Z-Net: A Novel Way of Lane Detection

Ziyuan Zhang
School of Vehicles and Energy, Yanshan University, Qinhuangdao 066000, China
1901110384@pku.edu.cn
zhangziyuan_1999@163.com

Abstract. In this paper we proposed a new neural network Z-net for lane detection. The novelty of this network is not the network itself but the detection method. After comprehensive investigation and analysis of the application of several used networks in lane detection, we analyzed the advantages and disadvantages of various networks and modified the existing network to get the Z-net. Under the influence of rapid development of lane detection, it is necessary to propose a neural network with fast detection speed and great accuracy. We found out that Z-net satisfies these two requirements very well.

1. Introduction
Autonomous driving has become the main focus and the most challenging field these years owing to the rapid development of computer vision and robotics control. We want to apply computer vision technology to driving cars in order to get a fully understanding of the traffic around us. By using digital sensors and sensitive control models, we may achieve our goals. Multi-lane detection is one of the vital driver-assist features in the vehicle. Handcraft digital features is an important step towards such environment perception as it allows the car to set the position of itself using the road lanes. It is also important for autonomous vehicles to be capable of estimating subsequent lane departure and trajectory lane departure.

Lane estimation (or detection) may seem trivial and simple at first, but it can be very tricky and challenging. Lane markings vary in shape and color and can be easily affected by the weather and the surrounding environment, adverse situation. It is necessary to retrieve from the environment not only the shape of the line markings but also the position of the car. Since the lanes vary with terrain and road conditions, it is very difficult to predict.

In recent years, researchers have proposed various methods to help to lane detection. From the initial traditional methods like using hand-craft features to applying computer vision and image processing as a medium, the lines can be detection can be quickly and accurately achieved by changing the architecture of the neural network.

2. Related Works
Traditional lane detecting methods depends on highly-specialized, hand-crafted features to identify lane segments. Popular features including color-based features, the structure features and the bar filter and the ridge features. Which are usually using Hough line grouping to group the nodes we get from the camera. Followed by an advanced RANSAC line fitting called RANSAC spline fitting turning these segments into smooth curve then post-processing, the lines appear longer and localized on the screen, to coincide the curve of the actual road. Some others use parallax information, the binocular feature points are extracted and match to estimate the road surface equation, calculate the disparity.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd
map and the vanishing point and then using dynamic programming to detect the road lanes. For a detailed overview of the lane detection system we refer the reader to [1]. However, the traditional approaches can be easily affected owing to the road variations and are sometimes difficult to be model.

More recent methods have placed the traditional hand-craft features with deep neural networks to learn dense predictions. At first, researchers did some adjustment like “changes in classifier level” and “simplify the convolution layer and adopt concept of mask detector with simplification.” with the neural networks that are already existed, and directly applied them to lane detection [2]. However, the lane maybe blocked or missing due to road conditions, some propose the method of using convolution neural networks to extract robustness from the road images, to make this happened, we need to train the neural network with new road data, then trained extra trees regression models to get lane features [3]. To make predictions more accurate, some brought up the ideal of DVCNN, which is a dual-view network contain both the front view and the top view. The front view can eliminate the misjudgements caused by moving vehicles, fences and road boundaries, while the top view image is obtained through IPM, which can remove some unnecessary elements like the arrows printed on the road, after a global optimization, we can get the lane we want [4]. (writer) proposes the task called VPG Net, which is a end-to-end multitask convolution network based on the vanishing point principle that can identify lanes and road markings simultaneously, the author established a database of road and lane markings, including various traffic conditions [5]. All attempts made above is trying to make the network to segment out the lane markings better. However, at the final step of all methods, the generated binary lane segmentations still need to be disentangled into different lane instances.

This problem is usually solved in two ways. Post processing techniques that rely on heuristics, which is usually guided by geographical properties. Another line of work casts the lane detection problem as a multi-class segmentation, in which each lane forms its own class. By doing this, the output of the network contains disentangled binary maps for each lane that can be trained end to end [6]. However, they both have their own downside. The heuristic method is computationally expensive and prone to robustness issues due to road variations. The multitask segmentation problem is limited to detecting only the preset, fixed numbers of lanes and could not cope well with the lane changes.

