School of Informatics

THE UNIVERSITY OF EDINBURGH

Informatics Research Review
2D Vision-based Lane Detection Methods in Autonomous Driving Based on deep learning

[redacted]

January 2021

Abstract

Autonomous driving is becoming a promising tendency in the future. As one of the most part of it, real time and accurate lane detection results assist the vehicles to locate themselves and plan routes so that they are able to drive faster and more securely. Therefore, lane detection has been studied by many researchers using deep learning related algorithms to achieve these goals and among them, 2D vision based designs are most researched and encouraging. In this literature review, a number of the state of the art deep learning lane detection methods depend on 2D vision will be presented.

Date: Friday 22^{nd} January, 2021

Supervisor: [redacted]

1 Introduction

In recent years, there has been an increasing interest in intelligent transport, especially autonomous driving, which has become a research hotpot along with popular topics, such as object detection on roads, traffic signs recognition, path planning and lane detection. Among these tasks, lane detection, which intends to detect lane areas or lane lines, is an important component and plays a key role in the driving assistance system. With the real time and accurate lane detection results, the vehicles can manage to complete lane departure warnings, lane-keeping and lane-changing assistance, etc., all of which are vital instruments in supporting the self-driving cars to function.

For successful autonomous driving, hardware basis, including millimeter-wave radars, ultrasonic sensors, LiDAR (Light Detection And Ranging) and cameras, are indispensable to drive safely. They enable vehicles to see and sense everything on the road like human beings, and gather the information they need, send it to CPU(Central Processing Unit) to process and then maintain safe driving. Compared with the other sensors, cameras draw more attention of research as they are the most low-cost ones and they function like the human eyes with the images they collect containing much more abundant messages. Just like each one of us has a pair of eyes, 3D(3 dimensions) cameras can capture stereoscopic images with richer information but meanwhile, it is also more complex to extract the effective data. Therefore, this review will focus on the lane detection algorithms that analyse 2D images. It is beyond the scope of this IRR to examine the 3D camera based methods.

As the result of the improvement of computing power and GPU(Graphics Processing Unit), in the last few years, deep learning (DL) algorithms have shown great potential and excellent performances in many fields, especially in computer vision. Lane detection task is also no exception since DL algorithms have been set up on vehicle platforms and various DL methods have been applied on lane detection problems. Different from the conventional image processing methods which use classic operators for example, Hough Transform, to recognise the straight lines in the images , DL algorithms for lane detection adopt neural networks, especially CNN(convolutional neural networks) to extract the useful features in the images and output the detection results beyond the limitations of illumination and scene changes. As a result, it is not the task of this IRR to examine the traditional algorithms to extract lane lines.

However, some issues are quick to appear following the 2D image DL algorithms. First of all, the performance of a DL method is greatly determined on the dataset which it is trained on. The larger and richer the dataset is, the detection results tend to be better. Secondly, a perfect detector, which is required to be fast and precise, still need more efforts. It is as plain as the nose on your face that the detection results will be more accurate with more valid information included as input and at the same time, this means more processing steps and computations involved and then it is harder to get the real time output. It can be seen from this that processing images need time and these algorithms still have room for improvement.

To sum up, this IRR will focus only on 2D vision-based deep learning lane detection algorithms proposed recently. We will make a survey of their methodologies, the limitations and the potential improvement.We do not engage with the other sensors such as LiDAR and radar based methods and 3D cameras. Besides, it is not the task of this IRR to review all the datasets in computer vision fields, only algorithm related. The readers should also bear in mind that the traditional ways to analyse the lane lines in the pictures are excluded.

2 Literature Review

The introduction section gives a overview of the lane detection task. The remaining of this IRR proceeds as follows. The section 2.1 will examine the available datasets that are used in lane detection tasks, the volume, their scenes, annotation ways and limitations. The section 2.2 and 2.3 are concerned with the methodologies used to get accurate detection results and to achieve real time detection. Both sections analyse their premises, techniques and conclusions. Eventually, this IRR ends up with the summary and conclusion section to state the key findings and justify a research proposal.

