

# A Review of Lane Detection Based on Semantic Segmentation

Shi Jiaqi

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, China  
E-mail: shijiaqi2019@163.com

Zhao Li

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, China  
E-mail: 332099732@qq.com

**Abstract**—With the introduction of full convolutional neural product networks, semantic segmentation networks have also been widely used in the field of deep learning. Most lane detection tasks are currently done on the basis of semantic segmentation networks, so the development of semantic segmentation also directly determines the progress of lane detection. **Methods:** The development of semantic segmentation networks and the performance comparison between different model frames are used to summarize the improvement points as well as the advantages and disadvantages of each approach. Current lane detection network models with good performance based on semantic segmentation networks are described and the performance between the models is compared. **Result:** The current development of deep learning-based lane detection methods has been very fruitful, with significant improvements in network performance, but they cannot yet be applied in practice. For example, lightweight networks are not stable enough in extracting features, while deep neural networks are too ineffective in real time. **Conclusion:** Lane detection is of high research value as a key technology for unmanned driving. However, most of the current neural network methods have not been studied from a practical point of view, and there are few methods that use multiple frames as a basis for research. Therefore, in the future how to efficiently use continuous images for lane detection is a key direction to be researched in the future.

**Keywords**—*Semantic Segmentation; Lane Detection; Deep Learning; Neural Networks*

## I. INTRODUCTION

Lane detection is an integral step in the field of driverlessness, allowing cars to identify lanes so that vehicles know which direction they are travelling in and avoid them pulling out of their

lanes. Lane detection was first done based on the feature approach, which extracts features and fits them based on lane line image features (e.g. colour, shape). However, feature-based methods are susceptible to poor feature extraction due to factors such as light and obstacle occlusion, and the algorithm for fitting lanes requires a range of parameters based on lane characteristics, often with many limitations. Therefore, feature-based algorithms are not suitable for practical applications.

### A. Traditional methods

Due to developments in computer vision, lane detection based on model algorithms has been proposed and this method is mainly divided into straight line detection and curve detection. Most algorithms for straight line detection use the Hough transform to perform this method, which equates straight line detection to coordinate statistics, simplifying detection, but frequent coordinate mapping will increase the complexity of the algorithm and cause a reduction in real-time efficiency. A number of improved algorithms have been subsequently introduced to address this algorithm. For example, the maximum length straight line based lane line detection algorithm proposed by Xie Mei et al. This algorithm connects broken straight lines by setting a maximum straight line gap, selecting the maximum length straight line in the vertical direction on either side of the vertical centre of the image, using the maximum length straight lines on each side as edges, binarising the interior of the edges, and subjecting the interior image to Hough

straight line detection, with the line closest to the vertical centre being the final detected lane line. This method greatly reduces the search area, simplifies the difficulty of the algorithm and speeds up detection efficiency.

There are also many different detection methods for bend detection, most algorithms use different line shapes to fit the lanes and rely on different models, the higher the complexity of their models, the better the fit to the lanes, but taking into account the efficiency of the algorithm also requires a streamlined model. Some of the better known lane modelling methods are the B-spline model and the IPM model (Inverse Perspective Transformation Model)[5]. The IPM model converts the monocular vision image into a bird's eye view by applying an inverse perspective transformation, converting the lanes from far to near into parallel lanes, which reduces the difficulty of lane detection. However, this method requires knowledge of the camera's internal parameters, and then determines the transformation matrix for the inverse perspective based on the specific parameters, so when the camera's internal parameters are not known, the inverse perspective transformation model is not very widely used. The B-spline model uses multiple control points to fit the lane lines, also based on parallel perspective technology, and the algorithm is highly accurate but has poor real-time performance; moreover, the method divides the lane lines into multiple areas for separate detection, especially in the presence of false lane lines or lane wear, and the accuracy of the algorithm is not guaranteed, and the lane line jump is serious.

