IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 19, 2019, accepted October 11, 2019, date of publication October 15, 2019, date of current version October 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947574

Multi-Lane Detection Based on Deep Convolutional Neural Network

FAN CHAO, SONG YU-PEI, AND JIAO YA-JIE
College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
Corresponding author: Song Yu-Pei (yupei_song@126.com)

This work was supported in part by the Natural Science Research Program of Henan Provincial Department of Education under Grant 14A510019, and in part by the Natural Science Project of Henan Science and Technology Department under Grant 162300410062.

ABSTRACT In order to solve the problem of the poor robustness for extracting the multi-lane marking, a multi-lane detection algorithm based on deep convolutional neural network is proposed. Since the results of the feature extraction of lane boundary were affected by various factors, a deep convolutional neural network based on FCN network is used for lane boundary feature extraction, and the neural network can be used to classify the lane images at the pixel level. Then the network parameters are trained on the public Tusimple data set and evaluated at Caltech Lanes data set. The last part combines a linear model with a curved model to realize the establishment of the lane marking equation, Hough transform is used to determine the fit interval and the least square method is used to fit lane marking. The experimental results show that the average accuracy of the algorithm for identifying lane on the Tusimple data set is 98.74%, and the accuracy rate on the Caltech Lanes data set is 96.29%.

INDEX TERMS Multi-lane detection, fully convolutional networks, Hough transform, least squares method, perspective transformation.

I. INTRODUCTION

According to the survey, the great majority of road traffic accidents are caused by the driver's poor driving behavior. Therefore, the study of advanced driver assistance systems (ADASs) can significantly reduce the probability of traffic accidents, and lane detection is one of the important functions of the intelligent assisted driving system [1].

At present, the sensors used for the acquisition of road scenes mainly include cameras (monocular/binocular camera) and radars (millimeter wave, laser and ultrasonic radar). The most commonly used are monocular cameras and laser radar. Monocular cameras were used in [2]–[4] as a road information acquisition sensor, and binocular cameras were used in [5]–[8] generally combine the detection of parallax maps and vanishing points to achieve drivable road detection. In [9] three monocular cameras were used for road scene detection. Laser radar has been demonstrated its effectiveness in experimental tests [10], [11], the biggest advantage of it compared to the camera is small environmental constraints and can be used day or night, but it still has defects such as high cost, big volume and easy to be affected by rain and snow. Therefore, cameras are more suitable as sensors for road environment perception than laser radar so far.

In general, there are two main problems for Lane detection by using monocular camera in road images. The first one is the illumination variation which is often caused by the street lamps, vehicles headlights, taillights at night and the water of road surface after rain, etc. All these conditions will definitely affect the accurate identification of the lane markings. Additionally, it is also becoming more difficult to distinguish the lane boundary from the background under the bad weather, such as the strong gale, the heavy fog, and so on, which results in poor visibility and affecting the detection of the lane markings. Another problem is the diversity of the road surface. For example, lane borders are unclear, damaged, discontinuous or with poor contrast, in addition, the shadows of plants and fences on both sides of the road, or the vehicles on the road will also make the lane detection become a big challenge.

In the past few years, in order to solve above two problems, several algorithms have been proposed by researchers. Hough transform [12] is one of the most extensively lane detection algorithms because of its simplicity. However, the drawbacks of the Hough transform are its lots requirement of memory and calculating time, and due to the complexity of the actual

The associate editor coordinating the review of this manuscript and approving it for publication was Junhui Zhao.

road conditions, the algorithm based on Hough transform is difficult to be robust under various conditions because it is too much relies on the feature extraction image. Recently, Convolutional Neural Network (CNN) [13] has been adopted to enhance and extract lane points before lane fitting [14], which proves that the classification effect of CNN on objects is better than SVM. Another example is the CNN based on the LeNet structure was adopted in [9], to reduce the training time of CNN based on the LeNet, the Extreme Learning Machin (ELM) was inserted this structure in [15], which also significantly reduces the computing time of the algorithm. In recent years, with the significant help of computing speed increase on computer, deep convolutional neural networks have made breakthroughs in the field of computer vision. Visual Geometry Group 16 (VGG-16) is one of the most popular deep CNN, which participated in the 2014 ImageNet image classification and positioning challenge, achieved excellent results and was widely used for image classification. However, VGG-16 has a lower recognition rate for small objects such as lane line pixels, and the fully connected layer occupies 60% of the total parameters. After that, the Full Connected Network (FCN) based on VGG-16 was designed to realize the segmentation of the image [16], which replaced of the fully connected layer by the fully convolution layer. Since the jump structure of different depth layer feature maps was adopted to FCN network, the output of the FCN network was the same as the input size and realized end-to-end image segmentation. In this paper, FCN network was used for the extraction of feature points of the lane marking. Each pixel in the input image can be used as a separate classification experiment. The original full convolution layer was removed and the extracted feature map was directly restored. Network parameters were trained on the Tusimple data set. In order to achieve multi-lane fitting, the Hough transform was combined with the least squares method, the fitting interval is determined by Hough transform, and the lane line fitting is performed in the fitting interval by the least square method. The experimental results show that the lane boundary pixel extraction using FCN network is better than other neural networks, and the recognition rate of lane lines is also better than other algorithms.

The remainder of this paper is organized as follows: Section 2 described the related work in feature extraction and lane detection. Details of the proposed feature extraction net architecture and the algorithm of lane detection in Section 3. Section 4 provided experimental results and evaluate the performance of the proposed method on public lane detection data sets and the paper is finally concluded in Section 5.

