Viser Technology Pte. Ltd.
Mechanical Engineering Science
DOI: 10.33142/mes.v5i2.12721

Research Article Open Access

A Review of Lane Detection Based on Deep Learning Methods

Yunzuo ZHANG [1*], Zhiwei TU [1], Fenfen LYU [2]

1. School of Information science and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, 050043, China
2. Sports Technical School of Hebei Sports Bureau, Shijiazhuang, Hebei, 050041, China

[*]***Corresponding Author:*** *Yunzuo ZHANG, E-mail: zhangyunzuo888@sina.com*

Abstract

Lane detection is an important aspect of autonomous driving, aiming to ensure that vehicles accurately understand road structures as well as improve their ability to drive in complex traffic environments. In recent years, lane detection tasks based on deep learning methods have made significant progress in detection accuracy. In this paper, we provide a comprehensive review of deep learning-based lane detection tasks in recent years. First, we introduce the background of the lane detection task, including lane detection, the lane datasets and the factors affecting lane detection. Second, we review the traditional and deep learning methods for lane detection, and analyze their features in detail while classifying the different methods. In the deep learning methods classification section, we explore five main categories, including segmentation-based, object detection, parametric curves, end-to-end, and keypoint-based methods. Then, some typical models are briefly compared and analyzed. Finally, in this paper, based on the comprehensive consideration of current lane detection methods, we put forward the current problems still faced, such as model generalization and computational cost. At the same time, possible future research directions are given for extreme scenarios, model generalization and other issues.

Keywords: *Deep learning; Lane detection; Image segmentation; Object detection; Parametric curves*

1 Introduction

With the continuous development of autonomous driving technology, lane detection, as a key link in application scenarios such as autonomous driving and intelligent parking, plays a crucial role in ensuring that the vehicle accurately understands the road structure and travels in complex traffic environments. After more than a decade of development, lane detection methods [1-2] have evolved through an evolutionary process, from early traditional methods, including image processing [3-4] and feature processing [5], to current deep learning methods [6-9]. Despite the continuous emergence of new detection methods, accuracy and robustness are still the focus of attention in lane detection tasks. With the introduction of deep learning methods, the performance of lane detection has been improved compared to traditional methods, but due to the dataset and other influences, the resulting accuracy is still limited. Although some scholars have proposed corresponding solutions, there are still some limitations.

This paper is devoted to a comprehensive review of lane detection tasks based on deep learning methods in recent years, aiming to provide insights into the latest advances in the field. First, this paper will introduce the background of the lane detection task, including the lane dataset, factors affecting lane detection, as well as traditional lane detection methods and related deep learning methods. By categorizing deep learning methods and analyzing their advantages and disadvantages, it will provide readers with a clear research background.

In the deep learning methods section, this paper will explore segmentation [10], object detection [11], end-to-end, parametric curves and keypoint [12-14] based methods, so as to provide a comprehensive understanding of the characteristics of various lane detection methods. Through a brief comparative analysis of some typical models, this paper will highlight their advantages and limitations in different aspects.

Finally, based on the comprehensive consideration of the current lane detection methods, this paper presents the problems still faced by the current methods, such as computational cost and model generalization, and puts forward some suggestions and solutions to provide the readers with insightful thoughts on the future research direction in this field. Through this review, this paper expects to provide more comprehensive and in-depth references and insights for the research and practical applications in the field of lane detection.

2 Background

As one of the popular tasks in the field of

Received on November 12, 2023; Accepted on December 14, 2023

computer vision, lane detection is widely used in application scenarios such as autonomous driving. With the rapid development of computer vision, lane detection methods have achieved remarkable success in terms of accuracy and robustness. Therefore, an introduction to lane detection and common datasets as well as factors affecting the datasets can help to gain a deeper understanding of the importance and progress of this task.

2.1 Lane detection

In a lane detection task, the vehicle system must be able to accurately capture and represent the shape characteristics of the lane. This usually involves the accurate detection and modeling of various types of lane lines on the road, including solid lines, dashed lines, curves, etc. Accurate lane detection not only contributes to the stability of the vehicle while traveling, but also provides critical decision support such as keeping the vehicle in the correct lane, vehicle turning, etc.

In the process of lane prediction, the shape features of lane lines need to be extracted and represented with high accuracy. Also, for different types of lane lines such as solid and dashed lines, the vehicle system needs to be able to correctly recognize their start and end points as well as curvature information.

Different environmental conditions need to be considered in order to obtain accurate lane shapes, such as lighting changes, bad weather, and nighttime driving. Lane lines may exhibit different visual characteristics in different environments, so the detection system needs to be robust to environmental changes so that it can adapt to a variety of complex real-world situations.

2.2 Common lane datasets

A lane dataset is a set of image data used to train and evaluate lane detection algorithms. These datasets usually contain images of road scenes, which include lane lines, road signs, other traffic signs and environmental information. Such datasets are critical for research and development of lane line detection algorithms because they provide the diversity and complexity of lane lines in the real world. Four common datasets are shown below:

(1) CULane, a dataset provided by the University of Hong Kong, which extracts 133,235 images from 55 hours of Beijing road video data collected by a vehicle-mounted camera and contains image data of traffic scenes under various weather conditions.

(2) TuSimple, 6408 images were extracted by collecting lane images of foreign highways under general weather conditions.

(3) ApolloScape, a dataset provided by Baidu's Apollo autonomous driving team. It provides large-scale image and point cloud data containing images from different scenes such as cities and highways.

(4) BDD100K, a dataset created by Berkeley Deep Learning Lab for autonomous driving scenarios. The dataset contains diverse urban traffic scenarios. The dataset contains a large number of diverse images and videos totaling over 10,000 samples covering different weather conditions, traffic scenarios, and driving environments. The content of the dataset covers various road types such as city roads and highways, and provides rich annotation information including the location and shape of lane lines.