### *B. Deep learning methods*

Research on lane detection based on deep learning neural networks has been conducted in recent years, and the results have been a great improvement compared to traditional algorithms. Due to the variability of the practical situation, most scholars have transformed the lane detection problem into a semantic segmentation problem. Convolutional neural networks have had great success in image detection and recognition, so convolutional-based semantic segmentation networks also have a wide range of applications in

lane detection. The laneNet network proposed by Davy [1] et al. converts lane processing into an end-to-end instance segmentation problem, using a lightweight ENet network as the main structure and adding instance segmentation branches to classify different lanes into different categories. XingGang Pan[2] et al. proposed a spatially based deep neural network SCNN (Spatial CNN), which was trained to classify the network for the poorly conditioned dataset CULane, and the network performance was substantially improved in lane detection compared to the traditional convolutional network. The ENet-SAD[3] network is based on the lightweight neural network model ENet incorporating elements of SAD, Self-Attention Knowledge Distillation, which has 20 times fewer parameters, runs and is 10 times faster and more accurate than the state-of-the-art SCNN. While domestic scholars have paid much attention to the diverse road conditions, an improved YOLOv3 model was proposed by Zhang Xiang [4] to improve the adaptive and accuracy problems of lane detection technology in complex road environments, where complex road problems refer to road potholes, rugged mountain roads and other problems. A multi-scale MFCN model was proposed by Shuaihua Wang et al. to solve the lane line sample inhomogeneity problem, using a weighted loss function to solve the lane line inhomogeneity problem. For sharp turns, over curved lanes, CurveLane-NAS, a lane sensitive architecture search framework combining NAS with curved lane detection algorithms proposed by Huawei Noah's Ark Lab and Sun Yat-sen University [6], can automatically capture long-distance coherent and accurate short-distance curve information to solve the problem of curved lane detection.

The neural network methods described above are all based on semantic segmentation networks for end-to-end lane line detection, i.e. the lane detection problem is converted into a multi-category segmentation problem where each lane belongs to one category, which enables the end-to-end training of a well-classified binary graph. This paper therefore focuses on describing the current state of development of lane detection based on semantic segmentation networks.

## II. SEMANTIC SEGMENTATION NETWORK

There are many applications of neural networks in the field of computer vision, such as image classification [11], target detection [12], semantic segmentation [14], and instance segmentation [13]. One important problem in computer vision is the semantic segmentation network, as its work is much more complex than the classification and detection tasks. Semantic segmentation of images means that each pixel of the input image is assigned a semantic category to it, thus obtaining a dense classification for each pixel. That is, it requires learning the contour of the object, the location of the object and the class of the object from high-level semantic information and local location information, and thus scholars in general view the semantic segmentation problem as a pixel-level target segmentation.

Traditional semantic segmentation is generally classified into threshold-based segmentation methods [8], region-based segmentation methods [9], edge-based segmentation methods [10] and so on. The threshold segmentation method is one of the commonly used segmentation techniques, which in essence automatically determines the optimal threshold value based on certain criteria and uses these pixels according to the grey level in order to achieve clustering. Region-based segmentation is a segmentation technique based on the direct search for new regions and can be divided into two basic extraction methods: region growing and region splitting and merging. Region growth is based on individual pixel points, which are aggregated together to form regions with similar features, and is computationally simple and works well for uniformly distributed images. Region splitting and merging starts from the overall image and obtains each sub-region by splitting between pixel points, the quadtree decomposition method is a typical representative method. Edge detection-based segmentation methods segment images by detecting the edges of different regions. The simplest edge detection method is the parallel differential operator method, which uses the nature of discontinuous pixel

values in adjacent regions and uses derivatives to detect edge points. Most traditional methods work by extracting low-level semantics of the image, such as size, texture, colour, etc. In complex environments, the response capability and accuracy is far from adequate.

With the development of deep learning, the proposal of convolutional neural networks has allowed significant progress to be made in combining semantic segmentation and neural networks. Because of the powerful generalisation ability of convolutional networks to acquire image features, they have shown excellent performance in different areas of image and video such as image classification, target detection, visual tracking and action recognition. The following subsections describe the development of semantic segmentation networks based on deep learning.

### A. *Derivation of the semantic segmentation network model*

A turning point in the development of semantic segmentation based on deep learning was the FCN, a fully convolutional neural network for end-to-end segmentation, proposed by Jonathan Long [14] et al. in 2014, when a major breakthrough in semantic segmentation was achieved. It upsampling the local information loss caused by the convolutional neural network with a deconvolution operation that restores the feature map to the original image size, hence the current general semantic segmentation network architecture is an encoder-decoder structure. Where the encoder is usually a pre-trained classification network, the task of the encoder is to semantically project the discriminable features learned by the encoder onto the pixel space to obtain dense classification.