II. RELATED WORK

In general, the process of lane detection consists of preprocessing, feature extraction and lane detection. Preprocessing helps to reduce calculating time and the noise point, which consists of smoothing operation, enhanced the grayscale image and edge detection. Mean filter [17], median filter [18] or Gaussian filter [19] is used to perform smoothing operation. To overcome different illumination change, Otsu's [20] algorithm of adaptive threshold is performed. Self clustering and fuzzy c-means algorithm are used in [21] to overcome the condition of glare.

A. FEATURE EXTRACTION

Edge detection is used to extract the lane boundaries. Sobel [22], Canny [23] is the two most used edge detection operators. After that, morphological [24] algorithm is used to further reduce the number of noise point. In addition, predict Region of Interest (ROI) [25] of the image is used on several papers to reduce redundant image data. In [26], the line where the vanishing point is located as the upper boundary line of the ROI, and the top line of the car hood as the upper boundary line of the ROI.

However, the lane feature extraction using the above algorithm requires artificially setting the parameters of the filter, the starting and ending areas of the ROI, and accurately calculating the position of the vanishing point. Under complex road conditions, the artificial parameters provided are difficult to have strong robustness. The latest and powerful method of lane feature extraction algorithm focuses on the classification of pixel points, rather than artificially setting parameters to filter noise points. CNN is one of the most powerful tools for object classification, which obtains the feather map through several convolution layers and pooling layers, and it shares the parameters between the hidden layers. The parameters of the convolution kernel confirmed through multiple iterations. In [27], CNN is designed for object classification and the framework of the algorithm is based on the LeNet. However, the recognition rate will be low for detecting small target objects such as lane pixels. In our paper, the FCN was used to realize pixel-level lane classification. Compared with the above enumerated algorithms, our proposed algorithm does not need to manually design some complicated constraints to extract lane boundary feature points, and to our surprise, the classification performance is better than the existing CNN networks. We made a test on the Tusimple dataset and the Caltech Lanes dataset, and the results show that the average accuracy of lane recognition is 98.74% and 92.29%, which are superior to other algorithms. So the algorithm will provide technical support for the design of intelligent assistant systems.

B. LANE FITTING

Lane fitting is used to obtain mathematical description of lane. Straight line is the most popular lane model for lane detection. Hough transform is used to detect straight lines. But its limitation is that can't present all lane, such as the lane marking in front of the turning car. So in this case, the curve model will be a better choice of lane detection algorithm. In order to fit the lanes more accurately, the parabolic model [28], the B-spline curve model [19], and the compound model which combined the linear model and the curved model [18] had proposed by relevant scholars to fit the lane curve. But, higher-order curve equations increase the

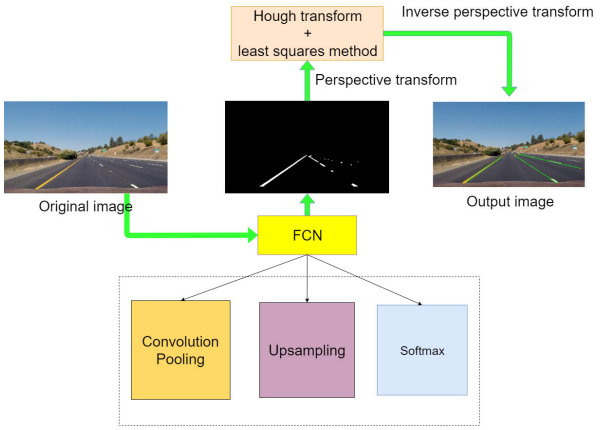

FIGURE 1. Overall framework of the algorithm.

calculating time. To solve the problem, Hough transform and the least square method combined with the perspective transformation is used as lane detection algorithm. The Hough transform and the perspective transformation are used to fitting the lane fitting area, then the least square method is used in the area to get the lane mathematical parameter.

III. PROPOSED METHOD

The structure of the method in this paper is shown in Fig.1. The lane image captured by the monocular camera is input into the FCN network. It will calculate by the convolution layer, the pooling layer, the deconvolution layer and the softmax classification layer in the neural network, and finally, the binary image including only the lane boundary points is output. Then, the output image of the neural network is perspective-transformed, and the lane fitting is performed on the perspective transformation map by Hough transform and the least square method. Finally, the fitted lane is marked on the original image after the inverse perspective-transformed.

A. DETAILS OF LANE FEATURE POINT EXTRACTION NETWORK BASED ON FCN

In order to realize high accuracy and robustness lane detection, we have modified the FCN network to classify the lane marking pixels. The structure of the network is shown in the Fig.2, which was divided into two parts: an encoder and a decoder. Three type network models of the encoder part were adopted according to the number of superimposed feature maps. The first model is the decoded result obtained from the feature map outputted by the 5th pooling layer. The second one combines the feature maps outputted by the 4th pooling layer and the 5th pooling layer. The third network combines the feature maps outputted by the 3th pooling layer, 4th pooling layer and the 5th pooling layer. Under the third decoder network structure, the entire network contains 14 convolution layers, 5 pooling layers, 3 deconvolution layers (deconv1, deconv2, deconv3) and one Softmax classification layer. The convolution kernel size of the two convolutional layers (conv6, conv7) of the decoder portion is 1×1. Compared with the original FCN network structure, the full convolution layer behind the fifth pooling layer is removed, and the feature map extracted by the encoder partial convolution layer is directly restored. The FCN network performs parameter training on the Tusimple dataset. In the training phase, the original image and the label image are input into the network, which will output a predict image by the initialized parameters. The loss function was used to calculate the difference between the label image and the predict image, and then the back propagation was used to update the network parameters. This process will stop when the loss function reaches a minimum value and the network parameters will be saved. During the test phase, the network only needs to input an original image and output a predict image.

