Chapter 11
Road Environment Perception for Safe and Comfortable Driving

Sicen Guo, Yu Jiang, Jiahang Li, Dacheng Zhou, Shuai Su, Mohammud Junaid Bocus, Xingyi Zhu, Qijun Chen, and Rui Fan

Abstract With the ongoing evolution of autonomous driving technology, road environment perception systems have become a significant focus of research. However, there is currently a paucity of comprehensive survey articles that provide a systematic overview of state-of-the-art (SoTA) computer vision techniques and vibration methods for road defect detection, particularly with regards to deep learning methods. This chapter aims to fill this gap by describing the sensing technologies for vision-based

S. Guo · J. Li · D. Zhou · S. Su · Q. Chen · R. Fan (✉)
The Robotics and Artificial Intelligence Laboratory (RAIL), The Department of Control Science and Engineering, Frontiers Science Center for Intelligent Autonomous Systems, and State Key Laboratory of Intelligent Autonomous Systems, Tongji University, Shanghai 201804, People's Republic of China
e-mail: rfan@tongji.edu.cn

S. Guo
e-mail: guosicen@tongji.edu.cn

J. Li
e-mail: 2230745@tongji.edu.cn

D. Zhou
e-mail: zhoudacheng20@tongji.edu.cn

S. Su
e-mail: sushuai@tongji.edu.cn

Q. Chen
e-mail: qjchen@tongji.edu.cn

Y. Jiang
CTO Office, ClearMotion, Inc. Billerica, MA 01821, USA
e-mail: yjiang@clearmotion.com

M. J. Bocus
University of Bristol, Bristol, United Kingdom
e-mail: junaid.bocus@bristol.ac.uk

X. Zhu
The Department of Roads and Airport Engineering and Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 200092, People's Republic of China
e-mail: zhuxingyi66@tongji.edu.cn

R. Fan et al. (eds.), *Autonomous Driving Perception*, Advances in Computer Vision and Pattern Recognition, https://doi.org/10.1007/978-981-99-4287-9_11

and vibration-based road environment data acquisition, summarizing several public datasets for pothole and crack detection, and providing a comprehensive review of SoTA road defect detection algorithms. Additionally, this chapter also discusses the core competencies of autonomous vehicle software systems, such as planning and control. Finally, we offer a glimpse into the future of autonomous driving systems, envisioning the integration of both vision and motion sensors. We believe that this survey will act as a helpful guide for advancing road defect detection technology, providing strategic advice and practical guidance to those involved in developing such systems.

11.1 Introduction

Over the past few decades, the global vehicle population has significantly increased, resulting in greater pressure on road surfaces and a corresponding rise in road defects. These harsh road conditions not only damage vehicles but also impede smooth driving and contribute to traffic accidents. In fact, one-third of the approximately 33,000 yearly crashes are attributed to poor road conditions [1]. A study by The Pacific Institute for Research and Evaluation, which analyzed crash data from various government agencies over 18 months, found that road issues such as potholes and icy conditions on highways are responsible for over 42,000 fatalities annually [2].

Road defects such as potholes can significantly increase the risk of accidents and vehicle repair costs, particularly in poor or substandard conditions. The impact of hitting a pothole can cause damage to the vehicle's suspension and can also result in the driver losing control of the vehicle. When a vehicle's tyres encounter a pothole, the forces acting on the tyres become imbalanced, leading to weight shifting to the lower tyres and causing the vehicle to tilt. As the tyre strikes the pothole's edge, a concentrated force impacts the tyre directly, which can result in tyre deformation, breakage, and even rim bending. These impacts not only cause damage to the vehicle, but also interfere with the driving experience, making it difficult to drive the vehicle in a straight line.

Autonomous driving, also known as robot-assisted driving, has garnered significant research attention in recent years. The utilization of autonomous vehicles is projected to have a crucial impact on the future of urban transportation systems by enhancing safety, productivity, accessibility, road efficiency, and environmental sustainability. The pursuit of fully autonomous vehicles has instigated intense competition among leading companies, including Google, Toyota, and Ford, to advance their respective autonomous robotic vehicle concepts [4–6]. Autonomous ground vehicles have been recognized for their prospects to provide crucial benefits to society, such as reducing the frequency of road accidents caused by human factors, optimizing fuel consumption, and enhancing driving convenience.

Despite the significant milestone that autonomous vehicles represent in the advancement of intelligent automation, concerns about their safety and reliability continue to be major issues. Autonomous vehicles require robust detection of their surroundings and interpretation of a wide range of contextual information to avoid collisions, particularly from a safety perspective. Road environment perception capabilities play a crucial role in ensuring the safe deployment of autonomous vehicles. For example, ClearMotion [7] has developed a cloud software called RoadMotion. The collection and analysis of road surface mapping data by autonomous vehicles enable the vehicle subsystems to make informed decisions in real-time. This data facilitates safer control of steering, stopping, starting, and absorbing shocks based on the actual road conditions. This requires the vehicle sensors to recognize various road defects, as well as road signs such as lane lines, speed humps, and pedestrian crossing. In the contemporary realm of autonomous driving, the detection and localization of roadway defects have garnered paramount importance in ensuring safe and efficient transportation. To this end, sophisticated algorithms have been developed, which predominantly employ a global mapping approach, harnessing the power of Global Positioning Systems (GPSs), Light Detection and Ranging (LIDARs), and cameras. Such algorithms meticulously detect and precisely localize a given defect within the vehicle's coordinate system, subsequently transforming this information into global coordinates. Leveraging this accurate spatial information, the algorithm scrutinizes the global map and identifies the optimal path for the vehicle, guaranteeing smooth navigation on the road [8, 9].

In the last two decades, the field of road perception has witnessed a remarkable transformation, marked by the advent of advanced acquisition and detection techniques. Yet, most recent surveys in this field do not comprehensively cover the SoTA advancements in computer vision, including 3-D point cloud modeling, segmentation, and machine/deep learning. Furthermore, the relatively nascent field of vibration-based road detection remains largely unexplored. In light of this, we present a meticulous and comprehensive review of the SoTA road perception systems, discussing both computer vision-based and vibration-based road defect detection algorithms. To facilitate a comprehensive understanding of this domain, we begin by providing a systematic overview of available systems and methods, as illustrated in Fig. 11.1. We begin by providing a comprehensive introduction to vision-based road imaging methods (see Sect. 11.2.1), followed by an introduction of vibration-based road environment acquisition techniques (see Sect. 11.2.2). We then introduce the road defect datasets obtained through the aforementioned methodologies (see Sect. 11.3). To ensure a clear and coherent presentation, we categorize road defect detection algorithms into two main groups: vision-based (see Sect. 11.4.1) and vibration-based (see Sect. 11.4.2). Subsequently, we describe how road information can be utilized in autonomous driving systems for vehicle planning and control (see Sect. 11.5). Finally, we conduct a critical evaluation of the limitations of existing methods and explore future research directions in this fast-evolving field (see Sect. 11.6). Our comprehensive survey promises to equip researchers and practitioners alike with a nuanced and in-depth understanding of the latest developments in road perception systems.

11.2 Sensing Technologies for Road Environment Perception

11.2.1 Vision Sensors

11.2.1.1 2-D Imaging

The collection of road data through 2-D imaging methods, such as digital imaging or digital image acquisition, was first initiated in the early 1990s [10, 11]. However, the limitations of 2-D images make it difficult to effectively depict the physical structure of roads [12]. During that time, cameras and range sensors were commonly used as the primary sensing devices for acquiring road information. Nonetheless, image segmentation algorithms that are applied to gray-scale/RGB road defect images can

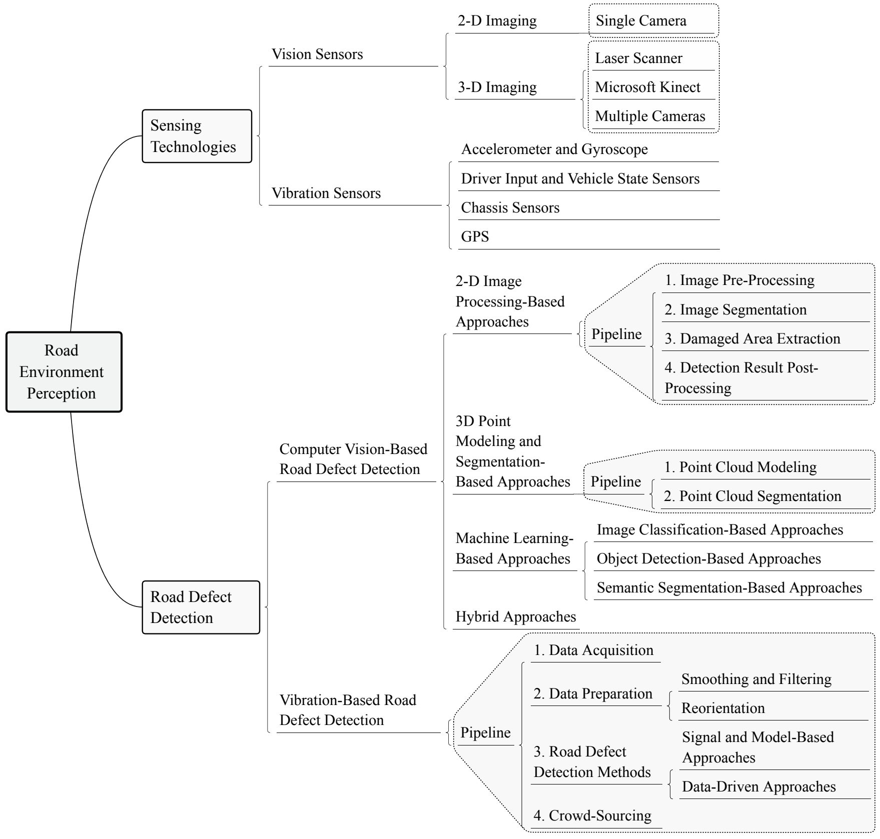

Fig. 11.1 A systematic overview of road environment perception systems created based on our previous taxonomy introduced in [3]

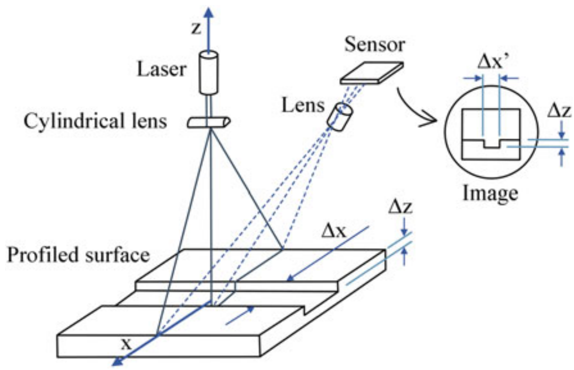

Fig. 11.2 Laser triangulation [19]. By placing the sensor at a predetermined distance from the laser's light source, the reflection angle of the laser illumination can be calculated, enabling the derivation of precise 3-D information about the road surface

be easily influenced by a wide range of factors, including unfavorable illumination conditions [13]. To overcome these limitations and simultaneously enhance the accuracy of road defect detection, researchers have increasingly turned to 3-D imaging technologies [12, 14, 15].