A number of scholars have since proposed a number of sophisticated network frameworks, but most have been studied on the basis of fully convolutional networks. In this paper, we only discuss semantic segmentation networks that are applicable to lane detection, and the research in recent years is shown in the following Table 1:

TABLE I. COMPARISON OF IMAGE SEMANTIC SEGMENTATION NETWORKS

<i>Methods</i>	<i>Features</i>	<i>Advantages</i>	<i>Disadvantages</i>
<b>FCN[14]</b>	Proposes novel end-to-end network architecture ; Encoder-decoder architecture ; Fully connected output classification.	Images of any size can be split.	The large number of parameters and the pooling operation caused a loss of spatial information in the images and a low accuracy rate.
<b>SegNet[15]</b>	Symmetrical Encoder-Decoder architecture ; up-sampling to recover image size at the decoding stage using unpooling; full convolutional layer output classification.	The small number of parameters compared to FCN maintains the integrity of the HF information.	The computational effort is too large to meet the real-time requirements of lane detection. The up-sampling operation also loses adjacent information.
<b>Unet[16]</b>	Symmetrical structure; connects each stage to the encoder feature map with the upsampled feature map of the decoder.	Can be trained end-to-end from very small data sets; fast.	More suitable for segmentation of medical images
<b>ENet[17]</b>	Consisting of Bottleneck modules; with a large encoder-small decoder structure.	Greatly reduces the number of parameters and floating point operations, takes up less memory and has high real time performance.	Increases the number of calls to the kernel function; not very precise and unstable results.
<b>PSPNet[18]</b>	Improving ResNet structures using null convolution ; A pyramid pooling module has been added.	The segmentation accuracy exceeds that of models such as FCN, DPN and CRF-RNN.	Obscured situations between targets are not handled well and the edges are not segmented accurately enough.
<b>ERFNet[19]</b>	ENet network improvements; the adoption of factorized convolutions;	Non-bottleneck is more accurate to bottleneck.	High calculation volume compared to Enet.
<b>DeepLab V3+[20]</b>	Uses a modified version of Xception as the base network; uses atrous[19] convolutional kernels.	More accurate segmentation of target edges; considers global information, eliminates noise interference and improves segmentation accuracy.	The model does not run at a high speed and has a high storage space requirement.
<b>FPN[21]</b>	Combining FCN and Mask R-CNN[13] using rich multi-scale features.	Semantic segmentation and instance segmentation tasks can be solved simultaneously.	Increased inference time; larger memory footprint; use of image pyramids only in the testing phase.

### B. Limitations of semantic segmentation networks

While semantic segmentation web techniques are currently achieving good segmentation results, there is currently no universal algorithm that is applicable to all domains. In practical segmentation tasks, it is necessary to choose the segmentation method flexibly depending on the application scenario, and in some cases it is even necessary to use a combination of segmentation methods to obtain the best results. Therefore semantic segmentation still has some challenges: 1) network training requires a large dataset and pixel-level image quality is difficult to guarantee due to the extensive use of strongly supervised

segmentation-based methods that rely on manual data tagging and are less adaptable to unknown scenes; 2) segmentation of small-sized targets is not accurate enough; 3) segmentation algorithms are computationally complex; and 4) interactive segmentation cannot be achieved, which hinders the implementation, application and promotion of segmentation techniques.

### III. SEMANTIC SEGMENTATION NETWORK BASED LANE DETECTION METHOD

With the rapid development in the field of unmanned vehicles, scholars have proposed many sophisticated lane detection network models in recent years. The current mainstream lane line detection networks are basically based on semantic

segmentation networks to complete the detection task, so this subsection focuses on the current better performance of the semantic segmentation-based lane detection network models.

### A. Related methods

#### 1) SCNN

In a classical CNN, each convolutional layer takes input from the previous layer, applies convolution and nonlinear activation and then passes the output to the subsequent layers. XingGang Pan [2] et al. proposed the SCNN model based on CNN with a spatial attention element. The SCNN model views the rows and columns of the feature map as layers, and also uses convolution plus nonlinear activation to achieve a spatially deep neural network. This allows spatial information to be propagated between different neurons in the same layer, enhancing spatial information and thus being particularly effective for identifying structured objects.