2.3 Factors affecting lane detection

In the field of lane detection, the key factors affecting model performance mainly originate from lane datasets containing information about various road scenes such as city roads, highways, and rural roads [15-16]. While providing training samples, these datasets also carry rich realistic background information, which involves various aspects such as illumination changes [17], object occlusion, and road deformation. In this subsection, we will delve into the factors affecting lane detection in order to fully understand the realistic context of model training. The common influencing factors are as follows:

(1) Influence of lighting changes. Changes in lighting conditions are one of the factors that cannot be ignored in lane detection. Lighting conditions such as strong light, low light, and shadows may lead to changes in the appearance of lane lines, thus affecting the visibility of lane lines.

(2) The effect of object occlusion. In the actual road, vehicles, pedestrians or other obstacles may obscure part of the lane line, making it partially or completely invisible, resulting in the model is difficult to detect the lane completely.

(3) The effect of road deformation. Over time, roads can change the appearance of lane lines due to wear and tear, dirt, rain or snow, increasing the complexity of detection.

(4) The effect of weather changes. Different weather conditions, such as rain, snow or fog, can affect the appearance of the lane or completely cover the lane line, thus affecting the visibility of the lane line.

(5) The impact of complex lanes. Some roads may contain multiple lane lines, intersections or roundabouts, and other complexities, which puts higher demands on lane detection.

3 Traditional Lane Detection Methods

In the development history of lane detection, a series of image processing techniques were used in the early stage, mainly including two aspects of image processing and feature extraction. This section will introduce these two aspects one after another.

3.1 Image processing

Image processing focuses on the alteration of the overall image to adjust the appearance of the image or enhance specific information by using a variety of predefined operations such as filtering, smoothing and edge detection. This approach is robust in manipulating the image as a whole, especially in different environments and scenes. However, it is relatively rigid and not flexible enough to adapt to changes in different tasks, requires manual selection and adjustment of image processing operations, and usually relies on the experience of domain experts. However, with the development of computer vision field, feature extraction methods gradually replace the traditional image processing methods. Among the image processing methods, the main ones are:

(1) Edge detection[18], which detects changes in pixel values in an image in order to determine the edge structure in the image. In lane detection, edge detection helps in identifying the road edges and provides the underlying geometric information.

(2) Color filtering[19], an image processing method that uses color information to emphasize or suppress colors in an image by filtering a specific color channel or setting a color threshold, which helps to emphasize the color of road markings.

(3) Hough transform[20-21], which helps to extract the geometric information of the road by mapping the pixels in the image to the parameter space so as to find the points in the parameter space that share the maximum accumulation to represent the geometric shapes in the image.

(4) Optical flow method[21-22], a technique to obtain motion information by analyzing the motion of pixels in an image between successive frames. In lane detection, the optical flow method can be used to recognize the motion of vehicles and the environment, thus providing information about the location of the lane lines.

(5) Curve fitting[23-24], using a geometric model, is performed by mathematical representation such as polynomials, and optimization techniques such as least squares are used for curve fitting.

3.2 Feature extraction

Feature extraction is concerned with extracting informative local features from an image to capture key information in the image and provide effective input for subsequent tasks. The design of feature extraction is more flexible in that the features used to represent the information in an image can be selected and designed manually. Although manual design of features is required, it is more flexible compared to image processing and does not necessarily require expert level experience. In addition, feature extraction can be computationally expensive depending on the complexity required for the design and computation of the selected features, but is robust in the case of relatively stable local features. Among the feature extraction methods, the main ones are:

(1) Scale-invariant feature transform (SIFT), SIFT is a method for image feature extraction by detecting key points in an image and using local feature descriptors around these key points. In lane detection, SIFT can be used to detect key points in road structure and to describe road texture.

(2) Histogram of oriented gradients (HOG), HOG is a method for extracting image features by analyzing the gradient direction of local regions in an image. In lane detection, HOG can be used to describe the edges and textures in the image to provide strong support for lane detection.

(3) Region of interest (ROI) extraction, Feature extraction is performed by defining and selecting regions of interest in an image and focusing attention on these regions. In lane detection, road regions can be identified by a priori knowledge or image segmentation techniques to reduce computational complexity and increase the reliability of features.

A comparison of image processing and feature extraction features is given in Table 1. The table shows the differences between the two methods by comparing the features of image processing and feature extraction.

Table 1 Features Comparison of Image Processing and Feature Extraction

Characteristics	Image Processing	Feature Extraction
Target	Focus on geometric structure and edge extraction	Capture of local features and textures in images
Robustness	Need to deal with varying environmental conditions	Robust to some local feature stabilizing
Design	Using predefined operations such as filters, edge detectors, etc.	Manual selection and design of feature representations
Cost	Usually faster to compute	Depends on the features design
Flexibility	Not flexible enough to adapt to different tasks	Flexibility to adapt to different tasks
Dependency	Requires human selection and adaptation of image processing operations	Features have to be designed manually
Generalizability	Constrained by a priori knowledge and design	Some generalization capability when local features are stable

4 Deep Learning Based Lane Detection

The emergence of convolutional neural networks [25] marks the shift from traditional methods to deep learning methods [26] in the field of lane detection. Convolutional neural networks bring significant performance improvements to lane detection tasks with their powerful feature extraction capabilities and end-to-end learning. In the face of complex road situations and different environmental changes, CNNs are able to capture and learn the key information in the image more effectively, laying the foundation for accurate and robust lane detection.

4.1 Overview of network model classification

In this subsection, the network models for lane detection based on different methods, including segmentation, object detection, end-to-end, parametric curves and keypoint methods, are explored in depth to reveal in more detail the wide range of applications and advantages of convolutional neural networks in lane detection.

(1) Segmentation methods, through deep learning networks, such as semantic segmentation models, classify images at the pixel level to achieve fine-grained recognition and localization of lane lines. Through convolution and pooling layers, segmentation methods are able to accurately delineate lane regions in an image, providing more precise information for subsequent decision making.