In this paper, we decided to do the lane detection based on neural networks using a new and creative method called edge detection. The network structure used in our method is basically the same as the structure of FCOS,[7] but we modified the structure in the head section. Alternation in the head section will be explained in detail in the neural network architecture section of the method. The reason we continue using FCOS framework is because compared with original methods, this method has following advantages: First, it is anchor free and bbox free, which means we do not need to use a detection frame, it is relatively simple to implement. Secondly, although it is a single-stage detection algorithm, it still has high detection accuracy which meets the requirement for lane detection in the field of autonomous driving. Most importantly, by modifying the output branch of FCOS, we can achieve many functions like instance segmentation and keypoint detection. This is a great source of inspiration for the new method we proposed.

We also refer to the mask-assembly method [8]. The network automatically locate endpoints on the processed image, we extracted the essence of that method. The difference is that instead of trying to obtain an image, we use the approach of emitting rays to connect dots and lane segments together. Such a new attempt on lane detection is not only novel but also have slight improvement on the detection effect.

3. Method
We are inspired by Polar mask method and create a novel method for lane detection. We also made some structural adjustments to the FCOS network. By modifying the output branch of the FCOS network, we can locate the dot and the edge point on either side of the discontinuous lane segments. Through a special means, we can connect these dots and short lane segments in a straight line according to the direction of the actual road lane. The experimental outcome shows that this method has achieved good results.
3.1. Dataset
In this paper we use CUlane dataset. Using our network to detect lanes in the dataset. The dataset is used mainly include the following aspects. Firstly, data collected by this dataset is very comprehensive. They were collected by cameras loaded on 6 cars that contains more than 55 hours of videos and 133235 frames. Secondly, the dataset includes different type of roads, dividing into normal and 8 challenging categories. This makes the results of our road detection have a certain generality. They annotate the traffic lanes with the cubic splines and annotated the lanes blocked by houses and vehicles which reduced the difficulty of neural network identifying lane lines. This dataset emphasizes the detection of the barriers on the road and pay great attention to practical application. Therefore it only pays great attention to the detection of the four road lanes. This is similar to our concept, which is the main reason we prefer this dataset.

3.2. Neural network architecture

![Diagram of Z-Net](image)

Fig.1 Neural network architecture of Z-Net

In this article, the structure of the neural network is roughly similar to FCOS, but we did some innovation in the head section. In the backbone section, we apply the feature pyramid, which is a bottom-up structure. The input data is fed into the backbone network in order to obtain the feature map. We did some changes to the head, instead of having regression and center-ness processed in parallel at the end, we removed the original content of the head section and added 12 regression branches. The reason for this change will be explained below.

FCOS is designed for object detection, but we use the regression branch for those discontinuous lines. The different targets cause to the different design, and the structure of FCOS shows great success in point detection, so we follow it. Compared with other lane line detection networks based on semantic segmentation, our detection based network is light and fast. To the best of our knows, our work is the first one try to use networks like FCOS for lane line detection.

3.3. Connection method
We input the image in the network, the output image obtained some discrete points and short lane segments, scattered on the edge of both sides of the road lane. The neural network automatically recognizes and locates these dots and dots on both ends of the lane segment. Each point evenly emits 12 rays to all sides. The 12 regression branches represent 12 different scores, rays with largest score value will attract and connect with each other. By setting the score value of the rays emitted by each point, we can connect dots and lane segments according to the shape and direction of the road lane.
After this process, we identified the edge of the lane and complete lane detection. Dots inside and outside the road lane are not included.

![Image](87x585 to 520x691)

**Fig.2** The left and right picture are the road lanes before and after recognition

### 3.4. Loss Function

We will introduce the loss function used in Z-net in this section. The loss function used in this paper is not fabricated, but combines the basic loss function principle with the detection method of this paper. For a given input $x$, we will have a corresponding $\hat{y}$ used as the prediction of the output $y$. Suppose here we assume that $\hat{y}$ represents the prediction of $y=1$.

$$ y = P(y = 1 \mid x) (0 \leq y \leq 1) \quad (1) $$

**output $\hat{y}$**

$$ \sigma(W^T + b) (W/b \text{ are parameters}) \quad (2) $$

By calculating the difference between $y$ and $\hat{y}$, so called “Loss”, we can easily measure the quality of the neural network. In order to properly calculate this loss, we need to find a suitable calculation method. After previous exploration, a function using logistic regression was determined. The basic form of the equation is:

$$ L(y, \hat{y}) = - (y \log \hat{y} + (1 - y) \log (1 - \hat{y})) \quad (3) $$

The reason the equation takes this form is because only by using this form can we avoid nonconvex in optimization process and find the global optimal after gradient decent.

