Lane Line Detection based on Improved U-Net network

Yanqiang Li[*a], Shulin Zhang[a], Jing Ma[a,b], Yong Wang[a]
[a]Institute of automation, Qilu University of Technology (Shandong Academy of Sciences), Shandong Academy of Sciences, Jinan, Shandong China 250014
[b]Department Computer Science & Software Engineering, Auckland University of Technology, Auckland, New Zealand 1010
[*]Corresponding author: liyq@sdas.org

Abstract

The lane line detection and recognition are crucial research area for automatic driving. It aims at solving the problem of fuzzy feature expression and low time-sensitives of lane line detection based on semantic segmentation. This paper proposes to remove irrelevant background by dynamic programming region of interest while improving the lightweight neural network (U-Net). A group-by-group convolution and depth wise separable convolution in the backbone network are introduced, simplifies the branches of the backbone network, and atrous convolution is introduced into the enhanced path network with multi-level skip connection structure to retain the underlying coarse-grained semantic feature information. The full-scale skip connection fusion mechanism of the decoder is preserved, while capturing the fine-grained semantics and coarse-grained semantics of the feature map at full scale. The introduction of skip connections between the decoder and the encoder can enhance the lanes without increasing the size of the receptive field. The ability to extract line features and the ability to extract context improves the accuracy of lane lines. The experimental results show that the improved neural network can obtain good detection performance in complex lane lines, and effectively improve the accuracy and time-sensitives of lane lines.

Keywords: Driverless vehicle, Lane line recognition, Assisted driving, Image segmentation

1. Introduction

Driverless technology has gradually become a research hotspot in the field of modern automobile industry. In the process of unmanned detection of the surrounding road environment, accurate detection of lane lines is one of the basic requirements of intelligent driving. The detection and recognition of lane lines are often of great significance to the decision-making of automobiles. However, due to the wide variety of lanes in the real scene, different degrees of natural wear, and different degrees of illumination changes in different seasons, these problems affect the overall study of lane lines and bring difficulties to the detection and classification of lane lines. At present, lane line detection algorithms can be divided into two categories: traditional feature-based image segmentation methods and neural network-based image segmentation methods.

Based on the traditional image processing method, the recognition speed is fast, and the calculation efficiency is high. However, many parameters in the detection process need to be adjusted by the staff according to experience. Too much manual work, it is difficult to extract features on complex and unclearly labeled roads, and the robustness is relatively poor. Therefore, many studies have proposed improved methods for the traditional stages Tingting Gui [1] proposed an improved image grayscale method for image preprocessing in lane line recognition. Based on this method, lane line information can be accurately extracted on complex roads. Chen Yingfo [2] and others proposed to use the improved Hough transform to detect lane lines based on the detection operator to extract the lane edge to enhance the robustness of the lane line detection method. Based on traditional lane line detection, Haris [3]and others proposed to perform dynamic area planning for the region of interest to improve the robustness and real time performance of the algorithm.

In recent years, with the vigorous development of deep learning and neural networks and the significant improvement of computer performance, deep learning has made significant progress in the field of image processing, and detection methods based on deep learning have gradually become a research hotspot in the field of lane line recognition. Seokju Lee [5] proposed a unified end-to-end trainable multi-task network to solve the detection of vanishing points under severe weather conditions, which was transformed into the localization network and segmentation network. This staged and sub-task algorithm successfully improved the detection robustness. Awesome. Li Bing, Yan [4] proposed a multi-task multi-stage

hybrid cascade structure network, which alternately performs bounding box regression and mask prediction at each stage, enhances connection paths between adjacent mask branches, and provides mask branches. The information flow between them improves the performance of instance segmentation. Chen Wei Wei [6] proposed a new lightweight fully convolutional semantic segmentation algorithm (Seg Lane Net), which not only simplifies the parallel hollow convolution branches, but also adds a skip connection structure and deep feature fusion. The branch of the backbone network meets the real-time requirements of lane detection. Wang Z [7] compresses the network structure based on the classic lane line detection method model Lane net to improve the speed of lane line detection. Compared with traditional methods, deep learning-based methods have greatly improved accuracy and robustness, but there is still a low utilization rate of image semantic information. Further research is needed to solve the weak contextual connection of images and poor real-time performance.

