ISPRS Journal of Photogrammetry and Remote Sensing 197 (2023) 122–144

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

# Road object detection for HD map: Full-element survey, analysis and perspectives

Zhipeng Luo [a,b,c], Lipeng Gao [d], Haodong Xiang [b], Jonathan Li [c,e,*]

[a] School of Computer Science, Minnan Normal University, Zhangzhou 363000, Fujian, China
[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong 999077, China
[c] Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Informatics, Xiamen University, Xiamen 361005, China
[d] School of Software, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China
[e] Departments of Geography and Environmental Management, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

*Keywords:*
High-Definition map
Autonomous driving
Road surface information extraction
Road object detection
3D point clouds

ABSTRACT

As the key part of autonomous driving (AD), High-Definition (HD) map provides more precise location and rich semantic information than the traditional map. With the development of AD, it is eager to construct HD map accurately and efficiently. Road scene perception, especially the road object detection, plays important role in HD map generation. Recently, numerous methods have been undertaken for different tasks in constructing HD maps. To stimulate further research, this paper presents a comprehensive review of recent progress in road surface's information extraction and road object detection. For road surface information extraction, it covers three main categories: road pavement, marking and line. For road object detection, it focuses on the full-element road object detection, including sign-objects, pole-like objects, guardrails, street trees and moving objects detection. In addition, the common public datasets and corresponding results are provided. Besides, for the future research directions, we also summarize the current development trends and topics.

## 1. Introduction

Recently, autonomous driving (AD) has attracted more and more attention from industries and researchers. Correspondingly, it is becoming more and more eager to construct High-Definition (HD) map accurately and efficiently (Elhousni et al., 2020; He et al., 2021; Mi et al., 2021a). As an essential part for AD, HD map provides rich location and semantic information, which enables autonomous vehicles to obtain the perception of surroundings (Li et al., 2021) and localize themselves (Vivacqua et al., 2018; Engel et al., 2021).

Generally, as shown in Fig. 1, there are four main steps for the HD maps construction, i.e., data acquisition, data preprocessing, road object detection and road surface information extraction, and storing the information and visualization. Mobile laser scanning (MLS) systems and the camera systems are usually used for HD map data acquisition. Data preprocessing is the key step. Generally, it includes density filtering, noise reduction, unwanted points deletion, down sampling and registration. The point cloud registration stage usually contains the alignment, stitching, calibration and coordinate registration. More details can be obtained from the existing review works (Xie et al., 2019; Huang et al., 2021a). Apart from the point cloud registration, there is image-to-point cloud registration. Since the detected objects from images cannot get the precision location, in data preprocessing, image-to-point cloud registration is usually conducted to resolve this issue. As the pioneering work, Feng et al. (2019) provides the general image-to-point cloud registration framework, the 2D3D-MatchNet. It firstly extracts image and point cloud key-points, then a siamese-like network is utilized to construct the deep model. Another work is Yu et al. (2020) that takes the 2D-3D line correspondences between images and prior LiDAR maps. In addition, authors in (Pham et al., 2020; Cattaneo et al., 2020) provided image-to-point cloud registration methods for place recognition and retrieval tasks. Recently, Li and Lee (2021) proposed a new framework, DeepI2P, for the image-to-point cloud registration, which does not rely on another localization system. As to the road object detection and road surface information extraction, they are the two key tasks in HD maps construction. After objects detection, all the related information must be stored and visualized. The common software frameworks are OpenDrive, Navigation Data Standard(NDS), and OpenStreetMap(OSM).

In this work, we will mainly introduce the road surface information extraction and road object detection. The road surface information

* Corresponding author at: Departments of Geography and Environmental Management, University of Waterloo, Waterloo, Canada.
*E-mail address:* junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.isprsjprs.2023.01.009
Received 30 June 2022; Received in revised form 4 December 2022; Accepted 13 January 2023
Available online 7 February 2023

includes different kinds of road pavement, road markings, and driving lines. While road objects information can provide rich description of road targets, such as the traffic signs, road signs, traffic lights and guardrails. As shown in Fig. 2, for road surface information extraction, we mainly review three categories: the road pavement, markings and road lines. For road object detection, as shown in Fig. 3, elements in HD map can be divided into three main categories: traffic elements (including traffic sign and traffic light), road elements (including road sign, green belt, and guardrail), and support elements (including the street lamp and tree). Besides, although moving objects are not part of HD map, removing the moving objects can make the algorithms to extract useful information more robust. Therefore, moving object detection should also be paid attention. More specifically, the traffic element is the key part of HD map. Traffic signs and lights provide important information for autonomous vehicle to obtain its surrounding perception. Road elements are the important components for HD map. Road signs provide the road information to guide the vehicles driving in right directions. Green belts and guardrails indicate the drivable area for vehicles. Support elements ensure the safety of automatic driving, while the moving elements must to be removed from the road environments to obtain a clean HD map. Therefore, detecting the above elements efficiently and accurately is of importance for HD map construction.

In the last decade, numerous works have been undertaken to detect the above elements for HD map. These works usually focus on a special single kind element, such as the sign, the street lamp, and the guardrail. Besides, some common datasets have also been released (Geiger et al., 2012; Houston et al., 2020; Sun et al., 2020; Caesar et al., 2020), which boosted the development of HD map. Several survey works have also been provided on special road element detection, such as the moving object detection (Guo et al., 2021b) and the traffic sign detection (Liu et al., 2019a). Different from these works, this paper provides a comprehensive review for full-element object detection. Fig. 2 presents a taxonomy of existing road information extraction and road object detection methods for HD map. It is worth noting that, in this work, methods that provide classic frameworks, which are utilized by several later methods as the backbones, will be selected as the milestone methods. In addition, we will also select the recent novel methods with good/ competitive performance. Compared with the existing survey works, this paper enjoys the following major contributions:

1) To the best of our knowledge, this is the *first full-element* survey paper for HD-map to comprehensively cover the important existing road surface information extraction and road object detection methods.
2) This paper covers the *most recent and advanced* works in HD map related fields, which can provide readers with the comprehensive and the state-of-the-art methods.
3) *Insightful observations and analysis* of these existing methods are provided, and deep discusses for the future research directions are also presented.

This paper is organized as follows. Section 2 provides a survey for the related datasets and common evaluation metrics. Section 3 introduces the road surface information extraction. Sections 4-8 review road object detection methods with full-element, including the signs, guardrails, pole-likes, street trees, and moving objects. It is worth noting that these five parts do not stand alone. Some methods for detecting vehicles can be used to detect trees, while some methods for detecting trees can be used to detect pole-like objects. Especially for image based detection tasks, many methods can be shared. In this work, we choose these methods mainly according to what targets the author aims at when proposing these methods. Section 9 concludes this paper.

## 2. Dataset and evaluation metric

### 2.1. Dataset for road surface information extraction

In this section, we will introduce several datasets for road pavement, markings and road lines extraction.

Several datasets are released to further support the development of road pavement extraction. Images in these datasets are usually collected from DigitalGlobe WorldView satellites or Google Earth images. The ground truth of roads is manually marked by the experts in the field of remote sensing image interpretation. Table 1 provides detail information for these datasets.

The common elements of road marking include solid lines, dashed lines, zebra crossing lines, turning arrow, and so on. As shown in Table 1, Wu and Ranganathan (2012) provides a dataset with 11 categories. It consists of 1443 road images with a pixel resolution of 800 × 600 containing manually marked road markings. Additionally, Ye et al. (2020) presents a 3D point clouds datasets, which was collected in Xiamen, China. It contains 2.961 billion points with a size of 19.7 GB and stored in 17 LAS files.

Several datasets have also been provided to evaluate the performance of road lane extraction methods. There are four main datasets: Cityscape, Vistas, Apollo and Tusimple dataset. Table 1 presents the specific information. Cityscape dataset provides semantic segmentation labels, but no lane information. Vistas dataset provides scene labels and some general categories of lane markings. Apollo dataset provides very detailed pixelated lane information and lane attributes, including 6 types of demarcation markings, 4 types of guide markings, 2 stop lines, 12 types of turn markings. Tusimple dataset is a widely used open-source dataset consisting of 3626/2782 training/test images.

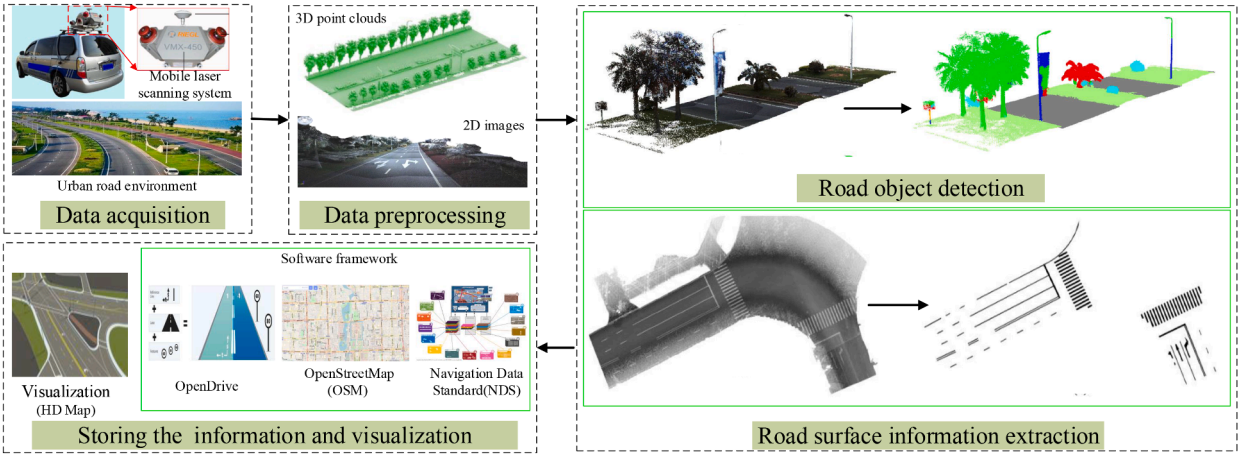



**Fig. 1.** Four main steps for the HD maps construction.

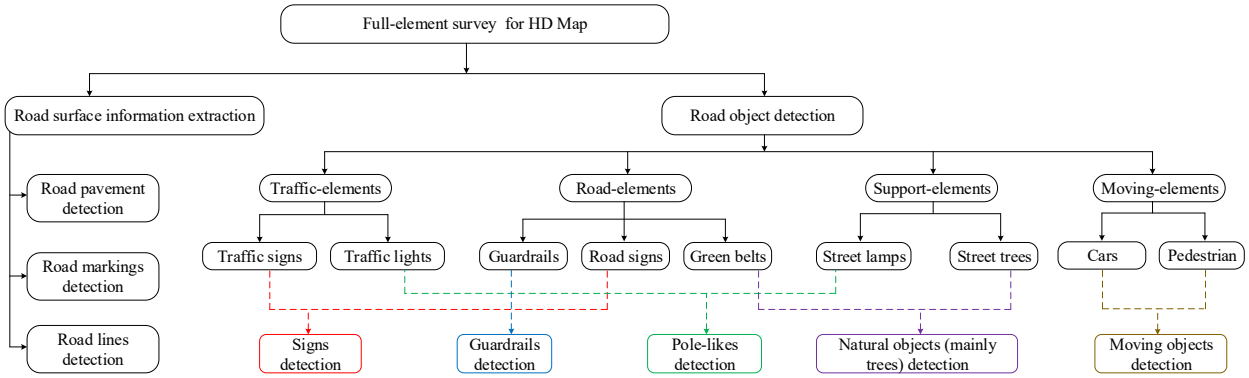



**Fig. 2.** The taxonomy of road surface information extraction and object detection methods for HD map with full-element.

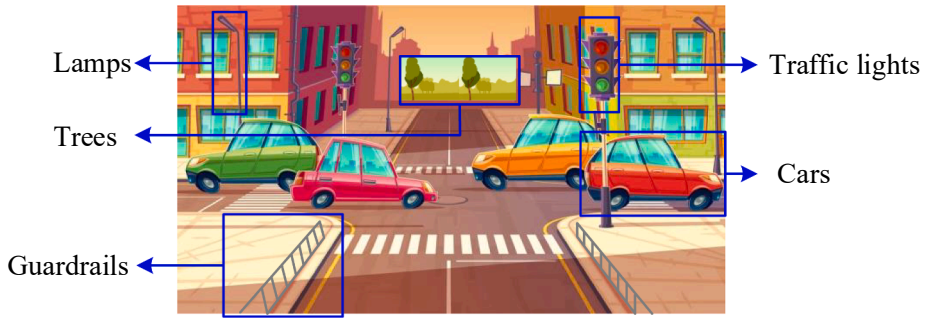



**Fig. 3.** Road elements for HD map: lamps, trees, guardrails, traffic lights, cars, etc.

**Table 1**
Common datasets used for road surface information extraction.

| Name(reference) | Task | Modality | size (distance/ area/images) |
|---|---|---|---|
| Massachusetts Roads Dataset (Mnih, 2014) | Pavement detection | Image | 2600 $km^2$ |
| TorontoCity Dataset (Wang, et al, 2016) | Pavement detection | Image + point cloud | 8439 km/~700 $km^2$ |
| NWPU-RESISC45 dataset ( Cheng et al., 2017a) | Pavement detection | Image | 31,500 images |
| DeepGlobe (Demir et al., 2018) | Pavement detection | Image | 2220 $km^2$ /8570 images |
| ISPRS semantic labeling benchmark (Potsdam, 2018) | Pavement detection | Image | 4.8 $km^2$ |
| SpaceNet 3 (Van Etten et al., 2019) | Pavement detection | Image | 8677 km/3011 $km^2$ |
| Wu and Ranganathan (2012) | Markings extraction | Image | 800 × 600 |
| Ye et al. (2020) | Markings extraction | Point cloud | 19.7 GB |
| Cityscape (Cordts et al., 2016) | Lines extraction | Image | 1024 × 2048 |
| Vistas (Neuhold et al., 2017) | Lines extraction | Image | varies |
| Apollo (Huang et al., 2018) | Lines extraction | Image | 2710 × 3384 |
| Tusimple (Shirke and Udayakumar, 2019) | Lines extraction | Image | 1280 × 720 |

### 2.2. Dataset for road object detection

In this section, we first introduce the existing traffic sign detection datasets. Several datasets were used for the evaluation of TSD methods. These datasets can be divided into three classes: image-based datasets, point cloud-based datasets, and multi-modality (point cloud and image)-based datasets. Traffic signs in image/point-cloud based datasets are shown with different resolution images/point-clouds, while objects in multi-modality datasets contain both image and point cloud representations, which would provide additional information, such as the location and size of traffic sign. Considering that the sign recognition is of great importance in practice, some of these datasets are also collected for recognizing different kinds of traffic signs. Table 2 shows several typical datasets, with summarizing their attributes. Then we introduce the existing datasets for other road objects detection, as shown in Table 3. Some datasets are released to further support the development of guardrails detection. These datasets can be divided into three categories: image-based datasets, 3D point cloud-based datasets, and multi-modal-based datasets. Table 3 provided detailed information for these datasets. For image-based datasets, Seibert et al. (2013) provided a dataset for guardrails classification. In Scharwchter et al. (2014), a large-scale dataset is presented for guardrails detection. This dataset consists of 1968 annotated images with 4934 individual regions marked as guardrails. For 3D point cloud-based datasets, by using a Velodyne HDL-32E LiDAR, in Masuda et al. (2013), authors provided a 3D point cloud-based dataset for guardrail detection. They used a MMS X-640, which has 4 laser scanners (Sick LMS 291), and can measure front upward, front downward, rear upward, and rear downward directions. In addition, Jiang et al. (2016) provided a dataset in 3D point clouds, which contains 42,317 scans corresponding to the approximately 30 km vehicle trajectory. Recently, authors in Gao et al. (2020) also proposed a dataset based on MLS point clouds. They acquired this dataset using a RIEGL VMX-450 scanning system. For multi-modal based datasets, Zhu and Guo (2018) developed a dataset in both images and 3D point clouds representation, which is generated from KITTI dataset. In Matsumoto et al. (2019), authors provided a dataset with images and 3D point clouds. The 3D point clouds are collected by the Mitsubishi MMS-X, which is equipped by a RIEGL VQ 250 laser scanner.

Since most existing pole-like object detection (POD) methods are designed for 3D point clouds, several 3D point cloud datasets have been provided for the evaluation of these methods. Classical datasets include

**Table 2**
The description for details of existing TSD datasets in images, point clouds, and multi-modality representations. The "size" means the resolution of image. "Det." and "Rec." denote as detection and recognition, respectively.

| Datasets for TSD in image representation | | | | | | | |
|---|---|---|---|---|---|---|---|
| Name/reference | Year | # Images (training/test) | # Signs (training/test) | # Class | Size | Country | Det./ Rec. |
| RUG (Grigorescu and Petkov, 2003) | 2003 | 48 | 60 | 3 | 360 × 270 | Netherland | ✓/✓ |
| BTSD (Timofte et al., 2009) | 2009 | 25,634 (21,950/3684) | 4627 (2682/1945) | 62 | 1628 × 1236 | Belgium | ✓/× |
| BTSC (Timofte and Gool, 2011) | 2011 | 7125 (4591/2534) | 7125 (4591/2534) | 62 | 26 × 26 to 527 × 527 | Belgium | ×/✓ |
| STS (Larsson and Felsberg, 2011) | 2011 | 20,000 | 3488 | 7 | 1280 × 960 | Sweden | ✓/✓ |
| Stereopolis (Paparoditis et al., 2012) | 2012 | 847 | 251 | 10 | 960 × 1280 | France | ✓/✓ |
| GTSRB (Stallkamp et al., 2012) | 2012 | 51,840 (39,210/12,630) | 51,840 (39,210/12,630) | 43 | 15 × 15 to 250 × 250 | Germany | ×/✓ |
| GTSDB (Houben et al., 2013) | 2013 | 900 (600/300) | 1206 (846/340) | 43 | 1360 × 800 | Germany | ✓/× |
| LISA (Mgelmose et al., 2015) | 2015 | 5638 (3307/2331) | 6094 (3672/2422) | 35 | 1280 × 960 | United States | ✓/✓ |
| TT100K (Zhu et al., 2016) | 2016 | 100,000 | 30,000 | 45 | 2048 × 2048 | China | ✓/✓ |
| ETSD (Gmez Serna and Ruichek, 2018) | 2018 | 82,476 | 82,476 | 164 | 6 × 6 to 780 × 780 | Europe | ×/✓ |
| **Datasets for TSD in 3D point clouds** | | | | | | | |
| Name/reference | Year | Distance (km) | # Points (billion) | # Signs | Size (GB) | Country | Det./ Rec. |
| XMD (Huang et al., 2017) | 2017 | 23.68 | 1.5 | 274 | 47.38 | China | ✓/× |
| **Datasets for TSD in joint-image-point cloud** | | | | | | | |
| Name/reference | Year | Length (km) | # Points (billion) | # Signs/Image | Size (GB) | Country | Det./ Rec. |
| HDR (You et al., 2019) | 2018 | 5.03 | 571 | 138/3011 | 14.89 | China | ✓/✓ |
| WPR (You et al., 2019) | 2018 | 4.45 | 522 | 125/2572 | 14.06 | China | ✓/✓ |
| XHR (You et al., 2019) | 2018 | 11.13 | 1030 | 266/8220 | 33.7 | China | ✓/✓ |

the TIQmulus and TerraMobilita Contest (TITMC) dataset (Vallet et al., 2015), Youyi square (YYS) datasets (Huang et al., 2021b), and Xinxi library (XXL) datasets (Huang et al., 2021b) in Wuhan University. TITMC dataset is large and complex, with high-quality ground truth for each point. It is the most common dataset used for the evaluation of POD detection methods. YYS dataset mainly contains street lamps and various trees, with dimensions of 128 m × 153 m. The number of points is more than 5,000,000, with a density of approximately 279 points/$m^2$. For XXL dataset, two main pole-like objects are street lamps and sycamore trees. The number of points is more than 12,000,000 and the density is about 2822 points/$m^2$.

