International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 8385-8397

https,//doi.org/10.70517/ijhsa463717

Research on Semantic Segmentation Algorithm for Lane Lines Based on Multiscale Deep Feature Fusion

Rao Li[1,*], Yaxiong Tao[1] and Lingfeng Chen[2]
[1] College of Communication Engineering, Chongqing Polytechnic University of Electronic Technology, Chongqing, 401331, China
[2] School of Information Engineering, Chongqing Vocational and Technical University of Mechatronics, Chongqing, 400000, China
Corresponding authors: (e-mail: lirao@cqcet.edu.cn).

Abstract Lane line detection is a key technology to realize autonomous driving, which is a fundamental and challenging task in autonomous driving. In this paper, a semantic segmentation algorithm for lane lines based on multi-scale deep feature fusion is proposed. By analyzing the spatial structural properties of continuous elongated lane lines, we design a multimorphic CASPP module, which combines the mutual quality null rate with 1D convolutional branching to enhance the context-awareness of elongated linear features. The DeepLab-ERFC model is further constructed to introduce the enhanced boundary learning of ER Loss based on Hausdorff distance, combined with dynamic gradient correction to alleviate the category imbalance problem, and optimize the prediction boundary using the post-processing of fully-connected CRFs. Experiments on TuSimple, VPG and tvtLANE datasets show that the model significantly outperforms mainstream methods in both accuracy and speed, with average intersection and merger ratios of mIoU reaching 64.62%, 68.79% and 64.62%, respectively, which is an improvement of 2.12-8.31 percentage points over models such as DANet and PSPNet. In terms of real-time, the inference speed reaches 89.34 FPS, which is more than 2.6 times higher than the comparison model. The ablation experiment verifies the effectiveness of the multi-module synergistic optimization, with the CASPP module increasing the mIoU by 5.75%, the ER Loss with gradient correction by a further 6.86%, and the CRFs post-processing finally pushing the mIoU to 64.62%. Under extreme scenarios (e.g., sudden changes in tunnel light, vehicle occlusion, rain and snow interference), the average accuracy of the model improves by 3.8-21.3 percentage points over the suboptimal method, demonstrating strong robustness. The model constructed in the article significantly improves the accuracy, stability and real-time performance of lane line detection, thus realizing safer and more efficient autonomous driving technology.

Index Terms multi-scale deep feature fusion, lane line detection, semantic segmentation algorithm, CASPP module, ER Loss

I. Introduction

Lane lines, as key traffic signs on the road, assume the important roles of dividing lanes, indicating the direction of travel, and providing navigation for pedestrians, which are crucial to ensure the safe driving of motor vehicles [1]. In the field of intelligent driving, lane line detection, as one of the core technologies, is widely used in advanced assisted driving systems (ADAS) to realize the functions of lane departure warning, lane keeping assistance, and forward collision warning [2]-[4]. However, in real-world driving environments, lane lines may be missing or discontinuous due to long-term wear and tear, occlusion by pedestrians and vehicles, and their visibility is also affected by a variety of factors such as climate, lighting, road shadows, and wear and tear of the lane lines themselves [5]. In addition, the lane line detection task also needs to meet the real-time requirements, which brings many challenges to the lane line detection technology.

In the environment sensing system of autonomous driving, realizing fast and accurate lane line detection is crucial to ensure the safe and reliable driving of autonomous vehicles. There are many existing lane detection methods, including computer vision methods based on camera sensors, three-dimensional information acquisition methods based on LiDAR, and GPS methods [6]-[9]. In recent years, given that computer vision methods have advantages such as low cost, high adaptability and good real-time performance, they have been widely used in lane line detection tasks such as highways and urban roads [10]. Meanwhile, through the subsequent development of computer technology, the optimization from the perspective of computer vision algorithms can help to further improve the accuracy and real-time performance of lane line detection technology.

Techniques are widely used before lane lines, for example, traditional image processing methods such as the Hough transform are the main methods for dealing with lane line detection [11]. For example, Bi W et al [12] proposed a color edge road lane line extraction algorithm based on quaternion Hardy filter, which first inputs a color road image, followed by a quaternion Hardy filter based edge enhancement method to get a smooth road image, then a color gradient detector is used to record the edges, and finally the lane lines are extracted by combining with the Hough transform. Huang, Q and Liu, J [13] pointed out that the traditional current lane detection algorithm based on Hough transform has certain defects in challenging scenarios such as different lighting conditions, and thus the Hough transform algorithm needs to be improved. Subsequently, the key feature information is refined by an adaptive algorithm, and finally the Hough transform algorithm is utilized to effectively detect the location of lane lines [14]. Wei, Y and Xu, M [15] improved the lane line detection method of Hough transform by replacing the Canny operator with the Robert operator, this move reduces the detection time and improves the real-time detection performance, which improves the safety and efficiency of self-driving cars to some extent. Javeed, M. A et al [16] proposed a fast Hough transform method based on ossu thresholding and canny edge detection aiming at high lane detection accuracy for autonomous vehicles. In addition, Tian, J et al [17] proposed a lane line detection and tracking technique for ADAS that incorporates a line segment detector (LSD), an adaptive angle filter, and a dual Kalman filtering system, focusing on analyzing the limitations of the traditional lane line detection techniques and pointing out that these techniques are usually effective only under specific environmental conditions, which is rooted in the lack of complex and dynamic scene and dynamic scenarios. In general, traditional lane line detection methods usually have good detection results on well-lit highways with clear lane lines and fewer vehicles. However, the lane line features are easily interfered by external complex environmental factors and have poor robustness. Moreover, the model is complex and computationally intensive, which affects the real-time lane line detection. It can be seen that the traditional lane line detection methods are difficult to cope with the current increasingly complex and changing road scenes.

