IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 17, 2019, accepted September 27, 2019, date of publication October 2, 2019, date of current version October 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944993

Road Marking Segmentation Based on Siamese Attention Module and Maximum Stable External Region

WEIWEI ZHANG[1], ZEYANG MI[1], YAOCHENG ZHENG[1], QIAOMING GAO[2], AND WENJING LI[1]

[1]School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
[2]School of Mechanical and Transportation Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China

This work was supported in part by the National Natural Science Foundation of China under Grant 51805312, in part by the Shanghai Sailing Program under Grant 18YF1409400, in part by the Training and Funding Program of Shanghai College Young Teachers under Grant ZZGCD15102, in part by the Scientific Research Project of Shanghai University of Engineering Science under Grant 2016-19, in part by the Science and Technology Commission of Shanghai Municipality under Grant 19030501100, and in part by the Shanghai University of Engineering Science Innovation Fund for Graduate Students under Grant 18KY0613.

ABSTRACT Lane detection severs as one of the pivotal techniques to promote the development of local navigation and HD Map building of autonomous driving. However, lane detection remains an unresolved problem for the challenge of detection accuracy in diverse driving scenarios and computational limitation in on-board devices, let alone other road guidance markings. In this paper, we go beyond aforementioned limitations and propose a segmentation-by-detection method for road marking extraction. The architecture of this method consists of three modules: pre-processing, road marking detection and segmentation. In the pre-processing stage, image enhancement operation is used to highlight the contrast especially between road markings and road background. To reduce the computational complexity, the road region will be cropped by vanishing point detection algorithm in this module. Then, a lightweight network is dedicated designed for road marking detection. In order to enhance the network sensitivity to road markings and improve the detection accuracy, we further incorporate a Siamese attention module by integrating with the channel and spatial maps into the network. In the segmentation module, different from the method of semantic segmentation by neural network, our segmentation method is mainly based on conventional image morphological algorithms, which is less computational and also can achieve pixel-level accuracy. Additionally, the sliding search box and maximum stable external region (MSER) algorithms are utilized to compensate for missed detection and position error of bounding boxes. In the experiments, our proposed method delivers outstanding performances on cross datasets and achieves the real-time speed on the embedded devices.

INDEX TERMS Lane detection, segmentation-by-detection, lightweight network, Siamese attention module.

I. INTRODUCTION

In the development of autonomous driving, the most essential part is to make the car have the ability to perceive the surrounding environment. The road is inherent with plentiful local navigation guides and warnings (i.e. pedestrians, vehicles, traffic signs, road signs, etc.), each of them can effectively benefit autonomous vehicles. Various road markings are spread over common structured roads, accurate perception of these markings is conducive to vehicle positioning and can be a guidance of intelligent driving. Therefore, the research on the road marking detection has been one of the most popular research directions for advanced assisted driving. However, most of the current research is only focused on the detection and modeling of lanes, which can only provide guidance for the basic assist driving functions like lane keeping and lane departure warning on highway. It is not capable for high-level assisted driving to understand more comprehensive environment, such as recognizing

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Shorif Uddin.

the intersection turn signs and taking appropriate steering operations on urban roads.

There are two main challenges in the detection of road markings. Due to the limitations of low-cost and hardware deployment, the high processing cost and real-time requirements is the prime challenge. Then, the complexity of the road environment and the high precision requirements of the detection results grow into another challenge. Traditional lane detection methods mainly rely on a combination of highly-specialized, hand-crafted features and utilize Hough transform or Kalman filters, which is vulnerable to the complexity road environment [1]. In recent years, more methods based on deep neural networks, especially convolutional neural networks (CNN) simulated a promising direction [2]. However, considering the severe limited computation resources on vehicle-based system, methods based on CNN encounter a speed bottleneck and cannot meet with the real-time requirements.

Accordingly, to strive for a generalized, low computational cost, and real-time on-board method, a novel segmentation-by-detection (SBD) method is proposed for universal road markings extraction, which can better facilitate the self-driving vehicle. Different from the common road marking segmentation by training semantic network, the road marking segmentation problem is regarded as the target detection task with each road marking serve as an object. Using the lightweight network and simple operation to achieve efficient and high-accuracy detection, supplemented by simple post-process, the segmentation of road markings can be completed. Several reasons can account for this strategy: First, the extremely simple appearance (mono color) of road markings provides no complicated or distinctive features for road marking segmentation, which would dramatically increase the false positive risk. Adversely, the diversity of road markings, such as dash, solid, turning, and crosswalk, can be the excellent and distinctive features for different markings detection. Secondly, the SBD method has higher segmentation precision robustness than direct semantic segmentation. Since detection network has stronger adaptability for the harsh conditions such as shadows, defects and occlusions than direct segmentation. Moreover, adopting the lightweight network in the first stage can greatly reduce the computational cost, which enables our proposed method runs in a real-time way on the embedded platform.

In summary, we highlight our main contributions below:

- We propose a novel method, named segmentation-by-detection, to extract road markings for autonomous vehicles, and extend road markings to 31 categories.
- We fuse a Siamese attention module into lightweight network to detection road marking accurately and robustly. The dedicated network achieves a competitive precision even in terrible conditions.
- We develop an effective dataset (4812 images and 31 categories) for road marking detection and other related applications.

The remaining of this paper is organized as follows. The latest related work is reviewed on various methods on lane detection and their limitations in Section 2. Our proposed method is fully presented in Section 3. Network training strategy and extensive experiments are elaborated in Section 4. Finally, conclusions and future work have been present in Section 5.

II. RELATED WORK

The main sensors currently available for road marking recognition are on-board vision systems and LIDAR. Due to expensive manufacturing cost and complicated usage process, LIDAR has no priority to be the preferred sensor for this task. This paper mainly discusses image-based road marking detection methods. The edge, geometry, and texture of road markings are the most obviously features in road marking detection, which are the basic index for local positioning and navigation. Once the ROI of road markings are located, the road markings can be clearly detected and segmented. In terms of the different principles of road marking detection algorithm, the road marking detection method on structured road can be classified into traditional image processing method and CNN-based method.