There are several datasets that are used for the evaluation of tree detection methods. They can be divided into two main categories: image- and point cloud- based datasets. Table 3 lists some typical datasets with images or 3D point clouds representation. There are relatively few image-based datasets for street tree detection. The recently proposed dataset in image presentation comes from Xie et al. (2020). Authors captured images from 22 roads in the City of Nanjing in China by using a street-view collection vehicle equipped with a camera with five fisheye lenses. This dataset is labeled manually, assisted with the *Labelimg*. It includes 2919 labeled images, with 8297 ground truth bounding boxes. For 3D point cloud-based dataset, Yue et al. (2015) provided a dataset, which was collected using the vehicular 3D movement measurement system V-Surs of the Qingdao mountain mobile measuring company. Vo et al. (2015) provided a dataset. It was acquired from different areas of Shanghai. Li et al. (2016b) provided a dataset for street tree detection. The area is on a street called'California Sunshine Food Street' in Optics Valley in Wuhan. This street is approximately 300 m long, and has a slight slope with about 2.4 m elevation difference between the two ends of the street surface. A total of 66 trees, which comprise several species, are presented along both sides of the street. Researchers from Xiamen University also provided some datasets for street tree detection. In Rastiveis et al. (2020), authors used a MLS system to acquire the 3D point cloud in Xiamen, China. There are many trees that are close to street light poles, which causes that the extraction of trees is difficult. Besides, authors in Zhong et al. (2017) provided two datasets acquired by SSW Mobile Laser Modeling and Surveying System developed by the Chinese Academy of Surveying and Mapping. Other classic datasets are provided in Zhong et al. (2013) and Yadav et al. (2018).

Several 3D point cloud datasets have been provided for the evaluation of moving target extraction methods. The most popular one is the KITTI dataset (Geiger et al. 2012), which contains three main categories, cars, pedestrians, and cyclists with three levels of difficulty: easy, moderate, and hard. Recently, several works also provide important benchmarks, including Houston et al. (2020), Sun et al. (2020) and Caesar et al. (2020). Most of these datasets are mainly for scene perception, so they often contain a lot of additional information, such as the images and radar waves. Table 3 provides the details for these datasets.

### 2.3. Common evaluation metric

To evaluate the road information extraction and road object detection results, several metrics have been provided. Precision, recall, F1-values, and IoU (Intersection over Union) are the most common evaluate criteria:

$$\begin{aligned} Precision &= \frac{TP}{TP+FP}\times 100\%, \\ Recall &= \frac{TP}{TP+FN}\times 100\%, \\ F1\_score &= \frac{2TP}{2TP+FP+FN}\times 100\%, \\ IoU &= \frac{Area\ of\ predicting\ bounding\ boxes}{Area\ of\ ground-truth\ bounding\ boxes}\times 100\%, \end{aligned} \tag{1}$$

where TP, FP, TN, and FN denote the true positive, false positive, true negative, and false negative, respectively.

In addition, the completeness and correctness are used in road pavement extraction. They are defined as follows:

$$\begin{aligned} Completeness &= \frac{TP}{TP+FN}\times 100\%, \\ Correctness &= \frac{TP}{TP+FP}\times 100\%, \\ Quality &= \frac{TP}{TP+FP+FN}\times 100\%, \end{aligned} \tag{2}$$

where the ground truth centerline pixels are defined as the true TP if they lie within a buffer of width H (a given width, for example, H = 5 pixels) around the estimated centerline, and FN otherwise. Estimated centerline pixels are TP if they lie within H pixels of the ground truth

**Table 3**
Common datasets used for road object detection (including guardrails, pole-like objects, street trees and moving objects).

| Name(reference) | Task | Modality | Description |
|---|---|---|---|
| Seibert et al. (2013) | Guardrail detection | Image | N/A |
| Scharwchter et al. (2014) | Guardrail detection | Image | 1968 images |
| Masuda et al. (2013) | Guardrail detection | Point clouds | N/A |
| Jiang et al. (2016) | Guardrail detection | Point clouds | 42,317 scans |
| Gao et al. (2020) | Guardrail detection | Point clouds | 18 pieces |
| Zhu and Guo (2018) | Guardrail detection | Image + Point clouds | 1000 frames |
| Matsumoto et al. (2019) | Guardrail detection | Image + Point clouds | 3912 + 777 |
| TITMC (Vallet et al., 2015) | Pole-like detection | Point clouds | By Wuhan University |
| YYS (Huang et al., 2021b) | Pole-like detection | Point clouds | 279 points/$m^2$ |
| XXL (Huang et al., 2021b) | Pole-like detection | Point clouds | 2822 points/$m^2$ |
| Xie et al. (2020) | Street tree detection | Image | 2900 images |
| Zhong et al. (2013) | Street tree detection | Point clouds | More than 10 categories |
| Yue et al. (2015) | Street tree detection | Point clouds | 1,203,881 laser spots |
| Vo et al. (2015) | Street tree detection | Point clouds | 95 points/m2 |
| Vo et al. (2015) | Street tree detection | Point clouds | 240 m long; 113 points/m2 |
| Li et al. (2016b) | Street tree detection | Point clouds | 300 m long; 123 points/m2 |
| Yue et al. (2015) | Street tree detection | Point clouds | 120 m long; 306 points/m2 |
| Rastiveis et al. (2020) | Street tree detection | Point clouds | 311 trees |
| Zhong et al. (2017) | Street tree detection | Point clouds | 2821 pts/m2; |
| Zhong et al. (2017) | Street tree detection | Point clouds | 222 pts/m2; |
| Yadav et al. (2018) | Street tree detection | Point clouds | 57 m long; 1,240 points/m2; |
| KITTI (Geiger et al., 2012) | Moving object detection | Image + Point clouds | 50 Scenes |
| Waymo (Sun et al., 2020) | Moving object detection | Image + Point clouds | 1000 Scenes |
| nuScenes (Caesar et al., 2020) | Moving object detection | Image + Point clouds | 1000 Scenes |
| Houston (Houston et al., 2020) | Moving object detection | Image + Point clouds | 170 k Scenes |

centerline, or FP otherwise.

## 3. Road surface information extraction

Road surface information is an important part of the HD map. It mainly includes road pavement, markings, and road lanes.

### 3.1. Road pavement extraction

#### 3.1.1. Methods and analysis

Heuristic-based methods are the mainstream in the early years, and they extract road features using the prior of road regions. The road features include geometric feature, spectral feature, topological feature, and texture feature. Heuristic-based methods first extract line features or point features of lane from the images by expert knowledge or regions of interest, then filter or enhance the detected lane features using Kalman filtering, particle filtering, or Gaussian filtering, and extract the boundary information of the lanes (Lee et al., 2017a; Shin et al., 2015; Kortli et al., 2016). Expert knowledge or empirical features have a large impact on the results. We subdivide these methods into two types: segmentation-based and tracking-based.

**Segmentation based methods** consider road pavement extraction as the segmentation problem. The path trajectory point and the angle-based texture feature of a particular pixel were defined to quantify road probability based on shape (Hu et al., 2007; Zhang and Lin, 2013). Das et al. (2011) adopted the spectral and local linear features of multispectral road images. In this paper, the author classifies the areas by combining the probabilistic support vector machine (PSVM) method, the dominant singular value method, the local gradient function, and the vertical central axis transformation method, detects the road edge, breaks the connection, and eliminates the non-road areas. Wegner et al. (2013) developed a higher-order CRF formulation for road pavement extraction based on superpixel segmentation, in which the prior is represented by long-range cliques with robust PN-Potts potentials. Cheng et al. (2014) firstly used SVM soft classification to divide the image into road and non-road; then, the probability of each pixel belonging to the road was obtained simultaneously. The final road was acquired through the graph cut method. Cheng et al. (2016) proposed to use fused multiscale collaborative representation and graph cuts to firstly segment the image. The initial contour of the road was then obtained by filtering the road shape. Finally, the road centerline was obtained through tensor voting. Alshehhi and Marpu (2017) introduced a hierarchical graph-based method based on color and shape features. However, these methods usually work well in multispectral images and detect main roads in urban areas. Extracting roads from areas with dense buildings or no-urban areas remains a challenge.

**Tracking-based methods** regard road pavement extraction as the tracking problem of road seed points. Hu et al. (2004) proposed a segmented parabolic model to delineate road centerline networks. Miao et al. (2014) proposed a kernel density estimation method combined with the geodesic method to decrease the number of seed points required for road pavement extraction. Zhou et al. (2006) used particle filtering to track road segments between seed points. However, particle filtering is limited by its incapability to effectively deal with road branches. To extend the generalization capability of particle filtering to complex scenarios, Movaghati et al. (2010) integrated particle filtering with extended Kalman filtering. Poz et al. (2012) proposed a semi-automatic method to extract urban and suburban roads from stereoscopic satellite images. This method uses seed points to construct the road model in the object space. Optimal road segments between seed points are then generated through dynamic programming. Lv et al. (2017) proposed a multiply feature sparsity based model that can utilize multiple features complementation to extract roads from high resolution imagery. To improve the efficiency of road seed point connection, Gao et al. (2018b) proposed to treat road seed point connection as the shortest path problem. Since this method uses edge features as constraints, it can mitigate the influence of factors such as shadows on road centerline extraction and improve the accuracy of road pavement extraction, but the method needs to be improved for the extraction of regions where road features are not obvious. To obtain the road centerline, Miao et al. (2019) applied the geodesic method on the hue saturation value color space instead of red green blue color space, and achieved better performance. These methods can achieve high precision. However, many required seed points will affect the efficiency.

In addition, DL-based methods have developed rapidly in the past decades and attracted the attention of experts in the field of road pavement extraction. Fig. 4 shows several important methods. It can be observed that most of the recently developed algorithms are based on FCNs with an encoder-to-decoder architecture. Inspired by autonomous mobile systems, Wang et al. (2015) presented a neural-dynamic tracking

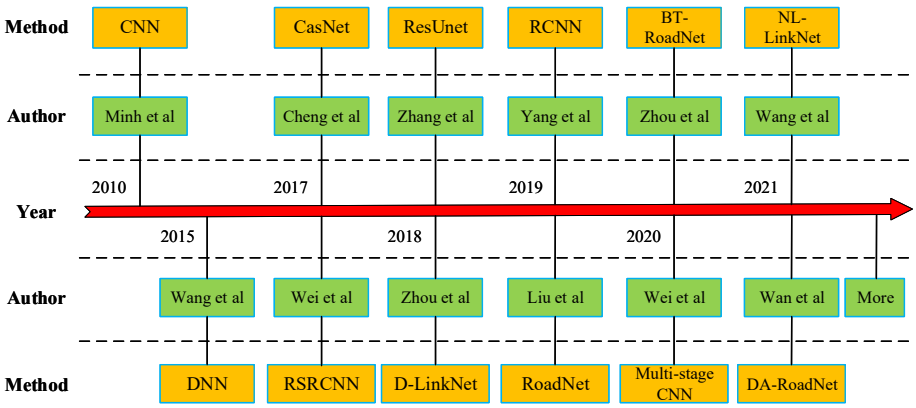



**Fig. 4.** Some representative road pavement extraction methods based on DL.

framework to extract road networks. Although it has achieved good results, a lot of work is consumed in the generation of training data sets. To simplify the training of deep neural networks, Zhang et al. (2018) proposed the residual U-Net, which combines the advantages of U-Net with residual blocks. To extract roads of various widths, Gao et al. (2018a) proposed a multi-feature pyramid network (MFPN). Chen et al. (2021) proposed a reconstruction bias U-Net for road pavement extraction from high-resolution optical remote sensing images. D-LinkNet (Zhou et al., 2018) is an efficient method in comprehensive performance of road pavement extraction, which won first place in the 2018 DeepGlobe Road Extraction Challenge. Despite its advantages, the down sampling process of the encoder module usually leads to a reduction in the boundary accuracy of the road pavement extraction results. To further improve the accuracy, scholars proposed the cascaded neural networks based on multi-task learning, such as (Zhou et al., 2020; Liu et al., 2019b; Shao et al., 2021; Lu et al., 2019). These approaches use mutually constrained multiple tasks to achieve further improvement in road pavement extraction accuracy. However, in the case where the roads are obscured by vegetation or shadows, the contextual information modeling approach based on local receptive fields cannot build the topological relationships between the road segments separated by trees or shadows, which leads to interrupted results of road pavement extraction.

Several methods with post-process have also proposed. One of the solutions for the interruption of road pavement extraction results is to post-process the pixel-level segmentation results of roads. Chen et al. (2018) proposed a two-stage approach combining road edge features and area features, which applies connection analysis to discrete line elements with directional consistency to extract potential road objects; potential road objects are then evaluated by shape features to refine the road pavement extraction results. Gao et al. (2019) first used a semantic segmentation model to obtain pixel-level road segmentation results, and then used tensor voting (TV) to connect the breaks. Inspired by human behavior of annotating roads, Batra et al. (2019) proposed a stacked multi-branch convolution model and connectivity optimization method based on Orientation Learning. Tao et al. (2019) focused on the modeling of road context information and put forward a well-designed spatial information reasoning structure. Wei et al. (2020) introduced a DL-based multistage framework to accurately extract the road pavement and road centerline simultaneously. The idea of generative adversarial networks (GANs) is also adopt in several works. Zhang et al. (2019b) developed a novel road pavement extraction method based on an GAN. Zhang et al. (2019c) proposed a multi-supervised GAN, which learns how to reconstruct obscured roads based on the relationship between visible road areas and road center lines. Although these methods can solve part of the road occlusion problem, it is still difficult to obtain accurate and complete road information using only a single data source in the face of complex road background.

#### 3.1.2. Summary

As described above, the existing road pavement extraction approaches are not smart enough to fulfill practical applications. Each method has both advantages and disadvantages. It is difficult to use only one approach to get the high recognition accuracy. Therefore, the methods should be studied with a variety of ways and combine different approaches. In addition, the key problem is how to describe the road features. Most of the existing methods describe a road as a linear or narrow bright band and can get a good detection result. However, with the increasing image resolution, more detailed road features along with the more noise interference (buildings, shadows, road obstructions) would appear, so the road objects should be described precisely. Hence, how to establish a good road model and extract the road quickly and accurately is a common concern of many researchers.

### 3.2. Road marking extraction

Road markings are symbols and texts displayed on the road surface (Bailo et al., 2017). It mainly consists of solid lines, crosswalks, dotted lines, arrows, and characters are common forms of road marking (Soheilian et al., 2010; Qin et al., 2013). As a key feature of HD maps, road markings provide important information for autonomous vehicles (Guan et al., 2014; Woo et al., 2008).

#### 3.2.1. Methods and analysis

As road markings are highly reflective objects, high reflectance of intensity has been widely used to extract road markings. The intensity of the reflected laser pulse also depends on the incident angle of the emitted laser beam, material properties of the target, and the distance between the measured target and the laser scanner. Generally, the laser pulse intensity value decreases with increasing distance and angle of incidence (Yu et al., 2015b). Based on the data source used, the extraction methods can be divided into three categories: 2D vision-, geo-referenced feature (GRF) image- and point-based methods.

**2D vision-based methods** mainly focus on the color and pattern changes between the pavement background and marking materials. These methods usually rely on the fact that road markings are light (white or yellow) on a dark road background. Many previous studies have been developed based on various image processing techniques, such as edge detection, color segmentation, template matching, Hough transform, and RANSAC (Ying and Li, 2016; Ye et al., 2017; Mammeri et al., 2014). However, they are seriously affected by distortion, poor ambient lighting conditions, occlusion caused by pedestrians and vehicles, and shadows cast by buildings and trees (Yu et al., 2015b). Currently, several DL-based methods are developed to obtain better performance. Considering that road markings are generally textual or closed geometries, there are also many works that look at road marking extraction as a target detection task. Among them, deep learning-based

methods are the most effective methods available (Ye et al., 2017). Chen et al. (2015) proposed a road marking detection algorithm based on PCANet (Chan et al., 2015). Bailo et al. (2017) used a well-designed shallow CNN to classify road markings. This method needs to extract a fixed number of candidate regions for each frame, which usually leads to computational redundancy. Another disadvantage is that road markings are not accurately located, and the derived bounding box usually includes other unrelated objects.

**GRF image-based methods** use multi-threshold segmentation to solve the uneven distribution and fluctuation of intensities. Guan et al. (2014) and Jaakkola et al. (2008) used radiometric correction for the normalization of MLS intensity data. Vosselman (2009) proposed a distance dependent intensity normalization method by which several types of road markings can be identified. In addition, lateral range estimation was used to divide the GRF images into blocks and then determine the intensity threshold for each block based on the distance of the target from the laser scanner (Kumar et al., 2014; Guan et al., 2015). The accuracy of the GRF image-based method is relatively high. However, the extracted markings are incomplete and contain distinguishable noise. This is because the information loss when the road markings are projected onto 2D plane. False positives and false negative errors of candidate road marker points in GRF image-based methods can be effectively eliminated by median filtering or multiscale tensor voting (Kumar et al., 2014; Cheng et al., 2017b). To fix incomplete road markings, Kumar et al. (2014) converted the marked GRF images to binary images. A linear shape structure operation was then used to expand the road marking region and optimize the extraction results.

**Point-based methods** directly analyze 3D pavement point clouds and detect candidate road marking points, which provides a promising high-precision method for road marking extraction. Yang et al. (2013) proposed a scan line segmentation method for pavement points based on scan angle and detected the edge points of road markings by dynamic median filter. Since some elements on the road surface may have high reflectivity and thus lead to false extraction, it is necessary to refine the extracted candidate road marking points. For example, some false alarms and false negative errors can be eliminated by dividing scan lines or spatial density filtering. To solve the problem of uneven intensity distribution, Yu et al. (2015a) first used a multi-threshold strategy to segment the pavement point cloud and then extracted the road marking points. In addition, Wen et al. (2019) proposed a deep learning framework for road marking extraction, classification and completion from 3D MLS point clouds. Compared with other extraction methods, this method is more robust to data quality. Similarly, Mi et al. (2021b) proposed a two-stage approach to extract road marking by using MLS point clouds. Different from the above two methods, Zhang et al. (2019a) used the 3D data to extract the road marking. They designed a step-shaped-based method to convolve the profile data to extract the marking edges. Then the geometric information of selected region and the continuity of the convolution features were used to extract the road marking.

#### 3.2.2. Summary

As described above, although the existing road marking extraction methods can achieve more than 90 % accuracy, the predefined shapes used to fit the road marking segments can lead to incomplete road marking extraction. Also, most algorithms use segmentation based on global thresholding, which can introduce a lot of noise and thus lead to less efficient road marking extraction. In addition, road markings are drawn using materials that have a high reflectivity to laser pulses, and the intensity of the road marking points is higher than the surrounding road points. However, due to the attenuation of road markings and obstacles on the road, the intensity of some road marking points may be lower, and some road markings may be damaged. In this case, errors may occur in the road marking extraction. These issues are a challenge to the existing methods.

### 3.3. Road lane extraction

Accurate lane detection is the key function of lane keeping, lane change automation and lane departure warning. Lane detection can be realized by using monocular camera, stereo camera, LiDAR, etc. (Bar Hillel et al., 2014).

#### 3.3.1. Methods and analysis

In the past two decades, a lot of research has been conducted by scholars in the field of lane detection (Jefri Muril et al., 2020; Bar Hillel et al., 2014; Li et al., 2014; Yenikaya et al., 2013). These methods can be broadly classified into two categories: traditional and DL-based methods. In this section, we provide a brief overview of each category.

**Traditional methods** can be divided into three classes: the geometric model, Hough transform (HT) model and energy minimization model. Geometric modelling basically follows a two-step solution, with edge detection followed by line fitting (Zhou et al., 2010; McCall and Trivedi, 2006; Aly, 2008; Borkar et al., 2011). Various types of gradient filters are used in edge detection, such as Gaussian filter, Steerable filter, and Gabor filter, etc. To test the extraction of rails from videos, Selver et al. (2016) preprocessed the images with Gabor filtering using kernels of different orientations and then post-processed them using morphological operations to link gaps and remove unwanted blocks. Choi and Oh (2010) first used template matching to find lane candidates and then extracted lanes by color clustering. The Hough transform (HT) model is often used in line fitting. Curve fitting method, such as B-snake and Catmull-Rom Spline, can be used either as a refinement of the HT results or as a direct replacement of the HT (Deng and Han, 2013). As for the energy minimization model, Conditional Random Field (CRF) is used to detect multiple lanes by establishing the best association of multiple lane markings in a complex scene (Hur et al., 2013). Energy minimization can also be embedded in the search for the best modeling results in lane fitting. Kang et al. (1996) developed an active line model with an energy function that contains two components. In addition, Kalman filter and Particle filter are also widely used in road lane tracking, which can track multiple lanes (Mammeri et al., 2014).