(2) Object detection methods, such as YOLO [27] and Faster R-CNN [28], accurately locate and recognize the lane portion of an image by region detection through convolutional and fully connected layers. object-based detection methods can be categorized as two-stage, single-stage. and anchor-based methods. Choosing the appropriate method can effectively solve the problem of detecting different types of lanes and improve the adaptability and robustness of the model.

(3) The end-to-end methods, by constructing an end-to-end neural network structure, learns the feature representation of lane lines directly from the original image, avoiding the manual feature extraction step in the traditional approach. This approach simplifies the model design process and makes lane detection more intuitive and efficient.

(4) The parametric curves approach, a key step in optimizing the performance of deep learning models. By adjusting the parameters such as the number of layers and nodes of the convolutional neural network, the model is better adapted to different road conditions, thus improving the accuracy of lane detection.

(5) Keypoint methods, which is used to recognize the key feature points of the lane lines in the image, including the start point and the end point. Through convolution and pooling layers, the key point method can effectively capture the important information of lane lines in images.

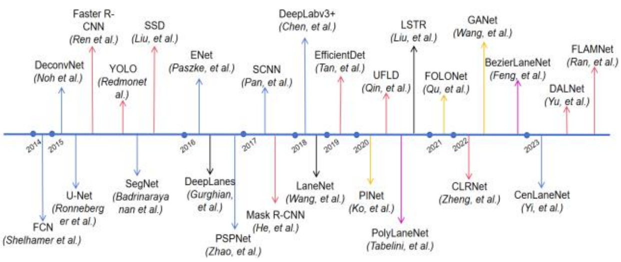

Figure 2 Major methods of deep learning-based network models, including segmentation (in blue), object detection (in red), end-to-end (in black), curve fitting (in pink) and keypoints (in yellow)

In the field of lane lines, many excellent network models have appeared and played an important role. From the traditional convolutional neural networks to the Transformer [29] model that has emerged in recent years, these models provide a variety of options for achieving accurate and efficient lane detection. Figure 2 illustrates typical approaches based on deep learning network models and lists the network models covered in this paper.

4.2 Segmentation methods

Segmentation based methods are based on dividing different pixels into different parts and labeling them according to classification rules. These methods usually include semantic segmentation, instance segmentation and panoramic segmentation.

4.2.1 Semantic segmentation

Semantic segmentation [30-32] aims to assign each pixel in an image to a specific semantic category so that pixels of the same category are labeled as the same region. In lane segmentation, the goal of semantic segmentation is to classify each pixel in an image as belonging to a lane or background. Common semantic segmentation networks include SegNet, U-Net, FCN, and DeconvNet, which achieve pixel-level prediction through convolution and upsampling operations.

In the early days, the use of CNNs greatly simplified the process of extracting image features. In 2014, Jonathan Long et al [33] advanced the field of semantic segmentation by proposing fully convolutional networks. FCN [33] successfully achieved pixel-level semantic segmentation labeling of the entire image through convolution and up-sampling operations. In the following year, a series of network models with encoder-decoder structure emerged to further optimize the semantic segmentation task. Among them, models such as SegNet, DeconvNet, and U-Net became important representatives. SegNet, proposed by Vijay et al [34], utilizes the encoder-decoder structure to regularize the up-sampling and down-sampling operations and improve the accuracy of segmentation. DeconvNet, proposed by Hyeonwoo Noh et al [35], and Olaf U-Net proposed by Ronneberger et al [36] also adopt this structure.

4.2.2 Instance segmentation

The goal of instance segmentation [37] is to split different instances (e.g., different lane lines) in an image and assign unique identifiers to each instance. This method can be used to distinguish between multiple parallel lane lines.

Common lane instance segmentation models include ENet, DeepLabv3+, and LaneNet. ENet was proposed by Adam Paszke et al [38] in 2016, and successfully achieved fast and accurate instance segmentation in resource-constrained scenarios by designing efficient network structures and operations. In the following years, DeepLabv3+ and LaneNet appeared, where DeepLabv3+

is based on an improved version of the DeepLab family, which employs null convolution to expand the sensory field so that it can better capture the contextual information of images. And LaneNet [39], an instance segmentation network focusing on the lane detection task, cleverly combines the ideas of semantic segmentation and instance segmentation, and achieves accurate segmentation of lane lines by assigning each pixel in the image to a lane category and identifying each lane instance. The LaneNet architecture, as depicted in Figure 3, comprises a segmentation branch that creates a binary lane mask, and an embedding branch that generates an N-dimensional embedding for each lane pixel. The embeddings ensure that pixels from the same lane are close together and those from different lanes are far apart. After removing background pixels using the segmentation map, lane embeddings are clustered around their respective centers for identification. Existing models based on instance segmentation still face some challenges on complex scenes in terms of robustness, Yi Sun et al [40] proposed CenLaneNet for robustness. CenLaneNet achieves improved robustness and performance by combining instance segmentation and lane center estimation.

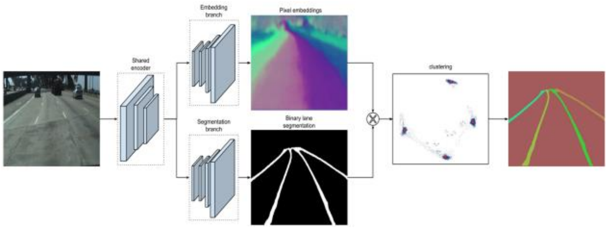

Figure 3 LaneNet architecture framework. The embedding branch is located at the top, while the segmentation branch is located at the bottom. Different lane instances are distinguished using different colors

4.2.3 Panoramic segmentation

Panoramic segmentation [41-42] is the segmentation of an image at a higher level. In lane detection, panoramic segmentation can be used to separate different regions of the entire road (e.g., lanes, sidewalks, lawns, etc.). This helps to understand the road scene more comprehensively. PSPNet [43] is able to capture different scale contextual information and enhance the understanding of the panoramic scene by introducing a pyramid pooling module. The DeepLab family, including DeepLabv3 and DeepLabv3+, allows for detailed panoramic segmentation of the road scene through techniques such as null convolution. UNet employs an encoder -decoder structure that can be used to segment road areas and other environmental elements for panoramic scene understanding.