2.1 Datasets

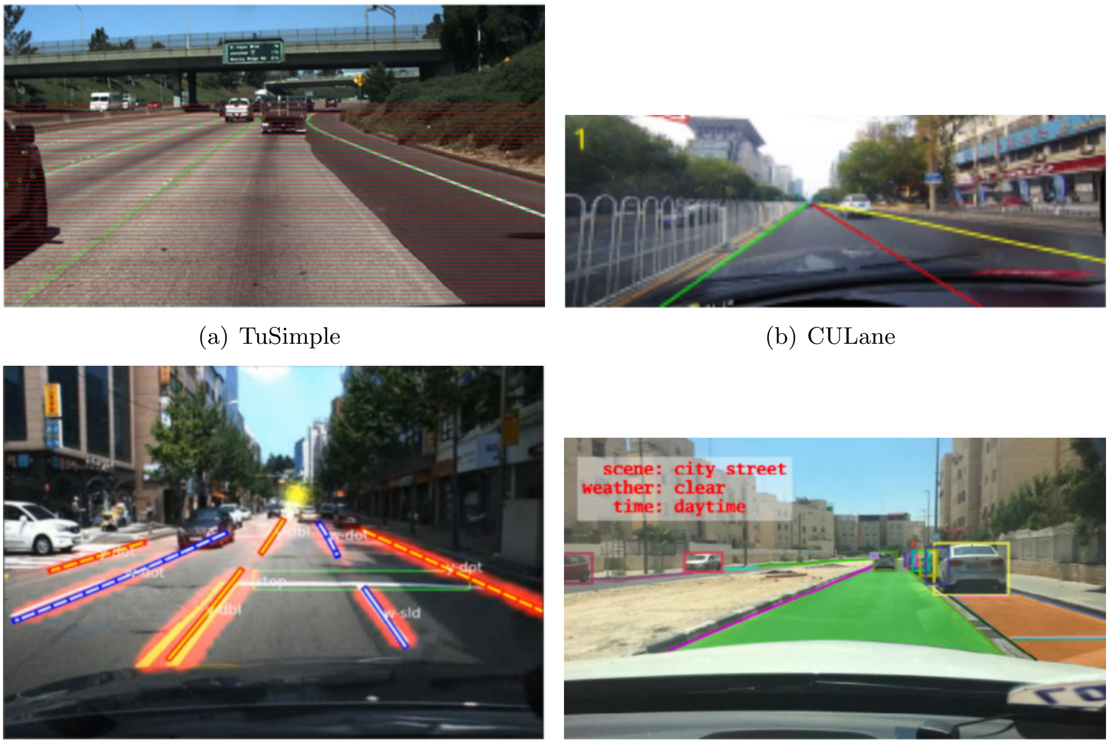
For deep learning algorithms, dataset is an indispensable and crucial part to train the neural networks in order to get excellent performance. For lane detection, there are already a few public datasets used by resaerchers. Figure 1 shows some examples of these datasets.

TuSimple benchmark, published by the self-driving truck company TuSimple, is composed of 7632 labeled images and it annotates lane markings respectively which means each lane is an instance. It contains 7,000 one-second-long video clips of 20 frames each and the last frame of each clip contains labelled lanes which can be used to train the networks. There are various numbers of lane lines, such as 2 ,3, 4 lines or more in the image. However, its traffic scenario is uncomplicated, on highways, with light traffic and clear lane markings, while in real life, these things can be occluded or interfered by human. Besides, the weather conditions are not worse, mainly sunny and cloudy with good illumination. It also contains a small number of low light images and different traffic conditions.

CULane dataset [1] , which is collected by cameras set up on six different cars by various drivers, has more complicated scenes including normal scene, night, no lines, shadows and so on. It has bigger volumes, with 133,235images in total. For each frame, the traffic lanes are manually annotated with cubic splines. The intact lane lines are labeled according to the context in spite of the occlusions by other vehicles or other reasons. However, they do not annotate the lanes at the other side of the barriers as the vehicles are not supposed to drive there so that ignoring them are beneficial to plan the routes. One of their limitations is that only four lane markings are preserved and to achieve a better performance no matter at day or night, the data for later is still not sufficient.