A. TRADITIONAL IMAGE PROCESSING METHOD

The traditional methods can only handle the road marking detection task in some limited scenarios. Mammeri *et al.* [3] used MSER (Maximum Stable Extremal Region) and gradual probability Hough transform to detect lane line. However, this method is vulnerable to occlusion blocked by obstacles and vehicles. Subsequently, Satzoda and Trivedi [4] proposed a system called ELVIS (Efficient Lane and Vehicle Detection with Integrated Synergies) to overcome this problem through lane and vehicle detection integration. Huang *et al.* [5] combined inverse perspective transformation and feature voting mechanism to detect and fit straight line. In which, Kalman filter is utilized to optimize and track the lane position. However, they proposed approach fit all lanes to straight line which would result in a large positioning error. Niu *et al.* [6] proposed a two-stage lane detection method, which applied a modified HT (Hough Transform) to extract small line segments and modeled the lanes by curve fitting to enhance robustness. The method is difficult to deal with the complex lane detection and fitting at intersections and ramps. Ma *et al.* [7] put forward a multiple lane detection algorithm based on optimized dense disparity map estimation. Song *et al.* [8] presented a stereo vision-based driving lane detection and classification system to obtain ego-car's lateral position for advanced assisted driving. Nevertheless, both of the above methods use binocular vision, which is expensive and computational costly.

B. CNN-BASED METHOD

The convolutional neural network demonstrated superior performance for image classification and lane detection [9].

Lee *et al.* [10] proposed a unified end-to-end trainable network to tackle lane and road marking classification in adverse weather conditions. Neven *et al.* [1] considered the lane detection problem as an instance segmentation problem, divided the segmentation task into two branches, lane detection branch and lane embedded branch. Similarly, Wang *et al.* [2] proposed a lane detection method based on deep neural network, named LaneNet. Which is consist of lane edge proposal module and lane line localization module. The network shows robust performance to both highway and urban road scenarios without depending on any assumptions on lane line patterns and lane numbers. In order to generate HD map, Liang *et al.* [11] put forward a CNN architecture with a novel prediction layer and a zoom module, named LineNet. However, their dataset relies on GPS signal as precision, resulting in a final error of approximately 31.3 cm. Chen *et al.* [12] proposed Lane Marking Detection (LMD) based on another CNN mechanism to extract lane marking features and adopted the dilated convolution to reduce algorithm complexity. This is an ordinary process including semantic segmentation and post processing. Recently, Garnett *et al.* [13] introduced a network that can directly predict the 3D layout of lanes in a road scene from a single static image. Their approach can handle complex situations such as lane merging and splitting. Zhang and Mahale [14] used a Global Convolution Networks (GCN) model to tackle lane classification and localization simultaneously. However, this color-based GCN model is subject to dynamic lighting. Pan *et al.* [15] proposed a Spatial CNN (SCNN) method which is suitable for continuous structure and large targets with less surface features, such as lane lines. All of the aforementioned methods provide a promising performance of lane detection, but they mostly designed only for lane lines and be of computational complexity. Therefore, we propose a novel SBD method which can achieve real-time road marking segmentation on the vehicle-based system, and covers almost all types of road markings, which can provide convenience for more advanced automatic driving.

III. PROPOSED METHOD

For the sake of monotonous color and succinct surface texture of road marking, it is no useful and distinct features for semantic segmentation. So, it does not work well when carrying out semantic segmentation directly which will give rise to high false positive rate. Oppositely, various road markings inherent with abundant structural features, which is favorable for classification and detection algorithm. Accordingly, to strive for a generalized and real-time on-board solution, we consider the semantic segmentation problem of road marking as an object detection task, and propose a segmentation-by-detection method. The general algorithm pipeline consists of three submodules which is shown in the Fig. 1.

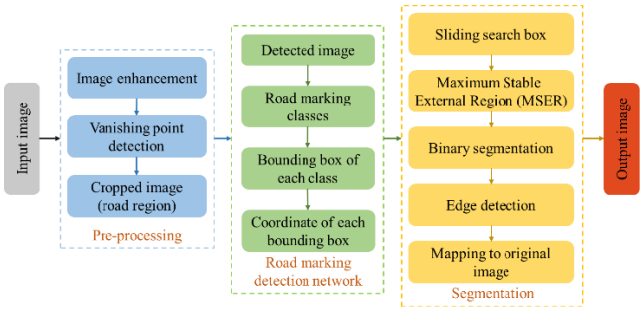

FIGURE 1. The general pipeline of our proposed segmentation-by-detection method.

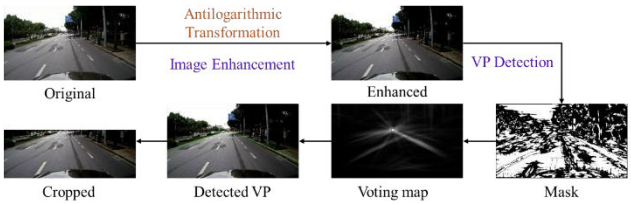

FIGURE 2. Pre-processing pipeline.

A. PRE-PROCESSING MODULE

1) IMAGE ENHANCEMENT

The purpose of image enhancement is to improve the visual effect of the image, suppress the useless interference information, highlight the meaningful information for machine analysis, and transform it into a more suitable form of expression for machine detection. We enhance the contrast of the image to strengthen the white road markings from black road background. In terms of the convex property of logarithmic function, the logarithmic transformation (as shown in (1)) method can improve the brightness of the low gray portion in the image. Inversely, in the road marking detection task, the features of road marking need highlighted. Hence, the method of antilogarithmic transformation (the concave property) is adopted, which can brighten the high gray part and dim the low gray part, as (2) shows.

$$s = c \cdot log_{v+1}(1 + v \cdot r), \quad r \in [0, 1] \tag{1}$$

$$s = \frac{((v+1)^r - 1)}{v}, \quad r \in [0, 1] \tag{2}$$

where the c is a constant, r is the normalized pixel value, the output range of s is [0, 1]. And finally, the value of s will be restored to the original image gray value interval [0, 255], the effect is shown in Fig. 2. In addition, v represents the transformation intensity coefficient, and the larger v, the more obvious the contrast enhancement.

2) VANISHING POINT DETECTION

As the distance increases, the visibility of the road markings decreases. Inspired by this, a road vanishing point (VP) detection algorithm was designed, which can be used to provide a

global geometric constraint and guide robust road marking detection in appropriate area.

