# Semi-supervised Real-time Roadway Detection Based on ML−ELM

**Wenjuan Zhang*1, Feng Jin*2, Lingjun Meng*3**

*1Academy of Automation, Beijing Institute of Technology, Beijing 100081, P.R. China
E-mail: wjzhang056@163.com
*2Academy of Automation, Beijing Institute of Technology, Beijing 100081, P.R. China
E-mail: jinfeng226@163.com
*3Academy of Automation, Beijing Institute of Technology, Beijing 100081, P.R. China
E-mail: menglingjun@cloudcver.com

**Abstract. To meet the autopilot in non-standard highway of real-time and robustness requirement, this paper proposes a half supervision based on ML-ELM road driving regional rapid segmentation methods. This method contains image driving regional clustering and the pavement decision-making in two stages. In the clustering stage, the first to drop pavement grayscale sample build multi-scale image pyramid, decision-making based on scale factor in each layer to find the best contrast images, and then to clustering of grey value, to search connected domain in the output image of clustering, Inhibit the abnormal characteristics of the connected domain and within its domain sampling according to the rules for decision-making sub-graph. In road area of decision-making, using multilayer ELM in the process of the training as a classifier, treat decision-making sub-graph for binary classification, clustering diagram decision will result feedback to the connected domain category attributes, namely road may exercise area pixel level segmentation. Experimental results show this method is compared with the traditional FCM algorithm speed improved, compared with the semantic network FCN segmentation based on deep learning, Multi-net has similar accuracy, effectively ruled out the lane line, pavement crack and other non-standard interference pixels of interference, It has better robustness and real-time performance on curves and multi-mode pavement.**

**Keywords: Lane Segmentation, ML-ELM, Scale Decision Factor**

## 1. INTRODUCTION

The advancement of artificial intelligence technology is rapidly promoting the development of intelligent transportation. The research on unmanned driving and ADAS (automotive assisted driving system) has become a hot research topic. Among them, the driving area determination is the basic function of the ADAS system to determine the current The relative relationship between the vehicle and the lane, as well as the judgment basis for automatic parking and lane-changing decisions. There are currently two main types of vision-based driving area detection methods [1], which are based on prior knowledge of lane width and lane detection technology based on deep learning.

Based on the prior knowledge of lane width, the detected multiple lane lines are segmented. The traditional methods of lane line detection include image feature method and model matching method. The image feature method analyzes the underlying image features, such as lane color, edge, and texture, as the basis for lane line segmentation and extraction [2-3]. The model matching method mainly uses different road models, such as straight line model [4], curve model [5], B-spline model [6], etc. to achieve structured lane detection. This method has a good effect on fixed structure lanes. The recognition effect is poor, but the adaptability is poor, and the expected detection effect cannot be achieved in the case of changes in illumination, noise interference, non-standardized curves or discontinuous lane lines.

Lane detection technology based on deep learning [7-9] uses a multi-layer convolutional neural network without artificially defining features, and uses a small area around the pixel as CNN input for training and prediction. For example, FCN replaces the fully connected layer behind the traditional network with a convolutional layer, so that the network output is no longer a category but a heatmap; at the same time, in order to solve the impact of convolution and pooling on the image size, it is proposed to use upsampling to restore To the original size [10]. Such as Multi-net, Multi-net is a segmentation network proposed by the original author to participate in ISBI Challenge, which can adapt to a small training set (about 30 pictures). Both Multi-net and FCN are small segmented networks with simple structure [11]. This type of method can accept input images of any size (without a fully connected layer) and is more efficient, avoiding the problems of repeated calculation and space waste caused by the use of neighborhoods. The disadvantage is that the computational complexity is high and the need for stronger performance GPU can do the work.

It is difficult for the existing road detection algorithms to balance the real-time robustness and the increased complexity in multiple scenarios. This paper proposes a novel semi-supervised (no manual feature) real-time drivable area detection algorithm architecture based on ML-ELM, which can effectively reduce the number of manual feature designs and improve the real-time robustness of the algorithm.

The rest of this article is organized as follows. Section 2 introduces the background and ideas of the semi-supervised real-time drivable road detection algorithm based on ML-ELM. In Section 3, we create a road detection data set and use ML-ELM to do some experiments and comparisons. Section 4 summarizes the conclusions of this article.