Based on the above research background, this paper transforms the research on lane detection into an example segmentation problem of continuous slender regions, adopts dynamic division of regions of interest and then clipping pictures, and based on the U-net3+ network model, introduces deep separable convolution into the backbone network to improve the characteristic pyramid network, and introduces hole convolution into the enhancement path connected to the decoder to improve the context extraction ability of the enhancement path, Improve the environmental adaptability of the lane line of the network model [12]. Figure 1 shows the whole process.

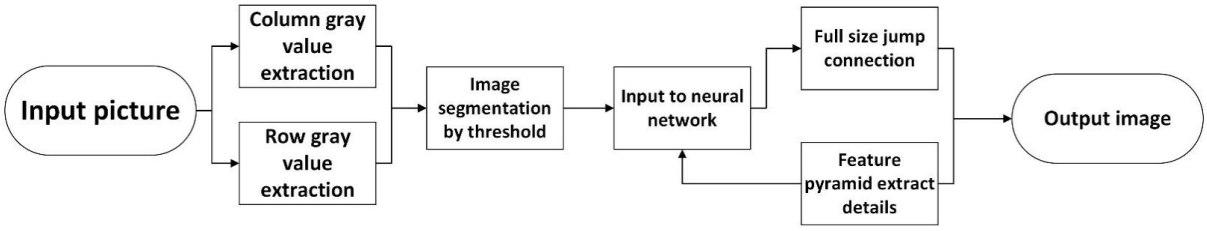

Figure 1. Overall structure

2. Network Structure

U-Net3+ is a lightweight semantic segmentation network proposed in 2020. The multi-stage hybrid cascade network of this network shows excellent segmentation performance in image segmentation tasks. As shown in Figure 1, the network performs feature extraction in the encoder stage, down sampling to reduce the feature map at each stage and establishes a full-scale skip connection to increase the ability of full-scale exploration of information. Hierarchical semantic information, which enhances the performance of instance segmentation, employs full-scale deep supervision in the decoding stage, producing a segmented side output from each decoder stage and supervised by the ground truth. In this paper, some modules and the overall network structure of U-Net3+ have been improved. The overall structure of the network mainly adopts a general decoding and encoding architecture. The morphological structure of the lane line is slender, so it is necessary to obtain the overall structure information of the lane line and the detailed information of the lane at the same time when retrieving the lane line.

The improved network structure of lane line detection is shown in Figure 2. The improved backbone network contains 5 incompletely repeated convolution modules [X1,X2,X3,X4,X5]. Each convolutional layer in the convolutional module uses a depth wise separable convolution to replace the original conventional 3×3 convolution and a 2×2 pooling layer with a stride of 2. The 5 convolutional layers extract and output 5 different sizes of feature map. The feature map continuously increases the receptive field and continuously enhances feature extraction to form a feature pyramid. This bottom-up network structure shortens the information transfer path so that the backbone network can contain as much lane detail information as possible. As shown in Figure 2, the depth wise separable convolution splits the spatial dimension and channel correlation, reducing the volume of the number of parameters required for the product improves the parameter usage efficiency of the convolution kernel.