First, the texture of image is analyzed. The Gabor wavelet filter is widely used in image texture analysis for its sensitivity of the image local edges, which provides good direction and magnitude response. In our work, the joint 4-direction Gabor filter proposed by Moghadam *et al.* [16] is improved to reduce the computational cost. The 4-direction Gabor filter bank is defined as follows:

$$G_{\varphi,\omega}(x,y) = \frac{\omega}{\sqrt{2\pi}c} e^{-\frac{\omega^2\left(4a^2+b^2\right)}{8c^2}} \left(e^{-ia\omega} - e^{-\frac{c^2}{2}}\right) \quad (3)$$

where, $a = xcos\varphi + ysin\varphi$, $b = -xsin\varphi + ycos\varphi$, φ denotes the direction angle, $\varphi \in \{0°, 45°, 90°, 135°\}$. $\omega = 2\pi/\lambda$ denotes the radial frequency, frequency multiplication constant $c = \pi/2$, spatial frequency $\lambda = 4\sqrt{2}$. Convolving the image $I(x, y)$ with a Gabor filter with a direction angle of φ, the magnitude response of the pixel $p(x, y)$ can be obtained as follows:

$$E_{\varphi,\omega}(p) = I(p) \times G_{\varphi,\omega}(p) \quad (4)$$

Since E contains both the real and imaginary components, the square root of the real and imaginary parts is calculated to obtain the total magnitude response as shown in (5).

$$M_{\varphi,\omega}(p) = \sqrt{Re\left(E_{\varphi,\omega}(p)\right)^2 + Im\left(E_{\varphi,\omega}(p)\right)^2} \quad (5)$$

Then, the main direction of pixel texture is calculated. In order to judge whether the pixel p has obvious directivity and calculate its texture main direction, the four magnitudes obtained by (3) are first sorted in descending order $M^1_{\varphi,\omega}(p) > M^2_{\varphi,\omega}(p) > M^3_{\varphi,\omega}(p) > M^4_{\varphi,\omega}(p)$. If the ratio of $M^1_{\varphi,\omega}(p)$ to $M^4_{\varphi,\omega}(p)$ is larger, the directivity of the pixel p is more obviously. Moghadam *et al.* did not consider the influence of weak textured pixels. If weak texture points are used for voting algorithm, the computational cost will be significantly improved, which will also affect voting accuracy. To this end, we introduce confidence to rule out weak texture pixels. The confidence of pixel p is defined as follows:

$$Conf(p) = 1 - \frac{M^4_{\varphi,\omega}(p)}{M^1_{\varphi,\omega}(p)}, \quad M^1_{\varphi,\omega}(p) > M_{th} \quad (6)$$

where, M_{th} is a threshold constant. The pixel confidence with a maximum magnitude less than M_{th} is set to zero. The pixels with a confidence lower than 0.5 will be deleted and the remaining texture points as valid voting points.

At last, using the sparse texture voting method to examine the relationship between each pixel and the texture direction, and with reference to the confidence of each valid point to construct a voting map, as shown in Fig. 2. Iterating through each voting point to locate the point with the most votes as road vanishing point, and then crop the image area below the VP and feed it to the road marking detection network. After the VP been detected, the region of interest (ROI) can be narrowed, which would result in computation and interference reducing for subsequent algorithm.

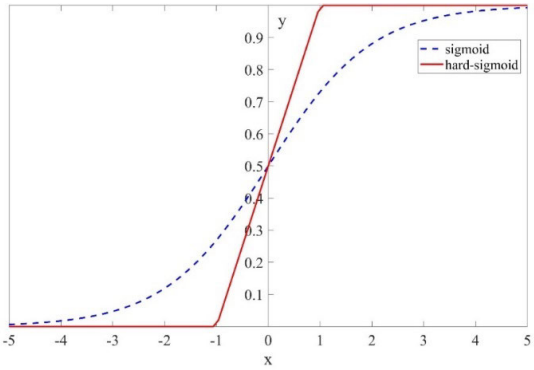

FIGURE 3. Comparison of sigmoid and hard-sigmoid curves.

B. ROAD MARKING DETECTION MODULE

It is of top priority to design a high-efficiency backbone network with less computation and fewer parameters for the deployment of deep CNN on vehicle-based system. So, we designed a dedicated lightweight network RMN (Road Marking Network) with 14 convolution layers, 2 max-pooling layers and 2 avg-pooling layers for the road marking detection. Each convolutional layer bundles with a BN operation, which can improve the speed of the model convergence, achieve a certain regularization effect, and reduce the risk of over-fitting. The max-pooling layer is inserted to eliminate interference features in the image (such as dirty and damaged road surface) while reducing computational complexity. In the later stage of convolution, the image size is small (12x12, 3x3) and carries a lot of high dimensional features. The utility of avg-pooling is to maintain the feature of each region and achieve the effect of model compression.

1) ACTIVATION FUNCTION

Sigmoid function plays an important role in neural network. However, its derivation in backpropagation involves the division operation, which will bring in a huge amount of computation. Therefore, to further compress the network parameters, the approximate function hard-sigmoid (HS) function was adopted (as shown in Fig. 3 and (7)), which can greatly reduce the calculation amount and ensure the real-time capability of this network.

$$y = f(x) = max\left(0, min\left(1, \frac{x+1}{2}\right)\right) \quad (7)$$

2) LOSS FUNCTION

In terms of the sample imbalance in the training dataset, a coefficient was used to describe the importance of different samples in loss function. For a small number of samples (such as arrows and zebra crossing), we enhance its contribution to loss. While for common samples like lane lines, we suppress its contribution to loss. According to the statistics of different samples in the dataset, the distribution ratio of the lane line to the remaining road markings is about 10:3. So the WCE function shown in (8) is utilized in the network and the loss

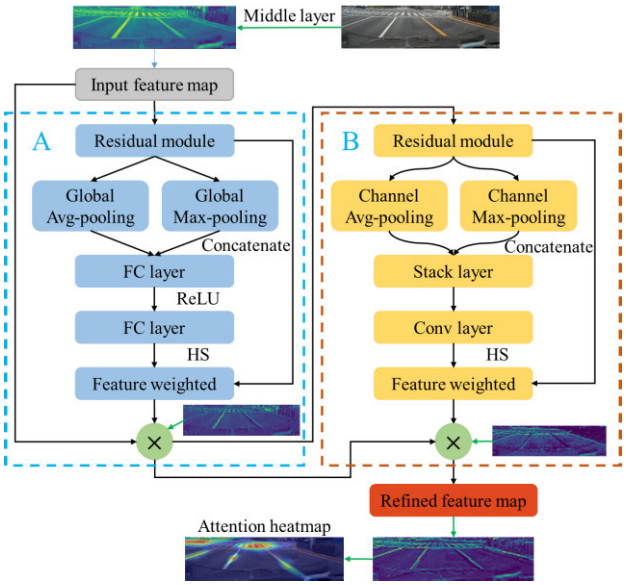

FIGURE 4. Siamese attention module: channel attention (A) and spatial attention (B).

of lane line is multiplied by 0.3,

$$loss = -\sum_{i}^{N} [\omega_i y_i \, log\,(P_i) + (1 - y_i) \; log\,(1 - P_i)] \quad (8)$$

where ω_i is weighted coefficient for different classes of N, y_i denotes ground truth, P_i denotes predicted probability.