B. FEATURE MAP CALULATION AND REDUCTION

VGG-16 is a deep neural network. The local feature of the image is acquired by the shallow convolution layers, such as color, texture, line segments and so on. More global feature can be learned by the deeper convolution layer, such as a face, a complete car, and so on. In order to confirm the feasibility and effectiveness of the FCN network for lane marking feature extraction, Fig.4, 5 show the partial layer feature map visualization of the trained neural network. The feature map channel numbers increase and the size becomes smaller as the depth of the convolution layer. In order to clearly understand what each layer of the neural network is doing, due to the limited space, we show nine channels of each layer's feature map. All the channels of the feature map of the display layer are superimposed to obtain a fusion map.

The input to the FCN network was a lane marking image captured by a monocular camera. Then, it will be calculated by the first stage of two convolution layers (conv1_1, conv1_2) and one pooling layer (pool1). The convolution kernel size is 3×3, the stride is 1 and the number of neurons in each convolution layer is 64. The activation function of the convolution layer is ReLU. Each convolution kernel is equivalent to a receptive field, and the trained convolution kernel can extract the information of the feature points of the lane marking in its receptive field. In order to reduce the amount of calculation and overfitting, the max-pooling layer is used as pool1 and the kernel size is 2×2. Thus the size of feature map is reduced doubled when passed through the max-pooling layer. It can be seen from the output of the pool1 in the Fig.4 that the feature maps obtained by different convolution kernels are quite different. Some convolution kernels acquire color information, while others extract edge information from different objects. From the fusion map output from the pool1 layer in Fig.5, it can be seen that the edge information of various objects in the input image are retained such as cars, lane boundaries, trees, etc. In order to further extract lane boundary feature points and reduce noise points, perform the next set of convolution calculations. Then the output from the first stage is used as the input of the second stage. The difference from the first stage is that

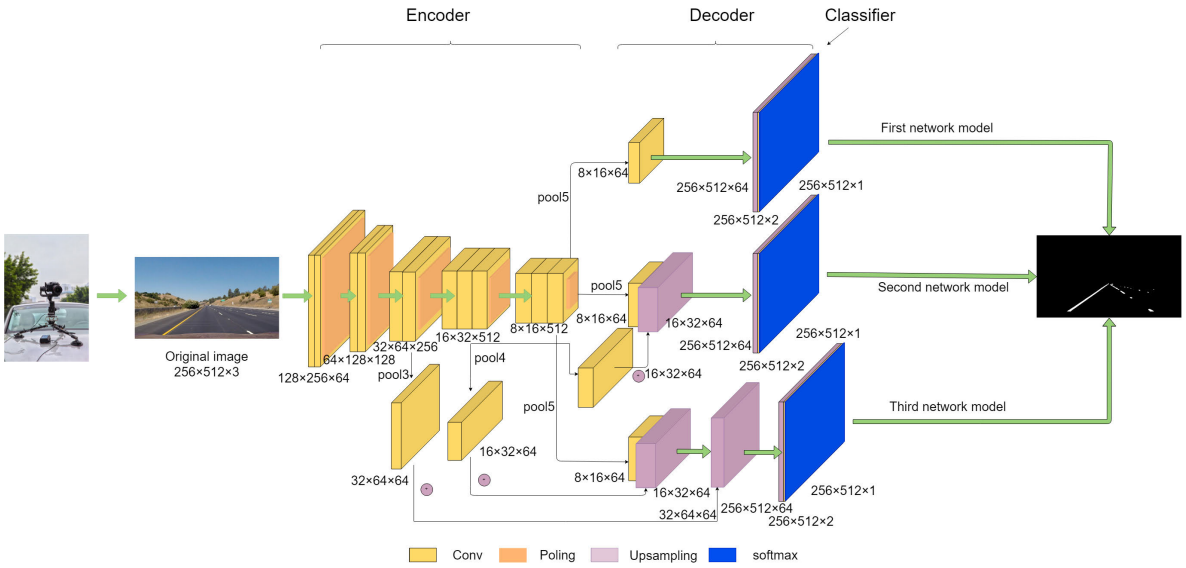

FIGURE 2. Overall framework of the FCN.

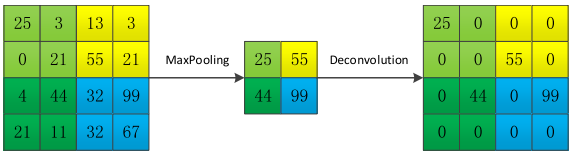

FIGURE 3. Principle of deconvolution layer.

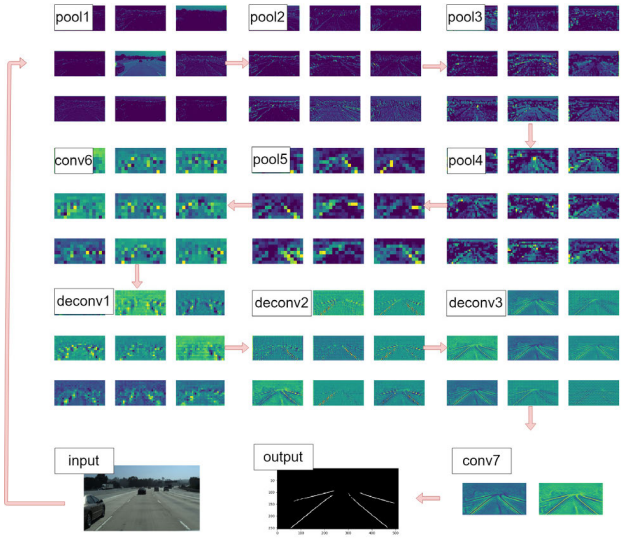

FIGURE 4. Visualization of part layers.