11.2.1.2 3-D Imaging

The first successful use of 3-D imaging methods to collect road surface data was in 1997 [16]. This section aims to provide a comprehensive discussion of the contemporary 3-D imaging technologies that are currently being employed for the purpose of acquiring road data.

Laser scanning is a widely adopted imaging method for precise 3-D road surface data acquisition [17]. Nonetheless, it requires laser sensors to be mounted on dedicated detection vehicles like the Georgia Institute of Technology Sensing Vehicle [18]. As a result, this method is not widely used due to the high cost of equipment acquisition and long-term maintenance expenses [15] (Fig. 11.2).

Microsoft Kinect [13] was originally developed for motion sensing games on the Xbox-360 platform, but have since been utilized for a variety of other applications. These sensors are equipped with various components such as an infrared (IR) sensor/camera, an IR emitter, an RGB camera, accelerometers, a tilt motor for motion tracking capabilities, and microphones [17]. Microsoft Kinect is a ideal technology for road surface imaging when installed on a vehicle, with a working range of 800–4000 mm. Previous studies have used Kinect sensors for 3-D road data acquisition and defect detection [13, 20, 21]. However, the reliability of Kinect sensors can be challenged by the IR saturation issue, particularly when exposed to direct sunlight [22] (Fig. 11.3).

Multiple images captured from different views can also be used to reconstruct the 3-D geometry of a road surface [12, 24]. This technique is based on the theory of *multi-view geometry* [25]. The initial proposal for an automatic 3-D reconstruction and modeling system for road signs was introduced in [26]. The authors employed

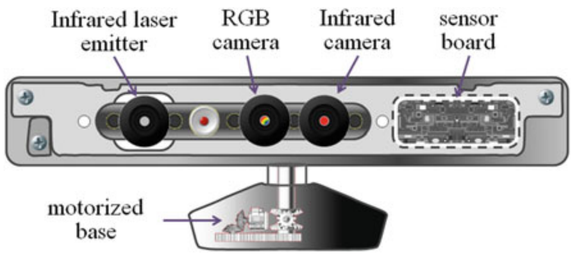

Fig. 11.3 Microsoft Kinect sensor [23]

a multi-view constrained 3-D reconstruction approach that incorporates geometric constraints on the shape of the road signs to obtain an optimal 3-D silhouette, achieving centimeter-level accuracy.

11.2.2 Vibration Sensors

Vibration-based road defect detection methods utilize non-vision-based sensors, such as inertial sensors (e.g., accelerometer and gyroscope sensors), vehicle speed, vehicle suspension states, drivers' inputs, etc. Vibration-based road defect detection methods have gained popularity due to their cost-effectiveness and relatively light computational burden. Despite being tied to vehicle dynamics and potentially affected by vehicle parameters, these methods provide more robust measurements as they are less influenced by weather or lighting conditions.

One of the earlier attempts at using mobile inertial sensors to detect potholes was conducted by a research group in CASIL at MIT in 2008. In [28], Eriksson et al. created a system known as *pothole patrol* that used inertial sensors and GPS sensors from mobile devices deployed to a fleet of seven taxis in the Boston area. The orientation of each mobile sensor was fixed in the vehicle once installed. Therefore, a one-time calibration can be performed for each vehicle right after sensor installation. In [29], researchers at Microsoft Research India presented *Nericell*, an add-on system that can be carried together with smartphones by users in the normal course. Since the orientation of the mobile sensors may vary when placed inside a vehicle, a practical reorientation approach was developed to virtually align the smartphone's three axes with body frame of the car. In 2011, Perttunen et al. collected accelerometer data at 38Hz and GPS data at 1Hz from a Nokia N95 8GB mobile phone, and they were able to detect road defects by training and deploying an SVM detector [30]. Later, with the increasing programmability and sensor accessibility of smartphones, researchers started developing applications directly on smartphones to detect potholes and other types of road defects using built-in smartphone sensors (see Table 11.1 for a list of possible sensors or signals that can be used for road defect detection). Android smartphones with accelerometers were used in [31] for real-time pothole detection. Wu et al. [32] developed a customized Android application and the smartphone can be mounted without being fixed to the back seat. In [33], the authors proposed a

Table 11.1 Different types of smartphone sensor data for road defect detection

Signal/sensor name	Type	Unit	Description
Acceleration	Measured	m/s^2	Raw accelerometer measurements
Gyroscope	Measured	deg/s	Raw gyroscope measurements
Linear acceleration	Derived	m/s^2	Accelerometer measurements excluding gravity
Magnetometer	Measured	μT	Ambient geomagnetic field measurements
Gravity	Derived	m/s^2	Gravitational acceleration
Rotation	Derived	rad	Orientation of the sensor
GPS	Measured	deg	Location information of the sensor

Fig. 11.4 In-vehicle sensors for vibration signals collection [27]

model called *smart-patrolling* on Android phones to crowd-source data for pothole detection. A system called *Wolverine* was introduced in [34] and was implemented on a Google Nexus S. A road monitoring system was developed using a Nokia Lumia 820 mobile device to detect road defects from measured gyroscope data around gravity rotation and accelerometer sensor data [35]. Recently, Rajput et al. placed smartphones in public transportation buses with dense connectivity to achieve large-scale road monitoring [36]. While most research studies rely on accelerometer sensing as the primary indicator for road defects, the *RoadMonitor* system [35] proposed a cross-validation methodology for the detection and severity assessment of road bumps using gyroscope sensors.

Although smartphones can be used as vibration sensors, vehicle built-in sensors generally provide richer content. In [37], researchers constructed a signal set con-

Table 11.2 Pothole dataset

Pothole dataset	Sensors	Color/gray	Dataset size	Resolution
[43]	GoPro Hero 3 + camera	Color	628	2,760 × 3,680
[44]	–	Color	3,227 (2,475 for training and 752 for testing)	720 × 1,280
[45]	Smartphone mounted on a car	Color	9,053	600 × 600
[22]	ZED stereo camera	Color	67	800 × 1,312
[46]	ZED stereo camera	Color	600 (240 for training, 180 for validation and 180 for testing)	400 × 400
UDTIRI	Collected from the Internet or taken manually	Color	1,000 (600 for training, 100 for validation and 300 for testing)	Variety of resolutions

taining eighteen different signals for road surface defect detection. The signal set includes measurements from the Inertial Measurement Unit (IMU) located at the center of gravity of the vehicle, as well as measurements from the shock absorber and spindle signals. Similarly, in [27], vehicle measurements, including suspension position reading, were collected to conduct state estimation based on Jump-Diffusion Processes (JDPs). The sensor data collection setup utilized by [27] is presented in Fig. 11.4.

11.3 Public Datasets

Data plays a crucial role in deep learning algorithms. While there are many large-scale datasets available for general-purpose object detection, the small number and size of existing datasets have become a bottleneck for the development of defect detection systems. This section provides an overview of common road imaging datasets, as well as pothole and crack datasets, which are summarized in Tables 11.2 and 11.3.

11.3.1 Road Imaging Datasets

Road scene imaging datasets are crucial for road defect detection. While datasets such as KITTI [38] and Cityscape [39], which are widely used for computer vision

Table 11.3 Crack dataset

Crack dataset	Sensors	Color/gray	Dataset size	Resolution
[48]	Nine CMOS line camera	Gray	38,000	6,144 × 1,024
[49]	Aviiva SM2 CL 1010 Camera	Color	–	1,024 × 1,714
[50]	DSLR camera (Nikon D5200)	Color	332	277 images with 4,928 × 3,264 and 55 images with 5,888 × 3,584
[51]	16 MP Nikon digital camera	Color	56,092	256 × 256,64 × 64
[52]	Collected from the Internet or taken manually	Color	200	Variety of resolutions
[53]	Digital camera	Color	1,328	4,032 × 3,016
[54]		Color	600	224 × 144,976 × 640

tasks, also include road scenes, they provide less information about the road surface and are primarily focused on street scenes. Specifically, there are datasets dedicated to road surface defect detection, such as those used for 3D reconstruction of road surfaces [12, 15].

For the road surface 3-D reconstruction, [12] creates three datasets[1] (91 stereo image pairs). Datasets 1 and 2 focus on road scenes, while dataset 3 assists researchers in evaluating with sample models. Each dataset includes both uncalibrated and calibrated left and right images. The calibrated images are serve as estimation of the disparity map, while the uncalibrated images and calibration parameters are used to reconstruct the 3-D road geometry. Additionally, they have created a road pothole 3-D geometry reconstruction dataset[2] [15]. The accuracy of the road pothole 3-D geometry reconstructed using stereo vision technology is 2.23 mm.