**DL-based methods** can be divided into three groups. Encoder-decoder CNN model treats lane extraction as a segmentation problem. Kim and Park (2017) proposed an end-to-end encoder-decoder CNN network for lane detection. Neven et al. (2018) proposed a two-decoder structure lane extraction network. One decoder is used to detect lanes and the other one is used to segment the road. Finally, based on the feature map, a clustering and curve fitting algorithm is used to generate the final results. In response to the lack of training data, Bruls et al. (2018) proposed a new weakly supervised CNN network that uses redundant sensors to generate many annotated images for the training. As for the FCN model, Kim et al. (2017) proposed to use a CNN with fully connected layer as the original decoder and a hat-shaped kernel to infer lane edges, and applied RANSAC to optimize the results. He et al. (2016) proposed a dual-view CNN framework and applied a weighted hat filter to the top view to find lane candidates. Gurghian et al. (2016) implemented lane detection using a CNN with a fully connected layer and softmax classification. To achieve the goals of lane detection, lane marker recognition and vanishing point extraction, Lee et al. (2017b) proposed to use the VPGNet model together with optimization algorithms such as clustering and subsampling. To improve the performance of CNN in detecting long continuous shape structures, a spatial CNN was proposed, it extended the traditional layer-by-layer convolution to slice by slice convolution based on feature maps, which improved the performance of lane extraction (Pan et al., 2018). Besides, GANs are also used for lane extraction. For example, Goodfellow et al. (2014) proposed an embedding loss GAN for semantic segmentation of driving scenes, where lanes were predicted by a generator and judged by a discriminator with shared weights.

#### 3.3.2. Summary

Based on the description of the above methods, it can be found that although the existing methods can predict the road lanes more finely and accurately, it is still difficult to cope with the presence of occlusions or shadows. Meanwhile, the existing methods do not make full use of the correlation information between different frames of video data, which makes the real-time performance of lane detection low, which poses a challenge to the construction of high-definition maps for autonomous driving. RNN and Transformer methods with sequential data analysis capability can be considered in the next research to improve the lane detection performance of the existing methods by introducing them into CNN.

## 4. Sign detection

For HD map generation, the signs usually contain the traffic sign and road sign. These two kinds of signs are often detected together. Therefore, it is reasonable and convenient to class the traffic and road signs as the same category. For convenience, we use traffic sign in the following, which denotes both the traffic and road signs.

### 4.1. Traffic signs detection methods

Traditionally, traffic signs detection (TSD) methods contain two stages. Manual-designed descriptors are firstly provided to mine features (e.g. color and shape) from raw images. These features are then fed into off-the-shelf classifiers, such as threshold classification, Random forest and SVMs. With the development of machine learning, deep learning (DL) has shown its advantage in feature extracting. Recently, several DL-based methods have been undertaken for TSD and obtain excellent performance. Therefore, according to the way of extracting features, there are two categories: the handcrafted feature- and DL -based methods. In addition, for TSD from 3D point clouds, some works have been undertaken. Fig. 5 shows the classification of TSD methods on image and 3D point clouds. Fig. 6 illustrates several milestone methods.

#### 4.1.1. Handcrafted feature-based methods

Generally, traffic signs are represented with several discriminative features (especially the color and shape). For example, the warning sign and prohibition sign are with the red color, while the warning sign and temporary sign are with triangle shape. Therefore, we can further categorize these methods into four classes: color-, shape-, color and-shape-, and multi-attribute- based methods. The statistical distribution for these methods is shown in Fig. 7. From Fig. 7 (a) and (b) we can find that the color-based and shape based methods got more attention before 2010, and multi-attribute-based methods have become the main research idea after 2015. This is mainly because color-based and shape-based methods are relatively straightforward, so they got more attention in the early stage. With a further understanding of the traffic sign, color-shape-based methods are favored by more researchers because of their better performance. Besides, from Fig. 7 (c) we can find that the multi-attribute-based methods have the highest proportion. This is mainly because the multi-attribute-based methods make full use of the advantages of different attributes, that is, different attributes complement each other to make the performance better than single-attribute-based methods, such as the color- and shape- based methods. In addition, Fig. 8 shows the common framework for TSD from images. Most existing image-based TSD methods follow the framework.

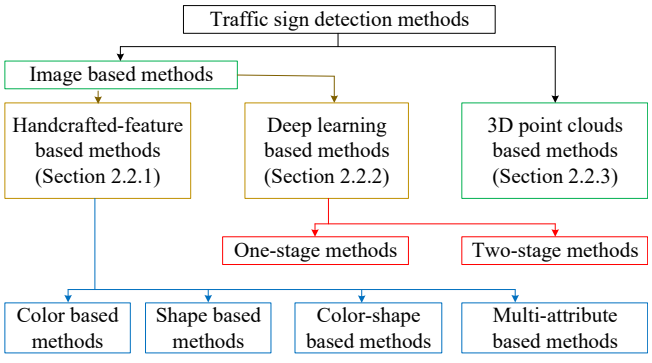



**Fig. 5.** Classification of existing TSD methods on images and 3D point clouds.

**Color-based methods**. Considering that the traffic sign has specific characteristic colors, such as the red, blue, and yellow, several methods take them as the basic features, and design different color spaces to improve the performance of TSD. Generally, there are three main space-based methods: RGB, HSV, and YUV spaces.

*RGB space*: It is an intuitive way to utilize the channels in RGB space. As the classic TSD method, De la Escalera et al. (1997) adopt a threshold-based method in the RGB space. However, it may not work for the occlusion case. Ruta et al. (2010) took the filtering operation as the basic framework, and used the RGB channels to enhance the color with maximum and minimum operations. Three maps are generated by authors from the input images. However, a mandatory sign usually has a dark or bright illumination. Therefore, it has a similar value in blue or green channel, and would not work for the extraction in blue color. To resolve this issue, Salti et al. (2015) changed the enhanced blue channels by interest regions extraction. Similarly, Lim et al. (2009) normalized three RGB channels firstly, followed by constructing four new image channels according to the RGB transformation formula provided in Itti et al. (1998).

*HSV space*: Due to the lighting variation, it may be not discriminative enough to extract the sign from the image by just using the threshold in RGB spaces. HSV (hue, saturation and value) color space can also be used to represent the image. In Vitabile et al. (2001), authors utilized a dynamic threshold aggregation-based method to detect the road sign from real world scenes. However, it takes high time consumption. Similarly, by utilizing the achromatic decomposition, Maldonado-Bascon et al. (2007) combined thresholds on both H and S components. However, it is easily disturbed by the illumination change and signal deterioration. Wang et al. (2014a) also extracted features from HSV space. Authors used the values in HSV space of neighbor pixels to design a threshold-based method. In Yakimov (2016), authors obtained the optimal threshold value for red color with experimental results. After segmentation, they used a filter algorithm to remove the noise and finally determine the regions of interest.

*YUV space*: YUV space consists of three components, which are computed by RGB information. In Miura et al. (2000), a threshold-based method in YUV space is proposed to detect the rectangular information panel, then a gradient based projection method is provided for the final result. Similarly, in Broggi et al. (2007), authors also operated on YUV space to detect road signs.

**Shape-based methods**. Since the traffic sign usually enjoys the conventional shape, such as triangle, circle, rectangle, and octagon, several methods take this property as the basic features. There are mainly-four classes for these methods: standard shape-, shape matching-, Fourier transformation-, and key point-based methods.

*Standard shape-based methods*. Standard shapes are the key characteristics for the traffic signs. Several existing methods usually used the common transforming algorithm as the foundational detection algorithm, such as the Hough and radial symmetry transformation. Barnes and Loy (2006) utilized the fast-radial symmetry to detect traffic signs with polygonal shapes. It obtained promising results. Besides, based on the fast radial symmetry, Barnes et al. (2008) proposed a TSD model, which can robustly detect the un-occluded shapes with high efficiency. In addition, Hough transform is also a common technology. The classic Hough transform-based sign detection method is Yakimov and Fursov (2015). Authors proposed a novel TSD method by modifying the general Hough transform with high efficiency. However, the preprocessing stage, including the red color detection and noise deleting, may take

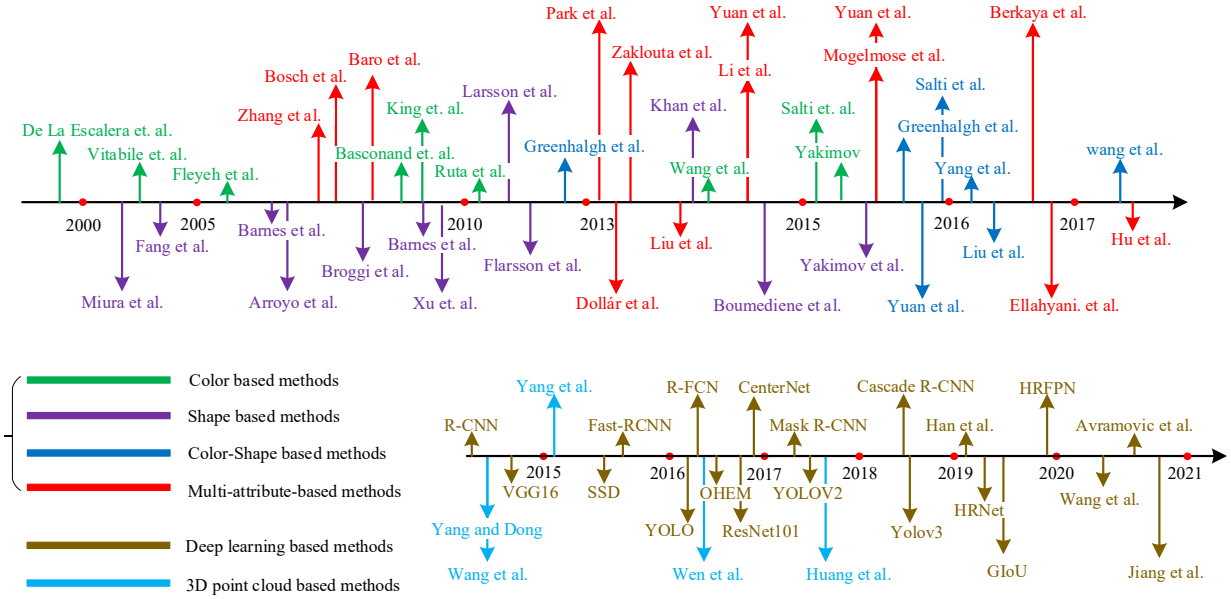
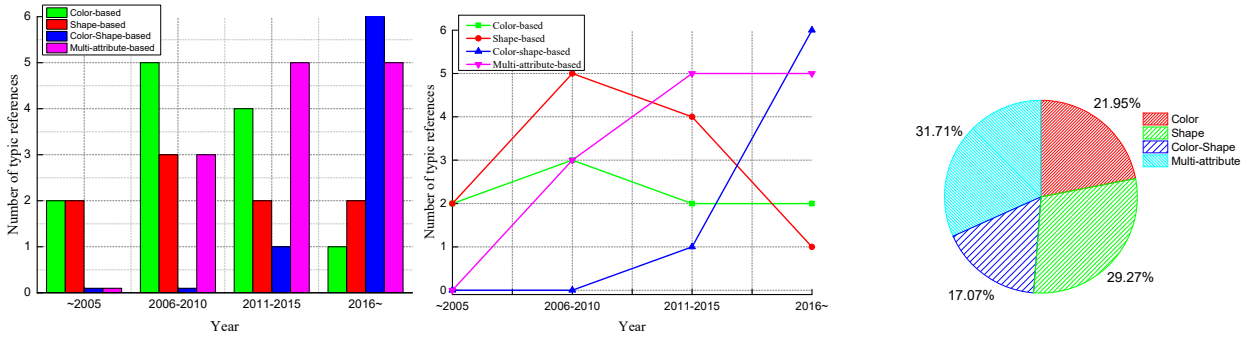



**Fig. 6.** Chronological overview of the most relevant TSD methods. The color/shape/color-shape / multi-attribute-based methods belong to the handcrafted-feature based methods.




**(a)** Distribution histogram **(b)** Trend curves **(c)** Distribution map

**Fig. 7.** Statistical distribution for handcrafted-feature based methods.

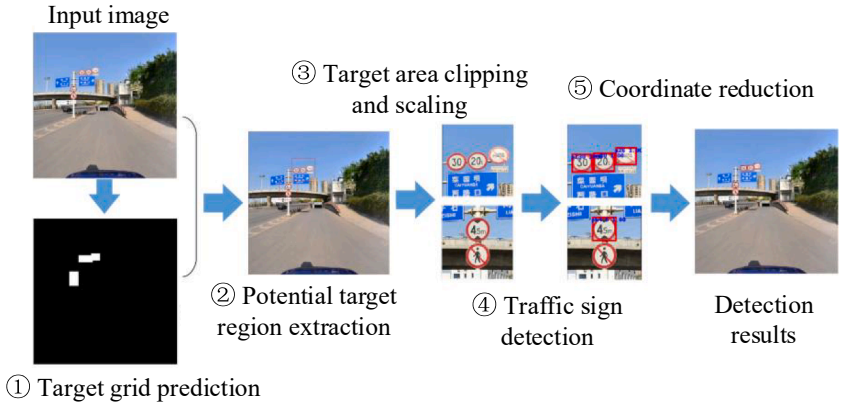



**Fig. 8.** The common framework for TSD from images (Liu et al., 2016a).

high time consumption.

*Shape matching-based methods*. These methods detect the signs with significant edges by matching and analyzing different shapes. By designing different complex shape models, Fang et al. (2003) proposed a matching method to detect the circular, triangular and octagonal signs. However, since these models depend heavily on the distinct edges, they usually suffer from the noise and shape changes. To improve the robustness, Xu (2009) provided a decomposition method. Specifically, a more complex shape can be represented by several simpler components. Then a maximal supported convex arcs-based method is used to select

the potential components.

*Fourier transformation-based methods*. These methods use the Fourier transformation to describe the traffic sign. Gil Jimnez et al. (2008) represented the shapes of different traffic signs via Fast Fourier Transformation, then developed a triangle normalization and reorientation to obtain the final results. Larsson and Felsberg (2011) designed a Fourier descriptor to represent the sign, following by a local contour based segmentation method to extract different traffic signs. In addition, Larsson et al. (2011) provided Fourier transformation-based method to detect traffic signs. However, since the Fourier transformation tends to capture the global information, these methods usually ignore the details of the traffic and road signs.

*Key point-based methods*. Key point-based idea can be applied in detecting the singularities and angular edges of traffic signs. Since the scale-invariant feature transform (SIFT) is scale and rotation invariant, it has been the one of most popular key point-based descriptors. By focusing on the corner detection, Boumediene et al. (2014) designed a method to detect the corners of traffic signs. More specifically, a Harris corner based detector is proposed to obtain the candidate regions, followed by a neighborhood-based selection strategy to generate the final results. Khan et al. (2014) proposed a Gabor filter based method to detect the local feature, following by a clustering-based algorithm to obtain the target points.

**Color-and-Shape-based methods**. Generally, these methods take the color and shape as the basic features. Broadly, they can be divided into three classes: extreme regions-, high contrasted margin regions-, and saliency-based methods.

*Extreme regions-based methods*. As a popular region-based method, the maximally stable extremal regions (MSERs) algorithm is usually used as the basic technology in detecting the high-contrast region, which would include the approximately uniform gray tone or different shapes. Greenhalgh and Mirmehdi (2012) and Greenhalgh and Mirmehdi (2015) used the MSERs based methods to detect traffic signs. In Greenhalgh and Mirmehdi (2015), authors took the hue, saturation, and value color to obtain the final results, while in Greenhalgh and Mirmehdi (2012), authors shared the same idea, but detected the traffic sign from color enhancement images, not the grayscale one. In addition, Salti et al. (2015) first enhanced the images in RGB channels, and used the MSERs and a wave-based method for the extraction of the traffic sign with a promising detection rate. Although these methods have achieved high accuracy, they are far from satisfactory in robustness and efficiency.

*High contrasted margin regions (HCRE) based methods*. Generally, HCRE was used for the detection of the high contrasted margin region. Based on HCRE, several methods have been undertaken for TSD. Liu et al. (2016a) designed the HCRE method for the extraction of ROIs, then a split-flow cascade tree detector is provided to extract traffic signs. This method can achieve a good balance between the extraction rate and detection rate.

*Saliency-based methods*. The saliency is usually considered as an important property in traffic and road signs. Focused on detecting traffic signs from complex environments, Yuan et al. (2015) proposed a graph-based detection method. It contains three stages: saliency measure, graph-based ranking, and multi-threshold segmentation. In the saliency measure stage, a formula is designed to compute the saliency, and the nodes with high saliency are considered as the traffic signs regions. Wang et al. (2017) provided a TSD method in a fast run speed. They designed a saliency test and neighboring scale awareness strategy. To improve the run speed, a saliency based algorithm is designed on mid-level features. This method was demonstrated to be more robust and effective than other methods.

**Multi-attribute-based methods**. Generally, this kind of methods take multiply attributes as the features. According to the difference of classifiers, they are further divided into two main classes: the AdaBoost and SVM based methods.

*AdaBoost-based methods* use the AdaBoost to select features to express different objects, following by the cascade-based method for object detection. In Baro et al. (2009), a boosted detectors cascade is designed by combining with a novel evolutionary version of AdaBoost in a large feature spaces. As a popular rectangular feature used in face detection and recognition, Multi-Block Local Binary Pattern (MB-LBP) feature (Zhang et al., 2007) is also used as the basic feature in traffic signs. Based on MB-LBP, Liu et al. (2014) provided a high efficient TSD method. They designed two features, the multi-block normalization LBP and the titled MN-LBP features to represent different kinds of signs. Using the integral channel features (ICF), Mgelmose et al. (2015) provided different methods for traffic signs detection and have achieved good performance. In addition, unlike traditional AdaBoost based structures with only gray information, authors in (Dollr et al., 2014) designed a fast feature pyramids detection method by using a cascade of three frameworks, the Aggregate Channel Features (ACF), integral channel features (ICF), and HOG features. Hu et al. (2016) provided a single-stage learning-based detection framework, which contains a dense feature extractor and a detector. For the improvement of robustness and image deformations, authors developed spatially pooled features to support the aggregated channel features.

*SVM-based methods* usually combine with the HOG-likes feature descriptors. Features generated from HOG-like descriptors are fed into SVM-based classifiers. The classic methods include Salti et al. (2015) and Zaklouta and Stanciulescu (2014). In these methods, the HOG-SVM frameworks are used as the basic technologies. Several similar works include Bosch et al. (2007), Yuan et al. (2014) and Kaplan Berkaya et al. (2016). Some descriptive handcraft features are also designed to capture 3D shapes. As one of the most popular features, pyramid histogram of oriented gradients (PHOG) feature provided in Bosch et al. (2007) has been used in several works, such as Park and Kim (2013), Yuan et al. (2014), Li et al. (2015), Ellahyani et al. (2016) and Kaplan Berkaya et al. (2016). Different with AdaBoost-based methods, SVM-based methods usually need less training samples. Therefore, in practical application, SVM-based method can be an appropriate way when there are few training samples.

#### 4.1.2. DL-based methods for TSD

Generally, according to the number of stages in detecting objects, DL-based TSD methods from images can be further broadly classified as two classes: two- and single- stage methods.

**Two-stage methods**. The two-stage algorithms include R-CNN (Girshick et al., 2014), Fast-RCNN (Girshick, 2015), Faster R-CNN (Ren et al., 2017), R-FCN (Jifeng et al., 2016), Mask R-CNN (He et al., 2017), etc. Focusing on small traffic sign detection, Han et al. (2019) proposed a small region proposal generator by removing the pool4 layer of VGG16 (Simonyan and Zisserman, 2015) and adopting dilation for ResNet. Cao et al. (2021) innovatively adopted HRNet (Sun et al., 2019) to improve the feature extractor of Faster R-CNN (Girshick, 2015), and adopted PISA (Cao et al., 2020) strategy to optimize the learning strategy of Faster R-CNN. Focusing on the detection of different kinds of signs, Tabernik and Skoaj (2020) designed an automatic method to detect and recognize the signs by utilizing the mask R-CNN. Liu et al. (2020) designed a coarse-to-fine framework for the detection of traffic signs efficiently and accurately. More specifically, they sequentially detect the targets in grid-level and image level. Generally, similar with other detection tasks from images, the above two-stage TSD method can often achieve high accuracy with low efficiency.