Another dataset, VPGNet [2], which is aimed at rainy and low light conditions, involves roughly 20,000 images under four different weather and illumination conditions: no rain, rain, heavy rain, and evening. These two conditions are relatively more complex and less researched as wet areas on the roads can cause light reflections and alter the appearance of road markings and lanes while low light may induce color distortion, both of which are to the disadvantage of detection. Not only the lane lines, but also the road markings and a vanishing point are annotated to combine these information to output an accurate decision. On the other hand, more information means more time to process. As real time detection is one of the most crucial requirement, it needs a great balance when it is put into use. Besides, this dataset can be used on other tasks other than lane detection, for example road marking recognition.

Besides weather diversity advantage, the new dataset BDD100K [3] owns geographic and environmental variety, which is beneficial to train models which pursues to be robust to condition changes. It collected 100K driving videos, and each of them is around 40 seconds long, 720p, and 30 fps. The videos also appear with GPS/IMU information to display the rough driving



(a) TuSimple

(b) CULane

(c) VPGNet

(d) BDD100K

Figure 1: Datasets examples

trajectories. Sunny, overcast, and rainy three weather conditions, as well as day and night two different times of day are all covered in this dataset. Every the 10th second from each video is annotated as keyframe. Similar with VPGNet, except for lane lines,lane markings and drivable areas are labeled in the meantime with additional information including road object bounding boxes, image tagging, and full-frame instance segmentation, which are useful to multiple computer vision tasks.

After taking a look at those datasets, now we can dive into the specific DL lane detection algorithms.

2.2 For performance

Most of the research attention in this field are drew on improving the overall performance to gain complete and correct lane line information as far as possible. In an ideal scene, the images captured by cameras are pretty clean, with less traffic participants, bright light which neither being exposed nor being dim, clear, unbroken and not occluded lane lines. However, there are various adverse impacts shown in real scenario, such low light at night, reflection caused by wet roads, covered lane lines by cars or other obstacles. Therefore, we need the specific algorithms to tackle the corresponding problems and most important of all, a perfect method than can cope with all kinds of these troubles is expected. This IRR reviews several papers focusing on the general quality and each individual problem mentioned above. Table 1 concludes the performance for accuracy of these works.

Common CNNs(convolutional neural networks) are normally structured by loading convolution process layer-by-layer, which use less spatial relationship among pixels across rows and columns, and to some extend, they do not make full use of the information in images. Spatial CNN (SCNN) [1] proposed a new way to construct the convolutions by slice-by-slice to generate feature maps through slice-by-slice convolutions, which permits message passing between pixels across columns and rows to get better scene understanding so that it can obtain more accurate lane results. A special architecture is designed to propagating the spatial relation. The input tensor is divided into slices according to the image height and the output of a convolution layer is combined with the next slice until the final slice is revised. This structure can be simply incorporated into another neural networks to increase accuracy.

[4] assumes that one of the challenges for lane detection is that the annotations of ground truth images used for train the neural networks are scarce and understated. To use contextual information as supervision, it presents a novel knowledge distillation approach, Self Attention Distillation (SAD) to output lane detection results with exceptional performance. In addition to the feature maps, attention maps are also derived from the network, rather than some additional labels or extra supervision. There are two types of attention maps, gradient-based[5] and activation-based[5]. This efficient technique can be easily integrated into any feed forward networks and the inference time does not grow.

Treating the lane detection problem as an segmentation task, PINet[6] leverages key points estimation and consists of some stacked hourglass networks which are trained in parallel. Input image data is squeezed by the resizing network before being predicted. Several hourglass parts form the predicting network, which we can adjust the length in line with the requirements of computational resources. There are three output sections for each module in the prediction part, computing confidence of outputs, offset and embedding branches to ensure the segmentation effects. It produces key points of the lanes and then classify them into each different instance(individual lane lines are separate instances). This work achieves remarkable accuracy on TuSimple and Culane datasets. However, it shows limited ability to deal with local occlusions or vague traffic lanes. Here is the corresponding public code https://github.com/koyeongmin/PINet new.