3) SIAMESE ATTENTION MODULE

For the object detection task, the extraction network pays different attentions to the major feature region of various objects. If each feature map is treated with the same weight at the initial training, it will increase the time for network convergence. Since the position accuracy and the classification accuracy are mutually promoted, the detection algorithm benefits from the precise and distinct features compared to the classification network. Moreover, according to Woo *et al.* [17], attention mechanism is a better choice to adaptively adjust the weight of different feature maps. In addition, the introduced parameter quantity by attention module does not affect the real-time performance of the network and obtains mAP improvement.

Different from Woo *et al.*, we designed a Siamese attention module with channel attention and spatial attention and replaced the sigmoid function with hard-sigmoid in each attention module to strive for computational efficiency. The relationship of spatial features can also be used to model for the supplying of positional information that the channel attention module cannot obtain. On this basis, the entire attention module (Fig. 4) simultaneously filters and concatenates the feature information of the channel and space.

In the Siamese attention module, adding global avg-pooling can compensate for the residual information overlooked by the global max-pooling. For a feature map from the middle layer, the attention module will infer the first-order channel attention map and the second-order spatial attention map. The original feature map is continuously multiplied with the inferred first-order and second-order feature map to obtain the final refined feature map. The details of the network are shown in the Table 1. At the end of the pipeline, the attention heatmap produces more succinct visual explanations and more accurately exposes the network's behavior [18]. At the end, the softmax is used to classify 31 road markings.

TABLE 1. Road Marking Network (RMN) details. SAM denotes whether there is a Siamese attention module in this layer. NL denotes the type of nonlinearity used. Here, RE denotes ReLU and HS denotes hard-sigmoid.

Input	Operator	Filters	Stride	Pad	SAM	NL
544^2x3	Conv 3x3	16	1	1	N	RE
544^2x16	Maxpool 2x2	/	2	0	N	HS
272^2x16	Conv 3x3	32	1	1	Y	RE
272^2x32	Maxpool 2x2	/	2	0	N	HS
136^2x32	Conv 3x3	64	1	1	Y	RE
136^2x64	Conv 5x5	48	3	2	N	HS
45^2x48	Conv 1x1	64	1	0	Y	RE
45^2x64	Conv 3x3	64	2	1	N	HS
23^2x64	Conv 5x5	80	1	2	Y	RE
23^2x80	Conv 5x5	112	2	2	N	HS
12^2x112	Conv 3x3	144	1	1	Y	RE
12^2x144	Conv 3x3	224	1	1	N	RE
12^2x224	Conv 5x5	224	1	2	N	RE
12^2x224	Conv 1x1	256	1	0	N	RE
12^2x256	Avgpool 2x2	/	2	0	N	HS
6^2x256	Conv 3x3	512	2	1	N	HS
3^2x512	Avgpool 3x3	/	1	0	N	HS
1^2x512	Conv 1x1	1024	1	0	N	
1^2x1024	Softmax	Classifier				

C. SEGMENTATION MODULE

Segmenting road marking directly on the global image is a challenging task, often accompanied with a waste of computation and an increase in false detection. In this paper, after the image is detected by RMN, the accurate local region of road markings can be retrieved. These regions retain vivid road markings and elegant road background without strong interference, and the segmentation task is turn to be a binary task in nature. The specific pipeline and the visualization of segmentation procedure are referred to Fig. 5.

Each detected road marking has a regression box and corresponding coordinates $(x, y, w, h)_i, i \in \{0, 1, 2, \cdots\}$. We extract the center point of all regression boxes of the same class and calculate the gradient:

$$d_c = \frac{\Delta y_c}{\Delta x_c} = \frac{y_j - y_i}{x_j - x_i}, \quad c \in \{(i, j) | \in (0, 1, 2, \cdots), i \neq j\} \quad (9)$$

between each two center points. According to the principle of gradient similarity s_c (in (10)), points with similar gradients are clustered into the same cluster, which denotes the same class of road marking.

$$s_c = \frac{|d_c - d_{c+n}|}{|\Delta y_c| + |\Delta x_c| + |\Delta y_{c+n}| + |\Delta x_{c+n}|} \leq 0.05, \quad n \in (1, 2, 3, \cdots) \quad (10)$$

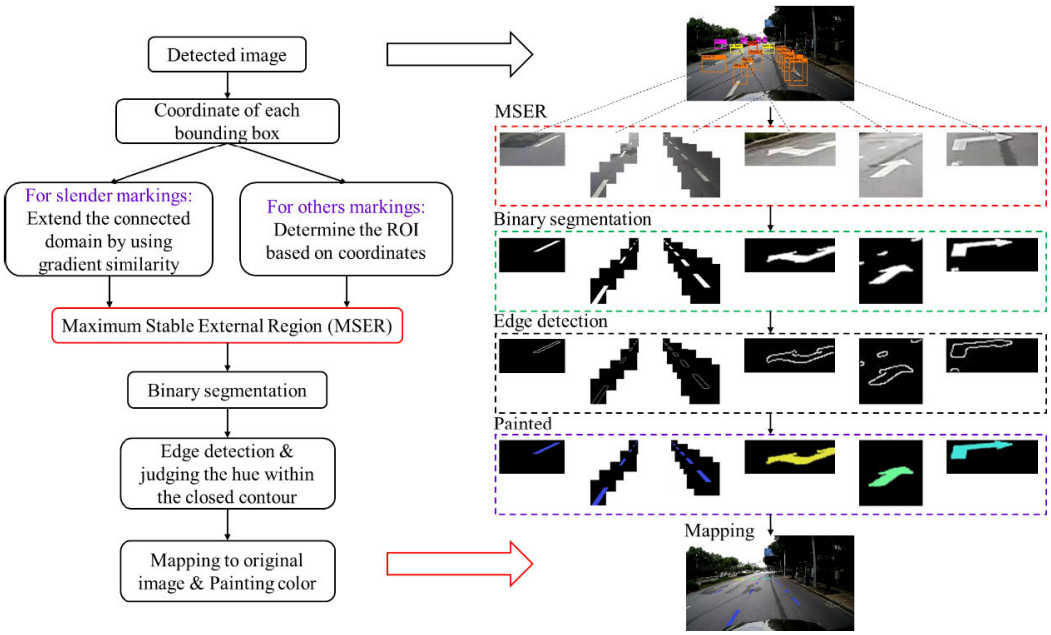

FIGURE 5. Pipeline of segmentation module.