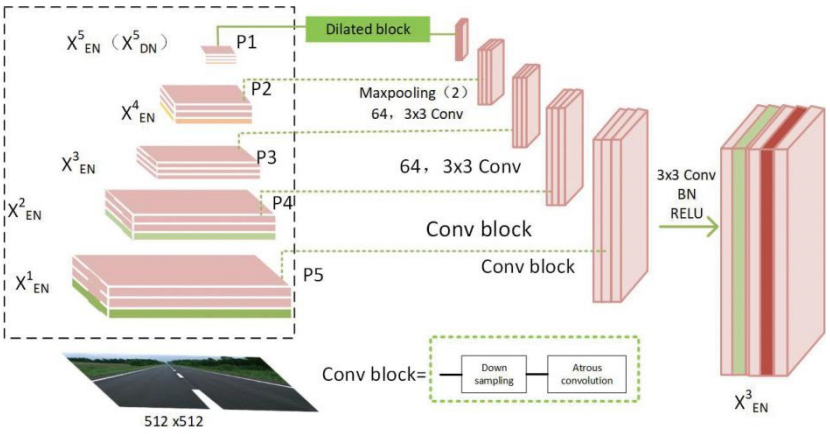

Figure 2. Improve structure of the network

The conventional information transmission in the original U-Net3+ neural network lacks the retention and fusion of target information. In response to this problem, Wei Wei [6] used Res2Net to fuse information between feature maps. The Res2Net structure is a 1×1 convolution, and the feature map is evenly divided into s number of feature map subsets, denoted by X, S. It is the control parameter of the scale size. When the number of input feature channels s is larger, the ability of the model to express multi-scale is stronger. Each X corresponds to a 3×3 convolution, and the output represented by y is collected in an incremental method k inside. The output fusion of different numbers and different receptive field sizes can improve the efficiency of feature processing.

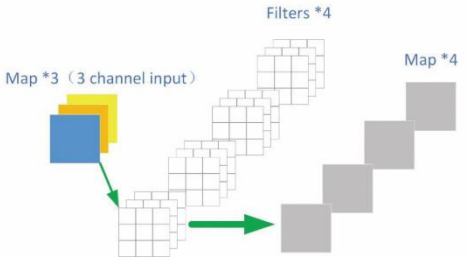

Figure 3. Depthwise separable convolution

The feature extraction network uses full-scale skip connections to build enhanced information paths, captures the full-scale information of feature maps, and provides fine-grained semantics and coarse-grained semantics for the decoding level. In the two coding modules X^1_{En} and X^2_{En}, the receptive field is small, which can better retain the fine-grained semantic information of the lane line. The part of the fusion of fine-grained semantic information adopts non-overlapping maximum pooling to detect the surrounding semantic information to supplement the information. completeness. In the X^3_{En} and X^4_{En} coding modules, larger receptive fields can provide coarse-grained semantic information. The backbone network filters different precisions step by step, and the feature method reduces the noise interference of complex background lanes and improves the lane line accuracy. In order to further make full use of the lane line detail information contained in the bottom layer of the feature pyramid, it is better to retain the low-level feature information. In this paper, the path extraction methods of X^1_{En} and X^2_{En} are enhanced, and atrous convolution is used to enhance the localization ability of these two modules to improve the detection lane line accuracy under the premise of saving parameters.

3. Branch Structure

In the conventional convolutional neural network, due to the huge number of parameters carried by the convolution calculation, it is easy to lead to the lack of time-sensitives of target detection. Therefore, the different information transmission paths [p1, p2, p3, p4, p5] of the feature pyramid network based on this paper are all composed of 3×3 atrous convolutions with a stride of 1. Increasing the receptive field of network features tends to increase the receptive range of the image and obtain more feature information at the semantic level and global level. This paper uses a small atrous convolutional layer to reduce operational parameters and alleviate the problems of accuracy degradation and loss of details. The formula for calculating the receptive field is as follows:

$$R_{i+1} = R_i + (\acute{k} - 1) \times S_i \tag{1}$$

R_{i+1} represents the receptive field of the current layer, R_i represents the receptive field of the upper layer，$\acute{k}$ represents the size of the equivalent convolution kernel , S_i represents the strided convolution of all previous layers except this layer. In addition, the equivalent calculation method of the equivalent convolution kernel of the atrous convolution is:

$$\acute{k} = k + (\mathrm{k} - 1) \times (\mathrm{d} - 1) \tag{2}$$

$\acute{k}$ indicates the size of the convolution kernel, indicates the number of holes. As shown in Figure 4, as the size of the convolution kernel, the size of the input feature map and the step size change, the size of the output feature map also changes, and the use of hole convolution and changing the number of holes does not change the receptive field, so this paper uses the hole volume. The product retains the details of the lane lines.