With the continuous improvement of computer arithmetic power and datasets, more and more semantic segmentation algorithms based on deep learning come into being, and they have gained significant progress in segmentation accuracy and speed, and semantic segmentation has achieved extraordinary performance on many datasets, which has greatly contributed to the improvement of lane line detection accuracy [18]-[20]. For example, Chougule, S et al [21] considered the lane line detection and classification problem as a convolutional neural network (CNN) regression task, and designed their network to classify only a few points on the lane line boundary at the pixel level and output parameterized lane line boundaries in the form of image coordinates.This approach eliminates the stringent criterion for correctly classifying each pixel point, and improves the segmentation of lane line boundary Accuracy. Yousri, R et al [22] proposed a benchmarking framework for lane detection in complex dynamic road scenes, which combines computer vision techniques with deep learning, and it demonstrated high performance in a variety of complex scenes and lighting conditions. Zou Q et al [23] utilized the inter-frame relationship of continuous images to effectively integrate CNN and recurrent neural network (RNN) to construct an end-to-end network structure, which effectively improves the robustness of off-road lane detection for complex roads without reducing the detection speed. Al Mamun, A et al [24] proposed a deep learning instance segmentation method based on the U-net framework and the VGG16 architecture, aiming to improve the segmentation accuracy of lane markings under various environmental conditions. Overall, deep learning-based lane line detection methods can automatically learn features and perceive road scenes well without complex pre-processing and post-processing operations. However, under complex road conditions such as lane line degradation or occlusion, lane line detection is easily interfered by various external factors, resulting in weak model generalization ability and inadequate or inaccurate extraction of lane line features. Meanwhile, the currently proposed deep learning-based lane line detection model generally has a large number of parameters and high complexity, which is not conducive to real-time lane line detection. Therefore, further optimization is needed.

In this paper, a semantic segmentation algorithm based on multi-scale deep feature fusion is proposed for the lane line detection task. Lane line detection can be regarded as a special image semantic segmentation task, the core of which is to make full use of the inherent spatial structure features of lane lines. The article first starts from the geometric properties of lane lines and analyzes their continuous slender linear structure. By decomposing the image coordinate system into x-axis and y-axis, the probabilistic correlation model of lane lines in the row direction and column direction is established respectively. The model shows that whether a pixel belongs to a lane line or not is closely related to the pixel states of its neighboring rows or columns. Then the polymorphic CASPP module is proposed. By extending the subregion aggregation of feature descriptors (maximum and global average pooling), the contextual information utilization is improved. And one-dimensional convolutional branching is introduced to adapt the linear structure of lane lines. Finally, the checkerboard effect is mitigated by adopting mutual mass null rate. Based on CASPP module, DeepLAB-ERFC model is constructed to balance the computational efficiency and feature extraction capability by combining void convolution and depth-separable convolution through encoding-decoding architecture. Two key improvements, ER Loss and Gradient Correction, are further proposed: the boundary loss based on Hausdorff distance (ER Loss) is introduced to strengthen the boundary learning, and the category weights are dynamically adjusted in conjunction with the ratio of labeled area to alleviate the category imbalance problem. Fully Connected Conditional Random Fields (CRFs): fully connected CRFs are introduced in the decoding stage, modeling the inter-pixel position and color relationship by Gaussian kernel function, eliminating voids in the prediction results and refining the boundaries.

II. Lane line semantic segmentation algorithm based on multi-scale deep feature fusion

II. A.Spatial structure of lane lines

For a lane line, its shape and its texture structure are roughly a continuous thin straight line or nearly straight line. In the lane line detection task, this inherent shape-structure feature of the lane line itself can be fully utilized as an a priori information. The lane line detection method in this paper is based on a semantic segmentation approach, which treats the lane line detection task in an image as a special kind of image semantic segmentation task.

Specifically, given an image to be detected, the goal is to determine whether each pixel in the image belongs to a lane pixel. In this paper, the horizontal and vertical directions of the image are taken as the x-axis and y-axis under the image coordinate system, respectively.

For the representation of lane line pixels (x, y), this paper uses a lane line structure representation similar to FastDraw. From the y-axis direction, a continuous lane line can be represented by a series of consecutive sets of pixel points $R \in \{r_1 = (x_1, y_1), r_2 = (x_2, y_2), ..., r_n = (x_n, y_n)\}$ constitute. $P(r_n)$ denotes the probability that a pixel point r_n belongs to a lane line pixel, then $P(r_n)$ can be expressed as:

$$P(r_n) = P(r_1)\prod_{i=1}^{n-1} P(r_{i+1} | r_i) \tag{1}$$

According to Equation (1), the probability that a pixel point in the i th row belongs to a lane line pixel is associated with the probability distribution of pixel points in the previous $i-1$ rows. Similarly, a continuous lane line viewed from the x-axis direction can consist of a series of consecutive sets of pixel points $C \in \{c_1 = (x_1, y_1), c_2 = (x_2, y_2), \ldots, c_n = (x_n, y_n)\}$ constitutes. $P(c_n)$ denotes the probability that pixel point r belongs to a lane line pixel, then $P(c_n)$ can be expressed as:

$$P(c_n) = P(c_1)\prod_{j=1}^{n-1} P(c_{j+1} | c_j) \tag{2}$$

According to Eq. (2), the probability that a pixel point in the j th column belongs to a lane line pixel is correlated with the probability distribution of the pixel points in the first $j-1$ th column.

Based on the above analysis of the lane line structural feature representation, the spatial structural features of the lane lines in the image can be roughly analyzed here. For the i th row pixel point, whether it belongs to the lane line pixel point or not is related to the state of its previous $i-1$ th row pixel point. Similarly, for the j th column pixel, whether it belongs to the lane line pixel point is related to the state of its previous $j-1$ th column pixel point. Therefore, this spatial structural feature of lane lines can be fully utilized as a kind of spatial information in the neural network structure module.

II. B.CASPP, an improved polymorphic module based on ASPP

Based on the modeling and analysis of the spatial structure of lane lines, this paper finds that the traditional ASPP module has significant defects in capturing the elongated morphology of lane lines. For this reason, this section proposes a polymorphic CASPP module, which incorporates the spatial a priori information of lane lines into the multi-scale feature extraction process by improving the feature aggregation method with the introduction of one-dimensional convolutional branches.

Deep convolutional neural networks generally obtain a larger receptive field by downsampling operation, but at the same time reduce the resolution of the feature maps, which leads to the loss of detailed information in the image.The Deeplab series uses dilated convolution to obtain a larger receptive field while maintaining the resolution of the image, and further proposes the ASPP module to incorporate the multiscale information.The structure of the ASPP is shown in Fig. 1.

