Deep Learning-Based Lane Detection: An In-depth Analysis

Yuhang Wu *

Department of Systems Engineering, City University of Hong Kong, Hong Kong, China

* Corresponding author: yuhangwu5-c@my.cityu.edu.hk

Abstract. Lane detection technology is mainly studied in the traditional method at first, but this method is not suitable for practice. Because in some unfavorable conditions usually do not have the required robustness of autonomous driving systems. Later, Deep learning-based techniques that could increase recognition speed and accuracy were put forth. However, in the complex and changing environment, the recognition accuracy would be reduced, so the current methods need to be strengthened. This paper compares the classical methods based on image segmentation and object detection. The classical object detection algorithms such as VPGNet, Spatial CNN and spatial-temporal deep learning are introduced. In addition, the classical image segmentation methods such as Lanenet, UNet-ConvLSTM and MMA-Net are also introduced. By comparing the performance of these methods on different data sets, the advantages and problems of these methods are analyzed, and the development trend of these methods in this field is predicted. Overall, at the end of this review, the full text is summarized and prospected.

Keywords: Lane detection, image segmentation, target detection.

1. Introduction

In modern society, transportation, as an essential component of peoples' daily lives, plays a vital role in ensuring traffic safety and improving traffic efficiency. As a kind of road traffic sign, lane line is really important to the driver's navigation, the regulation of vehicle running and the maintenance of traffic flow. Due to the quick advancements in intelligent transportation systems and automated driving technology, the accurate detection and recognition of lane lines has become one of the bases for realizing automatic driving and intelligent transportation. Therefore, the review of lane line research algorithms is of great significance.

The study of lane recognition has advanced significantly during the last few decades. Early methods were mainly based on traditional computer vision techniques, such as edge detection, color segmentation, etc. However, these methods were greatly affected by environmental factors such as lighting and weather, and could not meet the precise needs of complex road conditions. Lane detection has advanced significantly thanks to the adoption of deep learning techniques like convolutional neural networks (CNN), which have dramatically increased the detection accuracy and robustness.

However, despite deep learning techniques' enormous success in lane line research, there are still some challenges. First of all, complex and changeable environmental conditions (such as bad weather, night, etc.) put forward higher requirements for lane detection and recognition, requiring the algorithm to have stronger adaptability. Secondly, there are various types of lane lines in the real road, including solid line, dotted line, intersection marking, etc. Another pressing issue that needs to be tackled is the precise identification of various lane lines. In addition, the interaction of lane lines with other traffic participants (such as pedestrians, bicycles, etc.) also needs to be better handled to ensure the safety of traffic.

Therefore, this work intends to provide a review of deep learning-based lane algorithms, as well as a summary and analysis of the benefits and drawbacks of various lane detection and recognition approaches. In addition, this paper will also discuss the challenges of lane detection in different environments and conditions, and explore possible future development directions, with a view to providing valuable reference and guidance for scholars and engineers in the field of lane detection.

2. Lane Recognition Algorithm Based on Target Detection

One of the main issues in the subject of computer vision is the work of object detection, which entails finding all items of interest in the image and determining their category and location. Historically, object detection has been the most difficult task in the field of computer vision because of the wide range in appearance, form, and attitude of different objects as well as the interference of light, occlusion, and other elements during imaging.

There are four tasks in object detection: classification, location, detection and segmentation. Identification of the category that each image or video belongs to is referred to as "classification". Positioning is the area that signifies a little deviation from this aim. The goal of detection is to identify the target and pinpoint its location. To answer the question of "which object or scene does each pixel belong to," segmentation is further separated into instance segmentation and scene segmentation.