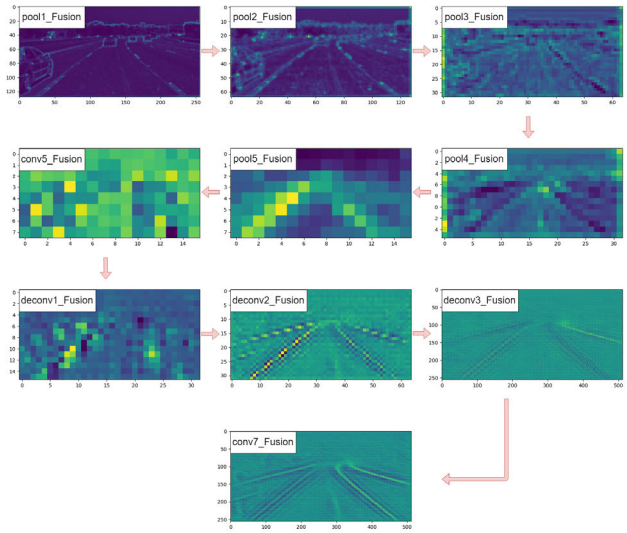

FIGURE 5. Visualization of part layers after fusion.

the number of channels in the convolutional layer becomes 128. It can be seen from the output of the pool2 in the Fig.4 that the feature map extracted in the second stage is similar to the first stage. However, due to the compression of the pooling layer, the resolution of the image is reduced, and the feature extraction condition cannot be visually observed. Then continue the calculations for the third, fourth and fifth phases. The number of channels in the convolutional layer is changed to 256, 512, and 512. In addition, the convolutional layer is increased from 2 layers to 3 layers in the fourth and fifth stages. As can be seen from the fusion map in Fig.5, the number of edge pixels of the car slowly disappears, and the lane boundary pixels are getting brighter and brighter.

After passing through the convolution layers and polling layers, the feature image was extracted from the original lane image by the neural network, but which is the low-resolution image with high-dimensional as we can see from Fig.4 pool5. For example, after the fifth merged layer, the size of the output feature map is 8×16, which is 1/32 of the original image. In order to display the extracted feature image, the deconvolution layer is introduced behind the encoder. The decoder portion mainly includes the deconvolution layers, the convolution layers, and a softmax layer. The feature image size is restored by the 3 deconvolution layers, it was multiple by calculation of the each deconvolution layer, and the

feature map dimension was reduced by the convolution layers. The convolution kernel size of the decoder convolution layer changes from 3 × 3 to 1 × 1, and the main function of the layer is to reduce the dimension of the feature map. The implementation principle of the deconvolution layer is shown in Fig.3. After passing through each max-pooling layer, the maximum index value is retained in the convolution kernel area. When the deconvolution operation is performed, the maximum index value is restored and the values of other regions are set to 0. The output of the convolutional network is the result of the segmentation of an image. The output of Conv7_fusion in Fig.5 obtains the result of restoring the feature maps outputted by the fifth, fourth, and third pooling layers. The range of values for each pixel is 0-1, indicating whether the neural network predicts is a lane boundary point. In order to realize the classification of each pixel in the original image, the softmax layer is added at the end of the network to determine the label value of each pixel of the feature map, and the value of the predicted line pixel is set to 1, otherwise, it is set to 0. Finally, a 256 × 512 × 1 binary map is obtained.

C. DEFINING THE LOSS FUNCTION

In order to improve the classification accuracy of FCN network for the lane images, it is necessary to obtain an optimal network parameters through training. Thus, an evaluation criterion is defined, and after several iterations, the network parameters are continuously updated until they are close to the optimal parameters. Usually the evaluation function is defined by the distance between the predicted output value of the model and the actual value. So the cost function of the FCN network is defined as follows:

$$\mathrm{L}(\tilde{y}_i, \mathbf{W}_j) = \frac{1}{N_{class}} \sum_i L_{class}(\tilde{y}_i, y_i) + \lambda \sum_{j=1}^{k} L_{tensor}(\mathbf{W}_j) \quad (1)$$

The cost function consists of two parts, the first part is the classification error and the second part is the tensor error of $\mathbf{W}$. Where $\tilde{y}_i$ is the predicted value of the model for the output result; y_i is the value of the real result; N_{class} represents the number of points whose values are labeled as 1 on the ith original image; λ is the balance weight of these two errors, and it takes as 0.001 in the test; k represents the layer number of the convolution kernel contained in the model; $\mathbf{W}_j$ is the parameter value of the jth convolution layer. The cross entropy function of the classification error L_{class} is used to calculate the cost value:

$$L_{class}(\tilde{y}_i, y_i) = -\log[\tilde{y}_i \times y_i + (1 - \tilde{y}_i)(1 - \tilde{y}_i)] \quad (2)$$

Because the structure of the FCN network is complex, and there are many layers and parameters, to prevent the model from over-fitting due to numerous parameters, the method of optimizing regular parameter is used.

$$L_{tensor}(\mathbf{W}_j) = \frac{1}{2} \sum_p (\mathbf{W}_j)^2 \quad (3)$$

where p is the number of elements in $\mathbf{W}_j$.