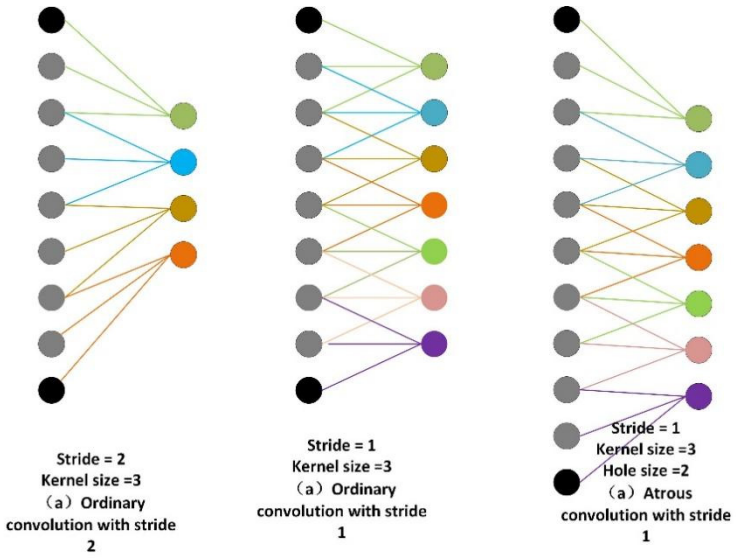

Figure 4. Atrous convolution 1D description

The feature pyramid network structure mainly includes three basic processes: feature generation of different dimensions from top to bottom, top-down supplement and enhancement, extraction of association table between each dimension of the backbone network layer and final output, and prediction output using different scale features. In lane line detection, as shown in Figure 3, the low-level feature layer mainly includes the contour range and geometric shape of lane line edge, and the high-level feature layer mainly includes the semantic information of lane line, which can better help lane line classification. Usually, the bottom-up information needs to undergo multiple convolution operations, which will also cause the information to disappear. For the detection of lane lines, it is necessary to obtain the global structure relationship and the precise details at the same time. With the same amount of calculation, hole convolution can provide a larger receptive field for the network, enhance the ability to extract spatial details, and improve the detection accuracy of lane lines.

$$X_{De}^{i} = \begin{cases} X_{En}^{i}, i = N \\ H\left\{\left[\underset{Scales: 1^{th} \sim i^{th}}{C\left(D\left(X_{En}^{k}\right)\right)_{k=1}^{i=1}, C\left(X_{En}^{i}\right)} \quad \underset{Scales: (i+1)^{th} \sim N^{th}}{C\left(u\, X_{De}^{k}\right)_{k=i+1}^{N}}\right], i = 1, \ldots, N-1\right\} \end{cases} \tag{3}$$

The decoder connects and receives the information of the backbone network through full-scale hopping, and preserves the same depth feature map of up sampling and deconvolution and information enhancement path to form a new feature map. As the above formula, X_{En}^{i} indicates the ith decoding module, i is the lower sampling layer along the coding direction, N indicates the number of decoders, Function C is the aggregation operation, function H is the feature aggregation mechanism, function d represents the up sampling operation, function u represents the down sampling operation

respectively, and the channel dimensions in brackets are fused.

In order to measure the lane detection performance of the network model, this paper uses the multi-scale structure similar loss function, including the end-to-end training of the network based on the cross entropy loss function, including calculating the target classification and giving higher weight to the fuzzy boundary. If the regional distribution of lane lines is different, the similarity index will be higher.

4. Pretreatment Stage

To verify the effectiveness of this research method, the lane line detection performance of the improved network model is evaluated. Based on the lane line detection performance, the mixed lane line data set based on Baidu is used for network training and testing. The mixed data set contains complex and diverse lighting conditions (strong light, shadow, night), various weather conditions (sunny, rainy, foggy), road conditions of different scenes (such as suburban, urban, rural, mountain roads, etc.), and different conditions of lane wear and occlusion. After the lane line is preprocessed, the data set is expanded by data enhancement method to enrich the diversity and complexity of the scene and the adaptability of lighting and improve the robustness of the algorithm. By turning the image left and right, it enriches the lack of image angle. By increasing or decreasing the brightness, it simulates the change of light intensity in the scene road scene and improves the ability of the network model to change the complex and changeable light intensity. By adding noise changes, the ability of anti-interference information of the network is improved. The lane line enhancement methods are as follows: horizontal turning, brightness enhancement, brightness reduction, and noise addition.