Toronto-3-D[3] [40] is a sizable dataset of urban outdoor point cloud collected by a Mobile Laser Scanning (MLS) system in Toronto, Canada. It contains approximately 78.3 million points and covers around 1 km of road.

The RadarScenes dataset[4] [41] consists of recordings collected from a measurement vehicle equipped with four automotive radar sensors and a documentary camera facing the front. It includes point cloud data collected by the radar sensors and seman-

[1] https://github.com/ruirangerfan/road_surface_3d_reconstruction_datasets.

[2] https://github.com/ruirangerfan/rethinking_road_reconstruction_pothole_detection.

[3] https://www.kaggle.com/datasets/priteshraj10/point-cloud-lidar-toronto-3d.

[4] https://www.kaggle.com/datasets/aleksandrdubrovin/the-radarscenes-data-set.

tic annotations on a point-wise level. The dataset was recorded in Ulm, Germany over a period of three years from 2016 to 2018 and spans over 4 hours in length.

RoadSaW [42] is a recently introduced dataset that aims to facilitate the estimation of road surface and wetness. The dataset contains 720,000 bird's-eye-view patches captured on a test track, high-resolution videos, and accurate synchronized water film height measurements. The dataset covers various surface types such as asphalt, concrete, and cobblestone. The FLIR Blackfly 3.2 MP, 30 fps camera is mounted behind the truck's windscreen at a height of 2.66 meters. The primary dataset consists of 250 videos recorded on five different days and is split almost evenly among the training (70%), validation (20%), and test (10%) sets.

11.3.2 Pothole Datasets

[45] presented a sizable road defect dataset,[5] containing 9,053 color road images with a resolution of 600 × 600 pixels collected in Japan. The images, which show 15,435 road defects, were captured using a smartphone installed on a car in diverse weather and lighting conditions.

A comprehensive dataset[6] for instance-level pothole detection has been compiled by Rath et al. [43]. The dataset comprises a training set, a test set, and an annotation CSV file. The training set includes 2,658 color images of pothole-free roads and 1,119 color images of roads with potholes, while the test set contains 628 color images. These images, with a resolution of 2,760 × 3,680 pixels, were taken by a GoPro Hero 3+ camera, with a 0.5-s time-lapse mode, while the car moved at a 40 km/h average speed and scanned the road surface. The dataset size is approximately 2.70 GB.

The first multi-modal road pothole detection dataset[7] to facilitate the development and evaluation of more advanced and accurate pothole detection algorithms is presented in [22]. This dataset includes 55 groups of data, consisting of (1) RGB images, (2) subpixel disparity images, (3) transformed disparity images, and (4) pixel-level pothole annotations, with an image resolution of 800 × 1,312 pixels. Additionally, the Pothole-600 dataset[8] [46] provides two types of vision sensor data: color images and transformed disparity images. The transformed disparity images are obtained by applying the disparity transformation algorithm [47] to dense subpixel disparity images estimated using the stereo matching algorithm [12].

UDTIRI dataset[9] focuses on object detection, semantic segmentation, and instance segmentation tasks in the field of pothole detection. The dataset consists of images collected from online search engines as well as a collection of real images captured

[5] https://github.com/sekilab/RoadDamageDetector.

[6] https://kaggle.com/sovitrath/road-pothole-images-for-pothole-detection.

[7] https://github.com/ruirangerfan/stereo_pothole_datasets.

[8] https://sites.google.com/view/pothole-600.

[9] https://www.udtiri.com/.

in China to increase the diversity of image sources. The potholes in the UDTIRI dataset vary in scale, with the smallest pothole covering 0.3% of the image area and the largest covering 92%. On average, potholes cover 11.2% of the image area. The dataset also provides pixel-level statistics, with the average pothole bounding box having a width of 500 pixels, a height of 248 pixels, and an average area of 224,892 pixels.

11.3.3 Crack Datasets

Cha et al. proposed a crack dataset[50], which was obtained using a hand-held DSLR camera (Nikon D5200) in a complex building at the University of Manitoba. The dataset comprises 332 raw images, including 277 images with a resolution of 4,928 $\times$ 3,264 pixels and 55 images with a resolution of 5,888 $\times$ 3,584 pixels. The objects in the images were located at distances ranging from approximately 1.0 to 1.5 m, although some images were taken at distances below 0.1 m for testing purposes. Additionally, the lighting intensities in the images were significantly different. The dataset was divided into training and validation subsets, consisting of 277 images, and a testing subset consisting of 55 images.

Zhang et al. presented a dataset [51] consisting of 56,092 images of manually annotated concrete bridge decks that include cracks of varying widths, ranging from as narrow as 0.06 mm to as wide as 25 mm. The images have a size of 256 $\times$ 256 pixels and are subdivided into smaller sub-images with dimensions of 64 $\times$ 64 pixels to improve detection accuracy. From an initial pool of 4,800 images, 4,300 images are partitioned into 17,200 sub-images, and blurry images or those with corner cracks are excluded. As a result, a total of 16,789 images are used as the dataset for this study. Moreover, to evaluate the classifier's generalization ability, 500 bridge deck images are randomly combined to form 20 images with a size of 1,280 $\times$ 1,280 pixels for testing.

Choi et al. introduced a dataset [52] comprising of 200 digital images sourced from the internet or captured manually. Each image was captured under distinct conditions such as varying distances, lighting intensity, field of view (FOV), and image quality. The images have spatial dimensions between 513 and 1,920 pixels in any axis. The minimum size of the images is 513 $\times$ 513 pixels, while the maximum size is approximately that of high definition images, which is 1,920 $\times$ 1,080 pixels.

A digital single lens reflex camera was used to capture a total of 409 photos (4,032 $\times$ 3,016 pixels) in a tunnel in Huzhou, Zhejiang Province, China, under various lighting circumstances [53]. Since training convolutional neural networks (CNNs) on large images may cause GPU memory overflow, the large images were cropped to a size of 512 $\times$ 512 pixels. The resulting dataset contains 919 cropped images, which were randomly divided into a training set and a test set with a ratio of 4:1. The training set was used to optimize the model parameters, and the test set was used to evaluate the model's accuracy.

11.4 Road Defect Detection

11.4.1 Computer Vision-Based Road Defect Detection

An overview of the taxonomy of cutting-edge computer vision-based road defect detection algorithms is presented in Fig. 11.1. Classical 2-D image processing-based algorithms employ explicit programming to process road RGB or disparity/depth images, involving various techniques such as image enhancement, compression, transformation, and segmentation [55]. On the other hand, segmentation-based methods and 3-D road point cloud modeling involve fitting geometric models, such as planar or quadratic surfaces, to the road point cloud and then segmenting the cloud by comparing observed and fitted surfaces. Machine/deep learning-based algorithms utilize techniques such as image classification, object recognition, or semantic segmentation to detect road potholes. Hybrid methods combine two or more methods to increase overall performance. Recent studies have shown that deep learning approaches hold potential for road defect detection [56–58].

11.4.1.1 2-D Image Processing-Based Approaches

The classical 2-D image processing techniques have been widely studied for road defect detection, and they generally follow four fundamental steps: (1) image preprocessing, (2) image segmentation, (3) defect extraction, and (4) post-processing. Ryu et al. [59] proposed a classical 2-D image processing technique that uses a histogram shape-based thresholding technique [60] to binarize the grayscale road data. The segmented images then undergo post-processing operations to eliminate noise, including median filter and morphology operations. The median filter smooths the image while preserving edges, and morphology operation helps in reducing redundant noise. Finally, the histogram of pixel intensity is evaluated to identify the regions of interest containing road defects.

Researchers have developed various image segmentation algorithms for detecting road defects using depth/disparity images, which have led to higher accuracy in intelligent road condition assessment systems [13, 18, 61]. For instance, [13] used the Microsoft Kinect sensor to capture depth images of pavements, which were segmented using the wavelet transform algorithm [62]. Tsai et al. introduced a method [18] that utilizes a quite precise laser scanner placed on a road inspection vehicle to identify road defects via depth images. Initially, a high-pass filter is applied to the depth images to make the depth values of undamaged road pixels more similar. The processed depth images are then segmented using the watershed method [63] to detect road defects. Similarly, Fan et al. processed dense road disparity images using the disparity transformation algorithm [61], which made road defects more distinct. They then applied a histogram-based thresholding technique to the converted disparity images to identify road faults.

11.4.1.2 3-D Point Modeling and Segmentation-Based Approaches

In recent years, 3-D point clouds have shown great promise for road pothole detection and segmentation. The techniques developed for processing 3-D road point clouds generally follow two procedures [64, 65]. Firstly, the 3-D point cloud is interpolated into an explicit geometric model, typically a planar or quadratic surface, in order to obtain a structured representation of the road surface. Secondly, the 3-D point cloud is segmented by comparing it with the geometric interpolation model, which enables the identification of road potholes with greater accuracy and efficiency [64, 65].

An example of utilizing the 3-D point cloud technique for road defect detection and segmentation is shown in [64]. This method employs least-squares fitting to approximate the dense 3-D road point clouds with quadratic surfaces, and calculates the elevation difference between the observed surfaces and the fitted surfaces to extract road defects (potholes) effectively.

The accuracy of least-squares fitting will be significantly impaired by the outliers in input data [22]. To address this challenge, [66] introduced the robust least-squares approximation with bi-square weights for modeling road point clouds. Reference [47] further improved the robustness of quadratic surface fitting by employing the random sample consensus (RANSAC) algorithm [67]. In addition, [22, 68] improved the accuracy of road defect detection by introducing surface normal information into the quadratic surface fitting procedure. These techniques have demonstrated their effectiveness in accurately modeling road surfaces and detecting road defects with a high level of precision.

Compared to other approaches, 3-D point cloud modeling and segmentation methods are relatively uncommon, mainly because real-world roads are often highly irregular and uneven, making these techniques sometimes impractical.