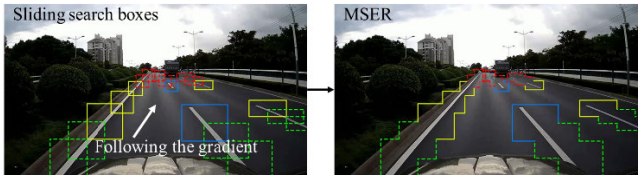

FIGURE 6. Visualization of sliding search box algorithm. Solid boxes denote the detection result, while the dashed boxes denote the sliding search boxes. These search boxes slide following the gradient to extend valid connected regions.

FIGURE 7. Image acquisition platform (left) and Annotation example (right).

Then merging the bounding boxes of the same cluster to form a connected region via maximum stable external region (MSER) algorithm:

- For the slender markings (e.g. lane lines): the sliding search box algorithm is used to extend the connected region along the gradient direction. Selecting the center point of the top and bottom bounding box as the initial position, respectively. The size of the up-sliding search box is set as to the top box, the other is set as to the bottom box. Then along the gradient direction, perform sliding twice with 0.5 scale ratio at both ends. The dynamic sliding search process is visualized in Fig. 6. At last, combine all the bounding boxes into a MSER. By doing this, we can refine the object region where the detection module overlooked due to occlusions and farther distance.
- For other markings (e.g. stop line and zebra crossing): the bounding box contains the entire structure of these road marking, and it can be converted directly into MSER.

After the MSER been determined, the binary operation was executed these regions, and then the internal contour edges can be easily obtained from these binary images. Integrated with the color information of the original image to identify these closed contours to determine whether it is a road marking or other disturbance. Eliminating the small, disorder closed contours, the entire road marking pattern can be obtained. The segmented road markings will be mapped back to the original image with a certain color painted on.

IV. EXPERIMENTS

In this section, datasets used in network training will be firstly introduced. Then we presented network training strategy. Moreover, the hardware platform was elaborated in our experiments and illustrate some results. Finally, our approach was compared with other state-of-the-art lane detection methods and demonstrated the effectiveness of our method in the embedded devices.

A. DATASETS

We gathered 4812 images (1920 × 1080 resolution) from our road scene collection platform with four cameras and one Lidar, as shown in Fig. 7. We first recorded hundreds of kilometers of road videos, then split all the videos into

TABLE 2. Details of each class.

ID	Label	Full Name	ID	Label	Full Name
0	car	car	16	lor	Left or right
1	bus	bus	17	s40	Speed limit 40
2	truck	truck	18	s50	Speed limit 50
3	zebra crossing	zebra crossing	19	s60	Speed limit 60
4	stop line	stop line	20	s80	Speed limit 80
5	solid ll	solid lane line	21	s90	Speed limit 90
6	dash ll	dash lane line	22	s100	Speed limit 100
7	guide ll	guide lane line	23	s120	Speed limit 120
8	straight line	straight line	24	rop	Roadside no parking
9	left turn	left turn	25	cross hatch	Cross hatch
10	right turn	right turn	26	brt lane	BRT lane
11	sl turn	Straight or left turn	27	bus lane	Bus lane
12	sr turn	Straight or right turn	28	diversion line	Diversion line
13	turn around	Turn around	29	no passing	No passing
14	sta	Straight or turn around	30	slow down	Slow down
15	lta	Left or turn around			

frames. After that, several people selected high-quality road marking images from these frames, and finally 4812 valid images are remained. These collected images cover a wide variety of scenes, such as day, night, rain, cloudy and so on. These images are elaborated into a road marking dataset with 31 categories (Table 2), which is already available on *https://github.com/namemzy/Road-Marking-Segmentation.* Among the dedicated dataset, the data is shuffled first, 3812 images are used to train our road marking detection network and the rest for validation.

B. NETWORK TRAINING

According to Smith *et al.* [19], the strategy of learning rate in our training process is increasing first and then decreasing. Specifically, the first five epochs are warm up with learning rates increase from 0 to 0.05 and then decrease at the rate of 2e-04/epoch. The momentum optimization algorithm was used and set it to 0.9. The training process is executed on 4 GPUs in PyTorch framework, and the loss during training is shown in the Fig. 8. From the loss curve of training, it can be concluded that under the same epochs, the Siamese attention module can accelerate the network training and make it converge at a smaller loss. Meanwhile, this module can also enhance the network learning capability, which makes the network achieve higher accuracy.

C. COMPARES WITH OTHER METHODS

Due to most of the public methods are currently only for lane line detection, in this section, we compare the performance of lane detection with the existing five methods on five datasets, regardless of the other road markings detection. The evaluation of lane detection is based on the pixel-level accuracy. In order to judge whether a lane line is successfully

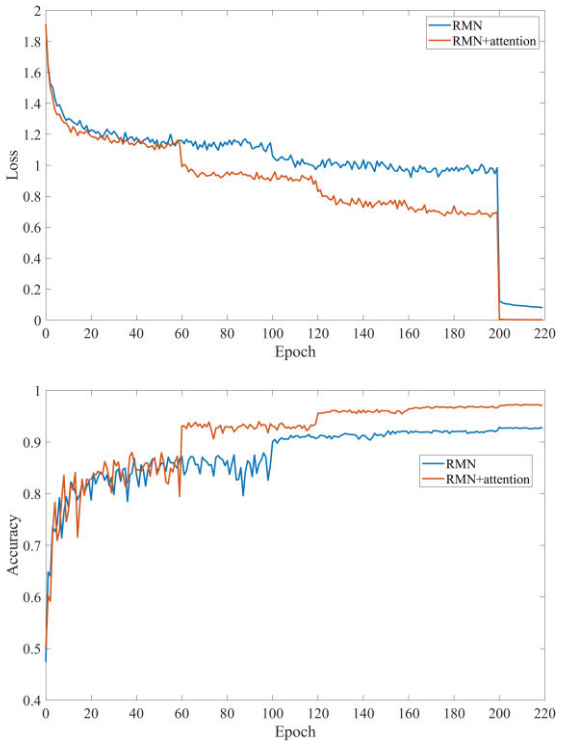

FIGURE 8. Loss curve (top) and accuracy curve (bottom).

TABLE 3. Comparison of F1-score on public datasets.