**Single-stage methods**. Single-stage object detection algorithms are represented by YOLO (Redmon et al., 2016), SSD (Liu et al., 2016b), etc. Such algorithms directly predict objects' location and category without region proposal. To improve the descriptiveness, Jinhong et al. (2020) utilized the Focal Loss Focal Loss (Lin et al., 2017) and GIoU (Rezatofighi et al., 2019) to replace the way of computing loss in Yolov3. However, the YOLO based methods mainly focus on the run speed, and the detection accuracy usually worse than the two-stage based methods. In Xiaozhu and Cheng (2017) considered the traffic signs detection as the region classification problem and designed a center-point estimation

based method, which can achieve a lightweight network. Taking the advantages of high run speed, straightforward architecture and high accuracy in general object detection, in Avramovi et al. (2020), authors used multiply YOLO architectures as baseline detectors, and obtain the final results in high efficiency.

#### 4.1.3. 3D point cloud-based methods for TSD

As the key data structure of HD map, point clouds can provide a more precise location description than images. Different from the TSD for images, traffic and road signs in 3D point clouds are more susceptible to noise, such as the occlusion, self-occlusion, outliers, and the distribution of points. Therefore, it requires a different idea from images processing to solve this problem. Currently, existing TSD methods from 3D point clouds take the segmentation and handcrafted-features as the basic technology and features, respectively.

Fig. 9 provides a common framework for TSD from 3D point clouds. Most of existing methods follow this framework. Yang and Dong (2013) designed a shape-based method to segment the point clouds. More specifically, after obtaining the optimal neighborhood size for each point, a rule-guide algorithm is designed to segment the point clouds. By merging the segmentation with topological connectivity, the meaningful geometrical abstraction can be generated. Focused on the efficiency and robustness, authors in Yang et al. (2015) proposed a multi-scale super-voxel based method utilizing the point attributes, such as the color and intensity, as the improvement of Yang and Dong (2013). In the detection stage, they designed a hierarchical order strategy with ranking the saliency of the segments, to extract the final results. Wang et al. (2014b) proposed a Hough Forest framework to detect objects from 3D point clouds. More specifically, they extract the targets at the peak regions in the 3D Hough voting space. In Yu et al. (2016), authors provided a classic TSD framework by designing a novel bag-of-visual-phrases-based algorithm, followed by a aussian-Bernoulli deep Bolitzmann machine-based hierarchical classifier. They first designed a bag-of-visual-phrases based method to represent the 3D point clouds, then used the Gaussian-Bernoulli deep Bolitzmann machine method to provide a hierarchical classifier for the recognition of the traffic sign. Fig. 9 shows the pipeline used in this method, which is considered as the basic schematic in follow-up works. Wen et al. (2016) described a spatial-associated algorithm for TSD using mobile laser scanning data. In detection step, they designed the geospatial relations between the traffic signs and road environments to obtain the final result. Similarly, Huang et al. (2017) proposed a detection method considering both reflectance and geometric features with the region growth strategy to capture traffic signs.

### 4.2. Summary

The above mentioned methods have made contributions in traffic sign detection field. Several observations can be drawn as follows:

- Most existing image-based methods, especially the DL-based methods, can obtain excellent performance on both accuracy and efficiency in some public datasets. However, more attention should be paid in the robustness. This is because in road environments, as one of the small targets, the visual range of traffic sign is usually narrow. For this reason, methods with strong robustness must be considered.
- It can be observed that most of the 3D point cloud-based methods are still in the stage, i.e., using the handcraft based features with traditional classifiers to detect traffic signs. Applying more descriptive feature mining technologies, such as the 3D deep learning, to obtain more effective features is a considering and encouraging attempt.
- It is rare to combine both image and 3D point clouds to detect traffic signs. It must be noted that images can provide rich type information for traffic signs, while the 3D point clouds capture the detailed shape, size and location of traffic signs. It would be a worth way to generate a more powerful description for the traffic signs by using both image and 3D point clouds.

## 5. Guardrail detection

The guardrail is the key part of HD map elements. It provides important information to guide vehicles driving in appropriate areas. Generally, as shown in Fig. 10, guardrails can be divided into two categories: anti-fall and lane-separating guardrails.

### 5.1. Methods and analysis

Several works have been undertaken to resolve the guardrail detection. According the representation of input data, existing guardrail detection algorithms are classified into three categories: image-based, point clouds based (Jiang et al., 2016; Masuda et al., 2013) and multi-modal based methods (Zhu and Guo, 2018; Matsumoto et al., 2019). Fig. 11 provides several common methods for guardrail detection from images, 3D point clouds and multi-modal representations.

#### 5.1.1. Image-based methods

These methods extract guardrails from images, which is similar to the common detection task from images. Authors in Seibert et al. (2013) extracted road objects by extending an existing commercial detection system. More specifically, they designed two complementing frameworks. In the first step, the Local Binary Patterns are utilized to evaluate the quality and ability of the borders, and then obtain the location of these borders. In the second step, authors provided an algorithm for the detection of guardrails and walls by using the Harris features and Lukas and Kanade image tracking. Although this method can extract the objects quickly, it is highly dependent on the guardrail's position. This would lead to partly extraction. In Scharwchter et al. (2014), authors designed a novel vision-based system for guardrail detection. They used a stereo camera setup to obtain full view images, then developed an excellent feature encoding method with full use of the geometry information. For the potential location of guardrails, authors proposed Hough-based method to guide each detected patch to fulfill linearity in depth and certain height expectations. Finally, to leverage the appearance information, an efficient bag-of-features representation are exploited, and the randomized clustering forests are also used. However, since the camera system is sensitive to the noise in the dark, this method only works in the daytime, which may bring limitations of the

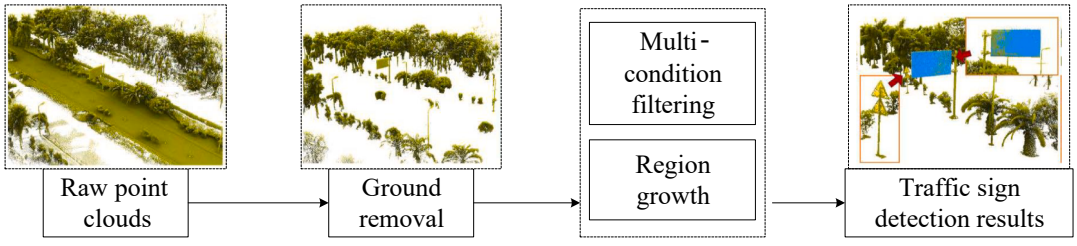



**Fig. 9.** Traffic sign detection with multi-condition filtering (Huang et al., 2017).

**Fig. 10.** Guardrails on roads. left: lane-separating; right: anti-fall guardrails (Gao et al., 2020).

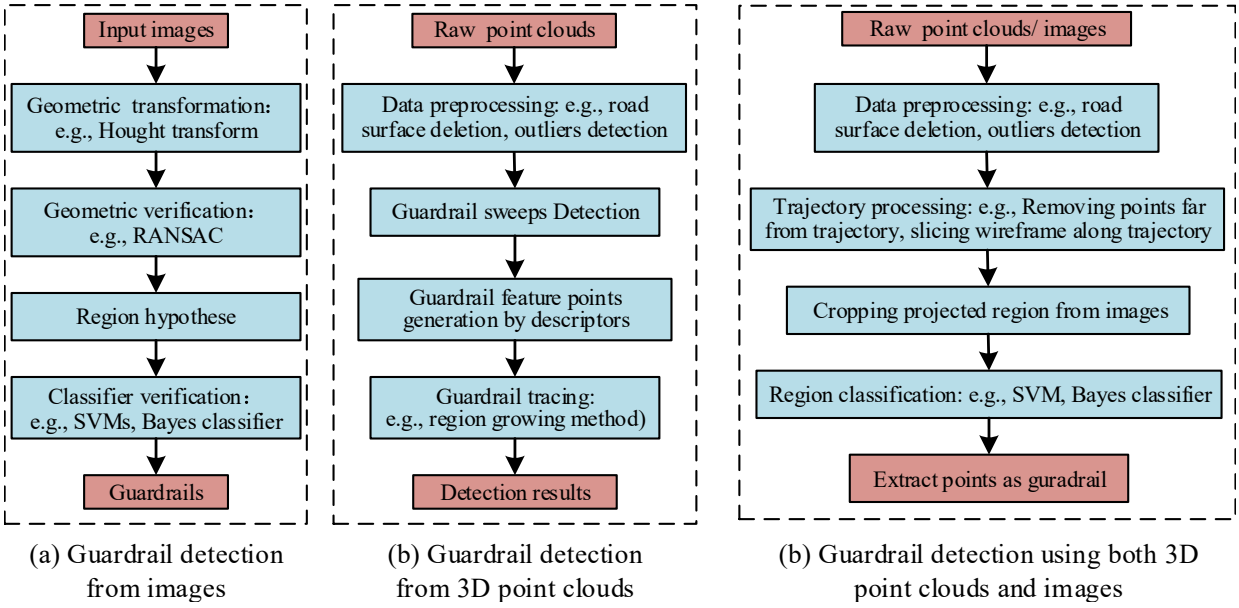



**Fig. 11.** Common frameworks for guardrails detection from (a) images, (b) 3D point clouds, and (c) multi-modals, respectively.

universality.

#### 5.1.2. Point clouds based methods

They extract guardrails from the 3D point cloud directly. Jiang et al. (2016) provided a novel automatic guardrail detection method with high efficiency and strong robustness for HD map generation using the MLS. More specifically, the potential guardrail sweeps are first extracted via a segment based strategy, then they predict the potential guardrail points by designing corresponding features. Finally, a guardrail tracking algorithm is proposed for the refinement of the potential target point, as well as generate the detection result. Although this method can achieve excellent performance, it just focuses on one type of guardrails, which may limit its universality. Zhu and Guo (2018) proposed a beam guardrail detection algorithm from 3D point clouds. Corner features and height features are designed to predict the potential guardrail points, while a clustering-based method is utilized to remove outliers. Different from Jiang et al. (2016) and Zhu and Guo (2018), Gao et al. (2020) provided a rapid extraction method for both anti-fall and lane-separating guardrails. Authors first designed a multiply levels-based method for the detection of the road surface and the deletion of the outliers. Then they developed a spatial cluster method based on the density and proposed a modified clustering method to extract guardrails with four-step screening. This method can adapt to detect different kinds guardrails and rough slope road, and has achieved high precision on two test datasets. However, these methods are all handcraft-feature-based, which means they may be limited in the universality.

#### 5.1.3. Multi-modal based methods

Multi-modal based methods mainly focus on image and point clouds representations. Authors in Matsumoto et al. (2019) made full use of both the image and 3D point cloud. The 3D point cloud is first segmented into small parts, then the image is cropped to obtain the corresponding image patches. Then the images patches are labeled into potential guardrail pixels and no-guardrail pixels using the convolutional neural networks. Finally, by superimposing points of unit shapes, authors created dense point clouds for guardrails.

### 5.2. Summary

The above mentioned methods have made contributions in guardrail detection. Some observations can be drawn.

- Most existing image-based methods can obtain satisficed results in some datasets. However, they are usually sensitive to disturbance. For example, some methods are highly dependent on the guardrail's position. They would obtain partly extraction without accuracy position. In addition, since the camera system is sensitive to the noise in the dark, these methods tend to only work in the daytime, which may bring limitations of the universality.
- It can be observed that most of point cloud-based methods take handcraft feature-based idea. The advantage lies in that they have good model Interpretability. However, utilizing the handcraft-feature based idea means they may be limited in the universality. It is encouraged to apply more descriptive feature mining technologies, such as the 3D deep learning frameworks.

## 6. Pole-like object detection

### 6.1. Methods

Pole-like objects, including the pole parts of the traffic/road sign, the street lamp, and trees, are important in urban road environments. They

play a crucial role since they can provide different information for the pedestrians and vehicles. Therefore, it is necessary to extract pole-like objects for HD map generation.

Detecting pole-like objects from images is of great importance for HD map generation. Generally, there are two main categories for image-based detection methods: one-stage and two-stage methods. Since detecting pole-like objects from images is similar with other object detection tasks from images, there are relatively few image-based methods designed for street tree detection specifically. The newest related work is the method in Lee et al. (2022). It uses the YOLOv3 algorithm and the transfer learning to construct a deep learning based method for the road facilities detection. By using the crowdsourcing framework, this method can update HD maps. However, as pointing out by authors, some other advanced algorithms can be applied to improve the road facilities detection.

Conversely, several works have been undertaken to extract the street trees in 3D point clouds. They can be broadly divided into four categories: shape-, verticality-, density-, and continuity-based methods. Fig. 12 shows several classic methods in chronological order. Fig. 13 provides the common flowchart for POD.

#### 6.1.1. Shape-based detection

Considering that the pole-like parts of different road objects are usually represented by the linear-like shape, it is feasible and straightforward to extract these parts as the pole-like objects. Lam et al. (2010) utilized the context-based strategy and existing prior-knowledge about the city infrastructure to build a road object detection method with high efficiency. Especially, to detect the pole-like objects, authors used the RANSAC and least squares fitting algorithms. Huang and You (2015) provided a detecting and classifying method for POD in the 3D point cloud urban road environments with three main steps: localization, segmentation and classification. Taking the advantage of the unique pole-like objects' properties in road scene, a slicing clustering method is designed to obtain the seed potential pole-like points. To extract target region, a region growing algorithm integrated with a bucket-shaped neighborhood is proposed. Finally, features with six attributes are designed to classify points to obtain the final pole-like objects. From the furniture detection perspective, authors in Li et al. (2018) provided a method with three steps to detect road furniture. The third step is de the pole-like objects are considered as the results of the third step. Some other methods also share the same idea, such as Tombari et al. (2014) and A. Husain et al. (2018). In addition, Yang et al. (2019a) designed a skeleton-based hierarchical method to automatically detect the pole-like objects from MLS point clouds. A coarse-to-fine strategy is proposed to achieve this goal. More specifically, the building facades is first extracted to analyze the occlusion. Then, the slice-based Euclidean clustering method is adopted to obtain the potential pole-like points. Finally, a region growing method with Voronoi-constrained strategy is used to generate the pole-like objects.

#### 6.1.2. Verticality-based detection

Generally, the verticality can be used as a unique attribute that distinguishes the pole-like object from other objects, which has been considered in several works. Fukano and Masuda (2015) provided a method for both the detection and classification. In the detection stage, they transformed 3D point clouds into wireframe, then calculated cross-sections between wireframe models and horizontal cutting planes. Yan et al. (2017) took the idea of space division along the axis. Authors designed a three-step method, which contains data preprocessing, pole-like object detection and pole-like object classification. In the detection stage, they utilized the filers combined with both prior and shape information to extract the target objects. Similar to Yan et al. (2017), a hierarchical extraction method is provided in Hao et al. (2018) for the extraction of pole-like objects. More specifically, the common Euclidean distance based clustering method is utilized to obtain the initial clusters, which are then segmented to different parts along the z-axis. Finally, authors designed a horizontal distance-based algorithm to extract the final pole-like objects. In addition, several related methods are proposed with other goals, such as the classification of pole-like object (Mori et al., 2018).

#### 6.1.3. Density-based detection

Considering that the density of points of pole-like objects may be much higher than other kinds of objects, it is available to use this property to extract pole-like objects. Naturally, these methods should project 3D point clouds into 2D space first. Yan et al. (2016) proposed a four-step workflow to detect light poles with DBSCAN as the clustering algorithm. In the key stage, the third step, similar to Li et al. (2016a), they also used the DBSCAN as the basic unsupervised clustering algorithm. This may overcome the limitation of this kinds of algorithms, such as the k-means and ISODATA that only considered the Euclidean distance. Similarly, Zheng et al. (2016) proposed a DBSCAN based POD method. More specifically, after projecting 3D point clouds into 2D grid, the seed points of potential pole-like objects are generated and finally, by using the DBSCAN, they extract the points of pole-like objects. Hu et al. (2011) proposed an approach to extract the street lamps from vehicle-borne laser data. In the detection of the pole-like part, they developed the conventional method, the density of projected points (DoPP) to achieve low time consumption. Focusing on the pole-like street furniture decomposition in mobile laser scanning data, Li et al. (2016a) also proposed a 2D-point-density based method combined with the RANSAC line fitting algorithm to extract poles. However, the key preprocessing step, i.e., the projection-based method, would lose some important information, which may reduce the accuracy.

#### 6.1.4. Continuity-based detection

For 3D point clouds, the continuity characterizes the important topological properties of the distribution of pole-like objects in 3D space. Based on the vertical continuity, Wu et al. (2013) proposed a voxel-based algorithm to extract the street trees from MLS point clouds by designing a neighborhood searching method. By designing six technical

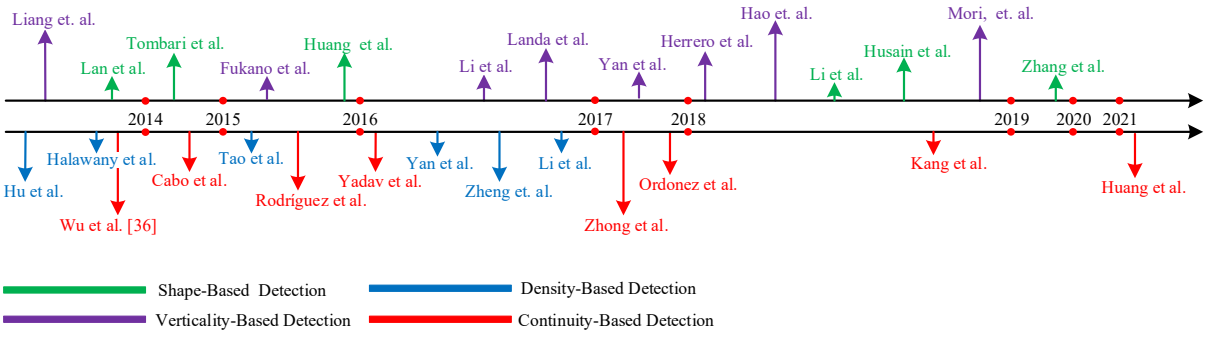



**Fig. 12.** The chronology for common POD methods.

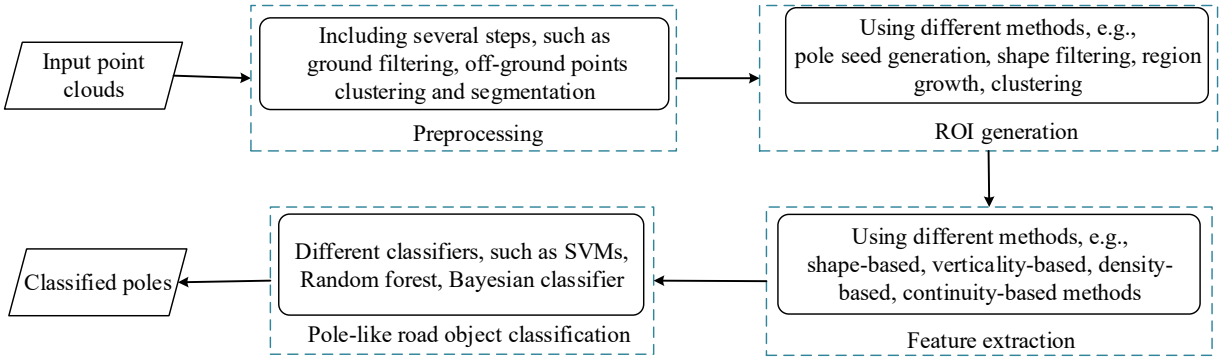



**Fig. 13.** The common flowchart for POD.

components, authors separate the pole-like objects from trees efficiently. Yadav et al. (2016) provided a POD method by utilizing the voxel data structure. They developed an eigenvalues-based algorithm to predict the voxel clouds associated with linear geometrical structure like objects as the final results. Utilized the spatial topological connection, Zhong et al. (2017) provided an octree based segmentation method to detect individual trees using both TLS and MLS point clouds. For the detection of trunks, authors designed a top-down hierarchical segmentation algorithm, combing with the connectivity-based spatial clustering and stem-based initial segmentation. Ordez et al. (2017) transformed the 3D point cloud into voxel data structures, and voxels with vertical continuity and a minimum height are selected as the pole-like objects. Some other similar works include Kang et al. (2018) and Rodrguez-Cuenca et al. (2015). Kang et al. (2018) designed a detection and classification POD method based on the voxel data structure. More specifically, a well-designed model with an adaptive radius and vertical region growing algorithm is used to extract the pole-like objects. Different from the above methods, authors in Rodrguez-Cuenca et al. (2015) developed an anomaly detection algorithm to detect and classify the pole-like objects. They also utilized the continuity in each voxel to determine whether it belongs to pole-like object. It is also an available idea to project the 3D point cloud into 2D space. Using the continuity in the 2D representation, Cabo et al. (2014) developed a strategy to separate the pole-like points and others.