A comparison of the segmentation-based approaches is shown in Figure 4. the respective goals of interest for semantic, instance and panoramic segmentation are distinguished by color.

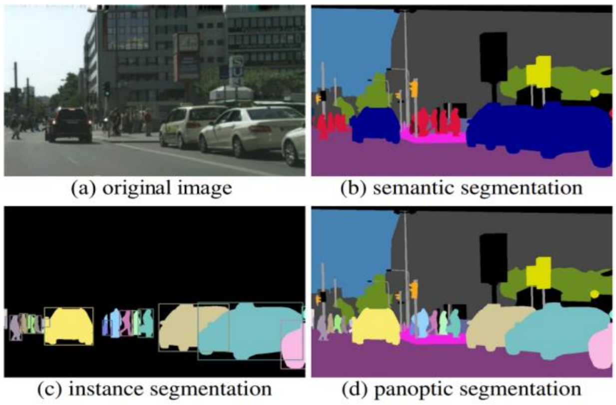

Figure 4 Comparison plot based on segmentation methods. (a) Original image. (b) Region of interest for semantic segmentation. (c) Region of interest for instance segmentation. (d) Region of interest for panoptic segmentation

4.3 Object Detection Methods

4.3.1 Single-stage

Single-stage object detection methods are known for their simplicity and efficiency. These methods usually predict the location and class of the object directly through a single neural network model. YOLO, SSD and EfficientDet are representative algorithms among them.

YOLO [27], as a classical single-stage network, divides the image into grids and performs object detection and localization on each grid while predicting the category of the object. By accomplishing all tasks in a single forward propagation. SSD proposed by W. Liu et al [44] simultaneously predicts objects of different sizes by convolving them at different scales, allowing it to efficiently deal with multi-scale objects. EfficientDet proposed by Mingxing Tang et al [45] focuses on the design of the network in terms of its lightweighting and performance enhancement, which allows it to perform well in the perform well in environments with limited computational resources. It achieves a significant performance improvement on the object detection task by optimizing the network depth and width, as well as adopting a bi-directional feature network structure.

4.3.2 Two-stage

Two-stage object detection methods adopt a staged strategy for object localization and classification. A typical two-stage includes first generating candidate regions and then classifying and regressing the locations of these regions. Mask R-CNN and Faster R-CNN are representatives of two-stage methods.

Mask R-CNN [46],as depicted in Figure 5, achieves instance segmentation based on object detection by introducing additional branches to generate a mask of the object. This enables the model to understand the objects in the image at a finer granularity. Faster R-CNN is a classical two-stage object detection model that generates candidate regions by introducing Region Proposal Network (RPN) and performs classification and location regression in the subsequent stages. It effectively combines the two tasks of candidate generation and object detection, providing a reliable basis for accurate object detection.

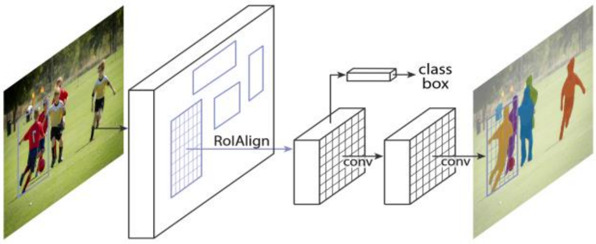

Figure 5 Mask R-CNN framework for instance segmentation

4.3.3 Anchor-based method

In recent years, the application of anchor-based methods has also attracted attention. The main idea of anchor methods is to predefine a set of anchor frames or anchor points through which object detection and localization can be performed. Tu Zheng et al [47] proposed a CLRNet model that first detects lanes with high-level semantic features by combining the lane a priori information, and then performs refinement based on the low-level features [47]. Hao Ran et al [48] proposed FLAMNet, as depicted in Figure 6, with a flexible line anchor mechanism, which enhances local detail extraction and global semantic information modeling by continuously correcting the position of line anchors in order to improve detection performance and computational efficiency. Zichen Yu et al [49] proposed DALNet, a dynamic anchor line-based detection network. This network introduces an innovative dynamic anchor line generator that dynamically generates appropriate anchor lines for each track instance based on the position and shape of the tracks in the input image. These dynamically generated anchor lines not only reflect the positions of the tracks more accurately, but also outperform the predefined anchor lines as positional references. Zequn Qin et al [50] considered lane detection as an anchor-based sequential classification problem, which effectively expands the sensory field and optimizes the complex scene problem by using hybrid anchors to represent the coordinates of the lanes.

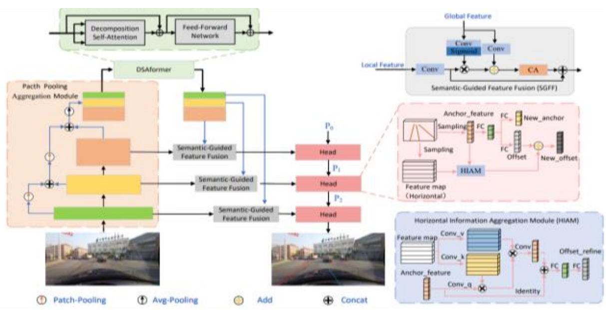

Figure 6 FLAMNet Overview. Features from various layers are pooled and aggregated via the PPAM module. These features then undergo global dependency modeling with the DSAformer, while local and global information fusion is performed using the SGFF module

4.4 End-to-end methods

The end-to-end lane detection method avoids the manual feature extraction step in traditional methods by constructing an end-to-end network model that learns the feature representation of lane lines directly from the original image. This approach enables the model to directly output information such as the location and shape of lane lines.

In recent years, several end-to-end lane detection methods have emerged. Among them, DeepLanes, proposed by Alexandru Gurghian et al [51], employs deep neural networks to reliably estimate the location of lane markers in an end-to-end manner through a classification architecture. On the other hand, Davy Neven et al [52] proposed an instance segmentation method to train an end-to-end detection network containing two branches by transforming the lane detection problem into instance segmentation. In addition, Ruijin Liu et al [53] viewed lane detection as an approximate curve problem, and proposed an end-to-end method that directly outputs the lane shape model parameters, which utilizes the self-attention mechanism of Transformers and is able to learn richer contextual information.