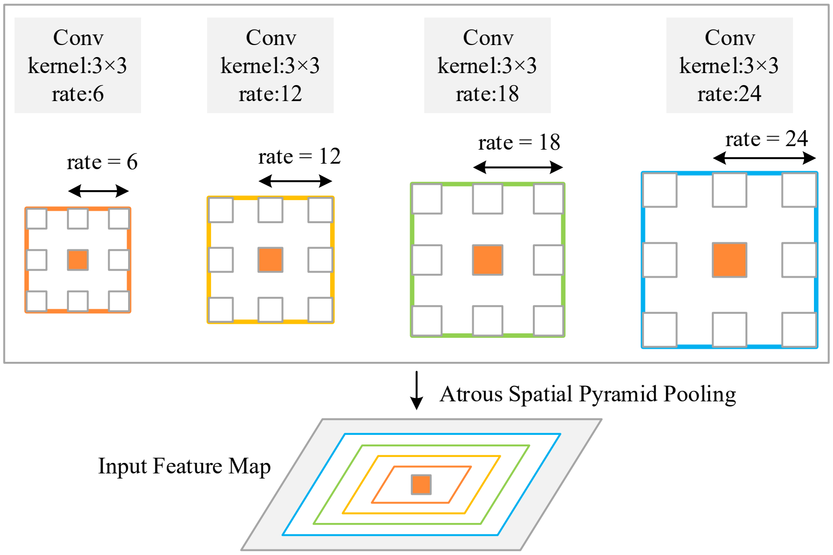

Figure 1: ASPP structure

ASPP operates on a given input with null convolution at different sampling rates, i.e., capturing the contextual information of the image at different scales. Finally fusing this multi-scale information generates the final result. The module constructs convolution kernels with different receptive fields by different null rates, which are used to acquire multi-scale object information.

The aim of ASPP is to extract multi-scale objects through parallel null convolution, which has been proved to be effective by the Deeplab series of articles, but there are still some problems in lane line detection:

(1) While it is true that convolution with different null rates can increase the sensing field, it is also true that only a fraction of the pixels can be sensed, and the sampling is not dense. Specifically, for the ASPP structure shown in Fig. 1, assuming that the input ASPP feature map X, with a size of H × W × C, can be regarded as H × W C-dimensional descriptors, then the feature descriptors used in the ASPP structure are only 3 × 3 × 3-2 = 25. For feature map X, the total number of descriptors is much larger than this. Assuming that the size of feature map X is 64×64, the total number of descriptors is 4096, from which we can calculate the proportion of descriptors used by the ASPP structure is 25/4096≈0.0061, which means that the utilization rate of descriptors is only 0.61%.

(2) The receptive field is a circular Gaussian kernel that resembles outward diffusion, and its shape does not match the rectangular shape of the image or the linear shape of the lane lines, making it less efficient.

For Problem 1, this paper will draw on the paper's solution. As shown in Fig. 2 expanding the description sub-region. Unlike ASPP which only considers 25 feature descriptors, this paper will consider subregions of the 25 original descriptors, with the size of the subregion set to k × k. An aggregation operation is performed on each subregion in the feature map X, and each subregion is aggregated into a new descriptor. This still results in 25 feature descriptors, but more contextual information is obtained for each feature descriptor. There are various ways of aggregation, and in this paper, we chose to conflate the maximum value and the global mean. In addition, in order to avoid the checkerboard effect caused by the null convolution, the null rate is set to mutually prime numbers such as 3, 5, 9, and 17.

Setting the null rate to 3 and 5 is used to capture smaller target features, and 9 and 17 are used to learn larger target features.

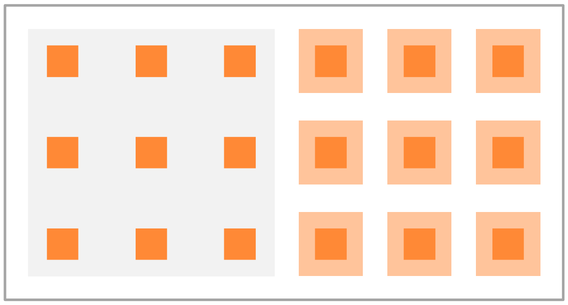

Figure 2: Expanded descriptor region

For Problem 2, considering the elongated shape of lane lines, this paper will add branching structure to the improved ASPP module. The branching structure consists of one-dimensional convolutional operations, and the corresponding convolutional kernel sizes are 3 × 1 and 1 × 3. The purpose is to allow the network to learn the lane line information better.

Based on the above ideas, this section introduces the ASPP module and improves it into a polymorphic CASPP module. After getting the feature maps at different scales it is also necessary to perform fusion processing on these feature maps. To this end, the following operations are carried out in this section: first, the collapsed feature maps are subjected to global average pooling operation to model the feature maps at different scales, and then the weights of the feature maps at different scales are obtained by passing them through 1 × 1 convolutional layers, layer normalization, activation function, and 1 × 1 convolutional layers, and then the corresponding elements are summed up by using the broadcasting mechanism, and the output is the final feature map of the CASPP module map.

II. C.DeepLAB-ERFC

The design of CASPP module provides a foundation for multi-scale fusion of lane line features. In order to further improve the boundary learning ability and prediction accuracy of the model, this section proposes the DeepLAB-ERFC model, which combines a novel loss function and post-processing strategy to construct a complete semantic segmentation framework for lane lines.

II. C. 1) Model structure

The DeepLab-ERFC model structure is shown in Fig. 3. Inspired by the Fusion Lane model, Deep Lab-ERFC uses a simple and effective encoding-decoding structure: in the encoding phase, it uses the cavity convolution instead of the traditional convolution in ResNet-101, which enhances the sense field of convolutional computation. Then the ASPP module based on Depth wise Separable Convolution is used to perform different number of convolution operations on the same feature map, which on the one hand enhances the ability of extracting semantic features at different levels, and on the other hand reduces the complexity of convolutional computation; in the decoding stage, both low-level features containing target details and high-level features containing deep semantic information are used to enhance the model on the original size. In the decoding stage, both low-level features containing target details and high-level features containing deeper semantic information are used to enhance the details of the model's prediction results at the original size, and finally reduced to the same size as the input image by up-sampling. During model training, both CELoss and ERLoss are used to weight the final training loss. The prediction accuracy of the model is further improved by calculating Fully-Connected CRFs before the final prediction results are output.