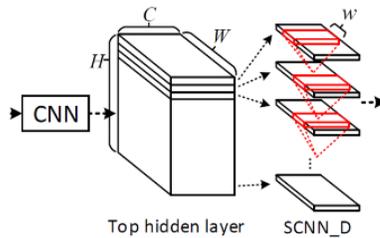


Figure 1. SCNN\_D module

As shown in Figure 1, the SCNN\_D module, where SCNN is applied to a 3D tensor  $C \times H \times W$ , with  $C$ ,  $H$ ,  $W$  representing the number of channels, length and width respectively. To achieve spatial information transfer, the tensor is sliced into  $H$

slices, the first slice is sent to the convolution layer of size  $C \times W$ , and the output of this slice is summed to the next slice as a new slice. Then the next slice continues to apply convolution until all slices have been processed and input to the next module. Three similar modules follow, each convolving the feature map from a different direction in three dimensions. SCNN does not acquire global elements when passing information, but passes them sequentially, thus simplifying the structure of information passing and accelerating the efficiency of the model.

#### 2) LaneNet

LaneNet [1], on the other hand, transforms the lane detection problem into an instance partitioning problem, where each lane line forms a separate instance, but all belong to the lane line category. The authors propose an end-to-end multitasking network with branching structure, consisting of a lane embedding branch and a lane embedding branch. One of the lane segmentation branches outputs two categories: background and lane lines, represented by a binarised segmentation map; the corresponding pixels of each lane line are concatenated to construct the binarised segmentation map, which has the advantage that the network can predict the lane position even if the lane lines are obscured. The lane embedding branch further separates the segmented lane lines into different lane instances. This branch is based on the one-shot method for distance metric learning, which is easily integrated into standard feedforward networks and can be used for real-time processing. The network structure is shown in Figure 2:

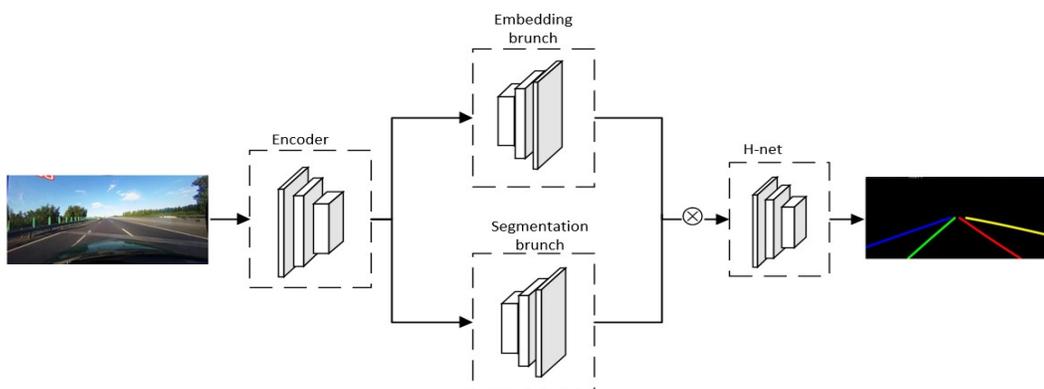


Figure 2. LaneNet network framework

The network then adopts a type of network called H-Net network for predicting the transpose matrix H, which solves the error caused by the traditional transpose matrix in the case of uneven ground planes such as slopes. Finally, the CNN learns the transformation matrix to perform an angular transformation to make the lane lines parallel in order to fit a reliable lane to different pictures or horizon transformations in the pictures.

3) DCNN+DRNN

The lane line features consist of continuous lines and feature extraction by the current frame alone is not sufficient information representation. Therefore QIN Zou [7] et al. proposed lane detection by successive frames, where the information of each frame is extracted by the CNN module and the CNN of multiple successive frames maintains temporal continuity and is fed to the RNN module as feature learning and lane line detection. CNNs have the advantage of being able to process a large number of images, extracting the input image into a small-sized feature map through operations such as convolution and pooling. RNN has the advantage of continuous signal processing, temporal feature extraction and integration, and can be used for lane detection and prediction.

