SLDNet: A Branched, Spatio-Temporal Convolution Neural Network for Detecting Solid Line Driving Behavior

Zhanke Zhou*, Ruiqi Li*, Yayu Gao(✉), Chengwei Zhang and Xiaojun Hei
Huazhong University of Science and Technology, Wuhan, China
{zhankezhou, rlee, yayugao, zhangcw, heixj}@hust.edu.cn

***Abstract*—Solid line driving is known as one of the major driving violations in China. In our work, we proposed a branched, spatio-temporal convolution neural network (SLDNet) to recognize these violations acts from photographs captured by surveillance camera and trained it on Pingxiang solid-line-driving dataset. Such neural network can achieve performance better than human in both metrics of accuracy and recall, and will be put into use in near future.**

I. Introduction

A vast number of traffic accidents happened every year in China. To maintain road traffic order and prevent these traffic accidents, the law of China on road traffic safety has specified a number of forbidden driving behaviors of *vehicles*, one of which is known as solid line driving. In the past few years, more and more surveillance cameras are set up in city transportation systems, to monitor road traffic status. These cameras capture images of vehicles with suspicious violation, and send these images to the traffic control center for policemen to review, while in practice it is hard to review these suspicious violation image manually. Thus it is necessary to design a method to recognize solid line driving violation from image information captured by camera. We design a branched deep learning network integrating different neural network architectures to solve this question, and achieve 91% in accuracy with 92% in recall. Our algorithm will practically come into service in near future. The complete literature review of this is available on [2].

II. Problem Description

Discussion above indicates that it is necessary to design an accurate and steady method to decide whether an image captured by surveillance camera contained solid line driving vehicle or not, and our work use deep learning method to design such an algorithm. *Fig. 1* shows a typical positive labeled image in Pingxiang solid-line-driving dataset, captured by surveillance cameras of Pingxiang city, Jiangxi province. An image in dataset contains four photographs. The photo in bottom right is a zoom-in of suspicious vehicle, and the rest three photo form a time sequence. To our knowledge, the general idea for solving the problem is that the location of vehicle and lane line should be detected. Specifically, we addressed the problem by determining which pixels composed vehicle and which pixel composed lane line in the image. Combining the location information of both vehicle and lane line, it could be inferred that whether the vehicle is violating or not.

Fig. 1 Original image from Dataset

III. The Algorithms

In this section, we first present the overview of SLDNet and then show more details about each part of pipeline.

A. Pipeline Overview

We design a two-part, branched pipeline to determine whether an original image contains solid line driving behavior or not. As illustrated in *Fig. 2* the front-end of pipeline consisting of two branches, is devised to detect lane lines and vehicles located in given input images and output the segmentation map of these lane lines and vehicles. As to the back-end part of pipeline, it takes the segmentation map as input, and output the ultimate judging result, working as a binary classifier.

In a nutshell, our main idea is to eliminate noise and lower the complexity of traffic intersection images via keeping lane lines and vehicles and getting rid of the rest information, and then proceed binary classification on images of segmentation.

B. Pipeline Front-end

1) Lane Line Dection

As for the lane detection branch located in the front-end of our detection pipeline, a novel architecture proposed by Neven et al [2], which is called LaneNet, is adopted here in order to detect lane and output the binary lane segmentation (see the red part in *Fig. 2*). Unlike other traditional methods which are prone to robustness issues due to road scene variations, LaneNet handles lane changes and allows the inference of an arbitrary number of lanes, casting the lane detection problem as an instance segmentation task.

---

* *indicates equal contributions.*
✉ *indicates corresponding authors.*

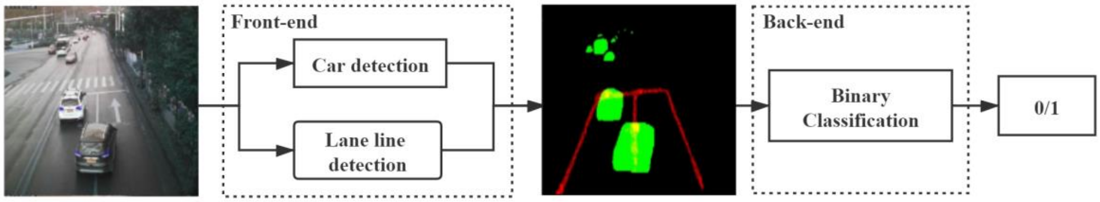

Fig. 2 Pipeline of the proposed SLDNet

In order to speed up inference process, we simplified the architecture of LaneNet and kept only the lane segmentation branch. As the output of LaneNet, the binary lane segmentation is going to be stitched with vehicle segmentation, forming the final segmentation output of the front-end part in our detection pipeline.