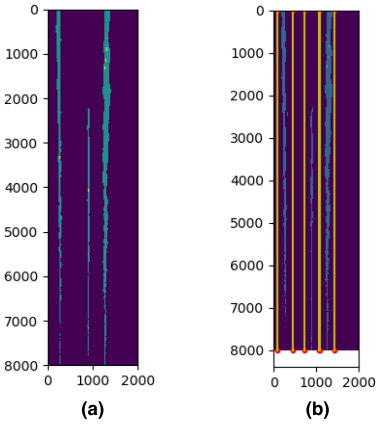

FIGURE 6. Determination of the fitting interval: (a) Result of perspective transformation; (b) Fit interval determined by Hough transform.

D. MULTI-LANE FITTING BASED IN HOUGH TRANSFORM AND THE LEAST SQUARE METHOD

After the network is constructed and trained, a binary map which just containing the lane pixels is output when the original lane image was input into the neural network. In order to detect the lane and obtain the equation of the lane the parameters of the curve equation are fitted through the lane model, which means that it is necessary to fit the left and right lanes separately for the images just containing two lanes.

But in many cases, three or more lane markings are captured to achieve the multi-lane detection, the perspective transformation is first done for the output binary image of the FCN network. To realize the perspective transformation, a 3 × 3 matrix $\mathbf{P}$ is defined, and the output binary map of the neural network is denoted as $\mathbf{S}$, then the transformation can be expressed as $\mathbf{S}_{out} = \mathbf{SP}$, transformed image is shown in Fig.6 (a), in which all the lane are parallel to each other. Then, the perspective image is further processed by the Hough transform, and the number of lane is determined based on extreme number at Hough space. Based on the abscissa of the midpoint of each straight line, we have two basic line by offset k pixels left and right, then the fit areal of each line is determined, as shown in Fig.6 (b). Then, the curve line is fitted by the least square method in the fitting areal to obtain the fitting equation of each lane. Finally, the fitting result obtained by inversely perspective transformation: $\mathbf{S}_{final} = \mathbf{S}'_{out}\mathbf{P}^{-1}$.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. DATA SET

The Tusimple dataset is used as the network training dataset, which contains 72,520 training images and 55,640 testing images. Our algorithm was also tested in the Caltech Lanes dataset, which is a classic lane dataset from the California Institute of Technology that included four shots around the streets of Pasadena, California at different times of the day, it has 1225 frames of images.

B. EXPERIMENTAL SETUP

The operating environment of the computer is Windows 10, software configuration is Python 3.6.8 and TensorFlow 1.13.1, and the hardware configuration of GPU is GTX-1070.

In order to reduce the training time and make the model fast and stable, some parameters of VGG-16 trained on the ImageNet data set are loaded as initial parameters. In the course of the training process, the original image and the manually labeled image simultaneously input into the model.

In the case of backpropagation, the momentum algorithm is used to update the network parameters, which is based on the stochastic gradient descent algorithm, but the oscillation of the algorithm can be suppressed effectively.

$$v_t = \gamma v_{t-1} + \alpha \nabla_W J(\mathbf{W}) \tag{4}$$

$$\mathbf{W} = \mathbf{W} - v_t \tag{5}$$

where v_t represents the update velocity of the parameter, who is positive when the value of the parameter is increasing, otherwise, it is negative.v_{t-1} is the changing value of the previous iteration, γ is a fixed parameter of 0.9, α is the learning rate whose initial value is 0.0005, and $\nabla_W J(\mathbf{W})$ is the reciprocal of the parameter $\mathbf{W}$ with respect to the function $J(\mathbf{W})$. Compared with the stochastic gradient descent algorithm, a term γv_{t-1} is added. Once the gradient maintains one direction for a long time (positive or negative), the learning rate α is increased to increase the update range of $\mathbf{W}$, and vice versa, the parameter update range is to be reduced. Thus, the convergence can be speed up significantly and the shock is reduced largely.

C. THE EFECT OF THE NUMBER OF POINTS ON LANE DETECTION

The effect of the number of fitting points on the lane fitting result during the lane fitting process is showed in Fig.7. From Fig.7 (a-c) more than 1000 points, 50-100 points and 20 points are selected as the fitting points respectively. By the effect of perspective transformation on the binary image, the number of lane pixels increase and the lines become thicker. Based on the fitting results, it is easy to find that the fitted curve can't represent the lane perfectly when the candidate points of the lane are selected too much. So the more points that model selected, the more easily over fitted will get. In addition, due to the number of selected pixels is large, it will also result in a large amount of computation and a large memory footprint. In order to avoid over fitting, 20 points are uniformly selected from the top to the bottom in the range of the image as the lane pixel, the fitting result is better and the speed is faster, which is shown in Fig.7 (c).

D. INFLUENCE OF THE THREE KINDS OF NETWORK STRUCTURES ON FEATURE EXTRACTION AND FITTING

Fig.8 (a-c) are the feature extraction results using three network structures for the same lane image. The first network only restores the feature map after the 5fth pooling layer, the second network merges the feature maps after the 5fth and 4th pooling layers, and the third network combines the output feature map after the 5th, 4th, and 3rd pooling layers. Comparing the output results of the same image through three networks, it can be seen that the lane pixels can be preserved completely in the feature maps for all these three networks. However, it can be clearly seen from the Fig.8 (a) that in addition to the lane pixels, there are more noise points, and the output of the neural network appears to be over-fitting. The difference between the feature maps of the networks 2 and 3 is small and almost no noise points are contained.