Figure 5. Data Enhancement

Generally, a lane line without background division has a large amount of background information. Dividing the region of interest can effectively segment the key information area and the lane line with less effective information. Through this background division, the range of the image to be processed can be limited and reduced to a great extent, and the running time of the algorithm can be greatly reduced. Generally, the method of dividing the region of interest is to cut off 1/3 of the image, After the lane line is binarized by dynamically dividing the region of interest, this paper detects whether the gray value of the row converted from 255 to 0 exceeds a certain threshold to judge whether there is irrelevant noise such as sky in the picture, and continues to calculate the gray value of the image column to eliminate the information such as scenery on both sides of the image, mathematical expression:

$$\begin{cases} image.at < uchar > \{i,j\} = 0 \\ image.at < uchar > \{i-1,j\} = 255 \\ return(i) \end{cases} \tag{4}$$

I and j represent the row coordinates and column coordinates of the image respectively image. Uchar represents the pixel gray value of {i, j} of image pixels. When the continuous pixels change and exceed the preset threshold, the picture boundary can be determined. As shown in Figure 6, the lane line image with irrelevant background is removed.

Figure 6. Grayscale image

5. Experiment and Result

After preprocessing, the images are learned through the network, and the average accuracy is used to evaluate the network performance to calculate the recall and accuracy of lane detection. AP refers to the area enclosed by the P-R curve with recall and accuracy as horizontal lines. The calculation formula for the combination of recall rate and accuracy rate:

$$\begin{cases} p = \dfrac{TP}{TP + FP} \\ R = \dfrac{TP}{TP + FN} \end{cases} \tag{5}$$

TP is the number of correctly detected positive samples, FP is the number of incorrectly detected samples, FN is the number of missed positive samples, when the IOU threshold range is 0.5 to 0.95, the AP value is calculated at an interval of 0.05s, and ap0.5 represents the AP value when the IOU threshold is 0.5 . To verify the lane detection performance of the improved network, the model structure with the best detection performance is determined. Based on a small range of data, the feature extraction training and testing of backbone networks with different convolution types are carried out. The experimental environment is Ubuntu 18.04, the GPU adopts NVIDIA's geforce RTX 2080, the deep learning framework adopts pytorch, and the integrated development environment uses pycharm to verify the lane detection performance of the network with different degrees of improvement and determine the best network model structure. When the deep separable convolution is used in the backbone network to replace the conventional convolution, the network model is trained and tested based on the constructed lane line. The comparison of lane line detection accuracy is shown in the figure. When the backbone network only uses two deep separable convolutions to replace the ordinary convolution network, the map decline is the least obvious.

Table 1 lane line detection results based on different feature extraction methods of backbone network

Backbone network			MAP	$AP_{0.5}$	$AP_{0.95}$
X2	X3	X4			
Conv	conv	conv	58.5	94.5	64.4
Conv	conv	Dep conv	58.2	94.3	64.2
Conv	Dep conv	Dep conv	57.6	94.0	64.1
Dep conv	Dep conv	Dep conv	56.1	93.5	63.1

In the improved network training process, the batch size is set to 1, the random gradient descent method is adopted, the weight attenuation factor is set to 0.005, the number of training iterations is 9*10000, the initial learning rate is set to 0.0035, and it decreases to 0.0025 when iterating 5000 to 8000 times, and to 0.001 after iterating 8000 times. The pre training weight file parameter in the data set is loaded as the initial value of the training weight of the network model. The loss function in the training process decreases gradually with the increase of the number of iterations, and finally stabilizes at 0.1. In order to verify the effectiveness of using void convolution instead of convolution for feature extraction in the

feature pyramid enhancement path, the comparison between ordinary convolution network enhancement and the introduction of void convolution is carried out. The comparison of lane line detection is shown in the table below. It can be seen that the precision of lane line detection is improved without increasing the network model parameters.

Table 2 test results of different enhancement paths

Enhanced path	$M/10^6$	$AP_{0.5}$	$AP_{0.75}$
Convolution	100.76	94.3	63.5
Void Convolution	100.76	94.5	64.1

In order to verify the effectiveness of using void convolution instead of conventional convolution in feature pyramid enhancement path, a comparative experiment between improved path enhancement strategy and ordinary path enhancement was carried out. In Table 2, M is the number of parameters used to measure the memory occupation of the network model. It can be seen from table 2 that in the same model structure, the introduction of hole convolution can improve the detection accuracy of lane lines without increasing the number of network model parameters.