11.4.1.3 Machine Learning-Based Approaches

The year 2012 marked a significant turning point in the field of machine learning, as neural network-based methods surpassed traditional hand-designed feature approaches in the ImageNet classification competition. Since then, the development of machine learning methods has progressed rapidly, resulting in many advanced algorithms for tasks such as image classification, object detection, and semantic segmentation, which consistently outperform traditional methods. The use of deep learning methods has become standard practice in many fields due to their ability to provide broader coverage and higher performance for solving more complex problems. In this section, we will review road defect detection applications of excellent learning-based methods, including image classification, object detection, and semantic segmentation.

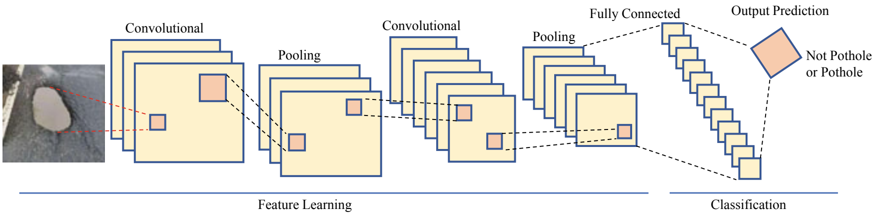

Fig. 11.5 Architecture of image classification progress [69]

Image Classification Approaches

The increasing availability of large training datasets and computational resources has facilitated the extensive use of Deep Convolutional Neural Networks (DCNNs) in detecting road defects. Compared to classical support vector machine (SVM) approaches, DCNNs can learn more abstract and hierarchical visual features, resulting in better detection performance [61]. Various DCNN-based approaches have been proposed for road defect detection [69–72], and they have been extensively studied [69, 71]. A study by Pereira et al. [69] presented a DCNN architecture consisting of four convolutional-pooling layers and a fully connected (FC) layer. The model was able to accurately classify defect and non-defect images in road data obtained from Timor-Leste. Similarly, Ye et al. [17] proposed a DCNN for road pothole classification [71]. The network consists of a pre-pooling layer, three convolutional-pooling layers, a sigmoid layer, and two fully connected layers. The pre-pooling layer is tailored to eliminate features that are irrelevant to road potholes, resulting in improved classification performance of road images. The experimental results demonstrate that the network can effectively identify road defects in different lighting conditions.

As shown in Fig. 11.5, image classification involves two steps: generating a feature map and performing feature learning through convolutional networks. The final output layer is a fully connected layer that flattens the input received from preceding layers and provides the result. Different classification networks vary in the architecture of the feature learning network used. For example, Bhatia et al. [72] proposed a DCNN based on the widely used residual network architecture [73]. Their proposed model was tested on thermal road images collected in challenging weather conditions and outperformed previous methods [59, 70, 74] in accurately classifying road images into pothole and non-pothole images.

Compared to pothole classification, road crack detection has received more attention using classification networks, mainly due to the ability of CNNs to learn discriminative visual features without explicit feature engineering [75]. For instance, [76] introduced a CNN for road crack detection that takes RGB road images as input. The learned visual features are then mapped to a scalar value using a fully connected layer, representing the probability of the input image containing road cracks. Similarly, [56] proposed a deep CNN to classify road images as containing or not

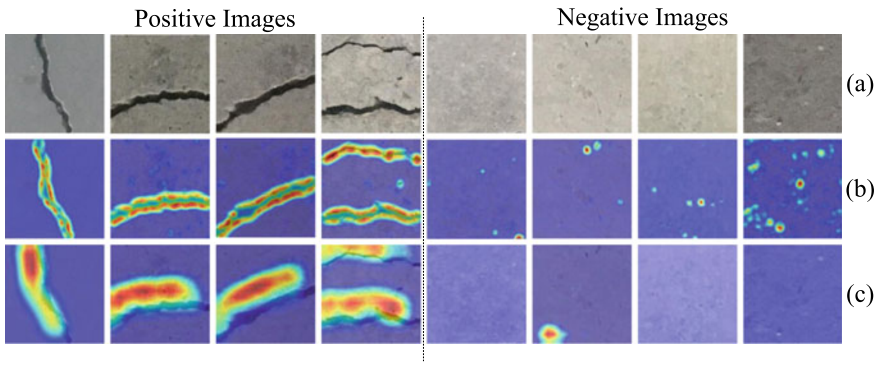

Fig. 11.6 Examples of Grad-CAM++ [79] results [80]: **a** road crack images, **b** the class activation maps of ResNet-50 [81], **c** the class activation maps of SENet [82]. The warmer the color is, the more attention the CNN pays

containing road cracks. Besides image classification, an image thresholding method was applied to segment the road images for pixel-level road crack detection.

A comprehensive evaluation of 30 SoTA CNNs was conducted by [77] for the purpose of road crack detection. The results of the experiments suggest that image classification is not a significantly challenging task for road crack detection, and the performance of deep CNNs in this area is highly similar. Furthermore, it was found that only 10,000 images are needed to train a well-performing CNN for this task. However, pre-trained CNNs often perform poorly on additional test sets, making unsupervised domain adaptation (UDA) [78] an important area of research that requires more attention. Additionally, observable artificial intelligence (AI) algorithms like Grad-CAM++ [79], have become essential for applications involving image classification to better understand CNN decision-making. Figure 11.6 shows an example of Grad-CAM++ results.

Object Detection Approaches

In 2015, Girshick introduced an advanced object detection algorithm named Fast R-CNN [83], which builds upon the R-CNN approach. This method employs CNNs to directly analyze the input image, producing a Convolutional Feature Map (CFM). Then, selective search is applied to identify region proposals from the CFM, and the proposals are reshaped to a fixed size using a Region of Interest (RoI) pooling layer before being fed into a fully connected layer. Finally, a softmax layer predicts the classes of the region proposals. However, the reliance of R-CNN [84] and Fast R-CNN [83] on selective search to generate region proposals is computationally expensive. To tackle this problem, Ren et al. presented Faster R-CNN [85], which introduces a region proposal network to learn the region proposals. Compared to R-CNN, Faster R-CNN is 250 times faster, and 11 times faster than Fast R-CNN.

Unlike the R-CNN series, the YOLO series takes a distinct approach to object detection. YOLOv1 [86] views the issue of object detection as a regression problem and divides the input image into an $S \times S$ grid, where each grid cell predicts B bounding boxes with their corresponding class probabilities and offsets. However, YOLOv1 [86] suffers from a high number of localization errors and relatively low recall. To address these issues, YOLOv2 [87] introduces several improvements, including the addition of batch normalization to all convolutional layers, fine-tuning the classification network at full resolution, and replacing the fully connected layers with anchor boxes for predicting bounding boxes.

Regarding the implementation of object detectors in road defect detection, several studies have utilized popular algorithms to detect road defects. For instance, Suong et al. [88] trained a YOLOv2 [87] model to recognize road potholes, while [89] utilized a Faster R-CNN [85] to accomplish the same goal. Reference [90] employed three kinds of different YOLOv3 [91] architectures for detecting road defects, while [92] introduced YOLOv3 [91] to detect road potholes from RGB images obtained by a car-mounted smartphone. The proposed algorithm has been implemented on a Raspberry Pi 2 model B using TensorFlow, and the reported results show a mean average precision (mAP) of 68.83% and an inference time of 10 ms. Moreover, in a recent study[10] [93], the performances of YOLOv2 [87] and Mask R-CNN [94] for road defect detection have been compared and analyzed. However, these methods use object detection networks that are not specifically designed for the task of road defect detection and are limited to providing instance-level predictions. Consequently, in recent years, semantic image segmentation has emerged as a more desirable technique for detecting road defects, as it can provide pixel-level predictions. It should be noted that Dhiman et al. (2019) compared the performances of YOLOv2 and Mask R-CNN for road defect detection, but these methods still lack the specificity of semantic image segmentation.

Semantic Segmentation Approaches

Semantic segmentation is an image processing technique for labeling each pixel in an image with a corresponding class [95]. In recent years, semantic segmentation has gained significant popularity for road defect detection [96]. Cutting-edge semantic image segmentation networks can be broadly classified into two categories based on the number of encoders used: (1) single-modal [57] segmentation networks that use only one kind of vision sensor data, such as RGB or depth images, and (2) data-fusion segmentation networks that use multiple types of vision sensor data, such as color and transformed disparity images [97] or color and surface normal information [98]. Both types of networks have been employed by researchers for the detection of road defects and defects, and they all have achieved successful outcomes.

Recent years have seen significant progress in the development of state-of-the-art solutions for semantic segmentation of road cracks. For instance, Choi et al. proposed

[10] https://vimeo.com/337886918.

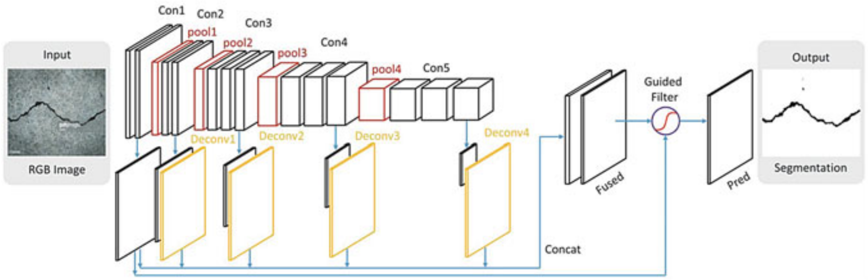

Fig. 11.7 DeepCrack architecture [58] (with permission for resue)

an approach for the segmentation of concrete cracks called SDDNet [52], which uses a combination of standard convolutions, Densely connected Separable convolution (DenSep), and a modified Atrous Spatial Pyramid Pooling (ASPP) module. Similarly, Liu et al. introduced DeepCrack [58], which is a robust CNN-based model that utilizes a deep hierarchical feature learning architecture. As illustrated in Fig. 11.7, DeepCrack avoids using fully connected layers and instead inserts side-output layers after the convolutional layers. Deep supervision is applied at every side-output layer to generate a final fused output that can capture features at different scales and levels. Finally, guided filtering is used to refine the resulting fused prediction with the first side-output layer.