4.5 Parametric curves methods

In lane detection, parametric curve or curve fitting is a common method that describes the shape of lane lines by fitting mathematical curves.

PolyLaneNet, as depicted in Figure 7, proposed by Lucas Tabelini et al [18] uses the idea of polynomial fitting, which describes the curve shape of lane lines by learning polynomial coefficients. This method is more flexible and can adapt to different shapes of lane lines, thus improving the accuracy of detection. LSTR proposed by Ruijin Liu et al [53] combines the ideas of lane line segmentation and tracking, and realizes the parametric curve fitting of lane lines by regression method. The model focuses on modeling the temporal information of lane lines, which in turn improves the adaptability to dynamic scenes. BezierLaneNet proposed by Zhengyang Feng et al [54] uses Bezier curves for fitting lane lines, and represents the shape of the lane lines by learning the control points of the Bezier curves. This method is more flexible in expressing curves and is especially suitable for irregularly shaped lane lines.

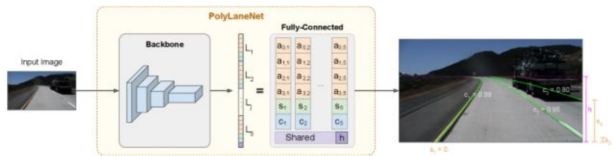

Figure 7 Overview of the PolyLaneNet. From left to right: the model receives as input an image from a forward-looking camera and outputs information about each lane marking in the image

4.6 Keypoint methods

Key point method is a method to achieve lane detection by identifying important feature points of lane lines in an image, such as start and end points. These particular points are critical for understanding the shape and direction of the lane lines, hence the name keypoints. The goal of the keypoint method is to detect and localize these keypoints automatically by using computer vision algorithms or deep learning models to provide detailed information about the location and shape of the lane lines.

In recent years, Yeongmin Ko et al [12] proposed PINet, a deep learning based lane detection method, whose main feature is that it introduces the idea of location awareness and iteration to better capture the location information of lane lines. Through many iterations, PINet gradually optimizes the detection performance for lane lines. In addition, FOLOLane proposed by Z. Qu et al [55] adopts a flow field guided approach to combine the lane detection problem with the flow field task. This unique design allows the model to better understand the dynamic information in the image and improve the accurate detection of lane lines. GANet proposed by Jinsheng Wang et al [56], introduces a novel perspective to lane detection. It directly links each key point to the starting point of its lane line, improving efficiency by eliminating point-by-point extension. This association is achieved by predicting offsets to global lane starting points, enabling parallel processing. Additionally, the Lane Perception Feature Aggregator (LFA) captures local correlations between adjacent key points, enhancing global associations with local information. The GANet architecture is shown in Figure 8.

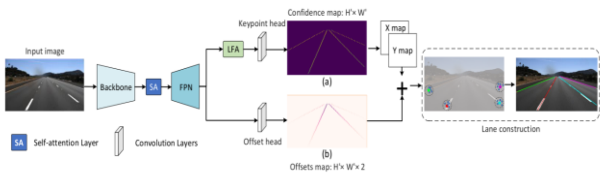

Figure 8 GANet architecture. Confidence map and offset map are combined into group, each representing a lane instance

4.7 Summary of various lane detection algorithms based on deep learning

In this section, we give summarize the approaches, advantages and limitations of representative deep learning based lane detection algorithms. In this paper, we categorize these algorithms into segmentation-based, object detection, end-to-end, parametric curves, and keypoint detection methods. Table 2 demonstrates the representative segmentation methods.

In segmentation-based lane detection methods, this paper divides segmentation methods into three categories according to the goal of processing: semantic segmentation, instance segmentation, and panoramic segmentation. The representative FCN model replaces the fully connected layers with convolutional layers to form a fully convolutional structure, which allows the model to receive input images of arbitrary size and generate dense pixel-level predictions. However, it is worth noting that FCN does not fully consider the relationships between pixels. After the FCN model, the encoder-decoder structure is widely used for semantic segmentation tasks. The SegNet model and the U-Net model are based on the FCN model, which employs an encoder-decoder structure for semantic segmentation tasks. This structure effectively improves the ability to model inter-pixel relationships. DeconvNet employs a convolution-inverse convolution structure, similar to the encoder-decoder structure. For instance segmentation tasks, ENet is a lightweight segmentation network that focuses on high real-time and low latency. However, due to its lightweight structure, for some complex tasks, ENet may perform poorly in terms of adaptability. DeepLabv3+ improves segmentation accuracy by employing null convolution and decoder modules, but requires more computational resources to obtain results. The LaneNet model and CenLaneNet model focus on autonomous driving scenarios, utilizing a multi-branching approach to achieve low Latency. In the panoramic segmentation task, the PSPNet model, based on the FCN model, introduces the pyramid pooling module to establish the connection between local and global features, thus improving the segmentation accuracy. However, the pyramid pooling module increases the computational complexity and memory utilization of the network, resulting in the model requiring more computational resources.