#### 6.1.5. Other methods

There are also some other methods using different ideas. Authors in Tu et al. (2021) provided an extract method based on plane filtering algorithm from MLS point clouds. Focusing on the vehicle localization improvement, authors in Lee et al. (2020) constructed a pole feature map by utilizing the pole-shape characteristic. Recently, from the pole-like 3D distribution perspective. Huang et al. (2021b) proposed a clustering method for the extraction of the trunk parts accurately by adopting a divergence-incorporated idea.

### 6.2. Summary

The above different kinds of methods have been proposed and achieved excellent performance in some datasets or urban road scenes. According to the above review, several observations can be drawn:

- Shape-based and density-based methods usually can enjoy stronger robustness, while the verticality and continuity based methods can achieve higher accuracy. The reason mainly lies in that shape- and density-based methods would lose some important information when projecting the raw 3D point clouds into regular data structures, which may limit their descriptiveness. The verticality and continuity based methods directly process the points. This framework may make them more sensitive to noise.
- Due to depending heavily on the handcrafted features, all these methods tend to capture the shallow feature. This may limit their descriptiveness and universality. More attention must be paid to design models to capture more descriptive characters. An available idea is to apply the deep learning technologies in 3D point clouds, which have been shown the power in mining deep features.

## 7. Tree detection

As an important part of the supports in HD map, street trees are of great help in pavement maintenance and minimizing the erosion. In addition, they can also serve as the natural guardrails and provide supporting information to guide the driving (Safaie et al., 2021). Therefore, extracting street trees accurately is one of the important contents for HD maps in urban areas.

### 7.1. Methods

Detecting objects from images is of great importance in image processing field. Generally, there are two main categories for image-based detection methods: one-stage and two-stage methods. Considering that detecting street trees from images is similar with other object detection tasks from images, there are relatively few image-based methods designed for street tree detection specifically. The newest related work is the method in Xie et al. (2020). Focusing on the tree detection with low illumination and heavily occlusion conditions, to reduce the false detection, author designed a loss function to guide the model, as well as a tree part-attention module. Similar to the common methods for the object detection in image, authors adopt the R-CNN as the basic framework. The designed method achieves excellent detection performance on the dataset. However, it is highly related to the set of images, which would limit its generalization.

Conversely, several works have been undertaken to extract the street trees in 3D point clouds. Generally, there are two main classes for the existing methods: region growing- and clustering- based methods. Fig. 14 provides some classic works, while Fig. 15 presents two typical frameworks for these two kinds of categories.

#### 7.1.1. Region growing-based methods

As one of the most classic categories, region growing-based methods take the segmentation as basic technology and obtain satisfactory performance. The classic method is provided in Besl and Jain (1988). Authors proposed a coarse-to-refine segmentation method. In the first step, they utilized the mean and Gaussian curvature of points for coarse segmentation, while in the second step, a variable order bivariate surface fitting ideal is designed to support the iterative region growing model and then obtain a refinement segmentation. Several later works have followed this basic ideal, such as Nurunnabi et al. (2012), Tvri and Pfeifer (2005), and Gorte (2002). In Gorte (2002), authors used the triangulated irregular network (TIN) to obtain a region growing

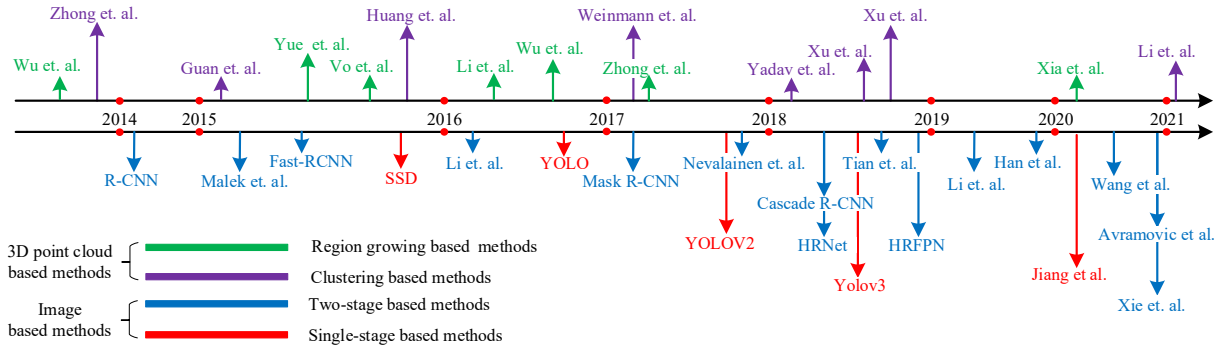



**Fig. 14.** Chronological overview of the typical methods for street tree detection.

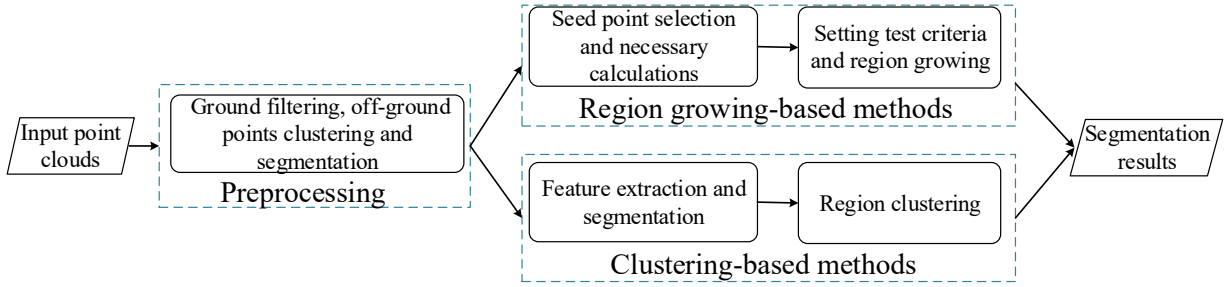



**Fig. 15.** Typical flowchart for region growing algorithm (up branch) and clustering algorithm (down branch), respectively.

segmentation for the first step. In Tvri and Pfeifer (2005), the seed region is firstly randomly picked from the raw dataset, and the normal vectors and distance between different kinds of points are utilized for the fusion of different seed regions. While in Nurunnabi et al. (2012), authors also shared the same idea and obtained excellent results. Similar works include Vieira and Shimada (2005), Ning et al. (2009) and Deschaud and Goulette (2010). Over the past decade, several region growing methods have also proposed. Based on the region growing idea, Li et al. (2012) provided a segmentation method with the region growing idea in a designed top-to-down growing way. In Vega et al. (2014), authors proposed PTrees for the segmentation of the individual tree from raw point clouds, and achieved excellent performance. Tao et al. (2015) designed the shortest-path method for the segmentation of the over-lapping region of trees. In addition, by designing a dual-growing strategy, Li et al. (2016b) proposed a segment method to separate trees from other objects.

Since the above methods usually require high time consumption, several works have been undertaken by using the octrees data structure or voxel grid structure to improve the processing efficiency. As a pioneering work, Woo et al. (2002) takes the voxel data structure into account. Authors in this work first divided the raw data into voxels recursively with a given smaller resolution. These divided smaller voxels are merged to generate leaf node, and then obtain the final region. Based on Woo et al. (2002), utilizes the residual distance to improve the performance. Wang and Tseng (2011) takes the splitting-merging idea to generate the region area in high efficiency. Zhang et al. (2013) proposed a surface growing-based methods with a low computational burden. Combining the normalized cut segmentation and voxelization data transformation, Reitberger et al. (2009) segment the individual tree in excellent accuracy and efficiency. Yu et al. (2015b) provides a semi-automated algorithm to extract the street light poles from MLS point clouds. As the development of voxel, the supervoxel is also introduced for the processing of 3D point clouds. In Aijazi et al. (2013), a subsequent supervoxel segmentation-based method is proposed to organize the 3D data. In Xu et al. (2018c), authors developed a three-step supervoxel approach to segment street trees from raw MLS with high efficiency. By utilizing the spatial information of points along with vertical and horizontal directions, authors in Wu et al. (2013) designed a supervoxel-based morphological model for the detection of the individual tree. They proposed a voxel method with marked neighborhood searching strategy to identify street trees and obtained 98 % accuracy for tree detection. Although this kinds of methods can obtain excellent results, they are usually sensitive to the seed points or initial regions. In addition, they tend to have a higher time consumption.

#### 7.1.2. *Clustering-based methods*

Different from the region growing-based methods, clustering-based methods apply the related clustering algorithms directly. The common algorithms include the mean shift method, the hierarchical and k-means clustering methods. Since the mean shift algorithm enjoys several conveniences, such as operating without defining the number of expected models, some works (Ferraz et al., 2012; Schmitt et al., 2013; Yao et al., 2013; Shahzad et al., 2015) have taken this clustering algorithm as the basic technology. In addition, several other methods have also undertaken. Based on the commercial TerraScan software, Lin and Hyyppa (2012) proposed a clustering-based method to extract tree-related point clusters from multi-echo-recording laser scanning point clouds. Mongus and alik (2015) provided a clustering algorithm for the segmentation of the individual tree in LiDAR data. Specifically, authors used the graph theory to design a unified mathematical framework, and utilized a marker-controlled clustering algorithm to segment the individual trees. In addition, Hamraz et al. (2016) designed a non-parametric model to extract trees by using the crown shape and height of the vegetation. In Weinmann et al. (2017), authors provided a novel method to extract the individual tree from point clouds with two steps. By designing a top-down segmentation method, Zhong et al. (2017) used a three-step strategy, the connectivity spatial clustering, stem-based segmentation, and normalized cut refinement, to extract individual trees with high accuracy. However, the performance of this method is easy to be affected by the point density. Conversely, to detect clusters belonging to the no-photosynthetic component, authors in Xu et al. (2018b) presented a hierarchical method with a bottom-up way.

#### 7.1.3. Other methods

The above two kinds of methods can achieve excellent performance, while several other schemes have also been developed. Lin et al. (2012) designed a RD-schematic algorithm with a three-level canopy-surface-model to segment individual trees from MLS. Liang et al. (2012) utilized the spatial distribution pattern of laser points to present an automatic stem-mapping algorithm. Reitberger et al. (2009) and Yu et al. (2015b) provided their models to separate individual trees from 3D point clouds with normalized cut (Ncut). However, since the normalized cut method is difficult to resolve the segmentation with more than two trees, it may not work for these methods to obtain the complete individual tree.

### 7.2. Summary

- According to the performance of several classic methods, there are two main issues for region growing based methods, related low robustness and extensive computing time. The corresponding two reasons may lie in that, 1) they are sensitive to the seed points or initial regions, which means these methods strongly depend on the selection of such points or regions, and 2) the point/region-wise growing strategy may require high time consumption when processing large-scale scenes.
- Unlike region growing methods, clustering algorithm enjoys stronger robustness since it operates without seed points or initial regions. However, clustering-based methods strongly depend on the features that used for clustering operation. These features should be well designed by selecting appropriate size of neighborhood and the density of points. Furthermore, clustering based methods usually need special-designed strategies for refinement phase, which would reduce the universality.

## 8. Moving objects detection

Moving objects must be removed from the road environments when generating HD mappings. Generally, moving elements mainly contain the pedestrians and vehicles. Guo et al. (2021b) provided a complete review of 3D object detection for common task, which contains the moving objects detection from 3D point clouds. More details can be found in Guo et al. (2021b). Since the pedestrian and vehicle detection from images is similar to other common detection tasks in image processing field, we mainly review the existing methods in 3D point cloud for HD map generation. Similar to the target extraction from the image, generally, the moving object detection from point clouds has two main classes: two-stage based and single-stage based methods.

### 8.1. Two-stage-based methods

As like the two-stage based detection methods in images, two-stage based methods for 3D point clouds first mine several potential regions for target objects, then compute semantic features to classify these regions. According to the way of extracting potential region, there are three classes for these methods: multiply views-, segmentation-, and frustum-based methods (Guo et al., 2021b). Fig. 16 provides three typical frameworks for different kinds of methods.

**Multi-view-based methods**. Transforming point clouds into the image space then processing these images using developed technologies is a straightforward way, which is the basic idea of this kind of methods. More specifically, these methods extract features from multiply views, such as the bird's view and LiDAR front view for the generation of the potential bounding boxes. As the classic method, Chen et al. (2017) used the bird's view to obtain the potential regions, then extract features from these regions to determine the final 3D boxes. However, for this method, mining features from each view would require highly computational cost. To improve the efficiency, several works have been undertaken. Ku et al. (2018) provided a multi-modality model by fusing both bird's-image view information. Liang et al. (2019) developed a multi-task method by utilizing multiply sensors, and obtained significant improvement on all tasks. In Zeng et al. (2018), authors designed a pre-ROI pooling convolution and have achieved excellent performance with high accuracy and efficiency. In addition, works (Liang et al., 2018; Luo et al., 2018) also bring improvement from different aspects.

**Segmentation-based methods**. Different from the above methods,

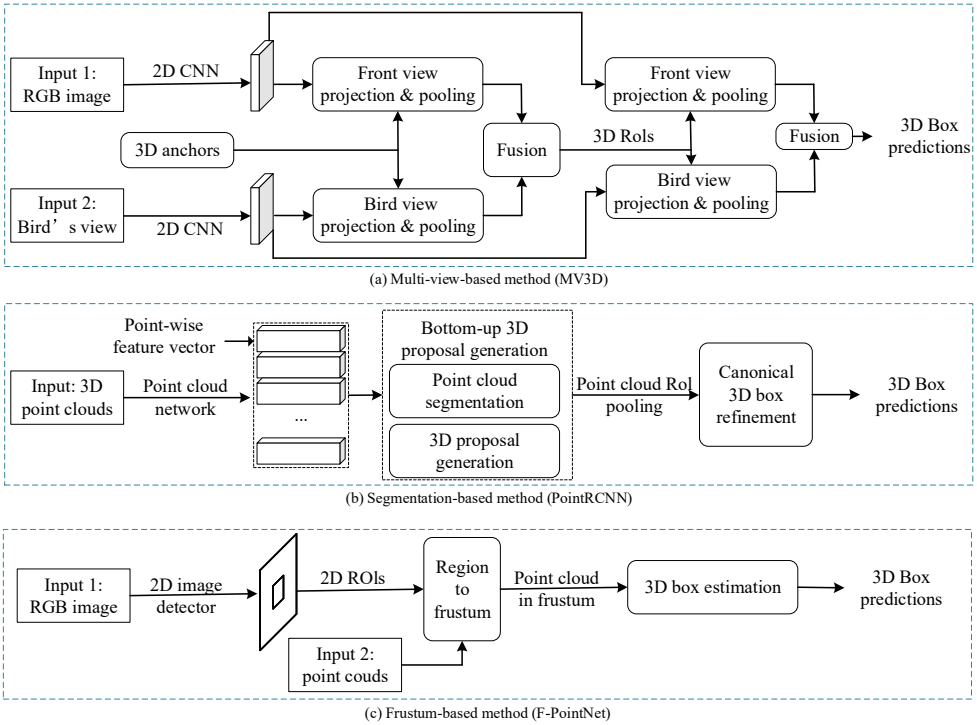



**Fig. 16.** Typical frameworks for (a) multi-view-based method (Chen et al., 2017), (b) segmentation-based method (Shi et al., 2019), and (c) frustum-based method (Qi et al., 2018), respectively.

segmentation based methods process directly the 3D point clouds. Generally, they first extract the semantic features for each point, then obtain high potential target points for the object regions. Compared to multi-view methods, these methods enjoy highly computing efficiency and robustness. Yang et al. (2018c) provided the IPOD, a well-designed intensive point-based object detector with high efficiency. Based on Yang et al. (2018c), Shi et al. (2019) developed a famous model, the PointRCNN, which segments the 3D point clouds directly to obtain the potential points and fuses these points to form the final 3D boxes. Inspired by Shi et al. (2019), authors in Zarzar et al. (2019) provided the Graph Convolution Networks (GCNs) based framework to obtain more accurate bounding boxes. By developing a spherical anchor, Yang et al. (2019b) introduced an efficient model combining a well-designed PointPool layer and a parallel IoU branch, with high accuracy detection performance. Besides, other methods are also proposed to improve the detection performance. Several classic works include Vora et al. (2020), Lang et al. (2019) and Zhou and Tuzel (2018).

**Frustum-based methods**. As one kind of the very effective methods, frustum-based detection strategy generates the 3D frustum regions by using 2D object detectors, which would bring a significant improvement in efficiency. As the pioneering work, F-PointNet (Qi et al., 2018) designed the frustum detection framework. It uses the basic point-wise feature detector, such as the PointNet (Qi et al., 2017a) or PointNet++ (Qi et al., 2017b), to mine the semantic feature. Following F-PointNet, Zhao et al. (2019) combined PointSIFT (Jiang et al., 2018) to enhance the feature description, and proposed the PointSENet to improve the utilization of features. To accurately predict the 3D boxes, authors in Xu et al. (2018a) provided a well-designed dense module to fuse the global features of frustum points and image features. Utilizing the idea of sequence data, Wang and Jia (2019) proposed a sequence-frustum based model by using the corresponding image. With the support of 2D images, Shin et al. (2019) first operates in 2D images to obtain the 2D boxes and 3D poses of objects, and then predict the 3D boxes by extracting point-level features.

In addition, several other kinds of methods are also provided. Chen et al. (2019) take the point and voxel representations as inputs, and obtain a high speed run performance. Shi et al. (2020) provides a similar work, the PointVoxel-RCNN (PV-RCNN), which has obtained excellent performance in KITTI dataset. Utilizing the Hough voting algorithm in 2D image processing, authors in Qi et al. (2019) developed the VoteNet, which can directly compute the center point and generate the 3D bounding boxes with high accuracy. Several follow-up methods are also proposed, such as the Shi et al. (2021) and Qi et al. (2020).

### 8.2. Single-stage-based methods

These methods directly predict the 3D boxes. Thus they usually enjoy a better efficiency. Similarly, there are three main classes for these methods: bird's-view-, discretization-, and point-based methods. Fig. 17 provides typical frameworks for different kinds of methods.

**Bird's-view-based methods**. Projecting disordered and discrete 3D point clouds into a regular data structure is a feasible idea. As the classic methods, authors in Yang et al. (2018b) transformed the point cloud into a regular representation with given subspace resolution. Several following methods are then proposed, including the Vote3Deep (Engelcke et al., 2017), 3D-FCN (Li, 2017), and VeloFCN (Li et al., 2016a,b,c). In addition, by using the information from HD maps, authors in Yang et al. (2018a) combined the geometric and prior knowledge to obtain excellent performance with strong robustness. To further improve this method, Beltrn et al. (2018) designed a normalization map to support 3D point clouds come from different LiDAR sensors.

**Discretization-based methods**. Transforming the discrete 3D point cloud into regular voxel data, followed by applying the 3D CNN to extract the features from voxel data structure, can also be considered as a feasible idea. Li et al. (2019) had taken this basic idea to process the 3D object detection task and obtain acceptable results with low computational and memory burden. As the common basic framework, VoxelNet (Zhou and Tuzel, 2018) is based on the voxel data structure. Built on VoxelNet, Yan et al. (2018) proposed an end-to-end trainable method by using a sparse convolutional network and a well-designed sine-error angle loss. Using both the 2D images and 3D data information, Sindagi et al. (2019) extended the VoxelNet into a more efficient method. Different from the above voxel-based frames, Lang et al. (2019) designed the PointPillars, which discretizes the 3D point clouds into a vertical columns data structure.

**Point-based methods**. It is a worth trying way to processing directly the point clouds to detect 3D objects. The pioneering work of these methods is the 3DSSD (Yang et al., 2020), which provides the well-designed fusion sampling strategy to obtain lower time consuming. Compared with some two-stage point-based method, this method performs with a significant improvement, while maintaining a high efficiency.