The current target detection algorithms are mainly based on three categories, namely Faster R-CNN (Faster Region-based Convolutional Neural Networks) [1], YOLO (You Only Look Once) [2] and SSD (Single Shot MultiBox Detector) [3]. The structure diagrams of these three methods are shown in Fig. 1, 2, and 3. To achieve a balance of speed and accuracy, Faster R-CNN combines an object classification network and a region proposal network (RPN). And may provide end-to-end training, which streamlines the procedure. In addition, it can learn the feature representation suitable for the task, avoiding the trouble of manual feature engineering. However, the computational complexity is high and there may be more false detection. Compared with Faster RCNN, YOLO has a simple structure. The network only includes conv, relu, pooling and full connection layer, as well as the detect layer used to synthesize information at last. YOLO class detection is fast, suitable for dense targets, and can run on embedded devices. However, its accuracy is not high enough when detecting small targets, and the precision of targets at various scales is also insufficient. The foundational network of SSD is based on the conventional image classification network and uses a portion of the vgg16 network as the primary network. The SSD series combines multi-scale feature map detection to achieve a good balance between speed and accuracy. Moreover, it has different scale prediction layers, which can better adapt to different size targets. However, its detection accuracy for small targets is limited and more training data is required.

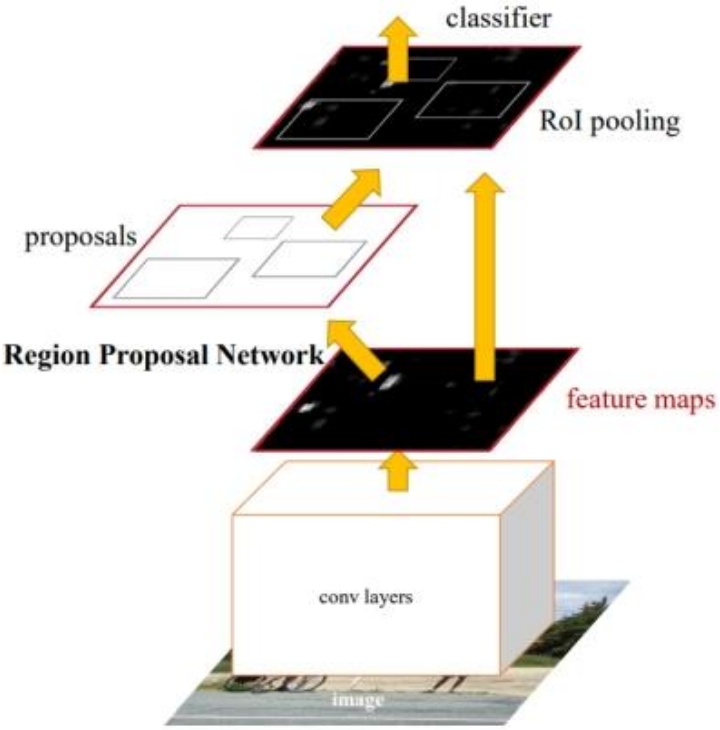


Figure 1. Structure of faster R-CNN [1]