Table 2 Representative segmentation methods

Method	Network	Strengths	Limitations
Segmentation		Semantic Segmentation	
	Fcn[33]	Using Full Convolutional Structure	Does Not Consider Pixel Relations
	Segnet[34]	Encoder-Decoder Structure	Sensitive To Complex Scenes
	Deconvnet	Up-Sampling Using Deconvolutional Layer	Overfitting
	U-Net[36]	U-Shaped Structure With Encoding And Decoding Processes	Requires More Training
		Instance Segmentation	
	Enet[38]	Lightweight Network Structure	Less Generalizable To Complex Scenes
	Deeplabv3+	Introduces Decoder Module	Restricted By Resources
	Lanenet[39]	Focuses On Lane Line Segmentation	Low Scene Generalization
	Cenlanenet [40]	Incorporates Center Line Extraction Idea	Road Detection Needs Improvement
		Panoramic Segmentation	
	Pspnet[43]	Introduces Pyramid Pooling Module	Restricted By Resources
	Deeplabv3+	Panoramic Segmentation Performs Well	Restricted By Resources

In the lane detection method based on object detection, in this paper, we will categorize the lane detection methods based on object detection into single-stage and two-stage approaches, and additionally consider anchor-based methods. Typical single-stage models such as YOLO have fast detection speeds and are relatively weak in small object detection due to the fact that each mesh can only predict one category. SSD employs multi-scale feature maps for detecting objects of different sizes to enhance small object detection, but still performs relatively weakly. EfficientDet uses network scaling to balance the size and performance of the model. Two-stage methods have higher detection accuracy than single-stage methods, but are usually slower and require more computational resources. Typical two-stage methods are Faster R-CNN and Mask R-CNN. Anchor-based method are CLRNet, UFLD and so on. Table 3 demonstrates representative object detection methods.

Table 3 Representative object detection methods

Method	Network	Strengths	Limitations
Object detection		Single-stage	
	YOLO[27]	Fast scenes with high real-time performance	Poor performance on small object inspection
	SSD[44]	Multi-scale feature map detection for variable-size objects	Poor performance on small object inspection
	EfficientDet[45]	Combining efficiency and performance	Poorer performance than some complex models
		Two-stage	
	Mask R-CNN[46]	Introduction of mask segmentation	Complex, high computational requirements
	Faster R-CNN[28]	Classic two-stage	Relatively slow, high hardware requirements
		Anchor-based method	
	CLRNet[47]	Incorporating a priori lane information	May need further adaptability for other object inspection tasks
	FLAMNet[48]	Using line anchors to improve detail and global modeling capabilities	May need further generalization for special scenarios
	DALNet[49]	Allows model to handle different tasks concurrently	Multi-task combination may increase model complexity
	UFLD[50]	Uses lightweight design for model streamlining	Slightly slower to perform in some scenarios

In end-to-end lane detection based approaches, the model maps directly from learning inputs to outputs, simplifying the model tuning process. Representative approaches such as DeepLanes focus on convolutional neural network structure, which reduces manual feature design but is limited by computational resources and dataset size. LaneNet focuses on the lane detection task, which improves the real-time performance of the model, but the training and computational costs are relatively high. LSTR, on the other hand, improves the adaptability to dynamic scenes by introducing a tracking mechanism, but requires more data to train the model and adjust the parameters. Table 4 shows representative end-to-end detection methods.

Table 4 Representative end-to-end detection methods

Method	Network	Strengths	Limitations
End-to-End	DeepLanes[51]	Introduces deep neural network for end-to-end	Variety of training data and scenarios can limit performance
	LaneNet[39]	Focus on lane line detection	Requires more data and computing resources
	LSTR[53]	Directionally output lane shape model parameters	Requires more training data for complex scenarios

Curve fitting methods construct lane shapes by modeling them mathematically. A typical PolyLaneNet model uses polynomials for fitting lanes. Although PolyLaneNet can adapt well to different shapes of lane shapes, the polynomial fitting performance is not flexible enough in complex scenarios such as excessive curvature. Compared with the implementation, BezierLaneNet uses Bessel curves to fit the lanes, which can flexibly adapt to various curve shapes. However, the computation is relatively complex and requires more computational resources. LSTR, on the other hand, can better focus on contextual information to capture lane shapes by introducing a self-attention mechanism, but the model also requires more data to optimize the parameters. Table 5 shows representative parametric curve methods.

Keypoint methods detect lane shape and direction by constructing special points. Typical models include PINet, FOLONet and GANet.

Among them, PINet introduces location awareness and iteration, but there is still room for improvement in its adaptability. It can perform well in ordinary scenes but may be limited when dealing with complex scenes. FOLONet adopts flow field guidance to improve detection accuracy by considering the motion relationship between pixels. Although flow field guidance helps to better understand the dynamic changes in the image, the performance may be low in special scenes. On the other hand, GANet employs generative adversarial networks to improve the robustness and generalization of the model. Although this method performs better in handling complex scenes, the corresponding computational cost is high, which makes it potentially limited in some resource-constrained applications. Table 6 demonstrates representative keypoint methods.

Table 5 Representative parametric curve methods

Method	Network	Strengths	Limitations
Parametric curve	PolyLaneNet[18]	Polynomial fitting of lanes with different shapes	Requires more tuning and optimization for complex scenarios
	LSTR[53]	Focus on global information using self-learning mechanism	Requires more data sets to tune parameters
	BezierLaneNet[54]	Flexibility to fit different curved line shapes	Requires more computing resources when dealing with large data sets

Table 6 Representative key point approach

Method	Network	Strengths	Limitations
Keypoints	PINet[12]	Introduction of location awareness and iteration to incrementally optimize lane line detection performance	Needs to be optimized for complex scenarios
	FOLONet[55]	Introduction of flow field guidance to improve detection accuracy	Lower performance is possible in some special scenarios
	GANet[56]	Introduces generative adversarial network to improve robustness and generalization ability	Higher hardware requirements and higher computational costs

5 Problems and Pspects

Lane detection still faces the impact of datasets and other factors that will directly or indirectly limit the ability of lane detection methods to address complex scenarios, as well as their potential to improve generalization capabilities, real-time performance, and adaptability. The issues of complex scenarios, data diversity, real-time requirements, model generalization, and computational cost are explored in detail below:

(1) In complex scenarios, complex road conditions, multi-lane intersections, road construction and other complex scenarios make the lane detection task more complicated. The lane lines may be curved, bifurcated, widened, etc., which increases the difficulty of the detection algorithms to adapt to irregular shapes. In addition, the existence of a large number of intersections, pedestrian crossings and other scenarios require the lane detection model to maintain high accuracy in these cases, which increases the complexity of the algorithm.