Method	Caltech Lanes	BDD [22]	TuSimple	CULane [15]	Ours
LaneNet	0.875	0.787	0.784	0.806	0.903
LineNet	**0.955**	/	/	0.731	/
VPGNet	0.884	0.844	0.813	0.727	0.946
Neven et al. [1]	0.903	0.826	0.792	0.751	0.919
SCNN	0.927	0.791	0.828	0.713	0.928
RMN+SAM (Ours)	0.934	**0.883**	**0.836**	**0.842**	**0.974**

segmented, 1000 images were randomly selected from each dataset. When the intersection-over-union (IoU) of lane pixels between the prediction and ground truth is larger than 0.5, this prediction is regard as a true positive (TP). Then, we employ the widely used evaluation metric F1-score to measure each method as (11) shown. Where the $P = TP/(TP + FP)$ and $R = TP/(TP + FN)$, F1 denotes the F1-score.

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{11}$$

The experimental results are shown in Table 3. The Caltech Lanes [20] dataset was captured by the fisheye lens. The image was distorted and the resolution was not distinguishing enough, which led to a decline in the performance of our method. As we all known, although the TuSimple [21] is the most difficult dataset for lane detection, our method attains competitive results and is slightly superior to SCNN. In general, our proposed method achieved the satisfactory

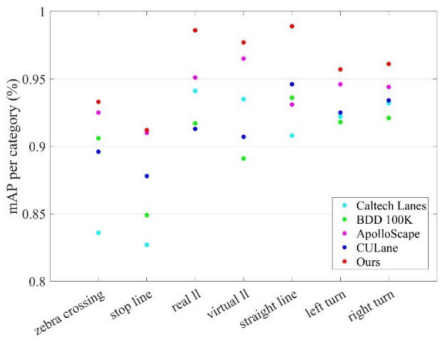

FIGURE 9. The mAP of 7 common road markings on five datasets.

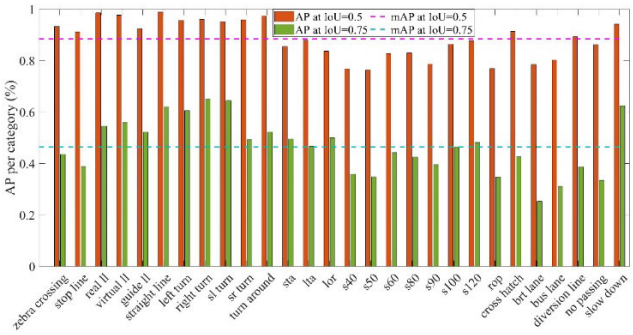

FIGURE 10. AP at 0.5 and 0.75 IoU of each category.

performance on most datasets, which demonstrates the effectiveness on diverse real-world scenarios.

D. TEST ON PUBLIC DATASETS

In terms of the imbalanced distribution of road markings in above five datasets, some categories are rare or not appear in some datasets. For example, there are few speed limit markings on the roads in the United States and Germany. Therefore, we only tested 7 categories that exist in great quantity of each dataset. As shown in the Fig. 9, the mAP of these 7 categories at IoU = 0.5 are displayed. From the scatter picture, red points denote the detection mAP on our dataset, which hit the highest accuracy in the experiments. Globally, the accuracy of all datasets is above 80%, which proves the robustness of our method.

To further evaluate the effectiveness of our proposed method, we tested the segmentation accuracy of 28 road markings (except car, bus, truck) on our test set and the results is shown in the Fig. 10. According to Fig. 10, our SBD method achieved excellent mAP = 0.885 at IoU = 0.5. However, due to the imbalance of categories, the number of speed limit markings is too small compared to the lane lines and arrows, resulting in a significant drop in mAP (0.465) when the IoU is 0.75. It will be improved through data augmentation and scene synthesis in the future work. Statistically, our method achieved great segmentation effect in most categories, which has met the detection and segmentation of the main road markings in various scenarios.

To intuitively demonstrate the effectiveness of our SBD method, here in Fig.11, Fig. 12 and Fig. 13 some segmentation results on the cross datasets are shown. The proposed method SBD can effectively detect horizontal curves, such as stop lines and zebra lines, since these classes have been annotated in advance. The bounding box positions of the lane lines on the roads are accurate regardless the shapes and the number of them. And the segmentation results are fined, which demonstrate the effectiveness of our proposed method to diverse real-world scenarios. From Fig. 11, although the effect at night and rain is a slight weak, the segmentation of the ego-car lane is still very accurate. Moreover, the other road markings are properly estimated, which we believe it is attributed to the ability of our designed Siamese attention module to exert the holistic attention on the road markings from the channel and spatial features.

From Fig. 12, it can be found that our proposed method can handle extreme scenes such us intersections, ramps and occlusions by another car. But the segmentation result is a little worse at the road crossing as the road markings are highly dense and messy. In addition, the dash line sometimes make the bounding boxes separated and discontinuous. When the distribution of dash line is dense, the bounding boxes are generally continuous. When the dashed lines are scattered, the bounding boxes are independent of each other. In this paper, in order to segment the lane lines, it is consider feasible if the regression boxes completely cover the dash line. Therefore, there is no optimization of the bounding box discontinuity for the dash line. In our future work, when the lane line needs to be parametric fitted, we can optimize the MSER algorithm by adding the connectivity verification and merging the discontinuous bounding boxes to obtain the integral dash line area.

E. ALGORITHM DEPLOYMENT

The heterogeneous computing platform was employed to accelerate our algorithms, CPU for morphological operations and GPU for vanishing point detection and network inference. Firstly, CPU loads the image and performs an image enhancement operation, which takes less than 1ms. The enhanced image will be pushed onto CUDA for vanishing point detection and network inference. Then, the results of the network detection return back to the CPU for the operation of segmentation module. The CPU multi-thread will be created according to the maximum number of stable external areas. In other words, one thread is responsible for a MSER. At last, all segmentation results are projected to the original image by CPU. Benefiting from the heterogeneous computing mode, the GPU performs network inference for the second frame, and the CPU performs the segmentation operation of former frame, simultaneously. The vanishing point detection and network inference are performed on the GPU, and the time consumption is about 0.41ms and 9.9ms, respectively. The segmentation module is a series of simple morphological operation, which takes about 7ms to finish under the acceleration of opencv and CPU multithreading.

