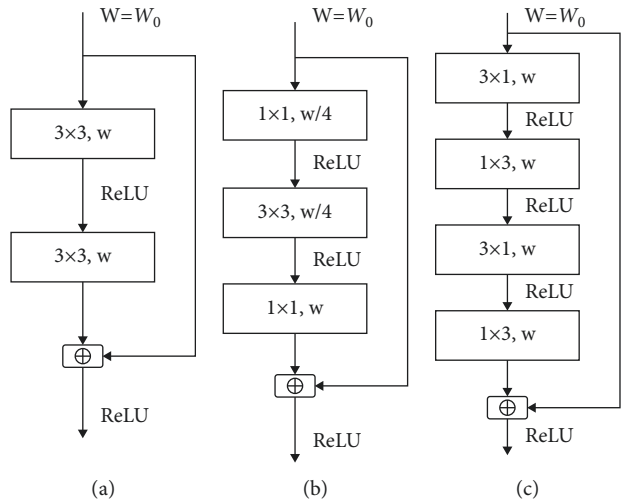


Figure 6: Different designed residual layers: (a) Nonbottleneck. (b) Bottleneck. (c) Nonbottleneck 1D.

Table 1: ERFNet model architecture distribution.

	Serial number	Module type	Number of output channels	Down-sampling multiple
Encoder	1	Downsample block	16	2
	2	Downsample block	64	4
	3–7	Non-bottleneck-1D × 5	64	4
	8	Downsample block	128	8
	9	Nonbottleneck-1D(2-dilated)	128	8
	10	Nonbottleneck-1D(4-dilated)	128	8
	11	Nonbottleneck-1D(8-dilated)	128	8
	12	Nonbottleneck-1D(16-dilated)	128	8
	13	Nonbottleneck-1D(2-dilated)	128	8t
	14	Nonbottleneck-1D(4-dilated)	128	8
	15	Nonbottleneck-1D(8-dilated)	128	8
	16	Nonbottleneck-1D(16-dilated)	128	8
Decoder	17	Upsample block	64	4
	18–19	Nonbottleneck-1D × 2	64	4
	20	Upsample block	32	2
	21–22	Nonbottleneck-1D × 2	32	2
	23	Upsample block	C	1

However, as the number of layers increases, the accuracy of nonbottleneck is higher, and bottleneck still has the problem of degradation. Therefore, combined with their advantages and disadvantages, ERFNet designed the nonbottleneck-1d module, which can obtain higher precision with lower computation, as shown in Figure 6(c). Nonbottleneck-1d module uses 3×1 convolution kernel and 1×3 convolution kernel instead of 3×3 convolution kernel, which can reduce the parameter quantity by about 30% without affecting the accuracy. The design of 1D convolution kernel can greatly reduce calculation consumption, improve model compactness and learning ability, and keep the same accuracy as 2D convolution kernel.

In order to give consideration to the accuracy and real-time of the detection model, ERFNet adopts a more orderly architecture. The encoder extracts and encodes the feature information of the ground semantics to generate a down-sampled feature map, and the decoder upsamples the feature map to the input resolution, so as to obtain a finer semantic segmentation result. See Table 1 for details of the architecture distribution of the ERFNet model. The encoder consists of 1 to 16 layers, and Downsample Block and a Nonbottleneck-1D. Among them, inspired by ENet, Downsample Block contains 2×2 maximum pool layer and 3×3 convolution kernel. With the increase of down-sampling multiple, the number of channels of output feature map also increases. At the same time, when the down-sampling times are 8, nonbottleneck-1D uses different sizes of expansion convolutions (2-dilated, 4-dilated, 8-dilated, and 16-dilated) in different modules alternately, so as to avoid oversampling the feature map and obtain a larger receptive field at the same time, so as to obtain more context information and enter the next layer. The decoder is 1723 layer, including Upsample Block and Nonbottleneck-1D. The Upsample Block is a simple deconvolution operation with a step size of 2, and the feature map is gradually restored to the original resolution. The final output size of the model is $H \times W \times C$, where H is the height of the input image, W is

Figure 7: Ground semantic segmentation result.

the width of the input image, and C is the number of classification categories. In addition, nonbottleneck-1D also uses Dropout, which can effectively alleviate overfitting and achieve regularization effect to a certain extent.