In this paper, our loss function is formed based on this, however, we have made minor changes to make it more practical for this paper. The equation is shown below:

$$ L = - \frac{n}{m} [ y \log y' + (1 - y) \log (1 - y') ] \quad (4) $$

we set the total number of points identified on the image as $m$, the number of points locate on both edge of the lane are denoted as $n$. The ratio $n/m$ can represent the accuracy of lane detection. The larger the ratio, the better the lane detection effect. We use this ratio as the coefficient and the new weight of the new equation to get the loss function. The equation $L$ is used for single training example we continue to introduce the cost function that can cope with multi-example:

$$ J(w, b) = - \frac{n}{m} \sum_{i=1}^{n} L(y^{(i)}, y^{(i)}) = - \frac{n}{m} \sum_{i=1}^{n} [ y^{(i)} \log y^{(i)} + (1 - y^{(i)}) \log (1 - y^{(i)}) ] $$

Using this cost function, we can calculate multi-example training loss.

### 4. Experiment

**Table 1** Comparison of the three networks

| Network         | Inference Time(ms) | Frames per second(fps) | Model size(MB) |
|-----------------|--------------------|------------------------|----------------|
Judging from the result of the experiment, it has made some progress compared with segnet and LMD. As the chart shows, we use three indicators to compare these three neural networks, which are Inference time, frame per second and model size. The three columns in the chart are the comparison of the three network data from top to bottom. Inference Time reflects the speed of network operation from a certain angle. FPS reflects the speed of image transmission, therefore higher FPS can improve the speed of image acquisition and improve the agility of lane detection. Model size reflects the size of memory occupied by the network. Generally, the smaller the size, the faster the detection speed. Through comparison, we can find that Z-net has the fastest inference time, the largest FPS and the smallest model size, which means the Z-net has faster image transmission and road lane detection speed. To sum up, Z-net has the advantages of being more precise, faster and more lightweight, means that the neural network we built is meaningful and valuable.

5. Conclusion
This paper proposes a new network structure called Z-net. It is based on the FCOS network structure and has been modified to a certain extent. Our innovation is that we have modified the head part and added two links, edge-detection and regression. This change has an intuitive manifestation in the lane line recognition and connection method. Our network can greatly improve the speed of detection without affecting the accuracy. The network based on detection rather than segmentation has good robustness and is not easily affected by occlusion. Our work is an attempt in this direction, and this direction needs more improvements, including integrity and continuity of testing, which we believe will receive more attention in the future.

References
[1] Zhang F, Stähle H, Chen C, et al. A lane marking extraction approach based on random finite set statistics[C]//2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2013: 1143-1148.
[2] Huval B, Wang T, Tandon S, et al. An empirical evaluation of deep learning on highway driving[J]. arXiv preprint arXiv:1504.01716, 2015.
[3] John V, Liu Z, Guo C, et al. Real-time lane estimation using deep features and extra trees regression[C]//Image and Video Technology. Springer, Cham, 2015: 721-733.
[4] He B, Ai R, Yan Y, et al. Accurate and robust lane detection based on dual-view convolutional neural network[C]//2016 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2016: 1041-1046.
[5] Lee S, Kim J, Shin Yoon J, et al. Vpynet: Vanishing point guided network for lane and road marking detection and recognition[C]//Proceedings of the IEEE international conference on computer vision. 2017: 1947-1955.
[6] Neven D, De Brabandere B, Georgoulis S, et al. Towards end-to-end lane detection: an instance segmentation approach[C]//2018 IEEE intelligent vehicles symposium (IV). IEEE, 2018: 286-291.
[7] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE international conference on computer vision. 2019: 9627-9636.
[8] Xie E, Sun P, Song X, et al. Polarmask: Single shot instance segmentation with polar representation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12193-12202.