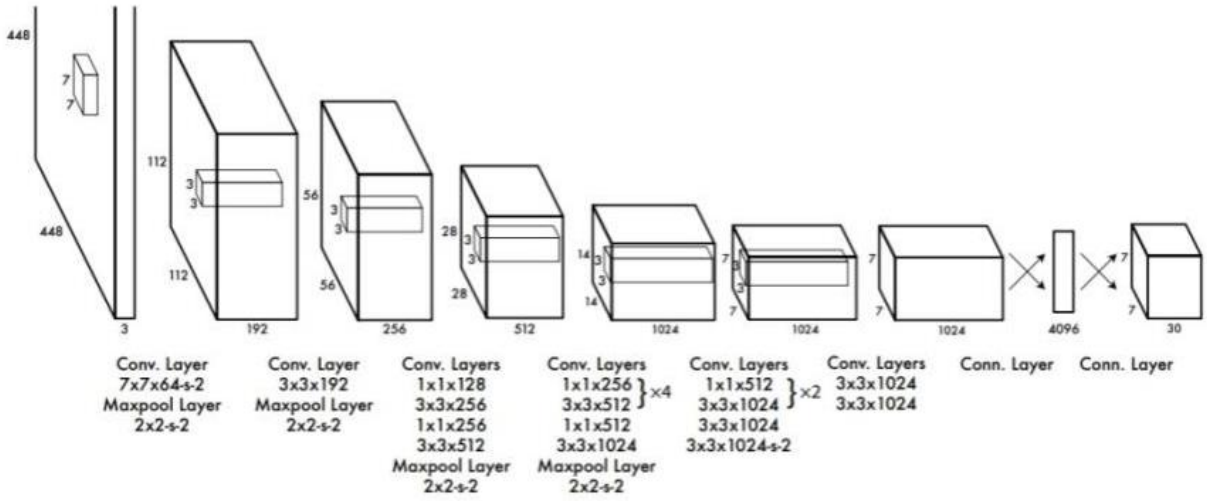


Figure 2. Structure of faster YOLO series [2]

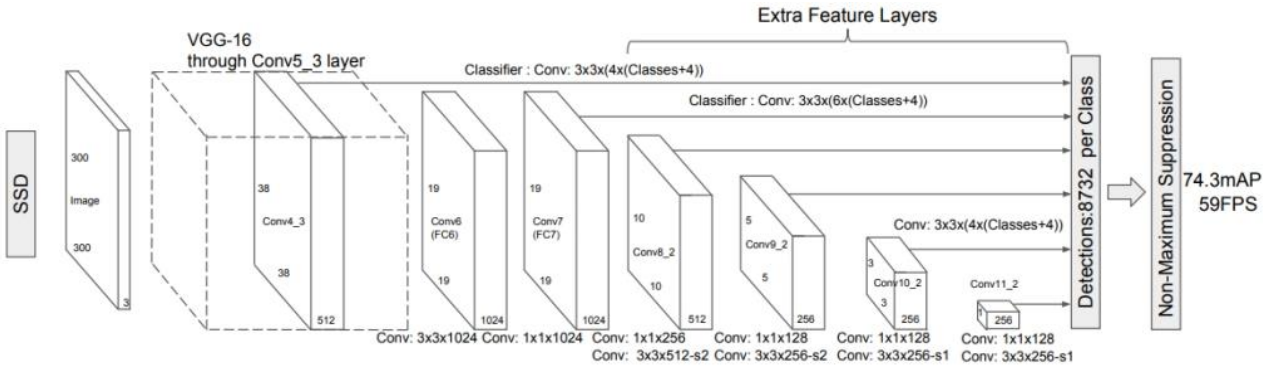


Figure 3. Structure of faster SSD [3]

In 2017, LEE et al. proposed VPGNet (Vanishing Point Guided Network), which uses an improved vanishing point location method and is mainly used to detect roads and road markings [3]. For the purpose of addressing the issue of color distortion caused by insufficient lighting conditions at night, the scheme establishes the road and road marking data set under different scenes [4]. The model will predict the vanishing point of the road, then classify the lanes and road markings, and finally use the inverse perspective transform to classify the adjacent sampling points to get the lane lines. This method can identify lane lines under rainy weather conditions and low lighting conditions, but it needs to carry out complex post-processing operations.

In 2017, a new design for a neural network used for lane recognition has been presented by Pan et al [5], named Spatial CNN. Compared with the traditional layer by layer convolution operation structure, the biggest feature of this structure is that convolution operations are carried out in the order of up, down, right and left, and batch feature maps are obtained after multiple operations. The upsampled pixels of lane lines are segmented by the batch feature map. The structure of this convolution operation is more conducive to the identification of objects that are long in space, and the lane lines belong to this type of objects.

In addition, in 2018, Huang et al. introduced a space-time deep learning system for lane boundary recognition that can successfully identify lane borders in real time even in challenging traffic and weather settings [6]. The technology has good performance in well-lit, open line of sight places, but it is more affected by the environment, in low-light places, cannot effectively identify the lane.

3. Lane recognition algorithm based on image segmentation

The procedure of segmenting an image into individual regions with distinct properties and extracting the item of interest is known as image segmentation. Because lane detection is different from ordinary detection problems, lane lines occupy very few pixels, so it will be very difficult to classify the problem as target detection. Therefore, Academics frequently convert the lane detection issue into an image cutting problem. Based on picture segmentation, the lane lines and background pixels are designated into several classes, and the detection outcomes may be categorized at the pixel level, including semantic segmentation and instance segmentation.