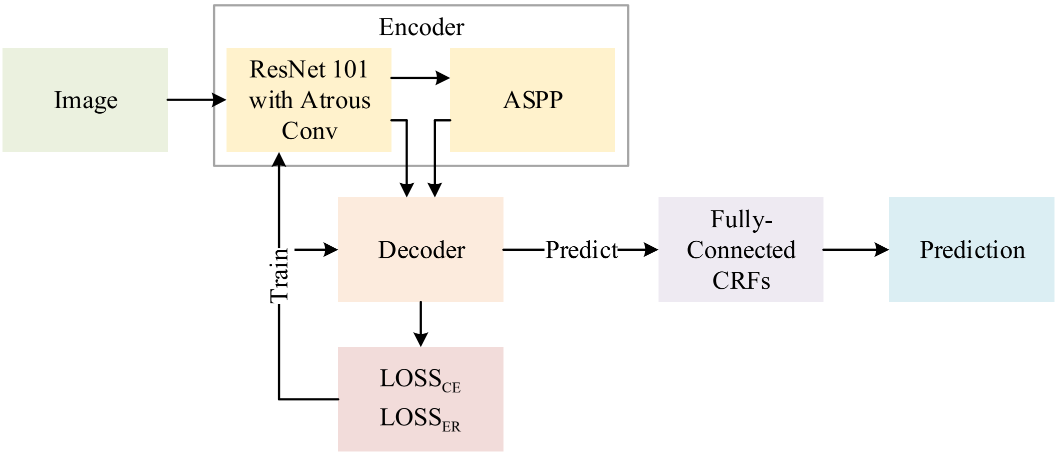

Figure 3: Model Structure of DeepLab-ERFC

II. C. 2) ER Loss and Gradient Correction

CELoss is a loss function commonly used to train semantic segmentation models, which can efficiently estimate the degree of similarity between two sets through the matrix operation of cross entropy, but given the geometric characteristics of thin and narrow lane lines, the model should be trained with an appropriate tendency to learn predictions for boundaries as well as harder-to-train categories. For this reason, this subsection proposes the use of ERLoss and a gradient correction method based on labeled area.

(1) ERLoss

ERLoss is a boundary loss function for estimating the Hausdorff distance (HD), which has the advantage of being less computationally intensive. Firstly, the prediction segmentation matrix is defined as p and the true segmentation matrix is defined as g, and their value domains are both [0, 1], where the part of pixels with value 0 is the background and the part of pixels with value 1 is the detection target. Therefore, the part of the sum matrix of the two segmentation matrices with the value of 1 is the non-overlapping part, defined as $p\nabla g$. ERLoss proposes to perform an erosion operation on the non-overlapping part of the predicted and true labels, and to use the radius of the structural element E that can completely erode out the part, r, as an approximate estimation of the HD. The corrosion operation for a target O on the grid G is defined as shown in equation (3).

$$ER(O,r)=\{p\in G\,|\,B(p,r)\in O\} \tag{3}$$

where $B(p,r)$ denotes the structural element B whose midpoint is at pixel p and has radius r. Finally, the ERLoss is defined using the form of relaxation loss function as shown in equation (4).

$$Loss_{ER}(p,g)=\frac{1}{|G|}\sum_{G}\sum_{r=1}^{R}(ER((p-g)^2,r)r^{\alpha}) \tag{4}$$

where the parameter α determines the degree of correction for larger segmentation errors and is set to 2.0 by default.

(2) Gradient correction

Inspired by the MaxSquare model that uses the pixel frequency of each category for gradient correction, this subsection proposes a gradient correction method based on the area ratio of each category in the dataset annotation. First, the average labeled area A_c with the area of each category in each image as the basic statistic is calculated based on the coordinates of the boundary points of each category provided in the dataset annotation file, and then the mapping ratio of each category is calculated based on the largest area $A_{\max}$ and the smallest area $A_{\min}$, and finally, based on the ratio of the The value domain of the weights of each category during training is limited to $[1,W_{\max}]$. The final formula for the corresponding weights of each category is shown in equation (5).

$$w_c=W_{\max}-\frac{W_{\max}-1}{A_{\max}-A_{\min}}(A_c-A_{\min}) \tag{5}$$

In the actual training, this subsection also uses the commonly used CELoss and applies the w_c calculated in Eq. (5) to this loss function, and the final complete loss function is defined as shown in Eq. (6).

$$Loss=\frac{1}{|C|}\sum_{c=1}^{C}(w_c(1-w_{ER})\times Loss_{CE}(p,g))+w_{ER}\times Loss_{ER}(p,g) \tag{6}$$

where w_{ER} denotes the weight of ERLoss in the calculation of the loss function. Considering that in the process of learning the boundaries by the reinforcement model, the boundaries of each category are mutually influential, the gradient correction of each category is not applied to the ERLoss.

II. C. 3) Fully-Connected CRFs

In modern DCNNs architectures, the commonly used Conditional Random Fields (CRFs) are mainly classified into two categories: short-range CRFs, which are used for smoothing the target boundary; and local-range CRFs, which can obtain more detailed information about the boundary, but still lose some fine structures. This subsection uses a more effective method to recover the details of the target boundary, Fully-ConnectedCRFs, which defines an energy function as shown in Equation (7).

$$E(x)=\sum_{i}(\varphi_i(x_i)+\sum_{j,j\neq i}\varphi_{ij}(x_i,x_j)) \tag{7}$$

where x denotes all pixels in the annotation, and the connection between pixels is divided into two cases: one is the self-connection of pixels, denoted as $\varphi_i(x_i)=-\log P(x_i)$, where $P(x_i)$ is the classification probability of pixel i computed by DCNN; The other is the connection between different pixels, denoted by $\varphi_{ij}(x_i,x_j)=\sum_{c=1}^{K}(w_c\times k_c(f_i,f_j))$, where K denotes the number of categories, w_c denotes the weights set on the

Gaussian kernel function, and f_i denotes the features extracted at pixel i. In the update strategy proposed in the article, CRFs processing is used in the prediction results after performing Softmax and only one category is computed at a time, so $P(x_i)=1$ and $K=2$. The Gaussian kernel function k_c integrates the positional relationship as well as the color intensity and is defined as shown in equation (8).

$$w_1 exp(-\frac{\| p_i - p_j \|^2}{2\mu_\alpha^2} - \frac{\| l_i - l_j \|^2}{2\mu_\beta^2}) + w_2 exp(\frac{\| p_i - p_j \|^2}{2\mu_\gamma^2}) \tag{8}$$

where p denotes the pixel position, I denotes the color intensity, and the domain of the Gaussian kernel is regulated by the parameters μ_α, μ_β and μ_γ, respectively. From the factors affecting the two Gaussian kernels in Eq. the first kernel considers inter-pixel position and color intensity, while the second kernel only considers pixel position, i.e., position information plays a major role in k_c.