(2) The diversity of data from different geographic locations, road types and traffic environments requires lane detection models to have strong generalization capabilities. Lack of sufficient data training for various complex scenarios can easily lead to model performance degradation in new scenarios, and more samples and more comprehensive datasets are needed to improve the robustness of the model.

(3) Real-time applications such as autonomous driving require high real-time lane detection. Timely and accurate acquisition of lane line information is crucial for decision making and control. Therefore, the model needs to complete the image processing and prediction in a limited time, which puts higher demands on the computational efficiency and speed.

(4) The generalization performance of the model in unknown environments is a key issue in the field of lane detection. Due to changes in actual road and traffic conditions, the model needs to maintain accuracy under different conditions. Fully considering the generalizability of the model so that it can operate stably in various real-world scenarios is an urgent problem to be solved at present.

(5) With the continuous development of deep learning models, some advanced models may have higher requirements on computational resources. The rising computational cost may make it difficult to realize real-time processing in practical applications, especially on embedded devices. Therefore, researchers need to seek more efficient model structures and computational optimization methods while maintaining model performance.

Facing these problems, researchers are constantly seeking innovative solutions. From the algorithmic perspective, some novel deep learning structures and attention mechanisms have been introduced to better handle complex scenarios and improve generalization. In terms of data, more comprehensive and diverse datasets are constructed for training to improve model adaptability. Meanwhile, computational efficiency improvement is also a hotspot in current research, reducing the computational burden through model pruning and lightweight network design.

In future research, more attention needs to be paid to the robustness of the model under extreme conditions, such as extreme weather and light conditions. In addition, further improvement of real-time performance, innovation of data enhancement techniques, and optimization of computational efficiency of deep learning models will be important directions to solve the current problems.

6 Conclusion

This paper provides an exhaustive summary of deep learning-based lane detection algorithms, which is divided into three main parts: First, the background of lane detection is introduced, including lane detection, lane line datasets, and factors affecting lane detection. Second, the traditional methods for lane detection and network models for different deep learning methods are presented, including segmentation-based, object detection, end-to-end, parametric curve, and keypoint methods. In the deep learning methods section, the advantages and limitations of various models are listed in detail. Finally, the complex scenarios, data diversity, real-time requirements, model generalization, and computational costs faced by lane detection are discussed in depth.

Overall, the researchers proposed various innovative solutions, including novel deep learning structures, attention mechanisms, more comprehensive dataset construction, and computational efficiency improvement. Future research directions will focus on addressing model robustness in complex scenarios, improving real-time performance, innovative data augmentation techniques, and optimizing the computational efficiency of deep learning models.

Acknowledgements: This work is funded by the Science and Technology Project of Hebei Education Department (No. ZD2022100).

References

[1] Song Ruitao, Li Bin. Surrounding Vehicles'lane change maneuver prediction and detection for intelligent vehicles: a comprehensive review[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23:6046-6062.

[2] Wang Zhangu, Jun Zhan, C. Duan, et al.. A review of vehicle detection techniques for intelligent vehicles[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023,34:3811-3831.

[3] Lin Lianghao. Research on structured road recognition algorithm based on image processing[D]. Harbin University of Science and Technology, 2021.

[4] Zhong Zebin. An image graying method for lane line recognition[J]. Journal of Tongji University (Natural Science Edition), 2019,47:178-182.

[5] Ahmed Hashem, T. Schlechter. Feature-based lane detection algorithms for track following: a comparative study[C]. 2022 International Conference on Mechatronics and Robotics Engineering, 2022.

[6] Yu Fuxing, Wu Yafeng, Suo Yina, et al.. Shallow detail and semantic segmentation combined bilateral network model for lane detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2023,24:8617-8627.

[7] Guo Rui. Research and exploration of lane line detection based on deep learning[C]. International Conference on Image Processing & Computer Vision (IPCV), 2023.

[8] Zhong Haomai. Research and application of lane line detection method based on deep learning[D]. Xi'an University of Technology, 2023.

[9] Tian Sheng, Zhang Jianfeng, Zhang Yutian, et al.. Lane line detection algorithm based on expanded convolutional pyramid network[D]. Journal of Southwest Jiaotong University, 2020,55:386-392.

[10] Li Xiaoxiao, X. Hu, Z. Wang. Advances in instance segmentation research based on deep learning[J]. Computer Engineering and Applications, 2021,57:60-67.

[11] X. Liu, Hu Xiaoguang, Wang Ziqiang, et al.. Progress of anchor-less frame-based object detection method and its application in complex scenes[J]. Journal of Automation, 2023,49:1369-1392.

[12] Yeongmin Ko, Jun Jiwon, Donghwuy Ko, et al.. Key points estimation and point instance segmentation approach for lane detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2021,23:8949-8958.

[13] Lv Chuan. Research on lane line recognition and tracking method based on key point detection[D]. Northwest A&F University, 2021.

[14] Wang Jinsheng, Ma Yinchao, Huang Shaofei, et al.. A keypoint-based global association network for lane detection[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[15] S. Ettinger, Chen Shuyang , Benjamin Caine, Large scale interactive motion forecasting for autonomous driving: the waymo open motion dataset[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021.

[16] Zhang Youcheng, Lu Zongqing, Zhang Xuechen, et al.. Deep learning in lane marking detection: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022,23:5976-5992.

[17] Wang Xi. Lane line detection algorithm in abnormal driving environment[D]. Beijing Jiaotong University, 2022.

[18] L. Tabelini, R. Berriel, T. Paixao. Polylanenet: lane estimation via deep polynomial regression[C]. Proceedings of 2020 25th International Conference on Pattern Recognition, 2020.

[19] Teng Zhu, J. Kim, D. Kang. Real-time lane detection by using multiple cues[C]. International Conference on Control, Automation and Systems, 2010.

[20] S. Kumar, M. Jailia, S. Varshney. An efficient approach for highway lane detection based on the hough transform and kalman filter[J]. Innovative infrastructure solutions, 2022,7(5):290.