## 2. ALGORITHM

### 2.1. Background and problem analysis

In automatic driving, the position of the image sensor is near the center line of the front of the vehicle, so that the collected road image contains a lot of prior information. There will be clear lane lines, traffic signs, left and right guardrails in the images of standard roads, long-distance straights on highways and fixed road widths, which will enable road images to manually design features based on prior knowledge to get better experiments result. However, road images in non-standard situations will be less robust. The projection of the guardrail on the pavement under the illumination shown in Fig. 1, the interference of road cracks, uneven pavement asphalt repair areas in Fig. 2, and fuzzy or irregular lane lines in Fig. 3 will reduce the generalization ability of the algorithm.

To obtain better results, more features must be added. Therefore, the design of such algorithm ideas is too cumbersome, and superimposing a large number of features will increase the amount of calculation and reduce real-time performance, and due to the performance difference of different sensors and different road conditions and different natural conditions, such prior information is not stable. Therefore, algorithms based on such features are not reliable for road detection in unknown road conditions. How to find a stable prior and reduce the feature description for special road conditions has become the problem that this article focuses on.

Traditional target recognition semantic segmentation, describing the target based on the feature descriptor, and then handing it over to the classifier for feature classification. With the development of convolutional neural networks, features are automatically extracted through convolutional networks, and automatic classification methods are gradually replacing target recognition methods based on feature self-description, and semantic segmentation methods based on convolutional neural networks when the data set scene is complete Stronger robustness and good generalization ability. However, the calculation of convolutional neural networks relies on the parallel computing power of the arithmetic unit, and even requires dedicated computing modules, such as NVIDIA's GPU, to obtain high-speed parallel computing capabilities. For vehicle-mounted processors, there is no independent GPU, and the performance of the CPU is not strong. Most of them use ARM cores. Therefore, if you want to use convolutional neural networks, you cannot guarantee real-time requirements.

In response to the problems of the above algorithms, this paper proposes a novel semi-supervised (no manual feature) real-time drivable area detection algorithm architecture based on ML-ELM, which can effectively reduce the number of manual feature designs and improve the real-time robustness of the algorithm.

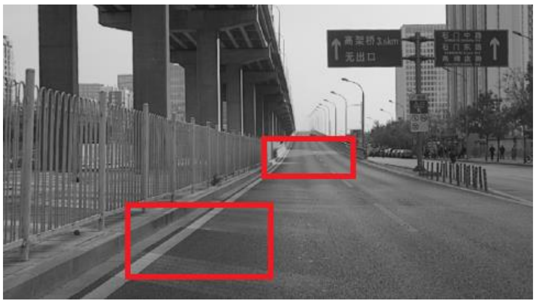

**Fig.** 1 Shadow and light interference

**Fig.2** Surface defect interference

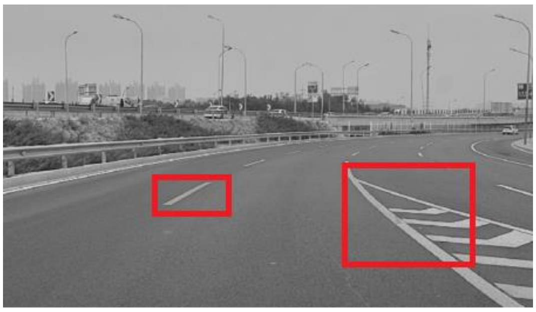

**Fig.3** Irregular lane lines

### 2.2. Algorithm ideas

Image sensor equipment in automatic driving is relatively concentrated in the center line area of the front of the vehicle, and the road surface area that can be driven occupies a large proportion in the full frame. Therefore, the amount of information contained in the image is higher in the road surface information. This is an automatic Stable prior information of the driving scene. There is a certain contrast between other elements in the image and the amount of information in the drivable area. The attenuation of the amount of information under the same amplitude can increase the contrast between the target area and the non-target area.

This paper proposes an algorithm idea based on the scale decision factor to find the best scale image to improve the contrast between the target area and the environment, reduce the complexity of the problem and reduce the amount of program calculations, and then use the method of clustering and classification to achieve weak supervision and semi-supervision Detection method of road drivable area, As shown in Fig. 4.


**Fig.4** Algorithm ideas

## 3. EXPERIMENTS

### 3.1. Image preprocessing

The data in this article comes from a vehicle-mounted grayscale image sensor, a binocular stereo vision system composed of two CMOS cameras, and the two left and right grayscale images collected at the same time are used for depth measurement. This article uniformly takes the left picture.