III. Experimental validation of lane line semantic segmentation algorithm and analysis of multimorphic feature fusion

The multimorphic CASPP module and DeepLab-ERFC model proposed in Chapter 2 provide a theoretical framework for semantic segmentation of lane lines by fusing multiscale features and boundary optimization strategies. To verify its effectiveness, Chapter 3 launches experiments based on three types of datasets, TuSimple, VPG and tvtLANE, to systematically evaluate the practical application value of the algorithms, from the basic performance comparison, the robustness analysis of complex scenarios to the module ablation study.

III. A. Experimental setup

III. A. 1) Data sets

The dataset used in this thesis research work is constructed based on the TuSimple lane dataset.The TuSimple lane dataset, whose main collection area is on a foreign highway, is filmed in an angle direction that is close to the direction of the car's travel, and consists of 4,172 video clips for the training set and 3,226 video clips for the test set. Each video clip contains 20 consecutive frames collected within one second. For each video clip the last frame, i.e., the 20th image carries an annotation. The lane lines are labeled with points, and each line is actually a collection of coordinates of a sequence of points rather than a collection of regions.

The VPG dataset has a total of 22674 images with a resolution of 1288×728, of which 15872 are in the training set and 6802 are in the test set. In order to speed up the training process, the experiments in this section reduce the image resolution to 640×480 and extract 1500 images from the original dataset as the dataset used for the experiments, and according to the division ratio of 7:3, we get the specific number of training set and test set divided into 1050 and 450. The six categories of white solid lines, white dashed lines, double yellow lines, yellow solid lines, stop lines, and crosswalks, which appear most frequently in the traffic scene, are selected as segmentation targets for the experiment.

The tvtLANE dataset: the TuSimple lane dataset is newly added to the TuSimple lane dataset, which consists of 1308 sequences of rural driving scenarios collected in China, constructed from 10 challenging driving situations, for robustness evaluation.

III. A. 2) Experimental parameterization

In the experiments, the lane detection images were sampled with a resolution of 256 × 128.The lane line detection experiments were performed on a processor AMD Ryzen 5 3550H 2.10 GHz computer, developed using Python 3.5. The optimizer was selected ADAM, the initial learning rate was set to 5e-4, the momentum parameter 0.9, and the data precision type was FP32/FP16.

III. A. 3) epoch and Batch_Size settings

Suitable epoch and Batch_Size settings are the basis of network training. epoch needs to be decided according to the training sample size, generally speaking epoch obtained too small will lead to underfitting, too large will lead to overfitting. For the training data, the epochs reach basic stability at 100 iterations, and the training loss is shown in Figure 4.

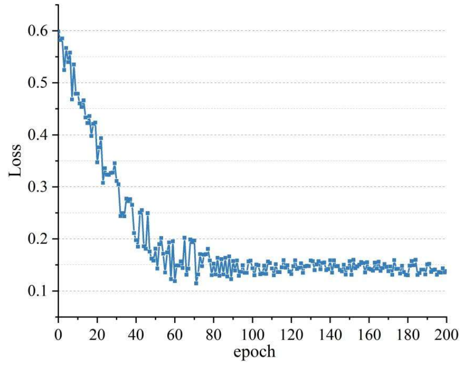

Figure 4: Loss training

Batch_Size has an impact in terms of model convergence speed, generalization ability, and stochastic gradient noise. Because the batch data is too small, resulting in insufficient input data for the network, if the parameters of the network are adjusted only for a few features, it will lead to a slower convergence speed of the model. A large amount of batch data not only generates gradient noise, but also occupies a large amount of computational space, thus affecting the learning effect of the model. When epoch is set to 100, the model training process performs well when Batch_Size is set to 16 according to the amount of training data.

III. A. 4) Performance evaluation indicators

In this paper, we adopt the average intersection and merger ratio mIoU, which is the most commonly used in semantic segmentation, as a metric to judge the accuracy of segmentation, and the number of frames per second transmitted as a metric to measure the speed of segmentation. The formula for mIoU can be expressed as:

$$mIoU=\frac{1}{K}\sum_{i=0}^{K}\frac{T_{TP}}{F_{FN}+F_{FP}-T_{TP}} \tag{9}$$

where, T_{TP} denotes a true example, F_{FP} denotes a false positive example, F_{FN} denotes a false negative example, and K denotes the number of categories.
The calculation of FPS starts from the uploading of the image to the GPU and the calculation formula can be expressed as:

$$FPS=\frac{n}{\sum_{i=1}^{n}t_i} \tag{10}$$

where, n denotes the number of predicted image sheets and ti denotes the time used to predict the ith sheet.

FLOPS: Floating point operations per second, which is used to measure the performance of the computing device, is used to measure the computational complexity of the algorithm or model.

To better evaluate the performance of the proposed method, the model is quantitatively evaluated using Accuracy, Precision, Recall and F1-measure.

III. B. Comparative Experiments

On the basis of completing the experimental parameter settings and dataset preprocessing, this section verifies the comprehensive advantages of DeepLab-ERFC in terms of accuracy, speed and adaptability to complex scenes by comparing the performance of mainstream models, such as DANet and PSPNet, on the Tusimple, VPG and tvtLANE datasets.

III. B. 1) Comparison experiments on the Tusimple dataset

In order to illustrate the accuracy and timeliness of the lane line segmentation network designed in this paper, compared with other mainstream semantic segmentation networks on the Tusimple dataset, this subsection conducts comparative experiments with the same experimental parameters for the current mainstream segmentation networks replicated with the specific models of DANet, PSPNet, and DeeplabV3plus. The quantitative comparison results of the experiments on the Tusimple the quantitative comparison results of the experiments on the dataset are shown in Fig. 5 and Table 1, respectively.