In 2018, Neven et al. regarded lane detection as an instance segmentation task and each lane line as an instance [7]. The clustering loss function developed for one-time instance segmentation was combined with the advantages of binary lane segmentation, for end-to-end training, two decoders were chosen: one for embedding and the other for segmentation. In this network, an embedded branch is added to the output branch of the semantic segmentation network to separate the lanes into different instances. H-net network is proposed to predict the transformation matrix H in each picture. H-Net network has a good effect in dealing with the errors caused by road slope.

Since the images the automobile camera records during continuous driving are continuous, the lanes in one frame cross over with the lanes in the preceding frame, and lane detection can be carried out in the time series prediction framework, Z et al proposed the CNN-LSTM network structure in 2020 [8]. This is a hybrid depth network structure that combines CNN and recurrent neural network (RNN). Between the encoder and the decoder, two LSTM layers are added, and time domain information is integrated. This method can also achieve good performance under vehicle occlusion, shadow, and bad weather conditions.

For real-time video processing, Zhang et al proposed MMA-Net in 2021 to segment video instance lane lines [9] and enhance the features of target frames by integrating local and global memory features. The performance of this method on VIL-100 data set is better than the most advanced lane line detection methods.

4. Experimental analysis and comparison

4.1. Common data set

The commonly used data sets for lane detection mainly include CULane, Tusimple, KITTI, Caltech Lanes and Apollos cape. CULane is a large-scale data set for lane detection, containing traffic scenes in different cities in China. The data set contains thousands of image sequences, covering a variety of complex traffic scenes, including urban roads, highways, etc. The TuSimple isan extensive dataset that is also utilized for lane detection. It contains more than 10,000 images covering a variety of different traffic environments, such as cities, villages and highways. The KITTI dataset is a widely used dataset for computer vision research that contains images, liDAR, and camera data from a variety of scenarios. The data set contains lane marking, but compared with other data sets, the lane marking data is less. The Caltech Lanes dataset contains images of traffic scenes under many different weather conditions, as well as lane markings. The data set covers different scenarios such as cities and highways. ApolloScape data set is a multi-modal automated driving data set provided by Baidu, which includes lane tagging, liDAR data, camera images, etc. The scene diversity in the data set is high. Table 1 displays their pros and downsides.

Table 1. Pros and downsides of different data sets

Data set	Advantages	Disadvantages
Tusimple	The dataset is large and covers a variety of scenarios and road types, which is suitable for training and evaluating lane detection algorithms.	Due to the large size of the data, the accuracy of the annotation may fluctuate, requiring data cleaning and correction.
CULane	It includes rich and diverse traffic scenarios, covering different road types such as cities and expressways, which helps train algorithms to adapt to various environments.	There may be some errors in the annotation, especially in complex scenes, which requires certain pre-processing of the data.
KITTI	Due to the large size of the data, the accuracy of the annotation may fluctuate, requiring data cleaning and correction.	Compared with other tasks such as vehicle detection, KITTI data set has less lane line data, which restricts the training algorithm's efficiency.
ApolloScape	The data set is multimodal and includes data from different sensors to better simulate the autonomous driving environment.	The data sets are large, processing and training costs can be high, and the complexity of the annotation is also high.
Caltech Lanes	The dataset covers a variety of weather and traffic scenarios, helping to train the algorithm to adapt to a variety of different environments.	The relatively small size of the data set may limit the generalization ability of the model.

4.2. Analysis and comparison of different methods

As shown in Table 2, in real applications, the efficiency of the approach using image detection is typically superior than that of the method relied on picture segmentation. The model of the image detection approach can produce more accurate findings when there are enough data sets. According to Table 3, in the combination of dynamic environment and bad weather environment, the MMA-NET model based on image segmentation method has excellent performance in this respect, because it enhances the features of t-he target frame by integrating local and global memory features, so as to improve the feature extraction.