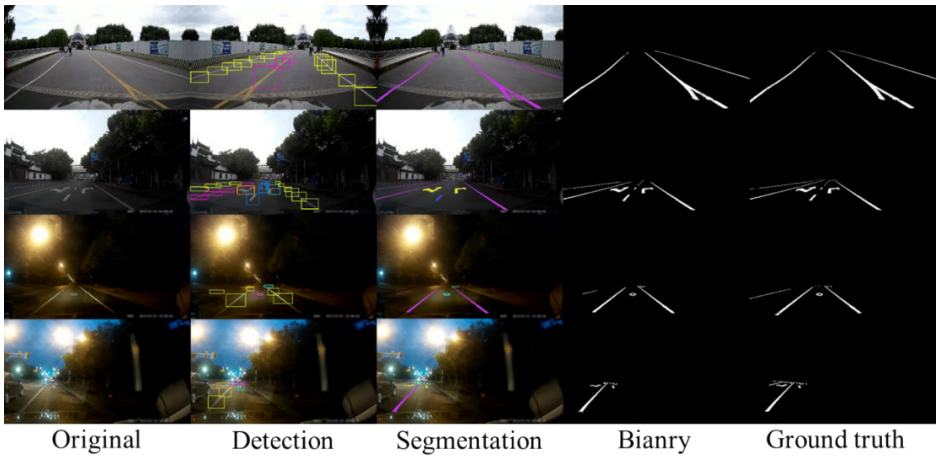

FIGURE 11. Detection and segmentation results on our dataset. From top to bottom is sunny, cloudy, night and rainy scene.

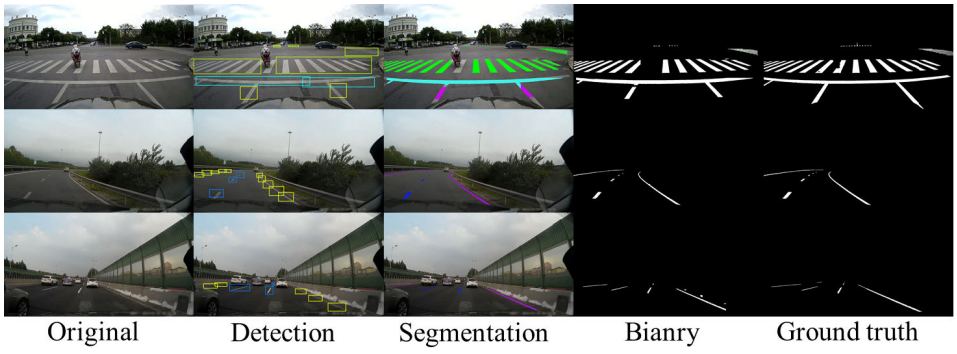

FIGURE 12. Illustration of some extreme scenes on our dataset. From top to bottom is intersection, ramp and occlusion.

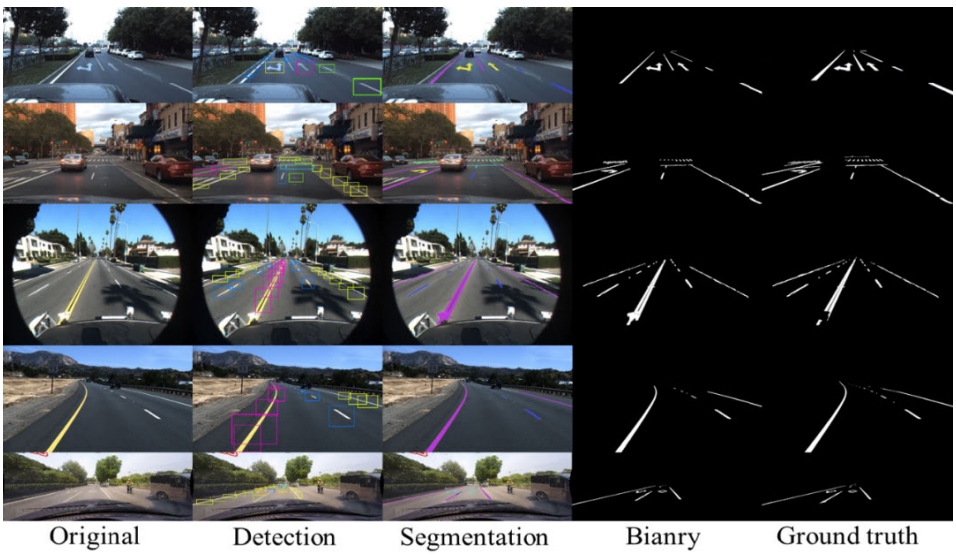

FIGURE 13. Illustration of detection and segmentation results on five public datasets. From top to bottom is ApolloScape [23], BDD 100K, Caltech Lanes, TuSimple and CULane.

Since the running time of the image on the CPU is less than the network inference time on the GPU, the cost time of the entire pipeline depends on the network inference. When our approach was tested on a workstation equipped with Intel Xeon E5 2.4GHz CPU and an NVIDIA GTX 1080 GPU. The average runtime for a 640x480 input is about 10.3ms (97 fps).

TABLE 4. Comparison of the real-time performance of our approach on three embedded devices.

Input	Jeston Tx1		Kirin 970		RK3399Pro	
	ms	fps	ms	fps	ms	fps
640x480	51.3	19.5	32.3	31	22.2	45
800x600	62.5	16	37	27	27.8	36
1080x720	105.3	9.5	62.5	16	41.6	24

F. TEST ON EMBEDDED DEVICES

In high-speed driving scenarios, the vehicle travels at speeds about 120km/h (33.3m/s). It is a strict requirement for the speed of road marking detection algorithm. To ensure the safe driving of the vehicles, the detection speed should not slower than once for every two meters. In other words, the detection algorithm is imperative to achieve a speed of 15 fps in order to achieve real-time performance. Therefore, in our implementation, we test the real-time performance of the proposed method on different embedded platforms, including the NVIDIA Jeston Tx1, Kirin 970 and TB-RK3399Pro. These embedded devices are optimized for deep learning, allowing us to take full advantage of the CPU and GPU computing capability. The detailed test results have been summarized in the Table 4. The images are cropped by pre-processing stage then resized to 544 $\times$ 544 for network input. For the large input (1080 $\times$ 720), it is a challenge for Jeston Tx1, but it can be handled well and achieve the real-time speed by the other two devices. Both the high segmentation accuracy and the real-time running speed further enable the deployment of our road marking segmentation method on vehicles.

G. ABLATION STUDY

In section III.B, we propose a Siamese attention module to make contributions to network learning and inference. The two different methods RMN and RMN with attention module was trained individually. It can be seen from the Fig. 8, the method RMN with attention module converges faster and achieves higher inference accuracy. Which proves the effectiveness of the Siamese attention module in accelerating the network training. From the test results, RMN with attention module retains more feature information and makes conducive to the network inference in our overall approach.