In a recent study, Wang et al. [97] proposed a data-fusion convolutional neural network (CNN) for detecting road defects. The authors conducted a comprehensive evaluation of different modalities of visual features, including RGB images, disparity images, surface standard images [99], elevation images, HHA images, and transformed disparity images [61]. The experimental results showed that the transformed disparity image is the most informative visual feature for road defect detection. The implications of their findings are significant for improving the effectiveness of road defect detection techniques.

11.4.1.4 Hybrid Approaches

Hybrid approaches for road defect detection usually involve the integration of multiple algorithms mentioned earlier. By combining different methods, the system can achieve greater robustness and produce more accurate detection results.

Jog et al. proposed a hybrid approach for road pothole detection in 2012 [100]. The proposed method combines 2-D object recognition and 3-D road geometry reconstruction to accurately detect road potholes. High definition camera video frames are initially utilized for road pothole recognition, while sparse 3-D road geometry reconstruction is concurrently performed using the same video. The combination of these two modalities allows for accurate detection of road potholes and reduces

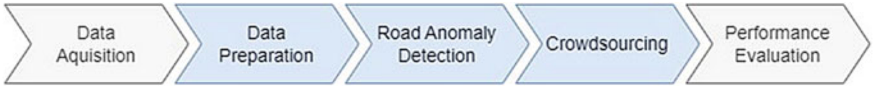

Fig. 11.8 Typical vibration-based road defect detection workflow

the number of falsely detected potholes. This hybrid method provides an effective solution for road defect detection.

In 2017, Kang et al. presented a hybrid road pothole detection system that utilizes 2-D LiDAR data and RGB road images for automated detection [101]. This system overcomes the limitations caused by electromagnetic waves and poor road conditions, and benefits from the complementary strengths of these two types of data. The 2-D LiDAR data provides road profile information, while the RGB road images offer road texture information. Two LiDARs are used to obtain accurate and large road area coverage. By combining the advantages of different vision data sources, this hybrid system is able to enhance the overall road pothole detection accuracy.

In 2019, [22] also proposed a hybrid network framework, which employs a hybrid framework that combines 2-D image processing and 3-D road surface modeling algorithms to enhance the detection of road potholes. Initially, potential undamaged road areas are extracted by applying disparity transformation (in Sect. 11.4.1.2) and Otsu's thresholding [102]. Subsequently, a quadratic surface is fitted to the original disparity image using the RANSAC algorithm to improve the robustness of surface fitting. The difference between the actual and fitted disparity images is then analyzed to detect potholes. This approach shows promising results in terms of improving road defect detection accuracy.

11.4.2 Vibration-Based Road Defect Detection

As discussed in Sect. 11.2.2, vibration-based techniques utilize non-vision sensors, such as inertial sensors, to gather vehicle data, which is then analyzed using detection algorithms to identify road defects. Vibration-based methods are generally more cost-effective than computer vision-based solutions, and have gained popularity among research groups in recent years. However, it is important to note that accurately estimating the dimensions of road defects is more challenging with vibration-based approaches compared to vision-based methods.

Typically, the workflow of road defect detection approaches follows the steps depicted in Fig. 11.8. It should be noted that the first step, i.e., data acquisition approaches, has been discussed in Sect. 11.2.2. The last step, performance evaluation, can employ any generic method and is not specific to vibration-based methods. Hence, our discussion here will concentrate on the three intermediate steps, namely, data preparation, road defect detection, and crowd-sourcing of detection results.

11.4.2.1 Data Preparation

Like other machine learning workflows, it is important to preprocess the raw data collected from mobile and built-in vehicle sensors to obtain a curated dataset before analysis. To ensure precise road defect detection, the raw sensor data typically undergoes smoothing and filtering. Additionally, if mobile sensors such as smartphones are employed, it is important to virtually re-orientate the sensor data to correspond with the vehicle's body frame.

Smoothing and Filtering

Smoothing and filtering of raw signals can be achieved by performing convolution in the time domain with finite impulse response (FIR) filters, such as a moving-average filter. Also, one can pass the raw signals through infinite impulse response (IIR) filters which are usually designed in the frequency domain. In both cases, low-pass filters (e.g., [103, 104]) and/or high-pass filters (e.g., [28]) are typically used to remove the high-frequency sensor noise and/or the low-frequency bias, respectively. Band-pass filters are frequently employed to isolate signal content within a specific frequency range. The speed signal is often utilized to accept or reject windows for further processing. Windows where the vehicle is moving slowly or not moving at all are commonly discarded to eliminate events like door slams [28]. When speed can only be inferred from GPS sensors, smoothing and estimation methods like Kalman filters are used to enhance speed estimation and remove outliers [30]. In particular, accelerometer measurements can be combined with GPS to overcome time-to-first-fix (TTFF) delays [105], and a technique called speed-dependency removal is used to segment the data. Speed-dependency removal is also addressed in [30].

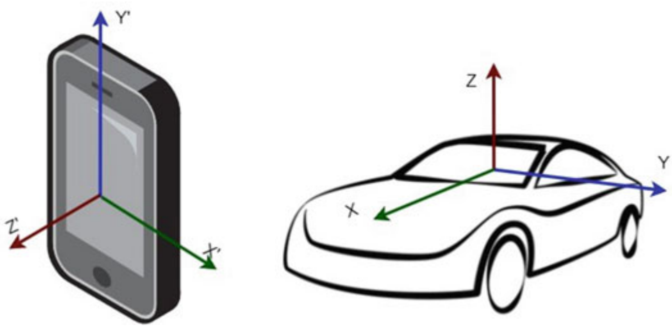

Fig. 11.9 Smartphone sensor coordinate and vehicle body coordinate

Reorientation

Another common practice in data pre-processing is reorientation, which is referred to as the process of reorienting sensor data values from a device coordinate system to the vehicle body coordinate system (see Fig. 11.9). This challenge exists in almost all smartphone-based solutions since the displacement of a smartphone in a vehicle can be arbitrary. The development of *Nericell* by Mohan et al. involved extensive investigation into this issue [29]. They realized that while it is theoretically possible to infer the angles of rotation about each axis, such a framework yields multi-factor trigonometric equations that are complex to solve. Alternatively, they resorted to an equivalent framework based on Euler angles. Under this framework, a pre-rotation about Z and a tilt angle about Y can be estimated from a short window of data collected from the steady motion of the vehicle. Then, a post-rotation angle, again about Z, can be derived from the braking and acceleration of a vehicle traveling in a straight line. These three Euler angles can easily represent any orientation, and this practical reorientation approach has been quickly adopted and extended by other research studies (e.g., [33, 106]). More reorientation strategies can be found in [107].

11.4.2.2 Road Defect Detection Methods

In the literature, there are various methodologies reported for road defect detection using vibration signals. Generally, these algorithms can be divided into two categories: signal-based and data-driven approaches. The signal-based methods aim at filtering or estimating signals with physical meanings, and identifying outliers by imposing metrics and thresholds on them, whereas the data-driven approaches focus more on extracting distinctive information from the data and training machine learning classifiers to detect road defects.

Signal and Model-Based Approaches

The researchers at CSAIL, MIT, developed *Pothole Patrol* [28] which utilized a threshold on Z-peak (vertical peak acceleration) along with longitudinal acceleration and vehicle speed to filter potential defects. They trained a simple parameterized model using five different types of manually selected road events. The system reported a 0.2% false positive rate for pothole detection. The *Nericell* system presented by Microsoft Research India utilized Windows phones carried by users to perform rich sensing. The authors used empirical calibration to determine parameters and thresholds for detecting bumps and defects based on vertical acceleration. The system was evaluated using data collected from Bangalore and Seattle, and reported promising results.

In [31], the authors explored four different metrics on accelerometer sensor data for pothole detection. The first three metrics focused on the peak, standard deviation, and changing rate of the vertical acceleration. The fourth metric, called G-ZERO, detected the moment when the acceleration measurements in all three axes are close to zero. This was based on the assumption that when a vehicle enters a pothole, it experiences a temporary free fall. It is noteworthy that reorientation was not required with the G-ZERO approach. The authors reported a high detection accuracy of up to 90% using these simple metrics.

In 2015, Li et al. analyzed pothole characteristics, and broke down the dynamics of hitting a pothole into different stages [108]. A multi-phase dynamic model was created as a switching system with three modes to capture the pothole-hitting response. They compared the results with F-Tire in a simulation environment and demonstrated that their model is capable of generating similar results and trends as in F-Tire. A detection algorithm was proposed by recursively computing the confidence of each mode with Bayesian estimation and Unscented Kalman Filtering. In [27], a state estimation framework was developed to detect road defects. These measurements from multiple vehicle state sensors were studied with both a full-vehicle model and half-model, and an optimal observer based on jump-diffusion processes (JDPs) was constructed to estimate known state and input signals, such as the road velocity input. In this work, it has also been demonstrated that the JDP-base estimator works better than a Kalman filter when jumps, such as potholes, cracks and bumps. With the estimated states, a threshold-based road defect detection algorithm was developed, and a false negative rate of 1.78% was reported.

Data-Driven Approaches

Data-driven methods involve training machine learning models using either extracted features or raw sensor signals to produce results close to labeled ground truth. A variety of signal-processing techniques can be applied to extract features from rich sensor data. In the time domain, features can be derived from statistics such as standard deviation, mean, entropy, variance, root mean square, median, integral square, and range [109]. Another method used for feature extraction in the time domain is dynamic time warping (DTW), which is primarily employed in research on speech recognition. DTW calculates the degree of similarity between two datasets by comparing incoming signal data to established templates. The similarity can then be used as a feature or compared against a threshold for detection purposes. In [33], DTW was compared to other machine learning methods such as support vector machines (SVMs), hidden-Markov models (HMMs), and general neural networks. In the frequency domain, the power spectral density (PSD) of vibration signals reveals useful insights that could be used to distinguish different road conditions.