Fig.9 (b-d) shows the fitting results for the same input image based on the three network structures. It can be seen that three lanes can be identified correctly by different network structures. However, for the first network structure, only the feature map of the 5th pooling layer is restored, a large number of error points are contained in the output. The second network structure combines the feature maps of the 4th and 5th pooling layers, which significantly reduces the erroneous results in Fig.9 (b). However, compared with

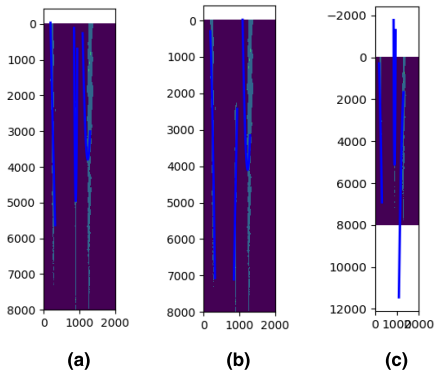

FIGURE 7. The effect of the number of fit points on the fit results: (a) Select more than 1000 points;(b) Select 50-100 points; (c) Select 20 points.

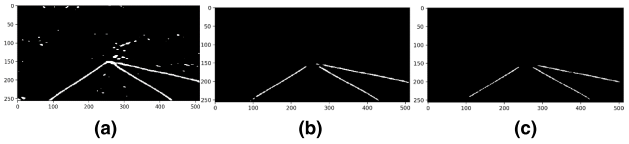

FIGURE 8. Output of the neural network: (a) First network output; (b) Second network output; (c) Third network output.

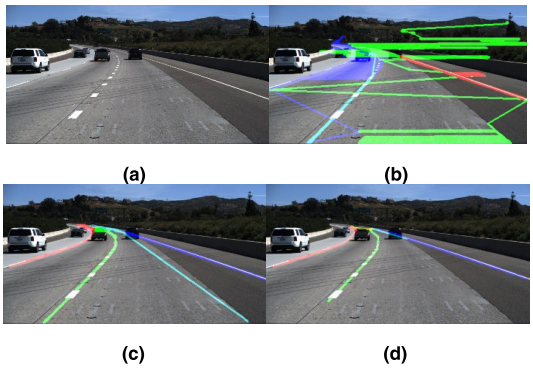

FIGURE 9. Detection results of the same image by three networks: (a) Original image; (b) Fit result of the first network; (c) Fit result of the second network; (d) Fit result of the third network.

TABLE 1. Fitting results of three networks in the Tusimple test set.

Model	Misidentification or missing identification number	Number of lanes	Accuracy
Network 1	Ignore	9624	Ignore
Network 2	199	9624	97.93%
Network 3	121	9624	98.74%

TABLE 2. Detection result comparison on tusimple dataset.

Algorithm	Precision	Recall	F1 sore
CNN+RNN [9]	0.61	0.52	0.56
Fast R-CNN [33]	0.7	0.5	0.58
Faster R-CNN [34]	0.73	0.54	0.62
ELCNN [15]	0.81	0.6	0.69
Our approach	0.89	0.66	0.76

the third network structure, the fitting result of the second network structure has one more output than the fitting result of the third network structure, and which obviously is a wrong output in Fig.9 (c). Because of the feature maps are merged multiple times after different pooling layers, the pixel value of the lane point is enhanced, and the contrast between the lane boundary and the background is improved. Thus, the pixel of the lane can be distinguished easily for the classifier. The superior comparison between network 2 and network 3 can also be derived from Table 1.

The results were tested by randomly selecting 2,617 images from the Tusimple dataset. The detection accuracy of lane by using three different network structures is shown in Table 1. Since the first type of network structure has too many noise points, the detection accuracy is not counted. The test results are compared between the second and third network structure in Table 1. It can be seen from the table that the average accuracy of the third network structure obtained by the feature maps generated by the 3rd, 4th, and 5th pooling layers is 98.74%, which is better than that of the second network structure.

The experiments have shown that, the fitting results will not be improved much when the feature maps of the first and second pooling layer are added, at the same time, with the adding of the pooling layer, the computational cost of the network will increase much.

E. EVALUATION ON THE TUSIMPLE DATA SET AND CALTECH DATA SET

The evaluation criteria of lane boundary point classification are F1 score, which is calculated by precision and recall. Precision is the proportion of all of the examples detected that are positive ground truth, and recall is the proportion of positive ground truth images detected in the sample.

Table 2 shows the lane extraction performance of our approach compared with other networks. ELCNN won second place, because the extreme learning machine (ELM) was embedded in the CNN structure, which significantly reduced training time and possessed smaller loss values. Our approach obtained the best result by using the structure of FCN, which has the ability to achieve pixel level classification.

TABLE 3. Test results of the algorithm in the Caltech dataset.

Road condition	Number of images	Number of lanes	Number of errors or missed measurements	Accuracy
cordova1	234	788	25	97.83%
cordova2	366	848	23	98.29%
washington1	330	1191	179	95.03%
washington2	222	882	88	94.02%
Overall	1152	3709	350	96.29%

TABLE 4. Test results of the algorithm in the Caltech dataset.