The preprocessing part of the algorithm in this paper suppresses uneven illumination and weakens the interference caused by the projection of light and obstacles. At the same time, the method of morphology is used to eliminate many isolated noises produced by non-uniform illumination suppression. Construct an image pyramid on the processed image for subsequent operations.

### 3.2. Optimal scale factor

The first stage of the algorithm is to find the best scale image in this vision system scene based on the scale decision factor. In order to measure the contrast between target and non-target regions in the image, this paper proposes the best scale factor for finding the best scale image. Use the images at all levels of the image pyramid to calculate the gray distribution histogram. Based on prior knowledge, the sky and the ground occupy the largest proportion in the vehicle image, so at least two relatively tall waveforms should be stable in the gray distribution histogram, and the optimal scale factor is used to describe the distance between the center gray values of different waveforms. It is then used to decide which level of image is the best contrast image.

### 3.3. Clustering strategy and sampling rules

Clustering algorithm is a typical unsupervised learning algorithm, which is mainly used to automatically group similar samples into a category. The biggest difference between clustering algorithm and classification algorithm is: clustering algorithm is an unsupervised learning algorithm, while classification algorithm is a supervised learning algorithm.

This paper chooses Kmeans clustering based on pixel gray value. K-Means algorithm is a clustering algorithm based on distance similarity. By comparing the similarity between samples, similar samples are classified into the same category. The main idea is: given K value and K initial cluster center points, divide each point (that is, data record) into the cluster represented by the nearest cluster center point, all points After the allocation is completed, the center point of the cluster is recalculated according to all the points in the cluster (take the average value), and then the steps of assigning points and updating the center point of the cluster are performed iteratively until the center point of the cluster changes greatly. Is small, or reaches the specified number of iterations.

Kmeans clustering is performed by the gray value and the cluster number k is set to 3. After several iterations of the objective function convergence, as shown in Fig. 5, the image is divided into three types of regions after the clustering experiment, and the regions cross each other There are a large number of isolated and irregularly shaped areas, see Fig. 6. The clustering of the algorithm in the gray scale image completes a process of automatic feature extraction. The experimental results show that there are a large number of situations where the sky and the ground converge together, and under non-standard roads, the walls and railings on both sides of the road also converge with the road surface. Therefore, the clustering results need to be further subdivided, and the positive and negative decisions of each connected domain under each category (positive decisions are considered to be roads).