In order to tackle with lane detection under the poor light conditions, [7] presented a dataset enhancement way to enlarge the proportion of night scene images, as in its view, the reason why the neural networks fail in low illumination scenes is there is not adequate data to train the model. It uses style transfer Generative Adversarial Networks (GANs) to generate low-light images. Based on the CULane dataset, it select the day and night image pairs to train this GAN to transfer images with acceptable light to images in poor light scenes without the need to change labels. The enhanced dataset is verified by ERFNet[8] and the adaptability for environment is improved. This idea may be extended to other severe conditions such as rainy days and snowy settings. It should be noted that the dataset pairs used to train the style transfer network need much efforts to single out and this requires some techniques.

Another bad condition we usually encounter is rain, especially heavy rain as it may shield the camera lens and obfuscate the view, ruining the collected images. In other words, the content of images are distorted from a better and easy state. In addition, the water wholes left because of the uneven road surface, which is common in real world, after the rain may cause exposure and the relevant information in those areas is lost. There are two approaches to handle this situation. One is like the above-mentioned style transfer way develop new datasets to train the model as it is almost impossible to collect image pairs from the same view of both sunny and rainy conditions. The other one is like [9], remove rain before the next step prediction, which definitely increase the run time. There are few works on dealing with this problem at present.

Table 1: Performance Evaluation

Method	PINet(4H)	SCNN	ENet-SAD
Dataset	TuSimple	TuSimple	TuSimple
Accuracy	96.75	96.53	96.64
F1 score	97.20	95.97	95.92

2.3 For real time

A perfect performance is necessary but not sufficient. In other words, one 100 percent accurate result in one minute is way behind an 80 percent detection in one millisecond. In this way, a real time detector is required to drive safely, not to mention the time spent on operate the brake and other things. There are several works about a trade-off between high quality and immediate outputs. A run time evaluation summary of these methods can be found at table 2.

LaneNet[10], tackles the lane detection task as an instance segmentation problem. It believes that lane lines are lack of distinctive features hence the algorithms are apt to be confused by other targets sharing similar appearance. In this case, a lane edge proposal network is introduced before the line localisation model. The false objects like the arrows and signs on the roads are filtered and the suggested edges that may constitute lanes are selected at this first stage. Then localisation network processes the information from last step and output the coordinates of the lane lines. It is able to run at 250 fps(frame per second). Nevertheless, this method takes an 'bird's-eye view' image, Inverse Perspective Mapping (IPM) of the front view camera as input. Once the camera is installed, the transformation is fixed and if there is any kind of shake that leads to the movement of the camera view, this network may fail with accuracy decline. The adapted version [11] employs a trained perspective transformation network and can be trained end-to-end. One thing led to another, with the network getting deeper, this version can only run at 50 fps.

Not sitting only on one of its sides, ERFNet[8] put forward an end-to-end deep neural network that can run 80 FPS and maintain a good performance meanwhile, which is implemented by introducing novel residual connections[12] and factorized convolutions. The non-bottleneck residual component is redesigned in this paper with only 1D filters, which is decomposed from 2D convolutions. In the end, the performance gets improved and run time is accelerated. As a semantic segmentation task, this architecture combines an encoder to extract feature map and a decoder to recover the abstract data, which is popular is some similar works like SegNet[13]. It achieves significantly more accurate goal while only slightly slower than the fastest network. While comparing with the top one accurate architecture, it maintains similar accuracy and much faster than it. This research can be one of the best among the state of the art algorithms in lane detection in general both for speed and accuracy. The public code of this work can be found at https://github.com/Eromera/erfnet

Besides, a few techniques are recommended in order to obtain the lane detection results real-time, for instance, adopting SAD[4] technique with lightweight models, lane detection can run 10 times faster than SCNN.

Compared with the focus on increasing accuracy of lane detection, there are less works attempting to speed up the running. On the other hand, this also makes sense as the power of computational resources can make a significant influence on it.