The $H \times W \times C$ results output by the model are processed by softmax function to obtain the probability that each pixel belongs to each semantic category, namely:

$$q_i = y(c_i) = \frac{e^{x_i}}{\sum_{j=1}^{C} e^{x_j}}, \tag{2}$$

where x_i is the input value of I channel of a pixel in the last layer.

As the proportion of training samples in different semantic categories is quite different, and there are some semantic categories that are difficult to train. In order to alleviate the imbalance of multiclass samples and improve the training efficiency, this paper uses the weighted cross entropy loss function as the optimization function, and its output is as follows:

$$\text{Loss} = -\sum_{j=1}^{C} \lambda_j p_j \log\left(q_j\right). \tag{3}$$

Among them, p_j is the label of class j. If the pixel belongs to class j, $p_j = 1$, otherwise $p_j = 0$. λ_j is a super parameter of category j, which is preset according to the proportion of training samples. For the categories of edge lanes or road signs that are difficult to detect or less distributed, higher weights can be set to help the model focus on learning some information that is difficult to learn.

After the probability map output by the model is dyed, the segmentation result of ground semantics in the road sign dataset ApolloScape is shown in Figure 7. Categories include single solid line, double solid line, dotted line, crosswalk, left-turn arrow, straight arrow, right-turn arrow, stop line, etc.

3.2. Lane Detection Based on Case Segmentation. In this paper, the task of lane line detection is realized based on example segmentation, in order to solve the problems of imbalance between positive and negative samples and the inability to distinguish different lane lines in lane line detection. The lane detection method is divided into two steps: deep learning model and postprocessing. The pixel-level probability distribution of multilane lines is obtained through a deep learning model, and the final lane detection result is obtained through postprocessing.

In the lane detection model based on deep learning, a two-stage method is usually adopted, which is divided into the deep learning model and postprocessing part. The deep learning model predicts multiple points of each lane line, and the final lane line detection result is obtained by postprocessing the fitting curve. However, due to the serious imbalance between positive and negative samples in the lane line, direct prediction of multiple points in the lane line will lead to a large number of predicted backgrounds, thus affecting the learning effect of the model. Therefore, in the scene where only lane line detection is needed, we regard lane line detection as a segmentation problem, and transform the label of lane line into a curve with a certain pixel width, thus increasing the positive sample ratio of lane line and alleviating the imbalance between positive and negative samples to a certain extent. At the same time, this paper instantiates each lane line to distinguish lane lines in different positions.

There is a line prediction branch, which is used to guide the model to learn and converge better and to predict whether there is a lane line at the corresponding position. The lane detection model includes a coding layer and branch layer, and its model framework is shown in Figure 8. Among them, the branch layer includes the branch of decoder and lane line prediction. The decoder outputs the lane line prediction probability map with the size of $h \times w \times c$. The lane existence prediction branch predicts whether there is a

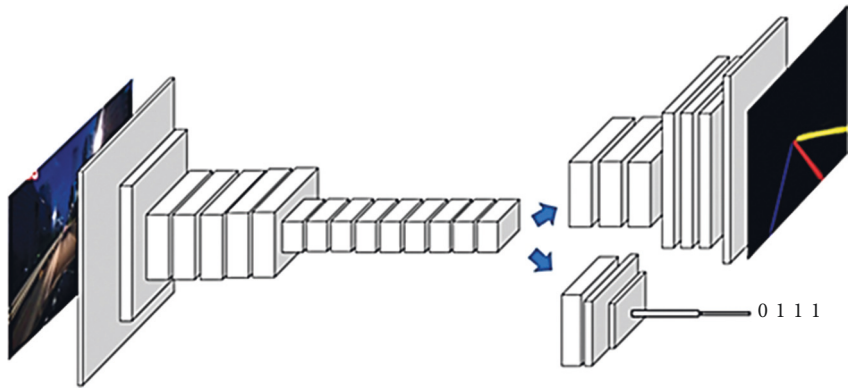


Figure 8: Lane detection model.

lane line at a specific position, and the output result size is $1 \times 1 \times C$, 1 means that the lane line at the corresponding position exists, and 0 means that it does not exist. Where h is the height of the input image, W is the input image width, and c is the number of specific positions of the lane line to be predicted. Here $C = 4$, only two lane lines, left lane line and right lane line of the main vehicle lane are predicted.