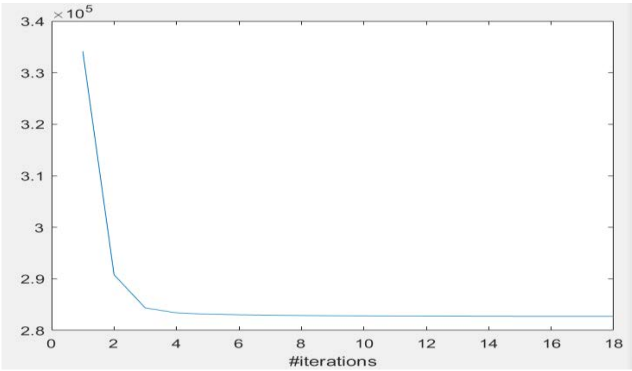


**Fig.5** Convergence of objective function

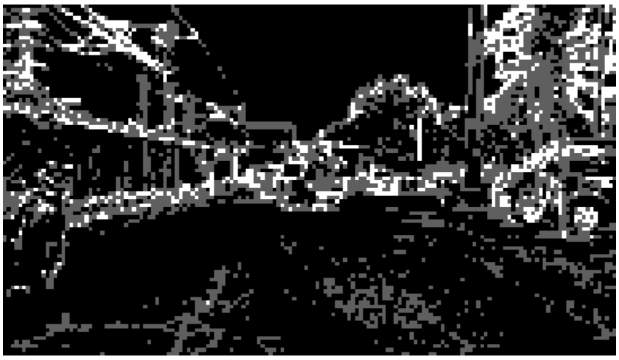

**Fig.6** Visualization after image gray-scale clustering

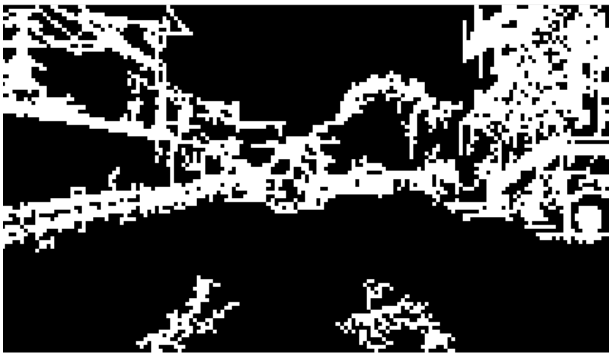

**Fig.7** Connected domain suppression result graph

### 3.4. Driven road area decision

According to the result of the previous step, there are a large number of connected domains that contain each other in the picture. In this paper, the area of connected domains is suppressed, and the isolated small area exclusion method is shown in Fig. 7. Sampling in the connected domain to obtain the sub-regions to be classified, the clustering results can be more detailed to distinguish the drivable road area. Due to the presence of a large number of interference elements, pedestrians, vehicles, guardrails, trees, etc., the clustering results will produce many discontinuous irregular small areas. And the classification area that contains each other. But the drivable area is a large continuous area, so the small connected domain formed by the misclassification caused by interference can be eliminated by sampling and reclassification.

The strategy adopted in this paper is that if the area of the connected domain is too small, the final classification result will be determined by voting for the classification results of all large areas in its neighborhood, that is, the class with the largest number of large areas of the same type nearby determines the small area. The classification result of the connected domain. Image cutting, due to the installation requirements of the ADAS system, the angle of the camera needs to be adjusted in advance, so that the road surface area in the image cannot exceed one-half of the image. Therefore, this article does not perform clustering results for more than one-half of the image. deal with.

### 3.5. Regional sampling

Sampling in connected domains, because the shape of the drivable area itself is irregular, and the connected domain shape formed after clustering is also uncontrollable, this paper divides the obtained shapes into long strips and relatively concentrated irregular shapes. As shown in Fig. 8.

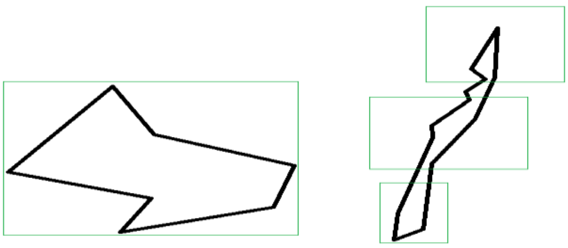

**Fig.8** Example of regional sampling rules

For these two types of shapes, this article uses two sampling methods. The strip structure is divided into three segments along the pixel axis of the strip structure. Each segment is searched for the smallest circumscribed moment, saved as a sample to be classified and finally this The result of the final classification decision of the long-bar structure is determined by the majority of the classification results in the three samples. For irregular areas with relatively concentrated pixels, directly find the minimum circumscribed moment, and save it as a sample to be classified, and the classification result is the decision result of this area classification.

### 3.6. Classifier, ML-ELM

ELM (Extra-Limited Learning Machine), from the perspective of theory and practical application, the over-limit learning machine can produce better generalization capabilities than support vector machines and their variants. The over-limit learning machine is also better than CNN in many scenarios.

This article uses ML-ELM as a variant of the classifier (multi-layer over-limit learning machine). The effect of a single-layer ELM is similar to that of an independent perceptron, but the effect on high-dimensional data like images is not obvious. Therefore, a multi-layer ELM stack is used to form a perceptual network to classify images.

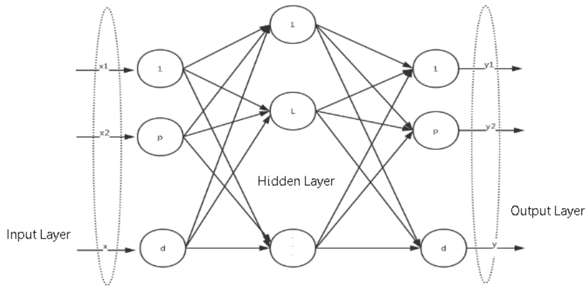


**Fig. 9** ELM structure diagr

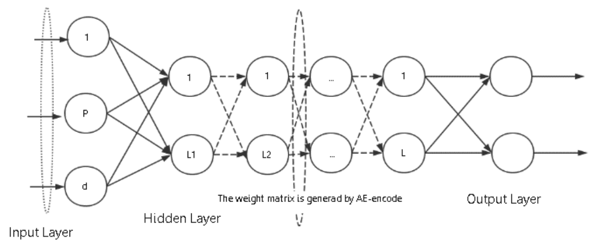


**Fig.10** ML-ELM structure diagram

ELM is a new type of fast learning algorithm. For a single hidden layer neural network, ELM can initialize input weights and biases randomly and get the corresponding output weights. For a single hidden layer neural network as shown in Fig. 9, suppose there are N arbitrary samples $(X_i, t_i)$, among them, $X_i = [x_{i1}, x_{i2}, \ldots x_{in}]^T \in R^n$, $t_i = [t_{i1}, t_{i2}, \ldots t_{im}]^T \in R^m$. For a single hidden layer neural network with L hidden layer nodes, it can be expressed as