Table 2: Run time Evaluation

Method	ERFNet	LaneNet	LaneNet + HNet
Platform	Titan X	Titan X	1080 Ti
Speed(FPS)	83	250	50

3 Summary & Conclusion

This IRR reviews the 2D vision based deep learning algorithms to complete the lane detection task. These methods suggest that there are various angles to start with, for example, enrich the datasets serving for neural network training to involve data under all conditions that we may run into in practical use, come up with special techniques that can be easily integrated into other networks or algorithms to improve accuracy or propose novel neural network architecture design to make the full use of more valid information to obtain exceptional results. Two typical goals for this task exist. One is optimising the accuracy performance, the other one is accelerate run speed of the neural network. In other words, recognising all the lanes in camera view and without any other makings in milliseconds level (considering the time spent on message passing, order given and operation taken), is required for the vehicles to drive safely no matter on highway or at urban scenes.

Notwithstanding these works, there are still some application scenarios expected to be deeply researched. The first one is at rainy days as collecting images at the same view with only difference weather condition is nearly impossible. This needs much more efforts and time so no one yet has done it. After all, deep learning calls for more than thousands of images to train the network before a satisfying result can be gained. Rain removal is another method to consider this problem. However, the same trouble is still there for datasets. Some studies generate simulated datasets by analysing the differences between rainy images and sunny images and modeling the formatting of rainy images to remove rain layer and leave clean sunny layer. Other similar cases are at snowy or foggy days. In contrast to transparent rain drops, snow and fog is almost opaque and it will shield the camera view. On top of lane detection, prediction is one of the requirements to cope with occlusion cases. In addition to these weather conditions, low illumination is another challenge for lane detection. In summary, great efforts are still needed to ensure the algorithms are able to adapt different environments.

When the day comes that we can totally rely on the autonomous driving technology like we trust the phone nowadays, every function of the vehicles is supposed to work at one hundred percent accuracy and be fast, which is affected by the computational power as well, to ensure the whole process successful. As for lane line detection step and not limited to it, what we should pursue is 100 percent accurate results and even faster than human beings. We wish tragedies caused by negligence of human drivers and accidents will no longer exist and this technology can free us from this kind of manual labor to enjoy more leisure time. Those are the points to study this topic so as to let autonomous driving techniques overrun standards of driving of mankind and replace human drivers.

We are still looking forward and working on a flawless algorithm to handle every single different expected and unexpected condition and can still run in truly real time.

References

[1] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial CNN for traffic scene understanding. In *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 2018.

[2] Seokju Lee, Junsik Kim, Jae Shin Yoon, Seunghak Shin, Oleksandr Bailo, Namil Kim, Tae Hee Lee, Hyun Seok Hong, Seung Hoon Han, and In So Kweon. VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, 2017.

[3] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020.

[4] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. Learning lightweight lane detection CNNS by self attention distillation. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2019-October, 2019.

[5] Sergey Zagoruyko and Nikos Komodakis. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, 2017.

[6] Yeongmin Ko, Jiwon Jun, Donghwuy Ko, and Moongu Jeon. Key points estimation and point instance segmentation approach for lane detection, 2020.

[7] Tong Liu, Zhaowei Chen, Yi Yang, Zehao Wu, and Haowei Li. Lane detection in low-light conditions using an efficient data enhancement: Light conditions style transfer, 2020.

[8] Eduardo Romera, Jose M. Alvarez, Luis M. Bergasa, and Roberto Arroyo. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1), 2018.

[9] Hao Sun, Marcelo H. Ang, and Daniela Rus. A Convolutional Network for Joint Deraining and Dehazing from A Single Image for Autonomous Driving in Rain. In *IEEE International Conference on Intelligent Robots and Systems*, 2019.

[10] Ze Wang, Weiqiang Ren, and Qiang Qiu. LaneNet: Real-time lane detection networks for autonomous driving, 2018.

[11] Davy Neven, Bert De Brabandere, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Towards End-to-End Lane Detection: An Instance Segmentation Approach. In *IEEE Intelligent Vehicles Symposium, Proceedings*, volume 2018-June, 2018.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-December, 2016.

[13] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2017.