Research of Lane Detection Method Based on Attention Mechanism

Lin Wang and Rongyu Lv(✉)

School of Automation and Information Engineering, Xi'an University of Technology, Xian 710048, China
lvry625@163.com

Abstract. Facing the problems of low accuracy and poor real-time performance in lane detection, a lane detection method based on deep learning was proposed. ENet, a lightweight semantic segmentation network, is used as the backbone of the detection method. In view of the slender characteristics of lane lines, an improved spatial attention module is introduced to enhance the ability to extract lane line features. Then the final detection result is obtained by post-processing operation. Compared with SCNN and ENet, the improved algorithm has better accuracy and meets the requirements of real-time detection.

Keywords: Deep learning · Semantic segmentation · Spatial attentional mechanism · Lane detection

1 Introduction

As an emerging technical force in the field of automotive development, deep learning is making an important contribution to the development of intelligent driving technology.

1.1 Background and Significance

With the rapid development of the automobile industry, there are also many problems, among which the most prominent is the frequent occurrence of automobile safety accidents, which poses a great threat to the stability of people's life and property.

In order to solve these problems, intelligent driving vehicles emerged and began to explore Advanced Driver Assistance Systems (ADAS) and autonomous driving systems. They mainly use different sensor devices carried or loaded by the vehicle itself to monitor the environment around the vehicle. When the vehicle senses potential danger, it can avoid danger through alarm or active error correction and other measures, and guide and control the vehicle with reliable information, so as to provide users with safer driving experience.

As an important traffic sign in road environment, lane line contains abundant scene information. Lane detection is one of the basic work and key technologies in the field of intelligent driving. At present, many intelligent functions such as lane keeping detection, deviation warning and collision avoidance warning have been realized. Lane line

R. Sun and J. Zhang (Eds.): ICAICC 2022, ACSR 97, pp. 60–66, 2023.
https://doi.org/10.2991/978-94-6463-002-2_8

detection is still very challenging and practical in current autonomous driving research, because vehicles will run at a relatively high speed under various complex and changeable road conditions. This requires that in the actual driving environment, lane line detection should meet the requirements of real-time and robust. And through the research of this technology, it is of great significance to real-time identification of passers-by, vehicles and road lanes in the driving environment, improve driving safety and reduce the occurrence of road traffic accidents.

1.2 Research Status

Lane line detection usually uses vehicle-mounted sensors to collect data and image processing methods to detect the specific location of lane lines. However, these technologies have some disadvantages, such as high cost, poor real-time performance, low accuracy and easy to be interfered by the environment. With the continuous improvement of computer processing power and the development of artificial intelligence technology, there are more and more research methods of deep learning. Lane detection, as an important part of intelligent transportation, plays an increasingly important role in vehicle navigation and unmanned driving. How to identify the road quickly and effectively is one of the key problems. At present, lane detection methods are mainly divided into traditional detection methods and deep learning based detection methods.

The traditional method is based on the characteristics of lane lines [1, 2] and model [3] to conduct in-depth research. Compared with the road background, lane lines have obvious characteristics of color, edge, shape and gradient change [4]. The feature-based detection method is to manually set the experience value based on the difference between pixels to realize the recognition and positioning of lane lines. This method is suitable for simple driving environment, that is, in the case of clear and complete lane lines and obvious edge features, the detection effect is good. This type of algorithm mainly extracts shallow features of lanes, which are prone to environmental interference, such as shadows of trees or buildings, shadows of other vehicles on the road, tire tracks and traffic signs, which will make the robustness of lane line detection worse. Model-based detection methods mainly use the structure and shape characteristics of lane lines [5], such as whether the lane lines are continuous, parallel to the lane lines, or whether there is an infinite distance at a certain point, etc., to establish mathematical models to describe the lane lines, so as to achieve the purpose of lane line detection. This method does not require complex prior knowledge, and has good real-time performance, which can meet the real-time requirements. According to the actual application situation, a reasonable selection of the model can obtain a better detection effect.

At present, the smooth application of deep learning technology in the field of computer vision has produced a large number of detection methods based on deep learning. This is mainly due to the strong feature extraction ability of convolutional neural network, which has been widely used in image recognition and classification tasks. Different from traditional detection methods, deep learning builds complex network structure by learning a large number of samples, and trains corresponding network parameters on this basis, so as to complete data understanding and analysis. This method does not depend on any prior knowledge and has strong adaptability and robustness. Lane detection methods based on deep learning are often transformed into semantic segmentation or instance segmentation tasks, which classify the pixels of images by semantic or instance.

At present, such detection methods still have some problems, such as complicated network structure and large number of calculation parameters. Therefore, this paper studies its detection performance from the perspective of semantic segmentation to achieve the robustness and real-time performance of the algorithm in complex driving environment.