### 8.3. Stereo image-based methods

As the pioneering work, Pseudo-LiDAR (Wang et al., 2019) makes full use of the depth information of the stereo image. Specifically, it first uses the common framework to obtain the corresponding depth map from stereo images. Then the pseudo LiDAR is obtained by combining the original image with the depth information. Finally, the pseudo LiDAR is used to replace the original point clouds, and the 3D target detection model is constructed in the form of bird's eye view. It has made outstanding achievements in the data set KITTI. Several following methods (Qian et al., 2020; Ma et al., 2020; Andrea et al., 2021) take the same framework. Addition, authors in Chen et al. (2020b) proposed a deep stereo geometry network, DSGN, for 3D object detection. It takes stereo images as input, extracts the feature through a Siamese network and constructs the plane sweep volume (PSV). Then, it learns pixel

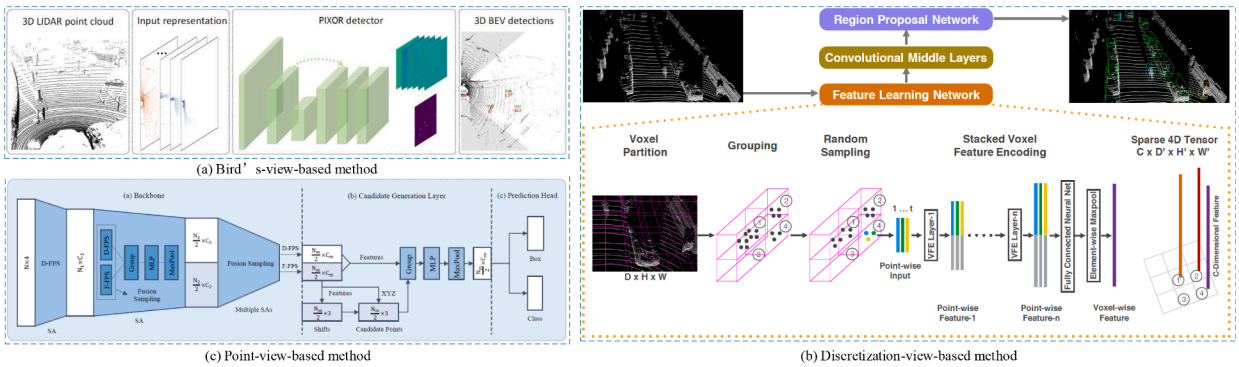



**Fig. 17.** Typical frameworks for (a) bird's-view-based methods (Yang et al., 2018b), (b) discretization-based methods (Yang et al., 2018b), and (c) point-based methods (Yang et al., 2020), respectively.

correspondence in volume to convert PSV to stereo geometry representation to establish 3D geometry model in 3D space. It achieves comparable performance on the KITTI dataset. Different from Chen et al. (2020b), authors in Guo et al. (2021a) designed a LiDAR geometry aware representations by mining the stereo image information. They use a designed LiDAR based detector to guide the stereo based detector. Beside, in Qin et al. (2019), by utilizing the triangulation, authors proposed a stereo triangulation learning network (TLNet) to detect 3D objects from stereo images. Different from other stereo image based methods, it does not need to calculate pixel level depth maps and can be easily integrated into the baseline monocular detector. In order to enhance the information channel of view specific RoI features, they also proposed a feature re-evaluation strategy, which is conducive to triangulation learning by focusing the network attention on the key parts of an object. All these methods take the stereo image as the important input data, and have shown impressive performance in 3D object detection. However, it would also be aware that to obtain the precise location, they should combine the 3D point clouds. In practice, such methods can be regarded as a useful supplement to the point cloud-based methods.

### 8.4. Other methods

Several methods that enjoy interesting ideas have also been undertaken. By utilizing the hotspot representation, Chen et al. (2020a) designed a new method, which can achieve comparable performance when processing sparse scenes. Meyer et al. (2019b) provided a novel probability distribution selecting strategy to improve the efficiency, while Meyer et al. (2019a) combined the RGB images and LiDAR points to extend the method in Meyer et al. (2019b) to be a more efficiency-one. In addition, by exploiting the graph neural network, Shi and Rajkumar (2020) provided the Point-GNN as the 3D object detector.

### 8.5. Summary

- For two-stage based methods, the multi-view based branch can achieve excellent performance by fusing different direction views, which also means that the efficiency would be related low. In addition, in the complicated environments, such as the highly occluded and crowded objects, these methods may not work well. Compared with multiply views-based algorithms, segmentation-based methods can obtain higher detection accuracy and can enjoy better robustness when processing complicated environments. Frustum-based methods can make full use of the 2D image processing technologies, which also means that these methods may heavily depend on 2D image technologies, so would bring limitations for their performance.
- Compared to two-stage methods, single-stage methods would enjoy a higher run speed. The bird's-view-based methods can obtain excellent performance when the point clouds are dense. However, facing the sparse point clouds, the value in each pixel for the view is hard to determine accurately. Discretization-based methods can use the developed technologies for regular data structure to obtain related high accuracy results. But inevitably, they may cost more computing resources and lose some spatial structure information.

## 9. Conclusion

This paper provides a contemporary full-element survey of road object detection methods for HD map. It contains sign-objects detection, pole-like object detection, guardrails detection, street tress detection and moving object detection. We present the merits and demerits for all these methods. In addition, the performance comparison on several public datasets have also been provided. Although there are many methods that have been undertaken for full-element detection, several important issues should be discussed in depth.

Firstly, except for the moving object detection, most 3D point cloud detection methods are based on traditional idea. By designing feature descriptors for specific object, they build detection models based on traditional classifiers, such as SVMs, random forest and Bayesian classifier. The reason for this may be that the method based on deep learning requires a large amount of labeled data. At present, except for mobile object detection (i.e. vehicles and pedestrians), there are too few labeled public data sets. The reason for this may be that the deep learning based methods require a large amount of labeled data. In addition, it is well known that labeling 3D point cloud data usually requires a lot of effort. Therefore, it is feasible to utilize the transfer learning, such as the domain adaptation, to resolve this problem. For example, if we can obtain 3D point cloud annotation data of virtual scenes through some simulation softwares, we may alleviate the above problem to some extent. In addition, the use of generative models may also reduce the effort to annotate data.

Secondly, we may have to think deeply about the question, i.e., is it better to construct HD maps using an end-to-end method for all road information extraction and road object detection? It is well known that designing different detection methods for different elements will reduce the efficiency of generating HD maps. In addition, utilizing the multi-stage detection strategy will also increase the probability of errors. Therefore, it is reasonable to think about how to build an end-to-end approach. However, at present, to the best of our knowledge, there is no such method. In our opinion, there may be two main important reasons. First, it is difficult to design a large model that can simultaneously extract the road information and detect road objects. It is difficult to design a full-element detection model due to the large differences between the road surface information and the road objects in terms of distribution and shape. Second, at present, there are few full-element annotation data. It may require significant labeling costs to obtain such full-element annotation data. However, since it has great potential value in improving the accuracy and efficiency of HD map generation, we still believe that this may be a problem worth studying by all researchers in the coming years.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 41871380. The authors would like to acknowledge the anonymous reviewers for their valuable comments in improving the paper.

## References

Aijazi, A.K., Checchin, P., Trassoudaine, L., 2013. Segmentation based classification of 3D urban point clouds: A super-voxel based approach with evaluation. Remote Sens. (Basel) 5 (4), 1624–1650.

Alshehhi, R., Marpu, P.R., 2017. Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. ISPRS J. Photogramm. Remote Sens. 126, 245–260.

Aly, M., 2008. Real time detection of lane markers in urban streets. IEEE Intelligent Vehicles Symp. 7–12.

Andrea, S., Samuel, R., Lorenzo, P., Peter, K., Elisa, R., 2021. Are we missing condence in pseudo-lidar methods for monocular 3d object detection. In: 2021 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 3225–3233.

Avramovi, A., Sluga, D., Tabernik, D., Skoaj, D., Stojni, V., Ilc, N., 2020. Neural-network-based traffic sign detection and recognition in high-definition images using region focusing and parallelization. IEEE Access 8, 189855–189868.

Bailo, O., Lee, S., Rameau, F., Yoon, J.S., Kweon, I.S., 2017. Robust road marking detection and recognition using density-based grouping and machine learning techniques. IEEE, Santa Rosa, CA, USA, pp. 760–768.

Bar Hillel, A., Lerner, R., Levi, D., Raz, G., 2014. Recent progress in road and lane detection: a survey. Mach. Vis. Appl. 25 (3), 727–745.

Barnes, N., Loy, G., 2006. Real-time regular polygonal sign detection. in: Field and Service Robotics. Berlin, Heidelberg, pp. 55–66.

Barnes, N., Zelinsky, A., Fletcher, L.S., 2008. Real-time speed sign detection using the radial symmetry detector. IEEE Trans. Intell. Transp. Syst. 9 (2), 322–332.
Baro, X., Escalera, S., Vitria, J., Pujol, O., Radeva, P., 2009. Traffic sign recognition using evolutionary adaboost detection and forest-ecoc classification. IEEE Trans. Intell. Transp. Syst. 10 (1), 113–126.
Batra, A., Singh, S., Pang, G., Basu, S., Jawahar, C. V., Paluri, M., 2019. Improved road connectivity by joint learning of orientation and segmentation. pp. 10385–10393.
Beltrn, J., Guindel, C., Moreno, F. M., Cruzado, D., Garca, F., De La Escalera, A., 2018. BirdNet: A 3D object detection framework from lidar information. In: International Conference on Intelligent Transportation Systems (ITSC). pp. 3517–3523.
Besl, P., Jain, R., 1988. Segmentation through variable-order surface fitting. IEEE Trans. Pattern Anal. Mach. Intell. 10 (2), 167–192.
Borkar, A., Hayes, M., Smith, M.T., 2011. Polar randomized hough transform for lane detection using loose constraints of parallel lines. IEEE 1037–1040.
Bosch, A., Zisserman, A., Munoz, X., 2007. Representing shape with a spatial pyramid kernel. in: Proceedings of the 6th ACM international conference on Image and video retrieval (CIVR). pp. 401–408.
Boumediene, M., Lauffenburger, J.-P., Daniel, J., Cudel, C., Ouamri, A., 2014. Multi-roi association and tracking with belief functions: Application to traffic sign recognition. IEEE Trans. Intell. Transp. Syst. 15 (6), 2470–2479.
Broggi, A., Cerri, P., Medici, P., Porta, P.P., Ghisio, G., 2007. Real time road signs recognition. IEEE Intelligent Vehicles Symp. 2007, 981–986.
Bruls, T., Maddern, W., Morye, A.A., Newman, P., 2018. Mark yourself: Road marking segmentation via weakly-supervised annotations from multimodal data. IEEE, Brisbane, QLD, pp. 1863–1870.
Cabo, C., Ordoez, C., Garca-Corts, S., Martnez, J., 2014. An algorithm for automatic detection of pole-like street furniture objects from mobile laser scanner point clouds. ISPRS J. Photogramm. Remote Sens. 87, 47–56.
Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2020. nuScenes: A multimodal dataset for autonomous driving. in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11618–11628.
Cao, Y., Chen, K., Loy, C. C., Lin, D., 2020. Prime sample attention in object detection. in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11580–11588.
Cao, J., Zhang, J., Huang, W., 2021. Traffic sign detection and recognition using multi-scale fusion and prime sample attention. IEEE Access 9, 3579–3591.
Cattaneo, D., Vaghi, M., Fontana, S., Ballardini, A.L., Sorrenti, D.G., 2020. Global visual localization in lidar-maps through shared 2D–3D embedding space. In: 2020 AAAI Conference on Articial Intelligence. IEEE, pp. 4365–4371.
Chan, T.H., Jia, K., Gao, S., Lu, J., Zeng, Z., Ma, Y., 2015. PCANet: A simple deep learning baseline for image classification? IEEE Trans. Image Process. 24 (12), 5017–5032.
Chen, T., Chen, Z., Shi, Q., Huang, X., 2015. Road marking detection and classification using machine learning algorithms. pp. 617–621.
Chen, Y., Liu, S., Shen, X., Jia, J., 2019. Fast point R-CNN. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9774–9783.
Chen, Y., Liu, S., Shen, X., Jia, J., 2020b. DSGN: Deep stereo geometry network for 3D object detection. In: 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 12536–12545.
Chen, X., Ma, H., Wan, J., Li, B., Xia, T., 2017. Multi-view 3D object detection network for autonomous driving. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6526–6534.
Chen, Q., Sun, L., Wang, Z., Jia, K., Yuille, A., 2020a. Object as hotspots: An anchor-free 3D object detection approach via firing of hotspots. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 68–84.
Chen, Z., Wang, C., Li, J., Xie, N., Han, Y., Du, J., 2021. Reconstruction bias u-net for road extraction from optical remote sensing images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2284–2294.
Chen, L., Zhu, Q., Xie, X., Hu, H., Zeng, H., 2018. Road extraction from VHR remote-sensing imagery via object segmentation constrained by Gabor features. ISPRS Int. J. Geo Inf. 7 (9), 362.
Cheng, G., Wang, Y., Gong, Y., Zhu, F., Pan, C., 2014. Urban road extraction via graph cuts based probability propagation. IEEE 5072–5076.
Cheng, G., Zhu, F., Xiang, S., Wang, Y., Pan, C., 2016. Accurate urban road centerline extraction from vhr imagery via multiscale segmentation and tensor voting. Neurocomputing 205, 407–420.
Cheng, G., Han, J., Lu, X., 2017a. Remote sensing image scene classification: Benchmark and state of the art. Proc. IEEE 105 (10), 1865–1883.
Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., Pan, C., 2017b. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Trans. Geosci. Remote Sens. 55 (6), 3322–3337.
Choi, H.-C., Oh, S.-Y., 2010. Illumination invariant lane color recognition by using road color reference and neural networks. IEEE 1–5.
Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. IEEE, Las Vegas, NV, USA, pp. 3213–3223.
Das, S., Mirnalinee, T.T., Varghese, K., 2011. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. IEEE Trans. Geosci. Remote Sens. 49 (10), 3906–3931.
De la Escalera, A., Moreno, L., Salichs, M., Armingol, J., 1997. Road traffic sign detection and classification. IEEE Trans. Ind. Electron. 44 (6), 848–859.
Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images. IEEE, Salt Lake City, UT, pp. 172.01–172.09.
Deng, J., Han, Y., 2013. A real-time system of lane detection and tracking based on optimized ransac b-spline fitting. ACM Press, Montreal, Quebec, Canada, pp. 157–164.
Deschaud, J.-E., Goulette, F., 2010. A fast and accurate plane detection algorithm for large noisy point clouds using filtered normals and voxel growing. In: International Symposium on 3D Data Processing, Visualization and Transmission. Paris France, pp. 1–8.
Dollr, P., Appel, R., Belongie, S., Perona, P., 2014. Fast feature pyramids for object detection. IEEE Trans. Pattern Anal. Mach. Intell. 36 (8), 1532–1545.
Elhousni, M., Lyu, Y., Zhang, Z., Huang, X., 2020. Automatic building and labeling of HD maps with deep learning. In: The Thirty-Second Innovative Applications of Artificial Intelligence Conference (AAAI), pp. 13255–13260.
Ellahyani, A., Ansari, M.E., Jaafari, I.E., 2016. Traffic sign detection and recognition based on random forests. Appl. Soft Comput. 46, 805–815.
Engel, N., Belagiannis, V., Dietmayer, K., 2021. Attention-based vehicle self-localization with HD feature maps. arXiv preprint, arXiv: 2107.07787.
Engelcke, M., Rao, D., Wang, D.Z., Tong, C.H., Posner, I., 2017. Vote3deep: Fast object detection in 3D point clouds using efficient convolutional neural networks. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 1355–1361.
Fang, C.-Y., Chen, S.-W., Fuh, C.-S., 2003. Road-sign detection and tracking. IEEE Trans. Veh. Technol. 52 (5), 1329–1341.
Feng, M., Hu, S., Ang, M.H., Lee, G.H., 2019. 2D3D-Matchnet: Learning to match keypoints across 2D image and 3d point cloud. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp. 4790–4796.
Ferraz, A., Bretar, F., Jacquemoud, S., Gonalves, G., Pereira, L., Tom, M., Soares, P., 2012. 3-D mapping of a multi-layered mediterranean forest using ALS data. Remote Sens. Environ. 121, 210–223.
Fukano, K., Masuda, H., 2015. Detection and classification of pole-like objects from mobile mapping data. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W5, 57–64.
Gao, J., Chen, Y., Junior, J.M., Wang, C., Li, J., 2020. Rapid extraction of urban road guardrails from mobile lidar point clouds. IEEE Trans. Intell. Transp. Syst. 1–6.
Gao, L., Shi, W., Miao, Z., Lv, Z., 2018a. Method based on edge constraint and fast marching for road centerline extraction from very high-resolution remote sensing images. Remote Sens. (Basel) 10 (6), 900.
Gao, L., Song, W., Dai, J., Chen, Y., 2019. Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network. Remote Sens. (Basel) 11 (5), 552.
Gao, X., Sun, X., Zhang, Y., Yan, M., Xu, G., Sun, H., Jiao, J., Fu, K., 2018b. An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. IEEE Access 6, 39401–39414.
Geiger, A., Lenz, P., Urtasun, R., 2012. Are we ready for autonomous driving? the KITTI vision benchmark suite. in: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361.
Gil Jimnez, P., Bascn, S.M., Moreno, H.G., Arroyo, S.L., Ferreras, F.L., 2008. Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2d homographies. Signal Process. 88 (12), 2943–2955.
Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 580–587.
Girshick, R., 2015. Fast R-CNN. in: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1440–1448.
Gmez Serna, C., Ruichek, Y., 2018. Classification of traffic signs: The European dataset. IEEE Access 6, 78136–78148.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial Nets. Advances in neural information processing systems 27.
Gorte, B., 2002. Segmentation of tin-structured surface models. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-6, 1–5.
Greenhalgh, J., Mirmehdi, M., 2012. Real-time detection and recognition of road traffic signs. IEEE Trans. Intell. Transp. Syst. 13 (4), 1498–1506.
Greenhalgh, J., Mirmehdi, M., 2015. Recognizing text-based traffic signs. IEEE Trans. Intell. Transp. Syst. 16 (3), 1360–11039.
Grigorescu, C., Petkov, N., 2003. Distance sets for shape filters and shape recognition. IEEE Trans. Image Process. 12 (10), 1274–1286.
Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B., 2014. Using mobile laser scanning data for automated extraction of road markings. ISPRS J. Photogramm. Remote Sens. 87, 93–107.
Guan, H., Li, J., Yu, Y., Ji, Z., Wang, C., 2015. Using mobile lidar data for rapidly updating road markings. IEEE Trans. Intell. Transp. Syst. 16 (5), 2457–2466.
Guo, X., Shi, S., Wang, X., Li, H., 2021a. LIGA-Stereo: Learning lidar geometry aware representations for stereo-based 3D detector. In: 2021 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 3153–3163.
Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2021b. Deep learning for 3D point clouds: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (12), 4338–4364.
Gurghian, A., Koduri, T., Bailur, S.V., Carey, K.J., Murali, V.N., 2016. Deeplanes: End-to-end lane position estimation using deep neural networks. IEEE, Las Vegas, NV, USA, pp. 38–45.
Hamraz, H., Contreras, M.A., Zhang, J., 2016. A robust approach for tree segmentation in deciduous forests using small-footprint airborne lidar data. Int. J. Appl. Earth Obs. Geoinf. 52, 532–541.
Han, C., Gao, G., Zhang, Y., 2019. Real-time small traffic sign detection with revised faster-rcnn. Multimed. Tools Appl. 78 (10), 13263–13278.
Hao, W., Wang, Y., Li, Y., Shi, Z., Zhao, M., Liang, W., 2018. Hierarchical extraction of pole-like objects from scene point clouds. Opt. Eng. 57 (8), 1–11.