Method	Accurancy
Perspective transformation based detection(Multi-lanes)[30]	90.89%
straight line based detection (Double lanes)[31]	90.00%
Vanishing point based detection(Double lanes)[32]	95.90%
Our apporach (Multi-lanes)	96.29%

Table 3 shows the lane recognition results for the third network structure under the untrained Caltech Lanes dataset. The Caltech Lanes dataset consist of four collections that correspond to four different road conditions. Under these four road conditions, since the dataset cordova1 and cordova2 contain the highway roads and good weather mostly, the difference between them with the Caltech Lanes dataset is small, which results higher detection accuracy. But for the other two sets of washington1 and washington2, multiple lanes are contained and the roadway line shadows are severely occluded. The algorithm is not particularly good in the two sets because it is quite different from the Tusimple dataset.

Table 4 gives the results of the average accuracy of the algorithm in the Caltech Lanes dataset compared to other algorithms. The author of the perspective-based algorithm created the Caltech Lanes dataset, which has a recognition accuracy of 90.89% under multi-lane conditions and 96.34% accuracy under two-lane conditions. The algorithm based on the vanishing point is more accurate than two above algorithms, but which is more suitable for the recognition of two lanes, and the recognition rate is still lower than us. The algorithm we proposed can achieve multi-lane detection, and the accuracy is 96.29%, which is higher than the algorithm based on the vanishing point. The three algorithms of comparison do not use the neural network for feature extraction. Therefore, under complex road conditions, the result of lane detection is easy to receive interference from noise points. So, the algorithm proposed in this paper is superior to the other three algorithms. Also, as a training set in the Tusimple dataset, there is almost no large amount of shadow occlusion in the washington1 collection. Therefore, if more different types of lane training sets are input to the algorithm model, the performance of the convolutional neural network will be more excellent.

F. PART OF THE TEST RESULTS OF THE ALGORITHM UNDER THE DATA SET

Fig.10 shows the partial detection results of lanes for the Tusimple test set and the Caltech Lanes dataset. Due to space limitations, several representative images were selected as the

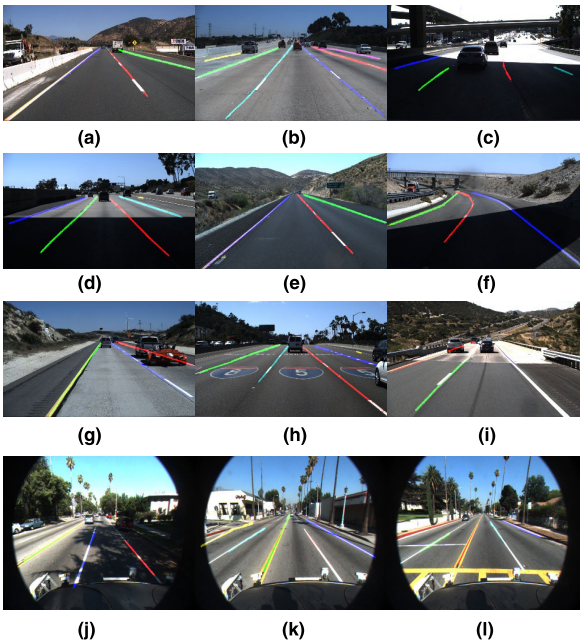

FIGURE 10. Fit result map of partial lane image: (a) Fence interference; (b) Multi-lanes; (c) Shadow; (d) High intensity of illumination; (e) Color interference; (f) Shadow; (g) vehicle occlusion; (h) Interference with road signs; (i) High intensity of illumination; (j) Shadow; (k) Multi-lanes; (l) Interference with road signs.

results. Fig.10 (a-i) show the detection results in the Tusimple dataset, and Fig.10 (j-l) give the results in the Caltech Lanes dataset. From the detection results, the proposed method can accurately identify the lane under the conditions of fence interference, multi-lane, shadow occlusion, strong illumination, color interference, road marker interference, and vehicle occlusion.

V. CONCLUSION

1) In this paper, a deep convolutional neural network based on FCN network is proposed. The FCN network can directly classify lane images and achieve end-to-end detection of multi-lane images.

2) The fitting of multi-lanes is achieved by a combination of Hough variation and the least square method.

3) The experimental results show that the accuracy of the proposed algorithm on the Tusimple dataset is 98.74%, and the accuracy of the Caltech Lanes dataset is 96.29%, which is a kind of robustness lane detection algorithm.

REFERENCES

[1] S. P. Narote, P. N. Bhujbal, A. S. Narote, and D. M. Dhane, "A review of recent advances in lane detection and departure warning system," *Pattern Recognit.*, vol. 73, pp. 216–234, Jan. 2018.

[2] D. C. Andrade, F. Bueno, F. R. Franco, and R. A. Silva, "A novel strategy for road lane detection and tracking based on a vehicle's forward monocular camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1497–1570, Apr. 2019.

[3] X. Zhang and X. Zhu, "Autonomous path tracking control of intelligent electric vehicles based on lane detection and optimal preview method," *Expert Syst. Appl.*, vol. 121, pp. 38–48, May 2019.

[4] C.-B. Wu, L.-H. Wang, and K.-C. Wang, "Ultra-low complexity block-based lane detection and departure warning system," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 582–593, Feb. 2019.

[5] Y. Su, Y. Zhang, T. Lu, J. Yang, and H. Kong, "Vanishing point constrained lane detection with a stereo camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2739–2744, Aug. 2018.

[6] 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.

[7] R. Fan and N. Dahnoun, "Real-time stereo vision-based lane detection system," *Meas. Sci. Technol*, vol. 29, no. 7, 2018, Art. no. 074005.