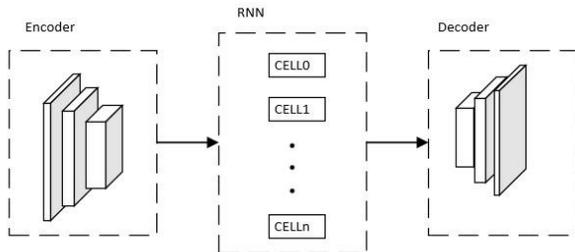


Figure 3. DCNN+DRNN network framework

To fuse CNN and RNN networks into an end-to-end training network, the authors used the classical network structure of semantic segmentation, with the encoding-decoding structure as the main framework. The images are fed into the coding module to obtain a temporal feature map; the feature map is then passed as input to the RNN network to predict lane line information; the output of the RNN is then passed back to the decoding module to obtain a probability map of lane prediction. Experiments have shown that this network performs better than

a network based on single-frame feature extraction, and has more stable performance especially under some complex road conditions. And the longer the continuous input sequence the better the performance, further proving that multi-frame images are more helpful than single-frame images.

B. Performance comparison

Deep learning based lane line detection requires a large amount of well labeled lane line training data to train the convolutional neural network model. Early lane line datasets were generally small in size and the scenarios were relatively homogeneous for deep learning lane line detection and the amount of data was too small to achieve a good model. With the rise of lane line detection technology, lane line datasets have evolved rapidly. The CULane dataset contains 133,235 images, of which 88,880 are in the training set, 9,675 in the validation set and 34,680 in the test set. It includes urban, rural and motorway scenarios as well as a variety of weather, heavy lane shading, lane wear and tear, etc. Road conditions are complex and variable, so many networks use the CULane dataset to reflect the performance and strengths and weaknesses of the network.

The key indicator of lane detection is the accuracy rate. Generally, the calculation of the accuracy rate first requires the calculation of the overlap between the true value of the lane T and the predicted value H as a percentage of the true value IoU, and the calculation formula is shown in (1). If IoU is greater than the set threshold, the lane line is considered to be accurately detected and the number of predicted lanes correct TP is added to 1, otherwise the number of predicted lanes incorrect FP is added to 1; the formula for calculating the accuracy rate is shown in (2).

$$IoU = \frac{H \cap T}{T} \times 100\% \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

To better illustrate the performance comparison between networks, the results of the above

network performance are presented in Table 2 below using the CULane dataset the results in the table show that the accuracy of the DCNN+DRNN hybrid neural network is basically better than the other network models in various scenarios. The ability to extract feature information in congestion and bends is slightly weaker than the LaneNet network. As the hybrid neural network model

extracts features in multiple consecutive frames, generally the slow movement of vehicles in congestion or large angular shifts in road direction cause long periods of time when road features cannot be extracted or are too different from the features of the previous consecutive frames, which can cause bad results.

TABLE II. ACCURACY COMPARISON OF LANE LINE DETECTION IN DIFFERENT SCENES OF CULANE

Methods	Normal	Crowded	Night	NoLine	Shadow	Arrow	DazzleLight	Curve	Crossroad	Total
SCNN	0.906	0.696	0.661	0.434	0.669	0.841	0.585	0.644	0.532	0.716
LaneNet	0.921	<b>0.708</b>	0.714	0.563	0.697	0.850	0.635	<b>0.746</b>	0.591	0.742
DCNN+DRNN	<b>0.984</b>	0.652	<b>0.797</b>	<b>0.724</b>	<b>0.840</b>	<b>0.852</b>	<b>0.774</b>	0.731	<b>0.787</b>	<b>0.782</b>

Accuracy, a key metric for lane detection network performance, is then real-time. Real-time performance is evaluated in terms of processing speed and the amount of memory consumed, but processing speed is relatively more important. As can be seen from the Tabel 3 below, the LaneNet network has the best real time performance, as it uses a lightweight network as the base framework, so the processing speed is significantly better than the other frameworks.