He, B., Ai, R., Yan, Y., Lang, X., 2016. Accurate and robust lane detection based on dual-view convolutional neutral network. IEEE, Gotenburg, Sweden, pp. 1041–1046.
He, K., Gkioxari, G., Dollr, P., Girshick, R., 2017. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988.
He, L., Jiang, S., Liang, X., Wang, N., Song, S., 2021. Diff-Net: Image feature difference based high-definition map change detection. In: arX-iv,doi:2107.07030.
Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., Igel, C., 2013. Detection of traffic signs in real-world images: The german traffic sign detection benchmark. In: The 2013 International Joint Conference on Neural Networks (IJCNN). pp. 1–8.
Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Chen, L., Jain, A., Omari, S., Iglovikov, V., Ondruska, P., 2020. One thousand and one hours: Self-driving motion prediction dataset. arXiv preprint, arXiv: 2006.14480.
Hu, Y., Li, X., Xie, J., Guo, L., 2011. A novel approach to extracting street lamps from vehicle-borne laser data. in: International Conference on Geoinformatics. pp. 1–6.
Hu, Q., Paisitkriangkrai, S., Shen, C., van den Hengel, A., Porikli, F., 2016. Fast detection of multiple objects in traffic scenes with a common detection framework. IEEE Trans. Intell. Transp. Syst. 17 (4), 1002–1014.
Hu, J., Razdan, A., Femiani, J.C., Cui, M., Wonka, P., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE Trans. Geosci. Remote Sens. 45 (12), 4144–4157.
Hu, X., Zhang, Z., Tao, C.V., 2004. A robust method for semi-automatic extraction of road centerlines using a piecewise parabolic model and least square template matching. Photogramm. Eng. Remote Sens. 70 (12), 1393–1398.
Huang, J., You, S., 2015. Pole-like object detection and classification from urban point clouds. in: 2015 IEEE International Conference on Robotics and Automation (ICRA). pp. 3032–3038.
Huang, P., Cheng, M., Chen, Y., Luo, H., Wang, C., Li, J., 2017. Traffic sign occlusion detection using mobile laser scanning point clouds. IEEE Trans. Intell. Transp. Syst. 18 (9), 2364–2376.
Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y., Yang, R., 2018. The apolloscape dataset for autonomous driving. IEEE, Salt Lake City, UT, pp. 1067–10676.
Huang, X., Mei, G., Zhang, J., Abbas, R., 2021a. A comprehensive survey on point cloud registration. arXiv preprint arXiv:2103.02690.
Huang, Y., Ma, P., Ji, Z., He, L., 2021. Part-based modeling of pole-like objects using divergence-incorporated 3-D clustering of mobile laser scanning point clouds. IEEE Trans. Geosci. Remote Sens. 59 (3), 2611–2626.
Hur, J., Kang, S.-N., Seo, S.-W., 2013. Multi-lane detection in urban driving environments using conditional random fields. IEEE 1297–1302.
Husain, A., Vaishya, R.C., Sarif, M.O., 2018. A moving window search method for detection of pole like objects using mobile laser scanner data. Int. J. Comput. Sci. Eng. 6 (3), 1–6.
Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20 (11), 1254–1259.
Jaakkola, A., Hyypp, J., Hyypp, H., Kukko, A., 2008. Retrieval algorithms for road surface modelling using laser-based mobile mapping. Sensors 8 (9), 5238–5249.
Jefri Muril, M., Abdul Aziz, N.H., Ghani, H.A., Ab Aziz, N.A., 2020. A review on deep learning and nondeep learning approach for lane detection system. IEEE, Melaka, Malaysia, pp. 162–166.
Jiang, Y., He, B., Liu, L., Ai, R., Lang, X., 2016. Effective and robust corrugated beam guardrail detection based on mobile laser scanning data. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). pp. 1540–1545.
Jiang, M., Wu, Y., Zhao, T., Zhao, Z., Lu, C., 2018. PointSIFT: A sift-like network module for 3D point cloud semantic segmentation. arXiv preprint, arXiv:1807.00652.
Jifeng, D., Yi, L., Kaiming, H., Jian, S., 2016. R-FCN: object detection via region-based fully convolutional networks. In: Proceedings of the 30th International Conference on Neural Information Processing SystemsDecember. pp. 379–387.
Jinhong, J., Shengli, B., Wenxu, S., Zhenkun, W., 2020. Improved traffic sign recognition algorithm based on yolo v3 algorithm. J. Comput. Appl. 40 (8), 2472.
Kang, D.J., Choi, J.W., Kweon, I.S., 1996. Finding and tracking road lanes using "line-snakes". IEEE, Tokyo, Japan, pp. 189–194.
Kang, Z., Yang, J., Zhong, R., Wu, Y., Shi, Z., Lindenbergh, R., 2018. Voxel-based extraction and classification of 3-D pole-like objects from mobile lidar point cloud data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (11), 4287–4298.
Kaplan Berkaya, S., Gunduz, H., Ozsen, O., Akinlar, C., Gunal, S., 2016. On circular traffic sign detection and recognition. Expert Syst. Appl. 48, 67–75.
Khan, J., Bhuiyan, S., Adhami, R., 2014. Hierarchical clustering of emd based interest points for road sign detection. Opt. Laser Technol. 57, 271–283.
Kim, J., Kim, J., Jang, G.-J., Lee, M., 2017. Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection. Neural Netw. 87, 109–121.
Kim, J., Park, C., 2017. End-to-end ego lane estimation based on sequential transfer learning for self-driving cars. IEEE, Honolulu, HI, USA, pp. 1194–1202.
Kortli, Y., Marzougui, M., Atri, M., 2016. Efficient implementation of a real-time lane departure warning system. IEEE, Hammamet, Tunisia, pp. 1–6.
Ku, J., Mozifian, M., Lee, J., Harakeh, A., Waslander, S.L., 2018. Joint 3D proposal generation and object detection from view aggregation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1–8.
Kumar, P., McElhinney, C.P., Lewis, P., McCarthy, T., 2014. Automated road markings extraction from mobile laser scanning data. Int. J. Appl. Earth Obs. Geoinf. 32, 125–137.
Lam, J., Kusevic, K., Mrstik, P., Harrap, R., Greenspan, M., 2010. Urban scene extraction from mobile ground based lidar data. in: 3D Data Processing Visualization and Transmission (3DPVT). pp. 1–8.
Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O., 2019. PointPillars: Fast encoders for object detection from point clouds. in: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12689–12697.
Larsson, F., Felsberg, M., 2011. Using Fourier descriptors and spatial models for traffic sign recognition. In: Image Analysis. Berlin, Heidelberg, pp. 238–249.
Larsson, F., Felsberg, M., Forssen, P.E., 2011. Correlating Fourier descriptors of local patches for road sign recognition. IET Comput. Vis. 5 (4), 244–254.
Lee, T.-Y., Jeong, M.-H., Peter, A., 2022. Object detection of road facilities using YOLOv3 for high-definition map updates. Sens. Mater. 34 (1), 251–260.
Lee, S., Kim, J., Yoon, J.S., Shin, S., Bailo, O., Kim, N., Lee, T.-H., Hong, H.S., Han, S.-H., Kweon, I.S., 2017b. Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. IEEE, Venice, pp. 1965–1973.
Lee, S.-W., Lin, P.-W., Fu, Y.-T., Hsu, C.-M., Chan, C.-Y., Lin, J.-H., Chiang, Y.-H., 2020. Improving vehicle localization using pole-like landmarks extracted from 3-D lidar scans. In: IEEE Intelligent Vehicles Symposium (IV), pp. 2052–2057.
Lee, D.-K., Shin, J.-S., Jung, J.-H., Park, S.-J., Oh, S.-J., Lee, I.-S., 2017a. Real-time lane detection and tracking system using simple filter and kalman filter. IEEE, Milan, pp. 275–277.
Li, B., 2017. 3D fully convolutional network for vehicle detection in point cloud. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1513–1518.
Li, Q., Chen, L., Li, M., Shaw, S.-L., Nuchter, A., 2014. A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios. IEEE Trans. Veh. Technol. 63 (2), 540–555.
Li, B., Zhang, T., Xia, T., 2016. Vehicle Detection from 3D Lidar Using Fully Convolutional Network. arXiv e-prints, arXiv:1608.07916.
Li, F., Oude Elberink, S., Vosselman, G., 2018. Pole-like road furniture detection and decomposition in mobile laser scanning data based on spatial relations. Remote Sensing 10 (4), 531.1–531.28.
Li, X., Guivant, J. E., Kwok, N., Xu, Y., 2019. 3d backbone network for 3D object detection, arXiv:1901.08373.
Li, W., Guo, Q., Jakubowski, M., Kelly, M., 2012. A new method for segmenting individual trees from the lidar point cloud. Photogramm. Eng. Remote Sens. 78, 75–84.
Li, J., Lee, G.H., 2021. Deep I2P: Image-to-Point cloud registration via deep classication. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 15960–15969.
Li, L., Li, D., Zhu, H., Li, Y., 2016c. A dual growing method for the automatic extraction of individual trees from mobile laser scanning data. ISPRS J. Photogramm. Remote Sens. 120, 37–52.
Li, Y., Ma, L., Zhong, Z., Liu, F., Chapman, M.A., Cao, D., Li, J., 2021. Deep learning for lidar point clouds in autonomous driving: A review. IEEE Trans. Neural Networks Learn. Syst. 32 (8), 3412–3432.
Li, F., Oude Elberink, S., Vosselman, G., 2016b. Pole-like street furniture decompostion in mobile laser scanning data. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial. Inf. Sci. III-3, 193–200.
Li, H., Sun, F., Liu, L., Wang, L., 2015. A novel traffic sign detection method via color segmentation and robust shape matching. Neurocomputing 169, 77–88.
Liang, X., Litkey, P., Hyyppa, J., Kaartinen, H., Vastaranta, M., Holopainen, M., 2012. Automatic stem mapping using single-scan terrestrial laser scanning. IEEE Trans. Geosci. Remote Sens. 50 (2), 661–670.
Liang, M., Yang, B., Wang, S., Urtasun, R., 2018. Deep continuous fusion for multi-sensor 3D object detection. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 1–16.
Liang, M., Yang, B., Chen, Y., Hu, R., Urtasun, R., 2019. Multi-task multi-sensor fusion for 3D object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7337–7345.
Lim, K.H., Ang, L.-M., Seng, K.P., 2009. New hybrid technique for traffic sign recognition. In: International Symposium on Intelligent Signal Processing and Communications Systems, pp. 1–4.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollr, P., 2017. Focal loss for dense object detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007.
Lin, Y., Hyypp, J., Jaakkola, A., Yu, X., 2012. Three-level frame and RD-schematic algorithm for automatic detection of individual trees from MLS point clouds. Int. J. Remote Sens. 33 (6), 1701–1716.
Lin, Y., Hyyppa, J., 2012. Multiecho-recording mobile laser scanning for enhancing individual tree crown reconstruction. IEEE Trans. Geosci. Remote Sens. 50 (11), 4323–4332.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016b. SSD: Single shot multibox detector. In: European Conference on Computer Vision, pp. 21–37.
Liu, C., Chang, F., Chen, Z., 2014. Rapid multiclass traffic sign detection in high-resolution images. IEEE Trans. Intell. Transp. Syst. 15 (6), 2394–2403.
Liu, C., Chang, F., Chen, Z., Liu, D., 2016a. Fast traffic sign recognition via high-contrast region extraction and extended sparse representation. IEEE Trans. Intell. Transp. Syst. 17 (1), 79–92.
Liu, C., Li, S., Chang, F., Wang, Y., 2019a. Machine vision based traffic sign detection methods: Review, analyses and perspectives. IEEE Access 7, 86578–86596.
Liu, L., Wang, Y., Li, K., Li, J., 2020. Focus first: Coarse-to-fine traffic sign detection with stepwise learning. IEEE Access 8, 171170–171183.
Liu, Y., Yao, J., Lu, X., Xia, M., Wang, X., Liu, Y., 2019b. Roadnet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images. IEEE Trans. Geosci. Remote Sens. 57 (4), 2043–2056.

Lu, X., Zhong, Y., Zheng, Z., Liu, Y., Zhao, J., Ma, A., Yang, J., 2019. Multi-scale and multi-task deep learning framework for automatic road extraction. IEEE Trans. Geosci. Remote Sens. 57 (11), 9362–9377.
Luo, W., Yang, B., Urtasun, R., 2018. Fast and furious: Real time end-to-end 3D detection, tracking and motion forecasting with a single convolutional net. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3569–3577.
Lv, Z., Jia, Y., Zhang, Q., Chen, Y., 2017. An adaptive multifeature sparsity-based model for semiautomatic road extraction from high-resolution satellite images in urban areas. IEEE Geosci. Remote Sens. Lett. 14 (8), 1238–1242.
Ma, X., Liu, S., Xia, Z., Zhang, H., Zeng, X., Ouyang, W., 2020. Rethinking pseudo-lidar representation. In: 2020 IEEE International Conference on Computer Vision (ECCV). IEEE, pp. 1–17.
Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., Lopez-Ferreras, F., 2007. Road-sign detection and recognition based on support vector machines. IEEE Trans. Intell. Transp. Syst. 8 (2), 264–278.
Mammeri, A., Boukerche, A., Lu, G., 2014. Lane detection and tracking system based on the mser algorithm, hough transform and kalman filter. ACM Press, Montreal, QC, Canada, pp. 259–266.
Masuda, H., Oguri, S., Jun, H., 2013. Shape reconstruction of poles and plates from vehicle-based laser scanning data. In: International Symposium on Mobile Mapping Technology, pp. 1–6.
Matsumoto, H., Mori, Y., Masuda, H., 2019. Extraction and shape reconstruction of guardrails using mobile mapping data. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLII-2/W13, 1061–1068.
McCall, J., Trivedi, M., 2006. Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation. IEEE Trans. Intell. Transp. Syst. 7 (1), 20–37.
Meyer, G.P., Charland, J., Hegde, D., Laddha, A., Vallespi-Gonzalez, C., 2019a. Sensor fusion for joint 3D object detection and semantic segmentation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1230–1237.
Meyer, G.P., Laddha, A., Kee, E., Vallespi-Gonzalez, C., Wellington, C.K., 2019b. Lasernet: An efficient probabilistic 3D object detector for autonomous driving. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12669–12678.
Mgelmose, A., Liu, D., Trivedi, M.M., 2015. Detection of U.S. traffic signs. IEEE Trans. Intell. Transp. Syst. 16 (6), 3116–3125.
Mi, L., Zhao, H., Nash, C., Jin, X., Gao, J., Sun, C., Schmid, C., Shavit, N., Chai, Y., Anguelov, D., 2021a. HDMapGen: A hierarchical graph generative model of high definition maps, arXiv:2106.14880.
Mi, X., Yang, B., Dong, Z., Liu, C., Zong, Z., Yuan, Z., 2021. A two-stage approach for road marking extraction and modeling using mls point clouds. ISPRS J. Photogramm. Remote Sens. 180, 255–268.
Miao, Z., Wang, B., Shi, W., Zhang, H., 2014. A semi-automatic method for road centerline extraction from vhr images. IEEE Geosci. Remote Sens. Lett. 11 (11), 1856–1860.
Miao, Z., Gao, L., He, Y., Wu, L., Shi, W., Samat, A., Liu, S., Li, J., 2019. Use of colour transformation and the geodesic method for road centreline extraction from vhr satellite images. Int. J. Remote Sens. 40 (10), 4043–4058.
Miura, J., Kanda, T., Shirai, Y., 2000. An active vision system for real-time traffic sign recognition. in: ITSC2000. 2000 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.00TH8493). pp. 52–57.
Mnih, V., 2014. Machine learning for aerial image labeling. Ph.D. thesis, Ottawa.
Mongus, D., Alik, B., 2015. An efficient approach to 3D single tree-crown delineation in lidar data. ISPRS J. Photogramm. Remote Sens. 108, 219–233.
Mori, Y., Kohira, K., Masuda, H., 2018. Classification of pole-like objects using point clouds and images captured by mobile mapping systems. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2, 731–738.
Movaghati, S., Moghaddamjoo, A., Tavakoli, A., 2010. Road extraction from satellite images using particle filtering and extended kalman filtering. IEEE Trans. Geosci. Remote Sens. 48 (7), 2807–2817.
Neuhold, G., Ollmann, T., Bulo, S.R., Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. IEEE, Venice, pp. 5000–5009.
Neven, D., Brabandere, B.D., Georgoulis, S., Proesmans, M., Gool, L.V., 2018. Towards end-to-end lane detection: an instance segmentation approach. IEEE, Changshu, pp. 286–291.
Ning, X., Zhang, X., Wang, Y., Jaeger, M., 2009. Segmentation of architecture shape information from 3D point cloud. in: VRCAI09: Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry. pp. 127–132.
Nurunnabi, A., Belton, D., West, G., 2012. Robust segmentation in laser scanning 3D point cloud data. in: 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA). pp. 1–8.
Ordez, C., Cabo, C., Sanz-Ablanedo, E., 2017. Automatic detection and classification of pole-like objects for urban cartography using mobile laser scanning data. Sensors 17 (7). https://doi.org/10.3390/s17071465.
Pan, X., Shi, J., Luo, P., Wang, X., Tang, X., 2018. Spatial as deep: Spatial cnn for traffic scene understanding. Vol. 32(1). pp. 7276–7283.
Paparoditis, N., pierre Papelard, J., Devaux, R., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology. pp. 69–80.
Park, J.-G., Kim, K.-J., 2013. Design of a visual perception model with edge-adaptive gabor filter and support vector machine for traffic sign detection. Expert Syst. Appl. 40 (9), 3679–3687.
Pham, Q.-H., Uy, M. A., Hua, B.-S., Nguyen, D. T., Roig, G., Yeung, S.-K., 2020. LCD: learned cross-domain descriptors for 2D-3D matching. In: 2020 AAAI Conference on Articial Intelligence. IEEE, pp. 11856–11864.
Potsdam, I., 2018. 2d semantic labeling dataset, 1–17.
Poz, A.P.D., Gallis, R.A.B., da Silva, J.F.C., Martins, r.F.O., 2012. Object-space road extraction in rural areas using stereoscopic aerial images. IEEE Geosci. Remote Sens. Lett. 9 (4), 654–658.
Qi, C. R., Su, H., Kaichun, M., Guibas, L. J., 2017a. Pointnet: Deep learning on point sets for 3D classification and segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 77–85.
Qi, C. R., Yi, L., Su, H., Guibas, L. J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems. Vol. 30.
Qi, C. R., Liu, W., Wu, C., Su, H., Guibas, L. J., 2018. Frustum pointnets for 3D object detection from rgb-d data. in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 918–927.
Qi, C. R., Litany, O., He, K., Guibas, L., 2019. Deep hough voting for 3d object detection in point clouds. in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9276–9285.
Qi, C. R., Chen, X., Litany, O., Guibas, L. J., 2020. Imvotenet: Boosting 3d object detection in point clouds with image votes. in: 2020 IEEE/CVF 1239 Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4403–4412.
Qian, R., Garg, D., Wang, Y., You, Y., Belongie, S., Hariharan, B., Campbell, M., Weinberger, K. Q., Chao, W.-L., 2020. End-to-End pseudo-lidar for image-based 3D object detection. in: 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 5881–5890.
Qin, Z., Wang, J., Lu, Y., 2019. Triangulation learning network: from monocular to stereo 3D object detection. in: 2019 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 7607–7615.
Qin, B., Liu, W., Shen, X., Chong, Z.J., Bandyopadhyay, T., Ang, M.H., Frazzoli, E., Rus, D., 2013. A general framework for road marking detection and analysis. IEEE 619–625.
Rastiveis, H., Shams, A., Sarasua, W.A., Li, J., 2020. Automated extraction of lane markings from mobile lidar point clouds based on fuzzy inference. ISPRS J. Photogramm. Remote Sens. 160, 149–166.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 779–788.
Reitberger, J., Schnrr, C., Krzystek, P., Stilla, U., 2009. 3D segmentation of single trees exploiting full waveform lidar data. ISPRS J. Photogramm. Remote Sens. 64 (6), 561–574.
Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149.
Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., 2019. Generalized intersection over union: A metric and a loss for bounding box regression. in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 658–666.
Rodrguez-Cuenca, B., Garca-Corts, S., Ordez, C., Alonso, M.C., 2015. Automatic detection and classification of pole-like objects in urban point cloud data using an anomaly detection algorithm. Remote Sens. (Basel) 7 (10), 12680–12703.
Ruta, A., Li, Y., Liu, X., 2010. Real-time traffic sign recognition from video by class-specific discriminative features. Pattern Recogn. 43 (1), 416–430.
Safaie, A.H., Rastiveis, H., Shams, A., Sarasua, W.A., Li, J., 2021. Automated street tree inventory using mobile lidar point clouds based on hough transform and active contours. ISPRS J. Photogramm. Remote Sens. 174, 19–34.
Salti, S., Petrelli, A., Tombari, F., Fioraio, N., Di Stefano, L., 2015. Traffic sign detection via interest region extraction. Pattern Recogn. 48 (4), 1039–1049.
Scharwchter, T., Schuler, M., Franke, U., 2014. Visual guard rail detection for advanced highway assistance systems. in: 2014 IEEE Intelligent Vehicles Symposium Proceedings. pp. 900–905.
Schmitt, M., Brueck, A., Schoenberger, J., Stilla, U., 2013. Potential of airborne single-pass millimeterwave insar data for individual tree recognition. 33. Wissenschaftlich-Technische Jahrestagung der DGPF. No. 22.
Seibert, A., Hhnel, M., Tewes, A., Rojas, R., 2013. Camera based detection and classification of soft shoulders, curbs and guardrails. In: 2013 IEEE Intelligent Vehicles Symposium (IV). pp. 853–858.
Selver, M.A., Er, E., Belenlioglu, B., Soyaslan, Y., 2016. Camera based driver support system for rail extraction using 2-d gabor wavelet decompositions and morphological analysis. IEEE, Birmingham, United Kingdom, pp. 270–275.
Shahzad, M., Schmitt, M., Zhu, X. X., 2015. Segmentation and crown parameter extraction of individual trees in an airborne tomosar point cloud. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS) XL-3-W2, 205–209.
Shao, Z., Zhou, Z., Huang, X., Zhang, Y., 2021. Mrenet: Simultaneous extraction of road surface and road centerline in complex urban scenes from very high-resolution images. Remote Sens. (Basel) 13 (2), 239.
Shi, W., Rajkumar, R., 2020. Point-gnn: Graph neural network for 3d object detection in a point cloud. in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1708–1716.
Shi, S., Wang, X., Li, H., 2019. Pointrcnn: 3D object proposal generation and detection from point cloud. in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–779.
Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., Li, H., 2020. PV-RCNN: Point-voxel feature set abstraction for 3D object detection. in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10526–10535.