Features can also be extracted in the wavelet domain (see, for example, [105, 110, 111]). In her PhD thesis [110], Griffiths carefully analyzed Mexican Hat, Mortlet, Haar, as well as Daubechies 6 and 10. She came to the conclusion that road vehicle vibrations may be adequately analyzed using the Mortlet wavelet as well as the Daubechies 6 and 10 wavelets. Seraj et al. utilised wavelet decomposition analysis for signal processing of inertial sensor signals from smartphones. In [111], Li et al. applied continuous wavelet transform (CWT) to extract features from sensor data and were able to estimate the length of potholes by converting wavelet scales to physical scales. Additionally, Fourier transform and wavelet transform can be used to generate virtual images from 1D sensor data, making them compatible with convolutional neural networks (CNNs) [112].

As for classification, numerous machine learning techniques have been applied to detect road defects from vibration data. In [32], classifiers such as SVM, linear regression (LR), and random forest (RF) were trained and used for prediction. They reported that RF had the best performance as compared to other classifiers, and a precision of 88.5% was achieved. In [105], SVM was utilized in a real-time multi-class detector that achieved consistent detection accuracy of approximately 90% for severe road defects, regardless of vehicle type or road location. Du et al. [103] used an improved Gaussian model for road-surface recognition and the k-nearest neighbor (kNN) method for classification. Their system operated at 400Hz for data sampling, and a detection accuracy rate of 96.03% was reported. In [104], the authors utilized a $C4.5$ decision tree algorithm to classify sensor data from smartphones for detecting road conditions such as potholes and bumps. They developed an Android application called *RoadSense* and evaluated its performance with experimental data, which showed a consistent detection rate of around 98.6%. In [113], the authors proposed a two-step method. First, a random forest filter was applied to quickly distinguish normal windows from defective windows. Then, data windows with different lengths were analyzed using DTW together with the kNN to determine the type of defects. Self Organizing Map (SOM) was applied in [36] to handle data collected from buses.

Road surface defect detection has also been approached using deep learning techniques, such as deep feed-forward networks (DFN), recurrent neural networks (RNNs), CNNs, and long-short-term memories (LSTMs). One advantage of deep learning techniques is that they can directly process the raw data without requiring any feature extraction process. However, when using CNNs for 1D signals, such as accelerometer sensor data, it is necessary to transform the data into a 2D virtual image, which can be achieved using various methods. In [114], various deep learning models such as CNN, LSTM, and reservoir computing models were examined for road surface defect detection. In [112], Baldini et al. proposed using short-time Fourier Transform (STFT) and continuous wavelet transform (CWT) on inertial sensors to generate virtual images for CNNs to classify road defects, achieving an accuracy of 97.2%. Luo et al. compared DFNs, CNNs, and RNNs in [37], using a range of sensor data, including suspension states, and found that RNNs outperformed DFNs and CNNs while using fewer model parameters.

Seraj et al. proposed a method to detect road defects by analyzing driver behaviors in [115], which is different from previous studies that focused on the detection of different types of pavements and/or types of defects. They were able to reliably detect swerves for analyzing driver behavior by studying features of the road curves and using a simple machine learning technique. Signal analysis was performed using Stationary Wavelet Transformation (SWT) decomposition. In 2017, a similar approach of detecting road defects from the behavior of vehicles was presented in [116], where the sum-of-squares (SOS) polynomial was used as a metric to evaluate the normality of vehicle behavior.

11.4.2.3 Crowd-Sourcing

When it comes to road defect detection without vision sensors, aggregating detection results from different drives or crowd-sourced data can be more challenging compared to vision-based solutions, as low-cost GPS sensors are usually the only source of location information. In [28], a distance grid-based spatial clustering algorithm was introduced to remove false positives that were unlikely to occur at the exact same location. A similar spatial clustering method was implemented in [115] to reduce smartphone GPS error, enabling precise aggregation of swerving behavior at the same location. In 2016, [117] proposed a clustering algorithm that incorporates Mahalanobis distance to quantify the similarity between newly reported and existing clusters. This method can effectively localize isolated defects and compress information for densely distributed ones.

11.5 Planning and Control

Planning and control are crucial aspects of autonomous vehicles that help ensure safe and efficient navigation. Planning refers to the process of determining a safe and efficient path for the vehicle to follow, taking into account its surroundings, environmental conditions, and any constraints such as traffic laws and regulations. The vehicle's planning system considers factors such as traffic patterns, road conditions, and the position and movement of other objects in the environment to create a map of the driving space. Based on this information, the vehicle's onboard computer generates a path that minimizes risk and optimizes efficiency and updates it continuously as the vehicle moves and its surroundings change. Control, on the other hand, refers to the processes that ensure the vehicle follows the planned path. The control system of an autonomous vehicle receives information from various sensors and actuators, such as cameras, lidar, radar, and GPS, to understand its position, speed, and orientation. Based on this information, the control system makes decisions about how to control the vehicle's movement, including steering, accelerating, and braking. The control system is also responsible for monitoring the vehicle's performance and making any necessary adjustments to keep it on track and ensure it is operating safely.

Different from other planning layers, a motion planner is a component of the planning system in an autonomous vehicle that is responsible for generating smooth and safe vehicle trajectories in the environment. The motion planner takes into account the vehicle's surroundings, including other vehicles, road conditions, and any constraints such as traffic laws, to generate a feasible and safe path for the vehicle to follow (see [118, 119], and references therein). However, the majority of recently reported motion planning techniques are 2D-based and only the vehicle's in-plane (longitudinal and lateral) motions will take into account. Previous studies on autonomous vehicles have investigated the potential of utilizing preview information of upcoming terrain, such as road grade, for optimizing their performance (e.g., [120–122]). However, these studies did not address how vertical trajectories can be planned to improve the comfort and safety of autonomous vehicles. In contrast, active suspension technology, which is a suspension system that can adjust in real-time based on road conditions and driving dynamics, has been extensively studied (e.g., [123–125]). When applied to autonomous vehicles, active suspension technology can enhance the driving experience by improving ride comfort, stability and control, safety, and sensor performance.

In [126], a new framework referred to as XYZ motion planning was introduced by Jiang et al. . It is shown that with the knowledge of road surface defects, such as potholes or large dips, as well as the ability to actively control the vertical motion of an autonomous vehicle, a 3-D path can be planned for the vehicle, and can be executed by speed, steering, and active suspension control systems. It was also shown via simulation that this new planning and control framework provides increased ride comfort and safety.

It has been reported that Tesla's 2022.20 update includes a new feature that allows a vehicle to predict rough roads and raise the suspension accordingly for Model S and Model X vehicles, using data collected from other Teslas. This feature has been seen as a possible step towards the pothole-avoiding feature that Tesla CEO Elon Musk promised in 2020 [127]. However, unlike active suspension systems, this feature is based on predictive modeling and does not involve real-time adjustments to the suspension system. In contrast, Ford developed a semi-active suspension system for their 2017 Fusion V6 Sport that has a feedback control-based system installed to lessen the discomfort experienced by drivers when traveling on a road with potholes [128].

11.6 Existing Challenges and Future Insights

In the era before the advent of deep learning in 2012, the primary research focus in the field of road perception were classical 2-D image processing-based methods. However, these explicit programming-based methods are computationally demanding and sensitive to environmental factors, particularly changes weather in and illumination conditions [13]. Additionally, the irregular shapes of road defects make it infeasible to rely on the geometric assumptions made in these methods. To overcome these limitations, 3D point cloud modeling and segmentation-based techniques were introduced in 2013, which greatly improved the accuracy of pothole detection [64]. However, these approaches have a limited field of view, assuming a single-frame 3D road point cloud forms a planar or quadratic surface. Despite efforts to enhance the robustness of road point cloud modeling, such as using the RANSAC algorithm [22], these methods require many parameters to guarantee satisfactory performance, making them challenging to apply to new scenarios.

Autonomous driving perception systems require detailed geometric information about road defects and obstacles, such as width, depth, and volume. As a result, hybrid methods that combine 3-D road reconstruction and semantic segmentation have become a promising area of research. While deep learning networks have shown excellent performance, they require large amounts of annotated data, making unsupervised or self-supervised stereo matching algorithms an attractive alternative for road surface 3-D reconstruction. Furthermore, the development of more comprehensive datasets for road perception and innovative data fusion methods, such as those proposed by [97, 98, 129], are essential for further progress in this field.

In recent years, advancements in computing power and the decreasing cost of sensing and computing technologies have led to significant progress in the research of autonomous driving systems. As a result, various companies are actively researching related technologies. Advanced driver assistance systems, such as Autopilot [130] and Super Cruise [131], integrate traffic-aware cruise control and lane-keeping assist systems to improve safety and reduce driver workload. Autopilot, as implemented in Tesla cars, adjusts the vehicle's speed to match surrounding traffic and centers the vehicle in a clearly marked lane. Similarly, Super Cruise includes adaptive cruise control that maintains a safe distance from the vehicle ahead and a lane-keeping system that keeps the vehicle centered in the lane, even around curves. Additionally, Super Cruise employs automatic emergency braking to help prevent collisions. Overall, these features enhance safety by reducing driver workload and assisting in potentially dangerous situations.

Although significant progress has been made toward fully autonomous vehicles, there is still a long way to go before driverless cars become the norm. According to a poll by Partners for Automated Vehicle Education (PAVE) [132], Americans remain skeptical of current AV technology, but are slightly more optimistic about the availability of safe autonomous driving technology in the future. Nearly three in four Americans believe that AV technology is not ready for widespread use. Additionally, 48% of Americans say they would never ride in a taxi or ride-share vehicle that was being driven autonomously. However, 58% believe that safe AVs will be available in ten years, while 20% believe that they will never be safe. The industry needs to work on gaining the trust and confidence of drivers before autonomous driving cars can become mainstream.