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = o_j,\ j = 1, \cdots, N \tag{1}$$

among them, $g(x)$ is the activation function, $W_i = [w_{i1}, w_{i2}, \ldots,\ w_{in}]^T$ is the input weight, $\beta_i$ is the output weight, $b_i$ is the bias of the i-th hidden layer unit. $W_i \bullet X_i$ represents the inner product of $W_i$ and $X_i$. The goal of single hidden layer neural network learning is to minimize the output error, which can be expressed as

$$\sum_{j=1}^{N} \| o_j - t_j \| = 0 \quad (2)$$

That is, there are $\beta_i$ , $W_i$ and $b_i$ such that

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = t_j, \; j = 1, \cdots, N \quad (3)$$

Can be expressed as a matrix

$$H\beta = T \quad (4)$$

Among them, $H$ is the output of the hidden layer node, $\beta$ is the output weight, and $T$ is the expected output.

$$H(W_1, \cdots, W_L, b_1, \cdots b_L, X_1 \cdots X_L) = \begin{bmatrix} g(W_1 \cdot X_1 + b_1) & \cdots & g(W_L \cdot X_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(W_1 \cdot X_N + b_1) & \cdots & g(W_L \cdot X_N + b_L) \end{bmatrix}_{N \times L} \quad (5)$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{N \times m} \quad (6)$$

In order to be able to train a single hidden layer neural network, we hope to get $\hat{W}_i$, $\hat{b}_i$ and $\hat{\beta}_i$ such that

$$\| H(\hat{W}_i, \hat{b}_i)\hat{\beta}_i - T \| = \min_{W,b,\beta} \| H(W_i, b_i)\beta_i - T \| \quad (7)$$

Some traditional algorithms based on the gradient descent method can be used to solve such problems, but the basic gradient-based learning algorithm needs to adjust all parameters in the iterative process. In the ELM algorithm, once the input weight $W_i$ and the bias $b_i$ of the hidden layer are randomly determined, the output matrix $H$ of the hidden layer is uniquely determined. Training a single hidden layer neural network can be transformed into solving a linear system $H\beta = T$ . And the output weight $\beta$ can be determined. Among them, $H^+$ is the Moore-Penrose generalized inverse of matrix $H$ .

$$\hat{\beta} = H^+ T \quad (8)$$

Using generalized Moore inverse method for parameter iteration instead of back propagation error for parameter learning, the algorithm complexity is far lower than convolutional neural network, and it is suitable for running on ARM cores with low computing power. Over-limit learning machine is an excellent perceptron algorithm following deep learning and convolutional neural networks. Use a multi-layer over-limit learning machine to build a classifier that features autonomous learning and autonomous classification. Classify the samples to be classified in the above experiment.

### 3.7. Data set

This paper uses a vehicle-mounted camera to collect image data under various road conditions in Beijing for algorithm testing and evaluates the algorithm in this paper on KITTI Vision BenchmarkSuite [12]. A total of 1,000 self-created data sets and manual pixel-level annotations are shown in Fig. 11. The first row of data sets and the second row of experimental results. According to the figure, it can be seen that the experimental results are in good agreement with the data set. And use the pixel accuracy score and the running time of the algorithm to process a single image to detect the performance of the algorithm.

**Fig.11** Experimental result graph

### 3.8. Experimental results and analysis

By comparing the road detection algorithm based on traditional area segmentation, the algorithm in this paper has a significant improvement in pixel accuracy and processing speed of processing a single image, which fulfilled the experimental expectations. As shown in Table 1. Compared with the neural network-based semantic segmentation algorithm, the pixel accuracy is not much different from the current highest effect, and the accuracy is significantly improved compared with the classic neural network algorithm FCN.