[8] C. Yuan, H. Chen, J. Liu, D. Zhu, and Y. Xu, "Robust lane detection for complicated road environment based on normal map," *IEEE Access*, vol. 6, pp. 49679–49689, 2018.

[9] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 690–703, Mar. 2017.

[10] X. Liu, Z. Deng, and G. Yang, "Drivable road detection based on dilated FPN with feature aggregation," in *Proc. IEEE 29th Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2017, pp. 1128–1134.

[11] S. Gu, T. Lu, Y. Zhang, J. M. Alvarez, J. Yang, and H. Kong, "3-D LiDAR + monocular camera: An inverse-depth-induced fusion framework for urban road detection," *IEEE Trans. Intell. Veh.*, vol. 3, no. 3, pp. 351–360, Sep. 2018.

[12] V. Gaikwad and S. Lokhande, "Lane departure identification for advanced driver assistance," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 910–918, Apr. 2015.

[13] G. Huang, X. G. Wang, W. Q. Wu, H, Zhou, and Y. Y. Wu, "Real-time lane-vehicle detection and tracking system," in *Proc. Chin. Control Decis. Conf. (CCDC)*, May 2016, pp. 4438–4443.

[14] C. Hou, J. Hou, and C. C. Yu, "An efficient lane markings detection and tracking method based on vanishing point constraints," in *Proc. 35th Chin. Control Conf. (CCC)*, Jul. 2016, pp. 6999–7004.

[15] 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.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[17] X. J. An, E. Shang, J. Z. Song, J. Li, and H. G. He, "Real-time lane departure warning system based on a single FPGA," *EURASIP J. Image Vide Process.*, vol. 38, pp. 1–18, Dec. 2013.

[18] K. Zhao, M. Meuter, C. Nunn, D. Müller, S. Müller-Schneiders, and J. Pauli, "A novel multi-lane detection and tracking system," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 1084–1089.

[19] P.-Y. Hsiao, C.-W. Yeh, S.-S. Huang, and L.-C. Fu, "A portable vision-based real-time lane departure warning system: Day and night," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 2089–2094, May 2009.

[20] H. J. Xu and H. F. Li, "Study on a robust approach of lane departure warning algorithm," in *Proc. 2nd Int. Conf. Signal Process. Syst.*, Jul. 2010, pp. 201–204.

[21] J. G. Wang, C.-J. Lin, and S.-M. Chen, "Applying fuzzy method to vision-based lane detection and departure warning system," *Expert Syst. Appl.*, vol. 37, pp. 113–126, Jan. 2010.

[22] C. Mu and X. Ma, "Lane detection based on object segmentation and piecewise fitting," *TELKOMNIKA Indonesian J. Elect. Eng.*, vol. 12, no. 5, pp. 3491–3500, May 2014.

[23] P.-C. Wu, C.-Y. Chang, and C. H. Lin, "Lane-mark extraction for automobiles under complex conditions," *Pattern Recognit.*, vol. 47, no. 8, pp. 2756–2767, Aug. 2014.

[24] C.-J. Lin, J.-G. Wang, S.-M. Chen, and C.-Y. Lee, "Design of a lane detection and departure warning system using functional-link-based neuro-fuzzy networks," in *Proc. Int. Conf. Fuzzy Syst.*, Jul. 2010, pp. 1–7.

[25] C. Lee and J.-H. Moon, "Robust lane detection and tracking for real-time applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 4043–4048, Dec. 2018.

[26] H.-Y. Cheng, C.-C. Yu, C.-L. Lin, H.-C. Shih, and C.-W. Kuo, "Ego-lane position identification with event warning applications," *IEEE Access*, vol. 7, pp. 14378–14386, 2019.

[27] G. L. Oliveira, C. Bollen, W. Burgard, and T. Brox, "Efficient and robust deep networks for semantic segmentation," *Int. J. Robot. Res.*, vol. 37, pp. 472–491, Apr. 2018.

[28] J. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations (ICLR)*, Sep. 2014, pp. 1–14.
[30] M. Aly, "Real time detection of lane markers in urban streets," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 7–12.
[31] T. M. Hoang, H. Hong, H. Vokhidov, and K. R. Park, "Road lane detection by discriminating dashed and solid road lanes using a visible light camera sensor," *Sensors*, vol. 16, no. 8, p. 1313, 2016.
[32] A. Irshad, A. A. Khan, I. Yunus, and F. Shafait, "Real-time lane departure warning system on a lower resource platform," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov./Dec. 2017, pp. 1–8.
[33] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
[34] S. Q. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

FAN CHAO was born in Xinyang, Henan, China, in 1976. He received the B.S. and M.S. degrees in computer science from the Xi'an University of Technology, Shanxi, in 1999 and 2004, respectively, and the Ph.D. degree in computer engineering from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Shanxi, in 2007.

Since 2007, he has been an Assistant Professor with the College of Information Science and Engineering, Henan University of Technology, Henan. He is the author of two books, more than 40 articles, and eight national inventions. His research interests include information processing, statistical analysis and grain yield prediction, and so on.

SONG YU-PEI was born in Linzhou, Henan, China, in 1994. He received the B.S. degree in electronic information engineering from the Henan University of Technology, in 2017, where he is currently pursuing the master's degree in information processing. His research interests include image processing, machine learning, lane detection, and so on.

JIAO YA-JIE was born in Zhengzhou, Henan, China, in 1994. She received the B.S. degree in electronic information engineering from the Henan University of Technology, in 2017, where she is currently pursuing the master's degree in information processing. Her research interests include the information processing and algorithm analysis.