The intermediate result output by the decoder gets the probability that each pixel belongs to each lane line through the softmax function, that is,

$$q_i = y(c_i) = \frac{e^{x_i}}{\sum_{j=1}^{4} e^{x_j}}, \tag{4}$$

where x_i is the input value of the I-th channel of a pixel in the last layer.

As the proportion of training samples of different lane lines is quite different, and the inspection of left lane line and right lane line, there is a big challenge in lane detection. Focal Loss is used as the optimization function in lane detection model. Calloss loss function is improved on the basis of standard cross entropy loss, which can reduce the weight of easy-to-classify samples and make the model pay more attention to hard-to-classify samples in training. Its output is

$$L_{\text{seg}} = -\sum_{j=1}^{4} p_j a_j \left(1 - q_j\right)^{\gamma} \log\left(q_j\right). \tag{5}$$

Among them, p_j is the label, if the pixel belongs to the lane line j, then $p_j = 1$, otherwise $p_j = 0$. The hyper parameter α is used to control the shared weight of positive and negative samples to L_{seg}, and the modulation coefficient γ is used to reduce the weight of easy-to-classify samples so that the model can focus more on hard-to-classify samples during training. Through the introduction of focal loss, problems such as uneven distribution of positive and negative samples and imbalance of difficult and easy samples can be alleviated.

The output result of lane existence prediction branch adopts the binary cross entropy loss function as the optimization function, and its output is

$$L_{\text{exit}} = -\sum_{j=1}^{4} \left[y_i \log\left(\hat{y}_j\right) + (1 - y_i) \log\left(1 - \hat{y}_j\right) \right], \tag{6}$$

y_j is the label of the J-th lane line, and $y_j$is the output of predicting whether the J-th lane line exists or not. The total loss value of includes L_{seg} and L_{exit}, as follows:

$$\text{Loss} = \lambda_1 L_{\text{seg}} + \lambda_2 L_{\text{exit}}. \tag{7}$$

Among them, λ_1 and λ_2 are super parameters of preset values, which are used to balance L_{seg} and L_{exit}. The output result of lane detection model encoder is dyed, which is in the lane line dataset CULane [13].

The results are shown in the second column of Figure 9. Different colors are used to distinguish lane lines in different positions, among which green is the left lane line, blue is the left lane line, red is the right lane line, and yellow is the right lane line.

4. Experiment

The dataset used in the lane line detection experiment is the lane line detection dataset CULane. The batch size used for training is 12, and the number of training times is 40 epochs. The training initial learning rate size is 0.015 and adopts the training strategy of linear decline of the learning rate, and the optimizer adopts stochastic gradient descent. Similar to the ground-based semantic segmentation experiments, a pre-trained model trained on the Cityscapes dataset is used for initialization, and data augmentation methods such as random cropping, random flipping, and random translation are used.

The ERFNet method in this paper has obvious improvement in some challenging scenes, such as low-light scenes, the F1-Measure of night scenes is increased by 1.7%, and the F1-Measure of shadow scenes is increased by 1.4%. In addition, other challenging scenes have different degrees of improvement, such as the F1-Measure of the wireless scene is increased by 1.6%, and the F1-Measure of the curve scene is increased by 1.7%. Other current mainstream lane line detection methods, such as SCNN [13], ENetSAD [14] and ResNet¬101¬SAD [14], achieve F1-Measure of 71.6%, 70.8%, and 71.8%, respectively. The benchmark method in this paper and the proposed EAF-ERFNet are superior to other current mainstream lane line detection methods. At the same time, on the graphics card 2080Ti, the operation speed of ERFNet can reach 98.0fps, and the operation speed