TABLE III. COMPARISON OF DETECTION SPEED OF VARIOUS NETWORK MODELS

Methods	Time(ms)	fps
SCNN	42	23.8
LaneNet	<b>19</b>	<b>52.6</b>
DCNN+DRNN	58	17.2

#### IV. CONCLUSION

##### A. Summaries

This paper focuses on the development of lane detection tasks in terms of semantic segmentation. There is still a lot of room for development of semantic segmentation networks, both in terms of the cost of training and the complexity of computation, which are not yet up to the requirements of practical applications. The current lane detection network is basically based on the semantic segmentation network, so the development of the semantic segmentation network has a direct impact on the progress of lane detection. Although deep learning-based lane

detection is more adaptable to unknown environments than traditional methods, it is still unable to achieve both accuracy and real-time performance. Although the LaneNet network uses a lightweight network, the detection results are not as good as compared to the DCNN+DRNN hybrid neural network in some poor road conditions. And although the hybrid neural network outperformed the other models in all aspects, the processing speed was clearly not up to the practical requirements.

##### B. Prospects

Although lane detection technology based on deep learning methods is still in the development stage and is still some distance away from practical applications, the trend of development will become faster and faster. There has also been a great deal of progress in deep learning methods, but most of them are based on feature extraction on a single frame basis, and there are still few methods that have been studied on a video basis, while in practical applications the images captured by the camera are often continuous. Therefore, in the future the focus of research on lane detection tasks should be on how to efficiently segment and detect lane lines using continuous frames. In the future deep learning based lane line detection will soon be used in practice.

#### REFERENCES

- [1] Neven D, Brabandere B D, Georgoulis S, et al. Towards End-to-End Lane Detection: an Instance Segmentation Approach [J]. IEEE, 2018.

- [2] Pan X, Shi J, Luo P, et al. Spatial As Deep: Spatial CNN for Traffic Scene Understanding. 2017.
- [3] Hou Y, Ma Z, Liu C, et al. Learning Lightweight Lane Detection CNNs by Self Attention Distillation [J]. 2019
- [4] An improved YOLOv3 model based on skipping connections and spatial pyramid pooling [J]. *Systems Science & Control Engineering*, 2021, 9(S1).
- [5] Chun-yang CHENG, Min LI, Xue-wu ZHANG, Yu-bo XIE, Yan XIANG, Jin-bao SHENG. A Lane Detection Algorithm under Complex Scenes [A]. *Advanced Science and Industry Research Center. Proceedings of 2017 2nd International Conference on Computer, Mechatronics and Electronic Engineering(CMEE 2017)*[C]. *Advanced Science and Industry Research Center: Science and Engineering Research Center*, 2017:5.
- [6] Xu H, Wang S, Cai X, et al. CurveLane-NAS: Unifying Lane-Sensitive Architecture Search and Adaptive Point Blending [J]. 2020.
- [7] Q. Zou, H. Jiang, Q. Dai, et al. Robust Lane Detection From Continuous Driving Scenes Using Deep Neural Networks [J]. 2019.
- [8] Zhihuan Wu, Yongming Gao, Lei Li, Junshi Xue, Yuntao Li. Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold [J]. *Connection Science*, 2019, 31(2).
- [9] Vadim Romanuke. A Prototype Model for Semantic Segmentation of Curvilinear Meandering Regions by Deconvolutional Neural Networks [J]. *Applied Computer Systems*, 2020, 25(1).
- [10] Chu He, Shenlin Li, Dehui Xiong, Peizhang Fang, Minsheng Liao. Remote Sensing Image Semantic Segmentation Based on Edge Information Guidance [J]. *Remote Sensing*, 2020, 12(9).
- [11] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.
- [12] Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// CVPR. IEEE, 2014.
- [13] He K, Gkioxari G, P Dollár, et al. Mask R-CNN [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017.
- [14] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
- [15] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017:1-1.
- [16] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer International Publishing, 2015.
- [17] Paszke A, Chaurasia A, Kim S, et al. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation [J]. 2016.
- [18] ZHAO H S, SHI J P, (I X J, et al. Pyramid network[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017:2881-2890. DOI: 10.1109/CVPR.2017.660.
- [19] Romera E, Alvarez J M, Bergasa L M, et al. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, PP(1):1-10.
- [20] CHEN L C, ZHU Y, PAPANDEOU G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [EB/OL]. [2018-08-09]. <https://arxiv.org/pdf/1802.0261v1.pdf>. 20.FPN
- [21] Lin T Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.