11.7 Conclusion

Researchers are highly engaged in road environment perception systems as autonomous driving technology advances. However, there is currently a lack of comprehensive surveys on the latest computer vision and vibration-based defect detection techniques, particularly those involving deep learning methods. We first discussed sensing technologies for acquiring data from vision-based and vibration-based sensors. Following that, we summarized several public datasets about potholes and cracks. Then, a detailed analysis of SoTA road defect detection algorithms was presented. Computer vision-based methods include (1) 2-D image processing-based approaches, (2) 3-D point modeling and segmentation, and (3) deep learning-based approaches. Road defect detection methods based on vibration were also extensively discussed. This chapter also discussed the other core competencies of autonomous vehicle software systems: Planning and Control. Finally, we envisage that future autonomous driving systems will combine vision and motion sensors.

Acknowledgements This work was supported by the National Key R&D Program of China under Grant 2020AAA0108100, the National Natural Science Foundation of China under Grant 62233013, and the Science and Technology Commission of Shanghai Municipal under Grant 22511104500.

References

1. The pothole facts. https://www.pothole.info/the-facts/
2. Accidents and injuries caused by bad road conditions. https://www.gregcolemanlaw.com/bad-road-damages-and-effects.html
3. Ma N et al (2022) Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. Transp Safety Environ 4(4):tdac026
4. Winn J, Shotton J (2006) The layout consistent random field for recognizing and segmenting partially occluded objects. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR), vol 1. IEEE, pp 37–44
5. Pandey G et al (2011) Ford campus vision and LIDAR data set. Int J Robot Res 30(13):1543–1552
6. Korosec K (2027) Toyota is betting on this startup to drive its self driving car plans forward. https://fortune.com/2017/09/27/toyota-self-driving-car-luminar/, November 2027
7. Proactive chassis controls powered by road surface fingerprinting software. https://www.clearmotion.com/roadmotion
8. Roberts J, Corke P (2020) Obstacle detection for a mining vehicle using a 2D laser. Proceedings of the Australian conference on robotics and automation 2000:185–190
9. Lakhotia A et al (2006) CajunBot: architecture and algorithms. J Field Robot 23(8):555–578
10. Mahler DS et al (1991) Pavement distress analysis using image processing techniques. Comput-Aided Civil Infrastruct Eng 6(1):1–14
11. Koutsopoulos HN, Downey A (1993) Primitive-based classification of pavement cracking images. J Transp Eng 119(3):402–418
12. Fan R et al (2018) Road surface 3D reconstruction based on dense subpixel disparity map estimation. IEEE Trans Image Process 27(6):3025–3035
13. Jahanshahi MR et al (2013) Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor. J Comput Civ Eng 27(6):743–754
14. Fan R et al (2019) Real-time dense stereo embedded in a UAV for road inspection. In: 2019 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). IEEE Computer Society, pp 535–543
15. Fan R et al (2022) Rethinking road surface 3-D reconstruction and pothole detection: from perspective transformation to disparity map segmentation. IEEE Trans Cybern 52(7):5799–5808
16. Laurent J et al (1997) Road surface inspection using laser scanners adapted for the high precision 3D measurements of large flat surfaces. In: Proceedings international conference on recent advances in 3-D digital imaging and modeling (Cat. No. 97TB100134). IEEE, pp 303–310
17. Mathavan S et al (2015) A review of three-dimensional imaging technologies for pavement distress detection and measurements. IEEE Trans Intell Transp Syst 16(5):2353–2362
18. Tsai Y, Chatterjee A (2018) Pothole detection and classification using 3d technology and watershed method. J Comput Civ Eng 32(2):04017078
19. Guo S et al (2022) Digital transformation for intelligent road condition assessment. In: Intelligent systems in digital transformation: theory and applications. Springer, pp 511–533

20. Joubert D et al (2011) Pothole tagging system. In: The robotics and mechatronics conference of South Africa, CSIR International Conference Centre, Pretoria, pp 23–25
21. Moazzam I et al (2013) Metrology and visualization of potholes using the microsoft kinect sensor. In: 16th international IEEE conference on intelligent transportation systems (ITSC 2013). IEEE, pp 1284–1291
22. Fan R et al (2019) Pothole detection based on disparity transformation and road surface modeling. IEEE Trans Image Process 29:897–908
23. Warade S et al (2012) Real-time detection and tracking with Kinect. In: International conference on computer and information technology, pp 86–89
24. Liu C-W et al (2023) Stereo matching: fundamentals, state-of-the-art, and existing challenges. Springer
25. Andrew AM (2021) Multiple view geometry in computer vision. Kybernetes
26. Soheilian B et al (2013) Detection and 3D reconstruction of traffic signs from multiple view color images. ISPRS J Photogramm Remote Sens 77:1–20
27. Li Z et al (2017) Optimal state estimation for systems driven by jump-diffusion process with application to road anomaly detection. IEEE Trans Control Syst Technol 25(5):1634–1643
28. Eriksson J et al (2008) The pothole patrol: using a mobile sensor network for road surface monitoring. In: Proceedings of the 6th international conference on mobile systems, applications, and services, Breckenridge, pp 29–39
29. Mohan P et al (2008) Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In: Proceedings of the 6th ACM conference on embedded network sensor systems, Raleigh, NC, November 2008, pp 323–336
30. Perttunen M et al (2011) Distributed road surface condition monitoring using mobile phones. In: Ubiquitous intelligence and computing. Springer, pp 64–78
31. Mednis A et al (2011) Real time pothole detection using android smartphones with accelerometers. In: 2011 international conference on distributed computing in sensor systems and workshops (DCOSS), Barcelona, Spain, 2011, pp 1–6
32. Wu C et al (2020) An automated machine-learning approach for road pothole detection using smartphone sensor data. Sensors 20(19)
33. Singh G et al (2017) Smart patrolling: an efficient road surface monitoring using smartphone sensors and crowdsourcing. Pervasive Mob Comput 40:71–88
34. Bhoraskar R et al (2012) Wolverine: traffic and road condition estimation using smartphone sensors. In: 2012 fourth international conference on communication systems and networks (COMSNETS 2012), Bangalore, India, January 2012, pp 1–6
35. Mohamed A (2015) Roadmonitor: An intelligent road surface condition monitoring system. In: Proceedings of the 7th IEEE international conference intelligent systems IS'2014, September 24–26, 2014, Warsaw, Poland, Volume 2: tools, architectures, systems, applications. Springer, pp 377–387
36. Rajput P et al (2022) Road condition monitoring using unsupervised learning based bus trajectory processing. Multimodal Transp 1(4):100041
37. Luo D et al (2020) Road anomaly detection through deep learning approaches. IEEE Access 8:117 390–117 404
38. Geiger A et al (2013) Vision meets robotics: the KITTI dataset. Int J Robot Res 32(11):1231–1237
39. Cordts M et al (2016) The cityscapes dataset for semantic urban scene understanding. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 3213–3223
40. Raj P et al (2020) Point cloud LIDAR (Toronto 3D). https://www.kaggle.com/datasets/priteshraj10/point-cloud-lidar-toronto-3d
41. Dubrovin A et al (2020) The radarscenes data set. https://www.kaggle.com/datasets/aleksandrdubrovin/the-radarscenes-data-set
42. Cordes K et al (2022) RoadSaW: a large-scale dataset for camera-based road surface and wetness estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 4440–4449

43. Rath SR (2020) Road pothole images for pothole detection. https://www.kaggle.com/datasets/sovitrath/road-pothole-images-for-pothole-detection, September 2020
44. Bombay YI (2021) Semantic segmentation datasets of indian roads. https://www.kaggle.com/datasets/eyantraiit/semantic-segmentation-datasets-of-indian-roads, November 2021
45. Maeda H et al (2018) Road damage detection using deep neural networks with images captured through a smartphone. Comput Res Repos (CoRR). arXiv:1801.09454
46. Fan R (2020) We learn better road pothole detection: from attention aggregation to adversarial domain adaptation. In: Computer vision-ECCV 2020 workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16. Springer, pp 285–300
47. Fan R et al (2018) A novel disparity transformation algorithm for road segmentation. Inf Process Lett 140:18–24
48. Zhang W et al (2014) Automatic crack detection and classification method for subway tunnel safety monitoring. Sensors 14(10):19 307–19 328
49. Medina R et al (2017) Crack detection in concrete tunnels using a gabor filter invariant to rotation. Sensors 17(7):1670
50. Cha Y et al (2017) Deep learning-based crack damage detection using convolutional neural networks. Comput-Aided Civil Infrastruct Eng 32(5):361–378
51. Maguire M et al (2018) SDNET2018: a concrete crack image dataset for machine learning applications
52. Choi W, Cha Y-J (2019) SDDNet: Real-time crack segmentation. IEEE Trans Industr Electron 67(9):8016–8025
53. Ren Y et al (2020) Image-based concrete crack detection in tunnels using deep fully convolutional networks. Constr Build Mater 234:117367
54. Chen F, Jahanshahi MR (2020) ARF-Crack: rotation invariant deep fully convolutional network for pixel-level crack detection. Mach Vis Appl 31(6):47
55. Koch C et al (2015) A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv Eng Inform 29(2):196–210
56. Fan R (2019) Road crack detection using deep convolutional neural network and adaptive thresholding. In: IEEE intelligent vehicles symposium (IV). IEEE, pp 474–479
57. Fan R et al (2022) Learning collision-free space detection from stereo images: homography matrix brings better data augmentation. IEEE/ASME Trans Mechatron 27(1):225–233
58. Liu Y et al (2019) DeepCrack: a deep hierarchical feature learning architecture for crack segmentation. Neurocomputing 338:139–153
59. Ryu S et al (2015) Image-based pothole detection system for ITS service and road management system. Math Prob Eng 2015
60. Koch C, Brilakis I (2011) Pothole detection in asphalt pavement images. Adv Eng Inform 25(3):507–515
61. Fan R, Liu M (2020) Road damage detection based on unsupervised disparity map segmentation. IEEE Trans Intell Transp Syst 21(11):4906–4911
62. Beylkin G et al (2009) Fast wavelet transforms and numerical algorithms. In: Fundamental papers in wavelet theory. Princeton University Press, pp 741–783
63. Najman L, Schmitt M (1994) Watershed of a continuous function. Signal Process 38(1):99–112
64. Zhang Z (2013) Advanced stereo vision disparity calculation and obstacle analysis for intelligent vehicles. PhD dissertation, University of Bristol
65. Wu R et al (2021) Scale-adaptive road pothole detection and tracking from 3D point clouds. In: 2021 IEEE international conference on imaging systems and techniques (IST). IEEE, pp 1–5
66. Li Y et al (2018) Road pothole detection system based on stereo vision. In: NAECON 2018-IEEE national aerospace and electronics conference. IEEE, pp 292–297
67. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
68. Ozgunalp U (2016) Vision based lane detection for intelligent vehicles. PhD dissertation, University of Bristol