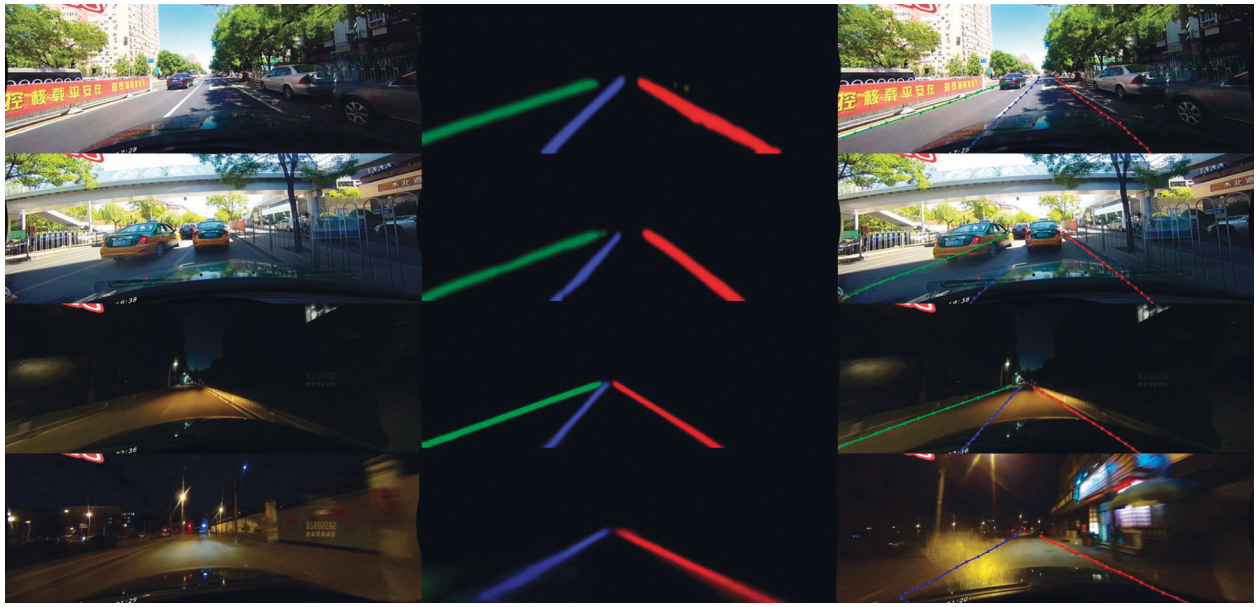


Figure 9: Lane detection result.

Table 2: Comparison of lane line detection methods.

Scenes	ERFNet	SCNN [13]	ENet-SAD [14]	UFAST [15]
Normal	**91.8**	90.6	90.1	90.7
Crowded	**72.3**	69.7	68.8	70.2
Night	**68.8**	66.1	66.0	66.7
Wireless	**46.7**	43.4	41.6	44.4
Shadow	**72.7**	66.9	65.9	69.3
Arrow	**87.7**	84.1	84.0	85.7
Glare	65.1	58.5	60.2	59.5
Curve	68.0	64.4	65.7	**69.5**
Intersection	2254	**1990**	1998	2037
Total	**74.0**	71.6	70.8	72.3
Frame rate	62.6	7.5	74.6	**175.4**

of EAF¬ERFNet can reach 62.6fps, which takes into account the requirements of accuracy and real-time performance (Table 2).

5. Conclusion

Aiming at the difficulty of balancing the real-time and accuracy of the task model of ground semantic segmentation, the large-scale difference of ground semantic targets and the challenges of detection in unfriendly environment, this paper proposes a road sign recognition method based on segmentation and attention mechanism. First, the ground semantic segmentation is realized by the lightweight semantic segmentation model ERFNet. When only lane detection is needed, the prediction branch of lane existence is introduced on the basis of lightweight semantic segmentation model ERFNet to realize the instantiation cognition of lane, and lane postprocessing is realized through point extraction and curve fitting to obtain the final lane detection result. When only lane line detection is needed, the prediction branch of lane line existence is introduced based on the lightweight semantic segmentation model ERFNet to realize lane line instantiation cognition, solve the imbalance of positive and negative samples of lane line detection, and obtain the final lane line detection result through post-processing. Deep features were used to guide shallow layers to extract semantic features at high resolution, and the model performance was further optimized without increasing the inference cost.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] A. I. Abu Bakar, M. A. Abas, M. F. Muhamad Said, and T. A. Tengku Azhar, "Synthesis of autonomous vehicle guideline for public road-testing sustainability," *Sustainability*, vol. 14, no. 3, p. 1456, 2022.

[2] C. P. Hsu, B. Li, B. Solano-Rivas et al., "A review and perspective on optical phased array for automotive LiDAR," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 27, no. 1, pp. 1–16, 2020.

[3] M. Parker, B. Quinn, J. Bates et al., "Exploring cold regions autonomous operations," *Journal of Terramechanics*, vol. 96, pp. 159–165, 2021.