2) Vehicle Detection

For the vehicle detection branch, we adopted two-stage object detection method Mask R-CNN [1] which is a classic and widely used object detection and instance segmentation method. It is able to output not only the bounding boxes of vehicles but also their accurate segmentation due to its multi-task network. The time cost of inference is acceptable for us and we do not need to build a real-time processing detection system of vehicle violation, hence we value more on the accuracy but inference speed.

C. Pipeline Back-end

Recall that each original image in our dataset contains 4 sub-images. To proceed binary classification, two methods are experimented in the back-end of pipeline.

1) Method 1. Single Image Classification + Voting

Intuitively we apply mature scheme such as single image classification to conduct binary classification. We choose 34-layer Resnet [4] as backbone and add dense layers to build classifier network. Each original image is split to four sub-images and marked correspondingly, based on which well-trained front-end of pipeline is used to generate the dataset for training binary classifier. An voting schema is then adopted to calculate the score of each original image. For example, if the binary classifier believe that two sub-images out of four in an original image contain offending vehicles, the score is given as two.

2) Method 2. Sequence Modeling

Another network for binary classification is designed through sequence modeling which takes four segmentation images of four sub-images as input at one time. Two reasons for applying sequence modeling are: firstly we are capable to obtain the features of vehicle trajectory through sequence modeling. Secondly, the effect of lane detection strikingly decreases when lane lines are covered by vehicles or other stuff especially in complicated intersections, hence causes performance degradation.

We employed 3D convolution as well as Bidirectional LSTM to extract temporal features, and 2D convolution to

TABLE 1
EVALUATION RESULT

Method of Pipeline Back-end	Accuracy	Recall	Parameter	FPS
Single Image Classification + Voting	0.81	0.70	22.4M	192.8
Sequence Modeling	0.91	0.92	19.5M	166.7

Voting threshold is set to two, which means an original image is considered to contain solid line driving behavior if its score if not less than two.

IV. EVALUATION

We randomly single out 4000 original images from dataset to make up validate set, and evaluate two kinds of back-end network on the same validate set. As shown in TABLE 1, method of sequence modeling surpasses method of single image classification + voting in both metrics of accuracy and recall.

Compared with single image classification method which processes each sub-image separately, sequence modeling method has better ability to complement the obscured lane lines and extract temporal features through taking four sub-images into computing at a time, hence makes it helpful to make final determination and bring promotion in performance.

V. CONCLUSION

In this paper, we proposed a branched, spatio-temporal convolution neural network (SLDNet) for detecting solid line driving behavior. In the front-end of network, lane lines and vehicles are detected, forming the segmentation image for each sub-image. After that, the back-end captures the general spatio-temporal dependencies through sequence modeling and output the final determination result. Experimental results on Pingxiang solid-line-driving dataset show that SLDNet can achieve considerable performance in both metrics of accuracy and recall, and will be put into use in near future.

REFERENCE

[1] Z. Zhou, R. li, Y. Gao, C. Zhang, X. Hei, "The Literature Review of SLDNet", available on https://rlee.netlify.com/files/SLD_paper_literature_review.pdf, 2020.

[2] D. Neven, B. D. Brabandere, S. Georgoulis, M. Proesmans and L. V. Gool, "Towards End-to-End Lane Detection: an Instance Segmentation Approach," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, 2018.

[3] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN", in *IEEE International Conference on Computer Vision*, Venice, 2017.

[4] K. He, X. Zhang, S. Ren, J. Sun. "Deep Residual Learning for Image Recognition", in *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016.