Table 2. Comparison of the various approaches' detection accuracy

Type	Method	Dataset	Accuracy
Object detection	Faster R-CNN [10]	KITTI	73.2%
	SSD [10]	KITTI	75.9%
	VPGNET	Caltech Lanes	88.4%
	Space-time based deep learning methods	Caltech Lanes	96.98%
	Spatial CNN	Cityscape	97%
Image segmentation	Lanenet	Tusimple	96.4%
	UNet-ConvLSTM	Tusimple	86.89%
	MMA-Net	VIL-100	91.0%

Table 3. Detection effects of diverse methods on the VIL-100 dataset [9]

Method	Accuracy
LaneNet [7]	85.8%
SCNN [5]	90.7%
ENet-SAD [11]	88.6%
LSTR [12]	88.4%
STM [13]	90.2%
MMA-ConvLSTM[9]	91.0%

5. Conclusion

The current lane detection techniques using deep learning are categorized in this research into object detection techniques and image segmentation techniques. and then some classical algorithms of these two methods are summarized respectively, and the results of these algorithms are compared and analyzed, and it is concluded that some algorithms based on object detection can achieve better results after improvement in practical application. Then, some data sets commonly used in lane detection tasks are introduced, their characteristics are explained, and their advantages and disadvantages are summarized.

At present, Both the target detection-based lane detection method and the picture segmentation-based algorithm have advanced significantly. However, in complex driving scenarios, the impact of surrounding vehicles and adjacent lanes on current vehicles cannot be completely eliminated, so the next work should go deeper into the research of target identification in driving scenarios and lane change of vehicles, mainly target identification of surrounding vehicles and other traffic facilities, and special post-processing optimization of lane line examples. The network framework of target recognition algorithm and lane detection algorithm is constructed to realize the synchronous detection of the two algorithms.

References

[1] Shaoqing R, Kaiming H, Ross G, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE transactions on pattern analysis and machine intelligence, 2017, 39 (6).

[2] Redmon J, Divvala K S, Girshick B R, et al. You Only Look Once: Unified, Real-Time Object Detection. CoRR, 2015, abs/1506.02640.

[3] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector. CoRR, 2015, abs/1512.02325.

[4] Lee S, Kim J, Yoon J S, et al. VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition, 2017 IEEE International Conference on Compute Vision. IEEE, 2017.

[5] Pan X, Shi J, Luo P, et al. Spatial as Deep: Spatial CNN for Traffic Scene Understanding. IEEE, 2017, 1 (1).

[6] Huang Y, Chen S, Chen Y, et al. Spatial-temproal based lane detection using deep learning, IFIP International conference on artificial Intelligence applications and innovations. Rhodes: Springer, Cham, 2018: 143 - 154.

[7] Neven D, Brabandere B D, Georgoulis S, et al. Towards End-to-End Lane Detection: An Instance Segmentation Approach, 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018.

[8] Zou Q, Jiang H, Dai Q, et al. Robust Lane Detection Froms Continuous Driving Scenes Using Deep Neural Networks. IEEE Transactions on Vehicular Technology, 2020, 69 (1): 41 - 54.

[9] Zhang Y, Zhu L, Feng W, et al.VIL-100: A New Dataset and A Baseline Model for Video Instance Lane Detection, arXiv. arXiv, 2021. DOI: 10.48550/arXiv.2108.08482.

[10] Tian Y, Gelernter J, Wang X, et al. Lane marking detection via deep convolutional neural network. Neurocomputing, 2018, 280.

[11] Hou Y, Ma Z, Liu C, et al. Learning Lightweight Lane Detection CNNs by Self Attention Distillation. CoRR, 2019, abs/1908.00821.

[12] Ruijin Liu, Zejian Yuan, Tie Liu, and Zhiliang Xiong. End-to-end lane shape prediction with transformers. In Winter Conference on Applications of Computer Vision, 2021, 3694 – 370.

[13] Seoung Wug Oh, Joon-Young Lee, Ning Xu, and Seon JooKim. Video object segmentation using space-time memory networks. International Conference on Computer Vision, 2019, 9226 – 9235, 2019.