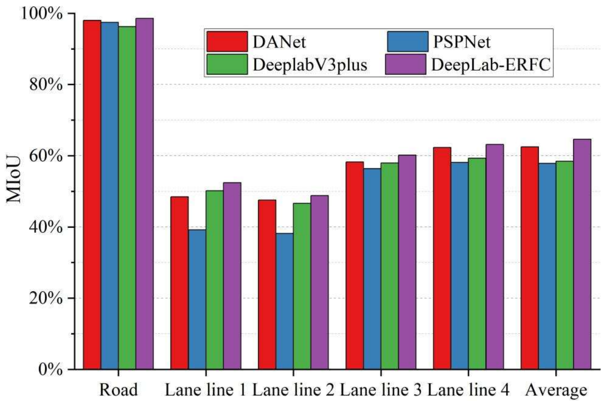

Figure 5: The MIoU performance of each model on the Tusimple dataset

Table 1: Comparative experimental results on Tusimple

	DANet	PSPNet	DeeplabV3plus	DeepLab-ERFC
FLPOs(G)	199.67	176.51	175.32	206.11
FPS(frame/s)	27.08	33.52	26.78	89.34
#params(M)	50.88	48.55	45.63	55.84

On the Tusimple dataset, the DeepLab-ERFC model proposed in this paper shows significant advantages in all the metrics. In terms of segmentation accuracy, its average mIoU reaches 64.62%, which is 2.12, 6.75 and 6.15 percentage points higher than the comparison models DANet's 62.50%, PSPNet's 57.87% and DeeplabV3plus's 58.47%, respectively. Specifically for different lane line categories, the model's improvement in detecting secondary lane lines, such as lane line 1 and lane line 2, is particularly obvious, e.g., the mIoU of lane line 1 is improved from 48.48% to 52.39% in DANet, indicating that the model enhances the feature extraction capability of the elongated structure through the polymorphic CASPP module and the one-dimensional convolutional branching. In terms of real-time performance, the FPS of DeepLab-ERFC is as high as 89.34 fps, which is much higher than that of other models, such as PSPNet's 33.52 fps, thanks to the efficient computational design of deeply separable convolution in the encoding-decoding architecture. Although the model's FLPOs of 206.11G and parametric count of 55.84M are slightly higher than the comparison models, it excels in the balance of accuracy and speed, verifying the effectiveness of the multi-scale feature fusion and post-processing strategy.

III. B. 2) Comparison experiments on VPG dataset

In the experiments in the previous section, the types of lane lines were not differentiated, and in order to further validate the superiority of the semantic segmentation network of lane lines designed in this chapter and to consider the semantic guidance of different types of lane lines for the vehicle assisted driving system to determine the drivable area, this section uses the VPG dataset to conduct further tests. Figure 6 shows the results of the comparison experiments with the VPG dataset.

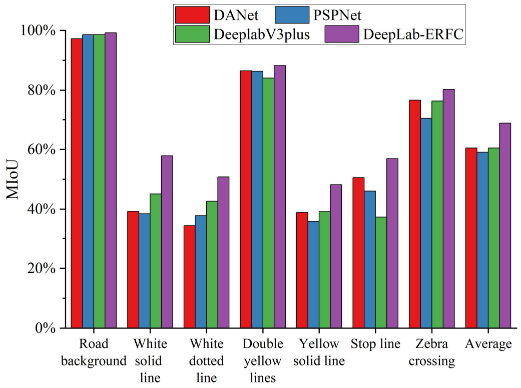

Figure 6: Comparative experimental results on VPG

In more complex VPG datasets, DeepLab-ERFC's generalization ability is further highlighted. Its average mIoU is 68.79%, an improvement of more than 8 percentage points from 60.50% in DANet and 60.48% in DeeplabV3plus. In the subdivided lane line category, the mIoU of white solid line and white dashed line reaches 57.86% and 50.81%, respectively, which is an improvement of 18.61 and 16.49 percentage points compared with DANet, indicating that the reinforcement of boundary learning by ER Loss significantly improves the detection of fine line-like targets. In addition, the model's detection accuracy of 80.25% for crosswalk is also better than other methods, which verifies the adaptability of multimorphic feature fusion to complex textures. Notably, the detection accuracy of double yellow lines (88.26%) is close to that of the background category (99.21%), indicating that the model is more capable of modeling regular linear structures.DeepLab-ERFC significantly improves the robustness of semantic segmentation of multi-category lane lines by fusing spatial prior information with dynamic gradient correction.

III. B. 3) Comparison experiments on the tvtLANE dataset

In order to better validate the superiority of the DeepLab-ERFC lane line semantic segmentation algorithm proposed in this paper, model performance comparisons are investigated under a number of challenging driving scenarios. These challenging scenarios cover a wide range of situations containing severe vehicle occlusion, poor lighting conditions (e.g., shadows, dimness), tunnel situations, and dirt road conditions. Even in some extremely difficult scenarios, e.g., where the entire lane is completely obscured by cars, other objects, or shadows, and where the lane deviates from the structural joints of the roadway, which are difficult for a human being to recognize, the proposed model is still able to accurately recognize them.The 10 challenging driving situations are (1): severe vehicular obstruction (e.g., large trucks, buses that completely block the lane lines); (2): alternating shadow and bright lighting conditions (e.g., alternating shade and direct sunlight covering the lane); (3): dim or low-light conditions (e.g., nighttime, dawn, or dusk); (4): dramatic lighting changes in tunnels (sudden changes in light and darkness when entering and exiting tunnels); (5): dirt or unpaved roads (lack of a clear lane line or ambiguous roadway boundaries); (6): lanes that are completely obstructed (completely covered by other vehicles, obstacles, or snow) ; (7): lanes deviating from the structural joints of the road (e.g., ambiguous demarcation between the edge of the road and the grass or shoulder); (8): blurred lane lines due to rain or snow (rain washout, snow cover, or reflections on slippery surfaces); (9): interference from glare reflections (e.g., reflections on water surfaces, glass, or metal surfaces affecting the identification of the lane lines); and (10): road construction or temporary rerouting (confusing lane lines, temporary markings overlapping with original markings overlap).

Table 2 shows the model performance comparison of the 10 scenarios that will be challenging in the tvtLANE dataset.

Table 2: Performance in 10 challenging scenarios(%)

	DANet		PSPNet		DeeplabV3plus		DeepLab-ERFC	
	Precision	F1	Precision	F1	Precision	F1	Precision	F1
Scene 1	49.85	46.14	45.61	42.13	51.96	50.08	53.46	50.72
Scene 2	52.49	50.34	58.61	56.51	59.87	57.35	71.38	68.59
Scene 3	65.94	63.43	55.12	52.81	67.59	65.62	71.32	68.55
Scene 4	81.85	78.19	74.03	72.49	67.85	64.97	87.14	85.43
Scene 5	45.98	43.34	62.96	60.97	56.64	55.10	64.37	61.04
Scene 6	52.38	50.13	69.25	65.60	68.59	66.31	72.35	70.33
Scene 7	49.11	45.58	39.43	35.72	38.73	36.50	53.94	52.29
Scene 8	57.42	55.88	52.45	50.52	56.49	53.21	62.67	61.10
Scene 9	56.41	54.38	73.32	70.24	75.66	73.83	76.94	75.36
Scene 10	55.71	54.05	75.55	72.04	70.92	68.48	77.32	74.70

The DeepLab-ERFC model shows significant advantages under 10 extreme driving scenarios in the tvtLANE dataset. Overall, its average accuracy Precision and F1 scores are higher than those of the comparison models in all scenarios, verifying the robustness of the algorithm in complex environments.