[4] Y. Chen, P. K. Wong, and Z.-X. Yang, "A new adaptive region of interest extraction method for two-lane detection," *International Journal of Automotive Technology*, vol. 22, no. 6, pp. 1631–1649, 2021.

[5] G. Deng and Y. Wu, "Double lane line edge detection method based on constraint conditions Hough transform," in *Proceedings of the 17th International Symposium on Distributed*

Computing and Applications for Business Engineering and Science (DCABES), Wuxi, China, October 2018.

[6] J. Hao, S. Luo, and L. Pan, "Computer-aided intelligent design using deep multi-objective cooperative optimization algorithm," *Future Generation Computer Systems*, vol. 124, pp. 49–53, 2021.

[7] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Keep your eyes on the lane: real-time attention-guided lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 294–302, Nashville, TN, USA, 2021.

[8] M. Haris and A. Glowacz, "Lane line detection based on object feature distillation," *Electronics*, vol. 10, no. 9, p. 1102, 2021.

[9] Y. Yenİaydin and K. W. Schmidt, "A lane detection algorithm based on reliable lane markings," in *Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, Izmir, Turkey, May 2018.

[10] L. Yan, S. Hu, and C. Zhang, "Lane detection based on improved FCN," in *Proceedings of the 2021 33rd Chinese control and decision conference (CCDC)*, pp. 887–892, Kunming, China, May 2021.

[11] B. De Brabandere, W. Van Gansbeke, D. Neven, M. Proesmans, and L. Van Gool, "Endtoend lane detection through differentiable leastsquares fitting," 2019, https://arxiv.org/abs/1902.00293.

[12] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards endtoend lane detection: an instance segmentation approach," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium*, pp. 286–291, Changshu, China, June 2018.

[13] X. Pan, J. Shi, P. Luo, and X. Tang, "Spatial as deep: spatial cnn for traffic ccene understanding," *Thirty Second AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[14] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning lightweight lane detection cnns by self attention distillation," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, no. 1, pp. 1013–1021, 2019.

[15] Z. Qin, H. Wang, and X. Li, *Ultra fast structure aware deep lane detection*, Springer, Cham, 2020.

[16] R. Liu, Z. Yuan, T. Liu, and Z. Xiong, "End to end lane shape prediction with transformers," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3694–3702, Waikoloa, HI, May 2021.

[17] P. Shyam, K. J. Yoon, and K. S. Kim, "Weakly supervised approach for joint object and lane marking detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2885–2895, Montreal, BC, Canada, October 2021.

[18] S. Xu, J. Wang, P. Wu, W. Shou, X. Wang, and M. Chen, "Vision-based pavement marking detection and condition assessment-A case study," *Applied Sciences*, vol. 11, no. 7, p. 3152, 2021.

[19] N. Ahmed, S. Rabbi, S. Rabbi, T. Rahman, R. Mia, and M. Rahman, "Traffic sign detection and recognition model using support vector machine and histogram of oriented gradient," *International Journal of Information Technology and Computer Science*, vol. 13, no. 3, pp. 61–73, 2021.

[20] L. Yao, C. Qin, Q. Chen, and H. Wu, "Automatic road marking extraction and vectorization from vehicle-borne laser scanning data," *Remote Sensing*, vol. 13, no. 13, p. 2612, 2021.

[21] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN,", in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 1, no. 1, pp. 2961–2969, Bordeaux, France, October 2017.

[22] Z. He and L. Zhang, "Multi-adversarial faster-rcnn for unrestricted object detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, vol. 1, no. 1, pp. 6668–6677, 2019.

[23] W. Sun and R. Wang, "Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with DSM," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 3, pp. 474–478, 2018.

[24] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[25] J. Liu, X. Xu, Y. Shi, C. Deng, and M. Shi, "RELAXNet: residual efficient learning and attention expected fusion network for real-time semantic segmentation," *Neurocomputing*, vol. 474, pp. 115–127, 2022.

[26] H. Byun, J. Kim, D. Yoon, I.-S. Kang, and J.-J. Song, "A deep convolutional neural network for rock fracture image segmentation," *Earth Science Informatics*, vol. 14, no. 4, pp. 1937–1951, 2021.

[27] S. Zhang, Z. Ma, G. Zhang, T. Lei, R. Zhang, and Y. Cui, "Semantic image segmentation with deep convolutional neural networks and quick shift," *Symmetry*, vol. 12, no. 3, p. 427, 2020.