Shi, S., Wang, Z., Shi, J., Wang, X., Li, H., 2021. From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network. IEEE Trans. Pattern Anal. Mach. Intell. 43 (8), 2647–2664.
Shin, K., Kwon, Y. P., Tomizuka, M., 2019. RoarNet: A robust 3D object detection based on region approximation refinement. in: 2019 IEEE Intelligent Vehicles Symposium (IV). pp. 2510–2515.
Shin, B.-S., Tao, J., Klette, R., 2015. A superparticle filter for lane detection. Pattern Recogn. 48 (11), 3333–3345.
Shirke, S., Udayakumar, R., 2019. Lane datasets for lane detection. IEEE, Chennai, India, pp. 0792–0796.
Simonyan, K., Zisserman, A., Ma, 2015. Very deep convolutional networks for large-scale image recognition. in: International Conference on Learning Representations. pp. 1–14.
Sindagi, V. A., Zhou, Y., Tuzel, O., 2019. Mvx-net: Multimodal voxelnet for 3D object detection. in: 2019 International Conference on Robotics and Automation (ICRA). pp. 7276–7282.
Soheilian, B., Paparoditis, N., Boldo, D., 2010. 3d road marking reconstruction from street-level calibrated stereo pairs. ISPRS J. Photogramm. Remote Sens. 65 (4), 347–359.
Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C., 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. Neural Netw. 32, 323–332.
Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5686–5696.
Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., Vasudevan, V., Han, W., Ngiam, J., Zhao, H., Timofeev, A., Ettinger, S., Krivokon, M., Gao, A., Joshi, A., Zhang, Y., Shlens, J., Chen, Z., Anguelov, D., 2020. Scalability in perception for autonomous driving: Waymo open dataset. in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2443–2451.
Tabernik, D., Skoaj, D., 2020. Deep learning for large-scale traffic-sign detection and recognition. IEEE Trans. Intell. Transp. Syst. 21 (4), 1427–1440.
Tao, C., Qi, J., Li, Y., Wang, H., Li, H., 2019. Spatial information inference net: Road extraction using road-specific contextual information. ISPRS J. Photogramm. Remote Sens. 158, 155–166.
Tao, S., Wu, F., Guo, Q., Wang, Y., Li, W., Xue, B., Hu, X., Li, P., Tian, D., Li, C., Yao, H., Li, Y., Xu, G., Fang, J., 2015. Segmenting tree crowns from terrestrial and mobile lidar data by exploring ecological theories. ISPRS J. Photogramm. Remote Sens. 110, 66–76.
Timofte, R., Gool, L. V., 2011. Sparse representation based projections. in: Proc. BMVC. pp. 61.1–61.12.
Timofte, R., Zimmermann, K., Van Gool, L., 2009. Multi-view traffic sign detection, recognition, and 3D localisation. in: 2009 Workshop on Applications of Computer Vision (WACV). pp. 1–8.
Tombari, F., Fioraio, N., Cavallari, T., Salti, S., Petrelli, A., Di Stefano, L., 2014. Automatic detection of pole-like structures in 3D urban environments. in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 4922–4929.
Tu, J., Yao, J., Li, L., Zhao, W., Xiang, B., 2021. Extraction of street pole-like objects based on plane filtering from mobile lidar data. IEEE Trans. Geosci. Remote Sens. 59 (1), 749–768.
Tvri, D., Pfeifer, N., 2005. Segmentation based robust interpolation- a new approach to laser data filtering. in: ISPRS WG III/3, III/4, V/3 Workshop 1324 Laser scanning 2005. pp. 12–14.
Vallet, B., Brdif, M., Serna, A., Marcotegui, B., Paparoditis, N., 2015. Terramobilita iqmulus urban point cloud analysis benchmark. Comput. Graph. 49, 126–133.
Van Etten, A., Lindenbaum, D., Bacastow, T. M., 2019. Spacenet: A remote sensing dataset and challenge series. arXiv:1807.01232 [cs].
Vega, C., Hamrouni, A., El Mokhtari, S., Morel, J., Bock, J., Renaud, J.-P., Bouvier, M., Durrieu, S., 2014. Ptrees: A point-based approach to forest tree extraction from lidar data. Int. J. Appl. Earth Obs. Geoinf. 33, 98–108.
Vieira, M., Shimada, K., 2005. Surface mesh segmentation and smooth surface extraction through region growing. Comput. Aided Geom. Des. 22 (8), 771–792.
Vitabile, S., Pollaccia, G., Pilato, G., Sorbello, F., 2001. Road signs recognition using a dynamic pixel aggregation technique in the hsv color space. in: Proceedings 11th International Conference on Image Analysis and Processing. pp. 572–577.
Vivacqua, R.P.D., Bertozzi, M., Cerri, P., Martins, F.N., Vassallo, R.F., 2018. Self-localization based on visual lane marking maps: An accurate low-cost approach for autonomous driving. IEEE Trans. Intell. Transp. Syst. 19 (2), 582–597.
Vo, A.-V., Truong-Hong, L., Laefer, D.F., Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. ISPRS J. Photogramm. Remote Sens. 104, 88–100.
Vora, S., Lang, A. H., Helou, B., Beijbom, O., 2020. Pointpainting: Sequential fusion for 3D object detection. in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4603–4611.
Vosselman, G., 2009. Advanced point cloud processing. Vol. 9. University of Stuttgart Stuttgart, Germany, pp. 137–146.
Wang, Z., Jia, K., 2019. Frustum ConvNet: Sliding frustums to aggregate local point-wise features for amodal 3D object detection. in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1742–1749.
Wang, Y., Chao, W., Garg, D., Hariharan, B., Campbell, M., Weinberger, K. Q., 2019. Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving. in: 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 8445–8453.
Wang, D., Hou, X., Xu, J., Yue, S., Liu, C.-L., 2017. Traffic sign detection using a cascade method with fast feature extraction and saliency test. IEEE Trans. Intell. Transp. Syst. 18 (12), 3290–3302.
Wang, G., Ren, G., Jiang, L., Quan, T., 2014a. Hole-based traffic sign detection method for traffic signs with red rim. Vis. Comput. 30 (5), 539–551.
Wang, J., Song, J., Chen, M., Yang, Z., 2015. Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens. 36 (12), 3144–3169, 6.
Wang, M., Tseng, Y.-H., 2011. Incremental segmentation of lidar point clouds with an octree-structured voxel space. Photogram. Rec. 26 (133), 32–57.
Wang, H., Wang, C., Luo, H., Li, P., Cheng, M., Wen, C., Li, J., 2014b. Object detection in terrestrial laser scanning point clouds based on hough forest. IEEE Geosci. Remote Sens. Lett. 11 (10), 1807–1811.
Wegner, J.D., Montoya-Zegarra, J.A., Schindler, K., 2013. A higher-order crf model for road network extraction. IEEE, Portland, OR, USA, pp. 1698–1705.
Wei, Y., Zhang, K., Ji, S., 2020. Simultaneous road surface and centerline extraction from large-scale remote sensing images using cnn-based segmentation and tracing. IEEE Trans. Geosci. Remote Sens. 58 (12), 8919–8931.
Weinmann, M., Weinmann, M., Mallet, C., Brdif, M., 2017. A classification-segmentation framework for the detection of individual trees in dense mms point cloud data acquired in urban areas. Remote Sens. (Basel) 9 (3).
Wen, C., Li, J., Luo, H., Yu, Y., Cai, Z., Wang, H., Wang, C., 2016. Spatial-related traffic sign inspection for inventory purposes using mobile laser scanning data. IEEE Trans. Intell. Transp. Syst. 17 (1), 27–37.
Wen, C., Sun, X., Li, J., Wang, C., Guo, Y., Habib, A., 2019. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. ISPRS J. Photogramm. Remote Sens. 147, 178–192.
Woo, S., Hong, D., Lee, W.-C., Chung, J.-H., Kim, T.-H., 2008. A robotic system for road lane painting. Autom. Constr. 17 (2), 122–129.
Woo, H., Kang, E., Wang, S., Lee, K.H., 2002. A new segmentation method for point cloud data. Int. J. Mach. Tool Manu. 42 (2), 167–178.
Wu, T., Ranganathan, A., 2012. A practical system for road marking detection and recognition. pp. 25–30.
Wu, B., Yu, B., Yue, W., Shu, S., Tan, W., Hu, C., Huang, Y., Wu, J., Liu, H., 2013. A voxel-based method for automated identification and morphological parameters estimation of individual street trees from mobile laser scanning data. Remote Sens. (Basel) 5 (2), 584–611.
Xiaozhu, X., Cheng, H., 2017. Object detection of armored vehicles based on deep learning in battlefield environment. in: 2017 4th International Conference on Information Science and Control Engineering (ICISCE). pp. 1568–1570.
Xie, D., Zhu, W., Rong, F., Xia, X., Shang, H., 2019. Registration of point clouds: A survey. in: 2021 International Conference on Networking Systems of AI (INSAI). IEEE, pp. 136–142.
Xie, Q., Li, D., Yu, Z., Zhou, J., Wang, J., 2020. Detecting trees in street images via deep learning with attention module. IEEE Trans. Instrum. Meas. 69 (8), 5395–5406.
Xu, S., 2009. Robust traffic sign shape recognition using geometric matching. IET Intelligent Transport Systemsy 3 (1), 10–18.
Xu, D., Anguelov, D., Jain, A., 2018a. Pointfusion: Deep sensor fusion for 3D bounding box estimation. in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 244–253.
Xu, S., Xu, S., Ye, N., Zhu, F., 2018b. Automatic extraction of street trees' nonphotosynthetic components from mls data. Int. J. Appl. Earth Obs. Geoinf. 69, 64–77.
Xu, S., Ye, N., Xu, S., Zhu, F., 2018c. A supervoxel approach to the segmentation of individual trees from lidar point clouds. Remote Sens. Lett. 9 (6), 515–523.
Yadav, M., Khan, P., Singh, A. K., Lohani, B., 2018. Generating GIS database of street trees using mobile lidar data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-5, 233–237.
Yadav, M., Lohani, B., Singh, A.K., Husain, A., 2016. Identification of pole-like structures from mobile lidar data of complex road environment. Int. J. Remote Sens. 37 (20), 4748–4777.
Yakimov, P., 2016. Traffic signs detection using tracking with prediction. In: E-Business and Telecommunications. Cham, pp. 454–467.
Yakimov, P., Fursov, V., 2015. Traffic signs detection and tracking using modified hough transform. in: 2015 12th International Joint Conference on e-Business and Telecommunications (ICETE). Vol. 05. pp. 22–28.
Yan, Y., Mao, Y., Li, B., 2018. Second: Sparsely embedded convolutional detection. Sensors 18 (10), 3337.1–3337.17.
Yan, L., Li, Z., Liu, H., Tan, J., Zhao, S., Chen, C., 2017. Detection and classification of pole-like road objects from mobile lidar data in motorway environment. Opt. Laser Technol. 97, 272–283.
Yan, W.Y., Morsy, S., Shaker, A., Tulloch, M., 2016. Automatic extraction of highway light poles and towers from mobile lidar data. Opt. Laser Technol. 77, 162–168.
Yang, B., Dong, Z., 2013. A shape-based segmentation method for mobile laser scanning point clouds. ISPRS J. Photogramm. Remote Sens. 81, 19–30.
Yang, B., Dong, Z., Zhao, G., Dai, W., 2015. Hierarchical extraction of urban objects from mobile laser scanning data. ISPRS J. Photogramm. Remote Sens. 99, 45–57.
Yang, B., Liang, M., Urtasun, R., 2018. Hdnet: Exploiting hd maps for 3d object detection. in: Proceedings of The 2nd Conference on Robot Learning. Vol. 87 of Proceedings of Machine Learning Research. pp. 146–155.
Yang, B., Luo, W., Urtasun, R., 2018b. Pixor: Real-time 3d object detection from point clouds. in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7652–7660.
Yang, Z., Sun, Y., Liu, S., Shen, X., Jia, J., 2018c. IPOD: Intensive point-based object detector for point cloud. arXiv preprint, arXiv: 1812.05276.

Yang, Z., Sun, Y., Liu, S., Shen, X., Jia, J., 2019b. STD: Sparse-to-dense 3d object detector for point cloud. in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1951–1960.
Yang, Z., Sun, Y., Liu, S., Jia, J., 2020. 3DSSD: Point-based 3D single stage object detector. in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11037–11045.
Yang, J., Kang, Z., Akwensi, P.H., 2019a. A skeleton-based hierarchical method for detecting 3-D pole-like objects from mobile lidar point clouds. IEEE Geosci. Remote Sens. Lett. 16 (5), 801–805.
Yang, B., Wei, Z., Li, Q., Li, J., 2013. Semiautomated building facade footprint extraction from mobile lidar point clouds. IEEE Geosci. Remote Sens. Lett. 10 (4), 766–770.
Yao, W., Krzystek, P., Heurich, M., 2013. Enhanced detection of 3d individual trees in forested areas using airborne full-waveform lidar data by combining normalized cuts with spatial density clustering. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-5-W2, 349–354.
Ye, Y. Y., Chen, H. J., Hao, X. L., 2017. Lane marking detection based on waveform analysis and cnn. Singapore, Singapore, pp. 1044316.1–1044316.5.
Ye, C., Li, J., Jiang, H., Zhao, H., Ma, L., Chapman, M., 2020. Semi-automated generation of road transition lines using mobile laser scanning data. IEEE Trans. Intell. Transp. Syst. 21 (5), 1877–1890.
Yenikaya, S., Yenikaya, G., Dven, E., 2013. Keeping the vehicle on the road: A survey on on-road lane detection systems. ACM Comput. Surv. 46 (1), 1–43.
Ying, Z., Li, G., 2016. Robust lane marking detection using boundary-based inverse perspective mapping. IEEE, Shanghai, pp. 1921–1925.
You, C., Wen, C., Wang, C., Li, J., Habib, A., 2019. Joint 2D and 3D traffic sign landmark data set for geo-localization using mobile laser scanning data. IEEE Trans. Intell. Transp. Syst. 20 (7), 2550–2565.
Yu, H., Zhen, W., Yang, W., Zhang, J., Scherer, S., 2020. Monocular camera localization in prior lidar maps with 2D-3D line correspondences. in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 4588–4594.
Yu, Y., Li, J., Guan, H., Jia, F., Wang, C., 2015a. Learning hierarchical features for automated extraction of road markings from 3-d mobile lidar point clouds. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (2), 709–726.
Yu, Y., Li, J., Guan, H., Wang, C., Yu, J., 2015b. Semiautomated extraction of street light poles from mobile lidar point-clouds. IEEE Trans. Geosci. Remote Sens. 53 (3), 1374–1386.
Yu, Y., Li, J., Wen, C., Guan, H., Luo, H., Wang, C., 2016. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data. ISPRS J. Photogramm. Remote Sens. 113, 106–123.
Yuan, X., Hao, X., Chen, H., Wei, X., 2014. Robust traffic sign recognition based on color global and local oriented edge magnitude patterns. IEEE Trans. Intell. Transp. Syst. 15 (4), 1466–1477.
Yuan, X., Guo, J., Hao, X., Chen, H., 2015. Traffic sign detection via graph-based ranking and segmentation algorithms. IEEE Trans. Syst., Man, Cybernetics: Systems 45 (12), 1509–1521.
Yue, G., Liu, R., Zhang, H., Zhou, M., 2015. A method for extracting street trees from mobile lidar point clouds. Open Cybernet. Syst. J. 9 (1), 204–209.
Zaklouta, F., Stanciulescu, B., 2014. Real-time traffic sign recognition in three stages. Rob. Auton. Syst. 62 (1), 16–24.
Zarzar, J., Giancola, S., Ghanem, B., 2019. PointRGCN: Graph convolution networks for 3d vehicles detection refinement. arXiv preprint, arXiv:1911.12236.
Zeng, Y., Hu, Y., Liu, S., Ye, J., Han, Y., Li, X., Sun, N., 2018. RT3D: Real-time 3-d vehicle detection in lidar point cloud for autonomous driving. IEEE Rob. Autom. Lett. 3 (4), 3434–3440.
Zhang, L., Chu, R., Xiang, S., Liao, S., Li, S. Z., 2007. Face detection based on multi-block lbp representation. In: Advances in Biometrics. Berlin, Heidelberg, pp. 11–18.
Zhang, J., Lin, X., Ning, X., 2013. Svm-based classification of segmented airborne lidar point clouds in urban areas. Remote Sensing 5 (8), 1451 3749–3775.
Zhang, D., Xu, X., Lin, H., Gui, R., Cao, M., He, L., 2019a. Automatic road-marking detection and measurement from laser-scanning 3d profile data. ISPRS Journal of Photogrammetry and Remote Sensing 108, 102957.1–102957.14.
Zhang, X., Han, X., Li, C., Tang, X., Zhou, H., Jiao, L., 2019b. Aerial image road extraction based on an improved generative adversarial network. Remote Sensing 11 (8), 930.1–930.19.
Zhang, Y., Xiong, Z., Zang, Y., Wang, C., Li, J., Li, X., 2019c. Topology-aware road network extraction via multi-supervised generative adversarial networks. Remote Sensing 11 (9), 1017.1–1017.19.
Zhang, J., Lin, X., 2013. Semi-automatic extraction of straight roads from very high resolution remotely sensed imagery by a fusion method. Sens. Lett. 11 (6), 1229–1235.
Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual u-net. IEEE Geosci. Remote Sens. Lett. 15 (5), 749–753.
Zhao, X., Liu, Z., Hu, R., Huang, K., 2019. 3D object detection using scale invariant and feature reweighting networks. Proceedings of the AAAI Conference on Artificial Intelligence 33 (01), 9267–9274.
Zheng, H., Tan, F., Wang, R., 2016. Pole-like object extraction from mobile lidar data. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B1, 729–734.
Zhong, L., Cheng, L., Xu, H., Wu, Y., Chen, Y., Li, M., 2017. Segmentation of individual trees from TLS and MLS data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (2), 774–787.
Zhong, R., Wei, J., Su, W., Chen, Y.F., 2013. A method for extracting trees from vehicle-borne laser scanning data. Math. Comput. Model. 58 (3), 733–742.
Zhou, Y., Tuzel, O., 2018. VoxelNet: End-to-end learning for point cloud based 3d object detection. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4490–4499.
Zhou, J., Bischof, W.F., Caelli, T., 2006. Road tracking in aerial images based on human computer interaction and bayesian filtering. ISPRS J. Photogramm. Remote Sens. 61 (2), 108–124.
Zhou, S., Jiang, Y., Xi, J., Gong, J., Xiong, G., Chen, H., 2010. A novel lane detection based on geometrical model and gabor filter. IEEE 59–64.
Zhou, M., Sui, H., Chen, S., Wang, J., Chen, X., 2020. Bt-roadnet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 168, 288–306.
Zhou, L., Zhang, C., Wu, M., 2018. D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. IEEE, Salt Lake City, UT, USA, pp. 192–1924.
Zhu, H., Guo, B., 2018. A beam guardrail detection algorithm using lidar for intelligent vehicle. in: 2018 IEEE 8th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). pp. 1398–1402.
Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., Hu, S., 2016. Traffic-sign detection and classification in the wild. in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2110–2118.