The model performs especially well in scenes with drastic changes in illumination. For example, the Precision of Scene 4 (sudden change of light and darkness in the tunnel) reaches 87.14%, which is 5.29, 13.11, and 19.29 percentage points higher than the 81.85% of DANet, the 74.03% of PSPNet, and the 67.85% of DeeplabV3plus, respectively, indicating that the polymorphic CASPP module is effective in capturing sudden changes of illumination under the lane line features. In Scene 2 (alternating shadows and brights), the Precision of DeepLab-ERFC is 71.38%, which is improved by more than 10 percentage points compared with other models, indicating that the enhancement of boundary learning by ER Loss significantly improves the model's ability to adapt to the interference of alternating brights and darks.

In the complex road condition and occlusion scenarios, the Precision of DeepLab-ERFC is 53.46% and 72.35% in Scene 1 (severe vehicle occlusion) and Scene 6 (lane completely occluded), respectively, which are 1.5 and 3.76 percentage points higher than the suboptimal models, thanks to the fusion of global contextual information by the encoding-decoding architecture. In addition, in Scenario 7 (lane deviation from roadway joint), both model Precision (53.94%) and F1 score (52.29%) are significantly ahead of each other, indicating that the fully-connected CRFs effectively refine the prediction results with fuzzy boundaries.

For Scenario 5 (Dirt or unpaved road), the Precision of DeepLab-ERFC is, 64.37% although it is only 1.41 percentage points higher than that of PSPNet's 62.96%, but the F1 score (61.04%) still maintains the advantage, indicating that the dynamic gradient correction alleviates the problem of category imbalance. In Scenario 9 (glare reflective interference) and Scenario 10 (road construction rerouting), the model Precision reaches 76.94% and 77.32%, respectively, which is improved by 1.28 and 1.77 percentage points compared with the suboptimal model, verifying the adaptability of multi-scale feature fusion to temporary marking and reflective interference.

DeepLab-ERFC maintains the lead in all 10 challenging scenarios, with the average Precision and F1 scores improved by 3.8~21.3 percentage points compared with the comparison model, and the advantage is especially significant in scenarios with dynamic occlusion, sudden lighting changes and complex boundaries. This is attributed to the contextual enhancement of the polymorphic CASPP module, the boundary optimization of ER Loss, and the post-processing refinement of fully-connected CRFs, which verifies the practicability and robustness of the algorithm in real complex driving environments.

III. C. Ablation experiments

Comparison experiments show that DeepLab-ERFC outperforms in multiple datasets, but the specific attribution of its performance improvement still needs to be further analyzed. To this end, this section decouples the contributions of the polymorphic CASPP module, ERLoss and CRFs post-processing layer by layer through ablation experiments to reveal the core drivers of the model improvement.

To validate the effectiveness of each improvement module in the DeepLab-ERFC model, the following groups of ablation experiments are designed on the Tusimple dataset, with controlled variables to analyze the effects of the polymorphic CASPP module, ERLoss with gradient correction, and fully connected CRFs on the model performance. The experiments are divided into 5 parts, (1) Baseline: DeepLabV3 + original architecture (standard ASPP module + cross-entropy loss); (2) Baseline + polymorphic CASPP module (3) Baseline + CASPP + ERLoss; (4) Baseline + CASPP + ERLoss + labeled area-based gradient correction (5) Baseline+CASPP+ERLoss+ gradient correction + post-processing of fully connected CRFs. Each evaluation index of the ablation experiment is shown in Table 3.

Table 3: Evaluation indicators of the ablation experiment

Group	Accuracy/%	Precision/%	F1/%	MIoU/%
Baseline	68.62	71.15	68.58	49.22
Baseline+CASPP	73.57	77.68	72.49	54.97
Baseline+CASPP + ERLoss	84.29	85.42	82.85	59.74
Baseline+CASPP + ERLoss+Gradient correction	90.07	92.41	90.32	61.60
Baseline+CASPP + ERLoss+Gradient correction+Fully-ConnectedCRFs	97.75	96.45	95.`3	64.62

Table 3 demonstrates the contribution of each improvement module in the DeepLab-ERFC model to the performance improvement. The accuracy, precision, F1 score, and average intersection and merger ratio mIoU of the baseline model are 68.62%, 71.15%, 68.58%, and 49.22%, respectively. After adding the polymorphic CASPP module, the four metrics are significantly improved to 73.57%, 77.68%, 72.49% and 54.97%, indicating that the polymorphic CASPP effectively enhances the extraction of elongated lane line features through the reciprocal nulling rate and the one-dimensional convolutional branching. After further introduction of ERLoss, the model accuracy is 85.42% vs. 59.74% of mIoU is substantially improved, which verifies the enhancement effect of boundary loss based on Hausdorff distance on lane line edge learning. On this basis, combined with the gradient correction based on labeled area, the model accuracy is improved to 90.07% versus 90.32% for the F1 score, indicating that the dynamic gradient correction alleviates the category imbalance problem. Finally, by adding the post-processing of fully-connected CRFs, the model metrics reached the optimum, and the mIoU was improved to 64.62% with an accuracy of 97.75%, indicating that the CRFs significantly optimized the boundary details and global consistency of the prediction results by modeling the inter-pixel location and color relationships. The ablation experimental data show that the layer-by-layer superposition of the polymorphic CASPP module, ERLoss, gradient correction and CRFs post-processing systematically improves the model's adaptability to the semantic segmentation task of lane lines.

IV. Conclusion

In this paper, a semantic segmentation algorithm based on multi-scale deep feature fusion is proposed for the lane line detection task in autonomous driving, which significantly improves the accuracy, real-time performance and robustness of lane line detection by systematically improving the network architecture and optimization strategy.