V. CONCLUSION

In this paper, we proposed a segmentation-by-detection (SBD) method composed of pre-processing, road marking detection and segmentation for road marking segmentation. The pre-processing module provides a cropped and enhanced image to feed the road marking detection network. The network involves the approximate activation function and the Siamese attention module to reduce the computation and improve the detection accuracy respectively. We find that integrating attention module increases the accuracy, at the modest increase of number of parameters, and no discernible latency cost. Using the bounding boxes generated by road marking detection network endows our sliding search box and MSER algorithms obtaining entire road marking pattern with less interference. As a result, extensive experiments demonstrate the robustness, accuracy and real-time performance of our SBD method to employed on mobile devices for diverse driving scenarios. Nevertheless, in terms of the data imbalance, the mAP of our approach will cause a slight drop especially on rare road markings. It will be tackled through data augmentation and scene synthesis in the future. At the intersections, the intricate lane lines may disorder our segmentation result, which is also a direction of improvement in our future work.

REFERENCES

[1] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: An instance segmentation approach," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2018, pp. 286–291.

[2] Z. Wang, W. Ren, and Q. Qiu, "LaneNet: Real-time lane detection networks for autonomous driving," 2018, *arXiv:1807.01726*. [Online]. Available: https://arxiv.org/abs/1807.01726

[3] A. Mammeri, A. Boukerche, and Z. Tang, "A real-time lane marking localization, tracking and communication system," *Comput. Commun.*, vol. 73, pp. 132–143, Jan. 2016.

[4] R. K. Satzoda and M. M. Trivedi, "Efficient lane and vehicle detection with integrated synergies (ELVIS)," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, vol. 1, Jun. 2014, pp. 708–713.

[5] Y. Huang, Y. Li, X. Hu, and W. Ci, "Lane detection based on inverse perspective transformation and Kalman filter," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 2, pp. 643–661, 2018.

[6] J. Niu, J. Lu, M. Xu, P. Lv, and X. Zhao, "Robust lane detection using two-stage feature extraction with curve fitting," *Pattern Recognit.*, vol. 59, pp. 225–233, Nov. 2016.

[7] H. Ma, Y. Ma, J. Jiao, M. U. M. Bhutta, M. J. Bocus, L. Wang, M. Liu, and R. Fan, "Multiple lane detection algorithm based on optimised dense disparity map estimation," in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2018, pp. 1–5.

[8] W. Song, Y. Yang, M. Fu, Y. Li, and M. Wang, "Lane detection and classification for forward collision warning system based on stereo vision," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5151–5163, Jun. 2018.

[9] J. Kim, J. Kim, G.-J. Jang, and M. Lee, "Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection," *Neural Netw*, vol. 87, pp. 109–121, Mar. 2017.

[10] S. Lee, J. Kim, J. S. Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. S. Hong, S.-H. Han, and I. S. Kweon, "VPGNet: Vanishing point guided network for lane and road marking detection and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1947–1955.

[11] D. Liang, Y. Guo, S. Zhang, S.-H. Zhang, P. Hall, M. Zhang, and S. Hu, "LineNet: A Zoomable CNN for crowdsourced high definition maps modeling in urban environments," 2018, *arXiv:1807.05696*. [Online]. Available: https://arxiv.org/abs/1807.05696

[12] P.-R. Chen, S.-Y. Lo, H.-M. Hang, S.-W. Chan, and J.-J. Lin, "Efficient road lane marking detection with deep learning," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[13] N. Garnett, R. Cohen, T. Pe'er, R. Lahav, and D. Levi, "3D-LaneNet: End-to-end 3D multiple lane detection," 2018, *arXiv:1811.10203*. [Online]. Available: https://arxiv.org/abs/1811.10203

[14] W. Zhang and T. Mahale, "End to end video segmentation for driving : Lane detection for autonomous car," 2018, *arXiv:1812.05914*. [Online]. Available: https://arxiv.org/abs/1812.05914

[15] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," 2017, *arXiv:1712.06080*. [Online]. Available: https://arxiv.org/abs/1712.06080

[16] P. Moghadam, J. A. Starzyk, and W. S. Wijesoma, "Fast vanishing-point detection in unstructured environments," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 425–430, Jan. 2012.

[17] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 11211. Cham, Switzerland: Springer, 2018, pp. 3–19.

[18] J. Kim and J. Canny, "Interpretable learning for self-driving cars by visualizing causal attention," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2942–2950.
[19] L. N. Smith, "Cyclical learning rates for training neural networks," Apr. 2015, *arXiv:1506.01186*. [Online]. Available: https://arxiv.org/abs/1506.01186
[20] M. Aly, "Real time detection of lane markers in urban streets," 2008, *arXiv:1411.7113*. [Online]. Available: https://arxiv.org/abs/1411.7113
[21] "Tusimple lane detection challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2017. [Online]. Available: https://github.com/TuSimple/tusimple-benchmark/wiki
[22] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling," May 2018, *arXiv:1805.04687*. [Online]. Available: https://arxiv.org/abs/1805.04687
[23] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The ApolloScape dataset for autonomous driving," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 954–960.

WEIWEI ZHANG received the Ph.D. degree in mechanical engineering from Hunan University, in 2015. He is currently a Lecturer with the Shanghai University of Engineering Science. His research direction is the technology of intelligent vehicle. His team currently undertakes several major projects from renowned Chinese companies. His current research interests include the technology of image processing, intelligent vehicle, and power train of vehicle.

ZEYANG MI received the B.E. degree from the Changshu Institute of Technology, Changshu, China, in 2016. He is currently pursuing the master's degree with the Shanghai University of Engineering Science, Shanghai, China. His research direction focuses on the computer vision. His current research interests include lane detection and intelligent vehicle.

YAOCHENG ZHENG received the B.E. degree from the Huaiyin Institute of Technology, Huai'an, China, in 2017. He is currently pursuing the master's degree with the Shanghai University of Engineering Science, Shanghai. His current research interests include computer vision, intelligent transportation systems, and intelligent automobiles.

QIAOMING GAO received the Ph.D. degree in transportation and the Doctor degree in vehicle engineering from the Beijing University of Aeronautics and Astronautics. He is currently an Associate Professor and a Senior Engineer with the School of Mechanical and Transportation Engineering, Guangxi University of Science and Technology. His main research interests include special vehicle designs, control systems, and experimental equipment development.

WENJING LI was born in 1986. She received the master's degree in automotive electronics from the Shanghai University of Engineering and Technology.