Although compared with FCN, the algorithm of Multi-net has a significant difference in computing speed. This is because the platforms on which the two algorithms run are essentially different. The GPU is far superior to the CPU in processing image data. The algorithm in this paper is aimed at a vehicle-mounted computing platform, so the speed of the algorithm in this paper can meet real-time requirements. The segmentation of the driving area on the road has a good segmentation effect, and the detection of edges, especially for special road scenes such as curves, also has a better segmentation effect.

**Table. 1** Data comparison

| | PA | Time (s) CPU |
|---|---|---|
| **FCM** | 67.30 | 0.5038 |
| **Multi-net** | 94.88 | 0.0716 (GPU) |
| **FCN-8s** | 89.10 | 0.1756 (GPU) |
| **ML-ELM** | 92.11 | 0.2743 |

## 4. CONCLUSION

In this paper, in order to meet the real-time and robustness requirements of autonomous driving on non-standard roads, we propose a semi-supervised method for rapid segmentation of road drivable areas based on ML-ELM.

Experiments show that compared with other algorithms, the speed of this method is greatly improved, and the accuracy is also higher. At the same time, the algorithm has the following advantages: 1) Use scale decision factors to find the best scale image in the image pyramid, which improves the contrast between the target area and the background, reduces the data scale of image processing, speeds up the detection rate, and reduces The noise of non-standard jammers improves the contrast between the target and the background, which helps to improve the accuracy of target detection. 2) The method of first clustering and then sampling and classifying images is used to replace the design process of image feature description, so as to obtain advantages similar to convolutional neural networks, that is, there is no need to manually define and design the advantages of features, while maintaining high accuracy. 3) The classification network uses the ML-ELM structure and does not need to use the backpropagation algorithm for iterative optimization. It directly obtains the local extremum by calculating the generalized Moore inverse to achieve the optimization, which makes the algorithm more efficient and suitable for application in embedded systems.

In the future, semantic segmentation of drivable roads will be widely used in unmanned driving systems, and more and more deep learning will be used to solve road traffic problems. If necessary, we are willing to disclose our data and algorithms for others to study and use.

**Acknowledgements**

I would like to thank our partner for providing us with some images and data. I would like to thank my supervisor and my senior fellow apprentice for taking time to review this work.

**REFERENCES:**

[1] CHEN Han-shen, YAO Ming-hai, CHEN Zhi-hao and YANG Zhen, Efficient Method of Lane Detection Based on Multi-frame Blending and Windows Searching. Computer Science, Vol. 45, pp. 255-260, 2018

[2] JEONG P, NEDEVSCHIS. Efficient and robust classification method using combined feature vector for lane detection. Circuits and Systems for Video Technology, 2005, 15(4): 528-53 7.

[3] R. Gopalan, T. Hong, M. Shneier and R. Chellappa, "A Learning Approach Towards Detection and Tracking of Lane Markings," in IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 3, pp. 1088-1098, Sept. 2012.

[4] R. K. Satzoda, S. Sathyanarayana, T. Srikanthan and S. Sathyanarayana, "Hierarchical Additive Hough Transform for Lane Detection," in IEEE Embedded Systems Letters, vol. 2, no. 2, pp. 23-26, June 2010.

[5] CHEN Q,WANG H. real-time lane detection algorithm based on a hyperbola-air model. Intelligent Vehicles Symposium. Tokyo: IEEE. 2006: 510-515.

[6] WU B F, LIN C T, CHEN YL. Dynamic calibration and occlusion handling algorithms for lane tracking. IEEE Transactions on Industrial Electronics,2009, 56(5): 1757-1773.

[7] A. Gurghian, T. Koduri, S. V. Bailur, K. J. Carey and V. N. Murali, "DeepLanes: End-To-End Lane Position Estimation Using Deep Neural Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, 2016, pp. 38-45.

[8] LI J, MEI X, PROKHOROV D. Deep Neural Network for Structural Prediction and Lane Detection in Traffic Scene. IEEE Transactions on Neural Networks & LearningSystems, 2016, 28(3):1-14.

[9] C. Chen, A. Seff, A. Kornhauser and J. Xiao, "DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 2722-2730.

[10] M. Teichmann, M. Weber, M. Zöllner, R. Cipolla and R.Urtasun, "MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving," 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, 2018, pp.1013-1020.

[11] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3431-3440.

[12] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, “Extreme Learning Machine: Theory and Applications”, Neurocomputing, vol. 70, pp. 489-501, 2006.