69. Pereira V et al (2018) A deep learning-based approach for road pothole detection in timor leste. In: 2018 IEEE international conference on service operations and logistics, and informatics (SOLI). IEEE, pp 279–284
70. An KE et al (2018) Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving. In: 2018 IEEE international conference on consumer electronics (ICCE). IEEE, pp 1–2
71. Ye W et al (2021) Convolutional neural network for pothole detection in asphalt pavement. Road Mater Pavement Design 22(1):42–58
72. Bhatia Y et al (2022) Convolutional neural networks based potholes detection using thermal imaging. J King Saud University-Comput Inf Sci 34(3):578–588
73. He K et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
74. Hoang N (2018) An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable filter-based feature extraction. Adv Civil Eng 2018
75. LeCun Y et al (2015) Deep learning. Nature 521(7553):436–444
76. Zhang L (2016) Road crack detection using deep convolutional neural network. In: IEEE international conference on image processing (ICIP). IEEE, pp 3708–3712
77. Fan J et al (2021) Deep convolutional neural networks for road crack detection: qualitative and quantitative comparisons. In: 2021 IEEE international conference on imaging systems and techniques (IST). IEEE, pp 1–6
78. Hoffman J et al (2018) Cycada: cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning. PMLR, pp 1989–1998
79. Chattopadhay A et al (2018) Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 839–847
80. Fan R et al (2023) Computer-aided road inspection: systems and algorithms. In: Recent advances in computer vision applications using parallel processing. Springer, pp 13–39
81. He K et al (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
82. Hu J et al (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), pp 7132–7141
83. Girshick R (2015) Fast R-CNN. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 1440–1448
84. Girshick R et al (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) 2014:580–587
85. Ren S et al (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst (NeurIPS) 28:91–99
86. Redmon J et al (2016) You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) 2016:779–788
87. Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) 2017:7263–7271
88. Suong LK et al (2018) Detection of potholes using a deep convolutional neural network. J Univers Comput Sci 24(9):1244–1257
89. Wang W et al (2018) Road damage detection and classification with faster R-CNN. In: 2018 IEEE international conference on big data (big data). IEEE, pp 5220–5223
90. Ukhwah EN (2019) Asphalt pavement pothole detection using deep learning method based on YOLO neural network. In: International seminar on intelligent technology and its applications (ISITIA). IEEE, pp 35–40
91. Redmon J, Farhadi A (2018) YOLOv3: an incremental improvement. Comput Res Repos (CoRR). arXiv:1804.02767
92. Camilleri N, Gatt T (2020) Detecting road potholes using computer vision techniques. In: 2020 IEEE 16th international conference on intelligent computer communication and processing (ICCP). IEEE, pp 343–350

93. Dhiman A, Klette R (2019) Pothole detection using computer vision and learning. IEEE Trans Intell Transp Syst 21(8):3536–3550
94. He K et al (2017) Mask R-CNN. Proceedings of the IEEE international conference on computer vision (ICCV) 2017:2961–2969
95. Fan R et al (2023) Computer stereo vision for autonomous driving: Theory and algorithms. In: Recent Advances in Computer Vision Applications Using Parallel Processing. Springer International Publishing, 2023, pp 41–70
96. Yang J et al (2023) Semantic segmentation for autonomous driving. Springer
97. Wang H et al (2021) Dynamic fusion module evolves drivable area and road anomaly detection: a benchmark and algorithms. IEEE Trans Cybern 52(10):10 750–10 760
98. Fan R et al (2020) SNE-RoadSeg: incorporating surface normal information into semantic segmentation for accurate freespace detection. In: European conference on computer vision (ECCV). Springer, pp 340–356
99. Fan R et al (2021) Three-filters-to-normal: An accurate and ultrafast surface normal estimator. IEEE Robot Autom Lett 6(3):5405–5412
100. Jog G et al (2012) Pothole properties measurement through visual 2D recognition and 3d reconstruction. Comput Civil Eng 2012:553–560
101. Kang B-H, Choi S-I (2017) Pothole detection system using 2D LiDAR and camera. In: 2017 ninth international conference on ubiquitous and future networks (ICUFN). IEEE, pp 744–746
102. Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66
103. Du R et al (2020) Abnormal road surface recognition based on smartphone acceleration sensor. Sensors 20(2):451
104. Allouch A et al (2017) Roadsense: Smartphone application to estimate road conditions using accelerometer and gyroscope. IEEE Sens J 17(13):4231–4238
105. Seraj F et al (2016) RoADS: a road pavement monitoring system for anomaly detection using smart phones. Big data analytics in the social and ubiquitous context. Springer International Publishing, Cham, pp 128–146
106. Sattar S et al (2021) Developing a near real-time road surface anomaly detection approach for road surface monitoring. Measurement 185:109990
107. Carlos MR et al (2016) Evaluating reorientation strategies for accelerometer data from smartphones for ITS applications. Ubiquitous computing and ambient intelligence. Springer International Publishing, Cham, pp 407–418
108. Li Z et al (2015) Road anomaly estimation: model based pothole detection. American control conference (ACC) 2015:1315–1320
109. Martinez-Ríos EA et al (2022) A review of road surface anomaly detection and classification systems based on vibration-based techniques. Appl Sci 12(19)
110. Griffiths KR (2012) An improved method for simulation of vehicle vibration using a journey database and wavelet analysis for the pre-distribution testing of packaging. PhD dissertation, University of Bath, Bath
111. Li X et al (2019) Embracing crowdsensing: an enhanced mobile sensing solution for road anomaly detection. ISPRS Int J Geo Inf 8(9):412
112. Baldini G et al (2020) On the application of time frequency convolutional neural networks to road anomalies' identification with accelerometers and gyroscopes. Sensors 20(22):6425
113. Zheng Z et al (2020) A fused method of machine learning and dynamic time warping for road anomalies detection. IEEE Trans Intell Transp Syst 23(2):827–839
114. Varona B et al (2020) A deep learning approach to automatic road surface monitoring and pothole detection. Pers Ubiquit Comput 24:519–534
115. Seraj F et al (2015) A smartphone based method to enhance road pavement anomaly detection by analyzing the driver behavior. In: Adjunct proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2015 ACM international symposium on wearable computers, Osaka, Japan, September 2015, pp 1169–1177

116. Shu D et al (2017) A sum-of-squares polynomial approach for road anomaly detection using vehicle sensor measurements. In: Dynamic systems and control conference, vol 58288, Tysons, VA, USA, November 2017, p V002T17A004
117. Li Z et al (2016) A new clustering algorithm for processing gps-based road anomaly reports with a mahalanobis distance. IEEE Trans Intell Transp Syst 18(7):1980–1988
118. Paden B et al (2016) A survey of motion planning and control techniques for self-driving urban vehicles. IEEE Trans. Intell. Veh. 1(1):33–55
119. Claussmann L et al (2020) A review of motion planning for highway autonomous driving. IEEE Trans Intell Transp Syst 21(5):1826–1848
120. Fröberg A (2008) Efficient simulation and optimal control for vehicle propulsion. PhD dissertation, Department of Electrical Engineering, Linköping University
121. Ward JW et al (2022) A method of optimal control for class 8 vehicle platoons over hilly terrain. J Dyn Syst Meas Contr 144(1):1–18
122. Sciarretta A et al (2015) Optimal ecodriving control: energy-efficient driving of road vehicles as an optimal control problem. IEEE Comput Sci Eng 35(5):71–90
123. Tseng HE et al (2015) State of the art survey: active and semi-active suspension control. Veh Syst Dyn 53(7):1034–1062
124. Anderson Z et al (2020) Self-driving vehicle with integrated active suspension. USA Patent US10 828 953B2, Nov 10, 2020
125. Anderson ZM et al (2020) Active vehicle suspension system. USA Patent US9 702 349B2, July 11, 2017
126. Jiang Y et al (2023) On XYZ-motion planning for autonomous vehicles with active suspension systems. In: American control conference. San Diego, SD
127. Teslas can now use adaptive suspension to automatically smooth the ride over rough roads. https://www.carscoops.com/2022/07/teslas-can-now-use-adaptive-suspension-to-automatically-smooth-the-ride-over-rough-roads. Accessed: 2023-02-13
128. Technology in 2017 ford fusion helps protect against pothole damage. https://www.cbsnews.com/detroit/news/technology-in-2017-ford-fusion-helps-protect-against-pothole-damage/, Accessed: 2023-02-13
129. Wang H et al (2020) Applying surface normal information in drivable area and road anomaly detection for ground mobile robots. In: 2020 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 2706–2711
130. Souman J et al (2021) Human factors guidelines report 2: driver support systems overview. TNO Human Factors Research Institute
131. What Do I Need To Use Super Cruise? https://www.cadillac.com/ownership/vehicle-technology/super-cruise
132. PAVE Poll: Americans wary of AVs but say education and experience with technology can build trust. https://pavecampaign.org/pave-poll-americans-wary-of-avs-but-say-education-and-experience-with-technology-can-build-trust/