Figure 7 shows the lane line detection results of the method in the test set. It can be seen that this method can effectively identify the category of lane lines, and has good detection performance under different lighting conditions and different photographing directions . Aiming at the problem of lane detection in complex environment, this paper proposes a lane detection method based on improved neural network instance segmentation. This method introduces deep separable convolution into the backbone network to reduce the parameters of the backbone network and improve the detection speed. Based on the feature pyramid network, a bottom-up feature transfer path is added, and a hole convolution is introduced into the enhancement network to improve the detection accuracy, so as to make up for the details of lane.

Figure 7. Lane image

6. Conclusions

According to the real-time requirements of lane detection, a hybrid cascade detection network based on U-net3+ is proposed in this paper. In this method, the images are put into the neural network for learning, the lane line problem is regarded as a segmentation problem, and the data set is supplemented by data enhancement to meet the training requirements and improve the generalization ability of the network model. Enhance the coarse-grained information of lane lines in the enhanced path. Experiments show that the improved lane detection network can be applied to various vehicle scenes, has good detection accuracy, is better than the original algorithm, and is more in line with the actual needs.

Acknowledgments

This paper has been supported by the research and demonstration application of key technologies of autonomous driving test and evaluation, and the project (introduction of innovation team) funded by the "20 colleges and universities" in Jinan (Award:2020GXRC029). Supported by the National Natural Science Foundation of China: project support for research on Intelligent Vehicle decision driven by driver's cognitive mechanism (Award:52131201). Sincerely thank the experts.

References

[1] Tingting Gui, Xing Gu, Yanli Shi. “Gray-scale Image Colorization based on Conditional Deep Convolution

Generation Adversarial Network[J].” International Core Journal of Engineering, 2021, 7(9).
[2] Chen Yingfo,Wong Pak Kin,Yang Zhi Xin. “A New Adaptive Region of Interest Extraction Method for Two-Lane Detection[J].” International Journal of Automotive Technology,2021, 22(6).
[3] Haris, Malik,Hou, Jin,Wang, Xiaomin. “Lane line detection and departure estimation in a complex environment by using an asymmetric kernel convolution algorithm[J].” The Visual Computer, 2022(prepublish).
[4] Li Bing,Yan Qiu-Rong,Wang Yi-Fan,Yang Yi-Bing,Wang Yu-Hao. “A binary sampling Res2net reconstruction network for single-pixel imaging.[J] .” The Review of scientific instruments, 2020, 91(3).
[5] Lee S , Kim J , Yoon J S , et al. “ VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition[J].” IEEE, 2017.
[6] Wei Wei, Xin Junchang. “ Curvelet-based Image Super-Resolution via Res2Net Multi-Scale Network[J].” THIRTEENTH INTERNATIONAL CONFERENCE ON DIGITAL IMAGE PROCESSING (ICDIP 2021), 2021, 11878.
[7] Wang Z , Ren W , Qiu Q . “LaneNet: Real-Time Lane Detection Networks for Autonomous Driving[J] .” 2018.
[8] ZHAO C L, MA X, ZHANG C T, et al., “RRT-connect Path Planning Algorithm based on Gravity Field Guidance.” Electronic Measurement Technology, 2021, 44(22):44-49.
[9] Chen Yu,Zheng Yunan,Xu Zhenyu,Tang Tianhang,Tang Zixin,Chen Jie,Liu Yiguang.“Cross-Domain Few-Shot Classification based on Lightweight Res2Net and Flexible GNN[J] .” Knowledge-Based Systems,2022,247.
[10] Zhang Zhongqiang,Liu Danhua,Gao Dahua,Shi Guangming. “A novel spectral-spatial multi-scale network for hyperspectral image classification with the Res2Net block[J].” International Journal of Remote Sensing,2022,43(3).
[11] Li Bing,Yan Qiu-Rong,Wang Yi-Fan,Yang Yi-Bing,Wang Yu-Hao. “A binary sampling Res2net reconstruction network for single-pixel imaging.[J].” The Review of scientific instruments,2020,91(3).
[12] Wei Wei,Xin Junchang. “Curvelet-based image super-resolution via Res2Net multi-scale network[P] .” Northeast Univ. (China);Nanyang Technological Univ. (Singapore);Gifu Univ. (Japan),2021.