2 Method

Lane detection is an important part of intelligent driving and the basis of normal driving, lane departure warning and lane navigation. In complex road environment, lane detection will be affected by rain and snow weather, obstacles and road shadows, but the traditional lane detection algorithm mostly adopts manual experience to judge, the accuracy is low. At present, the lane detection method based on convolutional neural network has strong feature extraction ability and end-to-end characteristics, without manual intervention, and has high accuracy and robustness. However, the general model has large depth, high computation requirements and poor real-time performance. This paper uses the network model based on semantic segmentation to design.

Traditional methods often ignore the influence of image details, resulting in some errors in the results. In order to solve the above problems, a detection algorithm based on semantic segmentation is proposed. In practical application, due to the model parameter setting is not uniform or the influence of mobile terminal is not taken into account, the results tend to be biased. Therefore, ENet [6] is selected as the backbone network to realize lane line detection. ENet is one of the most successful methods based on semantic segmentation at present. Its size is only 0.7M, which meets the speed requirements of lane line detection under high-speed driving conditions. In this paper, the original ENet is improved at the cost of less time consumption, so as to improve its detection accuracy on lane lines.

ENet is a lightweight semantic segmentation network for street View proposed by Paszke et al. in 2016, which adopts an asymmetric large encoder small decoder structure. Compared with other semantic segmentation networks, it not only has fewer parameters and faster computing speed, but also can meet the requirements of street view segmentation task well. At the same time, it has certain structural plasticity, which can be further improved and optimized. Its network structure is shown in Table 1.

The lane line itself has its particularity. First, the shape of the lane line is thin and long, and the pixel points of the lane line in the image account for a small proportion. However, because the lane line is relatively long, its scope in the image is relatively large. Second, in reality, lane lines are divided into solid lines and dashed lines, but in data sets, dashed lines on the same road are usually represented by continuous curves. Third, the blocking of some vehicles on the road caused too little display of the lane line, or even unable to see the lane line. Based on the unique situation of these lane lines, we hope that the lane line detection network has a larger receptive field, can see more details, and has a strong ability to extract global features. Therefore, the spatial attention mechanism is added into the network to enhance the representation ability of feature maps.

Table 1. ENet network structure

	Name	Type	Output
	Initial		16 × 256 × 256
Encoder-1	Bottleneck 1.0	Downsampling	64 × 128 × 128
	4 × Bottleneck1.x		64 × 128 × 128
Encoder-2	Bottleneck 2.0	Downsampling	128 × 64 × 64
	Bottleneck 2.1		128 × 64 × 64
	Bottleneck 2.2	Dilated 2	128 × 64 × 64
	Bottleneck 2.3	Asymmetric 5	128 × 64 × 64
	Bottleneck 2.4	Dilated 4	128 × 64 × 64
	Bottleneck 2.5		128 × 64 × 64
	Bottleneck 2.6	Dilated 8	128 × 64 × 64
	Bottleneck 2.7	Asymmetric 5	128 × 64 × 64
	Bottleneck 2.8	Dilated 16	128 × 64 × 64
Encoder-3	Repeat Section 2, without Bottleneck 2.0		
Decoder-1	Bottleneck 4.0	Upsampling	64 × 128 × 128
	Bottleneck 4.1		64 × 128 × 128
	Bottleneck 4.2		64 × 128 × 128
Decoder-2	Bottleneck 5.0	Upsampling	16 × 256 × 256
	Bottleneck 5.1		16 × 256 × 256
	fullconv		C × 512 × 512

Because the lane lines are very slender and the span of space is relatively large, the network is required to have good global feature extraction ability. Therefore, this paper improves the traditional spatial attention module. Average pooling gives equal attention to all eigenvalues, while maximum pooling concentrates all attention on the largest, which is not optimal. In order to extract more features, a method of adding adaptive weight coefficients λ and μ after maximum and average pooling is proposed. These two weight parameters are iterated and updated continuously through network training to select the optimal collocation scheme. The improved spatial attention model is shown in Fig. 1.

The specific implementation process is as follows: the maximum and average pooling of the input feature graph is carried out, and the high-level feature graph is generated after fusion. Sigmoid function is used to obtain the weight of each channel (0–1) of the input feature layer. After obtaining this weight, multiply the weight by the original input feature layer to get the output feature result graph. Among them, the formula for feature fusion is as follows.

$$F_1 = \lambda F_{\max} + \mu F_{mean} \tag{1}$$

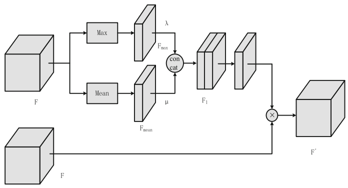

Fig. 1. Improved spatial attention module

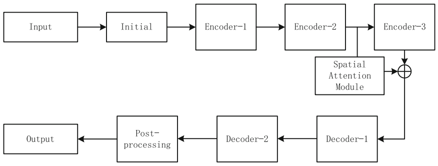

Fig. 2. S-ENet model structure

After the attention module is placed on the encoder that generates high-level semantic features, it is beneficial to extract more rich features.The improved model structure S-ENet is shown in Fig. 2.