On TuSimple, VPG and tvtLANE datasets, the average mIoU of the proposed DeepLab-ERFC model reaches 64.62%, 68.79% and 64.62%, respectively, which is an improvement of 2.12-21.3 percentage points compared with the mainstream methods, such as DANet, PSPNet, etc., and the inference speed reaches 89.34 FPS, with the real-time performance improved by 2.6 times more. Especially in extreme scenarios (e.g., sudden change of tunnel light, vehicle occlusion, rain and snow interference), the average accuracy of the model is improved by 3.8~21.3 percentage points compared with the suboptimal methods, demonstrating strong environmental adaptability.

The ablation experiments show that the CASPP module improves the model's mIoU by 5.75% on the TuSimple dataset, and the introduction of full connectivity condition random field CRFs post-processing effectively eliminates voids in the prediction results and refines the boundaries by modeling the inter-pixel position and color relationships. The final model achieves an mIoU of 64.62% on the TuSimple dataset, which is a 15.4 percentage point improvement over the baseline model.

Funding

This work was supported by the Science and Technology Research Program of the Chongqing Education Commission (No. KJQN202403106).

References

[1] Cao, J., Song, C., Song, S., Xiao, F., & Peng, S. (2019). Lane detection algorithm for intelligent vehicles in complex road conditions and dynamic environments. Sensors, 19(14), 3166.

[2] Farag, W., & Saleh, Z. (2018, November). Road lane-lines detection in real-time for advanced driving assistance systems. In 2018 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT) (pp. 1-8). IEEE.

[3] Bilal, H., Yin, B., Khan, J., Wang, L., Zhang, J., & Kumar, A. (2019, July). Real-time lane detection and tracking for advanced driver assistance systems. In 2019 Chinese control conference (CCC) (pp. 6772-6777). IEEE.

[4] Möller, D. P., Haas, R. E., Möller, D. P., & Haas, R. E. (2019). Advanced driver assistance systems and autonomous driving. Guide to Automotive Connectivity and Cybersecurity: Trends, Technologies, Innovations and Applications, 513-580.

[5] Chen, J., Ruan, Y., & Chen, Q. (2018). A precise information extraction algorithm for lane lines. China Communications, 15(10), 210-219.

[6] Nieto, M., Vélez, G., Otaegui, O., Gaines, S., & Van Cutsem, G. (2016). Optimising computer vision based ADAS: vehicle detection case study. IET Intelligent Transport Systems, 10(3), 157-164.

[7] Haque, M. R., Islam, M. M., Alam, K. S., Iqbal, H., & Shaik, M. E. (2019). A computer vision based lane detection approach. International Journal of Image, Graphics and Signal Processing, 10(3), 27.

[8] Kassem, N. S., Masoud, A. M., & Aly, A. M. B. (2021, October). Driver-in-the-Loop for computer-vision based ADAS testing. In 2021 3rd Novel Intelligent and Leading Emerging Sciences Conference (NILES) (pp. 252-256). IEEE.

[9] BV, S. S., & Karthikeyan, A. (2018, March). Computer vision based advanced driver assistance system algorithms with optimization techniques-a review. In 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 821-829). IEEE.

[10] Waykole, S., Shiwakoti, N., & Stasinopoulos, P. (2021). Review on lane detection and tracking algorithms of advanced driver assistance system. Sustainability, 13(20), 11417.

[11] Luo, S., Zhang, X., Hu, J., & Xu, J. (2020). Multiple lane detection via combining complementary structural constraints. IEEE Transactions on Intelligent Transportation Systems, 22(12), 7597-7606.

[12] Bi, W., Cheng, D., & Kou, K. I. (2021). A Robust Lane Detection Associated with Quaternion Hardy Filter. arXiv preprint arXiv:2108.04356.

[13] Huang, Q., & Liu, J. (2021). Practical limitations of lane detection algorithm based on Hough transform in challenging scenarios. International Journal of Advanced Robotic Systems, 18(2), 17298814211008752.

[14] Wei, X., Zhang, Z., Chai, Z., & Feng, W. (2018, August). Research on lane detection and tracking algorithm based on improved hough transform. In 2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE) (pp. 275-279). IEEE.

[15] Wei, Y., & Xu, M. (2021). Detection of lane line based on Robert operator. Journal of Measurements in Engineering, 9(3), 156-166.

[16] Javeed, M. A., Ghaffar, M. A., Ashraf, M. A., Zubair, N., Metwally, A. S. M., Tag-Eldin, E. M., ... & Jiang, X. (2023). Lane line detection and object scene segmentation using otsu thresholding and the fast hough transform for intelligent vehicles in complex road conditions. Electronics, 12(5), 1079.

[17] Tian, J., Liu, S., Zhong, X., & Zeng, J. (2021). LSD-based adaptive lane detection and tracking for ADAS in structured road environment. Soft Computing, 25(7), 5709-5722.

[18] Jo, Y. H., & Lee, D. J. (2022). Real-time lane detection techniques using optimal estimator and deep learning-based lane segmentation for self-driving vehicles. Journal of Institute of Control, Robotics and Systems (in Korean), 28(4), 353-361.

[19] Yang, Q., Ma, Y., Li, L., Su, C., Gao, Y., Tao, J., ... & Jiang, R. (2023). Lightweight lane line detection based on learnable cluster segmentation with self-attention mechanism. IET Intelligent Transport Systems, 17(3), 522-533.

[20] Al Mamun, A., Em, P. P., Hossen, M. J., Jahan, B., & Tahabilder, A. (2023). A deep learning approach for lane marking detection applying encode-decode instant segmentation network. Heliyon, 9(3).

[21] Chougule, S., Koznek, N., Ismail, A., Adam, G., Narayan, V., & Schulze, M. (2018). Reliable multilane detection and classification by utilizing cnn as a regression network. In Proceedings of the European conference on computer vision (ECCV) workshops (pp. 0-0).

[22] Yousri, R., Elattar, M. A., & Darweesh, M. S. (2021). A deep learning-based benchmarking framework for lane segmentation in the complex and dynamic road scenes. IEEE Access, 9, 117565-117580.

[23] Zou, Q., Jiang, H., Dai, Q., Yue, Y., Chen, L., & Wang, Q. (2019). Robust lane detection from continuous driving scenes using deep neural networks. IEEE transactions on vehicular technology, 69(1), 41-54.

[24] Al Mamun, A., Poh Ping, E., & Hossain, M. J. (2021). A Deep Learning Instance Segmentation Approach for Lane Marking Detection. International Journal Of Computing and Digital System.