[21] G. Taubel, R. Sharma, J. Yang. An experimental study of a lane departure warning system based on the optical flow

and hough transform methods [J]. WSEAS Transaction on Systems, 2014,13:105.
[22] W. Kruger, W. Enkelmann, S. Rossle. Real-time estimation and tracking of optical flow vectors for obstacle detection [C].Proceedings of the Intelligent Vehicles' 95. Symposium, 1995.
[23] J. Niu, J. Lu, M. Xu, et al.. Robust lane detection using two-stage feature extraction with curve fitting[J]. Pattern Recognition, 2016,59:225-233.
[24] W. Van Gansbeke, B. Brabandere, D. Neven. End-to-end lane detection through differentiable least-squares fitting[C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019.
[25] A. Krizhevsky, X. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2012,60:84-90.
[26] S. Yalabaka, C. Prasad, P. Sanjana. Lane detection using deep learning techniques[c]. Proceedings of 2022 8th International Conference on Signal Processing and Communication (ICSC), 2022.
[27] J. Redmon, S. Divvala, R. Girshick. You only look once: unified, real-time object detection[C]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[28] Ren Shaoqin, He Kaiming, R. Girshick, et al.. Faster r-cnn: towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015,39:1137-1149.
[29] A. Vaswani , N. Shazeer, N. Parmar. Attention is all you need[J]. Neural Information Processing Systems, 2017: 5998-6008.
[30] Wang Biyao, Han Yi, Cui Hangbin, et al.. Image-based road semantic segmentation detection method[J]. Journal of Shandong University, 2023,53:37-47.
[31] K. Muhammad, T. Hussain, H. Ullah. Vision-based semantic segmentation in scene understanding for autonomous driving: recent achievements, challenges, and outlooks[J]. IEEE Transactions on Intelligent Transportation Systems, 2022(23):22694-22715.
[32] Tian Xuan, Wang Liang, Ding Qi. A review of deep learning based image semantic segmentation methods[J]. Journal of Software, 2019,30:440-468.
[33] E. Shelhamer, J. Long, T. Darrell. Fully convolutional networks for semantic segmentation[C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[34] V. Badrinarayanan, A. Kendall, R. Cipolla. Segnet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017,39:2481-2495.
[35] H. Noh, S. Hong, B. Han. Learning deconvolution network for semantic segmentation[C]. Proceedings of the IEEE international conference on computer vision, 2015.
[36] O. Ronneberger, P. Fischer, T. Brox. U-Net: convolutional networks for biomedical image segmentation[J]. Proceedings of Medical Image Computing and Computer-Assisted Intervention, 2015:234-241.
[37] Yang Shuqin, Ma Yuhao, Fang Mingyu, et al.. Instance segmentation based lane line detection for complex environments[J]. Journal of Zhejiang University (Engineering Edition), 2022,56:809-815.
[38] A. Paszke, A. Chaurasia,S. Kim. Enet: a deep neural network architecture for real-time semantic segmentation[J]. arXiv, 2016,1606.02147.
[39] Wang Ze, Ren Weiqiang. Qiu Qiang. Lanenet: real-time lane detection networks for autonomous driving[J]. arXiv, 2018,1807.01726.
[40] Yi. S, Jun Yu Li, Xu Xin. Adaptive multi-lane detection based on robust instance segmentation for intelligent vehicles[J]. IEEE Transactions on Intelligent Vehicles, 2022,8:888-899.
[41] Xu Pengbin, Qu Anguo, Wang Kunfeng. A review of panoramic segmentation research[J]. Journal of Automation, 2021,47: 549-568.
[42] Wang Hai, Qiu Meng, Cai Yingfeng, et al.. Sparse u-pdp: a unified multi-task framework for panoptic driving perception[J]. IEEE Transactions on Intelligent Transportation Systems, 2023,24:11308-11320.
[43] Zhao Hengshuang, Shi Jianpin, Qi Xiaojuam, et al.. Pyramid scene parsing network[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
[44] W. Liu, D. Anguelov, D. Erhan. Ssd: single shot multibox detector[C]. European Conference on Computer Vision (ECCV), 2016.
[45] Tan Mingxing, Pang Ruoming, Q. Le. Efficientdet: scalable and efficient object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020.
[46] He Kaiming, G. Gkioxar, P. Dollár. Mask r-cnn. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)[C]. 2017.
[47] Zheng Tu, Huang Yifei, Liu Yang, et al.. Clrnet: cross layer refinement network for lane detection[C]. Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
[48] Ran Hao, Yin Yunfei, Huang Faliang, et al.. Flamnet: a flexible line anchor mechanism network for lane detection[C]. IEEE Transactions on Intelligent Transportation Systems, 2021.
[49] Yu Zichen, Liu Quanli, Wang Wen, et al.. Dalnet: A rail detection network based on dynamic anchor line[J]. arXiv, 2023,2308.11381.
[50] Qin Zequn, Zhang Pengyi, Li Xi. Ultra fast deep lane detection with hybrid anchor driven ordinal classification[J]. IEEE transactions on pattern analysis and machine intelligence, 2022:1-14.
[51] A. Gurghian, T. Koduri, S. Bailur. Deeplanes: end-to-end lane position estimation using deep neural networks[C]. 2016 IEEE Conference on Computer Vision and Pattern

Recognition Workshops, 2016.
[52] D. Neven, B. De, S. Georgoulis. Towards end-to-end lane detection: an instance segmentation approach[C]. 2018 IEEE intelligent vehicles symposium(IV), 2018.
[53] Liu Tuijin, Yuan Zejian, Liu Tie, et al.. End-to-end lane shape prediction with transformers[C]. Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021.
[54] Feng Zhengyang, Guo Shaohua, Tan Xin, et al.. Rethinking efficient lane detection via curve modeling[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
[55] Z. Qu, Jin Huan, Zhou Yang, et al.. Focus on local: detecting lane marker from bottom up via key point[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
[56] Wang Jinsheng, Ma Yinchao, Huang Shaofei. A keypoint-based global association network for lane detection[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.