In order to obtain the final detection results, post-processing operations are required. In this paper, clustering and fitting methods are used for post-processing operations. Mean shift clustering [7] algorithm is used to divide feature points belonging to the same lane line. The advantage of this algorithm is that it does not need to set the number of clustering in advance. For fitting, the least square method [8], the most common method in mathematical modeling, is used to fit lines and curves simultaneously.

3 Experiment and Result

In this paper, TuSimple dataset, a public data set for lane detection, is selected to test the model. This data set is collected for different driving scenarios (such as multi-lane, expressway, etc.) and different traffic conditions, which has a certain complexity. Due to the high resolution of images in the dataset in this paper, putting them directly into the lane segmentation model for training will waste more GPU resources. Therefore, the nearest neighbor interpolation method [9] is used in this paper to reduce the size of input images.

In this paper, the evaluation index in TuSimple Lane Line Challenge was used to evaluate, and Acc (Accuracy) was taken as the final evaluation index. Since lane lines on the TuSimple dataset are usually labeled as discrete points, the accuracy is represented by the average of predicted correct lane line points on each image, with the following formula.

$$Acc = \sum_{i} \frac{C_i}{S_i} \tag{2}$$

The visualization results obtained after training on the TuSimple dataset are shown in Fig. 3. Dotted lines in the figure are lane lines, and solid lines are detected lane lines. As

Fig. 3. Visual result

Table 2. ENet network structure result

Method	Acc (%)	Speed (fps)
SCNN	96.5	23.8
ENet	95.7	52.6
S-ENet	97.2	70.4

can be seen from the figure, there are some shadow interference and lane line damage in each figure, but the model in this paper can still successfully segment lane lines, and the segmentation results are relatively fine, indicating a high robustness. For the normal driving scenario in the figure, the model shows good accuracy; For the images with vehicle occlusion and severe shadow occlusion, the prediction results of S-ENet model are slightly wrong, with some missed detection. However, since most of the lane line pixels have been successfully segmented, a few missed detection points have little influence on the overall results.

In order to verify the effectiveness of the experiment, ENet, SCNN [10] and the improved S-ENet were tested on TuSimple data set, and the experimental results of these algorithms were listed as shown in Table 2.

4 Conclusion

This paper studies the background and development status of lane line detection technology, and selects the deep learning method to detect lane lines. In view of the particularity of lane shape, detection is regarded as a segmentation problem. On the basis of lightweight network ENet, the improved spatial attention module is added to enable the network to extract richer features and enhance the ability to represent context information. Experiments on Tusimple data set also show the reliability of the algorithm. By comparing ENet and SCNN networks, the advantages of the improved algorithm are demonstrated.

Although the model in this paper has achieved good results on the TuSimple dataset, there are still some imperfections in the model. In actual driving scenarios, lane line detection is usually varied. In the experiment of this paper, all lane lines were grouped into one class without distinguishing different lane lines. In addition, different data sets need to be added to make the model generalizing.

References

1. Dang, L., Tewolde, G., Zhang, X., et al.: Reduced resolution lane detection algorithm. In: 2017 IEEE AFRICON, pp. 1459–1464. IEEE Computer Society, Washington, DC (2017)
2. Liu, Y., Zhou, C., Liu, Y., et al.: Traffic lane detection based on edge feature points clustering. Sci. Technol. Eng. **19**(27), 247–252 (2019)
3. Wu, L., Yu, Q.: Research on a fast and accurate unstructured road detection method. Comput. Simul. **33**(09), 174–178 (2016)
4. Chiu, K.Y., Lin, S.F.: Lane detection using color-based segmentation. In: 2005 IEEE Proceedings of the Intelligent Vehicles Symposium, pp. 706–711. IEEE (2005)
5. Low, C.Y., Zamzuri, H., Mazlan, S.A.: Simple robust road lane detection algorithm. In: 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, pp. 1–4. IEEE (2014)
6. Paszke, A., Chaurasia, A., Kim, S., et al.: ENet: a deep neural network architecture for real-time semantic segmentation. arXiv arXiv:1606.02147 (2016)
7. Cho, H., Kang, S.J., Cho, S.I., et al.: Image segmentation using linked mean-shift vectors and its implementation on GPU. IEEE Trans. Consum. Electron. **60**(4), 719–727 (2015)
8. Wei, J., Fu, J.: J. Commun. Univ. Chin. (Nat. Sci. Edn.) **27**(05), 72–78 (2020)
9. You, Y., Zhou, X.: Chin. Space Sci. Technol. **25**(3), 14–18 (2005)
10. Pan, X., Shi, J., Luo, P., et al.: Spatial as deep: spatial CNN for traffic scene understanding. In: Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, USA, vol. 32, no. 1, pp. 7276–7283 (2018)